

1. Let X_1, \dots, X_n be iid from $\text{Gamma}(a, b)$ where a is known.

(a) Express the likelihood function $f(X_1, \dots, X_n|b)$ in terms of $\eta = -\frac{1}{b}$ and find the conjugate prior for η .

The density of a $\text{Gamma}(a, b)$ distribution is

$$f(x|\eta) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b} = \frac{1}{\Gamma(a)} x^{a-1} (-\eta)^a e^{\eta x} = \frac{1}{\Gamma(a)} x^{a-1} e^{\eta x - (-a) \log(-\eta)}.$$

The likelihood function is thus

$$f(\mathbf{X}|\eta) = \prod_{i=1}^n f(X_i|\eta) = \left(\frac{1}{\Gamma(a)} \right)^n \left(\prod_{i=1}^n X_i \right)^{a-1} \exp \left(\eta \sum_{i=1}^n X_i - n(-a) \log(-\eta) \right).$$

The conjugate prior family is

$$\pi(\eta|k, \mu) = c(k, \mu) \exp(k\eta\mu - k(-a) \log(-\eta))$$

where μ can be thought of as a prior mean and k is proportional to a prior variance.

(b) Using the prior for η in (a), find the Bayes estimator under the losses (i) $L(b, \delta) = (b - \delta)^2$ and (ii) $L(b, \delta) = (1 - \delta/b)^2$.

The posterior distribution is

$$\begin{aligned} \pi(\eta|\mathbf{x}, k, \mu) &\propto \exp \left(\eta \sum_{i=1}^n x_i - n(-a) \log(-\eta) \right) \exp(k\eta\mu - k(-a) \log(-\eta)) \\ &= \exp(\eta(n\bar{x} + k\mu) - (n+k)(-a) \log(-\eta)), \end{aligned}$$

(i) Using the conclusion of Problem 3.9 in Theory of Point Estimation, we know that

$$\mathbb{E} \left(\frac{\partial(-a) \log(-\eta)}{\partial \eta} | \mathbf{x}, k, \mu \right) = \mathbb{E} \left(-\frac{a}{\eta} \right) = \frac{n\bar{x} + k\mu}{n+k}.$$

Under $L(b, \delta) = (b - \delta)^2$, using Corollary 1.2, we know the Bayes estimator of b is

$$\delta(\mathbf{x}) = \mathbb{E}(b|\mathbf{x}) = \mathbb{E} \left(-\frac{1}{\eta} | \mathbf{x} \right) = \frac{n\bar{x} + k\mu}{a(n+k)}.$$

(ii) Under $L(b, \delta) = (1 - \delta/b)^2 = (b - \delta)^2/b^2$, using Corollary 1.2, we know the Bayes estimator of b is

$$\delta(\mathbf{x}) = \frac{\mathbb{E}(1/b|\mathbf{x})}{\mathbb{E}(1/b^2|\mathbf{x})} = \frac{\mathbb{E}(-\eta|\mathbf{x})}{\mathbb{E}(\eta^2|\mathbf{x})}.$$

Since the posterior distribution of η is proportional to

$$\exp(\eta(n\bar{x} + k\mu) - (n+k)(-a) \log(-\eta)),$$

we know $-\eta|\mathbf{x}$ follows Gamma distribution with shape $a(n+k) + 1$ and scale $1/(n\bar{x} + k\mu)$ so

$$\mathbb{E}(-\eta|\mathbf{x}) = \frac{a(n+k) + 1}{n\bar{x} + k\mu},$$

and

$$\text{Var}(-\eta|\mathbf{x}) = \frac{a(n+k) + 1}{(n\bar{x} + k\mu)^2} \implies \mathbb{E}(\eta^2|\mathbf{x}) = \frac{(a(n+k) + 1) + (a(n+k) + 1)^2}{(n\bar{x} + k\mu)^2},$$

so

$$\delta(\mathbf{x}) = \frac{\mathbb{E}(-\eta|\mathbf{x})}{\mathbb{E}(\eta^2|\mathbf{x})} = \frac{a(n+k) + 1}{n\bar{x} + k\mu} \frac{(n\bar{x} + k\mu)^2}{(a(n+k) + 1) + (a(n+k) + 1)^2} = \frac{n\bar{x} + k\mu}{a(n+k) + 2}.$$

(c) The problem is misleading, in fact it lies in the Chapter 4 (3.9) of the indicated book.

