STAT 713 hw 3

Bayesian estimators, MLEs, MoMs, bias and mean squared error

Do problems 7.19, 7.23, 7.50 from CB. In addition:

1. Suppose $X_1, \dots, X_n \overset{\text{ind}}{\sim} f_X(x; \alpha, \beta) = \beta\alpha^\beta x^{-(\beta+1)}\mathbf{1}(x > \alpha)$.

   (a) Give expressions for α and β in terms of the τ_1 and τ_2 quantiles ξ_{τ_1} and ξ_{τ_2}.

   We find that the cdf corresponding to the density $f_X(x; \alpha, \beta)$ is given by $F_X(x; \alpha, \beta) = 1 - (x/\alpha)^{-\beta}$ for $x > \alpha$. Writing $u = 1 - (x/\alpha)^{-\beta}$ and solving for u gives $x = \alpha(1 - u)^{-1/\beta}$. Therefore we may write

   $$\xi_{\tau_1} = \alpha(1 - \tau_1)^{-1/\beta}$$
   $$\xi_{\tau_2} = \alpha(1 - \tau_2)^{-1/\beta}.$$

   Solving the system of equations for α and β gives

   $$\beta = \frac{\log(1 - \tau_1) - \log(1 - \tau_2)}{\log(\xi_{\tau_2}) - \log(\xi_{\tau_1})}$$
   $$\alpha = \xi_{\tau_1} \exp\left[\frac{\log(\xi_{\tau_2}) - \log(\xi_{\tau_1})}{\log(1 - \tau_1) - \log(1 - \tau_2)} \cdot \log(1 - \tau_1)\right].$$

   (b) (Optional) Run a simulation with 10,000 datasets to obtain (an approximation of) the MSE of the quantile estimators of α and β corresponding to your work in part (a) under $\tau_1 = 0.1$ and $\tau_2 = 0.9$ when $\alpha = 1$, $\beta = 2$, and $n = 50$.

   The following code runs the simulation and returns a Monte Carlo estimate of the MSEs.

```
n <- 50
alpha <- 1
beta <- 2
tau1 <- 0.1
tau2 <- 0.9

S <- 10000
beta.hat <- alpha.hat <- numeric(S)
for(s in 1:S){

  U <- runif(n)
  X <- sort(alpha*(1 - U)^(-1/beta))

  xi1 <- X[ceiling(tau1*n)]
  xi2 <- X[ceiling(tau2*n)]
  beta.hat[s] <- (log(1-tau1) - log(1-tau2)) / (log(xi2) - log(xi1))
```

```
  alpha.hat[s] <- xi1*(1 - tau1)^(1/beta.hat[s])

}

mean((beta.hat - mean(beta.hat))^2)
mean((alpha.hat - mean(alpha.hat))^2)
```

The estimator of α had an MSE of 0.000622 and the estimator of β had an MSE of 0.1696.

2. Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} f(x;\theta) = \theta x^{\theta-1}\mathbf{1}(0 < x < 1)$ for $\theta > 0$.

   (a) Find the method of moments estimator of θ.

   We have $m_1 = \int_0^1 x\theta x^{\theta-1}dx = \theta/(\theta+1)$, so $\theta = m_1/(1-m_1)$. The method of moments estimator of θ is therefore $\bar{\theta} = \hat{m}_1/(1-\hat{m}_1)$.

   (b) Use Jensen's inequality to show that this estimator is biased.

   The function $g(z) = z/(1-z)$ on $z \in (0,1)$ is convex, therefore $\mathbb{E}\bar{\theta} = \mathbb{E}g(\hat{m}_1) \leq g(\mathbb{E}\hat{m}_1) = g(m_1) = \theta$. Moreover, since the function g is strictly convex, the inequality is a strict inequality, so $\mathbb{E}\bar{\theta} < \theta$.

3. Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Gamma}(\alpha_0, \beta)$, $\beta > 0$ with α_0 known.

