MPO1: Quantitative Research Methods

Session 3: Normal distribution, Estimators, Sampling distributions of estimators, Tests of hypotheses

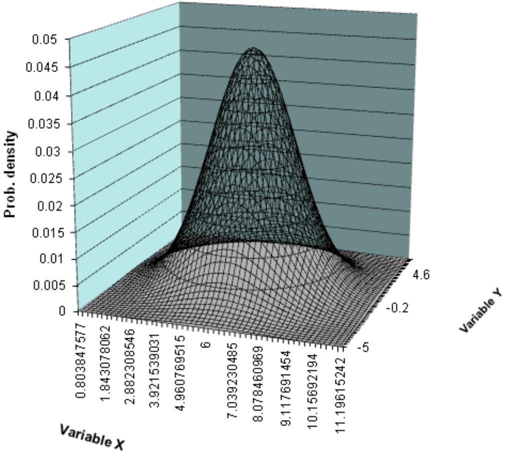
Thilo Klein

University of Cambridge
Judge Business School

CAMBRIDGE Judge Business School

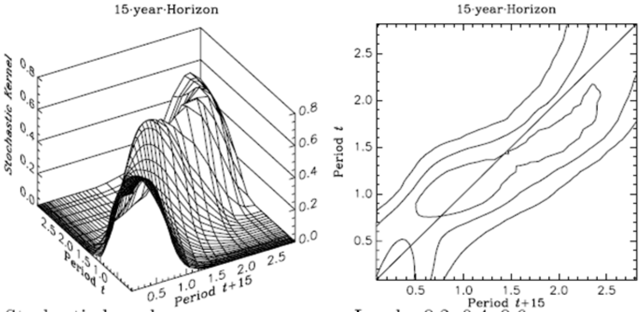
Joint Distributions, review

Bivariate Normal distribution, independent r.v.s



Joint distribution of two continuous random variables

Danny Quah, 2000



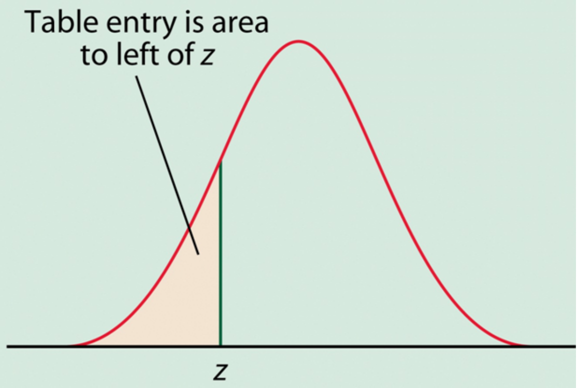
Stochastic kernel Levels: 0.2, 0.4, 0.6

Fig. 3: Distribution dynamics across countries (Relative output per worker) The right panel contains contour plots of the 15-year stochastic kernel in the left panel.

CAMBRIDGE Judge Business School

Using the Standard Normal Distribution (SND)

