Intermediate Probability: A Computational Approach

Marc S. Paolella

Swiss Banking Institute, University of Zürich

Intermediate Probability: A Computational Approach

Intermediate Probability: A Computational Approach

The Moment Generating Function

- The **moment generating function**, abbreviated **m.g.f.**, of random variable X is the function $\mathbb{M}_X \colon \mathbb{R} \mapsto \mathbb{X}_{\geq 0}$ given by $t \mapsto \mathbb{E}\left[e^{tX}\right]$. It is a **function of real number t**.
- We will see below that $\mathbb{M}'_X(t)\big|_{t=0} = \mathbb{E}[X]$.
- As a simple example from the text, let $X \sim \mathrm{DUnif}(\theta)$ with p.m.f. $f_X(x;\theta) = \theta^{-1}\mathbb{I}_{\{1,2,\ldots,\theta\}}(x)$. The m.g.f. of X is

$$\mathbb{M}_X(t) = \mathbb{E}\left[e^{tX}\right] = \frac{1}{\theta}\sum_{j=1}^{\theta} e^{tj},$$

and

$$\mathbb{M}'_X(t) = \frac{1}{\theta}\sum_{j=1}^{\theta} j e^{tj}, \qquad \mathbb{E}[X] = \mathbb{M}'_X(0) = \frac{1}{\theta}\sum_{j=1}^{\theta} j = \frac{\theta+1}{2}.$$

The Moment Generating Function (2)

- Again: The **moment generating function**, abbreviated **m.g.f.**, of random variable X is the function $\mathbb{M}_X \colon \mathbb{R} \mapsto \mathbb{X}_{\geq 0}$ given by $t \mapsto \mathbb{E}\left[e^{tX}\right]$. It is a **function of real number** t.
- The m.g.f. $\mathbb{M}_X$ **exists** if it is finite on a neighborhood of zero, i.e., if there is an $h > 0$ such that, $\forall t \in (-h, h)$, $\mathbb{M}_X(t) < \infty$.
- If $\mathbb{M}_X$ exists, then the largest (open) interval $\mathcal{I}$ around zero such that $\mathbb{M}_X(t) < \infty$ for $t \in \mathcal{I}$ is referred to as the **convergence strip (of the m.g.f.) of** X.
- If $\mathbb{M}_X$ exists, then all positive absolute moments of X exist.

If $\mathbb{M}_X$ exists, then $\forall r \in \mathbb{R}_{>0}$, $\mathbb{E}\left[|X|^r\right] < \infty$.

The Moment Generating Function (3)

- If the m.g.f. exists, we can exchange infinite sum and expectation:

$$\mathbb{M}_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbb{E}\left[X^k\right].$$

- We can also write

$$\mathbb{M}_X^{(j)}(t) = \frac{d^j}{dt^j}\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\frac{d^j}{dt^j}e^{tX}\right] = \mathbb{E}\left[X^j e^{tX}\right]$$

to get:

$$\mathbb{M}_X^{(j)}(t)\Big|_{t=0} = \mathbb{E}\left[X^j\right].$$

- In general, if $\mathbb{M}_Z(t)$ is the m.g.f. of Z and $X = \mu + \sigma Z$, then

$$\mathbb{M}_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t(\mu+\sigma Z)}\right] = e^{t\mu}\mathbb{M}_Z(t\sigma).$$

Exercise (Example 1.2 from the text)

Let $U \sim \mathrm{Unif}\,(0,1)$.

- Show that

$$\mathbb{M}_U(t) = \frac{e^t - 1}{t}, \quad t \neq 0.$$

- Verify that $\mathbb{M}_U(t)$ is continuous at zero.
- From a Taylor series expansion around zero, show that

$$\mathbb{M}_U(t) = \sum_{j=0}^{\infty} \frac{1}{j+1} \frac{t^j}{j!}$$

  and thus that

$$\mathbb{E}\left[U^r\right] = (r+1)^{-1}, r = 1, 2, \ldots.$$

- For $r = 1$ and $r = 2$, verify the previous formula by direct calculation of the moments.

The Cumulant Generating Function

- The **c.g.f.** is defined as $\mathbb{K}_X(t) = \log \mathbb{M}_X(t)$.
- The terms κ_i in the series expansion $\mathbb{K}_X(t) = \sum_{r=0}^{\infty} \kappa_r \frac{t^r}{r!}$ are referred to as the **cumulants** of X, so that the ith derivative of $\mathbb{K}_X(t)$ evaluated at $t = 0$ is κ_i, i.e.,

$$\kappa_i = \mathbb{K}_X^{(i)}(t)\Big|_{t=0} .$$

- It is straightforward to show that

$$\kappa_1 = \mu, \qquad \kappa_2 = \mu_2, \qquad \kappa_3 = \mu_3 \qquad \text{and} \qquad \kappa_4 = \mu_4 - 3\mu_2^2.$$

Exercise

Do so for the first three cumulants.

Solution

- Define $j_i = j_i(s) = \mathbb{E}\left[X^i e^{sX}\right]$ for $i \geq 0$ and observe that $j_i\,\big|_{s=0} = \mathbb{E}\left[X^i\right] = \mu_i'$, where $\mu_0' := 1$ and $\mathrm{d}j_i / \mathrm{d}s = j_{i+1}$.
- Then
$$\frac{\mathrm{d}\mathbb{K}_X(s)}{\mathrm{d}s} = \frac{1}{\mathbb{E}\left[e^{sX}\right]} \frac{\mathrm{d}}{\mathrm{d}s} \mathbb{E}\left[e^{sX}\right] = \frac{j_1}{j_0},$$
which, evaluated at $s = 0$ is $\mu_1' = \mu = \mathbb{E}[X]$.
- Next,
$$\frac{\mathrm{d}^2\mathbb{K}_X(s)}{\mathrm{d}s^2} = \frac{\mathrm{d}}{\mathrm{d}s} \frac{j_1}{j_0} = \frac{j_0 j_2 - j_1^2}{j_0^2},$$
which, at $s = 0$, is $\mu_2' - \mu^2 = \mu_2 = \mathbb{V}(X)$.
- Similarly,
$$\frac{\mathrm{d}^3\mathbb{K}_X(s)}{\mathrm{d}s^3} = \frac{j_0^2 \left(j_0 j_3 + j_2 j_1 - 2 j_1 j_2\right) - \left(j_0 j_2 - j_1^2\right) 2 j_0 j_1}{j_0^4},$$
which, at $s = 0$, simplifies to $\mu_3' - 3\mu\mu_2' + 2\mu^3 = \mu_3$.

