**Summary of Random Variable Concepts** **April 19, 2000**

This is a list of important concepts we have covered, rather than a review that derives or explains them.

### The first and primary viewpoint:

A *random process* is an indexed collection of random variables.

That is, a random process is $\{X(t): t \in I\}$, where $I$ is an infinite collection of times (or time indices). For each time t in I, there is a random variable denoted $X(t)$. Sometimes we write $X_t$ instead of $X(t)$.

The random process itself is denoted $\{X(t): t \in I\}$, or by one of the shorthands: $\{X(t)\}$, $X(t)$, $X$ or any of the previous with $X_t$ replacing $X(t)$.

### Discrete and Continuous Time Random Processes

*Discrete-time* random processes have $I = \{0,1,2,...\}$, $I = \{1,2,3, ...\}$ or $I = \{...,-2,-1,0,1,2,...\}$.

For discrete-time processes we often use the following notations $\{X(n)\}$, $\{X_n\}$ or replace n by some other letter such as i,j,k,m that suggests an integer.

*Continuous-time* random processes have $I = [a,b]$, where $-\infty \le a < b \le \infty$.

### Probability Distribution of the Random Process $\{X(t)\}$

To know the *probability distribution* of a random process is to know the joint distribution of every finite collection of its random variables; i.e. to know the joint distribution of $X(t_1), X(t_2),...,X(t_n)$ for every n and $t_1,...,t_n \in I$. For example, it suffices to know the joint cdf, pdf or cmf of $X(t_1), X(t_2),...,X(t_n)$.

Knowing the probability distribution of a random process, we can compute the probability of any event involving the random process, or any conditional probability, or any expected value.

### Partial Characterizations of the Probability Distribution of a Random Process $\{X(t)\}$

The probability distribution of a random process is an awful lot to have to know or to specify. Consequently, we often work with one or more of the following partial descriptions of its probability distribution.

1. First-order distribution: This consists of the marginal distributions of every individual random variable $X(t)$, $t \in I$, as specified, for example, by cdf, pdf or pmf of every individual random variable.

2. Second-order distribution: This consists of the joint distribution of every pair of random variables $(X(t),X(s))$ $t,s \in I$, as specified, for example by the joint cdf, pdf or pmf of every pair of random variables.

3. nth-order distribution: This consists of the joint distribution of every collection of n random variables $(X(t_1),...,X(t_n))$, $t_1,...,t_n \in I$, as specified, for example by specifying the joint cdf, pdf or pmf of collection of n random variables.

4. mean function:

$$m_X(t) = E\, X(t)$$

5. power function:

$$P_X(t) = E\, X^2(t)$$

6. autocorrelation function:

$$R_X(t_1,t_2) = \text{correlation between } X(t_1) \text{ and } X(t_2) = E\, X(t_1)\, X(t_2)$$

7. autocovariance function:

$$C_X(t_1,t_2) = \text{covariance of } X(t_1) \text{ and } X(t_2) = E\left[(X(t_1)-m_X(t_1))(X(t_2)-m_X(t_2))\right]$$

$$= R_X(t_1,t_2) - m_X(t_1)m_X(t_2) \quad \text{(we derived this formula in class)}$$

**Some Discrete-Time Examples**

1. IID Random Process (Independent and Identically Distributed):

   Like the name suggests, the random variables are independent and identically distributed. For IID random processes the complete probability distribution is determined from the probability distribution of just one random variable.

2. Bernoulli Random Process:

   A binary IID random process.

3. Moving Average (MA) Random Process:

$$Y_n = \sum_{i=0}^{M-1} b_i \ X_i$$

   where $\{X_n\}$ is an IID process, $M \geq 2$ is an integer, and $b_0,...,b_{M-1}$ are parameters.

4. Autoregressive Random (AR) Process:

$$Y_n = \sum_{i=1}^{N} a_i \ Y_{n-i} + X_N$$

   where $\{X_n\}$ is an IID process, $X_n$ is independent of $Y_{n-1},Y_{n-2},...$, $N \geq 1$ is an integer, and $a_0,...,a_N$ are parameters.

5. Autoregressive Random Moving Average (ARMA) Process:

$$Y_n = \sum_{i=1}^{N} a_i \ Y_{n-i} + \sum_{i=0}^{M-1} b_i \ X_i$$

   where $\{X_n\}$ is an IID process, $X_n$ is independent of $Y_{n-1},Y_{n-2},...$, $N \geq 1$ is an integer, $a_0,...,a_N$ are parameters, $M \geq 2$ is an integer, and $b_0,...,b_{M-1}$ are parameters.

6. Binomial Random Process:

$$Y_n = \sum_{i=1}^{n} X_i$$

   where $\{X_n\}$ is a Bernoulli process, with $X_n = 0$ or 1.

7. Gaussian Random Process:

   This means that every finite collection of the random variables is jointly Gaussian.

**Some Continuous-Time Examples**

1. Sinusoidal random process:

   $X(t) = A\cos(\omega t + \Theta)$ where $A$ and $\Theta$ are independent random variables, and $\Theta$ is uniformly distributed on $[0,2\pi)$.

