Notes for IE 599: Stochastic Processes for Industrial Engineering

*created by Joyce W. Yen*

*Autumn 2002*


**Stochastic Process:**

The difference between the realization of a stochastic process and a probability:

**Examples:**
1 Probability Review

1.1 Probability spaces

Definition 1.1 An experiment is a specific set of actions the results of which cannot be predicted with certainty.

Definition 1.2 Each possible results of the experiment defines a sample point or outcome, $\omega$.

Definition 1.3 A probability space $(\Omega, A, P)$ consists of three elements:

- sample space $\Omega$: the set of possible outcomes, must be mutually exclusive and exhaustive.

- events $A$: the set of all subsets of $\Omega$, where an event is said to occur if and only if one of the outcomes in that event occurred.

- probability function $P$: a function that assigns a probability to each event in $A$, that is $P : A \mapsto \mathbb{R}$ must satisfy:

1. $P(\Omega) = 1$,

2. $0 \leq P(A) \leq 1$ $\forall A \in A$,

3. Countable Additivity: $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ whenever $A_i \cap A_j = \emptyset$ for $i \neq j$.

Lemma 1.1 The 3 axioms given in the previous definition result in the following:

1. $P(\emptyset) = 0$,

2. If $A \subseteq B$, then $P(A) \leq P(B)$,

3. $P(A^c) = 1 - P(A)$,

4. (Boole’s inequality) $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$. 

1
1.2 Random Variables

Definition 1.4 A random variable $X$ is a quantification of outcomes, i.e. $X : \Omega \rightarrow \mathbb{R}$.

Random variables are generally denoted with a capital letter, i.e. $X, Y, Z$.

Examples:

Definition 1.5 We specify the probability distribution of a random variable (r.v.) $X$ by its cumulative distribution function (cdf) $F$, where

1. $F(x) = (X \leq x), \forall x \in \mathbb{R}$,

2. $\lim_{x \to -\infty} F(x) = 0$,

3. $\lim_{x \to \infty} F(x) = 1$,

4. $F(x)$ increases in $x$.

Definition 1.6 A discrete r.v. is a r.v. that can take on at most a countable number of distinct values, can specify probability distribution by pmf (probability mass function):

$$p(x) = P(X = x) \text{ for } x = x_1, x_2, x_3, \ldots$$

Example: cdf for discrete r.v.
Definition 1.7 A **continuous r.v.** is a r.v. whose cdf $F$ can be written as

$$F(x) = \int_{-\infty}^{x} f(y) dy,$$

for some function $f$. Here $f$ is the **probability density function** of $X$.

Example: cdf for continuous r.v.

---

Definition 1.8 **Stieltjes Integral:** We define the integral

$$\int_{A} g(x) dF(x) = \begin{cases} \sum_{x \in A} g(x) p(x) & \text{if } X \text{ is discrete with pmf } p, \\ \int_{A} g(x) f(x) dx & \text{if } X \text{ is continuous with pdf } f, \end{cases}$$

Definition 1.9 The **expectation** of a random variable $X$ is given by

$$E[X] = \int_{-\infty}^{\infty} x dF(x) = \begin{cases} \sum_{i} x_{i} p(x_{i}) & \text{if } X \text{ is discrete with pmf } p, \\ \int_{-\infty}^{\infty} x f(x) dx & \text{if } X \text{ is continuous with pdf } f, \end{cases}$$

Note also that $E[g(X)] = \int_{-\infty}^{\infty} g(x) dF(x)$. Show that $dF(x) \approx P(X \approx x)$. This definition holds since $dF(x) \approx P(X \approx x)$. 

---

3
Definition 1.10 **Variance** is a measure of spread and calculated the average squared deviation from the mean. Variance is given by

\[ \text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2. \]

Definition 1.11 For two random variables, \( X \) and \( Y \), we specify their joint probability distribution with a joint cdf as

\[ F(x, y) \equiv P(X \leq x, Y \leq y) \forall -\infty < x, y < \infty. \]

For joint continuous r.v. \( X \) and \( Y \),

\[ F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x', y')dx'dy'. \]

For joint discrete r.v. \( X \) and \( Y \),

\[ F(x, y) = \sum_{y' \leq y} \sum_{x' \leq x} p(x', y'), \]

where \( p(x', y') = P(X = x', Y = y') \) is the joint pmf of \( X \) and \( Y \).

Definition 1.12 The marginal distribution of a r.v. \( X \) or r.v. \( Y \) is given by

\[ P(X \leq x) = H(x) = \lim_{y \to \infty} F(x, y), \]

\[ P(Y \leq y) = G(y) = \lim_{x \to \infty} F(x, y). \]

Definition 1.13 Two r.v.’s \( X \) and \( Y \) are independent if \( F(x, y) = H(x)G(y) \forall x, y. \)

Lemma 1.2 If \( X \) and \( Y \) are independent, then \( E[l(X)w(Y)] = E[l(X)]E[w(Y)]. \)

Definition 1.14 The covariance of \( X \) and \( Y \) is defined as

1.3 Moment generating, Characteristic functions, and Laplace Transforms

Definition 1.15 The &textit;moment generating function&; of a r.v. $X$ is given by

$$
\Psi_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} dF(x),
$$

for all real $t$. When a moment generating function exists in a neighborhood of 0 it uniquely determines the distribution of $X$.

Definition 1.16 The &textit;characteristic function&; of a r.v. $X$ is given by

$$
\Phi_X(u) = E[e^{iuX}] = \int_{-\infty}^{\infty} e^{iux} dF(x),
$$

for all real $u$ (where $i = \sqrt{-1}$). The characteristic function uniquely defines the probability distribution $F$ and vis versa.

