Probability Distributions and Functions

POSC 3410 - Quantitative Methods in Political Science

Steven V. Miller

Department of Political Science

Goal for Today

Discuss probability distributions.

Introduction

Last lecture discussed probability and counting.

- While abstract, these are important foundation concepts for what we're doing in applied statistics.

Today, we're going to talk about probability distributions.

- Our most prominent tool for statistical inference makes assumptions about parameters given a known (i.e. normal) distribution.

Refresher

Recall the choose notation (aka **combination**):

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \tag{1}$$

The exclamation marks indicate a factorial.

- e.g. $5! = 5 * 4 * 3 * 2 * 1$.

Binomial Theorem

The most common use of a choose notation is the **binomial theorem**.

- Given any real numbers X and Y and a nonnegative integer n,

$$(X+Y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \tag{2}$$

A special case occurs when $X = 1$ and $Y = 1$.

$$2^n = \sum_{k=0}^{n} \binom{n}{k} \tag{3}$$

Binomial Theorem

This is another theorem with an interesting history.

- Euclid knew of it in a simple form.
- The Chinese may have discovered it first (Chu Shi-Kié, 1303)
- General form presented here owes to Pascal in 1654.

Binomial Theorem

The binomial expansion increases in polynomial terms at an interesting rate.

$$
\begin{aligned}
(X+Y)^0 &= 1 \\
(X+Y)^1 &= X+Y \\
(X+Y)^2 &= X^2+2XY+Y^2 \\
(X+Y)^3 &= X^3+3X^2Y+3XY^2+Y^3 \\
(X+Y)^4 &= X^4+4X^3Y+6X^2Y^2+4XY^3+Y^4 \\
(X+Y)^5 &= X^5+5X^4Y+10X^3Y^2+10X^2Y^3+5XY^4+Y^5
\end{aligned} \tag{4}
$$

Notice the symmetry?

Pascal's Triangle

The coefficients form **Pascal's triangle**, which summarizes the coefficients in a binomial expansion.

$$
\begin{array}{lccccccccccc}
n=0: & & & & & & 1 & & & & & \\
n=1: & & & & & 1 & & 1 & & & & \\
n=2: & & & & 1 & & 2 & & 1 & & & \\
n=3: & & & 1 & & 3 & & 3 & & 1 & & \\
n=4: & & 1 & & 4 & & 6 & & 4 & & 1 & \\
n=5: & 1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

Pascal's Triangle

Beyond the pyramidal symmetry, Pascal's triangle has a lot other cool features.

- Any value in the table is the sum of the two values diagonally above it.
- The sum of the kth row (counting the first row as zero row) can be calculated as $\sum_{j=0}^{k} \binom{k}{j} = 2^k$
- If you left-justify the triangle, the sum of the diagonals form a Fibonacci sequence.
- If a row is treated as consecutive digits, each row is a power of 11 (i.e. magic 11s).

There are many more mathematical properties in Pascal's triangle. These are just the cooler/more famous ones.

Binomial Mass Function

Beyond cool math stuff, these have a purpose for statistics.

Let's start basic: how many times could we get heads in 10 coin flips?

- The sample space $S = \{ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 \}$
- We expect 10 heads (or no heads) to be unlikely, assuming the coin is fair.

This is a combination issue.

- For no heads, *every* flip must be a tail.
- For just one head, we have more combinations.

What's the probability of a series of coin flips with just one head?

- For a small number of trials, look at Pascal's triangle.
- For 5 trials, there is 1 way to obtain 0 heads, 5 ways to obtain 1 head, 10 ways to obtain 2 and 3 heads, 5 ways to obtain 4 heads, and 1 way to obtain 5 heads.

Binomial Mass Function

This is also answerable by reference to the **binomial mass function**, itself derivative of the **binomial theorem**.

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \tag{5}$$

where:

- $x =$ the count of "successes" (e.g. number of heads in a sequence of coin flips)
- $n =$ the number of trials.
- $p =$ probability of success in any given trial.

Binomial Mass Function

What's the probability of getting five heads on ten fair coin flips.

$$
\begin{aligned}
p(x = 5 \mid n = 10, p = .5) &= \binom{10}{5}(.5)^5(1-.5)^{10-5} \\
&= (252) * (.03125) * (.03125) \\
&= 0.2460938 \qquad (6)
\end{aligned}
$$

In R:

```
> dbinom (5,10,.5)
[1] 0.2460938
```

An Application: Everyone Hates Congress

Congress is doing nothing at a historic rate. This much we know.

