Chap. 5: Joint Probability Distributions

- **Probability modeling of several RV's**
- **We often study relationships among variables.**
  - **Demand on a system = sum of demands from subscribers ($D = S_1 + S_2 + \ldots. + S_n$)**
  - **Surface air temperature & atmospheric CO_2**
  - **Stress & strain are related to material properties; random loads; etc.**
- **Notation:**
  - **Sometimes we use X_1 , X_2 ,…., X_n**
  - **Sometimes we use X, Y, Z, etc.**

Sec 5.1: Basics

- **First, develop for 2 RV (X and Y)**
- **Two Main Cases**

I. Both RV are discrete

II. Both RV are continuous

I. (p. 185). <u>Joint Probability Mass Function</u> (pmf) of X and Y is defined for all pairs (x,y) by

$$p(x, y) = P(X = x \text{ and } Y = y)$$

$$= P(X = x, Y = y)$$

- **pmf must satisfy:**

$$p(x,y) \geq 0 \text{ for all } (x,y)$$

$$\sum_x \sum_y p(x,y) = 1$$

- **for any event A,**

$$P\big((X,Y) \in A\big) = \sum_{(x,y)\in A} p(x,y)$$

Joint Probability Table:

Table presenting joint probability distribution:

			y	
		1	2	3
x	1	.10	.15	.22
	2	.30	.10	.13

- **Entries:** $p(x, y)$
- **P(X = 2, Y = 3) = .13**
- **P(Y = 3) = .22 + .13 = .35**
- **P(Y = 2 or 3) = .15 + .10 + .35 =.60**

- **The <u>marginal pmf</u> X and Y are**

$$p_X(x) = \sum_y p(x, y) \text{ and } p_Y(y) = \sum_x p(x, y)$$

			y		
		1	**2**	**3**	
x	**1**	**.10**	**.15**	**.22**	**.47**
	2	**.30**	**.10**	**.13**	**.53**
		.40	**.25**	**.35**	

x	**1**	**2**
$\mathbf{p_X(x)}$	**.47**	**.53**

y	**1**	**2**	**3**
$\mathbf{p_Y(y)}$	**.40**	**.25**	**.35**

II. Both continuous (p. 186)

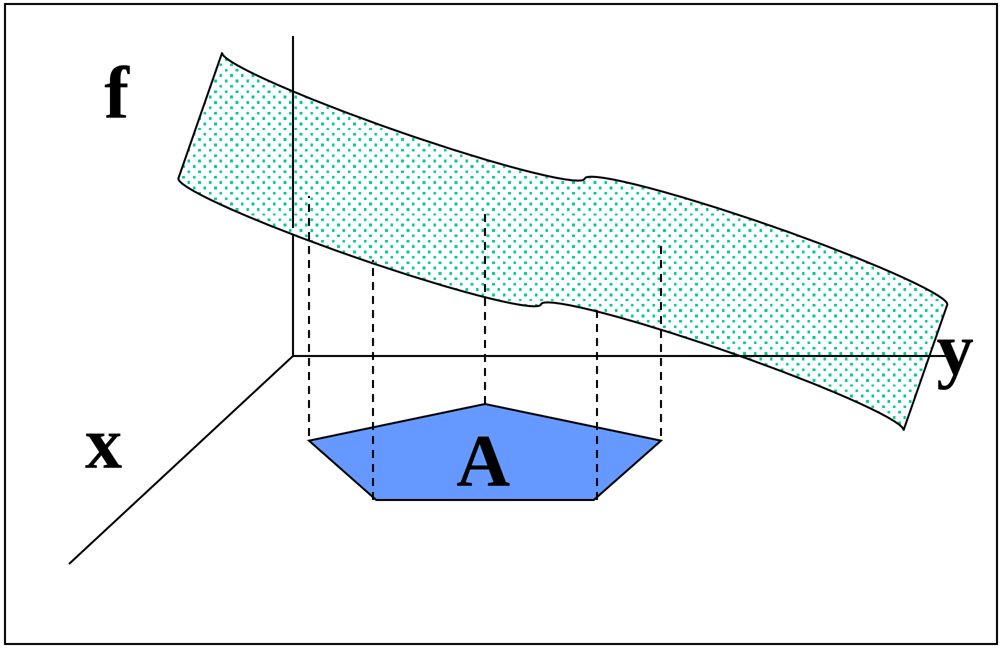
A <u>joint</u> <u>probability density function</u> (pdf) of X and Y is a function f(x,y) such that

- **f(x,y) $\geq$ 0 everywhere**
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$

and $P[(X,Y) \in A] = \iint_A f(x,y)\,dxdy$

pdf f is a surface above the (x,y)-plane

- **A is a set in the (x,y)-plane.**
- $P[(X,Y)\in A]$ **is the volume of the region over A under f. (Note: It is *not* the area of A.)**



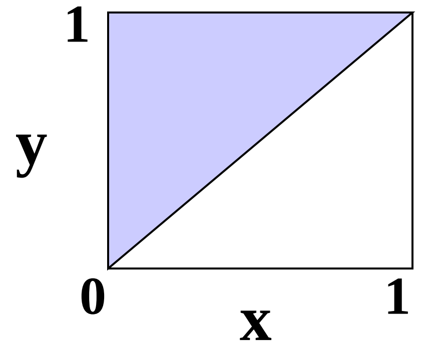
Ex) X and Y have joint PDF

f(x,y) = c x y^2 if 0 < x < y < 1

= 0 elsewhere.

- **Find c. First, draw the region where f > 0.**

$$1 = \int_0^1 \int_0^y cxy^2 \, dx\, dy$$



$$= \int_0^1 \int_x^1 cxy^2 \, dy\, dx$$

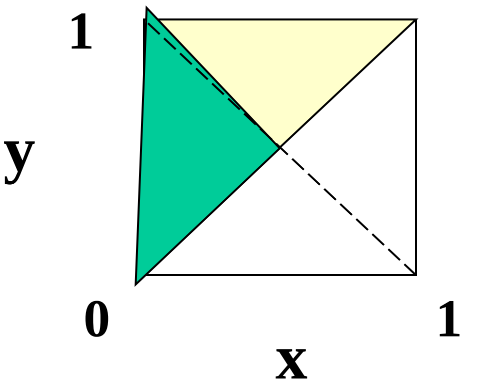
(not $\int_x^1 \int_0^y cxy^2 \, dx\, dy$

$$\int_0^1 \int_0^y cxy^2 \, dx\,dy = c\int_0^1 y^2 [.5x^2 \,|_0^y] dy = c\int_0^1 .5y^4 dy = c/10$$

so, c = 10

- **Find P(X+Y<1)**

First, add graph of x + y =1



$$P(X+Y<1) = \int_0^{.5}\int_0^y 10xy^2 \, dx\,dy + \int_{.5}^1 \int_0^{1-y} 10xy^2 \, dx\,dy$$

$$= \int_0^{.5}\int_x^{1-x} 10xy^2 \, dy\,dx = 10\int_0^{.5} x\left(\frac{y^3}{3}\right]_x^{1-x} dx =$$

$$(10/3)\int_0^{.5} x((1-x)^3 - x^3)dx = .135$$

Marginal pdf (p. 188)

Marginal pdf of X:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx$$

Marginal pdf of Y:

Ex) X and Y have joint PDF

f(x,y) = 10 x y^2 if 0 < x < y < 1 , and 0 else.

For 0 < y < 1:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx = \int_0^y 10xy^2\,dx = 10y^2 \int_0^y x\,dx = 5y^4$$

and $f_Y(y) = 0$ otherwise.

marginal pdf of Y:

$$f_Y(y) = 5y^4 \text{ for } 0 < y < 1 \text{ and is 0 otherwise.}$$

marginal pdf of Y: you check

$$f_X(x) = (10/3)x(1 - x^3) \text{ for } 0 < x < 1$$

$$\text{and is 0 otherwise.}$$

Notes:

1. **x cannot appear in** $f_Y(y)$ **(y can't be in** $f_X(x)$**)**
2. **You must give the ranges; writing** $f_Y(y) = 5y^4$ **is *not* enough.**

Math convention: writing $f_Y(y) = 5y^4$ **with no range means it's right for all y, which is very wrong in this example.**

Remark: Distribution Functions

- **For any pair of jointly distributed RV, the joint <u>distribution function (cdf) of X and Y</u> is**

$$F(x,y) = P(X \le x, Y \le y)$$

defined for all (x,y).

- **For X,Y are both continuous:**

$$f(x,y) = \frac{\delta^2}{\delta x \delta y} F(x,y)$$

wherever the derivative exists.

Extensions for 3 or more RV: by example

X, Y, Z are discrete RV with joint pmf

$$p(x,y,z) = P(X = x, Y = y, Z = z)$$

<u>marginal pmf</u> of X is

$$p_X(x) = \sum_y \sum_z p(x,y,z) \quad \textbf{(= P(X = x))}$$

<u>(joint)</u> <u>marginal pmf</u> of X and Y is

$$p_{XY}(x,y) = \sum_z p(x,y,z) \quad \textbf{(= P(X=x, Y = y))}$$

X, Y, Z are continuous RV with joint pdf f(x,y,z):

marginal pdf of X is

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y, z)\,dydz$$

(joint) marginal pmf of X and Y is

$$f_{XY}(x, y) = \int_{-\infty}^{\infty} f(x, y, z)\,dz$$

Conditional Distributions & Independence

Marginal pdf of X:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

Marginal pdf of Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

Conditional pdf of X given Y=y (h(y) > 0)

$$f(x\,|\,y) = f(x,y)/h(y)$$

Conditional prob for X for y fixed

$$P(X \in A\,|\,Y = y) = \int_A f(x\,|\,y)\,dx$$

Conditional Distributions & Independence

Review from Chap. 2:

- **For events A & B where P(B) > 0, define P(A|B) to be the <u>conditional prob</u>. that A occurs *given* B occurred:**

$$P(A \mid B) = P(A \cap B) / P(B)$$

- **Multiplication Rule:** $P(A \cap B) = P(A)\,P(B|A)$
$= P(B)\,P(A|B)$

- **Events A and B are <u>independent</u> if**

$$P(B|A) = P(B)$$

or equivalently $P(A \cap B) = P(A)\,P(B)$

Extensions to RV

- **Again, first, develop for 2 RV (X and Y)**
- **Two Main Cases**

I. Both RV are discrete

II. Both RV are continuous

I. (p. 193). <u>Conditional Probability Mass Function</u> (pmf) of Y given X = x is

$$p_{Y|X}(y \mid x) = \frac{p(x, y)}{p_X(x)} = \frac{\text{joint}}{\text{marginal of condition}}$$

as long as $p_X(x) > 0.$

- **Note that idea is the same as in Chap. 2**

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

as long as $P(X = x) > 0.$

- **However, keep in mind that we are defining a (conditional) prob. dist for Y for a fixed x**

Example:

			y		
		1	2	3	
x	1	.10	.15	.22	.47
	2	.30	.10	.13	.53
		.40	.25	.35	

x	1	2
$p_X(x)$	.47	.53

y	1	2	3
$p_Y(y)$	.40	.25	.35

Find cond'l pmf of X given Y = 2:

$$p_{X|Y}(x\,|\,y) = \frac{p(x,y)}{p_Y(y)} \quad \textbf{gives} \quad p_{X|Y}(x\,|\,2) = \frac{p(x,2)}{p_Y(2)}$$

So

x	1	2
$p_{X\|Y}(x\|2)$	.15/.25=.60	.10/.25=.40

II. Both RV are continuous

(p. 193). <u>Conditional Probability Density Function</u> (pdf) of Y given X = x is

$$f_{Y|X}(y\,|\,x) = \frac{f(x,y)}{f_X(x)} = \frac{\text{joint pdf}}{\text{marginal pdf of condition}}$$

as long as $f_X(x) > 0.$

The point: $P(Y \in A \,|\, X = x) = \int_A f_{Y|X}(y\,|\,x)dy$

Remarks

- **ALWAYS: for a cont. RV, prob it's in a set A is the integral of its pdf over A:**

  no conditional; use the marginal pdf

  with a condition; use the right cond'l pdf

- **Interpretation: For cont. X, P(X = x) = 0, so by Chap 2 rules,** $P(Y \in A \mid X = x)$ **is meaningless.**
  - **There is a lot of theory that makes sense of this**
  - **For our purposes, think of it as an approximation to** $P(Y \in A \mid X \approx x)$

    that is "given X lies in a tiny interval around x"

Ex) X, Y have pdf

$f(x,y) = 10\,x\,y^2$ if $0 < x < y < 1$, and $= 0$ else.

