MFin Econometrics I

Session 1: Normal distribution, Estimators, Sampling distributions of estimators

Thilo Klein

University of Cambridge
Judge Business School

Timetable

Timetable

- Lectures
  7, 14, 21, 28 Nov, 9-12.30pm @ LT2
- Lab Sessions EViews and RExcel software
  Despo Malikkidou
  11, 18, 23, 30 Nov, 2-4pm @ Computer Lab
  Jerry He
  11, 18, 23, 30 Nov, 2-4pm @ (W2.01, LT2, LT2, LT2)

CAMBRIDGE Judge Business School

Deadlines

Contest (Multiple Choice Exercises)

	Sheet 1	Sheet 2	Sheet 3	Sheet 4
Submit on	13 Nov	17 Nov	20 Nov	22 Nov
Weight	8 %	10 %	12 %	14 %

	Sheet 5	Sheet 6	Sheet 7	–
Submit on	27 Nov	29 Dec	6 Dec	–
Weight	16 %	18 %	22 %	–

Assessment (Workbooks)

	Book 1	Book 2
Handed out	30 Nov	30 Nov
Submit on	12 Dec	12 Dec
Weight	50 %	50 %

CAMBRIDGE Judge Business School

Objectives

Objectives of the module

- "Introduction" to *applied* statistical methods
- Mathematical sophistication $\sim$ simpler research journal papers in finance/strategy/marketing ...
- Learning by doing - do many exercises
- Should enable you to estimate useful, insightful and *exciting* regression models and make careful inferences.

Motivation

Patterns and relationships in Finance / Management / Economics with important strategic and policy implications:

- Do financial intermediaries reduce information asymmetries on online lending platforms?
- Does management advice improve productivity and performance of firms?
- Does microfinance reduce poverty?
- How much are people willing to pay for different hospital care packages?
- Does smoking lead to lung cancer?
- Do smaller class sizes lead to better test score performance?

CAMBRIDGE Judge Business School

Causal Effects

Using data to measure causal effects (1)

- Ideally we should do experiments:
  - e.g., experiment to estimate effect of access to microcredit on small enterprise revenue / household consumption / savings, etc.
- But almost always have to make do with *observational* (non-experimental) data
  - At best, data from "natural experiments"
  - Increasingly, behavioural finance, economics, management data come from class room experiments

CAMBRIDGE Judge Business School

Causal Effects

Using data to measure causal effects (2)

- What is the difficulty in using observational data when we wish to estimate causal effects?
  - Notion: Data generating process: empirical observations are outcome of (natural) "experiments"
  - The same experiment performed by "nature", leads to different outcomes (some randomness)
  - And, we have no control over the experiment of interest

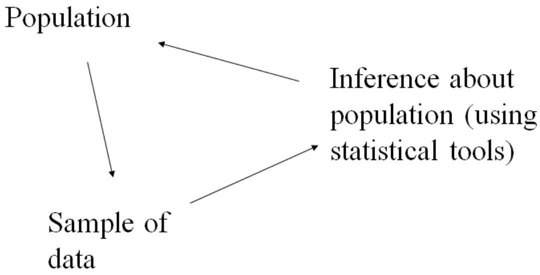
  We need to:
    - identify the causes and factors relevant to the outcome of interest
    - To disentangle effects of the different causes on the outcome
    - To come to conclusions about these effects with some assurance about their level of accuracy, i.e., quantifying our uncertainty about conclusions

Learning points

You will (learn) ...

- Statistics studies *sets* of objects/entities/things (firms, individuals, households ...)
- Statistics studies "causes of variation": If there is no variation, one individual describes the population
- You will learn
  - How to exploit variation (between observations in data) to estimate causal effects
  - Hands-on experience of regression with focus on applications - theory only as needed
  - How to evaluate the other people's analysis - understand empirical papers critically

Quantitative research



Paradigm

So, you (should) have a *useful* theory about the phenomenon of interest. You need to solve:

1. the *Specification* problem - specify a model from (your) theory. The mathematical form you think governs the population. You do not know (and will never *know*) the *parameters* of this
2. the *Estimation* problem - choose methods to *estimate* the unknown parameters governing the population, using sample data
3. the *Inference* problem - quantify the degree of uncertainty attached to these estimates, given that they are based on just one (random) sample

