1 Double-Check Your Intuition Again

Note 20
Note 21

(a) You roll a fair six-sided die and record the result X. You roll the die again and record the result Y.

(i) What is $\text{cov}(X+Y, X-Y)$?

(ii) Prove that $X+Y$ and $X-Y$ are not independent.

For each of the problems below, if you think the answer is "yes" then provide a proof. If you think the answer is "no", then provide a counterexample.

(b) If X is a random variable and $\text{Var}(X) = 0$, then must X be a constant?

(c) If X is a random variable and c is a constant, then is $\text{Var}(cX) = c\,\text{Var}(X)$?

(d) If A and B are random variables with nonzero standard deviations and $\text{Corr}(A,B) = 0$, then are A and B independent?

(e) If X and Y are not necessarily independent random variables, but $\text{Corr}(X,Y) = 0$, and X and Y have nonzero standard deviations, then is $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$?

The two subparts below are **optional** and will not be graded but are recommended for practice.

(f) If X and Y are random variables then is $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$?

(g) If X and Y are independent random variables with nonzero standard deviations, then is

$$\text{Corr}(\max(X,Y), \min(X,Y)) = \text{Corr}(X,Y)?$$

Solution:

(a) (i) Using bilinearity of covariance, we have

$$\begin{aligned}
\text{cov}(X+Y, X-Y) &= \text{cov}(X,X) + \text{cov}(X,Y) - \text{cov}(Y,X) - \text{cov}(Y,Y) \\
&= \text{cov}(X,X) - \text{cov}(Y,Y), \\
&= 0
\end{aligned}$$

where we use that $\text{cov}(X,Y) = \text{cov}(Y,X)$ to get the second equality.

(ii) Observe that $\mathbb{P}[X+Y=7, X-Y=0] = 0$ because if $X-Y=0$, then the sum of our two dice rolls must be even. However, both $\mathbb{P}[X+Y=7]$ and $\mathbb{P}[X-Y=0]$ are nonzero, so $\mathbb{P}[X+Y=7, X-Y=0] \neq \mathbb{P}[X+Y=7] \cdot \mathbb{P}[X-Y=0]$.

(b) Yes. If we write $\mu = \mathbb{E}[X]$, then $0 = \text{Var}(X) = \mathbb{E}[(X-\mu)^2]$ so $(X-\mu)^2$ must be identically 0 since perfect squares are non-negative. Thus $X = \mu$.

(c) No. We have $\text{Var}(cX) = \mathbb{E}[(cX - \mathbb{E}[cX])^2] = c^2\,\mathbb{E}[(X-\mathbb{E}[X])^2] = c^2\,\text{Var}(X)$ so if $\text{Var}(X) \neq 0$ and $c \neq 0$ or $c \neq 1$ then $\text{Var}(cX) \neq c\,\text{Var}(X)$. This does prove that $\sigma(cX) = c\sigma(X)$ though.

(d) No. Let $A = X+Y$ and $B = X-Y$ from part (a). Since A and B are not constants then part (b) says they must have nonzero variances which means they also have nonzero standard deviations. Part (a) says that their covariance is 0 which means they are uncorrelated, and that they are not independent.

Recall from lecture that the converse is true though.

(e) Yes. If $\text{Corr}(X,Y) = 0$, then $\text{cov}(X,Y) = 0$. We have $\text{Var}(X+Y) = \text{cov}(X+Y, X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{cov}(X,Y) = \text{Var}(X) + \text{Var}(Y)$.

(f) Yes. For any values x, y we have $\max(x,y)\min(x,y) = xy$. Thus, $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$.

(g) No. You may be tempted to think that because $(\max(x,y), \min(x,y))$ is either (x,y) or (y,x), then $\text{Corr}(\max(X,Y), \min(X,Y)) = \text{Corr}(X,Y)$ because $\text{Corr}(X,Y) = \text{Corr}(Y,X)$. That reasoning is flawed because $(\max(X,Y), \min(X,Y))$ is not always equal to (X,Y) or always equal to (Y,X) and the inconsistency affects the correlation. It is possible for X and Y to be independent while $\max(X,Y)$ and $\min(X,Y)$ are not.

