Computer Science

CSC196: Analyzing Data

Mathematical Expectation

Jason Pacheco and Cesim Erten

Outline

- Mean of a Random Variable
- Variance & Covariance of a Random Variable
- Chebyshev’s Theorem

Outline

- Mean of a Random Variable
- Variance & Covariance of a Random Variable
- Chebyshev’s Theorem

Measures of Location

Recall the sample mean is the arithmetic mean of the data.

Suppose that the observations in a sample are $x_1, x_2, \ldots, x_n$. The **sample mean**, denoted by $\bar{x}$, is

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

We often refer to this as the <u>average</u> of the data.

Example: Sample Mean

Suppose we toss 2 fair coins a total of 15 times and observe:

- 0 heads 4 times
- 1 head 7 times
- 2 heads 5 times

Frequency of each event

$$(0)\left(\frac{4}{16}\right) + (1)\left(\frac{7}{16}\right) + (2)\left(\frac{5}{16}\right) = 1.06.$$

Value of each event

Notice we can never see 1.06 heads

What happens if we do this 1,000 times? 100,000 times?

Moments of RVs: Expected Value

Definition *The expected value of a RV* X*, denoted by* $\mathbf{E}[X]$*, is:*

$$\mu_X = \mathbf{E}[X] = \sum_x x\, P(X = x)$$

if X is discrete. If X is continuous then the expected value is given by:

$$\mu_X = \mathbf{E}[X] = \int x\, p(X = x)\, dx$$

Summation / integral over all values in domain of X

Example Let X be the sum of two fair dice, then:

$$\mathbf{E}[X] = 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{18} + \ldots + 12 \cdot \frac{1}{36} = 7$$

Example: Expected Value

Example 4.1: A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

***Solution*:** Let X represent the number of good components in the sample. The probability distribution of X is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \qquad x = 0, 1, 2, 3.$$

Simple calculations yield $f(0) = 1/35$, $f(1) = 12/35$, $f(2) = 18/35$, and $f(3) = 4/35$. Therefore,

$$\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1.7.$$

Thus, if a sample of size 3 is selected at random over and over again from a lot of 4 good components and 3 defective components, it will contain, on average, 1.7 good components.

Moments of RVs: Expected Value

Theorem (Linearity of Expectations) *If a and b are constants then,*

$$\mu_{aX+b} = \mathbf{E}[aX + b] = a\mathbf{E}[X] + b$$

Proof:

$$\begin{aligned}\mu_{aX+b} &= \int (ax + b)p(X = x)\,dx \\ &= a\int xp(X = x) + b\,dx \\ &= a\mathbf{E}[X] + b\end{aligned}$$

Corollary Setting a=0 we see that E[b]=b.

Corollary Setting b=0 we see that E[aX]=aE[X]

Moments of RVs: Expected Value

Theorem (Linearity of Expectations) *For any finite collection of discrete RVs* $X_1, X_2, \ldots, X_N$ *with finite expectations,*

$$\mathbf{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbf{E}[X_i]$$

Note This holds for any collection of random variables. They do not need to be independent.

Expectations of Functions

Definition *The expected value of a function* $g(X)$ *denoted by* $\mathbf{E}[g(X)]$ *, is:*

$$\mathbf{E}[g(X)] = \int g(x)\, p(X = x)\, dx$$

- The summation is over all values x
- Any function g(X) of a random variable X is also a random variable
- Therefore, any function g(X) has a probability distribution p(g(X)=g)

Example: Expected Value of a Function

Example 4.4: Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

x	4	5	6	7	8	9
$P(X = x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(X) = 2X - 1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

***Solution*:** By Theorem 4.1, the attendant can expect to receive

$$E[g(X)] = E(2X - 1) = \sum_{x=4}^{9} (2x - 1) f(x)$$

$$= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right)$$

$$+ (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right) = \$12.67.$$

Expected Value

Definition *Consider a function of two random variables denoted,*

$$g(X, Y)$$

The expected value is given by,

$$\mathbf{E}[g(X,Y)] = \sum_x \sum_y g(x,y)\, P(X = x, Y = y)$$

Example

Example 4.6: Let X and Y be the random variables with joint probability distribution indicated in Table 3.1 on page 96. Find the expected value of $g(X, Y) = XY$. The table is reprinted here for convenience.

