Chapter 5. Joint Probability Distributions and Random Sample

Weiqi Luo (骆伟祺)
School of Data & Computer Science
Sun Yat-Sen University
Email：luoweiqi@mail.sysu.edu.cn

Chapter 5: Joint Probability Distributions and Random Sample

- 5.1. Jointly Distributed Random Variables
- 5.2. Expected Values, Covariance, and Correlation
- 5. 3. Statistics and Their Distributions
- 5.4. The Distribution of the Sample Mean
- 5.5. The Distribution of a Linear Combination

5.1. Jointly Distributed Random Variables

- The Joint Probability Mass Function for Two Discrete Random Variables

Let X and Y be two discrete random variables defined on the sample space $\boldsymbol{S}$ of an experiment. The joint probability mass function $p(x,y)$ is defined for each pair of numbers (x,y) by

$$p(x, y) = P(X = x \quad and \quad Y = y)$$

5.1. Jointly Distributed Random Variables

- Let A be any set consisting of pairs of (x,y) values. Then the probability $P[(X,Y)\in A]$ is obtained by summing the joint pmf over pairs in A:

$$p[(X,Y)\in A]=\sum_{(x,y)\in A}\sum p(x,y)$$

- Two requirements for a pmf

$$p(x,y)\geq 0 \qquad \sum_{x}\sum_{y} p(x,y)=1$$

5.1. Jointly Distributed Random Variables

- Example 5.1

A large insurance agency services a number of customers who have purchased both a homeowner's policy and an automobile policy from the agency. For each type of policy, a deductible amount must be specified. For an automobile policy, the choices are \$100 and \$250, whereas for a homeowner's policy the choices are 0, \$100, and \$200.

Suppose an individual with both types of policy is selected at random from the agency's files. Let X = the deductible amount on the auto policy, Y = the deductible amount on the homeowner's policy

Joint Probability Table

			y	
	$p(x,y)$	0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

5.1. Jointly Distributed Random Variables

- Example 5.1 (Cont')

	$p(x,y)$	y: 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

p(100,100) =P(X=100 and Y=100) = 0.10

P(Y ≥ 100) = p(100,100) + p(250,100) + p(100,200) + p(250,200) = 0.75

5.1. Jointly Distributed Random Variables

- The marginal probability mass function

  The marginal probability mass functions of X and Y, denoted by $p_X(x)$ and $p_Y(y)$, respectively, are given by

$$p_X(x) = \sum_y p(x,y); \quad p_Y(y) = \sum_x p(x,y)$$

$\mathbf{p_Y}$

	$\mathbf{Y_1}$	$\mathbf{Y_2}$	…	$\mathbf{Y_{m-1}}$	$\mathbf{Y_m}$
X_1	$p_{1,1}$	$p_{1,2}$		$p_{1,m-1}$	$p_{1,m}$
X_2	$p_{2,1}$	$p_{2,2}$		$p_{2,m-1}$	$p_{2,m}$
…					
X_{n-1}	$p_{n-1,m}$	$p_{n-1,m}$		$p_{n-1,m}$	$p_{n-1,m}$
X_n	$p_{n,m}$	$p_{n,m}$		$p_{n,m}$	$p_{n,m}$

$\mathbf{p_X}$

5.1. Jointly Distributed Random Variables

- Example 5.2 (Ex. 51. Cont')

  The possible X values are x=100 and x=250, so computing row totals in the joint probability table yields

	$p(x,y)$	y = 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

$p_x(100)=p(100,0)+p(100,100)+p(100,200)=0.5$

$p_x(250)=p(250,0)+p(250,100)+p(250,200)=0.5$

$$p_x(x)=\begin{cases}0.5, x=100, 250\\ 0, otherwise\end{cases}$$

5.1. Jointly Distributed Random Variables

- Example 5.2 (Cont')

	$p(x,y)$	y = 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

$p_y(0)=p(100,0)+p(250,0)=0.2+0.05=0.25$

$p_y(100)=p(100,100)+p(250,100)=0.1+0.15=0.25$

$p_y(200)=p(100,200)+p(250,200)=0.2+0.3=0.5$

$$p_Y(y)=\begin{cases}0.25, y=0,100\\ 0.5, y=200\\ 0, otherwise\end{cases}$$

$P(Y \geq 100) = p(100,100) + p(250,100) + p(100,200) + p(250,200)$
$= p_Y(100)+p_Y(200) = 0.75$

5.1. Jointly Distributed Random Variables

- The Joint Probability Density Function for Two Continuous Random Variables

Let X and Y be two continuous random variables. Then f(x,y) is the joint probability density function for X and Y if for any two-dimensional set A

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

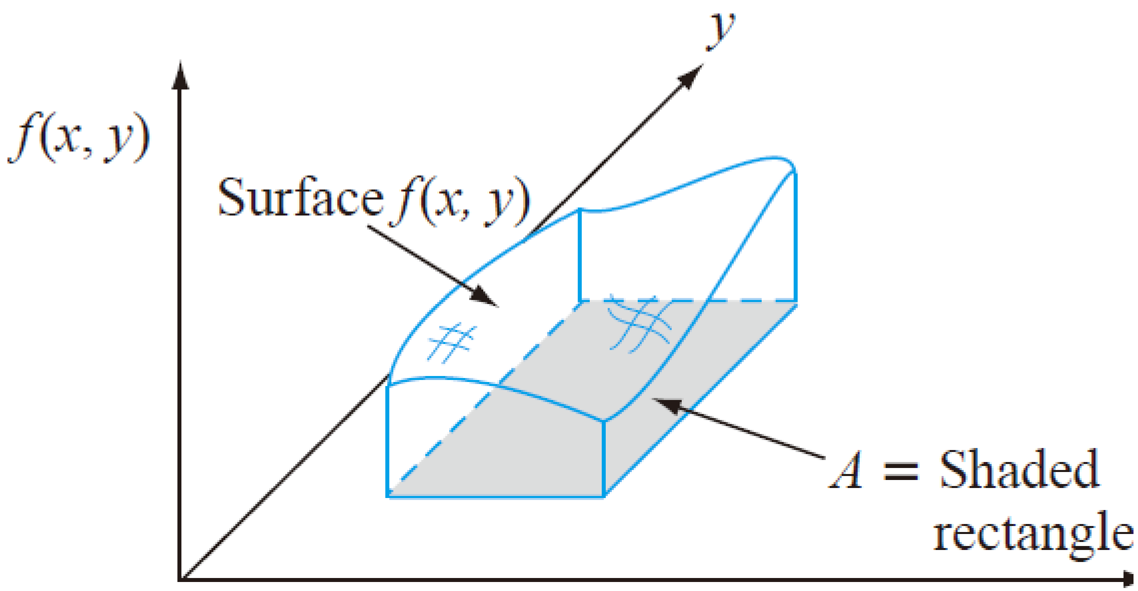
Two requirements for a joint pdf

1. $f(x,y) \geq 0$; for all pairs (x,y) in R^2
2. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dxdy = 1$

5.1. Jointly Distributed Random Variables

- In particular, if A is the two-dimensional rectangle $\{(x,y): a \le x \le b,\ c \le y \le d\}$, then

$$P[(X,Y) \in A] = P(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f(x,y)\,dy\,dx$$



5.1. Jointly Distributed Random Variables

- Example 5.3

A bank operates both a drive-up facility and a walk-up window. On a randomly selected day, let X = the proportion of time that the drive-up facility is in use, Y = the proportion of time that the walk-up window is in use. Let the joint pdf of (X,Y) be

$$f(x,y)=\begin{cases}\frac{6}{5}(x+y^2) & 0\le x\le 1, 0\le y\le 1\\ 0 & otherwise\end{cases}$$

1. Verify that $f(x,y)$ is a joint probability density function;
2. Determine the probability $P(0\le X\le \frac{1}{4}, 0\le Y\le \frac{1}{4})$

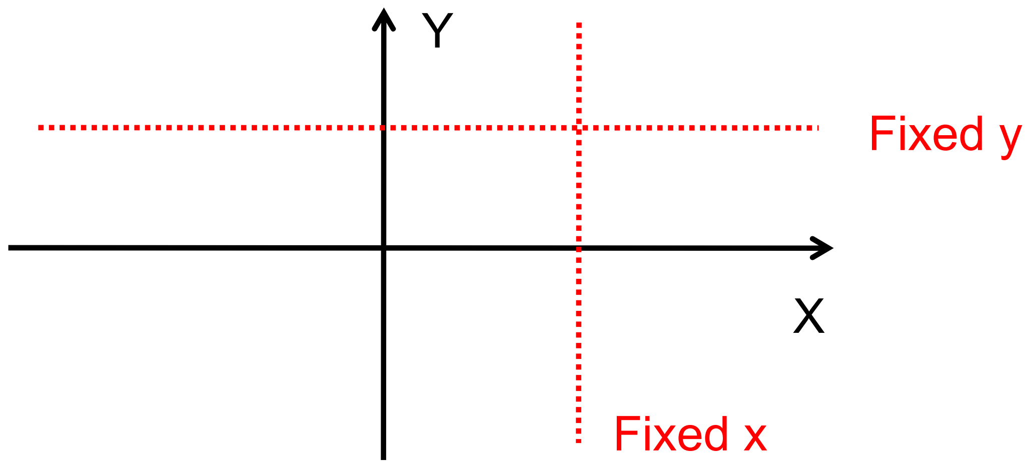
5.1. Jointly Distributed Random Variables

- Marginal Probability density function

The marginal probability density functions of X and Y, denoted by $f_X(x)$ and $f_Y(y)$, respectively, are given by

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

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



5.1. Jointly Distributed Random Variables

- Example 5.4 (Ex. 5.3 Cont')

The marginal pdf of X, which gives the probability distribution of busy time for the drive-up facility without reference to the walk-up window, is

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \int_0^1 \frac{6}{5}(x+y^2)dy = \frac{6}{5}x + \frac{2}{5}$$

for x in (0,1); and 0 for otherwise.

$$f_Y(y) = \begin{cases} \frac{6}{5}y^2 + \frac{3}{5} & 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

