14 2025-02-12 | Week 5 | Lecture 13

Please read 3.4, 3.6. Skip 3.5.

In the homework, we say the following problem:

Problem 42. Two cards are randomly drawn from a deck of cards. Let A be the event that at least one ace is drawn. Let A_s be the event that the ace of spades is chosen. And let B be the event that both cards are aces. Compute the following conditional probabilities:

(a) $\mathbb{P}[B \mid A_s]$

(b) $\mathbb{P}[B \mid A]$

We computed that $\mathbb{P}[B \mid A_s] = \frac{1}{17} \approx .06$ and that $\mathbb{P}[B \mid A] = \frac{1}{33} \approx .03$. The first of these is twice as big as the second. Why on earth does knowing A_s increase the probability of B, compared to knowing A?!

14.1 Variance

Recall, the *variance* of a random variable X is

$$\mathrm{Var}(X) := \sum_x (x-\mu)^2 p(x)$$

where $\mu = \mathbb{E}[X]$ and x ranges over all possible values that X can take. The variance of a random variable is often denoted σ^2.

Important formula: There is a shortcut for computing variance

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 \tag{9}$$

Also, we have the following useful properties. Let a, b be fixed, nonrandom numbers. Then

- $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
- $\mathrm{Var}(X + b) = \mathrm{Var}(X)$

Moreover, if X, Y are independent random variables, then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

(Of course, we always have $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, even if X and Y are not independent. This is called *linearity of expectation*. But for variance, we need X and Y to be independent to do something similar.)

Example 43. Suppose

$$U = \begin{cases} 2 & : \text{ with probability } \alpha \\ 3 & : \text{ with probability } 1-\alpha \end{cases}$$

What is $\mathbb{E}[U]$ and $\mathrm{Var}(U)$?

Solution: First, we will compute the expectation:

$$\begin{aligned} \mathbb{E}[U] &= 2\alpha + 3(1-\alpha) \\ &= 3 - \alpha \end{aligned} \tag{10}$$

Next, we will compute $\mathrm{Var}(U)$. We will use the formula Eq. (9). In particular we wil need $\mathbb{E}\left[U^2\right]$, so let's compute that:

$$\begin{aligned} \mathbb{E}\left[U^2\right] &= 2^2\alpha + 3^2(1-\alpha) && \text{by Theorem 40} \\ &= 4\alpha + 9(1-\alpha) \\ &= 9 - 5\alpha \end{aligned} \tag{11}$$

We can now plug the values from Eqs. (10) and (11) into Eq. (9) to get

$$\begin{aligned}\mathrm{Var}(U) &= \mathbb{E}\left[U^2\right] - (\mathbb{E}\left[U\right])^2 \\ &= 9 - 5\alpha - (3-\alpha)^2\end{aligned}$$

For example, if $\alpha = \frac{1}{2}$, then we get $\mathbb{E}\left[U\right] = 3 - \frac{1}{2} = 2.5$ and

$$\mathrm{Var}(U) = 9 - \frac{5}{2} - \frac{25}{4} = \frac{1}{4}$$

Now, someone asked "how do we interpret this 1/4?" The short answer is, it's complicated. There isn't a nice interpretation that works for every random variable, so often the best we can do is understand this as some measure of how much our random variable tends to deviate from its mean. In this case, the 1/4 is pretty small, so the random variable doesn't seem to deviate all that much from its mean.

End of Example 43. □

While variance does give us some information, it's only a single number and so the amount of information it gives us about the random variable is actually very limited, and it is impossible to interpret it precisely without additional information about the random variable. This is illustrated in the following example.

Example 44. Let $M = 1{,}000{,}000$. And let

$$X = \begin{cases} 0 & : \text{ with probability } .99 \\ 10M & : \text{ with probability } .01 \end{cases}$$

The random variable X is like winning the lottery.

