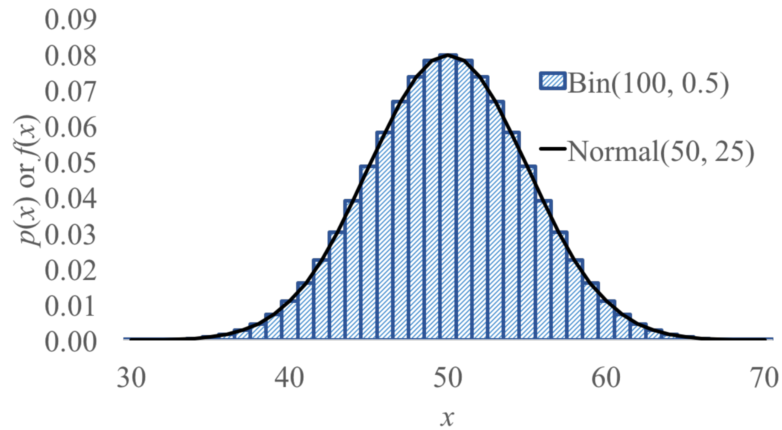
Chris Piech
CS109

Lecture #11
Oct 18th, 2017

Binomial Approximation and Joint Distributions

Binomial Approximation

For certain values, a normal can be used to approximate a Binomial. Let's take a side by side view of a normal and a binomial:



Lets say our binomial is a random variable $X \sim \text{Bin}(100, 0.5)$ and we want to calculate $P(X \geq 55)$. We could cheat by using the closest fit normal (in this case $Y \sim N(50, 25)$). How did we chose that particular Normal? Simply select one with a mean and variance that matches the Binomial expectation and variance. The binomial expectation is $np = 100 \cdot 0.5 = 50$. The Binomial variance is $np(1-p) = 100 \cdot 0.5 \cdot 0.5 = 25$.

You can use a Normal distribution to approximate a Binomial $X \sim Bin(n, p)$. To do so define a normal $Y \sim (E[X], Var(X))$. Using the Binomial formulas for expectation and variance, $Y \sim (np, np(1-p))$. This approximation holds for large n and moderate p. Since a Normal is continuous and Binomial is discrete we have to use a continuity correction to discretize the Normal.

$$P(X = k) \sim P\left(k - \frac{1}{2} < Y < k + \frac{1}{2}\right) = \Phi\left(\frac{k - np + 0.5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - np - 0.5}{\sqrt{np(1-p)}}\right)$$

You should get comfortable deciding what continuity correction to use. Here are a few examples of discrete probability questions and the continuity correction:

Discrete (Binomial) probability question	Equivalent continuous probability question
$P(X = 6)$	$P(0.5 < X < 6.5)$
$P(X \geq 6)$	$P(X > 5.5)$
$P(X > 6)$	$P(X > 6.5)$
$P(X < 6)$	$P(X < 5.5)$
$P(X \leq 6)$	$P(X < 6.5)$

Example 3

100 visitors to your website are given a new design. Let X = # of people who were given the new design and spend more time on your website. Your CEO will endorse the new design if $X \geq 65$. What is

$P(\text{CEO endorses change}|\text{it has no effect})$?

$E[X] = np = 50$. $Var(X) = np(1-p) = 25$. $\sigma = \sqrt{(Var(X))} = 5$. We can thus use a Normal approximation: $Y \sim \mathcal{N}(50, 25)$.

$$P(X \geq 65) \approx P(Y > 64.5) = P\left(\frac{Y-50}{5} > \frac{64.5-50}{5}\right) = 1 - \Phi(2.9) = 0.0019$$

Example 4

Stanford accepts 2480 students and each student has a 68% chance of attending. Let X = # students who will attend. $X \sim Bin(2480, 0.68)$. What is $P(X > 1745)$?

$E[X] = np = 1686.4$. $Var(X) = np(1-p) = 539.7$. $\sigma = \sqrt{(Var(X))} = 23.23$. We can thus use a Normal approximation: $Y \sim \mathcal{N}(1686.4, 539.7)$.

$$P(X > 1745) \approx P(Y > 1745.5) = P\left(\frac{Y-1686.4}{23.23} > \frac{1745.5-1686.4}{23.23}\right) = 1 - \Phi(2.54) = 0.0055$$

Joint Distributions

Often you will work on problems where there are several random variables (often interacting with one another). We are going to start to formally look at how those interactions play out.

For now we will think of joint probabilities with two events X and Y.

