Probability & Stochastic Processes for Communications:

A Gentle Introduction

Shivkumar Kalyanaraman

Outline

- Please see my experimental networking class for a longer video/audio primer on probability (not stochastic processes):
  - http://www.ecse.rpi.edu/Homepages/shivkuma/teaching/fall2006/index.html
- Focus on Gaussian, Rayleigh/Ricean/Nakagami, Exponential, Chi-Squared distributions:
  - Q-function, erfc(),
  - Complex Gaussian r.v.s,
  - Random vectors: covariance matrix, gaussian vectors
  - …which we will encounter in wireless communications
- Some key bounds are also covered: Union Bound, Jensen's inequality etc
- Elementary ideas in stochastic processes:
  - I.I.D, Auto-correlation function, Power Spectral Density (PSD)
  - Stationarity, Weak-Sense-Stationarity (w.s.s), Ergodicity
  - Gaussian processes & AWGN ("white")
  - Random processes operated on by linear systems

Elementary Probability Concepts
(self-study)

Probability

- Think of probability as modeling an experiment
  - Eg: tossing a coin!
- The set of all possible *outcomes* is the *sample space*: S

- Classic "Experiment":
- Tossing a die: S = {1,2,3,4,5,6}
  - Any subset A of S is an *event*:
  - A = {*the outcome is even*} = {2,4,6}

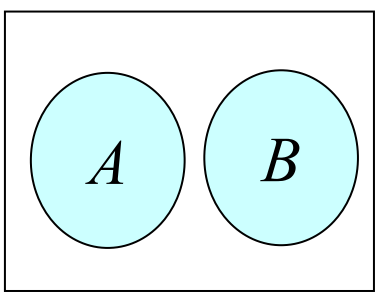
Probability of Events: Axioms

- P is the Probability Mass function if it maps each event A, into a real number ***P(A)***, and:

  i.) $P(A) \geq 0$ for every event $A \subseteq S$

  ii.) *P(S)* = 1

  iii.)If A and B are mutually exclusive events then,



$A \cap B = \phi$

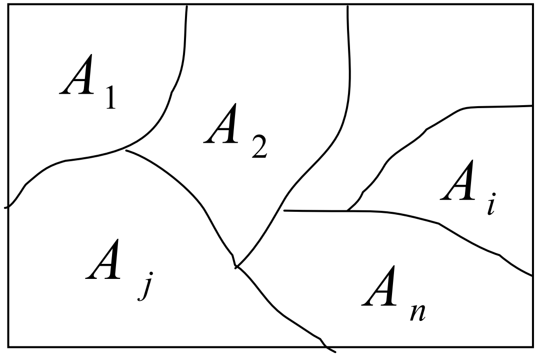
$$P(A \cup B) = P(A) + P(B)$$

Probability of Events

…In fact for any sequence of pair-wise-mutually-exclusive events, we have

$$A_1, A_2, A_3, \ldots \quad \text{(i.e. } A_i A_j = 0 \text{ for any } i \neq j\text{)}$$

$$A_i \cap A_j = \phi, \quad \text{and} \quad \bigcup_{i=1} A_i = \mathrm{S}$$

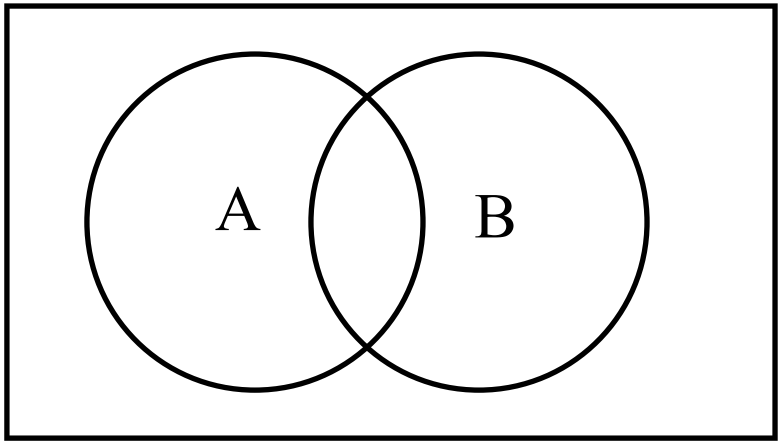


$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Detour: Approximations/Bounds/Inequalities

Why? A large part of information theory consists in finding *bounds* on certain performance measures

Approximations/Bounds: Union Bound



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

$$\mathbf{P(A_1 \cup A_2 \cup \ldots A_N) \leq \Sigma_{i=1..N} P(A_i)}$$

- Applications:
  - Getting bounds on BER (bit-error rates),
  - In general, bounding the tails of prob. distributions
- We will use this in the analysis of error probabilities with various coding schemes

  (see chap 3, Tse/Viswanath)

Approximations/Bounds: log(1+x)

- **$\log_2(1+x) \approx x$ for <u>small</u> x**

- Application: Shannon capacity w/ AWGN noise:
  - Bits-per-Hz = C/B = $\log_2(1+\gamma)$
  - If we can increase SNR (γ) linearly when γ is small (i.e. very poor, eg: cell-edge)…
  - … we get a <u>linear</u> increase in capacity.

- When γ is **<u>large</u>**, of course increase in γ gives only a diminishing return in terms of capacity: $\log(1+\gamma)$

Approximations/Bounds: Jensen's Inequality

Definition 6 A function $f(x)$ is said to be **convex** over an interval (a, b) if for every $x_1, x_2 \in (a, b)$ and $0 \le \lambda \le 1$,

$$f(\lambda x_1 + (1-\lambda)x_2)) \le \lambda f(x_1) + (1-\lambda)f(x_2)$$

A function is **strictly convex** if equality holds only if $\lambda = 0$ or $\lambda = 1$. □

Example 0.1 Convex x^2, e^x, $|x|$, $x \log x$.

Concave: $\log x$, $\sqrt{x}$

Second derivative > 0

Theorem 6 *If f is a convex function and X is a r.v. then*

$$Ef(X) \ge f(EX).$$

Put another way,

$$\sum_x p(x)f(x) \ge f\left(\sum_x p(x)x\right)$$

If f is strictly convex then equality in the theorem implies that $X = EX$ w.p. 1.
If f is concave then

$$Ef(X) \le f(EX).$$

Schwartz Inequality & Matched Filter

- Inner Product ($\mathbf{a}^T\mathbf{x}$) <= Product of Norms (i.e. |**a**||**x**|)
  - Projection length <= Product of Individual Lengths
- This is the *Schwartz Inequality*!
  - Equality happens when **a** and **x** are in the same direction (i.e. $\cos\theta = 1$, when $\theta = 0$)

- Application: "matched" filter
  - Received vector $\mathbf{y} = \mathbf{x} + \mathbf{w}$ (zero-mean AWGN)
  - Note: **w** is infinite dimensional
  - Project **y** to the subspace formed by the finite set of transmitted symbols **x**: y'
  - y' is said to be a "**sufficient statistic**" for detection, i.e. reject the noise dimensions outside the signal space.
  - This operation is called "*matching*" to the signal space (projecting)
  - Now, pick the x which is closest to y' in distance (ML detection = nearest neighbor)

Back to Probability…

Conditional Probability

- $P(A|B)$ = (conditional) probability that the outcome is in *A* given that we know the outcome in *B*

$$P(A|B) = \frac{P(AB)}{P(B)} \qquad P(B) \neq 0$$

- **Example: Toss one die.**

$$P(i = 3 \,|\, \text{i is odd}) =$$

- **Note that:** $P(AB) = P(B)P(A|B) = P(A)P(B|A)$

What is the value of knowledge that B occurred ?
How does it reduce uncertainty about A?
How does it change P(A) ?

Independence

- Events A and B are <u>independent</u> if ***P(AB) = P(A)P(B).***
- Also: $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$
- <u>Example</u>: A card is selected at random from an ordinary deck of cards.
  - A=event that the card is an ace.
  - B=event that the card is a diamond.

$P(AB) =$

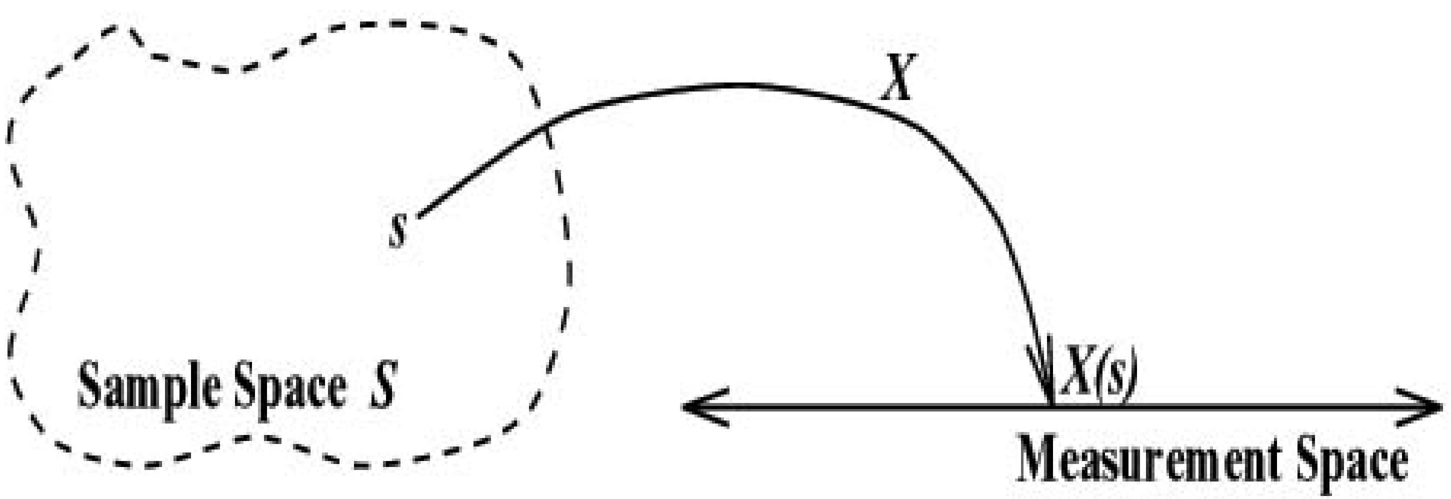
$P(A) =$

$P(B) =$

$P(A)P(B) =$

Random Variable as a Measurement

- Thus a random variable can be thought of as a <u>measurement</u> (yielding a real number) on an experiment
  - Maps "*events*" to "*real numbers*"
  - We can then talk about the pdf, define the mean/variance and other moments



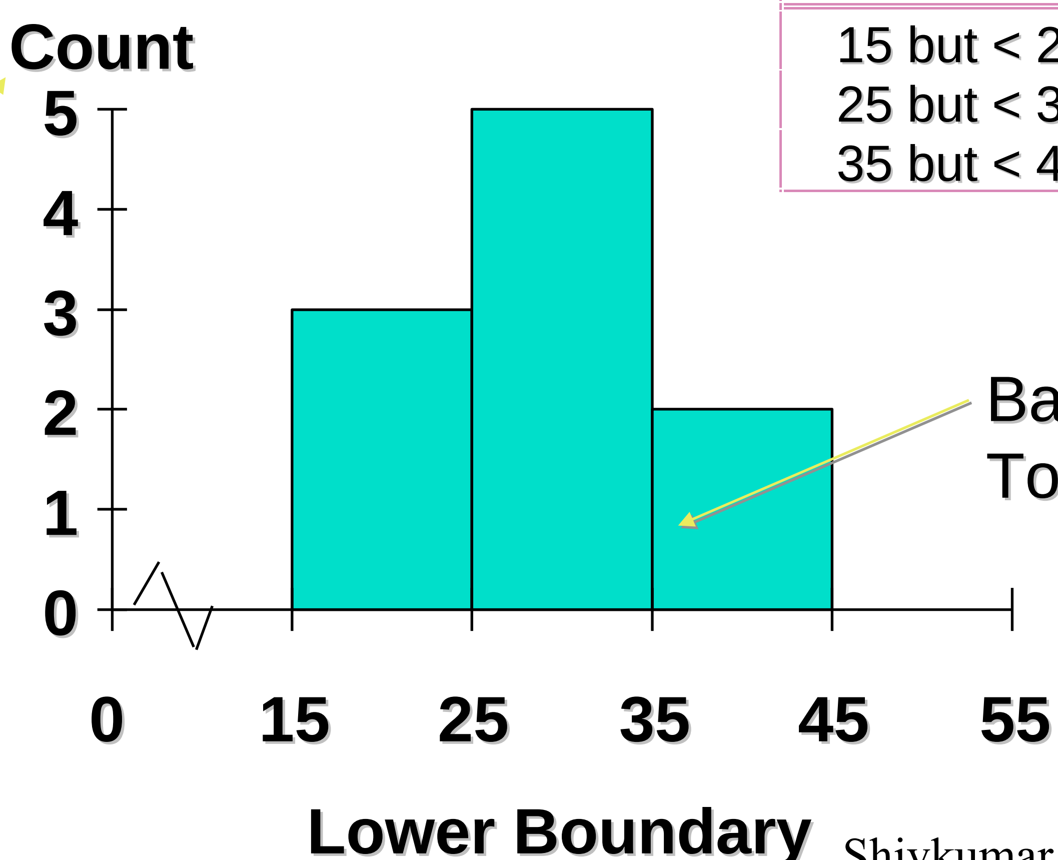
Histogram: Plotting Frequencies

Class	Freq.
15 but < 25	3
25 but < 35	5
35 but < 45	2

Frequency

Relative Frequency

Percent



Probability Distribution Function (pdf): continuous version of histogram

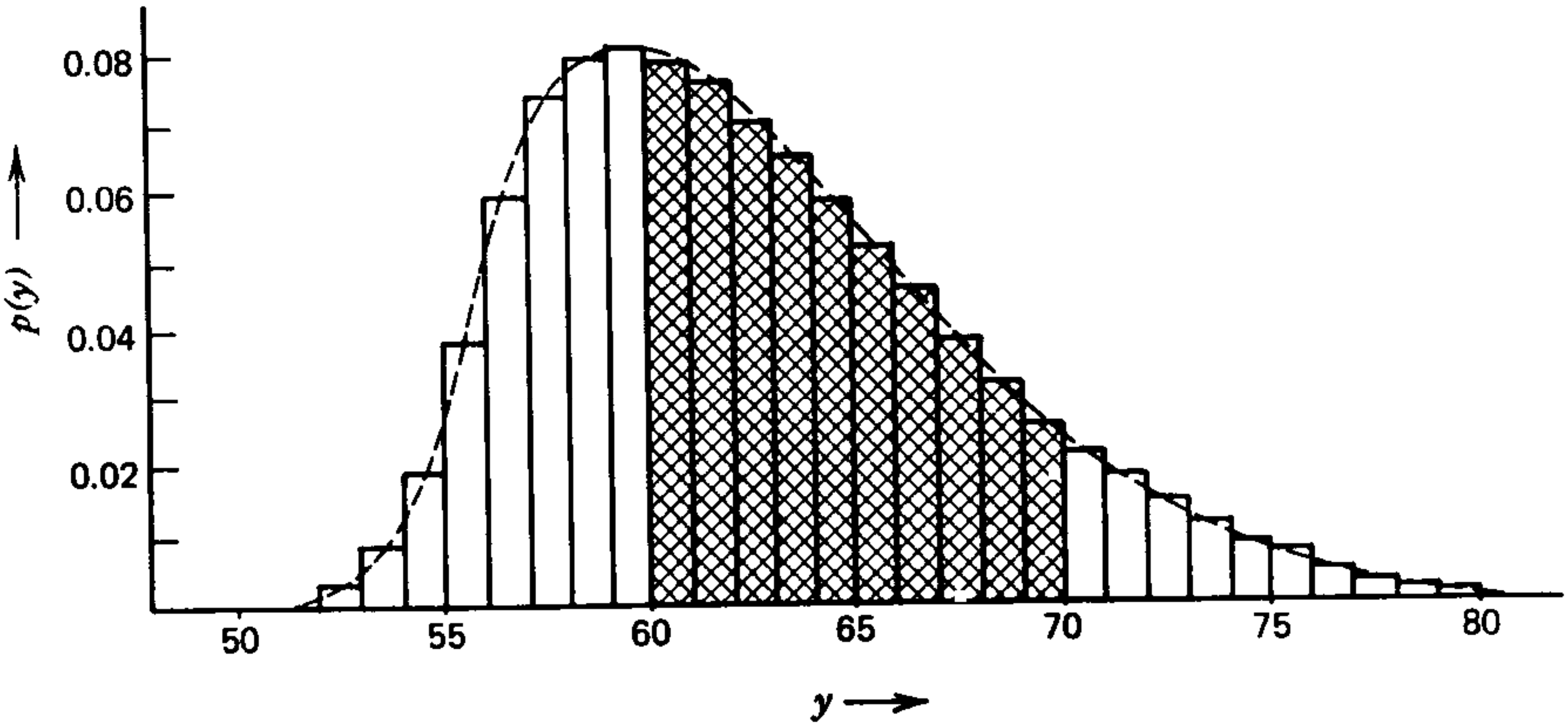


FIGURE 2.4. Probability distribution for a conceptual population of yield values.

a.k.a. frequency histogram, p.m.f (for discrete r.v.)

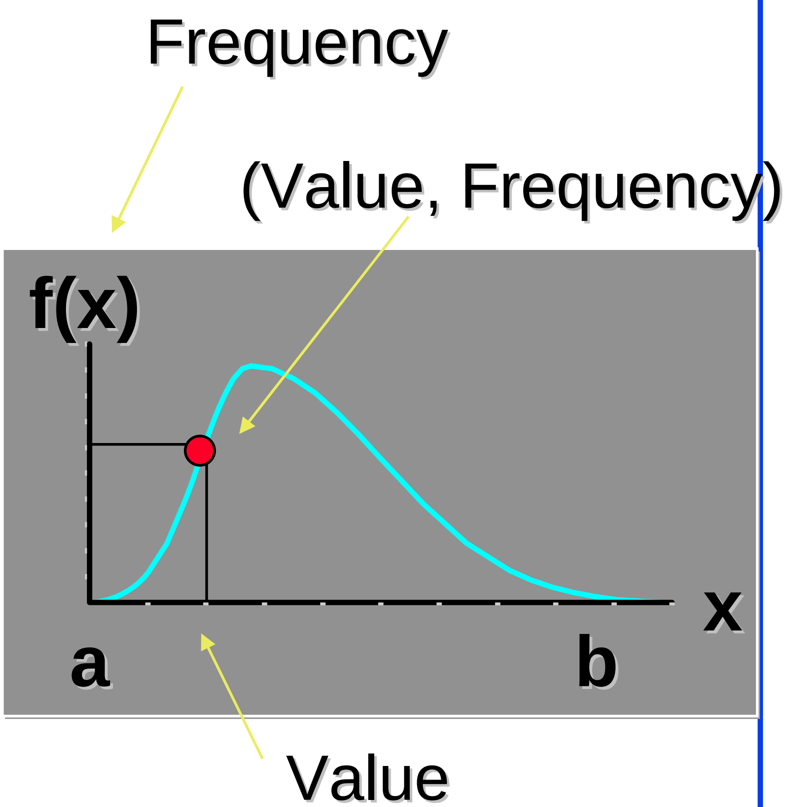
Continuous Probability Density Function

- 1. Mathematical Formula
- 2. Shows All Values, x, & Frequencies, f(x)
  - f(X) Is ***Not*** Probability
- 3. Properties

$$\int_{\text{All } X} f(x)dx = 1$$

(Area Under Curve)

$$f(x) \geq 0, \ a \leq x \leq b$$



Cumulative Distribution Function

- The cumulative distribution function (CDF) for a random variable ***X*** is

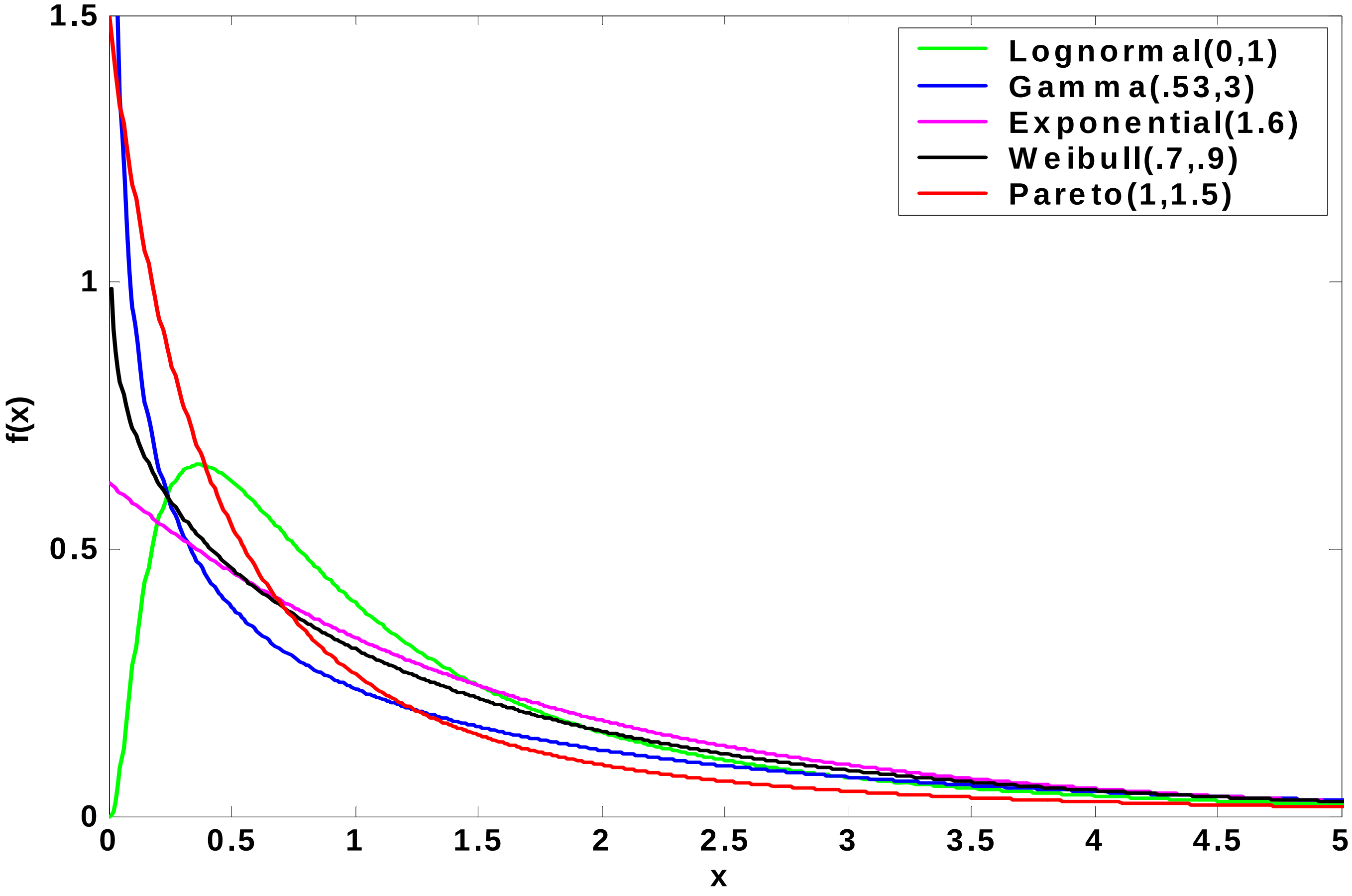
$$F_X(x) = P(X \le x) = P(\{s \in S \mid X(s) \le x\})$$

