Bayesian Evidence Synthesis:opioid crisis

Hyeongcheol Park* & Paul Gustafson* & Micheal A Irvine*[1]
April 12, 2020

Contents

List of Figures

List of Tables

Abstract

1 Introduction

The opioid crisis is a major issue in North America including British Columbia, Canada. British Columbia is in the midst of a drug overdose crisis due to illicit opioids and the circumstances are getting worse. According to the BC Coroners Service, there were more than 930 apparent illicit drug overdose deaths in BC from Jan. 1 to Dec. 31, 2016. This compares to 518 in 2015, an increase of 79.2%. [fDC19] The goal of this project is to apply Bayesian evidence synthesis to better understand the opioid crisis in British Columbia, Canada.

One difficulty in coping with the opioid crisis is that the total number of overdoses is unknown. The reason is that the use of prescription opioids often leads to use of illicit opioids and the extent of illicit usage cannot be known. Many of those who become addicted to opioids do so after initially receiving a prescription. The highly addictive nature of these pain relievers makes it easy for the human brain to crave more. It is only after their prescription ends that many users realize they've become dependent on the effects of opioids to function "normally." At that point, they are either forced to get clean and endure the pain that comes with the withdrawal symptoms of opioids or look for another means of getting their high. This is often the time where people will turn to illicit drugs or other analogues. Because prescription opioids are so expensive, many users turn to heroin. It is often cheaper, more potent, and easier to locate than what they were taking before. In fact, about 80% of people using heroin started with a prescription to another opioid. After using heroin, however, 23% of individuals develop opioid addiction.[Smi19]

Since the total number of overdoses is unknown, this number needs to be estimated. Bayesian statistics can be a way to approach the problem and give us a good estimate of the number with a defensible range of uncertainty. Bayesian statistics is a theory in the field of statistics based on the Bayesian interpretation of probability where probability expresses a degree of belief in an event.

[wik20] Further explanation is provided at the following section. [todo: explain Bayesian statistics a little bit more here]

2 Methods

[I guess its done; phrase differently: the prior and the data relate to "all the unknown parameter", but as a last stage you focus on one target.]

[done?; following the comment, the first attempt:]

The number of overdoses is our ultimate target of estimation. To achieve the goal in context of Bayesian statistics, we use our prior belief and available data set about the target of estimation. The posterior belief is coming from both of these sources of the information. Let θ denote all the unknown variables and let Y denote all the related data sets. The following equation express the idea of Bayesian generative models in a mathmatical form:

$$p(\theta|Y) = \frac{p(\theta,Y)}{p(y)} = \frac{p(\theta)p(Y|\theta)}{p(Y)} \propto p(\theta)p(Y|\theta) \tag{1}$$

[done?: to make this flow better, maybe you want a version of (1) which generically describes parameter and latent varaibles given observed data.] where $p(\cdot)$ is probability density (or mass) function. The equation 1 states that the posterior belief of unobserved variables, $p(\theta|Y)$, is proportional to the mixture of the prior belief of the varaibles, $p(\theta)$, and the related data sets given the prior belief, $p(Y|\theta)$. Notice that the vector θ from equation 1 includes parameters and latent variables, and hence it contains the total number of overdoses.

Let us assume that we are interested in the opioid crisis from month 1 to month T. Let the subscript 1:T for any variable denote the variable of every target months and subscription t for any

variable denote the variable of the month t. Then the target variable, the total number of overdoses from month 1 to month T in British Columbia, is denoted by $O_{1:T}$. In Bayesian concepts, the variable of interest $O_{1:T}$ is a vector of fixed number but the numbers are yielded from a certain distribution where the distribution describes knowledge about $O_{1:T}$; that is, each O_t is considered to be a sample of a random variable. The simplest conceptual model is to take an underlying log-rate z_t for each month t where z_t is independent and identically distributed across months according to a normal distribution with mean μ and variance σ^2. [IM19] Then let λ_t^{OD} denote the rate of overdose at month t. It is assumed that the number of overdoses in the month, O_t, follows Poisson distribution with parameter $\lambda_t^{OD}N$ where the population of the region of interest is N. Note that the population size, N, is unknown and vary over time in reality but it is assumed known and constant for the investigator here for simplicity. The mechanism is named as overdose model and is illustrated in mathmatical form below:

$$\left.\begin{aligned} z_t &\sim N(\mu, \sigma^2) \\ \lambda_t^{OD} &= \exp(z_t) \\ O_t &\sim Poi(\lambda_t^{OD} N) \end{aligned}\right\} \text{ overdose model} \tag{2}$$