Discrete Case

In the discrete case a joint probability mass function tells you the probability of any combination of events $X = a$ and $Y = b$:

$$p_{X,Y}(a,b) = P(X = a, Y = b)$$

This function tells you the probability of all combinations of events (the "," means "and"). If you want to back calculate the probability of an event only for one variable you can calculate a "marginal" from the joint probability mass function:

$$p_X(a) = P(X = a) = \sum_y P_{X,Y}(a,y)$$
$$p_Y(b) = P(Y = b) = \sum_x P_{X,Y}(x,b)$$

In the continuous case a joint probability density function tells you the relative probability of any combination of events $X = a$ and $Y = y$.

In the discrete case, we can define the function $p_{X,Y}$ non-parametrically. Instead of using a formula for p we simply state the probability of each possible outcome.

Multinomial Distribution

Say you perform n independent trials of an experiment where each trial results in one of m outcomes, with respective probabilities: $p_1, p_2, \ldots, p_m$ (constrained so that $\sum_i p_i = 1$). Define X_i to be the number of trials with outcome i. A multinomial distribution is a closed form function that answers the question: What is the probability that there are c_i trials with outcome i. Mathematically:

$$P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m) = \binom{n}{c_1, c_2, \ldots, c_m} p_1^{c_1} p_2^{c_2} \ldots p_m^{c_m}$$

Example 1

A 6-sided die is rolled 7 times. What is the probability that you roll: 1 one, 1 two, 0 threes, 2 fours, 0 fives, 3 sixes (disregarding order).

$$\begin{aligned}
P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3) &= \frac{7!}{2!3!} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^3 \\
&= 420 \left(\frac{1}{6}\right)^7
\end{aligned}$$

Fedaralist Papers

In class we wrote a program to decide whether or not James Madison or Alexander Hamilton wrote Fedaralist Paper 49. Both men have claimed to be have written it, and hence the authorship is in dispute. First we used historical essays to estimate p_i, the probability that Hamilton generates the word i (independent of all previous and future choices or words). Similarly we estimated q_i, the probability that Madison generates the word i. For each word i we observe the number of times that word occurs in Fedaralist Paper 49 (we call that count c_i). We assume that, given no evidence, the paper is equally likely to be written by Madison or Hamilton.

Define three events: H is the event that Hamilton wrote the paper, M is the event that Madison wrote the paper, and D is the event that a paper has the collection of words observed in Fedaralist Paper 49. We would like to know whether $P(H|D)$ is larger than $P(M|D)$. This is equivalent to trying to decide if $P(H|D)/P(M|D)$ is larger than 1.

The event $D|H$ is a multinomial parameterized by the values p. The event $D|M$ is also a multinomial, this time parameterized by the values q.

Using Bayes Rule we can simplify the desired probability.

$$\begin{aligned}
\frac{P(H|D)}{P(M|D)} &= \frac{\frac{P(D|H)P(H)}{P(D)}}{\frac{P(D|M)P(M)}{P(D)}} = \frac{P(D|H)P(H)}{P(D|M)P(M)} = \frac{P(D|H)}{P(D|M)} \\
&= \frac{\binom{n}{c_1, c_2, \ldots, c_m} \prod_i p_i^{c_i}}{\binom{n}{c_1, c_2, \ldots, c_m} \prod_i q_i^{c_i}} = \frac{\prod_i p_i^{c_i}}{\prod_i q_i^{c_i}}
\end{aligned}$$

This seems great! We have our desired probability statement expressed in terms of a product of values we have already estimated. However, when we plug this into a computer, both the numerator and denominator come out to be zero. The product of many numbers close to zero is too hard for a computer to represent. To fix this problem, we use a standard trick in computational probability: we apply a log to both sides and apply some basic rules of logs.

$$\begin{aligned}
\log\Big(\frac{P(H|D)}{P(M|D)}\Big) &= \log\Big(\frac{\prod_i p_i^{c_i}}{\prod_i q_i^{c_i}}\Big) \\
&= \log(\prod_i p_i^{c_i}) - \log(\prod_i q_i^{c_i}) \\
&= \sum_i \log(p_i^{c_i}) - \sum_i \log(q_i^{c_i}) \\
&= \sum_i c_i \log(p_i) - \sum_i c_i \log(q_i)
\end{aligned}$$

This expression is "numerically stable" and my computer returned that the answer was a negative number. We can use exponentiation to solve for $P(H|D)/P(M|D)$. Since the exponent of a negative number is a number smaller than 1, this implies that $P(H|D)/P(M|D)$ is smaller than 1. As a result, we conclude that Madison was more likely to have written Fedaralist Paper 49.