ECE 5654 Lecture 2
Review of Probability theory & Random Processes

Robert McGwier & Ravi Tandon

Virginia Tech

January 27, 2014

Learning Objectives

At the end of this lecture, the student should be able to:

- Describe random variables and random processes and explain why they are useful in the study of digital communications.
- Define the major ways of describing a random variable.
- Describe the relationship between random processes and random variables.

Why are Random Variables and Processes Important ?

- Random Variables and Processes let us talk about and analyze quantities and signals which are unknown in advance.
- Examples:
  - The binary data sent through a communication system is modeled as random binary values.
  - The noise, interference, and fading introduced by the channel can all be modeled as random processes.
  - Even the measure of performance (Probability of Bit Error) is expressed in terms of a probability.

Random Events as Sets

- The result of a random experiment/trial is called an outcome.
- The set of all possible outcomes is called the **sample space**: S.
- Eg. a coin is tossed. 2 possible outcomes: H or T, so $S = \{H, T\}$.
- We may be interested in the possible occurrences of following events:
  - the outcome is a head: $A = \{H\}$.
  - the outcome is either a head or a tail: $A = \{H\} \cup \{T\}$.
  - the outcome is not a head: $A = \{H\}^c$.
  - the outcome is both head and a tail: $A = \{H\} \cap \{T\}$. (unlikely!)
- Thus, events can be specified as subsets A of the sample space S.
- The complementary event: $A^c = S - A$.
- The certain event: S
- The null event: ϕ
- Transmitting a data bit is also an experiment.

Probability of an Event

- The probability $P(A)$ is a positive real number which measures the likelihood of the event A.

Axioms of Probabilty

- Probability of sample space is 1: $P(S) = 1$.
- No event has probability less than zero: $P(A) \geq 0$.
- Let A and B be two mutually exclusive events ($A \cup B = \phi$), then

$$P(A \cup B) = P(A) + P(B)$$

- All other laws of probability follow from these axioms.

Relationship between Random Events

- Joint probability: $P(A, B) = P(A \cap B)$
  -probability that both A and B occur.
- Conditional probability: $P(A|B) = \frac{P(A,B)}{P(B)}$
  -probability that A will occur given that B has occurred.
- Statistical Independence:
  - Events A and B are statistically independent if:

$$P(A, B) = P(A) \cdot P(B)$$

  - If A and B are independent then:

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

Random Variables

- A random variable $X(s)$ is a real valued function of the underlying sample space $s \in S$.
- A random variable may be:
  - Discrete-valued: range is finite (e.g. $\{0{,}1\}$) or countably infinite (e.g. $\{1, 2, 3, ...\}$
  - Continuous-valued: range is uncountably infinite (e.g., $\mathcal{R}$)
- A random variable is described by:
  - A name: X
  - It's range: $X \in \mathcal{R}$
  - A description of its distribution

Cumulative Distribution Function (CDF)

- Definition: $F(x) = P(X \leq x)$.
- Properties of CDF:
  - $F(x)$ is monotonically nondecreasing.
  - $F(-\infty) = 0$
  - $F(\infty) = 1$
  - $P(a < X \leq b) = F(b) - F(a)$
- While CDF completely defines the distribution of a r.v., we usually work with the probability density function (PDF) or the probability mass function (PMF)

Probability Density Function (PDF)

- Definition: $p(x) = \frac{dF(x)}{dx}$
- Interpretation: PDF measures how fast the CDF is increasing (its slope) or how likely a random variable is to take a particular value.
- Properties:
  - $p(x) \geq 0$
  - $\int_{-\infty}^{\infty} p(x)\, dx = 1$
  - $P(a < X \leq b) = \int_a^b p(x)\, dx$ ("area" between a and b)

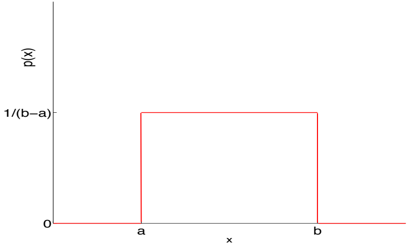
Expectation Operator

- Expectation operator:

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

- Here, $g(X)$ could be any function of the random variable X
- Most important examples are:
  - Mean: $E(X) = m_x = \int_{-\infty}^{\infty} xp(x)\,dx$
  - Variance: $E\left([X - m_x]^2\right) = V_x = \int_{-\infty}^{\infty} (x - m_x)^2 p(x)\,dx$
  - Verify: $V_x = E(X^2) - m_x^2$
  - Standard deviation: $\sigma_x = \sqrt{V_x}$

Eg. 1: Uniform Random variable

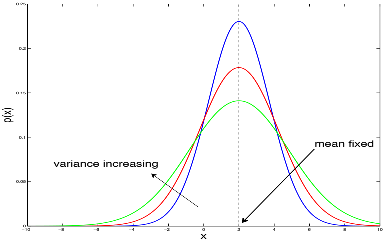


- $p(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & elsewhere. \end{cases}$

- $m_x = \int_a^b x(\frac{1}{b-a})\,dx = \frac{a+b}{2}$

- $V_x = \int_a^b (x - m_x)^2(\frac{1}{b-a})\,dx = \frac{(b-a)^2}{12}$

Eg. 2: Gaussian Random variable



- $p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ (denoted as $\mathcal{N}(\mu,\sigma^2)$)
- A Gaussian r.v. is completely characterized by its mean and variance.
- Standard normal distribution, $\mathcal{N}(0,1)$: $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- $N(\mu,\sigma^2)$ in terms of $\phi(x)$: $\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$
  (operations: translating by μ, stretching by σ, and scaling by $1/\sigma$)

Complementary Error Function & Gaussian CDF

- Error function:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$

- Complementary error function: $erfc(x) = 1 - erf(x)$

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\, dt$$

- CDF of Standard Normal $\mathcal{N}(0,1)$: $\Phi(x) = P(X \le x)$

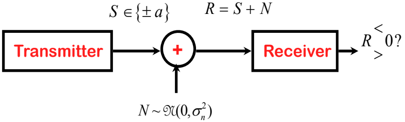
$$\Phi(x) = \frac{1}{2}\left[1 + erf\left(\frac{x}{\sqrt{2}}\right)\right]$$

Q-function

- Q-function: $P(X > x)$ ["tail" of a standard Normal distribution]

$$Q(x) = 1 - \Phi(x) = \frac{1}{2} erfc\left(\frac{x}{\sqrt{2}}\right)$$

- Q-function is a standard form for expressing error probabilities without a closed form expression.
- For a non-standard Gaussian r.v. $\mathcal{N}(\mu, \sigma^2)$
- CDF: $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- $P(X > x) = 1 - F(x) = Q\left(\frac{x-\mu}{\sigma}\right)$



- Conditional error event: $R > 0$ given $S = -a$.
- Given $S = -a$, $R = -a + N$, hence $R_{|-a} \sim \mathcal{N}(-a, \sigma_n^2)$.
- Conditional error prob. $= P(R > 0| - a) = P(\mathcal{N}(-a, \sigma_n^2) > 0)$
- $\Rightarrow Q\left(\frac{0-(-a)}{\sigma_n}\right) = Q(a/\sigma)$

$$\begin{aligned}
\text{Prob. of error} &= P(a)P(R > 0| - a \text{ sent}) + P(-a)P(R \leq 0|a \text{ sent}) \\
&= \underbrace{(P(a) + P(-a))}_{=1} Q(a/\sigma) \\
&= Q(a/\sigma)
\end{aligned}$$