Since estimation of O_t is not straightforward as some of the variables (μ, σ) determining O_t is not known, we assume there was a survey conducted to help to support the inference about O_t. The survey asked witnesses of opioid overdoses whether someone called an ambulance for each month. This provides an estimate of the proportion of ambulance call $p_{A,t}$ among the overdoses. For the moment $p_{A,t}$ is assumed constant across time for simplicity. Let $n_{A,t}$ be the sample size of the survey at month t, and $x_{A,t}$ be the total number who confirmed that an ambulance was called at the month t. It is assumed that $x_{A,t}$ follows a Binomial distribution with number of trials, $n_{A,t}$, and success probability for each trial, $p_{A,t}$. The mechanism is named as ambulance call-outs model and is illustrated in mathmatical form below:

$$x_{A,t} \sim Bin(n_{A,t}, p_{A,t}) \text{ ambulance call-outs model} \tag{3}$$

Now the estimated $p_{A,t}$ can help to infer O_t if there is a data regarding the number of ambulance-attended overdoses. Let U_t be the number of ambulance-attended overdoses at a time point t. In general, the data of ambulance-attended overdoses U_t can be obtained. It is assumed that U_t follows a Binomial distribution:

$$U_t \sim Bin(O_t, p_{A,t}) \tag{4}$$

Note that O_t can be estimated because $p_{A,t}$ can be infered from the survey data and the data regarding U_t is given. Figure 1 shows an example of how the estimation can be proceeded. The variables that can be observed (N, $U_{1:T}$, and $x_{A,1:T}$) are drawn in the square boxes and the parameters (μ, σ^2) and the latent variables $(\lambda_{1:T}, O_{1:T}, p_{A,1:T})$ that are not observarble from possible data sets are drawn in the circle boxes. Variables that need prior distributions have asterisks on their draws. The arrows indicate which variables affects which variables; for example, there is an arrow from μ, σ^2 to $\lambda_{1:T}$ since μ, σ^2 is the parameters of $\lambda_{1:T}$, and the two arrows from $\lambda_{1:T}$ and N to $O_{1:T}$ indicate that both N and $\lambda_{1:T}$ affects the value of $O_{1:T}$ since the product of $\lambda_{1:T}$ and N is the parameter of the distribution of $O_{1:T}$.

We suggest a simple model in figure 1 as a start where the model only combines Overdose Model 2 and Ambulance Call-outs Model 3 with the relation 4. The next step is to run some simulations to figure out how different types of inputs lead some changes of output.

2.1 Generative Models

[**done?: comments: Before moving on to simulation, you should write out how the generative(generable) model factors, and mention Fig1 again.**] The amalgamation of the prior distribution and the statistical model can be referred to as the generative model.[reference] A generative

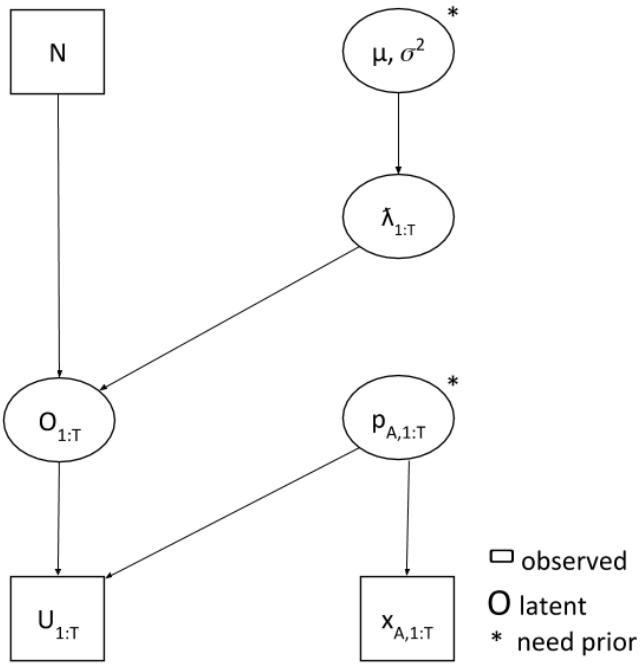


Figure 1: Example of estimating the total number of overdoses indirectly given two data sets: ambulance attended overdoses and survey data.

