Sketched Solutions 11/8/11

Master's Exam – Fall 2011
October 27, 2011
1:00 pm - 5:00 pm

TEST NUMBER:__________________ SEAT NUMBER:__________________

A. The number of points for each problem is given.

B. There are problems with varying numbers of parts.

Problems 1 - 5 Probability (60 points)

Problems 6 - 9 Statistics (60 points)

C. Write your answers on the exam paper itself. If you need more room you may use the extra sheets provided. Answer as many questions as you can on each part. Tables are provided.

Good Luck!

1. (a) (5 pts) X has the Beta density $f(x) = 12x^2(1-x), 0 < x < 1$. $P(X > .5) =$ __________

$$\int_{.5}^{1} (12x^2 - 12x^3)\,dx = 4x^3 - 3x^4 \Big|_{.5}^{1}$$

$$= (4-3) - \left(\frac{4}{8} - \frac{3}{16}\right) = 1 - \frac{5}{16} = \frac{11}{16}$$

(b) (7 pts) X and Y are independent random variables. X is distributed Binomial $n = 100, p = .6$ and Y is distributed Poisson mean 30. Use normal approximation to estimate

$P(X + Y \geq 100) =$ $.09$

$E(X+Y) = 60 + 30 = 90$

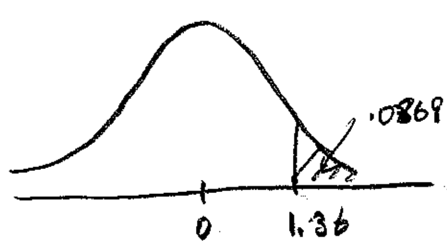
$V(X+Y) = 24 + 30 = 54$

$SD(X+Y) = \sqrt{54} = 7.34$

$$z = \frac{99.5 - 90}{7.34} = 1.29$$

or

$$z = \frac{100 - 90}{7.34} = 1.36$$



2. (6 pts) Let X have the density function $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$. Derive the density function of $Y = X^2$.

$$Y = X^2$$

Let $y > 0$.

$$F(y) := P(Y \le y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$f(y) = F'(y) = 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} \frac{d}{dy} \sqrt{y}$$

$$= \frac{2}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2} y^{-1/2}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{e^{-y/2}}{y^{1/2}}, \quad y > 0$$

$$f(y) = 0, \quad y \le 0$$

3. (12 pts) Consider the model where X is has a uniform distribution on the interval (0, 12) and conditional on $X = x$, Y is uniformly distributed on the interval from $(x/2, 12)$. (In answering you may use may use the fact that the variance of the uniform distribution on the unit interval is 1/12.)

$E(Y) =$ 7.5 and $V(Y) =$ 7.75.

$$E(Y) = E\,E(Y|X) = E\left(\tfrac{1}{2}\left(\tfrac{X}{2}+12\right)\right) = E\left(\tfrac{X}{4}+6\right)$$

$$= \frac{6}{4} + 6 = 7.5$$

$$V(Y) = E\,V(Y|X) + V\,E(Y|X)$$

$$= E\,\frac{\left(12-\frac{X}{2}\right)^2}{12} + V\left(\frac{X}{4}+6\right)$$

$$= E\left(12 - X + \frac{X^2}{48}\right) + \frac{1}{16}V(X)$$

$$= 12 - 6 + \frac{1}{48}E(X^2) + \frac{1}{16}(12)$$

$$EX^2 = V(X) + (EX)^2 = 12 + 6^2 = 48$$

$$V(Y) = 12 - 6 + 1 + \frac{3}{4} = 7\tfrac{3}{4}$$

4. Consider the family of probability densities on the intervals $[\theta, \infty)$

$$f_{\alpha,\theta}(x) = \frac{\cancel{(1-\alpha)}\overset{\alpha-1}{}\theta^{\alpha-1}}{x^{\alpha}}, x \geq \theta \ where \ \ \theta > 0, \alpha > 1.$$

(a) (5 pts) Show that $f_{\alpha,\theta}$ is a probability density.