For a concrete example, suppose X is either 0 or 1 with probability $1/2$ each and Y is independently drawn from the same distribution. Then $\text{Corr}(X,Y) = 0$ because X and Y are independent. Even though X never gives information about Y, if you know $\max(X,Y) = 0$ then you know for sure $\min(X,Y) = 0$.

More formally, $\max(X,Y) = 1$ with probability $3/4$ and 0 with probability $1/4$, and $\min(X,Y) = 1$ with probability $1/4$ and 0 with probability $3/4$. This means

$$\mathbb{E}[\max(X,Y)] = 1 \cdot \frac{3}{4} + 0 \cdot \frac{1}{4} = \frac{3}{4}$$

and

$$\mathbb{E}[\min(X,Y)] = 1 \cdot \frac{1}{4} + 0 \cdot \frac{3}{4} = \frac{1}{4}.$$

Thus,

$$\begin{aligned}\text{cov}(\max(X,Y), \min(X,Y)) &= \mathbb{E}[\max(X,Y)\min(X,Y)] - \frac{3}{16} \\ &= \frac{1}{4} - \frac{3}{16} = \frac{1}{16} \neq 0\end{aligned}$$

We conclude that $\text{Corr}(\max(X,Y), \min(X,Y)) \neq 0 = \text{Corr}(X,Y)$.

2 Dice Games

Note 21

(a) Alice rolls a die until she gets a 1. Let X be the number of total rolls she makes (including the last one), and let Y be the number of rolls on which she gets an even number. Compute $\mathbb{E}[Y \mid X = x]$, and use it to calculate $\mathbb{E}[Y]$.

(b) Bob plays a game in which he starts off with one die. At each time step, he rolls all the dice he has. Then, for each die, if it comes up as an odd number, he puts that die back, and adds a number of dice equal to the number displayed to his collection. (For example, if he rolls a one on the first time step, he puts that die back along with an extra die.) However, if it comes up as an even number, he removes that die from his collection.

Compute the expected number of dice Bob will have after n time steps. (Hint: compute the value of $\mathbb{E}[X_k \mid X_{k-1} = m]$ to derive a recursive expression for X_k, where X_i is the random variable representing the number of dice after i time steps.)

Solution:

(a) Let's compute $\mathbb{E}[Y \mid X = x]$. If Alice makes x total rolls, then before rolling a 1, she makes $x-1$ rolls that are not a 1. Since these rolls are independent, Y follows a binomial distribution with $n = x-1$ and $p = 3/5$, and $\mathbb{E}[Y \mid X = x] = \frac{3}{5}(x-1)$.

Now, we'd like to compute $\mathbb{E}[Y]$. With total expectation, we have

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_x \mathbb{E}[Y \mid X = x]\,\mathbb{P}[X = x] \\
&= \sum_x \frac{3}{5}(x-1)\,\mathbb{P}[X = x] \\
&= \frac{3}{5}\sum_x x \cdot \mathbb{P}[X = x] - \frac{3}{5}\sum_x \mathbb{P}[X = x] \\
&= \frac{3}{5}\mathbb{E}[X] - \frac{3}{5}
\end{aligned}$$

Since X follows a geometric distribution with $p = 1/6$, $\mathbb{E}[X] = 6$, and

$$\mathbb{E}[Y] = \frac{3}{5}\mathbb{E}[X] - \frac{3}{5} = \frac{3}{5} \cdot 6 - \frac{3}{5} = 3.$$

(b) Let X_k be a random variable representing the number of dice after k time steps. In particular, this means that $X_0 = 1$. To compute the number of dice at step k, we first condition on $X_{k-1} = m$. Each one of the m dice is expected to leave behind 2 in its place, since there's a $\frac{1}{2}$ probability that it leaves behind 0 dice, a $\frac{1}{6}$ probability for each of 2, 4, and 6 dice, corresponding to rolling a 1, 3, and 5 respectively.

