Appendix B

Elements on random variables

This Appendix summarizes a few elements of probability theory, with a focus on random variables, adopting the point of view of a physicist interested in results more than in formal proofs.[(76)]

B.1 Definition

The notion of a *random variable*—or *stochastic variable*—X relies on two elements:

a) The set Ω—referred to as *sample space*, *universe* or *range*—of the possible values x (the *realizations*) describing the outcome of a "random experiment".

This set can be either discrete or continuous, or even partly discrete and partly continuous. Besides, the sample space can be multidimensional. Accordingly, one speaks of discrete, continuous or multidimensional random variables. The latter will in the following often be represented as vectors $\mathbf{X}$.

A physical instance of discrete resp. continuous one-dimensional random variable is the projection of the spin of a particle on a given axis, resp. the kinetic energy of a free particle. Examples of continuous 3-dimensional stochastic variables are the three components of the velocity $\vec{v}$ or those of the position $\vec{x}$ of a Brownian particle at a given instant.

b) The *probability distribution* on this set.

Consider first a continuous one-dimensional random variable defined on a real interval (or on a union of intervals) $\mathcal{I}$. The probability distribution is specified through a *probability density*, that is a nonnegative function $p_X(x)$

$$p_X(x) \geq 0 \quad \forall x \in \mathcal{I} \tag{B.1a}$$

normalized to 1 over its range of definition

$$\int_{\mathcal{I}} p_X(x)\, \mathrm{d}x = 1. \tag{B.1b}$$

$p_X(x)\, \mathrm{d}x$ represents the probability that X takes a value between x and $x + \mathrm{d}x$.

To account for the possible presence of discrete subsets in the sample space, the probability distribution may involve Dirac distributions:

$$p_X(x) = \sum_n p_n \delta(x - x_n) + \tilde{p}_X(x), \tag{B.2a}$$

with the normalization condition

$$\sum_n p_n + \int \tilde{p}_X(x)\, \mathrm{d}x = 1, \tag{B.2b}$$

[(76)] The presentation is strongly inspired by Chapter I of van Kampen's[(bo)] classic book [53].

[(bo)] N. G. van Kampen, 1921–2013

with $p_n > 0$ and $\tilde{p}_X$ a nonnegative function. If $\tilde{p}_X = 0$ identically over the range, then X is simply a discrete random variable. The corresponding probability density is then replaced by a *probability mass function*, which associates finite positive probabilities p_n to the respective realizations x_n.

The generalization to the case of a multidimensional stochastic variable is straightforward and involves D-dimensional integrals. The corresponding D-dimensional infinitesimal volume around a point $\mathbf{x}$ will hereafter be denoted by $\mathrm{d}^D\mathbf{x}$.

Remark: Physical quantities often possess a dimension, like length, mass, time... As a consequence, the probability density p_G for the distribution of the values g of such a quantity G must also have a dimension, namely the inverse of that of G, to ensure that the probability $p_G(g)\,\mathrm{d}g$ be dimensionless. This property can easily be checked on the various probability densities introduced in Sec. B.3.

In formal probability theory, one distinguishes between the sample space Ω—the set of all possible "outcomes" of a random experiment—and a set $\mathcal{F}$, which is a subset of the power set (set of all subsets) of Ω. The elements of $\mathcal{F}$, called "events", represent the events(!) that can be observed. Eventually, one introduces a function, the "probability measure", $\mathcal{P}$ from $\mathcal{F}$ in the real interval $[0,1]$, which associates to each event $A \in \mathcal{F}$ a probability $\mathcal{P}(A)$ fulfilling the conditions

- $\mathcal{P}(\Omega) = 1$ [normalization, cf. Eq. (B.1b) or (B.2b)],
- $\forall A, B \in \mathcal{F}$, $\mathcal{P}(A\cup B) = \mathcal{P}(A)+\mathcal{P}(B)$ if $\mathcal{P}(A\cap B) = 0$—in particular when $A\cap B = \emptyset$—and otherwise $\mathcal{P}(A \cup B) < \mathcal{P}(A) + \mathcal{P}(B)$.

