Notes on Probability Theory for FinTech
DATA 5690-6690: Computational Methods for FinTech
Spring Semester 2026

1 Introduction

- **Probability model.** A mathematical model for the study of nondeterministic situations.
- The term *stochastic* (derived from the Greek word *stochos*, meaning "guesses") is sometimes used instead of the word *probabilistic*.

2 Sample Spaces

Definition 2.1 (Sample Space)

The set of all possible outcomes of an experiment is called the *sample space*, denoted by Ω. Note that one and only one of the possible outcomes will occur on any given trial of the experiment.

- A sample space is said to be **finite** if it consists of a finite number of outcomes, say $\Omega = \{e_1, e_2, \ldots, e_N\}$, and is said to be **countably infinite** if its outcomes can be put into a one-to-one correspondence with the positive integers, say $\Omega = \{e_1, e_2, \ldots\}$.

Definition 2.2 (Discrete Sample Space)

If a sample space Ω is either finite or countably infinite, then it is called a *discrete sample space*.

3 Events

Definition 3.1 (Event)

An *event* is a subset of the sample space Ω. If A is an event, then A has occurred if it contains the outcome that occurred.

- We will consider the whole sample space Ω as a special type of event, called the *sure event*.
- We will also include the empty set $\varnothing$ as an event, called the *null event*.

Definition 3.2 (Elementary Event)

An event is called an *elementary event* if it contains exactly one outcome of the experiment.

Definition 3.3 (Mutually Exclusive Events)

Two events A and B are called *mutually exclusive* if

$$A \cap B = \varnothing.$$

Definition 3.4 (Pairwise Mutually Exclusive Events)

Events $A_1, A_2, A_3, \ldots$ are said to be *mutually exclusive* if they are pairwise mutually exclusive. That is, if

$$A_i \cap A_j = \varnothing \quad \text{whenever } i \neq j.$$

4 Relative Frequency and Statistical Regularity

Relative Frequency

If $m(A)$ represents the number of times that the event A occurs among M trials of a given experiment, then

$$f_A = \frac{m(A)}{M}$$

represents the *relative frequency* of occurrence of A on those trials of the experiment.

Statistical Regularity

If, for an event A, the limit of f_A as M approaches infinity exists, then one can assign probability to A by

$$P(A) = \lim_{M \to \infty} f_A.$$

This expresses a property known as *statistical regularity.*

Subjective Measure of Belief

- Although the relative frequency approach may not always be adequate as a practical method of assigning probabilities, it is the way that probability usually is interpreted.
- Many think this is too restrictive (or unnecessarily metaphysical).
- By regarding probability as a *subjective measure of belief* that an event will occur, they are willing to assign probability in any situation involving uncertainty without assuming properties such as repeatability or statistical regularity.

5 Definition of Probability

Probability as a Set Function

Mathematically, we can think of $P(A)$ as a set function. In other words, it is a function whose domain is a collection of sets (events), and the range of which is a subset of the real numbers.

> ***Definition 5.1 (Probability)***
>
> For a given experiment, Ω denotes the sample space, and $A_1, A_{01}, A_{02}, \ldots$ represent possible events. A set function that associates a real value $P(A)$ with each event A is called a *probability set function*, and $P(A)$ is called the *probability of* A, if the following properties are satisfied:
>
> 1. $0 \leq P(A)$ for every A.
> 2. $P(\Omega) = 1$.
> 3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), \quad$ if $A_1, A_2, \ldots$ are pairwise mutually exclusive events.

- These 3 properties are typically referred to as the **Probability Axioms**.
- One consequence is that the probability of the null set is zero, or $P(\varnothing) = 0$.

6 Random Variables

> ***Definition 6.1 (Random Variable)***
>
> A *random variable*, say X, is a function defined over a sample space, Ω, that associates a real number, $X(e) = x$, with each possible outcome e in Ω.

We will denote random variables with capital letters like X, Y, Z.

> ***Definition 6.2 (Discrete Random Variable and Probability Density Function)***
>
> If the set of all possible values of a random variable X is a countable set, $x_1, x_2, \ldots, x_n$, or $x_1, x_2, \ldots$, then X is called a *discrete random variable.* The function
>
> $$f(x) = P[X = x], \quad x = x_1, x_2, \ldots$$
>
> that assigns the probability to each possible value x will be called the *discrete probability density function* (discrete pdf).

- If it is clear from context that X is discrete, then we will simply say *pdf.*
- Another common terminology is *probability mass function* (pmf).

- The possible values x_i are called *mass points*.

7 Cumulative Distribution Function (CDF)

Definition 7.1 (Cumulative Distribution Function (CDF))

The cumulative distribution function (CDF) of a random variable X is defined for any real x by

$$F(x) = P[X \leq x].$$

- The function $F(x)$ often is referred to simply as the *distribution function* of X, and the subscripted notation $F_X(x)$ is sometimes used.
- For brevity, it is common to use a short notation to indicate that a distribution of a particular form is appropriate.
- If we write $X \sim f(x)$ or $X \sim F(x)$, this will indicate that the random variable X has pdf $f(x)$ and CDF $F(x)$.

