## 1 Tellers

Note 17

Imagine that $X$ is the number of customers that enter a bank at a given hour. To simplify everything, in order to serve $n$ customers you need at least $n$ tellers. One less teller and you won't finish serving all of the customers by the end of the hour. You are the manager of the bank and you need to decide how many tellers there should be in your bank so that you finish serving all of the customers in time. You need to be sure that you finish in time with probability at least 95%.

(a) Assume that from historical data you have found out that $\mathbb{E}[X] = 5$. How many tellers should you have?

(b) Now assume that you have also found out that $\text{Var}(X) = 5$. Now how many tellers do you need?

**Solution:**

(a) We should apply Markov's, since we only have access to the expected value. Aiming for the probability of finishing $t$ in time to be least 95% is equivalent to aiming to limit the probability of not finishing (or in other words, taking more time to finish $t$ customers) as 5%. So, we want

$$\mathbb{P}[X \geq t] \leq 0.05$$

Using Markov's, we know $\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$. Therefore we want $\frac{\mathbb{E}[X]}{t} = 0.05$ for the above inequality to hold. Therefore, $t = 100$ to be able to limit the probability of not finishing under 5%. The inequality becomes $\mathbb{P}[X \geq 100] \leq 0.05$ as wanted. So 99 tellers are needed.

(b) Now that we have access to the variance as well, we can try apply Chebyshev's. Note that Markov's inequality is still correct, but Chebyshev's gives us a tigher bound here. Same as part (a), aiming for the probability of finishing $t$ in time to be least 95% is equivalent to aiming to limit the probability of not finishing (or in other words, taking more time to finish $t$ customers) as 5%. So, we want

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq t] \leq 0.05$$

Using Chebyshev's we know $\mathbb{P}[|X - \mathbb{E}[X]| \geq t] \leq \frac{\text{Var}(X)}{t^2}$. Plugging in $\mathbb{E}[X] = 5$ and $\text{Var}(X) = 5$, we get $\mathbb{P}[|X - 5| \geq t] \leq \frac{5}{t^2}$. Since we want to limit $\mathbb{P}[|X - 5| \geq t] \leq 0.05$, we get $\frac{5}{t^2} = 0.05$. Thus $t^2 = 100$ and $t = 10$. Now plugging $t = 10$:

$$\mathbb{P}[|X - 5| \geq 10] = \mathbb{P}[X \geq 5 + 10] = \mathbb{P}[X \geq 15] \leq 0.05$$

as wanted. Thus, 14 tellers are needed this time.

## 2 Just One Tail, Please

Note 17 Let $X$ be some random variable with finite mean and variance which is not necessarily non-negative. The *extended* version of Markov's Inequality states that for a non-negative function $\varphi(x)$ which is monotonically increasing for $x > 0$ and some constant $\alpha > 0$,

$$\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[\varphi(X)]}{\varphi(\alpha)}$$

Suppose $\mathbb{E}[X] = 0$, $\mathrm{Var}(X) = \sigma^2 < \infty$, and $\alpha > 0$.

(a) Use the extended version of Markov's Inequality stated above with $\varphi(x) = (x+c)^2$, where $c$ is some positive constant, to show that:

$$\mathbb{P}[X \geq \alpha] \leq \frac{\sigma^2 + c^2}{(\alpha + c)^2}$$

(b) Note that the above bound applies for all positive $c$, so we can choose a value of $c$ to minimize the expression, yielding the best possible bound. Find the value for $c$ which will minimize the RHS expression (you may assume that the expression has a unique minimum).

We can plug in the minimizing value of $c$ you found in part (b) to prove the following bound:

$$\mathbb{P}[X \geq \alpha] \leq \frac{\sigma^2}{\alpha^2 + \sigma^2}.$$

This bound is also known as Cantelli's inequality.

(c) Recall that Chebyshev's inequality provides a two-sided bound. That is, it provides a bound on $\mathbb{P}[|X - \mathbb{E}[X]| \geq \alpha] = \mathbb{P}[X \geq \mathbb{E}[X] + \alpha] + \mathbb{P}[X \leq \mathbb{E}[X] - \alpha]$. If we only wanted to bound the probability of one of the tails, e.g. if we wanted to bound $\mathbb{P}[X \geq \mathbb{E}[X] + \alpha]$, it is tempting to just divide the bound we get from Chebyshev's by two.

