Bayesian Sensitivity Analysis Using E-value

Wu Xue

George Washington University
marcoxue516@gwu.edu

Abbas Zaidi

Facebook
abbaszaidi@fb.com

Abstract

Statistical inference in the presence of missing outcome data is an inevitability in almost any application such as those in the social sciences or medical research. However, the quality of inference in these settings rests on strong but unfortunately untestable assumptions on the missingness mechanism. In order to ensure that inference is reliable, *Sensitivity Analysis* is a necessary step to assess robustness against violations of untestable assumptions. Using motivating examples from *Facebook* conversion data, we present methodology for conducting an E-value based Sensitivity Analysis at scale with three novel contributions. First, we develop a means for the Bayesian estimation of sensitivity parameters from privacy focused noisy aggregates with empirically derived and objective priors. Second, resting on the estimation of the sensitivity parameters we develop a mechanism for posterior inference via simulation of the E-value. Finally, we derive closed form distributions for the E-value (under a range of assumptions) to make direct inference possible for cases where posterior simulation may be infeasible due to computational constraints. We demonstrate gains in performance over asymptotic inference of the E-value using a data-based simulation supplemented by a case-study on partially missing Facebook conversion data.

Keywords: Sensitivity Analysis; E-value; Missing data; Empirical Bayes

1 Introduction

An increasing number of statistical methods have been developed to handle missing data; an inevitability in applications such as those in the social sciences and medical research. Some of these methods are ad hoc solutions relying on strong assumptions about the missingness mechanism that are implausible in practice, such as the analysis of complete cases. Model-based approaches such as multiple imputation (MI) and inverse probability weighting (IPW) are proposed to deal with missing data problems under less restrictive assumptions. Both techniques leverage the *Missing At Random* (MAR) assumption on the missingness mechanism which states that missingness is *random* conditional on the observed features.

While it is reasonable to assume MAR, the possibility of data being *Missing Not at Random* (MNAR) can never be fully excluded; When data is potentially MNAR, methods such as IPW would yield inconsistent (and non-identified) estimates. However, since the missingness mechanism is inherently a statement about *unobserved* data, direct validation is not possible; this necessitates indirect evaluation to understand the impacts of MNAR on conclusions. One mechanism for doing so is *Sensitivity Analysis* [24].

A myriad of research [13, 22] has been undertaken on Sensitivity Analysis for missing data problems. The representative monograph [18] introduces several methods for doing so, such as the pattern mixture model and selection model approaches as the dominant classifications of methods. Related techniques have also been used in Causal Inference [3, 6, 25] with some applications that may be extended to missing data mechanisms. For example, [25] proposed to report the E-value to show how robust the causal effect estimates are against potentially unmeasured confounding. Treating the MAR assumption as a form of ignorability, we will discuss how the E-value is applicable to missing data problems [23].

In order to effectively utilize the E-value to understand the implications of MAR violations, estimating its uncertainty is crucial; an area that is both underdeveloped and necessary. In this paper we propose to cast the estimation and uncertainty quantification of the E-value as a *Bayesian Inference* problem. Conceptually, these results are rooted in the Bayesian estimation of the *sensitivity parameter* using a combination of *noisy benchmarks* and *prior information.* These benchmarks are becoming increasingly common in applications where aggregated information is used as a means of protecting privacy at the unit-of-observation level (the objective of inference). Examples of these include but are not limited to *Google's Privacy Sandbox* [10], *differentially private aggregates* [20] or Facebook's *Aggregated Event Measurement* (AEM) system. Under this overarching theme, the Bayesian estimation of the sensitivity parameter induces a posterior distribution for the E-value that can either be approximated using simulation [9] or analytically determined under a series of testable assumptions where simulation is infeasible.

The remainder of the paper proceeds as follows. Section 2 provides a brief overview of conducting sensitivity analysis in missing data problems. In Section 3, a Bayesian approach with objective and empirically elicited subjective priors for the sensitivity parameter is proposed to conduct sensitivity analysis. Section 4.1 presents a simulation study to evaluate the performance of the proposed method. In Section 4.2, this technique is applied to the validation process for IPW estimates from *Facebook* data; an example motivated by methodology used on the platform for tackling missing data. Finally, section 5 concludes the paper with a discussion of our findings as motivation for future work.

2 Background

Pattern Mixture Models and Selection Models are two widely used approaches when assessing whether inference resting on MAR assumptions are robust against mechanisms truly being MNAR. Both approaches are based on factorization of joint distribution of the measurement and the missingness mechanism. Pattern Mixture Models factorize the aforementioned joint distribution as the product of conditional distribution of the measurement given the missingness mechanism and the marginal distribution of the missingness mechanism [16, 17, 19]. By comparison, Selection Models express the joint distribution as the product of conditional distribution of missingness mechanism given the measurement and marginal distribution of the measurement [5, 11, 12].

When the underlying missingness mechanism is potentially MNAR, explicit modeling of the mechanism (parametric or nonparametric) needs to be taken into account. Selection Models are appealing since they represent the marginal distribution of the full measurement in the model. However, this requires assuming an explicit but untestable model for the missingness mechanism. Pattern-mixture models by comparison place assumptions about the missingness mechanism via restrictions on the sensitivity parameter [15], which are more explicit than with Selection Models. One of the drawbacks of Pattern Mixture Models is that marginal distribution of the full measurement is not explicitly represented and may induce more complexity in derivation. It warrants mention that assumptions about the missingness mechanism necessary for either approach are not directly verifiable from the data; Pattern Mixture Models are uniquely suited for our motivating applications [24].

Earlier work on Sensitivity Analysis using Pattern Mixture Models relied on expert knowledge to determine the plausible range of the sensitivity parameter values i.e. on average how much we expect the identifying assumption to be violated. To circumvent inappropriately selecting the range of sensitivity parameters, Bayesian methods have been applied in sensitivity analysis. [7] proposed to incorporate Bayesian shrinkage on the mean and dependence parameter to share information across different missing patterns. [13] introduced Bayesian approach to analyze outcome from exponential distribution family with missing values that are potentially MNAR. [21] proposed an Bayesian approach to deal with missing data when estimating causal effect in randomized clinical trials. Although there is a rich literature on using Bayesian approaches to assess missing data, inference on the sensitivity parameter is largely limited to subjective priors. Given the scale of our applications for Facebook data relying on expert information elicit priors or choosing hyperparameters manually at scale is infeasible.

In this work, we propose a possible approach to conducting Sensitivity Analysis at scale using Bayesian estimates of the sensitivity parameters from the noisy, imperfect data. We propose applying either empirically derived subjective priors (when noisy but collectively useful data is available) or objective priors (when high quality data with strong unit information is available); both techniques allow automation. These Bayesian estimates of the sensitivity parameter can be used to compute the E-value of [25], to summarize sensitivity to MAR violations using the induced uncertainty under the Bayesian paradigm. Furthermore, since E-value is a function of the sensitivity parameter, we derive analytical forms of the distribution function of E-value based on the posterior distribution of the sensitivity parameter. These provide further possibilities for scalable sensitivity analysis where simulation based posterior inference may not be viable.

3 Methodology

3.1 Notation and Assumptions

For units of observation $i = 1, \ldots, n$ let Y_i and $\mathbf{X}_i$ denote the continuous outcome and covariates respectively. Furthermore, let R_i denote the missing status with $R_i = 1$ if Y_i are observed and $R_i = 0$ otherwise. For this paper, we will assume that the objective is to estimate the population mean i.e. $\mathbb{E}[Y] = \mu$.

If the missingness mechanism is *missing completely at random* i.e. MCAR ($Y_i \perp\!\!\!\perp R_i$) , one can estimate μ by $\hat{\mu} = \left(\sum_{i=1}^{n} R_i\right)^{-1} \sum_{i=1}^{n} R_i Y_i$ consistently; this assumption is unlikely to hold in practice. A more relaxed condition assumes that the mechanism is *missing at random* i.e. MAR ($Y_i \perp\!\!\!\perp R_i | \mathbf{X}_i$); μ can be estimated by techniques such as IPW or MI. In this paper, we will focus on the IPW estimator i.e. $\hat{\mu} = n^{-1} \sum_{i=1}^{n} R_i Y_i / \pi(\mathbf{X}_i)$ where $\pi(\mathbf{x}) = \mathrm{pr}(R = 1 | \mathbf{X} = \mathbf{x})$ is the true propensity score. In practice, propensity scores are unknown but estimable; with a consistent estimate of the propensity score μ can be consistently estimated.

