CT111 Introduction to Communication Systems
Lecture 3: Probability and Estimation

Yash M. Vasavada

Associate Professor, DA-IICT, Gandhinagar

8th January 2019

Overview of Today's Talk

Overview of Today's Talk

1. Reading Material
2. A Review of Probability Theory
   - Random Variables (RVs)
     - Definition of RV
   - Characterization of RVs
3. Detection and Decision

Overview of Today's Talk

Reading Assignment

1. Stark and Woods (Probability and Random Processes with Applications to Signal Processing)
   - Chapter 1: Section 1.1 to 1.7
2. Proakis:
   - Chapter 2: Section 2.1
3. DJC MacKay:
   - Chapter 2: Sections 2.1 to 2.3
   - Chapter 3: Sections 3.1, 3.2, 3.4. Also do Exercises in Section 3.5.

Reading Pointers

Stark and Woods provide

- an enlightening comparison of different types of probabilities (intuition, ratio of favorable to total outcomes, relative frequency, etc.)
- paradoxes involved in probabilities
- and many example problems

Reading Pointers

DJC MacKay provides

- an alternate description of the meaning of probability
- a detailed treatment of forward and inverse probabilities, and Bayes' Theorem
- and many example problems

Why to Study Random Processes?

Random variables and processes let us talk about quantities and signals which are *not* known in advance:

- Data sent through the communication channel is best modeled as a random sequence
- Noise, interference and fading affecting this data transmission are all random processes
- Receiver performance is measured probabilistically.

Random Events

- Outcome of any random event can be viewed as a member of a set.
- This set is a set of all possible outcomes of this random event or random experiment.
  - → Roll a six-sided die:
    - ▷ Set of all possible outcomes: $S = \{1, 2, \ldots, 6\}$
  - → Toss a coin:
    - ▷ Set of all possible outcomes: $S = \{\text{Head, Tail}\}$
  - → Transmit one bit in a noisy channel:
    - ▷ Set of all possible outcomes: $S = \{\text{Bit is received correctly, Bit is received incorrectly}\}$

Random Events

- An event is a subset of set S.
  - → Roll a six-sided die:
    - ▷ An event: $A = \{1, 2\}$
  - → Toss a coin:
    - ▷ An event: head turns up $A = \{\text{Head}\}$
  - → Transmit one bit in a noisy channel:
    - ▷ An event: $A = \{\text{Bit is received correctly}\}$
- Set S of all possible outcomes is a certain event. Its probability is 1.
- Set $\emptyset$ is a null event. Its probability is zero.
- Subset A of set S denotes a probable event. Its probability is a variable between 0 and 1.

Axioms of Probability

- Probability $P(A)$ is a number which measures the likelihood of event A.
- Following are three axioms of probability:
  1. $P(A) \geq 0$ (i.e., no event has a probability less than zero).
  2. $P(A) \leq 1$, and $P(A) = 1$ only if $A = S$, i.e., if A is a certain event.
  3. Let A and B be two events such that $A \bigcap B = \emptyset$. In this case, $P\left(A \bigcup B\right) = P(A) + P(B)$ (i.e., probabilities of mutually exclusive events *add*).
- All other theorems of probability follow from these three axioms.

Rules of Probability

- Joint probability $P(A, B) = P\left(A \bigcap B\right)$ is the probability that both A and B occur
- Conditional probability $P(A|B) = \dfrac{P(A, B)}{P(B)}$ is the probability that A will occur given B has occurred
- Statistical independence: events A and B are statistically independent if $P(A, B) = P(A) \times P(B)$
  - $\rightarrow$ If A and B are statistically independent, $P(A|B) = P(A)$ and $P(B|A) = P(B)$

Random Variables

- A random variable (RV) $X(s)$ is a real-valued function of the underlying event space $s \in S$
- Typically we omit the notation s and just denote the RV as X
- A random variable may be:
  - $\rightarrow$ Discrete-valued with either finite range (e.g., $[0, 1]$) or infinite range
  - $\rightarrow$ Continuous-valued (e.g., range can be the set $\mathcal{R}$ of real numbers)
- A random variable is described by its name, its range and its distribution