SND Table area



Using the Standard Normal distribution

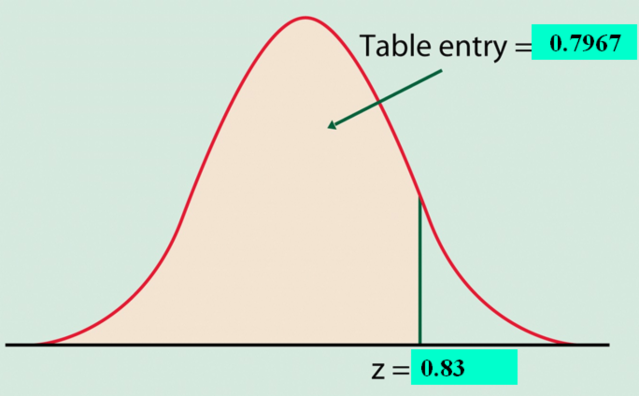
SND Table

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

CAMBRIDGE
Judge Business School

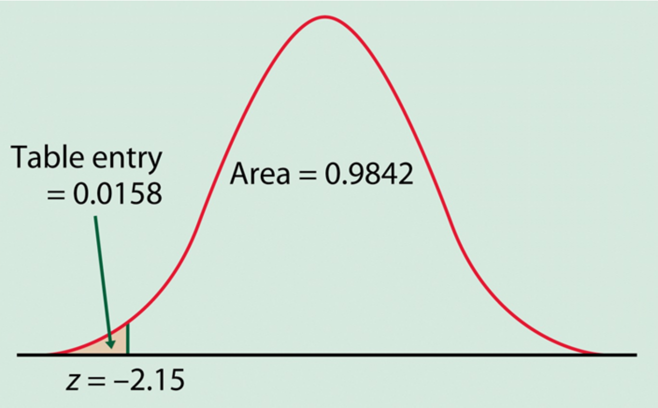
Proportion smaller than 0.83?

What proportion of observations are smaller than 0.83?



Proportion greater than −2.15?

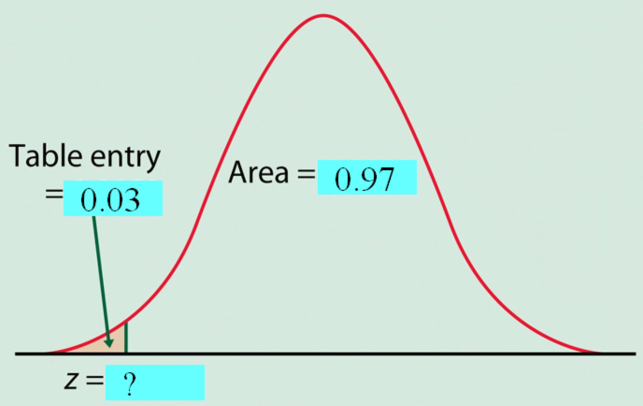
What proportion of observations are greater than −2.15?



Inverse of SND

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

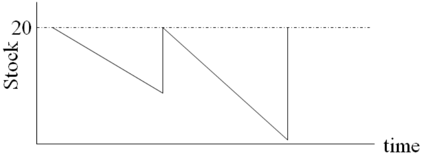
Z Value that cumulates 3% of probability



Example

Inventories in a dealership

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



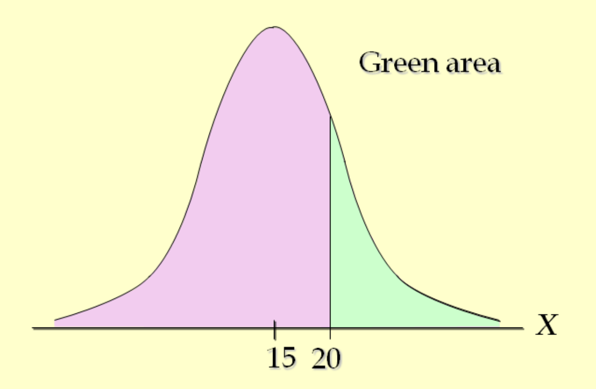
- Sales are lost due to stockouts
- Known that demand (X) within the month is normally distributed with a mean of 15 and a standard deviation of 6
- What is the probability of a stockout?

CAMBRIDGE Judge Business School

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* $z = 0.83$

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

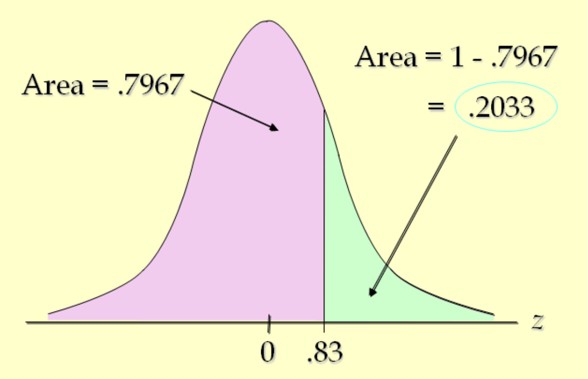
- Probability of stockout $= Pr(X > 20) = 0.2$

CAMBRIDGE Judge Business School

Using the Standard Normal distribution (cont'd)

