

Lecture #4: Probability Spaces & Discrete Random Variables

MATH 321 – Mathematical Statistics I

Torin Quinlivan

Knox College
tquinlivan@knox.edu

October 17, 2026



Review: Axioms of Probability

Definition: A **probability space** is a triple $(\Omega, \mathcal{F}, P)$, where Ω is the **sample space**, $\mathcal{F}$ is the **event space**, and P is a **probability law (or measure)**.

Review: Axioms of Probability

Definition: A **probability space** is a triple $(\Omega, \mathcal{F}, P)$, where Ω is the **sample space**, $\mathcal{F}$ is the **event space**, and P is a **probability law (or measure)**.

Definition: A **probability law** is a function $P : \mathcal{F} \rightarrow [0, 1]$ that maps an event A to a real number in $[0, 1]$. The function must satisfy the **axioms of probability**:

1. **Non-Negativity:** $P(A) \geq 0$, for any $A \in \mathcal{F}$.
2. **Normalization:** $P(\Omega) = 1$.
3. **Additivity:** For any pairwise disjoint events $\{A_1, A_2, A_3, \dots\}$, it must be true that

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Review: Important Properties and Results

Corollary: Let $A \in \mathcal{F}$ be an event. Then:

1. $P(\overline{A}) = 1 - P(A)$.
2. $P(A) \leq 1$.
3. $P(\emptyset) = 0$.

Review: Important Properties and Results

Corollary: Let $A \in \mathcal{F}$ be an event. Then:

1. $P(\overline{A}) = 1 - P(A)$.
2. $P(A) \leq 1$.
3. $P(\emptyset) = 0$.

Corollary (Unions of Two Non-Disjoint Sets): For any A and B in $\mathcal{F}$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Review: Important Properties and Results

Corollary (Inequalities): Let A and B be two events in $\mathcal{F}$. Then:

1. $P(A \cup B) \leq P(A) + P(B)$. (Union Bound)
2. If $A \subseteq B$, then $P(A) \leq P(B)$.

Review: Important Properties and Results

Corollary (Inequalities): Let A and B be two events in $\mathcal{F}$. Then:

1. $P(A \cup B) \leq P(A) + P(B)$. (Union Bound)
2. If $A \subseteq B$, then $P(A) \leq P(B)$.

Theorem (Bayes' Theorem): For any two events A and B such that $P(A) > 0$ and $P(B) > 0$,

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}.$$

Review: Important Properties and Results

Theorem (Law of Total Probability): Let $\{A_1, A_2, A_3, \dots\}$ be a partition of Ω , i.e., $A_1, A_2, A_3, \dots$ are disjoint and $\Omega = A_1 \cup A_2 \cup \dots$. Then, for any $B \in \mathcal{F}$,

$$P(B) = \sum_{i \in I} P(B | A_i)P(A_i).$$

Review: Important Properties and Results

Theorem (Law of Total Probability): Let $\{A_1, A_2, A_3, \dots\}$ be a partition of Ω , i.e., $A_1, A_2, A_3, \dots$ are disjoint and $\Omega = A_1 \cup A_2 \cup \dots$. Then, for any $B \in \mathcal{F}$,

$$P(B) = \sum_{i \in I} P(B | A_i) P(A_i).$$

Corollary: Let $\{A_1, A_2, A_3, \dots\}$ be a partition of Ω , i.e., $A_1, A_2, A_3, \dots$ are disjoint and $\Omega = A_1 \cup A_2 \cup \dots$. Then, for any $B \in \mathcal{F}$,

$$P(A_j | B) = \frac{P(B | A_j) P(A_j)}{\sum_{i \in I} P(B | A_i) P(A_i)}.$$

Prove You're a Probability Space!

Example: Let S be a non-empty *finite* set. Then $(S, \mathcal{P}(S), P)$ is a probability space, where $P(A) = \frac{|A|}{|S|}$ for any $A \in \mathcal{P}(S)$.

Prove You're a Probability Space!

