2

Probability

2.1 Random Variables

Flip a coin, throw some dice, measure the voltage of a noisy circuit, and the outcomes are realizations of random variables. Let X denote the random variable in question, which we assume takes on strictly real values, and let J denote the set of values X can assume. The set J may be discrete (as for dice), continuous (as for noisy voltages), or the union of the two. Repeated and independent measurements yield a sequence of possible values x_n, where $n = 1, 2, ..., N$, $N < \infty$, that sample the possible outcomes of X in an unbiased way. Approximate averages involving the x_n values can be obtained easily, such as

$$\begin{aligned}\overline{X} &= \frac{1}{N}\sum_{n=1}^{N} x_n \;, \\ \overline{X^2} &= \frac{1}{N}\sum_{n=1}^{N} x_n^2 \;,\end{aligned}$$

etc. Experience shows that the larger N becomes, the closer these values generally are to those values of the true averages for these quantities obtained ideally when $N \to \infty$.

We can use this data in yet another way, in particular to build a histogram, namely, an approximate frequency of distribution, which we may call $\sharp(n)$. For a single die, we can record the number of times each number comes up, i.e., $\sharp(n)$, $n \in \{1, 2, 3, ..., 6\}$, in a sample of N throws; see Fig. 2.1. With such information, we can compute the average $\overline{X}$ and $\overline{X^2}$ as

$$\begin{aligned}\overline{X} &= \frac{\sum_{n=1}^{6} n\,\sharp(n)}{\sum_{n=1}^{6} \sharp(n)} \;, \\ \overline{X^2} &= \frac{\sum_{n=1}^{6} n^2\,\sharp(n)}{\sum_{n=1}^{6} \sharp(n)} \;.\end{aligned}$$

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_2,
© Springer Science+Business Media, LLC 2011

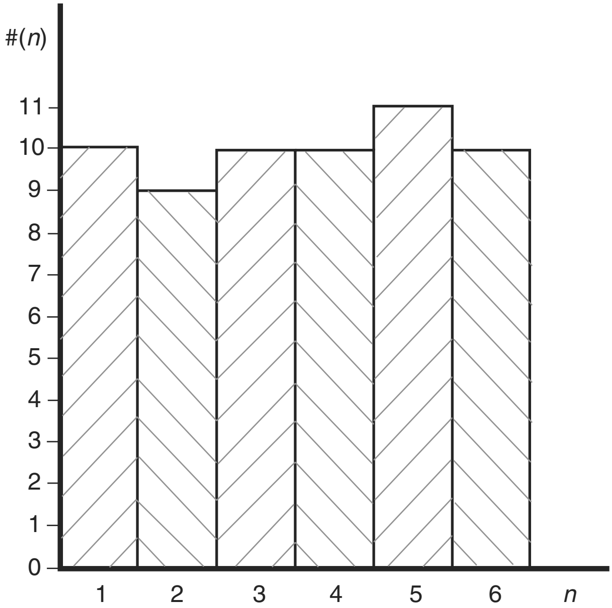

Fig. 2.1. Possible distribution of distinct die values 1-6 obtained in 60 separate and independent throws

Naturally, the larger N is, the closer the result for $\overline{X^m}$ to its idealized value is expected to be. For discrete outcomes, like a coin or a die, it is actually possible that a precise and unbiased histogram arises from a finite number of drawings. If, however, the distribution for X is continuous—at least partly—it is not possible to get an accurate histogram with any finite number of drawings.

This introductory discussion suggests that we choose another but closely related way to characterize the properties of a random variable.

2.1.1 Probability distributions

To cover all possibilities, we suppose that the values that X may assume lie between $-\infty$ and $+\infty$. We introduce [DMS]

$$P(X \leq x) \equiv \mu(x)$$

as the probability that the random variable X takes values between $-\infty$ and x (*including* x itself); the important function $\mu(x)$ is called a *probability measure.* Just from this definition, we have

$$\begin{array}{lll} \text{a)} & \mu(-\infty) = 0 \,, & \mu(\infty) = 1 \\ \text{b)} & \mu(x) \leq \mu(x+h) \,, & h > 0 \end{array}$$

c) $$\mu(x_-) \le \mu(x) = \mu(x_+) \; , \qquad x_\pm = \lim_{\epsilon \to 0^+} x \pm \epsilon \; .$$

The functions $\mu(x)$ that satisfy these conditions are called *probability measures.* It turns out that there are *three* qualitatively different types of functions that obey these conditions. Rather than experimentally determine this distribution, as we discussed above, we now require that μ be given as part of the definition of X in the first place.

The names of the three forms of probability measures μ are (i) absolutely continuous (ac), (ii) discrete (d), and (iii) singular continuous (sc). For applications to the subject of this monograph, the first two forms are by far the most important; however, we briefly discuss the third variety for completeness. For clarity, we discuss examples that are purely of the given variety. More generally, a probability distribution can be a normalized convex linear combination of all three kinds.