- Note that $F_X(x)$ is non-decreasing in ***x***, i.e.

$$x_1 \le x_2 \Rightarrow F_x(x_1) \le F_x(x_2)$$

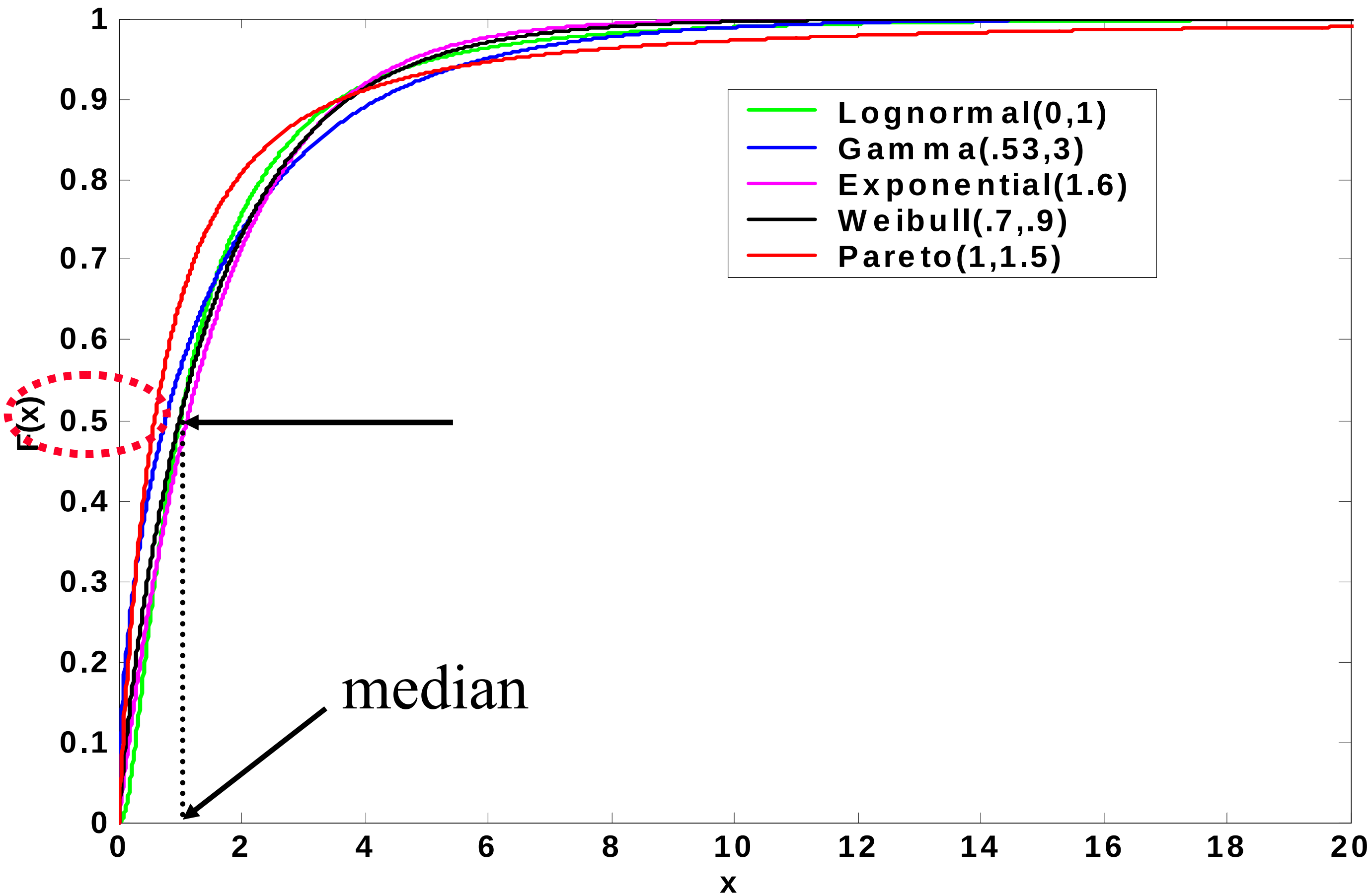
- Also $\lim_{x \to -\infty} F_x(x) = 0$ and $\lim_{x \to \infty} F_x(x) = 1$

Probability density functions (pdf)



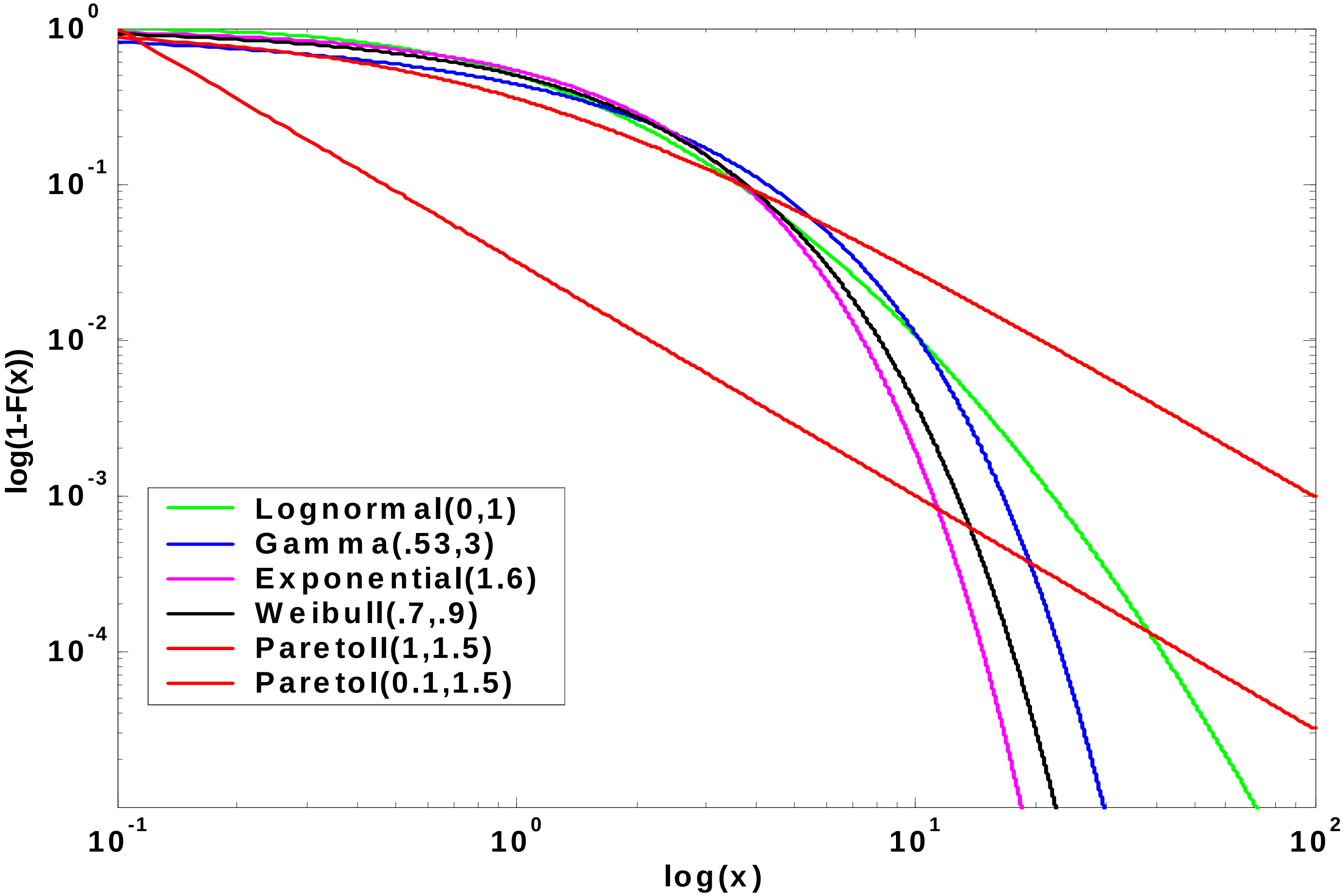
Emphasizes main body of distribution, frequencies, various modes (peaks), variability, skews

Cumulative Distribution Function (CDF)



Emphasizes skews, easy identification of median/quartiles, converting uniform rvs to other distribution rvs

Complementary CDFs (CCDF)



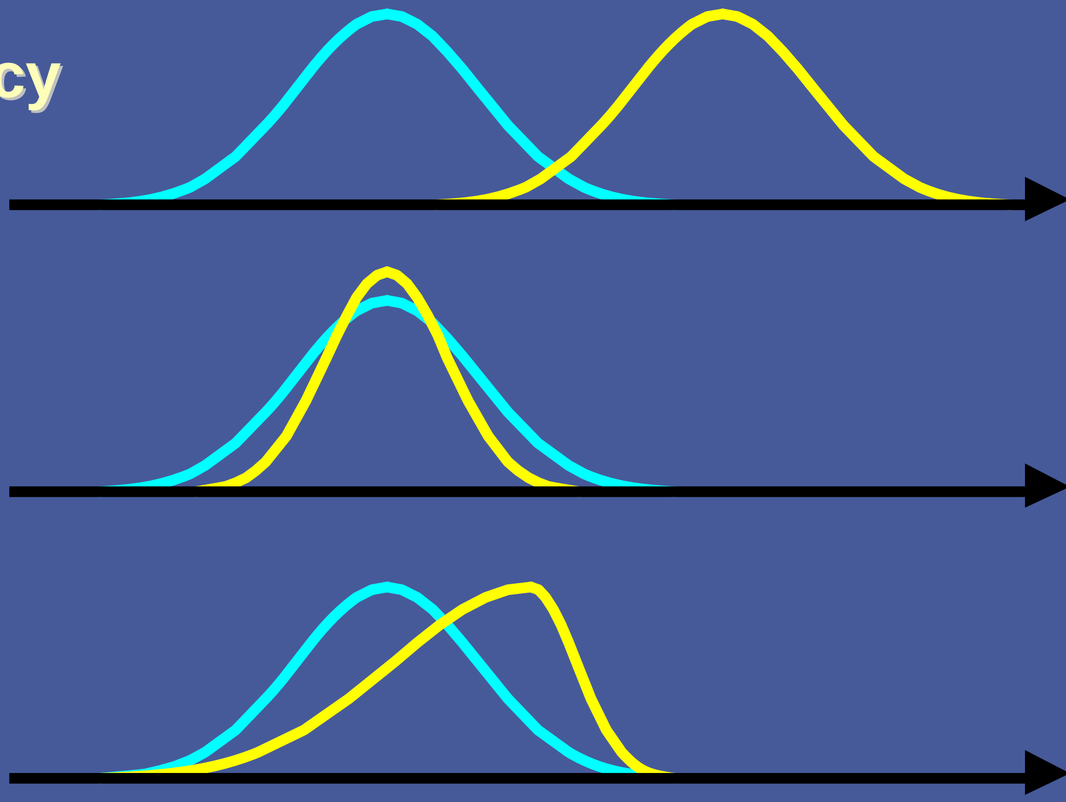
Useful for focussing on "tails" of distributions:
Line in a log-log plot => "heavy" tail

Numerical Data Properties

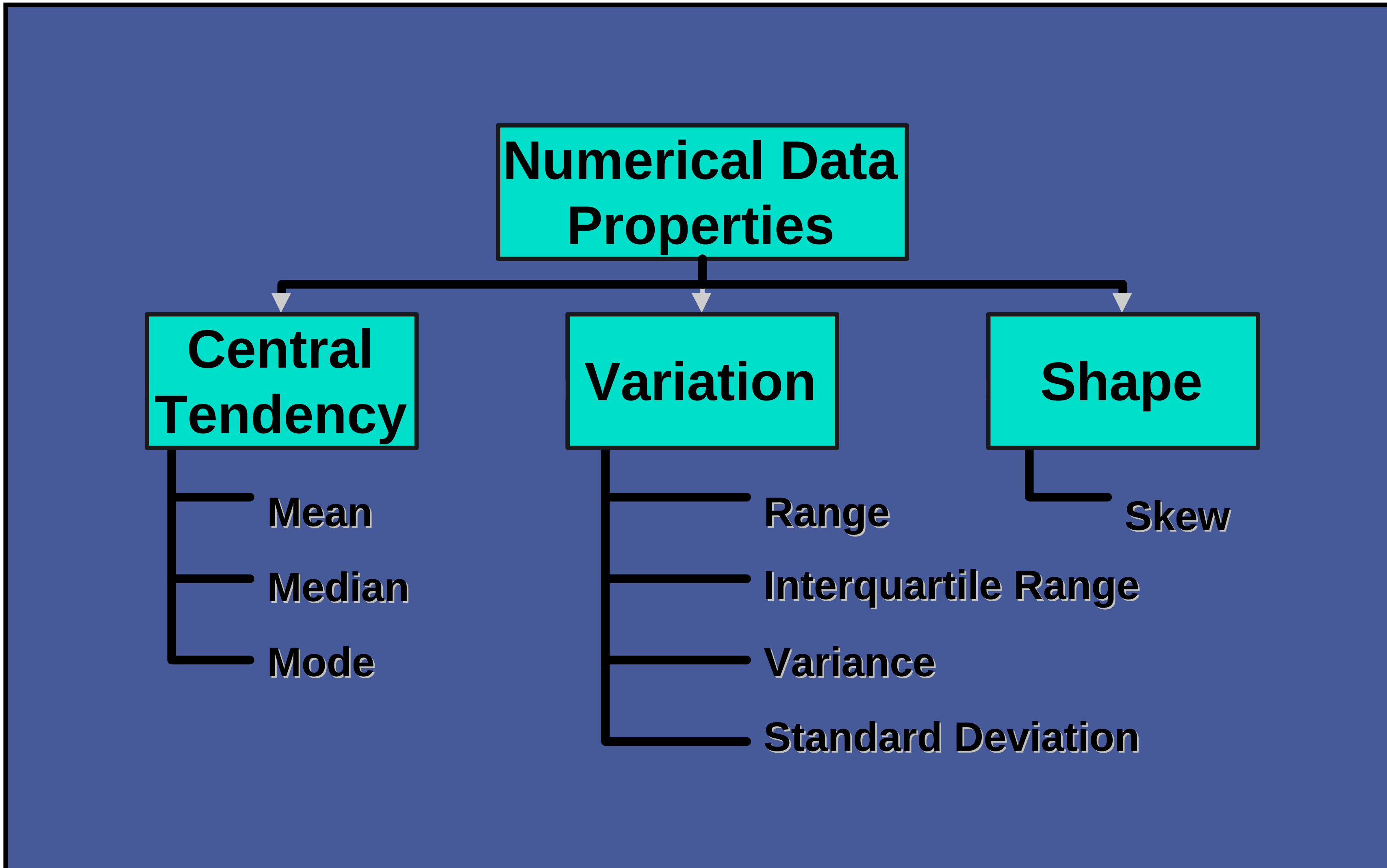
Central Tendency (Location)

Variation (Dispersion)

Shape



Numerical Data Properties & Measures



Expectation of a Random Variable: E[X]

- The <u>expectation</u> (average) of a (discrete-valued) random variable ***X*** is

$$\overline{X} = E(X) = \sum_{x=-\infty}^{\infty} xP(X = x) = \sum_{-\infty}^{\infty} xP_X(x)$$

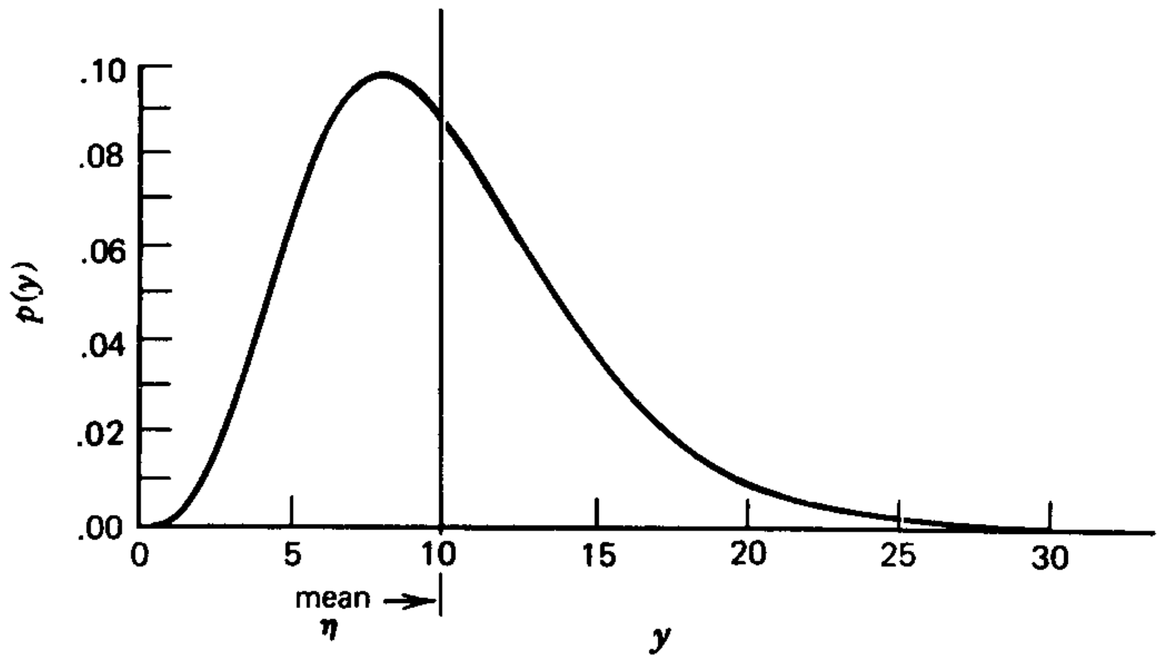


FIGURE 2.7. The mean $\eta = E(y)$ as the center of gravity of a distribution.

Continuous-valued Random Variables

- Thus, for a continuous random variable ***X***, we can define its probability density function (pdf)

$$f_x(x) = F'_X(x) = \frac{dF_X(x)}{dx}$$

- Note that since $F_X(x)$ is non-decreasing in ***x*** we have

$$f_X(x) \geq 0$$ for all ***x***.