	$f(x,y)$	$x = 0$	$x = 1$	$x = 2$	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

***Solution*:** By Definition 4.2, we write

$$E(XY) = \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x,y)$$

$$= (0)(0)f(0,0) + (0)(1)f(0,1)$$

$$+ (1)(0)f(1,0) + (1)(1)f(1,1) + (2)(0)f(2,0)$$

$$= f(1,1) = \frac{3}{14}.$$

Moments of RVs

Theorem: *If X and Y are independent then* $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$

Proof:

$$\mathbf{E}[XY] = \sum_x \sum_y (x \cdot y) p(X = x, Y = y)$$

$$= \sum_x \sum_y (x \cdot y) p(X = x) p(Y = y) \quad \textbf{(Independence)}$$

$$= \left(\sum_x x \cdot p(X = x) \right) \left(\sum_y y \cdot p(Y = y) \right) = \mathbf{E}[X]\mathbf{E}[Y] \quad \textbf{(Linearity of Expectation)}$$

Example *Let* $X_1, X_2 \in \{1, \ldots, 6\}$ *be RVs representing the result of rolling two fair standard die.* *What is the mean of their product?*

$$\mathbf{E}[X_1 X_2] = \mathbf{E}[X_1]\mathbf{E}[X_2] = 3.5^2$$

Outline

- Mean of a Random Variable
- Variance & Covariance of a Random Variable
- Chebyshev’s Theorem

