Probability

Kenneth A. Ribet

Math 10A
November 15, 2016

Friday's Blue Bottle Coffee visit

Yesterday's breakfast

Announcements

No Ribet office hour on Thursday. I'll be away, and James will take over the lecture.

Faculty Club "pop-in" lunches on Wednesday, November 16 and on Monday, November 21, both at high noon.

Breakfast Thursday, December 1 at 9AM (new event).

Breakfast Monday, December 5 at 9AM (new event).

Independence

In Math 10B, there is the notion of *independent events* in a probability space: if A and B are subsets of Ω, then A and B are *independent* if

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

The equation is automatically true when A or B has probability 0 because both sides of the equation are 0. So we can and probably should assume $P(A), P(B) > 0$.

Then (by definition) the condition of independence can be interpreted as the equality

$$P(A|B) = P(A),$$

where the left-hand side is the probability of "A given B." If A has positive probability, we can symmetrically rephrase the condition as

$$P(B|A) = P(B).$$

Example (10B)

We consider the (eight-element) space of outcomes of three tosses of a fair coin; one such outcome is `HTH`. Consider these two events:

> A: the outcome is mixed (not `TTT` or `HHH`);
> B: there is at most one `T` in the string.

The first event consists of six of the eight outcomes, so $P(A) = 3/4$. Similarly, there is one outcome with three $\mathtt{T}$'s and three with two $\mathtt{T}$'s; thus B has four elements and $P(B) = 1/2$.

The two are independent if and only if $p(A \cap B) = \dfrac{3}{4} \cdot \dfrac{1}{2} = \dfrac{3}{8}$, i.e., if and only if $A \cap B$ has three elements.

This is true because $A \cap B$ is the set of outcomes with exactly one $\mathtt{T}$, so $A \cap B$ has three elements.

I'll miss you guys too. (I won't be teaching 10B next semester.)

There is a more general definition for a finite set of events A_1, $A_2, \ldots, A_t$: they are *independent* if *even after reordering the events* we have

$$P(A_1 \cap A_2) = P(A_1)P(A_2),$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3),$$
$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2)P(A_3)P(A_4),$$

and so on.

Random Variables

Two random variables X_1 and X_2 are independent if the events $X_1 \le x_1$ and $X_2 \le x_2$ are independent for all numbers x_1 and x_2. This means:

$$P(X_1 \le x_1 \text{ and } X_2 \le x_2) = P(X_1 \le x_1)P(X_2 \le x_2)$$

for all x_1 and x_2. There's a similar definition for the independence of n random variables.

The probability on the left side of the equation is called $F_{X_1,X_2}(x_1, x_2)$ or simply $F(x_1, x_2)$; it's called the joint cumulative distribution function for X_1 and X_2. Independence then means that

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$$

for all x_1 and x_2.

People understand independence to mean "not having anything to do with each other." For example, our probability space could be the set of all possible outcomes of rolling a single die 42 times; these are strings $a_1 a_2 \cdots a_{42}$ in which each "coordinate" a_i is between 1 and 6.

For each i, let X_i be the value of the ith roll, i.e.,

$$X_i(a_1 a_2 \cdots a_{42}) = a_i.$$

It is "obvious" (at least intuitively) that the set of random variables $\{X_i\}$ is independent because the different rolls don't depend on each other.

When I taught 10B last semester, I wondered how hard it would be to verify the independence of the X_i with an honest check of the definition. I'm still wondering.

Expected values

If c is a constant (i.e., a number) and X is a random variable, then

$$E[cX] = cE[X].$$

If $X_1, \ldots, X_n$ are random variables, then

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots E[X_n].$$

Both statements are pretty obvious in the case when we're dealing with finite sums and then become true "in the limit" when we're dealing with continuous random variables.

If X is a continuous random variable with $f(x)$ as its PDF, what can you say about cX? If $c = 0$, $cX = 0$ is discrete. If $c \neq 0$, then cX is continuous, and its PDF is equal to what?

We did the computation in class and thought that the answer might be $\frac{1}{c} f(\frac{x}{c})$. What do you think?

If X and Y are random variables, is it true that

$$E[XY] \stackrel{?}{=} E[X]E[Y]?$$

Not always. For a simple example, flip a coin one time and let $X = 1$ if we get $\mathtt{H}$ and $X = 0$ if we get $\mathtt{T}$. Let $X = Y$. . . . The LHS is $\frac{1}{2}$, while the RSH is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

If X and Y are random variables, is it true that

$$E[XY] \stackrel{?}{=} E[X]E[Y]?$$

Not always. For a simple example, flip a coin one time and let $X = 1$ if we get $\mathtt{H}$ and $X = 0$ if we get $\mathtt{T}$. Let $X = Y \ldots$. The LHS is $\frac{1}{2}$, while the RSH is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

A theorem

If X and Y are independent, then

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

For discrete random variables, the proof comes about by manipulating sums in a straightforward way. You can see the computation, for example, on page 4 of computer science course notes that I found by googling. (The second author received her PhD from UC Berkeley. Go Bears!)

