PROBABILITY THEORY

BRAD BAXTER

Version: 200905281454

1. Introduction

The textbooks mentioned in the formal syllabus are fine, but the following pair are particularly useful:

- *Probability and Random Processes*, by G. Grimmett and D. Stirzaker.
- *Weighing the Odds*, by D. Williams.

Grimmett and Stirzaker is thorough, but somewhat dry for my taste. Williams' book is inspiring, but slightly eccentric. I hope to steer a safe path between the two. You can download an earlier set of older lecture notes, developed by my colleagues, A. Cartea and L. Sinnadurai, from my office linux server:

`http://econ109.econ.bbk.ac.uk/brad/PSM/`

This server also contains other useful information. The syllabus is unchanged from previous years, but I shall present the material in an alternative order.

These notes are provisional, and will be updated regularly; please note the date stamp in the heading above. Please check the webpage above for newer versions (`probnotes.pdf`) and let me know of any slips.

Brad Baxter [`b.baxter@bbk.ac.uk`]

1.1. **Some Fundamentals.**

A continuous random variable X is a real-valued random variable for which there is a *probability density function (PDF)* $p(s)$, for $s \in \mathbb{R}$, for which

$$\mathbb{P}(a \le X \le b) = \int_a^b p(s)\,ds, \tag{1.1}$$

for any real numbers $a \le b$. Its *expected value* is then given by

$$\mu = \mathbb{E}X = \int_{-\infty}^{\infty} sp(s)\,ds \tag{1.2}$$

and its *variance* by

$$\operatorname{var} X = \mathbb{E}\left([X-\mu]^2\right) = \int_{-\infty}^{\infty} (s-\mu)^2 p(s)\,ds. \tag{1.3}$$

if the integral exists[1]

Exercise 1.1. *Show that*

$$\operatorname{var} X = \mathbb{E}(X^2) - (\mathbb{E}X)^2\,.$$

[1] As we shall soon see, this integrability requirement *fails* for the Cauchy distribution.

Exercise 1.2. *We say that* U *is* uniformly distributed *on the interval* $[a, b]$ *if its PDF is given by*

$$p(s) = \begin{cases} \frac{1}{b-a} & \text{if } a \le s \le b, \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $\mathbb{E}U$ *and* $\operatorname{var} U$.

The variance of a sum of independent random variables is simply the sum of their variances, but the general case is slightly more complicated.

Proposition 1.1. *We have*

$$\operatorname{var} X + Y = \operatorname{var} X + \operatorname{var} Y + 2\left[\mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)\right]. \tag{1.4}$$

Thus $\operatorname{var} X + Y = \operatorname{var} X + \operatorname{var} Y$ *if and only if* $\mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = 0$.

Proof. We have

$$\begin{aligned} \operatorname{var} X + Y &= \mathbb{E}\left[(X+Y)^2\right] - (\mathbb{E}X + Y)^2 \\ &= \mathbb{E}\left[X^2 + Y^2 + 2XY\right] - (\mathbb{E}X)^2 - (\mathbb{E}Y)^2 - 2(\mathbb{E}X)(\mathbb{E}Y) \\ &= \mathbb{E}(X^2) - (\mathbb{E}X)^2 + \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 + 2\left[\mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)\right] \\ &= \operatorname{var} X + \operatorname{var} Y + 2\left[\mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)\right]. \end{aligned}$$

Thus $\operatorname{var} X + Y = \operatorname{var} X + \operatorname{var} Y$ if and only if $\mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = 0$. □

The quantity $\mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$ is sufficiently important to merit a name: it's called the *covariance* of X and Y.

Definition 1.1. *The covariance* $\operatorname{Cov}(X, Y)$ *is defined by*

$$\operatorname{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right]. \tag{1.5}$$

Further, if $\operatorname{Cov}(X, Y) = 0$*, then we say that* X *and* Y *are* uncorrelated.

Exercise 1.3. *Show that* $\operatorname{Cov}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$.

Thus we have shown that the variance of a sum of uncorrelated random variables is the sum of their variances. We come now to one of the most fundamental definitions in Probability Theory.

Definition 1.2. *We shall say that random variables* X *and* Y *are* independent *if*

$$\mathbb{P}\left(a \le X \le b \text{ and } c \le Y \le d\right) = \mathbb{P}\left(a \le X \le b\right) \cdot \mathbb{P}\left(c \le Y \le d\right),$$

for any real numbers $a \le b$ *and* $c \le d$.

Independence is a stronger condition than uncorrelated, but the key slogan is "Independence means multiply!".

Proposition 1.2. *If* X *and* Y *are independent random variables, then*

$$\mathbb{E}(f(X)g(Y)) = (\mathbb{E}f(X))(\mathbb{E}g(Y)),$$

for any continuous functions f *and* g.

Proof. This is not examinable in its full generality. However, one possible proof begins with the observation that $\mathbb{E}(X^pY^q) = \mathbb{E}(X^p)\mathbb{E}(Y^q)$, for any non-negative integers p and q, which implies that the theorem is true when f and g are polynomials. □

Proposition 1.3. *If X and Y are independent random variables, then they're uncorrelated, that is,* $\mathrm{Cov}(X,Y) = 0$.

Proof. This is "Independence means multiply!" again:

$$\begin{aligned}\mathbb{E}\left[(X-\overline{X})(Y-\overline{Y})\right] \\ &= \mathbb{E}(X-\overline{X})\cdot\mathbb{E}(Y-\overline{Y}) \\ &= 0.\end{aligned}$$

□

Exercise 1.4. *Prove that* $\mathbb{E}(aX+bY) = a\mathbb{E}X + b\mathbb{E}Y$*. Show further that* $\mathrm{var}(cX) = c^2\,\mathrm{var}\,X$*, for any* $c \in \mathbb{R}$.