2. Page 242, Corollary 3.3 in Chapter 4, Lehmann and Casella (1998): If $\mathbf{X} = (X_1, \dots, X_p)$ has the density

$$p_{\boldsymbol{\eta}}(\mathbf{x}) = \exp \left(\sum_{i=1}^p \eta_i x_i - A(\boldsymbol{\eta}) \right) h(\mathbf{x})$$

and $\boldsymbol{\eta}$ has prior density $\pi(\boldsymbol{\eta})$, the Bayes estimator of $\boldsymbol{\eta}$ under the loss $L(\boldsymbol{\eta}, \delta) = \sum (\eta_i - \delta_i)^2$ is given by

$$\mathbb{E}(\eta_i | \mathbf{x}) = \frac{\partial}{\partial x_i} \log m(\mathbf{x}) - \frac{\partial}{\partial x_i} \log h(\mathbf{x}).$$

We prove Theorem 3.2 first: If \mathbf{X} has density

$$p_{\boldsymbol{\eta}}(\mathbf{x}) = \exp \left(\sum_{i=1}^p \eta_i T_i(\mathbf{x}) - A(\boldsymbol{\eta}) \right) h(\mathbf{x})$$

and $\boldsymbol{\eta}$ has prior density $\pi(\boldsymbol{\eta})$, then for $j = 1, \dots, n$,

$$\mathbb{E} \left(\sum_{i=1}^s \eta_i \frac{\partial T_i(\mathbf{x})}{\partial x_j} | \mathbf{x} \right) = \frac{\partial}{\partial x_j} \log m(\mathbf{x}) - \frac{\partial}{\partial x_j} \log h(\mathbf{x})$$

where $m(\mathbf{x}) = \int p_{\boldsymbol{\eta}}(\mathbf{x}) \pi(\boldsymbol{\eta}) d\boldsymbol{\eta}$ is the marginal distribution of \mathbf{X} .

Proof. Note that

$$\frac{\partial}{\partial x_j} \exp \left(\sum_{i=1}^s \eta_i T_i \right) = \left(\sum_{i=1}^s \eta_i \frac{\partial T_i}{\partial x_j} \right) \exp \left(\sum_{i=1}^s \eta_i T_i \right),$$

we have

$$\begin{aligned} & \mathbb{E} \left(\sum_{i=1}^s \eta_i \frac{\partial T_i(\mathbf{x})}{\partial x_j} | \mathbf{x} \right) \\ &= \int \left(\sum_{i=1}^s \eta_i \frac{\partial T_i(\mathbf{x})}{\partial x_j} \right) \pi(\boldsymbol{\eta} | \mathbf{x}) d\boldsymbol{\eta} \\ &= \frac{1}{m(\mathbf{x})} \int \left(\sum_{i=1}^s \eta_i \frac{\partial T_i(\mathbf{x})}{\partial x_j} \right) \exp \left(\sum_{i=1}^p \eta_i T_i(\mathbf{x}) - A(\boldsymbol{\eta}) \right) h(\mathbf{x}) \pi(\boldsymbol{\eta}) d\boldsymbol{\eta} \\ &= \frac{1}{m(\mathbf{x})} \int \left[h(\mathbf{x}) \frac{\partial}{\partial x_j} \exp \left(\sum_{i=1}^s \eta_i T_i \right) \right] \exp(-A(\boldsymbol{\eta})) \pi(\boldsymbol{\eta}) d\boldsymbol{\eta} \\ &= \frac{1}{m(\mathbf{x})} \int \left[\frac{\partial}{\partial x_j} \exp \left(\sum_{i=1}^s \eta_i T_i \right) h(\mathbf{x}) - \exp \left(\sum_{i=1}^s \eta_i T_i \right) \frac{\partial}{\partial x_j} h(\mathbf{x}) \right] \exp(-A(\boldsymbol{\eta})) \pi(\boldsymbol{\eta}) d\boldsymbol{\eta} \\ &= \frac{1}{m(\mathbf{x})} \frac{\partial}{\partial x_j} \int p_{\boldsymbol{\eta}}(\mathbf{x}) \pi(\boldsymbol{\eta}) d\boldsymbol{\eta} - \frac{1}{m(\mathbf{x}) h(\mathbf{x})} \frac{\partial}{\partial x_j} h(\mathbf{x}) \int p_{\boldsymbol{\eta}}(\mathbf{x}) \pi(\boldsymbol{\eta}) d\boldsymbol{\eta} \\ &= \frac{\partial}{\partial x_j} \log m(\mathbf{x}) - \frac{\partial}{\partial x_j} \log h(\mathbf{x}). \end{aligned}$$

