Statistics Review Part 2

Distributions, Sampling, Estimation

Review: Expected Values

- Think of the **expected value** (or **mean**) of a RV as the long-run average value of the RV over many repeated trials
- You can also think of it as a measure of the "middle" of a probability distribution, or a "good guess" of the value of a RV
- Denoted $E(X)$ or μ_X
- More precisely, $E(X)$ is a *probability-weighted average of all possible outcomes of X*
- Example: rolling a die
  - $f(1) = f(2) = f(3) = f(4) = f(5) = f(6) = 1/6$
  - $E(X) = 1*(1/6) + 2*(1/6) + 3*(1/6) + 4*(1/6) + 5*(1/6) + 6*(1/6)$
    $= 1/6 + 2/6 + 3/6 + 4/6 + 5/6 + 6/6$
    $= 21/6 = 3.5$
- interpretation?

Review: More about E(X)

- The general case for a discrete RV
  - Suppose RV X can take k possible values $x_1, x_2, \ldots, x_k$ with associated probabilities $p_1, p_2, \ldots, p_k$ then

$$E(X) = \sum_{i=1}^{k} p_i x_i$$

- The general case for a continuous RV involves an integral
- $E(X)$ is a "mathematical operator" (like +, -, *, /).
  - It is a **linear** operator, which means we can pass it through addition and subtraction operators
  - That is, if a and b are constants and X is a RV,

$$E(a + bX) = a + bE(X)$$

Review:Conditional Distributions

- The distribution of a random variable Y conditional on another random variable X taking a specific value is called the **conditional distribution of Y given X**.
- The conditional probability that Y takes value y when X takes value x is written $\Pr(Y = y \mid X = x)$.
- In general,

$$\Pr(Y = y \mid X = x) = \frac{\Pr(Y = y, X = x)}{\Pr(X = x)}$$

- Intuitively, this measures the probability that $Y = y$ and $X=x$, **given that** $X = x$.

Review: Conditional Expectation

- The mean of the conditional distribution of *Y* given *X* is called the **conditional expectation** (or **conditional mean**) **of *Y* given *X***.
- It's the expected value of *Y*, given that *X* takes a particular value.
- It's computed just like a regular (unconditional) expectation, but uses the conditional distribution instead of the marginal.
  - If *Y* takes one of *k* possible values $y_1, y_2, \ldots, y_k$ then:

$$E(Y \mid X = x) = \sum_{i=1}^{k} y_i \Pr(Y = y_i \mid X = x)$$

Review: Table of Means

Immigrant Cohort	N(WAGES)	mean(WAGES)
Before 1950	66	27561
1950s	573	30682
1960s	1,260	31760
1970s	2,679	33365
1980s	2,416	27478
1990s	5,412	20283
2000s	2,463	17103
Cdn-Born	17,730	38317
Temporary	608	21737

This is a Table of Means. But, we can interpret it as an (estimate of) a table of conditional expectations.

What is *Y* in the conditional expectation formula? What is *X*?

What are the probabilities: $\Pr(Y = y_i \mid X = x)$

What are we summing over?

Review: Uniform Distribution

- uniform distribution is completely characterised by two parameters: *a,b*
- if *X~U(a,b)*, ("the r.v. *X* is uniformly distributed between *a* and *b*") then
  - *f(x)=1/(b-a)* and *F(x)=(x-a)/(b-a)*
  - special case: if *a=0* and *b=1* gives the "standard uniform"
    - *f(x)=1* and *F(x)=x*
- lots of things are uniform:
  - values of a roll of a single die; probability of rain falling on a particular part of the sidewalk;
- as with any distribution, P[y<x<z]=F(z)-F(y)
  - P[y<x<z]=*(z-a)/(b-a)-(y-a)/(b-a)=(z-y)/(b-a)*
  - draw pictures (tails, range), do calculations

Some Useful Probability Distributions

- There are four important probability distributions that we'll encounter repeatedly:
  - The **Normal distribution**
  - The **Chi-square distribution**
  - Student's **t distribution**
  - Snedecor's **F distribution**
  - the normal is the basis of all of these: the last 3 are derived from the first
- Why are these important?
  - Most theory regarding the classical linear regression model (CLRM) is developed in the context of the normal distribution. Doing so gives us exact results (you'll see what this means soon enough!)
  - When we get away from the exact distributional assumptions of the CLRM, we use large sample approximations. We know from the **central limit theorem** (remember this?) that many statistics have an approximately Normal distribution as the sample size gets large.
  - Consequently, test statistics that we care about turn out to have Normal, Chi-square, t, or F sampling distributions.