Variance / Dispersion

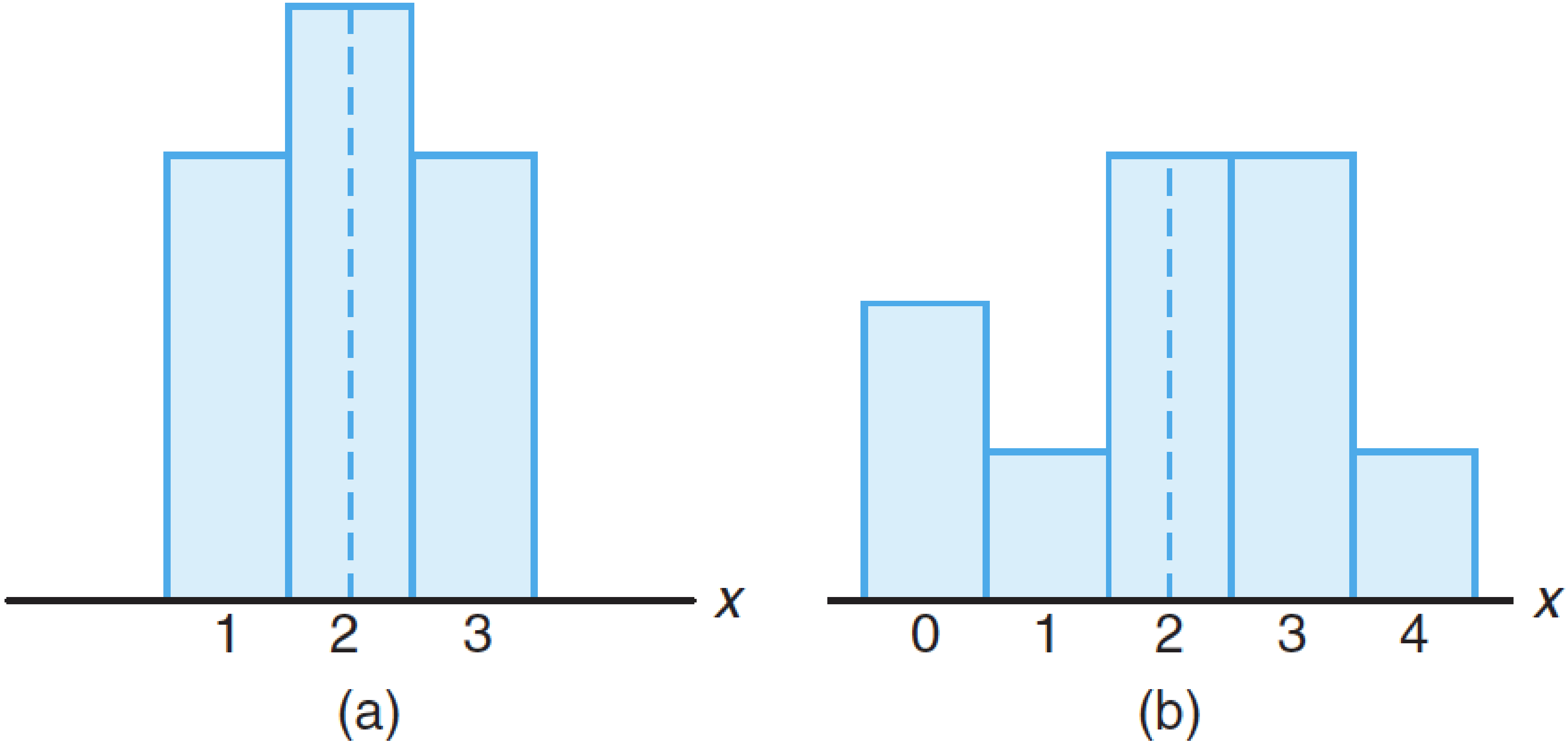


Figure 4.1: Distributions with equal means and unequal dispersions.

Moments of Discrete RVs: Variance

Definition *The variance of a RV X is defined as,*

$$\sigma_X^2 = \mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

(X-units)^2

First compute the mean,

$$\mu = \mathbf{E}[X] = \sum_x P(X = x) \cdot x$$

Then compute variance as,

$$\sigma_X^2 = \mathbf{Var}[X] = \sum_x (x - \mu)^2 \cdot P(X = x)$$

The standard deviation is $\sigma_X = \sqrt{\mathbf{Var}[X]}$ *.*

(X-units)

Example: Variance

Example 4.8: Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday. The probability distribution for company A [Figure 4.1(a)] is

x	1	2	3
$f(x)$	0.3	0.4	0.3

and that for company B [Figure 4.1(b)] is

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

Show that the variance of the probability distribution for company B is greater than that for company A.

Example: Variance

***Solution*:** For company A, we find that

$$\mu_A = E(X) = (1)(0.3) + (2)(0.4) + (3)(0.3) = 2.0,$$

and then

$$\sigma_A^2 = \sum_{x=1}^{3}(x-2)^2 = (1-2)^2(0.3) + (2-2)^2(0.4) + (3-2)^2(0.3) = 0.6.$$

For company B, we have

$$\mu_B = E(X) = (0)(0.2) + (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) = 2.0,$$

and then

$$\begin{aligned}\sigma_B^2 &= \sum_{x=0}^{4}(x-2)^2 f(x)\\ &= (0-2)^2(0.2) + (1-2)^2(0.1) + (2-2)^2(0.3)\\ &\quad + (3-2)^2(0.3) + (4-2)^2(0.1) = 1.6.\end{aligned}$$

Moments of Continuous RVs: Variance

Definition *The* *variance* *of a RV* X *is defined as,*

$$\sigma_X^2 = \mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

(X-units)^2

First compute the mean,

$$\mu = \mathbf{E}[X] = \int x\, p(X = x)\, dx$$

Then compute variance as,

$$\sigma_X^2 = \mathbf{Var}[X] = \int (x - \mu)^2\, P(X = x)\, dx$$

The *standard deviation* *is* $\sigma_X = \sqrt{\mathbf{Var}[X]}$.

