# Contents

1 Probability review
   1.1 Introduction .................................................. 1
   1.2 Random variables and distributions .......................... 3
   1.3 Convergence of random variables ............................. 6

2 Monte Carlo methods
   2.1 Introduction .................................................. 13
   2.2 Pseudo-random numbers ........................................ 14
   2.3 Tests for studying random data .............................. 16
   2.4 Monte Carlo integration ...................................... 18
      2.4.1 Basic Monte Carlo integration .......................... 18
      2.4.2 Variance reduction ...................................... 21
   2.5 Quasi-Monte Carlo methods ................................. 23
      2.5.1 Quasi-random numbers .................................. 24

3 Nonlinear filtering
   3.1 Introduction .................................................. 27
   3.2 Kalman filter .................................................. 28
   3.3 Kalman filter extensions ..................................... 33
      3.3.1 Extended Kalman filter ................................. 33

Bibliography ......................................................... 39
Chapter 2

Probability review

2.1 Introduction

We start by briefly reviewing basic probability concepts. This introduction loosely follows [2].

Definition 2.1. The set of all possible outcomes of an experiment is called the sample space and is denoted by $\Omega$.

An experiment may be as simple as throwing a die once; in which case we have $\Omega = \{1, 2, 3, 4, 5, 6\}$. Another experiment could consist in measuring a patient’s temperature at a given time, $\Omega$ is then an entire interval.

Definition 2.2. An event is a subset of $\Omega$.

Consider the event “getting an even number” when throwing the above die; the corresponding event is $A = \{2, 4, 6\}$. If the above patient has a fever, the event is $A = [99.5, \infty)$. There is an obvious difference between our two experiments: in one case, $\Omega$ is countable, in the other, it is not. This, unfortunately, plays a role when we start trying to “measure” events, i.e., to define their “probability”. In the case of uncountable sample spaces $\Omega$, such measures will only be defined on specific subsets of $\Omega$ called $\sigma$-algebras.

Definition 2.3. A collection $\mathcal{F}$ of subsets of $\Omega$ is a $\sigma$-algebra if

1. $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$,

2. if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$,

3. if $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$,

where $A^c$ denotes the complement of $A$ in $\Omega$. 
We would like to be able to analyze the likelihood of the occurrence of events. For instance, let us throw the above die (assumed fair) \( N \) times and look for the event \( A = \{1\} \). If \( N(A) \) stands for the number of times \( A \) occurs during these \( N \) trials, we expect \( N(A)/N \) to converge to \( 1/6 \). This “limit” is what we will call the probability \( P(A) \) that \( A \) occurs at any particular trial.

**Definition 2.4.** A probability measure \( P \) is a function \( P : \mathcal{F} \to [0, 1] \) satisfying

1. \( P(\Omega) = 1 \),
2. if \( A_1, A_2, \ldots \) is a collection of disjoint members of \( \mathcal{F} \), i.e., \( A_i \cap A_j = \emptyset, i \neq j \), then \( P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \).

In the case of the die, we have \( \Omega = \{1, 2, 3, 4, 5, 6\} \) and we can take \( \mathcal{F} \) as the powerset of \( \Omega \), i.e., \( \mathcal{F} = 2^\Omega \) = set of all the subsets of \( \Omega \). If \( A_i \) is the event \( \{i\} \), then \( P(A_i) = p_i = p, i = 1, \ldots, 6 \), since the die is fair. Further

\[
1 = P(\Omega) = P(\bigcup_{i=1}^{6} A_i) = \sum_{i=1}^{6} P(A_0) = 6p,
\]

and thus \( P(A_i) = 1/6, i = 1, \ldots, 6 \). More generally, \( P(A) = \frac{|A|}{6} \) where \( |A| \) is the cardinality of \( A \).

