Universitat de València OpenCourseWare

Multiple Regression Analysis

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$

- 3. Asymptotic Properties

OLS: Finite sample properties

- We have seen that in the MLR:

1. Under MLR.1 through MLR.5, OLS is BLUE: Efficient within the class of linear and unbiased estimators.
2. If we add MLR.6 (normality), then we can derive the *exact* sampling distributions of the OLS estimators (conditional on the ***x***´s). This allows us to do inference by means of the *t* and *F* distribution.

OLS: Finite sample properties

- This are called *finite sample*, *small sample*, or *exact* properties of the OLS estimators.
- The name, *finite sample*, comes because they hold for *any* sample size, *n*.
- Sometimes we are not able to proof such finite sample properties or we cannot obtain the exact sampling distributions of the estimators, so the inference should be only approximate.

OLS: Asymptotic properties

- **Asymptotic properties** or **large sample properties** of estimators and test statistics are not defined for a particular sample size; rather, they are defined as the sample size grows without bound, as $n \rightarrow \infty$.
- Under our assumptions, OLS has satisfactory large sample properties.
- One practically important finding is that even without normality, MLR.6, *t* and *F* statistics have approximately *t* and *F* distributions, at least in large samples.

Consistency

- Unbiasedness of estimators is a finite sample property.
- Unbiasedness, while important, cannot always be achieved.
- In those cases, we switch attention to **consistent estimators**.
- While not all useful estimators are unbiased, virtually all economists agree that **consistency** is a minimum requirement for an estimator.

Consistency

- There are a few different ways to describe consistency, here we focus on an intuitive understanding.
- Consistency means that as $n \rightarrow \infty$, the distribution of the estimator collapses to the parameter value.
- This can be illustrated by the sample mean, $\bar{x}$, as an estimator of the population mean, μ, given a random sample.

Consistency of the sample mean

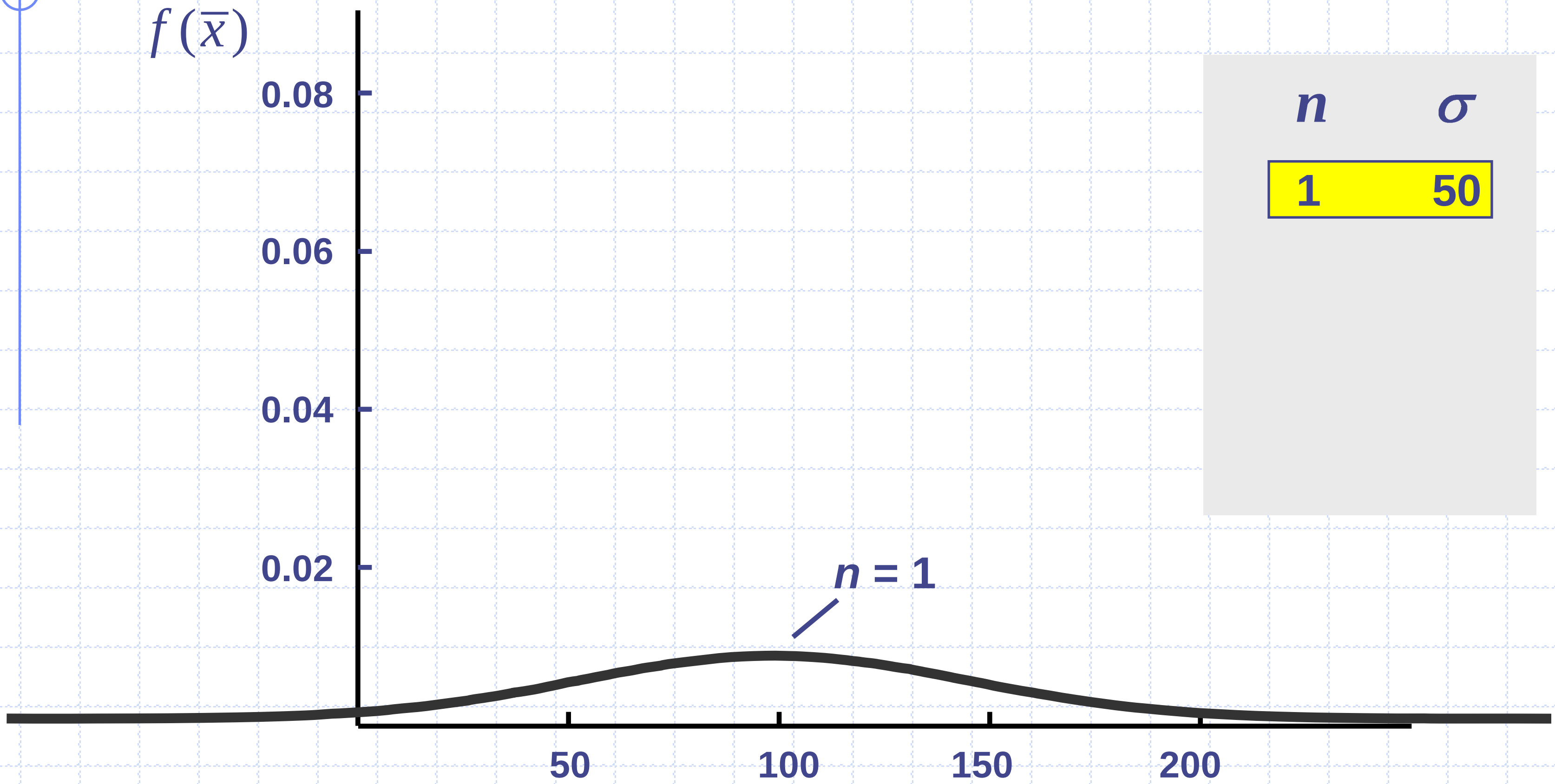
- It is well known that if $x_i \sim$ i.i.d.(μ,σ^2), then

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \Rightarrow \quad \begin{cases} E(\bar{x}) = \mu \\ Var(\bar{x}) = \dfrac{\sigma^2}{n} \end{cases}$$

  so the probability density function of $\bar{x}$, whatever it is, is centered around μ with an standard deviation that decreases at the rate $\sqrt{n}$

- Lets see the effect of increasing *n*.