It warrants mention that any statement about the missingness mechanism is a statement about *unknown unknowns*, untestable from the observed data; as a consequence missingness not-at random i.e. MNAR can never be fully excluded from possibility. Therefore we need to conduct sensitivity analysis to estimate the robustness of IPW estimator for μ against potential MAR violations.

Let δ denote the sensitivity parameter in the underlying pattern mixture model (our mode of analysis) which represents the degree to which MNAR is induced. As we noted earlier, this parameter is usually selected based on substantive assumptions. An alternative is to estimate it as $\hat{\delta} = \mathbb{E}(\hat{Y}) - \mathbb{E}(Y)$ where $\mathbb{E}(Y)$ can be observed from a variety of sources but may be contaminated by noise. In our motivating example from Facebook data, we focus on the scenarios where there are **two** distinct sources of obtaining values of $\mathbb{E}(Y)$.

Furthermore, let $\boldsymbol{\delta}_j = (\delta_{j1}, \delta_{j2})$ denote the sensitivity parameter estimates for some group $j = 1, \ldots, m$ in the population. Across m groups, Assume that the likelihood function $f(\boldsymbol{\delta}_j)$ is bivariate normal with mean vector $\delta\mathbf{1}$ and covariance matrix $\boldsymbol{\Sigma}$. Treating this as the data, under the Bayesian paradigm, we consider both the subjective and objective priors for $(\delta, \boldsymbol{\Sigma})$.

3.2 Subjective Prior

We choose Normal-Inverse-Wishart distribution as the subjective prior over $(\delta, \mathbf{\Sigma})$,

$$\pi(\delta|\mathbf{\Sigma}) \sim N(\delta_0, \phi_0), \phi_0 = \left(\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}\right)^{-1}$$
$$\pi(\mathbf{\Sigma}) \sim W^{-1}(\mathbf{\Psi}, \nu).$$

Then the joint density $f(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m, \delta, \mathbf{\Sigma})$ is given by

$$f(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m, \delta, \mathbf{\Sigma}) = \frac{1}{(2\pi)^n |\mathbf{\Sigma}|^{n/2}} e^{-\frac{1}{2}tr(\mathbf{S}_0\mathbf{\Sigma}^{-1})} \frac{1}{\sqrt{2\pi\phi_0}} e^{\frac{(\delta-\delta_0)^2}{2\phi_0}} \frac{|\mathbf{\Psi}|^{\nu/2}}{2^\nu \Gamma_2(\nu/2)} |\mathbf{\Sigma}|^{-\frac{\nu+3}{2}} e^{-\frac{1}{2}tr(\mathbf{\Psi}\mathbf{\Sigma}^{-1})}, \tag{1}$$

where $\mathbf{S}_0 = \sum_{j=1}^m (\boldsymbol{\delta}_j - \delta\mathbf{1})(\boldsymbol{\delta}_j - \delta\mathbf{1})'$. Then the marginal likelihood of $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m$ is given by integrating out δ and $\mathbf{\Sigma}$,

$$\begin{aligned} P(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m) &= \iint f(\hat{\delta}_1, \hat{\delta}_2, \delta, \mathbf{\Sigma}) d\delta d\mathbf{\Sigma} \\ &= \iint \frac{1}{(2\pi)^m |\mathbf{\Sigma}|^{n/2}} e^{-\frac{1}{2}tr(\mathbf{S}_0\mathbf{\Sigma}^{-1})} \frac{1}{\sqrt{2\pi\phi_0}} e^{\frac{(\delta-\delta_0)^2}{2\phi_0}} \frac{|\mathbf{\Psi}|^{\nu/2}}{2^\nu \Gamma_2(\nu/2)} |\mathbf{\Sigma}|^{-\frac{\nu+3}{2}} e^{-\frac{1}{2}tr(\mathbf{\Psi}\mathbf{\Sigma}^{-1})} d\delta d\mathbf{\Sigma} \end{aligned}$$

We first find the posterior distribution of δ and $\mathbf{\Sigma}$,

$$\begin{aligned} \pi(\delta, \mathbf{\Sigma}|\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m) &\propto \exp\left[-\frac{1}{2}\mathrm{tr}\mathbf{\Sigma}^{-1}\left\{\mathbf{\Psi} + \sum_{i=1}^n (\boldsymbol{\delta}_i - \delta\mathbf{1})(\boldsymbol{\delta}_i - \delta\mathbf{1})' + (\delta\mathbf{1} - \delta_0\mathbf{1})(\delta\mathbf{1} - \delta_0\mathbf{1})'\right\}\right] \\ &\propto \exp\left[-\frac{1}{2}\mathrm{tr}\mathbf{\Sigma}^{-1}\left\{\mathbf{\Psi} + \mathbf{S} + n(\bar{\boldsymbol{\delta}} - \delta\mathbf{1})(\bar{\boldsymbol{\delta}} - \delta\mathbf{1})' + (\delta\mathbf{1} - \delta_0\mathbf{1})(\delta\mathbf{1} - \delta_0\mathbf{1})'\right\}\right] \end{aligned}$$

Therefore, we have

$$\pi(\delta, \mathbf{\Sigma}|\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m) \sim NIW\left(\tilde{\delta}, \tilde{\Psi}, \tilde{\nu}\right)$$

where

$$\begin{aligned} \tilde{\delta} &= \frac{m\delta_0 + m\bar{\boldsymbol{\delta}}'\mathbf{1}}{m+1} \\ \tilde{\mathbf{\Psi}} &= \mathbf{\Psi} + \mathbf{S} + \frac{m}{m+1}(\bar{\boldsymbol{\delta}} - \delta_0\mathbf{1})(\bar{\boldsymbol{\delta}} - \delta_0\mathbf{1})' \\ \tilde{\nu} &= \nu + m \\ \mathbf{S} &= \sum_{j=1}^m (\boldsymbol{\delta}_j - \bar{\boldsymbol{\delta}})(\boldsymbol{\delta}_j - \bar{\boldsymbol{\delta}})' \end{aligned}$$

Then the marginal likelihood is the ratio of joint density $f(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m, \delta, \boldsymbol{\Sigma})$ to the posterior distribution $\pi(\delta, \boldsymbol{\Sigma} | \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m)$,

$$\begin{aligned} m(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m) &= (2\pi)^{-m} \frac{|\boldsymbol{\Psi}|^{\nu/2}}{2^{\nu}\Gamma_2(\nu/2)} \frac{2^{\tilde{\nu}}\Gamma_2(\tilde{\nu}/2)}{|\tilde{\boldsymbol{\Psi}}|^{\tilde{\nu}/2}} \\ &= \frac{1}{\pi^m} \frac{|\boldsymbol{\Psi}|^{\nu/2}}{|\tilde{\boldsymbol{\Psi}}|^{\tilde{\nu}/2}} \frac{\Gamma_2(\tilde{\nu}/2)}{\Gamma_2(\nu/2)} \end{aligned} \tag{2}$$

Take the negative logarithm of the marginal likelihood function yields the cost function with respect to $\delta, \boldsymbol{\Psi}$ and ν,

$$\mathcal{L}(\delta, \boldsymbol{\Psi}, \nu) = m \log(\pi) - \frac{\nu}{2} \log |\boldsymbol{\Psi}| + \frac{\tilde{\nu}}{2} \log |\tilde{\boldsymbol{\Psi}}| + \log \frac{\Gamma_2(\nu/2)}{\Gamma_2(\tilde{\nu}/2)} \tag{3}$$

Proposition 3.1 *Let $\Psi \in \mathbb{R}^{2\times 2}$ be a symmetric matrix. The cost function $\mathcal{L}(\delta, \boldsymbol{\Psi}, \nu)$ is convex with respect to δ_0 and Ψ when $\nu > Cm$ where C is some constant.*

Let $(\boldsymbol{\Psi}^*, \nu^*, \delta_0^*)$ be the global minimizer of equation (3). Substituting $(\boldsymbol{\Psi}^*, \nu^*, \delta_0^*)$ into the posterior distribution of $(\delta, \boldsymbol{\Sigma})$ and integrating out $\boldsymbol{\Sigma}$ will lead to the marginal posterior distribution of δ,

$$\begin{aligned} P(\delta) &\propto \left| (\delta - \delta_0^*)^2 \mathbf{1}\mathbf{1}' + \boldsymbol{\Psi}^* + \sum_{j=1}^{m} (\boldsymbol{\delta}_j - \delta\mathbf{1})(\boldsymbol{\delta}_j - \delta\mathbf{1})' \right|^{-(m+\nu^*)/2} \\ &\propto \left[1 + (j+1)(\bar{\mathbf{y}} - \delta\mathbf{1})'\mathbf{U}^{-1}(\bar{\mathbf{y}} - \delta\mathbf{1})\right]^{-(m+\nu^*)/2}, \end{aligned} \tag{4}$$

where

$$\begin{aligned} \mathbf{U} &= \boldsymbol{\Psi}^* + \mathbf{S} + \frac{m}{m+1}(\bar{\boldsymbol{\delta}} - \delta_0^*\mathbf{1})(\bar{\boldsymbol{\delta}} - \delta_0^*\mathbf{1})' \\ \bar{\mathbf{y}} &= \frac{m\bar{\boldsymbol{\delta}} + \delta_0\mathbf{1}}{m+1} \end{aligned}$$