Absolutely continuous: In this case

$$\mu_{ac}(x) = \int_{-\infty}^{x} \rho(y)\, dy \; ,$$

where $\rho(y) \ge 0$. The function $\rho(y)$ is called the *probability density* and it is necessarily normalized so that

$$\mu_{ac}(\infty) = \int_{-\infty}^{\infty} \rho(y)\, dy = 1 \; .$$

In addition, the function $\mu_{ac}(x)$ is differentiable and $\mu_{ac}(x)' = \rho(y)$, a.e., which stands for ‘almost everywhere’, namely, up to a set of measure zero (e.g., a set consisting of finitely many points).

Discrete: The prototypical example of this kind of probability distribution is given by $\mu_d(x) = H(x)$, where $H(x)$ is the Heaviside function,

$$H(x) = 1 \; , \qquad x \ge 0 \; ,$$
$$H(x) = 0 \; , \qquad x < 0 \; .$$

It follows that

$$\mu_d(x) = \int_{-\infty}^{x} dH(x) = H(x) \; .$$

In modern terms, this measure is often loosely stated as

$$\mu_d(x) = \int_{-\infty}^{x} \delta(x)\, dx \; , \qquad \textit{informal!} \; ,$$

where $\delta(x)$ denotes the Dirac δ-function defined by the requirement that $\delta(x) = 0$ for $x \ne 0$, and $\int \delta(x)\, dx = 1$ so long as the range of integration includes $x = 0$. However, this latter expression begs the question as to what is the behavior of $\mu_d(x)$ as $x \to 0^+$, $x \to 0^-$, and especially what is the value of $\mu_d(0)$. These delicate issues are all avoided by using the Heaviside function $H(x)$.

The most general discrete probability measure is of the form

$$\mu_d(x) = \sum_{n=1}^{\infty} p_n H(x - x_n) ,$$

where the weights $p_n \geq 0$, $\Sigma_{n=1}^{\infty} p_n = 1$, and x_n, for each n, is an arbitrary jump location.

Singular continuous: It is clear that a sequence of absolutely continuous distributions, $\mu_{ac\,n}(x)$, $n = 1, 2, ...$, can converge to a discrete distribution, $\mu_d(x)$. It is less clear, but nevertheless true, that a sequence of discrete distributions, $\mu_{d\,n}(x)$, $n = 1, 2, ...$, can converge to an absolutely continuous distribution, $\mu_{ac}(x)$. A common way to compute the third kind of distribution, the singular continuous distribution, $\mu_{sc}(x)$, is as a suitable limit of a sequence of discrete distributions, $\mu_{d\,n}(x)$. This sequence will be more transparent when we introduce *characteristic functions* in the next section; here, we can only state the canonical example of a singular continuous distribution without making its connection with the other distributions clear.

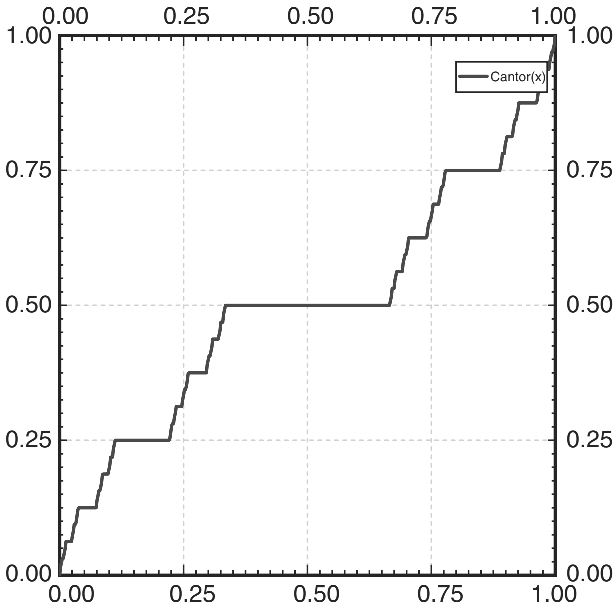

Fig. 2.2. Cantor function

The standard example we have in mind is based on the *Cantor function* $\mu_C(x) = \overline{C}(x)$ illustrated in Fig. 2.2 [Leb07]. The Cantor function is defined as follows: Let x, $0 \le x \le 1$, be given in a *tertiary* representation, i.e., $x = t_1/3 + t_2/3^2 + t_3/3^3 + \cdots$ and map such points to the function $\overline{C}(x)$, $0 \le x \le 1$, given in a *binary* representation, i.e., $\overline{C} = b_1/2 + b_2/2^2 + b_3/2^3 + \cdots$ according to the following rules: if $t_j = 0$, then $b_j = 0$; if $t_j = 2$, then $b_j = 1$; and finally if $t_j = 1$, then $b_j = 1$ *and all further* t_{j+p}*,* $p > 0$*, are mapped to* $b_{j+p} = 0$. For example, $x \equiv (t_1, t_2, t_3, \ldots)_3 = (2, 1, \ldots)_3$ is mapped to $\overline{C} \equiv (b_1, b_2, b_3, \ldots)_2 = (1, 1, \text{all } 0)_2$. Likewise, $(1, \ldots)_3 \to (1, \text{all } 0)_2$; $(0, 2, 2, 0, 1, \ldots)_3 \to (0, 1, 1, 0, 1, \text{all } 0)_2$, etc. The function $\overline{C}(x)$ has a derivative which is almost everywhere zero. Hence $\overline{C}(x) \ne \int_0^x \overline{C}'(y)\,dy = 0$! Nevertheless, $\overline{C}(x)$ is a continuous function rising from $\overline{C}(0) = 0$ to $\overline{C}(1) = 1$.