CAMBRIDGE Judge Business School

Method

Step by step

- Formulate a model (based on hypotheses about the population)
- Gather data - sample
- Estimate the model - estimate population parameters
- Make inferences - test hypotheses about the population
- Interpret results, in terms of the theory

Review

Topics today:

- Data description: *Statistics* that summarise data - these are always "estimates" of the unknown population parameters
- Probability principles: how can the world be described in terms of random variables and probability distributions (i.e., probability models)
- Next: Introduction to statistical inference: drawing conclusions about the *population* from only one sample, using probability principles

CAMBRIDGE Judge Business School

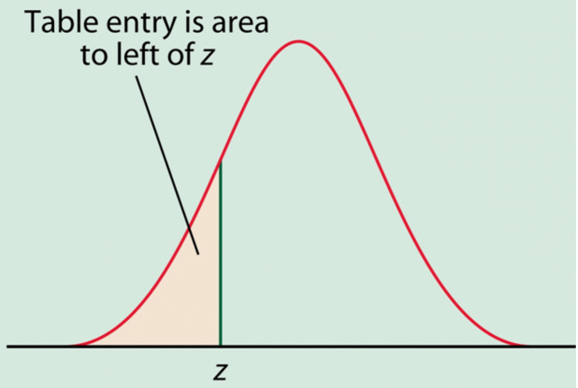
Review (cont'd)

Further on...

- Estimation procedures for regression models: why and how they work
- Inference after regression: how to test hypothesis

Using the Standard Normal Distribution (SND)

SND Table area



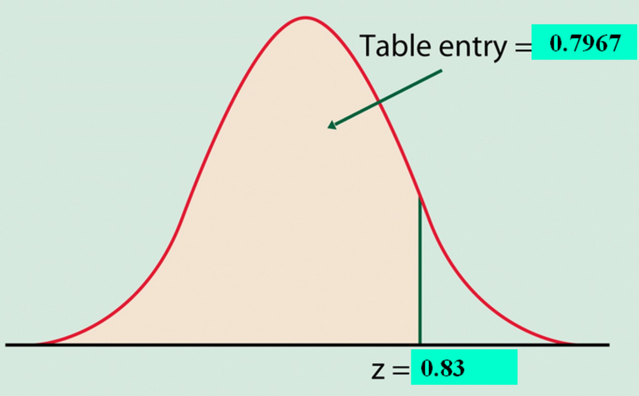
Using the Standard Normal distribution

SND Table

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

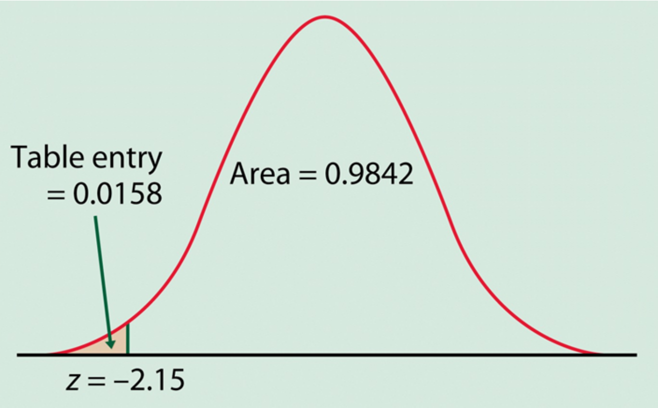
Proportion smaller than 0.83?

What proportion of observations are smaller than 0.83?



Proportion greater than −2.15?

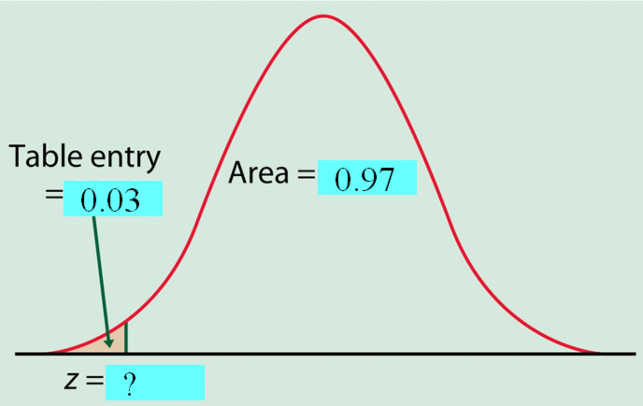
What proportion of observations are greater than −2.15?