model described in figure 1 at a month t can be written as follows given some a priori assumptions:

$$\begin{aligned} p(\theta,Y) &= p(\mu,\sigma^2,\lambda_t,N,O_t,p_{(A,t)},n_{(A,t)},x_{(A,t)},U_t) \\ &= p(\mu)p(\sigma^2|\mu)p(\lambda_t|\mu,\sigma^2)p(N)p(O_t|\lambda_t,N)p(p_{(A,t)})p(n_{(A,t)})p(x_{(A,t)}|p_{(A,t)},n_{(A,t)})p(U_t|O_t,p_{(A,t)}) \\ &\propto p(\mu)p(\sigma^2)p(\lambda_t|\mu,\sigma^2)p(O_t|\lambda_t,N)p(p_{(A,t)})p(x_{(A,t)}|p_{(A,t)},n_{(A,t)})p(U_t|O_t,p_{(A,t)}) \end{aligned} \tag{5}$$

where $\theta = (\mu,,\sigma^2,\lambda_t,O_t,p_{(A,t)})$ and $Y = (N,n_{(A,t)},x_{(A,t)},U_t)$. The chain rule for random variables [Cha20] is used to the joint distribution of all variables at the first line of equation 5 such that the second line of the equation is derived. A priori independence assumptions and conditional independence assumptions were presumed so that the second line of the equation is simplified.

For example, it is asumed that σ^2 and μ are independent to each other and $p_{A,t}$ and also all the variables from overdose model 2, $(\mu, \sigma^2, \lambda_t, N, O_t)$, are independent to each other. Note that as a function of unknown variables, the function from the second line of the equation 5 is proportional to the function from the third line of the equation since N and $n_{(A,t)}$ are presumed to be known and constants such that the density functions of the known variables are merely constants. Therefore the third line of the equation is the ultimate generative model of interest.

2.2 Simulation

Simulation can be used to predict the performance of the Bayesian models. In order to measure the performance of the simplified model, a data set was simulated and the process is described in this chapter.

2.2.1 Likelihood

There exist two data sets: the survey data $(n_{A,1:T}, x_{A,1:T})$, and the ambulance attended overdose data $(U_{1:T})$. The two data sets were simulated as follows. It was assumed that the two data sets were collected for a year (T=12). In terms of the survey data, it is assumed that we have 12 survey data sets where each survey data set was collected for each month . The true value of $p_{A,t}$ was set as $p_{A,t} = 0.8$ and the survey sample size $n_{A,t}$ was set as $n_{A,t} = 1000$ for every months (t=1:T). The $x_{A,t}$ values for each month were independentally generated from the Binomial distribution (3). In terms of the overdose data, it is assumed that the total number of population were $N = 10000$ the true values of parameters for overdose model were $\mu = \log 0.05,\ \sigma = 1$. The vector of O_t was generated following the overdose model (2). The vector of U_t was generated from the Binomial relation of the two variables (4). The two generated vectors have the same length with the survey data (T=12). Note that only U_t and $x_{A,t}$ are the only known as the data set in real world and $p_{A,t}$

needs to be estimated first in order to estimate O_t which is the ultimate interest of the research.

2.2.2 Prior Distributions

Noninformative prior distributions are presumed as a start for simplicity.

$$p(p_A) \sim Beta(1,1) \text{ noninformative prior for ambulance model} \tag{6}$$

$$\left.\begin{aligned} \mu &\sim U(-10,0) \\ \sigma &\sim U(0,5) \end{aligned}\right\} \text{ noninformative prior for overdose model} \tag{7}$$

This leads the posterior distribution of variables of interest to heavily depend on the likelihood. Later, the noninformative priors will be changed and the impact of the changes to posteriors will be investigated.

2.2.3 Markov Chain Monte Carlo

[Why do we need MCMC? Because we don't have platinum, golden, and silver grails.]

