# EE 5654 - Digital Communications Spring 2005

Instructor: R. Michael Buehrer

Lecture #2 - Review of Probability and Random Processes

# Why are Random Processes important?

- Random Variables and Processes let us talk about quantities and signals which are unknown in advance
- Examples:
  - The data sent through a communication system is modeled as random
  - 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

- When we conduct a random experiment, we can use set notation to describe possible outcomes.
- Example: Roll a six-sided die.

  Possible Outcomes: $S = \{1,2,3,4,5,6\}$
- An event is any subset of possible outcomes: $A = \{1,2\}$
- The complementary event: $\overline{A} = S - A = \{3,4,5,6\}$
- The set of all outcomes is the certain event: $S$
- The null event: $\phi$
- Transmitting a data bit is also an experiment

# Probability

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

**Axioms of Probability:**

- No event has probability less than zero: $P(A) \geq 0$

$$P(A) \leq 1 \text{ and } P(A) = 1 \Leftrightarrow A = S$$

- Let *A* and *B* be two events such that: $A \cap B = \phi$

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

- All other laws of probability follow from these axioms

# Relationships Between Random Events

- Joint Probability: $P(A,B) = P(A \cap B)$
  - Probability that both A and B occur
- Conditional Probability: $P(A|B) = \dfrac{P(A,B)}{P(B)}$
  - Probability that A will occur given that B <u>has</u> 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 event 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. $\Re$ )
- A random variable may be described by:
  - A name: *X*
  - It’s range: $X \in \Re$
  - A description of its distribution

# Cumulative Distribution Function (CDF)

- Also called Probability Distribution Function
- Definition: $F_X(x) = F(x) = P(X \le x)$
- Properties:

  $F(x)$ is monotonically nondecreasing

  $F(-\infty) = 0$

  $F(\infty) = 1$

  $P(a < X \le b) = F(b) - F(a)$
- While the CDF completely defines the distribution of a random variable, we will usually work with the probability density function (pdf) or the probability mass function (pmf)

# Probability Density Function (pdf)

- Defn: $p_X(x) = \frac{dF_X(x)}{dx}$ or $p(x) = \frac{dF(x)}{dx}$
- Interpretations:
  - pdf measures how fast CDF is increasing or how likely a random variable is to lie at a particular value
- Properties: $p(x) \geq 0$

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

$$P(a < X \leq b) = \int_{a}^{b} p(x)dx$$

# Expected Values

- Expected values are a shorthand way of describing a random variable
- The most important examples are:
  - Mean: $$E(X) = m_X = \int_{-\infty}^{\infty} xp(x)dx$$
  - Variance: $$E\left([X - m_X]^2\right) = \int_{-\infty}^{\infty} (x - m_X)^2 p(x)dx$$
- The expectation operator works with any function:

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

# Chebyshev Inequality

- Let $X$ be a random variable with mean: $m_X$ and variance: $\sigma_X^2$
- Then for any $\delta$ , $P(|X - m_X| \geq \delta) \leq \dfrac{\sigma_X^2}{\delta^2}$
- The size of the variance determines how probable it is for a random variable to lie close to it's mean value
- This bound can be used to determine confidence intervals of a simulation

# Chernoff Bound

- Let $Y$ be a random variable
- Then, for any value of $v>0$ and $\delta>0$:

