Statistics 582, Problem Set 5 Solutions

Wellner; 2/9/2011

1. Suppose that $X_n \equiv X \sim \text{Multinomial}_k(n, \underline{\theta})$.
(a) Suppose that the prior distribution on θ is given by a Dirichlet distribution, Dirichlet($\underline{\alpha}$):

$$\lambda(\underline{\theta}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\prod_{j=1}^k \Gamma(\alpha_j)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1} 1_{[\underline{\theta}:\ \sum \theta_i = 1]}\,.$$

Verify the computation of the Bayes estimator for squared error loss given in example 4.3.4
(b) What is the posterior distribution for θ? Find the mode of the posterior distribution (along the lines of our computations of the MLE of the multinomial) and compare it with the MLE.
(c) Find a minimax estimator d_M of $\underline{\theta}$.

Solution: (a) If $\underline{\theta} \sim$ Dirichlet($\underline{\alpha}$) then $\theta_j \sim$ Beta($\alpha_j, \sum_{j' \neq j} \alpha_{j'}$), and hence from our computations of the mean of a Beta, $E(\theta_j) = \alpha_j / \sum_{i=1}^k \alpha_i$, and as a vector $E(\underline{\theta}) = \underline{\alpha} / \sum_{i=1}^k \alpha_i$ Since the posterior distribution of $\underline{\theta}$ is Dirichlet($\underline{\alpha} + \underline{X}$), the posterior mean is

$$d_\Lambda(\underline{X}) = E(\underline{\theta} | \underline{X}) = (\underline{\alpha} + \underline{X}) / (\sum_i \alpha_i + n).$$

(b) As noted in A, the posterior density is Dirichlet($\underline{\alpha} + \underline{X}$):

$$\lambda(\underline{\theta} | \underline{X}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k + n)}{\prod_{j=1}^k \Gamma(\alpha_j + X_j)} \theta_1^{\alpha_1 + X_1 - 1} \cdots \theta_k^{\alpha_k + X_k - 1} 1_{[\underline{\theta}:\ \sum \theta_j = 1]}.$$

To find the mode of the posterior, we need to find the value of $\underline{\theta}$ which maximizes $\lambda(\underline{\theta} | \underline{X})$ over the set $\sum_j \theta_j = 1$, or equivalently which maximizes

$$\sum_{j=1}^k (\alpha_j + X_j - 1) \log \theta_j + c(\sum_{j=1}^k \theta_j - 1).$$

Thus we need to solve

$$\frac{\alpha_j + X_j - 1}{\theta_j} + c = 0, \qquad j = 1, \ldots, k. \tag{1}$$

and

$$\sum_{j=1}^{k} \theta_j = 1. \qquad (2)$$

The first equation yields

$$\theta_j^{mode} = \frac{\alpha_j + X_j - 1}{-c}, \qquad j = 1, \ldots, k;$$

substitution of this into (2) yields

$$1 = \sum_{j=1}^{k} \theta_j^{mode} = \frac{1}{-c} \left\{ \sum_{j=1}^{k} \alpha_j + n - k \right\},$$

and hence $-c = \sum_j \alpha_j + n - k$. Thus the mode of the posterior is given by

$$\underline{\theta}^{mode} = \frac{\underline{\alpha} + \underline{X} - \underline{1}}{\sum \alpha_j + n - k}.$$

When $\underline{\alpha} = \underline{1}$ (the vector of all 1's), then the mode of the posterior equals the MLE $\hat{\theta} = \underline{X}/n$. Note that $\underline{\alpha} = \underline{1}$ yields a uniform prior over θ.

(c) As shown in class, if $\underline{X} \sim \text{Mult}_k(n; \underline{\theta})$ and $\underline{\theta} \sim \text{Dirichlet}(\underline{\alpha})$, then the Bayes estimator of $\underline{\theta}$ for squared error loss is $d_\Lambda(\underline{X}) = (\underline{\alpha} + \underline{X})/(\sum \alpha_i + n)$ For $\alpha_1 = \cdots = \alpha_k = \alpha$, this yields the Bayes estimator

$$d_\Lambda(\underline{X}) \;=\; \frac{\alpha \underline{1} + \underline{X}}{k\alpha + n} = \frac{k\alpha}{k\alpha + n} \frac{\underline{1}}{k} + \frac{n}{k\alpha + n} \frac{\underline{X}}{n}.$$

