Journal of Modern Applied Statistical Methods

Volume 15 | Issue 1 Article 15

5-2016

Bayesian Estimation of the Size of a Street-Dwelling Homeless Population

Lawrence C. McCandless
Faculty of Health Sciences, Simon Fraser University, lmccandl@sfu.ca

Michelle L. Patterson
Faculty of Health Sciences, Simon Fraser University

Lauren B. Currie
Faculty of Health Sciences, Simon Fraser University

Akm Moniruzzaman
Faculty of Health Sciences, Simon Fraser University

Julian M. Somers
Faculty of Health Sciences, Simon Fraser University

Recommended Citation

McCandless, Lawrence C.; Patterson, Michelle L.; Currie, Lauren B.; Moniruzzaman, Akm; and Somers, Julian M. (2016) "Bayesian Estimation of the Size of a Street-Dwelling Homeless Population," *Journal of Modern Applied Statistical Methods*: Vol. 15 : Iss. 1 , Article 15.
DOI: 10.22237/jmasm/1462076040

Bayesian Estimation of the Size of a Street-Dwelling Homeless Population

Cover Page Footnote

Lawrence McCandless if funded by a Discovery Grant from the National Sciences and Engineering Research Council (NSERC), Canada.

Journal of Modern Applied Statistical Methods
May 2016, Vol. 16, No. 1, 276-298.

ISSN 1538 − 9472

Bayesian Estimation of the Size of a Street-Dwelling Homeless Population

Lawrence C. McCandless
Simon Fraser University
Burnaby, British Columbia

Michelle L. Patterson
Simon Fraser University
Burnaby, British Columbia

Lauren B. Currie
Simon Fraser University
Burnaby, British Columbia

Akm Moniruzzaman
Simon Fraser University
Burnaby, British Columbia

Julian M. Somers
Simon Fraser University
Burnaby, British Columbia

A novel Bayesian technique is proposed to calculate 95% interval estimates for the size of the homeless population in the city of Edmonton using plant-capture data from Toronto, Canada. The probabilities of capture in Edmonton and Toronto are modeled as exchangeable in a hierarchical Bayesian model, and Markov chain Monte Carlo is used to sample from the posterior distribution. Guidelines are recommended for applying the method to assess the accuracy of homeless counts in other cities.

Keywords: Bayesian statistics, capture-recapture studies, Markov chain Monte Carlo, homelessness

Introduction

Estimating the size of street-dwelling homeless populations is important for city planning. However, it is a daunting task that is fraught with methodological and statistical challenges. One strategy is to use a homeless count with the help of volunteers. These volunteers serve as census takers, and their job is to walk throughout the city on predetermined walking routes and interview and count homeless people. For example, in the city of Edmonton, Canada, homelessness counts are conducted every 2 years during a single day in October. Table 1 describes the eight consecutive homeless counts in Edmonton between 1999 and 2012 (Homeward Trust Edmonton, 2012). Figure 1a plots the total number of homeless people that were counted during each year.

An astonishing fact about homelessness counts is that interval estimates (e.g. Bayesian 95% credible intervals (CIs)) for the true population size are rarely

Dr. McCandless is an Associate Professor in the Faculty of Health Sciences. Email him at: lmccandl@sfu.ca.

provided. For example, the most recent 2012 Edmonton homeless count identified a total of 1070 street-dwelling homeless individuals (see Table 1). However, no interval estimate was provided. Furthermore, homeless counts are known to be notoriously inaccurate because they underestimate the population size (Hopper, Shinn, Laska, Meisner, & Wanderling, 2008; US Department of Housing and Urban Development, 2008). Homeless people can remain hidden and out of sight. The volunteers can make errors in judgement in determining who is homeless. The street count walking routes may not be sufficiently comprehensive and the number of volunteers may be too few. Variation in counts may also be related to the experience of volunteers and how they are trained. Thus plotted curve in Figure 1a should be interpreted with extreme scepticism because there is no uncertainty assessment, and it is difficult to judge the accuracy of the estimation.

An important strategy for counting homeless people is to use plant-capture studies (Schwarz & Seber, 1999; Laska & Meisner, 1993; Martin, Laska, Hopper, Meisner, & Wanderling, 1997; Goudie, Jupp, & Ashbridge, 2007; Hopper et al, 2008). It is a variation of capture-recapture that requires only a single capture. Fake homeless individuals called plants are placed at random locations across the city. The plants are trained to dress and behave in a manner that does not draw attention to themselves so they can blend in with the homeless population. They are assumed to be indistinguishable from other homeless individuals, so that their probability of capture is the same. After the homeless count is complete, the proportion of plants that were counted is examined, and these data are used to estimate the size of the entire homeless population. The plant-capture design is recommended by the United States Department of Housing and Urban Development (2008), which develops guidelines for counting homeless people in American cities.

The validity of the plant-capture methodology depends on several assumptions, and these are reviewed by Laska and Meisner (1993) and Martin et al. (1997). A stable, closed population of individuals is required, with no entry or exits. In practice, this is achieved by conducting the homeless count over a short period of time. Plants should have the same probability of capture as other homeless individuals, and, in particular, the presence of plants should not affect the probability of capture. Plant-capture studies also depend on the accuracy of the data collection. The volunteers must respect the study protocol regarding whom to approach and how to conduct the interview to ascertain homeless status. They should have access to all parts of the street walking routes and a clear understanding of the geography of the city and time restrictions. See Martin et al. (1997) for a review of the assumptions for homeless street counts.

Table 1. Description of homeless counts in Edmonton and Toronto

	Edmonton in 1999, 2000, …, 2012,	**Toronto in 2006, 2009, 2013**
# of street-dwelling homeless who were counted	1070 in 2012, 1533 in 2010; 1862 in 2008; 1774 in 2006; 1452 in 2004; 1213 in 2002; 650 in 2000; 611 in 1999	447 in 2013; 362 in 2009; 735 in 2006
Definition of homelessness	Asking individuals the question: Do you have a permanent residence to return to tonight?"	Any individual sleeping outdoors on the night of the survey
Description	A street count that involved approaching individuals along predetermined walking routes where homeless are known to congregate.	An outdoor survey where teams were instructed to stop everyone they encountered to ask screening questions that establish housing status.
Date, time, temperature and weather conditions	2012: October 16, 05:00 to 22:00, 11.5C, Clear skies;2010: October 5, 05:00 to 22:00, 10.5C, Clear skies; 2008:October 21, 05:00 to 22:00, 6C, Cloudy skies; 2006: 05:00 October 17 to 05:00 October 18, 0.4C, Clear skies; 2004: 04:30 October 19 to 04:30 October 20, 2.5C, Cloudy skies; 2002: 04:30 October 23 to 05:00 October 24, -3.5C, Clear skies; 2000: September 14, 24 hour period, Temperature and weather unknown; 1999: November 17, 24 hour period, Temperature and weather unknown	2013: April 17, 19:00 to 01:00, 7.5C, Rain showers; 2009: April 15, 19:30 to 11:59, 9C, No precipitation; 2006: April 19, 20:30 to 11:59, 13C, No precipitation
# volunteer enumerators	300 in 2012; 300 in 2010; 220 in 2008; 300 in 2006; 157 in 2004; 200 in 2002; 100 in 2000; 100 in 1999	569 in 2013; 458 in 2009; 750 in 2006
Population of city in 2006	739000	2500000
Area of city in 2006	684 km^2	1749 km^2
Plant capture study?	No	Yes