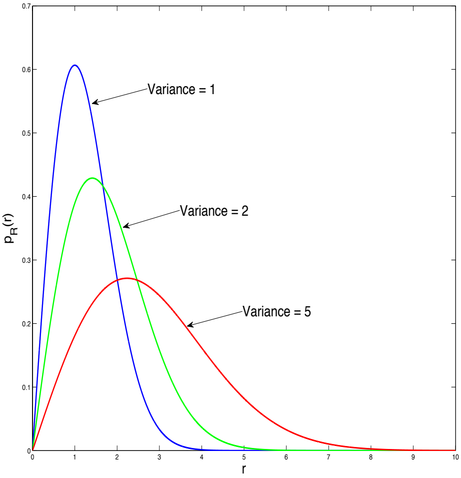
Eg. 3: Rayleigh Random variable

- Let $R = \sqrt{X_1^2 + X_2^2}$,
  where X_1 and X_2 are Gaussian and distributed as $\mathcal{N}(0, \sigma^2)$.
- Then R is a Rayleigh random variable with the pdf:

$$p_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

- Rayleigh pdf's are frequently used to model fading when no line of sight signal is present.

Rayleigh PDF



Probability Mass Function (PMF)

- A discrete random variable can be described by a pdf if we allow impulse functions.
- We usually use probability mass function (pmf): $p(x) = P(X = x)$
- Properties are analogous to pdf:
  - $\sum_x p(x) = 1$
  - $p(x) \geq 0$
  - $P(a \leq X \leq b) = \sum_{x=a}^{b} p(x)$

Eg. 4: Bernoulli distribution

- $p(x) = \begin{cases} 1-p, & x=0 \\ p, & x=1 \end{cases}$
- This is frequently used to model binary data.
- Mean $= 0 \times p(0) + 1 \times p(1) = p$
- Variance $= p(1-p)$
- If X_1 and X_2 are independent binary random variables, then

$$p_{X_1,X_2}(a,b) = p_{X_1}(a) \cdot p_{X_2}(b)$$

for all $a, b \in \{0,1\}$.

Eg. 5: Binomial Distribution

- Let $Y = \sum_{i=1}^{n} X_i$, where X_i are independent and identically distributed as $p(x) = \begin{cases} 1-p, & x=0 \\ p, & x=1 \end{cases}$
- Then, $p_Y(y) = \binom{n}{y} y^p (n-y)^{1-p}$
  (probability that y elements are 1s and $(n-y)$ are 0s)
- Mean $= np$
- Variance $= np(1-p)$

Eg. 5 (continued)

- Suppose that we transmit a 31 bit sequence with error correction capable of correcting upto 3 errors.
- If the probability of a bit error is $p = 0.001$, what is the probability that codeword is received in error ?

$$
\begin{aligned}
&= 1 - \sum_{e=0}^{3} \binom{31}{e} (1-p)^{31-e} p^e \\
&\approx 3.0793e^{-08}
\end{aligned}
$$

- If no error correction is used, the error probability is:

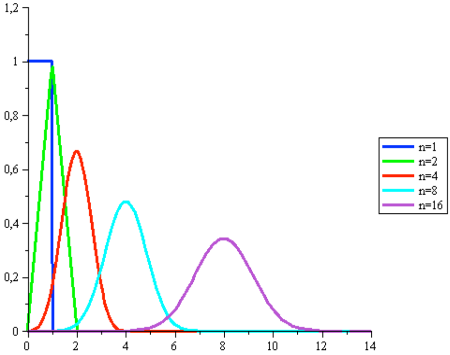
$$= 1 - (1-p)^{31} \approx 0.0305$$

Central Limit Theorem

- Let $X_1, X_2, \ldots, X_N$ be a set of independent random variables with identical pdfs.
- Let $Y = \sum_{i=1}^{N} X_i$
- Then, as $N \to \infty$, the distribution of Y converges to a Gaussian distribution.
- In practice, $N = 10$ is usually enough to see this effect.
- Thermal noise results from the random movement of many electrons- this is well modeled by a Gaussian distribution.