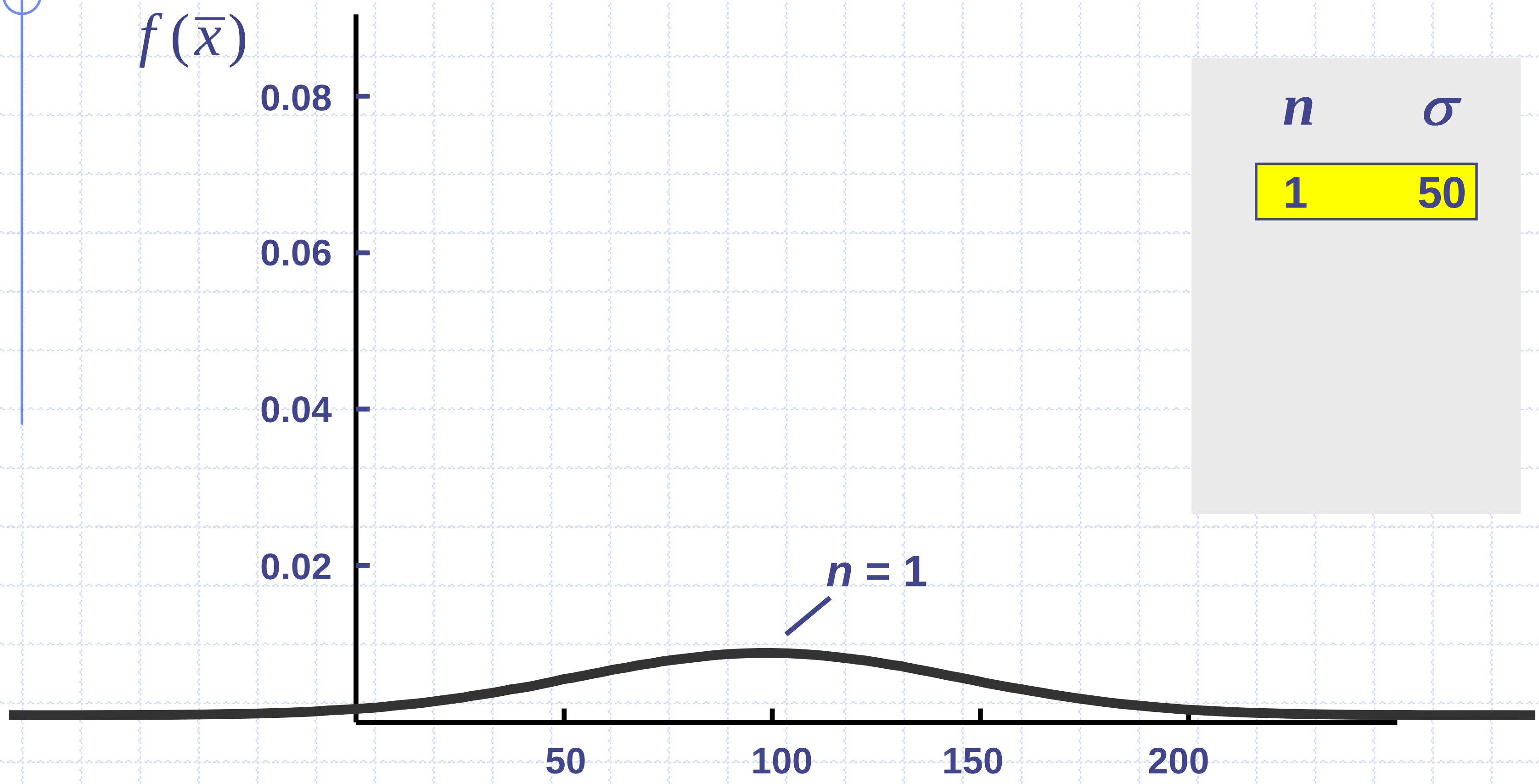
Consistency of the sample mean



n	σ
1	50

Assume μ = 100 and σ = 50. Suppose that we do not know the population mean and we use the sample mean to estimate it. The probability density function of the sample mean will have the same expected value as *x*, but its standard deviation will be $50/\sqrt{n}$.