Let $u = \mathbf{1}'\mathbf{U}^{-1}\bar{\mathbf{y}}, z = \mathbf{1}'\mathbf{U}^{-1}\mathbf{1}$ and $w = \mathbf{y}'\mathbf{U}^{-1}\mathbf{y}$, the marginal posterior distribution of δ follows a non-central student's t-dsitrbution with $m + \nu^* - 1$ degrees of freedom,

$$P(\delta) \propto \left[1 + \frac{(m+1)z\left(\delta - \frac{u}{z}\right)^2}{1 + (m+1)w - (m+1)u^2 z^{-1}}\right]^{-(m+\nu^*)/2} \tag{5}$$

The relationship between the likelihood, empirically motivated subjective prior and the posterior distributions is given in Figure 1.

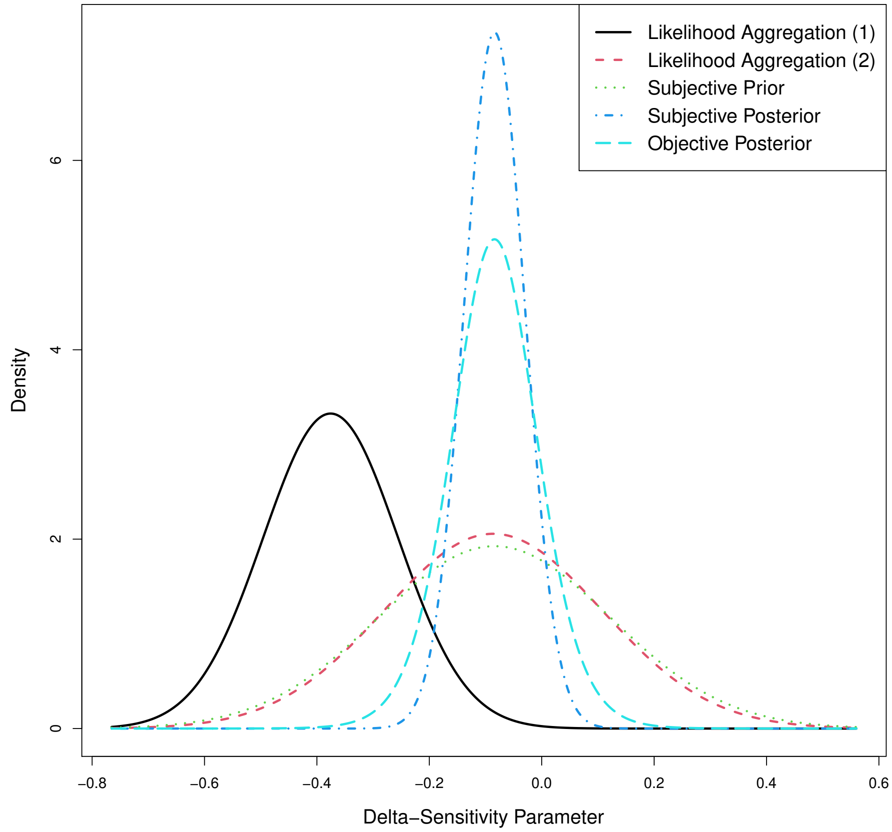

Figure 1: The likelihood function, the empirically motivated subjective prior and the posterior distributions

3.3 Objective Priors

As for the objective Bayesian approach, we choose the independent Jeffreys prior, $\pi_{IJ} = |\mathbf{\Sigma}|^{-(p+1)/2}$. Since $p = 2$ in our motivating application, independent Jeffreys prior has the form,

$$\pi_{IJ}(\delta, \mathbf{\Sigma}) = |\mathbf{\Sigma}|^{-3/2} \tag{6}$$

Then the joint density $f(\hat{\delta}_1, \hat{\delta}_2, \delta, \mathbf{\Sigma})$ is given by

$$\begin{aligned} f(\hat{\delta}_1, \hat{\delta}_2, \delta, \Sigma) &= \frac{1}{(2\pi)^m |\mathbf{\Sigma}|^{m/2}} e^{-\frac{1}{2} tr\{\sum_{j=1}^m (\boldsymbol{\delta}_i - \delta \mathbf{1})(\boldsymbol{\delta}_i - \delta \mathbf{1})' \mathbf{\Sigma}^{-1}\}} |\mathbf{\Sigma}|^{-3/2} \\ &= \frac{1}{(2\pi)^m} e^{-\frac{1}{2} tr\{\sum_{j=1}^m (\boldsymbol{\delta}_j - \delta \mathbf{1})(\boldsymbol{\delta}_j - \delta \mathbf{1})' \mathbf{\Sigma}^{-1}\}} |\mathbf{\Sigma}|^{-(m+3)/2}, \end{aligned} \tag{7}$$

and the marginal posterior distribution of δ is given by integrating out $\mathbf{\Sigma}^{-1}$

$$\begin{aligned} P(\delta) &= \int f(\hat{\delta}_1, \hat{\delta}_2, \delta, \mathbf{\Sigma}) d\mathbf{\Sigma} \\ &= \int \frac{1}{(2\pi)^m} e^{-\frac{1}{2} tr\{\sum_{j=1}^m (\boldsymbol{\delta}_j - \delta \mathbf{1})(\boldsymbol{\delta}_j - \delta \mathbf{1})' \mathbf{\Sigma}^{-1}\}} |\mathbf{\Sigma}|^{-(m+3)/2} d\mathbf{\Sigma} \\ &\propto \left| \sum_{j=1}^m (\boldsymbol{\delta}_j - \delta \mathbf{1})(\boldsymbol{\delta}_j - \delta \mathbf{1})' \right|^{-m/2}. \end{aligned}$$

Now let

$$\bar{\boldsymbol{\delta}} = m^{-1} \sum_{j=1}^m \boldsymbol{\delta}_j \quad \text{and} \quad \mathbf{S} = \sum_{j=1}^m (\boldsymbol{\delta}_j - \bar{\boldsymbol{\delta}})(\boldsymbol{\delta}_j - \bar{\boldsymbol{\delta}})'$$

so that

$$\sum_{j=1}^m (\boldsymbol{\delta}_j - \delta \mathbf{1})(\boldsymbol{\delta}_j - \delta \mathbf{1})' = \mathbf{S} + m(\bar{\boldsymbol{\delta}} - \delta \mathbf{1})(\bar{\boldsymbol{\delta}} - \delta \mathbf{1})'.$$

Recall that

$$\left| I + m(\bar{\boldsymbol{\delta}} - \delta \mathbf{1})(\bar{\boldsymbol{\delta}} - \delta \mathbf{1})' + \mathbf{S}^{-1} \right| = 1 + m(\bar{\boldsymbol{\delta}} - \delta \mathbf{1})' \mathbf{S}^{-1} (\bar{\boldsymbol{\delta}} - \delta \mathbf{1})$$

Define $u = \mathbf{1}' \mathbf{S}^{-1} \bar{\boldsymbol{\delta}}$, $z = \mathbf{1}' \mathbf{S}^{-1} \mathbf{1}$ and $w = \bar{\boldsymbol{\delta}}' \mathbf{S}^{-1} \bar{\boldsymbol{\delta}}$, we have

$$P(\delta) \propto \left[1 + \frac{mz \left(\delta - \frac{u}{z} \right)^2}{1 + mw - mu^2 z^{-1}} \right]^{-m/2}$$

Therefore, the marginal posterior distribution of δ is a *non-central* Student's t-distribution with $m-1$ degrees of freedom.

3.4 Bayesian Inference for E-value

[25] introduced sensitivity analysis using the E-value for the difference in continuous measurements. This technique rests on a *standardized effect size* i.e. a scaled difference

between $\hat{\mu}$ and $\hat{\mu}_\delta$, where $\hat{\mu}_\delta$ denotes the estimate incorporating the sensitivity parameter. This can be used to approximate the risk ratio and which in turn yields the E-value. In applications of missing data, we ideally want the E-value to be statistically indistinguishable from its reference value 1: this indicates that there is no difference between μ and μ_δ.