- **Conditional pdf of X given Y= y:**

$$f_{X|Y}(x \mid y) = f(x,y) / f_Y(y)$$

We found $f_Y(y) = 5y^4$ for $0<y<1$, $= 0$ else. So

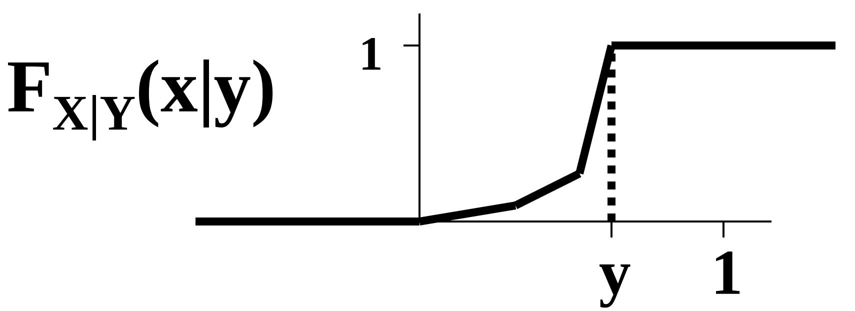
$$f(x \mid y) = \frac{10xy^2}{5y^4} \text{ if } 0 < x < y < 1$$

***Final Answer*: For a fixed y, $0 < y < 1$,**

$f_{X|Y}(x \mid y) = 2x / y^2$ if $0 < x < y$, and $= 0$ else.

(f(x | y) = 2x / y^2 0 < x < y, and = 0 else.)

- **P(X < .2 | Y = .3)** $= \int_0^{.2} 2x / .3^2 dx$
- **P(X < .35 | Y = .3) = 1**
- $F_{X|Y}(x \mid y) = \int_0^x 2t / y^2 dt = x^2 / y^2$ **if 0 < x < y**



Last Time: X, Y have pdf f(x,y)

Marginal pdf of X:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

Marginal pdf of Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Conditional pdf of X given Y=y **($f_Y(y) > 0$)**

$$f_{X|Y}(x|y) = f(x,y) / f_Y(y)$$

Conditional prob for X for y fixed

$$P(X \in A | Y = y) = \int_A f_{X|Y}(x|y)dx$$

Three More Topics

1. **<u>Multiplication Rule for pdf</u>:**

$$f(x,y) = f_{X|Y}(x \mid y)\, f_Y(y) = f_{Y|X}(y \mid x)\, f_X(x)$$

[For events P(A ∩ B) = P(A) P(B|A) = P(B) P(A|B)]

- **Extension, by example:**

$$f(x,y,z) = f_X(x)\, f_{Y|X}(y \mid x)\, f_{Z|XY}(z \mid x,y)$$

[Chap 2: P(A ∩ B ∩ C) = P(A) P(B|A) P(C|A ∩ B)]

2. Independence

Chap. 2: A, B are independent if P(A|B)=P(A) or equivalently, P(A ⋂ B) = P(A)P(B)

- **X and Y are <u>independent RV</u> if and only if**

$$f_{X|Y}(x\,|\,y) = f_X(x)$$

***for all* (x,y) for which f(x,y)>0, or**

$$f(x,y) = f_X(x)\,f_Y(y)$$

***for all* (x,y) for which f(x,y)>0.**

- **i.e., the joint is the product of the marginals.**

2. Independence

More general: $X_1, X_2, \ldots, X_n$ are independent if *for every subset* of the n variable, their joint pdf is the product of their marginal pdf's.

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \ldots f_n(x_n)$$

$$\text{and } f(x_1, x_7, x_{28}) = f_1(x_1) f_7(x_7) f_{28}(x_{28}) \text{ etc.}$$

3. Bayes' Thm

Chap. 2:

$$P(B_r \mid A) = \frac{P(A \mid B_r)P(B_r)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)}$$

Bayes' Theorem for Disc RV's: For $p_X(x) > 0$

$$p_{Y|X}(y \mid x) = \frac{p(x,y)}{p_X(x)} = \frac{p(x,y)}{\sum_y p(x,y)}$$

Note: pmf's are prob's, so this is Bayes's Thm in disc RV notation

3. Bayes' Thm

$$p_{Y|X}(y \mid x) = \frac{p(x,y)}{p_X(x)} = \frac{p(x,y)}{\sum_y p(x,y)}$$

Bayes' Theorem for Cont RV's: For $f_X(x) > 0$

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dy}$$

Note: pdf 's are not prob's but the formula works

Ex) X, Y have PDF $f(x,y) = c(x^2 - y^2)e^{-x}$

if $0 < x < \infty, -x < y < x$ **and 0 elsewhere.**

Find the conditional PDF of Y given X=x:

$$f_{Y|X}(y \mid x) = f(x,y) / f_X(x)$$

$$f_X(x) = \int_{-x}^{x} c(x^2 - y^2)e^{-x}dy = ce^{-x}[x^2 y - y^3/3|_{-x}^{x}$$

$$= (4c/3)x^3 e^{-x}, 0 < x < \infty$$

means c=1/8 (we'll see why later))

hence

$$f_{Y|X}(y\,|\,x) = (3/4)\frac{(x^2 - y^2)e^{-x}}{x^3 e^{-x}}, -x < y < x$$

and 0 elsewhere

- **Partial check, integrate this and verify we get 1**

$$\int_{-x}^{x} f_{Y|X}(y\,|\,x)dy = (3/4)\int_{-x}^{x} \frac{(x^2 - y^2)}{x^3}dy$$

$$= (3/x^3 4)[x^2 y - y^3/3|_{-x}^{x} = 1$$

- **Don’t need c** $f(y\,|\,x) = f(x,y) / \int_{-\infty}^{\infty} f(x,y)dy$

Conditional Independence & Learning (beyond the text)

- **Ex) Very large population of people. Let Y be the unknown proportion having a disease (D). Sample 2 people "at random", without replacement & check for D.**
- **Define X_1 = 1 if person 1 has D, X_1 =0 if not. Define X_2 for person 2 the same way.**
- **Note: the X's are discrete, Y is continuous**

Model assumptions

1. **Given Y= y, X_1 and X_2 are conditionally independent, with PD's**

$$p_1(x_1 \mid y) = y^{x_1}(1-y)^{1-x_1}, x_1 = 0,1$$

$$p_2(x_2 \mid y) = y^{x_2}(1-y)^{1-x_2}, x_2 = 0,1$$

Hence, $p_{12}(x_1, x_2 \mid y) = p_1(x_1 \mid y) p_2(x_2 \mid y)$

$$= y^{x_1+x_2}(1-y)^{2-(x_1+x_2)}, x_1 \,\&\, x_2 = 0,1$$

2. **Suppose we know little about Y: Assume $f_Y(y)=1$, $0<y<1$, and 0 elsewhere.**

Learn about Y after observing X_1 & X_2?

Answer

$$f_{Y|X_1X_2}(y \mid x_1, x_2) = p_{12}(x_1, x_2 \mid y)\, f_Y(y) / p_{12}(x_1, x_2)$$

where

$$p_{12}(x_1, x_2) = \int_0^1 y^{x_1+x_2}(1-y)^{2-(x_1+x_2)}\,dy,$$

$$x_1 \,\&\, x_2 = 0,1$$

Note: X_1 & X_2 are unconditionally dependent.

Learn about Y after observing X_1 & X_2?

Answer

$$f_{Y|X_1X_2}(y\,|\,x_1,x_2) = p_{12}(x_1,x_2\,|\,y)\,f_Y(y)\,/\,p_{12}(x_1,x_2)$$

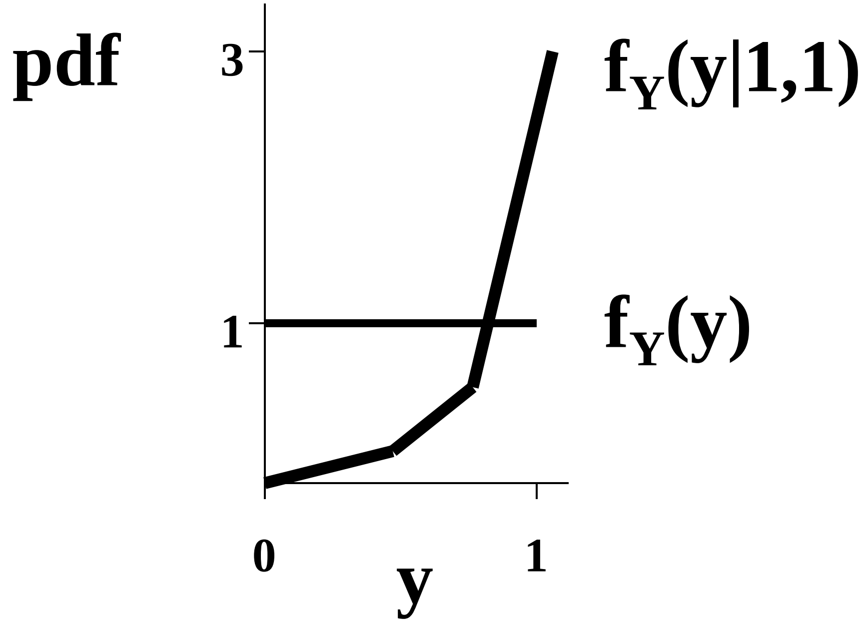
Ex) Observe $X_1 = X_2 = 1$. Then

$$p_{12}(1{,}1) = \int_0^1 y^2 dy = 1/3$$

and $$f(y\,|\,1{,}1) = \frac{y^2}{\int_0^1 y^2 dy} = 3y^2, 0 < y < 1$$

Summary

- **Before data, we had no idea about Y, our "*prior* pdf" was $f_Y(y)=1$, $0<y<1$. After seeing 2 out of 2 people sampled have D, we update to the "*posterior* pdf" $f_Y(y|1,1) = 3y^2$, $0<y<1$.**



Sect. 5.2 Expected Values

(p. 197) For discrete RV's *X*, *Y* with joint pmf *p*, the <u>expected value of *h*(*X*,*Y*)</u> is

$$E[h(X,Y)] = \Sigma_x \Sigma_y h(x,y)\, p(x,y)$$

***if finite*.**

For continuous RV's *X*, *Y* with joint pdf *f*, the <u>expected value of *h*(*X*,*Y*)</u> is

$$E[h(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)\, f(x,y)\, dx\, dy$$

***if finite*.**

- **Ex) X and Y have pdf $f(x,y) = x + y$, $0 < x < 1$, $0 < y < 1$, and 0 else. Find $E(XY^2)$.**

$$E(XY^2) = \int_0^1 \int_0^1 xy^2 (x + y) dxdy = 17/72$$

(Check my integration)

- **Extensions, by example:**

$$E[h(X,Y,Z)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y,z) f(x,y,z) dxdydz$$

Important Result:

If $X_1, X_2, \cdots, X_n$ are independent RV's, then

$$E[h_1(X_1)h_2(X_2)\cdots h_n(X_n)] = E[h_1(X_1)]E[h_2(X_2)]\cdots E[h_n(X_n)]$$

- **“Under independence, the expectation of product = product of expectations”**
- **Proof: (for n=2 in the continuous case.)**

$$E[h_1(X_1)h_2(X_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x_1)h_2(x_2)\, f(x_1, x_2)\,dx_1 dx_2$$

by indep.

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x_1)h_2(x_2)\, \overbrace{f_1(x_1)\, f_2(x_2)}\,dx_1 dx_2$$

$$= \int_{-\infty}^{\infty} h_1(x_1)\, f_1(x_1)\,dx_1 \int_{-\infty}^{\infty} h_2(x_2)\, f_2(x_2)\,dx_2$$

$$= E[h_1(X_1)]E[h_2(X_2)]$$

Ex: X and Y have pdf

$$f(x,y) = (1/8)xe^{-0.50(x+y)}, x > 0, y > 0$$

and 0 elsewhere. Find E(Y/X).