Why do Things get Normal?

- **Central Limit Theorems** typically say that if you add up enough random variables from non-normal distributions, their sum (or average) looks pretty much like a normal distribution.
  - Uniform random variables (like a single die) are not normal---there is no hump in the pdf.
  - But the sum of 2 dice has a point (its pdf looks like a triangle).
  - the sum of 2 identical continuous uniforms is *triangular* (its pdf is a triangle).
  - The sum of 3 dice has a hump (derive it).

Review: One Die, pdf and cdf

- The pdf for one die is uniform.

		Outcome (value of roll of single die)				
		2	**3**	**4**	**5**	**6**
pdf	1/6	1/6	1/6	1/6	1/6	1/6
cdf	1/6	1/3	1/2	2/3	5/6	1

Review: Sum of Two Dice

pdf and cdf, *in 36ths*

	2	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
pdf	1	2	3	4	5	6	5	4	3	2	1
cdf	1	3	6	10	15	21	26	30	33	35	36

The Normal Distribution

- A continuous RV with a normal distribution has a bell-shaped pdf.
  - It is symmetric around its mean.
  - It is **completely** characterized by two parameters: its mean (μ) and variance (σ^2).
  - 95% of its probability density lies between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$
  - (draw a picture)
- Usual notation is $N(\mu, \sigma^2)$.
  - To say "*X* is Normally distributed with mean μ and variance σ^2" we write $X \sim N(\mu, \sigma^2)$
- A special case is the **standard Normal** distribution, where $\mu = 0$ and $\sigma^2 = 1$, denoted $N(0,1)$.
  - Usual notation for the standard normal cdf is $\Pr(Z \leq c) = \Phi(c)$

More About the Normal Distribution

- Useful result 1: If $X \sim N(\mu, \sigma^2)$ then $a + bX \sim N(a + b\mu, b^2\sigma^2)$
- This implies that if $X \sim N(\mu, \sigma^2)$, we can **standardize** X by subtracting off the mean and dividing by the standard deviation: $Z = (X - \mu)/\sigma$.
  - After standardizing, $Z \sim N(0,1)$
- This is useful for computing probabilities. If $X \sim N(\mu, \sigma^2)$, Z is as above, c_1 and c_2 are constants, $d_1 = (c_1 - \mu)/\sigma$ and $d_2 = (c_2 - \mu)/\sigma$ then
  - $\Pr(X \leq c_1) = \Pr(Z \leq d_1) = \Phi(d_1)$
  - $\Pr(X \geq c_2) = \Pr(Z \geq d_2) = 1 - \Pr(Z \leq d_2) = 1 - \Phi(d_2)$
  - $\Pr(c_2 \leq X \leq c_1) = \Pr(d_2 \leq Z \leq d_1) = \Phi(d_1) - \Phi(d_2)$
- We can look up these probabilities in tables, e.g. Table B-7
- Useful result 2: If $X_1, X_2, \ldots, X_n$ are normally distributed RVs, then their sum (and any weighted sum) is also normally distributed.

The Chi-square distribution

- As we'll see soon enough, many important test statistics have a Chi-square distribution.
  - It is defined by a single parameter: the **degrees of freedom**, denoted *v*.
  - It is not symmetric – it is **positively skewed**, which means it has a very long tail in the positive direction – very large positive values can occur, though not "too often"
  - A RV with a Chi-square distribution takes positive values only.
  - (draw a picture)
- Standard notation: χ^2_v
- Its definition is based on the Normal distribution:
  - if $Z \sim N(0,1)$, then $Z^2 \sim \chi^2_1$.
- Furthermore, if X_1 and X_2 are **independent** χ^2_1 RVs, then $X_1 + X_2 \sim \chi^2_2$
- Likewise, if we add *v* independent χ^2_1 RVs, their sum is distributed χ^2_v