(X-units)

Moments of RVs

Lemma An equivalent form of variance is:

$$\sigma_X^2 = \mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$$

Proof Keep in mind that $E[X]$ is a constant,

$$\mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2] \quad \textbf{(Distributive property)}$$

$$= \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + \mathbf{E}[X]^2 \quad \textbf{(Linearity of expectations)}$$

$$= \mathbf{E}[X^2] - \mathbf{E}[X]^2 \quad \textbf{(Algebra)}$$

Example: Variance

Example 4.9: Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Using Theorem 4.2, calculate σ^2.

***Solution*:** First, we compute

$$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$$

Now,

$$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87.$$

Therefore,

$$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$$

Variance of a Function

Definition *The variance of a function* $g(X)$ *denoted by* $\mathbf{Var}[g(X)]$ *is:*

$$\sigma^2_{g(X)} = \mathbf{Var}[g(X)] = \sum_x (g(x) - \mathbf{E}[g(X)])^2 \cdot P(X = x)$$

Example 4.11: Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

Example: Variance of a Function

Solution: First, we find the mean of the random variable $2X+3$. According to Theorem 4.1,

$$\mu_{2X+3} = E(2X+3) = \sum_{x=0}^{3} (2x+3) f(x) = 6.$$

Now, using Theorem 4.3, we have

$$\sigma^2_{2X+3} = E\{[(2X+3) - \mu_{2x+3}]^2\} = E[(2X+3-6)^2]$$

$$= E(4X^2 - 12X + 9) = \sum_{x=0}^{3} (4x^2 - 12x + 9) f(x) = 4.$$

Moments of RVs: Covariance

Definition *The* *covariance* *of two RVs* X *and* Y *is defined as,*

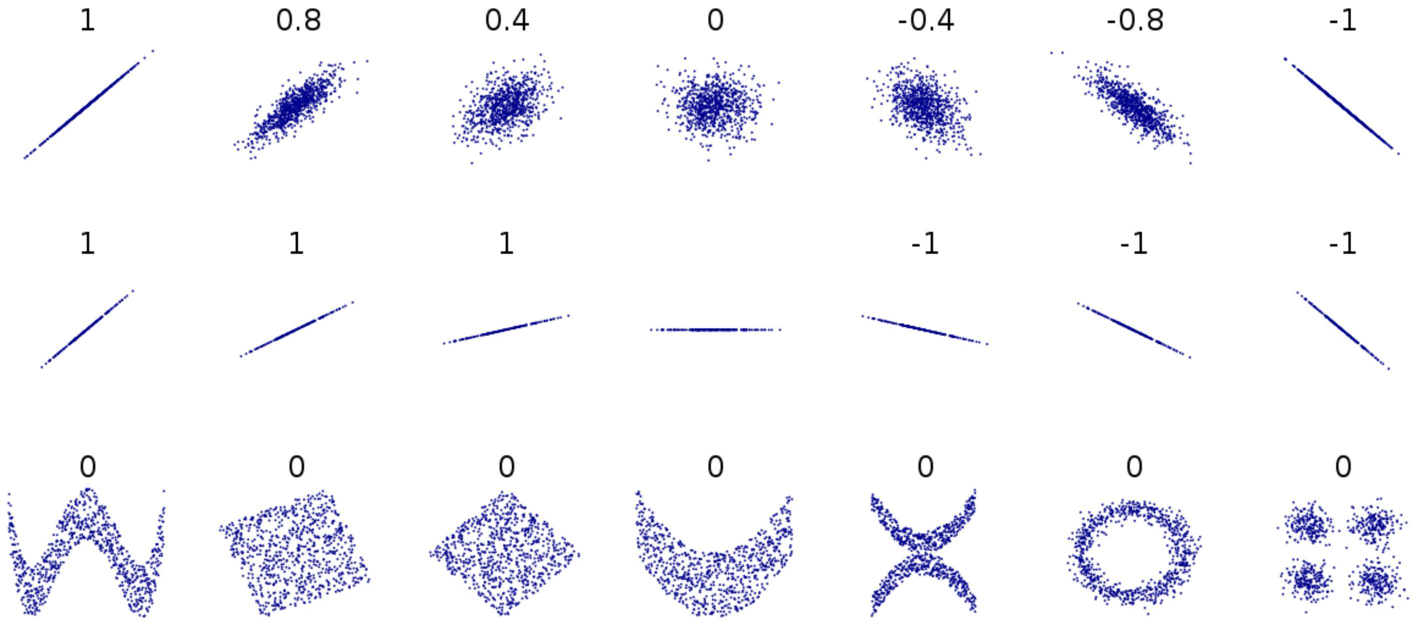
$$\sigma_{XY} = \mathrm{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

