# BAYESIAN EVIDENCE SYNTHESIS:OPIOID CRISIS

HYEONGCHEOL PARK* & PAUL GUSTAFSON* & MICHEAL A IRVINE*[1]

## CONTENTS



## LIST OF FIGURES

## LIST OF TABLES



## ABSTRACT

* *Department of Statistics, University of British Columbia, Vancouver, Canada*
[1] *Department of Mathmatics, University of British Columbia, Vancouver, Canada*

# 1 INTRODUCTION

Opioid crisis is one of major issues in North America continents including Canada. There were 1,490 deaths and 15,598 paramedic- attended overdose events during 2017 alone. [1] (need to know about bib in latex, change statistics to 2018 later) The goal of this project is to apply Bayesian evidence synthesis to help reduce the effect of opoid crisis in Vancouver, Canada.

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.

# 2 METHODS

The number of overdoses is our ultimate interest of estimation. Let $O_t$ the number of overdose in a given month t. Suppose there was a survey conducted to estimate the proportion of ambulance call $p_A$ among the subjects of overdoses. Let $n_A$ the sample size of the survey and $x_A$ to be the total number who confirmed they did call ambulance. It is assumed that $x_A$ follows Binomial distribution.

$$x_A \sim \mathrm{Bin}(n_A, p_A) \text{ ambulance call-outs model} \tag{1}$$

The total overdoses need to be modeled. The simplest conceptual model is to take an underlying log-rate $z_t$ that is independent and identically distributed according to a normal distribution with mean $\mu$ and variance $\sigma^2$. [1] Denote $\lambda_t$ the rate of overdose at time t. It is assumed that the total overdose $O_t$ follows Poission distribution where the population of the region of interest is N.

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

Estimation of $O_t$ is not straightforward since none of the variables ($\mu$, $\sigma$, N) determining $O_t$ is known. Hence $O_t$ should be inferred from using $U_t$ and $p_A$, where $p_A$ is the ambulance call out rate and $U_t$ is 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 Binomial distribution:

$$U_t \sim \mathrm{Bin}(O_t, p_A) \tag{3}$$

Now $O_t$ can be estimated as $p_A$ can be infered by survey data and the data regarding $U_t$ is given. We suggest a simple model as a start where the model only combines Ambulance Call-outs Model (1) and Overdose Model (2).

The next step is to run some simulations to figure out how different types of inputs lead some changes of output. To do so, the simple model illustrated below.

## 2.1 Simulation

The first simulation simplifies the assumptions of variables as much as possible; We assumed $N = 10000, n_A = 1000$. The assumptions will change later to see the impact of the likelihood over the posterior distributions of variables of interest; The total number of population for a region N could vary over time or it can be staritified for a better realization of the real world. $n_A$ can be vary as $n_A = 100$ or $n_A = 10000$.

#### 2.1.1 *Likelihood*

There exist two data sets; survey data $(n_A, x_A)$, and ambulance attended overdose data $(U_t)$. The two data set is simulated as follows. The true value of $p_A$ was set $p_A = 0.8$ for the survey data. It is assumed that the data was collected for a year (t=1,2,3, ..., 12) and $x_t$ values were independentally generated from the Binomial distribution (1). It is assumed that 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 gerated from the Binomial relation of the two variables (3). The two generated vectors have the same length with the survey data (t=1,2,3, ..., 12).

Note that only $U_t$ and $x_t$ are known as the likelihood and $p_A$ needs to be estimated first so as to estimate $O_t$ which is the ultimate interest of the research.

#### 2.1.2 *Prior Distributions*

Noninformative prior distributions are presumed as a start for simplicity.

$$p(p_A) \sim Beta(1,1) \text{ noninformative prior of ambulance model} \tag{4}$$

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

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 over posteriors will be investigated.

### 2.2 Early Result

The result from the simple case scenario is illustrated below.

