PhD Prelim Exam
THEORY
(Majors and Co-majors)

Summer 2008
(Given on 7/10/08)

1. (a) Define the following:
   (i) A probability space $(\Omega, \mathcal{F}, P)$.
   (ii) A real valued random variable X on $(\Omega, \mathcal{F}, P)$.
   (iii) Expectation of $f(X)$ with respect to P where $f : R \to R$ is a Borel measurable function.
   (iv) A sequence $\{X_n\}_{n\geq 1}$ of real valued random variables on $(\Omega, \mathcal{F}, P)$ being independent with respect to P.

   (b) Show that if there exist events $A_1, A_2, \ldots, A_k$, (k a finite positive integer) on a probability space $(\Omega, \mathcal{F}, P)$ that are independent with respect to P and $0 < P(A_i) < 1$ for all $i = 1, 2, \ldots k$, then Ω must have at least 2^k elements. Can this lower bound be achieved?

2. (a) A real valued random variable X is called *continuous* if $P(X = x) = 0$ for all x in R.
   (i) Give an example of a random variable X that is continuous but whose distribution is not *absolutely continuous.*
   (ii) Let X be a continuous random variable and $A \in \mathcal{B}(R)$ be such that $\mu_X(A) \equiv P(X \in A) > 0$. Show that for each $0 < \alpha < 1$, there is a set $B \subset A$, $B \in \mathcal{B}(R)$ such that $P(X \in B) = \alpha P(X \in A)$.
   (iii) Using (ii) conclude that there exists a sequence of disjoint sets $\{B_n\}_{n\geq 1}$ such that for each n, $B_n \subset A$ and $P(X \in B_n) > 0$.
   (iv) Conclude that there is a Borel measurable function $f : R \to R$ such that $f \geq 0$ on A and $f = 0$ on A^c and $E\left(f(X)I_A(X)\right) = \infty$. [Here $I_A(x)$ is the indicator function of the set A, i.e. $I_A(x)$ takes the value 1 if $x \in A$ and takes the value 0 otherwise.]

   (b) Let X and Y be random variables such that
   (i) X is continuous
   (ii) for any Borel measurable function $f : R \to R$

$$E|f(Y)| < \infty \Rightarrow E|f(X)| < \infty .$$

   Show that the probability measure $\mu_X(\cdot) \equiv P(X^{-1}(\cdot))$ is *dominated* by the probability measure $\mu_Y(\cdot) \equiv P(Y^{-1}(\cdot))$ on $(R, \mathcal{B}(R))$.

3. (a) State the following inequalities for random variables on a probability space:
   (i) Markov, (ii) Chebychev, (iii) Jensen, (iv) Hölder, (v) Minkowski.

   (b) Let X_1 and X_2 be two random variables such that (X_1, X_2) and (X_2, X_1) have the same distribution. Suppose that $P(X_1 = 0) = 0$. Show that $E|\frac{X_1}{X_2}| \geq 1$ with equality holding if and only if $P(X_1 = X_2) = 1$. (Hint: Use the Cauchy-Schwartz inequality.)

   (c) Using (b) above or otherwise, conclude that for any set of positive numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$,

$$\sum_{\substack{i=1 \\ j=1}}^{n} \frac{\alpha_i}{\alpha_j} \geq n^2$$

   with equality attained if and only if all α_is are equal.

4. (a) State (i) the Borel-Cantelli lemmas and (ii) Etemadi's strong law of large numbers.

   (b) Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that for $j = 0, 1, 2$, the variables $\{X_{3k+j}\}_{k\geq 1}$ are iid with cdf F_j. For $n \geq 1$, let

$$G_n(x) \equiv \frac{1}{n}\sum_{j=1}^{n} I(X_j \leq x) \ .$$

   Show that for each x in R, $\lim_{n\to\infty} G_n(x) \equiv G(x)$ exists with probability 1. Can the convergence be made uniform over R?

5. You may use the following result for this problem: If $\{\delta_i\}_{i\geq 1}$ are iid *Bernoulli* $\left(\frac{1}{2}\right)$ random variables then $U \equiv \sum_{i=1}^{\infty} \frac{\delta_i}{2^i}$ is a $Uniform[0,1]$ random variable.

   Let U be as defined above. Show that given a sequence $\{F_n\}_{n\geq 1}$ of cdf's on R there exist Borel measurable functions $f_n : R \to R$ such that $\{X_n \equiv f_n(U)\}_{n\geq 1}$ is an independent sequence with X_n having cdf F_n for each n.

   (Hint: Consider $U_n \equiv \sum_{i=1}^{\infty} \frac{\delta_{n_i}}{2^i}$, $n \geq 1$ where $n_i = (p_n)^i$ for $i = 1, 2, \ldots$ and $\{p_n\}_{n\geq 1}$ is the enumeration of the primes, i.e. $\{p_n\}_{n\geq 1} = \{2, 3, 5, 7, 11, 13, 17, 19, \ldots\}$.)

1. a) i – iv) Book work. All four parts get 3 points each.

b) For $\tilde{\delta}_k = (\delta_1, \delta_2, \ldots, \delta_k)$, $\delta_i \in \{0,1\}$

let $B_{\tilde{\delta}_k} = \bigcap_{i=1}^{k} A_i^{\delta_i}$ where $A_i^1 = A_i$, $A_i^0 = A_i^c$

Then if $\tilde{\delta}_k^{(1)} \neq \tilde{\delta}_k^{(2)}$ then $B_{\tilde{\delta}_k^{(1)}} \cap B_{\tilde{\delta}_k^{(2)}} = \phi$

Also $P(B_{\tilde{\delta}_k}) = \prod_{i=1}^{k} P(A_i^{\delta_i}) > 0$.

Since $\{B_{\tilde{\delta}_k} : \tilde{\delta}_k \in \{0,1\}^k\}$ are all disjoint and nonempty, Ω must have at least 2^k ~~$= |\{0,1\}^k|$~~ elements since $\{0,1\}^k$ has 2^k elements.

This lower bound can be achieved by defining $\Omega = \{0,1\}^k$, $A_i = \{\delta_i = 1\}$, and setting $0 < p_i = P(\delta_i = 1) < 1$.

2 a) i) Let $X = \sum_{i=1}^{\infty} \frac{\delta_i}{2^i}$

where $\{\delta_i\}_{i\geq 1}$ are i.i.d. Bernoulli (p)

with $p \neq 1/2$, $0 < p < 1$.

Then X has a continuous distribution

Since if $x \equiv \sum_{i=1}^{\infty} \frac{\eta_i}{2^i}$

where $\{\eta_i\}_{i\geq 1} \in \{0,1\}^{\infty}$

Then $\{X = x\}$ implies $\{\delta_i\}_{i\geq 1}$ ~~has~~ can equal

atmost two possible sequences. But each

such sequence has probability zero

Next, by SLLN ~~the set $B \equiv \{\{\delta_i\} \ni \frac{1}{n}\sum_1^n \delta_i \to p\}$~~

has $P(X \in B) = 1$ where $B \equiv \{\{\eta_i\} \ni \frac{1}{n}\sum_1^n \eta_i \to p$ as $n \to \infty\}$

But $m(B) = 0$ where $m(\cdot)$ is Lebesgue measure

So X has a singular distribution.

2a ii)

Fix $A \in \mathcal{B}(R) \ni P(X \in A) > 0$.

Let $\varphi_A(x) \equiv \dfrac{P(X \in A, \{X \le x\})}{P(X \in A)}$, $x \in R$.

Then since $P(X = x) = 0 \quad \forall x \in R$

$\varphi_A(\cdot)$ is a continuous function.

Also $\varphi_A(x) \uparrow 1$ as $x \uparrow \infty$

$\downarrow 0$ as $x \downarrow -\infty$

So for $\forall\ 0 < \alpha < 1$, $\exists\ x_\alpha \ni$

$$\varphi_A(x_\alpha) = \alpha$$

ie $P(A \cap \{X \le x_\alpha\}) = \alpha$.

Take $B_\alpha = A \cap \{X \le x_\alpha\}$.

2 iii) Fix $A \in \mathcal{B}(R) \ni P(X \in A) > 0$.

For $\alpha = 1/2$ let B_1 be $\ni P(X \in B_1) = \frac{1}{2} P(X \in A)$.

