Mathematics 60 - Probability Theory

Carter Kruse

Homework 5

Chapter 5.1: 32, 35, 39

Chapter 5.2: 2, 10, 31, 34, 37

Chapter 6.1: 7, 31

Chapter 5.1, Question 32

It is often assumed that the auto traffic that arrives at the intersection during a unit time period has a Poisson distribution with expected value m. Assume that the number of cars X that arrive at an intersection from the north in unit time has a Poisson distribution with parameter $\lambda = m$ and the number Y that arrive from the west in unit time has a Poisson distribution with parameter $\lambda = \bar{m}$. If X and Y are independent, show that the total number $X + Y$ that arrive at the intersection in unit time has a Poisson distribution with parameter $\lambda = m + \bar{m}$.

Let us say $Z = X + Y$, so we claim Z has a Poisson distribution with parameter $\lambda = m + \bar{m}$. We start by recognizing that

$$P\left(Z = z\right) = \sum_{j=0}^{z} P\left(X = j \ \ \& \ \ Y = z - j\right).$$

This ensures that $X + Y = z$. Following from this, as we know that X and Y are independent, we have

$$\sum_{j=0}^{z} P\left(X = j\right) P\left(Y = z - j\right).$$

Using the definition of Poisson distribution, which is

$$P\left(X = k\right) \approx \frac{\lambda^k}{k!} e^{-\lambda},$$

we have

$$\sum_{j=0}^{z} \left(\frac{m^j}{j!} e^{-m}\right) \left(\frac{\bar{m}^{(z-j)}}{(z-j)!} e^{-\bar{m}}\right).$$

When simplifying, this becomes

$$\sum_{j=0}^{z} \left(\frac{z!}{j!\left(z - j\right)!}\right) \left(\frac{\left(e^{-m} m^j\right)\left(e^{-\bar{m}} \bar{m}^{(z-j)}\right)}{z!}\right).$$

Using the form of binomial coefficients and factoring out $z!$ and $e^{-m} e^{-\bar{m}} = e^{-(m+\bar{m})} = e^{-\lambda}$, we have

$$\left(\frac{e^{-\lambda}}{z!}\right) \sum_{j=0} \binom{z}{j} m^j \bar{m}^{(z-j)}.$$

By using the reverse of binomial expansion, this reduces to

$$\left(\frac{e^{-\lambda}}{z!}\right)(m+\bar{m})^z\,,$$

which is equivalent to

$$\frac{e^{-\lambda}\lambda^z}{z!}.$$

Thus, we have demonstrated that if X and Y are independent, the total number $X+Y$ that arrive at the intersection in unit time has a Poisson distribution with parameter $\lambda = m + \bar{m}$.

Chapter 5.1, Question 35

A manufactured lot of brass turnbuckles has S items of which D are defective. A sample of s items is drawn without replacement. Let X be a random variable that gives the number of defective items in the sample. Let $p(d) = P(X = d)$.

(a) Show that

$$p(d) = \frac{\binom{D}{d}\binom{S-D}{s-d}}{\binom{S}{s}}.$$

Thus, X is hypergeometric.

In order to have d defective items in s items, you must choose d items out of D defective ones and the rest from $S-D$ good ones. In this situation, 'the rest' is given by the expression $s-d$, as s represents the total number of items in the sample. This gives us the numerator of the expression above.

The total number of samples points is the number of ways to choose s out of S. This gives us the denominator of the expression above.

Thus, we have shown that

$$p(d) = P(X = d) = \frac{\binom{D}{d}\binom{S-D}{s-d}}{\binom{S}{s}}$$

and so X is hypergeometric.

