Probability and Random Variables

Carrson C. Fung
Dept. of Electronics Engineering
National Chiao Tung University

Motivations

- Message/signals, noises, and systems are random in nature
  - CDF/PDF are used as models
    - Parametric vs. nonparametric
  - speech signal can be modeled as a Markov chain (process) - current state depends partly on the previous state, but no other states. Each state contains a Gaussian mixture
  - Color image signal can be modeled by a Gaussian mixtures – each mode represents different color, say, R, G, B
  - Signal detection: optimal (LR and GLR) detector is a function of signals' PDFs
  - Channel model
    - Power delay profile of a channel
      - Strictly speaking, this models the output of the channel, but it is used to characterize the system response (similar to using $h(t)$ to characterize the I/O response of a LTI system)
- Deterministic vs. Probabilistic model
  - Different, so no better or worse
  - Performance of design depends on
    - System model (D/P)
    - Problem model (D/P)
    - Application
  - Like any modeling problems, if the underlying model is incorrect, results will be incorrect no matter how good the model is. E.g. models having a million parameters

Definition of Probability

- Relative frequency

$$\Pr(A) \triangleq \lim_{N\to\infty} \frac{N_A}{N}, \quad N_A: \text{ \# of event } A \text{ outcomes}, \; N: \text{ total \# of trials}$$

  - Experimental and intuitive but infinite # of experiments are required – not possible
  - Only approximates a probability
  - Not satisfactory for mathematical analysis
- Axiomatic theory
  - Mathematical
  - Rigorous
  - Facilitate further derivation

Example: Tossing Two Fair Coins

Relative Frequency	Axiomatic Theory
Model: $\begin{cases} \text{Events: A, B, } \cdots \\ \text{Prob. of Events:} \\ \Pr(A), \Pr(B), \cdots \end{cases}$ A: event for observing HH B: event for observing HT Possible outcomes: HH, HT, TH, TT To find $\Pr(A)$ and $\Pr(B)$, repeat experiment N times $\Pr(A) = \lim_{N\to\infty} \frac{N_A}{N} = \frac{1}{4}$ $\Pr(A \text{ or } B) = \lim_{N\to\infty} \frac{N_{A\cup B}}{N} = \frac{1}{2}$	S: Sample space $= \{HH, HT, TH, TT\}$ E: class of events, e.g. $\begin{Bmatrix} HH, HT, \ldots, \\ \{HH \cup HT\}, \ldots, \phi \end{Bmatrix}$ $\Pr(\cdot)$: real-valued probability function defined on E, i.e. $E \to [0,1]$ **Axiom 1**: $\Pr(A) \geq 0$ for all events A in S **Axiom 2**: The probability of all possible events occurring is unity, i.e. $\Pr(S) = 1$ **Axiom 3**: If the A and B are mutually exclusive events, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ If $A_1, A_2, \ldots$ are mutually exclusive $\Rightarrow \Pr(A_1 \cup A_2 \cup \cdots) = \Pr(A_1) + \Pr(A_2) + \cdots$

Probability Relationships

- $A \cup \overline{A} = S,\ \overline{A}:$ the event "not A"

From Axiom 2, $\Pr(A) + \Pr(\overline{A}) = \Pr(S) = 1$

$\Rightarrow \Pr(\overline{A}) = 1 - \Pr(A)$

- Generalization of Axiom 3

  $A \cup B = A \cup (B \cap \overline{A})$ (verify by Venn Diagram)

  $\Rightarrow \Pr(A \cup B) = \Pr(A) + \Pr(B \cap \overline{A})$

- Also, $A \cap B$ and $\overline{A} \cap B$ are disjoint and $(A \cap B) \cup (\overline{A} \cap B) = B$

  $\Rightarrow \Pr(A \cap B) + \Pr(\overline{A} \cap B) = \Pr(B)$

- From previous two relationships

  $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

Probability Relationships

- Conditional probability

$$\Rightarrow \Pr(A|B) \triangleq \frac{P(A \cap B)}{P(B)} \quad \text{or} \quad \Pr(B|A) \triangleq \frac{\Pr(A \cap B)}{\Pr(A)}$$

$$\Rightarrow \quad \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$$

$$\Rightarrow \Pr(B|A) = \frac{\Pr(A|B)P(B)}{\Pr(A)} \quad \text{(Bayes' rule)}$$

- Statistical independence

occurrence or nonocurrence of one event influence the other event

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

So using Bayes' rule: $\Pr(A \cap B) = \Pr(A)\Pr(B)$

Example 6.3: Bayes' Rule

Suppose two fair coins are tossed simultaneously. Let

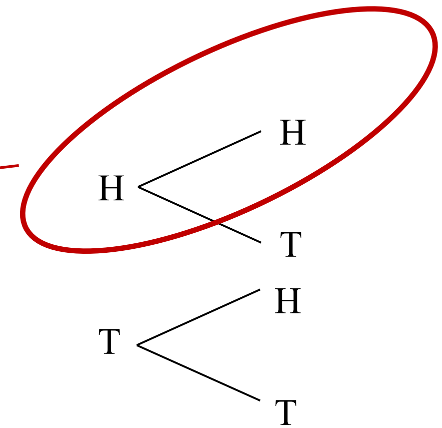
A: event that a least one head

B: event that there is a match

Find $\Pr(A)$, $\Pr(B)$, $\Pr(A|B)$, $\Pr(B|A)$, and $\Pr(A \cup B)$.

$$A: \text{HH, HT, TH} \Rightarrow \Pr(A) = \frac{3}{4}$$

$$B: \text{HH, TT} \Rightarrow \Pr(B) = \frac{1}{2}$$



$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/4}{1/2} = \frac{1}{2}$$

$$\Pr(B|A) = \frac{\Pr(B \cap A)}{\Pr(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Since $\Pr(A \cap B) = \Pr(A|B)\Pr(B) = \frac{1}{2} \bullet \frac{1}{2} = \frac{1}{4}$

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = \frac{3}{4} + \frac{1}{2} - \frac{1}{4} = 1$$

Example: Conditional Probability

Given $(x,y)=\{(1,0),\ (1,0),\ (2,0),\ (2,\ 1)\}$

What is $\Pr(x,y)$ and $\Pr(y|x)$?

$\Pr(x,y)$	y=0	y=1
x=1		
x=2		

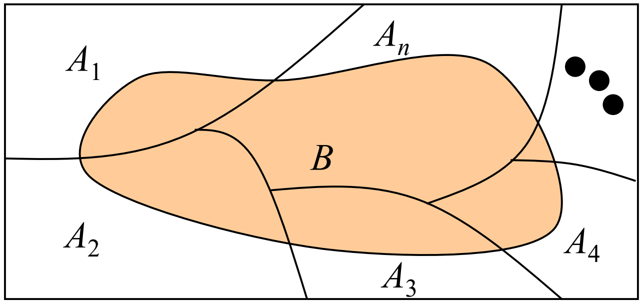
$\Pr(y|x)$	y=0	y=1
x=1		
x=2		

Multiplication Theorem for Conditional Probability (aka Chain Rule)

For any events $A_1, A_2, \ldots, A_n$

$$\begin{aligned}
\Pr\left(A_1 \cap A_2 \cap \cdots \cap A_n\right) &= \Pr\left(A_1\right)\Pr\left(A_2 \cap \cdots \cap A_n \middle| A_1\right) \\
&= \Pr\left(A_1\right)\Pr\left(A_2 \middle| A_1\right)\Pr\left(A_3 \cap \cdots \cap A_n \middle| A_1 \cap A_2\right) \\
&\;\;\vdots \\
&= \Pr\left(A_1\right)\Pr\left(A_2 \middle| A_1\right)\Pr\left(A_3 \middle| A_1 \cap A_2\right)\cdots\Pr\left(A_n \middle| A_1 \cap A_2 \cap \cdots \cap A_{n-1}\right)
\end{aligned}$$

Partitions and Total Probability



Suppose the events $A_1, A_2, \ldots, A_n$ form a partition of a sample space S, that is, the events A_i's are mutually exclusive and their union is S. Suppose B is any other event. Then

$$B = S \cap B = \left(\bigcup_{i=1}^{n} A_i \right) \cap B$$
$$= (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B),$$

where $A_i \cap B$ are also mutually exclusive. Then

$$\Pr(B) = \Pr(A_1 \cap B) + \Pr(A_2 \cap B) + \cdots + \Pr(A_n \cap B).$$

From the multiplication theorem,

$$\Pr(B) = \Pr(A_1)\Pr(B|A_1) + \Pr(A_2)\Pr(B|A_2) + \cdots + \Pr(A_n)\Pr(B|A_n).$$

This is known as the **law of total probability**.

Example 6.4: Statistical Independence

A single card is drawn at random from a deck of cards. Which of the following events are independent. (a) The card is a club, and the card is black. (b) The card is a king, and the card is black.

(a)

A: event that the card is a club

B: event that the card is black

There are 26 black cards, 13 of them are clubs $\Rightarrow \Pr(A|B) = \frac{13}{26} = \frac{1}{2}$

$\Pr(B) = \frac{26}{52} = \frac{1}{2}$, and $\Pr(A) = \frac{13}{52} = \frac{1}{4}$

So, $\Pr(A \cap B) = \Pr(A|B)\Pr(B) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$, but $\Pr(A)\Pr(B) = \frac{1}{8}$

Hence, A and B are not statistically independent. Why?

Example 6.4: Statistical Independence

(b)

A: The card is a king

B: event that the card is black

There are 26 black cards, 2 of them is a king $\Rightarrow \Pr\left(A|B\right) = \frac{2}{26} = \frac{1}{13}$

$$\Pr\left(B\right) = \frac{26}{52} = \frac{1}{2}, \text{ and } \Pr\left(A\right) = \frac{4}{52} = \frac{1}{13}$$

So, $\Pr\left(A \cap B\right) = \Pr\left(A|B\right)\Pr\left(B\right) = \left(\frac{1}{13}\right)\left(\frac{1}{2}\right) = \frac{1}{26}$, and $\Pr\left(A\right)\Pr\left(B\right) = \frac{1}{26}$

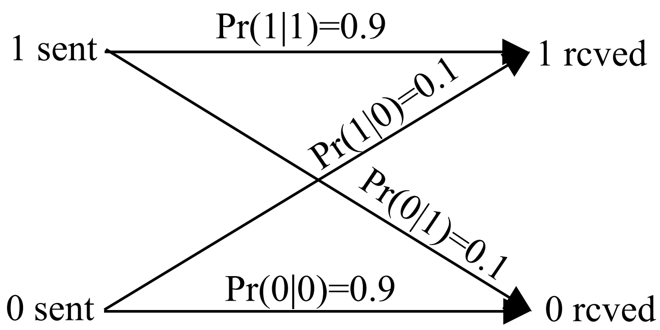
Hence, A and B are statistically independent. Why?

Example: Total Probability

Three machines A, B and C produce respectively 50%, 30%, and 20% of the total number of items of a factory. The percentages of defective output of these machines are 3%, 4%, and 5%. If an item is selected at random, find the probability that the item is defective.

Suppose an item is selected at random and is found to be defective. Find the probability that the item was produced by machine A; that is, find $\Pr\left(A \middle| D\right)$.

