Computational Aspects of Bayesian Solution Estimators in Stochastic Optimization

Danial Davarnia
Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA, USA, davarnia@iastate.edu

Burak Kocuk
Industrial Engineering Program, Sabancı University, Istanbul, Turkey, burakkocuk@sabanciuniv.edu

Gérard Cornuéjols
Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA, gc0v@andrew.cmu.edu

Appendix A: Omitted proofs from Section 2

Proof of Proposition 1 We write that

$$\mathcal{R}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})) = \mathbb{E}_\theta \mathbb{E}_{\bar{\xi}|\theta} \left[\mathcal{L}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})) \right] = \mathbb{E}_{\bar{\xi}} \mathbb{E}_{\theta|\bar{\xi}} \left[\mathcal{L}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})) \right],$$

where the first equality follows from the definition of the linear risk, and the second equality follows from exchanging the order of sequential expectations. According to the definition of Bayes solution estimator, we seek among all $\hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}) \in \mathcal{X}$ an estimator that minimizes the risk $\mathcal{R}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}))$. First, we claim that any minimizer $\hat{\boldsymbol{x}}^{J,L}(\bar{\boldsymbol{\xi}})$ of $\mathbb{E}_{\theta|\bar{\xi}} \left[\mathcal{L}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})) \right]$ is also a minimizer of the risk $\mathcal{R}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}))$. To prove this claim, consider any estimator $\hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}) \neq \hat{\boldsymbol{x}}^{J,L}(\bar{\boldsymbol{\xi}})$. It follows from the assumption that $\mathbb{E}_{\theta|\bar{\xi}} \left[\mathcal{L}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}^{J,L}(\bar{\boldsymbol{\xi}})) \right] \leq \mathbb{E}_{\theta|\bar{\xi}} \left[\mathcal{L}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})) \right]$. Taking the expectation $\mathbb{E}_{\bar{\xi}}[.]$ with respect to the marginal distribution of $\bar{\boldsymbol{\xi}}$ from both sides, we obtain that $\mathbb{E}_{\bar{\xi}} \mathbb{E}_{\theta|\bar{\xi}} \left[\mathcal{L}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}^{J,L}(\bar{\boldsymbol{\xi}})) \right] \leq \mathbb{E}_{\bar{\xi}} \mathbb{E}_{\theta|\bar{\xi}} \left[\mathcal{L}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})) \right]$. Hence, the claim follows from the chain relation in the first line. As a result, a Bayes solution estimator is a minimizer $\hat{\boldsymbol{x}}^{J,L}(\bar{\boldsymbol{\xi}})$ of $\mathbb{E}_{\theta|\bar{\xi}} \left[\mathcal{L}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})) \right] = \mathbb{E}_{\theta|\bar{\xi}} \mathbb{E}_{\xi|\theta}[f(\boldsymbol{\xi}, \boldsymbol{x}^*(\boldsymbol{\theta}))] - \mathbb{E}_{\theta|\bar{\xi}} \mathbb{E}_{\xi|\theta}[f(\boldsymbol{\xi}, \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}))]$. Recall that $\boldsymbol{x}^*(\boldsymbol{\theta})$ is independent of $\hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})$, i.e., the term $\mathbb{E}_{\theta|\bar{\xi}} \mathbb{E}_{\xi|\theta}[f(\boldsymbol{\xi}, \boldsymbol{x}^*(\boldsymbol{\theta}))]$ is fixed regardless of the value of $\hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})$. Therefore, any maximizer of $\mathbb{E}_{\theta|\bar{\xi}} \mathbb{E}_{\xi|\theta}[f(\boldsymbol{\xi}, \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}))]$ is also a minimizer of $\mathbb{E}_{\theta|\bar{\xi}} \left[\mathcal{L}^L(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})) \right]$. □

Proof of Proposition 2 Using the result of Proposition 1, it suffices to show that the objective function of (4) matches that of (5). Using Fubini's theorem of integration, the assumption that $\mathbb{E}_{\xi|\bar{\xi}}[|f(\boldsymbol{\xi}, \boldsymbol{x})|]$ is finite for any $\boldsymbol{x} \in \mathcal{X}$ allows for the interchange of the integration order. We have that