Definition 1.17 For non-negative r.v. $X$ the &textit;Laplace Transform&; is given by

$$
\tilde{F}(s) = E[e^{-sX}] = \int_{0}^{\infty} e^{-sx} dF(x),
$$

for all real $s > 0$.

Comment 1.1 Note that $\tilde{F}(s) = \Phi_X(is)$. 

5
Example: Suppose r.v. $X$ and $Y$ are independent. Determine the distribution of $X + Y$.

Example: Suppose $X \sim N(\lambda_1, \sigma_1^2)$, $Y \sim N(\lambda_2, \sigma_2^2)$, and $X$ and $Y$ are independent. What is the distribution of $X + Y$?
How to directly find the distribution of $X + Y$ where $X$ and $Y$ are independent? Suppose $X \sim F, Y \sim G$, and $X, Y$ are independent r.v. What is the cdf $H$ of $X + Y$?
Definition 1.18  The convolution of $F$ and $G$, written $F * G$, is

$$(F \otimes G)(z) = \int_{-\infty}^{\infty} F(z-y)dG(y).$$

Comment 1.2  We have $H(z) = (F \otimes G)(z) = (G \otimes F)(z)$ because summing is a symmetric operation.

Example:  Erlang distribution
1.4 Limit Theorems

Theorem 1.1 Strong Law of Large Numbers

Given \( X_1, X_2, X_3, \ldots \) iid (independent, identically distributed) random variables with finite expectation \( E[X_i] = \mu \), then

\[
\lim_{n \to \infty} \left( \frac{X_1 + X_2 + \ldots + X_n}{n} \right) = \mu \text{ with probability 1.}
\]

Example

Theorem 1.2 Central Limit Theorem

Given \( X_1, X_2, X_3, \ldots \) iid random variables with finite mean \( E[X_i] = \mu \) and finite variance \( \sigma^2 \) then

\[
P \left( \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma \sqrt{n}} \leq z \right) \to \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \text{ as } n \to \infty.
\]

That is for large values of \( n \), the standardized form of \( \sum_{i=1}^{n} X_i \) is distributed \( N(n\mu, n\sigma^2) \).

1.5 Some Useful Tools

Theorem 1.3 Fubini’s Theorem

1. If \( X_n \geq 0 \) a.s., then \( E \sum_{n=1}^{\infty} X_n = \sum_{n=1}^{\infty} EX_n \).

2. If \( E \sum_{n=1}^{\infty} |X_n| < \infty \), then \( E \sum_{n=1}^{\infty} X_n = \sum_{n=1}^{\infty} EX_n \).
3. If $X(t) \geq 0$, then $E \int_{-\infty}^{\infty} X(t)dt = \int_{-\infty}^{\infty} EX(t)dt$.

4. If $E \int_{-\infty}^{\infty} |X(t)|dt < \infty$, then $E \int_{-\infty}^{\infty} X(t)dt = \int_{-\infty}^{\infty} EX(t)dt$.

Note that an expectation is just a sum or integral, so the above results hold if the expectations are replaced by sums or integrals. indeed, Fubini’s theorem must be used whenever interchanging infinite sums or integrals.

**Definition 1.19** Define the **indicator variable** of the event $A$ to be

$$I_A = \begin{cases} 
1 & \text{if event } A \text{ occurs}, \\
0 & \text{if event } A \text{ does not occur}.
\end{cases}$$

Then $E[I_A] = P(A)$.

A useful tool is to multiply an expression by $I_A + I_{A^c}$ (which is identically equal to one) and then evaluate the terms individually.

**Lemma 1.3** Suppose $\{X_i\}$ are i.i.d. non-negative r.v.’s, independent of $N$, a positive integer valued r.v., then $E[\sum_{i=1}^{N} X_i] = E[N]E[X_i]$.

### 1.6 Conditional Probability and Expectations

**Definition 1.20** Let $X$ and $Y$ be discrete r.v’s

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

A visual representation of conditional probability:
Definition 1.21 The \textit{conditional expectation} of $X$ given $Y = y$ is

$$E[X|Y = y] = \sum_x xP(X = x|Y = y).$$

**Example:** Suppose we have 100 light bulbs and we choose 2 randomly. We want to test whether the light bulbs have defective elements, defective threads, or both defective element and thread. Let $X =$ number of light bulbs (out of 2) with a defective element. Let $Y =$ number of light bulbs (out of 2) with a defective thread. Suppose we have the following information:

<table>
<thead>
<tr>
<th>$P(X=x, Y=y)$</th>
<th>$y = 0$</th>
<th>$y = 1$</th>
<th>$y = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 0$</td>
<td>.7733</td>
<td>.0356</td>
<td>.0002</td>
</tr>
<tr>
<td>$x = 1$</td>
<td>.1244</td>
<td>.0562</td>
<td>.0012</td>
</tr>
<tr>
<td>$x = 2$</td>
<td>.0042</td>
<td>.0042</td>
<td>.0006</td>
</tr>
</tbody>
</table>

What is the expected number of defected elements given there is 1 defected thread light bulb? What is the expect number of defective elements?

Note: $E(X|Y)$ takes on the value $E(X|Y = y)$ with probability $P(Y = y)$.

