Chapter 4
Expectation

Instructor : Dr. Jing Liang
School of Information and Communication Engineering
liangjing@uestc.edu.cn

Outlines

- The Expectation of a $R.V.$
- The Expectation of typical distributions
- Properties of Expectations
- The mean and the Median
- Variance
- The Variance of typical distributions
- Moments
- Covariance and Correlation
- Conditional Expectation

Expectation of a Discrete Distribution-1

- **Ex1 (Book Ex4.1.2) Stock Price Change**. Suppose that the change in price of a stock is a *R.V.* X that can assume only the four different values −2, 0, 1, and 4, and that $\Pr(X=-2) = 0.1$, $\Pr(X = 0) = 0.4$, $\Pr(X = 1) = 0.3$, and $\Pr(X = 4) = 0.2$. Then the weighted avarage of these values is

$$-2(0.1) + 0(0.4) + 1(0.3) + 4(0.2) = 0.9.$$

0.9 is the ***average value***, or ***expected value***, or the ***mean***, or the ***expectation*** of X.

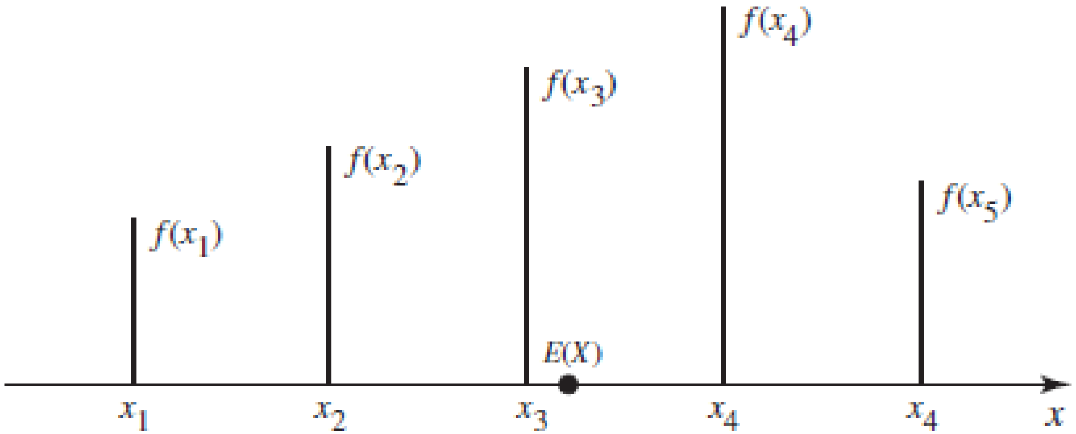
- **Definition 4.1.1 Mean of Bounded Discrete *R.V.*** Let X be a bounded discrete *R.V.* whose p.f. is f. The ***expectation of X***, denoted by $E(X)$, is a **number** defined:

$$E(X) = \sum_{\text{All } x} xf(x).$$

Q: Suppose X is a Bernoulli *R.V.*, what's $E(X)$?

Interpretation of the Expectation

- The mean of a distribution can be regarded as being the center of gravity of that distribution.



- x-axis may be regarded as a long weightless rod;
- $f(x_i)$ is the weight attached to this rod at each point x_i;
- The rod will be balanced if it is supported at the point $E(X)$.

Every two *R.V.*s that have the same distribution will have the same mean. However, not vice versa.

Expectation of a Discrete Distribution-2

- **Definition 4.1.2 Mean of General Discrete *R.V.*** Let X be a discrete $R.V.$ whose p.f. is f. Suppose that at least one of the following sums is finite:

$$\sum_{\text{Positive } x} xf(x), \quad \sum_{\text{Negative } x} xf(x).$$

Then the ***mean***, ***expectation***, or ***expected value*** of X is said to **exist** and is defined to be

$$E(X) = \sum_{\text{All } x} xf(x).$$

If both of the sums in are infinite, then $E(X)$ does not exist.
Why? Fails to converge or converge to many different values in different orders of the terms.

Expectation of a Discrete Distribution-2

◆ Ex2 (Book Ex4.1.4) The Mean of *X* Does Not Exist.

Let X be a $R.V.$ whose p.f. is

$$f(x)=\begin{cases}\dfrac{1}{2|x|(|x|+1)} & \text{if } x=\pm1,\pm2,\pm3,\ldots, \\ 0 & \text{otherwise.}\end{cases}$$

It can be verified that this function satisfies the conditions required to be a p.f. The two sums are

$$\sum_{x=-1}^{-\infty} x\frac{1}{2|x|(|x|+1)}=-\infty \text{ and } \sum_{x=1}^{\infty} x\frac{1}{2x(x+1)}=\infty;$$

hence, $E(X)$ does not exist.

Expectation of a Discrete Distribution-3

- **Ex3 (Book Ex4.1.5) An Infinite Mean.**

Let X be a $R.V.$ whose p.f. is

$$f(x)=\begin{cases}\dfrac{1}{x(x+1)} & \text{if } x=1,2,3,\dots,\\ 0 & \text{otherwise.}\end{cases}$$

The sum over negative values $\sum_{\text{Negative } x} xf(x)$ is 0,

so the mean of X exists and is

$$E(X)=\sum_{x=1}^{\infty} x\frac{1}{x(x+1)}=\infty.$$

We say that the mean of X is infinite in this case.

Mean of a Continuous Distribution-1

- **Definition 4.1.3 Mean of Bounded Continuous *R.V.***

Let X be a bounded continuous $R.V.$ whose p.d.f. is f. The expectation of X, denoted $E(X)$, is defined as follows:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

- **Ex4 (Book Ex4.1.6) Expected Failure Time**. An appliance has a maximum lifetime of one year. The time X until it fails is a $R.V.$ with a continuous distribution having p.d.f.

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$E(X) = \int_0^1 x(2x)dx = \left.\frac{2}{3}x^3\right|_0^1 = \frac{2}{3}.$$

Mean of a Continuous Distribution-2

Definition 4.1.4 Mean of General Continuous *R.V*.

Let X be a continuous random variable whose p.d.f. is f. Suppose that at least one of the following integrals is finite

$$\int_0^\infty xf(x)dx, \int_{-\infty}^0 xf(x)dx.$$

Then the ***mean***, ***expectation***, or ***expected value*** of X is said to exist and is defined to be

$$E(X) = \int_{-\infty}^\infty xf(x)dx.$$

If both of the integrals are infinite, then ***E(X) does not exist***.

The Expectation of a Function - 1

- **Functions of a Single *R.V*.** If X is a $R.V.$ for which the p.d.f. is f, then the expectation (if exists) of each real-valued function $Y=r(X)$ can be found by:

$$E[r(X)] = E(Y) = \int_{-\infty}^{\infty} yg(y)dy.$$

- **Ex5 (Book Ex4.1.11)** Suppose the p.d.f. of X is

$$f(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

$Y=1/X$. Find $E(Y)$.

Let $Y=r(X)=1/X$. Then $g(y) = \begin{cases} 3y^{-4} & \text{if } y > 1, \\ 0 & \text{otherwise.} \end{cases}$

The mean of Y is $E(Y) = \int_{-\infty}^{\infty} y3y^{-4}dy = \dfrac{3}{2}.$

The Expectation of a Function - 2

◆ **Theorem 4.1.1 Law of the Unconscious Statistician.**
Let X be a $R.V.$, and let r be a real valued function of a real variable. If X has a continuous distribution, then

$$E[r(X)] = \int_{-\infty}^{\infty} r(x) f(x) dx,$$

if the mean exists. If X has a discrete distribution, then

$$E[r(X)] = \sum_{\text{All } x} r(x) f(x),$$

if the mean exists.

Ex6 (Book Ex4.1.12) Failure Rate and Time to Failure.

$$E(Y) = \int_0^1 \frac{1}{x} 3x^2 dx = \frac{3}{2}.$$

The Expectation of a Function - 3

- **Note: In General, $E[g(x)]\neq g[E(x)]$.**
- e.g., In Ex4, we have that for the $R.V.$, X,

$$f(x)=\begin{cases}2x & \text{for } 0<x<1,\\ 0 & \text{otherwise.}\end{cases}$$

$$E(X)=\int_0^1 x(2x)dx=\left.\frac{2}{3}x^3\right|_0^1=\frac{2}{3}.$$

Suppose $g(X)=\sqrt{X}$.

$$E[g(X)]=\int_0^1 \sqrt{x}(2x)dx=2\int_0^1 x^{\frac{3}{2}}dx=\frac{4}{5}.$$

$$g[E(X)]=\sqrt{\frac{2}{3}}. \qquad E[g(X)]\neq g[E(X)].$$

A linear function g does satisfy $E[g(x)]=g[E(x)]$.

Functions of Several *R.V.*s - 1

- **Theorem 4.1.2 Law of the Unconscious Statistician.** Suppose that $X_1, \ldots, X_n$ are *R.V.*s with the joint p.d.f. $f(x_1, \ldots, x_n)$. Let r be a real-valued function of n real variables, and suppose that $Y = r(X_1, \ldots, X_n)$. Then $E(Y)$ can be determined directly from the relation

$$E(Y) = \int_{R^n} \cdots \int r(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \cdots dx_n,$$

if the mean exists. Similarly, if $X_1, \ldots, X_n$ have a discrete joint distribution with p.f. $f(x_1, \ldots, x_n)$, the mean of $Y = r(X_1, \ldots, X_n)$ is

$$E(Y) = \sum_{\text{All } x_1, \ldots, x_n} r(x_1, \ldots, x_n) f(x_1, \ldots, x_n),$$

if the mean exists.

Functions of Several *R.V.*s - 2

◆ **Ex6 (Book Ex4.1.16) Determining the Expectation of a Function of Two Variables**. Suppose that a point (X,Y) is chosen at random from the square S containing all points (x,y), such that $0\le x\le 1$ and $0\le y\le 1$. Determine the expected value of X^2+Y^2.

