

A Polya Tree Based Model for Unmarked Individuals in an Open Wildlife Population

Alex Diana, Jim Griffin and Eleni Matechou

Abstract Many ecological sampling schemes do not allow for unique marking of individuals. Instead, only counts of individuals detected on each sampling occasion are available. In this paper, we propose a novel approach for modelling count data in an open population where individuals can arrive and depart from the site during the sampling period. A Bayesian nonparametric prior, known as Polya Tree, is used for modelling the bivariate density of arrival and departure times. Thanks to this choice, we can easily incorporate prior information on arrival and departure density while still allowing the model to flexibly adjust the posterior inference according to the observed data. Moreover, the model provides great scalability as the complexity does not depend on the population size but just on the number of sampling occasions, making it particularly suitable for data-sets with high numbers of detections. We apply the new model to count data of newts collected by the Durrell Institute of Conservation and Ecology, University of Kent.

Key words: Bayesian nonparametrics, Polya Tree, Count Data, Statistical Ecology

Alex Diana
School of Mathematics Statistics and Actuarial Science, University of Kent, e-mail:
ad603@kent.ac.uk

Jim Griffin
School of Mathematics Statistics and Actuarial Science, University of Kent, e-mail: J.E.Griffin-
28@kent.ac.uk

Eleni Matechou
School of Mathematics Statistics and Actuarial Science, University of Kent, e-mail: E.Matechou
@kent.ac.uk

1 Introduction

Monitoring wildlife populations presents particular challenges. For example, it is typically not possible to perform a census of the population of interest by encountering all of the individuals. One of the most cost and time effective ways to monitor a wildlife population is to collect counts of the population on repeated sampling occasions (SO). This protocol is considerably easier to perform than a capture-recapture (CR) scheme as it does not require physical capture or unique identification of the individuals in the population. The most popular model for analysing count data in a frequentistic setting is the N-mixture model introduced in [6], which allows the estimation of population size and capture probability when the population is closed, that is the same individuals are present throughout the study period. However, when the data are sparse or detection probability is low, N-mixture models are known to suffer from parameter identifiability issues and may give rise to infinite estimates for population size [2]. In a Bayesian setting, the natural way to solve issues of parameter identifiability is to assume informative prior distributions on detection probability or on population size in order to obtain sensible posterior distributions.

In this paper we work in a Bayesian framework and we relax the assumption of population closure, allowing for individuals to enter and leave the site (thus becoming available or unavailable for detection) at random times, but still assuming emigration is permanent. The random arrival and departure times are sampled from a distribution with unknown parameters. However, the absence of closure makes it more challenging to separately estimate capture probability, population size and density of arrival/departure times. Hence, it is of primary importance to assume informative prior distributions to obtain ecologically sensible posterior distributions.

In order to allow for the posterior distribution to correctly adjust to the data without relying on parametric assumptions, we use a Bayesian nonparametric (BNP) approach for the prior distribution of the bivariate distribution of arrival and departure times. In particular, we work with Polya Trees, which in the BNP framework are the main alternative to Dirichlet process mixture models for modelling continuous distributions. More information on nonparametric priors can be found in [4].

2 The Polya Tree Prior

We model the joint density of arrival and departure times, using the Polya tree (PT) prior, defined in [5]. A PT has two parameters: the first is a sequence of nested partitions Π of the sample space Ω ($\Omega = \mathbb{R}^2$ in our case), while the second parameter, α , is a sequence of positive numbers associated with each set of each partition.

The partition at the first level, π_1 , is obtained by splitting the sample space in two sets, B_0 and B_1 . Then for the partition at the second level, π_2 , we split each of the two sets in two additional sets B_{00} , B_{01} and B_{10} , B_{11} , respectively.

$$\pi_2 = \{B_{00}, B_{01}, B_{10}, B_{11}\}, \quad B_{00} \cup B_{01} = B_0, \quad B_{10} \cup B_{11} = B_1.$$

The same process is repeated to generate the partitions at the remaining level. A visual representation of the scheme for $\Omega = [0, 1]$ is given in Fig. 1.



Fig. 1: Scheme of the partitions at the first three levels of the Polya tree.