Example 6.5: Binary Channel



Notation: Pr(r|s)
(likelihood probability)

Given:

$\Pr(0s) = 0.8 \quad \Rightarrow \quad \Pr(1s) = 1 - \Pr(0s) = 0.2$

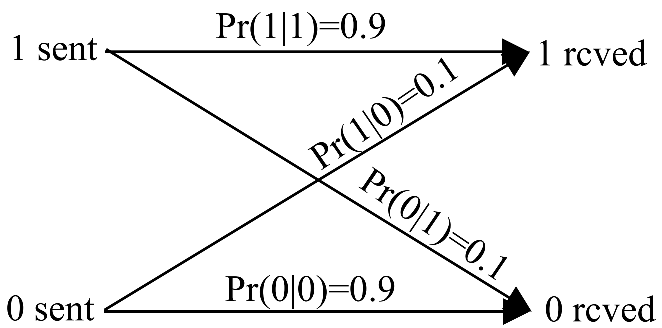
If 1 was received, what's the probability that 1 was sent?

$$\Pr(1s|1r) = \frac{\Pr(1s \cap 1r)}{\Pr(1r)} = \frac{\Pr(1r|1s)\Pr(1s)}{\Pr(1r)}$$

Note that
$$\begin{aligned}\Pr(1r) &= \Pr(1r \cap 1s) + \Pr(1r \cap 0s) \\ &= \Pr(1r|1s)\Pr(1s) + \Pr(1r|0s)\Pr(0s) \\ &= (0.9)(0.2) + (0.1)(0.8) \\ &= 0.26\end{aligned}$$

$$\Rightarrow \Pr(1s|1r) = \frac{\Pr(1s \cap 1r)}{\Pr(1r)} = \frac{\Pr(1r|1s)\Pr(1s)}{\Pr(1r)} = \frac{(0.9)(0.2)}{0.26} = 0.69$$

Example 6.5: Binary Channel



Notation: Pr(r|s)
(likelihood probability)

$$\Pr\left(0s \middle| 1r\right) = ?$$

$$\Pr\left(0s \middle| 0r\right) = ?$$

If 0 was received, what's the probability that 1 was sent?

$$\Pr\left(1s \middle| 0r\right) = \frac{\Pr\left(1s \cap 0r\right)}{\Pr\left(0r\right)} = \frac{\Pr\left(0r \middle| 1s\right)\Pr\left(1s\right)}{\Pr\left(0r\right)}$$

Note that $\Pr\left(0r\right) = \Pr\left(0r \cap 1s\right) + \Pr\left(0r \cap 0s\right)$

$$= \Pr\left(0r \middle| 1s\right)\Pr\left(1s\right) + \Pr\left(0r \middle| 0s\right)\Pr\left(0s\right)$$

$$= \left(0.1\right)\left(0.2\right) + \left(0.9\right)\left(0.8\right)$$

$$= 0.74$$

$$\Rightarrow \Pr\left(1s \middle| 0r\right) = \frac{\Pr\left(1s \cap 0r\right)}{\Pr\left(0r\right)} = \frac{\Pr\left(0r \middle| 1s\right)\Pr\left(1s\right)}{\Pr\left(0r\right)} = \frac{\left(0.1\right)\left(0.2\right)}{0.74} = 0.03$$

Random Variable (RV)

- A random variable is a function that assigns a numerical value each possible outcome in S, i.e. $S \rightarrow \Re$ (field of real number)
  - More convenient to work with a numerical value than nonnumerical value
- Can be discrete or continuous (example of discrete RV on top right, continuous RV on bottom right)
- Convention
  - Capital letters denote RVs
  - Lowercase letters denote values the RVs take on
    - E.g. $f_X(x)$ distribution function for RV X with value x

Table 5.2 Possible Random Variables (RV)

Outcome: S_i	RV No. 1: $X_1(S_i)$	RV No. 2: $X_2(S_i)$
S_1 = heads	$X_1(S_1) = 1$	$X_2(S_1) = \pi$
S_2 = tails	$X_1(S_2) = -1$	$X_2(S_2) = \sqrt{2}$

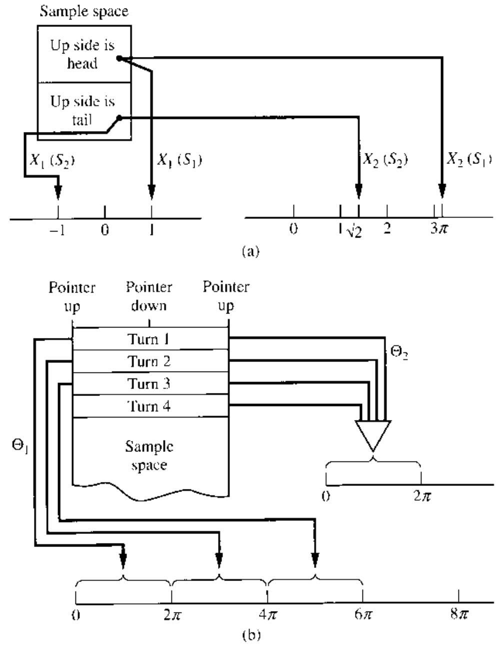


Figure 5.4
Pictorial representation of sample spaces and random variables.
(a) Coin-tossing experiment.
(b) Pointer-spinning experiment.

CDF and PDF

- Functions which relates the probability of an event to a numerical value assigned to an event
- Parameter vs. nonparameteric
  - There are several different parametric PDFs
  - Nonparametric
    - Estimated directly from data
    - Easily adaptable

Probability (Cumulative) Distribution Functions

- A way to probabilistically describe an RV

$$F_X(x) \triangleq \Pr(X \le x)$$

Properties of $F_X(x)$

1. $0 \le F_X(x) \le 1$, with $F_X(-\infty) = 0$, $F_X(\infty) = 1$
2. $F_X(x)$ is continuous from the right, that is,

$$\lim_{x \to x_0^+} F_X(x) = F_X(x_0)$$

3. $F_X(x)$ is a nondecreasing function of x, i.e.

$$F_X(x_1) \le F_X(x_2) \text{ if } x_1 < x_2$$

From 2., $F_X(x)$ is continuous from right, so the jump amount = P_0

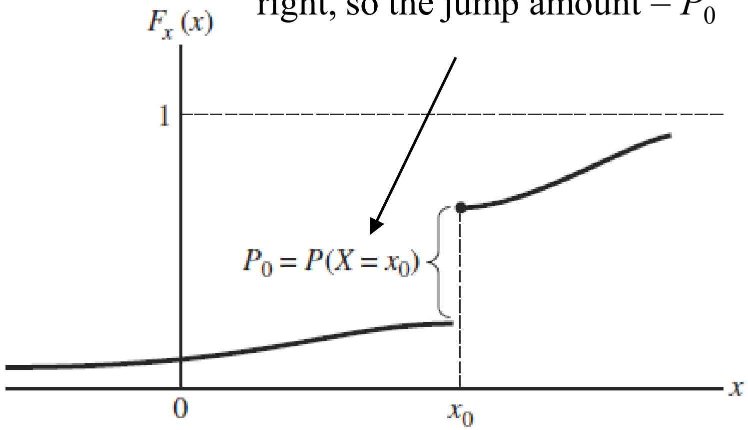


Figure 6.5
Illustration of the jump property of $f_X(x)$.

Probability Density Functions (PDF)

More convenient to express statistical averages using PDFs

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Properties of $f_X(x)$

1. $F_X(x) = \int_\eta f_X(\eta)\,d\eta \quad \Rightarrow f_X(x) = \frac{dF_X(x)}{dx} \geq 0$

2. $\int_x f(x)\,dx = 1$

3. $\Pr(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx$

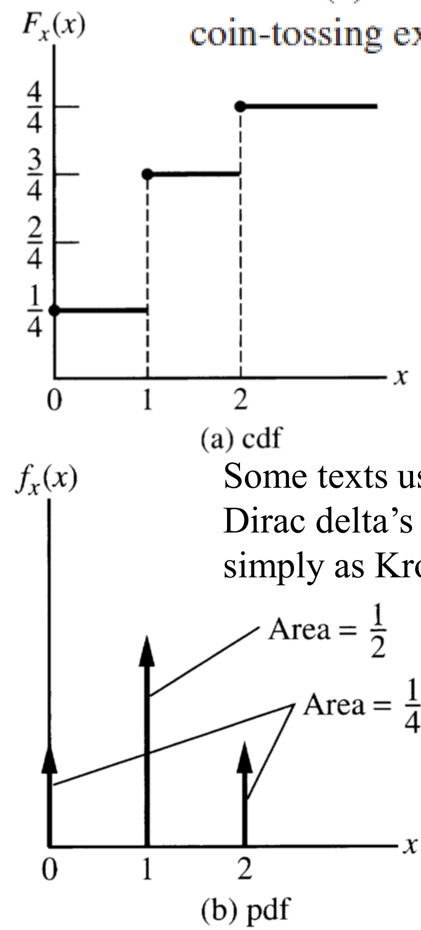
4. $f_X(x)\,dx = P(x - dx < X \leq x)$

Example 6.9 – Discrete PDF and CDF

- 2 fair coins are tossed
- X: # of heads

Outcome	X	$\Pr(X=x_j)$
TT	$x_1=0$	¼
TH HT	$x_2=1$	½
HH	$x_3=3$	¼

Figure 6.6
The cdf (a) and pdf (b) for a coin-tossing experiment.



Some texts use pmf where the Dirac delta's are represented simply as Kronecker delta's

Example 6.10: Cont. PDF and CDF

Consider the pointer-spinning experiment. Assume any one stopping point is not favored over any other and that the RV Θ is defined as the angle that the pointer makes with the vertical, modulo 2π. Thus Θ is limited to $[0,2\pi)$ and for any two angles θ_1 and θ_2 in $[0,2\pi)$, we have

$$\Pr(\theta_1 - \Delta\theta < \Theta \le \theta_1) = \Pr(\theta_2 - \Delta\theta < \Theta \le \theta_2) \quad \text{(equally likely assumption)}$$

$$\Rightarrow f_\Theta(\theta_1) = f_\Theta(\theta_2),\ 0 \le \theta_1, \theta_2 < 2\pi.$$

$$\Rightarrow f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & 0 \le \theta < 2\pi, \\ 0, & \text{otherwise} \end{cases}$$

Area under PDF curve is the probability.

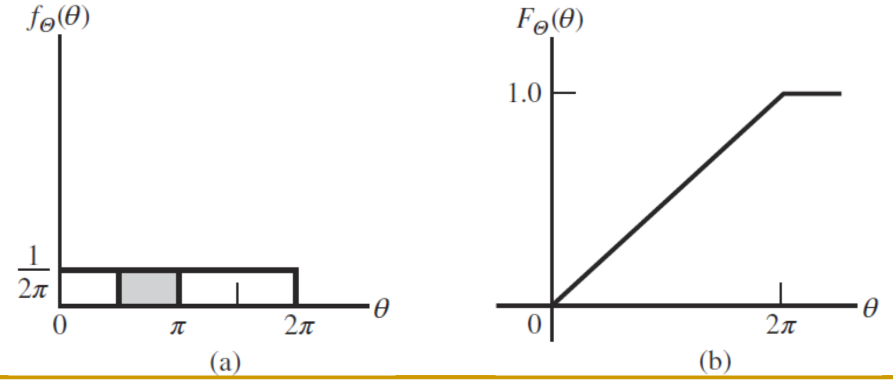


Figure 6.7
The pdf (a) and cdf (b) for a pointer-spinning experiment.

