Review of Probability Theory

Conditional Probability P{B|A} = P{event B occurs given that A has already occurred}

If B⊥A (i.e. event B is independent of A), then P{B|A}=P{B} Since occurrence of A will not affect the chances of occurrence of B, given that B is independent of A.

Baye's Rule Denoting P{AB} = P(events A and B both occur}, we get that -

P{AB} = P{A}P{B|A} = P{B}P{A|B}

This leads to Baye's Rule $\boxed{P\{B \mid A\} = \dfrac{P\{B\}P\{A \mid B\}}{P\{A\}}}$

This relationship is extremely useful in probability calculations such as in changing conditioning of events. (Example: Changing from *a priori* to *a posteriori* probabilities)

Mutually Exclusive Events Note that if A and B are *Mutually Exclusive* events then P{AB}=0 as in that case, probabilistically, events A and B do not occur together

We define P{AÇB} as the probability of the *union* of events A and B. This is the event when either A occurs or B occurs or both occur

By definition, P{AÇB} = P{A} + P{B} – P{AB}
= P{A} + P{B} *if A and B are mutually exclusive events*
= P{A} + P{B} – P{A}P{B} *if A^B, i.e. are independent events*

Complementary Event For an event A, the complementary event A^C refers to the event where A does not occur

P{A^C} = 1 – P{A}

Note also that, for any event B, P{B} = P{B|A}P{A} + P{B|A^C}P{A^C}

Discrete Random Variables A discrete random variable X takes on discrete values x_i with probabilities P{X=x_i}>0 for i=1, 2, 3, ... and P{X∈(x_1, x_2, ...)}=1

Examples of Distributions for Discrete Random Variables -

1. Binomial Distribution $P\{X = x\} = \binom{n}{x} p^x (1-p)^{n-x}$ x=0, 1, ..., n

2. Poisson Distribution $P\{X = x\} = e^{-l} \dfrac{l^x}{x!}$ for x=0, 1, 2, ...∝

Continuous Random Variables

In this case, the random variable X is not limited to discrete values but can take on any value x in a range [x_1, x_2], i.e. $x \in [x_1, x_2]$, where the probability of the random variable X lying between x and x+dx is given by $P\{x £ X £ x+dx\} = f_X(x)dx$ where $f_X(x)$ is referred to as the *Probability Density Function* (*pdf*) of the random variable X

The *Cumulative Distribution Function* $F_X(x)$ may also be used to describe the probability distribution of a continuous random variable. This is defined as $F_X(x) = P\{X \le x\} = \int_{-\infty}^{x} f_X(x)dx$

Note also that $f_X(x) = \dfrac{dF_X(x)}{dx}$

Examples of Distributions of Continuous Random Variables -

1. Normal Distribution - $f_X(x) = \dfrac{1}{\sqrt{2ps^2}} e^{-\frac{(x-m)^2}{2s^2}}$ for $-\infty \le x \le \infty$

2. Uniform Distribution - $f_X(x) = \dfrac{1}{b-a}$ for $a \le x \le b$
   $f_X(x) = 0$ otherwise

3. Exponential - $f_X(x) = me^{-mx}$ for $0 \le x \le \infty$
   $f_X(x) = 0$ otherwise

Memoryless Property

Let the random variable $x \ge 0$ be the length of service provided to a customer when service starts from the time instant t=0. Consider a customer who is still in service at time t and let {(X-t) | X>t} be the remaining service time for that customer. [Not that this random variable is the remaining service time when the customer is examined at time t, given (of course) that the customer is still in service at time t - i.e. the customer's service time X is greater than t]

Note that we can write P{(X-t)>x, X>t} = P{(X-t)>x | X>t}P{X>t}
and that trivially P{(X-t)>x, X>t} = P{(X-t)>x} since x and t are both positive

Therefore

$$P\{(X-t) > x \mid X > t\} = \frac{P\{(X-t) > x\}}{P\{X > t\}} = \frac{P\{(X > x+t\}}{P\{X > t\}} = \frac{1-F_X(t+x)}{1-F_X(t)}$$

If the service distribution is *memoryless*, then that implies that when we examine the customer (who started service at t=0 and is still in service) at time t, the service given in the past during the interval (0,t) is forgotten! If this is indeed the case, then it follows that P{(X-t)>x | X>t} = P{X>x}= 1- $F_X(x)$

Using this, we get that for a memory less distribution, $\boxed{[1-F_X(t+x)] = [1-F_X(x)][1-F_X(t)]}$

The *Exponential Distribution* is an example of a memory less distribution. Note that for this distribution,

$f_X(x) = me^{-mx}$ and $F_X(x) = 1 - e^{-mx}$ for $x \geq 0$

Therefore, $[1 - F_X(t+x)] = e^{-m(t+x)} = e^{-mx}e^{-mt} = [1 - F_X(x)][1 - F_X(t)]$ as required by the Memoryless Property

For integer valued random variable X={0, 1, 2, ...∝}, the corresponding memory less distribution is the *Geometric Distribution* where $P\{X=n\} = q^n(1-q)$ leading to $P\{X \geq x\} = q^x$ for x=0, 1, 2, ..., ∝

This may be verified by noting that $P\{X \geq x+N\} = q^{x+N} = q^x q^N = P\{X \geq x\}P\{X \geq N\}$ as required.