Note that $d_{\Lambda,i}(\underline{X}) = (\alpha + X_i)/(k\alpha + n)$ has

$$Var_{\underline{\theta}}(d_{\Lambda,i}(X) = \frac{n\theta_i(1 - \theta_i)}{(k\alpha + n)^2},$$
$$E_{\underline{\theta}}(d_{\Lambda,i}(X)) = \frac{\alpha + n\theta_i}{k\alpha + n},$$
$$\text{bias}_{\underline{\theta}}(d_{\Lambda,i}(X)) = \frac{\alpha - k\alpha\theta_i}{k\alpha + n}.$$

Thus the risk is

$$\begin{aligned}
R(\underline{\theta}, \underline{d}_\Lambda) &= E_{\underline{\theta}}|\underline{\theta} - \underline{d}_\Lambda(\underline{X})|^2 \\
&= \sum_{i=1}^{k} \left\{ Var_{\underline{\theta}}(d_{\Lambda,i}(\underline{X})) + \text{bias}^2_{\underline{\theta}}(d_{\Lambda,i}) \right\} \\
&= \frac{1}{(k\alpha + n)^2} \sum_{i=1}^{k} \left\{ n\theta_i(1-\theta_i) + (\alpha - k\alpha\theta_i)^2 \right\} \\
&= \frac{1}{(k\alpha + n)^2} \left\{ n - k\alpha^2 + (\alpha^2 k^2 - n) \sum_{i=1}^{k} \theta_i^2 \right\} \qquad \text{since} \quad \sum \theta_i = 1 \\
&= \frac{(1 - 1/k)}{(1+\sqrt{n})^2} \qquad \text{if} \;\; \alpha = \frac{\sqrt{n}}{k}.
\end{aligned}$$

which is constant in $\underline{\theta}$. Hence by corollary 5.6.3

$$\begin{aligned}
d_\Lambda(\underline{X}) &= \frac{\sqrt{n}}{\sqrt{n}+n}\frac{1}{k} + \frac{n}{\sqrt{n}+n}\frac{\underline{X}}{n} \\
&= (1-\lambda_n)\frac{\underline{1}}{k} + \lambda_n \hat{\underline{p}}_n
\end{aligned}$$

is minimax for estimation of $\underline{\theta}$.

2. Find the limit distribution of the minimax estimator d_M in problem 1 (i.e. $\sqrt{n}(d_M(X_n) - p) \to_d$ "something" and find "something"). Is d_M a regular estimator of p?

Solution: Note that $\sqrt{n}(1-\lambda_n) = \lambda_n \to 1$. Hence

$$\begin{aligned}
\sqrt{n}(d_M(\underline{X}_n) - \underline{\theta}) &= \sqrt{n}\{\lambda_n \hat{\underline{p}}_n + (1-\lambda_n)\frac{\underline{1}}{k} - (\lambda_n + 1 - \lambda_n)\underline{\theta}\} \\
&= \lambda_n \sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}) + \sqrt{n}(1-\lambda_n)(\frac{\underline{1}}{k} - \underline{\theta}) \\
&\to_d N_k(0, \Sigma) + \frac{\underline{1}}{k} - \underline{\theta} \\
&= N_k(\frac{\underline{1}}{k} - \underline{\theta}, \Sigma)
\end{aligned}$$

where $\Sigma = \text{diag}(\underline{\theta}) - \underline{\theta}\underline{\theta}^T$.

3. Suppose that $X_1, \ldots, X_n$ are i.i.d. Exponential(θ) (so the X's have density $p_\theta(x) = \theta e^{-\theta x} 1_{(0,\infty)}(x)$. with respect to Lebesgue measure on R, and that $\theta \sim \Gamma(\alpha, \beta)$:

$$\lambda(\theta) = \beta \frac{(\beta\theta)^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta\theta) 1_{[0,\infty)}(\theta).$$

(a) Find the Bayes rule $d_B(\underline{X})$ for estimation of θ with squared error loss $L(\theta, a) = |\theta - a|^2$. Find the Bayes rule $d_{Bw}(\underline{X})$ for estimation of θ with weighted squared error loss $L(\theta, a) = (\theta - a)^2/\theta$. Is the maximum likelihood estimator among either of these families of Bayes estimators?
(b) Are the Bayes estimators d_B and d_{Bw} consistent? What are the limit distributions of d_B and d_{Bw}? Compare them with the maximum likelihood estimator.
(c) Suppose that instead of the Gamma prior distribution, θ has the Pareto(θ_0, α) distribution with density λ given by

$$\lambda(\theta) = (\frac{\alpha}{\theta_0})(\frac{\theta_0}{\theta})^{\alpha+1} 1_{(\theta_0,\infty)}(\theta);$$

here $E(\theta) = \frac{\alpha}{\alpha-1}\theta_0$ where $\alpha > 1$ and $\theta_0 > 0$ are known. What can you say about the Bayes estimator for squared error loss with this prior? For what values of θ_0 is the Bayes rule consistent?

