Practice Problems Solutions

Rebecca C. Steorts

April 4, 2018

1. Let $X_1, \ldots, X_n$ be iid Poisson(θ) variables, where $\theta \in (0, \infty)$. Let $L(\theta, \delta) = (\theta - \delta)^2/\theta$. Assuming the prior,
$$g(\theta) = \frac{\exp\{-\theta\alpha\}\alpha^{\beta}\theta^{\beta-1}}{\Gamma(\beta)} I_{[\theta>0]},$$
where $\alpha > 0$ and $\beta > 0$ are given. Show that the Bayes estimator of θ is given by
$$h(\boldsymbol{X}) = \begin{cases} \dfrac{\sum_i X_i + \beta - 1}{n+\alpha} & \text{if } \sum_i X_i + \beta - 1 > 0 \\ 0 & \text{if } \textit{otherwise}. \end{cases}$$
Solution: Observe that $X|\theta \sim$ Poisson(θ) and $\theta \sim$ Gamma(β, α). By conjugacy, $\theta|X = x \sim$ Gamma($\beta + n\bar{x}, \alpha + n$). The posterior risk for decision a is:
$$\int_{\Theta} \frac{(\theta - a)^2}{\theta} f(\theta|x) d\theta \propto \int_{\Theta} \frac{(\theta-\delta)^2}{\theta} \theta^{\beta+n\bar{x}-1} e^{-\theta(\alpha+n)} d\theta =$$
$$= \int_{\Theta} (\theta - \delta)^2 \theta^{\beta+n\bar{x}-2} e^{-\theta(\alpha+n)} d\theta$$
If $\beta + n\bar{x} - 1 > 0$, $\theta^{\beta+n\bar{x}-2} e^{-\theta(\alpha+n)}$ is proportional to the pdf of a Gamma($\beta + n\bar{x} - 1, \alpha + n$). Hence, last expression is proportional to $E[(\theta - \delta)^2|X]$ when $\theta|X \sim$ Gamma($\beta + n\bar{x} - 1, \alpha + n$). We know this expression is minimized with $\hat{\delta} = E[\theta|X]$. That is, the Bayes Estimator is $\hat{\delta} = \frac{n\bar{x}+\beta-1}{n+\alpha}$.

   If $\beta + n\bar{x} - 1 \le 0$, then for every $\delta \neq 0$, $\int_{\Theta} (\theta - \delta)^2 \theta^{\beta+n\bar{x}-2} e^{-\theta(\alpha+n)} d\theta = \infty$. Similarly, if $\delta = 0$, the posterior risk is proportional to
$$\int_{\Theta} \theta^2 \theta^{\beta+n\bar{x}-2} e^{-\theta(\alpha+n)} d\theta = \int_{\Theta} \theta^{\beta+n\bar{x}+1-1} e^{-\theta(\alpha+n)} d\theta \propto$$
$$\propto \frac{\beta + n\bar{x} + 1}{\alpha + n} < \infty$$
Hence, in this case, the Bayes Estimator is $\hat{\delta} = 0$.

2. Suppose $X \mid p \sim$ Bin(n, p) and that $p \sim$ Beta(a, b).

(a) Show that the marginal distribution of X is the beta-binomial distribution with mass function

$$m(x) = \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(x+a)\Gamma(n+b-x)}{\Gamma(n+a+b)}.$$

(b) Show that the mean and variance of the beta-binomial is given by $EX = \dfrac{na}{a+b}$ and $\mathrm{V}X = n\left(\dfrac{a}{a+b}\right)\left(\dfrac{b}{a+b}\right)\left(\dfrac{a+b+n}{a+b+1}\right)$.

Hint: For part(b): Use the formulas for iterated expectation and iterated variance.

