Probability concepts

Math 10A

October 33, 2017

Serge Lang lecture

This year's Serge Lang Undergraduate Lecture will be given by Keith Devlin of Stanford University. The title is

> *When the precision of mathematics meets the messiness of the world of people.*

The lecture will be given at 4:10PM *today* in 60 Evans.

Please visit the Facebook event page for the lecture as well.

Dinners

Foothill DC dinner, Friday (Nov. 3) at 6:30PM.

Clark Kerr DC, Sunday (Nov. 5) at 6PM.

Yuge crowds at both dinners, OK? We want to have the biggest crowds ever.

So come.

Recall the fundamental concepts

- Probability space
- Random variable
- Probability density function (PDF)
- Cumulative distribution function (CDF)

Computing averages

Take a (biased) coin that comes up heads (`1`) 3/4 of the time and tails (`0`) 1/4 of the time. Flip the coin three times:

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

$$P(\mathtt{000}) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}, P(\mathtt{001}) = \frac{1}{4} \times \frac{1}{4} \times \frac{3}{4}, \text{ etc.}$$

We do *not* have a uniform probability space!

Random variable

As on Tuesday, let

$$X : \Omega \to \{ 0, 1, 2, 3 \}$$

be the random variable that takes each string in Ω to the number of `1`s in the string:

$$X(\texttt{000}) = 0, X(\texttt{001}) = 1, X(\texttt{111}) = 3, \text{ etc.}$$

Mean

What is the average value of X?

The average value is called the *mean* BTW.

In order to compute the mean, we have to say what we mean by the mean.

Mean

What is the average value of X?

The average value is called the *mean* BTW.

In order to compute the mean, we have to say what we mean by the mean.

Mean

What is the average value of X?

The average value is called the *mean* BTW.

In order to compute the mean, we have to say what we mean by the mean.

Mean's meaning, first try

We sum over all strings and weight each string by its probability. For each string we compute X(the string). We add the result:

$$\mu = X(\texttt{000})P(\texttt{000}) + X(\texttt{001})P(\texttt{001}) + \cdots$$
$$= 0 \cdot \frac{1}{4^3} + 1 \cdot \frac{3}{4^3} + \cdots .$$

There are eight terms in the sum.

$$\mu = X(\texttt{000})P(\texttt{000}) + X(\texttt{001})P(\texttt{001}) + \cdots$$
$$= 0 \cdot \frac{1}{4^3} + 1 \cdot \frac{3}{4^3} + \cdots .$$

There are eight terms in the sum, one where $X = 0$, three where $X = 1$, three where $X = 2$, one where $X = 3$.

The strings for which $X = 1$ (for example) are: `001`, `010` and `100`. Their contribution to μ is

$$1 \cdot P(\texttt{001}) + 1 \cdot P(\texttt{010}) + 1 \cdot P(\texttt{100}),$$

which we rewrite as

$$1 \cdot P(X = 1).$$

Mean's meaning, second try

$$\mu = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3)$$
$$= \sum_i i \cdot P(X = i)$$
$$= \sum_i i \cdot f(i).$$

In the last two equations, i ranges over the possible values of X, namely 0, 1, 2, 3 and 4. Also, f has the same meaning as in yesterday's discussion of coin tosses: it's the Math 10B analogue of the probability density function. Namely,

$$f(i) = P(X = i).$$

Mean's meaning

In the Math 10B world, if a random variable is distributed according to f, then

$$\text{mean of } X = \sum_x x \cdot f(x),$$

where x runs over all possible values of X. We could list the possible values of X: x_1, x_2,..., x_k and let

$$p_i = f(x_i) = P(X = x_i).$$

Then

$$\mu = \sum_{i=1}^{k} p_i x_i.$$

This is how Schreiber writes the mean.

Another example

What is the mean height of students in this class?

Say $\Omega =$ the set of students in this class and think of Ω as a uniform probability space. Let

$$X : \Omega \to \{ 20, 21, 22, \ldots, 200 \}$$

be the height function, where heights are measured in inches and rounded up or down to the nearest inch.

The first approach to computing the mean height is to sum up everyone's height and to divide by the number of students in the class—235, let's say.

The approach taken by the second formula (two slides ago) is to count the number of students with a given X-value, say $X = 65$. Imagine that 12 students have height 65; then $P(X = 65) = \frac{12}{235}$. Once we have tabulated the number of students with each possible height, we can compute the mean very easily by adding up numbers like $65 \cdot \frac{12}{235}$.

Expected value

According to Wikipedia:

> *In probability theory, the expected value of a random variable, intuitively, is the long-run average value of repetitions of the experiment it represents. . . . The expected value is also known as the expectation, mathematical expectation, EV, average, mean value, mean, or first moment.*

In other words:

$$\text{Expected value} = \text{mean}.$$

They are just synonyms.

Expected value

Consider the probability space consisting of the two possible outcomes of the flip of a fair coin: $\Omega = \{ \texttt{0}\ \texttt{1} \}$, each outcome occurring with probability $\frac{1}{2}$. Let

$$X(\texttt{0}) = 0, \quad X(\texttt{1}) = 1.$$

What is the expected value of X?

Answer:

$$\begin{aligned}\mu &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \\ &= 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.\end{aligned}$$

Expected value

Consider the probability space consisting of the two possible outcomes of the flip of a fair coin: $\Omega = \{\, \texttt{0}\ \texttt{1} \,\}$, each outcome occurring with probability $\frac{1}{2}$. Let

$$X(\texttt{0}) = 0, \quad X(\texttt{1}) = 1.$$

What is the expected value of X?