Example: The m.g.f. and c.g.f. of the Normal

- The m.g.f. of $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$ is shown below to be given by

$$\mathbb{M}_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}, \quad \mathbb{K}_X(t) = \mu t + \frac{1}{2}\sigma^2 t^2,$$

so

$$\begin{aligned} \mathbb{K}_X'(t) &= \mu + \sigma^2 t, \quad \mathbb{E}\left[X\right] = \mathbb{K}_X'(0) = \mu, \\ \mathbb{K}_X''(t) &= \sigma^2, \quad \mathbb{V}\left(X\right) = \mathbb{K}_X''(0) = \sigma^2 \quad \text{and} \\ \mathbb{K}_X^{(i)}(t) &= 0, \quad i \geq 3. \end{aligned}$$

- Thus,

$$\mu_3 = 0 \quad \text{and} \quad \mu_4 = \kappa_4 + 3\mu_2^2 = 3\sigma^4.$$

- This also shows that X has skewness $\mu_3/\mu_2^{3/2} = 0$ and kurtosis $\mu_4/\mu_2^2 = 3$.

Example: The m.g.f. of the Normal (cont.)

Let $Z \sim \mathrm{N}(0,1)$ and $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$. To compute $\mathbb{E}\left[e^{tZ}\right]$ and $\mathbb{E}\left[e^{tX}\right]$,

$$\mathbb{E}\left[e^{tZ}\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}z^2 + tz\right\} \mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(z^2 - 2tz\right)\right\} \mathrm{d}z$$

and, by completing the square as $z^2 - 2tz + t^2 - t^2 = (t-z)^2 - t^2$,

$$\begin{aligned}\mathbb{E}\left[e^{tZ}\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left((t-z)^2 - t^2\right)\right\} \mathrm{d}z \\ &= \exp\left\{\frac{t^2}{2}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z-t)^2\right\} \mathrm{d}z = \exp\left\{\frac{t^2}{2}\right\}.\end{aligned}$$

As $X = \mu + \sigma Z$ is a (location-scale) transformation of X,

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t(\mu+\sigma Z)}\right] = \exp\{t\mu\}\,\mathbb{M}_Z(t\sigma) = \exp\left\{t\mu + \frac{t^2\sigma^2}{2}\right\}.$$

Example: The m.g.f. and c.g.f. of the Binomial

- Let $X \sim \text{Bin}(n, p)$. From the binomial theorem

$$\sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n ,$$

and we have, defining $q = 1 - p$,

$$\begin{aligned} \mathbb{M}_X(t) &= \sum_{x=0}^{n} \mathrm{e}^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} \left(p\mathrm{e}^t\right)^x (1-p)^{n-x} = \left(p\mathrm{e}^t + q\right)^n \end{aligned}$$

and $\mathbb{K}_X(t) = n \log\left(p\mathrm{e}^t + q\right)$.

Example: The m.g.f. and c.g.f. of the Binomial (2)

- Thus,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{K}_X(t) = \frac{npe^t}{pe^t + q} \Rightarrow \mu = np,$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbb{K}_X(t) = \frac{npqe^t}{(pe^t + q)^2} \Rightarrow \mu_2 = npq.$$

- Further computations (see the solutions for details) yield

$$\frac{\mu_3}{\mu_2^{3/2}} = \frac{1-2p}{\sqrt{npq}} \qquad \text{and} \qquad \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{npq} - \frac{6}{n},$$

which, for any $p \in (0,1)$, converge to 0 and 3, respectively, as $n \to \infty$.

Practice (Examples 1.4 and 1.5 from the text)

- Let $X \sim \mathrm{Poi}(\lambda)$, with p.m.f. $f_X(x) = \mathrm{e}^{-\lambda}\lambda^x/x!$.
  Compute the c.g.f. of X and its first two moments.

- Let $Y \sim \mathrm{Gam}(a, b)$ with p.d.f. $f_Y(y) = b^a x^{a-1} \exp\{-bx\}/\Gamma(a)$,
  where $\Gamma(a) = \int_0^\infty x^{a-1}\mathrm{e}^{-x}\,\mathrm{d}x$.
  Compute the c.g.f. of Y and its first two moments.

Convergence in Distribution

- Let r.v.s X and Y be defined on $\{\mathbb{R}, \mathcal{B}, \Pr(\cdot)\}$.

  They are said to be **equal in distribution**, written $X \stackrel{d}{=} Y$, if $\Pr(X \in A) = \Pr(Y \in A) \quad \forall A \in \mathcal{B}$.
- For sequence of r.v.s X_n, $n = 1, 2, \ldots$, we say X_n **converges in distribution** to X, written $X_n \stackrel{d}{\to} X$, if $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$, for all points x such that $F_X(x)$ is continuous.
- To be brief, if X_n converges in distribution to a random variable which is, say, normally distributed, then we will write $X_n \stackrel{d}{\to} \mathrm{N}(\cdot, \cdot)$, where the mean and variance of the specified normal distribution are constants, and do not depend on n.
- Observe that $X_n \stackrel{d}{\to} \mathrm{N}\left(\mu, \sigma^2\right)$ implies that, for n sufficiently large, the distribution of X_n can be adequately approximated by that of a $\mathrm{N}\left(\mu, \sigma^2\right)$ random variable. We will denote this by writing $X_n \stackrel{\text{app}}{\sim} \mathrm{N}\left(\mu, \sigma^2\right)$.

Uniqueness and Convergence of the m.g.f.

Uniqueness

For r.v.s X and Y and some $h > 0$,

$$\mathbb{M}_X(t) = \mathbb{M}_Y(t) \, \forall \, |t| < h \quad \Rightarrow \quad X \stackrel{d}{=} Y.$$

Convergence

Let X_n be a sequence of r.v.s such that the corresponding m.g.f.s $\mathbb{M}_{X_n}(t)$ exist for $|t| < h$, for some $h > 0$, and all $n \in \mathbb{N}$.
If X is a random variable whose m.g.f. $\mathbb{M}_X(t)$ exists for $|t| \le h_1 < h$ for some $h_1 > 0$ and $\mathbb{M}_{X_n}(t) \to \mathbb{M}_X(t)$ as $n \to \infty$ for $|t| < h_1$, then $X_n \stackrel{d}{\to} X$.

Example

A sequence of Binomial distributed r.v.s.

Let X_n, $n = 1, 2, \ldots$ be a sequence of r.v.s such that $X_n \sim \mathrm{Bin}(n, p_n)$, with $p_n = \lambda/n$, for some constant value $\lambda \in \mathbb{R}_{>0}$, so that $\mathbb{M}_{X_n}(t) = (p_n \mathrm{e}^t + 1 - p_n)^n$, or

$$\mathbb{M}_{X_n}(t) = \left(\frac{\lambda}{n}\mathrm{e}^t + 1 - \frac{\lambda}{n}\right)^n = \left(1 + \frac{\lambda}{n}\left(\mathrm{e}^t - 1\right)\right)^n .$$

For all $h > 0$ and $|t| < h$, $\lim_{n\to\infty} \mathbb{M}_{X_n}(t) = \exp\{\lambda(\mathrm{e}^t - 1)\} = \mathbb{M}_P(t)$, where $P \sim \mathrm{Poi}(\lambda)$. That is, $X_n \xrightarrow{d} \mathrm{Poi}(\lambda)$.

