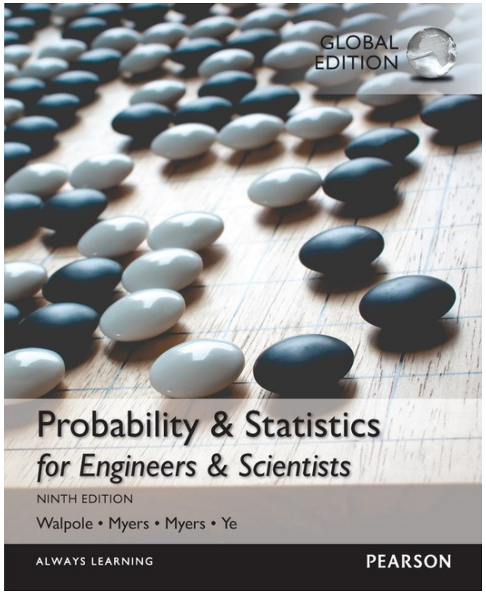
Chapter 4

Mathematical Expectation – part 2

School of Computing, Gachon Univ.
Joon Yoo



Outline

- Mean of a Random Variable
- Variance and **Covariance of Random Variables**
- Means and Variances of Linear Combinations of Random Variables
- Chebyshev's Theorem (advanced)

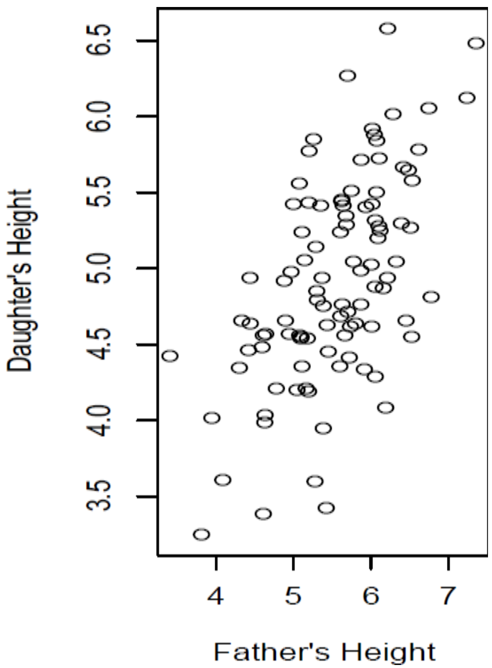
4.2 Variance and Covariance of Random Variables (2)

Question

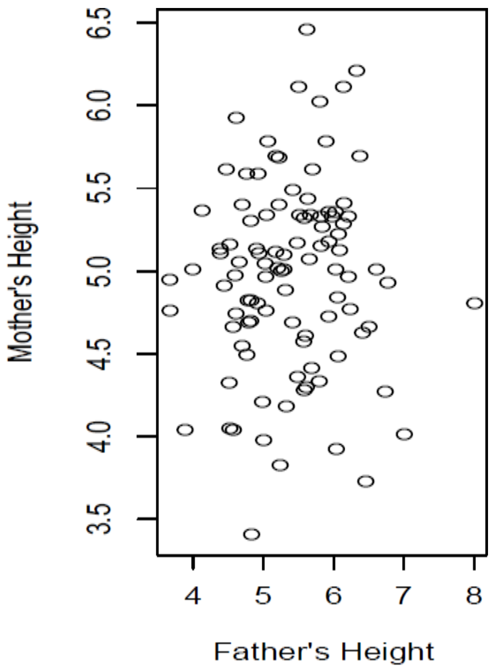
Does there exist a quantity to measure how much two variables change together or provide a measure of the strength of the **correlation** between two random variables?

Example

- Let X denote the father's height, Y the daughter's height, and Z the mother's height.



Cov(X,Y) = 0.3



Cov(X,Z) = 0.02

Covariance

- a measure of how much two random variables change together

Covariance

Definition 4.2:

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

If $g(X,Y) = (X - \mu_X)(Y - \mu_Y)$, where $\mu_X = E(X)$ and $\mu_Y = E(Y)$, Definition 4.2 yields an expected value called the **covariance** of X and Y, which we denote by σ_{XY} or $\mathrm{Cov}(X,Y)$.

