Economics 120A

7. Point Estimation

Graham Elliott

Overview of Statistics

Model/World

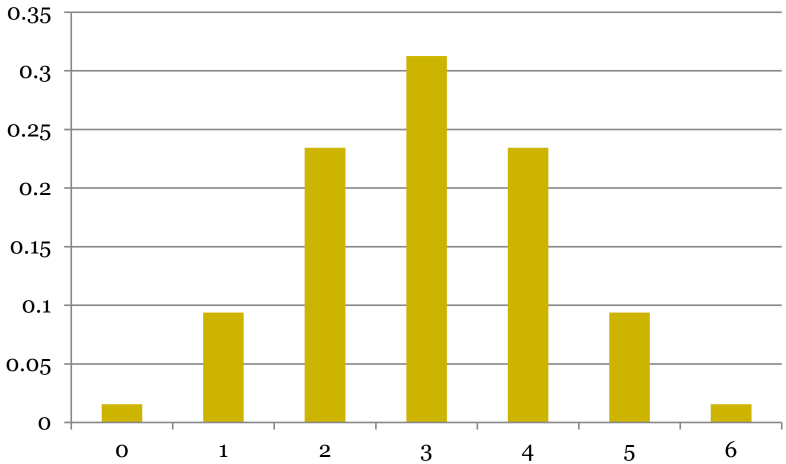
Data

Cards Example

6 draws

$x_i \begin{cases} 0 & \text{red} \\ 1 & \text{black} \end{cases}$

We now understand the cards example from the first class: each card can be modelled as a Bernoulli(0.5) random variable, the number of ~~red~~ black cards is Binomial (6,0.5) if n=6 is fixed ahead of time.



What do we do?

1. Point Estimation

What function of the data (sets of zeros and ones in the cards example) do I take to give my best answer to “What is the mean of X here?”

2. Hypothesis Testing

Suppose I think I know the mean (have a theory) does the data suggest that the theory is true or does it suggest that the theory is false?

3. Confidence Intervals

A middle ground answer – give me a range of possible values for the mean of X.

Point Estimation

We have data $(x_1, x_2, \ldots, x_n)$ which we consider to be outcomes from random variables $(X_1, X_2, \ldots, X_n)$.

Suppose that the mean of each of the random variables is μ, and we want to learn about μ.

What function of the data $U(x_1, x_2, \ldots, x_n)$ do we want to take to give our best estimate for μ?

