

STAT201A – Sec. 102  
Homework #3.

Steven Pollack  
24112977

**#1.** For  $1 \leq i \leq n$ , let  $U_i$  be i.i.d. uniform on the unit interval. Let  $M = \min \{U_1, U_2, \dots, U_n\}$ . Find the distribution, expectation and SD of  $M$ . On the same horizontal axis, sketch (roughly, by hand) the density of  $M$  for  $n = 10$  and  $n = 100$ . Give a brief intuitive explanation for how the two densities differ.

[Hint: to find the distribution, recall what we said in lecture about finding probabilities associated with extrema. To connect with the text, notice that this problem is 2.14.7 with many bells and whistles attached. The hint for that problem should seem very familiar.]

*Proof.* For  $U_i \stackrel{iid}{\sim} \text{Uniform}([0, 1])$ ,  $1 \leq i \leq n$ , if  $M = \min \{U_1, U_2, \dots, U_n\}$ , then

$$P(M > x) = P(U_i > x \text{ for all } 1 \leq i \leq n) = \prod_{i=1}^n P(U_i > x) = (1 - x)^n$$

wherever  $0 \leq x \leq 1$ . If  $x < 0$ ,  $P(M \leq x) = 0$ , and if  $x > 1$ ,  $P(M \leq x) = 1$ . Hence,

$$P(M \leq x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - (1 - x)^n & \text{if } 0 < x \leq 1 \\ 1 & \text{if } 1 < x \end{cases}$$

Now, disregarding continuity issues at  $x = 0, 1$ ,  $dP/dx = f(x) = n(1 - x)^{n-1}I_{[0,1]}$  and thus

$$E(M) = \int_0^1 t f(t) dt = n \int_0^1 t(1 - t)^{n-1} dt$$

One round of integration by parts yields

$$E(M) = \frac{n}{1} \frac{n-1}{2} \int_0^1 t^2 (1 - t)^{n-2} dt$$

Another yields

$$E(M) = \frac{n}{1} \frac{n-1}{2} \frac{n-2}{3} \int_0^1 t^3 (1 - t)^{n-3} dt$$

and a simple induction shows that the  $k^{\text{th}}$  round will provide

$$E(M) = \frac{\prod_{i=0}^{k-1} (n - i)}{\prod_{i=1}^k i} \int_0^1 t^k (1 - t)^{n-k} dt \quad (1 \leq k \leq n)$$

Hence, setting  $k = n$ , we have

$$E(M) = \frac{n(n-1)(n-2) \cdots 1}{1 \cdot 2 \cdot 3 \cdots n} \int_0^1 t^n dt = \frac{1}{n+1}$$

For  $SD(M)$ , we use integration by parts again, playing the exact same game. However, to spare you the gory details, we have the following algebraic result:

$$E(M^2) = \frac{\prod_{i=0}^k (n - i)}{\prod_{i=3}^{k+2} i} \int_0^1 t^{2+k} (1 - t)^{n-k-1} dt \quad (0 \leq k \leq n - 1)$$

Hence, setting  $k = n - 1$  we have

$$E(M^2) = \frac{\prod_{i=0}^{n-1} (n-i)}{\prod_{i=3}^{n+1} i} \int_0^1 t^{n+1} dt = 2 \times \frac{n!}{(n+1)!} \times \frac{1}{n+2} = \frac{2}{(n+1)(n+2)}$$

Combining this result, and the preceding expectation:

$$SD(M) = \sqrt{\text{var}(M)} = \sqrt{\frac{2}{(n+1)(n+2)} - \frac{1}{(n+1)^2}} = \sqrt{\frac{n}{(n+1)^2(n+2)}} = \frac{1}{n+1} \sqrt{\frac{n}{n+2}}$$

□

**#2.** Let  $X$  have the Poisson ( $\mu$ ) distribution. Find  $SD(X)$  by finishing the calculation that we started in class.

