Bayesian models for health related quality of life data: propagating uncertainty in data from the EQ-5D-3L questionnaire

Spyridon Poulimenos

Department of Statistical Science

University College London (UCL)

This thesis is submitted for the degree of

Doctor of Philosophy

2020

Declaration:

I, Spyridon Poulimenos, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

To God Almighty

Abstract

The EQ-5D-3L is a tool that is used for the measurement and valuation of the health of populations for research. It is a self-reported questionnaire, with five questions and describes 243 distinct health states. Each state has been assigned a score (or utility score) that reflects the relative ranking. These scores can then be used to estimate quality-adjusted life-years in economic evaluations. The derived utility scores of the UK EQ-5D-3L were estimated based on a regression model using standard classical (frequentist) statistical techniques. The scores represent point estimates of the quality of life associated with each health state. These point estimates tacitly ignore the uncertainty of the estimates due to the variability inherent in the underlying data. In order to address this, the objective of this thesis is to extend the original analysis and propagate the uncertainty of the UK EQ-5D-3L scores by constructing a Bayesian model, which assigns appropriate posterior probability distributions to each of the EQ-5D-3L health states. The data used are from a 1993 UK-representative survey in which respondents evaluated EQ-5D-3L health states. A Bayesian hierarchical model is built, which accounts for model-misspecification and the responses of the survey participants in order to assign a probability distribution to the utility score of every feasible EQ-5D-3L state experienced by a group of people such as clinical trial subjects. My methods are applied on the CoBalT trial as well as on simulated data. Markov Chain Monte Carlo (MCMC) samples of simulations are derived for the utility values of the EQ-5D-3L health states. The posterior utility distributions of the EQ-5D-3L health states are summarised as approximate three-component Normal distributions using numerical optimisation and the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm. The cost-effectiveness results of the proposed model are compared with those obtained under the regular approach and differences are observed especially for the results of the simulated data. I recommend the use of the presented approach in order to properly propagate the underlying uncertainty, as otherwise an important layer of uncertainty is not taken into consideration and this can lead to the wrong inference when conducting cost-effectiveness analysis. Furthermore, this approach provides the useful advantage of doing sensitivity analysis without making any further distributional assumptions about the utility scores of the EQ-5D-3L health states experienced by the clinical trial subjects. Similar methods can be applied to the EQ-5D-3L scores of other countries and to other health instruments such as the SF-6D. Other extensions include applications in the context of model-based economic evaluations and in the area of mapping utility scores across health instruments.

Impact Statement

The work presented in this thesis has a potential impact on both academic and industrial communities. The standard procedure for doing cost-effectiveness analysis using EQ-5D-3L involves the use of the results derived under the frequentist approach. These are point estimates, but I discuss that they tacitly ignore the uncertainty of the estimates due to the variability inherent in the underlying data. This is addressed in this thesis and I show that sometimes the results of the cost-effectiveness analysis under the methods introduced in this thesis can be different than the results under the standard approach.

Health economists in the pharmaceutical industry could take into consideration the methods of this piece of work and examine the impact on the results of the particular economic evaluations which they review. Notably, there is potential for some advisory frameworks which advise healthcare provides (such as the National Institute for Health and Care Excellence in the UK) to revise their recommendations on the conduction of cost-effectiveness analysis in order to consider the applications of the presented approach and its advantages.

There have been some other pieces of work related to the utility scores of the UK EQ-5D-3L, but according to my knowledge this is the first time that there is an attempt to summarise the posterior distributions in close-form as approximate mixtures of Normal distributions. Therefore, I provide a new tariff of utility scores which can be used both in academia and in industry. Hence, researchers wishing to assess the robustness of the cost-effectiveness of an intervention could use the derived tariff of distributions, as there is no need to make any further assumptions about the aforementioned distributions. Furthermore, these distributions could be used as a known reference point which eases the situation when having to deal with different assumptions made in separate economic evaluations.

In this thesis the focus is on the UK EQ-5D-3L utility scores. One of the benefits of EQ-5D-3L is that as a generic instrument it is used extensively in economic evaluations in the UK and thus my approach could be applied to many real-world scenarios. In fact, if my approach is used then sometimes the conclusions of the economic evaluation can be reversed as a further layer of uncertainty is accounted for. Nevertheless, there is also potential for extensions to value sets of different health instruments (such as the SF-6D), and to populations of different countries or kind, since my approach is suitably designed for applications of that nature, as well as other extensions in the context of model-based economic evaluations and in the area of mapping utility scores across health instruments.

Acknowledgements

First and foremost I would like to thank God. None of what I have achieved in my life (from birth until now) would have been possible without God's help. However, nothing is impossible with God. Glory to the Father and the Son and the Holy Spirit. Amen.

I would also like to thank my principal supervisor Gianluca Baio and my secondary supervisor Jeff Round; I am grateful for the valuable experience of working with them during my PhD! Furthermore, I would like to express my gratitude and appreciation to my family (including those who passed away), colleagues and friends who have supported me in this marathon journey.

On a technical note, the kind provision of the data by the UK Data Service (Health State Valuations from the British General Public, 1993) and Nicola Wiles (CoBalT trial) is appreciated.

Spyridon (Spyros) Poulimenos

Contents

Executive summary

Health economics is concerned with balancing the finite supply of resources with near infinite demand. In jurisdictions where healthcare is publicly funded, like in the UK, the decisions of which healthcare technologies to fund given the available resources are influenced by the governmental policies. The introduction of new healthcare technologies requires assessment of both their clinical effectiveness and value for money: funding a particular intervention implies less funding on other interventions in the same or different disease areas. Economic evaluations assist in making such decisions by using a systematic approach to compare the benefits and costs between interventions. These assessments are partly based on utilities. In particular, these are values which are assigned to individual health states experienced by individuals.

The estimation of utility scores from preference based measures of health related quality of life, such as the EQ-5D-3L, is a prerequisite for the calculation of quality-adjusted life-years in cost-utility analysis, the recommended way of performing economic evaluations (NICE, 2013). The tariffs of utility scores currently used in health economic evaluations have been produced by typically using frequentist regression techniques; individuals evaluate some particular health states and then estimates can be derived for the utility scores of all attainable health states.

The UK tariffs of the EQ-5D-3L were derived by the pioneering work of the MVH group, which is based on a UK-representative sample in which respondents ranked EQ-5D-3L states in 1993 using a time-trade-off exercise. However, these tariffs give single point estimate utility scores for each health state, tacitly ignoring the uncertainty of the estimates due to the variability inherent in the underlying data.

To address this, a Bayesian-equivalent of the original MVH model is constructed; we estimate and assign appropriate posterior probability distributions to each of the health states of the EQ-5D-3L in order to correctly account for the underlying uncertainty in the estimated tariff. Furthermore, my approach also accounts for model-misspecification. The posterior distributions are summarised in closed form so that the EQ-5D-3L scores can be modelled by researchers by using these distributions instead of the point estimates.

First, MCMC samples from the posterior utility distributions are obtained, which are summarised by constructing credible intervals. It is desired to directly model these distributions in closed form but this is not immediately feasible. Nevertheless, it is possible to use the MCMC samples in order to approximate the posterior distributions as multiple-component mixtures of Normals. Under the Kullback–Leibler divergence as a measure of fitness, the best attainable approximations are achieved by obtaining the parameters of the mixture distributions as the maximum-likelihood estimates. Numerical optimisation is required for this estimation of the maximum-likelihood estimates; the Broyden–Fletcher–Goldfarb–Shanno algorithm is used.

A new tariff is obtained which is described by three-component mixtures of normals for the 243 EQ-5D-3L states. By using this strategy, we are able to map each state to a closed-form probability distribution describing the posterior uncertainty from the original data and model. These can be used to model the EQ-5D-3L in a new study, instead of the plug-in point estimates. The derived approximations of the posterior distributions as three-component mixtures of Normal distributions are quite satisfactory as we get small Kullback–Leibler divergence values.

Next, we examine the impact of incorporating this uncertainty in economic evaluations using a worked example as well as simulated data. The aforementioned application also accounts for differences between trial arms in mean baseline values. Differences are observed between the results produced under our model and those under the standard approach. The differences, which are caused due to the different mathematical properties of each method, are even more noticeable for the results of the simulated data. The use of the standard approach means that bootstrap deals with trial-sampling variation, but it does not deal with the variability of the parameters of the MVH model, unlike my approach. I recommend the use of the presented approach when conducting cost-utility analysis because if such uncertainty is ignored, then this can lead to wrong inference.

Notably, the robustness of the derived mixture distributions is further checked by also doing cost-utility analysis based on samples which are drawn from the mixture distributions instead of using the original MCMC samples of the posterior utility distributions. The similarity of these results is in favour of the derivation of the mixture distributions being robust. Moreover, the overall results are robust to changes in priors, initial values and seeds.

Potential extensions include the development of similar techniques for other health instruments or other countries' EQ-3D-3L datasets. Furthermore, a statistical package could be created to ease the access of researchers to my techniques and my findings. Other extensions include applications in the context of model-based economic evaluations and in the area of mapping utility scores across health instruments.

Chapter 1

Introduction

This chapter introduces the problem addressed in this thesis and provides an insight to the health economics background essential for this work. Our objective is to explore the current literature on valuing health states and conduct a probabilistic approach in order to model the utility values of health states by assigning probability distributions to them. First, I discuss some background information and review the concepts of quality-adjusted life-years (QALYs) and time trade-off (TTO) in relation to my work. Next, relevant previous work is presented along with the general contribution of this project, its aims and its aspirations. The literature review is chiefly based on UK work, because the first large survey to value health was undertaken there, and the UK remains a current location of pioneering relevant research. Finally, having discussed the important background information necessary to understand the concepts of this project, there is some discussion about the importance of the project and some of its potential uses.

1.1 Health economics and evaluation of health states

1.1.1 Decision-making healthcare

Health is important to both society as a whole and to people as individuals. Determinants of health include, among others: education, employment, income, and lifestyle (Marmot and Allen, 2014). Here we focus on the effect of healthcare, i.e. the improvement or maintenance of health via the diagnosis, treatment and prevention of disease and illness as delivered by medical and other practitioners.

In the UK, the National Health Service (NHS) aims at providing healthcare to individuals in order to improve, or at least retain, the health of the people. However, in the last few

decades there have been substantial changes in healthcare. The demand for it has increased and recent technological advancements have made new medical technologies feasible. However, these medical advancements usually come at increasing costs. As a result, theoretically there exist unlimited potential healthcare improvements and corresponding ways for them to be distributed, but the healthcare budget is limited. Some priorities have to be set regarding how this budget is distributed by considering the costs and the benefits of each available intervention.

Results from randomised clinical trials (RCTs) are pivotal to exploring the effects of an intervention, but they have intrinsic flaws and weaknesses and thus they can be misleading to the decision-maker as they are prone to erroneous or biased conclusions and misinterpretations (Veith, 2011). The results of a RCT are not always generalisable to different locations. Furthermore, RCTs do not automatically justify the cost-effectiveness of a treatment (i.e. if the value of an intervention justifies its cost): a treatment may be clinically effective, but that does not necessarily mean it should be paid for.

Nevertheless, mathematical models are valuable in logically connecting clinical data and assumptions with the valued costs and benefits of the intervention (Weinstein et al., 2003). In particular, their aim is to structure evidence on a logical framework by linking costs and benefits of interest in a form that can assist health-care decision-makers. This has led to the establishment of health economics, which uses a systematic approach to provide economically fitting solutions with respect to the allocation of the accessible funds.

Health economics deals with the analysis of decision making in healthcare by applying theories and principles of economics. However, patients do not have the same behaviour of consumers and doctors do not have the same behaviour as firms (Arrow, 1963). Arrow (1963) underlies that the aim of the dominant healthcare providers is not to make profit. Thus specialised theories and methods have been developed, which are applied to health economics. Most governments intervene strongly by setting regulations on the provision and charging of healthcare services; healthcare is subsidised to some degree and it is funded by different sources of taxes. For example, the UK government dominates the funding of the NHS with the allocation of a fixed budget at each time period.

In such a case, the behaviour of healthcare providers and patients (the equivalent of consumers) cannot be predicted by the interplay of market forces (Morris, Devlin and Parkin, 2007), and thus the job of decision-makers (who try to choose between different ways of resource allocation) gets difficult as the cost and benefit of each intervention is not directly known to them. *Economic evaluation* (or economic appraisal) provides information to decision-makers in the healthcare sector, such as to assist them in choosing between different uses of finite resources.

Its main tasks are to identify, measure, value, and compare the costs and benefits of the potential interventions (Drummond et al., 2005).

In the UK, the National Institute for Health and Care Excellence (NICE) makes decisions and recommendations to the NHS about which health technologies to fund, which contribute to the construction of a system of healthcare prioritisation. In particular, if NICE considers an intervention to be cost-effective (they consider that the costs are justified by the benefits achieved), then they recommend it to be funded and thus NHS is legally obliged to make it available to those people it could help. There are also similar advisory frameworks in other countries, which advise healthcare providers, such as the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia, the Canadian Agency for Drugs and Technologies in Health (CADTH) and the Institute for Quality and Efficiency in Health Care in Germany. The decisions of these frameworks are made in the context of analytic modelling.

1.1.2 Health states and health related quality of life

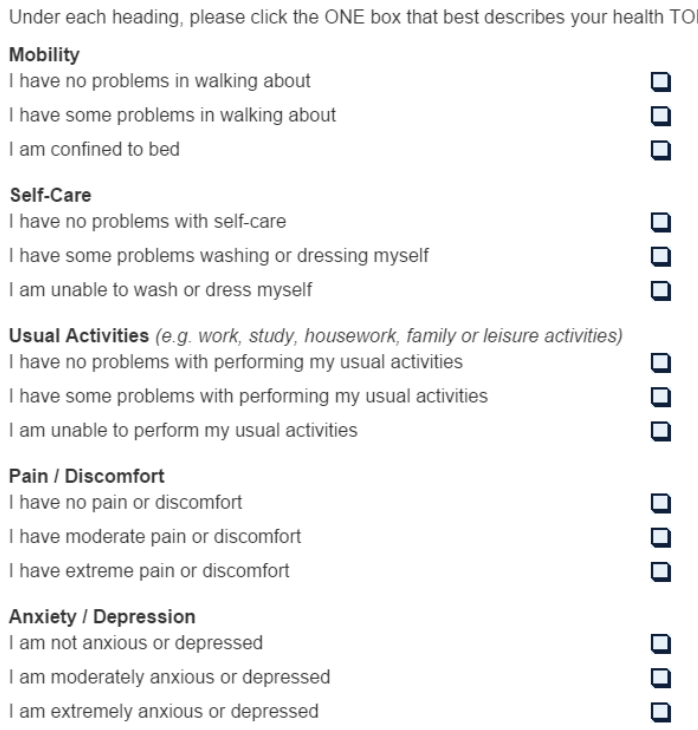
In the process of assessing the cost-effectiveness of an intervention, it is often required to identify the health states of the people in the study. A *health state* is the description of someone's quality of life with respect to their assessed level of well-being. It is within individuals' interests to have the best health state possible and to seek for potential improvements of the utility (an economics term used for the description of their well-being) and length of their life. The utility of a health state can be valued in terms of Health Related Quality of Life (HRQL), which is a quantitative measure of the impact of health on the quality of life of someone experiencing that health state. Relevant surveys can be conducted for the appraisal of the utilities of the health states of a health instrument. However, these evaluations are subjective and thus evaluations of health states are rather individual-related: there can be substantial differences between the health state evaluations of different groups of people, such as those with different experience of illness (Rosser and Watts, 1972).

HRQL scores are typically used as measures to represent the qualities of health states in economic evaluations. Clinicians are mainly interested in examining changes in the quality of life that are related only to changes in health and not to other aspects such as quality of the environment and income (Guyatt, 1993). There are many disease-specific and generic health state *instruments* (typically made up by a series of questions) which measure different dimensions of health. The quality of such a questionnaire is determined by specific criteria, such as validity (the extent to which the instrument measures the dimension which it intends to measure) and

reliability (the extent to which the instrument successfully remains stable over replications). A comprehensive review of these criteria can be found in Lohr et al. (1996). One generic instrument which complies with these criteria is the EQ-5D-3L, a health related questionnaire developed by the EuroQol Group (Brooks, 1996).

1.1.3 EQ-5D-3L

The EQ-5D-3L is a two-page self-report questionnaire: it consists of a descriptive system (illustrated in Figure 1.1) and a visual analogue scale also known as EQ VAS. The EQ-5D-3L measures HRQL for the dimensions of *mobility*, *self-care*, *usual activities*, *pain and discomfort*, *anxiety and depression* by five corresponding questions. In its most common version, each dimension is measured in terms of severity, with three levels. There are $3^5 = 243$ possible combinations of the different levels of these dimensions, and hence 243 possible health states. The EQ-5D-3L is popular because of its five-dimensional simplicity and it is used worldwide to evaluate health states (Cleemput et al., 2004); it is a generic instrument so it is easy to be used across different disease areas and it is the preferred tool of NICE for the measurement of the HRQL (NICE, 2013).


Under each heading, please click the ONE box that best describes your health TODAY.

Mobility
- I have no problems in walking about ☐
- I have some problems in walking about ☐
- I am confined to bed ☐

Self-Care
- I have no problems with self-care ☐
- I have some problems washing or dressing myself ☐
- I am unable to wash or dress myself ☐

Usual Activities *(e.g. work, study, housework, family or leisure activities)*
- I have no problems with performing my usual activities ☐
- I have some problems with performing my usual activities ☐
- I am unable to perform my usual activities ☐

Pain / Discomfort
- I have no pain or discomfort ☐
- I have moderate pain or discomfort ☐
- I have extreme pain or discomfort ☐

Anxiety / Depression
- I am not anxious or depressed ☐
- I am moderately anxious or depressed ☐
- I am extremely anxious or depressed ☐

Figure 1.1: The EQ-5D-3L self-report questionnaire (`http://www.euroqol.org`).

The five dimensions of the EQ-5D-3L (*mobility*; *self-care*; *usual activities*; *pain and discomfort*; *anxiety and depression*) are measured in terms of severity with three levels (as shown in Figure 1.1). The first level represents no health problems on that dimension, the second level means that there are some/moderate problems, and the third level represents severe problems. Therefore, a convenient way to abbreviate each health state is by the use of a five-digit number where each digit corresponds to the severity level of each of the EQ-5D-3L dimensions. For example, `32211` is the abbreviation for the EQ-5D-3L health state when there exist severe problems on the first EQ-5D-3L dimension, moderate problems on the second and the third dimensions, but not any health problems on the fourth and the fifth dimensions.

1.1.4 Valuing the EQ-5D-3L states

An individual's health can be measured by answering the questions of the EQ-5D-3L to determine their current health state. In particular, it is desirable to assign a weight to each of the EQ-5D-3L states as a means of quantifying the states' utilities compared to perfect health. The most common methods for assigning utility values to health states are *visual analogue scales* (VAS; Wewers and Lowe, 1990), *time trade-off* (TTO; Torrance, 1986), and *standard gamble* (SG; Garza, 2003); more details are contained in the following sections. These three methods attempt to elicit individuals' preferences for health states by asking them to undertake a specific task. Specifically, VAS requires the respondents to rank the health states on a thermometer-like scale, which ranges from the worst to the best imaginable health. Figure 1.2 illustrates the typical form of the visual analogue scale of the EQ-5D-3L self-report questionnaire (EQ VAS). The latter two methods attempt to mimic real life situations when individuals have to choose between outcomes and make decisions with respect to their health. SG requires the respondents to make choices between a stable non-perfect health state and a fictitious intervention which might improve their health but there is also some stated probability of resulting in death. TTO enables individuals to state their preferences with respect to trading life-time for quality of life. TTO is a useful and frequently used technique, although it should be noted that the time preference of the respondents might affect the results (Morris, Devlin and Parkin, 2007). In particular, individuals might consider that a health condition is valued differently at different time-points: one person might be more focused on their far future well-being while another person might care more for their current health.

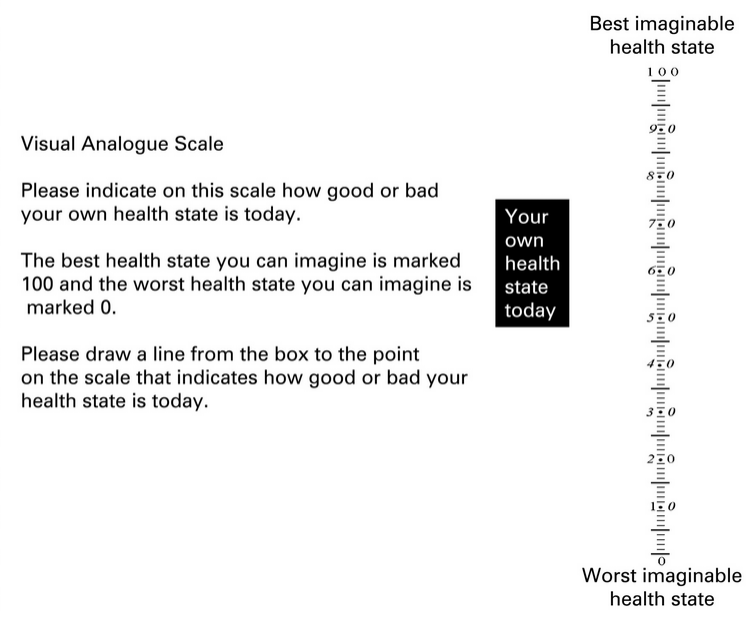


Figure 1.2: The EQ-5D-3L visual analogue scale (`http://www.euroqol.org`). Users can rank their current health state on a 0-100 scale.

1.2 Relevant UK work in the appraisal of health state indices

A health state *index* is a number representing the overall utility of a specific state in terms of HRQL. When index values have been given to all attainable states of a health instrument, then it is possible to get a clear view of how people evaluate the different states and we can make comparisons between them. A survey of 3 395 respondents organised by the Centre for Health Economics of the University of York (MVH Group, 1994; MVH Group, 1995) took place in 1993 in order to perform the evaluation of the EQ-5D-3L health states by the general UK population. The full name of the project is "The Measurement and Valuation of Health", but for the sake of simplicity it will be referred to as the "MVH project". The aim of the survey was to appraise an index value for each of the 243 health states, in addition to the states of *unconscious* and *dead*. It was the first ever national survey of this scale (in the UK and in the world) to value the EQ-5D-3L states. Survey participants had to make choices regarding some particular health states and three methods were used to obtain the indices of those health sates. The respondents had to undertake three main tasks. Fifteen health states were presented to the survey participants and the first task was to explicitly rank all of them in such an order which reflected their preference between health states. The remaining tasks were to rank these

health states on a VAS scale and then perform a TTO exercise.

Ideally, we would prefer respondents to evaluate all the possible health states. For other questionnaires with fewer total attainable states (such as the valuation matrix proposed by Rosser and Watts (1972)) there would be no problem to directly value all the health states. However, for the case of the EQ-5D-3L the number is too large, and such an aspiration is practically infeasible. As a result, the survey was conducted for the direct valuation of the indices of 45 distinct health states, where each survey participant evaluated a total of 15 states (more on this in Chapter 2).

Power calculations were performed in order to obtain the required sample size of the survey for detecting differences between the valuations given to different states. In particular, 3 235 participants were needed at the significance level of 0.05 for a power of 80% in order to detect a difference of 0.05 between the valuations given to different states. The cost was that indirect methods had to be used in order to value the utility values of the remaining states, effectively a prediction model. The final assessment of the states' utility values was undertaken using the results of a TTO exercise (Section 1.4): the TTO data and regression methods were used as a means of extrapolation (more details in Chapter 2). Notably, each health state was considered as a combination of attributes and so an additive regression model was estimated. The derived equation assigns a single number less than or equal to 1 to each health state, which represents its utility as it is valued by the UK population (Szende, Oppe and Devlin, 2007). Therefore, a value set is produced for the attainable health states of the EQ-5D-3L.

A similar value set was produced using a regression model for the data obtained from the VAS exercise. Other countries have also used regression techniques to generate health state value sets, using data from VAS based surveys (e.g. Devlin et al., 2003) and TTO based surveys (e.g. Shaw, Johnson and Coons, 2005). NICE (2013) recommends that the valuation of HRQL should be based on a choice-based method (SG or TTO), but since in the UK there has been no SG evaluation survey for the EQ-5D-3L, TTO value sets are preferred to be used for economic appraisals.

1.3 The computation of quality-adjusted life-years (QALYs)

1.3.1 The fundamentals of QALYs

There are different types of economic evaluation. Cost-utility analysis (CUA; Robinson, 1993) is a common type of economic evaluation (sometimes it is referred as part of "cost-effectiveness analysis"), where the benefits of healthcare interventions are usually measured in gained quality-adjusted life-years (QALYs). QALYs are a compound measure of HRQL and quantity of life lived. The concept of QALYs (which is generic as it includes the perception of utility) in CUA makes it possible to compare interventions of different natures. This makes CUA a popular and preferred technique over other measures of economic evaluations (Phelps, 2009). NICE (2013) underlies that CUA is preferred for economic evaluations and the benefits should be expressed in QALYs; they argue that the QALY is the preferred outcome because it is the most fitting measure which incorporates both HRQL and mortality. Thus, it is important to explore how to compute the QALYs.

The fundamental notion of QALYs is the representation of the combination of the quality (utility) and quantity (length of life) of someone's remaining life as a single index (Prieto and Sacristán, 2003). In order to compute QALYs, we have to to multiply the utility value of the individual's health state by the life-years experienced in that state. The individual's quantity of life is measured naturally in life-years. The quality of life (which is restricted to HRQL) is measured in terms of attributed utility to the health state of the individual. The derivation of the utility values assigned to each health state is conducted using techniques such as VAS, GS, and TTO (which will be examined further in the following section).

On calculating QALYs, we assume that perfect health has utility equal to 1, while death has a utility of 0. Health states worse than perfect health but better than death are assigned utilities between 0 and 1. However there are also health states which can be considered being worse than death (e.g. imagine state 33333), which have negative utilities. An intervention resulting to an additional year of life experienced in full health is worth 1 QALY (1 Utility value $\times$ 1 Year of Life = 1 QALY), which is better than an intervention resulting to an additional year experienced in half health (0.5 Utility value $\times$ 1 Year of Life = 0.5 QALYs), which is better than immediate death (0 Utility value $\times$ 0 Year of Life = 0 QALYs).

Figure 1.3 illustrates an example for the computation of the gains in QALYs of an intervention.

We assume that the QALYs of an individual's remaining life are represented by the light grey area of the graph. Initially, the individual's health state has a utility value of 0.5. He remains in this state for a year, at the end of which his quality of life reduces and it is described by another health state of 0.25 utility, in which he remains for the next two years until his death. By multiplying the utility value of each health state for the time length that he spent on it, we can deduce that his remaining life is worth of $0.5 \times 1 + 0.25 \times 2 = 1$ QALY. Suppose that a particular intervention has an immediate effect of improving the life of the individual to perfect health. After a year, he is on a new health state worth of 0.75 utility. Finally, after an additional year, his remaining life has a utility of 0.5 for the next two years until his death. This intervention improves both the individual's quality of life and extends his life expectancy. By computing, we find that the total worth of the remaining life of the individual due to this intervention is equal to 2.75 QALYs. Therefore, the gains in QALYs of this intervention are $2.75 - 1 = 1.75$ QALYs (represented by the dark grey area). A tacit assumption of this example is that we can continuously measure the quality of life of the patient; practically it is only feasible to measure the quality of life at discrete time-points.

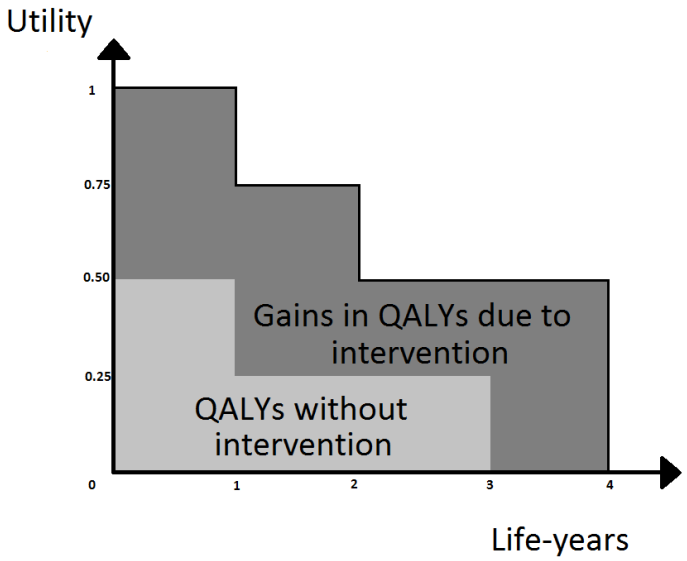


Figure 1.3: We assume that the QALYs of an individual's remaining life are represented by the light grey area of the graph. Next, we make a comparison with an imaginary intervention which is supposed to immediately improve the patient's life to perfect health (utility of 1), and then the utility of the patient is gradually reduced untill he dies after 4 years. It can be seen that the QALYs of an intervention are equal to the area under the curve representing the quality of the life of the individual. The dark grey area represents the patient's gains in QALYs due to that intervention.

When using EQ-5D-3L for the economic evaluation of a medical intervention, EQ-5D-3L is distributed to clinical trial participants at different specified time-points during the period in which the clinical trial takes place, including at its beginning (baseline) and at its end. At each of these specified time occurrences, the patients have to self-assess their current health states by answering the questionnaire. Hence, utility values can be assigned to the health states of patients at different time-points using a preference-based algorithm (which was derived using a method such as VAS, TTO or SG) with respect to the patients' answers to the questionnaire. It is then possible to compute the QALYs of that intervention by combining the information of the corresponding utility scores of the assessed health states and the time-points of the trial at which the states were experienced by the patients.

1.3.2 Computation in practice

QALYs can be calculated using individual-level data (Hunter et al., 2015). Specifically, in order to calculate the QALYs q_{ti} of an individual i $(i = 1, ..., n)$ between time-points $t-1$ and t $(t = 1, ..., T)$, we use the following formula:

$$q_{ti} = \frac{\left(u_{(t-1)i} - u_{ti}\right)}{2} \left(l_t - l_{t-1}\right), \tag{1.1}$$

where u_{ti} is the utility value of the state of individual i at time point t, and $l_t - l_{t-1}$ is the time length in years between time-points $t-1$ and t. Next, the total QALYs Q_i per individual i for the total duration T of the clinical trial are calculated by:

$$Q_i = \sum_{t=1}^{T} q_{ti}\,. \tag{1.2}$$

This method assumes that the connection between utility values of different time-points is linear, as shown in Figure 1.4. This is just an approximation because we only know the utility values at the time-points of the measurements and we do not necessarily know the exact utility values between time-points $t-1$ and t; the connection might not be perfectly linear and thus (1.2) can be considered an approximation of $Q_i = \int_{t=0}^{T} u_{ti} dt$. Finally, the total mean QALYs $\bar{Q}$ for the group of patients are calculated as follows:

$$\bar{Q} = \frac{\sum_{i=1}^{n} Q_i}{n}\,. \tag{1.3}$$

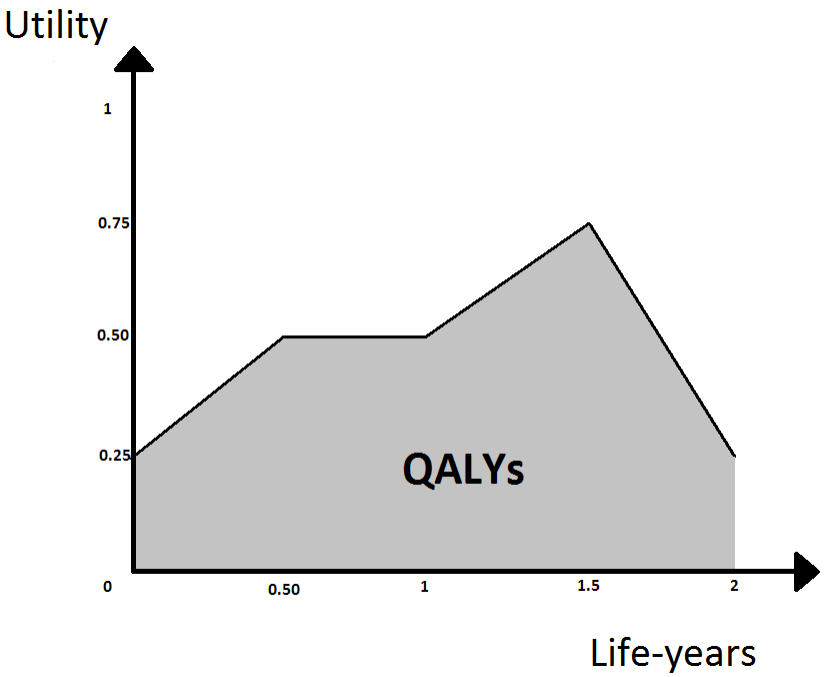


Figure 1.4: A hypothetical example for the individual-level data calculation of QALYs for a clinical data of a two-years duration. The patients had to self-assess their quality of health using the EQ-5D-3L at the beginning of the trial and then every 6 months until the end of the trial (including at its end). This Figure illustrates the QALYs (the area under the curve) of a patient whose health state initially had a utility of 0.25, 0.5 at 6 months and at 1 year, 0.75 after 1.5 years and 0.25 at the end of the trial. It can be seen that the utility scores are linearly connected.

1.3.3 QALYs criticism

QALYs are not the perfect measure of benefits. An implication of summarising the whole outcome as a single index in an economic evaluation is that some important health consequences are excluded (Phillips, 2009). For example, drug A might provide 40 additional years of life, each one worth a utility value of 0.2, and hence this intervention is worth 8 QALYs in total. Drug B might provide 13 additional life-years, each one worth a utility value of 0.6, and hence the total benefit of this intervention is worth 7.8 QALYs. It is implied that if both drugs have the same cost, drug A will always be preferred than drug B (because it worth more QALYS over B). However, it is not safe to consider that any individual in the world would consider the benefits of drug A to be better than the benefits of drug B (i.e. an intervention offering more gains in QALYs is not always to be intuitively preferred by any individual). For example, one individual could have preferred the health benefits of drug B (if the individual thinks that 40

life-years are not to be preferred if they have a utility of only 0.2).

Although not perfect measures, QALYs are considered by NICE (2013) and equivalent agencies of other countries (CADTH, 2006; PBAC, 2013) as the preferred choice among other alternatives for the measurement of benefits in CUA, which in turn is considered by NICE as the preferred technique for economic evaluations. QALYs have the useful property that they represent the benefits of an intervention by a single number. This makes possible the comparisons of interventions across different medical areas for which it is otherwise infeasible to directly compare the physical units of the benefits. Some researchers have highlighted some of the imperfections of QALYs and their calculation methods, and they have proposed their suggestions to deal with such problems (Nord et al., 1999; Prieto and Sacristán, 2003; ECHOUTCOME, 2013). For example, one notable case has been the European Consortium in Healthcare Outcomes and Cost-Benefit Research (ECHOUTCOME), which recommended the abandonment of the use of QALYs as an outcome in healthcare decision making in favour of epidemiological outcomes (ECHOUTCOME, 2013). Sir Andrew Dillon, the chief executive of NICE since 1999 has argued that the research of the ECHOUTCOME project was "rather limited" (Hill, 2013). Furthermore, he admitted that "Economists will argue about the precision of the QALY methods and it's not perfect," but that QALY is "based on solid research and uses a way of measuring how quality of life changes when using different treatments which is the best we have available."

This thesis also aspires to minimise the imperfections of the use of QALYs in CUA for economic evaluations: by assigning distributions to the indeces of health states, it is possible to conduct additional sensitivity analysis with respect to the gained QALYs of a medical intervention. Instead of assuming that the utility value of a state is always equal to a fixed number, we can sample different values (from the distributions of the indeces of the health states), and hence each time the gained QALYs of the intervention have a different total value. It is of interest to examine whether the results of an economic evaluation are sensitive to QALYs changes.

We have shown how the QALYs are calculated, and discussed some of the criticisms of the QALYs in the literature. Next, I discuss how the quality estimates used to calculate QALYs are derived, particularly in relation to the EQ-5D-3L.

1.4 Time trade-off (TTO)

1.4.1 Using TTO to assign utility values to health states

VAS values are calculated explicitly, because the respondents directly state their preferences with respect to a health state. In contrast to VAS, TTO (Torrance, 1986) is a technique which assigns health state utility values in an implicit way, using respondents' responses with respect to decision dilemmas. There are several TTO variations, but the focus here is on the variation used by the MVH project.

Consider the situation where the utility u_A of health state A is already known (e.g. if health state A is full health, then its utility value is equal to 1), and we are interested in using this information in order to value the utility u_B of state B. There are two different ways to achieve this: one is for health states considered better than death (death is assumed to have a utility value of zero) and one for health states worse than death (hence these states are regarded to have negative utility values). Initially, the respondent should state whether they consider this health state to be better or worse than death and then use the corresponding method.

First we consider the case where we only deal with states better than death. We could ask respondents if they prefer the option to spend l_B life-years in imperfect state B, or if they prefer the option to spend $l_A < l_B$ life-years in state A, which is perfect health. This can be viewed as a trade-off between quality and time. The main assumption of TTO is that, between these two options, the respondents will select the one that provides them the highest value of QALYs.

Assume that Q_A is equal to the QALYs produced by the first option, while Q_B represents the QALYs of the second option. When the respondent is indifferent between states A and B, it means that $Q_A = u_A \times l_A = u_B \times l_B = Q_B$, and so

$$u_B = \frac{u_A \times l_A}{l_B}. \tag{1.4}$$

Since all the values on the right hand side of equation (1.4) are known, we can compute the value of u_B. Figure (1.5) provides a graphical representation of this case: the QALYs of the two different scenarios are assumed to be equivalent so we use the available information in order to calculate the (currently unknown) utility of health state B.

In practice, respondents are asked to state how many l_A life-years they believe that one should spend in perfect health state A in order to be in an equivalent situation to someone who spent l_B (fixed) years on state B. The fewer years they believe that one should spend in perfect health, the worse the utility value of the state under evaluation is considered. In the MVH

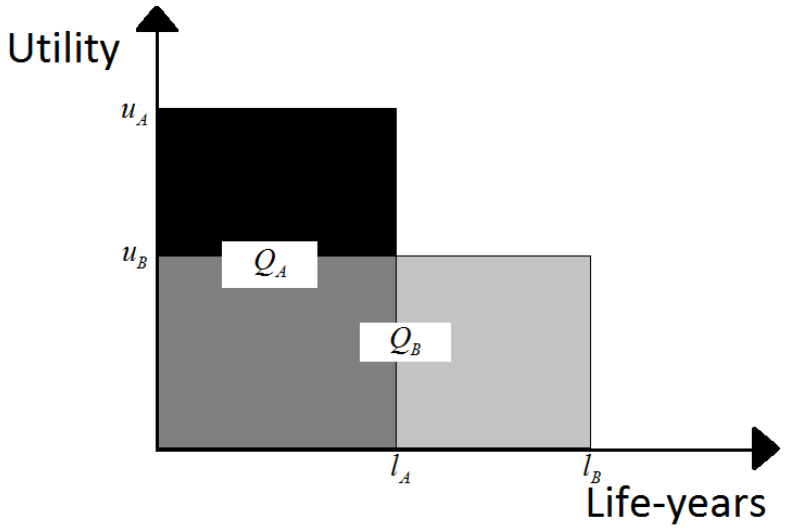


Figure 1.5: A time trade-off case for the valuation of a health state B which is considered to be better than death. The QALYs of state A (Q_A) are in black colour while the QALYs of state B (Q_B) are in light grey. Dark grey represents the area which is covered concurrently by black (QALYs of state A) and light grey (QALYs of state B) colour. The two rectangles representing the QALYs of states A and B have the same area.

project, respondents were asked to imagine that the state presented to them would last for 10 years. This is equivalent to considering that $l_B = 10$ in equation (1.4).

For the case when the health states are worse than death, individuals are asked to choose between the option of immediate death, and the option of living l_A life-years on health state B (which is worse than death, and hence it is associated with a negative utility) followed by $l_B - l_A$ years in perfect health, i.e. state A (Gudex, 1994).

Immediate death is equivalent to zero QALYs. As a result, if the individual considers these two options indifferent, then the QALYs of the second option should be equal to zero as well. Therefore, $0 = u_B \times l_A + u_A \times (l_B - l_A)$, or equivalently:

$$u_B = \frac{-u_A \times (l_B - l_A)}{l_A}. \tag{1.5}$$

Practically, in order to assign a utility value to a state which is regarded to be worse than death, respondents are asked to specify how many l_A life-years they can sustain living in this worse than death state B, which will be followed by a compensation of experiencing $l_B - l_A$ life-years in perfect health state A (where l_B is a fixed number). The more life-years required to experience in health state A as a compensation of experiencing the previous life-years in the worse than death state, the less the utility value of that state. In the MVH, respondents were asked to state the equivalent life-years they can sustain in the health state under evaluation,

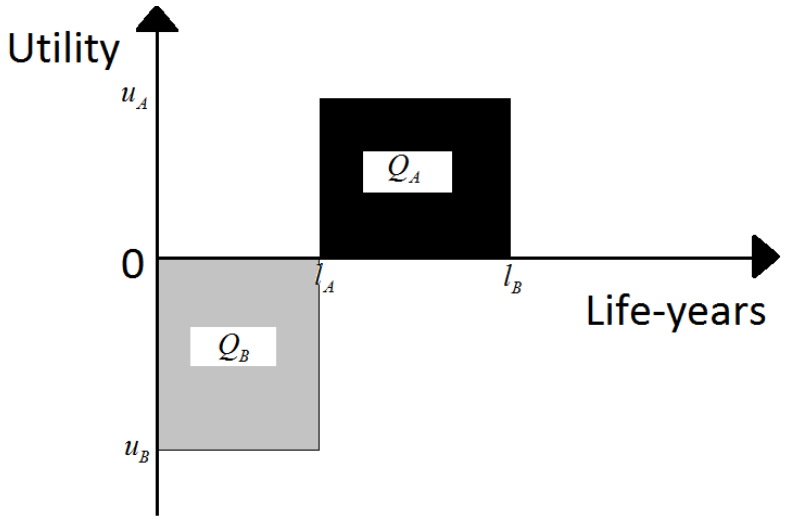


Figure 1.6: A time trade-off case for the valuation of a health state B which is considered to be worse than death. The QALYs of state A (Q_A) are in black colour while the QALYs of state B (Q_B) are in light grey.

knowing that these years will be followed by experiencing perfect health for 10 years minus this stated number of life-years. This is illustrated in Figure (1.6)

Although states superior than death are associated with utility values within the interval of $(0, 1]$, there is no lower bound for the negative utility values of health states worse than death. This asymmetry can be dealt with a suitable transformation after which the new values are within the interval of $[-1, 1]$, such as the one used by Patrick et al. (1994). The same transformation was used by the MVH project. In particular, the new value u^* is derived by the following equation:

$$u^* = \begin{cases} u, & \text{if } u \geq 0 \\ \frac{u}{1-u}, & \text{if u} < 0 \end{cases}, \tag{1.6}$$

where u is the raw value derived by the TTO survey.

1.4.2 TTO in comparison to VAS and SG

It should be noted that, similar to the VAS case, TTO values are also derived under conditions of certainty (Parkin and Devlin, 2006): unlike in SG exercises where the participants have to make decisions about uncertain events, TTO participants are told to imagine that the states in question will certainly hold for the specified time-length. This can be considered by some researchers as disadvantage similar to the VAS criticism by Drummond et al. (2005). Moreover, there have been some TTO surveys which generated counter-intuitive results. For

example, Dolan and Gudex (1995) found that the same imperfect health states were valued less by individuals when they were experienced for one month than when they were experienced for 10 years; they suggest that TTO surveys should only be used for chronic diseases with a hypothetical duration of at least 5 years because it may be deemed unreasonable to present respondents with scenarios in which they will be dead in a matter of months or even weeks.

Another factor that might affect the results of TTO surveys is the *time preference* of respondents, i.e. whether individuals prefer to face a potential burden now or in the future (Morris, Devlin and Parkin, 2007). Respondents have positive time preference when they prefer to have good health now and to delay the event of pain to be experienced in the future. In contrast, negative time preference is when respondents prefer to experience the undesirable event of bad health as soon as possible so that they can experience good health afterwards. The former case could exist when respondents believe that a cure could be found by the time that they experience the unpleasant event, while the latter could exist in cases where individuals want to face the pain quickly to avoid the experience of dread when awaiting for it (Dolan and Gudex, 1995). The time preference of respondents might add bias to the TTO values which should not be ignored; there have been some appealing relevant methods to correct the raw TTO values but this is not always practical because the required methods are usually impotent to perform the task (Attema et al., 2013).

NICE (2013) advises that health states indeces values should be derived using a choice-based method, i.e. a method where the respondents express their relative preference of a state compared to other states, such as TTO and SG. The MVH project has derived value sets for the valuation of the EQ-5D-3L states by UK general population using both VAS and TTO methods. My work takes into account this recommendation of NICE and hence we are more focused on the TTO data of the MVH.

SG is an alternative choice-based method, which is more complicated to use than TTO (Torrance, 1986); its results could be quite different than when using TTO (Martin et al., 2000), while sometimes it produces results similar to TTO (Torrance, 1976). The main difference with TTO is that the derivation of SG values is conducted under uncertainty conditions. The respondents are given props such as the one illustrated in Figure (1.7). Each prop represents a hypothetical scenario where respondents have to choose between two alternative options (one of them being certain, the other one being uncertain), with three outcomes in total, all of which for the same fixed length of time, after which the individual will (hypothetically) die. The MVH project did not include any SG assessment for the valuation of the EQ-5D-3L health states. It would be interesting to examine SG value sets by considering potential UK health

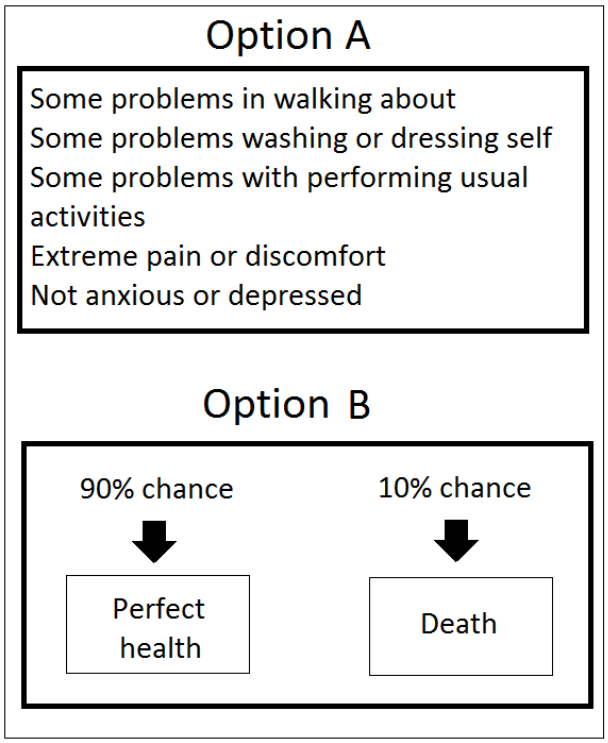


Figure 1.7: A prop of a standard gamble case for the valuation of a health state which is considered to be better than death.

states valuation surveys to include SG methods as well. This would allow the investigation of values which were generated under uncertainty.

1.5 The purpose of the project

The resulting value set for the EQ-5D-3L states computed by the York team is used in economic appraisals in the UK for the computation of QALYs. An index value is provided as a representative point estimate of the utility of the corresponding health state. This would have been a reasonable estimate if the data (i.e. the observed valuations of a state's utility value) were normally distributed. Nevertheless, the data appear to be far from normally distributed: there are signs of skewness and sometimes of multimodality. For example, Figure 1.8 illustrates a kernel density plot of the observed utility evaluations for state `11133`, which appear to be far from being normally distributed. In addition, the point estimates do not reflect the underlying variability in the estimates, as illustrated by the distribution in Figure 1.8. As a result, I believe that the information that has been obtained by a point estimate is not as useful as the description of a full probability distribution would be. Acquiring the mean value for the index of a state is helpful; however, merit should be given to individual evaluations which deviate

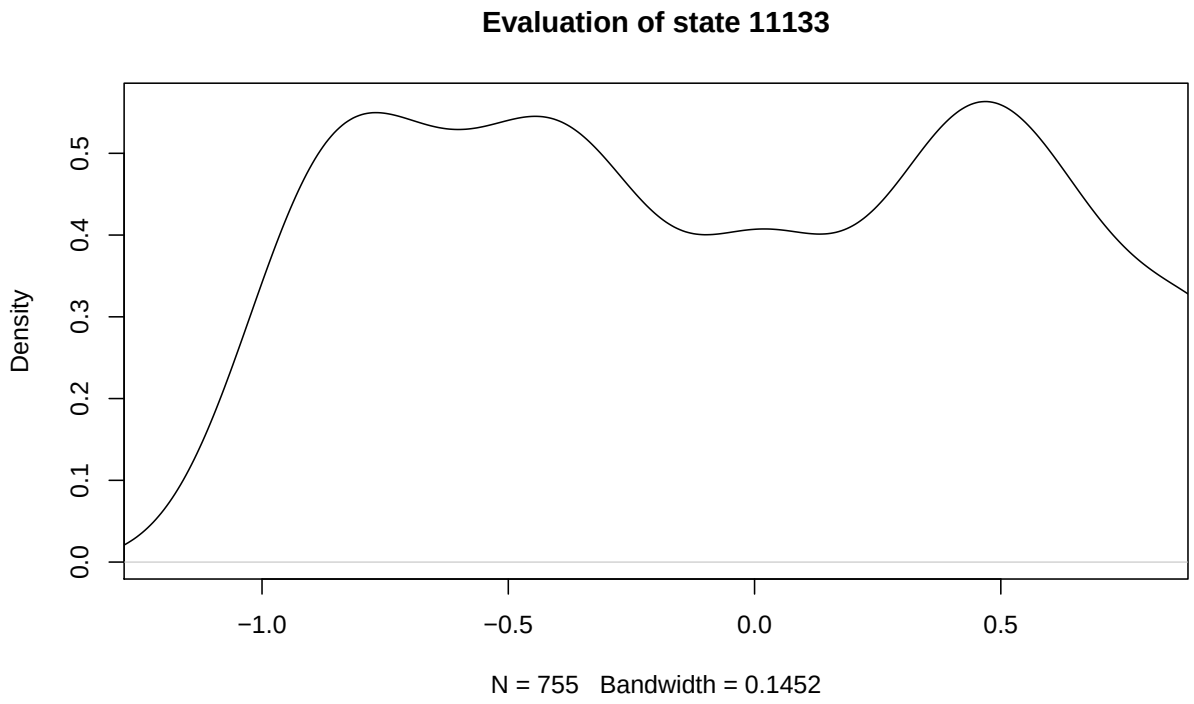


Figure 1.8: Kernel density plot of the observed evaluated utilities by 755 MVH survey participants for state 11133. Visual examination may suggest that the observations do not originate from a univariate normal distribution. Note: Observing positive values of probability outside of $[-1, 1]$ is just an artefact of using Kernel density estimation over $[-1, 1]$; however, the topic of having utility values outside of $[-1, 1]$ is dealt with later in the thesis.

from the mean. It would be better if this within-state variability could be expressed in the form of a particular distribution. Furthermore, as we will see later on, Bayesian credible intervals of distributions are intuitively easy to understand and interpret (Gelman et al., 2013; Bernardo and Smith, 1994). It would be interesting to examine the shape of the distribution of each health state index to receive informative knowledge about the overall state valuation by the UK population.

Previous work on the USA EQ-5D-3L value set from a Bayesian perspective has shown the importance of quantifying the uncertainty of the utility values (Pullenayegum, Chan and Xie, 2015). They found that if uncertainty is not accounted for, then the conventional standard error of the utility value of an EQ-5D-3L state substantially underestimates the value of the true standard error. Furthermore, some researchers have worked on similar health instruments, such as the SF-6D (Brazier, Roberts and Deverill, 2002); they focused on using the available valuation data in a more efficient way to produce value sets which are subject to less param-

eter uncertainty (Kharroubi, Brazier and O'Hagan, 2007; Kharroubi et al., 2007; Kharroubi, O'Hagan and Brazier, 2005). Nevertheless, there have not been any attempts yet in reporting the utility distributions of the distinct EQ-5D-3L states as explicitly specified probability distributions.

The aim of this report is to use the data obtained by the York team in the 1993 survey and extend the original MVH framework by constructing a Bayesian model which propagates uncertainty in the original evaluation through the estimated utility values. Specifically, I construct a Bayesian regression model preserving the original principles of the MVH project. Thus, the objective is to update the point estimate values by estimating and assigning appropriate probability distributions. Reporting these results will allow researchers to model each EQ-5D-3L state by using closed form probability functions instead of the point estimates which do not give information about the existing uncertainty. With the help of MCMC techniques (Gelman et al., 2013) simulations are derived from the posterior distributions of the utilities of the 243 EQ-5D-3L states. These distributions are first reported as credible intervals and kernel density plots. However, it is not feasible to directly obtain these distributions in closed forms. Therefore, in order to achieve the initial objective of reporting the posterior distributions in a closed form, we have to approximate them. This is done by using numerical optimisation techniques which provide satisfactory approximations of the underlying utility distributions as mixtures of appropriate distributions. Thus, a new tariff is produced by mapping the utility scores of the EQ-5D-3L health states to corresponding closed form probability functions.

1.6 Potential uses of the results

The results of this thesis are important in a variety of ways. First, the distributions of health state indeces could be used as an alternative descriptive summary of health states for a statistical analysis.

The calculation of HRQL is a prerequisite for the computation of QALYs. Nevertheless, if the indeces of the EQ-5D-3L states are modelled by distributions, then each distribution draw provides an alternative value set. It is intuitive to deduce that if an economic evaluation is performed with EQ-5D-3L value sets different than the ones which are currently used in economic appraisals, then the results might be in favour of a different decision.

Moreover, the results of this thesis will have an impact on sensitivity analysis. The robustness of the output of an economic evaluation is subject to model uncertainty which is examined by investigating how sensitive the results are to changes in the model. Sensitivity analysis examines

modelling uncertainty (e.g. changes to the structure of the model) and parameter uncertainty (e.g. changes to the values of the parameters, or assuming that the observations originate from some particular distributions). The more times the decision of an economic evaluation remains unchanged due to model changes, the more validity is given to the results of the evaluation. As a result of providing distributions for the indices of health states, these distributions can be incorporated to sensitivity analysis. Therefore, the results of an economic appraisal can be assessed on their sensitivity to the utility values' assumptions of the gained QALYs. In particular, I apply my approach on data from a clinical trial as well as some simulated datasets and investigate their effect in terms of sensitivity analysis.

This project focuses on the EQ-5D-3L value sets of the UK general population. However, this methodology can also be extended to value sets of different health instruments, and to populations of different countries or kind. The EQ-5D-5L is another version of the EQ-5D-3L, each of its dimensions is measured with five levels: no problems, slight problems, moderate problems, severe problems, extreme problems. Using five dimensions instead of three implies that there is more flexibility in distinguishing between different states (as there are more levels of severity), but the total number of attainable states increases to $5^5 = 3125$. Recently there have also been attempts to value the states of the EQ-5D-5L version (Feng et al., 2016), but here we focus on the EQ-5D-3L (the three-dimensions version), which we could also merely refer to as EQ-5D, as its value sets have been used in economic evaluations for several years and I try to investigate the impact of my methods on such appraisals. Nevertheless, a similar methodology could be applied to the EQ-5D-5L.

Version 4.0.3 of the statistical package `R` (R Core Team, 2020) and version 4.3.0 of the statistical package `JAGS` (Plummer, 2017) have been used for the implementation of this project. Lunn et al. (2012) provide an in-depth tutorial for running `JAGS` models. Relevant `JAGS` code is provided in the Appendix.

1.7 Summary

Health instruments are used to measure the quality of life with the use of methods such as TTO and VAS. Relevant work includes a UK survey of 3395 respondents which was undertaken in 1993 in order to evaluate the health states of the EQ-5D-3L. Due to the large number of the possible health states of the questionnaire, it was not feasible to directly value all of them. The health state indices are computed by a regression function. The aim of this project is to model the state indices as suitable posterior probability distributions. The resulting data

can be used to investigate alternative decisions for CUA and for further sensitivity analysis to examine whether the model is sensitive to changes in the values of gained QALYs. The methodology can be extended to other populations and questionnaires. Pullenayegum, Chan and Xie (2015) have demonstrated the importance of quantifying the uncertainty of the utility values of the EQ-5D states, whereas other researchers (Kharroubi et al., 2007; Kharroubi, O'Hagan and Brazier, 2005; Kharroubi, Brazier and O'Hagan, 2007) have focused on producing value sets which are subject to less parameter uncertainty.

The computation of QALYs takes into consideration the utility values of health states. In economic evaluations, QALYs are NICE's preferred tool for measuring the benefits of an intervention while TTO is the preferred way for the elicitation of health states utilities.

Drummond and McGuire (2010) and Briggs, Claxton and Sculpher (2011) describe the principles and details of economic evaluations. The problem in this thesis is addressed in a Bayesian framework. Spiegelhalter, Abrams and Myles (2011) and Baio (2013) advocate the use of Bayesian methods as a useful tool in the area of health economic evaluation.

Chapter 2

"The Measurement and Valuation of Health" (MVH) project

This chapter describes how the survey for the "The Measurement and Valuation of Health" (MVH) project was designed: how the survey was conducted and which health states were chosen to be valued by respondents. We also focus on the methodology used, the regression function for the valuation of the EQ-5D-3L states and the computed value set.

2.1 The framework of the MVH

The aim of the MVH was to assign utility values to the EQ-5D-3L health states, by taking into consideration the preferences of the UK general population. This implies that the general population's preferences are taken into account when the EQ-5D-3L is used for economic evaluations. The objective was achieved by conducting a survey which took place in the summer of 1993, following a pilot survey. This was also followed by a retest survey, which concluded that the overall survey responses were reliable. The aforementioned data of the MVH project can be obtained from the UK Data Service (Gudex, Dolan, Williams and Kind, 1995).
In particular, the survey participants were chosen in order to geographically represent England, Wales and Scotland. The selection of the participants was made in a three-stage approach (Erens, 1994), which finally produced 6080 addresses. An adult was randomly chosen at each of the addresses. In the end, the interviewers achieved a total of 3395 interviews, which were more than the targeted sample of 3235 individuals (as discussed in section 1.2).
The interviews were conducted as follows. Initially, the respondents were asked to rate their health using the EQ-VAS. Next, the three main tasks of each individual were about the eval-

uation of some particular EQ-5D-3L states. In particular, the first main task was that the individual had to directly rank some states in such an order that reflected their preference. Then, for the two following tasks the individuals were asked to evaluate those states using the VAS and the TTO method. Each of these three main tasks was undertaken for a predefined set of 15 health states for each survey participant; 45 health states were evaluated in total. Each individual was told that the 15 health states presented to them should be considered to last for 10 years with no changes, followed by immediate death. The final part of the survey was the collection of personal background data from the respondents.

The collected background data were compared with the 1991 Census and the 1992 General Household Survey (GHS). The comparison implied that the survey achieved a good representativeness of the UK general population in terms of geographical coverage, sex, age, education, tenure, social classes, accommodation types, economic positions, and social status.

The responses by some participants were excluded from further analysis, using stringent exclusion criteria with respect to missing data and logical inconsistency (e.g. if a state with severe problems on all dimensions was valued to be better than a state with moderate problems on all dimensions). However, the percentages of the final exclusions for each section of the survey were very low; for example, for the case of the TTO, the responses of only 1.7% of the survey participants were excluded. It was decided that respondents with any missing VAS or TTO data would be excluded. As a result, the final dataset consisted of the responses of 2997 individuals.

2.2 The selection of the evaluated health states

Section 1.2 discussed the limitations of directly appraising all the states of the EQ-5D-3L; a decision was taken for appraising 45 health states in the overall survey of the MVH and each individual would evaluate 15 of these states. For the sake of simplicity, in order to discuss the selections of these states, numerical abbreviations are used for the EQ-5D-3L states.

The MVH researchers chose to select 45 states to be valued by the survey respondents. They aimed to select states across different levels of severity and dimensions, as well as all possible combinations of levels across the dimensions. However, one concern among their considerations was that they should not choose states which sounded unrealistic, because individuals would have difficulties to perceive the existence of such states and value them. For example, it is not intuitive to imagine the existence of state `33133`, since it is difficult to conceive that someone who is confined to bed, unable to wash or dress, has extreme pain and is extremely anxious has

no problems with performing usual activities such as working, studying or doing housework. Furthermore, the "unconscious" and "dead" could not be estimated by the results of other states, since they are not in the 243 combinations of different severity levels across dimensions. As a result, these two states had to be among those chosen to be valued by all the respondents. Moreover, it was decided that states 11111 and 33333 would act as reference points among all survey participants and hence they had to be valued by everyone as well. Another decision taken by the researchers was that everyone had to value approximately 2 of the 5 mildest states (i.e. 21111, 12111, 11211, 11121, and 11112). The remaining 36 states were chosen to be three groups with 12 states each: one with 12 mild states, one with 12 states of moderate severity, and one of 12 severe states, in order to keep a balance among different severity levels. Precisely, for the selection of the rest of the 9 states which were valued by each individual: from each of these three groups, three states were randomly chosen, as shown in Table 2.1.

Group		
Mild states	**Moderate states**	**Severe states**
12211	13212	33232
11133	32331	23232
22121	13311	23321
12121	22122	13332
22112	12222	22233
11122	21323	22323
11312	32211	32223
21312	12223	32232
21222	22331	33321
21133	21232	33323
11113	32313	23313
11131	22222	33212

Table 2.1: The MVH survey participants had to value the states "unconscious", "dead", 11111 and 33333. They also had to value two states among 21111, 12111, 11211, 11121, and 11112. In addition to them, each respondent had to value an additional of 9 states from these 36 shown in this table, categorised in three groups in terms of severity. Each respondent had to value 3 mild states, 3 moderate states and 3 severe states, which were randomly selected by the survey conductors from the first, second, and third groups respectively.

2.3 The methodological framework of the MVH

With the exclusion of "unconscious" and "dead", all of the other health states are a combination of different levels of the five dimensions of the instrument. Furthermore, state `11111` can be assumed to have a utility score of one. Therefore, we focus on the remaining 42 states (out of the original 45 which were chosen to be directly evaluated); each survey participant evaluated 12 of these states. It is feasible to use the information of these 42 directly valued states in order to value all EQ-5D-3L states, including those which were not included in the MVH survey.
In particular, a regression model can estimate the mean utility value of a health state (response variable), taking into account the characteristics of that health state (covariates), such as how severe the problems are (if any) of each health dimension. Individual data were used for the calculation of the regression function for the estimation of the means of 243 EQ-5D-3L states. It was decided to perform regression with fixed and random effects in which the functional form was additive. Moreover, it is reasonable to assume that observations (valued states) of each individual are related to each other (i.e. an individual who assigned a much lower than the population average utility value to one state has probably assigned low utility values to other states as well). Hence, a random effect term was used to incorporate this concept. More specifically, the regression to estimate the utility values was based on 35,964 (= 12 states per each of the 2,997 respondents) observations. The model used for the utility value y_{pq} of the q-th ($q = 1, ..., 12$) evaluation of the p-th ($p = 1, ..., 2997$) individual is as follows:

$$1 - y_{pq} = \boldsymbol{X}_{pq}^{T}\boldsymbol{\beta} + \omega_p + \varepsilon_{pq}, \tag{2.1}$$

$$\omega_p \sim N\left(0, \sigma_{\omega}^{2}\right),$$

$$\varepsilon_{pq} \sim \mathrm{N}\left(0, \sigma_{\varepsilon}^{2}\right),$$

where $1 - y_{pq}$ is the corresponding disutility, $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_0, \beta_1, ..., \beta_D)$, where the β coefficients (of size D) are the parameters of interest, $\boldsymbol{X}_{pq}^{T} = (V_{1pq}, V_{2pq}, V_{3pq}, ..., V_{Dpq})$, where the elements of $\boldsymbol{X}_{pq}$ are variables (which take specific values according to their definitions) associated with the q-th evaluated health state of the p-th individual. This concept of the correlation of evaluations by the same individual is incorporated by the addition of the ω_p term, which is the random effect associated with the deviation of the p-th respondent from the overall intercept. Furthermore, ε_{pq} is an error term associated with that particular observation. Notably, both the ω_p's and the ε_{pq}'s are independent and identically distributed; the ω_p's are independent of the ε_{pq}'s.
After estimating the $\boldsymbol{\beta}$ regression coefficients, $\hat{\boldsymbol{\beta}}$ can be used to estimate the utility score of a

person (such as a clinical trial subject) experiencing a particular health state at a given time point. For clinical trial data, the utility score u_{ijt} of the EQ-5D-3L health state experienced by the i-th subject ($i = 1, ..., n_j$) of the j-th group of the clinical trial ($j = 1, ..., J$) at time point t ($t = 1, ..., T$) is calculated as:

$$u_{ijt} = 1 - \boldsymbol{X}_{ijt}^{*T}\hat{\boldsymbol{\beta}}, \tag{2.2}$$

where $\boldsymbol{X}_{ijt}^{*}$ is the column vector with the variables corresponding to the state experienced by the i-th subject of the j-th group of the clinical trial at time point t. Notably, we can see that the lack of an intercept term in equation (2.1) means that if the model is structured in this way then the V variables can be defined in such a fashion that they are all equal to 0 when the state under evaluation is the state of perfect health. As a result, this provides the flexibility to force the expectation of state's `11111` disutility to be 0 (in other words the expectation of state's `11111` utility to be 1) in accordance with the fundamental assumptions about the state of perfect health which were discussed earlier.

The size D of variables V (and corresponding β coefficients) as well as the definitions the V variables are dependent to the nature of each candidate regression model. There are many plausible models which can be tested, but it is sensible to define the response variables using the ordinal nature of the descriptive system of the EQ-5D-3L (see section 1.1). For example, we may consider a dummy variable which takes the value 1 if the health state has at least one dimension at level 3 and otherwise it takes the value 0.

The different candidate regression models were judged and compared. For each model, the researchers considered: its goodness-of-fit; its parsimony (how simple the model is); its logical consistency (e.g. the estimated value of state `22223` should be lower than the estimated value of state `22222`, because it is not logical to expect the reverse). As it has already been discussed, the raw TTO data are asymmetrical since the upper limit is unity whereas there exist negative values less than -1. The work of Patrick et al. (1994) implies that in such a case, the estimated values would be highly affected by some few respondents with very negative utility values. This asymmetry was dealt with by applying equation (1.6) to modify the data in order to constrain the resulting utility values between -1 and 1.

2.4 Estimating the regression model

The researchers of the MVH project were mostly interested in the value set derived using the TTO method since they found evidence that it performed better than VAS (Dolan et al., 1996), such as in terms of logical consistency. Hence, we focus on the coefficients of the TTO model

for the estimation of the mean utilities of the EQ-5D-3L states.

The chosen model is an additive one with some dummy variables V. Specifically, there is variable ALL which takes value 1 if there there is at least one dimension at level 2 or 3 (i.e. for any state other than perfect health). Similarly, the variable N_3 (often referred to as the "N3 term") takes value 1 if there is at least one dimension at level 3. Furthermore, for each dimension there is a variable which takes the value 0 if the dimension is at level 1, the value 1 if the dimension is at level 2, otherwise it takes the value 2 if the dimension is at level 3 (MO,SC,UA,PD,AD). Finally, there is one dummy variable for each dimension; these variables take the value 1 if the dimension is at the third level, otherwise it takes the value 0 ($MO2$,$SC2$,$UA2$,$PD2$,$AD2$). The derived regression for the computation of the utility u_s of the s-th state among the 243 aforementioned EQ-5D-3L health states is as follows:

$$\begin{aligned} u_s = 1 &- 0.081 \times ALL_s - 0.069 \times MO_s - 0.104 \times SC_s - 0.036 \times UA_s \\ &- 0.123 \times PD_s - 0.071 \times AD_s - 0.176 \times MO2_s - 0.006 \times SC2_s \\ &- 0.022 \times UA2_s - 0.140 \times PD2_s - 0.094 \times AD2_s - 0.269 \times N_{3s}. \end{aligned} \tag{2.3}$$

This model has an adjusted R^2 of 0.46, which, given the type of the cross-sectional data, is considered high enough (Dolan, 1997). Nevertheless, equation (2.3) is not the one commonly found in the literature; the utility value of a health state is usually reported using the more simple equation (2.4), which consists of only (more easily interpreted) twelve dummy variables. The coefficient of M_3 is derived by the sum of the coefficient of $MO2$ and twice the coefficient of MO. Similar reasoning applies to the computation of the coefficients of S_3,U_3,P_3,A_3. Thus, (2.3) can be expressed alternatively as follows:

$$\begin{aligned} u_s = 1 &- 0.081 \times ALL_s - 0.269 \times N_{3s} - 0.069 \times M_{2s} - 0.314 \times M_{3s} \\ &- 0.104 \times S_{2s} - 0.214 \times S_{3s} - 0.036 \times U_{2s} - 0.094 \times U_{3s} \\ &- 0.123 \times P_{2s} - 0.386 \times P_{3s} - 0.071 \times A_{2s} - 0.236 \times A_{3s}. \end{aligned} \tag{2.4}$$

The York team used version 6.0 of the `LIMDEP` software (Greene, 1992) to run the regression and derive the results. I used `R` and in particular the Restricted Maximum Likelihood (REML) method within the `lmer` function (Bates, Maechler, Bolker and Walker, 2020) in order to replicate the analysis of the MVH study and the same results were obtained.

The definitions and the possible values of the dummy variables of equation (2.4) are summarised in Table 2.2. For example, state `12113` is assumed to have a utility value of 0.31 (1-0.081-0.269-0.104-0.236=0.31). Furthermore, we can double check that in accordance with the definitions of the variables given in Table 2.2, equation (2.4) gives 1 as the expected utility value of state

11111. Variable ALL is equal to 1 for all states except the state of perfect health (in which case it takes the value of 0). Therefore, we could interpret the corresponding coefficient 0.081 as the intercept which represents any move away from full health.
Finally, the state of unconscious is not a combination of different levels of the five dimensions of the instrument. In particular, it is assumed to have a utility value of -0.402, which is the mean of the utility values assigned to this state by the survey respondents.

Variable	**Values**	
ALL	1 if there is at least one dimension at level 2 or 3	0 otherwise
N_3	1 if there is at least one dimension at level 3	0 otherwise
M_2	1 if the mobility dimension is at level 2	0 otherwise
M_3	1 if the mobility dimension is at level 3	0 otherwise
S_2	1 if the self-care dimension is at level 2	0 otherwise
S_3	1 if the self-care dimension is at level 3	0 otherwise
U_2	1 if the usual activities dimension is at level 2	0 otherwise
U_3	1 if the usual activities dimension is at level 3	0 otherwise
P_2	1 if the pain/discomfort dimension is at level 2	0 otherwise
P_3	1 if the pain/discomfort dimension is at level 3	0 otherwise
A_2	1 if the anxiety/depression dimension is at level 2	0 otherwise
A_3	1 if the anxiety/depression dimension is at level 3	0 otherwise

Table 2.2: The definition of the variables of equation (2.4).

2.5 Summary

This Chapter described the design of the MVH project and the statistical methods used to derive the results. Following the exclusion of some observations, the responses of 2997 respondents entered the final analysis of the data. A total of 45 health states were directly valued by the survey participants while each one of them assigned utility values to a set of 15 states. The choice of these 45 states made it possible to derive utility estimates for all the EQ-5D-3L states using regression techniques. The derived model was an additive one with 12 dummy variables. Finally, although not all the concepts of the MVH are ideal and some assumptions are not perfectly backed-up, we can take its core principles for granted to extend its original framework as shown in detail in the next chapter.

Chapter 3

The derivation of the distributions of the EQ-5D-3L utility scores

This chapter reviews the motivation of my approach, the methods used, the derivation and discussion of the obtained results. First, a less-technical description of the motivation to my approach is presented. Next, the Bayesian model is constructed and we obtain MCMC samples from the posterior distributions. Since it is not feasible to directly model the posterior distributions in closed forms, I discuss how to approximate them using mixtures. Numerical optimisation techniques are used to estimate the parameters of the mixture distributions. The criterion used to examine the quality of the approximate distributions is the Kullback–Leibler divergence; I illustrate how to empirically compute it in order to assess the obtained mixture distributions. My core objective is achieved by presenting the new tariff of the EQ-5D-3L scores by mapping each state to a parametric form of a three-component mixture of Normal distributions.

3.1 Motivation

The primary motivation of this project was not to criticize the methodology of the MVH project, but, taking its core principles for granted, to extend its framework from a Bayesian point of view paying particular attention to the uncertainty of the utility values. Therefore, the main assumptions in this project are in agreement with those of the MVH project.
We stick to the concept of perfect health having a utility value of 1 while the state of dead has a utility of 0 (as well as using the transformation of (1.6) to ensure that the utility values will be between -1 and 1) and we try to evaluate the remaining indeces of the EQ-5D-3L

using the MVH data, where participants directly evaluated the utility values of a sub-set of the EQ-5D-3L states. An important assumption of the MVH regression techniques is that the utility values assigned to a specific state by different survey participants come from a normal distribution. This does not seem as a perfectly-backed assumption given that the distributions of most of the 45 states under evaluation in the survey seem to be multi-modal (e.g. as seen in Figure 1.8). Following the same principles, however, the underlying objective is to build and extend a Bayesian equivalent of equation (2.4).

Furthermore, the MVH model is a nonsaturated one (i.e. there are fewer parameters than the number of health states which are evaluated); although a model with fewer parameters generally estimates the parameters more precisely, there is a risk of model misspecification. Pullenayegum, Chan and Xie (2015) have quantified the extent of model misspecification in the US-model of EQ-5D-3L and they concluded that model misspecification is responsible for a substantial portion of observed prediction errors. It is essential to account for the presence of important model misspecification and thus I handle this as an extension to the framework of the model by considering an additional (model misspecification) term.

3.2 Bayesian modelling

3.2.1 Model structure and Bayesian methods

The problem in this thesis is addressed in a Bayesian context; in section 1.5 I discussed the rationale of using Bayesian methods and the associated benefits. For the construction of the Bayesian model, I use the same MVH survey responses and formulate a Bayesian model in order to make posterior inference about the distributions of the EQ-5D-3L state utilities. The model is written as:

$$1 - y_{pq} \sim \mathrm{N}\left(\mu_{pq}, \sigma_{\varepsilon}^2\right),$$

$$\mu_{pq} = \boldsymbol{X}_{pq}^T \boldsymbol{\beta} + \omega_p + \xi_{pq},$$

$$\begin{cases} \xi_{pq} = 0 & \text{, for state } \texttt{11111}, \\ \xi_{pq} \sim N\left(0, \sigma_{\xi}^2\right) & \text{, otherwise,} \end{cases}$$

$$\omega_p \sim N\left(0, \sigma_{\omega}^2\right),$$

where y_{pq} is the utility value of the q-th EQ-5D-3L state which was evaluated by the p-th survey respondent. Notably, ξ_{pq} (which refers to the q-th EQ-5D-3L state which was evaluated by the p-th survey respondent) is the term which accounts for model misspecification. The

values of the twelve dummy variables of $\boldsymbol{X}_{pq}^{T}$ (which refer to the q-th EQ-5D-3L state which was evaluated by the p-th survey respondent) are in accordance with the definitions given in Table 2.2.

The priors of the regression coefficients, the unknown within-state variance and the random effect variance are set as follows:

$$\beta_d \sim N\left(0, 10\right), \text{ for } d = 0, 1,$$

$$\beta_d \sim N\left(0, 1\right), \text{ for } d = 2, 3, ..., 11,$$

$$\sigma_\xi \sim U\left(0, 1\right),$$

$$\sigma_\varepsilon \sim U\left(0, 1\right),$$

$$\sigma_\omega \sim U\left(0, 1\right).$$

The motivation for using vague uniform priors for the dispersion σ parameters is "to let the data speak for themselves". Furthermore, β coefficients are centred around 0 as without prior knowledge we do not know if we expect the coefficients to be negative or positive. The prior variance of the β's associated with all indicator variables except ALL and $N3$ is chosen to be 1 to reflect the prior uncertainty about the coefficients which are related to one dimension of the EQ-5D-3L. The prior variance of the coefficients of the ALL and $N3$ variables, which are related to multiple EQ-5D-3L dimensions is chosen to be 10 because of the wider uncertainty. The values of the β coefficients are used to compute the EQ-5D-3L utility scores deterministically; since the values of the utility scores are expected to be between -1 and 1, the assigned prior distributions for the β coefficients do not provide strong prior information. Moreover, my priors are in agreement with those used in other research work related to EQ-5D-3L modelling (Pullenayegum, Chan and Xie, 2015).

However, we could also use different priors to do sensitivity analysis and examine the robustness of the model. A different choice of priors for the dispersion σ parameters is as follows:

$$log(\sigma_\xi) \sim N\left(0, 10^6\right),$$

$$log(\sigma_\varepsilon) \sim N\left(0, 10^6\right),$$

$$log(\sigma_\omega) \sim N\left(0, 10^6\right).$$

This time Normal distributions are assigned to the natural logarithms of the σ parameters, where the corresponding standard deviations 10^6 are quite large. However, we could also consider the following priors:

$$1/\sigma_\xi^2 \sim Gamma\left(0.001, 0.001\right),$$

$$1/\sigma_{\varepsilon}^2 \sim Gamma\,(0.001, 0.001)\,,$$

$$1/\sigma_{\omega}^2 \sim Gamma\,(0.001, 0.001)\,.$$

The motivation behind this is that this prior should be "similar" to the improper distribution $1/\sigma^2 \sim Gamma\,(0,0)$,but the prior $1/\sigma^2 \sim Gamma\,(0.001, 0.001)$ actually favours small values of the standard deviation σ.

In terms of alternative priors for the β coefficients, we could use different Normal distributions with even larger standard deviations. In particular, the β priors which we could use for the conduction of sensitivity analysis are as follows:

$$\beta_d \sim N\left(0, 10^6\right), \text{ for } d = 0, 1, ..., 11.$$

These distributions are even less-informative compared to the original choice of prior distributions.

Next, the calculation of the utility score u_s of the s-th distinct EQ-5D-3L state is done as follows:

$$\begin{aligned} u_s = 1 - \beta_0 \times ALL_s - \beta_1 \times N_{3s} - \beta_{2s} \times M_{2s} - \beta_3 \times M_{3s} - \beta_4 \times S_{2s} \\ -\beta_5 \times S_{3s} - \beta_6 \times U_{2s} - \beta_7 \times U_{3s} - \beta_8 \times P_{2s} - \beta_9 \times P_{3s} \\ -\beta_{10} \times A_{2s} - \beta_{11} \times A_{3s} - \xi_s^{new}, \end{aligned} \tag{3.1}$$

where ALL_s, N_{3s}, M_{2s}, M_{3s}, S_{2s}, S_{3s}, U_{2s}, U_{3s}, P_{2s}, P_{3s}, A_{2s}, A_{3s} are the dummy variables corresponding to the s-th EQ-5D-3L state in accordance with the definitions given in Table 2.2. Furthermore, $\xi_s^{new} = 0$ if s corresponds to state `11111`, otherwise $\xi_s^{new} \sim N\left(0, \sigma_{\xi}^2\right)$; so that the mean of the utilities is $\boldsymbol{X}_s^T\boldsymbol{\beta}$, but with a variance that better reflects the true uncertainty in the data.

In Bayesian analysis, many of the distributions which we try to compute are not analytically tractable. Therefore we can instead simulate the random variable and obtain a sample of values originating from that variable, using Markov Chain Monte Carlo (MCMC) techniques (Gilks, Richardson and Spiegelhalter, 1996; Gamerman, 1997; Robert and Casella, 2010; Gelman et al., 2013). Using `JAGS`, MCMC simulations are obtained from the posterior distributions of the utility values of the EQ-5D-3L states (including those which were not directly evaluated by the MVH survey participants) of the Bayesian model. Kernel density plots (Sheather, 2004), which are a generalisation and smooth improvement of the underlying histograms of these simulations provide a visual inspection of the form of the posterior distributions, using the MCMC samples.

For each β and ξ^{new}, MCMC initially generates a sample of size C: $\beta^{(1)}, \beta^{(2)}, ..., \beta^{(C)}$ and

$\xi^{new(1)}, \xi^{new(2)}, ..., \xi^{new(C)}$ in respect. Then, being a deterministic function of $\boldsymbol{\beta}$'s and $\xi^{(new)}$'s (equation (3.1)), we can deterministically derive the sample of size C for the utility u_s of the s-th EQ-5D-3L state: $u_s^{(1)}, u_s^{(2)}, ..., u_s^{(C)}$. Furthermore, for clinical trial data, the utility value $u_{ijt}^{(c)}$ of the state experienced by a specific individual at a given point for the c-th ($c = 1, ..., C$) MCMC iteration is computed as

$$u_{ijt}^{(c)} = 1 - \boldsymbol{X}_{ijt}^{*T} \boldsymbol{\beta}^{(c)} - \xi_{ijt}^{new(c)},$$

where $\boldsymbol{\beta}^{(c)}$ is the vector of the β coefficients with the values which were computed under the c-th MCMC iteration and $\xi_{ijt}^{new(c)}$ is the misspecification term (which refers to the state of the i-th subject of the j-th group of the clinical trial at time point t) the value of which was computed under the c-th MCMC iteration.