Expectation of a Continuous Random Variable

- The expectation (average) of a continuous random variable ***X*** is given by

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

- Note that this is just the continuous equivalent of the discrete expectation

$$E(X) = \sum_{x=-\infty}^{\infty} x P_X(x)$$

Other Measures: Median, Mode

- **Median** = F^{-1} (0.5), where F = CDF
  - Aka 50% percentile element
  - I.e. Order the values and pick the middle element
  - Used when distribution is skewed
  - Considered a "robust" measure

- **Mode:** Most frequent or highest probability value
  - Multiple modes are possible
  - Need not be the "central" element
  - Mode may not exist (eg: uniform distribution)
  - Used with categorical variables

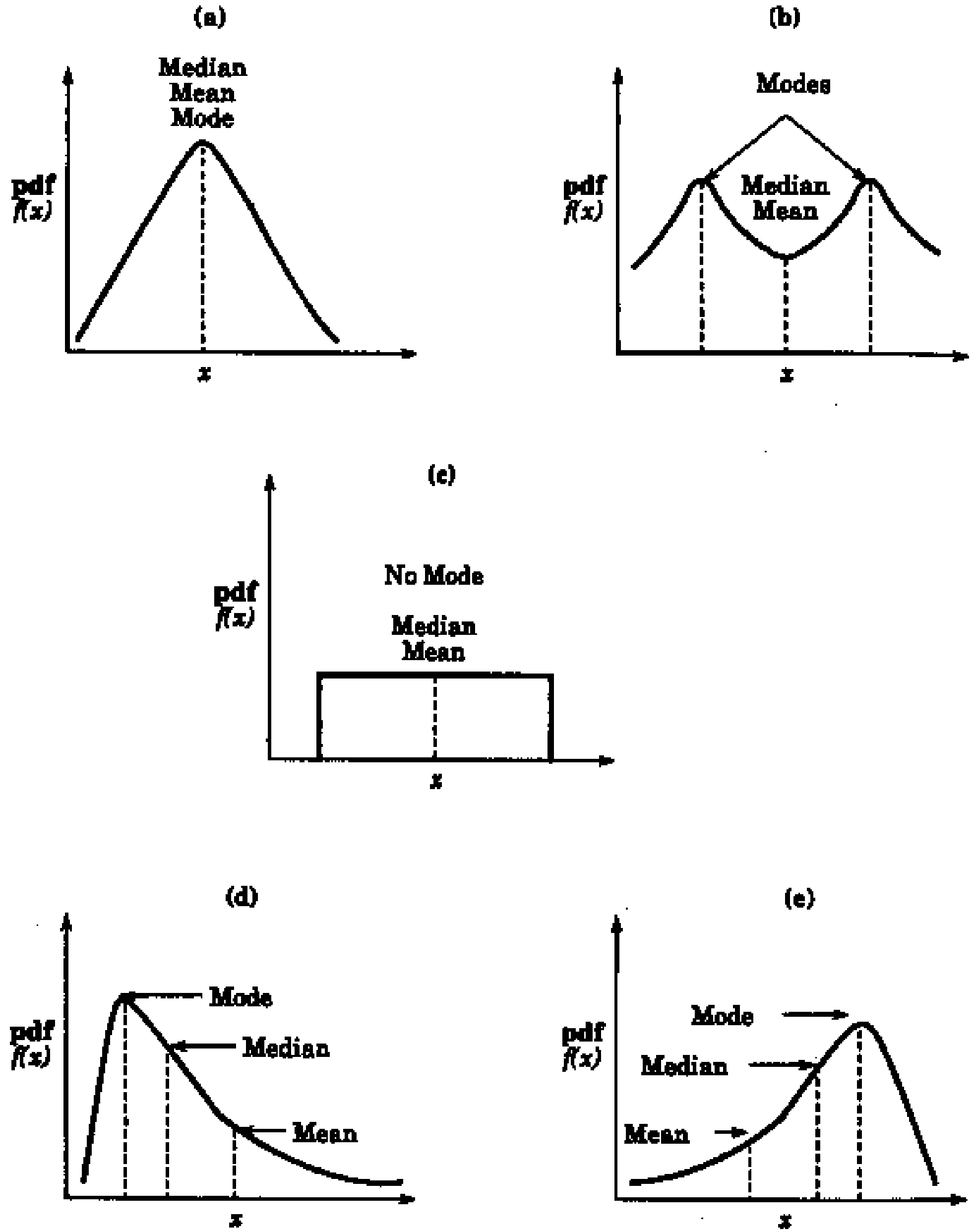


FIGURE 12.1 Five distributions showing relationships among mean, median, and mode.

Indices/Measures of Spread/Dispersion: Why Care?

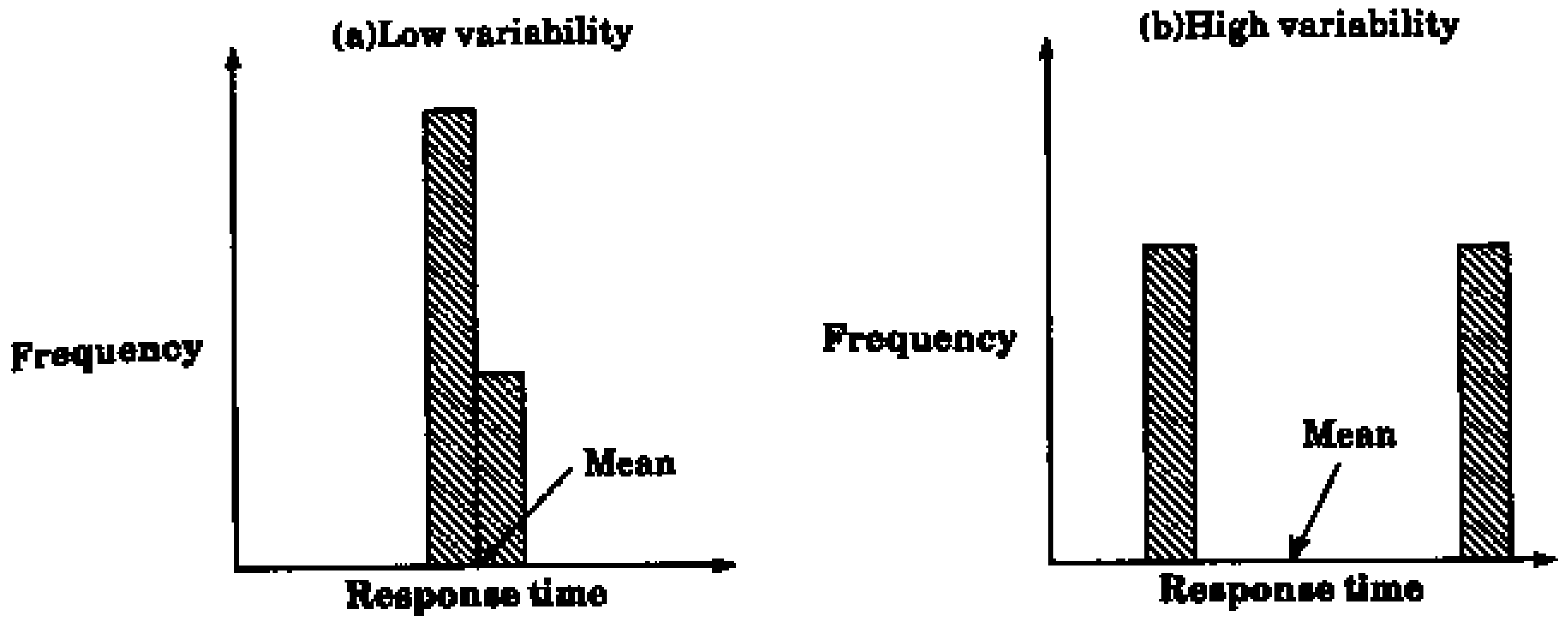


FIGURE 12.3 Histograms of response times of two systems.

You can drown in a river of average depth 6 inches!

<u>Lesson</u>: The measure of uncertainty or dispersion may matter more than the index of central tendency

Standard Deviation, Coeff. Of Variation, SIQR

- **Variance**: second moment around the mean:
  - $\sigma^2 = E[(X-\mu)^2]$
- **Standard deviation** = σ

$$\mathrm{stdv}(x) = \sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \sqrt{\mu_2' - \mu^2},$$

- **Coefficient of Variation (C.o.V.)=** σ/μ
- **SIQR**= Semi-Inter-Quartile Range (used with median = 50th percentile)
  - (75th percentile – 25th percentile)/2

Covariance and Correlation: Measures of Dependence

- **<u>Covariance</u>:** $\langle (x_i - \mu_i)(x_j - \mu_j) \rangle = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle ,$
  - For i = j, covariance = variance!
  - Independence => covariance = 0 (not vice-versa!)

- **<u>Correlation (coefficient)</u>** is a normalized (or scaleless) form of covariance:

$$\mathrm{cor}(x_i, x_j) \equiv \frac{\mathrm{cov}(x_i, x_j)}{\sigma_i \sigma_j},$$

  - Between –1 and +1.
    - Zero => no correlation (uncorrelated).
    - Note: uncorrelated <u>DOES NOT</u> mean independent!

Random Vectors & Sum of R.V.s

- Random Vector = $[X_1, \ldots, X_n]$, where Xi = r.v.
- Covariance Matrix:
  - **K** is an nxn matrix…
  - $K_{ij} = Cov[X_i, X_j]$
  - $K_{ii} = Cov[X_i, X_i] = Var[X_i]$

- Sum of *independent* R.v.s
  - $Z = X + Y$
  - PDF of Z is the *convolution* of PDFs of X and Y

    $p_Z(z) = p_X(x) * p_Y(y)$. Can use transforms!

Characteristic Functions & Transforms

- Characteristic function: a special kind of expectation

The distribution of a random variable X can be determined from its *characteristic function*, defined as

$$\phi_X(\nu) \triangleq \mathbf{E}[e^{j\nu X}] = \int_{-\infty}^{\infty} p_X(x) e^{j\nu x} dx. \quad \text{(B.10)}$$

- Captures all the *moments*, and is related to the *IFT of pdf*:

We see from (B.10) that the characteristic function $\phi_X(\nu)$ of $X(t)$ is the inverse Fourier transform of the distribution $p_X(x)$ evaluated at $f = \nu/(2\pi)$. Thus we can obtain $p_X(x)$ from $\phi_X(\nu)$ as

$$p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(\nu) e^{-j\nu x} dx. \quad \text{(B.11)}$$

This will become significant in finding the distribution for sums of random variables.

Important (Discrete) Random Variable: Bernoulli

- The simplest possible measurement on an experiment:
  - **Success** (***X* = *1*)** or **failure *(X = 0)***.
- Usual notation:

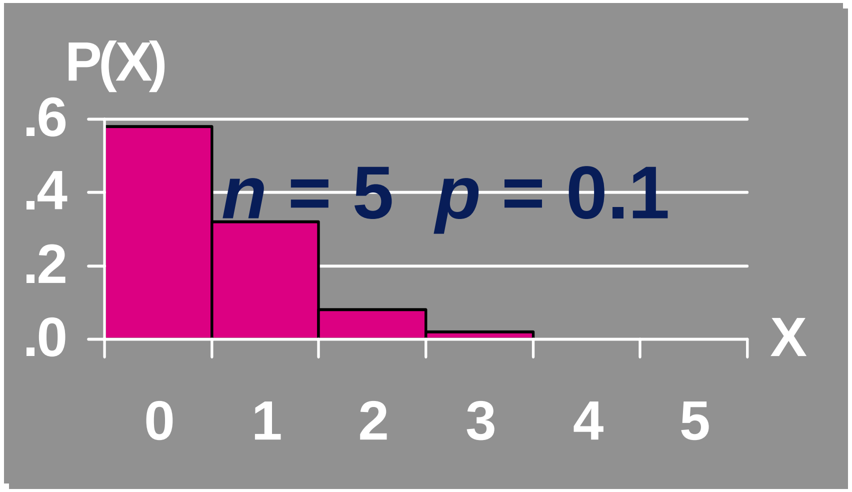
$$P_X(1) = P(X = 1) = p \qquad P_X(0) = P(X = 0) = 1 - p$$

  - ***E(X)=***

Binomial Distribution

$$p(Y=k) = \binom{n}{k} p^k (1-p)^{n-k},$$

$$\binom{n}{k} \triangleq \frac{n!}{k!(n-k)!}.$$

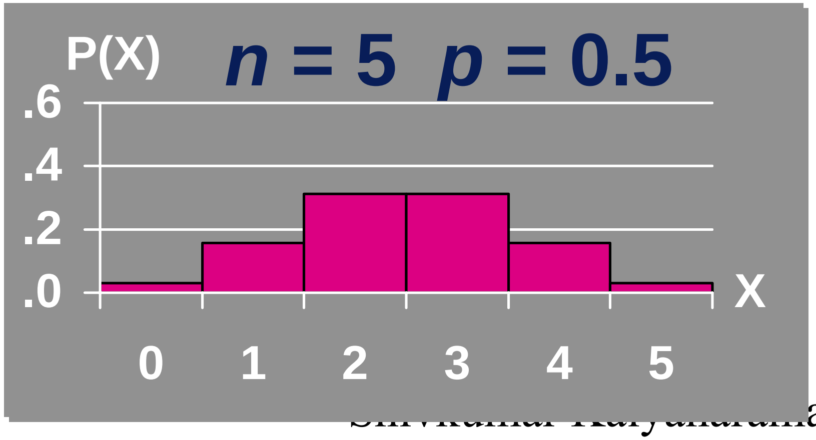


Mean

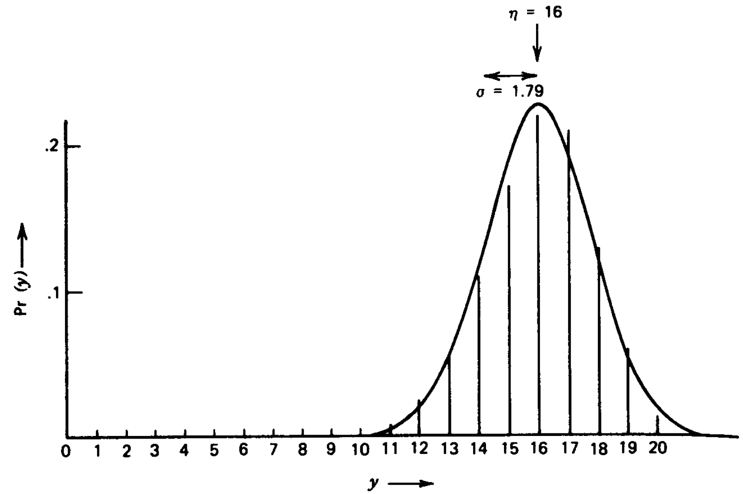
$$\mu = E(x) = np$$

Standard Deviation

$$\sigma = \sqrt{np(1-p)}$$

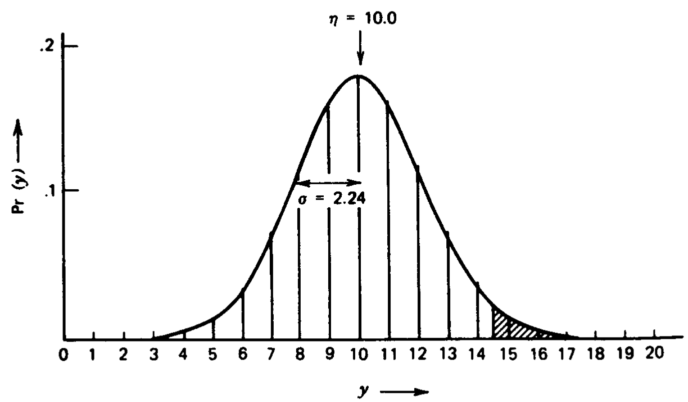


Binomial can be skewed or normal



(c) Binomial distribution with mean $p = 0.8$ and $n = 20$.

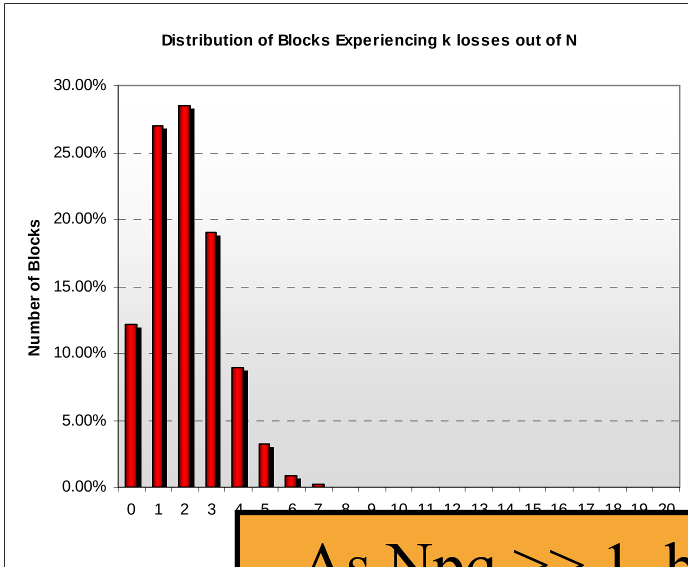
Depends upon p and n !

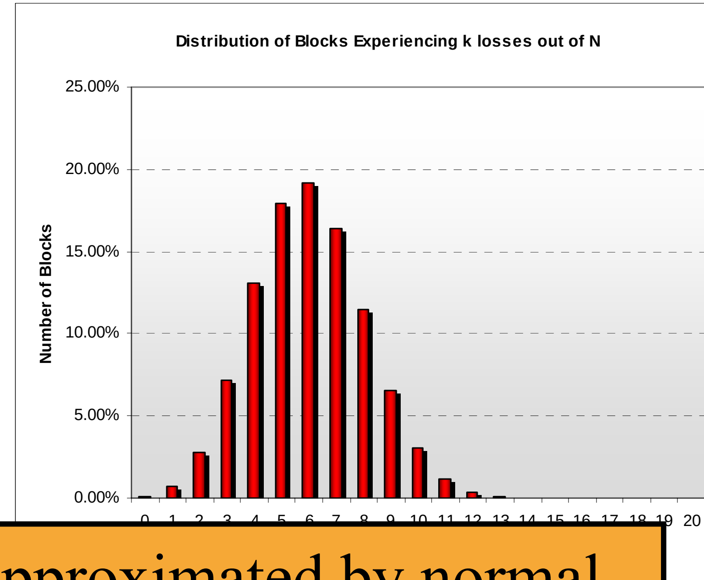


(d) Binomial distribution with mean $p = 0.5$ and $n = 20$.

FIGURE 5.4. (continued)

Binomials for different p, N =20

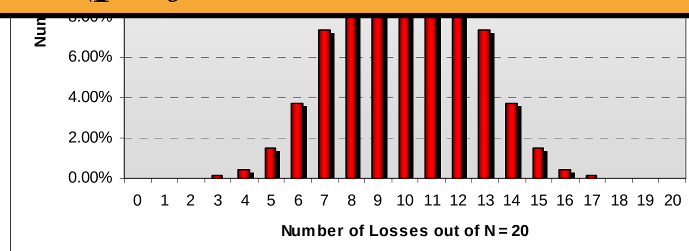




As Npq >> 1, better approximated by normal distribution (esp) near the mean:

$\Rightarrow$*symmetric, sharp peak at mean, exponential-square (e^{-x^2}) decay of tails*

(pmf concentrated near mean)



Important Random Variable: Poisson

- A Poisson random variable ***X*** is defined by its PMF: (limit of binomial)

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad x = 0, 1, 2, \ldots$$

Where λ > 0 is a constant

- Exercise: Show that

$$\sum_{x=0}^{\infty} P_X(x) = 1$$

and ***E(X)*** = λ

- Poisson random variables are good for *counting frequency of occurrence*: like the number of customers that arrive to a bank in one hour, or the number of packets that arrive to a router in one second.

Important Continuous Random Variable: Exponential

- Used to represent time, e.g. until the next arrival
- Has PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

for some $\lambda > 0$

- Properties:

$$\int_0^\infty f_X(x)dx = 1 \quad \text{and} \quad E(X) = \frac{1}{\lambda}$$

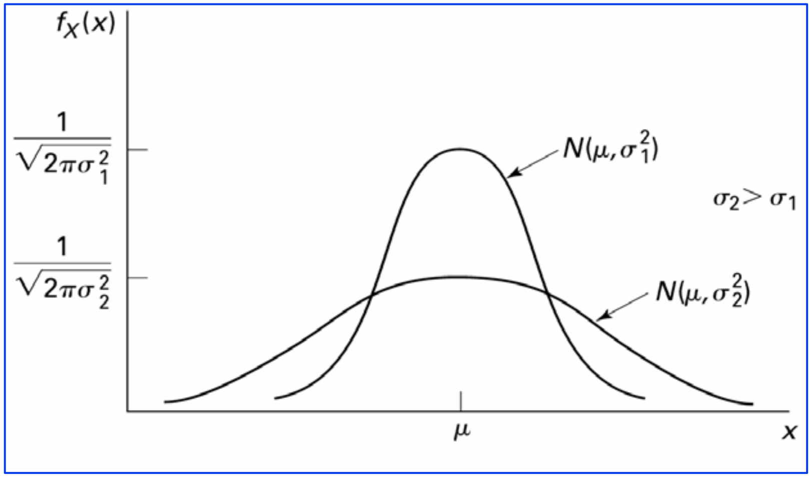
  - Need to use integration by Parts!