More generally, we have the relation

$$\mathbb{E}f(X) = \int_{-\infty}^{\infty} f(s)p(s)\,ds. \tag{1.6}$$

1.2. **Random Vectors.**

Theorem 1.4. *Let $X_1, X_2, \ldots, X_n$ be independent continuous random variables with PDFs $p_1, p_2, \ldots, p_n$. Then the* random vector

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

has PDF

$$q(\mathbf{s}) = p_1(s_1)p_2(s_2)\cdots p_n(s_n), \qquad \text{for } \mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} \in \mathbb{R}^n.$$

Proof. TBW □

In general the components of a random vector $\mathbf{X}$ are not independent random variables, in which case its *covariance matrix* M is extremely useful. Specifically, we define $M_{jk} = \mathrm{Cov}(X_j, X_k)$, for $1 \le j,k \le n$. In matrix notation, we write

$$M = \equiv \mathrm{Cov}(\mathbf{X}) = \mathbb{E}\left(\mathbf{X} - \mathbb{E}\mathbf{X}\right)\left(\mathbf{X} - \mathbb{E}\mathbf{X}\right)^T.$$

The full derivation of (2.7) is not required, but the key observation is as follows.

Proposition 1.5. *Let*

$$U = [a_1,b_1]\times[a_2,b_2] \equiv = \left\{(x_1,x_2)\in\mathbb{R}^2 : a_1 \le x_1 \le b_1 \text{ and } a_2 \le x_2 \le b_2\right\}. \tag{1.7}$$

Then

$$\mathbb{P}\left(\mathbf{X} \in U\right) = \int_{a_1}^{b_1} ds_1 \int_{a_2}^{b_2} ds_2 p_1(s_1)p_2(s_2). \tag{1.8}$$

Proof. We have

$$\begin{aligned}\mathbb{P}(\mathbf{X} \in U) &= \mathbb{P}(a_1 \le X_1 \le b_1 \text{ and } a_2 \le X_2 \le b_2) \\ &= \mathbb{P}(a_1 \le X_1 \le b_1) \cdot \mathbb{P}(a_2 \le X_2 \le b_2) \\ &= \left(\int_{a_1}^{b_1} p_1(s_1)\, ds_1\right)\left(\int_{a_2}^{b_2} p_2(s_2)\, ds_2\right) \\ &= \int_{a_1}^{b_1} ds_1 \int_{a_2}^{b_2} ds_2 p_1(s_1) p_2(s_2),\end{aligned}$$

as required. □

1.3. **The Gaussian.**

Definition 1.3. *We say that a continuous random variable W is* Gaussian *(or, equivalently,* normal*) if its PDF is the Gaussian function*

$$p(s) = (2\pi\sigma^2)^{-1/2} \exp\Big(\frac{-(s-\mu)^2}{2\sigma^2}\Big), \qquad s \in \mathbb{R}. \tag{1.9}$$

As we shall see shortly, $\mathbb{E}W = \mu$ and $\operatorname{var} W = \sigma^2$. The standard notation is to write $W \sim N(\mu, \sigma^2)$.

One crucial integral is as follows.

Lemma 1.6.

$$I(c) = \int_{-\infty}^{\infty} e^{-cx^2}\, dx = \sqrt{\frac{\pi}{c}}, \tag{1.10}$$

for any $c > 0$.

Proof. If we use the change of variable $s = \sqrt{c}x$ in (1.10), then we obtain $I(c) = I(1)/\sqrt{c}$, so we only need to calculate $I(1)$. Now

$$\begin{aligned}I(1)^2 &= \left(\int_{\mathbb{R}} e^{-s^2}\, ds\right)^2 \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-(s^2+t^2)}\, ds\ dt \\ &= \int_0^{\infty}\int_0^{2\pi} e^{-r^2} ddrd\theta \\ &= \pi \int_0^{\infty} 2re^{-r^2}\, dr \\ &= \pi.\end{aligned}$$

Thus $I(1) = \sqrt{\pi}$. □

Lemma 1.7. *Furthermore,* (1.10) *is valid for any $c \in \mathbb{C}$ with positive real part.*

Proof. This proof is not examinable, but is not difficult if you've taken a first course in complex analysis. Firstly, the integral is well-defined if the real part of c is positive. Secondly, we have already shown that $I(c) = \sqrt{\pi/c}$ for $c > 0$. The identity principle for analytic functions therefore implies the result. □

1.4. **Multivariate Gaussians.** If $Z_1, \ldots, Z_n$ are independent $N(0,1)$ Gaussian random variables, then the random vector

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}$$

has PDF

$$q(\mathbf{s}) = (2\pi)^{-n/2} e^{-\|\mathbf{s}\|^2/2}, \qquad \text{for } \mathbf{s} \in \mathbb{R}^n.$$

We say that $\mathbf{Z}$ is a *normalized Gaussian random vector*. We see that $\mathbb{E}Z_k = 0$, $\mathbb{E}Z_k^2 = 1$, and $\mathbb{E}Z_k Z_\ell = 0$, for $k \neq \ell$. In matrix notation, we have $\mathbb{E}\mathbf{Z} = 0$, and its covariance matrix $\mathbb{E}\mathbf{Z}\mathbf{Z}^T = I$, the identity matrix; we write $\mathbf{Z} \sim N(0, I)$.

