ECE4110: Random Processes

Probability Review

Qing Zhao
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14850
qz16@cornell.edu

Outline

- Probability space.
- Conditional probability.
- Total probability theorem and Bayes' rule.
- Independence of events.
- Random variables, distributions and densities.
- Expectation and variance.
- Jointly distributed random variables.
- Conditional distribution and conditional expectation.

Random Experiment

Random Experiment:

- The outcome cannot be pre-determined;
- Repeating the experiment may not lead to the same outcome.

Examples of Random Experiment

R1. Toss a fair coin twice in succession

R2. Throw a die

R3. Take ECE4110 until you pass

R4. Throw a dart to a unit disk

Outcomes of a Random Experiment:

R1. Get two heads.

R2. 2 is thrown.

R3. Pass in the second try.

R4. Hit point $(0.5, 0.5)$

Events of a Random Experiment:

R1. $A \triangleq$ get the same outcome in two coin tosses.

R2. $B \triangleq$ an even number is thrown.

R3. $C \triangleq$ pass in no more than three tries.

R4. $D \triangleq$ hit the inner disk with radius $1/2$.

Probability Measure:

How likely does a particular event happen?

Probability Space

Probability Space:

A probability space is defined by a triplet $(\Omega, \mathcal{F}, \Pr)$:

- Ω is the sample space that contains all possible outcomes.

- $\mathcal{F}$ is a σ-field of subsets of Ω (events) satisfying
  - (i) $\Omega \in \mathcal{F}$.
  - (ii) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
  - (iii) If $A_i \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

- $\Pr : \mathcal{F} \to [0,1]$ is a function on $\mathcal{F}$ satisfying
  - (i) $0 \leq \Pr(A) \leq 1$ for all $A \in \mathcal{F}$.
  - (ii) $P(\Omega) = 1$.
  - (iii) If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events in $\mathcal{F}$ (i.e., $A_i A_j = \emptyset$ for all $i \neq j$), then

$$\Pr(\bigcup_{i=1}^{\infty} A_i) = \sum \Pr(A_i)$$

Interpretation of the σ-Field

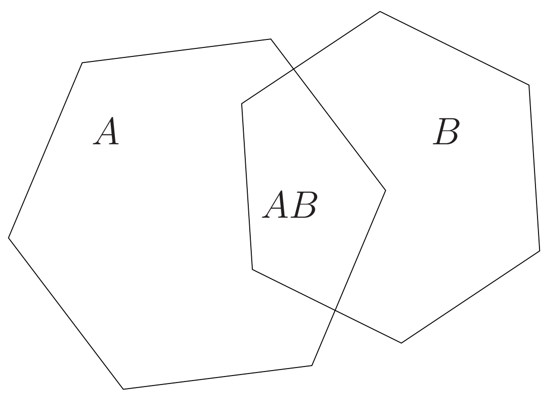
Interpretation of $\mathcal{F}$:

- The σ-field $\mathcal{F}$ can be interpreted as a way to specify how fine our observations about the random experiment of interest are. For example, for the case of throwing a die, $\Omega = \{1, 2, 3, 4, 5, 6\}$. If $\mathcal{F} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$, then it specifies a probabilistic model where we can only observe (or only care about) whether the outcome is an even or an odd number.
- If Ω is finite, one possible $\mathcal{F}$ is the power set $\mathcal{F} = 2^{\Omega}$ where $\mathcal{F}$ includes all possible subsets of Ω. This gives the finest observation model.
- $\mathcal{F} = \{\emptyset, \Omega\}$ is a trivial σ-field, corresponding to the coarsest observation model.

Properties

Properties:

- $\Pr(A^c) = 1 - \Pr(A)$.
- $\Pr(\emptyset) = 0$.
- If $A \subseteq B$, then $\Pr(A) \leq \Pr(B)$.
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(AB)$



- The union bound:

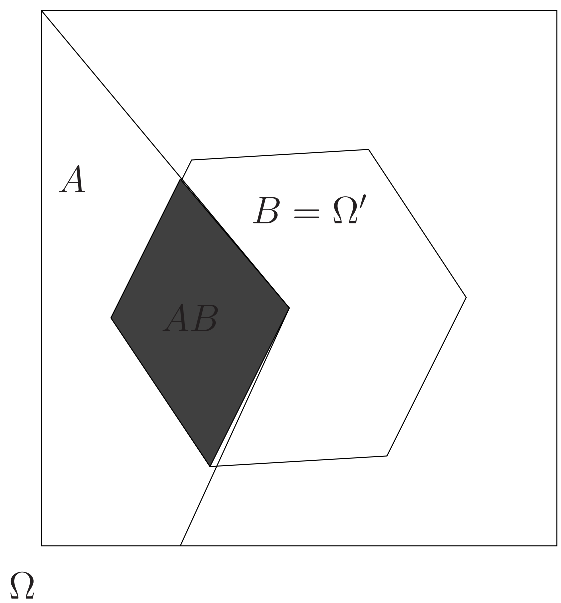
$$\Pr(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \Pr(A_i)$$

Conditional Probability

Conditional Probability:

If A and B are events and $\Pr(B) \neq 0$, then the conditional probability of A given B is defined as

$$\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)}$$



Interpretation:

We can think "conditioning" as generating a new probability model from the old by treating B as the new sample space Ω'. The probability measure for Ω' should thus be normalized by $\Pr(B)$.

Total Probability Theorem and Bayes' Rule

Total Probability Theorem:

If $\{E_1, \ldots, E_k\}$ partition Ω, *i.e.,*

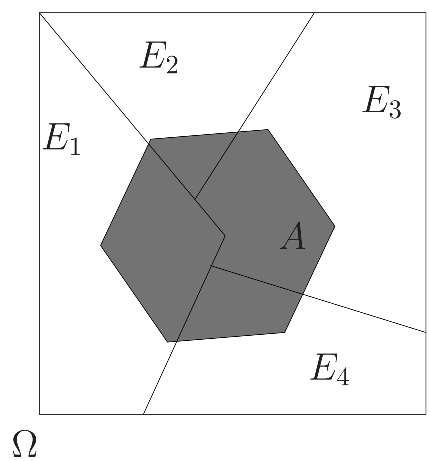
$$E_i \cap E_j = \emptyset \text{ for } i \neq j, \qquad \cup_{i=1}^{k} E_i = \Omega,$$

then

$$\Pr(A) = \sum_{i=1}^{k} \Pr(AE_i).$$

If $\Pr(E_i) \neq 0$ for all i, we further have

$$\Pr(A) = \sum_{i=1}^{k} \Pr(A|E_i)\Pr(E_i).$$



Bayes' Formula:

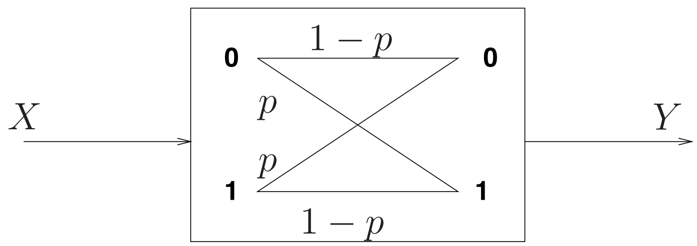
$$\Pr(E_i|A) = \frac{\Pr(A|E_i)\Pr(E_i)}{\sum_{j=1}^{k}\Pr(A|E_j)\Pr(E_j)}$$

Example: Binary Symmetric Channel

Binary Symmetric Channel:

Defined by the conditional probability

$$\Pr(Y=0|X=0) = \Pr(Y=1|X=1) = 1-p$$
$$\Pr(Y=1|X=0) = \Pr(Y=0|X=1) = p$$



Prior Probability:

Suppose that $\Pr(X=0) = \Pr(X=1) = \frac{1}{2}$.

Posterior Probability:

$$\begin{aligned}\Pr(X=0|Y=0) &= \frac{\Pr(X=0, Y=0)}{\Pr(Y=0)} = 1-p \\ \Pr(X=1|Y=0) &= \frac{\Pr(X=1, Y=0)}{\Pr(Y=0)} = p\end{aligned}$$

where

$$\Pr(Y=0) = \Pr(X=0)\Pr(Y=0|X=0) + \Pr(X=1)\Pr(Y=0|X=1) = \frac{1}{2}$$

Detection:

Suppose that $Y=0$ is received at the channel output, what is the detection of X that minimizes the probability of error?