There are of course many more examples of singular continuous measures, but this one example serves to give the basic idea.

****

Note that the most general probability measure $\mu(x)$ is the combination

$$\mu(x) = A\mu_{ac}(x) + B\mu_d(x) + C\mu_{sc}(x) \ ,$$

where $A \ge 0$, $B \ge 0$, and $C \ge 0$, as well as $A + B + C = 1$.

2.2 Characteristic Functions

Let us define the function

$$C(t) = \int_{-\infty}^{\infty} e^{itx}\, d\mu(x) \ ,$$

for all real t, $-\infty < t < \infty$, as the *characteristic function* associated with each probability measure $\mu(x)$ [Luk70]. This function has several important properties,

a) $C(0) = 1$,

b) $|C(t)| \le 1$,

c) $\Sigma_{j,k=1}^{K,K} \alpha_j^* \alpha_k\, C(t_k - t_j) \ge 0$, $\quad \alpha_j \in \mathbb{C}$,

d) $C(t)$ is continuous.

The first three properties are evident, so we focus on the fourth property, d). Consider

$$|C(t) - C(s)| \le \int_{-\infty}^{\infty} |e^{itx} - e^{isx}|\, d\mu(x) = \int_{-\infty}^{-A} |e^{itx} - e^{isx}|\, d\mu(x)$$

$$+\int_{-A}^{B} |e^{itx} - e^{isx}|\, d\mu(x) + \int_{B}^{\infty} |e^{itx} - e^{isx}|\, d\mu(x)$$
$$\leq 2\left(\int_{-\infty}^{-A} + \int_{B}^{\infty}\right) d\mu(x) + \int_{-A}^{B} |e^{i(t-s)x} - 1|\, d\mu(x)\,.$$

The goal is to show that if $|C(t) - C(s)| < \epsilon$, $\epsilon > 0$, then $|t-s| < \delta(\epsilon)$, $\delta > 0$. Since $\int_{-\infty}^{\infty} d\mu(x) = 1$, it follows that we can choose A and B so large that the first term in the last line is less than $\epsilon/2$. With A and B thus fixed and finite, we can clearly choose $|t-s|$ small enough to make the second term in the last line less than $\epsilon/2$. In particular, since

$$|e^{i(t-s)x} - 1| = |\textstyle\int_0^{(t-s)x} e^{iy}\, dy| \leq |t-s||x| \leq |t-s|(|A| + |B|)\,.$$

Thus it suffices to choose $\delta = \epsilon/2(|A| + |B|)$. This concludes the proof that $C(t)$ is continuous.

We note in passing that if $C(-t) = C(t)$, then it follows that $C(t)$ is real.

Bochner's Theorem: It is especially noteworthy that the four properties satisfied by $C(t)$, which are listed above, *automatically* ensure that

$$C(t) = \int_{-\infty}^{\infty} e^{itx}\, d\mu(x)\,,$$

where $\mu(x)$ is a probability measure [Luk70].

2.2.1 Convergence properties - 1

The Fourier transform of a measure leads to its characteristic function,

$$C(s) \equiv \int e^{isx}\, d\mu(x)\,.$$

Whatever the type of measure involved, the characteristic function is always continuous. However, the three types of measures can be distinguished in the following way. Besides the characteristic function $C(s)$, let us also introduce

$$C_T(s) \equiv T^{-1}\int_0^T C(s+t)\, dt\,, \qquad T > 0\,,$$

which amounts to a partial averaging of the function $C(s)$. Then it follows [Luk70] that

(ac) For purely absolutely continuous measures,

$$\lim_{s\to\infty} C(s) = 0\,, \quad \text{and} \quad \lim_{s\to\infty} C_T(s) = 0\,.$$

(d) For purely discrete measures,

$$\lim_{s\to\infty} C(s) \neq 0\,, \quad \text{and} \quad \lim_{s\to\infty} C_T(s) \neq 0\,.$$

(sc) For purely singular continuous measures,

$$\lim_{s\to\infty} C(s) \neq 0\,, \quad \text{while} \quad \lim_{s\to\infty} C_T(s) = 0\,.$$

2.2.2 Convergence properties - 2

It is often extremely convenient to study probability distributions through a study of their corresponding characteristic functions. One aspect deals with the convergence of sequences. Suppose we deal with the sequence

$$C_n(t) = \int_{-\infty}^{\infty} e^{itx}\, d\mu_n(x) \;, \qquad\qquad n = 1, 2, 3, \ldots \;.$$

If

$$\lim_{n\to\infty} C_n(t) = C(t) \;,$$

and the limiting function $C(t)$ obeys all the required four properties—the first three are trivial, so continuity is the only real issue—then $C(t)$ is the characteristic function of *some* probability measure $\mu(x)$. In this case, one says that the sequence $\{\mu_n\}$ of measures converges "weakly" to the measure μ, a property that is denoted by the equation

$$\mathrm{w-}\lim_{n\to\infty} \mu_n(x) = \mu(x) \;.$$

For example, if

$$C_n(t) = e^{-t^2/4n} \;,$$

then

$$\lim_{n\to\infty} C_n(t) = 1 = C(t) \;.$$

This fact shows that the sequence

$$d\mu_n(x) = \sqrt{\pi n}\, e^{-nx^2}\, dx \;, \qquad n = 1, 2, 3, \ldots \;,$$

converges weakly as $n \to \infty$ to the Heaviside measure $\mu(x) = H(x)$, providing an example of a sequence of absolutely continuous measures that converges weakly to a discrete measure.