Inverse of SND

Inverse of SND: $F^{-1}(.3) = ?$

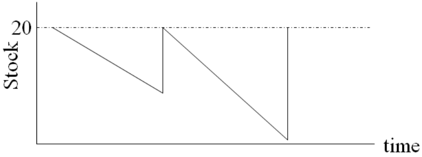
Z Value that cumulates 3% of probability



Example

Inventories in a dealership

An inventory or resource management problem: A dealership's stock of new autos is replenished to 20 every month.

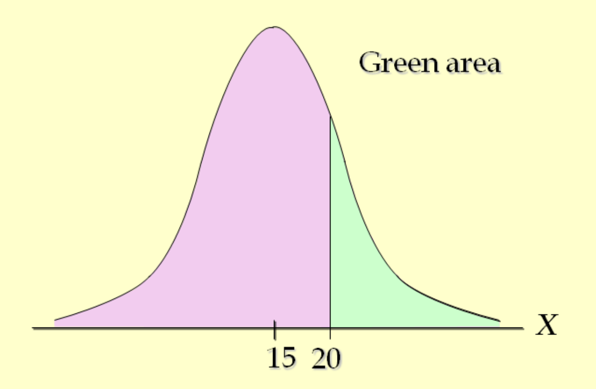


- Sales are lost due to stockouts
- Known that demand (X) within the month is normally distributed with a mean of 15 and a standard deviation of 6
- What is the probability of a stockout?

Using the Standard Normal distribution

Solving for the stockout probability

$P(X > 20)$given $X \sim N(15, 6^2)$



Using the Standard Normal distribution (cont'd)

Solving for the stockout probability (cont'd)

- Convert $x = 20$ to its standard normal value

$$\begin{aligned} z &= (x-\mu)/\sigma \\ &= (20-15)/6 \\ &= 0.83 \end{aligned}$$

- Find area under SND *to the right of* $z = 0.83$

$$\begin{aligned} Pr(z > 0.83) &= 1 - F(0.83) \\ &= 1 - 0.797 \\ &= 0.20 \end{aligned}$$

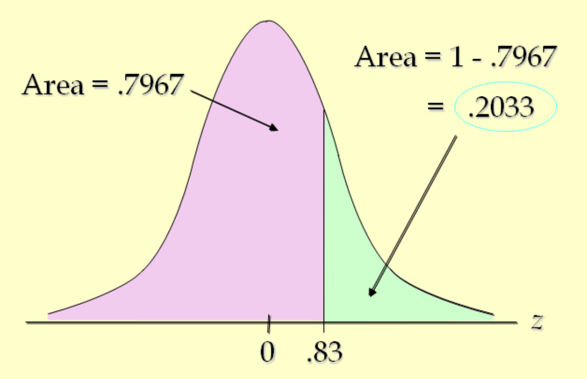
- Probability of stockout $= Pr(X > 20) = 0.2$

CAMBRIDGE Judge Business School

Using the Standard Normal distribution (cont'd)

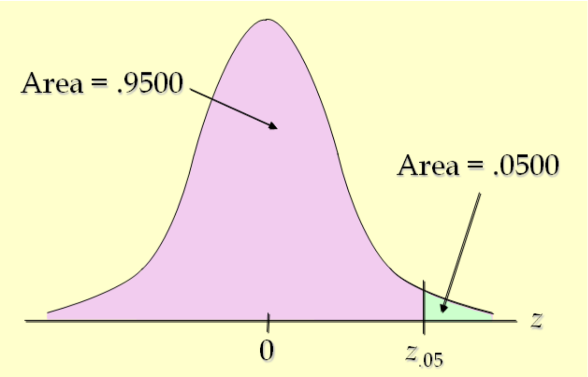
Solving for the stockout probability (cont'd)

If the probability of stockout is to be no more than 5%, what should the reorder point be?