$$f_{\alpha,\theta}(x) = \frac{(\alpha-1)\theta^{\alpha-1}}{x^{\alpha}} \geq 0 \text{ for } x \geq \theta \text{ since } \alpha > 1$$

$$\int_{\theta}^{\infty} (\alpha-1)\theta^{\alpha-1}x^{-\alpha}\,dx = (\alpha-1)\frac{\theta^{\alpha-1}x^{1-\alpha}}{1-\alpha}\Big|_{\theta}^{\infty}$$

$$= -\frac{(\alpha-1)\theta^{\alpha-1}\theta^{1-\alpha}}{1-\alpha} = 1$$

(b) (5 pts) For what α, θ do the first two moments exist? The first two moments exist if and only if

$$\int_{\theta}^{\infty} x^2 x^{-\alpha}\,dx \text{ is finite.}$$

$$\int_{\theta}^{\infty} \frac{1}{x^{\alpha-2}}\,dx \text{ is finite if } \alpha-2>1\text{, that is, } \alpha>3.$$

5. Let U and V be independent random variables with U distributed Poisson with parameter α and V distributed Poisson with parameter β. The probability mass function of the Poisson distribution with parameter $\theta > 0$ is

$$p_\theta(k) = e^{-\theta}\frac{\theta^k}{k!}, k = 0, 1, 2, \ldots$$

(a) (5 pts) Prove that $Y = U + V$ is distributed Poisson with parameter $\alpha + \beta$.

$$P(Y=y)=P(U+V=y) = \sum_{u=0}^{y} P(U=u,\ V=y-u) = \sum_{u=0}^{y} e^{-\alpha}\frac{\alpha^y}{u!} e^{-\beta}\frac{\beta^{y-u}}{(y-u)!}$$

$$= e^{-(\alpha+\beta)}\frac{1}{y!}\sum_{u=0}^{y}\binom{y}{u}\alpha^u\beta^{y-u}$$

$$= e^{-(\alpha+\beta)}\frac{(\alpha+\beta)^y}{y!}$$

So $Y \sim \text{Poisson}(\alpha+\beta)$.

(b) (5 pts) Prove that the conditional distribution of U given $U + V = n$ where $n > 0$ is Binomial with n trials and chance of success $\pi = \frac{\alpha}{\alpha+\beta}$.

$$P(U=u \mid U+V=n) = \frac{P(U=u,\ U+V=n)}{P(U+V=n)}$$

$$= \frac{P(U=u,\ V=n-u)}{P(U+V=n)}$$

$$= \frac{e^{-\alpha}\frac{\alpha^u}{u!} e^{-\beta}\frac{\beta^{n-u}}{(n-u)!}}{e^{-(\alpha+\beta)}\frac{(\alpha+\beta)^n}{n!}} = \binom{n}{u}\left(\frac{\alpha}{\alpha+\beta}\right)^u\left(\frac{\beta}{\alpha+\beta}\right)^{n-u}$$

So $U \mid U+V=n \sim \text{Binomial}\left(n, \frac{\alpha}{\alpha+\beta}\right)$.

5. **(cont)**

(c) Let X_i be independent random variables having Poisson distributions with parameters $i\lambda$, $i = 1, 2, 3, 4$. Define the random variables $Y_i := X_i + X_{i+1}, i = 1, 2, 3$.

(5 pts) Compute the covariance matrix of (Y_1, Y_2, Y_3).

$$Y_1 = X_1 + X_2 \sim \text{Poisson}(3\lambda)$$
$$Y_2 = X_2 + X_3 \sim \text{Poisson}(5\lambda)$$
$$Y_3 = X_3 + X_4 \sim \text{Poisson}(7\lambda)$$

$$\begin{bmatrix} 3\lambda & 2\lambda & 0 \\ 2\lambda & 5\lambda & 3\lambda \\ 0 & 3\lambda & 7\lambda \end{bmatrix}$$

(5 pts) Compute variance of $Y_1 + 2Y_2 - 4Y_3$.