Sol: the joint p.d.f. of X and Y is

$$f(x,y)=\begin{cases}1 & \text{for } (x,y)\in S,\\ 0 & \text{otherwise.}\end{cases}$$

$$E(X^2+Y^2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x^2+y^2)f(x,y)dxdy$$

$$=\int_0^1\int_0^1(x^2+y^2)f(x,y)dxdy=\frac{2}{3}$$

The expectation of Typical *R.V.*s-1

- **Definition 5.2.1 Bernoulli Distribution**

A *R.V.* X has the ***Bernoulli distribution with parameter p*** ($0 \le p \le 1$) if X can take **only the values 0 and 1** and the probabilities are

$\Pr(X=1)=p$ and $\Pr(X=0)=1-p$.

The p.f. of X is

$$\boxed{f(x \mid p) = \begin{cases} p^x (1-p)^{1-x} & \text{for } x = 0,1, \\ 0 & \text{otherwise.} \end{cases}}$$

The expectation of X is

$$E(X) = 0 \times (1-p) + 1 \times p = p.$$

$$E(X^2) = 0^2 \times (1-p) + 1^2 \times p = p.$$

The expectation of Typical *R.V.*s-2

- **Definition 5.2.3 Binomial Distribution**

A *R.V. X* has the binomial distribution with parameters *n* and *p* if *X* has a discrete distribution with the p.f.

$$f(x \mid n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0,1,2,\ldots,n, \\ 0 & \text{otherwise.} \end{cases}$$

Where $0 \le p \le 1$.

$$E(X) = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = np \sum_{i=1}^{n} \binom{n-1}{i-1} (1-p)^{(n-1)-(i-1)} p^{i-1}$$

$$= np[p + (1-p)]^{n-1} = np.$$

The expectation of Typical *R.V.*s-3

- **Theorem 5.2.1** If the *R.V.*s $X_1, \ldots, X_n$ form n Bernoulli trials with parameter p, and if $X = X_1 + \ldots + X_n$, then X has the binomial distribution with parameters n and p.

- **Two conditions of Theorem 5.2.1**

1) All X_i are mutually independent.

2) All X_i have the same parameter p.

- In this case

$$E(X) = \sum_{i=1}^{n} E(X_i) = np.$$

The expectation of Typical *R.V.*s-4

Definition 5.4.1 Poisson Distribution.

Let λ>0, a *R.V. X* has the ***Poisson distribution with mean λ*** if the p.f. of *X* is as follows

$$\boxed{f(x\mid\lambda)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & x=0,1,2,\cdots,\\ 0 & otherwise.\end{cases}}$$

$$E(X)=\sum_{x=0}^{+\infty}x\frac{e^{-\lambda}\lambda^{x}}{x!}=\lambda e^{-\lambda}\sum_{x=1}^{+\infty}\frac{\lambda^{x-1}}{(x-1)!}$$

$$=\lambda e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^{k}}{k!}\qquad (k=x-1)$$

$$=\lambda e^{-\lambda}\cdot e^{\lambda}=\lambda$$

The expectation of Typical *R.V.*s-5

- **Theorem 3.2.1 Uniform Distribution p.d.f.** If X has the uniform distribution on an interval $[a,b]$, then the p.d.f of X is

$$\boxed{f(x\mid a,b)=\begin{cases}\dfrac{1}{b-a} & \text{if } a\le x\le b,\\ 0 & \text{otherwise.}\end{cases}}$$

$$E(X)=\int_{-\infty}^{\infty}xf(x)dx=\frac{1}{b-a}\int_a^b xdx=\frac{a+b}{2}$$

The expectation of Typical *R.V.*s-6

◆ **Definition 5.6.1** A *R.V. X* has the ***normal distribution*** with mean μ and variance σ^2 ($-\infty<\mu<\infty$ and $\sigma>0$) if X has a continuous distribution with the following p.d.f.:

$$\boxed{f(x\mid\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\quad \text{for } -\infty<x<\infty}$$

$$E(X)=\int_{-\infty}^{+\infty}x\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}xe^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

$$\overset{t=\frac{x-\mu}{\sigma}}{=}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}(\mu+\sigma t)e^{-\frac{t^2}{2}}dt=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\sigma te^{-\frac{t^2}{2}}dt+\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\mu e^{-\frac{t^2}{2}}dt$$

$$=0+\mu\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\mu.$$

The expectation of Typical *R.V.*s-7

- **Definition 5.7.3** A *R.V. X* has ***the exponential distribution with parameter β*** (*β >0*) if *X* has a continuous distribution with the following p.d.f.:

$$f(x)=\begin{cases}\beta e^{-\beta x} & \text{for } x>0,\\ 0 & \text{for } x\le 0.\end{cases}$$

$$E(X)=\int_0^{\infty} x\beta e^{-\beta x}dx \overset{\text{set } u=x\beta}{=} \frac{1}{\beta}\int_0^{+\infty} ue^{-u}du$$

$$=\frac{1}{\beta}e^{-x}(-x-1)\Big|_0^{+\infty}=\frac{1}{\beta}.$$

Properties of Expectations - 1

- **Theorem 4.2.1 Linear Function**. If $Y = aX + b$, where a and b are finite constants, then $E(Y) = aE(X) + b$.

Proof: first assume, for convenience, that X has a continuous distribution for which the p.d.f. is f. Then

$$E(Y) = E(aX+b) = \int_{-\infty}^{\infty}(ax+b)f(x)dx$$

$$= a\int_{-\infty}^{\infty} xf(x)dx + b\int_{-\infty}^{\infty} f(x)dx$$

$$= aE(x) + b.$$

A similar proof can be given for a discrete distribution.

- **Corollary 4.2.1** If X=c with probability 1, then $E(X)$=c.

Properties of Expectations - 2

- **Theorem 4.2.2** If there exists a constant such that $\Pr(X \ge a)=1$, then $E(X) \ge a$. If there exists a constant b such that $\Pr(X \le b)=1$, then $E(X) \le b$.

Proof. For convenience, assume X is continuous and p.d.f. f, and $\Pr(X \ge a)=1$.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{a}^{\infty} xf(x)dx$$

$$\ge \int_{a}^{\infty} af(x)dx = a\int_{a}^{\infty} f(x)dx$$

$$= a\Pr(X \ge a) = a.$$

The proof of the other part is similar.

Properties of Expectations - 3

- **Theorem 4.2.3** Suppose that $E(X) = a$ and that either $Pr(X \geq a) = 1$ or $Pr(X \leq a) = 1$. Then $Pr(X = a) = 1$.

- **Theorem 4.2.4** If $X_1, \ldots, X_n$ are n $R.V.s$ such that each expectation $E(Xi)$ is finite ($i = 1, \ldots, n$), then $E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n)$.
Here regardless of whether $X_1, \ldots, X_n$ are independent or not, regardless of what the joint distribution is.

- **Corollary 4.2.2** Assume that $E(X_i)$ is finite for $i = 1, \ldots, n$. For all constants $a_1, \ldots, a_n$ and b, $E(a_1X_1 + \ldots + a_nX_n + b) = a_1E(X_1) + \ldots + a_nE(X_n) + b$.

Properties of Expectations - 3

- Theorem 4.2.4 Proof: assume n=2, also X_1 and X_2 have a continuous joint distribution for convenience.

$$E(X_1+X_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1+x_2)f(x_1,x_2)dx_1dx_2$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1f(x_1,x_2)dx_1dx_2+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_2f(x_1,x_2)dx_1dx_2$$

$$=\int_{-\infty}^{\infty}[\int_{-\infty}^{\infty}x_1f(x_1,x_2)dx_2]dx_1+\int_{-\infty}^{\infty}[\int_{-\infty}^{\infty}x_2f(x_1,x_2)dx_1]dx_2$$

$$=\int_{-\infty}^{\infty}x_1[\int_{-\infty}^{\infty}f(x_1,x_2)dx_2]dx_1+\int_{-\infty}^{\infty}x_2[\int_{-\infty}^{\infty}f(x_1,x_2)dx_1]dx_2$$

$$=\int_{-\infty}^{\infty}x_1f(x_1)dx_1+\int_{-\infty}^{\infty}x_2f(x_2)dx_2$$

$$=E(X_1)+E(X_2)$$

Q: X and Y are two standard normal $R.V.$s, what's $E(X\text{-}Y)$?

$E(X\text{-}Y)=E(X)\text{-}E(Y)=0$.

Properties of Expectations - 4

- **Ex7 (Book Ex4.2.4) Sampling without Replacement.**

Suppose that a box contains red balls and blue balls. The proportion of red balls in the box is p ($0 \le p \le 1$). Suppose that n balls are selected from the box at random without replacement, and let X denote the number of red balls that are selected. Determine $E(X)$.

Sol: Define n $R.V.s$ $X_1, X_2, \ldots, X_n$. For $i=1,\ldots,n$, let $X_i=1$ if the ith selected ball is red, and $X_i=0$ if it is blue. Since the n balls are selected without replacement, $X_1, X_2, \ldots, X_n$ are dependent ($\because$ $X=X_1+X_2+\ldots+X_n$, which is the total number of red balls that are selected).