**Theorem 1.4** $E[X] = E[E(X|Y)]$ if the expectation of $X$ is finite or $X \geq 0$ a.s.
Proof: (discrete case)

Comment 1.3 To compute \( E[X] \), first condition on \( Y \) (steps 1 and 2 below) and then uncondition on \( Y \) (step 3). That is

1. Find \( g(y) = E(X|Y = y) \).

2. Set \( E(X|Y) = g(Y) \).

3. Compute \( E[X] = E[g(Y)] \).
**Example:** Suppose a manufacturer produces a lot of $n$ items. Each item is defective with probability $P$, where $P$ is a rv. with cdf $F$ and mean $p_0$. What is the expected number of defective items in the lot?
Example: What is \( E \sum_{i=1}^{N} X_i \) where \( X_1, X_2, \ldots \) are i.i.d. r.vs with \( E[X_i] = \mu \) and are independent of \( N \), whose mean is \( m \)?
Example: Given a sequence of independent Bernoulli trials with probability \( p \) and \( 1 - p \) of success and failure, respectively, what is the expected number of trials until the first failure?
Example: Thief of Baghdad
A thief is trying to escape a chamber. Unfortunately he doesn’t remember what door he chose before. He has three doors from which to choose. If he chooses door number 1, he has his freedom. If he chooses door number 2, he will travel for one day and then return to his chamber. If he chooses door number 3, he will have 3 days of travel and then return to his chamber. What is the expected number days of travel for this thief?
Comment 1.4  Conditional probabilities are a special case of conditional expectations. Given any event $A$ and related r.v. $X \sim F$, we want to find $P(A)$. Define an indicator variable $I_A$ to be 1 if $A$ occurs and 0 otherwise.
Theorem 1.5 Law of Total Probability

Let $A$ be some event and $Y$ some r.v. with cdf $F$, then

$$P(A) = \int_y P(A|Y = y)dF(y).$$

Note $P(A) = E(P(A|Y))$. More commonly, may have seen the special case

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

Example Suppose we’re given two r.v.’s $X \sim F$ and $Y \sim G$ which are independent. What is $P(X \leq Y)$?

Example Given 2 r.v. $X \sim F$ and $Y \sim G$ which are independent. What is $P(X + Y \leq z)$?
1.7 Probability Inequalities

Lemma 1.4 Markov’s Inequality
If $X$ is a nonnegative r.v., then for any $x > 0$, $P(X \geq x) \leq EX/x$.

Lemma 1.5 Chernoff Bounds
If $X$ is a r.v. with moment generating function $M(t) = E(e^{tX})$, then for $x > 0$,

\[
P(X \geq x) \leq e^{-tx}M(t) \quad \text{for all } t > 0,
\]

\[
P(X \leq x) \leq e^{-tx}M(t) \quad \text{for all } t < 0.
\]

Lemma 1.6 Jensen’s Inequality
If $f$ is a convex function, and $EX, Ef(X) < \infty$, then

\[
Ef(X) \geq f(EX).
\]

1.8 Types of Convergence

Definition 1.22 A sequence of r.v. $(X_n : n \geq 1)$ is said to converge \textit{almost surely} (written as a.s.) or \textbf{with probability 1} to $X$ if $P(X_n \to X) = 1$.

Definition 1.23 A sequence of r.v. $(X_n : n \geq 1)$ is said to converge \textit{in probability} to $X$ (written as $X_n \xrightarrow{P} X$) if for all $\epsilon > 0, \lim_n P(|X_n - X| \geq \epsilon) = 0$.

Definition 1.24 A sequence of r.v. $(X_n : n \geq 1)$ is said to converge \textit{in distribution} to $X$ (written as $X_n \Rightarrow X$) if for all $x$ such that $P(X = x) = 0, \lim_n P(X_n \leq x) = P(X \leq x)$.

Theorem 1.6 If $X_n \to X$ a.s. then $X_n \xrightarrow{P} X$. If $\xrightarrow{P} X$, then $X_n \Rightarrow X$. 

19
2 Stochastic Processes and Brownian Motion

Definition 2.1 A **stochastic process** \( \{X(t), t \in T\} \) is a (time) indexed collection of random variables defined on the same sample space.

Definition 2.2 The r.v. \( X(t) \) is called the **state** of the stochastic process (s.p.) at time \( t \).

Definition 2.3 If \( T = \{t | t \geq 0\} \), the s.p. is called a **continuous time stochastic process**.

Definition 2.4 If \( T = \{0, 1, 2, \ldots\} \), the s.p. is called a **discrete time stochastic process**.

Definition 2.5 A realization of \( \{X(t), t \in T\} \) is a graph or a **sample path** of \( X(t) \) versus \( t \).

A picture

Theorem 2.1 **Kolmogorov’s Extension Theorem**

The distribution of a s.p. \( \{X(t), t \geq 0\} \) is uniquely specified by giving the joint distribution

\[
F_{X(t_1), X(t_2), \ldots, X(t_n)}(x_1, x_2, \ldots, x_n) = P(X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_n) \leq x_n) \text{ for all } t_1 \leq t_2 \leq \ldots \leq t_n \text{ for all } n.
\]

Note: Although the distribution is uniquely determined, the sample paths are not.
Two properties a stochastic process may possess

**Definition 2.6** A continuous time s.p. \( \{X(t), t \geq 0\} \) has **stationary increments** if \( X(t+s) - X(s) \) has the same probability distribution as \( X(t) - X(0) \) for all \( t, s \geq 0 \).

Note: If \( X(0) = 0 \), then under stationary increments, \( X(t + s) - X(s) \sim X(t) \) for all \( t, s \geq 0 \).

**Example**

**Definition 2.7** A continuous time s.p. \( \{X(t), t \geq 0\} \) has **independent increments** if for all \( t_0 < t_1 < t_2 < ... < t_n \) the r.v.’s \( X(t_1) - X(t_0), X(t_2) - X(t_1), ..., X(t_n) - X(t_{n-1}) \) are
Comment on Probabilistically restart or renews at time s

**Theorem 2.2** If the continuous time s.p. \( \{X(t), t \geq 0\} \) has independent and stationary increments and \( X(0) = 0 \), then the distribution of the r.v. \( X(t) \) for all \( t \geq 0 \) uniquely defines the distribution of the process \( \{X(t), t \geq 0\} \).

**Informal proof**
Definition 2.8 A the continuous time s.p. \( \{X(t), t \geq 0\} \) is a Brownian Motion Process if

1. \( \{X(t), t \geq 0\} \) has independent and stationary increments with \( X(0) = 0 \).

