1 Uniform Uniform Computation

Note 21

Suppose $X \sim \text{Uniform}[0,1]$ and $Y \sim \text{Uniform}[0,X]$. That is, conditioned on $X = x$, Y has a Uniform$[0,x]$ distribution.

(a) What is $\mathbb{P}[Y > 1/2]$?

(b) Calculate $\text{Cov}(X,Y)$.

Solution:

(a) First we compute $\mathbb{P}[Y > 1/2 \mid X = x]$. Conditioned on $X = x, Y$ has the Uniform$[0,x]$ distribution, and the tail probability of the uniform distribution is as follows:

$$\mathbb{P}\left[Y > \frac{1}{2} \,\middle|\, X = x\right] = \begin{cases} 0, & x < 1/2 \\ (x-1/2)/x, & x \geq 1/2 \end{cases}$$

So, integrate over values of $x \geq 1/2$ to find $\mathbb{P}[Y > 1/2]$ (note that the upper limit of integration is $x = 1$ since $X \sim \text{Uniform}[0,1]$).

$$\begin{aligned}\mathbb{P}[Y > 1/2] &= \int_{-\infty}^{\infty} \mathbb{P}\left[Y > \frac{1}{2} \,\middle|\, X = x\right] f_X(x)\,\mathrm{d}x = \int_{1/2}^{1} \left(1 - \frac{1}{2x}\right) \mathrm{d}x \\ &= \left[x - \frac{1}{2}\ln x\right]_{x=1/2}^{x=1} = \frac{1}{2}(1 - \ln 2).\end{aligned}$$

(b) To compute $\mathbb{E}[XY]$, note we integrate over the region defined by the triangle with endpoints $(0,0),(1,0),(1,1)$.

$$\begin{aligned}\mathbb{E}[XY] &= \int_0^1 \int_0^x xy f_{Y|X=x}(y) f_X(x)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \int_0^x xy \frac{1}{x}\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^1 \int_0^x y\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \left[\frac{y^2}{2}\right]_0^x \mathrm{d}x = \int_0^1 \frac{x^2}{2}\,\mathrm{d}x \\ &= \left[\frac{x^3}{6}\right]_0^1 = \frac{1}{6}.\end{aligned}$$

$\mathbb{E}[X] = \frac{1}{2}$ because it is a Uniform$[0,1]$ random variable. To compute $\mathbb{E}[Y]$, we can integrate y

over the same triangle as before, scaled by the joint pdf.

$$\begin{aligned}\mathbb{E}[Y] &= \int_0^1 \int_0^x y f_{Y|X=x}(y) f_X(x)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \int_0^x y\frac{1}{x}\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^1 \int_0^x \frac{y}{x}\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \left[\frac{y^2}{2x}\right]_0^x \mathrm{d}x = \int_0^1 \frac{x}{2}\,\mathrm{d}x \\ &= \left[\frac{x^2}{4}\right]_0^1 = \frac{1}{4}.\end{aligned}$$

An intuitive but less rigorous way of arriving at $\mathbb{E}[Y]$ is that Y will on average be half of X, and on average X is $\frac{1}{2}$, so Y is on average $\frac{1}{4}$, using the fact we are working with uniform random variables. Therefore the covariance is

$$\mathrm{Cov}(X,Y) = \frac{1}{6} - \left(\frac{1}{2}\right)\left(\frac{1}{4}\right) = \frac{1}{6} - \frac{1}{8} = \frac{1}{24}.$$

2 Moments of the Gaussian

Note 21

For a random variable X, the quantity $\mathbb{E}[X^k]$ for $k \in \mathbb{N}$ is called the *kth moment* of the distribution. In this problem, we will calculate the moments of a standard normal distribution.

(a) Prove the identity

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{tx^2}{2}\right)\mathrm{d}x = t^{-1/2}$$

for $t > 0$.

Hint: Consider a normal distribution with variance $\frac{1}{t}$ and mean 0.

(b) For the rest of the problem, X is a standard normal distribution (with mean 0 and variance 1). Use part (a) to compute $\mathbb{E}[X^{2k}]$ for $k \in \mathbb{N}$.

Hint: Try differentiating both sides with respect to t, k times. You may use the fact that we can differentiate under the integral without proof.

(c) Compute $\mathbb{E}[X^{2k+1}]$ for $k \in \mathbb{N}$.

