Methods of Mathematics

Kenneth A. Ribet

UC Berkeley

Math 10B
February 23, 2016

Office hours

Office hours Monday 2:10–3:10 and Thursday 10:30–11:30 in Evans; Tuesday 10:30–noon at the SLC

Breakfasts

Our next breakfasts will be on Thursday, March 3 and on March 18, both at 8AM. Send email if you'd like to come.

Crossroads lunch

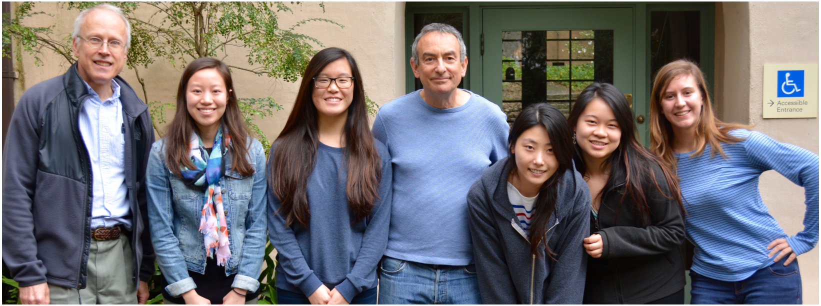
About a month ago, one of the students suggested a Crossroads lunch on

Thursday, January 25
at 12:15PM.

See you there!

Faculty Club lunch

Our next pop-up Faculty Club lunch will be at 12:30PM on Friday, February 26. So that I can keep track of the size of the group, I recommend that you let me know ahead of time if you plan on coming. *However, you can just show up!*



All are welcome.

Discrete probability distributions

We are interested in a random variable

$$X : \Omega \to \mathbf{R},$$

where Ω is a finite or countably infinite set. In the most interesting examples, the values of X are natural numbers (non-negative integers).

Concretely, we can let Ω be the set of outcomes of the experiment of flipping a coin until $\mathtt{H}$ comes up, so Ω consists of $\mathtt{H}$, $\mathtt{TH}$, $\mathtt{TTH}$, $\mathtt{TTTH}$, etc.

Last time, we introduced the random variable $X : \mathtt{T}^n\mathtt{H} \mapsto n + 1$ that takes each string to its length.

The possible values of X are the positive integers 1, 2, 3,. . . .

If X is a random variable and x is a value of X, we define

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

Said somewhat differently, $x \in \mathbf{R}$ corresponds to an event $A \subseteq \Omega$, namely

$$A = \{ \omega \in \Omega \mid X(\omega) = x \}.$$

Then $P(X = x) := P(A)$.

The theme of this class meeting is that the set of numbers $P(X = x_i)$ can be interesting!

This set of numbers is called a *probability distribution*. The sum of the numbers of a probability distribution is 1, as you can check easily.

Today we will focus on examples of probability distributions.

Geometric distribution

Let $\Omega = \{ \mathtt{T}^n\mathtt{H} \mid n \geq 0 \}$ be the set of outcomes when we flip a possibly biased coin until a head comes up. Define

$$X(\mathtt{T}^n\mathtt{H}) = n;$$

thus X (for today) is the number of tails before we get the head. (On Thursday, we looked at the random variable $X + 1$, which indicates the number of tosses, including the head toss.)

Then $P(X = n) = q^n p$, where q is the probability of getting a $\mathtt{T}$ when we flip the coin and $p = 1 - q$ is the probability of getting a $\mathtt{H}$.

Note here that the possible values of n are 0, 1, 2,. . . .

Bernoulli distribution

We go back a few steps and imagine flipping the coin only *once*. The possible outcomes are $\mathtt{T}$ and $\mathtt{H}$; thus Ω is the two-element set $\{\,\mathtt{T},\mathtt{H}\,\}$.

Let $X:\Omega\to\mathbf{R}$ be defined by

$$X(\mathtt{T})=0,\quad X(\mathtt{H})=1.$$

Then the values of X are 0 and 1, and they occur with probabilities q and p, respectively (same notation as for the geometric distribution).

Bernoulli trials

Imagine pulling balls out of a box that contains red and black balls. Suppose that 70% of the balls are red. If you pull a ball out of the box at random, the probability that it's red is 0.70.