Posterior distribution of the unseen variables, parameters $(\mu, \sigma^2, p_{A,1:T})$ and the latent variables $(\lambda_{1:T}, O_{1:T})$, given the seen variables, Y=$(N, n_{(A,1:T)}x_{(A,1:T)}, U_{1:T})$, is the key to make any inference for any of the unseen variables, θ. For the inference about the variable of interest, $O_{1:T}$, the marginal posterior distribution, $p(O_{1:T}|Y)$, is essential which can be obtained by integrating the joint posterior distribution, $p(\theta|Y)$, over all the other unseen variables except $O_{1:T}$. However, in general it is severely complicated and almost impossible to perform such multiple integrations because $p(\theta|Y)$ is a complex distribution of high dimensions. Moreover, $p(O_{1:T}|Y)$ may not have

closed form nor follow a known parametric distribution, which lead the general inference procedure laborious.

Markov Chain Monte Carlo (MCMC) is a detour to obtain posterior samples from $p(\theta|Y)$ without explicitly knowing the distribution. MCMC technique allows us to generate Monte Carlo realizations of samples, $\theta_{(i)}$, which are dependent, with the distribution of each sample, $p(\theta_{(i)})$, converging to $p(\theta|Y)$.[Gus20]

Generative model auto pilot conceptunally there's only one distribution of unseen stuff given the seen stuff we can use the distribution to make whatever inference claims that we need to make, such as point estimation, credibal interval, inference about parameters or latent variables.

[Two types of bronze grails, NUTS, metropolis]

[explain all the figures below was derived by NUTS, and python]

The achievement of the simulation was done by Markov chain Monte Carlo (MCMC). The technique is a popular method to obtain posterior samples from distributions where analytic forms do not exist or are hard to be derived. **[Further explanation is going to be added.]** [attempt:brief explaination of MCMC and related topics]

2.3 Initial Result

[**combine the two paragraph into one.**] All examples here were performed in Python 3.7 using the library pyMC (reference) and JAGS (reference). Training was performed using No U-Turn Sampling (NUTS) over two chains with 1000 iterations (is it sample size?). Fitting was performed on a GHz Intel Core i5 with 8GM of LPDD3 RAM and typically had wall times under ten minutes. Data processing was carried out using the Pandas and SciPy library [reference]. Data visualization was performed using the libraries Seaborn and Matplotlib [ref]. Code for all examples in this study are provided.

The result from the simple case scenario is illustrated below. Markov chain Monte Carlo pack-

ege PYMC3 3.8[JTC16] was used with Python 3.7 for the simulation. Figures were generated by Python 3.7. The results were verified with JAGS 4.3.0 and R 3.6.2.

2.3.1 Posterior Distribution

Figure 2 is the boxplot of posterior samples of O_t. It is shown that our posterior estimates of O_t are fairly accurate since (1) the boxplots contain actual values of O_t within their interquartile range (IQR) and (2) the ranges of IQR and 95% range seem narrow. Notice that the range of the boxplot from a higher O_t values (t=1, 7, 11) is wider than the other ranges of the boxplots from smaller estimates of O_t (all t values but 1, 7, 11)

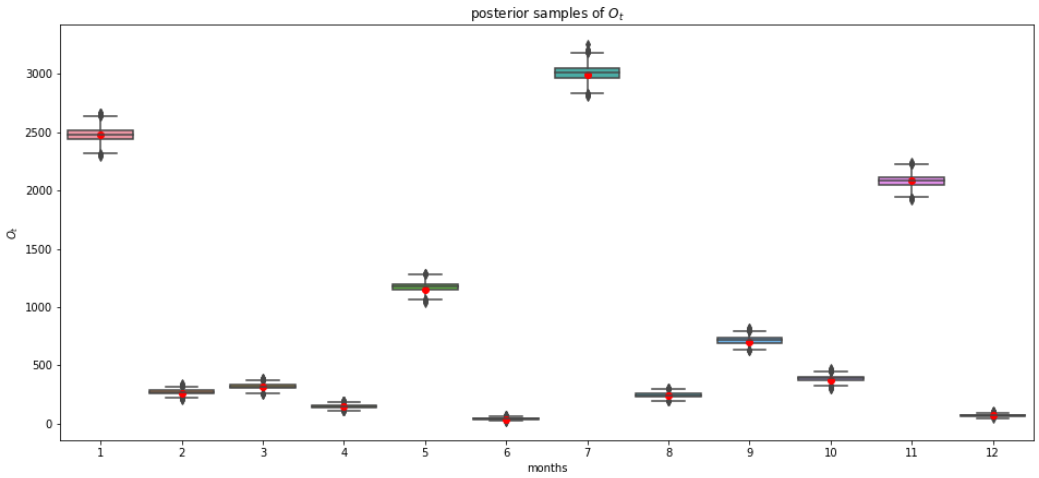


Figure 2: Boxplot of posterior samples of O_t (2000 samples for each month) with simulated O_t values as red dots.

