IEOR 6711: Stochastic Models I

Fall 2012, Professor Whitt

Solutions to Homework Assignment 1

The assignment consists of the following ten problems from Chapter 1: Problems 1.1-1.4, 1.8, 1.11, 1.12, 1.15, 1.22 and 1.37. You need not turn in problems with answers in the back. (Here those are 1.22 and 1.37.) Since you may not have the textbook yet, the problems are given again here and expanded upon. (However, there are plenty of copies of the textbook in the Columbia bookstore.) You are to do the extra parts added in this expansion, as well as the problems from the book. The expansion illustrates that there may be more going on than you at first think.

1. Elaboration on Problem 1.1 on p. 46.

(a) Application

This problem is about the *tail-integral formula for expected value*,

$$EX = \int_0^\infty P(X > t)\, dt = \int_0^\infty \bar{F}(t)\, dt \ ,$$

where $F \equiv F_X$ is the *cumulative distribution function* (cdf) of the random variable X, i.e., $F(t) \equiv P(X \le t)$, under the assumption that X is a nonnegative random variable. Here $\equiv$ denotes "equality by definition." The function $\bar{F} \equiv 1 - F$ is the *complementary cumulative distribution function* (ccdf) or tail-probability function.

The tail-integral formula for expected value is an alternative to the standard formula (usual definition)

$$EX = \int_0^\infty x f(x)\, dx \ .$$

The tail-integral formula for expected value can be proved in at least two ways: (i) by converting it to an iterated double integral and changing the order of integration, and (ii) by integration by parts. Before considering the proof, let us see why the formula is interesting and useful.

Apply the tail integral formula for the expected value to compute the expected value of an exponential random variable with rate λ and mean $1/\lambda$, i.e., for the random variable with tail probabilities (ccdf)

$$\bar{F}(t) \equiv P(X > t) \equiv e^{-\lambda t}, \quad t \ge 0 \ ,$$

and density (probability density function or pdf)

$$f(t) \equiv f_X(t) \equiv \lambda e^{-\lambda t}, \quad t \ge 0 \ .$$

You are supposed to see how easy it is this way:

$$EX = \int_0^\infty \bar{F}(t)\, dt = \int_0^\infty e^{-\lambda t}\, dt = \frac{1}{\lambda} \ .$$

(b) Alternative Approaches

Compute the expected value of the exponential distribution above in two other ways:

(i) Exploit the structure of the *gamma distribution*: The gamma density with *scale parameter* λ and *shape parameter* ν is

$$f_{\lambda,\nu}(t) \equiv \frac{1}{\Gamma(\nu)} \lambda^{\nu} t^{\nu-1} e^{-\lambda t}, \quad t \geq 0,$$

where Γ is the *gamma function*, i.e.,

$$\Gamma(t) \equiv \int_0^{\infty} x^{t-1} e^{-x}\, dx\ ,$$

which reduces to a factorial at integer arguments: $\Gamma(n+1) = n!$. The important point for the proof here is that the gamma density is a proper probability density, and so integrates to 1.

This way is also easy, but it exploits special knowledge about the gamma distributions. We just write down the standard definition of the expected value, and then see the connection to the gamma density for $\nu = 2$. (The exponential distribution is the gamma distribution with $\nu = 1$.)

$$EX = \int_0^{\infty} t f(t)\, dt = \int_0^{\infty} t \lambda e^{-\lambda t}\, dt = \int_0^{\infty} (1/\lambda) f_{\lambda,2}(t)\, dt = \frac{1}{\lambda}$$

because the gamma density $f_{\lambda,2}$ integrates to 1.

(ii) Use *integration by parts*, e.g., see Feller II (1971), p. 150: Suppose that u is bounded and has continuous derivative u'. Suppose that f is a pdf of a nonnegative random variable with associated ccdf $\bar{F}$. (The derivative of $\bar{F}$ is $-f$.) Then, for any b with $0 < b < \infty$,

$$\int_0^b u(t) f(t)\, dt = -u(b)\bar{F}(b) + u(0)\bar{F}(0) + \int_0^b u'(t)\bar{F}(t)\, dt\ .$$

