Math 416/516: Stochastic Simulation

Haijun Li

lih@math.wsu.edu
Department of Mathematics
Washington State University

Week 1
Outline

1. Basic Concepts from Probability Theory
2. Random Variables
3. Random Vectors
4. Conditional Distribution
Notation

- Sample or outcome space $\Omega := \{\text{all possible outcomes } \omega \text{ of an experiment}\}$.
- Any element $\omega \in \Omega$ is called a sample point or simple event.
- Any $A \subseteq \Omega$ (subset of $\Omega$) is called an event.
- The empty set is denoted by $\emptyset$, which is also called the impossible event.
- The sample space $\Omega$ is also called the sure event.
- If $A, B$ are events, then union $A \cup B$, intersection $A \cap B$ (or $AB$), and complement $A^c$ are also events.
- $\sigma$-field $\mathcal{F}$ ( = operational class of observable events): A non-empty class of subsets of $\Omega$ that are closed under countable union, countable intersection and complements.
Probability Space

- $P(A) := \text{likelihood of event } A$
- Probability measure $P(\cdot) : \mathcal{F} \rightarrow [0, 1]$ is a (measurable) function of events.

Operational Axioms

1. $P(\Omega) = 1$.  
2. For any sequence of disjoint (or mutually exclusive) events \{ $A_1, A_2, A_3, \ldots$ \}; that is, $A_i \cap A_j = \emptyset$, $i \neq j$,  
   \[ P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i). \]

The triplet $(\Omega, \mathcal{F}, P)$ is called a probability space (or a probability model).
Example: S&P Rating

An investor is interested in buying bonds from a S&P rated bank with the following possible ratings and default probabilities.

<table>
<thead>
<tr>
<th>S&amp;P Rating</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default Prob</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td>50</td>
<td>300</td>
<td>1100</td>
<td>2800</td>
</tr>
</tbody>
</table>

- Sample space $\Omega = \{\text{AAA}, \text{AA}, \text{A}, \text{BBB}, \text{BB}, \text{B}, \text{CCC}\}$.
- $\sigma$-field $\mathcal{F}$ consists of all subsets of $\Omega$.
- The probability $P(\cdot)$ is given by default probabilities.
Properties of A Probability Measure

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

1. \(P(A^c) = 1 - P(A)\) for any event.
2. \(P(A \cup B) = P(A) + P(B) - P(A \cap B)\) for any two events \(A, B\).
3. For any sequence of events \(\{A_1, A_2, \ldots, A_n\}\) (may not be disjoint),
   
   \[
P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(\bigcap_{i=1}^{n} A_i).
   \]
Probability Given Some Information

Definition

Let $(\Omega, \mathcal{F}, P)$ denote a probability space. For any two events $A, B \in \mathcal{F}$, the conditional probability of $A$ given $B$ is given by

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

provided that $P(B) > 0$.

- For fixed $B$, $P(\cdot|B) = \text{Probability measure given that } B \text{ has occurred}$.
- $P(\emptyset|B) = 0$, and $P(\Omega|B) = 1$.
- $P(A \cup C|B) = P(A|B) + P(C|B)$ for any two disjoint events $A$ and $C$. 
Product Form & Independence

Let $A_1$ and $A_2$ be two events.

- Rewrite the conditional probability: $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$.
- If $B$ is another event, then

$$P(A_1 \cap A_2|B) = P(A_1|B)P(A_2|A_1 \cap B).$$

**Definition**

1. Two events $A$ and $B$ are **stochastically independent** if $P(A|B) = P(A)$ or

$$P(A \cap B) = P(A)P(B).$$

2. A collection of events $\{A_1, \ldots, A_n\}$ is said to be independent if

$$P(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_m}) = P(A_{k_1})P(A_{k_2})\cdots P(A_{k_m})$$

for any $1 \leq k_1 < k_2 < \cdots < k_m \leq n$. 
Remark

- The event $\emptyset$ is independent of any event (including itself and $\Omega$).
- The event $\Omega$ is independent of any event (including itself and $\emptyset$).
- Let two events $A$ and $B$ be independent, then the $\sigma$-field

$$(\emptyset, A, A^c, \Omega)$$

and the $\sigma$-field

$$(\emptyset, B, B^c, \Omega)$$

are independent.
Total Probability Law

Consider two events $A$ and $B$.

