Communication Systems
Chapter 7 Fundamentals of Probability Theory

Hunter Mills

April 6, 2024

The previous chapters have only focused on deterministic signals which cannot relay information since the information is already known. The higher the signal uncertainty, the more information it can convey. In addition to information bearing signals, noise signals in a system are also unpredictable (cannot be subtracted). The unpredictable message and noise signals are known as **random processes**.

1   Concept of Probability

Some terminologies needed for probability. **experiments** are used in probability to describe a process whose outcome cannot be fully predicted. An experiment may have server separately identifiable   **outcomes**. An **event** is a subset of outcomes that share some common characteristic. An event occurs if the outcome of an experiment belongs to the specific subset of outcomes defining the event.

Set theory is used in probability. The **sample space** $\mathcal{S}$ is a collection of all possible and distinct outcomes of an experiment. The sample space $\mathcal{S}$ specifies the experiment. Each outcome is an **element** or **sample point** of this space $\mathcal{S}$. In the experiment of rolling a die, the sample space consists of ζ_1, ζ_2, ζ_3, ζ_4, ζ_5, ζ_6, where ζ_i is the outcome when a 'number i is thrown'. An event on the other hand is a subset of $\mathcal{S}$, 'an odd number is thrown', denoted by A_{odd}.

$$A_{odd} = (\zeta_1, \zeta_3, \zeta_5) \qquad A_{even} = (\zeta_2, \zeta_4, \zeta_6) \tag{1}$$

An event, a number ≤ 4 is denoted as

$$B = (\zeta_1, \zeta_2 \zeta_3, \zeta_4) \tag{2}$$

The complement of any event A is denoted A^c and is the event containing all points not in A. For example, $A_{odd}^c = A_{even}$. An event that has no sample points is a **null event** and is denoted as $\phi = \mathcal{S}^c$.

The **union** of events A and B, denoted by $A \cup B$, is the event that contains all points in A and B (having an outcome of either A or B).

$$A_{odd} \cup B = (\zeta_1, \zeta_2 \zeta_3, \zeta_4, \zeta5) \tag{3}$$

The intersection of events A and B, denoted by $A \cap B$, or AB is the event that contains points common to A and B. This is the event, outcome is both A and B, and also is known as the **joint event**.

$$A_{even} B = (\zeta_2, \zeta_4) \tag{4}$$

$$A \cap B = B \cap A \tag{5}$$

If events A and B are such that

$$A \cap B = \phi \tag{6}$$

A and B are said to be **disjoint** or **mutually exclusive** (A and B cannot occur simultaneously).
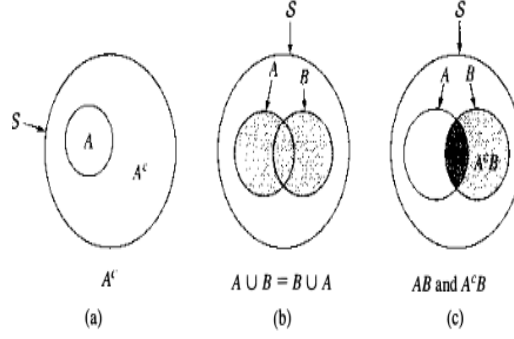
Figure 1: Complement, Union and Intersection

1.1 Relative Frequency and Probability

Although the outcome of a random experiment is unpredictable, there is a statistical regularity of the outcomes. For example a coin flipped many times has the relative frequency of heads/tails of $1/2$. Let A be one of the events of interest in an experiment and N independent trials, and if the event A occurs in $N(A)$ of the trials then the fraction

$$f(A) = \lim_{N \to \infty} \frac{N(A)}{N} \tag{7}$$

is the **relative frequency** of the event A. The probability of an event A has the same connotations as the relative frequency, this the probability of A is

$$P(A) = \lim_{N \to \infty} \frac{N(A)}{N} \tag{8}$$

and it follows that

$$0 \leq P(A) \leq 1 \tag{9}$$

If A and B are mutually exclusive, the event $A \cup B$ occurs in $N(A) + N(B)$ trials and