$\Pr(X_i=1)= p$, $\Pr(X_i=0)= 1-p$, $E(X_i)=1\times p+0\times(1-p)=p$.

$$E(X)=E(X_1)+\ldots+E(X_n)=np.$$

Compare Ex4.2.5

Properties of Expectations - 4

- **Ex7-continued** Suppose that a class contains 10 boys and 15 girls, and 8 students are to be selected at random from the class without replacement. Let *X* denote the number of boys that are selected, and let *Y* denote the number of girls that are selected. Find *E*(*X*-*Y*).

$$E(X) = np = 8 \cdot \frac{10}{25} = \frac{16}{5}.$$

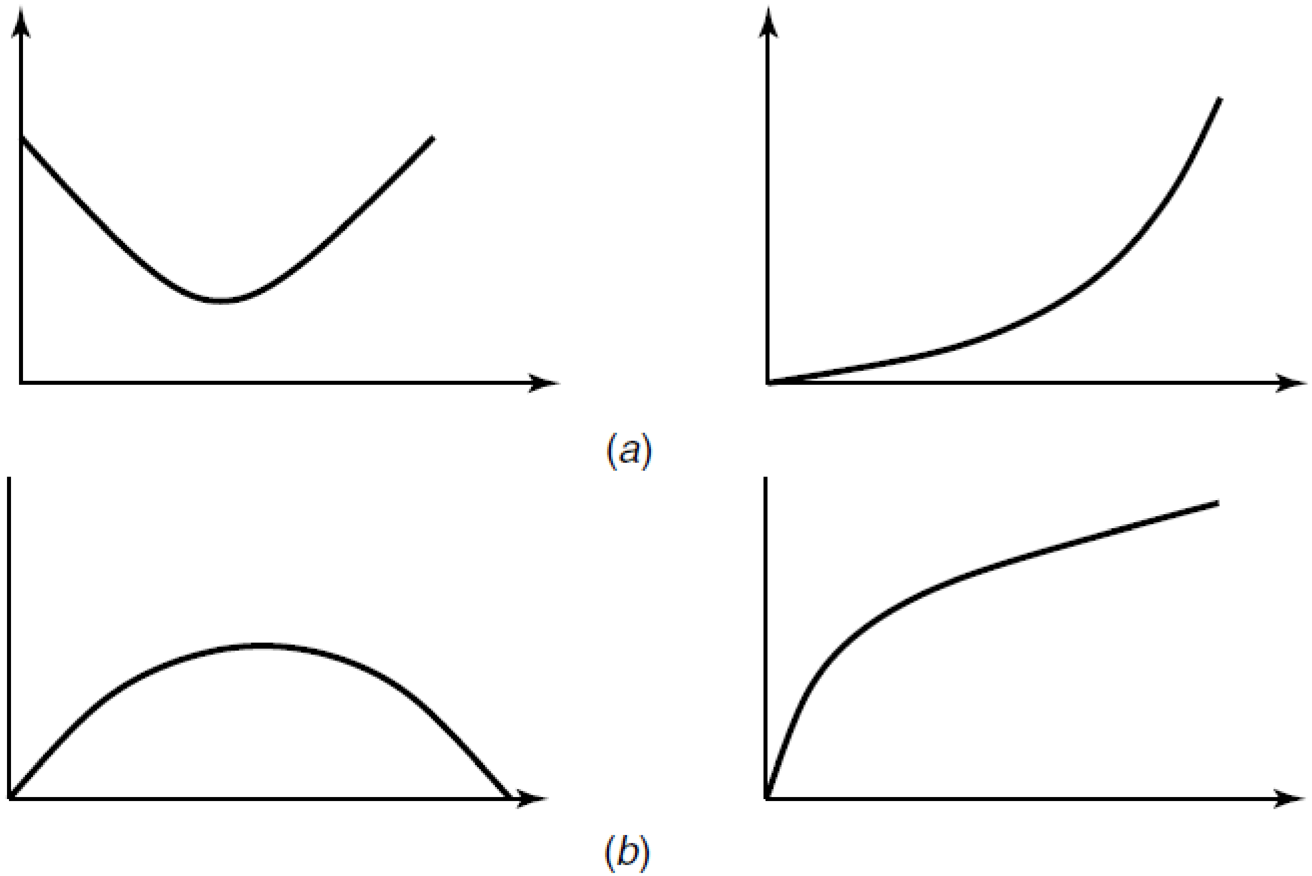
$$Y = 8 - X.$$

$$E(Y) = 8 - E(X) = \frac{24}{5}.$$

$$E(X-Y) = E(X) - E(Y) = -\frac{8}{5}.$$

Convex and Concave

Theorem If the function g has a second derivative that is nonnegative (positive) over an interval, the function is ***convex*** (strictly convex) over that interval.



(a) Convex
(b) Concave

Jensen's Inequality

- **Theorem 4.2.5** Let g be a convex function, and let $\underline{X}$ be a random vector with finite mean. Then

$$E[g(\underline{X})] \geq g[E(\underline{X})]$$

- **Definition 4.2.1** A function g of a vector argument is convex if, for every $\alpha \in (0,1)$, and every $\underline{x}$ and $\underline{y}$,

$$g[\alpha \underline{x} + (1-\alpha)\underline{y}] \leq \alpha g(\underline{x}) + (1-\alpha) g(\underline{y})$$

Textbook made a mistake!

Mean of a Product of Independent *R.V.*s - 1

- **Theorem 4.2.6** If $X_1, \ldots, X_n$ are n independent *R.V.*s such that each expectation $E(X_i)$ is finite ($i = 1, \ldots, n$), then

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i).$$

Proof : since $X_1, \ldots, X_n$ are independent, it follows that every point $(x_1, \ldots, x_n) \in R^n$,

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i).$$

$$E\left(\prod_{i=1}^{n} X_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} x_i\right) f(x_1, \cdots, x_n) dx_1 \cdots dx_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} x_i f(x_i)\right] dx_1 \cdots dx_n$$

$$= \prod_{i=1}^{n} \int_{-\infty}^{\infty} x_i f(x_i) dx_i = \prod_{i=1}^{n} E(X_i).$$

Mean of a Product of Independent *R.V.*s - 2

◆ **Ex8 (Book Ex4.2.7) Calculating the Expectation of a Combination of Random Variables.** Suppose that X_1,X_2, and X_3 are independent *R.V.*s such that $E(X_i) = 0$ and $E(X_i^2) = 1$ for $i = 1, 2, 3$. We shall determine the value of $E[X_1^2(X_2 - 4X_3)^2]$.

Sol: since X_1, X_2, and X_3 are independent, it follows that X_1^2 and $(X_2 - 4X_3)^2$ are also independent. Therefore,

$$
\begin{aligned}
E[X_1^2(X_2-4X_3)^2] &= E(X_1^2)E[(X_2-4X_3)^2] \\
&= E(X_2^2 - 8X_2X_3 + 16X_3^2) \\
&= E(X_2^2) - 8E(X_2X_3) + 16E(X_3^2) \\
&= 1 - 8E(X_2)E(X_3) + 16 \\
&= 17.
\end{aligned}
$$

The Mean and the Median - 1

- **Mean.→Expectation.**
- **Median. Definition 4.5.1** Let X be a $R.V.$ Every number m with the following property is called a median of the distribution of X:

$$\Pr(X \le m) \ge 1/2 \text{ and } \Pr(X \ge m) \ge 1/2.$$

A median is a point m that satisfies the following two requirements: 1) if m is included with the values of X to the left of m, then

$$\Pr(X \le m) \ge \Pr(X > m).$$

2) if m is included with the values of X to the right of m,

$$\Pr(X \ge m) \ge \Pr(X < m).$$

This means that the number m actually divide the total probability into two equal parts.

The Mean and the Median - 2

- **Ex9 (Book Ex4.5.1) The Median of a Discrete Distribution.** Suppose that X has the following discrete distribution:

$$\Pr(X = 1) = 0.1,\ \Pr(X = 2) = 0.2,$$
$$\Pr(X = 3) = 0.3,\ \Pr(X = 4) = 0.4.$$

Q: what's the median? What's the mean?

The value 3 is a median of this distribution because $\Pr(X \leq 3) = 0.6$, which is greater than 1/2, and $\Pr(X \geq 3) = 0.7$, which is also greater than 1/2.

Furthermore, 3 is the unique median of this distribution.

The mean is $E(X)$=1(0.1)+2(0.2)+3(0.3)+4(0.4)=3

The Mean and the Median - 3

- **Ex10 (Book Ex4.5.3) The Median of a Discrete Distribution for Which the Median is Not Unique.**

Suppose that X has the following discrete distribution:

$$\Pr(X = 1) = 0.1,\ \Pr(X = 2) = 0.4,$$
$$\Pr(X = 3) = 0.3,\ \Pr(X = 4) = 0.2.$$

Q: what's the median?

Here, $\Pr(X \le 2) = 1/2$, and $\Pr(X \ge 3) = 1/2$. Therefore, every value of m in the closed interval $2 \le m \le 3$ will be a median of this distribution.

The most popular choice of median of this distribution would be the midpoint 2.5

The Mean and the Median - 4

- **Ex11 (Book Ex4.5.4) The Median of a Continuous Distribution.** Suppose that X has a continuous distribution for which the p.d.f. is as follows:

$$f(x) = \begin{cases} 4x^3 & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

The unique median of this distribution will be the number m such that

$$\int_0^m 4x^3 dx = \int_m^1 4x^3 dx = \frac{1}{2}.$$

This number is $m = (1/2)^{1/4}$.

The Mean and the Median - 5

The median has one convenient property that the mean does not have.

- **Theorem 4.5.1 One-to-One Function.**

Let X be $R.V.$ that takes values in an interval I of real numbers. Let r be a one-to-one function defined on the interval I. If m is a median of X, then $r(m)$ is a median of $r(X)$.

Proof: Let $Y=r(X)$. Since r is one-to-one on the interval I, it must be either increasing or decreasing over I. (Wrong)

If r is increasing, then $Y \geq r(m)$ if and only if $X \geq m$.

Thus, $\Pr(Y \geq r(m)) = \Pr(X \geq m) \geq 1/2$.

Similarly, $Y \leq r(m)$ if and only if $X \leq m$.

So $\Pr(Y \leq r(m)) = \Pr(X \leq m) \geq 1/2$.

If r is decreasing, the proof is similar to the above.

Variance - 1

◆ **Definition 4.3.1 Variance/ Standard Deviation**. Let X be a $R.V.$ with finite mean $\mu=E(X)$. The variance of X, denoted by $Var(X)$, is defined as follows:

$$Var(X) = E[(X-\mu)^2]$$

If X has infinite mean or if the mean of X does not exist, we say that $Var(X)$ does not exist.

The ***standard deviation of X*** is the nonnegative square root of $Var(X)$ if the variance exists.

Variance - 2

- **Ex12 (Book Ex4.3.1&4.3.2) Stock Price Changes.**

Consider the prices *A* and *B* of two stocks at a time in the future. Assume that *A* has the uniform distribution on the interval [25, 35] and *B* has the uniform distribution on the interval [15, 45].

Q1: What are their means and variances, respectively?