Example: Let S be a non-empty *finite* set. Then $(S, \mathcal{P}(S), P)$ is a probability space, where $P(A) = \frac{|A|}{|S|}$ for any $A \in \mathcal{P}(S)$.

Example: Let $A \subset \Omega$, where $(\Omega, \mathcal{F}, P)$ is a probability space and $P(A) > 0$. Show $\hat{P}(B) = \frac{P(B \cap A)}{P(A)}$ defines a probability measure on $\mathcal{F}$.

Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Definition: A **random variable** X is a function $X : \Omega \rightarrow \mathbb{R}$ that maps an outcome $\xi \in \Omega$ to a number $X(\xi)$ on the real line.

Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Definition: A **random variable** X is a function $X : \Omega \rightarrow \mathbb{R}$ that maps an outcome $\xi \in \Omega$ to a number $X(\xi)$ on the real line.

Definition: A random variable X is said to be **discrete** if it can assume only a finite or countably infinite number of distinct values.

Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Definition: A **random variable** X is a function $X : \Omega \rightarrow \mathbb{R}$ that maps an outcome $\xi \in \Omega$ to a number $X(\xi)$ on the real line.

Definition: A random variable X is said to be **discrete** if it can assume only a finite or countably infinite number of distinct values.

Example: Let Ω consist of the outcomes of tossing a (fair) coin three times and let $X : \Omega \rightarrow \mathbb{R}$ be the number of heads that appears.

Probability Mass Functions

Definition: The **probability mass function (PMF)** of a random variable X is a function which specifies the probability of obtaining a number $X(\xi) = x$. We denote a probability mass function as:

$$p_X(x) = P(X = x).$$

The set of all possible states of X is denoted as $X(\Omega)$.

Probability Mass Functions

Definition: The **probability mass function (PMF)** of a random variable X is a function which specifies the probability of obtaining a number $X(\xi) = x$. We denote a probability mass function as:

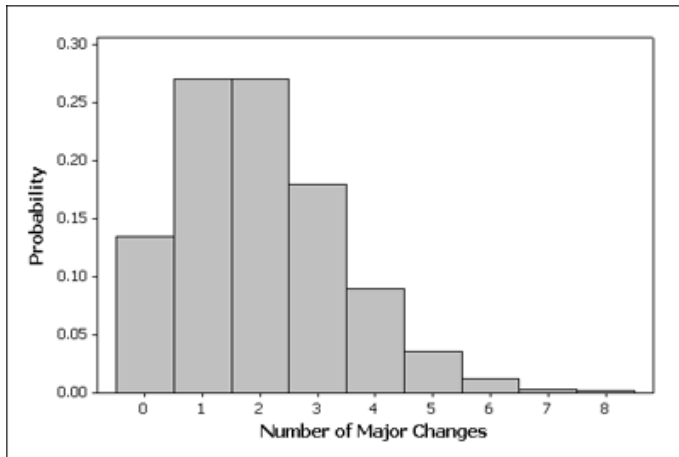
$$p_X(x) = P(X = x).$$

The set of all possible states of X is denoted as $X(\Omega)$.

Example: Let Ω consist of the outcomes of tossing a (fair) coin three times and let $X : \Omega \rightarrow \mathbb{R}$ be the number of heads that appears. What is the probability mass function (PMF) of this random variable?

Properties of Probability Mass Functions

Random variables have a graphical representation called a **probability histogram**.



Properties of Probability Mass Functions

Theorem: The function $Q(A) = \sum_{x \in A} p_X(x)$, defined for $A \subseteq X(\Omega)$, is a probability law on $X(\Omega)$.

Can you prove this?

The Expected Value of a Random Variable

Definition: Let X be a discrete random variable with the probability mass function $p_X(x)$. Then the **expected value (or mean)** of X , $E(X)$, is defined to be:

$$E(X) = \sum_x x p_X(x).$$

If $p_X(x)$ is an accurate characterization of the population frequency distribution, then $E(X) = \mu$, the population mean.