Informally speaking, the binomial distribution with increasing n and decreasing p, such that np is a constant, approaches a Poisson distribution.

Vector m.g.f.

The m.g.f. of **vector** $\mathbf{X} = (X_1, \ldots, X_n)'$ is given by $\mathbb{M}_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}[\mathrm{e}^{\mathbf{t}'\mathbf{X}}]$, where $\mathbf{t} = (t_1, \ldots, t_n)' \in \mathbb{R}^n$.

As in the univariate case, if $\mathbb{M}_{\mathbf{X}}$ exists, then it characterizes the distribution of $\mathbf{X}$ and, thus, **all the marginals as well**. In particular,

$$\mathbb{M}_{\mathbf{X}}((0, \ldots, 0, t_i, 0, \ldots, 0)) = \mathbb{E}\left[\mathrm{e}^{t_i X_i}\right] = \mathbb{M}_{X_i}(t_i)\,, \quad i = 1, \ldots, n.$$

Generalizing the univariate case,

$$\begin{aligned} \frac{\partial^k \mathbb{M}_{\mathbf{X}}(\mathbf{t})}{\partial t_1^{k_1} \partial t_2^{k_2} \cdots \partial t_n^{k_n}} \quad = \quad & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} \\ & \times \exp\left\{t_1 x_1 + t_2 x_2 + \cdots + t_n x_n\right\} \\ & \times f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_n, \end{aligned}$$

(provided the exchange of derivative and integral is valid).

Vector m.g.f. (2)

- Therefore, the integer product moments of $\mathbf{X}$, $\mathbb{E}\left[\prod_{i=1}^{n} X_i^{k_i}\right]$ for $k_i \in \mathbb{N}$, are given by

$$\left.\frac{\partial^k \mathbb{M}_{\mathbf{X}}(\mathbf{t})}{\partial t_1^{k_1} \partial t_2^{k_2} \cdots \partial t_n^{k_n}}\right|_{\mathbf{t}=\mathbf{0}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

  for $k = \sum_{i=1}^{n} k_i$ and such that $k_i = 0$ means that the derivative with respect to t_i is not taken.

- For example, if X and Y are r.v.s with m.g.f. $\mathbb{M}_{X,Y}(t_1, t_2)$, then

$$\mathbb{E}[XY] = \left.\frac{\partial^2 \mathbb{M}_{X,Y}(t_1, t_2)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=0} \quad \text{and}$$

$$\mathbb{E}\left[X^2\right] = \left.\frac{\partial^2 \mathbb{M}_{X,Y}(t_1, t_2)}{\partial t_1^2}\right|_{t_1=t_2=0} = \left.\frac{\partial^2 \mathbb{M}_{X,Y}(t_1, 0)}{\partial t_1^2}\right|_{t_1=0}.$$

Result from Sawa (1972) and (1978)

Let X_1 and X_2 be r.v.s such that $\Pr(X_1 > 0) = 1$, with joint m.g.f. $\mathbb{M}_{X_1,X_2}(t_1, t_2)$ which exists for $t_1 < \epsilon$ and $|t_2| < \epsilon$, for $\epsilon > 0$. Then, if it exists, the kth order moment, $k \in \mathbb{N}$, of X_2/X_1 is given by

$$\mathbb{E}\left[\left(\frac{X_2}{X_1}\right)^k\right] = \frac{1}{\Gamma(k)} \int_{-\infty}^{0} (-t_1)^{k-1} \left[\frac{\partial^k}{\partial t_2^k} \mathbb{M}_{X_1,X_2}(t_1, t_2)\right]_{t_2=0} \mathrm{d}t_1.$$

There are several important applications of this result in econometrics, some of which stem from its relation to the least squares estimator. Recent examples in finance includes a study of the popular two-pass test to assess the risk premium in asset pricing models[1] and the bias in popular models for predicting excess returns.[2]

[1] Raymond Kan and Chu Zhang (1999, p. 230), *Two-Pass Tests of Asset Pricing Models with Useless Factors*, The Journal of Finance, Vol. 54(1), pp. 203-235.

[2] Yakov Amihud and Clifford M. Hurvich (2004), *Predictive Regressions: A Reduced-Bias Estimation Method*, Stern School of Business, New York University, New York.

Quick Review of Complex Numbers

- The **imaginary unit** i is defined to be a "number" such that $i^2 = -1$.
- The set of all complex numbers is $\mathbb{C} := \{a + bi \mid a, b \in \mathbb{R}\}$, and is closed under addition and multiplication:

$$\begin{aligned}(a+bi)+(c+di) &= (a+c)+(b+d)i \\ (a+bi)\cdot(c+di) &= (ac-bd)+(bc+ad)i.\end{aligned}$$

- If $z = a + bi$, then $\mathrm{Re}(z) := a$ and $\mathrm{Im}(z) := b$ are the **real** and **imaginary** parts of z.
- For complex number $z = a + bi$, its **complex conjugate** is $\overline{z} = a - bi$.
- The product $z \cdot \overline{z} = (a + bi)(a - bi) = a^2 - b^2 i^2 = a^2 + b^2$ is always a non-negative real number.

Complex Numbers (2)

- The sum is

$$z + \overline{z} = (a + bi) + (a - bi) = 2a = 2\,\mathrm{Re}(z).$$

- The absolute value of z, or its **(complex) modulus**, is

$$|z| = |a + bi| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}.$$

- Short calculations show that, for $z_1, z_2 \in \mathbb{C}$,

$$|z_1 z_2| = |z_1||z_2|, \quad \overline{z_1 z_2} = \overline{z_1}\ \overline{z_2}, \quad |z_1 + z_2| \leq |z_1| + |z_2|.$$

Exercise

- Let $z = a + bi$. Compute an expression for $1/z =: c + di$. Hint: Recall that the inverse of an invertible 2×2 matrix is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

  Verify your answer for $z = 1 + 2i$ with $1/z = 0.2 - 0.4i$; and $1/i = -i$.

- Show that $|z_1 + z_2| \leq |z_1| + |z_2|$.