The triplet $(\Omega, \mathcal{F}, \mathcal{P})$ is called "probability space".

Consider then such a probability space. A one-dimensional random variable X is a function from Ω to $\mathbb{R}$ with the property

$$\forall x \in \mathbb{R}, \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F},$$

i.e. the set of all outcomes ω, whose realization $X(\omega)$ is smaller than x, is an event.

A *cumulative distribution function* F from $\mathbb{R}$ to $[0,1]$ is then associated to this random variable, which maps the real number x onto the probability $\mathcal{P}(X \leq x) \equiv \mathcal{P}(\{\omega \in \Omega \mid X(\omega) \leq x\})$. One then has

$$F(x) \equiv \mathcal{P}(X \leq x) = \int_{-\infty}^{x^+} p_X(x')\,\mathrm{d}x',$$

with $p_X(x)$ the probability density. (The notation x^+ means that when p_X contains a contribution $\delta(x)$, then the latter is also taken into account in the integral.)

B.2 Averages and moments

Besides the sample space Ω and the probability density p_X, other notions may be employed for the characterization of a random variable X. In this Section we restrict the discussion to one-dimensional stochastic variables—multidimensional ones will be addressed in Sec. B.4.

Consider a function f defined on the one-dimensional sample space Ω of a random variable X. The *expectation value* or *average value* of f is defined by

$$\langle f(X)\rangle \equiv \int_\Omega f(x)\,p_X(x)\,\mathrm{d}x. \tag{B.3}$$

The generalization of this definition to the case of multidimensional random variables is straightforward.

Remarks:

$*$ This expectation value is also denoted by $E\big(f(X)\big)$, in particular by mathematicians.

$*$ Averaging a function is a linear operation.

The m-th *moment*—or "moment of order m"—of a one-dimensional random variable X (or equivalently of its probability distribution) is defined as the average value

$$\mu_m \equiv \langle X^m \rangle . \tag{B.4}$$

In particular, μ_1 is the expectation value of the random variable. In analogy with the arithmetic mean, μ_1 is often referred to as "mean value".

In addition, the *variance* of the probability distribution is defined by

$$\sigma^2 \equiv \big\langle (X - \langle X \rangle)^2 \big\rangle = \mu_2 - \mu_1^2. \tag{B.5}$$

The positive square root σ is called *standard deviation*. The latter is often loosely referred to as "fluctuation", because σ constitutes a typical measure for the dispersion of the realizations of a random variable about its expectation value, i.e. for the scale of the fluctuations of the quantity described by the random variable.

Remarks:

$*$ The integral defining the m-th moment of a probability distribution might possibly diverge! See for instance the Cauchy–Lorentz distribution in § B.3.8 below.

$*$ In analogy to the variance (B.5), one also defines the m-th *central moment* (or *m-th moment about the mean*) $\big\langle (X - \langle X \rangle)^m \big\rangle$ for arbitrary m.

$*$ If the random variable possesses a physical dimension, then its moments are also dimensioned quantities.

Another useful notion is that of the *characteristic function*, defined for $k \in \mathbb{R}$ by

$$G_X(k) \equiv \big\langle \mathrm{e}^{\mathrm{i}kX} \big\rangle = \int \mathrm{e}^{\mathrm{i}kx} p_X(x)\, \mathrm{d}x. \tag{B.6a}$$

When all moments (B.4) of the probability distribution p_X exist, one easily checks that the Taylor expansion of $G_X(x)$ about $k = 0$ reads

$$G_X(k) = \sum_{m=0}^{\infty} \frac{(\mathrm{i}k)^m}{m!} \mu_m, \tag{B.6b}$$

i.e. the m-th derivative of the characteristic function at the point $k = 0$ is related to the m-th moment of the probability distribution.