Gaussian/Normal Distribution

References:
Appendix A.1 (Tse/Viswanath)
Appendix B (Goldsmith)

Gaussian/Normal

- **Normal Distribution:** Completely characterized by mean (μ) and variance (σ^2)

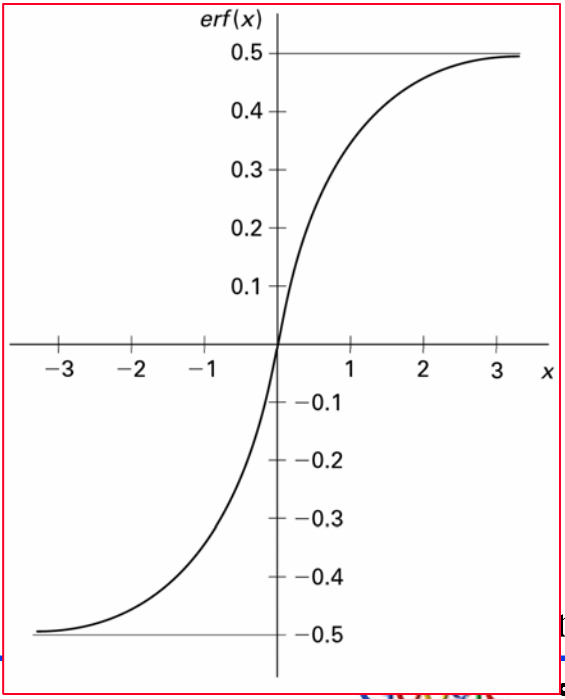


- **Q-function:** one-sided tail of normal pdf

$$Q(z) \triangleq p(x > z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

- **erfc():** two-sided tail.
- So:

$$Q(z) = \frac{1}{2} \operatorname{erfc}\left(\frac{z}{\sqrt{2}}\right)$$



Normal Distribution: Why?

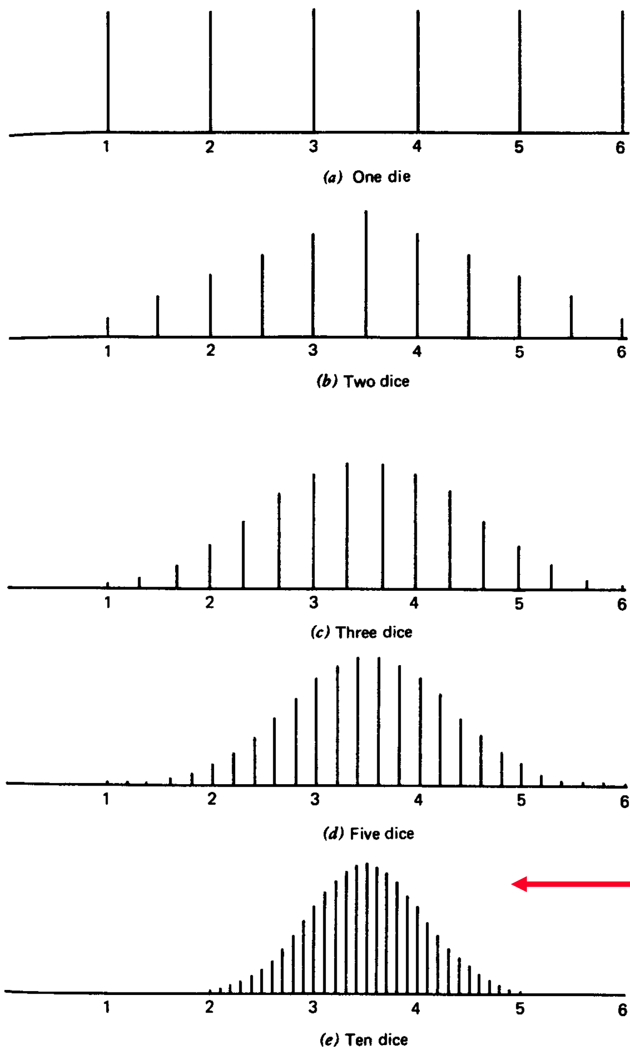


FIGURE 2.10. Distribution of average scores from throwing various numbers of dice.

Uniform distribution looks <u>nothing</u> like bell shaped (gaussian)! Large spread (σ)!

<u>CENTRAL LIMIT TENDENCY!</u>

<u>Sum of r.v.s</u> from a uniform distribution after <u>very few</u> samples looks remarkably normal

<u>BONUS</u>: it has decreasing σ !

Gaussian: Rapidly Dropping Tail Probability!

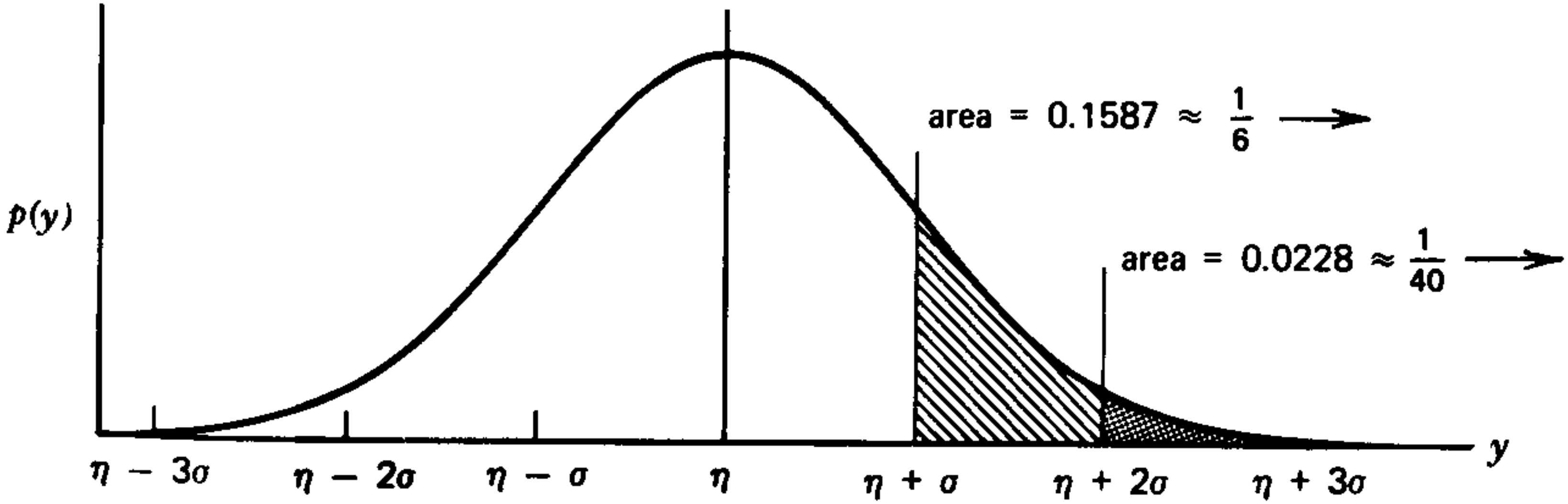


FIGURE 2.12. Tail areas of the normal distribution.

Why? Doubly exponential PDF (e^{-z^2} term…)
A.k.a: "Light tailed" (not heavy-tailed).
No skew or tail => don't have two worry
about > 2nd order parameters (mean, variance)
Fully specified with just mean and variance (2nd order)

Height & Spread of Gaussian Can Vary!

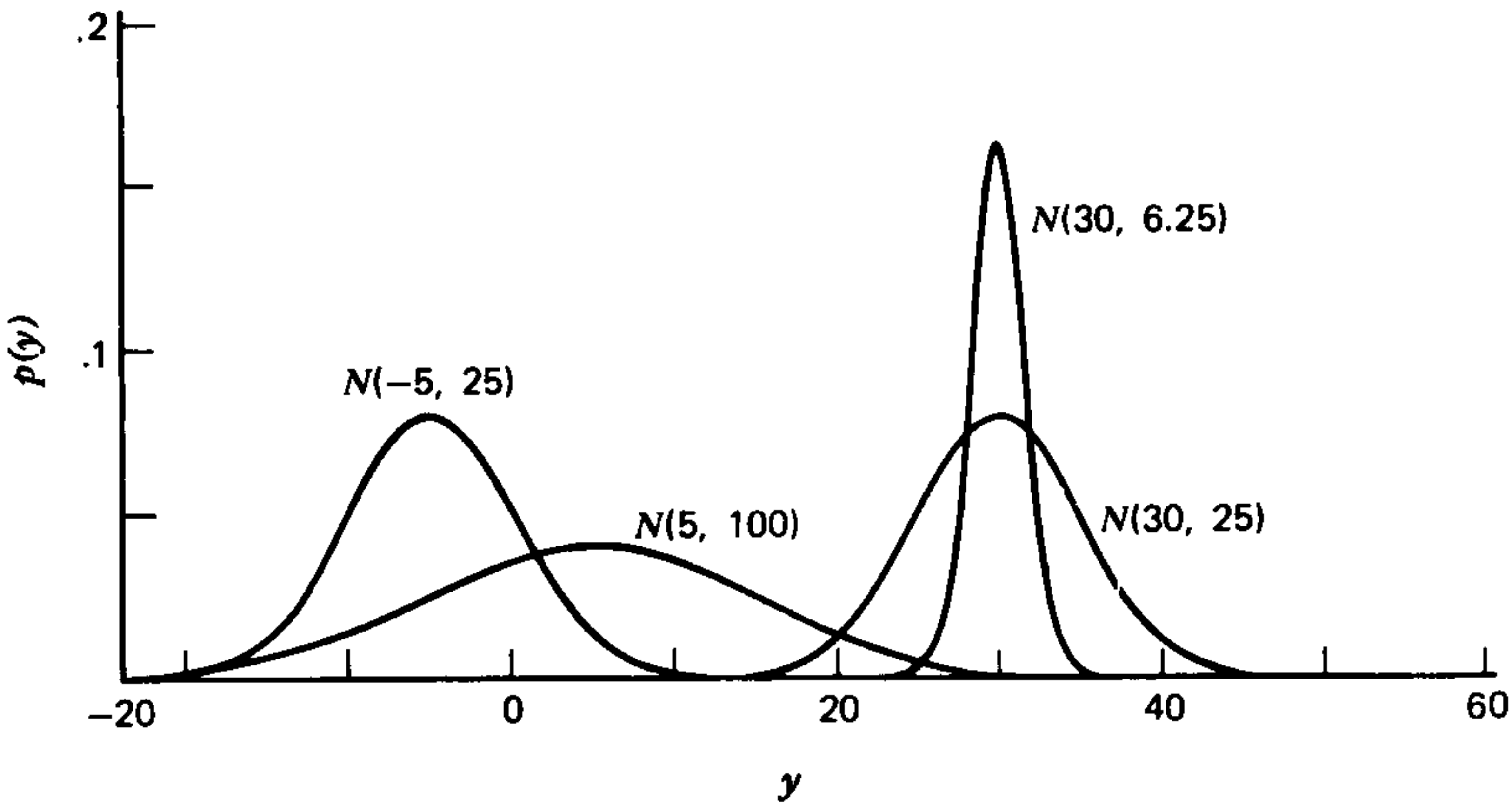
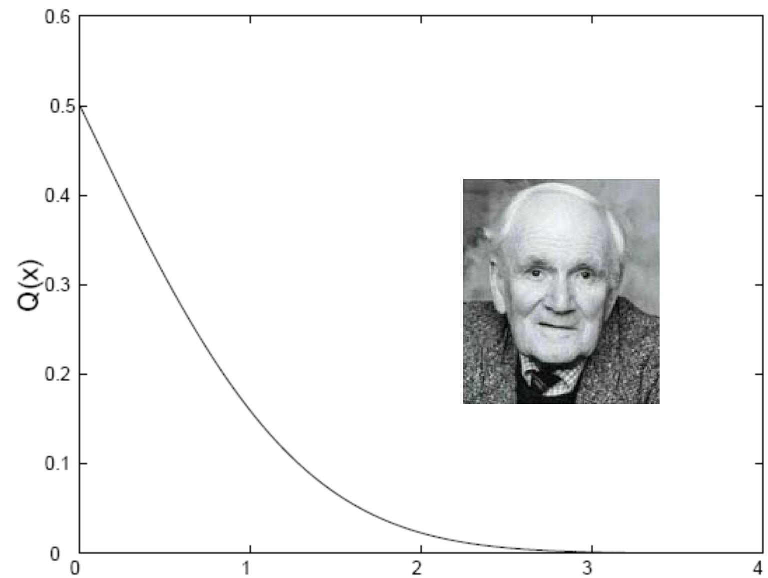


FIGURE 2.11. Normal distributions with different means and variances.

Gaussian R.V.

- **Standard Gaussian** $\mathcal{N}(0,1)$ $\quad f(w) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right), \qquad w \in \Re.$

- Tail: **Q(x)** $\quad Q(a) := \mathbb{P}\{w > a\}.$
  - tail decays exponentially!



$$\frac{1}{\sqrt{2\pi}a}\left(1-\frac{1}{a^2}\right)e^{-a^2/2} < Q(a) < e^{-a^2/2}, \qquad a > 1.$$

- Gaussian property preserved w/ linear transformations:

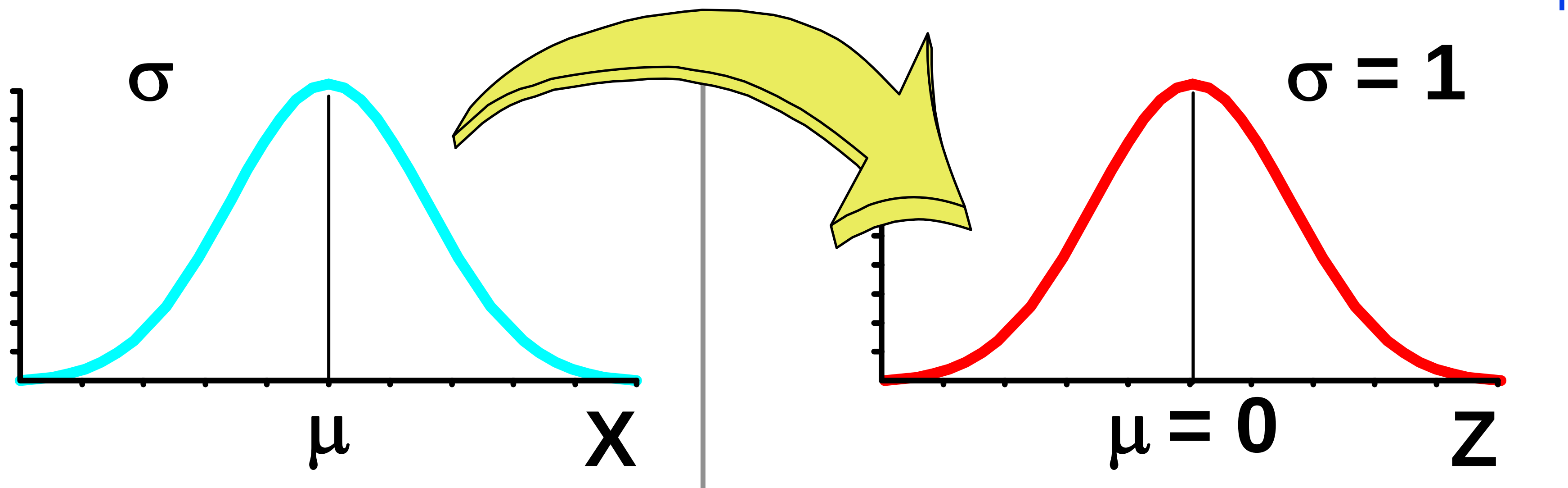
$$\sum_{i=1}^{n} c_i x_i \sim \mathcal{N}\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right).$$

Standardize the Normal Distribution

$$Z = \frac{X - \mu}{\sigma}$$

Normal Distribution

Standardized Normal Distribution

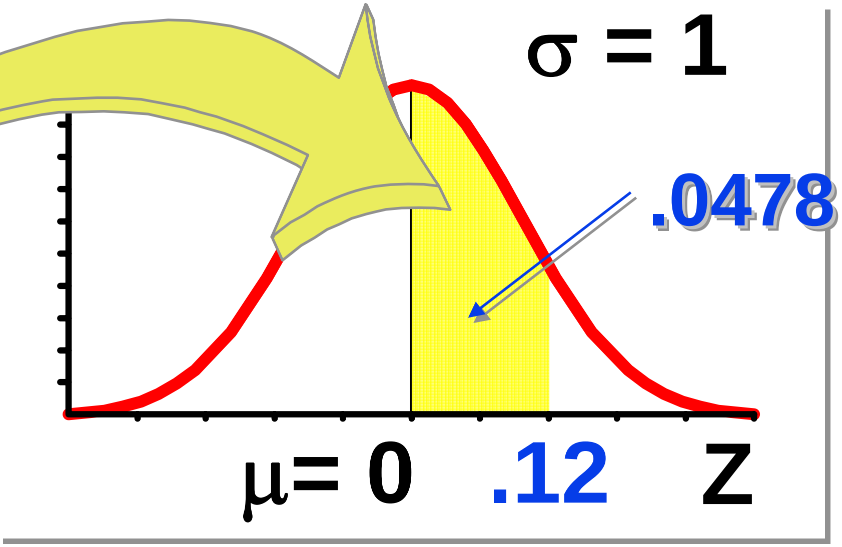


One table!

Obtaining the Probability

Standardized Normal Probability Table (Portion)

Z	.00	.01	.02
0.0	.0000	.0040	.0080
0.1	.0398	.0438	**.0478**
0.2	.0793	.0832	.0871
0.3	.1179	.1217	.1255

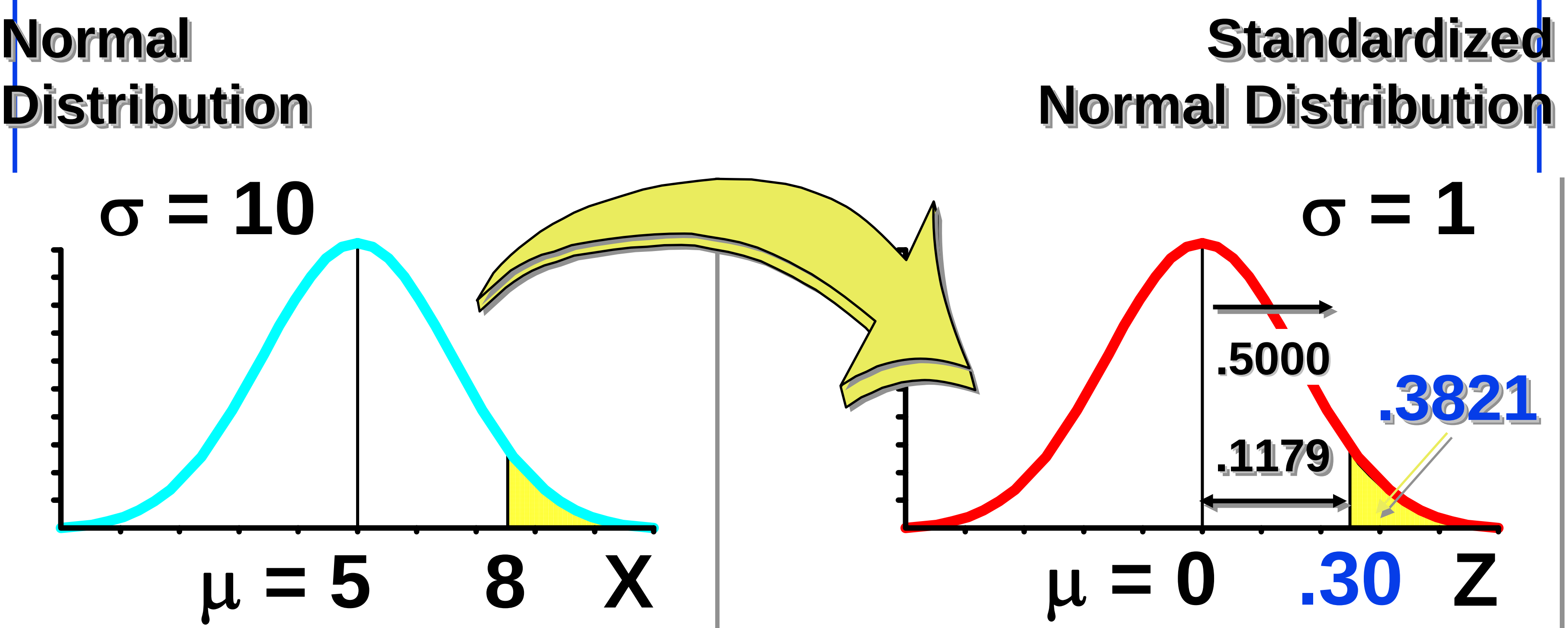


Probabilities

Shaded area exaggerated

Example
P($X \geq 8$)

$$Z = \frac{X - \mu}{\sigma} = \frac{8 - 5}{10} = .30$$



Shaded area exaggerated

Q-function: Tail of Normal Distribution

$Q(z) = P(Z > z) = 1 - P[Z < z]$

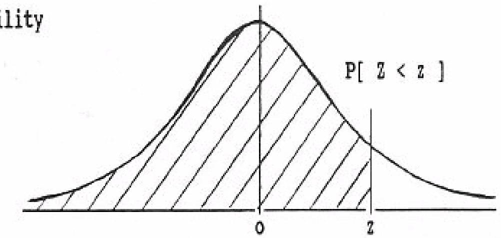
Rensselaer Polytechnic Institute

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z i.e.