$$P(A \cup B) = \lim_{N \to \infty} \frac{N(A) + N(B)}{N} \tag{10}$$

$$P(A \cup B) = P(A) + P(B) \quad \text{if} \quad A \cap B = \phi \tag{11}$$

For more than two mutually exclusive events where

$$A_i \cap A_j = \phi \quad \text{if} \quad i \neq j \tag{12}$$

$$P(\underset{i}{\cup} A_i) = \sum_i P(A_i) \tag{13}$$

An important concept in probability is that

$$P(\mathcal{S}) = 1 \tag{14}$$

1.2 Conditional Probability and Independent Events

1.2.1 Conditional Probability

It often happens that the probability of one event if influenced by another event. The **conditional probability**, $P(B|A)$ (probability of B given A), is the probability of event B when A has occurred.

$$P(A \cap B) = P(A)P(B|A) \tag{15}$$

2

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{16}$$

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} \tag{17}$$

The above equations are known as **Bayes Rule**.

1.2.2 Multiplication Rule for Conditional Probabilities

Given events A_1, A_2, ... , A_n and $A_1A_2...A_n \neq \phi$ then

$$P(A_1A_2...A_n) = \frac{P(A_1A_2...A_n)}{P(A_1A_2...A_{n-1})} * \frac{P(A_1A_2...A_{n-1})}{P(A_1A_2...A_{n-2})}... * \frac{P(A_1A_2)}{P(A_1)}P(A_1) \tag{18}$$

$$P(A_1A_2...A_n) = p(A_n|A_1A_2...A_{n-1}) * P(A_{n-1}|A_1A_2...A_{n-2})... * P(A_2|A_1)P(A_1) \tag{19}$$

1.2.3 Independent Events

Unlike the previous section some events occurrence does not influence another event. The event B is said to be **independent** iff

$$P(A \cap N) = P(A)P(B) \tag{20}$$

So it follows

$$P(B|A) = P(B) \tag{21}$$

1.3 Bernoulli Trials

In Bernoulli trials, if A occurs we call it a success and $P(A) = p$ where p if the probability of success. If q is the probability of failure the $q = 1 - p$. To find the probability of k successes in n (Bernoulli) trials then

$$P(k\text{successes in }n\text{trials}) = \binom{n}{k}p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!}p^k(q)^{n-k} \tag{22}$$

1.4 The Total Probability Theorem

$\mathcal{S}$ can be partitioned into n mutually exclusive events $A_1,...,A_n$. Thus

$$\bigcup_{i=1}^{n} A_i = S \quad \text{and} \quad A_i \cap A_j \quad \text{if} \quad i \neq j \tag{23}$$

The probability of event B is

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) \tag{24}$$

Bayes theorem for the total probability theorem is

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum P(B|A_i)P(A_i)} \tag{25}$$

The probability $P(A_j|B)$ is known as the posterior probability, the probability that B may be caused be A_j.

2 Random Variables

It is connivent to assign sample points real numbers, so $\chi(\zeta_i)$ would represent the sample points of ζ_i ($i = 1, 2...m$). This $\chi(.)$ if a function that maps sample points (ζ_i) to real numbers x_i. We now have a **random variable** that takes on the values x_i. The probability of a random variable (RV) χ taking a value of χ_i is denoted by $P_\chi(x_i) =$ probability of $\chi = x_i$.

2.1 Discrete Random Variables

A RV is discrete if there exists a distinct number of outcomes χ_i such that

$$\sum_i P_\chi(x_i) = 1 \tag{26}$$

This can be extended to two random variables χ and γ. The joint probability $P_{\chi\gamma}(x_i, y_i)$ is the probability that $\chi = x_i$ and $\gamma = y_i$. For the general case when χ can take values x_i and γ can take values y_j,

$$\sum_i \sum_j P_{\chi\gamma}(x_i, y_i) = 1 \tag{27}$$

2.2 Conditional Probabilities

If χ and γ are two RVs then the probability of $\chi = x_i$ given $\gamma = y_j$ is denoted as $P_{\chi|\gamma}(x_i|y_j)$ and

$$\sum_i P_{\chi|\gamma}(x_i|y_j) = \sum_j P_{\gamma|\chi}(y_j|x_i) = 1 \tag{28}$$

$$P_{\chi\gamma}(x_i, y_j) = P_{\chi|\gamma}(x_i|y_j)P_\gamma(y_j) = P_{\gamma|\chi}(y_j|x_i)P_\chi(x_i) \tag{29}$$

and from Bayes Rule

$$P_{\chi|\gamma}(x_i|y_j) = \frac{P_{\chi\gamma}(x_i, y_j)}{P_\gamma(y_j)} \tag{30}$$

