# Chapter 5 Part 2

STAT 5700: Probability

Prof. Katie Fitzgerald, PhD

Fall 2025

## Contents

## 5.6 & 5.7 Review

Recall the Theorems from 5.6.

**Example**:
Let $Y_1, Y_2, Y_3$ be three continuous random variables with joint pdf $f(y_1, y_2, y_3)$. Find an expression for $E(2Y_1 + 3Y_2^2 - Y_3)$

**Example**: Let $Y_1, Y_2, Y_3$ be three independent random variables with binomial distributions $b(4, 1/2), b(6, 1/3)$, and $b(12, 1/6)$ respectively. Find $E(Y_1Y_2Y_3)$

**Example**: Let X, Y be two independent random variables. Show that $Cov(X, Y) = 0$.

**Recall:** *independence* $\implies Cov(X, Y) = 0$, *but the converse is NOT true. It is possible to have* $Cov(X, Y) = 0$ *when* $X$ *and* $Y$ *are dependent.*

## 5.8 Expected value and variance of linear functions of r.v's

**Theorem 5.12** Let $Y_1, Y_2, \dots Y_n$ and $X_1, X_2, \dots X_m$ be random variables with $E(Y_i) = \mu_i$ and $E(X_j) = \xi_j$. Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{and} \quad U_2 = \sum_{j=1}^{m} b_j X_j$$

for constants $a_1, a_2, \dots a_n$ and $b_1, b_2, \dots b_m$. Then the following hold:

(a) $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$
(b) $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum \sum_{1 \le i \le j \le n} a_i a_j Cov(Y_i, Y_j)$, where the double sum os over all pairs $(i, j)$ with $i < j$
(c) $Cov(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(Y_i, X_j)$

Note, when $n = 2$, (b) is simply $V(a_1Y_1 + a_2Y_2) = a_1^2 V(Y_1) + a_2^2 V(Y_2) + 2a_1a_2Cov(Y_1, Y_2)$

**Example**: Let $Y_1$ and $Y_2$ be random variables with means $\mu_1$, $\mu_2$, variances $\sigma_1^2$, $\sigma_2^2$, and covariance $\sigma_{12}$ respectively. Suppose $U = 2Y_1 + 3Y_2$. Find $E(U)$ and $V(U)$.

**Example (cont'd)**: Return to the above example where $Y_1, Y_2, Y_3$ are three independent random variables with binomial distributions $b(4, 1/2), b(6, 1/3)$, and $b(12, 1/6)$ respectively. Find the mean and variance of $W = Y_1 + Y_2 + Y_3$.

**Example:**
Let the independent random variables $X_1$ and $X_2$ have respective means $\mu_1 = -4$ and $\mu_2 = 3$ and variances $\sigma_1^2 = 4$ and $\sigma_2^2 = 9$. Find the mean and variance of $Y = 3X_1 - 2X_2$.

If $X_3$ is a third random variable that has the same mean and variance as $X_2$ but is correlated with $X_1$ by the amount $\rho = 0.3$, what is the mean and variance of $W = 3X_1 - 2X_3$?

**Example** In the 5.2 gasoline example, we were interested in $X$ and $Y$ that had the joint and marginal pdfs below:

$$f(x,y) = 3x, \quad 0 \leq y \leq x \leq 1$$

$$f_x(x) = 3x^2, \quad 0 \leq x \leq 1$$

$$f_y(y) = 3/2(1-y^2), \quad 0 \leq y \leq 1$$

Supposed we are interested in $W = X - Y$, which in context represents the proportional amount of gasoline remaining at the end of the week. Find the mean and variance of $W$.

### Sample statistics

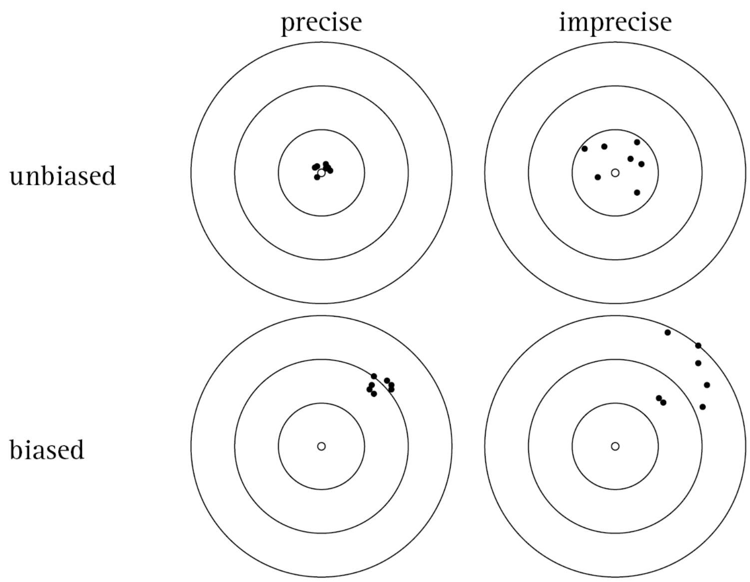
Any function of the sample observations $X_1, X_2, \ldots X_n$ is called a **statistic**
The **sample mean** of a random sample $X_1, X_2, \ldots X_n$ is given by

$$\overline{X} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

and the **sample variance** is given by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$\overline{X}$ and $S^2$ are two examples of **statistics**. We will see that $\overline{X}$ is an **unbiased estimator** for the population mean $\mu$ and $S^2$ is an **unbiased estimator** for the population variance $\sigma^2$




When $X_1, X_2, \ldots X_n$ are independent random variables with mean $\mu$ and variance $\sigma^2$, find the mean and variance of $\overline{X}$.