Then

$$P(\frac{1}{4} \le Y \le \frac{3}{4}) = \int_{1/4}^{3/4} f_Y(y)dy = 0.4625$$

5.1. Jointly Distributed Random Variables

- Example 5.5

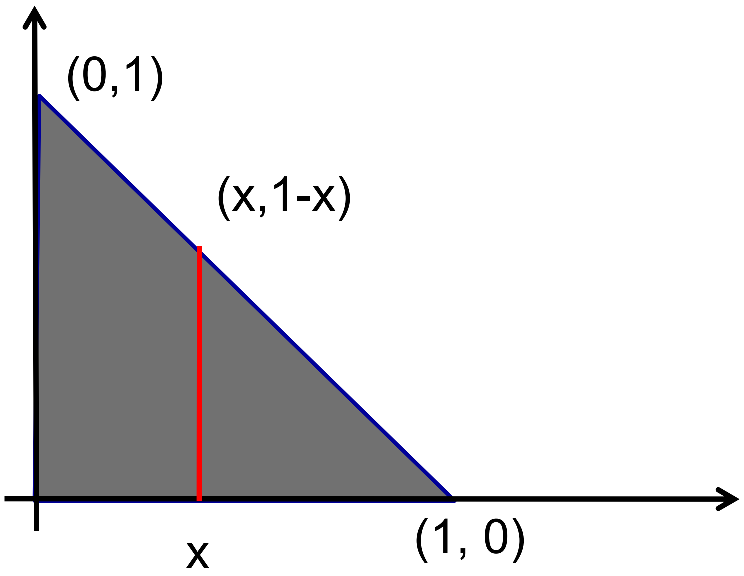
A nut company markets cans of deluxe mixed nuts containing almonds, cashews, and peanuts. Suppose the net weight of each can is exactly 1 lb, but the weight contribution of each type of nut is random. Because the three weights sum to 1, a joint probability model for any two gives all necessary information about the weight of the third type. Let X = the weight of almonds in a selected can and Y = the weight of cashews. The joint pdf for (X,Y) is

$$f(x,y)=\begin{cases}24xy & 0\le x\le 1, 0\le y\le 1, x+y\le 1\\ 0 & otherwise\end{cases}$$

5.1. Jointly Distributed Random Variables

- Example 5.5 (Cont')

$$f(x,y)=\begin{cases}24xy & 0\le x\le 1, 0\le y\le 1, x+y\le 1\\ 0 & otherwise\end{cases}$$



1: f(x,y) ≥ 0

$$2:\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dydx=\iint_D f(x,y)dydx$$

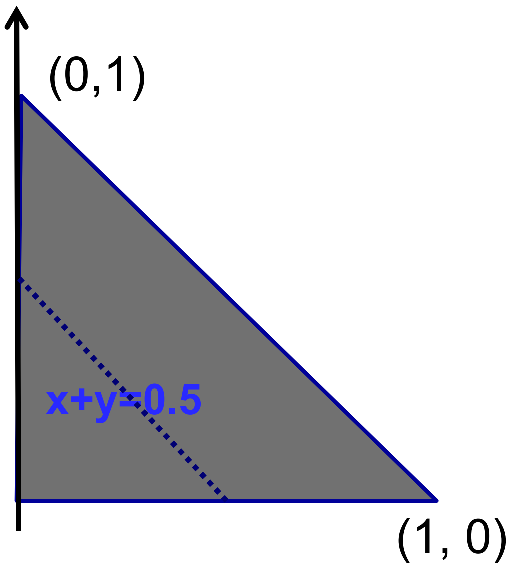
$$=\int_0^1\{\int_0^{1-x}(24xy)dy\}dx$$

$$=\int_0^1 12x(1-x)^2dx=1$$

5.1. Jointly Distributed Random Variables

- Example 5.5 (Cont')

Let the two type of nuts together make up at most 50% of the can, then A={(x,y); 0≤x ≤1; 0 ≤ y ≤ 1, x+y ≤ 0.5}



$$P((X,Y)\in A)=\iint_A f(x,y)dydx$$

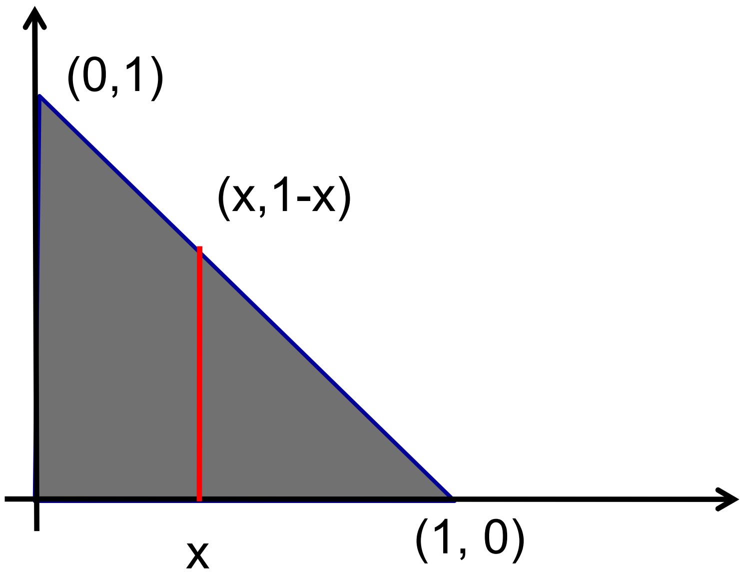
$$=\int_0^{0.5}\{\int_0^{0.5-x}(24xy)dy\}dx$$

$$=0.625$$

5.1. Jointly Distributed Random Variables

- Example 5.5 (Cont')

The marginal pdf for almonds is obtained by holding X fixed at x and integrating f(x,y) along the vertical line through x:



$$f_X(x) = \int_0^{0.5} f(x,y)dy$$

$$= \begin{cases} \int_0^{1-x} (24xy)dy = 12x(1-x)^2, 0 \le x \le 1 \\ 0, otherwise \end{cases}$$

5.1. Jointly Distributed Random Variables

- Independent Random Variables

Two random variables X and Y are said to be independent if for every pair of x and y values,

$$p(x, y) = p_X(x) \cdot p_Y(y) \quad \text{when } X \text{ and } Y \text{ are discrete}$$

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{when } X \text{ and } Y \text{ are continuous}$$

Otherwise, X and Y are said to be dependent.

Namely, two variables are independent if their joint pmf or pdf is the product of the two marginal pmf's or pdf's.

5.1. Jointly Distributed Random Variables

- Example 5.6

In the insurance situation of Example 5.1 and 5.2

$$p(100,100) = 0.1 \neq (0.5)(0.25) = p_X(100)\, p_Y(100)$$

	$p(x,y)$	y: 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

So, X and Y are not independent.

5.1. Jointly Distributed Random Variables

- Example 5.7 (Ex. 5.5 Cont')

  Because f(x,y) has the form of a product, X and Y would appear to be independent. However, although

$$f_X(x) = \int_0^{1-x} (24xy)dy = 12x(1-x)^2$$

$$f_Y(y) = \int_0^{1-y} (24xy)dx = 12y(1-y)^2$$

By symmetry

$$f_X(\frac{3}{4})f_Y(\frac{3}{4}) = \frac{9}{16}, f(x,y) = 0 \neq \frac{9}{16} \bullet \frac{9}{16}$$

5.1. Jointly Distributed Random Variables

- Example 5.8

Suppose that the lifetimes of two components are independent of one another and that the first lifetime, X_1, has an exponential distribution with parameter λ_1 whereas the second, X_2, has an exponential distribution with parameter λ_2. Then the joint pdf is

$$f(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) = \begin{cases} \lambda_1 \lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2} & x_1 > 0, x_2 > 0 \\ 0 & otherwise \end{cases}$$

Let λ_1 =1/1000 and λ_2=1/1200. So that the expected lifetimes are 1000 and 1200 hours, respectively. The probability that both component lifetimes are at least 1500 hours is

$$P(1500 \le X_1, 1500 \le X_2) = P(1500 \le X_1)P(1500 \le X_2)$$

5.1. Jointly Distributed Random Variables

- More than Two Random Variables

If $X_1, X_2, \ldots, X_n$ are all discrete rv's, the joint pmf of the variables is the function

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

If the variables are continuous, the joint pdf of X_1, X_2, …, X_n is the function $f(x_1, x_2, \ldots, x_n)$ such that for any n intervals $[a_1, b_1]$, …, $[a_n, b_n]$,

$$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_b) = \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1$$

5.1. Jointly Distributed Random Variables

- Independent

The random variables X_1, X_2, …X_n are said to be independent if for every subset X_{i1}, X_{i2},…, X_{ik} of the variable, the joint pmd or pdf of the subset is equal to the product of the marginal pmf's or pdf's.

5.1. Jointly Distributed Random Variables

- Multinomial Experiment

An experiment consisting of n independent and identical trials, in which each trial can result in any one of r possible outcomes. Let p_i=P(Outcome i on any particular trial), and define random variables by X_i=the number of trials resulting in outcome i (i=1,…,r). The joint pmf of $X_1,\dots,X_r$ is called the multinomial distribution.

$$p(x_1,\dots,x_r)=\begin{cases}\dfrac{n!}{(x_1!)(x_2!)\dots(x_r!)}p_1^{x1}\dots p_r^{xr}, x_i=0,1\dots with x_1+x_2\dots+x_r=n\\ 0\end{cases}$$

Note: the case r=2 gives the binomial distribution.

5.1. Jointly Distributed Random Variables

- Example 5.9

If the allele of each of then independently obtained pea sections id determined and p_1=P(AA), p_2=P(Aa), p_3=P(aa), X_1= number of AA's, X_2=number of Aa's and X_3=number of aa's, then

$$p(x_1, x_2, x_3) = \frac{10!}{(x_1!)(x_2!)\ldots(x_r!)} p_1^{x1} p_2^{x2} p_3^{x3}, x_i = 0,1,..and x_1 + x_2 + x_3 = 10$$

If p_1=p_3=0.25, p_2=0.5, then

$$P(x_1 = 2, x_2 = 5, x_3 = 3) = p(2,5,3) = 0.0769$$

5.1. Jointly Distributed Random Variables

- Example 5.10

  When a certain method is used to collect a fixed volume of rock samples in a region, there are four resulting rock types. Let X_1, X_2, and X_3 denote the proportion by volume of rock types 1, 2 and 3 in a randomly selected sample. If the joint pdf of X_1,X_2 and X_3 is

$$f(x_1, x2, x3) = \begin{cases} kx_1x_2(1-x_3), 0 \le x_1 \le 1, 0 \le x_2 \le 1, 0 \le x_3 \le 1, x_1 + x_2 + x_3 \le 1 \\ 0, otherwise \end{cases}$$

$$\iiint_{D_1} f(x_1, x2, x3) = 1, D_1 : -\infty \le x_i \le \infty, i = 1,2,3 \implies \text{k=144.}$$

$$\iiint_{D_2} f(x_1, x2, x3) = 0.6066, D_2 : X_1 + X_2 \le 0.5$$

5.1. Jointly Distributed Random Variables

- Example 5.11

  If $X_1, \ldots, X_n$ represent the lifetime of n components, the components operate independently of one another, and each lifetime is exponentially distributed with parameter, then

$$f(x_1, x_2, \ldots x_n) = (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2})\ldots(\lambda e^{-\lambda x_n})$$

$$= \begin{cases} \lambda^n e^{-\lambda \sum xi}, x_1 \geq 0; x_2 \geq 0; \ldots, x_n \geq 0; \\ 0, otherwise \end{cases}$$

5.1. Jointly Distributed Random Variables

- Example 5.11 (Cont')

  If there n components constitute a system that will fail as soon as a single component fails, then the probability that the system lasts past time is

$$P(X_1 > t, X_2 > t, \ldots, X_n > t) = \int_t^\infty \ldots \int_t^\infty f(x_1, x_2, \ldots, x_n) dx_1 \ldots dx_n$$

$$= \left(\int_t^\infty \lambda e^{-\lambda x_1} dx_1\right) \ldots \left(\int_t^\infty \lambda e^{-\lambda x_n} dx_n\right) = e^{-n\lambda t}$$

therefore,

$$P(systemlifetime \le t) = 1 - e^{-n\lambda t}, fort \ge 0$$

5.1. Jointly Distributed Random Variables

- Conditional Distribution

Let X and Y be two continuous rv's with joint pdf f(x,y) and marginal X pdf $f_X(x)$. Then for any X values x for which $f_X(x)>0$, the conditional probability density function of Y given that X=x is

$$f_{Y|X}(y\,|\,x)=\frac{f(x,y)}{f_X(x)},-\infty<y<\infty$$

If X and Y are discrete, then

$$f_{Y|X}(y\,|\,x)=\frac{p(x,y)}{p_X(x)},-\infty<y<\infty$$

is the conditional probability mass function of Y when X=x.

