ECE531: Principles of Detection and Estimation
Course Introduction

D. Richard Brown III

WPI

22-January-2009

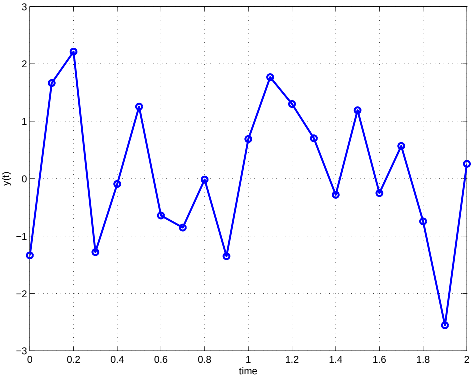
Lecture 1 Major Topics

1. Web page.
2. Syllabus and textbook.
3. Academic honesty policy.
4. Students with disabilities statement.
5. Course introduction.
6. Review of essential probability concepts.

Some Notation

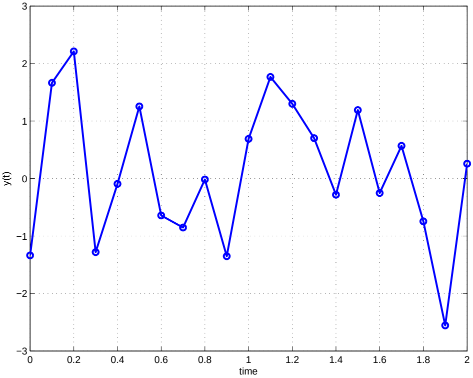
- A set with discrete elements: $\mathcal{S} = \{-1, \pi, 6\}$.
- The cardinality of a set: $|\mathcal{S}| = 3$.
- The set of all integers: $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$.
- The set of all real numbers: $\mathbb{R} = (-\infty, \infty)$.
- Intervals on the real line: $[-3, 1]$, $(0, 1]$, $(-1, 1)$, $[10, \infty)$.
- Multidimensional sets:
  - $\{a, b, c\}^2$ is shorthand for the set $\{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$.
  - $\mathbb{R}^2$ is the two-dimensional real plane.
  - $\mathbb{R}^3$ is the three-dimensional real volume.
- An element of a set: $s \in \mathcal{S}$.
- A subset: $\mathcal{W} \subseteq \mathcal{S}$.
- The probability of an event A: $\mathrm{Prob}[A] \in [0, 1]$.
- The joint probability of events A and B: $\mathrm{Prob}[A, B] \in [0, 1]$.
- The probability of event A conditioned on event B: $\mathrm{Prob}[A \mid B] \in [0, 1]$.

Typical Detection Problems



- Is this a sine wave plus noise, or just noise?
- Is the frequency of the sine wave 1Hz or 2Hz?
- Detection is about making smart choices (and the consequences).

Typical Estimation Problems



- What is the frequency, phase, and/or amplitude of the sine wave?
- What is the mean and/or variance of the noise?
- Estimation is about "guessing" values (and the consequences).

Joint Estimation and Detection

Suppose we have a binary communication system with an intersymbol interference channel. M symbols are sent through the channel and we observe

$$y_k = \sum_{\ell=0}^{L-1} h_\ell s_{k-\ell} + w_k$$

for $k \in \{0, \ldots, L+M-2\}$ where

- Unknown binary symbols $[s_0, \ldots, s_{M-1}] \in \{-1, +1\}^M$
- Unknown discrete-time impulse response of channel $[h_0, \ldots, h_{L-1}] \in \mathbb{R}^L$
- Unknown noise $[w_0, \ldots, w_{L+M-2}] \in \mathbb{R}^{L+M-1}$

In some scenarios, we may want know the bits that were sent and the channel coefficients. This is a joint estimation and detection problem. Why?

Consequences

To develop optimal decision rules or estimators, we need to quantify the **consequences** of incorrect decisions or inaccurate estimates.

Simple Example

It is not known if a coin is fair (HT) or double headed (HH). We are given one observation of the coin flip. Based on this observation, how do you decide if the coin is HT or HH?