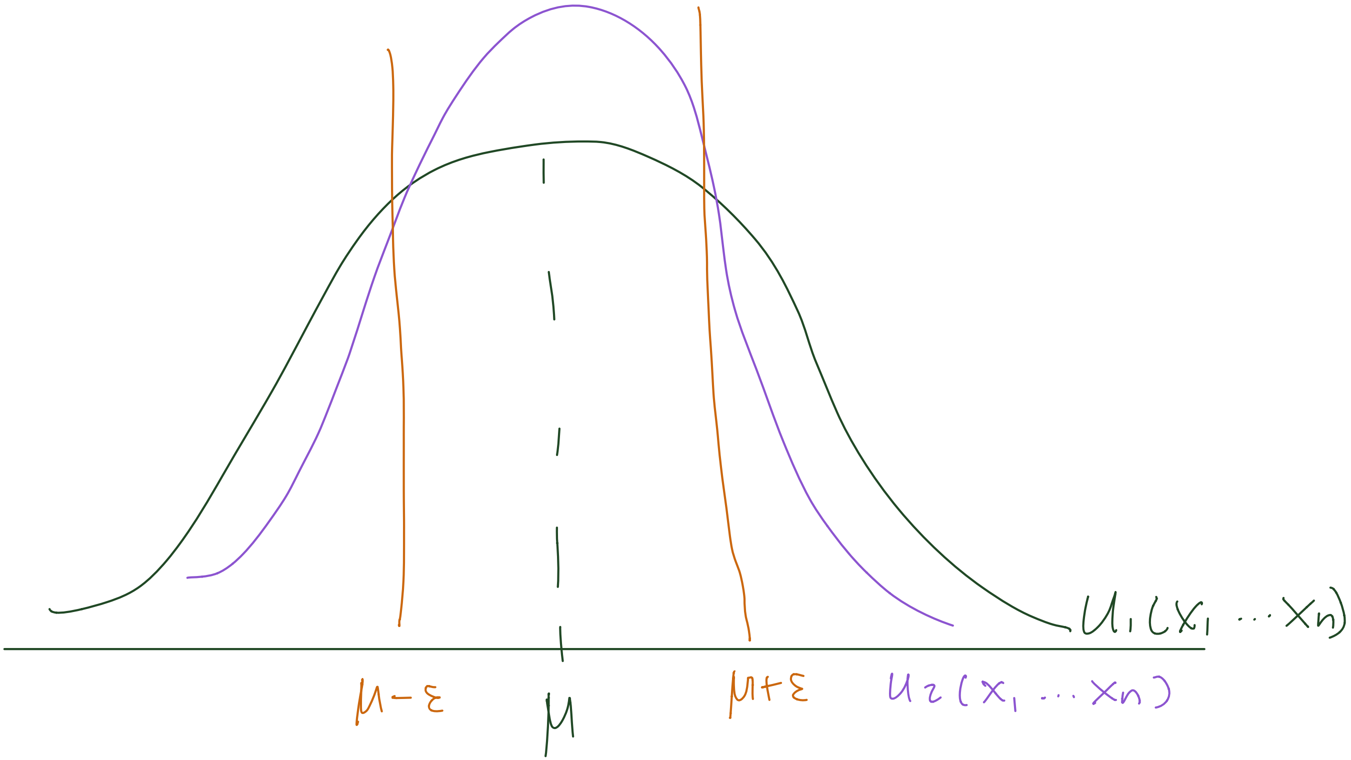
We will answer this question by realizing that $U(x_1, x_2, \ldots, x_n)$ is an outcome of the random variable $U(X_1, X_2, \ldots, X_n)$ which has a distribution describing the randomness of the likely estimates, which will have some relationship with what we are estimating, i.e. μ.

So what distributions are going to be good ones?

Point Estimation

What distributions are going to be good ones?

$P[\mu - \varepsilon < u_1 < \mu + \varepsilon] < P[\mu - \varepsilon < u_2 < \mu + \varepsilon]$



Point Estimation

We have a number of ways to think about what is “Good”.

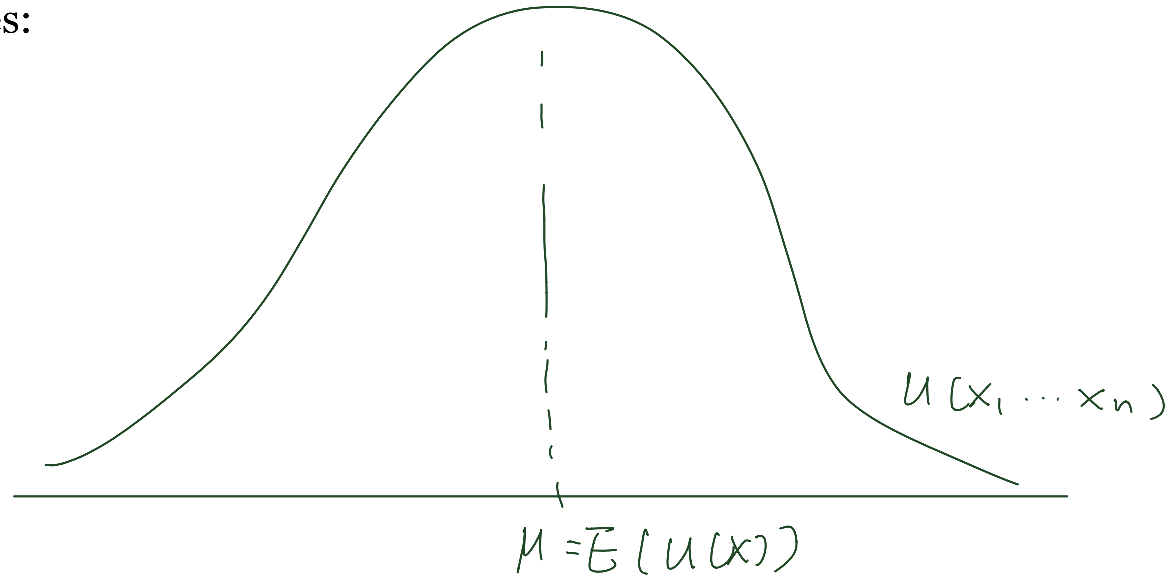
We will go through each of them, they are