Note that the memory less property of the *exponential* and *geometric* distributions make them easy to handle. These are therefore very useful in the analytical modeling of queuing systems and computer communications.

Joint Distributions The joint distribution for continuous random variables X and Y is given in the following form -

cumulative distribution function (cdf) $F_{XY}(x,y) = P\{X \leq x, Y \leq y\}$

probability density function (pdf)

$$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y) = P\{x \leq X \leq x + dx, y \leq Y \leq y + dy\}$$

Note that $F_X(x) = F_{XY}(x, \mu)$ $F_Y(y) = F_{XY}(\mu, y)$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy \text{ and } f_Y(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx$$

Note also that if X⊥Y, then $f_{XY}(x,y)=f_X(x)f_Y(y)$ and $F_{XY}(x,y)=F_X(x)F_Y(y)$

Functions of Random Variables For a random variable X, U=u(X) may be defined as another random variable which is a function of the random variable X.

If u(x) is differentiable and monotone, then the pdf of the random variable U may be easily found as $f_U(u)=f_X(x)/|u'(x)|$ or $f_U(u)|du|=f_X(x)|dx|$

If u(x) is not monotone then one has to be more careful as the function $X=u^{-1}(U)$ may have multiple roots and these should be accounted for while finding $f_U(u)$. If u(x) is not differentiable at some point in the range, then delta functions will arise in the pdf of U.

Example: Consider $u(x) = x^2$ to generate the random varaible U from the random variable X, where X∈[-1, 1]

From the form of the function u(x) and the range of X, we can see that -

$f_U(u)du = f_X(x)dx + f_X(-x)|dx|$

Since $du=2xdx$, we get that, *dx=du/(2Öu)*

Therefore, $$f_U(u) = \frac{f_X(\sqrt{u}) + f_X(-\sqrt{u})}{2\sqrt{u}}$$

If we consider the case where the random variable X is uniformly distributed in [-1, 1], then

$f_X(x)=0.5$	for *-1£x£1*
$= 0$	otherwise

This will then give $f_U(u) = \dfrac{1}{2\sqrt{u}}$ and $F_U(u) = \sqrt{u}$ for *0£u£1*

This approach may be extended using the *Jacobian* for the case of functions of more than one variables.

As an example of this, consider $z=g(x,y)$ and $w=h(x,y)$ [Note that if only one function is given then the second function may be arbitrarily defined.]

Let $g(x_i, y_i) = z$ and $h(x_i, y_i)=w$ i.e. (x_i, y_i) are the real solutions to these equations for given (z,w)

The *Jacobian* $J(x,y)$ is then defined as $J(x,y) = \begin{vmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \\ \dfrac{\partial h}{\partial x} & \dfrac{\partial h}{\partial y} \end{vmatrix}$

which may be used to get $f_{ZW}(z,w) = \dfrac{f_{XY}(x_1, y_1)}{J(x_1, y_1)} + \ldots\ldots\ldots\ldots + \dfrac{f_{XY}(x_n, y_n)}{J(x_n, y_n)} + \ldots\ldots$

where $\{x_i, y_i\}$ $i=1, 2,\ldots\ldots, n,\ldots\ldots$ are the roots of $z=g(x,y)$ and $w=h(x,y)$

Note that if there are no real solutions for some values of (z,w), then for these $f_{ZW}(z,w)=0$

Using Expectations and Transforms

Expectations If X is a random variable, then we can define expectations for various functions of X as -

$$\overline{g(X)} = E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x)dx \quad \text{for Continuous Random Variables}$$

$$= \sum_{x} g(x)P\{X = x\} \quad \text{for Discrete Random Variables}$$

Some useful results:

E{*cg(X)*} = *c*E{*g(X)*}
E{*g(X)*+*h(Y)*} = E{*g(X)*} + E{*h(Y)*}
For X⊥Y, E{*g(X)h(Y)*} = E{*g(X)*}E{*h(Y)*}

The n[th] moment of the random variable X is defined as $E\{X^n\} = \overline{X^n}$