Let $A_1 = A \cap B_1^c$. Then $P(X \in A_1) > 0$.

Apply (ii) to get $B_2 \ni P(X \in B_2) = \frac{1}{2} P(X \in A_1)$.

Let $A_2 = A_1 \cap B_2^c$. Apply (ii) to get $B_3 \ni$

$P(X \in B_3) = \frac{1}{2} P(X \in A_2)$.

This generates disjoint sets $\{B_n\}_{n \ge 1} \subset A \ni$

$P(X \in B_n) = \frac{1}{2^n}$, $n \ge 1$.

2 iv)

Gn: ~~$\exists f \geq 0$ on A, $\Rightarrow 0$~~

Let $f(x) = 2^n$ on B_n $n \geq 1$

$= 0$ on $\left(\bigcup_{n\geq 1} B_n\right)^c$

Then $E f(X) = \sum 2^n P(B_n) = \sum 1 = \infty$.

b) Gn: 1) ~~X cont~~, X, Y r.v. X cont

2) $E|f(Y)| < \infty \Rightarrow E|f(X)| < \infty$.

To show $\mu_X(\cdot) = P(X^{-1}(\cdot))$ is dominated by $\mu_Y(\cdot) = P(Y^{-1}(\cdot))$

Suppose not. Then $\exists A \in \mathcal{B}(\mathbb{R}) \ni$

$\mu_Y(A) = 0$ but $\mu_X(A) > 0$.

Since X is cont + $\mu_X(A) > 0$, by a (iv)

$\exists$ $\left. \begin{array}{l} f \geq 0 \text{ on } A \\ = 0 \text{ on } A^c \end{array} \right\} \ni$ $E(f(X) : X \in A) = \infty$

But $E|f(Y)| = E f(Y) = 0$

since $\mu_Y(A) = 0$.

Thus $E|f(Y)| < \infty$ but $E|f(X)| = \infty$

a contradiction.

3 a) Bookwork

b) Let $Y = \left|\frac{X_1}{X_2}\right|$. Then Y and $\frac{1}{Y}$ have the same distribution. Also $P(0<Y<\infty)=1$.

For any $0<y<\infty$, $\left(y+\frac{1}{y}\right) \geq 2$

since $y^2-2y+1 = (y-1)^2 \geq 0$, $=0$ iff $y=1$.

Thus $Y+\frac{1}{Y}-2$ is a nonnegative r.v.

Thus $E\left(Y+\frac{1}{Y}\right) \geq 2$, $=2$ iff $P(Y=1)=1$.

But $E\frac{1}{Y} = EY$. So

$2EY \geq 2$, ie $EY \geq 1$, $=1$ iff $P(Y=1)=1$.

c) For $\alpha_1, \alpha_2, \ldots, \alpha_n > 0$

$$\frac{1}{n^2}\sum_{\substack{i=1\\j=1}}^{n} \frac{\alpha_i}{\alpha_j} = E\frac{X_1}{X_2},$$

where X_1, X_2 are iid $\sim \left\{\begin{matrix}\alpha_1 & \ldots & \alpha_n \\ \frac{1}{n}, & \ldots & \frac{1}{n}\end{matrix}\right.$

So follows from b)

4. a) Book work.

c) For any n $\exists$ k $\ni$ $3k \leq n < 3(k+1)$

If $n = 3k+r$, $r = 0,1,2$

$$G_n(x) = \frac{3k}{n} G_{3k}(x) + \frac{1}{n}\left(I(X_{3k+1} \leq x) + I(X_{3k+2} \leq x)\right)$$

$$\sup_x \left| G_n(x) - G_{3k}(x) \right| \le \frac{4}{n}.$$

Also $$G_{3k}(x) = \frac{1}{3}\sum_{r=0}^{2} \frac{1}{k} \sum_{\substack{j=3l+r \\ l=0}}^{k-1} I(X_j \le x)$$

By Glivenko Cantelli

$\forall\ r = 0, 1, 2$

$$\sup_x \left| \frac{1}{k} \sum_{\substack{j=3l+r \\ l=0}}^{k-1} I(X_j \le x) - F_r(x) \right| \to 0 \quad \text{wp1}$$

Thus $$\sup_x \left| G_n(x) - \frac{1}{3}\sum_{r=0}^{2} F_r(x) \right| \to 0. \quad \text{wp1.}$$

Yes, the convergence is uniform.

5. For each j, the sequence

$\{ \delta_{ji} \equiv \delta_{p_j^i} \}_{i \ge 1}$ are iid Bernoulli(1/2)

and so U_j is a Uniform $(0,1)$ r.v.

Also $\{U_j\}_{j \ge 1}$ are iid

Let $X_j \equiv F_j^{-1}(U_j)$

where $F_j^{-1}(x) \equiv \inf\{y : F_j(y) \ge x\}$

Then $\forall$ each j, X_j has cdf F_j.

Also $\{X_j\}_{j \ge 1}$ are independent.

1. Let $X_1, \ldots, X_n$ be a simple random sample (i.e., iid random observations) from a Poisson density with mean (and variance) θ:

$$f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}, \quad \theta > 0, \ \ x = 0, 1, \ldots \tag{1}$$

Let $\tau(\theta) = P_\theta(X_1 = 0) = e^{-\theta}$ be the parameter of interest.

(a) Find the maximum likelihood estimator (MLE) of $\tau(\theta)$, denoted as $\hat{\tau}(\theta)$. Is the MLE unbiased? (Explain why or why not.)

(b) Find the asymptotic distribution of the MLE $\hat{\tau}(\theta)$ you identified in part (a). That is,

$$\sqrt{n}(\hat{\tau}(\theta) - \tau(\theta))$$

has what limiting distribution as $n \to \infty$? (Identify the parameters of the limiting distribution.)

Suppose we are interested in the mean θ of the above Poisson density in (1), and $\theta > 0$ has a gamma prior $\Gamma(1, 1/2)$ with a density $\rho(\theta) = 2e^{-2\theta}$.

(c) Derive the posterior distribution of θ.

(d) Derive the Bayes estimator of θ under squared-error loss, $\bar{\theta}_n$.

(e) Comment on the admissibility of the Bayes estimator $\bar{\theta}_n$.

2. Suppose X_1, X_2, and X_3 are iid Bernoulli random variables, i.e., Bernoulli(θ), $\theta \in \Theta = [0, 1]$. We are interested in estimating θ^2.

(a) What is the MLE of θ^2? Is it unbiased? (Explain why or why not.)

(b) Find a sufficient and complete statistic for θ and derive the UMVUE of θ^2, $\tilde{\theta}^2_{UMVUE}$.

(c) Suppose $\sum_{i=1}^3 X_i = 1$. What is the value of $\tilde{\theta}^2_{UMVUE}$? Is $\tilde{\theta}^2_{UMVUE}$ a reasonable estimator of θ^2? (Explain why or why not.)

(d) Calculate $P_\theta(\tilde{\theta}^2_{UMVUE} = 0)$.

(e) Suppose θ has a prior distribution uniform on $[0, 1]$. Derive the Bayes estimator for θ^2 under squared-error loss.

3. Consider a simple mixture model $\mathcal{P} = \{g_\theta(x) : \theta \in [0,1], x \in (-\infty, +\infty)\}$ for

$$g_\theta(x) = \theta f_1(x) + (1-\theta) f_0(x),$$

where θ is the mixing proportion and f_0 and f_1 are two different pdf's.

Suppose X is a single random observation from the mixture probability density $g_\theta(x)$.

(a) Let U and V be two random variables with pdf's f_0 and f_1, respectively. Find the mean and variance of X in terms of the first and second moments of U and V.

(b) Derive the Fisher information in X for θ evaluated at $\theta_0 \in (0,1)$ and show that it is bounded above by $[\theta_0(1-\theta_0)]^{-1}$. Explain when this upper bound might be achieved.

Suppose $X_1, X_2, \cdots, X_n$ are iid random observations from the mixture density $g_\theta(x)$, where $f_0(x)$ and $f_1(x)$ are positive for all $x \in (-\infty, +\infty)$.

(c) Show that the likelihood equation based on $X_1, X_2, \cdots, X_n$ has a unique root if and only if

$$\frac{1}{n}\sum_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} > 1 \text{ and } \frac{1}{n}\sum_{i=1}^{n} \frac{f_0(X_i)}{f_1(X_i)} > 1.$$

(d) Argue that the unique root of the likelihood equation, when it exists, is the MLE of θ.