Remarks:

$*$ More precisely, the *moment-generating function* is $\mathcal{G}_X(k) \equiv G_X(-\mathrm{i}k)$, whose successive derivatives at $k = 0$, when they exist, are exactly equal to the moments μ_m. This moment-generating function may be non-analytic at the origin—e.g. in the case of the Cauchy–Lorentz distribution—, so that Eq. (B.6b) makes no sense, while Eq. (B.6a) is always defined.

$*$ The logarithm of G_X (or $\mathcal{G}_X$) generates the successive *cumulants* κ_m of the probability distribution:

$$\ln G_X(k) = \sum_{m=1}^{\infty} \frac{(\mathrm{i}k)^m}{m!} \kappa_m, \tag{B.7}$$

which are sometimes more useful than the moments (see Sec. B.4). Again, $\ln G_X$ may not be analytic at $k = 0$, in which case the right-hand side of this equation is not well-defined. One easily checks for instance $\kappa_1 = \mu_1 = \langle X \rangle$, $\kappa_2 = \sigma^2$, $\kappa_3 = \big\langle (X - \langle X \rangle)^3 \big\rangle$.

B.3 Some usual probability distributions

In this Section, we list a few often encountered probability distributions together with some of their properties, starting with discrete ones, before going on with continuous densities.

B.3.1 Discrete uniform distribution

Consider a random variable X with the finite discrete sample space $\Omega = \{x_1, \ldots, x_N\}$ where $N \in \mathbb{N}^*$. The *discrete uniform distribution*

$$p_n = \frac{1}{N} \quad \forall n \in \{1, 2, \ldots, N\} \tag{B.8}$$

corresponds to the case where all realizations of the random variable are equally probable.

The expectation value is then $\langle X \rangle = \frac{1}{N}\sum_{n=1}^{N} x_n$ and the variance $\sigma^2 = \frac{1}{N^2}\sum_{n=1}^{N} x_n^2 - \frac{1}{N}\left(\sum_{n=1}^{N} x_n\right)^2$.

B.3.2 Binomial distribution

Let p be a real number, $0 < p < 1$,[(77)] and $N \in \mathbb{N}^*$ a positive integer.

The *binomial distribution* with parameters N and p is the probability distribution for a random variable with sample space $\Omega = \{0, 1, \ldots, n, \ldots N\}$ given by

$$p_n = \binom{N}{n} p^n (1-p)^{N-n}. \tag{B.9}$$

p_n is the probability that, when a random experiment with two possible outcomes ("success" and "failure", with respective probabilities p and $1-p$) is repeated N times, one obtains exactly n "successes".

The expectation value is $\langle X \rangle = pN$ and the variance $\sigma^2 = Np(1-p)$.

B.3.3 Negative binomial distribution

Let p be a real number, $0 < p < 1$,[(77)] and $r \in \mathbb{N}^*$ a positive integer.[(78)]

The *negative binomial distribution* with parameters r and p is the probability distribution for a random variable with sample space $\Omega = \mathbb{N}$, the space of nonnegative integers, given by

$$p_n = \binom{n+r-1}{n} p^r (1-p)^n. \tag{B.10}$$

p_n represents the probability that, in a random experiment with two possible outcomes ("success" and "failure", with respective probabilities p and $1-p$), one obtains n "failures" before attaining exactly r "successes".

The expectation value is $\langle X \rangle = \frac{r(1-p)}{p}$ and the variance $\sigma^2 = \frac{r(1-p)}{p^2}$.

B.3.4 Poisson distribution

Let λ be a positive real number. The *Poisson distribution* with parameter λ associates to the integer $n \in \mathbb{N} = \Omega$ (sample space) the probability

$$p_n = \mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}. \tag{B.11}$$

The corresponding average value and variance are $\langle X \rangle = \sigma^2 = \lambda$.

(77) The limiting cases $p = 0$ or $p = 1$ are trivial.
(78) The extension to $r = 0$ is trivial.