Solution:

(a) Using the definition of $m(x)$,

$$\begin{aligned} m(x) &= \int_\Theta f(x|\theta)\pi(\theta)d\theta \\ &= \int_\Theta \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}\theta^{b-1}d\theta = \\ &= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_\Theta \theta^{a+x-1}(1-\theta)^{b+n-x-1}d\theta = \\ &= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+x)\Gamma(b+n-x)}{\Gamma(a+b+n)} \end{aligned}$$

(b) By the Law of Iterated Expectation

$$E[X] = E[E[X|\theta]] = E[n\theta] = \frac{na}{a+b}$$

By the Law of Iterated Variance

$$\begin{aligned} Var[X] &= Var[E[X|\theta]] + E[Var[X|\theta]] = \\ &= Var[n\theta] + E[n\theta(1-\theta)] = \\ &= n^2Var[\theta] + nE[\theta] - nE[\theta^2] = \\ &= \frac{n^2ab}{(a+b)^2(a+b+1)} + \frac{ma}{a+b} - \frac{n(a+1)}{a+b+1} = \\ &= \frac{n^2ab + na(a+b)(a+b+1) - n(a+1)(a+b)^2}{(a+b)^2(a+b+1)} = \\ &= n\frac{nab + (a+b)(a(a+b+1) - (a+1)(a+b))}{(a+b)^2(a+b+1)} = \\ &= n\frac{nab + ab(a+b)}{(a+b)^2(a+b+1)} = n\frac{a}{a+b}\frac{b}{a+b}\frac{a+b+n}{a+b+1} \end{aligned}$$

3. DasGupta (1994) presents an identity relating the Bayes risk to bias, which illustrates that a small bias can help achieve a small Bayes risk. Let $X \sim f(x|\theta)$ and $\theta \sim \pi(\theta)$. The Bayes estimator under squared error loss is $\hat{\delta} = E(\theta|X)$. Show that the Bayes risk of $\hat{\delta}$ can be written

$$r(\pi, \hat{\delta}) = \int_\Theta \int_\mathcal{X} [\theta - \hat{\delta}(X)]^2 f(x|\theta)\pi(\theta) dx\, d\theta = -\int_\Theta \theta\, b(\theta)\pi(\theta)\, d\theta$$

where $b(\theta) = E[\hat{\delta}|\theta] - \theta$ is the bias of $\hat{\delta}$.

Solution: First, observe that

$$E[\theta E[\theta|X]] = E[E[(\theta E[\theta|X])|X]] = E[E[\theta|X]^2] \tag{1}$$

Now recall that

$$\begin{aligned} r(\pi, E[\theta|X]) &= E[(\theta - E[\theta|X])^2] = E[\theta^2] - 2E[\theta E[\theta|X]] + E[E[\theta|X]^2] =^{(*)} \\ &= E[\theta^2] - E[\theta E[\theta|X]] = E[\theta^2 - \theta E[\theta|X]] = \\ &= E[E[(\theta^2 - \theta E[\theta|X])|\theta]] = -E[\theta E[(E[\theta|X] - \theta)|\theta]] = \\ &= -E[\theta b(\theta)] \end{aligned}$$

The (*) equality follows from Equation **??**.

4. Suppose that

$$\begin{aligned} X|\theta &\sim f(x|\theta) \\ \theta|\lambda &\sim \pi(\theta|\lambda) \\ \lambda &\sim \pi(\lambda). \end{aligned}$$

Using the HB model above,

(a) prove that $E[\theta|x] = E[\ E[\theta|x, \lambda]\ ]$.

(b) prove that $V[\theta|x] = E[\ V[\theta|x, \lambda]\ ] + V[\ E[\theta|x, \lambda]\ ]$.

Remark: when proving (a) and (b) above, you may show this two ways, either by integrals in which say what you are integrating over or you may simply just use expectations (and in this case specifying what you are taking an expectation over).

(a) Let $f(x, \theta, \lambda)$ be the joint density function of the random variables,

$$
\begin{aligned}
E[E[\theta^k|x,\lambda]|x] &= \int\int \theta^k f(\theta|\lambda,x)d\theta f(\lambda|x)d\lambda = \\
&= \int\int \theta^k f(\theta|\lambda,x)f(\lambda|x)d\theta d\lambda = \\
&= \int\int \theta^k f(\theta,\lambda|x)d\theta d\lambda = \\
&= \int\int \theta^k f(\theta,\lambda|x)d\lambda d\theta = \\
&= \int \theta^k f(\theta|x)d\theta = E[\theta^k|x]
\end{aligned}
$$

Taking $k = 1$, the proof is complete.