The probability $P_\chi(x_i)$ and $P_\gamma(y_i)$

$$P_\gamma(y_j) = \sum_i P_{\chi\gamma}(x_i, y_j) \tag{31}$$

$$P_\chi(x_i) = \sum_j P_{\chi\gamma}(x_i, y_j) \tag{32}$$

are called **marginal probabilities**.

2.3 Cumulative Distribution Function

The **cumulative distribution function (CDF)** $F_\chi(x)$ of a RV χ is the probability that χ takes a value less than or equal to x.

$$F_\chi(x) = P(\chi \leq x) \tag{33}$$

The CDF has the following properties:

- **1** - $F_\chi(x) \geq 0$

- **2** - $F_\chi(\infty) = 1$

- **3** - $F_\chi(-\infty) = 0$

- **4** - $F_\chi(x)$ is a non-decreasing function

Point 4 leads to

$$F_\chi(x_1) \leq F_\chi(x_2) \quad \text{for} \quad x_1 \leq x_2 \tag{34}$$
$$F_\chi(x_2) = F_\chi(x_1) + P(x_1 < \chi < x_2) \tag{35}$$

2.4 Continuous Random Variables

A continuous RV χ can take any value in a certain interval. Since there is an infinite number of values in that interval

$$P_\chi(x_i) = 0 \tag{36}$$

The CDF of a continuous RV has the same equations that the discrete RV has.

4

2.4.1 Probability Density Function

The PDF is the probability that the RV will take a value in a given range. The PDF $p_\chi(x)$, is defined as

$$p_\chi(x) = \frac{dF_\chi(x)}{dx} \tag{37}$$

And the CDF $F_\chi(x)$ is defined as

$$F_\chi(x) = \int_{-\infty}^{x} p_\chi(u)du \tag{38}$$

Some other useful formulas are

$$P(x_1 \leq \chi \leq x_2) = F_\chi(x_2) - F_\chi(x_1) = \int_{x_1}^{x_2} \tag{39}$$

$$\int_{-\infty}^{\infty} p_\chi(x)dx = 1 \tag{40}$$

2.5 Discrete PDF

For an RV χ with values x_i and probabilities a_i the PDF is

$$p_\chi(x) = \sum_{r=1}^{n} a_r \delta(x - x_r) \tag{41}$$

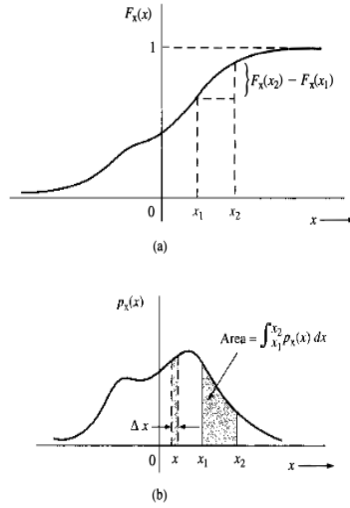since the CDF is a step function, the PDF (derivative of the CDF) are impulse functions.

Figure 2: CDF and PDF of Continuous RV

2.6   The Gaussian (Normal) Random Variable

The PDF

$$p_\chi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{42}$$

is known as the **Gaussian** or **Normal** distribution pdf. It has zero mean and variable of 1.

The CDF is

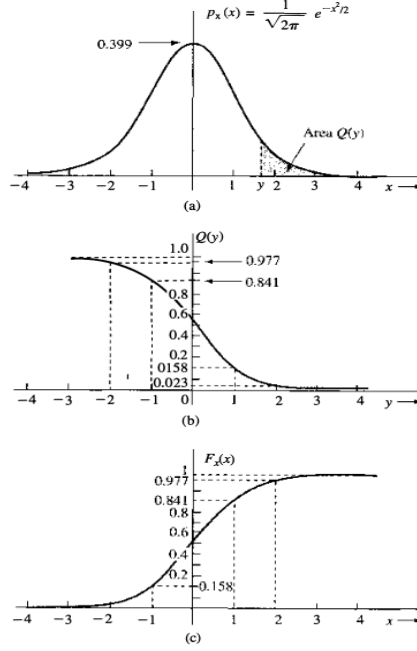$$F_\chi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx \tag{43}$$

