OM9103: Stochastic Process

Lectures 1 & 2:
A Review on Probability Theory

A. Probability

A.1 Three Key Concepts

(1) "Random experiment"
- an experiment whose outcome is uncertain

(2) "Sample space": S
- the set of all possible outcomes of the experiment

(3) "Event": E
- a subset of the sample space

Example: flip a coin, roll a die, …

A.2 Probability Function

Let Ω be the set of all possible events, $E \subseteq S$. Ω is in fact the power set of S, namely , $\Omega = 2^S$. A probability function P: $\Omega \rightarrow [0, 1]$ such that for $E \in \Omega$, $P(E)$ is the probability of the event E.

Axioms:

(1) $P(E) \in [0, 1]$, $\forall E \in \Omega$;
(2) $P(S) = 1$;
(3) For any sequence of events $\{E_i$: $i = 1, \ldots, \infty\}$ such that $E_i \cap E_j = \varnothing$, all $i \neq j$, then

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

Key properties:

(a) $E_1 \subseteq E_2 \Rightarrow P(E_1) \leq P(E_2)$;
(b) $P(E^C) = 1 - P(E)$;
(c) Boole's inequality: $P\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} P(E_i)$.

Continuity: limit of a sequence of nested events

- $E_1 \subseteq E_2 \subseteq \ldots \Rightarrow \lim_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n$;
- $E_1 \supseteq E_2 \supseteq \ldots \Rightarrow \lim_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} E_n$;

Proposition A.1: If $\{E_n\}$ is a monotone (increasing or decreasing) sequence of events, then

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

Proof. Let $\{E_n\}$ be an increasing sequence of events. By definition,

$$P(\lim_{n\to\infty} E_n) = P\left(\bigcup_{n=1}^{\infty} E_n\right).$$

Now define $F_1 = E_1$ and for $n > 1$, F_n is the set of these elements in E_n that are not in any of the earlier E_i, $i < n$. Then it follows:

(1) $F_n = E_n\left(\bigcup_{i=1}^{n-1} E_i\right)^C = E_n E_{n-1}^C$;

(2) $\{F_n\}$ are mutually exclusive events;

(3) $\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i$ and $\bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i$, for all $n \geq 1$.

Therefore,

$$P(\lim_{n\to\infty} E_n) = P\left(\bigcup_{n=1}^{\infty} E_n\right) = P\left(\bigcup_{n=1}^{\infty} F_n\right) = \lim_{n\to\infty} \sum_{i=1}^{n} P(F_i)$$
$$= \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} F_i\right) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} E_i\right) = \lim_{n\to\infty} P(E_n)$$

For a decreasing sequence $\{E_n\}$, we apply the above result to the sequence of their complements $\{E_i^C\}$:

$$P\left(\bigcup_{n=1}^{\infty} E_n^C\right) = \lim_{n\to\infty} P\left(E_n^C\right) \Rightarrow 1 - P\left(\bigcap_{n=1}^{\infty} E_n\right) = 1 - \lim_{n\to\infty} P\left(E_n\right)$$
$$\Rightarrow P\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} P\left(E_n\right).$$

A.3 Borel-Cantelli Lemma

Given a sequence of events $\{E_n\}$, what is the probability that a finite or an infinite number of these events occur?

$$\{\text{an infinite number of the } E_i \text{ occurs}\} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i = \limsup_{i\to\infty} E_i$$

i.e., for each n, at least one of the E_i occurs for $i \geq n$.

Borel-Cantelli Lemma: If $\sum_{i=1}^{\infty} P(E_i) < \infty$, then

$$P(E_i \text{ occurs infinitely often}) = 0.$$

Proof.

$$\begin{aligned}
&P(E_i \text{ occurs infinitely often}) \\
&= P(\limsup_{i\to\infty} E_i) = P\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} E_i\right) = P\left(\lim_{n\to\infty}\bigcup_{i=n}^{\infty} E_i\right) \\
&= \lim_{n\to\infty} P\left(\bigcup_{i=n}^{\infty} E_i\right) \le \lim_{n\to\infty}\sum_{i=n}^{\infty} P\left(E_i\right) = 0.
\end{aligned}$$