7.2 Unbiasedness
7.3 Efficiency
7.4 Mean Square Error
7.5 Consistency.

Unbiasedness

Definition: If a statistic U(X) is used to estimate θ, then the Bias of the statistic is E[U(X)-θ]. If this is equal to zero for all possible θ then we say the estimator is unbiased.

In pictures:



Be sure to remember that for any particular dataset, we have most likely that . It is as we said above, that this is a draw. Unbiasedness refers to what happens on average, not in a single draw.

Unbiasedness

The sample mean $\bar{X}$ when $(X_1, X_2, \ldots, X_n)$ are a VSRS is unbiased.

i.e. Bias = $E[\bar{X} - \mu] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right]$

$$= \frac{1}{n}\sum_{i=1}^{n} E x_i - \mu$$

$$= \frac{1}{n} \cdot n \cdot \mu - \mu$$

$$= 0$$

$\bar{x}$ unbiased for μ

Unbiasedness

Consider s^2 to estimate σ^2.

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

First, note that

$$E(X_i - \bar{X})^2 = E((x_i-\mu) - (\bar{x}-\mu))^2$$

$$= E(x_i-\mu)^2 + E(\bar{x}-\mu)^2 - 2E((x_i-\mu)(\bar{x}-\mu))$$

$$= \sigma^2 + \frac{\sigma^2}{n} - \frac{2}{n}E[(x_1-\mu)(x_i-\mu) + (x_2-\mu)(x_i-\mu)$$

$$+ \cdots + (x_n-\mu)(x_i-\mu)]$$

$$= \sigma^2 + \frac{\sigma^2}{n} - \frac{2}{n}E(x_i-\mu)^2$$

因为 $(E(x_i-\mu)(x_j-\mu)=0$
$i \neq j)$

$$= \sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n}$$

So

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

$$E\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\cdot n \cdot \sigma^2 \cdot \frac{n-1}{n} = \sigma^2$$