Statistical Independence

Independence:

Two events A_1 and A_2 are statistically independent if

$$\Pr(A_1A_2) = \Pr(A_1)\Pr(A_2).$$

In general, events $\{A_1, \cdots, A_n\}$ are statistically independent if

$$\Pr(A_{i_1}A_{i_2}\dots A_{i_k}) = \Pr(A_{i_1})\Pr(A_{i_2})\dots\Pr(A_{i_k})$$

for all $\{i_1, \cdots, i_k\} \subseteq \{1, \cdots, n\}$.

Pairwise Independence:

$\{A_1, \cdots, A_n\}$ are pairwise independent if

$$\Pr(A_iA_j) = \Pr(A_i)\Pr(A_j)$$

for all $i \neq j$.

$$\boxed{\text{Pairwise independence} \not\Rightarrow \text{independence}}$$

Interpretation:

If $\Pr(A_1A_2) = \Pr(A_1)\Pr(A_2)$, then $\Pr(A_1|A_2) = \Pr(A_1)$ and $\Pr(A_2|A_1) = \Pr(A_2)$, that is, the knowledge of the occurrence of one event does not alter the probability of the other event. This is the reason we call these two events "independent".

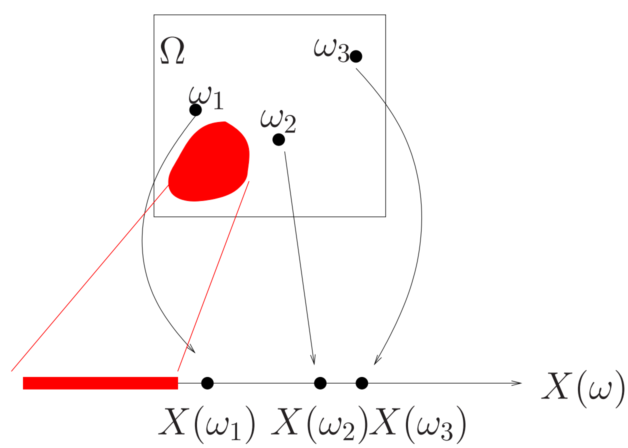
Random Variables

Random Variable:

Given a probability space $(\Omega, \mathcal{F}, \Pr)$, a random variable is a function

$$X : \Omega \to R$$

such that, for all x, $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$.



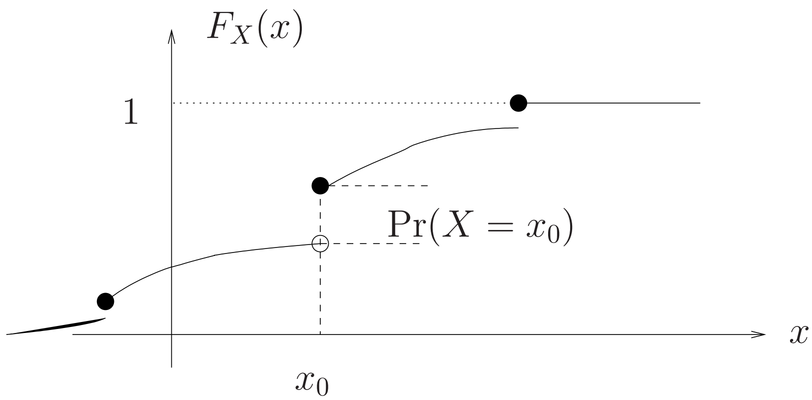
Notation:

Capital letters (X) denote random variables whereas lower-cased letters (x) indicate realizations of random variables (X).

Cumulative Distribution Function

The cumulative distribution function (CDF) of a random variable X is

$$F_X(x) = \Pr(X \leq x)$$



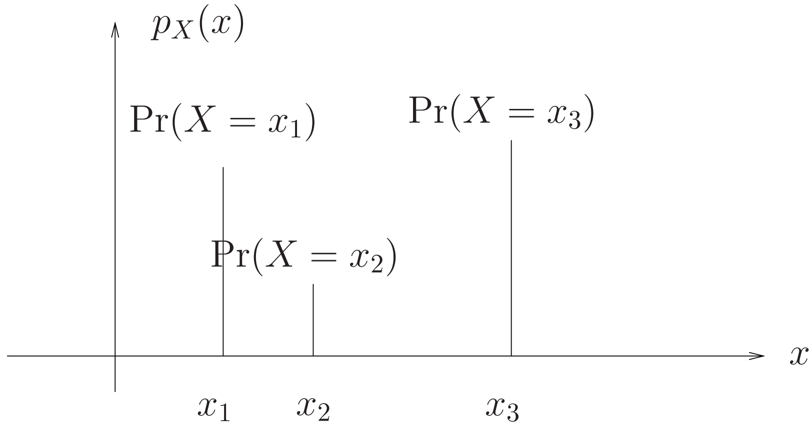
Properties:

1. $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$.
2. If $x < y$, then $F_X(x) \leq F_X(y)$.
3. F is right continuous, *i.e.*, $\lim_{\Delta\to0^+} F_X(x+\Delta) = F_X(x)$
4. $\Pr(x < X \leq y) = F_X(y) - F_X(x)$.
5. $\Pr(X = x_0) = F_X(x_0) - \lim_{y\uparrow x_0} F_X(y)$.

Probability Mass Function

For a discrete random variables X (*i.e.*, X takes values in a countable set $\{x_i\}$), the probability mass function (PMF) of X is given by

$$p_X(x) = \Pr(X = x)$$



Properties:

- The PMF is related to CDF by

$$\begin{cases} F_X(x) &= \sum_{u:u\leq x} p_X(u) \\ p_X(x) &= F_X(x) - F_x(x^-) \end{cases}$$

Probability Density Function

A random variable is continuous if its distribution function can be expressed as

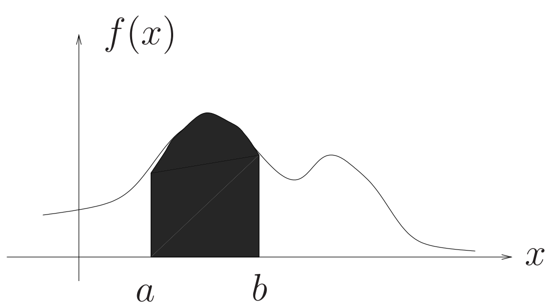
$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

for some integrable function $f_X : \mathcal{R} \to [0, \infty)$. Function $f_X(x)$ is the probability density function (PDF) of X:

$$f_X(x) = \frac{d}{dx}F_X(x).$$

Properties:

- $f_X(x) \geq 0$.
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$.
- $\int_a^b f_X(x)dx = \Pr(a < X \leq b)$.



Expectation of a Random Variable

Expectation:

- For a discrete random variable X with PMF $p_X(x)$

$$\mathbb{E}(X) = \sum_k x_k p_X(x_k)$$

- For a continuous random variable X with PDF $f_X(x)$

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

Properties:

- Law of the Unconscious Statistician (LOTUS):

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

- Linearity:

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$$

- Preservation of order:

$$\text{if } \Pr(X \geq Y) = 1, \text{ then } \mathbb{E}(X) \geq \mathbb{E}(Y)$$

- Integration by parts formula:

$$\mathbb{E}(X) = \int_{0}^{\infty} (1 - F_X(x)) dx - \int_{-\infty}^{0} F_X(x) dx$$

Variance and Moments

Variance and Standard Deviation:

- Variance: $\text{Var}(X) \triangleq \mathbb{E}\left[(X - \mathbb{E}(X))^2\right]$
- Standard deviation: $\sqrt{\text{Var}(X)}$

Moments and Central Moments:

- nth moment: $\mathbb{E}\left[X^n\right]$
- n-th central moment: $\mathbb{E}\left[(X - \mathbb{E}(X))^n\right]$

Properties:

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \geq 0$$

Jointly Distributed Random Variables

Joint Distribution:

- joint CDF:

$$F_{X_1X_2\ldots X_m}(x_1,\ldots,x_m) = \Pr(X_1 \le x_1,\ldots,X_m \le x_m).$$

- joint PMF for discrete random variables:

$$p_{X_1X_2\ldots X_m}(x_1,\ldots,x_m) = \Pr(X_1 = x_1,\ldots,X_m = x_m).$$

- joint PDF for continuous random variables:

$$f_{X_1X_2\ldots X_m}(x_1,\ldots,x_m) = \frac{\partial^m}{\partial x_1 \cdots \partial x_m} F_{X_1X_2\ldots X_m}(x_1,\ldots,x_m)$$