As a counterexample to weak convergence, consider the sequence

$$C_n(t) = e^{-nt^2}$$

for which

$$\lim_{n\to\infty} C_n(t) = D(t) \;,$$

where $D(0) = 1$, while $D(t) = 0$ if $t \neq 0$, which is *not* a characteristic function since it is not a continuous function.

2.2.3 Characteristic function for a Cantor-like measure

Consider the characteristic function given by

$$C_1(t) = \cos(\frac{t}{3}) \ ,$$

which consists of δ-function distributions located at $x = \pm 1/3$ each with a weight $1/2$. Likewise,

$$C_2(t) = \cos(\frac{t}{3}) \ \cos(\frac{t}{3^2}) = \frac{1}{2}\Big[\cos(\frac{t}{3}+\frac{t}{9}) + \cos(\frac{t}{3}-\frac{t}{9})\Big]$$

has four δ-functions at $x = \pm 1/3 \pm 1/9$ (independent signs), each with a weight $1/4$. A similar kind of discussion applies for

$$C_N(t) = \prod_{n=1}^{N} \cos(t/3^n) \ ,$$

namely, a sequence of δ-functions each with a weight $1/2^N$. Finally, we take the limit

$$C(t) = \lim_{N\to\infty} C_N(t) = \prod_{n=1}^{\infty} \cos(t/3^n) \ ;$$

here the result is infinitely many δ-functions, each with an equal weight: *zero!* This is the characteristic function appropriate to a Cantor-like function, i.e.,

$$\prod_{n=1}^{\infty} \cos(t/3^n) = \int e^{itx} \, d\mu_{Cantor-like}(x) = \int e^{itx} \, d\mu_{sc}(x) \ .$$

A word about infinite products is in order. One says that

$$\prod_{n=1}^{\infty} A_n$$

converges provided that

$$\sum_{n=1}^{\infty} |1 - A_n| < \infty \ .$$

If every $A_n \neq 0$ and yet $\Pi_{n=1}^{\infty} A_n = 0$, one says that the product "diverges to zero." In the case at hand, we want strict convergence, and

$$\sum_{n=1}^{\infty} |1 - \cos(t/3^n)| = \sum_{n=1}^{\infty} | \int_0^{t/3^n} \sin(u) \, du \, |$$
$$\leq \sum_{n=1}^{\infty} | \int_0^{t/3^n} du \, | \leq \sum_{n=1}^{\infty} \frac{|t|}{3^n} < \infty \ .$$

An interesting related example refers to

$$C_N(t) = \prod_{n=1}^{N} \cos(t/2^n) \ ,$$

which is also a set of δ-function distributions, all with equal weight 2^{-N}. However, in this case,

$$C(t) = \lim_{N\to\infty} C_N(t) = \prod_{n=1}^{\infty} \cos(t/2^n)$$
$$= \frac{\sin(t)}{t} = \frac{1}{2}\int_{-1}^{1} e^{itx}\,dx = \int e^{itx}\,d\mu_{ac}(x)\ .$$

In other words, the present case leads to an absolutely continuous distribution, while the former case led to a singular continuous distribution.

2.2.4 An application of the characteristic function

As an illustration of such a discussion, let us discuss a *general Gaussian random variable* X. To say that a random variable is Gaussian means that all moments are uniquely determined by the *mean*

$$\langle X\rangle = \int x\,d\mu(x)\ ,$$

and the *variance* $\langle X^2\rangle^c \equiv \langle X^2\rangle - \langle X\rangle^2$, where

$$\langle X^2\rangle = \int x^2\,d\mu(x)\ .$$

Explicitly, the characteristic function has the particular form given by

$$\langle e^{itX}\rangle = \int e^{itx}\,d\mu(x) = e^{it\langle X\rangle - \frac{1}{2}t^2\langle X^2\rangle^c}\ .$$

An expansion in powers of t determines the form of the moments in terms of the mean and the variance. In particular,

$$\langle X\rangle^c = \langle X\rangle\ , \qquad \langle X^2\rangle^c = \langle X^2\rangle - \langle X\rangle^2\ ,$$

etc.

A special case arises if the mean is zero, $\langle X\rangle = 0$, for which the variance is equal to the second moment, $\langle X^2\rangle^c = \langle X^2\rangle$. In that case,

$$\langle e^{itX}\rangle = e^{-\frac{1}{2}t^2\langle X^2\rangle}\ .$$

To distinguish this case, we shall sometimes refer to it as a *normal distribution*, although it is commonly the case in the literature that normal and Gaussian distributions generally refer to the same thing.

Let us next derive a useful property of normal variables X.