$$\begin{aligned} \mathbb{E}_{\theta|\bar{\xi}} \mathbb{E}_{\xi|\theta}[f(\boldsymbol{\xi}, \boldsymbol{x})] &= \int_{\boldsymbol{\theta}} \int_{\boldsymbol{\xi}} f(\boldsymbol{\xi}, \boldsymbol{x}) g(\boldsymbol{\xi}|\boldsymbol{\theta}) \Pi(\boldsymbol{\theta}|\bar{\boldsymbol{\xi}}) d\boldsymbol{\xi} d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\xi}} f(\boldsymbol{\xi}, \boldsymbol{x}) \left(\int_{\boldsymbol{\theta}} g(\boldsymbol{\xi}|\boldsymbol{\theta}) \Pi(\boldsymbol{\theta}|\bar{\boldsymbol{\xi}}) d\boldsymbol{\theta} \right) d\boldsymbol{\xi} = \int_{\boldsymbol{\xi}} f(\boldsymbol{\xi}, \boldsymbol{x}) h(\boldsymbol{\xi}|\bar{\boldsymbol{\xi}}) d\boldsymbol{\xi} = \mathbb{E}_{\xi|\bar{\xi}}[f(\boldsymbol{\xi}, \boldsymbol{x})], \end{aligned}$$

where the second equality is obtained by the interchange of integration order, and the third equality follows from the definition of posterior predictive distribution in Definition 1. □

Proof of Proposition 3 We have that $\mathcal{R}^Q(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})) = \mathbb{E}_\theta \mathbb{E}_{\bar{\xi}|\theta}[\mathcal{L}^Q(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}))] = \mathbb{E}_{\bar{\xi}} \mathbb{E}_{\theta|\bar{\xi}}[\mathcal{L}^Q(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}))]$, where the first equality holds because of (7), and the second equality follows from changing the order of sequential expectations. According to the definition of Bayes estimator, we seek among all $\hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}) \in \mathcal{X}$ an estimator that minimizes the risk $\mathcal{R}^Q(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}))$. We claim that any minimizer $\hat{\boldsymbol{x}}^{J,Q}(\bar{\boldsymbol{\xi}})$ of $\min_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}_{\theta|\bar{\xi}}[\mathcal{L}^Q(\boldsymbol{x}^*(\boldsymbol{\theta}), \boldsymbol{x})]$ is also a minimizer of the risk $\mathcal{R}^Q(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}))$. To prove this claim, consider any estimator $\hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}) \neq \hat{\boldsymbol{x}}^{J,Q}(\bar{\boldsymbol{\xi}})$. It follows from the assumption that $\mathbb{E}_{\theta|\bar{\xi}}[\mathcal{L}^Q(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}^{J,Q}(\bar{\boldsymbol{\xi}}))] \leq \mathbb{E}_{\theta|\bar{\xi}}[\mathcal{L}^Q(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}}))]$. Taking the expectation $\mathbb{E}_{\bar{\xi}}[.]$ with respect to the marginal distribution of $\bar{\boldsymbol{\xi}}$ from both sides we obtain the desired result due to the chain relation in the first line. □

Proof of Corollary 1 Since $\boldsymbol{x}^*(\boldsymbol{\theta})$ is unique, we obtain that $\mathcal{D}^2\big(\boldsymbol{x}^*(\boldsymbol{\theta}), \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})\big) = ||\boldsymbol{x}^*(\boldsymbol{\theta}) - \hat{\boldsymbol{x}}(\bar{\boldsymbol{\xi}})||^2$. To obtain the Bayes solution estimator, we need to find a minimizer of (8) which reduces to $\min_{\boldsymbol{x}\in\mathcal{X}}\ \mathbb{E}_{\theta|\bar{\xi}}[||\boldsymbol{x}^*(\boldsymbol{\theta}) - \boldsymbol{x}||^2]$. Define $\boldsymbol{w} = \mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{x}^*(\boldsymbol{\theta})]$. We write that

$$\begin{aligned}
\mathbb{E}_{\theta|\bar{\xi}}\left[||\boldsymbol{x}^*(\boldsymbol{\theta}) - \boldsymbol{x}||^2\right] &= \mathbb{E}_{\theta|\bar{\xi}}\left[||(\boldsymbol{x}^*(\boldsymbol{\theta}) - \boldsymbol{w}) + (\boldsymbol{w} - \boldsymbol{x})||^2\right] \\
&= \mathbb{E}_{\theta|\bar{\xi}}\left[||\boldsymbol{x}^*(\boldsymbol{\theta}) - \boldsymbol{w}||^2\right] + \mathbb{E}_{\theta|\bar{\xi}}\left[||\boldsymbol{w} - \boldsymbol{x}||^2\right] + 2\mathbb{E}_{\theta|\bar{\xi}}\left[(\boldsymbol{x}^*(\boldsymbol{\theta}) - \boldsymbol{w})^\intercal(\boldsymbol{w} - \boldsymbol{x})\right] \\
&= \mathbb{E}_{\theta|\bar{\xi}}\left[||\boldsymbol{x}^*(\boldsymbol{\theta}) - \boldsymbol{w}||^2\right] + ||\boldsymbol{w} - \boldsymbol{x}||^2,
\end{aligned}$$