2. Poisson Counting Process:

Let $T_1, T_2, \ldots$ be an IID random process where each $T_n$ is exponentially distributed with mean 1/a. $T_n$ is the nth interarrival time. Let $S_n = T_1+T_2+\ldots+T_n$ . $S_n$ is the nth arrival time.

$Y(t) = n$ if $S_n \leq t$ and $S_{n+1} > t$ , i.e. if the nth event has occurred by time t but not the n+1'th.

3. Random Process with Finite Number of Sample Functions:

Sample functions:

$X(t,1) = 1, \quad X(t,2) = -2, \quad X(t,3) = \sin \pi t, \quad X(t,4) = \cos \pi t.$

These have probabilities P(1), P(2), P(3), P(4)

4. Gaussian Random Process:

This means that every finite collection of the random variables is jointly Gaussian.

5. White Noise Ra\ndom Process:

This is a WSS random process with power spectral density that is constant with frequency.

**The Second, i.e. Alternative, View of a Random Process**

A random process is a randomly chosen *sample function*. More specifically, a random process is $\{X(t,s): t\in I, s\in S\}$, where I is a set of time-indices (as before) and S is the sample space of some underlying random experiment with probability law P.

For each t, $X(t,s)$ is a random variable. (Recall that model of a random variable as a function of an underlying random experiment. Note that the value of this model is that all random variables are viewed as functions of the same underlying experiment.)

For each $s\in S$, $X(t,s)$ is a function of t called a sample function.

One may think of the random process as being generated in the following way:

At the beginning of time, the underlying experiment is performed (whose probability law is P) resulting in an outcome s. The random process produces the sample function $X(t,s)$ (for this particular s).

This is called the sample-function viewpoint.
In principle, one can derive the probability distribution of the random process (as needed in the first viewpoint) by knowing the function $X(t,s)$ and the probability law P of the underlying experiment.

**Stationarity**

A random process $\{X(t): t\in I\}$ is (strictly) *stationary* if the probability distribution of $X(t_1+\tau),X(t_2+\tau),\ldots,X(t_n+\tau)$ does not depend on $\tau$ for every choice of n and $t_1,\ldots,t_n$.

That is, $F_{X(t_1+\tau),X(t_2+\tau),\ldots,X(t_n+\tau)}(x_1,\ldots,x_n)$ does not depend on $\tau$ and if the random process has joint pdf's or pmf's, the same holds for them.
The basic idea is that for a stationary r.p. the probability distributions of random variables (and vectors) do not change with time shifts. The probability of something happening at time t is the same as the probability of it happening at any other time.

The following are some of the consequences of stationarity:

$f_{X(t)}(x) = f_{X(s)}(x)$ all t,s,x

$f_{X(t)X(t+\tau)}(x_1,x_2) = f_{X(s)X(s+\tau)}(x_1,x_2)$ all $t, s, \tau, x_1,x_2$

$\mu_X(t)$ is the same for all t

$R_X(t,t+\tau)$ does not depend on t.

## Wide-sense Stationarity

A random process $\{X(t): t \in T\}$ is *wide-sense stationary* (WSS) if

$m_X(t)$ and $R_X(t,t+\tau)$ do not depend on $t$.

Stationarity $\Rightarrow$ wide-sense stationary. The converse is false.

Wide-sense stationarity is a weak kind of stationarity that is easier to check and to work with, since it only depends on the mean and autocorrelation functions.

Properties of the Autocorrelation Function of Wide-Sense Stationary (or Stationary) Random Processes

1. Symmetry: $R_X(-\tau) = R_X(\tau)$
2. $R_X(0) \geq |R_X(\tau)|$ for all $\tau$
3. $R_X(\tau) = R_{decay}(\tau) + R_m(\tau) + R_{periodic}(\tau)$ ,
   where $R_{decay}(\tau)$ is a function such that $R_{decay}(\tau) \to 0$ as $|\tau| \to \infty$, $R_m(\tau) = (m_X)^2$ is the term due to the mean of the random process, and $R_{periodic}(\tau)$ is a periodice function that is itself due to a periodic component of the sample functions.

**Ergodicity** (this will not be covered on the exam, but is included here for completeness)

Recall the law of large numbers. Does it hold for random processes other than IID? Sometimes yes, sometimes no. Processes for which it does are called **ergodic.**

Definition: (not the standard mathematical definition)

A discrete-time stationary random process $\{X(n): n = 1,2,...\}$ is *(strict-sense) ergodic* if

$$\frac{1}{n}\sum_{i=1}^{n} g(X(i+1),\ldots,X(i+m)) \to E\, g(X(1),...,X(m)) \quad \text{almost surely as } n \to \infty$$

for any $m$ and any function $g(x_1,...,x_m)$ such that $E\, g(X_1,\ldots,X_m)$ is well-defined.

A continuous-time stationary random process $\{X(t): t \in [0,\infty)\}$ is (strict-sense) ergodic if

$$\frac{1}{T}\int_0^T g(X(t+\tau_i),\ldots X(t+\tau_m)) \to E\, g(X(\tau_1),...,X(\tau_m)) \quad \text{almost surely as } T \to \infty$$

for any $m$, $\tau_1,...,\tau_m$ and any function $g(x_1,...,x_m)$ such that $E\, g(X(\tau_1),\ldots,X(\tau_m))$ is well-defined.

