

Class15-17

September 28, 2020

0.1 Intro

Last week we explored distributions of random variable that serve as building blocks for more complicated distributions. This week we'll explore three distributions: the binomial, normal, and poisson and see how they can be used in pop health applications.

0.2 The Binomial distribution and fun with counting

The binomial distribution is a discrete distribution assigning probabilities to a number of successes (usually assigned a value of 1) out of a total number of N . We assume each of the N trials are independent from one another, that only a success or failure can occur, and that a success occurs with a probability θ . The binomial distribution is used anytime you ask "what is the probability of x occurrences of the same event out of a total N number of tries?"

0.2.1 Definition

The probability mass function (discrete so we can assign probabilities to individual outcomes) is

$$p(x) = \binom{N}{x} \theta^x (1 - \theta)^{N-x} \quad (1)$$

The symbol $\binom{N}{x}$ is called the a binomial coefficient, sometimes said "N choose x". The binomial coefficient (we'll see below) is the number of times you can select x items from a total of N items without caring about the order you selected them.

0.2.2 Expectation and Variance

The expectation is

$$E(x) = N\theta \quad (2)$$

and variance is

$$Var(x) = N\theta(1 - \theta) \quad (3)$$

0.2.3 Counting

Lets pull apart the probability mass function of the binomial distribution.

$$p(x) = \binom{N}{x} \theta^x (1 - \theta)^{N-x} \quad (4)$$

Suppose we have a total number of N “trials” and we want to compute the probability of x successes. Well if there are x successes there must have been $N - x$ failures. Lets assign success the value 1 and failures the value 0. If we ordered the trial $1, 2, 3, \dots, N$, one way we could have x successes and $N - X$ failures is

$$(1, 1, 1, 1, 1, 1, 1, \dots, 1, 0, 0, 0, 0, \dots, 0, 0) \quad (5)$$

Since each of these N outcomes are independent we have a convenient way of writing the probability.

$$p(\text{first outcome is a 1, second outcome is a 1, } \dots, x^{\text{th}} \text{ outcome is a 1}) = \theta \times \theta \times \theta \times \dots \times \theta = \theta^x \quad (6)$$

In the same way we can compute the probability of $N - x$ failures

$$p(N-x \text{ failures}) = (1 - \theta)^{N-x} \quad (7)$$

And so the probability of x successes and $N - x$ failures is

$$p(x \text{ success and } N-x \text{ failures in order}) = \theta^x (1 - \theta)^{N-x} \quad (8)$$

but this is only one way we could have ended up with x successes and $N - x$ failures. The binomial distribution assigns a probability to all possible ways (all combinations) we could have had x successes and $N - x$ failures.

Permutations and combinations (the binomial coefficient)

Permutations A **permutation** counts the number of ways you can select s objects from a total of N objects—in order. For example, suppose i have 5 dogs that are fuzzy, furry, floofy, hairy, and shaggy. How many ways can i select two dogs from the pack of five in order (ie picking the shaggy then hairy dog is distinct from selecting the hairy then shaggy dog)?

Well to pick the first dog, we have 5 options. For each of those 5 options we have 4 options remaining after that so then there is a total of

$$5 \times 4 = 20 \quad (9)$$

ways we can select 2 dogs from a pack of 5 so that the order matters. By a similar line of thought, there are

$$5 \times 4 \times 3 = 60 \quad (10)$$

ways to choose 3 dogs from 5 in order. And we notice a pattern.

In general there are

$$nPr = n \times (n - 1) \times (n - 2) \times \dots \times (n - r + 1) \quad (11)$$

ways to select r items from a set of n total objects. The above nPr is said aloud “ n permute r ”.

Factorial Lets introduce a new mathematical symbol that will come in handy when we discuss combinations. A factorial is a function that takes as input an integer and returns the product of all values between 1 and the integer. We write this function as

$$N! = N \times (N - 1) \times (N - 2) \times \dots \times 2 \times 1 \quad (12)$$

We could rewrite our permutation in terms of factorials.

$$nPr = \frac{n!}{(n - r)!} \quad (13)$$

Combinations (the binomial coefficient) Combinations count the number of times we can chose r items from a set of n items, but unlike a permutation, the order of our selection doesn’t matter. In our 5 puppy dog example, selecting shaggy and then hairy would be the same as selecting hairy and then shaggy.

