Class15-17

September 30, 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

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

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

$$E(x) = N\theta \tag{2}$$

and variance is

$$Var(x) = N\theta(1-\theta) \tag{3}$$

0.2.3 Counting

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

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

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 $1, 2, 3, \cdots, N$, one way we could have x successes and $N-X$ failures is

$$(1, 1, 1, 1, 1, 1, 1, \cdots, 1, 0, 0, 0, 0, \cdots, 0, 0) \tag{5}$$

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

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

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

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

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

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

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.

We can define an events $O_1, O_2, \cdots, O_?$ for each way x 1's and $N-x$ 0's could be arranged. Since each event contains a single arrange of 1s and 0s they are disjoint. Then the probability of x 1s and $N-x$ 0s, arranged in any way possible is

$$p(O_1 \cup O_2 \cup O_3 \cup \cdots O_?) = p(O_1) + p(O_2) + p(O_3) + \cdots + p(O_?) \text{ (disjoint)} \tag{9}$$

for any arrangement of x 1s and $N-x$ 0s we can compute the probability. It is, from before, $\theta^x(1-\theta)^{N-x}$ and so our probability of any arrangement above is

$$
\begin{aligned}
p(O_1 \cup O_2 \cup O_3 \cup \cdots O_?) &= p(O_1) + p(O_2) + p(O_3) + \cdots + p(O_?) \text{ (disjoint)} \quad (10)\\
&= \theta^x(1-\theta)^{N-x} + \theta^x(1-\theta)^{N-x} + \theta^x(1-\theta)^{N-x} + \cdots + \theta^x(1-\theta)^{N-x} \quad (11)\\
&= \sum_{i=1}^{?} \theta^x(1-\theta)^{N-x} \quad (12)\\
&= ? \times \theta^x(1-\theta)^{N-x} \quad (13)
\end{aligned}
$$

where ? is the total number of arrangements of x 1s and $N - x$ 0s. The question is, how many arrangements are there?

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

$$5 \times 4 = 20 \quad (14)$$

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

$$5 \times 4 \times 3 = 60 \quad (15)$$

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

In general there are

$$\text{nPr} = n \times (n-1) \times (n-2) \times (n-r+1) \quad (16)$$

ways to select r items from a set of n total objects. The above n**Pr** 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

$$N! = N \times (N-1) \times (N-2) \times \cdots 2 \times 1 \quad (17)$$

We could rewrite our permutation in terms of factorials.

$$\text{nPr} = \frac{n!}{(n-r)!} \tag{18}$$

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.

$$\text{nPr} = \text{ select r objects (unordered) } \times \text{ the ways we can order these r objects} \tag{19}$$

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: n**C**r or more often $\binom{n}{r}$.

So then our formula above is

$$\text{nPr} = \text{nCr} \times \text{ the ways we can order these r objects} \tag{20}$$

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

$$\binom{n}{r} = \frac{\text{nPr}}{\text{the ways we can order these r objects}} \tag{21}$$

Assign the labels $1, 2, 3, \cdots, r$ 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.