where the first equality is obtained by adding and subtracting $\boldsymbol{w}$, the second equality follows from decomposition of the norm vector, and the last equality holds because $\mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{x}^*(\boldsymbol{\theta}) - \boldsymbol{w}] = \boldsymbol{0}$ by definition, and because $||\boldsymbol{w} - \boldsymbol{x}||^2$ does not depend on $\boldsymbol{\theta}$. Note in the last relation that the first term $\mathbb{E}_{\theta|\bar{\xi}}[||\boldsymbol{x}^*(\boldsymbol{\theta}) - \boldsymbol{w}||^2]$ does not contain $\boldsymbol{x}$. This gives the desired relation (9). For the next result, when $\mathcal{X}$ is convex, any convex combination of $\boldsymbol{x}^*(\boldsymbol{\theta})$ belongs to $\mathcal{X}$ as $\boldsymbol{x}^*(\boldsymbol{\theta}) \in \mathcal{X}$. We conclude that $\mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{x}^*(\boldsymbol{\theta})] \in \mathcal{X}$, and therefore the minimizer of (9) is attained at $\mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{x}^*(\boldsymbol{\theta})]$. ☐

Appendix B: Omitted proofs from Section 4

Proof of Proposition 4 It follows from the assumptions that $\mathbb{E}_{\xi|\theta}[f(\boldsymbol{\xi}, \boldsymbol{x})]$ can be written as $\sum_{k=1}^{K} h_k(\boldsymbol{x}) g_k(\boldsymbol{\theta})$ where $h_k(\boldsymbol{x}) : \mathbb{R}^m \to \mathbb{R}$ and $g_k(\boldsymbol{\theta}) = \prod_{i\in I_k} \theta_i$ for some $I_k \subseteq [n]$. For this function, the assumption also implies that for any $k \in \{1, \dots, K\}$, all variables ξ_l for $l \in I_k$ are independent, and so are all variables θ_l for $l \in I_k$. We compute the objective of the Joint-EO method as given in (4)

$$\mathbb{E}_{\theta|\bar{\xi}}\mathbb{E}_{\xi|\theta}[f(\boldsymbol{\xi}, \boldsymbol{x})] = \sum_{k=1}^{K} h_k(\boldsymbol{x}) \mathbb{E}_{\theta|\bar{\xi}}[g_k(\boldsymbol{\theta})] = \sum_{k=1}^{K} h_k(\boldsymbol{x}) \prod_{i\in I_k} \mathbb{E}_{\theta_i|\bar{\xi}_i}[\theta_i], \tag{37}$$

where the first equality follows from linearity of the expectation operator and the second equality holds because random variables and parameters are independent. As discussed before, the Bayes estimator for a parameter under the squared error loss is the posterior mean. Therefore for the Separate-EO method, each θ_i is replaced by its posterior mean $\mathbb{E}_{\theta_i|\bar{\xi}_i}[\theta_i]$ in (1). The resulting objective function is (37). This shows that the objective of the Joint-EO and Separate-EO methods are equal. Since both have the same constraint set $\mathcal{X}$, their optimal solutions are the same. ☐

Proof of Proposition 5 Under the assumption that $\boldsymbol{x}^*(\boldsymbol{\theta})$ is unique for any given $\boldsymbol{\theta}$, the Separate-EO solution estimator can be expressed as $\hat{\boldsymbol{x}}^S(\bar{\boldsymbol{\xi}}) = \boldsymbol{x}^*(\hat{\boldsymbol{\theta}}^B(\bar{\boldsymbol{\xi}}))$ where $\hat{\boldsymbol{\theta}}^B(\bar{\boldsymbol{\xi}})$ is the Bayes estimator of the unknown parameter $\boldsymbol{\theta}$ under the squared error loss. It follows from the discussion in Section 2.1 that $\hat{\boldsymbol{\theta}}^B(\bar{\boldsymbol{\xi}}) = \mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{\theta}]$. Therefore, the Separate-EO method yields $\hat{\boldsymbol{x}}^S(\bar{\boldsymbol{\xi}}) = \boldsymbol{x}^*(\mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{\theta}])$. We obtain that $\mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{x}^*(\boldsymbol{\theta})] = \boldsymbol{x}^*(\mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{\theta}])$ because of the linearity of the expectation operator and because of the independence of variables appearing in the products. As a result, $\mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{x}^*(\boldsymbol{\theta})] \in \mathcal{X}$. It follows from (9) in Corollary 1 that the Joint-EO method yields $\hat{\boldsymbol{x}}^{J,Q}(\bar{\boldsymbol{\xi}}) = \mathbb{E}_{\theta|\bar{\xi}}[\boldsymbol{x}^*(\boldsymbol{\theta})]$. ☐