Example of Central Limit Theorem

- Sum of uniformly distributed random variables (each uniform over $[0, 1]$)



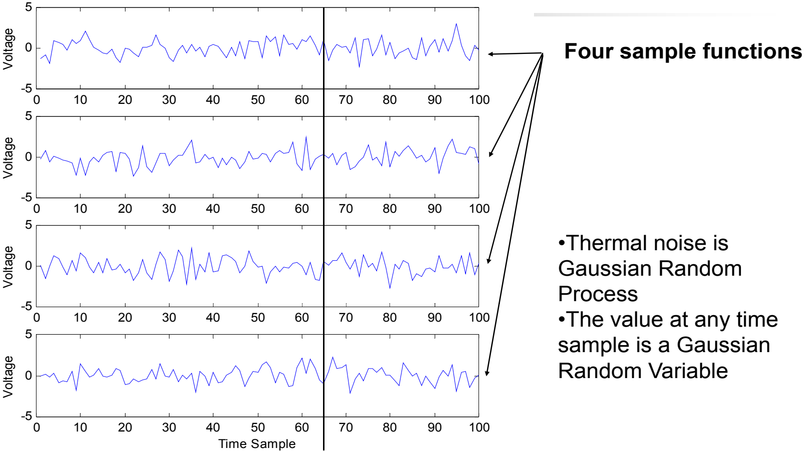
Random Processes

- A random variable has a single value. However, actual signals change with time.
- Random <u>variables</u> model unknown values (e.g., samples).
- Random <u>processes</u> model unknown signals.
- A random process is just a collection of random variables.
- **<u>Definition</u>**: A random process is an indexed set of functions of some parameter (usually time) that has certain statistical properties.
- If $X(t)$ is a random process, then $X(1), X(5), X(18.5)$ are all random variables for any specific time t

Random Processes

- A specific instance of a random process is termed a sample function
- The value of a random process at time t, i.e., $X(t)$ is a random variable.
- Thus, a random process is an indexed set of random variables that have some a cross-correlation and distribution that are determined by the underlying function.
- We can talk of ensemble averages and time averages:
  - Ensemble averages are the averages of all possible sample functions sampled at a specific time
  - Time averages are the averages taken of a specific sample function over all time

Example: Gaussian Random Process



•Thermal noise is Gaussian Random Process
•The value at any time sample is a Gaussian Random Variable

- Value at $t = 65$ is a random variable.

Eg. Random Process

$$X(t) = A\sin(\omega t + \theta)$$

- Let A and ω be known.
- θ is a random variable uniformly distributed on $[0, 2\pi)$
- $x_1(t) = A\sin(\omega t + \pi/5)$ is a sample function.
- The value at any time t_1, i.e., $x = X(t_1)$ is a random variable with distribution:

$$f(x) = \begin{cases} \frac{1}{\pi\sqrt{A^2 - x^2}}, & |x| \leq A \\ 0, & elsewhere. \end{cases}$$

Terminology Describing Random Processes

- A stationary random process has statistical properties which do not change at all with time (i.e., all joint pdfs do not change)
- A wide sense stationary (WSS) process has a mean and autocorrelation function which do not change with time (this is usually sufficient)
- A random process is ergodic if the time average always converges to the statistical average.
- Unless specified, we will assume that all random processes are WSS and ergodic.

Stationary Random Processes

- A stationary random process has statistical properties which do not change at all with time (i.e., all joint pdfs do not change)
  - First order: $f_X(x_1)$ where $x_1 = x(t_1)$ does not depend on the value of t_1
  - Second order: $f_{X_1,X_2}(x_1, x_2)$ where $x_1 = x(t_1), x_2 = x(t_2)$ don't depend on the values of t_1 and t_2 but only the difference $\tau = t_1 - t_2$
- A wide sense stationary (WSS) process has a mean and autocorrelation function which do not change with time (this is usually sufficient)
  1. $E[x(t)] = \overline{X}$
  2. $E[x(t_1)x(t_2)] = E[x(t)x(t+\tau)] = R_X(\tau)$

Ergodic Random Processes

- A random process is ergodic if the time averages (e.g., mean and autocorrelation) always converge to the statistical averages.
  - i.e., we can use time averages of a sample function to estimate the ensemble averages
  - In real life we can not obtain a sufficient number of sample functions, so we rely on time averages of a single sample function.
- Unless specified, we will assume that all random processes are WSS and ergodic.
- Note that all ergodic processes are stationary, but not all stationary processes are ergodic.