However, let us now consider the PDF for $\mathbf{W} = A\mathbf{Z}$, where $A \in \mathbb{R}^{n \times n}$ can be any nonsingular matrix (i.e. $\det A \neq 0$). Then, for any (measurable) set U in $\mathbb{R}^n$, we have

$$\mathbb{P}\left(\mathbf{W} \in U\right) = \mathbb{P}\left(\mathbf{Z} \in A^{-1}U\right),$$

where

$$A^{-1}U = \{A^{-1}\mathbf{u} : \mathbf{u} \in U\}.$$

hence

$$\begin{aligned} \mathbb{P}\left(\mathbf{W} \in U\right) &= \mathbb{P}\left(\mathbf{Z} \in A^{-1}U\right) \\ &= (2\pi)^{-n/2} \int_{A^{-1}U} \exp(-\|\mathbf{s}\|^2/2)\, d\mathbf{s} \\ &= (2\pi)^{-n/2} \int_{A^{-1}U} \exp(-\mathbf{s}^T\mathbf{s}/2)\, d\mathbf{s} \\ &= (2\pi)^{-n/2} \det(A^{-1}) \int_U \exp(-(A^{-1}\mathbf{t})^T(A^{-1}\mathbf{t})/2)\, d\mathbf{t} \\ &= (2\pi)^{-n/2} \det(A)^{-1} \int_U \exp(-\mathbf{t}^T(A^{-T})A^{-1}\mathbf{t})/2)\, d\mathbf{s} \end{aligned}$$

using the substitution $\mathbf{t} = A\mathbf{s}$, so that $d\mathbf{t} = |\det(A)| d\mathbf{s}$, together with the notation $A^{-T} \equiv (A^{-1})^T$. Hence the $\mathbb{E}\mathbf{W} = \mathbb{E}(A\mathbf{Z}) = A\mathbb{E}\mathbf{Z} = 0$ and its covariance matrix is given by

$$\begin{aligned} M &= \mathbb{E}\left(\mathbf{W}\mathbf{W}^T\right) \\ &= \mathbb{E}(A\mathbf{Z})(A\mathbf{Z})^T \\ &= A\mathbb{E}\mathbf{Z}\mathbf{Z}^T A^T \\ &= AA^T. \end{aligned}$$

Hence $M^{-1} = A^{-T}A^{-1}$, and $\mathbf{W}$ has PDF $q(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^n$, where

$$q(\mathbf{t}) = (2\pi)^{-n/2}(\det M)^{-1/2} \exp(-\mathbf{t}^T M^{-1}\mathbf{t}/2), \qquad \mathbf{t} \in \mathbb{R}^n.$$

We write $\mathbf{W} \sim N(0, M)$.

1.5. **Stirling's formula.** We have

$$\log(n-1)! \le \int_1^n \log x \, dx \le \log n!,$$

or

$$\log n! - \log n \le n(\log n - 1) \le \log n!.$$

Dividing by n, we obtain

$$\log((n!)^{1/n}) - \frac{\log n}{n} \le \log n - 1 \le \log((n!)^{1/n}).$$

Rearranging these inequalities, we have

$$0 \le \log((n!)^{1/n}) - \log n + 1 \le \frac{\log n}{n},$$

or

$$0 \le \log\left(\frac{(n!)^{1/n}}{n/e}\right) \le \frac{\log n}{n}.$$

Hence,

$$\lim_{n\to\infty} \log\left(\frac{(n!)^{1/n}}{n/e}\right) = 0,$$

or

$$\lim_{n\to\infty} \frac{(n!)^{1/n}}{n/e} = 1.$$

We need a less cumbersome notation to describe these properties. If two sequences (u_n) and (v_n) satisfy $\lim_{n\to\infty} u_n/v_n = 1$, then we shall write $u_n \sim v_n$. In other words, $u_n \sim v_n$ if the percentage error in estimating u_n by v_n tends to zero.

Exercise 1.5. *Show that $u_n = n^2$ and $v_n = (n+1)^2$ satisfy $u_n \sim v_n$, as $n \to \infty$, but that $v_n - u_n \to \infty$.*

In our new notation, we have shown that

$$(n!)^{1/n} \sim n/e,$$

as $n \to \infty$. However, a stronger result was derived by de Moivre and Stirling in the early Eighteenth century:

Theorem 1.8 (Stirling's Formula)**.** *It can be shown that*

$$\sqrt{2\pi n}(n/e)^n \le n! \le e^{1/(12n)}\sqrt{2\pi n}(n/e)^n, \tag{1.11}$$

for all positive n. In particular, we have $n! \sim \sqrt{2\pi n}(n/e)^n$.

Proof. See Section 2.9 of the excellent *An Introduction to Probability Theory and its Applications*, by William Feller. □

It can also be shown (see Feller again) that the percentage error in using the right hand side of (1.11) to estimate $n!$ is at most $9n^{-2}$ per cent. For example, when $n = 5$, the right hand side gives 120.01.

2. Sums of Random Variables

2.1. **Sums of Exponentially Distributed r.v.s.** We shall say that a random variable X is *exponentially distributed* if it has the PDF

$$p(s) = \begin{cases} e^{-s} & \text{if } s \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Let $X_1, X_2, \dots, X_n$ be independent exponentially distributed random variables. We shall calculate the PDF of their sum $X_1 + X_2 + \cdots + X_n$. To this end, we introduce the random vector

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^n, \tag{2.2}$$

which has PDF

$$q(\mathbf{s}) = e^{-(s_1+s_2+\cdots+s_n)}, \qquad \text{for } \mathbf{s} \geq 0. \tag{2.3}$$

Then

$$\begin{aligned} &\mathbb{P}\left(a \leq X_1 + X_2 + \cdots + X_n \leq b\right) \\ &\quad = \int_{a \leq s_1+\cdots+s_n \leq b} q(\mathbf{s})\, d\mathbf{s} \\ &\quad = \int\limits_{a \leq s_1+\cdots+s_n \leq b,\ \mathbf{s} \geq 0} e^{-(s_1+\cdots+s_n)}\, d\mathbf{s} \\ &\quad = \int_a^b e^{-u} C_n u^{n-1}\, du, \end{aligned}$$

where C_n is a constant depending only on the ambient dimension n. To calculate C_n, we set $a = 0$ and $b = \infty$, providing

$$1 = C_n \int_0^\infty e^{-u} u^{n-1}\, du,$$

or

$$C_n = \frac{1}{(n-1)!}.$$

Hence the PDF for $X_1 + \cdots + X_n$ is given by

$$r(u) = \begin{cases} \frac{1}{(n-1)!} e^{-u} u^{n-1} & \text{if } u \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