Specifically -

$$\overline{X} = mean$$

$$s_X^2 = \text{var}\,iance = E\{(X - \overline{X})^2\} = \overline{X^2} - \overline{X}^2$$

s_X is referred to as the *standard deviation* of X

Note that s_X is really indicative of the *dispersion* of the random variable X around its mean

Laplace Transform This is a convenient tool to use for continuous random variables X such that X≥0 and is defined as

$$\tilde{F}_X(s) = L(f_X(x)) = E\{e^{-sX}\} = \int_0^{\infty} e^{-sx} f_X(x)dx$$

Some useful properties of this transform are -

(a) Moment Generating Property $\overline{X^n} = (-1)^n \dfrac{d^n}{ds^n} \tilde{F}_X(s)\,|_{s=0}$

(b) Given a transform, inverting it will provide the corresponding *pdf* $f_X(x)$ of X

(c) *Multiplication* in the Transform Domain would correspond to *Convolution* in the r.v. domain and *vice-versa*.

For example

$$L\big(f_1(x) * f_2(x)\big) = L\left(\int_{-\infty}^{\infty} f_1(x) f_2(x - x)dx\right) = \tilde{F}_1(s)\tilde{F}_2(s)$$

$$\& \quad L^{-1}\big(\tilde{F}_1(s) * \tilde{F}_2(s)\big) = f_1(x) f_2(x)$$

(d) *Transform of the sum of independent random variables = Product of the individual transforms*
If random variables X and Y are such that X⊥Y, then we can show that -

$$\tilde{F}_{X+Y}(s) = E\{e^{-s(X+Y)}\} = \tilde{F}_X(s)\tilde{F}_Y(s)$$

Characteristic Function (Fourier Transform) This type of transform is useful for continuous random variables where X may take on negative values, i.e. $-\propto \le X \le \propto$

This is defined as $f_X(v) = E\{e^{jvX}\} = \mathcal{F}\,[f_X(x)]$ i.e. the Fourier Transform of $f_X(x)$

Properties similar to those described for Laplace Transforms above are also applicable here -

(a) $\overline{X^n} = (-j)^n \dfrac{d^n}{dv^n} f_X(v)|_{v=0}$ Moment Generating Property

(b) Multiplication in Transform Domain corresponds to Convolution in the random variable domain and vice versa

(c) The characteristic function of the sum of independent random variables is the product of the characteristic functions of the individual random variables

Generating Function or *Probability Generating Function* (Z-Transform of the Probability Distribution)

This transform is used for a discrete random variable X, such that X≥0. It is defined as -

$$G_X(z) = E\{z^X\} = \sum_{i=0}^{\infty} p_i z^i = \mathcal{Z}\,[P\{X=i\}] \qquad \text{where } p_i = P\{X = i\}$$

This also has properties similar to the transforms given earlier

(a) Moment Generating Property $\overline{X} = G_X'(1)$ $\overline{X(X-1)} = G_X''(1)$ .... etc.

(b) If $X_1 \perp X_2$ and $Y = X_1 + X_2$,

then $G_Y(z) = G_{X_1}(z)G_{X_2}(z)$

and $p_Y(y) = P\{Y = y\} = p_{X_1+X_2}(y) = \sum_{x_1=0}^{\infty} p_{X_1}(x_1)\, p_{X_2}(y - x_1)$

Covariance & Correlation Consider the random variables X and Y.

$$s_{X+Y}^2 = E\{(X + Y - \overline{X} - \overline{Y})^2\} = E\{(X - \overline{X})^2\} + E\{(Y - \overline{Y})^2\} + 2E\{(X - \overline{X})(Y - \overline{Y})\}$$
$$= s_X^2 + s_Y^2 + 2Cov(X,Y)$$

Note that the *Covariance* of X and Y, *Cov(X, Y)* is defined as -

$$Cov(X,Y) = E\{(X - \overline{X})(Y - \overline{Y})\} = \overline{XY} - \overline{X}.\overline{Y}$$

(a) If X⊥Y, then *Cov(X, Y) = 0* *since* $\overline{XY} = \overline{X}.\overline{Y}$

(b) If *Cov(X, Y) = 0*, then the random variables X and Y are *uncorrelated*

Note that X⊥Y, implies *Cov(X, Y) = 0*. However, *Cov(X, Y) = 0* does not imply X⊥Y but the much weaker condition that X and Y are uncorrelated

Stochastic (Random) Processes

Definition of a Stochastic Process: $\{X(t);\ t\in T\}$ is a stochastic process if X(t) is a random variable for each t in the index set T - usually t indicates time.

Different ways of classifying Stochastic Processes:

"*Continuous Time Processes*" when T is an interval and all times in that interval are possible choices for t.

"*Discrete Time Processes*" when T is a set of discrete time points with $X_n = X(t_n)$ constituting the process.

"*Continuous State Process*" when X(t) can have a continuum of values possibly within a fixed range.

