1 Short Answer

Note 21

(a) Let X be uniform on the interval $[0,2]$, and define $Y = 4X^2 + 1$. Find the PDF, CDF, expectation, and variance of Y.

(b) Let X and Y have joint distribution

$$f(x,y) = \begin{cases} cxy + \frac{1}{4} & x \in [1,2] \text{ and } y \in [0,2] \\ 0 & \text{otherwise.} \end{cases}$$

Find the constant c. Are X and Y independent?

(c) Let $X \sim \text{Exp}(3)$.

(i) Find probability that $X \in [0, 1]$.

(ii) Let $Y = \lfloor X \rfloor$. For each $k \in \mathbb{N}$, what is the probability that $Y = k$? Write the distribution of Y in terms of one of the famous distributions; provide that distribution's name and parameters.

(d) Let $X_i \sim \text{Exp}(\lambda_i)$ for $i = 1, \ldots, n$ be mutually independent. It is a (very nice) fact that $\min(X_1, \ldots, X_n) \sim \text{Exp}(\mu)$. Find μ.

Solution:

(a) Let's begin with the CDF. It will first be useful to recall that

$$F_X(t) = \mathbb{P}[X \le t] = \begin{cases} 0 & t \le 0 \\ \frac{t}{2} & t \in [0,2] \\ 1 & t \ge 2 \end{cases}.$$

Since Y is defined in terms of X, we can compute that

$$\begin{aligned}
F_Y(t) = \mathbb{P}[Y \le t] = \mathbb{P}[4X^2 + 1 \le t] \\
&= \mathbb{P}\left[X^2 \le \frac{t-1}{4}\right] \\
&= \mathbb{P}\left[X \le \frac{1}{2}\sqrt{t-1}\right] \\
&= F_X\left(\frac{1}{2}\sqrt{t-1}\right) \\
&= \begin{cases} 0 & t \le 1 \\ \frac{1}{4}\sqrt{t-1} & t \in [1, 17] \\ 1 & t \ge 17 \end{cases}
\end{aligned}$$

where in the third line we use that $X \in [0, 2]$, and in the final line we have used the PDF for X. We know that the PDF can be found by taking the derivative of the CDF, so

$$f_Y(t) = \frac{\mathrm{d}}{\mathrm{d}t} F_Y(t) = \begin{cases} \frac{1}{8\sqrt{t-1}} & t \in [1, 17] \\ 0 & \text{else} \end{cases}.$$

By linearity of expectation, we have $\mathbb{E}[Y] = \mathbb{E}[4X^2 + 1] = 4\,\mathbb{E}[X^2] + 1$. There are a couple ways to compute $\mathbb{E}[X^2]$.

One way is to use the fact that $\operatorname{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, so $\mathbb{E}[X^2] = \operatorname{Var}(X) + \mathbb{E}[X]^2$. Since $X \sim \text{Uniform}[0, 2]$, we know $\operatorname{Var}(X) = \frac{1}{3}$ and $\mathbb{E}[X] = 1$; this means

$$\mathbb{E}[X^2] = \operatorname{Var}(X) + \mathbb{E}[X]^2 = \frac{1}{3} + 1^2 = \frac{4}{3}.$$

Another way is to use LOTUS and integrate directly:

$$\mathbb{E}[X^2] = \int_0^2 t^2 f_X(t)\,\mathrm{d}t = \int_0^2 t^2 \cdot \frac{1}{2}\,\mathrm{d}t = \frac{1}{2}\left(\frac{1}{3}2^3\right) = \frac{4}{3}.$$

Plugging this in, we have $\mathbb{E}[Y] = 4\,\mathbb{E}[X^2] + 1 = 4 \cdot \frac{4}{3} + 1 = \frac{19}{3}$.

For the variance, we have $\operatorname{Var}(Y) = \operatorname{Var}(4X^2 + 1) = 16\operatorname{Var}(X^2) = 16(\mathbb{E}[X^4] - \mathbb{E}[X^2]^2)$. Here, we already know $\mathbb{E}[X^2] = \frac{4}{3}$, so we only need to compute $\mathbb{E}[X^4]$:

$$\mathbb{E}[X^4] = \int_0^2 t^4 f_X(t)\,\mathrm{d}t = \int_0^2 t^4 \cdot \frac{1}{2}\,\mathrm{d}t = \frac{1}{2}\left(\frac{1}{5}2^5\right) = \frac{16}{5}.$$

