Lecture II

Probabilities and Random Variables

Sourendu Gupta
Tata Institute of Fundamental Research,
Homi Bhabha Road, Mumbai 400004, India.
January–April, 2002.

1 Basic Probability

The most easily understood situations in life are when an action gives rise to a definite outcome— press a switch and a light comes on, push an object and it moves. In such situations physics makes quick progress, uncovering the laws of nature. A thrown object traces a parabola— whether it is a stone, a football or a ballistic missile. Deviations from these laws can be understood in terms of other laws— a swinging cricket ball involves the dynamics of a spinning sphere moving in a fluid: a complicated but solvable problem. Some influences however, are random— try playing badminton in an open space between tall buildings and you'll quickly realize that there are so many eddies and cross currents that your effort to send the shuttlecock in one direction does not always have the result that you expect. When such random influences cannot be removed by controlled experiments (such as always playing badminton matches in indoor courts), then we are driven to use the notion of probability.

Probabilities are needed to describe situations where the results of actions do not obviously depend on the causes. We start by enumerating all the possible outcomes. Then we use experiments to define a probability for each outcome. The remainder of this section sets out these definitions and follows out the basic common-sense logic that follows.

1.1 Event space and ensembles

Consider an experiment with the set of outcomes $\Omega = \{O_i | i = 1, \cdots, M\}$. Denote subsets of the event space, Ω, by $\mathbf{O}_\alpha = \{O_{k_i} | i = 1, \cdots, M_\alpha\}$. While the notation seems to imply that the **event space** is discrete, this need not be so. In order to develop the notions of probability, we will consider an **ensemble** of experiments, *i.e.*, a set of N completely independent identical copies of this experiment. By independent we shall for now mean that the outcome in one experiment does not influence that in another (we shall refine this concept later). By identical we mean identical in setup, and not in outcome.

Problem set II.1

1. *A coin is tossed before a cricket match. What is the event space?*

2. *A baby is born. What is the event space of its sex?*

3. *A game of badminton is being played outdoors on a very windy day. If the players are only interested in hitting the shuttle across the net, what is the event space?*

4. *Waves come randomly into the shore from the sea. What is the event space?*

1.2 Measuring and combining probabilities

The **probability** of an outcome set $\mathbf{O}_i \subset \Omega$ is

$$P(\mathbf{O}_i) \;=\; \lim_{N\to\infty} \frac{N_i}{N} \tag{1}$$

where N_i is the number of times the outcome $\mathbf{O}_i$ was observed in the ensemble of N identical and independent experiments. Clearly, $P(\mathbf{O}_i)$ is a real number and $0 \le P(\mathbf{O}_i) \le 1$. The two identities

$$\text{probability of something happening} \;\equiv\; P(\Omega) \;=\; 1, \tag{2}$$
$$\text{probability of nothing happening} \;\equiv\; P(\phi) \;=\; 0, \tag{3}$$

are also clear. Figure 1 shows a pictorial representation of these statements.

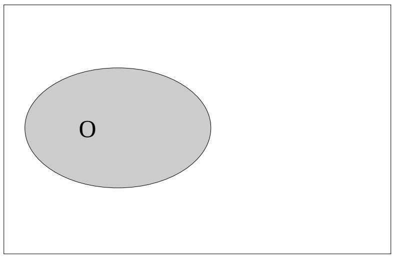

Figure 1: Think of the probability of the outcome set $\mathbf{O}$ as the area of the ellipse relative to the area of the box. The box is the event set Ω. The probability of Ω is clearly 1. The null set, ϕ, has zero area and hence zero probability.

Given two outcome sets $\mathbf{O}_1$ and $\mathbf{O}_2$, both subsets of Ω, the probability of occurrence of any event in the union of the two is given by

$$P(\mathbf{O}_1 \cup \mathbf{O}_2) \;=\; P(\mathbf{O}_1) + P(\mathbf{O}_2) - P(\mathbf{O}_1 \cap \mathbf{O}_2). \tag{4}$$

This is the primary law for combining probabilities of two events.