$$\binom{n}{r} = \frac{\text{nPr}}{r!} \tag{22}$$

$$= \frac{n!}{(n-r)!r!} \tag{23}$$

0.2.4 The point (back to the binomial)

We left off with constructing the binomial distribution here,

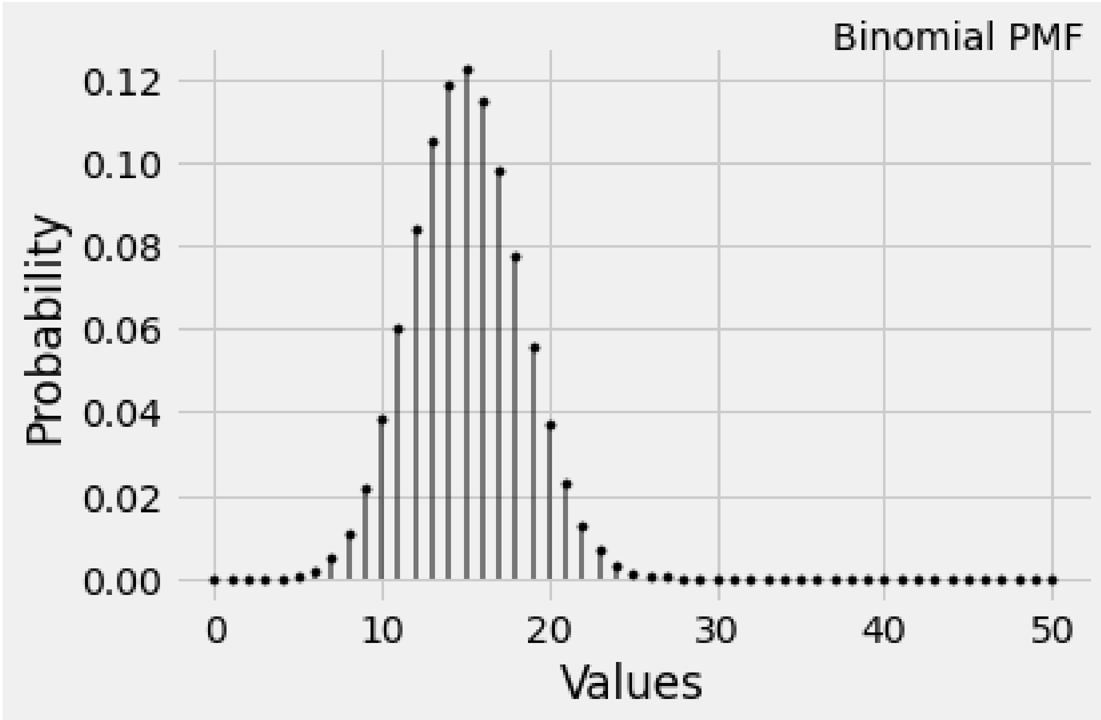
$$p(\text{x successes and N-x failures in order}) = \theta^x (1-\theta)^{N-x}, \tag{24}$$

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 $\binom{N}{x}$.

$$p(\text{x successes and N-x failures }) = \binom{N}{x} \theta^x (1-\theta)^{N-x} \tag{25}$$

0.2.5 The PMF function

[10]:
```
from scipy import stats

PMF = stats.binom(50,0.3).pmf # first number is N and seoncd number is p
domain = np.arange(0,50+1,1)  # a list from 0 to 50

plt.style.use("fivethirtyeight")
fig,ax = plt.subplots()

ax.scatter(domain, PMF(domain), s=10,color="k")
for d in domain:
    ax.plot([d]*2,[0,PMF(d)],'k-',lw=2,alpha=0.5)

ax.set_xlabel("Values")
ax.set_ylabel("Probability")

ax.text(0.99,0.99,"Binomial PMF",fontsize=14,transform=ax.transAxes,ha='right')

fig.set_tight_layout(True)
plt.show()
```



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 unimodel and defined over all values from negative infinity to positive infinity. Values close to the mean are much, much more likely that 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.

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma}} && (26) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma} \right\} && (27) \\
&= \left(2\pi\sigma^2\right)^{-1/2} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\} && (28)
\end{aligned}
$$

0.4.2 Expectation and variance

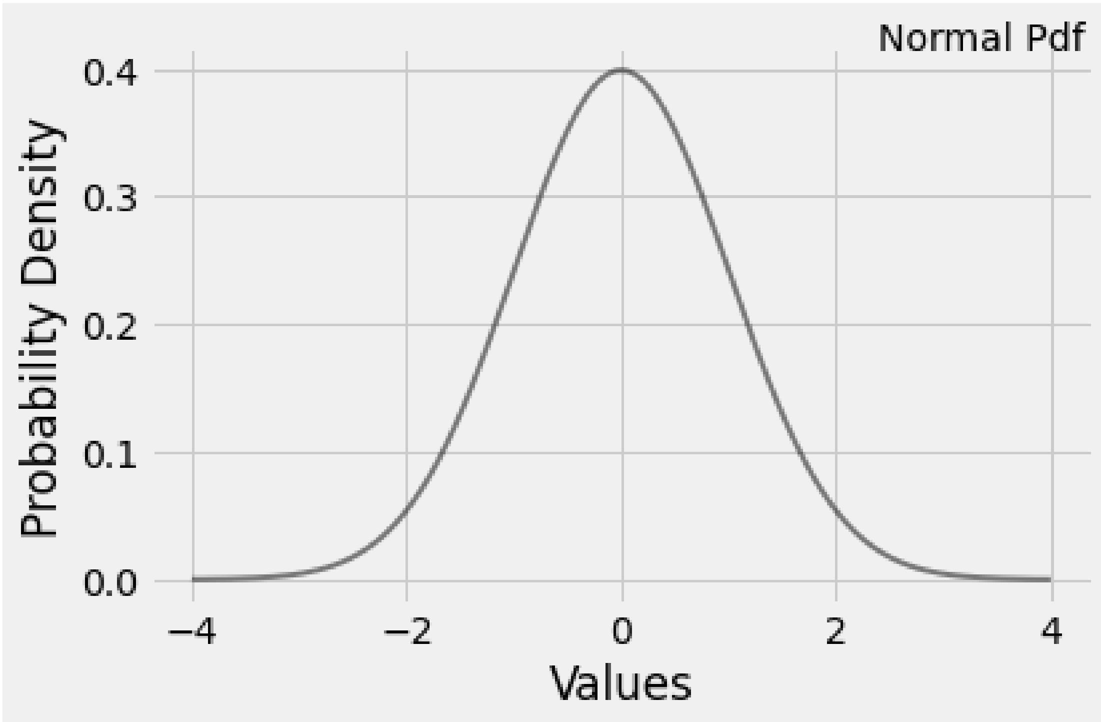
If X is a r.v. and normally distributed $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, the expectation is