2. The state of the process at time \( t \), \( X(t) \), is normally distributed, i.e. \( X(t) \sim N(\mu(t), \sigma^2(t)) \) for all \( t \geq 0 \).

Lemma 2.1 If \( f \) satisfies \( f(s+t) = f(s) + f(t) \) for all \( s, t \geq 0 \), then \( f \) (if measurable) must be of the form \( f(t) = ct \).

Lemma 2.2 If \( f(t+s) = f(s)f(t) \) for all \( s, t \geq 0 \), then \( f(t) \) must be of the form \( f(t) = e^{ct} \).

Proof of lemma 2:

Comment 2.1 Both the mean and the variance of \( X(t) \) must be a linear function of \( t \).

1. The mean \( \mu(t) = \mu t \). \( \mu \) is called the drift coefficient.

Proof:
2. The variance of $X(t), \sigma^2(t)$ must be of the form $\sigma^2(t) = \sigma^2 t$ where $\sigma^2 > 0$.

Proof:
3 Poisson Processes

Definition 3.1 \( \{N(t), t \geq 0\} \) is a \textbf{counting process} if \( N(t) = \) the number of events that have occurred through time \( t \), or more formally \( \{N(t), t \geq 0\} \) is non-negative, integer valued, and increasing in \( t \) with probability 1. Note \( N(0) = 0 \).

A picture

\hfill

\begin{center}
\textbf{Informal Statement of a Poisson Process}
\end{center}

A Poisson Process must satisfy four postulates. Informally, we have

1. The probability that at least one Poisson arrival occurs in a small period of time \( \Delta t \) is “approximately” proportional to \( \lambda \Delta t \).

2. The number of Poisson-type arrivals happening at any prespecified time interval of fixed length is \( nt \) dependent on the “starting time” of the interval or on the total number of Poisson arrivals recorded prior to the interval.

3. The number of arrivals happening in disjoint time intervals are mutually independent random variables.

4. Given that one Poisson arrival occurs at a particular time, the conditional probability that another occurs at exactly the same time is zero.
Examples of each postulate

Definition 3.2 A counting process \( \{N(t), t \geq 0\} \) is a Poisson process if

1. \( \{N(t), t \geq 0\} \) has independent and stationary increments

2. \( N(t) \sim \text{Poisson}(\lambda t) \) for all \( t \geq 0 \). i.e.,

\[
P(N(t) = n) = \frac{e^{-\lambda t} (\lambda t)^n}{n!},
\]

for \( n = 0, 1, 2, \ldots \).

Two observations
Definition 3.3 The function $f$ is said to be $o(t)$ if $\lim_{t \to 0} \frac{f(t)}{t} = 0$.

Examples

Theorem 3.1 A counting process $\{N(t), t \geq 0\}$ is a Poisson process if and only if

1. $\{N(t), t \geq 0\}$ has independent and stationary increments with $N(0) = 0$.

2. $P(N(t) \geq 2) = o(t)$.

3. $P(N(t) = 1) = \lambda t + o(t)$.

Observations and Proof
Comment 3.1  Let $X_n$ = the time between the $n^{th}$ and $(n - 1)^{th}$ events for $\{N(t), t \geq 0\}$ for $n \geq 2$. And $X_1$ = the time until the first event.

Every event is expressible in terms of $\{N(t), t \geq 0\}$ is expressible in terms of $\{X_n, n = 1, 2, 3, ...\}$ and vis versa.

What is the distribution of $X_1$? $X_2$?
Theorem 3.2  The inter-event times $X_1, X_2, ...$ are i.i.d. $\text{Expon}(\lambda)$ r.v.’s. (i.e. with mean $\frac{1}{\lambda}$.)

Comment

Theorem 3.3  Let $X$ be a nonnegative random variable. Then $X$ is memoryless if and only if $X$ is exponentially distributed.

Comment

Proof
Comment 3.2 Let $S_n = $ time of occurrence of $n^{th}$ event in a Poisson process, called the $n^{th}$ event time, $n = 1, 2, ...$ What is the distribution of $S_n$? Given $N(t) = 1$, what is the conditional distribution of $S_1$?
Definition 3.4 Consider any \( n \) random variables \( Y_1, Y_2, \ldots, Y_n \). The \( i^{th} \) order statistic, \( Y_{(i)} \) of these r.v.'s is the \( i^{th} \) smallest among them, \( i = 1, 2, \ldots, n \).

Comment

Lemma 3.1 Suppose \( Y_1, Y_2, \ldots, Y_n \) are i.i.d. continuous r.v. with common pdf \( f(y) \). Then the joint pdf of their ordered statistics \( Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)} \) is given by

\[
f_{Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}}(y_1, y_2, \ldots, y_n) = \begin{cases} 
  n! f(y_1) f(y_2) \ldots f(y_n) & \text{for } y_1 < y_2 < \ldots < y_n, \\
  0 & \text{otherwise}.
\end{cases}
\]

Theorem 3.4 Given \( N(t) = n \), the \( n \) event times \( S_1, S_2, \ldots, S_n \) are distributed the same as the ordered statistics of \( n \) i.i.d r.v's uniformly distributed over the interval \([0, t]\).

Proof
Corollary 3.1 Given $N(t) = n$, the $n$ unordered events are independent and uniformly distributed over $[0, t]$.

Theorem 3.5 *Decomposition of Poisson Process*

Suppose that events are being generated by a Poisson process of rate $\lambda$. Whenever an event occurs it is assigned to one of $k$ streams with the $i^{th}$ stream ($1 \leq i \leq k$) being chosen with probability $p_i$ ($\sum_{i=1}^{k} p_i = 1$), independently of any previous assignments. Then the $i^{th}$ stream is itself a Poisson process of rate $\lambda p_i$. Furthermore, the streams are independent of each other.