**Definition 2.5.** If \( P(B) > 0 \) then \( P(A|B) \), the conditional probability that \( A \) occurs given that \( B \) occurs, is

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

Let us throw two dice. Clearly, \( \Omega = \{1, 2, 3, 4, 5, 6\}^2 \), \( \mathcal{F} \) can be taken as the set of all subsets of \( \Omega \) and \( P(A) = |A|/36 \) for any \( A \subset \Omega \). Given that the first die shows 4, what is the probability that the total exceeds 6? Let \( B \) be the event that the first die shows 4, i.e., \( B = \{(4, b), b = 1, 2, 3, 4, 5, 6\} \) and let \( A \) be the event that the total exceeds 6, i.e., \( A = \{(a, b), a + b > 6\} \). Therefore

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B|}{|B|} = \frac{|\{(4, 3), (4, 4), (4, 5), (4, 6)\}|}{6} = 2/3.
\]

Let’s consider another example. Neonatal screening for congenital hypothyroidism (CH) is done routinely. In one study [5], 430,764 children were screened. The test was positive for 772 of them, however, only 224 actually had CH; that is, there were 548 false positives. Further, 13 children actually had CH but tested negative (false negative). In terms of conditional probabilities, we thus have

\[
P(+) = \frac{548}{430,764} \approx .0012722, \quad P(-|\text{well}) = \frac{13}{430,764} \approx 3.0179(5).
\]
2.2 Random variables and distributions

The conditional probability that an individual who has the condition tests positive is then

\[ P(+|\text{CH}) = 1 - P(-|\text{CH}) \approx .99997. \]

On the other hand, the conditional probability that an individual who tests positive actually has the condition is

\[ P(\text{CH}|+) = \frac{P(\text{CH} \cap +)}{P(+)} = \frac{P(\text{CH})P(+|\text{CH})}{P(+)} = \frac{224}{772} \left(1 - \frac{13}{430,764}\right) \approx .29. \]

The previous formula, i.e.,

\[ P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (2.1) \]

is usually referred to as Bayes Theorem, see also Exercise 1.4.

2.2 Random variables and distributions

A random variable \( X : \Omega \to \mathbb{R} \) is a function that maps \( \Omega \) into \( \mathbb{R} \) with the additional property that it is possible to assign probabilities the occurrence of the various values; in other words

\[ \{ \omega \in \Omega; X(\omega) \leq x \} \in \mathcal{F} \quad \forall x \in \mathbb{R}. \]

The distribution function of \( X \) is the function \( F : \mathbb{R} \to [0,1] \)

\[ F(x) = P(X \leq x). \]

Going back to the fair die example, we can define \( X(i) = i, \quad i = 1, \ldots, 6 \) and the corresponding distribution function \( F \) is displayed in Figure 1.1.

![Figure 2.1. Probability distribution function for a fair die.](image-url)
In the case of the feverish patient, if \( \omega \) is the event “the patient’s temperature is \( T \)”, then we can just take \( X(\omega) = T \). The corresponding distribution function is not immediately clear at this point. In this latter example, as opposed to the die example, the random variable is continuous.

For a discrete random variable \( X \), the **mass function** \( f : \mathbb{R} \to [0,1] \) is defined as

\[
f(x) = P(X = x).
\]

Its **expected value** is

\[
E(X) = \sum x f(x).
\]

**Example 2.6** A *Bernoulli* variable \( X \) takes value 1 and 0 with probability \( p \) and \( 1 - p \) respectively. We have

\[
f(0) = 1 - p, f(1) = p.
\]

**Example 2.7** If \( n \) independent Bernoulli variables are considered together with the total sum \( Y = X_1 + X_2 + \cdots + X_n \), then the mass function of \( Y \) is the **binomial distribution** is (exercise)

\[
f(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{2.2}
\]

**Example 2.8** The binomial distribution is a particular case of the multinomial distribution. Assume that instead of having just two categories as with the Bernoulli variables (0 or 1, heads or tails, etc...), we have \( k \) mutually exclusive categories. Assume further that some theory or *null hypothesis* gives us the probability \( p_i \) that an observation falls into the \( i \)-th category (of course \( \sum_{i=1}^{k} p_i = 1 \)). Then after \( n \) trials, the probability of having \( x_1 \) results in category 1, \( x_2 \) results in category 2, ..., \( x_k \) results in category \( k \) is given by

\[
\frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}. \tag{2.3}
\]

Let’s now turn to continuous random variables.

