PHOENICIA UNIVERSITY

Mathematical Expectations of Random Variables

Probability and Statistics in Engineering — GENG 207, Chapter 4

College of Engineering | Dr. Chireen Saghir | Spring 2025–2026

1 Chapter Overview

1. **Expected value (mean) of a random variable** — *the discrete and continuous cases*
2. **Mean of a function of a random variable, g(X)** — *and of a joint distribution g(X, Y)*
3. **Variance and covariance** — *measuring spread and joint variability*
4. **Correlation coefficient** — *a scale-free measure of linear association*
5. **Means and variances of linear combinations of random variables** — *the properties used throughout the rest of the course*
6. **Chebyshev's Rule** — *a distribution-free bound on probability*

2 Expected Value of a Discrete Random Variable

MOTIVATION

Recall from Chapter 1: for the 10 data values 10, 10, 12, 12, 13, 13, 10, 10, 15, 16, the sample mean can be written using relative frequencies (4/10, 2/10, 2/10, 1/10, 1/10) instead of individual values:

$$\bar{x} = 10(4/10) + 12(2/10) + 13(2/10) + 15(1/10) + 16(1/10)$$

The same idea defines the expected value of a random variable X: replace relative frequencies with probabilities.

EXAMPLE — TOSSING A FAIR COIN TWICE

Let X count the number of heads in two tosses of a fair coin, so $X \in \{0, 1, 2\}$ with $S = \{HH, HT, TH, TT\}$. The probability mass function is:

$$f(0) = 1/4, \quad f(1) = 1/2, \quad f(2) = 1/4$$

These probabilities behave exactly like the relative frequencies above — which motivates the definition of the mean of a discrete random variable.

$$\mu = E(X) = \sum_x x \cdot f(x)$$

3 Expected Value of a Continuous Random Variable

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Note: the mean of X need not itself be a value X can take, and it need not be an integer.

EXAMPLE 1

An inspector samples 3 components at random from a batch of 7, where 4 of the 7 are working and 3 are defective. Find the expected number of working components in the sample of 3.

This is solved with the hypergeometric distribution from Chapter 3 — compute $E(X) = \Sigma_x\, x \cdot f(x)$ directly from $f(x)$.

EXAMPLE 2

The lifetime X (in hours) of a certain electronic device has probability density function:

$$f(x) = 20{,}000 / x^3$$

for $x > 100$, and $f(x) = 0$ elsewhere. Find the expected lifetime of this type of device.

4 Worked Examples — Expected Value

EXAMPLE — EXPECTED COMMISSION

A sales rep has two client meetings today. She estimates a 70% chance of closing a $1,000 deal at the first meeting, and (independently) a 40% chance of closing a $1,500 deal at the second. Find her expected commission for the day.

Expected commission = 0.70(1000) + 0.40(1500) = $700 + $600 = $1,300.

EXAMPLE — EXPECTED LIFETIME (CONT.)

Using the density from the previous slide, $f(x) = 20{,}000/x^3$ for $x > 100$:

$$E(X) = \int_{100}^{\infty} x\cdot(20{,}000/x^3)\,dx = 200 \text{ hours}$$

So although X can take any value above 100 hours, its mean lifetime works out to exactly 200 hours — illustrating that the mean need not be a “typical” single outcome.

5 Mean of a Function of a Random Variable

Sometimes we care about a function $g(X)$ of a random variable X rather than X itself. Because $g(X)$ is built from X, it is itself a new random variable.

EXAMPLE

Let X be the number of phones a store sells on a given day, and define the revenue function $g(X) = 500X$ (dollars). Then $g(X)$ is a new random variable derived from X.

If $X = 2$, then $g(2) = \$1{,}000$. In general, more than one value of X can map to the same value of g — which is exactly why $g(X)$ needs its own probability distribution.

To find $E[g(X)]$, we do not need to derive the full distribution of $g(X)$ first — we can work directly from the distribution of X, as the next examples show.