- marginal CDF of X_1:

$$F_{X_1}(x) \stackrel{\Delta}{=} \Pr(X_1 \le x) = F_{X_1X_2\ldots X_m}(x,\infty,\ldots,\infty)$$

- marginal PDF of X_1:

$$f_{X_1}(x) = \int f_{X_1X_2\ldots X_m}(x,x_2,\ldots,x_m)dx_2\ldots dx_m$$

Independent Random Variables:

$X_1,\cdots,X_m$ are statistically independent if

$$F_{X_1X_2\ldots X_m}(x_1,\ldots,x_m) = F_{X_1}(x_1)\ldots F_{X_m}(x_m)$$

or equivalently,

$$\begin{aligned} p_{X_1X_2\ldots X_m}(x_1,\ldots,x_m) &= p_{X_1}(x_1)\ldots p_{X_m}(x_m) \quad \text{(discrete)} \\ f_{X_1X_2\ldots X_m}(x_1,\ldots,x_m) &= f_{X_1}(x_1)\ldots f_{X_m}(x_m) \quad \text{(continuous)} \end{aligned}$$

Conditioning on Random Variables

Conditional PMF:

Suppose that X and Y have a joint PMF $p_{XY}(x,y)$. The conditional PMF $p_{X|Y}(x|y)$ for y satisfying $p_Y(y) \neq 0$ is defined as

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

Conditional PDF:

Suppose that X and Y have a joint PDF $f_{XY}(x,y)$. The conditional PDF $f_{X|Y}(x|y)$ for y satisfying $f_Y(y) \neq 0$ is defined as

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

Conditional Expectation:

The conditional expectation of X given $Y = y$ is given by

$$\mathbb{E}(X|Y = y) = \sum_x x p_{X|Y}(x|y) \quad \text{(discrete)}$$

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \quad \text{(continuous)}$$

Remarks:

- If X and Y are independent, $p_{X|Y}(x|y) = p_X(x)$ or $f_{X|Y}(x|y) = f_X(x)$.
- $\mathbb{E}(X|Y)$ is a function of Y, thus a random variable with distribution determined by $F_Y(y)$.
- $\mathbb{E}(X) = \mathbb{E}_Y\left(\,\mathbb{E}(X|Y)\,\right)$.

Example: Two Coins

There are two coins. Coin A is a fair coin. Coin B has two heads. Consider the following random experiment. First flip coin A. If a head shows up, flip coin A again. If it is a tail, flip coin B.

Let X denote the outcome of the first coin flip with $X = 1$ for head and $X = 0$ for tail. Let Y denote the outcome of the second coin flip with a similar definition.

- What is the joint PMF of $p_{XY}(x, y)$?
- What are the marginal PMFs $p_X(x)$ and $p_Y(y)$?
- What are the conditional PMFs of Y given $X = 1$ and $X = 0$, and X given $Y = 1$ and $Y = 0$?
- What are $\mathbb{E}(Y|X = 1)$, $\mathbb{E}(Y|X = 0)$, $\mathbb{E}(X|Y = 1)$, and $\mathbb{E}(X|Y = 0)$?
- What are the PMFs of $\mathbb{E}(Y|X)$ and $\mathbb{E}(X|Y)$?
- Calculate $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ by averaging $\mathbb{E}(X|Y)$ and $\mathbb{E}(Y|X)$, respectively.

Example: Signal in Noise

Consider a binary signal $X \in \{-1, 1\}$ corrupted by additive Gaussian noise:

$$Y = X + W$$

where W is independent of X, and has a Gaussian distributed with zero mean and variance σ^2, i.e.,

$$f_W(w) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{w^2}{2\sigma^2}\}.$$

Assume that $\Pr(X = -1) = p$ and $\Pr(X = 1) = 1 - p$.

- What is the distribution of Y?
- What is $\mathbb{E}(Y)$?
- What is the distribution of Y given $X = -1$ $\big(X = 1\big)$?
- What are $\mathbb{E}(Y|X = -1)$ and $\mathbb{E}(Y|X = 1)$?
- What is the distribution of $\mathbb{E}(Y|X)$?
- Given a noisy observation of $Y = y$, what is the optimal detection of X that minimizes the probability of error?
- What is the probability of error given by the optimal detector?