Therefore, we have $\mathbb{E}[X_k \mid X_{k-1} = m] = 2m$, so with total expectation, we have

$$\begin{aligned}
\mathbb{E}[X_k] &= \sum_m \mathbb{E}[X_k \mid X_{k-1} = m]\,\mathbb{P}[X_{k-1} = m] \\
&= \sum_m 2m \cdot \mathbb{P}[X_{k-1} = m] \\
&= 2\sum_m m \cdot \mathbb{P}[X_{k-1} = m] \\
&= 2\,\mathbb{E}[X_{k-1}]
\end{aligned}$$

This means that we expect to have $\mathbb{E}[X_n] = 2\,\mathbb{E}[X_{n-1}] = 2^2\,\mathbb{E}[X_{n-2}] = \cdots = 2^n\,\mathbb{E}[X_0] = 2^n$ dice.

3 The Axe-pectation

Note 21

It is Big Game Week! The Axe travels between Cal and Stanfurd depending on who wins Big Game.

Suppose that if The Axe is currently at Cal, Cal wins with probability 0.8 (i.e. The Axe stays at Cal). On the other hand, if The Axe is currently at Stanfurd, Stanfurd wins with probability 0.4 (i.e. The Axe stays at Stanfurd).

Let X denote the number of games played until The Axe is at Cal.

(a) Suppose The Axe is currently at Stanfurd (S). Write an equation that would solve for $\mathbb{E}[X \mid S]$.

(b) Find $\mathbb{E}[X \mid S]$.

(c) Stanfurd pulls a trick up their sleeve! Now, if The Axe is at Stanfurd, Cal needs to win 2 games in a row before The Axe can travel to Cal. Write an equation that would solve for $\mathbb{E}[X \mid S]$, assuming The Axe just arrived at Stanfurd.

Hint: What happens when Cal wins the first game?

Solution:

(a) The Axe is currently at Stanfurd, so Stanfurd wins with 0.4 probability and Cal wins with 0.6 probability.

At least one game must be played for the Axe to be at Cal. If Stanfurd wins, then we are in the same situation that we started with. If Cal wins, we don't have to play any more games, since The Axe would now be at Cal.

Thus,

$$\begin{aligned}
\mathbb{E}[X \mid S] &= 0.4(1 + \mathbb{E}[X \mid S]) + 0.6(1 + 0) \\
&= 1 + 0.4\,\mathbb{E}[X \mid S]
\end{aligned}$$

(b) Solving the above equation gives $\mathbb{E}[X \mid S] = \frac{1}{0.6}$. Notice that this is the same as the expectation of a Geometric(0.6) random variable.

(c) We want to find $\mathbb{E}[X \mid S]$. The difference between this scenario and (b) is that if Cal wins (with probability 0.6), we still need to play more games until The Axe is back at Cal.

It would be helpful to represent when The Axe is at Stanfurd and Cal is on a 1 game winning streak: let S_1 denote this event.

Then,

$$\begin{aligned}\mathbb{E}[X \mid S] &= 0.4(1+\mathbb{E}[X \mid S]) + 0.6(1+\mathbb{E}[X \mid S_1]) \\ &= 1 + 0.4\,\mathbb{E}[X \mid S] + 0.6\,\mathbb{E}[X \mid S_1]\end{aligned}$$

where $\mathbb{E}[X \mid S_1]$ is the number of games played until The Axe is back at Cal, assuming that The Axe is currently at Stanfurd and Cal won the last game.

Then, in this case, if Cal wins, no more games are needed since The Axe is at Cal. But if Cal loses, then we are in the same situation that we started with. Thus,

$$\begin{aligned}\mathbb{E}[X \mid S_1] &= 0.4(1+\mathbb{E}[X \mid S]) + 0.6(1+0) \\ &= 1 + 0.4\,\mathbb{E}[X \mid S]\end{aligned}$$

So,

$$\mathbb{E}[X \mid S] = 1 + 0.4\,\mathbb{E}[X \mid S] + 0.6(1 + 0.4\,\mathbb{E}[X \mid S])$$

4 Estimating π

Note 22

In this problem, we discuss one way that you could probabilistically estimate π. We'll use a square dartboard of side length 2, and a circular target drawn inscribed in the square dartboard with radius 1. A dart is then thrown uniformly at random in the square. Let p be the probability that the dart lands inside the circle.

(a) What is p?