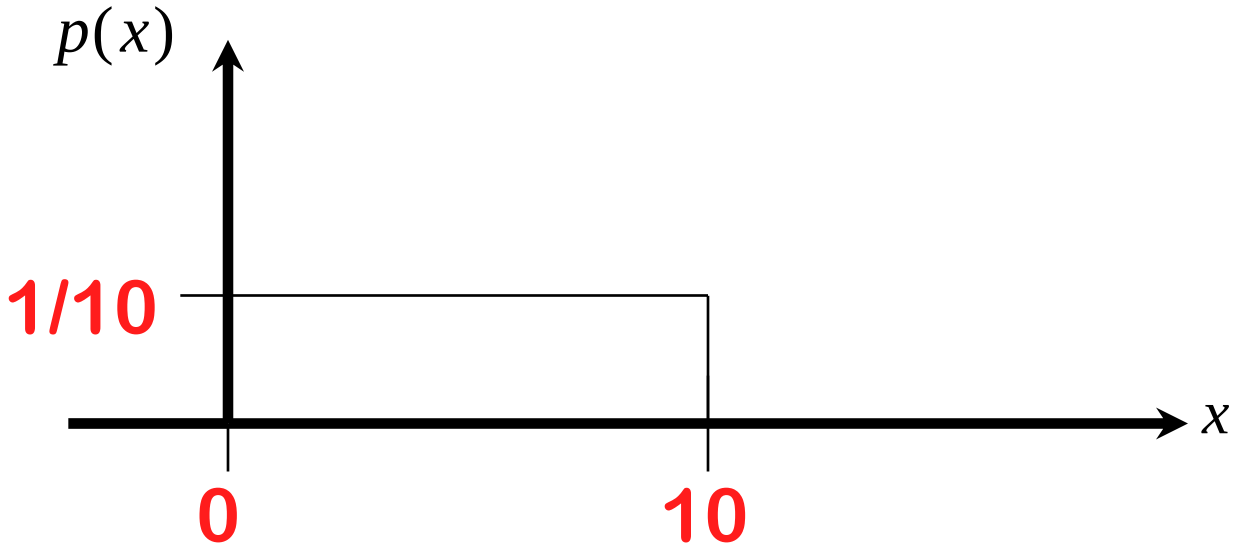
$$\Pr[Y \geq \delta] \leq E\left(e^{v(Y-\delta)}\right)$$

- Very useful for upper bounding low probability events on the tails of distributions
- Example: $p_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \quad \Rightarrow \Pr[Y \geq \delta] = Q(\delta) \leq e^{-\delta^2/2}$

  (choosing the optimal value of $v$)
- We will also see that the Chernoff bound gives us a useful upper bound on the error probability of convolutional codes

# Example #1: Uniform pdf

$$p(x) = \begin{cases} 1/10, & 0 \le x \le 10 \\ 0, & \text{else} \end{cases}$$

- 

# Example #1 (continued)

- Mean: $$m_x = \int_{-\infty}^{\infty} x \cdot p(x)dx = \int_{0}^{10} x \cdot \frac{1}{10}dx = \left[\frac{x^2}{20}\right]_0^{10} = 5$$
- Variance:

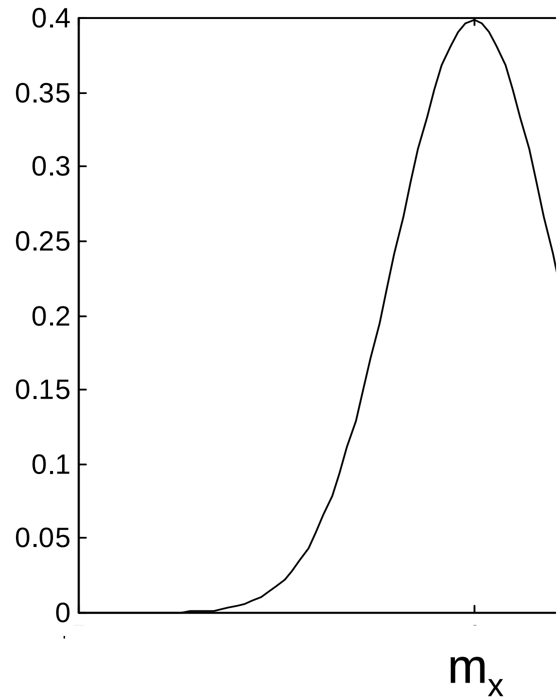
$$\sigma_x^2 = \int_{-\infty}^{\infty} (x-5)^2 \cdot p(x)dx = \int_{0}^{10} (x-5)^2 \cdot \frac{1}{10}dx = \frac{25}{3}$$