Observation	Rule 1	Rule 2	Rule 3	Rule 4
H	HH	HT	HH	HT
T	HH	HT	HT	HH

Suppose you have to pay $100 if you are wrong. Which decision rule is "optimum"?

Rule 1: Always decide HH

Note that the observation is ignored here.

- If the coin is HT (fair), the decision was wrong and you must pay \$100.
- If the coin is HH (double headed), the decision was right and you pay nothing.

The **maximum cost** (between HH or HT) for Rule 1 is \$100.
The **average cost** for Rule 1 is

$$\bar{C}_1 = \mathrm{Prob}[\mathsf{HT}] \cdot \$100 + \mathrm{Prob}[\mathsf{HH}] \cdot \$0$$

where $\mathrm{Prob}[\mathsf{HT}]$ and $\mathrm{Prob}[\mathsf{HH}]$ are the prior probabilities (the probability before any observations) on the coin being fair or double headed, respectively.
For purposes of illustration, lets assume $\mathrm{Prob}[\mathsf{HT}] = \mathrm{Prob}[\mathsf{HH}] = 0.5$ so that $\bar{C}_1 = \$50$.

Rule 2: Always decide HT

Again, the observation is being ignored. Same analysis as for Rule 1...

- If the coin is HT (fair), the decision was right and you pay nothing.
- If the coin is HH (double headed), the decision was wrong and you must pay \$100.

The **maximum cost** for Rule 2 is \$100.
The **average cost** for Rule 2 is

$$\bar{C}_2 = \mathrm{Prob}[\mathsf{HT}] \cdot \$0 + \mathrm{Prob}[\mathsf{HH}] \cdot \$100$$

If $\mathrm{Prob}[\mathsf{HT}] = \mathrm{Prob}[\mathsf{HH}] = 0.5$, then $\bar{C}_2 = \$50$.

Rule 3: Decide HH if H observed, HT if T observed

- If the coin is HT (fair), there is a 50% chance the observation will be H and you will decide HH. This will cost you \$100. There is also a 50% chance that the observation will be T and you will decide HT. In this case, you made the correct decision and pay nothing.

$$C_{\mathsf{HT}} = \mathrm{Prob}[\mathsf{H}|\mathsf{HT}] \cdot \$100 + \mathrm{Prob}[\mathsf{T}|\mathsf{HT}] \cdot \$0 = \$50$$

- If the coin is HH (double headed), what is our cost? \$0

The **maximum cost** for Rule 3 is \$50.
The **average cost** for Rule 3 is

$$\bar{C}_3 = \mathrm{Prob}[\mathsf{HT}] \cdot \$50 + \mathrm{Prob}[\mathsf{HH}] \cdot \$0$$

If $\mathrm{Prob}[\mathsf{HT}] = \mathrm{Prob}[\mathsf{HH}] = 0.5$, then $\bar{C}_3 = \$25$.

Rule 4: Decide HT if H observed, HH if T observed

Obviously, this is a bad rule.

- If the coin is HT (fair), there is a 50% chance the observation will be T and you will decide HH. This will cost you \$100. There is also a 50% chance that the observation will be H and you will decide HT. In this case, you made the correct decision and pay nothing.

$$C_{\mathsf{HT}} = \mathrm{Prob}[\mathsf{T}|\mathsf{HT}] \cdot \$100 + \mathrm{Prob}[\mathsf{H}|\mathsf{HT}] \cdot \$0 = \$50$$

- If the coin is HH (double headed), what is our cost? \$100

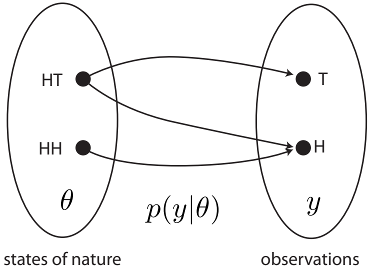
The **maximum cost** for Rule 4 is \$100.
The **average cost** for Rule 4 is

$$\bar{C}_3 = \mathrm{Prob}[\mathsf{HT}] \cdot \$50 + \mathrm{Prob}[\mathsf{HH}] \cdot \$100$$

If $\mathrm{Prob}[\mathsf{HT}] = \mathrm{Prob}[\mathsf{HH}] = 0.5$, then $\bar{C}_4 = \$75$.

