Green Trees Consulting Working Papers Series 2009

Professional Caregiver Insurance Risk
Health Care Finance & Practice Myths
Version 0.1

Thomas Cox PhD, MS, MSW, MSN, RN

Green Trees Consulting
1711 NW 55 Terrace
Gainesville, FL 32605
email: nurse.statistician at yahoo.com

June 16, 2009

Disclaimer

Green Trees Consulting and Thomas Cox make no warranty, expressed or implied, to users of these documents, and accept no responsibility for their use.

The mention of computer hardware, software, or other materials does not constitute endorsement by Green Trees Consulting or Thomas Cox, nor does it indicate that the products or materials are necessarily those best suited for the users intended purposes.

Contents

1 Probability theory for professional caregiver insurance risk **4**

1.1 Introduction . . . 4

1.2 Probability spaces and random variables . . . 5

1.3 Mean values of random variables . . . 6

1.4 Random vectors . . . 8

1.5 Marginal distributions of random variables . . . 8

1.6 Independent random variables . . . 9

1.7 Moments and variances of random variables . . . 9

1.8 Standard normal random variables . . . 10

1.9 Characteristic functions of random variables . . . 10

1.10 Characteristic functions of normal random variables . . . 11

1.11 Covariance and correlation of random variables . . . 12

1.12 Normal random vectors . . . 13

1.13 Convergence . . . 14

1.14 Continuity theorems . . . 15

1.15 The weak law of large numbers . . . 16

1.16 Classical central limit theorem . . . 17

1.17 The central limit theorem in health care finance . . . 17

Chapter 1

Probability theory for professional caregiver insurance risk

1.1 Introduction

The mathematical basis for professional caregiver insurance risk includes probability theory, real analysis, statistics, and finance. The core idea is that insurers manage risk well by writing large numbers of insurance policies. Very large insurers, assuming equally efficient operations, manage risk better than very small insurers.

This seems clear enough, and outside the realm of health insurance, managed care, and capitation, there is not an actuary in the world that would disagree with the idea that writing more policies reduces an insurer's portfolio risks, unless the risk assumption is inherently underpriced. Despite this, the most obvious violation of this defining characteristic of insurance is violated when entities that consumers think are insurers: Managed care organizations; Medicare; Medicaid; and many far more traditional insurers, engage in practices that transfer their insurance risks to health care providers.

The poster child for health insurance risk transfers is the global capitation contract. When a health care provider agrees to work under a capitation contract it accepts responsibility for a portfolio of insurance risks (The patients for whom the provider assumes responsibility). The provider becomes the *de facto* insurer, and the risk transferring entity eliminates the insurance risks for that portion of its portfolio.

If the risk transferring entity succeeds in writing capitation contracts with enough health care providers, it completely eliminates its role as an insurer. The insurance risks do not change at all. The only thing that changes is which entity bears them. Because the risk transferring entity contracts with many, many providers, each of the providers has a relatively small portion of the portfolio and is a very small, very inefficient, insurer. Ms. Jones is still going to develop breast cancer in October. Mr. Jones is still going to have knee surgery in March. Tiny Tim is still going to be hospitalized with pneumonia the day after Christmas. Every health event that was going to occur before the capitation contract is still going to occur after it. The only issue is who is responsible for the costs? Under global capitation the answer is clear. The risk assuming health care provider must cover the costs, not the entity that transferred the risks.

A full understanding of how insurance works is too involved for this paper. The book[1], explains, in depth, how insurance works, the economic and financial context of risk transfers, the ethical and legal issues involved, and what impact provider insurance risk assumption has on health care providers and consumer outcomes.

This document lays the ground work of a mathematical proof of concept. Herein, we explore the most

[1] $\sqrt{\frac{s^2}{n}}$: A Critique of Managed Care and Capitation Financed Health Care, (Copyright 2009, Thomas Cox PhD, RN)

fundamental concepts from mathematical probability theory leading up to the explanation for the greater efficiency of large insurers when sampling from the population of all possible policyholders. We shall see that it is the portfolio *standard error*, not the individual risk's *variance* or *standard deviation*, and the size of the portfolio, selected from an entirely different source, the sampling distibution of sample means, that helps us to understand insurance, risk management, and the flaws in managed care and capitation. Unlike the *variance* and *standard deviation*, whichare the same regardless of portfolio size, the risk (*standard error*) in a portfolio, grows arbitrarily smaller as the size of the insurer's portfolio increases.

As suggested above, insurers do not select individual risks. They select portfolios (collections of similar risks), of a given size. Risk assuming health care providers make very inefficient insurers because large portfolios produce more stable operations while small portfolios lead to wide fluctuations in an insurer's operating results. Insurers are, despite the growing number of failures, stability seeking entities. Two of the most risk seeking insurers failed: Reliance without a balance, and AIG despite massive infusions of capital from the federal government.

1.2 Probability spaces and random variables

A **probability space** is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and $\mathbb{P}$ is a σ-additive measure on $(\Omega, \mathcal{F})$ such that $\mathbb{P}(\Omega) = 1$. Measurable sets $A \in \mathcal{F}$ are called *events*.

A **random variable** is a measurable function $\xi \colon \Omega \to \mathbb{R}$, i.e., such that for an arbitrary Borel set $B \subset \mathbb{R}$ we have $\xi^{-1}(B) \in \mathcal{F}$.

The **probability distribution** of a random variable ξ is a Borel probability measure $\mathbb{P}_\xi$ on $\mathbb{R}$ defined by $\mathbb{P}_\xi(B) = \mathbb{P}\big(\xi^{-1}(B)\big)$ for every Borel set $B \subset \mathbb{R}$. We also call this the measure induced by ξ on $\mathbb{R}$.

The **distribution function** of a random variable ξ is defined by:

$$F_\xi(x) = \mathbb{P}(\xi \leq x) = \mathbb{P}\big(\xi^{-1}(-\infty, x]\big) = \mathbb{P}_\xi\big((-\infty, x]\big).$$

The following properties hold for all distribution functions:

- (a) F_ξ is monotonically increasing, i.e., $F_\xi(x) \geq F_\xi(y)$ if $x \geq y$
- (b) F_ξ is right-continuous at every point $x \in \mathbb{R}$
- (c) Its limit values are $F_\xi(\infty) = 1$ and $F_\xi(-\infty) = 0$.

NB: For any $a \in \mathbb{R}$ we have

$$\mathbb{P}(\xi = a) = \mathbb{P}_\xi\big(\{a\}\big) = F_\xi(a) - F_\xi(a - 0),$$

where $F_\xi(a-0)$ denotes the left-sided limit of F at a. In particular, we say that $\mathbb{P}(\xi = a) = 0$ if and only if (iff) $F(x)$ is continuous at a.

