Statistics 582, Problem Set 5 Solutions

Wellner; 2/11/2009

1. Let $\mathcal{X} = \{0,1\}$, $\mathcal{A} = \Theta = \{1,2\}$, and assume that the losses are given by $L(1,1) = L(2,2) = 0$, $L(1,2) = a$, $L(2,1) = b$. Suppose that the statistician can observe either X or Y where

$$
\begin{aligned}
p_1(1) &= P_1(X=1) = 2/3, \qquad & p_2(1) &= P_2(X=1) = 1/2,\\
p_1^*(1) &= P_1(Y=1) = 3/4, \qquad & p_2^*(1) &= P_2(Y=1) = 1/2.
\end{aligned}
$$

Let $\underset{\sim}{\lambda} = (\lambda, 1-\lambda)$, $\lambda \in [0,1]$ be the prior distribution over Θ.
(a) Find the Bayes risk when X is observed, and similarly for Y.
(b) In the case $a = b$, $\lambda = 1/2$, would the statistician prefer to observe X or Y?
(c) For general $a \ne b$, $\lambda \in (0,1)$ would the statistician prefer to observe X or Y?

Solution: Let d_i =probability of action 1 given that i is observed.
(i) The Bayes risks for observing X or Y are:

$$\mathcal{R}_X(\lambda, d) = \lambda a\{(1-d_0)\frac{1}{3} + (1-d_1)\frac{2}{3}\} + (1-\lambda)b\{d_0\frac{1}{2} + d_1\frac{1}{2}\},$$

and

$$\mathcal{R}_Y(\lambda, d) = \lambda a\{(1-d_0)\frac{1}{4} + (1-d_1)\frac{3}{4}\} + (1-\lambda)b\{d_0\frac{1}{2} + d_1\frac{1}{2}\}.$$

(ii) When $a = b$ and $\lambda = 1/2$,

$$
\begin{aligned}
\mathcal{R}_X(\lambda, d) &= a\frac{1}{2}\left\{1 + \frac{1}{6}d_0 - \frac{1}{6}d_1\right\},\\
\mathcal{R}_Y(\lambda, d) &= a\frac{1}{2}\left\{1 + \frac{1}{4}d_0 - \frac{1}{4}d_1\right\},
\end{aligned}
$$

and these are both minimized by choosing $d = (0,1) \equiv d_\lambda$. Then $\mathcal{R}_X(\lambda, d_\lambda) = (5/12)a > (3/8)a = \mathcal{R}_Y(\lambda, d_\lambda)$, so we would prefer to observe Y.
(iii) The risks are

$$
\begin{aligned}
R_X(1,d) &= a\left\{(1-d_0)\tfrac{1}{3} + (1-d_1)\tfrac{2}{3}\right\}, \qquad & R_X(2,d) &= b\left\{d_0\tfrac{1}{2} + d_1\tfrac{1}{2}\right\}\\
R_Y(1,d) &= a\left\{(1-d_0)\tfrac{1}{4} + (1-d_1)\tfrac{3}{4}\right\}, \qquad & R_Y(2,d) &= b\left\{d_0\tfrac{1}{2} + d_1\tfrac{1}{2}\right\}.
\end{aligned}
$$

Plotting these for $d_1 = (1,1)$, $d_2 = (1,0)$, $d_3 = (0,1)$, and $d_4 = (0,0)$ yields the following plot of the risk bodies:

