Class information

- Drew D. Creal
- Email: dcreal@chicagobooth.edu
- Office: 404 Harper Center
- Office hours: email me for an appointment
- Office phone: 773.834.5249

http://faculty.chicagobooth.edu/drew.creal/teaching/index.html
Course schedule

- Week # 1: Plotting and summarizing univariate data
- Week # 2: Plotting and summarizing bivariate data
- Week # 3: Probability 1
- Week # 4: Probability 2
- Week # 5: Probability 3
- Week # 6: In-class exam and Probability 4
- Week # 7: Statistical inference 1
- Week # 8: Statistical inference 2
- Week # 9: Simple linear regression
- Week # 10: Multiple linear regression
Outline of today’s topics

I. Standardization
II. Histograms and \textit{i.i.d.} draws
III. The Law of Large Numbers
IV. The Central Limit Theorem
Standardization
To standardize a random variable means to subtract the mean and divide by the standard deviation.

What does this do to the mean and variance?

Let $E[X] = \mu$ and $V[X] = \sigma^2$.

Then......

$Y = \frac{X - \mu}{\sigma} = \frac{X}{\sigma} - \frac{\mu}{\sigma}$

Our formulas for linear functions tell us that $E[Y] = 0$ and $V[Y] = 1$. 
Standardizing a numeric variable

- In many practical situations, it is also useful to standardize the data.
- To **standardize** a numeric variable means to subtract the sample mean and divide by the sample standard deviation.
- What are the sample mean and sample variance of the new variable?
Standardization

- **Standardizing** a random variable creates a new random variable with mean equal to zero and variance equal to 1.
- **Standardizing** a numeric variable in your dataset means to create a new variable with sample mean equal to zero and sample variance equal to 1.
- The new random variable is **unitless**.
- In both cases, the new variable can be interpreted as the number of standard deviations away from the mean.
- Let’s see an example!
Standardization: How unusual are some events?

Sometimes something weird or unusual happens and we want to quantify just how weird it is.

A typical example is a market crash.
How unusual is the crash?

1. The data up until the crash looks (approximately) normal.

2. Suppose we model it as: $N(0.03796, 0.7893)$.

3. The mean and variance were estimated using the data before the day of the crash.

4. The return on the day of the crash: -20.69%
The crash return was way out in the left tail.

We want to know the probability of this crash assuming the data is normal. To do this we will “standardize” the data.

We want to ask: if the value were from a standard normal, what would it be?
Standardization

We can think of our returns as:

\[ R_t = 0.03796 + \sqrt{0.7893}Z_t \]

\[ Z_t \sim N(0, 1) \]

The value \( Z_t \) corresponding to a generic \( R_t \) value is:

\[ Z_t = \frac{R_t - \mu}{\sigma} = \frac{R_t - 0.03796}{\sqrt{0.7893}} \]

The values of \( Z_t \) for \( t = 1, \ldots, T \) should look standard normal. Why?
So, how unusual is the crash return?

\[ Z_t = \frac{-20.69 - 0.03796}{\sqrt{0.7893}} = -21.84 \]

Its “z-value” is -21.84. It is like drawing a value of -21.84 from the standard normal. **No way!**

Plotted are the z-values for the previous months.
For $X \sim \mathcal{N}(\mu, \sigma^2)$, the **z value** corresponding to a value $x$ is

$$Z = \frac{x - \mu}{\sigma}.$$  

Any time someone says “z-value” or “z score,” they are just talking about how many standard deviations we are away from the mean under a bell curve.
Suppose a return is distributed $R \sim N(0.01, 0.04^2)$. What is the probability of a return between 0 and 0.05?