- **Note that f(x,y) "factors" (ranges on x,y are OK). Hence, X and Y are independent and**

$$f(x,y) = [(.25)xe^{-0.50x}][(.50)e^{-0.50y}], x > 0, y > 0$$

$$= f_X(x) f_Y(y)$$

Remark: If $f(x,y) = cxe^{-0.50(x+y)}, 0 < x < y$

then X and Y are dependent

E(Y/X) = E(Y) E(1/X)

$$= \int_0^\infty y(.50)e^{-0.50y}dy \int_0^\infty x^{-1}(.25)xe^{-0.50x}dx = (2)(.50) = 1$$

since,

- **Y ~ Exp(λ = .5) and mean of exponential is 1/λ**
- **X has a Gamma pdf, but**

$$\int_0^\infty x^{-1}(.25)xe^{-0.50x}dx = \int_0^\infty (.25)(.5)/(.5)e^{-0.50x}dx$$

$$= .25/.50 = .50$$

Important Concepts in Prob & *Stat* (p. 198)

1. **<u>Covariance of X and Y</u> = Cov(X,Y) =** σ_{XY}

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

Fact: $Cov(X,Y) = E(XY) - \mu_X \mu_Y$

Point: Cov(X,Y) measures dependence between X & Y

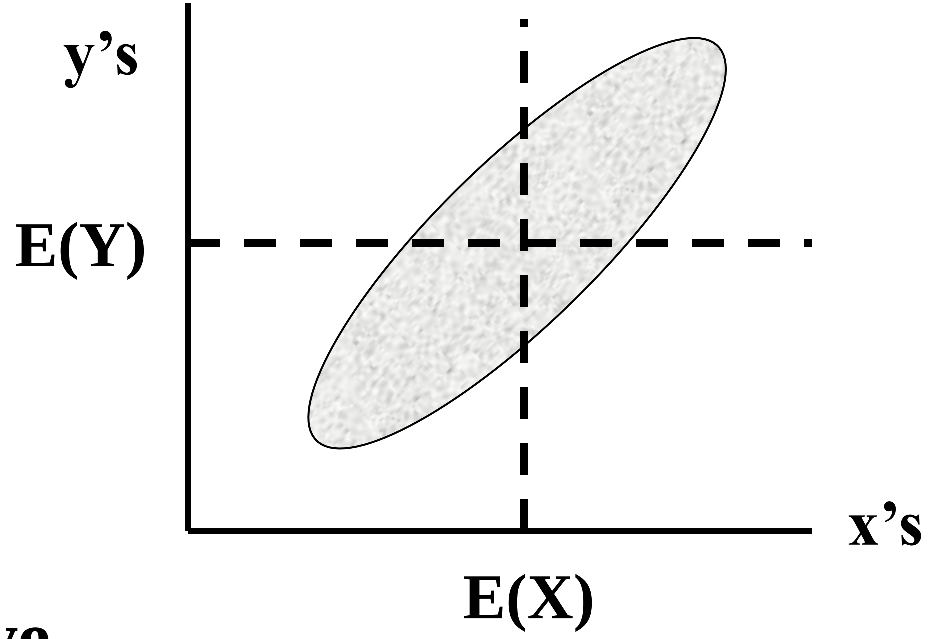
- **If X and Y are *independent*, then Cov(X,Y)= 0.**
  Why: E(XY) = E(X)E(Y) = $\mu_X \mu_Y$

  (indep. implies E(product)=product(E's))
- **But Cov(X,Y) = 0 does *not* imply independence.**

Intuition

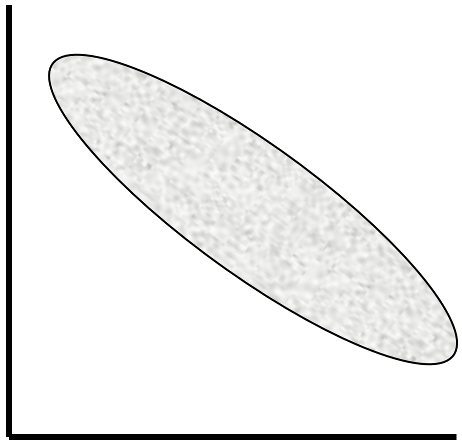
We observe and graph many pairs of (X,Y).

Suppose we get

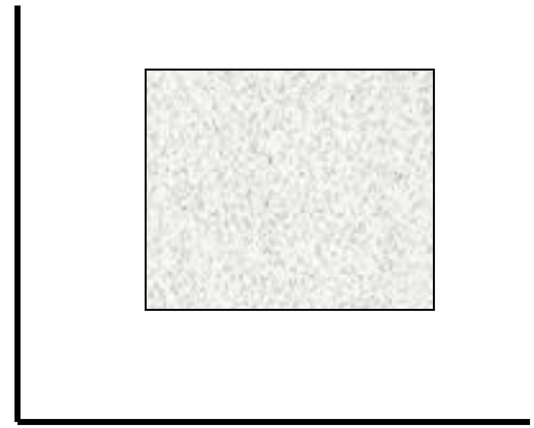


Then (X-E(X)) and (Y-E(Y)) tend to have the same sign, so the average of their product (i.e., covariance) is positive.

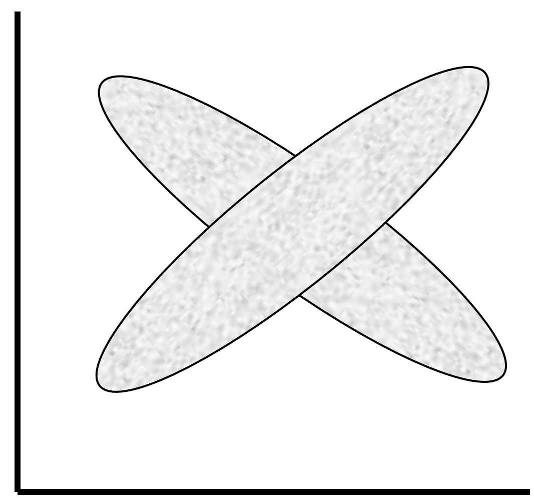
- **cov(X,Y) < 0**



cov(X,Y)=0 (& indep)



- **cov(X,Y)=0**
  (*Not* indep)



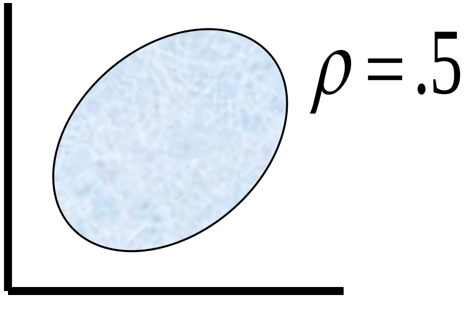
2. <u>**Correlation between X and Y**</u> **= Corr(X,Y) =** ρ_{XY}

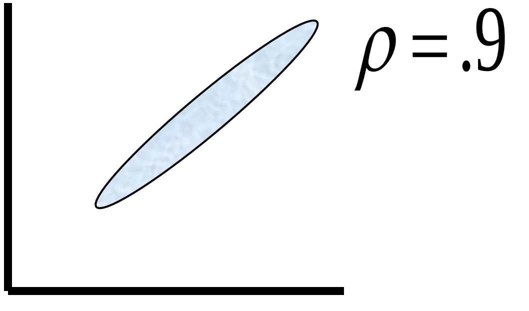
$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Theorem: $-1 \le \rho_{XY} \le 1$

Point: ρ_{XY} **measures dependence between X & Y in a fashion that does not depend on units of measurements**

- **Sign of** ρ_{XY} **indicates direction of relationship**
- **Magnitude of** $|\rho_{XY}|$ **indicates the strength of the** ***linear*** **relationship between X and Y**





Results for Linear Combinations of RV's

1. **Recall:** $V(aX+b)=a^2\sigma_X^2 \quad S.D.(aX+b)=|a|\sigma_X$
2. **Extensions:**

$$Cov(aX+b,cY+d)=acCov(X,Y)$$

Hence, $Corr(aX+b,cY+d)=Corr(X,Y)$

$$V(X+Y)=V(X)+V(Y)+2Cov(X,Y)$$

So, if X & Y are indep., $V(X+Y)=V(X)+V(Y)$

Thm: If $X_1, X_2, \cdots, X_n$ are indep.,

$$\mathrm{V}(X_1+X_2+\cdots+X_n)=\mathrm{V}(X_1)+V(X_2)+\cdots+V(X_n)$$

"Var(sum of indep. RV) = sum (their variances)"

Ex) Pistons in Cylinders

Let X_1 = diameter of cylinder, X_2 = diameter of piston. "<u>Clearance</u>" $Y = 0.50\,(X_1 - X_2)$. Assume X_1 and X_2 are independent and

$\mu_1 = 80.95$ cm, $\sigma_1 = .03$ cm;
$\mu_2 = 80.85$ cm, $\sigma_2 = .02$ cm

Find mean and SD of Y:

[$Y = 0.50\ (X_1 - X_2)$. $\mu_1 = 80.95$ cm, $\sigma_1 = .03$ cm;
$\mu_2 = 80.85$ cm, $\sigma_2 = .02$ cm]

1. $\mu_Y = E[.50\ (X_1 - X_2)]$

$= .50\ [\mu_1 - \mu_2] = .05$ cm

2. S.D.: Find V(Y), then square root
(there's no general shortcut)

$V(Y) = V[.50\ (X_1 - X_2)]$

$= (.5)^2\ V(X_1 - X_2)$

$= .25\ [(.03)^2 + (.02)^2] = 3.24 \times 10^{-4}$

so $\sigma_Y =$ .018 cm (*not* .5 (.03 + .02) = .025)

Ex), Cont'd

If Y is too small, the piston can't move freely; if Y is too big, the piston isn't tight enough to combust efficiently. Designers say a piston-cylinder pair will work if .01 < Y < .09. Assuming Y has a normal dist., find

P(.01 < Y < .09).

$$P(.01 < Y < .09) = P\left(\frac{.01 - .05}{.018} < Z < \frac{.09 - .05}{.018} \right)$$

$$= P(-2.22 < Z < 2.22) = .9736$$

FYI (not HW or Exams)

If the normality assumption can't be claimed, you can get a bound:

- $P(.01 < Y < .09) = P(.01 - .05 < Y - .05 < .09 - .05)$

  $= P(|Y - .05| < .04)$

  $= P(|Y - \mu_Y | < .04)$

- **Chebyshev's Inequality: For any constant k > 0,**

  $$P(|X - \mu| < k\,\sigma) \geq 1 - k^{-2}$$

- Set $.04 = k\,\sigma_Y$, so $k^{-2} = (.018 / .04)^2$

- **Conclusion:**

  $$P(.01 < Y < .09) = P(|Y - .05| < .04) \geq .797$$

no matter what the dist of Y is !!!!

FYI (not for HW or Exam): Conditional Expectation

Recall: X,Y have pdf f(x,y). Then

$f_{X|Y}(x|y) = f(x,y)/f_Y(y)$ and

$$P(X \in A \mid Y = y) = \int_A f_{X|Y}(x \mid y)dx$$

(p. 156) C<u>onditional expectation</u> of h(X) given Y=y

Discrete $$E[h(X) \mid y] = \Sigma_x h(x) p_{X|Y}(x \mid y)$$

Contin. $$E[h(X) \mid y] = \int_{-\infty}^{\infty} h(x) f_{X|Y}(x \mid y)dx$$

if they exist.

Special cases (Contin. case; discrete are similar)

1. **Conditional mean** of X given Y=y is

$$\mu_{X|y} = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

2. **Conditional variance** of X given Y=y is

$$\sigma^2_{X|y} = E[(X-\mu_{X|y})^2 | y] = E(X^2 | y) - \mu^2_{X|y}$$

Try not to let all this notation fox you. All definitions are the same, the conditioning just tells us what PD or pdf to use.

