Theory of Random Number Generation

Ji Yong-Hyeon

Department of Information Security, Cryptology, and Mathematics
College of Science and Technology
Kookmin University

December 12, 2023

Contents

Chapter 1

Introduction

Summary

- Required Properties for Random Bit Generator
  - **Unpredictability**, **Unbiasedness**, **Independence**
- Components of Cryptographically Secure Random Bit Generator
  - TRNG (Entropy Source) + PRNG (Pseudorandom Number Generator)
- Methods for Evaluating the Security of Random Bit Generator
  - Estimation of entropy for the output sequence from TRNG
  - Statistical randomness tests for the output sequence from RNG
- Types of Random Bit Generators
  - Hardware/Software-based Random Bit Generators
  - Operating System-based Random Bit Generators
  - Various Standard Pseudorandom Number Generators

Functions of RBG (Random Bit Generator)

Provides random numbers required for cryptographic systems An essential element (algorithm) for the operation of cryptographic systems and modules Required Properties: Unpredictability, Unbiasedness, Independence between bits

Ideally, the output should be akin to the results of "coin tossing." Applications of Random Bit Generator

Generation of Key and Initialization Vector (IV) used in symmetric-key cryptography (block/stream ciphers) Generation of various parameters in public-key cryptography: prime number generation, probabilistic public-key cryptography, etc. Generation of various parameters used in cryptographic protocols: nonce, salt, etc.

Chapter 2

Probability Theory

2.1 Introduction

Definition 2.1.

- An **experiment** is the process of observing a phenomenon that has variation in its outcomes.
- The **sample space** S associated with an experiment is the collection of all possible distinct outcomes of the experiment.
- An **event** A, B is the set of elementary outcomes possessing a designated feature. $(A, B \subseteq S)$

Remark 2.1.

- Union: $A \cup B$
- Complement: A^C
- Intersection: $A \cap B$ (simply, AB)
- A, B are mutually disjoint $\iff A \cap B = \emptyset$

2.2 Axioms of Probability

2.2.1 Kolmogorov's Axiom

Kolmogorov's Axiom

Axiom. The probability is a function $\Pr : 2^\Omega \to [0,1] \subseteq \mathbb{R}$ satisfies

(A1) $\forall$event A, $0 \le \Pr[A] \le 1$.

(A2) $\Pr[\Omega] = 1$.

(A3) (Countable Additivity) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P[A_i]$, where $\{A_1, A_2, \dots\}$ is a countable set.

Remark 2.2. A probability is a function $\Pr : 2^\Omega \to [0,1] \subseteq \mathbb{R}$.

Proposition 2.1. *Let* $A, B \subseteq \Omega$.

(1) $\Pr[A] = \Pr[AB^C] + \Pr[AB]$

(2) $\Pr[B] = \Pr[AB] + \Pr[A^C B]$

(3) $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[AB]$

(4) $\Pr[A \cup B] = \Pr[AB^C] + \Pr[AB] + \Pr[A^C B]$

(5) $\Pr[A^C] = 1 - \Pr[A]$

(6) $A \subseteq B \implies \Pr[A] \le \Pr[B]$

2.2.2 Conditional Probability and Independent

Conditional Probability

Definition 2.2. The **conditional probability** of A given B is denoted by $\Pr[A|B]$ and defined by the formula

$$\Pr[A|B] = \frac{\Pr[AB]}{\Pr[B]} \quad \text{with} \quad \Pr[B] > 0.$$

Equivalently, this formula can be written as **multiplication law of probability**:

$$\Pr[AB] = \Pr[A|B]\Pr[B].$$

Example 2.1.

(1) Start with a *shuffled deck of cards* and distribute all 52 cards to 4 player, 13 cards to each. What is the probability that each player gets an Ace?

(2) Next, assume that you are a player and you get a single Ace. What is the probability now that each player gets an Ace?

Solution.

(1) If any ordering of cards is equally likely, then any position of the four Aces in the deck is also equally likely. There are

$$\binom{52}{4} = \frac{52!}{4!48!}$$

possibilities for the positions (slots) for the 4 aces. Out of these, the number of positions that give each player an Ace 13^4 pick the first slot among the cards that the first player gets, then the second slot among the second player's card, then the third and the fourth slot. Therefore, the answer is $\frac{13^4}{\binom{52}{4}} \approx 0.1055$.

(2) After you see that you have a single Ace, the probability goes up the previous answer need to be divided by the probability that you get a single Ace, which is

$$\frac{13 \cdot \binom{39}{3}}{\binom{52}{4}} \approx 0.4388.$$

Note that

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}.$$

The answer then becomes $\frac{13^4}{13 \cdot \binom{39}{3}} \approx 0.2404$.

□

Independence

Definition 2.3. Two events A and B are **independent** if

$$\Pr[A|B] = \Pr[A]$$

Equivalent conditions are

$$\Pr[B|A] = \Pr[B] \quad \text{or} \quad \Pr[AB] = \Pr[A]\Pr[B]$$

Remark 2.3. $\Pr[A] = \Pr[A|B] = \dfrac{\Pr[AB]}{\Pr[B]} \implies \Pr[AB] = \Pr[A]\Pr[B]$.

Example 2.2. Suppose we roll a dice once. Let the universal set is $U = \{1, 2, 3, 4, 5, 6\}$.

(1) (Independent but Not Disjoint) Let

$$A = \{1, 3, 5\} \quad \text{and} \quad B = \{3, 6\}\,.$$

Then $A \cap B = \{3\} \neq \emptyset$, that is, A and B are not disjoint. Note that

$$\begin{aligned} \Pr[A] &= \frac{3}{6} = \frac{1}{2}, & \Pr[B] &= \frac{2}{6} = \frac{1}{3}, \\ \Pr[A \mid B] &= \frac{\Pr[AB]}{\Pr[B]} = \frac{1/6}{1/3} = \frac{1}{2}, & \Pr[B \mid A] &= \frac{\Pr[BA]}{\Pr[A]} = \frac{1/6}{1/2} = \frac{1}{3}. \end{aligned}$$

Thus, $\Pr[A|B] = \Pr[A]$ and $\Pr[B|A] = \Pr[B]$. That is, A and B are mutually independent.

(2) (Not Independent but Disjoint) Let

$$A = \{1, 3, 5\} \quad \text{and} \quad B = \{2, 4, 6\}\,.$$

Then $A \cap B = \emptyset$, that is, A and B are disjoint. Note that

$$\begin{aligned} \Pr[A] &= \frac{3}{6} = \frac{1}{2}, & \Pr[B] &= \frac{3}{6} = \frac{1}{2}, \\ \Pr[A \mid B] &= \frac{\Pr[AB]}{\Pr[B]} = \frac{0}{1/2} = 0, & \Pr[B \mid A] &= \frac{\Pr[BA]}{\Pr[A]} = \frac{0}{1/2} = 0. \end{aligned}$$

Thus, $\Pr[A|B] \neq \Pr[A]$ and $\Pr[B|A] \neq \Pr[B]$. That is, A and B are not independent.

Rule of Total Probailtiy

Proposition 2.2. *Let events* $A_1, \dots, A_n$ *are satisfies*

(1) $\Pr[A_i] > 0$ *for* $i = 1, \dots, n$

(2) $A_i \cap A_j = \emptyset$ *for* $i \neq j$

(3) $\bigcup_{i=1}^{n} A_i = \Omega$

Then

$$\begin{aligned}\Pr[B] &= \sum_{i=1}^{n} \Pr[B|A_i]\Pr[A_i] \\ &= \Pr[B|A_1]\Pr[A_1] + \Pr[B|A_2]\Pr[A_2] + \cdots + \Pr[B|A_n]\Pr[A_n].\end{aligned}$$

Proof. $B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^{n} A_i\right) = \bigcup_{i=1}^{n}(B \cap A_i)$. □

2.2.3 Bayes' Theorem

Bayes' Theorem

Theorem 2.3.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$$

The posterior probability of $\bar{B}$ *is then* $P(\bar{B}|A) = 1 - P(B|A)$.

Remark 2.4.

$$\Pr[B \mid A] = \frac{\Pr[A|B] \cdot \Pr[B]}{\Pr[A]} \iff \text{Posterior} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Evidence}}.$$

2.3 Random Variables

Random Variable

Definition 2.4. A **random variable** X is real-valued function on Ω the space of outcomes:

$$X : \Omega \to \mathbb{R}.$$

In other words, a random variable is a number whose value depends upon the outcome of a random experiment.

Remark 2.5. Sometimes, when convenient, we also allow X to have the value ∞ or, more rarely, $-\infty$.

2.3.1 Discrete Random Variables

Discrete Random Variable

Definition 2.5. A **discrete random variable** X has finitely or countably many values

$$x_i \quad \text{for} \quad i = 1, 2, \cdots$$

and

$$p(x_i) = P(X = x_i)$$

with $i = 1, 2, \cdots$, is called the **probability mass function of** X.

Remark 2.6. A probability mass function p has the following properties:

(1) $x = x_i, i \in I \implies p(x) = \Pr[X = x_i]$

(2) $0 \leq p(x) \leq 1$, $\sum_{x \in X} p(x) = 1$.

(3) $\Pr[a < X \leq b] = \sum_{a < x \leq b} p(x)$.

Discrete Probability Distribution

Definition 2.6. The **probability distribution** of a discrete of a random variable X is described as the function

$$f(x_i) = P(X = x_i)$$

which gives the probability for each value and satisfies:

1. $0 \leq f(x_i) \leq 1$ for each value x_i of X
2. $\displaystyle\sum_{i=1}^{k} f(x_i) = 1$

Expectation(Mean) and Standard Deviation of a Probability Distribution

Definition 2.7.

- The **mean** of X or **population mean**

$$\begin{aligned} E[X] &= \mu \\ &= \sum(\text{Value} \times \text{Probability}) = \sum x_i f(x_i) \end{aligned}$$

  Here the sum extends over all the distinct values x_i of X.

- The **Variance and Standard Deviation of X** is given by

$$\begin{aligned} \sigma^2 &= \mathrm{Var}[X] = \sum (x_i - \mu)^2 f(x_i) \\ \sigma &= \mathrm{sd}[X] = +\sqrt{\mathrm{Var}[X]} \end{aligned}$$

- **Alternative Formula for Hand calculation:**

$$\sigma^2 = \sum x_i^2 f(x_i) - \mu^2$$

Example 2.3 (Calculating a Population Variance and Standard Deviation). Calculate the variance and the standard deviation of the distribution of X that appears in the left two columns of below table.

x	$f(x)$	$xf(x)$	$(x-\mu)$	$(x-\mu)^2$	$(x-\mu)^2 f(x)$	$x^2 f(x)$
0	.1	0	-2	4	.4	0
1	.2	.2	-1	1	.2	0.2
2	.4	.8	0	0	.0	1.6
3	.2	.6	1	1	.2	1.8
4	.1	.4	2	4	.4	1.6
Total	1.0	$2.0 = \mu$			$1.2 = \sigma^2$	$5.2 = \sum x^2 f(x)$

$$\begin{aligned} \mathrm{Var}(X) &= \sigma^2 = 1.2 \\ \mathrm{sd}(X) &= \sigma = \sqrt{1.2} = 1.095 \end{aligned}$$

$$\begin{aligned} \sigma^2 &= 5.2 - (2.0)^2 = 1.2 \\ \sigma &= \sqrt{1.2} = 1.095 \end{aligned}$$

2.3.2 Bernoulli

Note.

- The sample space $S = \{ \text{S}, \text{ F} \}$.
- The probability of success $p = P(S)$, the probability of failure $q = P(F)$.
- $0 \le p \le 1$, $q = 1 - p$.

Binomial Distribution

Definition 2.8. The **binomial distribution** with n trails and success probability p is described by the function

$$f(x) = P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}$$

for the possible values $x = 0, 1, \cdots, n$.

Example 2.4 (An Example of the Binomial Distribution). The elementary outcomes of 4 samples, the associated probabilities, and the value of X are listed as follows.

FFFF	SFFF	SSFF	SSSF	SSSS
	FSFF	SFSF	SSFS	
	FFSF	SFFS	SFSS	
	FFFS	FSSF	FSSS	
		FSFS		
		FFSS		

Value of X	0	1	2	3	4
Probability of each outcome	q^4	pq^3	p^2q^2	p^3q	p^4
Number of outcomes	$1 = \binom{4}{0}$	$4 = \binom{4}{1}$	$6 = \binom{4}{2}$	$4 = \binom{4}{1}$	$1 = \binom{4}{4}$

Value x	0	1	2	3	4
Probability $f(x)$	$\binom{4}{0}p^0q^4$	$\binom{4}{1}p^1q^3$	$\binom{4}{2}p^2q^2$	$\binom{4}{1}p^3q^1$	$\binom{4}{4}p^4q^0$

Mean and Standard Deviation of the Binomial Distribution

Definition 2.9.

$$X = X_1 + X_2 + \cdots + X_n \sim B(n, p)$$

- $E[X] = E[X_1] + \cdots + E[X_n] = np$
- $\mathrm{Var}[X] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] = npq$

The binomial distribution with n trials and success probability p has

$$\begin{aligned} \text{Mean} &= np \\ \text{Variance} &= npq = np(1-p) \\ \text{sd} &= \sqrt{npq} \end{aligned}$$

Covariance and Correlation Coefficient of Two Random Variables

Definition 2.10. Let X, Y be a random variables. Then

1. The covariance of them:

$$Cov(X, Y) = E[(X - \mu_1)(Y - \mu_2)]$$

2. The correlation coefficient of them:

$$Corr(X, Y) = E\left[\left(\frac{X - \mu_1}{\sigma_1}\right)\left(\frac{Y - \mu_2}{\sigma_2}\right)\right] = \frac{Cov(X, Y)}{\mathrm{sd}(X)\mathrm{sd}(Y)}$$

Remark 2.7. Note that $-1 \leq Corr(X, Y) \leq 1$ and

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_1)(Y - \mu_2)] \\ &= E[XY - \mu_2 X - \mu_1 Y + \mu_1\mu_2] \\ &= E[XY] - \mu_2 E[X] - \mu_1 E[Y] + \mu_1\mu_2 \\ &= E[XY] - \mu_1\mu_2. \end{aligned}$$

That is, $Cov(X, Y) = E[XY] - \mu_1\mu_2$.

Proposition 2.4.