In lecture #5, we calculated this as:

$$P(0 < R < 0.05) = F_R(0.05) - F_R(0)$$

$$= 0.8413 - 0.4013$$

$$= 0.44$$

where we used “$=\text{NORMDIST}(0.0,0.01,0.04,\text{TRUE}) = 0.4013$”
Standardization

For $X \sim N(\mu, \sigma^2)$,

$$Pr(a < X < b) = Pr\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right).$$

When $Z \sim N(0, 1)$

- For a normal r. v., we can always calculate the probability of an interval $(a, b)$ by transforming the interval to $(\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma})$ and comparing it to a standard normal r.v..

- Before computers were common, we looked up probabilities in the tables at the back of a stats book!
An alternative way to do this is to first standardize the values 0 and 0.05.

This is equivalent to $Z$ being between $\frac{0-0.01}{0.04} = -0.25$ and $\frac{(0.05-0.01)}{0.04} = 1$. Using the normal CDF in Excel,

\[
\begin{align*}
&= \text{NORMDIST}(-0.25, 0, 1, \text{TRUE}) = 0.4013 \\
&= \text{NORMDIST}(1, 0, 1, \text{TRUE}) = 0.841 \\
\end{align*}
\]

$Pr(0 < R < 0.05) = Pr(-0.25 < Z < 1) = 0.84 - 0.4 = 0.44$
Histograms and IID draws
TOUR OF ACCOUNTING

OVER HERE WE HAVE OUR RANDOM NUMBER GENERATOR.

NINE NINE NINE NINE NINE

ARE YOU SURE THAT'S RANDOM?

THAT'S THE PROBLEM WITH RANDOMNESS: YOU CAN NEVER BE SURE.
Here is a histogram of 1000 draws from the standard normal distribution, i.e. $Z \sim N(0, 1)$.

The height of each bar tells us the number of observations in each interval.

All the intervals have the same width.
Histograms and IID Draws

If we divide the height of each bar by the width times 1000 the picture looks the same, but now the area of each bar equals the % of observations in the interval.

This is just a fancy way of “scaling” the histogram so that the total area of all the bars equals 1. It looks the same, but the vertical scale is different.
Histograms and IID Draws

For a large number of *i.i.d* draws, the observed percent in an interval should be close to the probability.

Note two things:

1. For the pdf, the area is the probability of the interval.
2. In the histogram, the area is the observed percent in the interval.
Histories and IID Draws

As the number of draws gets larger, the histogram gets closer to the pdf! It looks like a bell curve.

$n = 100$

$n = 500$

$n = 2000$

$n = 1 \text{ million}$
The (normalized) histogram of a “large” number of \( i.i.d. \) draws from any continuous distribution should look like the p.d.f.
Here is another example for uniform random variables $X \sim U(2, 5)$. 

- $n = 100$
- $n = 500$
- $n = 2000$
- $n = 1$ million
Histograms and IID Draws

Here is another example from a random variable with a skewed distribution.

$n = 100$

$n = 500$

$n = 2000$

$n = 1$ million
Can we use this to do probability calculations?.....YES!

For example, suppose $Z \sim N(0, 1)$ and we want to know $P(Z < -1.5)$.

**Step 1.** Using Excel, simulate 1,000 *i.i.d.* draws from the standard normal distribution.

**Step 2.** Determine the percentage of these draws that are less than -1.5.

**Step 3.** This is (approximately) the probability we are looking for.

(NOTE: This is also true for discrete random variables. And, the approximation gets better the larger the number of draws.)
The Law of Large Numbers
The Law of Large Numbers

- In lecture # 3, we learned that one possible interpretation of probability is “long run frequency.”
- In other words, if we were to repeat a random experiment over and over and over again, the probability of an event happening is the frequency that it happens after a large number of identical experiments.
The Law of Large Numbers

- Consider tossing a fair coin repeatedly.
- Let $Y_i = 1$ if the toss is a head and zero otherwise on the $i$-th toss.
- Let $X_1 = Y_1$.
- Let $X_2 = \frac{1}{2} (Y_1 + Y_2)$.
- Let $X_3 = \frac{1}{3} (Y_1 + Y_2 + Y_3)$.
  
  \[ \vdots \]

\[ \vdots \]

- Let $X_n = \frac{1}{n} (Y_1 + Y_2 + Y_3 + \ldots + Y_n) = \frac{1}{n} \sum_{i=1}^{n} Y_i$. 
Remember that \( X_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \). This is the plot of \( X_j \) for \( j = 1, \ldots, 5000 \).