- $P(A \cap B) = P(B)P(A|B)$.
- $P(A \cap B^c) = P(B^c)P(A|B^c)$.
- $P(A) = P(A \cap B) + P(A \cap B^c) = P(B)P(A|B) + P(B^c)P(A|B^c)$.

**Theorem**

Let the collection of disjoint events $\{B_k, k \geq 1\}$ be a partition of the sample space $\Omega$ (i.e., $\cup_{k \geq 1} B_k = \Omega$). Then for any event $A$,

$$P(A) = \sum_{k=1}^{\infty} P(B_k)P(A|B_k).$$
Example: S&P Rating

An investor has purchased bonds from five S&P AAA-rated banks and three S&P A-rated banks.

Table: Annual default probability in basis point (bp), 100 basis point = 1%

<table>
<thead>
<tr>
<th>S&amp;P Rating</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default Prob</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td>50</td>
<td>300</td>
<td>1100</td>
<td>2800</td>
</tr>
</tbody>
</table>

What is the probability that exactly one bank defaults?

Solution: Using the total probability law,

\[
\frac{5(0.9999)^4(0.0001)(0.9988)^3}{(0.9988)^2} + \frac{3(0.9999)^5(0.9988)^2(0.0012)}{(0.9988)^2} = 40.88 \text{bps}
\]

one AAA-bank defaults + one A-bank defaults
Random Variables

Consider a probability space \((\Omega, \mathcal{F}, P)\).

- Random variables \(X, Y, Z, \ldots\): Variables associated with random outcomes and random events.
- Random variable \(X : \Omega \rightarrow \mathbb{R}\) is a real-valued (measurable) function defined on \(\Omega\).
- “Measurable” means that these events are observable:
  \[X^{-1}(a, b) := \{a < X \leq b\} \in \mathcal{F}\quad \text{for all } a, b \in \mathbb{R};\]
- Cumulative distribution function (CDF) \(F(x) := P(X \leq x), \, x \in \mathbb{R}\).
  \[\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 1.\]
Continuous and Discrete Random Variables

- Random variable $X$ is said to be **continuous** if the distribution function $F$ has no jumps, that is,

$$\lim_{h \to 0^+} F(x - h) = F(x), \forall x \in \mathbb{R}.$$

Most continuous distributions of interest have a density $f(x) \geq 0$:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y) dy, \ x \in \mathbb{R}$$

where $\int_{-\infty}^{\infty} f(y) dy = 1$.

- Random variable $X$ is said to be **discrete** if the distribution function $F$ is a pure jump function:

$$F(x) = \sum_{x_k \leq x} P(X = x_k) = \sum_{x_k \leq x} p_k, \ x \in \mathbb{R}$$

where the probability mass function $\{p_k\}$ satisfies that $0 \leq p_k \leq 1$ and $\sum_{k=1}^{\infty} p_k = 1$. 
For a real-valued function \( g \), the expectation of \( g(X) \) is given by

\[
E[g(X)] = \int g(x) dF(x).
\]

- If \( X \) is continuous with density (PDF) \( f \), then

\[
E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx.
\]