NB: If two random variables, ξ and η, coincide with probability one, i.e., $\mathbb{P}(\xi = \eta) = 1$, then these random variables have the same probability distribution, $\mathbb{P}_\xi = \mathbb{P}_\eta$, and having the same probability distribution they also have the same distribution function, $F_\xi = F_\eta$.

A **discrete random variable** is a random variable ξ whose range $\xi(\Omega)$ is a finite or, at most, a countably infinite set (can be put into a 1-1 correspondence with the set of positive integers or the set of all rational numbers. We denote by $x_1, x_2, \ldots$ the values of a discrete random variable ξ and by $p_i = \mathbb{P}(\xi = x_i)$ the corresponding probabilities. Clearly, $\sum_i p_i = 1$. Generalizing this we will say that ξ is a discrete random variable if it has a finite or countable (mod 0) range, i.e., there exists a finite or countable set $R \subset \mathbb{R}$ such that $\mathbb{P}\big(\xi^{-1}(R)\big) = 1$.

We say that the probability distribution of a discrete random variable ξ is an atomic measure, i.e., $\mathbb{P}_\xi = \sum_i p_i \delta_{x_i}$, where δ_x denotes the delta-function concentrated at the point x. Distribution functions of discrete random variables are step functions with discontinuities (jumps) of size p_i at each x_i.

A **continuous random variable** is a random variable ξ such that $\mathbb{P}(\xi = a) = 0$ for all $a \in \mathbb{R}$. A random variable is continuous iff its distribution function is continuous.

An **absolutely continuous random variable** is a random variable ξ the distribution function of which $\mathbb{P}_\xi$ is absolutely continuous with respect to its Lebesgue measure, i.e., $\mathbb{P}_\xi(B) = 0$ for all Borel sets $B \subset \mathbb{R}$ with Lebesgue measure zero. Equivalently, we say that there exists a function $f_\xi \colon \mathbb{R} \to \mathbb{R}$ (the **probability density** of ξ) such that

$$F_\xi(b) - F_\xi(a) = \int_a^b f_\xi(x)\,dx \qquad \forall a < b.$$

This equivalence follows from the Radon-Nikodym theorem.

Clearly,

$$f_\xi(x) = \frac{d}{dx}F_\xi(x)$$

at the Lebesgue-almost every point $x \in \mathbb{R}$.

NB: *The density function f_ξ is not unique, since we can arbitrarily change it on a set of Lebesgue measure zero. The density function is, however, unique (mod 0).*

Density functions have two properties:

$f_\xi(x) \geq 0$ (mod 0) and $\int_{-\infty}^{\infty} f_\xi(x)\,dx = 1$. Any function $f(x)$ with these properties defines a density function of some random variable. Trivially we can define the distribution function by $F(x) = \int_{-\infty}^{x} f(u)\,du$).

NB: *Absolutely continuous random variables are continuous, but continuous random variables need not be absolutely continuous random variables.*

A **singular continuous random variable** is a continuous random variable ξ such that $\frac{d}{dx}F_\xi(x) = 0$ at the Lebesgue-almost every point $x \in \mathbb{R}$. Example: a random variable ξ whose distribution function F_ξ is the Cantor function.

The Cantor function is a singular function defined on the Cantor set and defined as: Let x be a real number in $[0, 1]$ with expansion $0.a_1a_2a_3\ldots$, and let $N = \infty$ if no $a_n = 1$, else let N be the smallest value such that $a_n = 1$. Let $b_n = \frac{1}{2}a_n$ for all $n < N$ and let $b_N = 1$. We can now define the Cantor function (or the Cantor ternary function) as

$$f(x) = \sum_{n=1}^{N} \frac{b_n}{2^n}.$$

The Cantor function is continuous and monotonic on $[0, 1]$ and $f'(x) = 0$ almost everywhere, i.e.. it is constant on the complement of the Cantor set), with $f(0) = 0$ and $f(1) = 1$.

The **uniform random variable** on $(a, b) \subset \mathbb{R}$ is an absolutely continuous random variable with density function

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{elsewhere} \end{cases}$$

1.3 Mean values of random variables

We define the **mean value** (**expectation**) of a random variable ξ as

$$\mathbb{E}(\xi) = \int_\Omega \xi \, d\mathbb{P}.$$

The right hand side is the Lebesgue integral of the function ξ on Ω with respect to the measure $\mathbb{P}$. The mean value $\mathbb{E}(\xi)$ exists iff $\xi \in L^1(\Omega)$.

Clearly,

$$\mathbb{E}(\xi) = \int_\mathbb{R} x \, d\mathbb{P}_\xi. \tag{1.1}$$

For any Borel measurable function $g \colon \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}\big[g(\xi)\big] = \int_\Omega (g \circ \xi) \, d\mathbb{P} = \int_\mathbb{R} g \, d\mathbb{P}_\xi = \int_{-\infty}^{\infty} g(x) \, dF_\xi(x). \tag{1.2}$$

If ξ is absolutely continuous with density f_ξ, then

$$\mathbb{E}\big[g(\xi)\big] = \int_{-\infty}^{\infty} g(x) f_\xi(x) \, dx.$$

If ξ is a discrete random variable with values $x_1, x_2, \ldots$ and probabilities $p_1, p_2, \ldots$, then

$$\mathbb{E}(\xi) = \sum_i x_i p_i.$$

NB: $\mathbb{E}(\xi)$ *exists iff this series converges absolutely.*

The mean value has the same properties as the Lebesgue integral:

- If ξ is constant, i.e., $\mathbb{P}(\xi = a) = 1$ for some $a \in \mathbb{R}$, then $\mathbb{E}(\xi) = a$
- $\mathbb{E}(\xi + \eta) = \mathbb{E}(\xi) + \mathbb{E}(\eta)$
- $\mathbb{E}(c\xi) = c\mathbb{E}(\xi)$ for any constant $c \in \mathbb{R}$
- If $\mathbb{P}(a \le \xi \le b) = 1$, then $a \le \mathbb{E}(\xi) \le b$
- If $\xi \ge 0$ (mod 0) and $\mathbb{E}(\xi) = 0$, then $\xi = 0$ (mod 0)

A **random vector** is a measurable function $\mathbf{X} \colon \Omega \to \mathbb{R}^n$, $n \ge 2$. For every $\omega \in \Omega$ we have $\mathbf{X}(\omega) = \big(X_1(\omega), \ldots, X_n(\omega)\big)$, thus its components, $X_1, \ldots, X_n$, are just random variables.