(i) Provide an example of a random variable $X$ (does not have to be zero-mean) and a constant $\alpha$ such that using this method (dividing by two to bound one tail) is not correct, that is, $\mathbb{P}[X \geq \mathbb{E}[X] + \alpha] > \frac{\mathrm{Var}(X)}{2\alpha^2}$ or $\mathbb{P}[X \leq \mathbb{E}[X] - \alpha] > \frac{\mathrm{Var}(X)}{2\alpha^2}$.

Now we see the use of the bound proven in part (b) - it allows us to bound just one tail while still taking variance into account, and does not require us to assume any property of the random variable. Note that the bound is also always guaranteed to be less than 1 (and therefore at least somewhat useful), unlike Markov's and Chebyshev's inequality!

(d) Let's try out our new bound on a simple example. Suppose $X$ is a positively-valued random variable with $\mathbb{E}[X] = 3$ and $\mathrm{Var}(X) = 2$.

(i) What bound would Markov's inequality give for $\mathbb{P}[X \geq 5]$?

(ii) What bound would Chebyshev's inequality give for $\mathbb{P}[X \geq 5]$?

(iii) What bound would Cantelli's Inequality give for $\mathbb{P}[X \geq 5]$? (*Note*: Recall that Cantelli's Inequality only applies for zero-mean random variables.)

**Solution:**

(a) Note that $\sigma^2 = \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2]$. Using the inequality presented in the problem, we have:

$$\begin{aligned}
\mathbb{P}[X \geq \alpha] &\leq \frac{\mathbb{E}[(X+c)^2]}{(\alpha+c)^2} \\
&= \frac{\mathbb{E}[X^2 + 2cX + c^2]}{(\alpha+c)^2} \\
&= \frac{\mathbb{E}[X^2] + 2c\,\mathbb{E}[X] + c^2}{(\alpha+c)^2} \\
&= \frac{\sigma^2 + c^2}{(\alpha+c)^2}
\end{aligned}$$

(b) We set the derivative with respect to $c$ of the above expression equal to 0, and solve for $c$.

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}c} \frac{\sigma^2 + c^2}{(\alpha+c)^2} \\
&= \frac{2c(\alpha+c)^2 - 2(\alpha+c)(\sigma^2+c^2)}{(\alpha+c)^4} \\
&= 2c(\alpha+c)^2 - 2(\alpha+c)(\sigma^2+c^2) \\
&= \alpha c^2 + (\alpha^2 - \sigma^2)c - \sigma^2\alpha \\
c &= \frac{\sigma^2}{\alpha}
\end{aligned}$$

To get the last step we use the quadratic equation and take the positive solution.

(c) It is possible for one of the tails to contain more probability than the other. One example of a random variable which demonstrates this is $X$, where $\mathbb{P}[X = 0] = 0.75$ and $\mathbb{P}[X = 10] = 0.25$, with $\alpha = 7$. Here, $\mathbb{E}[X] = 2.5$ and $\text{Var}(X) = 100 \cdot 0.25 \cdot 0.75$, so we have:

$$\mathbb{P}[X \geq \mathbb{E}[X] + 7] = 0.25 > \frac{\text{Var}(X)}{2 \cdot 7^2} \approx 0.19$$

(d) (i) Using Markov's:

$$\mathbb{P}[X \geq 5] \leq \frac{\mathbb{E}[X]}{5} = \frac{3}{5}$$

(ii) Using Chebyshev's:

$$\mathbb{P}[X \geq 5] \leq \mathbb{P}[|X - \mathbb{E}[X]| \geq 2] \leq \frac{\text{Var}(X)}{2^2} = \frac{1}{2}$$

(iii) Using bound shown above (Cantelli's):

Since we have the condition that this bound applies to zero-mean random variables, let us define $Y = X - \mathbb{E}[X] = X - 3$. Note that $\text{Var}(Y) = \text{Var}(X)$. With this, we get:

$$\mathbb{P}[X \geq 5] = \mathbb{P}[Y \geq 2] \leq \frac{\text{Var}(Y)}{2^2 + \text{Var}(Y)} = \frac{1}{3}$$

We see that Cantelli's inequality (the bound from part (b)) does better than Chebyshev's, which does better than Markov's (note that having a smaller upper bound is better)! This is a good demonstration on how we might derive better bounds using Markov's inequality, if we know further information about the random variable like its variance.