Sol: Both stocks have a mean price of 30.

$$Var(A) = \int_{25}^{35} (a-30)^2 \frac{1}{10} da = \frac{1}{10} \int_{-5}^{5} x^2 dx = \frac{1}{10} \frac{x^3}{3} \Big|_{x=-5}^{5} = \frac{25}{3},$$

$$Var(B) = \int_{15}^{45} (b-30)^2 \frac{1}{30} db = \frac{1}{30} \int_{-15}^{15} y^2 dy = \frac{1}{30} \frac{y^3}{3} \Big|_{y=-15}^{15} = 75.$$

$Var(B)$ is 9 times as large as Var(A).

$$\sigma_A = \sqrt{25/3} = 2.87, \text{ and } \sigma_B = \sqrt{75} = 8.66.$$

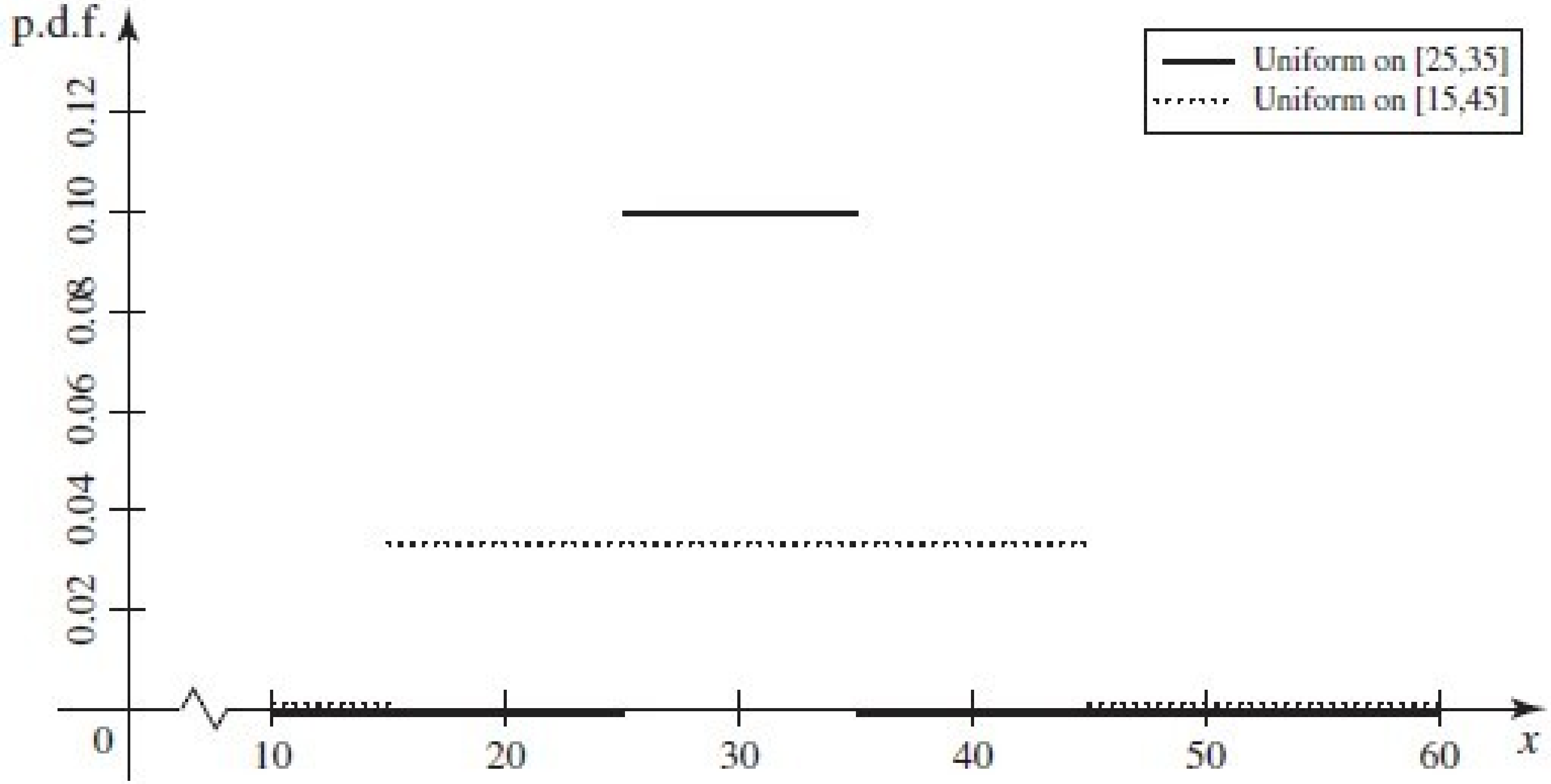


Figure 4.5 The p.d.f.'s of two uniform distributions in Example 4.3.1. Both distributions have mean equal to 30, but they are spread out differently.

Note: Variance depends only on the distribution.

Variance - 3

- **Ex13 (Book Ex4.3.3) Variance of a Discrete Distribution.** Suppose that a $R.V.$ X can take each of the five values −2, 0, 1, 3, and 4 with equal probability. We shall determine the variance and standard deviation of X.

$$E(X) = \frac{1}{5}(-2+0+1+3+4) = 1.2.$$

Let $\mu = E(X) = 1.2$, and define $W = (X - \mu)^2$. Then $Var(X) = E(W)$. We can easily compute the p.f. f of W:

x	-2	0	1	3	4
w	10.24	1.44	0.04	3.24	7.84
f(w)	1/5	1/5	1/5	1/5	1/5

$$\mathrm{Var}(X) = E(W) = \frac{1}{5}[10.24 + 1.44 + 0.04 + 3.24 + 7.84] = 4.56.$$

$$\sigma_X = 2.135.$$

Variance - 4

◆ **Theorem 4.3.1 Alternative Method for Calculating the Variance.** For every $R.V.\ X$,

$$Var(X) = E(X^2) - [E(X)]^2.$$

Proof: let $E(X) = \mu$. Then

$$\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= E[X^2 - 2\mu X + \mu^2] \\
&= E(X^2) - E[2\mu(X)] + E(\mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - [E(X)]^2
\end{aligned}$$

Variance - 5

- **Ex14 (Book Ex4.3.4) Variance of a Discrete Distribution.** Suppose that a $R.V.$ X can take each of the five values −2, 0, 1, 3, and 4 with equal probability. Determine the variance and standard deviation of X.

Sol: $$E(X^2) = \frac{1}{5}[(-2)^2 + 0^2 + 1^2 + 3^2 + 4^2] = 6.$$

Because $E(X) = 1.2$, Theorem 4.3.1 says that

$$Var(X) = 6 - (1.2)^2 = 4.56,$$

which agrees with the calculation in Example 4.3.3.

The variance of a distribution provides a measure of the **spread** or **dispersion** of the distribution around its mean. The small (large) the variance, the tightly concentration (wide spread) around μ.

Properties of the Variance - 1

- **Theorem 4.3.2** For each X, $Var(X) \geq 0$. If X is a bounded *R.V.*, then $Var(X)$ must exist and be finite.

- **Theorem 4.3.3** $Var(X) = 0$ if and only if there exists a constant c such that $\Pr(X = c) = 1$.

Proof: if there exists a constant c such that $\Pr(X = c) = 1$,
Then $E(X)=c$, and $\Pr[(X-c)^2=0]=1$. Therefore,
$Var(X)= E[(X-c)^2]=0$.
For only if, suppose that $Var(X)=0$. As $\Pr[(X-\mu)^2 \geq 0]=1$,
but $E[(X-\mu)^2]=0$. Based on Theorem 4.2.3 "Suppose that $E(X)=a$ and that either $\Pr(X \geq a)=1$ or $\Pr(X \leq a)=1$. Then $\Pr(X =a)=1$.", we have that $\Pr[(X-\mu)^2=0]=1$.
Hence, $\Pr(X=\mu)=1$.

Properties of the Variance - 2

- **Theorem 4.3.4** For constants a and b, let $Y = aX + b$.

Then $Var(Y) = a^2\,Var(X)$, and $\sigma_Y = |a|\sigma_X$.

Proof: let $E(X) = \mu$, then $E(Y) = a\mu + b$. Therefore,

$$Var(Y) = E[(aX + b - a\mu - b)^2]$$

$$= E[(aX - a\mu)^2]$$

$$= a^2 E[(X - \mu)^2]$$

$$= a^2 Var(X).$$

Q: What's $Var(X+b)$?

What's $Var(-X)$?

Properties of the Variance - 3

- **Theorem 4.3.5** If $X_1, \ldots, X_n$ are **independent** $R.V.s$ with finite means, then $Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n)$.

Proof: suppose first that n=2, $E(X_1)=\mu_1$, $E(X_2)=\mu_2$, then

$$E(X_1 + X_2) = \mu_1 + \mu_2.$$

$$Var(X_1 + X_2) = E[(X_1 + X_2 - \mu_1 - \mu_2)^2]$$

$$= E[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)]$$

$$= Var(X_1) + Var(X_2) + 2E[(X_1 - \mu_1)(X_2 - \mu_2)]$$

Since X_1 and X_2 are indepedent,

$$E[(X_1 - \mu_1)(X_2 - \mu_2)] = E(X_1 - \mu_1)E(X_2 - \mu_2) = 0$$

Therefore, $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$

Properties of the Variance - 4

- **Corollary 4.3.1** If $X_1, \ldots, X_n$ are independent $R.V.$s with finite means, and if $a_1, \ldots, a_n$ and b are arbitrary constants, then

$$Var(a_1X_1 + \cdots + a_nX_n + b) = a_1^2 Var(X_1) + \cdots + a_n^2 Var(X_n).$$

- **Ex15** Suppose that X and Y are independent $R.V.$s with $Var(X)=2$, $Var(Y)=3$. Find the values of $Var(2X-3Y+1)$.