Joint CDFs and PDFs

Characterized by two or more RVs

$$F_{XY}(x,y) = \Pr(X \le x, Y \le y)$$

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

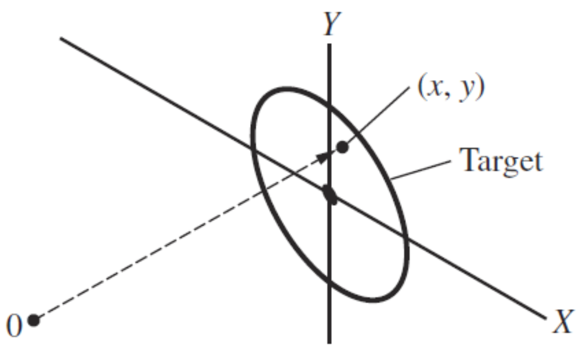


Figure 6.8
The dart-throwing experiment.

$$\Pr(x_1 < X \le x_2, y_1 < Y \le y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x,y)\,dxdy$$

$$\Rightarrow F_{XY}(\infty,\infty) = \int_y \int_x f_{XY}(x,y)\,dxdy = 1$$

$$\Rightarrow f_{XY}(x,y)\,dxdy = P(x - dx < X \le x, y - dy < Y \le y)$$

Marginal CDFs and PDFs

Can obtain cdf or pdf of one of the RVs from joint RVs

$$F_X(x,y) = \Pr(X \le x, Y \le \infty) = F_{XY}(x,\infty)$$

$$F_Y(x,y) = \Pr(X \le \infty, Y \le y) = F_{XY}(\infty, y)$$

$$F_X(x) = \int_{y'} \int_{-\infty}^{x} f_{XY}(x', y')\, dx' dy'$$

$$F_Y(y) = \int_{-\infty}^{y} \int_{x} f_{XY}(x', y')\, dx' dy'.$$

Since $f_X(x) = \dfrac{dF_X(x)}{dx}$ and $f_Y(y) = \dfrac{dF_Y(y)}{dy}$

$$\Rightarrow f_X(x) = \int_{y'} f_{XY}(x, y')\, dy' \text{ and } f_Y(y) = \int_{x'} f_{XY}(x', y)\, dx'$$

Conditional CDFs and PDFs

Conditional RV:

$$F_{X|Y}\left(x|Y\right)=F_{X|Y}\left(x|Y\le y\right)=\frac{F_{XY}\left(x,y\right)}{F_Y\left(y\right)}$$

$$f_{X|Y}\left(x|y\right)=\frac{\partial F_{X|Y}\left(x|Y=y\right)}{\partial x}=\frac{f_{XY}\left(x,y\right)}{f_Y\left(y\right)}$$

Bayes Theorem:

$$f_{X|Y}\left(x|y\right)=\frac{f_{XY}\left(x,y\right)}{f_Y\left(y\right)}=\frac{f_{Y|X}\left(y|X=x\right)f_X\left(x\right)}{f_Y\left(y\right)}=\frac{f_{Y|X}\left(y|x\right)f_X\left(x\right)}{f_Y\left(y\right)}$$

where $f_{Y|X}\left(y|x\right)dx=\Pr\left(y-dy<Y\le y \text{ given } X=x\right)$.

Statistical Independence

Two RVs are stat. independent if values one takes on do not influence the values that the other takes on.

$$\Rightarrow \Pr\left(X \le x, Y \le y\right) = \Pr\left(X \le x\right)\Pr\left(Y \le y\right) \quad \text{or}$$

$$F_{XY}\left(x, y\right) = F_X\left(x\right)F_Y\left(y\right)$$

$$f_{XY}\left(x, y\right) = f_X\left(x\right)f_Y\left(y\right)$$

If X and Y are not independent, then using Bayes' rule

$$f_{XY}\left(x, y\right) = f_X\left(x\right)f_{Y|X}\left(y|x\right) = f_Y\left(y\right)f_{X|Y}\left(x|y\right).$$

Example 6.11: Statistical Independence

Two RVs X and Y have joint pdf

$$f_{XY}(x,y)=\begin{cases} Ae^{-(2x+y)}, & x,y\ge 0 \\ 0, & \text{otherwise.} \end{cases}$$

A can be found by noting that

$$F_{XY}(\infty,\infty)=\int_y\int_x f_{XY}(x,y)\,dxdy=1$$

Since $\int_0^\infty\int_0^\infty Ae^{-(2x+y)}dxdy=1 \;\Rightarrow\; A=2$

$$f_X(x)=\int_y f_{XY}(x,y)\,dy=\begin{cases} \int_0^\infty 2e^{-(2x+y)}dy, & x\ge 0 \\ 0, & x<0 \end{cases}=\begin{cases} 2e^{-2x}, & x\ge 0 \\ 0, & x<0 \end{cases}$$

$$f_Y(y)=\int_x f_{XY}(x,y)\,dx=\begin{cases} e^{-y}, & y\ge 0 \\ 0, & y<0 \end{cases}$$

$$f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\begin{cases} 2e^{-2x}, & x\ge 0 \\ 0, & x<0 \end{cases}$$

$$f_{Y|X}(y|x)=\frac{f_{XY}(x,y)}{f_X(y)}=\begin{cases} e^{-y}, & y\ge 0 \\ 0, & y<0 \end{cases}$$

Conditional prob's are equal to respective marginals ➔ X and Y are independent.

Example 6.11: Statistical Independence

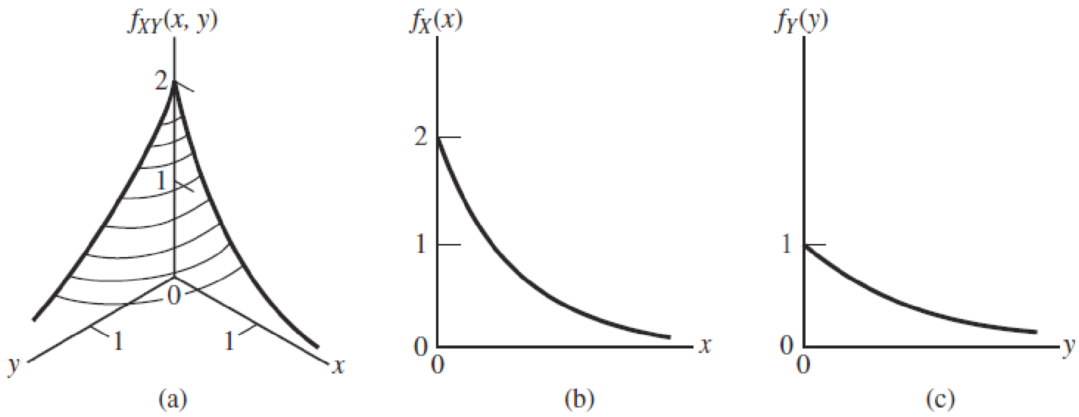


Figure 6.9
Joint and marginal pdfs for two random variables. (a) Joint pdf. (b) Marginal pdf for X. (c) Marginal pdf for Y.

$$f_{XY}(x,y)=\begin{cases} Ae^{-(2x+y)}, & x,y\ge 0 \\ 0, & \text{otherwise.}\end{cases}$$

$$f_X(x)=\int_y f_{XY}(x,y)\,dy=\begin{cases}\int_0^\infty 2e^{-(2x+y)}, & x\ge 0\\ 0, & x<0\end{cases}=\begin{cases}2e^{-2x}, & x\ge 0\\ 0, & x<0\end{cases}$$

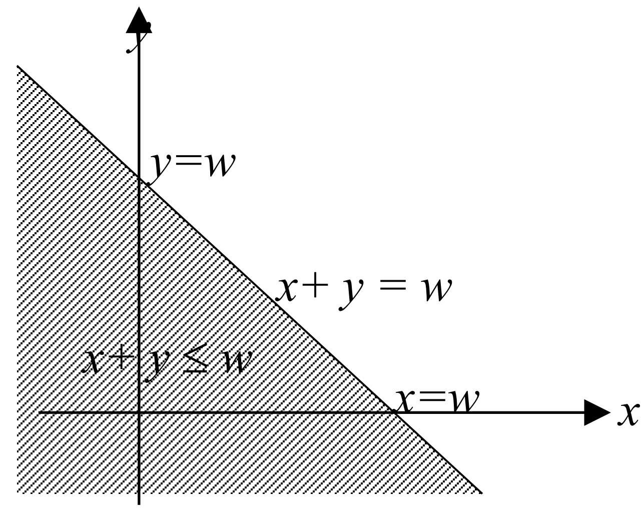
$$f_Y(y)=\int_x f_{XY}(x,y)\,dx=\begin{cases}e^{-y}, & y\ge 0\\ 0, & y<0\end{cases}$$

Sum of Two Statistically Indep. RVs

- The density of the sum of two statistically independent RVs is the convolution of their individual density functions.
- Suppose X, and Y are three independent RVS where $W = X + Y$, then

$$f_W(w) = \int_y f_Y(y) f_X(w-y)\, dy$$

$f_W(w)$, $f_X(x)$, and $f_Y(y)$ are pdfs of W, X, and Y, respectively



$$F_W(w) = P(W \le w) = P(X + Y \le w)$$

$$= \int_y \int_{x=-\infty}^{w-y} f_{X,Y}(x, y)\, dxdy$$

$$= \int_y f_Y(y) \int_{x=-\infty}^{w-y} f_X(x)\, dxdy \quad \text{(stat. indep.)}$$

Differentiating we get the result

Transformation of RVs (Monotonic)

Given a known X, define a second RV such that it is a function of the first:

$$Y = g(X).$$

Assume $g(X)$ monotonic. Probability that X lies in the range $(x - dx, x)$ is the same as the probability that Y lies in the range $(y - dy, y)$

$$\begin{cases} f_X(x)dx = f_Y(y)dy, & g(X) \text{ monotonically increasing} \\ f_X(x)dx = -f_Y(y)dy, & g(X) \text{ monotonically decreasing} \end{cases}$$

$$\Rightarrow f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|_{x=g^{-1}(y)}, \text{ for } x = g^{-1}(y)$$

$$\Rightarrow f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{dx}{dy}\right|_{x=g^{-1}(y)}.$$

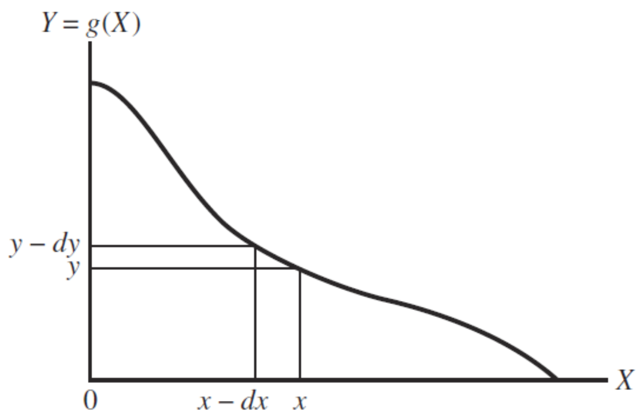


Figure 6.11
A typical monotonic transformation of a random variable.

Example 6.13

From Ex 6.10, suppose $f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi}, & 0 \le \theta \le 2\pi, \\ 0, & \text{otherwise.} \end{cases}$ Note that $f_\Theta(\theta) \sim Unif(0, 2\pi)$

Suppose Y is a transformed version of Θ according to

$$Y = -\left(\frac{1}{\pi}\right)\Theta + 1 \quad \Rightarrow \quad \theta = -\pi y + \pi$$

$$\Rightarrow f_Y(y) = f_\Theta(\theta)\Big|_{\theta=-\pi y+\pi} \left|\frac{d\theta}{dy}\right|_{\theta=-\pi y+\pi} = \begin{cases} \frac{1}{2}, & -1 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Since transformation is affine $\Rightarrow f_Y(y)$ is also uniform, $Unif(-1,1)$.