3.2.2 Results

The MCMC algorithm is run using `R` and `JAGS`. The `R2jags` package (Su and Yajima, 2020) is also used. Initially, the Raftery and Lewis's diagnostic (Raftery and Lewis, 1992) is considered from a pilot MCMC run: the number of iterations required in order to estimate the 24th permille to within an accuracy of +/− 0.005 with probability 0.95 is less than 4,000. In particular, the decision is to run two MCMC chains; the first 4,000 iterations of each chain are discarded as burn-in, and then a further $C = 4{,}000$ iterations are used in total for making inference on the posterior distributions of the parameters of interest. Table 3.1 provides a summary of posterior statistics of the β's. In order to assess potential lack of convergence of the MCMC run, I calculate the Geweke (1992) statistic, the Gelman and Rubin (1992) statistic, and the effective sample size. Table 3.1 illustrates for each of the β's the corresponding p-value of the Geweke statistic, the Gelman and Rubin statistic (also known as $\hat{R}$) and the effective sample size (which is calculated using the `R2jags` package (Su and Yajima, 2020); the values are rounded to the nearest 100). Large p-values of the Geweke statistic suggest no evidence of non-convergence; no evidence was found in favour of non-convergence at $\alpha = 5\%$. The main idea of the Gelman and Rubin statistic is to compare the between-chains to the within-chain variation for each scalar component of the vector of parameters of interest. If the value of the statistic is large (as a rule of thumb if it is greater than 1.1), then this suggests that there is no-convergence; all the computed $\hat{R}$ values are very small and thus there is no evidence of non-convergence. The effective sample size represents the number of equivalent independent observations that are associated with the actual sample size of the MCMC run and thus it is lower than or equal to the post-burn-in MCMC sample size. Equivalently, the higher the decay of the autocorrelation

Parameter	MVH	Mean	S.D.	2.5%	97.5%	p.Geweke	$\hat{R}$	E.S.S.
$\boldsymbol{\beta}_0$	0.0810	0.0921	0.0260	0.0419	0.1443	0.1925	1.0008	4000
$\boldsymbol{\beta}_1$	0.2690	0.2324	0.0342	0.1628	0.3002	0.3140	1.0008	4000
$\boldsymbol{\beta}_2$	0.0690	0.0658	0.0221	0.0195	0.1102	0.1478	1.0009	4000
$\boldsymbol{\beta}_3$	0.3140	0.3281	0.0280	0.2737	0.3846	0.1567	1.0008	4000
$\boldsymbol{\beta}_4$	0.1040	0.0970	0.0237	0.0512	0.1440	0.8263	1.0010	4000
$\boldsymbol{\beta}_5$	0.2140	0.2259	0.0286	0.1696	0.2832	0.6336	1.0007	4000
$\boldsymbol{\beta}_6$	0.0360	0.0357	0.0272	-0.0163	0.0885	0.8113	1.0008	4000
$\boldsymbol{\beta}_7$	0.0940	0.1083	0.0318	0.0456	0.1709	0.9496	1.0008	4000
$\boldsymbol{\beta}_8$	0.1230	0.1199	0.0231	0.0751	0.1646	0.0633	1.0011	3900
$\boldsymbol{\beta}_9$	0.3860	0.4072	0.0255	0.3569	0.4587	0.3238	1.0009	4000
$\boldsymbol{\beta}_{10}$	0.0710	0.0664	0.0235	0.0197	0.1132	0.9478	1.0008	4000
$\boldsymbol{\beta}_{11}$	0.2360	0.2568	0.0273	0.2041	0.3103	0.8424	1.0009	4000

Table 3.1: For each β: the values assigned to it in the MVH report; its posterior mean, standard deviation (S.D.), 25th permille, and 975th permille, as well as the p-value of the Geweke statistic, the Gelman and Rubin statistic and the effective sample size (E.S.S.), the values of which are rounded to the nearest 100.

with the increased number of simulations used, the higher the effective sample size; the values of the computed effective sample sizes are high and thus there is no suggestion of having auto-correlation issues. The aforementioned diagnostics did not find evidence of non-convergence; thus the assumption is that the 4,000 iterations are considered satisfactory enough for the model to converge.

Furthermore, the main interest is to summarise the distributions of the utilities of the distinct 243 EQ-5D-3L states. Figures 3.1, 3.2 and 3.3 illustrate 95% credible intervals for these posterior distributions. The expected utility of each state based on the original MVH project can also be seen as a small circle. The states are ordered with respect to their posterior median utility. Furthermore, Table 3.2 summarises the MCMC samples (the means, the 25th and the 975th permilles) of the posterior utility distributions. The fact that the MVH point estimate and the mean from the Bayesian model are almost always in agreement and the credible intervals have a sensible range (including the MVH point estimate) indicate that the results are in line with our expectations of building a sensible model.

	MVH mean	posterior mean	2.5%	97.5%
11112	0.84800	0.84050	0.71386	0.96483
11121	0.79600	0.78902	0.66010	0.91315
11211	0.88300	0.86961	0.73981	1.00000
12111	0.81500	0.81107	0.68794	0.93484
21111	0.85000	0.84401	0.71420	0.97135
11113	0.41400	0.41859	0.29279	0.54325
11131	0.26400	0.26983	0.14558	0.39803
11311	0.55600	0.56662	0.43672	0.69601
13111	0.43600	0.45045	0.31184	0.59087
31111	0.33600	0.34795	0.20755	0.48423
11122	0.72500	0.72210	0.59827	0.85106
11212	0.81200	0.80656	0.67933	0.93246
12112	0.74400	0.74556	0.61644	0.87488
21112	0.77900	0.77611	0.64954	0.90368
21121	0.72700	0.72308	0.59920	0.84974
12121	0.69200	0.69056	0.56627	0.81998
11221	0.76000	0.75363	0.61902	0.88850
12211	0.77900	0.77423	0.64321	0.90252
21211	0.81400	0.80594	0.67079	0.93979
22111	0.74600	0.74564	0.61864	0.87167
11123	0.29100	0.29868	0.17046	0.42578
11132	0.19300	0.20279	0.06906	0.33227
11312	0.48500	0.50098	0.37022	0.63462
13112	0.36500	0.38350	0.24279	0.52633
31112	0.26500	0.27950	0.14013	0.41893
31121	0.21300	0.22783	0.08565	0.37091
13121	0.31300	0.32995	0.18348	0.47234
11321	0.43300	0.44783	0.31279	0.57528
11231	0.22800	0.23305	0.10435	0.36365
11213	0.37800	0.38233	0.25534	0.51301
12113	0.31000	0.32174	0.19577	0.46105
12131	0.16000	0.17065	0.03544	0.30431
12311	0.45200	0.46983	0.34131	0.60054
13211	0.40000	0.41331	0.27867	0.54175
31211	0.30000	0.31277	0.17797	0.44703

	MVH mean	posterior mean	2.5%	97.5%
32111	0.23200	0.24969	0.10945	0.39125
23111	0.36700	0.38319	0.23976	0.52583
21311	0.48700	0.50188	0.36984	0.63096
21131	0.19500	0.20279	0.07474	0.33272
21113	0.34500	0.35410	0.22801	0.48606
11133	0.02800	0.01133	-0.11802	0.13829
11313	0.32000	0.31250	0.18775	0.44137
13113	0.20000	0.19191	0.05431	0.32725
31113	0.10000	0.08884	-0.04774	0.22326
31131	-0.05000	-0.06001	-0.19847	0.07589
13131	0.05000	0.04220	-0.09320	0.17059
11331	0.17000	0.15937	0.02892	0.28798
13311	0.34200	0.34035	0.21298	0.46220
31311	0.24200	0.23926	0.10944	0.37147
33111	0.12200	0.12148	-0.01895	0.26259
11222	0.68900	0.68595	0.55518	0.81822
12122	0.62100	0.62388	0.49291	0.75368
21122	0.65600	0.65579	0.52614	0.78054
12212	0.70800	0.70891	0.58085	0.83588
21212	0.74300	0.74053	0.61257	0.86962
22112	0.67500	0.67763	0.54685	0.81044
21221	0.69100	0.68665	0.54868	0.82072
22121	0.62300	0.62614	0.50183	0.75023
12221	0.65600	0.65611	0.52996	0.78309
22211	0.71000	0.71105	0.58185	0.84163
11223	0.25500	0.26391	0.13817	0.39462
11232	0.15700	0.16657	0.04328	0.29278
11322	0.36200	0.38106	0.24425	0.51285
13122	0.24200	0.26253	0.11980	0.40672
31122	0.14200	0.16012	0.01200	0.30053
12123	0.18700	0.20208	0.06689	0.33748
12132	0.08900	0.10469	-0.03329	0.24569
12312	0.38100	0.40357	0.26279	0.53451
13212	0.32900	0.34751	0.21806	0.47859
31212	0.22900	0.24358	0.11198	0.37452

	MVH mean	posterior mean	2.5%	97.5%
32112	0.16100	0.18398	0.04323	0.32908
23112	0.29600	0.31790	0.17836	0.45818
21312	0.41600	0.43643	0.31144	0.56531
21132	0.12400	0.13667	0.00891	0.26954
21123	0.22200	0.23533	0.10655	0.36917
21213	0.30900	0.31611	0.18212	0.44524
21231	0.15900	0.16782	0.03680	0.30280
21321	0.36400	0.38264	0.24968	0.51663
23121	0.24400	0.26425	0.12153	0.40833
32121	0.10900	0.13020	-0.01164	0.27636
31221	0.17700	0.19146	0.04860	0.33412
13221	0.27700	0.29424	0.15810	0.43149
12321	0.32900	0.35093	0.21539	0.48267
12231	0.12400	0.13526	0.00806	0.26476
12213	0.27400	0.28638	0.15659	0.42014
22113	0.24100	0.25688	0.12382	0.39061
22131	0.09100	0.10584	-0.02461	0.24381
22311	0.38300	0.40477	0.27833	0.53588
23211	0.33100	0.34891	0.21426	0.48095
32211	0.19600	0.21463	0.08633	0.34810
11233	-0.00800	-0.02289	-0.15231	0.10517
11323	0.19700	0.18909	0.06113	0.31315
13123	0.07700	0.07250	-0.06786	0.20998
31123	-0.02300	-0.02891	-0.16872	0.10526
31132	-0.12100	-0.12607	-0.26114	0.00919
13132	-0.02100	-0.02383	-0.15904	0.11196
11332	0.09900	0.09436	-0.03918	0.22638
13312	0.27100	0.27655	0.14595	0.40079
31312	0.17100	0.17178	0.04152	0.30607
33112	0.05100	0.05613	-0.08462	0.19449
31213	0.06400	0.05614	-0.07556	0.19021
31231	-0.08600	-0.09496	-0.22949	0.03955
31321	0.11900	0.11939	-0.02109	0.25617
33121	-0.00100	0.00127	-0.13877	0.14235
13321	0.21900	0.22204	0.09235	0.34922

	MVH mean	posterior mean	2.5%	97.5%
13231	0.01400	0.00625	-0.12564	0.14000
13213	0.16400	0.15713	0.01902	0.29427
12313	0.21600	0.21355	0.08012	0.34358
12331	0.06600	0.06289	-0.07238	0.19889
12133	-0.07600	-0.08518	-0.22075	0.05152
21133	-0.04100	-0.05627	-0.18288	0.06973
21313	0.25100	0.24441	0.11662	0.37106
23113	0.13100	0.12751	-0.01059	0.26307
32113	-0.00400	-0.00603	-0.14083	0.12697
32131	-0.15400	-0.15734	-0.28949	-0.02643
23131	-0.01900	-0.02250	-0.15999	0.11720
21331	0.10100	0.09382	-0.04062	0.22295
23311	0.27300	0.27442	0.14468	0.40355
32311	0.13800	0.14290	0.01318	0.27221
33211	0.08600	0.08515	-0.04215	0.21670
11333	-0.06600	-0.09556	-0.22979	0.03503
13133	-0.18600	-0.21460	-0.35442	-0.07498
31133	-0.28600	-0.31686	-0.44785	-0.18787
13313	0.10600	0.08511	-0.04279	0.20972
31313	0.00600	-0.01673	-0.15066	0.11700
33113	-0.11400	-0.13398	-0.27856	0.00893
31331	-0.14400	-0.16861	-0.30210	-0.03451
33131	-0.26400	-0.28593	-0.42218	-0.15064
13331	-0.04400	-0.06601	-0.19539	0.06539
33311	0.02800	0.01122	-0.11838	0.13873
12222	0.58500	0.58866	0.46168	0.71619
21222	0.62000	0.62209	0.49437	0.75159
22122	0.55200	0.55901	0.43227	0.68359
22212	0.63900	0.64307	0.51526	0.77173
22221	0.58700	0.58949	0.46234	0.71342
12223	0.15100	0.16652	0.03823	0.29185
12232	0.05300	0.07007	-0.06094	0.19971
12322	0.25800	0.28335	0.14397	0.42101
13222	0.20600	0.22662	0.09271	0.36104
31222	0.10600	0.12583	-0.00894	0.26261

	MVH mean	posterior mean	2.5%	97.5%
32122	0.03800	0.06485	-0.08495	0.21257
23122	0.17300	0.19659	0.04964	0.33861
21322	0.29300	0.31571	0.18473	0.45186
21232	0.08800	0.10075	-0.02751	0.22602
21223	0.18600	0.19792	0.06671	0.33435
22123	0.11800	0.13569	0.00469	0.26784
22132	0.02000	0.03937	-0.08828	0.17538
22312	0.31200	0.33614	0.20816	0.46698
23212	0.26000	0.28109	0.15259	0.40949
32212	0.12500	0.14961	0.02219	0.27559
22213	0.20500	0.22169	0.09446	0.35047
22231	0.05500	0.07021	-0.05558	0.20031
22321	0.26000	0.28337	0.14970	0.41159
23221	0.20800	0.22785	0.08867	0.36751
32221	0.07300	0.09377	-0.04187	0.22283
12233	-0.11200	-0.12072	-0.25057	0.00587
12323	0.09300	0.09445	-0.03483	0.22297
13223	0.04100	0.03527	-0.10131	0.17076
31223	-0.05900	-0.06508	-0.19880	0.06862
31232	-0.15700	-0.16128	-0.28640	-0.03001
13232	-0.05700	-0.05879	-0.18737	0.06879
12332	-0.00500	-0.00342	-0.13941	0.13381
13322	0.14800	0.15470	0.01772	0.28825
31322	0.04800	0.05402	-0.08816	0.19687
33122	-0.07200	-0.06566	-0.20312	0.07504
21233	-0.07700	-0.08937	-0.21867	0.04490
21323	0.12800	0.12477	-0.00483	0.25326
23123	0.00800	0.00815	-0.13433	0.14648
32123	-0.12700	-0.12786	-0.26084	0.00712
32132	-0.22500	-0.22360	-0.35763	-0.08675
23132	-0.09000	-0.08855	-0.22697	0.04531
21332	0.03000	0.02774	-0.10126	0.15641
23312	0.20200	0.20804	0.08141	0.33206
32312	0.06700	0.07583	-0.05636	0.21025
33212	0.01500	0.01884	-0.11648	0.14756

	MVH mean	posterior mean	2.5%	97.5%
32213	-0.04000	-0.04192	-0.17120	0.08567
32231	-0.19000	-0.19355	-0.32382	-0.06778
32321	0.01500	0.02192	-0.11001	0.15314
33221	-0.03700	-0.03407	-0.16761	0.10257
23321	0.15000	0.15623	0.02629	0.28684
23231	-0.05500	-0.06001	-0.19184	0.07180
23213	0.09500	0.09097	-0.03705	0.22183
22313	0.14700	0.14852	0.02351	0.27755
22331	-0.00300	-0.00179	-0.13147	0.12599
22133	-0.14500	-0.15013	-0.28171	-0.01979
12333	-0.17000	-0.19483	-0.33004	-0.06572
13233	-0.22200	-0.24944	-0.38746	-0.11890
31233	-0.32200	-0.35263	-0.48942	-0.21555
13323	-0.01700	-0.03373	-0.16507	0.09482
31323	-0.11700	-0.13691	-0.26820	-0.00439
33123	-0.23700	-0.25625	-0.39954	-0.12062
31332	-0.21500	-0.23525	-0.36976	-0.09693
33132	-0.33500	-0.35150	-0.49005	-0.21411
13332	-0.11500	-0.13170	-0.26542	0.00021
33312	-0.04300	-0.05420	-0.18420	0.06902
33213	-0.15000	-0.17167	-0.30319	-0.03872
33231	-0.30000	-0.32201	-0.44657	-0.19460
33321	-0.09500	-0.10777	-0.23842	0.02115
32313	-0.09800	-0.11524	-0.24763	0.01287
32331	-0.24800	-0.26573	-0.39857	-0.13558
32133	-0.39000	-0.41261	-0.54370	-0.27693
23133	-0.25500	-0.28133	-0.42478	-0.14315
23313	0.03700	0.01695	-0.11209	0.14221
23331	-0.11300	-0.13239	-0.26220	-0.00352
21333	-0.13500	-0.16257	-0.28863	-0.03351
13333	-0.28000	-0.32360	-0.45599	-0.19105
31333	-0.38000	-0.42585	-0.56076	-0.29111
33133	-0.50000	-0.54210	-0.67895	-0.40650
33313	-0.20800	-0.24411	-0.37258	-0.10811
33331	-0.35800	-0.39321	-0.51896	-0.26715

	MVH mean	posterior mean	2.5%	97.5%
22222	0.51600	0.52314	0.40111	0.64854
22223	0.08200	0.10090	-0.02564	0.22739
22232	-0.01600	0.00325	-0.12383	0.12727
22322	0.18900	0.21793	0.08816	0.34825
23222	0.13700	0.16096	0.02556	0.29278
32222	0.00200	0.02810	-0.10539	0.15984
22233	-0.18100	-0.18656	-0.31394	-0.06097
22323	0.02400	0.02692	-0.09814	0.15279
23223	-0.02800	-0.03010	-0.16105	0.10741
32223	-0.16300	-0.15996	-0.28316	-0.03584
32232	-0.26100	-0.25830	-0.38606	-0.12990
23232	-0.12600	-0.12590	-0.25121	0.00145
22332	-0.07400	-0.07000	-0.20155	0.05681
23322	0.07900	0.08926	-0.03596	0.21831
32322	-0.05600	-0.04703	-0.18289	0.08320
33222	-0.10800	-0.10195	-0.23410	0.03557
22333	-0.23900	-0.25911	-0.39059	-0.13234
23233	-0.29100	-0.31558	-0.45285	-0.18682
32233	-0.42600	-0.44949	-0.56947	-0.32274
23323	-0.08600	-0.10199	-0.22366	0.02189
32323	-0.22100	-0.23402	-0.35987	-0.10529
33223	-0.27300	-0.28919	-0.42583	-0.15720
32332	-0.31900	-0.33232	-0.46647	-0.19697
33232	-0.37100	-0.38930	-0.51676	-0.26271
23332	-0.18400	-0.19833	-0.32732	-0.07087
33322	-0.16600	-0.17288	-0.30140	-0.04523
23333	-0.34900	-0.38892	-0.52621	-0.25762
32333	-0.48400	-0.52323	-0.65577	-0.39249
33233	-0.53600	-0.57814	-0.71167	-0.44630
33323	-0.33100	-0.36372	-0.49231	-0.23962
33332	-0.42900	-0.46050	-0.59385	-0.33079
33333	-0.59400	-0.65041	-0.77951	-0.52241

Table 3.2: Summary of the MCMC samples of the posteriors of the EQ-5D-3L scores.

Figures 3.1, 3.2 and 3.3 show the states ranked in terms of severity (with respect to the values of the posterior medians). The MVH point estimates of the utilities of the states are within the 95% credible intervals. In general the pattern is that the order of the states has more or less remained the same as the MVH order. Furthermore, there appears to be no pattern with respect to the standard deviation of the utilities which take similar values across the states, the mean standard deviation being: 0.066714.

Moreover, the results of the derived MCMC samples are also robust to changes in priors using the alternative prior distributions which were discussed in section 3.2.1. When different starting values for the MCMC method or different priors are used, in the end we still obtain similar results for the posterior statistics of the β's. There might be some small changes in terms of the order in which some states are ranked, but such a switch is for states which have very similar utility values and overlapping credible intervals. For example, Figures 3.4 and 3.5 illustrate the posterior medians for the 14 EQ-5D-3L states which were ranked the lowest and their corresponding 95% credible intervals which were calculated under the original priors and the alternative priors respectively. Thus, we can see that states `33232` and `23333` have switched positions; states `31233` and `33132` have also switched positions when they are ranked with respect to their posterior medians. Table 3.3 shows the posterior means and medians of these states when both the original and the alternative priors are used for their derivation. It appears that there was a minimal change in the values of the posterior means/medians (and their posterior distributions) and thus the order in which they are ranked changed because the original posterior medians of these states were very similar in the first place. However, the general order of the EQ-5D-3L states, their posterior means and standard deviations remain the same and thus the overall results which are obtained are similar. In other words, in general the parameters are estimated precisely enough that the inferences are not sensitive to the starting MCMC values or the particular choice of prior distributions.

3.3 Approximating the posterior distributions

3.3.1 The framework of the approximation

For the posterior utility u_s of the s-th EQ-5D-3L state we can obtain an MCMC sample of size C. Although we obtain the sample $u_s^{(1)}, u_s^{(2)}, ..., u_s^{(C)}$, which is derived from the probability density function $f_{u_s}(\cdot)$ of u_s , we do not know the parametric form of $f_{u_s}(\cdot)$, because my model is not a conjugate or simplistic one. Kernel density estimates smooth the histogram of the MCMC

State	Posterior mean (original)	Posterior mean (alternative)	Posterior median (original)	Posterior median (alternative)
33232	-0.38930	-0.38787	-0.38971	-0.38778
23333	-0.38892	-0.38825	-0.38855	-0.38918
31233	-0.35263	-0.3534	-0.35229	-0.35273
33132	-0.35150	-0.35271	-0.35077	-0.35349

Table 3.3: The posterior means and medians of states 33232 and 23333, as well as states 31233 and 33132 when both the original and the alternative priors are used for their derivation. It appears that the original posterior medians of these pairs of states were very similar in the first place and they remain similar even when the alternative priors are used. However, they switched ranks due to the minimal changes caused by using the alternative priors.

sample giving us an first idea of the nature of the posterior distribution. Nevertheless, there is a greater interest in approximating these derived distributions with closed form distributions as this will allow to explicitly inform the parametric form of each distribution to researchers without having to show any graphs.

My objective is to find suitable parametric approximations $\hat{f}_{u_s}(\cdot)$ for each of the posterior EQ-5D-3L states utility distributions $f_{u_s}(\cdot)$ (which are deterministic functions of the β's and $\xi^{(new)}$'s). Specifically, due to graphical evidence of multi-modality it is reasonable to model them using mixtures of distributions which will be capable of approximating the true distribution better. The use of an algorithm for the approximation of multi-modal distributions with the use of standard probability distributions, such as normal, is a suitable solution to the problem of approximating the mixtures (Rubinshtein, 1993). These distributions are approximated as mixtures of normal probability functions with Z finite components:

$$\hat{f}_{u_s}(\cdot) = \sum_{z=1}^{Z} w_z N(\cdot\,|a_z, b_z). \tag{3.2}$$

The weights w_z's of the components as well as the components' means a_z's, and their corresponding standard deviations b_z's of (3.2) will have to be estimated separately. There is a trade-off between the quality of the approximation, which is increased by having further components, and the complexity of the algorithm as adding further components brings further computational intensity as well as difficulties for the algorithm to reach convergence. Similar work of Schmidli et al. (2014) argued that a parsimonious and convenient approximation is a three-component ($Z = 3$) mixture which satisfactory approximates the targeted distribution.

In order to attain a close approximation of the real distribution the Kullback–Leibler (KL) divergence is regarded, as it is considered the standard measure in inference problems (O'Hagan and Forster, 2004). The KL divergence between the target distribution $f_{u_s}(x)$ and the approximate (mixture) distribution $\hat{f}_{u_s}(x)$ is defined as:

$$\mathrm{KL}\left(f_{u_s}(x), \hat{f}_{u_s}(x)\right) = \int_x \log\left(f_{u_s}(x)\right) f_{u_s}(x)\, dx - \int_x \log\left(\hat{f}_{u_s}(x)\right) f_{u_s}(x)\, dx. \tag{3.3}$$

The lower the KL divergence (between the proposed and the actual distribution), the better the approximation. The ideal approximation (i.e. when the KL takes the minimum value, which theoretically is 0) is derived by selecting such weights and hyper-parameters (using numerical optimisation) to have a maximum in the second right term of (3.3): $\int_x \log\left(\hat{f}_{u_s}(x)\right) f_{u_s}(x)\, dx$. Using the MCMC sample $u_s^{(1)}, u_s^{(2)}, ..., u_s^{(C)}$ generated from the posterior distribution of u_s, we can deduce the Monte-Carlo estimate of this integral as: $\frac{1}{C}\sum_{c=1}^{C} \log\left(\hat{f}_{u_s}\left(u_s^{(c)}\right)\right)$. Moreover, this is the same as the mean log-likelihood of the parameters of the mixture $\hat{f}_{u_s}(\cdot)$ given the observed MCMC sample. Hence, this implies that in order to minimise the KL divergence, we have to maximise this log-likelihood.

Moreover, when using the maximum-likelihood estimators the underlying log-likelihood is maximised and thus the KL divergence is minimised. Therefore, the KL divergence is optimal when the weights and the hyper-parameters of (3.2) are equal to the maximum likelihood estimates. In other words, the problem of finding a good approximate distribution is simplified to deriving the maximum likelihood estimates.

3.3.2 Estimating the parameters of the approximate distributions

The objective is to estimate the parameters of the mixture as the maximum likelihood estimates given the observed MCMC sample, but since we have a multivariate case, numerical optimisation is required for the successful estimation of these parameters.

The concept of optimisation in general is to construct an algorithm which can help us obtain a value x_* for which: $g(x_*) = 0$, where $g(\cdot)$ is an at-least-once differentiable function with domain and range being the set of real numbers. Alternatively, we can search for a x_* which maximises the at-least-twice differentiable function $g(\cdot)$. Newton's method of maximisation (Nocedal and Wright, 1999) initiates a starting value x_0 and then tries to construct a sequence which converges towards x_* for which $g(x_*)$ is a maximum. In particular, in the $t+1$ step of the algorithm, the value x_{t+1} of the sequence is calculated as:

$$x_{t+1} = x_t - \frac{g'(x_t)}{g''(x_t)}, \tag{3.4}$$

where $g'(\cdot)$ and $g''(\cdot)$ are respecively the first and second order derivatives of $g(\cdot)$. Intuitively, this kind of optimisation algorithms can be considered as ways to find the "highest place" by "going uphill" until a place which is "flat" is found (i.e. a place where the derivative of the objective function is zero). Therefore, the stage of initiating the value of the algorithm is crucial as a bad choice might result in non-convergence of the algorithm (e.g. due the nature of the function, the rest of the sequence terms might go further and further away from the optimal value).

In our case we have a log-likelihood function $g_{u_s}(\cdot)$ of a mixture which approximates the utility of state s; $g_{u_s}(\cdot)$ is defined as follows:

$$g_{u_s}(\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}) = \log\left(\hat{f}_{u_s}\left(\boldsymbol{w}, w_Z, \boldsymbol{a}, \boldsymbol{b} \,\middle|\, u_s^{(1)}, ..., u_s^{(C)}\right)\right). \tag{3.5}$$

The function $\hat{f}_{u_s}(\cdot)$ is defined in (3.2), $u_s^{(1)}, ..., u_s^{(C)}$ is the MCMC sample of size C of $f_{u_s}(\cdot)$, whereas $\boldsymbol{w}^\top = (w_1, ..., w_{Z-1})$, $\boldsymbol{a}^\top = (a_1, ..., a_Z)$, and $\boldsymbol{b}^\top = (b_1, ..., b_Z)$ are the component weights, means and standard deviations of the mixture distribution as defined in (3.2). Moreover, since the weights of the components have to sum to unity: $w_Z = 1 - \sum_{z=1}^{Z-1} w_z$. The domain of $g_{u_s}(\cdot) : \mathbb{R}^{3Z-1} \mapsto \mathbb{R}$ has $3Z-1$ dimensions, whereas being a log-likelihood function, its output is of one dimension. The maximum likelihood estimates $\hat{\boldsymbol{w}}$, $\hat{\boldsymbol{a}}$, and $\hat{\boldsymbol{b}}$ of $\boldsymbol{w}$, $\boldsymbol{a}$, and $\boldsymbol{b}$ in respect, are the values which maximise the $g_{u_s}(\cdot)$. Although it is not possible to obtain the exact maximum likelihood estimates, the objective is to use optimisation techniques in order to approximate the maximum likelihood estimates by deriving: $\boldsymbol{w}_* \simeq \hat{\boldsymbol{w}}$, $\boldsymbol{a}_* \simeq \hat{\boldsymbol{a}}$, and $\boldsymbol{b}_* \simeq \hat{\boldsymbol{b}}$.

The J-dimensional (i.e. $\boldsymbol{x}^\top = (x_1, ..., x_J)$) equivalent of (3.4) is:

$$\boldsymbol{x_{t+1}} = \boldsymbol{x_t} - \alpha_t \boldsymbol{H_t^{-1}} \nabla g(\boldsymbol{x_t}), \tag{3.6}$$

where $\nabla g(\boldsymbol{x_t})$ is the gradient of the function evaluated at $\boldsymbol{x_t}$, that is:

$$(\nabla g(\boldsymbol{x_t}))^\top = \left(\left.\frac{\partial g(\boldsymbol{x})}{\partial x_1}\right|_{x_1=x_{t,1}}, ..., \left.\frac{\partial g(\boldsymbol{x})}{\partial x_J}\right|_{x_J=x_{t,J}}\right), \tag{3.7}$$

and $\boldsymbol{H}$ is the Hessian. The Hessian matrix at iteration t is defined as $\boldsymbol{H_t} = \nabla^2 g(\boldsymbol{x_t})$, or:

$$\boldsymbol{H_t} = \begin{pmatrix} \left.\frac{\partial^2 g(\boldsymbol{x})}{\partial x_1^2}\right|_{x_1=x_{t,1}} & \cdots & \left.\frac{\partial^2 g(\boldsymbol{x})}{\partial x_1 \partial x_J}\right|_{x_1=x_{t,1}, x_J=x_{t,J}} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial^2 g(\boldsymbol{x})}{\partial x_J \partial x_1}\right|_{x_J=x_{t,J}, x_1=x_{t,1}} & \cdots & \left.\frac{\partial^2 g(\boldsymbol{x})}{\partial x_J^2}\right|_{x_J=x_{t,J}} \end{pmatrix}. \tag{3.8}$$

The scalar α_t can be 1, or it can be chosen to satisfy some conditions such as the Wolfe conditions (Wolfe, 1971):

$$g(\boldsymbol{x_t} + \alpha_t \boldsymbol{p_t}) \leq g(\boldsymbol{x_t}) + c_1 \alpha_t \boldsymbol{p_t}^\top \nabla g(\boldsymbol{x_t}), \tag{3.9}$$

$$\boldsymbol{p_t}^\top \nabla g\left(\boldsymbol{x_t} + \alpha_t \boldsymbol{p_t}\right) \geq c_2 \boldsymbol{p_t}^\top \nabla g\left(\boldsymbol{x_t}\right), \tag{3.10}$$

where $\boldsymbol{p_t} = -\boldsymbol{H_t^{-1}} \nabla g\left(\boldsymbol{x_t}\right)$, and c_1 and c_2 are constants for which $0 < c_1 < c_2 < 1$. The most popular way of selecting the value of α_t is backtracking line search (Nocedal and Wright, 1999): we start by choosing $\alpha_t = 1$ and if it does not satisfy the Wolfe conditions (3.9) and (3.10), then we iteratively shrink the value of α_t until the Wolfe conditions are satisfied. Convergence can be checked by observing the norm of the gradient: $|\nabla g\left(\boldsymbol{x_t}\right)|$ (which, at the optimal $\boldsymbol{x}^*$, should be close to $\mathbf{0}$).

The higher dimensional the considered space, the more difficult it can be to compute the inverse of the Hessian, as it requires the derivation of second-order derivatives. However, the Hessian matrix does not need to be computed in quasi-Newton methods. The Hessian is updated by analysing successive gradient vectors instead. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, which was proposed by Broyden (1970), Fletcher (1970), Goldfarb (1970), and (Shanno, 1970) independently, uses matrix $\boldsymbol{B_t}$, where $\boldsymbol{B_t^{-1}}$ is an approximation to the inverse Hessian $\boldsymbol{H_t^{-1}}$ at iteration t, which is easier to calculate; it is the most efficient of the quasi-Newton methods (Dai, 2012).

In particular, the first step of the BFGS algorithm is to set starting values to the elements of the vector $\boldsymbol{x_0}$ and the matrix $\boldsymbol{B_0}$. The initialisation of $\boldsymbol{x_0}$ is done by trial and error by visually inspecting the data, whereas the decision is to choose $\boldsymbol{B_0} = \boldsymbol{I_J}$. Then, the following steps are repeated until the sequence of $\boldsymbol{x_t}$ converges to solution $\boldsymbol{x_*}$:

1. Compute $\boldsymbol{p_t}$ as: $\boldsymbol{p_t} = -\boldsymbol{B_t^{-1}} \nabla g\left(\boldsymbol{x_t}\right)$.

2. Perform backtracking line search in order to find an acceptable stepsize α_t which satisfies the Wolfe conditions (3.9) and (3.10). We have $c_1 = 10^{-4}$, and $c_2 = 0.9$, as proposed by Nocedal and Wright (1999).

3. Update $\boldsymbol{x_{t+1}}$ as: $\boldsymbol{x_{t+1}} = \boldsymbol{x_t} + \alpha_t \boldsymbol{p_t}$.

4. Set $\boldsymbol{\zeta_t} = \alpha_t \boldsymbol{p_t}$.

5. Compute $\boldsymbol{y_t} = \nabla g\left(\boldsymbol{x_{t+1}}\right) - \nabla g\left(\boldsymbol{x_t}\right)$.

6. Update $\boldsymbol{B_{t+1}} = \boldsymbol{B_t} + \frac{\boldsymbol{y_t} \boldsymbol{y_t}^\top}{\boldsymbol{y_t}^\top \boldsymbol{\zeta_t}} - \frac{\boldsymbol{B_t} \boldsymbol{\zeta_t} \boldsymbol{\zeta_t}^\top \boldsymbol{B_t}}{\boldsymbol{\zeta_t}^\top \boldsymbol{B_t} \boldsymbol{\zeta_t}}$.

In the first step, $\boldsymbol{B_{t+1}^{-1}}$ can be computed as: $\boldsymbol{B_{t+1}^{-1}} = \boldsymbol{B_t^{-1}} + \frac{\left(\boldsymbol{\zeta_t}^\top \boldsymbol{y_t} + \boldsymbol{y_t}^\top \boldsymbol{B_t^{-1}} \boldsymbol{y_t}\right)\left(\boldsymbol{\zeta_t} \boldsymbol{\zeta_t}^\top\right)}{\left(\boldsymbol{\zeta_t}^\top \boldsymbol{y_t}\right)^2} - \frac{\boldsymbol{B_t^{-1}} \boldsymbol{y_t} \boldsymbol{\zeta_t}^\top + \boldsymbol{\zeta_t} \boldsymbol{y_t}^\top \boldsymbol{B_t^{-1}}}{\boldsymbol{\zeta_t}^\top \boldsymbol{y_t}}$.

In our settings the BFGS algorithm is used, where the target function $g\left(\cdot\right)$ is defined by (3.5)

and $\boldsymbol{x}_{\boldsymbol{t}}^{\top} = (w_{t,1}, ..., w_{t,Z-1}, a_{t,1}, ..., a_{t,Z}, b_{t,1}, ..., b_{t,Z})$. The solution set $\boldsymbol{x}_*$ (which consists of a satisfactory approximation of the maximum likelihood estimates) is used to define the parameters of the approximate mixture distribution (3.2).

3.3.3 Computing the Kullback-Leibler divergence

In section 3.3.2 I showed how to estimate the parameters of the mixture of normals $\hat{f}_{u_s}(\cdot)$. The next step is to use the KL divergence as a measure to assess the approximation of the actual posterior distribution $f_{u_s}(\cdot)$ by the mixture $\hat{f}_{u_s}(\cdot)$. The underlying value of $KL\left(f_{u_s}(x), \hat{f}_{u_s}(x)\right)$ cannot be derived directly from (3.3) as we do not know the full parametric form of $f_{u_s}(\cdot)$. Thus, the KL divergence should be estimated.

First of all, since we cannot directly compute $f_{u_s}(x)$, we can use the kernel density estimate $f_{u_s}^{\text{ker}}(x)$ instead. We can set specified bins and evaluate the kernel density at E point coordinates $\boldsymbol{r}^{\top} = (r_1, ..., r_E)$ over the range of the vector $\left(u_s^{(1)}, ..., u_s^{(C)}\right)$. Then, we could also compute the closed-form density of $\hat{f}_{u_s}(\cdot)$ at the same coordinates $\boldsymbol{r}^{\top}$.

We can empirically compute the KL divergence $\widehat{KL}\left(f_{u_s}(\cdot), \hat{f}_{u_s}(\cdot)\right)$ as:

$$\widehat{KL}\left(f_{u_s}(\cdot), \hat{f}_{u_s}(\cdot)\right) = \frac{\max(\boldsymbol{r}) - \min(\boldsymbol{r})}{E-1} \cdot \sum_{i=1}^{E} f_{u_s}^{\text{ker}}(r_i) \cdot \log\left[f_{u_s}^{\text{ker}}(r_i) / \hat{f}_{u_s}(r_i)\right], \tag{3.11}$$

where the quantity before the sum is a scale factor.

3.3.4 Results

Each posterior utility distribution is approximated as a $Z = 3$ mixture of Normal distributions. The BFGS algorithm is used and the criterion used for the assessment of the approximations is the KL divergence which is empirically calculated as in (3.11), where I use $E = 512$.

Moreover, each posterior distribution is then described as a mixture of normal distributions. Figures 3.6 and 3.7 illustrate the kernel density plots of states `11212` and `12112` in respect, for the MCMC simulations, and the superimposed probability density functions of their approximations as mixtures of normal density functions. Similar figures for the approximations of all the EQ-5D-3L states by three-component mixtures of Normals can be found in the Appendix. Moreover, Table 3.4 displays a summary of all the approximate mixture distributions. In the Appendix, Figure 5 visually illustrates the KL divergence values for each state. It can be seen that all the KL divergence values are quite small; furthermore, Figure 6 (in the Appendix) visually illustrates the weights of each of the three components of mixtures for each state.

	Approximate mixture distribution	KL
11112	$0.07068 \cdot N(0.78101, 0.07688) + 0.83179 \cdot N(0.83716, 0.05592) + 0.09753 \cdot N(0.91373, 0.05372)$	0.00210
11121	$0.03685 \cdot N(0.66983, 0.04489) + 0.50141 \cdot N(0.78228, 0.05137) + 0.46174 \cdot N(0.80592, 0.06587)$	0.00192
11211	$0.29981 \cdot N(0.82265, 0.05896) + 0.53380 \cdot N(0.87451, 0.04856) + 0.16639 \cdot N(0.94340, 0.05651)$	0.00195
12111	$0.26632 \cdot N(0.80387, 0.07593) + 0.64833 \cdot N(0.81089, 0.05177) + 0.08535 \cdot N(0.83576, 0.07766)$	0.00237
21111	$0.15694 \cdot N(0.80601, 0.06287) + 0.75432 \cdot N(0.85011, 0.05863) + 0.08875 \cdot N(0.86261, 0.09031)$	0.00248
11113	$0.12430 \cdot N(0.38135, 0.06879) + 0.83987 \cdot N(0.41943, 0.05722) + 0.03583 \cdot N(0.52785, 0.04338)$	0.00216
11131	$0.14276 \cdot N(0.26433, 0.07129) + 0.85504 \cdot N(0.27016, 0.06244) + 0.00221 \cdot N(0.51974, 0.02594)$	0.00338
11311	$0.01106 \cdot N(0.48614, 0.00854) + 0.20661 \cdot N(0.57033, 0.04104) + 0.78233 \cdot N(0.56678, 0.07054)$	0.00231
13111	$0.02676 \cdot N(0.43296, 0.12133) + 0.79490 \cdot N(0.44336, 0.06464) + 0.17834 \cdot N(0.48467, 0.07279)$	0.00261
31111	$0.71003 \cdot N(0.33989, 0.06252) + 0.09639 \cdot N(0.41718, 0.04463) + 0.19358 \cdot N(0.34307, 0.09076)$	0.00209
11122	$0.48079 \cdot N(0.70193, 0.06066) + 0.44547 \cdot N(0.73174, 0.05379) + 0.07374 \cdot N(0.79534, 0.06447)$	0.00229
11212	$0.09551 \cdot N(0.76481, 0.07419) + 0.83416 \cdot N(0.80549, 0.05933) + 0.07032 \cdot N(0.87660, 0.06196)$	0.00191
12112	$0.14808 \cdot N(0.74500, 0.08554) + 0.84894 \cdot N(0.74594, 0.06219) + 0.00299 \cdot N(0.66691, 0.00233)$	0.00264
21112	$0.00376 \cdot N(0.60369, 0.02188) + 0.91939 \cdot N(0.77563, 0.06028) + 0.07685 \cdot N(0.79075, 0.09674)$	0.00199
21121	$0.28263 \cdot N(0.71965, 0.04914) + 0.56627 \cdot N(0.71740, 0.06847) + 0.15111 \cdot N(0.75074, 0.06154)$	0.00259
12121	$0.07788 \cdot N(0.65282, 0.07080) + 0.71699 \cdot N(0.68379, 0.05675) + 0.20513 \cdot N(0.72857, 0.06769)$	0.00181
11221	$0.31950 \cdot N(0.72512, 0.06972) + 0.41619 \cdot N(0.74504, 0.05293) + 0.26430 \cdot N(0.80161, 0.06342)$	0.00281
12211	$0.09387 \cdot N(0.71228, 0.06984) + 0.87497 \cdot N(0.77808, 0.05992) + 0.03116 \cdot N(0.85284, 0.07512)$	0.00178
21211	$0.11008 \cdot N(0.78092, 0.08198) + 0.80426 \cdot N(0.80348, 0.06279) + 0.08565 \cdot N(0.86206, 0.06895)$	0.00184
22111	$0.07503 \cdot N(0.81145, 0.05671) + 0.52169 \cdot N(0.73558, 0.06876) + 0.40328 \cdot N(0.74642, 0.04822)$	0.00163
11123	$0.11464 \cdot N(0.31584, 0.05139) + 0.78320 \cdot N(0.29498, 0.06720) + 0.10216 \cdot N(0.30787, 0.05182)$	0.00196
11132	$0.08713 \cdot N(0.15659, 0.07340) + 0.78543 \cdot N(0.20616, 0.06089) + 0.12743 \cdot N(0.21369, 0.08239)$	0.00193
11312	$0.37200 \cdot N(0.48914, 0.07219) + 0.61575 \cdot N(0.50494, 0.05772) + 0.01224 \cdot N(0.66206, 0.03201)$	0.00189
13112	$0.83804 \cdot N(0.38249, 0.07420) + 0.14786 \cdot N(0.38149, 0.04333) + 0.01410 \cdot N(0.46418, 0.10359)$	0.00201
31112	$0.37082 \cdot N(0.25053, 0.06848) + 0.44377 \cdot N(0.28980, 0.06063) + 0.18540 \cdot N(0.31278, 0.07689)$	0.00174
31121	$0.17092 \cdot N(0.22625, 0.09392) + 0.82105 \cdot N(0.22679, 0.06587) + 0.00803 \cdot N(0.36712, 0.01247)$	0.00233
13121	$0.18603 \cdot N(0.33099, 0.09112) + 0.67226 \cdot N(0.33922, 0.06283) + 0.14170 \cdot N(0.28457, 0.06604)$	0.00346
11321	$0.14853 \cdot N(0.43643, 0.09608) + 0.69842 \cdot N(0.46204, 0.05629) + 0.15306 \cdot N(0.39407, 0.04986)$	0.00212
11231	$0.21699 \cdot N(0.22764, 0.08616) + 0.45771 \cdot N(0.21354, 0.05371) + 0.32530 \cdot N(0.26412, 0.05490)$	0.00226
11213	$0.10356 \cdot N(0.36501, 0.09095) + 0.02722 \cdot N(0.32759, 0.01585) + 0.86922 \cdot N(0.38610, 0.06134)$	0.00219
12113	$0.09096 \cdot N(0.32075, 0.09978) + 0.88779 \cdot N(0.31859, 0.06063) + 0.02125 \cdot N(0.45726, 0.02478)$	0.00270
12131	$0.08095 \cdot N(0.08976, 0.06072) + 0.87030 \cdot N(0.17346, 0.06101) + 0.04875 \cdot N(0.25466, 0.07190)$	0.00192
12311	$0.14503 \cdot N(0.42116, 0.06972) + 0.63822 \cdot N(0.46044, 0.05129) + 0.21675 \cdot N(0.53003, 0.05492)$	0.00265
13211	$0.02299 \cdot N(0.36215, 0.01774) + 0.46107 \cdot N(0.41926, 0.05313) + 0.51594 \cdot N(0.41028, 0.07707)$	0.00236
31211	$0.23171 \cdot N(0.26903, 0.06684) + 0.72401 \cdot N(0.32054, 0.05984) + 0.04427 \cdot N(0.41439, 0.05457)$	0.00214
32111	$0.10416 \cdot N(0.21848, 0.09186) + 0.75362 \cdot N(0.24069, 0.06121) + 0.14222 \cdot N(0.32021, 0.06097)$	0.00176
23111	$0.13801 \cdot N(0.31737, 0.06763) + 0.71286 \cdot N(0.38797, 0.06287) + 0.14914 \cdot N(0.42122, 0.08034)$	0.00217
21311	$0.24217 \cdot N(0.47608, 0.07458) + 0.37415 \cdot N(0.49015, 0.05192) + 0.38368 \cdot N(0.52965, 0.06131)$	0.00336
21131	$0.07098 \cdot N(0.15747, 0.02891) + 0.78807 \cdot N(0.20278, 0.07023) + 0.14095 \cdot N(0.22566, 0.03187)$	0.00213
21113	$0.00271 \cdot N(0.27339, 0.11627) + 0.75556 \cdot N(0.34870, 0.06002) + 0.24173 \cdot N(0.37188, 0.07784)$	0.00247
11133	$0.17129 \cdot N(-0.01686, 0.07366) + 0.69820 \cdot N(0.01276, 0.05438) + 0.13051 \cdot N(0.04072, 0.07828)$	0.00257
11313	$0.22036 \cdot N(0.32236, 0.08084) + 0.48575 \cdot N(0.29106, 0.05573) + 0.29390 \cdot N(0.34052, 0.04993)$	0.00197
13113	$0.35258 \cdot N(0.16329, 0.06952) + 0.59360 \cdot N(0.20380, 0.05724) + 0.05383 \cdot N(0.24826, 0.09444)$	0.00255
31113	$0.27864 \cdot N(0.04702, 0.06881) + 0.52051 \cdot N(0.08374, 0.04617) + 0.20085 \cdot N(0.16007, 0.05238)$	0.00167
31131	$0.16622 \cdot N(-0.09751, 0.08166) + 0.77619 \cdot N(-0.05886, 0.05951) + 0.05759 \cdot N(0.03261, 0.05897)$	0.00235
13131	$0.21517 \cdot N(-0.00677, 0.06757) + 0.73118 \cdot N(0.05142, 0.05735) + 0.05365 \cdot N(0.11321, 0.06732)$	0.00231
11331	$0.14502 \cdot N(0.16861, 0.08688) + 0.85013 \cdot N(0.15708, 0.05996) + 0.00484 \cdot N(0.28485, 0.00660)$	0.00247
13311	$0.02983 \cdot N(0.21267, 0.04466) + 0.72468 \cdot N(0.32736, 0.05193) + 0.24549 \cdot N(0.39422, 0.05238)$	0.00256
31311	$0.16719 \cdot N(0.16630, 0.04972) + 0.74439 \cdot N(0.24479, 0.05330) + 0.08843 \cdot N(0.33067, 0.05083)$	0.00179
33111	$0.17690 \cdot N(0.08637, 0.07480) + 0.77652 \cdot N(0.12288, 0.06316) + 0.04658 \cdot N(0.23155, 0.05362)$	0.00172

	Approximate mixture distribution	KL
11222	$0.19330 \cdot N(0.64937, 0.06764) + 0.71558 \cdot N(0.68993, 0.05835) + 0.09111 \cdot N(0.73250, 0.08476)$	0.00203
12122	$0.17482 \cdot N(0.57064, 0.06526) + 0.54276 \cdot N(0.61919, 0.05303) + 0.28242 \cdot N(0.66588, 0.06026)$	0.00204
21122	$0.30506 \cdot N(0.64521, 0.06941) + 0.66375 \cdot N(0.66113, 0.05834) + 0.03118 \cdot N(0.64562, 0.10279)$	0.00246
12212	$0.27919 \cdot N(0.68027, 0.06681) + 0.71596 \cdot N(0.71873, 0.05938) + 0.00485 \cdot N(0.90638, 0.03657)$	0.00219
21212	$0.06049 \cdot N(0.78590, 0.04989) + 0.84726 \cdot N(0.73670, 0.06095) + 0.09225 \cdot N(0.74600, 0.09534)$	0.00229
22112	$0.74615 \cdot N(0.67697, 0.07035) + 0.03738 \cdot N(0.65573, 0.01872) + 0.21647 \cdot N(0.68363, 0.05535)$	0.00197
21221	$0.04729 \cdot N(0.63804, 0.02379) + 0.79490 \cdot N(0.68451, 0.07199) + 0.15781 \cdot N(0.71194, 0.03370)$	0.00176
22121	$0.31077 \cdot N(0.62757, 0.04802) + 0.68535 \cdot N(0.62551, 0.06768) + 0.00388 \cdot N(0.62826, 0.04914)$	0.00196
12221	$0.37969 \cdot N(0.64550, 0.05532) + 0.45974 \cdot N(0.65235, 0.07042) + 0.16057 \cdot N(0.69214, 0.05406)$	0.00204
22211	$0.01825 \cdot N(0.63196, 0.07439) + 0.97656 \cdot N(0.71169, 0.06385) + 0.00519 \cdot N(0.86571, 0.04971)$	0.00253
11223	$0.20824 \cdot N(0.25370, 0.08023) + 0.76346 \cdot N(0.26276, 0.05698) + 0.02830 \cdot N(0.37019, 0.05816)$	0.00263
11232	$0.00959 \cdot N(0.07569, 0.09583) + 0.96837 \cdot N(0.16488, 0.06081) + 0.02204 \cdot N(0.27981, 0.03695)$	0.00202
11322	$0.40795 \cdot N(0.36021, 0.04830) + 0.37327 \cdot N(0.37900, 0.08513) + 0.21877 \cdot N(0.42346, 0.04525)$	0.00172
13122	$0.08465 \cdot N(0.20671, 0.08133) + 0.90510 \cdot N(0.26561, 0.06765) + 0.01025 \cdot N(0.45064, 0.04108)$	0.00215
31122	$0.21520 \cdot N(0.15176, 0.09194) + 0.68438 \cdot N(0.15339, 0.06404) + 0.10042 \cdot N(0.22387, 0.04935)$	0.00220
12123	$0.02548 \cdot N(0.08566, 0.04125) + 0.90961 \cdot N(0.20398, 0.06329) + 0.06491 \cdot N(0.22120, 0.09592)$	0.00216
12132	$0.07152 \cdot N(0.03496, 0.07709) + 0.87851 \cdot N(0.10483, 0.06273) + 0.04997 \cdot N(0.20194, 0.06453)$	0.00187
12312	$0.02579 \cdot N(0.25982, 0.04623) + 0.78475 \cdot N(0.40059, 0.06035) + 0.18946 \cdot N(0.43549, 0.07372)$	0.00194
13212	$0.00402 \cdot N(0.22255, 0.12437) + 0.93964 \cdot N(0.34346, 0.06260) + 0.05634 \cdot N(0.42385, 0.05407)$	0.00284
31212	$0.23507 \cdot N(0.19636, 0.06521) + 0.53440 \cdot N(0.23993, 0.05098) + 0.23052 \cdot N(0.30019, 0.05856)$	0.00225
32112	$0.23411 \cdot N(0.12312, 0.05974) + 0.46214 \cdot N(0.17743, 0.05179) + 0.30375 \cdot N(0.24082, 0.06182)$	0.00190
23112	$0.35682 \cdot N(0.28226, 0.06594) + 0.48455 \cdot N(0.32889, 0.05665) + 0.15863 \cdot N(0.36456, 0.07928)$	0.00209
21312	$0.02335 \cdot N(0.43696, 0.09757) + 0.90229 \cdot N(0.42881, 0.05946) + 0.07435 \cdot N(0.52869, 0.03733)$	0.00198
21132	$0.09206 \cdot N(0.14510, 0.09417) + 0.62986 \cdot N(0.12755, 0.06067) + 0.27807 \cdot N(0.15454, 0.06350)$	0.00219
21123	$0.15122 \cdot N(0.17204, 0.05432) + 0.67859 \cdot N(0.23535, 0.05615) + 0.17019 \cdot N(0.29155, 0.06224)$	0.00215
21213	$0.31215 \cdot N(0.32772, 0.06323) + 0.07364 \cdot N(0.31517, 0.03187) + 0.61422 \cdot N(0.31032, 0.07042)$	0.00218
21231	$0.45479 \cdot N(0.15822, 0.07093) + 0.47245 \cdot N(0.16938, 0.05492) + 0.07276 \cdot N(0.21761, 0.08335)$	0.00173
21321	$0.06215 \cdot N(0.39733, 0.10866) + 0.77839 \cdot N(0.37812, 0.06434) + 0.15947 \cdot N(0.39899, 0.06344)$	0.00287
23121	$0.37676 \cdot N(0.25646, 0.07625) + 0.23792 \cdot N(0.26048, 0.05196) + 0.38532 \cdot N(0.27419, 0.07855)$	0.00262
32121	$0.27459 \cdot N(0.10826, 0.07337) + 0.51446 \cdot N(0.13361, 0.05977) + 0.21095 \cdot N(0.15056, 0.08760)$	0.00199
31221	$0.33580 \cdot N(0.16236, 0.07330) + 0.51465 \cdot N(0.19734, 0.06041) + 0.14956 \cdot N(0.23651, 0.07751)$	0.00208
13221	$0.38164 \cdot N(0.29189, 0.07642) + 0.55545 \cdot N(0.29415, 0.06011) + 0.06291 \cdot N(0.30925, 0.10061)$	0.00258
12321	$0.01446 \cdot N(0.19561, 0.04733) + 0.04083 \cdot N(0.29103, 0.02328) + 0.94471 \cdot N(0.35590, 0.06579)$	0.00334
12231	$0.27139 \cdot N(0.13369, 0.07764) + 0.17860 \cdot N(0.12832, 0.04790) + 0.55001 \cdot N(0.13829, 0.06284)$	0.00212
12213	$0.28106 \cdot N(0.26760, 0.07372) + 0.36781 \cdot N(0.28153, 0.04952) + 0.35112 \cdot N(0.30655, 0.06991)$	0.00236
22113	$0.31756 \cdot N(0.22894, 0.04367) + 0.41522 \cdot N(0.25403, 0.08370) + 0.26722 \cdot N(0.29450, 0.04500)$	0.00210
22131	$0.37498 \cdot N(0.09599, 0.04878) + 0.05129 \cdot N(0.16016, 0.02066) + 0.57373 \cdot N(0.10744, 0.07764)$	0.00184
22311	$0.31207 \cdot N(0.35878, 0.05478) + 0.52277 \cdot N(0.41403, 0.05077) + 0.16516 \cdot N(0.46231, 0.06394)$	0.00205
23211	$0.18406 \cdot N(0.32652, 0.07627) + 0.80537 \cdot N(0.35229, 0.06343) + 0.01057 \cdot N(0.48334, 0.05450)$	0.00210
32211	$0.33436 \cdot N(0.19206, 0.06545) + 0.28047 \cdot N(0.20396, 0.05362) + 0.38517 \cdot N(0.24200, 0.06478)$	0.00256
11233	$0.17071 \cdot N(-0.05212, 0.07169) + 0.76858 \cdot N(-0.01780, 0.06008) + 0.06070 \cdot N(-0.00512, 0.09077)$	0.00268
11323	$0.17233 \cdot N(0.15479, 0.06074) + 0.70754 \cdot N(0.19687, 0.05901) + 0.12013 \cdot N(0.19262, 0.08349)$	0.00241
13123	$0.13205 \cdot N(0.06066, 0.09946) + 0.18564 \cdot N(0.02426, 0.05166) + 0.68231 \cdot N(0.08792, 0.06091)$	0.00175
31123	$0.15232 \cdot N(-0.08149, 0.05985) + 0.43537 \cdot N(-0.02110, 0.07811) + 0.41231 \cdot N(-0.01776, 0.05095)$	0.00255
31132	$0.11237 \cdot N(-0.15474, 0.03552) + 0.65030 \cdot N(-0.12864, 0.07337) + 0.23733 \cdot N(-0.10547, 0.05539)$	0.00198
13132	$0.06213 \cdot N(-0.03538, 0.08495) + 0.77564 \cdot N(-0.02353, 0.06367) + 0.16223 \cdot N(-0.02089, 0.08093)$	0.00218
11332	$0.07404 \cdot N(0.03572, 0.08146) + 0.68863 \cdot N(0.08966, 0.05909) + 0.23732 \cdot N(0.12632, 0.06842)$	0.00202
13312	$0.01490 \cdot N(0.12556, 0.04503) + 0.95405 \cdot N(0.28060, 0.06257) + 0.03106 \cdot N(0.22368, 0.01981)$	0.00204
31312	$0.00382 \cdot N(-0.01793, 0.03972) + 0.28836 \cdot N(0.18682, 0.07851) + 0.70782 \cdot N(0.16665, 0.06171)$	0.00231
33112	$0.01482 \cdot N(-0.08370, 0.02664) + 0.87886 \cdot N(0.05742, 0.06459) + 0.10632 \cdot N(0.06503, 0.09526)$	0.00295

	Approximate mixture distribution	KL
31213	$0.10652 \cdot N(-0.00143, 0.05349) + 0.68812 \cdot N(0.06199, 0.05874) + 0.20536 \cdot N(0.06640, 0.08366)$	0.00204
31231	$0.05047 \cdot N(-0.19040, 0.05197) + 0.40784 \cdot N(-0.10274, 0.05426) + 0.54169 \cdot N(-0.08019, 0.06996)$	0.00183
31321	$0.03165 \cdot N(0.12806, 0.13133) + 0.93453 \cdot N(0.11538, 0.06542) + 0.03382 \cdot N(0.22228, 0.03514)$	0.00311
33121	$0.27360 \cdot N(-0.00542, 0.08511) + 0.03232 \cdot N(-0.02079, 0.02395) + 0.69408 \cdot N(0.00492, 0.06620)$	0.00234
13321	$0.41222 \cdot N(0.22156, 0.07697) + 0.52801 \cdot N(0.21337, 0.05149) + 0.05977 \cdot N(0.30197, 0.02760)$	0.00235
13231	$0.00687 \cdot N(-0.09675, 0.00508) + 0.97201 \cdot N(0.00626, 0.06614) + 0.02112 \cdot N(0.03823, 0.11894)$	0.00261
13213	$0.14420 \cdot N(0.11688, 0.07518) + 0.81404 \cdot N(0.16118, 0.06249) + 0.04176 \cdot N(0.21775, 0.09828)$	0.00269
12313	$0.02195 \cdot N(0.08406, 0.04412) + 0.94750 \cdot N(0.21437, 0.06220) + 0.03055 \cdot N(0.28090, 0.07911)$	0.00191
12331	$0.07129 \cdot N(0.01079, 0.08611) + 0.90796 \cdot N(0.06379, 0.06195) + 0.02075 \cdot N(0.20250, 0.04173)$	0.00240
12133	$0.09379 \cdot N(-0.11129, 0.09547) + 0.89652 \cdot N(-0.08406, 0.06181) + 0.00969 \cdot N(0.06373, 0.01168)$	0.00278
21133	$0.05098 \cdot N(-0.06224, 0.02908) + 0.94473 \cdot N(-0.05606, 0.06508) + 0.00428 \cdot N(-0.03064, 0.18332)$	0.00290
21313	$0.08265 \cdot N(0.23567, 0.08783) + 0.87225 \cdot N(0.24452, 0.05962) + 0.04510 \cdot N(0.25823, 0.08225)$	0.00294
23113	$0.41286 \cdot N(0.10646, 0.06685) + 0.54275 \cdot N(0.14363, 0.06408) + 0.04438 \cdot N(0.12625, 0.11065)$	0.00238
32113	$0.00798 \cdot N(-0.19146, 0.04058) + 0.14094 \cdot N(0.02043, 0.08135) + 0.85108 \cdot N(-0.00869, 0.06257)$	0.00204
32131	$0.02222 \cdot N(-0.26718, 0.06141) + 0.96410 \cdot N(-0.15660, 0.06354) + 0.01368 \cdot N(-0.03083, 0.05716)$	0.00178
23131	$0.21688 \cdot N(-0.03907, 0.07561) + 0.48869 \cdot N(-0.02477, 0.06095) + 0.29443 \cdot N(-0.00652, 0.07315)$	0.00343
21331	$0.30843 \cdot N(0.06985, 0.04019) + 0.50410 \cdot N(0.09149, 0.07688) + 0.18747 \cdot N(0.13958, 0.03945)$	0.00248
23311	$0.01883 \cdot N(0.16449, 0.02711) + 0.08347 \cdot N(0.27640, 0.09753) + 0.89771 \cdot N(0.27654, 0.06058)$	0.00206
32311	$0.28603 \cdot N(0.13461, 0.07456) + 0.62870 \cdot N(0.14299, 0.05788) + 0.08527 \cdot N(0.16998, 0.08007)$	0.00218
33211	$0.28864 \cdot N(0.04298, 0.04867) + 0.43028 \cdot N(0.10861, 0.05089) + 0.28108 \cdot N(0.09260, 0.08221)$	0.00213
11333	$0.31818 \cdot N(-0.11066, 0.07410) + 0.30759 \cdot N(-0.08845, 0.05257) + 0.37424 \cdot N(-0.08864, 0.06957)$	0.00236
13133	$0.34240 \cdot N(-0.25936, 0.06370) + 0.64138 \cdot N(-0.19481, 0.05676) + 0.01622 \cdot N(-0.05279, 0.02602)$	0.00183
31133	$0.31993 \cdot N(-0.32349, 0.08160) + 0.41342 \cdot N(-0.33706, 0.05060) + 0.26665 \cdot N(-0.27759, 0.04908)$	0.00169
13313	$0.29231 \cdot N(0.07246, 0.07016) + 0.65781 \cdot N(0.08492, 0.05727) + 0.04988 \cdot N(0.16156, 0.05584)$	0.00242
31313	$0.00541 \cdot N(-0.21822, 0.04552) + 0.67309 \cdot N(-0.02808, 0.06434) + 0.32149 \cdot N(0.01037, 0.06537)$	0.00229
33113	$0.71242 \cdot N(-0.13444, 0.07780) + 0.27616 \cdot N(-0.13262, 0.05349) + 0.01142 \cdot N(-0.13781, 0.00255)$	0.00312
31331	$0.11396 \cdot N(-0.21622, 0.03178) + 0.34181 \cdot N(-0.14766, 0.04477) + 0.54422 \cdot N(-0.17179, 0.07811)$	0.00224
33131	$0.21299 \cdot N(-0.32554, 0.06993) + 0.59557 \cdot N(-0.27712, 0.05742) + 0.19143 \cdot N(-0.26929, 0.08057)$	0.00229
13331	$0.01814 \cdot N(-0.21509, 0.02815) + 0.49761 \cdot N(-0.09323, 0.05107) + 0.48426 \cdot N(-0.03245, 0.05877)$	0.00200
33311	$0.00260 \cdot N(-0.06355, 0.11576) + 0.18213 \cdot N(-0.02853, 0.06132) + 0.81527 \cdot N(0.02033, 0.06231)$	0.00196
12222	$0.03190 \cdot N(0.58393, 0.00750) + 0.72059 \cdot N(0.57958, 0.06391) + 0.24751 \cdot N(0.61577, 0.06001)$	0.00292
21222	$0.06490 \cdot N(0.58322, 0.04286) + 0.62493 \cdot N(0.62474, 0.07231) + 0.31017 \cdot N(0.62487, 0.04565)$	0.00163
22122	$0.00169 \cdot N(0.37172, 0.04804) + 0.54834 \cdot N(0.56039, 0.06937) + 0.44997 \cdot N(0.55805, 0.05372)$	0.00251
22212	$0.03641 \cdot N(0.57921, 0.01797) + 0.82495 \cdot N(0.64399, 0.06838) + 0.13863 \cdot N(0.65432, 0.03309)$	0.00222
22221	$0.06373 \cdot N(0.49459, 0.04773) + 0.11338 \cdot N(0.64708, 0.06599) + 0.82289 \cdot N(0.58896, 0.05504)$	0.00190
12223	$0.11495 \cdot N(0.20578, 0.06339) + 0.05256 \cdot N(0.10965, 0.07696) + 0.83249 \cdot N(0.16468, 0.06050)$	0.00235
12232	$0.00477 \cdot N(-0.06854, 0.00295) + 0.91534 \cdot N(0.07062, 0.06119) + 0.07988 \cdot N(0.07195, 0.10129)$	0.00279
12322	$0.21857 \cdot N(0.27303, 0.04220) + 0.75842 \cdot N(0.28455, 0.07545) + 0.02301 \cdot N(0.34149, 0.01617)$	0.00172
13222	$0.08606 \cdot N(0.24009, 0.04888) + 0.75541 \cdot N(0.22658, 0.07168) + 0.15853 \cdot N(0.21963, 0.04693)$	0.00262
31222	$0.26759 \cdot N(0.10350, 0.05503) + 0.27305 \cdot N(0.11888, 0.08384) + 0.45936 \cdot N(0.14297, 0.06189)$	0.00238
32122	$0.06080 \cdot N(-0.05754, 0.05831) + 0.65598 \cdot N(0.04876, 0.05634) + 0.28322 \cdot N(0.12838, 0.06012)$	0.00176
23122	$0.12972 \cdot N(0.17100, 0.04124) + 0.11485 \cdot N(0.23177, 0.03636) + 0.75543 \cdot N(0.19563, 0.07849)$	0.00179
21322	$0.04729 \cdot N(0.28042, 0.02522) + 0.92011 \cdot N(0.31634, 0.06760) + 0.03259 \cdot N(0.34904, 0.10715)$	0.00226
21232	$0.00968 \cdot N(0.02878, 0.08363) + 0.63046 \cdot N(0.09871, 0.06855) + 0.35986 \cdot N(0.10623, 0.05149)$	0.00193
21223	$0.13187 \cdot N(0.17204, 0.03638) + 0.82704 \cdot N(0.20031, 0.07086) + 0.04110 \cdot N(0.23282, 0.01796)$	0.00187
22123	$0.54377 \cdot N(0.13673, 0.05750) + 0.43337 \cdot N(0.12938, 0.07261) + 0.02286 \cdot N(0.22689, 0.06967)$	0.00215
22132	$0.05640 \cdot N(-0.01539, 0.06689) + 0.70601 \cdot N(0.03144, 0.06114) + 0.23759 \cdot N(0.07587, 0.06858)$	0.00215
22312	$0.09852 \cdot N(0.33060, 0.09532) + 0.01113 \cdot N(0.23060, 0.01793) + 0.89035 \cdot N(0.33808, 0.06163)$	0.00187
23212	$0.03399 \cdot N(0.16830, 0.06028) + 0.87854 \cdot N(0.27706, 0.05695) + 0.08746 \cdot N(0.36533, 0.04737)$	0.00189
32212	$0.02419 \cdot N(0.05935, 0.07139) + 0.68860 \cdot N(0.15565, 0.06467) + 0.28721 \cdot N(0.14276, 0.05538)$	0.00220

	Approximate mixture distribution	KL
22213	$0.13839 \cdot N(0.19465, 0.07514) + 0.85460 \cdot N(0.22454, 0.06143) + 0.00700 \cdot N(0.40733, 0.05830)$	0.00223
22231	$0.01880 \cdot N(0.02338, 0.01139) + 0.76514 \cdot N(0.06995, 0.05817) + 0.21605 \cdot N(0.07518, 0.08384)$	0.00177
22321	$0.01533 \cdot N(0.19760, 0.07795) + 0.73109 \cdot N(0.28424, 0.06932) + 0.25359 \cdot N(0.28604, 0.05204)$	0.00224
23221	$0.00956 \cdot N(0.33588, 0.11877) + 0.45526 \cdot N(0.22090, 0.07742) + 0.53518 \cdot N(0.23183, 0.06208)$	0.00291
32221	$0.03215 \cdot N(0.00320, 0.06640) + 0.65522 \cdot N(0.09220, 0.06315) + 0.31263 \cdot N(0.10634, 0.06804)$	0.00207
12233	$0.33368 \cdot N(-0.14167, 0.05870) + 0.32499 \cdot N(-0.11515, 0.07510) + 0.34133 \cdot N(-0.10555, 0.05309)$	0.00208
12323	$0.01054 \cdot N(-0.01068, 0.01184) + 0.55170 \cdot N(0.09779, 0.05476) + 0.43776 \cdot N(0.09276, 0.07643)$	0.00241
13223	$0.08649 \cdot N(-0.04400, 0.06540) + 0.74039 \cdot N(0.02989, 0.05726) + 0.17312 \cdot N(0.09794, 0.05922)$	0.00210
31223	$0.15198 \cdot N(-0.09314, 0.06254) + 0.73036 \cdot N(-0.06025, 0.06139) + 0.11765 \cdot N(-0.05869, 0.09383)$	0.00224
31232	$0.01501 \cdot N(-0.23784, 0.00346) + 0.19831 \cdot N(-0.15142, 0.08306) + 0.78668 \cdot N(-0.16237, 0.06130)$	0.00331
13232	$0.25276 \cdot N(-0.05633, 0.07853) + 0.07497 \cdot N(-0.08495, 0.03306) + 0.67227 \cdot N(-0.05680, 0.06004)$	0.00270
12332	$0.19679 \cdot N(-0.01940, 0.07877) + 0.58437 \cdot N(0.00050, 0.05976) + 0.21884 \cdot N(0.00051, 0.07859)$	0.00250
13322	$0.24011 \cdot N(0.09363, 0.05730) + 0.49823 \cdot N(0.15823, 0.05010) + 0.26166 \cdot N(0.20405, 0.06044)$	0.00262
31322	$0.04082 \cdot N(-0.02601, 0.08903) + 0.93737 \cdot N(0.05388, 0.06453) + 0.02181 \cdot N(0.21022, 0.04123)$	0.00216
33122	$0.00109 \cdot N(-0.11804, 0.07265) + 0.60610 \cdot N(-0.06691, 0.06482) + 0.39281 \cdot N(-0.06364, 0.07948)$	0.00276
21233	$0.16972 \cdot N(-0.11798, 0.07189) + 0.81585 \cdot N(-0.08613, 0.06071) + 0.01443 \cdot N(0.06367, 0.03941)$	0.00175
21323	$0.00272 \cdot N(0.08384, 0.06291) + 0.87269 \cdot N(0.12526, 0.05885) + 0.12460 \cdot N(0.12220, 0.09702)$	0.00255
23123	$0.16223 \cdot N(-0.04536, 0.06736) + 0.76817 \cdot N(0.01598, 0.06274) + 0.06960 \cdot N(0.04637, 0.08880)$	0.00241
32123	$0.61810 \cdot N(-0.13233, 0.07023) + 0.03348 \cdot N(-0.10103, 0.11755) + 0.34842 \cdot N(-0.12275, 0.05699)$	0.00239
32132	$0.32940 \cdot N(-0.26831, 0.06021) + 0.55970 \cdot N(-0.21341, 0.05441) + 0.11090 \cdot N(-0.14222, 0.05591)$	0.00213
23132	$0.25467 \cdot N(-0.13617, 0.06286) + 0.63438 \cdot N(-0.07515, 0.06047) + 0.11095 \cdot N(-0.05598, 0.07376)$	0.00197
21332	$0.01422 \cdot N(-0.04177, 0.01175) + 0.49630 \cdot N(0.02745, 0.07425) + 0.48948 \cdot N(0.03006, 0.05312)$	0.00245
23312	$0.00297 \cdot N(0.06074, 0.10286) + 0.80172 \cdot N(0.20326, 0.06485) + 0.19530 \cdot N(0.22988, 0.05617)$	0.00210
32312	$0.01799 \cdot N(0.08755, 0.11512) + 0.11726 \cdot N(0.06286, 0.03894) + 0.86475 \cdot N(0.07734, 0.06834)$	0.00222
33212	$0.33454 \cdot N(-0.02500, 0.06201) + 0.44097 \cdot N(0.02001, 0.04382) + 0.22449 \cdot N(0.08187, 0.05343)$	0.00218
32213	$0.14031 \cdot N(-0.08199, 0.07026) + 0.85340 \cdot N(-0.03657, 0.05974) + 0.00630 \cdot N(0.12370, 0.02604)$	0.00191
32231	$0.21572 \cdot N(-0.19011, 0.03885) + 0.77477 \cdot N(-0.19565, 0.06973) + 0.00951 \cdot N(-0.10042, 0.00385)$	0.00279
32321	$0.68818 \cdot N(0.01341, 0.06155) + 0.09875 \cdot N(0.01614, 0.08970) + 0.21307 \cdot N(0.05213, 0.05884)$	0.00284
33221	$0.00301 \cdot N(0.03384, 0.15951) + 0.90899 \cdot N(-0.03477, 0.06963) + 0.08800 \cdot N(-0.02911, 0.03431)$	0.00222
23321	$0.46651 \cdot N(0.13431, 0.05064) + 0.18555 \cdot N(0.14270, 0.08659) + 0.34793 \cdot N(0.19281, 0.05588)$	0.00187
23231	$0.54675 \cdot N(-0.06916, 0.05411) + 0.05963 \cdot N(0.01712, 0.03248) + 0.39362 \cdot N(-0.05898, 0.07994)$	0.00166
23213	$0.78187 \cdot N(0.07774, 0.06288) + 0.12719 \cdot N(0.12954, 0.04597) + 0.09094 \cdot N(0.15083, 0.07132)$	0.00186
22313	$0.39241 \cdot N(0.14775, 0.07200) + 0.17040 \cdot N(0.13980, 0.04190) + 0.43720 \cdot N(0.15262, 0.06335)$	0.00254
22331	$0.24798 \cdot N(-0.00337, 0.07983) + 0.73953 \cdot N(-0.00268, 0.05834) + 0.01249 \cdot N(0.08177, 0.01141)$	0.00218
22133	$0.01416 \cdot N(-0.25997, 0.08576) + 0.85505 \cdot N(-0.15067, 0.06231) + 0.13079 \cdot N(-0.13470, 0.08069)$	0.00200
12333	$0.07220 \cdot N(-0.18181, 0.10072) + 0.76557 \cdot N(-0.20453, 0.06287) + 0.16223 \cdot N(-0.15480, 0.05146)$	0.00205
13233	$0.65719 \cdot N(-0.25242, 0.07460) + 0.10879 \cdot N(-0.29654, 0.03210) + 0.23402 \cdot N(-0.21919, 0.04796)$	0.00240
31233	$0.66898 \cdot N(-0.36499, 0.05870) + 0.16468 \cdot N(-0.35044, 0.09375) + 0.16634 \cdot N(-0.30501, 0.05132)$	0.00221
13323	$0.19646 \cdot N(-0.04264, 0.05144) + 0.49574 \cdot N(-0.02546, 0.06090) + 0.30780 \cdot N(-0.04133, 0.08034)$	0.00222
31323	$0.14920 \cdot N(-0.17376, 0.06599) + 0.77311 \cdot N(-0.13300, 0.06235) + 0.07769 \cdot N(-0.10495, 0.08240)$	0.00201
33123	$0.31989 \cdot N(-0.25558, 0.05609) + 0.66255 \cdot N(-0.25810, 0.07586) + 0.01756 \cdot N(-0.19901, 0.01124)$	0.00298
31332	$0.04402 \cdot N(-0.32388, 0.07420) + 0.78042 \cdot N(-0.24406, 0.05996) + 0.17556 \cdot N(-0.17382, 0.06392)$	0.00197
33132	$0.11431 \cdot N(-0.38290, 0.09084) + 0.73728 \cdot N(-0.35724, 0.06112) + 0.14842 \cdot N(-0.29877, 0.06309)$	0.00210
13332	$0.09940 \cdot N(-0.19559, 0.07260) + 0.65506 \cdot N(-0.14319, 0.05407) + 0.24554 \cdot N(-0.07525, 0.05837)$	0.00224
33312	$0.16394 \cdot N(-0.10076, 0.06507) + 0.59470 \cdot N(-0.04824, 0.05546) + 0.24136 \cdot N(-0.03723, 0.06964)$	0.00177
33213	$0.26436 \cdot N(-0.19785, 0.07323) + 0.57119 \cdot N(-0.16741, 0.05632) + 0.16445 \cdot N(-0.14434, 0.07652)$	0.00289
33231	$0.05946 \cdot N(-0.40220, 0.02928) + 0.24169 \cdot N(-0.32024, 0.08384) + 0.69884 \cdot N(-0.31581, 0.05472)$	0.00246
33321	$0.11487 \cdot N(-0.09960, 0.09559) + 0.73432 \cdot N(-0.11858, 0.05867) + 0.15081 \cdot N(-0.06134, 0.04910)$	0.00286
32313	$0.01726 \cdot N(-0.24989, 0.02983) + 0.92363 \cdot N(-0.11500, 0.06013) + 0.05911 \cdot N(-0.07977, 0.09903)$	0.00201
32331	$0.14842 \cdot N(-0.34445, 0.05203) + 0.73738 \cdot N(-0.26195, 0.05174) + 0.11421 \cdot N(-0.18789, 0.05325)$	0.00204

	Approximate mixture distribution	KL
32133	$0.33443 \cdot N(-0.45352, 0.06188) + 0.63107 \cdot N(-0.39794, 0.05439) + 0.03450 \cdot N(-0.28413, 0.04492)$	0.00199
23133	$0.04201 \cdot N(-0.35062, 0.01860) + 0.68362 \cdot N(-0.28283, 0.07714) + 0.27437 \cdot N(-0.26698, 0.04836)$	0.00271
23313	$0.21444 \cdot N(-0.02111, 0.06785) + 0.67624 \cdot N(0.02156, 0.05437) + 0.10932 \cdot N(0.06304, 0.07061)$	0.00324
23331	$0.07275 \cdot N(-0.18565, 0.03119) + 0.55542 \cdot N(-0.12271, 0.05115) + 0.37183 \cdot N(-0.13643, 0.08218)$	0.00178
21333	$0.04865 \cdot N(-0.21302, 0.07492) + 0.93138 \cdot N(-0.16207, 0.06097) + 0.01997 \cdot N(-0.06246, 0.06643)$	0.00204
13333	$0.07425 \cdot N(-0.39404, 0.07825) + 0.67669 \cdot N(-0.33598, 0.05359) + 0.24906 \cdot N(-0.26901, 0.05792)$	0.00228
31333	$0.04880 \cdot N(-0.47333, 0.08483) + 0.94467 \cdot N(-0.42453, 0.06489) + 0.00653 \cdot N(-0.26394, 0.04066)$	0.00216
33133	$0.16666 \cdot N(-0.59806, 0.04658) + 0.39964 \cdot N(-0.52288, 0.04814) + 0.43370 \cdot N(-0.53831, 0.08032)$	0.00250
33313	$0.24998 \cdot N(-0.29486, 0.05894) + 0.52298 \cdot N(-0.24187, 0.04975) + 0.22704 \cdot N(-0.19340, 0.06443)$	0.00272
33331	$0.01261 \cdot N(-0.47188, 0.09176) + 0.97676 \cdot N(-0.39375, 0.06074) + 0.01063 \cdot N(-0.24833, 0.04342)$	0.00241
22222	$0.04625 \cdot N(0.47515, 0.07877) + 0.79965 \cdot N(0.51803, 0.05584) + 0.15409 \cdot N(0.56407, 0.06883)$	0.00170
22223	$0.01787 \cdot N(-0.04015, 0.03621) + 0.86318 \cdot N(0.09511, 0.05678) + 0.11895 \cdot N(0.16429, 0.05859)$	0.00218
22232	$0.09526 \cdot N(-0.07291, 0.05906) + 0.75070 \cdot N(0.00304, 0.05380) + 0.15404 \cdot N(0.05103, 0.06054)$	0.00241
22322	$0.30627 \cdot N(0.20776, 0.07471) + 0.68765 \cdot N(0.22108, 0.06111) + 0.00607 \cdot N(0.37663, 0.04555)$	0.00266
23222	$0.34467 \cdot N(0.12536, 0.06226) + 0.42518 \cdot N(0.17424, 0.05131) + 0.23015 \cdot N(0.18978, 0.07521)$	0.00210
32222	$0.61359 \cdot N(0.01921, 0.06156) + 0.36044 \cdot N(0.04253, 0.06844) + 0.02598 \cdot N(0.03813, 0.11999)$	0.00256
22233	$0.01349 \cdot N(-0.31934, 0.06309) + 0.95365 \cdot N(-0.18691, 0.06057) + 0.03286 \cdot N(-0.12131, 0.07637)$	0.00238
22323	$0.04983 \cdot N(-0.02494, 0.03561) + 0.56343 \cdot N(0.03116, 0.05048) + 0.38674 \cdot N(0.02743, 0.07816)$	0.00227
23223	$0.09123 \cdot N(-0.11897, 0.04121) + 0.61289 \cdot N(-0.02961, 0.05257) + 0.29588 \cdot N(-0.00372, 0.07751)$	0.00245
32223	$0.03717 \cdot N(-0.21487, 0.08132) + 0.87440 \cdot N(-0.16243, 0.05888) + 0.08843 \cdot N(-0.11249, 0.06713)$	0.00212
32232	$0.29868 \cdot N(-0.28671, 0.06441) + 0.67638 \cdot N(-0.25063, 0.05556) + 0.02495 \cdot N(-0.12613, 0.04117)$	0.00174
23232	$0.11675 \cdot N(-0.15577, 0.08283) + 0.45020 \cdot N(-0.13707, 0.05380) + 0.43304 \cdot N(-0.10624, 0.06279)$	0.00246
22332	$0.03217 \cdot N(-0.15234, 0.07774) + 0.81851 \cdot N(-0.06955, 0.06160) + 0.14933 \cdot N(-0.05477, 0.07073)$	0.00184
23322	$0.14606 \cdot N(0.08540, 0.09010) + 0.54940 \cdot N(0.07865, 0.05785) + 0.30454 \cdot N(0.11028, 0.05958)$	0.00258
32322	$0.30664 \cdot N(-0.03213, 0.05894) + 0.69109 \cdot N(-0.05433, 0.06989) + 0.00228 \cdot N(0.17152, 0.04436)$	0.00159
33222	$0.00195 \cdot N(-0.21569, 0.13577) + 0.32233 \cdot N(-0.10488, 0.05135) + 0.67572 \cdot N(-0.10021, 0.07353)$	0.00200
22333	$0.16926 \cdot N(-0.29530, 0.06959) + 0.77906 \cdot N(-0.25614, 0.05910) + 0.05168 \cdot N(-0.18525, 0.05752)$	0.00193
23233	$0.07744 \cdot N(-0.36285, 0.08247) + 0.82097 \cdot N(-0.31186, 0.06679) + 0.10159 \cdot N(-0.30953, 0.03466)$	0.00223
32233	$0.79231 \cdot N(-0.46415, 0.05840) + 0.14425 \cdot N(-0.41045, 0.03882) + 0.06344 \cdot N(-0.35512, 0.04972)$	0.00207
23323	$0.05762 \cdot N(-0.12744, 0.07184) + 0.93879 \cdot N(-0.10108, 0.06109) + 0.00359 \cdot N(0.06964, 0.02635)$	0.00199
32323	$0.44804 \cdot N(-0.24188, 0.05283) + 0.54082 \cdot N(-0.22942, 0.07192) + 0.01114 \cdot N(-0.14136, 0.00451)$	0.00338
33223	$0.02154 \cdot N(-0.34871, 0.08614) + 0.97127 \cdot N(-0.28915, 0.06436) + 0.00719 \cdot N(-0.11822, 0.02950)$	0.00304
32332	$0.33587 \cdot N(-0.34034, 0.07582) + 0.60567 \cdot N(-0.33478, 0.05989) + 0.05846 \cdot N(-0.26078, 0.06179)$	0.00208
33232	$0.68202 \cdot N(-0.38900, 0.07053) + 0.10994 \cdot N(-0.41474, 0.03352) + 0.20804 \cdot N(-0.37685, 0.04588)$	0.00224
23332	$0.11023 \cdot N(-0.29256, 0.03902) + 0.56234 \cdot N(-0.20582, 0.04524) + 0.32743 \cdot N(-0.15373, 0.05890)$	0.00160
33322	$0.23858 \cdot N(-0.17465, 0.05500) + 0.62938 \cdot N(-0.17340, 0.06278) + 0.13204 \cdot N(-0.16748, 0.08876)$	0.00258
23333	$0.01268 \cdot N(-0.54096, 0.02589) + 0.81506 \cdot N(-0.38905, 0.05960) + 0.17227 \cdot N(-0.37718, 0.08691)$	0.00243
32333	$0.17000 \cdot N(-0.55692, 0.07824) + 0.77438 \cdot N(-0.52246, 0.05799) + 0.05562 \cdot N(-0.43092, 0.05409)$	0.00222
33233	$0.34602 \cdot N(-0.58819, 0.04448) + 0.57299 \cdot N(-0.58239, 0.07439) + 0.08099 \cdot N(-0.50502, 0.03991)$	0.00230
33323	$0.78169 \cdot N(-0.36584, 0.07023) + 0.05851 \cdot N(-0.38592, 0.02067) + 0.15980 \cdot N(-0.34523, 0.03845)$	0.00298
33332	$0.24959 \cdot N(-0.50918, 0.06052) + 0.60808 \cdot N(-0.45190, 0.05494) + 0.14232 \cdot N(-0.41189, 0.07178)$	0.00206
33333	$0.28427 \cdot N(-0.65405, 0.08379) + 0.23017 \cdot N(-0.69842, 0.04168) + 0.48556 \cdot N(-0.62551, 0.04864)$	0.00209

Table 3.4: Summary of the three-component mixture distributions which approximate the posterior distributions of the EQ-5D-3L scores.