- Definition 4.4:

Let X and Y be random variables with joint probability distribution $f(x,y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_y) f(x,y)$$

if X and Y are discrete, and

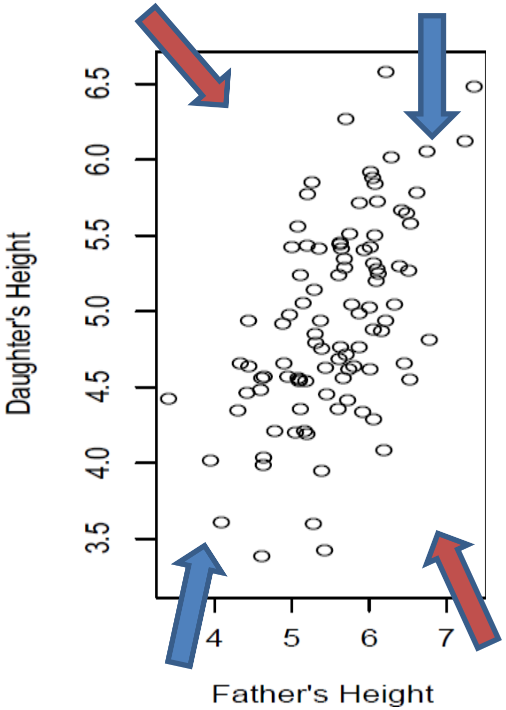
$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_y) f(x,y)\, dx\, dy$$

if X and Y are continuous.

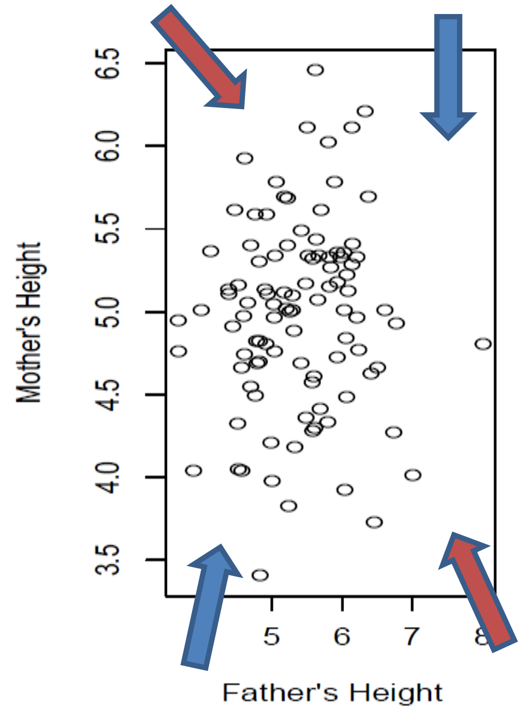
Wireless Intelligent Networking LAB

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_X)(y-\mu_y) f(x,y)$$

$f(x,y)$



$f(x,y)$



The probability that the values of X and Y are large together, or small together is high!

GWINL Wireless Intelligent Networking LAB

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_X)(y-\mu_y) f(x,y)$$

- The covariance between two random variables is
  - a measurement of the nature of the association between the two.
- If large values of X often result in large values of Y or small values of X result in small values of Y, positive $X-\mu_X$ will often result in positive $Y-\mu_Y$ and negative $X-\mu_X$ will often result in negative $Y-\mu_Y$.
- Thus, the product $(X-\mu_X)(Y-\mu_Y)$ will tend to be positive.
- On the other hand,
  → for tendency of (large X, small Y), (small X, large Y) → the product $(X-\mu_X)(Y-\mu_Y)$ will be negative

- **Theorem 4.4:**

The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

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

- **proof.** For the discrete case, we can write

Definition 4.4:
$$\begin{aligned}\sigma_{XY} &= \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x,y) \\ &= \sum_x \sum_y xy f(x,y) - \mu_X \sum_x \sum_y y f(x,y) \\ &\quad - \mu_Y \sum_x \sum_y x f(x,y) + \mu_X \mu_Y \sum_x \sum_y f(x,y).\end{aligned}$$

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. ∎

GWINL Wireless Intelligent Networking LAB

- ## Example 4.14
  - The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x,y)=\begin{cases}8xy, & 0\le y\le x\le 1,\\ 0, & \text{elsewhere.}\end{cases}$$

Find the covariance of X and Y.

