# Math 431 - Intro to Probability
# Comprehensive Review Notes

Prof. Claude

Spring 2024

## Probability Spaces and Basic Properties

We begin by defining the fundamental concept of a probability space, which provides a rigorous mathematical foundation for studying random phenomena.

**Definition 1.** *A* ***probability space*** *is a triple* $(\Omega, \mathcal{F}, P)$ *where:*

- $\Omega$ *is the sample space*
- $\mathcal{F}$ *is a* $\sigma$*-algebra of subsets of* $\Omega$ *(the events)*
- $P$ *is a probability measure on* $(\Omega, \mathcal{F})$*, satisfying: (i)* $P(\Omega) = 1$*, (ii)* $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ *for disjoint* $A_i \in \mathcal{F}$*.*

*The axioms of probability ensure that probability behaves in a consistent and intuitive manner. They form the basis for deriving many important properties and theorems.*

**Property 1** (Probability Axioms)**.** *For events* $A, B \in \mathcal{F}$*:*

1. $P(A^c) = 1 - P(A)$
2. *If* $B \subset A$ *then* $P(B) \leq P(A)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

### Discrete Probability Spaces

In discrete probability spaces, the sample space $\Omega$ is countable, and probability is assigned to individual outcomes using a probability mass function.

**Definition 2.** *For a discrete sample space $\Omega$, a **probability mass function** is a function $p : \Omega \to [0, 1]$ satisfying $\sum_{x \in \Omega} p(x) = 1$. Then $P(A) = \sum_{x \in A} p(x)$.*

### Continuous Probability Spaces

In continuous probability spaces, the sample space $\Omega$ is uncountable, and probability is assigned using a probability density function.

**Definition 3.** *A random variable $X$ has a **probability density function** $f$ if $f \geq 0$, $\int_{-\infty}^{\infty} f(x)\, dx = 1$, and $P(a \leq X \leq b) = \int_a^b f(x)\, dx$.*

*The probability density function allows us to compute probabilities for continuous random variables by integrating over intervals. Note that for continuous random variables, $P(X = x) = 0$ for any specific value $x$.*

## Conditional Probability and Independence

Conditional probability allows us to update probabilities based on new information, while independence describes situations where knowledge of one event does not affect the probability of another.

**Definition 4.** *The **conditional probability** of $A$ given $B$ is*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \text{ when } P(B) > 0$$

*Conditional probability is a fundamental concept that enables us to reason about probabilities in the presence of additional information. It forms the basis for important results like the multiplication rule, law of total probability, and Bayes' theorem.*

**Theorem 1** (Multiplication Rule)**.**

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_2 \cap A_1) \ldots P(A_n \mid A_{n-1} \cap \cdots \cap A_1)$$

**Theorem 2** (Law of Total Probability)**.** *If $\{B_1, \ldots, B_n\}$ is a partition of $\Omega$ with $P(B_i) > 0$ for all $i$, then*

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$$

**Theorem 3** (Bayes' Theorem)**.**

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)}$$

*Bayes' theorem is a powerful result that allows us to update probabilities based on observed evidence. It is widely used in inference and decision-making problems.*

**Definition 5.** *Events $A$ and $B$ are **independent** if $P(A \cap B) = P(A)P(B)$.*

*Independence is a crucial concept in probability theory. It simplifies many calculations and is a fundamental assumption in various statistical models.*

## Important Discrete Distributions

Discrete probability distributions describe the probabilities of outcomes in a countable sample space. Here are some commonly encountered discrete distributions:

- **Bernoulli($p$)**: Models a single trial with binary outcomes. $P(X = 1) = p$, $P(X = 0) = 1 - p$.
- **Binomial($n, p$)**: Models the number of successes in a fixed number of independent Bernoulli trials. $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$.
- **Geometric($p$)**: Models the number of trials until the first success. $P(X = k) = (1-p)^{k-1} p$, $k = 1, 2, \ldots$.
- **Negative Binomial($r, p$)**: Models the number of trials until the $r$-th success. $P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$, $k = r, r+1, \ldots$.
- **Hypergeometric($N, K, n$)**: Models the number of successes in a fixed number of draws without replacement. $P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$, $\max(0, n - N + K) \leq k \leq \min(n, K)$.
- **Poisson($\lambda$)**: Models the number of rare events in a fixed interval. $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$.