Notice that the sample mean \( X_n \) is getting closer to the true mean \( p = 0.5 \) as we increase \( n \).
The Law of Large Numbers

As $n$ becomes large

$$E[X] \approx \frac{1}{n} \sum_{i=1}^{n} x_i$$

where the $x_i$'s are outcomes from $i.i.d.$ draws all having the same distribution as $X$.

This is an example of the Law of Large Numbers.
Law of Large Numbers: why it works

Remember our example from Lecture #3 where we tossed two coins. Let $X$ equal the number of heads in two tosses.

Suppose we toss two coins ten times. Each time we record the number of heads:

$$1 \ 0 \ 2 \ 1 \ 0 \ 1 \ 2 \ 0 \ 2 \ 0$$

Question: what is the average value?

$$\bar{x} = \frac{(4 \times 0 + 3 \times 1 + 3 \times 2)}{10} = 0.9$$
Law of Large Numbers: why it works

Now suppose we toss two coins 1000 times. What is the sample mean?

| 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 0 | 2 | 1 | 1 | 2 | 1 | 2 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 | 1 | 0 | 2 | 1 | 1 | 0 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 |
| 0 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 |
| 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 2 | 2 | 0 | 1 | 1 | 0 | 1 |
| 2 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 0 | 2 | 1 | 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 0 | 2 | 2 | 0 | 0 | 1 | 0 | 2 | 2 | 2 | 1 | 1 | 0 | 1 | 1 |
| 1 | 0 | 2 | 2 | 0 | 1 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 2 | 1 | 2 | 1 | 1 |
| 0 | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| 1 | 0 | 2 | 1 | 0 | 1 | 0 | 1 | 1 | 2 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 | 1 | 1 | 2 | 1 | 0 | 0 | 1 | 0 | 2 | 1 | 1 | 2 | 1 | 1 |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| 2 | 2 | 0 | 1 | 1 | 0 | 2 | 0 | 1 | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 2 | 2 | 1 | 1 | .................. |
Law of Large Numbers: why it works

Well, of course we can just have the computer figure it out, but let us think about this for a minute.

What should the mean be?

Let $n_0$, $n_1$, and $n_2$ be the number of 0’s, 1’s, and 2’s, respectively.

Then, the average would be:

$$\frac{n_0}{n} \times 0 + \frac{n_1}{n} \times 1 + \frac{n_2}{n} \times 2$$

This appears similar to the expectation $E[X]$ but we’re just weighting each outcome by their “frequencies” instead of the probabilities!
Law of Large Numbers: why it works

\[ \frac{n_0}{n} \times 0 + \frac{n_1}{n} \times 1 + \frac{n_2}{n} \times 2 \]

Now note that the possible outcomes of each experiment are

\textit{i.i.d.} draws from the discrete distribution:

\[
\begin{array}{c|c}
 x & P(x) \\
\hline
 0 & 0.25 \\
 1 & 0.50 \\
 2 & 0.25 \\
\end{array}
\]
Law of Large Numbers: why it works

As the number of draws $n$ gets larger, we should have:

$$\frac{n_0}{n} \approx 0.25 \quad \frac{n_1}{n} \approx 0.5 \quad \frac{n_2}{n} \approx 0.25$$

Hence, the average should be about:

$$0.25 \times 0 + 0.5 \times 1 + 0.25 \times 2 = 1$$

but, this is just the expected value of the random variable $X$. 
The actual sample mean is from the 1000 tosses was:

\[ \bar{x} = 1.0110 \]

Hence, with a very, very, ...large number of tosses we would expect the sample mean to be very close to 1 (the expected value).