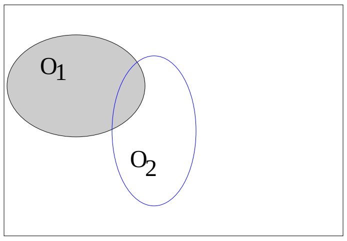

Figure 2: Eq. (4) gives the area of the union of the two sets in terms of their individual areas and the area of their intersection.

Problem set II.2

1. *What is the probability of finding a male in the event space of this class?*

2. *What is the probability that there are at least 10 people ahead of you in the queue for lunch between 1 and 1:10 PM in the west canteen?*

3. *What is the probability of finding a lift on the ground floor of the institute? Does it depend on the time of the day?*

1.3 Conditional probabilities and independence

The conditional probability of an outcome being in $\mathbf{O}_1$, given that it is also in $\mathbf{O}_2$, is

$$P(\mathbf{O}_1|\mathbf{O}_2) \;=\; \frac{P(\mathbf{O}_1 \cap \mathbf{O}_2)}{P(\mathbf{O}_2)}. \tag{5}$$

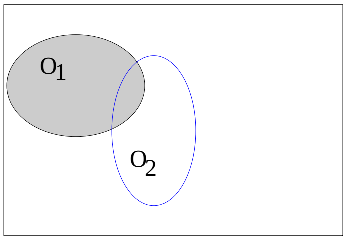

Figure 3: The conditional probability $P(\mathbf{O}_1|\mathbf{O}_2)$ is the area of the intersection $\mathbf{O}_1 \cap \mathbf{O}_2$ relative to the area of $\mathbf{O}_2$.

Two event sets $\mathbf{O}_1$ and $\mathbf{O}_2$ are **independent** when $P(\mathbf{O}_1|\mathbf{O}_2) = P(\mathbf{O}_1)$. The most important ingredient in the notion of independence of two pieces of information is whether one piece gives some deductive power about the second. This notion can be tricky, as shown in the following example.

Example II.1 A fair die is cast. The event space of outcomes is $\Omega = \{1, 2, 3, 4, 5, 6\}$, and each event has probability $1/6$. Take two events—

$$\mathbf{O}_1 = \{3, 4, 5, 6\} \qquad \text{and} \qquad \mathbf{O}_2 = \{1, 3, 5\}.$$

Clearly $P(\mathbf{O}_1) = 2/3$ and $P(\mathbf{O}_2) = 1/2$. Is $\mathbf{O}_2$ independent of $\mathbf{O}_1$? Yes, since $P(\mathbf{O}_2|\mathbf{O}_1) = P(\mathbf{O}_2 \cap \mathbf{O}_1)/P(\mathbf{O}_1) = (2/6)/(2/3) = 1/2 = P(\mathbf{O}_2)$!

How can two events be independent although they depend on one single cast of a die? The situation is worth analysing in a different way. Before the die was cast, the probability of getting an odd number (the event $\mathbf{O}_2$) was

1/2. After the die was rolled and you were told that neither 1 nor 2 appeared (the event was in $\mathbf{O}_1$), but the probability of getting an odd number did not change. As a result the events remained independent.

The following, more informal, example may make the notion even clearer. A coin is to be tossed. You guess the probability of a head to be 1/2. After the coin is tossed someone informs you that it was a silver coin. This does not change the probability of it coming up heads.

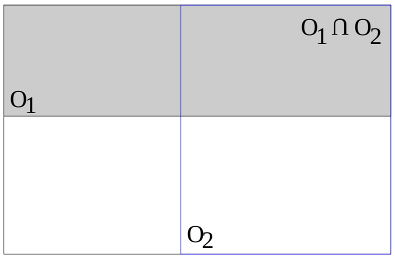

Figure 4: $\mathbf{O}_1$ and $\mathbf{O}_2$ are independent if the area of $\mathbf{O}_1 \cap \mathbf{O}_2$ is the same fraction of $\mathbf{O}_2$ as the area of $\mathbf{O}_1$ is of Ω.

