Probability Distributions and Functions

POSC 3410 – Quantitative Methods in Political Science

Steven V. Miller

Department of Political Science

CLEMSON UNIVERSITY

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.

$n = 0$:						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

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 *k*th 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.

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.

Binomial Mass Function

This is a combination issue.

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

How many ways can a series of coin flips land on 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.

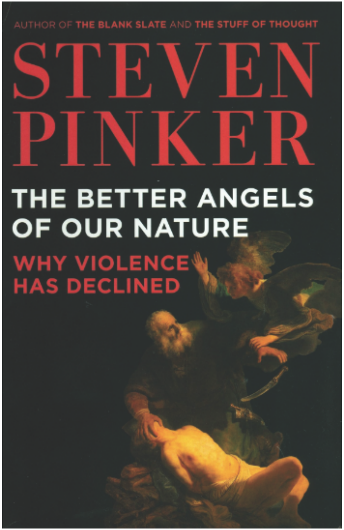
$$
\begin{aligned}
p(x = 5 \,|\, n = 10, p = .5) &= \binom{10}{5}(.5)^5(1-.5)^{10-5} \\
&= (252) * (.03125) * (.03125) \\
&= 0.2460938
\end{aligned} \tag{6}
$$

In R:

```
dbinom (5,10,.5)
```

```
## [1] 0.2460938
```

An Application: The Decline of War?



The Decline of War?

Pinker (2011) argues the absence of world wars since WW2 shows a decline of violence. But:

- This kind of war is fantastically rare.
- Gibler and Miller (Forthcoming) code 1,958 confrontations from 1816 to 2014.
- Of those: 84 are wars (p = .042)
- Of the wars, only 24 are wars we could think of as “really big” (p = .012)

The Decline of War?

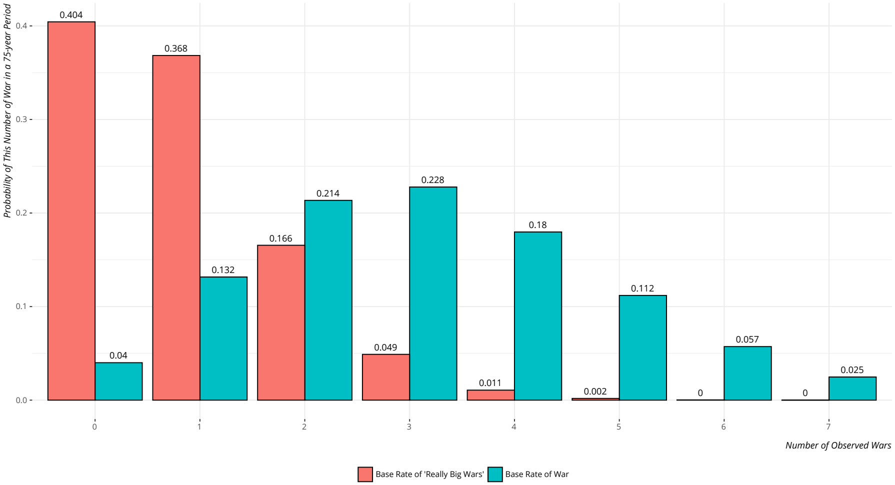
The year is 2022. We haven't observed a World War II in, basically, 75 years. What is the probability of us *not* observing this where:

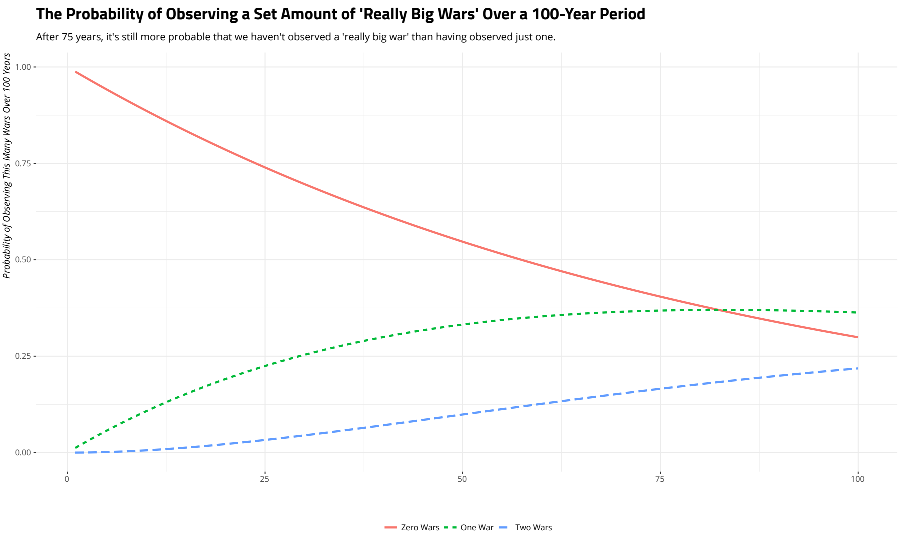
- p = .042, the overall base rate of war vs. not-war?
- p = .012, the overall base rate of a "really big war"?

```r
tibble(num_wars = seq(0:7)-1,
       base = dbinom(num_wars, 75, .042),
       rbw = dbinom(num_wars, 75, .012))

tibble(num_wars = rep(c(0, 1, 2), 100)) %>%
  arrange(num_wars) %>%
  mutate(period = rep(seq(1:100), 3),
         p = dbinom(num_wars, period, 0.012))
```