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2}\,\langle e^{itX}\rangle\,dt &= \langle e^{-\frac{1}{2}X^2}\rangle \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2}\,e^{-\frac{1}{2}t^2\langle X^2\rangle}\,dt \\ &= \frac{1}{\sqrt{1+\langle X^2\rangle}}\;.\end{aligned}$$

Now, suppose that $\langle X^2\rangle = \infty$. Then it follows that

$$\langle e^{-\frac{1}{2}X^2}\rangle = 0\;.$$

In turn, this equation implies that $X^2 = \infty$ for almost all X values. Instead if $\langle X^2\rangle < \infty$, then it follows—just from this fact—that $X^2 < \infty$ for almost all X values. In summary, if X is a normal variable, then

$$\begin{aligned}\langle X^2\rangle < \infty &\Longleftrightarrow \quad X^2 < \infty\;,\text{a.e.}\;,\\ \langle X^2\rangle = \infty &\Longleftrightarrow \quad X^2 = \infty\;,\text{a.e.}\;.\end{aligned}$$

This property will find good use later on.

2.3 Infinitely Divisible Distributions

Let us examine the product of two characteristic functions

$$C_1(t)C_2(t)\;.$$

This function clearly satisfies all the axioms to be a new characteristic function. In particular, the third axiom follows when we observe that

$$C_1(t)C_2(t) = \langle e^{itX_1}\rangle\langle e^{itX_2}\rangle = \langle e^{it(X_1+X_2)}\rangle\;,$$

where the final average is over the so-called *product measure*

$$d\mu_1(x_1)\,d\mu_2(x_2) = \rho_1(x_1)\,\rho_2(x_2)\,dx_1\,dx_2\;,$$

assumed absolutely continuous for clarity. The new random variable $X = X_1 + X_2$ is the sum of two *independent* random variables, and it has a distribution determined by the convolution integral:

$$\begin{aligned}\rho(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x-x_1-x_2)\,\rho_1(x_1)\,\rho_2(x_2)\,dx_1\,dx_2\\ &= \int_{-\infty}^{\infty} \rho_1(x-y)\,\rho_2(y)\,dy = \int_{-\infty}^{\infty} \rho_2(x-y)\,\rho_1(y)\,dy\;.\end{aligned}$$

As a special case, the two characteristic functions, C_1 and C_2, could of course be the same. More generally, one could take any integral power of a characteristic function to make a new one; that is, if $C(t)$ is a characteristic function, then so too is

$$C_{(p)}(t) \equiv C^p(t) \equiv [C(t)]^p ,$$

for every positive integer p. An immediate example of such a discussion involves a Gaussian distribution for which

$$C_{(p)}(t) = [\, e^{it\langle X\rangle - t^2\langle X^2\rangle^c/2}\,]^p = e^{ipt\langle X\rangle - pt^2\langle X^2\rangle^c/2}$$

yields a new Gaussian distribution for which

$$\begin{aligned} \langle X\rangle_{new} &= p\langle X\rangle_{old} , \\ \langle X^2\rangle_{new} &= p\langle X^2\rangle_{old} + (p^2 - p)\langle X\rangle^2_{old} . \end{aligned}$$

2.3.1 Divisibility

Interestingly enough, the pth root—instead of the pth power—applied to a Gaussian characteristic function yields a valid (Gaussian) characteristic function again [Luk70], in particular,

$$C_{(1/p)}(t) = [\, e^{it\langle X\rangle - t^2\langle X^2\rangle/2}\,]^{1/p} = e^{it\langle X\rangle/p - t^2\langle X^2\rangle/2p}$$

also defines a (Gaussian) characteristic function for all $p \in \{1, 2, 3, \ldots\}$. Such a distribution is said to be *infinitely divisible*. As far as random variables go, it means that the original Gaussian random variable may be realized as

$$X = \sum_{k=1}^{p} Y_k ,$$

where $Y_1, Y_2, \ldots, Y_p$ denote p *independent, identically distributed* (here Gaussian) random variables for arbitrary positive integer p.

Not all random variables are infinitely divisible—in fact, not all random variables are divisible at all. If a random variable is divisible by two, then

$$X = Y_1 + Y_2 ,$$

where Y_1 and Y_2 are two independent and identically distributed random variables each with the same characteristic function, say, $D(t)$. To illustrate a random variable X that is *not* divisible, let X have just two equally weighted values, say, at $x = \pm 1$. Then it follows that the relevant characteristic function $C(t) = \cos(t)$. However, if $X = Y_1 + Y_2$, as suggested, then $C(t) = \cos(t) = D(t)^2$. Since $D(t)$, like $C(t)$, is even they both must be real. However, there is no real $D(t) = \sqrt{\cos(t)}$ for all t.

2.3.2 Infinite divisibility

Let us determine the characteristic function for random variables X that are infinitely divisible. We will see that we can find a fairly explicit formula for these expressions. On the other hand, relatively little is known about the forms of the probability distributions $\mu(x)$ that belong to the class of infinitely divisible distributions [Luk70].

