PhD Prelim Exam
THEORY
(Majors and Co-majors)

Summer 2007
(Given on 7/12/07)

In the questions below, $I\!R$ denotes the set of all reals and $\mathcal{B}(S)$ denotes the Borel sigma-field of the space S.

1. (a) Let $\{f_n\}_{n\geq 1}$ be a sequence of *nonnegative* measurable functions defined on a measure space $(\Omega, \mathcal{F}, \mu)$. Define

$$g(\omega) = \liminf_{n\to\infty} f_n(\omega),\ \omega \in \Omega.$$

(i) Prove $g : I\!R \to I\!R$ is a $\mathcal{F}-$measurable function.

(ii) If $\limsup_{n\to\infty} \int f_n d\mu < \infty$, then prove that $g \in L^1(\Omega, \mathcal{F}, \mu)$.

(b) Let Z be a *nonnegative* random variable with $Var(Z) < \infty$. Prove:

$$P(Z > 0) \geq \frac{[E(Z)]^2}{E(Z^2)}.$$

[Hint: Use Cauchy-Schwarz Inequality]

(c) Let $f : I\!R \times [0,1] \to I\!R$ be a *continuous* function (in both variables) such that $|f(t,x)| \leq g(x)$ for all t, x, for some $g \in L^1([0,1], \mathcal{B}([0,1]), m)$. Here m denotes the Lebesgue measure on $([0,1], \mathcal{B}([0,1]))$. Define for $t \in I\!R$, $h(t) \doteq \int f(t,x)dm(x)$. Prove that h is a continuous function.

2. (a) Let μ_1, μ_2 be $\sigma-$finite measures on $(I\!R, \mathcal{B}(I\!R))$, and m denotes the Lebesgue measure on $(I\!R, \mathcal{B}(I\!R))$. Assume $\alpha > 0, \beta > 0$ are real numbers and $f, g \in L^1(I\!R, \mathcal{B}(I\!R), m)$ are *nonnegative* functions.

(i) Give definitions of the measures $(\alpha\mu_1 + \beta\mu_2)$ and $\mu_1 * \mu_2$ and the function $f * g$.

(ii) Assume $\mu_1(A) = \int I_A f dm$, for $A \in \mathcal{B}(I\!R)$ and $\frac{d\mu_2}{dm} = g$. Prove that $(\alpha\mu_1 + \beta\mu_2) \ll m$, $(\mu_1 * \mu_2) \ll m$ and

$$\frac{d(\alpha\mu_1 + \beta\mu_2)}{dm} = \alpha f + \beta g, \quad \frac{d(\mu_1 * \mu_2)}{dm} = f * g.$$

(b) Let X, Y, Z be three *independent* random variables defined on some probability space $(\Omega, \mathcal{F}, P)$ such that $X \sim Normal(0,1)$, $Y \sim \chi^2(7)$, $Z \sim Bernoulli(0.4)$. Let ϕ and ψ denote the "density function" of $Normal(0,1)$ and $\chi^2(7)$ respectively. For a random variable V, let μ_V denote the *law* or the *probability distribution* of V on $(I\!R, \mathcal{B}(I\!R))$, i.e. $\mu_V(A) = P(V \in A)$ for $A \in \mathcal{B}(I\!R)$.

(i) Define the random variable $W = XZ + Y(1-Z)$. Express the *law* of W (i.e μ_W) in terms of μ_X and μ_Y. Obtain the Lebesgue decomposition of μ_W w.r.t m and express the Radon-Nikodym derivative of its absolutely continuous part in terms of ϕ, ψ.

(ii) Define the random variable $S = X + Y$. Express the *law* of S (i.e μ_S) in terms of μ_X and μ_Y. Obtain the Lebesgue decomposition of μ_W w.r.t m and express the Radon-Nikodym derivative of its absolutely continuous part in terms of ϕ, ψ.

3. (a) *Weak law of large numbers*: Let $\{Z_n\}_{n\geq 1}$ be a sequence of random variables on some probability space such that,

$$(i)\ E(Z_i^2) < \infty,\ i \geq 1;\ \ (ii)\ E(Z_iZ_j) = E(Z_i)E(Z_j),\ j \neq i;\ \ (iii)\ \frac{1}{n^2}\sum_{k=1}^{n} \sigma_k^2 \to 0,\ \text{as } n \to \infty,$$

where $\sigma_i^2 = Var(Z_i)$. Then for $\bar{Z}_n = \frac{1}{n}\sum_{k=1}^{n} Z_k$ and $\bar{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} E(Z_k)$, prove the following

$$\bar{Z}_n - \bar{\mu}_n \to 0 \text{ in probability, as } n \to \infty.$$

(b) Suppose for some integer $k \geq 1$, $g : I\!R^k \to I\!R$ be continuous at $a = (a_1, \ldots, a_k) \in I\!R^k$. Also assume that for each $i \geq 1$, the sequence of random variables $\{X_n^{(i)}\}_{n\geq 1}$ converges *in probability* to a real number a_i as $n \to \infty$, for each $i = 1, \ldots, k$. Prove that

$$g(X_n^{(1)}, \ldots, X_n^{(k)}) \to g(a_1, \ldots, a_k), \text{ in probability as } n \to \infty.$$

(c) *Consistency of the regression estimators:* Let $\{\epsilon_i\}_{i\geq 1}$ be a sequence of i.i.d random variables defined on some probability space $(\Omega, \mathcal{F}, P)$ each with mean 0, variance $\sigma^2 \in (0, \infty)$ and $E(\epsilon_1^4) \equiv \gamma < \infty$. Let $\{x_i\}_{i\geq 1}$ be a sequence of real numbers such that $\frac{1}{n}\sum_{k=1}^{n} x_k^2 \to c \in (0, \infty)$, as $n \to \infty$. For some $\beta \in I\!R$, define a new sequence of random variables on $(\Omega, \mathcal{F}, P)$:

$$Y_i(\omega) = x_i\beta + \epsilon_i(\omega),\ \ \omega \in \Omega,\ i \geq 1.$$

Using the sequence $\{Y_i\}_{i\geq 1}$, define the following two random variables on $(\Omega, \mathcal{F}, P)$ for each $n \geq 1$

$$\hat{\beta}_n(\omega) = \frac{\sum_{i=1}^{n} x_iY_i(\omega)}{\sum_{i=1}^{n} x_i^2},\ \ \hat{\sigma}_n^2(\omega) = \frac{1}{n}\sum_{i=1}^{n}\Big(Y_i(\omega) - \hat{\beta}_n(\omega)x_i\Big)^2,\ \ \omega \in \Omega.$$

Prove that the (sequence of) random variables $\hat{\beta}_n, \hat{\sigma}_n^2$ converges *in probability* to β, σ^2 respectively, as $n \to \infty$.

[Hint: (a): start with Chebychev's Inequality. (c): show $\frac{\sum x_i\epsilon_i}{n} \to 0, \frac{\sum \epsilon_i^2}{n} \to \sigma^2$ *in probability.]*

1. (a) Note that $g(\omega) = \lim_{n\to\infty} [\inf_{m\geq n} f_m(\omega)] = \sup_{n\geq 1} [\inf_{m\geq n} f_m(\omega)]$, $\omega \in \Omega$. Define $g_n(\omega) = \inf_{m\geq n} f_m(\omega)$, $\omega \in \Omega, n \geq 1$. For each $a \in I\!R$

$$\{\omega : g_n(\omega) \geq a\} = \cap_{m\geq n}\{\omega : f_m(\omega) \geq a\} = \cap_{m\geq n} f_m^{-1}[a,\infty) \in \mathcal{F}$$

since all f_n's are $\mathcal{F}$-measurable. This proves g_n is $\mathcal{F}$-measurable for each $n \geq 1$. Hence, for each $b \in I\!R$

$$\begin{aligned}\{\omega : g(\omega) \leq b\} &= \{\omega : \sup_{n\geq 1} g_n(\omega) \leq b\} = \cap_{n\geq 1}\{\omega : g_n(\omega) \leq b\} \\ &= \cap_{n\geq 1}\ g_n^{-1}(-\omega, b] \in \mathcal{F}.\end{aligned}$$

This proves that g is $\mathcal{F}$-measurable. Part (ii) follows from non-negativity of f_n and Fatou's Lemma.