Using the Standard Normal distribution (cont'd)

Solving for the reorder point



CAMBRIDGE Judge Business School

Using the Standard Normal distribution (cont'd)

Solving for the reorder point (cont'd)

- We know from the SND that $z_{0.05} = 1.645$
- We are interested in the corresponding x value

$$
\begin{aligned}
x &= \mu + z_{0.05}\sigma \\
&= 15 + 1.645 \times 6 \\
&= 24.9
\end{aligned}
$$

- Reorder point of 25 automobiles will keep probability of stockout at slightly less than 0.05
- By increasing reorder point from 20 to 25 the probability of stockout falls from .2 to 0.05

CAMBRIDGE Judge Business School

Estimators

From the dist. of r.v. X, to the dist. of estimators

- Begin with a r.v. X and its probability distribution, $f(X, \theta)$ or $f_X(x; \theta_1, \cdots, \theta_L)$, characteristic of the population
- Parameter (θ) is the fixed, but unknown value (or set of values) that describes the popln. distribution, e.g.: true mean and variance of a price distribution
- The number of parameters depends on the distribution. The Normal has two
- Note: Distributions have generating mechanisms
  - The Central Limit Theorem is an example of a generating process: a *stochastic* process that underlies the r.v. (average, in this case)
- A random *vector* variable $(X_1, X_2, \cdots, X_n)$ is characterized by its *joint distribution*: $f_{X_1,\cdots,X_n}(x_1, \cdots, x_n; \theta_1, \cdots, \theta_K)$, e.g., a multivariate normal distribution

CAMBRIDGE Judge Business School

Estimators

Definitions, contd.

- A statistic is any given function of observable values, which can be evaluated from a sample, e.g., $m = max(X_1, ..., X_n)$
- As a function of random variables, a statistic is itself a random variable
- An estimator $(\hat{\theta})$ is the sample counterpart of a(n unknown) population parameter (θ). It is a statistic, i.e., it can be calculated from observed values
- An estimate is the numerical value obtained when the estimator is applied to a specific sample
- Sampling distribution is the prob. distribution over values taken by estimates across all possible samples of the same size from the population

CAMBRIDGE Judge Business School

Estimators

Unbiasedness

- An estimator $\hat{\theta}$, is unbiased if $E(\hat{\theta}) = \mu_{\hat{\theta}} = \theta$
- If not, the estimator is biased

$$\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

- Q: Is the sample mean an unbiased estimator of the population mean?
- How can we find out whether $E[\bar{X}] = \theta$?

CAMBRIDGE Judge Business School

Estimators

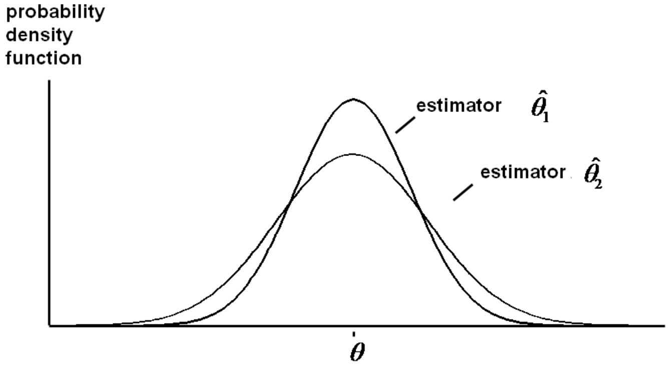
Efficiency

- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ
- Estimator $\hat{\theta}_1$ is the more efficient of the two if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$
- Among unbiased estimators, the one with the smallest variance is called the best unbiased estimator

CAMBRIDGE Judge Business School

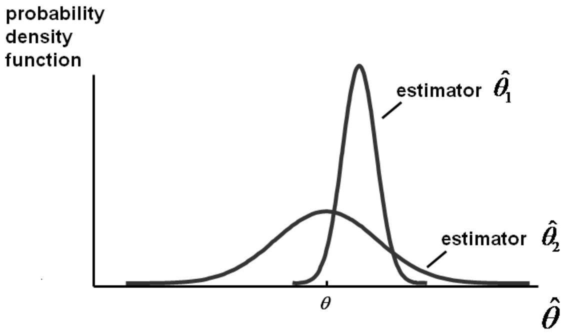
Estimators

Unbiasedness and Efficiency



Estimators

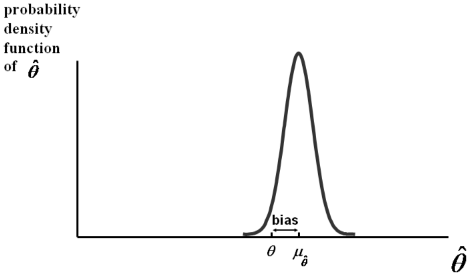
Conflict between unbiasedness and efficiency