t Distribution

- A very important test statistic -- called the "*t* statistic" (not a coincidence) -- has a probability distribution called **Student's *t* distribution** (or simply a ***t* distribution**).
  - It is defined by a single parameter: the degrees of freedom *v*.
  - The *t* distribution is very similar to the Normal, but with slightly thicker tails.
  - As *v* gets large, the *t* distribution approaches the Normal.
- Standard notation: t_v
- Its definition is based on the Normal and Chi-squared distributions:
  - If $Z \sim N(0,1)$, $X \sim \chi^2_v$, and Z and X are **independent**, then

$$\boxed{\frac{Z}{\sqrt{X/v}} \sim t_v}$$

F distribution

- The (Snedecor's) ***F* distribution** is another derived distribution that is very important for inference.
  - "*F* test" statistics have an *F* distribution
  - Like the Chi-square, RVs with an *F* distribution take positive values only & the distribution is positively skewed
  - It is defined by two degree of freedom parameters: *v1* and *v2*
- Standard notation: $F_{v1,v2}$
- Its definition is based on the Chi-square:
  - If X_1 and X_2 are **independent** Chi-square RVs with *v1* and *v2* degrees of freedom, respectively, then

$$\boxed{\frac{X_1 / v1}{X_2 / v2} \sim F_{v1,v2}}$$

Learn About the Population Using a Sample

- Our objective as econometricians is to learn something about a **population** of interest. This is called **inference.**
- The population can be **almost any** group of people, businesses, plants, animals, electrons, etc. that we are interested in, e.g.,
  - all Canadian adults
  - all firms (businesses)
  - all publicly-traded firms
  - the thirty largest firms traded on the New York Stock Exchange (i.e.,the DJIA)
- Exactly **what** we hope to learn depends on the specific question we hope to answer.
  - what is the average labour income in Canada? What is its variance? What proportion of Canadian adults earn over $100,000 per year?
  - what is the relationship between educational attainment and income?
  - what is the elasticity of demand for product X?
  - what is the probability that the price of stock X will increase in the next year?
  - what is the expected change in the price stock X over the next year?
  - what is the expected price of stock X one year from now if the price of oil increases to $75/barrel?
- **In general, we want to learn *something about the probability distribution of a variable of interest,* or about *the joint distribution of a group of variables*.**

Sampling

- In theory, we could measure the quantity we care about using the whole population.
- But we almost never do, because it's expensive ($, time, etc.)
- e.g., a VERY expensive way to measure the average income of Canadians is to contact every one of them and ask them how much they earn.
- StatCan *almost* does this in the Census of Population
  - Every household gets a census form to complete
  - 4/5 get "short forms" – these only record who lives at that address, their age, sex, marital status, and official language knowledge
  - 1/5 get "long forms" – in addition, these ask how much you earn, what industry you work in, etc. (unless proposed changes are adopted!)
- Collecting this information from 1/5 of the population is so expensive we only do it every 5 years
- But this illustrates the basic idea: rather than contacting everyone in the population, a cheaper alternative is to contact a **small**, representative group of individuals and ask them how much they earn
- this group is called a **SAMPLE**

Populations and Samples

- ECONOMETRIC INFERENCE ABOUT A POPULATION IS **ALMOST ALWAYS** BASED ON A SAMPLE!
- How do we choose which population members to sample?
- In a nutshell: choose them **randomly**.
- Example: Suppose I'm interested in the probability distribution of my commuting time to campus. Rather than recording my commuting time *every day*, I could randomly select five days each month to record my commuting time.
  - Population: every day
  - Sample: the days I record my commuting time
  - Use the sample data to estimate the population mean, variance, etc.
- Example: Political pollsters try to predict election outcomes. They ask questions like "If there was an election today, which of these candidates would you vote for?" Rather than asking *everyone in the country*, they randomly select a group of individuals to answer the question.
  - Population: everyone in the country
  - Sample: the group selected to answer the question
  - Use the sample to estimate the population mean, variance, etc.

Random Sampling

- How is a random sample selected?
- The easiest way is a **Simple Random Sample (SRS):** randomly choose *n* members of the population, each member of the population is equally likely to be selected. (like drawing names out of a hat)
- Most surveys are actually NOT simple random samples.
  - in a small sample, small groups may not be represented
    - e.g., in a SRS of 1000 Canadians, you are very unlikely to select anyone from PEI because not many people live there ... but the population you care about is "all Canadians." Consequently many surveys **oversample** small groups (e.g., minorities) to ensure the sample includes all subgroups of interest.
  - usually a SRS is more expensive than a **cluster sample**
    - if you're going door-to-door with surveys, it's cheapest to survey people/businesses that are close together. In a SRS of 1000 Canadians, they're likely to be spread out all over the place. So an alternative is to randomly sample some cities/towns, and then randomly sample some streets/blocks in those towns, and then survey everyone on that street/block
- These kind of samples are common in practice and a little harder to work with than a SRS.

