1. Let f and g be two probability density functions (pdf) on $\mathbb{R}$ with respect to a σ-finite measure μ. Let $D(f\|g) = \int f \log \frac{f}{g} d\mu$ denote the Kullback-Leibler divergence between f and g.

(i) Show $D(f\|g) \geq 0$

(ii) Let g_i, $1 \leq i \leq k$ be k pdf's with respect to μ. Show that

$$D(f\|\frac{1}{k}\sum_{i=1}^{k} g_i) \leq \log k + \min_{1\leq i\leq k} D(f\|g_i).$$

(Hint: Use the monotone property of the log function.)

2. Let $X_1, ..., X_n$ be independent and identically distributed real-valued observations from a population with pdf $f_\theta(x)$ with respect to a σ-finite measure μ. Here θ is an unknown parameter in the parameter space Θ. Let d be a metric on Θ. For estimating θ, let the action space be Θ and consider the loss function $L(t;\theta) = d(t,\theta)$ defined on $\Theta \times \Theta$. Let

$$R_n = \inf_{\widehat{\theta}} \sup_{\theta\in\Theta} E_\theta d(\widehat{\theta}, \theta)$$

be the minimax risk for estimating θ under the loss function $L(t;\theta)$, where the infimum is taken over all estimators. We are interested in deriving an appropriate lower bound for R_n.

Let $\Theta_0 = \{\theta_1, ..., \theta_N\}$ be a finite subset of Θ of size N. Assume that for any two distinct values θ and θ' in Θ_0, $d(\theta', \theta) > \epsilon_n$ for a positive constant ϵ_n (Note that n continues to denote the sample size). For any given estimator $\widehat{\theta}$ of θ based on $X_1, ..., X_n$, let

$$\widetilde{\theta} = \arg\min_{\theta\in\Theta_0} d(\widehat{\theta}, \theta),$$

i.e., $\widetilde{\theta}$ minimizes $d(\widehat{\theta}, \theta)$ over $\theta \in \Theta_0$ (please ignore the issue of possible ties in minimization here). Then $\widetilde{\theta}$ is an estimator of θ that takes values in Θ_0.

(i) Show that for $\theta \in \Theta_0$,

$$d(\widehat{\theta}, \theta) \geq \frac{1}{2} d(\widetilde{\theta}, \theta) I_{\{\theta\neq\widetilde{\theta}\}} \geq \frac{\epsilon_n}{2} I_{\{\theta\neq\widetilde{\theta}\}},$$

where $I_{\{\theta\neq\widetilde{\theta}\}}$ equals 1 if $\theta \neq \widetilde{\theta}$ and equals 0 otherwise.

(Hint: Use the triangle inequality.)

(ii) Justify the following inequalities:

$$\begin{aligned}
R_n &\geq \inf_{\widehat{\theta}} \sup_{\theta\in\Theta_0} E_\theta d(\widehat{\theta}, \theta) \\
&\geq \frac{\epsilon_n}{2} \inf_{\widehat{\theta}} \max_{\theta\in\Theta_0} P_\theta(\theta \neq \widetilde{\theta}) \\
&\geq \frac{\epsilon_n}{2} \inf_{\widehat{\theta}} \left(\frac{1}{N} \sum_{i=1}^{N} P_{\theta_i}(\theta_i \neq \widetilde{\theta}) \right).
\end{aligned}$$

It can be shown that

$$\inf_{\widehat{\theta}} \left(\frac{1}{N} \sum_{i=1}^{N} P_{\theta_i}(\theta_i \neq \widetilde{\theta}) \right) \geq 1 - \frac{\frac{1}{N}\sum_{i=1}^{N} D(f_{\theta_i}^n \| q^n) + \log 2}{\log N},$$

where f_θ^n is defined by $f_\theta^n(x_1, x_2, ..., x_n) = \Pi_{j=1}^n f_\theta(x_j)$ and where $q^n(x_1, x_2, ..., x_n)$ is any pdf on $\mathbb{R}^n$. Let $\Theta_1 = \{\theta^1, \theta^2, ..., \theta^M\}$ be a finite subset of Θ of size M with the property that for any $\theta \in \Theta$, there exists $\theta' \in \Theta_1$ such that $D(f_\theta \| f_{\theta'}) \leq \eta_n$ for some positive constant η_n. Now take q^n to be $q^n(x_1, x_2, ..., x_n) = \frac{1}{M}\sum_{j=1}^{M} \Pi_{i=1}^n f_{\theta^j}(x_i)$.