Assume the distribution function is such that \( F'(x) \) exists. Then we define the **probability density function (PDF)** as

\[
f(x) = F'(x).
\]

By definition, we clearly have (exercise)

\[
F(x) = \int_{-\infty}^{x} f(y) \, dy, \quad \int_{-\infty}^{\infty} f(y) \, dy = 1, \quad P(a \leq X \leq b) = \int_{a}^{b} f(y) \, dy.
\]
Example 2.9 The most important example of a continuous distribution is the normal (or Gaussian) distribution

\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty, \tag{2.4} \]

which has two parameters \( \mu \) and \( \sigma \). If the random variable \( X \) admits (1.4) as its PDF then we’ll say \( X \) is \( N(\mu, \sigma^2) \).

Example 2.10 If \( X \) admits

\[ f(x; k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}, \quad x \geq 0, \tag{2.5} \]

where \( k \) is a positive integer, as a PDF, then \( X \) is said to have the chi-squared distribution \( \chi^2(k) \) with \( k \) degrees of freedom.

The expectation (or expected value or mean) of \( X \) with PDF \( f \) is by definition

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

Let \( g \) be a continuous function; then if \( X \) is a random variable, so is \( g(X) \) and

\[ E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) \, dx. \]

The special cases

\[ E(X^n) = \int_{-\infty}^{\infty} x^n f(x) \, dx \quad \text{and} \quad E((X - E(X))^n) = \int_{-\infty}^{\infty} (x - E(X))^n f(x) \, dx, \]

are respectively the \( n \)-th moment and the \( n \)-th centered moment of \( X \). The variance of \( X \) is by definition

\[ \text{var}(X) = E((X - E(X))^2). \]

For instance, if \( X \) is \( N(\mu, \sigma^2) \) then \( E(X) = \mu \) and \( \text{var}(X) = \sigma^2 \), \( \sigma \) being the standard deviation, whereas if \( X \) is \( \chi^2(k) \), then \( E(X) = k \) and \( \text{var}(X) = 2k \), see Exercise 1.7. The two above distributions are displayed in Figure 1.2. It can be shown that they are closely linked to each other, see Exercise 1.8.

Two events \( A \) and \( B \) are called independent if \( P(A \cap B) = P(A)P(B) \). Two random variables \( X \) and \( Y \) are called independent if the events \( \{\omega \in \Omega; X(\omega) \leq x\} \) and \( \{\omega \in \Omega; Y(\omega) \leq y\} \) are independent for all \( x \) and \( y \).

The joint distribution function of two random variables \( X \) and \( Y \) is

\[ F(x, y) = P(X \leq x, Y \leq y). \]

If the second mixed partial derivative \( \frac{\partial^2 F}{\partial x \partial y} \) exists then

\[ F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) \, dudv, \]
where \( f \) is the joint density function. Clearly, \( X \) and \( Y \) are independent if and only if \( F(x, y) = F_X(x)F_Y(y) \) or if and only if \( f(x, y) = f_X(x)f_Y(y) \) where \( F_X \) and \( f_X \) are the distribution and PDF of \( X \) and similarly for \( Y \).

The covariance of \( X \) and \( Y \) is defined as

\[
\text{cov}(X, Y) = E((X - E(X))(Y - E(Y))).
\]

Equivalently, we have \( \text{cov}(X, Y) = E(XY) - E(X)E(Y) \). By definition, two variables \( X \) and \( Y \) are uncorrelated if \( \text{cov}(X, Y) = 0 \).

One can of course consider more than two random variables together; this is the multivariate case. For instance, let \( X : \Omega \rightarrow \mathbb{R}^n \) be the random vector \( X = (X_1, \ldots, X_n) \). The vector \( X \) is said to have multivariate normal distribution if

\[
f(x) = \frac{1}{\sqrt{(2\pi)^n \det V}} \exp\left(-\frac{1}{2}(x - \mu)V^{-1}(x - \mu)^T\right),
\]

where \( V \) is a symmetric positive definite matrix called the covariance matrix since \( V_{ij} = \text{cov}(X_i, X_j) \), see Exercise 1.9.

### 2.3 Convergence of random variables

Many of the methods to be discussed below (Monte Carlo, etc...) involve sequences of random variables. What can be said about the convergence?