- If \( X \) is discrete with mass function (PMF) \( \{p_k\} \), then

\[
E[g(X)] = \sum_{k=1}^{\infty} g(x_k) p_k = \sum_{k=1}^{\infty} g(x_k) P(X = x_k).
\]
The $k$-th moment of $X$ is given by $E(X^k) = \int x^k dF(x)$. The mean $\mu$ (or “center of gravity”) of $X$ is the first moment:

$$E(X) = \int x \, dF(x).$$

The variance (or “spread out”) of $X$ is defined as

$$\sigma^2 = \text{var}(X) := E(X - \mu)^2,$$

and

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

If the variance exists, then the Chebyshev inequality holds:

$$P(|X - \mu| > k\sigma) \leq k^{-2}, \quad k > 0.$$

That is, the probability of tail regions that are $k$ standard deviations away from the mean is bounded by $1/k^2$. 
Normal (or Gaussian) Density Distribution (PDF)

- Normal random variable $X \sim N(\mu, \sigma^2)$:
  $$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$  

- $E(X) = \mu = $ location parameter, $\text{var}(X) = \sigma^2 = $ scale parameter.
Gaussian Distribution $\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma^2}(y)\,dy$ (CDF)
Poisson Probability Mass Function (PMF)

- Poisson random variable $X \sim \text{Poi}(\lambda)$ (PMF):
  \[ P(X = k) = p_k = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \ldots \]

- $E(X) = \text{var}(X) = \lambda$. 

![Graphs for upper tail case](image-url)
Poisson Distribution $F_X(x) = \sum_{x_k \leq x} p_k$ (CDF)
Example: Binomial Model for Stock Prices

Let \( d \) and \( u \) be the positive constants such that \( d < 1 < u \).

- \( S_k \) = price of a certain stock at the end of the \( k \)-th time period.
- At the next time step, the price moves up to \( uS_k \) with probability \( p \).
- At the next time step, the price moves down to \( dS_k \) with probability \( 1 - p \).
- Given the initial stock price \( S_0 = x \), find the distribution of \( S_n \).

**Solution:** All the possible values of \( S_n \) are of the form \( u^k d^{n-k} x \), \( 0 \leq k \leq n \), where the stock moves up \( k \) times and down \( n - k \) times during the \( n \) time periods. Using the binomial distribution, we have,

\[
P(S_n = u^k d^{n-k} x) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n.
\]

\[
E(S_n) = np, \quad \text{Var}(S_n) = np(1 - p).
\]
Example: Value-at-Risk from Log-Normal

- $X \geq 0$ has a log-normal distribution if $\ln(X)$ has a normal distribution $N(\mu, \sigma^2)$.
- Denoted by $X \sim \text{LogN}(\mu, \sigma^2)$.
- $P(X \leq x) = P(\ln(X) \leq \ln(x)) = \Phi \left( \frac{\ln(x) - \mu}{\sigma} \right)$.
Example: Value-at-Risk

- Value-at-Risk, \( \text{VaR}_{1-\alpha}(X) \), of \( X \) at confidence level \( 1 - \alpha \), \( 0 \leq \alpha \leq 1 \), is the \( 100\alpha \)th percentile such that

\[
P(X \leq \text{VaR}(X)) = \alpha.
\]

- That is, \( \text{VaR}_{1-\alpha}(X) \) is the threshold which would be exceeded with probability \( 1 - \alpha \). Derive the VaR for the log-normal distribution.

Solution: Since

\[
\alpha = P(X \leq \text{VaR}(X)) = \Phi \left( \frac{\ln(\text{VaR}(X)) - \mu}{\sigma} \right),
\]

we have for \( 0 < \alpha \leq 0.5 \),

\[
\frac{\ln(\text{VaR}(X)) - \mu}{\sigma} = -z_\alpha
\]

leading to \( \text{VaR}(X) = e^{\mu - z_\alpha \sigma} \).
Random Vectors

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

- \(\mathbf{X} = (X_1, \ldots, X_d) : \Omega \rightarrow \mathbb{R}^d\) denotes a \(d\)-dimensional random vector, where its components \(X_1, \ldots, X_d\) are real-valued random variables.