Estimators

Mean square error: resolving trade-off between bias and inefficiency

- Think in terms of a *loss function*, which reflects the cost of making errors, positive or negative, of different sizes
- A widely used loss function : Mean square error (MSE) of the estimator = $E($ square of deviation of estimator from true $)$
- $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$, which is $= {\sigma_{\hat{\theta}}}^2 + (\mu_{\hat{\theta}} - \theta)^2$



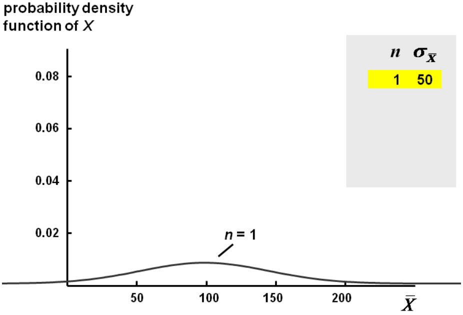
Asymptotic properties of Estimators

Large sample (asymptotic) properties of estimators

- The *finite sample* distribution of an estimator may often not be known
- Even so, statisticians are often able to figure out the sampling distribution of estimators when n is large enough
- e.g., Central limit theorem
- One relevant concept here is Consistency of the estimator

Asymptotic properties of Estimators

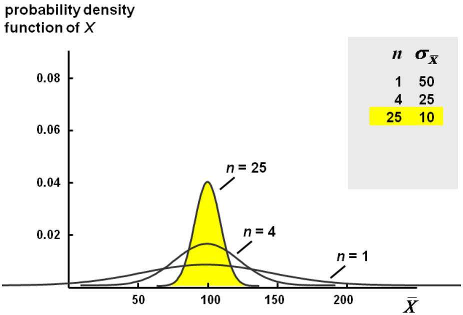
Effect of increasing the sample size on the distribution of $\bar{X}$



- Assume $E(X) = \mu_X = 100$ and $s.d.(X) = \sigma_X = 50$
- We do not know these population parameters
- We use the sample mean to estimate the population mean

Asymptotic properties of Estimators

Increasing sample size and the distribution of $\bar{X}$ (cont'd)

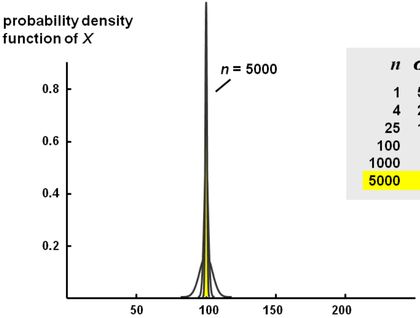


- How does the shape of the distribution change as the sample size is increased?
- The distribution is more concentrated about the pop. mean

CAMBRIDGE Judge Business School

Asymptotic properties of Estimators

Increasing sample size and the distribution of $\bar{X}$ (cont'd)



n	$\sigma_{\bar{X}}$
1	50
4	25
25	10
100	5
1000	1.6
5000	0.7

- The distribution collapses to a spike at the true value
- $\sigma^2_{\bar{X}} \to 0$
- The sample mean is a consistent estimator of the population mean.

Asymptotic properties of Estimators

Large sample (Asymptotic) properties of any estimator $\hat{\theta}$ is to do with:

- How the sampling distribution of $\hat{\theta}_n$, where n is the size of the sample, changes when n increases towards infinity?
- $\hat{\theta}$ is a consistent estimator for θ if:

$$plim(\hat{\theta}) = \theta$$

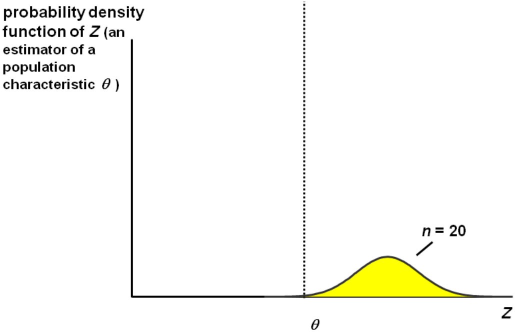
i.e.,

$$Prob(\theta - \epsilon \leq \hat{\theta}_n \leq \theta + \epsilon) = 1 \text{ as } n \to \infty$$