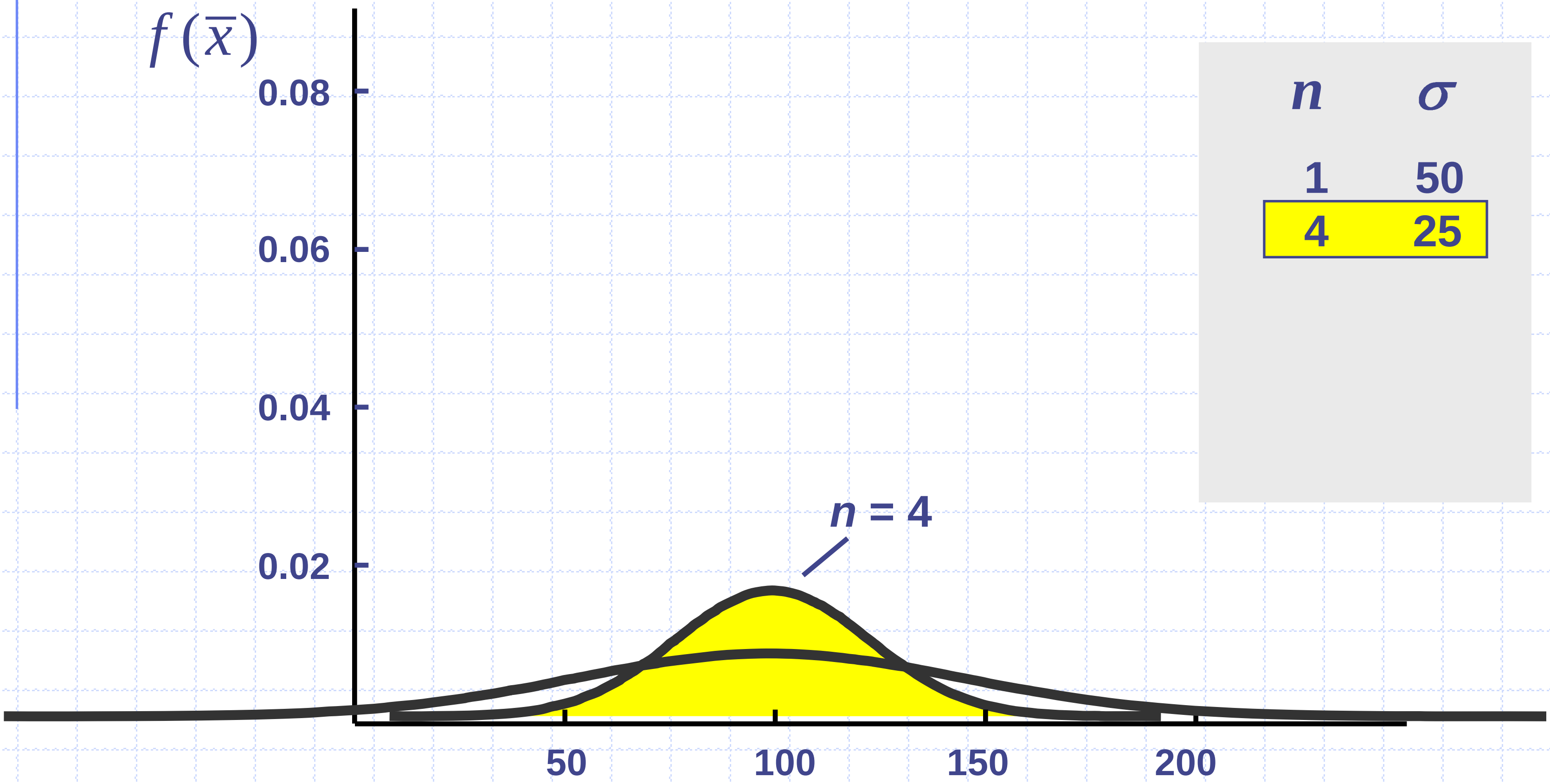
Consistency of the sample mean



The larger is the sample, the smaller will be the standard deviation of the sample mean.

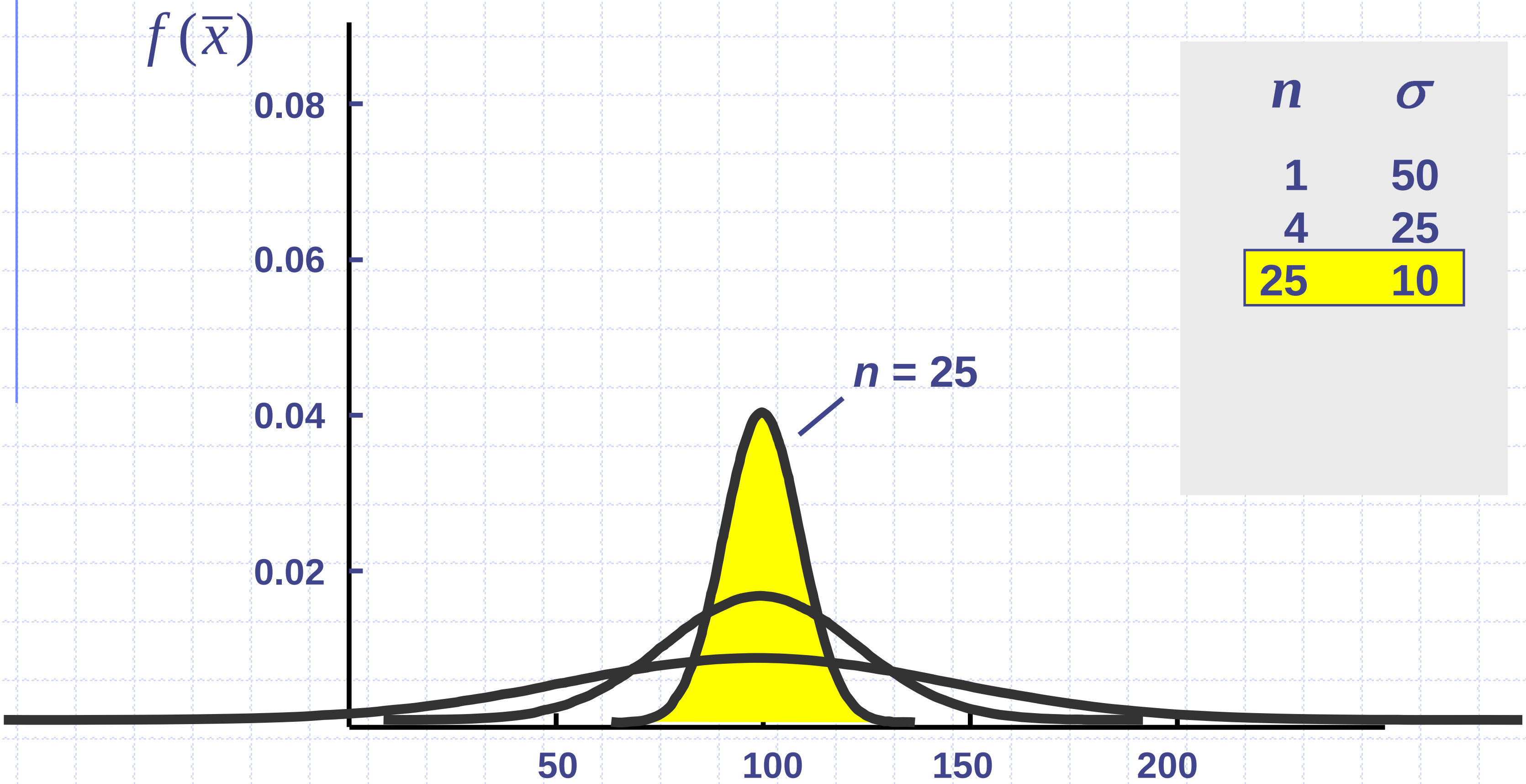
If n = 1, the sample consists of a single observation. $\bar{x}$ is the same as x and its standard deviation is 50.