Remarks

- The notion of **maximum cost** is the maximum over the possible "states of nature" (HH and HT in our example), but averaged over the probabilities of the observation.



- In our example, we could always lose \$100, irrespective of the decision rule. But the maximum cost of Rule 3 was \$50.
- Is Rule 3 optimal?

Probability Basics: Events

Let A be a possible (or impossible) outcome of a random experiment. We call A an "event" and $\text{Prob}[A] \in [0, 1]$ is the probability that A happens.

Examples:

- $A =$ tomorrow will be sunny in Worcester, $\text{Prob}[A] = 0.4$.
- $A =$ a 9 is rolled with two fair 6-sided dice, $\text{Prob}[A] = \frac{4}{36}$.
- $A =$ a 13 is rolled with two fair 6-sided dice, $\text{Prob}[A] = 0$.
- $A =$ an odd number is rolled with two fair 6-sided dice,

$$\text{Prob}[A] = \frac{2}{36} + \frac{4}{36} + \frac{6}{36} + \frac{4}{36} + \frac{2}{36} = \frac{1}{2}$$

- $A =$ any number but 9 is rolled with two fair 6-sided dice, $\text{Prob}[A] = 1 - \frac{4}{36} = \frac{32}{36}$

The last result used the fact that $\text{Prob}[A] + \text{Prob}[\bar{A}] = 1$, where $\bar{A}$ means "not event A" and $\text{Prob}[\bar{A}]$ is the probability that A doesn't happen.

Probability Basics: Random Variables

Definition

A random variable is a mapping from random experiments to real numbers.

Example: Let X be the Dow Jones average at the close on Friday.

We can easily relate events and random variables.

Example: What is the probability that $X \geq 8000$?

- X is the random variable. It can be anything on the interval $[0, \infty)$.
- The event is $A =$ "X is no less than 8000".

To answer these types of questions, we need to know the probabilistic distribution of the random variable X. Every random variable has a **cumulative distribution function** (CDF) defined as

$$F_X(x) := \text{Prob}[X \leq x]$$

for all $x \in \mathbb{R}$.

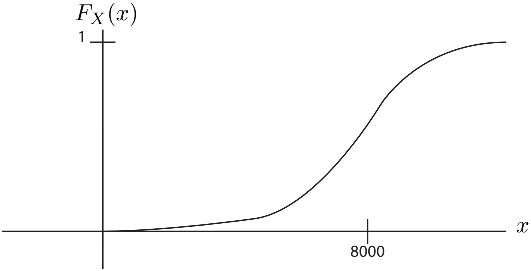
Probability Basics: Properties of the CDF

$$F_X(x) := \mathrm{Prob}[X \le x]$$

The following properties are true for any random variable X:

- $F_X(-\infty) = 0$.
- $F_X(\infty) = 1$.
- If $y > x$ then $F_X(y) \ge F_X(x)$.

Example: Let X be the Dow Jones average at the close on Friday.



Probability Basics: The PDF

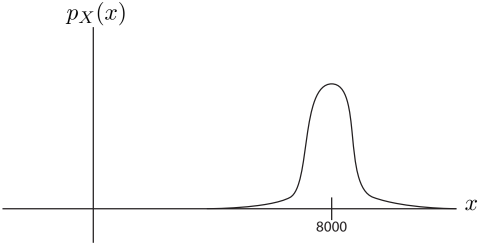
The **probability density function** (PDF) of the random variable X is

$$p_X(x) := \frac{d}{dx} F_X(x)$$

The following properties are true for any random variable X:

- $p_X(x) \geq 0$ for all x.
- $\int_{-\infty}^{\infty} p_X(x)\, dx = 1$.
- $\mathrm{Prob}[a < X \leq b] = \int_a^b p_X(x)\, dx = F_X(b) - F_X(a)$.