5.1. Jointly Distributed Random Variables

- Example 5.12 (Ex.5.3 Cont')

X= the proportion of time that a bank's drive-up facility is busy and Y=the analogous proportion for the walk-up window. The conditional pdf of Y given that X=0.8 is

$$f_{Y|X}(y\,|\,0.8) = \frac{f(0.8, y)}{f_X(0.8)} = \frac{1.2(0.8+y^2)}{1.2(0.8)+0.4} = \frac{1}{34}(24+30y^2), 0<y<1$$

The probability that the walk-up facility is busy at most half the time given that X=0.8 is then

$$f_{Y|X}(y \le 0.5\,|\,X=0.8) = \int_{-\infty}^{0.5} f_{Y|X}(y\,|\,0.8)dy = \int_{-\infty}^{0.5} \frac{1}{34}(24+30y^2)dy = 0.39$$

5.1. Jointly Distributed Random Variables

- Homework

  Ex. 8, Ex.12, Ex.18, Ex.20

- The Expected Value of a function h(x,y)

Let X and Y be jointly distribution rv's with pmf $p(x,y)$ or pdf $f(x,y)$ according to whether the variables are discrete or continuous. Then the expected value of a function $h(X,Y)$, denoted by $E[h(X,Y)]$ or $\mu_{h(X,Y)}$, is given by

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)\cdot p(x,y), X \,\&\, Y : discrete \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\cdot f(x,y)dxdy, X \,\&\, Y : continuous \end{cases}$$

- Example 5.13

Five friends have purchased tickets to a certain concert. If the tickets are for seats 1-5 in a particular row and the tickets are randomly distributed among the five, what is the expected number of seats separating any particular two of the five?

$$p(x, y) = \begin{cases} \dfrac{1}{20} & x = 1, \dots, 5; y = 1, \dots, 5; x \neq y \\ 0 & \text{otherwise} \end{cases}$$

The number of seats separating the two individuals is

$$h(X,Y)=|X-Y|-1$$

- Example 5.13 (Cont')

$h(x,y)$		x 1	2	3	4	5
	1	--	0	1	2	3
	2	0	--	0	1	2
y	3	1	0	--	0	1
	4	2	1	0	--	0
	5	3	2	1	0	--

$$E[h(X,Y)] = \sum_{(x,y)}\sum h(x,y)\cdot p(x,y)$$

$$= \sum_{x=1}^{5}\sum_{\substack{y=1 \\ x\neq y}}^{5} (|x-y|-1)\cdot\frac{1}{20} = 1$$

- Example 5.14

In Example 5.5, the joint pdf of the amount X of almonds and amount Y of cashews in a 1-lb can of nuts was

$$f(x,y)=\begin{cases}24xy & 0\le x\le 1, 0\le y\le 1, x+y\le 1\\ 0 & \text{otherwise}\end{cases}$$

If 1 lb of almonds costs the company \$1.00, 1 lb of cashews costs \$1.50, and 1 lb of peanuts costs \$0.50, then the total cost of the contents of a can is

$$h(X,Y)=(1)X+(1.5)Y+(0.5)(1-X-Y)=0.5+0.5X+Y$$

- Example 5.14 (Cont')

The expected total cost is

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

$$= \int_{0}^{1}\int_{0}^{1-x} (0.5 + 0.5x + y)\cdot 24xy\,dy\,dx = \$1.10$$

Note: The method of computing $E[h(X_1,\ldots, X_n)]$, the expected value of a function $h(X_1, \ldots, X_n)$ of n random variables is similar to that for two random variables.

5.2 Expected Values, Covariance, and Correlation

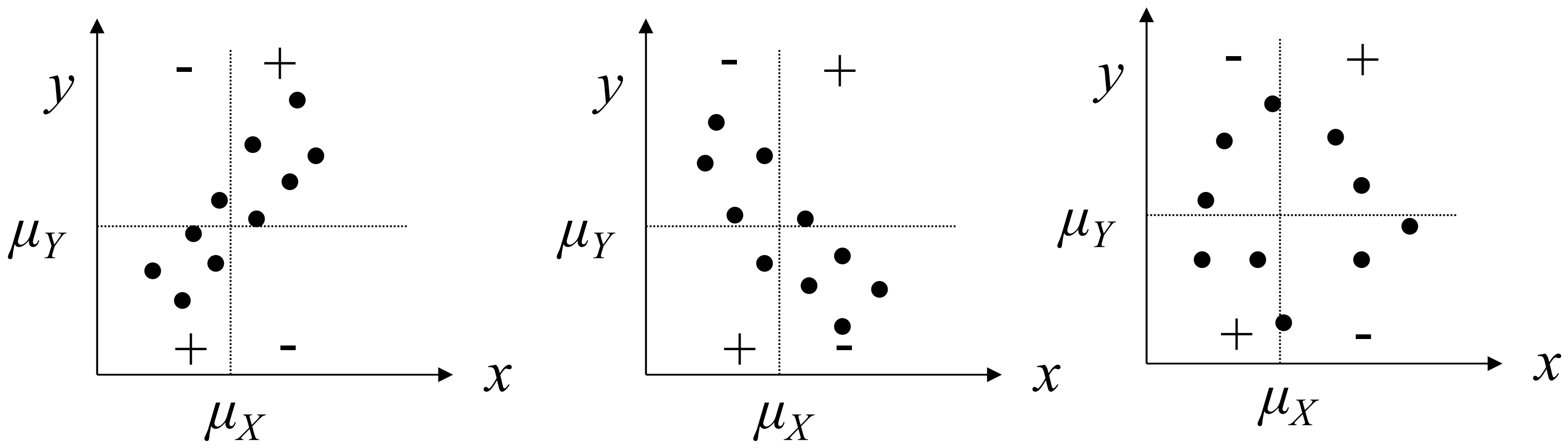
- Covariance

  The Covariance between two rv's X and Y is

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

$$= \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y) & X,Y \text{ discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f(x,y)dxdy & X,Y \text{ continuous} \end{cases}$$

- Illustrates the different possibilities.



(a) positive covariance (b) negative covariance; (c) covariance near zero

Here: P(x, y) =1/10

- Example 5.15

The joint and marginal pmf's for X = automobile policy deductible amount and Y = homeowner policy deductible amount in Example 5.1 were

$p(x,y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

x	100	250
$p_X(x)$	.5	.5

y	0	100	250
$p_Y(y)$	.25	.25	.5

From which $\mu_X=\sum x p_X(x)=175$ and $\mu_Y=125$. Therefore

$$\mathrm{Cov}(X,Y)=\sum_{(x,y)}\sum (x-175)(y-125)\,p(x,y)$$

$$=(100-175)(0-125)(0.2)+\ldots+(250-175)(200-125)(0.3)\ =1875$$

- Proposition

$$\text{Cov}(X,Y) = E(XY) - \mu_X \mu_Y$$

Note: $$\text{Cov}(X,X) = E(X^2) - \mu^2_X = V(X)$$

- Example 5.16 (Ex. 5.5 Cont')

  The joint and marginal pdf's of X = amount of almonds and Y = amount of cashews were

$$f(x,y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1, x+y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- Example 5.16 (Cont')

$$f_X(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$f_Y(y)$ can be obtained through replacing x by y in $f_X(x)$. It is easily verified that $\mu_X = \mu_Y = 2/5$, and

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy \quad = \int_0^1\int_0^{1-x} xy \cdot 24xydydx = 8\int_0^1 x^2(1-x)^3 dx = 2/15$$

Thus Cov(X,Y) = 2/15 - $(2/5)^2$ = 2/15 - 4/25 = -2/75. A negative covariance is reasonable here because more almonds in the can implies fewer cashews.

- Correlation

  The correlation coefficient of X and Y, denoted by Corr(X,Y), $\rho_{X,Y}$ or just ρ, is defined by

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

The normalized version of Cov(X,Y)

- Example 5.17

  It is easily verified that in the insurance problem of Example 5.15, $\sigma_X = 75$ and $\sigma_Y = 82.92$. This gives

$$\rho = 1875/(75)(82.92)=0.301$$

- Proposition

1. If a and c are either both positive or both negative

$$\mathrm{Corr}(aX+b,\ cY+d) = \mathrm{Corr}(X,Y)$$

2. For any two rv's X and Y, $-1 \le \mathrm{Corr}(X,Y) \le 1$.

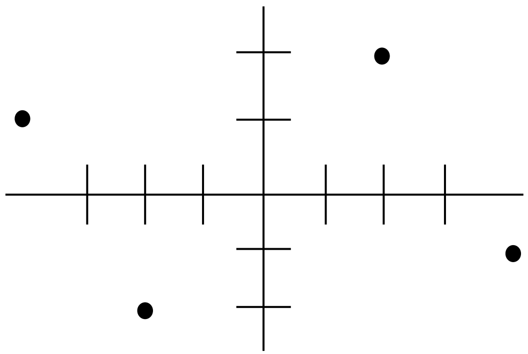
3. If X and Y are independent, then $\rho = 0$, but $\rho = 0$ does not imply independence.

4. $\rho = 1$ or -1 **iff** $Y = aX+b$ for some numbers a and b with $a \neq 0$.

- Example 5.18

  Let X and Y be discrete rv's with joint pmf

$$p(x,y)=\begin{cases}\dfrac{1}{4} & (x,y)=(-4,1),(4,-1),(2,2)(-2,-2)\\ 0 & \text{otherwise}\end{cases}$$



It is evident from the figure that the value of X is completely determined by the value of Y and vice versa, so the two variables are completely dependent.