Lemma 3.2 Let $\{N(s), s \geq 0\}$ be a Poisson process at rate $\lambda$. Suppose every event is, independently of all other events, classified as one of two mutually exclusive categories. The probability of an event occurring at time $S$ is classified into category 1 or 2 is given by $P(S)$ and $1 - P(S)$, respectively. Let $N_i(t)$ be the number of events classified type $i$ through time $t$ for $i = 1, 2$. Then $N_1(t)$ and $N_2(t)$ are independent Poisson distributed r.v.'s with mean $\lambda tp$ and $\lambda t(1 - p)$, respectively, where

$$p = \frac{1}{t} \int_{0}^{t} P(s)ds.$$

Proof
Theorem 3.6  If $X_1, X_2, ..., X_n$ are independent exponential r.v’s with parameters $\lambda_1, \lambda_2, ..., \lambda_n$, respectively, then $M = \min(X_1, X_2, ..., X_n)$ is exponentially distributed with parameter $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_n$. Also, $P(M = X_j)$ equals $\lambda_j/\lambda, 1 \leq j \leq n$.

Theorem 3.7  Recomposition of Poisson Process
Suppose we have $k$ streams of events each generated independently according to a Poisson process with the $i^{th}$ stream ($1 \leq i \leq k$) having rate $\lambda_i$. Whenever an event in one of these streams occurs, it is assigned to a new combined stream. This new stream is itself a Poisson process of rate with events being generate with rate $\lambda = \sum_{i=1}^{k} \lambda_i$. The probability of an event in the combined stream coming from stream $i$ is $\lambda_i/\lambda, 1 \leq i \leq k$. 
4 Introduction to Queueing Theory

Definition 4.1 A **queueing system** consists of an arrival stream of **customers** and a series of **servers**. When there are more customers than servers, the remaining customers are said to wait in a **queue**. If there is only a finite amount of space for waiting customers, then customers arriving to a full system depart without receiving services, and we speak of a **loss system**.

A picture

Assumptions
- arrival stream is a counting process with i.i.d. interarrival times
- service times are i.i.d.
- service time and arrival stream are independent

Notation
$A/B/s/K$ is the general notation describing a queueing system. $A$ stands for the arrival distribution ($D = \text{deterministic}, M = \text{exponential (memoryless)}, E_k = \text{Erlang-k},$ and $G = \text{general}$) and $B$ stands for the service time distribution ($D = \text{deterministic}, M = \text{exponential (memoryless)}, E_k = \text{Erlang-k},$ and $G = \text{general}$). The number of servers in
parallel is $s$ and $K$ is the number of customers the system can hold. If $K$ is not given, then it is assumed to be infinite.

**Comment:** Unless otherwise stated, assume First-In-First-Out (FIFO) service order (also known as First-Come-First-Serve). Other comment service disciplines include Last-Come-First-Serve (LCFS) and Shortest Expected Processing Time (SEPT).

**Comment 4.1 $M/G/\infty$ queuing system**

*Transient Analysis:* What is the distribution of the number of customers in the system at time $t$, $X(t)$?

*Steady-state Analysis:* What is the limiting distribution of $X(t)$ as $t \to \infty$?
Theorem 4.1 Consider an M/G/∞ queue with arrival rate λ. Service times have distribution function G and mean 1/µ < ∞. Let X(t) be the number of customers in the system at time t. Let

\[ p_t = \frac{1}{t} \int_0^t (1 - G(s))ds. \]

Then X(t) is distributed as a Poisson r.v. with mean λtp_t. The number of customers that have left the system after deceiving service by time t is independent of X(t) and is distributed as a Poisson r.v. with mean λt(1 − p_t). The limiting distribution of X(t) as t → ∞ is Poisson(λ/µ).
5 Renewal Theory

Examples
Definition 5.1  Let \( \{X_n, n = 1, 2, \ldots\} \) be a sequence of non-negative i.i.d. r.v.s with cdf \( F \). Let \( S_n = \sum_{i=1}^{n} X_i \) and set \( S_0 \equiv 0 \) and let \( N(t) = \max\{n | S_n \leq t\} \). Then \( \{N(t), t \geq 0\} \) is a renewal process.

Some Observations
Example: St. Petersburg Paradox
Let $X = 2^N$ where $N =$ number of tosses of a fair coin until the first head. What is $E[X]$?
More Observations

Let \( N(t) = \max \{ n | S_n \leq t \} \).

**Question 1** Is \( N(t) \) finite w.p.1?

**Question 2** What is the distribution of \( N(t) \)?
Proposition 5.1 Let \( \{N(t), t \geq 0\} \) be a renewal process, then \( N(t) < \infty \) for all \( t \geq 0 \) a.s. and we can write \( N(t) = \max\{n : S_n \leq t\} \).

Proposition 5.2 Let \( \{N(t), t \geq 0\} \) be a renewal process with renewal distribution \( F \). Let \( F_n \) be the \( n \)-fold convolution of \( F \) with itself. Then, \( P(N(t) \leq n) = 1 - F_{n+1}(t) \).

The expected value of \( N(t) \), \( E[N(t)] \) is called the renewal function, \( m(t) \).

Question 3 What is \( m(t) \) in terms of \( F \)?
Theorem 5.1

\[ m(t) = \sum_{n=1}^{\infty} F_n(t) \]

and

\[ m(t) < \infty \text{ for all } 0 \leq t < \infty. \]

Theorem 5.2 The renewal function \( m(t) \) uniquely determines the distribution of the renewal process, i.e. \( F \).

Example

Proposition 5.3 \( \lim_{t \to \infty} N(t) = \infty \) w.p.1.