(b) Apply Cauchy-Schwarz to $ZI_{\{Z>0\}}$ to obtain

$$E[ZI_{\{Z>0\}}] \leq [E(Z^2)]^{1/2}[P(Z>0)]^{1/2} \tag{1}$$

using the fact that $I^2_{\{Z>0\}} = I_{\{Z>0\}}$ and $E[I_{\{Z>0\}}] = P(Z>0)$. Also, since Z is nonnegative,

$$E(Z) = E(ZI_{(Z=0)}) + E(ZI_{(Z>0)}) = E(ZI_{(Z>0)}). \tag{2}$$

Combining (1) and (2) and squaring both sides, we get

$$\begin{aligned}[E(Z)]^2 &\leq E(Z^2)P(Z>0) \\ P(Z>0) &\geq \frac{[E(Z)]^2}{E(Z^2)}.\end{aligned}$$

(c) Fix any $t_0 \in I\!R$. It is enough to prove that $h(t_n) \to h(t_0)$ as $n \to \infty$, for all $\{t_n\}$ such that $t_n \to t_0$ as $n \to \infty$. Fix one such sequence $\{t_n\}$ such that $t_n \to t_0$. Let

$$f_n(x) \doteq f(t_n, x),\ f_0(x) \doteq f(t_0, x),\ x \in [0,1],\ n \geq 1$$

Then, by continuity of f (in both variables, here we need in x), one gets

$$f_n \to f_0 \text{ everywhere (for all } x \in [0,1])$$

and

$$|f_n(x)| \leq g(x)\ \ \forall x \in [0,1] \text{ and } \int g\, dm < \infty,$$

so, by dominated convergence theorem, we have

$$h(t_n) \equiv \int f_n(x)dm(x) \to \int f_0(x)dm(x) \equiv h(t_0) \text{ as } n \to \infty.$$

This completes the proof.

2. (a) (i) For $A \in \mathcal{B}(I\!R)$,

$$\begin{aligned}(\alpha\mu_1 + \beta\mu_2)(A) &\doteq \alpha\mu_1(A) + \beta\mu_2(A) \\ (\mu_1 * \mu_2)(A) &\doteq \int\int I_A(x+y)d\mu_1(x)d\mu_2(y)\end{aligned}$$

For $x \in I\!R$

$$(f*g)(x) \doteq \int f(x-u)g(u)du.$$

(ii) We know $\mu_1(A) = \int I_A f dm$, $\mu_2(A) = \int I_A g dm$, for all $A \in \mathcal{B}(I\!R)$. Hence, from the definition above, we have for all $A \in \mathcal{B}(I\!R)$,

$$\begin{aligned}(\alpha\mu_1 + \beta\mu_2)(A) &= \alpha \int I_A\, f\, dm + \beta \int I_A\, g\, dm \\ &= \int I_A(\alpha f + \beta g) dm \text{ (by linearity of integration).}\end{aligned}$$

Hence, $m(A) = 0 \Rightarrow (\alpha\mu_1 + \beta\mu_2)(A) = 0$, since α, β, f, g non-neg. i.e., $(\alpha\mu_1 + \beta\mu_2) << m$ and $\frac{d(\alpha\mu_1+\beta\mu_2)}{dm} = \alpha f + \beta g$. For the second part, again from the definition in (i) we get for any $A \in \mathcal{B}(I\!R)$,

$$\begin{aligned}(\mu_1 * \mu_2)(A) &= \int\int I_A(x+y) d\mu_1(x) d\mu_2(y) \\ &= \int [\int I_A(x+y) f(x) dm(x)] d\mu_2(y) \\ &= \int [\int I_A(z) f(z-y) dm(z)] d\mu_2(y) \\ &\quad \text{(by change of variable } z = x+y \\ &\quad \text{and translation invariance of } m) \\ &= \int\int I_A(z) f(z-y) g(y) dm(y) dm(z) \\ &\quad \text{(using Tonelli's theorem)} \\ &= \int I_A(z) [\int f(z-y) g(y) dm(y)] dm(z) \\ &= \int I_A(z) (f * g)(z) dm(z) \text{ (from our definition in (i))}\end{aligned}$$

Hence $m(A) = 0 \Rightarrow (\mu_1 * \mu_2)(A) = 0$ (since, by defn $f * g \geq 0$) i.e. $(\mu_1 * \mu_2) << m$ and $\frac{d(\mu_1 * \mu_2)}{dm} = f * g$.

(b) (i) For $A \in \mathcal{B}(I\!R)$

$$\begin{aligned}P(W \in A) &= P(\{W \in A\} \cap \{Z = 1\}) + P(\{W \in A\} \cap \{Z = 0\}) \\ &= P(\{X \in A\} \cap \{Z = 1\}) + P(\{Y \in A\} \cap \{Z = 0\}) \\ &\quad \text{(since } W(\omega) = X(\omega) \text{ when } Z(\omega) = 1 \text{ etc.)} \\ &= P(X \in A)(0.4) + P(Y \in A)(0.6) \quad \text{(by independence)}\end{aligned}$$

So $\mu_W(A) = P \circ W^{-1}(A) = P(W \in A) = (0.4)\mu_X(A) + (0.6)\mu_Y(A)$ ie, $\mu_W = (0.4)\mu_X + (0.6)\mu_Y$ and from part (a), $\mu_W << m$, and so, the Legesgue decomposition of $\mu_W = \mu_W + 0$, where 0 is the singular part. And, $\frac{d\mu_W}{dm} = (0.4)\frac{d\mu_X}{dm} + (0.6)\frac{d\mu_Y}{dm} = (0.4)\phi + (0.6)\psi$

(ii) For $A \in \mathcal{B}(I\!R)$

$$\begin{aligned}P(S \in A) &= P(X + Y \in A) \\ &= \int P(x + Y \in A) d\mu_X(x) \\ &= \int [\int I_A(x+y) d\mu_Y(y)] d\mu_X(x) \\ &= (\mu_X * \mu_Y)(A)\end{aligned}$$

[Alternatively, for $h(x,y) = x+y, x,y \in I\!R$,

$$\begin{aligned} P(S \in A) &= P(h(X,Y) \in A) = P((X,Y) \in h^{-1}(A)) \\ &= (\mu_X \times \mu_Y)(h^{-1}(A))(\text{Since } X, Y \text{ indep, their joint law is the product law}) \\ &= \int I_{h^{-1}(A)}(x,y) d(\mu_X \times \mu_Y)(x,y) \\ &= \int \int I_A(x+y) d\mu_Y(y) d\mu_X(x) (\text{by Tonelli}) \ = (\mu_X * \mu_Y)(A) \end{aligned}$$]

Hence $\mu_S = \mu_X * \mu_Y$, and using (a)(ii) we have $\mu_S << m$ and

$$\frac{d\mu_S}{dm} = \phi * \psi.$$

3. (a) Fix $\epsilon > 0$. We need to show:

$$P(|\bar{Z}_n - \bar{\mu}_n| > \epsilon) \to 0 \text{ as } n \to \infty.$$

Using the fact that $E(\bar{Z}_n) = \bar{\mu}_n$, we get that

$$\begin{aligned} E[(\bar{Z}_n - \bar{\mu}_n)^2] &= var(\bar{Z}_n) = \frac{1}{n^2}\left[\sum_{i=1}^{n} var(Z_i) + \sum_{i \neq j}^{n} cov(Z_i; Z_j)\right] \\ &= \frac{1}{n^2}\sum_{k=1}^{n} \sigma_k^2 + 0, \end{aligned}$$

Since by (ii), $cov(Z_i, Z_j) = 0 \ \forall i \neq j$. Hence, using Chebychev's Inequality, we get from (iii)

$$P(|\bar{Z}_n - \bar{\mu}_n| > \epsilon) \leq \frac{1}{\epsilon^2} E[(\bar{Z}_n - \bar{\mu}_n)^2] = \frac{1}{\epsilon^2}\frac{1}{n^2}\sum_{k=1}^{n} \sigma_k^2 \to 0 \text{ as } n \to \infty.$$

This completes the proof.