Solution:

(a) Note that a normal distribution with mean 0 and variance t^{-1} has the density function

$$f(x) = \frac{\sqrt{t}}{\sqrt{2\pi}} \exp\left(-\frac{tx^2}{2}\right),$$

and since the density must integrate to 1, we see that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{tx^2}{2}\right)\mathrm{d}x = t^{-1/2}.$$

(b) Differentiating the identity from (a) k times with respect to t, we obtain a LHS of

$$\begin{aligned}
\frac{\mathrm{d}^k}{\mathrm{d}t^k}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{tx^2}{2}\right)\mathrm{d}x\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\mathrm{d}^k}{\mathrm{d}t^k}\left[\exp\left(-\frac{tx^2}{2}\right)\right]\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(-1)^k\frac{x^{2k}}{2^k}\exp\left(-\frac{tx^2}{2}\right)\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\frac{(-1)^k}{2^k}\int_{-\infty}^{\infty}x^{2k}\exp\left(-\frac{tx^2}{2}\right)\mathrm{d}x
\end{aligned}$$

Here, we use the fact that everything involving x is a constant with respect to t.

Looking at the RHS, we have

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\left[t^{-1/2}\right] = (-1)^k\frac{1\cdot 3\cdots(2k-3)\cdot(2k-1)}{2^k}t^{-(2k+1)/2}$$

Together, this means that

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}}\frac{\cancel{(-1)^k}}{\cancel{2^k}}\int_{-\infty}^{\infty}x^{2k}\exp\left(-\frac{tx^2}{2}\right)\mathrm{d}x &= \cancel{(-1)^k}\frac{1\cdot 3\cdots(2k-3)\cdot(2k-1)}{\cancel{2^k}}t^{-(2k+1)/2} \\
\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^{2k}\exp\left(-\frac{tx^2}{2}\right)\mathrm{d}x &= (1\cdot 3\cdots(2k-3)\cdot(2k-1))\,t^{-(2k+1)/2}
\end{aligned}$$

If we set $t = 1$, we get

$$\mathbb{E}[X^{2k}] = \int_{-\infty}^{\infty}x^{2k}\cdot\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x = \prod_{i=1}^{k}(2i-1).$$

This is sometimes denoted $(2k-1)!!$. Note that we can also write the result as

$$\mathbb{E}[X^{2k}] = (2k-1)!! = \frac{(2k)!}{2\cdot 4\cdots(2k-2)\cdot(2k)} = \frac{(2k)!}{2^k k!}.$$

(c) $\mathbb{E}[X^{2k+1}] = 0$, since the density function is symmetric around 0.

3 Exponential Median

Note 21

(a) Prove that if $X_1, X_2, \ldots, X_n$ are mutually independent exponential random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\min(X_1, X_2, \ldots, X_n)$ is exponentially distributed with parameter $\sum_{i=1}^{n}\lambda_i$.

Hint: Recall that the CDF of an exponential random variable with parameter λ is $1 - e^{-\lambda t}$.

(b) Given that the minimum of three i.i.d exponential variables with parameter λ is m, what is the probability that the difference between the median and the smallest is at least s? Note that the exponential random variables are mutually independent.

(c) What is the expected value of the median of three i.i.d. exponential variables with parameter λ?

Hint: Part (b) may be useful for this calculation.

Solution:

(a) In order to prove that $X := \min(X_1, \ldots, X_n)$ is exponentially distributed with parameter $\lambda := \sum_{i=1}^{n} \lambda_i$, we just need to show that the CDF matches. Hence, we would like to find $\mathbb{P}(X \le t)$, but it turns out to be easier to find $\mathbb{P}(X > t)$ first. This is because in order for X to be larger than t, you need every X_i to be larger than t, so

$$\begin{aligned}\mathbb{P}(X > t) &= \mathbb{P}(X_1 > t \cap \ldots \cap X_n > t) \\ &= \mathbb{P}(X_1 > t) \cdot \ldots \cdot \mathbb{P}(X_n > t)\end{aligned}$$

where the second equality comes from the X_is being mutually independent. Now we know that

$$\begin{aligned}\mathbb{P}(X_i > t) &= 1 - \mathbb{P}(X_i \le t) \\ &= 1 - \left(1 - e^{-\lambda_i t}\right) \\ &= e^{-\lambda_i t}\end{aligned}$$

where the second equality comes from plugging in the CDF of an exponential random variable. Thus, we get that

$$\begin{aligned}\mathbb{P}(X > t) &= e^{-\lambda_1 t} \cdot \ldots \cdot e^{-\lambda_n t} \\ &= e^{(-\lambda_1 t) + \ldots + (-\lambda_n t)} \\ &= e^{-(\lambda_1 + \ldots + \lambda_n)t} \\ &= e^{-\lambda t}\end{aligned}$$

Thus, the CDF of X is $1 - e^{-\lambda t}$, which matches the CDF of an exponential random variable with parameter λ. Since the CDF uniquely determines the distribution, this allows us to conclude that X is indeed exponentially distributed with parameter λ.