(e) What is the MLE of θ when the likelihood equation has no root?

(f) Find a $\sqrt{n}$-consistent moment estimator of θ.

(g) Using a $\sqrt{n}$-consistent moment estimator as a starting value, specify an iterative algorithm to numerically compute the MLE when it is the unique root of the likelihood equation.

1. (a) The MLE of θ is the sample mean $\overline{X}$, which is unbiased. By the invariance property of the MLE, the MLE of $\tau(\theta) = e^{-\theta}$ is $\hat{\tau}(\theta) = \exp(-\overline{X})$, which is biased by Jensen's inequality.

   (b) The MLE of θ is the sample mean $\overline{X}$, which is unbiased. By the CLT,
$$\sqrt{n}(\overline{X} - \theta) \ \to_d \ N(0, \sigma_X^2).$$
   Note that for the Poisson(θ) density, it is well known that
$$EX = Var(X) = \theta.$$
   Thus
$$\sqrt{n}(\overline{X} - \theta) \ \to_d \ N(0, \theta).$$
   Since $\tau'(\theta) = -e^{-\theta}$, by the delta method, the limiting distribution of $\hat{\tau}(\theta)$ is normal with mean 0 and variance
$$\frac{[\tau'(\theta)]^2}{I(\theta)} = \theta e^{-2\theta}.$$
   That is,
$$\sqrt{n}(e^{-\overline{X}} - e^{-\theta}) \ \to_d \ N(0, \theta e^{-2\theta}).$$

   (c) Let $T = \Sigma X_i$ and $\mathbf{X} = (X_1, \ldots, X_n)$. The likelihood function is
$$\ell(\mathbf{X}|\theta) = \frac{\theta^T}{\prod_i X_i!} e^{-n\theta}.$$
   Since θ has a gamma prior $\rho(\theta) = 2e^{-2\theta}$, the posterior distribution of θ is
$$\pi(\theta|\mathbf{X}) = \frac{\ell(\mathbf{X}|\theta)\rho(\theta)}{\int \ell(\mathbf{X}|\theta)\rho(\theta)d\theta} = \frac{\theta^T (n+2)^{T+1}}{\Gamma(T+1)} e^{-(n+2)\theta} \sim \Gamma(T+1, 1/(n+2)).$$

   (d) Recall that for Gamma(α, β), the mean is $\alpha\beta$, and the mode is $(\alpha - 1)\beta$. Thus the mean of the posterior gamma distribution, which minimizes the Bayes risk with squared error loss (i.e., $L(\theta, \delta) = (\theta - \delta)^2$), is
$$\bar{\theta}_n = \frac{T+1}{n+2}.$$

   (e) Note that $\bar{\theta}_n$ is the unique Bayes estimator under a proper prior density, and is thus admissible.

2. (a) Since $f(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, $i = 1, 2, 3$, the MLE of θ is $\hat{\theta} = \sum_{i=1}^3 x_i/3$. By the invariance of MLE, the MLE of θ^2 is $\hat{\theta}^2 = [\sum_{i=1}^3 x_i/3]^2$.

Since the MLE of θ, $\hat{\theta}$, is unbiased, by Jensen's inequality, the MLE of θ^2, $\hat{\theta}^2$, has a positive bias. In fact,

$$E[\sum_{i=1}^{3} x_i/3]^2 = (\theta + 2\theta^2)/3 > \theta^2$$

unless $\theta = 0$ or 1.

(b) Since the Bernoulli family is an exponential family, $\sum X_i$ is complete and sufficient. Note that $E\sum_{i=1}^3 X_i = 3\theta$, $E[\sum_{i=1}^3 X_i]^2 = 3\theta + 6\theta^2$, and $E(\frac{\sum_{i=1}^3 X_i)^2 - \sum_{i=1}^3 X_i}{6}) = \theta^2$. Thus $\frac{1}{6}(\sum_{i=1}^3 X_i)^2 - \frac{1}{6}\sum_{i=1}^3 X_i$ is the UMVUE of θ^2.

(c) If there is exactly one success in these three observations, $\tilde{\theta}^2_{UMVUE} = \frac{1}{6}\cdot 1^2 - \frac{1}{6}\cdot 1 = 0$. Here, the UMVUE being 0 is not reasonable, because if $\theta^2 = 0$, the sum of X_i's should always equal 0.

(d) We have $P_\theta(\tilde{\theta}^2_{UMVUE} = 0) = P(\sum_{i=1}^3 X_i = 1 \text{ or } 0) = 3\theta(1-\theta)^2 + (1-\theta)^3$, which is positive for all $\theta \neq 1$.

(e) Let $\mathbf{X} = (X_1, X_2, X_3)$. When θ has a uniform prior, the joint distribution of $(\mathbf{X}, \theta)$ is

$$\theta^{\sum_{i=1}^3 x_i}(1-\theta)^{3-\sum_{i=1}^3 x_i} \cdot I\{0 \leq \theta \leq 1\}$$

Thus the posterior distribution of $\theta|\mathbf{X}$ would be $Beta(\sum_{i=1}^3 X_i + 1, 4 - \sum_{i=1}^3 X_i)$. Then under squared-error loss, the Bayes estimator of θ^2 would be

$$\begin{aligned}\tilde{\theta}^2_{Bayes} &= E_{\theta^2|\mathbf{X}}(\theta^2)\\ &= E_{\theta|\mathbf{X}}(\theta^2)\\ &= \frac{(\sum_{i=1}^3 X_i + 1)(4 - \sum_{i=1}^3 X_i)}{25 \times 6} + \frac{(\sum_{i=1}^3 X_i + 1)^2}{25}\end{aligned}$$

As we can see, $\tilde{\theta}^2_{Bayes} \in (0,1)$ with probability 1.

3. (a) By definition,

$$EX = \int x g_\theta(x)dx$$

$$= \theta \int x f_1(x)dx + (1-\theta)\int x f_0(x)dx$$

$$EX^2 = \int x^2 g_\theta(x)dx$$

$$= \theta \int x^2 f_1(x)dx + (1-\theta)\int x^2 f_0(x)dx$$

Thus

$$\text{Var}(X) = EX^2 - (EX)^2.$$

(b) The Fisher information in X for θ evaluated at any point θ_0 is

$$I_X(\theta_0) = \frac{1}{\theta_0(1-\theta_0)}\left[1 - \int_{-\infty}^{+\infty} \frac{f_1(x)f_0(x)}{\theta_0 f_1(x) + (1-\theta_0)f_0(x)}\,dx\right],$$

which is bounded above by $\frac{1}{\theta_0(1-\theta_0)}$. This upper bound is achieved when $f_0(x)$ and $f_1(x)$ have disjoint support sets, i.e., $f_0(x)f_1(x) = 0$ with probability 1. This upper bound is not achieved when f_0 and f_1 are the normal densities of $N(0,1)$ and $N(1,1)$.

(c) The likelihood equation has at most one root because it is easy to show that the likelihood function is concave. Let $\mathbf{X} = (x_1, \ldots, x_n)$. We have

$$\begin{aligned} L(\theta|\mathbf{X}) &= \Pi_{i=1}^n g_\theta(x_i) \\ &= \Pi_{i=1}^n[\theta f_1(x_i) + (1-\theta)f_0(x_i)] \\ \ell(\theta|\mathbf{X}) &= \sum_{i=1}^n \log(\theta f_1(x_i) + (1-\theta)f_0(x_i)). \end{aligned}$$

The likelihood equation is

$$\frac{d\ell(\theta|\mathbf{X})}{d\theta} = \sum_{i=1}^n \frac{f_1(x_i) - f_0(x_i)}{\theta f_1(x_i) + (1-\theta)f_0(x_i)} = 0. \tag{1}$$

Note that

$$\frac{d^2\ell(\theta|\mathbf{X})}{d\theta^2} = \sum_{i=1}^n \frac{-(f_1(x_i) - f_0(x_i))^2}{(\theta f_1(x_i) + (1-\theta)f_0(x_i))^2} < 0 \ \forall\theta. \tag{2}$$

So the likelihood equation has at most one solution. Thus

$$\frac{d\ell(\theta|\mathbf{X})}{d\theta}|_{\theta=0} > 0 \text{ and } \frac{d\ell(\theta|\mathbf{X})}{d\theta}|_{\theta=1} < 0$$

are the sufficient and necessary conditions for the existence of a solution, i.e.,

$$\frac{1}{n}\sum_{i=1}^n \frac{f_1(x_i)}{f_0(x_i)} > 1 \ \text{ and } \ \frac{1}{n}\sum_{i=1}^n \frac{f_0(x_i)}{f_1(x_i)} > 1$$

(d) The unique root of the likelihood equation, when it exists, is the MLE because of the concavity of the likelihood function.