Solution to $1/z$

We need

$$1 + 0i = z \cdot \frac{1}{z} = (a + bi)(c + di) = (ac - bd) + (bc + ad)\, i$$

or, as an equation for c and d,

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} ac - bd \\ bc + ad \end{bmatrix} = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}.$$

This is

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

with solution

$$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{a^2 + b^2} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

or

$$\frac{1}{z} = c + di = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}\, i.$$

Solution to $|z_1 + z_2| \le |z_1| + |z_2|$

To show this, let $z_1 = a + bi$ and $z_2 = c + di$, and since both sides are positive, it is equivalent if we show that $|z_1 + z_2|^2 \le (|z_1| + |z_2|)^2$. Then

$$\begin{aligned} |z_1 + z_2|^2 &= |(a + c) + (b + d)\,i|^2 = (a + c)^2 + (b + d)^2 \\ &= a^2 + b^2 + c^2 + d^2 + 2ac + 2bd \end{aligned}$$

and

$$\begin{aligned} (|z_1| + |z_2|)^2 &= |z_1|^2 + |z_2|^2 + 2\,|z_1|\,|z_2| \\ &= a^2 + b^2 + c^2 + d^2 + 2\sqrt{a^2 + b^2}\sqrt{c^2 + d^2}, \end{aligned}$$

and we wish to show $ac + bd < \sqrt{(a^2 + b^2)(c^2 + d^2)}$ or, squaring and expanding,

$$a^2c^2 + b^2d^2 + 2abcd < a^2c^2 + b^2d^2 + a^2d^2 + b^2c^2$$

or if $2abcd < a^2d^2 + b^2c^2$, which is true, because

$$a^2d^2 + b^2c^2 - 2abcd = (ad - bc)^2 > 0.$$

Complex Numbers (3)

- The exponential function is

$$\exp(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!}, \quad z \in \mathbb{C},$$

and, as in $\mathbb{R}$, $\exp(z_1 + z_2) = \exp(z_1)\exp(z_2)$ for every $z_1, z_2 \in \mathbb{C}$.

- From the series expansions of sin and cos, we get:

Euler's formula

$$\exp(it) = \cos(t) + i\sin(t).$$

- From Euler's formula, we get

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} \quad \text{and} \quad \cos z = \frac{e^{iz} + e^{-iz}}{2},$$

and, using $t = \pi$, we find $e^{i\pi} + 1 = 0$.

Complex Numbers (4)

- From Euler's formula, $\exp(2\pi i) = 1$, so that, for $z \in \mathbb{C}$,

$$\exp(z + 2\pi i) = \exp(z)\exp(2\pi i) = \exp(z).$$

- Calculation shows that $\exp(\bar{z}) = \overline{\exp(z)}$.
- The Riemann integral of a complex-valued function is the sum of the Riemann integrals of its real and imaginary parts. For example, calculation shows that

$$\int \mathrm{e}^{st}\,\mathrm{d}t = s^{-1}\mathrm{e}^{st}, \quad s \in \mathbb{C} \setminus 0,$$

which appears just like the result for $s \in \mathbb{R} \setminus 0$.

Exercise

Use Euler's formula to show that $i^i = \mathrm{e}^{-\pi/2}$. Hint: Take $t = \pi/2$.

Solution to Exercise

With $t = \pi/2$, Euler's formula implies $\mathrm{cis}\,(\pi/2) = \cos(\pi/2) + i\sin(\pi/2) = i$ or $e^{i\pi/2} = i$. Raising both sides to the power i, $\left(e^{i\pi/2}\right)^i = e^{i^2\pi/2} = e^{-\pi/2}$, so that $i^i = e^{-\pi/2} \approx 0.2079$.

Geometric Approach to Complex Numbers

- We can represent a complex number as a vector in the plane, with the real term on the horizontal axis, and the imaginary term on the vertical axis.
- Thus, the sum of two complex numbers is the sum of two vectors, and from Pythagoras, the modulus of $z \in \mathbb{C}$ is the length from 0 to z in the complex plane.
- The **unit circle** is the circle in the complex plane of radius 1 centered at 0, and includes all complex numbers of absolute value 1, i.e., such that $|z| = 1$.
- If $t \in \mathbb{R}$, then the number $\exp(it)$ is contained in the unit circle, because, from Euler's formula,

$$|\exp(it)| = \sqrt{\cos^2(t) + \sin^2(t)} = 1, \quad t \in \mathbb{R}.$$

Geometric Approach to Complex Numbers (2)

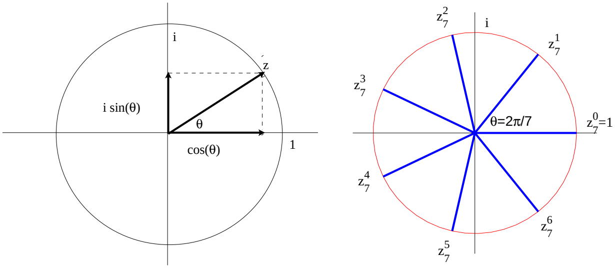


Figure: **Left**: $z = \cos(\theta) + i\sin(\theta)$. **Right**: Powers of $z_n = \exp(2\pi i/n) = e^{i\theta}$ for $n = 7$, demonstrating that $\sum_{j=0}^{n-1} z_n^j = 0$.

- For example, if $z = a + bi \in \mathbb{C}$, then

$$\exp(z) = \exp(a+bi) = \exp(a)\exp(bi) = \exp(a)\left[\cos(b) + i\sin(b)\right],$$

$$\text{and} \quad |\exp(z)| = |\exp(a)||\exp(bi)| = \exp(a) = \exp(\mathrm{Re}(z)).$$

Geometric Approach to Complex Numbers (3)

- If $z \neq 0$, let $r = |z| = \sqrt{a^2 + b^2}$ the modulus of z.
  Define the **argument** of z, denoted $\arg(z)$, to be the angle (in radians, measured counterclockwise from the positive real axis, modulo 2π), say θ, such that $a = r\cos(\theta)$ and $b = r\sin(\theta)$, i.e., for $a \neq 0$, $\arg(z) := \arctan(b/a)$.
- From Euler's formula,
  $$z = a + bi = r\cos(\theta) + ir\sin(\theta) = re^{i\theta},$$
  and, as $r = |z|$ and $\theta = \arg(z)$,
  $$\mathrm{Re}(z) = a = |z|\cos(\arg z) \quad \text{and} \quad \mathrm{Im}(z) = b = |z|\sin(\arg z).$$
- If $z_j = r_j \exp(i\theta_j)$, then $z_1 z_2 = r_1 r_2 \exp(i(\theta_1 + \theta_2))$, so that
  $$\arg(z_1 z_2 \cdots z_n) = \arg(z_1) + \arg(z_2) + \cdots + \arg(z_n)$$
  and, in particular, $\arg z^n = n\arg(z)$.