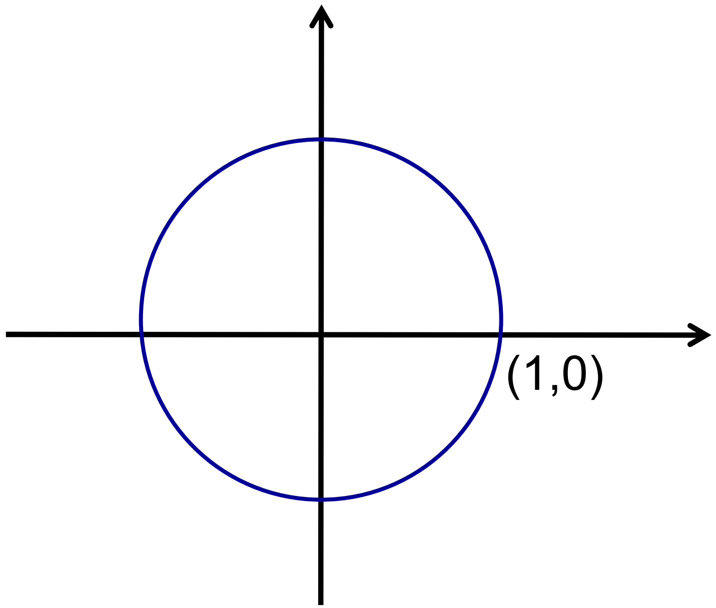
However, by symmetry $\mu_X = \mu_Y = 0$ and $E(XY)$ = (-4)1/4 + (-4)1/4 + (4)1/4 + (4)1/4 = 0, so Cov(X,Y) = $E(XY)$ - $\mu_X\mu_Y$ = 0 and thus $\rho_{XY} = 0$.

Although there is perfect dependence, there is also complete absence of any linear relationship!

- Another Example

X and Y are uniform distribution in an unit circle

$$p(x,y)=\begin{cases}\frac{1}{\pi}, x^2+y^2\le 1\\ 0, otherwise\end{cases}$$



Obviously, X and Y are dependent.
However, we have

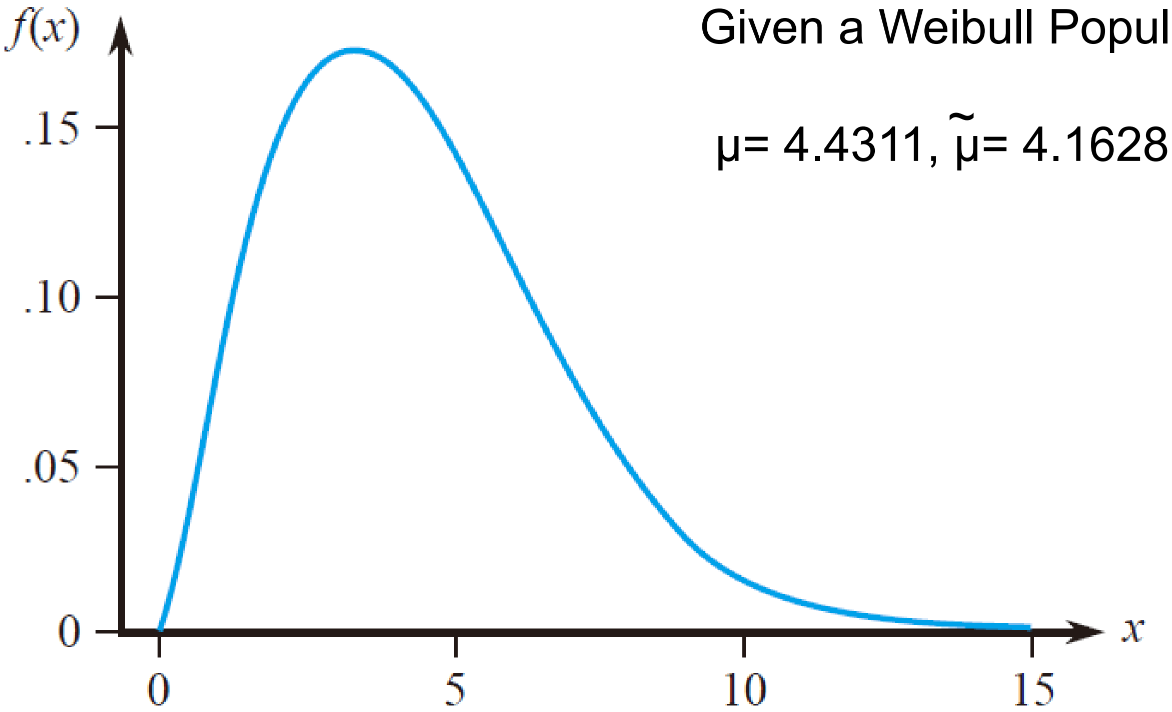
$$Cov(X,Y)=0$$

- Homework

  Ex. 23, Ex. 26, Ex. 33, Ex. 35

5.3 Statistics and Their Distributions

- Example 5.19



Given a Weibull Population with $\alpha=2$, $\beta=5$

$\mu = 4.4311$, $\tilde{\mu} = 4.1628$, $\delta = 2.316$

5.3 Statistics and Their Distributions

- Example 5.19 (Cont')

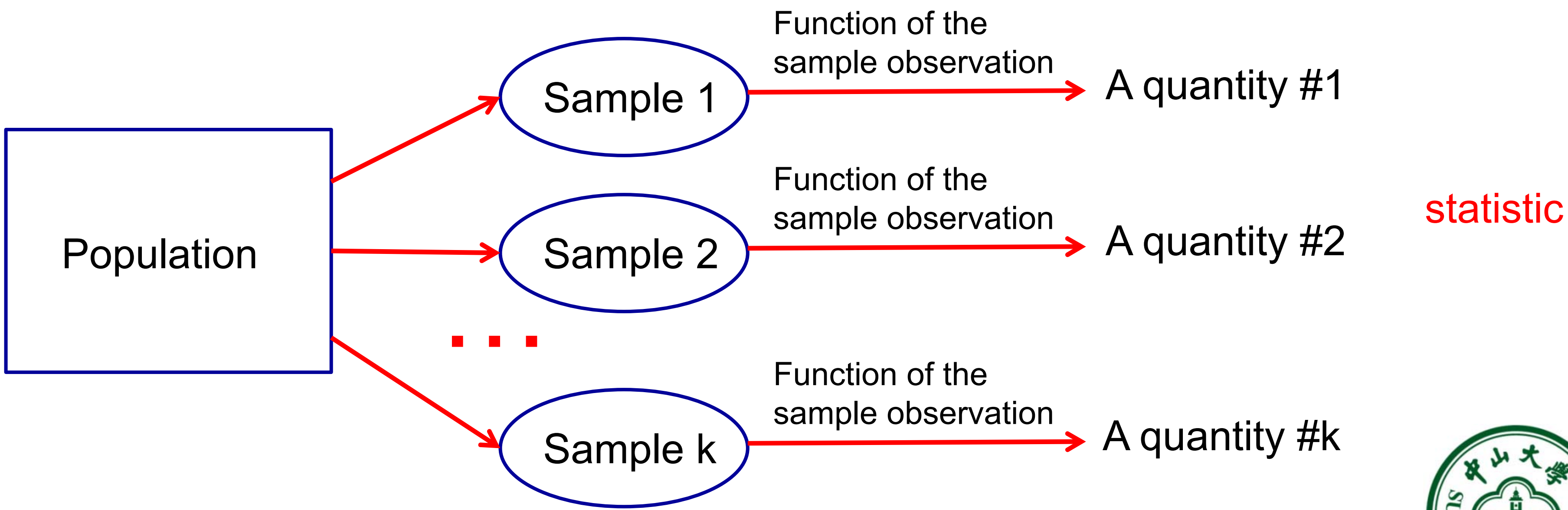
Table 5.1 Samples from the Weibull Distribution of Example 5.19

Sample	1	2	3	4	5	6
1	6.1171	5.07611	3.46710	1.55601	3.12372	8.93795
2	4.1600	6.79279	2.71938	4.56941	6.09685	3.92487
3	3.1950	4.43259	5.88129	4.79870	3.41181	8.76202
4	0.6694	8.55752	5.14915	2.49759	1.65409	7.05569
5	1.8552	6.82487	4.99635	2.33267	2.29512	2.30932
6	5.2316	7.39958	5.86887	4.01295	2.12583	5.94195
7	2.7609	2.14755	6.05918	9.08845	3.20938	6.74166
8	10.2185	8.50628	1.80119	3.25728	3.23209	1.75468
9	5.2438	5.49510	4.21994	3.70132	6.84426	4.91827
10	4.5590	4.04525	2.12934	5.50134	4.20694	7.26081
$\bar{x}$	4.401	5.928	4.229	4.132	3.620	5.761
$\tilde{x}$	4.360	6.144	4.608	3.857	3.221	6.342
s	2.642	2.062	1.611	2.124	1.678	2.496

5.3 Statistics and Their Distributions

- Example 5.19 (Cont')

Sample	1	2	3	4	5	6
Mean	4.401	5.928	4.229	4.132	3.620	5.761
Median	4.360	6.144	4.608	3.857	3.221	6.342
Standard Deviation	2.642	2.062	1.611	2.124	1.678	2.496



5.3 Statistics and Their Distributions

- Statistic

  A statistic is any quantity whose value can be calculated from **sample data** (with a function).

- Prior to obtaining data, there is **uncertainty** as to what value of any particular statistic will result. Therefore, a statistic is a random variable. A statistic will be denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value of the statistic.

- The probability distribution of a statistic is sometimes referred to as its **sampling distribution**. It describes how the statistic varies in value across all samples that might be selected.

5.3 Statistics and Their Distributions

- The probability distribution of any particular statistic depends on

1. The population distribution, *e.g.* the normal, uniform, etc. , and the corresponding parameters
2. The sample size n
3. The method of sampling, *e.g.* sampling with replacement or without replacement

5.3 Statistics and Their Distributions

- Example

  Consider selecting a sample of size $n = 2$ from a population consisting of just the three values 1, 5, and 10, and suppose that the statistic of interest is the sample variance.

- If sampling is done "**with replacement**", then $S^2 = 0$ will result if $X_1 = X_2$.
- If sampling is done "**without replacement**", then S^2 can not equal 0.

5.3 Statistics and Their Distributions

- Random Sample

The rv's $X_1, X_2, \ldots, X_n$ are said to form a (simple) random sample of size n if

1. The Xi*'s* are independent rv's.
2. Every Xi has the same probability distribution.

When conditions 1 and 2 are satisfied, we say that the X_i's are independent and identically distributed (**i.i.d**)

Note: Random sample is one of commonly used sampling methods in practice.

5.3 Statistics and Their Distributions

- Random Sample

- Sampling with replacement or from an infinite population is random sampling.

- Sampling without replacement from a finite population is generally considered not random sampling. However, if the sample size n is much smaller than the population size N ($n/N \leq 0.05$), it is approximately random sampling.

Note: The virtue of random sampling method is that the probability distribution of any statistic can be more easily obtained than for any other sampling method.