□

Under the sum of square loss, the Bayesian estimator of $\boldsymbol{\eta}$ is given by the posterior means $\mathbb{E}(\eta_i | \mathbf{x})$ for $i = 1, \dots, s$. To see this, note that the value $\delta(\mathbf{x})$ minimizing

$$\mathbb{E}[L(\boldsymbol{\eta}, \delta(\mathbf{x})) | \mathbf{x}] = \mathbb{E}[\mathbb{E}[L(\boldsymbol{\eta}, \delta(\mathbf{x})) | \eta_{[-1]}] | \mathbf{x}]$$

must have $\mathbb{E}(\eta_1 | \mathbf{x})$ as its first component since the squared loss is used. Similar arguments can be applied on

$$\mathbb{E}[\mathbb{E}[L(\boldsymbol{\eta}, \delta(\mathbf{x})) | \eta_{[-1]}] | \mathbf{x}] = \mathbb{E}[\mathbb{E}[\mathbb{E}[L(\boldsymbol{\eta}, \delta(\mathbf{x})) | \eta_{[-2]}] | \eta_{[-1]}] | \mathbf{x}]$$

and so on to reach the conclusion.

Use Theorem 3.2, we know

$$\mathbb{E} \left(\sum_{i=1}^s \eta_i \frac{\partial T_i(\mathbf{x})}{\partial x_j} | \mathbf{x} \right) = \frac{\partial}{\partial x_j} \log m(\mathbf{x}) - \frac{\partial}{\partial x_j} \log h(\mathbf{x}),$$

and since $T_i(\mathbf{x}) = x_i$, we have $\partial T_i / \partial x_j = x_i \mathbf{1}(j = i)$ so

$$\mathbb{E}(\eta_i | \mathbf{x}) = \frac{\partial}{\partial x_j} \log m(\mathbf{x}) - \frac{\partial}{\partial x_j} \log h(\mathbf{x}).$$

3. Example 3.4 in Chapter 4, Lehmann and Casella (1998): **Multiple normal model**. For

$$X_i|\theta_i \sim \mathcal{N}(\theta_i, \sigma^2), \quad i = 1, \dots, p, \text{ independent,}$$

$$\Theta_i \sim \mathcal{N}(\mu, \tau^2), \quad i = 1, \dots, p, \text{ independent,}$$

where σ^2 , τ^2 , and μ are known, $\eta_i = \theta_i/\sigma^2$ and the Bayes estimator of θ_i is

$$\mathbb{E}(\Theta_i|x) = \sigma^2 \mathbb{E}(\eta_i|x) = \sigma^2 \left[\frac{\partial}{\partial x_i} \log m(\mathbf{x}) - \frac{\partial}{\partial x_i} \log h(\mathbf{x}) \right] = \frac{\tau^2}{\sigma^2 + \tau^2} x_i + \frac{\sigma^2}{\sigma^2 + \tau^2} \mu.$$

Example 3.6 in Chapter 4, Lehmann and Casella (1998): **Continuation of Example 3.4**. To evaluate the risk of the Bayes estimator, we also calculate

$$\frac{\partial^2}{\partial x_i^2} \log m(x) = -\frac{1}{\sigma^2 + \tau^2},$$

and hence by Theorem 3.5,

$$R[\boldsymbol{\eta}, \mathbb{E}(\boldsymbol{\eta}|\mathbf{X})] = R[\boldsymbol{\eta}, -\nabla \log h(\mathbf{X})] - \frac{2p}{\sigma^2 + \tau^2} + \sum_i \mathbb{E}_{\boldsymbol{\eta}} \left(\frac{X_i - \mu}{\sigma^2 + \tau^2} \right)^2.$$

The best unbiased estimator of $\eta_i = \theta_i/\sigma^2$ is

$$-\frac{\partial}{\partial X_i} \log h(\mathbf{X}) = \frac{X_i}{\sigma^2}$$

with risk $R[\boldsymbol{\eta}, -\nabla \log h(\mathbf{X})] = p/\sigma^2$.