2.2. **Sums of Gaussians.** Now suppose we have a pair of independent $N(0,1)$ Gaussian random variables Z_1 and Z_2 and want to calculate the PDF of their sum $Z_1 + Z_2$. The Gaussian random vector

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \in \mathbb{R}^2$$

has PDF

$$p(\mathbf{s}) = (2\pi)^{-1} e^{-\|\mathbf{s}\|^2/2}, \qquad \mathbf{s} \in \mathbb{R}^2.$$

Thus

$$\mathbb{P}\left(a \le X_1 + X_2 \le b\right) = \int_{a \le s_1 + s_2 \le b} (2\pi)^{-1} e^{-(s_1^2+s_2^2)/2}\, d\mathbf{s}. \tag{2.4}$$

We now change coordinates, using $u_1 = s_1 + s_2$ and $u_2 = s_1 - s_2$, that is,

$$\mathbf{u} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \mathbf{s}.$$

Then

$$d\mathbf{u} = \left| \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right| \mathbf{s} = 2d\mathbf{s}$$

and $s_1 = (u_1 + u_2)/2$, $s_2 = (u_1 - u_2)/2$. Therefore

$$s_1^2 + s_2^2 = \frac{1}{2}(u_1^2 + u_2^2),$$

which implies that

$$\begin{aligned} \mathbb{P}(a \le Z_1 + Z_2 \le b) &= \frac{1}{2} \int_a^b du_1 \int_{-\infty}^{\infty} du_2 (2\pi)^{-1} e^{-(u_1^2+u_2^2)/4} \\ &= \int_a^b du_1 (2\pi.2)^{-1/2} e^{-u_1^2/(2.2)}, \qquad (2.5) \end{aligned}$$

that is we have shown that $Z_1 + Z_2 \sim N(0, 2)$. We now prove the more general result.

Proposition 2.1. *Let Z_1 and Z_2 be independent $N(0, 1)$ random variables. Then their linear combination $w_1 Z_1 + w_2 Z_2 \sim N(0, w_1^2 + w_2^2)$, for any $w_1, w_2 \in \mathbb{R}$.*

Proof. To calculate the PDF of the sum, we proceed as before:

$$P = \mathbb{P}\left(a \le w_1 Z_1 + w_2 Z_2 \le b\right) = \int_{a \le w_1 s_1 + w_2 s_2 \le b} (2\pi)^{=1/2} e^{-(s_1^2+s_2^2)/2}\, d\mathbf{s}. \tag{2.6}$$

The problem here is to rotate our coordinate system so that the integral becomes simple. If we let

$$\mathbf{u}_1 = \frac{1}{\sqrt{w_1^2 + w_2^2}} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \quad \text{and} \quad \mathbf{u}_2 = \frac{1}{\sqrt{w_1^2 + w_2^2}} \begin{pmatrix} -w_2 \\ w_1 \end{pmatrix},$$

then these are orthogonal unit vectors. Our new coordinates t_1, t_2 are defined by

$$\mathbf{s} = t_1 \mathbf{u}_1 + t_2 \mathbf{u}_2,$$

or

$$\mathbf{s} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix} \mathbf{t},$$

where $\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix}$ denotes the 2×2 matrix whose columns are $\mathbf{u}_1$ and $\mathbf{u}_2$, whilst

$$\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \in \mathbb{R}^2.$$

Then

$$\begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Hence

$$\mathbf{t} = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{pmatrix} \mathbf{u}$$

and

$$\det\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix} = 1.$$

Finally

$$\begin{aligned}\{\mathbf{s}\in\mathbb{R}^2 : a \le w_1 s_1 + w_2 s_2 \le b\} &= \{\mathbf{s}\in\mathbb{R}^2 : \frac{a}{\|\mathbf{w}\|} \le \frac{w_1 s_1 + w_2 s_2}{\sqrt{w_1^2+w_2^2}} \le \frac{b}{\|\mathbf{w}\|}\} \\ &= \{\mathbf{s}\in\mathbb{R}^2 : \frac{a}{\|\mathbf{w}\|} \le \mathbf{u}_1^T\mathbf{s} \le \frac{b}{\|\mathbf{w}\|}\} \\ &= \{\mathbf{t}\in\mathbb{R}^2 : \frac{a}{\|\mathbf{w}\|} \le t_1 \le \frac{b}{\|\mathbf{w}\|}\}.\end{aligned}$$

Further $s_1^2 + s_2^2 = t_1^2 + t_2^2$, so that

$$\begin{aligned} P &= \int_{a/\|\mathbf{w}\|}^{b/\|\mathbf{w}\|} dt_1 \int_{-\infty}^{\infty} dt_2 (2\pi)^{-1} e^{-(t_1^2+t_2^2)/2} \\ &= \int_{a/\|\mathbf{w}\|}^{b/\|\mathbf{w}\|} dt_1 (2\pi)^{-1/2} e^{-t_1^2/2} \int_{-\infty}^{\infty} dt_2 (2\pi)^{-1/2} e^{-t_2^2/2} \\ &= \int_{a/\|\mathbf{w}\|}^{b/\|\mathbf{w}\|} (2\pi)^{-1/2} e^{-t_1^2/2}\, dt_1 \\ &= \int_a^b (2\pi\|\mathbf{w}\|^2)^{-1/2} e^{-t_1^2/(2\|\mathbf{w}\|^2)}\, dt_1, \end{aligned}$$

using the change of variable $v_1 = t_1/\|\mathbf{w}\|$. Hence $w_1 Z_1 + w_2 Z_2 \sim N(0, \|\mathbf{w}\|^2)$. □