From the logic of the previous paragraph, or the interpretation in Figure 4, it is clear that

$$P(\mathbf{O}_1 \cap \mathbf{O}_2) = P(\mathbf{O}_1)P(\mathbf{O}_2) \text{ if and only if } \mathbf{O}_1 \text{ and } \mathbf{O}_2 \text{ are independent.} \quad (6)$$

Independence of a collection of event sets $\mathbf{O}_k \subset \Omega$, $1 \leq k \leq \mathcal{M}$ is defined by the relation $P(\mathbf{O}_1 \cap \cdots \cap \mathbf{O}_{\mathcal{M}}) = \prod_{i=1}^{\mathcal{M}} P(\mathbf{O}_i)$.

Problem set II.3

1. *Is the sex of twins independent events? Is the sex of identical twins independent? Is the sex of non-identical twins independent?*
2. *Is the probability of any two people finding at least 10 people ahead of them in the lunch queue independent of each other?*
3. *Consider repeated tosses of a fair coin. The event space is* $\Omega = \{\mathrm{H}, \mathrm{T}\}$ *and* $P(\mathrm{H}) = P(\mathrm{T}) = 1/2$.
   (a) *What is the event space for three successive tosses? What is the probability of getting H on the first two tosses and T on the third? What is the probability of getting exactly two H in any order?*

(b) *In successive tosses, if there is a run of 83 H, then what is the probability of getting a T on the 84-th toss?*

1.4 Bayes' theorem

For two events $\mathbf{O}_1$ and $\mathbf{O}_2$ in Ω we have the two identities

$$P(\mathbf{O}_1 \cap \mathbf{O}_2) = P(\mathbf{O}_1)P(\mathbf{O}_2|\mathbf{O}_1), \qquad P(\mathbf{O}_1 \cap \mathbf{O}_2) = P(\mathbf{O}_2)P(\mathbf{O}_1|\mathbf{O}_2). \quad (7)$$

Both follow from eq. (5). Now equating these two expressions we get Bayes' theorem—

$$P(\mathbf{O}_2|\mathbf{O}_1) = \frac{P(\mathbf{O}_2)P(\mathbf{O}_1|\mathbf{O}_2)}{P(\mathbf{O}_1)}. \quad (8)$$

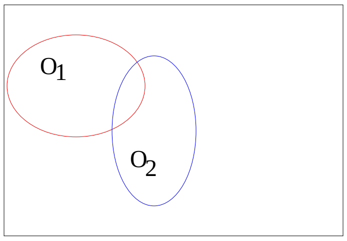

Figure 5: Bayes' theorem relates the area of $\mathbf{O}_1 \cap \mathbf{O}_2$ as a fraction of $\mathbf{O}_1$ to the same area as a fraction of $\mathbf{O}_2$.

Example II.2 (Amitabh Bachchab coins) A coin is to be tossed, and you have to figure out whether this is an ordinary coin with a head and tail (both having equal chances of coming up in a toss) or an Amitabh Bachchan coin with heads on both sides. I suspect that there is a probability $P(A) \neq 0$ that the coin is an Amitabh Bachchan coin. The probability of it being ordinary is therefore $P(O) = 1 - P(A)$. Let us write $\omega = P(O)/P(A)$. $P(A)$ and $P(O)$ are each called a **prior**, because these are expectations before any observation is made.