X and Y have pdf

$$f(x,y) = \lambda^2 e^{-\lambda y}, 0 < x < y$$

Find $\sigma^2_{X|y}$

1. Need $f_{X|Y}(x|y) = f(x,y)/f_Y(y)$

$$f_Y(y) = \int_0^y \lambda^2 e^{-\lambda y} dx = \lambda^2 e^{-\lambda y} y, y > 0$$

so

$$f_{X|Y}(x|y) = \frac{\lambda^2 e^{-\lambda y}}{y\lambda^2 e^{-\lambda y}} = 1/y, 0 < x < y$$

2. Find $\mu_{X|y} = \int_0^y x(1/y)dx = 0.50y$

$$E(X^2 \mid y) = \int_0^y x^2(1/y)dx = y^2/3$$

3. So $\sigma^2_{X|y} = y^2/3 - (y/2)^2 = y^2/12$

Sec 5.3 - 5.4

Last material for this course

Lead-in to

- **<u>statistical inference</u>: drawing conclusions about population based on sample data**
- **we state our inferences and judge their value: based on probability**

Key Definitions and Notions

A. Target of Statistical Inference:

1. **<u>Population</u>: Collection of units or objects of interest.**
2. **<u>Pop Random Variable (RV)</u>: Numerical value X associated with each unit.**
3. **<u>Pop Dist</u>.: Dist. of X over the pop.**
4. **<u>Parameters</u>: Numerical summaries of pop. (mean, variance, proportion, …)**

B. Inputs to Statistical Inference

1. **Sample**: Subset of the population
2. **Sample Data**: X_1 , X_2 , …, X_n for the *n* units in sample
3. **Statistic**: Function of the data

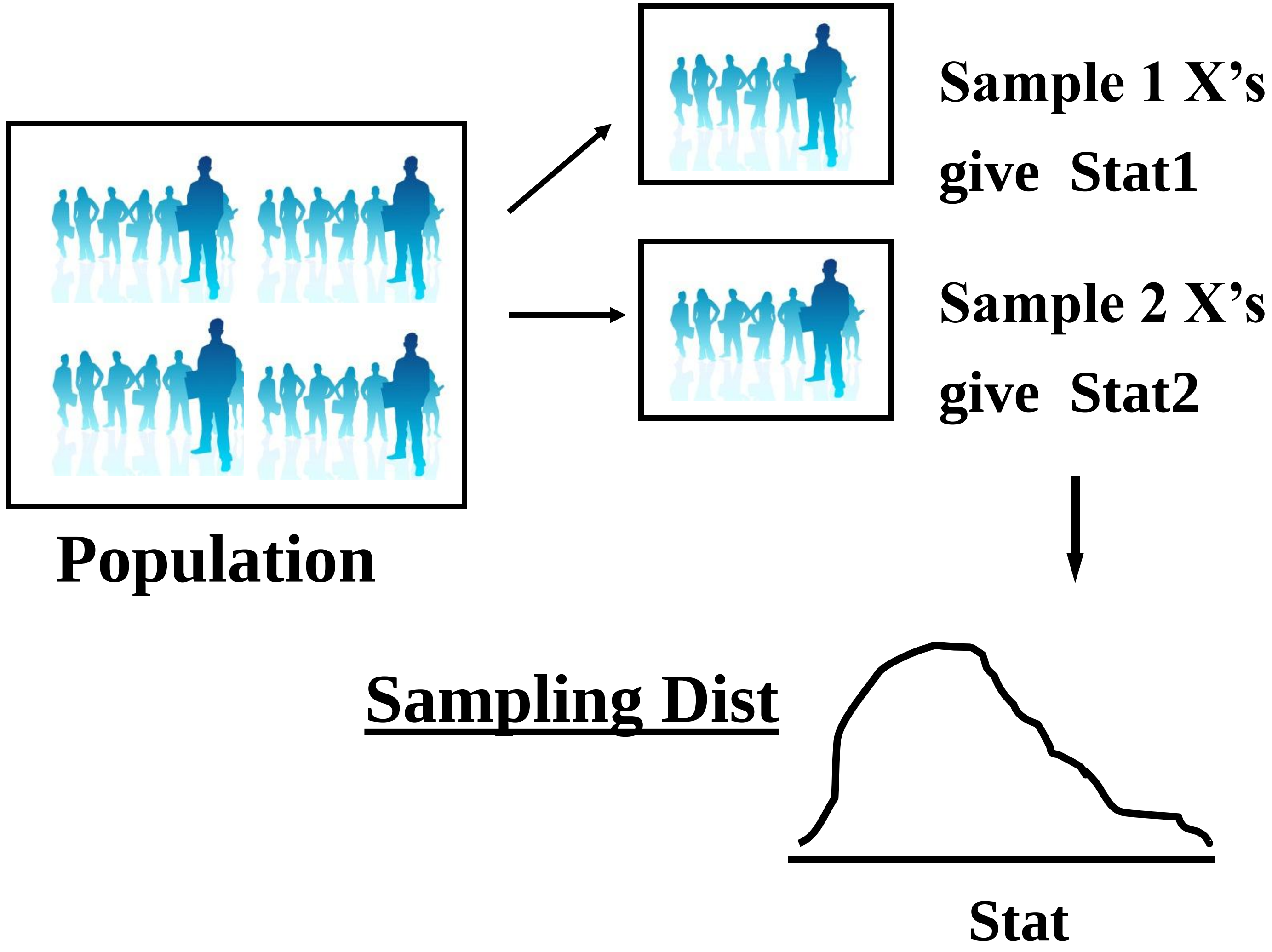
Main ex) "sample mean" $\bar{X} = \sum_{i=1}^{n} X_i \Big/ n$

4. **Sampling variability**: Different samples give different values of a statistic. That is,

Statistics are RV's

5. **Sampling distribution**: Probability distribution of a statistic.

Idea


Sample 1 X’s
give Stat1
Sample 2 X’s
give Stat2
Population
Sampling Dist
Stat

Key Point: Sampling Design

- **Form of sampling dist. depends on how we pick samples**
- **In most cases, we want samples to be *representative* of pop.**

  (i.e., not biased or special in some way).

If X_1 , X_2 , …, X_n are *independent and identically distributed* (i.i.d.), each having the pop. dist., they form a <u>random sample</u> from the population

- **Finding sampling dist:**

  (1) Simulation and (2) Prob. theory results

Remark

Alternative notion of representative samples:

Simple random sample (SRS):

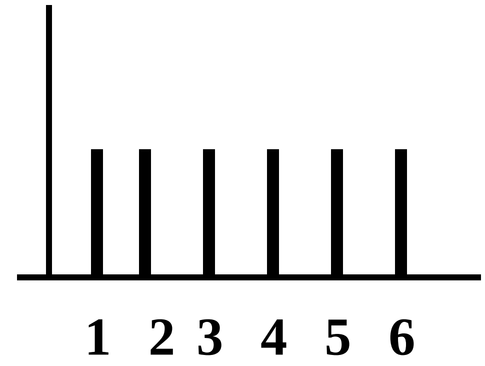
Sample of n units chosen in a way that all samples of n units have equal chance of being chosen

- **Sampling without replacement: observations are dependent.**
- **When sampling from huge populations, SRS are approximately Random Samples**

Sampling dist. via grunt work

X = the result of the roll of a fair die

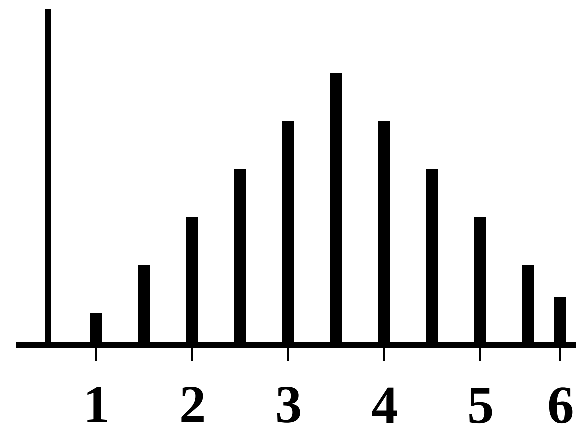
x :	1	2	3	4	5	6
p(x):	1/6	1/6	1/6	1/6	1/6	1/6



X_1, X_2 be results of 2 indep. rolls. Dist of $\bar{X}$:

$\bar{x}$:	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
$p(\bar{x})$:	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

	1	2	3	4	5	6
1	1	1.5	2	2.5	3	3.5
2	1.5	2	2.5	3	3.5	4
3	2	2.5	3	3.5	4	4.5
4	2.5	3	3.5	4	4.5	5
5	3	3.5	4	4.5	5	5.5
6	3.5	4	4.5	5	5.5	6



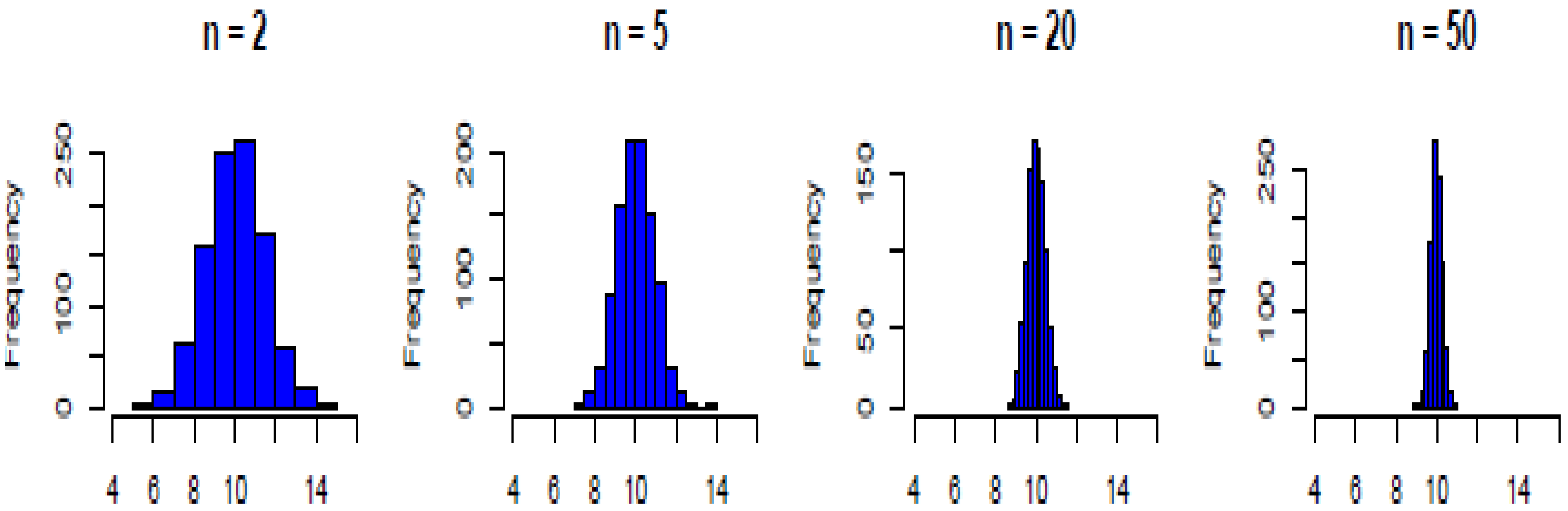
Simulation: Fundamental in modern applications

Math/Stat has developed numerical methods to "generate" realizations of RV's from specified dist.

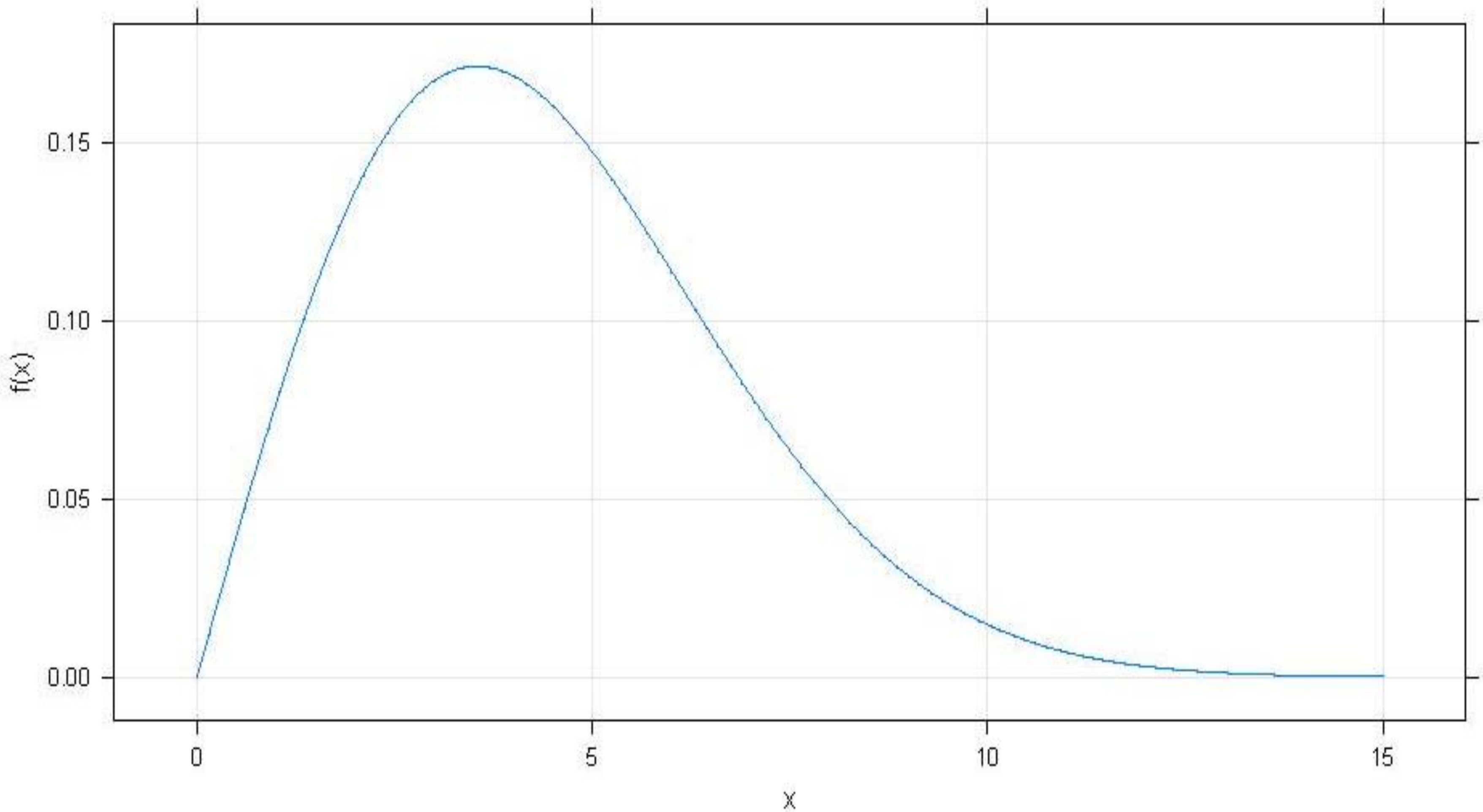
- **Select a dist. and a statistic**
- **Simulate many random samples**
- **Compute statistic for each**
- **Examine the dist of the simulated statistics: histograms, estimates of their density, etc.**