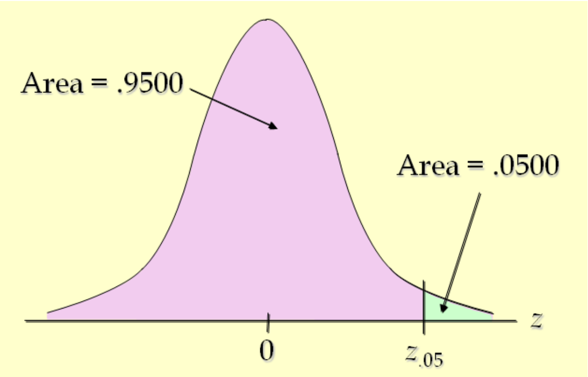
Solving for the stockout probability (cont'd)

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



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

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

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

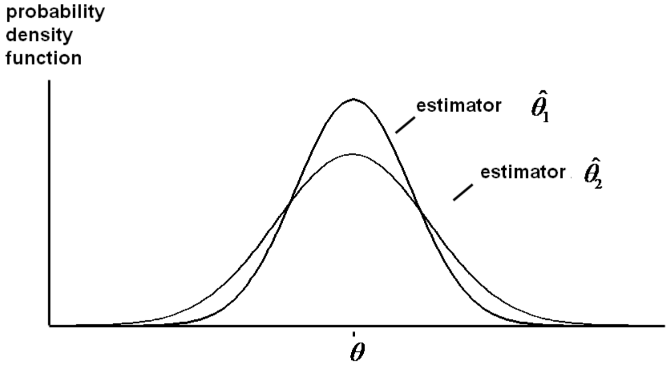
Efficiency

- 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

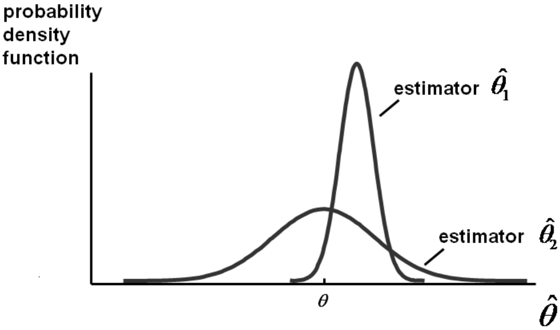
Estimators

Unbiasedness and Efficiency



Estimators

Conflict between unbiasedness and efficiency

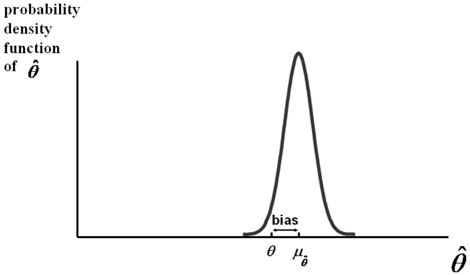


CAMBRIDGE Judge Business School

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

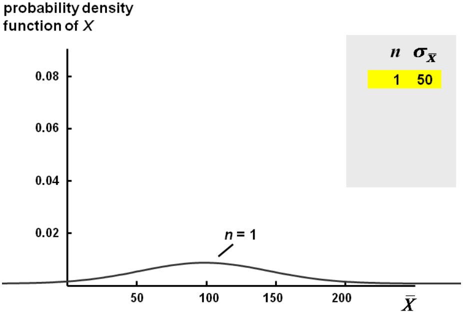
Large sample (asymptotic) properties of estimators

- 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

CAMBRIDGE Judge Business School

Asymptotic properties of Estimators

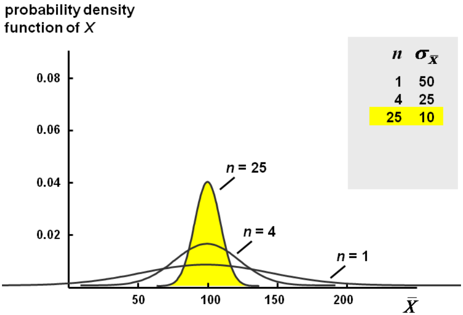
Effect of increasing the sample size on the distribution of $\bar{X}$



- Assume $E(X) = \mu_X = 100$ and $\sigma_X = {\sigma_X}^2 = 50$
- 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)

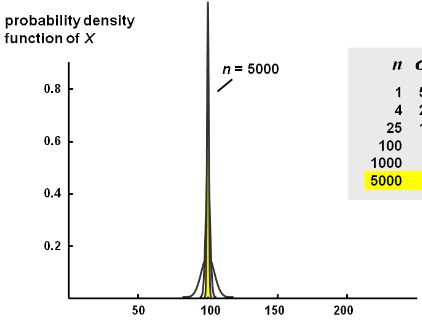


n	$\sigma_{\bar{X}}$
1	50
4	25
25	10

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

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
- $\sigma^2_{\bar{X}} \rightarrow 0$
- The sample mean is a consistent estimator of the population mean.