For inference, the posterior distribution of the E-value can then be approximated by simulation using the following scheme [4, 8].

For the purpose of Sensitivity Analysis, we first decompose μ using the *Law of Iterated Expectations*,

$$\mu = \mathbb{P}(R = 1)\mathbb{E}(Y|R = 1) + \mathbb{P}(R = 0)\mathbb{E}(Y|R = 0).$$

Pattern mixture model approach relates the unobserved $\mathbb{E}(Y|R = 0)$ with the observed $\mathbb{E}(Y|R = 1)$ by introducing the sensitivity parameter and assumes that

$$\mathbb{E}(Y|R = 0) = \mathbb{E}(Y|R = 1) + \delta.$$

Substituting this back into the decomposition of μ will give us

$$\mu_\delta = \mathbb{P}(R = 1)\mathbb{E}(Y|R = 1) + \{1 - \mathbb{P}(R = 1)\}\{\delta + \mathbb{E}(Y|R = 1)\}. \tag{8}$$

Let $\mu_{missing}$ denote the standardized effect size, then $\mu_{missing}$ can be calculated as

$$\begin{aligned}\mu_{missing} &= \frac{\mu_{\delta=\delta} - \mu_{\delta=0}}{\sqrt{\mathrm{Var}(Y)}} \\ &= \frac{\{1 - \mathbb{P}(R = 1)\}\delta}{\sqrt{\mathrm{Var}(Y)}},\end{aligned} \tag{9}$$

and the corresponding risk ratio (RR) can be approximated by

$$RR = \exp(0.91 \times \mu_{missing}),$$

and E-value can be obtained by

$$\text{E-value} = RR + \sqrt{RR * (RR - 1)}.$$

In addition to inference via posterior simulation, under certain assumptions, we can also derive the analytic approximate distribution functions of E-value under the framework in [25]. For brevity, let V denote the E-value; Using the formulation presented earlier, we have the following theorems on the distribution of V i.e. $f_V(v)$.

Theorem 3.2 *Suppose that $\mathbb{P}(R=1)=p$ and $Var(Y)=\sigma_Y$ are known, and that δ follows a normal distribution $N(\eta,\tau^2)$. Then the density function of V is*

$$f_V(v)=\begin{cases}\frac{1}{\sqrt{2\pi}\sigma_{RR}}\exp\left\{-\frac{\left(\ln\frac{v^2}{2v-1}-\mu_{RR}\right)^2}{2\sigma_{RR}^2}\right\}\left\{\frac{1}{v}-\frac{1}{v}\frac{1}{2v-1}\right\} & \textit{if } RR>1,\\ \frac{1}{\sqrt{2\pi}\sigma_{RR}}\exp\left\{-\frac{\left(\ln\frac{2v-1}{v^2}-\mu_{RR}\right)^2}{2\sigma_{RR}^2}\right\}\left\{\frac{1}{v}-\frac{1}{v}\frac{1}{2v-1}\right\} & \textit{if } 0<RR<1,\end{cases} \tag{10}$$

where $\mu_{RR}=0.91(1-p)\eta/\sigma_Y$ and $\sigma_{RR}=0.91(1-p)\tau/\sigma_Y$.

Theorem 3.3 *Suppose that $Var(Y)=\sigma_Y$ are known, that $q=1-\mathbb{P}(R=1)$ follows a normal distribution $N(\mu_q,\sigma_q^2)$, and that δ follows a normal distribution $N(\eta,\tau^2)$. Moreover, let $\rho_1=\sigma_q/\mu_q, \rho_2=\tau/\eta$. Assume that ρ_1 and ρ_2 are arbitrarily small, then the density function of V can be approximated by*

$$f_V(v)=\begin{cases}\frac{1}{\sqrt{2\pi}\sigma_{RR}}\exp\left\{-\frac{\left(\ln\frac{v^2}{2v-1}-\mu_{RR}\right)^2}{2\sigma_{RR}^2}\right\}\left\{\frac{1}{v}-\frac{1}{v}\frac{1}{2v-1}\right\} & \textit{if } RR>1,\\ \frac{1}{\sqrt{2\pi}\sigma_{RR}}\exp\left\{-\frac{\left(\ln\frac{2v-1}{v^2}-\mu_{RR}\right)^2}{2\sigma_{RR}^2}\right\}\left\{\frac{1}{v}-\frac{1}{v}\frac{1}{2v-1}\right\} & \textit{if } 0<RR<1,\end{cases} \tag{11}$$

where $\mu_{RR}=0.91\mu_q\eta/\sigma_Y$ and $\sigma_{RR}=0.91(\mu_q^2\tau^2+\eta^2\sigma_q^2+\sigma_q^2\tau^2)^{-1/2}/\sigma_Y$.

Theorem 3.4 *Suppose that σ_Y follows an inverse gamma distribution $IG(\alpha,\beta)$, that $q=1-\mathbb{P}(R=1)$ follows a normal distribution $N(\mu_q,\sigma_q^2)$, and that δ follows a normal distribution $N(\eta,\tau^2)$. Let $\rho_1=\sigma_q/\mu_q, \rho_2=\tau/\eta$, and $\rho_3=(\mu_q^2\tau^2+\eta^2\sigma_q^2+\sigma_q^2\tau^2)^{-1/2}/(\mu_q\eta)$. Assume that ρ_1,ρ_2, and ρ_3 are arbitrarily small, then the density function of V can be approximated by*

$$f_V(v)=\begin{cases}\frac{\beta_V^\alpha}{\Gamma(\alpha)}\exp\left\{-\beta_V\ln\frac{v^2}{2v-1}\right\}\left(\ln\frac{v^2}{2v-1}\right)^{\alpha-1}\left\{\frac{1}{v}-\frac{1}{v}\frac{1}{2v-1}\right\} & \textit{if } RR>1,\\ \frac{\beta_V^\alpha}{\Gamma(\alpha)}\exp\left\{-\beta_V\ln\frac{2v-1}{v^2}\right\}\left(\ln\frac{2v-1}{v^2}\right)^{\alpha-1}\left\{\frac{1}{v}-\frac{1}{v}\frac{1}{2v-1}\right\} & \textit{if } RR<1,\end{cases} \tag{12}$$

where $\beta_V=\mu_q\eta/(0.91\beta)$.

4 Results on Real and Simulated Data

To empirically demonstrate the advantages and limitations of ideas that we have presented in section 3, we present results on simulated data to study its properties in section 4.1. The results on our motivating application are provided in section 4.2. For both sets of results comparisons are made against *asymptotic* estimators that rely on large sample theory to demonstrate gains in inference quality using the proposed techniques. E-value variance is estimated using (1) a Taylor Series estimate of variance (see chapter 5 of [18]) and (2) Poisson Sampling Theory [14] using guidance from [25] to construct uncertainty intervals. On the other hand, the two Bayesian approaches calculate the credible intervals for E-value based on its posterior distribution.

4.1 Simulation

In order to evaluate our methodology, we present a simulation study to evaluate its performance. Each simulated dataset contains $i = 1, \ldots, 2500$ independent units of observation. The missing status for each subject i was generated as $R_i \sim Bernoulli(0.05)$. The outcomes Y_i and covariates $\mathbf{X}_i$ and estimated propensity scores are simulated via subsampling from *Facebook* data so that the simulated data mimics the properties of motivating application. The sensitivity parameter estimates $\boldsymbol{\delta}_j (j = 1, \ldots, 15)$ are generated from a bivariate normal distribution $\boldsymbol{\delta}_j \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.0025 & 0.0004 \\ 0.0004 & 0.0025 \end{pmatrix}\right)$. Therefore in this simulation study, the true sensitivity parameter is on average zero but observed with some noise. By simulating $T = 10000$ datasets in this scheme we estimate the coverage rate of 95% uncertainty intervals from the aforementioned four types of methods.

The coverage rates of the 95% uncertainty intervals from the four types of methods are given in Table 1. These results indicate that as the sample size of observed data increases, the coverage rates of confidence intervals from frequentist approaches come to attain expected coverage. On the other hand, although the coverage rate for the subjective Bayesian approach is marginally lower than the desired level, it is robust against different observed data sample sizes, indicating that subjective Bayesian approach is also applicable when the observed data size is relatively small, a concern often encountered in real data applications. The objective Bayesian approach is also robust against different observed datas sizes, but it may be too conservative particularly when large samples are available; at small sample sizes it also out-performs the asymptotic approaches.