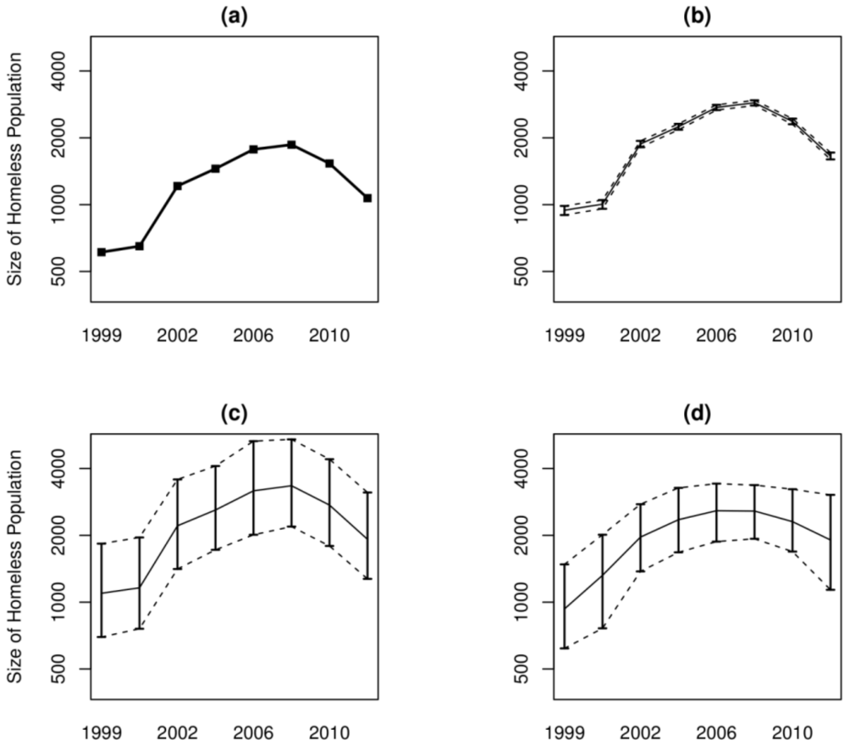

Figure 1. Estimated size of the homeless population in Edmonton

No plant-capture study has ever been done in Edmonton nor is any planned for the future. Homelessness counts are politically contentious, and controversy surrounds the costs and optics of paying individuals to pretend to be homeless. Thus when interpreting Figure 1a, the analyst is left with a basic research question: Is it possible to build interval estimates to quantify uncertainty in the population size? Is there data that allows us to estimate the proportion of homeless people that were counted during each year?

In this article, a novel Bayesian technique is proposed to calculate 95% interval estimates for the size of the homeless population in Edmonton using external data in the form of plant-capture studies from Toronto, Canada. The Bayesian approach is particularly well-suited to settings where multiple sources of information are available (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002; Gelman et al., 2013). Synthesizing data into a single model allows propagation of evidence and uncertainty about unknown quantities (Sweeting, De Angelis, Hickman, & Ades, 2008). This approach is an example of

Table 2. Plant-capture studies for the Toronto homelessness counts in 2006, 2009, and 2013

Year	Region	Plants Deployed	Plants Found	Proportion of Plants Found
2006	Toronto East-York	24	21	88%
	North York	13	7	54%
	Etobicoke	4	4	100%
	Scarborough	8	6	75%
	Total	50	26	52%
2009	Toronto East-York	17	9	53%
	North York	10	4	40%
	Etobicoke	6	5	83%
	Scarborough	12	8	67%
	Total	45	26	58%
2013	Toronto East-York	18	7	39%
	North York	10	7	70%
	Etobicoke	12	7	58%
	Scarborough	10	5	50%
	Total	49	38	78%

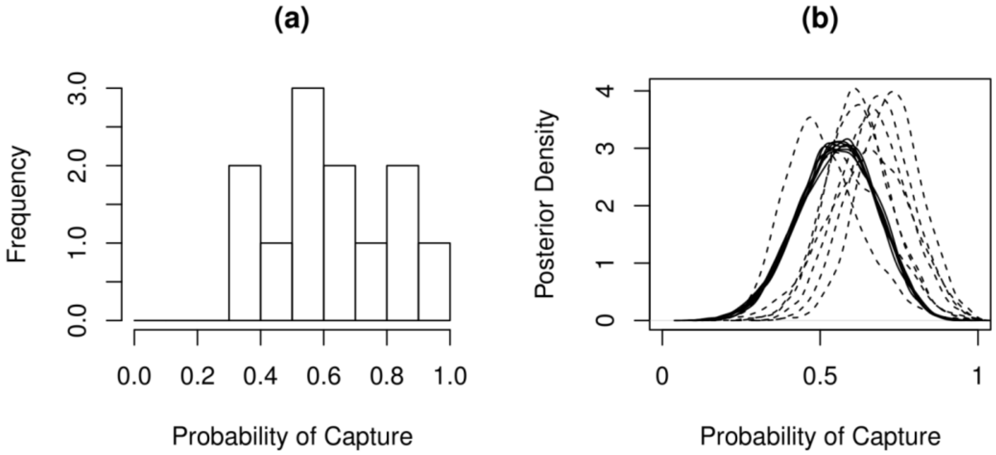

Figure 2. (a) Frequency histogram of the 12 proportions in Table 2. (b) Posterior distribution of the eight quantities $\mathbf{p_H} = (p_{H_{1999}},\ldots, p_{H_{2012}})$ calculated from the Bayesian analysis versus the Bayesian analysis with nonparametric regression for the population size

multiparameter evidence synthesis, which combines information from different datasets in order to estimate unknown parameters (Ades & Sutton, 2006).

To outline the proposed Bayesian methodology, consider the Toronto homeless data that is presented in Tables 1 and 2. In 2006, 2009 and 2013, homeless counts were conducted in Toronto, and they included plant-capture studies to estimate the probabilities of capture in Toronto (Toronto Shelter, Support and Housing Administration, 2013). Table 1 describes the homeless counts, and Table 2 summarizes the results of the plant-capture studies. Table 2 shows the number of plants that were deployed to each region of Toronto, by year, and it shows the proportion of plants that were captured. Figure 2a gives a histogram of the 12 proportions from Table 2. The proportions are heterogeneous and range from as low as 39% to as high as 100%. The mean is 65% and standard deviation is 19%. The heterogeneity is due to random error from the small number of plants in each region of Toronto, and additionally, due to variation in the probabilities of capture across space and time.