Theorem 5.3

\[ \lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{\mu} \quad \text{w.p.1.} \]

Proof
Example Consider an $M/G/1$ queue with balking, so that if the server is busy when a customer arrives, the customer departs without receiving service, i.e. the customer is lost. What is the long-run fraction of customers that are lost?
5.1 Stopping Times

Some things to ponder
Definition 5.2 A positive integer-valued random variable \( N \) is called a stopping time for the sequence of random variables \( X_1, X_2, ... \) if the event \( \{ N = n \} \) is dependent only on \( X_1, X_2, ..., X_n \) (i.e. whether or not \( N = n \) can be determined based solely on observing \( X_1, X_2, ..., X_n \)). (definition from page 298)

Example: Gambler’s Ruin
Each play, I win $1 with probability \( p \) and lose $1 to my opponent with probability \( 1 - p \). I start with a fortune of $1 and my opponent starts with $2. We play until one of us is broke.

Let
\[
X_i = \begin{cases} 
1 & \text{if I win on the } i^{th} \text{ day,} \\
-1 & \text{if I lose on the } i^{th} \text{ day.}
\end{cases}
\]

Let \( N \) = number of plays until “ruin”, i.e. until the end of the game.

Question Is \( N \) a stopping time for \( X_1, X_2, ... \)?
Theorem 5.4  \textit{Wald’s Equation}

If $X_1, X_2, \ldots$ are i.i.d. r.v.s with finite expectation and if $N$ is a stopping time for $X_1, X_2, \ldots$ and $E[N] < \infty$, then

\[ E \sum_{n=1}^{N} X_n = ENEX. \]

Proof

Definition 5.3  Let $N$ be a non-negative integer-valued r.v. Then $N$ is a \textit{generalized stopping time} with respect to the sequence of r.v. $(X_n : n \geq 0)$ if the event $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}, \ldots$ for all $n = 0, 1, \ldots$.

Remark about Wald’s equation and generalized stopping time
Examples revisited

Gambler’s Ruin example continued
5.2 Back to Renewal Theory

Question: Is $N(t)$ a stopping time for $X_1, X_2, ...$?

**Corollary 5.1** Let $\{N(t), t \geq 0\}$ be a renewal process. Then

$$E[S_{N(t)+1}] = E \left[ \sum_{i=1}^{N(t)+1} X_i \right] = \mu(m(t) + 1).$$

**Theorem 5.5** *Elementary Renewal Theorem*

$$\lim_{t \to \infty} \frac{m(t)}{t} = \frac{1}{\mu}.$$

**Proof**
Definition 5.4 A non-negative r.v. $X$ is **lattice** if there exists a constant $d > 0$ such that

$$\sum_{n=0}^{\infty} P(X = nd) = 1,$$

i.e. all possible values of $X$ are non-negative integer multiples of $d$. The largest such constant $D$ is called the **period** of the r.v.

Examples

Theorem 5.6 **Blackwell’s Theorem**

1. If $F$ is non-lattice, then

$$\lim_{t \to \infty} m(t + a) - m(t) = \frac{a}{\mu}.$$ 

2. If $F$ is lattice with period $d$, then for $k = 1, 2, ...$

$$\lim_{n \to \infty} m(nd + kd) - m(nd) = \frac{kd}{\mu},$$

For $k = 1$, this implies that

$$\lim_{n \to \infty} E[\text{number of renewals at time } nd] = \frac{d}{\mu}.$$
Theorem 5.7 Key Renewal Theorem

If $F$ is not lattice and $h(t)$ is a directly Riemann integrable function, then

$$\lim_{t \to \infty} \int_0^t h(t-x)dm(x) = \frac{1}{\mu} \int_0^\infty h(t)dt.$$ 

A sufficient condition for $h(t)$ to be directly Riemann integrable is that

1. $h(t) \geq 0$ for all $t \geq 0$

2. $h(t)$ is nonincreasing in $t$

3. $\int_0^\infty h(t)dt < \infty$

Definition 5.5 An equation of the form

$$g(t) = h(t) + \int_0^t g(t-x)dF(x),$$

for $t > 0$ is called a renewal type equation.
Proposition 5.4 A renewal type equation has solution

\[ g(t) = h(t) + \int_0^t h(t - x)dm(x), \]

where \( m(x) = \sum_{n=1}^\infty F_n(x), \) and \( F_n(x) \) is the convolution of \( F(x) \) with itself \( n \) times.

Proof

Theorem 5.8 The Basic Renewal Theorem

If \( F \) is not lattice and \( h(t) \) is directly Riemann integrable and \( g(t) \) satisfies the renewal type equation, then

\[ \lim_{t \to \infty} g(t) = \frac{1}{\mu} \int_0^\infty h(t)dt. \]

Proof
Comment: Finding $\lim_{t \to \infty} g(t)$

**Theorem 5.9 Alternating Renewal Process**

Consider a system that can be in one of two states, called “on” and “off.” Initially it is on for a random time $X_1$, followed by being off for a random time $Y_1$. Is is then on for $X_2$ and off for $Y_2$, etc. Suppose the $X_i$’s are i.i.d. r.v.s with c.d.f $F$ and the $Y_i$’s are i.i.d. r.v.s with c.d.f $G$. Although $X_i$ is allowed to be dependent on $Y_i$, the pairs $(X_i, Y_i), i = 1, 2, ...$ are i.i.d. Let $p(t)$ be the probability that the system is on at time $t$. Then if $E[X_1 + Y_1] < \infty$ and $X_1 + Y_1$ is not lattice, then

$$\lim_{t \to \infty} p(t) = \frac{E[X_1]}{E[X_1] + E[Y_1]}.$$ 

**Proof**
Application of Alternating Renewal Process Theorem to M/G/1 queue
Application of Alternating Renewal Process Theorem to Battery Replacement
Definition 5.6 Consider a renewal process. Let $A(t) = t - S_{N(t)}$, then $A(t)$ is referred to as the age of the renewal process at time $t$. Let $Y(t) = S_{N(t)+1} - t$, then $Y(t)$ is referred to as the excess (or residual) life of the renewal process at time $t$. Define $X(t) = X_{N(t)+1}$ then $X(t) = A(t) + Y(t)$. Note that $X(t)$ does not generally have the same distribution as $X_1$.