Denote the characteristic function of interest, as usual, by

$$C(t) = \int e^{itx} \, d\mu(x) \, .$$

By definition, $C^{1/q}(t)$ is also a characteristic function for all $q \in \{1, 2, 3, \ldots\}$, which is the definition of infinite divisibility. Integral powers always lead to characteristic functions so $C^{p/q}(t)$ is also a characteristic function. Let the rational ratio p/q tend as a sequence to a nonnegative real number r, and by continuity we learn that $C^r(t)$ is still a characteristic function for all real $r \geq 0$. As such $C^r(t)$ must be nonzero since, for any $P > 1$, it follows that $[C^{r/P}(t)]^P = C^r(t)$ and $C^{r/P}(t) \to 1$ as $P \to \infty$. But if $C^r(t) \neq 0$ for any t, then so too is $C(t)$ itself. Consequently, we learn that $C(t)$ *cannot vanish.* If that is the case, it follows that

$$C(t) = e^{-L(t)} = \int e^{itx} \, d\mu(x) \equiv \langle e^{itX} \rangle \, ,$$

where L is continuous, $L(0) = 0$, and $\Re \, L(t) \geq 0$.

Next we consider

$$C^r(t) = e^{-r\,L(t)} = \int e^{itx} \, d\mu_r(x) \, ,$$

or stated otherwise,

$$r^{-1}[1 - e^{-r\,L(t)}] = \int [1 - e^{itx}] \, d(r^{-1}\mu_r(x)) \, .$$

By assumption, the limit $r \to 0$ exists and leads to

$$\begin{aligned}
L(t) &= \lim_{r \to 0} r^{-1}[1 - e^{-r\,L(t)}] \\
&= \lim_{r \to 0} \int [1 - e^{itx}] \, d(r^{-1}\mu_r(x)) \\
&= \lim_{r \to 0} \int [1 - e^{itx} + itx/(1+x^2)] \, d(r^{-1}\mu_r(x)) \\
&\qquad - it \lim_{r \to 0} \int [x/(1+x^2)] \, d(r^{-1}\mu_r(x)) \\
&= -ita + \lim_{r \to 0} \int_{|x| \leq r} [1 - e^{itx} + itx/(1+x^2)] \, d(r^{-1}\mu_r(x)) \\
&\qquad + \lim_{r \to 0} \int_{|x| > r} [1 - e^{itx} + itx/(1+x^2)] \, d(r^{-1}\mu_r(x)) \, .
\end{aligned}$$

Finally,

$$L(t) = -ita + bt^2 + \int_{|x|>0} [1 - e^{itx} + itx/(1+x^2)]\, d\sigma(x) \ ,$$

where $a \in \mathbb{R}$, $b \geq 0$, and $\sigma(x)$ is a nonnegative measure that satisfies

$$\int_{|x|>0} [x^2/(1+x^2)]\, d\sigma(x) < \infty \ ,$$

although it is quite possible that $\int_{|x|>0} d\sigma(x) = \infty$.

The final expression for $L(t)$ is the most general version that leads to a characteristic function for an infinitely divisible distribution. It follows that the general infinitely divisible random variable X is composed of two parts:

$$X \equiv X_G + X_P \ .$$

The linear and quadratic factors in t correspond to a Gaussian random variable (X_G); the remaining term corresponds to a Poisson random variable (X_P). More specifically, (i) if $\sigma(x)$ has a one-point support, e.g., $d\sigma(x) = k\,\delta(x-c)\,dx$, then X_P is a *Poisson variable*; (ii) if $\int d\sigma(x) < \infty$ with wider support, then X_P is a *compound Poisson variable*; and (iii) if $\int d\sigma(x) = \infty$, then X_P is a *generalized Poisson variable.*

For many problems, there is a symmetry such that $C(-t) = C(t)$, and thus $L(-t) = L(t)$. Under such circumstances

$$L(t) = bt^2 + \int [1 - \cos(tx)]\, d\sigma(x) \ .$$

If we assume that σ is absolutely continuous, then in this case $d\sigma = U(x)\,dx$, where $U(-x) = U(x) \geq 0$.

2.4 Central Limit Theorem—and Its Avoidance

Consider the random variable

$$X = \sum_{p=1}^{N} Y_p$$

made up of $N < \infty$ independent, identically distributed random variables, Y_p. Let Y denote any one of the identical random variables Y_p, $1 \leq p \leq N$, and for simplicity assume the variable Y is purely absolutely continuous, that all moments $\langle Y^n \rangle$, $n \in \mathbb{N} \equiv \{1, 2, 3, \ldots\}$, exist, and that all *odd* moments vanish:

$$\langle Y^{2l+1} \rangle = 0 \ , \qquad l = 0, 1, 2, \ldots \ .$$

This leads to what is called a symmetric distribution. Moreover, we also assume that the probability distribution for Y, $\mu(y)$, depends on the total number of variables N; thus $\mu(y) = \mu_N(y)$. Likewise the averaging process $\langle(\cdot)\rangle$ over Y depends on N, and making that explicit leads to $\langle(\cdot)\rangle = \langle(\cdot)\rangle_N$.

The characteristic function of X reads

$$\begin{aligned}\langle e^{itX}\rangle_N &= \langle e^{it\Sigma_{p=1}^N Y_p}\rangle_N = [\langle e^{itY}\rangle_N]^N \\ &= [1 - t^2\langle Y^2\rangle_N/2! + t^4\langle Y^4\rangle_N/4! - \cdots]^N .\end{aligned}$$

We wish to study the limit of this expression as $N \to \infty$. To obtain a meaningful answer, it is necessary, for large N, that

$$\langle Y^2\rangle_N \propto 1/N .$$

This result can be obtained in *two fundamentally different ways*.