Table 1: Coverage rate of 95% uncertainty interval for E-value with different sample sizes of missing data

m	Taylor Series	Poisson Sampling	Subjective Bayesian	Objective Bayesian
$1\times$	1.0000	1.0000	0.9448	0.9810
$3\times$	1.0000	1.0000	0.9448	0.9810
$6\times$	0.9996	0.9999	0.9448	0.9810
$9\times$	0.9973	0.9988	0.9448	0.9810
$12\times$	0.9910	0.9943	0.9448	0.9810
$15\times$	0.9811	0.9865	0.9448	0.9810
$18\times$	0.9656	0.9742	0.9448	0.9810

We compare the average width of uncertainty intervals from the four types of methods with respect to different sample sizes of observed data in Table 2. The average width of the uncertainty intervals from the two asymptotic approaches are extremely large when the sample size of observed data is relatively small (close to what we observe in the real observed data). The average width from these two methods decrease rapidly when the sample size of observed data increases. By comparison the average widths of the credible intervals from the two Bayesian approaches are stable against the sample size of observed data, making them more reliable than their asymptotic counterparts when the sample size of observed data is relatively small.

Table 2: Average width of 95% uncertainty interval for E-value with different sample sizes of missing data

m	Taylor Series	Poisson Sampling	Subjective Bayesian	Objective Bayesian
1×	7.8255×10^5	2.1685×10^9	0.1380	0.1504
3×	88.4024	1.3056×10^3	0.1469	0.1601
6×	2.0216	3.8857	0.1513	0.1649
9×	1.0315	1.2356	0.1531	0.1670
12×	0.7996	0.8909	0.1542	0.1682
15×	0.6824	0.7407	0.1548	0.1689
18×	0.6078	0.6506	0.1553	0.1694

4.2 Motivating Application: ***Facebook***

Facebook relies on handling missing outcome data for conversion information (e.g. whether an item was purchased or not). In these settings, IPW methods may be utilized to ensure that bias from self-selection can be eliminated in the estimation of population averages.

We apply the proposed methods to study the robustness of these IPW estimators of population averages against the mechanism being MNAR due to misspecified *weighting* models being used in construction. For each observation in the data we collect the following information:

- `conversions`: Number of a certain type of events from a single user after accessing the advertisement (e.g. Purchases).
- `event name`: Type of the event, taking 14 levels including *Start Trial, Submit Application, Contact, Add To Cart, Add Payment Information, Search, View Content, Complete Registration, Initiate Checkout, Purchase, Schedule, Lead, Add To Wishlist.*
- `propensity scores`: Estimated propensity score of being missing.

The outcome of interest here is the conversions. Out of consideration for user data privacy and compliance with regulatory reform, Facebook utilizes aggregated conversions from advertisements as the approximate ground truth values for the population mean. In this work, we take the differences between average conversions from IPW estimators and approximate ground truth average values as the estimated sensitivity parameters. The sensitivity parameter estimates are then calculated for different types of events.

To apply the proposed Bayesian approaches, we obtain the marginal posterior distribution of the sensitivity parameter based on the data. We then apply Gibbs sampling [8] method to draw a sample of sensitivity parameters from its posterior distribution and calculate corresponding E-values. For reference, the distributions of the sensitivity parameters under both prior choices are presented in Figure 2.

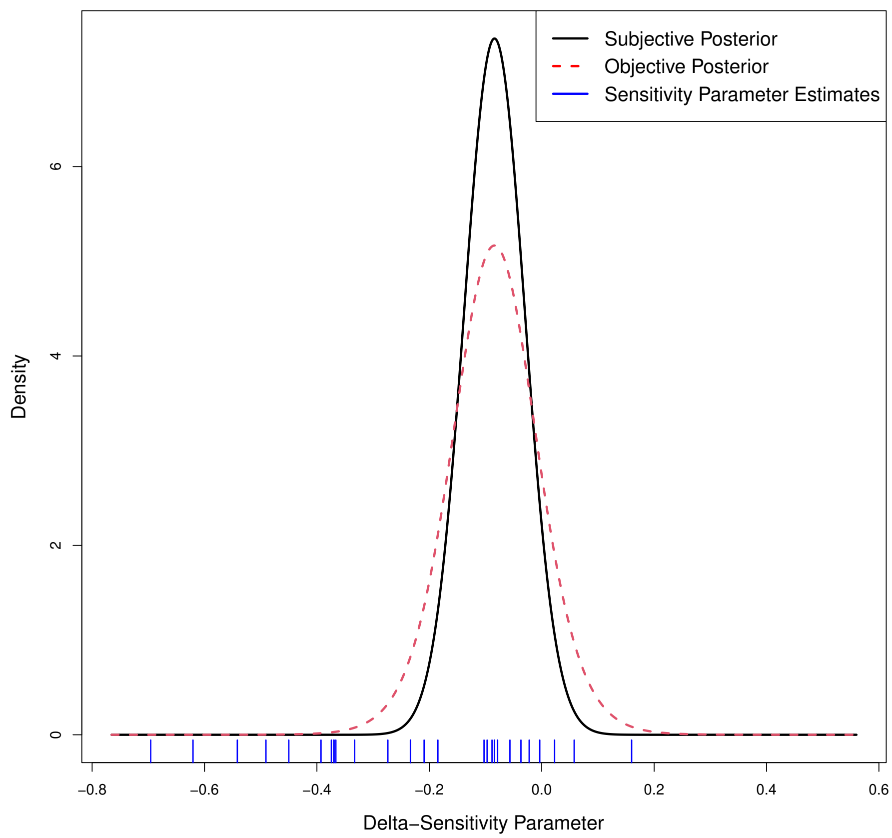

Figure 2: Estimated sensitivity parameters and posterior distributions of the sensitivity parameters under both subjective and objective priors

The 95% credible intervals for E-value are obtained by taking the 95% percentiles from the samples drawn. The 95% confidence intervals from the frequentist approaches and the 95% credible intervals from the two Bayesian approaches are summarized in Table 3 and 4.

Table 3: 95% uncertainty interval for E-value for different types of events from in-application data

Event Name	Taylor Series	Poisson Sampling	Subjective Bayesian	Objective Bayesian
Start Trial	(1.9825, 2.8795)	(1.9816, 2.8616)	(1, 1.6215)	(1, 1.7806)
Submit Application	(1.6015, 2.5573)	(1.5893, 2.5514)	(1, 1.6029)	(1, 1.7557)
Contact	(1.2941, 2.1228)	(1.2890, 2.1112)	(1, 1.5976)	(1, 1.7487)
Add To Cart	(1.9844, 2.2117)	(1.9836, 2.2081)	(1, 1.6023)	(1, 1.7549)
Add Payment Info	(2.0237, 3.0502)	(2.0124, 3.0424)	(1, 1.6076)	(1, 1.7620)
Search	(2.4783, 2.8783)	(2.4739, 2.8751)	(1, 1.5951)	(1, 1.7453)
View Content	(2.1910, 2.2559)	(2.1902, 2.2554)	(1, 1.5961)	(1, 1.7467)
Complete Registration	(1.5749, 1.8631)	(1.5726, 1.8599)	(1, 1.6059)	(1, 1.7597)
Initiate Checkout	(1.4163, 1.8987)	(1.4135, 1.8921)	(1, 1.6032)	(1, 1.7561)
Purchase	(1.9800, 2.1216)	(1.9788, 2.1201)	(1, 1.6074)	(1, 1.7618)
Schedule	(1.5842, 2.5758)	(1.5844, 2.5547)	(1, 1.6083)	(1, 1.7629)
Subscribe	(1.3760, 2.6290)	(1.3524, 2.6275)	(1, 1.6128)	(1, 1.7690)
Lead	(1.9731, 2.1028)	(1.9722, 2.1012)	(1, 1.6057)	(1, 1.7594)
Add To Wishlist	(1.3687, 2.2694)	(1.3522, 2.2675)	(1, 1.6018)	(1, 1.7543)

We find that the credible intervals from Bayesian approaches are more stabilized across different types of events when compared with their counterparts from frequentist approaches. This is consistent with what we observed in the simulation study where Bayesian credible intervals are not affected by the sample size of the observed data. Moreover, there exist inconsistencies between the results from frequentist and Bayesian approaches. For example, for events *Search* and *Purchase* as in Table 4. For *Search*, the confidence intervals from the two frequentist approaches do not include one while their Bayesian variants result in conservative conclusions. For *Purchase*, the confidence interval from the Taylor series approximation approach is solely significant while the others are more conservative. We prefer the Bayesian approach since they remain stable across event types and that they have better performance when the sample size of observed data is relatively small (which is the case for *Search* and *Purchase*).