Proposition 5.5 If the inter-event distribution of a renewal process is not lattice and $\mu < \infty$ then

$$\lim_{t \to \infty} P(Y(t) \leq x) = \lim_{t \to \infty} P(A(t) \leq x) = \frac{1}{\mu} \int_0^x 1 - F(y) dy.$$
**Proposition 5.6** If the inter-event distribution of a renewal process is not lattice and and $E[X_i^2] < \infty$ then

$$\lim_{t \to \infty} E[Y(t)] = \lim_{t \to \infty} E[A(t)] = \frac{E[X_i^2]}{2\mu}.$$ 

**Proposition 5.7** If the inter-event distribution of a renewal process is not lattice and $\mu < \infty$ then

$$\lim_{t \to \infty} P(X(t) \leq x) = \frac{1}{\mu} \int_0^x ydF(y),$$

and

$$\lim_{t \to \infty} E[X(t)] = \frac{E[X_i^2]}{\mu}.$$
6 Discrete Time Markov Chains

Definition 6.1 A Markov Process is a stochastic process \( \{X(t), t \in T\} \) such that for all \( t_1 < t_2 < \ldots < t_{n+1} \in T \)

\[
P(X(t_{n+1}) \leq x | X(t_n) = x_n, \ldots, X(t_1) = x_1) = P(X(t_{n+1}) \leq x | X(t_n) = x_n)
\]

for all \( x_1, x_2, \ldots, x_n, x \). Note that the above property is called the Markov property and in effect says that given the present, the future is independent of the past.

Theorem 6.1 Every stochastic process with independent increments and \( X(0) = 0 \) is a Markov process.

Proof

Comment and a Picture
Definition 6.2 \( \{X_n, n = 0, 1, 2, \ldots\} \) is a **discrete time Markov Chain** if the state space \( S \) (set of all possible values of \( X_n \)) is discrete, i.e. finite or countably infinite.

Markov property in this case:

Definition 6.3 If \( P(X_{n+1} = j | X_n = i) \) does not depend upon \( n \), the Markov chain is said to be **homogeneous**. We will assume time homogeneity for all future chains.

Notation

Definition 6.4 Let \( P_{ij} = P(X_{n+1} = j | X_n = i), i, j = 0, 1, 2, \ldots \). Then \( P_{ij} \) is said to be the **1-step transition probability from state** \( i \) **to state** \( j \). The \( P_{ij} \)'s are often written as a single matrix \( P \) where the \( (i, j)^{th} \) element of \( P \) is \( P_{ij} \). \( P \) is called the **1-step transition probability matrix**.
Definition 6.5 The 1-step transition matrix is often drawn pictorially as a \textit{transition diagram}. In this diagram the nodes represent the different states and the arcs are labelled with the transition probabilities between those two states.

Definition 6.6 Let $a_i = P(X_0 = i)$ for all $i = 0, 1, \ldots$. The \textit{initial probability row vector} $a = (a_0, a_1, \ldots)$ is called the \textit{initial state distribution}.

Theorem 6.2 Together $a$ and $P$ completely determine the distribution of the Markov chain.

Proof
Example: Discrete Time Discrete Random Walk
More Examples
Examples from Queueing Theory
Definition 6.7 \( P^{(n)}_{ij} = P(X_n = j|X_0 = i) = P(X_{m+n} = j|X_m = i) \) is the probability that if a Markov chain starts in state \( i \) then after \( n \) transitions it will be in state \( j \). This is called the \textit{n-step or n\textsuperscript{th}-order transition probability}. Just like the 1-step transition, these \( n \)-step transitions are often written in matrix form as \( P^{(n)} \).

Chapman-Kolmogorov Equation
Proposition 6.1 The $n$-step transition matrix may be found by multiplying $P$ (the 1-step transition matrix) by itself $n$ time, i.e. $P^{(n)} = P^n$.

Proposition 6.2 The probability the Markov chain is in state $j$ after $n$ transition is

$$a_j^{(n)} = \sum_{i=0}^{\infty} p_{ij}^{(n)} a_i$$

and the vector of state probabilities $a^{(n)}$ may be found by $a^{(n)} = a P^n$.

Example

Definition 6.8 A state $j$ is said to be accessible from state $i$, written $j \leftarrow i$ (or $i \rightarrow j$), if $P_{ij}^n > 0$ for some $n \geq 0$.

Example: Gambler’s Ruin
Definition 6.9 Two states, $i$ and $j$, are said to communicate, written $j \leftrightarrow i$ (or $i \leftrightarrow j$) if $i \rightarrow j$ and $j \rightarrow i$.

Example: Gambler’s Ruin

Theorem 6.3 Communication is an equivalence relation, i.e.

1. $i \leftrightarrow i$ (reflexive)

2. if $i \leftrightarrow j$, then $j \leftrightarrow i$ (symmetric)

3. if $i \leftrightarrow k$ and $k \leftrightarrow j$, then $i \leftrightarrow j$ (transitive)

Proof
Definition 6.10 We may partition the state space into mutually exclusive and exhaustive classes such that two states communicate if and only if they are in the same class. We do this by starting with any state $i$ and forming the class $C_i$ of all states that communicate with $i$. Then we repeat for any state not in $C_i$, and so on. (The $C_i$’s are called communicating classes.)