#### 2.2.1 *Posterior Distribution*

Figure 1 is the boxplot of posterior samples of $O_t$. It is shown that our posterior estimates of $O_t$ is 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 covering the actual values of $O_t$. Notice that the range of the boxplot from a higher $O_t$ values (t=4) is wider than the other ranges of the boxplots from smaller estimates of $O_t$ (all t values but 4)

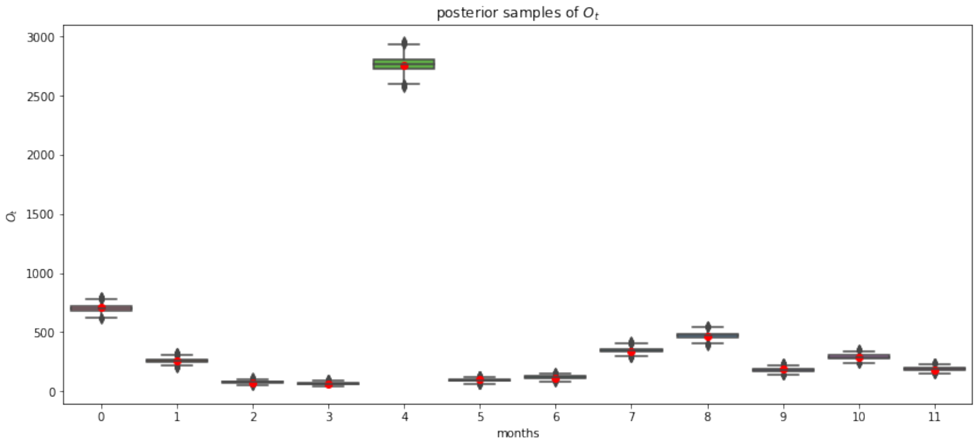



**Figure 1:** Boxplot of posterior samples of $O_t$ (2000 samples for each month) with actual data points of simulated $O_t$ values. The simulated values are shown as red dots.

### 2.2.2 *Posterior Predictive Check*

Figure2 is the boxplot of posterior predictive samples of $U_t$. It is shown that the posterior predictive estimates of $U_t$ is failry accurate with 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)
.

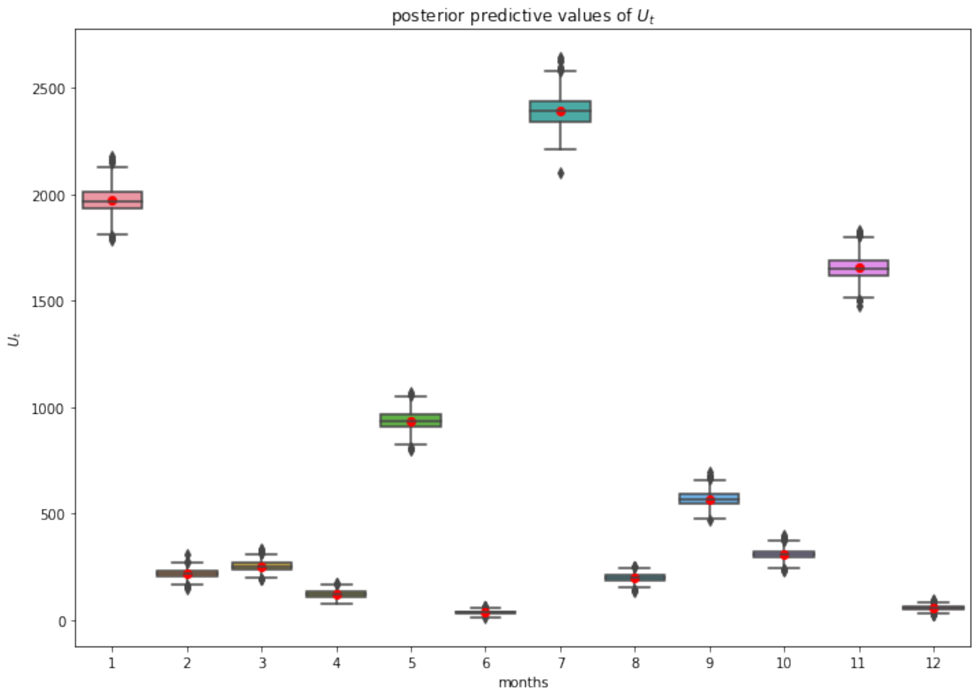



**Figure 2:** Boxplot of posterior predictive samples of $U_t$ (2000 samples for each month) with actual data points of simulated $U_t$ values. The simulated values are shown as red dots.

Figure 3 is the boxplot of posterior predictive samples of $x_A$. It is shown that the posterior predictive estimates of $x_A$ is failry accurate with the same two reasons regarding the accuracy of the posterior distribution of $x_A$.