(iii) Show that for any $\theta \in \Theta$,

$$\begin{aligned} D(f_\theta^n \| q^n) &\leq \log M + \min_{1 \leq j \leq M} D(f_\theta^n \| f_{\theta^j}^n) \\ &= \log M + n \min_{1 \leq j \leq M} D(f_\theta \| f_{\theta^j}) \\ &\leq \log M + n\eta_n. \end{aligned}$$

(iv) Deduce that

$$R_n \geq \frac{\epsilon_n}{2}\left(1 - \frac{\log M + n\eta_n + \log 2}{\log N}\right).$$

Solution

1\. (i) $D(f\|g) = E_f \log \frac{f(X)}{g(X)} = -E_f \log \frac{g(X)}{f(X)}$

$-\log x$ is a convex function $\geq -\log\left(E_f \frac{g(X)}{f(X)}\right) = -\log 1 = 0$

(ii) $D(f \| \frac{1}{k}\sum_{i=1}^{k} g_i) = \int f \log \frac{f}{\frac{1}{k}\sum_{i=1}^{k} g_i}\, d\mu$

$\leq \int f \log \frac{f}{\frac{1}{k} g_i}\, d\mu$ (since $\log \frac{1}{x}$ is $\downarrow$)

$= \log k + D(f\|g_i)$

Since this holds for each i, the assertion follows

2\. (i). For $\theta_0 \in \Theta_0$, if $\tilde{\theta} \neq \theta_0$, then

① .... $d(\hat{\theta}, \theta_0) \geq d(\hat{\theta}, \tilde{\theta})$ by definition of $\tilde{\theta}$.

By triangle inequality,

② .... $d(\hat{\theta}, \theta_0) + d(\hat{\theta}, \tilde{\theta}) \geq d(\tilde{\theta}, \theta_0)$.

Thus when $\tilde{\theta} \neq \theta_0$, $d(\hat{\theta}, \theta_0) \underset{\text{by ① and ②}}{\geq} \frac{d(\tilde{\theta}, \theta_0)}{2} \geq \frac{\varepsilon}{2}$ assumption on Θ_0.

When $\tilde{\theta} = \theta_0$, the inequalities are obvious.

(ii). The first inequality is clear since

$\sup_{\theta \in \Theta_0} E_\theta\, d(\hat{\theta}, \theta) \leq \sup_{\theta \in \Theta} E_\theta\, d(\hat{\theta}, \theta)$ $(\Theta_0 \subset \Theta)$

The second inequality follows ~~from~~ (i) with

$$E_\theta I\{\theta \neq \tilde{\theta}\} = P_\theta(\theta \neq \tilde{\theta}).$$

The third inequality follows from the fact that maximum is no smaller than the average.

(iii). The first inequality follows from 1. (ii).

The ~~second~~ identity follows from the observation

$$D(f_\theta^n \| f_{\theta^i}^n) = \int f_\theta(x_1) \cdots f_\theta(x_n) \log \frac{f_\theta(x_1) \cdots f_\theta(x_n)}{f_{\theta^i}(x_1) \cdots f_{\theta^i}(x_n)} \mu(dx_1) \cdots \mu(d$$

$$= \sum_{i=1}^n \int f_\theta(x_1) \cdots f_\theta(x_n) \log \frac{f_\theta(x_i)}{f_{\theta^i}(x_i)} \mu(dx_1) \cdots \mu(dx_n)$$

$$= \sum_{i=1}^n \int f_\theta(x_i) \log \frac{f_\theta(x_i)}{f_{\theta^i}(x_i)} \mu(dx_i)$$

$$\equiv n\, D(f_\theta \| f_{\theta^i}).$$

The second inequality then follows from the property of Θ_i.

(iv). From (iii), $\max_{\theta \in \Theta_0} D(f_\theta^n \| g^n) \leq \log M + n\eta_n$.

It follows $\frac{1}{N} \sum_{i=1}^N D(f_{\theta_i}^n \| g^n) \leq \log M + n\eta_n$

(average is no bigger than maximum).

The assertion then follows.

(With appropriate choices of ε_n and η_n, the above inequality can yield useful lower bounds on the minimax risk R_n.)

Let $X_1, X_2, \ldots$ be iid random variables with cumulative distribution function F where

$$\lim_{x \to -\infty} F(x)/[|x|^{-\alpha}(\log|x|)^{-2}] = (1-\beta)/2$$

and

$$\lim_{x \to \infty} [1 - F(x)]/[|x|^{-\alpha}(\log|x|)^{-2}] = (1+\beta)/2$$

for some $-1 < \beta < 1$ and $1 \leq \alpha \leq 2$. Also, let $\bar{X}_n = n^{-1}(X_1 + \ldots + X_n), n \geq 1$ and let $\mu = EX_1$.