Figure 3: Normal PDF, Q(y) and CDF

Which cannot be evaluated and must be computed numerically. It is connivent to use the function $Q(.)$

$$Q(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-x^2/2} dx \tag{44}$$

The shaded area in figure 3 is $Q(y)$.

$$Q(-y) = 1 - Q(y) \tag{45}$$

$$F_\chi(x) = 1 - Q(x) \tag{46}$$

The more general Normal Distribution takes two parameters m, σ,

$$p_\chi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2} \tag{47}$$

$$F_\chi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-m)/\sigma} e^{-z^2/2} dz = 1 - Q(\frac{x-m}{\sigma}) \tag{48}$$

$$P(\chi > x) = Q(\frac{x-m}{\sigma}) \tag{49}$$

$$P(\chi \le x) = 1 - Q(\frac{x-m}{\sigma}) \tag{50}$$

The Gaussian distribution is important in communications since it can model most noise.

2.7 Joint Distribution

For two RV's χ and γ, the CDF $F_{\chi\gamma}(x,y)$ is defined as

$$F_{\chi\gamma}(x,y) \doteq P(\chi \le x \text{ and } \gamma \le y) \tag{51}$$

and the joint PDF $p_{\chi\gamma}(x,y)$ as

$$p_{\chi\gamma}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{\chi\gamma}(x,y) \tag{52}$$

The probability of observing χ as $(x, x + \Delta y]$ and γ as $(y, y + \Delta y]$ is given by the volume under the joint PDF.

$$P(x_1 \leq \chi \leq x_2, y_1 \leq \gamma \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} p_{\chi\gamma}(x, y) dx dy \tag{53}$$

The independent probabilities, $p_\chi(x)$ and $p_\gamma(y)$ can be obtained from $p_{\chi\gamma}(x, y)$ and are called **marginal densities**. The PDF and CDF for χ is shown below, but still holds for γ.

$$p_\chi(x) = \int_{-\infty}^{\infty} p_{\chi\gamma}(x, y) dy \tag{54}$$

$$F_\chi(x) = F_{\chi\gamma}(c, \infty) \tag{55}$$

2.8 Conditional Densities

Bayes rule for continuous conditional probabilities is

$$p_{\chi|\gamma}(x|y) = \frac{p_{\gamma|\chi}(y|x)p_\chi(x)}{p_\gamma(y)} \tag{56}$$

2.9 Independent Random Variables

The continuous RV's χ and γ are independent if

$$p_{\chi|\gamma}(x|y) = p_\chi(x) \tag{57}$$

and implies

$$p_{\chi\gamma}(x, y) = p_\chi(x)p_\gamma(y) \tag{58}$$

and

$$F_{\chi\gamma}(x, y) = F_\chi(x)F_\gamma(y) \tag{59}$$

3 Statistical Means

The statistical mean $\bar{\chi}$ is defined as

$$\bar{\chi} = \sum_{i=1}^{N} x_i P_\chi(x_i) \tag{60}$$

and is also called the **average** or **expected value**. It can also be denoted as

$$\bar{\chi} = E[\chi] \tag{61}$$

If χ is continuous then

$$\bar{\chi} = \int_{-\infty}^{\infty} x p_\chi(x) dx \tag{62}$$

For a Gaussian RV the mean is $E[\chi] = m$.