   (a) Find the MLE $\hat{\tau}$ of $\tau = 1/\beta$.

   The log-likelihood function for β is given by

   $$\ell(\beta;\mathbf{X}) = -n\alpha_0 \log\beta - n\log\Gamma(\alpha_0) - (\alpha_0 - 1)\sum_{i=1}^n \log X_i - n\bar{X}_n/\beta,$$

   which is minimized at $\hat{\beta} = \bar{X}_n/\alpha_0$. The MLE for $\tau = \tau(\beta) = 1/\beta$ is therefore $\hat{\tau} = \alpha_0/\bar{X}_n$.

   (b) Find the constant c such that $c\hat{\tau}$ is unbiased for τ.

   We can use mgfs to show that $\bar{X}_n \sim \Gamma(n\alpha_0, \beta/n)$, and further, that

   $$\mathbb{E}\hat{\beta} = \mathbb{E}(\alpha_0/\bar{X}_n) = \frac{n\alpha_0}{n\alpha_0 - 1}\frac{1}{\beta}.$$

   An unbiased estimator for τ is therefore given by

   $$\hat{\tau}_{\text{unbiased}} = \frac{n\alpha_0 - 1}{n\alpha_0}\hat{\beta}.$$

(c) Find the constant c that minimizes the mean squared error of $c\hat{\tau}$.

Use the fact that

$$\mathbb{E}\hat{\tau}^2 = \mathbb{E}(\alpha_0/\bar{X}_n)^2 = \frac{n^2\alpha_0^2}{(n\alpha_0 - 1)(n\alpha_0 - 2)}\frac{1}{\beta^2},$$

we can write

$$\begin{aligned}\text{MSE } c\hat{\tau} &= \mathbb{E}(c\hat{\tau} - \tau)^2 \\ &= \mathbb{E}(c^2\hat{\tau}^2 - 2c\hat{\tau}\tau + \tau^2)^2 \\ &= c^2\frac{n^2\alpha_0^2}{(n\alpha_0 - 1)(n\alpha_0 - 2)}\frac{1}{\beta^2} - 2c\frac{n\alpha_0}{n\alpha_0 - 1}\frac{1}{\beta^2} + \frac{1}{\beta^2}.\end{aligned}$$

Now we have

$$\frac{\partial}{\partial c}\text{MSE } c\hat{\tau} = 2c\frac{n^2\alpha_0^2}{(n\alpha_0 - 1)(n\alpha_0 - 2)}\frac{1}{\beta^2} - 2c\frac{n\alpha_0}{n\alpha_0 - 1}\frac{1}{\beta^2}.$$

Setting the above equal to zero and solving for c gives

$$c = \frac{n\alpha_0 - 2}{n\alpha_0}$$

as the optimal value of c for minimizing the mean squared error of $c\hat{\tau}$.

4. (Optional) Consider the Bayesian hierarchical model

$$\begin{aligned} X_1, \ldots, X_n | \theta &\overset{\text{ind}}{\sim} \text{Normal}(\theta, \sigma^2) \\ \theta &\sim \pi(\theta) = \exp(-|\theta|/\lambda)/(2\lambda), \end{aligned}$$

for some known constants $\lambda > 0$ and $\sigma > 0$. Find the posterior mode of $\theta | X_1, \ldots, X_n$.