(b) Suppose we throw N darts uniformly at random in the square. Let $\hat{p}$ be the proportion of darts that land inside the circle. Create an unbiased estimator X for π using $\hat{p}$.

(c) Using Chebyshev's Inequality, compute the minimum value of N such that your estimate is within ε of π with $1-\delta$ confidence. Your answer should be in terms of ε and δ. Note that since we are estimating π, your answer should not have π in it.

Solution:

(a) The total area is 4, and the area of the circle is π. The throw is uniform, so $p = \frac{\pi}{4}$.

(b) We have that $\mathbb{E}[\hat{p}] = p = \frac{\pi}{4}$, so we also have that $\mathbb{E}[4\hat{p}] = \pi$. Thus, $X = 4\hat{p}$ is an unbiased estimator for π.

(c) We have

$$\begin{aligned}\mathbb{P}[|X-\pi| \geq \varepsilon] &= \mathbb{P}\left[\left|\hat{p} - \frac{\pi}{4}\right| \geq \frac{1}{4}\varepsilon\right] \\ &\leq \frac{\mathrm{Var}(\hat{p})}{\left(\frac{1}{4}\varepsilon\right)^2}\end{aligned}$$

by Chebyshev's Inequality and using the fact that $X = 4\hat{p}$. We want our estimate to have confidence $1-\delta$, so we want $\frac{\mathrm{Var}(\hat{p})}{\left(\frac{1}{4}\varepsilon\right)^2} < \delta$. Since $N\hat{p}$ is a $\mathrm{Binomial}(N, p)$ variable, it has variance $Np(1-p)$ and therefore $\hat{p}$ has variance $\frac{Np(1-p)}{N^2} = \frac{p(1-p)}{N}$. Since we are estimating π, we should not assume anything about the value of p in our calculations. Thus, we should use the greatest possible value of the variance, which is $\frac{1}{4}$ (when $p = \frac{1}{2}$). Then

$$\frac{\frac{p(1-p)}{N}}{\left(\frac{1}{4}\varepsilon\right)^2} < \delta \implies N > \frac{16p(1-p)}{\delta\varepsilon^2} = \frac{4}{\delta\varepsilon^2}.$$

5 Deriving the Chernoff Bound

Note 22

We've seen the Markov and Chebyshev inequalities already, but these inequalities tend to be quite loose in most cases. In this question, we'll derive the *Chernoff bound*, which is an *exponential* bound on probabilities.

The Chernoff bound is a natural extension of the Markov and Chebyshev inequalities: in Markov's inequality, we utilize only information about $\mathbb{E}[X]$; in Chebyshev's inequality, we utilize only information about $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ (in the form of the variance). In the Chernoff bound, we'll end up using information about $\mathbb{E}[X^k]$ for *all* k, in the form of the *moment generating function* of X, defined as $\mathbb{E}[e^{tX}]$. (It can be shown that the kth derivative of the moment generating function evaluated at $t=0$ gives $\mathbb{E}[X^k]$.)

In several subparts, we'll ask you to express your answer as a single exponential function, which has the form $e^{f(t)} = \exp(f(t))$ for some function f.

Here, we'll derive the Chernoff bound for the binomial distribution. Suppose $X \sim \mathrm{Binomial}(n, p)$.

(a) We'll start by computing the *moment generating function* of X. That is, what is $\mathbb{E}[e^{tX}]$ for a fixed constant $t > 0$? (Your answer should have no summations.)

Hint: It can be helpful to rewrite X as a sum of Bernoulli RVs.

(b) A useful inequality that we'll use is that

$$1 - \alpha \leq e^{-\alpha},$$

for any α. Since we'll be working a lot with exponentials here, use the above to find an upper bound for your answer in part (a) as a single exponential function. (This will make the expressions a little nicer to work with in later parts.)

Note: Make sure the inequality still holds if you manipulate it (i.e., suppose you square both sides). What must be true about $1-\alpha$?

(c) Use Markov's inequality to give an upper bound for $\mathbb{P}[e^{tX} \geq e^{t(1+\delta)\mu}]$, for $\mu = \mathbb{E}[X] = np$ and a constant $\delta > 0$.