Cumulative Distribution Function (CDF)

- Definition: $F_X(x) = F(x) = P(X \leq x)$
- Properties:
  1. $F(x)$ is monotonically nondecreasing
  2. $F(-\infty) = 0$
  3. $F(\infty) = 1$
  4. $P(a < X \leq b) = F(b) - F(a)$
- CDF completely defines the probability distribution of a RV
- Alternate specifications are called PDF (Probability Density Function - for continuous variables) or PMF (Probability Mass Function - for discrete variables)

Probability Density Function (PDF)

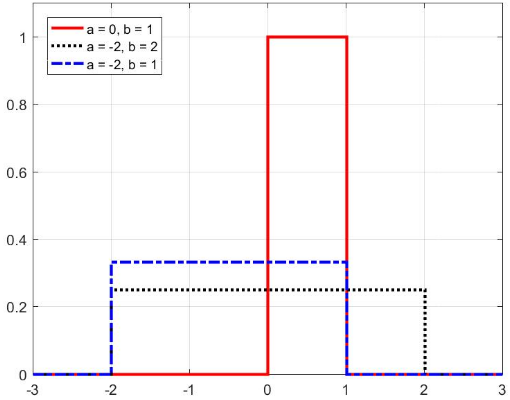
- Definition: $p_X(x) = \dfrac{dF(x)}{dx}$
- Interpretations: PDF measures
  - $\rightarrow$ how fast CDF is increasing
  - $\rightarrow$ how likely a random variable is to lie at a particular value
- Properties
  1. $p(x) \geq 0$
  2. $\displaystyle\int_{-\infty}^{\infty} p(x)dx = 1$
  3. $\displaystyle P(a < X \leq b) = \int_{a}^{b} p(x)dx$

Expected Values

- Sometimes the PDF is cumbersome to specify, or it may not be known
- Expected values are shorthand ways of describing the behavior of RVs
- Most important examples are:
  - → Mean: $E(x) = m_x = \int_{-\infty}^{\infty} x\, p(x)\, dx$
  - → Variance: $E\left((x - m_x)^2\right) = \int_{-\infty}^{\infty} (x - m_x)^2\ p(x)\, dx$
- Expectation operator works with any function $Y = g(X)$.
  - → $E(Y) = E\left(g(X)\right) = \int_{-\infty}^{\infty} g(x)\, p(x)\, dx$

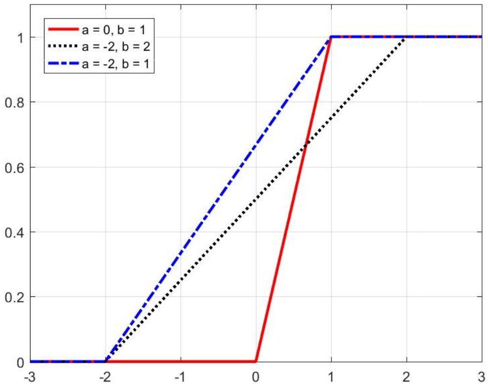
Example RVs
Uniform PDF

- $p(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{else} \end{cases}$



Example PDFs
Uniform CDF

- $F(x) = \begin{cases} 0, & x \leq 0 \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x \geq b \end{cases}$



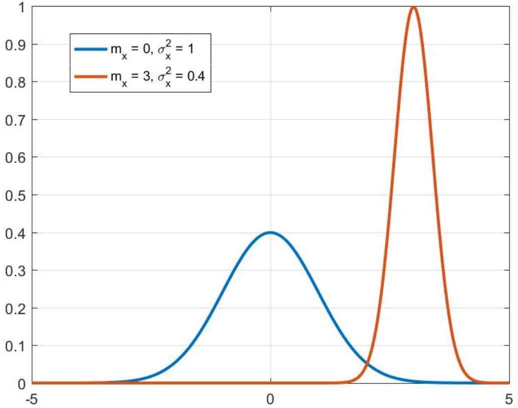
Example PDFs
Uniform RV

- Mean: $m_x = \int_a^b x\, p(x)\, dx = \frac{1}{b-a} \int_a^b x\, dx = \frac{a+b}{2}$
- Variance: $\sigma_x^2 = \int_a^b (x - m_x)^2\, p(x)\, dx = \frac{(b-a)^2}{12}$
- Probability:

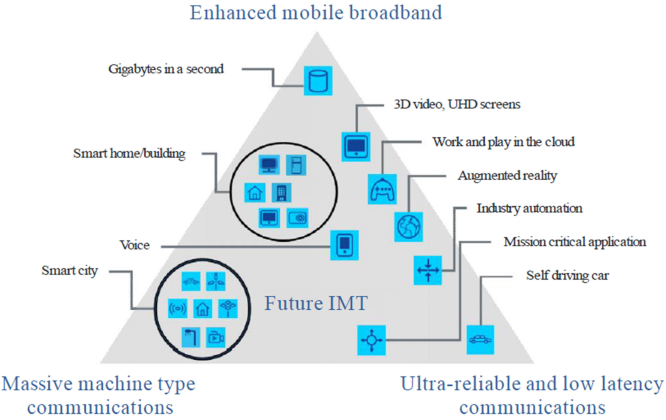
$$P(a_1 \le x < b_1) = \int_{a_1}^{b_1} p(x)\, dx = \frac{b_1 - a_1}{b - a}, \quad a < a_1, b_1 < b$$

Example RVs
Gaussian PDF

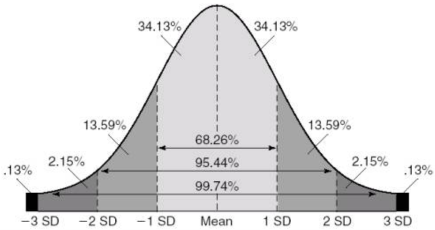
- $p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x-m_x)^2}{2\sigma_x^2}\right)$



Example RVs
Gaussian CDF



Example RVs
Gaussian PDF



z-scores	−3	−2	−1	0	1	2	3
Percentile ranks	0.1	2	16	50	84	98	99.9
IQ	55	70	85	100	115	130	145
SAT	200	300	400	500	600	700	800

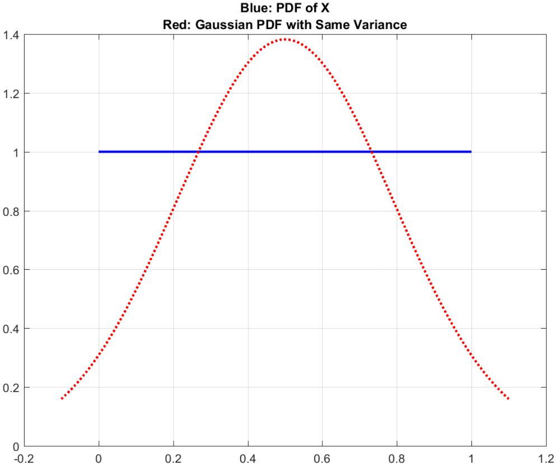
FIGURE 15.8 Percentile ranks and standard scores in relation to the normal curve.