(1) $Cov(aX + b, cY + d) = ac \cdot Cov(X, Y)$

(2) $Corr(aX + b, cY + d) = \begin{cases} Corr(X, Y) & : ac > 0 \\ -Corr(X, Y) & : ac < 0 \end{cases}$

Proof. (1)

$$\begin{aligned} Cov(aX+b, cY+d) &= E[(aX+b)-(a\mu_x+b)\cdot(cY+d-(c\mu_y+d))] \\ &= E[a(X-\mu_x)\cdot c(Y-\mu_y)] = acE[(X-\mu_x)(Y-\mu_y)] \\ &= ac\cdot Cov(X,Y). \end{aligned}$$

(2) Note that $\sigma_{aX+b} = \sqrt{\mathrm{Var}(aX+b)} = \sqrt{a^2\mathrm{Var}(X)} = |a|\,\sigma_X$. Similarly $\sigma_{cY+d} = |c|\,\sigma_Y$.

$$Corr(aX+b, cY+d) = \frac{Cov(aX+b, cY+d)}{\sigma_{aX+b}\sigma_{cY+d}} = \frac{ac\cdot Cov(X,Y)}{|a|\,\sigma_X|c|\,\sigma_Y} = \frac{ac}{|ac|}Corr(X,Y).$$

Hence, $Corr(aX+b, cY+d) = \begin{cases} Corr(X,Y) & \text{if } ac > 0 \\ -Corr(X,Y) & \text{if } ac < 0 \end{cases}$

□

Distribution of Sum of Two Probability Variables

Proposition 2.5.

(1) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

(2) $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

Two Probability Variables are Independent

Proposition 2.6.

(1) $E[XY] = E[X]\cdot E[Y]$

(2) $Cov(X,Y) = 0, Corr(X,Y) = 0$

(3) $Var(X \pm Y) = Var(X) + Var(Y)$

Proof. (1)

$$\begin{aligned} E[XY] &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j p(x_i, y_j) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j p_1(x_i) p_2(y_j) \\ &= \sum_{i=1}^{\infty} x_i p_1(x_i) \sum_{j=1}^{\infty} y_j p_2(y_j) \\ &= E[X]\cdot E[Y]. \end{aligned}$$

(2) $Cov(X,Y) = E[XY] - E[X]\cdot E[Y] = 0.$

□

2.3.3 Continuous Random Variables

Probability Density Function

Definition 2.11. The **probability density function** $f(x)$ describes the distribution of probability for a continuous random variable. It has the properties:

(1) The total area under the probability density curve is 1.

(2) $P[a \leq X \leq b]$ = area under the probability density curve between a and b.

(3) $f(x) \geq 0$ for all x.

Remark 2.8. With a continuous random variable, the probability that $X = x$ is **always** 0. It is only meaningful to speak about the probability that X lies in an interval.

Remark 2.9. $p(x)$ is called **probability density function** of continuous random variable X if $p(x)$ satisfies:

(i) $p(x) \geq 0, \displaystyle\int_{-\infty}^{\infty} p(x)\, dx = 1,$

(ii) $P(a \leq X \leq b) = \displaystyle\int_{a}^{b} p(x)\, dx.$

Note that

- For any constant c, $\displaystyle\int_{c}^{c} p(x)\, dx = 0.$
- $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$

Expectation of a Continuous Random Variable

Definition 2.12.

- Expectation(or Mean) of a Random Variable X
  - a Discrete Random variable: $E[X] = \sum_{i=1}^{\infty} x_i p(x_i)$
  - a Continuous Random variable: $E[X] = \int_{-\infty}^{\infty} xp(x)\, dx$
- Expectation and Median of a Continuous Random Variable X
  - Expectation($\mu = E[X]$): the balance point of the probability mass.
  - Median: the value of X that divides the area under the curve into halves.

2.3.4 Normal Random Variable

Normal Random Variable

Definition 2.13. A random variable is **Normal with parameter** $\mu \in \mathbb{R}$ **and** $\sigma^2 > 0$ or, in short, X **is** $N(\mu, \sigma^2)$, if its density is the function given below.

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

where $x \in (-\infty, \infty)$.

Proposition 2.7.

(1) For $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$, $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

(2) $EX = \mu$.

(3) $Var(X) = \sigma^2$.

2.3.5 The Normal Approximation to the Binomial

The Normal Approximation to the Binomial

Theorem 2.8. *When np and $np(1-p)$ are both large, say, greater than 15, the binomial distribution is well approximated by the normal distribution having mean $= np$ and sd $= \sqrt{np(1-p)}$. That is,*

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \text{ is approximately } N(0,1).$$

Mean and Standard Deviation of $\overline{X}$

Definition 2.14. The distribution of the sample mean, based on a random sample of size n, has

$$E[\overline{X}] = \mu \qquad (= \text{Population mean})$$

$$\text{Var}[\overline{X}] = \frac{\sigma^2}{n} \qquad \left(= \frac{\text{Population variance}}{\text{Sample size}}\right)$$

$$\text{sd}[\overline{X}] = \frac{\sigma}{\sqrt{n}} \qquad \left(= \frac{\text{Population standard deviation}}{\sqrt{\text{Sample size}}}\right)$$

2.4 Central Limit Theorem

2.4.1 CLT

Central Limit Theorem

Theorem 2.9. *Assume that $X, X_1, X_2, \ldots$ are independent, identically distributed random variables, with finite $\mu = EX$ and $\sigma^2 = Var[X]$. Then,*

$$\lim_{n\to\infty} \Pr\left[\frac{\sum_{i=1}^n X_i - \mu n}{\sigma\sqrt{n}} \leq x\right] = \Pr\left[Z \leq x\right],$$

where Z is standard Normal.

2.4.2 Laws of Large Numbers

Weak Law of Large Numbers

Theorem 2.10. *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$ and variance σ^2. Then, for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \Pr\left[\left|\frac{\sum_{i=1}^n X_i}{n} - \mu\right| \geq \varepsilon\right] = 0.$$

Strong Law of Large Numbers

Theorem 2.11. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with a finite first moment, $\mathbb{E}[X_i] = \mu$. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \quad \textit{almost surely as} \quad n \to \infty.$$

2.5 Problem: RBG $\longrightarrow$ RNG

Exercise 2.1 (Uniform Distribution)**.** Consider an algorithm

Step 1: Drive the RBG independently 4 times to generate a 4-bit integer value r.
Step 2: **If** $r < 10$ **then**
  return r
 else
  go to Step 1

Prove that

$$\Pr[\text{ouput} = n] = \frac{1}{10}$$

for $n = 0, 1, 2, \ldots, 9$.

Solution. Let

- $n \leq 2^k = m$ with $n = 10$, $k = 4$ and $m = 16$
- Output digit $= r \in [0, 9]$.

Then

output	1st iteration	2nd iteration	$\cdots$	step iteration
0	$\Pr[0] = \frac{1}{m}$	$\Pr[0] = \frac{1}{m} \cdot \frac{m-n}{m}$	$\cdots$	$\Pr[0] = \frac{1}{m} \cdot \left(\frac{m-n}{m}\right)^{\text{step}-1}$
1	$\Pr[1] = \frac{1}{m}$	$\Pr[1] = \frac{1}{m} \cdot \frac{m-n}{m}$	$\cdots$	$\Pr[1] = \frac{1}{m} \cdot \left(\frac{m-n}{m}\right)^{\text{step}-1}$
2	$\Pr[2] = \frac{1}{m}$	$\Pr[2] = \frac{1}{m} \cdot \frac{m-n}{m}$	$\cdots$	$\Pr[2] = \frac{1}{m} \cdot \left(\frac{m-n}{m}\right)^{\text{step}-1}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$n-1$	$\Pr[n-1] = \frac{1}{m}$	$\Pr[n-1] = \frac{1}{m} \cdot \frac{m-n}{m}$	$\cdots$	$\Pr[n-1] = \frac{1}{m} \cdot \left(\frac{m-n}{m}\right)^{\text{step}-1}$

Thus,

$$\begin{aligned}
\Pr[\text{output} = r] &= \frac{1}{m} + \frac{1}{m} \cdot \frac{m-n}{m} + \cdots + \frac{1}{m} \cdot \left(\frac{m-n}{m}\right)^s + \cdots \\
&= \frac{1}{m} \sum_{s=0}^{\infty} \left(\frac{m-n}{m}\right)^s = \frac{1}{m} \sum_{s=0}^{\infty} \left(1 - \frac{n}{m}\right)^s \\
&= \frac{1}{m} \cdot \frac{1}{1 - \left(1 - \frac{n}{m}\right)} = \frac{1}{m} \cdot \frac{m}{n} \\
&= \frac{1}{n}.
\end{aligned}$$

□

Chapter 3

Markov Chains

Markov Chain

Definition 3.1. Let

$$\langle X_n \rangle_{n\geq 0} := \{X_n : n = 0, 1, 2, \cdots\}$$

be a stochastic process over a countable set S. Let $\Pr[X]$ is the probability of the random variable X. Then $\langle X_n \rangle_{n\geq 0}$ satisfies **Markov property** if

$$\Pr[X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n] = \Pr[X_{n+1} = x_{n+1} \mid X_n = x_n]$$

for all $n \geq 0$ and all $x_0, \ldots, x_{n+1} \in S$. Then $\langle X_n \rangle_{n\geq 0}$ is a **Markov chain**.

Remark 3.1.

(1) The conditional probability of X_{i+1} is dependent only upon X_i, and upon no earlier values of $\langle X_n \rangle$

(2) the state of $\langle X_n \rangle$ in the future is unaffected by its history.

(3) The set S is called the **state space** of the Markov chain.

(4) The conditional probabilities $\Pr[X_{n+1} = y \mid X_n = x]$ are called the **transition probabilities**.

(5) Markov chain having **stationary transition probabilities**, i.e., $\Pr(X_{n+1} = y \mid X_n = x)$, is independent of n.

Example 3.1 (*The general two-state Markov chain*). There are two states 0 and 1 with transitions

- $0 \to 1$ with probability p
- $0 \to 0$ with probability $1-p$
- $1 \to 0$ with probability q
- $1 \to 1$ with probability $1-q$.

Thus we have

$$\Pr\left[X_{n+1}=1 \mid X_n=0\right]=p,$$
$$\Pr\left[X_{n+1}=0 \mid X_n=1\right]=q,$$

and $\Pr[X_0=0]=\pi_0(0)$. Since there are only two states, 0 and 1, it follows immediately that

$$\Pr\left[X_{n+1}=0 \mid X_n=0\right]=1-p,$$
$$\Pr\left[X_{n+1}=1 \mid X_n=1\right]=1-q,$$

and $\pi_0(1)=\Pr[X_0=1]=1-\pi_0(0)$. The transition matrix has two parameters $p, q \in [0,1]$:

$$\begin{bmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{bmatrix} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$

Note that

- $\Pr[A]=\Pr[B\cap A]+\Pr[B^C \cap A]$
- $\Pr[A\cap B]=\Pr[A]\cdot\Pr[B \mid A]$

Then we observe that

$$\begin{aligned}
\Pr\left[X_{n+1}=0\right] =& \Pr\left[X_n=0 \wedge X_{n+1}=0\right]+\Pr\left[X_n=1 \wedge X_{n+1}=0\right]\\
=& \Pr[X_n=0]\Pr[X_{n+1}=0 \mid X_n=0]\\
&+\Pr[X_n=1]\Pr[X_{n+1}=0 \mid X_n=1]\\
=& \Pr[X_n=0]\cdot(1-p)+\Pr[X_n=1]\cdot q\\
=&(1-p)\cdot\Pr[X_n=0]+q\cdot\big(1-\Pr[X_n=0]\big)\\
=&(1-p-q)\cdot\Pr[X_n=0]+q.
\end{aligned}$$

Now

$$\begin{aligned}
\Pr[X_0=0] &= \pi_0(0),\\
\Pr[X_1=0] &= (1-p-q)\pi_0(0)+q,\\
\Pr[X_2=0] &= (1-p-q)\Pr[X_1=0]+q\\
&= (1-p-q)^2\pi_0(0)+q\left(1+(1-p-q)\right),\\
\Pr[X_3=0] &= (1-p-q)\Pr[X_2=0]+q\\
&= (1-p-q)^3\pi_0(0)+q\left(1+(1-p-q)+(1-p-q)^2\right),\\
&\ \vdots\\
\Pr[X_n=0] &= (1-p-q)^n\pi_0(0)+q\sum_{j=0}^{n-1}(1-p-q)^j.
\end{aligned}$$

In the trivial case $p = q = 0$, it is clear that for all n

$$\Pr[X_n = 0] = \pi_0(0) \quad \text{and} \quad \Pr[X_n = 1] = \pi_0(1).$$

Suppose that $p + q > 0$. By the formula $\sum_{j=0}^{n-1} r^j = \frac{1-r^n}{1-r}$ for the sum of a finite geometric progression,

$$\sum_{j=0}^{n-1}(1-p-q)^j = \frac{1-(1-p-q)^n}{p+q}.$$

Thus,

$$\begin{aligned}
\Pr[X_n = 0] &= \frac{q}{p+q} + (1-p-q)^n \cdot \left(\pi_0(0) - \frac{q}{p+q}\right),\\
\Pr[X_n = 1] &= \frac{p}{p+q} + (1-p-q)^n \cdot \left(\pi_0(1) - \frac{p}{p+q}\right).
\end{aligned}$$

Suppose that $p, q \notin \{0, 1\}$. Then

$$0 < p + q < 2 \implies \left|1 - p - q\right| \le 1.$$

Then

$$\lim_{n\to\infty} \Pr[X_n = 0] = \frac{q}{p+q} \quad \text{and} \quad \lim_{n\to\infty} \Pr[X_n = 1] = \frac{p}{p+q}.$$

Suppose we want to choose $\pi_0(0)$ and $\pi_0(1)$ so that $\Pr[X_n = 0]$ and $\Pr[X_n = 1]$ are independent of n. To do this, we should choose $\pi_0(0) = q/(p+q)$ and $\pi_0(1) = p/(p+q)$. Thus if $\langle X_n\rangle_{n\ge 0}$ start with the initial distribution

$$\pi_0 = \Pr[X_n = 0] = \frac{q}{p+q} \quad \text{and} \quad \pi_0(1) = \frac{p}{p+q},$$

then for all n

$$\Pr[X_n = 0] = \frac{q}{p+q} \quad \text{and} \quad \Pr[X_n = 1] = \frac{p}{p+q}.$$

Example 3.2. Let $n = 2$ and $x_0, x_1, x_2 \in \{0, 1\}$. Then

$$\begin{aligned}
&\Pr[X_0 = x_0, X_1 = x_1, X_2 = x_2] \\
=&\Pr[X_0 = x_0, X_1 = x_1] \cdot \Pr[X_2 = x_2 \mid X_0 = x_0, X_1 = x_1] \\
=&\Pr[X_0 = x_0] \Pr[X_1 = x_1 \mid X_0 = x_0] \cdot \Pr[X_2 = x_2 \mid X_0 = x_0, X_1 = x_1].
\end{aligned}$$

If the Markov property is satisfied, then

$$\Pr[X_2 = x_2 \mid X_0 = x_0, X_1 = x_1] = \Pr[X_2 = 2 \mid X_1 = x_1],$$

which is determined by p and q. In this case

$$\Pr[X_0 = x_0, X_1 = x_1, X_2 = x_2] = \Pr[X_0 = x_0] \Pr[X_1 = x_1 \mid X_0 = x_0] \Pr[X_2 = x_2 \mid X_1 = x_1].$$

Recall that the transition matrix with $p, q \in [0, 1]$:

$$\begin{bmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{bmatrix} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$

Then

x_0	x_1	x_2	$\Pr[X_0 = x_0, X_1 = x_1, X_2 = x_2]$
0	0	0	$\pi_0(0)(1-p)^2$
0	0	1	$\pi_0(0)(1-p)p$
0	1	0	$\pi_0(0)pq$
0	1	1	$\pi_0(0)p(1-q)$
1	0	0	$(1-\pi_0(0))q(1-p)$
1	0	1	$(1-\pi_0(0))qp$
1	1	0	$(1-\pi_0(0))(1-q)q$
1	1	1	$(1-\pi_0(0))(1-q)^2$

Chapter 4

Examples of Markov chains

4.1 Random Walk

Let $\xi_1, \xi_2, \ldots$ be independent integer-valued random variables having common density f. Let X_0 be an integer-valued random variable that is independent of the ξ's and set $X_n = X_0 + \xi_1 + \cdots + \xi_n$. The sequence X_n, $n \geq 0$, is called a *random walk*. It is a Markov chain whose state space is the integers and whose transition function is given by

$$p(x, y) = f(y - x).$$

For all $k \in \mathbb{Z}$, we have $\Pr[\xi_i = k] = f(k)$.