(b) Prove the following identity, known as *Euler's formula*:

$$\sum_{d=0}^{\min(D,s)} \binom{D}{d}\binom{S-D}{s-d} = \binom{S}{s}.$$

If we sum up the probability $P(X = j)$ over all the values of j, the value should equal one. This is expressed as

$$\sum_{j=0}^{\min(D,s)} P(X = j) = 1.$$

Thus, as we know

$$P(X = j) = \frac{\binom{D}{j}\binom{S-D}{s-j}}{\binom{S}{s}},$$

we find

$$\sum_{j=0}^{\min(D,s)} \binom{D}{j}\binom{S-D}{s-j} = \binom{S}{s}.$$

Chapter 5.1, Question 39

Suppose that N and k tend to ∞ in such a way that k/N remains fixed. Show that

$$h(N,k,n,x) = b(n,k/N,x).$$

The hypergeometric distribution with parameters N,k,n,x with N and k tending to ∞ in such a way that k/N remains fixed has a distribution as follows:

$$P(X=x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

which may be expressed as

$$\frac{k!}{(k-x)!x!} \cdot \frac{(N-k)!}{(n-x)!(N-k-n+x)!} \cdot \frac{n!(N-n)!}{N!}.$$

Color coding is used to keep track of where the various terms in the next couple of steps are coming from. The above expression simplifies to

$$\binom{n}{x}\frac{k!}{(k-x)!} \cdot \frac{(N-k)!}{(N-k-n+x)!} \cdot \frac{(N-n)!}{N!},$$

which may be alternatively expressed as

$$\binom{n}{x}\frac{k(k-1)(k-2)\cdots(k-x+1)}{N(N-1)(N-2)\cdots(N-x+1)} \cdot \frac{(N-k)(N-k-1)\cdots(N-k-(n-x)+1)}{(N-x)(N-x-1)\cdots(N-n-1)}.$$

In the last denominator, we recognize that $N-n+1 = N-x-(n-x)+1$. Now we take the limit, noticing that

$$\frac{k(k-1)(k-2)\cdots(k-x+1)}{N(N-1)(N-2)\cdots(N-x+1)} \approx \frac{k^x}{N^x} = \left(\frac{k}{N}\right)^x,$$

since x is kept finite while k and N diverge. By the same argument, where n is kept finite, we notice

$$\frac{(N-k)(N-k-1)\cdots(N-k-(n-x)+1)}{(N-x)(N-x-1)\cdots(N-n-1)} \approx \frac{(N-k)^{n-x}}{N^{n-x}} = \left(\frac{N-k}{N}\right)^{n-x} = \left(1-\frac{k}{N}\right)^{n-x}.$$

Taking all of these factors into account, we compute the product to be

$$\lim_{N,k\longrightarrow\infty} P(X=x) = \binom{n}{x}\left(\frac{k}{N}\right)^x\left(1-\frac{k}{N}\right)^{n-x}.$$

This demonstrates that $h(N,k,n,x) = b(n,k/N,x)$ when considering that N and k tend to ∞ in such a way that k/N remains fixed. $\square$

Chapter 5.2, Question 2

Choose a number U from the interval $[0,1]$ with uniform distribution. Find the cumulative distribution and density for the random variables.

(a) $Y = 1/(U+1)$

The cumulative distribution function is

$$F_Y(y) = P(Y \le y),$$

which is equivalent to

$$P\left(\frac{1}{U+1} \le y\right).$$

By rearranging the interior expression, we have

$$P\left(U \ge \frac{1}{y} - 1\right),$$

which is equivalent to

$$1 - \left(\frac{1}{y} - 1\right) = 2 - \frac{1}{y}.$$

The bounds of this cumulative distribution function are determined by considering $\left(\frac{1}{y} - 1\right) \in [0, 1]$, so that $y \in \left[\frac{1}{2}, 1\right]$.