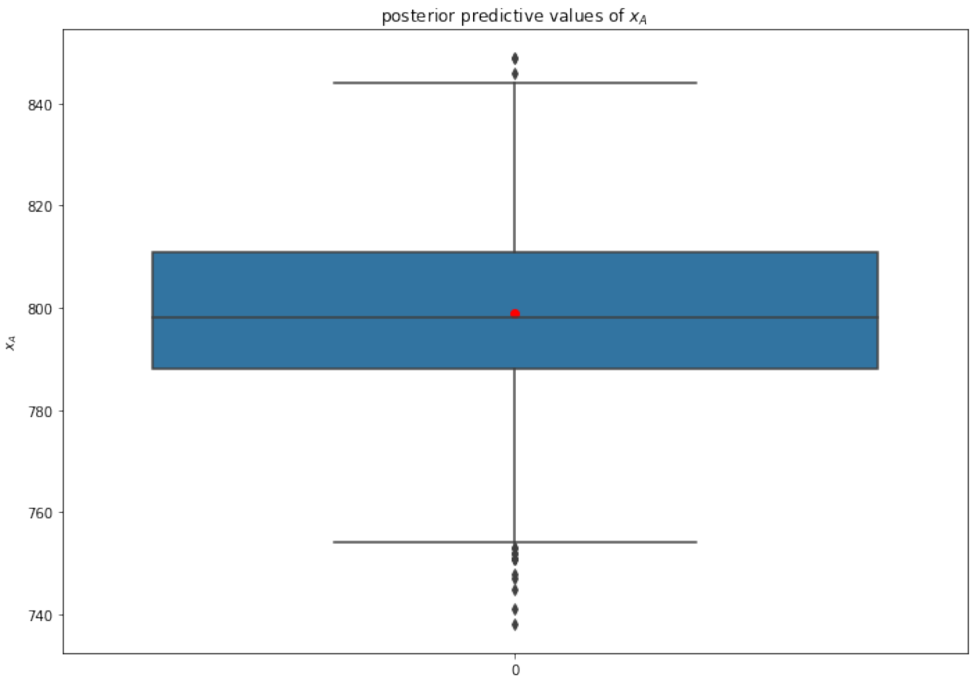



**Figure 3:** Boxplot of posterior predictive samples of $x_A$ (1000 samples) with the actual data point of simulated $x_A$ value. The simulated value is shown as a red dot.

## 2.3 Early Result: Contamination of $p_A$

One of the attention of this research project is to investigate how robust the model is from a contamination of the data set. 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 overdose.

Both of underestimation and overestimation were conducted 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 the same. 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.

### 2.3.1 *Posterior Distribution*

Figure 4 is the boxplot of posterior predictive samples where $p_A$ is underestimated from the survey data. It is seen that the underestimation of $p_A$ leads to an overestimation of $O_t$ as the boxplots are above the red dots. This is justifiable considering the given data sets (likelihoods) and the relationship between the two models (3); $O_t$ is generated by multiplying $U_t$ and the inverse of $p_A$ where $p_A$ is underestimated. This leads overestimated inverse of $p_A$ so that $O_t$ becomes overestimated. Figure 5 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.