2.3. **General Sums of Random Variables.** Let X_1 and X_2 be independent random variables with PDFs f_1 and f_2, respectively. Then the PDF of the random vector

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \in \mathbb{R}^2$$

is given by

$$p(\mathbf{s}) = f_1(s_1) f_2(s_2), \qquad \text{for } \mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \in \mathbb{R}^2. \tag{2.7}$$

Thus

$$\mathbb{P}(a \le X_1 + X_2 \le b) = \int_{a \le s_1 + s_2 \le b} f_1(s_1) f_2(s_2)\, dvs.$$

Again using the change of variables

$$\mathbf{u} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \mathbf{s}.$$

we have

$$d\mathbf{u} = \left|\det\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\right| \mathbf{s} = 2d\mathbf{s}$$

and

$$\mathbb{P}(a \le X_1 + X_2 \le b) = \frac{1}{2}\int_a^b du_1 \int_{-\infty}^{\infty} du_2 f_1(\frac{u_1+u_2}{2}) f_2(\frac{u_1-u_2}{2}).$$

Therefore the PDF for $X_1 + X_2$ is given by

$$q(u_1) = \frac{1}{2} \int_{-\infty}^{\infty} f_1 \left(\frac{u_1 + u_2}{2} \right) f_2 \left(\frac{u_1 - u_2}{2} \right) du_2.$$

If we now make the further change of variable $y = (u_1 + u_2)/2$, then $(u_1 - u_2)/2 = u_1 - y$ and $dy = (1/2)du_2$, and the integral becomes

$$q(u_1) = \int_{-\infty}^{\infty} f_1(y) f_2(u_1 - y)\, dy, \qquad \text{for } u_1 \in \mathbb{R},$$

so this particular integral provides the PDF for $X_1 + X_2$ when X_1, X_2 are independent continuous random variables with PDFs f_1, f_2 — it's sufficiently important to merit a standard notation:

Definition 2.1. *The* convolution $f_1 * f_2$ *is defined by*

$$f_1 * f_2(u) = \int_{-\infty}^{\infty} f_1(y) f_2(u - y)\, dy, \qquad \textit{for } u \in \mathbb{R}.$$

We now summarize our findings.

Theorem 2.2. *If X_1 and X_2 are independent continuous random variables with PDFs f_1 and f_2, respectively, then their sum $X_1 + X_2$ has PDF $f_1 * f_2$.*

Corollary 2.3. *If $X_1, X_2, \ldots, X_n$ are independent continuous random variables, with associated PDFs $f_1, f_2, \ldots, f_n$, then their sum $X_1 + \cdots + X_n$ has PDF $f_1 * \cdots * f_n$. [Here $f_1 * \cdots * f_n = (f_1 * \cdots * f_{n-1}) * f_n$.]*

Corollary 2.4. *If f_1 and f_2 are the probability density functions any two independent continuous random variables X_1 and X_2, then the order of their convolution is irrelevant, that is, $f_1 * f_2 = f_2 * f_1$.*

Proof. We have

$$\begin{aligned} f_1 * f_2 &= \text{PDF of } X_1 + X_2 \\ &= \text{PDF of } X_2 + X_1 \\ &= f_2 * f_1. \end{aligned}$$

□

The great advantage of the convolution integral is that it avoids multivariate integration, although sometimes the integration is the easier route – that's why you learn both. This is best illustrated by an example.

Example 2.1. *Suppose $X_1, X_2, \ldots, X_n$ are independent exponentially distributed random variables. Let's calculate the PDF $p_n(s)$ of their sum $X_1 + \cdots + X_n$ via the convolution integral. Using*

$$p_1(s) = \begin{cases} e^{-s} & \textit{if } s \geq 0, \\ 0 & \textit{otherwise,} \end{cases} \tag{2.8}$$

we deduce that the PDF $p_2(s)$ of $X_1 + X_2$ is given by the convolution integral

$$p_2(s) = p_1 * p_1(s) = \int_{-\infty}^{\infty} p_1(z) p_1(s - z)\, dz.$$

Now $p_1(z) = 0$ *if* $z < 0$ *and* $p_1(s-z) = 0$ *if* $s - z < 0$. *In other words,* $p_1(z)p_1(s-z) > 0$ *if and only if* $0 < z < s$. *Hence*

$$\begin{aligned} p_2(s) &= \int_0^s p_1(z)p_1(s-z)\,dz \\ &= \int_0^s e^{-z}e^{-(s-z)}\,sz \\ &= se^{-s}, \end{aligned}$$

for $s \geq 0$*; obviously* $p_2(s) = 0$ *for* $s < 0$.

Similarly, we have

$$\begin{aligned} p_n(s) &= p_1 * p_{n-1}(s) \\ &= \int_{-\infty}^{\infty} p_1(z)p_{n-1}(s-z)\,dz. \end{aligned}$$

Exercise 2.1. *Using the terminology of the last example, using proof by induction to show that*

$$p_n(s) = \begin{cases} \frac{s^{n-1}}{(n-1)!}e^{-s} & \text{if } s \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

3. Characteristic Functions

It's extremely tedious computing convolution integrals. Fortunately, there is an extremely clever alternative due to Fourier.