This is the straightforward approach, but it is somewhat complicated. First,

$$EX = \int_0^{\infty} t f(t)\, dt = \lim_{b\to\infty} \int_0^b t f(t)\, dt\ .$$

Next apply integration by parts, in the form above, to get

$$\begin{aligned}\int_0^b t f(t)\, dt &= -u(b)\bar{F}(b) + u(0)\bar{F}(0) + \int_0^b u'(t)\bar{F}(t)\, dt \\ &= -be^{-\lambda b} + 0e^{-\lambda 0} + \int_0^b 1e^{-\lambda t}\, dt \\ &= -be^{-\lambda b} + \frac{1}{\lambda}(1 - e^{-\lambda b})\ .\end{aligned}$$

Now let $b \to \infty$ to get

$$EX = \lim_{b\to\infty} \int_0^b t f(t)\, dt = \lim_{b\to\infty} \{-be^{-\lambda b} + \frac{1}{\lambda}(1 - e^{-\lambda b})\} = \frac{1}{\lambda}\ .$$

The same approach produces a proof of the general formulas.

(c) Interchanging the order of integrals and sums.

Relatively simple proofs of the results to be proved in Problem 1.1 of Ross follow from interchanging the order of integrals and sums. We want to use the following relations:

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{i,j} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} a_{i,j}$$

and

$$\int_0^{\infty}\int_0^{\infty} f(x,y)\,dx\,dy = \int_0^{\infty}\int_0^{\infty} f(x,y)\,dy\,dx$$

It is important to know that these relations are *usually valid*. It is also important to know that these relations are *not always valid*: In general, there are *regularity conditions*. The complication has to do with infinity and limits. There is no problem at all for finite sums:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j} = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{i,j} \ .$$

That is always true. Infinite sums involve limits, and integrals are defined as limits of sums.

The interchange property is often used with expectations. In particular, it is used to show that the expectation can be taken inside sums and integrals:

$$E[\sum_{i=1}^{\infty} X_i] = \sum_{i=1}^{\infty} EX_i$$

and

$$E[\int_0^{\infty} X(s)\,ds] = \int_0^{\infty} E[X(s)]\,ds \ .$$

These relations are of the same form because the expectation itself can be expressed as a sum or an integral. However, there are regularity conditions, as noted above.

(i) Compute the two iterated sums $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{i,j}$ and $\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} a_{i,j}$ for

$$a_{i,j} = 2 - 2^{-i} \quad \text{for} \quad i = j, \quad a_{i,j} = -2 + 2^{-i} \quad \text{for} \quad i = j+1, \quad j \geq 1 \ ,$$

and $a_{i,j} = 0$ otherwise. What does this example show?

In this example the two iterated sums are not equal. This example shows that regularity conditions are needed. It would suffice to have either: (i) $a_{i,j} \geq 0$ for all i and j (Tonelli) or (ii) One of the iterated sums, or the double sum, be finite when $a_{i,j}$ is replaced by its absolute value $|a_{i,j}|$ for all i and j (Fubini).

In particular, here

$$\sum_{i=1}^{\infty} a_{i,j} = -2^{-(j+1)} \ ,$$

so that

$$\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{i,j} = -1/2 \ .$$

On the other hand,

$$\sum_{j=1}^{\infty} a_{i,j} = 0 \quad \text{for} \quad i \ge 2 \quad \text{and} \quad \sum_{j=1}^{\infty} a_{1,j} = 2 - 2^{-1} = 3/2 \ .$$

Hence,

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{i,j} = 3/2 \quad \text{while} \quad \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{i,j} = -1/2 \ .$$

(ii) Look up Tonelli's theorem and Fubini's theorem in a book on measure theory and integration, such as *Real Analysis* by H. L. Royden. What do they say about this problem?

Tonelli's theorem says that the interchange is valid for nonnegative summands and (measurable) integrands. There are also conclusions about measurability, but those are relatively minor technical issues. Fubini's theorem says that the interchange is valid if the summands and integrands are summable or integrable, which essentially (aside from measurability issues) means that the absolute values of the summands and integrands are summable and integrable, respectively. Neither condition is satisfied in the example above, so that the overall sum depends on the order in which we perform the summation. We could even get $\infty - \infty$. (There are infinitely many $+2's$ and infinitely many $-2's$.)

(d) Do the three parts to Problem 1.1 in Ross (specified here below).