Putting this together, we have

$$\operatorname{Var}(Y) = 16(\mathbb{E}[X^4] - \mathbb{E}[X^2]^2) = 16\left(\frac{16}{5} - \frac{16}{9}\right) = \frac{1024}{45}.$$

(b) To find the correct constant, we use the fact that a PDF must integrate to one. In particular,

$$1 = \int_1^2 \int_0^2 (cxy + 1/4)\,\mathrm{d}y\,\mathrm{d}x = 3c + \frac{1}{2},$$

so $c = 1/6$. In order to check independence, we need to first find the marginal distributions of X and Y:

$$\begin{aligned} f_X(x) &= \int_0^2 f(x,y)\,\mathrm{d}y = 1/2 + x/3 \\ f_Y(y) &= \int_1^2 f(x,y)\,\mathrm{d}x = 1/4 + y/4. \end{aligned}$$

Since

$$f_X(x) f_Y(y) = \frac{1}{8} + \frac{y}{8} + \frac{x}{12} + \frac{xy}{12} \neq \frac{1}{4} + \frac{xy}{6} = f(x,y),$$

the random variables are not independent.

(c) (i) Since $X \sim \mathrm{Exp}(3)$, the CDF of X is $F(x) = 1 - e^{-3x}$. Thus we have

$$\mathbb{P}[X \in [0,1]] = \int_0^1 f(x)\,\mathrm{d}x = F(1) - F(0) = (1 - e^{-3}) - (1 - e^0) = 1 - e^{-3}.$$

(ii) Similarly, if $Y = \lfloor X \rfloor$, then $Y = k$ exactly when $X \in [k, k+1)$, so

$$\begin{aligned} \mathbb{P}[Y = k] &= \mathbb{P}[X \in [k, k+1)] \\ &= \int_k^{k+1} f(x)\,\mathrm{d}x \\ &= F(k+1) - F(k) \\ &= (1 - e^{-3(k+1)}) - (1 - e^{-3k}) \\ &= e^{-3k} - e^{-3(k+1)} \\ &= e^{-3k}\left(1 - e^{-3}\right) = (e^{-3})^k \left(1 - e^{-3}\right). \end{aligned}$$

In other words, $Y = W - 1$ for $W \sim \mathrm{Geometric}(1 - e^{-3})$.

(d) Since the X_i are independent,

$$\begin{aligned} \mathbb{P}[\min(X_1, \ldots, X_n) \le t] &= 1 - \mathbb{P}[X_1 > t, X_2 > t, \cdots\cdot X_n > t] \\ &= 1 - \mathbb{P}[X_1 > t] \cdot \mathbb{P}[X_2 > t] \cdot\cdots\cdot \mathbb{P}[X_n > t] \qquad \text{(by independence)} \\ &= 1 - e^{-\lambda_1 t} e^{-\lambda_2 t} \cdot\cdots\cdot e^{-\lambda_n t} \\ &= 1 - e^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)t}. \end{aligned}$$

This is exactly the CDF of an $\mathrm{Exp}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$ random variable, so $\mu = \lambda_1 + \cdots + \lambda_n$.

2 Uniform Estimation

Note 17
Note 21

Let $U_1, \ldots, U_n \stackrel{\text{iid}}{\sim} \text{Uniform}(-\theta, \theta)$ for some unknown $\theta \in \mathbb{R}$, $\theta > 0$. We wish to estimate θ from the data $U_1, \ldots, U_n$.

(a) Why would using the sample mean $\overline{U} = \frac{1}{n}\sum_{i=1}^{n} U_i$ fail in this situation?

(b) Find the PDF of U_i^2 for $i \in \{1, \ldots, n\}$.

(c) Consider the following variance estimate:

$$V = \frac{1}{n}\sum_{i=1}^{n} U_i^2.$$

Show that for large n, the distribution of V is close to one of the famous ones, and provide its name and parameters.

(d) Use part (c) to construct an unbiased estimator for θ^2 that uses all the data.

(e) Let $\sigma^2 = \text{Var}[U_i^2]$. We wish to construct a confidence interval for θ^2 with a significance level of δ, where $0 < \delta < 1$.

(i) Without any assumption on the magnitude of n, construct a confidence interval for θ^2 with a significance level of δ using your estimator from part (d).

(ii) Suppose n is large. Construct an approximate confidence interval for θ^2 with a significance level of δ using your estimator from part (d). You may leave your answer in terms of Φ and Φ^{-1}, the normal CDF and its inverse.

Solution:

(a) The sample mean would not work well as an estimator for θ because it has expected value 0, not θ.