Answer:

$$\begin{aligned}\mu &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \\ &= 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.\end{aligned}$$

Expected value

The expected value $\frac{1}{2}$ of X is not a value of X.

We do not expect the expected value of X because it is not a value of X.

The terminology "expected value" is misleading and therefore bad.

Back to Math 10A

Regrades have been enabled for MT#2.

Please do not request a regrade on a problem unless you have already discussed your solution with your GSI. Your request should contain a statement along the lines of "My GSI Ken Ribet said this was okay."

If X has PDF equal to f, then

$$\text{mean of } X = \int_{-\infty}^{\infty} x \cdot f(x)\, dx.$$

Pareto

Fix $p > 1$ and note that

$$\int_1^\infty \frac{1}{x^p}\,dx = -\frac{1}{p-1}\,\frac{1}{x^{p-1}}\Big]_1^\infty = \frac{1}{p-1}.$$

The function

$$f(x) = \begin{cases} 0 & \text{if } x < 1, \\ \dfrac{p-1}{x^p} & \text{if } x \geq 1 \end{cases}$$

is then a PDF.

What is the mean of this PDF?

Pareto

The mean in this case is

$$\int_1^\infty x \cdot \frac{p-1}{x^p}\,dx = (p-1)\int_1^\infty \frac{1}{x^{p-1}}\,dx.$$

The integral

$$\int_1^\infty \frac{1}{x^q}\,dx$$

converges only for $q > 1$, when its value is $\dfrac{1}{q-1}$.

Hence a Pareto-distributed X has finite mean only for $p > 2$.
The mean in this case is $\dfrac{p-1}{p-2}$.

Comparisons

When we work with PDFs and CDFs, and when we compute means, we deal with improper integrals like $\int_1^\infty \frac{1}{x^p}\,dx$. This integral has the form

$$\int_a^\infty g(x)\,dx,$$

where $g(x)$ is positive.

An integral like this is either *finite* (convergent) or *infinite* (divergent).

Consequently, if $h(x) \leq g(x)$ and both h and g are positive, we have

$$\int_a^\infty g(x)\,dx < \infty \Longrightarrow \int_a^\infty h(x)\,dx < \infty.$$

If the integral of a "big" function g is convergent, so is the integral of a smaller function. That's comparison for you.

Logically, this means that if the integral of the smaller function is infinite, then so is the integral of the larger function:

$$\int_a^\infty h(x)\,dx = \infty \Longrightarrow \int_a^\infty g(x)\,dx = \infty.$$

HW Example

The integral $\int_0^\infty e^{-x}\,dx$ is easily evaluated: it's 1, so it's finite.

Consequently, $\int_0^\infty \frac{1}{1+e^x}\,dx$ is convergent: indeed, $1+e^x > e^x$, so $\frac{1}{1+e^x} < \frac{1}{e^x} = e^{-x}$. We take $h(x) = \frac{1}{1+e^x}$, $g(x) = e^{-x}\ldots.$

This is §7.2 #11, but I changed the lower limit of integration from 1 to 0.

HW Example

The integral $\int_0^\infty e^{-x}\,dx$ is easily evaluated: it's 1, so it's finite.

Consequently, $\int_0^\infty \frac{1}{1+e^x}\,dx$ is convergent: indeed, $1+e^x > e^x$, so $\frac{1}{1+e^x} < \frac{1}{e^x} = e^{-x}$. We take $h(x) = \frac{1}{1+e^x}$, $g(x) = e^{-x}$....

This is §7.2 #11, but I changed the lower limit of integration from 1 to 0.

Want to know more?

An antiderivative of $\frac{1}{1+e^x}$ turns out to be $x - \ln(1+e^x)$; this will be easy to check on the doc camera. Consequently,

$$\int_0^\infty \frac{1}{1+e^x}\,dx = \left(x - \ln(1+e^x)\right)\Big]_0^\infty$$
$$= \lim_{x\to\infty}\left((x - \ln(1+e^x)\right) + \ln(2).$$

The limit is 0 because $e^{\text{the limit}} = 1$ by l'Hôpital's rule. Hence the value of the integral is $\ln 2 \approx 0.693$. It is validating that 0.693 is less than 1.

A good limit problem

Find the limit as $x \to \infty$ of the difference

$$x - \ln(1 + e^x).$$

This is a quintessential $\infty - \infty$ example.

The integral test

While discussing comparisons, we should highlight this fact:

Take a decreasing function $f(x)$ on $[1, \infty)$ with the property that $\lim_{x \to \infty} f(x) = 0$. Typical examples: $f(x) = \frac{1}{\sqrt{x}}$, $f(x) = \frac{1}{x^2}$; more generally, $f(x) = \frac{1}{x^p}$ with $p > 0$. Then

$$f(1) + f(2) + f(3) + \cdots < \infty$$

$$\Updownarrow$$

$$\int_1^\infty f(x)\, dx < \infty.$$

Example

For $p > 0$, the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if $p > 1$.

In particular, the harmonic series

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

diverges; we saw that before.

Why is there a comparison between series and integrals?

This is based on the diagrams that we drew for left- and right-endpoint approximations to integrals. I'll redraw a few diagrams on the document camera.

Come to dinner

Dinner tonight in the math department after Keith Devlin's talk.

Dinner on Friday at Foothill DC (6:30PM).

Dinner on Sunday at CKC DC (6PM).

Let's do this, Bears!