The Expected Value of a Function of a Random Variable

Definition: Let X be a discrete random variable with probability mass function $p_X(x)$ and $g(X)$ be a real-valued function of X . Then the expected value of $g(X)$ is given by:

$$E(g(X)) = \sum_x g(x) p_X(x).$$

Moments of a Random Variable

A specific case of an expected value of a function of a random variable are the **moments** of the random variable.

Moments of a Random Variable

A specific case of an expected value of a function of a random variable are the **moments** of the random variable.

The n^{th} **moment** of a random variable X is defined as:

$$E(X^n) = \sum_x x^n p_X(x).$$

Moments of a Random Variable

A specific case of an expected value of a function of a random variable are the **moments** of the random variable.

The n^{th} **moment** of a random variable X is defined as:

$$E(X^n) = \sum_x x^n p_X(x).$$

The n^{th} **central moment** of a random variable X is defined as:

$$E((X - \mu)^n) = \sum_x (x - \mu)^n p_X(x),$$

where $\mu = E(X)$.

Properties of the Expected Value

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and c be a constant. Then $E(c) = c$.

Properties of the Expected Value

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and c be a constant. Then $E(c) = c$.

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$, $g(X)$ be a function of X , and c be a constant. Then:

$$E[c \times g(X)] = c \times E[g(X)].$$

Properties of the Expected Value

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and c be a constant. Then $E(c) = c$.

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$, $g(X)$ be a function of X , and c be a constant. Then:

$$E[c \times g(X)] = c \times E[g(X)].$$

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and $g_1(X), g_2(X), \dots, g_k(X)$ be k functions of X . Then:

$$E\left[\sum_{i=1}^k g_i(X)\right] = \sum_{i=1}^k E[g_i(X)].$$

Variance of a Random Variable

$E(X)$ gives the mean of the population with distribution given by $p_X(x)$. Now let's define the variance and standard deviation of this population.

Variance of a Random Variable

$E(X)$ gives the mean of the population with distribution given by $p_X(x)$. Now let's define the variance and standard deviation of this population.

Definition: The **variance** of a set of measurements is the average of the square of the differences between the values in a set of measurements and their mean, denoted as σ^2 .

Variance of a Random Variable

$E(X)$ gives the mean of the population with distribution given by $p_X(x)$. Now let's define the variance and standard deviation of this population.

Definition: The **variance** of a set of measurements is the average of the square of the differences between the values in a set of measurements and their mean, denoted as σ^2 .

Definition: The **standard deviation** of X is the positive square root of the variance, denoted as σ .

Formulas for the Variance

Definition: If X is a random variable with mean $\mu = E(X)$, the **variance** of a random variable X is defined to be the expected value of $(X - \mu)^2$. That is,

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

Formulas for the Variance

Definition: If X is a random variable with mean $\mu = E(X)$, the **variance** of a random variable X is defined to be the expected value of $(X - \mu)^2$. That is,

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

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and finite mean $E(X) = \mu$; then:

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

Formulas for the Variance

Definition: If X is a random variable with mean $\mu = E(X)$, the **variance** of a random variable X is defined to be the expected value of $(X - \mu)^2$. That is,

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

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and finite mean $E(X) = \mu$; then:

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

This second formula is a lot easier to use!

Properties of the Variance

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and c be a constant. Then $V(c) = 0$.

Properties of the Variance

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and c be a constant. Then $V(c) = 0$.

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and c be a constant. Then:

$$V[c \times X] = c^2 \times V[X].$$

Properties of the Variance

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and c be a constant. Then $V(c) = 0$.

Theorem: Let X be a discrete random variable with probability mass function $p_X(x)$ and c be a constant. Then:

$$V[c \times X] = c^2 \times V[X].$$

Theorem: Let $X_1, X_2, X_3, \dots$ be discrete independent random variables with probability mass functions $p_{X_1}(x), p_{X_2}(x), p_{X_3}(x), \dots$. Then:

$$V \left[\sum_{i=1}^k X_i \right] = \sum_{i=1}^k V[X_i].$$