(b)

$$
\begin{aligned}
&E[V[\theta|x,\lambda]|x] + V[E[\theta|x,\lambda]|x] = \\
&= E[E[\theta^2|x,\lambda]|x] - E[E[\theta|x,\lambda]^2|x] + E[E[\theta|x,\lambda]^2|x] - (E[E[\theta|x,\lambda]|x])^2 =^{(*)} \\
&= E[\theta^2|x] - E[\theta|x]^2 = V(\theta|x)
\end{aligned}
$$

The (*) equality follows from part (a) for $k = 2$ and $k = 1$.

5. Albert and Gupta (1985) investigate theory and application of the hierarchical model

$$
\begin{aligned}
X_i|\theta_i &\overset{ind}{\sim} \text{Bin}(n,\theta_i),\ i = 1,\dots,p \\
\theta_i|\eta &\sim \text{Beta}(k\eta, k(1-\eta)),\ k \text{ known} \\
\eta &\sim \text{Uniform}(0,1).
\end{aligned}
$$

(a) Show that

$$
E(\theta_i|x) = \frac{n}{n+k}\frac{x_i}{n} + \frac{k}{n+k}E(\eta|x)
$$

and

$$
V(\theta_i|x) = \frac{x_i(n+k-x_i) + E(\eta|x)k(n+k-2x_i) - k^2E(\eta^2|x)}{(n+k)^2(n+k+1)} + \frac{k^2V(\eta|x)}{(n+k)^2}.
$$

Note that $E(\eta|x)$ and $V(\eta|x)$ are not expressible in a simple form.

(b) Unconditionally on η, the θ_i's have conditional covariance

$$
\text{Cov}(\theta_i,\theta_j|x) = \left(\frac{k}{n+k}\right)^2 V(\eta|x) \ \text{ for } i \neq j.
$$

Show this.

(c) Ignoring the prior on η, show how to construct an EB estimator of θ_i. Again, this is not expressible in a simple form. That is, simply derive the marginal distribution and then explain using software how you would find an estimator for η. Then give a *simple* construction for the EB estimator.

Solution: Let $E_m[N]$ denote $E[N|M=m]$, $V_m(N) = V(N|M=m)$ and $\text{Cov}_m[N,O] = \text{Cov}[N,O|M=m]$.

(a) Using the model conditional independencies, observe that, for every $i \neq j$, θ_i is conditionally independent of X_j given η. Hence, $\theta_i|\eta, x$ is identically distributed as $\theta_i|\eta, x_i$. Using conjugacy, observe that $\theta_i|\eta, x \sim \text{Beta}(k\eta + x_i, k(1-\eta)+n-x_i)$. Hence, $E_x[\theta_i|\eta] = \frac{k\eta+x_i}{n+k}$. By the Law of Total Expectation, $E_x[\theta_i] = E_x[E_x[\theta_i|\eta]]$,

$$E_x[\theta_i] = E_x\left[\frac{k\eta + x_i}{n+k}\right] = \frac{x_i}{n+k} + \frac{kE_x[\eta]}{n+k} = \frac{n}{n+k}\frac{x_i}{n} + \frac{k}{n+k}E[\eta|x]$$