Use this to deduce an upper bound on $\mathbb{P}[X \geq (1+\delta)\mu]$ for any constant $\delta > 0$. (Your bound should be a single exponential function, where f should also depend on $\mu = np$ and δ.)

(d) Notice that so far, we've kept this new parameter t in our bound—the last step is to optimize this bound by choosing a value of t that minimizes our upper bound.

Take the derivative of your expression with respect to t to find the value of t that minimizes the bound. Note that from part (a), we require that $t > 0$; make sure you verify that this is the case!

Use your value of t to verify the following Chernoff bound on the binomial distribution:

$$\mathbb{P}[X \geq (1+\delta)\mu] \leq \exp(-\mu(1+\delta)\ln(1+\delta) + \delta\mu).$$

Note: As an aside, if we carried out the computations without using the bound in part (b), we'd get a better Chernoff bound, but the math is a lot uglier. Furthermore, instead of looking at the binomial distribution (i.e. the sum of independent and identical Bernoulli trials), we could have also looked at the sum of independent but not necessarily identical *Bernoulli trials as well; this would give a more general but very similar Chernoff bound.*

(e) Let's now look at how the Chernoff bound compares to the Markov and Chebyshev inequalities. Let $X \sim \text{Binomial}(n = 100, p = \frac{1}{5})$. We'd like to find $\mathbb{P}[X \geq 30]$.

  (i) Use Markov's inequality to find an upper bound on $\mathbb{P}[X \geq 30]$.

  (ii) Use Chebyshev's inequality to find an upper bound on $\mathbb{P}[X \geq 30]$.

  (iii) Use the Chernoff bound from part (d) to find an upper bound on $\mathbb{P}[X \geq 30]$.

  (iv) Now use a calculator to find the exact value of $\mathbb{P}[X \geq 30]$. How did the three bounds compare? That is, which bound was the closest and which bound was the furthest from the exact value?

Solution:

(a) Note that we can write $X = \sum_{i=1}^{n} X_i$, where each X_i is an independent and identical Bernoulli trial with probability p. This means that we have

$$\begin{aligned}
\mathbb{E}[e^{tX}] &= \mathbb{E}\left[e^{t\sum_{i=1}^n X_i}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^n e^{tX_i}\right] \\
&= \prod_{i=1}^n \mathbb{E}[e^{tX_i}] && \text{(independence)} \\
&= \prod_{i=1}^n \left(e^t \cdot \mathbb{P}[X_i = 1] + e^0 \cdot \mathbb{P}[X_i = 0]\right) && \text{(LOTUS)} \\
&= \prod_{i=1}^n \left(pe^t + 1 - p\right) \\
&= (pe^t + 1 - p)^n
\end{aligned}$$

Alternate Solution: We can also evaluate the expectation directly; using LOTUS, we have

$$\begin{aligned}
\mathbb{E}[e^{tX}] &= \sum_{k=0}^n e^{tk} \cdot \mathbb{P}[X = k] \\
&= \sum_{k=0}^n e^{tk} \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\
&= (pe^t + 1 - p)^n
\end{aligned}$$

In the last step, we used the binomial theorem in reverse:

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k},$$

for $a = pe^t$ and $b = 1 - p$.

(b) With $\alpha = p - pe^t = p(1 - e^t)$, we have

$$(pe^t + 1 - p)^n = (1 - p(1 - e^t))^n \le \exp(-np(1 - e^t)) = \exp(-\mu(1 - e^t)).$$

Notice that $(1 - p(1 - e^t))^n \le \exp(-np(1 - e^t))$ is valid, even when n is even. This is because $1 - \alpha = 1 - p(1 - e^t)$ is non-negative. Since $e^t > 0$, $1 - e^t < 1$, so $\alpha = p(1 - e^t) < p$, and $1 - \alpha > 1 - p$. Since $p \le 1$ (Binomial parameter), $1 - p \ge 0$, and so is $1 - \alpha$.