CAMBRIDGE Judge Business School

Asymptotic properties of Estimators

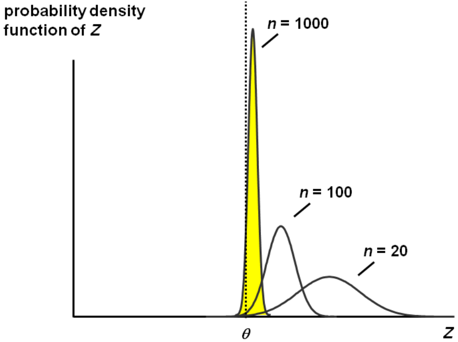
Example: Estimator biased in finite samples but consistent



- $\hat{\theta}$ is an estimator of a population characteristic θ
  From the probability distribution of $\hat{\theta}$, $\hat{\theta}$ is biased upwards
- We will see soon that the sample variance (if measured as $\sum(X_i - \bar{X})^2/n$ is biased downwards

Asymptotic properties of Estimators

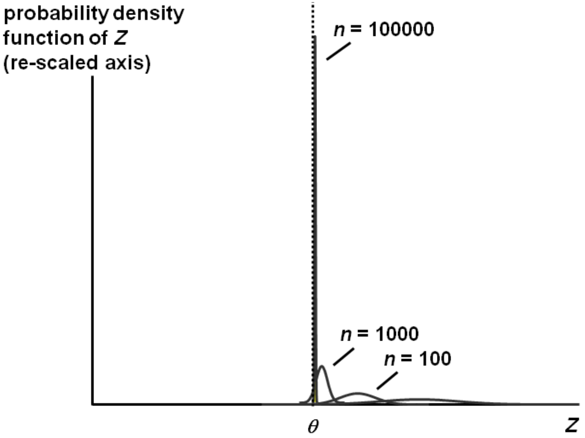
Example: biased in finite samples but consistent (cont'd)



The distribution collapses to a spike with larger samples

Asymptotic properties of Estimators

Example: biased in finite samples but consistent (cont'd)



Example?

Sampling and Sampling distribution

Distribution of a sample, $Y_1, ..., Y_n$, under random sampling

- Under simple random sampling:
  - We choose an individual (firm, household, stock, entity ...) at random from the population
  - Prior to sample selection, the value of Y is random because the individual is to be selected randomly
  - Once the individual is selected, the value of Y is observed, and Y is not random
  - The data set is $(Y_1, Y_2, .., Y_n)$, $\quad Y_i =$ is the value of the r.v. pertaining to the i^{th} entity sampled

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

Distribution of $Y_1, \ldots, Y_n$ under simple random sampling

- Because individuals i and j are selected at random, the value of Y_i has no information on the value of Y_j (independent events)
  - Y_i and Y_j are independently distributed
- Because Y_i and Y_j come from the same distribution
  - Y_i and Y_j are identically distributed
- So under simple random sampling, Y_i and Y_j are independently and identically distributed (i.i.d.)
- More generally, under simple random sampling, $\{Y_i\}$, $i = 1, \ldots, n$ are i.i.d.
- Probability theory makes statistical inference about moments of population distributions simple when samples drawn from the population are *random*

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

The sampling distribution of $\bar{Y}$

- $\bar{Y}$ is a random variable, and its properties are given by the sampling distribution of $\bar{Y}$
  - The individuals in the sample are drawn at random; so the vector $(Y_1, ..., Y_n)$ is random
  - So functions of $(Y_1, ..., Y_n)$, such as $\bar{Y}$, are random. Different samples, different $\bar{Y}$ values
  - The distribution of $\bar{Y}$ over each of the different possible samples of size n is the sampling distribution of $\bar{Y}$
  - The mean and variance of $\bar{Y}$ are the mean and variance of its sampling distribution: $E(\bar{Y})$ and $Var(\bar{Y})$
  - The concept of sampling distribution underpins statistical analysis

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

Things we want to know about the sampling distribution

- What is the mean of $\bar{Y}$?
  - If $E(\bar{Y}) = \mu_Y$, then $\bar{Y}$ is an *unbiased* estimator of μ_Y
- What is the variance of $\bar{Y}$?
  - If the variance of $\bar{Y}$ is lower than that of another estimators of μ, then $\bar{Y}$ estimator is the more *efficient*
  - How does $Var(\bar{Y})$ depend on n?
    Does $\bar{Y}$ tend to fall closer to μ as n grows large?
  - if so, $\bar{Y}$ is a *consistent* estimator of μ
- Can we pin down the probability distribution (i.e., the sampling distribution) of $\bar{Y}$?