(b) Without loss of generality, let X_1 be the minimum of the three random variables. We also know the median is the second smallest of the random variables, so we can think of it as the minimum of the remaining two random variables. Then:

$$\begin{aligned}\mathbb{P}(\min(X_2, X_3) - X_1 > s) &= \mathbb{P}(\min(X_2, X_3) > m + s \mid X_1 > m) \\ &= \mathbb{P}(\min(X_2, X_3) > m + s) \\ &= \mathbb{P}(X_2 > m + s, X_3 > m + s) \\ &= \mathbb{P}(X_2 > m + s)\mathbb{P}(X_3 > m + s) \\ &= e^{-\lambda(m+s)} e^{-\lambda(m+s)} \\ &= e^{-2\lambda(m+s)}\end{aligned}$$

Intuitively, suppose you knew that the value of the minimum of the three random variables was m. Then asking for the probability that the difference is at least s is exactly asking for the probability that the minimum of the two remaining random variables is at least $s+m$. But you know that they're both at least m (since m is the minimum), so the memoryless property gives us that the distribution of this difference is exactly the same as the distribution of the minimum of two exponential random variables! Then, the distribution is exponential with parameter 2λ, and the probability is $e^{-2\lambda(m+s)}$.

(c) By linearity of expectation, the expected value of the median is the expectation of the minimum plus the expected difference between the median and the minimum. From part a, we know that the minimum is exponentially distributed with parameter 3λ, so its expectation is $\frac{1}{3\lambda}$. Then, from part b, that the expectation of the difference is exponentially distributed with parameter 2λ, so its expectation is $\frac{1}{2\lambda}$.

Putting these two together, we have that the expected value of the median is $\frac{1}{3\lambda}+\frac{1}{2\lambda}=\frac{5}{6\lambda}$.

. . . .

In case you are not satisfied with the above explanation, then here is a more formal proof of the fact: If $X_{(1)} := \min\{X_1,X_2,X_3\}$ and $X_{(2)}$ is the median of X_1,X_2,X_3, then $X_{(2)}-X_{(1)}$ is exponentially distributed with parameter 2λ.

To prove that $X_{(2)}-X_{(1)}$ is exponentially distributed, we will calculate $\mathbb{P}\{X_{(2)}-X_{(1)}\geq x\}$ and show that it matches the tail probability of an exponential distribution. So, let $x>0$. First, we will use conditioning to write our event $\{X_{(2)}-X_{(1)}\geq x\}$ in terms of our original random variables X_1, X_2, and X_3.

$$\begin{aligned}\mathbb{P}\{X_{(2)}-X_{(1)}\geq x\} &= 3\mathbb{P}\{X_{(2)}-X_{(1)}\geq x,\, X_{(1)}=X_1\} \qquad \text{(by symmetry)}\\ &= 3\mathbb{P}\{\min(X_2,X_3)-X_1\geq x,\, \min(X_2,X_3)\geq X_1\}\end{aligned}$$

This almost looks like what we want to apply the memoryless property of the exponential distribution. Our next step is to condition on the value of X_1. Let f denote the density function for the exponential distribution.

$$\begin{aligned}&= 3\int_0^\infty \mathbb{P}\{\min(X_2,X_3)\geq x+X_1,\, \min(X_2,X_3)\geq X_1 \mid X_1=x_1\}f(x_1)\,\mathrm{d}x_1\\ &= 3\int_0^\infty \mathbb{P}\{\min(X_2,X_3)\geq x+x_1,\, \min(X_2,X_3)\geq x_1 \mid X_1=x_1\}f(x_1)\,\mathrm{d}x_1\\ &= 3\int_0^\infty \mathbb{P}\{\min(X_2,X_3)\geq x+x_1,\, \min(X_2,X_3)\geq x_1\}f(x_1)\,\mathrm{d}x_1\end{aligned}$$

Here, we dropped the conditioning because the random variables $\min(X_2, X_3)$ and X_1 are independent. Now, we can apply the memoryless property.