1. Show that $E|X_1|^\gamma < \infty$ for all $0 < \gamma \leq \alpha$.

   [Hint: $E|X_1|^\gamma = \gamma \int_0^\infty t^{\gamma-1} P(|X_1| > t)dt$.]

2. Suppose $\alpha = 2$. Then, $E|X_1|^2 < \infty$. Show that

   $$\sqrt{n}(\bar{X}_n - \mu)/s_n \to^d N(0,1)$$

   where $s_n^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, n \geq 2$ and $\to^d$ denotes convergence in distribution.

3. Next, suppose that $1 \leq \alpha < 2$. In this case, $E|X_1| < \infty$. but $EX_1^2 = +\infty$.

   (a) Show that

   $$n^{-1} \sum_{i=1}^n X_i^2 \to \infty \qquad a.s.$$

   [Hint: Apply the SLLN to $Y_i \equiv \min\{X_i^2, M\}, i \geq 1$ for $M > 0$ and let $M \to \infty$.]

   (b) Show that

   $$s_n^2 \equiv n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \to \infty \qquad a.s.$$

3.(c) It was shown in class that in this case,

$$n(\bar{X}_n - \mu)/[n^{1/\alpha}(\log n)^{-2/\alpha}] \to^d Z_\alpha \tag{1}$$

where Z_α is a random variable having the stable distribution of order α. In view of 3(b), it is not possible to use a function of s_n^2 to Studentize $\bar{X}_n$ in the infinite variance case. Here we develop an alternative method of finding a scaling sequence that does not involve the parameter α. *In the following, suppose that n is an even integer, say $n = 2k$. Write*

$$M_n = \max\{|X_i| : 1 \leq i \leq k\}.$$

Also, for simplicity, suppose that

$$P(|X_1| > x) = x^{-\alpha}(\log x)^{-2}$$

for all $x > 3$.

i. Show that for any $t > 0$,

$$\lim_{n\to\infty} P(M_n \leq k^{1/\alpha}(\log k)^{-2/\alpha} t) = \exp(-t^{-\alpha} \cdot \alpha^2).$$

That is, $M_n/[k^{1/\alpha}(\log k)^{-2/\alpha}] \to^d W$ where the cdf of W is

$$F_W(t) = \begin{cases} 0 & t \leq 0 \\ \exp(-t^{-\alpha}\alpha^2) & t > 0. \end{cases}$$

ii. Let $\bar{X}_{2,n} = (X_{k+1} + \ldots + X_n)/k$. Now, using (1) and the iid-property of $X_1, \ldots, X_n$, show that

$$k(\bar{X}_{2,n} - \mu)/M_n \to^d Z_\alpha/W.$$

[Hint: If $U_n \to^d U$ and $V_n \to^d V$ and U_n and V_n are independent for each $n \geq 1$, then $(U_n, V_n) \to^d (U, V)$, with U and V independent.]

iii. Assuming that all quantiles of the variable Z_α/W are available to you (which may be approximately found by applying the bootstrap method), indicate how you would construct a two-sided 90% confidence interval for μ based on a sample of size $n = 100$.

Ph.D. Prelim

SOLUTION / Theory - II

Spring 2001

1. By the hint,

$$E|X_1|^\gamma = \gamma \int_0^\infty t^{\gamma-1} \, P(|X_1| > t)\, dt$$

$$\leq \gamma \cdot \left[\int_0^M t^{\gamma-1}\, dt + \int_M^\infty t^{\gamma-1} P(|X_1|>t)\, dt \right]$$

where $M > 1$ is such that for $x > M$,

$$F(-x) + 1 - F(x) \leq 2|x|^{-\alpha} (\log|x|)^{-2}.$$

$$\leq \left[M^\gamma + 2 \int_M^\infty t^{\gamma-\alpha-1} (\log t)^{-2}\, dt \right]$$

$$\leq \left[M^\gamma + 2 \int_M^\infty t^{-1} (\log t)^{-2}\, dt \right] < \infty.$$

2. By the CLT, $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$

and by the SLLN, $S_n^2/\sigma^2 \longrightarrow_p 1$.

$$\Longrightarrow \quad \frac{S_n}{\sigma} \longrightarrow_p 1$$

Hence, the result follows from Slutsky's Theorem

3 (a) Fix $M > 0$. Then, $Y_i = \min\{X_i^2, M\} \in (0, M)$

is bounded by M and hence, $E|Y_1| < \infty$.