<u>Example</u>: Consider a sequence of random variables X_1, X_2, … such that $P(X_n = 0) = 1/n^2$ and $P(X_n = 1) = 1 - 1/n^2$. Let $E_n = \{X_n = 0\}$. Note that

$$\sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty}\frac{1}{n^2} < \infty$$

$$\Rightarrow P(X_n = 0 \text{ infinitely often}) = 0$$

which implies that for sufficiently large n, $X_n = 1$. Hence with probability 1,

$$\lim_{n\to\infty} X_n = 1.$$

<u>Borel-Cantelli Lemma (Converse)</u>: If $\{E_n\}$ are independent events such that $\sum_{i=1}^{\infty} P(E_i) = \infty$, then $P(E_i \text{ occurs infinitely often}) = 1$.

<u>Proof</u>.

$$P(E_i \text{ occurs i.o.}) = \lim_{n\to\infty} P\left(\bigcup_{i=n}^{\infty} E_i\right) = 1 - \lim_{n\to\infty} P\left(\bigcap_{i=n}^{\infty} E_i^C\right)$$

$$= 1 - \lim_{n\to\infty}\prod_{i=n}^{\infty} P\left(E_i^C\right) = 1 - \lim_{n\to\infty}\prod_{i=n}^{\infty}\left(1 - P\left(E_i\right)\right)$$

Now,

$$\prod_{i=n}^{\infty}(1 - P(E_i)) \le \prod_{i=n}^{\infty} e^{-P(E_i)} = \exp\left(-\sum_{i=n}^{\infty} P(E_i)\right) = 0, \forall n.$$

$$\Rightarrow P(E_n \text{ occurs i.o.}) = 1.$$

<u>Notes</u>:

(1) The converse of the Borel-Cantelli Lemma requires *the independence assumption*, while the original lemma is true in general.

(2) For <u>any</u> sequence of independent events $\{E_n\}$,

$$P(E_n \text{ occurs i.o.}) = 0 \text{ or } 1.$$

(This is known as the "0-1 property".)

B. Random Variables

B.1 Definition

A *random variable* X is real-valued function defined on the sample space, i.e., $X: S \to R^1$. For any event $A \subseteq R^1$, $\{X \in A\}$ defines an event in S, i.e., $\{X \in A\} \subseteq S$. Hence $\{X \in A\} \equiv X^{-1}(A)$ has a probability value, i.e.,

$$\Pr(X \in A) = P(X^{-1}(A)), \ \forall A \subseteq R^1.$$

B.2 Cumulative Distribution Function (c.d.f.)

$$F(x) \equiv \Pr(X \le x) = \Pr(X \in (-\infty, x]) = P(X^{-1}(-\infty, x])$$
$$\bar{F}(x) = 1 - F(x) = \Pr(X > x) \text{ (tail probability)}$$

(1) If S is countable, X is a discrete random variable and

$$F(x) = \sum_{y \le x} \Pr(X = y)$$

(2) X is a continuous random variable if there exists a function $f(x)$ as probability density function (p.d.f.) such that

$$P(X \in B) = \int_B f(x)ds, \ \forall B \subseteq R^1; \ \Rightarrow F(x) = \int_{-\infty}^{x} f(t)dt.$$

B.3 Joint Distribution

Consider two r.v.s X and Y. Their joint d.f. $F(x, y)$ is defined as follow:

$$F(x, y) = \Pr(X \le x, Y \le y).$$

Given a joint d.f. for (X, Y), the underlying d.f. for X and Y respectively is known as the marginal d.f.:

$$F_X(x) = F(x, \infty) = \lim_{y \to \infty} F(x, y)$$
$$F_Y(y) = F(\infty, y) = \lim_{x \to \infty} F(x, y)$$

B.4 Independent R.V.s

<u>Definition</u>: X and Y are said to be independent if $F(x, y) = F_X(x)\, F_Y(y)$.

If X, Y are continuous r.v.s, then there exists a joint density function $f(x, y)$ such that

$$\Pr(X \in A, Y \in B) = \int_A \int_B f(x, y)dxdy, \ \forall A \subseteq R^1, B \subseteq R^1$$

Two continuous r.v.s X and Y are independent if and only if $f(x, y) = f_X(x)\, f_Y(y)$, where $f_X(x) = F_X'(x)$ and $f_Y((y) = F_Y'(y)$.