Let O_+ denote the total overdose per year ($O_+ = \sum_{t=1}^{12} O_t$). Figure 3 is the histogram of posterior samples of the total number of overdoses O_+ per year. 2000 samples were collected and shown as the histogram. It is seen that the plot is bell shaped and not skewed. The total number of overdoses per year is inferred to be around 10900 to 11000. [mention about credible interval]

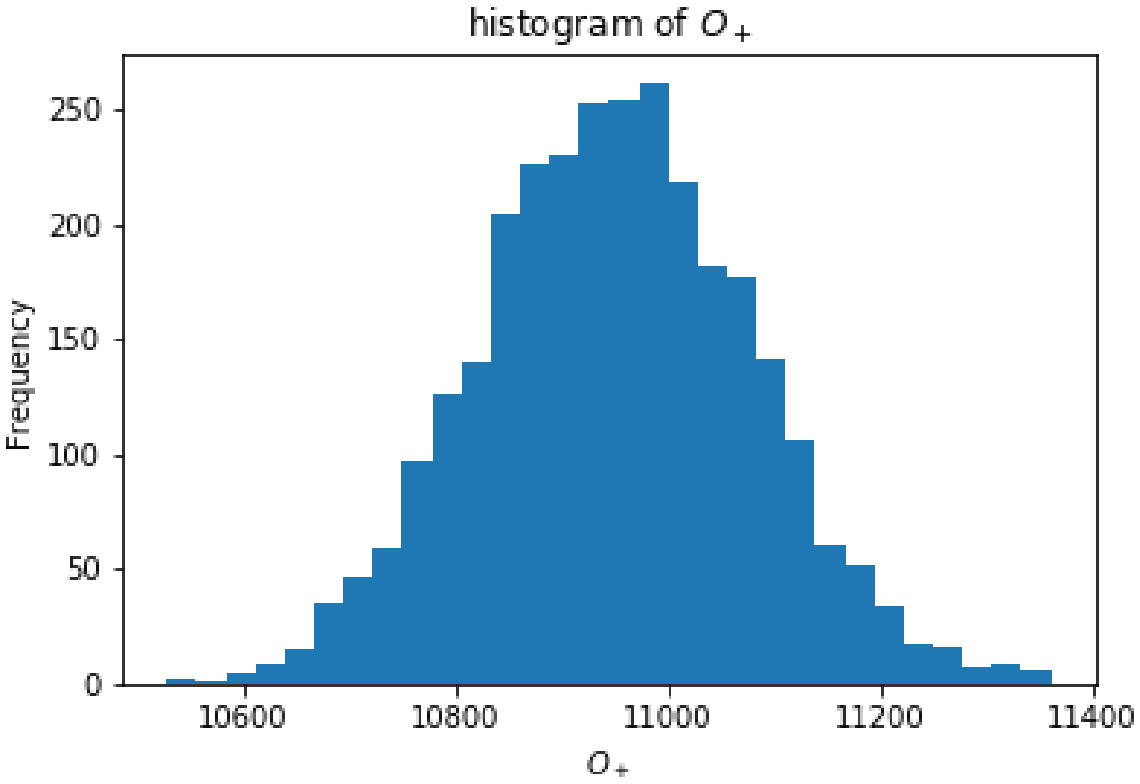
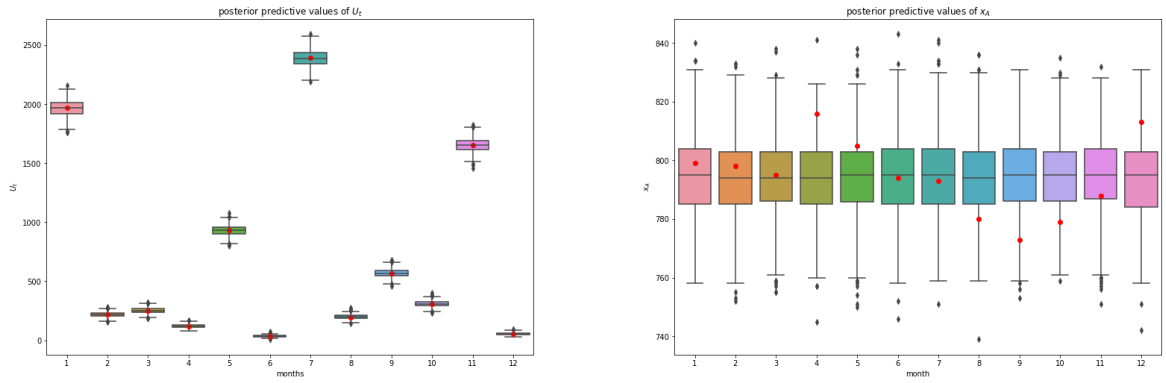