By the SLLN, ~~for any~~

$$\limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} X_i^2$$

$$\geq \limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} Y_i = M \quad \text{a.s.}$$

Since $M > 0$ is arbitrary, this shows that

$$n^{-1} \sum_{i=1}^{n} X_i^2 \longrightarrow \infty. \quad \text{a.s. (P)}$$

3 (b). Again, by the SLLN, $E|X_1| < \infty \Rightarrow$

$\bar{X}_n \longrightarrow EX_1$ a.s. ~~(P)~~. Hence,

$$S_n^2 = n^{-1} \sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2 \longrightarrow \infty \quad \text{as } n \to \infty, \text{ a.s.}$$

3(c)(i) Using the independence of the X_i's, for any $x > 0$,

$$P(m_n \leq x) = [P(|X_1| \leq x)]^K = \cancel{(F(x))}^K$$

Since $K^{1/\alpha} (\log K)^{-2/\alpha} \uparrow \infty$ as $n \to \infty$, it follows that using the tail-conditions on $F(\cdot)$, for any $t > 0$,

$$\lim_{n \to \infty} P\left(m_n \leq K^{1/\alpha} (\log K)^{-1/\alpha} \cdot t \right)$$

$$= \lim_{n \to \infty} \left[P\left(|X_1| \leq K^{1/\alpha} (\log K)^{-1/\alpha} t \right) \right]^K$$

$$\cancel{= \lim_{n \to \infty} \left[1 - \left(1 - F\left(K^{1/\alpha} (\log K)^{-1/\alpha} t \right) \right) \right]^K}$$

$$= \lim_{n \to \infty} \left[1 - P\left(|X_1| > K^{1/\alpha} (\log K)^{-1/\alpha} t \right) \right]^K$$

$$= \lim_{n \to \infty} \left[1 - \left[K^{1/\alpha} (\log K)^{-2/\alpha} t \right]^{-\alpha} \left(\log \left[K^{1/\alpha} (\log K)^{-2/\alpha} t \right] \right)^{-2} \right]^K$$

$$= \lim_{k\to\infty} \left[1 - k^{-1} \left(t^{-\alpha} \cdot \frac{(\log k)^2}{\left[\frac{1}{\alpha}\log k - \frac{2}{\alpha}\log\log k + \log t\right]^2} \right) \right]^k$$

$$= \exp(- t^{-\alpha} \cdot \alpha^2)$$

C(ii) Since $X_1, X_2, \ldots$ are ~~iid~~ ind. and, identically distributed

$$\bar{X}_{2,n} \stackrel{d}{=} \bar{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i$$

Hence, by (1),

$$Z_n \equiv \frac{k(\bar{X}_{2n} - \mu)}{k^{1/\alpha}(\log k)^{-2/\alpha}} \xrightarrow{d} Z_\alpha \longrightarrow (*)$$

Also, by part 3(c)(i),

$$W_n \equiv \frac{M_n}{k^{1/\alpha}(\log k)^{-2/\alpha}} \xrightarrow{d} W \longrightarrow (**).$$

By the independence of $\{X_1, \ldots, X_k\}$ and $\{X_{k+1}, \ldots X_n\}$, it follows that Z_n and W_n are ind. for each $n \geq$

Hence, $\begin{pmatrix} Z_n \\ W_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_\alpha \\ W \end{pmatrix}$, whence

$Z_n / W_n \xrightarrow{d} Z_\alpha / W$, by the continuous mapping theorem. But $Z_n/W_n = k(\bar{X}_{2,n} - \mu)/M_n$.

3 c (iii)

Let $t_{.05}$ and $t_{.95}$ respectively denote the .05 and .95 quantiles of $T = Z_\alpha / W$. Then,

$$\lim_{n\to\infty} P\left(t_{.05} \le \frac{Z_n}{W_n} \le t_{.95} \right) = P\left(t_{.05} \le T \le t_{.95}\right)$$

$$= .95 - .05 = .90.$$

$$\Rightarrow \quad P\left(t_{.05} \le \frac{Z_{100}}{W_{100}} \le t_{.95} \right) \approx .90$$

$$\Rightarrow \quad P\left(t_{.05} \le \frac{50(\bar{X}_{2,100} - \mu)}{M_{100}} \le t_{.95} \right) = .90$$

$$\Rightarrow \quad P\left(\bar{X}_{2,100} - \frac{t_{.95} \cdot M_{100}}{50} \le \mu \le \bar{X}_{2,100} - t_{.05} \cdot \frac{M_{100}}{50} \right) = .90.$$

Hence, an approximate 90% CI for μ is:

$$\left(\bar{X}_{2,100} - \frac{t_{.95} \cdot M_{100}}{50} \;,\; \bar{X}_{2,100} - t_{.05} \cdot \frac{M_{100}}{50} \right).$$

1. For parameters θ in a parameter space Θ and behavioral decision rules ϕ with risk functions $R(\theta, \phi)$, what does it mean for a rule ϕ^* to be

a) admissible?

b) Bayes versus a prior G?

c) minimax?