CAMBRIDGE Judge Business School

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:

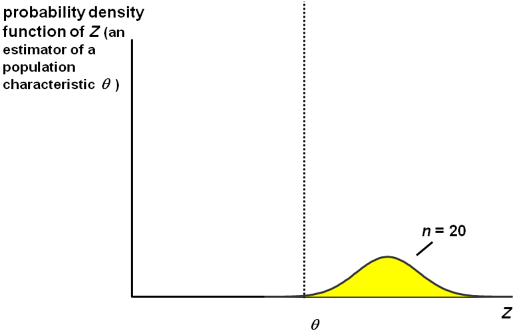
$$plim(\hat{\theta}) = \theta$$

i.e.,

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

Asymptotic properties of Estimators

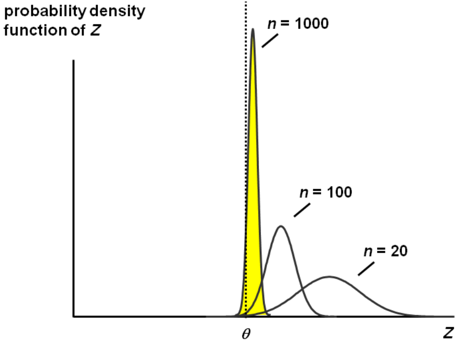
Example: Estimator biased in finite samples but consistent



- $\hat{\theta}$ 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

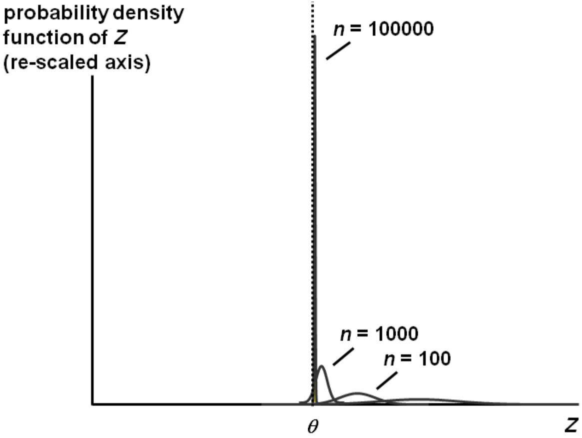
Example: biased in finite samples but consistent (cont'd)



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 $(Y_1, Y_2, .., Y_n)$, $Y_i =$ is the value of the r.v. pertaining to the i^{th} entity sampled

Sampling and Sampling distribution

Distribution of $Y_1, ..., Y_n$ 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, $\{Y_i\}$, $i = 1, ..., n$ 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)

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

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[(Y_i - \mu_Y)(Y_j - \mu_Y)\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(Y_i, Y_j) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_Y^2$$

$$= \frac{\sigma_Y^2}{n}$$

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

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

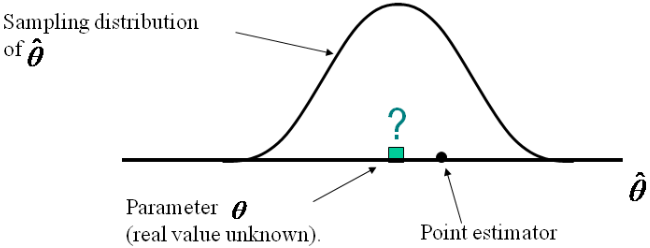
Point and Interval estimation

Point estimation

- A point estimator estimates the value of an unknown parameter in a population using a *single value*
- But we deal with random variables and therefore cannot have certainty
- Way forward?