Thus, the cumulative distribution for Y is given by

$$F_Y(y) = 2 - \frac{1}{y} \text{ where } \frac{1}{2} \le y \le 1.$$

To calculate the density function of Y we simply differentiate the cumulative distribution function as follows:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} 2 - \frac{1}{y} = \frac{1}{y^2}.$$

Thus, the density for Y is given by

$$f_Y(y) = \frac{1}{y^2}.$$

(b) $Y = \log(U+1)$

The cumulative distribution function is

$$F_Y(y) = P(Y \le y),$$

which is equivalent to

$$P(\log(U+1) \le y).$$

By rearranging the interior expression, we have

$$P(U \le e^y - 1),$$

which is equivalent to

$$e^y - 1.$$

The bounds of this cumulative distribution function are determined by considering $e^y - 1 \in [0, 1]$, so that $y \in [0, \log(2)]$.

Thus, the cumulative distribution for Y is given by

$$F_Y(y) = e^y - 1 \text{ where } 0 \le y \le \log(2).$$

To calculate the density function of Y we simply differentiate the cumulative distribution function as follows:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} e^y - 1 = e^y.$$

Thus, the density for Y is given by

$$f_Y(y) = e^y.$$

Chapter 5.2, Question 10

Let U, V be random numbers chosen independently from the interval $[0,1]$. Find the cumulative distribution and density for the random variables.

(a) $Y = \max(U, V)$

The cumulative distribution function is

$$F_Y(y) = P(Y \leq y),$$

which is equivalent to

$$P(\max(U, V) \leq y).$$

By rearranging the interior expression, we have

$$P(U \leq y \cap V \leq y),$$

which is equivalent to

$$P(U \leq y) P(V \leq y) = y^2$$

The bounds of this cumulative distribution function are determined by considering $y^2 \in [0,1]$, so that $y \in [0,1]$.

Thus, the cumulative distribution for Y is given by

$$F_Y(y) = y^2 \text{ where } 0 \leq y \leq 1.$$

To calculate the density function of Y we simply differentiate the cumulative distribution function as follows:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} y^2 = 2y.$$

Thus, the density for Y is given by

$$f_Y(y) = 2y.$$

(b) $Y = \min(U, V)$

The cumulative distribution function is

$$1 - F_Y(y) = P(Y > y),$$

which is equivalent to

$$P(\min(U, V) > y).$$

By rearranging the interior expression, we have

$$P(U > y \cap V > y),$$

which is equivalent to

$$P(U > y) P(V > y) = (1 - y)^2$$

The bounds of this cumulative distribution function are determined by considering $1 - (1-y)^2 \in [0,1]$, so that $y \in [0,1]$.

Thus, the cumulative distribution for Y is given by

$$F_Y(y) = 1 - (1-y)^2 \text{ where } 0 \leq y \leq 1.$$

To calculate the density function of Y we simply differentiate the cumulative distribution function as follows:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} 1 - (1-y)^2 = 2(1-y).$$

Thus, the density for Y is given by

$$f_Y(y) = 2(1-y).$$

Chapter 5.2, Question 31

Let U be a uniformly distributed random variable on $[0, 1]$. What is the probability that the equation

$$x^2 + 4Ux + 1 = 0$$

has two distinct real roots x_1 and x_2?

For the equation $x^2 + 4Ux + 1 = 0$ to have real roots, $(4U)^2 - 4 \geq 0$. For the roots to be distinct, $U \neq \pm\frac{1}{2}$, as this would result in a repeated root at $x = \mp 1$, respectively.

In this case, as U is uniformly distributed on $[0, 1]$, we do not need to consider $U = -\frac{1}{2}$, though we should consider $U = \frac{1}{2}$.

Solving $(4U)^2 - 4 \geq 0$, we have $16U^2 \geq 4$, which implies $U^2 \geq \frac{1}{4}$, which further implies $U \geq \frac{1}{2}$. Taking into consideration that $U = \frac{1}{2}$ does not yield two distinct real roots x_1 and x_2, we have $U > \frac{1}{2}$.

As U is a uniformly distributed random variable on $[0, 1]$, we find that the probability that the equation $x^2 + 4Ux + 1 = 0$ has two distinct real roots x_1 and x_2 is $\frac{1}{2}$.