To make our calculations simpler, we will use the fact that $\bar{X}_n$ is a sufficient statistic for θ. This gives $\pi(\theta|X_1, \ldots, X_n) = \pi(\theta|\bar{X}_n)$. Since $\bar{X}_n \sim \text{Normal}(\theta, \sigma^2/n)$, we may write

$$\begin{aligned}\pi(\theta|X_1, \ldots, X_n) &\propto \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma/\sqrt{n}}\exp\left[-\frac{(\bar{X}_n - \theta)^2}{2\sigma^2/n}\right]\frac{1}{2\lambda}\exp\left[-\frac{|\theta|}{\lambda}\right] \\ &\propto \exp\left[-\frac{(\bar{X}_n - \theta)^2 + \lambda^{-1}(2\sigma^2/n)|\theta|}{2\sigma^2/n}\right].\end{aligned}$$

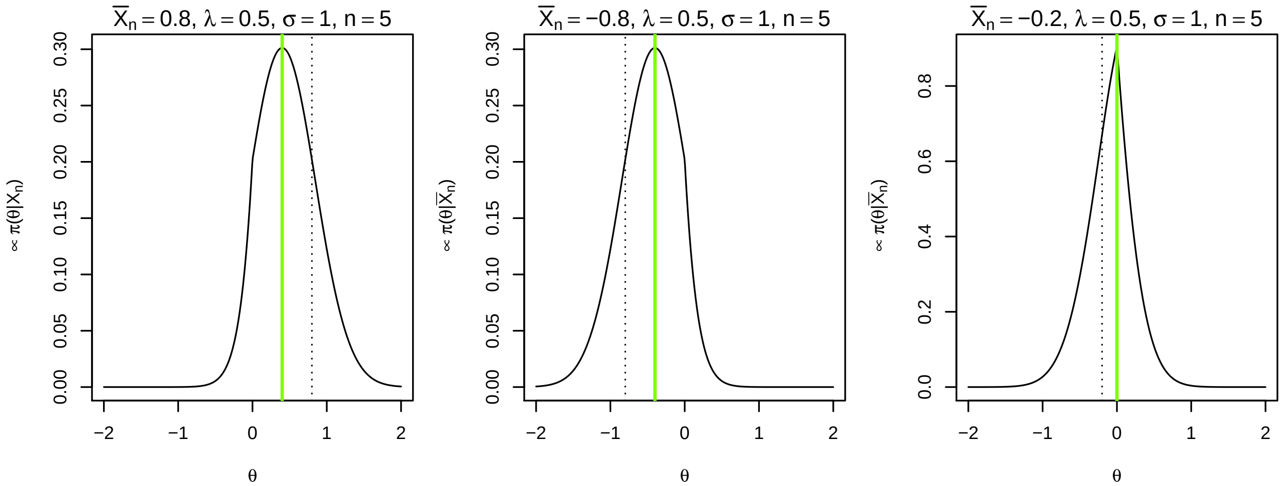
The posterior mode is the value of θ which maximizes the above expression. This is the minimizer of

$$Q(\theta) = (\bar{X}_n - \theta)^2 + 2\frac{\sigma^2}{\lambda n}|\theta|.$$

The value of θ which maximizes $Q(\theta)$ is given by

$$\hat{\theta} = \begin{cases} \bar{X}_n + \frac{1}{\lambda}\frac{\sigma^2}{n}, & \bar{X}_n < -\frac{1}{\lambda}\frac{\sigma^2}{n} \\ 0, & |\bar{X}_n| \leq \frac{1}{\lambda}\frac{\sigma^2}{n} \\ \bar{X}_n - \frac{1}{\lambda}\frac{\sigma^2}{n}, & \bar{X}_n > \frac{1}{\lambda}\frac{\sigma^2}{n}. \end{cases}$$

This is somewhat tricky to work out. One can plot pictures of the function Q as an aid:



The best way is to use subgradients. A *subgradient* of a convex function f at the point x is any value $b \in \mathbb{R}$ such that $f(y) \geq f(x) + b(y - x)$ for all y, and the *subdifferential* ∂f of f at x is the set of such values b. That is

$$\partial f(x) = \{b \in \mathbb{R} : f(y) \geq f(x) + b(y - x) \text{ for all } y \in \mathbb{R}\}.$$