Point and Interval estimation

Point estimation - illustration



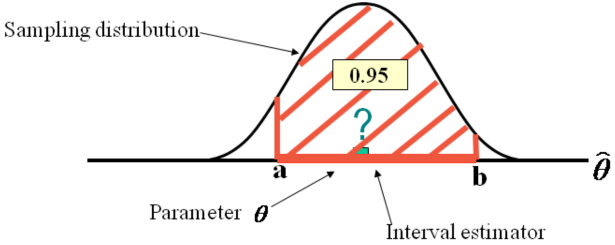
Point and Interval estimation

Interval estimation

- An interval estimator estimates the unknown parameter using a (small) *interval*, ...
- ...and the associated (high) probability that the population parameter is contained in that interval
- This takes account of sampling. The sample (to which the estimator is applied to obtain an estimate) is random
- Q: What is the the smallest interval with a sufficiently high probability - the most informative *interval*?

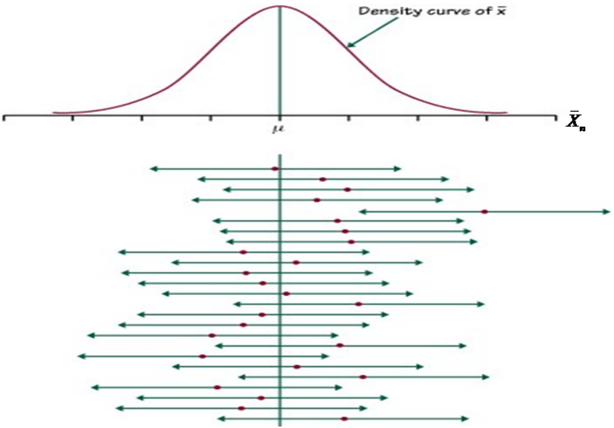
Point and Interval estimation

Interval estimation - illustration



Point and Interval estimation

Interval estimation - illustration 2



Point and Interval estimation

Q: Interval estimation

- Question: Why not define an interval we can be *certain* of containing the true value?
- The only certain interval is $[-\infty, +\infty]$!

Point and Interval estimation

Confidence interval for a parameter θ ...

- ... is an interval on the line (the space in which θ can lie) that, given the sampling distribution of the estimator $\hat{\theta}$, contains θ with a specified (sufficiently high) probability
  - e.g., What is the interval $[a, b]$ that will contain θ with probability of, say, 0.95 (i.e., $a \leq \theta \leq b$ with probability 0.95)?
  - Find a and b, and you have an interval estimate: $[a, b]$ is the 95% confidence interval for θ
  - The price of defining a (small) interval $[a, b]$, and not $(-\infty, \infty)$ is the (5%) probability that you necessarily allow that your interval estimate may be wrong (does not contain θ)
  - This probability split is yours to make: $99\% - 1\%$ or $90\% - 10\%$, any other, depending on the *probability of being wrong* that you can live with

CAMBRIDGE Judge Business School

Point and Interval estimation

Confidence interval and Critical region for a test of hypothesis

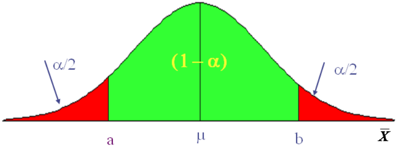
- So, to test your (well reasoned) hypothesis about the unknown θ, you need to fix a and b;
- the region in the parameter space (the real line) outside $[a, b]$ is the critical region for your test
  - What you are really asking is: Is the difference between your hypothesized θ and the estimated $\hat{\theta}$ attributable to the randomness of sampling?
  - Or is the difference between θ and $\hat{\theta}$ too large for it to be merely due to sampling variation?
  - If so, what should you do with your pet theory?

CAMBRIDGE Judge Business School

Point and Interval estimation

Confidence interval for μ

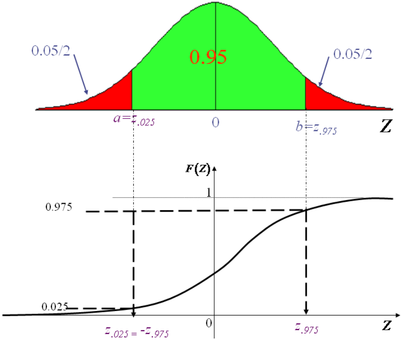
- Assume $X \sim N(\mu, \sigma)$ and that σ is known (or that n is large)
- $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ So the test statistic $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = Z \sim N(0,1)$
- What probability (α) that your best interval estimate is wrong can you live with? (in testing hypotheses, α will be referred to as the size of the test). Let us fix $\alpha = 5\%$
- What are that values a and b, (with reference to $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = Z \sim N(0,1)$) such that $Pr(a \leq \mu \leq b) = 1 - \alpha$?