Proof. Let π_0 denote the distribution of X_0. Then

$$\begin{aligned}
\Pr(X_0 = x_0, \ldots, X_n = x_n) &= \Pr(X_0 = x_0, \xi_1 = x_1 - x_0, \ldots, \xi_n = x_n - x_{n-1}) \\
&= \Pr(X_0 = x_0) \Pr(\xi_1 = x_1 - x_0) \cdots \Pr(\xi_n = x_n - x_{n-1}) \\
&= \pi_0(x_0) f(x_1 - x_0) \cdots f(x_n - x_{n-1}) \\
&= \pi_0(x_0) \Pr(x_0, x_1) \cdots \Pr(x_{n-1}, x_n).
\end{aligned}$$

□

Remark 4.1. Suppose a "particle" moves along the integers according to this Markov chain. Whenever the particle is in x, regardless of how it got there, it jumps to state y with probability $f(y - x)$.

As a special case, consider a *simple random walk* in which $f(1) = p$, $f(-1) = q$, and $f(0) = r$, where p, q, and r are nonnegative and sum to one. The transition function is given by

$$p(x, y) = \begin{cases} p & : y = x + 1 \\ q & : y = x - 1 \\ r & : y = x \\ 0 & \text{elsewhere} \end{cases}, \quad \begin{cases} \Pr[\xi_i = 1] = f(1) = p \\ \Pr[\xi_i = -1] = f(-1) = q \\ \Pr[\xi_i = 0] = f(0) = r \\ p + q + r = 1. \end{cases}$$

Let a particle undergo such a random walk. If the particle is in state x at a given observation, then by the next observation it will have jumped to state $x + 1$ with probability p and to state $x - 1$ with probability q; with probability r it will still be in state x.

$X_0, X_1, X_2, \ldots$ are i.i.d. random variables with $P(X_i = +1) = P(X_i = -1) = 1/2$.

$$S_n = X_0 + X_1 + \ldots + X_n$$

$\{S_n, n \geq 1\} = \{S_1, S_2, \ldots\}$: simple random walk (Markov chain)

Definition. The discrete random variables $X_1, X_2, \ldots$ on $\mathbb{Z}^d$ are called steps of the random walk and have the following probability distribution:

$$P(X_i = e) = \begin{cases} \frac{1}{2d} & \text{if } e \in \mathbb{Z}^d \text{ and } \|e\| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Definition. $S_0 = 0$ in $\mathbb{Z}^d$ and $S_n = X_1 + \ldots + X_n$ for $n \in \mathbb{N}$ is called the position of the random walk at time n.

Pólya's Theorem. Simple random walks of dimension $d = 1, 2$ are recurrent, and of $d \geq 3$ are transient.

4.2 Ehrenfest Chain

Example 2. [Ehrenfest chain]. The following is a simple model of the exchange of heat or of gas molecules between two isolated bodies. Suppose we have two boxes, labeled 1 and 2, and d balls labeled 1, 2, ..., d. Initially some of these balls are in box 1 and the remainder are in box 2. An integer is selected at random from $1, 2, \ldots, d$, and the ball labeled by that integer is removed from its box and placed in the opposite box. This procedure is repeated indefinitely with the selections being independent from trial to trial. Let X_n denote the number of balls in box 1 after the n-th trial. Then $X_n, n \geq 0$, is a Markov chain on $\mathcal{S} = \{0, 1, 2, \ldots, d\}$.

The transition function of this Markov chain is easily computed. Suppose that there are x balls in box 1 at time n. Then with probability $\frac{x}{d}$ the ball drawn on the $(n+1)$-th trial will be from box 1 and will be transferred to box 2. In this case there will be $x - 1$ balls in box 1 at time $n + 1$. Similarly, with probability $\frac{d-x}{d}$ the ball drawn on the $(n+1)$-th trial will be from box 2 and will be transferred to box 1, resulting in $x + 1$ balls in box 1 at time $n + 1$. Thus the transition function of this Markov chain is given by

$$p(x, y) = \begin{cases} x/d & : y = x - 1 \\ 1 - x/d & : y = x + 1 \\ 0 & \text{elsewhere} \end{cases}$$

Note that the Ehrenfest chain can in one transition only go from state x to $x - 1$ or $x + 1$ with positive probability.

4.3 Gambler's Ruin Chain

A state of a Markov chain is called an *absorbing state* if $P(a, a) = 1$ or, equivalently, if $P(a, y) = 0$ for $y \neq a$. The next example uses this definition.

Suppose a gambler starts out with a certain initial capital in dollars and makes a series of one dollar bets against the house. Assume that he has respective probabilities p and $q = 1 - p$ of winning and losing each bet, and that if his capital ever reaches zero, he is ruined and his capital remains zero thereafter. Let X_n denote the gambler's capital at time n. This is a Markov chain in which 0 is an absorbing state, and for $x \geq 1$

$$p(x, y) = \begin{cases} q & : y = x - 1 \\ p & : y = x + 1 \\ 0 & \text{elsewhere} \end{cases}$$

Such a chain is called a *gambler's ruin chain* on $\mathcal{S} = \{0, 1, 2, \ldots\}$. We can modify this model by supposing that if the capital of the gambler increases to d dollars he quits playing. In this case 0 and d are both absorbing states, and (15) holds for $x = 1, \ldots, d - 1$.

4.4 Birth and Death Chain

Example 4. [Birth and death chain]. Consider a Markov chain either on $\mathcal{S} = \{0, 1, 2, \ldots\}$ or on $\mathcal{S} = \{0, 1, \ldots, d\}$ such that starting from x the chain will be at $x - 1$, x, or $x + 1$ after one step. The transition function of such a chain is given by

$$p(x, y) = \begin{cases} q_x & : y = x - 1 \\ r_x & : y = x \\ p_x & : y = x + 1 \\ 0 & \text{elsewhere} \end{cases}$$

where p_x, q_x, and r_x are nonnegative numbers such that $p_x + q_x + r_x = 1$. The Ehrenfest chain and the two versions of the gambler's ruin chain are examples of birth and death chains. The phrase "birth and death" stems from applications in which the state of the chain is the population of some living system. In these applications a transition from state x to state $x + 1$ corresponds to a "birth," while a transition from state x to state $x - 1$ corresponds to a "death."

4.5 Queuing Chain

Example 5. [Queueing chain]. Consider a service facility such as a checkout counter at a supermarket. People arrive at the facility at various times and are eventually served. Those customers that have arrived at the facility but have not yet been served form a waiting line or queue. There are a variety of models to describe such systems. We will consider here only one very simple and somewhat artificial model.

Let time be measured in convenient periods, say in minutes. Suppose that if there are any customers waiting for service at the beginning of any given period, exactly one customer will be served during that period, and that if there are no customers waiting for service at the beginning of a period, none will be served during that period. Let ξ_n denote the number of new customers arriving during the n-th period. We assume that $\xi_1, \xi_2, \ldots$ are independent nonnegative integer-valued random variables having common density f.

Let X_0 denote the number of customers present initially, and for $n \geq 1$, let X_n denote the number of customers present at the end of the n-th period. If $X_n = 0$, then $X_{n+1} = \xi_{n+1}$; and if $X_n \geq 1$, then $X_{n+1} = X_n + \xi_{n+1} - 1$. It follows without difficulty from the assumptions on $\xi_n, n \geq 1$, that $X_n, n \geq 0$, is a Markov chain whose state space is the nonnegative integers and whose transition function P is given by

$$P(0, y) = f(y)$$

and

$$P(x, y) = f(y - x + 1), \quad x \geq 1.$$

$$\begin{aligned} p(x, y) &= p(x_{n+1} = y \mid x_n = x) \\ &= p(x_n + \xi_{n+1} - 1 = y \mid x_n = x) \\ &= p(\xi_{n+1} = y - x + 1) \\ &= f(y - x + 1) \end{aligned}$$

4.6 Branching Chain

Example 6. [Branching chain]. Consider particles such as neutrons or bacteria that can generate new particles of the same type. The initial set of objects is referred to as belonging to the 0th generation. Particles generated from the nth generation are said to belong to the $(n+1)$th generation. Let X_n, $n \geq 0$, denote the number of particles in the nth generation.

Nothing in this description requires that the various particles in a generation give rise to new particles simultaneously. Indeed at a given time, particles from several generations may coexist.

A typical situation is illustrated in Figure 1: one initial particle gives rise to two particles. Thus $X_0 = 1$ and $X_1 = 2$. One of the particles in the first generation gives rise to three particles and the other gives rise to one particle, so that $X_2 = 4$. We see from Figure 1 that $X_3 = 2$. Since neither of the particles in the third generation gives rise to new particles, we conclude that $X_4 = 0$ and consequently that $X_n = 0$ for all $n \geq 4$. In other words, the progeny of the initial particle in the zeroth generation become extinct after three generations.

In order to model this system as a Markov chain, we suppose that each particle gives rise to ξ particles in the next generation, where ξ is a non-negative integer-valued random variable having density f. We suppose that the number of offspring of the various particles in the various generations are chosen independently according to the density f.

Under these assumptions $X_n, n \geq 0$, forms a Markov chain whose state space is the nonnegative integers. State 0 is an absorbing state. For if there are no particles in a given generation, there will not be any particles in the next generation either. For $x \geq 1$

$$P(x, y) = P(\xi_1 + \cdots + \xi_x = y),$$

where $\xi_1, \xi_2, \ldots$ are independent random variables having common density f. In particular, $P(1, y) = f(y)$, $y \geq 0$.

If a particle gives rise to $\xi = 0$ particles, the interpretation is that the particle dies or disappears. Suppose a particle gives rise to ξ particles, which in turn give rise to other particles; but after some number of generations, all descendants of the initial particle have died or disappeared (see Figure 1). We describe such an event by saying that the descendants of the original particle eventually become extinct. An interesting problem involving branching chains is to compute the probability p of eventual extinction for a branching chain starting with a single particle, or, equivalently, the probability that a branching chain starting at state 1 will eventually be absorbed at state 0. Once we determine p, we can easily find the probability that in a branching chain starting with x particles the descendants of each of the original particles eventually become extinct. Indeed, since the particles are assumed to act independently in giving rise to new particles, the desired probability is just p^x.

Example 7. Consider a gene composed of d subunits, where d is some positive integer and each subunit is either normal or mutant in form. Consider a cell with a gene composed of m mutant subunits and $d - m$ normal subunits. Before the cell divides into two daughter cells, the gene duplicates. The corresponding gene of one of the daughter cells is composed of d units chosen at random from the $2m$ mutant subunits and the $2(d-m)$ normal subunits. Suppose we follow a fixed line of descent from a given gene. Let X_0 be the number of mutant subunits initially present, and let $X_n, n \geq 1$, be the number present in the n-th descendant gene. Then $X_n, n \geq 0$, is a Markov chain on $\mathcal{S} = \{0, 1, 2, \ldots, d\}$ and

States 0 and d are absorbing states for this chain.

Chapter 5

Statistical Inferences

Chapter 6

Statistical Tests for Randomness

- Some statistical tests are designed to measure the *quality* of a generator purported to be a random bit generator.
- While it is impossible to give a **mathematical proof** that a generator is indeed a random bit generator, the statistical tests help **detect certain kinds of weaknesses the generator may have**.
- This is accomplished by taking a sample output sequence of the generator and subjecting it to various statistical tests.
  - Each statistical test determines whether the sequence possesses a certain attribute that a truly random sequence would be likely to exhibit; the conclusion of each test is not definite, but rather *probabilistic*.
  - If the sequence is *deemed* (regarded) to have failed any one of the statistical tests, the generator may be rejected as being non-random; alternatively, the generator may be subjected to further testing.
  - On the other hand, if the sequence passes all of the statistical tests, the generator is *accepted* as being random. More precisely, the term "accepted" should be replaced by "not rejected", since passing the tests merely provides *probabilistic evidence* that the generator produces sequences with certain characteristics of random sequences.

6.1 Periodic Pseudo-Noise Sequence

Golomb's Randomness Postulates are historical attempts to define necessary conditions for periodic pseudorandom sequences to appear random. They are not sufficient conditions for randomness but were among the first efforts to systematically address the randomness in sequences. These postulates serve as a fundamental basis for more complex tests and are critical in understanding the nature of pseudorandom sequences and their applications.

Note (Symbol)**.** Let $s = s_0, s_1, s_2, \ldots$ be an infnite sequence. The subsequence consisting of the first n terms of s is denoted by

$$s^n = s_0, s_1, \cdots, s_{n-1}.$$

Especially, s is the bit sequence if $s_i \in \{0, 1\}$.