- Joint cumulative distribution function (CDF)
  \[ F(\mathbf{x}) := P(X_1 \leq x_1, \ldots, X_d \leq x_d), \quad \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d. \]
Continuous and Discrete Random Vectors

- A vector of continuous random variables \((X, Y)\) is described by the joint density (PDF) \(f(x, y)\)

\[
P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) \, dy \, dx.
\]

- A vector of discrete random variables \((X, Y)\) is described by the joint mass function (PMF) \(p(x_i, y_j)\)

\[
P(a \leq X \leq b, c \leq Y \leq d) = \sum_{a \leq x_i \leq b} \sum_{c \leq y_j \leq d} p(x_i, y_j).
\]
A Bivariate Normal (Gaussian) Density (PDF)
Bivariate Gaussian Distribution (CDF)
Bivariate Gaussian Contour Plot
Bivariate Discrete Distribution

Figure: graphs for upper tail case
Expectation, Variance, and Covariance

- The expectation or mean value of $X = (X_1, \ldots, X_d)$ is denoted by $EX := (E(X_1), \ldots, E(X_d)) = (\mu_1, \ldots, \mu_d)$.

- The covariance matrix of $X$ is defined as

  $$\Sigma := (\text{cov}(X_i, X_j); i, j = 1, \ldots, d)$$

  where the covariance of $X_i$ and $X_j$ is defined as

  $$\sigma_{ij} = \text{cov}(X_i, X_j) := E[(X_i - \mu_i)(X_j - \mu_j)] = E(X_iX_j) - \mu_i\mu_j.$$  

- The correlation coefficient of $X_i$ and $X_j$ is denoted by

  $$\text{corr}(X_i, X_j) := \frac{\text{cov}(X_i, X_j)}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$  

  It follows from the Cauchy-Schwarz inequality that $-1 \leq \text{corr}(X_i, X_j) \leq 1$. 
Independence

- Recall: The events (subsets) $A_1, \ldots, A_n$ are independent if for any $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,

$$P(\cap_{j=1}^k A_{i_j}) = \prod_{j=1}^k P(A_{i_j}).$$

- For example, three events $A, B, C$ are independent if

$$P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C), \quad P(B \cap C) = P(B)P(C),$$

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

- The random variables $X_1, \ldots, X_n$ are independent if for any nice (Borel) sets $B_1, \ldots, B_n$, the events

$$\left\{ X_1 \in B_1 \right\}_{A_1} \ldots \left\{ X_n \in B_n \right\}_{A_n}$$

are independent.
The random variables $X_1, \ldots, X_n$ are independent if and only if
\[ F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \text{ for all } (x_1, \ldots, x_n) \in \mathbb{R}^n. \]

The random variables $X_1, \ldots, X_n$ are independent if and only if
\[ E[\prod_{i=1}^{n} g_i(X_i)] = \prod_{i=1}^{n} E g_i(X_i) \]
for any real-valued functions $g_1, \ldots, g_n$.

In the continuous case, the random variables $X_1, \ldots, X_n$ are independent if and only if
\[ f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), \text{ for all } (x_1, \ldots, x_n) \in \mathbb{R}^n. \]
Dependence VS Independence

Daily Return (%)

Bi-variate standard normal
100 cycles with Box-Muller
Multivariate Normal (Gaussian) Distributions

Let \( \mathbf{X} = (X_1, \ldots, X_d) \sim N(\mu, \Sigma) \) have a \( d \)-dimensional Normal Gaussian distribution, where \( \mu \in \mathbb{R}^d \) and \( \Sigma \) is a symmetric positive-definite \( d \times d \) matrix.