**Lemma 2.11 (Markov’s inequality).** If \( X \) is a random variable with finite mean then

\[
P(|X| \geq a) \leq \frac{1}{a}E(|X|), \quad \text{for any } a > 0.
\]
2.3. Convergence of random variables

**Proof.** Let \( \chi_A \) be the indicator function of the event \( A \), i.e.,

\[
\chi_A(\omega) = \begin{cases} 
1 & \text{if } \omega \in A, \\
0 & \text{if } \omega \in A^c.
\end{cases}
\]

Now, let \( A = \{|X| \geq a\} \). By direct inspection, we have

\[
a \chi_A \leq |X|.
\]

Taking the expectation proves the result. \( \Box \)

**Lemma 2.12 (Chebyshev’s inequality).** Let \( X \) be a random variable with mean \( \mu \) and variance \( \sigma^2 \). Then

\[
P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.
\]

**Proof.** We have

\[
P(|X - \mu| \geq a) = P((X - \mu)^2 \geq a^2).
\]

Applying Lemma 1.11 achieves the proof. \( \Box \)

Chebyshev’s inequality says that if the variance of a random variable is small, then the random variable is concentrated around its mean. This is important, for instance, if we are using \( X \) as an estimator of \( E(X) \). In the case of a random variable with small variance, \( X \) is thus a good estimator of \( E(X) \).

This “concentration around the mean” can also be observed if we have a sequence of independent identically distributed (i.i.d.) random variables.

**Theorem 2.13 (Weak law of large numbers).** Let \( X_1, X_2, \ldots \) be a sequence of i.i.d. random variables, each with mean \( \mu \) and variance \( \sigma^2 \). Then for every \( \epsilon > 0 \)

\[
\lim_{n \to \infty} P\left( \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| > \epsilon \right) = 0.
\]

**Proof.** By elementary properties of the expected value and the variance, we have

\[
E\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu \quad \text{and} \quad \text{var}\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{var}(X_i) = \frac{\sigma^2}{n}.
\]

We can then apply Lemma 1.12 to the random variable \( \frac{1}{n} \sum_{i=1}^{n} X_i \) and get

\[
P\left( \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| > \epsilon \right) \leq \frac{1}{\epsilon^2} \frac{\sigma^2}{n}.
\]

Taking the limit \( n \to \infty \) achieves the proof. \( \Box \)
The typical example is here that of a fair coin tossed repeatedly. If $X_i$ is the random variable taking value 1 if the coin shows heads at the $i$-th toss and taking value 0 in case of tails, we clearly expect the average of the $X_i$'s to converge to $1/2$. Note that Theorem 1.13 deals with a very specific notion of convergence.

**Definition 2.14.** A sequence of random variables $\{X_n\}$ is said to converge in probability to the random variable $X$, written $X_n \overset{p}{\rightarrow} X$, if for any $\epsilon > 0$

$$\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0.$$  

Other notions of convergence of random variables are useful. Statisticians sometimes care more the distribution of a random variable than the random variable itself. Hence the

**Definition 2.15.** A sequence of random variables $\{X_n\}$ is said to converge weakly or in distribution or in law to the random variable $X$, written $X_n \overset{D}{\rightarrow} X$, if the distribution of $X_n$ converges weakly to the distribution of $X$, i.e.,

$$F_n(x) \to F(x) \quad \text{whenever } F \text{ is continuous at } x,$$

where $F_n$ and $F$ are the distribution functions of $X_n$ and $X$ respectively.

Note that convergence in probability is stronger than convergence in distribution, i.e.,

$$X_n \overset{p}{\rightarrow} X \Rightarrow X_n \overset{D}{\rightarrow} X,$$

see [2], § 7.2. The converse does not hold, see however Exercise 1.10.

The weak law above says that the sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ converges to $\mu$. It says nothing, however, on the distribution of $\bar{X}_n$. The Central Limit Theorem does (see [2], § 5.10 for a proof).

**Theorem 2.16 (Central Limit Theorem).** Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, each with mean $\mu$ and variance $\sigma^2$. Let

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}}.$$  

Then $Z_n \overset{D}{\rightarrow} Z$ where $Z$ is $N(0,1)$.