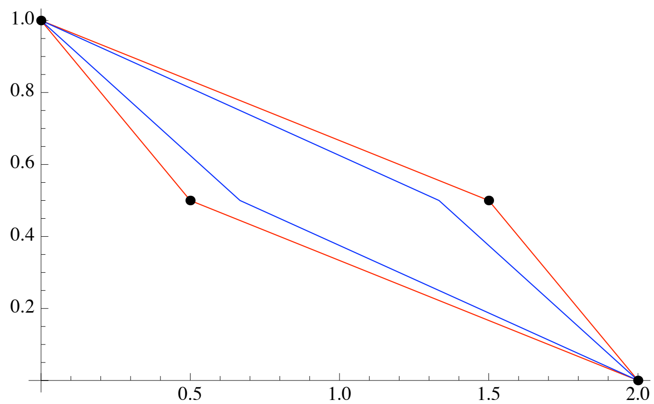

Figure 1: Comparison of the risk bodies $\mathcal{R}_Y$ (red), $\mathcal{R}_X$ (blue), for $a = 2$, $b = 1$

This plot makes it clear that we will always prefer to observe Y. To confirm this, let $r = a/b$, and write

$$\mathcal{R}_X = a\lambda + b\left\{(1-\lambda)\frac{1}{2} - r\lambda\frac{1}{3}\right\} d_0 + b\left\{(1-\lambda)\frac{1}{2} - r\lambda\frac{2}{3}\right\} d_1,$$
$$\mathcal{R}_Y = a\lambda + b\left\{(1-\lambda)\frac{1}{2} - r\lambda\frac{1}{4}\right\} d_0 + b\left\{(1-\lambda)\frac{1}{2} - r\lambda\frac{3}{4}\right\} d_1.$$

First consider $\mathcal{R}_X(\lambda, d)$:
For $0 \leq \lambda \leq (1+4r/3)^{-1}$, both coefficients are > 0, so $d_\lambda = (0,0)$ and $\mathcal{R}_X(\lambda, d_\lambda) = a\lambda$.
For $(1+4r/3)^{-1} \leq \lambda(1+2r/3)^{-1}$, $d_\lambda = (0,1)$ and $\mathcal{R}_X(\lambda, d_\lambda) = a\lambda/3 + b(1-\lambda)/2$.
For $(1+2r/3)^{-1} \leq \lambda \leq 1$, $d_\lambda = (1,1)$ and $\mathcal{R}_X(\lambda, d_\lambda) = b(1-\lambda)$.

Now consider $\mathcal{R}_Y(\lambda, d)$:
For $0 \leq \lambda \leq (1+2r/3)^{-1}$, both coefficients are > 0, so $d_\lambda = (0,0)$ and $\mathcal{R}_Y(\lambda, d_\lambda) = a\lambda$.
For $(1+3r/2)^{-1} \leq \lambda(1+r/2)^{-1}$, $d_\lambda = (0,1)$ and $\mathcal{R}_Y(\lambda, d_\lambda) = a\lambda/4 + b(1-\lambda)/2$.
For $(1+r/2)^{-1} \leq \lambda \leq 1$, $d_\lambda = (1,1)$ and $\mathcal{R}_Y(\lambda, d_\lambda) = b(1-\lambda)$.

These Bayes risks are plotted in the following figure. Thus the Bayes risk is *always* small for Y.

2. Let $\Theta = \mathcal{A} = \{0,1\}$, assume the (hypothesis testing) loss function $L(0,0) = L(1,1) = 0$, $L(0,1) = L(1,0) = 1$. Suppose that we observe the random variable

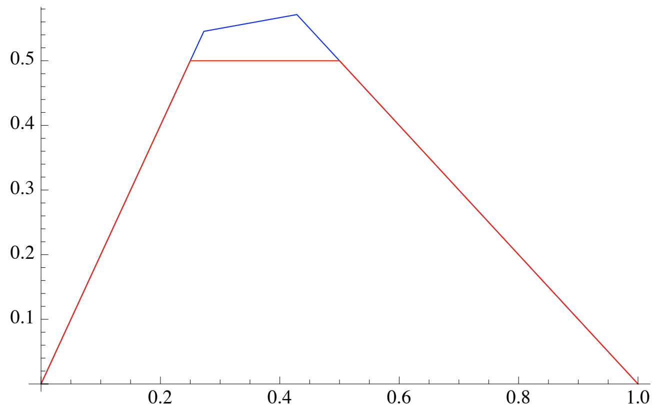

Figure 2: Comparison of the Bayes risks $\mathcal{R}_Y(\Lambda, d_\Lambda)$ (red), $\mathcal{R}_X(\Lambda, d_\Lambda)$ (blue), for $a = 2$, $b = 1$