Simple Ex)

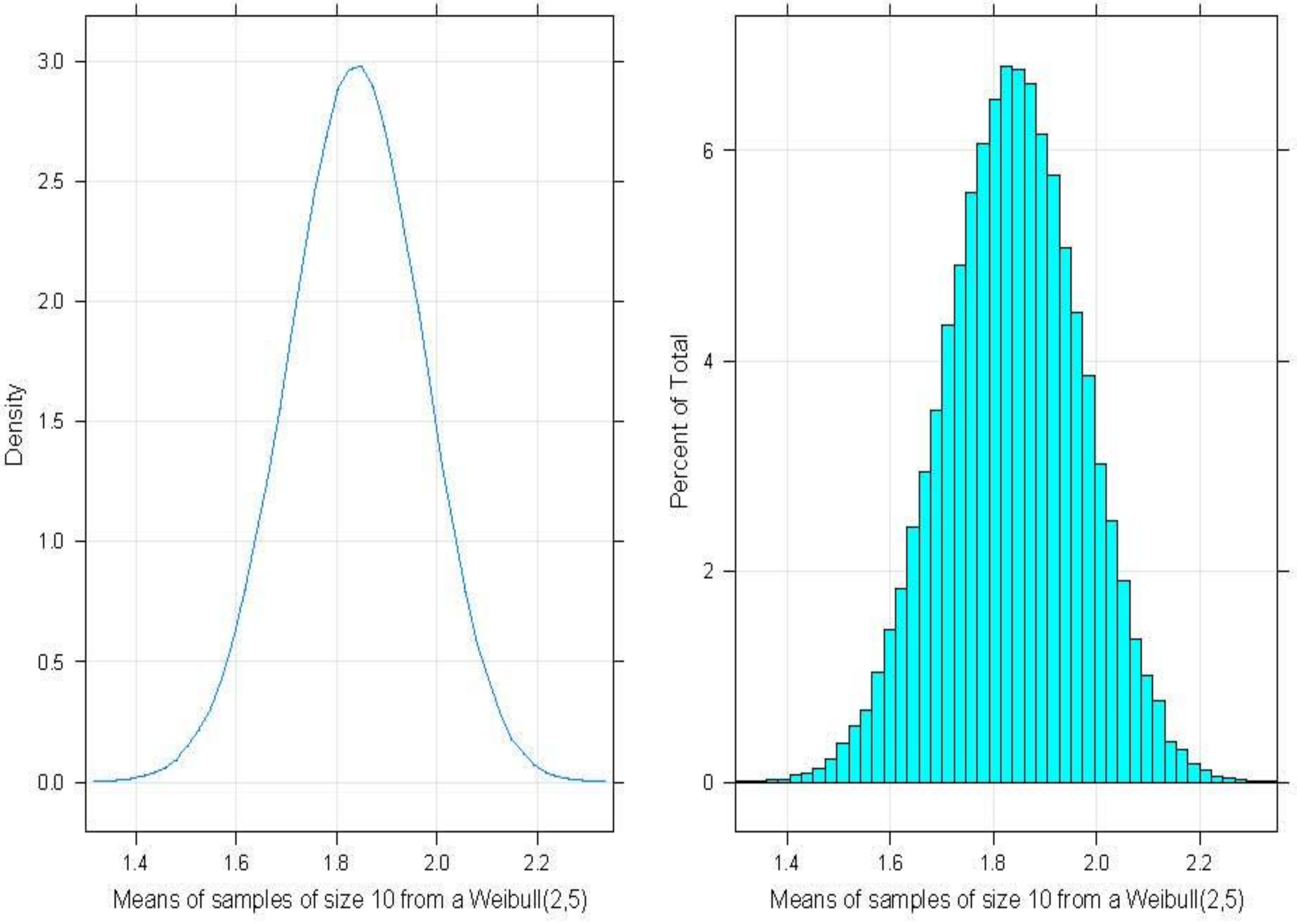
- **Pop. Dist: X ~ N(10, 4). Statistic** $\bar{X}$
- **Draw k = 1000 random samples (in practice, use much larger k); compute means and make histograms for four different sample sizes**



Consider the Weibull distribution with parameters $\alpha = 2$ (the shape parameter) and $\beta = 5$ (the scale parameter) shown below.

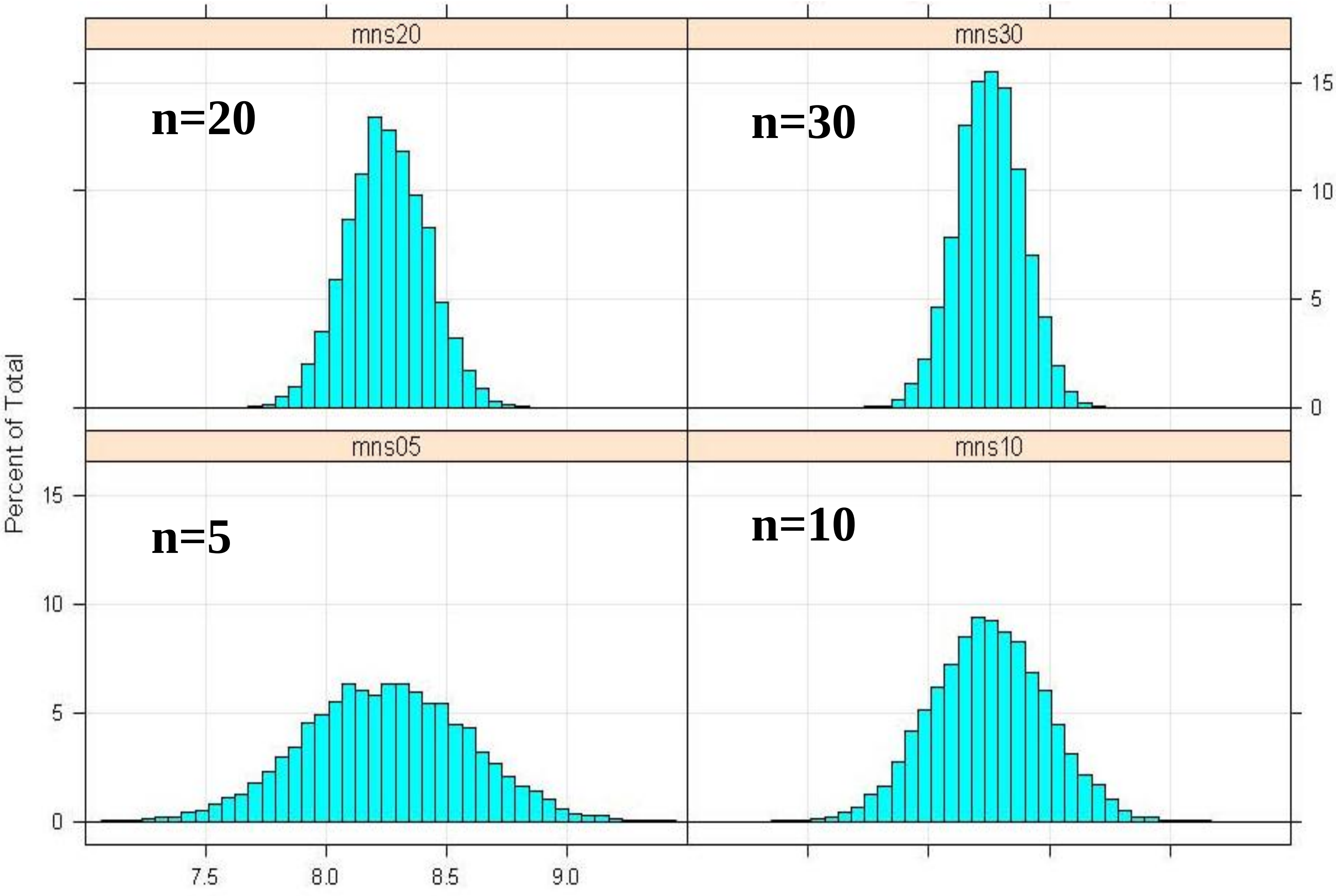


n=10, k = 50,000

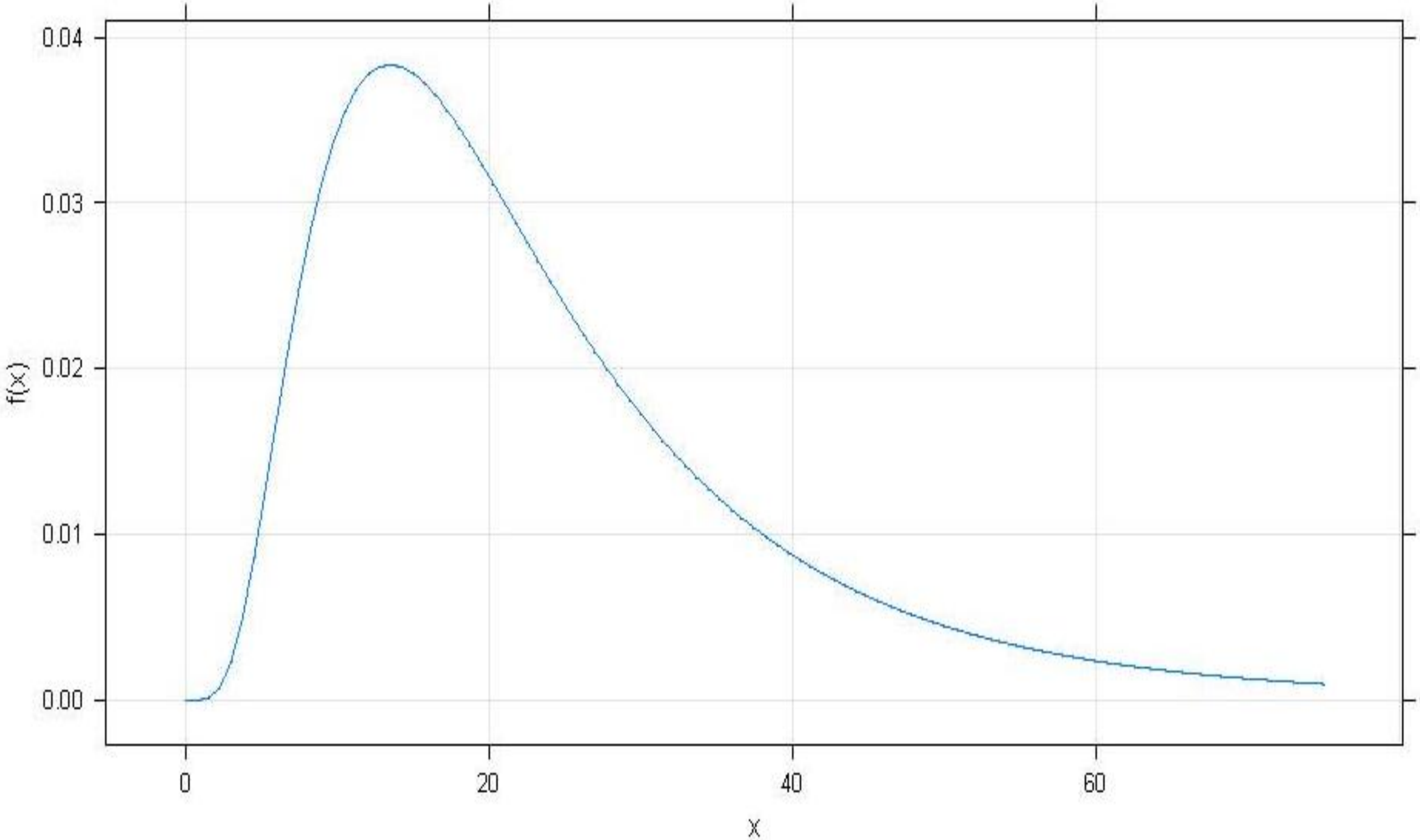


3.0
2.5
2.0
1.5
1.0
0.5
0.0
Density
1.4
1.6
1.8
2.0
2.2
Means of samples of size 10 from a Weibull(2,5)
6
4
2
0
Percent of Total
1.4
1.6
1.8
2.0
2.2
Means of samples of size 10 from a Weibull(2,5)