  - Hint:
    - To compute the means, first compute the marginal density functions

$$f(x,y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

- Solution
  - We first compute the marginal density functions:

- DEFINITION 3.10.

$$\boxed{g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx}$$

$$g(x) = \begin{cases} 4x^3, & 0 < x < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

$$h(y) = \begin{cases} 4y(1-y^2), & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

$$\mu_X = E(X) = \int_0^1 4x^4 dx = \frac{4}{5};$$

$$\mu_Y = \int_0^1 4y^2(1-y^2)dy = \frac{8}{15}.$$

$$E(XY) = \int_0^1 \int_y^1 8x^2y^2 dx dy = \frac{4}{9}$$

Theorem 4.4:

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

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{4}{9} - \frac{4}{5}\cdot\frac{8}{15} = \frac{4}{225}.$$

Definition 4.2:

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

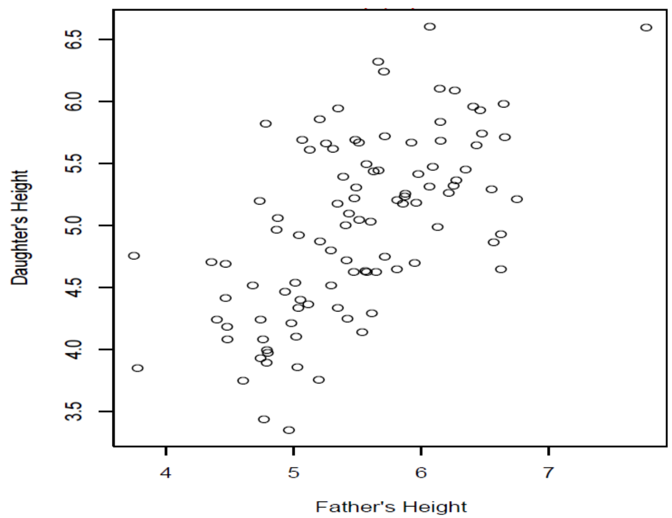
Wireless Intelligent Networking LAB

Question

- A larger value of the covariance between two random variables means closer correlation ?

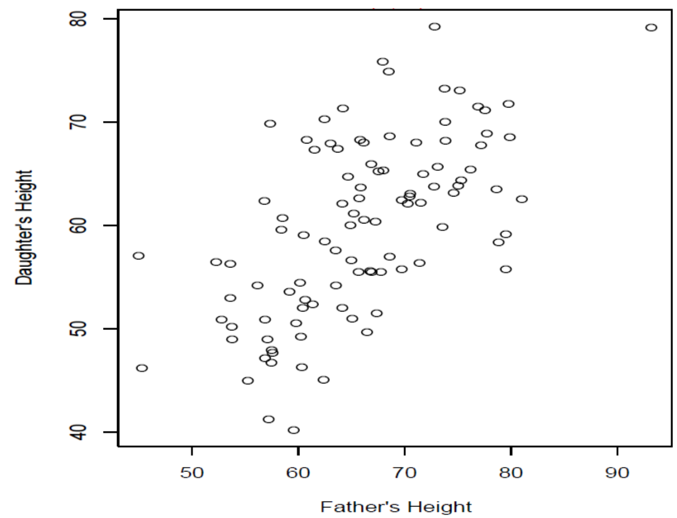
Example

- Let X denote the father's height, Y the daughter's height, and Z the mother's height.



Unit: feet

Cov(X,Y) = 0.3



Unit: inch

Cov(X,Y) = 30

Length

170 Centimeter = 5.57743 Foot

Length

170 Centimeter = 66.9291 Inch

- Although the covariance between two random variables does provide information regarding the nature of the relationship,
  the magnitude of σ_{XY} *does* ***not*** *indicate anything regarding the strength of the relationship*, since σ_{XY} is not scale-free. Its magnitude will depend on the units used to measure both X and Y .

- There is a scale-free version of the covariance called the **correlation coefficient** that is used widely in statistics.

Correlation Coefficient

- Definition 4.5.

Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient of X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

- **Notice:**
  - $-1 \leq \rho_{XY} \leq 1$
  - When $Y = a + bX$, then (an exact linear dependency)
    - If $b > 0$, $\rho_{XY} = 1$
    - If $b < 0$, $\rho_{XY} = -1$

- Example 4.16
  - Find the correlation coefficient of *X* and *Y in Example 4.14*

$$f(x,y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{4}{9} - \frac{4}{5} \cdot \frac{8}{15} = \frac{4}{225}.$$

Because

$$E(X^2) = \int_0^1 4x^5 \, dx = \frac{2}{3} \text{ and } E(Y^2) = \int_0^1 4y^3(1-y^2) \, dy = 1 - \frac{2}{3} = \frac{1}{3},$$

Theorem 4.2:

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

we conclude that

$$\sigma_X^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75} \text{ and } \sigma_Y^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}.$$

Hence,

Definition 4.5:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\rho_{XY} = \frac{4/225}{\sqrt{(2/75)(11/225)}} = \frac{4}{\sqrt{66}}.$$

GWINL Wireless Intelligent Networking LAB

4.3 Means and Variances of Linear Combinations of Random Variables

- Theorem 4.5

If a and b are constants, then

$$E(aX + b) = aE(X) + b.$$

- proof

$$\begin{aligned} E(aX+b) &= \int_{-\infty}^{\infty} (ax+b) f(x) dx \\ &= a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx \\ &= aE(X) + b. \end{aligned}$$

- Corollary 4.1

Setting $a = 0$, we see that $E(b) = b$.

- Corollary 4.2

Setting $b = 0$, we see that $E(aX) = aE(X)$.

- Example 4.17
  - Let X be a random variable with density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2; \\ 0, & \text{elsewhere}; \end{cases}$$

  - Find the expected value of **g**(**X**) = 4**X** +3.

- Solution 1 (Example 4.5)

Definition 4.2:

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

$$E(4X+3) = \int_{-1}^{2}(4x+3)\frac{x^2}{3}dx = \frac{1}{3}\int_{-1}^{2}(4x^3+3x^2)dx = 8$$

- Solution 2 (Example 4.17)

Theorem 4.5:

$$\boxed{E(aX+b) = aE(X)+b.}$$

$$E(X) = \int_{-1}^{2} x\left(\frac{x^2}{3}\right)dx = \int_{-1}^{2}\frac{x^3}{3}\ dx = \frac{5}{4}.$$

$$E(4X+3) = (4)\left(\frac{5}{4}\right)+3 = 8$$

- Theorem 4.6
  - The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is,

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)]$$

  - Proof.

$$\begin{aligned} E[g(X) \pm h(X)] &= \int_{-\infty}^{\infty} (g(x) \pm h(x)) f(x) dx \\ &= \int_{-\infty}^{\infty} g(x) f(x) dx \pm \int_{-\infty}^{\infty} h(x) f(x) dx \\ &= E[g(X)] \pm E[h(X)]. \end{aligned}$$

- Example 4.19
  - Let X be a random variable with probability distribution as follows:

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

  - Find the expected value of $Y = (X - 1)^2$.

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

- Solution:

Applying Theorem 4.6 to the function $Y = (X-1)^2$, we can write

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

From Corollary 4.1, $E(1) = 1$, and by direct computation,

$$E(X) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (2)(0) + (3)\left(\frac{1}{6}\right) = 1 \text{ and}$$

$$E(X^2) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (4)(0) + (9)\left(\frac{1}{6}\right) = 2.$$

Hence,

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

GWINL Wireless Intelligent Networking LAB

- ## Theorem 4.7
  - The expected value of the sum or difference of two or more functions of random variables X and Y is the sum or difference of the expected values of the functions. That is,

$$E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)]$$

  - Proof.

$$\begin{aligned} E[g(X,Y) \pm h(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(g(x,y) \pm h(x,y))f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy \\ &\pm \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)dxdy \\ &= E[g(X,Y)] \pm E[h(X,Y)]. \end{aligned}$$

- Theorem 4.8
  - Let X and Y be two **independent** random variables. Then

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

  - Proof.