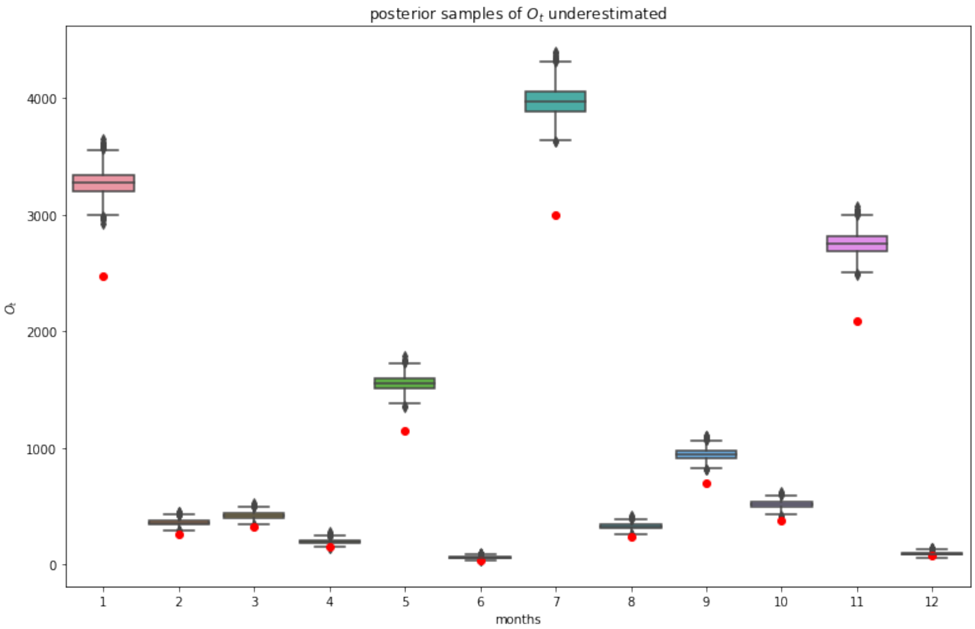



**Figure 4:** Boxplot of posterior samples of $x_A$ (1000 samples) where survey data is contaminated (underestimation of $p_A$.) The actual data point of simulated $x_A$ values are shown as red dots.

### 2.3.2 *Posterior Predictive Check*

Figure 6 and 7 are the boxplots of posterior predictive samples where $p_A$ is underestimated and overestimated respectively from the survey data. It is seen that the none of the contaminations of $p_A$ leads an effect $U_t$.

Figure 8 and 9 are the boxplots of posterior predictive samples where $p_A$ is underestimated and overestimated respectively from the survey data.

# 3 TEMPLATE

We first focus on the simplest sitauation that can describe the data sets and the general idea.

1. First item in a list
2. Second item in a list
3. Third item in a list

## 3.1 Paragraphs

**PARAGRAPH DESCRIPTION**

**DIFFERENT PARAGRAPH DESCRIPTION**

## 3.2 Math

$$\cos^3\theta = \frac{1}{4}\cos\theta + \frac{3}{4}\cos 3\theta \tag{6}$$

**Definition 1** (Gauss)**.** To a mathematician it is obvious that $\int_{-\infty}^{+\infty} e^{-x^2}\,dx = \sqrt{\pi}$.

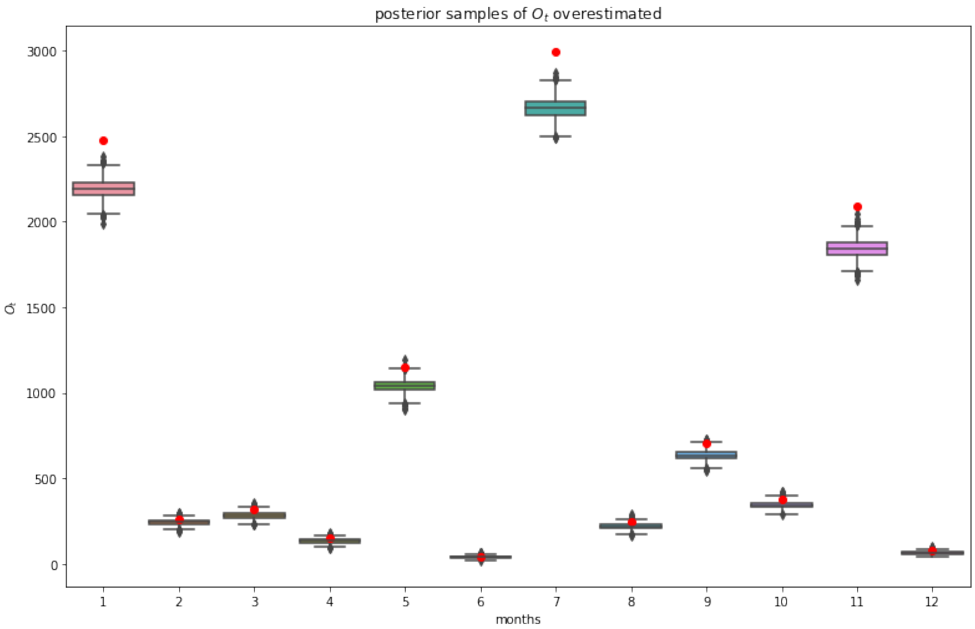



**Figure 5:** Boxplot of posterior samples of $x_A$ (1000 samples) where survey data is contaminated (overestimation of $p_A$). The actual data points of simulated $x_A$ values are shown as red dots.