*Proof.* Given  $X \sim \text{Poisson}(\mu)$ ,

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{\infty} x^2 e^{-\mu} \frac{\mu^x}{x!} = e^{-\mu} \sum_{x=0}^{\infty} (x(x-1) + x) \frac{\mu^x}{x!} \\ &= e^{-\mu} \left( \sum_{x=2}^{\infty} \frac{\mu^x}{(x-2)!} + \sum_{x=1}^{\infty} \frac{\mu^x}{(x-1)!} \right) \\ &= e^{-\mu} \left( \mu^2 \sum_{x=0}^{\infty} \frac{\mu^x}{x!} + \mu \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \right) \\ &= \mu^2 + \mu \end{aligned}$$

Hence,

$$\text{var}(X) = E(X^2) - E(X)^2 = \mu^2 + \mu - \mu^2 = \mu$$

□

**#3.** There are many close links between the binomial and the Poisson. You've seen one in class. Here's another.

A coin lands heads with probability  $p$ . It is tossed  $N$  times where  $N$  is Poisson with parameter  $\lambda$ . Let  $X$  be the number of heads. Find the distribution of  $X$ . Identify this as a well known distribution and find its parameter or parameters.

[*Hint:* find the possible values, carefully. For each possible value  $x$ , partition the event  $\{X = x\}$  according to the value of  $N$ ; and then use the two great rules: multiplication and addition.]

*Proof.* Because of the variable nature of  $N$  it's clear that the range of  $X$  is all of  $\mathbb{N}$ . To determine the distribution, however, we start with the law of total probability:

$$P(X = x) = \sum_{n=0}^{\infty} P(X = x | N = n) P(N = n)$$

However, it's pretty clear that  $P(X = x | N = n) = 0$  whenever  $n < x$  and, by definition,

$$P(X = x | N = n) = \binom{n}{x} p^x q^{n-x} \quad (x \leq n)$$

So,

$$\begin{aligned} P(X = x) &= \sum_{n=x}^{\infty} P(X = x | N = n) P(N = n) \\ &= \sum_{n=x}^{\infty} \binom{n}{x} p^x q^{n-x} e^{-\lambda} \frac{\lambda^n}{n!} \\ &= p^x q^{-x} e^{-\lambda} \sum_{n=x}^{\infty} \frac{n!}{(n-x)! x!} \frac{(q\lambda)^n}{n!} \\ &= \frac{p^x}{q^x x! e^{\lambda}} \sum_{n=0}^{\infty} \frac{(q\lambda)^{n+x}}{n!} \\ &= \frac{(pq\lambda)^x}{q^x x! e^{\lambda}} \times e^{q\lambda} \\ &= e^{-(1-q)\lambda} \frac{(p\lambda)^x}{x!} = e^{-p\lambda} \frac{(p\lambda)^x}{x!} \end{aligned}$$

Which demonstrates that  $X \sim \text{Poisson}(p\lambda)$ . □

**#4.** A coin lands heads with probability  $p$ . If it is tossed  $n$  times, where  $n$  is a fixed number, then clearly the number of heads and the number of tails are dependent random variables. But as you've seen before, randomizing a parameter can really affect dependence and independence. With that as preamble, do 2.14.11(b).

[Hint: write the event  $\{X = x, Y = y\}$  in terms of  $X$  and  $N$ . The result of Problem 2 will be very useful]

[2.14.11(b) reads: suppose we toss a coin once and let  $p$  be the probability of heads. Let  $X$  denote the number of heads and let  $Y$  denote the number of tails. (b) Let  $N \sim \text{Poisson}(\lambda)$  and suppose we toss a coin  $N$  times. Let  $X$  and  $Y$  be the number of heads and tails. Show that  $X$  and  $Y$  are independent.]

*Proof.* As in exercise #3., we write

$$P(X = x, Y = y) = \sum_{n=0}^{\infty} P(X = x, Y = y | N = n) P(N = n)$$

however, now we note that for  $y \neq n - x$ ,  $P(X = x, Y = y | N = n) = 0$ . Hence,

$$P(X = x, Y = y) = P(X = x, Y = n - x | N = n) P(N = n) = \binom{n}{x} p^x q^{n-x} e^{-\lambda} \frac{\lambda^n}{n!}$$

So, using the fact that  $1 = p + q$ , and  $x + y = n$  we may write

$$\begin{aligned} P(X = x, Y = y) &= \binom{x+y}{x} p^x q^y e^{-\lambda(p+q)} \frac{\lambda^{x+y}}{(x+y)!} \\ &= \frac{(p\lambda)^x e^{-\lambda p}}{x!} \cdot \frac{(q\lambda)^y e^{-\lambda q}}{y!} \\ &= P(X = x) \cdot P(Y = y) \end{aligned}$$

Where we used the result of exercise #3. in the last equation. □

**#5.** (Exercise 3.8.7) This is the “tail integral” formula for expectation, analogous to tail sums in the discrete case. The hint is useful because the basic formula for expectation is the integral of a product. Use the fact that  $f$  is the derivative of  $F$ , and the derivative of  $x$  is a really easy function. . .