http://turtleinvestor888.blogspot.in/2012_01_01_archive.html

Example RVs
Central Limit Theorem

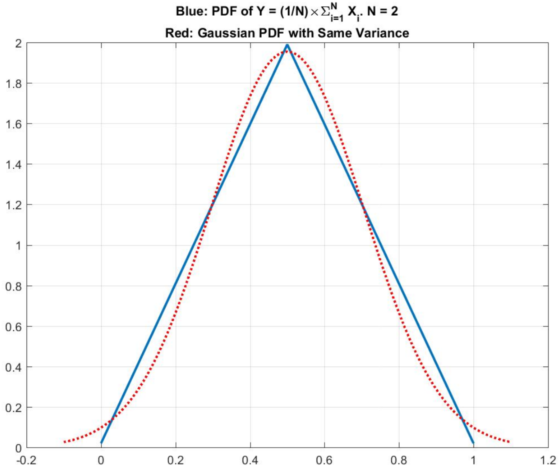
- Let $X_1, X_2, \ldots, X_N$ be N independent RVs with identical PDFs
- Let $Y = \sum_{i=1}^{N} X_i$
- A theorem of probability theory called Central Limit Theorem or CLT: as $N \to \infty$, distribution of Y tends to a Gaussian distribution
  - $\rightarrow$ In practice, $N = 10$ is sufficient to see the tendency of Y to follow the Gaussian PDF
- Importance of CLT:
  - $\rightarrow$ Thermal noise results from random movements of many electrons, and it is well modeled by the Gaussian PDF
  - $\rightarrow$ Interference from many identically distributed interferers in a CDMA system tends toward the Gauassian PDF