If you pull a ball out, examine it and record the color, and then replace the ball, you return the situation to its initial state. If you repeat the exercise again and again, you're basically flipping a biased coin. To do something repeatedly when there are two outcomes and the probabilities of the two outcomes do not change is to engage in Bernoulli trials.

Binomial distribution

Now imagine that we flip a biased coin n times and record the results. Then Ω has 2^n elements; each element is a string of length n made from the characters $\texttt{T}$ and $\texttt{H}$.

For $\omega \in \Omega$, let $X(\omega)$ be the number of $\texttt{H}$s that occur in ω. For example, $X(\texttt{THTTTHHTHTH}) = 5$ (if I counted correctly).

Then $P(X = k)$ is the product of $p^k q^{n-k}$ with the number of strings that have k $\texttt{H}$s and $(n-k)$ $\texttt{T}$s. That number is $\binom{n}{k}$, so

$$P(X = k) = p^k q^{n-k} \binom{n}{k}.$$

(The sum of these numbers is $(p+q)^n = 1^n = 1$.)

Hypergeometric distribution

Imagine now that our box of balls has m red balls and $N - m$ black balls; thus it has N balls in total. Suppose that we pull out n balls (sequentially or all at once) and ask how many of these balls are red. Let

$$X = \text{the number of red balls.}$$

Then

$$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}, \quad k = 0, \dots, n.$$

Analogue: Instead of a box of balls, we could have a pack of 52 cards, 26 of which are red and 26 of which are black. We could take $n = 5$; then we're choosing a poker hand and asking how many cards in the hand are red.

Sanity check

If the previous slide is right, then the sum of the probabilities has to be 1. This means:

$$\sum_{k=0}^{n} \binom{m}{k} \binom{N-m}{n-k} = \binom{N}{n},$$

as long as n and m are no bigger than N.

Can you think of a combinatorial proof of this identity?

THIS PAGE INTENTIONALLY LEFT BLANK

Poisson distribution

This is the first time that we try to impose a distribution on values of a random variable without trying to prove that this distribution is correct through examination of Ω and the precise random variable X that is in question. The distribution is defined by

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad k = 0, 1, 2, \ldots,$$

where λ is a parameter. Because of the factor $e^{-\lambda}$, the sum of the probabilities is 1 (as it needs to be).

The interpretation of λ is that it is the *average* value of X. (On Thursday, we will use the term "expected value"!) The average is by definition the sum

$$\sum_{k=0}^{\infty} P(X=k)k = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda},$$

which works out to be λ: the sum may be rewritten

$$\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!},$$

which becomes

$$\lambda e^{-\lambda}\sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{\ell!} = \lambda.$$

Wikipedia examples

If the average number of pieces of mail you receive per day is λ, then it is reasonable to expect that the number of pieces of mail that you receive per day is distributed according to the Poisson distribution.

“On a particular river, overflow floods occur once every 100 years on average. Calculate the probability of 0, 1, 2, 3, 4, 5, or 6 overflow floods in a 100-year interval, assuming the Poisson model is appropriate.”

Derivation of the Poisson model

Imagine that I receive λ pieces of mail per day on average and I want to estimate the probability of receiving *k* pieces of mail on a given day.

My model is that all *n* people (including corporations, of course) in the US flip a biased coin every day to decide whether to send me mail. Each coin comes up $\mathtt{H}$ ("send Ribet mail") with probability *p*. On average, each person then sends me *p* pieces of mail per day, so (on average) I receive *np* pieces of mail per day all told. We conclude that $p = \lambda/n$ (a minuscule number) because I started by saying that I receive λ pieces of mail per day on average.

The probability that I receive *k* pieces of mail on a given day is then known from the binomial distribution.

Specifically, the probability that I receive k pieces of mail on a given day is

$$\binom{n}{k}p^k(1-p)^{n-k} = \frac{1}{k!}\lambda^k(1-\frac{\lambda}{n})^n J,$$

where J (for “junk”) is the product

$$1(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})\cdot(1-\frac{\lambda}{n})^{-k}.$$

We have $J \to 1$ as $n \to \infty$.

The term $(1-\frac{\lambda}{n})^n$ approaches $e^{-\lambda}$ by first-semester calculus (Math 10A?). Hence as $n \to \infty$, the probability that I receive k pieces of mail on a given day approaches $\lambda^k e^{-\lambda}\frac{1}{k!}$, which is the Poisson number. Details via document camera or white board.