The target distributions have been approximated using three-component mixtures. Fewer than three components may not provide a satisfactory approximation to the original distribution.

If more than three components are used, then the algorithm allows for a better approximation, but the computational complexity increases. The choice of the starting values is crucial as well, because a poor choice of them might result in unsatisfactory results. In particular, when using fewer than three components or more than three components, the algorithm faces some issues: in the first case because the few components are not enough for the mixture to approximate the form of the target distribution, whilst in the second case the algorithm might become very sensitive to the selection of the starting values, which will regularly result into non-convergence. For example, Figure 3.8 provides a visual inspection for the improvement of the approximation of the distribution of state `31113` starting with one component, then with three, and finally with five components. The inclusion of the third component substantially improves the approximation of the mixture distribution (the KL divergence for the case of the three-component distribution is almost one third of the KL divergence for the case of the one-component distribution). Nevertheless, even when adding two extra components (fourth and fifth component), although the complexity of the algorithm increases, the further improvement of the approximation is relatively small. Given the objective to use as many components as possible to obtain good approximations without having too many of them for the algorithm to successfully converge for every state, the decision was to use three components; this is in agreement with the conclusions of Schmidli et al. (2014). Figure 3.9 also visually shows the KL divergence for the case of one, three, and five components mixtures for state `31113`.

3.4 Note: regarding the correlation between states

A new tariff for the EQ-5D-3L utility scores was derived in this chapter. In particular, I have shown that there are two ways to summarise and use the findings of this chapter. The first one is to use the $4,000 \times 243$ MCMC values that were derived in section 3.2. The second one is to consider the mixture distributions which were derived in section 3.3 (e.g. Table 3.4 represents the 243 three-component mixture distributions which approximate the posterior distributions of the EQ-5D-3L scores).

Given the way the Bayesian model was defined in section 3.2, it would be reasonable to expect the existence of some correlation between the different EQ-5D-3L states. This is accounted in the main Bayesian model (which was described in section 3.2), as equation (3.1) shows how to obtain the utility score of the s-th EQ-5D-3L state in particular, but in fact a vector of utility scores (whose size is 243) is obtained at each MCMC iteration c (the vector is computed using the c-th MCMC values of the β and $\xi^{(new)}$ parameters). Hence, there is a dependence between

the health state utilities which were obtained under the c-th iteration of the main Bayesian model. Nevertheless, this is not the case for the three-component mixture model in which each EQ-5D-3L state utility is sampled independently from its distribution. However, if there is no substantial dependence and if it does not matter to account for correlation between different states, then this would be in favour of using the results derived from the the three-component mixture model as they would be considered to not be substantially different than the results derived by the main Bayesian model.

The impact of the correlation can be examined by considering a hypothetical two-arms randomised trial which lasts for 1 year and the EQ-5D-3L is administrated at the end of the year. Let $\theta_{\mathbf{j}}^{\mathbf{T}} = (\theta_{1,j}, ..., \theta_{S,j})$ capture the probabilities of individuals of the j-th group of the trial (for $j = 1, 2$) falling into each of the EQ-5D-3L health states, where $\theta_{s,j}$ is the probability of an individual in group j falling into health state s (for $s = 1, ..., 243$). Then, $\boldsymbol{\Delta}_\theta = \theta_1 - \theta_2$ and let $\bar{\mathbf{u}}^T = (\bar{u_1}, ..., \bar{u_S})$, where $\bar{u_s}$ is the mean utility for the s-th EQ-5D-3L state. Furthermore, Pullenayegum, Chan and Xie (2015) report that if Δ_e is the difference in mean QALYs between the groups, then we have: $\Delta_e = \boldsymbol{\Delta_\theta}^T \cdot \bar{\mathbf{u}}$. Moreover, for the variance of Δ_e given $\boldsymbol{\Delta}_\theta$, we have: $var(\Delta_e|\boldsymbol{\Delta}_\theta) = \boldsymbol{\Delta_\theta}^T \cdot var(\bar{\mathbf{u}}) \cdot \boldsymbol{\Delta}_\theta$, where $var(\bar{\mathbf{u}})$ can be computed by the $4,000 \times 243$ matrix of utilities that was derived in section 3.2. Specifically, the aforementioned impact of the correlation can be examined by comparing VAR_1 with VAR_2, where $\text{VAR}_1 = \boldsymbol{\Delta_\theta}^T \cdot var(\bar{\mathbf{u}}) \cdot \boldsymbol{\Delta}_\theta$ and $\text{VAR}_2 = \boldsymbol{\Delta_\theta}^T \cdot diag(var(\bar{\mathbf{u}})) \cdot \boldsymbol{\Delta}_\theta$, while $diag(var(\bar{\mathbf{u}}))$ is a diagonal matrix with the same dimensions as matrix $var(\bar{\mathbf{u}})$ and the same diagonal entries as matrix $var(\bar{\mathbf{u}})$.

In particular, since the probability vectors θ_1 and θ_2 will vary from one trial to another, $1,000,000$ pairs of θ_1 and θ_2 are simulated from the Dirichlet distribution, where the concentration parameters are $1/S$,...,$1/S$. Therefore, we obtain $1,000,000$ pairs of VAR_1 and VAR_2. The summary statistics of $\text{VAR}_2/\text{VAR}_1$ are shown in Table 3.5. Figure 3.10 illustrates the kernel density plot of $\text{VAR}_2/\text{VAR}_1$.

From the summary statistics and from the density plot we can see that the values of $\text{VAR}_2/\text{VAR}_1$ are substantially close to 1 for most of the cases. The values of $\text{VAR}_1 - \text{VAR}_2$ are substantially close to 0. The interpretation is that VAR_1 and VAR_2 are found to be approximately equal to each other for most of the cases. Hence, the conclusion is that in general the three-component mixtures derived in section 3.3 can be used and produce results which most of the times are similar to those produced when we use the main Bayesian model of section 3.2. This remains an approximation of the results obtained from the main Bayesian model, though, and thus the use of the main Bayesian model could be the first preference.

Summary statistics of VAR_2/VAR_1							
Min.	25th Prm.	250th Prm.	Median	Mean	750th Prm.	975th Prm.	Max
0.5477	0.8564	0.9712	1.0361	1.0422	1.1071	1.2609	1.5851

Table 3.5: The summary statistics of VAR_2/VAR_1: minimum (Min.), 25th permille (25th Prm.), 250th permille (250th Prm.), median, mean, 750th permille (750th Prm.), 975th permille (975th Prm.), and maximum (Max).

3.5 Note: dealing with utility values greater than 1

As it stated previously, we assumed that the utility value of the state of perfect health is equal to 1 and that all utility values of the EQ-5D-3L states fall within $[-1, 1]$. In this chapter the utility distributions of the EQ-5D-3L health states were derived as unbounded distributions. There is a subtlety that if a state is considered to have utility values close to the bounds of $[-1, 1]$, it could be possible that if we sample from such a distribution we might obtain a value outside of $[-1, 1]$. Regarding the distributions of the states which were evaluated as the least desirable ones (i.e.those that are the furthest away from perfect health), even the lowest values of the MCMC samples are considerably away from the lower bound of $[-1, 1]$. Nevertheless, regarding the distributions of the states which were evaluated to be close to perfect health, there are a couple of states for which it is likely that in rare cases we might sample a utility value which is just slightly greater than 1. These "upper extreme" values from these distributions, which are slightly greater than 1 are also counter-weighted by the "lower extreme" values which can be sampled from the same distributions. In any case, the results of this thesis were found to be negligible to whether these aforementioned very few utility values were bounded by 1 or not. However, the decision was to prevent the possibility of using utility values which are greater than 1.

Therefore, the statistics of the MCMC samples (which are presented in tables, figures, etc) are derived by converting any rare values which are slightly greater than 1 to 1. This is done for consistency reasons even though the changes to the results due to potentially not bounding the values are negligible. Similarly, in the rare case that a value is sampled directly from the approximate posterior distributions which have been derived and it is found to be greater than 1, then it is converted to 1 before it is used in any of our further calculations.

Section 3.3 is an exception, though, as the values of $u_s^{(1)}, u_s^{(2)}, ..., u_s^{(C)}$ are not bounded by 1

during the procedure for the derivation of state's s approximate utility distribution and when the corresponding KL divergence is calculated. The approximate distributions which were obtained in section 3.3 are unbounded and so it is theoretically possibly to sample utility values greater than 1 for a couple of states. The reason why the utility values are not bounded for the aforementioned methods in section 3.3 is because of the desire to achieve the best-approximation possible of the "original" distribution before "changing" it by bounding values by 1 (which could unnecessary influence the convergence of the algorithm). Nevertheless, as it was already stated, once the approximate distributions are obtained, in the rare case that a value is sampled which is greater than 1, then it is converted to 1 before it is used for any of our further calculations. As a result, in the entire thesis all the results are based on utility values which (in the end) lie within $[-1, 1]$.

3.6 Summary

In this chapter I discussed the Bayesian model, the approximation of the posteriors by mixtures of normal distributions, the derivation of the parameters of these mixtures via optimisation techniques, and the computation of the KL divergence to assess the quality of the approximations. MCMC samples were generated from the posterior distributions of the EQ-5D-3L state utilities; statistics and credible intervals were provided. The posterior distributions were described in a parametric form as three-component mixtures of Normals. These distributions were mostly considered to be a good way to approximate the results of the MCMC algorithm. Furthermore, we noted that any values sampled from such distributions should be bounded by 1. The KL divergence values between the actual posterior distributions and their approximations by mixtures were calculated empirically. Chapter 4 shows how these methods can be applied in practice when doing CUA.

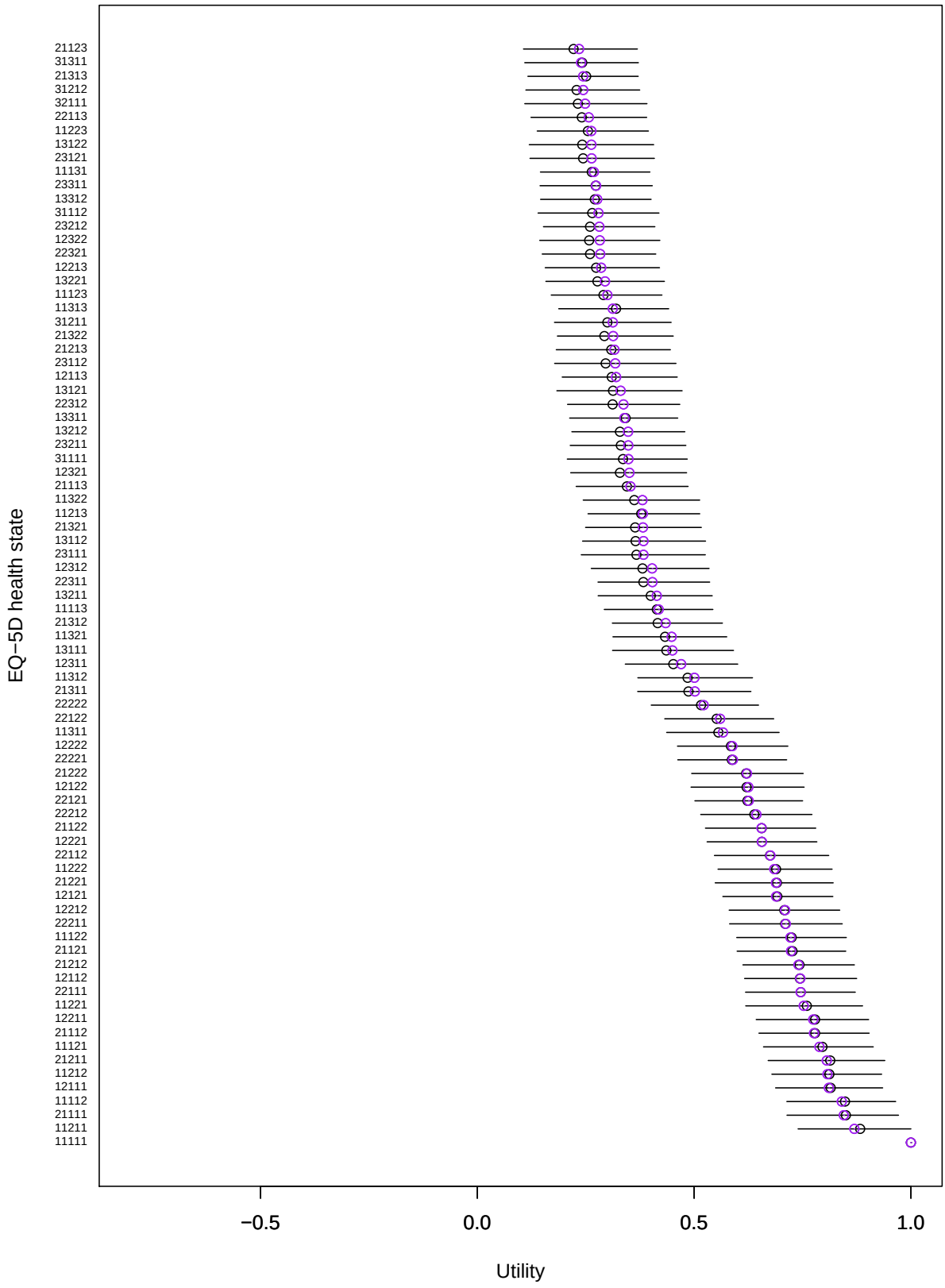


Figure 3.1: 95% credible intervals for the posterior distributions of the 81 EQ-5D-3L health states which were valued the highest ordered with respect to the posterior medians, shown as purple circles. The MVH point estimates of the utilities of the states are shown as black circles.

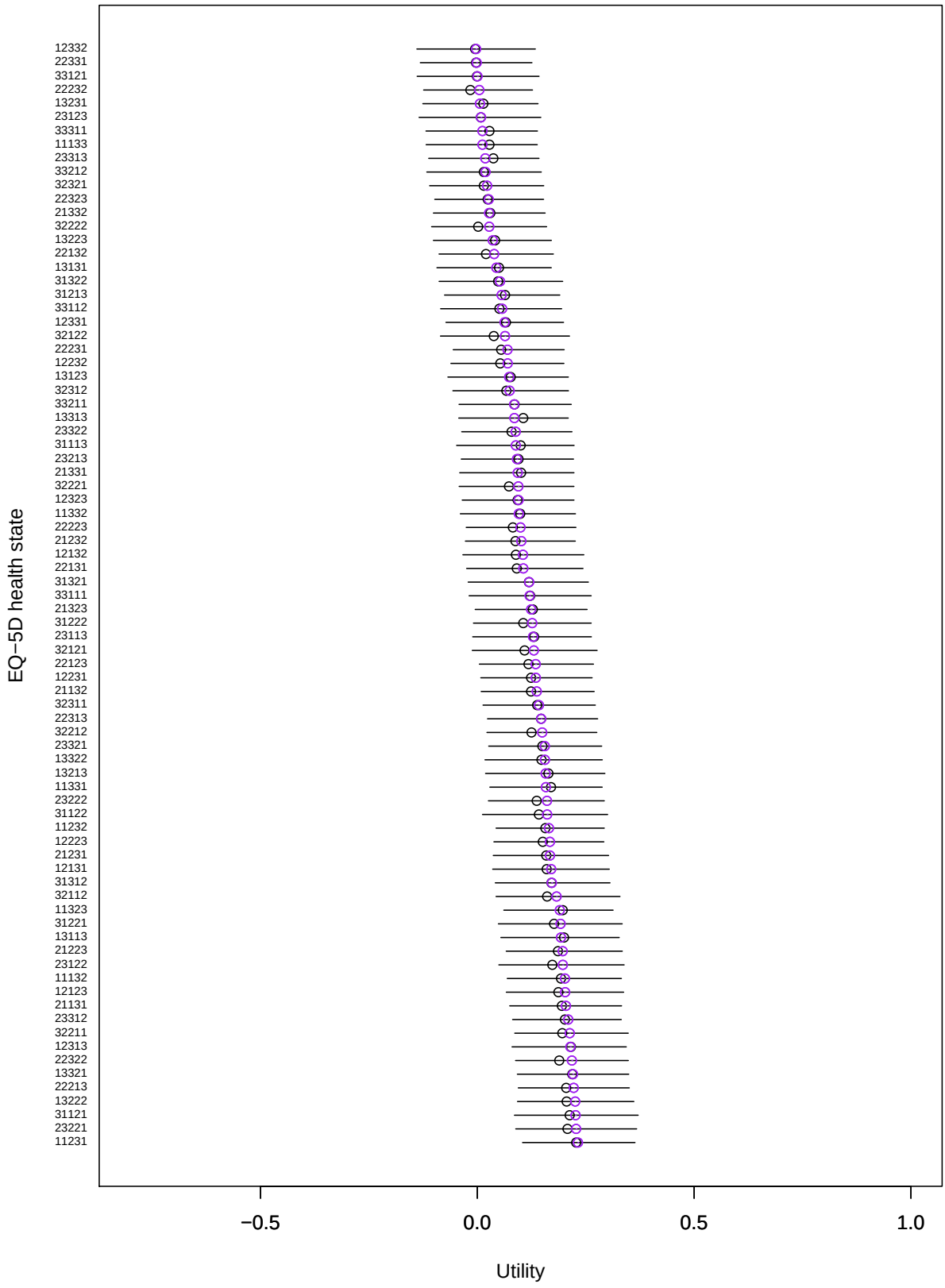


Figure 3.2: 95% credible intervals for the posterior distributions of the 81 EQ-5D-3L health states which were valued in the middle ordered with respect to the posterior medians, shown as purple circles. The MVH point estimates of the utilities of the states are shown as black circles.

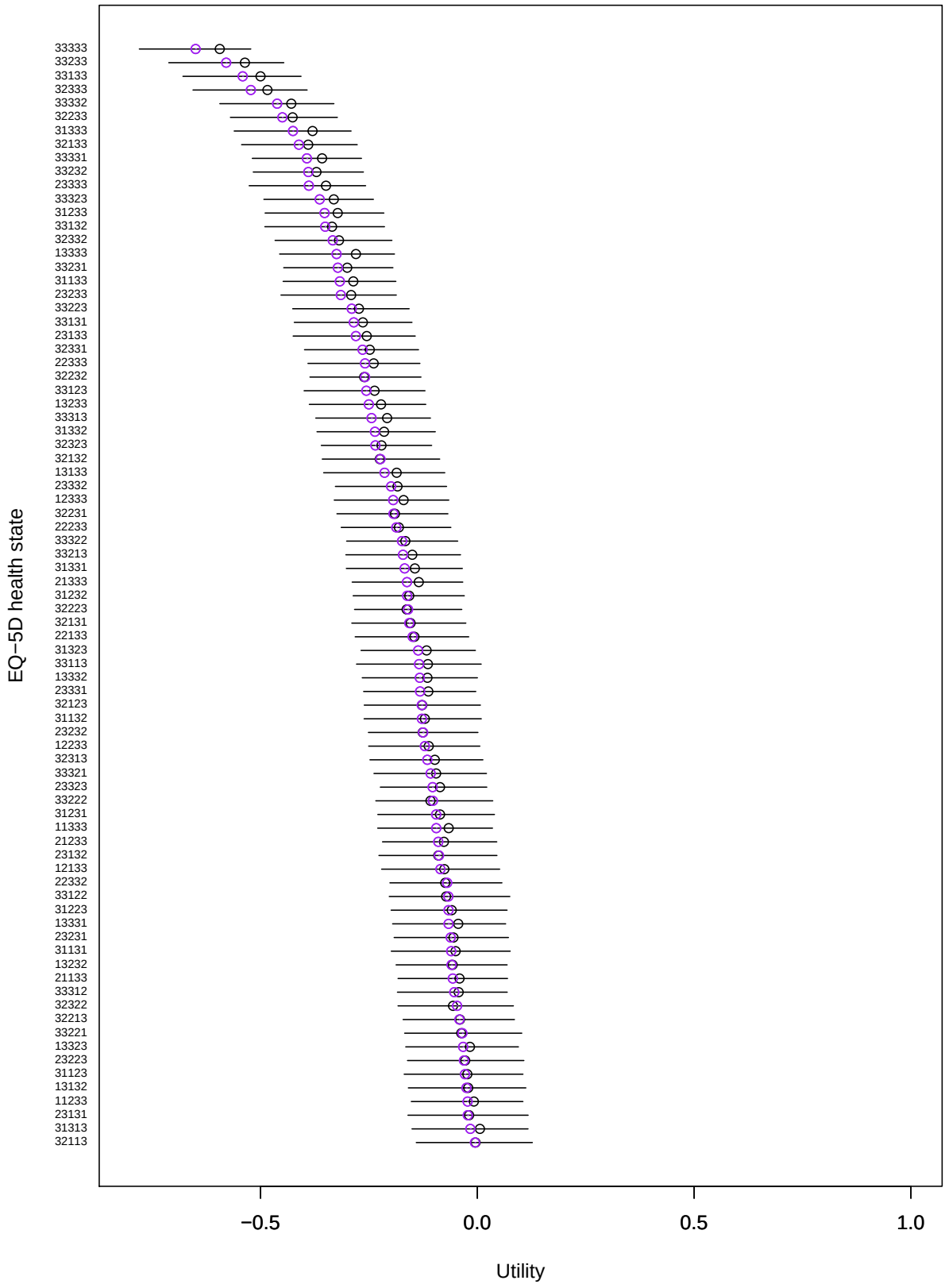


Figure 3.3: 95% credible intervals for the posterior distributions of the 81 EQ-5D-3L health states which were valued the worst ordered with respect to the posterior medians, shown as purple circles. The MVH point estimates of the utilities of the states are shown as black circles.

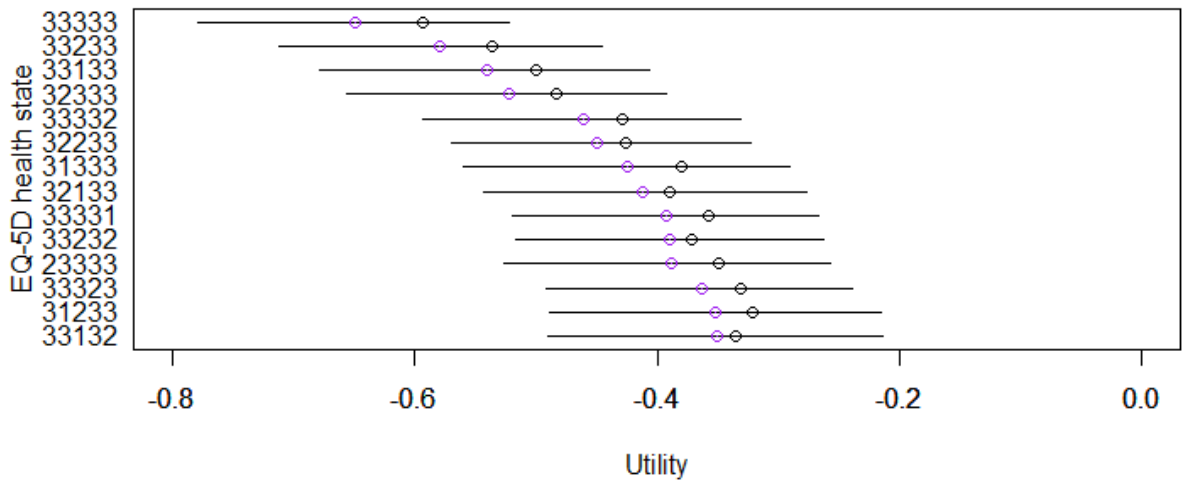


Figure 3.4: 95% credible intervals for the posterior distributions of the 14 EQ-5D-3L health states which were ranked the lowest ordered with respect to the posterior medians, shown as purple circles, when the original prior distributions are used. The MVH point estimates of the utilities of the states are shown as black circles.

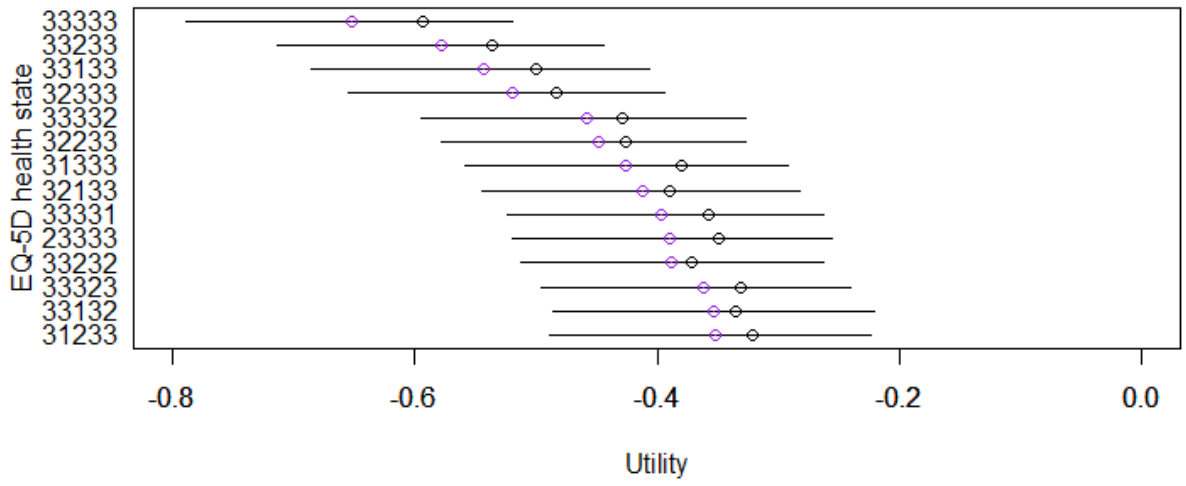


Figure 3.5: 95% credible intervals for the posterior distributions of the 14 EQ-5D-3L health states which were ranked the lowest ordered with respect to the posterior medians, shown as purple circles, when the alternative prior distributions are used. The MVH point estimates of the utilities of the states are shown as black circles.

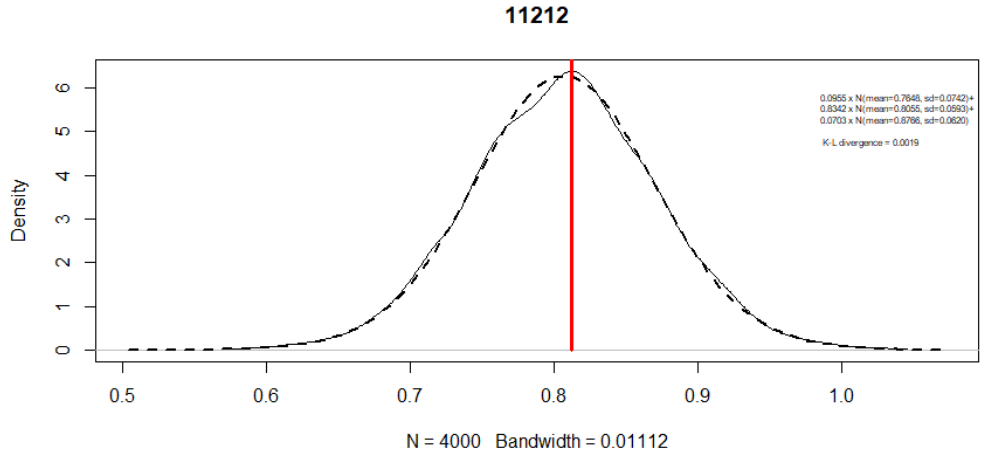


Figure 3.6: Kernel density plot for the MCMC simulations of state `11212` (solid line) and the superimposed probability density function (dotted line) of its approximation as a three-component mixture of normals. The MVH point estimate of the utility of that state is denoted as a vertical red line.

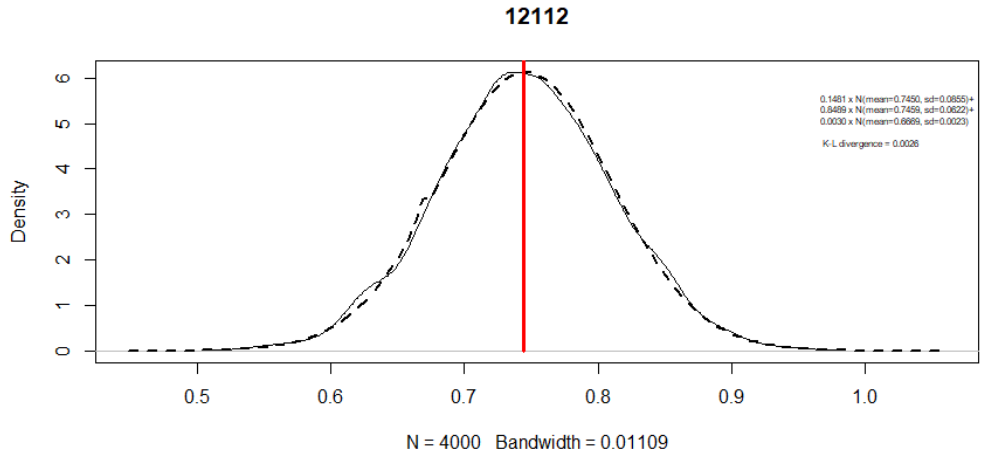


Figure 3.7: Kernel density plot for the MCMC simulations of state `12112` (solid line) and the superimposed probability density function (dotted line) of its approximation as a three-component mixture of normals. The MVH point estimate of the utility of that state is denoted as a vertical red line.

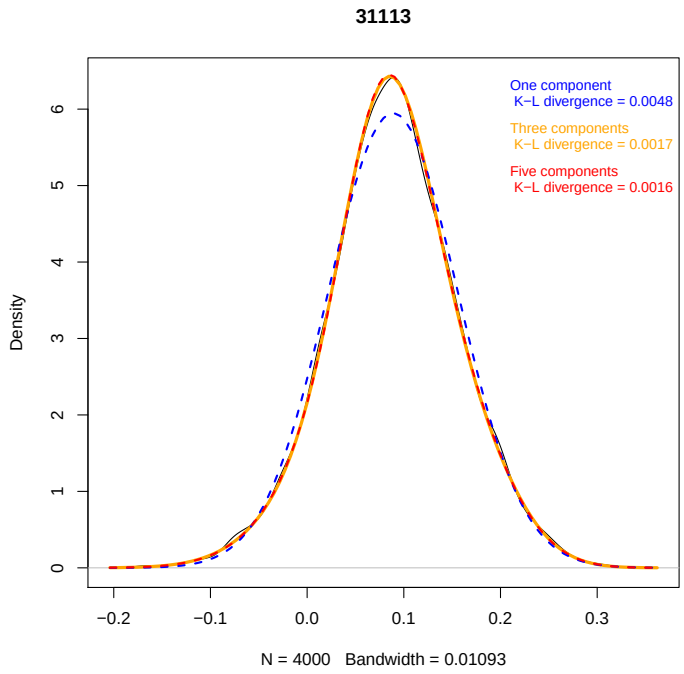


Figure 3.8: Kernel density plot for the MCMC simulations of state `31113` (solid line) and the superimposed probability distributions (dotted lines) of its approximation as one (blue), three (orange), and five (red) components mixture of normals. It can be seen that there is a substantial improvement of the approximation once we have three components, nevertheless, the improvement of the approximation by five components is not that noticeable.

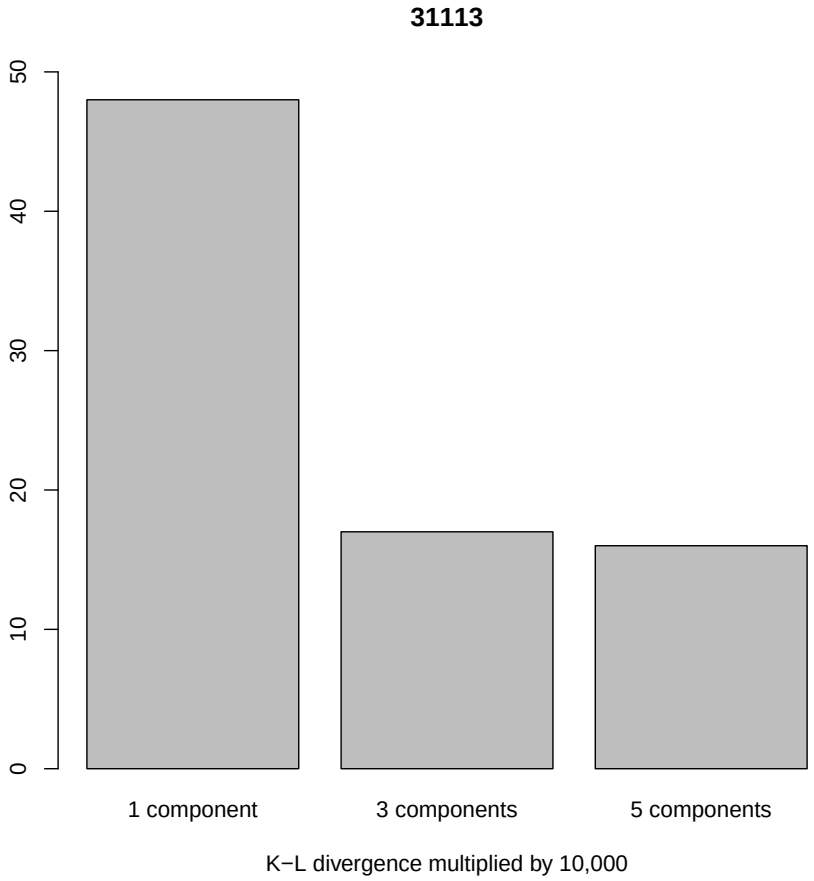


Figure 3.9: KL divergence (in 10^{-4}) for the case of two, three, and five components mixtures for state `31113`. Having three components instead of a single component means that the KL divergence is approximately 65% smaller. However, even when we add two extra components (i.e. the fourth and the fifth component) the value of the KL divergence barely changes.

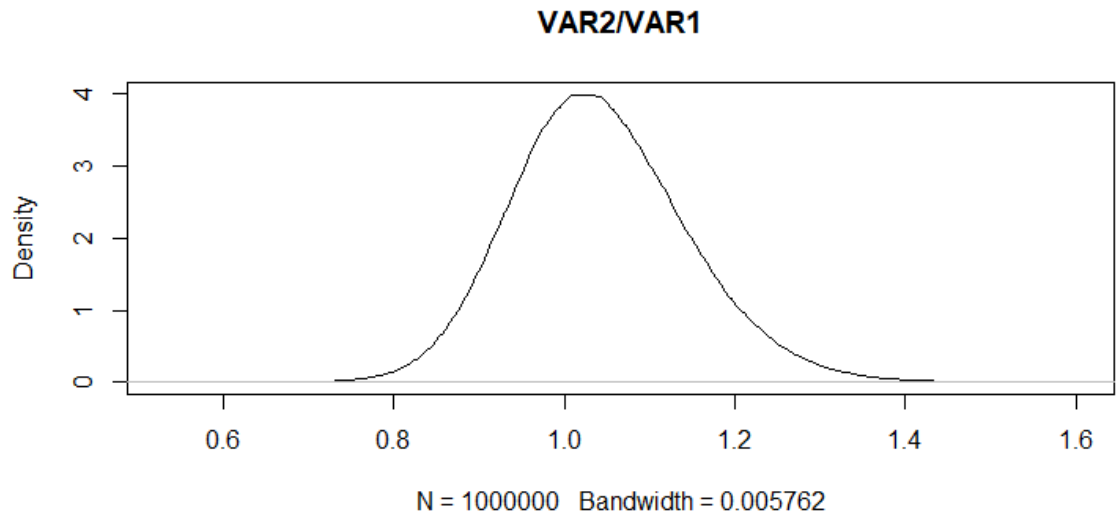


Figure 3.10: Kernel density plot for the $1,000,000$ derived values of $\mathrm{VAR}_2/\mathrm{VAR}_1$.

Chapter 4

Cost-utility analysis in practice

So far we have reviewed how to construct the presented Bayesian model to obtain the utility distributions of the EQ-5D-3L states. In this chapter, we will see how these can be used in practice when conducting cost-utility analysis. I begin by introducing the theoretical framework and how CUA is conducted in the frequentist setting; then I show how the presented techniques can be applied to do CUA. I proceed with a practical example of data from a clinical trial and then I also use simulated datasets.

4.1 CUA: Incremental cost-effectiveness ratio and useful plots

4.1.1 Incremental cost-effectiveness ratio and its use in CUA

In chapter 1 the importance of QALYs was discussed: they are the preferred tool of NICE for doing cost-utility analysis. Being a composite measure of length of time and utility means that computation of the QALYs is influenced by the use of the utility tariff which was derived by the Bayesian model. Different values of QALYs have different implications on the CUA results through the calculation of the *Incremental cost-effectiveness ratio (ICER)*. Fundamentally, the ICER is used in CUAs because of the concept that when there is consideration of investing to a new intervention, it is good to compare it with its best alternative (or the one currently in use).

Specifically, the ICER is the ratio of the difference of the total costs between the potential intervention and its best alternative, to the difference of their total benefits. More precisely:

$$\text{ICER} = \frac{C_1 - C_2}{E_1 - E_2}, \tag{4.1}$$

where C_1 and C_2 are the total costs of the potential intervention and its baseline intervention in respect, whilst E_1 and E_2 are the interventions' respective total benefits. The numerator of equation (4.1) is also called *incremental cost* (or *differential cost* Δ_c), while the denominator is also known as *incremental benefit*. In CUAs the effects can be measured in QALYs and thus the incremental benefit of an intervention is expressed in terms of the additional QALYs gained due to that intervention; hence, the denominator of the ICER is equal to the *differential QALYs* Δ_e.

There is a theory of decision rules (Morris, Devlin and Parkin, 2007) on how to decide between different interventions using ICER for CUA (where the benefits are expressed as QALYs). Briggs, Claxton and Sculpher (2011) and Baio (2013) describe the statistical principles of decision making (under uncertainty) in healthcare by considering the expected utility values of the available interventions. Remarkably, NICE sets an *ICER threshold* in terms of cost per QALY (Appleby, Devlin and Parkin, 2007), so the threshold represents the *willingness-to-pay*. If an intervention has an ICER less than this threshold, then NICE accepts it as cost-effective (it considers that the additional costs are justified by the additional benefits achieved). NICE sets thresholds between the range of £20,000 to £30,000 (McCabe, Claxton and Culyer, 2008). Claxton et al. (2013) believe that the acceptance threshold should be several thousands of pounds lower, namely £12,936. In contrast, these findings are rejected by Barnsley et al (2013) and Raftery (2014) as they think that the approval of the £12,936 threshold needs the acceptance of too many suppositions, while it is sensitive to alternative justifiable assumptions.

4.1.2 Useful plots in CUA

To help disentangle the difficulties in the interpretation of the CUA results for the comparison between two interventions, it is helpful to consider the graphical representation provided by some useful plots. Namely, these are the plots of the: cost-effectiveness plane, expected incremental benefit, cost-effectiveness acceptability curve and expected value of information. These plots are introduced here, as they are used later in this chapter (in section 4.3) for the presentation of the CUA results.

The cost-effectiveness plane consists of a four-quadrant diagram: the horizontal axis represents the incremental level of effectiveness of an outcome (e.g. in terms of utility) and the vertical axis represents the additional total cost of implementing this outcome. The further right we move on the horizontal axis, the more effective the outcome. Similarly, the further up we move on the vertical axis, the more costly the outcome. The ICER for a comparison of two interven-

tions can be represented as a point, the coordinates of which are equal to the corresponding differential QYALYs and differential cost. If the ICER is in the southeastern quadrant, (negative cost differential but positive differential QALYs) then it means that the new intervention is considered better than the one currently in used. If the ICER is located in the northwestern quadrant (positive cost differential but negative differential QALYs) then the new intervention is not preferred as it costs more but results in fewer QALYs than the one currently in use. However, the interpretation of the ICERs located in the rest of the quadrants is less clear. If the ICER is in the southwestern quadrant, it means that new intervention costs less but also results in fewer QALYs so it is not immediately clear whether this intervention should be adopted. Similarly, if the ICER is in the northeastern quadrant (which is actually the most common case of a CUA), then it is not immediately clear whether the new intervention (which costs more but results in more QALYs) should be adopted. A diagonal line can represent the willingness-to-pay threshold, so that if the ICER falls below that line then the new intervention is preferred compared to the currently used intervention, whereas if the ICER falls above that line then the new intervention is not preferred. Being below the diagonal line means that the cost for each additional QALY is less than the value which the decision-maker is willing to pay to acquire an additional QALY. If we want to do sensitivity analysis try to quantify the uncertainty of the ICER, then multiple pair-values of differential cost and differential QALYs can be derived (e.g. as described later in this chapter). Then, by representing the ratios of these pairs of values as points in the cost-effectiveness plane we can see how they are distributed and also check how many of these points would fall below the willingness-to-pay threshold. We calculate ICER $= \frac{E(\Delta_c)}{E(\Delta_e)}$, and we can quantify the uncertainty of the ICER by examining how the aforementioned points are distributed in the cost-effectiveness plane.

Assuming that there are multiple derived values of differential costs and differential QALYs, the expected incremental benefit is defined as $EIB = E[k\Delta_e - \Delta_c] = kE[\Delta_e] - E[\Delta_c]$, where k is the willingness-to-pay parameter. We can easily see that if $EIB > 0$ then $k > \frac{E(\Delta_c)}{E(\Delta_e)} = ICER$, and therefore interventions for which the ICER is less than the willingness-to-pay threshold are considered cost-effective. The expected incremental benefit plot depicts EIB (vertical axis) upon varying the willingness-to-pay parameter k (horizontal axis). From the graph it is possible to identify the break-even point, which is the value of k^* for which the optimal decision changes (i.e. for threshold values less than k^* the EIB is negative and thus the new intervention is not preferable, whereas for threshold values greater than k^* the EIB is positive and so the new intervention is cost-effective.

The cost-effectiveness acceptability curve is defined as $CEAC = P(k\Delta_e - \Delta_c) > 0$, which de-

pends on the willingness-to-pay parameter k. In other words the cost-effectiveness acceptability curve is computed as the proportion of the (Δ_c, Δ_e) pairs for which $k\Delta_e - \Delta_c > 0$. In accordance with what discussed so far, when $k\Delta_e - \Delta_c$ is positive, the new intervention is considered cost-effective and so higher values of the cost-effectiveness acceptability curve indicate that, for a given budget that the decision-maker is willing to invest, the probability that the intervention is more preferable than the currently used intervention is large. The main advantage of the cost-effectiveness acceptability curve plot is that it summarises the aforementioned probability of cost-effectiveness (vertical axis) upon varying the willingness-to-pay parameter (horizontal axis).

The expected value of (perfect) information (EVPI) is the upper limit of the price that a decision-maker would be willing to pay in order to gain access to supplemental information that is certain (perfect), regarding all factors that influence which intervention is preferred. In other words, it is the value (in money terms) of omitting all uncertainty from the analysis, since perfect information can theoretically eliminate the possibility of making the wrong decision. It can be calculated as $E[k\Delta_e - \Delta_c | k\Delta_e - \Delta_c > 0] \times CEAC - max(EIB, 0)$. The expected value of information plot shows the expected value of information (vertical axis) in monetary terms upon varying the willingness-to-pay parameter (horizontal axis). It can represent the average opportunity loss deriving by using the current most cost-effective intervention, instead of further investigating to reduce the uncertainty in the parameters, for each value of the willingness-to-pay parameters k. Higher values for the expected value of information mean that, for a given budget that the decision-maker is willing to invest, the value of the additional research is large. EVPI changes its shape around the break-even point k^*, because the optimal decision is reversed beyond that threshold.

4.2 CUA modelling

4.2.1 The standard procedure to conduct CUA

A "standard" procedure which can be followed to conduct CUA in a frequentist context is described. For the case of dealing with clinical trial data when conducting CUA using the UK EQ-5D-3L as an instrument, a utility value is assigned to each individual in the trial at each of the pre-determined time-points, at which they recorded their EQ-5D-3L states. The standard procedure is to assign these values using equation (2.2), where the TTO results of the MVH study (Szende, Oppe and Devlin, 2007) are used for the estimates of the β coefficients. Next,

u_{ijt} (utility of the i-th subject of the j-th group of the clinical trial at time point t) are then used to calculate the QALYs for the clinical trial participants.

An extension of equation (1.1) (and the concepts which were reviewed in section 1.3.2) using the appropriate notation for clinical trial data means that QALYs of the ij individual can be calculated as the AUC as follows:

$$AUC_{ij} = \sum_{t>1} \frac{\left(u_{ijt} + u_{ij(t-1)}\right)}{2} (l_t - l_{t-1}).$$

However, in order to calculate the differential QALYs (the difference between the mean QALYs per group) Δ_e, a regression-based adjustment (Manca, Hawkins and Sculpher, 2005) can be used to control for baseline utility differences between the trial arms. This is formalised as:

$$AUC_{ij} = \gamma_1 + \gamma_2 u_{ij0} + \gamma_3 \delta_i + \eta_i, \tag{4.2}$$

$$\eta_i \sim \mathrm{N}\left(0, \sigma_\eta^2\right),$$

where δ_j is an indicator variable taking the value of 0 if the individual belongs to the control group, and the value of 1 if the individual belongs to the intervention group; u_{ij0} is the utility of the i-th patient of the j-th group at baseline. An estimator of Δ_e is $\hat{\gamma}_3$. Alternatively, an estimator of the mean QALYs of the control group is $\hat{\gamma}_1 + \hat{\gamma}_2 \overline{u_{ij0}}$, where $\overline{u_{ij0}}$ is the mean utility (of all individuals) at baseline. Similarly, an estimator of the mean QALYs of the intervention group is $\hat{\gamma}_1 + \hat{\gamma}_2 \overline{u_{ij0}} + \hat{\gamma}_3$. Thus we could either get the estimates of the mean QALYs of the individual trial groups and subtract them to find the estimate of the differential QALYs or we could use $\hat{\gamma}_3$ directly.

The estimate of the differential QALYs can be combined with the estimate of the differential cost to calculate the ICER, which can be compared with the cost-effectiveness threshold set by NICE to determine the cost-effectiveness of the intervention. A cost-effectiveness plane (Morris, Devlin and Parkin, 2007; Cohen and Reynolds, 2008) can be used to get a visual illustration of the distribution of the ICER.

An approach to quantify uncertainty about the ICER in a frequentist setting is to use bootstrapping where the distribution of the ICER is generated by repeatedly sampling from the data (Khan, 2016). In particular:

- First we generate F bootstrap samples each one of size L, where L is the total number of trial participants. A bootstrap sample randomly selects each individual's set of data (costs, effects, trial group membership) and replaces the individual before selecting another individual's data. It is possible that the same individual may be (randomly) selected again.

- Once we have sampled L individuals with their entire data, we run the regression of equation (4.2) to estimate the differential QALYs. Since we have F bootstrap samples, we will have F differential QALYs. Similarly, we can derive F differential costs.
- For each sample we can compute the ICER using equation (4.1). Thus, we will have F ICERs.
- Now, from the F ICERs we can generate the required bootstrap statistics: such information can be illustrated by using figures such as the cost-effectiveness plane.

Another approach to quantify uncertainty about the ICER would be to make individual assumptions about each of the model parameters and then take samples from those distributions from which the ICER is computed. As a consequence, the derived ICER distribution gives us information about the underlying uncertainty but it is based on assumptions and it is not the "true" ICER distribution.

4.2.2 CUA adaptation of the Bayesian model

The structure of the model described in section 3.2.1 can be extended in a Bayesian context to find the ICER when we have clinical trial data. Using MCMC techniques we can obtain MCMC samples from the posterior distributions of the β coefficients and $\xi^{(new)}$. Next, because they are a deterministic function of utility score and time, QALYs can easily be calculated once we have the MCMC sample of the utility scores by considering the Area Under the Curve.
The utility value $u_{ijt}^{(c)}$ of the state experienced by a specific individual at a given point for the c-th ($c = 1, ..., C$) MCMC iteration is computed as

$$u_{ijt}^{(c)} = 1 - \boldsymbol{X}_{ijt}^{*T}\boldsymbol{\beta}^{(c)} - \xi_{ijt}^{new(c)}, \tag{4.3}$$

where $\boldsymbol{\beta}^{(c)}$ is the vector of the β coefficients with the values which were computed under the c-th MCMC iteration and $\xi_{ijt}^{new(c)}$ is the misspecification term (which refers to the state of the i-th subject of the j-th group of the clinical trial at time point t) the value of which was computed under the c-th MCMC iteration. Furthermore, the AUC of the subject at iteration c can be deterministically computed as:

$$AUC_{ij}^{(c)} = \sum_{t>1} \frac{\left(u_{ijt}^{(c)} + u_{ij(t-1)}^{(c)}\right)}{2} (l_t - l_{t-1}). \tag{4.4}$$

In order to account for utility differences at baseline, there is a second part of the model, which

allows us to quantify a further layer of uncertainty for Δ_e. Let ${\Delta_e}^{(c)}$ be the differential QALYs derived by using the AUCs calculated under the c-th MCMC iteration of the first part of the model. For each MCMC iteration c, we use the values of the AUCs computed at iteration c and construct a regression model from which we can obtain an MCMC sample of size C^* from the posterior distribution of the differential QALYs ${\Delta_e}^{(c)}$. We have:

$$AUC_{ij}^{(c)} \sim N\left(\nu_{ij}^{(c)}, \sigma_\eta^2\right), \text{ for } c = 1, ..., C,$$

$$\nu_{ij}^{(c)} = \gamma_1^{(c)} + \gamma_2^{(c)} u_{ij0}^{(c)} + \gamma_3^{(c)} \delta_i, \text{ for } c = 1, ..., C,$$

where $u_{ij0}^{(c)}$ is the utility of the i-th patient of the j-th group at baseline, computed under the c-th simulated values. Then, ${\Delta_e}^{(c)}$ is equal to ${\gamma_3}^{(c)}$. In particular, each time that we run such a regression (for a fixed value of c, i.e. at iteration c of the first part of the model which was presented in chapter 3), we derive c^* MCMC values of ${\gamma_3}^{(c)}$. Since the first part of the model has C iterations in total, the obtained overall MCMC sample of the differential QALYs will be a matrix of size $C \times C^*$ (C rows and C^* columns). Moreover, we could estimate the mean QALYs per group, instead. Similarly, the obtained MCMC sample of the differential costs will be of size $C \times C^*$ (more on how the costs are modelled in section 4.3.1).

The priors of the γ regression coefficients are Normal distributions with mean 0 and standard deviation 1, whereas $\sigma_\eta \sim U(0, 1)$. Alternative priors similar to those presented in section 3.2.1 (e.g. $N(0, 10^6)$ for the γ parameters and $Gamma(0.001, 0.001)$ for $1/\sigma_\eta$) can also be used for the conduction of sensitivity analysis.

Due to the additional complexity of the model (in order to account for baseline utility differences) the $C \times C^*$ MCMC sample of differential QALYs Δ_e (or differential costs) which we obtain can be of large size. As a result, further analysis of these values (which could be of larger size than conventionally needed) could be storage-demanding and computationally intensive. Nevertheless, for each one of these parameters of interest (e.g. differential QALYs and differential costs) we could try to obtain a shorter vector of size G which summarises the information described in the big matrix of size $C \times C^*$ and can be assumed to consist of values coming from the actual distribution of interest. This can be done by randomly sampling G values without replacement from the big matrix (i.e. we choose G matrix entries randomly) and thus we obtain a vector which has a summarised version of the information of interest without altering its nature. In other words, the proportion of the $C \times C^*$ values falling within a particular interval is expected to be similar to the proportion of the G values falling within that particular interval. Thus, the information that we possess remains the same.

I choose to obtain G = 10,000 values from each $C \times C^*$ matrix. Specifically, each time that

matrix entry (c, c^*) (where $c = 1, ..., C$ and $c^* = 1, ..., C^*$) is chosen for a matrix we wish to summarise, same-indices entries (c, c^*) are also used for the rest of the matrices we desire to summarise as well. For example, let's assume that the first element of the vector containing the summarised information of the differential QALYs happens to be the value of the differential QALYs which was derived under the 5th MCMC iteration of the first part of the model and the 10th MCMC iteration of the second part of the model (i.e. entry $(5, 10)$ of the $C \times C^*$ matrix of differential QALYs). Then, in that case the first element of the vector containing the summarised information of the differential costs is the value of the differential costs which was also derived under the 5th MCMC iteration of the first part of the model and the 10th MCMC iteration of the second part of the model (i.e. entry $(5, 10)$ of the $C \times C^*$ matrix of differential costs). The connection is the same for the second, third,..., G-th elements of the aforementioned vectors. Hence, in the end we obtain vectors of size G which contain the information needed (e.g. differential QALYs and differential costs) in order to compute the ICER and summarise the results of the CUA.

4.3 CUA applications

4.3.1 The CoBalT trial application

The objective is to apply the Bayesian model to the CoBalT trial and conduct cost-effectiveness analysis. Then, we can compare the results of the Bayesian model with those produced under the frequentist approach. The techniques described in this section can also be used for any simulated randomised control trial data.

The CoBalT trial is a randomised control trial about people with treatment-resistant depression (Hollinghurst et al., 2013). They recruited patients aged 18–75 years who had adhered to antidepressants for 6 weeks or more and had substantial depressive symptoms (Beck Depression Inventory (Beck, Steer and Brown, 1996) [BDI-II] score $\geq$14 and met the ICD-10 depression criteria (Wiles et al., 2013)). The control group consists of 235 patients who take unrestricted usual care from their general practitioners (GPs), including anti-depressant treatment and referrals. The intervention group consists of 234 patients, who receive the usual care as well as Cognitive–behavioural therapy (CBT): they were offered between 12 and 18 one-hour sessions. Measurements of costs and EQ-5D-3L states have been recorded at baseline, 6 and 12 months after baseline.

The next step is to define the final dataset which will be analysed and set the model outcomes.

After excluding any observations with missing values (i.e. missing cost or EQ-5D-3L score) we end up with 182 observations in the control group and 186 in the intervention one. For the calculation of the outcomes, the QALYs are calculated over the one-year period of the trial. In both the frequentist and the Bayesian models we consider that the utility of the EQ-5D-3L state of perfect health is always equal to 1. For the calculation of the costs in the frequentist model we use the face values of the costs computed under the National Health Service (NHS) and Personal Social Services (PSS) perspective without any discounting as the analysis is restricted to one-year period (discounting is the mathematical method by which economists devalue future costs and benefits to reflect the time preferences of individuals; see Drummond et al. (2005) for information on the concept of discounting and its use in economic evaluations). In the Bayesian models, the cost c_{ij} of the i-th patient of the j-th arm is modelled as:

$$c_{ij} \sim N\left(m_{cj}, \sigma_{cj}^2\right),$$

where the prior distributions are set as: $m_{cj} \sim U\left(1, 3000\right)$ and $\sigma_{cj} \sim U\left(1, 3000\right)$.

The statistical packages `R` and `JAGS` are used to do the statistical computations. In particular, for the first part of the main Bayesian model (which was described in section 3.2.1) MCMC is run with two chains; each chain is of size 2,000 and they were preceded by a burn-in sample of 4,000 iterations in total. Then the second part of the model is run (which was described in section 4.2.2) for $C = 2 \times 2,000 = 4,000$ times and each of these times we obtain an MCMC sample of size $C^* = 4,000$ for the differential QALYs. Hence, in the end we have a total MCMC sample of $C \times C^* = 4,000 \times 4,000 = 16,000,000$ simulations of the differential QALYs.

Moreover, we would also like to check further the robustness of the approximate posterior utility distributions (mixtures) which were derived in section 3.3.4). Hence, we could use such models and compare their results with the results of the main model. In particular, in these cases we will not have to run the first-part of the model (to derive MCMC samples for the posterior utility distributions) since we have already assigned approximate utility distributions (mixtures) to each state. Instead, we will sample some values from these mixtures and then use these values for the second part of the model.

For the case of the Bayesian models based on the approximate mixture distributions, we consider one, three and five components mixtures of Normals. For each of these models we also sample $C = 4,000$ values from each mixture distribution which approximates a utility distribution of an EQ-5D-3L state of interest. In the second part of the model (which accounts for baseline utility differences) C^* is also set equal to 4,000.

Finally, we also use bootstrap. The bootstrap results will be compared with those obtained

under each Bayesian model. In the analysis done under the standard frequentist approach, we estimated the differential QALYs as 0.055. The mean cost (in £) of the control group is 799.449 and 1,809.977 for the intervention group, implying an incremental cost of 1,010.529. Therefore, the incremental cost-effectiveness ratio (ICER) is £18,361.022 per QALY gained. This is substantially lower than the threshold of £30,000 per QALY.

We are interested in the results of the cost-effectiveness analyses of the CoBalT trial based on bootstrap and the various Bayesian models. For each Bayesian model we have obtained a very large MCMC sample of differential QALYs of size 16,000,000; we use the method described in section 4.2.2 to obtain a summary vector of differential QALYs which has a size of $G = 10,000$ and use it for the `BCEA` package (Baio, Berardi and Heath, 2019) to report the results of the cost-effectiveness analysis. Table 4.1 provides a summary of statistics of the cost-effectiveness analyses of the various methods.

method	mean Δ_e	standard deviation of Δ_e	ICER
Point-estimates	0.0550	NA	18,361
Bootstrap	0.0554	0.0180	18,218
Main Bayesian	0.0542	0.0194	18,662
One-component approximations	0.0541	0.0198	18,671
Three-component approximations	0.0542	0.0194	18,673
Five-component approximations	0.0543	0.0195	18,586

Table 4.1: Statistics of the cost-effectiveness analyses of the different models.

The results of the main Bayesian model and the three models based on the approximate distributions are similar. Although the MCMC results of these four models were derived under different seeds, the fact that they are similar enough is desirable. It means that good approximations have been achieved in section 3.3.

Plots such as the cost-effectiveness plane, expected incremental benefit, cost-effectiveness acceptability curve, and expected value of information are useful tools for reporting the results of the cost-effectiveness analysis. These plots are constructed using `R` and the `BCEA` package (see Baio (2013) for a discussion on these kind of plots and how to construct them using `BCEA`). Figure 4.1 shows the cost-effectiveness plane, the expected incremental benefit, cost-effectiveness acceptability curve, and expected value of information respectively, for the main Bayesian model (in black) compared to bootstrap (in red).

We notice that the ICER of the main Bayesian model and the ICER of the bootstrap are not

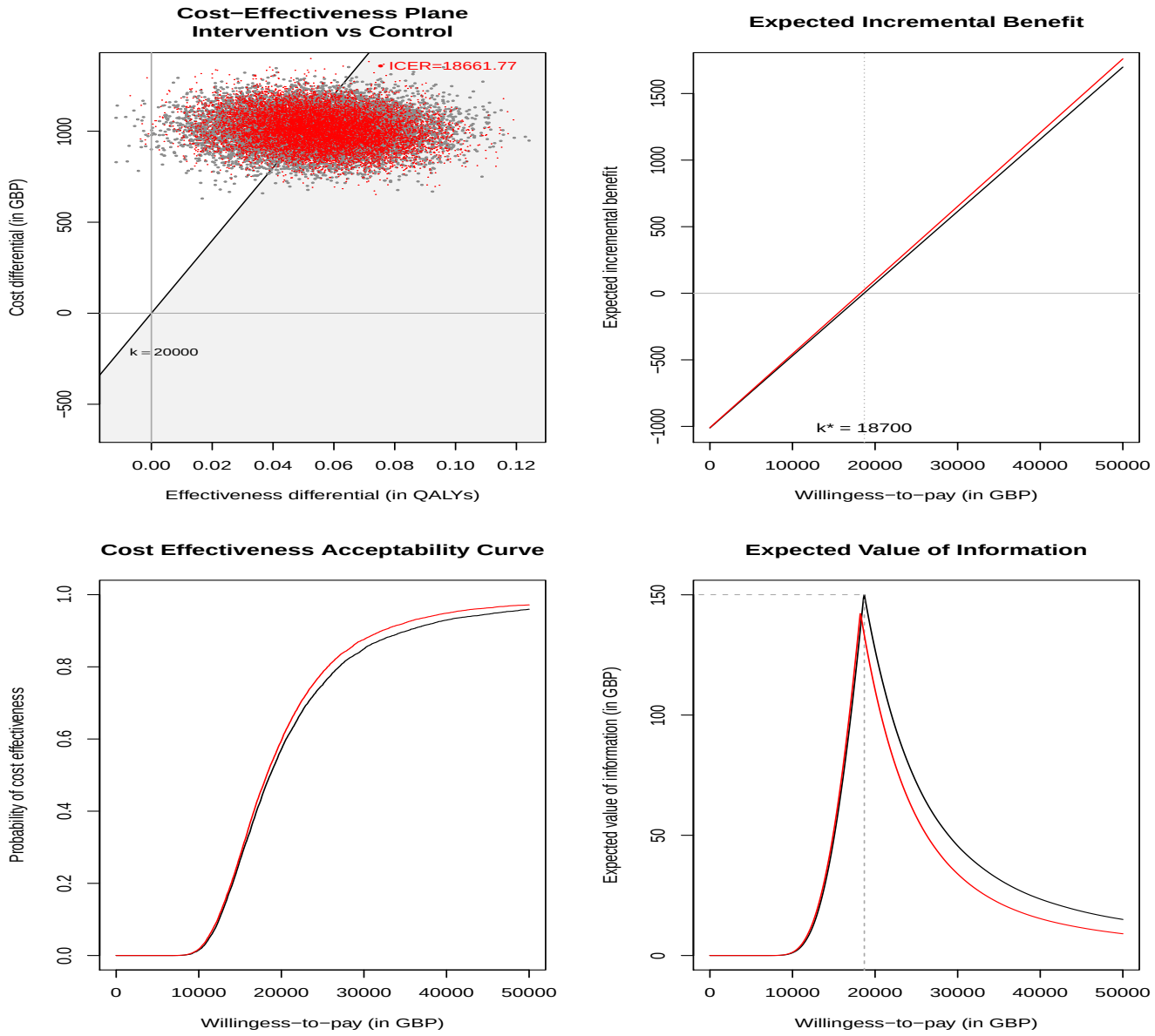


Figure 4.1: Cost-effectiveness analysis of the CoBalT data. From top-left to bottom-right we have the plots of the following: cost-effectiveness plane, expected incremental benefit, cost-effectiveness acceptability curve and expected value of information. The results of the main Bayesian model are in black and the results of the bootstrap are in red. The costs are measured in British pounds (GBP), k is the willingness-to-pay parameter, whereas k^* is the break-even point.

very different. This means that at at a given ICER threshold (which will probably be higher than £20,000, whilst the derived ICERs are less than that) , the intervention will be considered cost-effective under both the Bayesian and the bootstrap approach. Moreover, the plot of the expected incremental benefit shows that the two lines are not identical; there are some visual differences yet they are relatively small.

However, there appear to be some further differences between the results produced under my approach and those produced from bootstrapping. The plot of the cost-effectiveness plane shows that the black points are slightly more spread out compared to the red-points, implying that there is wider uncertainty associated with the results which were produced when my approach was used. Furthermore, this increased uncertainty is also reflected in the rest of the plots such as the cost-effectiveness acceptability curve where the dissimilarities of the curves show that, for a fixed willingness-to-pay value, it's less likely for the intervention to be cost-effective when my approach is used compared to when bootstrapping is used. Moreover, the plot of the expected value of information shows that for a given budget which is similar to that of the bootstrap approach, when my approach is used the expected value of information is actually larger, meaning that the value of the additional research is larger.

In conclusion, when we analyse the results of the CoBalT trial and compare the two methods, we notice that there are some differences. In particular, the results produced under my approach show that there is an increased level of uncertainty which is accounted for compared to the results produced under the bootstrap approach. This is important information to be known by the decision maker, because in addition to making their decision in favour of one intervention or another, they also desire to know the uncertainty associated with each decision. Nevertheless, it appears that for the case of the CoBalT trial the differences between the results of the two approaches are relatively small.

4.3.2 Simulated datasets

We saw that for the CoBalT trial it appears that there are not very big differences between the results derived under the bootstrap approach and those of my model. However, we could simulate some randomised control datasets to observe if some the times the results of the two methods are substantially more different.

For the construction of these simulated datasets multiple factors are taken into account, such as the number of different states experienced by the trial subjects, the extend to which the health states of trial participants generally improve (or not) substantially between the beginning and the end of the trial and the nature of populations (e.g having populations with utilities close to perfect health or the opposite). The simulated datasets have EQ-5D-3L indeces at three different time-points: baseline, six months after baseline, and one year after baseline. Hence, we assume that there is no need to consider any discounting. For the derivation of realistic simulated datasets which have similar properties to those of real-life datasets, their character-

istics are based on descriptions of clinical trial data as discussed by Hamashima (2002), Lee et al. (2003), Norum (1996), Schneider et al. (2000), Stafford et al. (2012) and van Roijen et al. (1997).

Upon examining the results of the CUA for the datasets which were simulated, I notice that sometimes there appears to be a small difference between my approach and the frequentist method. However, I deduced that for some of them the results of the two methods were substantially more different than the previous cases: in particular there was more variability in the values of the differential effects produced under the Bayesian approach.

The main Bayesian and bootstrap results of the cost-effectiveness analysis for such a selected dataset are illustrated in Figure 4.2. The costs and the EQ-5D-3L states experienced by the individuals of the simulation at each of the three aforementioned time-points can be found in the Appendix. Under the bootstrap approach the ICER is £23,913, whereas under the Bayesian approach it is £25,055 per QALY gained. The standard deviation of the differential QALYs for the bootstrap and main Bayesian model is 0.01003221 and 0.01363324 respectively.

This time, in Figure 4.2 we can see that the visual differences between the results produced under the two methods are even more apparent in relation to the differences we observed when we used the CoBalT dataset. In the cost-effectiveness plot, the black dots (which represent the results under the Bayesian approach) are more spread out than the red dots (which represent the results under the bootstrap approach). The standard deviation of the differential QALYs derived under the Bayesian approach is approximately a third greater than the standard deviation of the differential QALYs derived under the bootstrap approach. The plot of the expected incremental benefit shows noticeable differences between the two lines and we can clearly see that there are different thresholds at which the new intervention would be considered to be cost-effective under each of the approaches. Similarly, the expected value of information plot illustrates the different willingness-to-pay values at which the expected value of information is increased under each of the approaches, specifying when the value of additional research is large. Important differences can be spot by examining the cost-effectiveness acceptability curve plot: up to a certain willingness-to-pay value it's more likely for the intervention to be cost-effective when my approach is used compared to when bootstrapping is used, and then after that value, for a fixed willingness-to-pay value, it's substantially less likely for the intervention to be cost-effective when my approach is used compared to when bootstrapping is used.

Moreover, Table 4.2 provides a summary of statistics of the cost-effectiveness analysis of the main Bayesian model as well as those produced under the bootstrap approach, the point estimates, and the mixture distributions. Specifically, the results of the main Bayesian model and

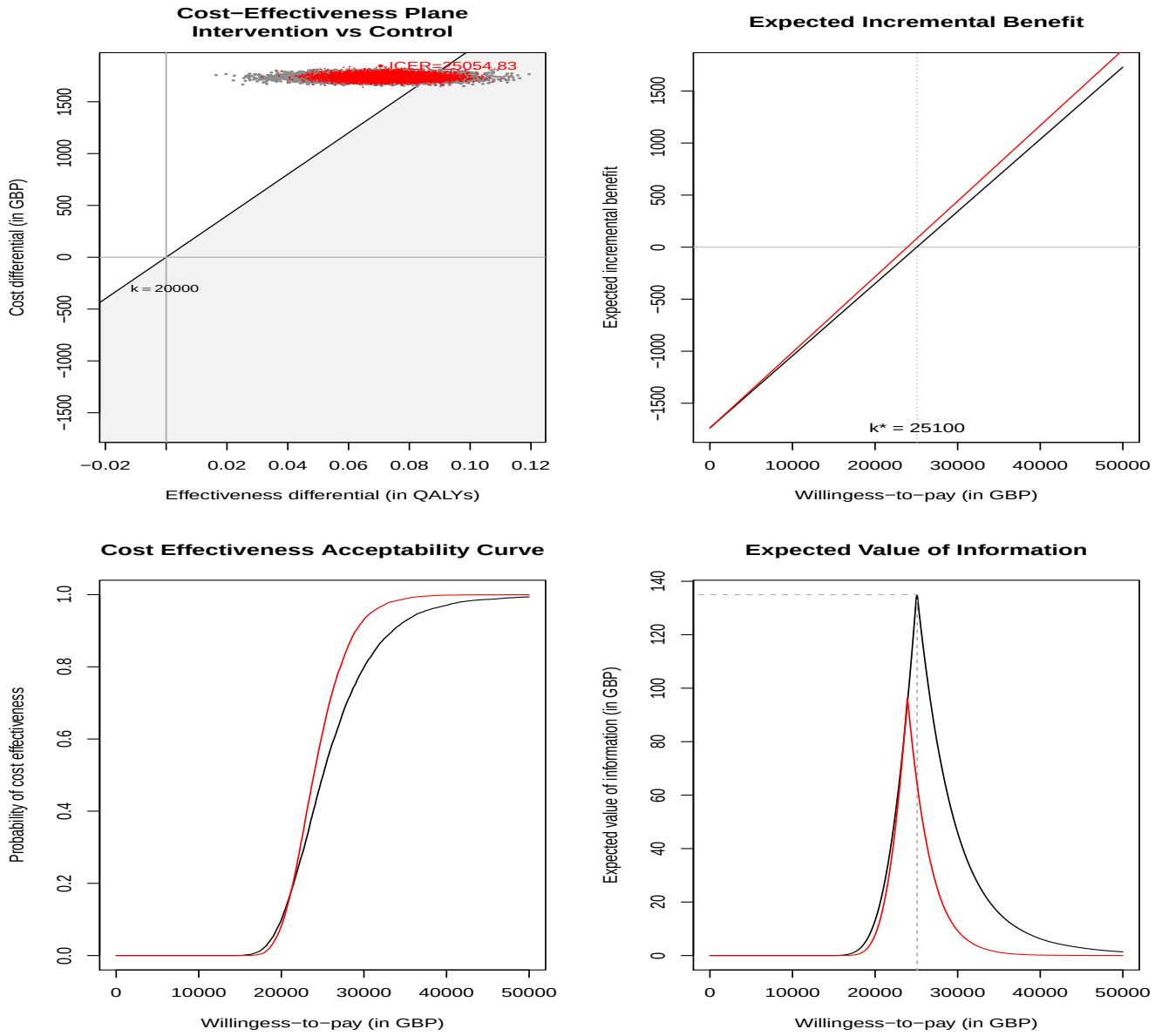


Figure 4.2: Cost-effectiveness analysis of the simulated dataset. From top-left to bottom-right we have the plots of the following: cost-effectiveness plane, expected incremental benefit, cost-effectiveness acceptability curve and expected value of information. The results of the main Bayesian model are in black and the results of the bootstrap are in red. The costs are measured in British pounds (GBP), k is the willingness-to-pay parameter, whereas k^* is the break-even point.

the three models based on the approximate distributions are similar. Although the MCMC results of these four models were derived under different seeds, the fact that they are similar enough is desirable. It means that good approximations have been achieved in section 3.3. Therefore, due to the different mathematical properties of the two methods (i.e. the main Bayesian approach and the bootstrap approach), sometimes there are datasets for which there

method	mean Δ_e	standard deviation of Δ_e	ICER
Point-estimates	0.0727	NA	23,907
Bootstrap	0.0727	0.0100	23,913
Main Bayesian	0.0693	0.0136	25,055
One-component approximations	0.0691	0.0138	25,138
Three-component approximations	0.0694	0.0137	25,027
Five-component approximations	0.0694	0.0137	25,040

Table 4.2: Statistics of the cost-effectiveness analyses of the different models.

appear to be substantial differences between the results of the two methods. The results of my approach show that sometimes there is a change (for a fixed willingness-to-pay threshold) in whether the intervention is considered to be cost-effective or not. Furthermore, there is a change in how much uncertainty is associated with each decision the decision-maker has to make. The Bayesian model of my approach accounts for uncertainty which wouldn't be captured if only the standard approach to conduct CUA had been used instead.

4.4 Summary

In this chapter we saw what a practical way is to use my model to conduct CUA and how this is compared to the standard way in which uncertainty is reported by using bootstrap. In order to properly account for baseline utility differences, we have to use a second stage for our model in which we consider a regression which further complicates the model as we have to account for a further level of uncertainty. For the case of the application of my methods on the CoBalT trial we saw that there appear to be differences between the results of the two methods. Although the differences between the results of the two methods appear to not be very large for the CoBalT data, we saw that there exist cases of simulated datasets for which there were even more noticeable differences between the methods and this higher level of uncertainty would have been ignored if the presented Bayesian approach had not been used. Furthermore, CUA was also conducted based directly on the approximate utility distributions which were derived in section 3.3.4 and we got evidence in favour of the robustness of the derivation of these mixtures.

Chapter 5

Discussion and future work

In terms of its structure, this chapter consists of two sections in which there is a discussion about the methodology, the derived results, the potential limitations of the project and the proposed future work. The first paragraph of the first section reiterates the research problem and summarises the major findings. Next, the key topics of this thesis are discussed in the order in which they were originally presented in the previous chapters. For each of these topics, when relevant, I interpret the results, discuss the implications, acknowledge the limitations and state my recommendations. The work that has been done is situated in the big picture. The final section is dedicated to plans for future work.

5.1 Discussion

The EQ-5D-3L is a popular instrument which is commonly used in economic evaluations; its wide spread requires the uncertainty associated with the utility scores to be accounted for. In the UK, the utility scores of the EQ-5D-3L states are usually calculated as the point estimates which were derived by the MVH project. However, these frequentist point estimates tacitly ignore the uncertainty of the estimates due to the variability inherent in the underlying data. This piece of work presented a Bayesian hierarchical model which accounts for model misspecification and the responses of the survey participants of the original evaluation survey of the MVH in order to assign a probability distribution to the utility score of every feasible EQ-5D-3L health state. Hence, this Bayesian model propagates the uncertainty by dealing with the variability of the parameters of the model inherent in the original evaluation survey. MCMC samples were obtained and credible intervals were computed to summarise the posterior utility distributions. Using numerical optimisation techniques (by considering the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm and the KL divergence as the criterion

of assessment), the posteriors of the utilities were summarised as three-component mixtures of Normal distributions and thus a new tariff was provided. This allows the conduction of sensitivity analysis by directly assigning a closed-form distribution to the utility score of an individual instead of having to make any further assumptions about the distributions of the benefits. The impact of the presented approach was examined by conducting CUA and applying my methods on the CoBalT trial as well as on simulated data whilst also accounting for differences between trial arms in mean baseline values. Differences were observed between the CUA results derived under my approach and those derived under the standard approach, especially for the simulated data. This shows that if the presented approach is not used and an important layer of uncertainty is not taken into consideration, it can lead to wrong conclusions when conducting CUA. Even in cases when the intervention is still considered to be cost-effective, there are differences with respect to the level of uncertainty associated with that decision; such information is useful to decision-makers and thus it is important to properly propagate the underlying uncertainty. Moreover, the model was found to be robust to changes in priors, initial values, seeds and there is a variety of potential extensions.

The principal aim of this report is to discuss a Bayesian model which extends the original framework of the MVH project by assigning probability distributions to each of the EQ-5D-3L states rather than mere point estimates. I set up a Bayesian model and MCMC samples were derived for the posterior utility distributions of the EQ-5D-3L states. The statistics show that there is no evidence of non-convergence for my model and it is fairly robust in changes in priors. Credible intervals provide a summary of the MCMC samples; the health state utility means calculated when using the standard frequentist approach of replicating the MVH are included in the credible intervals, providing evidence in favour of my model as good extension of the original MVH techniques.