"*Discrete State Process*" when X(t) can only have one of a discrete set of values.

Note that X(t)=x implies that the random process is in state x at time t. The Stochastic Process X(t) is said to be *well defined* if the joint distribution of the random variables $X(t_1)$, $X(t_2)$, ..., $X(t_k)$ can be determined for every finite set of time instants $t_1, t_2, ..., t_k$

In a stochastic sense, the random process is completely specified if the joint distribution $F_{\tilde{X}}(\tilde{x},\tilde{t})$

$$F_{\tilde{X}}(\tilde{x},\tilde{t}) = P\{X(t_1)\le x_1,...............X(t_n)\le x_n\}$$

exists for (a) all $\tilde{x} = (x_1,............x_n)$

(b) all $\tilde{t} = (t_1,...........t_n)$

(c) all possible values of *n*

If $F_{\tilde{X}}(\tilde{x},\tilde{t})$ is known as above, then all possible stochastic dependencies between sample values of X(t) may be found. This, however, is usually hard to get. Typically, only some limited dependencies will be known and different *Stochastic Processes* are classified based on these known dependencies. Examples of this leading to special processes are given next.

(a) Independent Process: For this type of process, $\{X(t_n)\}$ or $\{X_n\}$ are independent random variables,

and therefore, $$f_{\tilde{X}}(\tilde{x};\tilde{t}) = \prod_{i=1}^{n} f_{X_i}(x_i;t_i)$$

Note that this is really a trivial case of a process as there are no dependencies between the various X_is.

(b) Stationary Processes: These are processes where the joint distribution of the random variables corresponding to a set of time points is invariant to a time shift of all the time points. The process is considered *Strictly Stationary* if the property holds for any choice of the number of time points. If the property holds for any choice of *n* time points or less but not for any choice of *n+1* time points then the process is referred to as being *Stationary of Order n*. The process is referred to as being *Wide Sense Stationary (WSS)* if (a) E{X(t)} is independent of t and (b) $E\{X(t)X(t+\tau)\}$ depends only on τ and not on t.

Stationary Processes will not be used in our description of queues and will not be considered further here.

(c) Markov Processes: Markov Processes are ones for which the *Markov Property* (given below) holds. This property states that -

$$P\{X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, ..., X(t_1) = x_1\} = P\{X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n\}$$

Note that when this property is satisfied, the state of the process/system at time instant t_{n+1} depends only on the state of the process/system at the previous instant t_n and *not* on any of the earlier time instants.

Alternatively, one can say that a process is termed a *Markov Process* if, given the present state of the process, its future evolution depends is independent of the past of the process. This effectively implies a *one-set dependence feature* for the *Markov Process* where older values are forgotten. Restricted versions of this property leads to special cases, such as -

(a) Markov Chains over a Discrete State Space

(b) Discrete-Time and Continuous- Time Markov Processes and Markov Chains

Markov Chains: The discrete random variables $\{X_n\}$ form a Markov Chain if the probability that the next state is x_{n+1} depends only on the current state x_n and not on any previous values.

For the *Discrete Time* case, state changes are pre-ordained to occur only at the integer points 0, 1, 2, ..., n (that is at the time points t_0, t_1, t_2,......, t_n). For the *Continuous Time* case, state changes may occur anywhere in time.

Homogenous Markov Chain: A Homogenous Markov Chain is one where the transition probabilities $P\{X_{n+1}=j \mid X_n=i\}$ is the same for all n. Note that one can then write that -

Transition Probability from state i to state j = p_{ij} = $P\{X_{n+1}=j \mid X_n=i\}$ "n

It should be noted that for a homogenous Markov chain, the transition probability depends only on the terminal states (i.e. the initial state *i* and the final state *j*) but does not depend on when actually the transition (*i* ® *j*) occurs.

(d) Semi-Markov Processes: In a Semi-Markov Process, the distribution of time spent in a state can have an arbitrary distribution but the one-step memory feature of the Markovian property is retained. We will find processes of this type useful in some of our analyses.

(e) Birth-Death Process: A Birth-Death Process is a special type of discrete-time or continuous-time Markov Chain with the restriction that at each step, the state transitions, if any, can occur only between neighboring states.

(g) Renewal Processes: These are related to random walks except that our interest here lies in counting the number of transitions that take place as a function of time.

State at time t = Number of transitions in (0, t)
where $(t_i - t_{i-1})\ \forall\ i$ are i.i.d. random variables

Let $X_i = (t_i - t_{i-1})\ \forall\ i$ be a set of i.i.d. random variables. Subject to the conditions that they are independent and have identical distributions, the random variables $\{X_i\}$ can have any distribution. Note that this corresponds to a *Semi-Markov Process*.