In this investigation, the histogram in Figure 2a is used to construct a prior distribution for the probability of capture for homeless people in Edmonton. Building on the work of Castledine (1981) and George and Robert (1992), the capture probabilities in Edmonton and Toronto are modelled as exchangeable in a hierarchical Bayesian model. They are treated as a random sample from a Beta distribution with unknown hyperparameters (Coull & Agresti, 1999; Pledger, 2005).The prior distribution expresses our initial beliefs about the probabilities of capture. It is updated using plant-capture studies from Toronto in order to obtain the posterior distribution for the unknown model parameters, including the size of the homeless population in Edmonton during each year.

This article describes the first example of a Bayesian analysis of plant-capture data, and it builds on the Bayesian literature for capture-recapture studies (e.g. Castledine, 1981; Smith, 1991; George & Robert, 1992; Fienberg, Johnson, & Junker, 1999; Basu & Ebrahimi, 2001; King & Brooks, 2001; Tardella, 2002; Manrique-Vallier & Fienberg, 2008; Corkrey et al., 2008). This article is organized as follows: First, the authors describe the methodology and modelling assumptions. An important issue in capture-recapture studies is understanding the role of heterogeneity in probability of capture between individuals (Burnham & Overton, 1978; Coull & Agresti, 1999; Link, 2003; Dozario & Royle, 2003; Pledger, 2005; Hwang & Huggins, 2005; Holzmann, Munk, & Zucchini, 2006; Farcomeni & Tardella, 2012), and this is discussed in the Statistical Models and Methods section. Next, the Results section is presented. The authors describe 95% CIs for the size of the homeless population in Edmonton. Further, the results of a simulation are presented that examines the sensitivity of the choice of prior distribution on the analysis results, including the coverage probability of interval estimates. A limitation of the analysis is that it ignores the fact that the size of the homeless population should change smoothly over time. Accordingly, in the final section of the Results, the authors incorporate a nonparametric regression model for the population size and study how this impacts uncertainty about the probability of capture. The article concludes with the Discussion section, and we provide guidelines for applying the method to assess the accuracy of homeless counts in other cities.

Statistical Models and Methods

Following Laska and Meisner (1993) and Martin et al. (1997), let H_i for $i \in \{1999, 2000, 2002, 2004, 2006, 2008, 2010, 2012\}$ denote the size of the finite population of homeless people in Edmonton during the homeless count in year i. Let n_{Hi} denote the number of homeless people who were counted in year i. Thus $n_{Hi} \leq H_i$. The quantity n_{Hi} is known, whereas H_i is unknown. The objective is to estimate H_i. The values of n_{Hi} are plotted over time in Figure 1a, and they are listed in the first row of Table 1. For example, $n_{H2012} = 1070$. Write $\mathbf{H}$ and $\mathbf{n_H}$ to denote vectors of the quantities H_i and n_{Hi} over i. Following Laska and Meisner (1993) and Martin et al. (1997), we model n_{Hi} using a Binomial distribution

$$n_{H_i} \sim \text{Binomial}\left(H_i, p_{H_i}\right) \tag{1}$$

with size H_i and proportion p_{Hi}. Let $\mathbf{p_H}$ denote the vector of p_{Hi} over index i.

The quantity p_{Hi} is defined as the average of the individual-level probabilities of capture among the H_i homeless people in Edmonton during year i. An important issue in the analysis of plant-capture data is understanding the role of heterogeneity in probability of capture between individuals (e.g. Burnham & Overton, 1978; Coull & Agresti, 1999; Link, 2003; Pedger, 2005). To illustrate the idea of heterogeneity, consider a hypothetical finite population of homeless individuals of size N. Suppose that each individual has only one opportunity for capture. Let $X_l = 1$ or 0, for $l = 1$ to N, be an indicator variable that indicates whether the lth individual was captured. Define $n = \sum_{l=1}^{N} X_l$ as the total number of homeless individuals who were captured. Additionally, let $\mathrm{P}(X_l = 1) = p_l$ denote the individual-level probability of capture, so that $X_l \sim \text{Bernoulli}(p_l)$. Further, suppose that the quantities $p_1, \ldots, p_N$ are independent and identically distributed with expected value $\mathrm{E}[p_l]$.

Then marginally, averaging over the probability distribution of p_l, we have X_l ~ Bernoulli(E[p_l]) and n ~ Binomial(N, E[p_l]).

Consequently, if one assumes that the detection of homeless individuals in Edmonton are treated as independent events, then this implies that the analyst can model the total number of homeless individuals who are counted in each year using (1), which is a binomial distribution with proportion p_{Hi} and no overdispersion. The quantity p_{Hi} depends on the calendar year i because the proportion of the population that is counted can vary from one year to the next. The Edmonton data are unique because each homeless individual has only one opportunity for capture in year i. In contrast, unmodelled heterogeneity in individual-level capture probabilities can greatly affect estimates of population size in capture-recapture studies because the same individual has multiple opportunities for capture (Burnham & Overton, 1978; Coull & Agresti, 1999; Link, 2003; Pledger, 2005). It can overstate precision about the population size (Link, 2003), and it can produce downward bias due to ignoring individuals with lower capture probabilities (Hwang & Huggins, 2005).

From (1), the conditional probability P(n_{Hi} | H_i, p_{Hi}) is

$$\mathrm{P}\left(n_{H_i} \mid H_i, p_{H_i}\right) = \binom{H_i}{n_{H_i}} p_{H_i}{}^{n_{H_i}} \left(1 - p_{H_i}\right)^{H_i - n_{H_i}} \tag{2}$$

The quantity H_i is large. If p_{Hi} is far from zero or one, then we can replace (2) with the normal approximation to the binomial distribution. The quantity n_{Hi} is modelled as normally distributed with mean $H_i p_{Hi}$ and variance $H_i p_{Hi}(1 - p_{Hi})$ which gives

$$\mathrm{P}\left(n_{H_i} \mid H_i, p_{H_i}\right) \propto \left\{2H_i p_{H_i}\left(1 - p_{H_i}\right)\right\}^{-1/2} \exp\left\{\begin{array}{c} -\left(2H_i p_{H_i}\left(1 - p_{H_i}\right)\right)^{-1} \\ \times\left(n_{H_i} - H_i p_{H_i}\right)^2 \end{array}\right\} \tag{3}$$

This Gaussian approximation can be used to accelerate Markov chain Monte Carlo (MCMC) computation.

The objective is to estimate H_i. A Bayesian approach is used to assign a hierarchical prior distribution to the capture probabilities p_{Hi} over i. To illustrate, write the joint probability density of the quantities (n_{Hi}, H_i, p_{Hi}) as

$$\mathrm{P}\left(n_{H_i}, H_i, p_{H_i}\right) = \mathrm{P}\left(n_{H_i} \mid H_i, p_{H_i}\right) \times \mathrm{P}\left(H_i, p_{H_i}\right)$$

Where P(H_i, p_{Hi}) is the joint prior distribution for H_i and p_{Hi}. Following George and Robert (1992) and Tardella (2002), the quantities H_i and p_{Hi} are assumed to be marginally independent a priori (i.e. that P(H_i, p_{Hi}) = P(H_i)P(p_{Hi})). There is no reason to believe that the probability of capture depends on the size of the homeless population.