X with the discrete distribution $P_\theta(X = x) = 2^{-(x+\theta)} 1_{\mathbb{Z}^+ \cap [1-\theta, \infty)}(x)$.
(a) Describe the set of all non-randomized decision rules.
(b) Plot the risk set $\mathcal{R}$ in the plane. Which non-randomized rules are admissible? Why?
(c) Can you find a non-randomized minimax rule?
(d) What decision rules result from a Neyman-Pearson approach?
Hint: every number $z \in [0, 1]$ has a diadic representation of the form $\sum_{x=1}^{\infty} d(x) 2^{-x}$ where $d(x) = 0$ or 1.

Solution: Let $d(x) =$ probability of choosing $\theta 1$ given that $X = x$ is observed. Thus an arbitrary rule $d = (d(0), d(1), d(2), \ldots)$, and the nonrandomized rules have 0 or 1 in each coordinate. Hence the nonrandomized rules consist of all infinite sequences of 0's and 1's. For any rule d

$$\begin{aligned} R(0, d) &= \sum_{x=1}^{\infty} d(x) 2^{-x}, \\ R(1, d) &= \sum_{x=0}^{\infty} (1 - d(x)) 2^{-(x+1)} = \frac{1}{2} \sum_{x=0}^{\infty} (1 - d(x)) 2^{-x} \\ &= 1 - \frac{1}{2} d(0) - \frac{1}{2} R(0, d). \end{aligned}$$

Note that as d ranges over all non-randomized rules, $R(0, d) = \sum_1^{\infty} d(x) 2^{-x}$ ranges over all real numbers between 0 and 1; every $z \in [0, 1]$ has a diadic expansion. Hence the non-randomized rules have risk points on the lines obtained by taking $d(0) = 0$ or $d(0) = 1$; Those with $d(0) = 0$ are inadmissible since the rule with $d(0) = 1$ is better (i.e. has strictly smaller $R(1, d)$); note that $X = 0$ is never observed when $\theta = 0$. When $d(0) = 1$, $R(1, d) = 2^{-1}(1 - R(0, d))$, and hence the minimax risk is $R(0,d) = R(1,d) = 1/3$. This risk is attained by the non-randomized rule $d_M(2x) = 1$, $d_M(2x+1) = 0$, $x \ge 0$. In fact, there are infinitely many randomized minimax rules; e.g. $(1, 2/3, 0, \ldots)$, $(1, 0, 1, 2/3, 0, \ldots)$, $(1, 0, 1, 0, 1, 2/3, 0, \ldots)$, and so forth.

To find a least-favorable prior, for rules d with $d(0) = 1$, write the Bayes risk as

$$\begin{aligned}\mathcal{R}(\lambda, d) &= \lambda R(0,d) + (1-\lambda)R(1,d) = \lambda R(0,d) + (1-\lambda)\frac{1}{2}(1 - R(0,d)) \\ &= \frac{1}{2}(1-\lambda) + \frac{1}{2}(3\lambda - 1)R(0,d).\end{aligned}$$

For $0 \le \lambda < 1/3$, $3\lambda - 1 < 0$, so the Bayes risk is minimized by rules d with $R(0,d) = 1$, e.g. $d = (1, 1, \ldots)$, and the Bayes risk is λ, $0 \le \lambda \le 1/3$.
For $1/3 < \lambda \le 1$, $3\lambda - 1 > 0$, so the Bayes risk is minimized by rules d with $R(0,d) = 0$; e.g. $d = (1, 0, 0, \ldots)$, with Bayes risk $\mathcal{R}(\lambda, d_\lambda) = (1-\lambda)/2$.
When $\lambda = 1/3$, $3\lambda - 1 = 0$, and all rules with $d(0) = 1$ are Bayes wrt λ, with Bayes risk $\mathcal{R}(\lambda, d_\lambda) = 1/3$. Hence the least favorable prior is $\lambda = (1/3, 2/3)$.

