Basic Probability Summer 2009
NYU Courant Institute
SOLUTIONS to the Practice Problems for the Final Exam

1. Let X be a continuous random variable with a density function which is symmetric about 0. Show that $E(X) = 0$.

Solution: Since $f(x)$ is symmetric about 0, we have $f(-x) = f(x)$. This means $xf(x)$ is an odd function. It immediately follows that

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = 0.$$

2. Each morning a person leaves her house and goes for a run. She is equally likely to leave either from the front door or the back door. Upon leaving she chooses a pair of running shoes from among those sitting by the door, or goes running barefoot if there are no shoes there. On her return, she is equally likely to enter by the front door or the back, and she removes her shoes, if she is wearing any , and leaves them by the door. If she owns a total of n pairs of running shoes, what proportion of the time does she run barefoot? (Hint: set this up as a Markov chain and compute the stationary distribution.)

Solution: For the sake of simplicity, let's start by assuming that she owns a total of 2 pairs of shoes. The general case is similar. Define a Markov chain where the states are $S = \{0, 1, 2\}$ referring to the number of pairs of shoes at the front door. Thus X_k is the number of pairs of shoes at the front door after the kth run. Now we see that the Markov matrix is:

$$\begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/4 & 3/4 \end{pmatrix}$$

The stationary distribution $\pi = (\pi_0, \pi_1, \pi_2)$ is computed by solving the equations $\pi P = \pi$ and $\pi_0 + \pi_1 + \pi_2 = 1$. We get $\pi = (1/3, 1/3, 1/3)$. This means that 1/3 of the time there are no shoes at the front door and also one third of the time there are no shoes at the back door. So 2/3 of the time, there are no shoes at one of the doors. Each time there is a door with no shoes at it, there is a probability of 1/2 she will run barefoot. So she runs barefoot a total of 1/3 of the time.

For the general case, assume she has n pairs of shoes. Again, let X_k be the number of shoes left at the front door after the kth run, so $S = \{0, 1, \ldots, n\}$. The Markov matrix is an $(n+1) \times (n+1)$ matrix. The first row consists of 3/4 followed by 1/4 followed by zeros. The second row consists of 1/4, followed by 1/2, followed by 1/4, followed by zeroes. The next row is the same as the previous row, but starting one further to the right. Finally the last row is all zeros, followed by 1/4, followed by 3/4.

Solving the equations we get that $\pi = (\frac{1}{n+1}, \frac{1}{n+1}, \ldots, \frac{1}{n+1})$ so she is barefoot $\frac{1}{n+1}$ of the time.

3. Let $X_1, X_2, \ldots, X_{100}$ be independent random variables with the common density $f(x) = 2 - 2x$, $0 \leq x \leq 1$. Let $S = X_1 + X_2 \ldots X_{100}$. Use the Central Limit Theorem to estimate $P(S \leq 35)$.

Solution: We have for each i:

$$E(X_i) = \int_0^1 x(2-2x)\, dx = \frac{1}{3},$$

$$E(X_i^2) = \int_0^1 x^2(2-2x)\, dx = \frac{1}{6}, \text{ and}$$

$$Var(X_i) = \frac{1}{6} - (\frac{1}{3})^2 = \frac{1}{18}.$$

Thus $E(S) = \frac{100}{3}$ and the standard deviation of S is $\frac{10}{3\sqrt{2}}$. By the Central Limit Theorem, S is approximately normal, and we get

$$P(S \leq 35) \approx P(Z \leq \frac{35 - \frac{100}{3}}{\frac{10}{3\sqrt{2}}}) = P(Z \leq .7071) = .7602.$$

4. Compute the moment generating function of a geometric random variable, and use it to compute the mean and the variance.

Solution: Let X be geometric with parameter p:

$$M(t) = \Sigma_{x=1}^{\infty} e^{tx} p(1-p)^{x-1} = \frac{pe^t}{1-(1-p)e^t}.$$

Differentiating and evaluating at 0 gives $E(X) = 1/p$ and $Var(X) = \dfrac{1-p}{p^2}$.

5. In a certain community, 4 percent of all adults over the age of 50 have tuberculosis (T.B.). A health service in this community correctly diagnoses 98 percent of all persons with T.B. as having the disease, and incorrectly diagnoses 3 percent of all persons without T.B. as having the disease. Find the probabilities that,

a) the community health service will diagnose an adult over 50 as having T.B.,

Solution: Let PD = positive diagnosis, ND = negative diagnosis, TB = has TB, NTB = doesn't have TB. Then we have $P(PD|TB) = .98$, $P(PD|NTB) = .03$, $P(TB) = .04$, and $P(NTB) = .96$. and

$$P(PD) = P(PD|TB)P(TB) + P(PD|NTB)P(NTB) = (.98)(.04) + (.03)(.96) = .068.$$

b) a person over 50 diagnosed by the health service as having T.B. actually has the disease.

$$P(TB|PD) = \frac{P(TB \cap PD)}{P(PD)} = \frac{P(PD|TB)P(TB)}{P(PD)} = \frac{(.98)(.04)}{.068} = .5765.$$

6. Suppose the joint probability density of X and Y is given by

$$f(x,y) = \begin{cases} 24y(1-x-y) & \text{for } x > 0,\ y > 0,\ x+y < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

a) find the marginal density of X.