2. Suppose that a parameter space Θ, a sample space $\mathcal{X}$, and an action space $\mathcal{A}$ are all finite. For $\theta \in \Theta$ suppose that P_θ is a distribution on $\mathcal{X}$ with pmf p_θ and it is the case that for each $x \in \mathcal{X}$ $\exists$ at least one $\theta \in \Theta$ with $p_\theta(x) > 0$. Let $L(\theta, a) \geq 0$ be a loss function for a decision about θ based on an observation X taking values in $\mathcal{X}$.

Suppose that Θ is the disjoint union of Θ_1 and Θ_2, and G_1 and G_2 are prior distributions on Θ with respective supports Θ_1 and Θ_2. Define

$$\mathcal{X}_1 = \{x \in \mathcal{X} | \exists\ \theta \in \Theta_1 \text{ with } p_\theta(x) > 0\}$$

and

$$\mathcal{X}_2 = \mathcal{X} - \mathcal{X}_1 \ .$$

Suppose further that for each $x \in \mathcal{X}_1$ $\exists$ a **unique** action $a_1(x)$ minimizing

$$\sum_\theta L(\theta, a) G_1(\theta) p_\theta(x) \qquad (*)$$

over choices of $a \in \mathcal{A}$, and that if $\mathcal{X}_2$ is non-void, for each $x \in \mathcal{X}_2$ $\exists$ a **unique** action $a_2(x)$ minimizing

$$\sum_\theta L(\theta, a) G_2(\theta) p_\theta(x) \qquad (**)$$

over choices of $a \in \mathcal{A}$.

It is a standard (easy) result in finite Θ problems that if $\Theta_2 = \emptyset$, a Bayes rule versus G_1 is admissible. Here we consider an extension of that result in this finite ($\mathcal{X}$ and $\mathcal{A}$) context. Let

$$\delta(x) = \begin{cases} a_1(x) & \text{if } x \in \mathcal{X}_1 \\ a_2(x) & \text{if } x \in \mathcal{X}_2 \end{cases} .$$

We wish to establish that δ is admissible. To do so, consider any behavioral rule ϕ that is at least as good as δ.

a) Argue carefully that ϕ must be Bayes versus G_1 and agree with δ on $\mathcal{X}_1$.

b) Argue carefully that risk functions $R(\theta, \delta)$ and $R(\theta, \phi)$ must agree on Θ_1.

c) Argue carefully that Bayes risks $R(G_2, \delta)$ and $R(G_2, \phi)$ must be related as $R(G_2, \delta) \leq R(G_2, \phi)$.

(Hint: Use part a).)

d) Finally, argue carefully that risk functions $R(\theta, \delta)$ and $R(\theta, \phi)$ must agree on Θ_2 (and hence that δ is admissible).

3. Consider now a particular finite decision problem with $\Theta = \mathcal{A} = \{1, 2, 3, 4\}$,

$$L(\theta, a) = \begin{cases} 0 & \text{if } a = \theta \\ 1 & \text{if } a \neq \theta \end{cases}$$

and pmfs given in the table

$\theta \backslash x$	0	1	2	3	4
1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0
2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0	0
3	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	0
4	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

For purposes of setting notation, let a generic prior distribution G in this problem have pmf g.

a) Bayes rules in this problem are not necessarily unique. Identify 2 different Bayes rules versus a prior G with $g(1) = .6, g(2) = .4$ and $g(3) = g(4) = 0$, at least one of which is admissible. (Argue carefully that one of your rules is admissible.)

b) Consider the behavioral decision rule ϕ defined by

$$\phi_0(\{1\}) = .8 \qquad \phi_1(\{2\}) = .7 \qquad \phi_2(\{3\}) = .767$$
$$\phi_0(\{2\}) = .2 \qquad \phi_1(\{3\}) = .3 \qquad \phi_2(\{4\}) = .233$$

$$\phi_3(\{4\}) = 1.0 \qquad \phi_4(\{4\}) = 1.0$$

i) Find a prior against which this rule is Bayes.

(Hint: A Bayes rule can only randomize between posterior equally attractive actions.)

ii) Argue carefully that ϕ is minimax and compute the minimax risk in this problem.