Appendix C: Omitted example from Section 4

Example 5. Assume the setting of Example 4. Consider an instance of stochastic program (1) where the objective function is $\mathbb{E}_{\xi|\mu}[f(\boldsymbol{\xi},\boldsymbol{x})]=\sum_{i=1}^n \mu_i x_i$ and $\mathcal{X}$ is a unit-ball in $\mathbb{R}^n$, i.e.,

$$\max\left\{\sum_{i=1}^n \mu_i x_i \,\middle|\, \sum_{i=1}^n x_i^2 \le 1\right\}. \tag{38}$$

For any $\boldsymbol{\mu}\in\mathbb{R}^n\setminus\{0\}$, the unique optimal solution of the problem is $\boldsymbol{x}^*(\boldsymbol{\mu})=\frac{\boldsymbol{\mu}}{||\boldsymbol{\mu}||}$. As a result, the Separate-EO method yields the solution estimator $\boldsymbol{x}^S(\bar{\boldsymbol{\xi}})=\frac{\boldsymbol{\eta}}{||\boldsymbol{\eta}||}$. Now we calculate the estimator obtained from the Joint-EO method under the quadratic loss. Since the optimal solution of (38) is unique and its feasible region is convex, it follows from Corollary 1 that $\hat{\boldsymbol{x}}^{J,Q}(\bar{\boldsymbol{\xi}})=\mathbb{E}_{\mu|\bar{\xi}}[\boldsymbol{x}^*(\boldsymbol{\mu})]=\mathbb{E}_{\mu|\bar{\xi}}\left[\frac{\boldsymbol{\mu}}{||\boldsymbol{\mu}||}\right]$. In particular, the Joint-EO solution is a convex combination of normalized vectors $\frac{\boldsymbol{\mu}}{||\boldsymbol{\mu}||}$ for all possible values of $\boldsymbol{\mu}$ taken according to the posterior distribution of $\boldsymbol{\mu}|\bar{\boldsymbol{\xi}}$. When this distribution is non-degenerate (can assume multiple distinct values), the expected vector $\hat{\boldsymbol{x}}^{J,Q}(\bar{\boldsymbol{\xi}})$ belongs to the interior of the unit-ball. On the other hand, the Separate-EO solution $\boldsymbol{x}^S(\bar{\boldsymbol{\xi}})$ is always on the boundary of the unit-ball. We conclude that the two solution estimators can never be equal. Moreover, they can achieve a maximum distance in the unit-ball. For instance, assume that the parameters $\boldsymbol{\lambda}$, $\boldsymbol{\sigma}$, $\boldsymbol{\delta}$ and the observation $\bar{\boldsymbol{\xi}}$ are such that $|\eta_i|=\left|\frac{\sigma_i^2}{\sigma_i^2+\delta_i^2}\lambda_i+\frac{\delta_i^2}{\sigma_i^2+\delta_i^2}\bar{\xi}_i\right|\le\epsilon$ for all $i\in[n]$ and a sufficiently small but positive ϵ. Due to the symmetry of the feasible region and the posterior distribution of $\boldsymbol{\mu}|\bar{\boldsymbol{\xi}}$ around the origin, the Joint-EO solution estimator is sufficiently close to the origin, while the Separate-EO solution estimator is always on the boundary. This yields the maximum distance of the Separate-EO solution from the Joint-EO solution.

For a numerical illustration, assume that $n=2$, $\bar{\xi}_1=1.001$, $\bar{\xi}_2=-1$, $\lambda_1=-1,\lambda_2=1$, $\sigma_1=\sigma_2=\delta_1=\delta_2=1$. We compute $\eta_1=0.0005$, $\eta_2=0$ and $\zeta_1^2=\zeta_2^2=\frac{1}{2}$. It follows that the Bayes solution estimator for the Joint-EO is very close to the origin and the solution estimator for the Separate-EO is $\boldsymbol{x}=(1,0)$. ■

Appendix D: Omitted analyses from Section 5

D.1. Piecewise linear functions

In this section, we consider piecewise linear structures with exponential-gamma and geometric-beta conjugate pairs.

D.1.1. Exponential Likelihood with Gamma Prior Using Corollary 3, one can prove the following.

Proposition 12. *Consider the stochastic problem* (26). *Assume that the likelihood distribution is exponential with $\xi\sim Exp(\lambda)$ and the prior distribution is gamma with $\lambda\sim Gamma(\alpha,\beta)$. Assume further that the shape and rate hyperparameters α and β are known, and a realization $\bar{\xi}$ is observed. Then, we have*

(i) $\hat{x}^S(\bar{\xi})=\frac{\beta+\bar{\xi}}{\alpha+1}\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)$.

(ii) $\hat{x}^{J,L}(\bar{\xi})=(\beta+\bar{\xi})\left[\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)^{1/(\alpha+1)}-1\right]$.

(iii) $\hat{x}^{J,Q}(\bar{\xi})=\frac{\beta+\bar{\xi}}{\alpha}\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)$.