Example: Let X be the Dow Jones average at the close on Friday.



Probability Basics: Mean and Variance

Definition

The **mean** of the random variable X is defined as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x p_X(x)\, dx.$$

The mean is also called the **expectation**.

Definition

The **variance** of the random variable X is defined as

$$\mathrm{var}[X] = \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 p_X(x)\, dx.$$

Remark: The standard deviation of X is equal to $\mathrm{std}[X] = \sqrt{\mathrm{var}[X]}$.

Probability Basics: Properties of Mean and Variance

Assuming c is a known constant, it is not difficult to show the following properties of the mean:

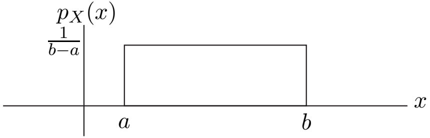
1. $\mathrm{E}[cX] = c\mathrm{E}[X]$ (by linearity)
2. $\mathrm{E}[X + c] = \mathrm{E}[X] + c$ (by linearity)
3. $\mathrm{E}[c] = c$

Assuming c is a known constant, it is not difficult to show the following properties of the variance:

1. $\mathrm{var}[cX] = c^2\mathrm{var}[X]$
2. $\mathrm{var}[X + c] = \mathrm{var}[X]$
3. $\mathrm{var}[c] = 0$

Uniform Random Variables

Uniform distribution: $X \sim \mathcal{U}(a,b)$ for $a \le b$.

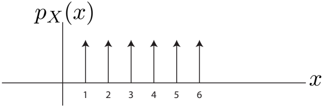


$$p_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

- Sketch the CDF.
- Suppose $X \sim \mathcal{U}(1,5)$.
  - What is $\mathrm{Prob}[X = 3]$?
  - What is $\mathrm{Prob}[X < 2]$?
  - What is $\mathrm{Prob}[X > 1]$?
  - What is $\mathrm{E}[X]$?
  - What is $\mathrm{var}[X]$?

Discrete Uniform Random Variables

Uniform distribution: $X \sim \mathcal{U}(\mathcal{S})$ where $\mathcal{S}$ is a finite set of discrete points on the real line. Each element in the set is equally likely. Example:



Given $\mathcal{S} = \{s_1, \dots, s_n\}$, then $\mathrm{Prob}[X = s_1] = \dots \mathrm{Prob}[X = s_n] = \frac{1}{n}$ and

$$p_X(x) = \frac{1}{n}\left(\delta(x - s_1) + \dots + \delta(x - s_n)\right)$$

- Sketch the CDF.
- What is $\mathrm{Prob}[X = 3]$?
- What is $\mathrm{Prob}[X < 2]$?
- What is $\mathrm{Prob}[X \leq 2]$?
- What is $\mathrm{E}[X]$?
- What is $\mathrm{var}[X]$?

Gaussian Random Variables

Gaussian distribution: $X \sim \mathcal{N}(\mu, \sigma^2)$ for any μ and σ.

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

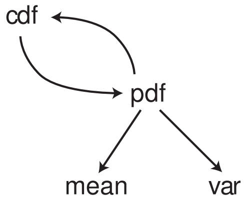
Remarks:

1. $\mathrm{E}[X] = \mu$.
2. $\mathrm{var}[X] = \sigma^2$.
3. Gaussian random variables are completely specified by their mean and variance.
4. Lots of things in the real world are Gaussian or approximately Gaussian distributed, e.g. exam scores, etc. The Central Limit Theorem explains why this is so.
5. Probability calculations for Gaussian random variables often require the use of erf and/or erfc functions (or the Q-function).

Final Remarks on Scalar Random Variables

1. The PDF and CDF completely describe a random variable.
   - If X and Y have the same PDF, then they have the same CDF, the same mean, and the same variance.
2. The mean and variance are only partial statistical descriptions of a random variable.
   - If X and Y have the same mean and/or variance, they might have the same PDF/CDF but not necessarily.