Transformation of RVs (Nonmonotonic)

Assume $g(X)$ nonmonotonic. $(y-dy, y)$ corresponds to three infinitesimal intervals: $(x_1 - dx_1, x_1)$, $(x_2 - dx_2, x_2)$, $(x_3 - dx_3, x_3)$

Probability that X lies in any of these intervals equal to the probability that Y lies in $(y-dy, y)$.

Generalizing to N disjoint intervals:

$$\Pr(y - dy < Y \le y) = \sum_{i=1}^{N} P(x_i - dx_i < X \le x_i).$$

Since $\Pr(y - dy < Y \le y) = f_Y(y)|dy|$ and

$$\Pr(x_i - dx_i < X \le x_i) = f_X(x_i)|dx_i|, \qquad \text{for } x_i = g_i^{-1}(y)$$

$$\Rightarrow f_Y(y) = \sum_{i=1}^{N} f_X\left(g_i^{-1}(y)\right)\left|\frac{dx_i}{dy}\right|_{x_i = g_i^{-1}(y)}.$$

Absolute value is used to keep probability nonnegative.

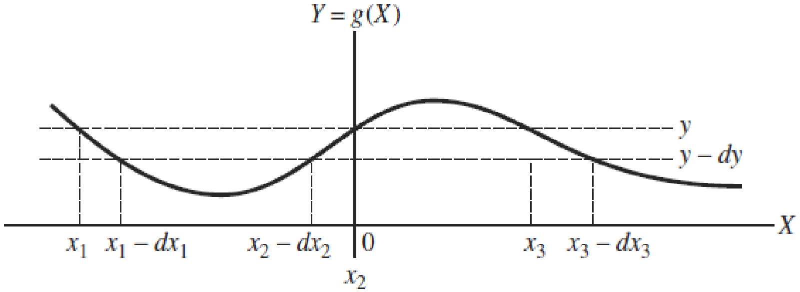


Figure 6.12
A nonmonotonic transformation of a random variable.

Example 6.14

Consider the transformation $y = x^2$. If $f_X(x) = 0.5e^{-|x|}$, find $f_Y(y)$.

There are 2 solutions for $y = x^2$: $\begin{cases} x_1 = \sqrt{y}, & x_1 \geq 0 \\ x_2 = -\sqrt{y}, & x_2 < 0 \end{cases}$, for $y \geq 0$.

$$\Rightarrow \begin{cases} \dfrac{dx_1}{dy} = \dfrac{1}{2\sqrt{y}}, & \text{for } x_1 \geq 0 \\ \dfrac{dx_2}{dy} = -\dfrac{1}{2\sqrt{y}}, & \text{for } x_2 < 0 \end{cases}$$

$$\Rightarrow f_Y(y) = \sum_{i=1}^{N} f_X\left(x_i = g_i^{-1}(y)\right) \left|\frac{dx_i}{dy}\right|_{x_i = g_i^{-1}(y)}$$

$$= \sum_{i=1}^{N} \frac{1}{2} e^{-\left|x_i = g_i^{-1}(y)\right|} \left|\frac{dx_i}{dy}\right|_{x_i = g_i^{-1}(y)}$$

$$= \frac{1}{2} e^{-\sqrt{y}} \left|-\frac{1}{2\sqrt{y}}\right| + \frac{1}{2} e^{-\sqrt{y}} \left|\frac{1}{2\sqrt{y}}\right|$$

$$= \left|\frac{1}{2\sqrt{y}}\right| e^{-\sqrt{y}}$$

$$= \begin{cases} \dfrac{1}{2\sqrt{y}} e^{-\sqrt{y}}, & \text{for } y > 0 \\ 0, & \text{for } y \leq 0 \end{cases}$$

Because Y can't be < 0

Transformation of >2 RVs (NonMonotonic)

For random vector $\mathbf{X}$, with pdf $f_{\mathbf{X}}(\mathbf{x})$, let $f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} = f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$, where $d\mathbf{y} = dy_1 dy_2 \cdots dy_n.$

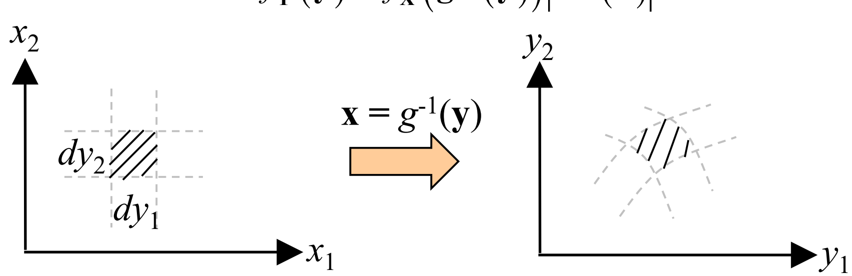
Assume $\mathbf{x} = \mathbf{g}^{-1}(\mathbf{y})$ has an inverse

$$\Rightarrow f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})\frac{d\mathbf{x}}{d\mathbf{y}}$$

$$\Rightarrow f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(\mathbf{g}^{-1}(\mathbf{y})\right)\frac{d\mathbf{x}}{d\mathbf{y}},$$

where $d\mathbf{x} = \left|\det(\mathbf{J})\right| d\mathbf{y}$, where $\mathbf{J} \triangleq \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_N}{\partial y_1} & \cdots & \frac{\partial x_N}{\partial y_n} \end{bmatrix}$ is the Jacobian matrix.

$$\Rightarrow f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(\mathbf{g}^{-1}(\mathbf{y})\right)\left|\det(\mathbf{J})\right|$$



Example 6.15

Consider the dart throwing example. Assume that joint PDF in terms of rectangular coordinates for the impact point is

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\left(x^2+y^2\right)\right],\ \text{for } -\infty<x,y,<\infty,$$

where σ^2 is a constant. Express $f_{XY}(x,y)$ using the poloar coordinate system where

$$R=\sqrt{X^2+Y^2},\ \text{and } \Theta=\tan^{-1}\left(\frac{Y}{X}\right),$$

for $0\le R<\infty,\ 0\le\Theta<2\pi.$

$$\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}=\begin{bmatrix}\frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta}\\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta}\end{bmatrix}\begin{bmatrix}\Delta r\\ \Delta \theta\end{bmatrix}$$

$$\begin{cases}X=R\cos\Theta=g_1^{-1}(R,\Theta)\\ Y=R\sin\Theta=g_2^{-1}(R,\Theta)\end{cases}\Rightarrow \mathbf{J}=\begin{bmatrix}\frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta}\\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta}\end{bmatrix}=\begin{bmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{bmatrix}$$

$$\Rightarrow\left|\det(\mathbf{J})\right|=\left|\det\begin{pmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{pmatrix}\right|=r\cos^2\theta+r\sin^2\theta=r$$

$$\Rightarrow f_{R\Theta}(r,\theta)=f_{\mathbf{X}}\left(\mathbf{g}^{-1}(\mathbf{y})\right)\left|\det(\mathbf{J})\right|$$

$$=\frac{r}{2\pi\sigma^2}\exp\left[-\frac{r^2}{2\sigma^2}\right],\quad \text{for } 0\le r<\infty$$

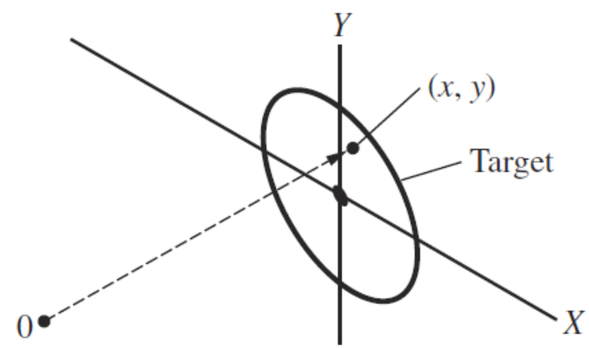


Figure 6.8
The dart-throwing experiment.

Example 6.15

Given $f_{R\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2}\exp\left[-\frac{r^2}{2\sigma^2}\right], \quad \text{for } 0 \le r < \infty$

Integrate $f_{R\Theta}(r,\theta)$ over θ, we have

$$f_R(r) = \frac{r}{\sigma^2}\exp\left[-\frac{r^2}{2\sigma^2}\right], \quad \text{for } 0 \le r < \infty$$

This is known as the **Rayleigh distribution function**.

We see that most probable distance for the dart to land from the bulleye is $R = \sigma$.

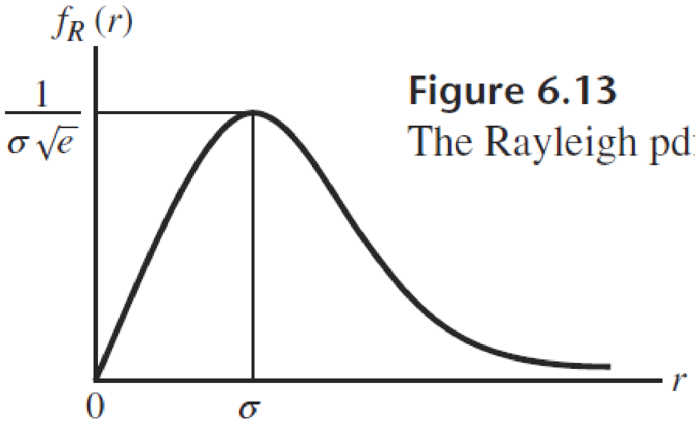


Figure 6.13
The Rayleigh pdf.

Rayleigh distribution is used to model distribution of power profile of wireless channels

$x = \sqrt{x_1^2 + x_2^2}$, where $x_1, x_2 \sim N(0, \sigma^2)$ indep. RVs

Statistical Averages

- Sometimes full description of RVs, i.e. knowing its CDF or PDF are not required
- Sometimes only partial information is needed
  - One type of partial information of a set of RVs statistical average or mean value

Average of Discrete RV

Expectation of M RVs, $x_1, \ldots, x_M$ with respective probabilities $P_1, \ldots, P_M$

$$\mu_x \triangleq E[X] = \sum_{j=1}^{M} x_j P_j$$

Justification:

Let experiment be perform N number of time, **with N large**

Arithmetic mean: $\dfrac{n_1 x_1 + \cdots + n_m x_m}{N} = \sum_{j=1}^{M} x_j \dfrac{n_j}{N}$

By relative frequency interpretation: $\lim_{N \to \infty} \dfrac{n_j}{N} = P_j$

$$\Rightarrow \frac{n_1 x_1 + \cdots + n_m x_m}{N} = \sum_{j=1}^{M} x_j P_j$$

Average of Cont. RV

Expectation of x_0 to x_M with pdf $f_X(x)$. Suppose we break up this interval into subintervals of size Δx (assume small). The probability that X lies between $x_i - \Delta x$ to x_i is

$$\Pr\left(x_i - \Delta x < X \le x_i\right) \approx f_X\left(x_i\right)\Delta x, \text{ for } i = 0, \ldots, M.$$

Hence, approximated X by a discrete RV that takes on values x_0 to x_M with probabilities $f_X\left(x_0\right)\Delta x, \ldots, f_X\left(x_M\right)\Delta x$.

$$\Rightarrow \mu_x \triangleq E\left[X\right] \approx \sum_{i=1}^{M} x_i f_X\left(x_i\right)\Delta x \overbrace{=}^{\lim\limits_{\Delta x \to 0}} \int_x x f_X\left(x\right) dx$$

Properties of Expectation

- $E[\cdot]$ is a linear operator
  - Sometimes need to perform $E(tr(\cdot))$. $tr(\cdot)$ is also linear operator ➔ $E(tr(\cdot)) = tr(E(\cdot))$
  - Additive
    - $E[X+Y] = E[X] + E[Y]$ for any 2 RVs
  - Homogeneity
    - $E[cX] = cE[X]$, for any constant c

Example

In a gambling game, the expected value E of the game is considered to be the value of the game to the player. The game is said to be favorable to the player if E is positive, and unfavorable if E is negative. If $E = 0$, the game is fair.