## 3 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 $c$ (Hint: remember that the PDF must integrate to 1). Are $X$ and $Y$ independent?

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

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

  (ii) Let $Y = \lfloor X \rfloor$, where the floor operator is defined as: $(\forall x \in [k, k+1))(\lfloor x \rfloor = k)$. 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 $\mu$.

**Solution:**

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

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

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 $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, so $\mathbb{E}[X^2] = \mathrm{Var}(X) + \mathbb{E}[X]^2$. Since $X \sim \mathrm{Uniform}[0, 2]$, we know $\mathrm{Var}(X) = \frac{1}{3}$ and $\mathbb{E}[X] = 1$; this means

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

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 $\mathrm{Var}(Y) = \mathrm{Var}(4X^2 + 1) = 16\,\mathrm{Var}(X^2) = 16(\mathbb{E}[X^4] - \mathbb{E}[X^2]^2)$. 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

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

(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 $Y$:

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

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 $X \sim \text{Exp}(3)$, 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 $W \sim \text{Geometric}(1 - e^{-3})$.

(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 $\text{Exp}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$ random variable, so $\mu = \lambda_1 + \cdots + \lambda_n$.

# 4 Uniform Distribution

Note 21

You have two fidget spinners, each having a circumference of 10. You mark one point on each spinner as a needle and place each of them at the center of a circle with values in the range $[0,10)$ marked on the circumference. If you spin both (independently) and let $X$ be the position of the first spinner's mark and $Y$ be the position of the second spinner's mark, what is the probability that $X \geq 5$, given that $Y \geq X$?

**Solution:**

First we write down what we want and expand out the conditioning:

$$\mathbb{P}[X \geq 5 \mid Y \geq X] = \frac{\mathbb{P}[Y \geq X \cap X \geq 5]}{\mathbb{P}[Y \geq X]}.$$

$\mathbb{P}[Y \geq X] = 1/2$ by symmetry. To find $\mathbb{P}[Y \geq X \cap X \geq 5]$, it helps a lot to just look at the picture of the probability space and use the continuous uniform law $\mathbb{P}[A] = (\text{area of } A)/(\text{area of } \Omega)$. We are interested in the relative area of the region bounded by $x < y < 10$, $5 < x < 10$ to the entire square bounded by $0 < x < 10$, $0 < y < 10$.

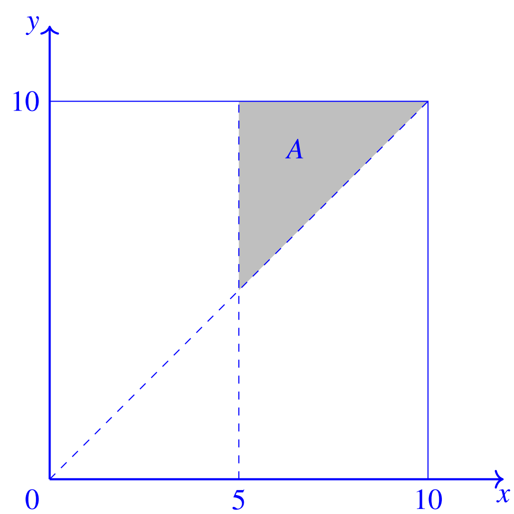


Figure 1: Joint probability density for the spinner.

Looking at the picture in Figure 1, we have

$$\mathbb{P}[Y \geq X \cap X \geq 5] = \frac{5 \cdot 5/2}{10 \cdot 10} = \frac{1}{8},$$

so $\mathbb{P}[X \geq 5 \mid Y \geq X] = (1/8)/(1/2) = 1/4$.

# 5 Darts with Friends

Note 21

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 $Y$.

(i) What's the cumulative distribution function of $X$?

(ii) What's the cumulative distribution function of $Y$?

(iii) What's the probability density function of $X$?

(iv) 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 $U = \max(X,Y)$?

**Solution:**

(a) (i) 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:

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

(ii) Using the same approach as the previous part:

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

(iii) We'll take the derivative of the CDF to get the following:

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

(iv) Using the same approach as the previous part:

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

(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 $u \in [0, 1]$:

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

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

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

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