- Probability Calculation:

$$P(6 \le x \le 9) = \int_{6}^{9} p(x)dx = \int_{6}^{9} \frac{1}{10}dx = 0.3$$

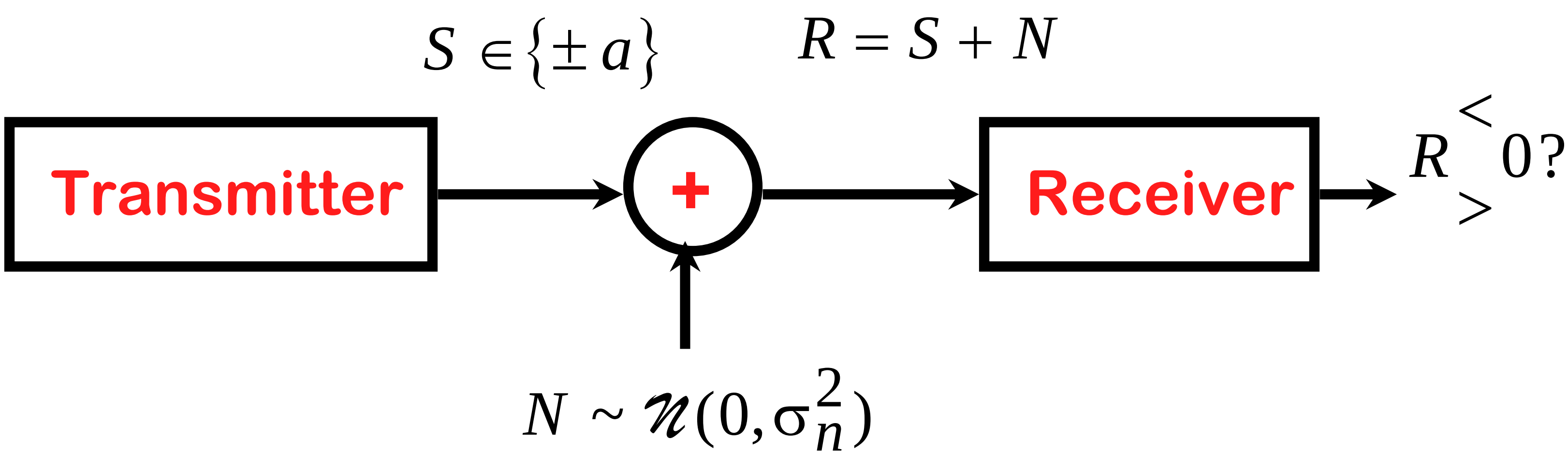
# Example #2: Gaussian pdf

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}$$




- A Gaussian random variable is completely determined by its mean and variance
- Also called a *normal* random variable

# A Communication System with Gaussian Noise




- The probability that the receiver will make an error is the probability that $R > 0$ given that $S = -a$:

$$P(R > 0 | S = -a) = \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(x+a)^2}{2\sigma_n^2}} = Q\left(\frac{a}{\sigma_n}\right)$$

# The *Q*-function

- The Q-function is a standard form for expressing error probabilities without a closed form:

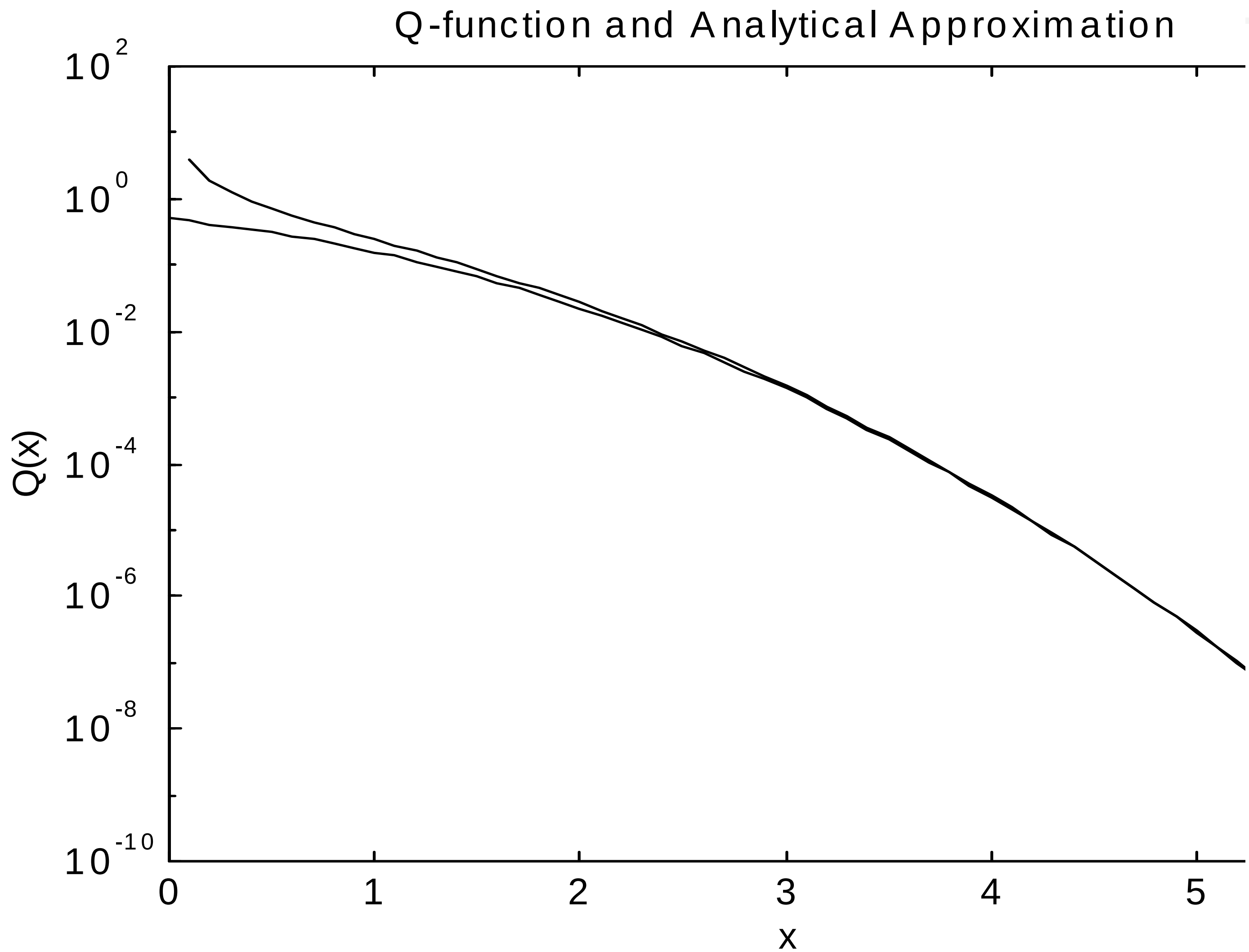
$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = \frac{1}{2}\operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)$$

- Numerical Calculation of *Q*-function:

$$Q(x) = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}\left[1 - \frac{1}{x^2} + \frac{1\cdot 3}{x^4} + \cdots + \frac{(-1)^n \cdot 1\cdot 3 \cdots (2n-1)}{x^{2n}}\right]$$

$$\approx \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \text{ for } x \geq 3$$