To specify a prior P(H_i), this paper builds on the work of George and Robert (1992), who investigate different prior distributions for sample size in capture-recapture studies, including uniform priors. The following prior distribution is assigned

$$\mathrm{P}\left(H_i\right) \sim \text{Uniform}\left(n_{H_i}, M = 10000\right)$$

which is a uniform distribution for H_i that ranges from n_{Hi} to 10000. This prior ensures that H_i cannot be less than n_{Hi}. Additionally, it has upper limit M = 10000 to reflect the prior belief that the size of the homeless population cannot be greater than 10000 individuals. It is important that the prior distribution P(H_i) penalize large values H_i. The reason is because during joint estimation of (p_{Hi}, H_i) the MCMC samplers may fail to converge when p_{Hi} and H_i simultaneously tend to zero and infinity, respectively. Other alternative priors for H_i include the Jeffreys prior P(H_i) ∝ 1/H_i (Smith, 1991; George & Robert, 1992) or Rissanen's prior (Tardella, 2002).

To formulate a prior for p_{Hi}, plant-capture data from Toronto is incorporated using a Bayesian hierarchical model. The Bayesian approach is well-suited to settings where multiple data sources are available (Spiegelhalter et al., 2002; Gelman et al., 2013). Referring to the data in Table 2, let R_j denote the number of plants that were deployed in region j ∈ {East York in 2006, North York in 2006, Etobicoke in 2006, Scarborough in 2006, East York in 2009, North York in 2009, Etobicoke in 2009, Scarborough in 2009, East York in 2013, North York in 2013, Etobicoke in 2013, Scarborough in 2013}. Similarly, let n_{Rj} denote the corresponding number of plants that were subsequently captured during the Toronto homeless count. So for example, Table 2 illustrates that $R_{\text{Etobicoke in 2009}}$ = 6 and $n_{\text{Etobicoke in 2009}}$ = 5. A binomial model is assigned to n_{Rj}, which can be written as

$$n_{R_j} \sim \text{Binomial}\left(R_j, p_{R_j}\right)$$

where p_{Rj} is the capture probability of the R_j plants. The quantity p_{Rj} depends on j to reflect the fact that the probability of capture may vary by calendar year and region. Let **R**, $\mathbf{n_R}$, and $\mathbf{p_R}$ denote the vector of quantities R_j, n_{Rj}, and p_{Rj} over j.

For the Toronto data, both R_j and n_{Rj} are known, whereas p_{Rj} is unknown. A prior distribution for the unknown capture probabilities p_{Rj} and p_{Hi} is assigned over i and j by modelling the quantities as exchangeable within a hierarchical Bayesian framework. Following Gelman et al. (2013), a common Beta prior distribution is assigned

$$p_{R_j}, p_{H_i} \sim \text{Beta}(\alpha, \beta) \tag{4}$$

for all i, j, with unknown hyperparameters α and β. Beta priors are common in Bayesian analysis of Binomial proportions because they are conditionally conjugate. If the prior distribution for p_{Rj} or p_{Hi} is a Beta, the posterior will also be a Beta. This allows rapid updating of parameters during MCMC computation.

To complete the specification, a prior distribution is required for the unknown hyperparameters α and β. Following Gelman et al. (2013, Section 5.3), the following prior is assigned

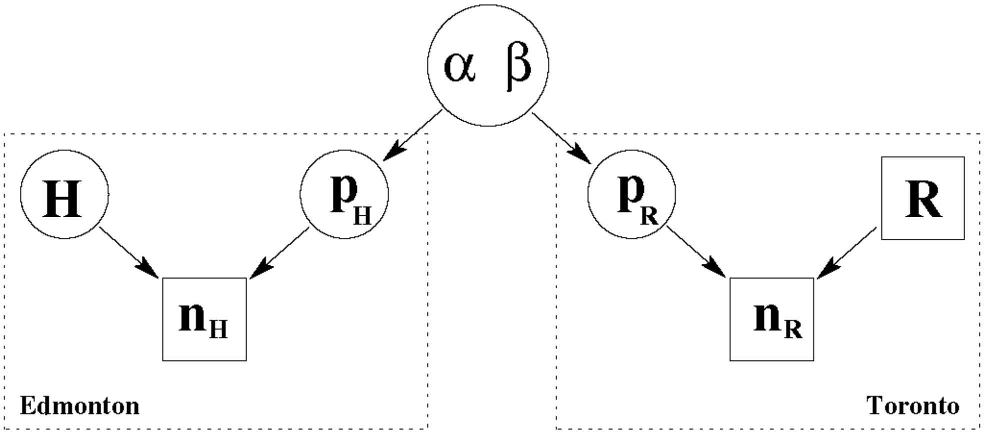

Figure 3. Probabilistic graphical model showing the conditional independence structure between data and unknown parameters in Edmonton and Toronto. Square boxes indicate quantities that are fixed and known, circles indicate unknown quantities. Our objective is to estimate **H** = (H_{1999},…, H_{2012}), the size of the homeless population in Edmonton for each year

$$\mathrm{P}(\alpha,\beta)\propto(\alpha+\beta)^{-\frac{5}{2}} \tag{5}$$

which yields a uniform prior on the standard deviation of the Beta distribution in (4).

The practical interpretation of our method is as follows: The collection of unknown probabilities of capture for Toronto and Edmonton is treated as a random sample from a Beta distribution with unknown hyperparameters α and β. The quantities α and β govern the shape of the distribution and, hence, the uncertainty of capture probabilities. Because one can estimate p_{Rj} for all j, this means that one can estimate α and β. Thus the hierarchical model imposes a probability distribution on p_H, which permits estimation of **H**. Figure 3 presents a probabilistic graphical model showing the conditional independence structure between data and unknown parameters in Edmonton and Toronto.

The full Bayesian model is written as follows: The joint probability density P($\mathbf{n_H}$, **H**, $\mathbf{p_H}$, $\mathbf{n_R}$, **R**, $\mathbf{p_R}$, α, β) is given by

$$\begin{aligned}\mathrm{P}(\mathbf{n_H},\mathbf{H},\mathbf{p_H},\mathbf{n_R},\mathbf{R},\mathbf{p_R},\alpha,\beta)=&\left[\prod_i \mathrm{P}(n_{H_i}\mid H_i,p_{H_i})\mathrm{P}(H_i)\mathrm{P}(p_{H_i}\mid\alpha,\beta)\right]\\&\times\left[\prod_j \mathrm{P}(n_{R_j}\mid R_j,p_{R_j})\mathrm{P}(R_j)\mathrm{P}(p_{R_j}\mid\alpha,\beta)\right]\\&\times\mathrm{P}(\alpha,\beta)\end{aligned}$$