5.3 Statistics and Their Distributions

- Deriving the Sampling Distribution of a Statistic

> Method #1: Calculations based on probability rules
>
> ***e.g.*** Example 5.20 & 5.21

> Method #2:
>
> Carrying out a simulation experiments
>
> ***e.g.*** Example 5.22 & 5.23

5.3 Statistics and Their Distributions

- Example 5.20

A certain brand of MP3 player comes in three configurations: a model with 2 GB of memory, costing \$80, a 4 GB model priced at \$100, and an 8 GB version with a price tag of \$120. If 20% of all purchasers choose the 2 GB model, 30% choose the 4 GB model, and 50% choose the 8 GB model, then the probability distribution of the cost X of a single randomly selected MP3 player purchase is given by

x	80	100	120
$p(x)$	.2	.3	.5

with $\mu = 106$, $\sigma^2 = 244$

- Suppose on a particular day only two MP3 players are sold. Let X_1 the revenue from the first sale and X_2 the revenue from the second.
- Suppose that X_1 and X_2 are independent,

5.3 Statistics and Their Distributions

- Example 5.20 (Cont')

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
80	80	.04	80	0
80	100	.06	90	200
80	120	.10	100	800
100	80	.06	90	200
100	100	.09	100	0
100	120	.15	110	200
120	80	.10	100	800
120	100	.15	110	200
120	120	.25	120	0

$\bar{x}$	80	90	100	110	120
$p_{\bar{X}}(\bar{x})$	.04	.12	.29	.30	.25

s^2	0	200	800
$pS^2(s^2)$	.38	.42	.20

Known the Population Distribution

5.3 Statistics and Their Distributions

- Example 5.20 (Cont')

$\bar{x}$	80	90	100	110	120
$p_{\bar{X}}(\bar{x})$	.04	.12	.29	.30	.25

N=2

$\bar{x}$	80	85	90	95	100	105	110	115	120
$p_{\bar{X}}(\bar{x})$	.0016	.0096	.0376	.0936	.1761	.2340	.2350	.1500	.0625

N=4

Example 5.21

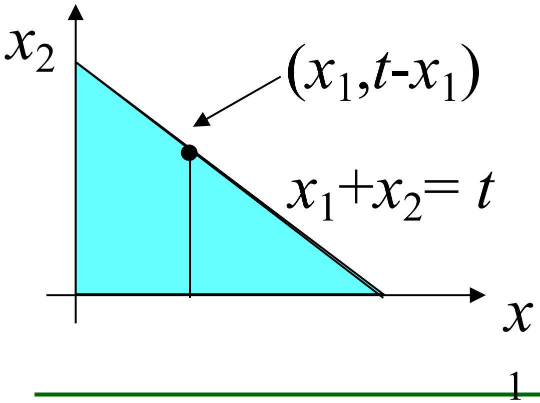
Service time for a certain type of bank transaction is a random variable having an exponential distribution with parameter λ. Suppose X_1 and X_2 are service times for two different customers, assumed independent of each other. Consider the total service time $T_o = X_1 + X_2$ for the two customers, also a statistic. What is the pdf of T_o?

The cdf of T_o is, for $t \geq 0$

$$F_{T_0}(t) = P(X_1 + X_2 \leq t) = \iint_{\{(x_1, x_2); x_1 + x_2 \leq t\}} f(x_1, x_2) dx_1 dx_2 = \int_0^t \int_0^{t-x_1} \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} dx_2 dx_1$$

$$= \int_0^t [\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t}] dx_1$$

$$= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}$$



5.3 Statistics and Their Distributions

- Example 5.21 (Cont')

The pdf of T_o is obtained by differentiating $F_{To}(t)$;

$$f_{T_0}(t) = \begin{cases} \lambda^2 t e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

This is a gamma pdf ($\alpha = 2$ and $\beta = 1/\lambda$).

The pdf of $\overline{X} = T_o/2$ is obtained from the relation $\{\overline{X} \leq \overline{x}\}$ iff $\{T_o \leq 2\overline{x}\}$ as

$$f_{\overline{X}}(\overline{x}) = \begin{cases} 4\lambda^2 \overline{x} e^{-2\lambda \overline{x}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

5.3 Statistics and Their Distributions

- Simulation Experiments

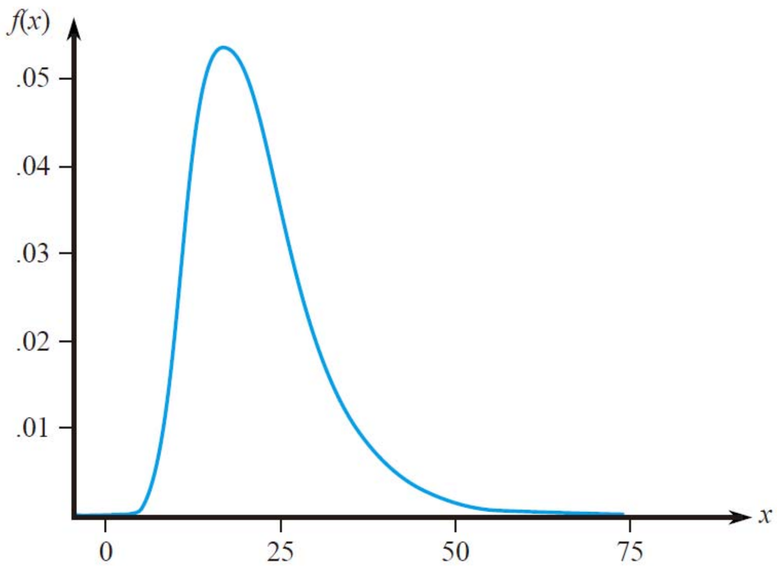
  This method is usually used when a derivation via probability rules is too difficult or complicated to be carried out. Such an experiment is virtually always done with the aid of a computer. And the following characteristics of an experiment must be specified:

- The statistic of interest (*e.g.* sample mean, S, etc.)
- The population distribution (normal with $\mu = 100$ and $\sigma = 15$, uniform with lower limit $A = 5$ and upper limit $B = 10$, etc.)
- The sample size n (*e.g.*, $n = 10$ or $n = 50$)
- The number of replications k (*e.g.*, $k = 500$ or 1000) (the actual sampling distribution emerges as $k \rightarrow \infty$)

5.3 Statistics and Their Distributions

- Example 5.23

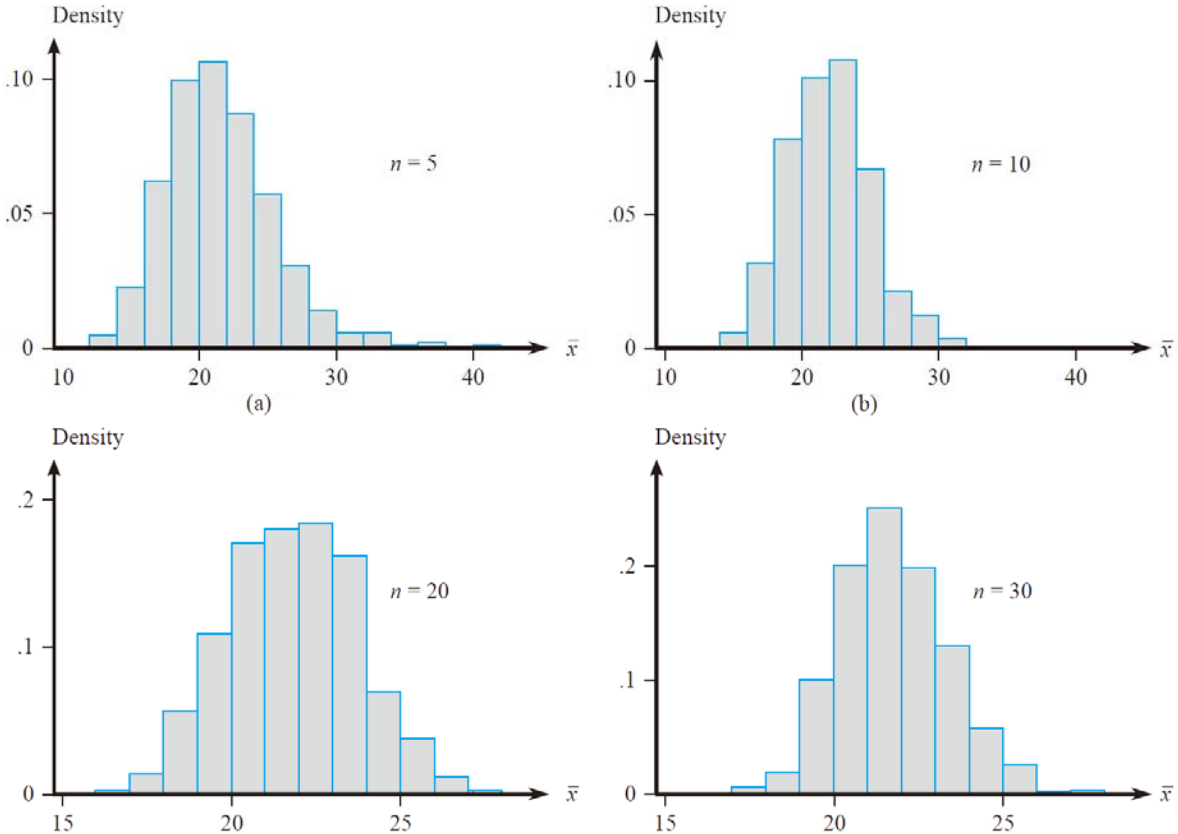
Consider a simulation experiment in which the population distribution is quite skewed. Figure shows the density curve of a certain type of electronic control (actually a lognormal distribution with $E(\ln(X)) = 3$ and $V(\ln(X))=.4$).



$E(X)=\mu=21.7584$, $V(X)=\sigma^2=82.1449$

5.3 Statistics and Their Distributions

- Example 5.23 (Cont')



1. **Center of the sampling distribution remains at the population mean.**
2. **As *n* increases:**
   - ✓ **Less skewed ("more normal")**
   - ✓ **More concentrated ("smaller variance")**

5.3 Statistics and Their Distributions

- Homework

  Ex.38, Ex.42, Ex. 43

5.4 The Distribution of the Sample Mean

- Proposition

Let X_1, X_2, …, X_n be a **random sample** (i.i.d. rv's) from a distribution with mean value μ and standard deviation σ. Then

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

$$V(\bar{X}) = \delta_{\bar{X}}^2 = \frac{\sigma^2}{n} \text{ and } \sigma_{\bar{X}} = \sigma / \sqrt{n}$$

In addition, with $T_o = X_1 + \ldots + X_n$ (the sample total),

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

Refer to 5.5 for the proof!

- Example 5.24

  In a notched tensile fatigue test on a titanium specimen, the expected number of cycles to first acoustic emission (used to indicate crack initiation) is μ = 28,000, and the standard deviation of the number of cycles is σ = 5000.

