Statistics and Probability Letters 118 (2016) 8–15

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Demystifying the bias from selective inference: A revisit to Dawid's treatment selection problem

Jiannan Lu*, Alex Deng

Analysis and Experimentation, Microsoft Corporation, United States

ARTICLE INFO

Article history:
Received 20 January 2016
Received in revised form 13 April 2016
Accepted 9 June 2016
Available online 16 June 2016

Keywords:
Bayesian inference
Posterior mean
Selection paradox
Multivariate truncated normal

ABSTRACT

We extend the heuristic discussion in Senn (2008) on the bias from selective inference for the treatment selection problem (Dawid, 1994), by deriving the closed-form expression for the selection bias. We illustrate the advantages of our theoretical results through numerical and simulated examples.

1. Introduction

Selective inference gained popularity in recent years (e.g., Lockhart et al., 2014; G'Sell et al., 2016; Reid and Tibshirani, 2016). To quote (Dawid, 1994), "…a great deal of statistical practice involves, explicitly or implicitly, a two stage analysis of the data. At the first stage, the data are used to identify a particular parameter on which attention is to focus; the second stage then attempts to make inferences about the selected parameter". Consequently, the results (e.g., point estimates, *p*-values) produced by selective inference are generally "cherry-picked" (Taylor and Tibshirani, 2015), and therefore it is of great importance for practitioners to conduct "exact post-selection inference" (e.g., Tibshirani et al., 2014; Lee et al., 2016).

To demonstrate the importance of "exact post-selection inference", in this paper we focus on the "bias" of the posterior mean associated with the most extreme observation (formally defined later, and henceforth referred to as "selection bias") in the treatment selection problem (Dawid, 1994), which is not only fundamental in theory, but also of great practical importance in, e.g., agricultural studies, clinical trials, and large-scale online experiments (Kohavi et al., 2013). In an illuminating paper, Senn (2008) provided a heuristic explanation that the existence of selection bias depended on the prior distribution, and upheld Dawid's claim that the fact that selection bias did not exist in some standard cases was a consequence of using certain conjugate priors. In this paper, we relax the modeling assumptions in Senn (2008) and derive the closed-form expression for the selection bias. Consequently, our work can serve as a complement of the heuristic explanation provided by Senn (2008), and is useful from both theoretical and practical perspectives.

The paper proceeds as follows. Section 2 reviews the treatment selection problem, defines the selection bias, and describes the Bayesian inference framework which the remaining parts of the paper are based on. Section 3 derives the closed-form expression for the selection bias. Section 4 highlights numerical and simulated examples that illustrates the advantages of our theoretical results. Section 5 concludes and discusses future directions.

* Correspondence to: Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, USA.
E-mail address: jiannl@microsoft.com (J. Lu).

http://dx.doi.org/10.1016/j.spl.2016.06.007
0167-7152/

2. Bayesian inference for treatment selection problem

2.1. Treatment selection problem and selection bias

Consider an experiment with $p \geq 2$ treatment arms. For $i = 1, \ldots, p$, let μ_i denote the mean yield of the ith treatment arm. After running the experiment, we observe the sample mean yield of the ith treatment arm, denoted as X_i. Let

$$i^* = \operatorname*{argmax}_{1 \leq i \leq p} X_i$$

denote the index of the largest observation. The focus of selective inference is on μ_{i^*}, which relies on $X_1, \ldots, X_p$. We let $\mathrm{E}(\mu_{i^*} \mid X_{i^*})$ be the posterior mean of μ_{i^*} as if it were selected before the experiment, and

$$\mathrm{E}(\mu_{i^*} \mid X_{i^*}, X_{i^*} = \max X_i)$$

be the "exact post-selection" posterior mean of μ_{i^*}, which takes the selection into account. Following Senn (2008), we define the selection bias as

$$\Delta = \mathrm{E}(\mu_{i^*} \mid X_{i^*}) - \mathrm{E}(\mu_{i^*} \mid X_{i^*}, X_{i^*} = \max X_i). \tag{1}$$

Having defined the selection bias, we briefly discuss the "selection paradox" in Dawid (1994), i.e., "since Bayesian posterior distributions are already fully conditioned on the data, the posterior distribution of any quantity is the same, whether it was chosen in advance or selected in the light of the data". In other words, if we define the selection bias as

$$\tilde{\Delta} = \mathrm{E}(\mu_{i^*} \mid X_1, \ldots, X_p) - \mathrm{E}(\mu_{i^*} \mid X_1, \ldots, X_p, X_{i^*} = \max X_i),$$

then indeed $\tilde{\Delta} = 0$.

2.2. The normal–normal model

Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ and $\boldsymbol{X} = (X_1, \ldots, X_p)'$. Following Dawid (1994), we treat them as random vectors. We generalize Senn (2008) and assume that

$$\boldsymbol{\mu} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0), \qquad \boldsymbol{X} \mid \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2}$$

where

$$\boldsymbol{\Sigma}_0 = \gamma^2 \boldsymbol{I}_p + (1 - \gamma^2)\mathbf{1}_p\mathbf{1}_p', \qquad \boldsymbol{\Sigma} = \sigma^2\{\eta^2 \boldsymbol{I}_p + (1 - \eta^2)\mathbf{1}_p\mathbf{1}_p'\}, \quad 0 \leq \gamma, \eta \leq 1. \tag{3}$$

To interpret (3) we let $X_i = \mu_i + \epsilon_i$, where μ_i is generated by

$$\phi \sim N\left(0, 1 - \gamma^2\right), \qquad \mu_i \mid \phi \sim N\left(\phi, \gamma^2\right),$$

and ϵ_i is generated by

$$\xi \sim N\{0, (1 - \eta^2)\sigma^2\}, \qquad \epsilon_i \mid \xi \sim N(\xi, \eta^2\sigma^2).$$

Note that $\eta = 1$ in Senn (2008), and we relax this assumption by allowing correlated errors.

2.3. Posterior mean

To derive the posterior mean of μ_p given $X_1, \ldots, X_p$, we rely on the following classic result.

Lemma 1 (*Normal Shrinkage*). *Let*

$$\mu \sim N(\mu_0, \nu^2), \qquad Z_i \mid \mu \overset{i.i.d.}{\sim} N(\mu, \tau^2) \quad (i = 1, \ldots, n).$$

Then the posterior mean of μ *is*

$$\mathrm{E}(\mu \mid Z_1, \ldots, Z_n) = \frac{\tau^2\mu_0 + \nu^2 \sum_{i=1}^{n} Z_i}{\tau^2 + n\nu^2},$$