For differentiable convex functions, the subgradient is always a singleton containing the derivative of the function. The subgradient of the absolute value function $|x|$ is defined as

$$\partial|x| \in \begin{cases} \{-1\}, & x < 0 \\ \{1\}, & x > 0 \\ [-1, 1], & x = 0. \end{cases}$$

Taking the derivative of $Q(\theta)$ and using the idea of subgradients, gives

$$\frac{\partial}{\partial\theta}Q(\theta) = -2(\bar{X}_n - \theta) + \frac{2\sigma^2}{\lambda n}\partial|\theta| = -2(\bar{X}_n - \theta) + \frac{2\sigma^2}{\lambda n}s,$$

where $s = 1$ if $\theta > 0$, $s = -1$ if $\theta < 0$, and $|s| \leq 1$ if $\theta = 0$. Setting the above expression equal to zero and going through the cases gives the result.

Problems 7.19, 7.23, and 7.50 from CB.

$\boxed{7.19}$ Let $Y_i = \beta \overset{\text{fixed}}{x_i} + \varepsilon_i$, $i=1,\dots,n$, $\varepsilon_1,\dots,\varepsilon_n \overset{\text{ind}}{\sim} \text{Normal}(0,\sigma^2)$.

(a) We have $Y_i \sim \text{Normal}(x_i\beta, \sigma^2)$ for $i=1,\dots,n$, $Y_1,\dots,Y_n$ independent.

So the likelihood (joint density) is given by

$$\begin{aligned}
L(\beta,\sigma^2;\underset{\sim}{Y}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left[-\frac{(Y_i - x_i\beta)^2}{2\sigma^2}\right] \\
&= (2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - x_i\beta)^2\right] \\
&= (2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n Y_i^2 + \frac{\beta}{\sigma^2}\sum_{i=1}^n x_iY_i - \frac{\beta^2}{2\sigma^2}\sum_{i=1}^n x_i^2\right] \\
&= (2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left[-\frac{\beta^2}{2\sigma^2}\sum_{i=1}^n x_i^2\right]\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n Y_i^2 + \frac{\beta}{\sigma^2}\sum_{i=1}^n x_iY_i\right].
\end{aligned}$$

From here we can see by the factorization theorem that

$$T(\underset{\sim}{Y}) = \left(\sum_{i=1}^n Y_i^2, \sum_{i=1}^n x_iY_i\right)$$

is a sufficient statistic for (β,σ^2).

(b) The log-likelihood is

$$\ell(\beta,\sigma^2;\underset{\sim}{Y}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - x_i\beta)^2.$$

We have

$$\begin{aligned}
\frac{\partial}{\partial\beta}\ell(\beta,\sigma^2;\underset{\sim}{Y}) &= \frac{1}{\sigma^2}\sum_{i=1}^n (Y_i - x_i\beta)x_i \overset{\text{set}}{=} 0 \\
&\iff \sum_{i=1}^n x_iY_i - \sum_{i=1}^n x_i^2\beta = 0 \\
&\iff \beta = \frac{\sum_{i=1}^n x_iY_i}{\sum_{i=1}^n x_i^2}.
\end{aligned}$$

So the MLE of β is $\hat\beta_{\text{mle}} = \dfrac{\sum_{i=1}^n x_iY_i}{\sum_{i=1}^n x_i^2}$.

(c) The moment generating function of $\hat{\beta}_{mle}$ is given by

$$M_{\hat{\beta}_{mle}}(t) = M_{\frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}}(t)$$

$$= \prod_{i=1}^n M_{Y_i}\left(x_i t \Big/ \sum_{j=1}^n x_j^2\right) \qquad M_{Y_i}(s) = e^{x_i \beta s + \frac{\sigma^2 s^2}{2}}$$

$$= \prod_{i=1}^n e^{x_i \beta \left(x_i t / \sum_{j=1}^n x_j^2\right) + \frac{\sigma^2}{2}\left(\frac{x_i t}{\sum_{j=1}^n x_j}\right)^2}$$

$$= \exp\left[\beta t + \frac{\sigma^2 t^2}{2 \sum_{j=1}^n x_j^2}\right],$$

which is the mgf of a Normal$\left(\beta, \sigma^2 \left(\sum_{j=1}^n x_j^2\right)^{-1}\right)$ distribution.