Chapter 5.2, Question 34

Jones puts in two new light bulbs: a 60 watt bulb and a 100 watt bulb. It is claimed that the lifetime of the 60 watt bulb has an exponential density with average lifetime 200 hours ($\lambda = 1/200$). The 100 watt bulb also has an exponential density but with average lifetime of only 100 hours ($\lambda = 1/100$). Jones wonders what is the probability that the 100 watt bulb will outlast the 60 watt bulb.

If X and Y are two independent random variables with exponential densities $f(x) = \lambda e^{-\lambda x}$ and $g(x) = \mu e^{-\mu x}$, respectively, then the probability that X is less than Y is given by

$$P(X < Y) = \int_0^\infty f(x)(1 - G(x))\, dx,$$

where $G(x)$ is the cumulative distribution function for $g(x)$. Explain why this is the case. Use this to show that

$$P(X < Y) = \frac{\lambda}{\lambda + \mu}$$

and to answer Jones's question.

To determine the probability that X is less than Y, we take the following:

$$P(X < Y) = \int_{x=0}^\infty \int_{y=x}^\infty f(x) f(y)\, dy\, dx,$$

where $f(y)$ is representative of $g(x)$ in this case, so as not to confuse the bounds of integration. In other words, we take the joint probability, only considering when the value of y is greater than the value of x, as set by the bounds of integration.

This simplifies to

$$P(X < Y) = \int_0^\infty f(x)(1 - G(x))\, dx,$$

where $G(x)$ is the cumulative distribution function for $g(x)$.

Using this, we may show that $P(X < Y) = \frac{\lambda}{\lambda+\mu}$ as follows:

$$P(X < Y) = \int_0^\infty \lambda e^{-\lambda x} \left(1 - \left(1 - e^{-\mu x}\right)\right) dx.$$

This expression is equivalent to

$$P(X < Y) = \int_0^\infty \lambda e^{-\lambda x} \left(e^{-\mu x}\right) \, dx,$$

which simplifies to

$$P(X < Y) = \lambda \int_0^\infty e^{-(\lambda+\mu)x} \, dx.$$

Computing the integral results in

$$P(X < Y) = \lambda \left[- \left(\frac{1}{\lambda + \mu} \right) e^{-(\lambda+\mu)x} \right]_0^\infty,$$

which evaluates to

$$P(X < Y) = \frac{\lambda}{\lambda + \mu},$$

as we hoped to achieve. □

Using this result, we answer Jones's question by inputting $\lambda = 1/200$ and $\mu = 1/100$ to find that the probability that the 100 watt bulb will outlast the 60 watt bulb is

$$P(X < Y) = \frac{1}{3}.$$

Chapter 5.2, Question 37

Let X be a random variable having a normal density and consider the random variable $Y = e^X$. Then Y has a log normal density. Find this density of Y.

Let the random variable X have the $N(0, 1)$ distribution for which the normal density function is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Now, when we consider the random variable $Y = e^X$, we find the probability density function of Y by finding the cumulative distribution function $F_Y(y)$ and then differentiating.

The value of Y is always positive, so $F_Y(y) = 0$ if $y \leq 0$. Suppose that $y > 0$. Then

$$F_Y(y) = P(Y \leq y) = P\left(e^X \leq y\right) = P(X \leq \ln(y)).$$

Using the integral definition for this expression, we have

$$\int_{-\infty}^{\ln(y)} f_X(x) \, dx = F_X(\ln(y)).$$

When we differentiate, we obtain the following expression for $f_Y(y)$:

$$f_Y(y) = F_x'(\ln(y)) \left(\frac{1}{y}\right) = f_x(\ln(y)) \left(\frac{1}{y}\right).$$

Using our expression for $f_X(x)$ (normal distribution function), this results in

$$f_Y(y) = \frac{1}{\sqrt{2\pi} y} e^{-\ln^2(y)/2} \text{ for } y > 0.$$

Chapter 6.1, Question 7

Show that, if X and Y are random variables taking on only two values each, and if $E(XY) = E(X)E(Y)$, then X and Y are independent.