B.3.5 Continuous uniform distribution

Consider a continuous random variable X whose sample space Ω is the real interval $]a,b]$ with $a<b$. A constant probability density

$$p_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a < x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{B.12}$$

on this range represents an instance of *continuous uniform distribution*. This is quite obviously the generalization to the continuous case of the discrete uniform distribution (B.8).

B.3.6 Gaussian distribution

Let $\Omega = \mathbb{R}$ represent the sample space for a continuous random variable X. The probability density

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{B.13}$$

is called *Gaussian distribution* (or *normal distribution*).

The average value is $\langle X \rangle = \mu$ and the variance σ^2. One can also easily check that the cumulants κ_m of all orders $m \geq 3$ vanish identically.

B.3.7 Exponential distribution

Let λ be a positive real number. A continuous random variable X with sample space $\Omega = \mathbb{R}_+$ is said to obey the *exponential distribution* with parameter λ if its probability density is given by

$$p_X(x) = \lambda\, \mathrm{e}^{-\lambda x}. \tag{B.14}$$

The expectation value is $\langle X \rangle = \dfrac{1}{\lambda}$ and the variance $\sigma^2 = \dfrac{1}{\lambda^2}$.

B.3.8 Cauchy–Lorentz distribution

Let x_0 and γ be two real numbers, with $\gamma > 0$. The *Cauchy–Lorentz distribution*, also called in physics (non-relativistic) *Breit*[(bp)]–*Wigner*[(bq)] *distribution* or shortly *Lorentzian*, is given by

$$p_X(x) = \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2 + \gamma^2} \tag{B.15}$$

for $x \in \Omega = \mathbb{R}$.

All moments of this distribution diverge! x_0 is the position of the maximum of the distribution—i.e. it corresponds to the most probable value of the random variable—, while 2γ represents the full width at half maximum (often abbreviated FWHM).

B.4 Multidimensional random variables

Let $\mathbf{X}$ be a D-dimensional random variable, whose components will be denoted $X_1, X_2, \ldots, X_D$. For the sake of brevity, we shall hereafter only consider the case of continuous variables.

B.4.1 Definitions

The probability density $p_{\mathbf{X}}(\mathbf{x})$ is also called *multivariate* or *joint probability density* of the D random variables. For commodity, we shall also denote this density by $p_D(x_1, \ldots, x_D) \equiv p_{\mathbf{X}}(\mathbf{x})$.

[(bp)] G. Breit, 1899–1981 [(bq)] E. P. Wigner, 1902–1995

B.4.1 a Moments and averages

The *moments* of a multivariate probability density are defined as the expectation values

$$\langle X_1^{m_1} X_2^{m_2} \cdots X_D^{m_D} \rangle \equiv \int x_1^{m_1} x_2^{m_2} \cdots x_D^{m_D}\, p_D(x_1,\ldots,x_D)\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_D. \tag{B.16}$$

They are generated by the *characteristic function*

$$G_{\mathbf{X}}(k_1,\ldots,k_D) \equiv \langle \mathrm{e}^{\mathrm{i}(k_1X_1+\cdots+k_DX_D)} \rangle, \tag{B.17}$$

with $k_1,\ldots,k_D$ real auxiliary variables. The logarithm of this characteristic function generates the corresponding (joint) *cumulants*.

Combinations of moments that play an important role are the *covariances*

$$\big\langle \big(X_i - \langle X_i\rangle\big)\big(X_j - \langle X_j\rangle\big) \big\rangle = \langle X_iX_j\rangle - \langle X_i\rangle\langle X_j\rangle \tag{B.18}$$

for every $i,\, j \in \{1,\ldots,D\}$. These are often combined into a symmetric *covariance matrix*, of which they constitute the entries. One easily checks that the covariances are actually the second-order cumulants of the joint probability distribution.