**Theorem 1** (Pythagoras)**.** *The square of the hypotenuse (the side opposite the right angle) is equal to the sum of the squares of the other two sides.*

*Proof.* We have that $\log(1)^2 = 2\log(1)$. But we also have that $\log(-1)^2 = \log(1) = 0$. Then $2\log(-1) = 0$, from which the proof. $\square$

# 4 RESULTS AND DISCUSSION

Reference to Figure 10 on page 11.

## 4.1 Subsection

### 4.1.1 *Subsubsection*

**WORD** Definition

**CONCEPT** Explanation

**IDEA** Text

- First item in a list
- Second item in a list
- Third item in a list

### 4.1.2 *Table*

Reference to Table 1 on the next page.

## 4.2 Figure Composed of Subfigures

Reference the figure composed of multiple subfigures as Figure 11 on page 12. Reference one of the subfigures as Figure 11b on page 12.

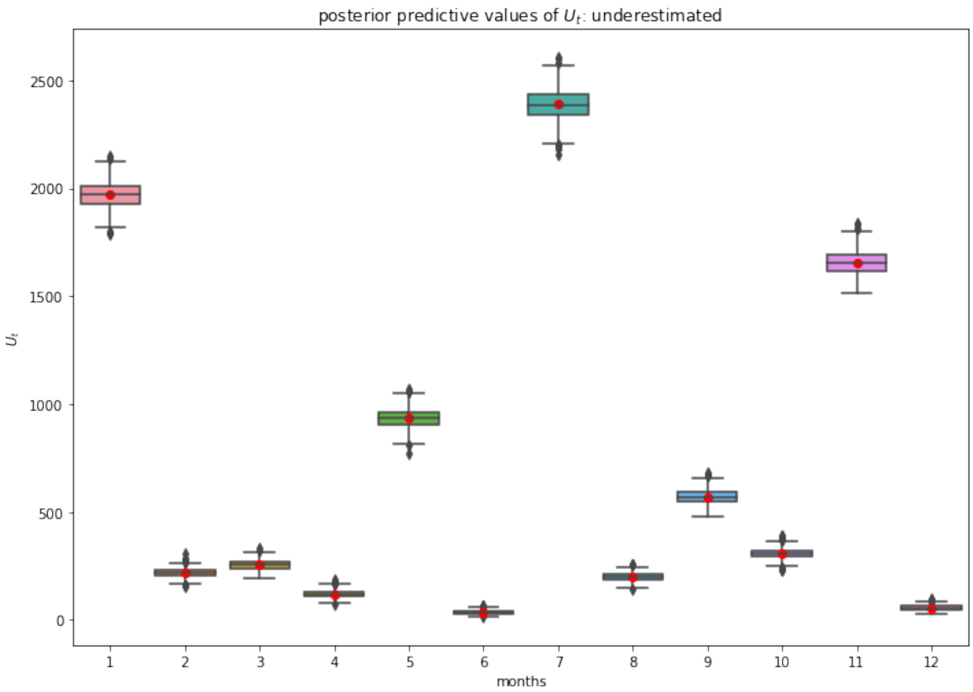



**Figure 6:** Boxplot of posterior predictive samples of $x_A$ (1000 samples) where survey data is contaminated (underestimation of $p_A$). The actual data point of simulated $x_A$ values are shown as red dots.

**Table 1:** Table of Grades

| Name | | |
|---|---|---|
| First name | Last Name | Grade |
| John | Doe | 7.5 |
| Richard | Miles | 2 |

## REFERENCES


[1] 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.

[2] A. J. Figueredo and P. S. A. Wolf. Assortative pairing and life history strategy - a cross-cultural study. *Human Nature*, 20:317–330, 2009.