1. Classwork

2. a) ϕ at least as good as $\delta \Rightarrow R(\theta,\phi) \le R(\theta,\delta)$ for all $\theta \in \Theta$. Then

$$\sum_\theta R(\theta,\phi) G_1(\theta) \le \sum_\theta R(\theta,\delta) G_1(\theta) = R(G_1)$$

and ϕ must be Bayes versus G_1. ϕ Bayes versus G_1 implies that for each $x \in \mathcal{X}_1$ ϕ_x must place probability 1 on the set of $a \in \mathcal{A}$ minimizing the conditional expected loss, i.e. minimizing (*). Since the minimizer of (*) is unique for $x \in \mathcal{X}_1$, we have that ϕ_x places probability 1 on $\delta(x)$ for $x \in \mathcal{X}_1$.

b) If $R(\theta,\delta)$ and $R(\theta,\phi)$ did not agree on Θ_1, the facts that ϕ is at least as good as δ and δ is Bayes versus a prior G_1 with support Θ_1 would produce a contradiction.

c)
$$R(G_2,\phi) - R(G_2,\delta)$$
$$= \sum_\theta \Big(R(\theta,\phi) - R(\theta,\delta) \Big) G_2(\theta)$$
$$= \sum_\theta \Big[\sum_x \Big(\sum_a L(\theta,a)\phi_x(a) \Big) p_\theta(x) - \sum_x L(\theta,\delta(x)) p_\theta(x) \Big] G_2(\theta)$$
$$= \sum_\theta \sum_x \Big(\sum_a L(\theta,a)\phi_x(a) - L(\theta,\delta(x)) \Big) p_\theta(x) G_2(\theta)$$

by part a)

$$\overset{\text{by part a)}}{=} \sum_{\theta} \sum_{x \in \mathcal{X}_2} \left[\sum_a L(\theta,a)\phi_x(a) - L(\theta, a_2(x)) \right] p_\theta(x) G_2(\theta)$$

$$= \sum_{x \in \mathcal{X}_2} \underbrace{\sum_{\theta} \left[\qquad \qquad \right] p_\theta(x) G_2(\theta)}_{\geq 0 \text{ by } (**)} \geq 0$$

d) From c) and the fact that ϕ is at least as good as δ, $R(G_2,\phi) = R(G_2,\delta)$. Then it must be the case (since G_2 has support Θ_2) that $R(\theta,\phi) = R(\theta,\delta)$ on Θ_2. This, together with b) implies that ϕ and δ are risk equivalent. Hence δ is admissible.

3. a) The G joint dsn of θ and X is as below

$\theta \backslash x$	0	1	2	
1	.3	.3	0	.6
2	.1	.2	.1	.4
	.4	.5	.1	

Bayesness versus G requires that for $x = 0,1,2$ we choose an action minimizing conditional expected loss. This means choosing a θ maximizing conditional probability of $\theta \mid X = x$. So Bayesness requires

$$\delta(0) = 1$$
$$\delta(1) = 1$$
$$\delta(2) = 2$$

$\delta(3)$ and $\delta(4)$ can be chosen freely if the only concern is Bayesness. If we want admissibility, we can use the result from part 2). For example, using G_2 with $g(3) = g(4) = \frac{1}{2}$, we see that the choice

$$\delta(0) = 1 \qquad \delta(3) = 4$$
$$\delta(1) = 1 \quad \text{and} \quad \delta(4) = 4$$
$$\delta(2) = 2$$

is an admissible one. Another (surely inadmissible) choice is

$$\delta(0) = 1 \qquad \delta(3) = 4$$
$$\delta(1) = 1 \quad \text{and} \quad \delta(4) = 1$$
$$\delta(2) = 2$$

b) i) G joint θ, X probabilities are –

$\theta \backslash x$	0	1	2	3	4	
1	$\frac{a}{2}$	$\frac{a}{2}$				$a = g(1)$
2	$\frac{b}{4}$	$\frac{b}{2}$	$\frac{b}{4}$			$b = g(2)$
3	$\frac{c}{8}$	$\frac{3c}{8}$	$\frac{3c}{8}$	$\frac{c}{8}$		$c = g(3)$
4	$\frac{d}{16}$	$\frac{d}{4}$	$\frac{3d}{8}$	$\frac{d}{4}$	$\frac{d}{16}$	$1-a-b-c = d = g(4)$

Then, if ϕ is to be Bayes we must have

$$\frac{a}{2} = \frac{b}{4} \quad \text{i.e.} \quad b = 2a$$

$$\frac{b}{2} = \frac{3c}{8} \quad \text{i.e.} \quad c = \frac{4}{3}b = \frac{8}{3}a$$

$$\frac{3c}{8} = \frac{3d}{8} \quad \text{i.e.} \quad d = c = \frac{8}{3}a$$

This requires

$$1 - a - 2a - \frac{8}{3}a = \frac{8}{3}a$$

i.e.

$$1 = 3a + \frac{16a}{3} = \frac{25}{3}a$$

i.e. $a = \frac{3}{25}$

So the prior is $g(1) = \frac{3}{25}$, $g(2) = \frac{6}{25}$; $g(3) = \frac{8}{25} = g(4)$

ii) This rule is easily seen to be an equalizer and from i) is Bayes! It is therefore minimax.

