Final Exam, Biostatistics 6341
12 December, 2016

Please show all your work and perform all calculations to whatever degree of exactness you are able. This test is closed book and no calculators are allowed. Please justify your answers. Partial credit will be given.

1. (45)
 a. DeMorgan's Law states that $(A \cup B)^C =$?
 b. Let $f(x) = cx^2$ for $0 < x < 1$. What value of c makes $f(x)$ a valid pdf?
 c. What is the cdf, $F(x)$?
 d. What is the inverse cdf, $F^{-1}(t)$?
 e. What is the median of X?
 f. What is $E(X)$?
 g. What is $Var(X)$?
 h. How would you generate random variables with pdf $f(x)$ using only the uniform(0,1) distribution?

2. (40)
 a. Let $X \sim N(\mu, \sigma^2)$. Show that the moment generating function of X is

$$exp(\mu t + \sigma^2 t^2/2).$$

 b. Derive $E(X^3)$.
 c. Let $X_i \sim^{iid} N(\mu, \sigma^2)$ for $i = 1, \ldots, n$. Prove that $\overline{X}_n \sim N(\mu, \sigma^2/n)$.
 d. Suppose $X_i \sim^{iid}$ Pareto(α, β) for $i = 1, \ldots, n$. That is, $f(x) = \beta\alpha^\beta/(x^{\beta+1})$ for $0 < \alpha < x < \infty, \beta > 0$, with $E(X) = \beta\alpha/(\beta - 1)$ and $Var(X) = \beta\alpha^2/((\beta - 1)^2(\beta - 2))$. What is an approximate distribution of $\overline{X}_n$?

3. (30) Let $X_i, i = 1, 2$, be independent Gamma$(\alpha_i, 1)$ random variables. Find the marginal distributions of $U = X_1/(X_1 + X_2)$ and $V = X_2/(X_1 + X_2)$.

4. (40) Let $X_i \sim^{iid}$ Unif(0,1) for $i = 1, \cdots, n$.
 a. What is the joint pdf of $X_{(1)}$ and $X_{(n)}$?
 b. What is the conditional pdf of $X_{(1)}$ given $X_{(n)}$?
 c. Prove that $X_{(1)} \to^p 0$.
 d. Prove that $nX_{(1)} \to^d Exp(1)$.

5. (45) There are 12 floors above the 2nd floor in 2525 West End (floors 3 to 15, skipping 13, the lobby is on the 2nd floor). Suppose n independent people enter the elevator on the 2nd floor and are going up, and suppose that each of them has an equal probability of getting off at each floor. (Throughout this problem,

assume that no person arriving at the elevator in the lobby is going to take it down to the 1st floor.)

a. What is the expected number of people getting off before the 11th floor? What is the variance?

b. Let X_3 be the number getting off on the 3rd floor and X_4 the number getting off on the 4th floor. What is the covariance of X_3 and X_4?

c. What is the probability that nobody will get off on the 5th floor?

d. What are the expected number of stops before the 11th floor? (more difficult – skip if you get stuck)

e. Now suppose that the number of people who show up at the elevator on the 2nd floor in t seconds, N, is a random variable distributed as Poisson(λt). Suppose that the time until the next elevator arrives, T, is exponentially distributed with mean β. There are no people waiting at the elevator when you push the up button. What is the expected number of people going in the next elevator with you? What is the variance? (Don't count yourself and assume that everyone who arrives goes up.)

f. I was going to ask a problem like the following: Suppose you show up at the elevator and there are 10 people (in addition to you) getting onto it. Should you get on the elevator or wait for the next one if you want to have the shortest expected time to reaching the 11th floor? Why? Describe your reasoning. This is a somewhat complicated problem, and I haven't given you all the information you need to solve it. However, please describe what additional information you might need to solve it and describe how you would solve it if you had that information.

1

a $(A \cup B)^c = A^c \cap B^c$

b $f(x) = cx^2$ for $0 < x < 1$.

$$\int_0^1 cx^2\,dx = \frac{cx^3}{3}\Big]_0^1 = \frac{c}{3} = 1 \quad \text{if valid pdf.}$$

$$\Rightarrow \quad c = 3.$$

c $F(x) = \int_0^x 3t^2\,dt = \frac{3t^3}{3}\Big]_0^x = x^3$ for $0 < x < 1$,

0 for $x \leq 0$

1 for $x \geq 1$

d $t = F(x) = x^3 \quad \Rightarrow \quad x = t^{1/3} = F_X^{-1}(t)$

e What is the Median of X?