CAMBRIDGE Judge Business School

Point and Interval estimation

95% Confidence interval for μ (2)



$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = Z \sim N(0,1); \qquad \alpha = 5\%;$$

find: $Pr(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\alpha/2}) = 0.95$

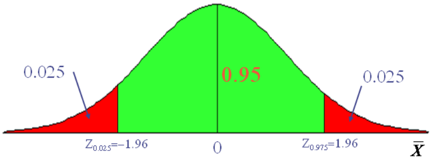
CAMBRIDGE Judge Business School

Point and Interval estimation

95% Confidence interval for μ (3)

$$Pr(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \bar{X} \leq \mu \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \bar{X}) = 0.95$$

- From the standard Normal Table:
  $z_{.025} = -1.96 \quad z_{.975} = 1.96$
- $Pr(-1.96\frac{\sigma}{\sqrt{n}} + \bar{X} \leq \mu \leq 1.96\frac{\sigma}{\sqrt{n}} + \bar{X}) = 0.95$



CAMBRIDGE Judge Business School

Tests of hypothesis

Old example to illustrate types of errors in testing hypotheses

- A rare disease infects 1 person in a 1000
- There is good but imperfect test
- 99% of the time, the test identifies the disease
- 2% of uninfected patients also return a positive test result
- Null Hypothesis H_0: Patient has the disease
- Alternate Hypothesis H_a: Patient does not
  - Q: Why not choose as H_0: Patient does not has the disease?

CAMBRIDGE Judge Business School

Tests of hypothesis

Test of hypotheses example: Joint distribution

	A: patient has disease	$\overline{A}$: patient does not have disease	
B: patient tests positive	0.00099	0.01998	$P(B) = 0.02097$
$\overline{B}$: patient does not test positive	0.00001	0.97902	$P(\overline{B}) = 0.97903$
	$P(A) = 0.001$	$P(\overline{A}) = 0.999$	1

H_0: Patient has the disease H_a: Patient does not

CAMBRIDGE Judge Business School

Tests of hypothesis

Test of hypotheses example: correct decisions

	A: Patient has disease	$\overline{A}$: Patient does not have disease
B: Tests positive	If you do not reject H_0: **Correct decision**	
$\overline{B}$: patient does not test positive		If you reject H_0: **Correct decision**

H_0: Patient has the disease H_a: Patient does not

CAMBRIDGE Judge Business School

Tests of hypothesis

Test of hypotheses example: Type I error

	A: Patient has disease	$\overline{A}$: Patient does not have disease
B: Tests positive	Correct decision	
$\overline{B}$: Does not test positive	If you rejected (the true) H_0, Type I error	Correct decision

H_0: Patient has the disease H_a: Patient does not

CAMBRIDGE Judge Business School

Tests of hypothesis

Test of hypotheses example: Type II error

	A: Patient has disease	$\overline{A}$: Patient does not have disease
B: Tests positive	Correct decision	If you did not reject (the false) H_0, Type II error
$\overline{B}$: Does not test positive	Type I error	Correct decision

H_0: Patient has the disease H_a: Patient does not

Tests of hypothesis

H_0: Patient has the disease H_a: Patient does not

	Patient has disease	Patient does not
B: Tests positive	Correct decision	Prob(Type II error) =0.01998/0.999 = .02 = 2%
$\overline{B}$: Does not test positive	Prob(Type I error) =0.00001/0.001 = 0.01 = 1%	Correct decision

- $P(\text{Error type I}) = P(\text{Reject } H_0 | H_0 \text{ true}) = \alpha = \textit{Size of test}$
- $P(\text{Error type II}) = P(\text{Not Reject } H_0 | H_0 \text{ false}) = \beta$
- $1 - P(\text{Error type } II) = 1 - \beta = \textit{Power of test}$