Suppose that $g(x)$ and $h(y)$ are the marginal distribution of X and Y, respectively. Since X and Y are independent, we may write $f(x, y) = g(x)h(y)$. By definition,

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyg(x)h(y)dxdy \\ &= \int_{-\infty}^{\infty} xg(x)dx \int_{-\infty}^{\infty} yh(y)dy \\ &= E(X)E(Y). \end{aligned}$$

Definition 3.12:

- statistically independent

$$\boxed{f(x, y) = g(x)h(y)}$$

Wireless Intelligent Networking LAB

Illustrative Example

Theorem 4.8: $E(XY) = E(X)E(Y).$

Theorem 4.8 can be illustrated for discrete variables by considering the experiment of tossing a **green die** and a **red die**. Let the **random variable X** represent the outcome on the **green die** and the **random variable Y** represent the outcome on the **red die**. Then **XY** represents the product of the numbers that occur on the pair of dice. In the long run, the **average of the products of the numbers** is **equal to the product** of the average number that occurs on the **green die** and the average number that occurs on the **red die**.

Question

- (Corollary 4.5)
  - Let X and Y be two independent random variables. Then σ_{XY} = ?

- **Theorem 4.4:**

The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

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

- Theorem 4.8
  - Let X and Y be two **independent** random variables. Then

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

Wireless Intelligent Networking LAB

- Example 4.21
  - Given the joint density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2, 0 < y < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

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

- Sol1.

$$E(XY) = \int_0^1 \int_0^2 \frac{x^2 y(1+3y^2)}{4}\, dx dy = \frac{5}{6}$$

$$E(X) = \frac{4}{3}, \text{ and } E(Y) = \frac{5}{8}.$$

Hence,

$$E(X)E(Y) = \left(\frac{4}{3}\right)\left(\frac{5}{8}\right) = \frac{5}{6} = E(XY)$$

- Sol2.
  - $g(x) = \frac{x}{2}$, $0 < x < 2$.
  - $h(y) = \frac{1+3y^2}{2}$, $0 < y < 1$.
  - $f(x,y) = g(x)h(y)$ for all x and y

So, X and Y are independent.

Wireless Intelligent Networking LAB

- **Theorem 4.12**
  - If a and b are constants, then

$$\sigma^2_{aX+b} = \text{Var}(aX+b) = a^2\text{Var}(X)$$

- Proof.

$$\begin{aligned}
\text{Var}(aX+b) &= E[(aX+b-\mu_{aX+b})^2] \\
&= E[(aX+b-(a\mu_X+b))^2] \\
&= E[(aX-a\mu_X)^2] \\
&= E[a^2\cdot(X-\mu_X)^2] \\
&= a^2E[(X-\mu_X)^2] = a^2\text{Var}(X).
\end{aligned}$$

Wireless Intelligent Networking LAB

- Example (4.12+)
  - Let X be a random variable with density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2; \\ 0, & \text{elsewhere}; \end{cases}$$

  - Find the variance of **g**(**X**) = 4**X** +3. Use Var(X) = 1/2.

- Sol)

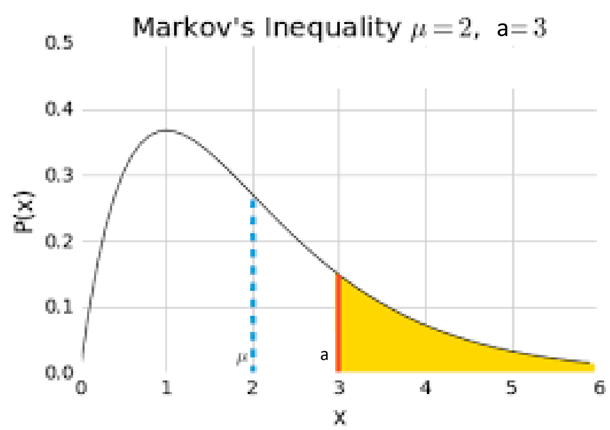
$$\text{Var}(4X + 3) = 4^2\text{Var}(X) = 16 \cdot \frac{1}{2} = 8.$$