A player tosses a fair die. If a prime number occurs he wins that number of dollars, but if a non-prime number occurs he loses that number of dollars. The possible outcomes x_i of the game with their respective probabilities $f(x_i)$ are as follows:

x_i	2	3	5	-1	-4	-6
$f(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

The negative numbers correspond to the fact that the player loses if a non-prime number occurs. The expected value of the game is

$$E = 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} - 1 \cdot \frac{1}{6} - 4 \cdot \frac{1}{6} - 6 \cdot \frac{1}{6}.$$

So the game is unfavorable to the player since the expected value is negative.

Average of a Function of a RV

Let $Y = g(X)$.

$$\mu_Y \triangleq E[Y] = \begin{cases} \sum_i y_i \Pr(y_i), & \text{discrete RV} \\ \int_y y f_Y(y)\,dx, & \text{cont. RV} \end{cases}.$$

r^{th} moment of X, for $r = 0, 1, 2, \ldots$. Let $Y = g(X) = X^r$

$$\xi_r \triangleq E\left[X^r\right] = \begin{cases} \sum_i x_i^r \Pr(x_i), & \text{discrete RV} \\ \int_x x^r f_X(x)\,dx, & \text{cont. RV} \end{cases}$$

r^{th} central moment of X, for $r = 0, 1, 2, \ldots$. Let $Y = g(X) = (X - \mu_X)^r$

$$m_r \triangleq E\left[(X - \mu_X)^r\right]$$

Special case: variance: $r = 2$

$$\text{var}[X] \triangleq m_2 \triangleq E\left[(X - \mu_X)^2\right] = E\left[X^2\right] - \mu_x^2 \triangleq \sigma_X^2$$

Properties of Variance

Let X be a random variable and k is a real number, then (i) $\operatorname{var}(X+k)=\operatorname{var}(X)$ and (ii) $\operatorname{var}(kX)=k^2\operatorname{var}(X)$.

Proof:

(i)
$$\operatorname{var}(X+k)=E\left[(X+k)(X+k)\right]-E^2(X+k)$$
$$=E(X^2)+2kE(X)+k^2-\left[E^2(X)+2kE(X)+k^2\right]$$
$$=E(X^2)-E^2(X)=\operatorname{var}(X)$$

(ii)
$$\operatorname{var}(kX)=E(k^2X^2)-E^2(kX)$$
$$=E(k^2X^2)-k^2E^2(X)$$
$$=k^2\left[E(X^2)-E^2(X)\right]=k^2\operatorname{var}(X)$$

Average of a Function of a RV

r^{th} joint moment of X and Y, for $i, j = 0,1,2,\ldots$

$$\xi_{ij} \triangleq E\left[X^i Y^j\right] = \begin{cases} \sum_{i,j} x_\ell^i y_m^j P(x_i, y_m), & \text{discrete RV} \\ \int_{x,y} x^i y^j f_{XY}(x,y)\,dxdy, & \text{cont. RV} \end{cases}$$

Correlation: $\xi_{11} \triangleq E[XY]$

Note:

Independent: $E_{XY}(XY) = E_X(X)E_Y(Y)$

Uncorrelated: $E_{XY}\left[(X-\mu_X)(Y-\mu_Y)\right] = 0$

Orthogonal: $E(XY) = 0$

Implications:

- If X and Y are independent and have zero mean, implies X and Y are uncorrelated and orthogonal.
- If X and Y are uncorrelated and have zero mean, implies they are orthogonal.
- Hence, independence is the strongest of the three properties.

Average of a Function of a RV

r^{th} joint central moment of X and Y, for $i, j = 0,1,2,\ldots$

$$m_{ij} \triangleq E\left[\left(X-\mu_X\right)^i\left(Y-\mu_Y\right)^j\right]$$

Covariance:

$$Cov\left[X,Y\right] \triangleq m_{11} \triangleq E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right] = E\left[XY\right]-\mu_X\mu_Y$$

Correlation coefficient for X and Y:

$$\rho \triangleq \frac{m_{11}}{\sqrt{m_{20}m_{02}}} = \frac{Cov\left[X,Y\right]}{\sqrt{\sigma_X^2\sigma_Y^2}}$$

Example

X\Y	4	10	Sum
1	1/4	1/4	1/2
3	1/4	1/4	1/2
Sum	1/2	1/2	

X'\Y'	4	10	Sum
1	0	1/2	1/2
3	1/2	0	1/2
Sum	1/2	1/2	

What is $E(XY)$ and $E(X'Y')$? What is Cov(X,Y) and Cov(X',Y')?

Conditional Expectation

Conditional expectation of X given $Y = y$

$$E\left[X \middle| Y\right] = E\left[X \middle| Y = y\right] = \int_x x f_{X|Y}\left(x \middle| Y = y\right) dx$$

Expectation of functions of X: $\quad Y = g\left(X\right)$

$$E\left[Y\right] = E\left[g\left(X\right)\right] = \int_x g\left(x\right) f_X\left(x\right) dx$$

Removing Conditional Expectation Via Expectation

Since $E_{X|Y}\left(X|Y\right)$ is a function of Y, it is also a RV.

$$
\begin{aligned}
E_Y\left[E_{X|Y}\left(X|Y\right)\right] &= \int_y \int_x x f_{X|Y}\left(x|y\right) dx\, f_Y\left(y\right)\, dy \\
&= \int_x x \int_y f_{X|Y}\left(x|y\right) f_Y\left(y\right) dy\, dx \\
&= \int_x x \int_y f_{XY}\left(xy\right) dy\, dx \\
&= \int_x x f_X\left(x\right) dx \\
&= E_X\left[X\right]
\end{aligned}
$$

Conditional Expectation

This is an "expectation" version of the total probability theorem.
In many cases, we can simplify a problem by conditioning or "fixing" one RV and performing an expectation. Then remove the conditioning in a second step by taking the expectation w.r.t. the conditioning RV.

More generally:

$$E\left[g\left(X\right)\right]=E_Y\left[E_X\left(g\left(X\right)\middle|Y\right)\right]$$

Example: Nonlinear MMSE Filter

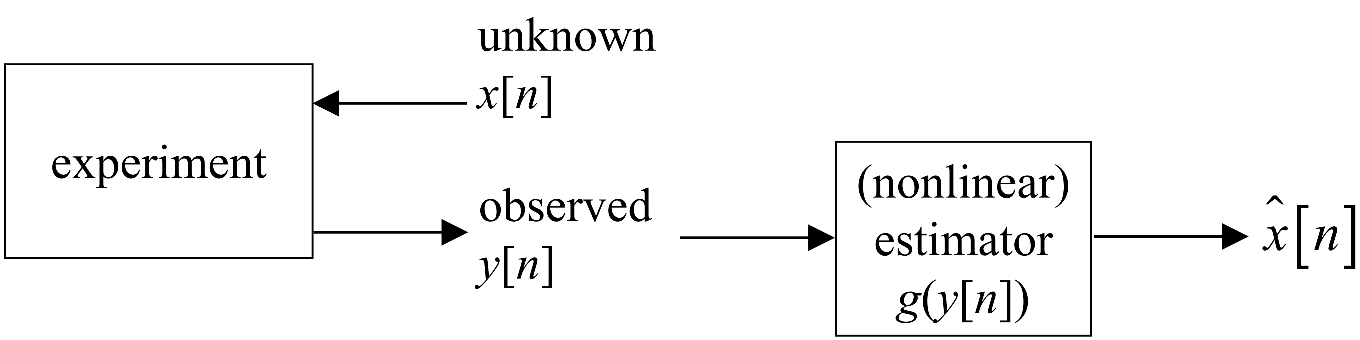
Suppose we want to recover the transmit signal $x[n]$ from the received signal $y[n]$ the MMSE filter $g(Y)$.
Cost function to consider

$$\mathrm{BMSE}(\hat{x}[n]) = E_{Y,X}\left[(x[n]-\hat{x}[n])^2\right] = \int\int (x[n]-\hat{x}[n])^2\, p(y,x)\, dx\, dy$$

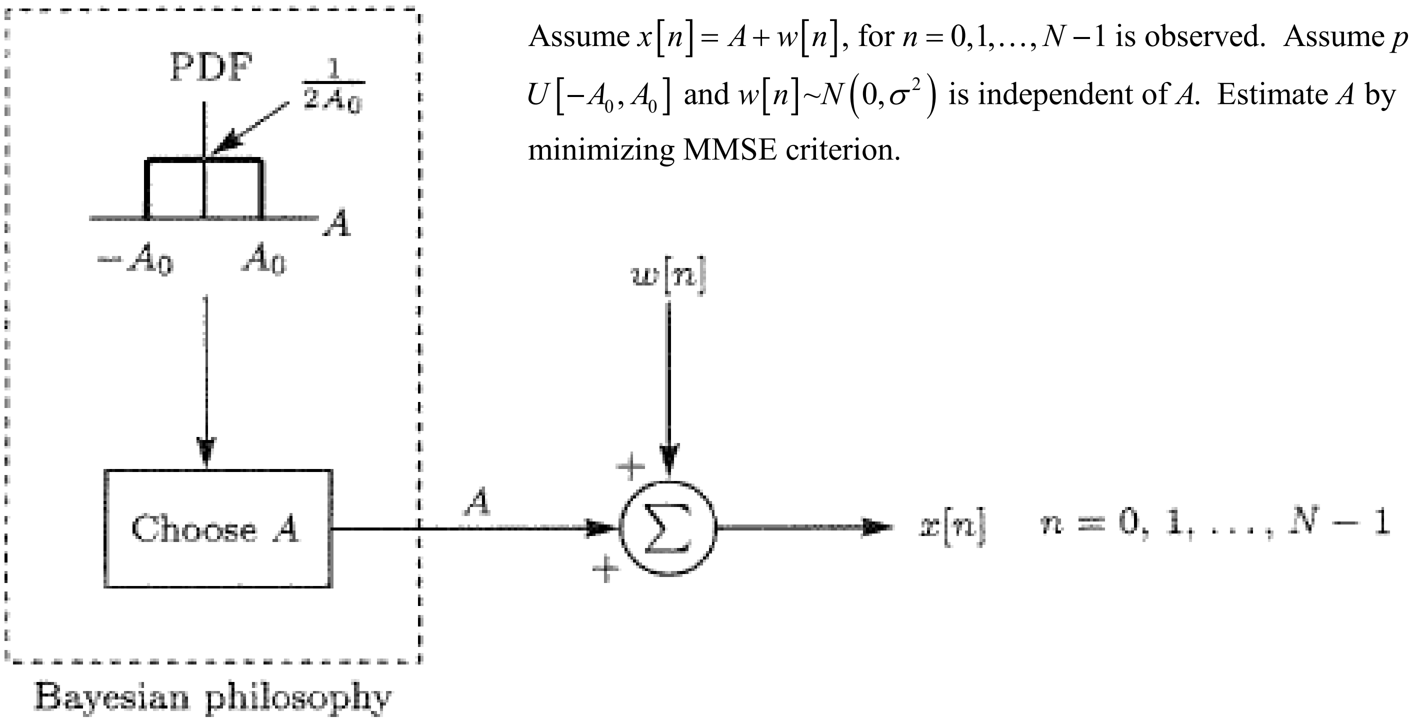
$$= \int\int (x[n]-\hat{x}[n])^2\, p(x|y)\, dx\, p(y)dy = \int E_{X|Y}\left[(x[n]-g(y[n]))^2 \middle| y[n]\right] p(y)dy$$

$$= E_Y\left\{E_{X|Y}\left[(x[n]-g(y[n]))^2 \middle| y[n]\right]\right\}$$

$$\Rightarrow \min_{g(y)} E_{X|Y}\left[(x[n]-g(y[n]))^2 \middle| y[n]\right]$$



BMSE Example



Assume $x[n] = A + w[n]$, for $n = 0, 1, \ldots, N-1$ is observed. Assume $p(A) = U[-A_0, A_0]$ and $w[n] \sim N(0, \sigma^2)$ is independent of A. Estimate A by minimizing MMSE criterion.