Figure 3: Histogram of posterior samples of total overdoses per year

2.3.2 Posterior Predictive Check

Posterior predictive checking is a model validation technique when we simulate some replicated data under the fitted model then compare the new data to the observed data. If the model fits, then replicated data generated under the model should look similar to observed data. To put it another way, the observed data should look plausible under the posterior predictive distribution. This is really a self-consistency check: an observed discrepancy can be due to model misfit or chance. [GCS+14] [attempted: brief explannation of what is PPC]

Figure 4a is the boxplot of posterior predictive samples of U_t. It is shown that the posterior predictive estimates of U_t are failry accurate, for the same two reasons regarding the accuracy of the posterior distribution of O_t. It is more obvious here that the range of the boxplots from higher O_t values (t=1, 7, 11) is wider than the other ranges of the boxplots from smaller estimates of O_t (all t values but 1,7, 11). Notice that the relative ranges of figure 4a follow the ones of figure 2; for months where U_t values are higher, the values of O_t are also higher than the average (t=1, 7, 11).



(a) ppc of U_t

(b) ppc of p_A

Figure 4: Boxplots of posterior predictive samples of U_t and x_A (1000 samples) with the actual data point of simulated U_t,x_A value. The simulated values are shown as a red dots.

Figure 4b is the boxplots of posterior predictive samples of x_A. It is shown that the posterior predictive estimates of x_A are tolerably accurate since (1), the boxplots contain actual values of O_t within their lines connecting the maximum and the minimum, and (2), the ranges of IQR seem narrow.

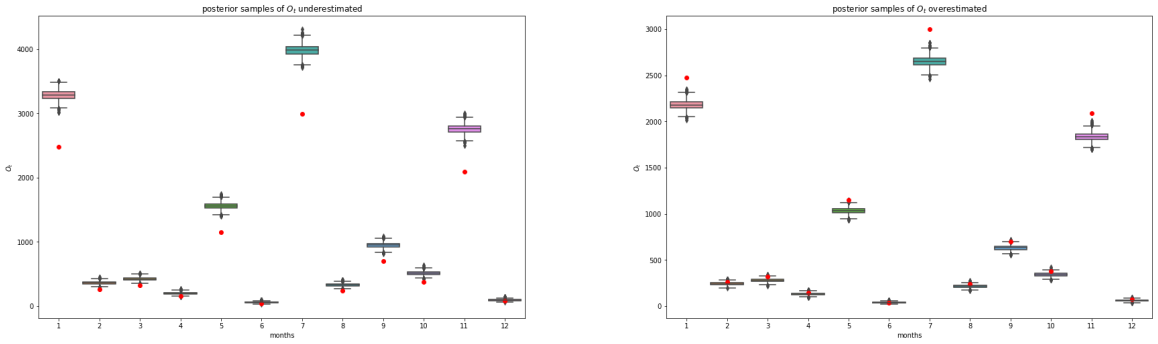
2.4 Initial Result: Contamination of p_A

One focus of this research project is to investigate how robust the model is to contaminations of the data sets. The first inspection is to check an impact of a contamination of p_A; what would happen if the estimation of p_A is biased? It is assumed that the survey data gives us a wrong estimate of p_A such that it would be underestimated or overestimated. We then want to see how the biased estimation of p_A affects the estimate of O_t, the total number of overdoses.