Therefore

$$\hat{\beta}_{mle} \sim \text{Normal}\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right).$$

$\boxed{7.23}$

Note: the χ^2_ν dist has pdf

$$f_W(w) = \frac{1}{\Gamma(\frac{\nu}{2})\, 2^{\nu/2}} w^{\nu/2 - 1} e^{-w/2} \mathbb{1}(w > 0).$$

So if $W \sim \chi^2_{n-1}$, then $V = cW$, $c > 0$, has density

$$f_V(v) = \frac{1}{c} \frac{1}{\Gamma(\frac{\nu}{2})\, 2^{\nu/2}} \left(\frac{v}{c}\right)^{\nu/2 - 1} e^{-\frac{v}{2c}} \mathbb{1}(v > 0) \qquad \left(v = cw \iff w = \frac{1}{c} v, \ \frac{dw}{dv} = \frac{1}{c}\right)$$

Therefore $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ means $S^2 \overset{D}{=} \frac{\sigma^2}{n-1} W$, $W \sim \chi^2_{n-1}$, so

$$f_{S^2}(s^2) = \left(\frac{n-1}{\sigma^2}\right) \frac{1}{\Gamma(\frac{n-2}{2})\, 2^{\frac{n-1}{2}}} \left(\frac{(n-1)}{\sigma^2} s^2\right)^{\frac{n-1}{2} - 1} e^{-\frac{(n-1)s^2}{2\sigma^2}} \mathbb{1}(s^2 > 0).$$

Now consider the hierarchical model

$$S^2 \mid \sigma^2 \sim f_{S^2}(s^2 \mid \sigma^2) = \left(\frac{n-1}{\sigma^2}\right)\frac{1}{\Gamma\left(\frac{n-2}{2}\right) 2^{\frac{n-1}{2}}}\left(\frac{(n-1)s^2}{\sigma^2}\right)^{\frac{n-1}{2}-1} e^{-\frac{(n-1)s^2}{2\sigma^2}} \mathbb{1}(s^2 > 0).$$

$$\sigma^2 \sim \pi(\sigma^2) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(\frac{1}{\sigma^2}\right)^{\alpha+1} e^{-\beta\sigma^2} \mathbb{1}(\sigma^2 > 0).$$

To find the posterior density of $\sigma^2 \mid S^2$, we write

$$\pi(\sigma^2 \mid s^2) \propto f_{S^2}(s^2 \mid \sigma^2)\,\pi(\sigma^2)$$

$$= \left(\frac{n-1}{\sigma^2}\right)\frac{1}{\Gamma\left(\frac{n-2}{2}\right) 2^{\frac{n-1}{2}}}\left(\frac{(n-1)s^2}{\sigma^2}\right)^{\frac{n-1}{2}-1} e^{-\frac{(n-1)s^2}{2\sigma^2}} \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(\frac{1}{\sigma^2}\right)^{\alpha+1} e^{-\frac{1}{\beta\sigma^2}}$$

$$\propto \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\right)^{\frac{n-1}{2}-1}\left(\frac{1}{\sigma^2}\right)^{\alpha+1} e^{-\frac{(n-1)s^2}{2\sigma^2}-\frac{1}{\beta\sigma^2}}$$

$$= \left(\frac{1}{\sigma^2}\right)^{\left(\frac{n-1}{2}+\alpha\right)-1} \exp\left[-\frac{1}{\sigma^2}\Big/\left(\frac{(n-1)s^2}{2}+\frac{1}{\beta}\right)^{-1}\right],$$

which is proportional to the pdf of the $IG\left(\frac{n-1}{2}+\alpha, \left(\frac{(n-1)s^2}{2}+\frac{1}{\beta}\right)^{-1}\right)$.