Now the coin is tossed five times and heads come up each time. Since $P(H^5|A) = 1$ and $P(H^5|O) = 1/2^5$, the unconditional probability $P(H^5) = P(A)P(H^5|A) + P(O)P(H^5|O) = P(A) + P(O)/2^5$. Finally,

$$P(A|H^5) = \frac{P(A)P(H^5|A)}{P(H^5)} = \frac{1}{1 + \omega/2^5}.$$

Our suspicion that the coin is doctored increases very rapidly as the number of heads increases. However, this happens only if we assigned initially $P(A) > 0$. Otherwise, $P(H^n|A) = 0$, no matter how many heads are thrown, because there is a probability, however small, of getting an arbitrarily long run of heads by tossing a normal coin. The moment a tail is seen, however, Bayes' theorem will reduce $P(A|H^nT)$ to zero.

Since the use of Bayes' theorem entails probabilities which exists before a measurement, *i.e.*, the priors, the use of such analysis is often cast in doubt. However, it remains a fact that much of our informal thinking about probabilities involves priors, and this theorem then gives us quantitative ways to think about the evolution of such priors.

Problem set II.4

1. *A certain DNA sequence occurs in a fraction* 10^{-6} *of a population. The test for this has a probability* 0.003 *of a false positive, i.e., of someone without this sequence testing positive. If the only evidence for a person being accused of a crime is such a test, what is the probability of this person actually having committed that crime?*

2. *You are offered a choice of opening one of three fridges. Two are empty. The third is full of beer. You want the beer, and make a random choice of fridges. After this, one of the other fridges is opened to show you that it is empty. Would it improve your chances if you change your choice now? Prove your answer.*

---

2 Probability distributions

A **probability distribution** is a list of the probability for each event in the event space Ω, when it is discrete. If Ω is continuous, then we instead define a **probability distribution function** $P(x)$ for $x \in \Omega$ such that the probability of any $O \subset \Omega$ is given by $P(O) = \int_O dx P(x)$.[1] Recall that $\int_\Omega dx P(x) = 1$ by definition, when $P(x)$ is the probability distribution function. Some commonly encountered distributions and distribution functions

[1]From now on we shall used the symbol $\int_O dx$ to mean an integral over a subspace O of the event space Ω ($x \in O$)if it is continuous, or a sum over O if it is discrete.

are listed in Table 1. Examples of where they may be encountered are given below.

The constant distribution is used as a model for coin tossing experiments: $\Omega = \{\mathrm{H}, \mathrm{T}\}$ and $P(\mathrm{H}) = P(\mathrm{T}) = 1/2$. For the throw of a fair die: $\Omega = \{1, 2, 3, 4, 5, 6\}$, and $P(i) = 1/6$ for $i \in \Omega$.

The binomial distribution with $p = 1/2$ follows for the probability distribution of independent multiple coin tosses, when the order of obtaining H or T is disregarded. Here n is the number of coins tossed and r is, for example, the number of H obtained. The binomial distribution is always obtained in repeated trials of an experiment with two outcomes when the order in which the trials are performed is immaterial.

The multinomial distribution with all $p_i = 1/6$ follows for the probability distribution of independent multiple rolls of a die, when the order of obtaining the scores is disregarded. Here n is the number times the die is rolled and r_i is the number of times a score i is obtained. The multinomial distribution