- Measures the relation between two random variables
- The *sign* of Cov(X,Y) indicates whether their relationship is positive or negative
- If X and Y are *independent* then Cov(X,Y)=0
- Cov(X,Y)=0 **does not** mean X and Y are independent

Correlation

Definition *The correlation of two RVs X and Y is given by,*

$$\mathbf{Corr}(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sigma_X \sigma_Y} \quad \textit{where} \quad \sigma_X = \sqrt{\mathbf{Var}(X)}$$



Like covariance, only expresses linear relationships!

Moments of RVs

Lemma *For any two RVs* X *and* Y *,*

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$$

e.g. variance is not a linear operator.

Proof $$\mathbf{Var}[X+Y] = \mathbf{E}[(X+Y-\mathbf{E}[X+Y])^2]$$

(Linearity of expectation) $$= \mathbf{E}[(X+Y-\mathbf{E}[X]-\mathbf{E}[Y])^2]$$

(Distributive property) $$= \mathbf{E}[(X-\mathbf{E}[X])^2 + (Y-\mathbf{E}[Y])^2 + 2(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$$

(Linearity of expectation) $$= \mathbf{E}[(X-\mathbf{E}[X])^2] + \mathbf{E}[(Y-\mathbf{E}[Y])^2] + 2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$$

(Definition of Var / Cov) $$= \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$$

Moments of RVs

Question: *What is the variance of the sum of independent RVs*

$$\mathbf{Var}[X_1 + X_2] = \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{Cov}(X_1, X_2)$$
$$= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{E}[(X_1 - \mathbf{E}[X_1])(X_2 - \mathbf{E}[X_2])]$$
$$= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{E}[(X_1 - \mathbf{E}[X_1])]\mathbf{E}[(X_2 - \mathbf{E}[X_2])]$$
$$= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\left(\mathbf{E}[X_1] - \mathbf{E}[X_1]\right)\left(\mathbf{E}[X_2] - \mathbf{E}[X_2]\right)$$
$$= \mathbf{Var}[X_1] + \mathbf{Var}[X_2]$$

E.g. variance is a *linear operator* for independent RVs

Theorem: *If* $X \perp Y$ *then* $\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$

Corollary: *If* $X \perp Y$ *then* $\mathbf{Cov}(X, Y) = 0$

Moments of RVs: Covariance

Theorem *The covariance of two RVs X and Y is given by,*

$$\sigma_{XY} = \mathrm{Cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

***Proof*:** For the discrete case, we can write

$$\begin{aligned}\sigma_{XY} &= \sum_x \sum_y (x - \mu_X)(y - \mu_Y) f(x, y) \\ &= \sum_x \sum_y xy f(x, y) - \mu_X \sum_x \sum_y y f(x, y) \\ &\quad - \mu_Y \sum_x \sum_y x f(x, y) + \mu_X \mu_Y \sum_x \sum_y f(x, y).\end{aligned}$$

Since

$$\mu_X = \sum_x x f(x, y), \quad \mu_Y = \sum_y y f(x, y), \text{ and } \sum_x \sum_y f(x, y) = 1$$

for any joint discrete distribution, it follows that

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y = E(XY) - \mu_X \mu_Y.$$

For the continuous case, the proof is identical with summations replaced by integrals. ∎

Example: Covariance

Example 4.13: Example 3.14 on page 95 describes a situation involving the number of blue refills X and the number of red refills Y. Two refills for a ballpoint pen are selected at random from a certain box, and the following is the joint probability distribution:

	$f(x, y)$	x = 0	x = 1	x = 2	$h(y)$
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
	$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Find the covariance of X and Y.