Sampled Objects are Random Variables

- Suppose we're interested in a variable X.
- We're going to select a sample of individuals/businesses or whatever and measure their value of X.
- The observed measurements of X that comprise our sample are called **observations**. All the observations together are our **data**.
- Usually, we denote the n observations in the sample $X_1, X_2, \ldots, X_n$
  - If X was annual earnings, X_1 is the first person's response, X_2 is the second, etc
- Because we randomly select objects into the sample, the **values** of the observations $X_1, X_2, \ldots, X_n$ **are random.**
  - We don't know what values of X we'll get in advance
  - If we had chosen different members of the population, their values of X would be different.
- Thus, given random sampling, we treat $X_1, X_2, \ldots, X_n$ as random variables.

iid Sampling

- In this class we'll assume a mathematically convenient kind of sample
- Suppose we care about some random variable X.
- Assume that the distribution of X, i.e., $f(X)$ is the same for **all** members of the population.
- Suppose we select a sample of people/businesses (or **respondents**, in general) of size n, and record their values of X.
- Thus our sample is $X_1, X_2, \ldots, X_n$
- Because each $X_1, X_2, \ldots, X_n$ comes from the same population distribution $f(X)$, each X_i has the same marginal distribution: also $f(X)$.
  - This is why we can use the sample to learn about the population.
- Because the X_i all have the same marginal distribution, we say they are **identically distributed**.
- Suppose further that the observations are drawn **independently** of one another
  - Knowing X_1 gives no information about X_2,or X_3 etc.
- Because the $X_1, X_2, \ldots, X_n$ are sample from the same population distribution and independently of one another, we say they are **independently and identically distributed**, or **iid**

Statistics and Sampling Distributions

- A **statistic** is any function of the sample data.
  - A (scalar-valued) **function** *f(x1,…xN)* is a single number associated with each set of values that x1,…,xN can take on.
- Because the sample data are random variables, so are statistics.
- We know that all random variables have probability distributions.

  ➔**All statistics have probability distributions (pdfs&cdfs)**.
- In fact we have a special name for the probability distribution of a statistic: we call it a **SAMPLING DISTRIBUTION.**
- **THIS IS THE MOST IMPORTANT CONCEPT IN THIS COURSE!!!**
- Every statistic has a sampling distribution because **if we drew a different sample, the data would take different values, and hence so would the statistic.**
- The sampling distribution represents **uncertainty** about the **population value** of the statistic because it is based on a sample, and not based on the whole population.

What the Sampling Distribution Tells Us

- Like any probability distribution, the sampling distribution tells us what values of the statistic are possible, and how likely the different values are.
- For instance, the **mean of the sampling distribution** tells us the expected value of the statistic.
  - It is a good measure of what value we expect the statistic to take.
  - It also tells us where the statistic's probability distribution is centered.
- The **variance of the sampling distribution** tells us how "spread out" the distribution of the statistic is.
  - It is usually a function of the sample size.
  - It has a special name: the **sampling variance** of the statistic (note: this is NOT THE SAME AS THE **SAMPLE VARIANCE**!)
  - If the sampling variance is large, then it is **likely** that the statistic takes a value "far" from the mean of the sampling distribution.
  - If the sampling variance is small, then it is **unlikely** that the statistic takes a value "far" from the mean of the sampling distribution.
  - Usually, the sampling variance gets smaller as the sample size gets bigger.
- A picture shows this.

