Photo snapped at the SLC by a Math 10A student

Variance, Independence

Math 10A

November 9, 2017

Today is the last day to request a regrade on one of your MT#2 problems.

Faculty Club events

Two new events on Friday, December 1 (last day of classes):

- 8AM breakfast—send email to sign up;
- pop-in lunch at high noon (just show up).

Chebyshev's inequality

The inequality in question concerns an arbitrary random variable X. Say that the mean of X is μ and that the standard deviation of X is σ. For integers $k \geq 1$, the inequality states:

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

In other words: the probability of being k or more standard deviations away from the mean is at most $\frac{1}{k^2}$. For example, the probability of being two or more standard deviations away from the mean is at most $1/4$.

Why is Chebyshev's inequality true?

The explanation is provided on page 553 of the book and also (of course!) in Wikipedia. The following slides summarize the argument.

Not on the exam

For simplicity, we'll assume that the expected value of X is 0; this just means shifting the line $x = \mu$ over to the y-axis. Then

$$\begin{aligned}
\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx &\geq \int_{-\infty}^{-k\sigma} x^2 f(x)\,dx + \int_{k\sigma}^{\infty} x^2 f(x)\,dx \\
&\geq \int_{-\infty}^{-k\sigma} (k\sigma)^2 f(x)\,dx + \int_{k\sigma}^{\infty} (k\sigma)^2 f(x)\,dx \\
&= k^2\sigma^2 \left(\int_{-\infty}^{-k\sigma} f(x)\,dx + \int_{k\sigma}^{\infty} f(x)\,dx \right).
\end{aligned}$$

Divide by $k^2\sigma^2$ to get

$$\frac{1}{k^2} \geq \left(\int_{-\infty}^{-k\sigma} f(x)\,dx + \int_{k\sigma}^{\infty} f(x)\,dx \right).$$

Not on the exam

The same inequality read differently:

$$\left(\int_{-\infty}^{-k\sigma} f(x)\,dx + \int_{k\sigma}^{\infty} f(x)\,dx \right) \leq \frac{1}{k^2}.$$

The left-hand sum represents the probability that X is to the right of $k\sigma$ plus the probability that X is to the left of $-k\sigma$. In other words, the left-hand term is the probability that X is k or more standard deviations from its mean.

Summary:

$$P(X \text{ is } k \text{ or more standard deviations from its mean}) \leq \frac{1}{k^2}.$$

$$P(X \text{ is within } k \text{ standard deviations of its mean}) \geq 1 - \frac{1}{k^2}.$$

Expected value of a product

As we saw on Thursday, the expected value of a product of two random variables is not necessarily the product of the expected values:

$$E[XY] \neq E[X]E[Y] \text{ usually.}$$

In face, typically

$$E[X^2] \neq E[X]^2; \text{ i.e., } \mathrm{Var}[X] \neq 0.$$

However:

If X and Y are *independent* random variables, then

$$E[XY] = E[X]E[Y].$$

Some questions

- What is the meaning of *independence*?
- What are some examples of independence?
- Why is $E[XY] = E[X]E[Y]$ if X and Y are independent?

Heuristically, “independence” means “having nothing to do with each other.” Two variables X and Y are independent if the value of X does not affect the value of Y.

This discussion could easily lead us to introduce the notion of conditional probability. We’ll have a very minimal discussion now but stay tuned for more in 10B.

Getting a 3 when you roll a die with your left hand is independent of getting a head when you flip a coin with your right hand. (Don't try that at home.) The two have nothing to do with each other.

When you flip a coin, you get heads half the time. When you roll a die, you get 3 one-sixth of the time. You get the two together–a 3 and a head—one-twelfth of the time:

$$\frac{1}{12} = \frac{1}{6} \cdot \frac{1}{2}.$$

In Math 10B, two random variables X and Y are independent if

$$P(X = a \text{ and } Y = b) = P(X = a) \times P(Y = b)$$

for all numbers a and b.

We understand that X and Y are functions on the same probability space Ω. Recall that each point $\omega \in \Omega$ has an associated probability $p(\omega)$. By definition,

$$P(X = a) = \sum_{\omega : X(\omega) = a} p(\omega).$$

The other probabilities $P(Y = b)$ and $P(X = a \text{ and } Y = b)$ are defined similarly.

Digression: I'm writting $p(\omega)$ for the probability of a point and will write $p(S)$ for the probability associated to a set.

I'm writing $P(X = a)$ and so forth for the probability that a random variable takes a specific value. Why is it sometimes "P" and sometimes "p"? Probably I should be using the same letter for everything.

Here's one way to re-phrase independence of X and Y. Let

$$A = \{\, \omega \in \Omega \,|\, X(\omega) = a \,\},$$
$$B = \{\, \omega \in \Omega \,|\, X(\omega) = b \,\}.$$

Then

$$A \cap B = \{\, \omega \in \Omega \,|\, X(\omega) = a \text{ and } X(\omega) = b \,\}.$$

For C a subset of Ω, it is natural to write $p(C) = \sum_{\omega \in C} p(\omega)$. The condition

$$P(X = a \text{ and } Y = b) = P(X = a) \cdot P(Y = b)$$

can then be written

$$p(A \cap B) = p(A)p(B).$$

Two subsets ("events") A and B of Ω are called *independent* if

$$p(A \cap B) = p(A)p(B).$$

So the random variables X and Y are independent if the sets $\{\, \omega \in \Omega \,|\, X(\omega) = a \,\}$ and $\{\, \omega \in \Omega \,|\, X(\omega) = b \,\}$ are independent for all numbers a and b.