B.5 Descriptive Statistics of R.V.s

(a) Expected Value ("Mean")

$$E(X) = \int_{-\infty}^{\infty} x dF(x) = \begin{cases} \int_{-\infty}^{\infty} x f(x) dx & X \text{ is continuous} \\ \sum_{x=-\infty}^{\infty} x \Pr(X = x) & X \text{ is discrete} \end{cases}$$

If $Y = h(X)$, then

$$E(Y) = \int_{-\infty}^{\infty} h(x) dF(x).$$

It is obvious that for any random variables X and Y,

$$E(X + Y) = E(X) + E(Y).$$

(b) Variances & Covariance

- Variance of X :

$$\sigma^2 = Var(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2$$

- Covariance of X and Y :

$$\sigma_{XY} = Cov(X,Y) = E[X - E(X)][Y - E(Y)] = E(XY) - E(X)E(Y)$$

- Correlation of X and Y :

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \in [-1,\ 1]$$

Remarks

- $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$;
- If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$, hence $\mathrm{Corr}(X, Y) = 0$;
- $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, Y_j).$

B.6 Examples

Example B.1: **Matching Problem**

There are n people and n car keys. Keys are randomly put on the table. Each person randomly picks a car key. What is the expected number of people who selects their own key?

Define

$$X_i = \begin{cases} 1 & \text{if person } i \text{ picks his/her own key} \\ 0 & \text{otherwise} \end{cases}$$

Then $X = X_1 + X_2 + \ldots + X_n$ is the number of people who end up with their own key.

Then we have,

$$P(X_i = 1) = \frac{1}{n}, \forall i \Rightarrow E(X_i) = \frac{1}{n}$$
$$\therefore E(X) = 1$$
$$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{1}{n}\left(1 - \frac{1}{n}\right)$$

And for $i \neq j$,

$$E(X_i X_j) = P(X_i = 1, X_j = 1)$$
$$= P(X_i = 1 \mid X_j = 1) P(X_j = 1) = \frac{1}{n-1}\frac{1}{n}$$
$$\therefore Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = \frac{1}{n^2(n-1)}$$

Example B.2: **Probability Methods (p.14)**

Solve deterministic problems by introducing a probabilistic structure and applying probabilistic reasoning. (Graph Theory: Paul Erdos)

Consider an arbitrary graph $G = (N, E)$: N = set of nodes and E = the set of all edges. Show that there is a subset of nodes A such that at least one-half of the edges have one of their nodes in A and the other in A^C (this is known as a bipartite subgraph of G).

Proof: Randomly select bipartite subgraph of G as follows:

- randomly choose each node with probability ½, resulting a random subset $B \subseteq N$ of selected nodes;
- now construct a bipartite graph by considering all edges having one node in B and other node in B^C.

Define X as the random variable to represent the number of edges in the random bipartite graph. Let

$$X_e = \begin{cases} 1 & \text{if edge } e \text{ has exactly one node in } B \\ 0 & \text{otherwise} \end{cases}$$

Then it is clear that

$$X = \sum_{e \in E} X_e \text{ and } E(X) = \sum_{e \in E} E(X_e)$$

Since $\Pr(X_e = 1) = \Pr(X = 0) = ½$, it follows that $E(X_e) = 1/2$. Therefore, $E(X) = 1/2\, m$, where $m = |E|$, the number of edges in graph G.

Example B.3: **Probabilistic Method – A Tournament Example (Exercise 1.9)**

There are n contestants and each of $\binom{n}{2}$ pairs plays each other exactly once. A potential ranking of contestants is represented by a Hamiltonian permutation (i_1, i_2, …, i_n) where player i_j beats player i_{j+1}. Show that there exists an outcome of the tournament such that the number of Hamiltonians is at least $\frac{n!}{2^{n-1}}$.

Proof. Define the following probabilistic model:

- Player i beats player j with probability ½, independent for all pairs.