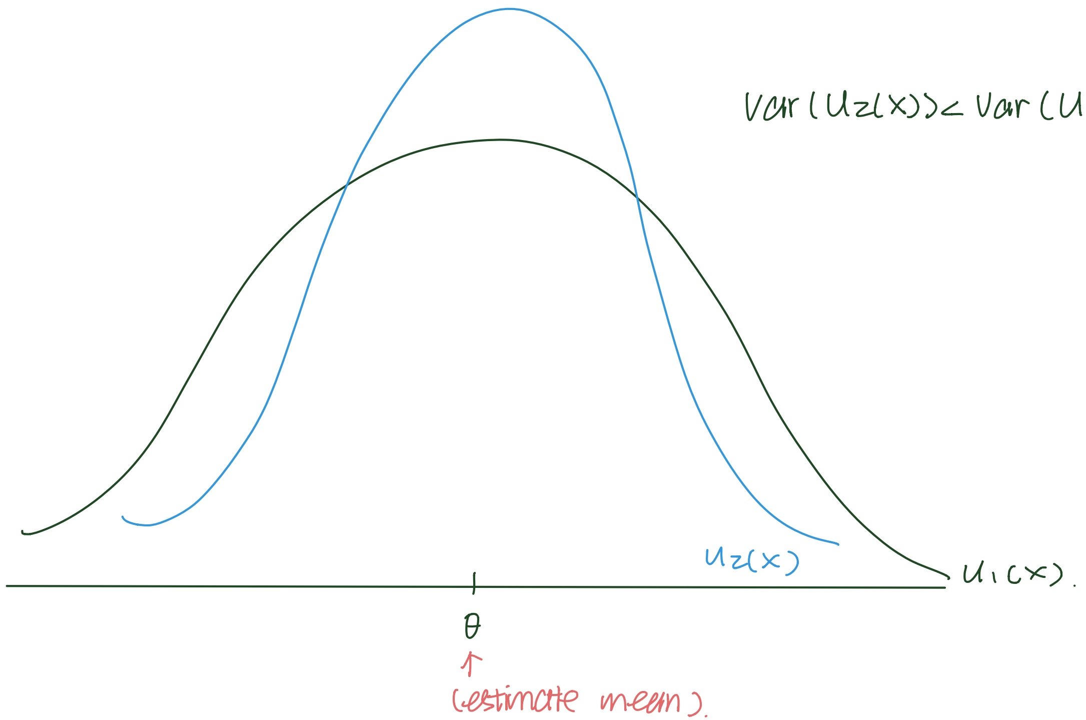
Efficiency

Suppose that we have two unbiased estimators – how do we choose?

Choose the one with the smallest variance, this is known as the efficient unbiased estimator.

In pictures:

$Var(U_2(X)) < Var(U_1(X))$



Mean Square Error of an Estimator

More often we want to compare estimators when one or more are biased.

In this case we might want to choose the one that has the smallest Mean Squared Error.

Mean Square Error

This is $MSE(u(x)) = \int_{-\infty}^{\infty} (u(x)-\theta)^2 p_x(x)\,dx$

$$= E[(u(x)-\theta)^2]$$

Notice we can rearrange this $E[(u(x)-\theta)^2]$

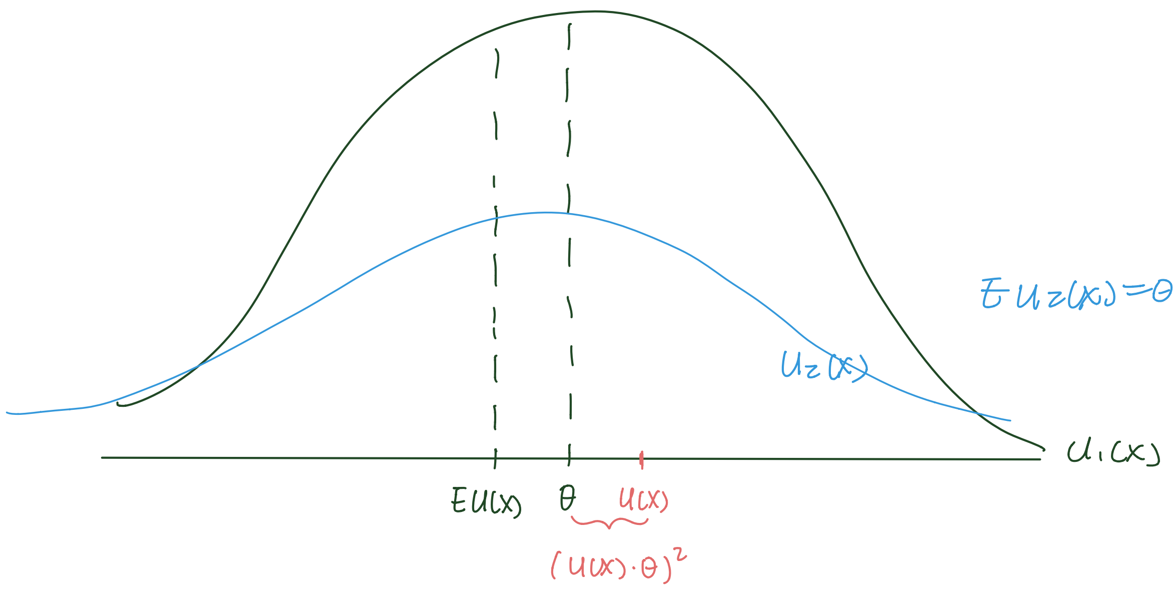
$$= E[((u(x)-Eu(x))+(Eu(x)-\theta))^2]$$

$$= \underbrace{E[(u(x)-Eu(x))^2]}_{Var(u(x))} + \underbrace{(E(u(x))-\theta)^2}_{(Bias)^2} + 2E(u(x)-\theta)\underbrace{E(u(x)-Eu(x))}_{Eu(x)-Eu(x)=0}$$

So there is a tradeoff between variance and bias.