Definition 3.1. *Let* X *be any random variable for which the expectation* $\mathbb{E}(|X|)$ *of its absolute value is finite. Then the* characteristic function*, of CF, of* X *is defined by*

$$\phi_X(z) = \mathbb{E}e^{izX}.$$

If you have encountered Fourier transforms previously, then I hope you've noticed that, for a continuous random variable X with PDF $p(s)$, the characteristic function $\phi_X(z)$ is simply the Fourier transform of the PDF, that is,

$$\phi_X(z) = \int_{-\infty}^{\infty} e^{izs}p(s)\,ds, \qquad z \in \mathbb{R}.$$

Example 3.1. *Suppose* $X \sim U[-1/2, 1/2]$*, i.e. it's uniformly distributed on the interval* $[-1/2, 1/2]$. *Thus its PDF is given by*

$$p(s) = \begin{cases} 1 & \text{if } -1/2 \leq s \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$
\begin{aligned}
\phi_X(z) &= \mathbb{E}e^{izX} \\
&= \int_{-1/2}^{1/2} e^{izs}\,ds \\
&= \int_{-1/2}^{1/2} \cos(zs)\,ds + i\int_{-1/2}^{1/2} \sin(zs)\,ds \\
&= 2\int_0^{1/2} \cos(zs)\,ds \\
&= \frac{\sin(z/2)}{z/2}.
\end{aligned}
$$

The key property of characteristic functions is the Convolution Theorem:

Theorem 3.1 (Convolution Theorem)**.** *Let X_1, X_2 be independent continuous random variables with CFs $\phi_{X_1}(z)$ and $\phi_{X_2}(z)$, respectively. Then the CF $\phi_{X_1+X_2}(z)$ corresponding to the sum of random variables $X_1 + X_2$ satisfies*

$$\phi_{X_1+X_2}(z) = \phi_{X_1}(z)\phi_{X_2}(z).$$

Proof. We have

$$
\begin{aligned}
\phi_{X_1+X_2}(z) &= \mathbb{E}e^{iz(X_1+X_2)} \\
&= \mathbb{E}e^{izX_1}\mathbb{E}e^{izX_2} \\
&= \phi_{X_1}(z)\phi_{X_2}(z).
\end{aligned}
$$

□

Example 3.2. *Let $X_1, X_2, \ldots, X_n$ be exponentially distributed random variables, their common PDF being given by*

$$p(s) = \begin{cases} e^{-s} & \textit{if } s \ge 0, \\ 0 & \textit{otherwise.} \end{cases} \tag{3.1}$$

Their common CF is then

$$
\begin{aligned}
\phi(z) &= \mathbb{E}e^{izX} \\
&= \int_0^\infty e^{izs}e^{-s}\,ds \\
&= \int_0^\infty e^{s(-1+iz)}\,ds \\
&= \left[\frac{e^{s(-1+iz)}}{-1+iz}\right]_{s=0}^\infty \\
&= \frac{-1}{-1+iz} \\
&= \frac{1}{1-iz}.
\end{aligned}
$$

The sum $S_n = X_1 + X_2 + \cdots + X_n$ *therefore has CF*

$$\begin{aligned}\phi_{S_n}(z) &= \phi_{X_1+\cdots+X_n}(z)\\ &= (\phi(z))^n\\ &= (1-iz)^{-n}.\end{aligned}$$

Lemma 3.2. *If* $c \in \mathbb{R}$ *and* $Z \sim N(0,1)$*, then*

$$\mathbb{E}e^{cZ} = e^{c^2/2}. \tag{3.2}$$

Proof.

$$\begin{aligned}\mathbb{E}e^{cZ} &= \int_{-\infty}^{\infty} e^{cs}(2\pi)^{-1/2}e^{-s^2/2}\,ds\\ &= \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[(s-c)^2-c^2\right]}\,ds\\ &= e^{c^2/2}\int_{-\infty}^{\infty}(2\pi)^{-1/2}e^{-\frac{1}{2}(s-c)^2}\,ds\\ &= e^{c^2/2}.\end{aligned}$$

□

Theorem 3.3. *Lemma 3.2 is valid for any complex number* $c \in \mathbb{R}$*.*

Not examinable. The function $F(c) = \mathbb{E}\exp(cZ)$ is well-defined and analytic for all $c \in \mathbb{C}$, and $F(c) = \exp(c^2/2)$, for all $c \in \mathbb{R}$, by Lemma 3.2. Therefore the identity theorem for analytic functions implies that the theorem is true. □

Corollary 3.4. *If* $W \sim N(0,1)$*, then*

$$\phi_W(z) = e^{-z^2/2}.$$

Proof. Setting $c = iz$ is Lemma 3.2, we obtain

$$\begin{aligned}\phi_W(z) &= \mathbb{E}e^{izW}\\ &= e^{(iz)^2/2}\\ &= e^{-z^2/2}.\end{aligned}$$

□

Corollary 3.5. *If* $V \sim N(\mu,\sigma^2)$*, then*

$$\phi_V(z) = e^{i\mu z-\sigma^2 z^2/2}.$$

Proof. We write $V = \mu + \sigma W$, where $W \sim N(0,1)$. Then

$$\begin{aligned}\phi_V(z) &= \mathbb{E}e^{iz(\mu+\sigma W)}\\ &= e^{iz\mu}\mathbb{E}e^{iz\sigma W}\\ &= e^{iz\mu-\sigma^2 z^2/2},\end{aligned}$$

on setting $c = iz\sigma$ in Lemma 3.2. □

Corollary 3.6. *If $Z_k \sim N(\mu_k, \sigma_k^2)$, for $k = 1, 2, \ldots, n$, are independent Gaussian random variables, then their sum is also Gaussian. Specifically, their sum $S_n = Z_1 + \cdots + Z_n$ satisfies*

$$S_n \sim N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2).$$

Proof. The CF $\phi_k(z)$ is given by

$$\phi_k(z) = e^{i\mu_k z - \sigma_k^2 z^2/2}, \qquad 1 \le k \le n.$$

Hence the CF of S_n is given by

$$\begin{aligned}\phi_{S_n}(z) &= \phi_1(z)\phi_2(z)\cdots\phi_n(z) \\ &= e^{i(\mu_1+\cdots+\mu_n)-\frac{1}{2}z^2(\sigma_1^2+\cdots+\sigma_n^2)}.\end{aligned}$$

