October 3, 2017

CS 361: Probability & Statistics

Random variables

Bernoulli random variables

Bernoulli random variables can take on two values: 0 and 1. Good for modeling an experiment which succeeds or fails or otherwise has only two kinds of interesting outcomes.

Bernoulli distribution

$$P(X = 1) = p$$
$$P(X = 0) = (1 - p)$$

We've derived its expectation and variance before

$E[X] = p$ $\qquad$ $Var[X] = p(1-p)$

Geometric random variables

If we have a biased coin where P(H) = p and we flip this coin again and again and stop the first time we observe heads, the number of flips required is a discrete random variable <u>taking integer values greater than or equal to 1</u>

To think of what the form of the distribution of this variable is, consider that it requires us to get n-1 tails each with probability 1-p and then one head with probability p. So we have

Distribution

$$P(\{X = n\}) = (1-p)^{n-1}p$$

p is a called **parameter** of the distribution

Expectation and variance

$$E[X] = \frac{1}{p} \qquad \mathrm{var}[X] = \frac{1-p}{p^2}$$

In homework, geometric series, hence the name

Geometric random variables

- Not really just about coins
- It can work for any situation where we have trials characterized by success and failure
- In some cases we want to see how to model the number of successes until a failure
- Or we may flip the logic and wish to model the number of failures until a success finally occurs

Binomial distribution

Here's one we've derived before, too

We have a biased coin with heads coming up with probability p. The binomial probability distribution gives the probability that we will see h heads in N flips

In N independent repetitions of an experiment with a binary outcome (ie heads or tails; 0 or 1; and so on) with $P(H) = p$ and $P(T) = 1 - p$, the probability of observing a total of h H's and $N - h$ T's is

$$\underline{P_b(h; N, p)} = \binom{N}{h} p^h (1-p)^{(N-h)}$$

as long as $0 \leq h \leq N$; in any other case, the probability is zero.

This can be read P(h) when the binomial distribution is parameterized by N and p

Binomial theorem and distribution

When we did the overbooking example, someone pointed out how the distribution function of the binomial random variable looked a bit like the binomial theorem. In case you don't remember:

Binomial Theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

$\longrightarrow$ x=1-p, y=p, n=N $\longrightarrow$

P_b(X=k; N, p)

$$\sum_{k=0}^{N} \boxed{\binom{N}{k} p^k (1-p)^{N-k}} = (1-p+p)^N$$

$= 1$

So our binomial distribution sums to 1 if we add P(X=h) for every $0 \leq h \leq N$ which it must to satisfy our probability laws

Binomial distribution: recursive

It's also worth noting that our binomial distribution can be written recursively

We can get h heads in N trials by:

$$P_b(h; N, p) = pP_b(h-1; N-1, p) + (1-p)P_b(h; N-1, p)$$

Getting heads on the Nth trial and then getting h-1 heads on the other N-1 trials

Getting tails on the Nth trial and getting all h heads in the other N-1 trials

Binomial and Bernoulli

Another way of thinking about the Binomial random variable with parameters N and p is as a sum of N independent Bernoulli random variables with parameter p

$$\overbrace{X}^{\text{Binomial}} = \underbrace{Y_1 + Y_2 + \ldots + Y_N}_{N \text{ independent copies of a Bernoulli}}$$

We will use this to calculate expectation and variance

Binomial expectation

Claim:
The expected value of a binomial random variable, X, with parameters N and p is Np

Proof

Let Y_i be the Bernoulli random variable associated with the i-th coin toss (recall Bernoulli is 1 when heads comes up and 0 otherwise and is parameterized by p) so that

$$X = \sum_{i=1}^{N} Y_i \quad \text{then} \quad E[X] = E\left[\sum_{i=1}^{N} Y_i\right]$$

Using linearity

$$E[X] = \sum_{i=1}^{N} E[Y_i]$$

But E[Y_i] = p, so

$$\sum_{i=1}^{N} E[Y_i] = \sum_{i=1}^{N} p$$

Thus

$$E[X] = Np$$

Binomial variance

Claim:
The variance of a binomial random variable with parameters N and p is Np(1-p)

Same trick as before

Let Y_i be the Bernoulli random variable associated with the i-th coin toss (recall Bernoulli is 1 when heads comes up and 0 otherwise and is parameterized by p) so that

And since these coin tosses are independent

$$\mathrm{var}\left[\sum_{i=1}^{N} Y_i\right] = \sum_{i=1}^{N} \mathrm{var}[Y_i]$$

var[Y_i] was p(1-p) so we have

$$\sum_{i=1}^{N} \mathrm{var}[Y_i] = \sum_{i=1}^{N} p(1-p)$$

Giving, as desired

$$\mathrm{var}[X] = Np(1-p)$$

Multinomial variables

Binomial variables let us count the number of heads when we flip coins many times
If we rolled a die N times instead, and wanted to keep track of how many of each face came up we could do that too.