(b) We will proceed by finding the CDF of U_i^2 first, and then taking the derivative after to get the PDF. Firstly, note that $0 \le U_i^2 \le \theta^2$, so we have that $\mathbb{P}[U_i^2 \le t] = 0$ when $t \le 0$ and $\mathbb{P}[U_i^2 \le t] = 1$ when $t \ge \theta^2$. When $0 < t < \theta^2$, we have that

$$\mathbb{P}[U_i^2 \le t] = \mathbb{P}[-\sqrt{t} \le U_i \le \sqrt{t}] = \frac{\sqrt{t}}{\theta},$$

hence the CDF of U_i^2 is

$$F(t) = \begin{cases} 0 & \text{if } t \le 0, \\ \frac{\sqrt{t}}{\theta} & \text{if } 0 < t < \theta^2, \text{ and} \\ 1 & \text{if } t \ge \theta^2. \end{cases}$$

Lastly, we take the derivative to get the PDF:

$$f(t) = F'(t) = \begin{cases} \frac{1}{2\theta\sqrt{t}} & \text{if } 0 < t < \theta^2, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

(c) We can see that

$$nV = \sum_{i=1}^{n} U_i^2,$$

so by the Central Limit Theorem, we know that for large n,

$$\frac{nV - n\mathbb{E}[U_1^2]}{\sqrt{n\,\mathrm{Var}(U_1^2)}} \xrightarrow{\text{in distribution}} \mathcal{N}(0,1).$$

Hence, multiplying and adding, we can see that

$$V \xrightarrow{\text{in distribution}} \mathcal{N}\left(\mathbb{E}[U_1^2], \frac{1}{n}\mathrm{Var}\left(U_1^2\right)\right).$$

Now, it remains to calculate both the expectation and variance of U_1^2. We have that

$$\mathbb{E}[U_1^2] = \mathrm{Var}(U_1) + \mathbb{E}[U_1]^2 = \mathrm{Var}(U_1) = \frac{\theta^2}{3},$$

and we have that

$$\mathrm{Var}\left(U_1^2\right) = \mathbb{E}[U_1^4] - \mathbb{E}[U_1^2]^2 = \int_{-\theta}^{\theta} \frac{t^4}{2\theta}\,dt - \frac{\theta^4}{9} = \frac{\theta^4}{5} - \frac{\theta^4}{9} = \frac{4\theta^4}{45},$$

so $V \sim \mathcal{N}\left(\frac{\theta^2}{3}, \frac{4\theta^4}{45n}\right)$.

Alternatively, we can do these calculations using the distribution for U_1^2 derived in a previous part. We have that

$$\mathbb{E}[U_1^2] = \int_0^{\theta^2} t \cdot \frac{1}{2\theta\sqrt{t}}\,dt = \int_0^{\theta^2} \frac{\sqrt{t}}{2\theta}\,dt = \frac{\theta^2}{3},$$

and we have that

$$\mathrm{Var}(U_1^2) = \int_0^{\theta^2} t^2 \cdot \frac{1}{2\theta\sqrt{t}}\,dt - \frac{\theta^4}{9} = \int_0^{\theta^2} \frac{t^{\frac{3}{2}}}{2\theta}\,dt - \frac{\theta^4}{9} = \frac{4\theta^4}{45},$$

so again, $V \sim \mathcal{N}\left(\frac{\theta^2}{3}, \frac{4\theta^4}{45n}\right)$.

(d) We can use $3V$ as our unbiased estimator, as $\mathbb{E}[3V] = \theta^2$ and $\mathrm{Var}(3V) = \frac{4\theta^4}{5n} \to 0$ as $n \to \infty$.

(e) (i) We will use Chebyshev's inequality to bound the probability of deviation from the mean. Firstly, we can compute that

$$\mathrm{Var}(3V) = 9\,\mathrm{Var}(V) = \frac{9\sigma^2}{n}.$$

Moving forward, we have that

$$\mathbb{P}[|3V - \theta^2| \geq c] \leq \frac{\mathrm{Var}(3V)}{c^2} = \frac{9\sigma^2}{nc^2},$$

so in order to guarantee that this probability is less than δ, we need to set

$$\frac{9\sigma^2}{nc^2} \leq \delta \implies c \geq \frac{3\sigma}{\sqrt{\delta n}},$$

so our confidence interval is thus $[3V - \frac{3\sigma}{\sqrt{\delta n}}, 3V + \frac{3\sigma}{\sqrt{\delta n}}]$.