$$E(x) = \mu \tag{29}$$

and variance is

$$Var(x) = \sigma^2 \tag{30}$$

[15]:
```
from scipy import stats

Pdf = stats.norm(0,1).pdf # first number is  mean and seoncd number is std dev
domain = np.linspace(-4,4,200) # 200 points linearly spaced between -3 and 3

plt.style.use("fivethirtyeight")
fig,ax = plt.subplots()

ax.plot(domain,Pdf(domain) ,'k-',lw=2,alpha=0.5)

ax.set_xlabel("Values")
ax.set_ylabel("Probability Density")

ax.text(0.99,0.99,"Normal Pdf",fontsize=14,transform=ax.transAxes,ha='right')

fig.set_tight_layout(True)
plt.show()
```



0.4.3 The expectedness of Normal distributions

The normal distribution's pdf is a decreasing exponential of a squared quantity. We know exponentials increase and decrease fast, and squaring the term inside the exp means they grow or shrink even fast. Because of this, high probabilities are assigned to intervals close to the mean. In some sense, a r.v. with a normal distribution assumes data are more or less expected to be near the mean.

To be more precise, a 0.68 probability is assigned to the interval within one standard deviation of the mean, 0.95 probability within 2 standard deviations and 0.997 within three standard deviations:

- $p[(\mu - \sigma, \mu + \sigma)] = 0.68$
- $p[(\mu - 2\sigma, \mu + 2\sigma)] = 0.95$
- $p[(\mu - 3\sigma, \mu + 3\sigma)] = 0.997$

It is important to keep in mind that a normal distribution characterizes values expected to be close to their mean—most values will fall within 2 standard deviations.

0.4.4 Z scores and "standardizing"

The normal distribution also plays a role in "standardizing" data. We can standardize a variable X with the following algorithm: 1. Compute the mean of X 2. Compute the standard deviation of X 3. For each value in X 1. subtract the mean 2. divide the value in 3.1. by the standard deviation

By standardizing, the values of X are put in terms of units of standard deviation. A value of 0 mean the value is the same as the mean of the data. A value of 1 is one standard deviation larger than the mean and so on.

Standardizing is important when you want to compare two variables that have different units, or that have different variances around their mean.

0.4.5 Application (and TL;DR)

Networks—nodes and their associated links—are an abstract way to look at a set of objects and their associated connections. A social network could define nodes as user accounts and links if two accounts register as "friends". An electrical network could define nodes as power sources and link as the physical links between two power sources.

One way to characterize networks is to define a r.v. that assigns a probability to the number of links a chosen node has (the number of links of a node is called it's degree): this distribution is called the degree distribution.

The degree distribution of many networks looks much much different than the Normal distribution. Lets look for example

```
[ ]: #<include example>
```

0.5 The Poisson distribution and incidence

0.5.1 Definition

A random variable X has a Poisson distribution—$X \sim \text{Pois}(\lambda)$— if it is discrete and it's probability mass function is

$$p(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \tag{31}$$

A Poisson distributed random variable assigns probabilities to the number of events that occur over a specific time period. The parameter λ is typically thought of as the rate of events per unit of time. For example, the number of deaths per month, number of phone calls per week, or number of emails per hour (so many).

0.5.2 Expected value and variance

The expected value is

$$E(X) = \lambda \tag{32}$$

and the variance is

$$Var(X) = \lambda \tag{33}$$

```
[ ]:
```