$$\text{Median}(X) = F_X^{-1}\left(\tfrac{1}{2}\right) = \left(\tfrac{1}{2}\right)^{1/3}$$

f $E(X) = \int_0^1 x\,3x^2\,dx = \int_0^1 3x^3\,dx = \frac{3x^4}{4}\Big]_0^1 = \frac{3}{4}$

g $Var(X) = E(X^2) - (E(X))^2$

$$E(X^2) = \int_0^1 x^2\,3x^2\,dx = \int_0^1 3x^4\,dx = \frac{3x^5}{5}\Big]_0^1 = \frac{3}{5}$$

$$\Rightarrow Var(X) = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{5} - \frac{9}{16}$$

$$= \frac{3(16) - 9(5)}{5(16)} = \frac{48 - 45}{80} = \frac{3}{80}$$

h Generate $U \sim \text{Unif}(0,1)$. Let $X = U^{1/3}$.

2 b

$$E(X^4) = \frac{\partial^4}{\partial t^4} \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

$$m_x(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

$$m_x'(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left(\mu + \sigma^2 t\right)$$

$$m_x''(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left(\mu + \sigma^2 t\right)^2 + \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\sigma^2$$

$$m_x'''(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left(\mu + \sigma^2 t\right)^3 + \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)2\left(\mu + \sigma^2 t\right)\sigma^2$$

$$+ \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\sigma^2\left(\mu + \sigma^2 t\right)$$

$$m_x'''(t)\Big|_{t=0} = \exp(0+0)(\mu+0)^3 + \exp(0)\,2(\mu+0)\sigma^2$$

$$+ \exp(0)\,\sigma^2(\mu+0)$$

$$= \mu^3 + 2\mu\sigma^2 + \sigma^2\mu$$

$$= \mu^3 + 3\mu\sigma^2$$

a

$$E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(x^2 - 2x\mu + \mu^2 - 2tx\sigma^2\right)\right)dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(x^2 - 2x(\mu + t\sigma^2) + (\mu+t\sigma^2)^2 - (\mu+t\sigma^2)^2 + \mu^2\right)\right)dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(x - (\mu+t\sigma^2)\right)^2\right) \exp\left(+\frac{1}{2\sigma^2}(\mu+t\sigma^2)^2 - \frac{1}{2\sigma^2}\mu^2\right)dx$$

$$= \exp\left(\frac{1}{2\sigma^2}\left(\mu^2 + 2\mu t\sigma^2 + t^2\sigma^4 - \mu^2\right)\right)$$

$$= \boxed{\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)}$$

$\boxed{c}$ $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$M_{\bar{X}}(t) = M_{\frac{\Sigma X_i}{n}}(t) = M_{\Sigma X_i}\left(\frac{t}{n}\right)$$

$$= E\left(e^{\Sigma X_i \frac{t}{n}}\right) = E\left(\prod e^{X_i \frac{t}{n}}\right)$$

$$= \prod\left(E\left(e^{X_i \frac{t}{n}}\right)\right) = \left(M_{X_i}\left(\frac{t}{n}\right)\right)^n$$

$$= \left(\exp\left(\mu \frac{t}{n} + \frac{\sigma^2 t^2}{2n^2}\right)\right)^n$$

$$= \exp\left(\frac{n\mu t}{n} + \frac{n\sigma^2 t^2}{2n^2}\right) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2n}\right)$$

$$= \text{mgf of } N\left(\mu, \frac{\sigma^2}{n}\right) \quad \blacksquare$$

$\boxed{d}$ $\sqrt{n}\left(\bar{X}_n - \frac{\beta\alpha}{\beta-1}\right) \xrightarrow{d} N\left(0, \frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)}\right)$

(3) $X_1 \sim \text{Gamma}(\alpha_1, 1)$ $\quad X_1 \perp X_2$

$X_2 \sim \text{Gamma}(\alpha_2, 1)$

$$f(x_1, x_2) = \frac{1}{\Gamma(\alpha_1)} x_1^{\alpha_1 - 1} e^{-x_1} \frac{1}{\Gamma(\alpha_2)} x_2^{\alpha_2 - 1} e^{-x_2}$$

$u = \dfrac{x_1}{x_1 + x_2}$; $\quad u = \dfrac{v}{v + x_2} \Rightarrow u(v + x_2) = v \Rightarrow uv + ux_2 = v$

$\Rightarrow ux_2 = v - uv$

$\Rightarrow x_2 = \dfrac{v - uv}{u} = \dfrac{v}{u} - v$

$= \dfrac{v}{u}(1 - u)$

$v = x_1 \qquad x_1 = v$

$\Rightarrow \dfrac{\partial x_1}{\partial u} = 0 \qquad \dfrac{\partial x_1}{\partial v} = 1$