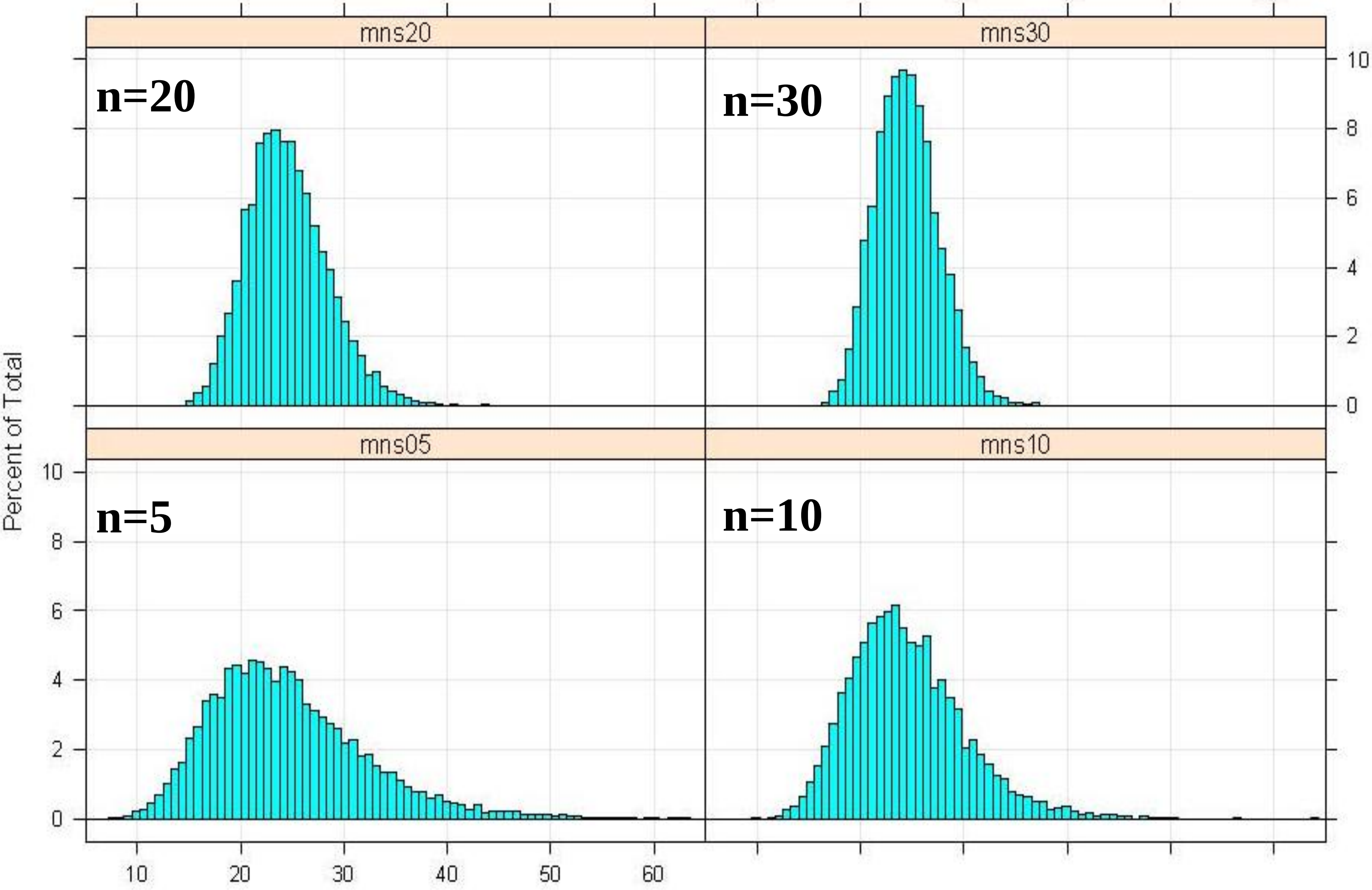
Various n


mns20
mns30
mns05
mns10
n=20
n=30
n=5
n=10
Percent of Total
0
5
10
15
7.5
8.0
8.5
9.0

In example 5.23 the means of samples of different sizes from a log-normal distribution with $E[\ln X] = 3$ and $\text{Var}(\ln X) = 0.4$ are simulated.

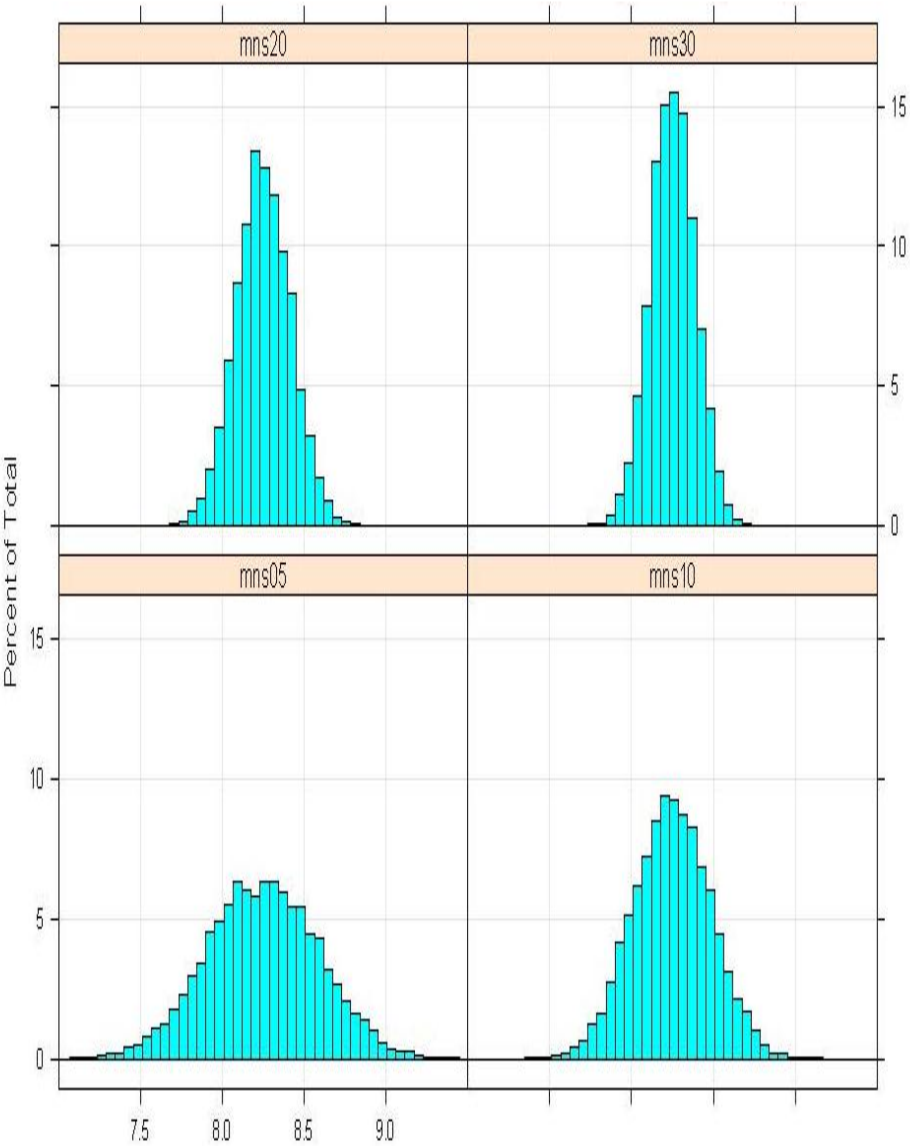


Various n

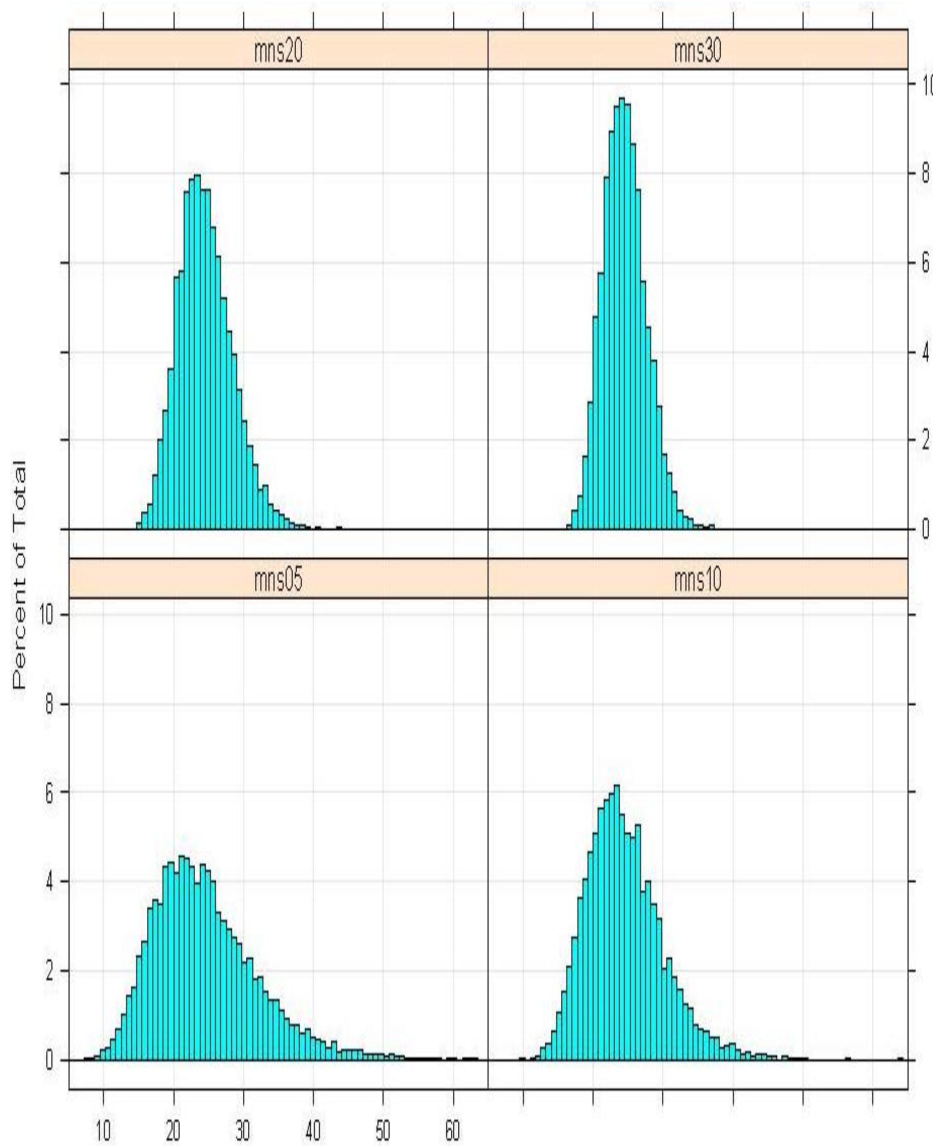


mns20
mns30
mns05
mns10
n=20
n=30
n=5
n=10
Percent of Total
0
2
4
6
8
10
10
20
30
40
50
60

Weibull



Lognormal



Last Time

1. **<u>Population</u>: Collection of objects of interest.**
2. **<u>Population RV</u>: X (value for each object in pop.)**
3. **<u>Population Dist</u>.: Dist. of RV X over the pop.**
4. **<u>Parameters</u>: Numerical summaries of pop. (ex. μ, σ)**
5. **<u>Sample</u>: Subset of the population**
6. **<u>Data</u>: X_1 , X_2 , …, X_n for the objects in sample**
7. **<u>Statistic</u>: Function of the data**

Key*: <u>Sampling variability</u>: different samples give different values of a statistic. *Statistics are RV's

8. **<u>Sampling distribution</u>: Distribution of a statistic.**

***Key*: Distribution of statistic depends on how sample is taken**

Sampling Design

In most cases, we want samples to be *representative* of pop. (i.e., not biased or special in some way).