(e) From part 3 (c), when the likelihood function has no root, there are two possibilities:
case (i):

$$\frac{d\ell(\theta|\mathbf{X})}{d\theta}|_{\theta=0} > 0 \text{ and } \frac{d\ell(\theta|\mathbf{X})}{d\theta}|_{\theta=1} > 0$$

The likelihood function is increasing, and thus the MLE is $\hat{\theta} = 1$.
case (ii):

$$\frac{d\ell(\theta|\mathbf{X})}{d\theta}|_{\theta=0} < 0 \text{ and } \frac{d\ell(\theta|\mathbf{X})}{d\theta}|_{\theta=1} < 0$$

The likelihood function is decreasing, and thus the MLE is $\hat{\theta} = 0$.

(f) As $EX_1 = \theta$, one moment estimator of θ would be $\bar{X}$, which is $\sqrt{n}$-consistent by the CLT. In fact,

$$\sqrt{n}(\bar{X} - \theta) \to_d N(0, \text{Var}(X_1)).$$

(g) Using a $\sqrt{n}$-consistent estimator as a starting value, e.g., set$T_0 = \bar{X}$, an iterative algorithm of getting MLE can be obtained based on the Newton-Raphson algorithm. Note that equations (1) and (2) give the explicit expressions of $\ell'(\theta|\mathbf{X})$ and $\ell''(\theta|\mathbf{X})$, respectively.

(i). Set

$$T_1 = T_0 - \left.\frac{\ell'(\theta|\mathbf{X})}{\ell''(\theta|\mathbf{X})}\right|_{\theta=T_0}$$

(ii). At the k-th step, set

$$T_k = T_{k-1} - \left.\frac{\ell'(\theta|\mathbf{X})}{\ell''(\theta|\mathbf{X})}\right|_{\theta=T_{k-1}}$$

(iii). Repeat step (ii), until the algorithm converges.

One can also use the iterative Fisher scoring method to compute the MLE after deriving the Fisher information explicitly as in part 3(b).

This question concerns a model and inference using that model for so-called "circular data." These are data that take values in the interval $[-\pi,\pi)$ (and can thus be identified with points on the unit circle). A useful standard model for circular data is the von Mises distribution with "direction parameter" $\mu\in[-\pi,\pi)$ and "concentration parameter" $\kappa\geq 0$. This distribution has probability density

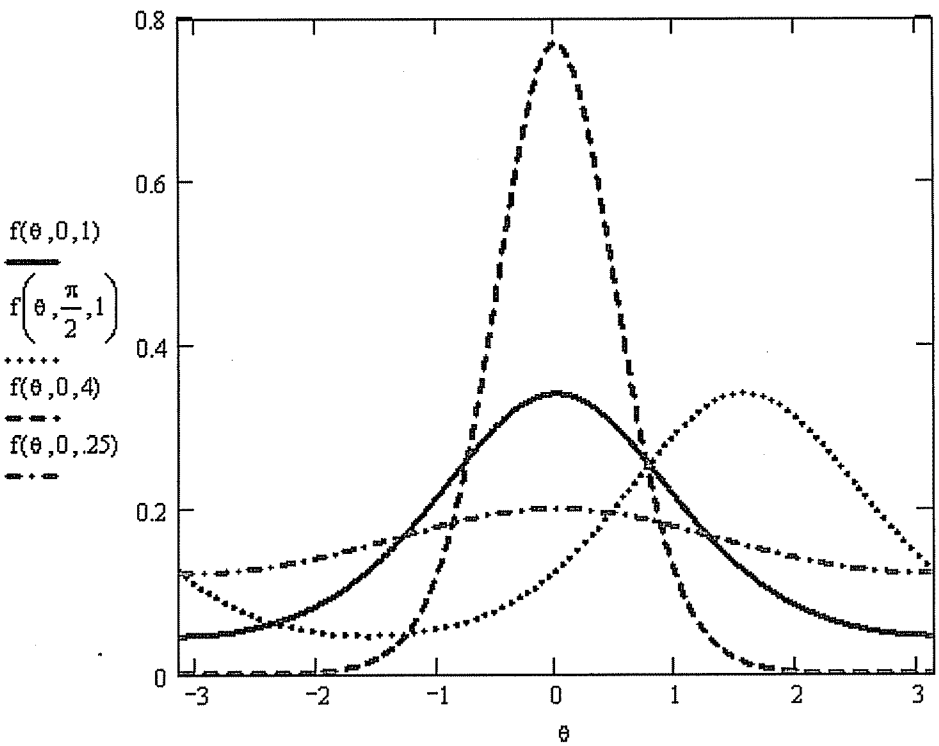
$$f(\theta\mid\mu,\kappa)=C(\kappa)\exp\left(\kappa\cos(\theta-\mu)\right)I[-\pi\leq\theta<\pi]$$

for normalizing constant

$$C(\kappa)=\left(\int_{-\pi}^{\pi}\exp(\kappa\cos\theta)d\theta\right)^{-1}$$

Plots of four von Mises probability densities are below in Figure 1.

Figure 1: Plots of Four Different von Mises Probability Densities



1) Carefully define convergence in distribution. (That is, for random variables X_n with cdfs F_n and X with cdf F, what does it mean for X_n to converge to X in distribution?) Then suppose that θ_n has the von Mises distribution with direction $\mu=0$ and concentration parameter $\kappa=\frac{1}{n}$. Argue carefully using your definition that the sequence $\{\theta_n\}$ converges in distribution. (You may use without proof the fact that if $\{g_n\}$ is a sequence of continuous functions on the finite closed interval $[a,b]$ converging to a function g and there is a positive constant A with $0\leq g_n(t)\leq A$ for all n and $t\in[a,b]$, then $\int_a^b g_n(t)dt\to\int_a^b g(t)dt$.)

2) For Y a random variable with pdf $h(y)$ on $\Re$ and positive constant c, what is the pdf of the random variable cY? Use your answer in the following. Suppose that θ_n has the von Mises distribution with direction parameter $\mu = 0$ and concentration parameter $\kappa = n$ and let $Z_n = \sqrt{n}\,\theta_n$. Argue that the pdf of Z_n, say g_n, approximates the standard normal pdf as $n \to \infty$. (Hint: It suffices to show that for large n, $\ln g_n(z) \approx d(n) - \frac{z^2}{2}$ for some $d(n)$ depending only upon n and not z. Consider Taylor's theorem.)

3) Completely describe an algorithm that you could use to simulate from the von Mises distribution with direction parameter $\mu = \pi/2$ and concentration parameter $\kappa = 1$. (Just naming such an algorithm is not sufficient. You must say exactly what needs computing and how to compute it.)

Suppose that $\theta_1, \theta_2, \ldots, \theta_n$ are iid with the von Mises distribution with direction parameter $\mu = 0$ and unknown concentration parameter $\kappa > 0$.

4) Find a minimal sufficient statistic for the parameter κ and carefully argue that it is minimal sufficient. Then argue carefully that the family of von Mises distributions with $\mu = 0$ has monotone likelihood ratio in the statistic that you identify.

5) Describe an optimal size $\alpha = .05$ test of $\mathrm{H}_0 : \kappa \ge 1$ versus $\mathrm{H}_a : \kappa < 1$ based on $\theta_1, \theta_2, \ldots, \theta_n$. Describe how you would find the appropriate cut-off value for your test using simulation. Then show how you could use numerical integration to evaluate some integrals and then a large sample normal approximation to *approximate* the cut-off value for your test.