Consistency of the sample mean



We will see how the shape of the distribution changes as the sample size increases.

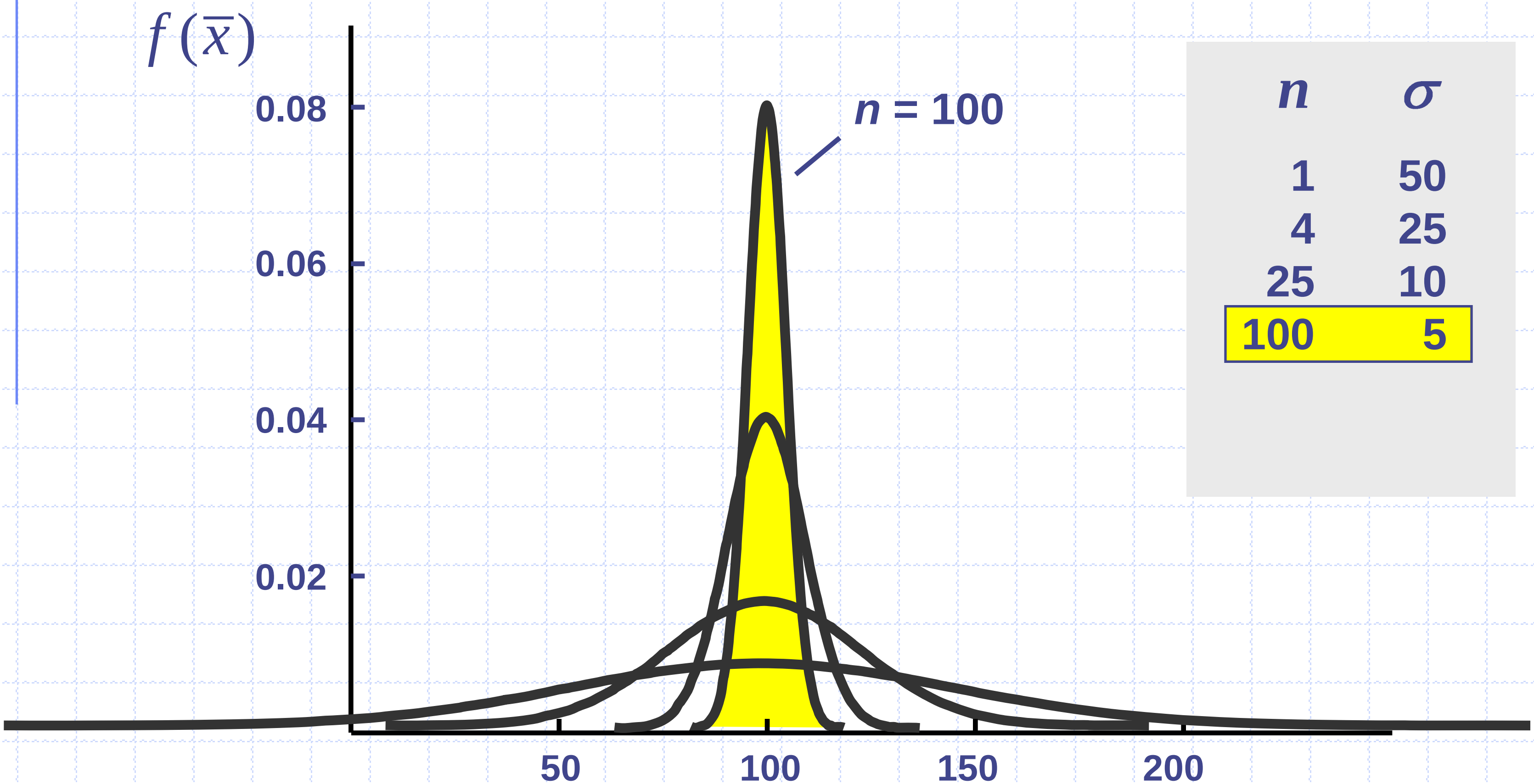
Consistency of the sample mean



n	σ
1	50
4	25
25	10

The distribution becomes more concentrated about the population mean.