Hint: Do the proofs by interchanging the order of sums and integrals. The first step is a bit tricky. We need to get an iterated sum or integral; i.e., we need to insert the second sum or integral. To do so for sums, write:

$$EN = \sum_{i=1}^{\infty} iP(N = i) = \sum_{i=1}^{\infty} [\sum_{j=1}^{i} 1] P(N = i) \ .$$

(i) For a nonnegative integer-valued random variable N, show that

$$EN = \sum_{i=1}^{\infty} P(N \ge i) \ .$$

(ii) For a nonnegative random variable X with cdf F, show that

$$EX = \int_0^{\infty} P(X > t) \, dt = \int_0^{\infty} \bar{F}(t) \, dt \ .$$

(iii) For a nonnegative random variable X with cdf F, show that

$$E[X^n] = \int_0^{\infty} n t^{n-1} \bar{F}(t) \, dt \ .$$

(i) Once we have the double sum, we can change the order of summation. The interchange is justified because the summand is nonnegative; we can apply Tonelli's theorem. A key practical step is to get the range of summation right:

$$\begin{aligned}
EN &= \sum_{i=1}^{\infty} iP(N=i) = \sum_{i=1}^{\infty} [\sum_{j=1}^{i} 1] P(N=i) \\
&= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} 1 P(N=i) \\
&= \sum_{j=1}^{\infty} P(N \geq j) \ .
\end{aligned}$$

(ii) Now consider the first integral. Then

$$\begin{aligned}
EX &= \int_0^{\infty} x f(x)\, dx = \int_0^{\infty} [\int_0^x 1\, dy] f(x)\, dx \\
&= \int_0^{\infty} 1 [\int_y^{\infty} f(x)\, dx]\, dy \\
&= \int_0^{\infty} \bar{F}(y)\, dy \ .
\end{aligned}$$

(iii) Finally, turning to the general case,

$$\begin{aligned}
E[X^n] &= \int_0^{\infty} x^n f(x)\, dx = E[\int_0^{\infty} [\int_0^x n y^{n-1}\, dy] f(x)\, dx \\
&= \int_0^{\infty} n y^{n-1} [\int_y^{\infty} f(x)\, dx]\, dy \\
&= \int_0^{\infty} n y^{n-1} [\bar{F}(y)]\, dy \ .
\end{aligned}$$

(e) What regularity property justifies the interchanges used in part (d)?

The summands and integrands are all nonnegative, so we can apply Tonelli's theorem.

(f) What happens to the tail-integral formula for expected value

$$EX = \int_0^{\infty} P(X > t)\, dt = \int_0^{\infty} \bar{F}(t)\, dt \ ,$$

when the random variable X is integer valued?

It reduces to the first formula given in Ross:

$$EN = \sum_{j=1}^{\infty} P(N \geq j) \ .$$

---

2. Elaboration of Problems 1.2 on p. 46 and 1.15 on p. 49.

(a) Application to simulation

(i) Suppose that you know how to generate on the computer a random variable U that is uniformly distributed on the interval $[0,1]$. (That tends to be easy to do, allowing for the necessary approximation due to discreteness.) How can you use U to generate a random variable X with an cdf F that is an arbitrary continuous and strictly increasing function? (Hint: Use the inverse F^{-1} of the function F.)

Since the function F is continuous and strictly increasing, it has an inverse, say F^{-1}, with the properties

$$F^{-1}(F(x)) = x \quad \text{and} \quad F(F^{-1}(t)) = t$$

for all x and t with $0 < t < 1$. Hence, we can use the random variable $F^{-1}(U)$ because, for all x,

$$P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x) \ .$$

(ii) Suppose that you know how to generate on the computer a random variable U that is uniformly distributed on the interval $[0,1]$. How can you use U to generate a random variable X with an exponential distribution with mean $1/\lambda$?

We are given the cdf $F(t) = 1 - e^{-\lambda t}$, $t \ge 0$. We want to find the inverse. For t given, $0 < t < 1$, it suffices to find the value of x such that $1 - e^{-\lambda x} = t$. That occurs when $1 - t = e^{-\lambda x}$. Taking logarithms, we find that

$$F^{-1}(t) = -\frac{1}{\lambda}\log(1-t) \ ,$$

where log is the natural (base e) logarithm. Hence, $-\frac{1}{\lambda}\log(1-U)$ has the given exponential distribution. Since $1-U$ has the same distribution as U, the random variable $-\frac{1}{\lambda}\log(U)$ has the same exponential distribution. (incidentally, this is Problem 1.15 (b).)

(b) Problem 1.2 for a cdf F having a positive density f.

Do the two parts of Problem 1.2 in Ross (specified here below) under the assumption that the random variable X concentrates on the interval (a,b) for $-\infty \le a < b \le +\infty$ and has a strictly positive density f there. That is, $P(a \le X \le b) = 1$,

$$F(x) \equiv P(X \le x) = \int_a^x f(y)\, dy \ ,$$

where $f(x) > 0$ for all x such that $a \le x \le b$.