# The Q-function and its Approximation



# Example #3 - Rayleigh pdf

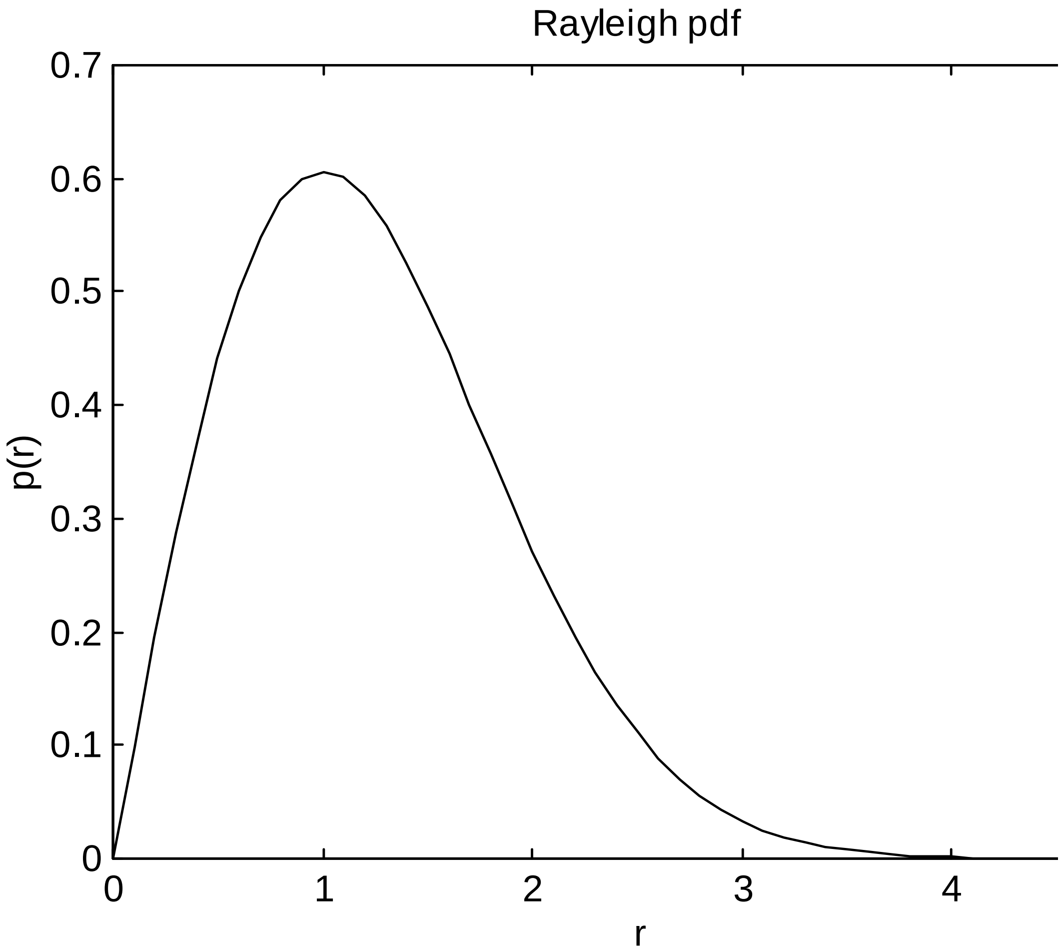
- Let: $R = \sqrt{X_1^2 + X_2^2}$

  where $X_1$ and $X_2$ are Gaussian with mean 0 and variance $\sigma^2$
- Then *R* is a Rayleigh random variable with pdf:

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

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

# Rayleigh pdf



# Probability Mass Functions (pmf)

- A discrete random variable can be described by a pdf if we allow impulse functions
- We usually use probability mass functions (pmf):

$$p(x) = P(X = x)$$

- Properties are analogous to pdf:

$$p(x) \geq 0$$

$$\sum_{X} p(x) = 1$$

$$P(a \leq X \leq b) = \sum_{x=a}^{b} p(x)$$

# Example #1: Binary Distribution

- $$p(x) = \begin{cases} 1/2, & x = 0 \\ 1/2, & x = 1 \end{cases}$$
- This is frequently used to model binary data
- Mean: $m_X = \sum_x x \cdot p(x) = 0 \cdot 1/2 + 1 \cdot 1/2 = 1/2$
- Variance:

$$\sigma_X^2 = \sum_x (x - m_X)^2 \cdot p(x) = (1/2)^2 \cdot 1/2 + (1/2)^2 \cdot 1/2 = 1/4$$