Also, since $\theta_i|\eta, x \sim \text{Beta}(k\eta + x_i, k(1-\eta)+n-x_i)$, $V_x[\theta_i|\eta] = \frac{(k\eta+x_i)(k(1-\eta)+n-x_i)}{(n+k+1)(n+k)^2}$. By the Law of Total Variance, $V_x[\theta_i] = E_x[V_x[\theta_i|\eta]] + V_x[E_x[\theta_i|\eta]]$ and

$$\begin{aligned}
V_x[\theta_i] &= E_x\left[\frac{(k\eta + x_i)(k(1-\eta)+n-x_i)}{(n+k+1)(n+k)^2}\right] + V_x\left[\frac{k\eta + x_i}{n+k}\right] \\
&= E_x\left[\frac{x_i(n+k-x_i) + \eta k(n+k-2x_i) - \eta^2 k^2}{(n+k+1)(n+k)^2}\right] + \frac{k^2}{(n+k)^2}V_x[\eta] \\
&= \frac{x_i(n+k-x_i) + E[\eta|x]k(n+k-2x_i) - E[\eta^2|x]k^2}{(n+k+1)(n+k)^2} + \frac{k^2}{(n+k)^2}V[\eta|x]
\end{aligned}$$

(b) By the model conditional independencies, (X_i, θ_i) is conditionally independent of $(X_j, \theta_j)_{j \neq i}$ given η. Hence, $\text{Cov}_x[\theta_i, \theta_j|\eta] = 0$ and $E_x[\text{Cov}_x[\theta_i, \theta_j|\eta]] = 0$. By the Law of Total Covariance,

$$\begin{aligned}
\text{Cov}_x[\theta_i, \theta_j] &= E_x[\text{Cov}_x[\theta_i, \theta_j|\eta]] + \text{Cov}_x[E_x[\theta_i|\eta], E_x[\theta_j|\eta]] = \\
&= \text{Cov}_x[E_x[\theta_i|\eta], E_x[\theta_j|\eta]] = \\
&= \text{Cov}_x\left[\frac{k\eta + x_i}{n+k}, \frac{k\eta + x_j}{n+k}\right] = \\
&= \text{Cov}_x\left[\frac{k\eta}{n+k}, \frac{k\eta}{n+k}\right] = \frac{k^2 V(\eta|x)}{(n+k)^2}
\end{aligned}$$

(c) Assume for the moment that we have an estimator for η. Let the estimator for η be $\hat{\eta}$. The EB estimator for θ_i corresponds to δ which minimizes the expected loss when η is known and equal to $\hat{\eta}$, that is minimizes $E[L(\theta_i, \delta_i)|x, \eta = \hat{\eta}]$. Consider that $L(\theta, \delta_i) = (\theta_i - \delta_i)^2$. In this case, we know that the EB estimator is $E[\theta_i|x, \eta = \hat{\eta}]$. Using part a, recall that $E[\theta_i|x_i, \eta = \hat{\eta}] = \frac{k\hat{\eta}+x_i}{n+k}$.
Now we need an estimator for $\hat{\eta}$. Observe that X_i are i.d. given η and

$$E[\bar{X}|\eta] = E[X_1|\eta] = E[E[X_1|\theta_1, \eta]|\eta] = E[n\theta_1|\eta] = n\eta$$

Hence, the MoM (Method of Moments) estimator for η is obtained solving for

$$\bar{X} = n\hat{\eta}$$
$$\hat{\eta} = \bar{X}/n$$

The EB estimator for θ_i using the MoM estimator for η is

$$\hat{\theta}_i \frac{k\bar{X}/n}{n+k} + \frac{x_i}{n+k}$$

Note: Another possible solution is the MLE (Maximum Likelihood Estimator) for η. That is to choose $\hat{\eta}$ which maximizes $f(x|\eta)$. By conditional independence, observe that

$$f(x|\eta) = \prod_{i=1}^{p} f(x_i|\eta)$$

Also observe from the Lecture Notes that $X_i|\eta$ is a Beta-Binomial$(n, k\eta, k(1-\eta))$:

$$f(x|\eta) \propto \prod_{i=1}^{p} \frac{\beta(k\eta + x_i, k(1-\eta) + n - x_i)}{\beta(k\eta, k(1-\eta))}$$

The MLE has no analytic solution. Hence, a numerical approximation would be necessary. A standard way to perform the maximization of the likelihood is through Gradient Ascent. In this problem, another alternative is the EM Algorithm (since we have the latent variables θ_i.