Statistics 394, Problem Set 6 Solutions

Wellner; 2/16/2000

1. K 3.4, # 2 (page 188). A fair coin is tossed four times; X is the number of heads that come up on the first three tosses, and Y is the number of heads that come up on tosses $2, 3$, and 4.
(a) What is the distribution of X? Of Y?
(b) Make a table of the join mass function of X and Y. [It has 16 entries.]
(c) Find the marginal mass functions and check that they agree with your answer to part a.
(d) Are X and Y independent?
(e) Find the conditional probability that $X = x$, given that $Y = 2$, for $x = 0, 1, 2, 3$.

Solution: (a) $X \sim Binomial(3, 1/2)$ and $Y \sim Binomial(3, 1/2)$: $P(X = x) = \binom{3}{x}(1/2)^3$ for $x \in \{0, 1, 2, 3\}$; thus $P(X = x) = 1/8, 3/8, 3/8, 1/8$ for $x = 0, 1, 2, 3$.
(b) The joint distribution of X and Y is given in the following table:

x/y	0	1	2	3	$p_X(x)$
0	1/16	1/16			2/16
1	1/16	3/16	2/16		6/16
2		2/16	3/16	1/16	6/16
3			1/16	1/16	2/16
$p_Y(y)$	2/16	6/16	6/16	2/16	1

(c) The marginals are given above. Note that they agree with the Binomial distribution in (a).
(d) X and Y are not independent:

$$P(X = 3, Y = 0) = 0 \neq (1/8)(1/8) = P(X = 3)P(Y = 0)\,.$$

(e) The conditional distribution of X given $Y = 2$ is given by

$$P(X = x|Y = 2) = \frac{P(X = x, Y = 2)}{P(Y = 2)} = \begin{cases} 0, & x = 0 \\ 2/6, & x = 1 \\ 3/6, & x = 2 \\ 1/6, & x = 3\,. \end{cases}$$

2. K 3.4, # 5 (page 188). Suppose that X and Y are independent Poisson random variables, with parameters 3 and 4 respectively. Find the probability that $X+Y = 5$.

Solution: Now

$$\begin{aligned}
P(X+Y=5) &= \sum_{x=0}^{5} P(X=x, Y=5-x) \\
&= \sum_{x=0}^{5} P(X=x)P(Y=5-x) \qquad \text{by independence of } X, Y \\
&= \sum_{x=0}^{5} e^{-3}\frac{3^x}{x!}e^{-4}\frac{4^{5-x}}{(5-x)!} \\
&= \frac{1}{5!}e^{-7}\sum_{x=0}^{5}\frac{5!}{x!(5-x)!}3^x 4^{5-x} \\
&= \frac{1}{5!}e^{-7}(3+4)^5 \qquad \text{by the binomial formula} \\
&= \frac{7^5}{5!}e^{-7}.
\end{aligned}$$

Note that this is $P(Poisson(7) = 5)$, as we would deduce immediately from the Poisson process interpretation of this problem, and in fact, if $X \sim Poisson(\lambda)$ and $Y \sim Poisson(\mu)$ are independent, then $X + Y \sim Poisson(\lambda + \mu)$ is always true.

3. K 3.4, # 10 (page 189). Breakdowns occur in a factory according to a Poisson process, with an average of λ breakdowns per week. The breakdowns are classified as damaging or nondamaging, depending on whether they hold up production for more than 5 minutes. The fraction of breakdowns that are damaging is p; that is, p is the probability that a given breakdown is damaging. Let X be the number of breakdowns in a given week, and let Y be the number of damaging breakdowns in that same week. We know the distribution of X; it is Poisson with parameter λ; we want the distribution of Y.
(a) Given that $X = x$, assume that the x breakdowns represent independent trials; the conditional probability that $Y = y$ given $X = x$ is the probability that y of the x breakdowns are damaging. Write down this probability.
(b) Find the joint mass function $p(x, y) = P(X = x, Y = y)$.
(c) For a fixed y, sum the mass function over all $x = y, y + 1, \ldots$, to get the unconditional mass function of Y.