Example of Ergodic Process

$$X(t) = A\sin(\omega t + \theta), \quad \theta \text{ is a uniform r.v. on } [0, 2\pi)$$

- Ensemble averages (i.e., average over all values of θ for a specific t):

$$\overline{x} = \int_0^{2\pi} \frac{1}{2\pi} A\sin(\omega t + \theta)\, d\theta = 0$$

$$\overline{x^2} = \int_0^{2\pi} \frac{1}{2\pi} A^2 \sin^2(\omega t + \theta)\, d\theta = \frac{A^2}{2}$$

- Time averages (i.e., average over a time interval T for a specific θ):

$$< x(t) > = \frac{1}{T}\int_0^{T} A\sin(\omega t + \theta)\, dt = 0$$

$$< x^2(t) > = \frac{1}{T}\int_0^{T} A^2 \sin^2(\omega t + \theta)\, dt = \frac{A^2}{2}$$

- Thus, we say that this random process is ergodic
- Would this be true if θ was a uniform r.v. over $[0, \pi]$?

Description of Random Processes

- Knowing the pdf of individual samples of the random process is not sufficient. We also need to know how how individual samples are related to each other.
- Two tools are available to describe this relationship
- Autocorrelation function
- Power spectral density function

Autocorrelation Function (AF)

- Autocorrelation measures how a random process changes with time.
- Intuitively, $X(1)$ and $X(1.1)$ will be more strongly related than $X(1)$ and $X(100000)$ (although it is possible to construct counter-examples). The autocorrelation function quantifies this.
- Definition (for WSS random processes):

$$R_X(\tau) = E[X(t)X(t+\tau)]$$

- Note that Power $= R_X(0)$.

Power Spectral Density (PSD)

- $S_X(f)$ is the power spectral density which tells us how much power is at each frequency.
- Wiener-Khinchine Theorem: $S_X(f) = \mathcal{F}\{R_X(\tau)\}$
- Power spectral density (PSD) and autocorrelation function (AF) are a Fourier Transform pair.
- Properties of Power Spectral Density:

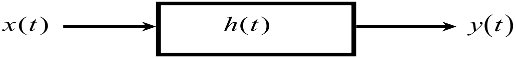
$$S_X(f) \geq 0$$
$$S_X(f) = S_X(-f)$$

- Power $= \int_{-\infty}^{\infty} S_X(f)\, df$

Gaussian Random Processes

- Gaussian Random Processes have several special properties
- If a Gaussian random process is wide-sense stationary, then it is also stationary.
- Any sample point from a Gaussian random process is a Gaussian random variable
- If the input to a linear system is a Gaussian random process, then the output is also a Gaussian random process

Linear Systems



- Input: $x(t)$
- Impulse Response: $h(t)$
- Ouput: $y(t)$

Computing the Output of Linear Systems

- Deterministic Signals:
  - Time domain: $y(t) = h(t) * x(t)$
  - Frequency domain: $Y(f) = \mathcal{F}(y(t)) = X(f)H(f)$
- For a random process, we can still relate the statistical properties of the input and output signal
  - Time domain: $R_y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$
  - Frequency domain: $S_Y(f) = S_X(f)|H(f)|^2$

Conclusions

- In this lecture we have examined the primary tools for analyzing random signals: random variables and random processes
- In general, in this class we will use random variables when examining the probability of bit error (i.e., performance), while we will use random processes when we examine spectra and bandwidth