Name	**Event set**	**Probability**
Constant	$\{i \mid i \in \mathbf{I} \text{ and } 0 \leq i \leq n\}$	$P(i) = 1/n$
Poisson	$\{i \mid i \in \mathbf{I} \text{ and } 0 \leq i \leq \infty\}$	$P_\lambda(i) = \lambda^i \mathrm{e}^{-\lambda}/i!$
Binomial	$\{(i_1, i_2, \cdots i_n) \mid i_\alpha \in Z_2\}$	$P_p(i_1 \cdots i_n) = \binom{n}{r} p^r (1-p)^{n-r}$ where r is the number of times 0 appears in the n-tuplet.
Multinomial	$\{(i_1, i_2, \cdots i_n) \mid i_\alpha \in Z_q\}$	$P_p(i_1 \cdots i_n) = \binom{n}{r_0, \cdots r_{q-2}} p_0^{r_0} p_i^{r_1} \cdots p_{q-1}^{r_{q-1}}$ where r_j is the number of times j appears in the n-tuplet, $\sum_{j=0}^{q-1} p_j = 1$ and $\sum_{j=0}^{q-1} r_j = n$.
Constant	$\{x \mid x \in \mathbf{R} \text{ and } 0 \leq x \leq A\}$	$P(x) = 1/A$
Exponential	$\{x \mid x \in \mathbf{R} \text{ and } 0 \leq x \leq \infty\}$	$P_\mu(x) = \exp(-x/\mu)/\mu$
Gaussian	$\mathbf{R}$	$P_{\mu,\sigma}(x) = \exp[-(x-\mu)^2/2\sigma]/\sqrt{2\pi\sigma^2}$
Cauchy	$\mathbf{R}$	$P(x) = [\pi(1+x^2)]^{-1}$

Table 1: Here $\mathbf{I}$ is the set of integers and $Z_q = \{0, 1, \cdots, q-1\}$. Also $\mathbf{R}$ is the set of real numbers.

is always obtained in repeated trials of an experiment with multiple finite outcomes when the order in which the trials are performed is immaterial.

The Poisson distribution is expected to occur when we count the number of words with spelling errors in a printed page such as this. Assume that the probability of a misspelling a word is $p = 1/2000$, and a book contains 500 words in each page. The probability of finding i mistakes in a given page is the Poisson distribution with $\lambda = 500 \times (1/2000) = 1/4$. For example, $P(i = 3) = \exp(-1/4)/(6.4^3) \simeq 0.002$. Interestingly, the probability of finding n such pages in a book of 200 pages is itself a Poisson distribution. The probability of finding 2 pages with 3 mistakes each in such a book is given by $\lambda = 0.002 \times 200 = 0.4$ and $P(n = 2) \simeq 0.05$.

Problem set II.5

1. *A coin is tossed and a die is rolled simultaneously. What is the event space and probability distribution for this experiment?*

2. *What is the probability distribution for the sexes of twins?*

3. *What is the event space of one of the lifts in the main foyer of the institute? What is the probability distribution?*

4. *For a Binomial distribution with $np = \lambda$ show that as n goes to ∞, the distribution of r H's is Poisson.*

2.1 Moments

An **expectation value** $\langle f(x) \rangle$ for $x \in \Omega$ is defined as

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} dx f(x) \mathcal{P}(x), \tag{9}$$

where $P(x)$ is the probability distribution function. **Moments** $\langle x^n \rangle$, where the corresponding integral exists. The **average** is just $\langle x \rangle$. We will denote the exact value of the average for any distribution by μ. **Central moments** are defined to be $\langle (x - \mu)^n \rangle$. The **variance** Var $x = \langle (x - \mu)^2 \rangle$.

Problem set II.6

1. *Two fair dice are rolled together. What is the expectation value of the sum of the face values? Is the expectation value of the sum of the squares the same as the expectation value of the square of the sum?*

2. *Find the average and variance of each distribution given in Table 1.*

3. *What are the first 10 moments of the Poisson distribution?*

4. *On the state space* $\Omega = \{i | i \in \mathbf{I} \text{ and } 1 \leq 9\}$ *the average of a probability distribution is 5 and the variance is zero. What is the distribution?*

5. *On the state space* $\Omega = \{i | i \in \mathbf{I} \text{ and } 1 \leq 5\}$ *the variance of a probability distribution is zero. Find all moments of the distribution.*