- If $X_1$ and $X_2$ are independent binary random variables, then

$$p_{X_1 X_2}(0,0) = p_{X_1}(0) \cdot p_{X_2}(0) = 1/2 \cdot 1/2 = 1/4$$

# Example #2: Binomial Distribution

- Let $Y = \sum_{i=1}^{n} X_i$ where $\{X_i, i = 1, \ldots, n\}$

are independent binary RVs with:

$$p_X(x) = \begin{cases} 1 - p, & x = 0 \\ p, & x = 1 \end{cases}$$

- Then $p_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}$ $\binom{n}{y} = \dfrac{n!}{y! \cdot (n-y)!}$

- Mean: $m_X = n \cdot p$
- Variance: $\sigma_X^2 = n \cdot p \cdot (1-p)$

# Example #2 (continued)

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

$$P(\text{codeword error}) = 1 - P(\text{correct codeword})$$

$$= 1 - \sum_{i=0}^{3} \binom{31}{i} (0.999)^{31-i} (0.001)^{i} \approx 3 \times 10^{-8}$$

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

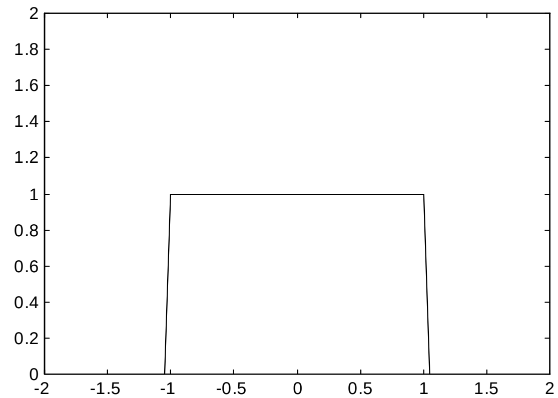
$$1 - (1 - 0.001)^{31} = 0.0305$$

# Central Limit Theorem

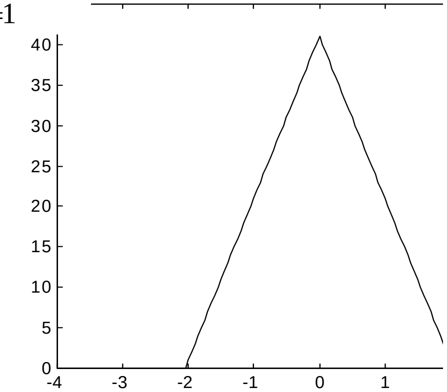
- Let $X_1, X_2, \ldots, X_N$ be a set of independent random variables with identical pdfs
- Let: $Y = \sum_{i=1}^{N} X_i$
- Then as $N \to \infty$, the distribution of will tend towards a Gaussian distribution
- In practice, $N = 10$ is usually enough to see this effect
- Thermal noise results from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

# Example of Central Limit Theorem:

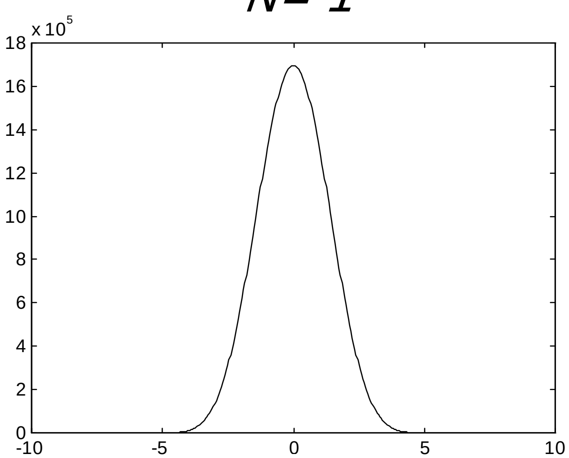
Probability Density Functions for $Y = \sum_{i=1}^{N} X_i$ where *$X_i$ is a uniform RV*