Proof: (i) Due to Corollary 3(i), we have

$$\hat{x}^S(\bar{\xi})=G^{-1}\left(\frac{\tilde{b}}{\bar{a}-\tilde{a}}\,\middle|\,\hat{\lambda}^B(\bar{\xi})\right)=-\frac{1}{\hat{\lambda}^B(\bar{\xi})}\ln\left(1-\frac{\tilde{b}}{\bar{a}-\tilde{a}}\right)=\frac{1}{\hat{\lambda}^B(\bar{\xi})}\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right),$$

where $G(\xi)$ is the cdf of an exponential random variable with parameter $\hat{\lambda}^B(\bar{\xi})=\frac{\alpha+1}{\beta+\bar{\xi}}$. Hence, we obtain $\hat{x}^S(\bar{\xi})$ as stated.

(ii) Due to Corollary 3(ii), we have

$$\hat{x}^{J,L}(\bar{\xi}) = H^{-1}\left(\frac{\tilde{b}}{\bar{a}-\tilde{a}}\middle|\bar{\xi}\right) = \beta'\left[\left(1-\frac{\tilde{b}}{\bar{a}-\tilde{a}}\right)^{-1/\alpha'}-1\right] = \beta'\left[\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)^{1/\alpha'}-1\right],$$

where $H(\xi)$ is the cdf of a Lomax random variable with the scale and shape parameters $\beta' := \beta + \bar{\xi}$ and $\alpha' = \alpha + 1$. Hence, the result follows.

(iii) We write that

$$\begin{aligned}\hat{x}^{J,Q}(\bar{\xi}) =& \int_0^\infty \frac{1}{\lambda}\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\Pi(\lambda|\bar{\xi})d\lambda \\ =& \ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\int_0^\infty \frac{1}{\lambda}\frac{(\beta+\bar{\xi})^{\alpha+1}}{\Gamma(\alpha+1)}\lambda^\alpha e^{-(\beta+\bar{\xi})\lambda}d\lambda \\ =& \ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)(\beta+\bar{\xi})\frac{\Gamma(\alpha)}{\Gamma(\alpha+1)}\int_0^\infty \frac{(\beta+\bar{\xi})^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-(\beta+\bar{\xi})\lambda}d\lambda \\ =& \ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\frac{(\beta+\bar{\xi})}{\alpha},\end{aligned}$$

where the first equality follows from Corollaries 1 and 3(iii), the second equality holds since the posterior distribution $\Pi(\lambda|\bar{\xi})$ is gamma with the shape and rate parameters $\alpha+1$ and $\beta+\bar{\xi}$ respectively, the third equality is obtained by factoring suitable terms out of the integral and the last equality follows from the facts that $\frac{\Gamma(\alpha)}{\Gamma(\alpha+1)} = \frac{1}{\alpha}$ and that the integral is equal to 1 as it represents a gamma distribution. □

We note that the complexity of obtaining the Separate-EO and Joint-EO estimators is the same as they all admit closed form solutions. Next, we discuss the risk difference between $\hat{x}^S(\bar{\xi})$ and $\hat{x}^{J,L}(\bar{\xi})$ as well as $\hat{x}^{J,Q}(\bar{\xi})$.

Proposition 13. *Let $\alpha > 1$. Under the assumptions of Proposition 12, we have*

$$\mathcal{R}^L\left(x^*(\mu), \hat{x}^S(\bar{\xi})\right) - \mathcal{R}^L\left(x^*(\mu), \hat{x}^{J,L}(\bar{\xi})\right) \le \frac{\tilde{b}\beta\alpha}{\alpha-1}\left\{\left[\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)^{1/(\alpha+1)}-1\right]-\frac{1}{\alpha+1}\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right\}.$$

Proof: Let us first compute an upper bound on the difference in the loss values of the estimators. We define $F(x) := \mathbb{E}_{\xi|\mu}[f(\xi,x)]$ as the objective function of (26) and write

$$\begin{aligned}&\mathcal{L}^L(x^*(\theta), \hat{x}^S(\bar{\xi})) - \mathcal{L}^L(x^*(\theta), \hat{x}^{J,L}(\bar{\xi})) = F(\hat{x}^{J,L}(\bar{\xi})) - F(\hat{x}^S(\bar{\xi})) \\ \le& F'(\hat{x}^S(\bar{\xi}))(\hat{x}^{J,L}(\bar{\xi}) - \hat{x}^S(\bar{\xi})) \\ \le& \tilde{b}(\beta+\bar{\xi})\left\{\left[\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)^{1/(\alpha+1)}-1\right]-\frac{1}{\alpha+1}\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right\},\end{aligned}$$