This problem further discusses Example 3.6 in Chapter 4, Lehmann and Casella (1998).

- (a) Show that if δ is a Bayes estimator of θ , then $\delta' = \delta/\sigma^2$ is a Bayes estimator of η , and hence $R(\theta, \delta) = \sigma^4 R(\eta, \delta')$.

If δ is a Bayes estimator of θ for the loss function $L(\theta, d) = \sum_{i=1}^p (\theta_i - d_i)^2$, then

$$\delta_i(\mathbf{x}) = \mathbb{E}[\theta_i|\mathbf{x}], \quad i = 1, 2, \dots, p.$$

Divide the equalities by σ^2 , we have

$$\frac{\delta_i(\mathbf{x})}{\sigma^2} = \mathbb{E}[\eta_i|\mathbf{x}], \quad i = 1, 2, \dots, p.$$

Hence under the loss function $L(\eta, d) = \sum_{i=1}^p (\eta_i - d_i)^2$, the Bayes estimator of η is given by $\delta' = \delta/\sigma^2$.

Knowing that $\delta' = \delta/\sigma^2$, we have

$$\mathbb{E}[L(\theta, \delta(X))|\theta] = \mathbb{E}\left[\sum_{i=1}^p (\theta_i - \delta_i(X))^2|\theta\right] = \sigma^4 \mathbb{E}\left[\sum_{i=1}^p \left(\eta_i - \frac{\delta_i(X)}{\sigma^2}\right)^2|\theta\right] = \sigma^4 \mathbb{E}[L(\eta, \delta'(X))|\theta],$$

which is equivalent to

$$R(\theta, \delta) = \sigma^4 R(\eta, \delta').$$

- (b) Show that the risk of the Bayes estimator of η is given by

$$\frac{p\tau^4}{\sigma^2(\sigma^2 + \tau^2)^2} + \left(\frac{\sigma^2}{\sigma^2 + \tau^2}\right)^2 \sum_i a_i^2,$$

with $a_i = \eta_i - \mu/\sigma^2$.

Using the results of Example 3.6,

$$\begin{aligned} R[\boldsymbol{\eta}, \mathbb{E}(\boldsymbol{\eta}|\mathbf{X})] &= R[\boldsymbol{\eta}, -\nabla \log h(\mathbf{X})] - \frac{2p}{\sigma^2 + \tau^2} + \sum_i \mathbb{E}_{\boldsymbol{\eta}} \left(\frac{X_i - \mu}{\sigma^2 + \tau^2} \right)^2 \\ &= \frac{p}{\sigma^2} - \frac{2p}{\sigma^2 + \tau^2} + \sum_i \mathbb{E}_{\boldsymbol{\eta}} \left(\frac{X_i - \mu}{\sigma^2 + \tau^2} \right)^2 \end{aligned}$$

$$\begin{aligned}
&= \frac{p(\tau^2 - \sigma^2)}{\sigma^2(\sigma^2 + \tau^2)} + \frac{1}{(\sigma^2 + \tau^2)^2} \sum_i \mathbb{E}_\eta(X_i^2 - 2\mu X_i + \mu^2) \\
&= \frac{p(\tau^4 - \sigma^4)}{\sigma^2(\sigma^2 + \tau^2)^2} + \frac{1}{(\sigma^2 + \tau^2)^2} \sum_i \left[\sigma^4 \left(\eta_i - \frac{\mu}{\sigma^2} \right)^2 + \sigma^2 \right] \\
&= \frac{p(\tau^4 - \sigma^4)}{\sigma^2(\sigma^2 + \tau^2)^2} + \frac{p\sigma^2}{(\sigma^2 + \tau^2)^2} + \frac{\sigma^4}{(\sigma^2 + \tau^2)^2} \sum_i \left(\eta_i - \frac{\mu}{\sigma^2} \right)^2 \\
&= \frac{p\tau^4}{\sigma^2(\sigma^2 + \tau^2)^2} + \left(\frac{\sigma^2}{\sigma^2 + \tau^2} \right)^2 \sum_i a_i^2,
\end{aligned}$$

with $a_i = \eta_i - \mu/\sigma^2$.