In this course (and most applications):

If X_1 , X_2 , …, X_n are *independent and identically distributed* (i.i.d.), each having the pop. dist., they form a <u>random sample</u> from the population

- **Finding sampling dist:**
  - **(1) Simulation (last time)**
  - **(2) Prob. theory results (today)**

Remark

Alternative notion of representative samples:

Simple random sample (SRS):

Sample of n units chosen in a way that all samples of n units have equal chance of being chosen

- **Sampling without replacement: observations are dependent.**
- **When sampling from huge populations, SRS are approximately Random Samples**

Main Statistics (Sect. 5.4; p. 229-230)

1. **Sample Mean**: $$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$
   or "X-bar"

2. **Sample Variance**: $$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$
   or "S-squared"

3. **Sample proportion: example later**

Sect. 5.4: Dist. of X-bar

***Proposition* (p. 213):**

If X_1 , X_2 , …, X_n is a (iid) random sample from a population distribution with mean μ and variance σ^2, then

1. $\mu_{\overline{X}} \overset{def}{=} E(\overline{X}) = \mu$

2. $\sigma^2_{\overline{X}} \overset{def}{=} V(\overline{X}) = \dfrac{\sigma^2}{n}$ and $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

Proof:

1. $$E(\bar{X}) = E\left(\sum_{i=1}^{n} X_i / n\right)$$

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

$$= \frac{n\mu}{n} = \mu$$

Constants come outside exp' and E(sum) = sum(E's)

2. $$\operatorname{var}(\bar{X}) = \operatorname{var}\left(\sum_{i=1}^{n} X_i / n\right)$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} \operatorname{var}(X_i)\right)$$

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

Constants come out of Var squared and

for *indep* RV, V(sum) =sum(V's)

Remarks:

1. In statistics, we typically use $\bar{X}$ to estimate μ and S^2 to estimate σ^2.

When $E($Estimator$)$ = target , we say the estimator is **unbiased**

(note: S is not unbiased for σ)

2. Independence of the X_i's is only needed for the variance result.

3. Results stated for sum's: Let $T_0 = \sum_{i=1}^{n} X_i$

Under the assumptions of the Proposition,

$$E(T_0) = n\mu, \quad V(T_0) = n\sigma^2 \text{ and } \sigma_{T_0} = \sqrt{n}\sigma$$

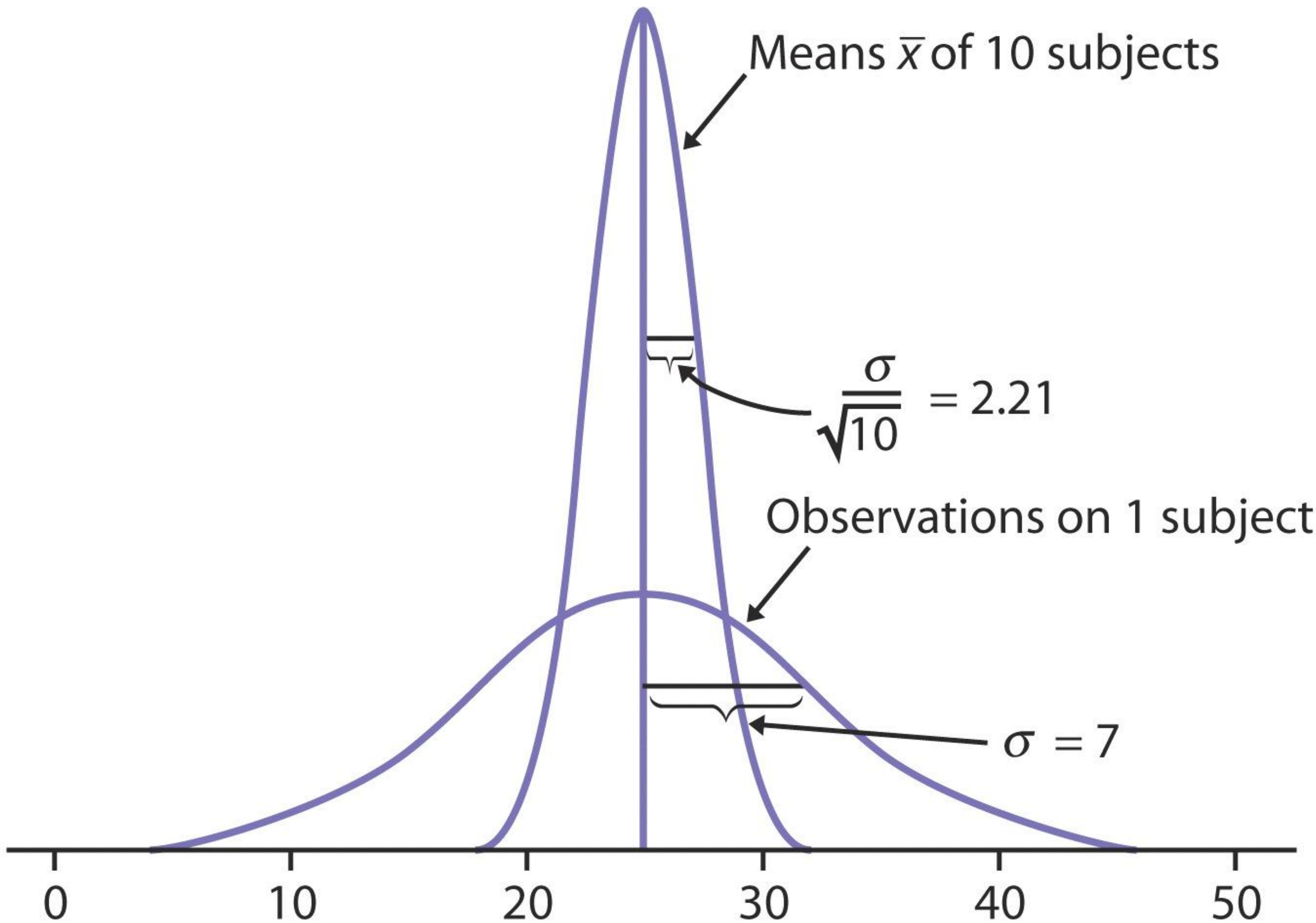
More results

Proposition **(p. 214): If X_1 , X_2 , …, X_n is an iid random sample from a population having a *normal distribution* with mean μ and variance σ^2, then $\bar{X}$ has a *normal distribution* with mean μ and variance σ^2/n**

That is,

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Proof: Beyond our scope (not really hard, just uses facts text doesn't cover)


Means $\bar{x}$ of 10 subjects
$\frac{\sigma}{\sqrt{10}} = 2.21$
Observations on 1 subject
$\sigma = 7$
0
10
20
30
40
50

Large sample (large n) properties of X-bar

Assume X_1, X_2, …, X_n are a random sample (iid) from a population with mean μ and variance σ^2 .
(*normality is not assumed*)

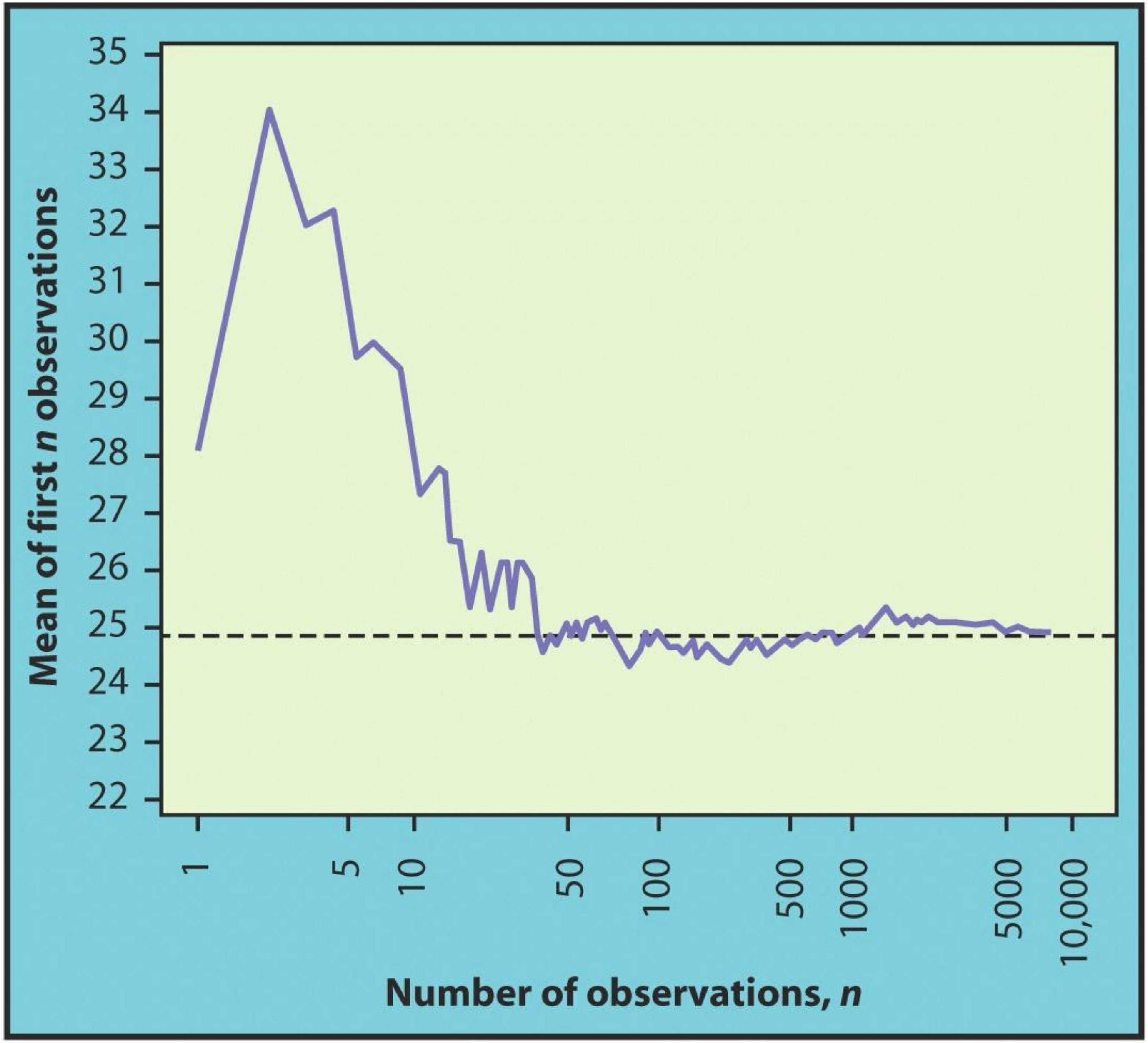
1. **<u>Law of Large Numbers</u>**

Recall that $\mu_{\overline{X}} = \mu$ **and note that**

$$n \to \infty \Rightarrow \sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \to 0$$

We can prove that with probability = 1,

$$n \to \infty \Rightarrow \overline{X} \to \mu$$



35
34
33
32
31
30
29
28
27
26
25
24
23
22
Mean of first n observations
1
5
10
50
100
500
1000
5000
10,000
Number of observations, n

2. <u>Central Limit Theorem</u> (CLT) (p. 215)

Under the above assumptions (iid random sample)

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z \sim N(0,1)$$

"converges in dist. to"
or "has limiting dist."

Point: For n large, $\bar{X} \approx N(\mu, \sigma^2/n)$

"approx dist as"

So for n large, we can approx prob's for *X*-bar even though we don't know the pop. dist.

2. <u>Central Limit Theorem</u> (CLT) For Sums

Under the above assumptions (iid random sample)

$$\frac{T_0 - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z \sim N(0,1)$$

Point: For n large,

$$T_0 \approx N(n\mu, n\sigma^2)$$

So for n large, we can approx prob's for T_0 even though we don't know the pop. dist.

Last Time:

Dist. of "X-bar": $\bar{X} = \sum_{i=1}^{n} X_i / n$

Three Main Results: ***Assumption common to all*****:**

> **X_1, X_2, …, X_n is a (iid) random sample from a pop. dist. with mean μ and variance σ^2.**

1. $\mu_{\bar{X}} = \mu$ and $\sigma^2_{\bar{X}} = \sigma^2 / n$

2. **If the pop. dist. is normal, then**

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

3. **If n is large**

$$\bar{X} \approx N(\mu, \sigma^2/n)$$

Restate for the sum: $T_0 = \sum_{i=1}^{n} X_i$

1. $\mu_{T_0} = n\mu$ and $\sigma_{T_0}^2 = n\sigma^2$

2. **If the pop. dist. is normal, then**

$$T_0 \sim N(n\mu, n\sigma^2)$$

(Theorem: sum of indep. normals is normal)

3. **If n is large**

$$T_0 \approx N(n\mu, n\sigma^2)$$

Ex) Estimate the average height of men in some population. Assume pop. σ is 2.5 in. We will collect an iid sample of 100 men. Find the prob. that their sample mean will be within .5 in of the pop. mean.