$R(1, \phi) = \frac{1}{2}(.2) + \frac{1}{2} = .6$

$R(2, \phi) = \frac{1}{4}(.8) + \frac{1}{2}(.3) + \frac{1}{4} = .2 + .15 + .25 = .6$

$R(3, \phi) = \frac{1}{8} + \frac{3}{8}(.7) + \frac{3}{8}(.233) + \frac{1}{8} = .25 + \frac{3}{8}(.933) = .6$

$R(4, \phi) = \frac{1}{16} + \frac{1}{4} + \frac{3}{8}(.767) = .6$

1. Suppose that $X \sim P_\theta$, where $\theta \in \Theta \subset \Re$ and the P_θ are dominated by some sigma finite measure μ on the sample space $\mathcal{X}$. Suppose further that the model $\{P_\theta\}$ is Fisher Information regular at the point $\theta_0 \in \Theta$.

a) Suppose $T : \mathcal{X} \to \mathcal{T}$.

i) What does it mean for $T(X)$ to be sufficient for θ?

ii) What does it mean for $T(X)$ to be minimal sufficient?

b) Define $I(\theta_0)$, the Fisher Information about θ contained in X at the point θ_0.

c) State and prove the Cramér-Rao Inequality for a real-valued function of X, $\delta(X)$.

2. Suppose that $Y_1, Y_2, \ldots, Y_n$ are independent Poisson random variables with means

$$\lambda_i = \exp(\alpha + \beta z_i)$$

where $z_1, z_2, \ldots, z_n$ are known distinct positive constants and $\alpha \in \Re$ and $\beta \in \Re$ are unknown parameters. We consider inference for the parameter vector (α, β).

a) Identify a two-dimensional minimal sufficient statistic in this problem. (Argue very carefully that your statistic really is minimal sufficient.)

b) This model is Fisher Information regular at any point $(\alpha_0, \beta_0) \in \Re^2$. Using the two-dimensional version of the Cramér-Rao Inequality, find a lower bound on the variance of any unbiased estimator of β. (You will need to evaluate the 2×2 Fisher Information matrix $I(\alpha, \beta)$ for $Y = (Y_1, Y_2, \ldots, Y_n)$ in order to do this.)

c) A Bayesian analysis in this problem might start with a prior distribution for (α, β) specified by a joint density wrt Lebesgue measure on $\Re^2$, $g(\alpha, \beta)$. In general it will not be possible to recognize a posterior distribution of (α, β) as having a standard form. Consider the improper/generalized prior distribution specified by the choice $g(\alpha, \beta) \equiv 1$.

i) Before attempting inference based on the "posterior" distribution for (α, β), it is important to know that indeed the "posterior" is "proper." Exactly what condition needs to be checked here before proceeding?

ii) Describe in as much detail as possible a simulation-based method of generating a "sample" from the posterior, say $(\alpha^*, \beta^*)_1, (\alpha^*, \beta^*)_2, \ldots, (\alpha^*, \beta^*)_M$.

(You may assume that methodology is available for generating observations from any distribution on $\Re$ with a density known up to a multiplicative constant, but specify the form of any densities you propose to use.)

iii) Describe how you would approximate a formal Bayes estimate of β under absolute error loss, based on the simulated sample referred to in ii).

iv) Describe how you would approximate a formal Bayes decision in a test of $H_0: \beta \leq 0$ versus $H_a: \beta > 0$ under 0-1 loss, based on the simulated sample referred to in ii).

Theory IV Key | Spring 2001 Prelim

1. Class work

2. a) The R-N derivative of the dsn of $\underset{\sim}{Y}=(Y_1,\dots,Y_n)$ wrt counting measure on $\{0,1,2,\dots,\}^n$ is

$$f_{\alpha,\beta}(\underset{\sim}{y}) = \prod_{i=1}^{n} \frac{e^{-e^{(\alpha+\beta z_i)}}\left(e^{\alpha+\beta z_i}\right)^{y_i}}{y_i!}$$

$$= \frac{e^{-\sum e^{(\alpha+\beta z_i)}} \; e^{\sum y_i(\alpha+\beta z_i)}}{\prod_{i=1}^{n} y_i!}$$

$$= \frac{e^{-\sum e^{(\alpha+\beta z_i)}} \; e^{\alpha \sum y_i + \beta \sum y_i z_i}}{\prod_{i=1}^{n} y_i!}$$