$$P[\,Z < z\,] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9999	0.9999	0.9999	1.0000

Sampling from Non-Normal Populations

- **Central Tendency**

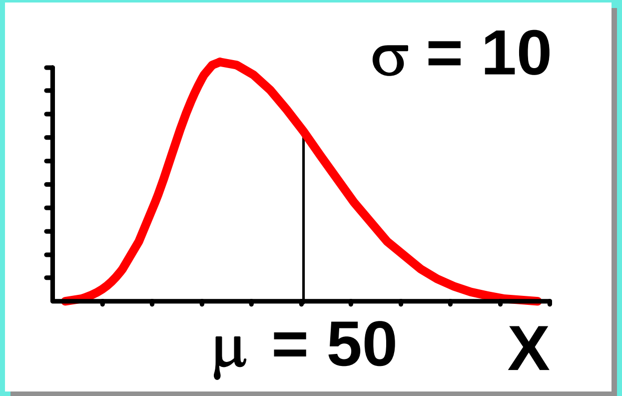
$$\mu_{\bar{x}} = \mu$$

- **Dispersion**

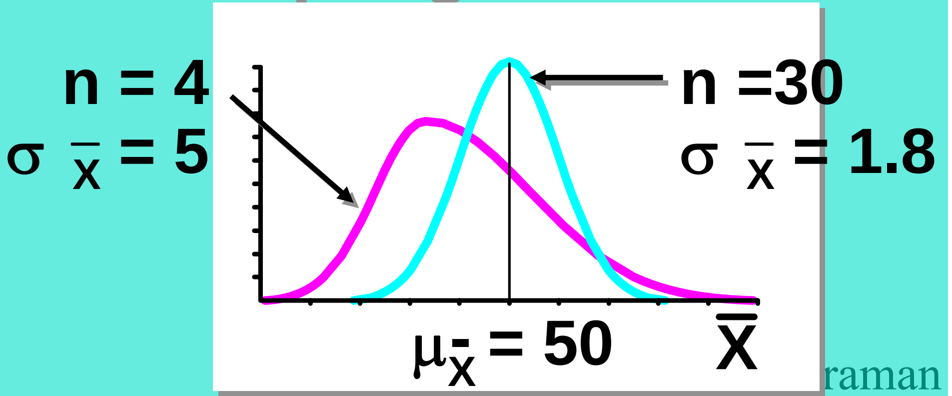
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Sampling with replacement

Population Distribution

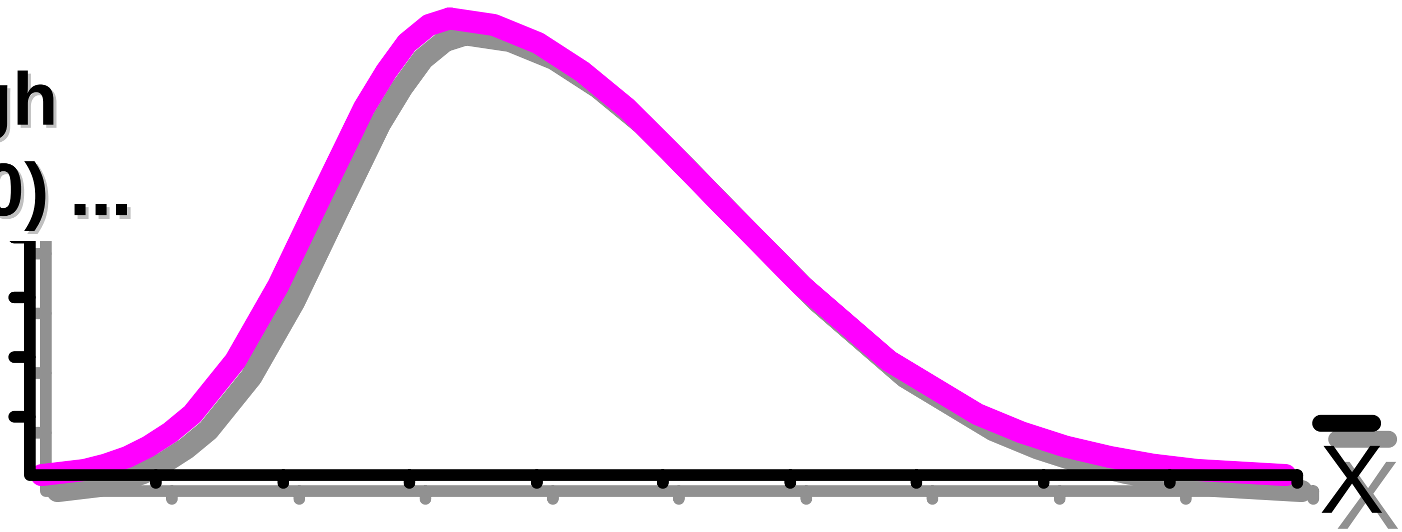


Sampling Distribution



Central Limit Theorem (CLT)

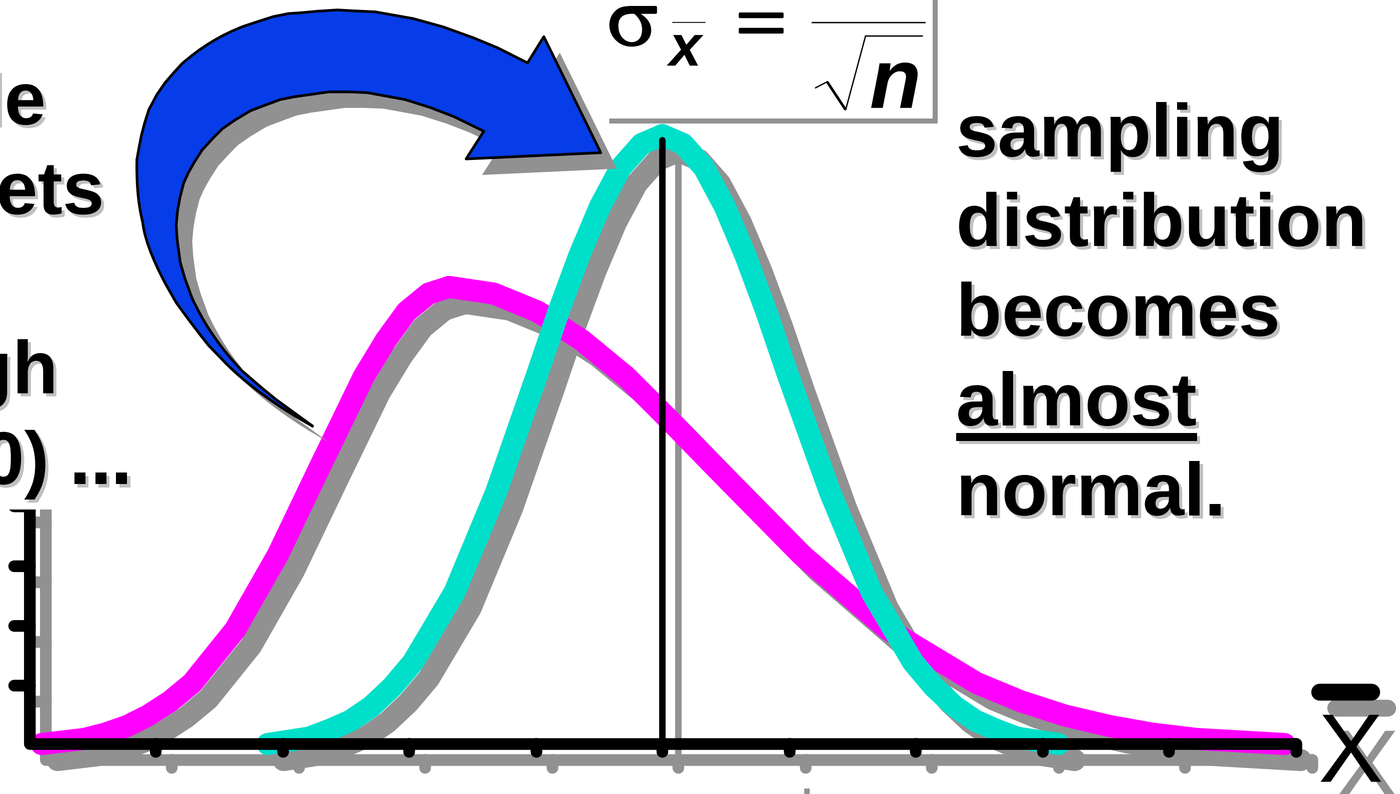
As sample size gets large enough ($n \geq 30$) ...



Central Limit Theorem (CLT)

As sample size gets large enough ($n \geq 30$) ...

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$



sampling distribution becomes <u>almost</u> normal.

$$\mu_{\bar{x}} = \mu$$

Aside: Caveat about CLT

- Central limit theorem works if original distribution are <u>not</u> heavy tailed
  - Need to have enough samples. Eg: with multipaths, if there is not rich enough scattering, the convergence to normal may have not happened yet
- Moments converge to limits
- Trouble with aggregates of "heavy tailed" distribution samples
- Rate of convergence to normal also varies with distributional skew, and dependence in samples

- Non-classical version of CLT for some cases (heavy tailed)…
  - Sum converges to stable Levy-noise (heavy tailed and long-range dependent auto-correlations)

Gaussian Vectors & Other Distributions

References:

Appendix A.1 (Tse/Viswanath)

Appendix B (Goldsmith)

Gaussian Vectors (Real-Valued)

- Collection of i.i.d. gaussian r.vs: $\mathbf{w} = (w_1, \dots, w_n)^t$

$$f(\mathbf{w}) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{\|\mathbf{w}\|^2}{2}\right), \qquad \mathbf{w} \in \Re^n.$$

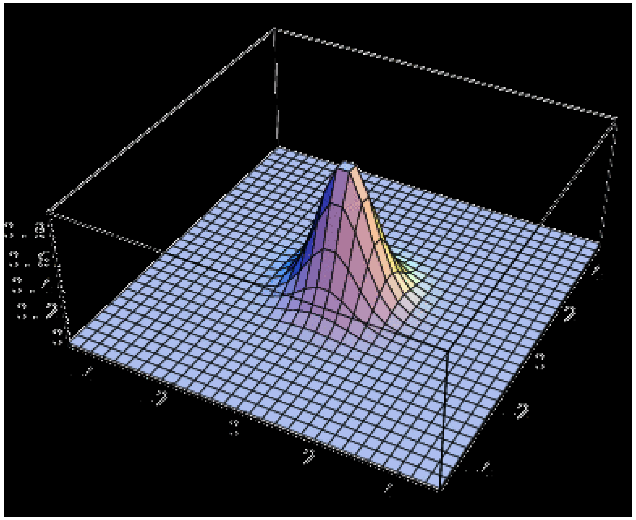
$$\|\mathbf{w}\| := \sqrt{\sum_{i=1}^n w_i^2}$$

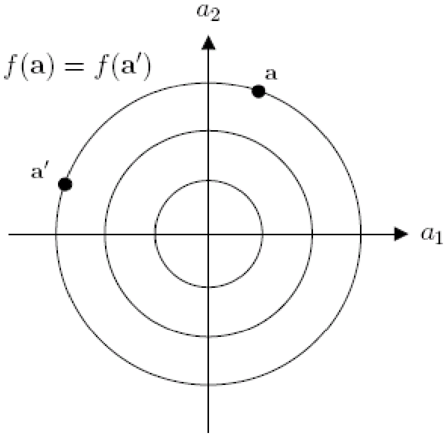
Euclidean distance from the origin to **w**

The density f(**w**) depends only on the magnitude of **w**, i.e. $\|\mathbf{w}\|^2$

Orthogonal transformation O (i.e., $O^tO = OO^t = I$) preserves the magnitude of a vector

2-d Gaussian Random Vector





Level sets (isobars) are circles

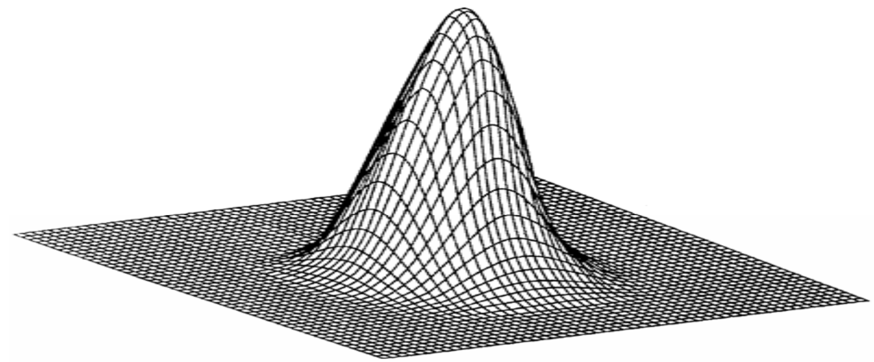
- **w** has the same distribution in *any orthonormal basis*.
- Distribution of w is *invariant to rotations and reflections i.e.* **Qw ~ w**
    - **w** does not prefer any specific direction ("*isotropic*")

Gaussian Random Vectors (Contd)

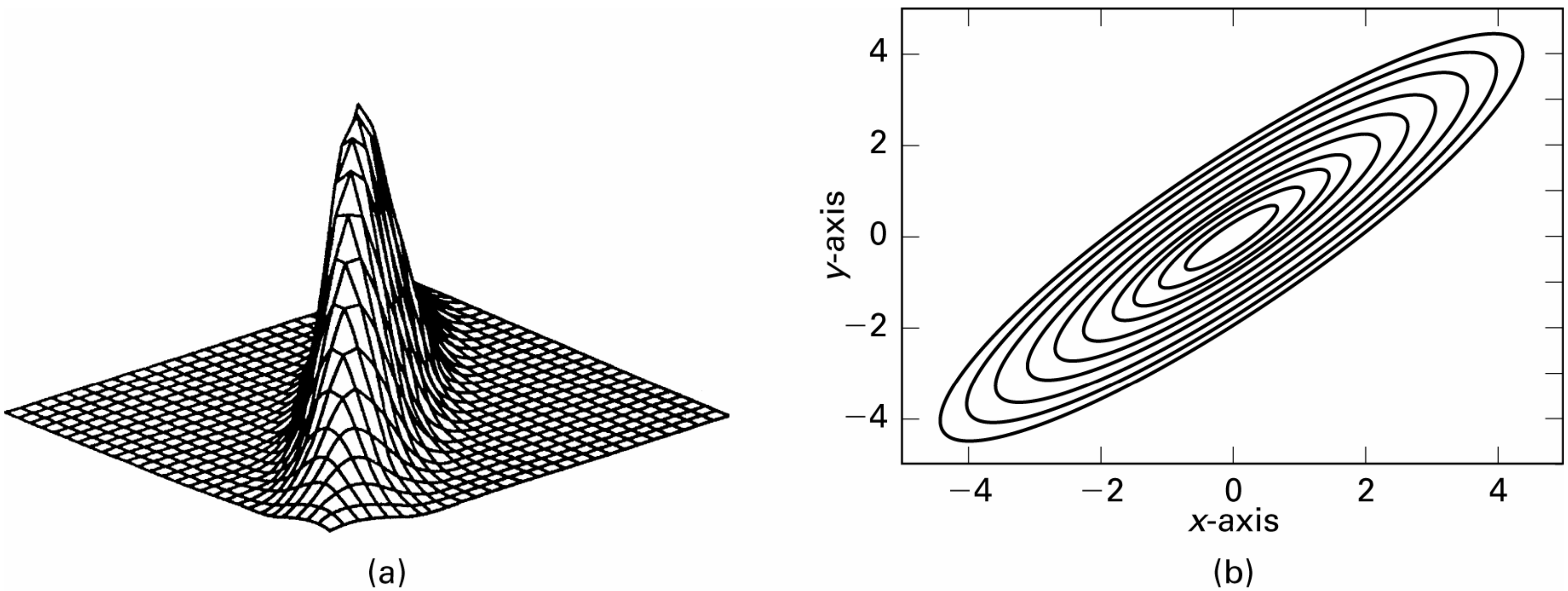
- Linear transformations of the standard gaussian vector: $\mathbf{x} = \mathbf{A}\mathbf{w} + \boldsymbol{\mu}.$

$$\mathbf{c}^t\mathbf{x} \sim \mathcal{N}\left(\mathbf{c}^t\boldsymbol{\mu}, \mathbf{c}^t\mathbf{A}\mathbf{A}^t\mathbf{c}\right);$$

Gaussian Random Vectors (uncorrelated vs correlated)



Graph of the joint Gaussian density



(a) Gaussian pdf with $\bar{X} = \bar{Y} = 0$, $\sigma_X = \sigma_Y = 2$, and $\rho = 0.9$; (b) Contours of constant density.

Complex Gaussian R.V: Circular Symmetry

- A complex Gaussian random variable X whose real and imaginary components are i.i.d. gaussian $\mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I$
- … satisfies a *circular symmetry* property:
  - $e^{j\phi}X$ has the same distribution as X for any ϕ.
  - $e^{j\phi}$ multiplication: rotation in the complex plane.
- We shall call such a random variable *circularly symmetric complex Gaussian*,
  - …denoted by $\mathcal{CN}(0, \sigma^2)$, where $\sigma^2 = E[|X|^2]$.

Complex Gaussian & Circular Symmetry (Contd)

For a circular symmetric complex random vector $\mathbf{x}$,

$$\mathbb{E}[\mathbf{x}] = \mathbb{E}\left[e^{j\theta}\mathbf{x}\right] = e^{j\theta}\mathbb{E}[\mathbf{x}]$$

for any θ; hence the mean $\boldsymbol{\mu} = 0$. Moreover

$$\mathbb{E}\left[\mathbf{x}\mathbf{x}^t\right] = \mathbb{E}\left[e^{j\theta}\mathbf{x}\left(e^{j\theta}\mathbf{x}\right)^t\right] = e^{j2\theta}\mathbb{E}\left[\mathbf{x}\mathbf{x}^t\right]$$

for any θ; hence the pseudo-covariance matrix $\mathbf{J}$ is also zero

Covariance matrix: $\mathbf{K}$ fully specifies the first and second order statistics

A collection of n i.i.d. $\mathcal{CN}(0,1)$ random variables forms a standard circular symmetric Gaussian random vector $\mathbf{w}$ and is denoted by $\mathcal{CN}(0,\mathbf{I})$. The density function of $\mathbf{w}$ can be explicitly written as, following from (A.7),

$$f(\mathbf{w}) = \frac{1}{\pi^n}\exp\left(-\|\mathbf{w}\|^2\right), \qquad \mathbf{w} \in \mathcal{C}^n. \tag{A.21}$$

$\mathbf{Uw}$ has the same distribution as $\mathbf{w}$, for any complex orthogonal matrix $\mathbf{U}$

Complex Gaussian: Summary (I)

Summary A.1 Complex Gaussian Random Vectors

- An n-dimensional complex Gaussian random vector $\mathbf{x}$ has real and imaginary components which form a $2n$-dimensional real Gaussian random vector.
- $\mathbf{x}$ is *circular symmetric* if for any θ,

$$e^{j\theta}\mathbf{x} \sim \mathbf{x}. \tag{A.24}$$

- A circular symmetric Gaussian $\mathbf{x}$ has zero mean and its statistics are fully specified by the covariance matrix $\mathbf{K} := \mathbb{E}[\mathbf{x}\mathbf{x}^*]$. It is denoted by $\mathcal{CN}(0, \mathbf{K})$.
- The scalar complex random variable $w \sim \mathcal{CN}(0, 1)$ has i.i.d. real and imaginary components each distributed as $\mathcal{N}(0, 1/2)$. The phase of w is uniformly distributed in $[0, 2\pi]$ and independent of its magnitude $|w|$, which is Rayleigh distributed:

$$f(r) = r \exp\left(-\frac{r^2}{2}\right), \qquad r \geq 0. \tag{A.25}$$

$|w|^2$ is exponentially distributed.