Both underestimation and overestimation were considered for the analysis. In terms of underestimation, the simulated survey data (n_A, x_A) was generated with $p_A = 0.6$ while the true value of p_A is 0.8, and all the other assumptions hold as before. That is, x_A is generated from $x_A \sim Bin(n_A, 0.6)$, while U_t is generated from $U_t \sim Bin(O_t, 0.8)$ for every t. For overestimation, the simulated survey data was generated with $p_A = 0.9$ while the true value of p_A is 0.8, and all the other assumptions hold the same.

2.4.1 Posterior Distribution

Figure 5 is the boxplots of posterior samples where p_A is contaminated from the survey data. It is seen that the underestimation of p_A (5a) leads to an overestimation of O_t as the boxplots are above



(a) ppc of O_t: underestimated p_A

(b) ppc of O_t: overestimated p_A

Figure 5: Boxplot of posterior samples of O_t (1000 samples) where survey data is contaminated. The actual data point of simulated O_t values are shown as red dots.

the red dots. This is explainable considering the given data sets (likelihoods) and the relationship between the two models (4); given the likelihood of the ambulance attended overdoses, U_t, a value of O_t is an unknown parameter of the binomial distribution, $Bin(O_t, p_A)$. Let o_t, u_t denote posterior samples from O_t and the observed value of U_t respectively. Then o_t is proportional to the quantity generated by multiplying u_t and the inverse of $\hat{p}_A$ where $\hat{p}_A$ is underestimated estimator of p_A; hence $\frac{1}{\hat{p}_A}$ is overestimated which leads overestimation of O_t:

$$o_t \propto u_t \frac{1}{\hat{p}_A} \tag{8}$$

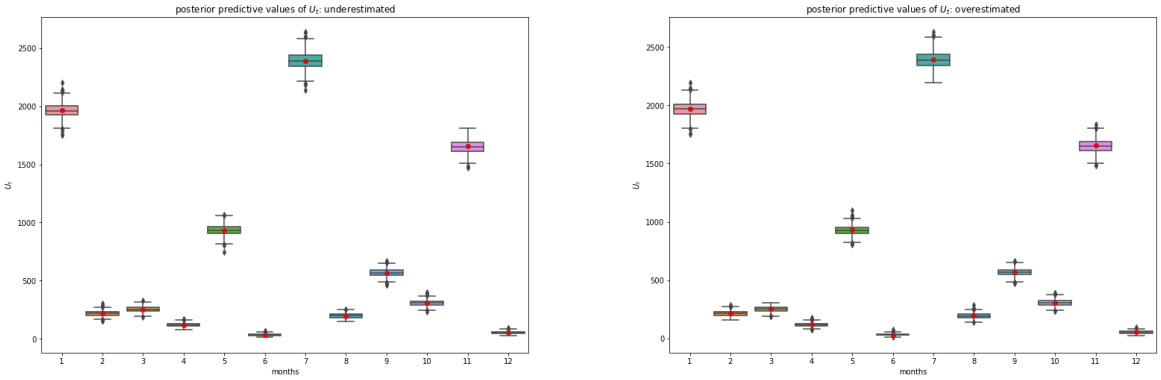
Figure 5b shows the opposite case. Overestimation of p_A leads underestimation of the inverse of p_A which causes underestimation of O_t. From both figures it is seen that the bias increases as the estimated values and the actual values get large.

2.4.2 Posterior Predictive Check

Figure 6 gives the boxplots of posterior predictive samples where p_A is underestimated and overestimated respectively from the survey data. It is seen that none of the contaminations of p_A leads to an contamination effect on U_t but only affects the O_t estimation.

The possible explanation is that the contaminations of p_A and O_t (from p_A) cancel out the bias component as combined so that U_t has no bias. Notice that U_t is a likelihood (data set) of O_t estimation. Hence it plays the role of y in Bayes theorem (**??**) and the likelihood, U_t, is used to obtain the biased posterior samples of O_t with the contaminated estimates of p_A. Here the pair of (O_t, P_A) are both biased and the pair is fitted by the likelihood, U_t. To obtain the posterior predictive samples, MCMC algorithm help obtain posterior samples of biased O_t and then the samples will be combined with biased p_A such that the new samples of $\tilde{U}_t$ from $p(\tilde{U}_t|U_t)$ is obtained. The result also matches with our intution. Posterior predictive checks use the existing data points twice; firstly it uses the data to obtain the posterior samples of paramaters, and secondly the samples are used to produce posterior predictive samples. Given the fact that there could be an overfitting issue due to using the data twice, the result shown here seems reasonable.