3.1 Mean of a Function of RV's

We may seek to find the mean value of γ where that is a function of χ, ie $\gamma = g(\chi)$.

$$\bar{\gamma} = \overline{g(\chi)} = \sum_{i=1}^{N} g(x_i) P_\chi(x_i) \tag{63}$$

And for the case of two continuous RV's

$$\overline{g(\chi, \gamma)} = \int \int g(x, y) p_{\chi\gamma}(x, y) dx dy \tag{64}$$

7

3.2 Mean of the Sum

$$E[\chi + \gamma] = E[\chi] + E[\gamma] \tag{65}$$

3.3 Mean of the Product of two Functions

$$\overline{g_1(\chi)g_2(\gamma)} = \int\int g_1(\chi)g_2(\gamma)p_{\chi\gamma}(x,y)dxdy \tag{66}$$

and if χ and γ are independent then

$$\overline{g_1(\chi)g_2(\gamma)} = \overline{g_1(\chi)}\,\overline{g_2(\gamma)} \tag{67}$$

3.4 Moments

The n_{th} moment of an RV χ is defined as the mean value of χ^n.

$$\overline{\chi^n} = \int x^n p_\chi(x)dx \tag{68}$$

and the nth central moment is

$$\overline{(\chi - \bar{\chi})^n} = \int (x - \bar{\chi})^n p_\chi(x)dx \tag{69}$$

The second central moment of an RV is called the **variance** and is denoted σ_χ^2 where σ_χ is the **standard deviation**.

$$\sigma_\chi = \overline{x^2} - \bar{\chi}^2 \tag{70}$$

3.5 Variance of a Sum of Independent RV's

If $\zeta = \chi + \gamma$ then

$$\sigma_\zeta^2 = \sigma_\chi^2 + \sigma_\gamma^2 \tag{71}$$

3.6 Chebyshev Inequality

The standard deviation is a measure of the width of the PDF. FOr a zero mean RV χ

$$P(|\chi| \le k\sigma_\chi) \ge 1 - \frac{1}{k^2} \tag{72}$$

4 Correlation

Correlation determines the nature of dependence between two RV's. The covariance $\sigma_{\chi\gamma}$ is defined as

$$\sigma_{\chi\gamma} = \overline{(\chi - \bar{\chi})(\gamma - \bar{\gamma})} \tag{73}$$

$$\sigma_{\chi\gamma} = \overline{\chi\gamma} - \bar{\chi}\bar{\gamma} \tag{74}$$

The correlation coefficient is

$$\rho_{\chi\gamma} = \frac{\sigma_{\chi\gamma}}{\sigma_\chi\sigma\gamma} \tag{75}$$

If χ and γ are uncorrelated then $\rho_{\chi\gamma} = 0$.

$$-1 \le \rho_{\chi\gamma} \le 1 \tag{76}$$

4.1 Independence vs Uncorrelatedness

Independent RV's are uncorrelated but uncorrelated RV's are not necessarily independent. **In one case independence and uncorrelatedness are equivalent when the RV χ and γ are jointly Gaussian.**

4.2 Mean Square of the Sum of Uncorrelated RV's

If χ and γ are uncorrelated then for $\zeta = \chi + \gamma$

$$\sigma_\zeta^2 = \sigma_\chi^2 + \sigma_\gamma^2 \tag{77}$$

That is, the variance of the sum is the sum of the variances for uncorrelated RV's.

5 Sum of RV's

It is useful to characterize an RV ζ as the sum of χ and γ.

$$\zeta = \chi + \gamma \tag{78}$$

$$F_\zeta(z) = P(\zeta \leq z) = P(\chi \leq \infty, \gamma \leq z - x) \tag{79}$$

$$F_\zeta(z) = \int dx \int_{-\infty}^{z-x} p_{\chi\gamma}(x, y) dy \tag{80}$$

and

$$p_\zeta(z) = \frac{dF_\zeta(z)}{dz} = \int p_{\chi\gamma}(x, z - x) dx \tag{81}$$

And if they are independent then

$$p_{\chi\gamma}(x, z - x) = p_\chi(x) p_\gamma(z - x) \tag{82}$$

and

$$p_\zeta(z) = \int p_\chi(x) p_\gamma(z - x) dx \tag{83}$$

6 Central Limit Theorem

Under certain conditions, the sum of a large number of RV's tends to be a Gaussian RV. Note that for a sum of RV,

$$\tilde{\chi}_n = \frac{\chi_1 + ... + \chi_N}{n} \tag{84}$$

is known as the sample mean. The sample mean of any distribution with nonzero finite variances converges to a Gaussian distribution with mean μ and a decreasing variance σ^2/n. In other words, regardless of the true distribution of χ_i, $\sum_i \chi_i$ can be approximated by a Gaussian distribution with mean $n\mu$ and variance $n\sigma^2$.