Solution: (a) The posterior distribution is Gamma$(\alpha + n, \beta + \sum X_i)$. Thus the Bayes rule for $L(\theta, a) = (\theta - a)^2$ is

$$d_B(\underline{X}) = \frac{\alpha + n}{\beta + \sum X_i}.$$

For $L(\theta, a) = (\theta - a)^2/\theta$, the Bayes rule is

$$d_{Bw}(\underline{X}) = \frac{E(\theta K(\theta)|\underline{X})}{E(K(\theta)|\underline{X})} = \frac{1}{E(1/\theta|\underline{X})} = \frac{\alpha + n - 1}{\beta + \sum X_i}$$

since, for $\theta \sim$ Gamma(α, β) we have

$$E(1/\theta) = \frac{\beta}{\alpha - 1}$$

if $\alpha > 1$. Thus the MLE $1/\overline{X}_n$ is *not* among either of these families of estimators.
(b) Both d_B and d_{Bw} are consistent and asymptotically equivalent to the MLE $1/\overline{X}_n$:

$$\begin{aligned} \sqrt{n}\left\{d_B(\underline{X}) - 1/\overline{X}_n\right\} &= \sqrt{n}\left\{\frac{1 + n^{-1}\alpha}{\overline{X}_n + n^{-1}\beta} - \frac{1}{\overline{X}_n}\right\} \\ &= n^{-1/2}\frac{\alpha\overline{X}_n - \beta}{\overline{X}_n(\overline{X}_n + n^{-1}\beta)} = O(n^{-1/2})O_p(1) = o_p(1), \end{aligned}$$

and similarly for d_{Bw}. Thus, for $d = d_B$ or $d = d_{Bw}$ we have, since $I(\theta) = \theta^{-2}$,

$$\sqrt{n}(d(\underline{X}) - \theta) = \sqrt{n}(\frac{1}{\overline{X}_n} - \theta) + o_p(1) \to_d N(0, 1/I(\theta)) = N(0, \theta^2).$$

(c) When the prior is Pareto(θ_0, α), the posterior density is of the form

$$\begin{aligned} \lambda(\theta|\underline{X}) &= \frac{\theta^n \exp(-\theta \sum X_i)(\alpha\theta_0^{-1})(\theta_0/\theta)^{\alpha+1} 1_{(\theta_0,\infty)}(\theta)}{\int_{\theta_0}^{\infty} s^n \exp(-s \sum X_i)(\alpha\theta_0^{-1})(\theta_0/s)^{\alpha+1} ds} \\ &= \frac{\theta^{n-\alpha-1} \exp(-\theta \sum X_i) 1_{(\theta_0,\infty)}(\theta)}{\int_{\theta_0}^{\infty} s^{n-\alpha-1} \exp(-s \sum X_i) ds}, \end{aligned}$$

which is concentrated on (θ_0, ∞). Thus the Bayes rule $d_B(\underline{X}) = E(\theta|\underline{X})$ takes values in (θ_0, ∞) a.s.. Similar to the argument in class in the Bernoulli(θ) example, $Z_n = d_B(\underline{X}) = E(\theta|X_1, \ldots, X_n)$ is a martingale and hence $Z_n = d_B(\underline{X}) \rightarrow E(\theta|X_1, X_2, \ldots)$. But $\widehat{\theta} = \overline{X}_n^{-1} \rightarrow_{a.s.} \theta$ for each fixed $\theta \in (0, \infty)$, and hence

$$P_\Lambda(\widehat{\theta}_n \rightarrow \theta) = \int P_\theta(\widehat{\theta}_n \rightarrow \theta) d\Lambda(\theta) = 1\,.$$

Hence $\widehat{\theta}_n \rightarrow \theta$ a.s. P_Λ, and this implies that θ is $\mathcal{F}_\infty \equiv \sigma(X_1, X_2, \ldots)$ measurable. Therefore $E(\theta|X_1, X_2, \ldots) = \theta$ a.s. and $d_B(\underline{X}) \rightarrow \theta$ a.s. P_Λ. This in turn implies that $d_B(\underline{X}) \rightarrow_{a.s.} \theta$ for Λ-a.e. θ. this suggests that d_B might be inconsistent for $\theta \in (0, \theta_0)$, and this is in fact the case since $d_B(\underline{X}) < \theta_0$. When the true $\theta < \theta_0$, it is possible to show that $d_B(\underline{X}) \rightarrow_{a.s.} \theta_0 > \theta$ and that the posterior distributions converge to point mass at θ_0.