Given any permutation σ, define

$$X_\sigma = \begin{cases} 1 & \text{if } \sigma \text{ is a Hamiltonian} \\ 0 & \text{otherwise} \end{cases}$$

Then $X = \sum_\sigma X_\sigma$ is the number of Hamiltonians. Since

$$E(X_\sigma) = \Pr(X_\sigma = 1) = \left(\frac{1}{2}\right)^{n-1}$$

and there are n! permutations, it follows that

$$E(X) = \sum_\sigma E(X_\sigma) = \frac{n!}{2^{n-1}}.$$

C. Conditional Probabilities and Expectations

C.1 Definitions

For any events E_1 and E_2, the conditional probability of E_1 given E_2 is defined as

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \quad \text{(provided } P(E_2) > 0)$$

Case 1: X and Y are discrete r.v.s

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \quad \text{given } P(Y = y) > 0.$$

The conditional d.f. of X given $Y = y$ is $F(x \mid y) = \Pr(X \le x \mid Y = y)$. Hence the conditional expectation of X given $Y = y$ is

$$E(X \mid Y = y) = \int x dF(x \mid y) = \sum_x xP(X = x \mid Y = y)$$

Case 2: X and Y are continuous r.v.s

The conditional p.d.f. of X given $Y = y$ is

$$f(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

The conditional c.d.f.:

$$F(x\,|\,y) = \Pr(X \le x\,|\,Y = y) = \int_{-\infty}^{x} f(s\,|\,y)ds$$

Conditional expectation:

$$E(X\,|\,Y = y) = \int_{-\infty}^{\infty} xdF(x\,|\,y) = \int_{-\infty}^{\infty} xf(x\,|\,y)dx$$

Property:

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

C.2 Examples

Example C.1: **Matching Problem Revisited**

Now consider that keys are randomly picked in a sequential manner, i.e., person 1 picks first, person 2 picks second, … Define

$$X_i = \begin{cases} 1 & \text{if person } i \text{ ends up with his/her own key} \\ 0 & \text{otherwise} \end{cases}$$

We need to find $\Pr(X_i = 1)$ for $i = 1, \ldots, n$.

$$\Pr(X_1 = 1) = \frac{1}{n};$$

$$\begin{aligned}\Pr(X_2 = 1) &= \Pr(X_2 = 1\,|\,\text{person \#1 got key \#2})\Pr(\text{person 1 got key \#2}) \\ &+ \Pr(X_2 = 1\,|\,\text{person \#1 did not get key \#2})\Pr(\text{person \#1 did not get key \#2}) \\ &= 0 + \frac{1}{n-1}\frac{n-1}{n} = \frac{1}{n}.\end{aligned}$$

Similarly, we can show that

$$\Pr(X_i = 1) = \frac{1}{n}, \text{ for all } i = 1, \ldots, n.$$

Example C.2: **Ballot Problem**

In an election, candidate A receives n notes and candidate B receives m votes, where $n > m$. Assume that all orderings are equally likely. Show that the probability that A is always ahead of B in the count of votes is $\frac{n-m}{n+m}$.

Proof. Let $P_{n,m}$ be the desired probability. Note that

$$\begin{aligned}P_{n,m} &= \Pr(\text{A always ahead|A receives the last vote})\frac{n}{n+m} \\ &+ \Pr(\text{A always ahead|B receives the last vote})\frac{m}{n+m}\end{aligned}$$

$$\Rightarrow P_{n,m} = P_{n-1,m}\frac{n}{n+m} + P_{n,m-1}\frac{m}{n+m}$$

Applying induction on $(n+m)$, we can prove

$$P_{n,m} = \frac{n-m}{n+m} \qquad (*)$$

(i) Consider $n + m = 1 \Rightarrow P_{1,0} = 1$, so (*) holds (note that $n > m$)

(ii) Assume that (*) holds for all n and m such that $n + m = k$.

(iii) Consider any n and m such that $n + m = k + 1$. Note

$$P_{n,m} = P_{n-1,m}\frac{n}{n+m} + P_{n,m-1}\frac{m}{n+m}$$
$$= \frac{(n-1)-m}{(n-1)+m}\frac{n}{n+m} + \frac{n-(m-1)}{n+(m-1)}\frac{m}{n+m} = \frac{n-m}{n+m}$$

implying that (*) holds for k+1.