(c) If $\sum_i a_i^2 = k$, a fixed constant, then the minimum risk is attained at $\eta_i = \mu/\sigma^2 + \sqrt{k/p}$.

If $\sum_i a_i^2 = k$, a fixed constant, then $R[\eta, \mathbb{E}(\eta|\mathbf{X})]$ is fixed. Hence the minimum risk is attained at $\eta_i = \mu/\sigma^2 + \sqrt{k/p}$ by the usual math trick.

4. If X_1, \dots, X_n are iid from a one-parameter exponential family, the Bayes estimator of the mean, under squared error loss using a conjugate prior, is of the form $a\bar{X} + b$ for constants a and b . Prove the followings:

(a) If $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$, then no matter what the distribution of the X_i 's, the mean squared error is $\mathbb{E}\{(a\bar{X} + b) - \mu\}^2 = a^2 \text{Var}(\bar{X}) + \{(a-1)\mu + b\}^2$.

We have

$$\begin{aligned}
\mathbb{E}\{(a\bar{X} + b) - \mu\}^2 &= \mathbb{E}\{(a\bar{X} + b)^2 - 2\mu(a\bar{X} + b) + \mu^2\} \\
&= \text{Var}(a\bar{X} + b) + [\mathbb{E}(a\bar{X} + b)]^2 - 2\mu \mathbb{E}(a\bar{X} + b) + \mu^2 \\
&= a^2 \text{Var}(\bar{X}) + (a\mu + b)^2 - 2\mu(a\mu + b) + \mu^2 \\
&= a^2 \text{Var}(\bar{X}) + (a^2 - 2a + 1)\mu^2 + 2(a-1)b\mu + b^2 \\
&= a^2 \text{Var}(\bar{X}) + \{(a-1)\mu + b\}^2,
\end{aligned}$$

regardless of the distribution of \mathbf{X} .

(b) If μ is unbounded, then no estimator of the form $a\bar{X} + b$ can have finite squared error for $a \neq 1$.

Since $\mathbb{E}\{(a\bar{X} + b) - \mu\}^2 = a^2 \text{Var}(\bar{X}) + \{(a-1)\mu + b\}^2$, if μ is unbounded, the squared error of $a\bar{X} + b$ will be unbounded for $a \neq 1$, so no estimator of the form $a\bar{X} + b$ can have finite squared error for $a \neq 1$.

(c) Can a conjugate-prior Bayes estimator in an exponential family have finite squared error?

Yes. Consider $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Binomial}(n, p)$ with p unknown and of interest, then the Bayes estimator of p , under squared error loss using a conjugate prior $\text{Beta}(a_0, b_0)$, is of the form $a\bar{X} + b$ for constants a and b and the mean squared error is finite.

5. Suppose that Θ follows a log-normal distribution with known hyperparameters $\mu_0 \in \mathbb{R}$ and $\sigma_0^2 > 0$ and that, given $\Theta = \theta$, (X_1, \dots, X_n) is an iid sample from $\text{Uniform}(0, \theta)$.

(a) What is the posterior distribution of $\log(\Theta)$?

We are given that

$$\eta := \log(\Theta) \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

and

$$\begin{aligned}
f(\mathbf{X}|\eta) &= f(\mathbf{X}|\theta) \\
&= \frac{1}{\theta^n} \mathbf{1}(X_{(0)} \geq 0) \mathbf{1}(X_{(n)} \leq \theta) \\
&= e^{-n \log \theta} \mathbf{1}(\log X_{(n)} \leq \log \theta) \mathbf{1}(X_{(0)} \geq 0) \\
&= e^{-n\eta} \mathbf{1}(\log X_{(n)} \leq \eta) \mathbf{1}(X_{(0)} \geq 0).
\end{aligned}$$