□

Example 3.3. *If $W \sim N(0,1)$, then we have already computed the PDF for $\chi_1^2 \equiv W^2$. However, it's easy to calculate its CF, as follows. We have*

$$\begin{aligned}\phi_{\chi_1^2}(z) &= \mathbb{E}e^{izW^2} \\ &= \int_{-\infty}^{\infty} e^{izs^2}(2\pi)^{-1/2}e^{-s^2/2}\,ds \\ &= (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{(-1/2)(-1-2iz)}\,ds \\ &= (1-2iz)^{-1/2}.\end{aligned}$$

Therefore the CF of χ_n^2 is given by

$$\phi_{\chi_n^2}(z) = (1-2iz)^{-n/2}.$$

We have see that the CF of a sum of independent random variables is simply the product of their CFs. It's easy to calculate the CF of cX, for any $c \in \mathbb{R}$. Specifically, we have

$$\phi_{cX}(z) = \mathbb{E}e^{iczX} = \phi_X(cz).$$

Corollary 3.7. *Let $X_1, X_2, \ldots, X_n$ be continuous, independent, identically distributed CFs with common CF $\phi(z)$. Then their average*

$$A_n = \frac{X_1 + \cdots + X_n}{n}$$

has CF

$$\phi_{A_n}(z) = \phi(z/n)^n.$$

Proof. Writing $A_n = S_n/n$, we obtain

$$\phi_{A_n}(z) = \phi_{S_n/n}(z) = \phi_{S_n}(z/n) = \phi(z/n)^n.$$

□

It's often useful to consider a slightly different average:

$$\widehat{A_n} = \frac{X_1 + \cdots + X_n}{\sqrt{n}}. \tag{3.3}$$

the idea here is that this normalization ensures that $\operatorname{var} A_n = \operatorname{var} X_k$, for all $k = 1, 2, \ldots, n$. I shall refer to (3.3) as the *variance-normalized average.*

Corollary 3.8. *Let $X_1, X_2, \dots, X_n$ be continuous, independent, identically distributed CFs with common CF $\phi(z)$. Then*

$$\widehat{A_n} = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

has CF

$$\phi_{\widehat{A_n}}(z) = \phi(z/\sqrt{n})^n.$$

Proof. Writing $\widehat{A_n} = S_n/n$, we obtain

$$\phi_{\widehat{A_n}}(z) = \phi_{S_n/\sqrt{n}}(z) = \phi_{S_n}(z/\sqrt{n}) = \phi(z/\sqrt{n})^n.$$

□

Exercise 3.1. *Let $X_k \sim N(0,1)$, for all k. Prove that $\widehat{A_n}$ is also $N(0,1)$, for all n.*

We're now very close to one form of the central limit theorem.

Theorem 3.9 (The Central Limit Theorem for CFs)**.** *Let $X_1, X_2, \dots$ be independent, identically distributed random variables satisfying $\mathbb{E}X_k = 0$ and $\mathbb{E}(X_k^2) = \sigma^2$, for all positive integer k. Then the CFs of the variance-normalized averages $\widehat{A_n}$ of* (3.3) *satisfy*

$$\lim_{n\to\infty} \phi_{\widehat{A_n}}(z) = e^{-z^2/2}. \tag{3.4}$$

In other words, the CFs of these variance-normalized averages converges pointwise to the CF for the normalized Gaussian.

Exercise 3.2. *A key assumption of the CLT is that the random variables must have finite variance. The CLT fails without this condition and, unsurprisingly, the counterexample is the Cauchy distribution. In this case, it can be shown that the CF corresponding to the PDF $p(s) = 1/(\pi(1+s^2))$, for $s \in \mathbb{R}$, is given by*

$$\phi(z) = e^{-|z|}, \qquad z \in \mathbb{R}.$$

Prove that $\phi_{A_n}(z) = \phi(z)$, for all positive integer n.

How do we recover the PDF from the CF?

Theorem 3.10. *Suppose the continuous PDF $p(s)$ has an* absolutely integrable *CF $\phi(z)$, that is,*

$$\int_{\infty}^{\infty} |\phi(z)|\, dz < \infty.$$

Then

$$p(x) = \frac{1}{2\pi} \int_{\infty}^{\infty} \phi(z) e^{-ixz}\, dz, \tag{3.5}$$

for all $x \in \mathbb{R}$.

4. The χ^2 distribution

Theorem 4.1. *Let $\mathbf{Z} \in \mathbb{R}^n$ be a normalized Gaussian random vector, that is, its components are independent $N(0,1)$ random variables. Then the probability density function $p_n(t)$ for the random variable $\|\mathbf{Z}\|^2 \equiv Z_1^2 + Z_2^2 + \cdots + Z_n^2$ is given by*

$$p_n(t) = \frac{e^{-t/2} t^{(n/2)-1}}{\int_0^\infty e^{-u/2} u^{(n/2)-1}\, du} = \frac{e^{-t/2} t^{(n/2)-1}}{2^{n/2}\Gamma(n/2)}, \qquad \text{for } t \geq 0. \tag{4.1}$$

[Of course, $p_n(t) = 0$ for $t < 0$, since the random variable $\|\mathbf{Z}\|^2$ is always non-negative.] We say that $\|\mathbf{Z}\|^2$ has the χ_n^2 distribution.