Example RVs
Central Limit Theorem: A Uniform Distribution



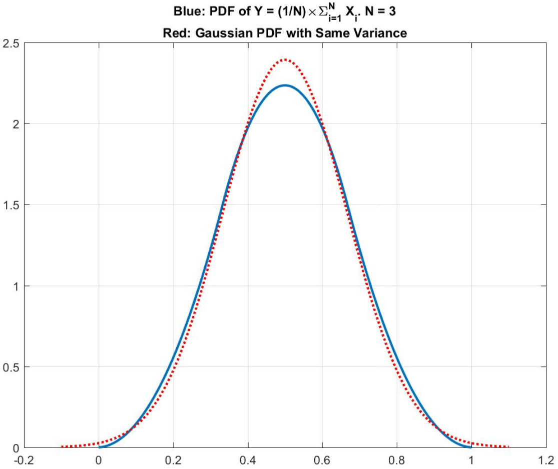
Example RVs

Central Limit Theorem. Average of $N = 2$ identically distributed Uniform RVs



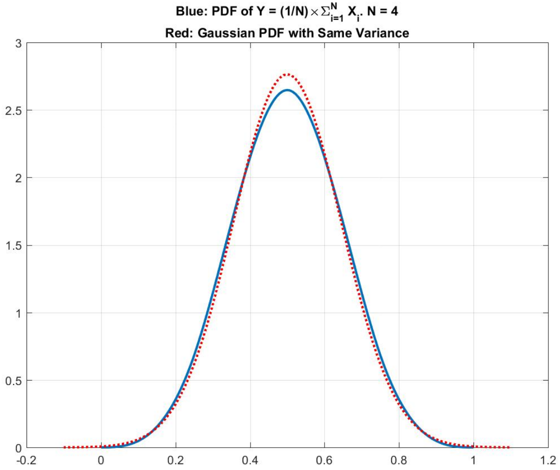
Example RVs

Central Limit Theorem. Average of $N = 3$ identically distributed Uniform RVs



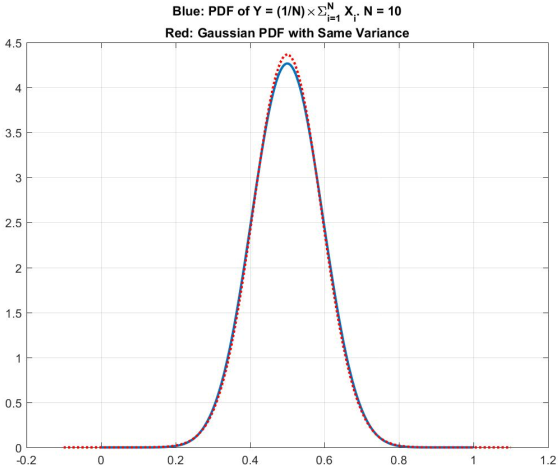
Example RVs

Central Limit Theorem. Average of $N = 4$ identically distributed Uniform RVs



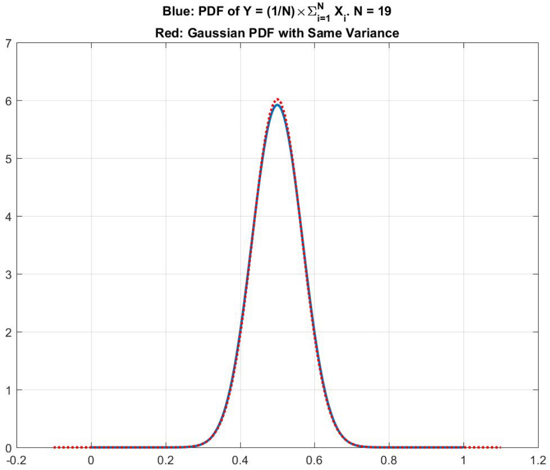
Example RVs

Central Limit Theorem. Average of $N = 10$ identically distributed Uniform RVs



Example RVs

Central Limit Theorem. Average of $N = 19$ identically distributed Uniform RVs



Example RVs
Gaussian PDF

- An application of Gaussian PDFs: signal level at the output of a digital communications receiver can often be given as $r = s + n$, where
  - $\rightarrow$ r is the received signal level,
  - $\rightarrow$ $s = -a$ is the transmitted signal level, and
  - $\rightarrow$ n is the Gaussian noise with mean 0 and variance σ_n^2
- Probability that the signal level $-a$ can be mistaken by the receiver as the signal level $+a$ is given as:

$$P(r > 0) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x+a)^2}{2\sigma_n^2}\right) dx = Q\left(\frac{a}{\sigma_n}\right)$$

- Definition: $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{v^2}{2}\right) dv$

Probability Mass Functions or PMFs

- For discrete random variables, the concept analogous to probability density function or PDF is PMF; $P(X = x) = p(x)$
- Properties are analogous to those of the PDFs:
    1. $p(x) \geq 0$
    2. $\sum_{X} p(x) = 1$
    3. $P(a < X \leq b) = \sum_{x=a}^{b} p(x)$

Example PMFs
Binary Distribution

- Outcome of the toss of a fair coin: $p(x) = \begin{cases} 1/2, & x = 0(\text{head}), \\ 1/2, & x = 1(\text{tail}) \end{cases}$
  - → Mean: $m_x = \sum_x x\, p(x) = 0 \times 1/2 + 1 \times 1/2 = 1/2$
  - → Variance: $\sigma_x^2 = 1/4$
- If X_1 and X_2 are independent binary random variables, $P_{X_1X_2}(x_1 = 0, x_2 = 0) = P_{X_1}(x_1 = 0)P_{X_2}(x_2 = 0)$

Example PMFs
Binomial Distribution

- Let $Y = \sum_{i=1}^{n} X_i$, where $\{X_i\}\,; i = 1, \ldots, N$ are independent binary RVs with $p(x) = \begin{cases} 1-p, & x = 0(\text{head}), \\ p, & x = 1(\text{tail}) \end{cases}$
- In this case, RV Y follows the Binomial Distribution given as $P_Y(y) = \binom{N}{y} p^y (1-p)^{N-y}$
- Mean $m_y = N \times p$
- Variance $\sigma_y^2 = N \times p \times (1-p)$

Estimation and Decision

- Cause x:
  - $\rightarrow$ $x = 1$: patient has (some) disease
  - $\rightarrow$ $x = 0$: patient does not have the disease
- Effect y:
  - $\rightarrow$ $y = 1$: the test that the patient takes is positive
  - $\rightarrow$ $y = 0$: test result is negative

Probability, Estimation and Decision

- Likelihood function of the test, i.e., $p(y|x)$, is known
- Marginal distribution of the cause x is also known

$p(y\|x)$	$y = 1$	$y = 0$	$p(x)$
$x = 1$	$p(y = 1\|x = 1) = 0.95$	$p(y = 0\|x = 1) = 0.05$	0.01
$x = 0$	$p(y = 1\|x = 0) = 0.05$	$p(y = 0\|x = 0) = 0.95$	0.99

Probability, Estimation and Decision

- Evidence: a patient takes the test and the result comes out positive, i.e., $y = 1$
- Question to be answered: what are the chances that the patient has the disease?
  - $\rightarrow$ Requires evaluation of $p(x = 1|y = 1)$.