## 5.9 The Multinomial Probability Distribution

Recall a binomial random variable results from an experiment with $n$ trials, with two possible outcomes per trial (success/failure). Sometimes, we're interested in situations where there are more than 2 outcomes per trial. For example, there are at least 4 possible blood types a person can have.

A **multinomial experiment** is a generalization of the binomial experiment.

A **multinomial experiment** possesses the following properties:

1. The experiment consists of $n$ identical trials
2. The outcome of each trial falls into one of $k$ classes or cells
3. THe probability that the outcomes of a single trial falls into cell $i$ is $p_i$, $i = 1, 2, \dots k$ and remains the same from trial to trial. Notice $p_1 + p_2 + \dots p_k = 1$
4. The trials are independent
5. The random variables of interest are $Y_1, Y_2, \dots Y_k$ where $Y_i$ equals the number of trials for which the outcome falls into cell $i$. Notice that $Y_1 + Y_2 + \cdots + Y_k = n$

The joint pmf for $Y_1, Y_2, \dots Y_k$ is given by

$$p(y_1, y_2, \dots y_k) = \frac{n!}{y_1! y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k},$$

where $\sum_{i=1}^{k} p_i = 1$ and $\sum_{i=1}^{k} y_i = n$

According to recent census figures, the proportions of adults (persons over 18 years of age) in the United States associated with five age catgories are as given in the following table.

| Age | Proportion |
|---|---|
| 18-24 | 0.18 |
| 25 - 34 | 0.23 |
| 35 - 44 | 0.16 |
| 45 - 64 | 0.27 |
| 65+ | 0.16 |

If these figures are accurate and five adults are randomly sampled, find the probability that the sample contains one person between ages of 18 and 24, two between the ages of 25 and 34, and two between the ages of 45 and 64.

**Theorem 5.13** If $Y_1, Y_2, \ldots Y_k$ have a multinomial distribution with parameters $n$ and $p_1, p_2, \ldots p_k$, then

1. $E(Y_i) = np_i$
2. $V(Y_i) = np_i q_i$
3. $Cov(Y_s, Y_t) = -np_s p_t$ if $s \neq t$

## 5.11 Conditional Expectation

Conditional distributions can be used to computed probabilities and expected values just as with any distribution.

**Definition 5.13** If $X$ and $Y$ are any two random variables, the *conditional expectation of $g(X)$ given that $Y = y$, is defined to be

$$E(g(X)|Y = y) = \int_{-\infty}^{\infty} g(x)f(x|y)dx$$

if $X$ and $Y$ are jointly continuous, and

$$E(g(X)|Y = y) = \sum_{x} g(x)p(x|y)$$

if $X$ and $Y$ are jointly discrete.

Refer back to our example in 5.3 with $f(x, y) = 1/2, 0 \leq x \leq y \leq 2$, where we found $f(x|y) = 1/y, \quad 0 < x \leq y$. Find the conditional expectation of the amount of sales, $X$, given that $Y = 1.5$.

**Theorem 5.14** Let $X$ and $Y$ denote random variables. Then

$$E(X) = E[E(X|Y)],$$

where on the right-hand side the inner expectation is with respect to the conditional distribution of $X|Y$ and the outside expectation is with respect to the distribution of $Y$.

PROOF

The conditional variance of $X$ given $Y$ is defined by

$$V(X|Y = y) = E(X^2|Y = y) - [E(X|Y = y)]^2$$

**Theorem 5.15** Let $X$ and $Y$ denote random variables. Then

$$V(X) = E[V(X|Y)] + V[E(X|Y)]$$

PROOF:

A quality control plan for an assembly line involves sampling $n = 10$ finished items per day and counting $Y$, the number of defectives. If $p$ denotes the probability of observing a defective, then $Y$ has a binomial distribution, assuming that a large number of items are produced by the line. But $p$ varies from day to day and is assumed to have a uniform distribution on the interval from 0 to 1/4. Find the expected value and variance of $Y$.

# Chapter 5 Group Work

### Problem 1

Show that for any constants $a$ and $b$, $Cov(a + X, b + Y) = Cov(X, Y)$. That is, shifting by a constant does not change the covariance. *Note: this fact will be useful on HW 5.110*

### Problem 2

To estimate a proportion of units that meet a given criteria, we often use the estimator $\hat{p} = \frac{Y}{n}$, where $Y \sim binomial(n, p)$. Find the expected value and variance of $\hat{p}$, assuming n is fixed.

### Problem 3

A learning experiment requires a rat to run a maze (a network of pathways) until it locates one of three possible exits. Exit 1 presents a reward of food, but exits 2 and 3 do not. (If the rat eventually selects exit 1 almost every time, learning may have taken place). Let $Y_i$ denote the number of times exit $i$ is chosen in successive runnings. For the following, assume that the rate chooses an exit at random on each run.

a. Find the probability that $n = 6$ runs result in $Y_1 = 3$, $Y_2 = 1$ and $Y_3 = 2$.
b. For general $n$, find $E(Y_1)$ and $V(Y_1)$.
c. For general $n$, find $Cov(Y_2, Y_3)$
d. To check for the fat's preference between exits 2 and 3, we may look at $Y_2 - Y_3$. Find $E(Y_2 - Y_3)$ and $V(Y_2 - Y_3)$ for general $n$.

### Problem 4

The number of defects per yard in a certain fabric, $Y$, is known to have a Poisson distribution with parameter $\lambda$. The parameter $\lambda$ is assumed to be a random variable with a pdf given by

$$f(\lambda) = e^{-\lambda}, \quad \lambda \geq 0$$

a. Find the expected number of defects per yard by first finding the conditional expectation of $Y$ for given $\lambda$.
b. Find the variance of $Y$.