Proof. Let us define the annulus

$$\mathrm{Ann}(r_1, r_2) = \{\mathbf{s} \in \mathbb{R}^n : r_1 \leq \|\mathbf{s}\| \leq r_2\}.$$

We first observe that

$$\begin{aligned}
\mathbb{P}\left(a \leq \|\mathbf{Z}\|^2 \leq b\right) &= \int_{\mathrm{Ann}(\sqrt{a},\sqrt{b})} (2\pi)^{-n/2} e^{-\|\mathbf{s}\|^2/2}\, d\mathbf{s} \\
&= \mathbb{P}\left(\mathbf{Z} \in \mathrm{Ann}(\sqrt{a}, \sqrt{b})\right) \\
&= \int_{\sqrt{a}}^{\sqrt{b}} C_n e^{-r^2/2} r^{n-1}\, dr \\
&= \int_a^b D_n e^{-t/2} t^{(n/2)-1}\, dt,
\end{aligned}$$

where I have used spherical polar coordinates to obtain the second line, and the change of variable $t = r^2$ to obtain the third. To find the constant D_n, we set $a = 0$ and $b = \infty$, in which case we find

$$1 = \mathbb{P}\left(0 \leq \|Z\|^2\right) = D_n \int_0^\infty e^{-u/2} u^{(n/2)-1}\, du,$$

or

$$D_n = \frac{1}{\int_0^\infty e^{-u/2} u^{(n/2)-1}\, du}.$$

If we set $v = u/2$ in the last integral, then we find

$$D_n^{-1} = 2\int_0^\infty e^{-v} (2v)^{(n/2)-1}\, dv = 2^{n/2}\Gamma(n/2).$$

□

Karl Pearson (1857–1936) introduced the chi-squared test and the name for it in "On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling," Philosophical Magazine, 50, (1900), 157–175. Pearson used χ to denote $\|\mathbf{Z}\|$, and such is the inertia of human habit that the notation is preserved: we still say that $\|\mathbf{Z}\|$ has the χ_n^2 distribution. Incidentally, Pearson founded the world's first university department wholly dedicated to statistics at UCL, a short walk from our lecture room.

The modern view that the χ^2 distribution is one of a family of distributions linked by the Gamma function was pioneered by another UCL statistician, R. A. Fisher.

5. An Introduction to Extreme Value Theory

The theme of this section is estimating the probability that a random variable can be very far from its mean (an *extreme value*).

Theorem 5.1 (Chebyshev's Inequality)**.** *Let X be any random variable, with mean μ and variance σ^2. Then, for any positive δ,*

$$\mathbb{P}\left(|X-\mu| \geq \delta\right) \leq \frac{\sigma^2}{\delta^2}.$$

Proof. We have

$$\sigma^2 = \mathbb{E}(X-\mu)^2 \geq \delta^2 \mathbb{P}(|X-\mu| \geq \delta).$$

□

Chebyshev's inequality leads immediately to the *Weak Law of Large Numbers*:

Theorem 5.2. *Let $X_1, X_2, \ldots$ be a sequence of independent random variables for which $\mathbb{E}X_k = 0$ and $\mathbb{E}X_k^2 = \sigma^2$, for all $k \geq 1$. Let*

$$A_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Then, for any $\delta > 0$,

$$\mathbb{P}\left(|A_n| \geq \delta\right) \leq \frac{\sigma^2}{n\delta^2}.$$

Proof. We have $\mathbb{E}A_n = 0$ and $\mathbb{E}A_n^2 = n\sigma^2/n^2 = \sigma^2/n$. Hence, by Chebyshev's inequality,

$$\mathbb{P}(|A_n| \geq \delta) \leq \frac{\sigma^2}{n\delta^2}.$$

□

In particular, $\lim_{n\to\infty} \mathbb{P}(|A_n| \geq \delta) = 0$, for any $\delta > 0$. The assigned work last year and this year provides further examples of extreme value theory.

6. Interesting Problems

6.1. **The Birthday Problem.** This is a traditional probabilistic problem which I shall treat in further detail than is common. Given n people, whose birthdays are uniformly distributed over the 365 days of the year (ignoring leap years), find

$$\mathbb{P}\left(\text{at least 2 people share a birthday}\right). \tag{6.1}$$

It's easier to observe that

$$\mathbb{P}\left(\text{at least 2 people share a birthday}\right) = 1 - p_n, \tag{6.2}$$

where

$$p_n = \mathbb{P}\left(\text{ all } n \text{ people have different birthdays}\right). \tag{6.3}$$

Now

$$p_n = \frac{365 \cdot 364 \cdot 363 \cdots (365-n+1)}{365^n}$$

$$= \prod_{k=1}^{n-1} \left(1 - \frac{k}{365}\right). \tag{6.4}$$

In a sense, we have now finished the problem, since we can easily evaluate (6.4) for any n. However, displaying the values of $p_2, p_3, \ldots, p_{100}$, we see that the graph strongly resembles a Gaussian. Moreover, we can see that $p_{20} \approx 1/2$; to be more precise, $p_{23} = 0.51$ and $p_{35} = 0.99$. Can we understand these observations?

Taking logarithms in (6.4), we obtain

$$\ln p_n = \sum_{k=1}^{n-1} \ln\left(1 - \frac{k}{365}\right). \tag{6.5}$$

Now it can be shown that

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots, \tag{6.6}$$

and the series is convergent for $|x| < 1$. Thus, for $x > 0$, we have

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots \le x. \tag{6.7}$$

In other words, for small positive x, we have both the approximation $\ln(1-x) \approx -x$ and the inequality $ln(1-x) \le -x$. Applying these observations to (6.5), we obtain the approximation

$$\ln p_n \approx -\sum_{k=1}^{n-1} \frac{k}{365} = -\frac{n(n-1)}{730} \tag{6.8}$$

and the bound

$$\ln p_n \le -\frac{n(n-1)}{730}. \tag{6.9}$$

Taking exponentials, we have

$$p_n \approx e^{-n(n-1)/730} \tag{6.10}$$

and

$$p_n \le e^{-n(n-1)/730}. \tag{6.11}$$

School of Economics, Mathematics and Statistics, Birkbeck College, University of London, Malet Street, London WC1E 7HX, England

E-mail address: b.baxter@bbk.ac.uk