Some of the original MVH assumptions can be considered imperfect, such as the assumption of normality for the observed evaluations of health states by the survey respondents. Moreover, other concepts which might be debated include the use of TTO as an appropriate method for eliciting utilities, the concept of cardinal utility and its application for EQ-5D-3L health states. Nevertheless, although not all the concepts of the MVH are ideal and some assumptions are not perfectly backed-up, the derived MVH tariff have been broadly used in economic appraisals in the UK for years and thus I do not create a new model from scratch which would use radically opposite assumptions, but one which respects the original assumptions and improves our knowledge about the EQ-5D-3L state utilities (from a Bayesian point of view) by accounting for the underlying uncertainty. My approach also accounts for model misspecification in line

with Pullenayegum, Chan and Xie (2015), who have emphasised the significance of accounting for the presence of important EQ-5D-3L model misspecification.

The initial attempt to replicate the results of the MVH project and to obtain the "final" dataset consisting of the responses of the 2997 individuals (after the exclusions) resulted in failing to obtain the same sample size with the York team. The first speculations were that this was potentially caused by numerical typographical errors in the original MVH literature, possibly in the parts where they describe the indeces of some participants (as some of the descriptions about data related to some particular survey participants appear to not necessarily match the actual data related to the individuals with those particular indeces which were mentioned). In fact, there have also been other attempts to replicate the results of the MVH project which faced the same issue; Gray et al. (2012) report that "after careful reading of the MVH and modelling documentation available, it was not possible to reach exactly the same sample size as the authors". In general, there have also been other instances of non-successful replications of published results; for example, the study of Camerer et al. (2016) found that approximately one-third of the experimental studies which they considered from economic journals failed to successfully replicate. Nevertheless, following an examination of the data and a careful consideration of the MVH literature (trying to find where the potential numerical typographical errors occurred), I was able to derive the correct sample size with what I believe to be the correct 2997 indeces and thus I was able to successfully replicate the results of the MVH project. Hence, these are the indeces which were used for the derivation of this thesis' results.

The use of different priors, starting values or seeds of the MCMC algorithm could produce results which are not identical to the the ones which were presented in the main body of the thesis (as it is usual in MCMC), yet they are still similar and there is no evidence of non-convergence. There are some changes in the order in which some particular states are ranked, but this is related to states the utilities of which were anyway considered to be very similar to each other. For example, by examining Figures 3.4 and 3.5 and Table 3.3 we can see that the values of the posterior medians of the states which switched ranks were almost equal to each other when both the original and the alternative priors were used. Their 95% credible intervals overlap (e.g. for example in Figure 3.4 we can see that the bounds of the credible intervals of states `31233` and `33132` are almost the same) and hence these states are considered to be more or less equal to each other in terms of their utility values regardless of whether some times one of them appears to be very slightly better than other and other times vice versa. Changes in the overall CUA results are negligible as the posterior means, medians and standard deviations of the EQ-5D-3L health state utility distributions take similar values and even the order of the

states in general more or less remains the same. Therefore, the overall results of my model are robust to changes in priors, initial values and seeds.

There is a greater interest in approximating the derived posteriors with closed form distributions, but this is not directly feasible to be done. Specifically, due to graphical evidence of multi-modality it is reasonable to model them using mixtures of distributions which will be capable of approximating the true distribution better. I follow the approach of similar studies (Schmidli et al., 2014) which approximate the posterior distributions as three-component mixtures. Maximum-likelihood optimisation techniques are used in order to estimate the parameters of the mixtures which provide the best fit to the derived simulations. In order to examine the quality of the obtained approximations, I use the KL divergence, the standard in the literature, as the assessment criterion.

Research work has examined aspects of KL-divergence-related topics, however, a degree of subjectivity is still required in answering the "how small is good" question. In general, the KL divergence values can be used to make comparisons between different approximate distributions and see which one appears to be a better approximation for the target distribution, but there is a also an interest to conclude whether the chosen distribution is considered similar enough to the target distribution. I propose a way to make an assessment of our satisfaction for the obtained KL divergence values.

In particular, we can set what we consider as an acceptable degree of dissimilarity between two distributions (i.e. the similarity is still "good enough") and then compute the theoretical value of the KL divergence between two such distributions. Since both of these distributions are known, the KL divergence value can be calculated directly from equation (3.3) and then we can use the derived KL divergence value as a reference point with regard to the KL divergence values which have been obtained for the approximations of the utility distributions of the EQ-5D-3L states. Specifically, for the case of one-component approximations, we can assume that two Gaussian distributions which have the same standard deviation of 0.071 (which is approximately the mean standard deviation of the univariate approximations which were derived) and an absolute means-difference value of 0.01 are still considered to be "similar enough". The theoretical KL divergence value for such Gaussian distributions is 0.009919, which is still larger than the largest KL-divergence value which were derived for the single-component approximations. Even if the absolute means-difference value is as low as 0.006, the theoretical KL divergence value is 0.003571 which is larger than the largest KL-divergence value of those derived for the three-component approximations. Therefore, we can claim that for the case of the three-component approximations, the dissimilarity of my derived distributions with the

target distributions is not worse than that of two Gaussian distributions which have the same standard deviation of 0.071 and an absolute mean difference of 0.006 (generally any two distributions which have a theoretical KL divergence difference of 0.003571). Figure 5.1 shows the probability density functions of two distributions which have a theoretical KL divergence difference of 0.003571 and we can see that the distributions appear to be quite similar.

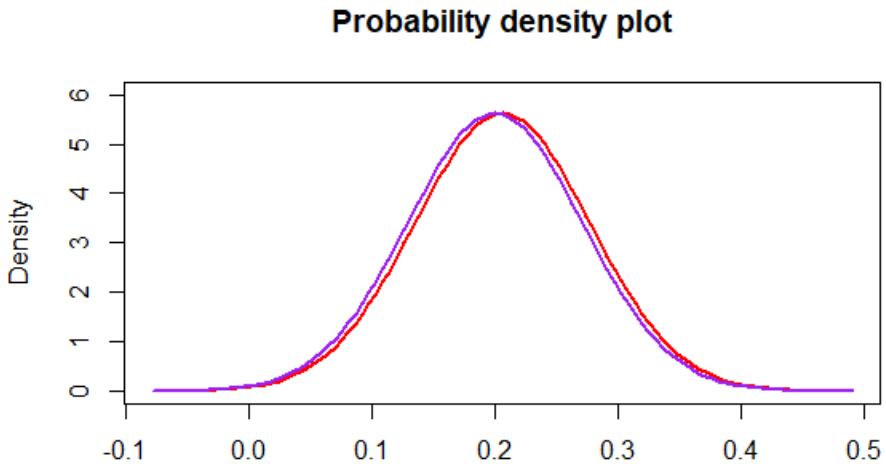


Figure 5.1: Probability density functions of two Gaussian distributions which have a theoretical KL divergence difference of 0.003571; in particular one distribution (in red) has a mean of 0.206 and the other distribution (in purple) has a mean of 0.2, whilst both distributions have a standard deviation of 0.071. The distributions appear to be quite similar. The achieved similarity of the derived three-component approximate distributions of this project and their corresponding target distributions is actually better than the similarity of the two distributions presented in this figure.

Next, after finding reasonable approximations of the distributions in a closed form, it is feasible to assign these to clinical trial participants, or any EQ-5D-3L respondents in general, according to their underlying EQ-5D-3L state. One main advantage of this thesis' innovation of knowing the precise utility distribution of each clinical trial subject is that we can perform sensitivity analysis with respect to the benefits of the intervention which are expressed in QALYS (a deterministic function of the patients health utility) and assess the robustness of the cost-effectiveness of the intervention. Cost-effectiveness analysis was conducted on the CoBalT trial and a simulated dataset using the Bayesian model which allowed us to directly sample from the posterior utility distributions of the EQ-5D-3L states experienced by the trial patients. We

compared the results with those obtained from the frequentist model, in which uncertainty is quantified using bootstrapping.

The innovations of these extensions of the original MVH work are beneficial for clinical trial researchers who can now assign close form utility distributions to trial subjects. Currently there is no literature for such utility distributions but I provide a tariff which has been derived by mapping from the MCMC samples of the posterior distributions of the Bayesian equivalent of the MVH model to closed form mixtures. Hence, the CUA results which assess the robustness of the cost-effectiveness of the intervention are based on these utility distributions. Moreover, there is no need to use bootstrapping or make any further assumptions about the distribution of the QALYs as we can sample directly from the derived distributions. Alternatively, these distributions could be used as a known reference point which eases the situation when having to deal with different assumptions made in separate economic evaluations.

On average, the KL divergence values decrease by approximately 40% when we use three instead of one component, whereas the corresponding KL divergence decrease is less than 10% when using five instead of three components. Since there is not big improvement when going from three to five components whereas the KL values are already relatively low, I recommend to using three component mixtures as they seem to provide a good approximation while still keeping the complexity of the algorithm relatively low compared to the increasing computational complexity of having more components.

Regarding the potential issue of lack of convergence (as it was presented in section 3.2.2) due to the algorithm being sensitive to the selection of the initial values: in general this is an issue when fitting finite mixture modules, which can suffer from local-optima problems (McLachlan and Peel, 2006; Shireman, Steinley and Brusco, 2016). Shireman, Steinley and Brusco (2015) contend that the accuracy of the final parameter estimates of the mixture models depends on the paramount choice of the initial values. My choice of initialising $\boldsymbol{B_0}$ as $\boldsymbol{I_J}$ in section 3.3.2 is in fashion with Nocedal and Wright (1999), but the initialisation of $\boldsymbol{x_0}^\top = (w_{0,1}, w_{0,2}, a_{0,1}, a_{0,2}, a_{0,3}, b_{0,1}, b_{0,2}, b_{0,3})$ is more heuristic as it is done by trial and error following a visual inspection of the target distribution. There are also other alternative ways of generating random starting values for fitting finite mixture modules using numerical optimisation techniques in general, some of which involve the generation of starting parameter values from a uniform distribution around some bounds informed by the data (Hipp and Bauer, 2006; Muthén and Muthén, 2012). Even if more complex ways are used for the selection of the starting values, we have to visually inspect the resulting mixture distributions to check if they have satisfactory captured the shape of the target distributions. My choice of the starting

values was also in line with the recommendations of IHS Markit (2019): a useful approach is to carefully consider the choice of the starting values so that they are as close as possible to what we believe that the optimum values might be whereas we should also avoid the use of starting values that are close to any regions in which the target function cannot be evaluated. In the end, with my choice of initial values the algorithm has successfully converged and the target distributions have been approximated with mixture distributions which are considered to be very good both in terms of closely capturing the shape of the target distributions (as it can be seen in the Figures of Appendix' section B) but also in terms of achieving small values of KL divergence.

Moreover, the CUA results based on the mixtures approximations were satisfactory and similar to those based on the main model which used the MCMC samples. This is in agreement with the derivation of the mixture distributions being robust.

The conversion of any utility values which were slightly greater than 1 (as it was described in section 3.5 for the MCMC values and for those derived from the mixtures) was based on theoretical grounds. However, the number of cases when there is a derivation of a utility score greater than 1 is very small and even then the derived values are very close to 1. Even without bounding the utilities by 1, the overall results and conclusions remain the same as the differences are negligible. Nevertheless, if an individual obtains any few values which are slightly greater than 1 for a limited number of states when sampling from the mixture distributions, they are advised to convert them to 1 for consistency reasons, same as the presented approach of this thesis. These greater-than-one values are also caused by the MCMC values of the β's, a limited number of which are negative (e.g for β_6 as it can be seen in table 3.1); these few slightly less-than-zero values are counter-weighted by the "upper extreme" values. The overall results of the entire thesis were bounded such that there are no utility values greater than 1 and no logical inconsistencies were observed in the overall ranking of the states (e.g. as it can be seen in Figures 3.1, 3.2 and 3.3) because any state which was ranked worse than another had at least one dimension at a worse level than the other better-ranked state.

The second part of the model (which was presented in section 4.2) could have been avoided if the decision was to not account for baseline utility differences. Without the regression adjustment of the second part of the model, my model would be a single-stage model: we would run the first part of the model (which was presented in chapter 3) and then we would just have to deterministically calculate the individual QUALYs as the AUC, by using equations (4.3) and (4.4). A single-stage model would be more easily implemented computationally, but I chose to have the full two-stage MCMC sampling instead, since I wanted to account for baseline utility

differences. The analysis of RCT's data is usually done on the assumption that the randomisation ensures that the characteristics of individuals are well-balanced between the trial arms, including baseline utilities. However, Manca, Hawkins and Sculpher (2005) argue that there will be some inevitable differences between groups in mean baseline values and in fact they demonstrate that in some cases the conclusions about the cost-effectiveness of the interventions can actually be opposite if no allowance is made for imbalance in baseline utility. Remarkably, a systematic review of QALY calculation methods and transparency of reporting found that only 23% of the studies they had sampled adjusted for baseline imbalance in mean utility (Richardson and Manca, 2004). Therefore, I decided to adapt the approach of the regression adjustment, which was also found to perform better compared to other alternatives (Manca, Hawkins and Sculpher, 2005). Thus, I recommend to use the two-stage approach in order to properly account for baseline utility differences but the methods of this thesis would also be applicable for the work of a researcher who would like to use these methods but they prefer to not adjust for baseline imbalance in mean utility.

The choice of values for C and C^* influences the size of the matrices of differential QALYs and costs (which have C rows and C^* columns each). The higher these values, the more computationally expensive (and time-consuming) the process is which can eventually be a problem for regular computers to handle in terms of RAM usage and data-storage. Dealing with `R` objects of large size can be an issue as well; some `R` packages might fail to work with very large `R` objects. In fact, even the `BCEA` package was unable to be functional with the initial size of the $C \times C^*$ matrices in section 4.3, which have 16,000,000 entries each. Then, summarising the information of such a matrix by a vector of size $G = 10,000$ (as it was described in section 4.2.2) is precise enough for our purposes. We desire that the values of C and C^* are large enough for the results to be precise whilst at the same time they should not be superfluously large, in order for regular computers and `R` packages to be able to successfully handle the task. Keeping in mind the results of the diagnostic statistics that were presented at the beginning of section 3.2.2 in order to assess the convergence of the MCMC, I chose the first part of the model to have $C = 4,000$ iterations. Even when the number of iterations of merely the first the part of the model (putting the second part of the model aside) was attempted to be increased by thousands, there were negligible differences and the diagnostic statistics provided no evidence of lack of convergence either. Hence, since there was no particular improvement by increasing even more the value of C and since that number would anyway have to be multiplied by C^* which can result in some very big matrices, the decision was to choose $C = 4,000$, which is considered to be sufficient for our purposes. A similar concept can be used for the second

part of the model to determine that C^* can be set equal to $4,000$. Therefore, the values of the aforementioned parameters were chosen in such a way so that the presented results are precise enough whilst still keeping these values within the limits of what a regular computer can handle.

A researcher who desires to conduct CUA using my approach has a choice about the second part of the model: they can either use the main Bayesian model or one of the models which use approximate distributions for the health state utilities (e.g. the three-component approximations). When the main Bayesian model is used, the values that were derived under the c-th iteration are used for the c-th step of the second part of the model (out of the total C times that the regression is run). When the model with the three-component approximations is used, the c-th values that were sampled from the approximate distributions are used for the c-th step of the second part of the model (out of the total C times that the regression is run). The main Bayesian model uses the entire MCMC sample which was derived from the first part of the model for the 243 EQ-5D-3L state utilities. Although in equation (3.1) we see how to obtain the utility score of the s-th EQ-5D-3L state in particular, in fact, at each MCMC iteration c we obtain a vector of utility scores (whose size is 243), which is computed using the c-th MCMC values of the β and $\xi^{(new)}$ parameters.

In other words, in the main Bayesian model there is a dependence between the health state utilities which were derived under the c-th iteration. However, for the three-component mixture model there is no such theoretical dependence between the health state utility scores which were sampled from their corresponding approximate distributions at step c, because each EQ-5D-3L state utility is sampled independently from its distribution. Nevertheless, it appears that it is not always a necessity to consider the aforementioned dependence, as it was discussed in section 3.4. Specifically, when we sample from the approximate distributions, most of the times the results of the CUA are similar to those of the main Bayesian model, even though the values were produced under different seeds. By evaluating the values of VAR_2/VAR_1 we can see how much % of an over- or under-estimation we have, since, VAR_1 is the "truth" and VAR_2 is the "approximation". From Table 3.5, we can see that 50% of the times we will have between a 2.9% of underestimation and a 10.7% of overestimation; whereas 95% of the times we will have between a 14.4% of underestimation and a 26.1% of overestimation. However, if one prefers to avoid the case of having a substantial under- or over-estimation, they could consider the results of the main Bayesian model or they could potentially consider an imputation approach such as the one presented by Chan, Xie, Willan and Pullenayegum (2016). Furthermore, given the way that the values of the health state utilities depend on the β's (as it can be seen in

equation (3.1)), it would be reasonable to assume that if the utility value derived for one state under the $c-$th step of the second part of the model was lower (or higher) than usual, then the utilities values derived for the remaining of the states under the $c-$ iteration would also be lower (or higher) than usual. Hence, another potential way to help minimising the effects of the aforementioned correlation issue if we decide to sample from the approximate distributions would be to use an approach such that when we sample the $c-$th set of utilities values from the 243 distributions: for each of the distributions we either sample from the lower- middle- or upper-part of the distribution.

When feasible, I recommend the use of the main Bayesian model, which is preferred for the consideration of all the information which is provided by the complete set of the MCMC samples which were obtained in the first part of the module. In fact, these MCMC samples of size $4,000 \times 243$ could be published (or they can be incorporated in an `R` package) so that the researchers can have direct access to them for their research, if they do not want to re-run the first part of the model. Otherwise, another option to avoid the use of the MCMC samples (and to re-run the first part of the model in general) for the researcher is to use the approximate mixture distributions which were presented in section 3.3.4, as we saw that most of the times the inference remains the same regardless of whether the full MCMC samples are used or the approximate distributions are used. Although the advantages of using the mixture distributions are discussed throughout this thesis, the reader should be reminded that this approach remains an approximation. Ultimately, it is a judgement call on the part of the researcher as to whether or not the stated approximation is accurate enough for their purposes.

Assuming ICER thresholds less than £20,000, the conclusions of the CoBalT cost-effectiveness analyses are the same for both the Bayesian approach introduced in this thesis and the bootstrap approach: the intervention is cost-effective. Nevertheless, there are differences between the results produced under these two methods. A decision-maker desires to know the level of uncertainty associated which each of their decisions, which is important information when reporting the CUA results and for the purposes of sensitivity analysis. The analysis of my approach shows that there is an increased level of uncertainty which is accounted for, unlike the standard approach. Therefore, I believe that it is better practice to use the approach presented in this thesis, in order to take into consideration the uncertainty of the estimates which would otherwise be ignored.

However, unlike the case of the CoBalT trial where the differences between the results produced under the bootstrap approach and the Bayesian model appear to not be very big, for the simulated dataset which I presented we observe even more substantial differences. This

supports further the argument that it is better to use the proposed approach of this thesis. It is important to use simulations when comparing the implications of the use of some methods and that's why I took into consideration multiple simulated datasets for the comparison of the two methods. In fact, Morris, White and Crowther (2019) note that "even if the choice is unlikely to materially affect the results, it may be useful to have unrealistically extreme data-generating mechanisms to understand when and how each method fails".

As a result of the different mathematical properties of the two methods, there exist datasets for which the produced results of the two approaches can be substantially different. The use of the standard approach means that bootstrap deals with trial-sampling variation, but it does not deal with the variability of the parameters of the MVH model. However, my approach was designed in such a way to properly propagate the uncertainty by dealing with the variability of the parameters of the model inherent in the original dataset of the MVH survey. This uncertainty would otherwise be ignored. The comparison of the two methods showed that when this uncertainty is ignored, there exist cases when the decision of which intervention to fund can be reversed or even if it is not reversed, the degree of certainty with which each decision is made can change substantially. Furthermore, the widths of the derived credible intervals in section 3.2.2 are large enough, given the fact that minimum important difference for EQ-5D-3L has been reported to be as low as 0.03 and generally ranges from 0.03 to 0.08 (Le, Doctor, Zoellner and Feeny, 2013; Pickard, Neary and Cella, 2007; Soer et al., 2012; Walters and Brazier, 2005). Considering all of these, sensitivity analysis is an essential part of CUA, I believe that it is important to not ignore this aforementioned layer of uncertainty and thus I recommend the approach presented in this thesis as the desired approach to use when conducting CUA.

Other research work on the USA EQ-5D-3L value set from a Bayesian perspective has argued about the importance of quantifying the uncertainty of the utility values (Pullenayegum, Chan and Xie, 2015). Moreover, some researchers have worked on similar health instruments, such as the SF-6D, whilst focusing on using the available valuation data in a more efficient way in order to produce value sets which are subject to less parameter uncertainty (Kharroubi, Brazier and O'Hagan, 2007; Kharroubi et al., 2007; Kharroubi, O'Hagan and Brazier, 2005), whilst others have attempted to use a bootstrap approach for dealing with the underlying uncertainty (Gray et al., 2012). Nevertheless, there have not been any other attempts yet in reporting the utility distributions of the distinct UK EQ-5D-3L states as explicitly specified probability distributions.

Chan, Xie, Willan and Pullenayegum (2016) used a multiple imputation type approach in order to deal with data related to the US EQ-5D-3L health state utilities. Assuming that their

imputed values are published and available to researchers, researches working on US EQ-5D-3L data could use their findings without the need to re-run an entire model for scratch, which is convenient. This is also an advantage of my work in terms of using the derived mixture distributions. However, the results of Chan, Xie, Willan and Pullenayegum (2016) can sometimes lead to logical inconsistencies (e.g. when one health state which is supposed to be no worse than another particular state is actually valued as the worst one); theoretically, this can also occur with the results of my approach, for instance if extreme values are sampled from the lower or upper end of the distributions of two states. Nevertheless, the individual imputed value set is not meant to be used in isolation, but all the imputed sets together would provide an estimation of the appropriate degree of uncertainty of the health instrument. Furthermore, another advantage of my approach is that, theoretically, there is no upper limit to how many times a researcher can choose to sample values from the derived mixture distributions, which is useful if the objective is to use more values than the fixed number of published values. Furthermore, in terms of accessibility it is easier to describe the EQ-5D-3L health state utilities by presenting the derived utility distributions instead of presenting a large number of value sets.

Extensions of the methods presented in this thesis include the construction of similar models about other health instruments. These methods can also be applied on populations of different countries.

The methods of this project can also be extended in the context of decision analytic modelling for economic evaluation. A decision analytic model is the fundamental tool of decision analysis in which decisions under uncertainty are made using a systematic, quantitative and transparent approach (Kuntz et al., 2013); usually a decision tree or a Markov model is used. Probabilistic sensitivity analysis (PSA) (York Health Economics Consortium, 2016) allows the modeller to assign probability distributions to the parameters of a decision analytic model and thus quantify the level of confidence in the output of the analysis in relation to uncertainty in the model inputs by sampling a set of input parameter values from these distributions. It has become standard practice to use MCMC for performing probabilistic sensitivity analysis in order to explore the uncertainty of input parameters (Claxton et al., 2005). Different distributions are assigned to different types of variables, such as the utility associated with a specific health state which is part of the decision analytic model. Nevertheless, Ara and Wailoo (2012) believe that there is a theme of lack of transparency and clarity in reports which describe the methodologies used when applying health state utility values in decision analytic models.

In the case that the modeller has to value the utility of a particular non-EQ-5D-3L health state of their decision analytic model using the EQ-5D-3L as an instrument (which is recommended

by NICE), they have to estimate it. Data from a clinical trial can be used: if the trial subjects who experience that particular health state of the decision analytic model have answered the EQ-5D-3L questionnaire then the corresponding point estimates of the utility values of their EQ-5D-3L states can be computed and then the average of these values can be used in order to estimate the decision analytic model's utility value of interest. Under this approach, when conducting probabilistic sensitivity analysis, the parameters of a distribution referring to a utility value can be set in such a way that the uncertainty expressed by the resulting distribution is based on error estimates from data sources. However, this practice deals with trial-sampling variation and not with the uncertainty of the parameters of the MVH project, unlike the approach of this thesis.

Specifically, my methods can be extended in the area of model-based economic evaluations even if the health states of the decision analytic model are not actual EQ-5D-3L states. This time, for the estimation of the utility of such a health state of the decision analytic model, we are not interested in obtaining information about the average of the point estimates of the EQ-5D-3L health state utilities of the participants of a clinical trial who experience a specific health state of the decision analytic model. Instead, we are interested in obtaining information about the actual EQ-5D-3L states of the trial participants who have stated that they experience the specific health state of interest of the decision analytic model; in particular we are interested in knowing the proportion ψ_s of each EQ-5D-3L state over the subjects who experience the specific health state of the decision analytic model. Then, we can estimate the utility of that heath state of interest at the c-th iteration ($c = 1, ..., C$) as a weighted average: $\psi_1^{(c)} u_1^{(c)} + \psi_2^{(c)} u_2^{(c)} + ... + \psi_S^{(c)} u_S^{(c)}$, where S is the number of EQ-5D-3L states and the values of the ψ parameters can be sampled from a distribution such as a Dirichlet distribution with concentration parameters informed from trial data as described earlier. The utility value of the s-th EQ-5D-3L state at iteration c can either be set equal to the c-th value of the MCMC sample which was obtained in section 3.2 or it can be sampled from the corresponding probability distribution that was derived in section 3.3. Similarly, the value of C can be chosen to either be (less than or) equal to the value of C used in section 3.2, or it can be set equal to a different value if there is particular preference to sample a specific amount of utility values.

Ara and Wailoo (2012) note that uncertainty around health state utility values is usually underreported, whereas frequently only mean values are used in decision analytic models. Probability distributions can be assigned to utility scores in the context of probabilistic sensitivity analysis but the choice of the parameters of the distributions is made by considering trial-sampling variation whereas in such a case MVH-parameter uncertainty is ignored. I have derived prob-

ability distributions of the utility scores of the EQ-5D-3L health states and these reflect the MVH-parameter uncertainty as well as they take model-misspecification into consideration; thus, I have provided some distributions for researchers to use without the need to make other assumptions on the forms of such distributions. Therefore, I support the extension of the methods of this project as it was discussed here in order to propagate the uncertainty of the parameters which would otherwise be ignored in model-based economic evaluations. In order to properly assess the impact of my proposed approach to consider MVH-parameter uncertainty in model-based economic evaluations, more work should be undertaken within the field of decision analytic modelling.

5.2 Plans for future work

I have derived the posterior distributions of the 243 EQ-5D-3L state utilities and I approximated them as three-component mixtures of normal distributions. The impact of this additive uncertainty on separate economic appraisals, which is currently ignored, can then be investigated further. Having explicitly described these posterior distributions in a closed form, researchers can use them in sensitivity analyses and examine whether the results of an economic evaluation are robust to changes in utility values and thus changes in QALYs. Each clinical trial individual can then be assigned a probability distribution which corresponds to their utility given the EQ-5D-3L state which they experience. Therefore, further worked examples can be done to apply the tariff of this thesis on clinical trials and further examine the impact of my work on economic appraisals. The tariff can also be used for model-based economic evaluations as in such a case the useful advantage is that we can use the derived probability distributions of this project without making any further assumptions about the distributions of the benefits of the economic evaluation.

Future work also includes the extension of these methods to other health instruments. For example, the SF-6D enjoys lots of popularity as well; it is frequently used in economic evaluations and it can be commonly found in literature. Therefore, it makes a good candidate as it will be of interest to apply my approach to it and compare the outcome with the available literature. Another obvious candidate is the five-levels version of the EQ-5D-3L: the EQ-5D-5L. The comparison can focus on spotting any pattern differences between the three and five levels versions of the health instrument.

In addition to extending the methods to alternative instruments, the approach can also be extended to instruments of different countries. Thus, we would be able to compare the results of

the UK EQ-5D-3L tariff with the derived tariffs of datasets of different countries. The primary point of interest in these comparisons would be the variability of the utility scores. The UK dataset is one of the biggest among those used in published reports. For instance, it would interesting to examine the variability of the derived results for a dataset with more survey participants (such as the US dataset) and for a dataset with fewer survey participants (such as the dataset of Japan).

Furthermore, the methods of this project could be applied in the area of mapping utility scores across health instruments. Sometimes there are clinical trials with individuals who have completed a disease-specific instrument, such as the Karnofsky Performance Status Scale for cancer (Karnofsky et al., 1948), but not a generic one such as the EQ-5D-3L. Nevertheless, having the patients' utilities scores on a generic instrument provides information which simplifies the comparison of cost-effectiveness across different disease areas. Researchers have developed frequentist techniques (for example Dixon, Dakin and Wordsworth, 2016; Teckle et al., 2013) for estimating a statistical relationship between instruments and predicting the health state utility values of one instrument based on the available responses of another. The application of my methods would provide the estimation of full probability distributions for the utilities when mapping across instruments, instead of the mere point estimates, which will bring the kind of advantages already discussed in this piece of work.

In addition, an `R`-package which assigns the derived probability distributions to individuals according to their health state will be helpful in sensitivity analyses projects. The package could also become available on-line in the form of an `R Shiny` web application (RStudio, Inc, 2020) which will make the use of the package more accessible and especially easier to those researchers unfamiliar with statistical programming.

5.3 Summary

Although the original concepts of the MVH project may not have been perfect, this project makes a contribution to the field by building a Bayesian adaptation of the MVH model and by going a step further and propagating the underlying uncertainty of the EQ-5D-3L by dealing with the variability of the parameters of the model inherent in the original valuation survey of the MVH project. My model also accounts for model-misspecification as well as for for differences between trial arms in mean utilities when it is applied on clinical trial data: the CoBalT trial and simulated data. The comparison of the CUA results produced under my approach with those under the standard approach shows that there can be substantial differences which can

lead to wrong inference. The presented approach properly accounts for the uncertainty which would otherwise be ignored and hence I recommend to using it when conducting CUA. MCMC samples were derived for the posterior utilities; the total size is $4,000 \times 243$. A new tariff was presented: three-component mixtures of Normal distributions were assigned as approximations to the utility scores of the 243 EQ-5D-3L states. The three-component approximations are considered satisfactory and the model is robust to changes in priors, initial values and seeds. Use of the MCMC samples is preferred, when feasible, otherwise the utility scores can also be sampled from the derived distributions when conducting CUA; the utilities used should be bounded by 1. These distributions can be used in economic evaluations for sensitivity analysis without making further assumptions about the distributions of the benefits. Similar techniques can be used for datasets of other countries, as well as other health instruments. These methods can also be applied when mapping across health instruments and in the context of model-based economic evaluations. `R` packages can exhibit the tariff of my work to other researchers and health economists in order to ease their tasks if they decide to use them.

The `JAGS` code used to produce the results, as well as the detailed tables and figures related to the posterior distributions of the 243 EQ-5D-3L states are aggregated in the Appendix. The Appendix also contains the information needed to replicate the simulated dataset.

References

Appleby, J., Devlin, N. and Parkin, D. (2007). NICE's cost effectiveness threshold. *BMJ*, 335(7616), pp.358-359.

Ara, R. and Wailoo, A. (2012). Using Health State Utility Values in Models Exploring the Cost-Effectiveness of Health Technologies. *Value in Health*, 15(6), pp.971-974.

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5), pp.941-973.

Attema, A., Edelaar-Peeters, Y., Versteegh, M. and Stolk, E. (2013). Time trade-off: one methodology, different methods. *The European Journal of Health Economics*, 14(S1), pp.53-64.

Baio, G. (2013). *Bayesian methods in health economics*. Boca Raton: Chapman & Hall/CRC.

Baio, G., Berardi, A. and Heath, A. (2019). *BCEA: Bayesian Cost Effectiveness Analysis.* R Package Version 2.3-1.1. [online] Available at: <http://cran.r-project.org/web/packages/BCEA/index.html>.

Barnsley, P., Towse, A., Karlsberg Schaffer, S. and Sussex, J. (2013). *Critique of CHE Research Paper 81: Methods for the Estimation of the NICE Cost Effectiveness Threshold*. Occasional Paper 13/01. London: Office of Health Economics.

Bates, D., Maechler, M., Bolker, B. and Walker, S. (2020). *Lme4: Linear Mixed-Effects Models Using Eigen And S4.* R Package Version 1.1-23. [online] Available at: <http://cran.r-project.org/web/packages/lme4/index.html>.

Beck, A., Steer, R. and Brown, G. (1996). *Manual For The Beck Depression Inventory-II.* San Antonio, TX: The Psychological Corporation.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory.* Chichester: Wiley.

Brazier, J., Roberts, J. and Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21(2), pp.271-292.

Briggs, A., Claxton, K. and Sculpher, M. (2011). *Decision modelling for health economic evaluation.* Oxford: Oxford University Press.

Brooks, R. (1996). EuroQol: the current state of play. *Health Policy*, 37(1), pp.53-72.

Broyden, C. (1970). The Convergence of a Class of Double-rank Minimization Algorithms. *IMA Journal of Applied Mathematics*, 6(3), pp.222-231.

Camerer, C., Dreber, A., Forsell, E., Ho, T., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M. and Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), pp.1433-1436.

Canadian Agency for Drugs and Technologies in Health, (2006). *Guidelines for the Economic Evaluation of Health Technologies: Canada.* 3rd ed. Ottawa.

Chan, K., Xie, F., Willan, A. and Pullenayegum, E. (2016). Underestimation of Variance of Predicted Health Utilities Derived from Multiattribute Utility Instruments. *Medical Decision Making*, 37(3), pp.262-272.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P. and Sculpher, M. (2013). *Methods for the Estimation of the NICE Cost Effectiveness Threshold.* CHE Research Paper 81. University of York: Centre for Health Economics.

Cleemput, I., Kesteloot, K., Moons, P., Vanrenterghem, Y., Van Hooff, J., Squifflet, J. and De Geest, S. (2004). The Construct and Concurrent Validity of the EQ-5D in a Renal Transplant Population. *Value in Health*, 7(4), pp.499-509.

Cohen, D. and Reynolds, M. (2008). Interpreting the Results of Cost-Effectiveness Studies. *Journal of the American College of Cardiology*, 52(25), pp.2119-2126.

Dai, Y. (2012). A perfect example for the BFGS method. *Mathematical Programming*, 138(1-2), pp.501-530.

Devlin, N., Hansen, P., Kind, P. and Williams, A. (2003). Logical inconsistencies in survey respondents' health state valuations - a methodological challenge for estimating social tariffs. *Health Economics*, 12(7), pp.529-544.

Dixon, P., Dakin, H. and Wordsworth, S. (2016). Generic and disease-specific estimates of quality of life in macular degeneration: mapping the MacDQoL onto the EQ-5D-3L. *Quality of Life Research*, 25(4), pp.935-945.

Dolan, P. (1997). Modeling Valuations for EuroQol Health States. *Medical Care*, 35(11), pp.1095-1108.

Dolan, P. and Gudex, C. (1995). Time preference, duration and health state valuations. *Health Economics*, 4(4), pp.289-299.

Dolan, P., Gudex, C., Kind, P. and Williams, A. (1996). Valuing health states: A comparison of methods. *Journal of Health Economics*, 15(2), pp.209-231.

Drummond, M. and McGuire, A. (2010). *Economic evaluation in health care.* Oxford: Oxford University Press.

Drummond, M., Sculpher, M., Torrance, G., O'Brien, B. and Stoddart, G. (2005). *Methods for the Economic Evaluation of Health Care Programmes.* 3rd ed. Oxford: Oxford University Press.

Erens, R. (1994). *Health Related Quality of Life: General Population Survey.* Technical report. London: Social and Community Planning Research.

European Consortium in Healthcare Outcomes and Cost-Benefit Research, (2013). *European Guidelines for Cost-Effectiveness Assessments of Health Technologies.* Switzerland.

Feng, Y., Devlin, N., Shah, K., Mulhern, B. and Van Hout, B. (2016). *New Methods for Modelling EQ-5D-5L Value Sets: An Application to English Data.* Research Paper 16/02. London: Office of Health Economics.

Fletcher, R. (1970). A new approach to variable metric algorithms. *JThe Computer Journal*, 13(3), pp.317-322.

Gamerman, D. (1997). *Markov chain Monte Carlo.* 1st ed. London: Chapman & Hall.

Garza, A. (2003). Health Utility Measures and the Standard Gamble. *Academic Emergency Medicine*, 10(4), pp.360-363.

Gelman, A. and Rubin, D. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4), pp.457-472.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D. (2013). *Bayesian Data Analysis.* 3rd ed. Boca Raton: Chapman & Hall/CRC.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *In Bayesian Statistics*, 4, pp.168-193.

Gilks, W., Richardson, S. and Spiegelhalter, D. (1996). *Markov chain Monte Carlo in practice.* 1st ed. London: Chapman & Hall.

Goldfarb, D. (1970). A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, 24(109), p.23.

Gray, A., Rivero Arias, O., Leal, J., Dakin, H. and Ramos Goni, J. (2012). *How Important Is Parameter Uncertainty Around The UK EQ-5D-3L Value Set When Estimating Treatment Effects?*. [online] Available at: <http://www.ces-asso.org/sites/default/files/Gray%20et%20al%20HESG%20submitted%201.pdf>.

Greene, W. (1992). *LIMDEP Version 6.0: User's Manual And Reference Guide*. New York, NY: Econometric Software Inc.

Gudex, C. (1994). *Time Trade-off User Manual: Props and self-completion methods*. University of York: Centre for Health Economics.

Gudex, C., Dolan, P., Williams, A.H., Kind, P. (1995). *Health State Valuations from the British General Public, 1993*. [data collection]. UK Data Service. SN: 3444. Available at: <http://doi.org/10.5255/UKDA-SN-3444-1>.

Guyatt, G. (1993). Measuring Health-Related Quality of Life. *Annals of Internal Medicine*, 118(8), p.622.

Hamashima, C. (2002). Long-term quality of life of postoperative rectal cancer patients. *Journal of Gastroenterology and Hepatology*, 17(5), pp.571-576.

Hill, A. (2013). NICE hits back at 'limited' EC study. *Pharmafile*. [online] Available at: <http://www.pharmafile.com/news/177270/nice-hits-back-limited-ec-study>.

Hipp, J. and Bauer, D. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, 11(1), pp.36-53.

Hollinghurst, S., Carroll, F., Abel, A., Campbell, J., Garland, A., Jerrom, B., Kessler, D., Kuyken, W., Morrison, J., Ridgway, N., Thomas, L., Turner, K., Williams, C., Peters, T., Lewis, G. and Wiles, N. (2013). Cost-effectiveness of cognitive-behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: economic evaluation of the CoBalT Trial. *The British Journal of Psychiatry*, 204(1), pp.69-76.

Hunter, R., Baio, G., Butt, T., Morris, S., Round, J. and Freemantle, N. (2015). An Educational Review of the Statistical Issues in Analysing Utility Data for Cost-Utility Analysis. *PharmacoEconomics*, 33(4), pp.355-366.

IHS Markit, (2019). *Command Reference : User-Defined Optimization : Technical Details.* [online] EViews. Available at: <http://www.eviews.com/help/helpintro.html#page/content%2Foptimize-Technical_Details.html%23ww239519>.

Karnofsky, D., Abelmann, W., Craver, L. and Burchenal, J. (1948). The use of the nitrogen mustards in the palliative treatment of carcinoma.With particular reference to bronchogenic carcinoma. *Cancer*, 1(4), pp.634-656.

Khan, I. (2016). *Design & analysis of clinical trials for economic evaluation & reimbursement: an applied approach using SAS & STATA*. Boca Raton, FL: Chapman & Hall/CRC.

Kharroubi, S., Brazier, J. and O'Hagan, A. (2007). Modelling covariates for the SF-6D standard gamble health state preference data using a nonparametric Bayesian method. *Social Science & Medicine*, 64(6), pp.1242-1252.

Kharroubi, S., Brazier, J., Roberts, J. and O'Hagan, A. (2007). Modelling SF-6D health state preference data using a nonparametric Bayesian method. *Journal of Health Economics*, 26(3), pp.597-612.

Kharroubi, S., O'Hagan, A. and Brazier, J. (2005). Estimating utilities from individual health preference data: a nonparametric Bayesian method. J *Royal Statistical Soc C*, 54(5), pp.879-895.

Kuntz, K., Sainfort, F., Butler, M., Taylor, B., Kulasingam, S., Gregory, S., Mann, E., Anderson, J. and Kane, R. (2013). *Decision And Simulation Modeling In Systematic Reviews.* Rockville, MD: Agency for Healthcare Research and Quality (US).

Le, Q., Doctor, J., Zoellner, L. and Feeny, N. (2013). Minimal clinically important differences for the EQ-5D and QWB-SA in Post-traumatic Stress Disorder (PTSD): results from a Doubly Randomized Preference Trial (DRPT). *Health and Quality of Life Outcomes*, 11(1), p.59.

Lee, S., Kim, D., Oh, J., Han, H., Yoo, K. and Kim, H. (2003). Validation of a Functional Evaluation System in Patients With Musculoskeletal Tumors. *Clinical Orthopaedics and Related Research*, 411, pp.217-226.

Lohr, K., Aaronson, N., Alonso, J., Audrey Burnam, M., Patrick, D., Perrin, E. and Roberts, J. (1996). Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clinical Therapeutics*, 18(5), pp.979-992.

Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2012). *The BUGS book*.

Manca, A., Hawkins, N. and Sculpher, M. (2005). Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Economics*, 14(5), pp.487-496.

Marmot, M. and Allen, J. (2014). Social Determinants of Health Equity. *American Journal of Public Health*, 104(S4), pp.S517-S519.

Martin, A., Glasziou, P., Simes, R. and Lumley, T. (2000). A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *International Journal of Technology Assessment in Health Care*, 16(1), pp.137-147.

McCabe, C., Claxton, K. and Culyer, A. (2008). The NICE Cost-Effectiveness Threshold: What it is and What that Means. *PharmacoEconomics*, 26(9), pp.733-744.

McLachlan, G. and Peel, D. (2006). *Finite Mixture Models*. New York, NY: John Wiley & Sons.

Morris, S., Devlin, N. and Parkin, D. (2007). *Economic analysis in health care*. Chichester: J. Wiley & Sons.

Morris, T., White, I. and Crowther, M. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), pp.2074-2102.

Muthén, L. and Muthén, B. (2012). *Mplus User's Guide*. 7th ed. Los Angeles, CA: Muthén & Muthén.

MVH Group, (1995). *The Measurement and Valuation of health. Final Report on the Modelling of Valuation Tariffs.* University of York: Centre for Health Economics.

MVH Group, (1994). *The Measurement and Valuation of health. First Report of The Main Survey*. University of York: Centre for Health Economics.

National Institute for Health and Care Excellence, (2013). *Guide to the methods of technology appraisal 2013*. Process and methods guides. [online] London. Available at: <http://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781>.

Nocedal, J. and Wright, S. (1999). *Numerical optimization*. New York: Springer.

Nord, E., Pinto, J., Richardson, J., Menzel, P. and Ubel, P. (1999). Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health Economics*, 8(1), pp.25-39.

Norum, J. (1996). Quality of life (QoL) measurement in economical analysis in cancer. *Oncology Reports*.

O'Hagan, A. and Forster, J. (2004). *Vol 2B Kendall's advanced theory of statistics Bayesian inference*. New York: Arnold.

Parkin, D. and Devlin, N. (2006). Is there a case for using visual analogue scale valuations in cost-utility analysis?. *Health Economics*, 15(7), pp.653-664.

Patrick, D., Starks, H., Cain, K., Uhlmann, R. and Pearlman, R. (1994). Measuring Preferences for Health States Worse than Death. *Medical Decision Making*, 14(1), pp.9-18.

Pharmaceutical Benefits Advisory Committee, (2013). *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee.* vers. 4.4 Canberra.

Phelps, C. (2009). *Health economics.* 4th ed. New York: Pearson

Phillips, C. (2009). *What is a QALY?.* 2nd ed. London: Hayward Medical Communications.

Pickard, A., Neary, M. and Cella, D. (2007). Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health and Quality of Life Outcomes*, 5(1).

Plummer, M. (2017). *Just Another Gibbs Sampler (JAGS).* [online] Available at: <http://mcmc-jags.sourceforge.net/>.

Prieto, L. and Sacristán, J. (2003). Problems and solutions in calculating quality-adjusted life years (QALYs). *Health and quality of life outcomes*, 1, p.80.

Pullenayegum, E., Chan, K. and Xie, F. (2015). Quantifying Parameter Uncertainty in EQ-5D-3L Value Sets and Its Impact on Studies That Use the EQ-5D-3L to Measure Health Utility: A Bayesian Approach. *Medical Decision Making*, 36(2), pp.223-233.

Raftery, A. and Lewis, S. (1992). [Practical Markov Chain Monte Carlo]: Comment: One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo. *Statistical Science*, 7(4), pp.493-497.

R Core Team, (2020). *The R Project For Statistical Computing.* [online] Available at: <http://http//www.R-project.org/>.

Raftery, J. (2014). NICE's Cost-Effectiveness Range: Should it be Lowered?. *PharmacoEconomics*, 32(7), pp.613-615.

Richardson, G. and Manca, A., (2004). Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. *Health Economics*, 13(12), pp.1203-1210.

Robert, C. and Casella, G. (2010). *Introducing Monte Carlo methods with R*. 1st ed. New York: Springer.

Robinson, R. (1993). Cost-utility analysis. *British Medical Journal*, 307(6908), pp.859-862.

Rosser, R. and Watts, V. (1972). The Measurement of Hospital Output. *International Journal of Epidemiology*, 1(4), pp.361-368.

RStudio, Inc, (2020). *Shiny: Easy Web Applications In R.* [online] Available at: <http://rstudio.com/products/shiny/>.

Rubinshtein, Y. (1993). Possibility of approximating multimodal distributions by mixtures of standard probability density functions. *Measurement Techniques*, 36(8), pp.858-864.

Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D. and Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4), pp.1023-1032.

Schneider, S., Pouget, I., Staccini, P., Rampal, P. and Hebuterne, X. (2000). Quality of life in long-term home enteral nutrition patients. *Clinical Nutrition*, 19(1), pp.23-28.

Shanno, D. (1970). Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24(111), p.647.

Shaw, J., Johnson, J. and Coons, S. (2005). US Valuation of the EQ-5D Health States: Development and Testing of the D1 Valuation Model. *Medical Care*, 43(3), pp.203-220.

Sheather, S. (2004). Density Estimation. *Statistical Science*, 19(4), pp.588-597.

Shireman, E., Steinley, D. and Brusco, M. (2015). Examining the effect of initialization strategies on the performance of Gaussian mixture modeling. *Behavior Research Methods*, 49(1), pp.282-293.

Shireman, E., Steinley, D. and Brusco, M. (2016). Local Optima in Mixture Modeling. *Multivariate Behavioral Research*, 51(4), pp.466-481.

Soer, R., Reneman, M., Speijer, B., Coppes, M. and Vroomen, P. (2012). Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. *The Spine Journal*, 12(11), pp.1035-1039.

Spiegelhalter, D., Abrams, K. and Myles, J. (2011). *Bayesian approaches to clinical trials and health-care evaluation*. Chichester: John Wiley & Sons.

Su, Y. and Yajima, M. (2020). *R2jags: Using R To Run 'JAGS'. R Package Version 2.3-1.1.* [online] Available at: <http://cran.r-project.org/web/packages/R2jags/index.html>.

Stafford, M., Hareendran, A., Ng-Mak, D., Insinga, R., Xu, R. and Stull, D. (2012). EQ-5D™-derived utility values for different levels of migraine severity from a UK sample of migraineurs. *Health and Quality of Life Outcomes*, 10(1), p.65.

Szende, A., Oppe, M. and Devlin, N. (2007). *EQ-5D value sets: Inventory, Comparative Review and User Guide*. Dordrecht: Springer.

Teckle, P., McTaggart-Cowan, H., Van der Hoek, K., Chia, S., Melosky, B., Gelmon, K. and Peacock, S. (2013). Mapping the FACT-G cancer-specific quality of life instrument to the EQ-5D and SF-6D. *Health and Quality of Life Outcomes*, 11(1), p.203.

Torrance, G. (1976). Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Economic Planning Sciences*, 10(3), pp.129-136.

Torrance, G. (1986). Measurement of health state utilities for economic appraisal: A review. *Journal of Health Economics*, 5(1), pp.1-30.

van Roijen, L., Nijs, H., Avezaat, C., Karlsson, G., Linquist, C., Pauw, K. and Rutten, F. (1997). Costs and effects of microsurgery versus radiosurgery in treating acoustic neuroma. *Acta Neurochirurgica*, 139(10), pp.942-948.

Veith, F. (2011). How Randomized Controlled Trials (RCTs) Can Be Misleading: Introduction. *Seminars in Vascular Surgery*, 24(3), pp.143-145.

Walters, S. and Brazier, J., (2005). Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research*, 14(6), pp.1523-1532.

Weinstein, M., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C. and Luce, B. (2003). Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value in Health*, 6(1), pp.9-17.

Wewers, M. and Lowe, N. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, 13(4), pp.227-236.

Wiles, N., Thomas, L., Abel, A., Ridgway, N., Turner, N., Campbell, J., Garland, A., Hollinghurst, S., Jerrom, B., Kessler, D., Kuyken, W., Morrison, J., Turner, K., Williams, C., Peters, T. and Lewis, G. (2013). Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. *The Lancet*, 381(9864), pp.375-384.

Wolfe, P. (1971). Convergence Conditions for Ascent Methods. II: Some Corrections. *SIAM Review*, 13(2), pp.185-188.

York Health Economics Consortium, (2016). *Probabilistic/Stochastic Sensitivity Analysis.* [online] Available at: <http://yhec.co.uk/glossary/probabilistic-stochastic-sensitivity-analysis/>.

Appendix

A JAGS-code & simulated dataset

The first half of the section contains some JAGS-code. At the second half of this section, there is information about the principal simulated dataset which was discussed in section 4.3.2.
First, we start with the presentation of the JAGS-code needed to replicate the results of the project.

JAGS code for the first part model, which was discussed in section 3.2.1.

```
model {

#### MVH
for (n in 1:number_of_rows)#for the MVH dataset
{
  disutility[n] ~ dnorm (mu[n], tau)
  mu[n]<- fixed[n]+random[n]
fixed[n] <- beta[1]*MVH_indicators[1,n]+beta[2]*MVH_indicators[2,n]
+beta[3]*MVH_indicators[3,n]+beta[4]*MVH_indicators[4,n]+beta[5]*MVH_indicators[5,n]
+beta[6]*MVH_indicators[6,n]+beta[7]*MVH_indicators[7,n]+beta[8]*MVH_indicators[8,n]
+beta[9]*MVH_indicators[9,n]+beta[10]*MVH_indicators[10,n]+beta[11]*MVH_indicators[11,n]
+beta[12]*MVH_indicators[12,n]
random[n]<-RE[new_respondent_index[n]]+delta.term[statenum[n]]
}

###### priors
for (j in 1:unique_respondents) {
        RE[j] ~ dnorm(0,upsilon)
    }
for (w in 1:12) {
beta[w] ~ dnorm(0,prec.beta[w])
}
tau <-1/pow(sigma,2)
sigma~dunif(0,1)
upsilon <-1/pow(phi,2)
phi~dunif(0,1)


delta.term[1]<-0
for (i in 2:243) {
delta.term[i] ~ dnorm(0,tausqd)
}
tausqd <-1/pow(sigmad,2)
sigmad ~ dunif(0,1)

###new_delta.term
new_delta.term[1]<-0
for (i in 2:243) {
new_delta.term[i]~dnorm(0,tausqd)
}


}#End of model
```

JAGS code for the second part model, which was discussed in section 4.2.2, when we use the main Bayesian model.

```
data {
  for (i in 1:total_number_of_subjects) {
     for (j in 1:nTime) {   # nTime=number of time-points at which measurements are taken
     u[i,j] <- min( (1-((beta.specific %*% all.trial.indicators[j,,i])
     +delta.mis.specific[statenum.trial_matrix[i,j]]) ) , 1)
   # ensuring that the utilities are not greater than 1
########thus, we used bounded (by 1) utility values
}
     for (j in 1:(nTime-1)) {
        tmp[i,j] <- (u[i,(j+1)]+u[i,j])*(delta[j]/2)
 # here delta[j] is the amount of time between consecutive measurements

     }
     auc[i] <- sum(tmp[i,])
# here computes the "area under the curve" for each individual in the trial
  }
}

model {
for (i in 1:total_number_of_subjects){

### differential QALYs
auc[i]~dnorm(mu.auc[i],prec.auc)
mu.auc[i]<-gamma[1]+gamma[2]*u[i,1]+gamma[3]*delta_group[i]
}

###### costs
for (i in 1:number_of_control_subjects){
cost_FREQUENTIST_control[i]~dnorm(cost_c.AND.i_mean_prior[1],cost_c.AND.i_precision_prior[1])
}
for (i in 1:number_of_intervention_subjects){
cost_FREQUENTIST_intervention[i]~dnorm(cost_c.AND.i_mean_prior[2],
cost_c.AND.i_precision_prior[2])
}

###### priors
prec.auc<-1/pow(sigma.auc,2)
sigma.auc~dunif(0,100)
for (i in 1:3){
gamma[i]~dnorm(0,1)
}

for (a in 1:2){
cost_c.AND.i_mean_prior[a]~dunif(1,3000)
cost_c.AND.i_precision_prior[a]<-1/pow(cost_c.AND.i_sd_prior[a],2)
cost_c.AND.i_sd_prior[a]~dunif(1,3000)
}
}
```

JAGS code for the second part model, which was discussed in section 4.2.2, when we use the one-component approximation model.

```
data {
            dist.ind[1]<-1
          for (z in 2:number_of_all.states){
          dist.ind[z]~dnorm(approximate.state.means[z],approximate.state.precisions[z])
#the (approximate) distribution of a particular EQ-5D state
            }
  for (i in 1:total_number_of_subjects) {
      for (j in 1:nTime) {   # nTime=number of time-points at which measurements are taken
#######u[i,j] <- dist.ind[statenum.trial_matrix[i,j]]
        u[i,j] <- min( dist.ind[statenum.trial_matrix[i,j]] , 1 )
        # ensuring that the utilities are not greater than 1
########thus, we used bounded (by 1) utility values
}
      for (j in 1:(nTime-1)) {
         tmp[i,j] <- (u[i,(j+1)]+u[i,j])*(delta[j]/2)
         # here delta[j] is the amount of time between consecutive measurements
      }
      auc[i] <- sum(tmp[i,])
      # here computes the "area under the curve" for each individual in the trial
  }
}

model {
for (i in 1:total_number_of_subjects){
### differential QALYs
auc[i]~dnorm(mu.auc[i],prec.auc)
mu.auc[i]<-gamma[1]+gamma[2]*u[i,1]+gamma[3]*delta_group[i]
}
###### costs
for (i in 1:number_of_control_subjects){
cost_FREQUENTIST_control[i]~dnorm(cost_c.AND.i_mean_prior[1],cost_c.AND.i_precision_prior[1])
}
for (i in 1:number_of_intervention_subjects){
cost_FREQUENTIST_intervention[i]~dnorm(cost_c.AND.i_mean_prior[2],
cost_c.AND.i_precision_prior[2])
}
###### priors
prec.auc<-1/pow(sigma.auc,2)
sigma.auc~dunif(0,100)
for (i in 1:3){
gamma[i]~dnorm(0,1)
}
for (a in 1:2){
cost_c.AND.i_mean_prior[a]~dunif(1,3000)
cost_c.AND.i_precision_prior[a]<-1/pow(cost_c.AND.i_sd_prior[a],2)
cost_c.AND.i_sd_prior[a]~dunif(1,3000)
}
}
```

JAGS code for the second part model, which was discussed in section 4.2.2, when we use the three-component approximation model.

```
data {
            util[1]<-1
          for (z in 2:number_of_all.states){
          for (w in 1:3){
          util.test[z,w]~dnorm(mixture.means_3[z,w],mixture.precisions_3[z,w])
               }
             sampled.n[z]~dcat(mix.weights_3[z,])
             util[z]<-util.test[z,sampled.n[z]]
            }
  for (i in 1:total_number_of_subjects) {
     for (j in 1:nTime) {   # nTime=number of time-points at which measurements are taken
###########u[i,j] <- util[statenum.trial_matrix[i,j]]
            u[i,j] <- min( util[statenum.trial_matrix[i,j]] , 1 )
            # ensuring that the utilities are not greater than 1
########thus, we used bounded (by 1) utility values
}
     for (j in 1:(nTime-1)) {
        tmp[i,j] <- (u[i,(j+1)]+u[i,j])*(delta[j]/2)
        # here delta[j] is the amount of time between consecutive measurements
     }
     auc[i] <- sum(tmp[i,])
     # here computes the "area under the curve" for each individual in the trial
  }
}
```

```
model {

for (i in 1:total_number_of_subjects){
### differential QALYs
auc[i]~dnorm(mu.auc[i],prec.auc)
mu.auc[i]<-gamma[1]+gamma[2]*u[i,1]+gamma[3]*delta_group[i]
}

###### costs
for (i in 1:number_of_control_subjects){
cost_FREQUENTIST_control[i]~dnorm(cost_c.AND.i_mean_prior[1],cost_c.AND.i_precision_prior[1])
}

for (i in 1:number_of_intervention_subjects){
cost_FREQUENTIST_intervention[i]~dnorm(cost_c.AND.i_mean_prior[2],
cost_c.AND.i_precision_prior[2])
}
```

```
###### priors
prec.auc<-1/pow(sigma.auc,2)
sigma.auc~dunif(0,100)
for (i in 1:3){
gamma[i]~dnorm(0,1)
}

for (a in 1:2){
cost_c.AND.i_mean_prior[a]~dunif(1,3000)
cost_c.AND.i_precision_prior[a]<-1/pow(cost_c.AND.i_sd_prior[a],2)
cost_c.AND.i_sd_prior[a]~dunif(1,3000)
}
}
```

JAGS code for the second part model, which was discussed in section 4.2.2, when we use the five-component approximation model.

```
data {
           util[1]<-1
         for (z in 2:number_of_all.states){
         for (w in 1:5){
         util.test[z,w]~dnorm(mixture.means_5[z,w],mixture.precisions_5[z,w])
              }
            sampled.n[z]~dcat(mix.weights_5[z,])
            util[z]<-util.test[z,sampled.n[z]]
           }

  for (i in 1:total_number_of_subjects) {
     for (j in 1:nTime) {   # nTime=number of time-points at which measurements are taken
###########u[i,j] <- util[statenum.trial_matrix[i,j]]
           u[i,j] <-  min( util[statenum.trial_matrix[i,j]] , 1 )
           # ensuring that the utilities are not greater than 1
########thus, we used bounded (by 1) utility values
}
     for (j in 1:(nTime-1)) {
        tmp[i,j] <- (u[i,(j+1)]+u[i,j])*(delta[j]/2)
        # here delta[j] is the amount of time between consecutive measurements

     }
     auc[i] <- sum(tmp[i,])
     # here computes the "area under the curve" for each individual in the trial
  }


}


model {

for (i in 1:total_number_of_subjects){

### differential QALYs
auc[i]~dnorm(mu.auc[i],prec.auc)
mu.auc[i]<-gamma[1]+gamma[2]*u[i,1]+gamma[3]*delta_group[i]
}

###### costs
for (i in 1:number_of_control_subjects){
cost_FREQUENTIST_control[i]~dnorm(cost_c.AND.i_mean_prior[1],cost_c.AND.i_precision_prior[1])
}
for (i in 1:number_of_intervention_subjects){
cost_FREQUENTIST_intervention[i]~dnorm(cost_c.AND.i_mean_prior[2],
cost_c.AND.i_precision_prior[2])
}
```

```
###### priors
prec.auc<-1/pow(sigma.auc,2)
sigma.auc~dunif(0,100)
for (i in 1:3){
gamma[i]~dnorm(0,1)
}

for (a in 1:2){
cost_c.AND.i_mean_prior[a]~dunif(1,3000)
cost_c.AND.i_precision_prior[a]<-1/pow(cost_c.AND.i_sd_prior[a],2)
cost_c.AND.i_sd_prior[a]~dunif(1,3000)
}

}
```

Next, we present the costs and the EQ-5D-3L states of the principal simulated dataset which was discussed in section 4.3.2, so that the aforementioned dataset can be replicated. Specifically, first we present the costs for the control group and then the costs for the intervention group. Then, we present the EQ-5D-3L states experienced at the three time-points of interest (i.e. at baseline, six months after baseline and one year after baseline) for the control group. This is followed by the presentation of the EQ-5D-3L states experienced at the three time-points of interest (i.e. at baseline, six months after baseline and one year after baseline) for the intervention group.

Control group costs (in GBP):

428 859 214 791 617 168 839 684 316 357 450 856 418 966 865 173 593 736 185 684 263 905 390 172 649 777 552 925 191 223 668 150 662 1,067 610 801 121 884 685 418 1,097 632 210 734 898 813 803 663 359 607 827 783 138 619 732 786 813 649 989 584 517 124 858 1,058 340 1,004 421 1,026 139 885 676 813 458 521 955 245 330 296 189 263 403 277 1,021 164 674 233 583 374 515 537 1,029 925 691 112 643 730 185 854 267 506 700 552 237 827 755 947 260 801 641 401 930 819 137 696 723 189 1,095 336 295 122 807 401 448 742 992 1,095 985 823 411 196 485 636 384 148 1,048 411 270 939 617 251 168 609 846 701 713 490 1,018 736 537 949 423 662 667 721 1,019 169 867 685 732 972 979 362 698 962 461 1,036 1,017 1,013 760 399 552 865 574 995 111 641 248 1,065 1,081 1,018 1,076 535 128 454 779 320 537 639 559 881 781 502 772 652 897 957 450 282 1,079 853 186 432 665 599 555 445 906 653 652 809 677 734 752 655 113 540 948 236 986 919 874 478 487 460 511 691 549 942 1,066 715 411 312 714 708 168 323 1,048 475 694 930 816 1,044 474 110 777 994 958 830 728 886 351 901 303 592 192 641 689 426 350 847 405 304 887 529 951 120 672 913 598 670 404 236 969 447 169 648 183 337 639 761 702 899 789 970 309 628 916 430 219 772 462 456 990 290 1,056 254 614 1,039 862 202

Intervention group costs (in GBP):

1,988 2,285 2,667 1,860 2,395 2,642 2,801 2,682 2,338 2,522 1,838 2,607 1,975 2,144 2,484 1,897
2,408 2,110 1,955 1,810 2,432 2,762 2,705 1,881 2,875 2,880 2,222 2,473 2,167 2,850 2,218 2,011
2,525 2,780 2,547 2,066 2,093 2,703 2,352 2,504 2,506 2,678 2,044 2,483 2,273 1,840 2,028 1,815
1,907 2,530 1,927 2,058 2,869 2,593 2,779 2,665 2,451 2,750 2,269 2,536 2,850 2,120 2,733 2,557
2,826 2,079 2,754 2,661 2,562 2,356 2,372 2,617 2,132 2,314 2,675 1,941 2,752 2,561 2,183 2,870
2,195 2,448 1,935 2,233 2,063 2,014 2,501 1,926 2,532 1,994 2,847 2,860 2,592 2,835 2,413 2,685
2,498 2,819 1,994 2,449 2,703 2,031 2,645 2,539 1,851 2,629 2,249 2,210 2,081 2,230 2,317 1,895
2,419 1,979 2,309 2,650 2,715 2,732 2,090 2,110 2,273 2,605 2,668 1,851 2,290 1,814 1,822 2,596
2,061 2,710 1,857 2,134 2,331 2,458 2,138 2,487 2,177 2,857 2,153 1,901 1,871 2,870 2,035 2,674
1,854 1,820 2,628 2,614 2,806 2,658 2,356 1,981 1,873 2,192 2,565 1,850 2,681 2,496 2,130 1,849
2,015 2,557 2,024 2,156 2,876 2,657 1,829 2,469 2,097 2,528 2,496 2,745 2,519 2,144 1,998 2,107
2,881 2,465 1,994 2,681 2,033 1,895 2,643 2,650 2,335 1,846 2,638 2,876 2,329 1,872 2,399 2,327
2,555 2,010 2,549 1,801 2,399 1,885 1,957 2,059 2,597 2,854 2,300 2,577 2,592 2,156 2,702 1,906
2,196 2,701 2,600 2,659 2,094 1,911 2,452 2,849 1,839 2,658 2,740 2,237 2,085 2,890 1,900 2,188
2,197 2,510 2,880 1,853 2,868 2,382 2,408 2,609 2,421 2,556 2,648 2,627 2,850 2,706 2,476 2,870
1,891 2,864 2,642 2,572 2,686 1,809 2,155 2,002 2,256 2,274 2,065 1,988 2,249 1,992 2,167 1,912
2,108 2,098 2,712 2,659 2,543 2,555 2,741 2,465 2,024 1,999 2,266 2,448 2,753 2,285 2,859 2,165
2,849 2,261 1,922 1,837 2,878 2,663 2,227 1,851 1,961 1,817 2,139 2,767 2,139 2,889 1,842 2,243
1,878 2,624 2,269 1,948 2,218 2,224 2,379 2,895 2,211 2,511 2,149 2,678

Control group EQ-5D-3L states at baseline:

21212 12113 11122 11222 23112 22312 11222 22312 12212 11122 11122
22312 11122 12121 11313 22312 22211 22211 23211 22111 12212 13121
11222 22211 22312 22211 11313 22211 22111 23112 23211 12121 11122
23112 23211 12113 23112 11222 11222 12113 11122 22111 22312 11122
11122 22111 23211 23112 11222 13121 12212 22312 12212 12121 12121
11222 22312 23112 23112 12113 23112 12113 23211 23211 22211 22211
12113 23112 22211 23211 11122 11313 11122 13121 12113 13121 12113
22211 22211 12121 12113 11313 11313 22111 11313 12121 12212 22111
22111 12121 11222 11313 12212 12121 11222 13121 11122 23112 21212
21212 11313 11122 22111 11313 23211 13121 11222 11313 22211 11222
22312 13121 11222 12121 23112 11313 11122 23211 11313 13121 21212
11313 11122 23211 23211 11313 12113 11222 22211 23112 22211 23112
11222 22312 11222 12212 23112 12121 11222 22312 11222 11313 11122
13121 22111 11222 12113 11313 11122 11313 12212 21212 12212 22312
23112 22312 12121 12113 23112 12121 13121 13121 23211 11313 21212
23112 22111 22111 22211 12113 22111 22111 12212 12212 12212 21212
13121 12121 22111 11122 11313 23211 12121 23112 23112 23112 11222
23112 22111 11122 12113 23112 21212 11313 22312 22312 23112 11222
21212 22111 21212 22111 11222 12212 11222 12212 23211 12121 13121
11122 22111 13121 11122 23211 21212 11313 21212 21212 21212 11222
22211 13121 12113 11222 22111 12121 22312 12121 22312 11122 23211
13121 13121 11313 21212 12212 23112 12121 12121 12212 12121 12121
11222 13121 22111 23211 22211 23112 11313 23112 22211 11222 22111
12212 22211 11222 12113 13121 23112 11222 11122 23112 11122 13121
23211 11122 22211 23112 23211 12121 11222 21212 12121 12121 13121
11313 22312 11313 11122 21212 21212 12212 12212 23112 21212 22211
22211 13121 11222 13121 12212 12113 11122 22111 22111 12113 22312
21212 11122 11313

Control group EQ-5D-3L states six months after baseline:

22122 11113 21213 11321 23112 22312 12111 12322 22221 11122 11122
22312 11122 12121 11313 22312 22222 22211 23211 12321 31211 12112
11222 22211 21121 22211 23221 22211 22111 11321 23211 12121 11122
23112 23211 32321 23112 11222 22312 22221 11122 22111 22312 11122
11113 22111 21232 23112 13111 13121 22221 22312 12212 12121 12121
11321 22312 23112 11133 33211 23112 12113 23211 23211 22211 11121
12113 23312 22211 23211 11122 11313 11122 13121 12113 13121 12113
22211 22211 12121 22113 11313 11313 13212 23322 12121 12212 22111
22111 12121 11222 11313 12212 12121 11222 23322 11122 23112 21212
21212 11313 11122 11211 11313 23211 13121 11222 21232 21222 11222
22312 13121 11222 21311 23112 11131 11122 12112 11313 13121 22312
23212 21121 23211 23211 23322 12113 11222 31211 23112 22211 23112
11222 21121 11222 12212 23112 21311 11321 11331 11222 21121 11122
13131 22112 13111 12113 11313 11122 11313 12113 21212 11112 21121
23112 22312 12121 12113 23312 12121 22321 21313 23211 11313 21212
23312 22111 21322 22211 12113 22111 21111 12212 12212 21111 21212
12112 12212 22111 11122 11313 13221 12121 23112 23112 11323 11222
23322 22111 11122 12113 23112 21212 11313 22312 21121 23112 11222
21212 22111 22312 12321 11222 11112 11222 12212 23211 12121 13121
21213 21122 13121 11122 12112 21212 11313 11211 21212 11211 11222
22211 13121 12113 11222 21111 11112 22312 12121 21212 11122 21322
13121 13121 11313 21212 11123 23112 12121 23212 12113 12121 12121
11222 12112 22111 23211 22211 23112 11313 23312 21213 12111 22111
12212 22211 22312 11122 13121 23112 11222 21211 23112 11122 13121
23211 11122 13212 32321 32221 12111 11222 22122 12121 12121 13121
11313 22312 11313 11122 21212 21212 12212 12212 23112 21212 22211
22211 13121 11222 13121 12212 12113 11122 22111 22111 12113 22312
21212 11122 11313

Control group EQ-5D-3L states one year after baseline:

22122 11121 21213 11321 11222 22312 12111 12322 22221 11122 11122
22312 21213 11312 11313 13222 31121 22222 23211 12321 31211 12112
11222 22211 21121 22211 23221 22211 22111 11321 23211 12121 12122
21211 21322 32321 23112 11222 22312 22221 11122 22111 11222 11122
11113 22111 21232 23112 13111 33211 22221 22312 12212 12121 12121
11321 22312 23112 11223 33211 23112 12113 23211 23211 22211 11121
21221 23312 22211 12212 11122 23221 11122 13121 12113 13121 12113
22211 22222 12121 22113 11313 12112 13212 33313 12121 22221 22111
22111 12121 11222 13322 21111 12121 11222 23322 21211 23112 21212
22122 11313 11122 11211 11313 23211 11211 11222 11313 21222 11222
22223 13121 11222 21311 23112 11131 11122 12112 21232 13121 21121
23212 21121 32231 21322 23322 11122 11222 22211 23322 22211 21221
31212 21121 11222 12212 11323 21311 11321 11331 11222 21121 21213
13131 22112 31111 12212 11313 11122 11313 12113 21212 12212 21121
23112 11223 11313 12113 23312 12121 22321 21332 23211 11313 22312
23312 22111 21322 22211 22113 22111 21111 12212 11112 21111 21212
22122 12113 22111 21213 11313 12331 23212 11222 23112 13231 11222
23322 11211 21111 12113 23112 21212 11313 22312 13211 23112 11222
22212 22111 22312 32122 11222 12112 11222 21112 23211 11312 13121
22213 23121 13121 11122 12112 21212 23212 11211 21212 11211 11222
31211 13121 12113 11222 12212 12212 22312 12121 21212 12113 21322
33211 31311 11313 11313 11123 23112 12121 11312 12113 12121 12121
11222 12112 13212 23211 21213 12121 11313 23312 31212 12111 22111
12212 22211 22312 11122 13121 23112 21211 21211 23112 11122 13121
33211 21211 13212 32321 32221 12211 11222 22122 12121 12121 31311
11313 22312 23322 11122 21212 21212 12212 12212 23112 21212 22211
22211 13121 11222 13121 11112 22331 11122 22111 22111 12113 22312
21212 11122 11313

Intervention group EQ-5D-3L states at baseline:

11222 11122 21212 12121 22312 11222 11122 13121 21212 22312 23211
11122 23211 22312 13121 13121 11313 11313 21212 12121 12113 12113
23112 22111 12113 12113 12121 11222 11222 12212 11122 11313 22111
22111 23112 23211 13121 22211 23112 11313 22211 22312 22312 22312
11313 12121 23211 21212 11313 12212 22312 22111 11222 12212 11313
12212 12113 22312 22111 13121 11122 22312 23112 11122 11313 23211
22312 21212 11122 12113 12113 21212 13121 13121 23112 23211 12113
22312 13121 12212 22312 22211 22312 12212 23211 11222 13121 22111
11122 11222 11122 22111 22111 13121 23211 12121 12113 23211 23211
23211 11122 12212 11122 12113 11222 12113 12113 12121 22312 22111
12113 22312 22211 12121 23211 22211 21212 12212 12212 22312 11122
11313 22312 11222 12212 12121 12113 11222 23211 23112 11122 22211
13121 12113 21212 12113 11222 12113 11313 22111 11222 23112 12121
11122 12113 11222 22312 22211 12121 22211 12113 23211 22312 11222
11313 12212 13121 21212 11122 11122 12113 23112 13121 12113 22211
12121 13121 11122 13121 11313 23112 23112 11222 11122 22111 12113
23211 11122 11222 22312 22111 11313 21212 23211 12121 12212 21212
21212 12113 11222 12113 12113 11122 21212 12113 11222 21212 23211
22312 22312 22111 22312 22211 11122 13121 12121 12121 22312 21212
22211 21212 22312 11122 11222 12121 12121 12212 22211 12113 22211
11122 23112 22211 13121 11313 11313 12212 12212 21212 22211 22111
12212 11313 12121 12113 22111 13121 22111 23112 22111 13121 11222
11122 22111 22211 13121 12212 23211 22211 21212 22312 11122 12121
12212 13121 12113 22211 23211 11122 12212 12212 13121 22312 11222
11222 13121 22111 12113 22312 21212 22211 22211 12121 13121 12121
23211 22312 22211 11222 23112 23211 11313 13121 11222 12212 23112
12113 11313 11313 11122 22312 11222 21212 12212 12121 21212 12212
11122 11313 12121

Intervention group EQ-5D-3L states six months after baseline:

12211 21111 11211 21212 21121 11111 11113 11112 11212 22312 11112
11122 13221 11222 22321 12321 12112 11111 11211 12121 11111 31322
31211 22111 11122 12212 11112 11222 11222 12212 11113 12121 22111
11211 11222 21111 13121 12212 21211 11313 11222 21212 13111 22312
12121 12111 21111 12111 11313 12212 11211 22111 11222 23211 12121
11111 11122 21121 21111 21211 21111 11222 11111 21211 11313 23211
11211 21212 11122 12113 21221 22212 21221 21121 23112 11111 22221
21212 11112 12212 11211 12111 21121 12212 12113 21211 11112 11211
21111 11222 11122 22111 13212 11112 13212 12121 12122 23211 22111
23211 21211 12111 11111 21111 11321 21221 21221 12121 11211 22111
21221 13111 21111 12111 22111 22211 11222 12212 11112 11222 11122
11211 13222 12111 12212 11111 11111 21211 12112 13111 21111 31112
13121 12212 12111 21312 11222 22221 11313 21111 11222 23112 11112
21111 12212 21211 21121 12111 11112 12111 21221 21122 21212 21211
23221 11111 21313 21212 21111 21211 11112 21312 13121 11122 21111
12121 13121 21111 21211 12112 11112 21212 21211 11113 21111 13121
21111 11113 11321 12111 22111 21212 22122 11211 12111 12212 12111
21212 21221 12111 11122 12113 11122 21212 12122 12111 11111 12221
22312 21211 22111 21212 12111 11111 21211 11112 11111 22312 21212
11222 22111 11222 21111 11222 11111 12111 21111 22211 21111 12111
11122 23112 12111 21221 23211 11211 12212 12212 12112 12211 21111
12212 11112 12121 31211 11211 13121 22112 11111 22111 11211 12111
11122 11111 12211 21121 11112 21112 21111 21212 22212 21111 21311
21111 13121 11112 11111 12222 21211 12212 12222 11111 22212 12113
21211 21121 12113 11122 12322 11111 12111 21111 12211 21121 12121
11111 11222 12111 11222 21221 22211 12121 11112 12111 11112 13113
12222 11313 21212 11121 22312 21211 21212 21221 12111 11111 12212
21211 11112 12121

Intervention group EQ-5D-3L states one year after baseline:

12111 21111 11211 11211 21121 12112 21111 11111 11212 12111 11112
11122 11322 12111 13121 11111 11211 11111 11111 12111 21111 31221
21111 11211 12311 12111 11111 11222 11222 12222 11113 12121 22111
11211 12111 11111 11112 22111 21211 21212 11222 12111 11222 11211
11111 12111 12212 12121 11112 21111 11211 11211 21211 11112 11112
11111 11113 21121 11111 21111 12212 21212 11211 21221 11311 22111
11211 12111 21211 12113 21211 12121 13121 21121 23112 11111 12112
11211 21112 12212 11211 12111 11313 12222 12113 11111 11112 11211
21111 23112 21211 21111 21222 11111 22111 21112 11221 11211 21232
23211 21111 11111 12111 21111 11133 21221 12111 12121 11211 22111
21211 21211 11111 12111 21212 12111 12111 11121 11112 11222 21211
21212 22312 21111 21111 11111 11111 11111 11311 11221 21111 21122
11111 11112 12211 21221 21211 22221 11313 21111 23112 11111 11112
21111 12212 21211 21121 11111 21112 22211 12113 31112 21212 11111
23221 11111 12221 21212 11121 11111 12212 21112 11111 11222 11111
12121 13121 11111 21211 12221 11112 21212 21211 11122 11111 21211
21111 21111 12121 12111 22112 11111 21212 12111 12111 21111 11111
11111 11112 11211 21211 21221 11111 11212 11122 12111 11111 12221
12322 11111 11211 11313 11111 21211 21211 11111 11111 11211 21212
11111 22111 11221 21111 11222 11111 12111 21111 12111 21111 21211
21111 11112 12111 21221 13221 11111 11112 22211 12112 12111 21111
11111 11111 11312 11112 11211 11112 11212 22211 21122 11211 11112
11122 11111 12211 21211 11111 21112 21111 11111 12211 21111 21311
21211 12112 11111 11111 11121 21211 11111 12212 11111 21212 12113
11111 11211 21221 11111 13111 11111 11111 11122 12111 11211 12121
11111 12111 12111 12111 13121 21211 11112 11111 11111 11112 11333
12222 12112 21212 11121 11222 11111 21212 11111 11111 11112 11112
21211 11111 12121

The principal simulated dataset which was discussed in section 3.2 was derived as follows: the costs of each group were drawn from a uniform distribution and the EQ-5D-3L states of each group at each time-point were obtained based on probabilities which represented how likely it was for each of the EQ-5D-3L dimensions of a subject's state to be at a specific level.
In particular, the distributions from which the costs were sampled are reported in Table 1. Regarding the health states of the dataset, some EQ-5D-3L states are randomly sampled at baseline, as shown in Table 2; all of them are sampled with equal probability. Then, the probabilities that the level of an EQ-5D-3L dimension improves, remains the same, or worsens from one time-point to another are reported in Table 3.

Trial group	**Distribution**
Control	$U(100, 1100)$
Intervention	$U(1800, 2900)$

Table 1: The distributions from which the costs of each group were sampled for the principal simulated dataset which was discussed in section 3.2.

EQ-5D-3L states at t=0				
12211	11221	22111	12112	21212
21121	11122	22211	12212	12121
21221	11222	22112		

Table 2: The EQ-5D-3L states which were randomly sampled to be the ones the trial subjects of both groups fall in at t=0 (i.e. at baseline) for the principal simulated dataset which was discussed in section 3.2. All of the states are sampled with equal probability.

Change of dimension level from t=0 to t=1		
Level of a dimension	Probability	
	Control group	Intervention group
Improves	0.05	0.35
Remains the same	0.9	0.6
Worsens	0.05	0.05
Change of dimension level from t=1 to t=2		
Level of a dimension	Probability	
	Control group	Intervention group
Improves	0.05	0.35
Remains the same	0.9	0.6
Worsens	0.05	0.05

Table 3: The probabilities (for each group) that the level of a dimension of a subject's state improves, remains the same, or worsens (the best level is level 1 and the worst level is level 3), from t=0 (i.e. at baseline) to t=1 (i.e. at 6 months after baseline) and from t=1 (i.e. at 6 months after baseline) to t=2 (i.e. at 1 year after baseline) for the principal simulated dataset which was discussed in section 3.2.

B Tables and figures

At the beginning of this section, there are three figures which illustrate the 95% credible intervals of the MCMC samples of the posterior utility distributions (first, second and last third of EQ-5D-3L states). Then, there is a figure with the empirical K-L divergence values for each state, and then a figure for the weights of each component for each state. Next, there are figures of Kernel density plots for the MCMC simulations (solid lines) of the EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line. Finally, there is the probability density function of each of the three components of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distribution.