$$\begin{aligned} Var(Y_1 + 2Y_2 - 4Y_3) &= Var(Y_1) + 4Var(Y_2) + 16Var(Y_3) + 4Cov(Y_1, Y_2) \\ &\quad - 8Cov(Y_1, Y_3) - 16Cov(Y_2, Y_3) \\ &= 3\lambda + 20\lambda + 112\lambda + 8\lambda - 48\lambda \\ &= 95\lambda. \end{aligned}$$

6. Consider the family of probability distributions on $\{0, 1, 2, \ldots, 5\}$ where

$$P_\theta(x) = \frac{\theta}{5}, x = 1, 2, \ldots, 5 \quad \text{and} \quad P_\theta(0) = 1 - \theta, \;\; 0 \leq \theta \leq 1.$$

(a) (4 pts) Determine a method of moments estimator $\hat{\theta}$ for θ based on $X_1, X_2, \ldots, X_n$ *iid* P_θ.

$$E_\theta(X) = \frac{\theta}{5}(1+2+3+4+5) + (1-\theta)(0) = 3\theta$$

$$E_\theta(\bar{X}) = 3\theta$$

$$\hat{\theta} = \frac{\bar{X}}{3}$$

(b) (4 pts) $E_\theta(\hat{\theta}) =$ $\underline{\quad \theta \quad}$ and $V_\theta(\hat{\theta}) =$ ____________.

$$V_\theta(X) = E_\theta X^2 - (E_\theta X)^2 = \frac{\theta}{5}(1+4+9+16+25) + (1-\theta)(0) - (3\theta)^2$$

$$= 11\theta - 9\theta^2$$

$$V_\theta(\hat{\theta}) = V_\theta\left(\frac{\bar{X}}{3}\right) = \frac{1}{9} \frac{V_\theta(X)}{n} = \frac{(\frac{11}{9} - \theta)\theta}{n}$$

6. (cont)

(c) (9 pts) Determine the maximum likelihood estimator $\hat{\theta}_{mle}$ for θ based on $X_1, X_2, \ldots, X_n$ *iid* P_θ.

$$p_\theta(x_1, x_2 \cdots x_n) = \left(\frac{\theta}{5}\right)^y (1-\theta)^{n-y} \quad , \; y = \text{No. of } x_i \in \{1,2,3,4,5\}$$

$$\ln p_\theta = y \ln\theta - y \ln 5 + (n-y)\ln(1-\theta)$$

$$\frac{\partial}{\partial\theta} = \frac{y}{\theta} - \frac{(n-y)}{1-\theta} \qquad \frac{\partial}{\partial\theta} = 0 \Leftrightarrow (1-\theta)y = \theta(n-y)$$

$$y = n\theta$$

$$\theta = \frac{y}{n}$$

$$\frac{\partial}{\partial\theta^2} = \frac{-y}{\theta^2} - \frac{(n-y)}{(1-\theta)^2} \le 0$$

$$\hat{\theta}_{mle} = \frac{y}{n}$$

(d) (4 pts) $E_\theta\left(\hat{\theta}_{mle}\right) =$ $\underline{\quad \theta \quad}$ and $V_\theta\left(\hat{\theta}_{mle}\right) =$ $\underline{\quad \frac{\theta(1-\theta)}{n} \quad}$.

$Y \sim B(n, \theta)$

$E(Y) = n\theta$

$V(Y) = n\theta(1-\theta)$

(e) (4 pts) Show that $V_\theta\left(\hat{\theta}_{mle}\right) < V_\theta\left(\hat{\theta}\right)$ for all $0 \le \theta \le 1$.

$$\frac{\theta(1-\theta)}{n} < \frac{\left(\frac{11}{9}-\theta\right)\theta}{n} \quad \text{iff} \quad 1-\theta < \frac{11}{9}-\theta$$

$$1 < \frac{11}{9}$$

7. Let $X_1, X_2, \ldots$ be a sequence of independent random variables with common probability density function

$$f_\theta(x) = e^{-(x-\theta)}, \qquad \theta \le x < \infty, -\infty < \theta < \infty.$$