Example: Random walk with Absorbing Barriers

Definition 6.11 A Markov chain is said to be irreducible if all the states communicate (i.e., there is only one communicating class); otherwise it is reducible. So a Markov chain is irreducible if, for all states $i$ and $j$, there is an $n \geq 0$ such that $P(X_n = j | X_0 = i) > 0$.

Definition 6.12 A property of a state is called a class property if either all members of the class have the property or if no members have the property.

Definition 6.13 The period of a state $j$, written $d(j)$, is the largest positive integer $d$ such that every positive $n$ with $P_{jj}^{(n)} > 0$ is an integer multiple of $d(j)$. In other words, $d(j)$ is the greatest common divisor (gcd) of all $n > 0$ such that $P_{jj}^{(n)}$. 

66
Examples
Theorem 6.4 If \( i \leftrightarrow j \) then \( d(i) = d(j) \), i.e. period is a class property.

Lemma 6.1 If \( n \) is gcd of positive integers \( k_1, k_2, ..., k_l \) and if \( m \) divides \( k_1, k_2, ..., k_l \), then \( m \) divides \( n \).

Example

Proof of Theorem 6.4
Definition 6.14 If the period of a state is 1, the state is called \textbf{aperiodic}.

Definition 6.15 Define the \textbf{first passage probabilities}

\[
f^n_{ij} = P(X_n = j, X_k \neq j, k = 1, 2, ..., n - 1 | X_0 = i),
\]

\(n = 0, 1, 2, ...\). Here \(f^n_{ij}\) is the probability that the Markov chain is in state \(j\) for the first time after \(n\) transitions, having started in state \(i\). Define \(f^0_{ij} = 0\).

Proposition 6.3 We have that for all \(i, j\) and all \(n \geq 0\), \(f^n_{ij} \leq P^n_{ij}\), for all \(i, j\) and all \(n \geq 1\),

\[
P^n_{ij} = \sum_{k=1}^{n} f^k_{ij} P^{n-k}_{jj}.
\]

Discussion
Definition 6.16 Let
\[ f_{ij} = \sum_{k=1}^{\infty} f_{ij}^k. \]
Then \( f_{ij} \) is the probability of ever going from \( i \) to \( j \). (Note: it could be zero.)

Definition 6.17 A state \( j \) is said to be \textbf{recurrent} if \( f_{jj} = 1 \), otherwise it is \textbf{transient}.

Proposition 6.4 State \( j \) is recurrent if and only if
\[ \sum_{n=1}^{\infty} P_{jj}^{(n)} = \infty. \]

Proof
Corollary 6.1 If state $j$ is transient then

$$\sum_{n=1}^{\infty} P_{ij}^{(n)} < \infty$$

for all initial states $i$.

Proof

Proposition 6.5 If $i \leftrightarrow j$ and $j$ is recurrent, then $i$ is recurrent (i.e. recurrence, and hence transience, are class properties.)

Proof
Example: Unrestricted Random Walk
Proposition 6.6 If \( i \leftrightarrow j \) and \( i \) is recurrent, then \( f_{ij} = 1 \).

Definition 6.18 Let \( \mu_{jj} \) be the expected number of transitions needed to return to state \( j \) from state \( j \), so that

\[
m_{jj} = \begin{cases} 
\infty & \text{if } j \text{ is transient,} \\
\sum_{n=1}^{\infty} n f_{jj}^n & \text{if } j \text{ is recurrent.}
\end{cases}
\]

Note that \( \mu_{jj} \) is also referred to as the \textit{mean recurrence time} for state \( j \).

Theorem 6.5 For any state \( j \),

\[
\lim_{n \to \infty} P_{jj}^{(nd(j))} = \frac{d(j)}{\mu_{jj}}.
\]

Note that if \( j \) is aperiodic and \( i \leftrightarrow j \), then this implies that

\[
\lim_{n \to \infty} P_{jj}^{(n)} = \frac{1}{\mu_{jj}}.
\]

Definition 6.19 Define the \textit{limiting probability} of state \( j \) as

\[
\pi_j = \lim_{n \to \infty} P_{jj}^{(nd(j))} = \frac{d(j)}{\mu_{jj}}.
\]

Definition 6.20 A recurrent state \( j \) is said to be \textbf{positive recurrent} if \( \pi_j > 0 \) (equivalently, if \( \mu_{jj} < \infty \)) and \textbf{null recurrent} if \( \pi_j = 0 \) (equivalently, if \( \mu_{jj} = \infty \)).

Proposition 6.7 Positive (null) recurrence is a class property.

Definition 6.21 A positive, aperiodic recurrent state is called \textbf{ergodic}. \((\pi_j > 0)\)

Definition 6.22 An irreducible, aperiodic, positive recurrent Markov chain is an \textbf{ergodic Markov chain}.

Definition 6.23 A class is \textbf{closed} if it is not possible to leave it. That is class \( C \) is closed if for all \( j \in C, i \notin C, P_{ji}^n = 0 \) for all \( n \).
Theorem 6.6  Every recurrent class is closed.

Outline of the proof

Definition 6.24  A probability distribution \( \hat{\pi} = (\hat{\pi}_0, \hat{\pi}_1, ...) \) is said to be a stationary probability distribution if

1. For all \( j \), \( \hat{\pi} = \sum_{i=0}^{\infty} \hat{\pi}_i P_{ij} \) (i.e. \( \pi = \pi P \))

2. \( \sum_{i=0}^{\infty} \hat{\pi}_i = 1 \) (i.e. \( \pi \) is a probability distribution over the states.)

Theorem 6.7  If a Markov chain is ergodic (i.e. irreducible, aperiodic, positive recurrent), then the limiting probabilities (\( \pi_0, \pi_1, ... \)) form a stationary probability distribution and there are no other stationary probability distributions.

“Proof”
Example: Social Mobility

**Theorem 6.8** If an irreducible aperiodic Markov chain is not positive recurrent, then no stationary probability distribution exists.