The two parts of Problem 1.2 are:

(i) If X is a random variable with a continuous cdf F, show that $F(X)$ is uniformly distributed on the interval $[0,1]$.

The positive density condition implies that the cdf is continuous and strictly increasing. Hence we are in the setting of the previous part: the cdf F has an inverse F^{-1}. We can exploit the inverse property to write

$$P(F(X) \leq t) = P(X \leq F^{-1}(t)) = F(F^{-1}(t)) = t ,$$

which implies that $F(X)$ is distributed the same as U, i.e., uniformly on $(0, 1)$.

(ii) If U is uniformly distributed on $(0, 1)$, then $F^{-1}(U)$ has cdf F.

Given the inverse property, this part is repeating the Problem 2. (a) (i) above.

(c) Right continuity.

A cdf F is a right-continuous nondecreasing function of a real variable such that $F(x) \to 1$ as $x \to \infty$ and $F(x) \to 0$ as $x \to -\infty$.

A real-valued function g of a real variable x is *right-continuous* if

$$lim_{y \downarrow x} g(y) = g(x)$$

for all real x, where $\downarrow$ means the limit from above (or from the right). A function g is *left-continuous* if

$$lim_{y \uparrow x} g(y) = g(x)$$

for all real x, where $\uparrow$ means the limit from below (or from the left).

A function g has a limit from the right at x if the limit $lim_{y \downarrow x} g(y)$ exists. Suppose that a function g has limits everywhere from the left and right. Then the *right-continuous version* of g, say g_+ is defined by

$$g_+(x) \equiv \lim_{y \downarrow x} g(y)$$

for all x. The *left-continuous version* of g, g_- is defined similarly.

Suppose that $P(X = 1) = 1/3 = 1 - P(X = 3)$. Let F be the cdf of X. What are $F(1)$, $F(2)$, $F(3)$ and $F(4)$? What are the values of the left-continuous version F_- at the arguments 1, 2, 3 and 4?

Right-continuity means that $F(x) = P(X \leq x)$, while left-continuity means that $F_-(x) = P(X < x)$. Hence,

$$F(1) = 1/3 = F(2), \quad \text{and} \quad F(3) = 1 = F(4) ,$$

while

$$F_-(1) = 0, \quad F_-(2) = 1/3 = F_-(3) \quad \text{and} \quad F_-(4) = 1 .$$

(d) The left-continuous inverse of a cdf F.

Given a (right-continuous) cdf F, let

$$F^{\leftarrow}(t) \equiv \inf\{x : F(x) \geq t\}, \quad 0 < t < 1 .$$

Fact: $F^{\leftarrow}$ is a left-continuous function on the interval $(0,1)$.

Fact: In general, $F^{\leftarrow}$ need not be right-continuous.

Fact:

$$F^{\leftarrow}(t) \le x \quad \text{if and only if} \quad F(x) \ge t$$

for all t and x with $0 < t < 1$.

Suppose that X is a random variable with a continuous cdf F. (We now do not assume that F has a positive density.) Show that $F(X)$ is uniformly distributed on the interval $(0,1)$. (Hint: use the fact that $F(F^{\leftarrow}(t)) = t$ for all t with $0 < t < 1$ when F is continuous.)

For all t, $0 < t < 1$,

$$P(F(X) \le t) = P(X \le F^{\leftarrow}(t)) = F(F^{\leftarrow}(t)) = t\ ,$$

using the continuity of F in the last step.

(e) The right-continuous inverse of a cdf F.

Given a (right-continuous) cdf F, let

$$F^{-1}(t) \equiv \inf\{x : F(x) > t\}, \quad 0 < t < 1\ .$$

Fact: It can be shown that F^{-1} is right-continuous, but in general is not left-continuous.

Show that the relation

$$F^{-1}(t) \le x \quad \text{if and only if} \quad F(x) \ge t$$

for all t and x with $0 < t < 1$ does *not hold* in general.

We give a counterexample: Let $P(X = 1) = P(X = 2) = 1/2$. Then $F(1) = 1/2$ and $F(2) = 1$. Note that $F(1) \ge 1/2$, but $F^{-1}(1/2) = 2$, which is not less than or equal to 1.

(f) Inverses. Let X be the discrete random variable defined above with $P(X = 1) = 1/3 = 1 - P(X = 3)$. Draw pictures (graphs) of the cdf F of X and the two inverses $F^{\leftarrow}$ and F^{-1}.