Note (N-Periodic)**.** The sequence $s = s_0, s_1, s_2, \ldots$ is said to be N**-periodic** if $s_i = s_{i+N}$ for all $i \geq 0$. If s is a N-periodic sequence, then the **cycle** of s is the subsequence s^N.

Run - Gap / Block

Definition 6.1. Let s be a sequence.

- A **run** of s is a subsequence of s consisting of consecutive 0's or 1's.
- A run of 0's and 1's are a **gap** and a **block**, respectively.

Autocorrelation Function

Definition 6.2. Let s be a N-periodic sequence. The **autocorrelation function** of s is the integer-value function $C(t) : \{s_i\} \longrightarrow \mathbb{Z}$ defined as

$$C(t) = \frac{1}{N} \sum_{i=0}^{N-1} (2s_i - 1) \cdot (2s_{i+t} - 1) \quad \text{for} \quad 0 \leq t \leq N - 1.$$

Remark 6.1. The autocorrelation function $C(t)$ measure the amount of similarity between the sequence s and a shift of s by t positions. If s is a random N-periodic sequence, then $|N \cdot C(t)|$ can be expected to be quite small for all vlaue of $t \in (0, N)$.

Golomb's Randomness Postulates

Definition 6.3. Let s be a N-periodic sequence. **Golomb's randomness postulates** are as follows:

R1 *Balance*: $|\#1(s^N) - \#0(s^N)| \leq 1$, where #1 and #0 count the number of 1's and 0's.

R2 *Run Distribution*: For runs of length l in s^N, at least $\frac{1}{2^l}$ of runs are of length l, for $l \geq 1$ and total runs > 1, with nearly equal numbers of gap and block.

R3 *Autocorrelation Property*: $C(t)$:

$$N \cdot C(t) = \sum_{i=0}^{N-1} (2s_i - 1) \cdot (2s_{i+t} - 1) = \begin{cases} N, & : t = 0, \\ K \in \mathbb{Z}, & : 1 \leq t \leq N - 1. \end{cases}$$

Pseudo-Noise Sequence (pn-sequence)

Definition 6.4. A binary sequence which satisfies Golomb's randomness postulates is called a **pseudo-noise sequence** (**pn-sequence**).

Remark 6.2. Pseudo-noise sequences arise in practice as output sequences of **maximum-length linear feedback shift registers (m-LFSR)**

Example 6.1. Consider the periodic sequence s of period $N = 15$ with cycle

$$s^{15} = 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1.$$

The following shows that the sequence s satisfies Golomb's randomness postulates.

R1: The number of 0's in s^{15} is 7, while the number of 1's is 8.

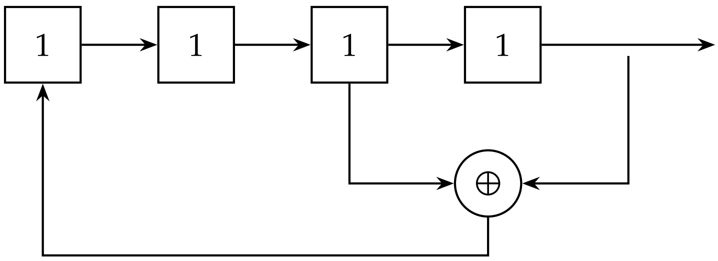
R2: s^{15} has 8 runs. There are 4 runs of length 1 (2 gaps and 2 blocks), 2 runs of length 2 (1 gap and 1 block), 1 run of length 3 (1 gap), and 1 run of length 4 (1 block).

R3: $C(0) = 15$ and $C(t) = -1$ for $1 \leq t \leq 14$.

Hence, s is a pn-sequence.

Note (LFSR (4, $1 + D^3 + D^4$**)).**

- Connection polynomial: $1 + D^3 + D^4$
- Initial state: $[1, 1, 1, 1]$



LFSR $(4, 1 + D^3 + D^4)$

State	$X(t-1)$	$X(t-2)$	$X(t-3)$	$X(t-4)$	$X(t)$
0	1	1	1	1	0
1	0	1	1	1	0
2	0	0	1	1	0
3	0	0	0	1	1
4	1	0	0	0	0
5	0	1	0	0	0
6	0	0	1	0	1
7	1	0	0	1	1
8	1	1	0	0	0
9	0	1	1	0	1
10	1	0	1	1	0
11	0	1	0	1	1
12	1	0	1	0	1
13	1	1	0	1	1
14	1	1	1	0	1
15	1	1	1	1	0

6.2 Five Basic Tests

The test aims to determine if the sequence s contains roughly the same number of 0's and 1's. In a random sequence, you would expect these numbers to be about equal.

6.2.1 Frequency Test (Monobit Test)

Monobit Test

Let $s = s_0, s_1, \ldots, s_{n-1}$ be a binary sequence of length n. Let $n_0 := \#0(s), n_1 := \#1(s)$. The statistic used is

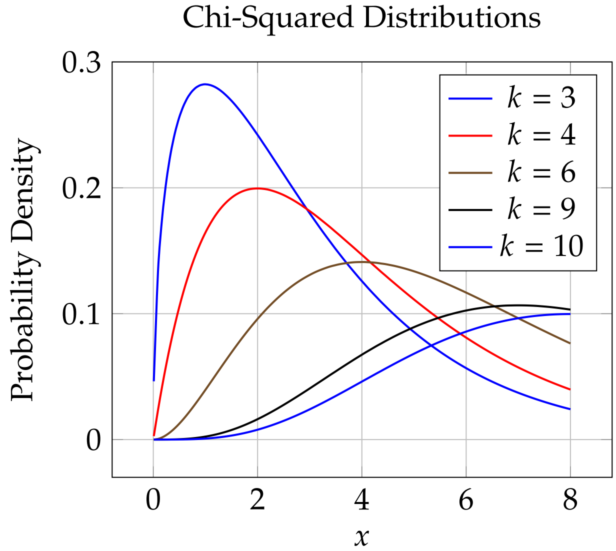
$$X_1 = \frac{(n_0 - n_1)^2}{n},$$

which approximately follows a χ^2 distribution with 1 degree of freedom if $n \geq 10$.

Remark 6.3 (Chi-Squared Distribution)**.** Let $Z_1, \cdots, Z_k$ be independent standard normal random variables. Then

$$Q := \sum_{i}^{k} Z_i^2 \sim \chi^2(k).$$

is said to have a chi-squared distribution with k degrees of freedom.



Remark 6.4 (**Interpretation**)**.**

- **Components:**
  - $n_0 - n_1$: The difference in the counts of 0's and 1's.
  - $(n_0 - n_1)^2$: Ensure the result is always non-negative.
  - n: This is the total length of the sequence s, i.e., $n_0 + n_1$.
- **High X_1 Value:** A high value of X_1 indicates a large difference between n_0 and n_1, suggesting that the sequence may not be random.
- **Low X_1 Value:** A low value of X_1 suggests that n_0 and n_1 are approximately equal, consistent with a random sequence.

6.2.2 Serial Test (Two-Bit Test)

The test assesses whether the sequence s is random by checking the frequency of adjacent pairs of bits $(00, 01, 10, 11)$.

Two-Bit Test

Let $n_{00}, n_{01}, n_{10}, n_{11}$ denote the number of occurrences of 00, 01, 10, 11 in s, respectively. These counts allow for overlapping. For example, in the sequence 00110, 00 and 01 overlap; they are not counted as separate, non-overlapping pairs. Then

$$n_{00} + n_{01} + n_{10} + n_{11} = (n - 1).$$

The statistic used is

$$X_2 = \frac{4}{n-1}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1$$

which approximately follows a χ^2 distribution with 2 degrees of freedom if $n \geq 21$.

Remark 6.5 (**Interpretation**).

- **Components:**
  - $\frac{4}{n-1}$: Scales the squared frequencies of the pairs relative to $n - 1$.
  - $n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2$: Sum of the squares of the frequencies of each pair.
  - $\frac{2}{n}$: Scales the squared frequencies of individual bits (0's and 1's).
  - $n_0^2 + n_1^2$: Sum of the squares of the frequencies of 0's and 1's.
  - The "+1" adjusts the statistic to align it with the χ^2 distribution.
- **High X_2 Value:** Indicates a significant imbalance among the counts of the subsequences, suggesting non-randomness.
- **Low X_2 Value:** Indicates that the counts of the subsequences are balanced, consistent with a random sequence.

Example 6.2. Let

$$s : \texttt{11100} \quad \texttt{01100} \quad \texttt{01000} \quad \texttt{10100} \quad \texttt{11101} \quad \texttt{11100} \quad \texttt{10010} \quad \texttt{01001}$$

with $n = 40$. Then

$$n_0 = 21,\ n_1 = 19 \longrightarrow X_1 = (21 - 19)^2/40 = 0.1$$

$$\begin{cases} n_{00} = 11 \\ n_{01} = 10 \\ n_{10} = 10 \\ n_{11} = 8 \end{cases} \longrightarrow X_2 = \frac{4}{39}(121 + 100 + 100 + 64) - \frac{2}{40}\left(21^2 + 19^2\right) + 1 \approx 0.387$$

6.2.3 Poker Test

Poker Test

Let m be a positive integer such that $\lfloor \frac{n}{m} \rfloor \geq 5 \cdot (2^m)$, and let $k = \lfloor \frac{n}{m} \rfloor$. Divide the sequence s into k non-overlapping parts each of length m, and let n_i be the number of occurrences of the i-th type of sequence of length m, $1 \leq i \leq 2^m$. The poker test determines whether the sequences of length m each appear approximately the same number of times in s, as would be expected for a random sequence. The statistic used is

$$X_3 = \frac{2^m}{k}\left(\sum_{i=1}^{2^m} n_i^2\right) - k$$

which approximately follows a χ^2 distribution with $2^m - 1$ degrees of freedom. Note that the poker test is a generalization of the frequency test: setting $m = 1$ in the poker test yields the frequency test.

Remark 6.6.

$$m = 1 \implies \begin{cases} k = n \\ n = n_0 + n_1 \end{cases} \implies X_3 = \frac{2}{n}\left(\sum_{i=1}^{2} n_i^2\right) - n = \frac{(n_0 - n_1)^2}{n} = X_1.$$

6.2.4 Runs Test

Runs Test

The purpose of the runs test is to determine whether the number of runs of various lengths in the sequence s is as expected for a random sequence. The expected number of gaps (or blocks) of length i in a random sequence of length n is

$$e_i = \frac{n - i + 3}{2^{i+2}}$$

for $1 \leq i \leq k$. Let k be equal to the largest integer i for which $e_i \geq 5$. Let B_i, G_i be the number of blocks and gaps, respectively, of length i in s for each $i, 1 \leq i \leq k$. The statistic used is

$$X_4 = \sum_{i=1}^{k} \frac{(B_i - e_i)^2}{e_i} + \sum_{i=1}^{k} \frac{(G_i - e_i)^2}{e_i}$$

which approximately follows a χ^2 distribution with $2k - 2$ degrees of freedom.

Remark 6.7.

$$\begin{aligned} \mathbb{E}[\text{ran of length } i] &= \frac{n - (i+1)}{2^{i+2}} + \frac{2}{2^{i+1}} \\ &= \frac{n - i + 3}{2^{i+2}} \end{aligned}$$

6.2.5 Autocorrelation Test

Autocorrelation Test

The purpose of this test is to check for correlations between the sequence s and (non-cyclic) shifted versions of it. Let d be a fixed integer, $1 \leq d \leq \lfloor \frac{n}{2} \rfloor$. The number of bits in s not equal to their d-shifts is $A(d) = \sum_{i=0}^{n-d-1} s_i \oplus s_{i+d}$, where $\oplus$ denotes the XOR operator. The statistic used is

$$X_5 = \frac{2(A(d) - \frac{n-d}{2})}{\sqrt{n-d}}$$

which approximately follows an $\mathcal{N}(0,1)$ distribution if $n - d \geq 10$. Since small values of $A(d)$ are as unexpected as large values of $A(d)$, a two-sided test should be used.

Remark 6.8.

$$\mathbb{E}[A(d)] = \frac{n-d}{2}, \quad \mathrm{Var}[A(d)] = \frac{n-d}{4}$$

Example 6.3. Consider the (non-random) sequence s of length $n = 160$ obtained by replicating the following sequence four times:

11100 01100 01000 10100 11101 11100 10010 01001.

(i) **Frequency test (monobit test):** $n_0 = 84$, $n_1 = 76$, and the value of the statistic X_1 is 0.4.

(ii) **Serial test (two-bit test):** $n_{00} = 44$, $n_{01} = 40$, $n_{10} = 40$, $n_{11} = 35$, and the value of the statistic X_2 is 0.6252.

(iii) **Poker test:** Here $m = 3$ and $k = 53$. The blocks 000, 001, 010, 011, 100, 101, 110, 111 appear 5, 10, 6, 4, 12, 3, 6, and 7 times, respectively, and the value of the statistic X_3 is 9.6415.

(iv) **Runs test:** Here $e_1 = 20.25$, $e_2 = 10.0625$, $e_3 = 5$, and $k = 3$. There are 25, 4, 5 blocks of lengths 1, 2, 3, respectively, and 8, 20, 12 gaps of lengths 1, 2, 3, respectively. The value of the statistic X_4 is 31.7813.

(v) **Autocorrelation test:** If $d = 8$, then $A(8) = 100$. The value of the statistic X_5 is 3.8933.

For a significance level of $\alpha = 0.05$, the threshold values for X_1, X_2, X_3, X_4, and X_5 are 3.8415, 5.9915, 14.0671, 9.4877, and 1.96, respectively.

Hence, the given sequence s passes the frequency, serial, and poker tests, but fails the runs and autocorrelation tests.

6.3 FIPS 140-1 Statistical Tests for Randomness

- FIPS (Federal Information Processing Standards) 140-1 specifies four statistical tests for randomness.
- Instead of making the user select appropriate significance levels for these tests, explicit bounds are provided that the computed value of a statistic must satisfy.
- A single bitstring s of length n output from a generator, is subjected to each of the following tests.
- If any of the tests fail, then the generator fails the test.
  - monobit test
  - poker test
  - runs test
  - long run test
- For high security applications, FIPS 140-1 mandates that the four tests be performed each time the random bit generator is powered up.
- FIPS 140-1 allows these tests to be substituted by alternative tests which provide equivalent or superior randomness checking.

Test details

(i) **monobit test:** The number of 1's in s should satisfy $9654 < n_1 < 10346$.