$$
\begin{aligned}
&= 3\int_0^\infty \mathbb{P}\{\min(X_2,X_3) \ge x_1\} \\
&\qquad \times \mathbb{P}\{\min(X_2,X_3) \ge x + x_1 \mid \min(X_2,X_3) \ge x_1\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\int_0^\infty \mathbb{P}\{\min(X_2,X_3) \ge x_1\}\mathbb{P}\{\min(X_2,X_3) \ge x\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\mathbb{P}\{\min(X_2,X_3) \ge x\}\int_0^\infty \mathbb{P}\{\min(X_2,X_3) \ge x_1\} f(x_1)\,\mathrm{d}x_1
\end{aligned}
$$

Now, we introduce X_1 back into the integral (trust me, this will work out).

$$
\begin{aligned}
&= 3\mathbb{P}\{\min(X_2,X_3) \ge x\}\int_0^\infty \mathbb{P}\{\min(X_2,X_3) \ge x_1 \mid X_1 = x_1\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\mathbb{P}\{\min(X_2,X_3) \ge x\}\int_0^\infty \mathbb{P}\{\min(X_2,X_3) \ge X_1 \mid X_1 = x_1\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\mathbb{P}\{\min(X_2,X_3) \ge x\}\mathbb{P}\{\min(X_2,X_3) \ge X_1\} \\
&= \mathbb{P}\{\min(X_2,X_3) \ge x\} \qquad \text{(by symmetry).}
\end{aligned}
$$

We have shown that the tail probabilities of $X_{(2)} - X_{(1)}$ matches that of $\min(X_2, X_3)$, that is, $X_{(2)} - X_{(1)}$ has the same distribution as $\min(X_2, X_3)$. Since $\min(X_2, X_3)$ has the exponential distribution with parameter 2λ, then so does $X_{(2)} - X_{(1)}$.

4 Chebyshev's Inequality vs. Central Limit Theorem

Note 17
Note 21

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the following distribution:

$$
\mathbb{P}[X_i = -1] = \frac{1}{12}; \qquad \mathbb{P}[X_i = 1] = \frac{9}{12}; \qquad \mathbb{P}[X_i = 2] = \frac{2}{12}.
$$

(a) Calculate the expectations and variances of X_1, $\sum_{i=1}^n X_i$, $\sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, and

$$
Z_n = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}.
$$

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n| \ge 2]$.

(c) Use b from the previous part to bound $\mathbb{P}[Z_n \ge 2]$ and $\mathbb{P}[Z_n \le -2]$.

(d) As $n \to \infty$, what is the distribution of Z_n?

(e) We know that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{P}[|Z| \le 2] = \Phi(2) - \Phi(-2) \approx 0.9545$. As $n \to \infty$, provide approximations for $\mathbb{P}[Z_n \ge 2]$ and $\mathbb{P}[Z_n \le -2]$.

Solution:

(a) Firstly, let us calculate $\mathbb{E}[X_1]$ and $\mathrm{Var}(X_1)$; we have

$$\mathbb{E}[X_1] = -\frac{1}{12} + \frac{9}{12} + \frac{4}{12} = 1$$
$$\mathrm{Var}(X_1) = \frac{1}{12} \cdot 2^2 + \frac{9}{12} \cdot 0^2 + \frac{2}{12} \cdot 1^2 = \frac{1}{2}.$$

Using linearity of expectation and variance (since $X_1, \ldots, X_n$ are independent), we find that

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n$$
$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n}{2}$$

Again, by linearity of expectation,

$$\mathbb{E}\left[\sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i - n\right] = n - n = 0.$$

Subtracting a constant does not change the variance, so

$$\mathrm{Var}\left(\sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])\right) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i - n\right) = \frac{n}{2},$$

as before.

Using the scaling properties of the expectation and variance, we finally have

$$\mathbb{E}[Z_n] = \mathbb{E}\left[\frac{\sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}\right] = \frac{1}{\sqrt{n/2}} \mathbb{E}\left[\sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])\right] = \frac{0}{\sqrt{n/2}} = 0$$
$$\mathrm{Var}(Z_n) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}\right) = \frac{1}{n/2} \mathrm{Var}\left(\sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])\right) = \frac{n/2}{n/2} = 1$$

(b) Using Chebyshev's, we have

$$\mathbb{P}[|Z_n| \geq 2] \leq \frac{\mathrm{Var}(Z_n)}{2^2} = \frac{1}{4}$$

since $\mathbb{E}[Z_n] = 0$ and $\mathrm{Var}(Z_n) = 1$ as we computed in the previous part.