6 Finding the Distribution of g(X)

EXAMPLE

Let X be a discrete random variable with values $x = -1, 0, 1, 2$, and define $g(X) = X^2$. Find the probability distribution of g.

Since $g(-1) = 1$, $g(0) = 0$, $g(1) = 1$, $g(2) = 4$, the value $g = 1$ can arise from two different values of X, so its probabilities combine:

$$P[g(X)=0] = f(0) \qquad P[g(X)=1] = f(-1) + f(1) \qquad P[g(X)=4] = f(2)$$

g(x)	0	1	4
P[g(X)=g(x)]	f(0)	f(−1)+f(1)	f(2)

A RELATED EXAMPLE

Let X be discrete with $f(k) = 1/5$ for $k = -1, 0, 1, 2, 3$, and let $Y = 2|X|$. Then Y takes values 0, 2, 4, 6, and by the same reasoning:

$$f_\gamma(0)=1/5,\ \ f_\gamma(2)=f^x(-1)+f^x(1)=2/5,\ \ f_\gamma(4)=f^x(2)=1/5,\ \ f_\gamma(6)=f^x(3)=1/5$$

7

The Expected Value of g(X)

$$\mu_{g(X)} = E[g(X)] = \Sigma_x\, g(x)\cdot f(x) \quad \textit{(discrete)}$$

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x)\cdot f(x)\,dx \quad \textit{(continuous)}$$

EXAMPLE — CAR WASH TIPS

The number of cars X passing through a car wash between 4 and 5 p.m. on a sunny Friday has the distribution below. Let $g(X) = 2X-1$ be the attendant's tip (in \$). Find the attendant's expected tip.

EXAMPLE — CONTINUOUS G(X)

Let X have density $f(x) = x^2/3$ for $-1 < x < 2$ (0 elsewhere). Find $E[g(X)]$ for $g(X) = 4X + 3$.

x	4	5	6	7	8	9
P(X=x)	1/12	1/12	1/4	1/4	1/6	1/6

$\rightarrow E[g(X)] = 2\cdot E(X)-1$

8

Mean of a Joint Distribution

Let X and Y be random variables with joint distribution $f(x,y)$. The mean of $g(X,Y)$ is defined the same way, summing (or integrating) over both variables:

$$\mu_{g(X,Y)} = E[g(X,Y)] = \Sigma_x \Sigma_y \, g(x,y) \cdot f(x,y) \quad \textit{(discrete)}$$

$$\mu_{g(X,Y)} = E[g(X,Y)] = \iint g(x,y) \cdot f(x,y) \, dx \, dy \quad \textit{(continuous)}$$

EXAMPLE — RECALL THE JOINT TABLE FROM CHAPTER 3

f(x,y)	x=0	x=1	x=2
y=0	3/28	9/28	3/28
y=1	3/14	3/14	0
y=2	1/28	0	0

Find $E[g(X,Y)]$ for $g(X,Y) = XY$.

9 Two Ways to Find E(X) from a Joint Distribution

Suppose we have a joint distribution f(X,Y) and want the expected value of the X random variable alone. There are two equivalent routes:

1. Find the marginal distribution

Sum out Y to get g(x), the marginal distribution of X, then apply the single-variable formula:

$$\mu_X = E(X) = \Sigma_x\, x \cdot g(x)$$

2. Work directly from the joint distribution

Define g(X)=X and apply the joint-distribution formula directly — no marginal needed:

$$\mu_{g(X)} = E[g(X)] = \Sigma_x \Sigma_y\, g(x) f(x,y)$$

Both routes give exactly the same value of E(X) — and the same equivalence holds in the continuous case, replacing sums with integrals.

10 Variance and Covariance of a Random Variable

The mean alone does not fully describe a distribution's shape — we also need to capture how spread out its values are around that mean.

(a) small spread

(b) large spread

Two distributions with the same mean but different dispersion around it

The best measure of this variability is the variance of X, denoted Var(X), σ^2a̧, or σ^2:

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$ *(discrete)*

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$ *(continuous)* *— the standard deviation σ is the positive square root*

11

A More Convenient Variance Formula

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

This is algebraically identical to the defining formula on the previous slide, but it is almost always faster to compute in practice.