(b) g is continuous at $(a_1, \ldots, a_k)$, i.e. given $\epsilon > 0, \exists \delta > 0$ such that

$$\| (x^{(1)}, \ldots, x^{(k)}) - (a_1, \ldots, a_k) \| \leq \delta \ \Rightarrow \ |g(x^{(1)}, \ldots, x^{(k)}) - g(a_1, \ldots, a_k)| \leq \epsilon$$

i.e.,

$$\begin{aligned} & P(|g(X_n^{(1)}, \ldots, X_n^{(k)}) - g(a_1, \ldots, a_k)| > \epsilon) \\ \leq \ & P(\| (X_n^{(1)}, \ldots, X_n^{(k)}) - (a_1, \ldots, a_k) \| > \delta) \\ \leq \ & P(|X_n^{(1)} - a_1|^2 + \ldots + |X_n^{(k)} - a_k|^2 > \delta^2) \\ \leq \ & P(|X_n^{(1)} - a_1|^2 > \delta^2/k) + \ldots + P(|X_n^{(k)} - a_k|^2 > \delta^2/k) \\ = \ & P(|X_n^{(1)} - a_1| > \delta/\sqrt{k}) + \ldots + P(|X_n^{(k)} - a_k| > \delta/\sqrt{k}) \to 0 \text{ as } n \to \infty. \end{aligned}$$

Hence $g(X_n^{(1)}, \ldots, X_n^{(k)})) \to g(a_1, \ldots, a_k)$ in probability.
[Note that this δ can depend on $(a_1, \ldots, a_k)$ as well as ϵ. So, this argument fails if the limit is random (instead of reals $(a_1, \ldots, a_k)$)]

(c) Note that by definition of Y_i's, we have

$$
\begin{aligned}
\hat{\beta}_n &= \frac{1}{\sum x_i^2}\left[\sum_{i=1}^n x_i(x_i\beta+\epsilon_i)\right] = \beta + \left[\frac{n}{\sum x_i^2}\right]\left(\frac{1}{n}\sum_{i=1}^n x_i\epsilon_i\right) \quad (3)\\
\hat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^n [x_i(\beta-\hat{\beta}_n)+\epsilon_i]^2\\
&= (\hat{\beta}_n-\beta)^2\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) + \frac{1}{n}\sum_{i=1}^n \epsilon_i^2 + 2(\hat{\beta}_n-\beta)\left(\frac{1}{n}\sum_{i=1}^n x_i\epsilon_i\right) \quad (4)
\end{aligned}
$$

Let $Z_i = x_i\epsilon_i, i \geq 1$. Note that $\mu_i \equiv E(Z_i) = x_iE(\epsilon_i) = 0$ and $\sigma_i^2 = E(Z_i^2) = x_i^2\sigma^2$. Hence, by our assumptions on x_i's, we have $\frac{1}{n^2}\Sigma\sigma_i^2 = \frac{1}{n}(\frac{1}{n}\Sigma x_i^2)\sigma^2 \to 0$ as $n \to \infty$. $E(Z_iZ_j) = x_ix_jE(\epsilon_i\epsilon_j) = 0$, since ϵ_i, ϵ_j are indep and hence uncorrelated. Since Z_i's satisfy all 3 conditions of part (a), we get from part (a):

$$\frac{1}{n}\sum_{i=1}^n x_i\epsilon_i \to 0 \text{ in probability, as } n \to \infty. \quad (5)$$

Now let $\tilde{Z}_i = \epsilon_i^2, i \geq 1$. Note that $E(\tilde{Z}_i) = \sigma^2$, $E(\tilde{Z}_i^2) = E(\epsilon_1^4) = \gamma < \infty$, $E(\tilde{Z}_i\tilde{Z}_j) = E(\epsilon_i^2\epsilon_j^2) = E(\epsilon_i^2)E(\epsilon_j^2)\forall i \neq j$, since ϵ_i's are independent. Also $\tilde{\sigma}_i^2 = var(\tilde{Z}_i) = E(\epsilon_1^4) - [E(\epsilon_1^2)]^2 = \gamma^4 - (\sigma^2)^2 \equiv \beta$, for all $i \geq 1$. Hence, $\frac{1}{n^2}\sum\tilde{\sigma}_i^2 = \frac{1}{n}\beta \to 0$ as $n \to \infty$. Since $\tilde{Z}_i$'s satisfy all 3 conditions of part (a) we get that

$$\frac{1}{n}\sum \epsilon_i^2 - \sigma^2 \to 0 \text{ in probability, as } n \to \infty. \quad (6)$$

Using (5)-(6) in (3)-(4) the proof is complete by part (b), since for some continuous g_1, g_2, we have

$$
\begin{aligned}
\hat{\beta}_n &= g_1\left(\frac{1}{n}\sum x_i^2,\ \frac{1}{n}\sum_{i=1}^n x_i\epsilon_i\right) \quad (7)\\
\hat{\sigma}_n^2 &= g_2\left(\hat{\beta}_n,\ \frac{1}{n}\sum x_i^2,\ \frac{1}{n}\sum_{i=1}^n x_i\epsilon_i,\ \frac{1}{n}\sum_{i=1}^n \epsilon_i^2\right).
\end{aligned}
$$

Please begin every answer to a part or sub-part of this question on a new sheet of paper!!

Part 1

Use a probabilistic argument to establish sufficient conditions on a sequence of positive real numbers $\{\theta_i\}$ so that $\exists\ \gamma > 0$ (depending upon the sequence) such that for every $t \in \Re$

$$\prod_{i=1}^{n} \frac{\sin\left(\frac{t\theta_i}{\sqrt{n}}\right)}{\frac{t\theta_i}{\sqrt{n}}} \xrightarrow{n\to\infty} \exp\left(-\gamma t^2\right)$$

(This will hold, e.g., for constant θ_i. But what more general condition will suffice to guarantee this convergence?)

Part 2

a) Suppose that the family of distributions $\{P_\theta\}$ for a random quantity X is dominated by a sigma-finite measure μ and that the family of distributions $\{Q_\gamma\}$ for a random quantity Y is dominated by another sigma-finite measure λ. Suppose that a statistic $T(X)$ is sufficient for $\{P_\theta\}$ and a statistic $S(Y)$ is sufficient for $\{Q_\gamma\}$. Prove or give a counter-example that statistic $(T(X), S(Y))$ is sufficient for $\{P_\theta \times Q_\gamma\}$.

b) Suppose that the family of distributions $\{P_\theta\}$ for the random quantity (X,Y) is dominated by a product of sigma-finite measures $\mu \times \lambda$. Suppose that the statistic $T(X,Y) = (T_1(X), T_2(Y))$ is sufficient for $\{P_\theta\}$. Prove or give a counter-example that the statistic $T_1(X)$ is sufficient for $\{P_\theta^X\}$, the set of marginal distributions of X.

Part 3

In the balance of this question we will consider versions of a decision problem where the parameter space is

$$\Theta = (0,\infty) \times (0,\infty),$$

the action space is

$$\mathcal{A} = \{1,2\}$$

and the loss function is

$$L(\boldsymbol{\theta}, a) = (\theta_2 - \theta_1) I[a=1] I[\theta_2 > \theta_1] + (\theta_1 - \theta_2) I[a=2] I[\theta_1 > \theta_2]$$

(The loss function is a "linear loss" appropriate to deciding which co-ordinate of the parameter vector $\boldsymbol{\theta}$ is larger.) For $\gamma > 0$ we will let P_γ be the exponential distribution with mean γ, i.e. the continuous distribution on $(0,\infty)$ with density with respect to Lebesgue measure of the form $f_\gamma(x) = \frac{1}{\gamma}\exp\left(-\frac{x}{\gamma}\right)$.

Consider first a version of the decision problem where the information available upon which to make a decision consists of

$$(X_1, X_2) \sim P_{\theta_1} \times P_{\theta_2}$$

(so that one has independent observations from both of the exponential distributions P_{θ_1} and P_{θ_2}).

a) Consider the non-randomized decision function

$$\delta(X_1, X_2) = 1 + I[X_2 \geq X_1].$$

Find the risk function for this decision rule. You may use without proof the fact that in this model

$$P[X_1 > X_2] = \frac{\theta_1}{\theta_1 + \theta_2}$$

b) What is the form of a non-randomized decision rule that is formal Bayes versus a prior on Θ that makes θ_1 and θ_2 *a priori* iid Inverse-$\Gamma(2,1)$? (You may wish to consider the difference $L(\boldsymbol{\theta},2) - L(\boldsymbol{\theta},1)$. Information about the Inverse-Γ distribution can be found on the last page of this question.)

c) Does the rule in **b)** have finite Bayes risk for the prior of part **b)**? (Show appropriate work to establish this one way or the other.)

d) Argue that the Bayes rule in part **b)** is admissible under this model. (If there is some detail that would need to be argued in a really careful answer here, say what it is and why you expect the condition to hold.)