The **probability distribution of** $\mathbf{X}$ is a probability measure $\mathbb{P}_\mathbf{X}$ on $\mathbb{R}^n$ defined by $\mathbb{P}_\mathbf{X}(B) = \mathbb{P}\big(\mathbf{X}^{-1}(B)\big)$ for every Borel set $B \subset \mathbb{R}^n$. It is also called the **joint probability distribution** of the variables $X_1, \ldots, X_n$.

The **distribution function of** $\mathbf{X}$ is defined by

$$F_\mathbf{X}(x_1, \ldots, x_n) = \mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n).$$

We shall also refer to this as the **joint distribution function** of the variables $X_1, \ldots, X_n$, denoted by $F_{X_1,\ldots,X_n}$.

1.4 Random vectors

An **absolutely continuous random vector** $\mathbf{X}$ is a random vector whose probability distribution $\mathbb{P}_{\mathbf{X}}$ is absolutely continuous with respect to the Lebesgue measure. Clearly, there exists a density function $f_{\mathbf{X}}(x_1, \ldots, x_n) \geq 0$ such that

$$\mathbb{P}(\mathbf{X} \in B) = \mathbb{P}_{\mathbf{X}}(B) = \int_B f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

for any Borel subset $B \subset \mathbb{R}^n$. Here $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $d\mathbf{x}$ is a shorthand notation for $dx_1\, dx_2 \cdots dx_n$. The function $f_{\mathbf{X}}(x_1, \ldots, x_n)$ is also called the **joint density function** of the variables $X_1, \ldots, X_n$ and denoted by $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$.

The **mean value of a random vector** $\mathbf{X}$ is a vector defined by

$$\mathbb{E}(\mathbf{X}) = \int_\Omega \mathbf{X}\, d\mathbb{P} = \big(\mathbb{E}(X_1), \ldots, \mathbb{E}(X_n)\big).$$

Generalizing (1.1) we obtain

$$\begin{aligned} \mathbb{E}(\mathbf{X}) &= \int_\Omega \mathbf{X}\, d\mathbb{P} = \int_{\mathbb{R}^n} (x_1, \ldots, x_n)\, d\mathbb{P}_{\mathbf{X}} \\ &= \left(\int_{\mathbb{R}^n} x_1 d\mathbb{P}_{\mathbf{X}}, \ldots, \int_{\mathbb{R}^n} x_n d\mathbb{P}_{\mathbf{X}} \right). \end{aligned} \tag{1.3}$$

Generalizing (1.2) we obtain for any Borel function $g \colon \mathbb{R}^n \to \mathbb{R}$

$$\mathbb{E}\big[g(\mathbf{X})\big] = \int_\Omega (g \circ \mathbf{X})\, d\mathbb{P} = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n)\, d\mathbb{P}_{\mathbf{X}}. \tag{1.4}$$

1.5 Marginal distributions of random variables

The **marginal distributions** of the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ are the probability distributions of its components $X_1, \ldots, X_n$, derived from the projection of the probability distribution $\mathbb{P}_{\mathbf{X}}$ on $\mathbb{R}^n$ onto the coordinate axes. That is, for every $i = 1, \ldots, n$, for any Borel set $B \subset \mathbb{R}$ on the ith coordinate axis, we have $\mathbb{P}_{X_i}(B) = \mathbb{P}_{\mathbf{X}}(B \times \mathbb{R}^{n-1})$; here $B \times \mathbb{R}^{n-1} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \colon x_i \in B\}$.

In general, for every subset $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ of indices, the k-dimensional marginal distribution is the distribution of the k-vector $(X_{i_1}, \ldots, X_{i_k})$, resulting from the projection of the probability distribution $\mathbb{P}_{\mathbf{X}}$ on $\mathbb{R}^n$ onto the k-dimensional subspace generated by the coordinate axes indexed by $i_1, \ldots, i_k$.

The **marginal density for a random vector** is characterized as follows. If a random vector $\mathbf{X}$ is absolutely continuous with density $f_{\mathbf{X}}(x_1, \ldots, x_n)$, then its components $X_1, \ldots, X_n$ are also absolutely continuous, and their densities (called **marginal densities**) are

$$f_{X_i}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n)\, dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_n, \tag{1.5}$$

.i.e.., when all other variables are integrated out.

Up to this point metric theory and probability theory are entirely consistent. Probability theory forks from measure theory when we introduce the definition of independence of sets (events).

1.6 Independent random variables

Two events $A, B \in \mathcal{F}$ are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

We can generalize this to n events, $A_1, \ldots, A_n$, and say that these events are independent if

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

for every subsequence of distinct indices $1 \le i_1 < i_2 < \cdots < i_k \le n$. We can also say this as, $A_1, \ldots, A_n$ are independent if

$$\mathbb{P}(A_1^{p_1} \cap \cdots \cap A_n^{p_n}) = \mathbb{P}(A_1^{p_1}) \cdots \mathbb{P}(A_n^{p_n})$$

for any combination $p_1, \ldots, p_n$, where each p_i is either "$+$" or "$-$", and A^+ stands for A while A^- stands for $A^c = \Omega \setminus A$.

Two random variables, ξ and η, are independent if for arbitrary Borel sets $A, B \subset \mathbb{R}$ the events $\xi^{-1}(A)$ and $\eta^{-1}(B)$ are independent. Several random variables $\xi_1, \ldots, \xi_n$ are independent if for any Borel subsets $A_1, \ldots, A_n \subset \mathbb{R}$ the events $\xi_1^{-1}(A_1), \ldots, \xi_n^{-1}(A_n)$ are independent. Random variables of an infinite sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ are independent if for any n the variables $\xi_1, \ldots, \xi_n$ are independent.

If $\xi_1, \ldots, \xi_n$ are independent random variables and $g_1, \ldots, g_n$ are arbitrary Borel functions (from $\mathbb{R}$ to $\mathbb{R}$), then the random variables $g_1(\xi_1), \ldots, g_n(\xi_n)$ are also independent.

If $X_1, \ldots, X_n$ are absolutely continuous independent random variables, then their joint density function satisfies

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \qquad \text{(mod 0)}.$$

If ξ and η are independent random variables, then

$$\mathbb{E}(\xi\eta) = \mathbb{E}(\xi) \cdot \mathbb{E}(\eta) \tag{1.6}$$

This formula extends to several independent random variables.

1.7 Moments and variances of random variables

The k^{th} **moment**, $M_k = \mathbb{E}(\xi^k)$, of the random variable ξ, where $k = 1, 2, \ldots$, is the expectation of the k^{th} power of the random variable.