Hence,

$$\begin{aligned} f(\eta|\mathbf{X}) &\propto f(\mathbf{X}|\eta)\pi(\eta) \\ &\propto e^{-n\eta} \mathbf{1}(\log X_{(n)} \leq \eta) \exp\left(-\frac{(\eta - \mu_0)^2}{2\sigma_0^2}\right) \\ &\propto \exp\left(-\frac{(\eta - (\mu_0 - n\sigma_0^2))^2}{2\sigma_0^2}\right) \mathbf{1}(\eta \geq \log X_{(n)}). \end{aligned}$$

The posterior distribution of $\log(\Theta)$ is, therefore, a truncated normal distribution with mean $\mu_0 - n\sigma_0^2$ and variance σ_0^2 and truncated below by $\log X_{(n)}$.

(b) Let δ_τ represent the Bayes estimator of θ under the loss

$$L(\theta, d) = \begin{cases} 0 & \text{if } \frac{1}{\tau} \leq \frac{\theta}{d} \leq \tau \\ 1 & \text{otherwise} \end{cases}$$

for fixed $\tau > 1$. Find a simple, closed-form expression for the limit of δ_τ as $\tau \rightarrow 1$.

The loss function can be equivalently expressed by

$$L(\theta, d) = \begin{cases} 0 & \text{if } |\log \theta - \log d| \leq \log \tau \\ 1 & \text{if } |\log \theta - \log d| > \log \tau \end{cases}.$$

By Corollary 1.2, the Bayes estimator $\delta_\tau(X)$ of θ is given by

$$\delta_\tau(X) = \arg \max_d \Pr(|\log \theta - \log d| \leq \log \tau).$$

We have the following two situations:

- i. If $\log X_{(n)} \leq \mu_0 - n\sigma_0^2 - \log \tau$, then we should choose $\log \delta_\tau(X) = \mu_0 - n\sigma_0^2$;
- ii. If $\log X_{(n)} > \mu_0 - n\sigma_0^2 - \log \tau$, then we should choose $\log \delta_\tau(X) = \log X_{(n)} + \log \tau$.

As $\tau \rightarrow 1$, we obtain that

$$\delta_\tau(X) \rightarrow \delta(X) = \exp(\max(\log X_{(n)}, \mu_0 - n\sigma_0^2)).$$

6. Consider a Bayesian inference setting in which the prior is continuous and the posterior mean $\mathbb{E}(\Theta|X = x)$ is finite for each x . Show that under the loss function

$$L(\theta, a) = \begin{cases} k_1|\theta - a| & \text{if } a \leq \theta \\ k_2|\theta - a| & \text{otherwise} \end{cases}$$

with $k_1, k_2 > 0$ constant and for p an appropriate function of k_1 and k_2 , every p -th quantile of the posterior distribution is a Bayes estimator.

Denote the p -th quantile of the posterior distribution by τ_p , with

$$\Pr(\theta \leq \tau_p|\mathbf{x}) = p.$$

The loss function $L(\theta, a)$ can be rewritten as

$$L(\theta, a) = k_1(\theta - a) \mathbf{1}(\theta \geq a) + k_2(a - \theta) \mathbf{1}(\theta < a),$$

so

$$\mathbb{E}(L(\theta, a)|\mathbf{x}) = \int_a^\infty k_1(\theta - a)dF(\theta|\mathbf{x}) + \int_{-\infty}^a k_2(a - \theta)dF(\theta|\mathbf{x}).$$

To find the Bayes estimator, take derivatives with respect to a :

$$\frac{\partial}{\partial a} \mathbb{E}(L(\theta, a)|\mathbf{x}) = -k_1 \Pr(\theta > a|\mathbf{x}) + k_2 \Pr(\theta \leq a|\mathbf{x}) = (k_1 + k_2) \Pr(\theta \leq a|\mathbf{x}) - k_1.$$

For p -th quantile of the posterior distribution to be a Bayes estimator, we must have

$$\left. \frac{\partial}{\partial a} \mathbb{E}(L(\theta, a)|\mathbf{x}) \right|_{a=\tau_p} = 0,$$

i.e.,

$$\Pr(\theta \leq \tau_p | \mathbf{x}) = \frac{k_1}{k_1 + k_2},$$

hence $p = k_1/(k_1 + k_2)$.