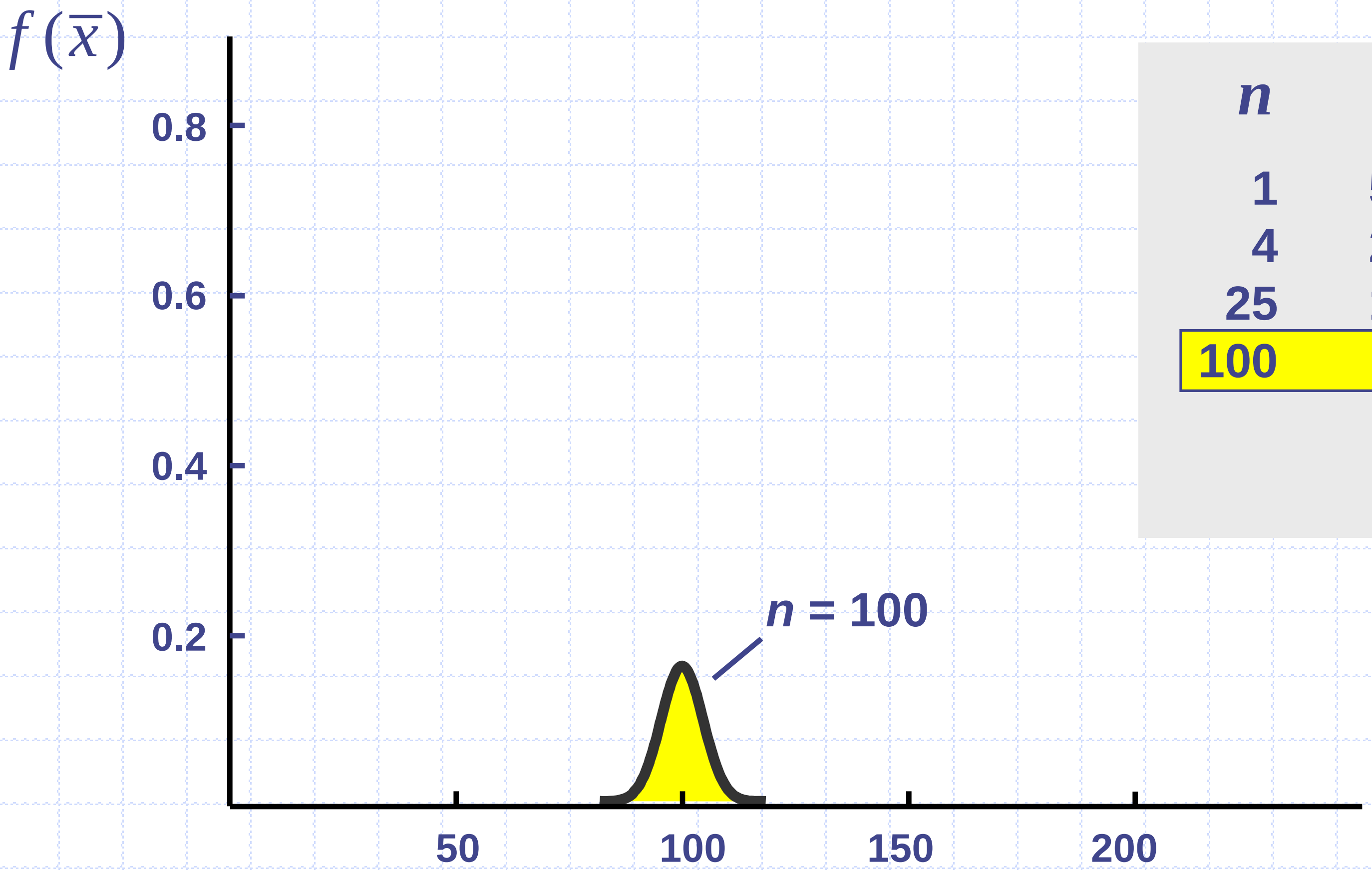
Consistency of the sample mean



n	σ
1	50
4	25
25	10
100	5

To see what happens for n greater than 100, we will have to change the vertical scale.

