Statistics 582, Problem Set 7 Solutions

Wellner; 2/24/2010

1. Continuation of problem 4, problem set 5: Suppose that $X_1, \ldots, X_n$ are i.i.d. Exponential(θ) (so the X's have distribution P_θ and density $p_\theta(x) = \theta e^{-\theta x} 1_{(0,\infty)}(x)$. with respect to Lebesgue measure on $\mathbb{R}$, and that $\theta \sim \Gamma(\alpha, \beta)$:

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

In problem set 5 we found the Bayes rules with respect to squared error loss $L(\theta, a) = (\theta - a)^2$ and weighted squared error loss $L(\theta, a) = (\theta - a)^2/\theta$.

(a) Prove a (conditional) limit theorem for the posterior distributions given $\underline{X}$.
(b) What does theorem 5.8.2 say about the limiting distribution of the Bayes rule for squared error loss (assuming that $X_1, \ldots, X_n$ are i.i.d. $P_{\theta_0} \equiv P$ with $\theta_0 \in (0, \infty)$)?

Solution: (a) There are several possible ways of proceeding here: (i) verify the hypotheses of Theorem 8.1 of the notes; (ii) verify the hypotheses of Theorem 10.1 of van der Vaart's *Asymptotic Statistics*; (iii) give a direct proof in this special case of convergence in distribution; or (iv) give a direct proof in this special case of convergence in total variation distance by showing that the densities converge pointwise followed by Scheffé's lemma. Both (i) and (ii) are made difficult by conditions B2 and B3 (in the case of Theorem 8.1) and by the "separation by tests" condition in van der Vaart's Theorem 10.1. Thus proceed directly as in (iii). First, note that

$$\begin{aligned}
\theta \sim \text{Gamma}(\alpha + n, \beta + \sum X_i) &=_d (\beta + \sum X_i)^{-1} \text{Gamma}(\alpha + n, 1) \\
&=_d (\beta + \sum X_i)^{-1} (Y_0 + \sum_{i=1}^n Y_i)
\end{aligned}$$

where $Y_0 \sim \text{Gamma}(\alpha, 1)$, and $Y_i \sim \text{Gamma}(1, 1) = \text{Exp}(1)$, $i = 1, \ldots, n$ are all independent. Thus conditionally on the X_i's we have, with $Z \sim N(0, 1)$ and with θ_0 the true value of θ,

$$\begin{aligned}
\sqrt{n}(\theta - E(\theta|\underline{X})) &=_d \sqrt{n} \frac{Y_0 + \sum_{i=1}^n Y_i - (\alpha + n)}{\beta + \sum_{i=1}^n X_i} \\
&= \sqrt{n}(\overline{Y}_n - 1) \frac{1}{\overline{X}_n + n^{-1}\beta} + \sqrt{n}(Y_0 - \alpha) \frac{1/n}{\overline{X}_n + n^{-1}\beta} \\
&\to_d Z \frac{1}{\theta_0^{-1}} \sim N(0, \theta_0^2)
\end{aligned}$$

almost surely with respect to the distribution of $X_1, X_2, \ldots$. Note that the posterior mean $E(\theta|\underline{X})$ can be replaced here by either the MLE $1/\overline{X}_n$ or by $T_n = \theta_0 + (nI(\theta_0))^{-1}\sum_{i=1}^n \dot{l}_\theta(X_i) = 2\theta_0 - \theta_0^2\overline{X}_n$ since

$$\sqrt{n}(E(\theta|\underline{X}) - 1/\overline{X}_n) = o_p(1)$$

and similarly with T_n in place of $1/\overline{X}_n$.
To show that the densities of $\sqrt{n}(\theta - E(\theta|\underline{X})$ converge pointwise, first consider the distribution function and density of the unscaled version of the lead term, $\sqrt{n}(\overline{Y}_n - 1)$: since $n\overline{Y}_n = \sum_1^n Y_i \sim \text{Gamma}(n, 1)$,

$$\begin{aligned} F_{\sqrt{n}(\overline{Y}_n - 1)}(z) &= P(\overline{Y}_n \leq 1 + n^{-1/2}z) = P(\sum_1^n Y_i \leq n + \sqrt{n}z) \\ &= \int_0^{n+\sqrt{n}z} \frac{t^{n-1}}{\Gamma(n)} \exp(-t)dt. \end{aligned}$$

Thus, differentiating and then using

$$\Gamma(n) = (n-1)! \sim \sqrt{2\pi(n-1)}\left(\frac{n-1}{e}\right)^{n-1}$$

by Stirling's formula, we find that

$$\begin{aligned} \sqrt{n}(\overline{Y}_n - 1) &= \frac{(n+\sqrt{n}z)^{n-1}}{\Gamma(n)} \exp(-(n+\sqrt{n}z))\sqrt{n} \\ &\sim \frac{(n+\sqrt{n}z)^{n-1}}{\sqrt{2\pi(n-1)}\left(\frac{n-1}{e}\right)^{n-1}} \cdot \exp(-(n+\sqrt{n}z)) \cdot \sqrt{n} \\ &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{n-1}} \cdot \left(1 + \frac{1+\sqrt{n}z}{n-1}\right)^{n-1} \cdot \exp(-(1+\sqrt{n}z)) \\ &\to \frac{1}{\sqrt{2\pi}} \exp(-z^2/2). \end{aligned}$$