Since X and Y are random variables taking on only two values each, we may choose a, b, c, d so that

$$U = \frac{X+a}{b} \qquad V = \frac{Y+c}{d},$$

and U, V take only values 0 and 1.

Thus, if $E(XY) = E(X)E(Y)$, then $E(UV) = E(U)E(V)$. This indicates that if U and V are independent, then so are X and Y. As such, it is sufficient to prove independence for U and V taking on values 0 and 1 with $E(UV) = E(U)E(V)$.

We have the following:

$$E(UV) = P(U=1, V=1) = E(U)E(V) = P(U=1)P(V=1)$$

and

$$\begin{aligned} P(U=1, V=0) &= P(U=1) - P(U=1, V=1) \\ &= P(U=1)(1 - P(V=1)) \\ &= P(U=1)P(V=0) \end{aligned}$$

Similarly,

$$P(U=0, V=1) = P(U=0)P(V=1)$$

$$P(U=0, V=0) = P(U=0)P(V=0)$$

Thus, U and V are independent, and hence X and Y are also.

Chapter 6.1 Question 31

A large number, N, of people are subjected to a blood test. This can be administered in two ways: (1) Each person can be tested separately, in this case N tests are required, (2) the blood samples of k persons can be pooled and analyzed together. If this test is negative, this one test suffices for the k people. If the test is positive, each of the k persons must be tested separately, and in all, $k+1$ tests are required for the k people. Assume that the probability p that a test is positive is the same for all people and that these events are independent.

(a) Find the probability that the test for a pooled sample of k people will be positive.

The probability that everyone is negative in a pool of k people is given by $(1-p)^k$. Thus, the probability that at least one person is positive in the group (which is equivalent to the probability that the test for a pooled sample of k people will be positive) is as follows:

$$1 - (1-p)^k$$

(b) What is the expected value of the number X of tests necessary under plan (2)? (Assume that N is divisible by k.)

To calculate the expected value of the number X of tests necessary under plan (2), we calculate a weighted average based on the probability of each occurrence. The probability that there will be

$(k+1)$ tests is given by part (a): $1-(1-p)^k$. The probability that there will be 1 test is given by the complement: $(1-p)^k$.

We know that for a grouping of N people with pools of k, there will be $\frac{N}{k}$ instances of the testing occurring, and we are given that N is divisible by k.

Thus, we calculate that the expected value of the number X of tests necessary under plan (2) is as follows:

$$\left[\left(1-(1-p)^k\right)(k+1)+(1-p)^k(1)\right]\left(\frac{N}{k}\right)$$

(c) For small p, show that the value of k which will minimize the expected number of tests under the second plan is approximately $\frac{1}{\sqrt{p}}$.

If p is small, then $(1-p)^k \sim 1-kp$. This is the key insight in this approximation.

Using $(1-kp)$ in place of $(1-p)^k$ in our expected value of the number of tests calculation, we have

$$\left[(1-(1-kp))(k+1)+(1-kp)(1)\right]\left(\frac{N}{k}\right),$$

which simplifies to

$$\left[kp(k+1)+(1-kp)\right]\left(\frac{N}{k}\right).$$

Further simplifying this results in

$$\left(k^2p+kp+1-kp\right)\left(\frac{N}{k}\right),$$

which is equivalent to

$$\left(kp+\frac{1}{k}\right)N.$$

To minimize this expression, we disregard N (which is fixed) and set the derivative of the expression equal to zero as follows:

$$\frac{d}{dk}\left(kp+\frac{1}{k}\right)=0$$

$$p-\frac{1}{k^2}=0$$

Solving for k results in

$$k=\frac{1}{\sqrt{p}}.$$

Therefore, the value of k which will minimize the expected number of tests under the second plan is approximately $\frac{1}{\sqrt{p}}$.