On the other hand, let

$$Y = \begin{cases} -10 & : \text{ with probability } \frac{1}{2} \\ 10 & : \text{ with probability } \frac{1}{2} \end{cases}$$

In the case of X, we have

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}\left[X^2\right] - (\mathbb{E}\left[X\right])^2 \\ &= \left[(10M)^2(0.01) + 0^2(0.99)\right] - \left[10M(0.01) + 0(0.99)\right]^2 \\ &= M^2 - \frac{M^2}{100} \\ &= \frac{99}{100}M^2\end{aligned}$$

This is a HUGE variance, because $M = 10{,}000{,}000$. But for the most part, intuitvely, few people would say that this r.v. "varies" all that much. After all, it's almost always 0.

On the other hand, what if we compute the variance of Y? In this, we have $\mathbb{E}\left[Y\right] = -10\frac{1}{2} + 10\frac{1}{2} = 0$, so

$$\begin{aligned}\mathrm{Var}(Y) &= \mathbb{E}\left[Y^2\right] - 0 \\ &= (-10)^2\frac{1}{2} + (10)^2\frac{1}{2} \\ &= 100\end{aligned}$$

In this case, the variance is pretty big—not huge, but pretty big. Yet the random variable Y doesn't really seem to "vary" all that much: after all, we know it is always either 10 more than the mean (which is zero), or 10 less than the mean.

End of Example 44. □

Hopefully this example doesn't give you the impression that variance is useless. It's actually a really important measure. The issue is just that random variables can "vary" in lots of different ways that can be hard to summarize as a single value.

Definition 45 (Binomial Distribution)**.** Flip a coin n times, and let $\alpha \in (0,1)$ be the probability of 'heads' on any given coin flip. The probability that you get exactly k heads is

$$\binom{n}{k}\alpha^k(1-\alpha)^{n-k}$$

A random variable with state space $\{0,1,2,\ldots,n\}$ is said to have *Binomial distribution with n trials and success probability α* if its pmf is

$$p(k) = \binom{n}{k}\alpha^k(1-\alpha)^{n-k}$$

for $k = 0,1,\ldots,n$. We write $X \sim \mathrm{Bin}(n,\alpha)$.

You've already seen binomial random variables before. For an example, the random variable X from Example 35 is an example of a random variable with $X \sim \mathrm{Bin}(3, \frac{1}{2})$.

14.2 Poisson convergence

Fix $\lambda \in (0,1)$. Suppose have a sequence of magic coins $1,2,\ldots$.

- Coin 1 has probability of heads λ
- Coin 2 has probability of heads $\frac{\lambda}{2}$
- Coin 3 has probability of heads $\frac{\lambda}{3}$
- And so forth.

Consider the following sequence of experiments:

- Experiment 1: Flip coin 1 once.
- Experiment 2: Flip coin 2 twice.
- Experiment 3: Flip coin 3 three times.
- Experiment n: Flip coin n, n times.

Now let us consider the nth experiment. Let $F_1,\ldots,F_n$ represent the n coin flips, with

$$F_1 = \begin{cases} 1 & : \text{ first coin flip is heads} \\ 0 & : \text{ first coin flip is tails} \end{cases}$$

$$F_2 = \begin{cases} 1 & : \text{ second coin flip is heads} \\ 0 & : \text{ second coin flip is tails} \end{cases}$$

and so forth for all n coin flips. So

$$X_n = F_1 + F_2 + \ldots + F_n$$

is the number of heads that you get.

Question: What is the expectation of X_n? We can use the linearity of expectation:

$$\begin{aligned} \mathbb{E}[X_n] &= \mathbb{E}[F_1 + F_2 + \ldots + F_n] \\ &= \mathbb{E}[F_1] + \mathbb{E}[F_2] + \ldots + \mathbb{E}[F_n] \\ &= \underbrace{\frac{\lambda}{n} + \frac{\lambda}{n} + \ldots + \frac{\lambda}{n}}_{n \text{ terms}} \\ &= \lambda. \end{aligned}$$

So the expected number of heads that we get doesn't change from experiment to experiment, even if we send $n \to \infty$. Think!—for each experiment, the probability of heads is $\frac{\lambda}{n}$ and that is getting smaller. However the total number of coin flips is getting larger and larger.

Question Is it possible that as $n \to \infty$, the random variable X_n converges to some new random variable, one that we haven't seen yet?

Yes.