Properties of Characteristic Functions

- Using Euler's formula,

$$\begin{aligned}\varphi_X(t) &= \int_{-\infty}^{\infty} e^{itx}\, dF_X(x)\\ &= \int_{-\infty}^{\infty} \cos(tx)\, dF_X(x) + i\int_{-\infty}^{\infty} \sin(tx)\, dF_X(x)\\ &= \mathbb{E}[\cos(tX)] + i\,\mathbb{E}[\sin(tX)].\end{aligned}$$

- As $\cos(-\theta) = \cos(\theta)$ and $\sin(-\theta) = -\sin(\theta)$, it follows that

$$\varphi_X(-t) = \bar{\varphi}_X(t),$$

where $\bar{\varphi}_X$ is the complex conjugate of φ_X.

- And as $z + \bar{z} = (a + bi) + (a - bi) = 2a = 2\,\mathrm{Re}(z)$,

$$\varphi_X(t) + \varphi_X(-t) = 2\,\mathrm{Re}(\varphi_X(t)).$$

Properties of Characteristic Functions

- As $|\exp(it)| = \sqrt{\cos^2(t) + \sin^2(t)} = 1$, $t \in \mathbb{R}$, and $|z_1 + z_2| \leq |z_1| + |z_2|$,

$$|\varphi_X(t)| = \left| \int_{-\infty}^{\infty} e^{itx} \, dF_X(x) \right| \leq \int_{-\infty}^{\infty} \left| e^{itx} \right| \, dF_X(x) = \int_{-\infty}^{\infty} dF_X(x) = 1,$$

  so that, contrary to the m.g.f., the c.f. will always exist.

Uniqueness Theorem

For r.v.s X and Y, $\qquad \varphi_X = \varphi_Y \quad \Leftrightarrow \quad X \stackrel{d}{=} Y.$

- Clearly, if φ_X is real, then $\int_{-\infty}^{\infty} \sin(tx) \, dF_X(x) = 0$. It can be proven that, for r.v. X with c.f. φ_X and p.m.f. or p.d.f. f_X:

φ_X is real iff f_X is symmetric about zero.

Relation between the m.g.f. and c.f.

Let X be a r.v. whose m.g.f. exists. At first glance, comparing the definitions of the m.g.f. and the c.f., it seems that we could just write

$$\varphi_X(t) = \mathbb{M}_X(it). \tag{1}$$

As an illustration how to apply this formula, consider the case $X \sim \mathrm{N}(0,1)$. Then $\mathbb{M}_X(t) = \mathrm{e}^{t^2/2}$. Plugging *it* into the right-hand side of this formula yields $\varphi_X(t) = \mathrm{e}^{-t^2/2}$, which is the correct answer.

Note that, formally, (1) doesn't make sense since we are plugging a complex variable into a function that only admits real arguments. See the text for more discussion.

Relation between the m.g.f. and c.f.

In the vast majority of real applications, such as the Gaussian case, this will work, because the m.g.f. has a functional form allowing for complex arguments, and so we use (1) without formally checking.

An example of a distribution for which it does not work is the skew normal (discussed in Chapter 7), with

$$\mathbb{M}_X(t) = 2\exp\left\{t^2/2\right\}\Phi(t\delta), \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}.$$

Inversion Formulae for Mass and Density Functions

- Let X be a discrete r.v. with support $\{x_j\}_{j=1}^{\infty}$, such that $x_j = x_k$ iff $j = k$, and mass function f_X. Then

$$f_X(x_j) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{e}^{-itx_j} \varphi_X(t) \, \mathrm{d}t.$$

- Far more useful for numeric purposes is

$$f_X(x_j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-itx_j} \varphi_X(t) \, \mathrm{d}t.$$

- If X is a continuous r.v. with pdf f_X and $\int_{-\infty}^{\infty} |\varphi_X(t)| \, \mathrm{d}t < \infty$, then

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-itx} \varphi_X(t) \, \mathrm{d}t.$$

Inversion Formulae for Mass and Density Functions (2)

- To prove the formula for the continuous case, first note that, for constants $k \in \mathbb{R} \setminus 0$ and $T \in \mathbb{R}_{\geq 0}$,

$$\int_{-T}^{T} \cos(kt)\, \mathrm{d}t = \frac{2\sin(kT)}{k} \quad \text{and} \quad \int_{-T}^{T} \sin(kt)\, \mathrm{d}t = 0.$$

- The first follows because (i) by definition, $\mathrm{d}\sin x/\,\mathrm{d}x = \cos x$, (ii) the fundamental theorem of calculus, (iii) a simple change of variable, and (iv) that

$$\cos(-\theta) = \cos(\theta) \quad \text{and} \quad \sin(-\theta) = -\sin(\theta).$$

- The second can be derived similarly and/or confirmed quickly from a plot of sin.
- In the proof following, we assume all interchanges of integrals and of limit and integral are valid.

Proof of the Inversion Formula for the p.d.f.

- Take $T > 0$. Then

$$\begin{aligned}\int_{-T}^{T} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t &= \int_{-T}^{T} \mathrm{e}^{-itx} \int_{-\infty}^{\infty} \mathrm{e}^{ity} f_X(y)\,\mathrm{d}y\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} f_X(y) \int_{-T}^{T} \mathrm{e}^{it(y-x)}\,\mathrm{d}t\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} f_X(y) \frac{2\sin T(y-x)}{y-x}\,\mathrm{d}y.\end{aligned}$$

- Then, with $A = T(y-x)$ and $\mathrm{d}y = \mathrm{d}A/T$,

$$\begin{aligned}\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t &= 2\lim_{T\to\infty} \int_{-\infty}^{\infty} f_X(y) \frac{\sin T(y-x)}{y-x}\,\mathrm{d}y \\ &= 2\lim_{T\to\infty} \int_{-\infty}^{\infty} f_X\left(x + \frac{A}{T}\right) \frac{\sin A}{A}\,\mathrm{d}A = 2\pi f_X(x),\end{aligned}$$

because $\int_0^{\infty} \frac{\sin x}{x}\,\mathrm{d}x = \frac{\pi}{2}$.