A is considered to be a random variable with a prior pdf. We attempt to estimate the realization of A

BMSE Example

Assume $x[n] = A + w[n]$, for $n = 0, 1, \ldots, N-1$ is observed. Assume $p(A) = U[-A_0, A_0]$ and $w[n] \sim N(0, \sigma^2)$ is independent of A. Estimate A by minimizing MMSE criterion.

Since $p_x(x[n] \| A) = p_w(x[n] - A | A) = p_w(x[n] - A)$

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

Hence $p(\mathbf{x}|A) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n] - A)^2\right]$

BMSE Example

$$p(A|\mathbf{x}) = \frac{p(A,\mathbf{x})}{p(\mathbf{x})} = \frac{p(\mathbf{x}|A)p(A)}{\int p(\mathbf{x}|A)p(A)dA}$$

Then the posterior pdf becomes

$$p(A|\mathbf{x}) = \begin{cases} \dfrac{\dfrac{1}{2A_0(2\pi\sigma^2)^{N/2}} \exp\left[-\dfrac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\int_{-A_0}^{A_0} \dfrac{1}{2A_0(2\pi\sigma^2)^{N/2}} \exp\left[-\dfrac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] dA}, & |A| \le A_0 \\ 0, & |A| > A_0 \end{cases}$$

$$= \begin{cases} \dfrac{\dfrac{1}{2A_0(2\pi\sigma^2)^{N/2}} \exp\left[-\dfrac{1}{2\sigma^2}\left(N(A-\bar{x})^2 + \sum_{n=0}^{N-1}x^2[n] - N\bar{x}^2\right)\right]}{\int_{-A_0}^{A_0} \dfrac{1}{2A_0(2\pi\sigma^2)^{N/2}} \exp\left[-\dfrac{1}{2\sigma^2}N(A-\bar{x})^2\right] dA \bullet \exp\left[-\dfrac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}x^2[n] - N\bar{x}^2\right)\right]}, & |A| \le A_0 \\ 0, & |A| > A_0 \end{cases}$$

BMSE Example

$$p(A|\mathbf{x}) = \begin{cases} \dfrac{\dfrac{1}{2A_0(2\pi\sigma^2)^{N/2}} \exp\left[-\dfrac{1}{2\sigma^2}\left(N(A-\bar{x})^2\right)\right]}{\int_{-A_0}^{A_0} \dfrac{1}{2A_0(2\pi\sigma^2)^{N/2}} \exp\left[-\dfrac{1}{2\sigma^2} N(A-\bar{x})^2\right] dA}, & |A| \le A_0 \\ 0, & |A| > A_0 \end{cases}$$

$$= \begin{cases} \dfrac{1}{c\sqrt{2\pi\dfrac{\sigma^2}{N}}} \exp\left[-\dfrac{1}{2\dfrac{\sigma^2}{N}}(A-\bar{x})^2\right], & |A| \le A_0 \\ 0, & |A| > A_0 \end{cases}$$

where

$$c = \int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}} N(A-\bar{x})^2\right] dA$$

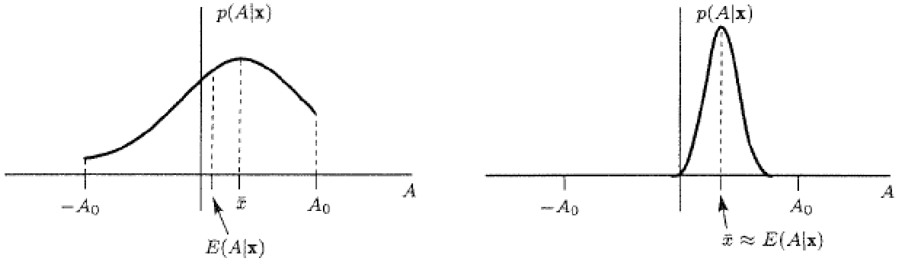
BMSE Example

The MMSE estimator

$$\hat{A} = E\left(A|\mathbf{x}\right) = \int_{-\infty}^{\infty} Ap\left(A|\mathbf{x}\right)\, dA$$

$$= \frac{\displaystyle\int_{-A_0}^{A_0} A \frac{1}{\sqrt{2\pi \frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}}\left(A-\overline{x}\right)^2\right] dA}{\displaystyle\int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi \frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}}\left(A-\overline{x}\right)^2\right] dA}$$

Cannot be evaluated in closed-form



(a) Short data record (b) Large data record

Remarks on BMSE

- As N increases, the MMSE estimator relies less and less on the prior knowledge and more on the data.
- Before observation, we assume a prior pdf $p(A)$. After observation, our state of knowledge about the parameter is summarized by the posterior pdf $p(A|\mathbf{x})$.
- The choice of a prior pdf is critical in Bayesian estimation. A wrong choice will result in a poor estimator.
- An optimal estimator is defined to be the one that minimizes the MSE when average over all realizations of θ and $\mathbf{x}$.

MMSE estimator: $\hat{\theta} = E(\theta|\mathbf{x}) = \int \theta p(\theta|\mathbf{x}) d\theta$

Special Average: Characteristic Function

Let $g\left(X\right)=e^{j\omega X}$

$$\Phi\left(\omega\right)\triangleq E\left[e^{j\omega X}\right]=\int_x f_X\left(x\right)e^{j\omega x}dx$$

$$f_X\left(x\right)=\frac{1}{2\pi}\int_v \Phi\left(\omega\right)e^{-j\omega x}dv$$

Note:

- This is Fourier transform of $f_X\left(x\right)$ if we have $e^{-j\omega X}$
- Sometimes it is more convenient to use the variable s in place of $j\omega$, the result becomes **moment generating function**.

Obtaining moments of a RV:

$$\frac{\partial\Phi\left(\omega\right)}{d\omega}=j\int_x xf_X\left(x\right)e^{jvx}dx$$

Set $\omega=0$: $\Rightarrow E\left[X\right]=\left(-j\right)\left.\frac{\partial\Phi\left(\omega\right)}{d\omega}\right|_{\omega=0}$

$$\Rightarrow E\left[X^n\right]=\left(-j\right)^n\left.\frac{\partial^n\Phi\left(\omega\right)}{d\omega^n}\right|_{\omega=0}$$

Chebyshev Inequality and the Law of Large Numbers (LLN)

Let X be a RV with mean μ_X and finite variance σ_X^2. Then for any $\delta > 0$,

$$\Pr\left(\left|X - \mu_X\right| \geq \delta\right) \leq \frac{\sigma_X^2}{\delta^2} \quad \text{(Chebyshev Inequality)}$$

Let $X_1, X_2, \ldots, X_N$ be i.i.d. (independent and identically distributed) RVs with mean μ_X and variance σ_X^2 each. Let the sample mean be

$$\hat{\mu}_X = \frac{1}{N}\sum_{i=1}^{N} X_i.$$

Then, for any fixed $\delta > 0$,

$$\lim_{N\to\infty} \Pr\left(\left|\mu_X - \hat{\mu}_X\right| \geq \delta\right) = 0. \quad \text{(LLN)}$$

Intuitively, this means the estimator, $\hat{\mu}_X$, will converge to μ_X in probability.

If the above limit equals 0, $\hat{\mu}_X$ is called a consistent estimator of μ_X.

Proof (Chebychev Inequality)

Given that $\sigma_X^2 = \sum_i (x_i - \mu_X)^2 f(x_i)$. Removing all the terms $|x_i - \mu_X| < \delta$. Then

$$\sum_i^* (x_i - \mu_X) f(x_i) \le \sigma_X^2,$$

where asterisk indicates the summation extends only over those i for which $|x_i - \mu_X| \ge \delta$.

Thus this new summation does not increase in value if we replace each $|x_i - \mu_X|$ by δ

so that $$\sum_i^* \delta^2 f(x_i) = \delta^2 \sum_i^* f(x_i) \le \sigma_X^2.$$

But $\sum_i^* f(x_i)$ is equal to $\Pr(|X - \mu_X| \ge \delta)$, hence

$$\delta^2 \Pr(|X - \mu_X| \ge \delta) \le \sigma_X^2. \qquad \square$$

Proof of LLN

Note that $E(\hat{\mu}_X) = \dfrac{E(X_1) + \cdots + E(X_N)}{N} = \dfrac{N\mu_X}{N} = \mu_X$.

Since $X_1, \ldots, X_N$ are independent, it follows that

$\mathrm{var}(X_1 + \cdots + X_N) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_N) = N\sigma_X^2$, then

$$\mathrm{var}(\hat{\mu}_X) = \mathrm{var}\left(\frac{X_1 + \cdots + X_n}{N}\right) = \frac{1}{N^2}\mathrm{var}(X_1 + \cdots + X_n) = \frac{1}{N^2}N\sigma_X^2 = \frac{\sigma_X^2}{N}.$$

So, from the Chebychev inequality, $\Pr\left(\left|\hat{\mu}_X - \mu_X\right| \geq \delta\right) \leq \dfrac{\sigma_X^2}{N\delta^2}$. Then as we take limit as $N \to \infty$ of the right hand side, it equals 0. □

Useful PDFs

- Discrete RVs
  - Binomial distribution
    - Related to chance experiments with two mutually exclusive outcomes with probability p and $1-p$
    - Model number of times event A has occurred in n trials (events are indep)
  - Poisson distribution
    - Related to chance experiment in which an event whose probability of occurrence in a very small time interval ΔT is $P=\alpha\Delta T$, where α is a constant
    - Model the probability of k events occurring in time T
    - Commonly used to model arrival time of packets in packet switching networks
- Continuous RVs
  - Normal (Gaussian) distribution
    - Commonly used to model large number of indep. random events when distribution of each event is unknown
    - Sum of large number of independent RVs converges to a Gaussian distribution
  - Rayleigh distribution
    - (see above)
  - Rician distribution
    - Commonly used to model distribution of power profile of wireless channel when direct line-of-sight (LOS) exists
    - $x = \text{sqrt}(x_1^2+x_2^2)$, where $x_1 \sim N(\mu_1,\sigma^2)$, $x_2 \sim N(\mu_2,\sigma^2)$ are indep. RV

Useful PDFs

- Continuous RVs
  - Chi-Squared (central and noncentral)
    - Commonly encounter in detector design

$$\chi_\nu^2 \text{ with } \nu \text{ degrees of freedom}$$

$$x = \sum_{i=1}^{\nu} x_i^2,\ x_i \sim N\left(0 \text{ or } \mu_i, 1\right) \text{ and indep.}$$

  - F-distribution (central and noncentral)
    - Commonly encounter in detector design

$$F \text{ PDF: ratio of 2 indep. } \chi_\nu^2 \text{ RVs}$$

$$x = \frac{x_1 / \nu_1}{x_2 / \nu_2},\ x_1 \sim \chi_{\nu_1}^2(\lambda),\ x_2 \sim \chi_{\nu_2}^2 \text{ and indep.}$$

$$\lambda = 0: \text{ central } F - \text{dist.}$$

Binomial Distribution

Consider repeated and independent trials of an experiment with two outcomes: success and failure. Probability of success equals p and failure equals $q \triangleq 1-p$. If we are interested in the number of successes and not in the order in which they occur, then the probability of exactly k successes in n repeated trials is given by

$$\Pr(K=k) \triangleq P_n(k) = \begin{cases} \binom{n}{k} p^k q^{n-k}, & \text{for } k = 0,1,\ldots,n \\ 0, & \text{otherwise} \end{cases},$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient, and K is an RV that equals to the number of successes.