Useful, dimensionless measures are then the *correlation coefficients* obtained by dividing the covariance of random variables X_i, X_j by the product of their standard deviations (B.5)

$$c_{ij} \equiv \frac{\langle X_iX_j\rangle - \langle X_i\rangle\langle X_j\rangle}{\sigma_{X_i}\sigma_{X_j}}. \tag{B.19}$$

Obviously, the diagonal coefficients c_{ii} are identically equal to 1.

If the covariance—or equivalently the correlation coefficient—of two random variables X_i and X_j vanishes, then these variables are said to be *uncorrelated*.

B.4.1 b Marginal and conditional probability distributions

Consider a nonnengative integer $r < D$ and choose r random variables among $X_1, X_2, \ldots, X_D$—for the sake of simplicity, the first r ones $X_1, \ldots, X_r$. The probability that the latter take values in the intervals $[x_1, x_1 + \mathrm{d}x_1], \ldots, [x_r, x_r + \mathrm{d}x_r]$, *irrespective* of the values taken by $X_{r+1}, \ldots, X_D$, is

$$p_r(x_1,\ldots,x_r)\, \mathrm{d}x_1 \cdots \mathrm{d}x_r = \left[\int p_D(x_1,\ldots,x_r,x_{r+1},\ldots,x_D)\, \mathrm{d}x_{r+1} \cdots \mathrm{d}x_D\right] \mathrm{d}x_1 \cdots \mathrm{d}x_r,$$

where the integral runs over the $(D-r)$-dimensional sample space of the variables $X_{r+1},\ldots,X_D$, which have thus been "integrated out". The density

$$p_r(x_1,\ldots,x_r) \equiv \int p_D(x_1,\ldots,x_r,x_{r+1},\ldots,x_D)\, \mathrm{d}x_{r+1} \cdots \mathrm{d}x_D \tag{B.20}$$

for the remaining random variables $X_1,\ldots,X_r$ is then called *marginal distribution*.

If the random variables $X_{r+1},\ldots,X_D$ take given realizations $x_{r+1},\ldots,x_D$, then one can consider the probability distribution for the remaining random variables under this condition. Accordingly, one introduces the corresponding *conditional probability density*

$$p_{r|D-r}(x_1,\ldots,x_r \,|\, x_{r+1},\ldots,x_D). \tag{B.21}$$

One has then the identities

$$\begin{aligned} p_D(x_1,\ldots,x_D) &= p_{r|D-r}(x_1,\ldots,x_r \,|\, x_{r+1},\ldots,x_D)\, p_{D-r}(x_{r+1},\ldots,x_D) \\ &= p_{D-r|r}(x_{r+1},\ldots,x_D \,|\, x_1,\ldots,x_r)\, p_r(x_1,\ldots,x_r). \end{aligned}$$

The first identity can be rewritten as

$$p_{r|D-r}(x_1,\dots,x_r\,|\,x_{r+1},\dots,x_D) = \frac{p_D(x_1,\dots,x_D)}{p_{D-r}(x_{r+1},\dots,x_D)}, \tag{B.22}$$

which constitutes *Bayes'*[br] *theorem*.

B.4.2 Statistical independence

When the identity

$$p_D(x_1,\dots,x_D) = p_r(x_1,\dots,x_r)\,p_{D-r}(x_{r+1},\dots,x_D) \tag{B.23}$$

holds for all realizations $x_1,\dots,x_r,x_{r+1},\dots,x_D$ of the random variables, then the sets of variables $\{X_1,\dots,X_r\}$ and $\{X_{r+1},\dots,X_D\}$ are said to be *statistically independent* (or shortly *independent*). In that case, the marginal probability distribution for X_1, …, X_r (resp. for X_{r+1}, …, X_D) equals the conditional distribution:

$$p_r(x_1,\dots,x_r) = p_{r|D-r}(x_1,\dots,x_r\,|\,x_{r+1},\dots,x_D).$$

Let X_1 and X_2 be two statistically independent random variables. For all functions f_1, f_2 defined on the respective sample spaces,[79] one has the identity $\langle f_1(X_1)\,f_2(X_2)\rangle = \langle f_1(X_1)\rangle\,\langle f_2(X_2)\rangle$. In particular, all moments—if defined—obey

$$\langle X_1^{m_1} X_2^{m_2}\rangle = \langle X_1^{m_1}\rangle\,\langle X_2^{m_2}\rangle \quad \forall m_1, m_2,$$

as one sees by considering the characteristic functions of the random variables.