Sol: $$\begin{aligned} Var(2X - 3Y + 1) &= Var(2X) + Var(-3Y) \\ &= 4Var(X) + 9Var(Y) \\ &= 4 \times 2 + 9 \times 3 \\ &= 35. \end{aligned}$$

Interquartile Range (IQR)

- **Definition 4.3.2** Let X be a $R.V.$ with quantile function $F^{-1}(p)$ for $0<p<1$. *The* ***interquartile range*** ***(IQR)*** is defined to be $F^{-1}(0.75) - F^{-1}(0.25)$.
- IQR is the **length of the interval** that contains the **middle half** of the distribution.
- **Ex16 (Book Ex4.3.9) The Cauchy Distribution.** Let X have the Cauchy distribution. The c.d.f. F of X is

$$F(x) = \int_{-\infty}^{x} \frac{dy}{\pi(1+y^2)} = \frac{1}{2} + \frac{\tan^{-1}(x)}{\pi},$$

where $\tan^{-1}(x)$ is the principal inverse of the tangent function, taking values from $-\pi/2$ to $\pi/2$ as x runs from $-\infty$ to ∞. The quantile function of X is then $F^{-1}(p) = \tan[\pi(p - 1/2)]$ for $0<p<1$. The IQR is

$F^{-1}(0.75) - F^{-1}(0.25) = \tan(\pi/4) - \tan(-\pi/4) = 2.$

The Variance of typical distributions - 1

- **The Bernoulli Distribution**. Let X have the Bernoulli distribution with parameter p, its c.d.f. is

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ 1 - p & \text{for } 0 \le x < 1 \\ 1 & \text{for } x \ge 1. \end{cases}$$

$$E(X) = 1 \times p + 0 \times (1 - p) = p,$$

$$E(X^2) = 1^2 \times p + 0^2 \times (1 - p) = p,$$

$$Var(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1 - p).$$

The Variance of typical distributions - 2

◆ **Ex17** Suppose that a box contains red balls and blue balls, and the proportion of red balls is p ($0\le p \le 1$). Suppose also that random sample of n balls is selected from the box with replacement. For i=1,…n. Let X_i=1if the ith selected ball is red and X_i=0 otherwise. If $X=\sum_{i=1}^{n} X_i$, what's the distribution of X?

X is a **binomial *R.V***. with parameter p and n. $Var(X)$=?

Since X_1,… X_n are independent,

$$Var(X)=\sum_{i=1}^{n} Var(X_i).$$

$\because E(X_i)=p$ for $i=1,\cdots,n$, and $X_i^2=X_i$ for each i.

$$Var(X_i)=E(X_i^2)-[E(X_i)]^2=p-p^2=p(1-p).$$

$$\therefore Var(X)=np(1-p).$$

The Variance of typical distributions - 3

- **The Poisson Distribution**. Let λ > 0. A *R.V. X* has the Poisson distribution with **mean λ**. Its p.f. is

$$f(x\mid\lambda)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & \text{for } x=0,1,2,\ldots,\\ 0 & \text{otherwise.}\end{cases}$$

$$E[X(X-1)]=\sum_{x=0}^{\infty}x(x-1)f(x\mid\lambda)=\sum_{x=2}^{\infty}x(x-1)\frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$=\lambda^{2}\sum_{x=2}^{\infty}\frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!}\overset{y=x-2}{=}\lambda^{2}\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^{y}}{y!}=\lambda^{2}$$

$$\because E[X(X-1)]=E(X^{2})-E(X)=E(X^{2})-\lambda\Rightarrow E(X^{2})=\lambda^{2}+\lambda$$

$$\therefore Var(X)=E(X^{2})-[E(X)]^{2}=\lambda$$

The Variance of typical distributions - 4

- **The Uniform Distribution.** If X has the uniform distribution on an interval $[a,b]$, then the p.d.f of X is

$$f(x|a,b)=\begin{cases}\dfrac{1}{b-a} & \text{if } a\le x\le b,\\ 0 & \text{otherwise.}\end{cases}$$

$$E(x)=\int_{-\infty}^{\infty} xf(x)dx=\frac{1}{b-a}\int_a^b xdx=\frac{a+b}{2}$$

$$E(X^2)=\int_{-\infty}^{\infty} x^2 f(x)dx=\frac{1}{b-a}\int_a^b x^2dx=\frac{1}{b-a}\cdot\frac{b^3-a^3}{3}=\frac{b^2+ab+a^2}{3}$$

$$Var(X)=E(X^2)-[E(X)]^2=\frac{b^2+ab+a^2}{3}-(\frac{a+b}{2})^2=\frac{(b-a)^2}{12}$$

The Variance of typical distributions - 5

- **The Normal Distribution**. A *R.V. X* has the ***normal distribution*** with **mean μ** and **variance σ^2** ($-\infty<\mu<\infty$ and $\sigma>0$) if X has a continuous distribution with the following p.d.f.:

$$f(x\mid\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\quad \text{for } -\infty<x<\infty.$$

$$Var(X)=\int_{-\infty}^{+\infty}(x-\mu)^2\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x-\mu)^2e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

$$\overset{t=\frac{x-\mu}{\sigma}}{=}\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^2e^{-\frac{t^2}{2}}dt=-\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t\,de^{-\frac{t^2}{2}}=\sigma^2.$$

The Variance of typical distributions - 8

- **The Exponential Distribution**. A *R.V.* X has ***the exponential distribution with parameter β*** ($\beta > 0$) if X has a continuous distribution with the following p.d.f.:

$$f(x) = \begin{cases} \beta e^{-\beta x} & \text{for } x > 0, \\ 0 & \text{for } x \le 0. \end{cases}$$

$$E(X) = \int_0^{\infty} x\beta e^{-\beta x} dx \overset{\text{set } u = x\beta}{=} \frac{1}{\beta}\int_0^{+\infty} u e^{-u} du = \frac{1}{\beta} e^{-x}(-x-1)\Big|_0^{+\infty} = \frac{1}{\beta}.$$

$$E(X^2) = \int_0^{\infty} x^2 \beta e^{-\beta x} dx = \frac{1}{\beta^2}\int_0^{\infty} (\beta x)^2 e^{-\beta x} d\beta x \overset{\text{set } u = x\beta}{=} \frac{1}{\beta^2}\int_0^{+\infty} u^2 e^{-u} du$$

$$= -\frac{1}{\beta^2} e^{-x}(x^2 + 2x + 2)\Big|_0^{\infty} = \frac{2}{\beta^2}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{2}{\beta^2} - \frac{1}{\beta^2} = \frac{1}{\beta^2}$$

PROPERTIES OF SPECIAL DISTRIBUTIONS

- **Bernoulli Distribution.**

  $E(X) = p, Var(X) = p(1-p).$
- **Binomial Distribution. $X \sim B(n,p)$**

  $E(X)=np, Var(X)=np(1\text{-}p).$
- **Poisson Distribution. $X \sim P(\lambda)$**

  $E(X)=\lambda, Var(X)=\lambda.$
- ***Uniform Distribution. $X \sim U(a,b)$***

  $E(X)=(a+b)/2, Var(X)=(b\text{-}a)^2/12.$
- ***Exponential Distribution.***

  $E(X)=1/\beta, Var(X)=1/\beta^2$.
- ***Normal Distribution. $X \sim N(\mu,\sigma^2)$***

  $E(X)=\mu, Var(X)=\sigma^2.$

M.S.E. vs. M.A.E. - 1

- **Definition 4.5.2 Mean Squared Error/M.S.E.** The number $E[(X-d)^2]$ is the ***mean square error (M.S.E.)*** of the prediction d.
- **Theorem 4.5.2** Let X be a $R.V.$ with finite variance σ^2 and let $\mu=E(X)$. For every value d,

$$E[(X-\mu)^2] \le E[(X-d)^2]$$

The number d for which the M.S.E. is minimized is $E(X)$.

- **Definition 4.5.3 Mean Absolute Error/ M.A.E.** The value $E(|X\text{-}d|)$ is the ***mean absolute error (M.A.E.)*** of the prediction d.
- **Theorem 4.5.3** Let X be a $R.V.$ with finite mean, and let m be a median of the distribution of X. For every d,

$$E(|X-m|) \le E(|X-d|).$$

M.S.E. vs. M.A.E. - 2

- **Ex18 (Book Ex4.5.9) Predicting a Discrete Uniform *R.V*.** Suppose that the probability is 1/6 that a *R.V.* X will take each of the following 6 values: 1,2,3,4,5,6. We shall determine the prediction for which the M.S.E. is minimum and the prediction for which the M.A.E is minimum.

Sol:

$$E(X)=\frac{1}{6}(1+2+3+4+5+6)=3.5$$

Thus, the M.S.E. will be minimized by the unique value d=3.5.

Every number m in the closed interval $3\leq m\leq 4$ is a median of the given distribution. Thus, the M.A.E. will be minimized by every value of d such that $3\leq d\leq 4$. In this case the mean is also a median of X.

Moments - 1

- **Definition <u>Moments</u>**. For each $R.V.\ X$ and every **positive integer *k***, the expectation ***E*(*X*^{k})** is called ***the kth moment of X***.
- The kth moment exists if and only if $E(|X|^k) < \infty$. It is possible that all moments of X exist even though X is not bounded.
- **Theorem 4.4.1** If $E(|X|^k) <\infty$ for some positive integer k, then $E(|X|^j) <\infty$ for every positive integer j such that $j<k$.

Theorem 4.4.1 says that if the kth moment of X exists, then all moments of lower order must also exist.

Moments - 2

- Proof: we assume, for convenience, that the distribution of X is continuous and the p.d.f. is f. Then

$$
\begin{aligned}
E(|X|^j) &= \int_{-\infty}^{\infty} |x|^j f(x)dx \\
&= \int_{|x|\le 1} |x|^j f(x)dx + \int_{|x|>1} |x|^j f(x)dx \\
&\le \int_{|x|\le 1} 1 \cdot f(x)dx + \int_{|x|>1} |x|^k f(x)dx \\
&\le \int_{|x|\le 1} 1 \cdot f(x)dx + \int_{-\infty}^{\infty} |x|^k f(x)dx \\
&= \Pr(|X| \le 1) + E(|x|^k).
\end{aligned}
$$

- If follows from Theorem 4.4.1 that if $E(X^2) < \infty$, then both the mean of X and the variance of X exist.

Moments - 3

- **Definition Central Moments.** Suppose that X is a $R.V.$ for which $E(X) = \mu$. For every positive integer k, the expectation $E[(X-\mu)^k]$ is called ***the kth central moment of X*** or ***the kth moment of X about the mean***.
- The variance of X is the 2nd central moment of X.
- For every distribution, the first central moment is ?