Table 4: 95% uncertainty interval for E-value for different types of events from advertiser server data

Event Name	Taylor Series	Poisson Sampling	Subjective Bayesian	Objective Bayesian
Start Trial	(1, 1.5646)	(1, 1.2293)	(1, 1.1064)	(1, 1.1254)
Submit Application	(1, 1.7366)	(1, 1.2706)	(1, 1.0995)	(1, 1.1171)
Add To Cart	(1, 1.2657)	(1, 1.5144)	(1, 1.1412)	(1, 1.1672)
Add Payment Info	(1, 1.7273)	(1, 1.3366)	(1, 1.1425)	(1, 1.1688)
Search	(1.2284, 1.4240)	(1.0532, 1.5202)	(1, 1.1496)	(1, 1.1773)
View Content	(1, 1.1658)	(1, 1.2253)	(1, 1.1420)	(1, 1.1681)
Complete Registration	(1, 1.3306)	(1, 1.0898)	(1, 1.1282)	(1, 1.1515)
Initiate Checkout	(1, 1.3566)	(1, 1.5983)	(1, 1.1372)	(1, 1.1624)
Purchase	(1.0479, 1.1843)	(1, 1.2697)	(1, 1.1297)	(1, 1.1532)
Subscribe	(1, 1.4992)	(1, 1.1472)	(1, 1.1156)	(1, 1.1363)
Lead	(1, 1.1922)	(1, 1.3087)	(1, 1.1304)	(1, 1.1541)
Add To Wishlist	(1, 1.3619)	(1, 1.4663)	(1, 1.1423)	(1, 1.1685)

5 Discussion

This paper introduces methodology for conducting Sensitivity Analysis for missing data problems using the approach of E-values under the Bayesian paradigm. The crucial sensitivity parameters are estimated as differences between multiple population level benchmarks (e.g. differentially private aggregates or meta-analyses) and their counterparts estimated from partially missing unit level outcomes. Treating these estimates of the sensitivity parameters as data and leveraging prior information over an assumed *true* sensitivity helps determine the robustness against violations of MAR while holistically representing uncertainty. We present results on real and simulated data motivated by applications at Facebook where missing unit level outcome data is commonplace.

This paper makes several novel contributions to the field of sensitivity analysis for missing data. To the best of our knowledge, we are the first to study the first to study the distribution function of the E-value in Sensitivity Analysis. Since that E-value can be presented as a function of the sensitivity parameter, we propose two novel Bayesian approaches to derive the posterior distribution of the sensitivity parameter and thus the distribution function of the E-value. Our theoretical findings are undergirded by empirical evidence of the benefits of this approach, demonstrating that we can draw higher quality inference on sensitivity to MAR violations under the proposed approach rather than relying on asymptotic guarantees.

With that said, we have also encountered challenges in this effort that warrant further study. First, the proposed methods rely on pooling together sensitivity parameter estimates that are likely to be similar due to representing, for instance, similar type of outcomes such as related conversions (e.g. adding an item to your cart and purchasing it). Stronger assumptions are needed if we want to pool together sensitivity parameters that are very likely to be dissimilar. In addition, our current analysis is restricted to cross-sectional analyses of data; longitudinal analyses will require careful understanding of how sensitivity parameters can be estimated and pooled. Hierarchical Bayesian modeling [2, 26] methods might be applicable when dissimilarity is significant among subgroups of sensitivity parameter estimates or there exists time series patterns among sensitivity parameter estimates. We intend to explore these ideas in our future work.

6 Appendix

6.1 Proof of Proposition 3.1

Take the negative logarithm of it and we get the cost function with respect to $\delta, \mathbf{\Psi}$ and ν,

$$\mathcal{L}(\delta, \mathbf{\Psi}, \nu) = m \log(\pi) - \frac{\nu}{2} \log |\mathbf{\Psi}| + \frac{\tilde{\nu}}{2} \log |\tilde{\mathbf{\Psi}}| + \log \frac{\Gamma_2(\nu/2)}{\Gamma_2(\tilde{\nu}/2)}$$

We first optimize over $\mathbf{\Psi}$,

$$\frac{\partial \mathcal{L}(\delta, \mathbf{\Psi}, \nu)}{\partial \mathbf{\Psi}} = -\frac{\nu}{2} \mathbf{\Psi}^{-1} + \frac{\tilde{\nu}}{2} \tilde{\mathbf{\Psi}}^{-1} \stackrel{\text{set}}{=} 0,$$

and we get

$$\mathbf{\Psi}^* = \frac{\nu}{m} \left(\mathbf{S} + \frac{m}{m+1} (\bar{\boldsymbol{\delta}} - \delta_0 \mathbf{1})(\bar{\boldsymbol{\delta}} - \delta_0 \mathbf{1})' \right).$$

To make it more rigorously, we need to check the convexity for the cost function to make sure that $\mathbf{\Psi}^*$ is the global minimum. To do that, we show that $f(\mathbf{\Psi}) = -\frac{\nu}{2} \log |\mathbf{\Psi}| + \frac{\tilde{\nu}}{2} \log |\tilde{\mathbf{\Psi}}|$ is convex for certain fixed values of ν by considering an arbitrary line given by $\mathbf{\Psi} + t\mathbf{V}$, where $\mathbf{\Psi}$ and $\mathbf{V}$ are positive definite matrices. We then define $g(t) = -\frac{\nu}{2} \log |\mathbf{\Psi} + t\mathbf{V}| + \frac{\tilde{\nu}}{2} \log |\tilde{\mathbf{\Psi}} + t\mathbf{V}|$ such that both $\mathbf{\Psi} + t\mathbf{V}$ and $\tilde{\mathbf{\Psi}} + t\mathbf{V}$ are positive definite matrices. Since $\mathbf{\Psi}$ and $\tilde{\mathbf{\Psi}}$ are positive definite, there exist $\mathbf{\Psi}^{1/2}$ and $\tilde{\mathbf{\Psi}}^{1/2}$ such that $\mathbf{\Psi} = \mathbf{\Psi}^{1/2} \mathbf{\Psi}^{1/2}$ and $\tilde{\mathbf{\Psi}} = \tilde{\mathbf{\Psi}}^{1/2} \tilde{\mathbf{\Psi}}^{1/2}$. We then have

$$
\begin{aligned}
g(t) &= -\frac{\nu}{2}\log|\mathbf{\Psi}+t\mathbf{V}| + \frac{\tilde{\nu}}{2}\log|\tilde{\mathbf{\Psi}}+t\mathbf{V}| \\
&= -\frac{\nu}{2}\log|\mathbf{\Psi}^{1/2}\mathbf{\Psi}^{1/2} + t\mathbf{\Psi}^{1/2}\mathbf{\Psi}^{-1/2}\mathbf{V}\mathbf{\Psi}^{-1/2}\mathbf{\Psi}^{1/2}| \\
&+ \frac{\nu+m}{2}\log|\tilde{\mathbf{\Psi}}^{1/2}\tilde{\mathbf{\Psi}}^{1/2} + t\tilde{\mathbf{\Psi}}^{1/2}\tilde{\mathbf{\Psi}}^{-1/2}\mathbf{V}\tilde{\mathbf{\Psi}}^{-1/2}\tilde{\mathbf{\Psi}}^{1/2}| \\
&= -\frac{\nu}{2}\log|\mathbf{\Psi}^{1/2}(\mathbf{I}+t\mathbf{\Psi}^{-1/2}\mathbf{V}\mathbf{\Psi}^{-1/2})\mathbf{\Psi}^{1/2}| \\
&+ \frac{\nu+m}{2}\log|\tilde{\mathbf{\Psi}}^{1/2}(\mathbf{I}+t\tilde{\mathbf{\Psi}}^{-1/2}\mathbf{V}\tilde{\mathbf{\Psi}}^{-1/2})\tilde{\mathbf{\Psi}}^{1/2}| \\
&= -\frac{\nu}{2}\left(\log|\mathbf{\Psi}| + \log|\mathbf{I}+t\mathbf{\Psi}^{-1/2}\mathbf{V}\mathbf{\Psi}^{-1/2}|\right) \\
&+ \frac{\nu+m}{2}\left(\log|\tilde{\mathbf{\Psi}}| + \log|\mathbf{I}+t\tilde{\mathbf{\Psi}}^{-1/2}\mathbf{V}\tilde{\mathbf{\Psi}}^{-1/2}|\right) \\
&= -\frac{\nu}{2}\left\{\log|\mathbf{\Psi}| + \log(1+t\lambda_1) + \log(1+t\lambda_2)\right\} \\
&+ \frac{\nu+m}{2}\left\{\log|\tilde{\mathbf{\Psi}}| + \log(1+t\eta_1) + \log(1+t\eta_2)\right\}
\end{aligned}
$$