Consider now a small "sequential" version of this problem, where

$$X_1 \text{ and } Y_1 \text{ are iid } P_{\theta_1}$$

independent of

$$X_2 \text{ and } Y_2 \text{ that are iid } P_{\theta_2},$$

and for $\Delta(X_1, X_2)$ a function that takes only the values 1 and 0,

$$Z = \Delta(X_1, X_2)Y_1 + (1 - \Delta(X_1, X_2))Y_2$$

and what is available for decision-making about $\boldsymbol{\theta}$ is the vector

$$(X_1, X_2, Z)$$

(one gets to see both of X_1 and X_2, and based on them, gets to choose whether to observe Y_1 or Y_2).

e) Consider the $\Delta(X_1, X_2) = I[X_1 > X_2]$ version of this model. Show explicitly how to find the Fisher information matrix

$$\mathcal{I}_{(X_1,X_2,Z)}(\boldsymbol{\theta}) = \left(-\mathrm{E}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln f_{\boldsymbol{\theta}}(X_1, X_2, Z)\right)_{\substack{i=1,2\\ j=1,2}}$$

for this choice. (You do not need to completely finish this calculation, but you do need to reduce it to terms involving integrals that you explain why it is possible to evaluate in closed form.)

f) For a fixed choice of $\Delta(X_1, X_2)$ and observed (X_1, X_2, Z), in as explicit terms as is possible, say what action is best if *a priori* θ_1 and θ_2 are iid Inverse-$\Gamma(2,1)$.

g) Outline how one would go about finding an optimal choice of the function $\Delta(X_1, X_2)$ if *a priori* θ_1 and θ_2 are iid Inverse-$\Gamma(2,1)$. (Exactly what would need to be optimized by choice of Δ?)

The Inverse-$\Gamma(\alpha,\beta)$ distribution is a distribution on $(0,\infty)$ with density

$$f(w|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{-(\alpha+1)} \exp\left(-\frac{\beta}{w}\right)$$

For $\alpha > 1$ this distribution has mean $\beta/(\alpha-1)$ and for $\alpha > 2$ it has finite variance.

Part 1

This is convergence of the characteristic function of $\sqrt{n}\,\bar{X}_n$ to that of a normal distribution for X_i's that are independent $U(-\theta_i, \theta_i)$. Note that $EX_i = 0$

$B_n^2 = \sum_{i=1}^n EX_i^2 = \frac{1}{3}\sum_{i=1}^n \theta_i^2$ and for $\delta > 0$

$$E|X_i|^{2+\delta} = 2\int_0^\theta \frac{1}{2\theta}\, x^{2+\delta}\, dx$$

$$= \frac{1}{\theta}\left. \frac{x^{3+\delta}}{3+\delta}\right|_0^\theta = \frac{\theta^{2+\delta}}{3+\delta}$$

So $\sum_{i=1}^n E|X_i|^{2+\delta} = \frac{1}{3+\delta}\sum_{i=1}^n \theta_i^{2+\delta}$ for $\delta > 0$. The Liapounov condition will suffice, i.e. that $\exists\ \delta > 0$ for which

$$\frac{\sum_{i=1}^n \theta_i^{2+\delta}}{\left(\sqrt{\sum_{i=1}^n \theta_i^2}\right)^{2+\delta}} \longrightarrow 0$$

Part 2

a) This is true. The factorization theorem says that we may write

$$\frac{dP_\theta}{d\mu}(x) = h_1(x)\, g_{1\theta}(T(x)) \quad \text{a.e. } \mu$$

and

$$\frac{dQ_\gamma}{d\lambda}(y) = h_2(y)\, g_{2\gamma}(S(y)) \quad \text{a.e. } \lambda$$

Then

$$\frac{d(P_\theta \times Q_\gamma)}{d(\mu \times \lambda)} = h_1(x)\, g_{1\theta}(T(x))\, h_2(y)\, g_{2\gamma}(S(y))$$

$$= \underbrace{h_1(x)\, h_2(y)}_{h(x,y)}\ \underbrace{g_{1\theta}(T(x))\, g_{2\gamma}(S(y))}_{g_{(\theta,\gamma)}(T(x),\, S(y))}$$

Part 2 b) This is not in general true. The sufficiency of T and the factorization Theorem give us

$$\frac{dP_\theta}{d(\mu\times\lambda)}(x,y) = h(x,y)\, g_\theta\,(T_1(x), T_2(y)) \quad \text{a.e. } \mu\times\lambda$$

But

$$\frac{dP_\theta^X}{d\mu} = \int \frac{dP_\theta}{d(\mu\times\lambda)}(x,y)\, d\lambda(y)$$

$$= \int h(x,y)\, g_\theta\,(T_1(x), T_2(y))\, d\lambda(y)$$

and there is just no guarantee that this integral produces proportional functions of θ $\forall$ x with a common $T_1(x)$ value. Simple discrete counter-examples are easy to make. Consider the joint discrete dsn with pmf specified by

x \ y	1	2	$f_\theta(x)$
1	θ	$\frac{1}{8}(1-\theta)$	$\frac{1}{8}+\frac{7}{8}\theta$
2	2θ	$\frac{1}{8}(1-\theta)$	$\frac{1}{8}+\frac{15}{8}\theta$
3	$\frac{1}{2}$	$\frac{1}{4}-\frac{22}{8}\theta$	$\frac{3}{4}-\frac{22}{8}\theta$

for $0<\theta<\frac{2}{22}$. Let $T_1(x) = \begin{cases} 0 & \text{if } x = 1 \text{ or } 2 \\ 7 & \text{if } x = 3 \end{cases}$

$T_2(y) = y$. $(T_1(X), T_2(Y))$ is sufficient for θ based on (X,Y) but $T_1(X)$ is not sufficient for θ based on X alone.

Part 3a)

$$R(\underline{\theta},\delta) = (\theta_2-\theta_1) I[\theta_2>\theta_1] P_{\underline{\theta}}[X_1>X_2]$$
$$+ (\theta_1-\theta_2) I[\theta_1>\theta_2] P_{\underline{\theta}}[X_2\geq X_1]$$
$$= I[\theta_2>\theta_1]\frac{(\theta_2-\theta_1)\theta_1}{\theta_1+\theta_2} + I[\theta_1>\theta_2]\frac{(\theta_1-\theta_2)\theta_2}{\theta_1+\theta_2}$$

Part 3b) Note that

$$L(\underline{\theta},2)-L(\underline{\theta},1) = (\theta_1-\theta_2)I[\theta_1>\theta_2] - (\theta_2-\theta_1)I[\theta_2>\theta_1]$$
$$= \theta_1-\theta_2$$

So then, the posterior mean of this difference is the difference in posterior means of θ_1 and θ_2 and so for a prior G, the Bayes rule has the form

$$\delta_G(X_1,X_2) = 1 + I\left[E_G[\theta_1|X_1,X_2] \leq E_G[\theta_2|X_1,X_2]\right]$$

Note that for the present prior, the joint pdf of $\theta_1,\theta_2,X_1,X_2$ is

$$f(\theta_1,\theta_2,x_1,x_2) \propto \frac{1}{\theta_1^3}e^{-\frac{1}{\theta_1}}\frac{1}{\theta_2^3}e^{-\frac{1}{\theta_2}}\frac{1}{\theta_1}e^{-\frac{x_1}{\theta_1}}\frac{1}{\theta_2}e^{-\frac{x_2}{\theta_2}}$$

from which it follows that the posterior of θ_i is

$$\text{Inv}-\Gamma(3, x_i+1)$$

and thus that the posterior mean of θ_i is $\frac{1}{2}(X_i+1)$. So the Bayes rule has the form

$$\delta_G(X_1,X_2) = 1 + I[X_2\geq X_1]$$

That is, the rule of 3a) is Bayes for this prior for (θ_1,θ_2)

Part 3c) This is the question of whether the risk function 3a) has a finite integral against the prior. But

$$|R(\underline{\theta},\delta)| \le |\theta_1-\theta_2| \le \theta_1+\theta_2$$

and thus since both θ_1 and θ_2 have finite prior means the answer is Yes!