The k^{th} **central moment**, The value $\mathbb{E}\Big(\big[\xi - \mathbb{E}(\xi)\big]^k\Big)$ is called the k-th central moment of ξ. The k-th moment and the k-th central moment exist iff $\xi \in L^k(\Omega)$.

The **variance** of a random variable ξ is its second central moment:

$$\begin{aligned}\mathsf{Var}(\xi) &= \mathbb{E}\Big(\big[\xi - \mathbb{E}(\xi)\big]^2\Big) \\ &= \mathbb{E}(\xi^2) - \big[\mathbb{E}(\xi)\big]^2.\end{aligned}$$

The **standard deviation**, $\sigma_\xi > 0$, of a random variable ξ is $\sigma_\xi = \sqrt{\mathsf{Var}(\xi)}$.

The variance and standard deviation have the following properties:
(a) $\mathsf{Var}(\xi) \ge 0$, and $\mathsf{Var}(\xi) = 0$ iff the random variable ξ is constant (mod 0), i.e., $\exists a \in \mathbb{R}\colon \mathbb{P}(\xi = a) = 1$.

(b) If $\eta = a + b\xi$, then $\mathsf{Var}(\eta) = b^2\mathsf{Var}(\xi)$ and $\sigma_\eta = |b|\sigma_\xi$.
(c) If ξ and η are independent, then

$$\mathsf{Var}(\xi + \eta) = \mathsf{Var}(\xi) + \mathsf{Var}(\eta).$$

The last property generalizes to any number of independent random variables.

If ξ is a random variable, then

$$\eta = \frac{\xi - \mathbb{E}(\xi)}{\sigma_\xi}$$

is called the **centered** and **normed** random variable obtained from ξ. It is a linear transformation of ξ such that $\mathbb{E}(\eta) = 0$ (.i.e..,, η is "centered") and $\mathsf{Var}(\eta) = 1$ (i.e., η is "normed").

A **normal random variable** is an absolutely continuous random variable ξ with density function

$$f_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Here $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are parameters. The normal random variable is denoted by $\mathcal{N}(\mu, \sigma^2)$. Note that $f_\xi(x) > 0$ for all $x \in \mathbb{R}$.

If ξ is $\mathcal{N}(\mu, \sigma^2)$ and a, b are real constants ($b \neq 0$), then $\eta = a + b\xi$ is also normal. More precisely, η is $\mathcal{N}(a + b\mu, b^2\sigma^2)$. Note that if ξ is $\mathcal{N}(\mu, \sigma^2)$, then $\frac{\xi-\mu}{\sigma}$, which is the centered and normed variable obtained from ξ, is a normal random variable $\mathcal{N}(0, 1)$. It plays a special role.

1.8 Standard normal random variables

A **standard normal random variable** is a normal random variable with $\mu = 0$ and $\sigma^2 = 1$, i.e., $\mathcal{N}(0, 1)$. It is customarily denoted by Z. Its density function is

$$f_Z(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}.$$

Its distribution function is customarily denoted by Φ, i.e.,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du.$$

NB: If Z is $\mathcal{N}(0, 1)$, then the random variables $\mu + \sigma Z$ and $\mu - \sigma Z$ are $\mathcal{N}(\mu, \sigma^2)$.

1.9 Characteristic functions of random variables

The **characteristic function** for a random variable ξ is defined to be

$$\varphi_\xi(t) = \mathbb{E}\big(e^{it\xi}\big) = \mathbb{E}\big(\cos(t\xi)\big) + i\,\mathbb{E}\big(\sin(t\xi)\big).$$

Here $t \in \mathbb{R}$ is the argument of the function, and $i = \sqrt{-1}$. This is a function from $\mathbb{R}$ to $\mathbb{C}$, it is well defined for *any* random variable and *any* $t \in \mathbb{R}$.

Characteristic functions have the following properties:

1. $\varphi_\xi(0) = 1$
2. $\varphi_\xi(-t) = \overline{\varphi_\xi(t)}$ $\qquad$ ($\overline{\varphi_\xi(t)}$ is the complex conjugate of $\varphi_\xi(t)$)

3. $|\varphi_\xi(t)| \le 1$ for all $t \in \mathbb{R}$

4. $\varphi_\xi(t)$ is uniformly continuous

5. If $\eta = a\xi + b$, then
$$\varphi_\eta(t) = \varphi_\xi(at) \cdot e^{ibt}$$

6. If ξ and η are independent, then
$$\varphi_{\xi\eta}(t) = \varphi_\xi(t)\,\varphi_\eta(t)$$

7. For any integer $k \ge 1$
$$\varphi_\xi^{(k)}(0) = i^k \mathbb{E}(\xi^k)$$

Hence, the derivatives of $\varphi_\xi(t)$ at $t = 0$ are related to the moments of the random variable ξ.

A function $\varphi\colon \mathbb{R} \to \mathbb{C}$ is said to be **positive definite** if for any $n \ge 1$ and any real numbers $t_1, \dots, t_n \in \mathbb{R}$ the $n \times n$ matrix with components $a_{ij} = \varphi(t_i - t_j)$ is Hermitian and positive semi-definite. Equivalently, for any complex numbers $z_1, \dots, z_n \in \mathbb{C}$ we have

$$\sum_{i,j=1}^{n} \varphi(t_i - t_j) z_i \bar{z}_j \ge 0. \tag{1.7}$$

The characteristic function $\varphi_\xi(t)$ is always positive definite. Indeed,

$$\begin{aligned}
\sum_{i,j=1}^{n} \varphi(t_i - t_j) z_i \bar{z}_j &= \sum_{i,j=1}^{n} \mathbb{E}\big[e^{i(t_i - t_j)\xi} z_i \bar{z}_j\big] \\
&= \mathbb{E}\bigg(\sum_{i,j=1}^{n} e^{it_i\xi} e^{-it_j\xi} z_i \bar{z}_j\bigg) \\
&= \mathbb{E}\bigg(\sum_{i=1}^{n} e^{it_i\xi} z_i\bigg)\bigg(\sum_{j=1}^{n} e^{-it_j\xi} \bar{z}_j\bigg) \\
&= \mathbb{E}\bigg(\bigg|\sum_{i=1}^{n} e^{it_i\xi} z_i\bigg|^2\bigg).
\end{aligned}$$

Bochner theorem. A function $\varphi\colon \mathbb{R} \to \mathbb{C}$ is a characteristic function of a random variable iff it satisfies three conditions: (a) $\varphi(0) = 1$, (b) $\varphi(t)$ is continuous, and (c) $\varphi(t)$ is positive definite.

The proof of this theorem is quite sophisticated, we derive a few simple facts from conditions (a)-(c).