4. Let $\Theta = (0, \infty)$, $\mathbf{A} = [0, \infty)$, let X have the discrete distribution

$$p(x, \theta) = \binom{r + x - 1}{x} \theta^x (\theta + 1)^{-(r+x)}, \qquad x = 0, 1, 2, \ldots$$

where r is some known positive integer; this is the negative binomial distribution reparametrized so that $E_\theta X = r\theta$. Suppose that

$$L(\theta, a) = \frac{(\theta - a)^2}{\theta(\theta + 1)}.$$

(a) Show that the usual estimator, $d_0(X) = X/r$ is an equalizer rule; i.e. show that it has a risk function $R(\theta, d_0)$ which is constant in θ.
(b) Show that the usual estimator d_0 is generalized Bayes with respect to Lebesgue measure on $(0, \infty)$ provided $r > 1$. (A generalized Bayes rule is a rule that minimizes the posterior Bayes risk even when starting with an improper prior; see e.g. Ferguson, MS, page 50.) (What happens if $r = 1$?)
(c) Find Bayes decision rules with respect to the prior distributions $\Lambda_{\alpha,\beta}$ with densities

$$\lambda_{\alpha,\beta}(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (\theta + 1)^{-(\alpha+\beta)} 1_{(0,\infty)}(\theta),$$

the distribution of $\theta = Z/(1-Z)$ where $Z \sim$ Beta(α, β).
(d) Show that $d(X) = X/(r+1)$ is minimax. [Note that d_0 is not minimax, hence not admissible.]

Solution: (a) First note that $E_\theta(X) = r\theta$ and $Var_\theta(X) = r\theta(\theta+1)$; this follows from the facts that if X has a negative binomial distribution with mass function

$$p(x;p) = \binom{x+r-1}{x} p^r q^x, \quad x \in \{0, 1, \ldots\},$$

then $EX = rq/p$ and $Var(X) = rq/p^2$ with $q \equiv 1-p$. Thus for the weighted squared error loss $L(\theta, a) = (\theta - a)^2/(\theta(\theta+1))$ the rule $d_0(X) = X/r$ has risk

$$R(\theta, d_0) = \frac{1}{\theta(\theta+1)} Var_\theta(X/r) = \frac{1}{r^2\theta(\theta+1)} r\theta(\theta+1) = \frac{1}{r};$$

since the risk function of the rule d_0 is constant in θ, it is "an equalizer rule".
(b) For $\lambda(\theta) = 1_{(0,\infty)}(\theta)$ (corresponding to Λ Lebesgue measure on $(0, \infty)$, the (generalized) Bayes rule is

$$d_\Lambda(X) = \frac{E\{K(\theta)\theta|X\}}{E\{K(\theta)|X\}} = \frac{E\{(\theta+1)^{-1}|X\}}{E\{\theta^{-1}(\theta+1)^{-1}|X\}}$$

where the posterior density is

$$\lambda(\theta|X) = \frac{\Gamma(X+r)}{\Gamma(X+1)\Gamma(r-1)} \theta^{X+1-1}(\theta+1)^{-(r+X)}.$$

Thus we compute the numerator as

$$\begin{aligned} &E\{(\theta+1)^{-1}|X\} \\ &= \int_0^\infty \theta^{X+1-1}(\theta+1)^{-(r+X+1)} \frac{\Gamma(X+r+1)}{\Gamma(X+1)\Gamma(r)} d\theta \cdot \frac{\Gamma(X+r)}{\Gamma(X+r+1)} \cdot \frac{\Gamma(r)}{\Gamma(r-1)} \\ &= \frac{r-1}{X+r}, \end{aligned}$$

and the denominator is

$$\begin{aligned} &E\{\theta^{-1}(\theta+1)^{-1}|X\} \\ &= \int_0^\infty \theta^{X-1}(\theta+1)^{-(r+X+1)} \frac{\Gamma(X+r+1)}{\Gamma(X)\Gamma(r+1)} d\theta \cdot \frac{\Gamma(X+r)}{\Gamma(X+r+1)} \cdot \frac{\Gamma(X)}{\Gamma(X+1)} \cdot \frac{\Gamma(r+1)}{\Gamma(r-1)} \\ &= \frac{1}{X+r} \cdot \frac{1}{X} \cdot r(r-1). \end{aligned}$$

Putting these together yields $d_\Lambda(X) = X/r = d_0(X)$. Thus d_0 is a "generalized Bayes rule" with respect to the (improper) prior given by Lebesgue measure on