Where λ_1, λ_2 are eigenvalues of $\mathbf{I} + t\mathbf{\Psi}^{-1/2}\mathbf{V}\mathbf{\Psi}^{-1/2}$ and η_1, η_2 are eigenvalues of $\mathbf{I} + t\tilde{\mathbf{\Psi}}^{-1/2}\mathbf{V}\tilde{\mathbf{\Psi}}^{-1/2}$. Notice that since $\mathbf{I} + t\mathbf{\Psi}^{-1/2}\mathbf{V}\mathbf{\Psi}^{-1/2}$ and $\mathbf{I} + t\tilde{\mathbf{\Psi}}^{-1/2}\mathbf{V}\tilde{\mathbf{\Psi}}^{-1/2}$ are also positive definite matrices. Then

$$
\begin{aligned}
g''(t) &= \frac{\nu}{2}\left\{\frac{\lambda_1^2}{(1+t\lambda_1)^2} + \frac{\lambda_2^2}{(1+t\lambda_2)^2}\right\} - \frac{\nu+m}{2}\left\{\frac{\eta_1^2}{(1+t\eta_1)^2} + \frac{\eta_2^2}{(1+t\eta_2)^2}\right\} \\
&= \frac{\nu}{2}\left[\left\{\frac{1}{(t+\lambda_1^{-1})^2} + \frac{1}{(t+\lambda_2^{-1})^2}\right\} - \left\{\frac{1}{(t+\eta_1^{-1})^2} + \frac{1}{(t+\eta_2^{-1})^2}\right\}\right] \\
&- \frac{m}{2}\left\{\frac{1}{(t+\eta_1^{-1})^2} + \frac{1}{(t+\eta_2^{-1})^2}\right\}
\end{aligned}
$$

Notice that $\tilde{\mathbf{\Psi}} = \mathbf{\Psi} + \mathbf{S} + \frac{m}{m+1}(\bar{\boldsymbol{\delta}} - \delta_0\mathbf{1})(\bar{\boldsymbol{\delta}} - \delta_0\mathbf{1})'$, it is easy to verify that $\lambda_1 > \eta_1$ and $\lambda_2 > \eta_2$, so we find the range of values of ν over which the cost function is convex as

$$
\nu > m\left[\left\{\frac{1}{(t+\lambda_1^{-1})^2} + \frac{1}{(t+\lambda_2^{-1})^2}\right\} - \left\{\frac{1}{(t+\eta_1^{-1})^2} + \frac{1}{(t+\eta_2^{-1})^2}\right\}\right]^{-1}\left\{\frac{1}{(t+\eta_1^{-1})^2} + \frac{1}{(t+\eta_2^{-1})^2}\right\}
$$

Then we substitute $\mathbf{\Psi}^*$ to the cost function and optimize over ν,

$$
\mathcal{L}(\delta, \mathbf{\Psi}^*, \nu) = \nu\log\frac{m+\nu}{\nu} + m\log\frac{m+\nu}{m} + \log\frac{\Gamma_2(\nu/2)}{\Gamma_2(\tilde{\nu}/2)} + \text{const}
$$

and

$$\frac{\partial \mathcal{L}(\delta, \mathbf{\Psi}^*, \nu)}{\partial \nu} = \log \frac{m+\nu}{\nu} + \frac{\partial}{\partial \nu} \log \frac{\Gamma_2(\nu/2)}{\Gamma_2(\tilde{\nu}/2)}$$
$$\frac{\partial^2 \mathcal{L}(\delta, \mathbf{\Psi}^*, \nu)}{\partial \nu^2} = -\frac{m}{m\nu+\nu^2} + \frac{\partial^2}{\partial \nu^2} \log \frac{\Gamma_2(\nu/2)}{\Gamma_2(\tilde{\nu}/2)} > 0$$

There is no closed form solution for ν but we know the objective function is convex since the second order derivative is positive.

Then we substitute $\mathbf{\Psi}^*$ and ν^* back in the cost function and optimize over δ_0,

$$\mathcal{L}(\delta_0, \mathbf{\Psi}^*, \nu^*) = \frac{m}{2} \log \left| \mathbf{S} + \frac{m}{m+1}(\bar{\boldsymbol{\delta}} - \delta_0 \mathbf{1})(\bar{\boldsymbol{\delta}} - \delta_0 \mathbf{1})' \right|.$$

And this objective function is convex since the matrix $\mathbf{S} + \frac{m}{m+1}(\bar{\boldsymbol{\delta}} - \delta_0 \mathbf{1})(\bar{\boldsymbol{\delta}} - \delta_0 \mathbf{1})'$ is positive definite. The plot of the values objective function versus δ_0 for certain given values of Ψ^* and ν^* is shown in Figure 3.

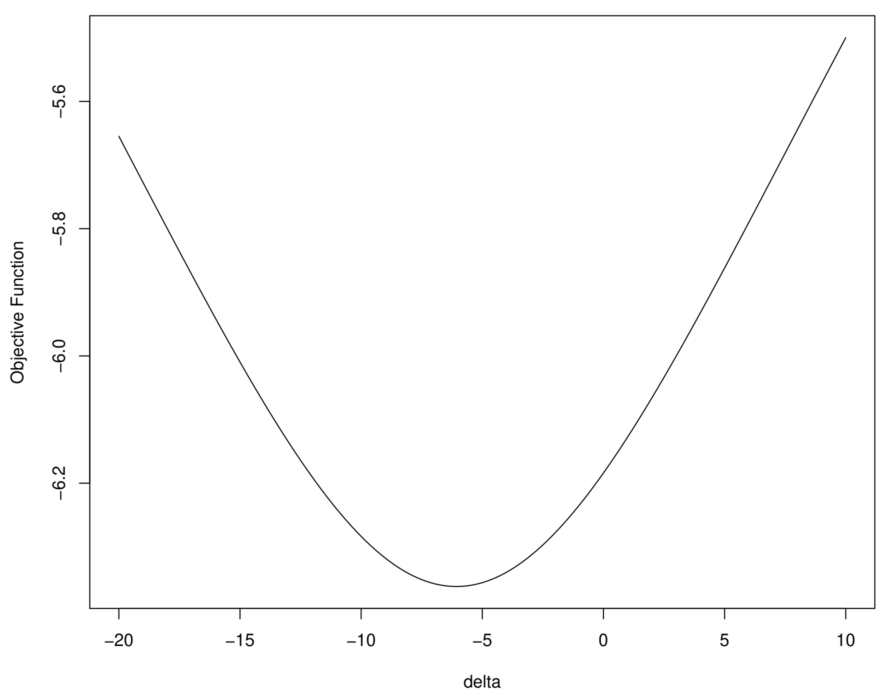

Figure 3: Plot of objective function for certain given values of Ψ^* and ν^*

6.2 Proof of Theorem 3.2

Let $\mu_{missing} = (1-p)\delta/\sigma_Y$ where $\delta \sim N(\eta, \tau^2)$, then $\mu_{missing} \sim N\left((1-p)\eta/\sigma_Y, (1-p)^2\tau^2/\sigma_Y^2\right)$. Then the risk ratio (RR) satisfies $RR = \exp(0.91 \times \mu_{missing})$ and $RR \sim \text{Lognormal}(\mu_{RR}, \sigma_{RR}^2)$ where $\mu_{RR} = 0.91(1-p)\eta/\sigma_Y$ and $\sigma_{RR} = 0.91(1-p)\tau/\sigma_Y$. Let V denote the E-value, we have

$$V = \begin{cases} RR + \sqrt{RR(RR-1)} & \text{if } RR > 1, \\ 1 & \text{if } RR = 1, \\ 1/RR + \sqrt{1/RR(1/RR-1)} & \text{if } 0 < RR < 1. \end{cases} \tag{13}$$

Since $\mathbb{P}(RR = 1) = 0$, we will only consider the cases where $RR > 1$ or $0 < RR < 1$. If $RR > 1$, then we have $RR = V^2/(2V-1)$ and then the density function of V is given by

$$\begin{aligned} f_V(v) &= f_{RR}(v^2/(2v-1))|RR'| \\ &= \frac{1}{\sqrt{2\pi}\sigma_{RR}} \exp\left\{-\frac{\left(\ln \frac{v^2}{2v-1} - \mu_{RR}\right)^2}{2\sigma_{RR}^2}\right\} \left\{\frac{1}{v} - \frac{1}{v}\frac{1}{2v-1}\right\}. \end{aligned}$$