The PT prior is defined in terms of the (random) mass associated with each set of each partition. By defining $\varepsilon_1 \dots \varepsilon_m$ as a generic sequence of 0s and 1s, $B_{\varepsilon_1 \dots \varepsilon_m}$ as a generic set of the partition and $\alpha_{\varepsilon_1 \dots \varepsilon_m}$ as the associated parameter, the mass associated to $B_{\varepsilon_1 \dots \varepsilon_m}$ by the Polya Tree is

$$G(B_{\varepsilon_1 \dots \varepsilon_m}) = \prod_{i=1}^m Y_{\varepsilon_1 \dots \varepsilon_i} \quad (1)$$

where $Y_{\varepsilon_1 \dots \varepsilon_{i-1}0}$ is a $\text{Beta}(\alpha_{\varepsilon_1 \dots \varepsilon_{i-1}0}, \alpha_{\varepsilon_1 \dots \varepsilon_{i-1}1})$ random variable and $Y_{\varepsilon_1 \dots \varepsilon_{i-1}1} = 1 - Y_{\varepsilon_1 \dots \varepsilon_{i-1}0}$. For example, $G(B_{01}) = Y_0(1 - Y_{00})$ where $Y_0 \sim \text{Beta}(\alpha_0, \alpha_1)$ and $Y_{00} \sim \text{Beta}(\alpha_{00}, \alpha_{01})$.

A conjugate scheme for a PT can be constructed if we assume a PT prior for a distribution G , and we have observations $y_1, \dots, y_n \sim G$, since the posterior distribution $G \mid y_1, \dots, y_n$ is still a PT. The parameters α_ε^* of the posterior distribution are computed as $\alpha_\varepsilon^* = \alpha_\varepsilon + n_\varepsilon$ where n_ε is the number of observations falling into B_ε .

A common choice is to center the PT on a pre-specified distribution G_0 , which means that, for every set B of the partition, $\mathbb{E}[G(B)] = G_0(B)$. In this paper we will set the $\alpha_{\varepsilon_1 \dots \varepsilon_{j-1}0}$ and $\alpha_{\varepsilon_1 \dots \varepsilon_{j-1}1}$ associated with the sets $B_{\varepsilon_1 \dots \varepsilon_{j-1}0}$ and $B_{\varepsilon_1 \dots \varepsilon_{j-1}1}$ to be proportional to the mass assigned to these sets from G_0 , that is $\alpha_\varepsilon = c_\varepsilon \times G_0(B_\varepsilon)$, where c_ε is a scale parameter tuning the overall variance around the mean distribution. Finally, we assume that G_0 has random parameter η , which has an additional prior on it, leading to a Mixtures of Polya Trees (MPT), as defined in [3].

3 Model

The data consist of the number of individuals, D_k , detected on SO k , with $k = 1, \dots, K$. We denote by N_k the (latent) number of individuals available for detection at SO k and by p the detection probability, assumed to be constant for each individual and each SO. Clearly $D_k \sim \text{Binomial}(N_k, p)$.

We do not assume that individuals are present throughout the study period but we instead assume that their arrival and departure times are random. These times are assumed to be sampled from a Poisson process, with intensity that can be written as $\omega \times \tilde{\nu}$ where ω is the overall mass of the process and $\tilde{\nu}$ is a probability density function. The MPT is employed as a prior for $\tilde{\nu}$ and we call P_0 the prior distribu-

tion on the hyperparameters of the centring distribution G_0 , which we define later. As departure is always greater than arrival, \tilde{v} is defined on $\{(x, y) \in \mathbb{R}^2 | x < y\}$.

Although the data depend only on the latent number of individuals N_k , we introduce additional latent variables to make the inference of the PT more efficient. Let $\{t_k\}_{k=1, \dots, K}$ be the times when samples are collected and take by convention $t_0 = -\infty$ and $t_{K+1} = \infty$. Additionally, let n_{ij} be the number of individuals having arrival times between t_i and t_{i+1} and departure times between t_j and t_{j+1} (with $n_{ij} = 0$ for $i > j$). The N_k can easily be obtained from the n_{ij} as $N_k = \sum_{j=k}^K \sum_{i=0}^{k-1} n_{ij}$.

We make the standard choices of a Beta prior distribution for detection probability and a Gamma prior distribution for the overall intensity of the process. The hierarchical structure of the model is the following:

$$\begin{aligned}
 D_k &\sim \text{Binomial}(N_k, p), & N_k &= \sum_{j=k}^K \sum_{i=0}^{k-1} n_{ij}, & k &= 1, \dots, K, \\
 n_{ij} &\sim \text{Poisson}(\omega \times \omega_{ij}), & i &= 0, \dots, K, \quad j = 0, \dots, K \quad i > j, \\
 \omega_{ij} &= \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} \tilde{v}(x, y) \, dx \, dy, & i &= 0, \dots, K, \quad j = 0, \dots, K, \quad i > j, \\
 \omega &\sim \text{Gamma}(a_\omega, b_\omega), & p &\sim \text{Beta}(a_0, b_0), \\
 \tilde{v} &\sim \text{PT}(\Pi, \alpha^\eta), & \eta &\sim P_0.
 \end{aligned}$$

In order to center the PT on a pre-specified distribution, we use the approach explained in Section 2 of using a fixed partition and choosing the α according to the value η of the parameters of the centring distribution. The dependence on η is thus only in the α .