Solution: (a) Since the damaging breakdowns occur independently with probability p, the total number of them in x trials has a Binomial(x, p) distribution. Hence

$$P(Y = y|X = x) = \binom{x}{y} p^y (1-p)^{x-y} \quad \text{for} \quad y = 0, \ldots, x\,.$$

(b) Since the marginal distribution of X is Poisson(λ), we have $P(X = x) = e^{-\lambda}\lambda^x/x!$ for $x = 0, 1, \ldots$. Thus the joint probability mass function of X, Y is given by

$$\begin{aligned}
p(x,y) &\equiv P(X=x, Y=y) = P(Y=y|X=x)P(X=x) \\
&= \binom{x}{y} p^y (1-p)^{x-y} e^{-\lambda}\frac{\lambda^x}{x!}
\end{aligned}$$

for $y \in \{0, \ldots, x\}$, $x = 0, 1, \ldots$.
(c) The marginal distribution of Y is obtained from the joint distribution of (X, Y) by summing over the possible values x of X: thus we compute

$$\begin{aligned} p_Y(y) &= \sum_{x=y}^{\infty} P(X = x, Y = y) \\ &= \sum_{x=y}^{\infty} \binom{x}{y} p^y (1-p)^{x-y} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \frac{\lambda^y p^y}{y!} e^{-\lambda} \sum_{x=y}^{\infty} \frac{(1-p)^{x-y} \lambda^{x-y}}{(x-y)!} \\ &= \frac{\lambda^y p^y}{y!} e^{-\lambda} \sum_{k=0}^{\infty} \frac{((1-p)\lambda)^k}{k!} \\ &= \frac{\lambda^y p^y}{y!} e^{-\lambda} e^{\lambda(1-p)} \\ &= \frac{(\lambda p)^y}{y!} e^{-\lambda p} . \end{aligned}$$

Thus we deduce that $Y \sim Poisson(\lambda p)$. (This result is connected with *thinning* of a Poisson process: if we start with a Poisson process with intensity ν if events are "thinned" (marked as "damaging breakdowns") independently with probability p, then the new process is a Poisson process with intensity νp.)

4. K 3.4, # 11 (page 189). Let X and Y be independent, each with the mass function $p(k) = pq^k$, $k = 0, 1, 2, \ldots$. (This is the geometric mass function with parameter p; as usual $0 < p < 1$ and $q = 1-p$.) [In our Bernoulli process notation, $X' \equiv X+1$, which has the mass function $p'(k) = pq^{k-1}$, $k = 1, 2, \ldots$, has a Geometric(p) distribution!] Let $Z = \min X, Y$. We want to find the mass function of Z.
(a) For $n = 0, 1, 2, \ldots$, find $P(X \geq n)$.
(b) For $n = 0, 1, 2, \ldots$, find $P(Z \geq n)$.
(c) Using the result of part (b), find $P(Z = k)$.

Solution: (a)

$$P(X \geq n) = \sum_{k=n}^{\infty} pq^k = pq^n(1 + q + q^2 + \cdots) = q^n .$$

(b) Since $[Z \geq n] = [X \geq n, Y \geq n]$,

$$P(Z \geq n) = P(X \geq n, Y \geq n) = P(X \geq n)P(Y \geq n) = q^n q^n = (q^{2n}) = (q^2)^n .$$

(c) In general for an integer-valued random variable Z, $P(Z = k) = P(Z \geq k) - P(Z \geq k+1)$. In our particular case,

$$\begin{aligned} P(Z = k) &= P(Z \geq k) - P(Z \geq k+1) \\ &= (q^2)^k - (q^2)^{k+1} = (q^2)^k (1 - q^2) , \quad k = 0, 1, 2, \ldots . \end{aligned}$$

Thus $Z' \equiv Z + 1 \sim$Geometric$(1 - q^2)$.