The latter equation shows that the statistical independence of two random variables implies that they are uncorrelated. The converse is not however true, although both notions are often taken as identical.

B.4.3 Addition of random variables

Consider again two random variables X_1, X_2 defined on the same sample space, whose joint probability density is denoted by $p_{\mathbf{X}}(x_1,x_2)$.

Their sum $Y = X_1 + X_2$ constitutes a new random variable with the expectation value

$$\langle Y\rangle = \langle X_1\rangle + \langle X_2\rangle \tag{B.24}$$

and more generally the probability density

$$\begin{aligned} p_Y(y) &= \int p_{\mathbf{X}}(x_1,x_2)\,\delta(y-x_1-x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int p_{\mathbf{X}}(x_1,y-x_1)\,\mathrm{d}x_1 = \int p_{\mathbf{X}}(y-x_2,x_2)\,\mathrm{d}x_2. \end{aligned} \tag{B.25}$$

This corresponds to the characteristic function

$$G_Y(k) = G_{X_1,X_2}(k,k), \tag{B.26}$$

where $G_{X_1,X_2}(k_1,k_2)$ is the generating function for $p_{\mathbf{X}}(x_1,x_2)$.

If X_1 and X_2 are statistically independent, Eq. (B.23) allows one to simplify Eq. (B.25) into

$$p_Y(y) = \int p_{X_1}(x_1)\,p_{X_2}(y-x_1)\,\mathrm{d}x_1 = \int p_{X_1}(y-x_2)\,p_{X_2}(x_2)\,\mathrm{d}x_2,$$

[79]... and whose product can be defined in some way, in case the functions are neither real- nor complex-valued, as e.g. the scalar product of two vectors.

[br]T. Bayes, 1702–1761

that is, into the convolution of p_{X_1} with p_{X_2}. In this case, the variance of Y is

$$\sigma_Y^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2. \tag{B.27}$$

This property generalizes to all cumulants of Y (however, not to its central moments!), as follows at once from the identity

$$G_Y(k) = G_{X_1}(k)\, G_{X_2}(k).$$

Remark: The latter equation shows at once that the sum of two Gaussian variables—and more generally, any linear combination of Gaussian variables—is itself a Gaussian variable.

B.4.4 Multidimensional Gaussian distribution

Let A be a positive definite, symmetric $D \times D$ matrix and $\mathbf{B}$ a D-dimensional vector. The multivariate Gaussian distribution for random variables $\mathbf{X} = (X_1, \ldots, X_D)$ is given by

$$p_{\mathbf{X}}(\mathbf{x}) = \sqrt{\frac{\det \mathsf{A}}{(2\pi)^D}}\, \mathrm{e}^{-\frac{1}{2}\mathbf{B}^T\cdot\mathsf{A}^{-1}\cdot\mathbf{B}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{D} A_{ij}x_i x_j - \sum_{i=1}^{D} B_i x_i\right], \tag{B.28}$$

with A_{ij} resp. B_i the elements of A resp. the components of $\mathbf{B}$, while $\mathbf{B}^T$ denotes the transposed vector of $\mathbf{B}$.

B.5 Central limit theorem

Consider a sequence $(X_1, X_2, \ldots, X_n, \ldots)$ of statistically independent one-dimensional random variables with the same sample space Ω and the same probability distribution.[(80)] One assumes that both the expectation value μ and the variance σ^2 of the distribution exist. Let

$$Z_N \equiv \frac{1}{N}\sum_{n=1}^{N} X_n \tag{B.29a}$$

denote the N-th partial sum of these random variables, multiplied with an adequate normalization factor. Following the results of § B.4.3, the expectation value of Z_N exactly equals μ while the variance is σ^2/N.

