Final – Math 5080 – Spring 2004

Name: ____________________________________

Instructions. READ CAREFULLY.:

(i) The work you turn in must be your own. You may not discuss the final with anyone, either in the class or outside the class. [You may of course consult with me for clarification of any of the problems.] *Failure to follow this policy will be considered cheating and will result in a course grade of E.*

(ii) You may consult the textbook and your notes. In particular, feel free to use any of the information in the tables of distributions in Appendix B, for example, the moment generating functions for specific distributions. You can use a **general mathematical reference**, for example a calculus text or a table of integrals. You *may not* use any statistical textbook or written source material concerning the specific subject matter of the course. *Failure to follow this policy will be considered cheating and will result in a course grade of E.*

(iii) Your final must be clearly written and legible. **I will not grade problems which are sloppily presented and such problems will receive a grade of** 0. If you are unable to write legibly and clearly, use of a word processor. You have at least 7 days to complete the final; budget time for writing up your solutions.

(iv) Think about your exposition. Someone (me) has to read what you have written. Your answer is only correct if I can understand what you have done. Style matters.

(v) Finals are due at 6 PM on Wednesday May 5, 2004. **Late finals will not be accepted**, except for reasons of death or serious illness. It is highly recommended that you turn in your exam to me personally. I will be in my office (LCB 209) at 6 PM May 5 2004 to accept exams then, and will check my box in the math department at that time. I cannot guarantee that exams left in my box will be received.

Sign here to indicate you have read and understand these instructions.

Problem 1. For each n, let $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$ be independent Bernoulli random variables. That is,

$$X_{n,i} = \begin{cases} 1 & \text{with probability } p_{n,i}\,, \\ 0 & \text{with probability } 1 - p_{n,i}\,. \end{cases}$$

Note that the $p_{n,i}$ are not assumed to be identical. Suppose that

$$\lim_{n\to\infty} \sum_{i=1}^{n} p_{n,i} = \mu\,, \quad \text{and} \quad \lim_{n\to\infty} \sum_{i=1}^{n} p_{n,i}^2 = 0\,. \tag{1}$$

Find a random variable Y so that $S_n = \sum_{i=1}^{n} X_{n,i}$ converges in distribution to Y. Prove your answer.

Solution. Let $M_{n,i}(t)$ be the mgf for $X_{n,i}$:

$$M_{n,i}(t) = E\big(e^{tX_{n,i}}\big) = e^t p_{n,i} + (1 - p_{n,i}) = 1 + p_{n,i}\left(e^t - 1\right)\,.$$

If $M_n(t)$ is the mgf for S_n, we have

$$\begin{aligned} M_n(t) &= \prod_{i=1}^{n} M_{n,i}(t) \\ &= \prod_{i=1}^{n} \left[1 + p_{n,i}(e^t - 1)\right] \\ \log M_n(t) &= \sum_{i=1}^{n} \log\left(1 + p_{n,i}(e^t - 1)\right)\,. \end{aligned}$$

Now $\log(1 + x) = x + \varepsilon(x)$, where $|\varepsilon(x)| \le x^2$. Thus

$$\begin{aligned} M_n(t) &= \sum_{i=1}^{n} \left[p_{n,i}(e^t - 1) + \varepsilon\left(p_{n,i}(e^t - 1)\right)\right] \\ &= (e^t - 1)\sum_{i=1}^{n} p_{n,i} + \sum_{i=1}^{n} \varepsilon\left(p_{n,i}(e^t - 1)\right)\,. \end{aligned}$$

Now,

$$\begin{aligned}
\left|\sum_{i=1}^{n} \varepsilon\left(p_{n,i}(e^t-1)\right)\right| &\le \sum_{i=1}^{n} \left|\varepsilon\left(p_{n,i}(e^t-1)\right)\right| \\
&\le \sum_{i=1}^{n} p_{n,i}^2 \left(e^t-1\right)^2 \\
&= \left(e^t-1\right)^2 \sum_{i=1}^{n} p_{n,i}^2 \,.
\end{aligned}$$

We conclude from the hypotheses (1) that

$$\log M_n(t) \to \mu(e^t-1)\,.$$

Thus, we conclude that S_n converges in distribution to a Poisson random variable with mean μ. □

Problem 2. Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with density

$$f(x;\theta) = \begin{cases} 2\theta^{-2}x & \text{if } 0 \le x \le \theta \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the MLE for θ.