- **Let μ be the population mean. Find**

$$P(|\bar{X} - \mu| < .5)$$

- **Applying CLT,** $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n} = \frac{2.5^2}{100} = \frac{2.5^2}{10^2} \right)$

so

$$P(|\bar{X} - \mu| < .5) \approx P(|Z| < .5/.25) = P(|Z| < 2) = .95$$

Ex) Application for "sums": Binomial Dist.

1. **Recall (p. 88): A <u>Bernoulli RV</u> *X* takes on values 1 or 0. Set P(*X* = 1) = *p* . Easy to check that**

$$\mu_{\mathrm{X}} = p \text{ and } \sigma_X^2 = p(1-p)$$

2. **X ~ Bin(*n*, *p*) is the sum of *n* iid Bernoulli's.**

Applying result $\mu_{T_0} = n\mu$ and $\sigma_{T_0}^2 = n\sigma^2$ **gives**

$$\mu_X = np \text{ and } \sigma_X^2 = np(1-p)$$

3. **CLT: For *n* large,**

$$X \approx N(np, np(1-p))$$

Remark: CLT doesn't include "continuity correction"

Ex) Continued. In practice, we may not know *p*

1. **Traditional estimator: <u>sample proportion</u>, p-hat**

$$\hat{p} = X/n$$

2. **Key: since X ~ Bin(*n*, *p*) is the sum of *n* iid Bernoulli's, p-hat is a *sample mean***

i.e., let B_i , i = 1,…, n denote the Bernoulli's:

$$\hat{p} = X/n = \sum_{i=1}^{n} B_i \Big/ n$$

3. **Apply CLT: For *n* large,**

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n} \right)$$

Ex) Service times for checkout at a store are indep., average 1.5 min., and have variance 1.0 min^2. Find prob. 100 customers are served in less than 2 hours.

- **Let X_i = service time of ith customer.**
- **Service time for 100 customers:** $T = \sum_{i=1}^{n} X_i$
- **Applying the CLT,**

$$T \approx N(n\mu = 100(1.5), n\sigma^2 = 100(1))$$

- **So,**

$$P(T < 120) = P\left(\frac{T - 150}{10} < \frac{120 - 150}{10} \right)$$

$$\approx P(Z < -3) = .0013$$

Ex) Common Class of Applications

- **In many cases, we know the distribution of the sum of independent RV's, but if that dist. is complicated, we may still want to use CLT approximations.**
- **Example:**

1) Theorem. If $X_1 \sim Poi(\lambda_1)$, …, $X_k \sim Poi(\lambda_k)$ are indep. Poisson RV's, then

$$T = \sum_{i=1}^{k} X_i \sim Poi(\lambda_T = \sum_{i=1}^{k} \lambda_i)$$

(i.e., "sum of indep. Poissons is Poisson")

Proof: not hard, but beyond the text

2) **Implication: Suppose *Y* ~ Poi(λ) where λ is very large. Recall** $\mu_Y = \lambda$ and $\sigma_Y^2 = \lambda$

3) **Slick Trick: pretend *Y* = sum of *n* iid Poisson's, each with parameter λ* where λ = *n*λ* and *n* is large. That is, X_i ~ Poi(λ*) for i = 1, …, n.**

By the Theorem:

$$Y = \sum_{i=1}^{n} X_i \sim Poi(\lambda = \sum_{i=1}^{n} \lambda^* = n\lambda^*)$$

4) **Apply CLT:**

$$\frac{Y - \lambda}{\sqrt{\lambda}} \approx N(0,1)$$

Ex) Number of flaws in a unit of material has a Poisson dist. with mean 2. We receive a shipment of 50 units.

a) **Find prob that the total number of flaws in the 50 units is less than 110.**
b) **Find prob that at least 20 of the 50 have more than 2 flaws.**
c) **Find prob that at least 2 of the 50 have more than 6 flaws.**

In the following solutions, we assume the number of flaws in the 50 units are *independent* RV's

Find prob that the total number of flaws in the 50 units is less than 110.

- **Since sum of indep Poisson's is Poisson:**

T = total number of flaws is Poi(λ = 2(50) = 100)

Since λ = 100 is large, use normal approx.:

$$P(T < 110) = P\left(\frac{T - 100}{10} < \frac{110 - 100}{10}\right)$$

$$\approx P(Z < 1) = .8413$$

Note, with cont. correction,

$$P(T < 110) \approx P\left(Z < \frac{109.5 - 100}{10}\right) = .8289$$

("Exact" Poisson calculation: .8294)

b) Find prob that at least 20 of the 50 have more than 2 flaws.

- **Let X = number units with more than 2 flaws. Since the units are indep., X is Bin(*n*=50, *p*) where *p* = P(Y >2) where Y is Poi(λ=2).**
- **Using Poisson pmf, check that *p* = .3233**
- **Apply Normal approx. to binomial:**
  $\mu_X = np = 50(.3233) = 16.165 \quad \sigma_X^2 = np(1-p) = 10.94$

so $P(X \geq 20) \approx P(Z \geq (20 - 16.165)/\sqrt{10.94}) = .123$

or, with cont. correction

$$P(X \geq 20) \approx P(Z \geq (19.5 - 16.165)/\sqrt{10.94}) = .1562$$

(Exact Binomial: .1566)

c) Find prob that at least 2 of the 50 have more than 6 flaws.

- **Now, $X \sim Bin(50, p)$ where $p = P(Y>6)$ and $Y \sim Poi(\lambda = 2)$. Poisson pmf gives p = .0045**
- **Hence, np = 50(.0045) = .225 which is way too small for normal approx.**
- **Use Poisson approx to Bin:**

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0,1)$$

Using $Poi(\lambda = .225)$, we get

$$\exp(-.225)\,(1.225) = .023$$

(Exact Binomial: .0215)

CLT: Assume X_1 , X_2 , …, X_n is a (iid) random sample from a distribution with mean μ and variance σ^2. If *n* is large, $\bar{X} \approx N(\mu, \sigma^2/n)$ **and the sum** $T_0 \approx N(n\mu, n\sigma^2)$

Note: In practice, we may need to estimate σ^2. Typical procedure: input the sample variance. Theorem:

$$n \to \infty \Rightarrow S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 \Big/ (n-1) \to \sigma^2 \text{ with prob } = 1$$

Remark: Rule of Thumb (p. 217) "If *n* $\geq$ 30, CLT can be used" is nonsense. This is only OK if the pop dist is reasonably symmetric.

(Also p. 217, text says use CLT approx. to bin. if *np* > 10, so if *p* = .10 and *n* = 50, *np* = 5 so don't use CLT

(though *n* > 30))

Two important settings/applications

***A. Sample Size Determination*.**

Estimate the unknown mean μ of some distribution.

Basic procedure:

i. **Choose a random sample (iid) of *n* observations.**

ii. **Use their sample mean $\bar{X}$ to estimate μ**

Idea: we know accuracy of estimate increases as *n* does, but so does cost of data collection.

How large should *n* be to obtain a desired accuracy?

***Quantification*: For specified choices of *M* and α, choose *n* large enough that**

$$P(|\bar{X} - \mu| \leq M) \geq 1 - \alpha$$

- ***M* is the margin of error**
- **α is the error rate; α small (.05 is a common choice)**

Apply CLT: for *n* large,

$$P(|\bar{X} - \mu| \le M) \approx P(|Z| \le \frac{M}{\sigma/\sqrt{n}}) \ge 1-\alpha$$

Conclusion:

α = .05: $\frac{M}{\sigma/\sqrt{n}} = 1.96$ so we need $n \ge (1.96\sigma/M)^2$

α = .01: $\frac{M}{\sigma/\sqrt{n}} = 2.576$ so we need $n \ge (2.576\sigma/M)^2$

Notes

1. **Of course, round up to an integer.**
2. **Procedure requires a guess at σ**
3. **Analysis is valid for all *n* if the population is normal.**
4. **Otherwise, if the answer turns out to be small, CLT does not apply, so analysis failed**

(Remark: you could use Chebyshev's Inequality, more conservative, but works for all distributions and all *n*)

Ex) Assess accuracy of lab scale. Weigh a sample known to weigh 100 gm repeatedly and estimate the mean μ = 100 + β where β (gm) denotes the scale's bias. Assume s = 10 gm. Find *n* so that we estimate β with *M* = 1 and α = .05. (Note: $\hat{\beta} = \bar{X} - 100$)

$$P(|\hat{\beta} - \beta| \le M) \ge 1 - \alpha$$

$$\frac{M}{\sigma / \sqrt{n}} = 1.96 \text{ so } n \ge (1.96 \times 10/1)^2 = 385$$

Note: Decimal points of accuracy can be expensive

M	**α = .05**	**α = .01**
1	**385**	**664**
0.1	**38,416**	**66,358**
0.01	**3,841,600**	**6,635,776**

Apply to Estimating a Pop. Proportion

Unknown proportion *p* typically estimated by sample proportion: $\hat{p} = X/n$

CLT: For large *n*, the sample proportion,

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n} \right)$$

For fixed *M*, how large should *n* be so that

$$P(|\hat{p} - p| \leq M) \approx P(|Z| \leq \frac{M}{\sqrt{p(1-p)}/\sqrt{n}}) \geq 1-\alpha$$

Note: we need a guess, say p^*, of *p*. Approaches:

1. **Based on past data and other information.**
2. **Choosing p^*= .5 is conservative.**

Example results: α = .05 and *p** = 0.50

$$P(|Z| \le \frac{M}{\sqrt{.50(.50)}/\sqrt{n}}) \ge .95$$

gives

$$\frac{M}{\sqrt{.5(1-.5)}/\sqrt{n}} = 2,$$

so $n \ge 1/M^2$

M	***n***	***M* in %**
.025	**1600**	**2.5%**
.02	**2500**	**2.0%**
.01	**10,000**	**1.0%**
.005	**250,000**	**0.50%**

- **This is why "statistics" works: increasing *n* from 10,000 to 250,000 reduces the *M* very little.**
- **If samples aren't representative, even millions of observations may be useless or misleading.**

B. Simulation --- Monte Carlo

1. Suppose *X* ~ *f(x)* (stated for cont. RV, but applies to discrete RV's too). We need *E*(*h*(*X*)) for some function *h*, but the calculation is very difficult.

i. Simulate X_1 , X_2 , …, X_n iid *f(x)*

ii. Compute *h*(X_1), *h*(X_2), …, *h*(X_n) and find

$$\bar{h} = \sum_{i=1}^{n} h(X_i)/n$$

iii. CLT: for n large,

$$\bar{h} \approx N(E(h(X)), \sigma_h^2/n)$$

iv. We need to estimate σ_h^2

Most use $s_h^2 = \sum_{i=1}^{n} (h(X_i) - \bar{h})^2 / (n-1)$

Recall: $n \to \infty \Rightarrow S_h^2 \to \sigma_h^2$ with prob $= 1$

v. We can apply the sample size calculations above to choose *n* to control accuracy.

2. Suppose $X \sim f(x) > 0$, for $0 < x < 1$, but it is hard to simulate from f. Recall that

$$E(h(X)) = \int_0^1 h(x)\,f(x)\,dx$$

Note: if Y ~ Uniform(0,1) [i.e, pdf = 1 on (0,1)]

$$E(h(X)) = \int_0^1 h(y)\,f(y)\,dy = E(h(Y)\,f(Y))$$

i. **Simulate Y_1, Y_2, …, Y_n iid Uniform(0,1)**

ii. **Compute $h(Y_1)f(Y_1)$, $h(Y_2)f(Y_2)$, …, $h(Y_n)f(Y_n)$ and find** $\overline{hf} = \sum_{i=1}^{n} h(X_i)\,f(X_i)/n$

iii. CLT: for *n* large,

$$\overline{hf} \approx N(E(h(X)), \sigma^2_{hf} / n)$$

Proceed as above

3. Numerical integration. Estimate the integral

$$I = \int_0^1 h(x)dx$$

Note that $I = E(h(X))$ **where X ~ Uniform(0,1)**

i. Simulate X_1 , X_2 , …, X_n iid Uniform(0,1)

ii. Compute h(X_1), h(X_2), …, h(X_n) and find

$$\hat{I} = \overline{h} = \sum_{i=1}^{n} h(X_i) / n$$

and proceed as above

Remarks

1. **Monte Carlo integration**
   i. **Purely deterministic problem approached via probabilistic methods.**
   ii. **Real value: Estimating high dimensional integrals**
2. **I've just scratched the surface of applications of Monte Carlo.**
3. **Key: We obtain estimates and probabilistic error bounds. When simulation is cheap, we can make these errors arbitrarily small with very high prob.**