EXAMPLE

Weekly demand X for a bottled-water brand at a small chain of convenience stores, measured in thousands of liters, is a continuous random variable with density:

$$f(x) = 2(x-1) \text{ for } 1 < x < 2, \quad f(x) = 0 \text{ elsewhere.}$$

Find the mean and variance of X.

Approach: compute $\mu = E(X) = \int x \cdot f(x)dx$, then $E(X^2) = \int x^2 \cdot f(x)dx$ over $1<x<2$, and finally $\sigma^2 = E(X^2) - \mu^2$.

12

Variance of a Function g(X)

$$\sigma^2_{g(X)} = E\{[g(X)-\mu_{g(X)}]^2\} = \sum_x [g(x)-\mu_{g(X)}]^2 f(x) \quad \textit{(discrete)}$$

$$\sigma^2_{g(X)} = \int_{-\infty}^{\infty} [g(x)-\mu_{g(X)}]^2 f(x)\,dx \quad \textit{(continuous)}$$

EXAMPLE 1

Find the variance of $g(X) = 2X + 3$, where X has the discrete distribution below.

x	0	1	2	3
f(x)	1/4	1/8	1/2	1/8

EXAMPLE 2

Let X have the density $f(x) = x^2/3$ for $-1 < x < 2$ (used earlier). Find the variance of $g(X) = 4X + 3$.

Approach: find $\mu_{g(X)} = E(4X+3)$, then apply the shortcut $\sigma^2_{g(X)} = E[g(X)^2] - [\mu_{g(X)}]^2$ using the continuous formula above.

13 Covariance of Two Random Variables

Covariance measures the nature of the association between X and Y — how variation in one relates to variation in the other.

Positive covariance

Large values of X tend to occur with large values of Y (and small with small), so $(X-\mu X)(Y-\mu Y)$ is usually positive — the variables move together.

Negative covariance

Large values of X tend to occur with small values of Y, so the product $(X-\mu X)(Y-\mu Y)$ is usually negative — the variables move in opposite directions.

Note: if X and Y are statistically independent, their covariance is zero — but the converse is not true; a zero covariance can still hide a non-linear relationship.

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)]$$

$= \Sigma_x \Sigma_y (x-\mu X)(y-\mu Y) f(x,y)$ if discrete, or the double integral if continuous.

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

14

Covariance — Properties and Examples

To prove:

- $Cov(X, X) > 0$
- $Cov(X, -X) < 0$
- $Cov(rX, sY) = r \cdot s \cdot Cov(X, Y)$
- Two statistically independent variables have zero covariance

EXAMPLE — PEN REFILLS

A box holds a mix of blue-ink and red-ink pen refills. Two are drawn at random. Let X = number of blue refills drawn and Y = number of red refills drawn, with the joint distribution recalled from Chapter 3. Find $Cov(X, Y)$.

EXAMPLE — MARATHON FINISHERS

The fractions X and Y of male and female finishers across a set of marathon races follow the joint density $f(x,y) = 8xy$ for $0 \le y \le x \le 1$ (0 elsewhere). Find $Cov(X, Y)$.

Both examples reduce to $\sigma_{XY} = E(XY) - \mu_X\mu_Y$ once the marginals (or the joint sums/integrals) are worked out.

15 Correlation Coefficient Between Two Variables

Covariance is not scale-free — it depends on the units of both variables. The correlation coefficient rescales it so it no longer depends on those units:

$$\rho_{XY} = \sigma_{XY} / (\sigma_X \sigma_Y)$$

The correlation coefficient is unitless, which is why it is preferred over covariance when comparing the strength of association across different pairs of variables.