One Random Variable



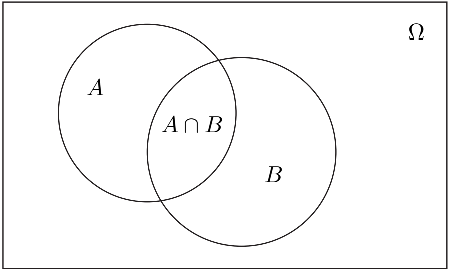
Joint Events

Suppose you have two events A and B. We can define a new event

$$C = \text{both } A \text{ and } B \text{ occur}$$

and we can write

$$\mathrm{Prob}[C] = \mathrm{Prob}[A \cap B] = \mathrm{Prob}[A, B]$$



Conditional Probability of Events

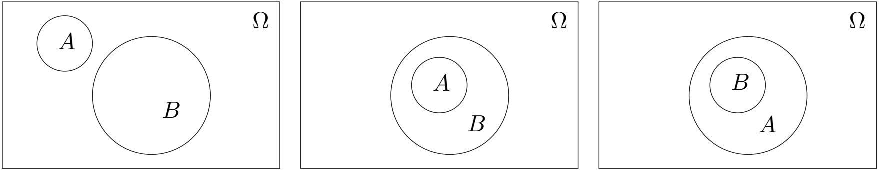
Suppose you have two events A and B. We can condition on the event B to write the probability

$$\mathrm{Prob}[A \,|\, B] = \text{the probability of event } A \text{ given event } B \text{ happened}$$

When $\mathrm{Prob}[B] > 0$, this conditional probability is defined as

$$\mathrm{Prob}[A \,|\, B] = \frac{\mathrm{Prob}[A \cap B]}{\mathrm{Prob}[B]} = \frac{\mathrm{Prob}[A, B]}{\mathrm{Prob}[B]}$$

Three special cases:



Conditional Probabilities: Our Earlier Example

It is not known if a coin is fair (HT) or double headed (HH). We can write the conditional probabilities of a one-flip observation as

$$\begin{aligned}
\mathrm{Prob}[\text{observe H} \,|\, \text{coin is HT}] &= 0.5 \\
\mathrm{Prob}[\text{observe T} \,|\, \text{coin is HT}] &= 0.5 \\
\mathrm{Prob}[\text{observe H} \,|\, \text{coin is HH}] &= 1 \\
\mathrm{Prob}[\text{observe T} \,|\, \text{coin is HH}] &= 0
\end{aligned}$$

Can you compute $\mathrm{Prob}[\text{coin is HH} \,|\, \text{observe H}]$?

We can write

$$\begin{aligned}
\mathrm{Prob}[\text{coin is HH} \,|\, \text{observe H}] &= \frac{\mathrm{Prob}[\text{coin is HH}, \text{observe H}]}{\mathrm{Prob}[\text{observe H}]} \\
&= \frac{\mathrm{Prob}[\text{observe H} \,|\, \text{coin is HH}]\mathrm{Prob}[\text{coin is HH}]}{\mathrm{Prob}[\text{observe H}]}
\end{aligned}$$

We are missing two things: $\mathrm{Prob}[\text{coin is HH}]$ and $\mathrm{Prob}[\text{observe H}]$....

Conditional Probabilities: Our Earlier Example

The term $\mathrm{Prob}[\text{coin is HH}]$ is called the **prior** probability, i.e. it is our belief that the coin is unfair **before we have any observations**. This is assumed to be given in some of the problems that we will be considering, so let's say for now that $\mathrm{Prob}[\text{coin is HH}] = 0.5$.

The term $\mathrm{Prob}[\text{observe H}]$ is the **unconditional probability** that we observe heads. Can we calculate this?