According to the *central limit theorem*,[(81)] the probability distribution for the random variable $\sqrt{N}(Z_N - \mu)$ converges for $N \to \infty$ towards the Gaussian distribution with expectation value $\mu_1 = 0$ and variance σ^2, i.e. for every real number z

$$p_{Z_N}(z) \underset{N\gg 1}{\sim} \frac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2/N}\right]. \tag{B.29b}$$

This theorem underlies the important role of the Gaussian distribution and is related to the law of large numbers. Since the variance of the distribution of Z_N becomes smaller with increasing N, the possible realizations z become more and more concentrated about the expectation value μ: the distribution approaches a δ-distribution at the point μ.

Remarks:

∗ The convergence in Eq. (B.29b) is actually a weak convergence, or "convergence in distribution", analogous to the pointwise convergence of "usual" (i.e. non-stochastic) sequences.

∗ There exist further analogous theorems (the version above is the theorem of Lindeberg[(bs)]–Lévy[(bt)]) for statistically independent random variables with different probability distributions, for "nearly independent" random variables...

[(80)] Such variables are referred to as "independent and identically distributed" (i.i.d.) random variables.
[(81)] ... in its simplest incarnation.

[(bs)] J. W. Lindeberg, 1876–1932 [(bt)] P. Lévy, 1886–1971

Elements of a proof:
The probability density for Z_N follows from the generalizations of Eqs. (B.25) and (B.23)

$$p_{Z_N}(z) = \int p_{X_1}(x_1)\cdots p_{X_N}(x_N)\,\delta\!\left(z - \frac{1}{N}\sum_{n=1}^{N} x_n\right) \mathrm{d}x_1\cdots\mathrm{d}x_N,$$

where $p_{X_1},\ldots,p_{X_N}$ actually all reduce to the same density p_X. Inserting the Fourier representation of the δ distribution, this becomes

$$\begin{aligned}p_{Z_N}(z) &= \int p_{X_1}(x_1)\cdots p_{X_N}(x_N)\,\exp\!\left[\mathrm{i}k\!\left(\frac{1}{N}\sum_{n=1}^{N} x_n - z\right)\right] \mathrm{d}x_1\cdots\mathrm{d}x_N\,\frac{\mathrm{d}k}{2\pi}\\ &= \int \mathrm{e}^{-\mathrm{i}kz}\prod_{n=1}^{N}\left(\int p_X(x_n)\,\mathrm{e}^{\mathrm{i}kx_n/N}\,\mathrm{d}x_n\right)\frac{\mathrm{d}k}{2\pi} = \int \mathrm{e}^{-\mathrm{i}kz}\left[G_X\!\left(\frac{k}{N}\right)\right]^N \frac{\mathrm{d}k}{2\pi},\end{aligned}$$

where G_X is the characteristic function (B.6a). A Taylor expansion of the latter at $k = 0$ yields

$$G_X\!\left(\frac{k}{N}\right) = 1 + \frac{\mathrm{i}k\mu}{N} - \frac{k^2\langle X^2\rangle}{2N^2} + \mathcal{O}\!\left(\frac{1}{N^3}\right),$$

i.e.

$$N\ln G_X\!\left(\frac{k}{N}\right) = \mathrm{i}k\mu - \frac{k^2\sigma^2}{2N} + \mathcal{O}\!\left(\frac{1}{N^2}\right).$$

This eventually gives

$$p_{Z_N}(z) \underset{N\gg 1}{\sim} \int \exp\!\left[-\frac{k^2\sigma^2}{2N} - \mathrm{i}k(z-\mu)\right]\frac{\mathrm{d}k}{2\pi} = \frac{1}{\sqrt{2\pi\sigma^2/N}}\exp\!\left[-\frac{(z-\mu)^2}{2\sigma^2/N}\right]. \qquad \square$$