Properties

- $-1 \leq \rho \leq 1$
- If $Y = a + bX$ (a perfect linear relationship): $\rho = 1$ when $b > 0$, and $\rho = -1$ when $b < 0$
- The closer $|\rho|$ is to 1, the stronger the linear relationship between X and Y

Example: find the correlation coefficient for the pen-refill and marathon examples from the previous slide.

16 Means of Linear Combinations of Random Variables

These properties (true for both discrete and continuous random variables) simplify mean and variance calculations used throughout the rest of the course. Let a, b, c be constants.

1. $E(aX + b) = a \cdot E(X) + b$ $\rightarrow$ if $b=0$: $E(aX)=a \cdot E(X)$ if $a=0$: $E(b)=b$

   Example: if $E(X) = 41/6$, find $E(2X-1) = 2(41/6) - 1 = 68/6$.

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

   Example: let X have f(x): $f(0)=1/3$, $f(1)=1/2$, $f(2)=0$, $f(3)=1/6$. Find $E[Y]$ for $Y = (X-1)^2$.

   Example: the weekly demand (in thousands of liters) for a drink is $g(X)=X^2+X-2$, where X has density $f(x)=2(x-1)$ for $1<x<2$. Find the expected weekly demand.

3. For two random variables X and Y with joint distribution $f(x,y)$:

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

   and, if X and Y are independent: $E(XY) = E(X)E(Y)$, and therefore $\sigma XY = 0$.

17

Variance of Linear Combinations

EXAMPLE

In a chip-fabrication process, let X be the gallium-to-arsenide ratio and Y the number of usable wafers produced in an hour — independent random variables with joint density $f(x,y) = x(1+3y^2)/4$ for $0<x<2$, $0<y<1$. Show that $E(XY) = E(X)E(Y)$.

4. If X, Y jointly distributed, a, b, c constants: $\sigma^2_{aX+bY+c} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}$

Setting b=0 gives $\sigma^2 aX+c=a^2\sigma^2$; setting a=1,b=0 gives $\sigma^2 X+c=\sigma^2$; setting b=0,c=0 gives $\sigma^2 aX=a^2\sigma^2$ — special cases of the same identity.

5–7. If X, Y are independent: $\sigma^2_{aX\pm bY} = a^2\sigma^2_X + b^2\sigma^2_Y$ *(covariance term vanishes)*

8. If $X_1,\ldots,X_n$ independent: $\sigma^2_{a_1X_1+\cdots+a_nX_n} = a_1^2\sigma^2_{X_1} + \cdots + a_n^2\sigma^2_{X_n}$

Example: if $\sigma^2 X=2$, $\sigma^2 Y=4$, and $\sigma XY=-2$, find the variance of $Z = 3X-4Y+8$.

These properties only hold for linear combinations. A non-linear function Y = g(X) needs its own methods and approximations, covered separately.

Chebyshev's Rule

If a random variable has a small variance, we expect most of its values to cluster tightly around the mean; a larger variance spreads values out more widely around the same mean.

(a) large spread *(b) small spread*

Values cluster more tightly around μ when the variance is small

If you know the mean and standard deviation, Chebyshev's Rule tells you the minimum proportion of values guaranteed to fall within k standard deviations of the mean — for any distribution at all:

Chebyshev's Theorem

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - 1/k^2$$

Key Takeaways

- The expected value E(X) generalizes the sample mean, replacing relative frequencies with probabilities; it need not be a value X can actually take.
- The same summation/integration idea extends directly to functions g(X), joint functions g(X,Y), variance, and covariance — they are all expectations of some function of the data.
- $\sigma^2 = E(X^2) - \mu^2$ is almost always the fastest way to compute variance in practice.
- Covariance captures the direction of a linear association; correlation ρ rescales it to a unitless value between -1 and 1.
- Independence guarantees zero covariance, but zero covariance does not guarantee independence.
- The linear-combination identities (properties 1–8) only apply to linear functions of random variables — non-linear g(X) needs separate treatment.
- Chebyshev's Rule works for any distribution, but only gives a (sometimes weak) lower bound — useful precisely when the distribution itself is unknown.