1.10 Characteristic functions of normal random variables

The **characteristic function of the standard normal random variable** is

$$\varphi_Z(t) = e^{-\frac{t^2}{2}}. \tag{1.8}$$

The **characteristic function of normal random variable** $\xi = \mathcal{N}(\mu, \sigma^2)$ is

$$\varphi_\xi(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

This follows from (1.8) and property 5 above.

Inversion formula: for any $a < b$ we have

$$\mathbb{P}_\xi\big[(a,b)\big] + \tfrac{1}{2}\,\mathbb{P}_\xi\big(\{a\}\big) + \tfrac{1}{2}\,\mathbb{P}_\xi\big(\{b\}\big) = \lim_{A\to\infty} \frac{1}{2\pi}\int_{-A}^{A} \frac{e^{-ita}-e^{-itb}}{it}\,\varphi_\xi(t)\,dt.$$

In particular, if a and b are points of continuity of the distribution function $F_\xi(x)$, then

$$F_\xi(b) - F_\xi(a) = \lim_{A\to\infty} \frac{1}{2\pi}\int_{-A}^{A} \frac{e^{-ita}-e^{-itb}}{it}\,\varphi_\xi(t)\,dt. \tag{1.9}$$

Uniqueness theorem. The distribution function $F_\xi(x)$ is uniquely determined by the characteristic function $\varphi_\xi(t)$.

Proof. The continuity points of $F_\xi(x)$ are dense in $\mathbb{R}$, and $F_\xi(x)$ can be uniquely determined at all those points by (1.9). The rest of $F_\xi(x)$ can be reconstructed by its right-continuity. $\square$

The **characteristic function of a random vector X** is defined by

$$\varphi_{\mathbf{X}}(t_1,\dots,t_n) = \mathbb{E}\big[e^{i(t_1X_1+\cdots+t_nX_n)}\big] = \mathbb{E}\big[e^{i\mathbf{t}^T\mathbf{X}}\big].$$

where we use shorthand notation $\mathbf{t} = (t_1,\dots,t_n)$. Note that $\varphi_{\mathbf{X}}$ has n arguments, i.e., it is a function $\mathbb{R}^n \to \mathbb{C}$. It is also called the **joint characteristic function** of the random variables $X_1,\dots,X_n$.

If random variables $X_1,\dots,X_n$ are independent, then

$$\varphi_{X_1,\dots,X_n}(t_1,\dots,t_n) = \varphi_{X_1}(t_1)\cdots\varphi_{X_n}(t_n).$$

1.11 Covariance and correlation of random variables

The **covariance** of two random variables ξ and η is defined by

$$\begin{aligned}\mathsf{Cov}(\xi,\eta) &= \mathbb{E}\Big(\big[\xi-\mathbb{E}(\xi)\big]\cdot\big[\eta-\mathbb{E}(\eta)\big]\Big)\\ &= \mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\cdot\mathbb{E}(\eta).\end{aligned}$$

The covariance exists if both ξ and η are in $L^2(\Omega)$, by the Cauchy-Schwartz inequality.

The **correlation coefficient** for two random variables ξ and η is defined by

$$\rho(\xi,\eta) = \frac{\mathsf{Cov}(\xi,\eta)}{\sigma_\xi\,\sigma_\eta}.$$

Covariance and correlation have the following properties:
(a) $\mathsf{Cov}(\xi,\xi) = \mathsf{Var}(\xi)$;
(b) $\mathsf{Var}(\xi+\eta) = \mathsf{Var}(\xi) + \mathsf{Var}(\eta) + 2\,\mathsf{Cov}(\xi,\eta)$;
(c) If ξ and η are independent, then $\mathsf{Cov}(\xi,\eta) = 0$ and $\rho(\xi,\eta) = 0$; the converse is false;
(d) $|\rho(\xi,\eta)| \le 1$. Moreover, $\rho(\xi,\eta) = \pm 1$ iff ξ and η there is a linear relation between ξ and η, i.e., $\exists a,b\in\mathbb{R}\colon \mathbb{P}(\xi = a + b\eta) = 1$.

”d” follows from Cauchy-Schwarz inequality.

If $\mathsf{Cov}(\xi,\eta) = 0$, ξ,η are **uncorrelated**. This is a weaker characteristic than independence, but often used in cases where independence cannot be assumed to hold.

The **covariance matrix** of a random vector $\mathbf{X} = (X_1,\dots,X_n)$ is

$$\begin{aligned}\mathsf{Var}(\mathbf{X}) &= \mathbb{E}\Big(\big[\mathbf{X}-\mathbb{E}(\mathbf{X})\big]\,\big[\mathbf{X}-\mathbb{E}(\mathbf{X})\big]^T\Big)\\ &= \mathbb{E}(\mathbf{X}\mathbf{X}^T) - \mathbb{E}(\mathbf{X})\,\mathbb{E}(\mathbf{X})^T.\end{aligned}$$

We assume vectors $\mathbf{X}$ and $\mathbb{E}(\mathbf{X})$ are column-vectors, and use **transposes** when we want to define row vectors, so that $\mathbf{X}^T$ and $\mathbb{E}(\mathbf{X})^T$, are row-vectors. Vector products, such as $\mathbf{X}\mathbf{X}^T$ and $\mathbb{E}(\mathbf{X})\,\mathbb{E}(\mathbf{X})^T$, are $n \times n$ matrices. Thus $\mathsf{Var}(\mathbf{X})$ is an $n \times n$ matrix.

The components of the matrix $\mathsf{Var}(\mathbf{X})$ are $\mathsf{Cov}(X_i, X_j)$. In particular, the **diagonal components** are $\mathsf{Var}(X_i)$. Clearly, $\mathsf{Var}(\mathbf{X})$ is a **symmetric matrix**, and it is not hard to show that it is **positive-semidefinite**. Moreover, it is positive-definite unless the vector $\mathbf{X}$ is degenerate. The latter means that some linear combination of its components X_i is a constant (mod 0), i.e.,

$$\mathbb{P}(a_1X_1 + \cdots + a_nX_n = a_0) = 1$$

for some constants $a_0, a_1, \ldots, a_n$ (such that $a_1^2 + \cdots + a_n^2 > 0$). In the latter case the values of $\mathbf{X}$ belong to a hyperplane with probability one, i.e., the range of $\mathbf{X}$ is essentially an $(n-1)$-dimensional space.