Figure 7 gives the boxplots of posterior predictive samples of x_A where p_A is underestimated and overestimated respectively from the survey data. It is seen that none of the contaminations of p_A leads an effect of contamination on x_A; the actual simulated points (red dots) are close to the medians from the two boxplots. However, notice that the range of the estimated values are different between the two boxplots; the median from Figure 7a is around 600 whereas the median from Figure 7b is around 900. [???? I think this should go to the previous of the previous paragraph] This is because overdose model (2) does not affect the resulf from the ambulance model (3); only the ambulance model has an effect on the overdose model.



(a) ppc of U_t: underestimated p_A

(b) ppc of U_t: overestimated p_A

Figure 6: Boxplot of posterior samples of U_t (1000 samples) where survey data is contaminated. The actual data point of simulated U_t values are shown as red dots.

2.5 Variying the Assumptions

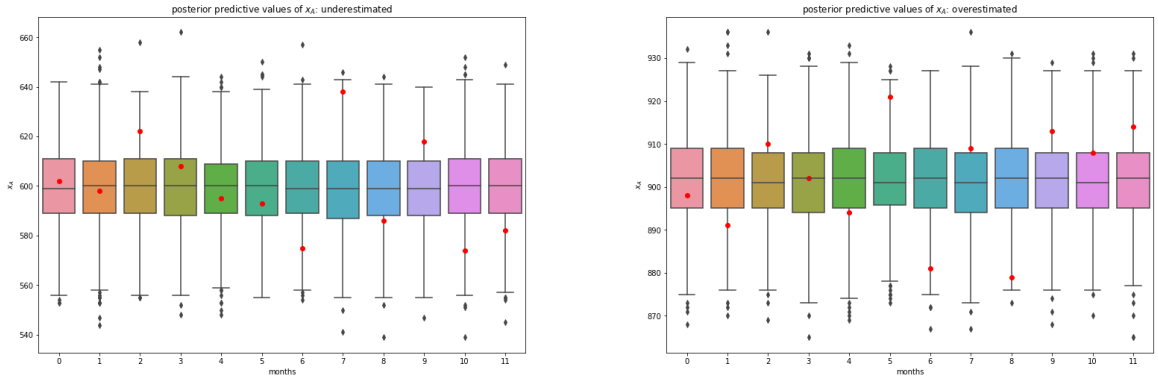
The assumptions will change later to see the impact of the likelihood over the posterior distributions of variables of interest; the total population size for a region, N, could vary over time or it can be staritified for a better realization of the reality. The survey size could also vary such as $n_A = 100$ or $n_A = 10000$ in later examples.

References

[Cha20] Chain rule (probability). Chain rule (probability) — Wikipedia, the free encyclopedia, 2020.

[fDC19] BC Center for Disease Control. BC Overdose Prevention Services Guide, 2019.

[GCS+14] Andrew Gelman, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin. *Bayesian Data Analysis*. CRC Press, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742, 2014.



(a) ppc of x_A: underestimated p_A

(b) ppc of x_A: overestimated p_A

Figure 7: Boxplot of posterior samples of x_A (1000 samples) where survey data is contaminated. The actual data point of simulated x_A values are shown as red dots.

[Gus20] Paul Gustafson. Lecture note 1 in STATISTICAL THEORY FOR THE DESIGN AND ANALYSIS OF CLINICAL STUDIES: Bayesian Inference in Latent Variable Models, March 2020.

[IM19] Buxton J Balshaw R Otterstatter M Macdougall L et al. Irvine MA, Kuo M. Modelling the combined impact of interventions in averting deaths during a synthetic-opioid overdose epidemic. *Addiction*, 2019.

[JTC16] Salvatier J., Wiecki T.V., and Fonnesbeck C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2016.

[Smi19] Cooper Smith. What Is the Opioid Epidemic?, 2019.

[wik20] Bayesian statistics— Wikipedia, the free encyclopedia, 2020.