Probability, Estimation and Decision

- From Bayes' rule

$$p(x=1|y=1) = \frac{p(y=1|x=1)p(x=1)}{p(y=1)}$$
$$= \alpha \times 0.95 \times 0.01 = 0.0095 \times \alpha$$

here, $1/p(y=1) = \alpha$.

- Similarly,

$$p(x=0|y=1) = \frac{p(y=0|x=1)p(x=0)}{p(y=1)}$$
$$= \alpha \times 0.05 \times 0.99 = 0.0495 \times \alpha$$

Probability, Estimation and Decision

- Even when $y = 1$, the only two possibilities for x are still 0 and 1
- Therefore, the probabilities $p(x = 1|y = 1)$ and $p(x = 0|y = 1)$ must sum to 1, which implies that $\alpha = \dfrac{1}{0.0095 + 0.0495}$.
- Substituting this value of α, we obtain that $p(x = 1|y = 1) = 0.16$ and $p(x = 0|y = 1) = 0.84$.
- Thus, even though the test came out positive, the chance of the disease is only 16%.

Bayes Rule

It's Everywhere!

- Toothache and Cavity:
  - $\rightarrow$ $x = 1$: there is a cavity in the teeth; $x = 0$: teeth are clean
  - $\rightarrow$ $y = 1$: patient has toothache; $y = 0$: no toothache
  - $\rightarrow$ $p(y = 1|x = 1) = 0.95, p(y = 1|x = 0) = 0.05, p(x = 1) = 0.01.$
  - $\rightarrow$ Does the patient have cavity if he complains of toothache?

Bayes Rule

It's Everywhere!

- Burglary and Car Alarm:
  - → $x = 1$: burglars are attempting to steal your car; $x = 0$: car is safe
  - → $y = 1$: car alarm goes off; $y = 0$: alarm is silent
  - → $p(y = 1|x = 1) = 0.95, p(y = 1|x = 0) = 0.05, p(x = 1) = 0.01.$
  - → Is my car being stolen if I hear the car alarm?

Bayes Rule
It's Everywhere!

- What color I am seeing of a dress in a dark room?
  - → $x = 1$: actual color is brown; $x = 0$: color is gray
  - → $y = 1$: color seems brown to me; $y = 0$: color seems gray to me
  - → $p(y = 1|x = 1) = 0.95, p(y = 1|x = 0) = 0.05, p(x = 1) = 0.01.$
  - → Is the color of the dress actually brown if it seems brown to me?

Bayes Rule

It's Everywhere!

- Suppose you're witness to a night-time hit and run accident involving a taxi in Gandhinagar. Of all taxis in Gandhinagar, 1% are light gray, and the rest are white colored. You swear, under oath, that the taxi was gray. Extensive testing has earlier shown that under dim light conditions, discrimination between the light gray and white is 95% reliable.
  - → $x = 1$: actual color is light gray; $x = 0$: color is white
  - → $y = 1$: color seems gray to me; $y = 0$: color seems white to me
  - → $p(y = 1|x = 1) = 0.95, p(y = 1|x = 0) = 0.05, p(x = 1) = 0.01.$
  - → I said under oath that the taxi was gray colored, however, was it?

Bayes Rule

It's Everywhere!

- Suppose you're a security guard at an underground nuclear facility. Only observation of the outside world you have is when you see the Director coming in. There is a 1% that it rains on any given day. You see the Director coming with an umbrella.
  - → $x = 1$: it's raining; $x = 0$: it's dry outside
  - → $y = 1$: Director came in with umbrella; $y = 0$: Director is without the umbrella
  - → $p(y = 1|x = 1) = 0.95, p(y = 1|x = 0) = 0.05, p(x = 1) = 0.01.$
  - → If I see the Director with the umbrella, is it raining outside?

Bayes Rule
It's Everywhere!

- In all the examples we have seen, both x and y are binary. In reality, they are quite frequently nonbinary. However, the same mathematics of Bayesian Inference still holds.