We can generalize the properties of the expectation and the variance to n dimensions. Let $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, where $\mathbf{A}$ is an $n \times n$ matrix and $\mathbf{b} \in \mathbb{R}^n$ is a vector ($\mathbf{A}$ and $\mathbf{b}$ are constant, i.e., not random). Then

$$\mathbb{E}(\mathbf{Y}) = \mathbf{A}\,\mathbb{E}(\mathbf{X}) + \mathbf{b}$$

and

$$\mathsf{Var}(\mathbf{Y}) = \mathbf{A}\,\mathsf{Var}(\mathbf{X})\,\mathbf{A}^T.$$

As well, if $\mathbf{X}$ is an absolutely continuous random vector with density $f_{\mathbf{X}}$ and $\mathbf{A}$ is non-singular, then $\mathbf{Y}$ is also absolutely continuous and has density

$$f_{\mathbf{Y}}(\mathbf{x}) = \frac{1}{|\det \mathbf{A}|} f_{\mathbf{X}}\big(\mathbf{A}^{-1}(\mathbf{x} - \mathbf{b})\big)$$

and we say $\mathbf{x} = (x_1, \ldots, x_n)$. This formula depends on the change of variables in multivariate calculus, and $\frac{1}{|\det \mathbf{A}|}$ is the Jacobian.

The characteristic function of $\mathbf{Y}$ is given by

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = \varphi_{\mathbf{X}}(\mathbf{A}^T\mathbf{t}) \cdot e^{i\mathbf{b}^T\mathbf{t}}. \tag{1.10}$$

As before, we say $\mathbf{t} = (t_1, \ldots, t_n)$.

1.12 Normal random vectors

Normal vectors. Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ be a random vector such that $Z_1, \ldots, Z_n$ are independent, standardized normal random variables. Then the density of this vector is

$$\begin{aligned} f_{\mathbf{Z}}(x_1, \ldots, x_n) &= \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\,(x_1^2 + \cdots + x_n^2)} \\ &= \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\,\mathbf{x}^T\mathbf{x}} \end{aligned}$$

and its characteristic function is

$$\begin{aligned} \varphi_{\mathbf{Z}}(t_1, \ldots, t_n) &= e^{-\frac{1}{2}\,(t_1^2 + \cdots + t_n^2)} \\ &= e^{-\frac{1}{2}\,\mathbf{t}^T\mathbf{t}}. \end{aligned}$$

As well, $\mathbf{t} = (t_1, \ldots, t_n)$. We note that $\mathbb{E}(\mathbf{Z}) = \mathbf{0}$, is the zero vector, and $\mathsf{Var}(\mathbf{Z}) = \mathbf{I}$, is the identity matrix.

If $\mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{b}$ with $\det \mathbf{A} \neq 0$, its density function is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{A}\mathbf{A}^T}}\; e^{-\frac{1}{2}\,(\mathbf{x}-\mathbf{b})^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{x}-\mathbf{b})}$$

and its characteristic function is

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{b}^T\mathbf{t} - \frac{1}{2}\,\mathbf{t}^T\mathbf{A}\mathbf{A}^T\mathbf{t}},$$

which follows from the above as a result of (1.10). We also note that

$$\mathbb{E}(\mathbf{X}) = \mathbf{b} \qquad \text{and} \qquad \mathsf{Var}(\mathbf{X}) = \mathbf{A}\mathbf{A}^T,$$

making the latter a positive definite matrix.

A random vector $\mathbf{X}$ is **normal** if there is a vector $\mathbf{a} \in \mathbb{R}^n$ and a symmetric positive-definite matrix $\mathbf{V}$ such that the density of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{V}}}\; e^{-\frac{1}{2}\,(\mathbf{x}-\mathbf{a})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{a})}. \tag{1.11}$$

If this is true, its characteristic function is

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{a}^T\mathbf{t} - \frac{1}{2}\,\mathbf{t}^T\mathbf{V}\mathbf{t}}.$$

We refer to this the normal random vector $\mathbf{X}$ as $\mathcal{N}(\mathbf{a}, \mathbf{V})$. Note that

$$\mathbb{E}(\mathbf{X}) = \mathbf{a} \qquad \text{and} \qquad \mathsf{Var}(\mathbf{X}) = \mathbf{V}.$$

NB: If $\mathbf{X}$ is a normal vector, then each of its components $X_1, \ldots, X_n$ are, themselves, normal random variables. This can be verified by direct integration, according to (1.5). In that case we also say that the components $X_1, \ldots, X_n$ of the vector $\mathbf{X}$ are **jointly normal**, i.e., have a **joint normal distribution**.

We note that it is not true that any n normal random variables $X_1, \ldots, X_n$ are jointly normal.

Let $\mathbf{X}$ be a normal vector $\mathcal{N}(\mathbf{a}, \mathbf{V})$ and $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ be its linear transformation. Then $\mathbf{Y}$ is also a normal random vector $\mathcal{N}(\mathbf{A}\mathbf{a} + \mathbf{b}, \mathbf{A}\mathbf{V}\mathbf{A}^T)$.

We note that the components $X_1, \ldots, X_n$ of a normal vector $\mathbf{X}$ are independent iff the matrix $\mathbf{V}$ is diagonal, i.e., $\mathbf{Cov}(X_i, X_j) = 0$ for all $i \neq j$. So, in the case of (jointly) normal random variables independence is *equivalent* to the lack of correlation.

The last property holds only for normal vectors. That is, let $\mathbf{X} = (X_1, X_2)$ be a random vector with independent components X_1 and X_2, and $\mathbf{Q}$ be a 2×2 orthogonal matrix with non-zero entries. If the components of $\mathbf{Y} = \mathbf{Q}\mathbf{A}$ are also independent, then $\mathbf{X}$ and $\mathbf{Y}$ are normal vectors. (Feller, Volume 2, Section III.4.)

1.13 Convergence

Weak convergence. We say that a sequence of distribution functions, $\{F_n\}$, weakly converges to a distribution function F, if $F_n(x) \to F(x)$ at every point $x \in \mathbb{R}$ where the limit function $F(x)$ is continuous. This is equivalent to the weak convergence of the sequence of probability distributions $\mathbb{P}_{F_n}$ corresponding to F_n to the probability distribution $\mathbb{P}_F$ corresponding to F. The latter means that

$$\int_{\mathbb{R}} g\, d\mathbb{P}_{F_n} \to \int_{\mathbb{R}} g\, d\mathbb{P}_F \tag{1.12}$$

for every bounded continuous function $g\colon \mathbb{R} \to \mathbb{R}$.

Convergence in distribution (in law). We say that a sequence of random variables ξ_n converges in distribution (in law) to a random variable ξ iff their distribution functions F_{ξ_n} weakly converge to F_ξ. Equivalently,

$$\int_{\mathbb{R}} g\, d\mathbb{P}_{\xi_n} \to \int_{\mathbb{R}} g\, d\mathbb{P}_\xi \tag{1.13}$$

for every bounded continuous function $g\colon \mathbb{R} \to \mathbb{R}$.