Solution: the joint density is zero except on the triangle bounded by the x-axis, the y-axis, and the line $x + y = 1$.

$$f_X(x) = \int_0^{1-x} 24y(1-x-y)\,dy = 24[\frac{y^2}{2}(1-x)-\frac{y^3}{3}]_0^{1-x} = 4(1-x)^3,\ 0 \le x \le 1.$$

b) find the marginal density of Y.

$$f_Y(y) = \int_0^{1-y} 24y(1-x-y)\,dx = 24y[x-\frac{x^2}{2}-yx]_0^{1-y} = 12y^3-24y^2+12y,\ 0 \le y \le 1.$$

c) determine if the two variables are independent.

From parts a) and b) above we see that $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$ so they are not independent.

7. Let X have variance σ^2 and let $m_i = E(X^i)$ denote the ith moment. The *skewness* of the random variable X is defined to be

$$\text{skw}(X) = E((X-m_1)^3)/\sigma^3.$$

a) Show that

$$\text{skw}(X) = \frac{m_3 - 3m_1m_2 + 2m_1^3}{\sigma^3},$$

b) Compute the skewness of an exponential variable with parameter λ and show that it doesn't depend on λ. (The best way to compute the moments is from the moment generating function.)

Solution: $(X-m_1)^3 = X^3 - 3X^2m_1 + 3Xm_1^2 - m_1^3$ so

$$\begin{aligned} E((X-m_1)^3) &= E(X^3 - 3X^2m_1 + 3Xm_1^2 - m_1^3) \\ &= E(X^3) - 3E(X^2m_1) + 3E(Xm_1^2) - E(m_1^3) \\ &= m_3 - 3m_1m_2 + 3m_1m_1^2 - m_1^3 = m_3 - 3m_1m_2 + 2m_1^3 \end{aligned}$$

and the result follows.

b) Compute the skewness of an exponential variable with parameter λ and show that it doesn't depend on λ. (The best way to compute the moments is from the moment generating function.)

If X is exponential with parameter λ, then $M_X(t) = \dfrac{\lambda}{\lambda - t}$. From this we get the moments $m_1 = \dfrac{1}{\lambda}$, $m_2 = \dfrac{2}{\lambda^2}$, and $m_3 = \dfrac{6}{\lambda^3}$. Substituting into the above formula we get $Skew(X) = 2$.

8. Suppose X is a Poisson variable, with parameter λ. For which value of k is $P(X = k)$ the greatest? Hint: compare successive values, i.e. look at

$$\frac{P(X = k)}{P(X = k-1)}.$$

Solution:

$$\frac{P(X = k)}{P(X = k-1)} = \frac{\lambda^k e^{-\lambda}}{k!} \frac{(k-1)!}{\lambda^{k-1} e^{-\lambda}} = \frac{\lambda}{k}.$$

So, if $k < \lambda$ the function is increasing, and if $k > \lambda$ the function is decreasing. So the maximum occurs at $k = [[\lambda]]$, the greatest integer less than or equal to λ.

9. Suppose you play a series of 100 independent games. If you win a game, you win 4 dollars. If you lose a game, you lose 4 dollars. The chances of winning each game is 1/2. Use the central limit theorem to estimate the chances that you will win more than 50 dollars.

Solution. Let X_i be a random variable which is 4 with probability 1/2 and -4 with probability 1/2. Then for each i:

$$E(X_i) = 0,$$

$$E(X_i^2) = Var(X_i) = 16.$$

Letting $S = X_1 + \ldots X_{100}$ Thus $E(S) = 0$, the variance of S is 1600, and the standard deviation of S is 40. By the Central Limit Theorem, S is approximately normal, and we get

$$P(S > 50) \approx P(Z > \frac{50 - 0}{40}) = P(Z > 1.25) = .1057.$$

10. Let X and Y be continuous random variables, having joint probability density function

$$f(x,y) = \begin{cases} 24xy & \text{for } 0 < x < 1,\ 0 < y < 1,\ 0 < x+y < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Find the probability density of $Z = X + Y$.

Solution: There are various ways to do this. One way is to first compute the cdf of Z, and then differentiate to get the pdf:

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = \int\int_R f(x,y)\,dxdy$$

where R is the triangular region bounded by the x-axis, the y-axis, and the line $x + y = z$. We get

$$\int\int_R f(x,y)\,dxdy = \int_0^z \int_0^{z-y} 24xy\,dx\,dy = z^4.$$

Now differentiate,

$$f_Z(z) = F'(z) = 4z^3.$$