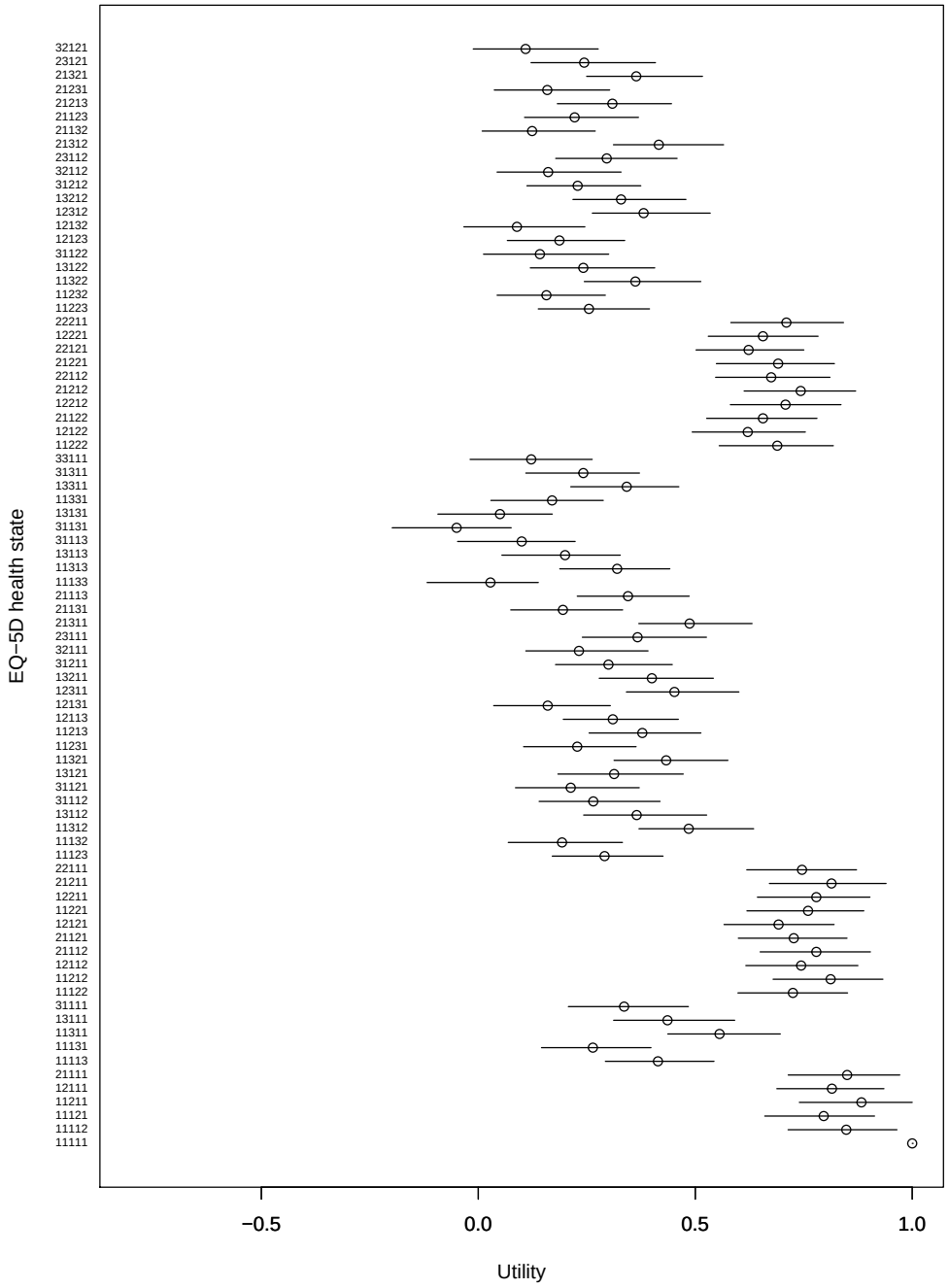


Figure 2: 95% credible intervals for the posterior distributions of 81 EQ-5D-3L health states (the first third of EQ-5D-3L states). The expected mean utilities of the states based on the original MVH report are shown as circles.

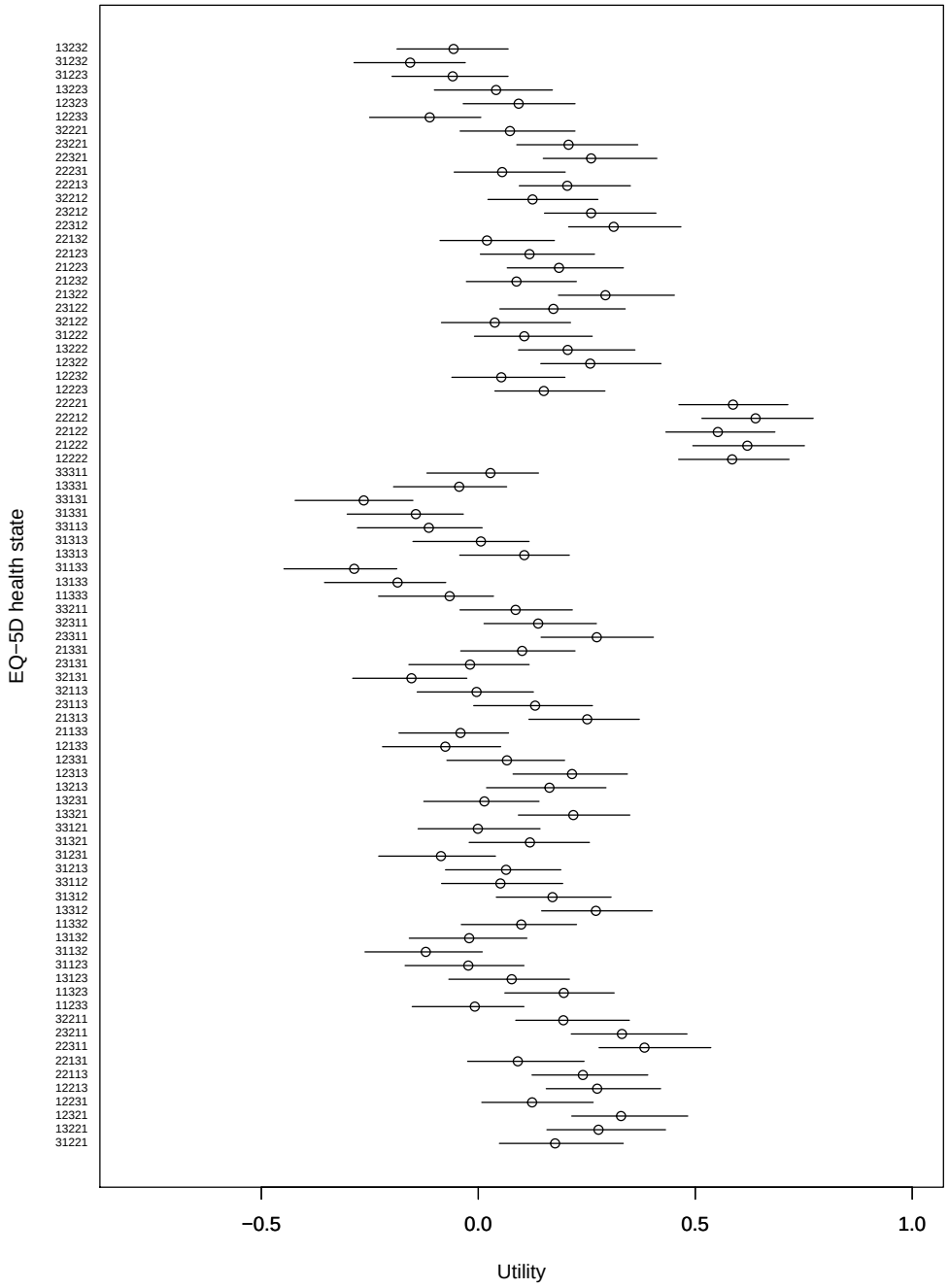


Figure 3: 95% credible intervals for the posterior distributions of 81 EQ-5D-3L health states (the second third of EQ-5D-3L states). The expected mean utilities of the states based on the original MVH report are shown as circles.

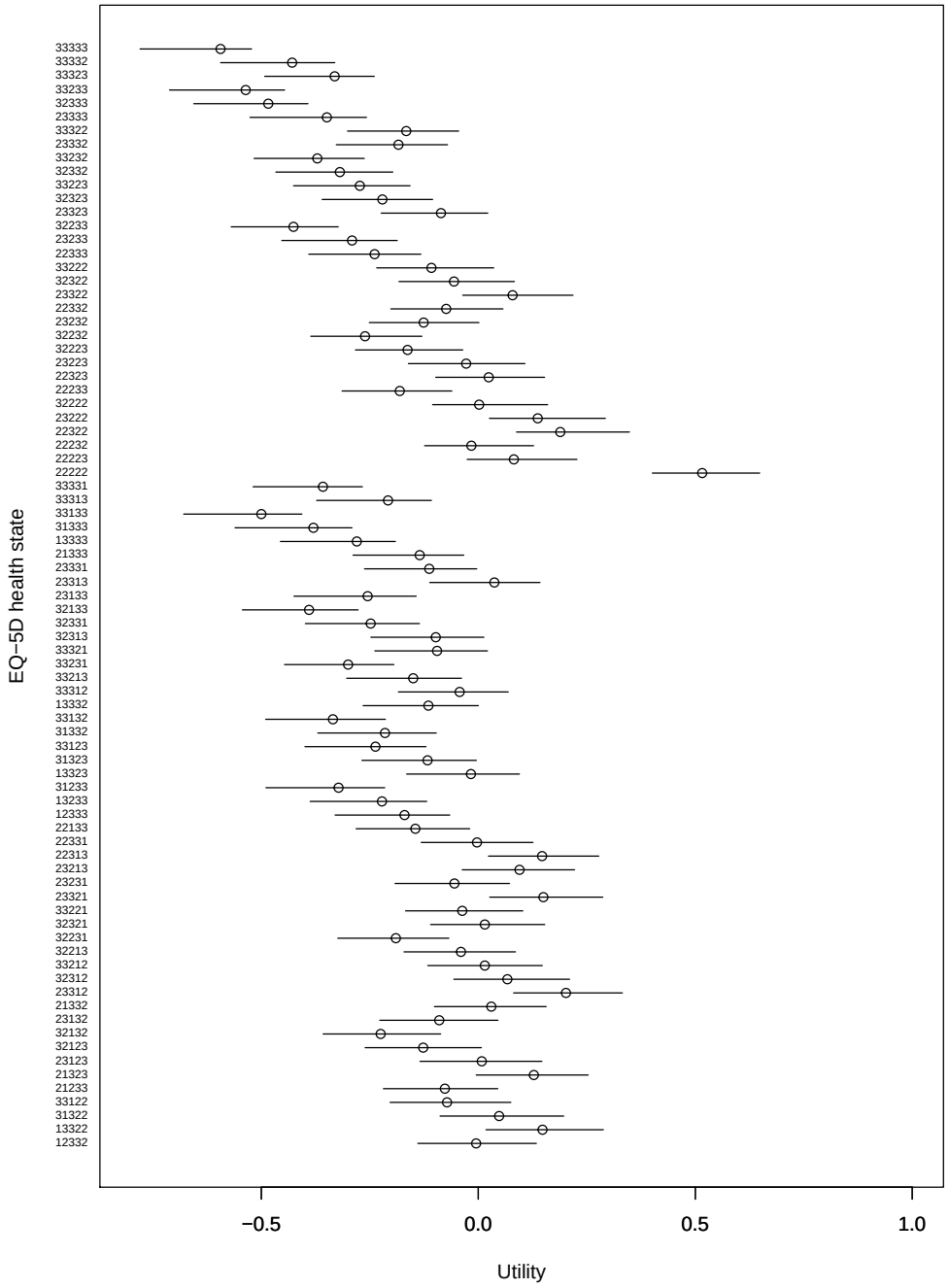


Figure 4: 95% credible intervals for the posterior distributions of 81 EQ-5D-3L health states (the last third of EQ-5D-3L states). The expected mean utilities of the states based on the original MVH report are shown as circles.

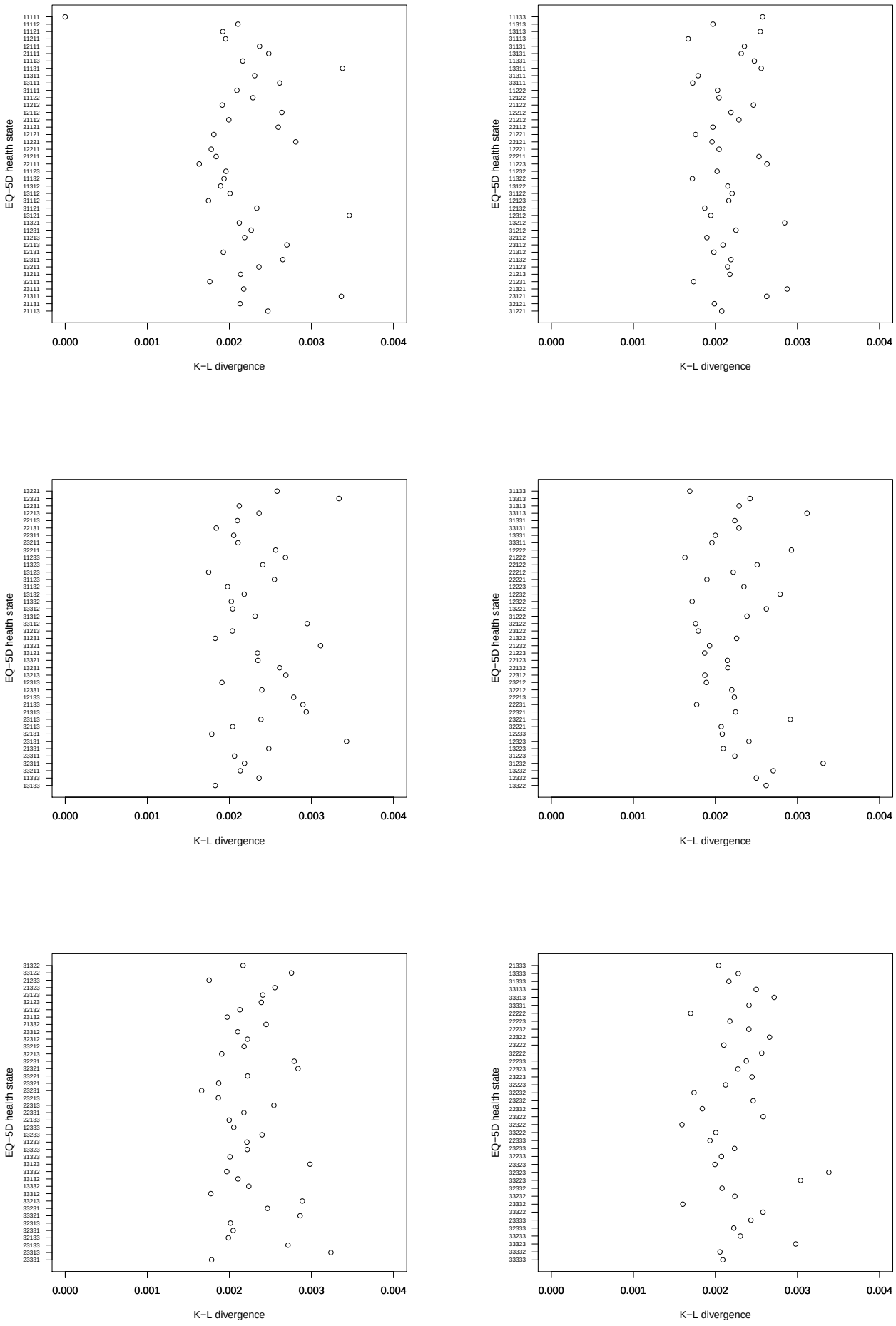


Figure 5: Empirical K-L divergence values per state for the three-component distributions.

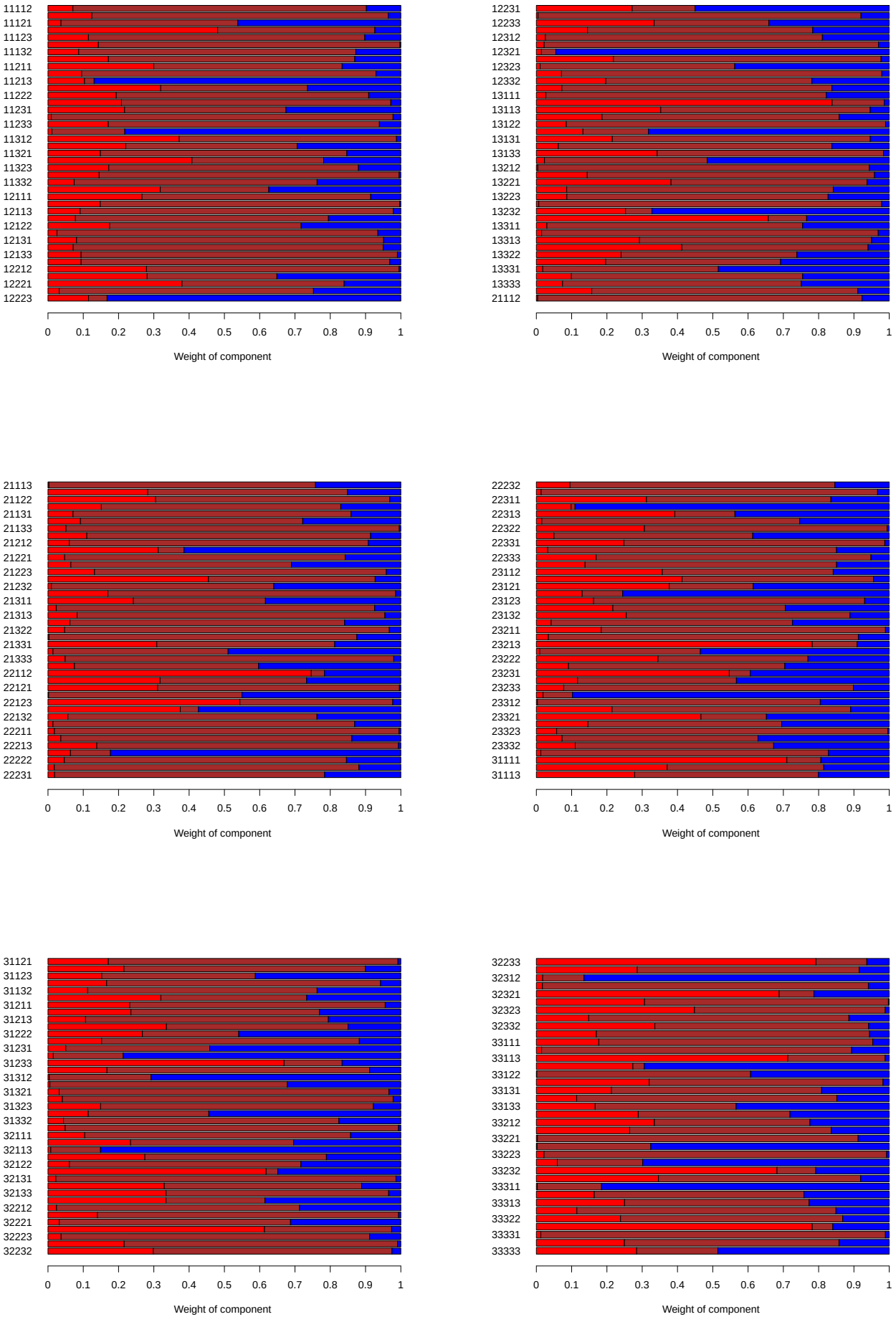


Figure 6: The weight of the first (red), second (brown) and third (blue) component.