The process used to create the partition is explained in Fig. 2. The last level corresponds to the partition defined by the sampling occasions. Since we use the latent variables n_{ij} and not the exact, on a continuous scale, individual arrival and departure times, it is not meaningful to build a finer level of the partitions, as no information is available about them.

We center the PT on a bivariate distribution with independent double exponential marginal distributions, with probability density function (pdf)

$$G_0(x_1, x_2; \mu_1, \mu_2, \lambda_1, \lambda_2) = \frac{1}{2\lambda_1} \exp\left(-\frac{|x_1 - \mu_1|}{\lambda_1}\right) \frac{1}{2\lambda_2} \exp\left(-\frac{|x_2 - \mu_2|}{\lambda_2}\right),$$

with the constraint that $G_0(x_1, x_2; \mu_1, \mu_2, \lambda_1, \lambda_2) = 0$ if $x_1 > x_2$.

The sets of the partition are squares and triangles, as shown in Fig. 2. The choice of the double exponential is motivated by the fact that integrals of this distribution on squares and triangles can be computed analytically, without resorting to numerical techniques. The hyperparameters (μ_1, μ_2) are given a bivariate normal prior distribution and λ_1 and λ_2 two independent Gamma prior distributions.

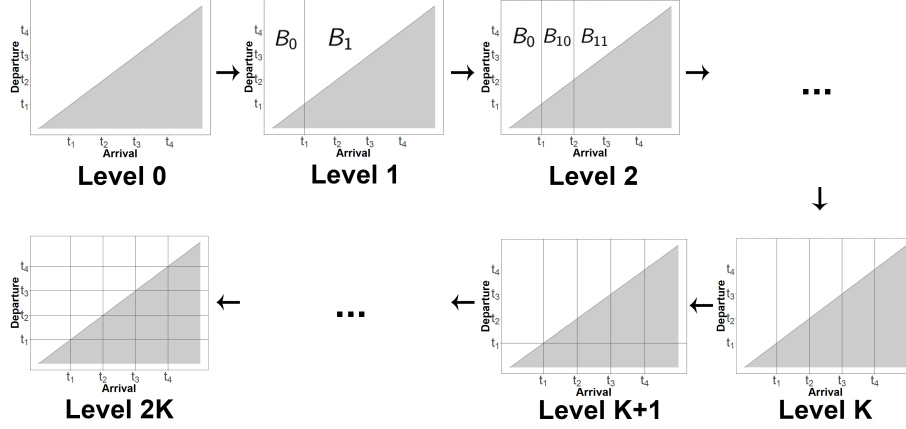


Fig. 2: Partitions of the Polya Tree. Observations occur only in the region above the diagonal. At step k , the set of individuals departing after the k -th sampling occasion is split into the individuals departing between the k -th and $(k + 1)$ -th sampling occasion and the individuals departing after the $(k + 1)$ -th sampling occasion. After having reached level K , each set is split according to the same procedure but with respect to the other dimension (arrival times).

4 Computational Notes

Posterior inference is performed using a Gibbs sampler. While this is straightforward for some parameters, such as the detection probability p and the intensity ω of the Poisson process thanks to conjugacy, for other parameters posterior inference is not straightforward given that we are working with a PT.

The conditional distribution of the parameters $(\{n_{ij}\}, \{\omega_{ij}\}, \eta)$ given p and ω can be written as

$$p(\{n_{ij}\}, \{\omega_{ij}\}, \eta | \{D_k\}, p, \omega) \propto p(\{D_k\} | \{N_k\}, p) p(\{n_{ij}\}, \omega, \{\omega_{ij}\}) p(\{\omega_{ij}\} | \alpha_\eta) p(\eta) \\ \propto \prod_{k=1}^K \text{Binomial}(D_k | N_k, p) \prod_{i=1}^{K+1} \prod_{j=1}^i \text{Poisson}(n_{ij} | \omega \times \omega_{ij}) p(\{\omega_{ij}\} | \Pi, \alpha_\eta) p(\eta)$$

where the distribution $p(\{\omega_{ij}\} | \alpha_\eta)$ is given from the PT.