The normal density is given by

\[
f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \mu)\Sigma^{-1}(\mathbf{x} - \mu)'} , \quad \mathbf{x} \in \mathbb{R}^d.
\]

The mean vector \( E(\mathbf{X}) = \mu \), and the covariance matrix \( \Sigma = (\sigma_{ij})_{d \times d} \), where

\[
\sigma_{ij} = \text{cov}(X_i, X_j).
\]
Properties of Multivariate Gaussian

\( \mathbf{X} = (X_1, \ldots, X_d) \sim N(\mu, \Sigma). \)

- For any \( d \times m \) matrix \( A \), the \( m \)-dimensional row vector \( \mathbf{XA} \) is also normally distributed.
  \[ \mathbf{XA} \sim N(\mu A, A' \Sigma A). \]

- The random variables \( X_1, \ldots, X_d \) are independent if and only if
  \( \text{corr}(X_i, X_j) = 0 \) for \( i \neq j \).
For non-Gaussian random vectors, however, independence and uncorrelatedness are not equivalent. Let $X$ be a standard normal random variable. Since both $X$ and $X^3$ have expectation zero, $X$ and $X^2$ are uncorrelated:

$$\text{cov}(X, X^2) = E(X^3) - E(X)E(X^2) = 0.$$ 

But $X$ and $X^2$ are clearly dependent (co-monotone). Since 

$$\{X \in [-1, 1]\} = \{X^2 \in [0, 1]\},$$ 

we obtain that

$$P(X \in [-1, 1], X^2 \in [0, 1]) = P(X \in [-1, 1])$$

$$> [P(X \in [-1, 1])]^2 = P(X \in [-1, 1])P(X^2 \in [0, 1]).$$
Example: Payoff of an Option

- The payoff $h(x_1, x_2)$ of an option depends on the prices $S_1$ and $S_2$ of two stocks.
- $h(S_1, S_2) = 1$ if $S_1 > k_1$ and $S_2 > k_2$, and zero otherwise.
- Assume that $S_1, S_2$ are independent, and $S_i$ has a log-normal distribution with parameters $\mu_i$ and $\sigma_i^2$.
- Find the expected payoff.

Solution: Since

$$P(S_i > k_i) = \Phi \left( -\frac{\ln(k_i) - \mu_i}{\sigma_i} \right),$$

then

$$E[h(S_1, S_2)] = P(S_1 > k_1, S_2 > k_2) = P(S_1 > k_1)P(S_2 > k_2)$$

$$= \Phi \left( -\frac{\ln(k_1) - \mu_1}{\sigma_1} \right) \Phi \left( -\frac{\ln(k_2) - \mu_2}{\sigma_2} \right).$$
 Conditional Distributions

- If $X$ and $Y$ are discrete random variables, then the conditional mass function given that $Y = y$ is given by

$$p(x|y) := P(X = x| Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \ x = x_1, x_2, \ldots$$

- If $X$ and $Y$ are continuous, then the conditional density function given that $Y = y$ is given by

$$f(x|y) = \frac{f(x, y)}{f(y)}, \ x \in \mathbb{R}.$$
There are \( m \) obligors, and the \( i \)-th obligor defaults if and only if his risk score \( X_i > x_i \) for some threshold \( x_i \).

The risk scores are assumed to take the form:

\[
X_i = \rho_i Z + \sqrt{1 - \rho_i^2} \epsilon_i, \quad -1 < \rho_i < 1,
\]

where \( Z, \epsilon_1, \ldots, \epsilon_m \) are independent, have standard normal \( N(0, 1) \).

\( Z \) is a common systematic risk factor affecting all obligors.

\( \epsilon_i \) is a specific risk factor affecting the \( i \)-th obligor.