The factorization theorem shows the statistic $(\sum Y_i, \sum z_i Y_i)$ to be sufficient for (α,β). $\mathbb{R}^2$ obviously contains an open rectangle and this is an exponential family. So $(\sum Y_i, \sum z_i Y_i)$ is complete and Bahadur's Theorem then guarantees that it is minimal sufficient.

b) For a single Y_i

$$\ln f_{\alpha,\beta}(y_i) = -\ln y_i! - e^{\alpha+\beta z_i} + \alpha y_i + \beta z_i y_i$$

$$\frac{\partial}{\partial\alpha}\left(\quad\right) = -e^{\alpha+\beta z_i} + y_i$$

$$\frac{\partial}{\partial\beta}\left(\quad\right) = -z_i e^{\alpha+\beta z_i} + z_i y_i$$

So the information in Y_i is

$$I_i(\alpha,\beta) = \begin{pmatrix} \operatorname{Var} Y_i & E\left(Y_i - e^{\alpha+\beta z_i}\right)(z_i Y_i) \\ \swarrow & \operatorname{Var} z_i Y_i \end{pmatrix}$$

$$= \begin{pmatrix} e^{\alpha+\beta z_i} & z_i\left(e^{\alpha+\beta z_i}\left(1+e^{\alpha+\beta z_i}\right) - e^{2(\alpha+\beta z_i)}\right) \\ \swarrow & z_i^2 e^{\alpha+\beta z_i} \end{pmatrix}$$

$$= e^{\alpha+\beta z_i}\begin{pmatrix} 1 & z_i \\ z_i & z_i^2 \end{pmatrix}$$

and the information in $\underline{Y}$ is thus

$$I(\alpha,\beta) = \begin{pmatrix} \sum e^{\alpha+\beta z_i} & \sum z_i e^{\alpha+\beta z_i} \\ \sum z_i e^{\alpha+\beta z_i} & \sum z_i^2 e^{\alpha+\beta z_i} \end{pmatrix}$$

So

$$I^{-1}(\alpha,\beta) = \frac{1}{I_{11}(\alpha,\beta)I_{22}(\alpha,\beta) - \left(I_{12}(\alpha,\beta)\right)^2}\begin{bmatrix} I_{22}(\alpha,\beta) & -I_{12}(\alpha,\beta) \\ I_{12}(\alpha,\beta) & I_{11}(\alpha,\beta) \end{bmatrix}$$

And the C-R lower bound for the variance of an unbiased estimator of β is thus

$$\frac{I_{11}(\alpha,\beta)}{I_{11}(\alpha,\beta)I_{22}(\alpha,\beta)-(I_{12}(\alpha,\beta))^2} = \frac{\sum e^{\alpha+\beta z_i}}{\left(\sum e^{\alpha+\beta z_i}\right)\left(\sum z_i^2 e^{\alpha+\beta z_i}\right)-\left(\sum z_i e^{\alpha+\beta z_i}\right)^2}$$

c) i) One really needs to know that

$$\int_{\mathbb{R}^2} f_{\alpha,\beta}(y)\, g(\alpha,\beta)\, d\alpha\, d\beta < \infty$$

Here this amounts to being sure that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\sum e^{(\alpha+\beta z_i)}} e^{\alpha\sum y_i+\beta\sum z_i y_i}\, d\alpha\, d\beta < \infty$$

ii) One might use a successive substitution / Gibbs sampling algorithm. Take $(\alpha^*,\beta^*)_0$ to be some arbitrary starting value (like $\alpha_0^* = \ln\bar{y}$ and $\beta_0^* = 0$). With $(\alpha^*,\beta^*)_j$ in hand, generate $(\alpha^*,\beta^*)_{j+1}$ as follows:

generate α^*_{j+1} from a dsn with density on $\mathbb{R}^1$ proportional to

$$e^{-\sum_i e^{(\alpha+\beta_j^* z_i)}} e^{\alpha\sum y_i+\beta_j^*\sum y_i z_i}$$

generate β^*_{j+1} from a dsn with density on $\mathbb{R}^1$ proportional to

$$e^{-\sum_i e^{(\alpha^*_{j+1}+\beta z_i)}} e^{\alpha^*_{j+1}\sum y_i+\beta\sum y_i z_i}$$

iii) Use $\hat{\beta} = \text{median}\{\beta_j^*\}$

iv) Decide in favor of H_0 iff $\#[\beta_j^* \le 0] \ge \#[\beta_j^* > 0]$