CAMBRIDGE Judge Business School

Tests of hypothesis

Hypothesis tests - two points

- First:
  - Probability of Type I error, is the size of the test, α and is 1% in this example
  - Can it be changed? How?
- Second:
  - We can never have enough evidence to *accept* a null hypothesis
  - We suspend judgement if the evidence is against the alternative
  - We can only *reject* or *not reject* the null

Test of hypothesis: an example

A hypothesis about the impact of discounts on sales of automobiles

- Increased sales of Citreons after discount
- $\mu = 1200$ *hypothesized* increase in UK sales of Citreons with discount
- $\sigma = 300$ *assumed known* population st. dev. of increase in sales with discount
- X random variable - increase in sales of Citreons after discount
- A sample of 100 discount episodes observed: $\bar{X} = 1265$
- Frame a test:

CAMBRIDGE Judge Business School

Test of hypothesis: an example

Possible hypothesis tests for the mean: One or two tailed

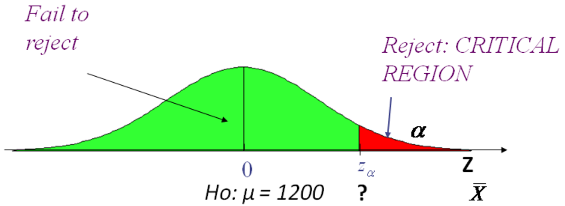
- $H_0: \ E(X) = \mu$ Vs. $H_a: \ E(X) > \mu$ (1-sided, $>$)
- $H_0: \ E(X) = \mu$ Vs. $H_a: \ E(X) < \mu$ (1-sided, $<$)
- $H_0: \ E(X) = \mu$ Vs. $H_a: \ E(X) \neq \mu$ (2-sided)

CAMBRIDGE Judge Business School

Test of hypothesis: an example

Q: Sales of automobiles

- $H_0 : \mu = 1200$, $H_a : \mu > 1200$
- Find a and b, using sample estimate $\bar{X}$, such that $Pr(a \leq \mu \leq b) = 1 - \alpha$
- i.e., a and b, such that Under H_0: Pr(test statistic *does not lie* in the critical region)=$1 - \alpha$
- $Pr\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_\alpha\right)$=1-$\alpha$ *i.e.,* $Pr\left(\frac{\bar{X}-1200}{\frac{300}{\sqrt{100}}} \leq 1.645\right)$=0.95



Test of hypothesis: an example

Q: Sales of automobiles

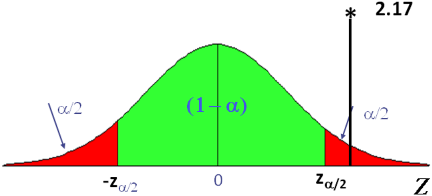
- The critical region is: $Pr\left(\frac{\bar{X}-1200}{\frac{300}{\sqrt{100}}} > 1.645\right)$
- Note that this is a one-tail test, so all 5% is on the right tail
- In this case: $(1265 - 1200)/30 = 2.17 > 1.645$
- So we reject the null hypothesis

CAMBRIDGE Judge Business School

Test of hypothesis: an example

Two tailed test: Sales of automobiles

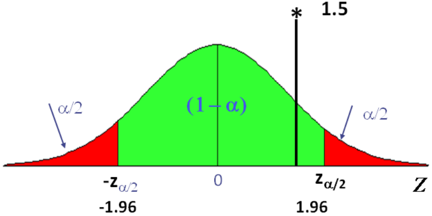
- $H_0 : \mu = 1200$, $H_a : \mu \neq 1200$
- From the SND table, $z_{\alpha/2} = z_{.025} = 1.96$



Test of hypothesis: an example

Two tailed test: Sales of automobiles (2)

- Suppose our sample estimate is different: $\bar{X} = 1245$



Test of hypothesis: an example

p-value

- The significance level of a test is the pre-specified probability of incorrectly rejecting the null, when the null is true
- e.g., if the pre-specified significance level is 5% (size of test):
    - you reject the null hypothesis in a two-tailed test if $|\text{standardised test statistic}| \geq 1.96$
- p-value = *probability* of drawing a statistic (e.g. $\bar{Y}$) *at least as adverse to the null* as the value actually computed with your data, assuming that the null hypothesis is true
    - If significance level is 5%, you reject the null hypothesis if $p \leq 0.05$
    - The p-value is sometimes called the *marginal significance level*
    - It is better to report the p-value, than simply whether a test rejects or not
    - p-value contains more information than “reject/not reject”