Part 3d) Since Θ is open, the prior puts positive probability on every open ball in Θ and the rule has finite Bayes risk against the prior; all we need to know is that risk functions are cont$\underline{s}$ in $\underline{\theta}$. These are of the form (for non-randomized ϕ)

$$R(\underline{\theta},\phi) = (\theta_1-\theta_2)\,I[\theta_1>\theta_2]\,P_{\underline{\theta}}[\phi(X_1,X_2)=2] + (\theta_2-\theta_1)\,I[\theta_2>\theta_1]\,P_{\underline{\theta}}[\phi(X_1,X_2)=1]$$

and so continuity of $P_{\underline{\theta}}[\phi(X_1,X_2)=i]$ in $\underline{\theta}$ suffices. This presumably holds because of the continuity of the joint pdf of (X_1,X_2) in $\underline{\theta}$.

Part 3e)

$$f(x_1,x_2,z\,|\,\theta_1,\theta_2) = \left(\frac{1}{\theta_1}e^{-\frac{x_1}{\theta_1}}\frac{1}{\theta_2}e^{-\frac{x_2}{\theta_1}}I[x_1>x_2]\frac{1}{\theta_1}e^{-\frac{z}{\theta_1}}\right)+\left(\frac{1}{\theta_1}e^{-\frac{x_1}{\theta_1}}\frac{1}{\theta_2}I[x_2>x_1]\frac{1}{\theta_2}e^{-\frac{z}{\theta_2}}\right)$$

$$= \frac{1}{\theta_1}e^{-\frac{x_1}{\theta_1}}\frac{1}{\theta_2}e^{-\frac{x_2}{\theta_2}}\left(\frac{1}{\theta_1}e^{-\frac{z}{\theta_1}}I[x_1>x_2] + \frac{1}{\theta_2}e^{-\frac{z}{\theta_2}}I[x_1<x_2]\right)$$

$$\ln f(x_1,x_2,z\,|\,\theta_1,\theta_2) = \left(-2\ln\theta_1-\ln\theta_2-\frac{x_1+z}{\theta_1}-\frac{x_2}{\theta_2}\right)I[x_1>x_2] + \left(-\ln\theta_1-2\ln\theta_2-\frac{x_1}{\theta_1}-\frac{x_2+z}{\theta_2}\right)I[x_1<x_2]$$

So $\frac{\partial}{\partial\theta_1}\ln f(x_1,x_2,z|\theta_1,\theta_2) = \left(-\frac{z}{\theta_1}+\frac{x_1+z}{\theta_1^2}\right)I[x_1>x_2]$

$+\left(-\frac{1}{\theta_1}+\frac{x_1}{\theta_1^2}\right)I[x_1<x_2] = \frac{x_1}{\theta_1^2}-\frac{1}{\theta_1}+I[x_1>x_2]\left(-\frac{1}{\theta_1}+\frac{z}{\theta_1^2}\right)$

Similarly, $\frac{\partial}{\partial\theta_2}\ln f(x_1,x_2,z|\theta_1,\theta_2) = \frac{x_2}{\theta_2^2}-\frac{1}{\theta_2}+I[x_1<x_2]\left(-\frac{1}{\theta_2}+\frac{z}{\theta_2^2}\right)$

So the mixed partials are 0 and we have

$\frac{\partial^2}{\partial\theta_1^2}\ln f(x_1,x_2,z|\theta_1,\theta_2) = -\frac{2x_1}{\theta_1^3}+\frac{1}{\theta_1^2}+I[x_1>x_2]\left(\frac{1}{\theta_1^2}-\frac{2z}{\theta_1^3}\right)$

and

$\frac{\partial^2}{\partial\theta_2^2}\ln f(x_1,x_2,z|\theta_1,\theta_2) = -\frac{2x_2}{\theta_2^3}+\frac{1}{\theta_2^2}+I[x_1<x_2]\left(\frac{1}{\theta_2^2}-\frac{2z}{\theta_2^3}\right)$

So consider

$E_{\theta_1,\theta_2}\frac{\partial^2}{\partial\theta_1^2}\ln f(X_1,X_2,Z|\theta_1,\theta_2) = E_{\theta_1,\theta_2}\left(E_{\theta_1,\theta_2}[\text{same}\,|\,X_1,X_2]\right)$

$= E_{\theta_1,\theta_2}\left[-\frac{2X_1}{\theta_1^3}+\frac{1}{\theta_1^2}+I[X_1>X_2]\left(\frac{1}{\theta_1^2}-\frac{2}{\theta_1^2}\right)\right]$

$= -\frac{1}{\theta_1^2}\left[1+\frac{\theta_1}{\theta_1+\theta_2}\right]$

and obviously, $E_{\theta_1,\theta_2}\frac{\partial}{\partial\theta_1^2}\ln f(X_1,X_2,Z|\theta_1,\theta_2) =$

$= -\frac{1}{\theta_2^2}\left[1+\frac{\theta_2}{\theta_1+\theta_2}\right]$

<u>Part 3f</u> The joint pdf for $(X_1,X_2,Z,\theta_1,\theta_2)$ is

$f(x_1,x_2,z,\theta_1,\theta_2) \propto f(x_1,x_2,z|\theta_1,\theta_2)\frac{1}{\theta_1^3}e^{-\frac{1}{\theta_1}}\frac{1}{\theta_2^2}e^{-\frac{1}{\theta_2}}$

so in the event that $\Delta(X_1,X_2)=1$ the posterior is proportional to

$$\frac{1}{\theta_1} e^{-\frac{X_1}{\theta_1}} \; \frac{1}{\theta_2} e^{-\frac{X_2}{\theta_2}} \; \frac{1}{\theta_1} e^{-\frac{Z}{\theta_1}} \; \frac{1}{\theta_1^3} e^{-\frac{1}{\theta_1}} \; \frac{1}{\theta_2^3} e^{-\frac{1}{\theta_2}}$$

$$= \frac{1}{\theta_1^5} e^{-\frac{1}{\theta_1}(X_1+Z+1)} \; \frac{1}{\theta_2^4} e^{-\frac{1}{\theta_2}(X_2+1)}$$

i.e. the product of

$$\text{Inv-}\Gamma(4, X_1+Z+1) \text{ and } \text{Inv }\Gamma(3, X_2+1)$$

and in the event that $\Delta(X_1,X_2)=0$ the posterior is the product of

$$\text{Inv-}\Gamma(3, X_2+1) \text{ and } \text{Inv-}\Gamma(4, X_1+Z+1)$$

So the Bayes rule is

$$\delta_G^\Gamma(X_1,X_2,Z) = \left(1+I\left[\frac{X_1+Z+1}{3} < \frac{X_2+1}{2}\right]\right)\Delta(X_1,X_2)$$

$$+\left(1+I\left[\frac{X_1+1}{2} < \frac{X_2+Z+1}{3}\right]\right)\left(1-\Delta(X_1,X_2)\right)$$

<u>Part 3g)</u> In principle, one computes the risk function of δ_G^Δ for each Δ

$$R(\underline{\theta}, \delta_G^\Delta) = (\theta_2-\theta_1)\, I[\theta_2>\theta_1]\, P_{\underline{\theta}}\left[\delta_G^\Delta(X_1,X_2,Z)=1\right]$$

$$+ (\theta_1-\theta_2)\, I[\theta_1>\theta_2]\, P_{\underline{\theta}}\left[\delta_G^\Delta(X_1,X_2,Z)=2\right]$$

finds the corresponding Bayes risk (the average of the above against G), and then attempts to minimize by choice of Δ.

Part I

Let $X_1, X_2, \ldots, X_n$ be iid random variable with uniform $(-\theta, \theta)$ distribution where $\theta > 0$ is an unknown parameter.

(1) Write down the likelihood function $L(X_1, X_2, \ldots, X_n, \theta)$ based on the data $X_1, X_2, \ldots, X_n$ and the parameter θ. Find the maximum likelihood estimator (MLE) $\hat{\theta}_n$ of θ based on the data $X_1, X_2, \ldots, X_n$.

(2) Find the probability distribution of $\hat{\theta}_n$ for each value of θ and show that the random variable $\hat{\theta}_n$ converges in distribution to θ as $n \to \infty$.