  Let $X_1, X_2, \ldots, X_{25}$ be a random sample of size 25, where each X_i is the number of cycles on a different randomly selected specimen. Then

$$E(\overline{X}) = \mu = 28,000, E(T_0) = n\mu = 25(28000) = 700,000$$

The standard deviations of $\overline{X}$ and T_o are

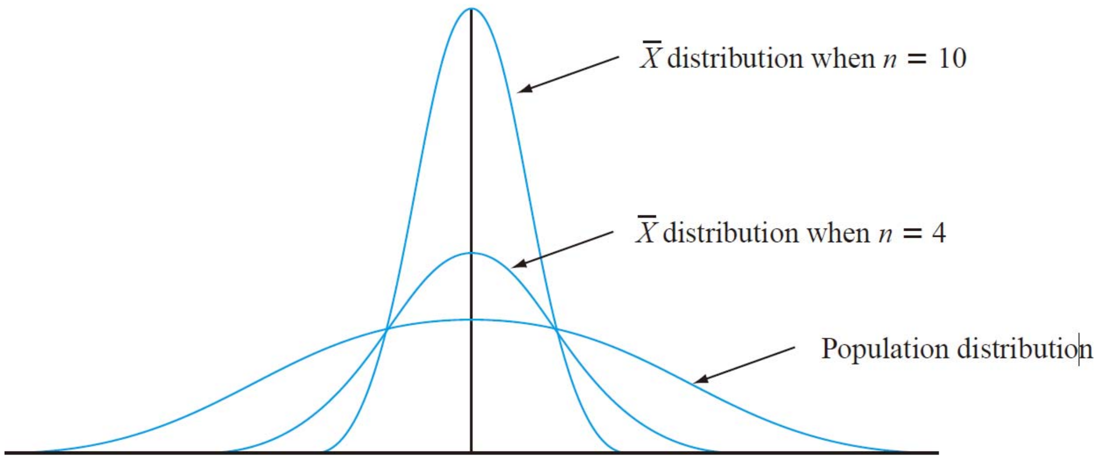
$$\sigma_{\overline{X}} = \sigma / \sqrt{n} = \frac{5000}{\sqrt{25}} = 1000$$

$$\sigma_{T_0} = \sqrt{n}\sigma = \sqrt{25}(5000) = 25,000$$

5.4 The Distribution of the Sample Mean

- Proposition

  Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and standard deviation σ. Then for **any *n***, $\bar{X}$ is normally distributed (with mean μ and standard deviation $\sigma/\sqrt{n}$), as is T_o (with mean $n\mu$ and standard deviation $\sqrt{n}\sigma$).



5.4 The Distribution of the Sample Mean

- Example 5.25

The time that it takes a randomly selected rat of a certain subspecies to find its way through a maze is a normally distributed rv with $\mu = 1.5$ min and $\sigma = .35$ min. Suppose five rats are selected. Let $X_1, X_2, \ldots, X_5$ denote their times in the maze. Assuming the X_i's to be a random sample from this normal distribution.

- **Q #1:** What is the probability that the total time $T_o = X_1+X_2+\ldots+X_5$ for the five is between 6 and 8 min?
- **Q #2:** Determine the probability that the sample average time $\bar{X}$ is at most 2.0 min.

5.4 The Distribution of the Sample Mean

- Example 5.25 (Cont')

A #1: T_o has a normal distribution with $\mu_{To} = n\mu = 5(1.5) = 7.5$ min and variance $\sigma_{To}^2 = n\sigma^2 = 5(0.1225) = 0.6125$, so $\sigma_{To} = 0.783$ min. To standardize T_o, subtract μ_{To} and divide by σ_{To}:

$$P(6 \le T_o \le 8) = P(\frac{6-7.5}{0.783} \le Z \le \frac{8-7.5}{0.783})$$

$$= P(-1.92 \le Z \le 0.64) = \Phi(0.64) - \Phi(-1.92) = 0.7115$$

A #2:

$$E(\bar{X}) = \mu = 1.5 \qquad \sigma_{\bar{X}} = \sigma / \sqrt{n} = 0.35 / \sqrt{5} = 0.1565$$

$$P(\bar{X} \le 2.0) = P(Z \le \frac{2.0-1.5}{0.1565})$$

$$= P(Z \le 3.19) = \Phi(3.19) = 0.9993$$

5.4 The Distribution of the Sample Mean

- The Central Limit Theorem (CLT)

  Let X_1, X_2, …, X_n be a **random sample** from a distribution (may or may not be normal) with mean μ and variance σ^2.

  Then if n is sufficiently large, $\bar{X}$ has approximately a normal distribution with

$$\mu_{\bar{X}} = \mu, \sigma_{\bar{X}}^2 = \sigma^2 / n$$

  T_o also has approximately a normal distribution with

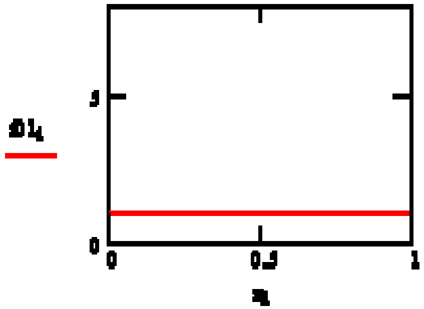
$$\mu_{T_0} = n\mu, \sigma_{T_0}{}^2 = n\sigma^2$$

The larger the value of n, the better the approximation

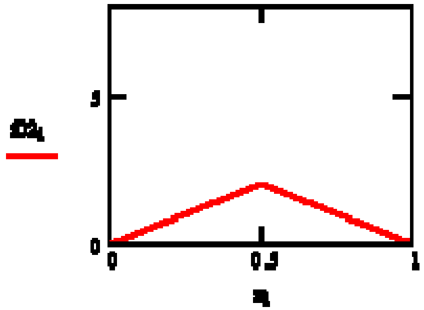
Usually, If *n* > 30, the Central Limit Theorem can be used.

5.4 The Distribution of the Sample Mean

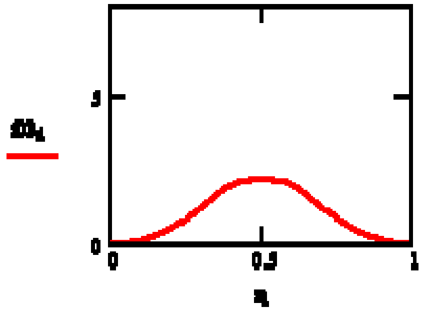
- An Example for Uniform Distribution



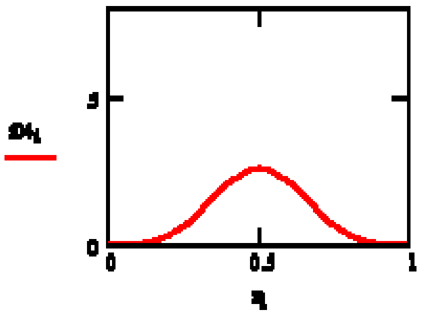
NonNormal Distribution of X



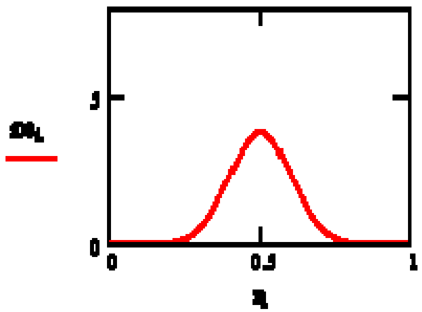
Distribution of Xbar when sample size is 2



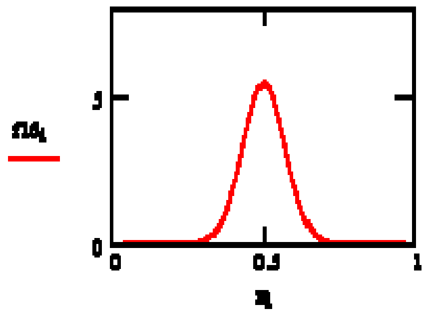
Distribution of Xbar when sample size is 3



Distribution of Xbar when sample size is 4



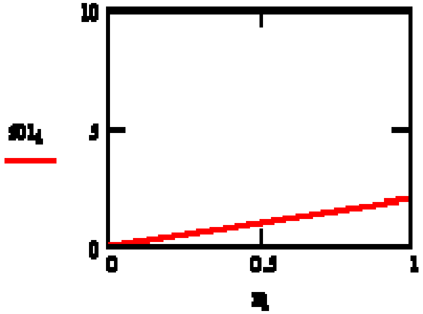
Distribution of Xbar when sample size is 8



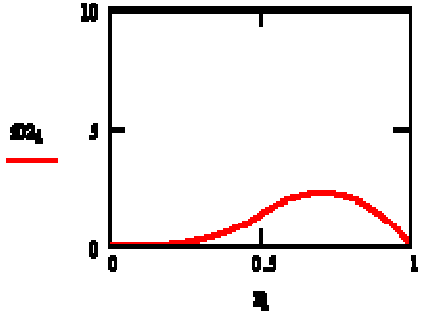
Distribution of Xbar when sample size is 16

5.4 The Distribution of the Sample Mean

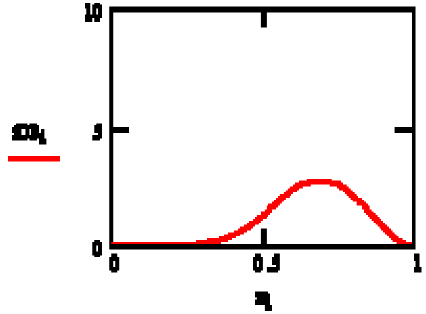
- An Example for Triangular Distribution



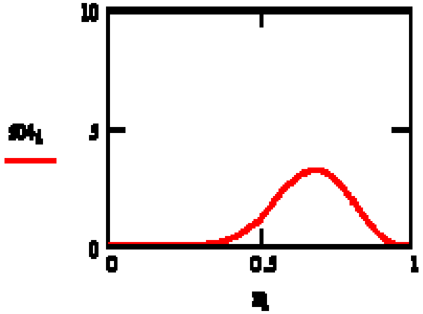
NonNormal Distribution of X



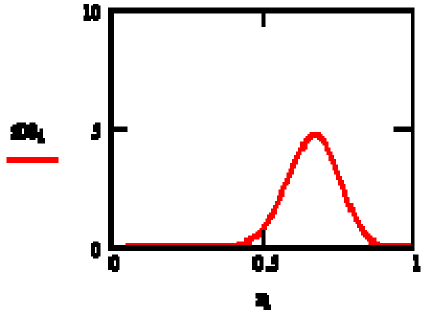
Distribution of Xbar when sample size is 2



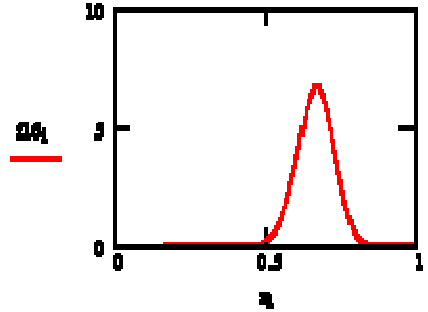
Distribution of Xbar when sample size is 3



Distribution of Xbar when sample size is 4



Distribution of Xbar when sample size is 8



Distribution of Xbar when sample size is 16

- Example 5.26

When a batch of a certain chemical product is prepared, the amount of a particular impurity in the batch is a random variable with mean value 4.0g and standard deviation 1.5g. If 50 batches are independently prepared, what is the (approximate) probability that the sample average amount of impurity $\bar{X}$ is between 3.5 and 3.8g?