Different types of convergence. As in real analysis, we say that a sequence of random variables ξ_n converges to a random variable ξ
(a) **pointwise** if $\xi_n(\omega) \to \xi(\omega)$ for every $\omega \in \Omega$
(b) **almost surely** if $\mathbb{P}(\xi_n \to \xi) = 1$
(c) **in probability** if $\mathbb{P}(|\xi_n - \xi| > \varepsilon) \to 0$ for every $\varepsilon > 0$.

1.14 Continuity theorems

The **continuity theorem** consists of two parts, the **direct case** and the **converse case**.

The direct **continuity theorem:** If a sequence of distribution functions F_n weakly converges to a distribution function F, then the corresponding characteristic functions converge pointwise, i.e., $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$. Equivalently, if a a sequence of random variables ξ_n converges in distribution (in law) to a random variable ξ, then $\varphi_{\xi_n}(t) \to \varphi_\xi(t)$ for all $t \in \mathbb{R}$.

The converse **continuity theorem:** If a sequence of characteristic functions $\varphi_n(t)$ converges pointwise to a continuous function $\varphi(t)$, then $\varphi(t)$ is a characteristic function, too. Furthermore, the corresponding distribution functions (which are uniquely defined by the uniqueness theorem) converge weakly.

The **proof** of the direct continuity theorem follows immediately from the definition of the characteristic function and (1.12)–(1.13).

A **proof** of the converse continuity theorem can be outlined as follows. let F_n be the distribution functions corresponding to φ_n. By using a standard diagonal procedure, we can find a subsequence F_{n_k} such that $F_{n_k}(x)$ converges at every rational point x. Hence, the limit values define monotonically increasing functions on the set of rationals. Since the rationals are dense in $\mathbb{R}$, we can define a limit function $F(x)$ on $\mathbb{R}$ by continuity. Obviously, $F(x)$ is monotonically increasing, and we can always redefine it to be right continuous. Clearly, $F_{n_k}(x)$ converges to $F(x)$ at every point x where x is continuous. It is also clear that $F(\infty) \leq 1$ and $F(-\infty) \geq 0$. However, $F(x)$ need not be a distribution function, so we need to prove that $F(\infty) = 1$ and $F(-\infty) = 0$.

Assume first that this is the case, i.e., $F(\infty) = 1$ and $F(-\infty) = 0$. F is a distribution function, and a weak limit for $\{F_{n_k}\}$. By part A $\varphi(t)$ is the characteristic function corresponding to $F(x)$. So, we get part B for the subsequence $\{F_{n_k}\}$. Finally, if the entire sequence $\{F_n\}$ does not weakly converge to F, we can find another limit function F' of the sequence $\{F_n\}$ (a weak limit of another subsequence of $\{F_n\}$), and by using the above argument φ also corresponds to F'. This contradicts the uniqueness theorem. Hence, $\{F_n\}$ weakly converges to F.

We must still prove that $F(\infty) = 1$ and $F(-\infty) = 0$. Clearly, no significant part of the measure ν_{F_n} goes to infinity as n grows (a form of compactness). We can see that $\varphi_n(0) = 1$, and hence $\varphi(0) = 1$.

Due to the continuity of φ, we have $\varphi(t) \approx 1$ on a small interval $(-\tau, \tau)$. Therefore,

$$\frac{1}{2\tau} \int_{-\tau}^{\tau} \varphi(t)\, dt \approx 1$$

Since the functions φ_n are uniformly bounded (by 1) and converge to φ pointwise, then

$$\frac{1}{2\tau} \int_{-\tau}^{\tau} \varphi_n(t)\, dt \approx 1$$

on $(-\tau, \tau)$ for sufficiently large n. We can force the integral arbitrarily close to 1 by choosing smaller values of τ. Now, by the definition of $\varphi_n(t)$ we get

$$\frac{1}{2\tau} \int_{-\infty}^{\infty} \left(\int_{-\tau}^{\tau} e^{itx}\, dt \right) dF_n(x) \approx 1$$

Interchanging integrals is justified since the function is uniformly bounded. We can compute the inner integral directly as $\frac{2}{x}\sin\tau x$. Hence,

$$\int_{-\infty}^{\infty} \frac{\sin\tau x}{\tau x}\, dF_n(x) \approx 1$$

We note that the integrand is bounded by 1, and for large x it is bounded by $\frac{1}{\tau x}$. The above integral captures the mass of the measure ν_{F_n} on a large, though finite, interval dependent on τ not n.

1.15 The weak law of large numbers

The **Weak law of large numbers**. Let $\{\xi_n\}$ be a sequence of independent, identically distributed (iid) random variables with common distribution functions $F(x)$. Denote by $S_n = \xi_1 + \cdots + \xi_n$ partial sums of this sequence.

We assume that $\mathbb{E}(\xi_n) = 0$, if not, we can always transform the variables so that this is true. The sequence $\{\xi_n\}$ satisfies the weak law of large numbers if $\forall\varepsilon > 0$

$$\mathbb{P}(|S_n/n| > \varepsilon) \to 0 \qquad \text{as} \quad n \to \infty \tag{1.14}$$

Denote by $G_n(x)$ the distribution function of the random variable S_n/n. The above weak law of large numbers is equivalent to

$$\lim_{n\to\infty} G_n(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases}$$

In other words, the distribution function $G_n(x)$ weakly converges to the function $F_\infty(x) = 1$ for $x > 0$ and 0 for $x \le 0$. The latter has the characteristic function $\varphi_\infty \equiv 1$.

To prove the weak law of large numbers, let $\varphi(t)$ be the characteristic function corresponding to $F(x)$. Since $\mathbb{E}(\xi_n) = 0$, we have $\varphi'(0) = 0$ (see Property 6). Hence, $\varphi(t) = 1 + o(t)$ for small t. Let $\psi_n(t)$ denote the characteristic function corresponding to $G_n(x)$. By Properties 5 and 4, we have

$$\psi_n(t) = [\varphi(t/n)]^n$$

Taking logarithm gives

$$\begin{aligned} \ln\psi_n(t) &= n\ln\varphi(t/n) \\ &= n\ln(1 + o(t/n)) \\ &= n \cdot o(t/n) \end{aligned}$$

Hence, $\psi_n(t) \to 1$ for every $t \in \mathbb{R}$. The continuity theorem (part B) completes the proof of the weak law of large numbers.