  0. Because $E(X- \mu)=E(X)-\mu=0$.
- If the distribution of X is symmetric with respect to its mean μ, and if the central moment $E[(X-\mu)^k]$ exists for a given odd integer k, then the value of $E[(X-\mu)^k]$ will be 0 because the positive and negative terms in this expectation will cancel on another.

$e.g.,\quad f(x) = ce^{-(x-3)^2/2} \; for -\infty < x < \infty.$

Moments - 4

◆ **Definition 4.4.1 Skewness.** Let X be a $R.V.$ with mean μ, standard deviation σ, and finite third moment. The skewness of X is defined to be $E[(X-\mu)^3]/\sigma^3$.

◆ It measures only the lack of symmetry rather than the spread of the distribution.

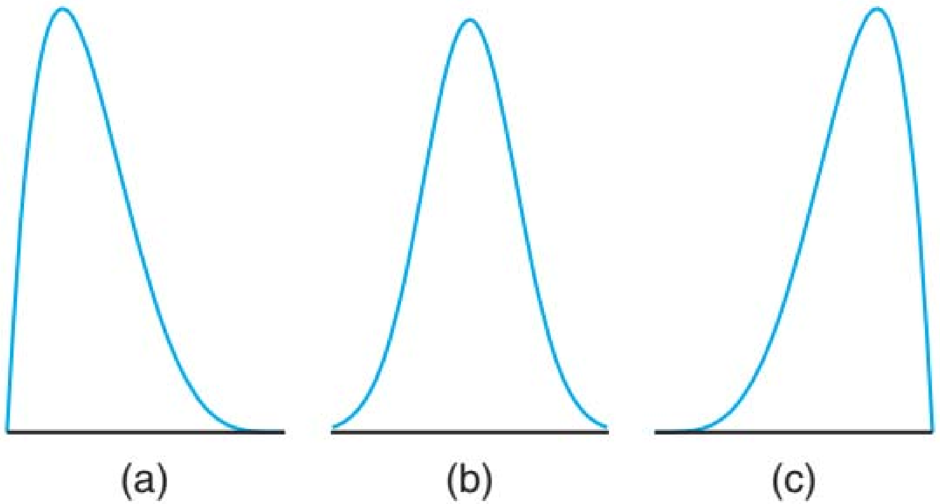


Figure 4.6 Skewness. (a) skewness<0; (b) skewness=0; (c) skewness>0.

Moment Generating Functions-1

◆ **Definition 4.4.2 Moment Generating Functions**.
Let X be a $R.V.$ For each real number t, define

$$\psi(t) = E(e^{tX}).$$

The function $\psi(t)$ is ***called the moment generating function*** (abbreviated **m.g.f.**) of X.

◆ If $R.V.\ X$ is bounded, then its m.g.f. must be finite for all values of t.

◆ For every $R.V.\ X$, $\psi(0) = E(1) = 1$.

◆ If $R.V.\ X$ is unbounded, then its m.g.f. might be finite for some values of t and might not be finite for others.

Moment Generating Functions-2

- **Theorem 4.4.2** Let X be a $R.V.$s whose m.g.f. $\psi(t)$ is finite for all values of t in some open interval around the point $t = 0$. Then, for each integer $n > 0$, the nth moment of X which is $E(X^n)$, is finite and equals the nth derivative $\psi^{(n)}(t)$ at $t = 0$. That is, $E(X^n) = \psi^{(n)}(0)$ *for* $n = 1, 2, \ldots$.

It can be shown that the derivative $\psi'(t)$ exists at the point t=0.

$$\psi'(0) = \left[\frac{d}{dt}E(e^{tX})\right]_{t=0} = E\left[\frac{d}{dt}(e^{tX})_{t=0}\right] = E\left[(Xe^{tX})_{t=0}\right] = E(X).$$

$$\psi^{(n)}(0) = \left[\frac{d^n}{dt^n}E(e^{tX})\right]_{t=0} = E\left[\frac{d^n}{dt^n}(e^{tX})_{t=0}\right] = E\left[(X^n e^{tX})_{t=0}\right] = E(X^n).$$

The derivative of m.g.f. is equal to the mean of the derivative.

Moment Generating Functions-3

- **Ex19 (Book Ex 4.4.3) Calculating an m.g.f.**

Suppose that X is a $R.V.$ with the p.d.f. as follows:

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

We shall determine the m.g.f. of X and also $Var(X)$.

Sol: for each real number t ,

$$\psi(t) = E(e^{tX}) = \int_0^\infty e^{tx} e^{-x} dx = \int_0^\infty e^{(t-1)x} dx.$$

$\psi(t)$ is finite only for $t < 1$. For such value of t, $\psi(t) = \dfrac{1}{1-t}$.

The first two derivatives of ψ are

$$\psi'(t) = \frac{1}{(1-t)^2} \text{ and } \psi''(t) = \frac{2}{(1-t)^3}.$$

Therefore, $E(X) = \psi'(0) = 1$ and $E(X^2) = \psi''(0) = 2$.

$Var(X) = \psi''(0) - [\psi'(0)]^2 = 1.$

Properties of m.g.f.-1

- **Theorem 4.4.3** Let X *be a R.V.* for which the m.g.f. is ψ_1; let $Y = aX + b$, where a and b are given constants; and let ψ_2 denote the m.g.f. of Y. Then for every value of t such that $\psi_1(at)$ is finite,

$$\psi_2(t) = e^{bt}\psi_1(at).$$

Proof: by the definition of an m.g.f.,

$$\psi_2(t) = E(e^{tY}) = E[e^{t(aX+b)}] = E[e^{taX}e^{tb}] = e^{tb}E(e^{taX})$$

$$= e^{bt}E(e^{atX}) = e^{bt}\psi_1(at).$$

Properties of m.g.f.-2

◆ **Ex20 (Book Ex4.4.4) Calculating the m.g.f. of a Linear Function**. Suppose that the p.d.f. of X is

$$f(x)=\begin{cases} e^{-x} & \text{for } x>0, \\ 0 & \text{otherwise.} \end{cases}$$

Let Y=3-2X. What's the m.g.f. of Y?

Sol: Previously we have obtained that the m.g.f. of X for t<1 is

$$\psi_1(t)=\frac{1}{1-t}.$$

If Y= 3-2X, the m.g.f. of Y is finite for t>-1/2 and will be

$$\psi_2(t)=e^{3t}\psi_1(-2t)=\frac{e^{3t}}{1+2t}.$$

Properties of m.g.f.-3

- **Theorem 4.4.4** Suppose that $X_1, \ldots, X_n$ are n **independent** $R.V.s$; and for $i = 1, \ldots, n$, let ψ_i denote the m.g.f. of X_i. Let $Y = X_1 + \ldots + X_n$, and let the m.g.f. of Y be denoted by ψ. Then for every value of t such that $\psi_i(t)$ is finite for $i = 1, \ldots, n$,

$$\psi(t) = \prod_{i=1}^{n} \psi_i(t).$$

Proof: by the definition of an m.g.f.,

$$\psi(t) = E(e^{tY}) = E[e^{t(X_1+X_2+\cdots+X_n)}] = E[\prod_{i=1}^{n} e^{tX_i}] = \prod_{i=1}^{n} E(e^{tX_i})$$

$$= \prod_{i=1}^{n} \psi_i(t).$$

Properties of m.g.f.-4

- m.g.f. is the expected value of a function of X, it depends only on the distribution of X. If X and Y have the same distribution, they have the same m.g.f.

Theorem 4.4.5 If the m.g.f.*s* of two *R.V.s* X_1 and X_2 are finite and identical for all values of t in an open interval around the point $t = 0$, then the probability distributions of X_1 and X_2 must be identical.

m.g.f.s for Bernoulli and Binomial *R.V.*s

- **Bernoulli m.g.f.** Let X have the Bernoulli distribution with parameter p, that is, assume that X takes only the two values 0 and 1 with $\Pr(X = 1) = p$. For $-\infty < t < \infty$,

$$\psi(t) = E(e^{tX}) = \Pr(X=1)e^{t\cdot 1} + \Pr(X=0)e^{t\cdot 0}$$
$$= pe^t + 1 - p.$$

- **Binomial m.g.f.** Theorem 5.2.1 says if the *R.V.*s $X_1,\ldots,X_n$ form n Bernoulli trials with parameter p, and if $X = X_1 + \ldots + X_n$, the X has the binomial distribution with parameters n and p.

For $-\infty < t < \infty$,

$$\psi(t) = \prod_{i=1}^{n} \psi_i(t) = (pe^t + 1 - p)^n.$$

m.g.f.s for Binomial *R.V.*s - 2

- **Theorem 4.4.6** If X_1 and X_2 are independent *R.Vs*, and if X_i has the binomial distribution with parameters n_i and p (i = 1, 2), then $X_1 + X_2$ has the binomial distribution with parameters $n_1 + n_2$ and p.

Proof: let ψ_i denote the m.g.f. of X_i for i=1,2. We know

$$\psi_i(t) = (pe^t + 1 - p)^{n_i}.$$

Let ψ denote the m.g.f. of $X_1 + X_2$. Then by Theorem 4.4.4,

$$\psi(t) = (pe^t + 1 - p)^{n_1 + n_2}.$$

The above function is the m.g.f. of the binomial distribution with parameters n_1+n_2 and p. Hence by Theorem 4.4.5, $X_1 + X_2$ has the binomial distribution.

m.g.f for Poisson Distribution-1

- **Theorem 5.4.3** The m.g.f. of the Poisson distribution with mean λ is

$$\psi(t) = e^{\lambda(e^t - 1)},$$

for all real t.

Proof: for every value $t(-\infty < t < \infty)$,

$$\psi(t) = E(e^{tX}) = \sum_{x=0}^{\infty} \frac{e^{tx} e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} = e^{-\lambda} e^{e^t \lambda}$$

$$= e^{\lambda(e^t - 1)}.$$

m.g.f for Poisson Distribution-2

- **Theorem 5.4.4** If the *R.V.*s $X_1,\ldots,X_k$ are independent and if X_i has the Poisson distribution with mean λ_i ($i=1,2,\ldots,k$), then the sum $X_1+\ldots+X_k$ has the Poisson distribution with mean $\lambda_1+\ldots+\lambda_k$.