Examples

Suppose we wish to obtain probability of k heads in n tosses of the coin and the probability of a head on a single toss equals p and that of a tail equals $q = 1 - p$. Possible sequence is

$$\underbrace{HH\cdots H}_{k \text{ heads}}\underbrace{TT\cdots T}_{n-k \text{ tails}}.$$

Under the assumption that the tosses are independent, the probability of this particular sequence is

$$\underbrace{p\cdot p\cdots p}_{k \text{ factors}}\underbrace{q\cdot q\cdots q}_{n-k \text{ tails}} = p^k q^{n-k}.$$

Since this is only the probability of this sequence. This sequence is one out of

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

possible sequences of having k heads in n tosses. Since all of these outcomes are mutually exclusive, the probability of exactly k heads in n tosses in any order

is $\binom{n}{k} p^k q^{n-k}$, $k = 0, 1, \ldots, n$

Example

A fair coin is tossed 6 times and head is called a success.

$\Rightarrow n = 6$ and $p = q = \dfrac{1}{2}$

(i) The probability that exactly two heads occur (i.e. $k = 2$) is

(ii) The probability of getting at least four heads (i.e. $k = 4, 5$ or 6) is

(iii) The probability of no heads (i.e. all failures):

(iv) The probability of at least 1 head:

Laplace Approximation to Binomial Distribution

Laplace approximation to binomial distribution

When $n \to \infty$, $|k - np| \leq \sqrt{npq}$

$$P_n(k) \approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq} \right)$$

Binomial Distribution (6 different parameters)

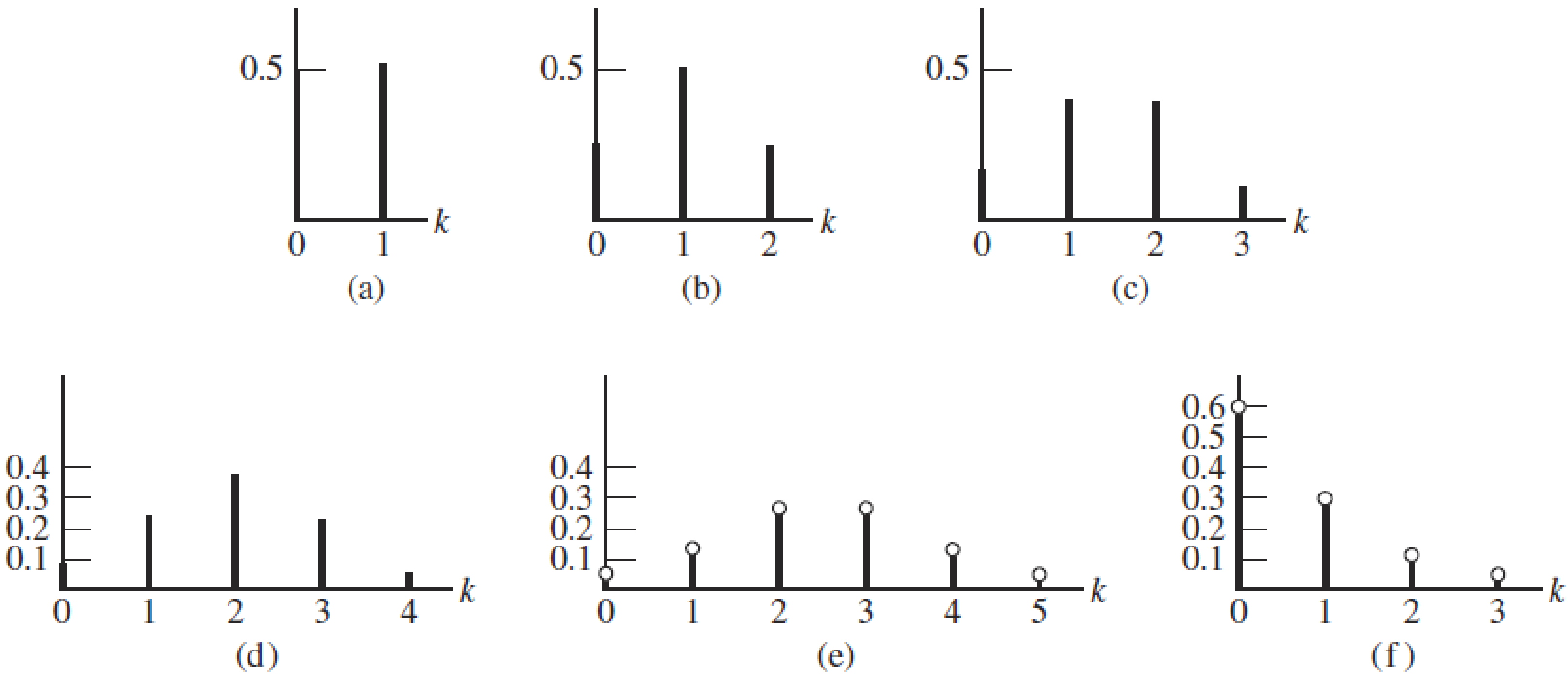


Figure 6.17
The binomial distribution with comparison to Laplace and Poisson approximations. (a) $n = 1, p = 0.5$. (b) $n = 2, p = 0.5$. (c) $n = 3, p = 0.5$. (d) $n = 4, p = 0.5$. (e) $n = 5, p = 0.5$. Circles are Laplace approximations. (f) $n = 5, p = \frac{1}{10}$. Circles are Poisson approximations.

Poisson Distribution

Consider a chance experiment in which an event whose probability of occurrence in a very small time ΔT is $P = \alpha \Delta T$, where α is a constant of proportionality. If successive occurrences are statistically independent, then the probability of k events in time T is

$$P_T(k) = \frac{(\alpha T)^k}{k!} e^{-\alpha T}, \quad k = 0, 1, 2, \ldots$$

- Can be used to model the number of telephone calls per minute
- Can be used to model the number of packets arriving at a router
- Can be used to approximate the binonomial distribution when n is large, and p is small, then $np \approx npq$

$$P_n(k) \approx \frac{(\bar{K})^k}{k!} e^{-\bar{K}},$$

$\bar{K} \triangleq E[K] = np$

Gaussian (Normal) Distribution

1 – dimensional:

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

where $\mu \triangleq E[X]$, $\sigma^2 \triangleq E\left[(X-\mu)^2\right]$

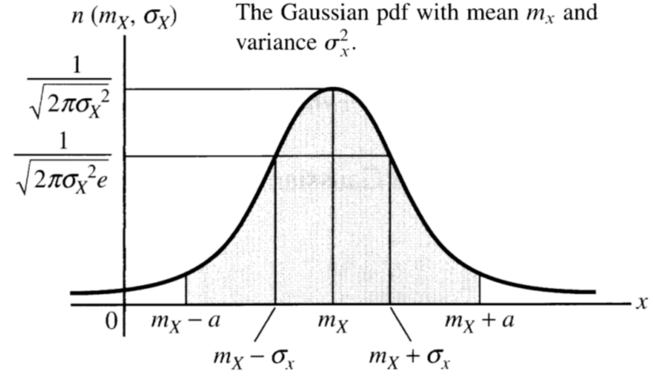


Figure 5.19
The Gaussian pdf with mean m_x and variance σ_x^2.

Joint CDFs and PDFs:

$$F_{XY}(x,y) = P(X \le x, Y \le y)$$

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

$$P(x_1 \le X \le x_2, y_1 \le Y \le y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{XY}(x,y)\,dxdy$$

Marginal distribution:

$$F_X(x) = F_{XY}(x,\infty) = F_{XY}(x, Y \le \infty)$$

$$F_Y(y) = F_{XY}(\infty, y) = F_{XY}(X \le \infty, y)$$

$$f_X(x) = \int_x f_{XY}(x,y)\,dy$$

2-D (Bivariate) Gaussian Distribution

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{\left[(x-\mu_x)/\sigma_x\right]^2-2\rho\left[(x-\mu_x)/\sigma_x\right]\left[(y-\mu_y)/\sigma_y\right]+\left[(y-\mu_y)/\sigma_y\right]^2}{2\left(1-\rho^2\right)}\right)$$

where

$$\mu_x=E[X],\quad \mu_y=E[Y],\quad \sigma_x^2=\mathrm{var}[X],\quad \sigma_y^2=\mathrm{var}[Y]$$

$$\rho=\frac{E\left[(X-\mu_x)E(Y-\mu_y)\right]}{\sigma_x\sigma_y}=\frac{Cov[X,Y]}{\sqrt{\sigma_x^2\sigma_y^2}}$$