The quantities ($\mathbf{n_R}$, **R**, $\mathbf{n_H}$) are observed, whereas (**H**, $\mathbf{p_H}$, $\mathbf{p_R}$, α, β) are unknown. The posterior distribution P(**H**, $\mathbf{p_H}$, $\mathbf{p_R}$, α, β | $\mathbf{n_R}$, **R**, $\mathbf{n_H}$) obeys the proportionality

$$\begin{aligned}\mathrm{P}(\mathbf{H},\mathbf{p_H},\mathbf{p_R},\alpha,\beta\mid\mathbf{n_R},\mathbf{R},\mathbf{n_H})\propto&\left[\prod_i \mathrm{P}(n_{H_i}\mid H_i,p_{H_i})\mathrm{P}(H_i)\mathrm{P}(p_{H_i}\mid\alpha,\beta)\right]\\&\times\left[\prod_j \mathrm{P}(n_{R_j}\mid R_j,p_{R_j})\mathrm{P}(p_{R_j}\mid\alpha,\beta)\right]\\&\times\mathrm{P}(\alpha,\beta)\end{aligned} \tag{6}$$

To fit the Bayesian model, MCMC is used in order to draw an approximate sample from the posterior distribution in (6). The yields a Markov chain with stationary distribution that is equal to the posterior distribution (Gelman et al., 2013). Using the MCMC sample the analyst can study the marginal posterior distribution of $\mathbf{H}$, denoted $P(\mathbf{H} \mid \mathbf{n_R}, \mathbf{R}, \mathbf{n_H})$, in order to estimate the size of the homeless population in Edmonton. Details of the MCMC algorithm are described in the Appendix. In the analyses that follow, the software R is used (R Development Core Team, 2013). Sampler convergence is assessed using multiple chains and diagnostic tools described by Gelman et al. (2013).

Results

Bayesian Estimation of the Size of the Homeless Population in Edmonton

A preliminary analysis is presented for the idealized scenario where the probability of capture is assumed to be exactly equal to 65% for each and every homeless count in Edmonton between 1999 and 2012. Recall from the Introduction that the value 65% is the sample average of the 12 proportions from Toronto listed in Table 2. Thus a naive estimate of $\mathbf{H}$ is obtained by ignoring uncertainty in the capture probabilities p_{Hi} and setting p_{Hi} = 65% for all i during MCMC computation. When p_{Hi} is fixed and known, then the analyst can sample from the posterior distribution in (6) by updating $\mathbf{H}$ from (A1) and ignoring $\mathbf{p_H}$ and (α, β) altogether.

Figure 1b gives posterior means and 95% highest posterior density CIs for $\mathbf{H}$. Recall that each component of $\mathbf{H}$ is the size of the homeless population in Edmonton during year i. Compared to Figure 1a, the resulting curve is shifted upwards to reflect that only 65% of the population was counted. The interval estimates are very narrow because we have fixed p_{Hi} = 65%.

Next, the full Bayesian analysis is fitted, which samples from the posterior distribution in (6) and estimates all unknown parameters. The results are plotted in Figure 1c, which depict posterior means and 95% CIs for $\mathbf{H}$. The point estimates are similar to those of Figure 1b, however the interval estimates are dramatically wider to reflect the uncertainty about the parameter vector $\mathbf{p_H}$.

To shed further light on the methodology, the solid curves in Figure 2b depict the posterior distribution of each of the eight quantities $\mathbf{p_H} = (p_{H1999},\ldots, p_{H2012})$, which are the average probabilities of capture in Edmonton during each of the eight homeless counts. The eight solid curves lie on top of one another, and they are a Beta approximation to the histogram in Figure 2a. The posterior mean of each quantity is roughly 55%, and the interquartile ranges are from 47% to 64%. Thus Figure 2 illustrates that the Bayesian method is working as expected. The uncertainty about the probabilities of capture in Edmonton translates into a broad range of uncertainty about the size of the homeless population, and this stretches the size of the interval estimates.

A Simulation Study to Examine the Sensitivity of the Prior Distribution on Analysis Results

A difficulty with the preceding analysis is that the results depend heavily on the prior distribution for $\mathbf{p_H}$. If the analyst chooses the “right prior” and the assumption of exchangeability between $\mathbf{p_H}$ and $\mathbf{p_R}$ is reasonable, then the interval estimates for the size of the homeless population in Edmonton will be suitably shifted towards the truth. However many things could go wrong. If the prior distribution for $\boldsymbol{p_H}$ is poorly chosen then the intervals will miss the truth entirely. Do 95% CIs have 95% frequentist coverage probability? To what extent will the coverage probability deteriorate through a careless choice of prior distribution for $\mathbf{p_H}$?

The coverage probability of 95% CIs is examined using a simulation study. In the Edmonton data example, the quantities $\mathbf{n_H}$, $\mathbf{R}$, and $\mathbf{n_R}$ are known. Suppose that $\mathbf{p_{H^*}}$ and $\mathbf{H^*}$ denote vectors of the true underlying probabilities of capture and true homeless population size for simulation purposes. A simulation is conducted as follows:

Table 3. Simulation study to examine the sensitivity of the prior distribution for $\mathbf{p_H}$ on the analysis results. Cells give the coverage probability of 95% CIs for the size of the homeless population in Edmonton for each year

Simulation #1 where the true capture probabilities are fixed as p_{Hi}* for each year

	Coverage probability of 95% CIs							
	1999	**2000**	**2002**	**2004**	**2006**	**2008**	**2010**	**2012**
Bayesian analysis assuming p_{Hi} = 65% and ignoring uncertainty	94.4%	95.0%	95.5%	95.2%	95.2%	95.0%	95.2%	93.9%
Bayesian analysis with hierarchical prior, which assumes that $\mathbf{p_H}$ and $\mathbf{p_R}$ are exchangeable	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Simulation #2 where the true capture probabilities are simulated as p_{Hi}* ~ Beta(α^* = 3.37, β^* = 1.84) for each year

	Coverage probability of 95% CIs							
	1999	**2000**	**2002**	**2004**	**2006**	**2008**	**2010**	**2012**
Bayesian analysis assuming p_{Hi} = 65% and ignoring uncertainty	12.1%	10.7%	7.9%	8.7%	6.6%	5.8%	7.3%	9.6%
Bayesian analysis with hierarchical prior, which assumes that $\mathbf{p_H}$ and $\mathbf{p_R}$ are exchangeable	83.4%	87.4%	81.6%	78.3%	80.5%	77.2%	79.3%	81.1%

1. Conduct a full Bayesian analysis of the homelessness data (**n$_{\mathbf{H}}$**, **R**, **n$_{\mathbf{R}}$**) to obtain 95% CIs, denoted I_{Hi}, for the size of the homeless population H_i in each year i.
2. For t from 1 to 1000:
   - a. For Simulation #1: Set the true probabilities of capture as $p_{H_i^*}^{(t)} = 65\%$ for each i.
   - b. For Simulation #2: Simulate $p_{H_i^*}^{(t)} \sim \text{Beta}\left(\alpha^* = 3.37, \beta^* = 1.84\right)$ for each i, which is a Beta distribution with mean 65% and standard deviation 19%.
   - c. Given $p_{H_i^*}^{(t)}$ and n_{Hi}, simulate the true homeless population size $H_i^{*(t)}$ from the conditional distribution $\mathrm{P}\left(H_i \mid p_{H_i^*}^{(t)} n_{H_i}\right)$ given in (A1) using MCMC.
   - d. Calculate the coverage indicator variable $Q_i^{(t)} = 1_{H_i^{*(t)} \in I_{H_i}}$ for each year i.
3. Calculate the average coverage probability $\left(1/1000\right)\sum_{t=1}^{1000} Q_i^{(t)}$ for each year i.