Some Statistics You Need to Know From BUEC 232

- Suppose we draw an iid sample of n observations, $X_1, X_2, \ldots, X_n$, from a population.
- The **sample mean** is:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  - it is a "good" estimate of the population mean μ.
- The **sample variance** is:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

  - it is a "good" estimate of the population variance σ^2.
  - the **sample standard deviation** is $s = \sqrt{s^2}$
- The **sample covariance** is:

$$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$$

  - it is a "good" estimate of the population covariance σ_{XY}
  - the **sample correlation is**:

$$r_{XY} = s_{XY} / s_X s_Y$$

Estimation

- An **estimator** is a statistic that is used to infer the value of an unknown quantity in a statistical model
- The sample mean, sample variance, and sample covariance are all statistics. But, they are also all called **estimators**, because they can be used to **estimate** population quantities.
- That is, the thing we care about is a population quantity like the population mean μ.
- We don't get to observe μ directly, and we can't measure its value in the population.
- So we draw a sample from the population, and **estimate** μ using the sample.
- One way to do this is to compute the **sample mean** $\bar{X}$ in our sample.
- It is a "good" estimate of the population mean, in a sense we'll now make precise.

Estimators and Their Properties: Bias

- There are lots and lots of estimators, but not all are equally "good."
  - The sample mean is an estimator of the population mean.
  - So is the median.
  - So is the value of one randomly selected observation.
- This is where the estimator's sampling distribution comes in – it tells us the estimator's properties.
  - Whether it gives "good" or "bad" estimates of a population quantity.
- Suppose we're interested in a population quantity Q and R is a sample statistic that we use to estimate Q.
  - e.g., Q might be the population mean, and R the sample mean
- We say R is an **unbiased estimator of Q** if $E(R) = Q$.

  ➔**if R is an unbiased estimator of Q, then Q is the mean of the sampling distribution of R**
- The **bias** of R is $E(R) - Q$. An **unbiased** estimator has bias = 0.
- DRAW A PICTURE!!

Estimators and Their Properties: Efficiency

- Unbiasedness is a nice property, but it is "weak."
  - There can be many unbiased estimators of a given population quantity.
  - Example: suppose we want to estimate the population mean μ. In an iid sample, the sample mean is an unbiased estimator of μ:

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu$$

  - because $E(X_i) = \mu$ for every observation.
  - Another unbiased estimator is the value of X_1, because $E(X_1) = \mu$.
- How do we choose between unbiased estimators?
  - We prefer the unbiased estimator with the smaller sampling variance. A picture shows the how the sampling distributions of the sample mean and a single observation's value differ.
  - Suppose we have two unbiased estimators of Q, call them R_1 and R_2. We say that R_1 is **more efficient** than R_2 if $Var(R_1) < Var(R_2)$.

Sampling Distribution of the Sample Mean

- Suppose $X_1, X_2, \ldots, X_n$ are an iid random sample of size n from a population with mean μ and variance σ^2.
- The sample mean is unbiased (we showed this already): $\boxed{E\left(\overline{X}\right)=\mu}$
- The **variance of the sampling distribution of the sample mean** (which we also call the **sampling variance of the sample mean**) is σ^2/n:

$$\boxed{\begin{aligned} Var\left(\overline{X}\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}Var\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\left[\sum_{i=1}^{n}Var\left(X_i\right)+2\sum_{i=1}^{n}\sum_{j\neq i}Cov\left(X_i,X_j\right)\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}Var\left(X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n} \end{aligned}}$$

- In fact, if $X_1, X_2, \ldots, X_n$ are iid draws from the $N(\mu, \sigma^2)$ distribution, then:

$$\boxed{\overline{X}\sim N\left(\mu,\sigma^2/n\right)}$$

  - Why? We already know the mean and variance of the sampling distribution. And we know the sampling distribution is normal because the sample mean is just a linear combinations of a bunch of $N(\mu, \sigma^2)$ random variables ... and we also know that linear combinations of normal RVs are also normally distributed (lecture 4).

The Sample Variance is Unbiased

$$E\left(s^2\right)=E\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)$$

$$=\frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i^2-2X_i\overline{X}+\overline{X}^2\right)\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-\sum_{i=1}^{n}2X_i\overline{X}+\sum_{i=1}^{n}\overline{X}^2\right)$$

$$=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-2\overline{X}\sum_{i=1}^{n}X_i+n\overline{X}^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-2n\overline{X}^2+n\overline{X}^2\right)$$

$$=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)=\frac{1}{n-1}\left(\sum_{i=1}^{n}E\left(X_i^2\right)-nE\left(\overline{X}^2\right)\right)$$

$$=\frac{1}{n-1}\left(n\left(\sigma^2+\mu^2\right)-n\left(\frac{\sigma^2}{n}+\mu^2\right)\right)=\sigma^2$$