The Probability of the Number of (Observed) Wars in 75 Years, Given Assumed Rates of War

Knowing how rare 'really big wars' are, it's highly probable (p = .404) that we haven't observed one 75 years after WW2.




The Probability of Observing a Set Amount of 'Really Big Wars' Over a 100-Year Period
After 75 years, it's still more probable that we haven't observed a 'really big war' than having observed just one.
Probability of Observing This Many Wars Over 100 Years
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
Zero Wars
One War
Two Wars

Normal Functions

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.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e\{-\frac{(x-\mu)^2}{2\sigma^2}\}, \tag{7}$$

where: μ = the mean, σ^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.

Normal Density Function

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).

Demystifying the Normal Density Function

Here is our normal density function.

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

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

Demystifying the Normal Density Function

When μ = 0 and σ^2 = 1, the normal density function is a bit simpler.

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

Let's plot it next in R.

```
ggplot(data.frame(x = c(-4, 4)), aes(x)) +
  theme_steve_web() + # from stevemisc
  stat_function(fun = dnorm, color="#522D80", size=1.5)
```

A Simple Normal Density Function

The mu parameter determines the central tendency and sigma-squared parameter determines the width.

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.

R Code

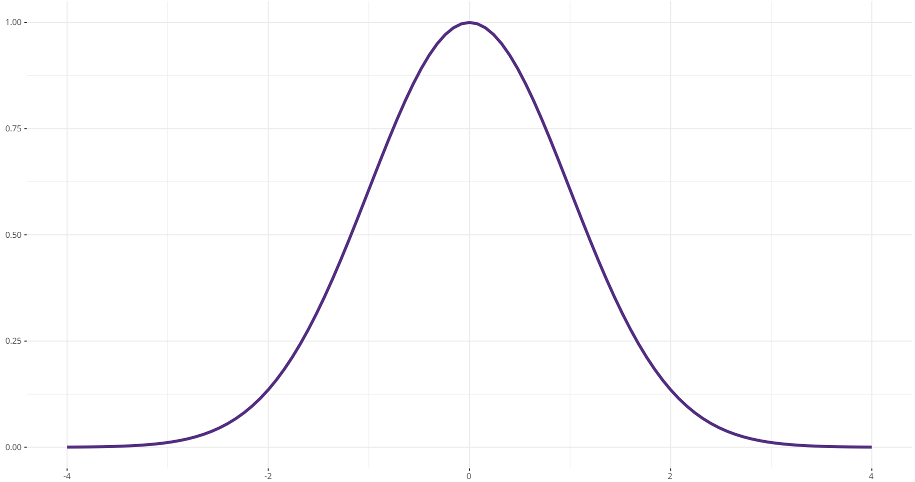
```
library(ggplot2)
parab <- function(x) {-x^2/2}
expparab <- function(x) {exp(-x^2/2)}

ggplot(data.frame(x = c(-4, 4)), aes(x)) +
  stat_function(fun = parab, color="#522d80", size=1.5) +
  theme_steve_web()

ggplot(data.frame(x = c(-4, 4)), aes(x)) +
  stat_function(fun = expparab, color="#522d80", size=1.5) +
  theme_steve_web()
```

A Basic Parabola

Notice the height is at 0 because the negative part flipped the parabola downward.

An Exponentiated Negative Parabola

Exponentiating squeezes the parabola, adjusts the height, and makes the tails asymptote to 0.



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*.
    - $log_b(x) = a \implies b^a = x$
- 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.