Complex Gaussian Vectors: Summary

- If the random vector $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I})$, then its real and imaginary components are all i.i.d, and $\mathbf{w}$ is *isotropic*, i.e., for any unitary matrix $\mathbf{U}$,

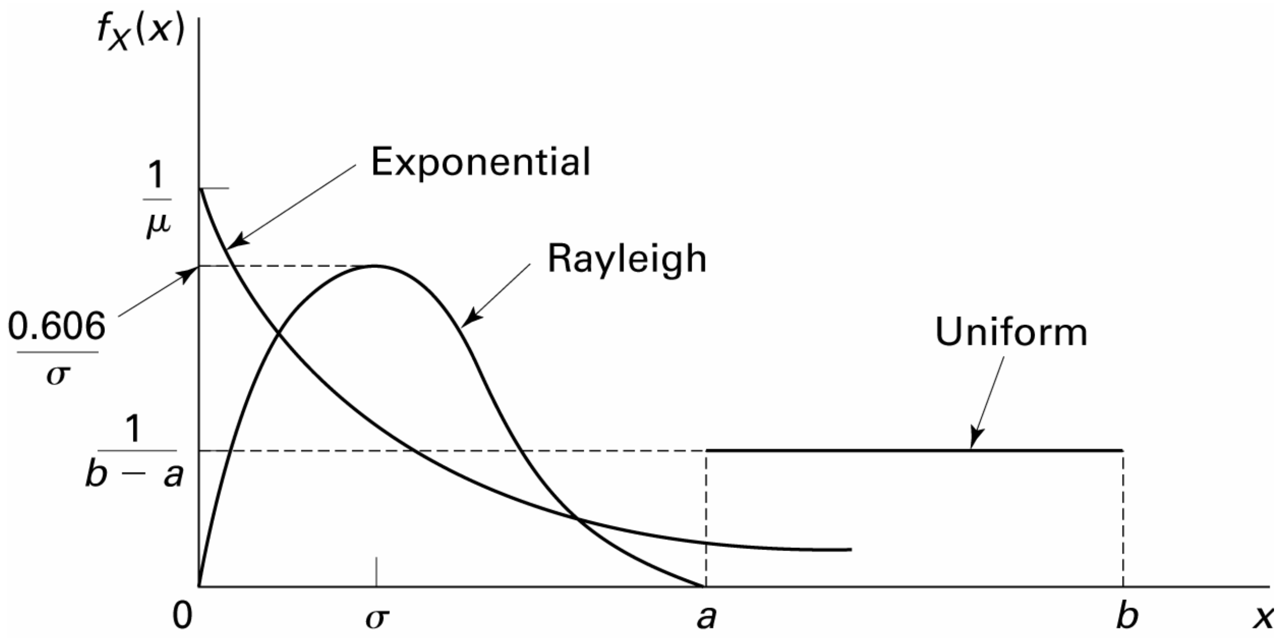
$$\mathbf{U}\mathbf{w} \sim \mathbf{w}. \tag{A.26}$$

  Equivalently, the projections of $\mathbf{w}$ onto orthogonal directions are i.i.d. $\mathcal{CN}(0, 1)$. The squared magnitude $||\mathbf{w}||^2$ is distributed as χ^2_{2n} with mean n.

- If $\mathbf{x} \sim \mathcal{CN}(0, \mathbf{K})$ and $\mathbf{K}$ is invertible, then the density of $\mathbf{x}$ is

$$f(\mathbf{x}) = \frac{1}{\pi^n \sqrt{\det \mathbf{K}}} \exp\left(-\mathbf{x}^* \mathbf{K}^{-1} \mathbf{x}\right), \qquad \mathbf{x} \in \mathcal{C}^n. \tag{A.27}$$

Related Distributions



The rayleigh, exponential, and uniform pdf 's.

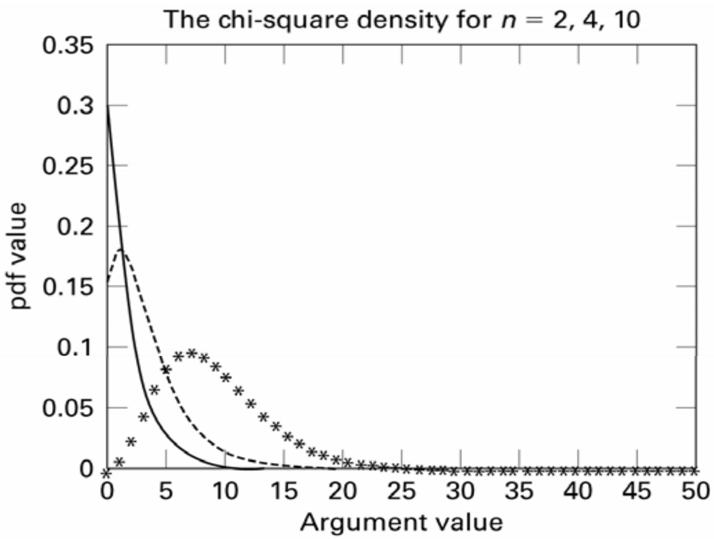
$X = [X_1, \ldots, X_n]$ is **Normal**

$||X||$ is **Rayleigh** { eg: *magnitude* of a complex gaussian channel $X_1 + jX_2$ }

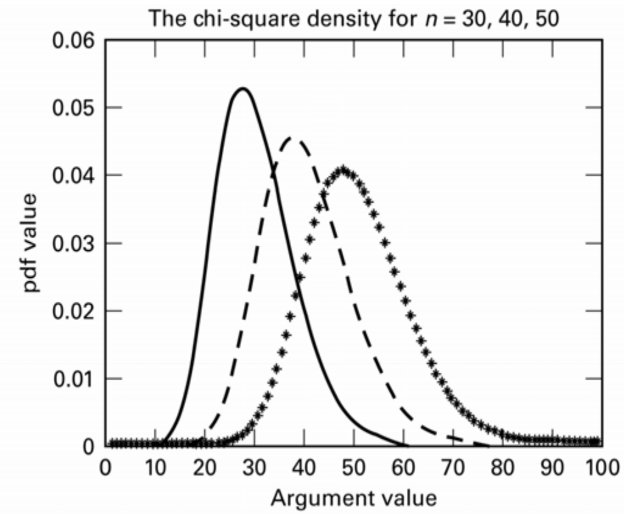
$||X||^2$ is **Chi-Squared w/ *n*-degrees of freedom**

When n = 2, chi-squared becomes **exponential**. {eg: *power* in complex gaussian channel: sum of squares…}

Chi-Squared Distribution



The Chi-square probability density function for $n = 2$ (solid), $n = 4$ (dashed), and $n = 10$ (stars). Note that for larger values of n, the shape approaches that of a Normal pdf positive mean-parameter μ.



The Chi-square pdf for three large values for the parameter n: $n = 30$ (solid); $n = 40$ (dashed); $n = 50$ (stars). For large values of n, the Chi-square pdf can be approximated by a normal $N(n, 2n)$ for computing probabilities not too far from the mean. For example, for $n = 30$, $P[\mu - \sigma < X < \mu + \sigma] = 0.6827$ assuming $X : N(30, 60)$. The value computed, using single-precision arithmetic, using the Chi-square pdf, yields 0.6892.

Sum of squares of *n* normal variables: *Chi-squared*

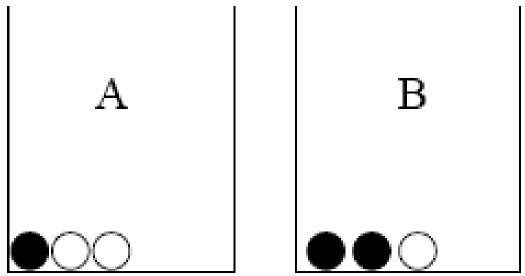
For *n =2*, it becomes an *exponential* distribution.

Becomes *bell-shaped for larger n*

Maximum Likelihood (ML) Detection: Concepts

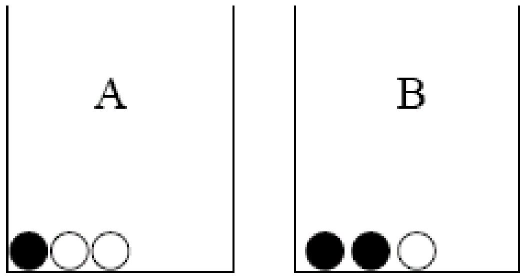
<u>Reference:</u>
Mackay, Information Theory,
http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html
(chap 3, online book)

Likelihood Principle



- Experiment:
  - Pick Urn A or Urn B at random
  - Select a ball from that Urn.
- The ball is black.
- What is the probability that the selected Urn is A?

Likelihood Principle (Contd)



- Write out what you know!
- **P(Black | UrnA) = 1/3**
- **P(Black | UrnB) = 2/3**
- **P(Urn A) = P(Urn B) = 1/2**
- We want **P(Urn A | Black).**
- Gut feeling: Urn B is more likely than Urn A (given that the ball is black). But by how much?
- This is an inverse probability problem.
  - Make sure you understand the inverse nature of the conditional probabilities!
- Solution technique: Use Bayes Theorem.

Likelihood Principle (Contd)

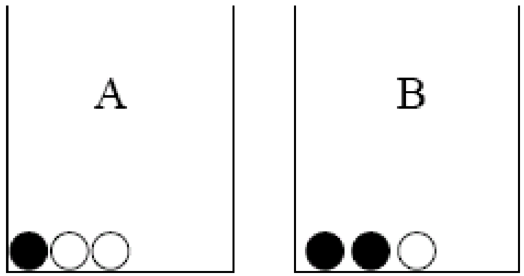
- **<u>Bayes manipulations:</u>**
- **P(Urn A | Black) =**
  - **P(Urn A and Black) /P(Black)**
- Decompose the numerator and denomenator in terms of the probabilities we know.

- **P(Urn A and Black) = P(Black | UrnA)*P(Urn A)**
- **P(Black) = P(Black| Urn A)*P(Urn A) + P(Black| UrnB)*P(UrnB)**

- We know all these values (see prev page)! Plug in and crank.
- **P(Urn A and Black) = 1/3 * 1/2**
- **P(Black) = 1/3 * 1/2 + 2/3 * 1/2 = 1/2**
- **P(Urn A and Black) /P(Black) = 1/3 = 0.333**
- Notice that it matches our gut feeling that Urn A is less likely, once we have seen black.

- *<u>The information that the ball is black has CHANGED !</u>*
  - From P(Urn A) = 0.5 to P(Urn A | Black) = 0.333

Likelihood Principle



- Way of thinking…
- Hypotheses: Urn A or Urn B ?
- Observation: "Black"
- Prior probabilities: P(Urn A) and P(Urn B)
- Likelihood of Black given choice of Urn: {aka *forward probability*}
  - P(Black | Urn A) and P(Black | Urn B)
- Posterior Probability: of each hypothesis given evidence
  - P(Urn A | Black) {aka *inverse probability*}
- Likelihood Principle (informal): All inferences depend ONLY on
  - The likelihoods P(Black | Urn A) and P(Black | Urn B), and
  - The priors P(Urn A) and P(Urn B)
- Result is a probability (or distribution) model over the space of possible hypotheses.

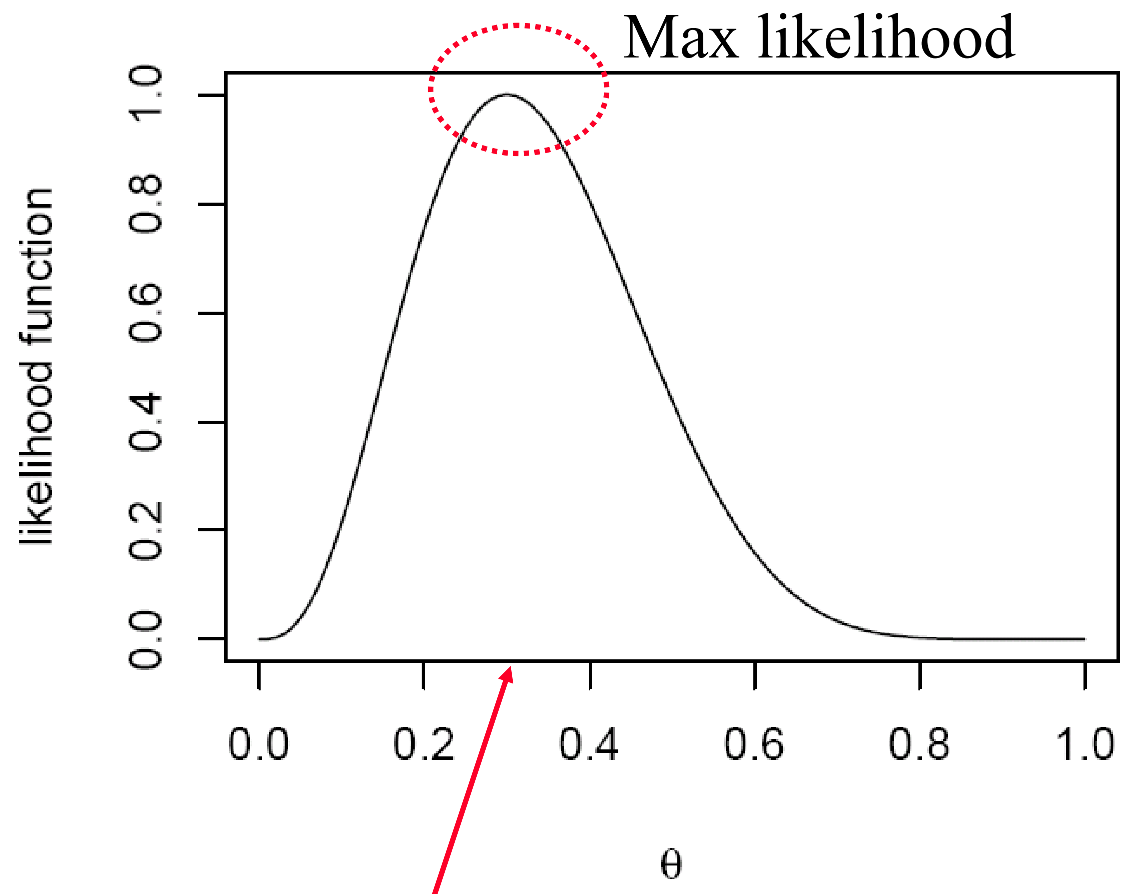
Maximum Likelihood (intuition)

- <u>Recall:</u>
- **P(Urn A | Black) = P(Urn A and Black) /P(Black) =**

  P(Black | UrnA)*P(Urn A) / P(Black)

- **P(Urn? | Black)** is maximized when **P(Black | Urn?)** is maximized.
  - Maximization over the hypotheses space (Urn A or Urn B)
- **P(Black | Urn?)** = “likelihood”
- => “*<u>Maximum Likelihood</u>*” approach to maximizing posterior probability

Maximum Likelihood: intuition



<u>This hypothesis</u> has the highest (maximum) likelihood of explaining the data observed

Maximum Likelihood (ML): mechanics

- **Independent Observations** (like Black): $\mathbf{X_1, \ldots, X_n}$
- **Hypothesis** θ
- **Likelihood Function**: $L(\theta) = P(X_1, \ldots, X_n \mid \theta) = \Pi_i P(X_i \mid \theta)$
    - {Independence => multiply individual likelihoods}
- **Log Likelihood** $\mathbf{LL(\theta) = \Sigma_i \log P(X_i \mid \theta)}$

Back to Urn example

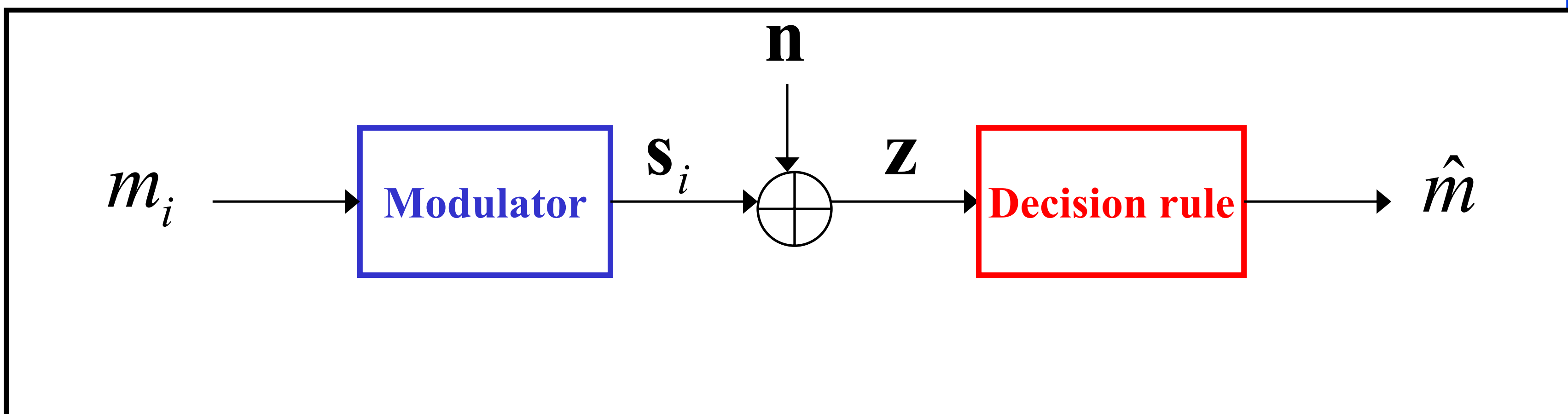
- In our urn example, we are asking:
  - Given the *observed* *data* "ball is black"…
  - …*which hypothesis* (Urn A or Urn B) has the *highest likelihood* of explaining this observed data?
  - Ans from above analysis: Urn B

- Note: this does not give the *posterior probability* P(Urn A | Black), but quickly helps us *choose the best hypothesis (Urn B)* that would explain the data…

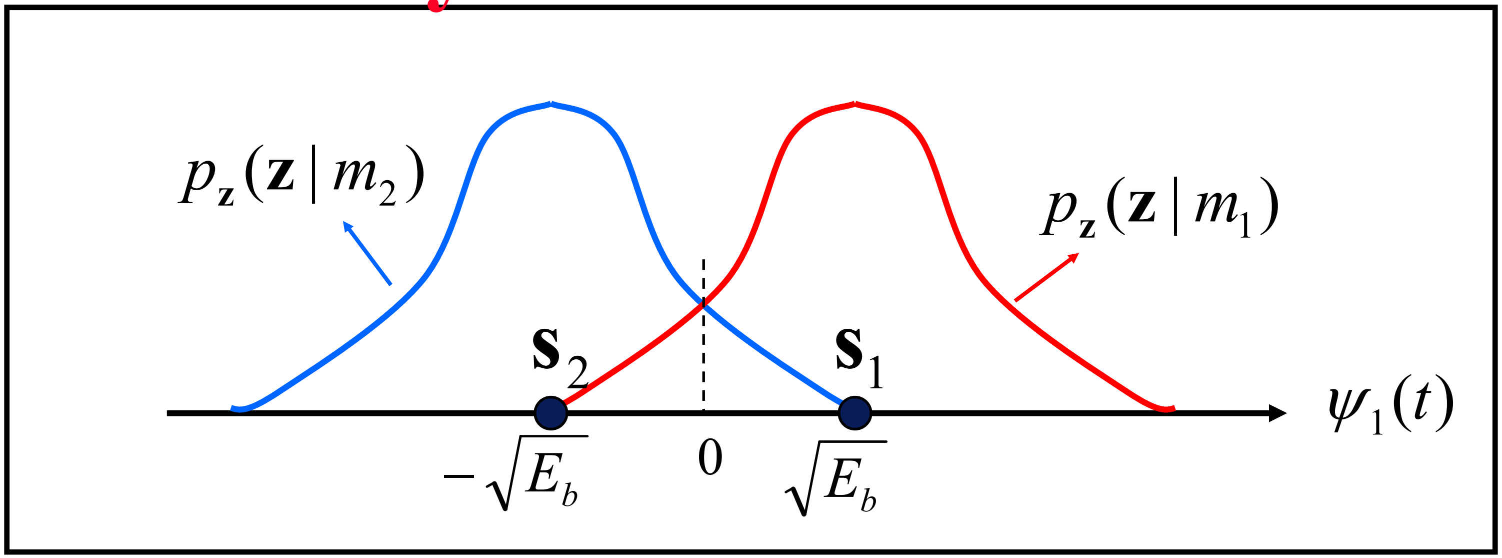
More examples: (biased coin etc)
http://en.wikipedia.org/wiki/Maximum_likelihood
http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html
(chap 3)

Not Just Urns and Balls:
Detection of signal in AWGN

- Detection problem:
  - Given the observation vector $\mathbf{z}$, perform a mapping from $\mathbf{z}$ to an estimate $\hat{m}$ of the transmitted symbol, m_i, such that the average probability of error in the decision is minimized.