Gaussian behavior: In the first way the probability density has the general shape illustrated in Fig. 2.3.

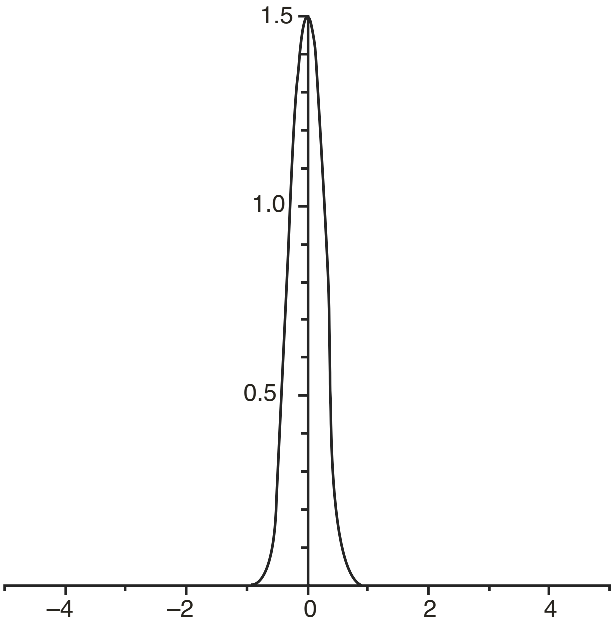

Fig. 2.3. Gaussian-like behavior

Here the principal property is that the width of the distribution should be "narrow," i.e., $\propto 1/\sqrt{N}$. Since there is unit area under the whole curve, it follows that the height is "tall," i.e., $\propto \sqrt{N}$. This shape implies that $\langle Y^2\rangle \propto 1/N$, but it also inevitably implies that $\langle Y^4\rangle_N \propto 1/N^2$, $\langle Y^6\rangle_N \propto 1/N^3$, etc.

As a consequence, the only contribution to $\langle e^{itX}\rangle_N$ that matters is the second moment, $\langle Y^2\rangle_N$, and therefore

$$\langle e^{itX}\rangle \equiv \lim_{N\to\infty} \langle e^{itX}\rangle_N = e^{-t^2 A/2} \; ,$$

where

$$A \equiv \lim_{N\to\infty} N\langle Y^2\rangle_N \; ,$$

which we assume exists. The resultant distribution, therefore, is a *normal* distribution; *all* features of the original distribution collapse into the single constant A! This behavior illustrates well the central property of the Central Limit Theorem.

Poisson behavior: The second way to achieve an acceptable behavior leads to a *non*-Gaussian result. In this second way the distribution $\rho_N(x) = \mu'_N(x)$ has the general shape illustrated in Fig. 2.4.

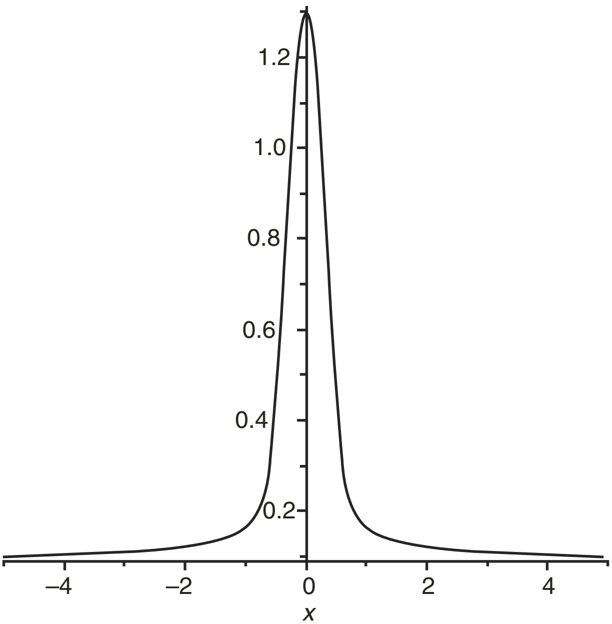

Fig. 2.4. Poisson-like behavior

The essence of this figure is the following: The central "spike" has a width $\propto 1/\sqrt{N}$, as before (or even narrower!), and this spike has an area of $1-O(1/N)$. The "broad shoulder" reaches out to Y values of order 1, i.e., independent of N. However, there is only an area $O(1/N)$ under this broad shoulder. In the present case, it follows that $\langle Y^2\rangle_N \propto 1/N$, certainly from the shoulder and part possibly from the spike. However, it also follows that $\langle Y^4\rangle_N \propto 1/N$, $\langle Y^6\rangle_N \propto 1/N$, etc., based entirely on the shoulder. Thus *all moments are uniformly small* $\propto 1/N$—save for $\langle 1\rangle_N = 1$. As a consequence,

$$\begin{aligned}\langle e^{itX}\rangle &= \lim_{N\to\infty}[\,1 - t^2\langle Y^2\rangle_N/2! + t^4\langle Y^4\rangle_N/4! - \cdots]^N \\ &= \exp[\,-t^2\langle Y^2\rangle'/2! + t^4\langle Y^4\rangle'/4! - \cdots]\,,\end{aligned}$$

where, for all $m \geq 1$,

$$\langle Y^{2m}\rangle' \equiv \lim_{N\to\infty} N\,\langle Y^{2m}\rangle_N\,.$$