(ii) With the assumption that n is large, we can claim via the CLT that $3V \sim \mathcal{N}(\theta^2, \frac{9\sigma^2}{n})$, so in particular, $\frac{\sqrt{n}(3V-\theta^2)}{3\sigma}$ is a standard normal. Thus, we have that

$$\mathbb{P}[|3V-\theta^2| > c] = \mathbb{P}\left[\frac{\sqrt{n}|3V-\theta^2|}{3\sigma} > \frac{c\sqrt{n}}{3\sigma}\right] = 1 - \Phi\left(\frac{c\sqrt{n}}{3\sigma}\right) + \Phi\left(-\frac{c\sqrt{n}}{3\sigma}\right).$$

We can further simplify the right hand side of this to

$$\mathbb{P}[|3V-\theta^2| > c] = 2\Phi\left(-\frac{c\sqrt{n}}{3\sigma}\right),$$

hence to get a significance level of δ, we can set

$$2\Phi\left(-\frac{c\sqrt{n}}{3\sigma}\right) = \delta \implies c = -\frac{3\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\delta}{2}\right).$$

Hence, our confidence interval is $\left[3V + \frac{3\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\delta}{2}\right), 3V - \frac{3\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\delta}{2}\right)\right]$.

3 Darts with Friends

Michelle and Alex are playing darts. Being the better player, Michelle's aim follows a uniform distribution over a disk of radius 1 around the center. Alex's aim follows a uniform distribution over a disk of radius 2 around the center.

(a) Let the distance of Michelle's throw from the center be denoted by the random variable X and let the distance of Alex's throw from the center be denoted by the random variable Y.

- What's the cumulative distribution function of X?
- What's the cumulative distribution function of Y?
- What's the probability density function of X?
- What's the probability density function of Y?

(b) What's the probability that Michelle's throw is closer to the center than Alex's throw? What's the probability that Alex's throw is closer to the center?

(c) What's the cumulative distribution function of $U = \max\{X,Y\}$?

(d) What's the cumulative distribution function of $V = \min\{X,Y\}$?

(e) What is the expectation of the absolute difference between Michelle's and Alex's distances from the center, that is, what is $\mathbb{E}[|X-Y|]$? [*Hint*: Use parts (c) and (d), together with the continuous version of the tail sum formula, which states that $\mathbb{E}[Z] = \int_0^\infty \mathbb{P}[Z \geq z] dz$.]

Solution:

(a)
- To get the cumulative distribution function of X, we'll consider the ratio of the area where the distance to the center is less than x, compared to the entire available area. This gives us the following expression:

$$\mathbb{P}[X \leq x] = \frac{\pi x^2}{\pi} = x^2, \qquad x \in [0,1].$$

- Using the same approach as the previous part:

$$\mathbb{P}[Y \leq y] = \frac{\pi y^2}{\pi \cdot 4} = \frac{y^2}{4}, \qquad y \in [0,2].$$

- We'll take the derivative of the CDF to get the following:

$$f_X(x) = \frac{\mathrm{d}\mathbb{P}[X \leq x]}{\mathrm{d}x} = 2x, \qquad x \in [0,1].$$

- Using the same approach as the previous part:

$$f_Y(y) = \frac{\mathrm{d}\mathbb{P}[Y \leq y]}{\mathrm{d}y} = \frac{y}{2}, \qquad y \in [0,2].$$

(b) We'll condition on Alex's outcome and then integrate over all the possibilities to get the marginal $\mathbb{P}(X \leq Y)$ as following:

$$\begin{aligned}\mathbb{P}[X \leq Y] &= \int_0^2 \mathbb{P}[X \leq Y \mid Y = y] f_Y(y)\,\mathrm{d}y = \int_0^1 y^2 \times \frac{y}{2}\,\mathrm{d}y + \int_1^2 1 \times \frac{y}{2}\,\mathrm{d}y \\ &= \frac{1}{8} + \frac{3}{4} = \frac{7}{8}.\end{aligned}$$

Note the range within which $\mathbb{P}[X \leq Y] = 1$. This allowed us to separate the integral to simplify our solution. Using this, we can get $\mathbb{P}[Y \leq X]$ by the following:

$$\mathbb{P}[Y \leq X] = 1 - \mathbb{P}[X \leq Y] = \frac{1}{8}$$

A similar approach to the integral above could be used to verify this result:

$$\mathbb{P}[Y \leq X] = \int_0^1 \mathbb{P}[Y \leq X \mid X = x] f_X(x)\,\mathrm{d}x = \int_0^1 \frac{x^2}{4} 2x\,\mathrm{d}x = \frac{1}{2}\int_0^1 x^3\,\mathrm{d}x = \frac{1}{8}.$$