Alternative Formulae

- Recall $\cos(-\theta) = \cos(\theta)$ and $\sin(-\theta) = -\sin(\theta)$, and Euler's formula $\exp(it) = \cos(t) + i\sin(t)$. From these, it follows that

$$\mathrm{e}^{-itx} = \cos(-tx) + i\sin(-tx) = \cos(tx) - i\sin(tx) = \overline{\mathrm{e}^{itx}}.$$

- Also: $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$, $\varphi_X(-t) = \bar{\varphi}_X(t)$, and $z + \overline{z} = 2\,\mathrm{Re}(z)$.
- Then, with $v = -t$,

$$\int_{-\infty}^{0} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t = \int_{0}^{\infty} \mathrm{e}^{ivx}\varphi_X(-v)\,\mathrm{d}v = \int_{0}^{\infty} \overline{\mathrm{e}^{-ivx}}\bar{\varphi}_X(v)\,\mathrm{d}v$$
$$= \int_{0}^{\infty} \overline{\mathrm{e}^{-ivx}\varphi_X(v)}\,\mathrm{d}v$$

so that, for X continuous, $f_X(x)$ is

$$\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t = \int_{0}^{\infty} \mathrm{e}^{-itx}\varphi_X(t)\,\mathrm{d}t + \int_{0}^{\infty} \overline{\mathrm{e}^{-ivx}\varphi_X(v)}\,\mathrm{d}v$$
$$= 2\int_{0}^{\infty} \mathrm{Re}\left[\mathrm{e}^{-itx}\varphi_X(t)\right]\,\mathrm{d}t.$$

Alternative Formulae (2)

- Thus, we get the alternative formulae

$$f_X(x_j) = \frac{1}{\pi} \int_0^{\pi} \mathrm{Re}\left[\mathrm{e}^{-itx_j} \varphi_X(t) \right] \, \mathrm{d}t \quad X \text{ discrete}$$

$$f_X(x) = \frac{1}{\pi} \int_0^{\infty} \mathrm{Re}\left[\mathrm{e}^{-itx} \varphi_X(t) \right] \, \mathrm{d}t \quad X \text{ continuous.}$$

- These could be more suitable for numeric work than their original counterparts if a 'canned' integration routine is available but which does not support complex arithmetic.

Ratios (Example 1.24)

Let X_1, X_2 be continuous r.v.s with joint p.d.f. f_{X_1,X_2}, joint c.f. φ_{X_1,X_2}, and such that $\Pr(X_2 > 0) = 1$ and $\mathbb{E}[X_2] = \mu_2 < \infty$. Geary (1944) showed that the density of $R = X_1/X_2$ can be written

$$f_R(r) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\left[\frac{\partial\varphi_{X_1,X_2}(s,t)}{\partial t}\right]_{t=-rs} ds.$$

The text details the proof. A nearly identical derivation shows that $f_R(r) = \mu_2 f_W(0)$, where W is a random variable with m.g.f.

$$\mathbb{M}_W(s) = \frac{1}{\mu_2}\frac{\partial}{\partial t}\mathbb{M}_{X_1,X_2}(s,t)\Big|_{t=-rs},$$

and the saddlepoint method can be applied to W.

Given the prevalence of ratios (as test statistics, estimators, etc.), these methods are very important. In particular, use of the latter result with a saddlepoint approximation (Chapter 5) leads to highly accurate yet trivially calculated approximations useful in practice.

Inversion Formulae for the c.d.f.

- Gil–Peleaz (1951) derived the inversion formula for continuous r.v. X with c.f. φ_X,

$$F_X(x) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{e^{itx}\varphi_X(-t) - e^{-itx}\varphi_X(t)}{it}\,dt.$$

- This can also be written as

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty g_x(t)\,dt,$$

where

$$g_x(t) = \frac{\mathrm{Im}\, z_x(t)}{t} = \frac{|z_x(t)|\sin(\arg z_x(t))}{t}$$

and

$$z_x(t) = e^{-itx}\varphi_X(t).$$

Using the FFT: Preliminaries

- The **fast Fourier transform** is the discrete Fourier transform, abbreviated FFT and DFT respectively, but calculated in a judicious manner which capitalizes on redundancies. The application of interest here is the fast numeric inversion of the characteristic function of a random variable for a (potentially large) grid of points.
- A **periodic function** is a function of time t, and its **period** is the time it takes for the function to repeat itself. Its **frequency** is the number of times the wave repeats in a given time unit. Let T be the smallest time that the function repeats, which is called its **fundamental period**, and its **fundamental frequency** is then $f = 1/T$.
- Let θ denote the angle in the unit circle, measured in radians (the arc length), of which the total length around the circle is 2π. If θ is the function of time t given by $\theta(t) = \omega t$, then ω is the **angular velocity**.

Using the FFT: Preliminaries (2)

- If we take $\omega = 2\pi/T$ and let t move from 0 to T, then $\theta(t) = \omega t$ will move from zero to 2π, and thus $\cos(\omega t)$ will make one complete cycle, and similarly for sin.
- Now with $n \in \mathbb{N}$, $\cos(n\omega t)$ will make n complete cycles, where the angular velocity is $n\omega$. We use $\cos(n\omega t)$ and $\sin(n\omega t)$, for $n = 0, 1, \ldots$, as a basis for decomposing periodic functions. That is, we wish to express a periodic function g with fundamental period T as

$$g(t) = \sum_{n=0}^{\infty} \left[a_n \cos(n\omega t) + b_n \sin(n\omega t)\right], \quad \omega = \frac{2\pi}{T},$$

where the a_i and b_i represent the **amplitudes** of the components.

Orthonormal Basis for $\mathbb{R}^n$

- For any $\mathbf{v} \in \mathbb{R}^n$ and any orthonormal basis of $\mathbb{R}^n$, say $\mathbf{e}_1, \ldots, \mathbf{e}_n$, we know that

$$\mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{e}_i \rangle \, \mathbf{e}_i .$$

- For example, take $n = 3$ and the non-orthogonal basis $\mathbf{a}_1 = (1, 0, 0)'$, $\mathbf{a}_2 = (1, 1, 0)'$, and $\mathbf{a}_3 = (1, 1, 1)'$. Using the **Gramm-Schmidt procedure** to orthonormalize gives, with $\mathbf{A}$ so defined,

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{O} = \begin{bmatrix} 0.7370 & 0.5910 & 0.3280 \\ 0.5910 & -0.3280 & -0.7370 \\ 0.3280 & -0.7370 & 0.5910 \end{bmatrix}.$$

  One could check that $\mathbf{O}'\mathbf{O} = \mathbf{O}\mathbf{O}' = \mathbf{I}_3$, so that the columns of $\mathbf{O}$, say $\mathbf{o}_1, \mathbf{o}_2, \mathbf{o}_3$, form an orthonormal basis for $\mathbb{R}^3$.
- For example, with $\mathbf{v} = [1, 5, -3]'$ and $w_i = \langle \mathbf{v}, \mathbf{o}_i \rangle$, $i = 1, 2, 3$, we get $w_1 = 2.7081$, $w_2 = 1.1620$ and $w_3 = -5.1299$, and indeed $w_1\mathbf{o}_1 + w_2\mathbf{o}_2 + w_3\mathbf{o}_3 = [1, 5, -3]' = \mathbf{v}$.