The results are given in Table 3. Simulation #1 considers the scenario where the true probabilities of capture are equal to 65% during each of the Edmonton homeless counts. As expected, the Bayesian analysis that correctly assumes p_{Hi} = 65% gives 95% CIs that have correct 95% coverage probability. The Bayesian analysis with hierarchical priors is too conservative and the coverage is 100% during each calendar year. In contrast, Simulation #2 describes the more realistic scenario where the true probabilities of capture **p$_{\mathbf{H*}}$** are heterogeneous and sampled from a Beta distribution with mean 65% and a standard deviation 20% (Gelman et al., 2013). Simulation #2 reveals that the hierarchical Bayesian model gives a large improvement in coverage probability compared to interval estimates that ignore uncertainty in the probability of capture.

Increasing Precision Using Bayesian Nonparametric Regression for the Population Size

One problem with Figure 1c is that the population sizes H_i are estimated independently. The inferences for H_i are driven entirely by n_{Hi} and p_{Hi} (see (A1)). But this ignores the reality that the population size should change smoothly over time. For example, if we know that H_{2004} = 2000, then can we not surmise that H_{2002} and H_{2006} are also close to 2000? This modelling information is ignored in Figure 1c. In other words, Figure 1c uses independent priors for each component of **H**.

To incorporate dependence in the prior for **H**, a model is required for the way in which the population size changes over time. Natural cubic splines are used (Gelman et al., 2013)

$$H_i \sim \mathrm{N}\left\{\sum_{k=1}^{3} \varphi_k \, \mathrm{g}_k\left(i\right), \sigma^2\right\}$$

with a single knot at i equal to the year 2005, which is the median of the collection of years. The quantities $\mathrm{g}_k(i)$ and φ_k are natural cubic spline basis functions and regression coefficients, respectively, and σ^2 is the unknown variance.

A relatively uninformative prior distributions is assigned to the regression parameters. The following prior is given to the coefficients

$$\varphi_1, \varphi_2, \varphi_3 \sim \mathrm{N}\left(0, 10^3\right)$$

and the variance is given the prior

$$\sigma^2 \sim \mathrm{Inv}\!-\!\chi^2\left(10^{-3}, 10^{-3}\right) \tag{7}$$

Write $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$. The posterior distribution becomes

$$\begin{aligned}
&\mathrm{P}\left(\mathbf{H}, \mathbf{p_H}, \mathbf{p_R}, \alpha, \beta, \boldsymbol{\varphi}, \sigma^2 \mid \mathbf{n_R}, \mathbf{R}, \mathbf{n_H}\right) \\
&\quad \propto \left[\prod_i \mathrm{P}\left(n_{H_i} \mid H_i, p_{H_i}\right) \mathrm{P}\left(H_i \mid \boldsymbol{\varphi}, \sigma^2\right) \mathrm{P}\left(p_{H_i} \mid \alpha, \beta\right)\right] \\
&\quad \times \left[\prod_j \mathrm{P}\left(n_{R_i} \mid H_i, p_{R_i}\right) \mathrm{P}\left(p_{R_i} \mid \alpha, \beta\right)\right] \\
&\quad \times \mathrm{P}\left(\alpha, \beta\right) \mathrm{P}\left(\boldsymbol{\varphi}\right) \mathrm{P}\left(\sigma^2\right)
\end{aligned} \tag{8}$$

To fit the regression model using MCMC, additional updates of $\boldsymbol{\varphi}$ and σ^2 are required. However, the required conditional distributions are available analytically using Bayesian linear regression. Details are given in the Appendix.

The results of fitting the model are given in Figure 1d. The posterior means of H are smoother than in Figure 1c because they have been shrunk together to fit the nonparametric curve. Interestingly, the interval estimates for H are sharply contracted compared to Figure 1c. When the analyst assumes that the population changes smoothly over time, then this gives more precise estimates of the population size because the regression model stabilizes the predictions.

Using a nonparametric curve to estimate the population size also implies a reduction in uncertainty about the probabilities of capture $\mathbf{p_H}$. This is illustrated in Figure 2b. The dashed curves plots the posterior distribution of each of the eight quantities $\mathbf{p_H} = (p_{H1999},\ldots, p_{H2012})$. The dashed curves are narrower than the solid curves. The locations of the curves are distorted to assist with fitting the nonparametric curve. This means that if the analyst assumes that the population size changes smoothly over time, then this induces a correlation among the p_{Hi} from one year to the next. The analysis with independent priors for **H** is too pessimistic about the magnitude of uncertainty about the probabilities of capture.

Discussion

The most recent homeless count in Edmonton occurred on October 17, 2012. A team of 300 volunteers found 1070 homeless people. Based on the Bayesian analysis that incorporates plant-capture data from Toronto, it is estimated that the true size of the homeless population is 2007 individuals with 95% Bayesian credible interval 1137 to 3042 (see Figure 1c). The city of Edmonton hopes to eliminate homelessness over the next decade, and an important question for government policy-makers is to determine whether the size of the homeless population is decreasing over time. The 2012 Edmonton Homeless Count Report states that "Between 2008 and 2012, the unsheltered homeless decreased by 30%" (Homeward Trust Edmonton, 2012). This calculation was based on the number of homeless people who were counted in 2012 (1070 individuals) versus 2010 (1533 individuals) because (1533-1070)⁄1533 = 30%. The estimation completely ignores uncertainty in the population size.

In contrast, the proposed Bayesian analysis directly contradicts this conclusion in the government report. The posterior mean of the ratio $(H_{2010} - H_{2012})/H_{2010}$, based on the Bayesian nonparametric regression analysis, is equal to 13% with 95% CI -40% to 58%. This implies a mere 13% reduction in the population size between 2010 and 2012, and there is a huge range of uncertainty and the interval estimate covers zero. Thus this analysis highlights the value of Bayesian uncertainty assessments when estimating the size of street-dwelling homeless populations. The failure to quantify uncertainty using posterior credible intervals can result in erroneous conclusions, which directly impact government policy decisions.