(3) Find an exact 95% confidence interval of the form (L_n, ∞) for θ based on the distribution of $\hat{\theta}_n$ in question #2.

(4) Derive the mean, variance and mean square error of $\hat{\theta}_n$.

(5) Show that $E(\hat{\theta}_n - \theta)^2 \to 0$ as $n \to \infty$.

(6) Derive the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$; does $\sqrt{n}(\hat{\theta}_n - \theta)$ converge in distribution to a normal distribution?

(7) Derive the distribution of $n(\hat{\theta}_n - \theta)$; does $n(\hat{\theta}_n - \theta)$ converge in distribution to a distribution? Identify the limiting distribution if it does converge, or give reason if otherwise.

Part II

Let $X_1, X_2, \ldots, X_n$, be a sequence of iid random variable with distribution function $F(\cdot)$. For each x let

$$F_n(x) \equiv \frac{1}{n} \sum_{j=1}^{n} I_{(-\infty, x]}^{(X_j)}$$

be the proportion of observations less than or equal to x among $X_1, X_2, \ldots, X_n$.

(7) Show that for each x, $F_n(x)$ converges w.p. 1 to $F(x)$.

(8) Show that for each x,

$$Y_n(x) \equiv \sqrt{n}(F_n(x) - F(x))$$

converges in distribution as $n \to \infty$ to a normal distribution. Identify the mean and variance of the limit distribution.

(9) Compute $\mathrm{Cov}(Y_n(x_1), Y_n(x_2))$ for $x_1, x_2 \in R$, $n \geq 1$.

(10) Show that $(Y_n(x_1), Y_n(x_2))$ converges in distribution as $n \to \infty$ to a bivariate normal distribution. Identify the covariance of the limit distribution.

Solution to 542/543 Question - Theory III
Majors/Co-majors

(1) $L_n(\theta) = \left(\frac{1}{2\theta}\right)^n I\left(-\theta \le X_{(1)} \le X_{(n)} < \theta\right)$

Hence $\hat{\theta}_n = \max\{X_{(n)}, -X_{(1)}\}$

(2) $P(\hat{\theta}_n < y) = P(-X_{(1)} < y,\ X_{(n)} < y) = \prod_{i=1}^{n} P(-y < X_i < y)$

$= \{F_X(y) - F_X(-y)\}^n$ where F_X is the cdf of X.

$$= \begin{cases} \left(\frac{y+y}{2\theta}\right)^n = \left(\frac{y}{\theta}\right)^n & \text{if } y \in [0, \theta] \\ 1 & \text{if } y > \theta \\ 0 & \text{if } y < 0 \end{cases}$$

$$F_X(y) = \begin{cases} \frac{y+\theta}{2\theta} & y \in [-\theta, \theta] \\ 1 & y > \theta \\ 0 & y < -\theta \end{cases}$$

The dist of θ is $F_\theta(y) = \begin{cases} 1 & \text{if } y \ge \theta \\ 0 & \text{o.w} \end{cases}$

As $P(\hat{\theta}_n < y) \to 0$ for any $y < \theta$, and $P(\hat{\theta}_n < y) \to 1$ if $y \ge \theta$

So. $\hat{\theta}_n \xrightarrow{d} \theta$.

(3) let $y_0 = \theta (0.95)^{1/n}$, then $P(\hat{\theta}_n < y_0) = 0.95$.

Hence $P\left(\theta > (0.95)^{-1/n} \hat{\theta}_n\right) = 0.95$. So, $L_n = (0.95)^{-1/n} \hat{\theta}_n$.

(4) $E(\hat{\theta}_n) = \int_0^\theta y \frac{n}{\theta} \left(\frac{y}{\theta}\right)^{n-1} dy = \frac{n}{\theta^n} \int_0^\theta y^n \, dy = \frac{n}{n+1}\theta$

$E(\hat{\theta}_n^2) = \frac{n}{\theta^n} \int_0^\theta y^{n+1} dy = \frac{n}{n+2}\theta^2$, $Var(\hat{\theta}_n) = \left[\frac{1}{n^2} + o\left(\frac{1}{n^2}\right)\right]\theta^2$

$= \left[\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right]\theta^2$

$$MSE(\hat{\theta}_n) = \left[\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right]\theta^2 + \frac{\theta^2}{(n+1)^2} = O(n^{-2})$$

This is a Super-rate !

(5) As $MSE(\hat{\theta}_n) \to 0$ as $n \to \infty$ (from (4)), $\hat{\theta}_n$ is the mean square consistent to θ.

(6) $P(\sqrt{n}(\hat{\theta}_n - \theta) < y) = P(\hat{\theta}_n < \theta + \frac{y}{\sqrt{n}})$

$$= \begin{cases} 1 & \text{if } y > 0 \\ \left(1 + \frac{y/\theta}{\sqrt{n}}\right)^n & \text{if } y \in (-\theta\sqrt{n}, 0] \\ 0 & \text{if } y \leq -\theta\sqrt{n} \end{cases} \not\to \text{ a Normal dist.}$$

(7) $P(n(\hat{\theta}_n - \theta) < y) = P(\hat{\theta}_n < \theta + y/n)$

$$= \begin{cases} 1 & \text{if } y > 0 \\ \left(1 + \frac{y}{\theta n}\right)^n & \text{if } y \in (-\theta n, 0] \\ 0 & \text{if } y \leq -\theta n \end{cases}$$

$$\to \begin{cases} 1 & \text{if } y > 0 \\ e^{y/\theta} & \text{if } y \leq 0 \end{cases}$$

which is a negative Exponential distribution with mean $-\theta$.

part II

(7) As $\{ I_{(-\infty, x]}(X_i) \}_{i=1}^{n}$ are bounded IID r.vs,

$E\, I_{(-\infty,x]}(X_i) = F(x)$, from the SLLN

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,x]}(X_i) \xrightarrow{wp1} F(x).$$

(8) use of the standard CLT as $\{ I_{(-\infty, x]}(X_i) \}_{i=1}^{n}$ has finite first two moments

(9) $$Cov(Y_n(x_1), Y_n(x_2)) = \frac{n}{n} Cov(I_{(-\infty, x_1]}(X_i), I_{(-\infty, x_2]}(X_i))$$

$$= F(\min(x_1, x_2)) - F(x_1) F(x_2)$$

(10) $$\begin{pmatrix} Y_n(x_1) \\ Y_n(x_2) \end{pmatrix} = \frac{\sqrt{n}}{n} \begin{pmatrix} \sum_{i=1}^{n} (I_{(-\infty, x_1]}(X_i) - F(x_1)) \\ \sum_{i=1}^{n} (I_{(-\infty, x_2]}(X_i) - F(x_2)) \end{pmatrix}$$

Use the Cramér-Wold device, $\forall\, a = (a_1, a_2)^T \in R^2$,

$$a' \begin{pmatrix} Y_n(x_1) \\ Y_n(x_2) \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ a_1 (I_{(-\infty, x_1]}(X_i) - F(x_1)) + a_2 (I_{(-\infty, x_2]}(X_i) - F(x_2)) \}$$

Which is the $\sqrt{n}$ x Average of IID r.vs., with zero mean and finite second moment $a^T V a$ where

$$V = \begin{bmatrix} F(x_1)(1-F(x_1)) & F(\min(x_1,x_2)) - F(x_1)F(x_2) \\ F(\min(x_1,x_2)) - F(x_1)F(x_2) & F(x_2)(1-F(x_2)) \end{bmatrix}$$

Hence $a' \begin{pmatrix} Y_n(x_1) \\ Y_n(x_2) \end{pmatrix} \xrightarrow{d} N(0, a^T V a) \quad \forall a \in \mathbb{R}^2$

$\Rightarrow \begin{pmatrix} Y_n(x_1) \\ Y_n(x_2) \end{pmatrix} \xrightarrow{d} N_2(0, V)$, a bivariate normal r. vector with zero mean & covariance V.

Recall that the probability density function (pdf) of a random variable Z with Gamma(α, β) distribution is

$$f(z) = \frac{1}{\Gamma(\alpha)\beta^\alpha} z^{\alpha-1} e^{-z/\beta},$$

where $z > 0$ and $\alpha, \beta > 0$ and $f(z) = 0$ otherwise. The mean and variance of Z are $\alpha\beta$ and $\alpha\beta^2$, respectively. Recall that the pdf of a random variable W with a Beta(α, β) distribution is

$$f(w) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1}(1-w)^{\beta-1},$$

where $0 < w < 1$ and $\alpha, \beta > 0$ and $f(w) = 0$ otherwise. The mean and variance of W are $\alpha/(\alpha+\beta)$ and $\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$, respectively.