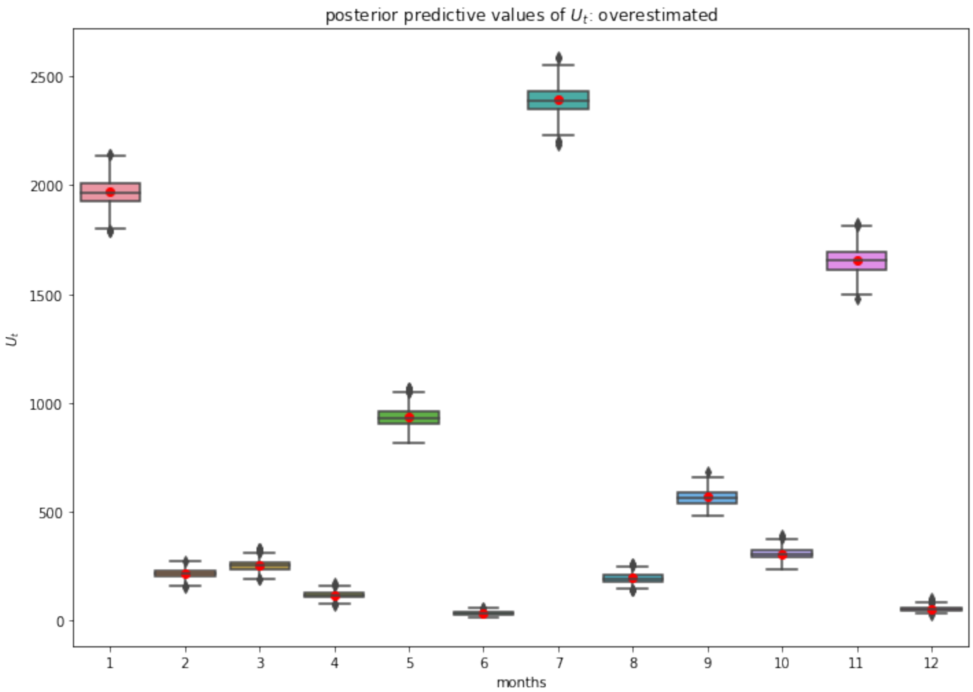



**Figure 7:** Boxplot of posterior predictive samples of $x_A$ (1000 samples) where survey data is contaminated (overestimation of $p_A$). The actual data point of simulated $x_A$ values are shown as red dots.

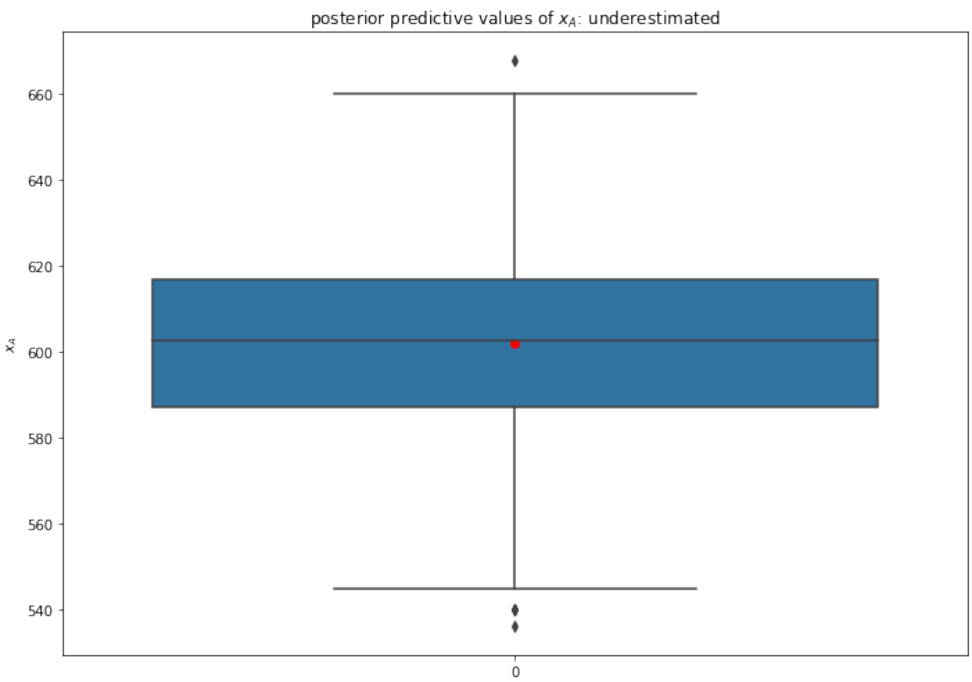



**Figure 8:** Boxplot of posterior predictive samples of $x_A$ (1000 samples) where survey data is contaminated (underestimation of $p_A$). The actual data point of simulated $x_A$ value is shown as the red dot.