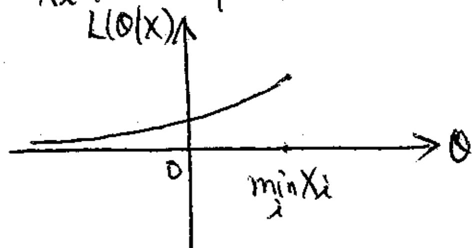
(a) (5 pts) Show that the maximum likelihood estimator of θ based on $X_1, \ldots, X_n$ is $\hat{\theta}_n := \min\{X_1, X_2, \ldots, X_n\}$.

The likelihood

$$L(\theta|X) = \prod_{i=1}^{n} e^{-(x_i-\theta)} I(X_i \ge \theta) = e^{-\sum_{i=1}^{n}(x_i-\theta)} I(\min_i X_i \ge \theta)$$

$$= e^{\theta - \sum_{i=1}^{n} X_i} I(\theta \le \min_i X_i),$$

is an increasing function of θ, which attain its maximum at $\theta = \min_i X_i$. Therefore, the maximum likelihood estimator of θ is $\hat{\theta}_n$.



(b) (5 pts) Show that $\hat{\theta}_n$ is a consistent estimator of θ. Clearly state any theorems that you use in the proof.

Firstly, the CDF of $\hat{\theta}_n$ is

$$F_{\hat{\theta}_n}(x) = P(\hat{\theta}_n \le x) = 1 - P(\hat{\theta}_n > x) = 1 - P(X_1 > x, \cdots, X_n > x)$$

$$= 1 - e^{-n(x-\theta)} \qquad x \ge \theta.$$

For any $\varepsilon > 0$,

$$P(|\hat{\theta}_n - \theta| > \varepsilon) = P(\hat{\theta}_n > \theta + \varepsilon) + P(\hat{\theta}_n < \theta - \varepsilon)$$

$$= e^{-n\varepsilon} + 0 \to 0 \qquad \text{as } n \to \infty.$$

Therefore, $\hat{\theta}_n$ is a consistent estimator of θ.

7. (cont)

(c) (10 pts) Construct the uniformly most powerful test of size α for testing H_0: $\theta = 4$ *vs.* H_1: $\theta > 4$ based on $X_1, \ldots, X_n$.

By Factorization theorem, we know $\hat{\theta}_n = \min_i \{x_1, \cdots x_n\}$ is a sufficient statistic.

For any $\theta_2 > \theta_1$, the likelihood ratio

$$\frac{L(\theta_2 | X)}{L(\theta_1 | X)} = \frac{e^{\theta_2 - \sum_{i=1}^n X_i} I(\hat{\theta}_n \geq \theta_2)}{e^{\theta_1 - \sum_{i=1}^n X_i} I(\hat{\theta}_n \geq \theta_1)} = e^{\theta_2 - \theta_1} \frac{I(\hat{\theta}_n \geq \theta_2)}{I(\hat{\theta}_n \geq \theta_1)}$$

is a monotone function of $\hat{\theta}_n$. So the uniformly most powerful test will have rejection region $\{\hat{\theta}_n > k\}$.

Set $\alpha = P_{\theta=4}(\hat{\theta}_n > k) = e^{-n(k-4)}$, we have $k = 4 - \frac{\log \alpha}{n}$.

Therefore, the uniformly most powerful test will reject the null hypothesis if $\{\hat{\theta}_n > 4 - \frac{\log\alpha}{n}\}$.

8. (5 pts) X is distributed binomial with n trials and chance of success π, that is, X has probability mass function

$$p_\pi(k) = \binom{n}{k}\pi^k(1-\pi)^{n-k}, k = 0, 1, \dots, n.$$

Consider the estimator $\hat{\pi} = \frac{X}{n}$ for π. Show that $P_\pi\left(|\hat{\pi} - \pi| \le \frac{1}{\sqrt{n}}\right) \ge .75$ for all π.