(ii) Compute the bias of the MLE for θ.

Solution. The likelihood function is

$$\begin{aligned}
L(\theta;\boldsymbol{x}) &= \prod_{i=1}^{n} \frac{2}{\theta^2}\mathbf{1}\left\{0 \le x_i \le \theta\right\} \\
&= \frac{2^n}{\theta^{2n}}\mathbf{1}\left\{x_{(n)} \le \theta\right\}\,.
\end{aligned}$$

By graphing this function, we see it is largest at $\theta = x_{(n)}$, and so the MLE is $\hat{\theta} = X_{(n)}$.

We need to find the distribution of $X_{(n)}$: For $0 \le x \le \theta$

$$\begin{aligned}
P\big(X_{(n)} \le x\big) &= F(x)^n \\
&= \frac{x^{2n}}{\theta^{2n}}\,,
\end{aligned}$$

and so

$$f_{X_{(n)}}(x) = \frac{2nx^{2n-1}}{\theta^{2n}}$$

for $x \in [0, \theta]$. Then

$$E\big(X_{(n)}\big) = \int_0^\theta x \frac{2nx^{2n-1}}{\theta^{2n}} dx = \frac{2nx^{2n+1}}{(2n+1)\theta^{2n}}\bigg|_0^\theta = \frac{2n}{2n+1}\theta$$

Thus

$$\beta = \theta - \frac{2n}{2n+1}\theta = \frac{\theta}{2n+1}\,.$$

□

Problem 3. Let X and Y be two independent random variables with respective density functions

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

$$g(x) = \begin{cases} e^{-x} & \text{if } 0 \le x < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Compute the density of $X + Y$.

Solution. Let $S = X + Y$. Using the convolution formula, for $s \ge 0$, and letting $a \wedge b =$

$\min\{a, b\}$,

$$\begin{aligned}
f_S(s) &= \int_{-\infty}^{\infty} f_X(x) f_Y(s-x) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{2} \mathbf{1}\left\{0 \le x \le 2\right\} e^{-(s-x)} \mathbf{1}\left\{0 \le s - x\right\} dx \\
&= \int_{0}^{2\wedge s} \frac{1}{2} e^{-(s-x)} dx \\
&= \frac{1}{2} e^{-(s-x)} \Big|_0^{2\wedge s} \\
&= \frac{1}{2} e^{-(s-2\wedge s)} - \frac{1}{2} e^{-s} .
\end{aligned}$$

In other words,

$$f_S(s) = \begin{cases} \frac{1}{2}\left(1 - e^{-s}\right) & \text{if } 0 \le s \le 2\,, \\ \frac{1}{2}\left(e^{-(s-2)} - e^{-s}\right) & \text{if } 2 \le 2 < \infty\,, \\ 0 & \text{if } s < 0\,. \end{cases}$$

Check that it integrates to one:

$$\begin{aligned}
\int_0^{\infty} f_S(s) ds &= \int_0^2 \frac{1 - e^{-s}}{2} ds + \int_2^{\infty} \frac{e^{-(s-2)} - e^{-s}}{2} ds \\
&= \frac{s + e^{-s}}{2} \Big|_0^2 + \frac{-e^{-(s-2)} + e^{-s}}{2} \Big|_2^{\infty} \\
&= \frac{1 + e^{-2}}{2} + \frac{1 - e^{-2}}{2} \\
&= 1\,.
\end{aligned}$$

□

Problem 4. Let X_1 and X_2 be independent random variables with the density function

$$f(x) = \begin{cases} e^{-x} & \text{if } 0 \le x < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Compute the joint density function of $Y_1 = X_1 + X_2$ and $Y_2 = 2X_1 - X_2$.