[3.8.7 reads: Let  $X$  be a continuous random variable with cdf  $F$ . Suppose that  $P(X > 0) = 1$  and that  $E(X)$  exists. Show that  $E(X) = \int_0^\infty P(X > x) dx$ . *Hint: Consider integrating by parts. The following fact is helpful: if  $E(X)$  exists then  $\lim_{x \rightarrow \infty} x[1 - F(x)] = 0$ .]*

*Proof.* Using the hint, let  $u(x) = 1 - F(x)$  and  $v(x) = x$ , then via integration by parts

$$\begin{aligned} \int_0^\infty 1 - F(x) dx &= x(1 - F(x)) \Big|_{x=0}^\infty - \int_0^\infty x(-f(x)) dx \\ &= (0 - 0) + \int_0^\infty x f(x) dx \\ &= E(X) \end{aligned}$$

Where in the second line, we used the fact that the existence of  $E(X)$  implies<sup>1</sup>

$$\lim_{x \rightarrow \infty} x(1 - F(x)) = 0$$

□

**#6.** For  $1 \leq i \leq n$ , let  $T_i$  be exponential with rate  $\lambda_i$  and assume the  $T_i$ ’s are independent.

a) Use the result of the previous problem to find  $E(T_1)$ .

b) Let  $M = \min \{T_i : 1 \leq i \leq n\}$ . Find the distribution (do you recognize it?) and expectation of  $M$ .

*Proof.* a) Using the result in exercise #5., and the fact that, for  $T \sim \text{Exp}(\lambda)$ ,  $P(T > t) = e^{-t/\lambda}$ , we have

$$E(T_1) = \int_0^\infty P(T_1 > t) dt = \int_0^\infty e^{-t/\lambda_1} dt = \lambda_1 e^{-t/\lambda_1} \Big|_{t=\infty}^0 = \lambda_1$$

(Note: I’m using the function  $f(x) = e^{-x/\lambda}/\lambda$  as the density for  $X \sim \text{Exponential}(\lambda)$  as defined in the back appendix of the book.)

---

<sup>1</sup>This result can be established via l’Hôpital’s rule and an argument that  $xf(x)$  must vanish at infinity for  $E(X) < \infty$ .

b) If  $M = \min \{T_i : 1 \leq i \leq n\}$  and the  $T_i$ 's are independent, then

$$P(M > t) = P(T_1 > t, T_2 > t, \dots, T_n > t) = \prod_{i=1}^n P(T_i > t) = \exp \left\{ -t \sum_{i=1}^n \lambda_i^{-1} \right\}$$

So  $M \sim \text{Exp}$  with rate  $\sum_{i=1}^n \lambda_i^{-1}$  and therefore has expectation  $\sum_{i=1}^n \lambda_i^{-1}$ . □

#7. Here you will derive two ways of computing covariance.

a) Prove the first statement in Theorem 3.19 that says that covariance is the “mean of the products minus the product of the means”. This formula should be treated like the basic formula for expectation and like the “variance is the mean of the square minus the square of the mean”. Use it only when the distributions of  $X, Y$  and  $XY$  are very simple. Occasionally, it is useful to prove some general fact about covariance. But, the key to covariance is part b).

b) (Exercise 3.8.14) but just do the easy case  $m = n = 2$ . The general case is proved by routine algebra that you don't have to do. The result is the most useful computational tool for covariance, and is called “bilinearity of covariance”.

[Exercise 3.8.14 reads: let  $X_1, X_2, \dots, X_n$  and  $Y_1, Y_2, \dots, Y_n$  be random variables and  $a_1, a_2, \dots, a_m$  and  $b_1, b_2, \dots, b_n$  be constants. Show that

$$\text{Cov} \left( \sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j)$$

*Proof.* a) From the definition of  $\text{Cov}(X, Y)$  in 3.18,

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E(XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y) \\ &= E(XY) - E(\mu_Y X) - E(\mu_X Y) + E(\mu_X \mu_Y) \\ &= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y \end{aligned}$$

b) To prove the  $n = m = 2$  case, consider the expansions

$$(a_1 X_1 + a_2 X_2)(b_1 Y_1 + b_2 Y_2) = a_1 b_1 X_1 Y_1 + a_1 b_2 X_1 Y_2 + a_2 b_1 X_2 Y_1 + a_2 b_2 X_2 Y_2$$

and

$$(a_1 \mu_{X_1} + a_2 \mu_{X_2})(b_1 \mu_{Y_1} + b_2 \mu_{Y_2}) = a_1 b_1 \mu_{X_1} \mu_{Y_1} + a_1 b_2 \mu_{X_1} \mu_{Y_2} + a_2 b_1 \mu_{X_2} \mu_{Y_1} + a_2 b_2 \mu_{X_2} \mu_{Y_2}$$