Ways to Characterize the Sampling Distribution

1. The easiest way to characterize a statistic's sampling distribution is to calculate some of its features, like its mean and variance.
   - We've already seen examples of this:
     - An estimator's bias depends on the mean of the sampling distribution.
     - Comparing the efficiency of two estimators involves comparing the variance of sampling distributions (i.e., comparing their sampling variances)
   - The **standard deviation of the sampling distribution of a statistic** has a special name. We call it the **standard error** of the statistic.
2. If we know the exact probability distribution of the population from which the sample is drawn (or if we assume one) we can go further and pin down the statistic's exact sampling distribution.

   Example: when sampling from a Normal population, the sample mean is normally distributed.
3. If we're unwilling or unable to assume an exact population distribution, we can rely on **asymptotic theory** to derive an **approximate** sampling distribution for the statistic as the sample size tends to infinity.
4. We can use a computer to simulate the sampling distribution.

The Law of Large Numbers and the Central Limit Theorem

- We have some powerful theorems we can use to describe the behavior of a sample mean as the sample size tends to infinity.
- Why do we care about sample means so much? It turns out that most statistics we care about can be written as the sample mean of *something*.
- Therefore, we can use these theorems to describe an **approximate** sampling distribution for many statistics.
  - Only approximate because our sample is always finite.
- The **law of large numbers** says that as the sample size n approaches infinity, the sample mean will be close to the population mean with very high probability.
  - If $R \to Q$ as $n \to \infty$, we say R is a **consistent** estimator of Q.
  - **The law of large numbers says the sample mean is a consistent estimator of the population mean**.
- The **central limit theorem** says that as the sample size n approaches infinity, the sampling distribution of the sample mean is approximately Normal with mean μ and variance σ^2/n.
  - **This is true no matter what the population distribution is.**

$$\boxed{\text{CLT}: \text{ as } n \to \infty,\ \bar{x} \to N\left(\mu, \sigma^2 / n\right)}$$

Simulating the Sampling Distribution

- When we can't pin down a statistic's exact sampling distribution, an alternative to asymptotic approximation is to **simulate** the sampling distribution using a computer.
- When we do this with "real" data, we call it **bootstrapping**. When we do it with "fake" data we call it **Monte Carlo** simulation.
- A Monte Carlo example.
  - Suppose we want to know the sampling distribution of the sample mean when sampling from a Chi-square distribution.
  - We could get the computer to generate 100 random numbers drawn from a Chi-square distribution with v degrees of freedom, and compute the sample mean of the randomly generated numbers.
  - If we repeat this many times (say 10,000) we get 10,000 estimates of the sample mean
  - The distribution of our 10,000 estimates is a good approximation to the sampling distribution of the sample mean
  - We could estimate the variance of the sampling distribution of the sample mean using the sample variance of the 10,000 estimates of the sample mean
  - We could plot the sampling distribution with a histogram.
  - We could compute the proportion of times the sample mean lies in an interval to estimate the probability the population mean lies in that interval.
  - etc.

Why do we care so much about sampling distributions?

- The point of statistical inference is to use the observed sample to learn about the population.
- We care about things like the population mean, the population variance, etc.
- But we don't observe population quantities – we only observe the sample.
- So we **estimate** the population quantities using sample statistics.
- Then, we usually want to **test hypotheses** about the population quantities
  - We might want to test whether the population mean is 10.
  - Or whether it is less than 6.
  - Or whether the population variance is 7.
- **Knowing the sampling distribution of a statistic allows us to test hypotheses like these.**

Hypothesis Testing

- Example: suppose we ask 100 randomly selected people how many times they go to the movies in a year.
- Suppose the sample mean of their responses is 7.
- This is probably an ok estimate of the population mean, but because the sample is “small,” we could be wrong.
- We might want to test the hypothesis that the true population mean is 8.
- **If we know the sampling distribution of the sample mean, we can compute the probability of finding a sample mean of 7 in a sample of 100 people, given that the true population mean is actually 8.**
- If this probability is “small,” then we can be fairly certain that the true population mean is not 8.
- If this probability is “big” then we cannot rule out the possibility that the true population mean is 8.
- We formalize this with a **hypothesis test** (next day).