By Chebyshev Inequality

$$P_\pi(|\hat{\pi} - E(\hat{\pi})| \le 2SD(\hat{\pi})) \ge .75$$

$E(\hat{\pi}) = \pi$ $\quad SD(\hat{\pi}) = \sqrt{\frac{\pi(1-\pi)}{n}} \le \sqrt{\frac{(.5)(.5)}{n}}$

so that $2\,SD(\hat{\pi}) \le \frac{1}{\sqrt{n}}$

$$\therefore P_\pi\left(|\hat{\pi} - \pi| \le \frac{1}{\sqrt{n}}\right) \ge .75.$$

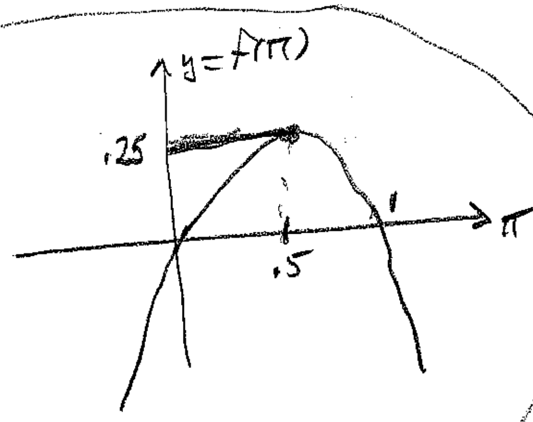
$f(\pi) = \pi(1-\pi)$

$f'(\pi) = 1 - 2\pi$

$f''(\pi) = -2$

Concave down,

Maximized at $\pi = .5$



Proof.

9. Three objects are to be weighed on a scale. There true weights are β_1, β_2 and β_3. The scale is subject to error and for Object i it gives a measurement $Y_i = \beta_i + \epsilon_i, i = 1, 2, 3$. For any true weight β, the scale returns the measurement $Y = \beta + \epsilon$ where $E(Y) = \beta$ and $V(Y) = \sigma^2$ where the variance σ^2 does not depend on β. We model the errors in repeated weighings as uncorrelated. The three objects are weighed separately and then all three objects are weighed together.

(a) (5 pts) Write the equations of the linear model for this experiment with $n = 4$ observations and $p = 3$ beta parameters.

$$Y_1 = \beta_1 + \epsilon_1$$
$$Y_2 = \beta_2 + \epsilon_2$$
$$Y_3 = \beta_3 + \epsilon_3$$
$$Y_4 = \beta_1 + \beta_2 + \beta_3 + \epsilon_4$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}}_{X} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

(b) (5 pts) Solve the normal equations to determine the least squares estimators for β_1, β_2 and β_3 and compute the variances of the estimators.

Normal Equations: $X'X\hat{\beta} = X'Y$

$$X'X = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \qquad X'Y = \begin{bmatrix} Y_1 + Y_4 \\ Y_2 + Y_4 \\ Y_3 + Y_4 \end{bmatrix}$$

Normal Equations:
$$\begin{cases} 2\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = Y_1 + Y_4 \\ \hat{\beta}_1 + 2\hat{\beta}_2 + \hat{\beta}_3 = Y_2 + Y_4 \\ \hat{\beta}_1 + \hat{\beta}_2 + 2\hat{\beta}_3 = Y_3 + Y_4 \end{cases}$$

$E(\hat{\beta}_i) = \beta_i$

Solution to Normal Equations:
$$\begin{cases} \hat{\beta}_2 = \frac{1}{4}[-Y_1 + 3Y_2 - Y_3 + Y_4] \\ \hat{\beta}_1 = \frac{1}{4}[3Y_1 - Y_2 - Y_3 + Y_4] \\ \hat{\beta}_3 = \frac{1}{4}[-Y_1 - Y_2 + 3Y_3 + Y_4] \end{cases}$$

$\therefore V(\hat{\beta}_i) = \frac{12}{16}\sigma^2 = \frac{3}{4}\sigma^2$