We can develop a formula for a combination by relating combinations to permutations (a formula we just came up with earlier). A permutation where we select r objects from a total of n could be done as follows: (i) select r objects (ii) for each choice of r objects, think of all the ways we could order them.

$$nPr = \text{select } r \text{ objects (unordered)} \times \text{the ways we can order these } r \text{ objects} \quad (14)$$

A permutation can count the number of ways we can select r objects, and for every selection, the number of ways we can order them. We see that the number of ways to select r object unordered

is what we want—our combination. Lets give the combination—the number of ways to select r objects from a total of n objects a symbol: nCr or more often $\binom{n}{r}$.

So then our formula above is

$$nPr = nCr \times \text{the ways we can order these } r \text{ objects} \quad (15)$$

All we need to know is the number of ways we can order r different objects. If we figure this out then our formula for $\binom{n}{r}$ will be

$$\binom{n}{r} = \frac{nPr}{\text{the ways we can order these } r \text{ objects}} \quad (16)$$

Assign the labels $1, 2, 3, \dots, r$ to our first, second, third, and so on selection. From our r selected and unordered objects, choose one. There are r positions we could assign this first object. Choose a second object. There are $r - 1$ positions we can assign the second object. Choose the third object. There are $r - 2$ positions we can assign the second object. A permutation! So then there are $r \times (r - 1) \times (r - 2) \cdots 1 = r!$ ways we can order our r objects.

$$\binom{n}{r} = \frac{nPr}{r!} \quad (17)$$

$$= \frac{n!}{(n - r)!r!} \quad (18)$$

0.2.4 The point (back to the binomial)

We left off with constructing the binomial distribution here,

$$p(x \text{ successes and } N-x \text{ failures in order}) = \theta^x (1 - \theta)^{N-x}, \quad (19)$$

a single way to get x successes and $N - x$ failures. But we want all possible ways. Well any choice of x trials from the total N trials available could have successes. How many ways can we pick x trials from N total trials without paying attention to their order? The binomial coefficient $\binom{N}{x}$.

$$p(x \text{ successes and } N-x \text{ failures}) = \binom{N}{x} \theta^x (1 - \theta)^{N-x} \quad (20)$$

0.3 Application

The [PARTNER trial](#) enrolled 358 patients with severe [aortic stenosis](#). Patients were randomized 1:1 to a control group (standard therapy) or to receive a trans-apical valve replacemant (TAVR). The proportion of patients in the control group, out of a total of 179, who experienced a stroke in

the first 30 days was estimated to be 5% while the proportion of patients in the TAVR group, out of a total of 179, experienced stroke in the first 30 days at a rate of 1%.

If we assume that every patient is independent of one another and has the same probability of a stroke within the first 30 days, we can model the probability of the number of patients experiencing a stroke in both groups with a binomial distribution.

Define the r.v. S_{control} to be the number of strokes in the control group and S_{TAVR} to be the number of strokes in the TAVR group. Assume S_{control} follows a Binomial distribution with $N=179$ and $\theta = 0.05$ —or $S_{\text{control}} \sim \text{Binom}(179, 0.05)$. Also assume S_{TAVR} follows a Binomial distribution with $N=179$ and $\theta = 0.01$ —or $S_{\text{TAVR}} \sim \text{Binom}(179, 0.01)$.

The expected value of $S_{\text{control}} = N * p = 179 \times 0.05 = 8.95$ and the expected value of $S_{\text{TAVR}} = N * p = 179 \times 0.01 = 1.79$. On average, we would expect patients who receive a TAVR to have $1.79/8.95 = 20\%$ of the proportion of strokes compared to control patients, an 80% reduction.

0.4 The normal distribution and “expectedness”

The Normal (or Gaussian) distribution describes the probability of a continuous random variable. This is the (likely familiar) bell curve.

0.4.1 Definition

The Normal distribution has two parameters: the mean (μ) and the standard deviation (σ). The pdf is symmetric and unimodal and defined over all values from negative infinity to positive infinity. Values close to the mean are much, much more likely than values further from the mean. Because of this, the Normal distribution describes phenomena or values of a r.v. that are more or less expected to be close to μ —surprises are not very likely.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (21)$$

$$= \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} \quad (22)$$

$$= (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right\} \quad (23)$$

0.4.2 Expectation and variance

If X is a r.v. and normally distributed $X \sim \mathcal{N}(\mu, \sigma^2)$, the expectation is

$$E(x) = \mu \quad (24)$$

and variance is

$$\text{Var}(x) = \sigma^2 \quad (25)$$

0.4.3 Z scores and “standardizing”

0.4.4 The expectedness of Normal distributions

0.5 The Poisson distribution and incidence