If there is a jump in F at x, $F(x)$ assumes the higher value, after the jump; that makes F right-continuous. The graph of $F^{\leftarrow}$ is just the graph of F with the x and y axes switched. That makes $F^{\leftarrow}$ left-continuous. Then F^{-1} is the right-continuous version of $F^{\leftarrow}$.

(g) Suppose that you know how to generate on the computer a random variable U that is uniformly distributed on the interval $[0,1]$. How can you use U to generate a random variable X with an *arbitrary* cdf F?

Let $X = F^{\leftarrow}(U)$. Then, for any x,

$$P(X \le x) = P(F^{\leftarrow}(U) \le x) = P(U \le F(x)) = F(x) \ ,$$

as desired.

(h) It can be shown that $P(F^{-1}(U) = F^{\leftarrow}(U)) = 1$. Given that result, what is the distribution of $F^{-1}(U)$?

The random variable $F^{-1}(U)$ also has cdf F.

---

3. Problem 1.3 in Ross:

Let the random variable X_n have a binomial distribution with parameters n and p_n, i.e.,

$$P(X_n = k) = \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} \ .$$

Show that

$$P(X_n = k) \to \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{as} \quad n \to \infty$$

if $np_n \to \lambda$ as $n \to \infty$. Hint: Use the fact that $(1 + (c_n/n))^n \to e^c$ as $n \to \infty$ if $c_n \to c$ as $n \to \infty$.

This limit is sometimes referred to as the *law of small numbers.*

$$\begin{aligned} P(X_n = k) &= \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} \\ &= \frac{n!}{k!(n-k)!n^k} (np_n)^k \frac{(1-(np_n/n)^n}{(1-p_n)^k} \\ &\to \frac{\lambda^k e^{-\lambda}}{k!} \end{aligned}$$

because $(np_n)^k \to \lambda^k$, $n!/(n-k)!n^k \to 1$, $(1-p_n)^k \to 1$ and $(1-(np_n/n)^n \to e^{-\lambda}$.

The limit above is an example of convergence in distribution. We say that random variables X_n with cdf's F_n *converge in distribution* to a random variable X with cdf F, and write $X_n \Rightarrow X$ or $F_n \Rightarrow F$, if

$$F_n(x) \to F(x) \quad \text{as} \quad n \to \infty \quad \text{for all} \quad x \quad \text{that are continuity points of } F \ .$$

A point x is a continuity point of F if F is continuous at x.

(b) Show that the limit above is indeed an example of convergence in distribution. What is the limiting distribution?

The limiting distribution is the Poisson distribution. We have shown that $P(X_n = k) \to P(X = k)$ for all k. We need to show that $P(X_n \le x) \to P(X \le x)$ for all x that are not continuity points of the limiting cdf $P(X \le x)$. That means for all x that are not nonnegative integers, but because all random variables are integer valued, it suffices to show that

$$P(X_n \le k) \to P(X \le k) \quad \text{for all} \quad k \ .$$

(All the cdf's remain unchanged between integer arguments; e.g., $F(x) = F(\lfloor x \rfloor)$, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x.) But that limit for the cdf's follows by induction from the property that the limit of a sum is the sum of the limits: If $a_n \to a$ and $b_n \to b$ as $n \to \infty$, then $a_n + b_n \to a + b$ as $n \to \infty$. (The property remains true for random variables defined on all integers, negative and positive, but that involves an extra argument.)

(c) Suppose that $P(X_n = 1 + n^{-1}) = P(X = 1) = 1$ for all n. Show that $X_n \Rightarrow X$ as $n \to \infty$ and that the continuity-point condition is needed here.

The main point here is that the continuity-point condition is needed, because $F_n(1) \equiv P(X_n \le 1) = 0$ for all n, while $F(1) \equiv P(X \le 1) = 1$. Since 1 is not a continuity point of F, the missing convergence does not cause a problem. For any $x > 1$, $F(x) = 1$. Moreover, there is an n_0 such that $1 + n_0^{-1} < x$, so that $F_n(x) = 1$ for all $n > n_0$. On the other hand, for any $x < 1$, $F_n(x) = F(x) = 0$. Hence, $F_n \Rightarrow F$.