(ii) **poker test:** The statistic X_3 defined by equation $X_3 = \frac{2^m}{k}\left(\sum_{i=1}^{2^m} n_i^2\right) - k$ is computed for $m = 4$. The poker test is passed if $1.03 < X_3 < 57.4$.

(iii) **runs test:** The number B_i and G_i of blocks and gaps, respectively, of length i in s are counted for each i, $1 \leq i \leq 6$. (For the purpose of this test, runs of length greater than 6 are considered to be of length 6.) The runs test is passed if the 12 counts $B_i, G_i, 1 \leq i \leq 6$, are each within the corresponding interval specified by the following table.

Length of run	Required interval
1	2267 – 2733
2	1079 – 1421
3	502 – 748
4	223 – 402
5	90 – 223
6	90 – 223

(iv) **long run test:** The long run test is passed if there are no runs of length 34 or more.

6.4 (Probabilistic) Randomness

A **random bit sequence** could be interpreted as:

- The result of the flips of an **unbiased "fair"** coin with sides that are labeled "0" and "1," with each flip having a **probability** of exactly $\frac{1}{2}$ of producing a "0" or "1."
- Furthermore, the flips are **independent** of each other: the result of any previous coin flip does not affect future coin flips.
- The unbiased "fair" coin is thus the **perfect random bit stream generator**, since the "0" and "1" values will be randomly distributed.
- All elements of the sequence are generated **independently** of each other, and the value of the next element in the sequence **cannot be predicted**, regardless of how many elements have already been produced.

The use of unbiased coins for cryptographic purposes is **impractical**.

- Nonetheless, the hypothetical output of such an idealized generator of a true random sequence serves as a benchmark for the evaluation of random and pseudorandom number generators.

6.5 (Cryptologic) Unpredictability

- Random and pseudorandom numbers generated for cryptographic applications should be **unpredictable**.
  - In the case of PRNGs, if the seed is unknown, the next output number in the sequence should be **unpredictable** in spite of any knowledge of previous random numbers in the sequence. This property is known as **forward unpredictability**.
  - It should also not be **feasible to determine** the seed from knowledge of any generated values, i.e., **backward unpredictability**
  - No correlation between a seed and any value generated from that seed should be evident.
- To ensure forward unpredictability, care must be exercised in obtaining **seeds**.
  - The values produced by a PRNG are **completely predictable** if the seed and generation algorithm are known.
  - Since in many cases the generation algorithm is publicly available, the **seed must be kept secret** and should not be derivable from the pseudorandom sequence that it produces. In addition, the seed itself must be unpredictable.

6.6 NIST SP 800-22

The publication is central to understanding the statistical test suite designed for evaluating random and pseudorandom number generators (RNGs) used in cryptographic applications.

Strategies for Statistical Analysis of an RNG

The strategy involves five key stages:

1. **Selection of a Generator:** Choosing a suitable hardware or software-based RNG.

2. **Binary Sequence Generation:** Generating a set of binary sequences using the selected RNG.

3. **Execution of the Statistical Test Suite:** Applying the NIST Statistical Test Suite to the generated sequences.

4. **Examination of P-values:** Analyzing the P-values obtained from the test suite to evaluate the quality of the sequences.

5. **Assessment: Pass/Fail Assignment:** Determining whether each sequence passes or fails the statistical tests based on P-value thresholds.

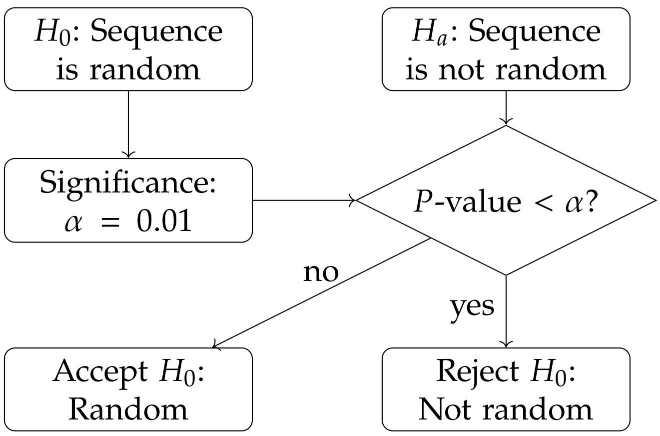
Remark 6.9.

- $H0$ (null hypothesis) : The sequence being tested is random.

- H_a (alternative hypothesis) : The sequence is not random.

- Level of significance : $\alpha = 0.01$ (α is chosen in the range $[0.001, 0.01]$.)

  - P-value $< \alpha = 0.01 \implies$ Reject H_0 (not random)
  - P-value $\geq \alpha = 0.01 \implies$ Accept H_0 (random)



6.6.1 Proportion of Sequences Passing a Test

Given the empirical results for a particular statistical test, compute the proportion of sequences that pass. For example, if 1000 binary sequences were tested (i.e., $n = 1000$), $\alpha = 0.01$ (the significance level), and 996 binary sequences had P-values > 0.01, then the proportion is $\frac{996}{1000} = 0.996$.

The range of acceptable proportions is determined using the confidence interval defined as,

$$\hat{p} \pm 4\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $\hat{p} = 1 - \alpha$, and n is the sample size. If the proportion falls outside of this interval, then there is evidence that the data is non-random. Note that other standard deviation values could be used.

6.7 Uniform Distribution of P-values

The distribution of P-values is examined to ensure uniformity. This may be visually illustrated using a histogram, whereby, the interval between 0 and 1 is divided into 10 sub-intervals, and the P-values that lie within each sub-interval are counted and displayed.

Uniformity may also be determined via an application of a χ^2 test and the determination of a P-value corresponding to the Goodness-of-Fit Distributional Test on the P-values obtained for an arbitrary statistical test (i.e., a P-value of the P-values). This is accomplished by computing

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where F_i is the number of P-values in sub-interval i, and s is the sample size. A P-value is calculated such that $P\text{-value}_{\chi^2} = \text{igamc}(\frac{k}{2}, \frac{\chi^2}{2})$. If $P\text{-value}_{\chi^2} < \alpha$, then the sequences can be considered to be uniformly distributed. Additionally, to provide statistically meaningful results, at least 55 sequences must be processed.

6.7.1 Useful Functions

- **Standard Normal (Cumulative Distribution) Function**

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du.$$

- **Complementary Error Function**

$$\text{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_{z}^{\infty} e^{-u^2}\, du.$$

(1) Starting Point: The given definition of $\Phi(z)$.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}}\, du.$$

(2) Change of Variable: To relate $\Phi(z)$ to $\text{erfc}(z)$, we make a change of variable in the integral. Let $v = u/\sqrt{2}$, which implies $u = \sqrt{2}v$ and $du = \sqrt{2}dv$.

$$\Phi(z) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{z}{\sqrt{2}}} e^{-v^2}\, dv.$$

(3) Expressing $\Phi(z)$ in terms of $\text{erfc}(z)$: he error function is defined as $\text{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_{z}^{\infty} e^{-u^2}\, du.$. Hence, the complementary error function is $\text{erfc}(z) = 1 - \text{erf}(z)$.

$$\text{erfc}(z) = 1 - \frac{2}{\sqrt{\pi}} \int_{0}^{z} e^{-u^2}\, du$$

(4) Connecting $\Phi(z)$ and $\text{erfc}(z)$: We observe that $\Phi(z)$ integrates from $-\infty$ to a positive value, while $\text{erfc}(z)$ integrates from a positive value to ∞. They are complementary in nature. Therefore, we can write:

$$\Phi\left(\frac{z}{\sqrt{2}}\right) = \frac{1}{2}\text{erfc}\left(-\frac{z}{\sqrt{2}}\right)$$

(5) Final Expression for $\text{erfc}(z)$: Rearranging the last equation for $\text{erfc}(z)$, we get:

$$\text{erfc}(z) = 2\left(1 - \Phi\left(\frac{z}{\sqrt{2}}\right)\right)$$

(6) Converting Back to Integral Form: Finally, substituting the integral form of $\Phi(z)$ into the equation for $\text{erfc}(z)$, we arrive at the desired expression:

$$\text{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_{z}^{\infty} e^{-u^2}\, du$$

6.8 The NIST Test

6.8.1 Frequency (Monobits) Test

- The most basic test is that of the null hypothesis: in a sequence of independent identically distributed Bernoulli random variables the probability of ones is 1/2.

- By the classic **De Moivre-Laplace theorem** (*Central Limit Theorem*), for a sufficiently large number of trials, the distribution of the binomial sum, normalized by $\sqrt{n}$, is closely approximated by a standard normal distribution.

- This test makes use of that approximation to assess the closeness of the fraction of 1's to 1/2.

- All subsequent tests are conditioned on having passed this first basic test.

- $$X = 2\varepsilon - 1, \quad S_n = X_1 + \cdots + X_n = 2(\varepsilon_1 + \cdots + \varepsilon_n) - n.$$

$$\mathbb{E}[S_n] = 0, \quad \mathrm{Var}(S_n) = n.$$

$$\lim_{n\to\infty} \Pr\left[\frac{S_n}{\sqrt{n}} \le z\right] = \Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-u^2/2}\,du, \quad \Pr\left[\frac{|S_n|}{\sqrt{n}} \le z\right] = 2\Phi(z) - 1.$$

- According to the test based on the statistic $s = |S_n|/\sqrt{n}$, evaluate the observed value $|s(\mathrm{obs})| = |X_1 + \ldots + X_n|/\sqrt{n}$, and then calculate the corresponding **P-value**, which is

$$2\left[1 - \Phi\left(|s(\mathrm{obs})|\right)\right] = \mathrm{erfc}\left(\frac{|s(\mathrm{obs})|}{\sqrt{2}}\right).$$

Here, erfc is the (complementary) error function

$$\mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}}\int_{z}^{\infty} e^{-u^2}\,du.$$

Remark 6.10. 1. Conversion to ±1: The zeros and ones of the input sequence (e) are converted to values of −1 and +1 and are added together to produce $S_n = X_1 + X_2 + \ldots + X_n$, where $X_i = 2e_i - 1$.

For example, if $e = 1011010101$, then $n = 10$ and $S_n = (-1) + 1 + (-1) + 1 + (-1) + 1 + (-1) + 1 + (-1) + 1 = 2$.

2. Compute the test statistic $s_{\mathrm{obs}} = \frac{|S_n|}{\sqrt{n}}$

For the example in this section, $s_{\mathrm{obs}} = \frac{|2|}{\sqrt{10}} \approx 0.632455532$.

3. Compute P-value $=$ erfc $\left(\frac{s_{\text{obs}}}{\sqrt{2}}\right)$, where erfc is the complementary error function as defined in Section 5.5.3.
   P-value $=$ erfc $\left(\frac{0.632455532}{\sqrt{2}}\right) \approx 0.527089 \geq 0.01 = \alpha$. PASS.

Example 6.4.

(input) $e = 11001010000011 \cdots 0011000100110001000110010100011001001 1100$

(input) $n = 100$

(processing) $S_{100} = -16$

(processing) $s_{\text{obs}} = 1.6$

(output) P-value $= 0.109599$

(conclusion) **Since P-value ≥ 0.01, accept the sequence as random.**

6.8.2 Frequency Test within a Block

The test seeks to detect localized deviations from the ideal 50% frequency of 1's by decomposing the test sequence into a number of overlapping subsequences and applying a chi-square test for a homogeneous match of empirical frequencies to the ideal $\frac{1}{2}$.

- Small P-values indicate large deviations from the equal proportion of ones and zeros in at least one of the substrings.
- The string of 0's and 1's (or equivalent -1's and 1's) is partitioned into a number of disjoint substrings.
- For each substring, the proportion of ones is computed.
- A chi-square statistic compares these substring proportions to the ideal $\frac{1}{2}$.
- The statistic is referred to a chi-squared distribution with the degrees of freedom equal to the number of substrings.

The parameters of this test are M and N, so that $n = MN$, i.e., the original string is partitioned into N substrings, each of length M.

- For each of these substrings, the probability of ones is estimated by the observed relative frequency of 1's, π_i, for $i = 1, \ldots, N$.

6.8.3 The reported P-value:

(where `igamc` is the incomplete gamma function)

The reported P-value is computed as:

$$\frac{\int_{\frac{\chi^2(\text{obs})}{2}}^{\infty} e^{-u} u^{(N/2)-1} du}{\Gamma(N/2)2^{N/2}} = \frac{\int_{\frac{\chi^2(\text{obs})}{2}}^{\infty} e^{-u} u^{N/2-1} du}{\Gamma(N/2)} = \texttt{igamc}\left(\frac{N}{2}, \frac{\chi^2(\text{obs})}{2}\right).$$

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

$$P(a,x) = \frac{\gamma(a,x)}{\Gamma(a)} = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt$$

where $P(a,0) = 0$ and $P(a,\infty) = 1$.

$$Q(a,x) = 1 - P(a,x) = \frac{\Gamma(a,x)}{\Gamma(a)} = \frac{1}{\Gamma(a)} \int_x^\infty e^{-t} t^{a-1} dt$$

where $Q(a,0) = 1$ and $Q(a,\infty) = 0$.

Remark 6.11. 1. Partition the input sequence into $N = \lceil \frac{n}{M} \rceil$ non-overlapping blocks. Discard any unused bits.

For example, if $n = 10$, $M = 3$ and $e = 0110011010$, 3 blocks ($N = 3$) would be created, consisting of 011, 001 and 101. The final 0 would be discarded.

2. Determine the proportion π_i of ones in each M-bit block using the equation

$$\pi_i = \frac{\sum_{j=1}^{M} e_{(i-1)M+j}}{M}$$

for $1 \le i \le N$.

For the example in this section, $\pi_1 = \frac{2}{3}$, $\pi_2 = \frac{1}{3}$, and $\pi_3 = \frac{2}{3}$.

3. Compute the χ^2 statistic:

$$\chi^2(\text{obs}) = 4M \sum_{i=1}^{N} \left(\pi_i - \frac{1}{2}\right)^2.$$

For the example in this section,

$$\chi^2(\text{obs}) = 4 \times 3 \left(\left(\frac{2}{3} - \frac{1}{2}\right)^2 + \left(\frac{1}{3} - \frac{1}{2}\right)^2 + \left(\frac{2}{3} - \frac{1}{2}\right)^2\right) = 1.$$

Compute P-value

Compute P-value = $\texttt{igamc}\left(\frac{N}{2}, \frac{\chi^2(\text{obs})}{2}\right)$, where `igamc` is the incomplete gamma function for $Q(a,x)$ as defined in Section 5.5.3.

Note: When comparing this section against the technical description in Section 3.2, note that $Q(a,x) = 1 - P(a,x)$.