Similarly, when $0 < RR < 1$ we have $RR = (2V-1)/V^2$ and the corresponding density function of V is

$$\begin{aligned} f_V(v) &= f_{RR}((2v-1)/v^2)|RR'| \\ &= \frac{1}{\sqrt{2\pi}\sigma_{RR}} \exp\left\{-\frac{\left(\ln \frac{2v-1}{v^2} - \mu_{RR}\right)^2}{2\sigma_{RR}^2}\right\} \left\{\frac{1}{v} - \frac{1}{v}\frac{1}{2v-1}\right\} \end{aligned}$$

6.3 Proof of Theorem 3.3

Let $\mu_{missing} = q\delta/\sigma_Y$ where $q \sim N(\mu_q, \sigma_q^2)$ and $\delta \sim N(\eta, \tau^2)$. By Theorem 2.5, 2.6 and 2.7 in [1], under the assumption that $\rho_1 = \sigma_q/\mu_q$ and $\rho_2 = \tau/\eta$ are arbitrarily small, we can approximate the distribution of $q\delta$ by a normal distribution with mean $\mu_q\eta$ and variance $\mu_q^2\tau^2 + \eta^2\sigma_q^2 + \sigma_q^2\tau^2$. Then we are able to derive the distribution of E-value via a similar approach as in proof of Theorem 3.2.

6.4 Proof of Theorem 3.4

To prove Theorem reftheorem: evalue3, we apply the normal approximation for the product distribution of $q = 1-p$ and δ as in the proof of Theorem 3.3 under the assumption that $\rho_1 = \sigma_q/\mu_q$ and $\rho_2 = \tau/\eta$ are arbitrarily small. Thus the characteristic function of $q\delta$ after normal approximation is $\exp(i\mu_q\eta t + (\mu_q^2\tau^2 + \eta^2\sigma_q^2 + \sigma_q^2\tau^2)t^2/2)$. Then the characteristic function of $\mu_{missing}$ is

$$\begin{aligned}\phi_{\mu_{missing}}(t) &= \mathbb{E}_{\sigma_Y}\left[\mathbb{E}\left\{e^{it/\sigma_Y q\delta}\middle|\sigma_Y\right\}\right] \\ &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\exp(-\beta y + i\mu_q\eta y + (\mu_q^2\tau^2 + \eta^2\sigma_q^2 + \sigma_q^2\tau^2)y^2/2)y^{\alpha-1}\, dy\end{aligned} \tag{14}$$

We additional assume that $\rho_3 = (\mu_q^2\tau^2 + \eta^2\sigma_q^2 + \sigma_q^2\tau^2)^{-1/2}/(\mu_q\eta)$ is arbitrarily small, then

$$\begin{aligned}\phi_{\mu_{missing}}(t) &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\exp(-\beta y + i\mu_q\eta y)y^{\alpha-1}\, dy \\ &= \frac{1}{(1 - i\mu_q\eta t/\beta)^\alpha}.\end{aligned} \tag{15}$$

Therefore, the distribution of $\mu_{missing}$ is approximated by a gamma distribution $G(\alpha, \mu_q\eta/\beta)$. Then similarly as in proof of Theorem 3.2, we can derive the approximation of the distribution function of E-value by applying change of variable twice. Specifically, when $RR > 1$, we have $RR = V^2/(2V-1)$ and then the density function of V is given by

$$\begin{aligned}f_V(v) &= f_{RR}\left(\frac{v^2}{2v-1}\right)|RR'| \\ &= \frac{\beta_V^\alpha}{\Gamma(\alpha)}\exp\left\{-\beta_V \ln\frac{v^2}{2v-1}\right\}\left(\ln\frac{v^2}{2v-1}\right)^{\alpha-1}\left\{\frac{1}{v} - \frac{1}{v}\frac{1}{2v-1}\right\}.\end{aligned} \tag{16}$$

Similarly, when $0 < RR < 1$ we have $RR = (2V-1)/V^2$ and the density function of V is given by

$$\begin{aligned}f_V(v) &= f_{RR}\left(\frac{2v-1}{v^2}\right)|RR'| \\ &= \frac{\beta_V^\alpha}{\Gamma(\alpha)}\exp\left\{-\beta_V \ln\frac{2v-1}{v^2}\right\}\left(\ln\frac{2v-1}{v^2}\right)^{\alpha-1}\left\{\frac{1}{v} - \frac{1}{v}\frac{1}{2v-1}\right\}.\end{aligned} \tag{17}$$

References

[1] L. A. Aroian. The probability function of the product of two normally distributed variables. *The Annals of Mathematical Statistics*, pages 265–271, 1947.

[2] L. M. Berliner. Hierarchical bayesian time series models. In *Maximum entropy and Bayesian methods*, pages 15–22. Springer, 1996.

[3] N. B. Carnegie, M. Harada, and J. L. Hill. Assessing sensitivity to unmeasured confounding using a simulated potential confounder. *Journal of Research on Educational Effectiveness*, 9(3):395–420, 2016.

[4] J. Diebolt and C. P. Robert. Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(2):363–375, 1994.

[5] P. Diggle and M. G. Kenward. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):49–73, 1994.

[6] A. Franks, A. D'Amour, and A. Feller. Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association*, 2019.

[7] J. Gaskins, M. Daniels, and B. Marcus. Bayesian methods for nonignorable dropout in joint models in smoking cessation studies. *Journal of the American Statistical Association*, 111(516):1454–1465, 2016.

[8] A. E. Gelfand, S. E. Hills, A. Racine-Poon, and A. F. Smith. Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association*, 85(412):972–985, 1990.

[9] A. E. Gelfand and A. F. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.

[10] D. Geradin, D. Katsifis, and T. Karanikioti. Google as a de facto privacy regulator: analysing the privacy sandbox from an antitrust perspective. *European Competition Journal*, pages 1–65, 2021.

[11] J. J. Heckman. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4*, pages 475–492. NBER, 1976.

[12] A. Heyting, J. Tolboom, and J. Essers. Statistical handling of drop-outs in longitudinal clinical trials. *Statistics in medicine*, 11(16):2043–2061, 1992.

[13] N. A. Kaciroti and T. Raghunathan. Bayesian sensitivity analysis of incomplete data: bridging pattern-mixture and selection models. *Statistics in medicine*, 33(27):4841–4857, 2014.

[14] P. S. Kott. Delete-a-group variance estimation for the general regression estimator under poisson sampling. *Journal of Official Statistics*, 22(4):759, 2006.

[15] R. Little. Selection and pattern-mixture models. In *Longitudinal data analysis*, pages 423–446. Chapman and Hall/CRC, 2008.

[16] R. J. Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 1993.

[17] R. J. Little. A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81(3):471–483, 1994.

[18] R. J. Little, R. D'Agostino, M. L. Cohen, K. Dickersin, S. S. Emerson, J. T. Farrar, C. Frangakis, J. W. Hogan, G. Molenberghs, S. A. Murphy, et al. The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, 367(14):1355–1360, 2012.

[19] R. J. Little and Y. Wang. Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, pages 98–111, 1996.

[20] X. Liu, W. Kong, S. Kakade, and S. Oh. Robust and differentially private mean estimation. *arXiv preprint arXiv:2102.09159*, 2021.

[21] A. J. Mason, M. Gomes, R. Grieve, P. Ulug, J. T. Powell, and J. Carpenter. Development of a practical approach to expert elicitation for randomised controlled trials with missing health outcomes: application to the improve trial. *Clinical Trials*, 14(4):357–367, 2017.

[22] L. C. McCandless, P. Gustafson, and A. Levy. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in medicine*, 26(11):2331–2347, 2007.

[23] G. Molenberghs, H. Thijs, I. Jansen, C. Beunckens, M. G. Kenward, C. Mallinckrodt, and R. J. Carroll. Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, 5(3):445–464, 2004.

[24] P. R. Rosenbaum and D. B. Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2):212–218, 1983.

[25] T. J. VanderWeele and P. Ding. Sensitivity analysis in observational research: introducing the e-value. *Annals of internal medicine*, 167(4):268–274, 2017.

[26] C. K. Wikle, L. M. Berliner, and N. Cressie. Hierarchical bayesian space-time models. *Environmental and ecological statistics*, 5(2):117–154, 1998.