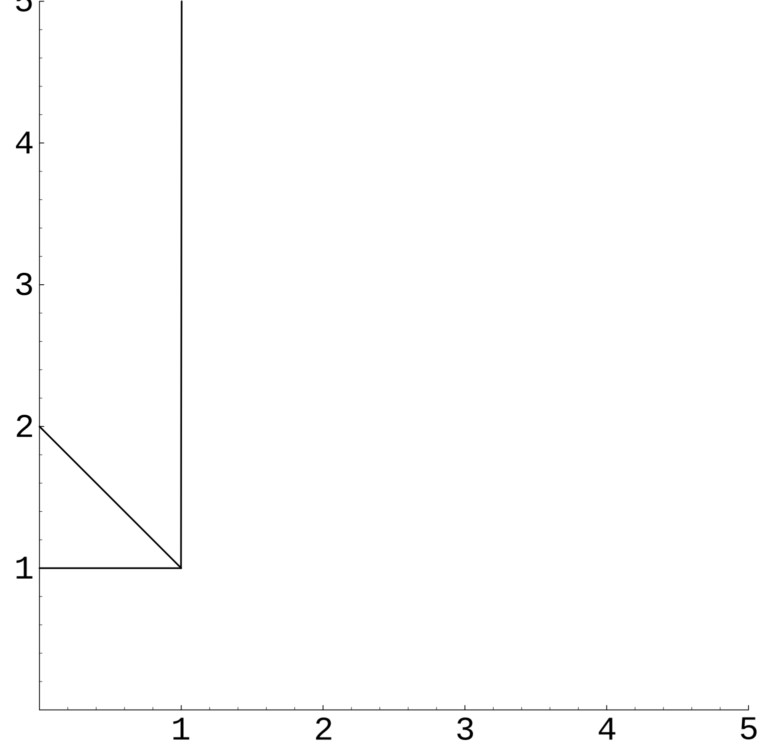

Figure 1: Regions for Problem 5.

5. K 3.5, # 4 (page 206). Let X and Y be independent, with $f_X(x) = 2x1_{(0,1)}(x)$, and $f_Y(y) = (2/y^3)1_{(1,\infty)}(y)$.
(a) Draw a picture of the set of possible values of the pair (X, Y).
(b) Write down the joint density.
(c) Find $P(X + Y \le 2)$.

Solution: (a) The possible values of (X, Y) are all in the set $\{(x, y) \in R^2 : \quad 0 < x < 1, 1 < y < \infty\}$. This is a strip of width 1 in the x direction, and going from 1 to infinity in the y direction: see Figure 1.

(b) Since X and Y are independent random variables, the joint density of (X, Y) is simply the product $f_X(x) \cdot f_Y(y)$:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{4x}{y^3} 1_{(0,1)\times(1,\infty)}(x, y) .$$

(c) The event $[X + Y \le 2]$ is shown by the shaded region in Figure 1. Thus we compute

$$\begin{aligned}
P(X + Y \le 2) &= \int\int_{x+y\le 2} f_{X,Y}(x, y) dx dy \\
&= 2\int_{y=1}^{2} \left(\int_{x=0}^{2-y} \frac{4x}{y^3} dy \right) dx \\
&= 2\int_{y=1}^{2} (2 - y)^2 y^{-3} dy \\
&= 2\int_{y=1}^{2} \left(\frac{4}{y^3} - \frac{4}{y^2} + \frac{1}{y} \right) dy \\
&= 2\left\{2(1 - 1/4) - 4(1 - 1/2) + \log(2)\right\} \\
&= 2\log(2) - 1 \doteq 0.3863 .
\end{aligned}$$

6. K 3.5, # 8 (page 207). Let X and Y be independent, each with the uniform distribution on the interval $[0,1]$. In the following calculations you do not need to do any integrals, because the set of possible values is a square in the plane and the joint density equals 1 on the square, and so integrals are just areas.
(a) Write down their joint density.
(b) Find $P(2X \leq Y)$.
(c) For a number a between 0 and 1, find the probability that the larger of X and Y is less than or equal to a. This is the CDF of $M = \max\{X,Y\}$. Differentiate it to get a density for M.
(d) Find $P(X+Y \leq a)$ for a number a between 0 and 1.
(e) Find $P(X+Y \leq a)$ for a number a between 1 and 2.
(f) Combine the results of parts (d) and (e) to get the CDF of $Z = X+Y$. Differentiate to get a density for Z.