Here the convergence follows by letting $a_n \equiv 1+\sqrt{n+1}z$ and noting that

$$\begin{aligned}
&\left(1+\frac{1+\sqrt{n+1}z}{n}\right)^n \cdot \exp(-(1+\sqrt{n+1}z)) \\
&= \left(1+\frac{a_n}{n}\right)^n \cdot \exp(-a_n) \\
&= \left\{\left(1+\frac{a_n}{n}\right) \cdot \exp(-a_n/n)\right\}^n \\
&\approx \left\{\left(1+\frac{a_n}{n}\right)\left(1-\frac{a_n}{n}+\frac{1}{2}\frac{a_n^2}{n}+O(n^{-3/2})\right)\right\}^n \\
&= \left\{1-\frac{1}{2}\frac{a_n^2}{n^2}+O(n^{-3/2})\right\}^n \\
&= \left\{1-\frac{1}{2}\frac{(1+\sqrt{n+1}z)^2/n}{n}\right\}^n \\
&\to \exp(-z^2/2).
\end{aligned}$$

Since $\overline{X}_n + n^{-1}\beta \to_{a.s.} \theta_0^{-1}$ and $\sqrt{n}(Y_0-\alpha)\frac{1/n}{\overline{X}_n+n^{-1}\beta} \to_{a.s.} 0$, it follows (via the convolution formula) that the density of $\sqrt{n}(\theta - E(\theta|\underline{X})$ converges pointwise to $\phi(z/\theta_0)/\theta_0$, the density of $N(0,\theta_0^2)$.

(b) In the present case Theorem 5.8.2 says that

$$\sqrt{n}(E(\theta|\underline{X})-\theta_0) \to_d N(0,1/I(\theta_0)) = N(0,\theta_0^2)$$

since $I(\theta_0)=1/\theta_0^2$. This also follows from a direct argument since

$$\begin{aligned}
\sqrt{n}(E(\theta|\underline{X})-\theta_0) &= \sqrt{n}\left(\frac{1+\alpha/n}{\overline{X}_n+\beta/n}-\theta_0\right) \\
&= \sqrt{n}(\alpha/n+1-\theta_0(\beta/n+\overline{X}_n))/(\beta/n+\overline{X}_n) \\
&= \left\{-\theta_0\sqrt{n}(\overline{X}_n-1/\theta_0)+n^{-1/2}(\alpha-\theta_0\beta)\right\}/(\beta_n+\overline{X}_n) \\
&\to_d \theta_0^2 N(0,\theta_0^{-2}) = N(0,\theta_0^2).
\end{aligned}$$

2. 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. a decision rule such that $R(\theta, d)$ is constant in θ).
(b) Show that the usual estimator d_0 is generalized Bayes with respect to Lebesgue measure on $(0, \infty)$ provided $r > 1$. (What happens if $r = 1$?) (An estimator is called a "generalized Bayes rule" if the posterior distribution (and any conditional expectations needed to define the Bayes rule) is (are) well defined, even though the prior distribution is not a "proper" prior with total mass 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

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

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

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

(d) The rule $d(X) = X/(r+1)$ corresponding to the limiting Bayes risk in (2) 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.

3. Specialize the decision rule in Theorem 5.2 of the course notes to the case when P_i is the normal distribution $N_d(\mu_i, I)$, $i = 1, \ldots, k$ where $\mu_1, \ldots, \mu_k$ are distinct vectors in $\mathbb{R}^d$, $\mu_i \neq \mu_j$ for $i \neq j$. What happens if we replace I by Σ?

Solution: When $P_i = N_d(\mu_i, I)$, the inequality $\lambda_i p_i(x) > \lambda_j p_j(x)$ can be written as

$$\lambda_i \exp(-\frac{1}{2}(x-\mu_i)^T(x-\mu_i)) > \lambda_j \exp(-\frac{1}{2}(x-\mu_j)^T(x-\mu_j)),$$

or, equivalently, assuming that $p_i \neq 0$ and $p_j \neq 0$,

$$(\mu_i - \mu_j)^T x > \frac{1}{2}(\mu_i^T\mu_i - \mu_j^T\mu_j) + \log(p_j/p_i).$$

When $\mu_i \neq \mu_j$, this set of x's corresponds to a half-space bounded by a hyperplane orthogonal to $\mu_i - \mu_j$. Moreover, if $p_i = p_j$, this hyperplane is the bisector of the line segment from μ_i to μ_j. When we consider all the $k-1$ mean vectors μ_j with $j \neq i$, it becomes clear that the set of x's for which $d(i|x) = 1$ is the intersection of $k-1$ half spaces, and this yields a convex polyhedron with at most $k-1$ faces. When $\lambda_i = 1/k$ for $i = 1, \ldots, k$, then we classify X as belonging to the (normal) distribution with mean μ_i that is closest to X. When the identity covariance matrix I is replaced by an arbitrary nonsingular covariance matrix Σ, this remains true with ordinary Euclidean distance replaced by $d_\Sigma^2(x,\mu) \equiv (x-\mu_i)^T\Sigma^{-1}(x-\mu_i)$.

4. 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

$$\frac{\alpha_j+X_j-1}{\theta_j} + c = 0, \qquad j=1,\ldots,k. \tag{3}$$

and

$$\sum_{j=1}^k \theta_j = 1. \tag{4}$$

The first equation yields

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

substitution of this into (4) 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 $\hat{\underline{\theta}} = \underline{X}/n$. 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

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

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 $\underline{\theta}$.

5. Find the limit distribution of the minimax estimator d_M in problem 4 (i.e. $\sqrt{n}(d_M(X_n) - \theta) \rightarrow_d$ "something" and find "something"). Is d_M a regular estimator of θ?

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

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

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