Binary PAM + AWGN Noise



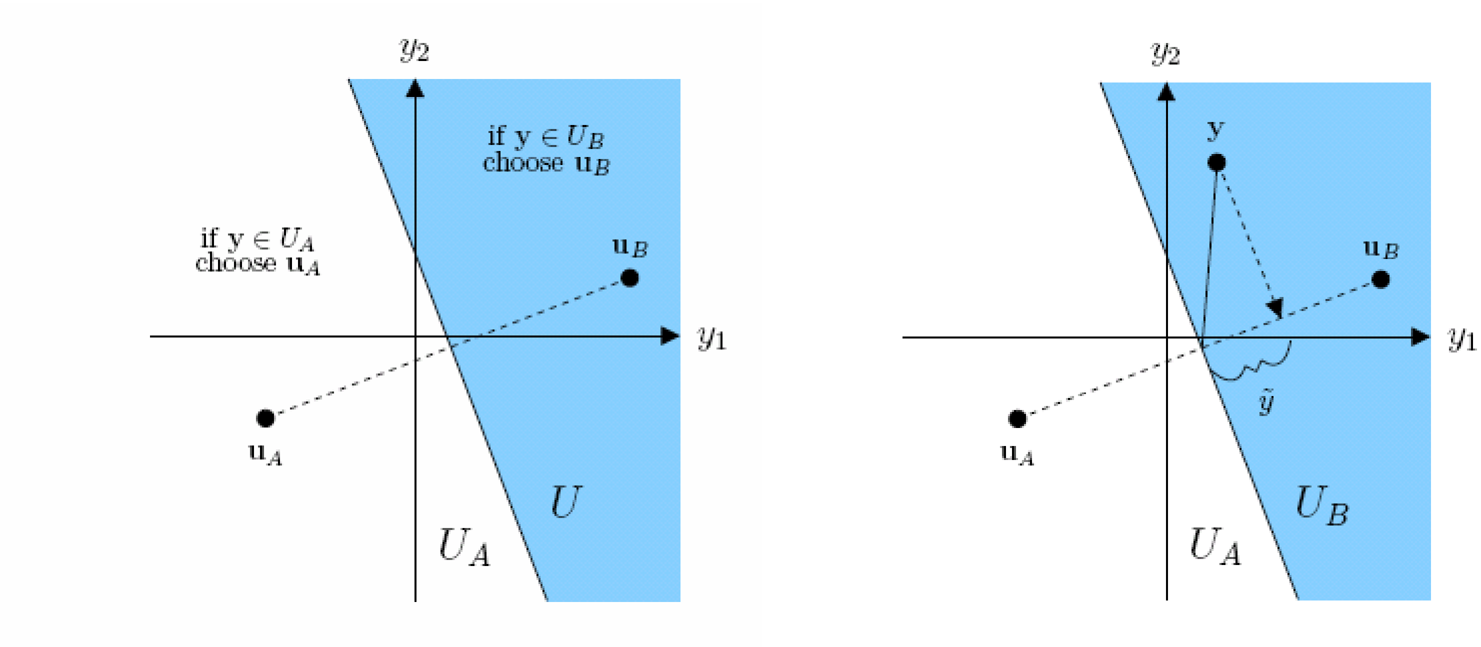
Signal s1 or s2 is sent. **z** is received

Additive white gaussian noise (AWGN) => the likelihoods are

$p_{\mathbf{z}}(\mathbf{z}|m_1)$ $p_{\mathbf{z}}(\mathbf{z}|m_2)$ bell-shaped pdfs around s1 and s2

MLE => at any point on the x-axis, see which curve (blue or red) has a higher (maximum) value and select the corresponding signal (s1 or s2) : simplifies into a "*nearest-neighbor*" rule

AWGN Nearest Neighbor Detection



- Projection onto the signal directions (subspace) is called *matched filtering* to get the "*sufficient statistic*"
- Error probability is the tail of the normal distribution (Q-function), based upon the mid-point between the two signals

$$\boxed{Q\left(\frac{\|\mathbf{u}_A - \mathbf{u}_B\|}{2\sqrt{N_0/2}}\right),}$$

Detection in AWGN: Summary

Summary A.2 | Vector Detection in Complex Gaussian Noise

Binary Signals:

The transmit vector $\mathbf{u}$ is either $\mathbf{u}_A$ or $\mathbf{u}_B$ and we wish to detect $\mathbf{u}$ from received vector

$$\mathbf{y} = \mathbf{u} + \mathbf{w}, \tag{A.52}$$

where $\mathbf{w} \sim \mathcal{CN}(0, N_0\mathbf{I})$. The ML detector picks the transmit vector closest to $\mathbf{y}$ and the error probability is:

$$Q\left(\frac{\|\mathbf{u}_A - \mathbf{u}_B\|}{2\sqrt{N_0/2}}\right). \tag{A.53}$$

Vector detection (contd)

Collinear Signals:

The transmit symbol x is equally likely to take one of a finite set of values in $\mathcal{C}$ (the *constellation* points) and the received vector is

$$\mathbf{y} = \mathbf{h}x + \mathbf{w}, \tag{A.54}$$

where $\mathbf{h}$ is a fixed vector.

Projecting $\mathbf{y}$ onto the unit vector $\mathbf{v} := \mathbf{h}/\|\mathbf{h}\|$ yields a scalar sufficient statistic:

$$\mathbf{v}^*\mathbf{y} = \|\mathbf{h}\|x + w. \tag{A.55}$$

Here $w \sim \mathcal{CN}(0, N_0)$.

If further the constellation is real-valued, then

$$\Re[\mathbf{v}^*\mathbf{y}] = \|\mathbf{h}\|x + \Re[w] \tag{A.56}$$

is sufficient. Here $\Re[w] \sim \mathcal{N}(0, N_0/2)$.

With antipodal signalling, $x = \pm a$, the ML error probability is simply

$$Q\left(\frac{a\|\mathbf{h}\|}{\sqrt{N_0/2}}\right). \tag{A.57}$$

Via a translation, the binary signal detection problem in the first part of the summary can be reduced to this antipodal signalling scenario.

Estimation

References:

- Appendix A.3 (Tse/Viswanath)
- Stark & Woods, Probability and Random Processes with Applications to Signal Processing, Prentice Hall, 2001
- Schaum's Outline of Probability, Random Variables, and Random Processes
- Popoulis, Pillai, Probability, Random Variables and Stochastic Processes, McGraw-Hill, 2002.

Detection vs Estimation

- In *detection* we have to ***decide*** which symbol was transmitted s_A or s_B
  - This is a binary (0/1, or yes/no) type answer, with an associated error probability
- In *estimation*, we have to output an *estimate* h' of a transmitted signal h.
  - This estimate is a complex number, not a binary answer.
  - Typically, we try to estimate the complex channel h, so that we can use it in coherent combining (matched filtering)

Estimation in AWGN: MMSE

$$\boxed{y = x + w.}$$

Need: estimate $\hat{x}$ of x

- Performance criterion: *mean-squared error (MSE)*

$$\boxed{\mathrm{MSE} := \mathbb{E}\left[(x - \hat{x})^2\right]}$$

- Optimal estimator is the "*conditional mean*" of x given the observation y
  - Gives *Minimum Mean-Square Error* (*MMSE*)

$$\boxed{\hat{x} = \mathbb{E}[x|y]}$$

- Satisfies *orthogonality* property:
  - Error independent of observation:

$$\boxed{\mathbb{E}[(\hat{x} - x)y] = 0}$$

- But, the conditional mean is a non-linear operator
  - It becomes *linear* if x is also *gaussian*.
  - Else, we need to find the best linear approximation (*LMMSE*)!

LMMSE

- We are looking for a linear estimate: x = cy
  - The best linear estimator, i.e. weighting coefficient c is:

$$\boxed{c = \frac{\mathbb{E}\left[x^2\right]}{\mathbb{E}\left[x^2\right] + N_0/2}.}$$

  - We are *weighting* the received signal y by the transmit signal energy as a fraction of the received signal energy.

- The corresponding error (MMSE) is:

$$\boxed{\mathrm{MMSE} = \frac{\mathbb{E}\left[x^2\right] N_0/2}{\mathbb{E}\left[x^2\right] + N_0/2}.}$$

LMMSE: Generalization & Summary

Summary A.3 Mean Square Estimation in a Complex Vector Space

The linear estimate with the smallest mean squared error of x from

$$y = x + w, \tag{A.80}$$

with $w \sim \mathcal{CN}(0, N_0)$, is

$$\hat{x} = \frac{\mathbb{E}\left[|x|^2\right]}{\mathbb{E}\left[|x|^2\right] + N_0} y. \tag{A.81}$$

To estimate x from

$$\mathbf{y} = \mathbf{h}x + \mathbf{w}, \tag{A.82}$$

where $\mathbf{w} \sim \mathcal{CN}(0, N_0\mathbf{I})$,

$$\mathbf{h}^*\mathbf{y} \tag{A.83}$$

is a sufficient statistic, reducing the vector estimation problem to the scalar one.

The best *linear* estimator is

$$\hat{x} = \frac{\mathbb{E}\left[|x|^2\right]}{\mathbb{E}\left[|x|^2\right]\|\mathbf{h}\|^2 + N_0} \mathbf{h}^*\mathbf{y}. \tag{A.84}$$

The corresponding minimum mean squared error (MMSE) is:

$$\mathrm{MMSE} = \frac{\mathbb{E}\left[|x|^2\right] N_0}{\mathbb{E}\left[|x|^2\right]\|\mathbf{h}\|^2 + N_0}. \tag{A.85}$$

In the special case when $x \sim \mathcal{CN}(\mu, \sigma^2)$, this estimator yields the minimum mean squared error among *all* estimators, linear or non-linear.

Random Processes

References:

• Appendix B (Goldsmith)

• Stark & Woods, Probability and Random Processes with Applications to Signal Processing, Prentice Hall, 2001

• Schaum's Outline of Probability, Random Variables, and Random Processes

• Popoulis, Pillai, Probability, Random Variables and Stochastic Processes, McGraw-Hill, 2002.

Random Sequences and Random Processes

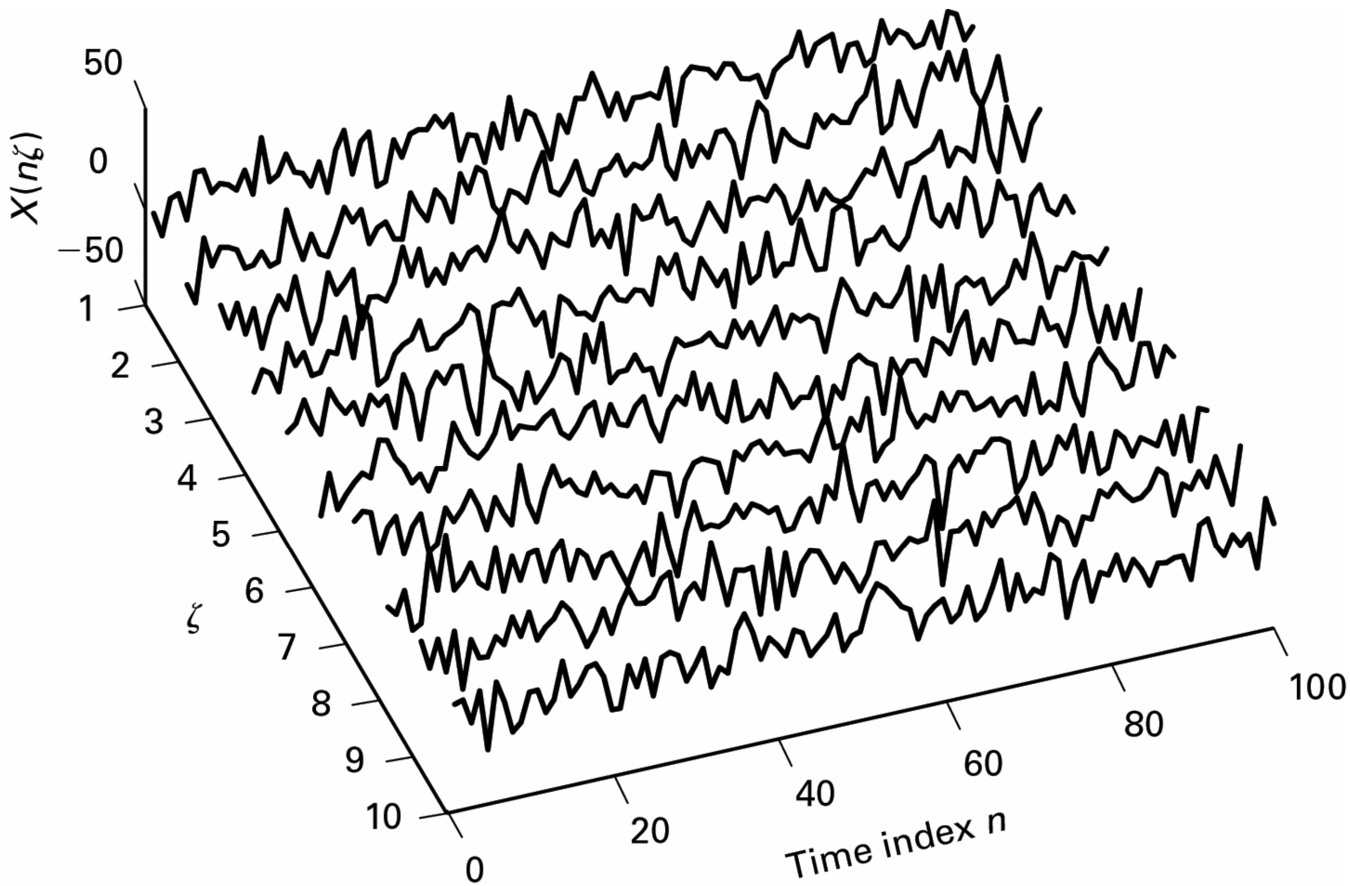
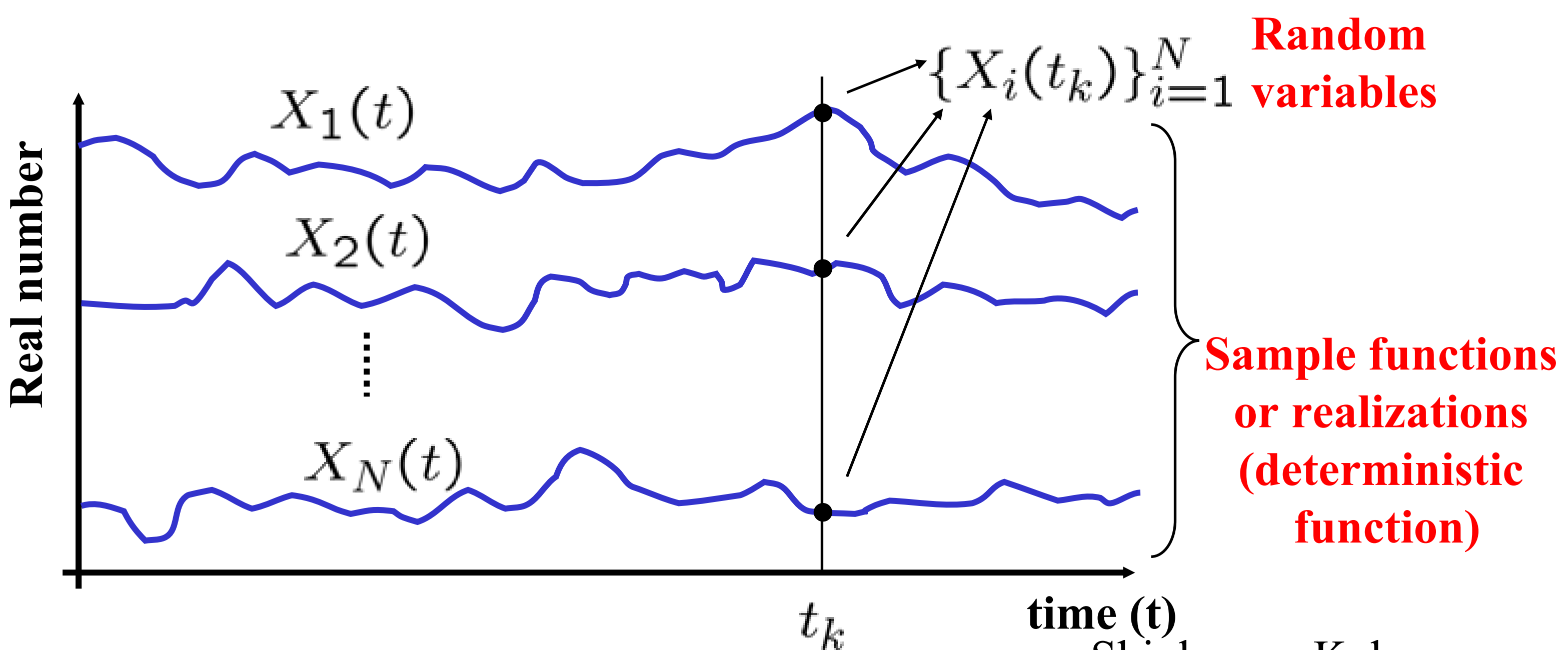


Illustration of the concept of *random sequence* $X(n, \zeta)$ where the ζ domain (i.e., the sample space Ω) consists of just 10 values. (Samples connected for plot.)

Random process

- A random process is a collection of time functions, or signals, corresponding to various outcomes of a random experiment. For each outcome, there exists a deterministic function, which is called a sample function or a realization.

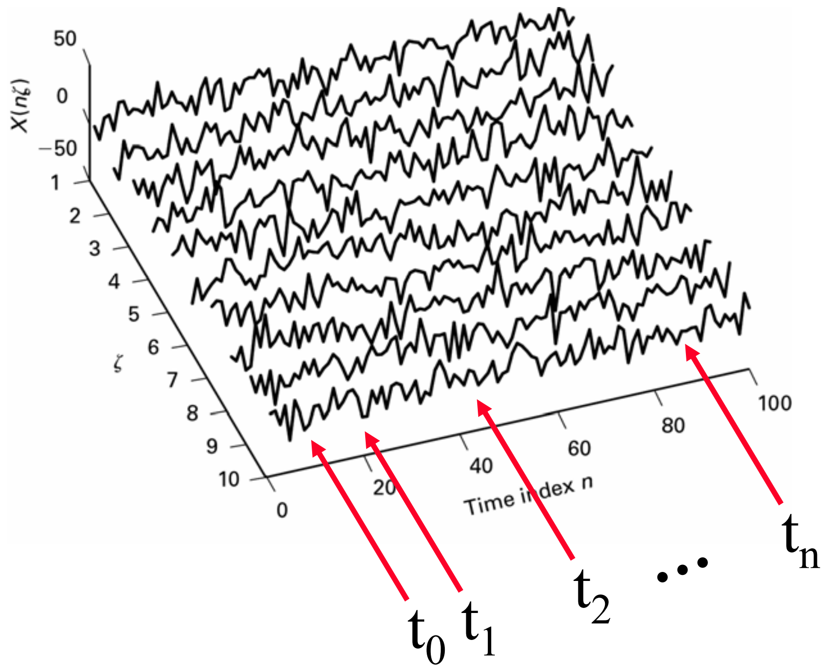


Specifying a Random Process

- A random process is defined by all its <u>joint CDFs</u>

$$p(X(t_0) \leq x_0, X(t_1) \leq x_1, \ldots, X(t_n) \leq x_n).$$

for *<u>all possible sets of sample times</u>* $\{t_0, t_1, \ldots, t_n\}$



Stationarity

- If time-shifts (any value T) do not affect its joint CDF

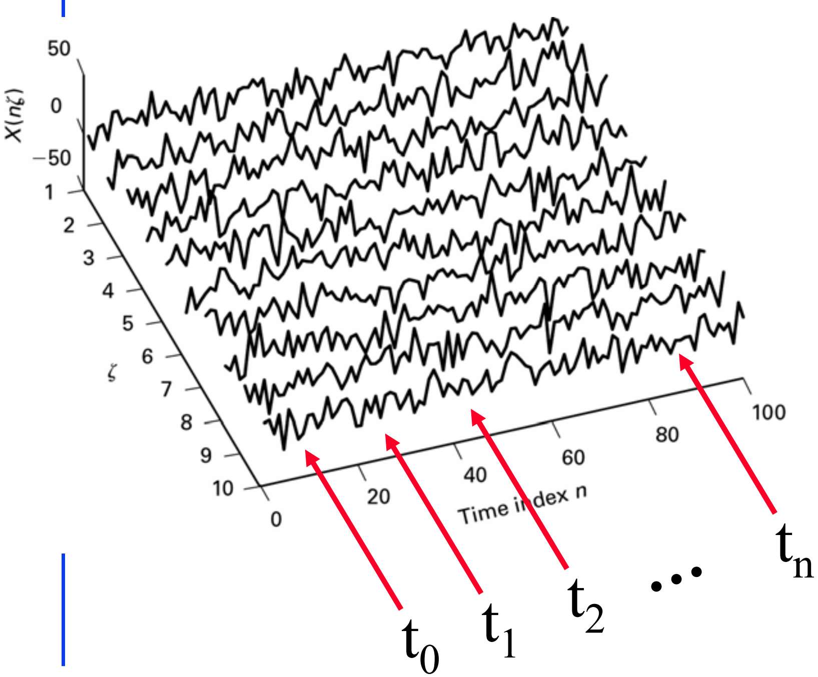
$$p(X(t_0) \le x_0, X(t_1) \le x_1, \ldots, X(t_n) \le x_n) =$$