Theorem 7.1

Let X be a discrete random variable with pdf $f(x)$ and CDF $F(x)$. If the possible values of X are indexed in increasing order, $x_1 < x_2 < x_3 < \cdots$, then $f(x_i) = F(x_i)$, and for any $i > 1$,

$$f(x_i) = F(x_i) - F(x_{i-1}).$$

Furthermore, if $x < x_1$, then $F(x) = 0$, and for any other real x,

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

where the summation is taken over all indices i such that $x_i \leq x$.

8 Expected Value

Definition 8.1 (Expected Value (Discrete))

If X is a discrete random variable with pdf $f(x)$, then the *expected value* of X is defined by

$$E(X) = \sum_x x\, f(x).$$

- Common notation for $E(X)$ is μ or μ_X.
- The terms *mean* or *expectation* are also used.

9 Continuous Random Variables

Definition 9.1 (Continuous Random Variable)

A random variable X is called a *continuous random variable* if there is a function $f(x)$, called the *probability density function* (pdf) of X, such that the CDF can be represented as

$$F(x) = \int_{-\infty}^{x} f(t)\, dt.$$

Definition 9.2 (Expected Value of a Continuous RV)

If X is a continuous random variable with pdf $f(x)$, then the expected value of X is defined by

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

if the integral in the equation is absolutely convergent. Otherwise we say that $E(X)$ does not exist.

As before, we often call $E(X)$ the *mean* or *expectation*.

10 Percentile

Definition 10.1 (Percentile)

If $0 < p < 1$, then a $100 \times p$th *percentile* of the distribution of a continuous random variable X is a solution x_p to the equation

$$F(x_p) = p.$$

- We can also think in terms of a proportion p, rather than a percentage $100 \cdot p$, of the population, and in this context x_p is called a pth *quantile* of the distribution.
- A *median* of the distribution of X is the 50th percentile, denoted by $x_{0.5}$ or m.

11 Mode and Symmetry

Definition 11.1 (Mode)

If the pdf has a unique maximum at $x = m_0$, say $\max f(x) = f(m_0)$, then m_0 is called the *mode* of X.

Definition 11.2 (Symmetric Distribution)

A distribution with pdf $f(x)$ is said to be *symmetric about* c if

$$f(c-x) = f(c+x) \quad \text{for all } x.$$

- The graph of $y = f(x)$ is a "mirror image" about the vertical line $x = c$.
- Asymmetric distributions are called *skewed*.
- If $f(x)$ is symmetric about c and μ exists, then $c = \mu$. Additionally, if $f(x)$ has a unique maximum at m_0 and a unique median m, then $\mu = m_0 = m$.

12 Variance and Moments

Definition 12.1 (Variance)

The *variance* of a random variable X is given by

$$\mathrm{Var}(X) = E\left[(X-\mu)^2\right].$$

Common notations for the variance are σ^2, σ_X^2, or $V(X)$.

A related quantity, called the *standard deviation* of X, is the positive square root of the variance:

$$\sigma = \sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Definition 12.2 (kth Moment)

The kth moment about the origin of a random variable X is

$$\mu_k' = E(X^k),$$

and the kth moment about the mean is

$$\mu_k = E[X - E(X)]^k = E(X-\mu)^k.$$

13 Special Discrete Distributions

13.1 Bernoulli Distribution

A *Bernoulli random variable* is one that assumes only the values 0 or 1. The experiment with only the two outcomes 0 or 1 is called a *Bernoulli trial*.

If an experiment can result only in "success" (E) or "failure" (E'), then the corresponding Bernoulli random variable is

$$X(e) = \begin{cases} 1 & \text{if } e \in E, \\ 0 & \text{if } e \in E'. \end{cases}$$

The pdf of X is given by $f(0) = 1-\theta$ and $f(1) = \theta$. The corresponding distribution is the *Bernoulli distribution*, and its pdf can be expressed as

$$f(x) = \theta^x (1-\theta)^{1-x}, \quad x = 0, 1.$$

13.2 Binomial Distribution

A *binomial experiment* can be constructed as a sequence of independent Bernoulli trials, where the quantity of interest is the number of successes on a certain number of trials.

In a sequence of n independent Bernoulli trials with probability θ (success) on each trial, let X represent the number of successes. The discrete pdf of X is given by

$$f(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \ldots, n.$$

The number of combinations of n distinct objects chosen r at a time is

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}.$$

This is called the *binomial coefficient*. It is also related to *Pascal's Triangle*.

Summary of the Binomial Distribution

Distribution	**pdf** $f(x)$	**Mean**	**Variance**
$X \sim \mathrm{BIN}(n, \theta)$	$\binom{n}{x} \theta^x (1-\theta)^{n-x}$	$n\theta$	$n\theta(1-\theta)$
$0 < \theta < 1, \quad q = 1-\theta$	$x = 0, 1, \ldots, n$		

Note: The Bernoulli distribution is the Binomial distribution for $n = 1$.