Remark. The weak law of large numbers holds when $\varphi'(0) = 0$, which is less restrictive than $\mathbb{E}(\xi_n) = 0$ (it may still hold when $\mathbb{E}(\xi_n)$ does not exist). However, the condition $\varphi'(0) = 0$ is a *necessary* condition for the weak law of large numbers.

The weak law of large numbers for a sequence of iid variables $\{\xi_n\}$ such that $\mu\colon = \mathbb{E}(\xi_n) < \infty$ reads $\forall\varepsilon > 0$

$$\mathbb{P}(|S_n/n - \mu| > \varepsilon) \to 0 \qquad \text{as} \quad n \to \infty$$

an immediate generalization of (1.14).

1.16 Classical central limit theorem

Let $\{\xi_n\}$ be a sequence of iid random variables with a common distribution function $F(x)$. Assume that $\mathbb{E}(\xi_n) = 0$ and $\mathbb{E}(\xi_n^2) = 1$. Denote by G_n the distribution function of the random variable $S_n/\sqrt{n}$, where $S_n = \xi_1 + \cdots + \xi_n$.

We say that $\{\xi_n\}$ satisfies the central limit theorem if $\forall x \in \mathbb{R}$ **Key Result**

$$G_n(x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du \tag{1.15}$$

as $n \to \infty$, i.e., the random variable $S_n/\sqrt{n}$ weakly converges to the standard normal random variable.

To prove the central limit theorem, denote by $\varphi(t)$ the characteristic function corresponding to $F(x)$. Since $\mathbb{E}(\xi_n) = 0$ and $\mathbb{E}(\xi_n^2) = 1$, we have $\varphi'(0) = 0$ and $\varphi''(0) = -1$ (see Property 6). Hence,

$$\varphi(t) = 1 - \frac{1}{2}t^2 + o(t^2)$$

for small t (Taylor expansion). Let $\psi_n(t)$ denote the characteristic function corresponding to $G_n(x)$. By Properties 5 and 4, we have

$$\psi_n(t) = \left[\varphi(t/\sqrt{n})\right]^n$$

Taking logarithms yields

$$\begin{aligned}
\ln \psi_n(t) &= n \ln \varphi(t/\sqrt{n}) \\
&= n \ln \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right) \\
&= -n\frac{t^2}{2n} + n \cdot o\left(\frac{t^2}{n}\right) \\
&\to -\frac{t^2}{2} \qquad \text{as} \quad n \to \infty
\end{aligned}$$

Hence, $\psi_n(t) \to e^{-t^2/2}$ for every $t \in \mathbb{R}$. The continuity theorem (part B) completes the proof of the central limit theorem. □

Returning to the classical central limit theorem, we now let $\{\xi_n\}$ be a sequence of iid random variables with a common distribution function $F(x)$. We assume that $\mathbb{E}(\xi_n) = \mu$ and $\mathbb{E}(\xi_n^2) = \sigma^2$ exist. Let G_n be the distribution function of the random variable **Key Result**

$$\eta_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

where $S_n = \xi_1 + \cdots + \xi_n$. Note that η_n is obtained from S_n by centering and norming. Now the central limit theorem says that $\forall x \in \mathbb{R}$

$$G_n(x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du \tag{1.16}$$

as $n \to \infty$, i.e., the random variable η_n weakly converges to the standard normal random variable.

1.17 The central limit theorem in health care finance

The meaning of the central limit theorem is that the partial sum S_n of a sequence of iid random variables deviates from its mean value $\mathbb{E}(S_n) = n\mu$ by a *random* amount, the size of which varies with the $\sqrt{n}$. This is the only right asymptotics of deviations, and if one normalizes the deviations by $\sqrt{n}$ then the deviations will have a well defined limit distribution. In insurance, we are typically working with a set of policyholders (insureds) who we can, for the sake of ease assume to be identically distributed before a claim is made, and we are interested in the size of the average loss S_n/n on such a set of policies. This portfolio average approaches μ for large n (by the weak law of large numbers) and the deviations from μ (errors between population loss ratio and portfolio loss ratio) are of order $1/\sqrt{n}$.

As a result, we can see that if we are comparing the magnitude of the random errors in two insurance portfolios, we would expect the ratio of the deviations from the true loss ratio for the population to vary by the square root of the ratio of the portfolio sizes. In particular, if we assume that the risk load (the adjustment to the pure premium to account for variation from expected losses) for providing insurance is α_{n_1} for Insurer A, with portfolio size of n_A, and that the risk load for providing insurance is α_{n_B} for Insurer B, with portfolio size of n_B, the risk load for insurer B can be expressed in terms of the risk load for insurer A, as:

$$(\text{Risk Load})_B = \sqrt{\frac{n_B}{n_A}} \times (\text{Risk Load})_A \tag{1.17}$$

In short, whenever insurance risks are transferred from large, efficient insurers, to smaller, less efficient insurers, the loss in efficiency is a function of the square root of the ratio of the portfolio sizes. When small health insurance risk portfolios are transferred from large insurers, to health care providers, the providers face substantially higher probabilities of large profits and large losses. While larger than expected profits are desirable, larger than expected losses clearly are not. In fact, larger than expected losses, especially for health care providers, can lead to financial failure, compromised care, and bankruptcy.

Health care providers, as insurers, must protect themselves from higher than expected losses. To protect themselves against the possibility that their contract costs will be higher than expected, health care providers must target their average costs significantly below the costs anticipated in the risk transfer premiums they will receive. The loss in actual consumer benefit is the difference between the provider's revenue and the provision it makes to manage the risk of higher than expected costs. This reduction in service capacity is due to the provider's inefficiency as an insurer. Once the contract ends, the provider will know what their actual costs were, but it is too late to go back and correct service shortfalls that may have occurred months earlier.

Patients of risk assuming health care providers must receive fewer services and/or lower quality services, than patients treated by equally clinically efficient health care providers who do not accept insurance risks. There are only two mechanisms that health care providers can employ to reduce their exposure to risk: Reduce the average value of services they will provide, denying all clients some level of service, or reducing the variability in their costs of providing services. This latter approach means that high cost services, the services health care providers need to provide to clients who need the greatest amount of care, must be reduced.

While denying needed services to a seriously ill or injured person saves large amounts of money, it is precisely the rare clients of insurers and providers that need their insurance benefits. It is easy to marginalize rare, high cost clients. But withholding insurance benefits for the rare people who require them defeats the reason consumers buy insurance.

An insurer or health care provider who routinely denies the highest cost, perfectly legitimate claims, for high cost care (say the highest 0.1% of claimants/patients, may still have excellent satisfaction surveys for 99.9% of its policyholders/patients. However, satisfying the top 0.1% of costliest claimants/patients is far more important in insurance, and health care, than satisfying the other 99.9%.