Solution: (a) Since X and Y are independent, the joint density is just the product of the marginal densities:

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)f_Y(y) = 1_{(0,1)}(x)1_{(0,1)}(y) \\ &= 1_{(0,1)\times(0,1)}(x,y) \\ &= \begin{cases} 1, & \text{if}(x,y) \in (0,1)^2 \times (0,1), \\ 0, & \text{if}(x,y) \notin (0,1)^2 \times (0,1)\,. \end{cases} \end{aligned}$$

(b) $P(2X \leq Y)$ is just the area of the triangle lying above the line $y = 2x$. Since this triangle has area $(1/2)(1)(1/2) = 1/4$, $P(2X \leq Y) = 1/4$.
(c) With $M = \max\{X,Y\}$,

$$\begin{aligned} F_M(a) = P(M \leq a) &= P(\max\{X,Y\} \leq a) = P(X \leq a, Y \leq) \\ &= \text{Area of the square } \ [0,a] \times [0,a] = a^2\,. \end{aligned}$$

for $0 \leq a \leq 1$. Thus $f_M(a) = 2a1_{[0,1]}(a)$.
(d) For a number a between 0 and 1, the event $[X+Y \leq a]$ is just the triangle bounded by the x and y axes and the line $y = a - x$. This triangle has area $a^2/2$, and hence we have

$$F_{X+Y}(a) = P(X+Y \leq a) = a^2/2 \ \text{ for } \ 0 \leq a \leq 1\,.$$

(e) For a number a between 1 and 2, the event $[X+Y \leq a]$ has complement $[X+Y > a]$, which is a triangle bounded by the lines $x = 1$, $y = 1$, and the line $y = a - x$; see Figure 3. The two sides are of length $2-a$, so the area of the triangle is $(2-a)^2/2$. Thus

$$F_{X+Y}(a) = P(X+Y \leq a) = 1 - P(X+Y > a) = 1 - (2-a)^2/2$$

for $1 \leq a \leq 2$. Differentiating the distribution function we find that the density of f_{X+Y} is given by

$$f_{X+Y}(a) = \begin{cases} a, & \text{for } 0 < a < 1 \\ 2-a, & \text{for } 1 \leq a < 2\,. \end{cases}$$

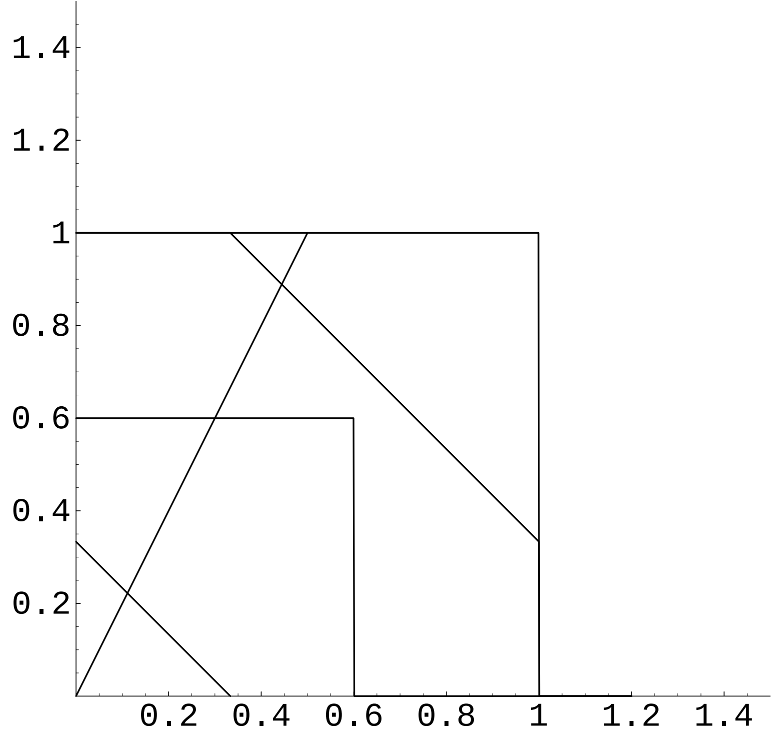

Figure 2: Regions for Problem 6.

Note that this is a triangular density on $[0, 2]$; this verifies a claim made in a lecture at the beginning of the quarter.