For the example in this section, P-value = $\texttt{igamc}\left(\frac{3}{2}, \frac{7.2}{2}\right) = 0.801252$.

6.9 Runs Test

This test assesses the randomness of a binary sequence by analyzing the frequency and pattern of changes between consecutive 1's and 0's. The test is useful in contexts where the binary sequence should be random, such as in random number generation.

This variant of a classic **nonparametric test**[1] looks at "runs" defined as substrings of consecutive 1's and consecutive 0's, and considers whether the oscillation[2] among such homogeneous substrings is too fast or too slow.

Runs Test

(i) Estimating the Proportion π

- **Proportion** $\pi = \frac{\sum_{i=1}^{n} \varepsilon_i}{n}$: Estimated proportion of 1's (or 0's) in the sequence.

(ii) Asymptotic Distribution

- As $n \longrightarrow \infty$, $\frac{V_n - 2n\pi(1-\pi)}{\sqrt{2n}\pi(1-\pi)}$ converges to the standard normal distribution $\Phi(z)$:

$$\lim_{n\to\infty} \Pr\left[\frac{V_n - 2n\pi(1-\pi)}{\sqrt{2n}\pi(1-\pi)} \le z\right] = \Phi(z).$$

(iii) The Test Statistic V_n

- **Total Number of Runs**: V_n is the total number of runs in the sequence:

$$V_n = \sum_{k=1}^{n-1} r(k) + 1 \quad \text{with} \quad r(k) = \begin{cases} 0 & : \varepsilon_k = \varepsilon_{k+1}, \\ 1 & : \varepsilon_k \neq \varepsilon_{k+1}. \end{cases}$$

(iv) P-Value

- **P-Value**: P-Value $= \operatorname{erfc}\left(\frac{|V_n(\text{obs}) - 2n\pi(1-\pi)|}{\sqrt{2n}\pi(1-\pi)}\right)$.
- Large $V_n(\text{obs})$: More oscillation than expected by chance.
- Small $V_n(\text{obs})$: Less oscillation than expected by chance.

Remark 6.12. $V_n$obs : The total number of runs (i.e., the total number of zero runs + the total number of one-runs) across all n bits. The reference distribution for the test statistic is a χ^2 distribution.

[1] A statistical test that does not assume the data comes from a particular distribution. Useful when normal distribution cannot be assumed.

[2] Assessing whether changes between 1's and 0's happen too frequently (fast oscillation) or too infrequently (slow oscillation).

Test Description

Note: The Runs test carries out a Frequency test as a prerequisite.

1. Compute the pre-test proportion π of ones in the input sequence: $\pi = \frac{\sum e_j}{n}$.

   For example, if $e = 1001101011$, then $n = 10$ and $\pi = \frac{6}{10} = \frac{3}{5}$.

2. Determine if the prerequisite Frequency test is passed: If it can be shown that $|\frac{\pi - 1/2}{\sqrt{n/4}}| < t$, then the Runs test need not be performed (i.e., the test should not have been run because of a failure to pass test 1, the Frequency (Monobit) test). If the test is not applicable, then the P-value is set to 0.0000. Note that for this test, $\frac{n}{4}$ has been pre-defined in the test code.

   For the example in this section, since $t = \frac{\pi - 1/2}{\sqrt{n/4}} = 0.63246$, then $|t - 1/2| = |3/5 - 1/2| = 0.1 < t$.

   Since the observed value π is within the selected bounds, the runs test is applicable.

3. Compute the test statistic $V_{n,\text{obs}}$:

   $$V_{n,\text{obs}} = \sum_{k=1}^{n-1} r(k+1), \text{ where } r(k) = 0 \text{ if } e_k = e_{k+1}, \text{ and } r(k) = 1 \text{ otherwise.}$$

   Since $e = 001101011$, then

   $$V_{10,\text{obs}} = (1 + 0 + 1 + 0 + 1 + 1 + 1 + 0 + 1) = 7.$$

4. Compute P-value = $\text{erfc}\left(\frac{|V_{n,\text{obs}} - 2n\pi(1-\pi)|}{2\sqrt{2n\pi(1-\pi)}}\right)$.

   For the example, P-value = $\text{erfc}\left(\frac{2 \cdot 10 \cdot \frac{3}{5} \cdot (1 - \frac{3}{5})}{2 \cdot \sqrt{2 \cdot 10 \cdot \frac{3}{5} \cdot (1 - \frac{3}{5})}}\right) \approx 0.147232$.

Input Size Recommendation

It is recommended that each sequence to be tested consist of a minimum of 100 bits (i.e., $n \geq 100$).

6.10 Binary Matrix Rank Test

- The focus of the test is the rank of disjoint sub-matrices of the entire sequence.
- The purpose of this test is to check for linear dependence among fixed length substrings of the original sequence.
  - Construct matrices of successive zeroes and ones from the sequence, and check for linear dependence among the rows or columns of the constructed matrices.
  - The deviation of the rank - or rank deficiency - of the matrices from a theoretically expected value gives the statistic of interest.
- The result states that the rank R of the $M \times Q$ random binary matrix takes values $r = 0, 1, 2, \ldots, m$, where $m = \min(M, Q)$, with probabilities

$$P_r = 2^{r(Q+M-r)-MQ} \prod_{i=0}^{r-1} \frac{(1-2^{i-Q})(1-2^{i-M})}{1-2^{i-r}},$$

  - The probability values are fixed in the test suite code for $M = Q = 32$.
  - The number M is then a parameter of this test, so that ideally $n = M^2N$, where N is the new "sample size".
- In practice, values for M and N are chosen so that the discarded part of the string, $n - M^2N$, is fairly small.

Discrete Fourier Transform (Spectral) Test

- The test described here is based on the **discrete Fourier transform**.
- It is a member of a class of procedures known as **spectral methods**.
- The Fourier test detects [text redacted] that would indicate a deviation from the assumption of randomness.

Test Purpose

- The focus of this test is the peak heights in the Discrete Fourier Transform of the sequence.
- The purpose of this test is to detect periodic features (i.e., *repetitive patterns* that are near each other) in the tested sequence that would indicate a deviation from the assumption of randomness.
- The intention is to detect whether the number of peaks exceeding the 95% threshold is significantly different than 5%.

Random Excursions Test

- The focus of this test is the number of cycles having exactly K visits in a cumulative sum random walk.
  - The cumulative sum random walk is derived from partial sums after the (0,1) sequence is transferred to the appropriate (-1,+1) sequence.
  - A cycle of a random walk consists of a sequence of steps of unit length taken at random that begin at and return to the origin.
- The purpose of this test is to determine if the number of visits to a particular state within a cycle deviates from what one would expect for a random sequence.
  - This test is actually a series of eight tests (and conclusions), one test and conclusion for each of the states: -4, -3, -2, -1 and +1, +2, +3, +4.
- This test is based on considering successive sums of the binary bits (plus or minus simple ones) as a one dimensional random walk.
  - The test detects deviations from the distribution of the number of visits of the random walk to a certain "state," i.e., any integer value.

Chapter 7

Evaluation of Entropy

7.1 Introduction to Entropy

We introduce the concept of **entropy** is a measure of the uncertainty of a random variable.

Entropy

Definition 7.1. The **entropy** $H(X)$ of a discrete random variable X is defined by

$$H : \mathbb{R}^{\Omega} \longrightarrow \mathbb{R}_{\geq 0} : H(X) = -\sum_{x \in X} p(x) \log_2 p(x).$$

We also write $H(p)$ for the above quantity. The log is to the base 2 and entropy is expressed in bits.

Example 7.1. For example, the entropy of a fair coin toss is 1 bit: note that

$$X : \Omega \longrightarrow \mathbb{R} : \begin{cases} X(\text{"heads"}) = 1 \\ X(\text{"tails"}) = 0 \end{cases}, \quad \begin{cases} \Pr[X = 1] = p(1) = \frac{1}{2} \\ \Pr[X = 0] = p(0) = \frac{1}{2}. \end{cases}$$

Since

$$p(0) \log_2 p(0) = \frac{1}{2} \log_2 \frac{1}{2} = -\frac{1}{2} \quad \text{and} \quad p(1) \log_2 p(1) = \frac{1}{2} \log_2 \frac{1}{2} = -\frac{1}{2},$$

we have

$$H(X) = -\sum_{x \in \{0,1\}} p(x) \log_2 p(x) = 1.$$

We will use the convention that

$$p(x) = 0 \implies 0 \log 0 := 0,$$

which is easily justified by continuity since

$$\lim_{x \to 0} x \log x = 0.$$

Adding terms of zero probability does not change the entropy.

Remark 7.1. The expected value of a function $g(x)$ of a random variable $X = x$ is given by

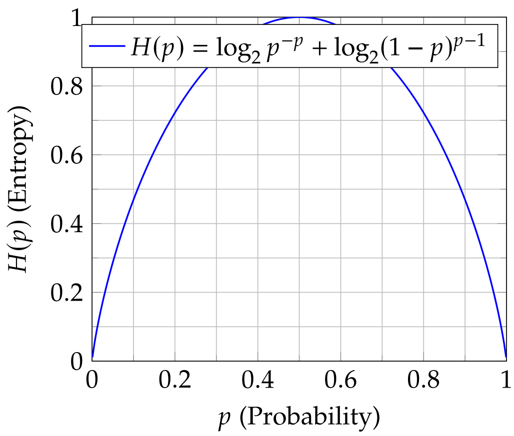
$$\mathbb{E}[g(x)] = \sum_{x \in X} p(x) g(x).$$

Let $g(x) := \log_2 \frac{1}{p(x)}$ then

$$\mathbb{E}\left[\log_2 \frac{1}{p(x)}\right] = \sum_{x \in X} p(x) \log_2 \frac{1}{p(x)} = -\sum_{x \in X} p(x) \log_2 p(x).$$

Therefore, entropy can be seen as the expected value of the information content of each outcome of a random variable. It quantifies the average amount of information (or uncertainty) inherent in the random variable's possible outcomes.

Example 7.2 (Entropy of a Binary Random Variable)**.**



Let

$$X = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases}$$

Then

$$H(X) = -\left(\log_2 p^p + \log_2 (1-p)^{1-p}\right) \stackrel{\text{def}}{=} H(p).$$

Example 7.3. Let

$$X = \begin{cases} a & \text{with probability } \frac{1}{2}, \\ b & \text{with probability } \frac{1}{4}, \\ c & \text{with probability } \frac{1}{8}, \\ d & \text{with probability } \frac{1}{8}. \end{cases}$$

The entropy of X is

$$H(X) = -\left(\frac{1}{2} \cdot (-1) + \frac{1}{4} \cdot (-2) + \frac{1}{8} \cdot (-3) + \frac{1}{8} \cdot (-3)\right) = \frac{4+4+3+3}{8} = \frac{14}{8} = \frac{7}{4} = 1.75 \text{ bits}.$$

7.2 Entropy in Cryptography

We want to look at what happens **as more and more plaintexts are encrypted using the same key**, and how likely a cryptanalyst will be able to carry out a successful **ciphertext-only attack**, given sufficient time.

The basic tool in studying this question is the idea of **entropy**, a concept from information theory introduced by Shannon in 1948. **Entropy** can be thought of as a mathematical measure of *information* or *uncertainty*, and is computed as a function of a probability distribution.

Suppose we have a discrete random variable X which takes values from a finite set according to a specified probability distribution. What is the *information* gained by the outcome of an experiment which takes place according to this probability distribution? Equivalently, if the experiment has not (yet) taken place, what is the *uncertainty* about the outcome? This quantity is called the **entropy of** X and is denoted by $H(X)$.

If $H(X)$ is high, it means there's a lot of uncertainty about the outcome of the experiment (or a lot of information to be gained by performing it). If $H(X)$ is low, it means most outcomes are quite predictable, with little new information to be gained.

In summary, the entropy $H(X)$ of a random variable X quantifies the expected amount of information gained—or equivalently, the uncertainty—about the outcome of an experiment modeled by X before the experiment is performed.

Example 7.4 (The Length of a Bit-string in Probability Encoding). Suppose we have a random variable X such that

$$\Pr[X = x_1] = \frac{1}{2}, \quad \Pr[X = x_2] = \frac{1}{4}, \quad \Pr[X = x_3] = \frac{1}{4}.$$

Suppose we encode the three possible outcomes as follows:

(i) x_1 is encoded as `0`,

(ii) x_2 is encoded as `10`, and

(iii) x_3 is encoded as 11.

Then the **(weighted) average number of bits in this encoding of** X is

$$\frac{1}{2} \times \text{“1-bit (0)”} + \frac{1}{4} \times \text{“2-bit (10)”} + \frac{1}{4} \times \text{“2-bit (11)”} = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 2 = \frac{3}{2}.$$

Note (Motivation of the Entropy). In binary encoding, an n-bit string represents 2^n different states or outcomes. If each of these 2^n outcomes is equally likely, the probability of any specific outcome is $\frac{1}{2^n}$. For an event with a probability p, we find the number of bits required n by setting $\frac{1}{2^n} = p$.

$$\begin{aligned} \frac{1}{2^n} = p &\implies \log_2 2^{-n} = \log_2 p \\ &\implies n = -\log_2 p. \end{aligned}$$

The bit-string length for encoding an event with probability p is approximately $-\log_2 p$. This principle is central in information theory, reflecting that less probable events, which carry more information, require longer bit strings for encoding.

We could imagine that an outcome occurring with probability p might be encoded by a bit-string of length approximately $-\log_2(p)$. Given an arbitrary probability distribution, taking on the values $p_1, p_2, \ldots, p_r$ for a random variable X, we take the **weighted average of the quantities** $-\log_2(p_i)$ to be our measure of information.