$\Rightarrow |J| = \left|\dfrac{-v}{u^2}\right| = \dfrac{v}{u^2}$

$\dfrac{\partial x_2}{\partial u} = -\dfrac{v}{u^2} \qquad \dfrac{\partial x_2}{\partial v} =$ something

$$\Rightarrow f(u, v) = \frac{1}{\Gamma(\alpha_1)} v^{\alpha_1 - 1} e^{-v} \frac{1}{\Gamma(\alpha_2)} \left(\frac{v}{u} - v\right)^{\alpha_2 - 1} e^{-\left(\frac{v}{u} - v\right)} \frac{v}{u^2} \text{ for } 0 < u < 1$$

$$f(u) = \int_0^\infty f(u, v)\, dv$$

$$= \int_0^\infty \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2) u^2} v^{\alpha_1 + 1 - 1} \left(v\left(\frac{1-u}{u}\right)\right)^{\alpha_2 - 1} e^{-v - v\left(\frac{1-u}{u}\right)} dv$$

$$= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2) u^2} \left(\frac{1-u}{u}\right)^{\alpha_2 - 1} \int_0^\infty v^{\alpha_1 + 1 + \alpha_2 - 1 - 1} e^{-v\left(1 + \frac{1-u}{u}\right)} dv$$

$$= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2) u^2} \left(\frac{1-u}{u}\right)^{\alpha_2 - 1} \Gamma(\alpha_1 + \alpha_2)\, u^{\alpha_1 + \alpha_2} \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha_1 + \alpha_2) u^{\alpha_1 + \alpha_2}} v^{\alpha_1 + \alpha_2 - 1} e^{-v\left(\frac{1}{u}\right)} dv}_{1}$$

$$= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{-2 - \alpha_2 + 1 + \alpha_1 + \alpha_2} (1 - u)^{\alpha_2 - 1}$$

$$= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1 - 1} (1 - u)^{\alpha_2 - 1} \Rightarrow U \sim \text{Beta}(\alpha_1, \alpha_2)$$

By symmetry, $V = \dfrac{x_2}{x_1 + x_2} \sim \text{Beta}(\alpha_2, \alpha_1)$

$\boxed{4}$ $X_i \overset{iid}{\sim} Unif(0,1) \quad \Rightarrow \quad f(x) = 1$ for $0 < x < 1$

$F(x) = x$ for $0 < x < 1$

$\boxed{a}$

$$f_{X_{(1)}, X_{(n)}}(x_1, x_n) = \frac{n!}{1!\,(n-2)!\,1!} f(x_1)\left(F(x_n) - F(x_1)\right)^{n-2} f(x_n)$$

$$= n(n-1) \cdot 1 \cdot (x_n - x_1)^{n-2} \cdot 1$$

$$= n(n-1)(x_n - x_1)^{n-2} \quad \text{for } 0 < x_1 < x_n < 1.$$

$\boxed{b}$

$$f_{X_{(1)}|X_{(n)}}(x_1|x_n) = \frac{f_{X_{(1)},X_{(n)}}(x_1, x_n)}{f(x_n)}$$

$$= \frac{n(n-1)(x_n - x_1)^{n-2}}{n x_n^{n-1}} = \frac{(n-1)(x_n - x_1)^{n-2}}{x_n^{n-1}}$$

for $0 < x_1 < x_n < 1$

$$f_{X_{(n)}}(x_n) = \frac{n!}{(n-1)!\,1!}\left(F(x_n)\right)^{n-1} f(x_n) = n x_n^{n-1}$$

$\boxed{c}$ Prove $X_{(1)} \xrightarrow{P} 0$. for some $\varepsilon > 0$

$$P(|X_{(1)} - 0| > \varepsilon) = P(X_{(1)} > \varepsilon) + P(-X_{(1)} > \varepsilon)$$

$$= P(X_{(1)} > \varepsilon) + P(X_{(1)} < -\varepsilon)$$

$$= P(X_{(1)} > \varepsilon) + 0$$

$$= P(\text{all } X_i > \varepsilon)$$

$$= \left(P(X_1 > \varepsilon)\right)^n \quad \text{because iid}$$

$$= (1 - F_X(\varepsilon))^n$$

$$= (1 - \varepsilon)^n$$

$$\lim P(|X_{(1)} - 0| > \varepsilon) = \lim (1 - \varepsilon)^n = 0 \quad \blacksquare$$