Theorem (Total Probability Theorem)

If the events $B_1, \dots, B_n$ are mutually exclusive, i.e. $\mathrm{Prob}[B_i, B_j] = 0$ for all $i \neq j$, and exhaustive, i.e. $\sum_{i=1}^n \mathrm{Prob}[B_i] = 1$, then

$$\mathrm{Prob}[A] = \sum_{i=1}^{n} \mathrm{Prob}[A \,|\, B_i] P[B_i].$$

So how can we use this result to compute $\mathrm{Prob}[\text{observe H}]$?

Independence of Events

Two events are independent if their joint probability is equal to the product of their individual probabilities, i.e

$$\mathrm{Prob}[A, B] = \mathrm{Prob}[A]\mathrm{Prob}[B]$$

Lots of events can be assumed to be independent. For example, suppose you flip a coin twice with $A =$"the first coin flip is heads", $B =$"the second coin flip is heads", and $C =$"both coin flips are heads".

- Are A and B independent?
- Are A and C independent?

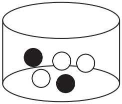
Note that when events A and B are independent,

$$\mathrm{Prob}[A \,|\, B] = \frac{\mathrm{Prob}[A, B]}{\mathrm{Prob}[B]} = \frac{\mathrm{Prob}[A]\mathrm{Prob}[B]}{\mathrm{Prob}[B]} = \mathrm{Prob}[A].$$

This should be intuitively satisfying since knowing B happened doesn't give you any useful information about A.

Conditional Probability Example

[Leon-Garcia p. 50] An urn contains two black balls and three white balls. Two balls are selected at random from the urn (without replacement).



1. What is the probability that the first ball you select from the urn is black?
2. Given that the first ball that you select from the urn is black, what is the probability that the second ball you select from the urn is also black?
3. What is the probability that both balls you select are black?

Another Conditional Probability Example

An urn contains one ball known to be either black or white with equal probability. A white ball is added to the urn, the urn is shaken, and a ball is removed randomly from the urn.

1. If the ball removed from the urn is black, what is the probability that the ball remaining in the urn is white?

2. If the ball removed from the urn is white, what is the probability that the ball remaining in the urn is white?

Joint Random Variables

When we have two random variables, we require a joint distribution. The joint CDF is defined as

$$F_{X,Y}(x,y) = \mathrm{Prob}[X \leq x \cap Y \leq y] = \mathrm{Prob}[X \leq x, Y \leq y]$$

and the joint PDF is defined as

$$p_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

If you know the joint distribution, you can get the marginal distributions:

$$\begin{aligned} F_X(x) &= F_{X,Y}(x,\infty) \\ F_Y(y) &= F_{X,Y}(\infty,y) \\ p_X(x) &= \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,dy \\ p_Y(y) &= \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,dx \end{aligned}$$

Marginals are not enough to specify the joint distribution, except in special cases.

Joint Statistics

Note that the means and variances are defined as usual for X and Y. When we have a joint distribution, we have two new statistical quantities:

Definition

The **correlation** of the random variables X and Y is defined as

$$\mathrm{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy p_{X,Y}(x,y)\,dx\,dy.$$

Definition

The **covariance** of the random variables X and Y is defined as

$$\mathrm{cov}[X,Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mathrm{E}[X])(y-\mathrm{E}[Y]) p_{X,Y}(x,y)\,dx\,dy.$$

Conditional Distributions

If X and Y are both discrete random variables (both are drawn from finite sets) with $\mathrm{Prob}[X = x] > 0$, then

$$\mathrm{Prob}[Y = y \,|\, X = x] \quad = \quad \frac{\mathrm{Prob}[X = x, Y = y]}{\mathrm{Prob}[X = x]}$$

If Y is a discrete random variable and X is a continuous random variable, then the conditional probability that $Y = y$ given $X = x$ is

$$\mathrm{Prob}[Y = y \,|\, X = x] = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

where $p_{X,Y}(x, y) := \lim_{h \to 0} \frac{\mathrm{Prob}[x-h<X\le x, Y=y]}{h}$ is the joint PDF-probability of the random variable X and the event $Y = y$.