Variance

If X and Y are independent then $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$. To verify this, we can replace X by $X - E[X]$ and Y by $Y - E[Y]$. This changes nobody's variance and makes $E[X]$, $E[Y]$ and $E[X+Y]$ all equal to 0. Then

$$\begin{aligned} \mathrm{Var}[X+Y] &= E[(X+Y)^2] = E[X^2] + 2E[XY] + E[Y^2] \\ &= E[X^2] + 2E[X]E[Y] + E[Y^2] \\ &= E[X^2] + 0 + E[Y^2] = \mathrm{Var}[X] + \mathrm{Var}[Y]. \end{aligned}$$

We used the independence when we equated $E[XY]$ and $E[X]E[Y]$.

Identically distributed random variables

We say that a bunch of random variables X_i are *identically distributed* if they have the same CDFs (equivalently: if they have the same PDFs). This means that $P(a \leq X_i \leq b)$ is the same for all the different i.

We are especially interested in the situation where all the random variables are (1) independent and (2) identically distributed. If you type "iid" into google, the suggested completion is iid random variables.

Now suppose that $X_1, \ldots, X_n$ are i.i.d. random variables. Then

$$\mathrm{Var}[X_1 + \ldots + X_n] = n\sigma^2,$$

where σ is the common std. deviation of all the X_i. Therefore

$$\mathrm{Var}\left[\frac{X_1 + \ldots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

because of the general formula $\mathrm{Var}[cX] = c^2 \mathrm{Var}[X]$.

Set $\overline{X} = \dfrac{X_1 + \ldots + X_n}{n}$; it's the average of the X_i. Then $\overline{X}$ has variance $\dfrac{\sigma^2}{n}$ and has standard deviation equal to $\dfrac{\sigma}{\sqrt{n}}$.

The interpretation of this formula is supplied by `Prob-Stat.pdf`:

> *...for large n, the average $\overline{X}$ is much "less random" than each individual random variable $X_1, X_2, \ldots, X_n$.*

Law of Large Numbers

See Wikipedia for a perfectly lucid discussion with some examples.

For the rest of this discussion, we imagine a sequence of i.i.d.'s

$$X_1, X_2, \ldots, X_{47}, \ldots$$

and write

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

Thus $\overline{X}_n$ is the $\overline{X}$ of the previous slide, but now we're considering an infinite sequence of random variables and take the average of the first n X_is for each n. We are thinking $n \to \infty$ and have the idea that the $\overline{X}_n$ become so little random that they converge to the obvious constant random variable.

The obvious constant is $\mu =$ the common expected value of all of the X_i.

The Law of Large Numbers states that indeed:

$$\overline{X}_n \longrightarrow \mu$$

as n approaches ∞.

The only question is what that means.

The Wikipedia page talks of the "weak" and "strong" laws and a discussion of the difference between the two.

Here's the strong law:

$$P(\lim_{n\to\infty} \overline{X}_n = \mu) = 1.$$

This is what (I think) it means:

For each point ω of the probability space Ω, we can consider the sequence of real numbers $\overline{X}_n(\omega)$. Does that sequence converge to μ as $n \to \infty$? Maybe yes, maybe no.

Let $A \subseteq \Omega$ be the set of ω for which the answer is "no." Then A is small in the sense that $P(A) = 0$.

Central Limit Theorem

We continue with a sequence X_i as in previous slides and introduce the averages $\overline{X}_n$ as on previous slides. These random variables still have average $\mu =$ the common expected value of the X_i; the standard deviation of $\overline{X}_n$ is $\dfrac{\sigma}{\sqrt{n}}$, as we saw some minutes ago. We introduce for each n:

$$\frac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma}.$$

These random variables have been rigged so as to have mean 0 and standard deviation 1…

…just like the standard normal variable, which has PDF equal to $\dfrac{1}{\sqrt{2}\pi} e^{-x^2/2}$.

Central Limit Theorem

We continue with a sequence X_i as in previous slides and introduce the averages $\overline{X}_n$ as on previous slides. These random variables still have average $\mu =$ the common expected value of the X_i; the standard deviation of $\overline{X}_n$ is $\frac{\sigma}{\sqrt{n}}$, as we saw some minutes ago. We introduce for each n:

$$\frac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma}.$$

These random variables have been rigged so as to have mean 0 and standard deviation 1…

…just like the standard normal variable, which has PDF equal to $\frac{1}{\sqrt{2}\pi}e^{-x^2/2}$.

The theorem

For real numbers a and b with $a \leq b$:

$$P\left(a \leq \frac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma} \leq b\right) \longrightarrow \frac{1}{\sqrt{2}\pi}\int_a^b e^{-x^2/2}\,dx$$

as $n \to \infty$.

The theorem is often paraphrased by the statement that the variables $\dfrac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma}$ are becoming more and more like a standard normal variable.