So

$$\sigma^2 \mid s^2 \sim IG\left(\frac{n-1}{2}+\alpha, \left(\frac{(n-1)s^2}{2}+\frac{1}{\beta}\right)^{-1}\right).$$

The mean of the $IG(\alpha,\beta)$ is given by

$$\int_0^\infty x \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(\frac{1}{x}\right)^{\alpha+1} e^{-\frac{1}{\beta x}}\,dx$$

$$= \frac{\Gamma(\alpha-1)}{\Gamma(\alpha)\beta}\underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha-1)\beta^{\alpha-1}}\left(\frac{1}{x}\right)^{(\alpha-1)+1} e^{-\frac{1}{\beta x}}\,dx}_{=1}$$

$$= \frac{1}{(\alpha-1)\beta}.$$

Therefore we have

$$\mathbb{E}\left[\sigma^2 \mid S^2\right] = \frac{1}{\left(\frac{n-1}{2}+\alpha-1\right)\left(\frac{(n-1)S^2}{2}+\frac{1}{\beta}\right)^{-1}}$$

$$= \frac{\frac{(n-1)S^2}{2}+\frac{1}{\beta}}{\frac{n-1}{2}+\alpha-1}$$

$$= S^2\left(\frac{\frac{n-1}{2}}{\frac{n-1}{2}+\alpha-1}\right) + \frac{1}{(\alpha-1)\beta}\left(\frac{\alpha-1}{\frac{n-1}{2}+\alpha-1}\right),$$

which we may use as a Bayesian estimator of σ^2.

$\boxed{7.50}$ Let $X_1,\dots,X_n \overset{ind}{\sim} \text{Normal}(\theta,\theta^2)$, $\theta>0$. Let $c = \frac{\sqrt{n-1}\,\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}$.
It is given that $\mathbb{E}\, cS = \theta$.

(a) We have

$$\mathbb{E}\left[a\bar{X}_n + (1-a)cS\right] = a\theta + (1-a)\theta = \theta.$$

(b)
$$\text{Var}\left(a\bar{X}_n + (1-a)cS\right) = a^2\,\text{Var}\,\bar{X}_n + (1-a)^2c^2\,\text{Var}\,S$$

$$= a^2\frac{\theta^2}{n} + (1-a)^2c^2\left[\mathbb{E}S^2 - (\mathbb{E}S)^2\right]$$

$$= a^2\frac{\theta^2}{n} + (1-a)^2c^2\left[\mathbb{E}S^2 - \frac{(\mathbb{E}cS)^2}{c^2}\right]$$

$$= a^2\frac{\theta^2}{n} + (1-a)^2c^2\left[\theta^2 - \frac{\theta^2}{c^2}\right]$$

$$= \theta^2\left[\frac{a^2}{n} + (1-a)^2(c^2-1)\right].$$

Now

$$\frac{\partial}{\partial a} \operatorname{Var}\left(a\bar{X}_n + (1-a)cS\right) = \theta^2\left[\frac{2a}{n} - 2(1-a)(c^2-1)\right] \overset{\text{set}}{=} 0$$

$$\iff a - n(1-a)(c^2-1) = 0$$

$$\iff a = n(c^2-1) - na(c^2-1)$$

$$\iff a(1+n(c^2-1)) = n(c^2-1)$$

$$\iff a = \frac{n(c^2-1)}{1+n(c^2-1)}.$$

So the choice $a = \dfrac{n(c^2-1)}{1+n(c^2-1)}$ minimizes the variance.

(c) The statistic $(\bar{X}_n, S_n^2)$ is not complete because

$\mathbb{E}(\bar{X}_n - cS_n) = 0$ but $P(\bar{X}_n - cS_n = 0) \neq 1.$