Suppose our die has k sides and we roll it N times and observe the sequence.
If side 1 appears with probability p_1 and comes up n_1 times, side 2 shows up with probability p_2 and comes up n_2 times, …, side k comes up n_k times and has probability p_k of showing up, then the probability of our sequence of N rolls is

$$p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

For a sequence of N rolls of our die, how many possible sequences would have n_1 appearance of face 1, n_2 appearances of face 2, …, n_k appearances of face k?

$$\frac{N!}{n_1! n_2! \ldots n_k!}$$

Multinomial variables

We put these two together to get the **multinomial distribution**. It tell us the probability if we roll a die N times, with what probability will see exactly n_1 of face 1, n_2 of face 2, …, and n_k of face k given that the die has probability p_1 of rolling a face 1, p_2 of rolling face 2, …, p_k of rolling face k. Or if we write it out

$$P_m(n_1, \ldots, n_k; N, p_1, \ldots, p_k) = \frac{N!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}.$$

Observe that the sum of all the n_i must be N and the sum of the p_j is 1

N and the p_j are the parameters of the distribution and it is a probability distribution over values over a tuple of n's

Example: multinomial

I throw five fair dice. What is the probability of getting two 2's and three 3's?

$$\frac{5!}{2!3!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^3$$

Poisson distribution

A distribution useful for modeling counts of events (not sets of outcomes, just colloquial events)

The number of calls a call center receives is a random variable. If we know that a call center receives on average 100 calls per hour. What is the probability it will receive 75 or 200 in a given hour?

If whatever we are counting
1) has a fixed rate and
2) the events occur independently

Then the **Poisson distribution** will be a good model

If the known rate of occurrence of the events per time interval is given by λ

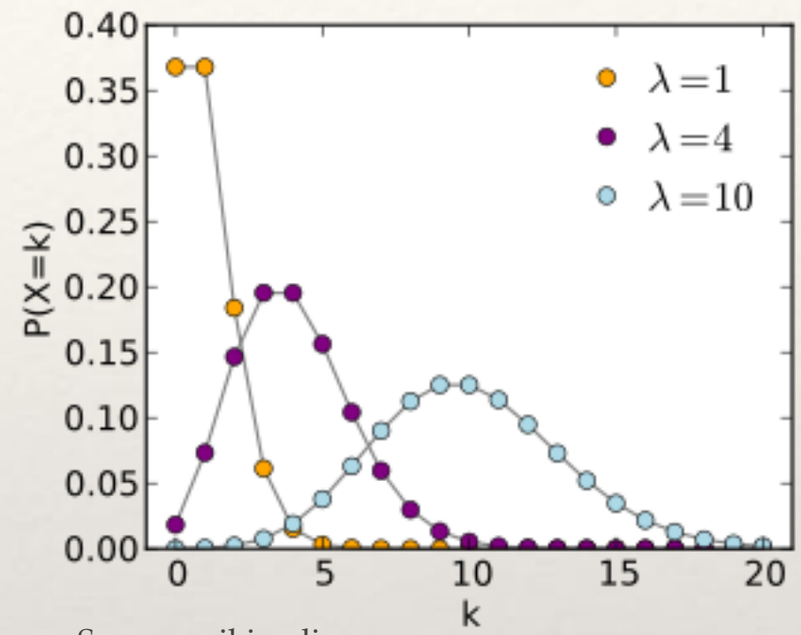
Then the probability that we will see k events in a length of time equal to the interval is given by

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

And we say that the count of events is a Poisson random variable with intensity λ

Poisson distribution

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



Source: wikipedia

1. The mean of a Poisson distribution with intensity λ is λ.
2. The variance of a Poisson distribution with intensity λ is λ (no, that's not an accidentally repeated line or typo).

Poisson examples

Number of decay events by a radioactive substance in a given time

Number of health insurance claims per month (assuming no disasters or epidemics)

We talked about counts of events in time, could just as easily use counts of events per interval of space

Roadkill per mile of road

Number of genetic mutations per 100 thousand bases as we traverse a strand of DNA

Observation:
If we double, or halve, the length of the interval, we will have to double, or halve, the intensity

Poisson point processes

A Poisson point process with intensity λ is a set of random points with the property that the number of points within an interval of length *s* is a Poisson random variable with intensity λs

Can easily generalize this to points on a plane or in space by saying that such a process in n dimensional space is one with random points in a region *D* and the number of points in any subset *s* of *D* is described by a Poisson random variable with intensity $\lambda m(s)$ where m(s) is the area or volume of *s*

Continuous uniform distribution

Uniform random variables take on any value within a range with equal probability

Density function for a random variable that's uniformly distributed in the range from l to u

$$p(x) = \begin{cases} 0 & x < l \\ 1/(u-l) & l \leq x \leq u \\ 0 & x > u \end{cases}$$

Exponential distribution

Suppose we have an infinite interval of time or space, with points randomly distributed on it. Assume these points form a Poisson point process with intensity λ

The distance between two consecutive points will be a random variable X described by an **exponential distribution** which has a distribution given by

$$p(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Useful for modeling failures. If failures form a Poisson point process in time and we have just observed a failure, the distribution of times until the next failure will be given by an exponential distribution

Exponential distribution

For an exponential random variable, we have

1. The mean is $\frac{1}{\lambda}$.
2. The variance is $\frac{1}{\lambda^2}$.

So if we are describing call center calls as a Poisson point process with intensity λ per hour. The number of calls per hour is given by the Poisson distribution. The expected number of calls per hour is λ. The time until the next call is given by an exponential distribution and the expected time until the next call is $\frac{1}{\lambda}$

Standard normal distribution

Definition: 6.6 *Standard Normal distribution*

The probability density function

$$p(x) = \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left(\frac{-x^2}{2}\right).$$

is known as the **standard normal distribution**

The random variable whose density is given by p is a **standard normal random variable**