(d) Use the techniques of Problem 2 above to prove the following *representation theorem*:

Theorem 0.1 *If $X_n \Rightarrow X$, then there exist random variables $\tilde{X}_n$ and $\tilde{X}$ defined on a common probability space such that $\tilde{X}_n$ has the same distribution F_n as X_n for all n, $\tilde{X}$ has the same distribution F as X, and*

$$P(\tilde{X}_n \to \tilde{X} \quad as \quad n \to \infty) = 1 \ .$$

Start with a random variable U uniformly distributed on $[0, 1]$. Let $\tilde{X}_n = F_n^{\leftarrow}(U)$ for all n and let $\tilde{X} = F^{\leftarrow}(U)$. By Problem 2, $\tilde{X}_n$ has the same distribution as X_n for all n and $\tilde{X}$ has the same distribution as X. It remains to show the w.p.1 convergence. Note that the same random variable U is used throughout. The assumed convergence means that $F_n(x) \to F(x)$ for all x that are continuity points of F. We need three technical lemmas at this point:

Lemma 0.1 *The convergence*

$$F_n^{\leftarrow}(t) \to F^{\leftarrow}(t) \quad as \quad n \to \infty$$

holds for for all t in $(0, 1)$ that are continuity points of $F^{\leftarrow}$ if and only if

$$F_n(x) \to F(x) \quad as \quad n \to \infty$$

holds for for all x in $\mathbb{R}$ that are continuity points of F.

This first lemma is tedious to prove, but it can be done. Supposing that $F_n \Rightarrow F$, you pick t such that t is a continuity point of $F^{\leftarrow}$. Then we show that indeed $F_n^{\leftarrow}(t) \to F^{\leftarrow}(t)$ as $n \to \infty$. That is a tedious ϵ and δ argument (or, at least, that is how I would proceed)..

Lemma 0.2 *If g is a nondecreasing real-valued function of a real variable, the number of discontinuity points of g is a countably infinite set.*

Lemma 0.3 *If A is a countably infinite subset of the real line $\mathbb{R}$ and X is a random variable with a probability density function, i.e., if*

$$F(x) = \int_{-\infty}^{x} f(y)\,dy, \quad -\infty < x < \infty \ ,$$

then

$$P(X \in A) = \int_A f(x)\,dx = 0 \ .$$

As a consequence of the three technical lemmas, the assumed convergence $F_n \to F$ implies that $F_n^{\leftarrow}(t) \to F^{\leftarrow}(t)$ as $n \to \infty$ for all t in $(0,1)$ except for a set of t of measure 0 under U. (We omit a full demonstration of this technical point.) Thus

$$P(F_n^{\leftarrow}(U) \to F^{\leftarrow}(U) \quad \text{as} \quad n \to \infty) = 1 \ .$$

---

4. Problem 1.4 in Ross:

Derive the mean and variance of a binomial random variable with parameters n and p. Hint: Use the relation between Bernoulli and binomial random variables.

Following the hint, use the fact that the binomial random variable is distributed as $S_n \equiv X_1 + \cdots + X_n$, where X_i are IID Bernoulli random variables, with $P(X_1 = 1) = p = 1 - P(X_1 = 0)$. Clearly,

$$EX_1 = p \quad \text{and} \quad Var(X_1) = p(1-p) \ .$$

Since the mean of a sum is the sum of the means (always) and the variance of the sum is the sum of the variances (because of the independence), we have

$$ES_n = np \quad \text{and} \quad Var(S_n) = np(1-p) \ .$$

See formula (1.3.4) on p. 10 of Ross for the formula of the variance of a sum of random variables in the general case in which independence need not hold.

---

5. Problem 1.8 in Ross:

Let X_1 and X_2 be independent Poisson random variables with means λ_1 and λ_2, respectively.

(a) Find the distribution of the sum $X_1 + X_2$.

The sum of independent Poisson random variables is again Poisson. That is easy to show with moment generating functions:

$$\phi_{X_i}(t) \equiv E[e^{tX_i}] = e^{\lambda_i(e^t - 1)} \ .$$

Because of the independence,

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)} \ .$$

The result is also not too hard to derive directly. We again exploit the independence and use the convolution formula for sums:

$$\begin{aligned}
P(X_1+X_2=k) &= \sum_{j=0}^{k} P(X_1=j)P(X_2=k-j) \\
&= \sum_{j=0}^{k} \frac{e^{-\lambda_1}\lambda_1^j}{j!}\frac{e^{-\lambda_2}\lambda_2^{k-j}}{(k-j)!} \\
&= \sum_{j=0}^{k} \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^j\lambda_2^{k-j}}{j!(k-j)!} \\
&= e^{-(\lambda_1+\lambda_2)}\sum_{j=0}^{k} \frac{\lambda_1^j\lambda_2^{k-j}}{j!(k-j)!} \\
&= e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^k\sum_{j=0}^{k} \frac{\frac{\lambda_1^j}{(\lambda_1+\lambda_2)^j}\frac{\lambda_2^{k-j}}{(\lambda_1+\lambda_2)^{(k-j)}}}{j!(k-j)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^k}{k!}k!\sum_{j=0}^{k} \frac{\frac{\lambda_1^j}{(\lambda_1+\lambda_2)^j}\frac{\lambda_2^{k-j}}{(\lambda_1+\lambda_2)^{(k-j)}}}{j!(k-j)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^k}{k!}\sum_{j=0}^{k} k!\frac{\frac{\lambda_1^j}{(\lambda_1+\lambda_2)^j}\frac{\lambda_2^{k-j}}{(\lambda_1+\lambda_2)^{(k-j)}}}{j!(k-j)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^k}{k!}
\end{aligned}$$

using the law of total probabilities for the binomial distribution in the last step.

(b) Compute the conditional distribution of X_1 given that $X_1+X_2=n$.

Use the definition of conditional probability and apply the necessary algebra:

$$\begin{aligned}
P(X_1=k|X_1+X_2=n) &= \frac{P(X_1=k \quad \text{and} \quad X_1+X_2=n)}{P(X_1+X_2=n)} \\
&= \frac{P(X_1=k \quad \text{and} \quad X_2=n-k)}{P(X_1+X_2=n)} \\
&= \frac{P(X_1=k)P(X_2=n-k)}{P(X_1+X_2=n)} \\
&= \frac{\frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}}{\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}} \\
&= \frac{n!}{k!(n-k)!}(\lambda_1/(\lambda_1+\lambda_2))^k(\lambda_2/(\lambda_1+\lambda_2))^{n-k}
\end{aligned}$$

So that $(X_1|X_1+X_2=n)$ has a binomial distribution with parameters n and $p=(\lambda_1/(\lambda_1+\lambda_2)$.

---

6. Problem 1.11 in Ross: generating functions.

Let X be a nonnegative integer-valued random variable. Then the generating function of X is

$$\hat{P}(z) \equiv E[z^X] = \sum_{j=0}^{\infty} z^j P(X=j) \ .$$

(a) Show that the k^{th} derivative of $\hat{P}(z)$ evaluated at $z=0$ is $k!P(X=k)$.

Use mathematical induction to justify the formula for all k. Differentiate term by term to get

$$\begin{aligned}
\frac{d^k \hat{P}(z)}{dz^k} &= \sum_{j=k}^{\infty} j(j-1)(j-2)\cdots(j-k+1)z^{j-k}P(X=j) \\
&= k!P(X=k) + \sum_{j=k+1}^{\infty} j(j-1)(j-2)\cdots(j-k+1)z^{j-k}P(X=j)
\end{aligned}$$

Evaluating at $z=0$ gives the desired result.

(b) Show that (with 0 being considered even)

$$P(X \text{ is even}) = \frac{\hat{P}(1)+\hat{P}(-1)}{2} \ .$$

$$\hat{P}(1)+\hat{P}(-1) = \sum_{j} P(X=j) + \sum_{j:even} P(X=j) - \sum_{j:odd} P(X=j) = 2\sum_{j:even} P(X=j) \ .$$

(c) Calculate $P(X$ is even $)$ when

(i) X is binomial (n,p);

When X is binomial(n,p),

$$\hat{P}(z) = (zp+1-p)^n$$

so that

$$P(X \text{ even}) = \frac{1^n + (1-2p)^n}{2} = \frac{1+(1-2p)^n}{2}$$

(ii) X is Poisson with mean λ;

When X is Poisson(λ),

$$\hat{P}(z) = e^{\lambda(z-1)}$$

so that

$$P(X \text{ even}) = \frac{e^0 + e^{-2\lambda}}{2} = \frac{1 + e^{-2\lambda}}{2}$$

(iii) X is geometric with parameter p, i.e., $P(X = k) = p(1-p)^{k-1}$ for $k \geq 1$.