*Each discrete distribution has its own unique characteristics and is suited for modeling different types of phenomena. Choosing the appropriate distribution depends on the nature of the problem and the assumptions about the data-generating process.*

## Important Continuous Distributions

Continuous probability distributions describe the probabilities of outcomes in an uncountable sample space. Here are some commonly encountered continuous distributions:

- **Uniform($a, b$)**: Models a variable that is equally likely to take any value in the interval $[a, b]$. $f(x) = \frac{1}{b-a}$, $a \leq x \leq b$.
- **Exponential($\lambda$)**: Models the waiting time until a rare event occurs. $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$.
- **Normal($\mu, \sigma^2$)**: Models a variable that is symmetrically distributed around its mean, with a bell-shaped curve. $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$.

*Continuous distributions are used to model a wide range of phenomena, from physical measurements to financial variables. The normal distribution, in particular, is ubiquitous due to the Central Limit Theorem.*

## Joint Distributions

Joint distributions describe the probabilities of outcomes for multiple random variables.

**Definition 6.** *The **joint probability mass function** of discrete random variables $X$ and $Y$ is $p_{X,Y}(x, y) = P(X = x, Y = y)$.*

**Definition 7.** *The **joint probability density function** of continuous random variables $X$ and $Y$ is a function $f_{X,Y}$ satisfying*

$$P((X, Y) \in A) = \iint_A f_{X,Y}(x, y)\, dy\, dx$$

*Joint distributions allow us to study the relationship between multiple random variables and to calculate probabilities of events involving these variables.*

**Definition 8.** *The **marginal pmf** of $X$ is $p_X(x) = \sum_y p_{X,Y}(x, y)$. The **marginal pdf** of $X$ is $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$.*

*Marginal distributions describe the probabilities of outcomes for a single random variable, ignoring the values of other variables. They can be obtained by summing or integrating the joint distribution over the other variables.*

**Definition 9.** *The **conditional pmf** of $Y$ given $X = x$ is $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$. The **conditional pdf** of $Y$ given $X = x$ is $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$.*

*Conditional distributions describe the probabilities of outcomes for one random variable, given the value of another variable. They are defined in terms of the joint and marginal distributions.*

**Theorem 4.** *$X$ and $Y$ are **independent** if and only if $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ (discrete case) or $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ (continuous case).*

*Independence of random variables is a stronger condition than independence of events. It means that the joint distribution factors into the product of the marginal distributions.*

## Functions of Random Variables

We are often interested in the distributions of functions of random variables.

**Definition 10.** *The **moment generating function** of $X$ is $M_X(t) = \mathbf{E}(e^{tX})$.*

*Moment generating functions (MGFs) are a powerful tool for studying the distributions of random variables. They uniquely determine the distribution (when they exist) and can be used to easily calculate moments and characterize the behavior of sums of independent random variables.*

**Theorem 5.** *If $X$ is discrete with pmf $p_X$ and $Y = g(X)$, then the pmf of $Y$ is*

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x)$$

**Theorem 6.** *If $X$ is continuous with pdf $f_X$ and $Y = g(X)$, where $g$ is monotone, then the pdf of $Y$ is*

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

*These theorems provide methods for finding the distributions of functions of discrete and continuous random variables, respectively. The discrete case involves summing the probabilities of all values of $X$ that map to each value of $Y$, while the continuous case involves a change of variables formula.*

**Theorem 7.** *If $X$ is continuous with cdf $F_X$, then $F_X(X) \sim$ Uniform$(0, 1)$.*

*This theorem, known as the Probability Integral Transform, is a fundamental result in probability theory. It allows us to simulate random variables with any continuous distribution using a uniform random number generator.*

## Sums of Independent Random Variables

The distribution of the sum of independent random variables is of great importance in probability theory and statistics.