Now consider the problem of inference for *both* κ and μ based on a sample of, say, $n = 20$ iid von Mises observations. In fact, suppose that one observes

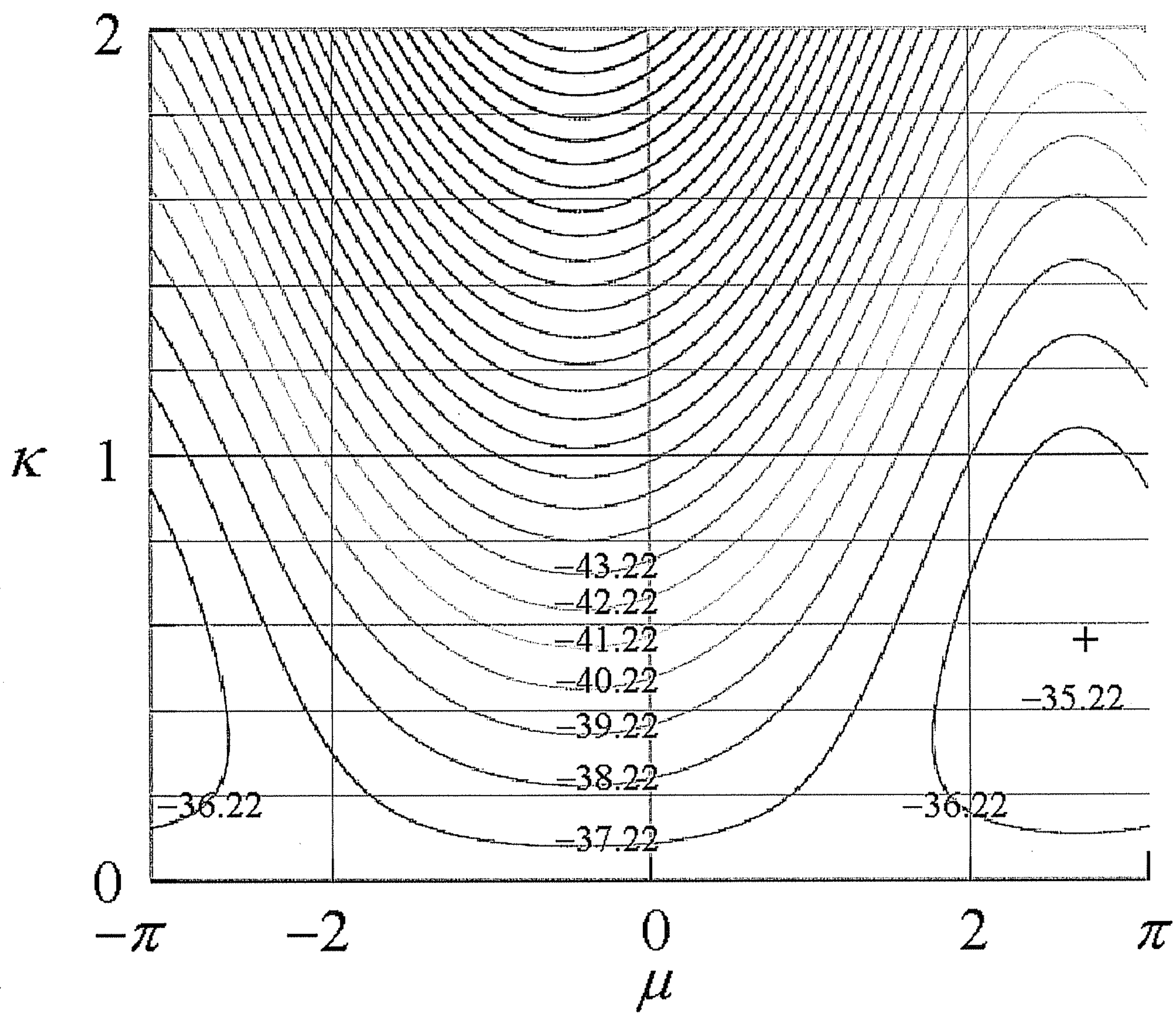
$$-3.135, -1.292, 1.605, .534, 2.405, -1.593, 2.026, 0.174, -2.542, -1.965,$$
$$3.083, -2.286, -3.094, 1.412, 1.663, -1.709, 1.080, -2.802, 2.269, 1.367$$

Figure 2 on the next page is a contour plot of the loglikelihood for this data set. This function has a maximum of about -35.22 at $(\mu, \kappa) = (2.701, .57)$. The matrix of 2nd partial derivatives of the loglikelihood at $(2.701, .57)$ is approximately

$$\begin{pmatrix} \frac{\partial^2}{\partial \mu^2}\text{loglikelihood} & \frac{\partial^2}{\partial \mu \partial \kappa}\text{loglikelihood} \\ \frac{\partial^2}{\partial \mu \partial \kappa}\text{loglikelihood} & \frac{\partial^2}{\partial \kappa^2}\text{loglikelihood} \end{pmatrix} = \begin{pmatrix} -3.128 & 0 \\ 0 & -8.882 \end{pmatrix}$$

6) Exactly what function of μ and κ has been plotted in Figure 2? Give a formula. (You may abbreviate the observed values -3.135 through 1.367 as $\theta_1, \theta_2, \ldots, \theta_{20}$.)

Figure 2: A Contour Plot of a Particular von Mises LogLikelihood Function



7) Consider the hypothesis $H_0 : (\mu, \kappa) = (0, .8)$. Does a likelihood ratio test of this hypothesis versus H_a : not H_0 (using a large sample approximation for the null distribution) reject this hypothesis at level $\alpha = .01$? Explain carefully.

8) What are a sensible point estimate for μ and a corresponding standard error for the estimate based on this likelihood function? In light of the fact that $\mu \in [-\pi, \pi)$ do these lead to an appropriate large sample confidence set for μ? Explain.

9) Identify appropriate (individual) approximate 99% confidence sets for both μ and κ based on profile loglikelihood functions corresponding to Figure 2. (Give numerical descriptions of these sets.)

Key Theory III / Statistics Prelim 2008

Note Title 6/23/2008

1) $F_n(t) \to F(t)$ at all continuity points $t \in \mathbb{R}$ (of the limit function F)

Here, take F to be the $U(-\pi,\pi)$ cdf. F is cont$^{\underline{s}}$ on all of $\mathbb{R}$.
For $t \le -\pi$ $F_n(t) = F(t) = 0$ for all n and for $t \ge \pi$ $F_n(t) = F(t) = 1$ for all n. So consider $t \in (-\pi, \pi)$.

$$F_n(t) = \frac{\int_{-\pi}^{t} \exp\left(\frac{1}{n}\cos\theta\right)d\theta}{\int_{-\pi}^{\pi} \exp\left(\frac{1}{n}\cos\theta\right)d\theta}$$

But $\frac{1}{n}\cos\theta \to 0$ $\forall\theta$. So the limit says that

$$\int_{-\pi}^{t} \exp\left(\frac{1}{n}\cos\theta\right)d\theta \longrightarrow t+\pi$$

and $$\int_{-\pi}^{\pi} \exp\left(\frac{1}{n}\cos\theta\right)d\theta \longrightarrow 2\pi$$

So, since the function $h(u,v) = \frac{u}{v}$ is cont$^{\underline{s}}$ except when $v = 0$, we have

$$F_n(t) \longrightarrow \frac{t+\pi}{2\pi} = F(t)$$

2) Y with pdf $h(y)$ and $c > 0 \Rightarrow W = cY$ has pdf

$$f(w) = \frac{1}{c} h\left(\frac{w}{c}\right)$$

So if θ_n has pdf $f(\theta|0,n) = c(n)\exp(n\cos\theta)\, I[-\pi \le \theta < \pi]$

Z_n has pdf

$$g_n(z) = \frac{1}{\sqrt{n}} c(n) \exp\left(n \cos\frac{z}{\sqrt{n}}\right) I\left[-\pi \le \frac{z}{\sqrt{n}} < \pi\right]$$

and as a function of z, for $z \in [-\pi, \pi)$

$$\ln g_n(z) = -\tfrac{1}{2}\ln n + \ln C(n) + n\cos\frac{z}{\sqrt{n}}$$

Now Taylor's theorem says that for small θ, $\cos\theta \approx 1 - \frac{\theta^2}{2}$, so for large n

$$\ln g_n(z) \approx -\tfrac{1}{2}\ln n + \ln C(n) + n - \frac{n}{2}\left(\frac{z}{\sqrt{n}}\right)^2$$
$$= \underbrace{-\tfrac{1}{2}\ln n + \ln C(n) + n}_{d(n)} - \tfrac{1}{2}z^2$$

3) There are several possibilities here. Among them are 1) using the rejection algorithm based on $U \sim U(-\pi,\pi)$ and 2) finding $F^{-1}(u)$ for $U \sim U(-\pi,\pi)$ (students need to show what equation must be solved if they follow this second option).