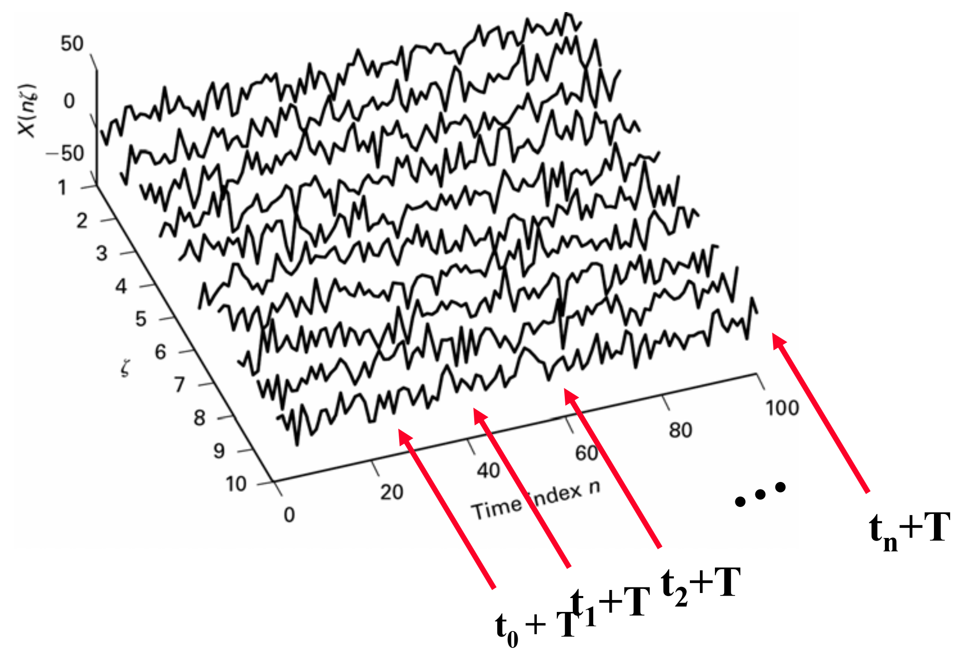
$$p(X(t_0 + T) \le x_0, X(t_1 + T) \le x_1, \ldots, X(t_n + T) \le x_n).$$

$$A_X(t, t+\tau) =$$

$$\boxed{\mathbf{E}[X(t)] = \mathbf{E}[X(t-t)] = \mathbf{E}[X(0)] = \mu_X.}$$

$$\boxed{\mathbf{E}[X(t-t)X(t+\tau-t)] = \mathbf{E}[X(0)X(\tau)] \triangleq A_X(\tau)}$$





Weak Sense Stationarity (wss)

$A_X(t, t+\tau) =$

$\boxed{\mathbf{E}[X(t)] = \mathbf{E}[X(t-t)] = \mathbf{E}[X(0)] = \mu_X.}$

$\boxed{\mathbf{E}[X(t-t)X(t+\tau-t)] = \mathbf{E}[X(0)X(\tau)] \triangleq A_X(\tau)}$

- Keep only above two properties (2nd order stationarity)…
  - Don't insist that higher-order moments or higher order joint CDFs be unaffected by lag T
- With LTI systems, we will see that WSS inputs lead to WSS outputs,
  - In particular, if a WSS process with PSD $S_X(f)$ is passed through a linear time-invariant filter with frequency response $H(f)$, then the filter output is also a WSS process with power spectral density $|H(f)|^2 S_X(f)$.
- Gaussian *w.s.s.* = Gaussian stationary process (since it only has 2nd order moments)

Stationarity: Summary

- **Strictly stationary:** If none of the statistics of the random process are affected by a shift in the time origin.
- **Wide sense stationary (WSS):** If the mean and autocorrelation function do not change with a shift in the origin time.
- **Cyclostationary:** If the mean and autocorrelation function are periodic in time.

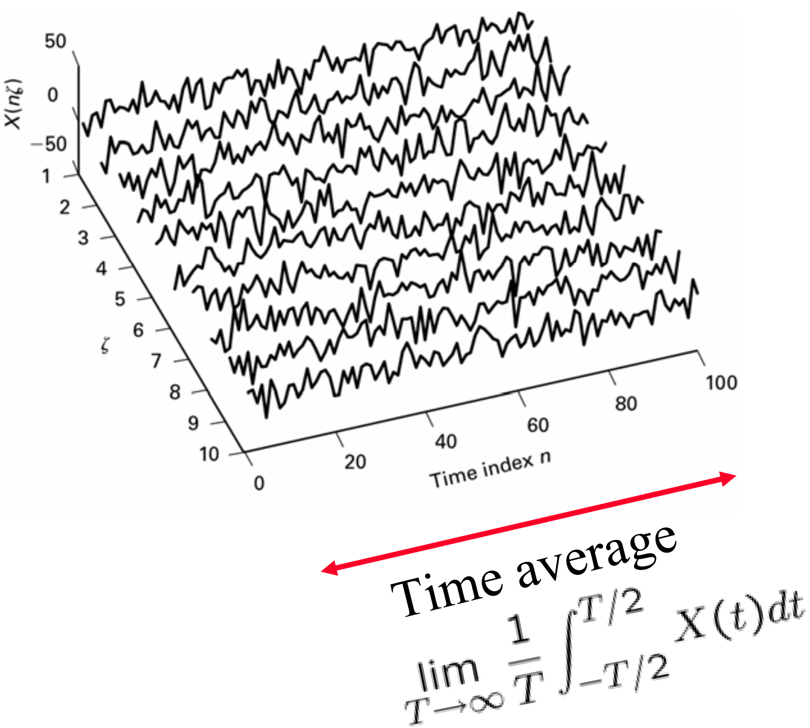
Ergodicity

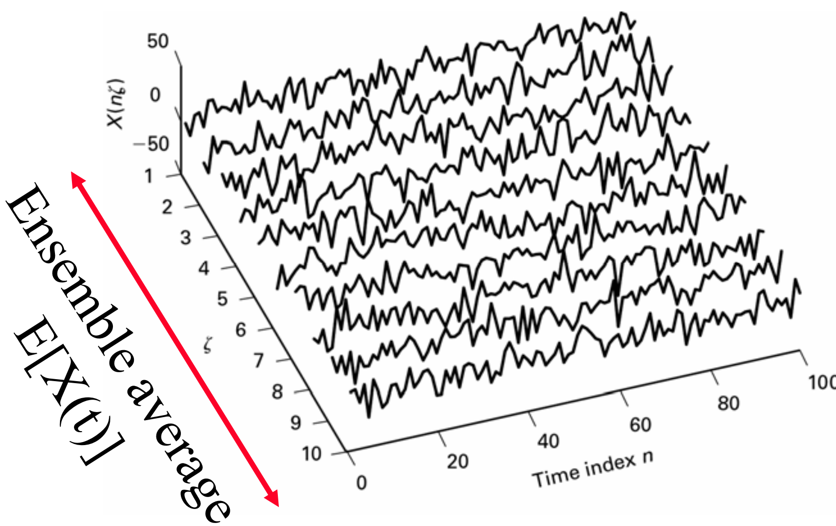
- Time averages = Ensemble averages

[i.e. "***ensemble***" averages like mean/autocorrelation can be computed as "***time-averages***" over a single realization of the random process]

- A random process: ergodic in *mean* and *autocorrelation* (like w.s.s.) if

$$m_X = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} X(t)dt \quad \text{and} \quad R_X(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} X(t)X^*(t-\tau)dt$$





Autocorrelation: Summary

- Autocorrelation of an energy signal

$$R_x(\tau) = x(\tau) \star x^*(-\tau) = \int_{-\infty}^{\infty} x(t)x^*(t-\tau)dt$$

- Autocorrelation of a power signal

$$R_x(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x^*(t-\tau)dt$$

  - For a periodic signal:

$$R_x(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t)x^*(t-\tau)dt$$

- Autocorrelation of a random signal

$$R_X(t_i, t_j) = \mathsf{E}[X(t_i)X^*(t_j)]$$

  - For a WSS process:

$$R_X(\tau) = \mathsf{E}[X(t)X^*(t-\tau)]$$

Power Spectral Density (PSD)

The power spectral density (PSD) of a WSS process is defined as the Fourier transform of its autocorrelation function with respect to τ:

$$S_X(f) = \int_{-\infty}^{\infty} A_X(\tau) e^{-j2\pi f\tau} d\tau. \tag{B.26}$$

The autocorrelation can be obtained from the PSD through the inverse transform:

$$A_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau} df. \tag{B.27}$$

The PSD takes its name from the fact that the expected power of a random process $X(t)$ is the integral of its PSD:

$$\mathbf{E}[X^2(t)] = A_X(0) = \int_{-\infty}^{\infty} S_X(f) df, \tag{B.28}$$

1. $S_X(f)$ is real and $S_X(f) \geq 0$
2. $S_X(-f) = S_X(f)$
3. $A_X(0) = \int S_X(\omega)\, d\omega$

Power Spectrum

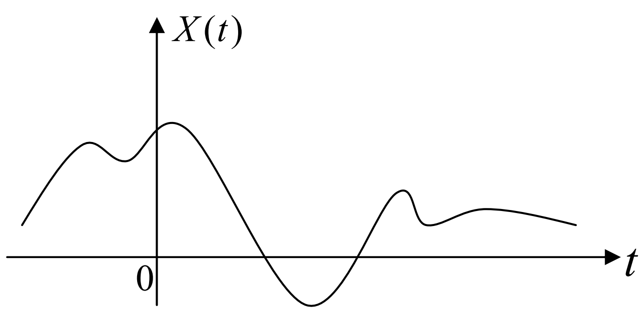
For a deterministic signal x(t), the spectrum is well defined: If $X(\omega)$ represents its Fourier transform, i.e., if

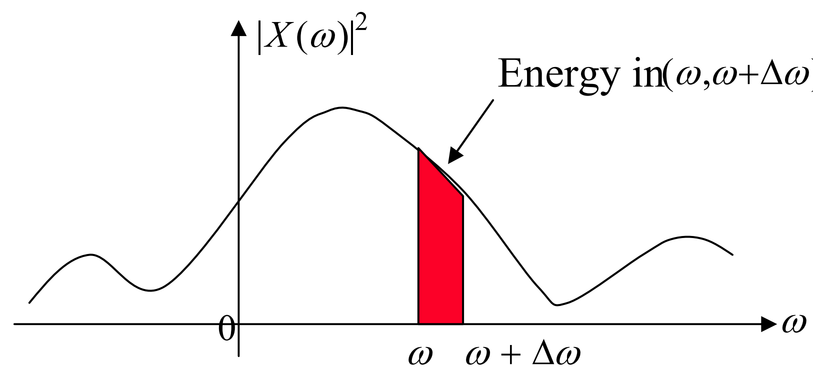
$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt,$$

then $|X(\omega)|^2$ represents its energy spectrum. This follows from Parseval's theorem since the signal energy is given by

$$\int_{-\infty}^{+\infty} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |X(\omega)|^2 d\omega = E.$$

Thus $|X(\omega)|^2 \Delta\omega$ represents the signal energy in the band $(\omega, \omega + \Delta\omega)$





Spectral density: Summary

- **Energy signals**:

$$E_x = \int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df \qquad X(f) = \mathcal{F}[x(t)]$$

  - Energy spectral density (ESD): $\boxed{\Psi_x(f) = |X(f)|^2}$

- **Power signals:**

$$P_x = \frac{1}{T_0} \int_{T_0/2}^{T_0/2} |x(t)|^2 dt = \sum_{n=-\infty}^{\infty} |c_n|^2 \qquad \{c_n\} = \mathcal{F}[x(t)]$$

  - ***Power spectral density (PSD):***

$$\boxed{G_x(f) = \sum_{n=-\infty}^{\infty} |c_n|^2 \delta(f - nf_0) \qquad f_0 = 1/T_0}$$

- **Random process**:
  - Power spectral density (PSD):

$$\boxed{G_X(f) = \mathcal{F}[R_X(\tau)]}$$

Note: we have used *f* for *ω* and *Gx* for *Sx*

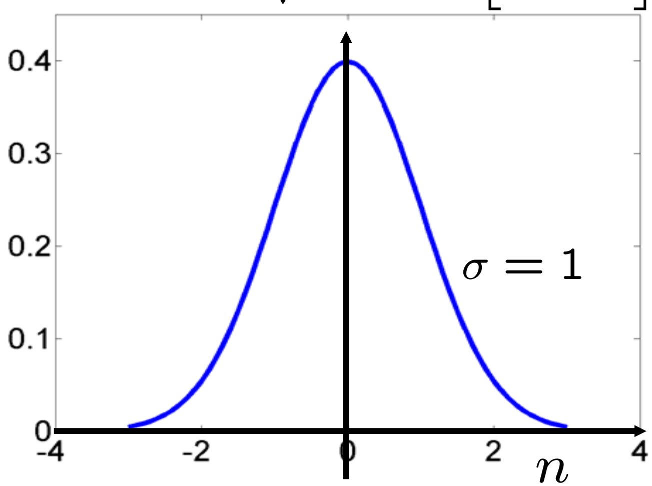
Properties of an autocorrelation function

- For real-valued (and WSS for random signals):
  1. Autocorrelation and spectral density form a Fourier transform pair. $R_X(\tau) \leftrightarrow S_X(\omega)$
  2. Autocorrelation is symmetric around zero. $R_X(-\tau) = R_X(\tau)$
  3. Its maximum value occurs at the origin. $|R_X(\tau)| \leq R_X(0)$
  4. Its value at the origin is equal to the average power or energy. $\mathbf{E}[X^2(t)] = A_X(0) = \int_{-\infty}^{\infty} S_X(f)df,$

Noise in communication systems

- Thermal noise is described by a zero-mean Gaussian random process, n(t).
- Its PSD is flat, hence, it is called white noise. IID gaussian.

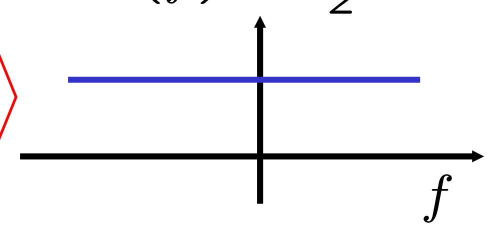
$$p(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{n^2}{2\sigma^2}\right]$$



Probability density function

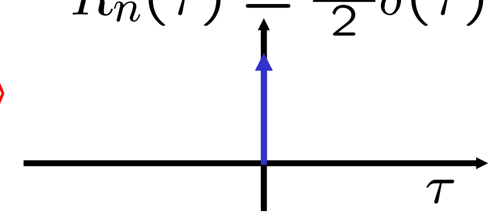
Power spectral density

$$G_n(f) = \frac{N_0}{2} \text{ [w/Hz]}$$



Autocorrelation function

$$R_n(\tau) = \frac{N_0}{2}\delta(\tau)$$



White Gaussian Noise

- White:
  - *Power spectral density (PSD)* is the same, i.e. *flat*, for all frequencies of interest (from dc to 10^{12} Hz)
  - Autocorrelation is a *delta* function => two samples no matter however close are uncorrelated.
    - $N_0/2$ to indicate two-sided PSD
    - Zero-mean gaussian completely characterized by its variance (σ^2)
    - Variance of filtered noise is finite = $N_0/2$
  - Similar to "white light" contains equal amounts of all frequencies in the visible band of EM spectrum
- Gaussian + uncorrelated => i.i.d.
  - Affects each symbol independently: memoryless channel
- Practically: if b/w of noise is much larger than that of the system: good enough
- Colored noise: exhibits correlations at positive lags

Signal transmission w/ linear systems (filters)

$$\text{Input}\quad \begin{matrix} x(t) \\ X(f) \end{matrix} \longrightarrow \boxed{\begin{matrix} h(t) \\ H(f) \end{matrix}} \longrightarrow \begin{matrix} y(t) \\ Y(f) \end{matrix} \quad \text{Output}$$

Linear system

- Deterministic signals: $\boxed{Y(f) = X(f)H(f)}$
- Random signals: $\boxed{G_Y(f) = G_X(f)|H(f)|^2}$

<u>Ideal distortion less transmission:</u>

• All the frequency components of the signal not only arrive with an <u>*identical time delay*</u>, but also <u>*amplified or attenuated equally*</u>.

$$\boxed{y(t) = Kx(t - t_0) \text{ or } H(f) = Ke^{-j2\pi f t_0}}$$

(Deterministic) Systems with Stochastic Inputs

A deterministic system[1] transforms each input waveform $X(t,\xi_i)$ into an output waveform $Y(t,\xi_i) = T[X(t,\xi_i)]$ by operating <u>only on the time variable *t*.</u> Thus a set of realizations at the input corresponding to a process *X*(*t*) generates a new set of realizations $\{Y(t,\xi)\}$ at the output associated with a new process *Y*(*t*).

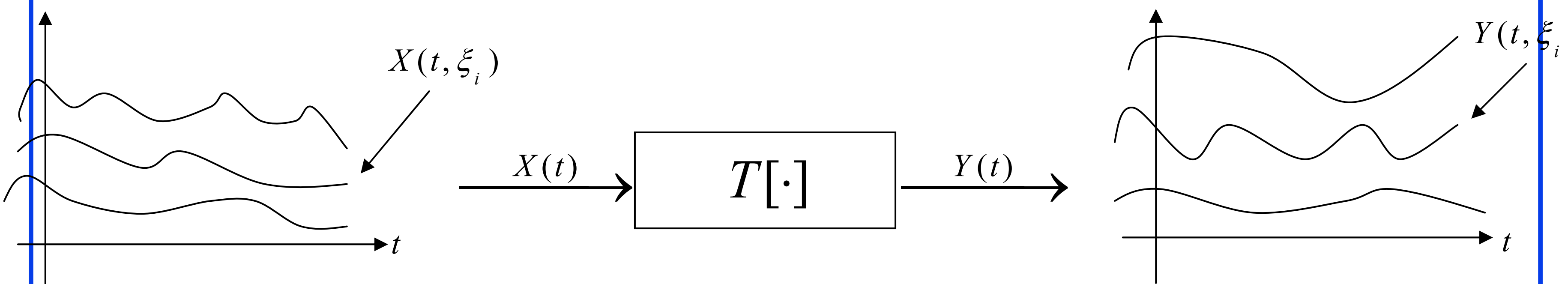


Fig. 14.3

Our goal is to study the <u>output process statistics</u> <u>in terms of the input process statistics and the system function</u>.

---

[1]A stochastic system on the other hand operates on both the variables t and ξ.

Deterministic Systems

Memoryless Systems

$$Y(t) = g[X(t)]$$

Systems with Memory

Time-varying systems

Time-Invariant systems

Linear systems

$$Y(t) = L[X(t)]$$

Linear-Time Invariant (LTI) systems

$$X(t) \longrightarrow \boxed{h(t)} \longrightarrow Y(t) = \int_{-\infty}^{+\infty} h(t-\tau) X(\tau) d\tau$$

$$= \int_{-\infty}^{+\infty} h(\tau) X(t-\tau) d\tau.$$

LTI system

LTI Systems: WSS input good enough

$X(t)$ wide-sense stationary process → LTI system $h(t)$ → $Y(t)$ wide-sense stationary process.

(a)

$X(t)$ strict-sense stationary process → LTI system $h(t)$ → $Y(t)$ strict-sense stationary process

(b)

$X(t)$ Gaussian process (also stationary) → Linear system → $Y(t)$ Gaussian process (also stationary)

(c)

White Noise Process & LTI Systems

$W(t)$ is said to be a white noise process if

$$R_{WW}(t_1, t_2) = q(t_1)\delta(t_1 - t_2),$$

i.e., $E[W(t_1)\, W^*(t_2)] = 0$ unless $t_1 = t_2$.

$W(t)$ is said to be *wide-sense stationary (w.s.s) white noise* if **$E[W(t)]$ = constant**, and

$$R_{WW}(t_1, t_2) = q\delta(t_1 - t_2) = q\delta(\tau).$$

If $W(t)$ is also a Gaussian process (white Gaussian process), then all of its samples are independent random variables

White noise $W(t)$ $\longrightarrow$ LTI $h(t)$ $\longrightarrow$ Colored noise $N(t) = h(t) * W(t)$

Summary

- Probability, union bound, bayes rule, maximum likelihood
- Expectation, variance, Characteristic functions
- Distributions: Normal/gaussian, Rayleigh, Chi-squared, Exponential
- Gaussian Vectors, Complex Gaussian
  - Circular symmetry vs isotropy
- Random processes:
  - stationarity, w.s.s., ergodicity
  - Autocorrelation, PSD, white gaussian noise
  - Random signals through LTI systems:
    - gaussian & wss useful properties that are preserved.
  - Frequency domain analysis possible