It is noteworthy that this expression may be written in the form

$$\langle e^{itX}\rangle = e^{\langle e^{itY} - 1\rangle'} \equiv e^{-\int[1-\cos(ty)]\,d\sigma(y)}\,,$$

which we recognize as the very same (compound, generalized) Poisson distribution that arose in the study of infinite divisibility. Coupled with the Gaussian answer for the Central Limit Theorem, the result we have obtained is the full range of possibilities found in the study of (even) infinitely divisible distributions. In fact, this should not be a surprise since in both cases we were seeking those random variables that admit the decomposition

$$X = \sum_{p=1}^{\infty} Y_p$$

into infinitely many independent, identically distributed random variables Y_p.

Exercises

2-1 Prove that

$$\prod_{n=1}^{\infty} \cos(t/2^n) = \frac{\sin(t)}{t}\,.$$

2-2 Expand the expression

$$\langle e^{itX}\rangle = \int e^{itx}\,d\mu(x) = e^{it\langle X\rangle - \frac{1}{2}t^2\langle X^2\rangle^c}$$

in powers of t to determine expressions for $\langle X^3\rangle$ and $\langle X^4\rangle$ in terms of the mean $\langle X\rangle$ and the variance $\langle X^2\rangle^c$.

2-3 When the mean vanishes, the Gaussian distribution becomes

$$\langle e^{itX} \rangle = \int e^{itx}\, d\mu(x) = e^{-\frac{1}{2}t^2 \langle X^2 \rangle} \ ,$$

and we have called X a normal variable. Find the general expression for $\langle X^p \rangle$ in terms of the variance, i.e., the second moment, $\langle X^2 \rangle$.

2-4 Find the characteristic function $C_{\overline{C}}\,(s)$ for the singularly continuous measure defined by the Cantor function. Show that

$$\lim_{s \to \infty} C_{\overline{C}}\,(s) \neq 0 \ ,$$

and show that the time-averaged expression

$$C_{\overline{C}\,T}\,(s) = \frac{1}{T} \int_0^T C_{\overline{C}}\,(s+t)\, dt \ , \qquad\qquad T > 0 \ ,$$

satisfies

$$\lim_{s \to \infty} C_{\overline{C}\,T}\,(s) = 0 \ .$$

2-5 Show that the function $F(t) = \exp(-Bt^4)$, with $B > 0$, cannot be the characteristic function of a probability distribution.

2-6 Show that the function $G(t) = \exp(-At^2 - Bt^4)$, with both $A > 0$ and $B > 0$, cannot be the characteristic function of a probability distribution.

2-7 For the characteristic function given, for $1 < \beta < 3$, by

$$C(t) = \exp\{-\textstyle\int [1 - \cos(tu)]\, du/|u|^{\beta}\} \equiv \displaystyle\int e^{itx}\, \rho_\beta(x)\, dx \ ,$$

determine the asymptotic functional form of the associated probability density function $\rho_\beta(x)$ as $x \to \infty$.

2-8 Let $\{C_n(t)\}$ denote a sequence of characteristic functions and $\{p_n\}$ a sequence of positive numbers. Build the functions

$$D_n(t) \equiv \exp\{p_n[C_n(t) - 1]\} \ .$$

Show that each of the functions $D_n(t)$ corresponds to a characteristic function of an *infinitely divisible* distribution. Show that *any* infinitely divisible distribution for a single random variable can be obtained as the limit of suitable sequences $\{C_n(t)\}$ and $\{p_n\}$. (This latter fact is known as De Finetti's Theorem [DMS].)

Table 2.1. Fertilizer amount (FA) vs. Growth rate (GR)

Sample	FA	GR
1	0.100	1.724
2	0.125	1.812
3	0.150	1.810
4	0.175	2.045
5	0.200	2.187
6	0.225	2.123
7	0.250	2.056
8	0.275	1.823
9	0.300	1.945
10	0.325	1.654

2-9 A farmer measures the growth rate (GR) in 10 different portions of her field as determined by 10 different amounts of fertilizer (FA). The results of the experiment are given in Table 2.1.

Find the average of GR for the experiment. Find the variance of GR, and the standard deviation (= square root of the variance) for the experiment.

If $1 \leq n \leq 10$ denotes the sample number, find the least squares fit of a linear fit to the data for GR given by

$$GR(n) = a + bFA(n) \; ,$$

for suitable constants a and b. Finally, find the least squares fit to a quadratic fit to the data for GR given by

$$GR(n) = c + dFA(n) + f\,[FA(n)]^2 \; ,$$

for suitable constants c, d, and f. On the basis of this information, make an educated guess for the optimal fertilizer amount to achieve the highest growth rate.

Springer
http://www.springer.com/978-0-8176-4790-2

A Modern Approach to Functional Integration
Klauder, J.R.
2011, XVI, 282 p. 9 illus., Hardcover
ISBN: 978-0-8176-4790-2
A product of Birkhäuser Basel