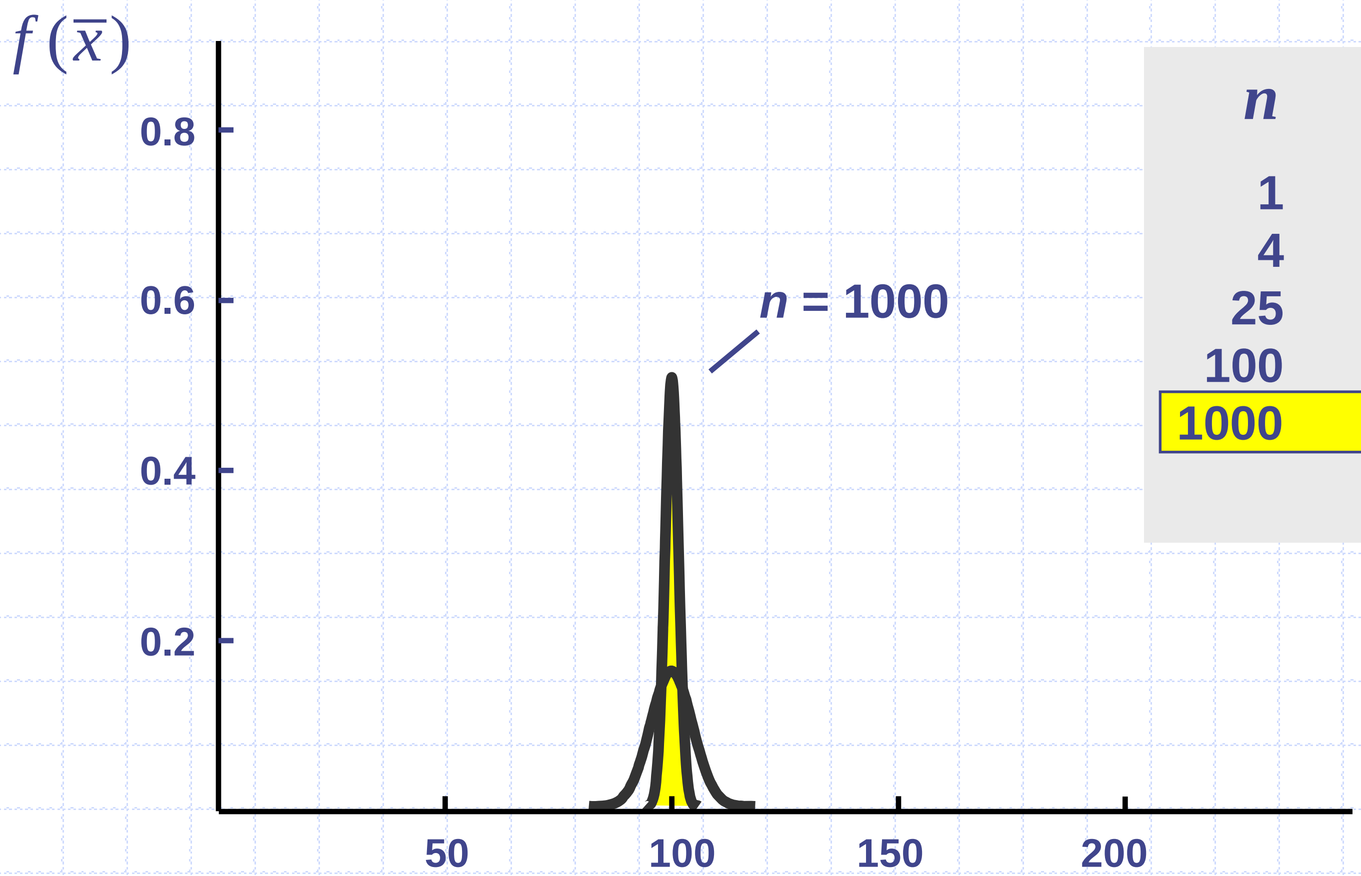
Consistency of the sample mean



n	σ
1	50
4	25
25	10
100	5

We have reduced the vertical scale by a factor of 10.

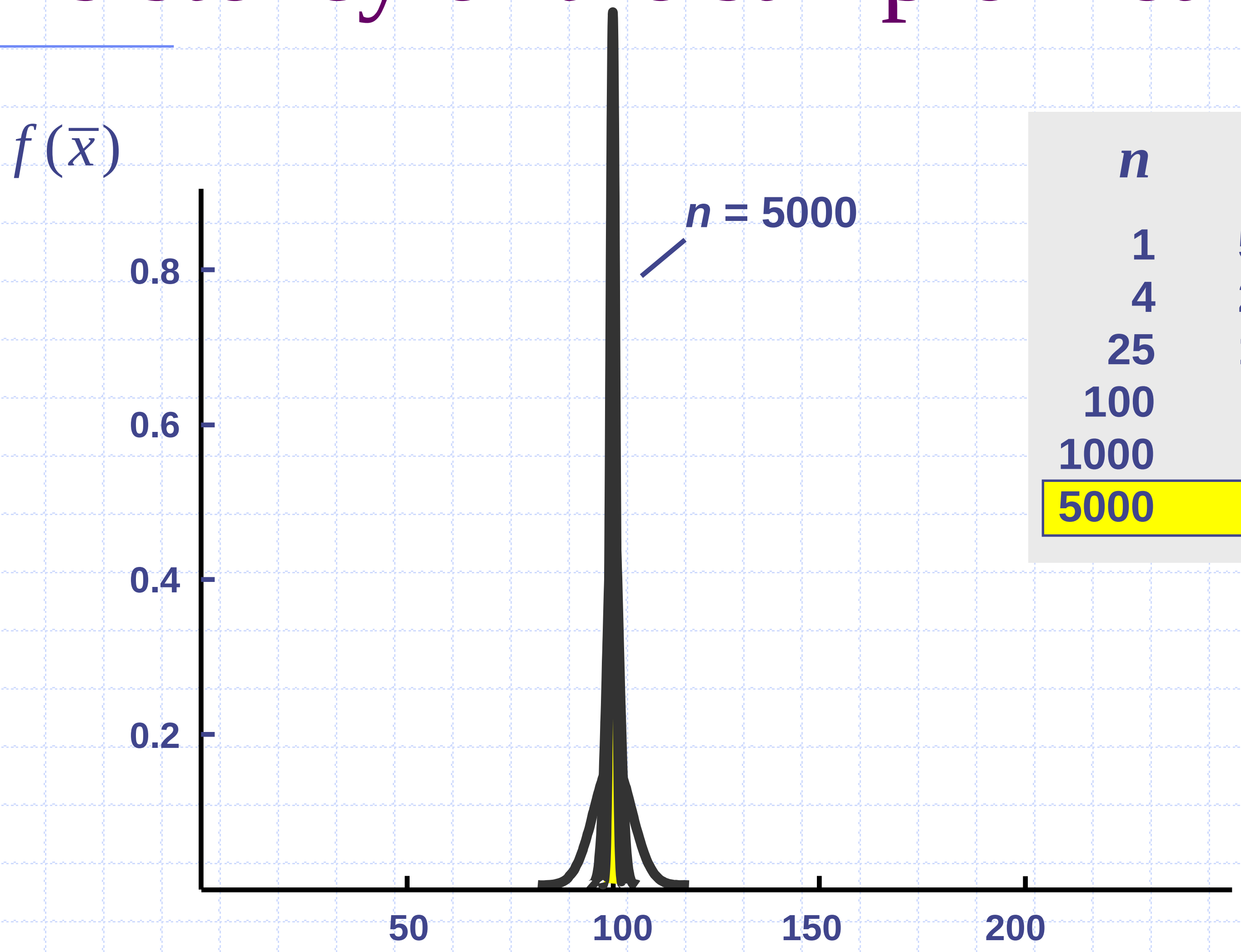
Consistency of the sample mean



n	σ
1	50
4	25
25	10
100	5
1000	1.6

The distribution continues to contract about the population mean.