**Theorem 8** (Convolution)**.** *If $X$ and $Y$ are independent, then:*

- *Discrete case:* $p_{X+Y}(z) = \sum_x p_X(x)p_Y(z-x)$
- *Continuous case:* $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx$

*The convolution formulas provide a way to calculate the distribution of the sum of independent random variables. In the discrete case, we sum the probabilities of all pairs of values of $X$ and $Y$ that add up to each possible value of $X+Y$. In the continuous case, we integrate the product of the densities over all pairs of values that sum to each possible value of $X+Y$.*

**Property 2** (Memoryless Property)**.**
- *If $X \sim$ Geometric$(p)$, then $P(X > m+n \mid X > m) = P(X > n)$.*
- *If $X \sim$ Exponential$(\lambda)$, then $P(X > s+t \mid X > s) = P(X > t)$.*

*The memoryless property is a unique characteristic of the geometric and exponential distributions. It states that the probability of waiting an additional amount of time (or trials) is independent of how long we have already waited. This property makes these distributions well-suited for modeling waiting times and arrival processes.*

## Limit Theorems

Limit theorems describe the behavior of sequences of random variables as the sample size grows large.

**Theorem 9** (Law of Large Numbers)**.** *If $X_1, X_2, \ldots$ are iid with $\mathbf{E}(X_i) = \mu$, then $\bar{X}_n \to \mu$ as $n \to \infty$.*

*The Law of Large Numbers states that the sample mean converges to the population mean as the sample size increases. This justifies the use of the sample mean as an estimate of the population mean and forms the basis for many statistical inference procedures.*

**Theorem 10** (Central Limit Theorem)**.** *If $X_1, X_2, \ldots$ are iid with $\mathbf{E}(X_i) = \mu$ and Var$(X_i) = \sigma^2 < \infty$, then*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty$$

*The Central Limit Theorem is one of the most important results in probability theory. It states that the standardized sum (or mean) of a large number of independent and identically distributed random variables converges to a standard normal distribution, regardless of the shape of the original distribution (under mild conditions). This theorem has far-reaching implications in statistics and is the foundation for many inferential procedures.*

**Example 1** (Poisson Limit of Binomial)**.** *If* $X_n \sim Bin(n, \lambda/n)$ *and* $\lambda$ *is fixed, then* $X_n \xrightarrow{d} Poisson(\lambda)$ *as* $n \to \infty$*.*

*This example illustrates another important limit theorem, which states that a binomial distribution with a small success probability and a large number of trials converges to a Poisson distribution. This provides a connection between the two distributions and justifies the use of the Poisson distribution as an approximation to the binomial distribution in certain situations.*

## Conclusion

This comprehensive review has covered the key concepts and results from an introductory course in probability theory, including:

- Probability spaces and axioms
- Discrete and continuous probability distributions
- Conditional probability and independence
- Joint, marginal, and conditional distributions
- Functions of random variables and transformations
- Sums of independent random variables and convolutions
- Limit theorems (Law of Large Numbers, Central Limit Theorem)

*By mastering these fundamental concepts and techniques, you have laid a solid foundation for further studies in probability, statistics, and related fields. The ideas and methods presented here are essential for understanding and analyzing a wide range of real-world phenomena, from science and engineering to economics and social sciences.*

As you continue your journey in probability and statistics, remember to:

- Practice solving problems to deepen your understanding and develop your skills.
- Explore the connections between different concepts and look for unifying themes.
- Apply your knowledge to real-world situations and data sets.
- Keep learning about more advanced topics and recent developments in the field.

*Probability theory is a rich and fascinating subject with a long history and numerous applications. By combining rigorous mathematical reasoning with intuitive interpretations and practical insights, it provides a powerful framework for making sense of the uncertain world around us. We hope that these review notes have helped you to appreciate the beauty and utility of probability theory, and we wish you the best in your future probabilistic endeavors!*