Solution. Let $g(x_1, x_2) = (x_1 + x_2, 2x_1 - x_2)$. We have

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

and since

$$\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^{-1} = \frac{1}{-3} \begin{bmatrix} -1 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} \end{bmatrix}$$

We have $g^{-1}(y_1, y_2) = (\frac{y_1}{3} + \frac{y_2}{3}, \frac{2y_1}{3} - \frac{y_2}{3})$, and

$$Dg^{-1}(y_1, y_2) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} \end{bmatrix}$$

$$|J_{g^{-1}}(y_1, y_2)| = \left| \frac{1}{9} - \frac{2}{9} \right| = \frac{1}{3}.$$

Thus

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f(g_1^{-1}(y_1, y_2)) f(g_2^{-1}(y_1, y_2)) |J_{g^{-1}}(y_1, y_2)| \\ &= e^{-\frac{y_1}{3} - \frac{y_2}{3}} \mathbf{1}\{y_1 + y_2 \ge 0\} e^{-\frac{2y_1}{3} + \frac{y_2}{3}} \mathbf{1}\{2y_1 - y_2 \ge 0\} \frac{1}{3} \\ &= \frac{e^{-y_1}}{3} \mathbf{1}\{-y_1 \le y_2 \le 2y_1\}. \end{aligned}$$

Note that if you don't indicate the region where the density is positive, your answer is wrong. In particular, it is *not* the case that $f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{3} e^{-y_1}$, as that does not integrate to one (it does not even integrate to a finite number!) □

Problem 5. Let $X_1, X_2, \ldots, X_6$ be independent random variables. We assume that X_1 and X_2 are Normal($\mu = 0, \sigma^2 = 2$), while X_3, X_4, X_5, X_6 are Normal($\mu = 0, \sigma^2 = 4$). Determine c such that

$$P\left(\frac{X_1 + X_2}{\sqrt{X_3^2 + \ldots + X_6^2}} \leq c\right) = 0.9\,.$$

Solution. Notice that $\frac{X_1+X_2}{2}$ is a Normal(0, 1) random variable. Also $X_i^2/4$ is χ_1^2 for $i = 3, \ldots, 6$ so $\frac{\sum_{i=3}^6 X_i^2}{4}$ is χ_4^2. Thus

$$\frac{\frac{X_1+X_2}{2}}{\sqrt{\frac{\sum_{i=3}^6 X_i^2}{4}/4}} = 2\frac{X_1 + X_2}{\sqrt{\sum_{i=3}^6 X_i^2}}$$

has a t_4 distribution. The 90the percentile of the t_4 distribution is 1.533. Thus since

$$P\left(\frac{X_1 + X_2}{\sqrt{X_3^2 + \ldots + X_6^2}} \leq c\right) = P\left(2\frac{X_1 + X_2}{\sqrt{X_3^2 + \ldots + X_6^2}} \leq 2c\right) = P(t_4 \leq 2c)\ ,$$

Setting $2c = 1.533$ makes the probability equal to 0.9 Thus $c = 0.7665$.

□

Problem 6. Let $X_1, X_2, \ldots, X_{120}$ be independent and identically distributed random variables with density function

$$f(x) = \begin{cases} 3x^2 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Get an approximate value of c so that

$$P(X_1 + \cdots + X_{120} \leq c) = 0.99\,.$$

Solution. We have

$$E(X_1) = \int_0^1 x3x^2dx = \frac{3}{4}x^4\Big|_0^1 = \frac{3}{4}\,,$$

and

$$E\big(X_1^2\big) = \int_0^1 x^2 3x^2 dx = \frac{3}{5}x^5\bigg|_0^1 = \frac{3}{5}\,.$$

Thus

$$\text{Var}\,(X_1) = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$$

Let $S_n = \sum_{i=1}^{120} X_i$. Then

$$E(S_n) = 120\frac{3}{4} = 90\,, \quad \text{SD}\,(S_n) = \sqrt{120\frac{3}{80}} = \frac{3}{\sqrt{2}}$$