(c) Getting the CDF of U relies on the insight that for the maximum of two random variables to be smaller than a value, they both need to be smaller than that value. Using this we can get the following result for $u \in [0,1]$:

$$\mathbb{P}[U \le u] = \mathbb{P}[X \le u]\mathbb{P}[Y \le u] = \left(u^2\right)\left(\frac{u^2}{4}\right) = \frac{u^4}{4}.$$

For $u \in [1,2]$ we have $\mathbb{P}[X \le u] = 1$; this makes

$$\mathbb{P}[U \le u] = \mathbb{P}[Y \le u] = \frac{u^2}{4}.$$

For $u > 2$ we have $\mathbb{P}[U \le u] = 1$ since CDFs of both X and Y are 1 in this range.

(d) Getting the CDF of V relies on a similar insight that for the minimum of two random variables to be greater than a value, they both need to be greater than that value. Taking the complement of this will give us the CDF of V. This allows us to get the following result. For $v \in [0,1]$:

$$\begin{aligned}\mathbb{P}[V \le v] &= 1 - \mathbb{P}[V \ge v] = 1 - \mathbb{P}[X \ge v]\mathbb{P}[Y \ge v] = 1 - \big(1 - \mathbb{P}[X \le v]\big)\big(1 - \mathbb{P}[Y \le v]\big) \\ &= 1 - \left(1 - v^2\right)\left(1 - \frac{v^2}{4}\right) = \frac{5v^2}{4} - \frac{v^4}{4}.\end{aligned}$$

For $v > 1$, we get $\mathbb{P}[X > v] = 0$, making $\mathbb{P}[V \le v] = 1$.

(e) We can subtract V from U to get this difference. Using the tail-sum formula to calculate the expectation, we can get the following result:

$$\begin{aligned}\mathbb{E}[|X - Y|] = \mathbb{E}[U - V] = \mathbb{E}[U] - \mathbb{E}[V] &= \int_0^2 \mathbb{P}[U \ge u]\,\mathrm{d}u - \int_0^1 \mathbb{P}[V \ge v]\,\mathrm{d}v \\ &= \int_0^1 \left(1 - \frac{u^4}{4}\right)\mathrm{d}u + \int_1^2 \left(1 - \frac{u^2}{4}\right)\mathrm{d}u - \int_0^1 \left(1 - \frac{5v^2}{4} + \frac{v^4}{4}\right)\mathrm{d}v \\ &= \frac{19}{20} + \frac{5}{12} - \frac{19}{30} = \frac{11}{15}.\end{aligned}$$

Alternatively, you could derive the density of U and V and use those to calculate the expectation. For $u \in [0,1]$:

$$f_U(u) = \frac{\mathrm{d}\mathbb{P}[U \le u]}{\mathrm{d}u} = u^3.$$

For $u \in [1,2]$:

$$f_U(u) = \frac{\mathrm{d}\mathbb{P}[U \le u]}{\mathrm{d}u} = \frac{u}{2}.$$

Using this we can calculate $\mathbb{E}[U]$ as:

$$\mathbb{E}[U] = \int_0^2 u f_U(u)\,\mathrm{d}u = \int_0^1 u^4\,\mathrm{d}u + \frac{1}{2}\int_1^2 u^2\,\mathrm{d}u = \frac{1}{5} + \frac{7}{6} = \frac{41}{30}.$$

To calculate $\mathbb{E}[V]$ we will use the following PDF for $v \in [0,1]$:

$$f_V(v) = \frac{\mathrm{d}\mathbb{P}[V \le v]}{\mathrm{d}v} = \frac{5v}{2} - v^3.$$

We can get the $\mathbb{E}[V]$ by the following:

$$\mathbb{E}[V] = \int_0^1 v f_V(v)\,\mathrm{d}v = \int_0^1 \left(\frac{5v^2}{2} - v^4\right)\mathrm{d}v = \frac{5}{6} - \frac{1}{5} = \frac{19}{30}.$$

Combining the two results gives us the same result as above:

$$\mathbb{E}[|X-Y|] = \mathbb{E}[U-V] = \mathbb{E}[U] - \mathbb{E}[V] = \frac{41}{30} - \frac{19}{30} = \frac{11}{15}.$$