When X is geometric(p),

$$\hat{P}(z) = \frac{zp}{(1 - z + zp)}$$

so that

$$P(X \text{ even}) = \frac{1 + \frac{-p}{2-p}}{2} = \frac{1-p}{2-p}$$

---

7. Problem 1.12 in Ross:

If $P(0 \leq X \leq a) = 1$, show that

$$Var(X) \leq a^2/4.$$

To get your intuition going, consider the obvious candidate distribution on the interval $[0, a]$ with large variance: Put mass 1/2 on 0 and a. The variance of that two-point distribution is $a^2/4$, which is our desired bound. So it looks right, to get started.

It is perhaps easier to bound $E[X^2]$ above, given EX: Note that $X^2 \leq aX$ under the condition. Then take expected values, getting

$$E[X^2] \leq aEX \ .$$

Then

$$Var(X) = E[X^2] - (EX)^2 \leq aEX - (EX)^2 = EX(a - EX) \ ,$$

but EX can be anything in the interval $[0, a]$, so let $EX = pa$ for some p with $0 \leq p \leq 1$. Then

$$Var(X) \leq EX(a - EX) = pa(a - pa) = a^2(p(1-p)) \leq \frac{a^2}{4} \ ,$$

because, by calculus, $p(1-p)$ is maximized at $p = 1/2$.

---

8. Problem 1.22 in Ross:

The *conditional variance* of X given Y is defined as

$$Var(X|Y) \equiv E[(X - E[X|Y])^2|Y] \ .$$

Prove the *conditional variance formula*:

$$Var(X) = E[Var(X|Y)] + Var(E[X|Y]) \ .$$

A key relation is $EX = E[E[X|Y]]$ for any random variables X and Y. It is easier to start with the two pieces and put them together. First,

$$\begin{aligned} E[Var(X|Y)] &= E[E[(X - E[X|Y])^2|Y]] \\ &= E[E[X^2 - 2XE[X|Y] + E[X|Y]^2|Y]] \\ &= E[E[X^2|Y] - 2E[X|Y]^2 + E[X|Y]^2] \\ &= E[E[X^2|Y] - E[X|Y]^2] \\ &= E[E[X^2|Y]] - E[E[X|Y]^2]] \\ &= E[X^2] - E[E[X|Y]^2]] \ . \end{aligned}$$

Next,

$$\begin{aligned} Var(E[X|Y]) &= E[E[X|Y]^2] - (E[E[X|Y]])^2 \\ &= E[E[X|Y]^2] - (E[X])^2 \ . \end{aligned}$$

Therefore,

$$E[Var(X|Y)]+Var(E[X|Y]) = (E[X^2]-E[E[X|Y]^2])+(E[E[X|Y]^2]-(E[X])^2) = E[X^2]-(E[X])^2 = Var(X) \ .$$

---

9. Problem 1.37 in Ross:

Let $\{X_n : n \geq 1\}$ be a sequence of independent and identically distributed (IID) random variables with a probability density function f. Say that a peak occurs at index n if $X_{n-1} < X_n > X_{n+1}$. (We stipulate that, then, a peak cannot occur for $n = 1$.) Show that the long-run proportion of indices at which a peak occurs is, with probability 1, equal to $1/3$. (Hint: Use the strong law of large numbers for partial sums of IID random variables.)

Since the random variables X_n are IID with a probability density function, the probability any two random variables are identically equal is 0. So we do not have to worry about ties. For $n \geq 2$, each of the three variables X_{n-1}, X_n and X_{n+1} is equally likely to be the largest of these three. So the probability that a peak occurs at time (index) n is $1/3$. But X_n and X_{n+1} are dependent. If n is a peak, then $n+1$ cannot be a peak.

Let $Y_n = 1$ if a peak occurs at time n and 0 otherwise. We know that $EY_n = P(Y_n = 1) = 1/3$ for all $n \geq 2$, but the successive Y_n variables are dependent. However, $\{Y_2, Y_5, Y_8, \ldots\}$, $\{Y_3, Y_6, Y_9, \ldots\}$ and $\{Y_4, Y_7, Y_{10}, \ldots\}$ are three identically distributed sequences of IID random variables. For each one, we can apply the strong law of large numbers for IID random variables and deduce that the long-run proportion within each sequence is $1/3$ with probability 1. For example, we have

$$n^{-1} \sum_{k=1}^{n} Y_{2+3k} \to \frac{1}{3} \quad \text{as} \quad n \to \infty \text{ w.p.1} \ .$$

Those three separate limits then imply that

$$n^{-1} \sum_{k=1}^{n} Y_k \to \frac{1}{3} \quad \text{as} \quad n \to \infty \text{ w.p.1 } .$$