Simple example

Let Ω be the set of possible outcomes of flipping a (possibly biased) coin twice. Then Ω has four elements. Let $X : \Omega \to \{ 0, 1 \}$ be 0 if the first flip is "tails" and 1 if it's "heads." Define Y analogously, using the second flip. Then X and Y are independent; heuristically, this means that the first flip of the coin has no link to the second flip.

To check that your understanding, verify (on your own, after the class) that X and Y are independent according to the definitions.

Weird example

Let Ω be the set of possible outcomes when you flip a fair coin three times:

$$\Omega = \{\, \texttt{000}, \texttt{001}, \texttt{010}, \texttt{011}, \texttt{100}, \texttt{101}, \texttt{110}, \texttt{111} \,\},$$

with each point having probability $\frac{1}{8}$. Let

$$A = \{\, \texttt{001}, \texttt{010}, \texttt{011}, \texttt{100}, \texttt{101}, \texttt{110} \,\},$$
$$B = \{\, \texttt{000}, \texttt{001}, \texttt{010}, \texttt{100} \,\}$$

Thus *A* consists of all outcomes with 1 or 2 heads while *B* consists of outcomes with 0 or 1 head.

Are the “events” A and B are independent?

Independence amounts to the equality

$$p(A \cap B) \stackrel{?}{=} p(A)p(B) = \frac{6}{8} \cdot \frac{4}{8} = \frac{3}{8},$$

which translates to the statement that $A \cap B$ has three elements.

The statement is true because $A \cap B$ is the of outcomes with exactly one head, and there are three such outcomes.

In words, “having at most one head” and “not having all heads or all tails” are independent conditions! This sounds really suspicious to me because both events reference numbers of heads. The events don’t seem independent—but they are.

Are the “events” A and B are independent?

Independence amounts to the equality

$$p(A \cap B) \stackrel{?}{=} p(A)p(B) = \frac{6}{8} \cdot \frac{4}{8} = \frac{3}{8},$$

which translates to the statement that $A \cap B$ has three elements.

The statement is true because $A \cap B$ is the of outcomes with exactly one head, and there are three such outcomes.

In words, “having at most one head” and “not having all heads or all tails” are independent conditions! This sounds really suspicious to me because both events reference numbers of heads. The events don’t seem independent—but they are.

Are the “events” A and B are independent?

Independence amounts to the equality

$$p(A \cap B) \stackrel{?}{=} p(A)p(B) = \frac{6}{8} \cdot \frac{4}{8} = \frac{3}{8},$$

which translates to the statement that $A \cap B$ has three elements.

The statement is true because $A \cap B$ is the of outcomes with exactly one head, and there are three such outcomes.

In words, “having at most one head” and “not having all heads or all tails” are independent conditions! This sounds really suspicious to me because both events reference numbers of heads. The events don’t seem independent—but they are.

In Math 10A, probabilities like $P(X = a)$ are typically 0, so we define independence to mean

$$P(X \leq a \text{ and } Y \leq b) = P(X \leq a) \cdot P(Y \leq b)$$

for all numbers a and b.

Why independence implies $E[XY] = E[X]E[Y]$

To see the idea of the proof, we consider the super simple Math 10B example where X and Y each take on the two values 1 and 2. Then XY has the values 1, 2 and 4. We have

$$
\begin{aligned}
E[X] &= 1 \cdot P(X=1) + 2 \cdot P(X=2) \\
E[Y] &= 1 \cdot P(Y=1) + 2 \cdot P(Y=2) \\
E[X]E[Y] &= 1 \cdot P(X=1)P(Y=1) \\
&\quad + 2\Big((P(X=1)P(Y=2) + P(X=2)P(Y=1)\Big) \\
&\quad + 4 \cdot P(X=2)P(Y=2) \\
E[XY] &= 1 \cdot P(XY=1) + 2 \cdot P(XY=2) + 4 \cdot P(XY=4).
\end{aligned}
$$

What remains to be seen is that the three pieces of $E[X]E[Y]$ can be equated with the three pieces of $[EY]$.

For example, why is it true that

$$P(XY = 2) \stackrel{?}{=} P(X = 1)P(Y = 2) + P(X = 2)P(Y = 1)?$$

There are two things to know. First, $XY = 2$ means either $X = 1$ and $Y = 2$, or else $X = 2$, $Y = 1$. Accordingly,

$$P(XY = 2) \;=\; P(X = 1 \text{ and } Y = 2) + P(X = 2 \text{ and } Y = 1).$$

Second, by independence

$$P(X = 1 \text{ and } Y = 2) = P(X = 1) \cdot P(Y = 2),$$

and similarly for the second term of the sum. This gives

$$P(XY = 2) \;=\; P(X = 1) \cdot P(Y = 2) + P(X = 2) \cdot P(Y = 1),$$

which is the equation at the top of this slide.