2.2 Generating functions

Several generating functions are of interest—

1. The **moment generating function**, $Z(s)$, for a probability distribution $P(x)$ is
$$Z(s) = \langle \mathrm{e}^{sx} \rangle = \int_{-\infty}^{\infty} dx \mathrm{e}^{sx} P(x). \tag{10}$$
If it exists, then the Taylor series expansion is $Z(s) = \sum_i \langle x^i \rangle s^i / i!$. Clearly,
$$\left\langle x^k \right\rangle = \left. \frac{d^k Z}{ds^k} \right|_{z=0}. \tag{11}$$

2. The **characteristic function**, $\varphi(s)$, of a probability distribution function, $P(x)$, is its Fourier transform—
$$\varphi(s) = \langle \exp(isx) \rangle = Z(is), \tag{12}$$
where the last equality holds only if $Z(s)$ exists. Existence of a characteristic function requires much weaker conditions than the existence of the moment generating function.

3. The **cumulant generating function** for a probability distribution $P(x)$ is given by the relation

$$F(s) = \log Z(s). \tag{13}$$

A **cumulant** of order k, is defined to be

$$\left[x^k\right] = \left.\frac{d^k F(s)}{ds^k}\right|_{s=0}. \tag{14}$$

The variance is the cumulant of order 2, Var $x = [x^2]$.

Example II.3 The exponential probability distribution An example can clarify some of these concepts. Consider the moments of the exponential probability distribution—

$$\langle x^n \rangle = \frac{1}{\mu}\int_0^\infty dx x^n \mathrm{e}^{-x/\mu} = \mu^n \int_0^\infty dy y^n \mathrm{e}^{-y} = n!\mu^n. \tag{15}$$

The moment generating function is given by

$$Z(s) = \langle \mathrm{e}^{sx} \rangle = \frac{1}{\mu}\int_0^\infty dx \mathrm{e}^{(s-1/\mu)x}. \tag{16}$$

The integral converges only for $s < 1/\mu$. In this case we obtain

$$Z(s) = \frac{1}{1-\mu s}. \tag{17}$$

Since the series converges for s near zero, we can obtain the moments by differentiating $Z(s)$ at $s = 0$. Alternatively, we could have written

$$Z(s) = \sum_{n=0}^{\infty} \langle x^n \rangle \frac{s^n}{n!} = \sum_{n=0}^{\infty} (\mu s)^n = \frac{1}{1-\mu s}. \tag{18}$$

It is obvious from this series expansion that the radius of convergence of the series is for $s < 1/\mu$.

The characteristic function is much more well behaved than the moment generating function. We find

$$\phi(s) = \left\langle \mathrm{e}^{isx} \right\rangle = \frac{1}{\mu}\int_0^\infty dx \mathrm{e}^{-(1/\mu - is)x} = \frac{1}{1-i\mu s}. \tag{19}$$

There is no convergence problem with the integral for real s, since $P(x)$ itself provides the damping at large x necessary for defining the integral. The characteristic function has a simple pole at $s = -i/\mu$ in the complex s plane, and is well defined everywhere else.

The cumulant generating function is

$$F(s) \;=\; \log(1 - \mu s). \tag{20}$$

The cumulants of this distribution grow almost as fast as the moments, as can be seen from the fact that

$$[x^n] \;=\; \left.\frac{dF(s)}{ds}\right|_{s=0} \;=\; (n-1)!\mu^n. \tag{21}$$

Problem set II.7

1. *For any probability distribution function $P(x)$ prove the following:*

   (a) $Z(0) = 1$

   (b) $F(0) = 0$.

   (c) *Write the expressions for $[x^i]$ in terms of $\langle x^i \rangle$.*

   (d) *If x has the characteristic function $\varphi(s)$, then $ax + b$ has the characteristic function $\exp(isb)\varphi(as)$.*

   (e) *If $x_1, \cdots, x_n$ each have characteristic function $\varphi(s)$, then the characteristic function of $S_n = \sum_{i=1}^n x_i$ is given by*

   $$\langle \exp(isS_n) \rangle \;=\; \prod_{i=1}^{n} \varphi(s). \tag{22}$$

2. *Find the moment generating function for each distribution Table 1. Also find the characteristic function and the cumulant generating function. What are the regions of definition of each? What are the moments and cumulants?*

3. *In a certain country each family is strictly monogamous and has at least two children. If neither of these is a daughter, then each family continues having further children until a daughter is born. What is the average number of children per family? What is the variance in the*

number of children per family? Is this higher than that in a Poisson distribution with the same mean? What is the ratio of males to females in each generation?