For "two-sided" discrete- and continuous-time random processes, the above averages are replaced by

$$\frac{1}{2n+1}\sum_{i=-n}^{n} \text{ and } \frac{1}{2T}\int_{-T}^{T}, \text{ respectively.}$$

The basic idea is that for ergodic processes, time averages converge to expected values.
As examples, the following are consequences of ergodicity

1. $\frac{1}{n}\sum_{i=1}^{n} X(i) \to EX, \qquad \frac{1}{T}\int_0^T X(t)\, dt \to EX$
2. $\frac{1}{n}\sum_{i=1}^{n} X^2(i) \to EX^2, \qquad \frac{1}{T}\int_0^T X^2(t)\, dt \to EX^2$

3. $\frac{1}{n}\sum_{i=1}^{n} X(i)X(i+1) \rightarrow R_X(1)$, $\quad \frac{1}{n}\sum_{i=1}^{n} X(i)X(i+m) \rightarrow R_X(m)$,

4. $\frac{1}{T}\int_0^T X(t)X(t+\tau)\,dt \rightarrow R_X(\tau)$

5. $\frac{n_A}{N} \rightarrow P(A)$ where $A$ is any event and $n_A$ is the number of times $A$ occurs in $X(1),...,X(n)$

For stationary processes that are not ergodic, time averages such as those above converge, but not to the expected value. Instead, all that we can say is

$$E\left(\frac{1}{n}\sum_{i=1}^{n} g(X(i+1),\ldots,X(i+m))\right) \rightarrow E\, g(X(1),...,X(m)) \quad \text{as } n\rightarrow\infty$$

$$E\left(\frac{1}{T}\int_0^T g(X(t+\tau_i),\ldots X(t+\tau_m)) \rightarrow E\, g(X(\tau_1),...,X(\tau_m))\right) \quad \text{as } T\rightarrow\infty$$

## Random Processes into Linear Filters

Here we focus only on continuous-time random processes and filters. The situation is basically the same for discrete-time random processes and filters, but we have not had the time to discuss it.

If the wide-sense stationary random process $\{X(t)\}$ with mean $m_X$ and autocorrelation function $R_X(\tau)$ is the input to a linear filter with impulse response $h(t)$ and frequency response $H(f)$, then the output of the filter is a wide-sense stationary random process $\{Y(t)\}$ with

$$m_Y = m_X \int_{-\infty}^{\infty} h(t)\,dt = m_X H(0)$$

$$R_X(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$$ recall the defn of convolution: $x(t)*y(t) = \int_{-\infty}^{\infty} x(u)\, y(t-u)\, du$

### Important Fact:

If the input to a linear filter is a Gaussian random process, then the output is a Gaussian random process.

## Power Spectral Density

The power spectral density of a WSS random process $\{X(t)\}$ is

$$S_X(f) = \mathcal{F}\{R_X(\tau)\} = \int_{-\infty}^{\infty} R_X(\tau)\, e^{-j2\pi f\tau} d\tau = \text{Fourier transform of } R_X(\tau)$$

Properties of the power spectral density:

1. $S_X(f) \geq 0$.

2. $S_X(-f) = S_X(f)$.

3. $\int_{-\infty}^{\infty} S_X(f)\, dx = $ power in $\{X(t)\}$

4. $2 \int_{f_1}^{f_2} S_X(f)\, dx$ = power in $\{X(t)\}$ in the frequency band $[f_1,f_2]$ .

This is why it is called a power spectral density; i.e. one integrates it over a frequency band to obtain the power in that band.

5. If $\{X(t)\}$ is the input to a linear filter with frequency response $H(f)$, then the output random process $\{Y(t)\}$ has power spectral density

$$S_Y(f) = S_X(f)\, H(f)\, H(-f) = S_X(f)\, |H(f)|^2$$

6. Alternate formula for the power spectral density (more difficult to work with)

$$S_X(f) = \lim_{T\to\infty} \frac{1}{2T} E \left|\mathcal{F}\{X_T(t)\}\right|^2, \quad \text{where} \quad X_T(t) = \begin{cases} X(t), & -T\le t\le T \\ 0\ , & \text{else} \end{cases}$$

7. Power spectral density of some simple processes:

a. $X(t) = A$, where $A$ is a random variable

$$S_X(f) = E A^2 \delta(f) \ .$$

b. $X(t) = A \cos(2\pi f t+\Theta)$ where $A$ and $\Theta$ are independent random variables and $\Theta$ is uniformly distributed on $[0,2\pi)$. Then

$$S_X(f) = \frac{1}{2} E A^2 \left[\delta(f-f_o) + \delta(f+f_o)\right]$$

c. If the input to a linear filter with frequency response $H(f)$ is white noise $X(t)$ with $S_Y(f) = c$, then the output has power spectral density

$$S_Y(f) = c\, |H(f)|^2$$