Proof : let $\psi_i(t)$ denote the m.g.f. of X_i for $i=1,2,\ldots,k$, and let $\psi(t)$ denote the m.g.f. of the sum $X_1+\ldots+X_k$. Since $X_1,\ldots,X_k$ are independent, it follows that, for $-\infty<t<\infty$,

$$\psi(t)=\prod_{i=1}^{k}\psi_i(t)=\prod_{i=1}^{k}e^{\lambda i(e^t-1)} = e^{(\lambda_1+\cdots+\lambda_k)(e^t-1)}.$$

This $\psi(t)$ is the m.g.f. of the Poisson distribution with mean $\lambda_1+\ldots+\lambda_k$. Hence, the distribution of $X_1+\ldots+X_k$ must be Poisson.

m.g.f for Normal Distribution-1

- **Theorem 5.6.2** The m.g.f. of the normal distribution is

$$\psi(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2) \quad \text{for } -\infty < t < \infty.$$

- **Theorem 5.6.3** The mean and variance of the normal distribution are μ and σ^2, respectively.

Proof: the first two derivatives of the normal m.g.f are

$$\psi'(t) = (\mu + \sigma^2 t)\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

$$\psi''(t) = [(\mu + \sigma^2 t)^2 + \sigma^2]\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

Plugging t=0 into each of these derivatives yields

$$E(x) = \psi'(0) = \mu, \;\; Var(x) = \psi''(0) - [\psi'(0)]^2 = \sigma^2.$$

m.g.f for Normal Distribution - 2

- **Theorem 5.6.7** If the *R.V.*s $X_1, \ldots, X_k$ are independent and if X_i has the normal distribution with mean μ_i and variance σ_i^2 ($i = 1, \ldots, k$), then the sum $X_1 + \ldots + X_k$ has the normal distribution with *mean* $\mu_1 + \ldots + \mu_k$ and variance $\sigma_1^2 + \ldots + \sigma_k^2$.

Proof: let $\psi_i(t)$ denote the m.g.f. of X_i for $i=1,2,\ldots,k$, and let $\psi(t)$ denote the m.g.f. of the sum $X_1+\ldots+X_k$. Since $X_1,\ldots,X_k$ are independent,

$$\psi(t) = \prod_{i=1}^{k} \psi_i(t) = \prod_{i=1}^{k} \exp(\mu_i t + \frac{1}{2}\sigma_i^2 t^2)$$

$$= \exp\left[\left(\sum_{i=1}^{k} \mu_i\right)t + \frac{1}{2}\left(\sum_{i=1}^{k} \sigma_i^2\right)t^2\right] \quad \text{for } -\infty < t < \infty.$$

m.g.f for Normal Distribution - 3

- **Corollary 5.6.1** If the *R.V.*s $X_1, \ldots, X_k$ are independent, if X_i has the normal distribution with mean μ_i and variance σ_i^2 $(i = 1, \ldots, k)$, and if $a_1, \ldots, a_k$ and b are constants for which at least one of the values $a_1, \ldots,$ a_k is different from 0, then the variable $a_1X_1 + \ldots + a_kX_k + b$ has the normal distribution with mean $a_1\mu_1 + \ldots + a_k\mu_k + b$ and variance $a_1^2\sigma_1^2 + \ldots + a_k^2\sigma_k^2$.

Sample Mean

- **Definition 5.6.3 Sample Mean**. Let $X_1, \ldots, X_n$ be $R.V.$s. The average of these n random variables,

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

is called their ***sample mean*** and is commonly denoted by $\bar{X}_n$.

- **Corollary 5.6.2** Suppose that the $R.V.s$ $X_1, \ldots, X_n$ form a random sample from the normal distribution with mean μ and variance σ^2, *and let* $\bar{X}_n$ *denote their sample mean. Then* $\bar{X}_n$ *has the normal distribution with mean* μ *and variance* σ^2/n.

m.g.f for Exponential Distribution

- **Theorem 5.7.8** The m.g.f for an exponential distribution with parameter β is

$$\psi(t) = E(e^{tX}) = \int_0^\infty e^{tx}\beta e^{-\beta x}dx = \int_0^\infty \beta e^{(t-\beta)x}dx$$

$$= \frac{\beta}{t-\beta} e^{(t-\beta)x}\Big|_0^\infty$$

$$\overset{\text{for } t<\beta}{=} \frac{\beta}{\beta - t}.$$

m.g.f for Uniform Distribution

- The m.g.f for the **uniform distribution** in the interval $[a, b]$ is :

$$\psi(t) = E(e^{tX}) = \int_a^b e^{tx} \frac{1}{b-a} dx = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

Covariance - 1

◆ **Definition 4.6.1 Covariance.** Let X and Y be $R.V.$s having finite means. Let $E(X) = \mu_X$ and $E(Y) = \mu_Y$ The covariance of X and Y , which is denoted by $Cov(X, Y)$, is defined as $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$, if the above expectation exists.

◆The value of $Cov(X, Y)$ can be positive, negative, or 0.

◆The covariance attempts to measure the dependence of two $R.V.$s, or the degree to which X and Y vary together. In particular, if $Cov(X, Y)>0$, it's more likely that $X>\mu_X$ and $Y>\mu_Y$ (or $X<\mu_X$ and $Y<\mu_Y$) occurs than that $X>\mu_X$ and $Y<\mu_Y$ (or $X<\mu_X$ and $Y>\mu_Y$).
If $Cov(X, Y)=0$, the prob. that X and Y on the same sides of their respective means are the same as the prob. that X and Y on the opposite sides of their means.

Covariance - 2

- **Ex21 (Book Ex4.6.2) Test Scores.** Let X and Y be the test scores and they have the joint p.d.f.

$$f(x,y)=\begin{cases}2xy+0.5 & \text{for } 0\le x\le 1 \text{ and } 0\le y\le 1,\\ 0 & \text{otherwise.}\end{cases}$$

Compute the covariance $Cov(X,\ Y)$.

Sol: The symmetry in the joint p.d.f. means that X and Y have the same marginal distribution; hence, $\mu_X=\mu_Y$.

$$\mu_X=\int_0^1 x\left[\int_0^1 (2xy+0.5x)dy\right]dx=\int_0^1\int_0^1 [2x^2y+0.5x]dydx$$

$$=\int_0^1 [x^2+0.5x]dx=\frac{1}{3}+\frac{1}{4}=\frac{7}{12},$$

$$Cov(X,\ Y)=\int_0^1\int_0^1 (x-\frac{7}{12})(y-\frac{7}{12})(2xy+0.5)dydx=\frac{1}{144}.$$

Covariance - 3

- **Theorem 4.6.1** For all *R.V.*s X and Y such that $\sigma_X^2 < \infty$ and $\sigma_Y^2 < \infty$, $Cov(X, Y) = E(XY) - E(X)E(Y)$.

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

$$= E(XY - X\mu_Y - \mu_X Y + \mu_X \mu_Y)$$

$$= E(XY) - \mu_Y \mu_X - \mu_X \mu_Y + \mu_X \mu_Y$$

$$= E(XY) - E(X)E(Y).$$

Ex22 For all *R.V.*s X and Y for all constants a, b, c, d, show

$$Cov(aX+b, cY+d) = ac\, Cov(X, Y).$$

$$Cov(aX+b, cY+d) = E[(aX+b)(cY+d)] - E(aX+b)\, E(cY+d)$$

$$= acE(XY) + adE(X) + bcE(Y) + bd - acE(X)E(Y) - adE(X)$$
$$- bcE(Y) - bd$$

$$= acCov(X, Y).$$

Correlation - 1

- **Definition 4.6.2 Correlation**. Let X and Y be $R.V.$s with finite variances $\sigma_X{}^2$ and $\sigma_Y{}^2$, respectively. Then the correlation of X and Y , which is denoted by $\rho(X, Y)$, is defined as follows:

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}.$$

- The magnitude of $Cov(X, Y)$ is influenced by the overall magnitudes of X and Y.
- Correlation is a measure of association between X and Y that is not driven by arbitrary changes in the scales of one or the other $R.V$.
- e.g., $Cov(2X,Y)=2Cov(X,Y)$.

Correlation - 2

- **Theorem 4.6.2 Schwarz Inequality**. For all *R.V.s* U and V such that $E(UV)$ exists,

$$[E(UV)]^2 \le E(U^2)E(V^2).$$

If, in addition, the right-hand side is finite, then the two sides are equal if and only if there are nonzero constants a and b such that $aU + bV = 0$ with probability 1.

- **Theorem 4.6.3 Cauchy-Schwarz Inequality.** Let X and Y be *R.V.s* with finite variance. Then

$$[Cov(X,Y)]^2 \le \sigma_X^2 \sigma_Y^2,$$

and $$-1 \le \rho(X, Y) \le 1.$$

Furthermore, the equality holds if and only if there are nonzero constants a and b and a constant c such that $aX + bY = c$ with probability 1.

Correlation - 3

- **Definition 4.6.3 Positively / Negatively Correlated / Uncorrelated.** It is said that X and Y are positively correlated if $\rho(X, Y) > 0$, that X and Y are negatively correlated if $\rho(X, Y) < 0$, and that X and Y are uncorrelated if $\rho(X, Y) = 0$.
- **Ex23 (Book Ex4.6.3) Test Scores.** Back in Ex4.6.2, we've already calculated that $Cov(X,Y)=1/144$. The variances of X and Y are both equal to 11/144, so the correlation is

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{1}{11}.$$

Properties of Covariance and Correlation-1

- **Theorem 4.6.4** If X and Y are independent $R.V.s$ with $0<\sigma_X^2<\infty$ and $0<\sigma_Y^2<\infty$, then $Cov(X, Y)=\rho(X, Y)=0$.

Proof: if X and Y are independent, then $E(XY)=E(X)E(Y)$. Therefore, $Cov(X, Y)=0$. Also, it follows that $\rho(X, Y) = 0$.