Consistency of the sample mean



n	σ
1	50
4	25
25	10
100	5
1000	1.6
5000	0.7

Consistency of the sample mean

- In the limit, the variance of the distribution tends to zero. The distribution collapses to a spike at the true value.
- The sample mean is therefore a consistent estimator of the population mean.
- This illustrates the difference between the concepts of unbiasedness and consistency.
- Unbiasedness is a finite-sample concept. The *expected* value of the sample mean is equal to the population mean, but in general its *actual* value will be different.
- Consistency is a large-sample concept. A consistent estimator becomes an increasingly accurate estimator of the population characteristic and in the limit becomes equal to it.

Consistency of the sample mean

Finite samples: $\bar{x}$ is an unbiased estimator of μ

Large samples: $\bar{x}$ is a consistent estimator of μ

$$\operatorname{plim} \bar{x} = \mu$$

As the sample size becomes large, the distribution of the sample mean collapses to a spike located at the true value.

The sample mean is therefore consistent as well as unbiased.

Consistency

- Consistency involves a thought experiment about what happen as the sample size gets large (while at the same time, we obtain numerous random samples for each sample size).
- If obtaining more and more data does not generally get us closer to the parameter value of interest, then we are using a poor estimation procedure.

Consistency of OLS.

- **THEOREM 5.1 CONSISTENCY OF OLS**

  Under assumptions MLR.1 through MLR.4, the OLS estimator $\hat{\beta}_j$ is consistent for β_j, for all $j = 0, 1, 2, \ldots, k$.

  PROOF: A general proof requires matrix algebra, see Appendices D & E.

Consistency of OLS.

- We can however prove consistency of the slope estimator in the SLR model.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x}) u_i}{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

We need to take plim´s to establish consistency.

Law of Large Numbers

- **LAW OF LARGE NUMBERS**:

Let $\{x_i\}_{i=1}^{n}$ a random sample with mean μ, then

$$\operatorname{plim} \bar{x} = \mu$$

In other words, sample moments converge to population moments.

Properties of plim´s

- **PROPERTY PLIM.1:**

  Assume W_n is a consistent estimator for θ, plim $(W_n) = \theta$.

  Then

  $$\text{plim } g(W_n) = g(\text{plim } W_n)$$

  for a continuous function $g(\bullet)$.

- **PROPERTY PLIM.2:**

  If plim $(T_n) = \alpha$ and plim $(U_n) = \beta$

  (i) plim $(T_n + U_n) = \alpha + \beta$

  (ii) plim $(T_n.U_n) = \alpha\beta$

  (iii) plim $(T_n/U_n) = \alpha/\beta$, provided $\beta \neq 0$.

Consistency of OLS.

- We can apply the law of large numbers to the numerator and denominator, which converge in probability to the population quantities, $Cov(x,u)$ and $Var(x)$ respectively.
- Provided $Var(x) \neq 0$, which is assumed in MLR.4, we can use the properties of the *probability limits* to get

$$\text{plim}\hat{\beta}_1 = \beta_1 + \frac{Cov(x,u)}{Var(x)} = \beta_1, \text{ since } Cov(x,u) = 0$$

using $E(u|x) = 0$, SLR 3.

Consistency of OLS.

- The previous arguments show that OLS is consistent in the SLR if we assume only zero correlation between u and x.
- This is also true in the general case, so assumption MLR.3 can be weakened.

Assumptions

- **MLR.3´: ZERO MEAN AND ZERO CORRELATION**

  $E(u) = 0$ and $\mathrm{Cov}(x_j,u) = 0$, for $j = 1,2,\ldots,k$

  For a random sample, this assumption implies that

  $$E(u_i) = 0 \text{ and } \mathrm{Cov}(x_{ij},u_i) = 0,$$
  $$\text{for } j = 1,2,\ldots,k \text{ and } i = 1,2,3,\ldots,n$$

Consistency of OLS.

- Assumption MLR.3 implies MLR.3´, but not vice versa.
- Interestingly, while OLS is unbiased under MLR.3, this is not the case under Assumption MLR.3´.
- This implies that estimators that are biased in finite samples can be consistent.

Inconsistencies in OLS

- Just as failure of $E(u|x_1,x_2,\ldots,x_k) = 0$ causes bias in the OLS estimators, correlation between u and *any* of $x_1, x_2,\ldots,x_k$ generally causes all of the OLS estimators to be inconsistent.
- This simple but important observation is often summarized as: *if the error is correlated with any of the independent variables, then OLS is biased and inconsistent*.
- This is very unfortunate because it means that any bias persist as the sample size grows.

Deriving an Inconsistency

- Consider the asymptotic analog of the omitted variable bias.

  Population:

  $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

  Estimated model:

  $$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

Deriving an Inconsistency

$$\tilde{\beta}_1 = \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2} + \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)u_i}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

then

$$\operatorname{plim}\tilde{\beta}_1 = \beta_1 + \beta_2\delta_1 \quad \text{where} \quad \delta_1 = \frac{Cov(x_1, x_2)}{Var(x_1)}$$

Thus, for practical purposes, we can view the inconsistency as being the same as the bias.

Inconsistencies in OLS

- An important point about inconsistency in OLS estimators is that, by definition, the problem does not go away by adding more observations to the sample.

Asymptotic normality

- Consistency of an estimator is an important property, but it alone does not allow us to perform statistical inference.
- Simply knowing that the estimator is getting closer to the population value as the sample size grows does not allow us to test hypothesis about the parameters.
- For testing, we need the sampling distribution of the OLS estimators.
- The exact normality of OLS estimators hinges crucially on the normality of the distribution of u in the population.