2-D (Bivariate) Gaussian Distribution

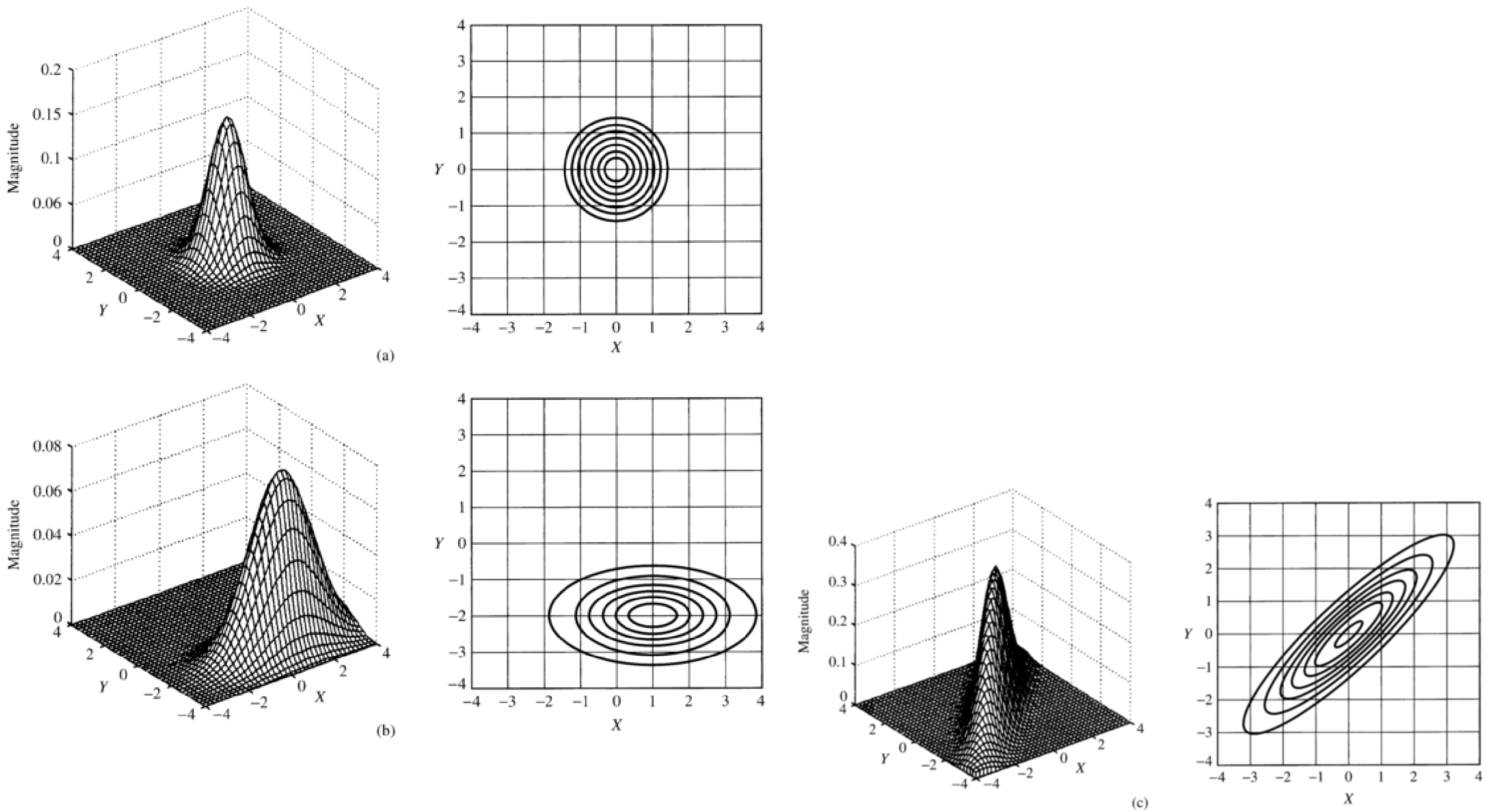


Figure 5.18

Bivariate Gaussian pdfs and corresponding contour plots. (a) $m_x = 0,\ m_y = 0,\ \sigma_x^2 = 1,\ \sigma_y^2 = 1$ and $\rho = 0$. (b) $m_x = 1,\ m_y = -2,\ \sigma_x^2 = 2,\ \sigma_y^2 = 1$, and $\rho = 0$.(c) $m_x = 0,\ m_y = 0,\ \sigma_x^2 = 1,\ \sigma_y^2 = 1$, and $\rho = 0.9$.

N-dimensional Gaussian Distribution

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}(\det \mathbf{C})^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})\right]$$

$$\boldsymbol{\mu}_{\mathbf{x}} \triangleq E[\mathbf{x}] = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_N) \end{bmatrix}$$

$$\mathbf{C} \triangleq E\left[(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})^T\right] \text{ (applied element-wise)}$$

Central Limit Theorem

Let $X_1, X_2, \ldots, X_N$ be indep. RVs with zero mean and variance $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$. Let $s_N^2 \triangleq \sigma_1^2 + \cdots + \sigma_N^2$. If for any fixed $\varepsilon > 0$, there exists a sufficient large N such that

$$\sigma_k^2 < \varepsilon s_N, \quad \text{for } k = 1, \ldots, N,$$

then the normalized RV

$$Z_N \triangleq \frac{X_1 + X_2 + \cdots + X_N}{s_N}$$

converges to the standard normal (Gaussian) PDF.

Q-Function

Gaussian Q-Function:

Normalized Normal distribution of $N\left(\mu_x, \sigma_x^2\right)$

Consider $P\left(\mu_x - a \le X \le \mu_x + a\right) = \int_{\mu_x - a}^{\mu_x + a} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{1}{2\sigma_x^2}\left(x-\mu_x\right)^2\right] dx$

(let $y = \frac{x-\mu_x}{\sigma_x}$) $= \int_{-a/\sigma_x}^{a/\sigma_x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$

$$= 2\int_{0}^{a/\sigma_x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

(since area under PDF=1) $= 1 - 2\int_{a/\sigma_x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$

$$= 1 - 2Q\left(\frac{a}{\sigma_x}\right)$$

where $Q(u) \triangleq \int_{u}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \approx \frac{1}{u\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right), \quad \text{for } u \gg 1$

has been computed numerically.

Gaussian PDF

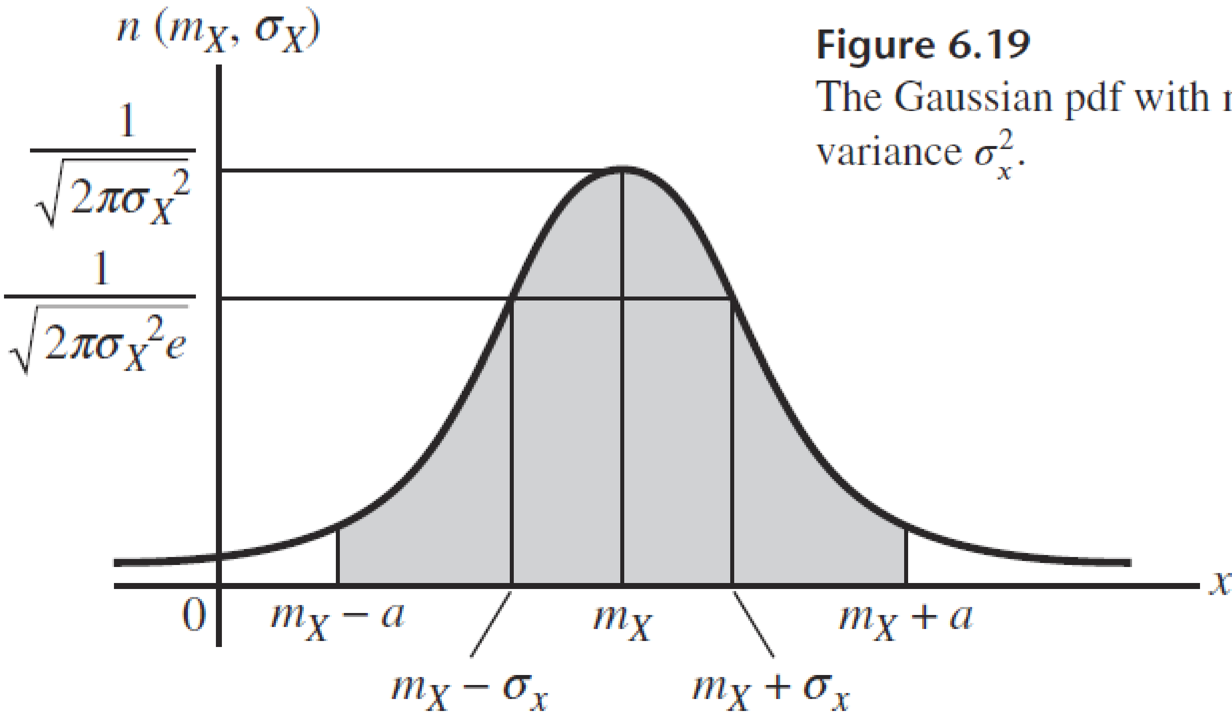


Figure 6.19
The Gaussian pdf with mean m_x and variance σ_x^2.

Normalized Distribution Function: $F(x)$ and $Q(x)$

Normalized cumulative distribution function: $\mu_x = 0,\ \sigma_x = 1$

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\xi^2/2} d\xi$$

$$F(-x) = 1 - F(x)$$

A related function: $F(x) = 1 - Q(x)$

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\xi^2/2} d\xi$$

$$Q(-x) = 1 - Q(x)$$

TABLE B-1
Values of $F(x)$ for $0 \leq x \leq 3.89$ in steps of 0.01

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	1.0000	1.0000	1.0000

Normalized cumaltive distribution function

$$F(x)$$

$$F(x) = 1 - Q(x)$$

Example

As the threshold changes, one error increases while the other decreases

$P(H_1;H_0)$ Probability of false alarm (P_{FA})

$P(H_1;H_1)=1-P(H_0;H_1)$ Probability of detection (P_D)

$$\begin{aligned}P_{FA} &= P(H_1;H_0)\\ &= \Pr\{x[0]>\gamma;H_0\}\\ &= \int_\gamma^\infty \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)dt\\ &= Q(\gamma)\end{aligned}$$

Suppose $P_{FA}=10^{-3}$, then $\gamma=3$. Then

$$\begin{aligned}P_D &= P(H_1;H_1)\\ &= \Pr\{x[0]>\gamma;H_1\}\\ &= \int_\gamma^\infty \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(t-1)^2\right)dt\\ &= Q(\gamma-1)=Q(2)=0.023\end{aligned}$$

$$F(2)=1-Q(2)=1-0.977$$

$$\Rightarrow Q(2)=0.023$$

Error Function

Error function:

$$erf(u) \triangleq \frac{2}{\sqrt{\pi}} \int_0^u \exp(-y^2)\, dy$$

Complementary error function:

$$erfc(u) \triangleq 1 - erf(u) = \frac{2}{\sqrt{\pi}} \int_u^\infty \exp(-y^2)\, dy$$

Note:

$$erf\left(\frac{u}{\sqrt{2}}\right) = \frac{1}{\sqrt{\pi}} \int_{-\frac{u}{\sqrt{2}}}^{\frac{u}{\sqrt{2}}} \exp(-y^2)\, dy$$

Thus

$$\frac{1}{2} erfc\left(\frac{u}{\sqrt{2}}\right) = \frac{1}{2}\left(1 - erf\left(\frac{u}{\sqrt{2}}\right)\right)$$

$$= \frac{1}{2}\left(1 - \frac{1}{\sqrt{\pi}} \int_{-\frac{u}{\sqrt{2}}}^{\frac{u}{\sqrt{2}}} \exp(-y^2)\, dy\right)$$

$$= \frac{1}{2}\left(\frac{2}{\sqrt{2\pi}} \int_{\frac{u}{\sqrt{2}}}^{\infty} \exp(-y^2)\, dy\right)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{u}{\sqrt{2}}}^{\infty} \exp(-y^2)\, dy$$

(let $y = \frac{s}{\sqrt{2}}$) $$= \frac{1}{\sqrt{2\pi}} \int_u^\infty \exp\left(-\frac{s^2}{2}\right) ds$$

$$= Q(u)$$

Example: Nonlinear MMSE Filter

Suppose we want to recover the transmit signal $x[n]$ from the received signal $y[n]$ the MMSE filter $g(Y)$.

Cost function to consider $E_{X|Y}\left[\left|x[n]-g(y[n])\right|^2 \middle| y[n]\right] \Rightarrow \min_{g(y)} E_{X|Y}\left[\left|x[n]-g(y[n])\right|^2 \middle| y[n]\right]$

$$E_{X|Y}\left[\left|x[n]-g(y[n])\right|^2 \middle| y[n]\right] = E_Y\left[E_{X|Y}\left(\left|x[n]-g(y[n])\right|^2 \middle| Y=y[n]\right)\right]$$

$$= \int_y E_{X|Y}\left(\left|x[n]-g(y[n])\right|^2 \middle| y[n]\right) f_Y(y[n])\, dy$$

Note:

- $f_Y(y[n]) > 0$ and it's not a function of $g(y)$, it can be ignored in computing $g(y)$
- $E_{X|Y}\left(\left|x[n]-g(y[n])\right|^2 \middle| y[n]\right) \geq 0$ since we are taking expectation of a non-negative quantity
- Integral of non-negative quantity w.r.t. y does not affect solution, can be ignored.

$$\Rightarrow \frac{\partial}{\partial g(y)} E_{X|Y}\left(\left|x[n]-g(y[n])\right|^2 \middle| y[n]\right) = -2E_{X|Y}\left(\left|x[n]-g(y[n])\right| \middle| y[n]\right) = 0$$

$$\Rightarrow E_{X|Y}\left(\left|x[n]\right| \middle| y[n]\right) = E_{X|Y}\left(\left|g(y[n])\right| \middle| y[n]\right) = g(y[n])$$

$\therefore$ MMSE filter to recover $x[n]$ from $y[n]$: $g(y[n]) = E_{X|Y}\left(\left|x[n]\right| \middle| y[n]\right)$