If X and Y are both continuous random variables, then the conditional PDF of Y given $X = x$ is

$$p_Y(y \,|\, X = x) = p_Y(y \,|\, x) = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

with $p_{X,Y}(x, y)$ as the usual joint distribution of X and Y.

Conditional Statistics

Definition

The **conditional mean** of a random variable Y given $X = x$ is defined as

$$\mathrm{E}[Y \,|\, x] \;=\; \int_{-\infty}^{\infty} y p_Y(y \,|\, x)\, dy.$$

The definition is identical to the regular mean except that we use the conditional PDF.

Definition

The **conditional variance** of a random variable Y given $X = x$ is defined as

$$\mathrm{var}[Y \,|\, x] \;=\; \int_{-\infty}^{\infty} (y - \mathrm{E}[Y \,|\, x])^2 p_Y(y \,|\, x)\, dx.$$

Independence of Random Variables

Two random variables are independent if their joint distribution is equal to a product of their marginal distributions, i.e.

$$p_{X,Y}(x,y) = p_X(x)p_Y(y).$$

If X and Y are independent, the conditional PDFs can be written as

$$p_Y(y\,|\,x) = \frac{p_{X,Y}(x,y)}{p_X(x)} = \frac{p_X(x)p_Y(y)}{p_X(x)} = p_Y(y)$$

and

$$p_X(x\,|\,y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x).$$

These results should be intuitively satisfying since knowing $X = x$ (or $Y = y$) doesn't tell you anything about Y (or X).

Jointly Gausian Random Variables

Definition: The random variables $\boldsymbol{X} = [X_1, \dots, X_k]^\top$ are jointly Gaussian if their joint density is given as

$$p_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; |2\pi \boldsymbol{P}|^{-1/2} \exp\left(\frac{(\boldsymbol{x}-\boldsymbol{\mu}_X)^\top \boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_X)}{2}\right)$$

where $\boldsymbol{\mu}_X = \mathrm{E}[\boldsymbol{X}]$ and $\boldsymbol{P} = \mathrm{E}[(\boldsymbol{X}-\boldsymbol{\mu}_X)(\boldsymbol{X}-\boldsymbol{\mu}_X)^\top]$.
Remarks:

1. $\boldsymbol{m}_X = [\mathrm{E}[X_1], \dots, \mathrm{E}[X_k]]^\top$ is a k-dimensional vector of means
2. $\boldsymbol{P}$ is a $k \times k$ matrix of covariances, i.e.,

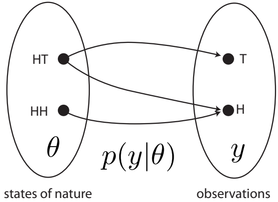
$$\boldsymbol{P} \;=\; \begin{bmatrix} \mathrm{E}[(X_1-\mu_{X_1})(X_1-\mu_{X_1})] & \dots & (X_1-\mu_{X_1})(X_k-\mu_{X_k})] \\ \vdots & \ddots & \vdots \\ \mathrm{E}[(X_k-\mu_{X_k})(X_1-\mu_{X_1})] & \dots & (X_k-\mu_{X_k})(X_k-\mu_{X_k})] \end{bmatrix}$$

Where We Are Heading

We are going to begin our study of detection and estimation by learning the fundamental concepts of **hypothesis testing**.

Hypothesis testing involves making inferences about unknown things (states of nature) from observations. Examples:

- Infer if the coin is fair or unfair after observing one or more flips.



- Infer if the airplane is friend or foe by observing a radar signature.

We will be working with lots of conditional probability expressions in our study of hypothesis testing. You should feel comfortable with this material.

Where We Will Be in April: Kalman Filtering

2008 Charles Stark Draper Prize: Dr. Rudolph Kalman

The Kalman Filter uses a mathematical technique that removes noise from series of data. From incomplete information, it can optimally estimate and control the state of a changing, complex system over time. Applications include target tracking by radar, global positioning systems, hydrological modeling, atmospheric observations, and time-series analyses in econometrics.

(paraphrased from http://www.nae.edu/)