- About 5% of bills that receive "some action" are ultimately passed.
  - Recall: many bills introduced die a quick death from inactivity.
  - This estimate says nothing about substantive importance of the bill.

Assume $p = .05$. What's the probability that Congresses passes three (x) of the next 20 (n) bills it gets?

An Application: Everyone Hates Congress

$$
\begin{aligned}
p(x = 3 \mid n = 20, p = .05) &= \binom{20}{3}(.05)^3(1-.05)^{20-3} \\
&= (1140) * (.000125) * (0.4181203) \\
&= 0.05958215 \qquad (7)
\end{aligned}
$$

In R:

```
> dbinom(3,20,.05)
[1] 0.05958215
```

Normal Functions

The binomial (and related: Bernoulli) are common density functions for modeling social/political phenomena.

A "normal" function is also quite common.

- Data are distributed such that the majority cluster around some central tendency.
- More extreme cases occur less frequently.

Normal Density Function

We can model this with a **normal density function**.

- Sometimes called a Gaussian distribution in honor of Carl Friedrich Gauss, who discovered it.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e\{-\frac{(x-\mu)^2}{2\sigma^2}\}, \tag{8}$$

where: $\mu =$ the mean, $\sigma^2 =$ the variance.

Normal Density Function

Properties of the normal density function.

- The tails are asymptote to 0.
- The kernel (inside the exponent) is a basic parabola.
  - The negative component flips the parabola downward.
- Denoted as a function in lieu of a probability because it is a continuous distribution.
- The distribution is perfectly symmetrical.
  - The mode/median/mean are the same values.
  - -*x* is as far from μ as *x*.

x is unrestricted. It can be any value you want in the distribution.

- μ and σ^2 are parameters that define the shape of the distribution.
  - μ defines the central tendency.
  - σ^2 defines how short/wide the distribution is.

Demystifying the Normal Density Function

Let's unpack this normal density function further (and use some R code).

Here is our normal density function.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e\{-\frac{(x-\mu)^2}{2\sigma^2}\} \tag{9}$$

Assume, for simplicity, $\mu = 0$ and $\sigma^2 = 1$.

Demystifying the Normal Density Function

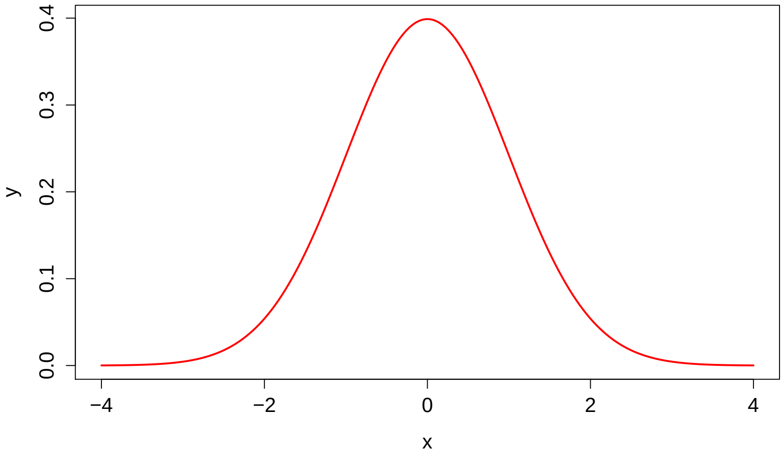
When $\mu = 0$ and $\sigma^2 = 1$, the normal density function is a bit simpler.

$$f(x) = \frac{1}{\sqrt{2\pi}} e \{-\frac{x^2}{2}\} \tag{10}$$

Let's plot it next in R.

```
> x=seq(-4,4,length=200)
> y=1/sqrt(2*pi)*exp(-x^2/2)
> plot(x,y,type="l",lwd=2,col="red")
```

The Normal Distribution

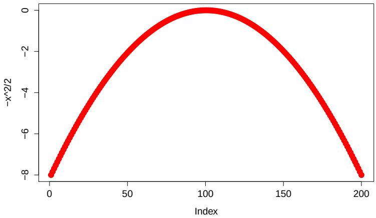


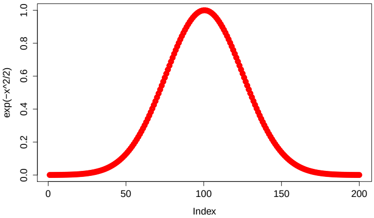
Demystifying the Normal Distribution

Let's look inside the exponent.