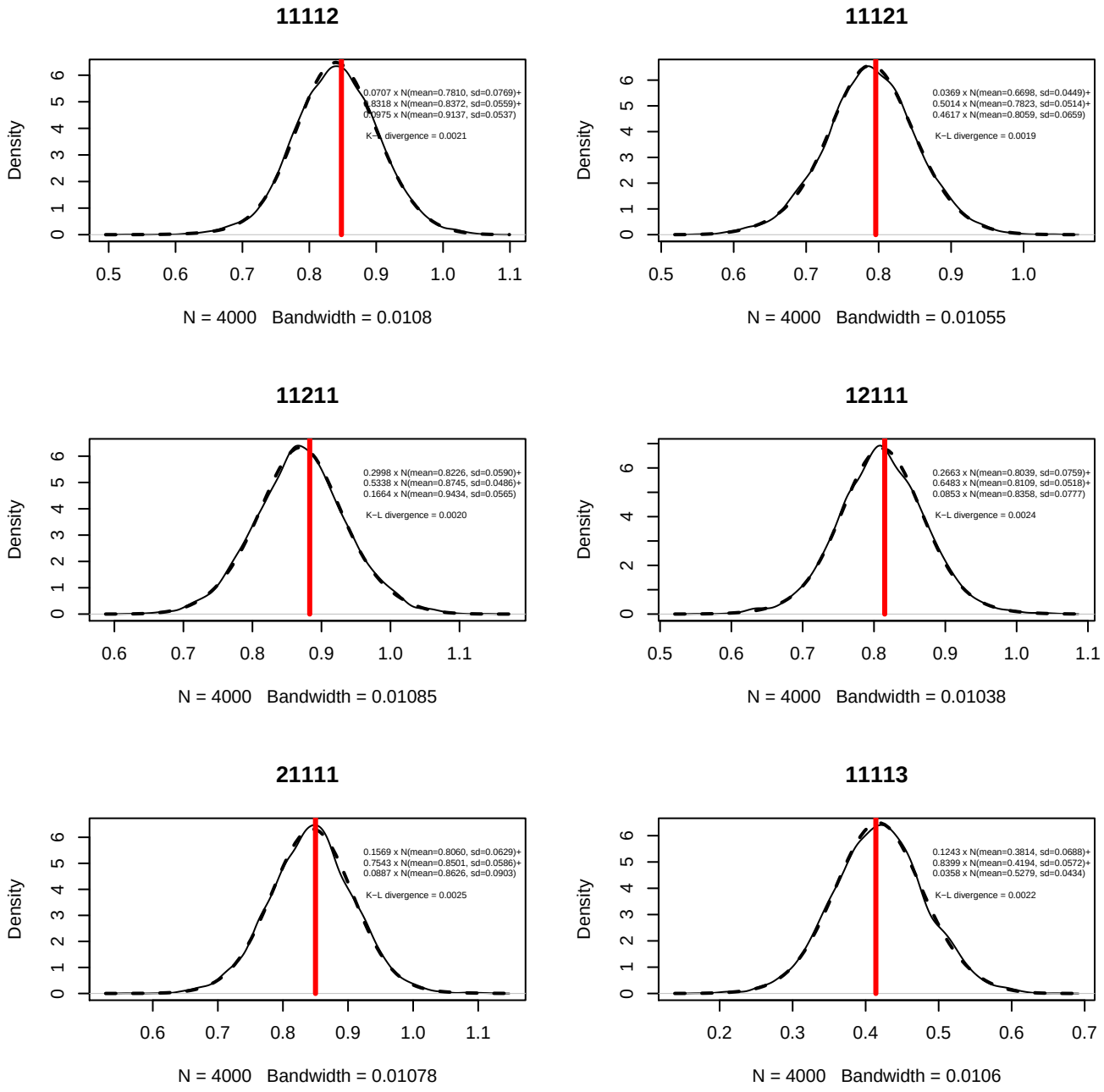


Figure 7: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

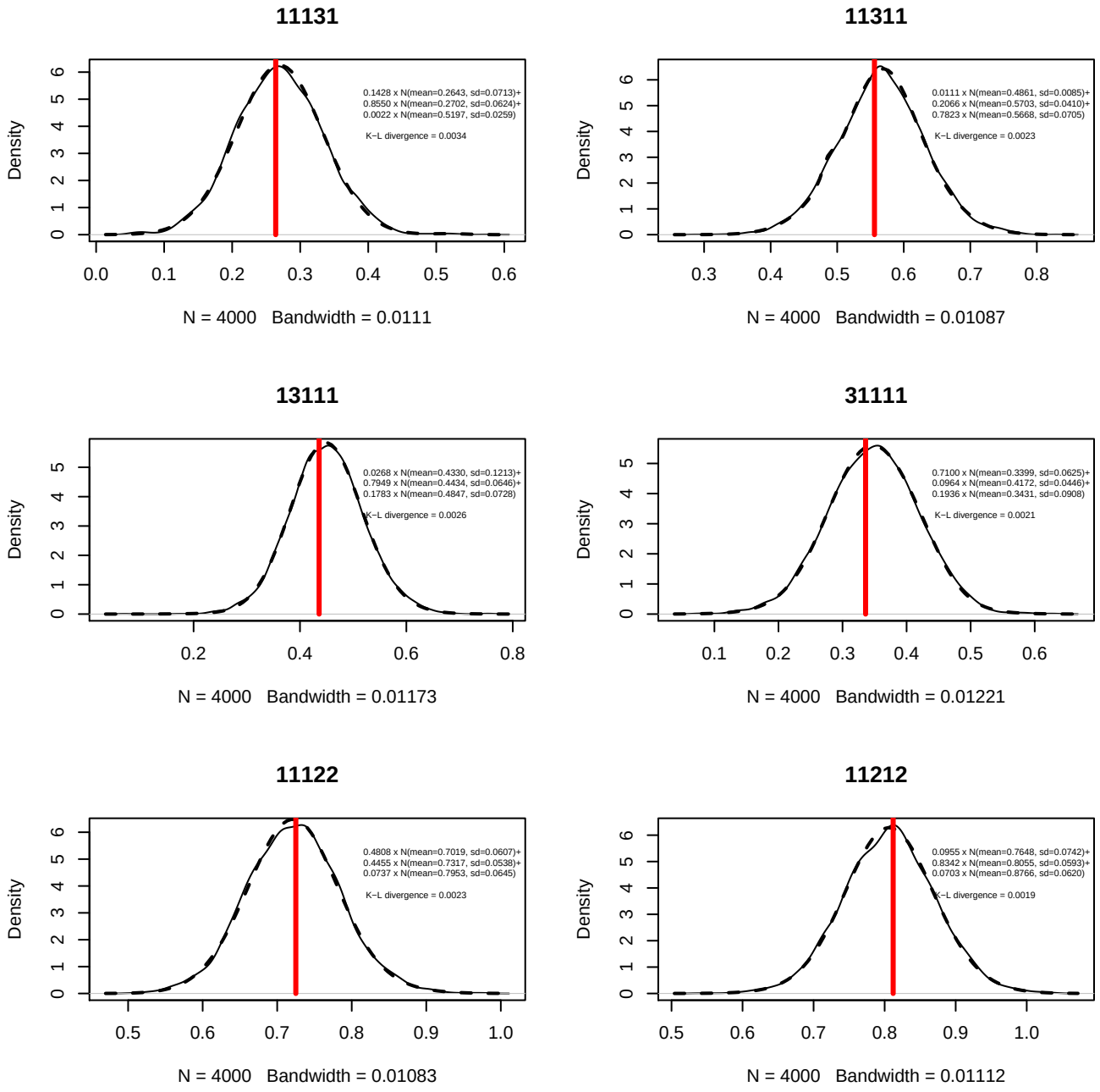


Figure 8: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

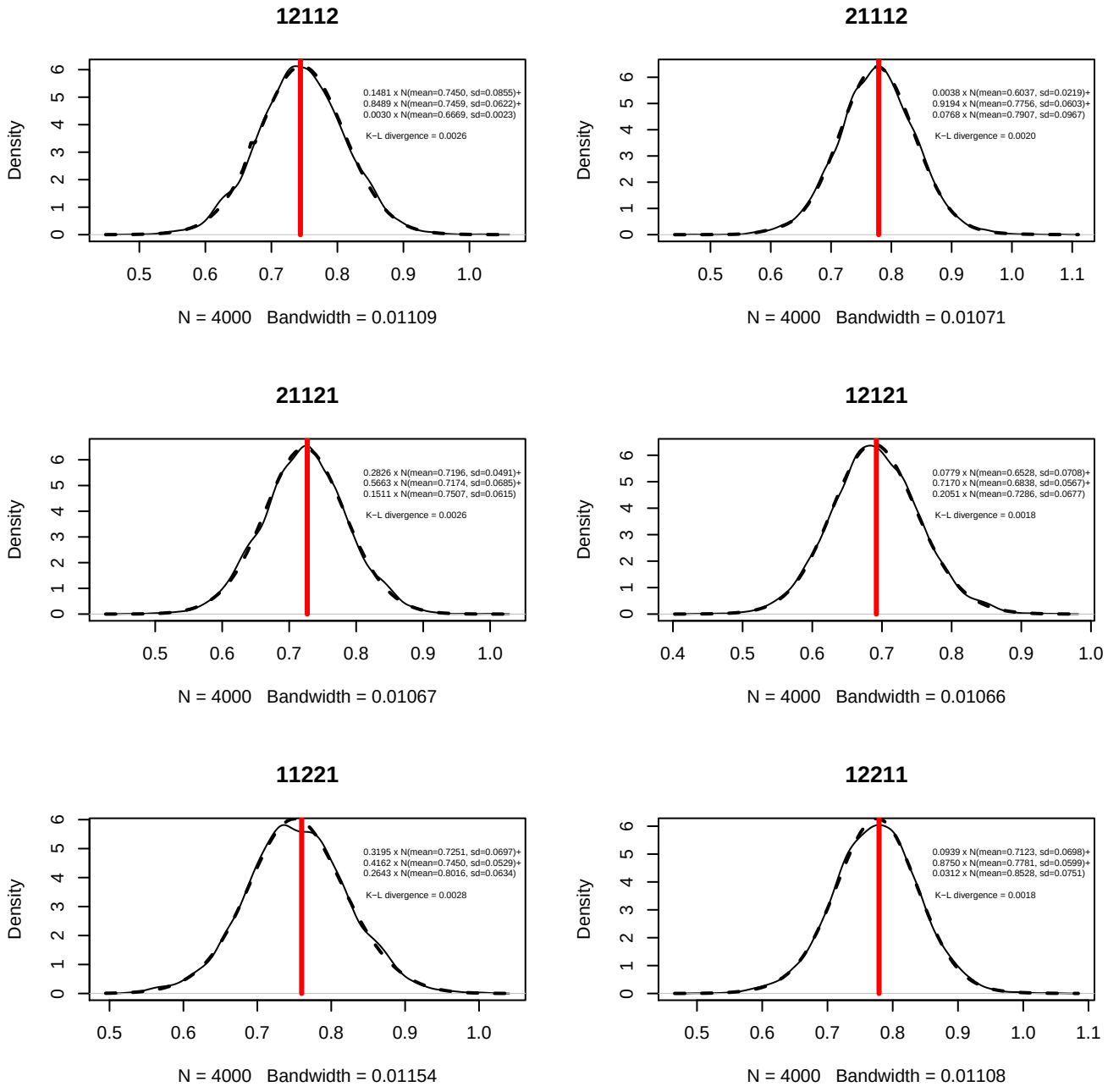


Figure 9: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

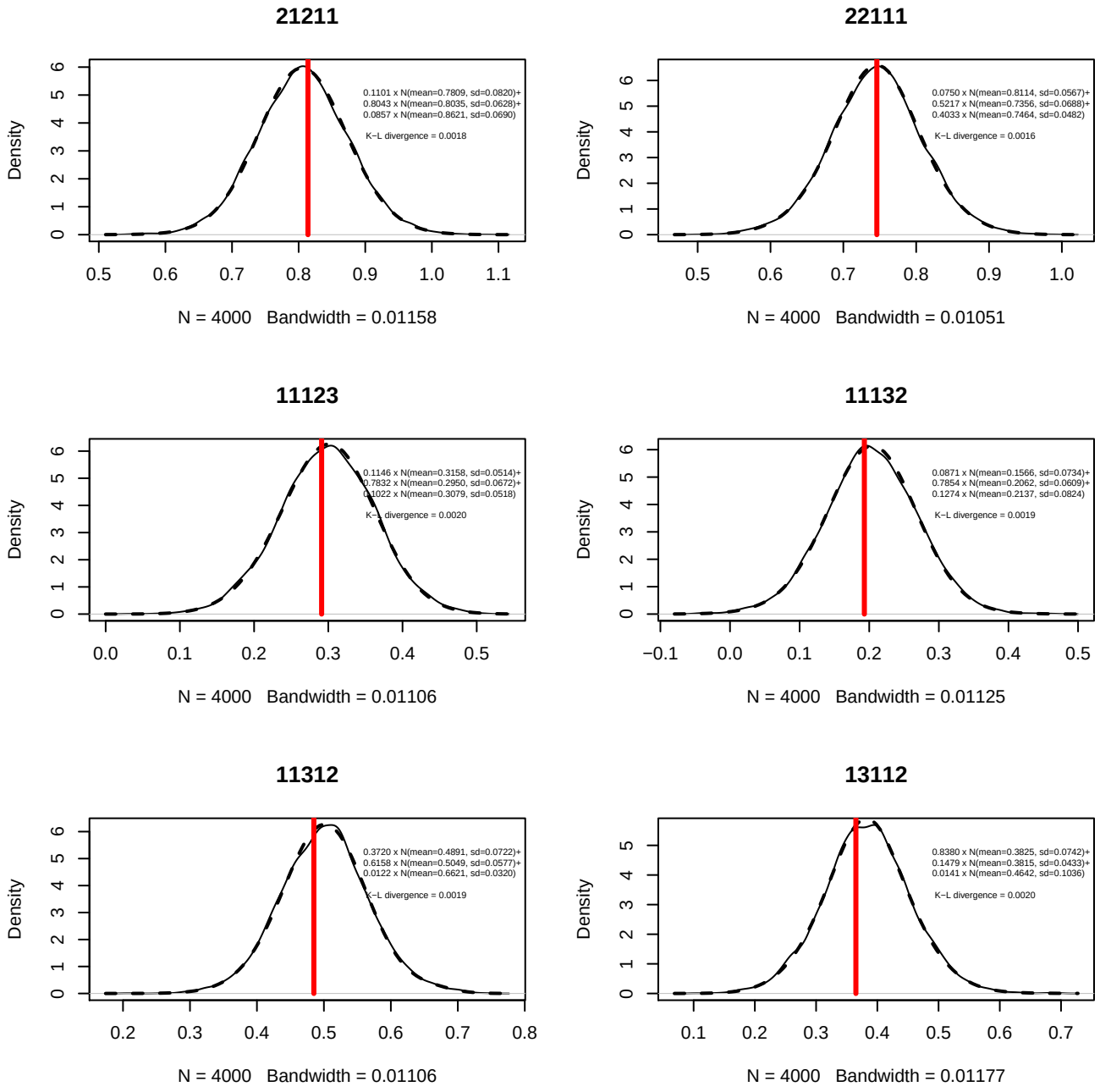


Figure 10: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

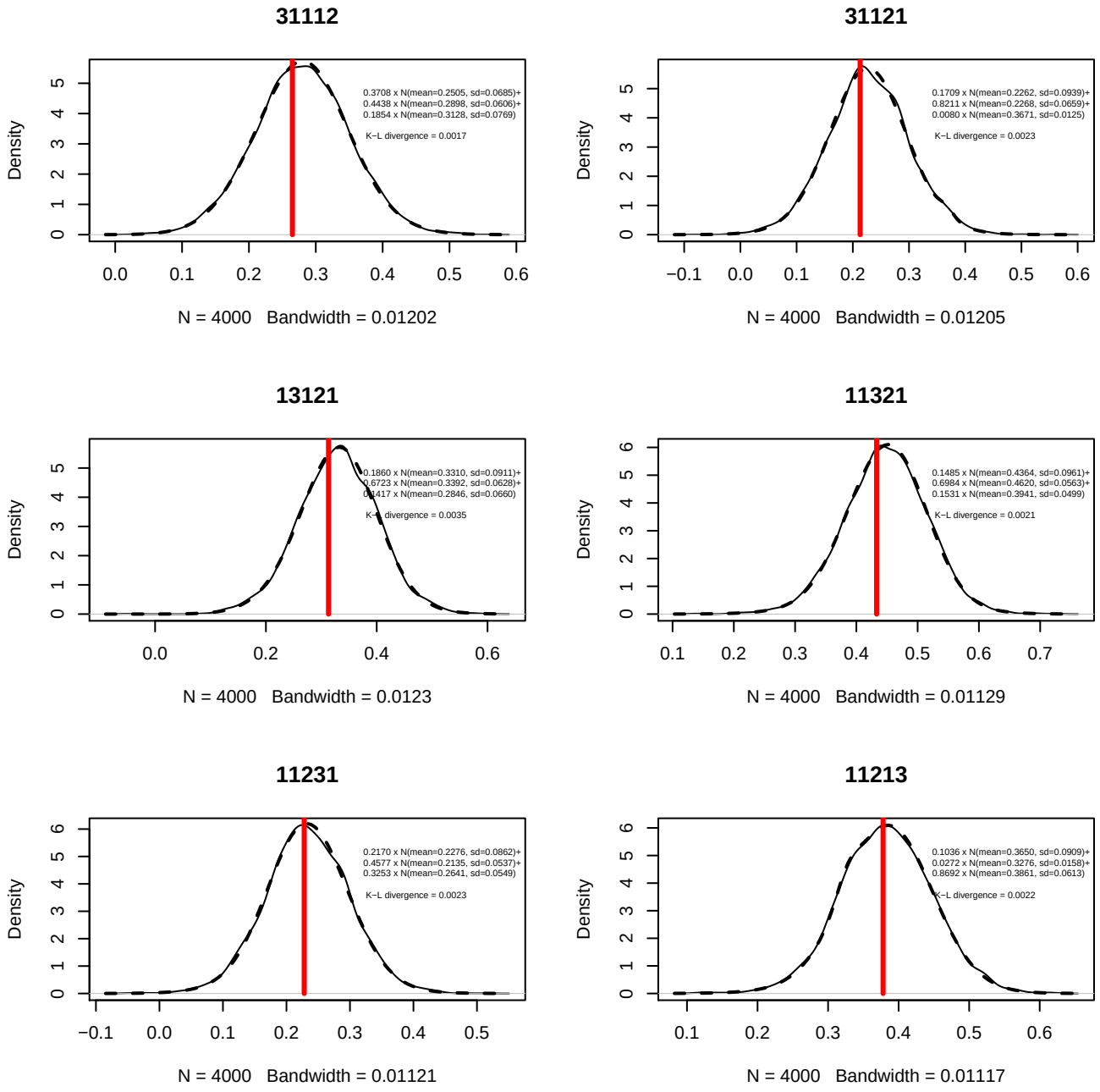


Figure 11: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

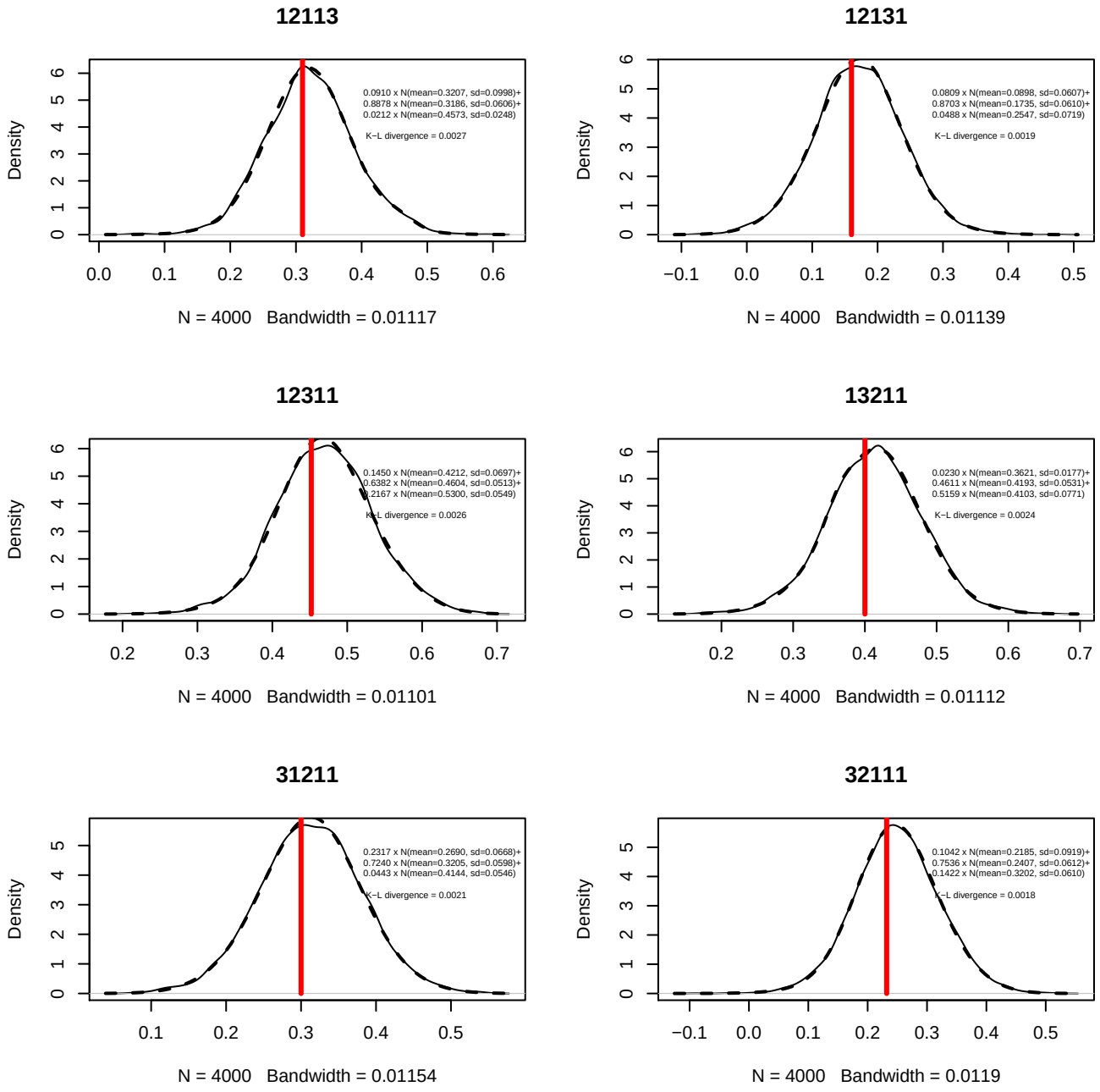


Figure 12: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

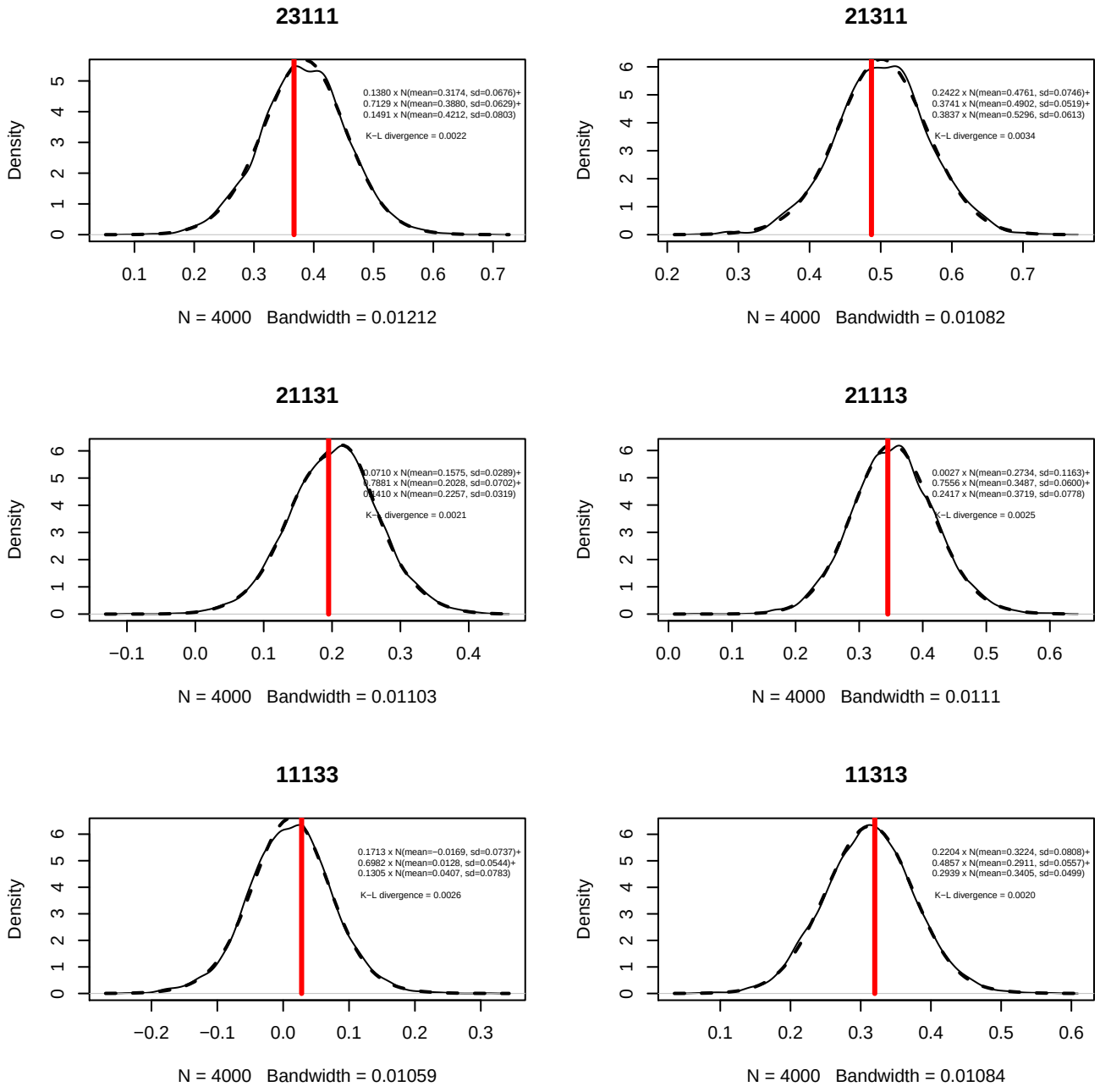


Figure 13: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

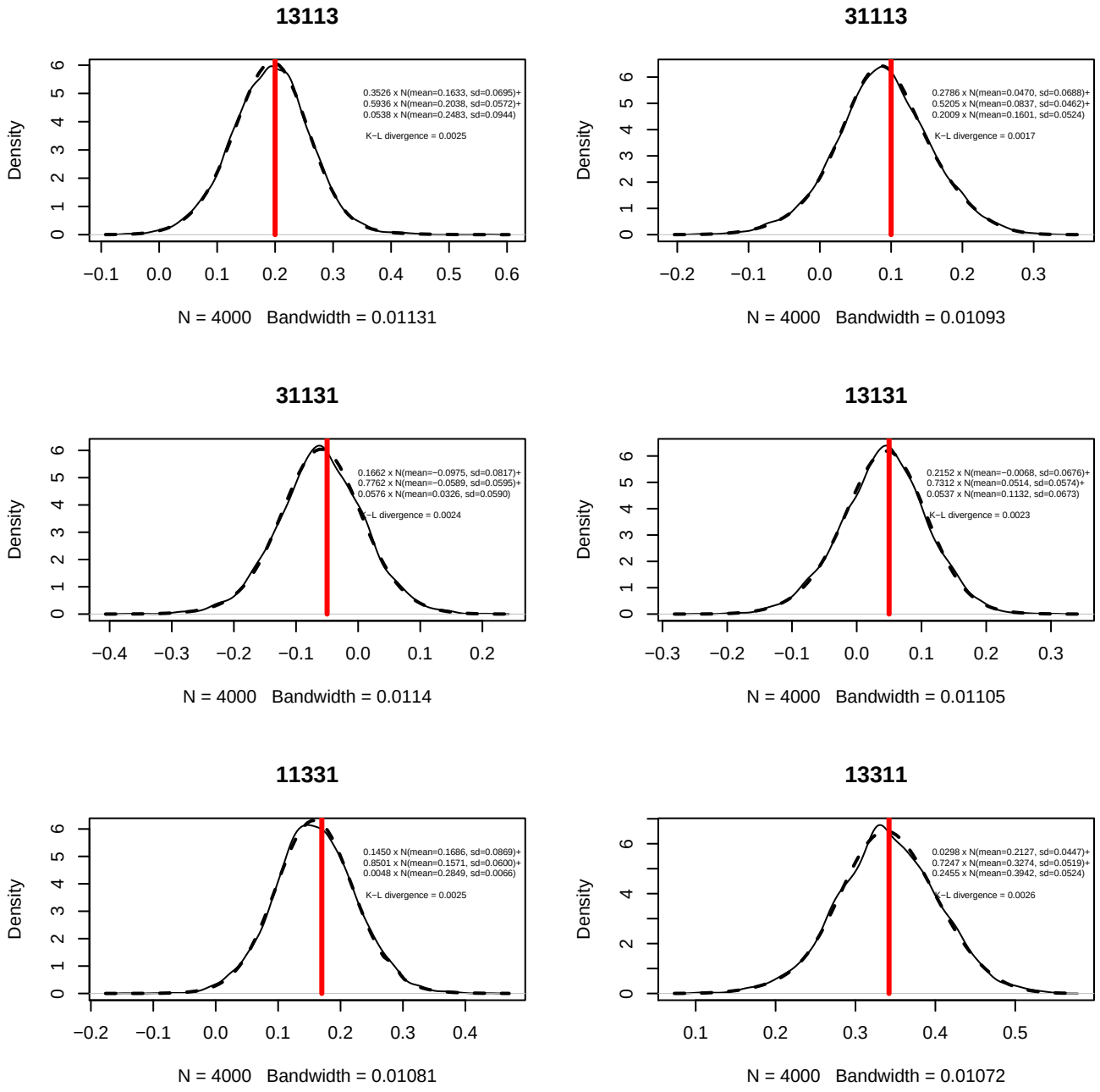


Figure 14: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

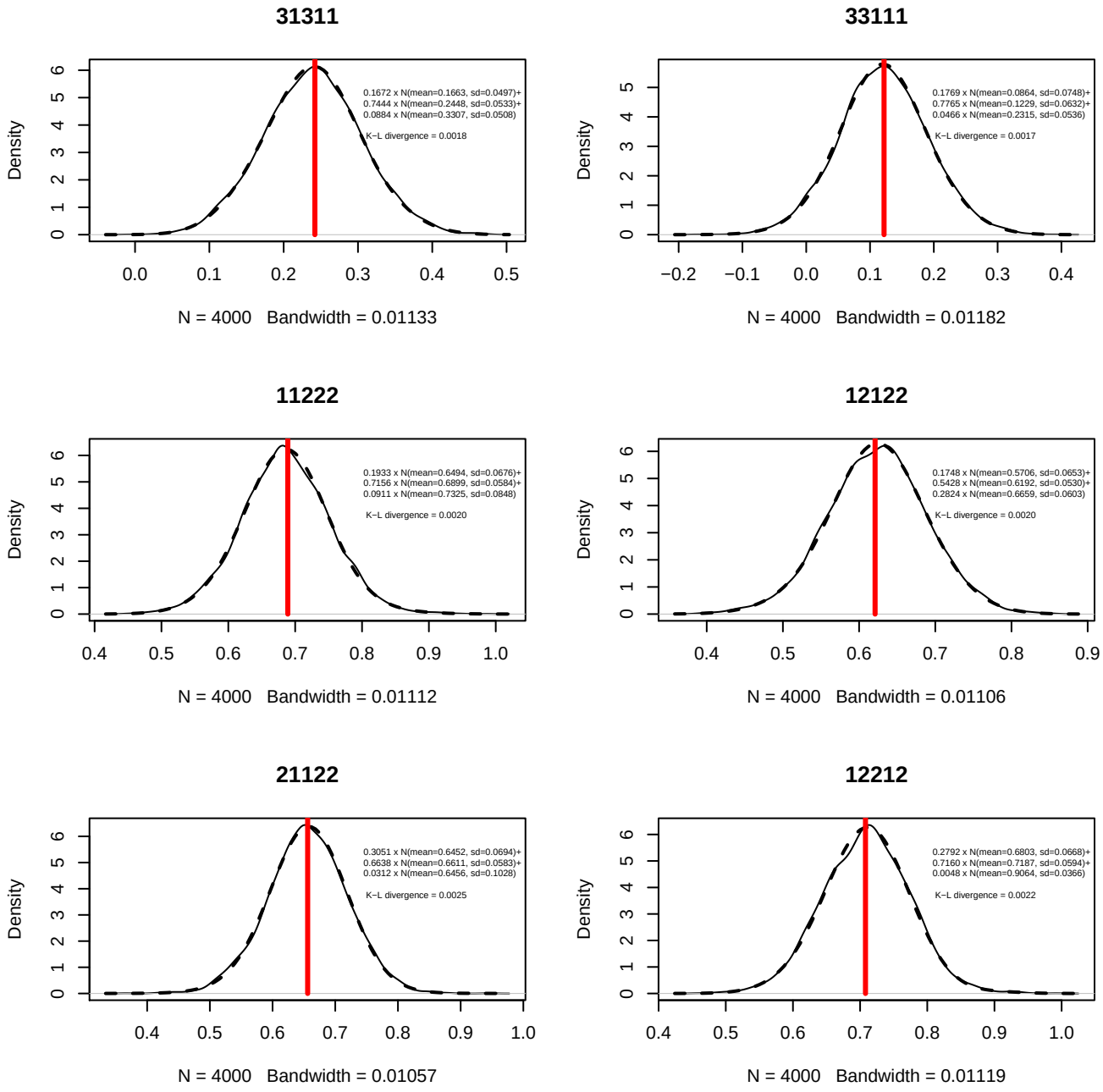


Figure 15: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

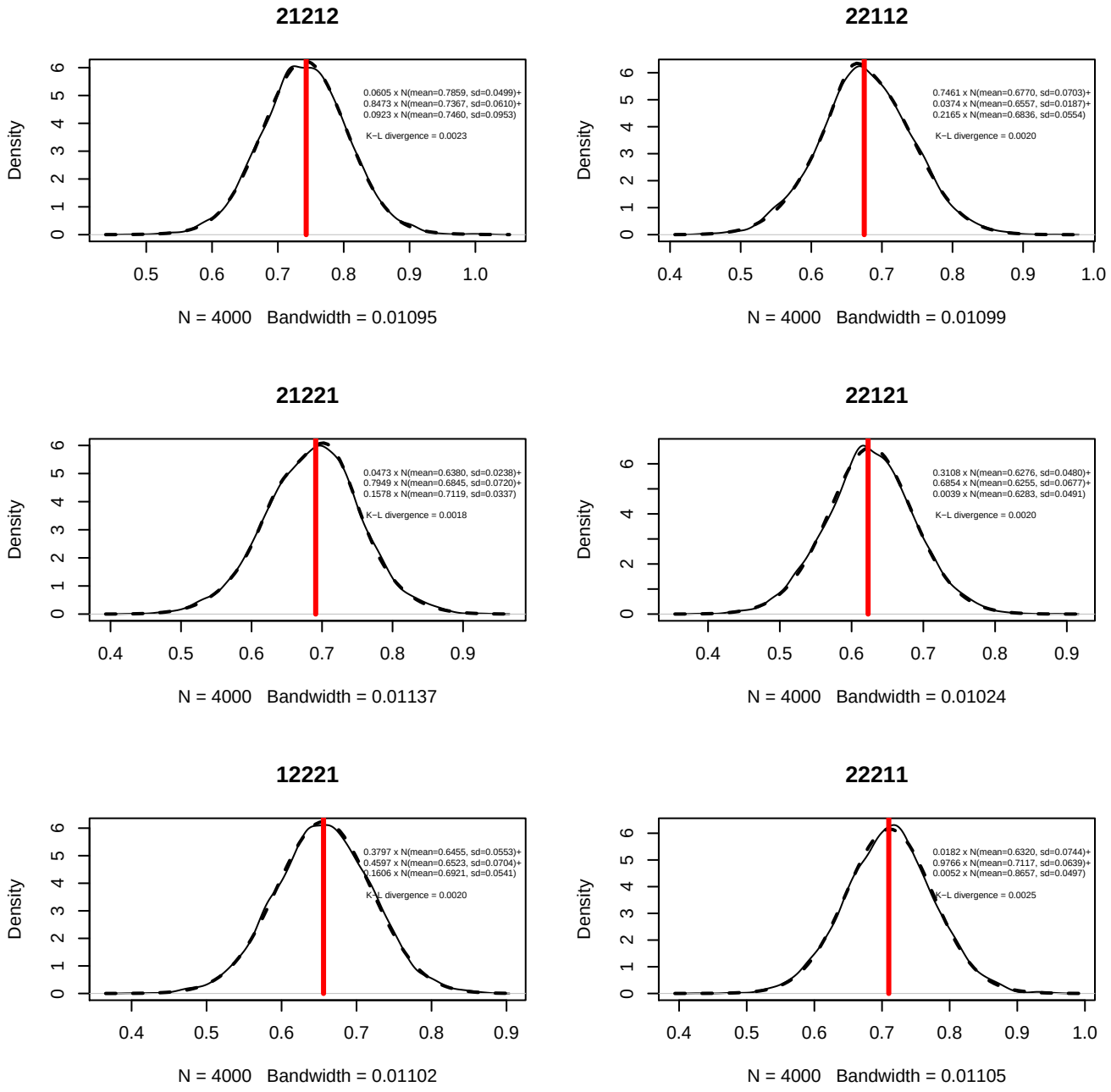


Figure 16: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

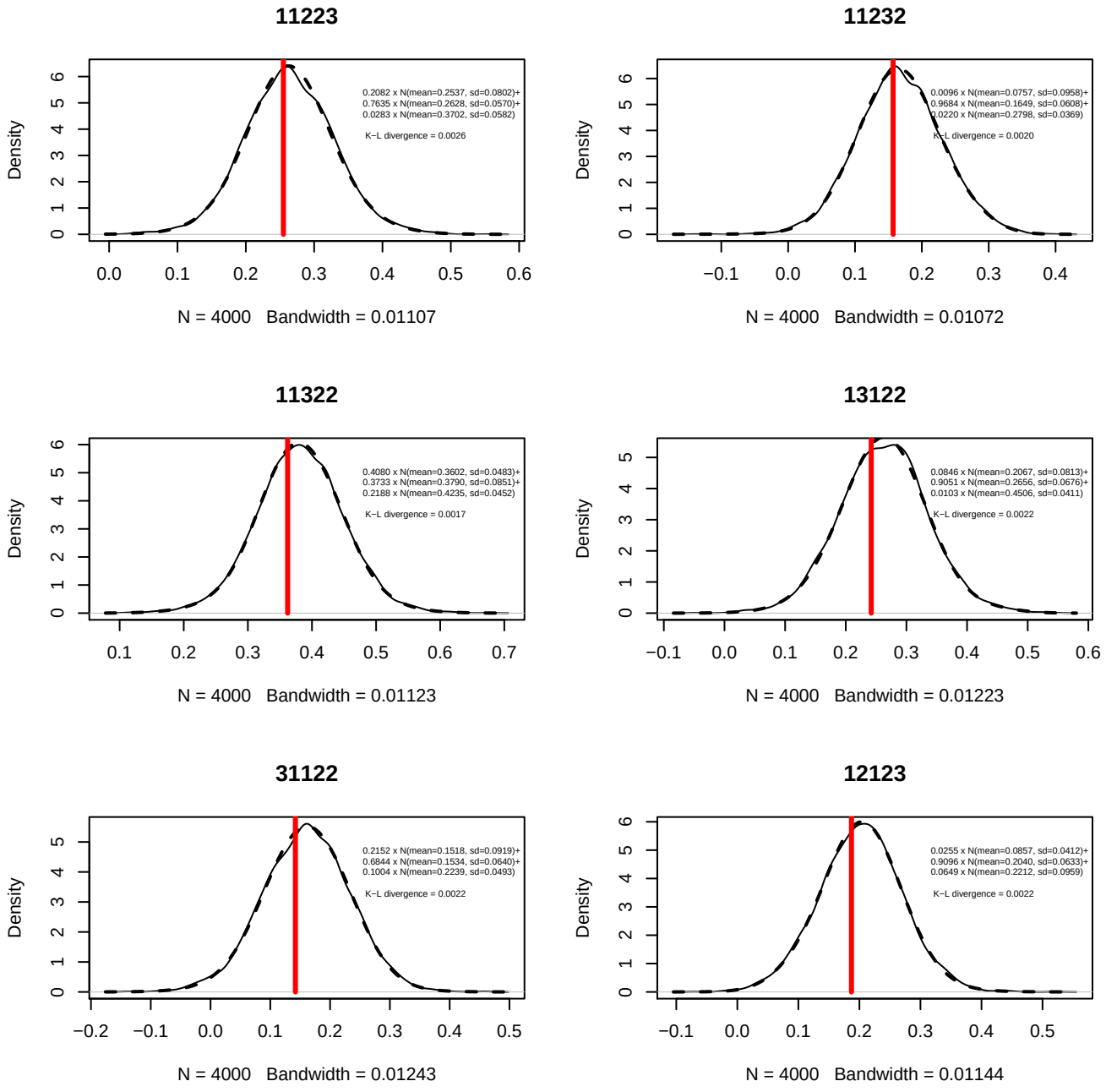


Figure 17: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

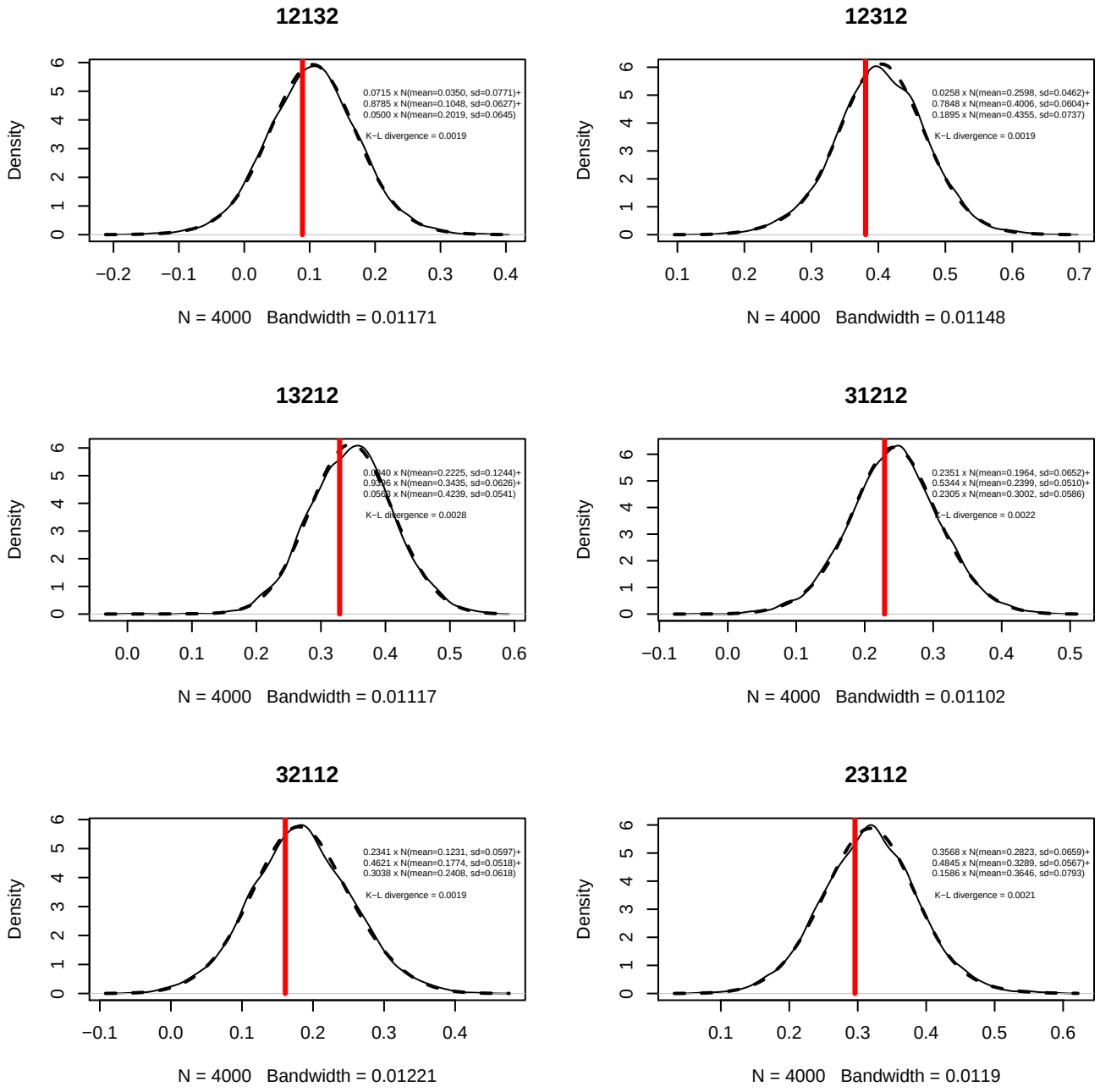


Figure 18: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

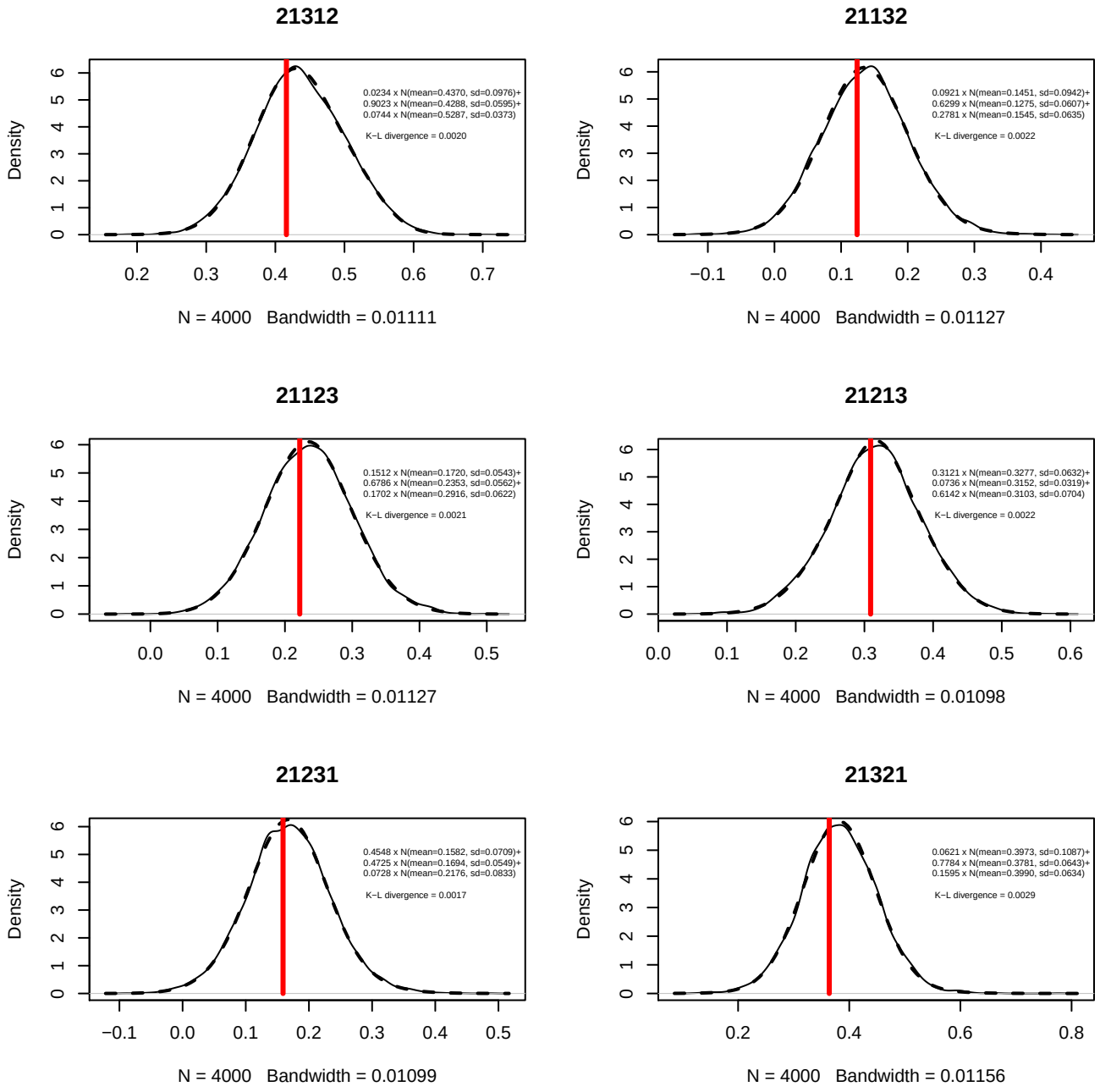


Figure 19: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

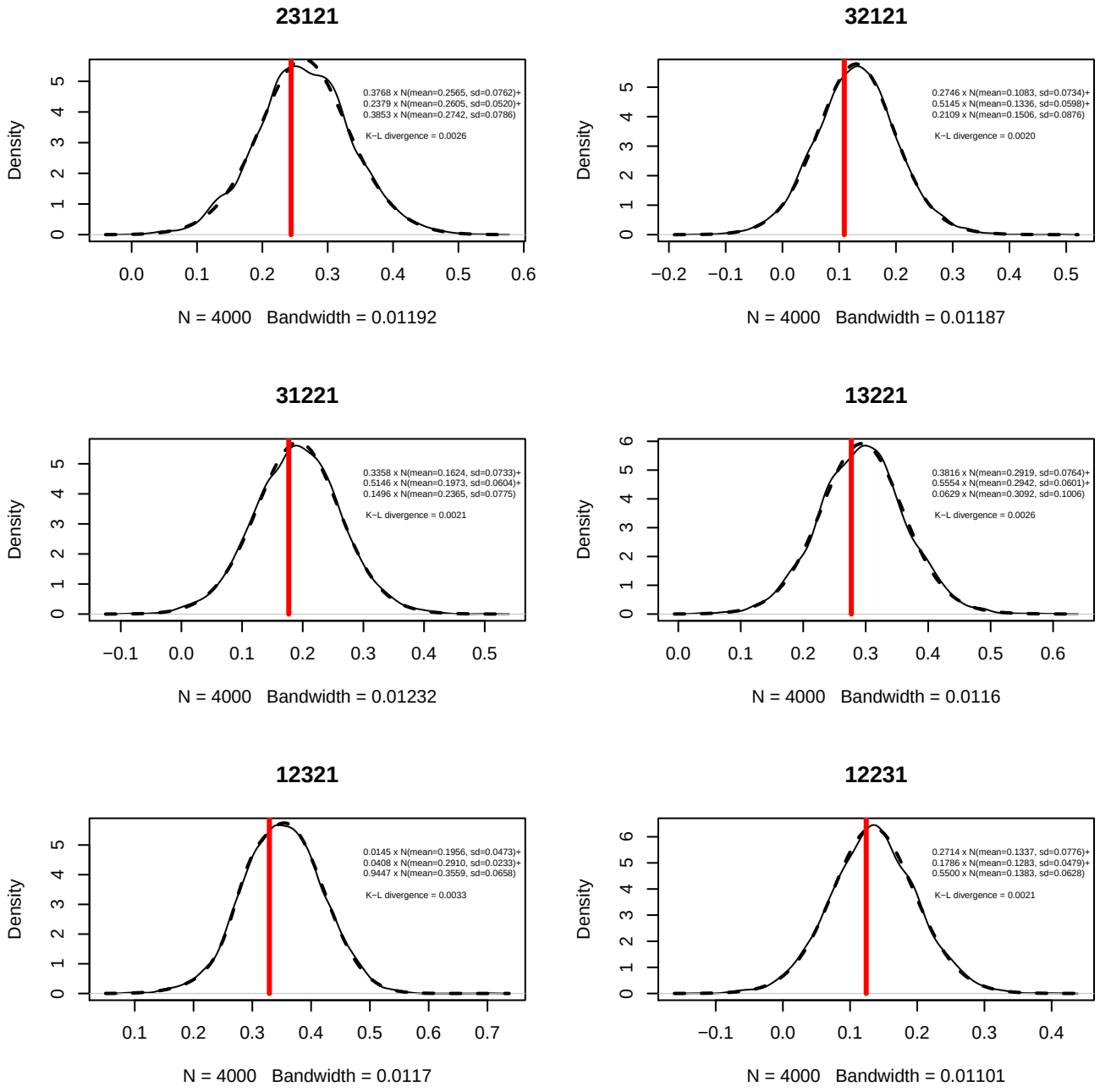


Figure 20: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

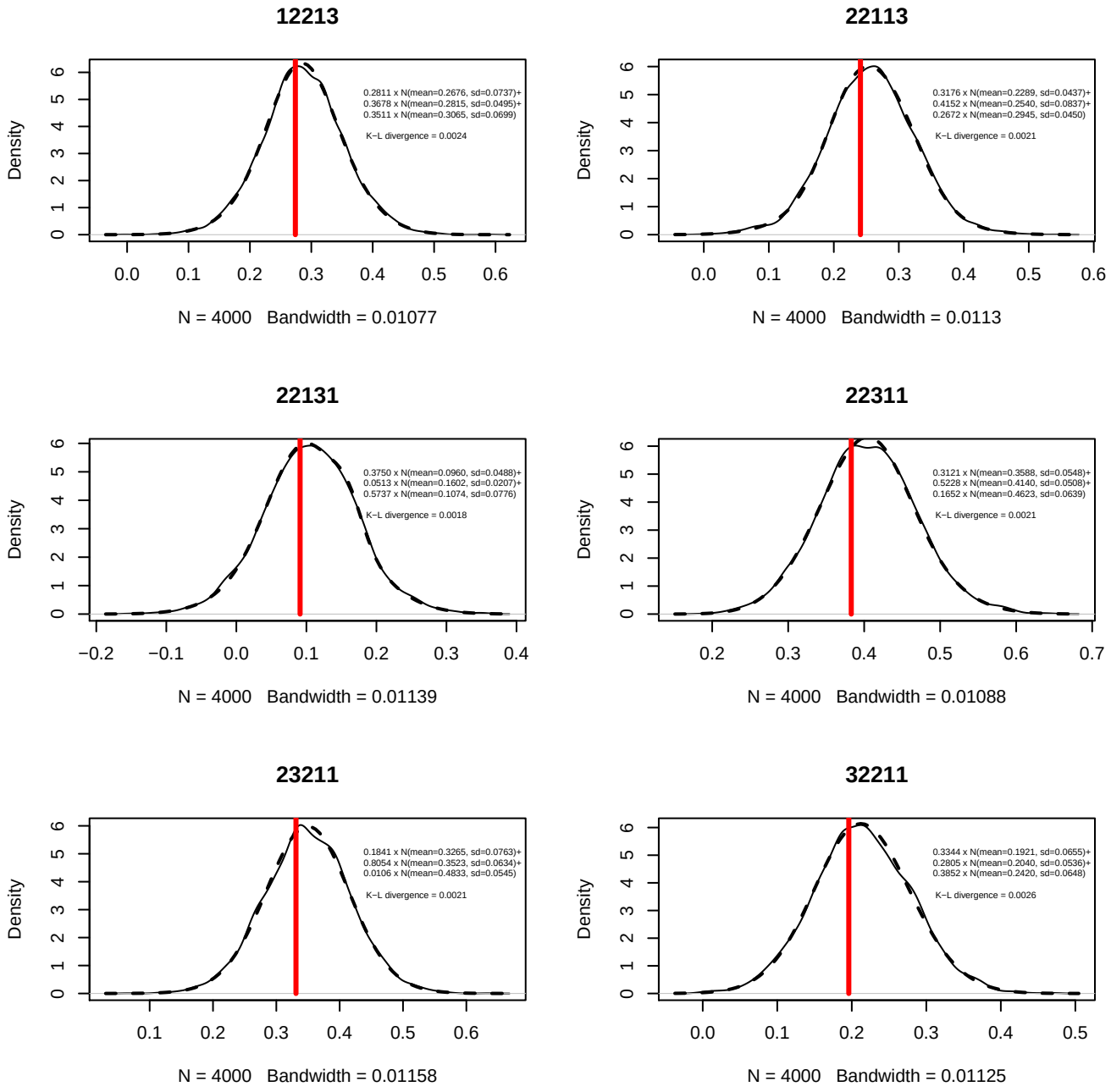


Figure 21: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

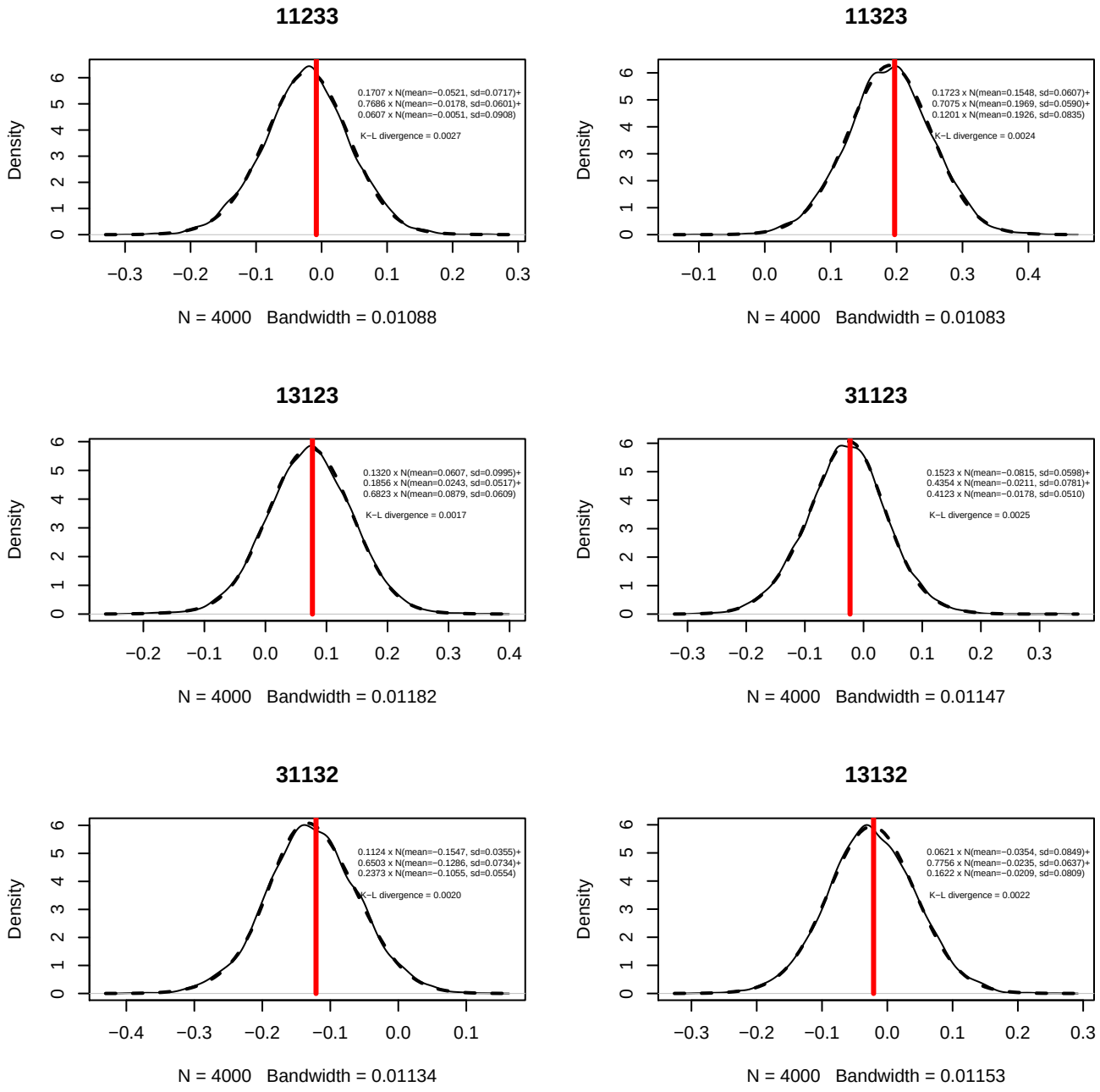


Figure 22: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

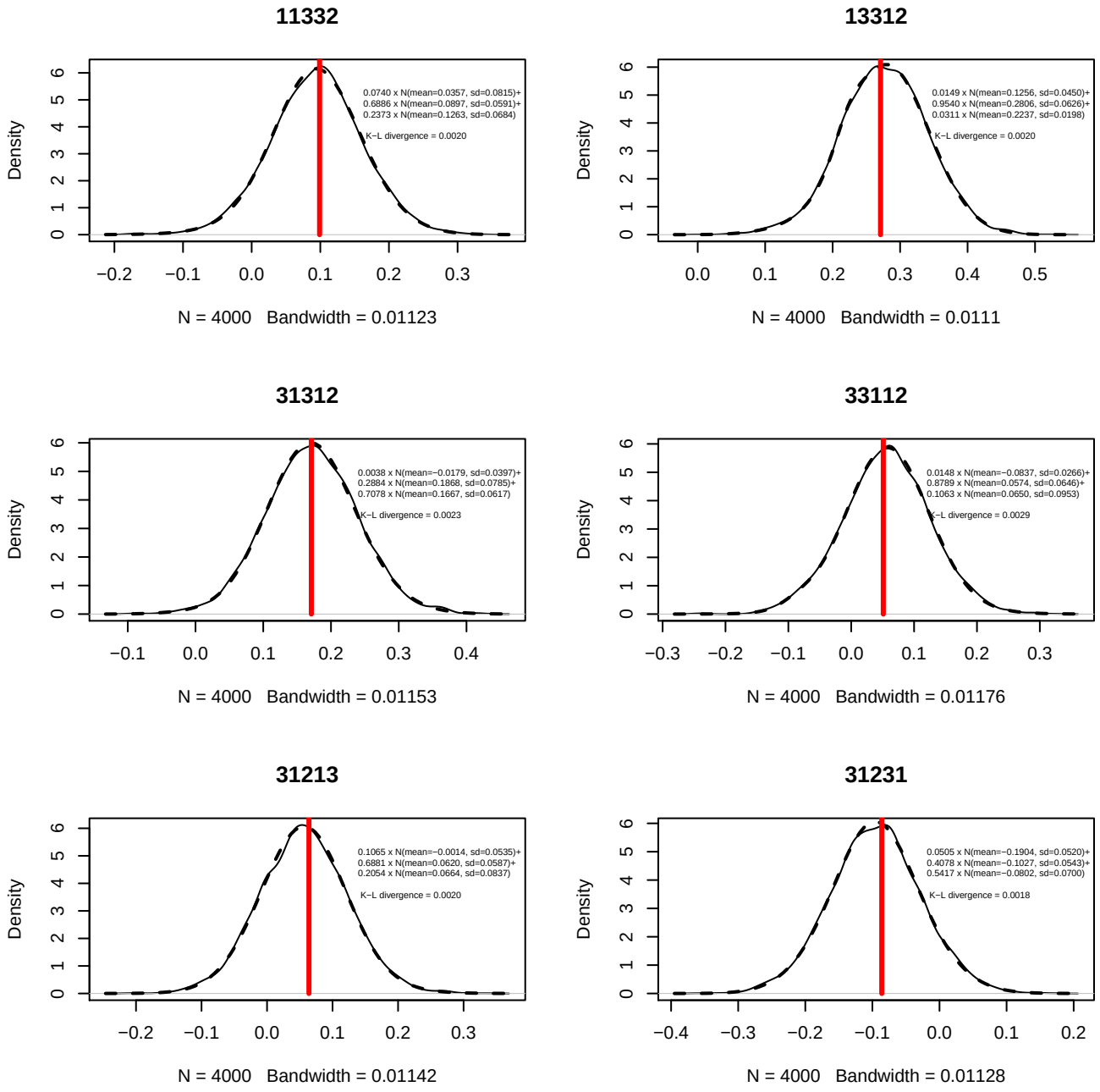


Figure 23: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

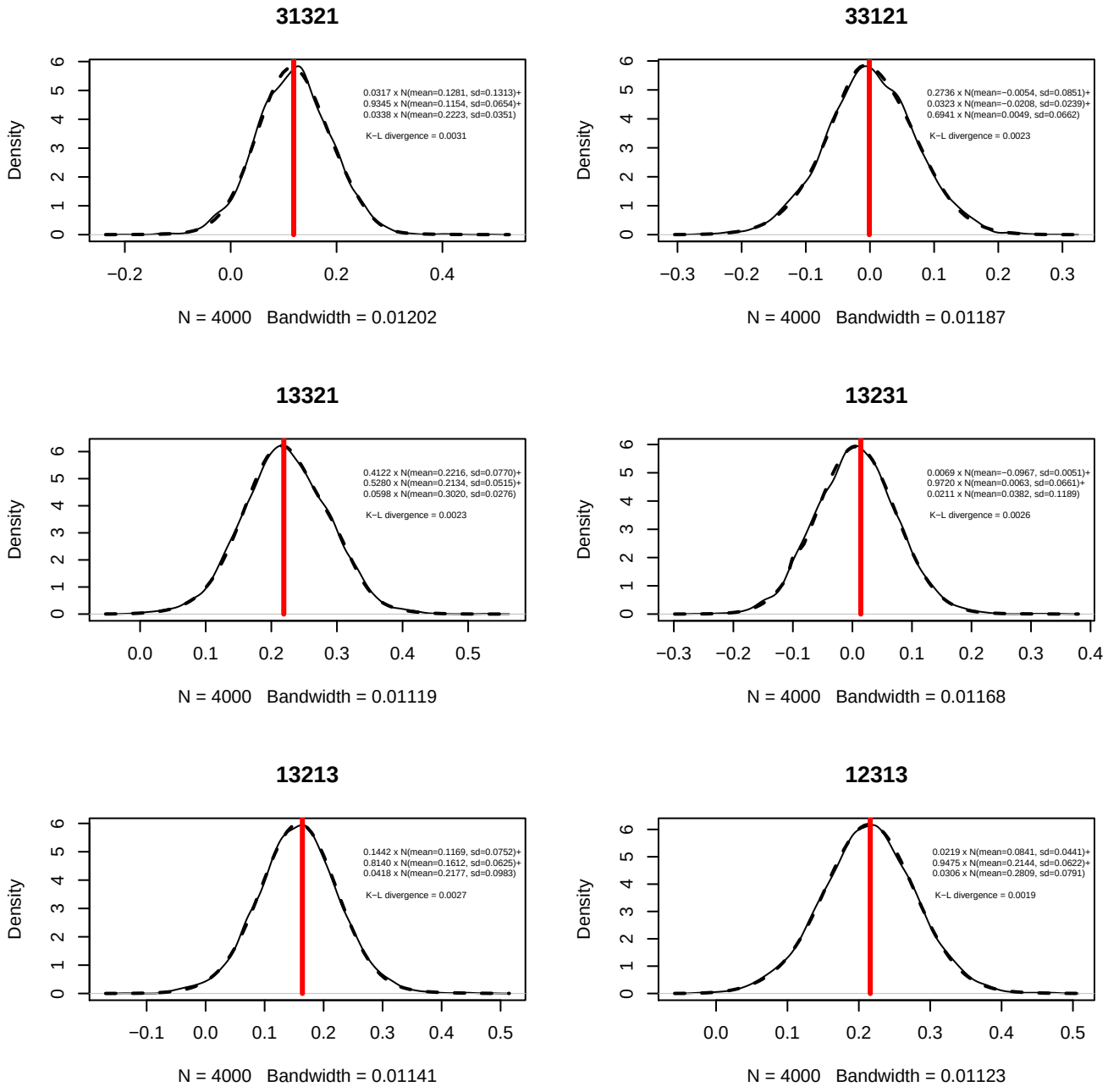


Figure 24: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

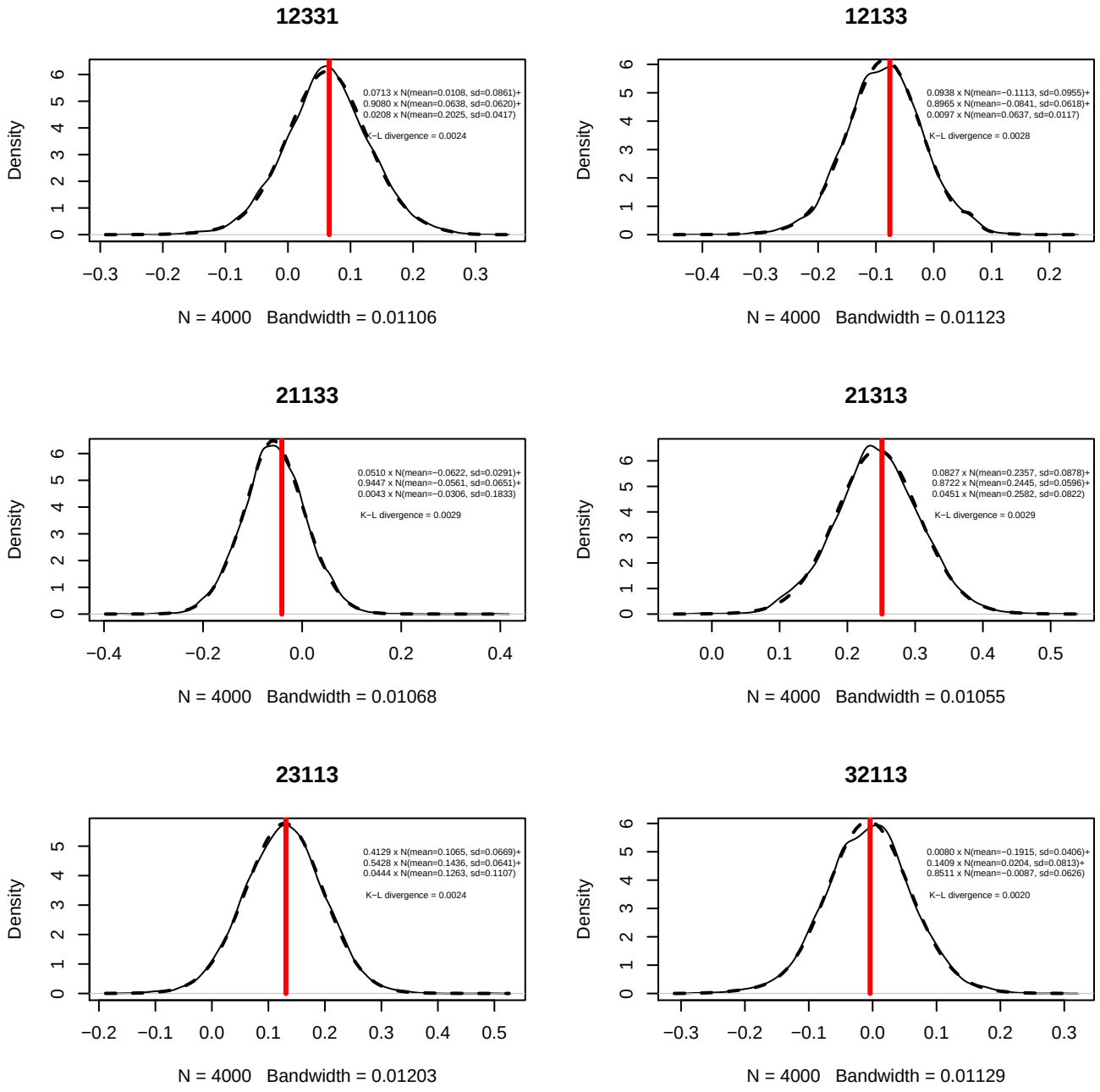


Figure 25: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

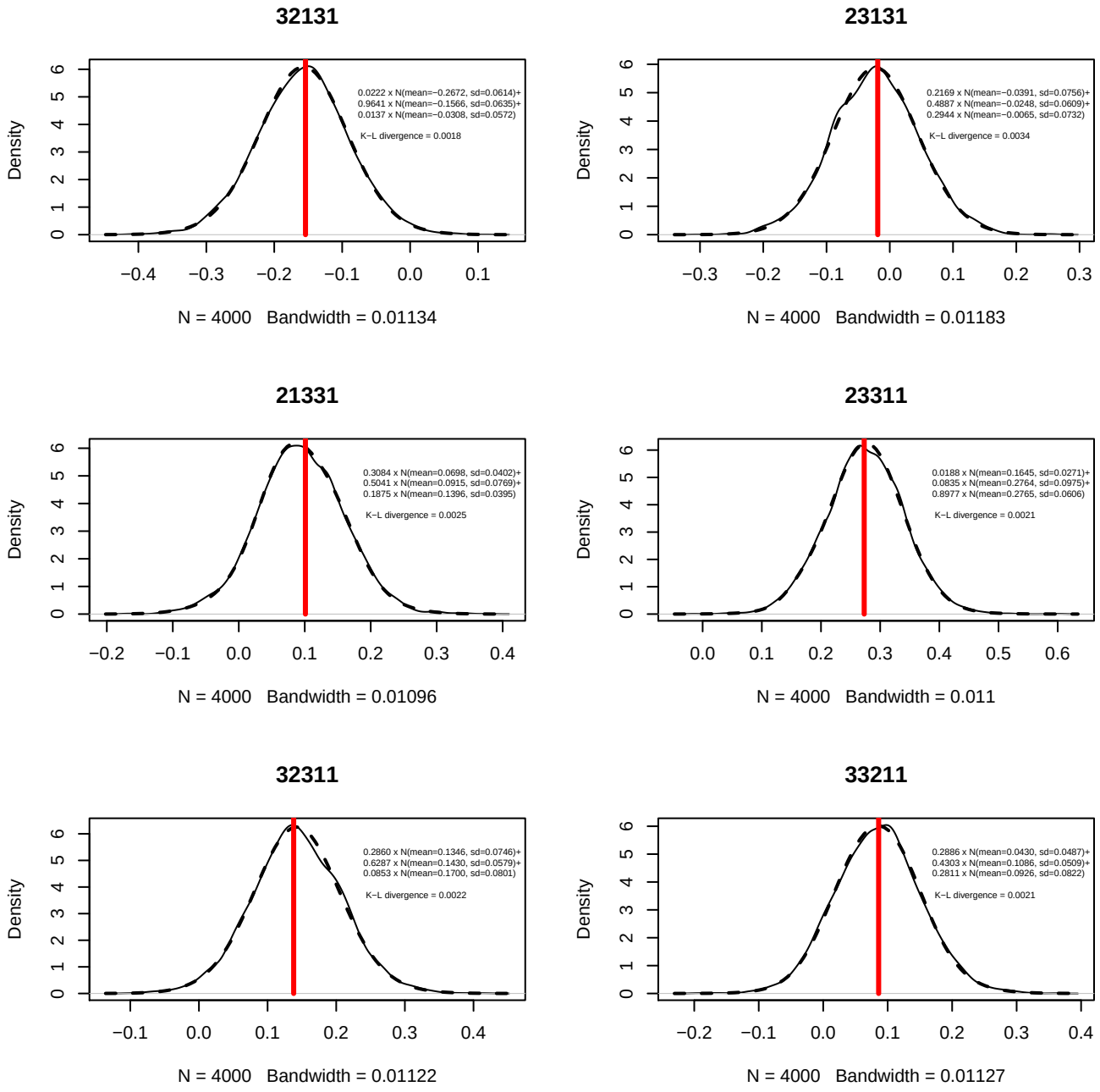


Figure 26: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

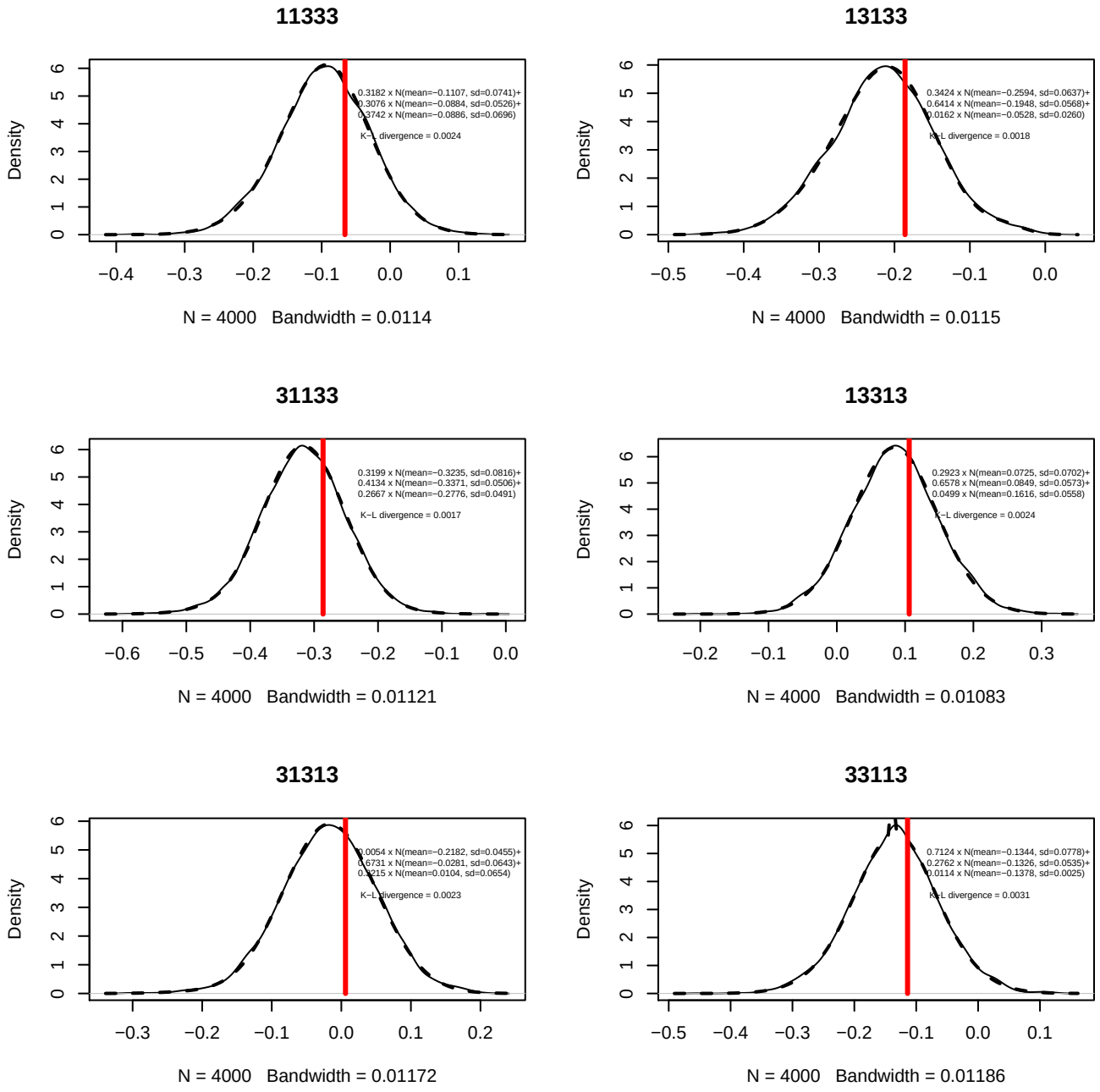


Figure 27: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

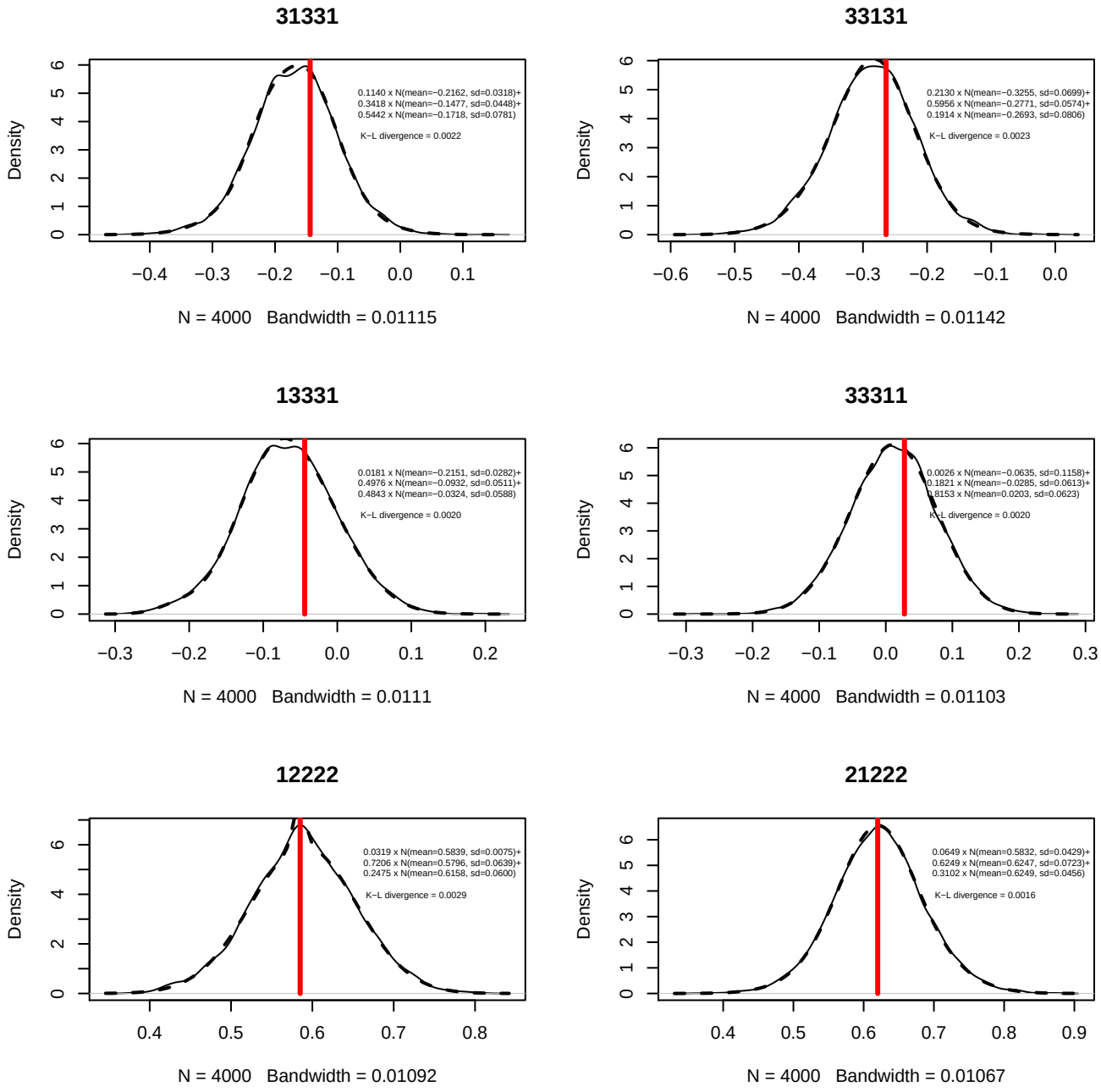


Figure 28: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