4) This is an exponential family where K is actually the natural parameter (belonging to the open set $(0,\infty)$). So

$$T(\underset{\sim}{\theta}) = \sum_{i=1}^{n}\cos(\theta_i)$$

is minimal sufficient. For $K_2 > K_1$

$$\frac{f(\underset{\sim}{\theta}\,|\,0,K_2)}{f(\underset{\sim}{\theta}\,|\,0,K_1)} = \left(\frac{C(K_2)}{C(K_1)}\right)^n \exp\left((K_2-K_1)\,T(\underset{\sim}{\theta})\right)$$

which is monotone in $T(\underset{\sim}{\theta})$.

5) Such a test rejects H_0 for small values of $T(\underset{\sim}{\theta})$. For a cut-off value c, the size of the test is

$$P_{K=1}\left[T(\underset{\sim}{\theta}) < c\right] = P_{K=1}\left[\sum_{i=1}^{n}\cos(\theta_i) < c\right]$$

So to find c via simulation, I could for some large number of repetitions (say R of them), simulate $\Theta_1, \Theta_2, \ldots, \Theta_n$ iid von Mises with $\mu = 0$ and $K = 1$ and compute $T(\underset{\sim}{\Theta})$. The lower 5% pt of the simulated values will serve as the cut point.

A second possibility is to use the CLT for the dsn of $T(\underset{\sim}{\Theta})$. That is,

$$T(\underset{\sim}{\Theta}) = \sum_{i=1}^{n} \cos(\Theta_i)$$

is a sum of iid values $\cos(\Theta_i)$ with means (supposing $K=1$)

$$m_1 = E_{K=1} \cos\theta = c(1) \int_{-\pi}^{\pi} (\cos\theta) \exp(\cos\theta)\, d\theta$$

and 2nd moments

$$m_2 = E_{K=1} (\cos\theta)^2 = c(1) \int_{-\pi}^{\pi} (\cos\theta)^2 \exp(\cos\theta)\, d\theta$$

m_1 and m_2 can be computed numerically and one can then use the fact that under $K=1$

$$\frac{T_n(\underset{\sim}{\Theta}) - m_1}{\sqrt{\frac{(m_2 - m_1^2)}{n}}} \xrightarrow{d} N(0,1)$$

to set $m_1 - 1.645\sqrt{\frac{m_2 - m_1^2}{n}}$ as an approximate cut-off.

6) The log likelihood is

$$\ln f(\underset{\sim}{\theta}, \mu, K) = 20 \ln(c(K)) + K \sum_{i=1}^{20} \cos(\theta_i - \mu)$$

7) The loglikelihood for this combination of parameters is less than -43.22 i.e.

$$\ln f(\underset{\sim}{\theta}, \underbrace{\hat{\mu}, \hat{\kappa}}_{\text{MLE's}}) - \ln f(\underset{\sim}{\theta}, 0, .8) \geq 8$$

So twice this difference is larger than 16. The upper 1% pt. of χ^2_2 is only 9.2. So H_0 would be rejected using the large sample χ^2 approximation to the null dsn of the likelihood ratio statistic.

8) A sensible point estimate is $\hat{\mu} = 2.701$ and since

$$-\begin{pmatrix} -3.128 & 0 \\ 0 & -8.882 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{3.128} & 0 \\ 0 & \frac{1}{8.882} \end{pmatrix} \approx \begin{pmatrix} .3197 & 0 \\ 0 & .1126 \end{pmatrix}$$

a sensible std error for this estimate is $\sqrt{.3197} = .5654$

These lead to a naive approximate/large sample 95% C.I. like

$$2.701 \pm \underbrace{1.96(.5654)}_{1.108}$$

which fails to honor the restriction that $\mu \in [-\pi, \pi)$

9) We can use for an approximate confidence set

$$S_\mu = \left\{ \mu \,\middle|\, \exists\, \kappa \text{ with } \underbrace{\ln f(\underset{\sim}{\theta} \mid \hat{\mu}, \hat{\kappa}) - \ln f(\underset{\sim}{\theta} \mid \mu, \kappa) < \tfrac{1}{2}\left(\chi^2_1\right)} \right\}$$

upper 1% pt of χ^2_1

$\frac{1}{2}(6.635) = 3.3$

and it's evident from the contour plot that every μ has a corresponding κ for which this holds. That is, $S_\mu = [-\pi, \pi)$. However, with

$$S_\kappa = \left\{ \kappa \mid \exists\, \mu \text{ with } \ln f(\underset{\sim}{\theta} \mid \hat{\mu}, \hat{\kappa}) - \ln f(\underset{\sim}{\theta} \mid \mu, \kappa) < 3.3 \right\}$$

we get something nontrivial $S_\kappa \approx [0, 1.5]$.

A random variable X that has a Gamma(α, β) distribution for $\alpha, \beta > 0$ has probability density function (pdf)

$$f_X(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta}$$

for $x > 0$ and $f_X(x) = 0$ otherwise. The Gamma(α, β) moment generating function (MGF) is $M_X(t) = (1 - \beta t)^{-\alpha}$.

Let $Y = e^X$. The random variable Y has a log Gamma(α, β) distribution.

1. What is the pdf of Y?
2. What are the mean and variance of Y? Hint: use the MGF of X.

Suppose random variables $X_i, i = 1, \ldots, n$ are iid Gamma(α, β).

3. What is the distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$?
4. Argue that $e^{\bar{X}_n}$ is a consistent estimator of $e^{\alpha\beta}$.
5. Identify a value θ such that $\sqrt{n}(e^{\bar{X}_n} - \theta)$ converges in distribution and then find the limit distribution. (Carefully specify parameters of the limit distribution.)
6. Let $Y_i = e^{X_i}$. Does $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$ converge to $e^{\alpha\beta}$? Why or why not?

Assume that λ has a log Gamma (α, β) distribution. Given a value of λ, assume W has an exponential(λ) distribution. Consequently, λ is the conditional mean of W.

7. What are the marginal mean and variance of W?

Suppose that V_1 and V_2 are independent exponential random variables with means λ_1 and λ_2, respectively.

8. Derive the probability that $V_1 + V_2 < c$. (You may leave your answer in the form of a double integral.)

A random variable X that has a Gamma(α, β) distribution has probability density function (pdf)

$$f_X(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} e^{-x/\beta}$$

for $x > 0$ ($\alpha, \beta > 0$) and $f_X(x) = 0$ otherwise. The moment generating function (MGF) is $M_X(t) = (1 - \beta t)^{-\alpha}$.

1. Let $Y = e^X$. The random variable Y has a log Gamma(α, β) distribtuion. What is the pdf of Y?

   $X = \log Y$, $dX/dY = 1/Y$,

   $$f_Y(y) = f_x(\log y)(1/y) = \frac{(\log y)^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} e^{-\log y/\beta}/y$$

   for $y > 0$.

2. What are the mean and variance of Y? Hint: use the MGF of X.

   $EY = E(e^X) = M_X(1) = (1 - \beta)^{-\alpha}$.
   $EY^2 = E((e^X)^2) = M_X(2) = (1 - 2\beta)^{-\alpha}$.
   $VY = EY^2 - (EY)^2$

   Suppose random variables $X_i, i = 1, \ldots, n$ be independent of one another. Let X_i have a Gamma(α, β) distribution.

3. What is the distribution of $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$?

   $M_{\text{sum}} = \Pi_{i=1}^{n} M_{X_i} = (1-\beta t)^{-\sum \alpha}$, so the sum is Gamma($n\alpha, \beta$). The average is then Gamma($n\alpha, \beta/n$).

4. Argue that $e^{\bar{X}_n}$ is a consistent estimator of $e^{\alpha\beta}$.

   Mean of X_i is $\alpha\beta$, variance is finite ($\alpha\beta^2$), so WLLN applies. The transformation is continuous so also have consistency.