Bayes Rule
The Main Formula

- Let x_i be the cause and y_j be the effect.
  - $\rightarrow$ $i \in \{1, 2, \ldots, M\}$ and $j \in \{1, 2, \ldots, N\}$
  - $\rightarrow$ i.e., the cause can be one of M different causes, and the observed effect can be one of N different effects
- It is often the case that the likelihood function $p(y_j|x_i)$, i.e., the likelihood of the evidence given the cause, is available.
- However, we are oftentimes interested in knowing or *infering*, given the observation y_j that we have, the most likely x_i that could have caused it.
- Since $p(y_j, x_i) = p(y_j|x_i) \times p(x_i) = p(x_i|y_j) \times p(y_j)$, we have:

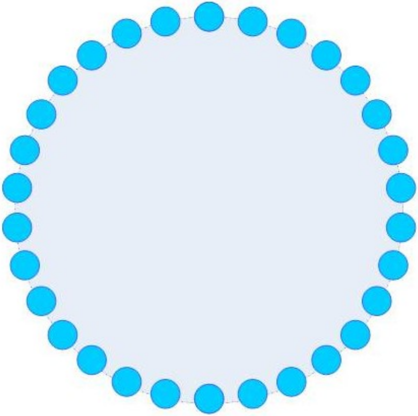
Theorem (Bayes')

$$p(x_i|y_j) = \frac{p(y_j|x_i) \times p(x_i)}{p(y_j)}$$

$$Posterior = \frac{Likelihood \times Prior}{Evidence}$$

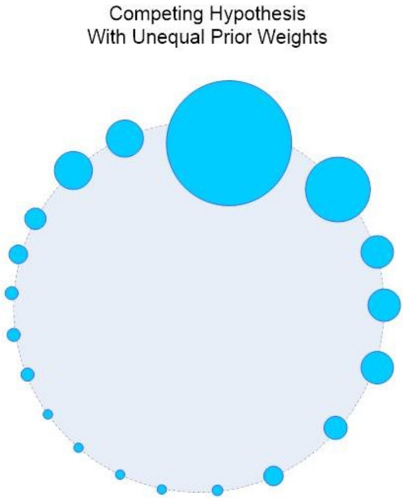
Bayesian Estimation Process
How to Think about It Intuitively

Competing Hypothesis
With Equal Prior Weights

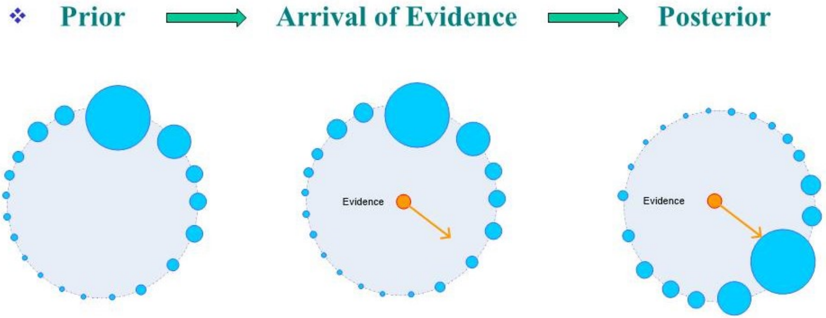


Bayesian Estimation Process

How to Think about It Intuitively



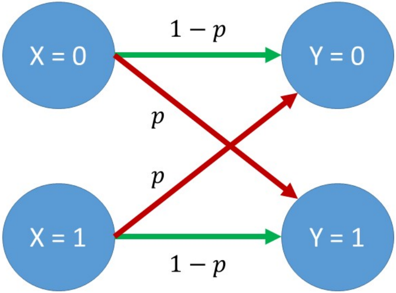
Bayesian Estimation Process
How to Think about It Intuitively