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

Mean of the sampling distribution of $\bar{Y}$

- General case - i.e., for Y_i, i.i.d. from any distribution:

$$E(\bar{Y}) = E(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} \mu_Y = \mu_Y$$

- $\bar{Y}$ is an unbiased estimator of μ_Y $\quad (E(\bar{Y}) = \mu_Y)$

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

Variance of the sampling distribution of $\bar{Y}$

$$
\begin{aligned}
Var(\bar{Y}) &= E[(\bar{Y} - \mu_Y)^2] \\
&= E\left[\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) - \mu_Y\right)^2\right] \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_Y)\right)^2\right] \\
&= E\left[\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_Y)\right] \times \left[\frac{1}{n}\sum_{j=1}^{n}(Y_j - \mu_Y)\right]\right]
\end{aligned}
$$

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

Variance of the sampling distribution of $\bar{Y}$ (2)

$$\begin{aligned} Var(\bar{Y}) &= E\left[\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_Y)\right] \times \left[\frac{1}{n}\sum_{j=1}^{n}(Y_j - \mu_Y)\right]\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[(Y_i - \mu_Y)(Y_j - \mu_Y)\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(Y_i, Y_j) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_Y^2 \\ &= \frac{\sigma_Y^2}{n} \end{aligned}$$

Note: $Cov(Y_i, Y_j) = 0$ for $i \neq j$; $Cov(Y_i, Y_j) = Var(Y_i)$ for $i = j$

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

Variance of the sampling distribution of $\bar{Y}$ - simpler

$$Var(\bar{Y}) = Var\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i)\right]$$
$$= \frac{1}{n^2}Var\left[\sum_{i=1}^{n}(Y_i)\right]$$

Recall: $V(Y_1 + Y_2) = V(Y_1) + V(Y_2) + 2Cov(Y_1, Y_2)$

But $Cov(Y_i, Y_j) = 0$ for $i \neq j$ (Why?)

So:

$$Var(\bar{Y}) = \frac{1}{n^2}nV(Y_i)$$
$$= \frac{\sigma(Y)^2}{n}$$

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

Mean and variance of sampling distribution of $\bar{Y}$

$$E(\bar{Y}) = \mu_Y$$

$$Var(\bar{Y}) = \frac{{\sigma_Y}^2}{n}$$

- $\bar{Y}$ is an unbiased estimator of μ
- $Var(\bar{Y})$ is inversely proportional to n
- the spread (st. dev.) of the sampling distribution is proportional to $\frac{1}{\sqrt{n}}$
- Larger samples, less uncertainty: Consistent

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

The sampling distribution of $\bar{Y}$ when n is large

- For small sample sizes, the distribution of $\bar{Y}$ is complicated, but if n is large, the sampling distribution is simple!
- **Law of Large Numbers**
  - If $(Y_1, ..., Y_n)$ are i.i.d. and ${\sigma_Y}^2 < \infty$, then $\bar{Y}$ is a consistent estimator of μ_Y : $\text{plim}(\bar{Y}) = \mu_Y$
  - $\bar{Y}$ converges in probability to μ_Y
  - i.e., as $n \to \infty$, $Var(\bar{Y}) = \frac{{\sigma_Y}^2}{n} \to 0$

CAMBRIDGE Judge Business School

Sampling and Sampling distribution

The Central Limit Theorem (CLT) statement

- If $(Y_1, ..., Y_n)$ are i.i.d. and $0 < {\sigma_Y}^2 < \infty$, then when n is *large*, the distribution of $\bar{Y}$ is approximated well by a normal distribution
  - $\bar{Y} \sim N(\mu_Y, \frac{{\sigma_Y}^2}{n})$ approximately
  - Standardized $\bar{Y} = \frac{\bar{Y} - \mu_Y}{\frac{\sigma_Y}{\sqrt{n}}} \sim N(0, 1)$ approximately
  - The larger is n, the better the approximation

CAMBRIDGE Judge Business School