4.4 Chebyshev's Theorem

Markov inequality

- Upper bound of tail distribution
  - Markov's inequality gives an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant.
  - If X is a random variable that takes only nonnegative values ($X \geq 0$), then for any value $a > 0$

$$P(X \geq a) \leq \frac{E(X)}{a}$$



Markov inequality

- Proof

$\boxed{X \geq 0}$

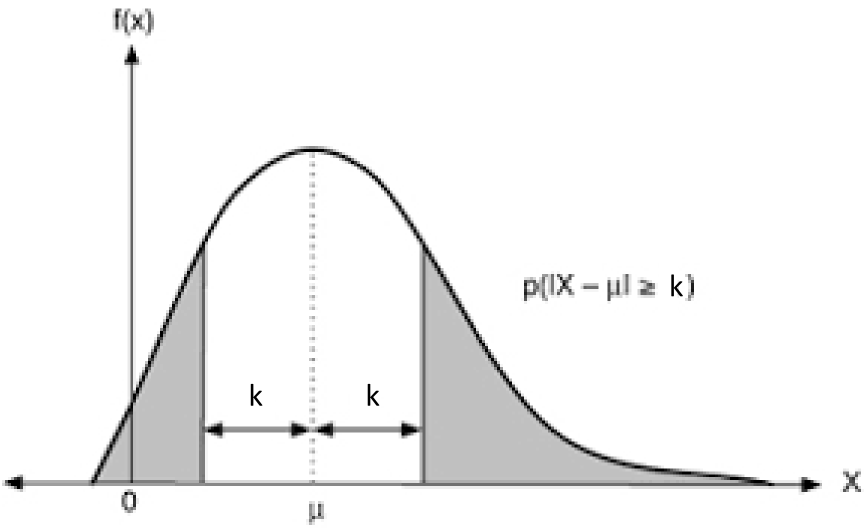
$$
\begin{aligned}
E[X] &= \int_0^\infty xf(x)\,dx \\
&= \int_0^a xf(x)\,dx + \int_a^\infty xf(x)\,dx \\
&\geq \int_a^\infty xf(x)\,dx \\
&\geq \int_a^\infty af(x)\,dx \\
&= a\int_a^\infty f(x)\,dx \\
&= aP\{X \geq a\} \quad \Longrightarrow \quad P(X \geq a) \leq \frac{E(X)}{a}
\end{aligned}
$$

WINL Wireless Intelligent Networking LAB

Chebyshev's inequality

- Upper bound of tail distribution
  - If X is a random variable with mean μ and variance σ^2, then for any value k>0

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$



Chebyshev's inequality

- Proof

$$\boxed{P(X \geq a) \leq \frac{E(X)}{a}}$$

Since $(X-\mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X-\mu)^2 \geq k^2\} \leq \frac{E[(X-\mu)^2]}{k^2}$$

But since $(X-\mu)^2 \geq k^2$ if and only if $|X-\mu| \geq k$,

$$P\{|X-\mu| \geq k\} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

Example

- Problem
  - Suppose that it is known that the number of items produced in a factory during a week is a random variable with **mean 50**.

(a) What can be said about the probability that this week's production will exceed 75?

By Markov's inequality $P(X \geq a) \leq \dfrac{E(X)}{a}$

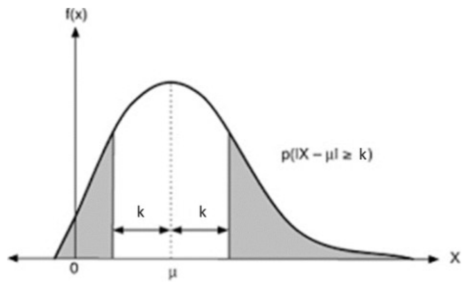
$$P\{X > 75\} \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

Example

(b) If the **variance** of a week's production is known to equal **25**, then what can be said about the probability that this week's production will be between 40 and 60?

By Chebyshev's inequality $P\{|X-\mu| \geq k\} \leq \dfrac{\sigma^2}{k^2}$

$$P\{|X-50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$



$$P\{|X-50| < 10\} \geq 1 - \frac{1}{4} = \frac{3}{4}$$

- The importance of Markov's and Chebyshev's inequalities
  - Enable us to ***derive bounds*** on probabilities when (i) only **the mean**, or (ii) both the **mean and the variance** are known.
  - Of course, if the actual distribution were known, then the desired probabilities could be exactly computed

GWINL Wireless Intelligent Networking LAB

https://www.psycom.net/bipolar-questions-answers