- The term inside the brackets ($-x^2/2$) is a parabola.
- Exponentiating it makes it asymptote to 0.

```
> x=seq(-4,4,length=200)
> plot(-x^2/2, lwd=6, col="red")
> plot(exp(-x^2/2),lwd=6, col="red")
```





Demystifying the Normal Distribution

When the numerator in the brackets is zero (i.e. $x = \mu$, here: 0), this devolves to an exponent of 0.

- $exp(0) = 1$ (and, inversely, $log(1) = 0$).
- A logarithm of x for some base b is the value of the exponent that gets b to x.
  - $log_b(x) = a \quad \Longrightarrow \quad b^a = x$
- Notice how the top of the curve was at 1 in the exponentiated parabola.

Demystifying the Normal Density Function

With that in mind, it should be clear that $\frac{1}{\sqrt{2\pi\sigma^2}}$ (recall: $\sigma^2 = 1$ in our simple case) determines the height of the distribution.

Observe:

```
> 1/sqrt(2*pi)
[1] 0.3989423
> dnorm(0,0,1)
[1] 0.3989423
```

The height of the probability distribution for $x = 0$ when $\mu = 0$ and $\sigma^2 = 1$ is .3989423.

Demystifying the Normal Distribution

Notice: we talked about the height and shape of the probability distribution as a *function*. It does not communicate probabilities.

- The normal distribution is continuous. Thus, probability for any one value is basically 0.

That said, the area *under* the curve is the full domain and equals 1.

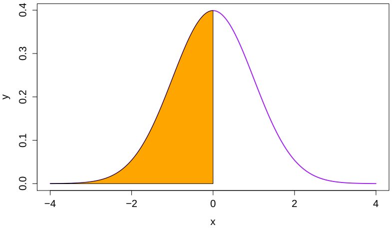
- The probability of selecting a number between two points on the x-axis equals the area under the curve *between* those two points.

Observe:

```
> pnorm(0, mean=0, sd=1)
[1] 0.5
```

Demystifying the Normal Distribution

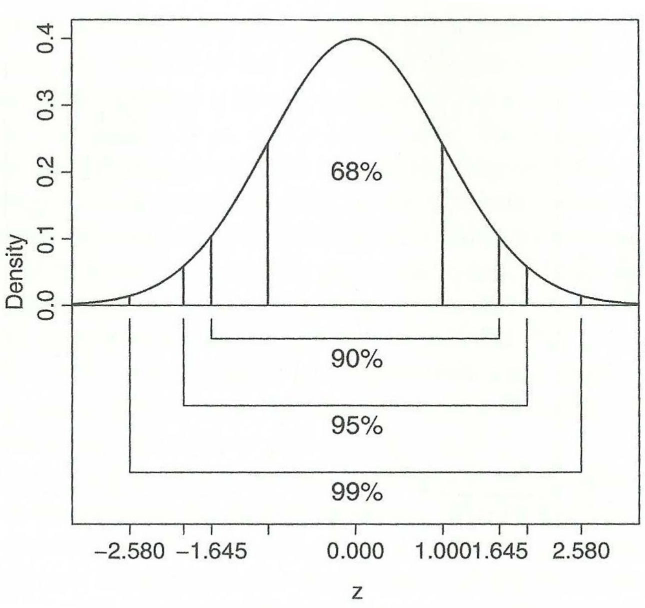
```
x=seq(-4,4,length=200)
y=dnorm(x, 0, 1)
plot(x,y,type="l", lwd=2, col="purple")
x=seq(-4,0,length=200)
y=dnorm(x, 0, 1)
polygon(c(-4,x,0),c(0,y,0),col="orange")
> pnorm(0, mean=0, sd=1)
[1] 0.5
```



68-90-95-99

```
> pnorm(1,mean=0,sd=1)-pnorm(-1,mean=0,sd=1)
[1] 0.6826895
> pnorm(1.645,mean=0,sd=1)-pnorm(-1.645,mean=0,sd=1)
[1] 0.9000302
> pnorm(1.96,mean=0,sd=1)-pnorm(-1.96,mean=0,sd=1)
[1] 0.9500042
> pnorm(2.58,mean=0,sd=1)-pnorm(-2.58,mean=0,sd=1)
[1] 0.99012
```

Areas under a Normal Curve



Conclusion

There are a lot of topics to digest in this lecture, all worth knowing.

- Probability and probability distributions are core components of the inferential statistics we'll be doing next.

Table of Contents