Orthonormal Basis for Fourier Series

- We are interested in the sequence of functions

$$\{2^{-1/2}, \cos(n\cdot), \sin(n\cdot),\ n \in \mathbb{N}\},$$

  because it forms an **infinite orthonormal basis** for the set of (piecewise) continuous functions on $[-\pi, \pi]$ with respect to the inner product

$$\langle f, g \rangle := \frac{1}{\pi} \int_{-\pi}^{\pi} f(t)\overline{g(t)}\, \mathrm{d}t,$$

  where $\bar{g}$ is the complex conjugate of function g.
- It is straightforward to verify orthonormality of these functions, the easiest of which is $\left\langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{1}{2}\, \mathrm{d}t = 1$.
- The others require the use of basic trigonometric identities. For example, after some substitution, $\langle \sin(nt), \sin(nt) \rangle = 1$ and, for $m \neq n$, $\langle \cos(mt), \cos(nt) \rangle = 0$.

Orthonormal Basis for Fourier Series (2)

- Thus, based on the analogous expression in $\mathbb{R}^n$, $\mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{e}_i \rangle \mathbf{e}_i$, we have, for a periodic function defined on $t \in [0, T]$ and such that $g(t) = g(t + T)$,

$$g(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(n\omega t) + b_n \sin(n\omega t) \right], \quad \omega = \frac{2\pi}{T},$$

with, for $n = 0, 1, 2, \ldots$,

$$a_n = \frac{2}{T} \int_0^T g(t) \cos(n\omega t) \, \mathrm{d}t, \quad b_n = \frac{2}{T} \int_0^T g(t) \sin(n\omega t) \, \mathrm{d}t.$$

Discrete Fourier Transform

- From Euler's formula $\exp(it) = \cos(t) + i\sin(t)$, it is easy to verify that
$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} \quad \text{and} \quad \cos z = \frac{e^{iz} + e^{-iz}}{2}.$$
From these, some work eventually yields the following **complex Fourier series expansion** for function g as
$$g(t) = \sum_{n=-\infty}^{\infty} C_n e^{in\omega t}, \quad C_n = \frac{1}{T}\int_0^T g(t)\, e^{-in\omega t}\, dt, \quad \omega = \frac{2\pi}{T}.$$
It seems natural to construct a discrete version of the C_n as
$$G_n = \frac{1}{T}\sum_{t=0}^{T-1} g_t e^{-2\pi i n t/T}, \quad n = 0, \dots, T-1,$$
where $g_t = g(t)$ (recall that g is evaluated at discrete points). This is the definition of the **(forward) discrete Fourier transform**.

Discrete Fourier Transform (2)

- For vector $\mathbf{g} = (g_0, \ldots, g_{T-1})$, we will write $\mathbf{G} = \mathcal{F}(\mathbf{g})$, where $\mathbf{G} = (G_0, \ldots, G_{T-1})$.
- The T values of g_t can be recovered from the **inverse discrete Fourier transform**, defined as

$$g_t = \sum_{n=0}^{T-1} G_n e^{2\pi i n t/T}, \quad t = 0, \ldots, T-1,$$

and we write $\mathbf{g} = \mathcal{F}^{-1}(\mathbf{G})$.

- Function $\mathcal{F}^{-1}$ is the inverse of $\mathcal{F}$ in the sense that, for a set of complex numbers $\mathbf{g} = (g_0, \ldots, g_{T-1})$, $\mathbf{g} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{g}))$.

Fast Fourier Transform

- The FFT is a recursive numerical procedure.
- The method will require approximately $T(\log_2(T)+1)$ complex-number multiplications, while the DFT requires about T^2.
- For $T = 2^6 = 64$ data points, the ratio $T/\log_2(T)$ is about 10, i.e., the FFT is 10 times faster.
- With $T = 2^{10} = 1024$ data points, the FFT is about 100 times faster.

Applying the FFT to c.f. Inversion

- Let X be a continuous random variable with p.d.f. f_X and c.f. φ_X, so that, for $\ell, u \in \mathbb{R}$ and $T \in \mathbb{N}$,

$$\varphi_X(s) = \int_{-\infty}^{\infty} \mathrm{e}^{isx} f_X(x)\, \mathrm{d}x \approx \int_{\ell}^{u} \mathrm{e}^{isx} f_X(x)\, \mathrm{d}x \approx \sum_{n=0}^{T-1} \mathrm{e}^{isx_n} P_n \ ,$$

with $P_n := f_X(x_n) \triangle x, \quad x_n := \ell + n(\triangle x), \quad \triangle x := \frac{u-\ell}{T}$. Higher accuracy results are obtained by taking larger values of T, $-\ell$ and u.

- Dividing both sides of the previous expression by $\mathrm{e}^{is\ell}$ and equating $sn(\triangle x)$ with $2\pi nt/T$ gives

$$\varphi_X(s)\, \mathrm{e}^{-is\ell} \approx \sum_{n=0}^{T-1} \mathrm{e}^{isn(\triangle x)} P_n = \sum_{n=0}^{T-1} \mathrm{e}^{2\pi int/T} P_n =: g_t,$$

yielding $\mathbf{g} = \mathcal{F}^{-1}(\mathbf{P})$, with $\mathbf{g} = (g_0, \ldots, g_{T-1})$, $\mathbf{P} = (P_0, \ldots, P_{T-1})$.

Applying the FFT to c.f. Inversion (2)

- As $s(\triangle x) = 2\pi t / T$, it makes sense to take a T–length grid of s–values as, say,

$$s_t = \frac{2\pi t}{T(\triangle x)}, \quad t = -\frac{T}{2}, -\frac{T}{2}+1, \ldots, \frac{T}{2}-1,$$

  so $g_t \approx \varphi_X(s_t)\, \mathrm{e}^{-i s_t \ell}$.
- The relation between $\mathcal{F}$ and $\mathcal{F}^{-1}$ implies $\mathbf{P} = \mathcal{F}(\mathbf{g})$, so applying the DFT (via the FFT) to $\mathbf{g}$ yields $\mathbf{P}$, the nth element of which is $P_n = f_X(x_n)\triangle x$.
- Finally, dividing this $\mathbf{P}$ by $\triangle x$ yields the p.d.f. of X at the T–length grid of x–values $\ell,\ \ell + (\triangle x),\ \ldots, \ell + (T-1)(\triangle x) \approx u$.

Exercise / Homework

Similar to Problem 1.12: Let X be an absolutely continuous random variable with support $\mathcal{S}_X$, characteristic function φ_X, and moment generating function $\mathbb{M}_X$ related by $\varphi_X(s) = \mathbb{M}_X(is)$. The p.d.f. at of X at $x \in \mathcal{S}_X$ can be computed by the usual inversion formula

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-itx} \varphi_X(t) \, \mathrm{d}t,$$

or, with $\mathbb{K}_X(s) = \ln \mathbb{M}_X(s)$ the c.g.f. of X, substituting $s = it$ gives, for $c = 0$,

$$f_X(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \exp\left\{\mathbb{K}_X(s) - sx\right\} \, \mathrm{d}s, \qquad (2)$$

and which, for any $c \in \mathbb{R}$, is the well-known Fourier–Mellin integral; see, e.g., Schiff (1999, Chap. 4).