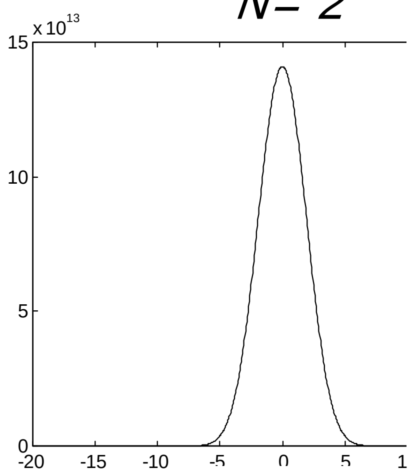

*N= 1*




*N= 2*




*N= 5*




*N= 10*

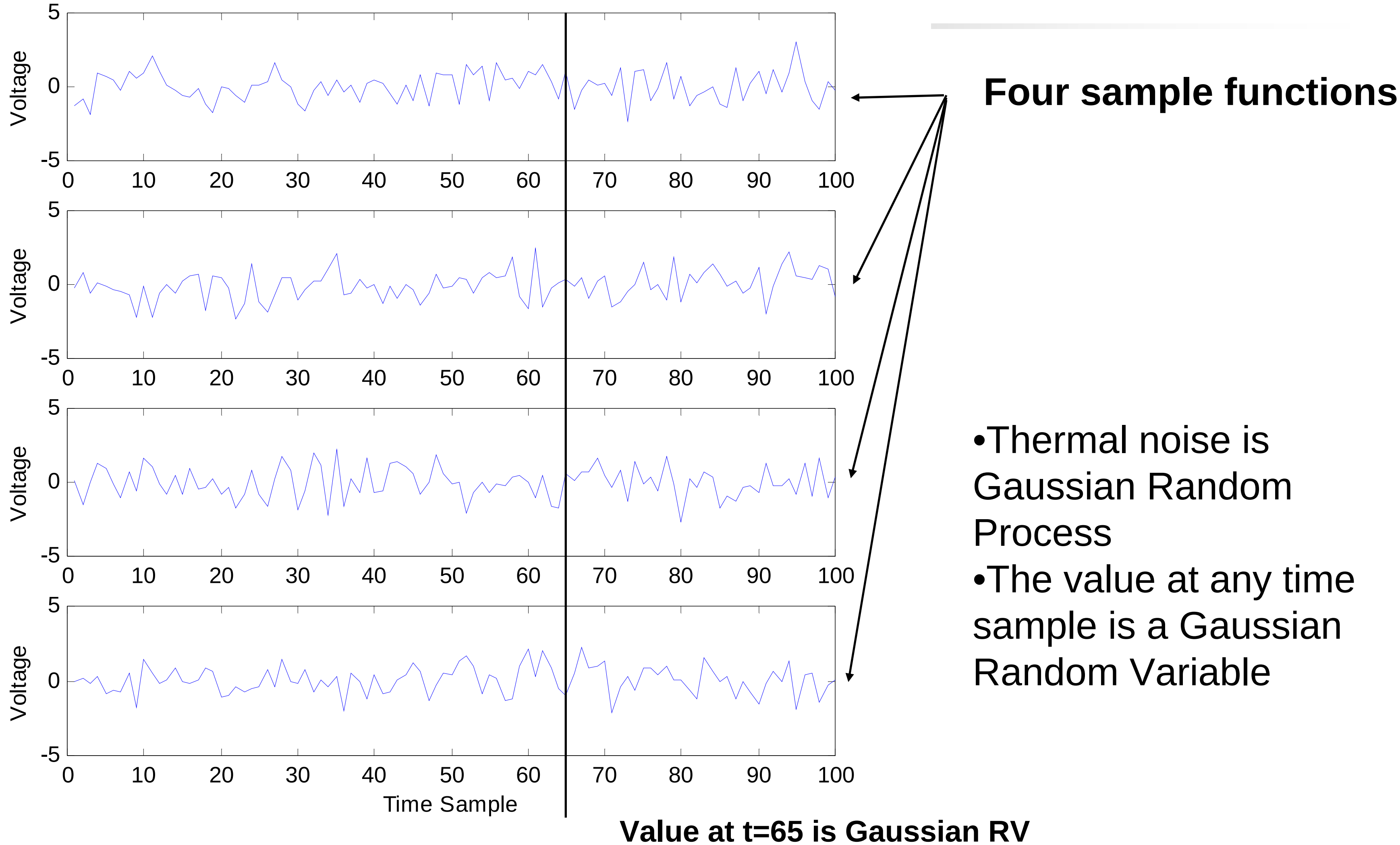
# Random Processes

- A random variable has a single value. However, actual signals change with time.
- Random <u>variables</u> model unknown events.
- Random <u>processes</u> model unknown signals.
- A random process is just a collection of random variables.
- ***Definition:*** 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 $X(1)$, $X(1.5)$, and $X(37.5)$ 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 $X(t_1)$ 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