4. *In a neighbouring small country each family is strictly monogamous and 50% of the time has at least one child. If a son is born, they stop immediately, otherwise in 50% of the cases they have another child, and the process is repeated. What is the average number of children per family? What is the variance in the number of children per family? Is this higher than that in a Poisson distribution with the same mean? What is the ratio of males to females in each generation?*

2.3 Multivariate probability distributions

For n-tuplets of random numbers $(x_1, \cdots, x_n)$, the probability density function is $P(x_1, \cdots, x_n)$. Let $\langle x_i \rangle = \mu_i$, where the expectation value is taken over P. The **covariance** of two variables is defined as

$$\text{Cov }(x_i, x_j) \;=\; \langle (x_i - \mu_i)(x_j - \mu_j) \rangle. \tag{23}$$

When Cov $(x_i, x_j) = 0$ the two variables are called **independent**. The **marginal distributions** of P are defined to be

$$P_i(x) \;=\; \int \prod_{j \neq i} dx_j P(x_1, \cdots, x_n). \tag{24}$$

If all variables are independent then $P(x_1, \cdots, x_n) = \prod P_i(x)$. However, even for independent variables, it does not follow that all the P_i are equal. The **correlation coefficient** is

$$C_{ij} \;=\; \frac{\text{Cov }(x_i, x_j)}{\sqrt{\text{Var } x_i \text{Var } x_j}}, \qquad (-1 \leq C_{ij} \leq 1). \tag{25}$$

Problem set II.8

1. *What is the event set and marginal distribution of one of the variables, i_α, in the binomial distribution? And in the multinomial distribution? What is the covariance of each pair of the variables in both the distributions?*

2. *Two numbers are drawn from an uniform distribution in $0 \leq x \leq 1$. What is the event space of their difference, and its probability distribution function?*

3. *Is the notion of independence introduced in eq. (25) above is the same as that introduced in the previous lecture.*

4. *Take the joint probability distribution for $\mathbf{x} = (x, y)$ given by*

$$\begin{aligned} G(\mathbf{x}) &= \sqrt{\frac{\det\sigma}{2(2\pi)^2}} \exp\left(-\frac{1}{2}\mathbf{x}\cdot\sigma\cdot\mathbf{x}\right) \\ \text{where} \qquad \sigma &= \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix} \quad (0 < v < 1). \end{aligned} \tag{26}$$

   (a) *Show that the marginal distributions of x and y are both Gaussian, and compute their variances.*

   (b) *Compute the covariance of x and y.*

5. *Construct the joint probability distribution for $0 \leq x, y \leq \infty$, both $x, y \in \mathbf{R}$, such that each is exponentially distributed with common mean μ, and their correlation coefficient is v.*

---

3 The Central Limit Theorem

Theorem II.1 (The Central Limit Theorem)
Let r_i $(i = 1, \cdots, N)$ be independent random variables with the same probability distribution with mean μ and finite second moment, $\sigma^2 > 0$. Also let $S_N = \sum_{i=1}^{N} r_i$. Then,

$$\lim_{N\to\infty} P\left(\frac{(S_N - \mu)}{\sigma\sqrt{N}} \leq \alpha\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} dt e^{-t^2/2}. \tag{27}$$

The idea of the proof is to show that the characteristic function of $S_N/\sigma\sqrt{N}$ tends to a Gaussian.
Proof: The proof involves the following steps:

1. Substituting $r_i - \mu$ for r_i we obtain the central limit theorem for $\mu = 0$. Hence we can set $\mu = 0$.

2. If the characteristic function of r is $\varphi(s)$, then the variable $S_N/\sigma\sqrt{N}$ has the characteristic function $\varphi^N(s/\sigma\sqrt{N})$. Now, in the Taylor series expansion $\varphi(s) = 1 - s^2\sigma^2/2 + s^2 B(s)$, $B(s) \to 0$ as $s \to 0$.