Our analysis depends on the assumptions that underlie plant-capture studies in general. See Laska and Meisner (1993) and Martin et al. (1997) for review. An important issue in the analysis of capture-recapture data is understanding the role of heterogeneity in probability of capture between individuals (Burnham & Overton, 1978; Coull & Agresti, 1999; Link, 2003; Pledger, 2005). In the analysis it is assumed that the homeless detections are independent events. As described the Statistical Models and Methods section, this assumption implies that the total number of homeless individuals who are counted in each year can be modelled using a binomial distribution with no overdispersion (see (1)). However, the assumption neglects the fact that homeless people usually live in groups (Martin et al., 1997). If homeless people aggregate into small groups, then the whole group is either spotted or lost. In principle, one could extend the modelling approach to model dependence in the probabilities of capture. For example, it is possible to model the probabilities of capture using a mixture of Beta distributions (Coull & Agresti, 1999). However, relaxing the independence assumption can cause the model to be nonidentifiable (Link, 2003).

More generally, the proposed Bayesian method can be used to quantify the accuracy of homeless counts in other cities. For example, Hopper et al. (2008) evaluated a plant-capture study of homelessness in New York City in 2005. The authors estimated the proportion of plants who were counted and, additionally, they conducted postcount interviews of homeless individuals to inquire about their whereabouts on enumeration night in order to establish if they were visible. A different example of plant-captures studies of homelessness is described by Martin (1992). In principle, these data could be used to assess the accuracy of homelessness counts in other American cities. Combining data from different cities requires a careful a careful examination of the exchangeability assumption.

References

Ades, A. E. & Sutton, A. J. (2006). Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society: Series A, 169*(1), 5-35. doi: 10.1111/j.1467-985X.2005.00377.x

Basu, S. & Ebrahimi, N. (2001). Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika*, *88*(1), 269-279. doi: 10.1093/biomet/88.1.269

Burnham, K. P. & Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, *65*(3), 625-633. doi: 10.1093/biomet/65.3.625

Castledine, B. J. (1981). A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika*, *68*(1), 197-210. doi: 10.1093/biomet/68.1.197

Corkrey, R., Brooks, S., Lusseau, D., Parsons, K., Durban, J. W., Hammond, P. S., & Thompson, P. M. (2008). A Bayesian capture–recapture population model with simultaneous estimation of heterogeneity. *Journal of the American Statistical Association*, *103*(483), 948-960. doi: 10.1198/016214507000001256

Coull, B. A. & Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture–recapture studies. *Biometrics*, *55*(1), 294-301. doi: 10.1111/j.0006-341X.1999.00294.x

Dorazio, R. M. & Royle A. J. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, *59*(2), 351-364. doi: 10.1111/1541-0420.00042

Farcomeni, A. & Tardella, L. (2012). Identifiability and inferential issues in capture-recapture experiments with heterogeneous detection probabilities. *Electronic Journal of Statistics*, *6*, 2602-2626. doi: 10.1214/12-EJS758

Fienberg, S. E., Johnson, M. S., & Junker B. W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society, Series A*, *162*(3), 383-405. doi: 10.1111/1467-985X.00143

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, A. B., Vehtai, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). New York, NY: Chapman Hall/CRC.

George, E. I. & Robert, C. P. (1992). Capture-recapture estimation via Gibbs sampling. *Biometrika*, *79*(4), 677-683. doi: 10.1093/biomet/79.4.677

Goudie, I. B. J., Jupp, P. E., & Ashbridge, J. (2007). Plant-capture estimation of the size of a homogeneous population. *Biometrika*, *94*(1), 243-248. doi: 10.1093/biomet/asm012

Holzmann, H., Munk, A., & Zucchini, W. (2006). On identifiability in capture–recapture models. *Biometrics, 62*(3), 934-936. doi: 10.1111/j.1541-0420.2006.00637_1.x

Hopper, K., Shinn, M., Laska, E. M., Meisner, M., & Wanderling, J. (2008). Estimating numbers of unsheltered homeless people through plant-capture and postcount survey methods. *American Journal of Public Health, 98*(8), 1438-1442. doi: 10.2105/AJPH.2005.083600

Homeward Trust Edmonton. (2012). 2012 Edmonton Homelessness Count. Retrieved from http://www.homewardtrust.ca/images/resources/2013-01-22-11-53FINAL%20%202012%20Homeless%20Count.pdf

Hwang, W. & Huggins, R. (2005). An examination of the effect of heterogeneity on the estimation of population size using capture-recapture data. *Biometrika*, *92*(1), 229-233. doi: 10.1093/biomet/92.1.229

King, R. & Brooks, S. P. (2001). On the Bayesian analysis of population size. *Biometrika*, *88*(2), 317-336. doi: 10.1093/biomet/88.2.317

Laska, E. M. & Meisner, M. (1993). A plant-capture method for estimating the size of a population from a single sample. *Biometrics, 49*(1), 209-220. doi: 10.2307/2532614

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics, 59*(4), 1123-1130. doi: 10.1111/j.0006-341X.2003.00129.x

Manrique-Vallier, D. & Fienberg, S. E. (2008). Population size estimation using individual level mixture models. *Biometrical Journal, 50*(6) 1051-1063. doi: 10.1002/bimj.200810448

Martin, E. (1992). Assessment of S-Night street enumeration in the 1990 Census. *Evaluation Review, 16*(4), 418-438. doi: 10.1177/0193841X9201600407

Martin, E., Laska, E. M., Hopper, K., Meisner, M., & Wanderling, J. (1997). Issues in the use of a plant-capture method for estimating the size of the street dwelling population. *Journal of Official Statistics, 13*(1), 59-74. Retrieved from http://www.jos.nu/Articles/abstract.asp?article=13159

Pledger, S. (2005). The performance of mixture models in heterogeneous closed population capture–recapture. *Biometrics*, *61*(3), 868-873. doi: 10.1111/j.1541-020X.2005.00411_1.x

Schwarz, C. J. & Seber, G. A. F. (1999). Estimating animal abundance: Review III. *Statistical Science, 14*(4), 427-456. Available from http://www.jstor.org/stable/2676809

Smith, P. J. (1991). Bayesian analyses for a multiple capture-recapture model. *Biometrika, 78*(2), 399-407. doi: 10.1093/biomet/78.2.399

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society: Series B, 64*(4), 583-639. doi: 10.1111/1467-9868.00353

Sweeting, M. J., De Angelis, D., Hickman, M., & Ades, A. E. (2008). Estimating hepatitis C prevalence in England and Wales by synthesizing evidence from multiple data sources. Assessing data conflict and model fit. *Biostatistics, 9*(4), 715-734. doi: 10.1093/biostatistics/kxn004

Tardella, L. (2002). A new Bayesian method for nonparametric capture-recapture models in presence of heterogeneity. *Biometrika, 89*(4), 807-817. doi: 10.1093/biomet/89.4.807

Toronto Shelter, Support and Housing Administration. (2013). *Street needs assessment results.* Retrieved from http://www.toronto.ca/legdocs/mmis/2013/cd/bgrd/backgroundfile-61365.pdf

United States Department of Housing and Urban Development. (2008). *A guide to counting unsheltered homeless people, 2nd Revision.* Retrieved from https://www.hudexchange.info/resources/documents/counting_unsheltered.pdf