**Exercises for Chapter 1**

1.1 Prove the following properties

1. $P(\emptyset) = 0$
2. $P(A^c) = 1 - P(A)$
2.3. Convergence of random variables

3. If \( A \subset B \) then \( P(B) = P(A) + P(B - A) \geq P(A) \)

4. \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \)

1.2 A family has two children. What is the probability that both are boys given that at least one of them is a boy? Hint: the answer is not 1/2.

1.3 For any events \( A \) and \( B \), prove that

\[
P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).
\]

1.4 1. Show that Bayes' Theorem can be rewritten as

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.\]

2. Show that if \( \{A_i\} \) is a partition of \( \Omega \) then Bayes' Theorem takes the form

\[
P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}.
\]

1.5 A gambler has \( k \)$. She has made the following agreement with a rich friend. She tosses a coin repeatedly; if comes up heads, the friend pays her 1$ however it comes up tails, then she pays the friend 1$. 

1. If \( p_k \) stands for the probability of ultimate ruin starting from \( k \), derive a difference equation for \( p_k \).

2. Solve the previous difference equation with boundary conditions \( p_0 = 1 \) and \( p_N = 0 \). Hint: try \( p_k = \theta^k \). Comment on the properties of the solution. You may have to differentiate between the case of a fair or biased coin.

1.6 Consider a two-dimensional random walk on a regular lattice.

1. Assuming equal probability of going up, down, left or right, write a MATLAB code generating such walks. Carefully explain how “decisions” are made (direction of the next step). Hint: use the MATLAB function \texttt{rand}.

2. Pólya [15] showed in 1921 that the probability of returning to the initial position is 1 (the problem is often described as that of a drunkard leaving the pub, walking “randomly” through town and eventually getting back to the pub with probability 1, assuming no mugging). Justify/investigate this assertion in any way you want. You have to convince the reader that the statement is correct; copying proofs from the literature is not an option.

3. Do the same in the three-dimensional case (where it can be proven that the probability of return is strictly less than 1, \( \approx 0.34\ldots, [1] \)).
1.7 1. Prove that (1.4) is a PDF.
2. Show that if $X$ is $N(\mu, \sigma^2)$ then $E(X) = \mu$ and $\text{var}(X) = \sigma^2$.
3. Show that if $X$ is $N(\mu, \sigma^2)$ then $aX + b$ is $N(a\mu + b, (a\sigma)^2)$.
4. Show that if $X$ is $\chi^2(k)$, then $E(X) = k$ and $\text{var}(X) = 2k$.

1.8 1. Let $X_1$ and $X_2$ be two independent $N(0, 1)$ variables. Show that $Z = X_1^2 + X_2^2$ is $\chi^2(2)$. Hint: express the distribution function of $Z$ as an integral and use polar coordinates.
2. Let $X_i$, $i = 1, \ldots, n$ be $n$ independent $N(0, 1)$ variables. Show that $Z = \sum_{i=1}^{n} X_i^2$ is $\chi^2(n)$.

1.9 Show that if the vector $X$ admits (1.6) as its PDF then
1. $E(X) = \mu$, i.e., $E(X_i) = \mu_i$, $i = 1, \ldots, n$.
2. $V_{ij} = \text{cov}(X_i, X_j)$.

1.11 Let $X$ be a Poisson variable with rate $\lambda$ ($\lambda$ is an integer), i.e., $X$ is a discrete random variable with mass function

$$f(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

1. Show that $E(X) = \text{var}(X) = \lambda$.
2. One can think of $X$ as the number of occurrences in an experiment that runs for a time $\lambda$. Justify the fact that $X$ can be alternatively viewed as $\lambda$ experiments running for time 1, i.e.,

$$X = \sum_{i=1}^{\lambda} X_i,$$

where each $X_i$ is Poisson with rate 1.
3. Show that for $\lambda$ large, $X$ is approximately normally distributed, more precisely, $X$ is approximately $N(\lambda, \lambda)$. Hint: show first that $\frac{X - \lambda}{\sqrt{\lambda}}$ is approximately $N(0, 1)$ using the Central Limit Theorem 1.16 and conclude by Exercise 1.7.3.