3. *Lemma:* For complex numbers z_i, w_i $(i = 1, \cdots, n)$ with modulus less than or equal to 1,

$$|z_1 z_2 \cdots z_n - w_1 w_2 \cdots w_n| \leq \sum_{i=1}^{n} |z_i - w_i|. \tag{28}$$

   The proof proceeds by induction from the identity

$$z_1 z_2 \cdots z_n - w_1 w_2 \cdots w_n = (z_1 - w_1)(z_2 \cdots z_n) + w_1(z_2 \cdots z_n - w_2 \cdots w_n). \tag{29}$$

4. Using the above lemma we find

$$\left|\varphi^N\left(\frac{s}{\sigma\sqrt{N}}\right) - \left(1 - \frac{s^2}{2N}\right)^N\right| \leq N\left|\varphi\left(\frac{s}{\sigma\sqrt{N}}\right) - \left(1 - \frac{s^2}{2N}\right)\right|. \tag{30}$$

   The right hand side of this inequality is equal to $(s/\sigma)^2 B(s/\sigma\sqrt{N})$ and hence tends to zero as $N \to \infty$ at any fixed s.

5. Therefore the characteristic function for the distribution of $S_N/\sigma\sqrt{N}$ tends to $(1 - s^2/2N)^N \to \exp(-s^2/2)$; and this is the characteristic function for a Gaussian.

This concludes the proof. ∎

Problem set II.9

1. *For independent random variates $r_i \in [0, 1]$, drawn from an uniform distribution, compute numerically the distribution $\mathcal{P}_N$ of the random variate defined by*

$$x_N = \frac{1}{N}\sum_{j=1}^{N} r_j. \tag{31}$$

   (a) *Plot the distribution $\mathcal{P}_N$.*

   (b) *Show that $\mathcal{P}_N$ goes to a Gaussian as $N \to \infty$ by plotting the cumulants of $\mathcal{P}_N$ against N.*

(c) *Show that the tail of the distribution tends to a Gaussian slower than the peak.*

2. *Using the independent random variates* $r_i \in [0,1]$*, drawn from an uniform distribution, construct the correlated variates*

$$\rho_1 = r_1 \tag{32}$$
$$\rho_i = (1-\epsilon)r_i + \epsilon\rho_{i-1}, \qquad (i > 1,\ \text{and } 0 < \epsilon \leq 1). \tag{33}$$

(a) *What are the means* $\langle\rho_i\rangle$ *and the variances* $\text{Var}\ \rho_i$*? What is the correlation coefficient* C_{ij} *between* ρ_i *and* ρ_j*?*

(b) *Compute numerically the distribution* $\mathcal{P}_N$ *for the random variates*

$$x_i = \frac{1}{N}\sum_{j=1}^{N}\rho_j. \tag{34}$$

(c) *Numerically test that when* ϵ *is independent of* N*, then, no matter how close* ϵ *is to 1, the central limit theorem holds.*

(d) *Check numerically that when* $\epsilon = \exp(-c/N)$ *(for any* $c > 0$*) the central limit theorem fails.*

3. *Repeat the above problem when the distribution of* r_i *is*

$$\mathcal{P}(r) = \begin{cases} 3.75 & (0 \leq r \leq 0.2) \\ 0.3125 & (0.2 < r \leq 1) \end{cases} \tag{35}$$

Quantify the difference in speeds in the approach to a Gaussian.

Index