$(0, \infty)$. This argument works when $r > 1$ (because of the factor $\Gamma(r-1)$ in the denominator). When $r = 1$ the corresponding posterior is

$$\lambda(\theta|X) = \frac{\Gamma(X+1)}{\Gamma(X+1)\Gamma(0)}\theta^{X+1-1}(\theta+1)^{-(1+X)} = 0$$

since $\Gamma(0) = \int_0^\infty x^{-1}e^{-x}dx = \infty$.
(c) By straightforward calculation the posterior density of θ for the given prior is

$$\lambda(\theta|X) = \frac{\Gamma(X+\alpha+r+\beta)}{\Gamma(X+\alpha)\Gamma(r+\beta)}\theta^{X+\alpha-1}(\theta+1)^{-(r+X+\alpha+\beta)}1_{(0,\infty)}(\theta).$$

The Bayes rule with respect to the loss function $L(\theta, a) = (\theta - a)^2/[\theta(\theta+1)] \equiv K(\theta)(\theta - a)^2$ is given by

$$d_\Lambda(X) = \frac{E\{K(\theta)\theta|X\}}{E\{K(\theta)|X\}} = \frac{E\{(\theta+1)^{-1}|X\}}{E\{\theta^{-1}(\theta+1)^{-1}|X\}}$$

By straightforward calculation the numerator and denominator are given by

$$E\{K(\theta)\theta|X\} = \frac{r+\beta}{X+\alpha+r+\beta},$$
$$E\{K(\theta|X\} = \frac{(r+\beta+1)(r+\beta)}{(X+\alpha+r+\beta)(X+\alpha-1)}.$$

Thus the Bayes rule with respect to this weighted loss function and prior Λ is

$$d_\Lambda(X) = \frac{X+\alpha-1}{r+\beta+1}.$$

Since $E_\theta d_\Lambda(X) = (r\theta+\alpha-1)/(r+\beta+1)$ and

$$Var_\theta(d_\Lambda(X)) = \frac{r\theta(\theta+1)}{(r+\beta+1)^2},$$

The (ordinary) risk of the rule d_Λ is

$$\begin{aligned} R(\theta, d_\Lambda) &= \frac{\frac{r\theta(\theta+1)}{(r+\beta+1)^2} + \left(\frac{r\theta+\alpha-1}{r+\beta+1} - \theta\right)^2}{\theta(\theta+1)} \\ &= \frac{1}{(r+\beta+1)^2}\left\{r + \frac{[\alpha-1-\theta(\beta+1)]^2}{\theta(\theta+1)}\right\} \\ &= \frac{1}{(r+\beta+1)^2}\left\{r + \frac{(\alpha-1)^2}{\theta(\theta+1)} - \frac{2(\alpha-1)(\beta+1)}{\theta+1} + \frac{\theta(\beta+1)^2}{\theta+1}\right\}. \end{aligned}$$

Thus after calculation of

$$\int_0^\infty \frac{1}{\theta(\theta+1)}\lambda(\theta)d\theta = \frac{\beta(\beta+1)}{\alpha(\alpha+\beta+1)},$$
$$\int_0^\infty \frac{1}{\theta+1}\lambda(\theta)d\theta = \frac{\beta}{\alpha+\beta}, \quad \text{and}$$
$$\int_0^\infty \frac{\theta}{\theta+1}\lambda(\theta)d\theta = \frac{\alpha}{\alpha+\beta},$$

we find the Bayes risk of the Bayes rule d_Λ to be

$$\begin{aligned}\mathcal{R}(\Lambda, d_\Lambda) &= \frac{1}{(r+\beta+1)^2}\left\{r + (\alpha-1)^2\frac{\beta(\beta+1)}{\alpha(\alpha+\beta+1)}\right. \\ &\qquad \left. - 2(\alpha-1)(\beta+1)\frac{\beta}{\alpha+\beta} + (\beta+1)^2\frac{\alpha}{\alpha+\beta}\right\} \\ &\to \frac{1}{(r+1)^2}\{r+1\} = \frac{1}{r+1} \qquad \text{as} \ \ \alpha \to 1, \ \beta \to 0. \qquad (3)\end{aligned}$$

(d) The rule $d(X) = X/(r+1)$ corresponding to the limiting Bayes risk in (3) has risk

$$R(\theta, d) = \frac{1}{(r+1)^2}\left\{r + \frac{\theta}{\theta+1}\right\}$$

with supremum risk

$$\sup_{\theta>0} R(\theta, d) = \frac{1}{r+1}.$$

Thus by theorem 6.2 the rule d is minimax.