Now, take the expectation of the former identity and subtract the second equation from it grouping terms by coefficients:

$$\begin{aligned} E[(a_1 X_1 + a_2 X_2)(b_1 Y_1 + b_2 Y_2)] - (a_1 \mu_{X_1} + a_2 \mu_{X_2})(b_1 \mu_{Y_1} + b_2 \mu_{Y_2}) \\ = a_1 b_1 (E(X_1 Y_1) - \mu_{X_1} \mu_{Y_1}) + a_1 b_2 (E(X_1 Y_2) - \mu_{X_1} \mu_{Y_2}) \\ + a_2 b_1 (E(X_2 Y_1) - \mu_{X_2} \mu_{Y_1}) + a_2 b_2 (E(X_2 Y_2) - \mu_{X_2} \mu_{Y_2}) \end{aligned}$$

Rewriting the above, using part a), we see that

$$\text{Cov}(a_1X_1 + a_2X_2, b_1Y_1 + b_2Y_2) = \sum_{i=1}^2 \sum_{j=1}^2 a_i b_j \text{Cov}(X_i, Y_j)$$

□

**#8.** Covariance has awful units. For example, if  $X$  is weight in pounds and  $Y$  is height in inches, the units of covariance are “inch pounds”. Nobody understands that. So, it’s a great idea to get rid of the units if possible. Read definitions 3.18. The first is covariance. The second is correlation, which divides covariance by the two SD’s to get rid of those awful units and create a pure number.

The goal of this problem is to prove the second statement in Theorem 3.19, that correlation is between -1 and 1. There are many ways of doing this but I’d like you to use the following steps.

- Let  $X^* = (X - \mu_X)/\sigma_X$  be “ $X$  in standard units”. Show that  $E(X^*) = 0$ ,  $SD(X) = 1$  and  $E(X^{*2}) = 1$ .
- Show that  $\rho(X, Y) = E(X^*Y^*)$ . There’s no real calculation here, just rewrite the definition of correlation in the right way.
- Both  $(X^* + Y^*)^2$  and  $(X^* - Y^*)^2$  are non-negative random variables, and hence so are their expectations. Compute the two expectations and apply the results of the previous two parts to show that  $-1 \leq \rho(X, Y) \leq 1$ .

*Proof.* a) Let  $E(X) = \mu$  and  $\sigma_X^2 = \text{var}(X)$ . Then,

$$E(X^*) = E\left(\frac{X - \mu}{\sigma_X}\right) = \sigma_X^{-1} E(X - \mu) = \sigma_X^{-1} \times 0 = 0$$

$$SD(X^*)^2 = \text{var}(X^*) = \text{var}\left(\frac{X - \mu}{\sigma_X}\right) = \sigma_X^{-2} \text{var}(X - \mu) = \sigma_X^{-2} \times \sigma_X^2 = 1$$

Hence,  $SD(X^*) = 1$  (since  $SD(X) \geq 0$ ). Finally,

$$E(X^{*2}) = E\left(\frac{(X - \mu)^2}{\sigma_X^2}\right) = \sigma_X^{-2} E[(X - \mu)^2] = \sigma_X^{-2} \times \sigma_X^2 = 1$$

b) Using the definition of  $\rho$  given in 3.12,

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = E\left[\frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y}\right] = E(X^*Y^*)$$

- c) Consider  $(X^* + Y^*)^2$ . Since this random variable is non-negative, its expectation is non-negative and we have

$$0 \leq E[(X^* + Y^*)^2] = E(X^{*2} + 2X^*Y^* + Y^{*2}) = 2 + 2\rho(X, Y)$$

Hence,  $-1 \leq \rho(X, Y)$ . Similarly,

$$0 \leq E[(X^* - Y^*)^2] = E(X^{*2} - 2X^*Y^* + Y^{*2}) = 2 - 2\rho(X, Y)$$

which implies that  $\rho(X, Y) \leq 1$ . Putting the two inequalities together, we have

$$-1 \leq \rho(X, Y) \leq 1$$

□