CAMBRIDGE Judge Business School

Test of hypothesis: an example

Different Confidence Intervals

1-α	Confidence interval
0.5	$(\bar{X} - 0.67\frac{\sigma}{\sqrt{n}}), \bar{X} + 0.67\frac{\sigma}{\sqrt{n}})$
0.9	$(\bar{X} - 1.64\frac{\sigma}{\sqrt{n}}), \bar{X} + 1.64\frac{\sigma}{\sqrt{n}})$
0.95	$(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}), \bar{X} + 1.96\frac{\sigma}{\sqrt{n}})$
0.99	$(\bar{X} - 2.57\frac{\sigma}{\sqrt{n}}), \bar{X} + 2.57\frac{\sigma}{\sqrt{n}})$
0.999	$(\bar{X} - 3.27\frac{\sigma}{\sqrt{n}}), \bar{X} + 3.27\frac{\sigma}{\sqrt{n}})$

The more the degree of certainty (lower Pr(Type I error)) needed, the larger the interval

CAMBRIDGE Judge Business School

Tests of hypothesis: Power of the test

Type II errors: simulation

- You commit a type II error if you *do not reject a false Null*
- Error type II occurs if you *do not reject a false Null*
- $P[\text{Reject} \quad H_0|H_0 \ \text{ false}] = 1 - P[\text{Error type II}] =$ Power of the test
- Experiment to illustrate:
  - Generate data through i.i.d. draws from $N(\mu, 1)$ (simple random sampling)
  - Keep $\sigma^2 = 1$; but different values of μ in the interval $[-2, 2]$
  - Always test the null: $H_0 : \mu = 0$ against alternative: $H_a : \mu \neq 0$
  - Aim: determine the power of the test, i.e., the prob of not making type II errors, prob. of not rejecting the Null when it is false

CAMBRIDGE Judge Business School

Tests of hypothesis: Power of the test

Probability of not making type II errors: simulation (2)

- Sample mean ($\bar{Y}$) is the estimator of μ
- 3 sample sizes: 10, 100 and 1000, used for estimating ($\bar{Y}$)
- Recall: Samples are from $N(\mu, 1)$ where μ is in $[-2, 2]$
- Critical region:
  - Size of the test fixed at 5%
  - We reject the null ($\mu = 0$) if $|\bar{Y}| > c$, where c is determined by $P[-c \leq \bar{Y} \leq c] = 0.95$, for $\mu = 0$
  - As $\sigma = 1$, and the test is for $\mu = 0$, we have $c = 1.96/\sqrt{n}$
- Note: in most cases in this experiment, the null is false
- 10000 runs of each test. The proportion of times when H_0 is rejected is reported

Tests of hypothesis: Power of the test

Pr(Reject H_0) reported in percentages

DGP: $Y_i \sim N(\mu, 1)$ for $\mu \in [-2, 2]$, including $\mu = 0$

$H_0 : \mu = 0; H_a : \mu \neq 0$

	population mean (actual)	n=10	n=100	n=1000
1-P(EII)	μ = -2	100	100	100
	μ = -1	89	100	100
	μ = -0.2	9.7	51.2	100
	μ = -0.1	6.4	17.3	88.5
	μ = -0.05	5.4	7.3	35.5
EI →	**μ = 0**	**5**	**4.7**	**4.8**
1-P(EII)	μ = 0.05	5.2	7.7	34.4
	μ = 0.1	6.4	16.7	88.3
	μ = 0.2	9.5	51.6	100
	μ = 1	88.4	100	100
	μ = 2	100	100	100

CAMBRIDGE Judge Business School

Tests of hypothesis: Power of the test

Probability of not making type II errors: simulation (3)

- The power of the test increases with the sample size
- The power of the test increases the further away is the true μ from the Null hypothesis μ
- For $n = 1000$ the null is rejected nearly always if DGP has $\mu < -0.1$ or $\mu > 0.1$
- Also: The smaller the probability of a Type 1 error, the greater the probability of Type II error (Show)
- Lesson: Choose the level of significance with care, and use as large a sample as possible

CAMBRIDGE Judge Business School