1.12 Let $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^m$ be jointly Gaussian random variables admitting the joint density function

$$f_{X,Y}(x, y) = \frac{1}{(2\pi)^{(n+m)/2} \det V} \exp \left( -\frac{1}{2} \begin{bmatrix} x - \mu_X \\ y - \mu_Y \end{bmatrix}^T V^{-1} \begin{bmatrix} x - \mu_X \\ y - \mu_Y \end{bmatrix} \right)$$

where $V = \begin{bmatrix} V_{XX} & V_{XY} \\ V_{YX} & V_{YY} \end{bmatrix}$.
The conditional density function \( f_{X|Y} \) is by definition

\[
f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.
\]

Check that

\[
f_{X|Y}(x, y) = \frac{1}{(2\pi)^{n/2}\sqrt{\det V_{X|Y}}} \exp\left( -\frac{1}{2} \left[ x - \mu_{X|Y} \right]^T V_{X|Y}^{-1} \left[ x - \mu_{X|Y} \right] \right),
\]

where

\[
\mu_{X|Y} = \mu_X + V_{XY} V_Y^{-1} (y - \mu_Y),
\]

\[
V_{X|Y} = V_{XX} - V_{XY} V_Y^{-1} V_Y X.
\]

1.13 Consider the bivariate normal case, i.e., \( n = m = 1 \) in Exercise 1.12. Let

\[
V = \begin{bmatrix}
\sigma_X^2 & \rho \sigma_X \sigma_Y \\
\rho \sigma_X \sigma_Y & \sigma_Y^2
\end{bmatrix}
\]

be the covariance matrix.

1. Show that

\[
f(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1 - \rho^2}} \exp\left\{ \frac{1}{1 - \rho^2} \left( \frac{(x - \mu_X)^2}{\sigma_X^2} - 2\rho \frac{1}{\sigma_X \sigma_Y} (x - \mu_X)(y - \mu_Y) + \frac{(y - \mu_Y)^2}{\sigma_Y^2} \right) \right\}.
\]

2. Assuming the pdf from point 1., show that
   - \( X \) is \( N(\mu_X, \sigma_X^2) \) and \( Y \) is \( N(\mu_Y, \sigma_Y^2) \),
   - the correlation between \( X \) and \( Y \) is \( \rho \),
   - \( X \) and \( Y \) are independent if and only if \( \rho = 0 \).

3. Show that if \( A \) and \( B \) are two \( N(0, 1) \) variables, then

\[
X = \mu_X + \sigma_X A,
\]

\[
Y = \mu_Y + \sigma_Y \left( \rho A + \sqrt{1 - \rho^2} B \right),
\]

have respectively mean \( \mu_X \) and \( \mu_Y \), standard deviation \( \sigma_X \) and \( \sigma_Y \) and correlation \( \rho \). In other words, \textit{this is a simple way to sample from the distribution from point 1}.

1.14 1. Let \( X \) be \( N(\mu, \sigma^2) \). Show by direct integration of the pdf that the probability \( P \) that a measurement falls within \( n \) standard deviation \( \sigma \) of the mean \( \mu \), i.e., within \([\mu - n \sigma, \mu + n \sigma]\), is given by

\[
P = \text{erf}\left( \frac{n}{\sqrt{2}} \right),
\]

where \( \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \) is the error function.
2. A confidence interval is an interval in which a trial falls corresponding to a given probability. Show that the probability-$P$ confidence interval centered at $\mu$ in units of $\sigma$ is given by

$$n = \sqrt{2} \text{erf}^{-1}(P).$$

Plot the size of the confidence interval as a function of $P$. Hint: you may want to consider the MATLAB function `norminv` and/or Exercise 2.6.

3. The natural generalization of confidence interval to bivariate variables are confidence ellipses. Write a MATLAB routine that given samples from a bivariate normal distribution plots the confidence ellipse for a given value of the probability $P$. 
Bibliography


[14] Karl Pearson, *On the criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, Philosophical Mag., 50 (1900), pp. 157–175.