\( X_i \sim N(0, 1), \quad 1 \leq i \leq m. \)

\( \text{cov}(X_i, X_j) = \rho_i \rho_j, \quad i \neq j. \)
Example: Compute No Default Probability

- \( P(\text{no default}) = \frac{P(X_1 \leq x_1, \ldots, X_m \leq x_m)}{\text{x}_1, \ldots, \text{x}_m \text{ are dependent}} \)
- Conditioning on \( Z = z \), \( X_1, \ldots, X_m \) become independent.

\[
P(\text{no default} \mid Z = z) = \prod_{k=1}^{m} P(X_k \leq x_k \mid Z = z) = \prod_{k=1}^{m} \Phi \left( \frac{x_k - \rho_i z}{\sqrt{1 - \rho_k^2}} \right).
\]

- The total probability law implies that

\[
P(\text{no default}) = \int_{-\infty}^{\infty} \prod_{k=1}^{m} \Phi \left( \frac{x_k - \rho_i z}{\sqrt{1 - \rho_k^2}} \right) \varphi_{0,1}(z) \, dz,
\]

where \( \varphi_{0,1}(z) \) is the standard normal density.
- No closed form! We need to use simulation to calculate it.
Conditional Expectations

- If $X$ and $Y$ are discrete random variables, then the conditional expectation given that $Y = y$ is given by
  \[
  E(X|Y = y) = \sum_{k=1}^{\infty} x_k p(x_k|y).
  \]

- If $X$ and $Y$ are continuous, then the conditional expectation given that $Y = y$ is given by
  \[
  E(X|Y = y) = \int_{-\infty}^{\infty} xf(x|y)dx.
  \]

- The random variable $E(X|Y)$ is called the conditional expectation of $X$ given $Y$. 
Properties of Conditional Expectations

Let $h : \mathbb{R} \to \mathbb{R}$ be any measurable function.

- $E(aX + bZ|Y) = aE(X|Y) + bE(Z|Y)$ for any constants $a$ and $b$.
- $E[h(Y)|Y] = h(Y)$.
- $E[h(Y)X|Y] = h(Y)E(X|Y)$.
- $E[h(X)|Y] = E[h(X)]$, if $X$ and $Y$ are independent.
- $E[E(X|Y)] = E(X)$ for any $X$ and $Y$. 
Let $Y$ denote the log-return of a stock over a fixed trading period. Assume that

$$Y = X + \sum_{i=1}^{N} Z_i.$$ 

$X \sim N(\mu, \sigma^2)$ represents the periodic jitter that causes minor fluctuations in stock prices.

$N$ is Poisson with rate $\lambda$, and $Z_1, Z_2, \ldots$ are independent, normal $N(0, \nu^2)$.

$$\sum_{i=1}^{N} Z_i = \text{larger price movements caused by major market upsets.}$$

$X, N, Z_1, Z_2, \ldots$ are all independent.

Use the conditional expectation to verify that $E(Y) = \mu$. 
Two Limiting Theorems

i.i.d. = independent and identically distributed

**Law of Large Numbers (LLN)**

If \( X_1, X_2, \ldots, X_n \) are i.i.d. taken from a population with finite mean \( \mu \), then

\[
\lim_{n \to \infty} \frac{\sum_{i=1}^{n} X_i}{n} = \mu, \text{ with probability 1.}
\]

**Central Limit Theorem (CLT)**

If \( X_1, X_2, \ldots, X_n \) are i.i.d. taken from a population with finite mean \( \mu \) and finite variance \( \sigma^2 \), then

\[
P \left( \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma}} \leq x \right) \to \Phi(x)
\]

as \( n \to \infty \), where \( \Phi(x) \) is a standard normal CDF.
Remarks

- LLN provides a first order approximation:
  \[ \sum_{i=1}^{n} X_i \approx n\mu. \]

- CLT provides a second order (random) approximation:
  \[ \sum_{i=1}^{n} X_i \approx n\mu + Z\sqrt{n}\sigma, \quad Z \sim N(0, 1). \]

- One has also developed second order deterministic approximations (various Laws of Iterated Logarithms).

- Higher order approximations have also been developed (e.g., Edgeworth Expansions).

- People who made contributions: Jacob Bernoulli, Pierre-Simon Laplace, Siméon Denis Poisson, Pafnuty Chebyshev, Émile Borel, Andrey Kolmogorov.