1. Let the joint density of X and Y be

$$f(x, y) = \frac{\alpha}{\Gamma(2\alpha)} \frac{x^{2\alpha-1}}{y^{\alpha+1}} e^{-x/y}$$

   for $0 < x < \infty$ and $0 < y < 1$ and $f(x, y) = 0$ otherwise. Assume $\alpha > 1$.

   (a) Find the variance of Y.

   (b) What is the conditional expectation of $XY/2$ given that $Y = y$?

   (c) What is the density of X/Y?

2. Suppose that random variables S has a Gamma($3\alpha, \beta$) distribution. Consider $S_1, S_2, \ldots, S_n$ an independent sample from S.

   (a) Demonstrate that $\bar{S}_n = \frac{1}{n}\sum_{i=1}^n S_i$ is consistent for, i.e., converges in probability to, the mean of S.

   (b) Propose an estimator $\hat{\tau}_n$ that is consistent for $\tau = \alpha^2\beta^2$. Demonstrate that your estimator is consistent for, i.e., converges in probability to, τ.

   (c) Completely specify the limiting distribution of $\sqrt{n}(\hat{\tau}_n - \tau)$, where $\hat{\tau}_n$ is the estimator from the previous part, as n goes to infinity.

3. Assume V has a Beta(α, 1) distribution.

   (a) Assume you can generate values from independent Uniform(0,1) random variables. Explain in detail how to generate a value v.

   (b) Instead of generating a value as described in the previous part, someone suggests generating a value from $V' \sim N(E(V), Var(V))$. What is wrong with this suggestion?

Co Major Theory I   Summer 2007   Solutions

page 1

$$f(x,y) = \frac{\alpha}{\Gamma(2\alpha)} \frac{x^{2\alpha-1}}{y^{\alpha+1}} e^{-x/y}$$

$0 < x < \infty$
$0 < y < 1$
$\alpha > 1$

1a $V(Y)$

Find $f_Y(y) = \int_0^\infty f(x,y)\,dx$

Recognize that $X|y \sim \text{Gamma}(2\alpha, y)$

$$= \frac{\alpha y^{2\alpha}}{y^{\alpha+1}} \int_0^\infty \frac{1}{\Gamma(2\alpha)} \frac{x^{2\alpha-1}}{y^{2\alpha}} e^{-x/y}\,dx$$

$$= \alpha y^{\alpha-1}$$

see that $Y \sim \text{Beta}(\alpha, 1)$

$EY = \frac{\alpha}{\alpha+1}$ $VY = \frac{\alpha}{(\alpha+1)^2(\alpha+2)} = EY^2 - (EY)^2$

$EY^2 = \frac{\alpha}{\alpha+2}$

1b $E(\frac{X}{2}Y|y)$

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{\Gamma(2\alpha)} \frac{x^{2\alpha-1}}{y^{2\alpha}} e^{-x/y}$$

$X|y \sim \text{Gamma}(2\alpha, y)$

$E X|y = 2\alpha y$

$E \frac{Xy}{2}|y = \alpha y^2.$

Co-major Theory I

(1c.) Density of X/Y

Let $V = X/Y$ $\quad X = VW$ $\quad J = \begin{vmatrix} w & v \\ 0 & 1 \end{vmatrix} = w$.

$W = Y$ $\quad Y = W$

$$f_{VW}(v,w) = f_{XY}(vw, w)\, w$$

$$= \frac{\alpha}{\Gamma(2\alpha)} \frac{(vw)^{2\alpha-1}}{w^{\alpha+1}} e^{-v} w$$

$$= \frac{\alpha}{\Gamma(2\alpha)} v^{2\alpha-1} e^{-v} w^{\alpha-1}$$

V and W are independent $\quad V \sim \text{Gamma}(2\alpha, 1)$

S_i iid Gamma $(3\alpha, \beta)$ $\quad ES_i = 3\alpha\beta$

$ES_i = 3\alpha\beta$ $\quad \bar{S} = \frac{1}{n}\sum_{i=1}^{n} S_i$ $\quad E\bar{S} = 3\alpha\beta$

$VS_i = 3\alpha\beta^2$ $\quad V\bar{S} = 3\alpha\beta^2/n$

2(a.) Since $E\bar{S} = 3\alpha\beta$ and $V\bar{S} = 3\alpha\beta^2/n$,
$E(\frac{1}{3}\bar{S}) = \alpha\beta$ and $V\frac{1}{3}\bar{S} = \alpha\beta^2/3n$,
by WLLN $\frac{1}{3}\bar{S}$ is consistent for $\alpha\beta$.

2(b) The function $g(t) = t^2$ is differentiable,
so $g(\frac{1}{3}\bar{S}) = \frac{1}{9}\bar{S}^2$ is consistent for $\alpha^2\beta^2$.

(2c) $\sqrt{n}\left(\frac{1}{3}\bar{S} - \alpha\beta\right) \xrightarrow{D} N(0, \alpha\beta^2/3)$

By delta method $[g'(t) = 2t \quad g'(t)^2 = 4t^2]$,

$$\sqrt{n}\left(\frac{1}{9}\bar{S}^2 - \alpha^2\beta^2\right) \xrightarrow{D} N\left(0, \frac{\alpha\beta^2}{3} 4\alpha^2\beta^2 = \frac{4\alpha^3\beta^4}{3}\right).$$

Co-Major Theory I

6. $V \sim \text{Beta}(\alpha, 1)$
$W | V=v \sim \text{Gamma}(2, v)$

Generate $u_1, u_2, u_3 \overset{iid}{\sim} \text{Unif}(0,1)$.

(3a) $f_V(v) = \alpha v^{\alpha-1}$ $\quad F_V(v) = v^\alpha$ $\quad (0<v<1)$
$F_V(v) = v^\alpha = u_1$
Let $v = u_1^{1/\alpha}$

$W | V=v \sim \text{Gamma}(2, v) = w_1 + w_2$ where
w_1 & w_2 are iid $\text{Exp}(v)$
$f_{w_i}(w) = \frac{1}{v} e^{-w/v}$ $\quad F_{w_i}(w) = 1 - e^{-w/v}$

Let $1 - e^{-w_i/v} = u_i$ $\quad i = 2, 3$
so $w_i = -v \log(1-u_i)$
Then $W = w_1 + w_2$.

V & W are not independent — correlation is not specified

(3b) Problems with ranges $0<v<1$, $0<w<\infty$
but normal range in $-\infty$ to ∞
Normals are symmetric but
Beta$(\alpha, 1)$ $[\alpha \neq 1]$ and Gamma$(2, v)$
are skewed.

Part 6 method might be more efficient
[3 uniform values] if Part 7 is implemented
simply [2 uniform values per normal deviate].

You may use the following two facts:

- If $Y_1, \ldots, Y_n$ are a random sample with a common cumulative distribution function (cdf) given by $F(y) = P(Y_1 \le y)$, $y \in \mathbb{R}$, then the cdf of $Y_{(1)} = \min_{1\le i\le n} Y_i$ and $Y_{(n)} = \max_{1\le i\le n} Y_i$ are given by

$$P(Y_{(1)} \le y) = 1 - [1 - F(y)]^n \quad \& \quad P(Y_{(n)} \le y) = [F(y)]^n, \quad y \in \mathbb{R}.$$

- The probability density function (pdf) of an exponential(1) random variable V is given by

$$f(v) = e^{-v}, \text{ if } v \ge 0 \qquad \& \qquad f(v) = 0, \text{ if } v < 0,$$

and the mean and variance of V equal 1.

For all following questions: Suppose $X_1, \ldots, X_n$, $n > 1$ are a random sample with common probability density function (pdf) and mean given by

$$f(x|\theta) = \begin{cases} e^{\theta - x} & x \ge \theta \\ 0 & \text{otherwise,} \end{cases} \qquad \theta > 0, \qquad \& \qquad \textit{mean } \mathrm{E}_\theta(X_i) = 1 + \theta.$$

Let $X_{(1)} = \min_{1\le i\le n} X_i$ and $X_{(n)} = \max_{1\le i\le n} X_i$. You may *assume* $X_{(1)}$ is complete.