Here $n = 50$ is large enough for the CLT to be applicable. $\bar{X}$ then has approximately a normal distribution with mean value $\mu_{\bar{X}} = 4.0$ and $\sigma_{\bar{X}} = 1.5/\sqrt{50} = 0.2121$, so

$$P(3.5 \le \bar{X} \le 3.8) \approx P\left(\frac{3.5-4.0}{0.2121} \le Z \le \frac{3.8-4.0}{0.2121}\right) = \Phi(-0.94) - \Phi(-2.36) = 0.1645$$

5.4 The Distribution of the Sample Mean

- Example 5.27

A certain consumer organization customarily reports the number of major defects for each new automobile that it tests. Suppose the number of such defects for a certain model is a random variable with mean value 3.2 and standard deviation 2.4. Among 100 randomly selected cars of this model, how likely is it that the sample average number of major defects exceeds 4?

Let X_i denote the number of major defects for the i^{th} car in the random sample. Notice that X_i is a discrete rv, but the CLT is applicable whether the variable of interest is discrete or continuous.

5.4 The Distribution of the Sample Mean

- Example 5.27 (Cont')

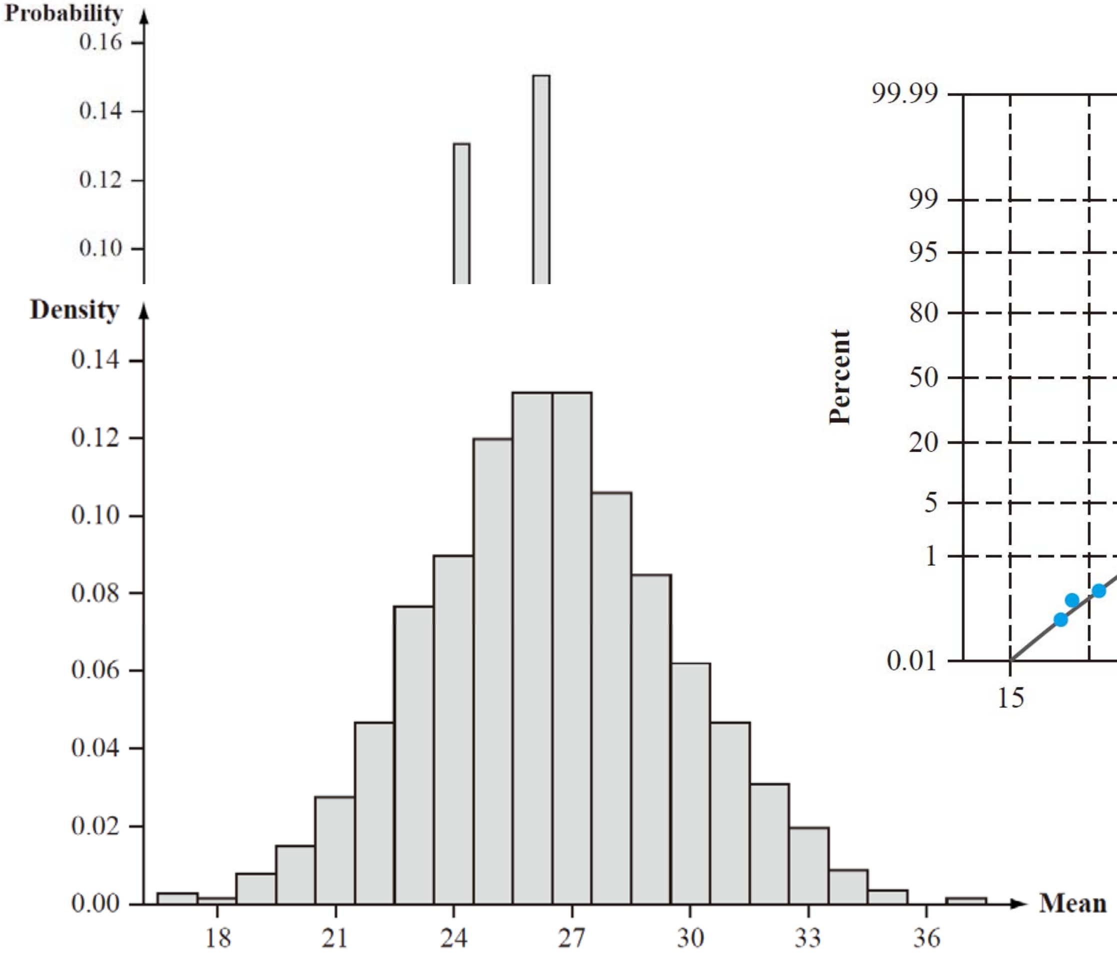
Using $\mu_{\overline{X}} = 3.2$ and $\sigma_{\overline{X}} = 0.24$

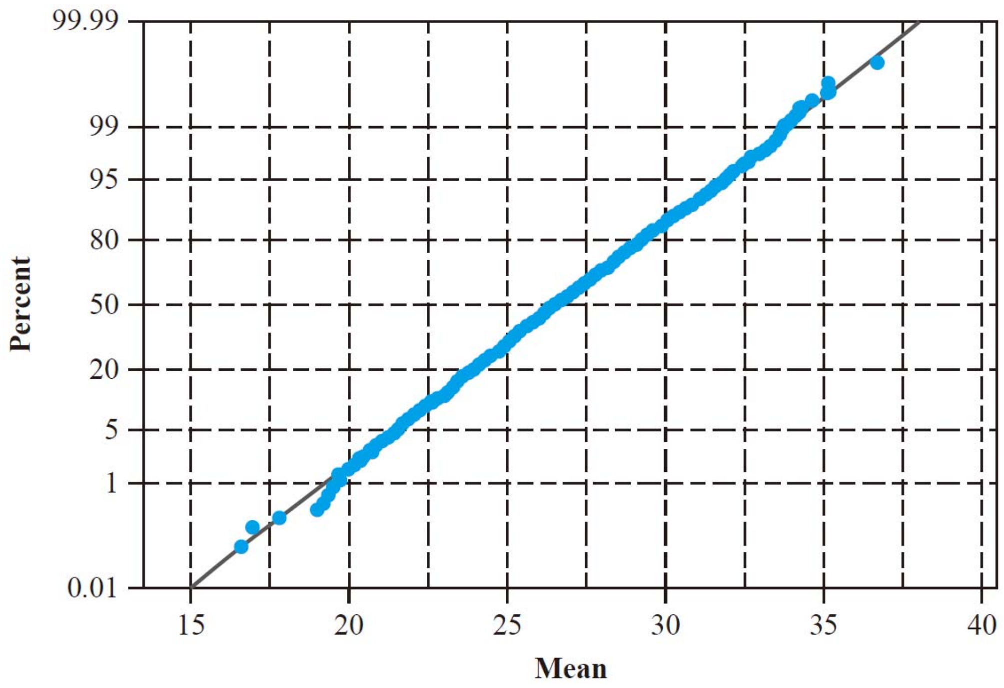
$$P(\overline{X} > 4) \approx P\left(Z > \frac{4-3.2}{0.24} \right)$$

$$= 1 - \Phi(3.33) = 0.0004$$

5.4 The Distribution of the Sample Mean

- Example 5.28





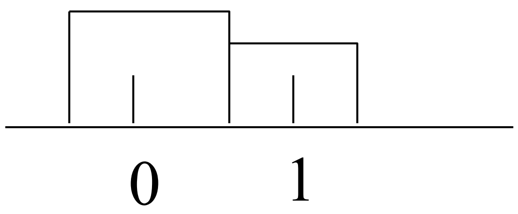
- Other Applications of the CLT

The CLT can be used to justify the normal approximation to the binomial distribution discussed in Chapter 4. Recall that a binomial variable X is the number of successes in a binomial experiment consisting of n independent success/failure trials with $p = P(S)$ for any particular trial. Define new rv's $X_1, X_2, \ldots, X_n$ by

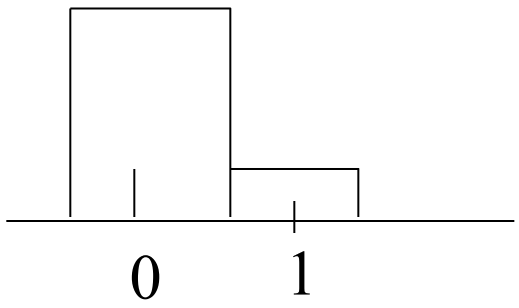
$$X_i = \begin{cases} 1 & \text{if the } i\text{th trial results in a success} \\ 0 & \text{if the } i\text{th trial results in a failure} \end{cases} \qquad (i = 1, \ldots, n)$$

5.4 The Distribution of the Sample Mean

- Because the trials are independent and $P(S)$ is constant from trial to trial to trial, the X_i's are *i.i.d* (a random sample from a Bernoulli distribution).
- The CLT then implies that if n is sufficiently large, both the sum and the average of the X_i's have approximately normal distributions. Now the binomial rv $X = X_1+\ldots+X_n$. X/n is the sample mean of the X_i's. That is, both X and X/n are approximately normal when n is large.
- The necessary sample size for this approximately depends on the value of p: When p is close to .5, the distribution of X_i is reasonably symmetric. The distribution is quit skewed when p is near 0 or 1.



p=0.4 (a)



p=0.1 (b)

Rule:
$np \geq 10$ & $n(1-p) \geq 10$
rather than
$n>30$

5.4 The Distribution of the Sample Mean

- Proposition

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution for which only positive values are possible [$P(X_i > 0) = 1$]. Then if n is sufficiently large, the product $Y = X_1 X_2 \cdot \ldots \cdot X_n$ has approximately a lognormal distribution.