Example C.3: **Sum of a Random Number of R.V.s**

Assume that X_1, X_2,… are independent and identically distributed (i.i.d.) random variables. N is a integer-valued random variable, independent of X_i's. Define

$$Y = \sum_{i=1}^{N} X_i$$

Show that $E(Y) = E(N)E(X_1)$.

Proof. Using conditional expectation,

$$E(y \mid N = n) = E\left(\sum_{i=1}^{n} X_i \mid N = n\right) = E\left(\sum_{i=1}^{n} X_i\right) = nE(X_1)$$
$$\Rightarrow E(Y) = E[E(Y \mid N)] = E[NE(X_1)] = E(N)E(X_1).$$

Example C.4: **Prisoner Problem** (similar to Example 1.5(B))

A prisoner in a cell has three doors: the first door leads to freedom, the second door leads to a tunnel which brings him back to the cell in one day, and third door leads to a longer tunnel which brings him back to the cell in 3 days. Upon returning to the cell, the prisoner is so disoriented that he has forgotten which door(s) he has chosen before. Find the prisoner's expected time to freedom.

Solution. Define

X – the number of days to freedom;

Y – the door initially chosen

Then

$E(X|Y = 1) = 0$; $E(X|Y = 2) = 1 + E(X)$; and $E(X|Y = 3) = 3 + E(X)$.

Therefore,

$$E(X) = E(E(X \mid Y)) = \frac{1}{3}\times 0 + \frac{1}{3}(1 + E(X)) + \frac{1}{3}(3 + E(X))$$
$$\Rightarrow E(X) = 4!$$

Exercise: If the prisoner marks the door(s) that he has chosen, namely, he will not repeat the same mistake, find his expected tike to freedom under this scenario.

Example C.4: **Time Until Multiple Successive Successes**

Consider a sequence of coin flips, each with a probability p of "success". Define N_k te number of flips until k successive successes. Find $E(N_k)$.

First, for k =1, N follows a geometric distribution, i.e.,

$$P(N_1 = i) = p(1\text{-}p)^{i\text{-}1},\ \ i \geq 1.$$

Hence

$$E(N_1) = p\sum_{i=1}^{\infty} i(1-p)^{i-1} = p\frac{d}{dx}\left(\sum_{i=0}^{\infty} x^i\right)|_{x=1-p}$$
$$= p\frac{d}{dx}\left(\frac{1}{1-x}\right)|_{x=1-p} = p\frac{1}{(1-x)^2}|_{x=1-p} = \frac{1}{p}.$$

Alternatively,

$$E(N_1) = E[E(N_1 \mid Y)]$$
$$= E(N_1 \mid Y=1)P(Y=1) + E(N_1 \mid Y=0)P(Y=0)$$
$$= 1p + (1+E(N_1))(1-p) = 1 + (1-p)E(N_1)$$
$$\Rightarrow E(N_1) = \frac{1}{p}.$$

For $k > 1$, note that

$$E(N_k) = E[E(N_k \mid N_{k-1})],$$

and

$$E(N_k \mid N_{k-1} = n_{k-1}) = (n_{k-1}+1) + p \times 0 + (1-p)E(N_k)$$

Hence

$$E(N_k) = E[E(N_k \mid N_{k-1})] = E(N_{k-1}) + 1 + (1-p)E(N_k)$$
$$\Rightarrow E(N_k) = \frac{1}{p}(1 + E(N_{k-1}))$$

Therefore, by induction, we have the following,

$$E(N_k) = \sum_{i=1}^{k} \frac{1}{p^i}.$$

Example C.5: **Monty Hall Problem**

Monty Hall is a game show host. You, as a contestant, have made to the final round. There are three doors, one of which hides a prize of \$1,000,000. You have to choose a door, and receive whatever is behind that door. However, after you pick a door, but before you open it, Monty Hall stops you and opens another door – one which does not hide the grand prize (randomly if he has a choice). With this information, should you stay with your original choice, or switch to the only remaining door (or you are indifferent)?

The answer is: it is optimal switch!

Intuition: The only situation in which you lose by switching is when your original choice hides the grand prize, which has a probability of 1/3. Therefore, by switching, you have a probability of 2/3 to win!