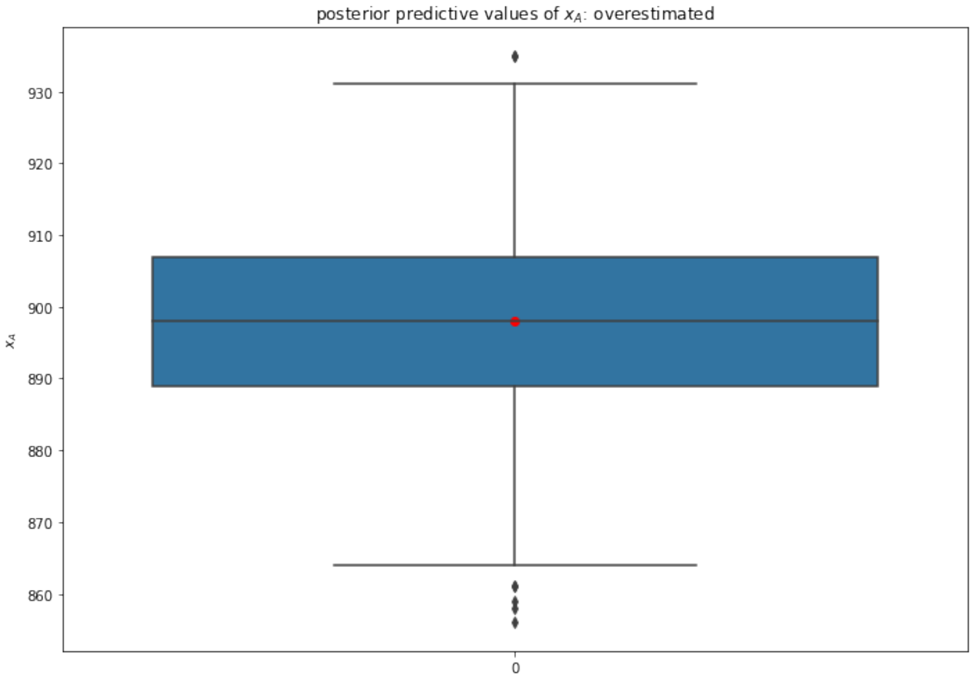



**Figure 9:** Boxplot of posterior predictive samples of $x_A$ (1000 samples) where survey data is contaminated (overestimation of $p_A$). The actual data point of simulated $x_A$ value is shown as the red dot.

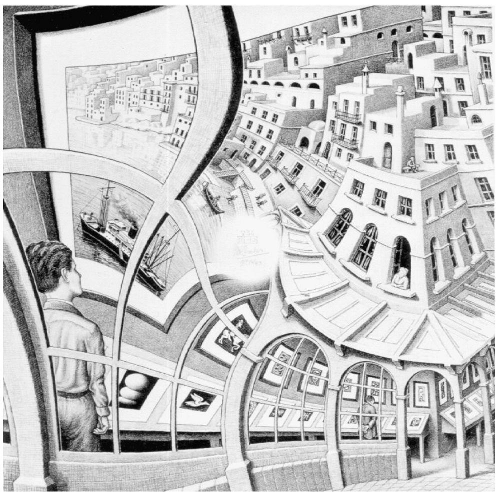
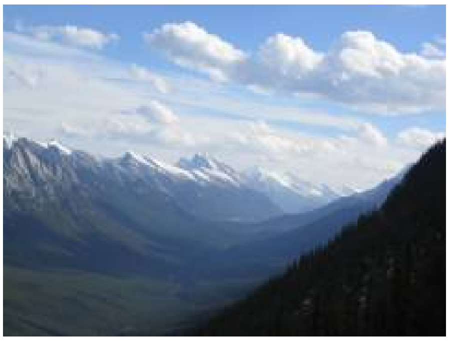


**Figure 10:** An example of a floating figure (a reproduction from the *Gallery of prints*, M. Escher, from `http://www.mcescher.com/`).

(a) A city market.

(b) Forest landscape.



(c) Mountain landscape.

(d) A tile decoration.

Figure 11: A number of pictures with no common theme.