To summarize, we can think of the expected value, which in this case is equal to:

\[ p_X(0) \times 0 + p_X(1) \times 1 + p_X(2) \times 2 = 1 \]

as the long run average of i.i.d. draws.
The Law of Large Numbers

The Law of Large Numbers is also true for functions $f(\cdot)$ of $X$.

$$E[f(X)] \approx \frac{1}{n} \sum_{i=1}^{n} f(x_i)$$

**Example:** Consider the function $f(x) = (x - \mu)^2$.

$$V[X] = E[f(X)] \approx \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

This implies that we can use the sample variance $s_x^2$ as an approximation of the true variance!
The Law of Large Numbers

Example: Let’s return to the example where we tossed 2 coins 1000 times.

The sample mean from the 1000 tosses was: \( \bar{x} = 1.0110 \)

The sample variance from the 1000 tosses was: \( s_x^2 = 0.51 \)

If \( X \) is the number of heads out of two coin tosses:

\[
V[X] = 0.25 \times (0 - 1)^2 + 0.55 \times (1 - 1)^2 + 0.25 \times (2 - 1)^2
\]
\[
= 0.5
\]
The Law of Large Numbers

Thus, for “large samples” the sample quantities that we can compute from our observed data should be similar to the quantities we talked about for random variables:

$$V[X] \approx \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\approx \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

This is true if we are taking i.i.d. draws!
The Central Limit Theorem
The Central Limit Theorem

The *central limit theorem (CLT)* says that the average of a large number of independent random variables is (approximately) normally distributed.

Another way of saying this is:

Suppose that $X_1, X_2, \ldots, X_n$ are *i.i.d.* random variables and let $Y = \frac{X_1 + X_2 + \ldots + X_n}{n}$. As $n$ gets large

$$Y \sim N(\mu_Y, \sigma_Y^2)$$
The Central Limit Theorem

What is so special about this?

Notice that although we did assume that the $X_i$’s are i.i.d., we DID NOT say what distribution they have.

That’s right! The CLT says:

The average of a large number of independent random variables is (approximately) normally distributed, no matter what distribution the individual random variables have!
The Central Limit Theorem

Example: Consider the binomial distribution. Define

\[ Y = X_1 + X_2 + \ldots + X_n \]

where \( X_i \sim \text{Bernoulli}(p) \) i.i.d.
The Central Limit Theorem

1. As we increase \( n \), the distribution of \( Y \) gets closer and closer to a normal distribution with the same mean and variance as the binomial.

2. In the graph on the right, I have plotted the binomial distribution (blue) on top of the normal distribution (red) with \( p = 0.2 \) and \( n = 100 \).
Your company is about to manufacture 100 parts.

Suppose defects are i.i.d. $X_i \sim \text{Bernoulli}(0.1)$.

Let $Y = X_1 + X_2 + \ldots + X_{100}$ be the number of defects.

$Y \sim \text{binomial}(100, 0.1)$.

$$
\begin{align*}
\mathbb{E}[Y] &= n \times p = 100 \times 0.1 = 10 \\
\sigma_Y &= \sqrt{n \times p \times (1 - p)} = \sqrt{100 \times 0.1 \times 0.9} = 3
\end{align*}
$$
How good is the approximation?

Even though $Y \sim \text{binomial}(100, 0.1)$, let us use the normal approximation, first.

Let the normal distribution have the same mean and variance.

$$Y \approx N(10, 9)$$

Based on the normal approximation, there is a 95% chance that the number of defects is in the interval:

$$(\mu - 2\sigma_Y, \mu + 2\sigma_Y) = 10 \pm 6 = (4, 16)$$
Example:

We can compare that to the exact answer based on the binomial probabilities.

What is the correct binomial probability of obtaining between 4 and 16 defective parts? If the normal approximation is good, the exact number should be close to .95. Let us see if this is the case...

\[
P(4 < Y < 16) = F(16) - F(4) = 0.9794 - 0.0237 = 0.9557
\]

The normal approximation appears to be pretty good.