5. State a value of θ to complete the following question, then answer the question.

   Find the limit of the distribution of $\sqrt{n}(e^{\bar{X}_n} - \theta)$ as $n \to \infty$.

   $\theta = e^{\alpha\beta}$. $\sqrt{n}(\bar{X}_n - \alpha\beta)$ has limiting distribution N(0, $\alpha\beta^2$) by CLT. If $g(x) = e^x$, $g'(x) = e^x$, so $\sqrt{n}(e^{\bar{X}_n} - e^{\alpha\beta})$ has limiting distribution N(0, $\alpha\beta^2 e^{2\alpha\beta}$).

6. Let $Y_i = e^{X_i}$ for $i = 1, \ldots, n$. Does $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ converge to $e^{\alpha\beta}$? Why or why not?

   No. That is the wrong value. It should be the mean of Y_i: see part 2: $(1-\beta)^{-\alpha}$.

7. Given a value of λ, assume W has an exponential(λ) distribution. Consequently, λ is the conditional mean of W. Assume that λ has a log Gamma (α, β) distribution. What are the mean and variance of W?

   $EW = E(E(W|\lambda)) = E(\lambda) = (1-\beta)^{-\alpha}$
   $VW = E(V(W|\lambda)) + V(E(W|\lambda)) =$
   $E(\lambda^2) + V(\lambda)$.

8. Let V_1 and V_2 be independently distributed as exponential(λ_i) $(i = 1, 2)$ random variables. Derive the probability that the sum of V_1 and V_2 is less than a value c.

   If $c \leq 0$, then the probability is zero, as both are positive. For $c > 0$,
   $P(V_1 + V_2 < c) = \int_0^c \int_0^{c-v_1} (1/\lambda_1)e^{v_1/\lambda_1}(1/\lambda_2)e^{v_2/\lambda_2}\,dv_2 dv_1$.

Potentially useful facts:

- A Poisson(λ) random variable Y with parameter $\lambda > 0$ has mean and variance given by $E(Y|\lambda) = \lambda = \text{Var}(Y|\lambda)$ and a probability mass function given by

$$P(Y = y|\lambda) = e^{-\lambda}\lambda^y/y!, \quad y = 0, 1, 2, \dots$$

- If $Y_1, \dots, Y_n$ are independent random variables and Y_i is Poisson(λ_i) distributed for $i = 1, \dots, n$, then $\sum_{i=1}^n Y_i$ is Poisson($\sum_{i=1}^n \lambda_i$) distributed with parameter $\sum_{i=1}^n \lambda_i$.

- A normal random variable with mean μ and variance σ^2 has moment generating function $e^{t\mu+2^{-1}\sigma^2t^2}$, $t \in \mathbb{R}$.

Consider random variables $Y_1, \dots, Y_n$, where Y_i represents the number of plants in an area of size $x_i > 0$. Assume $x_1, \dots, x_n$ are known and fixed constants; $Y_1, \dots, Y_n$ are independent random variables; and, for some fixed $\theta > 0$, each Y_i is Poisson(θx_i) distributed with parameter θx_i.

1. Identify a complete and sufficient statistic for θ based on $Y_1, \dots, Y_n$. Justify your answer.

2. Show, when the maximum likelihood estimator (MLE) $\hat{\theta}_n$ of θ exists, that $\hat{\theta}_n = \sum_{i=1}^n Y_i / \sum_{i=1}^n x_i$.

3. Another estimator of θ is given by $\tilde{\theta}_n = n^{-1} \sum_{i=1}^n Y_i/x_i$. Find the mean squared error (MSE) for both estimators $\tilde{\theta}_n$ and $\hat{\theta}_n$ (from question 2).

4. By applying either your result from question 3 or some moment inequality, show that $\dfrac{n}{\sum_{i=1}^n x_i} \le \dfrac{\sum_{i=1}^n x_i^{-1}}{n}$ for any positive constants $x_1, \dots, x_n$, .

5. Suppose that $\sum_{i=1}^n x_i = 5$ is known. Find an exact form of the most powerful (MP) test of size $\alpha = 9e^{-10}$ for testing $H_0 : \theta = 2$ vs $H_1 : \theta = 1$ based on $Y_1, \dots, Y_n$.

6. For some subset $\Theta_0 \subset \Theta = (0, \infty)$, the test in question 5 is also the uniformly most powerful (UMP) test of size $\alpha = 9e^{-10}$ for testing the claims $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \notin \Theta_0$, keeping in mind that $\theta > 0$. Identify Θ_0 and explain why the UMP property holds.

7. Let $a_n = \sum_{i=1}^n x_i$ and $T_{\theta,n} = a_n^{-1/2} \sum_{i=1}^n Y_i - a_n^{1/2}\theta$, for $\theta > 0$. The natural log of moment generating function $M_{T_{\theta,n}}(\cdot)$ of T_n is given by

$$\log\left\{M_{T_{\theta,n}}(t)\right\} = -t\theta\sqrt{a_n} + \theta a_n(e^{\frac{t}{\sqrt{a_n}}} - 1), \quad t \in \mathbb{R}.$$

   If $a_n \to \infty$ as $n \to \infty$, show that $\log\left\{M_{T_{\theta,n}}(t)\right\} \to \theta t^2/2$ as $n \to \infty$ for any fixed t.

   (*You may use the following:* For $x \in \mathbb{R}$ with $|x| < 1$, there is a function $R(\cdot)$ such that $e^x - 1 = x + x^2/2 + x^3R(x)$ holds, and $|R(x)| < 1$.)

8. Using the result from question 7, describe how to produce an approximate large sample 95% confidence interval for θ based on $T_{\theta,n}$. *(An explicit expression for this interval may be complicated and is not necessary to provide in great detail.)*

1. The joint density of $Y_1, \ldots, Y_n$ is given by

$$\prod_{i=1}^{n} f(y_i|x_i\theta) = \frac{e^{-a_n\theta}\theta^{s_n}\prod_{i=1}^{n} x_i^{y_i}\mathbb{I}_{\{y_i\in A\}}}{\prod_{i=1}^{n} y_i!}, \ \theta > 0 \quad \text{with } a_n \equiv \sum_{i=1}^{n} x_i, \quad s_n \equiv \sum_{i=1}^{n} y_i$$

and $A = \{0, 1, 2, \ldots\}$, and can be written in the form of an exponential family model as

$$\prod_{i=1}^{n} f(y_i|x_i\theta) = C(\theta)e^{s_n \log\theta}h(y_1, \ldots, y_n), \text{ for all } y_1, \ldots, y_n \in \mathbb{R}, \ \theta > 0$$

$$C(\theta) = e^{-a_n\theta}, \quad h(y_1, \ldots, y_n) = \prod_{i=1}^{n} x_i^{y_i}\mathbb{I}_{\{y_i\in A\}} / \prod_{i=1}^{n} y_i!$$

Checking that $\{\log\theta : \theta > 0\} = \mathbb{R}$ contains an open interval, we know from exponential family properties that $S_n = \sum_{i=1}^{n} Y_i$ must be complete and sufficient for θ.

2. Given $y_1, \ldots, y_n \in A$, the likelihood function for θ can be written

$$L(\theta) = \prod_{i=1}^{n} f(y_i|\theta x_i) = e^{-a_n\theta}\theta^{s_n}h(y_1, \ldots, y_n)$$

Then, taking the derivative of the log likelihood function for $\theta > 0$, we have

$$\frac{d\log L(\theta)}{d\theta} = -a_n + \frac{s_n}{\theta} \quad \& \quad \frac{d^2\log L(\theta)}{d\theta^2} = -\frac{s_n}{\theta^2}.$$

Assuming that $s_n > 0$, we may solve $d\log L(\theta)/d\theta = 0$ to find the MLE is $\hat{\theta}_n = s_n/a_n$.

On a technical note, when $s_n = 0$ then $\frac{d\log L(\theta)}{d\theta} = -a_n < 0$ for $\theta \in (0, \infty)$ so that the likelihood is decreasing on $(0, \infty)$; in this case, we would like to choose 0 as the MLE which, technically, is not allowed when the parameter space is $(0, \infty)$. That is, the MLE would not exist in this case but we could define the estimator $\hat{\theta}_n = 0 = s_n/a_n$ in this event.