Asymptotic normality

- If the errors $u_1, u_2, \ldots, u_n$ are random draws from some distribution other than the normal, the $\hat{\beta}_j$ will not be normally distributed, which means that the *t* statistics will not have *t* distributions and the *F* statistics will not have *F* distributions.
- This is a potentially serious problem because our inference hinges on being able to obtain critical values or *p*-values from the *t* of *F* distributions.
- We know that normality plays no role in the BLUE property of OLS under the Gauss-Markov assumptions. But exact inference requires MLR.6.

Asymptotic normality

- The question is: Can we perform inference without MLR.6?
- YES. Even though the *u*, o alternatively *y* conditional on ***x***´s, are not from a normal distribution, we can use the **central limit theorem** to conclude that the OLS estimators satisfy **asymptotic normality**, which means they are approximately normally distributed in large enough samples sizes.

Central Limit Theorem

- **CENTRAL LIMIT THEOREM**:

  Let $\{x_i\}_{i=1}^{n}$ a random sample with mean μ and variance σ^2, then

$$\sqrt{n}\,\frac{\bar{x}-\mu}{\sigma} = \frac{\bar{x}-\mu}{sd(\bar{x})} \overset{a}{\sim} N(0,1)$$

  In other words, standardized averages converge to the standard normal distribution.

Asymptotic Normality of OLS.

- **THEOREM 5.2 ASYMPTOTIC NORMALITY OF OLS**

  Under the Gauss-Markov assumptions MLR.1 through MLR.5,

(*i*) $\sqrt{n}\left(\hat{\beta}_j - \beta_j\right) \overset{a}{\sim} N\left(0, \frac{\sigma^2}{a_j^2}\right)$, where $\frac{\sigma^2}{a_j^2} > 0$ is the **asymptotic variance** of $\sqrt{n}\left(\hat{\beta}_j - \beta_j\right)$; for the slope coefficients,

$a_j^2 = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{r}_{ij}^2\right)$, where the $\hat{r}_{ij}$ are the residuals from regressing x_j on the other independent variables. We say that $\hat{\beta}_j$ is *asymptotically normally distributed*.

Asymptotic Normality of OLS.

(ii) $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2 = Var(u)$;

(iii) For each j,

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \overset{a}{\sim} N(0,1)$$

where $se(\hat{\beta}_j)$ is the usual OLS standard error.

Asymptotic Normality of OLS.

- Theorem 5.2 is useful because the normality assumption MLR.6 has been dropped; the only restriction on the distribution of the error is that it has finite variance, something we will always assume. We have also assumed zero conditional mean and homoskedasticity of u.
- Note that the distribution that appears in Theorem 5.2 is the *normal* not the t_{n-k-1} as in Theorem 4.2. However from a practical perspective, this difference is irrelevant and we can write

Asymptotic Normality of OLS.

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \overset{aprox}{\sim} t_{n-k-1}$$

since t_{n-k-1} approaches the standard normal distribution as the *df* gets large.

- This result tell us that t testing and the construction of confidence intervals are carried out *exactly* as under the CLM assumptions.

Asymptotic Normality of OLS.

- If n is not very large, then the t distribution can be a poor approximation to the normal when u is not normally distributed.
- Unfortunately, there are no general prescriptions on how big n must be before the approximation is good enough.
- Theorem 5.2 requires homoskedasticity. If $Var(y|\mathbf{x})$ is not constant, the usual t statistics and confidence intervals are invalid, no matter how large n is.
- From Theorem 5.2 $\operatorname{plim}\hat{\sigma}^2 = \sigma^2$, this implies $\operatorname{plim}\hat{\sigma} = \sigma$

Asymptotic Normality of OLS.

- Since the estimated variance of $\hat{\beta}_j$ is

$$V\hat{a}r(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{\mathrm{SST}_j\left(1 - R_j^2\right)}$$

$\hat{\sigma}$ appears in the standard error for each $\hat{\beta}_j$,

$$se(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{\mathrm{SST}_j\left(1 - R_j^2\right)}}$$

when u is not normal this is usually called **asymptotic standard error**.

Asymptotic Normality of OLS.

- It can be shown that

$$se(\hat{\beta}_j) \approx \frac{c_j}{\sqrt{n}}$$

where c_j is a positive constant that does not depend on n.

- This is a useful **rule of thumb**: standard errors can be expected to shrink at a rate that is the inverse of the *square root* of the sample size.

Asymptotic Normality of OLS.

- The asymptotic normality of the OLS estimators also implies that the F statistics have approximate F distributions in large sample sizes.
- Thus, for testing exclusion restrictions or other multiple hypothesis, nothing changes from what we have done before.

Asymptotic Efficiency of OLS

- We know that, under Gauss-Markov assumptions, the OLS estimators are BLUE.
- It can be shown that OLS is also **asymptotically efficient** within the class of the consistent and asymptotically normal estimators.

Asymptotic Efficiency of OLS

- **THEOREM 5.3 ASYMPTOTIC EFFICIENCY OF OLS**

  Under the Gauss-Markov assumptions, let $\tilde{\beta}_j$ denote the class of estimators that are consistent and asymptotically normal and let $\hat{\beta}_j$ denote the OLS estimators. Then for $j = 0, 1, 2, \ldots, k$, the OLS estimators have the smallest asymptotic variances:

$$AVar\sqrt{n}\;\hat{\beta}_j - \beta_j \;\leq AVar\sqrt{n}\;\tilde{\beta}_j - \beta_j$$

PROOF: A general proof requires matrix algebra and advance asymptotic analysis.