(c) $\frac{1}{4}$ for both, since we have

$$\mathbb{P}[Z_n \geq 2] \leq \mathbb{P}[|Z_n| \geq 2]$$
$$\mathbb{P}[Z_n \leq -2] \leq \mathbb{P}[|Z_n| \geq 2]$$

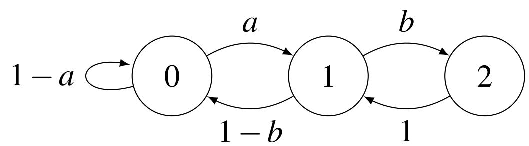
(d) By the Central Limit Theorem, we know that $Z_n \to \mathcal{N}(0,1)$, the standard normal distribution.

(e) Since $Z_n \to \mathcal{N}(0,1)$, we can approximate $\mathbb{P}[|Z_n| \geq 2] \approx 1 - 0.9545 = 0.0455$. By the symmetry of the normal distribution, $\mathbb{P}[Z_n \geq 2] = \mathbb{P}[Z_n \leq -2] \approx 0.0455/2 = 0.02275$.

It is interesting to note that the CLT provides a much smaller answer than Chebyshev. This is due to the fact that the CLT is applied to a particular kind of random variable, namely the (scaled) sum of a bunch of random variables. Chebyshev's inequality, however, holds for any random variable, and is therefore weaker.

5 Analyze a Markov Chain

Note 22 Consider a Markov chain with the state diagram shown below where $a, b \in (0,1)$.



Here, we let $X(n)$ denote the state at time n.

(a) Show that this Markov chain is aperiodic.

(b) Calculate $\mathbb{P}[X(1) = 1, X(2) = 0, X(3) = 0, X(4) = 1 \mid X(0) = 0]$.

(c) Calculate the invariant distribution.

Solution:

(a) The Markov chain is irreducible because $a, b \in (0,1)$. Also, $P(0,0) > 0$, so that

$$\gcd\{n > 0 \mid P^n(0,0) > 0\} = \gcd\{1,2,3,\dots\} = 1,$$

which shows that the Markov chain is aperiodic.

We can also notice from the definition of aperiodicity that if a Markov chain has a self loop with nonzero probability, it is aperiodic. In particular, a self loop implies that the smallest number of steps we need to take to get from a state back to itself is 1. In this case, since $P(0,0) > 0$, we have a self loop with nonzero probability, which makes the Markov chain aperiodic.

(b) As a result of the Markov property, we know our state at timestep n depends only on timestep $n-1$. Looking at the transition probabilities, we see that the final expression is

$$P(0,1) \times P(1,0) \times P(0,0) \times P(0,1) = a(1-b)(1-a)a.$$

(c) The balance equations are

$$\begin{cases} \pi(0) = (1-a)\pi(0) + (1-b)\pi(1) \\ \pi(1) = a\pi(0) + \pi(2) \end{cases} \implies \begin{cases} a\pi(0) = (1-b)\pi(1) \\ \pi(1) = a\pi(0) + \pi(2) \end{cases}$$

$$\implies \begin{cases} a\pi(0) = (1-b)\pi(1) \\ \pi(1) = a\left(\frac{1-b}{a}\pi(1)\right) + \pi(2) \end{cases}$$

$$\implies \begin{cases} a\pi(0) = (1-b)\pi(1) \\ b\pi(1) = \pi(2) \end{cases}$$

As a side note, these last equations express the equality of the probability of a jump from i to $i+1$ and from $i+1$ to i, for $i = 0$ and $i = 1$, respectively. These relations are also called the "detailed balance equations".

From these equations we find successively that

$$\pi(1) = \frac{a}{1-b}\pi(0) \qquad \pi(2) = b\pi(1) = \frac{ab}{1-b}\pi(0).$$

The normalization equation is

$$1 = \pi(0) + \pi(1) + \pi(2) = \pi(0)\left(1 + \frac{a}{1-b} + \frac{ab}{1-b}\right)$$
$$1 = \pi(0)\left(\frac{1-b+a+ab}{1-b}\right)$$

so that

$$\pi(0) = \frac{1-b}{1-b+a+ab}.$$

Thus,

$$\pi(0) = \frac{1-b}{1-b+a+ab} \qquad \pi(1) = \frac{a}{1-b+a+ab} \qquad \pi(2) = \frac{ab}{1-b+a+ab}$$

Or in vector form,

$$\pi = \frac{1}{1-b+a+ab}\begin{bmatrix}1-b & a & ab\end{bmatrix}.$$