Thus

$$\begin{aligned} P(X_1 + \cdots + X_{120} \leq c) &= P\left(\frac{S_n - 90}{\frac{3}{\sqrt{2}}} \leq \frac{c - 90}{\frac{3}{\sqrt{2}}}\right) \\ &\approx \Phi(z)\,, \end{aligned}$$

where $z = \frac{c-90}{\frac{3}{\sqrt{2}}}$. The 90th percentile of the standard normal distribution is 3.26. Thus solving

$$2.326 = \frac{c - 90}{\frac{3}{\sqrt{2}}}$$

for c gives $c = 94.9342$.

$\square$

Problem 7. Let $X_1, X_2, \ldots, X_n$ be i.i.d. Geometric(p) random variables:

$$P(X_i = x; p) = \begin{cases} p(1-p)^{x-1} & \text{if } x = 1, 2, \ldots\,, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the MLE of p.

(ii) Find the UMVU estimator of $1/p$. [Justify your answer.]

Solution. We have

$$f(\boldsymbol{x};p) = \prod_{i=1}^{n} p(1-p)^{x_i-1} = p^n(1-p)^{\sum x_i - n}$$
$$= \exp\Big(n\log p + (1-p)\sum x_i - n\log(1-p)\Big)$$
$$= \exp\left((1-p)\sum x_i + n\log\left(\frac{p}{1-p}\right)\right) .$$

Thus this is a regular exponential class, with sufficient statistic $\sum x_i$. It follows that $\sum x_i$ is a complete minimal sufficient statistic.

We have

$$\ell(p) = \log(1-p)\sum x_i + n\log\left(\frac{p}{1-p}\right)$$
$$\frac{\partial \ell}{\partial p} = -\frac{\sum x_i}{1-p} + \frac{n}{p(1-p)} .$$

Setting $\frac{\partial \ell}{\partial p} = 0$ and solving for p gives

$$0 = -p\sum x_i + n$$
$$p = \frac{n}{\sum x_i}$$
$$p = \frac{1}{\bar{x}} .$$

It can be checked that this is indeed a global maximum. Thus $\hat{p} = 1/\bar{X}$. By the invariance of the MLE, we have $\widehat{1/p} = \bar{X}$.

It is clear that $E\big(\bar{X}\big) = \frac{1}{p}$, so the MLE is unbiased.

Since $\bar{X}$ is a function of the complete minimum sufficient statistic, it is UMVU (by Lehmann-Scheffe Theorem). □

Problem 8. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be i.i.d. random vectors with the following density:

$$f(x,y;\rho) = \begin{cases} \frac{1}{\pi\rho^2} & \text{if } (x,y) \text{ is in the disc centered at 0 with radius } \rho\,, \\ 0 & \text{otherwise.} \end{cases}$$

Find the MLE of ρ.

Solution. The likelihood is

$$\begin{aligned} L(\rho) &= \prod_{i=1}^{n} \frac{1}{\pi\rho^2} \mathbf{1}\left\{\sqrt{X_i^2 + Y_i^2} \le \rho\right\} \\ &= \frac{1}{\pi^n \rho^{2n}} \mathbf{1}\left\{\max_i \sqrt{X_i^2 + Y_i^2} \le \rho\right\} \end{aligned}$$

Thus the MLE is

$$\hat{\rho} = \max_i \sqrt{X_i^2 + Y_i^2}\,.$$

Note that

$$\max_i \sqrt{X_i^2 + Y_i^2} \ne \sqrt{X_{(n)}^2 + Y_{(n)}^2}\,.$$

[Try the two points $(2,1)$ and $(1,2)$: In this case $x_{(2)} = y_{(2)} = 2$, so $\sqrt{x_{(2)}^2 + y_{(2)}^2} = \sqrt{8}$, while $\max \sqrt{x_i^2 + y_i^2} = \sqrt{5}$.] □