7. Bonus Problem: Look at the Bivariate Uniform Experiment at the Virtual Laboratories website:

`http://www.math.uah.edu/stat/dist/index.html`

(a) For the triangle part of this experiment, give a picture showing the region where the density is positive and where it is zero.
(b) Find the conditional density of $(Y|X = x)$ for x in the range -6 to 6.
(c) Verify that the blue line in the picture gives $r(x) \equiv E(Y|X = x)$.

Solution: (a) The density is positive on the triangle $T \equiv \{(x, y) : -6 < y < x < 6\}$; see Figure 3

(b) Since (X, Y) is distributed uniformly over T and the area of T is $12 \cdot 12/2 = 144/2 = 72$, the joint density is $f_{X,Y}(x, y) = (1/72)1_T(x, y)$. Thus the marginal density of X is

$$f_X(x) = \frac{1}{72}\int_{-6}^{x} dy = \frac{1}{72}(x + 6)1_{(-6,6)}(x) .$$

Note that the corresponding marginal distribution function is

$$F_X(x) = \int_{-\infty}^{x} f_X(x')dx' = \int_{-6}^{x} (1/72)(x' + 6)dx' = (x + 6)^2/144$$

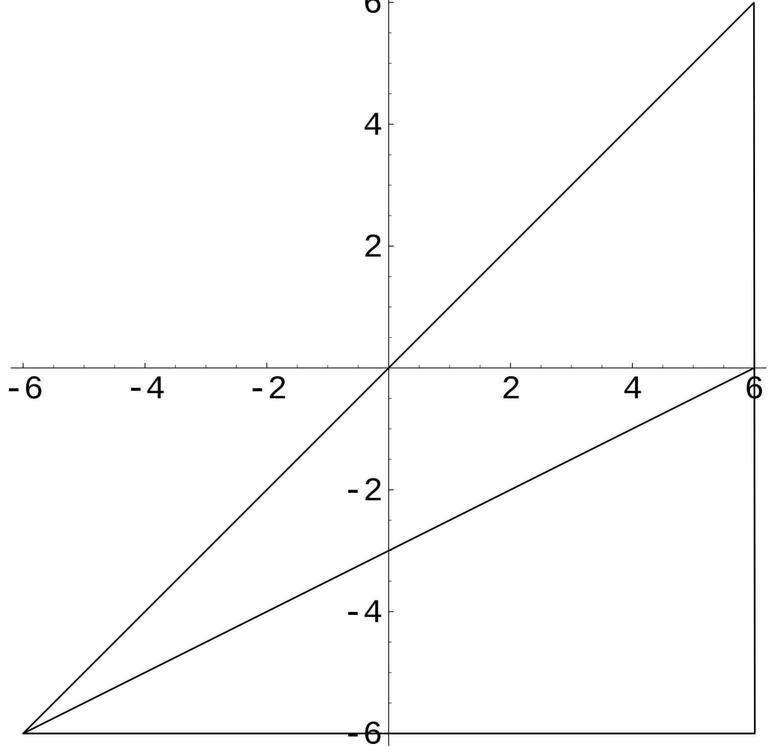

Figure 3: Regions for Bonus Problem 7.

for $-6 \leq x \leq 6$ which equals 1 when $x = 6$. Thus the conditional density of Y given $X = x$ is given by

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{x+6} 1_{(-6,x)}(y) \, ,$$

and this is just the uniform density on the interval $(-6, x)$: $(Y|X = x) \sim Uniform(-6, x)$.

(c)

$$r(x) = E(Y|X = x) = \frac{x-6}{2}, \qquad -6 < x < 6 \, .$$

This is exactly the function plotted as a blue line in the upper left for the experiment referred to above for the "triangle" version of the experiment. (Note that this is just a translation and scaling of the computation I did in class on 2/9 and 2/11: if we call the variables here (X', Y') and my variables (X, Y), then $X' = 12X - 6$, $Y' = 12Y - 6$. Since I computed $E(Y|X = x) = x/2$, it follows easily that

$$\begin{aligned} E(Y'|X' = x') &= E(12Y - 6|X = (x'+6)/12) \\ &= 12E(Y|X = (x'+6)/12) - 6 \\ &= (x'+6)/2 - 6 = (x'-6)/2 \, , \end{aligned}$$

in agreement with the answer obtained above by direct computation.