Formal Proof. Define the following events:

A = door 1 hides the prize
B = you initially choose door 1
C = Monty Hall opens door 2

Then,

$$\Pr(\text{stay put} = \text{win}) = P(A|\, B, C)$$

Note that

$$P(A\,|\,B,C) = \frac{P(A\cap B\cap C)}{P(B\cap C)} = \frac{P(C\,|\,A\cap B)P(A\cap B)}{P(C\,|\,B)P(B)}$$
$$= \frac{P(C\,|\,A\cap B)P(A\,|\,B)}{P(C\,|\,B)} = \frac{\frac{1}{2}\times\frac{1}{3}}{P(C\,|\,B)}.$$

To compute $P(C|B)$, define

$$A_j = \text{door } j \text{ hides the prize}, \; j = 1, 2, 3.$$

Then it follows that

$$P(C\,|\,B) = \sum_{j=1}^{3} P(C\,|\,B\cap A_j)P(A_j\,|\,B) = \sum_{j=1}^{3} P(C\,|\,B\cap A_j)P(A_j)$$
$$= \frac{1}{2}\times\frac{1}{3} + 0\times\frac{1}{3} + 1\times\frac{1}{3} = \frac{1}{2}.$$

Consequently,

$$\Pr(\text{stay put} = \text{win}) = 1/3.$$

D. Characteristics of Distributions, Inequalities and Limit Theorems

D.1 Moment Generating Functions

Definition: For any random variable X, its moment generating function (m.g.f.) is defined as

$$\psi(t) = E\left(e^{tX}\right) = \int e^{tx} dF(x)$$

It is easy to see that

$$\frac{d^n(\psi(t))}{dt^n} = \psi^{(n)}(t) = E(X^n e^{tX}) \Rightarrow \psi^{(n)}(0) = E(X^n).$$

Therefore, if the m.g.f. exists, it uniquely determines the distribution of X.

Fact: For any two independent r.v.s X and Y, let $Z = X + Y$. Then

$$\psi_Z(t) = \psi_X(t)\cdot\psi_Y(t)$$

Example: Consider two independent normal distributions,

$$X_i \sim N(\mu_i, \sigma_i^2), \quad i = 1, 2.$$

The corresponding m.g.f.s are

$$\psi_{X_i}(t) = \exp\{\mu_i t + \tfrac{1}{2}\sigma_i^2 t^2\}$$

For $Z = X_1 + X_2$, we have

$$\psi_Z(t) = \psi_{X_1}(t)\psi_{X_2}(t) = \exp\{(\mu_1 + \mu_2)t + \tfrac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2\}$$

$$i.e.,\ Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Remarks:

1. The m.g.f for a r.v. may not exist. We can have a more general concept, known as *characteristic function*:
$$\phi(t) = E(e^{itX}) \quad (i = \sqrt{-1} \text{ the complex unit})$$
2. $\phi(t)$ always exists and also uniquely determines the distribution of X.
3. For nonnegative r.v.s, another alternative is the Laplace transform, defined as
$$\tilde{F}(s) = E(e^{-sX}) = \int_0^\infty e^{-sx} dF(x)$$
This integral always exists for $s = a + ib$ ($a > 0$) and also uniquely determines the distribution of X.

D.2 Probability Inequalities

Jensen's Inequality: For any r.v. X and a converse function f, it is true that

$$E(f(X)) \geq f(E(X)).$$

Markov Inequality: For any nonnegative r.v. X and a constant $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Chernoff Bounds: For any r.v. X and given its m.g.f. $\psi(t)$, then for any $a > 0$,

$$P(X \geq a) \leq \min_{t>0} e^{-ta}\psi(t)$$

$$P(X \leq a) \leq \min_{t<0} e^{-ta}\psi(t)$$

Example: Consider a Poisson distribution with parameter λ,

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, k \geq 0.$$

Note that $\psi(t) = e^{\lambda(e^t - 1)}$. Then

$$P(X \geq j) \leq \min_{t>0} e^{\lambda(e^t-1)-tj}$$

$$P(X \leq j) \leq \min_{t<0} e^{\lambda(e^t-1)-tj}$$

Therefore,

$$\text{For } j > \lambda,\ P(X \geq j) \leq e^{-\lambda}\left(\frac{\lambda e}{j}\right)^j \quad (t^* = \ln\frac{j}{\lambda})$$

$$\text{For } j < \lambda,\ P(X \leq j) \leq e^{-\lambda}\left(\frac{\lambda e}{j}\right)^j \quad (t^* = \ln\frac{j}{\lambda})$$