where the first equality follows from the definition of the linear loss (2), the first inequality is obtained from the first order Taylor expansion of the concave function $F(x)$ at point $\hat{x}^{J,L}(\bar{\xi})$ about $\hat{x}^S(\bar{\xi})$, the second inequality follows from (i) $\hat{x}^{J,L}(\bar{\xi}) - \hat{x}^S(\bar{\xi}) = (\beta+\bar{\xi})\left[\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)^{1/(\alpha+1)}-1\right] - \frac{\beta+\bar{\xi}}{\alpha+1}\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)$ due to Proposition 12, (ii) $\hat{x}^{J,L}(\bar{\xi}) > \hat{x}^S(\bar{\xi})$ since $\kappa^{1/a} - 1 > \frac{\ln(\kappa)}{a}$ for all $\kappa, a > 1$, and (iii) $F'(x) \le \tilde{b}$ for all $x \in \mathbb{R}$ which is deduced from (25). To obtain the risk difference, we take the expectations $\mathbb{E}_\lambda\mathbb{E}_{\bar{\xi}|\lambda}$ from the last term, which yields the desired result using $\mathbb{E}_\lambda\mathbb{E}_{\bar{\xi}|\lambda}[\bar{\xi}] = \frac{\beta}{\alpha-1}$. □

For the quadratic loss, it is possible to compute the risk difference exactly.

Proposition 14. *Let $\alpha > 2$. Under the assumptions of Proposition 12, we have*

$$\mathcal{R}^Q\left(x^*(\mu), \hat{x}^S(\bar{\xi})\right) - \mathcal{R}^Q\left(x^*(\mu), \hat{x}^{J,Q}(\bar{\xi})\right) = \frac{\beta^2}{\alpha(\alpha+1)^2(\alpha-2)}\left[\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right]^2.$$

Proof: We write that

$$\begin{aligned}
\mathcal{R}^Q\left(x^*(\mu), \hat{x}^S\right) &= \left[\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right]^2 \mathbb{E}_\lambda \mathbb{E}_{\bar{\xi}|\lambda}\left[\left(\frac{1}{\lambda} - \frac{\beta+\bar{\xi}}{\alpha+1}\right)^2\right] \\
&= \left[\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right]^2 \mathbb{E}_\lambda \mathbb{E}_{\bar{\xi}|\lambda}\left[\frac{1}{\lambda^2} - \frac{2}{\lambda}\frac{\beta+\bar{\xi}}{\alpha+1} + \frac{(\beta+\bar{\xi})^2}{(\alpha+1)^2}\right] \\
&= \left[\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right]^2 \mathbb{E}_\lambda\left[\frac{1}{\lambda^2} - \frac{2}{\lambda}\frac{\beta+\frac{1}{\lambda}}{\alpha+1} + \frac{\beta^2+2\frac{\beta}{\lambda}+\frac{2}{\lambda^2}}{(\alpha+1)^2}\right] \\
&= \left[\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right]^2 \left\{\mathbb{E}_\lambda\left[\frac{1}{\lambda^2}\right] - \frac{2}{\alpha+1}\mathbb{E}_\lambda\left[\frac{\beta}{\lambda}+\frac{1}{\lambda^2}\right] + \frac{1}{(\alpha+1)^2}\left(\beta^2 + 2\mathbb{E}_\lambda\left[\frac{\beta}{\lambda}+\frac{1}{\lambda^2}\right]\right)\right\},
\end{aligned}$$

where the first equality follows from Proposition 12 and the definition (6) of the risk under quadratic loss, and the second equality holds because $\mathbb{E}_{\xi|\lambda}[\xi] = \frac{1}{\lambda}$ and $\mathbb{E}_{\xi|\lambda}[\xi^2] = \frac{2}{\lambda^2}$. Using similar arguments, we can compute

$$\mathcal{R}^Q\left(x^*(\mu), \hat{x}^{J,Q}(\bar{\xi})\right) = \left[\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right]^2 \left\{\mathbb{E}_\lambda\left[\frac{1}{\lambda^2}\right] - \frac{2}{\alpha}\mathbb{E}_\lambda\left[\frac{\beta}{\lambda}+\frac{1}{\lambda^2}\right] + \frac{1}{\alpha^2}\left(\beta^2 + 2\mathbb{E}_\lambda\left[\frac{\beta}{\lambda}+\frac{1}{\lambda^2}\right]\right)\right\}.$$