Demystifying the Normal Density Function

Observe:

```
1/sqrt(2*pi)
```

```
## [1] 0.3989423
```

```
dnorm(0,mean=0,sd=1)
```

```
## [1] 0.3989423
```

The height of the 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 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.

Demystifying the Normal Density Function

Observe:

```
pnorm(0, mean=0, sd=1)
```

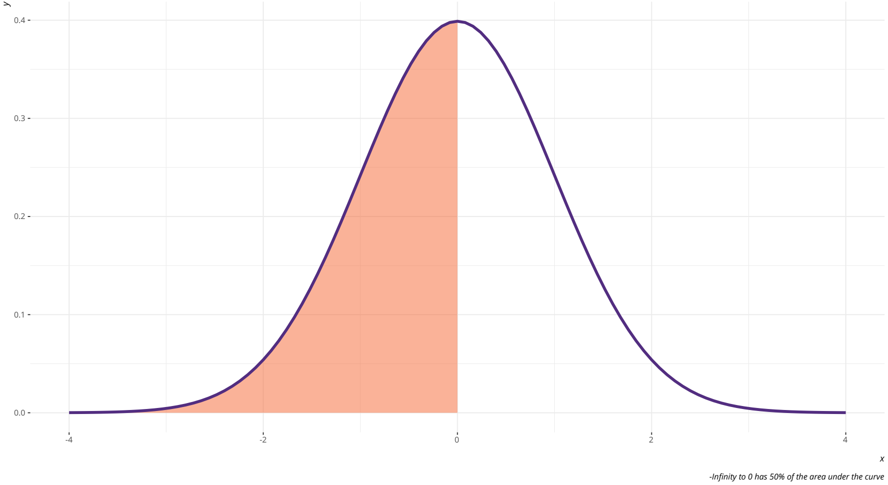
```
## [1] 0.5
```

Demystifying the Normal Distribution

```
ggplot(data.frame(x = c(-4, 4)), aes(x)) +
  theme_steve_web() +
  stat_function(fun = dnorm,
                xlim = c(-4,0),
                size=0,
                geom = "area", fill="#F66733", alpha=.5) +
    stat_function(fun = dnorm, color="#522D80", size=1.5)
```

A Standard Normal Distribution

Notice that half the distribution lies between negative infinity and 0.



68-90-95-99

```r
pnorm(1,mean=0,sd=1)-pnorm(-1,mean=0,sd=1)
```

```
## [1] 0.6826895
```

```r
pnorm(1.645,mean=0,sd=1)-pnorm(-1.645,mean=0,sd=1)
```

```
## [1] 0.9000302
```

```r
pnorm(1.96,mean=0,sd=1)-pnorm(-1.96,mean=0,sd=1)
```

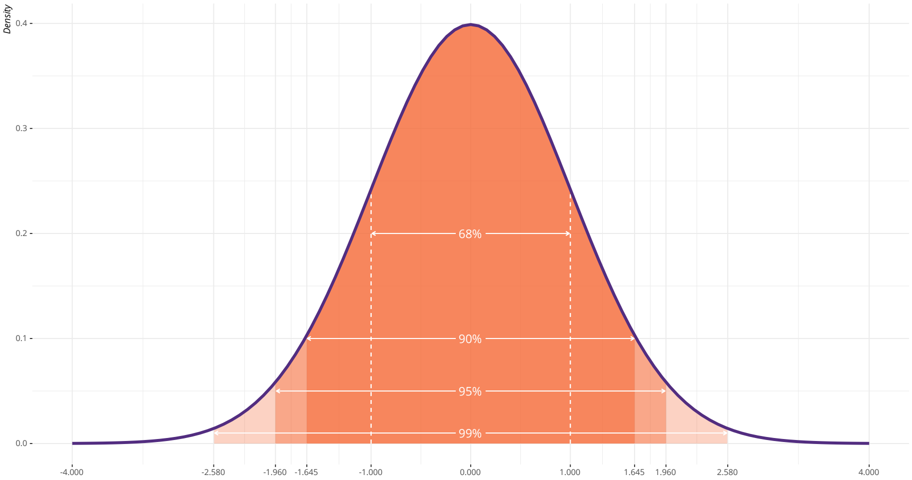
```
## [1] 0.9500042
```

```r
pnorm(2.58,mean=0,sd=1)-pnorm(-2.58,mean=0,sd=1)
```

```
## [1] 0.99012
```

The Area Underneath a Normal Distribution

The tails extend to infinity and are asymptote to zero, but the full domain sums to 1. 95% of all possible values are within about 1.96 standard units from the mean.



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