- **Ex24 (Book Ex4.6.4) Dependent but Uncorrelated *R.V.*s.** Suppose that $R.V.$ X can take only the three values -1, 0, and 1, and that each of these three values has the same probability. Also, let the $R.V.$ Y be defined by the relation $Y = X^2$. We shall show that X and Y are dependent but uncorrelated.

$$E(XY) = E(X^3) = E(X) = 0, \quad Cov(X,Y) = E(XY) - E(X)E(Y) = 0.$$

Here, X and Y are clearly dependent, but are uncorrelated.

The converse of Theorem4.6.4 is not true.

Properties of Covariance and Correlation-2

- **Theorem 4.6.5** Suppose that X is a $R.V.$ such that $0<\sigma_X{}^2<\infty$, and $Y=aX+b$ for some constants a and b, where $a\neq 0$. If $a>0$, then $\rho(X, Y)=1$. If $a<0$, then $\rho(X, Y)=-1$.

Proof: since $Y=aX+b$, then $\mu_Y= a\mu_X +b$. Y-$\mu_Y = a(X- \mu_X)$.

Therefore,

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = aE[(X-\mu_X)^2] = a\sigma_X^2$$

Since $\sigma_Y = |a|\sigma_X$,

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X\sigma_Y} = \frac{a\sigma_X^2}{|a|\sigma_X^2} = \begin{cases} 1 & \text{if } a>0 \\ -1 & \text{if } a<0 \end{cases}$$

$|\rho(X, Y)=1|$ implies that X and Y are linearly related.

There is a converse to Theorem 4.6.5.

Properties of Covariance and Correlation-3

- Proof: let $\mu_X=E(X)$, $\mu_Y=E(Y)$. Apply Theorem 4.6.2 (Schwarz Inequality) with $U=X-\mu_X$, $V=Y-\mu_Y$, then

$$Cov(X,Y)^2 \le Var(X)\,Var(Y). \qquad (1)$$

Now $|\rho(X, Y)=1|$ is equivalent to the equality in (1). According to Theorem 4.6.2, we obtain the quality in (1) if and only if these exists constants a and b such that $aU+bV=0$, that $a(X-\mu_X)+b(Y-\mu_Y)=0$ with probability 1. So $|\rho(X, Y)=1|$ implies that $aX+bY=a\mu_X+b\mu_Y$. Therefore, $|\rho(X, Y)=1|$ implies that X and Y are linearly related.

Properties of Covariance and Correlation-4

- **Theorem 4.6.6** If X and Y are $R.V.$s such that $Var(X) <\infty$ and $Var(Y) <\infty$, then

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

Proof: since $E(X+Y)=\mu_X+\mu_Y$, then

$$Var(X+Y) = E[(X+Y-\mu_X-\mu_Y)^2]$$

$$= E[(X-\mu_X)^2+(Y-\mu_Y)^2+2(X-\mu_X)(Y-\mu_Y)]$$

$$= Var(X)+Var(Y)+2\text{Cov}(X,Y).$$

Corollary 4.6.1 Let a, b, and c be constants. Under the conditions of Theorem 4.6.6,

$$Var(aX+bY+c)=a^2Var(X)+b^2Var(Y)+2abCov(X,Y).$$

$$Var(X-Y)=Var(X)+Var(Y)-2Cov(X,Y).$$

Properties of Covariance and Correlation-5

- **Ex25** Prove that

$$Cov(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j Cov(X_i, Y_j).$$

Proof:

$$Cov(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j) = E\left[\sum_{i=1}^{m} a_i (X_i - \mu_{X_i}) \sum_{j=1}^{n} b_j (Y_j - \mu_{Y_i})\right]$$

$$= E\left[\sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j (X_i - \mu_{X_i})(Y_j - \mu_{Y_i})\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j E\left[(X_i - \mu_{X_i})(Y_j - \mu_{Y_i})\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j Cov(X_i, Y_j).$$

Properties of Covariance and Correlation-6

- **Theorem 4.6.7** If $X_1, \ldots, X_n$ are $R.V.s$ such that $Var(X_i)<\infty$ for $i = 1, \ldots, n$, then

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum\sum_{i<j} Cov(X_i, X_j).$$

Proof: for every $R.V.Y$, Cov(Y,Y)=Var(Y). We can obtain

$$Var(\sum_{i=1}^{n} X_i) = Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j).$$

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + \sum\sum_{i\neq j} Cov(X_i, X_j).$$

$$= \sum_{i=1}^{n} Var(X_i) + 2\sum\sum_{i<j} Cov(X_i, X_j).$$

Properties of Covariance and Correlation-7

- **Corollary 4.6.2** If $X_1, \ldots, X_n$ are uncorrelated random variables (that is, if X_i and X_j are uncorrelated whenever $i = j$), then

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i).$$

Conditional Expectation - 1

- **Definition 4.7.1 Conditional Expectation/Mean** Let X and Y be $R.V.$s such that the mean of Y exists and is finite. The conditional expectation (or conditional mean) of Y given $X = x$ is denoted by $E(Y|x)$ and is defined to be the expectation of the conditional distribution of Y given $X = x$.

if Y has a continuous conditional distribution given $X = x$ with conditional p.d.f. $g_2(y|x)$, then

$$E(Y \mid x) = \int_{-\infty}^{\infty} y g_2(y \mid x) dy$$

If Y has a discrete conditional distribution given $X = x$ with conditional p.f. $g_2(y|x)$, then

$$E(Y \mid x) = \sum_{\text{All } y} y g_2(y \mid x).$$

Table 4.2 Joint p.f. $f(x, y)$ of X and Y in Example 4.7.2 together with marginal p.f.'s $f_1(x)$ and $f_2(y)$

y	x = 1	2	3	4	5	6	7	8	$f_2(y)$
0	0.040	0.028	0.012	0.008	0.008	0.004	0	0	0.100
1	0.048	0.084	0.100	0.120	0.100	0.060	0.020	0.004	0.536
2	0.004	0.020	0.040	0.060	0.080	0.044	0.020	0.012	0.280
3	0	0.008	0.012	0.020	0.020	0.012	0.008	0.004	0.084
$f_1(x)$	0.092	0.140	0.164	0.208	0.208	0.120	0.048	0.020	

Ex25 (Book Ex4.7.2) Calculate $E(Y|x)$.

Suppose x=4. $E(Y|x=4) = \sum_{\text{All } y} y g_2(y \mid x=4) = \sum_{\text{All } y} y \frac{f(4, y)}{f_1(4)}$

$$E(Y \mid 4) = 0 \times 0.0385 + 1 \times 0.5769 + 2 \times 0.2885 + 3 \times 0.0962 = 1.442.$$

x	1	2	3	4	5	6	7	8
$E(Y\|x)$	0.609	1.057	1.317	1.442	1.538	1.533	1.75	2

Conditional Expectation - 2

- **Definition 4.7.2 Conditional Means as *R.V.*s.** Let $h(x)$ stand for the function of x that is denoted $E(Y|x)$ in either

$$E(Y \mid x) = \int_{-\infty}^{\infty} y g_2(y \mid x) dy$$

or

$$E(Y \mid x) = \sum_{\text{All } y} y g_2(y \mid x).$$

Define the symbol $E(Y|X)$ to mean $h(X)$ and call it the ***conditional mean of Y given X***.

- $E(Y|X)$ is a $R.V.$ (a function of X) whose value when $X=x$ is $E(Y|x)$. $E(Y|X)$ is itself a $R.V.$ with its own probability distribution, which can be derived from X.
- Back in Ex25, what is $E(Y|X)$?

x	1	2	3	4	5	6	7	8
$E(Y\|x)$	0.609	1.057	1.317	1.442	1.538	1.533	1.75	2

Conditional Expectation - 3

- **Theorem 4.7.1 Law of Total Probability for Expectations.** Let X and Y be $R.V.$s such that Y has finite mean. Then $E[E(Y|X)] = \mathrm{E}(Y)$.

Proof: assume, for convenience, that X and Y have a continuous joint distribution

$$
\begin{aligned}
E[E(Y \mid X)] &= \int_{-\infty}^{\infty} E(Y \mid x) f_1(x) dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y g_2(y \mid x) f_1(x) dy dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y \frac{f(x,y)}{f_1(x)} f_1(x) dy dx \\
&= \int_{-\infty}^{\infty} y [\int_{-\infty}^{\infty} f(x,y) dx] dy \\
&= \int_{-\infty}^{\infty} y f_2(y) dy = E(Y).
\end{aligned}
$$

$E\{E[r(X,Y)|X]\} = E[r(X,Y)]$.

Conditional Expectation - 4

- **Ex26 (Book Ex4.7.6)**. Suppose that a point X is chosen in accordance with the uniform distribution on the interval [0,1]. After the value $X=x$ has been observed ($0<x<1$), a point Y is chosen in accordance with a uniform distribution on the interval $[x,1]$. Determine $E(Y)$.

Sol: for each given value of $x(0<x<1)$,

$$E(Y\mid x)=\frac{x+1}{2}. \qquad E(Y\mid X)=\frac{X+1}{2}.$$

$$E(Y)=E[E(Y\mid X)]=\frac{E(X)+1}{2}=\frac{\frac{1}{2}+1}{2}=\frac{3}{4}.$$

Conditional Expectation - 5

- **Ex26 (Book Ex4.7.7). Linear Conditional Expectation**. Suppose that $E(Y|X) = aX + b$ for some constants a and b. Determine the value of $E(XY)$ in terms of $E(X)$ and $E(X^2)$.
- Sol: we know that

$$E(XY) = E[E(XY \mid X)]$$

here X is considered to be given and fixed in the conditional expectation,

$$E[E(XY \mid X)] = E[XE(Y \mid X)] = E[aX^2 + bX]$$

$$= aE(X^2) + bE(X).$$

Conditional Expectation - 6

- **Definition 4.7.3 Conditional Variance**. For every given value x, let $Var(Y|x)$ denote the variance of the conditional distribution of Y given that $X=x$. That is

$$Var(Y \mid x) = E\{[Y - E(Y \mid x)]^2 \mid x\}.$$

We call $Var(Y|x)$ ***the conditional variance of Y given X=x***.
Similarly, $Var(Y|X)$ is a function of X and is called the ***conditional variance of Y given X***.