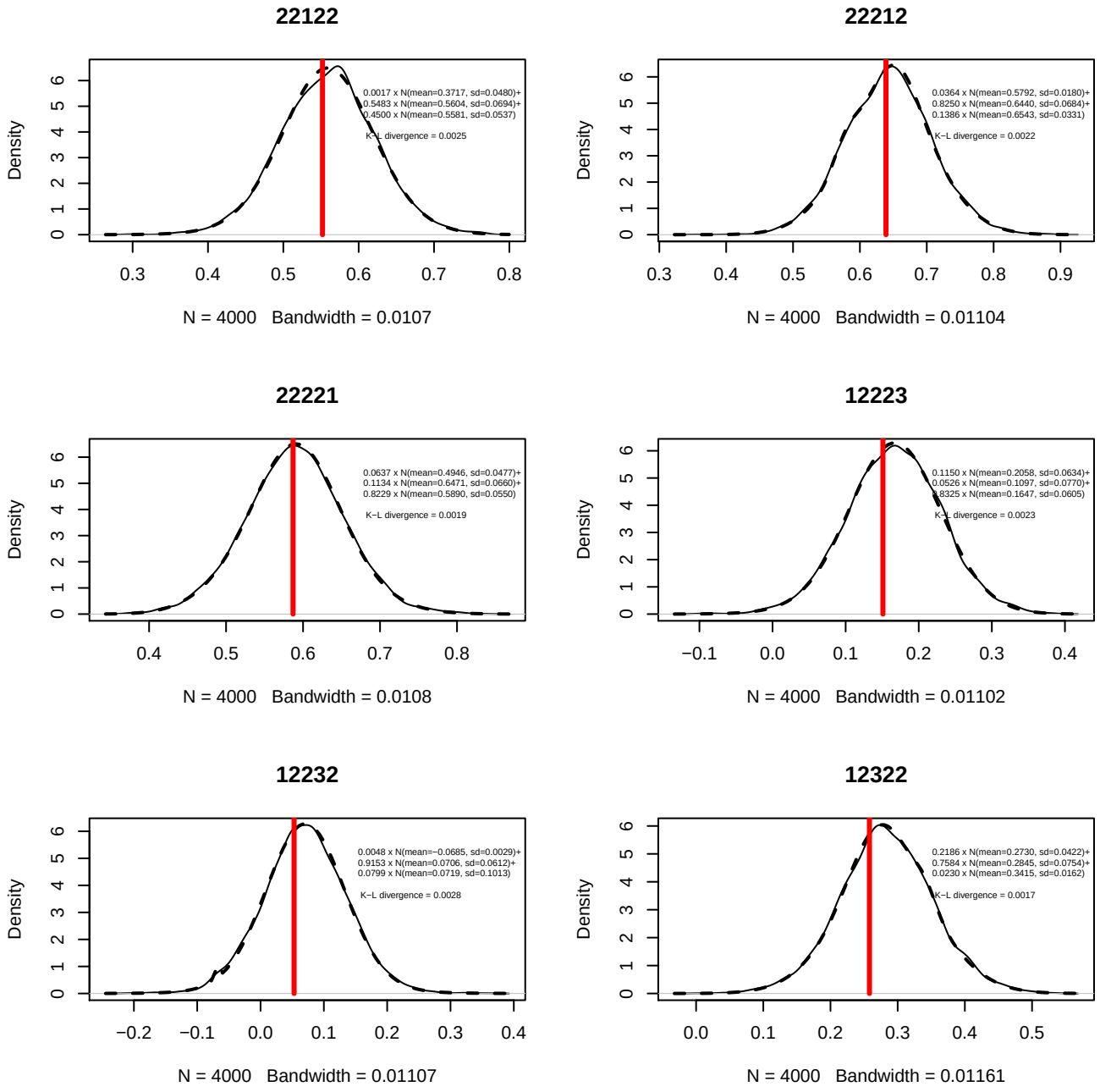


Figure 29: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

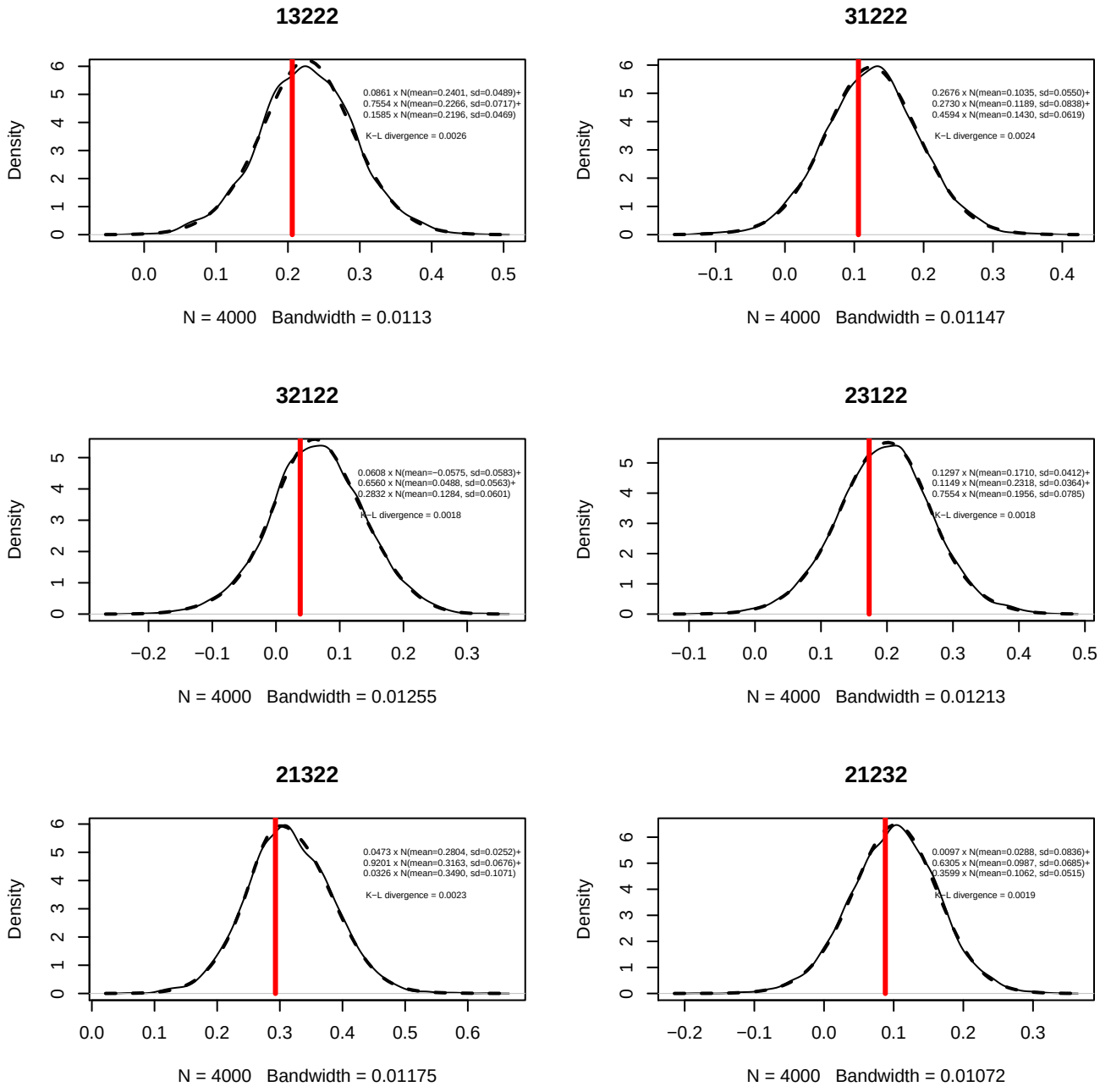


Figure 30: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

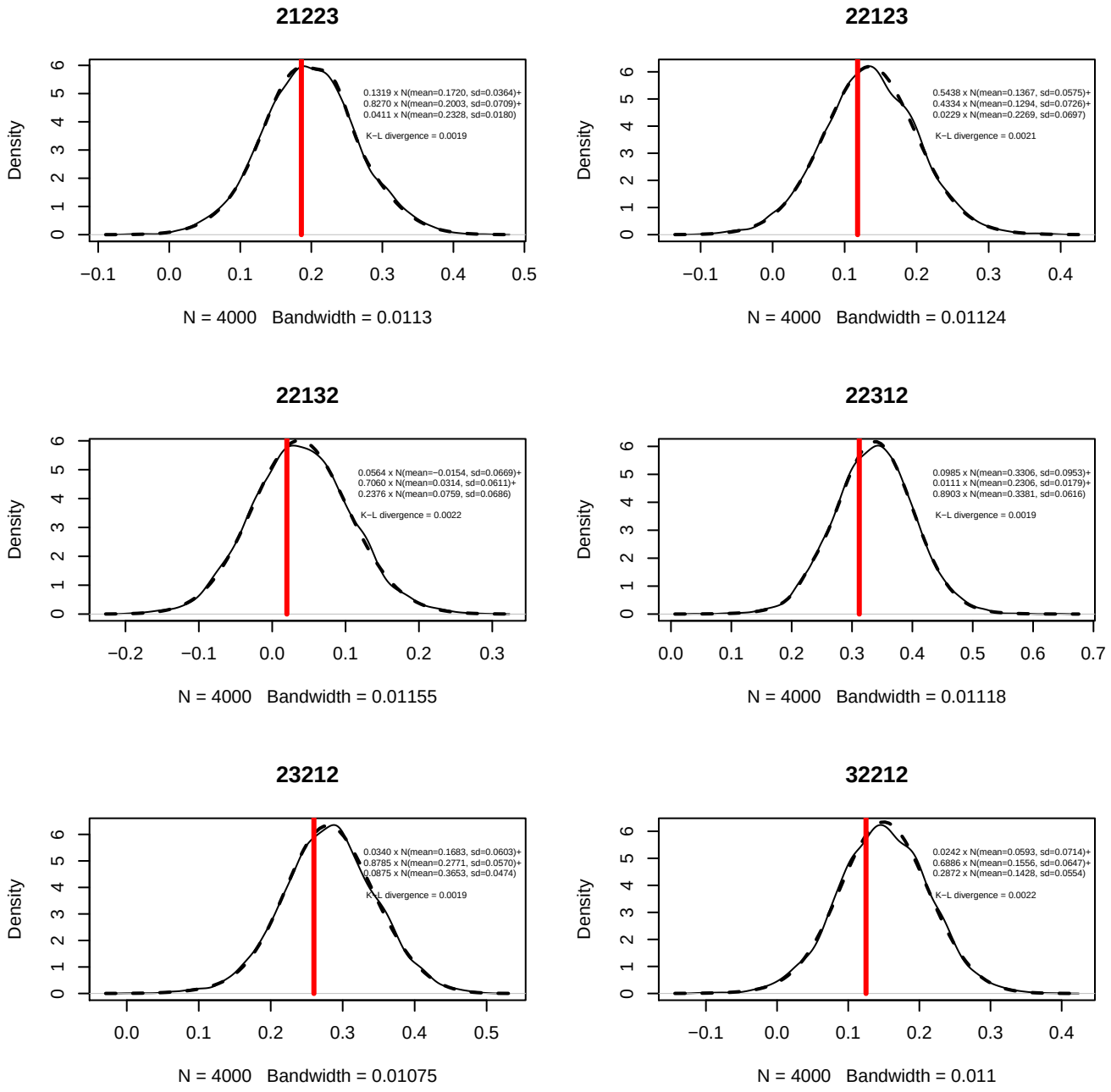


Figure 31: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

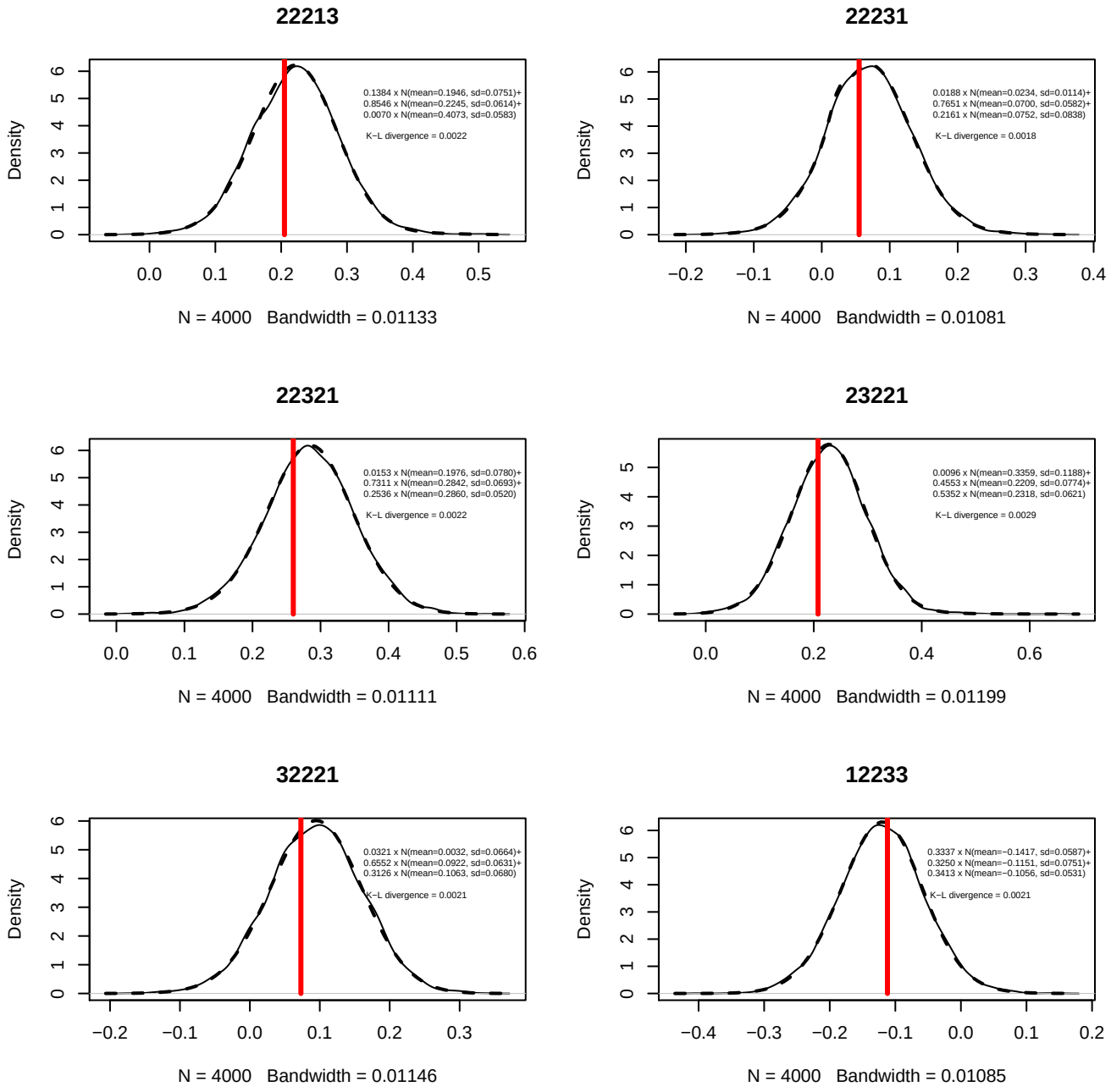


Figure 32: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

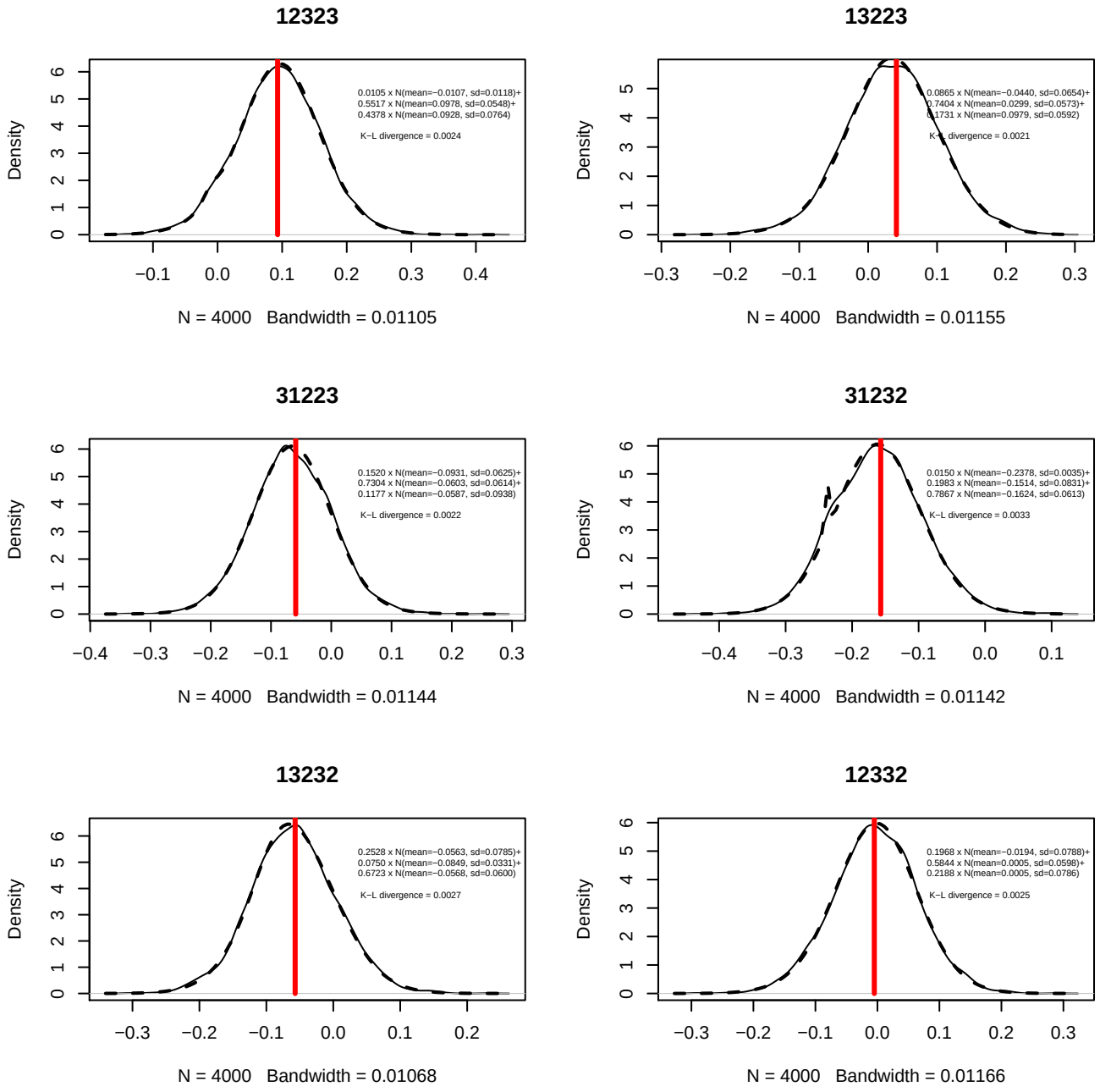


Figure 33: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

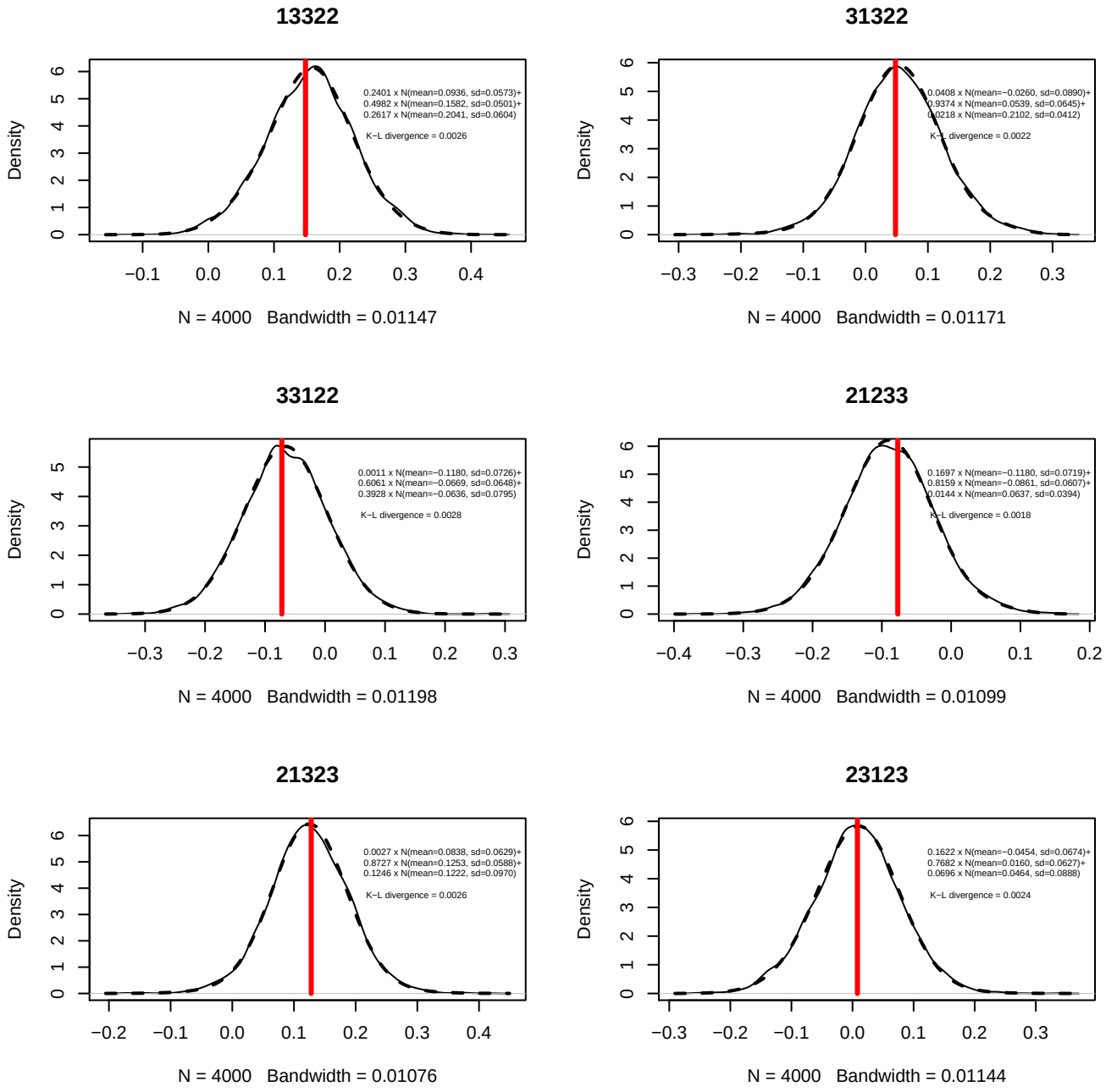


Figure 34: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

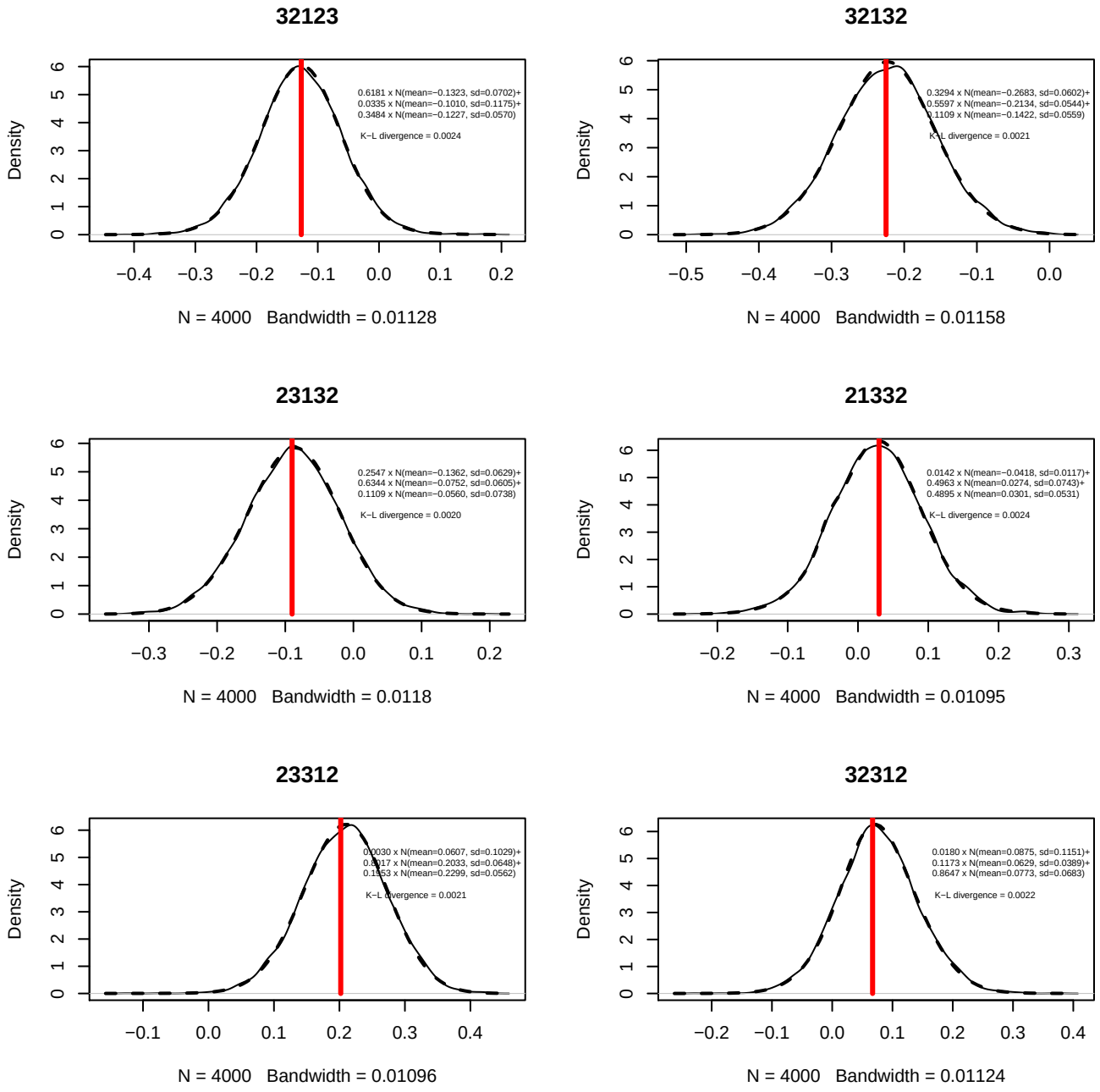


Figure 35: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

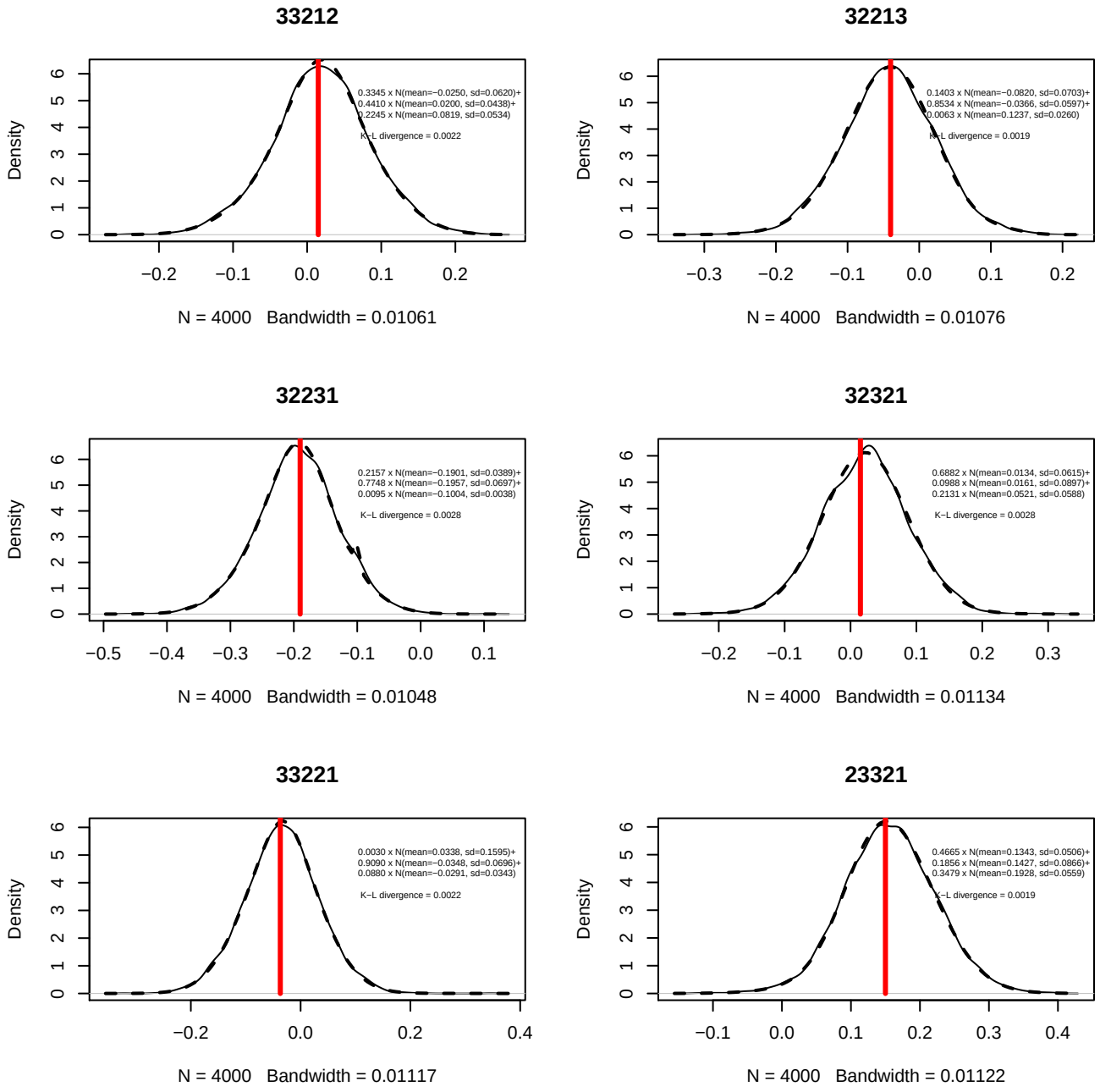


Figure 36: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

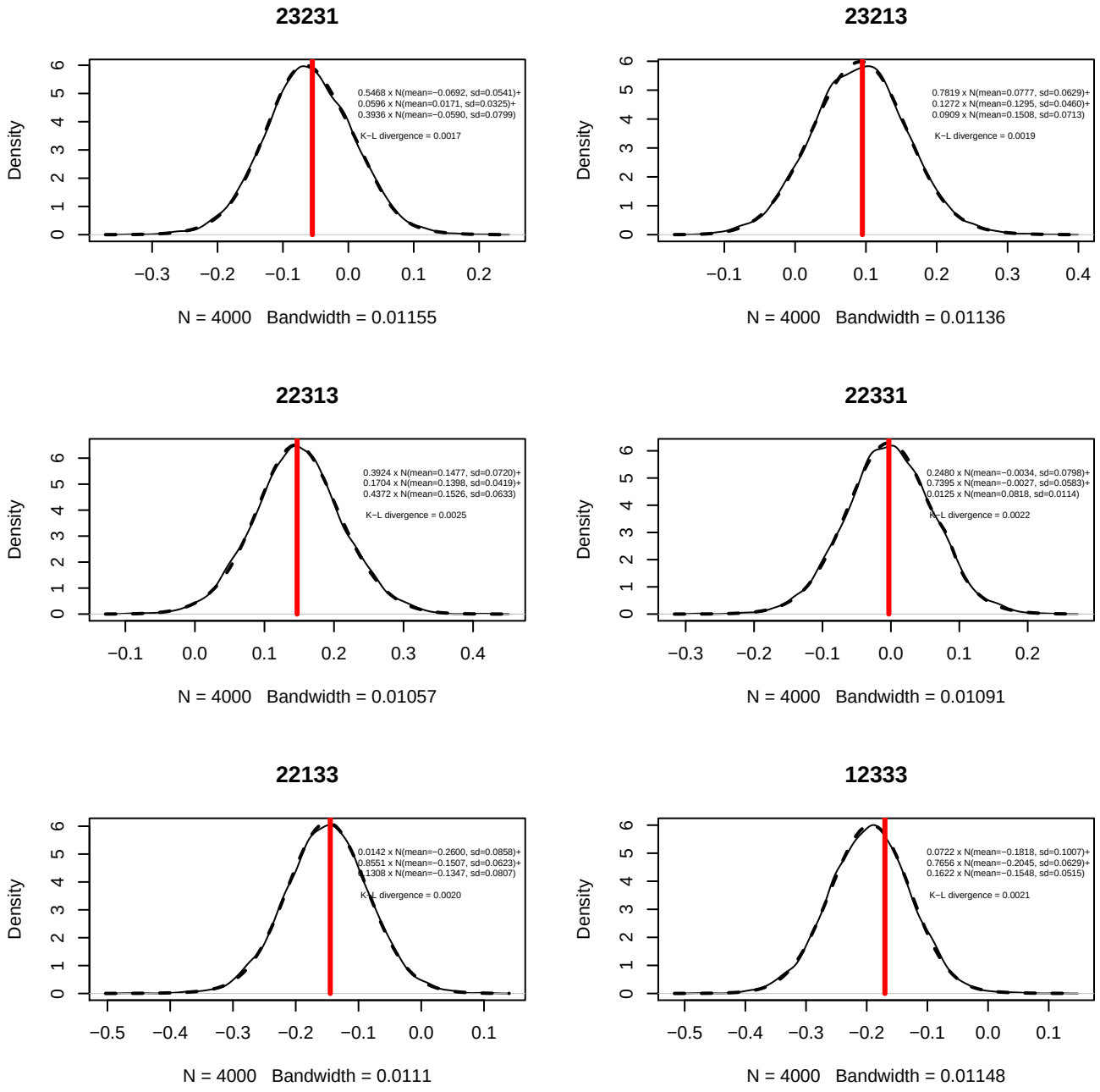


Figure 37: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

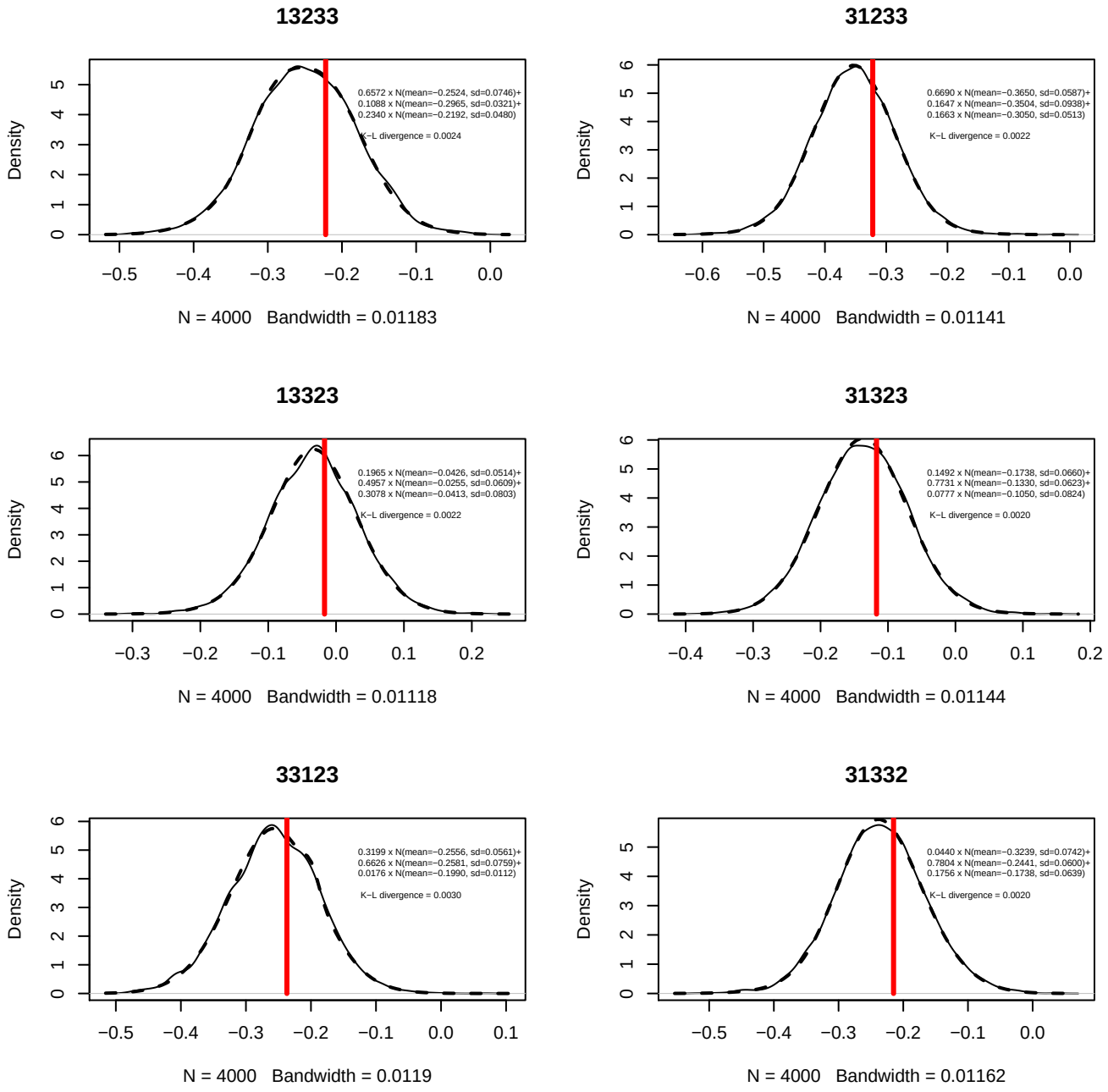


Figure 38: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

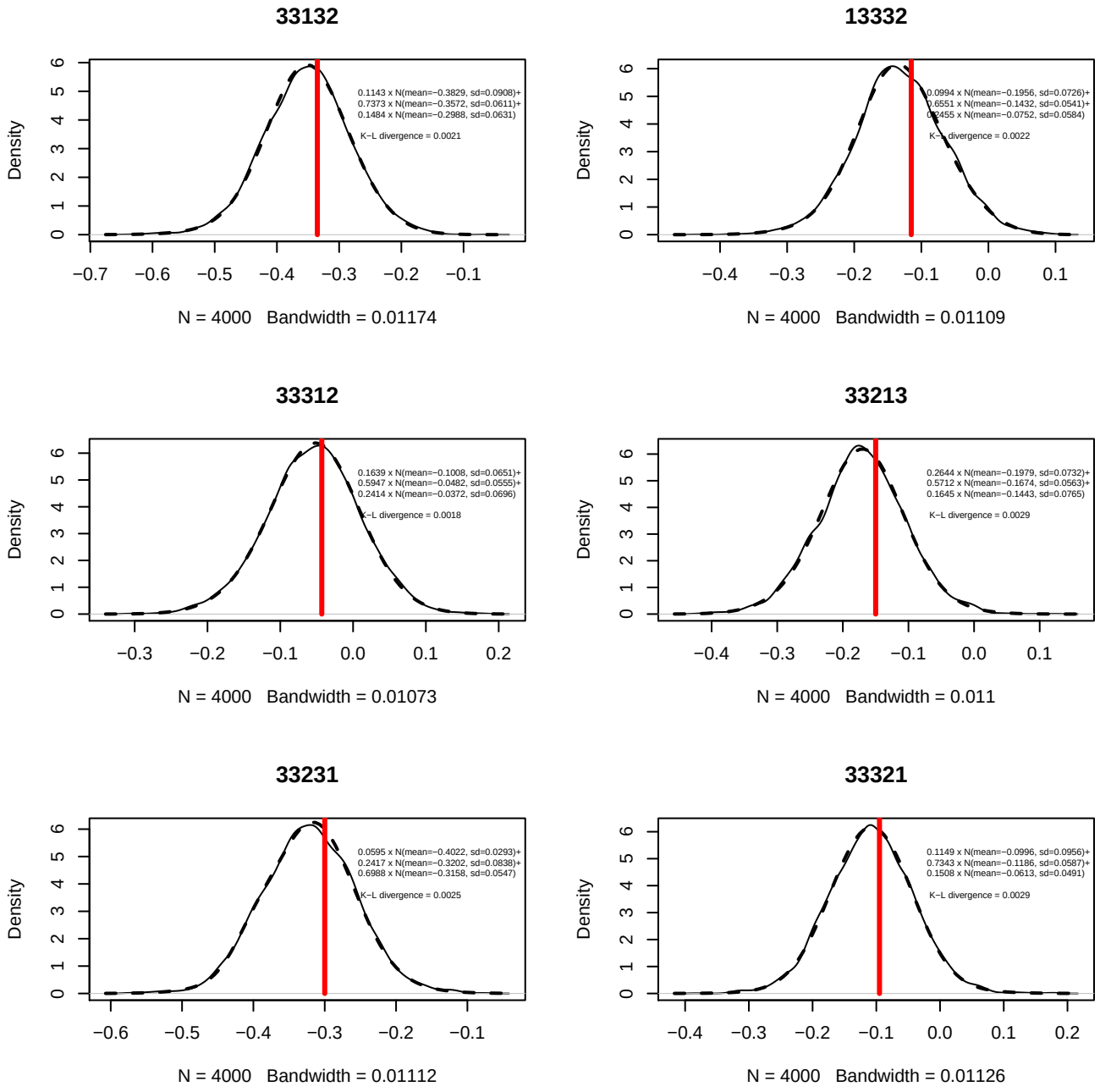


Figure 39: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

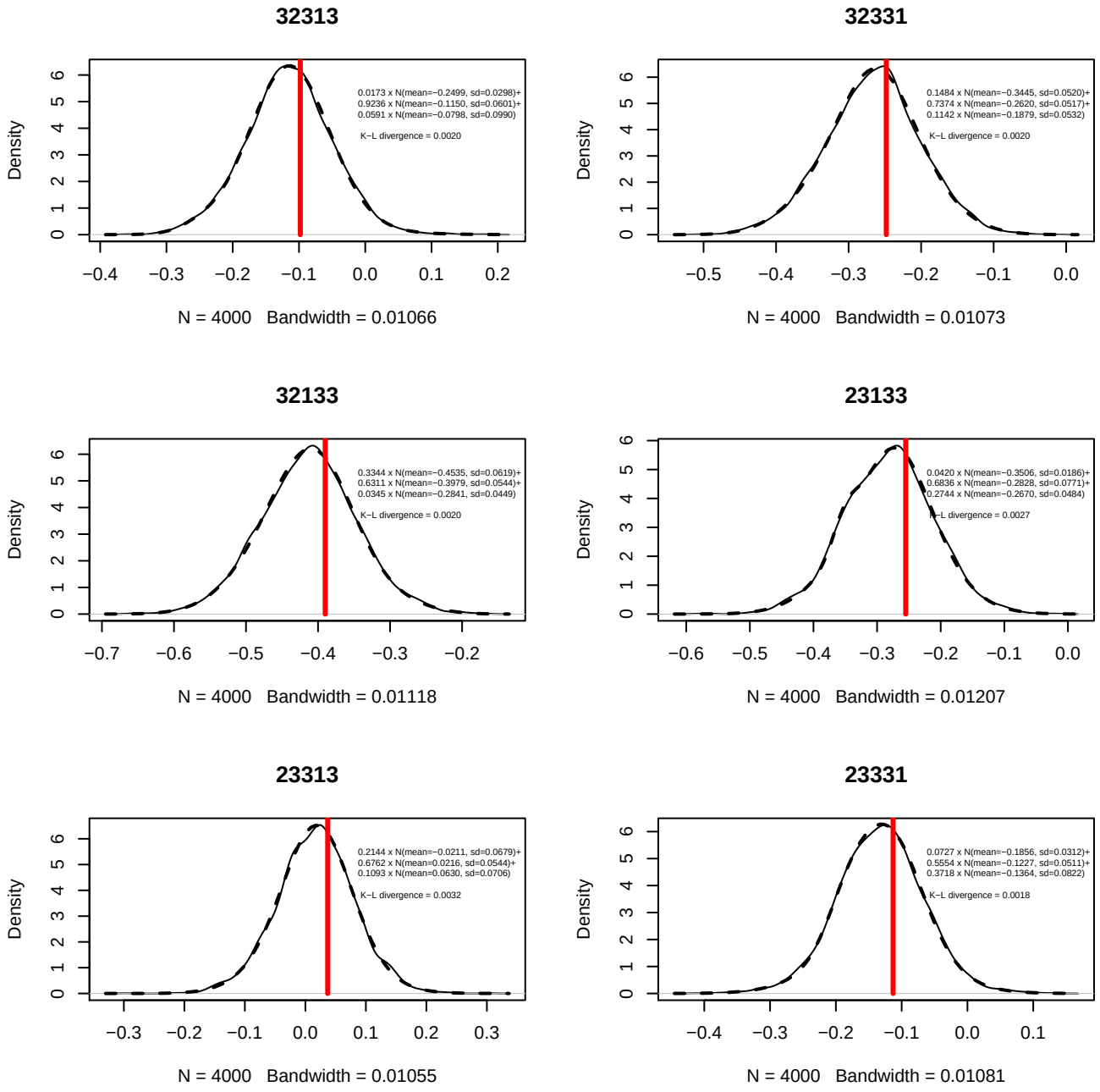


Figure 40: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

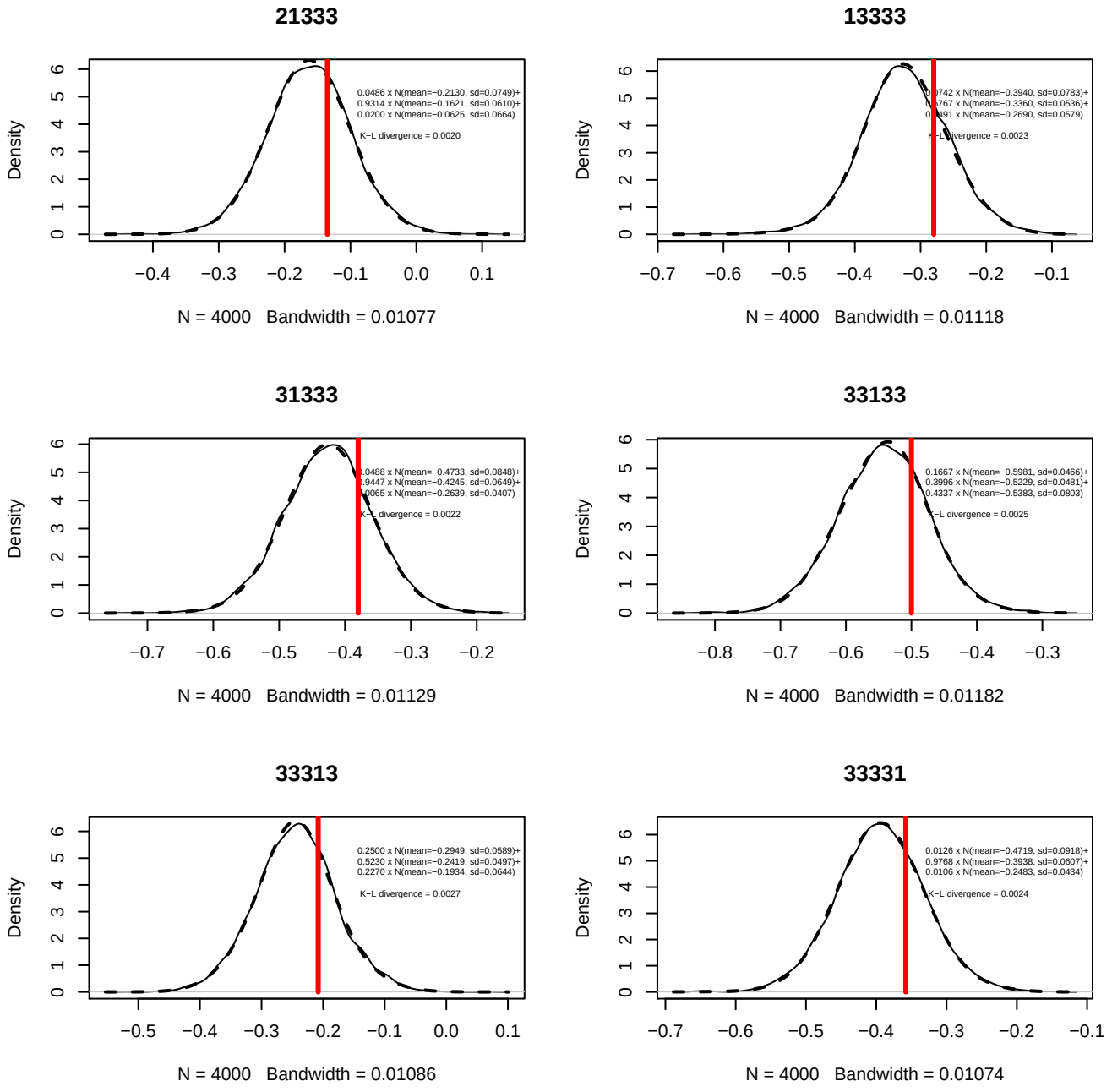


Figure 41: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

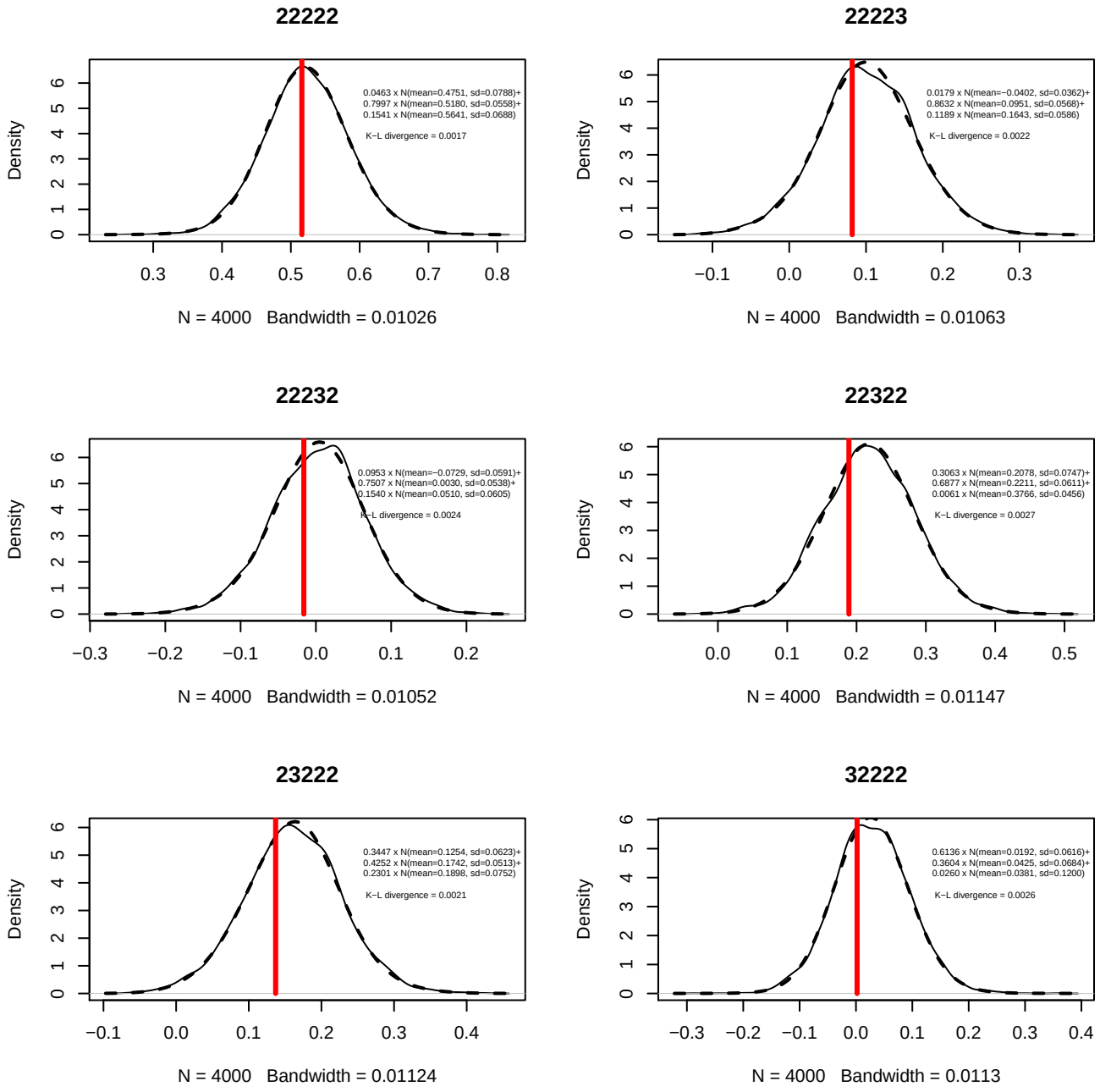


Figure 42: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

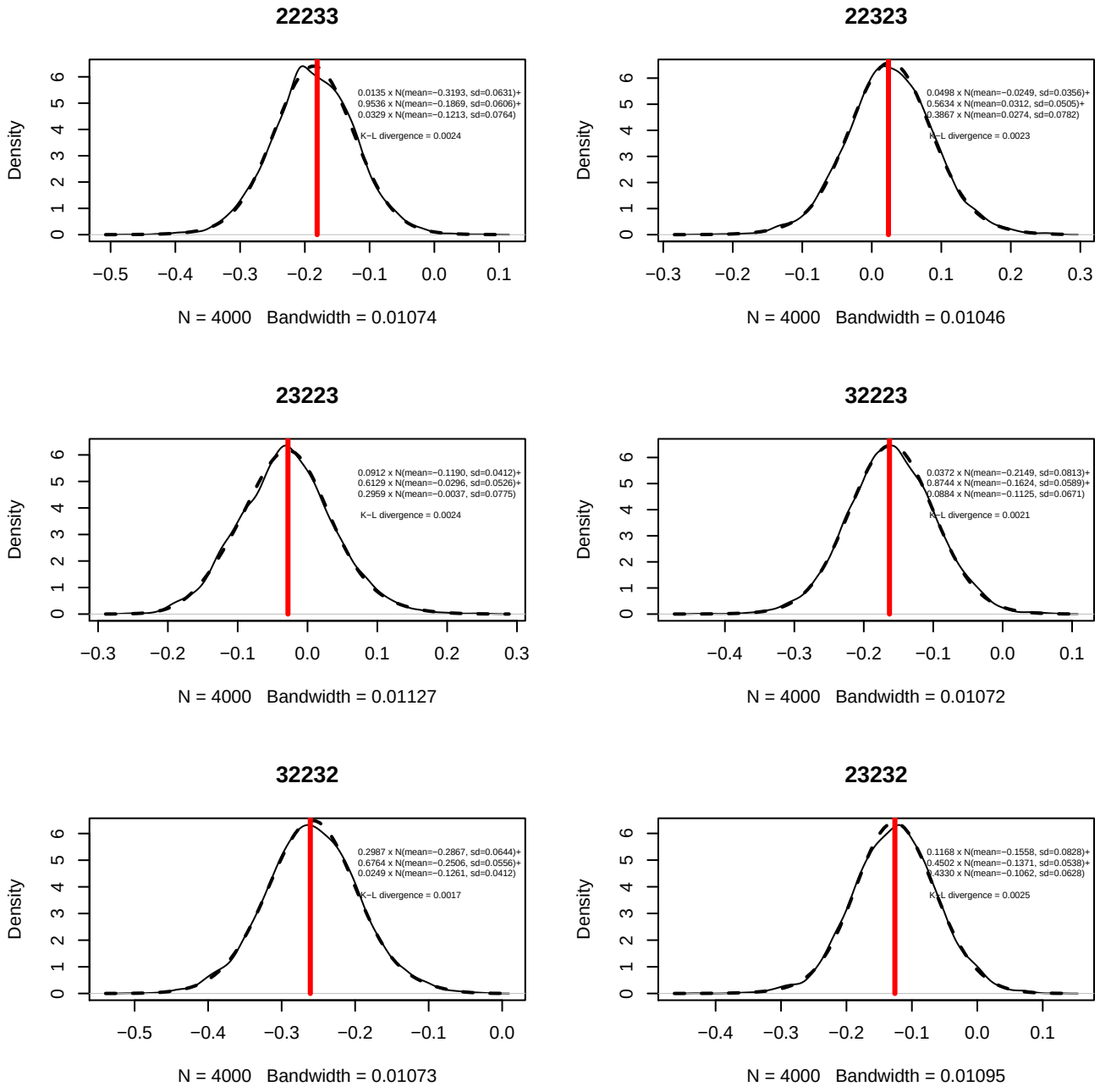


Figure 43: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

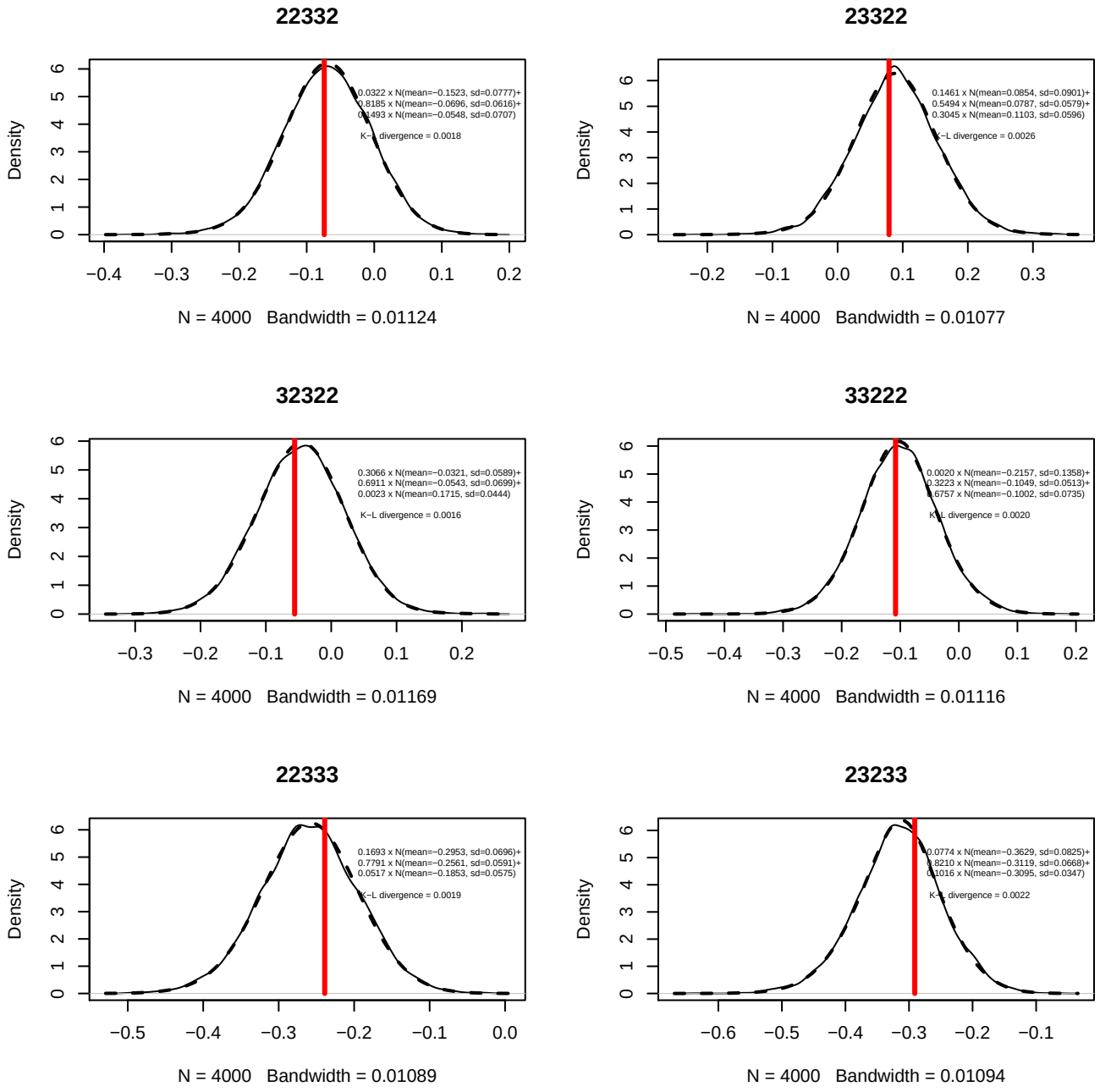


Figure 44: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

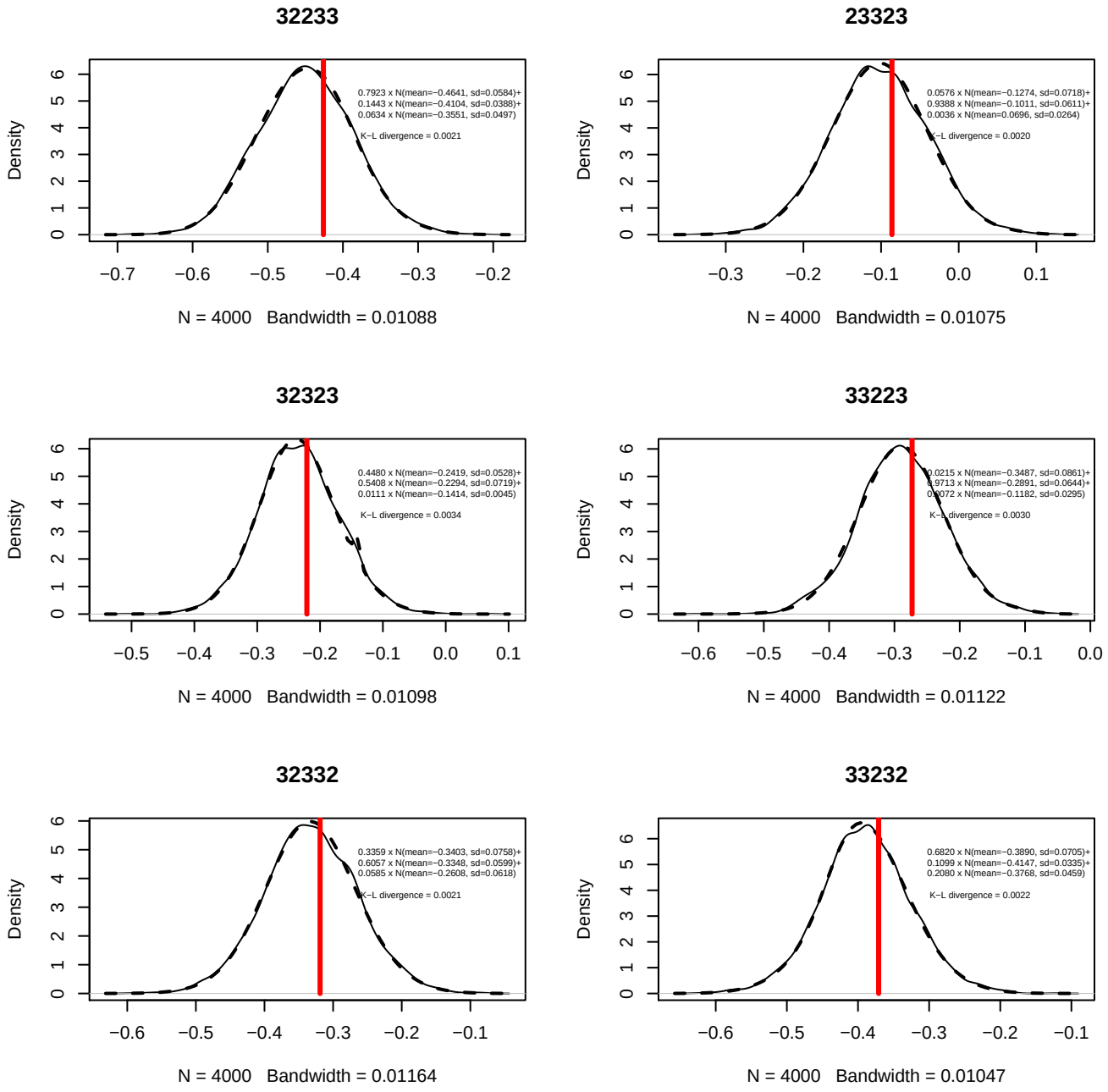


Figure 45: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

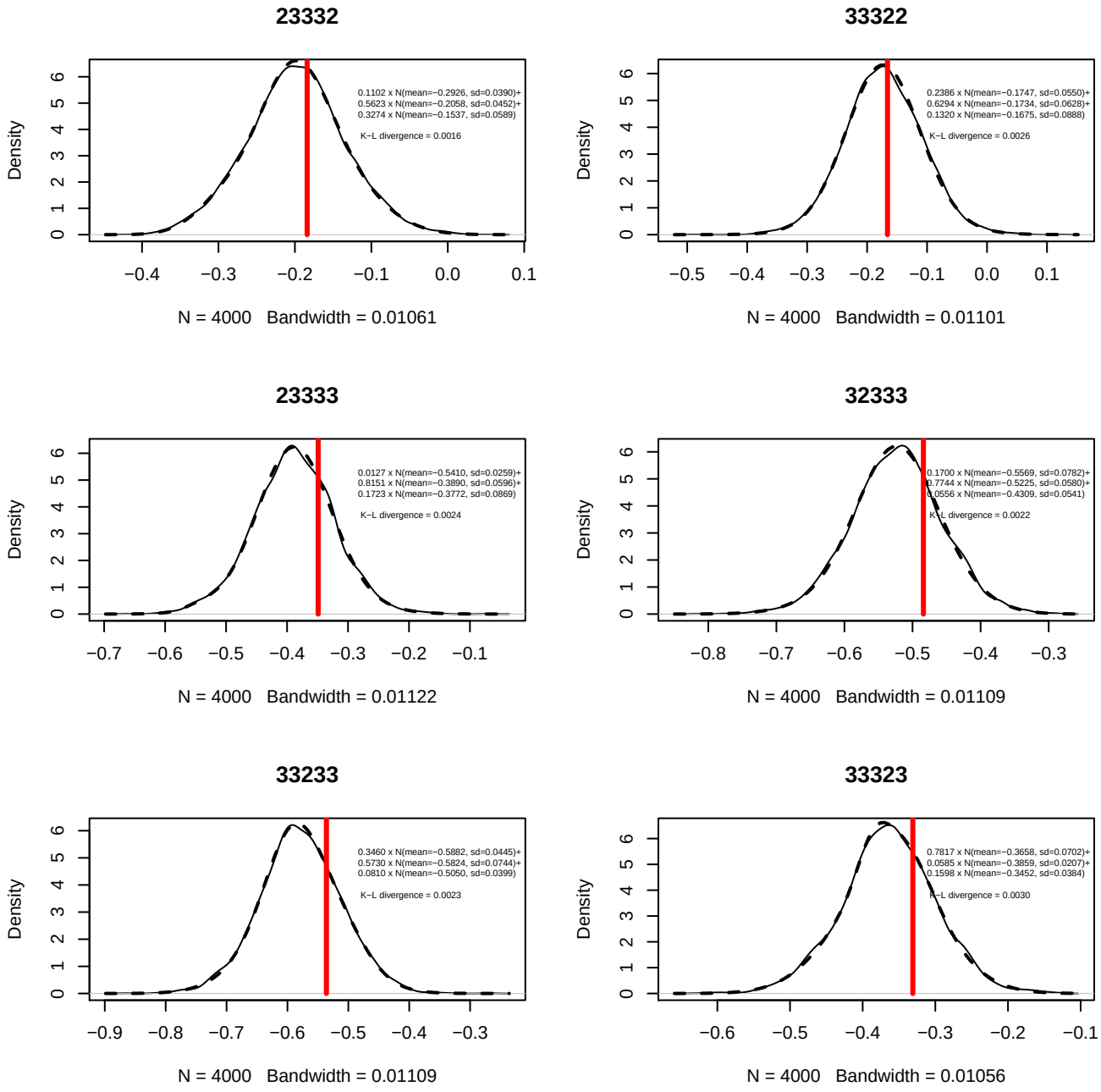


Figure 46: Kernel density plots for the MCMC simulations (solid lines) of six EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

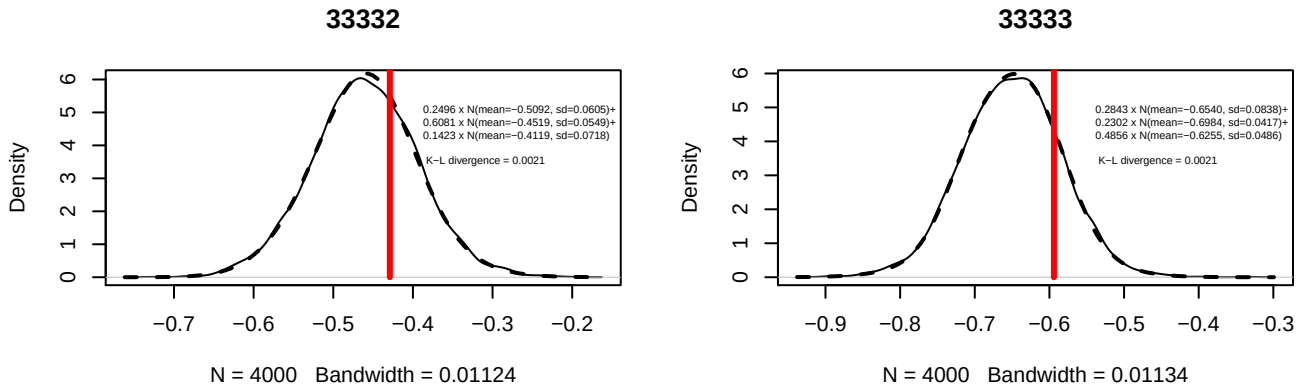


Figure 47: Kernel density plots for the MCMC simulations (solid lines) of two EQ-5D-3L states and the superimposed probability density functions (dotted lines) of their approximation as three-component mixtures of normals. The expected utility of each state based on the MVH project is denoted as vertical red line.