# Examples of Random Processes

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

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

$$f(x) = \begin{cases} \dfrac{1}{\pi\sqrt{A^2 - x^2}} & |x| \le A \\ 0 & else \end{cases}$$

# 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 –* $f_{X1X2}(x_1, x_2)$ *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)

$$1.\quad E\left[x(t_1)\right] = \bar{X}$$

$$2.\quad E\left[x(t_1)x(t_2)\right] = E\left[x(t)x(t+\tau)\right] = R_X(\tau)$$

# 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.

# Ergodic Process – Example

$$x(t) = A\sin(\omega_o t + \theta_o) \qquad \theta_o \textit{ is uniform R.V. on } [0,2\pi)$$

- First let us examine the ensemble averages (average over all values of θ for a specific *t* ):

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

$$= 0$$

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

$$= \frac{A^2}{2}$$

# Ergodic Process – Ex (cont.)

- Now let us examine the time averages for a specific value of θ:

$$\langle x(t) \rangle = \frac{1}{T_o}\int_0^{T_o} A\sin(\omega_o t + \theta)dt$$

$$= 0$$

$$\langle x^2(t) \rangle = \frac{1}{T_o}\int_0^{T_o} A^2 \sin^2(\omega_o t + \theta)dt$$

$$= \frac{A^2}{2}$$

- Thus, we say that this random process is *ergodic. Note that if* $\theta$ *were distributed over [0,*$\pi$*] the process would be neither stationary nor ergodic.*

# 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

- 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.
- Defn (for WSS random processes):

$$\phi_X(\tau) = E\left[X(t)X(t+\tau)\right]$$

- Note that Power $= \phi_X(0)$

# Power Spectral Density

- $\Phi(f)$ is the power spectral density which tells us how much power is at each frequency
- Wiener-Khinchine Theorem: $\Phi(f) = F\{\phi(\tau)\}$

  Power spectral density and autocorrelation are a Fourier Transform pair!
- Properties of Power Spectral Density:

$$\Phi(f) \geq 0$$

$$\Phi(f) = \Phi(-f)$$

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

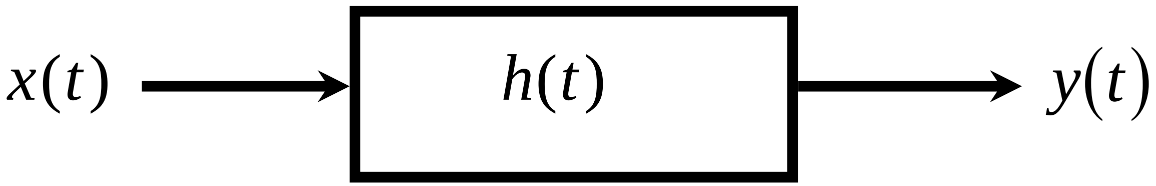
# Gaussian Random Processes

**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: $h(t)$
- Output: $y(t)$



# Computing the Output of Linear Systems

- Deterministic Signals:
  - Time Domain: $y(t) = h(t) * x(t)$
  - Frequency Domain: $Y(f) = F\{y(t)\} = X(f)H(f)$
- For a random process, we can still relate the statistical properties of the input and output signal
  - Time Domain:

$$\phi_Y(\tau) = \phi_X(\tau) * h(\tau) * h(-\tau)$$

  - Frequency Domain:

$$\Phi_Y(f) = \Phi_X(f) \cdot |H(f)|^2$$