Appendix

Bayesian Computation for Estimating the Homeless Population Size

The Metropolis Hastings algorithm is used to sample from the posterior distribution P(**H**, $\mathbf{p_H}$, $\mathbf{p_R}$, α, β | $\mathbf{n_R}$, **R**, $\mathbf{n_H}$) given in (6) by updating in blocks. This involves updating from the following conditional distributions

$$
\begin{aligned}
&\mathrm{P}\left(\mathbf{H}\,|\,\mathbf{n_H},\mathbf{p_H},\mathbf{n_R},\mathbf{R},\mathbf{p_R},\alpha,\beta\right)\\
&\mathrm{P}\left(\mathbf{p_H},\mathbf{p_R}\,|\,\mathbf{n_H},\mathbf{H},\mathbf{n_R},\mathbf{R},\alpha,\beta\right)\\
&\mathrm{P}\left(\alpha,\beta\,|\,\mathbf{n_H},\mathbf{H},\mathbf{p_H},\mathbf{n_R},\mathbf{R},\mathbf{p_R}\right)
\end{aligned}
$$

To update **H**, we have from (6)

$$
\begin{aligned}
&\mathrm{P}\left(\mathbf{H}\,|\,\mathbf{n_H},\mathbf{p_H},\mathbf{n_R},\mathbf{R},\mathbf{p_R},\alpha,\beta\right)\\
&\quad=\mathrm{P}\left(\mathbf{H}\,|\,\mathbf{n_H},\mathbf{p_H}\right)\\
&\quad\propto\mathrm{P}\left(\mathbf{n_H}\,|\,\mathbf{H},\mathbf{p_H}\right)\mathrm{P}\left(\mathbf{H}\right)\\
&\quad\propto\prod_i\mathrm{P}\left(n_{H_i}\,|\,H_i,p_{H_i}\right)\mathrm{P}\left(H_i\right)\\
&\quad\propto 1_{\left[n_{H_i},10000\right]}\times\prod_i\left\{2H_i p_{H_i}\left(1-p_{H_i}\right)\right\}^{-\frac{1}{2}}\\
&\qquad\times\exp\left\{-\left(2H_i p_{H_i}\left(1-p_{H_i}\right)\right)^{-1}\left(n_{H_i}-H_i p_{H_i}\right)^2\right\}
\end{aligned}
\tag{9}
$$

where the last line uses the Gaussian approximation to the binomial distribution given in (3). The quantity **H** is updated using a random walk Metropolis Hasting step with proposal distribution that is multivariate normal with mean that is a zero vector and variance that is equal to the identity matrix multiplied by a tuning parameter that is set by trial MCMC simulation runs. In principle, updating H could be improved by using a proposal distribution that approximates a negative binomial distribution (Castledine, 1981).

Updating $\mathbf{p_H}$ and $\mathbf{p_R}$ from P($\mathbf{p_H}$, $\mathbf{p_R}$ | $\mathbf{n_H}$, **H**, $\mathbf{n_R}$, **R**, α, β) is straightforward because the capture probabilities are conditionally conjugate under a Beta prior and Binomial model for $\mathbf{n_H}$ and $\mathbf{n_R}$. For all *i* and *j*, we have

$$
\begin{aligned}
\mathrm{P}\left(p_{H_i} \mid \mathbf{n_H}, \mathbf{H}, \mathbf{n_R}, \mathbf{R}, \alpha, \beta\right) &= \mathrm{P}\left(p_{H_i} \mid n_{H_i}, H_i, \alpha, \beta\right) \\
&= \mathrm{Beta}\left\{\alpha + n_{H_i}, \beta + \left(H_i - n_{H_i}\right)\right\} \\
\mathrm{P}\left(p_{R_j} \mid \mathbf{n_H}, \mathbf{H}, \mathbf{n_R}, \mathbf{R}, \alpha, \beta\right) &= \mathrm{P}\left(p_{R_j} \mid n_{R_j}, R_j, \alpha, \beta\right) \\
&= \mathrm{Beta}\left\{\alpha + n_{R_j}, \beta + \left(R_j - n_{R_j}\right)\right\}
\end{aligned}
$$

Hence updating $\mathbf{p_H}$ and $\mathbf{p_R}$ is accomplished by direct simulation from a vector of independent Beta random variables.

To update α and β, note that $\mathrm{P}(\alpha, \beta \mid \mathbf{n_H}, \mathbf{H}, \mathbf{p_H}, \mathbf{n_R}, \mathbf{R}, \mathbf{p_R}) = \mathrm{P}(\alpha, \beta \mid \mathbf{p_H}, \mathbf{p_R})$. Then

$$
\begin{aligned}
\mathrm{P}\left(\alpha, \beta \mid \mathbf{p_H}, \mathbf{p_R}\right) &\propto \mathrm{P}\left(\mathbf{p_H} \mid \alpha, \beta\right)\mathrm{P}\left(\mathbf{p_R} \mid \alpha, \beta\right)\mathrm{P}\left(\alpha, \beta\right) \\
&\propto \prod_i \left[\frac{\gamma(\alpha+\beta)}{\gamma(\alpha)\gamma(\beta)} p_{H_i}^{\alpha-1}\left(1-p_{H_i}\right)^{\beta-1}\right] \\
&\quad \times \prod_j \left[\frac{\gamma(\alpha+\beta)}{\gamma(\alpha)\gamma(\beta)} p_{R_j}^{\alpha-1}\left(1-p_{R_j}\right)^{\beta-1}\right] \\
&\quad \times (\alpha+\beta)^{-\frac{5}{2}}
\end{aligned}
$$

Given ($\mathbf{p_H}$, $\mathbf{p_R}$), the right hand side of this equation can be evaluated as a function of α and β. Updating from $\mathrm{P}(\alpha, \beta \mid \mathbf{n_H}, \mathbf{H}, \mathbf{p_H}, \mathbf{n_R}, \mathbf{R}, \mathbf{p_R})$ is achieved using a random walk Metropolis Hastings step with proposal distribution that is independent bivariate normal with mean zero and variance that is a tuning parameter set during initial MCMC runs.

Bayesian Computation for the Non-Parametric Regression Analysis

To sample from the posterior distribution in (8), the same MCMC procedure as the one described above is used except with additional updates of $\boldsymbol{\varphi}$ and σ^2. The required conditional distributions for $\boldsymbol{\varphi}$ and σ^2 are

$$
\begin{aligned}
&\mathrm{P}\left(\boldsymbol{\varphi} \mid \mathbf{H}, \mathbf{p_H}, \mathbf{p_R}, \alpha, \beta, \mathbf{n_R}, \mathbf{R}, \mathbf{n_H}, \sigma^2\right) \\
&\mathrm{P}\left(\sigma^2 \mid \mathbf{H}, \mathbf{p_H}, \mathbf{p_R}, \alpha, \beta, \mathbf{n_R}, \mathbf{R}, \mathbf{n_H}, \boldsymbol{\varphi}\right)
\end{aligned}
$$

Both of these distributions are conditionally conjugate based on the prior distributions in (7), and the analyst can sample from them directly using Bayesian linear regression (Gelman et al., 2013).