When writing the posterior distribution of the parameter η , we can integrate out the parameter ω_{ij} , by employing a different parametrisation of the PT. First, we define as n_ε the number of observations in set B_ε and as $q_{\varepsilon 0}$ the probability of assigning an observation in set $B_{\varepsilon 0}$ given that we are in B_ε which, according to the structure of the PT, has a Beta($\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1}$) prior distribution. The marginal likelihood of the n_ε can be expressed as

$$p(\{n_\varepsilon\}|\{\alpha_\varepsilon\}) = \prod_\varepsilon \int \text{Bin}(n_{\varepsilon 0}|n_\varepsilon, q_{\varepsilon 0}) \text{Beta}(q_{\varepsilon 0}|\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1}) dq_{\varepsilon 0}.$$

Integrating out the probabilities $q_{\varepsilon 0}$ gives as a result a beta-binomial distribution. Since the probability mass function of the beta-binomial is

$$f(k|n, \alpha, \beta) = \binom{n}{k} \frac{B(k + \alpha, n - k + \beta)}{B(\alpha, \beta)},$$

the marginal likelihood of the latent variable n_ε given the hyperparameter η is

$$p(\{n_\varepsilon\}|\eta) \propto \prod_\varepsilon \frac{B(\alpha_{\varepsilon 0} + n_{\varepsilon 0}, \alpha_{\varepsilon 1} + n_{\varepsilon 1})}{B(\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1})}.$$

The posterior distribution for the latent variable n_{ij} can be written as

$$p(n_{ij}|\{D_k\}, p, \omega, \omega_{ij}) \propto \text{Poisson}(n_{ij}|\omega \times \omega_{ij}) \prod_{k=1}^K \text{Binomial}(D_k|N_k, p).$$

The parameter is updated with a random walk with uniform proposal over $(n_{ij} - K_{ij}, \dots, n_{ij} + K_{ij})$, where K_{ij} is chosen according to the value of the n_{ij} chosen as a starting point for the MCMC. In our case, we choose K_{ij} to be 1/5 of the starting point of n_{ij} .

The parameters ω_{ij} correspond to the masses assigned by the distribution $\tilde{\nu}$ to the sets in the partition of the last level of the Polya tree. Hence, they can be sampled as a product of Beta distributions as in (1). The parameters of the PT are updated at each iteration conditional on the latent variable n_{ij} , using the standard update explained in Section 2.

5 Application

The data used in our application consist of weekly detections of great crested newts (GCN) (*Triturus cristatus*). This species generally start to migrate to ponds in late winter in order to breed. Subsequently, they leave the breeding site at the end of the summer and hibernate on land. Sampling took place in ponds located at the University of Kent with data collected between the end of February 2016 until the start of September 2016, which covers a large part of the breeding period of the newts. Samples were collected on weeks 1 – 22, 24 – 27 and 29 of the season.

GCN are uniquely identifiable and hence individual capture histories exist. A total of 69 individuals were captured during the study. However, in this case, the individual CR data have been collapsed to simple count data, which are obtained by recording the number of individual newts caught on each of the sampling weeks. It is believed that the population size is close to the sample size, and we choose a

prior distribution for ω with mean 76 to represent our belief that around 90% of the individuals have been detected at least once. The variance has been chosen to have a relative weakly informative prior, as the 95% prior mass includes up to 135 individuals. Finally, 95% of the prior mass for detection probability is placed on the $(0.05, 0.35)$ interval, based on previous analyses of data on the same population.

Prior knowledge suggests that a considerable number of individuals tend to arrive between the beginning of March and the end of April. Additionally, individuals depart between the end of May and the end of July. In order to translate this knowledge into prior distributions, we choose hyperpriors for $\eta = (\mu, \lambda)$ such that 95% of the prior mass of the arrival and departure density is in the aforementioned ranges.

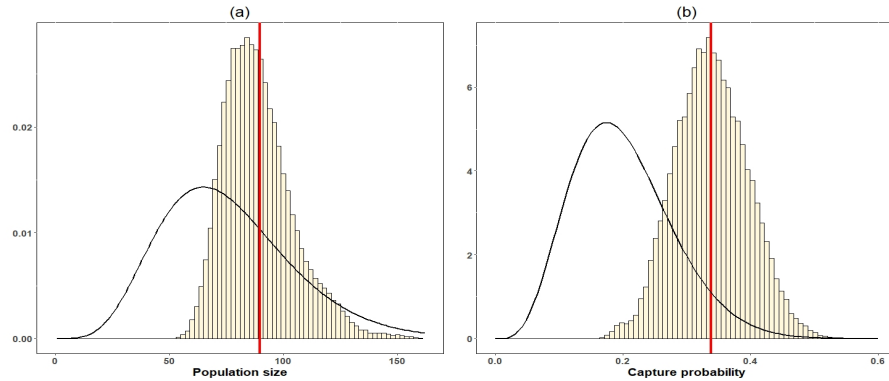


Fig. 3: Posterior distributions of population size (a) and detection probability (b), with red vertical lines showing the posterior means and black lines showing the prior distribution.