Here is another way: Note that for any minimax rule d_M, d_M is Bayes wrt $\lambda = (1/3, 2/3)$, and for this $\underline{\lambda}$ we have

$$\mathcal{R}(\lambda, d_M) = \frac{1}{3} = \sup_{\theta=0,1} R(\theta, d_M);$$

hence by Theorem 6.2, $\underline{\lambda} = (1/3, 2/3)$ is least favorable.

The Neyman - Pearson approach would be to find d so that $R(0,d) = \alpha < 1$, and so that $R(1,d) =$ probability of type 2 error is as small as possible. Thus we are lead to a rule of the form $d = (1, d_1, d_2, \ldots)$ where $\alpha = \sum_1^\infty d_x 2^{-x} = R(0,d)$, and then $R(1,d) = 2^{-1}(1-\alpha)$.

3. 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

$$\sum_{j=1}^k \theta_j = 1. \tag{2}$$

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

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

Thus the risk is

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

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

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

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

4. Find the limit distribution of the minimax estimator d_M in problem 3 (i.e. $\sqrt{n}(d_M(X_n) - p) \rightarrow_d$ "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

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

where $\Sigma = \text{diag}(\underline{\theta}) - \underline{\theta\theta}^T$. To see that $d_M(\underline{X}_n)$ is a regular estimator of θ, let $\theta_n = \theta_0 + tn^{-1/2}$ where $1't = 0$ (so that $1'\theta_n = 1$). Then since $\hat{p}_n$ is a regular estimator of θ with

$$\sqrt{n}(\hat{p}_n - \theta_n) \to_d Z \sim N_k(0, \text{diag}(\theta_0) - \theta_0\theta_0')$$

under P_{θ_n} (which follows from the Liapunov CLT together with the Cram'er-Wold device, or from contiguity theory), it follows that

$$\begin{aligned}
\sqrt{n}(d_M(\underline{X}_n) - \theta_n) &= \sqrt{n}((1-\lambda_n)(1/k) + \lambda_n \hat{p}_n - \theta_n) \\
&= \lambda_n \sqrt{n}(\hat{p}_n - \theta_n) + \sqrt{n}(1-\lambda_n)(1/k - \theta_n) \\
&\to_d 1 \cdot Z + 1 \cdot (1/k - \theta_0) \\
&\sim N_k((1/k - \theta_0), \text{diag}(\theta_0) - \theta_0\theta_0'),
\end{aligned}$$

where we used $\sqrt{n}(1-\lambda_n) = \lambda_n \to 1$ and $\theta_n \to \theta_0$. Since this limiting distribution does not depend on t, $d_M(\underline{X}_n)$ is regular.

5. 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

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

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

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

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

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

(corrected on 2/24/07, thanks to Krisztian Sebestyen), we find the Bayes risk of the Bayes rule d_Λ to be

$$\begin{aligned} \mathcal{R}(\Lambda, d_\Lambda) &= \frac{1}{(r+\beta+1)^2}\left\{r + (\alpha-1)^2\frac{\beta(\beta+1)}{(\alpha-1)(\alpha+\beta)}\right. \\ &\qquad \left. - 2(\alpha-1)(\beta+1)\frac{\beta}{\alpha+\beta} + (\beta+1)^2\frac{\alpha}{\alpha+\beta}\right\} \\ &= \frac{1}{r+\beta+1} \qquad (3) \\ &\to \frac{1}{r+1} \qquad \text{as} \quad \beta \to 0. \qquad (4) \end{aligned}$$

(d) The rule $d(X) = X/(r+1)$ corresponding to the limiting Bayes risk in (4) 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.