Chebyshev's Inequality: For any r.v. X with $E(X) = \mu$ and $Var(X) = \sigma^2$, then for any $\varepsilon > 0$,

$$P(|X-\mu|>\varepsilon) \le \frac{\sigma^2}{\varepsilon^2}.$$

D.3 Limit Theorems

Let $\{X_i\}$ be a sequence of i.i.d. random variable with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Define

$$\underline{S_n = \sum_{i=1}^{n} X_i, \quad n \ge 1.}$$

It is easy to check that

$$E\left(\frac{S_n}{n}\right) = \mu; \quad Var\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}.$$

A. Weak Law of Large Numbers (WLLN)

$$\underline{\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0, \quad \left(\frac{S_n}{n} \xrightarrow{P} \mu\right)}$$

Hint: This follows from Chebyshev's inequality immediately.

B. Strong Law of Large Numbers (SLLN)

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1, \quad \left(\frac{S_n}{n} \xrightarrow{a.s.} \mu\right) \quad \text{(almost surely)}$$

C. Central Limit Theorem (CLT)

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}s^2} ds.$$

E. Stochastic Processes

E.1 Definition

Definition: A sequence of random variables $\{X(t),\ t \in T\}$ is called a *stochastic process*.

Remarks

- t is called a time variable, $X(t)$ denotes the state of the process at time t;
- A realization of the stochastic process is called a sample path;
- When T is countable, it is called a discrete-time stochastic process; when T is a continuum, we call it a continuous-time stochastic process.

E.2 Independent Increments and Stationarity

Given $t_0 < t_1 < \ldots < t_n$, if the random variables $X(t_1) - X(t_0), \ldots, X(t_n) - X(t_{n-1})$ are independent, the process is said to have <u>independent increments</u>.

If the distribution of $X(t+s) - X(t)$ depends only on s (and not on t), the process is said to have <u>stationary increments</u>.

E.3 Examples

(1) A gambler bets the same amount on each bet. Let $X(t)$ be the gambler's wealth at time t. Does this process possess independent increments, stationary increments?

(2) If the gambler bets a fixed proportion of his wealth on each bet, then what can we say about $X(t)$?

(3) Consider the stochastic process that counts the number of cars that pass a certain intersection. Does this process possess independent increments, stationary increments?

E.4 Bernoulli Process

<u>(1) Definition</u>

Consider a sequence of i.i.d. Bernoulli trials with success probability p. Let Y_n be the number of successes in the first n trials. Then the process $\{Y_n, n \geq 1\}$ is a stochastic process. It is a counting process as well.

<u>Property</u>: $\{Y_n, n \geq 1\}$ has independent and stationary increments.

Let K be the number of trials until a success occurs. Then K follows a geometric distribution with p:

$$P(K = k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

The random variable K is commonly known as the <u>interarrival time</u>.

We can also view the stochastic process as a process in continuous time: $\{Y(t), t \geq 0\}$ with $Y(t) = Y_{[t]}$.

<u>(2) Exponential Distribution</u>

- c.d.f. $F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$
- p.d.f. $f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$

$$E(X) = 1/\lambda;\ Var(X) = 1/\lambda^2.$$

Memoryless Property of Exponential Distribution:

$$P(X > s + t \mid X > t) = P(X > s)$$

Note: exponential distribution is the only continuous distribution with this property.

E.5 Poisson Process

The Poisson process is the limiting process of the Bernoulli process when both the time between trials and the probability of success per trial converge to zero at the same rate, with constant ratio λ, which is the expected number of events per unit time.

	Bernoulli Process	Poisson Process
Time until next success	geometric	Exponential
# of success during a time period	binomial	Poisson

Poisson process can adequately describe situations in which an event is defined as the first occurrence of many improbable events. For example,

- the amount of time until the next sale of a particular product;
- the amount of time until the arrival of the next customer at the bank or ATM.