The posterior mean estimate of the population size is 89, while the posterior mean of the detection probability is 0.33. The two posterior distributions are shown in Fig. 3. The posterior distribution of the population size is different from the prior distribution, as more individuals are estimated to be at the site than expected by the ecologists. Moreover, in Fig. 4 we display the posterior mean of the latent number of individuals available at each sampling occasion together with the number of individuals counted. For some sampling occasions, the empirical estimated detection probability, estimated as the ratio between the estimated number of individuals available and the counted individuals, is outside the 95% posterior credible interval for detection probability. This suggests that detection probability is not constant across sampling occasions, as assumed in the model. According to expert knowledge, changes in detection probability might be due to differences in environmental conditions between sampling occasions, which affect behaviour of newts.

The posterior cumulative distribution functions (cdf) of arrival and departure are also shown in Fig. 4. As 95% of the individuals are estimated to arrive before any

individual has departed, the number of individuals is estimated to be fairly constant between sampling occasions 9 and 14.

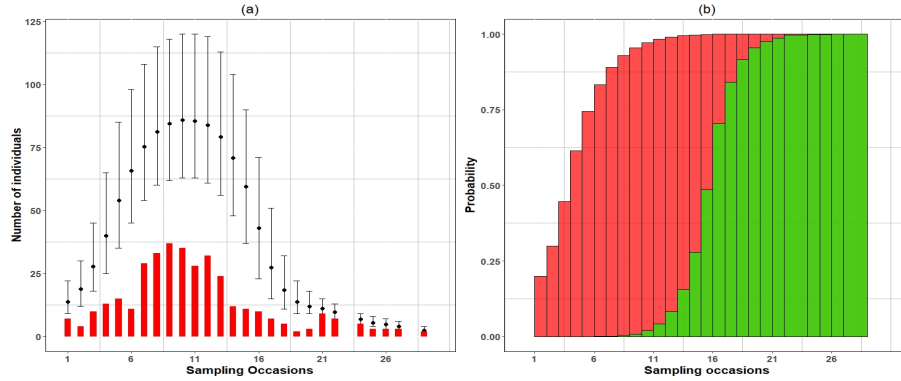


Fig. 4: (a) 95% posterior credible interval of the latent number of individuals available for detection, shown in black, and the number of individuals detected shown in red. (b) Posterior mean of the cdf of arrival times (red) and departure times (green).

6 Conclusion

In this paper we have presented a BNP model for count data on an open wildlife population consisting of individuals entering and exiting the site at random times. By assuming a PT prior, we make no parametric assumptions on the shape of the arrival and departure distribution. Moreover, the implementation is fast as the computational complexity does not depend on the number of individuals but on the levels of the PT, which depends on the number of sampling occasions. However, given the small amount of information provided in count data, it is important to assume meaningful and informative prior distributions in order to have sensible posterior distributions. In this paper, we assume informative prior distributions for detection probability and for population size, available thanks to expert knowledge.

As we mentioned in the introduction, another common sampling protocol is CR which, as opposed to count data, provides individual information that can improve estimation. Hence, a possible extension is to model count data and CR data jointly. Another useful extension is to model data collected at different sites, by replacing the Polya tree prior with a hierarchical Polya tree prior, defined in [1].

References

1. Christensen, J. Ma, Li.: A Bayesian hierarchical model for related densities using Polya trees. ArXiv 1710.01702v2
2. Dennis, E. B., Morgan, B. J., and Ridout, M. S.: Computational aspects of N-mixture models. *Biometrics*, 71(1):237-246. (2015)
3. Hanson T.: Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, 101:1548-1565. (2006)
4. Hjort, N. L., Holmes, C., Muller, P., Walker, S. G. *Bayesian Nonparametrics*, volume 28. Cambridge University Press (2010)
5. Lavine, M.: Some aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, pages 1222-1235. (1992)
6. Royle, J. A.: N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1):108-115. (2004)