Please note that :

$$\ln(Y) = \ln(X_1) + \ln(X_2) + \ldots + \ln(X_n)$$

Supplement: Law of large numbers

- **Chebyshev's Inequality**

Let X be a random variable (continuous or discrete) , then

$$P(|X-E(X)|\geq \varepsilon)\leq \frac{D(X)}{\varepsilon^2}, \forall \varepsilon > 0$$

Proof:

$$P(|X-E(X)|\geq \varepsilon) = P\left(\frac{|X-E(X)|}{\varepsilon}\geq 1\right) = P\left(\frac{(X-E(X))^2}{\varepsilon^2}\geq 1\right)$$

$$= \int_{\frac{(X-E(X))^2}{\varepsilon^2}\geq 1} p(x)dx \qquad B=\{X\leq E(X)-\varepsilon\}\cup\{X\geq E(X)+\varepsilon\}$$

$$\leq \int_B \frac{(X-E(X))^2}{\varepsilon^2} p(x)dx \qquad \frac{(X-E(X))^2}{\varepsilon^2}\geq 1$$

$$\leq \int_{-\infty}^{+\infty} \frac{(X-E(X))^2}{\varepsilon^2} p(x)dx = \frac{D(X)}{\varepsilon^2}$$

Supplement: Law of large numbers

- Khintchine law of large numbers

X_1, X_2, ... an infinite sequence of *i.i.d.* random variables with finite expected value $E(X_k) = \mu < \infty$ and variable $D(X_k) = \delta^2 < \infty$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right) = 1, \forall\varepsilon > 0$$

Proof:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad E(\overline{X}_n) = \mu;\ D(\overline{X}_n) = \frac{\delta^2}{n}$$

According to Chebyshev's inequality

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) = P\left(\left|\overline{X}_n - E(\overline{X}_n)\right| \geq \varepsilon\right) \leq \frac{D(\overline{X}_n)}{\varepsilon^2} = \frac{1}{n}\frac{\delta^2}{\varepsilon^2} \xrightarrow{n\to\infty} 0$$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right) = 1 - \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) = 1$$

Supplement: Law of large numbers

- Bernoulli law of large numbers

The empirical probability of success in a series of Bernoulli trials Ai will converge to the theoretical probability.

$$A_i = \begin{cases} 1, & A_i occurs \\ 0, & others \end{cases}$$

Ai	1	0
p	p	1-p

Let n(A) be the number of replication on which A does occur, then we have

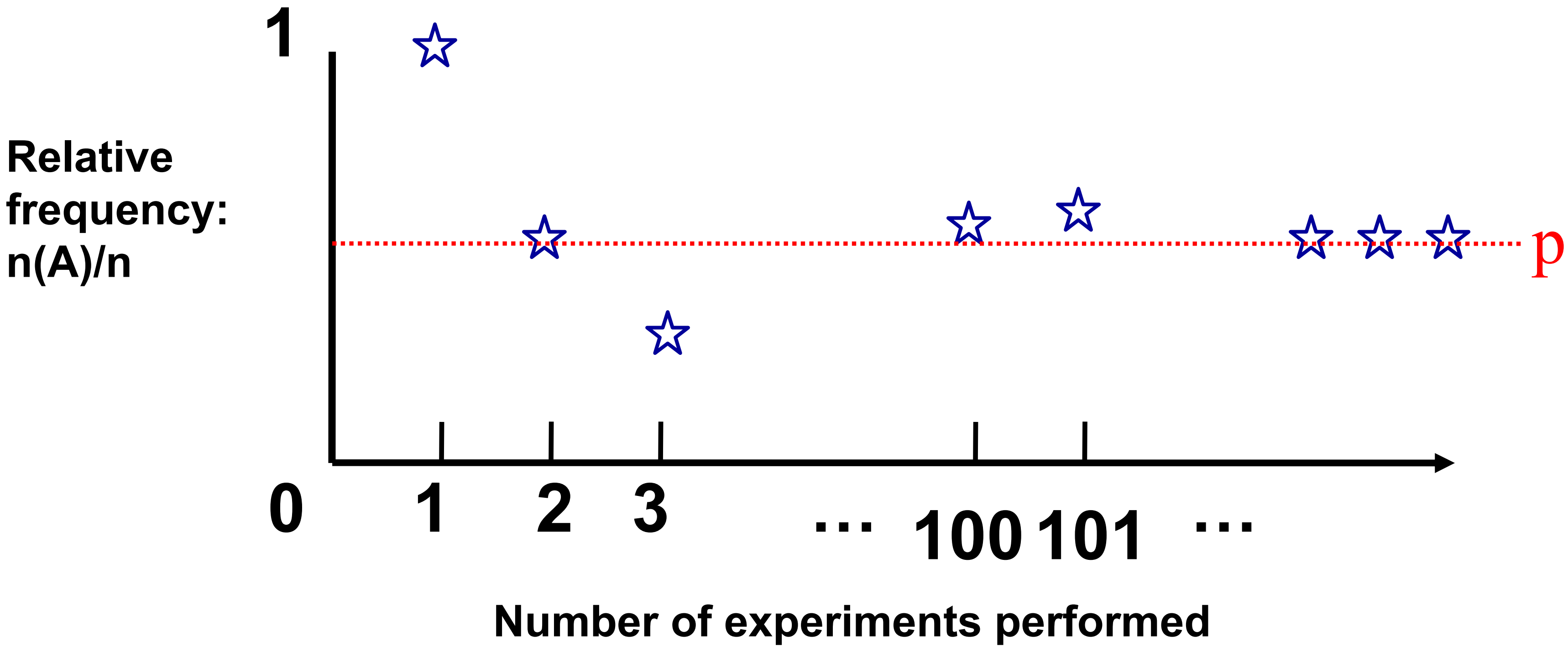
$$\frac{n(A)}{n} = \frac{1}{n}\sum_{i=1}^{n} A_i \qquad E\left(\frac{n(A)}{n}\right) = p \qquad D\left(\frac{n(A)}{n}\right) = \frac{\delta^2}{n} = \frac{p(1-p)}{n}$$

According to Chebyshev's inequality

$$P\left(\left|\frac{n(A)}{n} - p\right) \right| \geq \varepsilon) \leq \frac{1}{n}\frac{p(1-p)}{\varepsilon^2} \xrightarrow{n \to \infty} 0$$

$$\lim_{n\to\infty} P\left(\left|\frac{n(A)}{n} - p\right| < \varepsilon\right) = 1 - \lim_{n\to\infty} P\left(\left|\frac{n(A)}{n} - p\right| \geq \varepsilon\right) = 1$$

Supplement: Law of large numbers



5.4 The Distribution of the Sample Mean

- Homework

  Ex. 46, Ex. 48, Ex. 50, Ex. 56

5.5 The Distribution of a Linear Combination

- Linear Combination

Given a collection of n random variables $X_1, \ldots, X_n$ and n numerical constants $a_1, \ldots, a_n$, the rv

$$Y = a_1 X_1 + \cdots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

is called a linear combination of the X_i's.

5.5 The Distribution of a Linear Combination

Let X_1, X_2, …, X_n have mean values μ_1, …, μ_n respectively, and variances of σ_1^2, …., σ_n^2, respectively.

- Whether or not the X_i's are independent,

$$E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_i$$

- If X_1, X_2, …, X_n are independent,

$$V(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 V(X_i) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

$$\sigma_{a_1 X_1 + \cdots + a_n X_n} = \sqrt{a_1^2 \sigma_1^2 + \cdots a_n^2 \sigma_n^2}$$

- For any X_1, X_2, …, X_n,

$$V(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j)$$

5.5 The Distribution of a Linear Combination

- Proof: $E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_i$

For the result concerning expected values, suppose that X_i's are continuous with joint pdf $f(x_1,\ldots,x_n)$. Then

$$E(\sum_{i=1}^{n} a_i X_i) = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}(\sum_{i=1}^{n} a_i x_i) f(x_1,\ldots,x_n)dx_1\ldots dx_n$$

$$= \sum_{i=1}^{n} a_i \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_i f(x_1,\ldots,x_n)dx_1\ldots dx_n$$

$$= \sum_{i=1}^{n} a_i \int_{-\infty}^{\infty} x_i f_{X_i}(x_i)dx_i$$

$$= \sum_{i=1}^{n} a_i E(X_i)$$

5.5 The Distribution of a Linear Combination

- Proof: $V(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j)$

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = E\left[\left(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i \mu_i\right)^2\right]$$

$$= E\left\{\left[\sum_{i=1}^{n} a_i \left(X_i - \mu_i\right)\right]^2\right\} \quad = E\left\{\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j (X_i - \mu_i)(X_j - \mu_j)\right\}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j E[(X_i - \mu_i)(X_j - \mu_j)]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j)$$

When the X_i’s are independent, $\mathrm{Cov}(X_i, X_j) = 0$ for $i \neq j$, and

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j) \quad = \sum_{i=1}^{n} a_i^2 V(X_i)$$

5.5 The Distribution of a Linear Combination

- Example 5.29

A gas station sells three grades of gasoline: regular, extra, and super. These are priced at \$3.00, \$3.20, and \$3.40 per gallon, respectively. Let X_1, X_2, and X_3 denote the amounts of these grades purchased (gallons) on a particular day.

Suppose the X_i's are independent with $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$, and $\sigma_3 = 50$. The revenue from sales is $Y = 3.0X_1 + 3.2X_2 + 3.4X_3$, and

$$E(Y) = 3.0\mu_1 + 3.2\mu_2 + 3.4\mu_3 = \$5620$$

$$V(Y) = (3.0)^2\sigma_1^2 + (3.2)^2\sigma_2^2 + (3.4)^2\sigma_3^2 = 184{,}436$$

$$\sigma_Y = \sqrt{184{,}436} = \$429.46$$

5.5 The Distribution of a Linear Combination

- Corollary (the different between two rv's)

  $E(X_1\text{-}X_2) = E(X_1)$ - $E(X_2)$ and, if X_1 and X_2 are *independent*, $V(X_1\text{-}X_2) = V(X_1)+V(X_2)$.

- Example 5.30

  A certain automobile manufacturer equips a particular model with either a six-cylinder engine or a four-cylinder engine. Let X_1 and X_2 be fuel efficiencies for independently and randomly selected six-cylinder and four-cylinder cars, respectively. With $\mu_1 = 22$, $\mu_2 = 26$, $\sigma_1 = 1.2$, and $\sigma_2 = 1.5$,

$$E(X_1 - X_2) = \mu_1 - \mu_2 = 22 - 26 = -4$$

$$V(X_1 - X_2) = \sigma^2{}_1 + \sigma^2{}_2 = (1.2)^2 + (1.5)^2 = 3.69$$

$$\sigma_{X_1 - X_2} = \sqrt{3.69} = 1.92$$

5.5 The Distribution of a Linear Combination

- Proposition

  If $X_1, X_2, \ldots, X_n$ are **independent, normally distributed** rv's (with possibly different means and/or variances), then **any linear** combination of the X_i's also has a normal distribution.

- Example 5.31 (Ex. 5.29 Cont')

- The total revenue from the sale of the three grades of gasoline on a particular day was $Y = 3X_1+3.2X_2+3.4X_3$, and we calculated $\mu_Y = 5620$ and $\sigma_Y = 429.46$). If the X_i's are normally distributed, the probability that the revenue exceeds 4500 is

$$P(Y \geq 4500) = P(Z > \frac{4500-5620}{429.46})$$

$$= P(Z > -2.61) = 1 - \Phi(-2.61) = 0.9955$$

5.5 The Distribution of a Linear Combination

- Homework

  Ex. 58, Ex. 68, Ex. 70, Ex. 72