Entropy of Finite Random Variable

Definition 7.2. Suppose X is a discrete random variable that takes on values from a finite set X, say, $|X| = n$. The **entropy** of the random variable X is defined to be the quantity

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Remark 7.2. Note that if $|X| = n$ and $Pr[x] = 1/n$ for all $x \in X$, then

$$H(X) = -\sum_{i=1}^{n} p_i \log_2(p_i) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{n} = -n \cdot \frac{1}{n} \log_2 n^{-1} = \log_2 n.$$

Also, $H(X) \geq 0$ for any random variable X and

$$H(X) = 0 \iff \begin{cases} \Pr[X = x_0] = 1 & \text{for some } x_0 \in X \\ \Pr[X = x] = 0 & \text{for all } x \neq x_0 \end{cases}$$

Example 7.5. Let $\mathcal{P} = \{a, b\}$ with $p(a) = 1/4, p(b) = 3/4$. Let $\mathcal{K} = \{k_1, k_2, k_3\}$ with $p(k_i) = 1/2, p(k_2) = p(k_3) = 1/4$. Let $C = \{1, 2, 3, 4\}$, and suppose the encryption functions are defined to be $e_{k_1}(a) = 1, e_{k_1}(b) = 2; e_{k_2}(a) = 2, e_{k_2}(b) = 3$; and $e_{k_3}(a) = 3, e_{k_3}(b) = 4$. This cryptosystem can be represented by the following **encryption matrix**:

$\mathcal{K} \backslash \mathcal{P}$	a	b
K_1	1	2
K_2	2	3
K_3	3	4

Note that

$$p(1) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}, \quad p(2) = \frac{3}{8} + \frac{1}{16} = \frac{7}{16}, \quad p(3) = \frac{3}{16} + \frac{1}{16} = \frac{1}{4}, \quad p(4) = \frac{3}{16}.$$

We compute as follows:

- $H(\mathcal{P}) = -\left(\frac{1}{4} \cdot (-2) + \frac{3}{4}\left(\log_2 3 - (-2)\right)\right) = 2 - \frac{3}{4} \log_2 3 \approx 0.81.$
- $H(\mathcal{K}) = -\left(\frac{1}{2} \cdot (-1) + \frac{1}{4} \cdot (-2) + \frac{1}{4} \cdot (-2)\right) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 1.5.$
- $H(C) = -\left(\frac{1}{8} \cdot (-3) + \cdot \frac{7}{16} \cdot \left(\log_2 7 - (-4)\right) + \frac{1}{2} \cdot (-2) + \frac{3}{16}\left(\log_2 3 - (-4)\right)\right) \approx 1.85.$

7.3 Rényi Entropy

Rényi Entropy

Definition 7.3. The **Rényi entropy** of order α, where $\alpha \geq 0$ and $\alpha \neq 1$, is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^\alpha\right).$$

Here, X is a discrete random variable with possible outcomes $1, 2, \ldots, n$ and corresponding probabilities $p_i = \Pr[X = i]$ for $i = 1, \ldots, n$.

Remark 7.3.

- **Hartley or Max-entropy:** $H_0(X) = \log n = \log |X|$.
- **Shannon Entropy:** $H_1(X) = \lim_{\alpha \to 1} H_\alpha(X) = -\sum_{i=1}^n p_i \log p_i$

  Proof. Let $f(\alpha) := \log\left(\sum_{i=1}^n p_i^\alpha\right)$ and $g(\alpha) := 1 - \alpha$ Then

  (i)
  $$\frac{\mathrm{d}}{\mathrm{d}\alpha} f = \frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\log \sum_{i=1}^{n} p_i^\alpha\right) = \frac{\sum_{i=1}^n p_i^\alpha \log p_i}{\sum_{i=1}^n p_i^\alpha}.$$

  (ii)
  $$\frac{\mathrm{d}}{\mathrm{d}\alpha} g = \frac{\mathrm{d}}{\mathrm{d}\alpha}(1-\alpha) = -1.$$

  Thus,

  $$\lim_{\alpha \to 1} H_\alpha(X) = \lim_{\alpha \to 1} \frac{f(\alpha)}{g(\alpha)} = \lim_{\alpha \to 1} \frac{f'(\alpha)}{g'(\alpha)} = -1 \cdot \frac{\sum_{i=1}^n p_i \log p_i}{\sum_{i=1}^n p_i} = -1 \cdot \frac{\sum_{i=1}^n p_i \log p_i}{1} = -\sum_{i=1}^{n} p_i \log p_i.$$

  □

- **Collision entropy:** $H_2(X) = -\log \sum_{i=1}^n p_i^2 = -\log Pr[X = Y], \quad X, Y : \text{i.i.d.}$
- **Min-entropy:** $H_\infty(X) = \min_i(-\log p_i) = -(\max_i \log p_i) = -\log \max_i p_i$

Min-Entropy

Entropy is defined relative to one's knowledge of an experiment's output prior to observation, and reflects the **uncertainty** associated with predicting its value – the larger the amount of entropy, the greater the uncertainty in predicting the value of an observation.

There are many possible measures for entropy; NIST uses a very *conservative measure* known as **min-entropy**, which measures the effectiveness of the strategy of guessing the most likely output of the entropy source.

The min-entropy of an independent discrete random variable X that takes values from the set $\{x_1, x_2, \ldots, x_k\}$ with probability $Pr[X = x_i] = p_i$ for $i = 1, \ldots, k$ is defined as

$$H_\infty(X) = \min(-\log_2 p_i) = -\log_2 \max p_i.$$

If X has min-entropy H, then the probability of observing any particular value for X is no greater than 2^{-H}. The maximum possible value for the min-entropy of a random variable with k distinct values is $\log_2 k$, which is attained when the random variable has a uniform probability distribution, i.e., $p_1 = p_2 = \ldots = p_k = 1/k$.

Probabilistic Analysis for the Relationship Between Min-Entropy and Guessing Attack

Min-entropy is closely related to the *optimum guessing attack cost*.

$$\mathbb{E}[S_\delta] = W_k(P)$$

Abstract. Recently NIST has published the second draft document of recommendation for the entropy sources used for random bit generation. In this document NIST has provided a practical and detailed description about the fact that the min-entropy is closely related to the optimum guessing attack cost. However the argument lacks the mathematical rigor. In this paper we provide an elaborate probabilistic analysis for the relationship between the min-entropy and cost of optimum guessing attack. Moreover we also provide some simulation results in order to investigate the practicality of optimum guessing attack.

Keywords: Entropy source · Min-Entropy · Optimum guessing attack

Appendix D–Min-Entropy and Optimum Guessing Attack Cost

Suppose that an adversary wants to determine at least one of several secret values, where each secret value is independently chosen from a set of M possibilities, with probability distribution $P = \{p_1, p_2, \ldots, p_M\}$. Assume that these probabilities are sorted so that $p_1 \geq p_2 \geq \ldots \geq p_M$. Consider a guessing strategy: and assume successfully guessing as many secret values as possible... The adversary's goal would be to minimize the expected number of guesses per successful recovery. Such a strategy would consist of guessing a maximum of k possibilities for a given secret value, moving on to a new secret value when either a guess is correct, or k incorrect guesses for the current value have been made. In general, the optimum value of k can be anywhere in the range $1 \leq k \leq M$, depending on the probability distribution P. Note that when $k = M$, the M^{th} guess is considered valid (though trivial) guess. Regardless of the value of k chosen, it is clear that the guesser selected for a given secret value should be the k most likely possible values, in decreasing order of probability.

The expected number of guesses $W_k(P)$ is given by:

$$W_k(P) = p_1 + 2p_2 + \ldots + (k-1)p_{k-1} + k\left(1 - \sum_{i=1}^{k-1} p_i\right)$$

Entropy Source Model

Entropy Source Validation

- Data Collection
- Determining the track: IID track vs. non–IID track
- Initial Entropy Estimate
- Restart Tests
- Entropy Estimation for Entropy Sources Using a Conditioning Component
- Additional Noise Sources

Health Tests

- Repetition Count Test
- Adaptive Proportion Test

Testing the IID Assumption

Permutation Testing

Input: $S = (s_1, \ldots, s_n)$
Output: Decision on the IID assumption

1 For each test i

  a Assign the counters $C_{0,i}$ and $C_{1,i}$ to zero.
  b Calculate the test statistic T_i on S.

2 For $j = 1$ to 10,000

  1 Permute S using the Fisher-Yates shuffle algorithm.
  2 For each test i
    a Calculate the test statistic T on the permuted data.
    b If $(T > T_i)$, increment $C_{0,i}$. If $(T = T_i)$, increment $C_{1,i}$.

3 If $(C_{0,i} + C_{1,i})/10,000 > 0.9995$ for any i, reject the IID assumption; else, assume that the noise source outputs are IID.

Testing the IID Assumption

Permutation Testing

1. Excursion Test Statistic
2. Number of Directional Runs
3. Length of Directional Runs
4. Number of Increases and Decreases
5. Number of Runs Based on the Median
6. Length of Runs Based on Median
7. Average Collision Test Statistic
8. Maximum Collision Test Statistic
9. Periodicity Test Statistic
10. Covariance Test Statistic
11. Compression Test Statistic

Additional Chi-square Statistical Tests

1. Testing Independence for Non-Binary Data
2. Testing Goodness-of-fit for Non-Binary Data
3. Testing Independence for Binary Data
4. Testing Goodness-of-fit for Binary Data
5. Length of the Longest Repeated Substring Test

Estimating Min-Entropy

IID Track: Entropy Estimation for IID Data

- most common value estimate

Non-IID Track: Entropy Estimation for Non-IID Data

- The Most Common Value Estimate
- The Collision Estimate
- The Markov Estimate
- The Compression Estimate

- The t-Tuple Estimate
- The Longest Repeated Substring (LRS) Estimate
- The Multi Most Common in Window Prediction Estimate
- The Lag Prediction Estimate
- The MultiMMC Prediction Estimate
- The LZ78Y Prediction Estimate

The Most Common Value Estimate

This method first finds the proportion $\hat{p}$ of the most common value in the input dataset, and then constructs a confidence interval for this proportion. The upper bound of the confidence interval is used to estimate the min-entropy per sample of the source.

Given the input $S = (s_1, \ldots, s_L)$, where $s_i \in \{x_1, \ldots, x_k\}$,

1. Find the proportion of the most common value $\hat{p}$ in the dataset, i.e.,
$$\hat{p} = \max \frac{\#\{s_i\}}{L}.$$
2. Calculate an upper bound on the probability of the most common value p_u as
$$p_u = \min\left(1, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{L-1}}\right),$$
where z corresponds to the $Z_{(1-\alpha/2)}$ value.
3. The estimated min-entropy is $-\log_2(p_u)$.

Example: If the dataset is $S = (0,1,1,2,0,1,2,2,0,1,0,1,0,2,2,1,0,2,1)$, with $L = 20$, the most common value is 1, with $\hat{p} = 0.4$. $p_u = 0.4 + 2.576 \cdot \sqrt{0.012} = 0.6895$. The min-entropy estimate is $-\log_2(0.6895) \approx 0.5363$.

The Markov Estimate

This entropy estimation method is only applied to binary inputs.

Given the input $S = (s_1, \ldots, s_L)$, where $s_i \in \{0, 1\}$,

1. Estimate the initial probabilities for each output value, P_0 and $P_1 = 1 - P_0$.
2. Let T be the 2×2 transition matrix of the form
$$\begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix}$$
where the probabilities are calculated as
$$P_{00} = \frac{\#\{00 \text{ in } S\}}{\#\{0 \text{ in } S\} - 1}, \quad P_{01} = \frac{\#\{01 \text{ in } S\}}{\#\{0 \text{ in } S\}}$$
$$P_{10} = \frac{\#\{10 \text{ in } S\}}{\#\{1 \text{ in } S\}}, \quad P_{11} = \frac{\#\{11 \text{ in } S\}}{\#\{1 \text{ in } S\} - 1}$$

Find the probability of the most likely sequence of outputs of length 128, as calculated below.

Sequence	Probability
00...0	$P_0 \times P_0^{127}$
0101...01	$P_0^{64} \times P_1^{64}$
011...1	$P_0 \times P_1 \times P_1^{126}$
100...0	$P_1 \times P_0 \times P_0^{126}$
1010...10	$P_1^{64} \times P_0^{64}$
11...1	$P_1 \times P_1^{127}$

Let $p_{\max}$ be the maximum of the probabilities in the table given above. The min-entropy estimate is the negative logarithm of the probability of the most likely sequence of outputs, $p_{\max}$:

$$\text{min-entropy} = \min(-\log_2(p_{\max})/128, 1)$$

Example

For the purpose of this example[2], suppose that $L = 40$ and $S = (1, 0, 0, \ldots, 0, 1)$, with $L = 20$, the most common value is 1, with $p = 0.4$. The transition matrix is calculated as

$$\begin{bmatrix} 0 & 1 \\ P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix}$$

where the probabilities are calculated as

$$P_{00} = \frac{\#00 \text{ in } S}{\#0 \text{ in } S - 1}, \quad P_{01} = \frac{\#01 \text{ in } S}{\#0 \text{ in } S}$$
$$P_{10} = \frac{\#10 \text{ in } S}{\#1 \text{ in } S}, \quad P_{11} = \frac{\#11 \text{ in } S}{\#1 \text{ in } S - 1}$$

The probabilities of the possible sequences are

$$\begin{array}{rl} 00\ldots0 & 3.9837 \times 10^{-3} \\ 0101\ldots01 & 4.4381 \times 10^{-4} \end{array}$$

$$\begin{array}{rl} 011\ldots1 & 1.4202 \times 10^{-4} \\ 10\ldots0 & 6.4631 \times 10^{-3} \end{array}$$

$$\begin{array}{rl} 1010\ldots10 & 4.6288 \times 10^{-9} \\ 11\ldots1 & 1.0121 \times 10^{-4} \end{array}$$

The resulting entropy estimate is

$$\text{min-entropy} = \min(-\log_2(4.6288 \times 10^{-9})/128, 1) = \min(0.761, 1) = 0.761.$$

Chapter 8

2021-2 Final Exam

8.1 On Renyi Entropy

(a) Let us consider a discrete random variable X with a sample space $\{1, 2, \ldots, n\}$ and probability $p = P(X = i)$, for $1 \leq i \leq n$. Assuming $a > 0$, $a \neq 1$, the a-th order Renyi entropy is defined as:

$$H_a(X) = \frac{1}{1-a} \log\left(\sum_i p_i^a\right).$$

Show that, in the limit as a approaches 1, Renyi entropy converges to the Shannon entropy:

$$\lim_{a \to 1} H_a(X) = H(X) = -\sum_i p_i \log p_i.$$

(b) Considering a set of probabilities $\{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8\}$ where each p_i corresponds to an event in a sample space:

$$(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8) = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\right).$$

Calculate the Rényi entropies $H_0(X), H_1(X), H_2(X), H_\infty(X)$ for this distribution.

Solution.