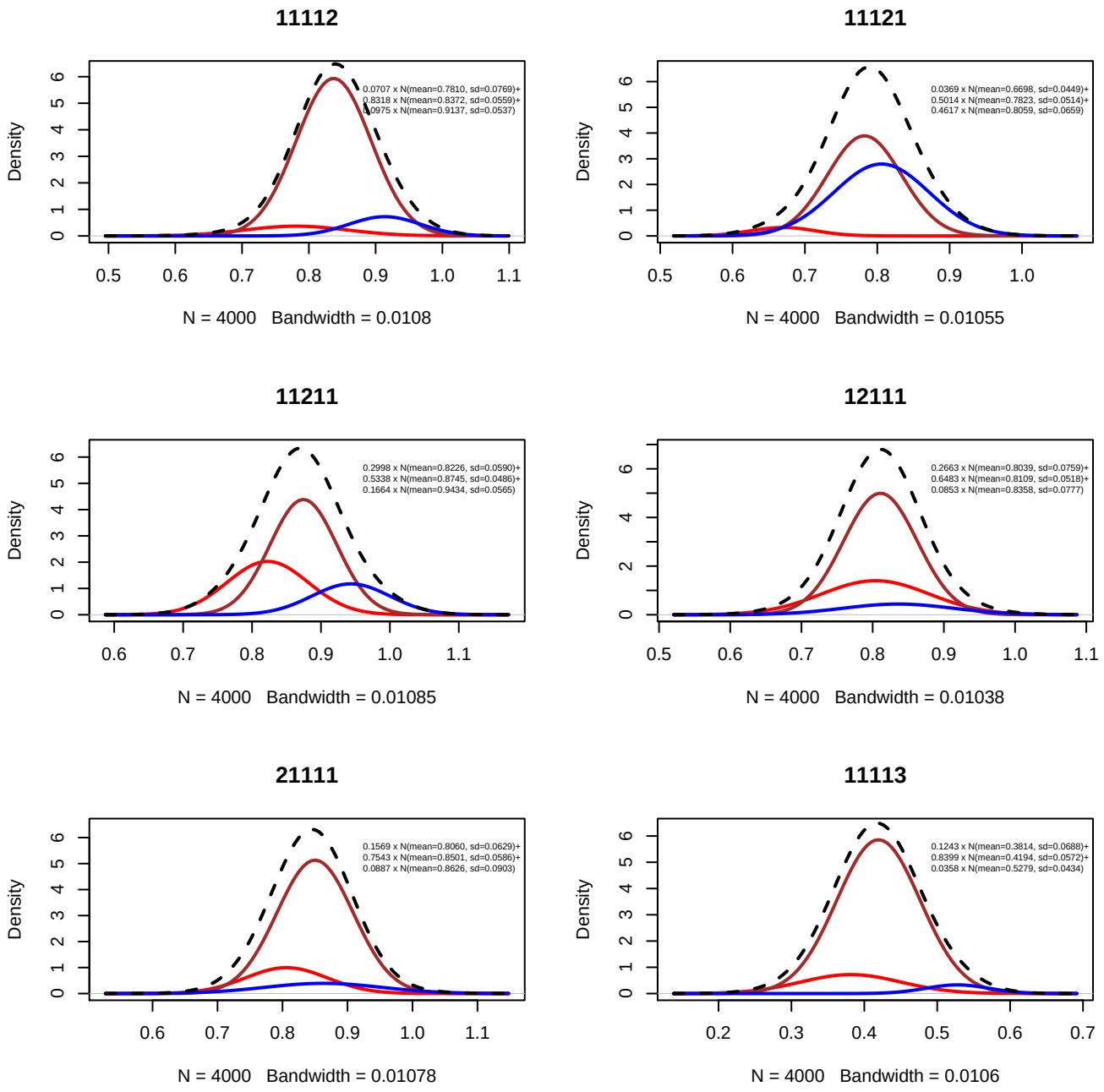


Figure 48: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

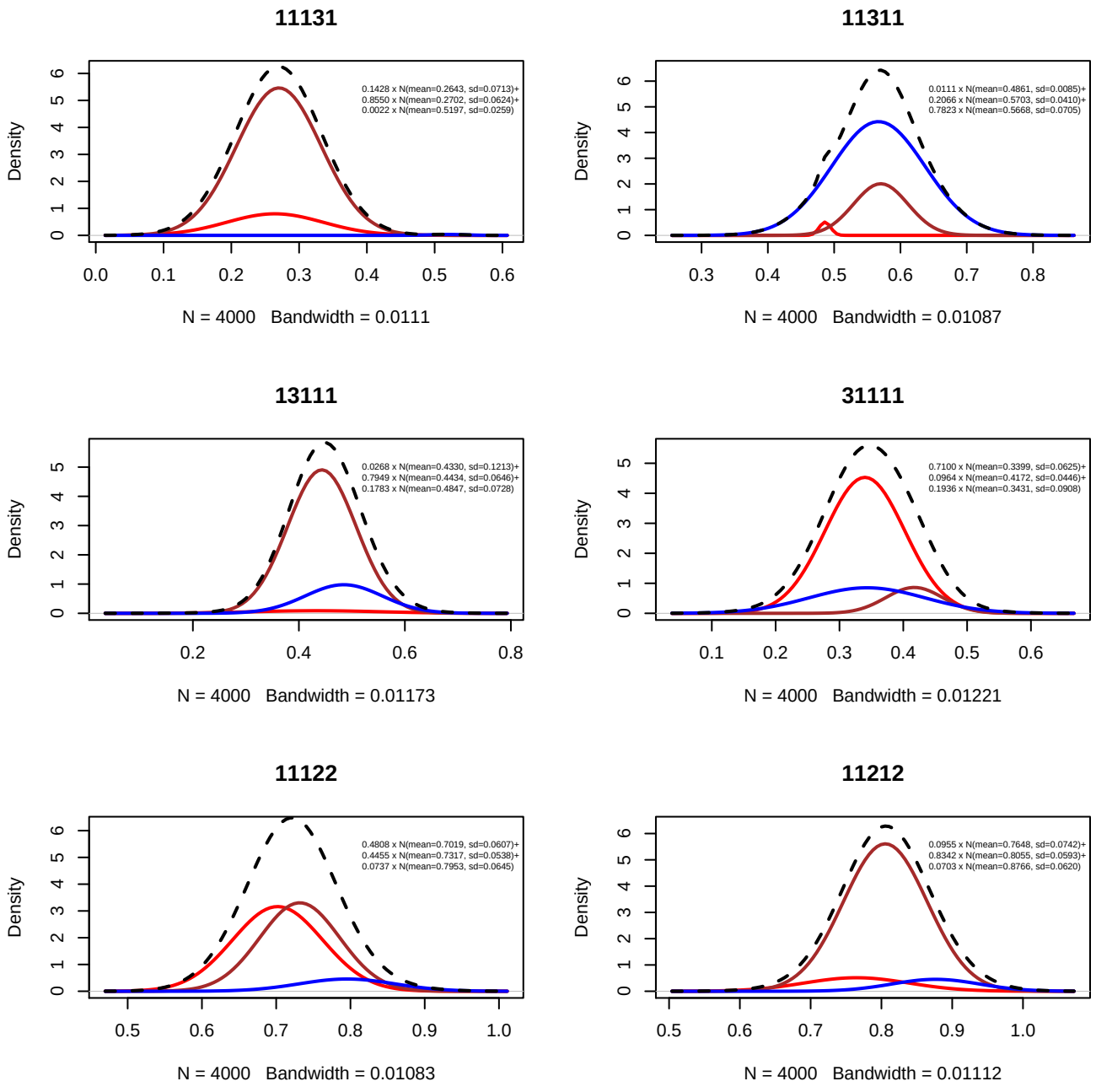


Figure 49: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

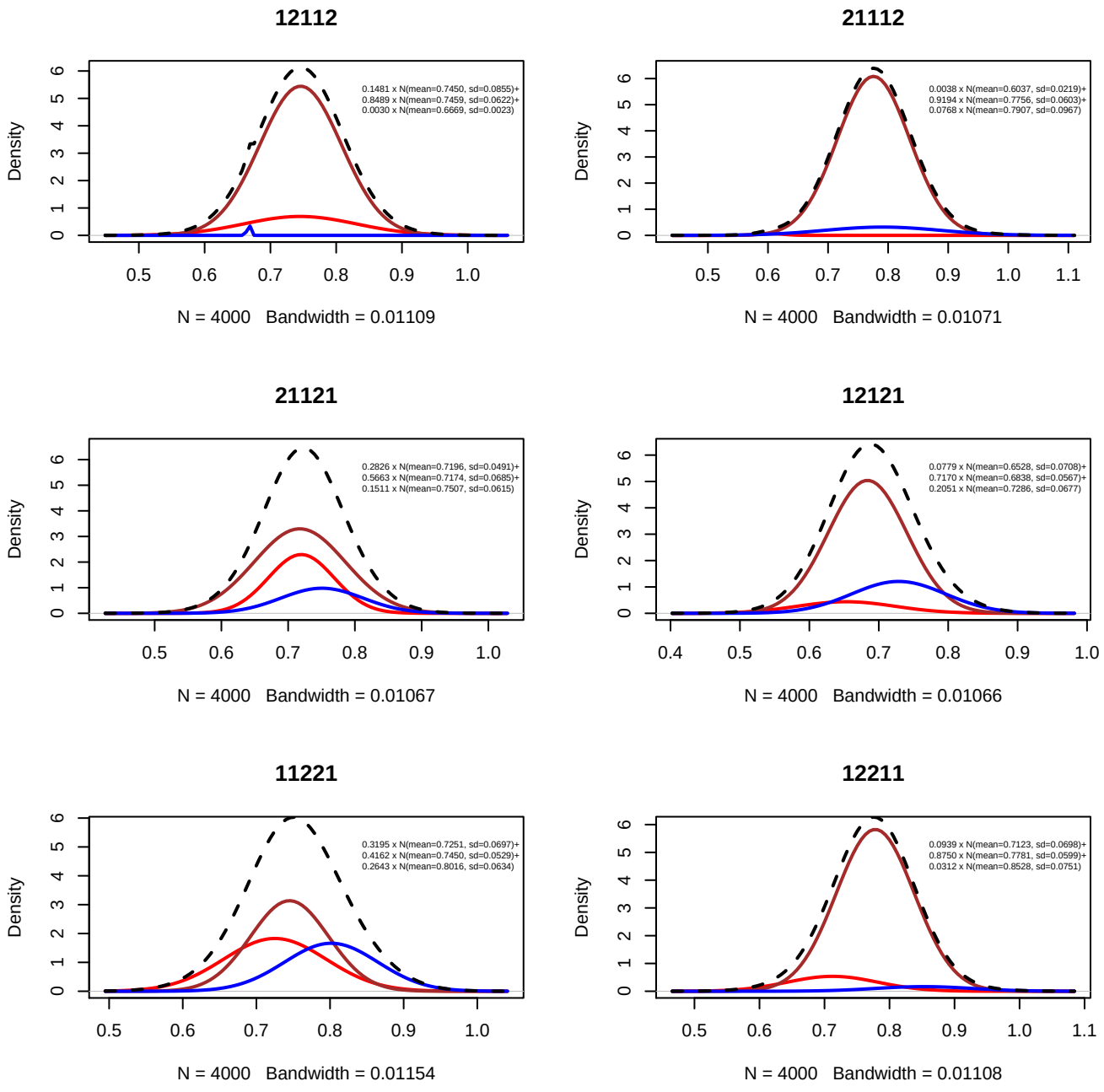


Figure 50: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

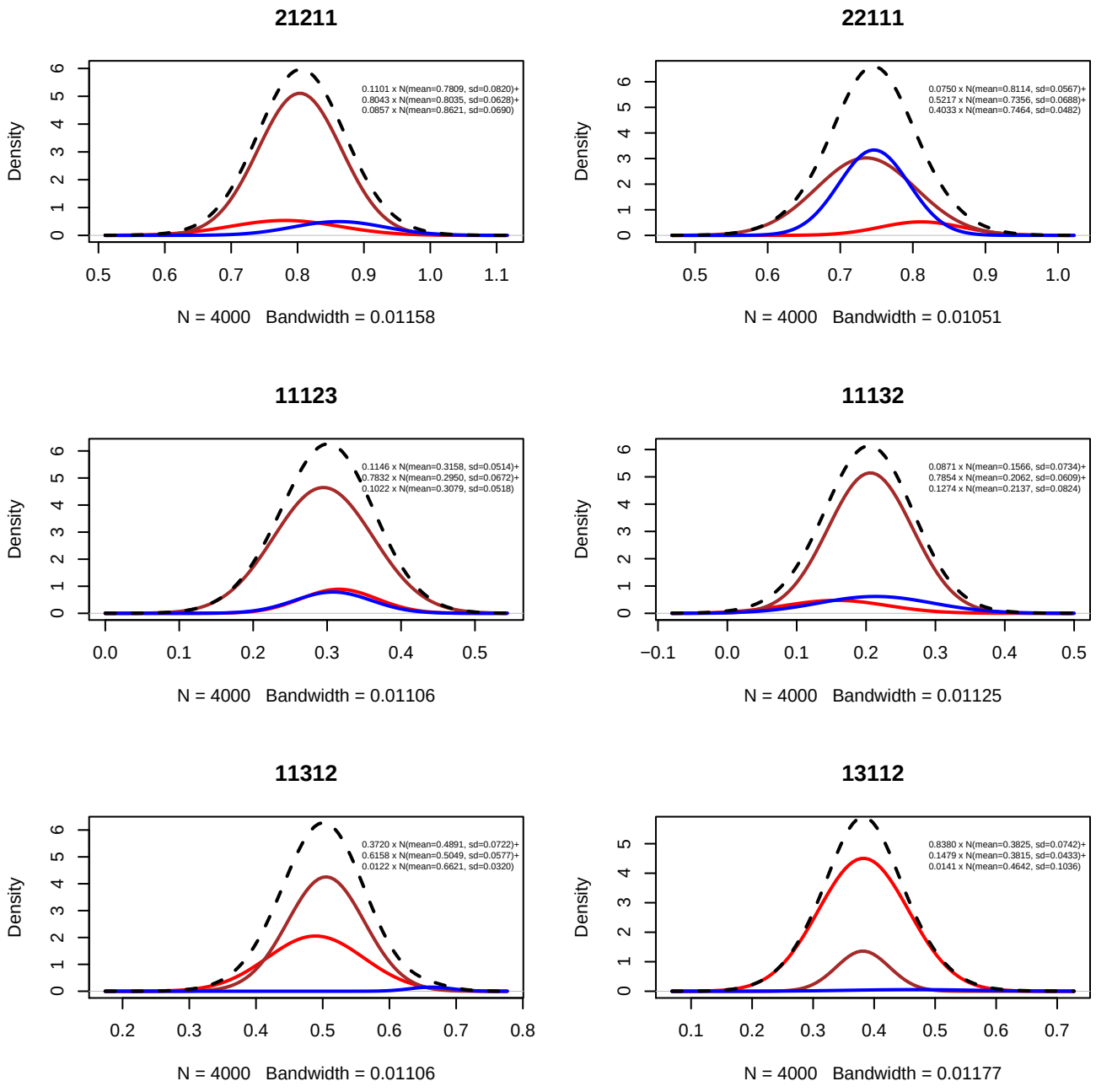


Figure 51: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

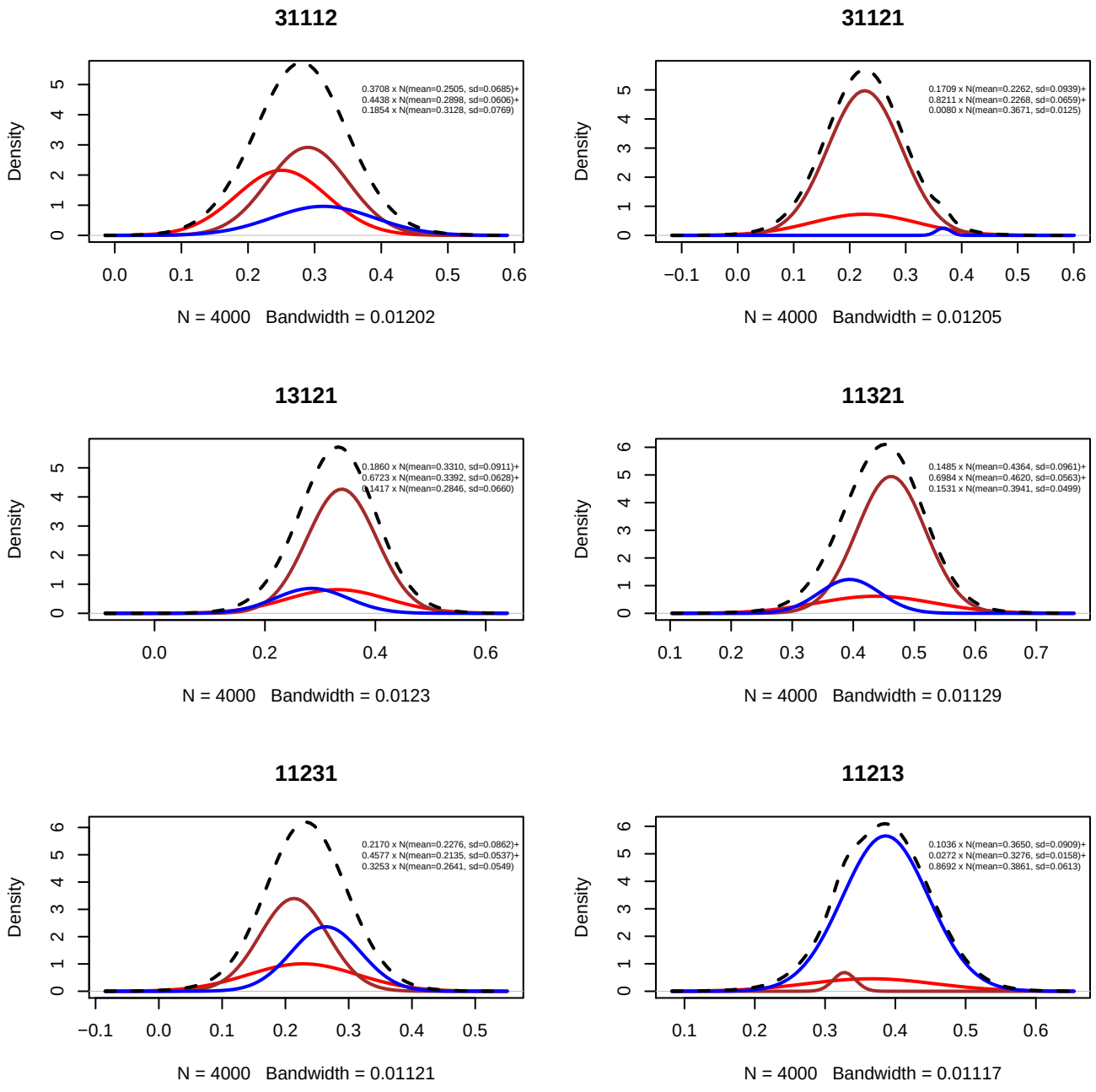


Figure 52: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

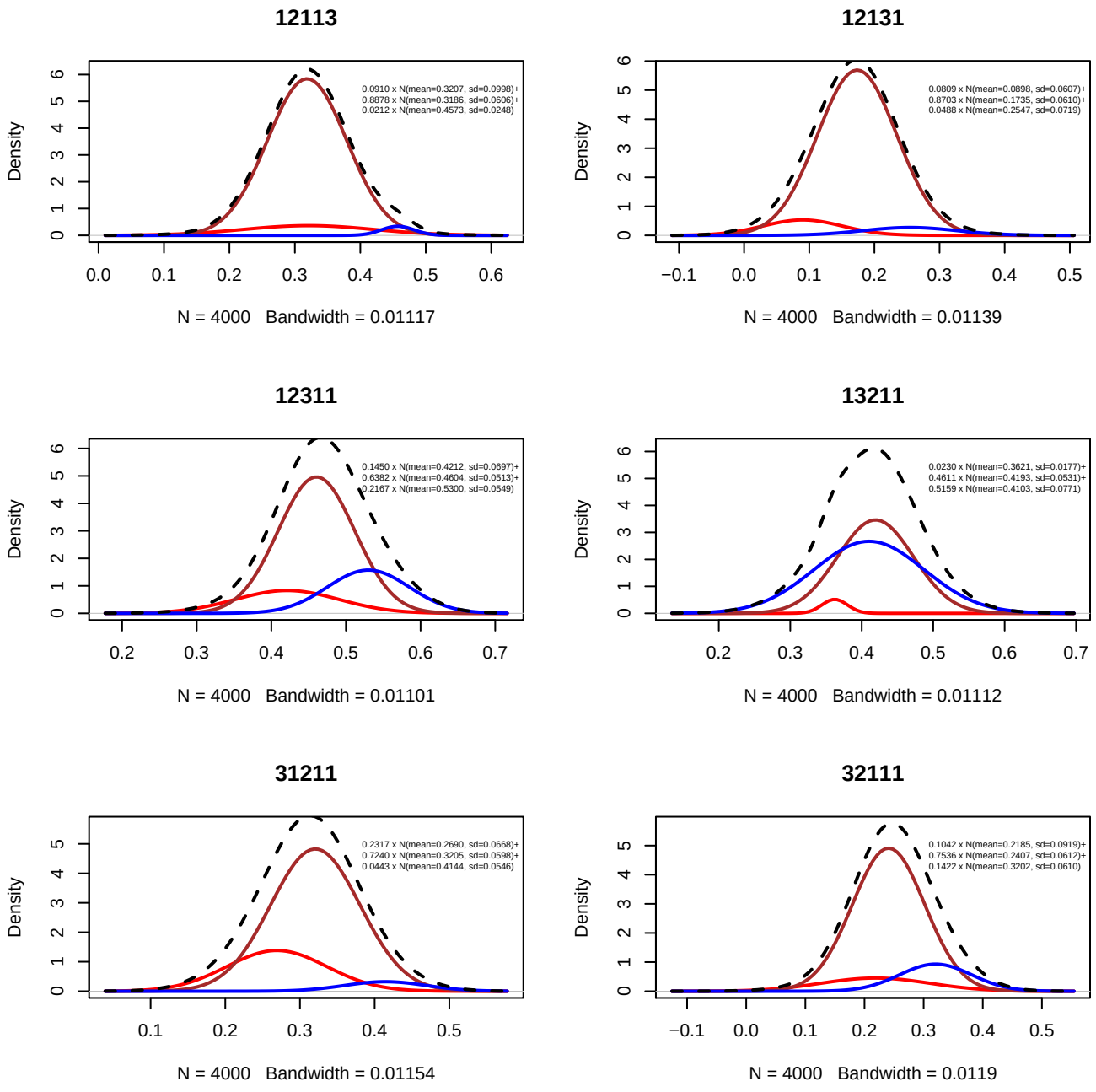


Figure 53: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

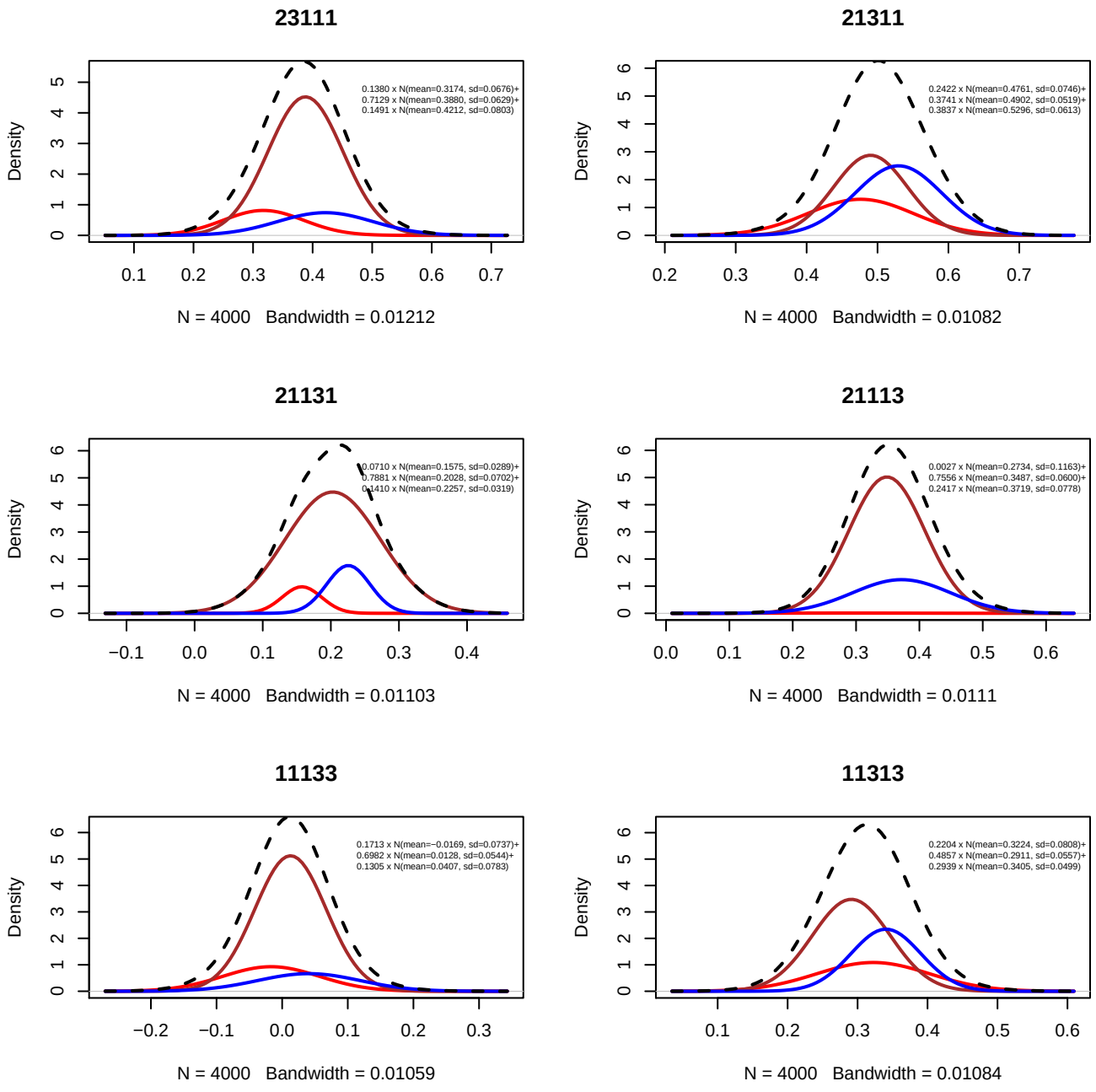


Figure 54: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

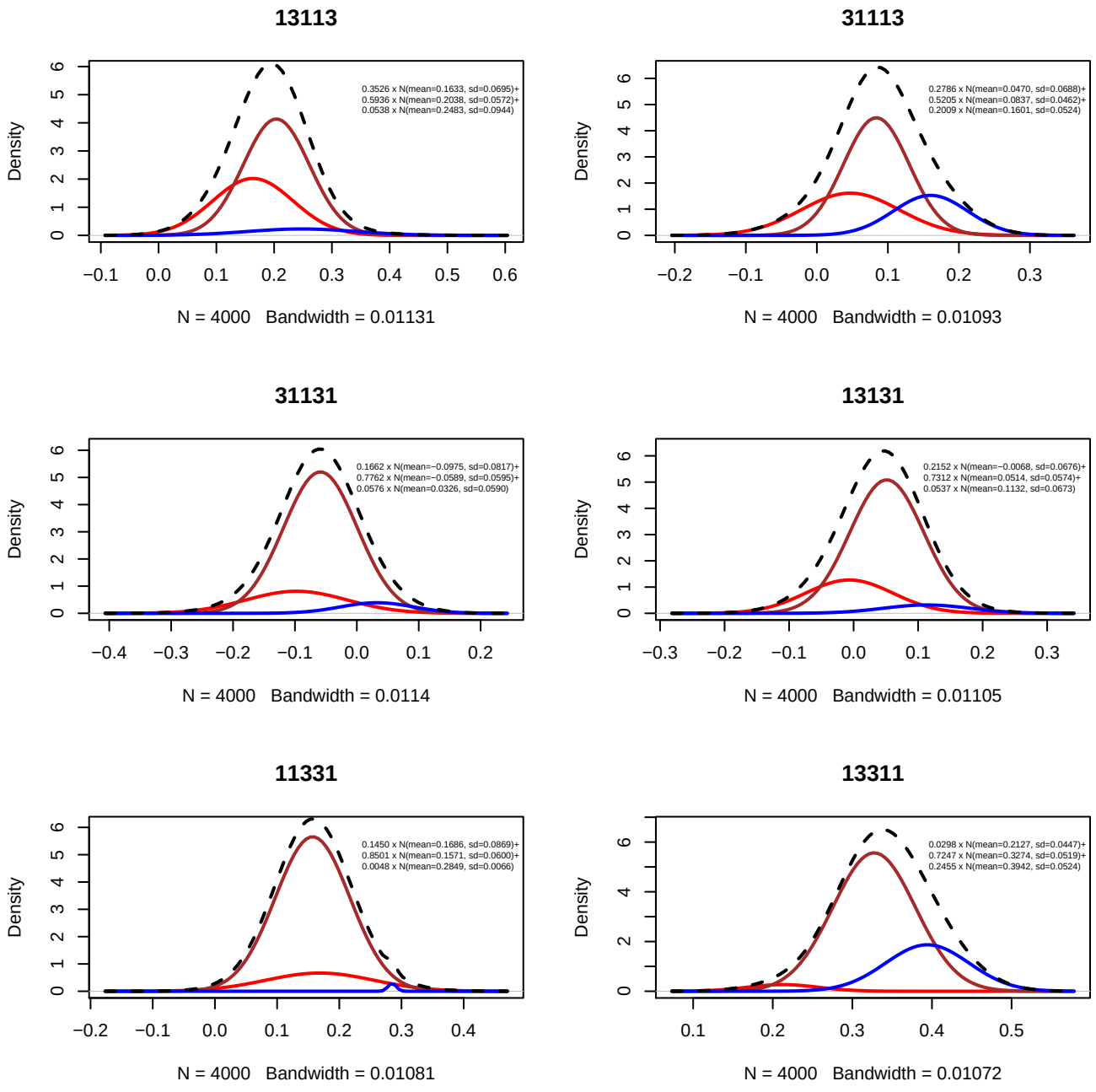


Figure 55: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

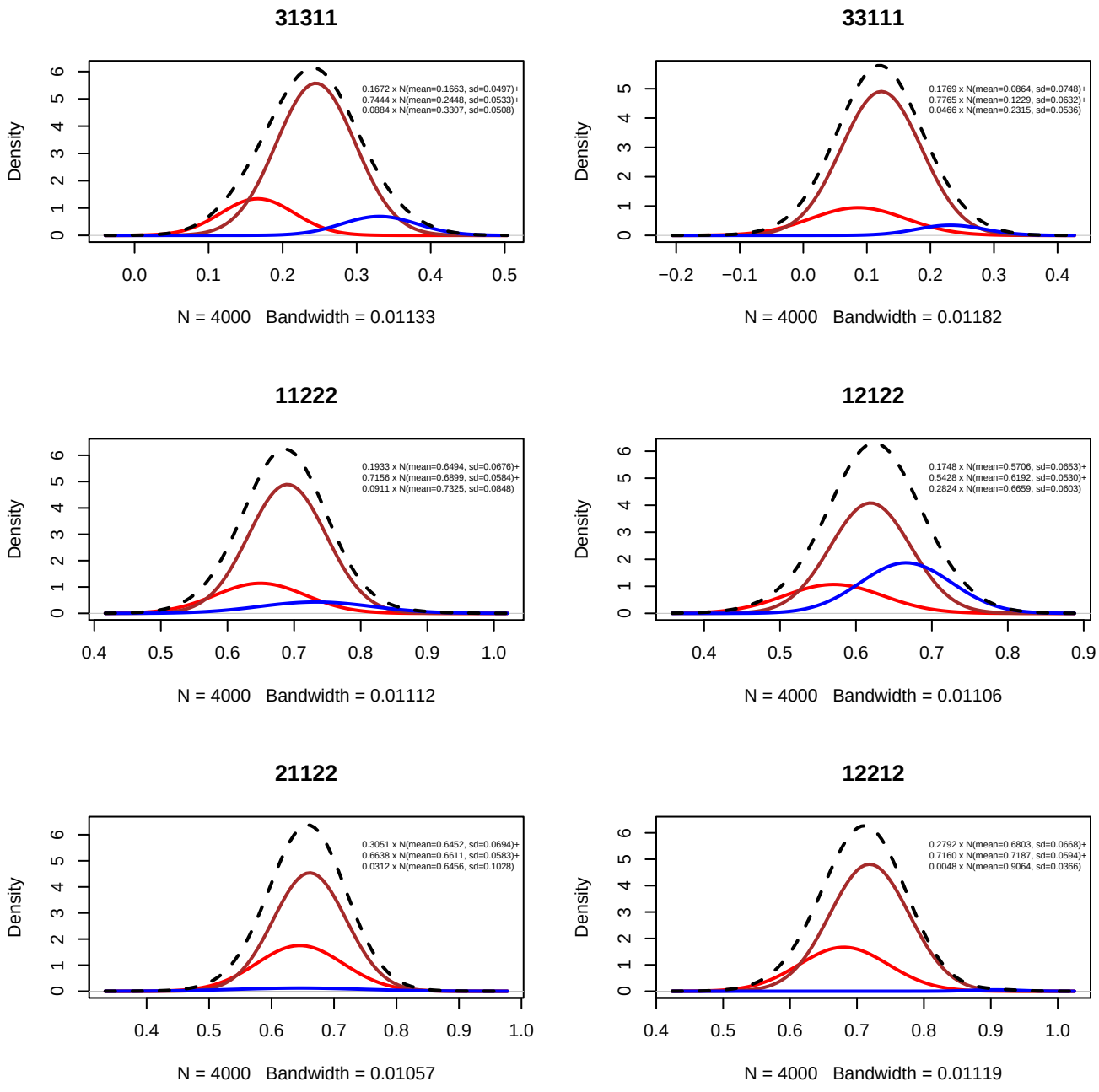


Figure 56: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

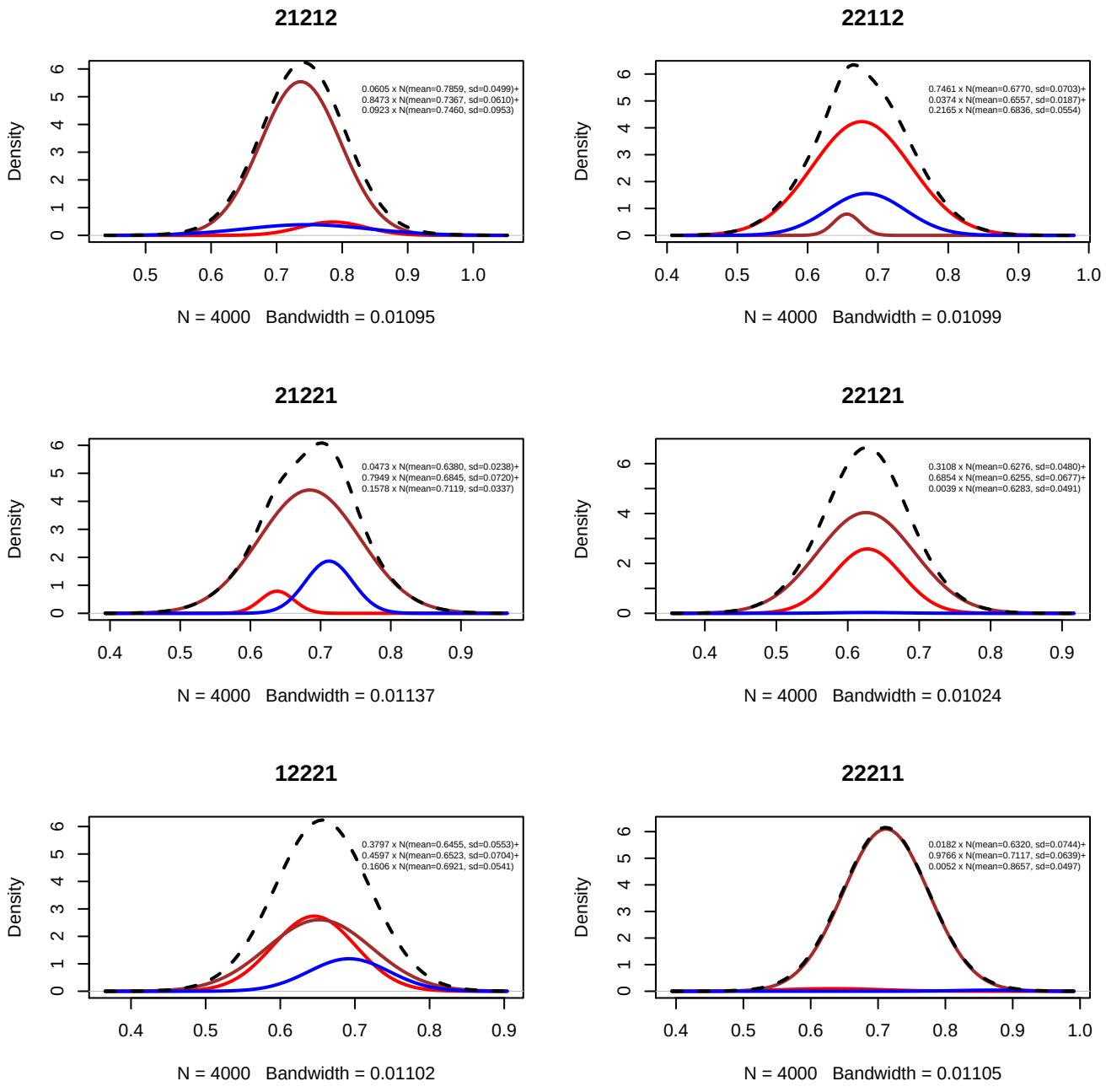


Figure 57: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

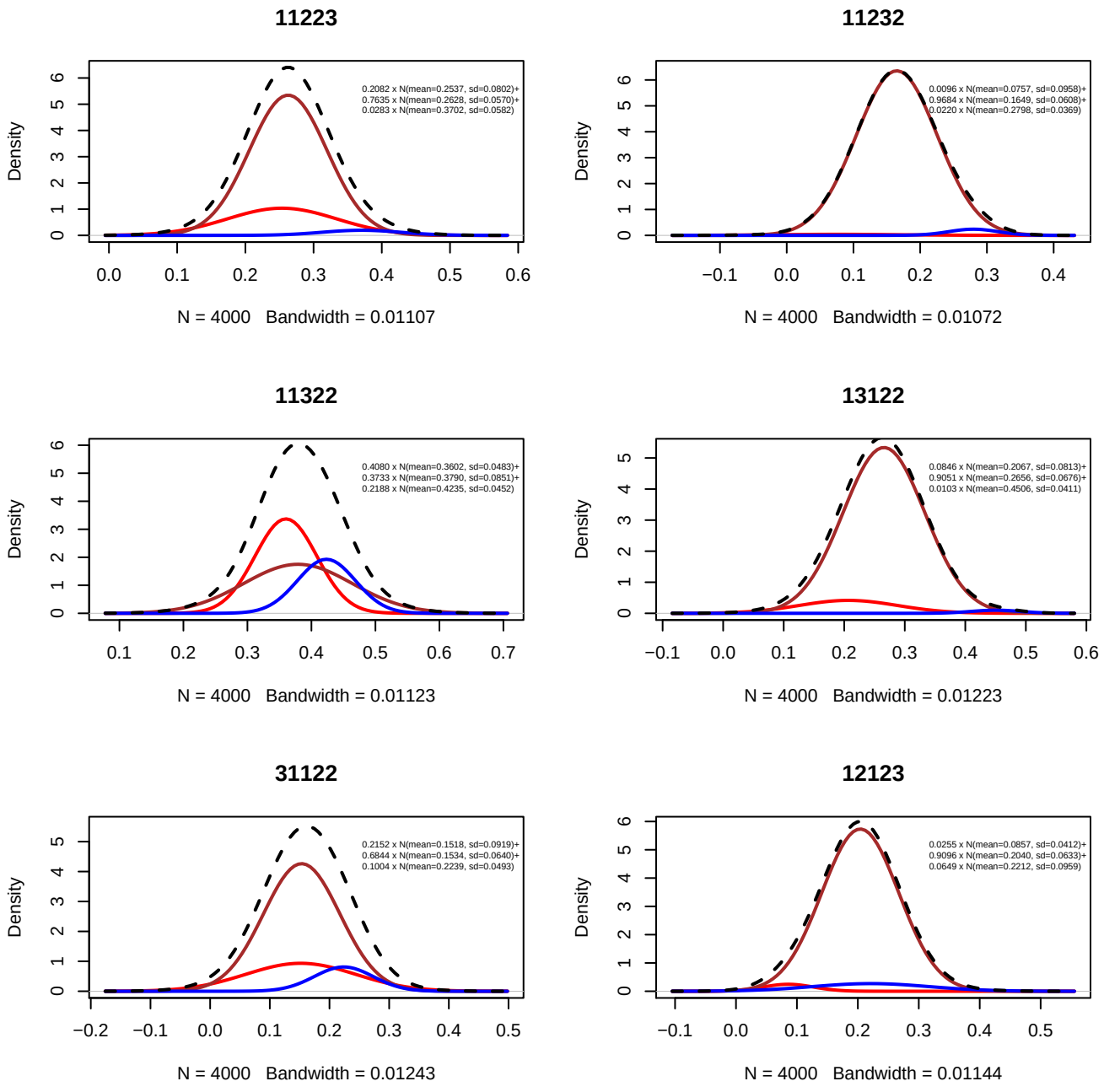


Figure 58: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

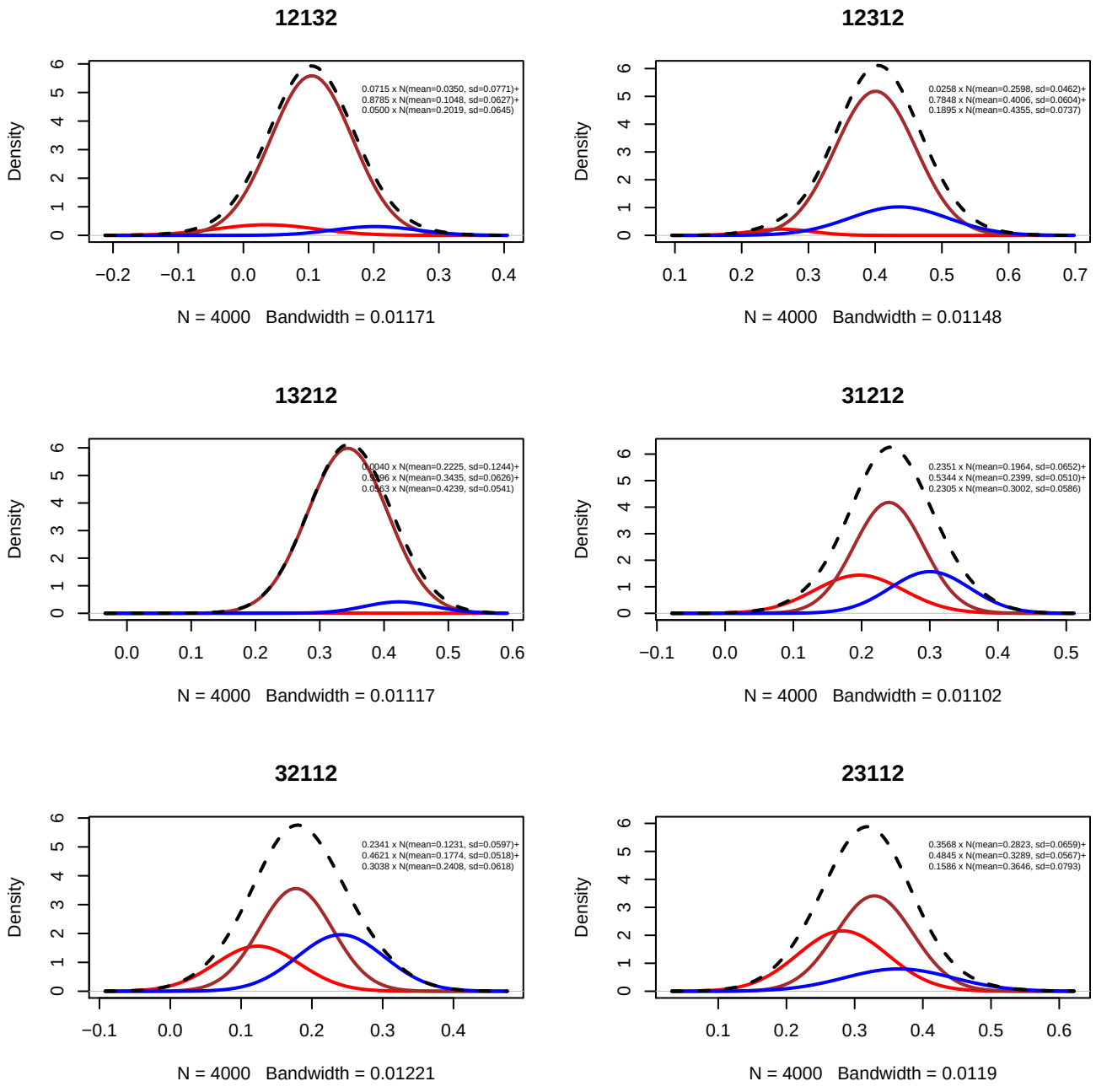


Figure 59: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

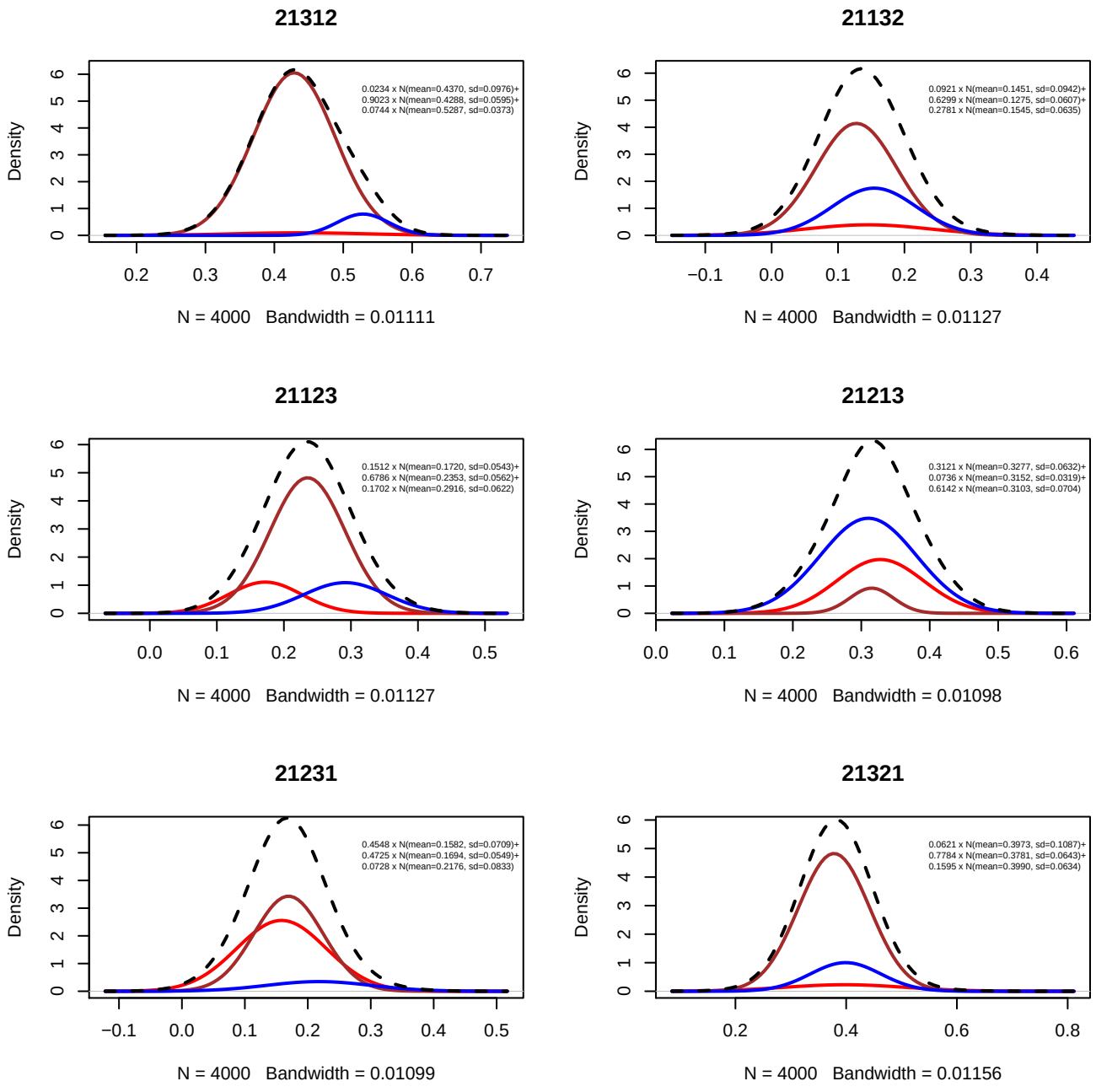


Figure 60: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

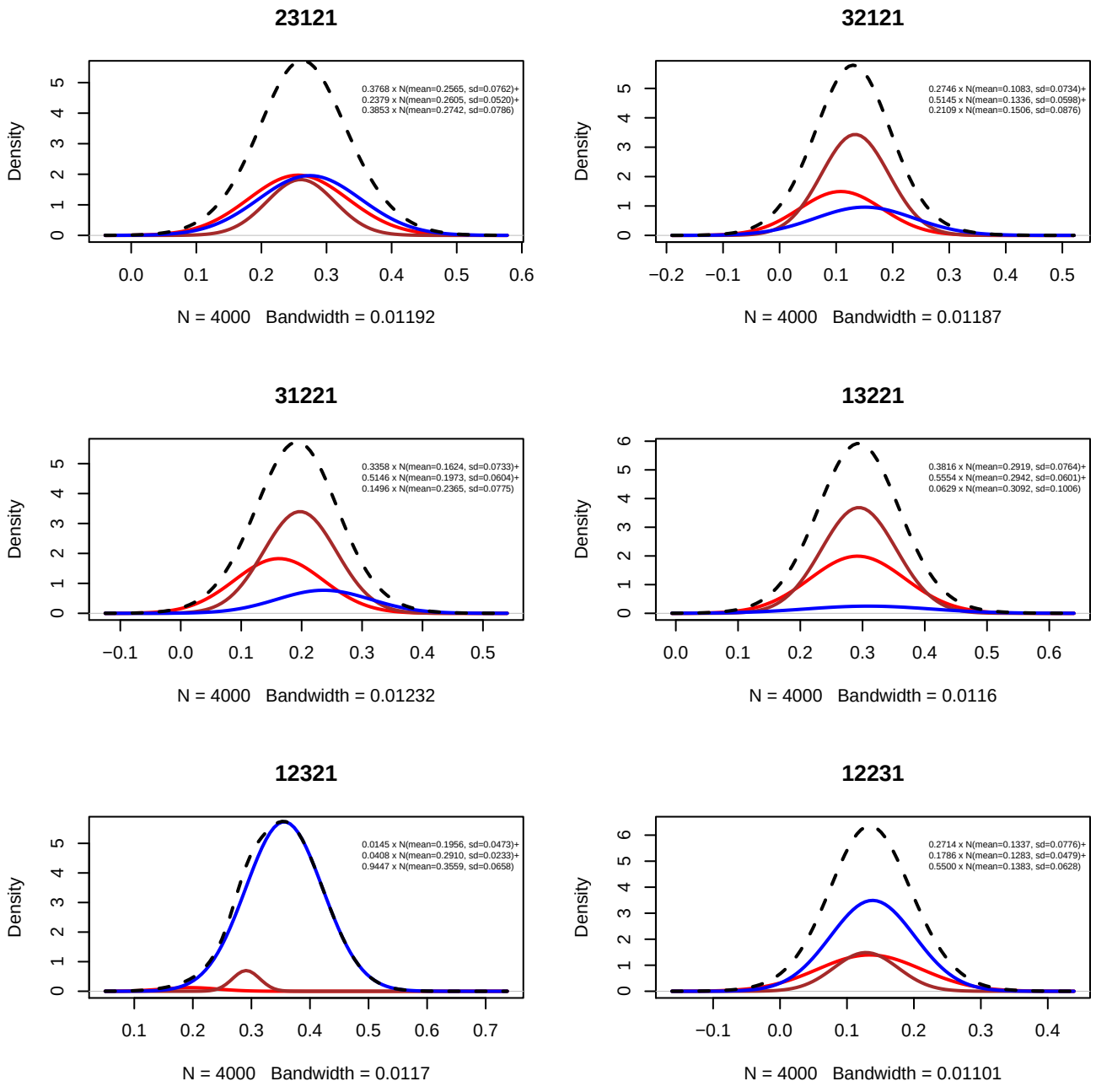


Figure 61: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

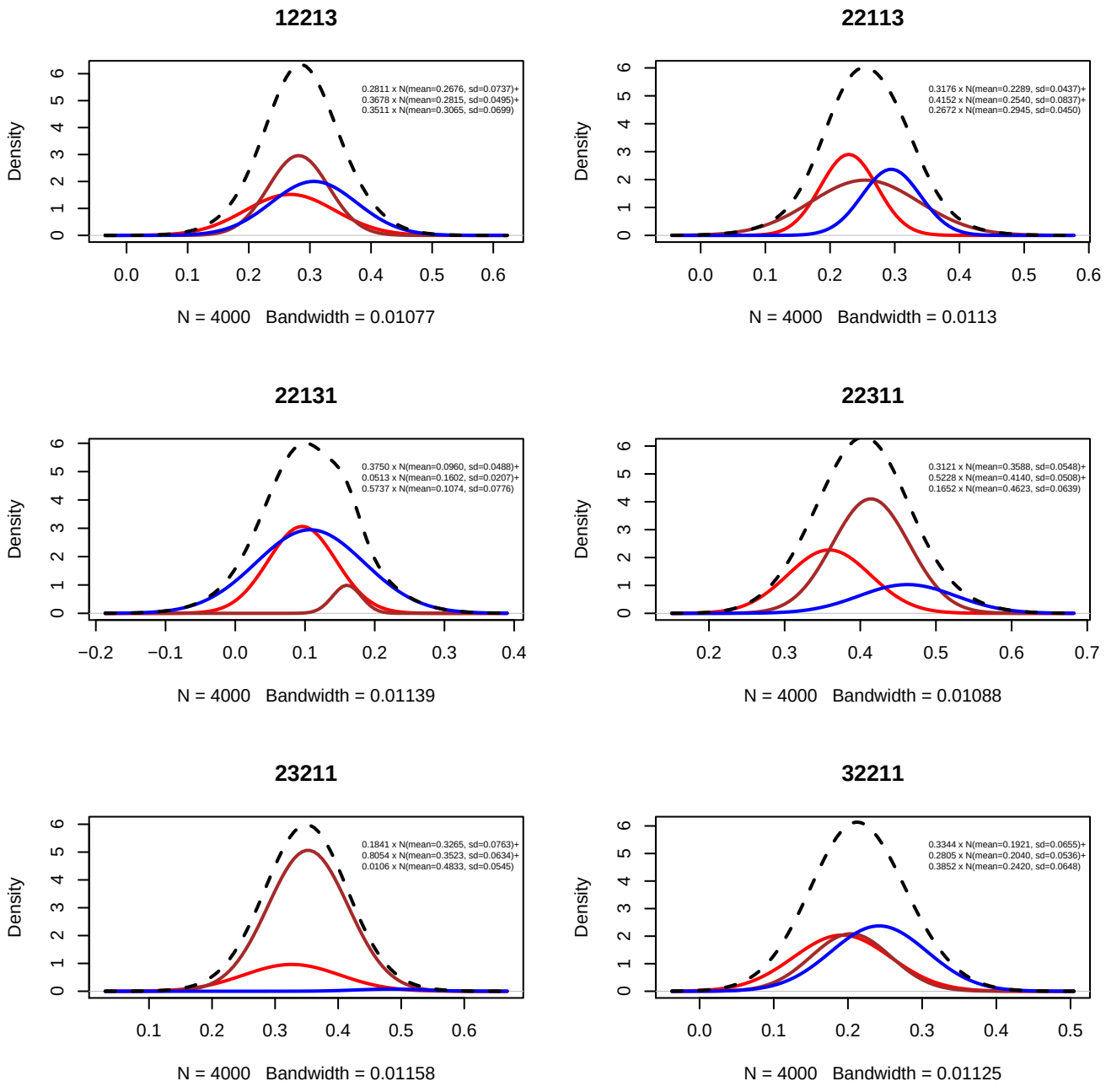


Figure 62: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

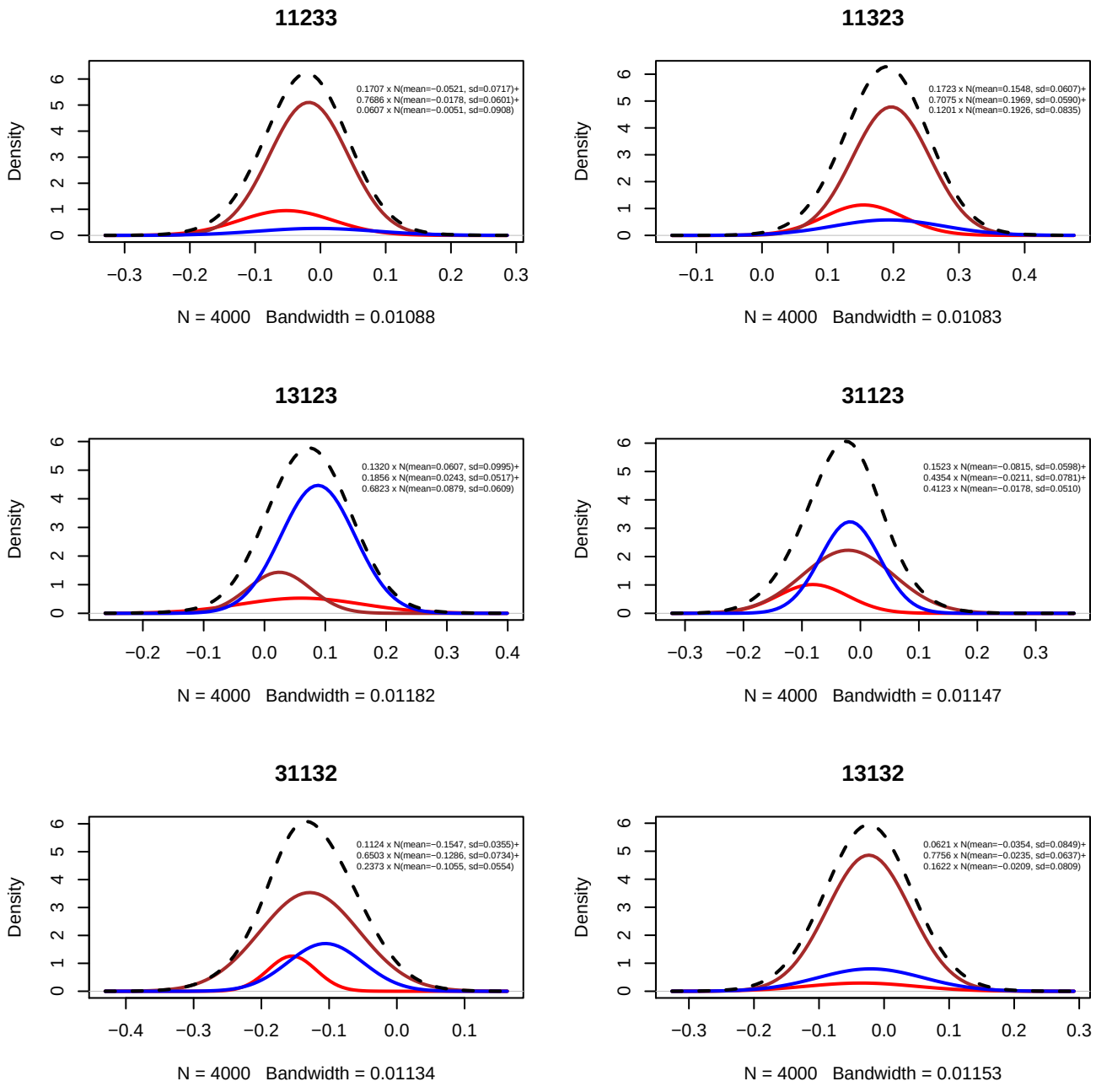


Figure 63: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

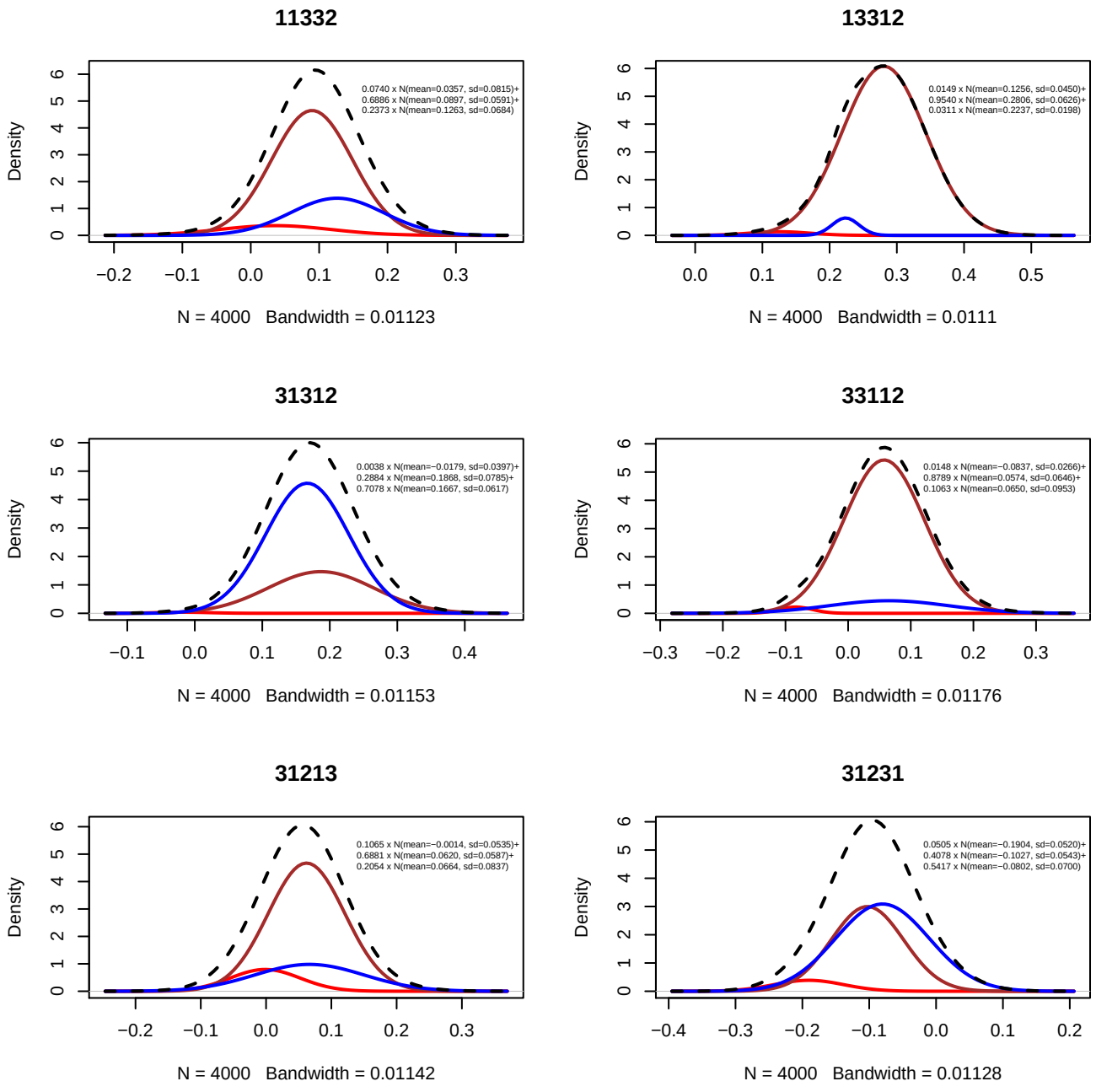


Figure 64: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

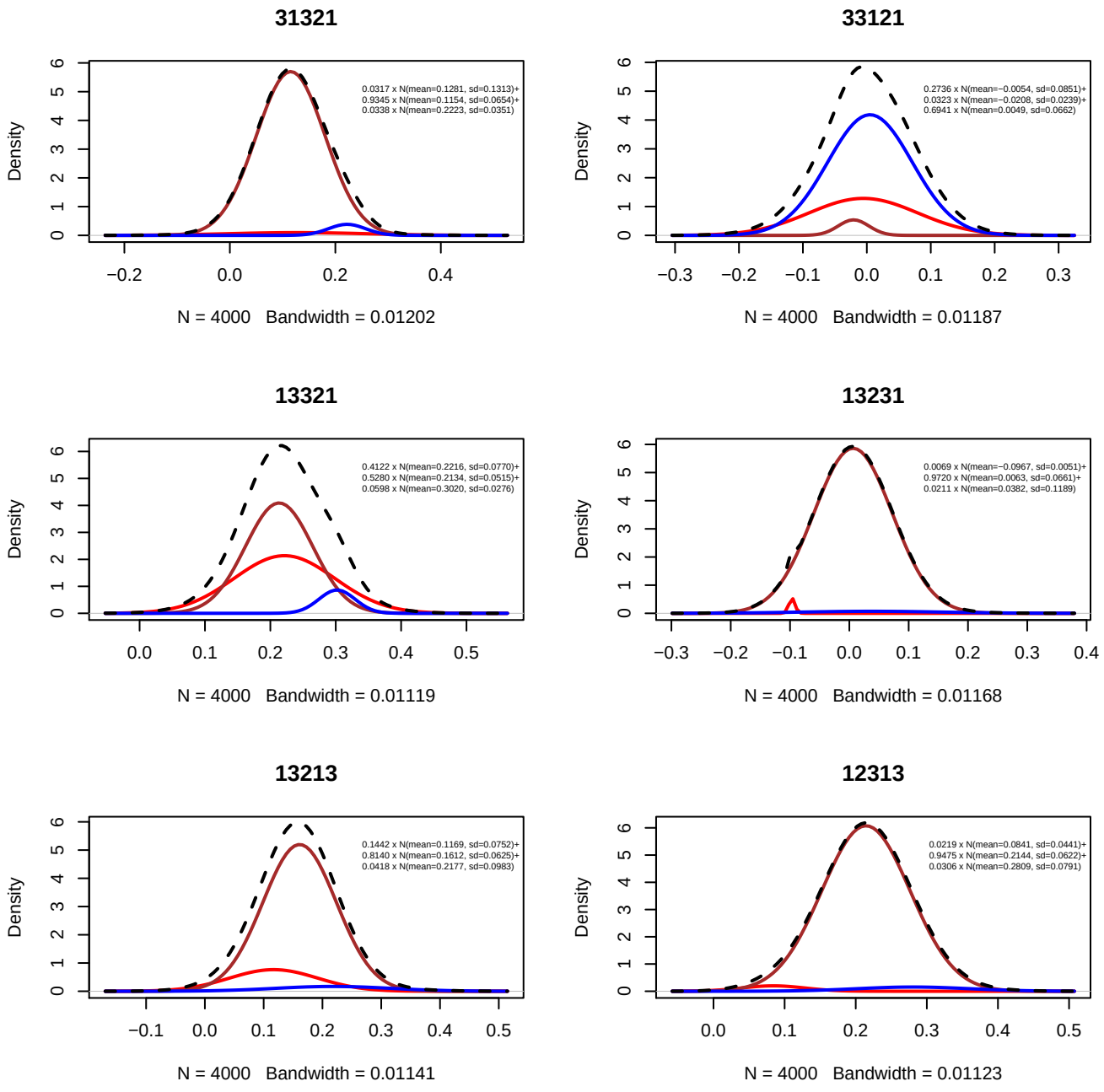


Figure 65: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

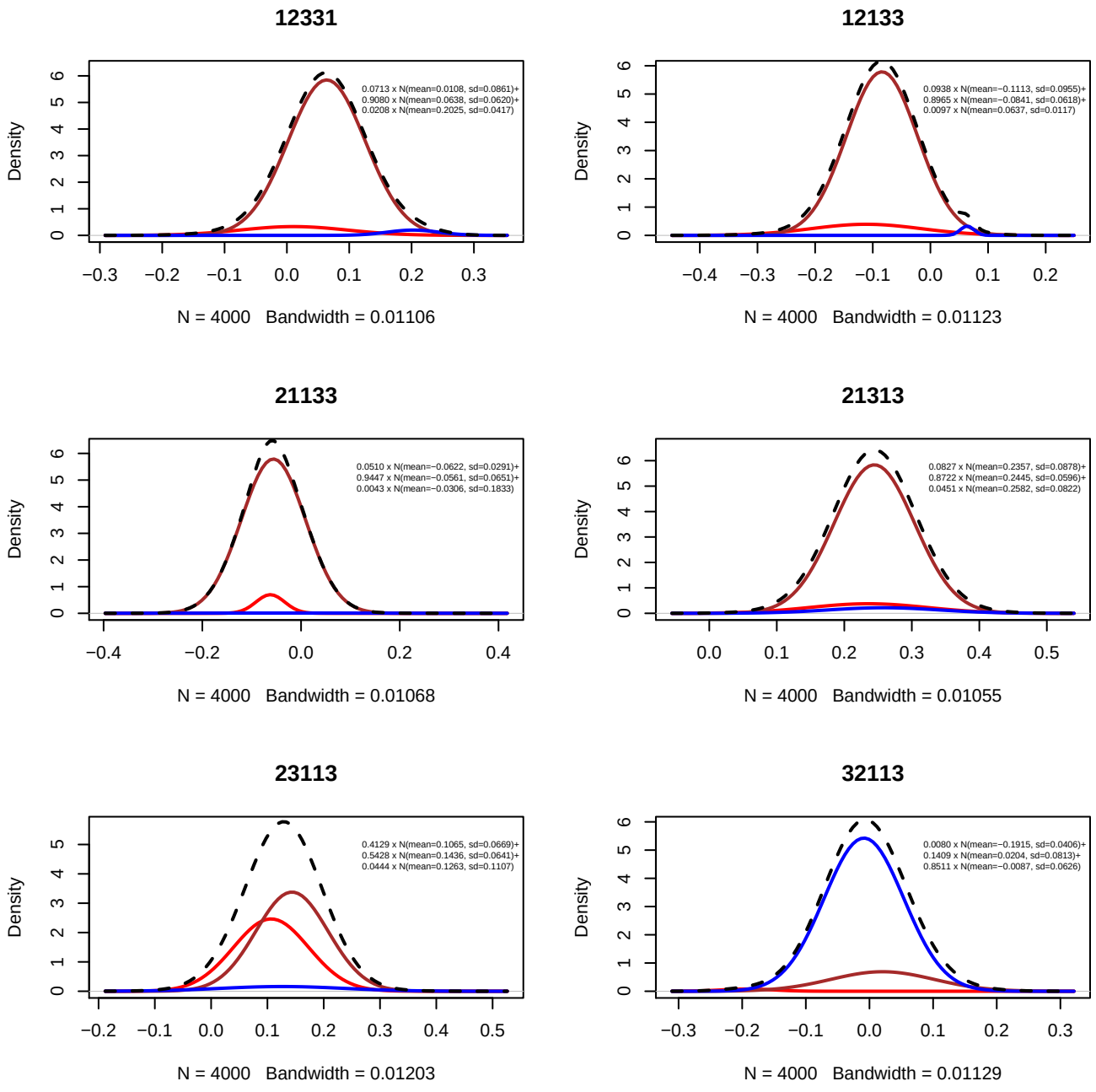


Figure 66: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

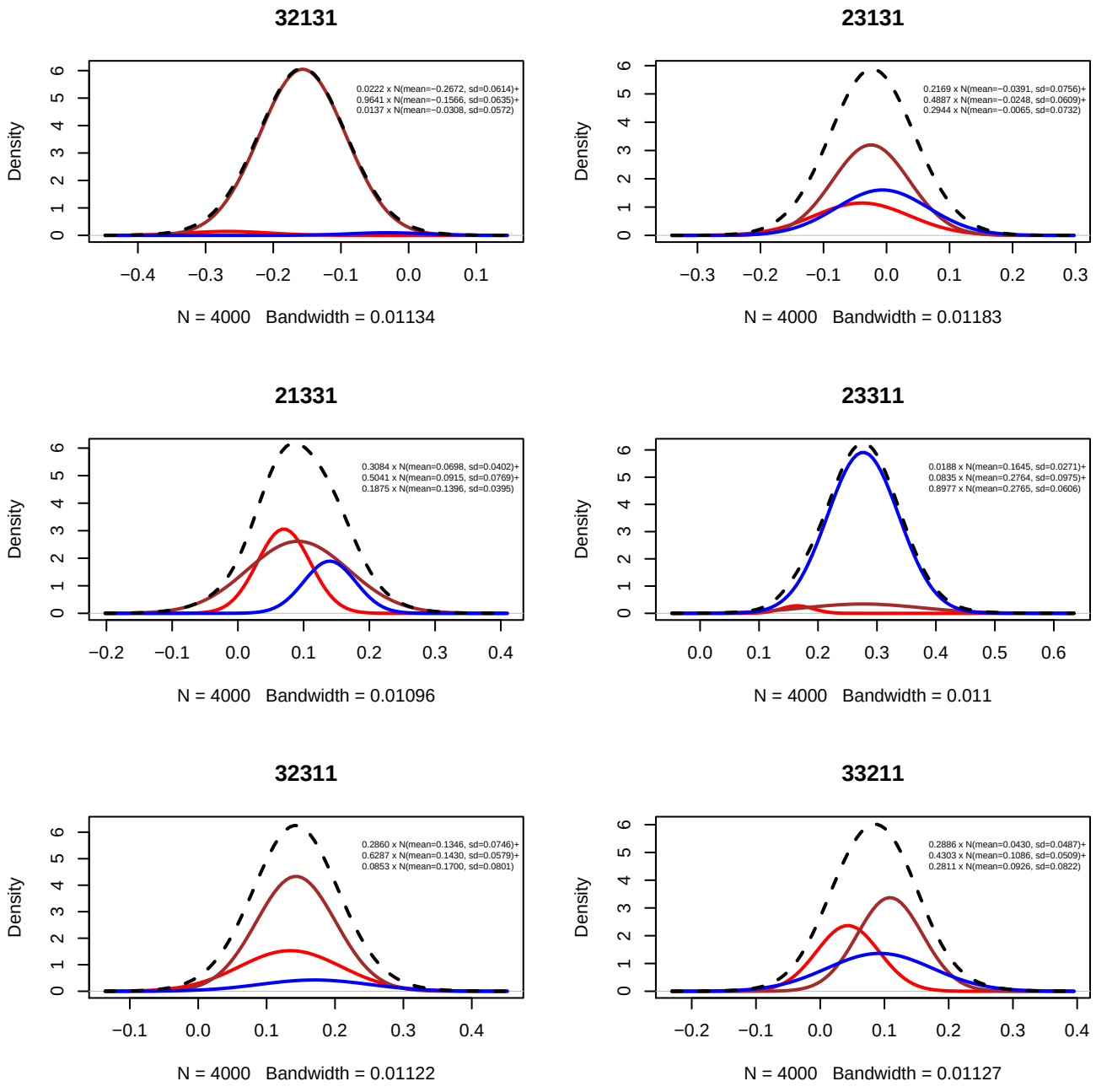


Figure 67: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

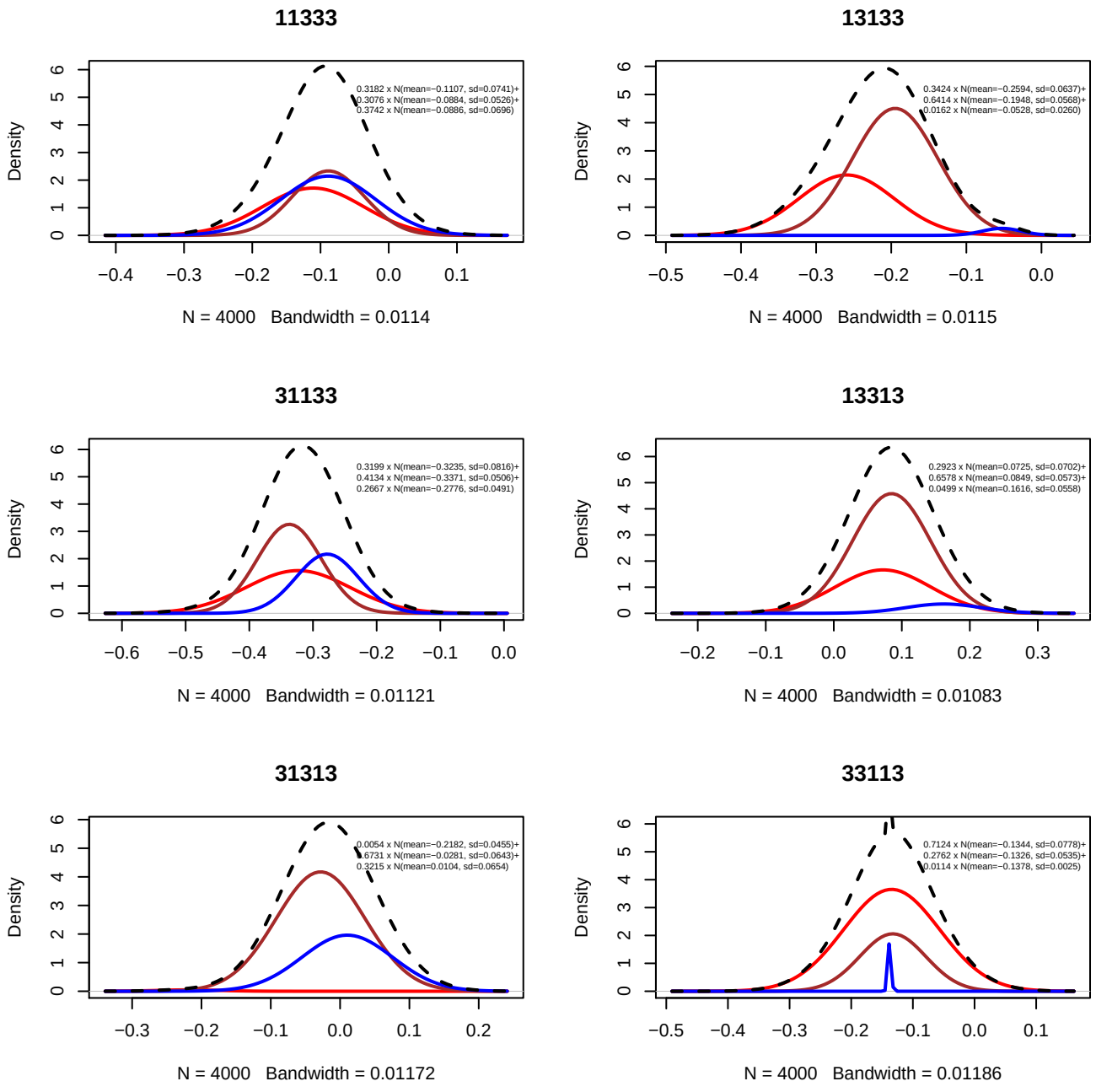


Figure 68: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

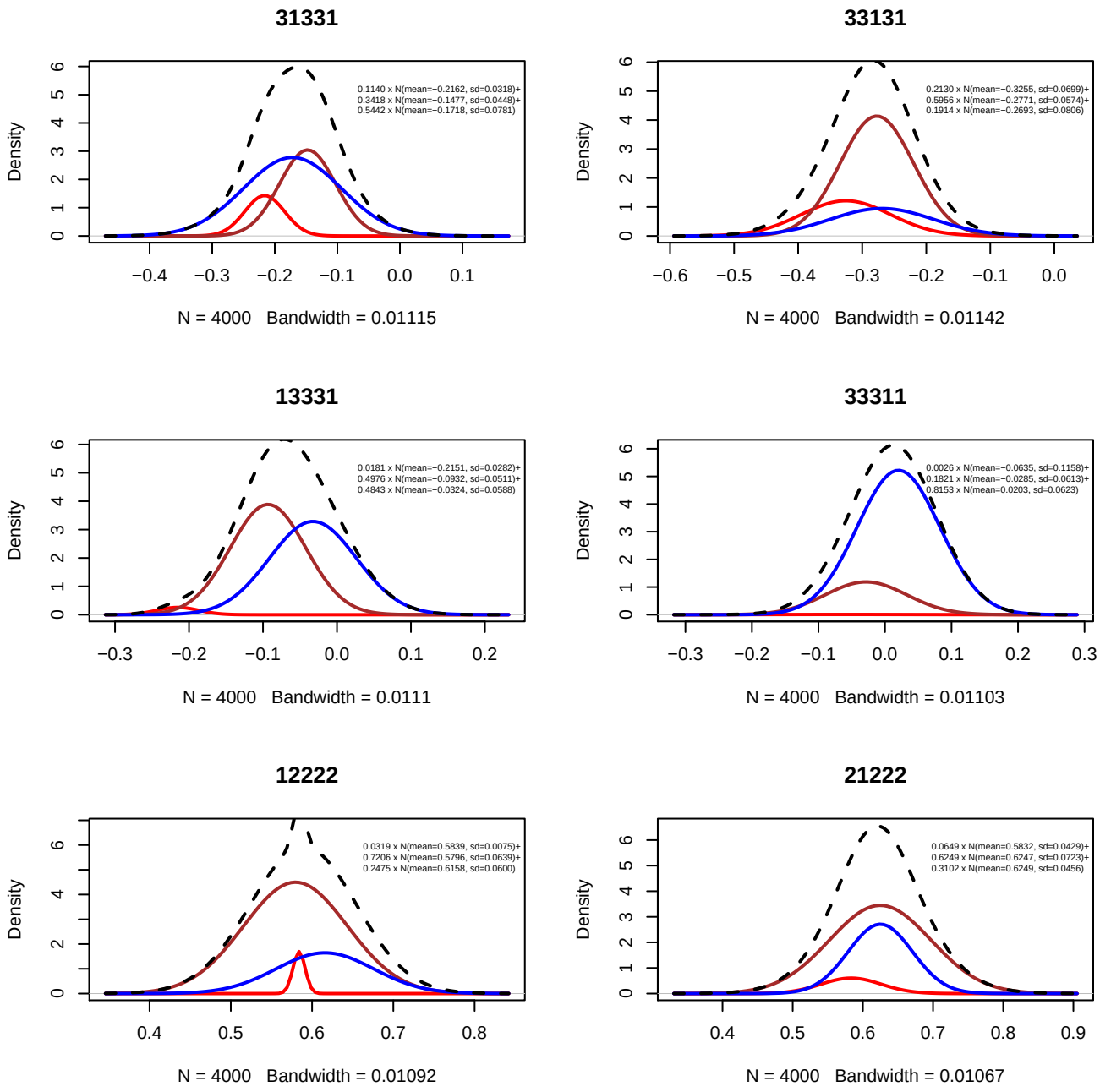


Figure 69: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

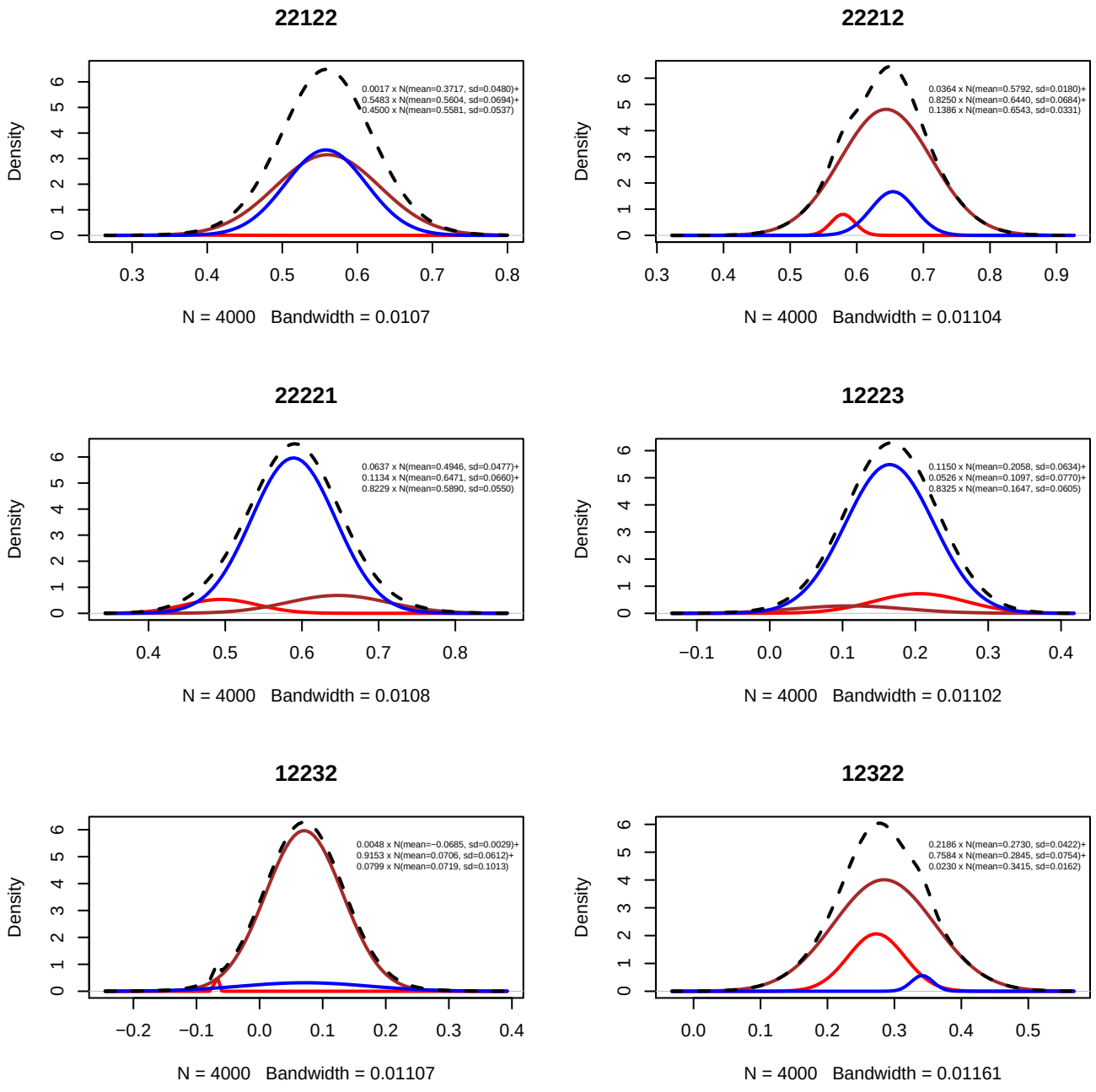


Figure 70: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

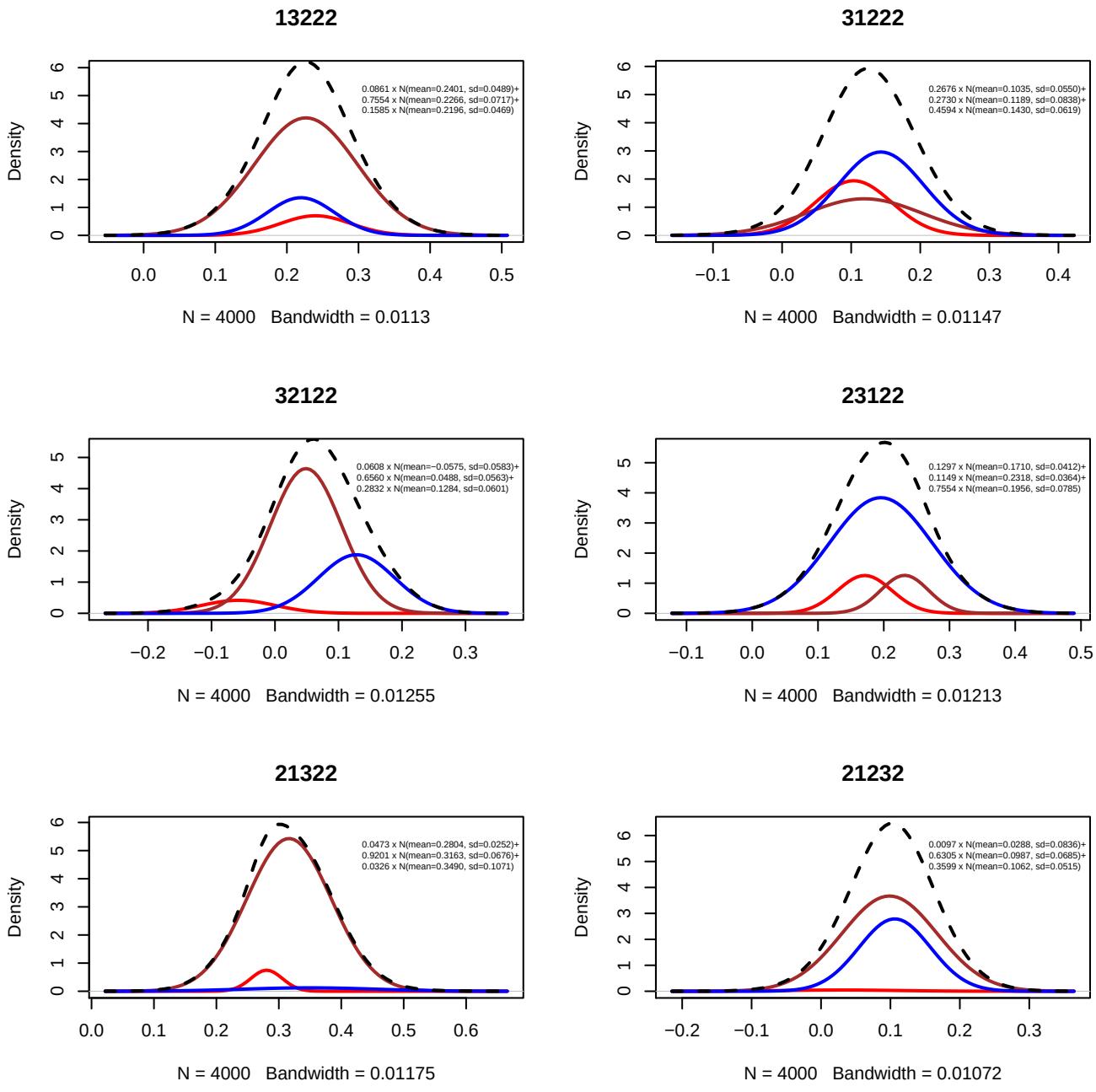


Figure 71: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

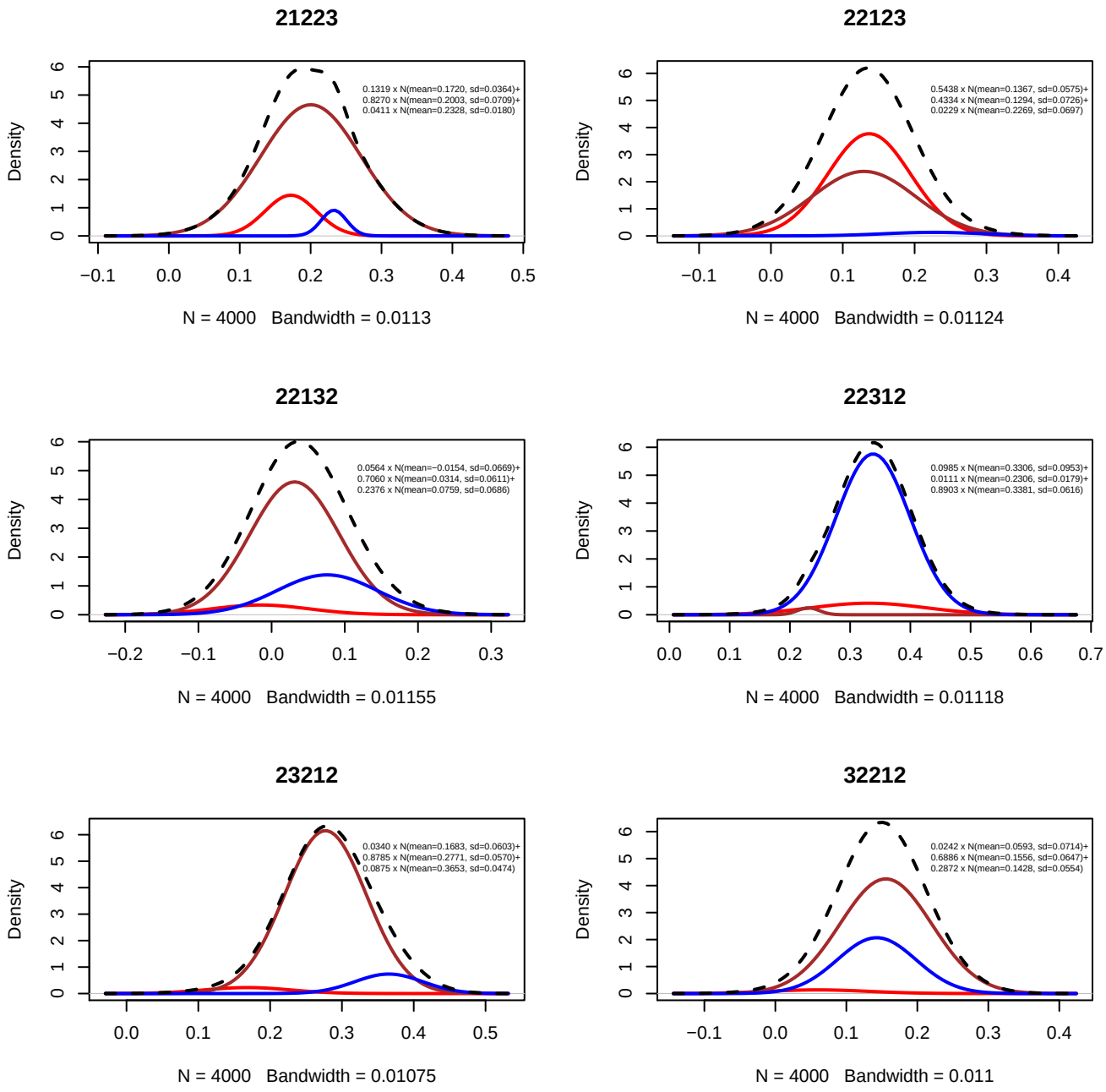


Figure 72: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

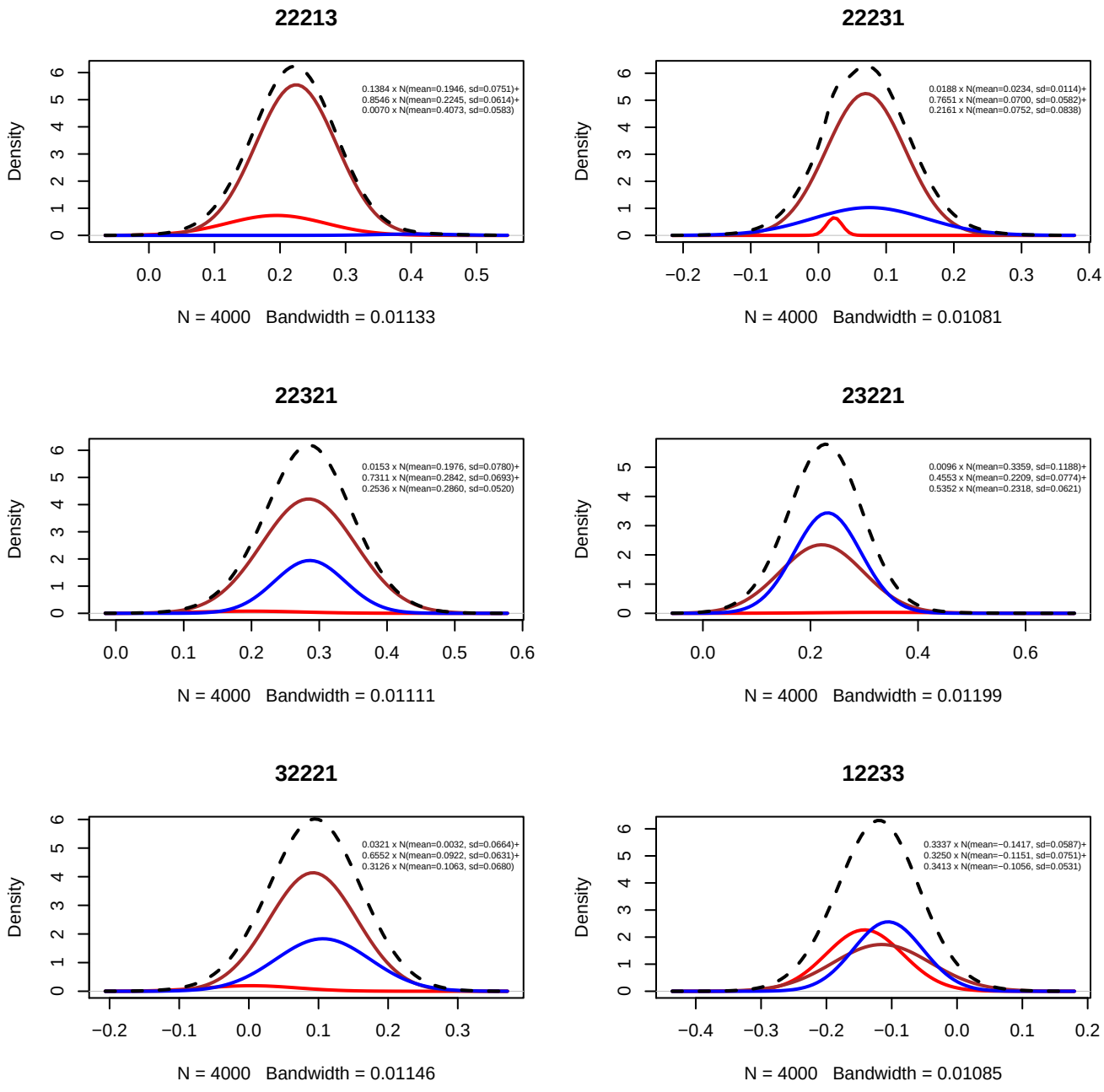


Figure 73: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

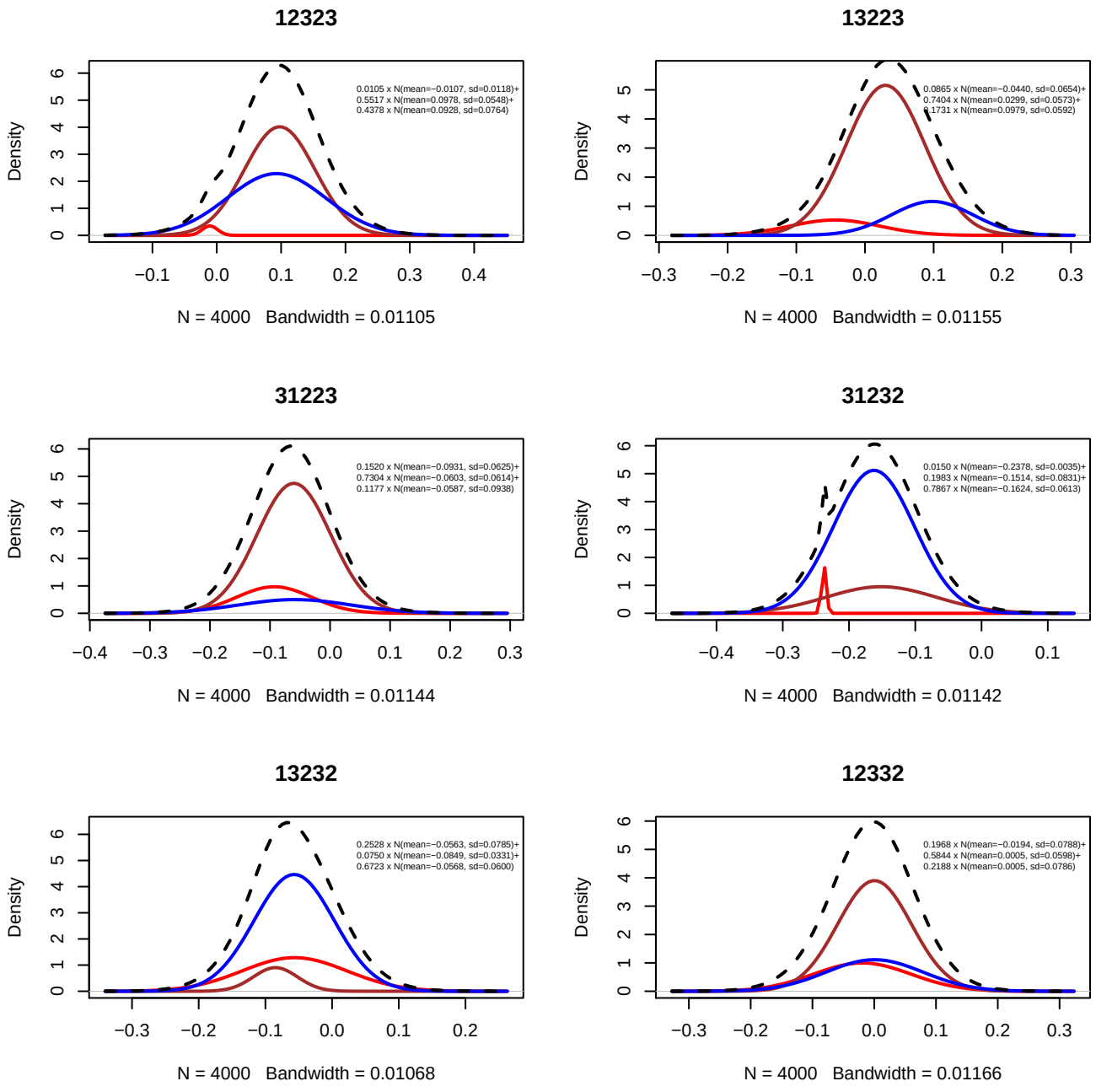


Figure 74: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

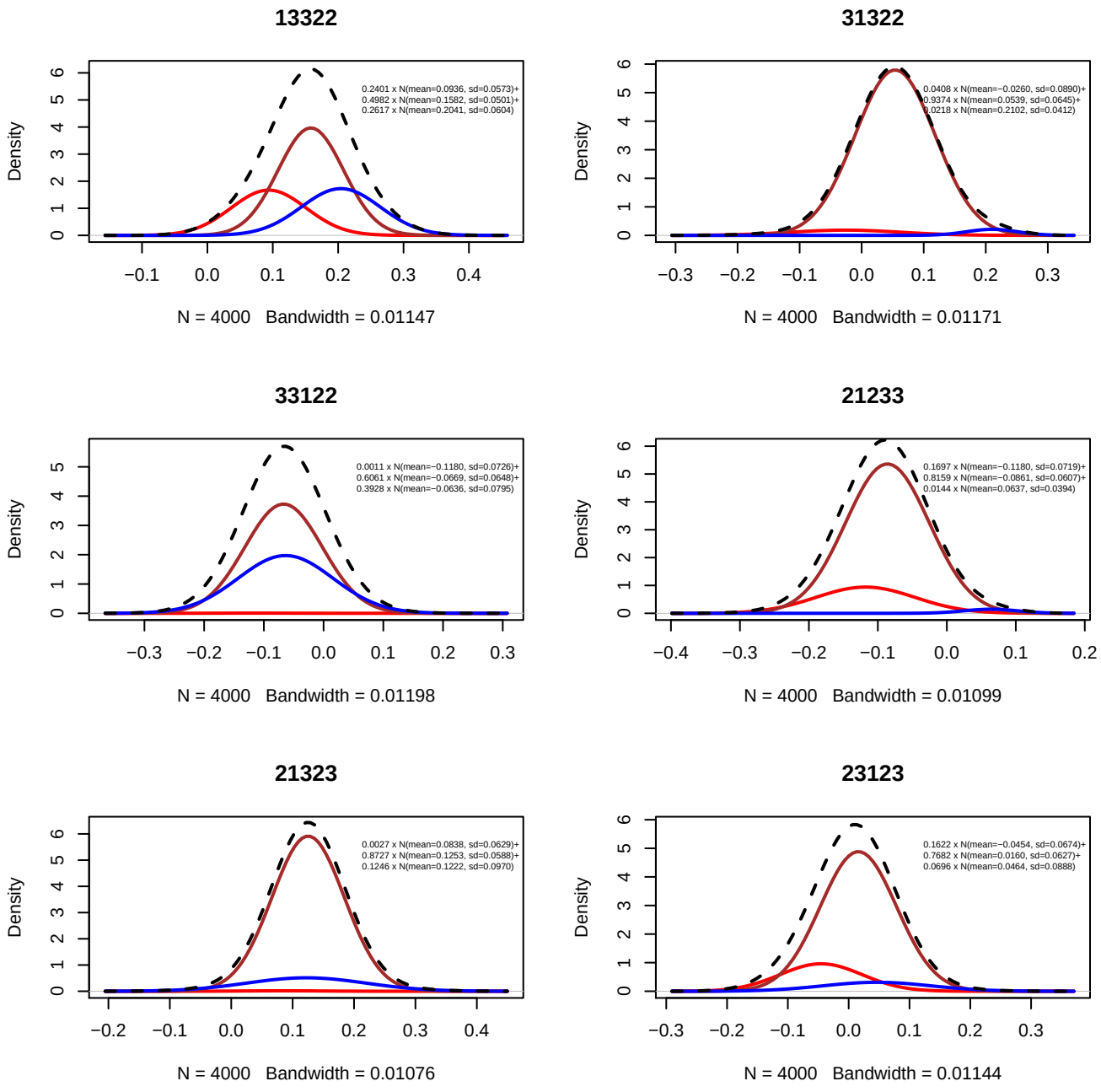


Figure 75: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

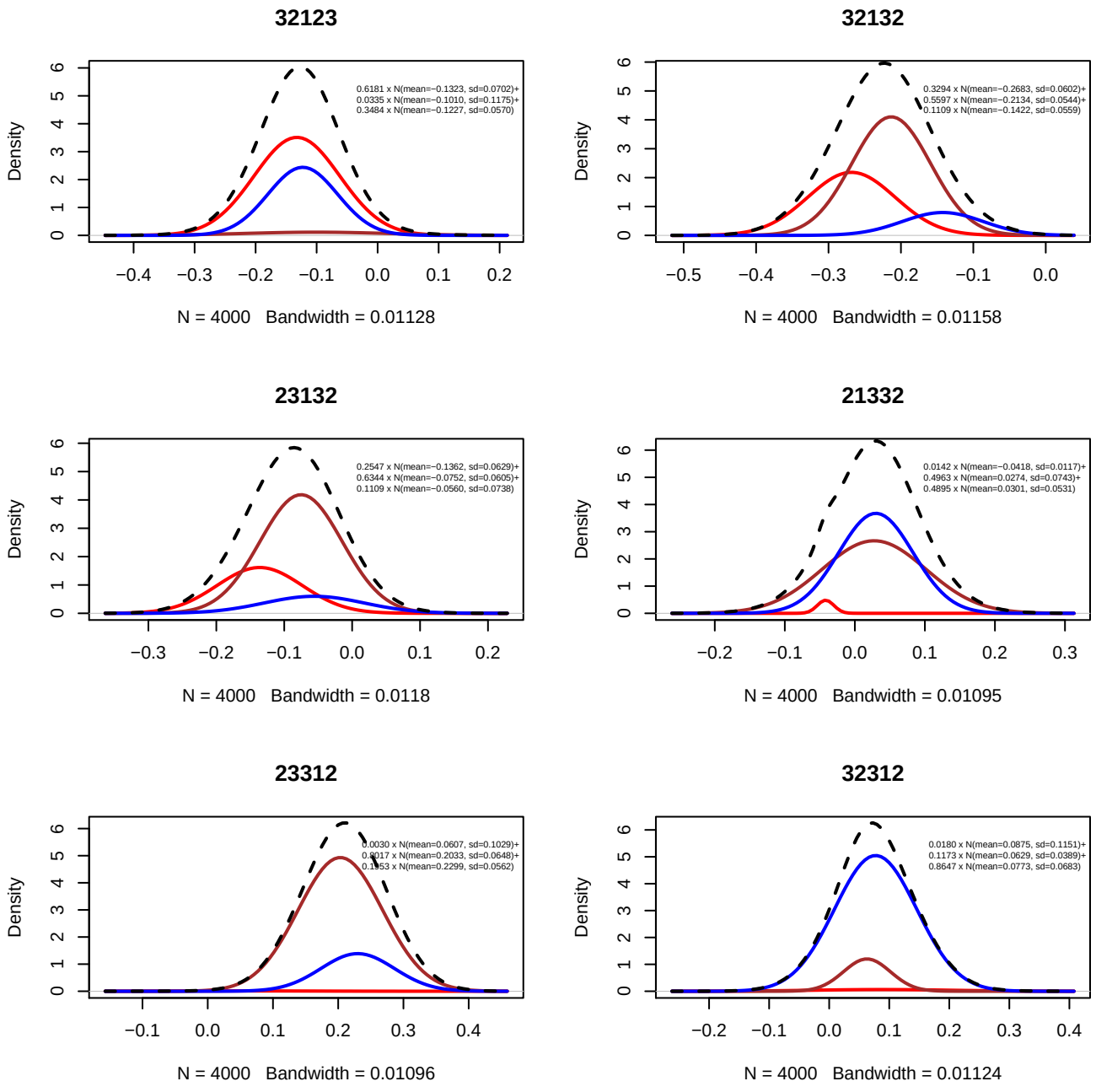


Figure 76: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

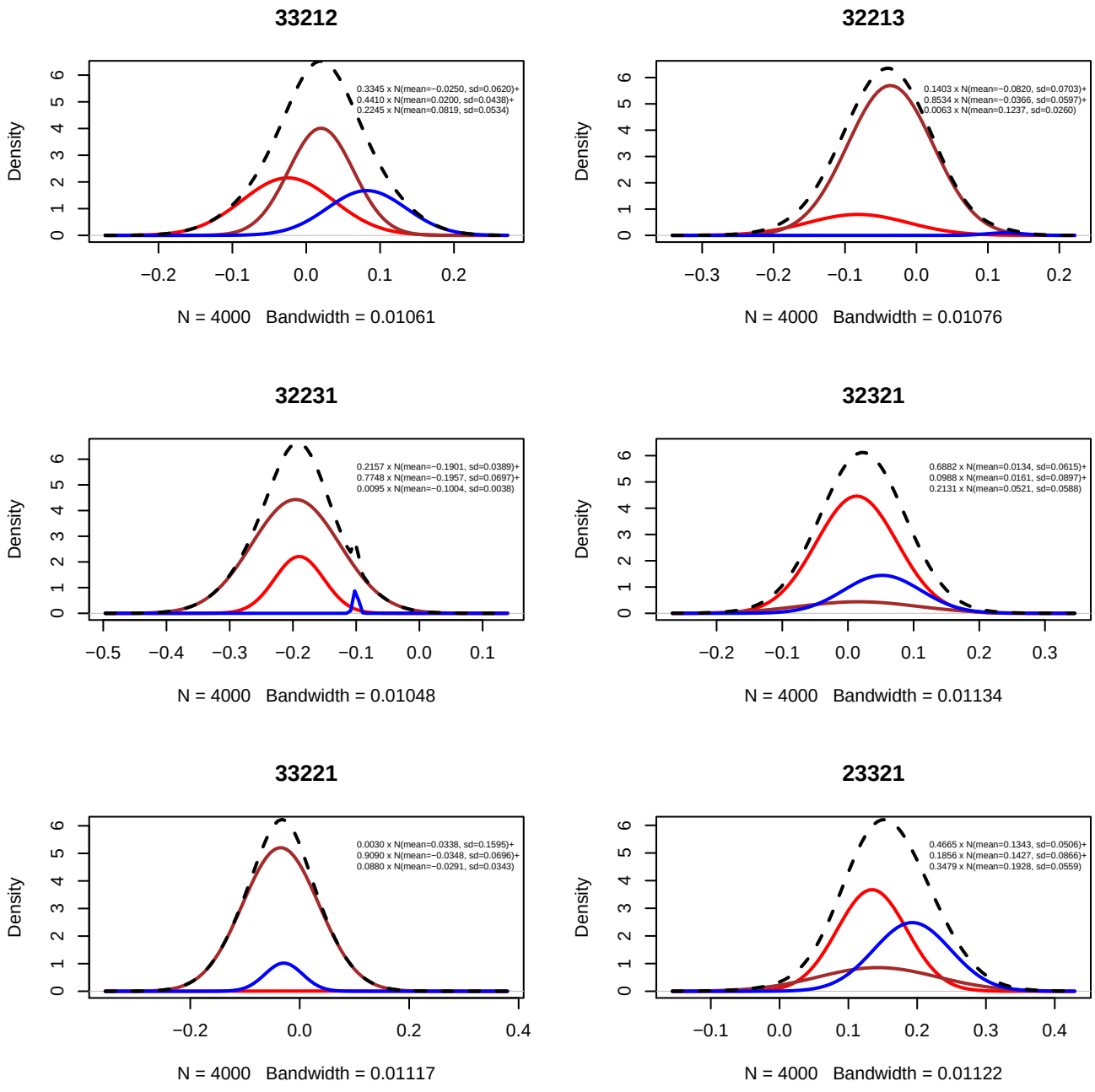


Figure 77: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

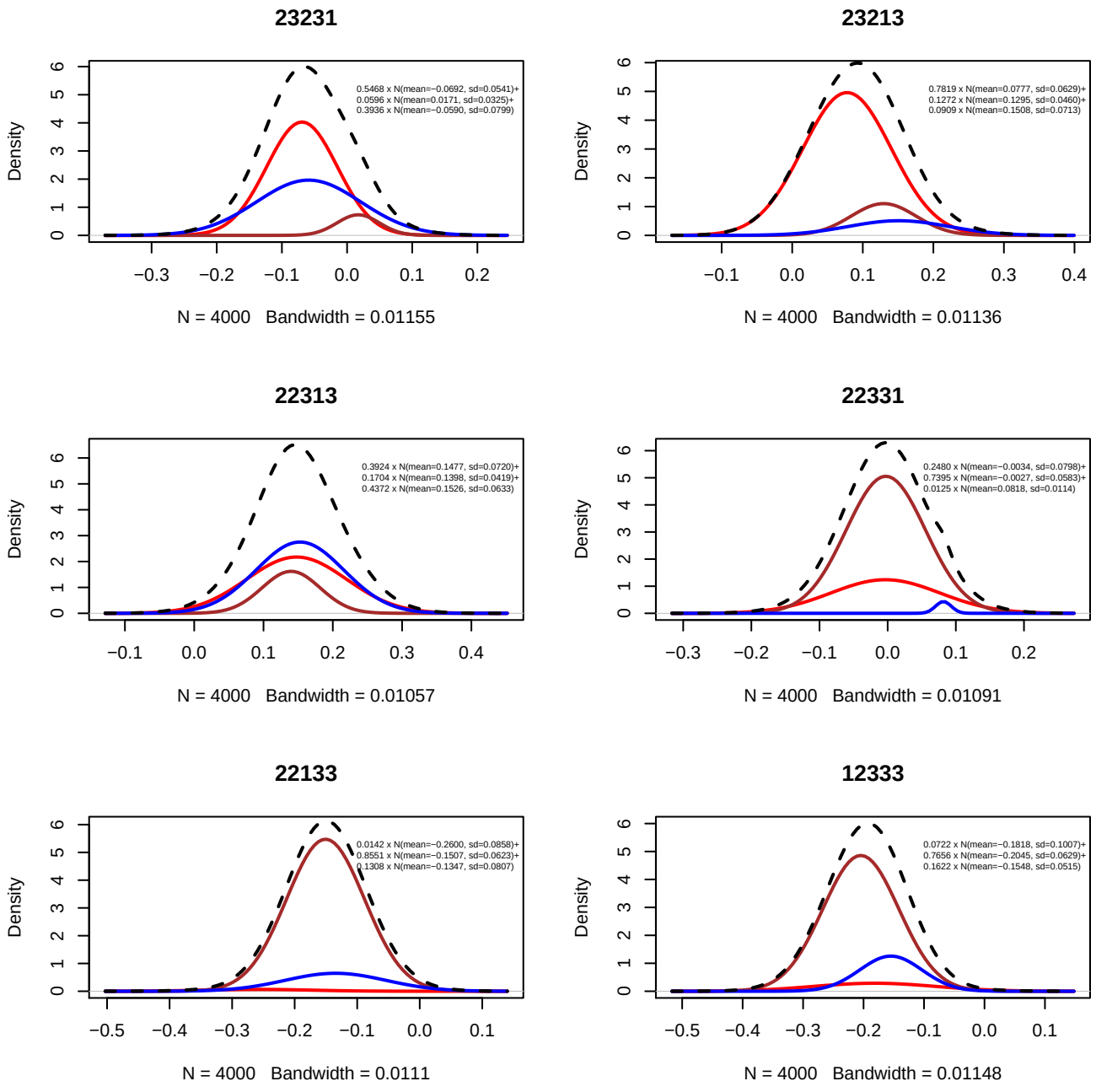


Figure 78: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

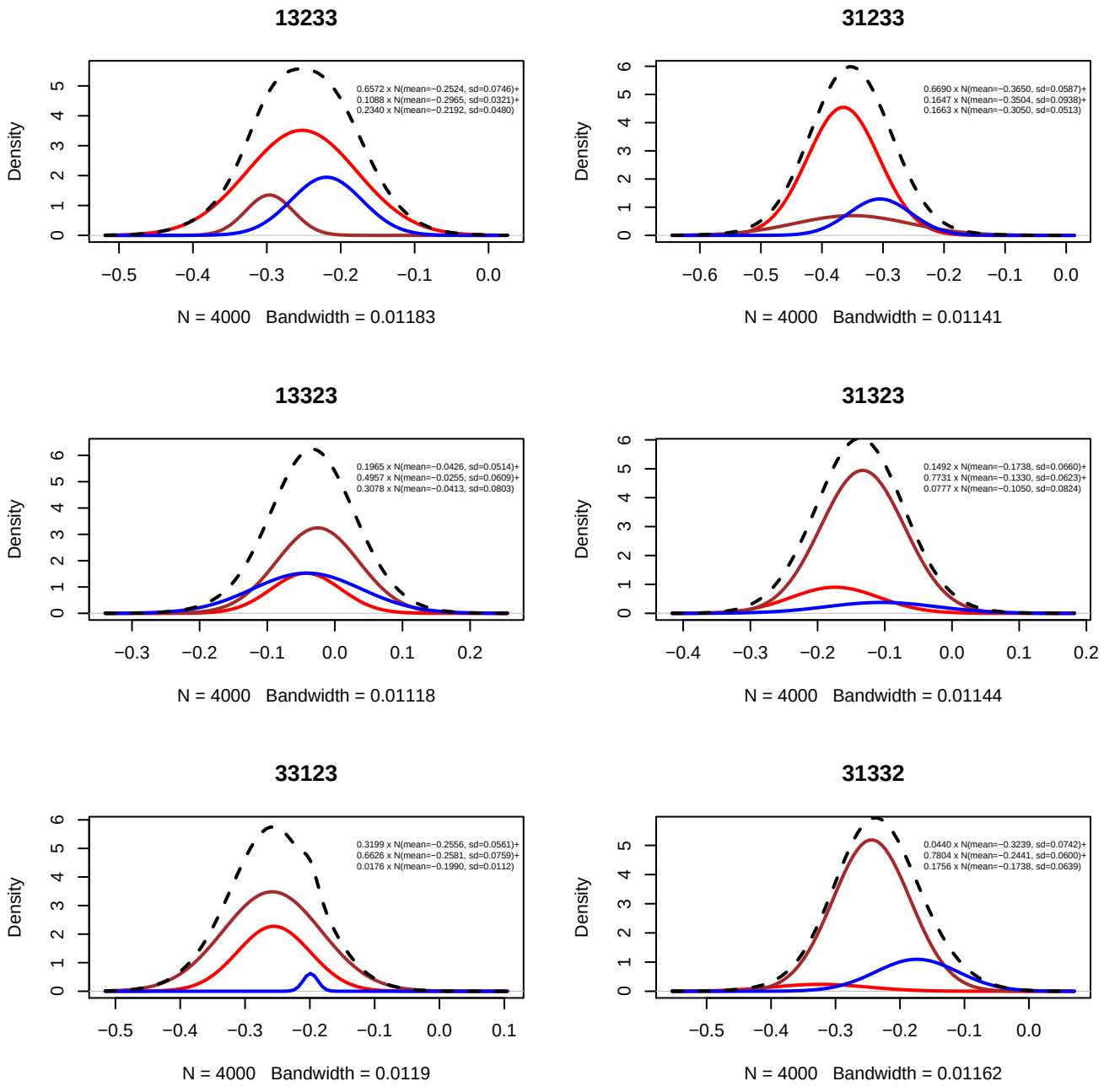


Figure 79: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

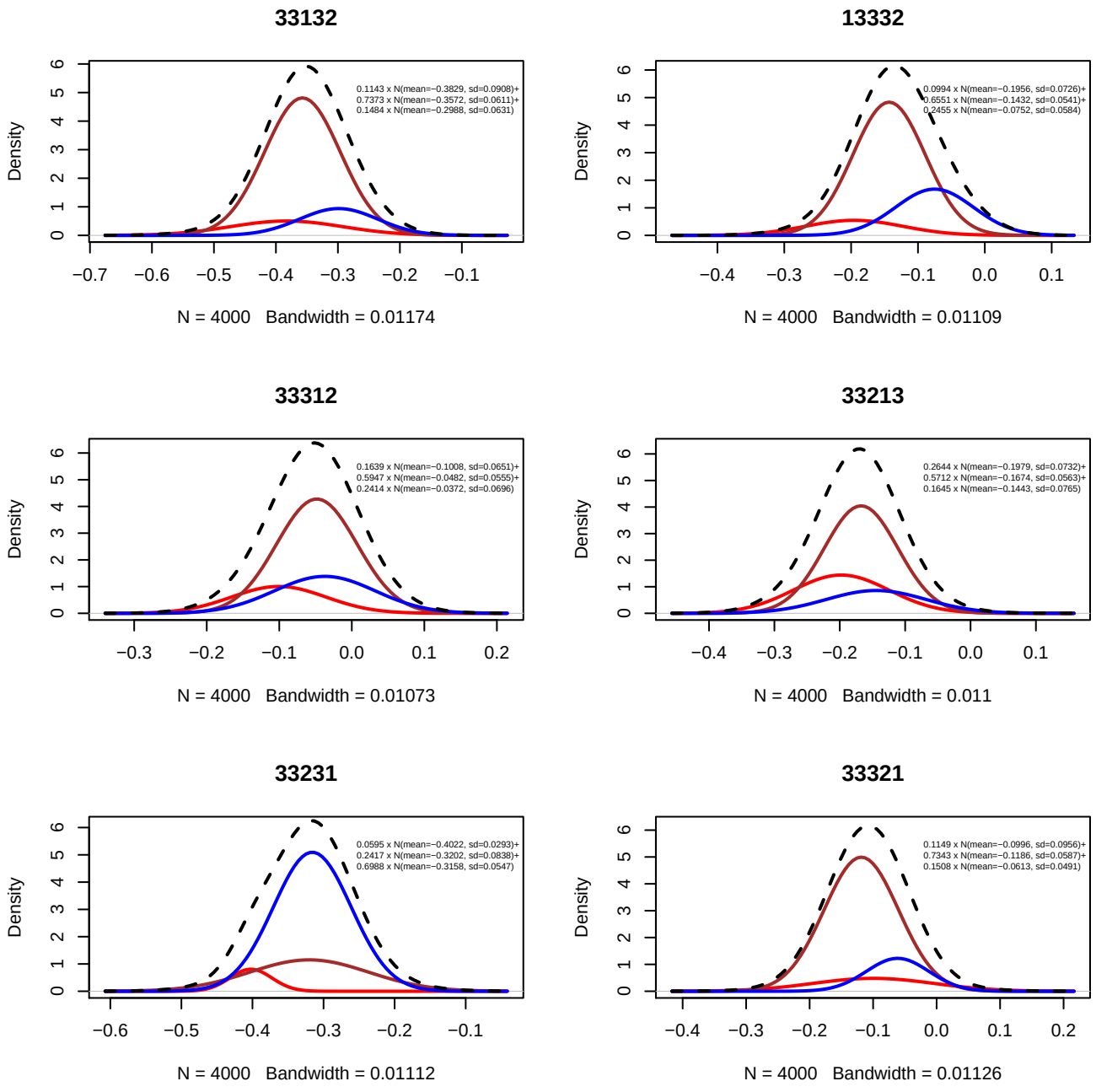


Figure 80: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

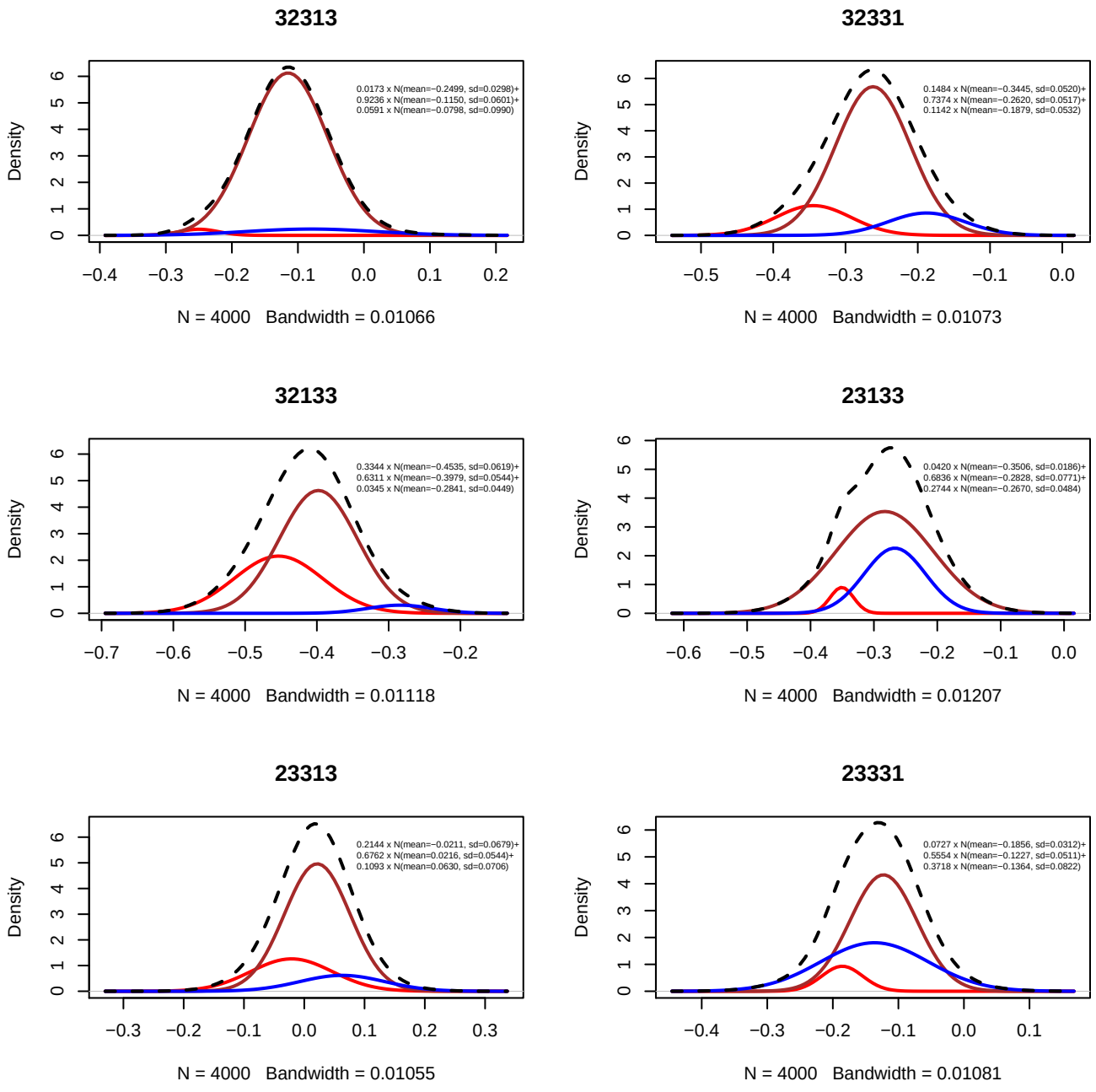


Figure 81: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

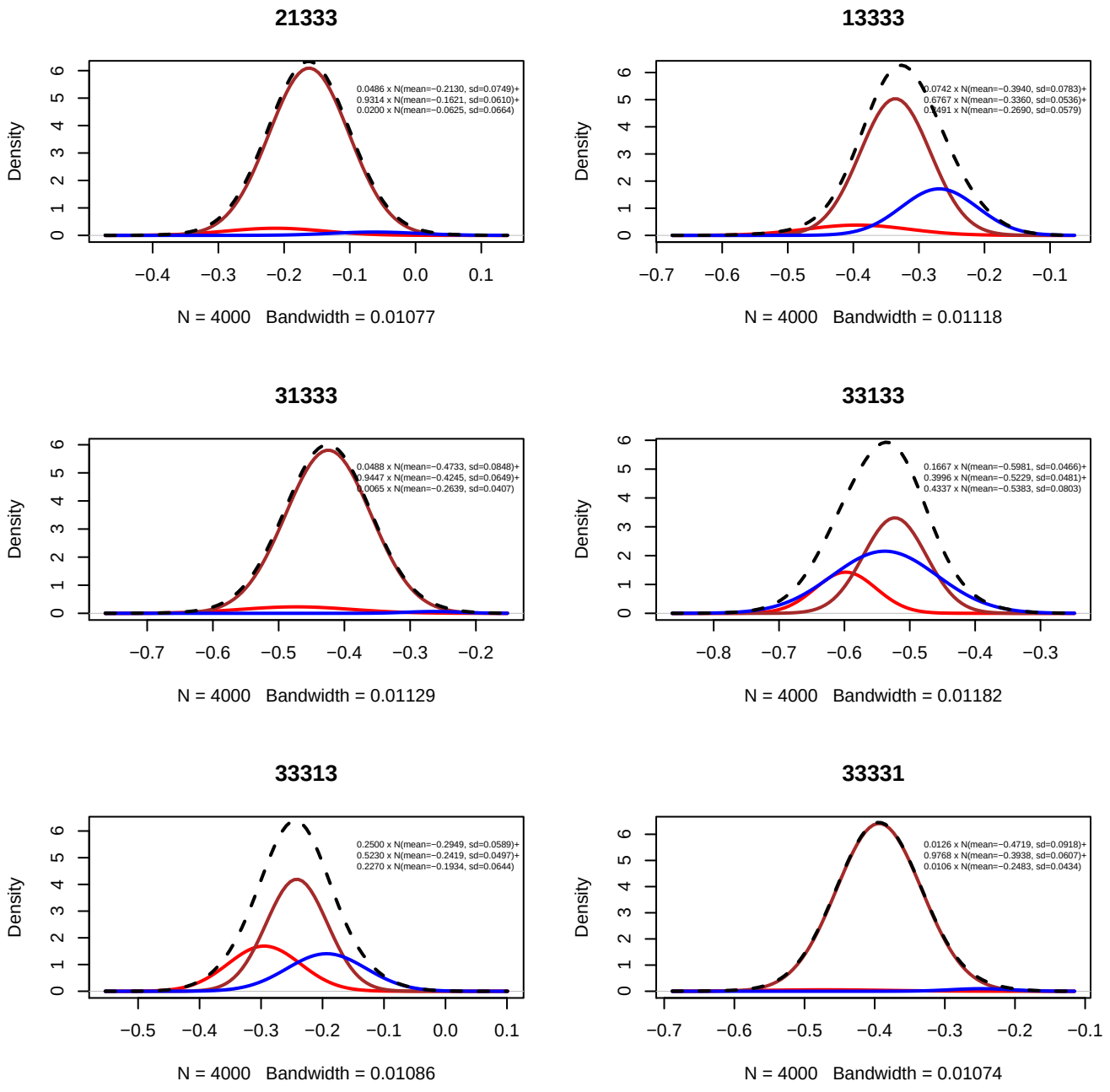


Figure 82: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

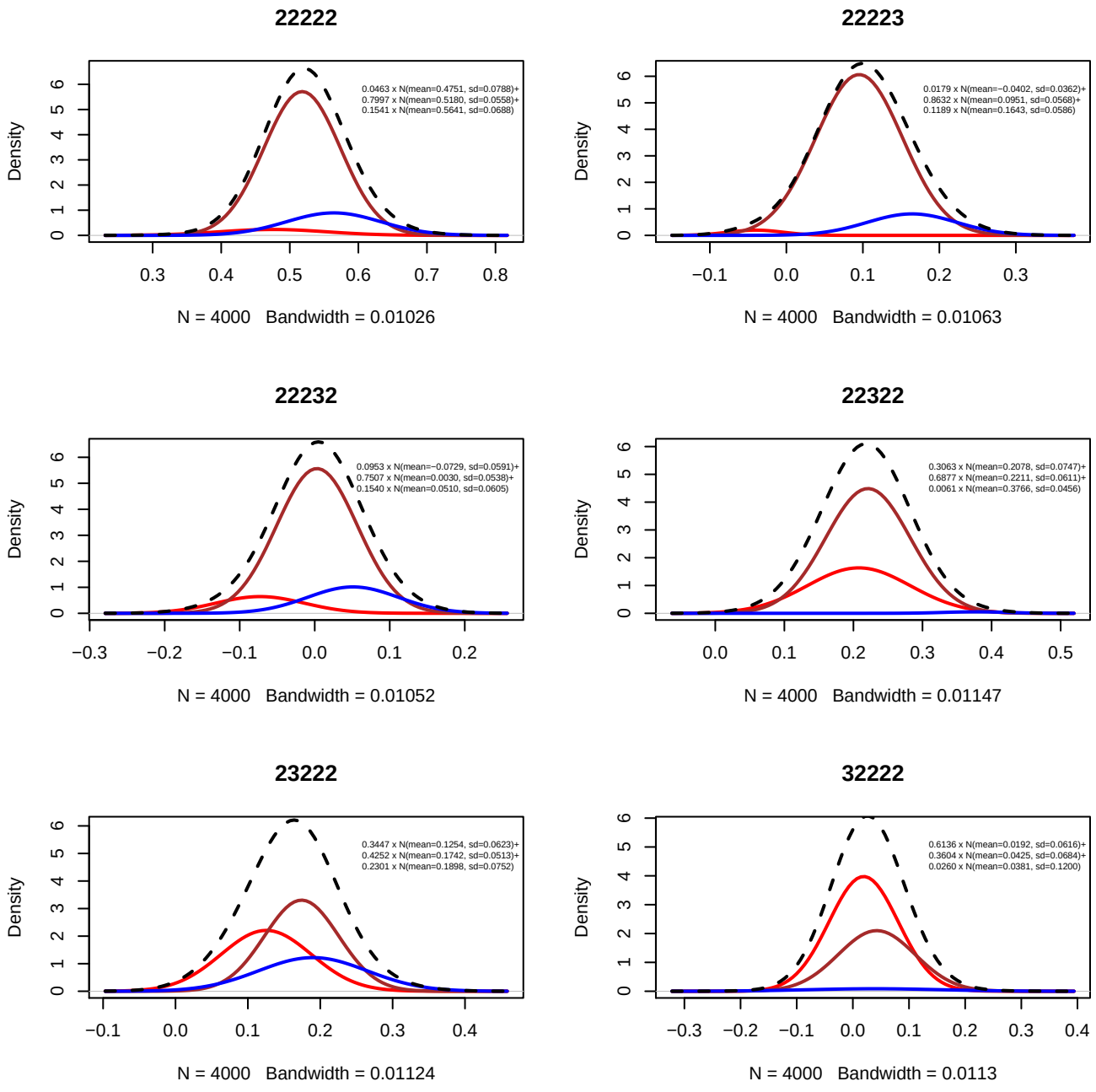


Figure 83: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

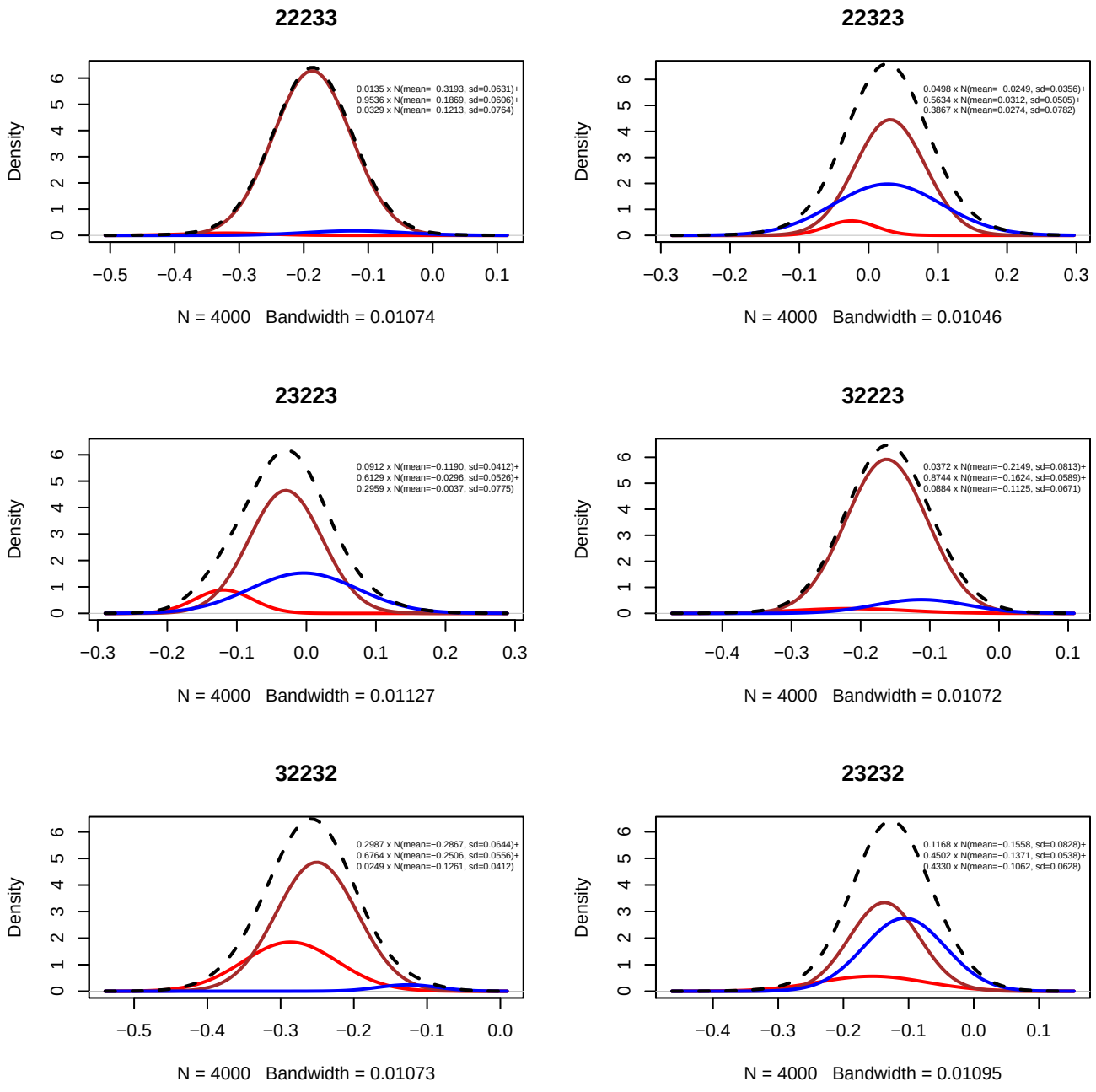


Figure 84: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

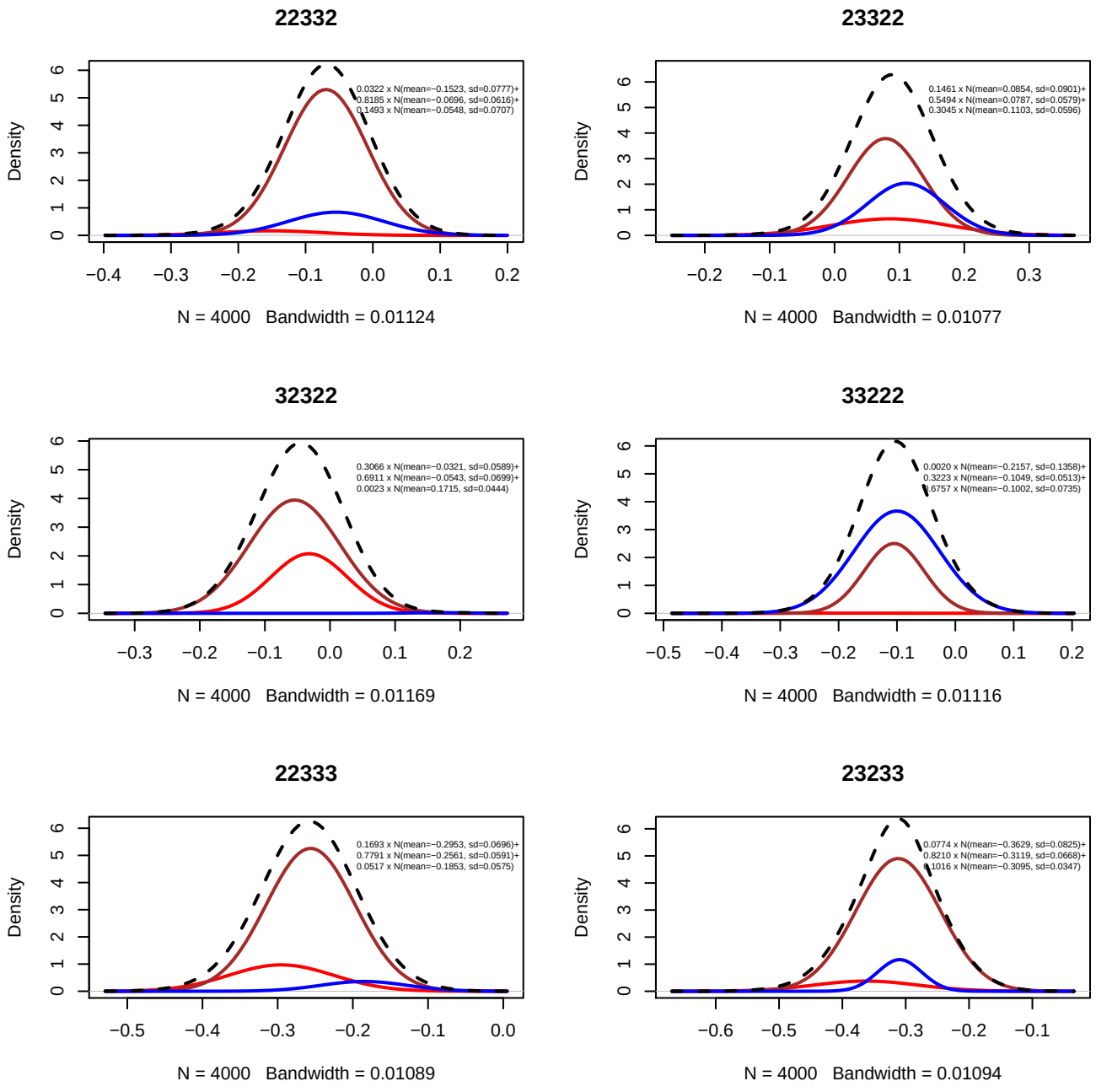


Figure 85: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

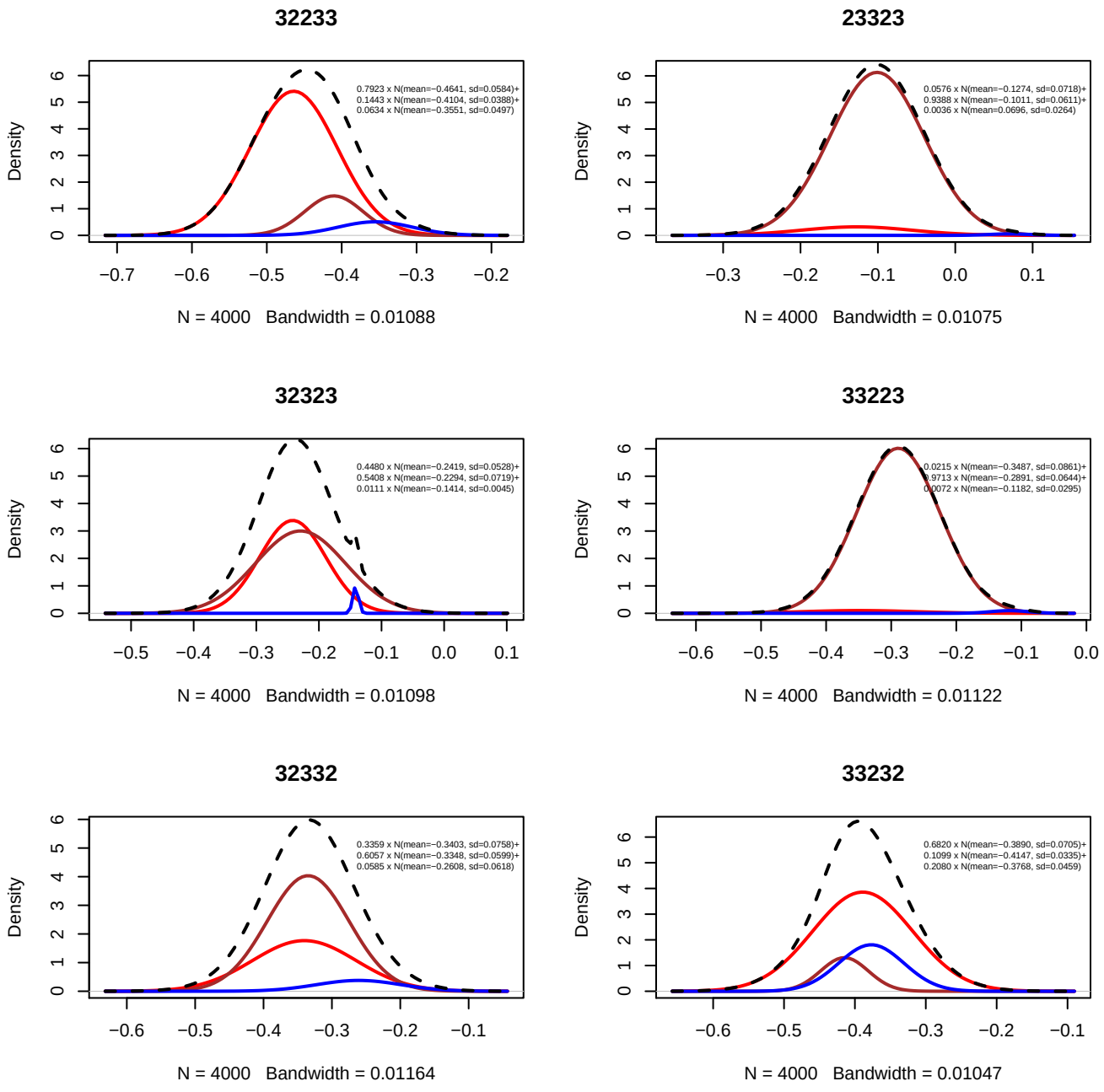


Figure 86: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

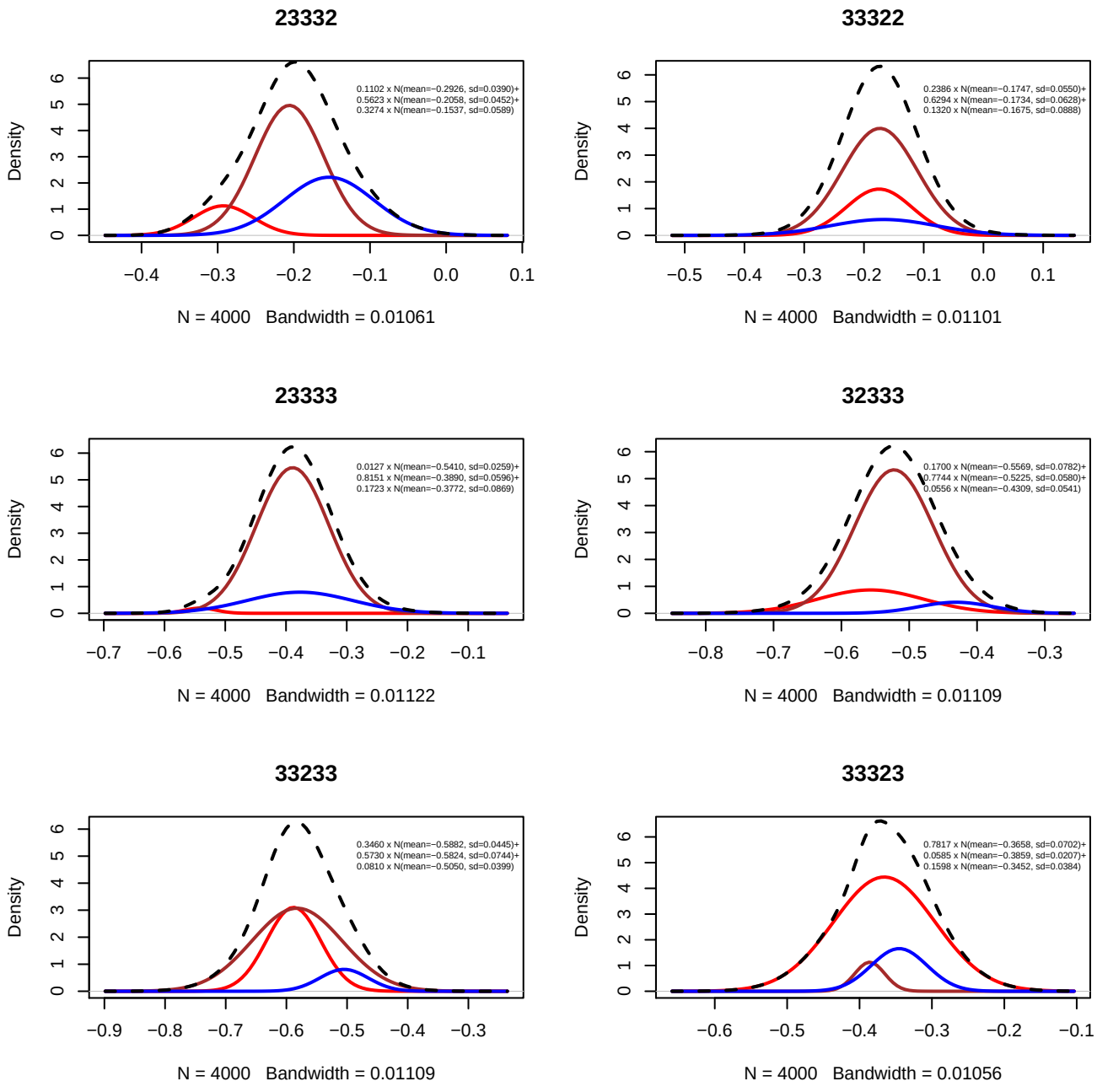


Figure 87: Plots for six EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.

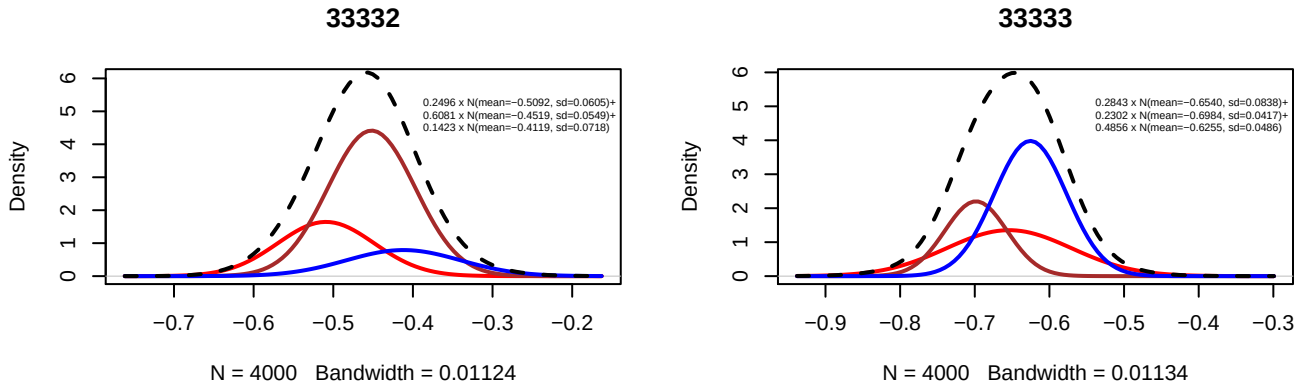


Figure 88: Plots for two EQ-5D-3L states with each of the three components' probability density function of the mixtures (red, brown and blue colour for the first, second and third component in respect) in relation to the mixture (dotted line) distributions.