Example: Covariance

***Solution*:** From Example 4.6, we see that $E(XY) = 3/14$. Now

$$\mu_X = \sum_{x=0}^{2} xg(x) = (0)\left(\frac{5}{14}\right) + (1)\left(\frac{15}{28}\right) + (2)\left(\frac{3}{28}\right) = \frac{3}{4},$$

and

$$\mu_Y = \sum_{y=0}^{2} yh(y) = (0)\left(\frac{15}{28}\right) + (1)\left(\frac{3}{7}\right) + (2)\left(\frac{1}{28}\right) = \frac{1}{2}.$$

Therefore,

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{3}{14} - \left(\frac{3}{4}\right)\left(\frac{1}{2}\right) = -\frac{9}{56}.$$

Variance of a Linear Function

Theorem *Let X and Y be random variables. The variance of the function aX + bY + c is given by,*

$$\sigma^2_{aX+bY+c} = \mathbf{Var}(aX + bY + c) = a^2\mathbf{Var}(X) + b^2\mathbf{Var}(Y) + 2ab\mathbf{Cov}(X, Y)$$

Corollary *Setting b=0 we see that,*

$$\mathbf{Var}(aX + c) = a^2\mathbf{Var}(X)$$

Corollary *Setting b=0 and c=0 we see that,*

$$\mathbf{Var}(aX) = a^2\mathbf{Var}(X)$$

So adding a constant c does not change the variance.

Example

Example 4.22: If X and Y are random variables with variances $\sigma_X^2 = 2$ and $\sigma_Y^2 = 4$ and covariance $\sigma_{XY} = -2$, find the variance of the random variable $Z = 3X - 4Y + 8$.

***Solution*:**

$$\sigma_Z^2 = \sigma_{3X-4Y+8}^2 = \sigma_{3X-4Y}^2 \quad \text{(by Corollary 4.6)}$$

$$= 9\sigma_X^2 + 16\sigma_Y^2 - 24\sigma_{XY} \quad \text{(by Theorem 4.9)}$$

$$= (9)(2) + (16)(4) - (24)(-2) = 130.$$

Outline

- Mean of a Random Variable
- Variance & Covariance of a Random Variable
- Chebyshev’s Theorem

An Intuition About Variance

We expect a random variable with smaller variance to be more concentrated around the mean…