Note that $\lambda \sim \text{Gamma}(\alpha,\beta)$ which implies that $\frac{1}{\lambda} \sim \text{Inverse-Gamma}(\alpha,\beta)$. Therefore, we obtain that $\mathbb{E}_\lambda[\frac{1}{\lambda}] = \frac{\beta}{\alpha-1}$ and $\mathbb{E}_\lambda[\frac{1}{\lambda^2}] = \frac{\beta^2}{(\alpha-1)(\alpha-2)}$, which yields $\mathbb{E}_\lambda[\frac{\beta}{\lambda}+\frac{1}{\lambda^2}] = \frac{\beta^2}{\alpha-2}$. Combining the above results, we obtain

$$\begin{aligned}
&\mathcal{R}^Q\left(x^*(\mu), \hat{x}^S(\bar{\xi})\right) - \mathcal{R}^Q\left(x^*(\mu), \hat{x}^{J,Q}(\bar{\xi})\right) \\
&= \left[\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right]^2 \left\{2\left(\frac{1}{\alpha} - \frac{1}{\alpha+1}\right)\frac{\beta^2}{\alpha-2} + \left(\frac{1}{(\alpha+1)^2} - \frac{1}{\alpha^2}\right)\left(\beta^2 + \frac{2\beta^2}{\alpha-2}\right)\right\} \\
&= \left[\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right]^2 \left\{\frac{2}{\alpha(\alpha+1)}\frac{\beta^2}{\alpha-2} + \frac{-2\alpha-1}{\alpha^2(\alpha+1)^2}\frac{\beta^2\alpha}{\alpha-2}\right\} \\
&= \left[\ln\left(\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}}\right)\right]^2 \frac{\beta^2}{\alpha(\alpha+1)(\alpha-2)}\left(2 - \frac{2\alpha+1}{\alpha+1}\right),
\end{aligned}$$

from which the result follows. □

D.1.2. Geometric Likelihood with Beta Prior Corollary 3 implies the following.

Proposition 15. *Consider the stochastic problem* (26)*. Assume that the likelihood distribution is geometric with $\xi \sim Geo(p)$ and the prior distribution is beta with $p \sim Beta(\alpha,\beta)$. Assume further that the shape parameters α and β are known, and a realization $\bar{\xi}$ is observed from the likelihood. Then, we have*

(i) $\hat{x}^S(\bar{\xi}) \approx \frac{\ln\left(\frac{\bar{a}-\tilde{a}-\tilde{b}}{\bar{a}-\tilde{a}}\right)}{\ln[(\beta+\bar{\xi}-1)/(\alpha+\beta+\bar{\xi})]}$.

(ii) $\hat{x}^{J,L}(\bar{\xi}) \approx \max\left\{x : \sum_{\xi=0}^{x} \frac{\alpha+1}{\alpha+\beta+\bar{\xi}+\xi+2}\ \frac{\beta+\bar{\xi}+\xi}{\alpha+\beta+\bar{\xi}+\xi+1}\ \frac{\beta+\bar{\xi}+\xi-1}{\alpha+\beta+\bar{\xi}+\xi} \le \frac{\tilde{b}}{\bar{a}-\tilde{a}}\right\}$.

(iii) $\hat{x}^{J,Q}(\bar{\xi}) \approx \frac{\ln\left(\frac{\bar{a}-\tilde{a}-\tilde{b}}{\bar{a}-\tilde{a}}\right)}{B(\alpha+1,\beta+\bar{\xi}-1)}\int_0^1 \frac{p^\alpha(1-p)^{\beta+\bar{\xi}-2}}{\ln(1-p)}dp$.

Proof: (i) Due to Corollary 3(i), we have

$$\hat{x}^S(\bar{\xi}) = G^{-1}\left(\frac{\tilde{b}}{\bar{a}-\tilde{a}}\middle|\hat{p}^B(\bar{\xi})\right) \approx \frac{\ln\left(\frac{\bar{a}-\tilde{a}-\tilde{b}}{\bar{a}-\tilde{a}}\right)}{\ln(1-\hat{p}^B(\bar{\xi}))},$$

where $G(\xi)$ is the cdf of a geometric random variable with parameter $\hat{p}^B(\bar{\xi}) = \frac{\alpha+1}{\alpha+\beta+\bar{\xi}}$. Hence, we obtain $\hat{x}^S(\bar{\xi})$ as stated.

(ii) Due to Corollary 3(ii), we have

$$\hat{x}^{J,L}(\bar{\xi}) = H^{-1}\left(\frac{\tilde{b}}{\bar{a}-\tilde{a}}\middle|\bar{\xi}\right),$$

where H is the cdf of posterior predictive distribution. The results follows due to the relationship between the cdf H and its pmf h.

(iii) We write that

$$\hat{x}^{J,Q}(\bar{\xi}) \approx \int_0^1 \frac{\ln\left(\frac{\bar{a}-\tilde{a}-\tilde{b}}{\bar{a}-\tilde{a}}\right)}{\ln(1-p)}\Pi(p|\bar{\xi})dp = \ln\left(\frac{\bar{a}-\tilde{a}-\tilde{b}}{\bar{a}-\tilde{a}}\right)\int_0^1 \frac{1}{\ln(1-p)}\frac{p^{\alpha}(1-p)^{\beta+\bar{\xi}-2}}{\mathrm{B}(\alpha+1,\beta+\bar{\xi}-1)}dp,$$

where the first relation follows from Corollaries 1 and 3(iii) and the second relation holds since the posterior distribution $\Pi(p|\bar{\xi})$ is beta with parameters $\alpha+1$ and $\beta+\bar{\xi}-1$. □

We note that $\hat{x}^S(\bar{\xi})$ can be approximated by a closed form expression while the approximations of $\hat{x}^{J,L}(\bar{\xi})$ and $\hat{x}^{J,Q}(\bar{\xi})$ require an algorithm and numerical integration, in general. This is an instance where computing the Joint-EO solutions is harder than computing the Separate-EO solutions.