(a) Is the pdf $f(x|\theta)$ an exponential family model? Explain.

(b) Show that the method of moments estimator $\tilde{\theta}_n$ of θ is mean-squared error consistent.

(c) Is the two-dimensional statistic $T = \{X_{(1)}, X_{(n)}\}$ minimally sufficient for θ? Explain.

(d) Show that $n(X_{(1)} - \theta)$ has an exponential(1) distribution.

(e) Noting that $\mathrm{E}_\theta(X_2) = 1 + \theta$, find the conditional expectation $\mathrm{E}(X_2|X_{(1)})$.
(*Hint:* do not attempt to directly use any pdf for this).

(f) For a fixed $\theta_0 > 0$, find the likelihood ratio test statistic $\lambda_n(\theta_0) \equiv \lambda(X_1, \ldots, X_n, \theta_0)$ for testing $H_0 : \theta = \theta_0$ vs. $H_0 : \theta \ne \theta_0$.

(g) Invert a likelihood ratio test, based on the test statistic in (f), to find a confidence interval for θ with confidence coefficient of $(1 - \alpha)$, for $\alpha \in (0, 1)$.

(h) When using an exponential(1) prior for θ, show that the posterior pdf of θ is

$$f(\theta|x_1, \ldots, x_n) = \begin{cases} \dfrac{(n-1)e^{\theta(n-1)}}{e^{x_{(1)}(n-1)} - 1} & 0 \le \theta \le x_{(1)} \\ 0 & \text{otherwise} \end{cases}$$

(i) Based on the prior for θ in (h), identify the endpoints L and U in the highest posterior density credible interval $[L, U]$ for θ of level $1 - \alpha$, for a given $\alpha \in (0, 1)$.

(a) No, the support of $f(x|\theta)$ depends on θ.

(b) The method of moments estimator $\tilde{\theta}_n = \bar{X}_n - 1$ is unbiased $\mathrm{E}_\theta(\tilde{\theta}_n) = \mathrm{E}_\theta(X_1) - 1 = \theta$ and $\mathrm{Var}_\theta(\tilde{\theta}_n) = \mathrm{Var}_\theta(\bar{X}_n) = \mathrm{Var}_\theta(X_1)/n \to 0$ as $n \to \infty$, since $\mathrm{Var}_\theta(X_1) < \infty$.

(c) No. Using the indicator function $I\{\cdot\}$, we may write the joint pdf as

$$f(\mathbf{x}|\theta) = \underbrace{e^{-\sum_{i=1}^n x_i}}_{h(\mathbf{x})} \cdot \underbrace{e^{n\theta} I\{x_{(1)} \geq \theta\}}_{g(x_{(1)},\theta)}, \quad \forall \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

By the factorization theorem, $X_{(1)}$ is sufficient. Since T is larger than $X_{(1)}$ by including $X_{(n)}$, T is not minimally sufficient. Also, $X_{(1)}$ is sufficient and (by assumption) complete, so $X_{(1)}$ must be minimally sufficient.

(d) The cdf of X_1 is $F(x) = P(X_1 \leq x) = 1 - e^{x-\theta}$ for $x \geq \theta$ ($F(x) = 0$ for $x < \theta$). Hence, for $x < 0$, $P(n(X_{(1)} - \theta) \leq x) = 0$. For $x \geq 0$, $P(n(X_{(1)} - \theta) \leq x) = P(X_{(1)} \leq n^{-1}x + \theta) = 1 - [1 - F(n^{-1}x + \theta)]^n = 1 - e^{-x}$. The cdf of $n(X_{(1)} - \theta)$ matches the cdf of an exponential(1).

(e) $X_{(1)}$ is sufficient and (by assumption) complete. Therefore, $\mathrm{E}(X_2|X_{(1)})$ is the uniform minimum variance unbiased estimator (UMVUE) for $\mathrm{E}_\theta(X_2) = 1 + \theta$ (Lehmann-Scheffé theorem). By part (c), $\mathrm{E}_\theta\{n(X_{(1)} - \theta)\} = 1 \Rightarrow \mathrm{E}_\theta(X_{(1)}) = n^{-1} + \theta$. So, $T^* = X_{(1)} + 1 - n^{-1}$ is unbiased for estimating $1 + \theta$ and a function of the complete/sufficient statistic $X_{(1)}$. That is, $T^* = \mathrm{E}(X_2|X_{(1)})$ is the UMVUE of $\mathrm{E}_\theta(X_2) = 1 + \theta$.

(f) Fix data $\mathbf{x} = (x_1, \ldots, x_n)$ and note the parameter space is $\Theta = (0, \infty)$. By part (c), the likelihood function $L(\theta) \equiv f(\mathbf{x}|\theta)$ equals 0 for $\theta > x_{(1)}$ and $L(\theta) \equiv f(\mathbf{x}|\theta) = e^{n\theta - \sum_{i=1}^n x_i} > 0$ for $\theta \in (0, x_{(1)}]$. Since $L(\theta)$ is increasing on $(0, x_{(1)}]$, the maximum likelihood estimator is $\hat{\theta}_n = x_{(1)}$. The likelihood ratio statistic is then

$$\lambda_n(\theta_0) = L(\theta_0)/L(\hat{\theta}_n) = e^{-n(x_{(1)} - \theta_0)} I\{\theta_0 \leq x_{(1)}\}.$$

(g) The size α likelihood ratio test (LRT) rejects H_0 if $\lambda_n(\theta_0) < k$ where $k \in (0,1)$ satisfies

$$\alpha = P_{\theta_0}(\lambda_n(\theta_0) < k) = P_{\theta_0}(n(X_{(1)} - \theta_0) > -\log(k)) = 1 - \int_0^{-\log(k)} e^{-x}dx = e^{\log(k)} = k.$$

So the acceptance region of the LRT of size α for $H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$ is $A(\theta_0) \equiv \{\mathbf{x} : \lambda_n(\theta_0) \geq \alpha\}$ and a $1 - \alpha$ confidence interval for θ is $\{\theta_0 > 0 : \lambda_n(\theta_0) \geq \alpha\}$. Since $\lambda_n(\theta_0) = 0$ for $\theta_0 > x_{(1)}$, the confidence interval is

$$\{\theta_0 > 0 : \lambda_n(\theta_0) \geq \alpha\} = \{\theta_0 \in (0, x_{(1)}] : e^{-n(x_{(1)} - \theta_0)} \geq \alpha\} = [n^{-1}\log(\alpha) + x_{(1)}, x_{(1)}]$$

(h) The posterior pdf $f(\theta|\mathbf{x})$ of θ is proportional to the product of the likelihood and the prior: that is, $f(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)e^{-\theta}I\{\theta \geq 0\} = e^{\theta(n-1) - \sum_{i=1}^n x_i} I\{0 \leq \theta \leq x_{(1)}\}$. So,

$$f(\theta|\mathbf{x}) = \begin{cases} \dfrac{e^{\theta(n-1) - \sum_{i=1}^n x_i}}{\int_0^{x_{(1)}} e^{\theta(n-1) - \sum_{i=1}^n x_i} d\theta} = \dfrac{(n-1)e^{\theta(n-1)}}{e^{x_{(1)}(n-1)} - 1} & 0 \leq \theta \leq x_{(1)} \\ 0 & \text{otherwise} \end{cases}$$

(i) The posterior pdf $f(\theta|\mathbf{x})$ is increasing in $\theta \in [0, x_{(1)}]$ (and $f(\theta|\mathbf{x})$ is otherwise zero) and we want a set $\{\theta > 0 : f(\theta|\mathbf{x}) > C\}$ for a $C > 0$. The highest posterior density credible interval for θ of size $1 - \alpha$ must then be of the form $[c, x_{(1)}]$ for some $0 < c < x_{(1)}$ where $1 - \alpha = \int_c^{x_{(1)}} f(\theta|\mathbf{x})d\theta = \{e^{(n-1)x_{(1)}} - e^{c(n-1)}\}/\{e^{(n-1)x_{(1)}} - 1\}$ or $e^{c(n-1)} = \alpha e^{(n-1)x_{(1)}} + (1 - \alpha)$ or $c = (n-1)^{-1}\log\{\alpha e^{(n-1)x_{(1)}} + (1 - \alpha)\}$.