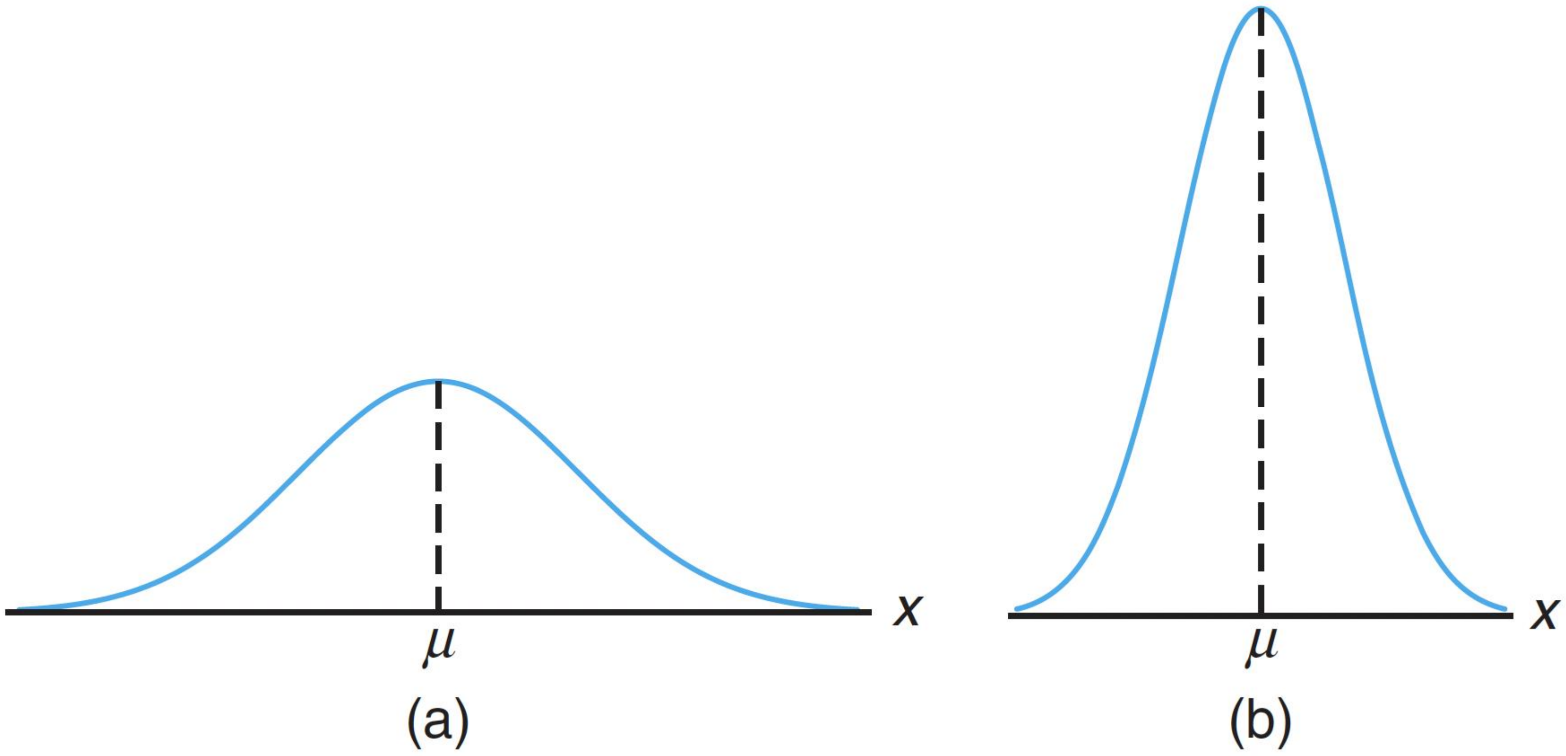


Figure 4.2: Variability of continuous observations about the mean.

Chebyshev's Theorem

Theorem 4.10: **(Chebyshev's Theorem)** The probability that any random variable X will assume a value within k standard deviations of the mean is at least $1 - 1/k^2$. That is,

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

- E.g. for k=2 the RV X has a probability of **at least** $1-1/2^2=3/4$ of falling within 2 standard deviations of the mean
- I.e. ¾ or more of the observations of **any distribution** lie in the interval $\mu \pm 2\sigma$

Example: Chebyshev's Theorem

Example 4.27: A random variable X has a mean $\mu = 8$, a variance $\sigma^2 = 9$, and an unknown probability distribution. Find

(a) $P(-4 < X < 20)$,

(b) $P(|X - 8| \geq 6)$.

Solution: (a) $P(-4 < X < 20) = P[8 - (4)(3) < X < 8 + (4)(3)] \geq \frac{15}{16}$.

(b) $P(|X - 8| \geq 6) = 1 - P(|X - 8| < 6) = 1 - P(-6 < X - 8 < 6)$

$$= 1 - P[8 - (2)(3) < X < 8 + (2)(3)] \leq \frac{1}{4}.$$

Chebyshev's theorem makes no assumptions about the distribution p(X), so the bounds can be rather weak.