3. Since $\mathrm{E}_\theta(Y_i) = x_i\theta$, both estimators are unbiased so that

$$MSE_\theta(\hat{\theta}_n) = \mathrm{Var}_\theta(\hat{\theta}_n) = \frac{\sum_{i=1}^{n}\mathrm{Var}_\theta(Y_i)}{(\sum_{i=1}^{n} x_i)^2} = \frac{\theta}{\sum_{i=1}^{n} x_i}$$

$$MSE_\theta(\tilde{\theta}_n) = \mathrm{Var}_\theta(\tilde{\theta}_n) = \frac{\sum_{i=1}^{n} x_i^{-2}\mathrm{Var}_\theta(Y_i)}{n^2} = \frac{\theta\sum_{i=1}^{n} x_i^{-1}}{n^2}$$

using $x_i\theta = \mathrm{Var}_\theta(Y_i)$.

4. Since the MLE is unbiased for θ and a function of the complete and sufficient statistic S_n, it follows that the MLE is the uniform minimum variance unbiased estimator of θ. Since $\tilde{\theta}$ is also unbiased, we must have $\mathrm{Var}_\theta(\tilde{\theta}_n) \le \mathrm{Var}_\theta(\hat{\theta}_n)$ for any $\theta \ge 0$, directly implying the inequality $n/(\sum_{i=1}^{n} x_i) \le n^{-1}\sum_{i=1}^{n} x_i^{-1}$.

   This inequality could also follow from Jensen's inequality $g(\mathrm{E}X) \le \mathrm{E}g(X)$, where $g(x) = 1/x$ is convex on $(0, \infty)$ and X is a discrete random variable with $P(X = x_i) = 1/n$ for $i = 1, \ldots, n$.

5. Using the Neyman-Pearson lemma and the likelihood function $L(\theta)$, the MP test $\phi(Y_1, \ldots, Y_n)$ has the form

$$\phi(Y_1, \ldots, Y_n) = \begin{cases} 1 & \text{if } L(1)/L(2) > k \\ \gamma & \text{if } L(1)/L(2) = k \\ 0 & \text{if } L(1)/L(2) < k \end{cases}$$

where $k \in (0, \infty]$ and $\gamma \in [0, 1]$ are chosen so that $\mathrm{E}_{\theta=2}\phi(Y_1, \ldots, Y_n) = \alpha = 9e^{-10}$. The likelihood ratio is $L(1)/L(2) = e^{-a_n}/(e^{-2a_n}2^{S_n}) = e^5 2^{-S_n}$ using that $a_n = 5$. So the likelihood ratio is a strictly decreasing function of $S_n = \sum_{i=1}^n Y_i$, which means that for the ratio $L(1)/L(2)$ to be greater than/equal to/less than an appropriate k, we need S_n to be less than/equal to/greater than an appropriate $\tilde{k} > 0$. That is, the MP test must have the following form in terms of S_n:

$$\phi(Y_1, \ldots, Y_n) = \begin{cases} 1 & \text{if } S_n < \tilde{k} \\ \gamma & \text{if } S_n = \tilde{k} \\ 0 & \text{if } S_n > \tilde{k} \end{cases}$$

for some $\tilde{k} > 0$ and $\gamma \in [0, 1]$ where

$$9e^{-10}\alpha = \mathrm{E}_{\theta=2}\phi(Y_1, \ldots, Y_n) = P_{\theta=2}(S_n < \tilde{k}) + \gamma P_{\theta=2}(S_n = \tilde{k}).$$

Since S_n is Poisson($\theta a_n = 10$) distributed when $\theta = 2$, we can find $P_{\theta=2}(S_n < 1) = P_{\theta=2}(S_n = 0) = e^{-10} < \alpha$ for $\tilde{k} = 1$ and

$$P_{\theta=2}(S_n < 2) = P_{\theta=2}(S_n = 0) + P_{\theta=2}(S_n = 1) = e^{-10} + e^{-10}10 > \alpha$$

for $\tilde{k} = 2$. So pick $\tilde{k} = 1$ and solve

$$9e^{-10} = \alpha = P_{\theta=2}(S_n < 1) + \gamma P_{\theta=2}(S_n = 1) = e^{-10} + \gamma e^{-10}10$$

to find $\gamma = 4/5$. The MP test of size $\alpha = 9e^{-10}$ is

$$\phi(Y_1, \ldots, Y_n) = \begin{cases} 1 & \text{if } S_n < 1 \\ 4/5 & \text{if } S_n = 1 \\ 0 & \text{if } S_n > 1 \end{cases}$$

6. The same test would have resulted in question 5 for testing $H_0 : \theta = 2$ vs $H_1 : \theta = \theta_1$ for any $\theta_1 < 2$ (there is nothing particularly special about $\theta_1 = 1$). So, intuitively, the alternative parameter space would be $\Theta \setminus \Theta_0 = (0, 2)$ and the null parameter space would be $\Theta_0 = [2, \infty)$.

More rigorously, the likelihood ratio $L(\theta_2)/L(\theta_1)$ is increasing in S_n whenever $\theta_2 > \theta_1 > 0$ (or as in question 5, $L(\theta_2)/L(\theta_1)$ is decreasing in S_n whenever $\theta_2 < \theta_1$). So the likelihood function has the monotone likelihood ratio property in S_n and the UMP test for $H_0 : \theta \geq 2$ vs $H_1 : \theta < 2$ would be

$$\phi(Y_1, \ldots, Y_n) = \begin{cases} 1 & \text{if } S_n < \tilde{k} \\ \gamma & \text{if } S_n = \tilde{k} \\ 0 & \text{if } S_n > \tilde{k} \end{cases}$$

for some $\tilde{k} > 0$ and $\gamma \in [0, 1]$ where $9e^{-10}\alpha = \mathrm{E}_{\theta=2}\phi(Y_1, \ldots, Y_n)$. This is the test in question 5.

7. Fix $t \in \mathbb{R}$. As a_n becomes large as $n \to \infty$, we have $|t/\sqrt{a_n}| < 1$ and we write

$$\begin{aligned}\log\left\{M_{T_{\theta,n}}(t)\right\} &= -t\theta\sqrt{a_n} + \theta a_n(e^{\frac{t}{\sqrt{a_n}}} - 1)\\ &= -t\theta\sqrt{a_n} + \theta a_n\left(\frac{t}{\sqrt{a_n}} + \frac{t^2}{2a_n} + \frac{t^3}{a_n^{3/2}}R(t/\sqrt{a_n})\right)\\ &= \frac{\theta t^2}{2} + \frac{\theta t^3}{a_n^{1/2}}R(t/\sqrt{a_n}) \to \frac{\theta t^2}{2}\end{aligned}$$

as $n \to \infty$ since t and θ are fixed, $|R(t/\sqrt{a_n})| < 1$, and $a_n \to \infty$.

8. By the result above, the moment generating function (MGF) of T_n must converge to the MGF of a normal variable with mean 0 and variance θ, so that $T_{\theta,n}$ converges to a normal$(0,\theta)$ in distribution as $n \to \infty$. Hence, $T_{\theta,n}/\sqrt{\theta}$ must converge in distribution to a standard normal. Based on $T_{\theta,n}/\sqrt{\theta}$ and a normal approximation involving standard normal quantiles $z_{0.975}, z_{0.025}$ that are symmetric around 0, i.e., $P(Z > z_{0.975}) = 0.025 = P(Z \le z_{0.025})$ with $z_{0.975} = -z_{0.025} > 0$, an approximate 95% confidence interval based on $T_{\theta,n}/\sqrt{\theta}$ is

$$\begin{aligned}\left\{\theta > 0 : -z_{0.975} \le \frac{T_{\theta,n}}{\sqrt{\theta}} \le z_{0.975}\right\} &= \left\{\theta > 0 : \left|\frac{T_{\theta,n}}{\sqrt{\theta}}\right| \le z_{0.975}\right\}\\ &= \left\{\theta > 0 : a_n^{-1}s_n^2 + a_n\theta^2 - 2s_n\theta \le \theta z_{0.975}^2\right\}\\ &= \left\{\theta > 0 : a_n\theta^2 - (2s_n - z_{0.975}^2)\theta + a_n^{-1}s_n^2 \le 0\right\}\end{aligned}$$

This would be enough for the problem and the quadratic formula could be used to find the roots of the polynomial in θ and thereby determine the endpoints of an interval for θ. Other standard normal quantiles are possible to use as well.