D.2. Sum of piecewise linear functions

In this section, we consider the sum of piecewise linear functions with exponential-gamma conjugate pairs.

Proposition 16. *Consider the one-dimensional stochastic median problem. Assume that, for each $i \in [n]$, the likelihood distribution is exponential with $\xi_i \sim Exp(\lambda_i)$ and the prior distribution is gamma with $\lambda_i \sim Gamma(\alpha_i, \beta_i)$. Assume further that the parameters α_i and β_i are known with $\alpha_i > 2$, and that a realization of locations $\bar{\xi}_i$ is observed for $i \in [n]$. Then, we have*

$$\mathcal{R}^L\left(x^*(\boldsymbol{\lambda}), \hat{x}^S(\bar{\xi})\right) - \mathcal{R}^L\left(x^*(\boldsymbol{\lambda}), \hat{x}^{J,L}(\bar{\xi})\right) \le n\Bigg(\max_i\left\{\frac{\alpha_i\beta_i(\sqrt[\alpha_i+1]{2}-1)}{\alpha_i-1}\right\}$$
$$+\sqrt{\frac{n-1}{n}\sum_{i=1}^n \frac{\beta_i^2\alpha_i(\sqrt[\alpha_i+1]{2}-1)}{(\alpha_i-1)^2(\alpha_i-2)}} - \min_i\left\{\frac{\alpha_i\beta_i\ln 2}{\alpha_i-1}\right\} + \sqrt{\frac{n-1}{n}\sum_{i=1}^n\frac{\beta_i^2\alpha_i(\ln 2)^2}{(\alpha_i-1)^2(\alpha_i-2)(\alpha_i+1)^2}}\Bigg).$$

Proof: Proposition 12 implies that $\hat{x}_i^S(\bar{\xi}_i) = \frac{\beta_i+\bar{\xi}_i}{\alpha_i+1}\ln 2$ and $\hat{x}_i^{J,L}(\bar{\xi}_i) = (\beta_i+\bar{\xi}_i)(\sqrt[\alpha_i+1]{2}-1)$ as $\frac{\bar{a}-\tilde{a}}{\bar{a}-\tilde{a}-\tilde{b}} = 2$ for $i \in [n]$. Since $\frac{\ln 2}{\alpha_i+1} < \sqrt[\alpha_i+1]{2}-1$ for $\alpha_i > 0$, we have $\hat{x}_i^S(\bar{\xi}_i) < \hat{x}_i^{J,L}(\bar{\xi}_i)$ for each $i \in [n]$. Applying Proposition 10, we obtain

$$\mathcal{R}^L\left(x^*(\boldsymbol{\lambda}), \hat{x}^S(\bar{\xi})\right) - \mathcal{R}^L\left(x^*(\boldsymbol{\lambda}), \hat{x}^{J,L}(\bar{\xi})\right) \le n\mathbb{E}_{\bar{\boldsymbol{\xi}}}[\max_i\{\hat{x}_i^{J,L}(\bar{\xi}_i)\} - \min_i\{\hat{x}_i^S(\bar{\xi}_i)\}]. \tag{39}$$

Now, we compute an upper bound on the right-hand-side using Aven (1985). Since $\bar{\xi}_i \sim \mathrm{Lomax}(\beta_i, \alpha_i)$ for $i \in [n]$, we have

$$\mathbb{E}[\xi_i] = \frac{\beta_i}{\alpha_i-1} \text{ and } \mathrm{Var}(\xi_i) = \frac{\beta_i^2\alpha_i}{(\alpha_i-1)^2(\alpha_i-2)}.$$

Since both $\hat{x}_i^S(\bar{\xi}_i)$ and $\hat{x}_i^{J,L}(\bar{\xi}_i)$ are affine transformations of ξ_i, we can easily obtain their mean and variance as well. Finally, plugging in the resulting bounds for $\mathbb{E}[\min_i\{\hat{x}_i^S(\bar{\xi}_i)\}]$ and $\mathbb{E}[\max_i\{\hat{x}_i^{J,L}(\bar{\xi}_i)\}]$ into (39) gives the desired result. □

References

Aven T (1985) Upper (lower) bounds on the mean of the maximum (minimum) of a number of random variables. *Journal of Applied Probability* 22(3):723–728.