Mean Square Error

In pictures:



Consistency

The notion of consistency of an estimator is that for large enough samples, we can have as accurate an estimate as we desire. Alternatively, we can think of this as being that when the sample size becomes infinitely large, we learn the true value of the parameter.

In pictures:

Consistency

Recall that for a VSRS, then $\bar{X}_n \sim^a N(\mu, \sigma^2/n)$.

So we have $P[|\bar{X}_n - \mu| > \varepsilon] = 1 - P[|\bar{X}_n - \mu| < \varepsilon]$

$$= 1 - P[-\varepsilon < \bar{X}_n - \mu < \varepsilon]$$

$$= 1 - P\left[-\frac{\sqrt{n}\,\varepsilon}{\sigma} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{\sqrt{n}\,\varepsilon}{\sigma}\right]$$

$$= 1 - P\left[-\frac{\sqrt{n}\,\varepsilon}{\sigma} < z < \frac{\sqrt{n}\,\varepsilon}{\sigma}\right]$$

$n \to \infty$ $\quad -\frac{\sqrt{n}\,\varepsilon}{\sigma} \to -\infty$

$n \to \infty$ $\quad \frac{\sqrt{n}\,\varepsilon}{\sigma} \to \infty$

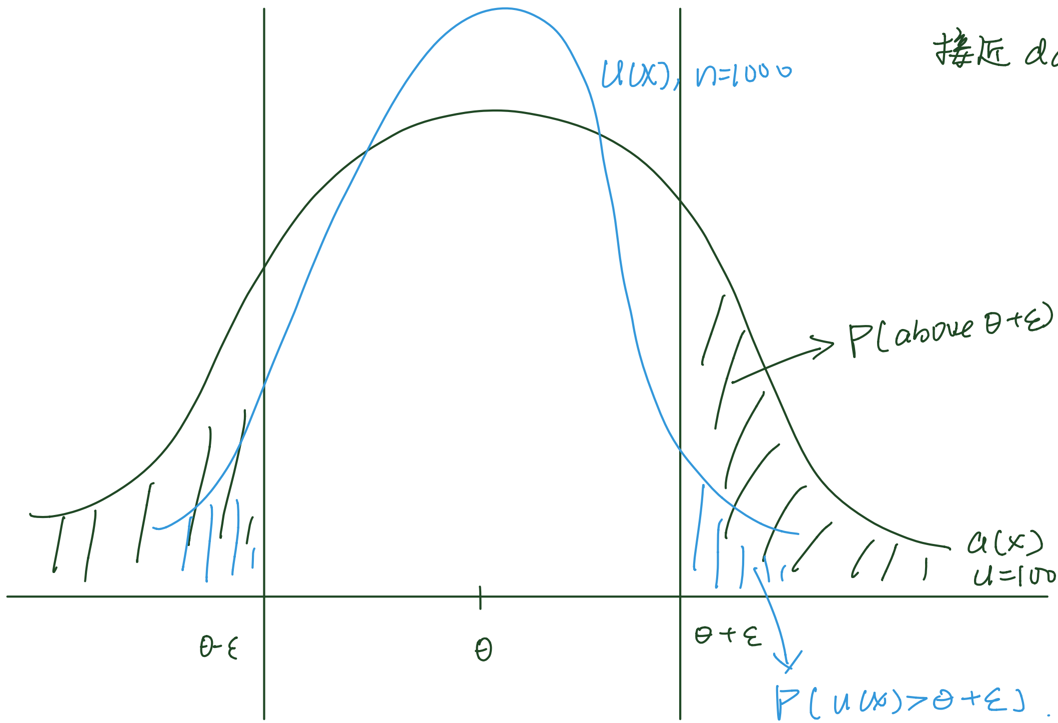
So as n becomes large, this probability goes to zero.

We say $\bar{X}_n \xrightarrow{P} \mu$ i.e $\bar{X}_n$ is consiscent for μ

Consistency

In pictures:

sample size 越大，
actual mean 越
接近 data



Discussion

Since a point estimate for any sample may or may not be close to what we are estimating, we really need more information.

How accurate is the estimate? For this we need to know the spread of the distribution, thus we need to introduce the variance.

Reporting simply the estimate is never enough.