(a) Let $f(\alpha) := \log\left(\sum_{i=1}^n p_i^\alpha\right)$ and $g(\alpha) := 1 - \alpha$ Then

(i)

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} f = \frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\log \sum_{i=1}^n p_i^\alpha\right) = \frac{\sum_{i=1}^n p_i^\alpha \log p_i}{\sum_{i=1}^n p_i^\alpha}.$$

(ii)

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} g = \frac{\mathrm{d}}{\mathrm{d}\alpha}(1 - \alpha) = -1.$$

Thus,

$$\lim_{\alpha \to 1} H_\alpha(X) = \lim_{\alpha \to 1} \frac{f(\alpha)}{g(\alpha)} = \lim_{\alpha \to 1} \frac{f'(\alpha)}{g'(\alpha)} = -1 \cdot \frac{\sum_{i=1}^n p_i \log p_i}{\sum_{i=1}^n p_i} = -1 \cdot \frac{\sum_{i=1}^n p_i \log p_i}{1} = -\sum_{i=1}^n p_i \log p_i.$$

(b) 1. $H_0(X)$ **(Rényi entropy for $a = 0$):**

The H_0 entropy, also known as the Hartley entropy, is defined as

$$H_0(X) = \log_2(N)$$

where N is the number of non-zero probabilities. Thus

$$H_0(X) = \log_2(8) = 3.0 \text{ bits}.$$

2. $H_1(X)$ **(Shannon entropy or Rényi entropy for $a = 1$):** The Shannon entropy, H_1, is calculated using the formula:

$$H_1(X) = -\sum_i p_i \log_2 p_i$$

The sum for our probabilities is:

$$\begin{aligned}
H_1(X) &= -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{16}\log_2\frac{1}{16} + 4\times\frac{1}{64}\log_2\frac{1}{64}\right) \\
&= -\left(\frac{1}{2}\cdot(-1) + \frac{1}{4}\cdot(-2) + \frac{1}{8}\cdot(-3) + \frac{1}{16}\cdot(-4) + 4\times\frac{1}{64}\cdot(-6)\right) \\
&= \frac{32+32+24+16+24}{64} \\
&= 2.0 \text{ bits}.
\end{aligned}$$

3. $H_2(X)$ **(Rényi entropy for $a = 2$):** For $a = 2$, the Rényi entropy is calculated as:

$$H_2(X) = -\log_2\left(\sum_i p_i^2\right)$$

This gives:

$$\begin{aligned}
H_2(X) &= -\log_2\left(\left(\frac{1}{2}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{8}\right)^2 + \left(\frac{1}{16}\right)^2 + 4\times\left(\frac{1}{64}\right)^2\right) \\
&= -\log_2\left(\frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} + 4\times\frac{1}{4096}\right) \\
&= -\log_2\left(\frac{256+64+16+4+1}{1024}\right) \\
&= \log_2\left(\frac{1024}{341}\right) \\
&= \log_2 1024 - \log_2 341 \\
&\approx 10 - 8.414 \\
&\approx 1.59.
\end{aligned}$$

Computing this sum and its logarithm gives approximately $H_2(X) = 1.59$ bits.

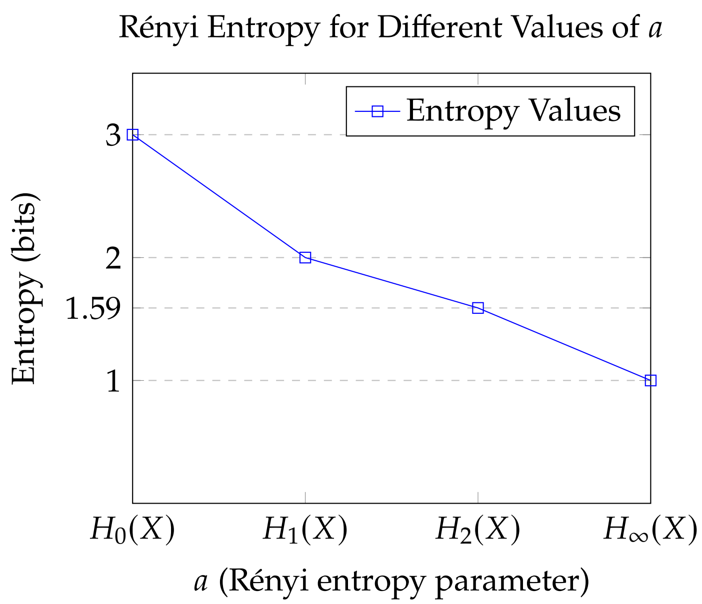
4. $H_\infty(X)$ **(Rényi entropy for $a = \infty$):** The H_∞ entropy is defined as

$$H_\infty(X) := \min(-\log_2 p_i) = -(\max \log_2 p_i) = -\log_2(\max(p_i))$$

Thus,

$$H_\infty(X) = -\log_2\left(\frac{1}{2}\right) = 1.0\text{bit}.$$

In summary,



□

8.2 Markov Chain and Transition Matrix

Consider a Markov chain with state space $S = \{0, 1, 2, 3, 4, 5, 6\}$ and the following transition matrix P for the process $(X_n, n = 0, 1, 2, \ldots)$:

$$P = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

(a) Represent the state space S by dividing it into the set S_T of transient states and the set S_R of recurrent states.

(b) Calculate the probability $\rho_{01} = P(T_1 < \infty)$.

Solution. (a) $S_T = \{1, 2, 3\}$ and $S_R = \{0, 4, 5, 6\}$

(b) $\rho_{01} = 0$.

□

8.3 Pseudo-Noise Sequence

Determine whether the sequence given below, with a cycle s^{15} for a period $N = 15$, is a pseudo-noise sequence.

$$s^{15} = \mathtt{0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 1,\ 1,\ 0,\ 1,\ 0,\ 1,\ 1,\ 1,\ 1.}$$

Solution. Recall that

Golomb's Randomness Postulates

Definition 8.1. Let s be a N-periodic sequence. **Golomb's randomness postulates** are as follows:

R1 *Balance*: $\left|\#1(s^N) - \#0(s^N)\right| \leq 1$, where #1 and #0 count the number of 1's and 0's.

R2 *Run Distribution*: For runs of length l in s^N, at least $\frac{1}{2^l}$ of runs are of length l, for $l \geq 1$ and total runs > 1, with nearly equal numbers of gap and block.

R3 *Autocorrelation Property*: $C(t)$:

$$N \cdot C(t) = \sum_{i=0}^{N-1} (2s_i - 1) \cdot (2s_{i+t} - 1) = \begin{cases} N, & : t = 0, \\ K \in \mathbb{Z}, & : 1 \leq t \leq N-1. \end{cases}$$

Given a sequence

$$0,0,0, \quad | \quad 1, \quad | \quad 0,0, \quad | \quad 1,1, \quad | \quad 0, \quad | \quad 1, \quad | \quad 0, \quad | \quad 1,1,1,1$$

R1 The sequence contains 7 zeros and 8 ones. It holds $|8 - 7| \leq 1$.

R2 Total number of runs: 8.

- Runs of length 1: 4 (2 gaps and 2 blocks).
- Runs of length 2: 2 (1 gap and 1 block).
- Runs of length 3: 1 (1 gap).
- Runs of length 4: 1 (1 block).
- Required number of runs of length 1: at least $\frac{4}{8} \geq \frac{1}{2}$ (satisfied).
- Required number of runs of length 2: at least $\frac{2}{8} \geq \frac{1}{2^2}$ (satisfied).
- Required number of runs of length 2: at least $\frac{1}{8} \geq \frac{1}{2^3}$ (satisfied).
- Required number of runs of length 2: at least $\frac{1}{8} \geq \frac{1}{2^4}$ (satisfied).

R3 Let $t = 0$ then

$$N \cdot C(t) = 15 \cdot C(0) = \sum_{i=0}^{14} (2s_i - 1)^2 = 7 \cdot (0-1)^2 + 8 \cdot (2-1)^2 = 15.$$

Let $1 \le t \le 14$ then

$$\begin{aligned}
15 \cdot C(1) &= \sum_{i=0}^{14} (2s_i - 1)(2s_{i+1} - 1) \\
15 \cdot C(2) &= \sum_{i=0}^{14} (2s_i - 1)(2s_{i+2} - 1) \\
15 \cdot C(3) &= \sum_{i=0}^{14} (2s_i - 1)(2s_{i+3} - 1) \\
&\vdots \\
15 \cdot C(14) &= \sum_{i=0}^{14} (2s_i - 1)(2s_{i+14} - 1).
\end{aligned}$$

For s^{15} = `0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1`, we know

$$2s_i - 1 = \{-1, -1, -1, 1, -1, -1, 1, 1, -1, 1, -1, 1, 1, 1, 1\}.$$

$$\begin{aligned}
15 \cdot C(1) &= [(2 \cdot 0 - 1)(2 \cdot 0 - 1) + (2 \cdot 0 - 1)(2 \cdot 1 - 1) + \ldots + (2 \cdot 1 - 1)(2 \cdot 1 - 1)] \\
&= -1
\end{aligned}$$

. Similar calculations are done for $C(t)$ for other values of t.

Hence s^{15} is a pn-sequence.

$\square$

8.4 Binary Sequence Analysis

A random number generator outputs a string of 100 bits as follows. This random number generator is subjected to testing with 100-bit strings.

```
01100 00001 00011 01001 10001 00110 00110 01010 00101 11000
10001 11001 00100 00111 11101 10101 01000 00001 01101 00011
```

Consider this as a sample sequence and apply the FIPS 140-1 statistical random number test.

(a) Determine the pass/fail of the monobit test according to the FIPS 140-1 statistical random number test.

(b) It is intended to apply the serial test to the above sample sequence of length 100. Calculate the value of the χ^2 statistic used at this time.

Solution.

(a) Recall that

Monobit Test

Let s be a binary sequence of length n. Let $n_0 := \#0(s), n_1 := \#1(s)$. The statistic used is

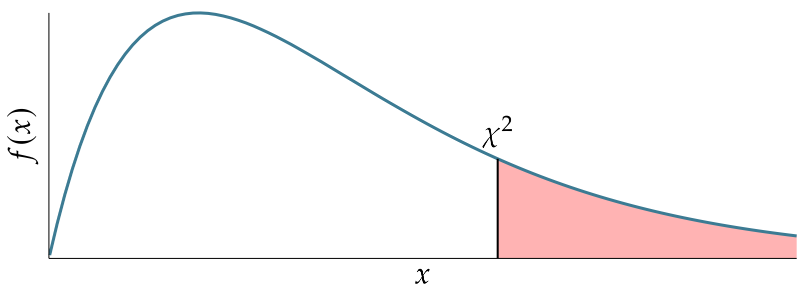
$$X_1 = \frac{(n_0 - n_1)^2}{n},$$

which approximately follows a χ^2 distribution with 1 degree of freedom if $n \geq 10$.

Since $n_0 = 58, n_1 = 42$ and $n = 100$, we have

$$X_1 = \frac{(58-42)^2}{100} = 2.56.$$

Note that



The colored area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$	$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$	$\chi^2_{0.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879

~~The critical value from the χ^2 distribution table at a significance level of 0.01(1%) for 1 degree is approximately 6.635. Since $X_1 = 2.56 < 6.635$, the test passes.~~

(b) Recall that

Two-Bit Test

Let $n_{00}, n_{01}, n_{10}, n_{11}$ denote the number of occurrences of 00, 01, 10, 11 in s, respectively. These counts allow for overlapping. For example, in the sequence 00110, 00 and 01 overlap; they are not counted as separate, non-overlapping pairs. Then

$$n_{00} + n_{01} + n_{10} + n_{11} = (n-1).$$

The statistic used is

$$X_2 = \frac{4}{n-1}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1$$

which approximately follows a χ^2 distribution with 2 degrees of freedom if $n \geq 21$.

Since $n_{00} = 33, n_{01} = 25, n_{10} = 24$ and $n_{11} = 17$, we have

$$X_2 = \frac{4}{99}\left(19^2 + 26^2 + 25^2 + 13^2\right) - \frac{2}{100}(58^2 + 42^2) + 1 \approx 2.64.$$

df	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$	$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$	$\chi^2_{0.005}$
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597

□

8.5 Runs Test for Binary Sequences

(a) Given a bit sequence $X_1, \ldots, X_n$ of length n, where each bit is independent and identically distributed according to a uniform distribution, define the number of runs of length k as the random variable $R_k^{(n)}$. Find the expected value $\mathbb{E}[R_k^{(n)}]$ of this random variable.

(b) When given a sample bit sequence of length 50 as:

```
01100 00001 00011 01001 10001 00110 00110 01010 00101 11000,
```

describe the process of calculating the value of the statistic V_{50}, which represents the total number of runs, according to the Runs Test in NIST SP 800-22.

Solution.

(a)

$$\begin{aligned}
\mathbb{E}[R_k^{(n)}] &= \sum_{i=1}^{n-k+1} P(\text{Run of length } k \text{ starting at position } i) \\
&= \sum_{i=1}^{n-k+1} P(X_i \neq X_{i-1}, X_i = X_{i+1} = \cdots = X_{i+k-1}, X_{i+k} \neq X_{i+k-1}) \\
&= \sum_{i=1}^{n-k+1} \left[P(X_i \neq X_{i-1}) \cdot P(X_i = X_{i+1} = \cdots = X_{i+k-1}) \cdot P(X_{i+k} \neq X_{i+k-1})\right] \\
&= \sum_{i=1}^{n-k+1} \left[\frac{1}{2} \cdot \left(\frac{1}{2}\right)^{k-1} \cdot \frac{1}{2}\right] \\
&= \sum_{i=1}^{n-k+1} \frac{1}{2^{k+1}} \\
&= \frac{n-k+1}{2^{k+1}}
\end{aligned}$$

(b) Recall that

Runs Test

(i) Estimating the Proportion $\pi = \frac{\sum_{i=1}^{n} \varepsilon_i}{n}$: proportion of 1's (or 0's)

(ii) Asymptotic Distribution $\lim_{n\to\infty} \Pr\left[\frac{V_n - 2n\pi(1-\pi)}{\sqrt{2n}\pi(1-\pi)} \le z\right] = \Phi(z)$.

(iii) The Test Statistic V_n is the total number of runs in the sequence:

$$V_n = \sum_{k=1}^{n-1} r(k) + 1 \quad \text{with} \quad r(k) = \begin{cases} 0 & : \varepsilon_k = \varepsilon_{k+1}, \\ 1 & : \varepsilon_k \neq \varepsilon_{k+1}. \end{cases}$$

(iv) P-Value $= \operatorname{erfc}\left(\frac{|V_n(\text{obs}) - 2n\pi(1-\pi)|}{\sqrt{2n}\pi(1-\pi)}\right)$.

Thus,

$$V_{50} = 24 + 1 = 25.$$

□

Definition 8.2.

Definition 8.3.

Domain Range

$$\boxed{\text{Probability distributions of } X} \xrightarrow{H(X)} \boxed{\text{Entropy in bits}}$$