(c) By Markov's inequality on the RV e^{tX} (which is always nonnegative), we have

$$\begin{aligned}
\mathbb{P}[e^{tX} \ge e^{t(1+\delta)\mu}] &\le \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu)}} \\
&\le e^{-t(1+\delta)\mu)} e^{-\mu(1-e^t)} \\
&= \exp(-t(1+\delta)\mu - \mu(1 - e^t))
\end{aligned}$$

where the second inequality comes from plugging in our answer from part (b).

As such, we have

$$\mathbb{P}[X \geq (1+\delta)\mu] = \mathbb{P}[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \exp(-t(1+\delta)\mu - \mu(1-e^t)).$$

(d) Taking the derivative of the exponential, we have

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left[\exp(-t(1+\delta)\mu - \mu(1-e^t)\right] \\ &= \left[\exp(-t(1+\delta)\mu - \mu(1-e^t)\right] \cdot \left(-(1+\delta)\mu + \mu e^t\right)\end{aligned}$$

This quantity is equal to zero when the last term is equal to zero (we can ignore the exponential, since it'll never be equal to 0). As such,

$$\begin{aligned}-(1+\delta)\mu + \mu e^t &= 0 \\ \mu e^t &= (1+\delta)\mu \\ e^t &= 1+\delta \\ t &= \ln(1+\delta)\end{aligned}$$

Since $\delta > 0$, we have that $t > 0$ as well, which satisfies our conditions on t.

Plugging this back in to our bound in part (c), we have

$$\begin{aligned}\mathbb{P}[X \geq (1+\delta)\mu] &\leq \exp(-t(1+\delta)\mu - \mu(1-e^t)) \\ &= \exp(-\mu(1+\delta)\ln(1+\delta) - \mu(1-(1+\delta))) \\ &= \exp(-\mu(1+\delta)\ln(1+\delta) + \delta\mu)\end{aligned}$$

as desired.

(e) Firstly, we'll compute a few statistics of X, which will be useful in these subparts:

$$\mathbb{E}[X] = np = 100 \cdot \frac{1}{5} = 20$$

$$\mathrm{Var}(X) = np(1-p) = 100 \cdot \frac{1}{5} \cdot \frac{4}{5} = 16$$

(i) Using Markov's inequality, we have

$$\mathbb{P}[X \geq 30] \leq \frac{\mathbb{E}[X]}{30} = \frac{20}{30} = \frac{2}{3} \approx 0.6666.$$

(ii) Using Chebyshev's inequality, we have

$$\begin{aligned}\mathbb{P}[X \geq 30] &= \mathbb{P}[X - 20 \geq 10] \\ &= \mathbb{P}[X - \mathbb{E}[X] \geq 10] \\ &\leq \mathbb{P}[|X - \mathbb{E}[X]| \geq 10] \\ &\leq \frac{\mathrm{Var}(X)}{10^2} \\ &= \frac{16}{100} = 0.16\end{aligned}$$

(iii) Using the Chernoff bound, we have

$$\begin{aligned}
\mathbb{P}[X \geq 30] &= \mathbb{P}\left[X \geq \left(1+\frac{1}{2}\right)\cdot 20\right] \\
&\leq \exp\left(-\mu\left(1+\frac{1}{2}\right)\ln\left(1+\frac{1}{2}\right)+\frac{1}{2}\mu\right) \qquad \text{(Chernoff with } \delta = \tfrac{1}{2}) \\
&= \exp(-30\cdot\ln(1.5)+10) \\
&\approx 0.1148
\end{aligned}$$

(iv) The exact value is

$$\mathbb{P}[X \geq 30] = \sum_{k=30}^{100} \mathbb{P}[X = k] = \sum_{k=30}^{100} \binom{100}{k}\left(\frac{1}{5}\right)^k\left(\frac{4}{5}\right)^{100-k} \approx 0.01124.$$

The Chernoff bound is the closest, followed by Chebyshev's inequality, and Markov's inequality is the furthest.

As an aside, this should be expected—the Markov bound utilizes the least amount of information, while the Chernoff bound utilizes the most. In particular, Markov's inequality only requires the expectation $\mathbb{E}[X]$, Chebyshev's requires the variance (which includes information about both $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$), and the Chernoff bound requires the moment generating function (which contains information about all *moments* of X, i.e. all $\mathbb{E}[X^k]$ for $k \geq 1$).