Binary Symmetric Channel or BSC

- Consider a simple channel shown on the next page for modeling the transmission of binary (zeros and ones) data
- Known as binary symmetric channel or BSC
  - ▷ Binary: this channel model does not work if the transmitted signal is nonbinary
  - ▷ Symmetric: the error affecting bit 0 has the same probability p (known as the cross-over probability of the BSC) as the error affecting bit 1; also, if an error occurs on transmitted bit 0, the receiver sees bit 1, and vice versa
  - ▷ Often parametrized by the probability p, and written as BSC(p)
- Notations: x denotes the binary random variable (RV) corresponding to the input and y is the RV corresponding to the output

Binary Symmetric Channel or BSC



Binary Symmetric Channel or BSC(p)

- Just like the binary information source x, the BSC channel itself can be thought of as a source of a binary digits
  - ▷ Bit 1 of this BSC string specifies the location where the error has occurred
  - ▷ Bit 0 specifies the locations that are error free
- Let the transmitted binary sequence be represented as a vector $\mathbf{x} = [x_0, x_1, \ldots, x_N]^T$
  - → A general notational remark: $[x_0, x_1, \ldots, x_N]$ denotes a *horizontal* row vector with dimensions $1 \times N$. The vectors are represented as column vectors by default. The superscript T denotes the *transposition* of vector that turns a $1 \times N$ row vector into an $N \times 1$ column vector
- the BSC noise string be represented as $\mathbf{e} = [e_0, e_1, \ldots, e_N]^T$,
- the received binary sequence $\mathbf{y} = [y_0, y_1, \ldots, y_N]^T$ can be written in the following two ways:
  - ▷ exclusive-OR of $\mathbf{x}$ and $\mathbf{e}$: $\mathbf{y} = \mathbf{x} \oplus \mathbf{e}$
  - ▷ sum (modulo-2) of $\mathbf{x}$ and $\mathbf{e}$: $\mathbf{y} = (\mathbf{x} + \mathbf{e})\ mod - 2$

Model of Channel Induced Bit Errors

for Binary Symmetric Channel

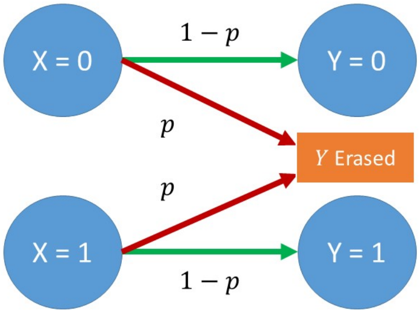
- Denoting the transmitted binary sequence as $\mathbf{x}$, and the BSC noise string as $\mathbf{e}$, the received binary sequence $\mathbf{y}$ can be written in the following two ways:
  - exclusive-OR of $\mathbf{x}$ and $\mathbf{e}$: $\mathbf{y} = \mathbf{x} \oplus \mathbf{e}$
  - sum (modulo-2) of $\mathbf{x}$ and $\mathbf{e}$: $\mathbf{y} = (\mathbf{x} + \mathbf{e})\ mod - 2$
- An example:
  - $x = 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 1$
  - $e = 0,\ 0,\ 1,\ 0,\ 1,\ 0,\ 0,\ 0$
  - $y = 0,\ 1,\ 1,\ 0,\ 0,\ 0,\ 0,\ 1$

Notice that where $\mathbf{e}$ is 1, the corresponding bit of $\mathbf{x}$ gets flipped in $\mathbf{y}$

Binary Erasure Channel or BEC

- An alternative channel model is where the binary (zeros and ones) information is received either correctly or it gets erased
- Known as binary erasure channel or BEC, and written as BEC(p)
- The channel can again be thought of as a vector **e** of binary digits, except that where **e** is 1, the corresponding bit of **x** gets *erased* in **y**

Binary Erasure Channel or BEC



An Application of Bayes' Theorem for Communication Engineering

- For BSC(p) and BEC(p), determine conditional probabilities of x given that the observed value of y is 0, i.e., $p(x = 0 \mid y = 0)$ and $p(x = 1 \mid y = 0)$. Determine the odds $o(y = 0) = \dfrac{p(x = 0 \mid y = 0)}{p(x = 1 \mid y = 0)}$
- Repeat for the case when y is observed to be 1.
- At the receiver, you observe a value of y, what will you decide regarding x?