$\boxed{4}$ $\boxed{d}$ Prove $nX_{(1)} \xrightarrow{d} Exp(1)$

$$P\left(X_{(1)} > \frac{t}{n}\right) = \left(1 - \frac{t}{n}\right)^n \quad \text{(following same steps as part c)}$$

$$\Rightarrow \lim_{n \to \infty} P\left(nX_{(1)} > t\right) = \lim_{n \to \infty} \left(1 - \frac{t}{n}\right)^n = \exp(-t)$$

$$= 1 - F_Z(t)$$

where $Z \sim Exp(1)$

$$\Rightarrow \lim_{n \to \infty} P\left(nX_{(1)} \le t\right) = 1 - \exp(t)$$

$$= F_Z(t)$$

☑

5 $(X_3, X_4, \dots, X_{15}) \sim \text{Multinomial}\left(n, (p_3, p_4, \dots, p_{15})\right)$

where $p_3 = p_4 = \dots = p_{15} = \frac{1}{12}$

a) Let $Y = X_3 + X_4 + \dots + X_{10}$

$$Y \sim \text{Bin}(n, p_3 + p_4 + \dots + p_{10}) = \text{Bin}\left(n, \tfrac{8}{12}\right) = \text{Bin}\left(n, \tfrac{2}{3}\right)$$

$$\boxed{E(Y) = \tfrac{2}{3}n} \qquad \text{Var}(Y) = \tfrac{2}{3}\left(\tfrac{1}{3}\right)n = \boxed{\tfrac{2}{9}n}$$

b) $$\text{Cov}(X_3, X_4) = -n p_3 p_4 = \boxed{-n\left(\tfrac{1}{12}\right)^2}$$

c) $$P(X_5 = 0) = (1 - p_5)^n = \left(1 - \tfrac{1}{12}\right)^n = \boxed{\left(\tfrac{11}{12}\right)^n}$$

d) $P(X_5 = 0) = \left(\frac{11}{12}\right)^n \Rightarrow P(X_5 = 1) = 1 - \left(\frac{11}{12}\right)^n$

Let Y_5 be indicator that elevator stops on floor 5.

$$Y_5 \sim \text{Bin}\left(1, 1 - \left(\tfrac{11}{12}\right)^n\right)$$

Similarly, let Y_3 be indicator elevator stops on floor 3.

$$Y_3 \sim \text{Bin}\left(1, 1 - \left(\tfrac{11}{12}\right)^n\right) \quad \text{------} \quad Y_{15} \sim \text{Bin}\left(1, 1 - \left(\tfrac{11}{12}\right)^n\right)$$

$Y_3, Y_4, \dots, Y_{15}$ are not independent, but it doesn't matter for our calculation.

$$E(\#\text{ stops before 11th floor}) = E(Y_3 + Y_4 + Y_5 + \dots + Y_{10})$$
$$= E(Y_3) + E(Y_4) + \dots + E(Y_{10})$$
$$= 8\,E(Y_5)$$
$$= \boxed{8\left[1 - \left(\tfrac{11}{12}\right)^n\right]}$$

5 e $N|T=t \sim Pois(\lambda t)$

$$T \sim Exp(\beta)$$

$$E(N) = E(E(N|T)) = E(\lambda T) = \boxed{\lambda\beta}$$

$$\begin{aligned} Var(N) &= E(Var(N|T)) + Var(E(N|T)) \\ &= E(\lambda T) + Var(\lambda T) \\ &= \boxed{\lambda\beta + \lambda^2\beta^2} \end{aligned}$$

f In part (d) we calculated the expected number of stops. That information together with the time per stop (say 10 seconds) and the time to go up a floor (say 1 second) could allow us to compute the expected time to reaching the 11th floor. It would be something like

$$8\left[1-\left(\tfrac{11}{12}\right)^n\right]\cdot \underbrace{10}_{\text{time per stop}} + \underbrace{9}_{\text{1 second for each floor}} \approx 55 \text{ seconds.}$$

In part (e) we computed the expected number of people on the next elevator before that elevator arrives is $(E(N) = \lambda\beta)$ and the expected time $E(T) = \beta$. Therefore, the expected time to getting to the 11th floor if you wait is something like

$$\beta + E\left[8\left[1-\left(\tfrac{11}{12}\right)^N\right]\right]\cdot 10 + 9$$

which is less than or equal to $\beta + 8\left[1-\left(\tfrac{11}{12}\right)^{\lambda\beta}\right]\cdot 10 + 9$

by Jensen's Inequality. I'd then need values of β and λ and compare expressions. Suppose $\beta = 5$ (next elevator comes in average of 5 seconds), and $\lambda = \frac{1}{5}$ (on average, 1 person shows up every 5 seconds), then →

the expected time if you wait for the next elevator is less than

$$5 + 8\left(1 - \left(\frac{11}{12}\right)^{1}\right) \cdot 10 + 9 \approx 21 \text{ seconds}$$

Definitely better to wait.

20	
19	
18	0 3 4
17	
16	0 3 4 6 9
15	1 5 6 8
14	
13	8
12	
11	
10	
9	
8	
7	8