Exercise / Homework (cont.)

We wish to derive an expression for $I_X(q;p) = \int_{-\infty}^{q} x^p f_X(x)\, \mathrm{d}x$, $p = 0, 1, 2$.

Substitute expression (2) into I_X, reverse the order of the integrals, and for the inner integral for $p = 1$, apply integration by parts.

Exercise / Homework Solution

For the $p = 1$ case, substituting (2) into the integral for the expected shortfall, $I_X(q) = \int_{-\infty}^{q} x f_X(x)\, \mathrm{d}x$, and reversing the order of the integrals gives

$$2\pi i I_X(q) = \int_{c-i\infty}^{c+i\infty} \int_{-\infty}^{q} x \exp\{\mathbb{K}_X(s) - sx\}\, \mathrm{d}x \mathrm{d}s.$$

For the inner integral, with $u = x$ and $\mathrm{d}v = \exp\{\mathbb{K}_X(s) - sx\}\, \mathrm{d}x$, integration by parts and restricting $c < 0$ such that c is in the convergence strip of $\mathbb{M}_X$ (so that the real part of s is negative), gives...

Exercise / Homework Solution

$$\begin{aligned}
&\int_{-\infty}^{q} x \exp\left\{\mathbb{K}_X(s) - sx\right\} \, \mathrm{d}x \\
= \quad & x \frac{\exp\left\{\mathbb{K}_X\left(s\right) - sx\right\}}{-s}\Bigg|_{-\infty}^{q} - \int_{-\infty}^{q} \frac{\exp\left\{\mathbb{K}_X\left(s\right) - sx\right\}}{-s} \, \mathrm{d}x \\
= \quad & q \frac{\exp\left\{\mathbb{K}_X\left(s\right) - sq\right\}}{-s} + \frac{1}{s}\int_{-\infty}^{q} \exp\left\{\mathbb{K}_X\left(s\right) - sx\right\} \, \mathrm{d}x \\
= \quad & -\frac{q}{s}\exp\left\{\mathbb{K}_X\left(s\right) - sq\right\} - \frac{1}{s^2}\exp\left\{\mathbb{K}_X\left(s\right) - sq\right\},
\end{aligned}$$

so that

$$2\pi i I_X\left(q\right) = -\int_{c-i\infty}^{c+i\infty} \left(\frac{q}{s} + \frac{1}{s^2}\right) \exp\left\{\mathbb{K}_X\left(s\right) - sq\right\} \, \mathrm{d}s.$$

Exercise / Homework Solution

Again using integration by parts with $u = \exp\{\mathbb{K}_X(s)\}$ and $\mathrm{d}v = \mathrm{e}^{-sq}\left(q/s + 1/s^2\right)\mathrm{d}s$, so that $\mathrm{d}u = \exp\{\mathbb{K}_X(s)\}\,\mathbb{K}'_X(s)\,\mathrm{d}s$ and $v = -\exp\{-qs\}/s$, gives

$$2\pi i I_X(q) = -\int_{c-i\infty}^{c+i\infty} \exp\{\mathbb{K}_X(s) - qs\}\,\mathbb{K}'_X(s)\,\frac{\mathrm{d}s}{s}.$$

Note that

$$\begin{aligned}
\lim_{s\to c+i\infty}\left|\frac{\exp\{\mathbb{K}_X(s) - qs\}}{s}\right| &= \mathrm{e}^{-qc}\lim_{k\to\infty}\left|\frac{\mathbb{M}_X(c+ik)\exp\{-iqk\}}{c+ik}\right| \\
&= \mathrm{e}^{-qc}\lim_{k\to\infty}\frac{|\mathbb{M}_X(c+ik)|}{\sqrt{c^2+k^2}} = 0.
\end{aligned}$$

Exercise / Homework Solution

For the $p = 2$ case: Similarly, for $I_X(q) = \int_{-\infty}^{q} x^2 f_X(x)\,\mathrm{d}x$, we need to evaluate

$$2\pi i I_X(q) = \int_{c-i\infty}^{c+i\infty} \int_{-\infty}^{q} x^2 \exp\{\mathbb{K}_X(s) - sx\}\,\mathrm{d}x\,\mathrm{d}s$$

and, with $u = x^2$ and $\mathrm{d}v = \exp\{\mathbb{K}_X(s) - sx\}\,\mathrm{d}x$,

$$\begin{aligned}
&\int_{-\infty}^{q} x^2 \exp\{\mathbb{K}_X(s) - sx\}\,\mathrm{d}x \\
=\;& x^2 \frac{\exp\{\mathbb{K}_X(s) - sx\}}{-s}\bigg|_{-\infty}^{q} - 2\int_{-\infty}^{q} x \frac{\exp\{\mathbb{K}_X(s) - sx\}}{-s}\,\mathrm{d}x \\
=\;& -\frac{q^2}{s}\exp\{\mathbb{K}_X(s) - sq\} + \frac{2}{s}\int_{-\infty}^{q} x\exp\{\mathbb{K}_X(s) - sx\}\,\mathrm{d}x \\
=\;& -\exp\{\mathbb{K}_X(s) - sq\}\left(\frac{q^2}{s} + \frac{2q}{s^2} + \frac{2}{s^3}\right).
\end{aligned}$$

Exercise / Homework Solution

And with $u = \exp\{\mathbb{K}_X(s)\}$ and $\mathrm{d}v = \mathrm{e}^{-sq}\left(q^2/s + 2q/s^2 + 2/s^3\right)\mathrm{d}s$, so that $\mathrm{d}u = \exp\{\mathbb{K}_X(s)\}\mathbb{K}'_X(s)\,\mathrm{d}s$ and

$$v = \int \mathrm{e}^{-sq}\left(q^2/s + 2q/s^2 + 2/s^3\right)\mathrm{d}s = -\frac{\mathrm{e}^{-qs}}{s^2}(1+qs),$$

we have

$$\begin{aligned} 2\pi i I_X(q) &= -\int_{c-i\infty}^{c+i\infty} \exp\{\mathbb{K}_X(s) - sq\}\left(\frac{q^2}{s} + \frac{2q}{s^2} + \frac{2}{s^3}\right)\mathrm{d}s \\ &= -\int_{c-i\infty}^{c+i\infty} (1+sq)\exp\{\mathbb{K}_X(s) - qs\}\mathbb{K}'_X(s)\frac{\mathrm{d}s}{s^2}. \end{aligned}$$

Intermediate Probability: A Computational Approach

m.g.f. of the Sum of Independent r.v.s

- Recall that, under certain conditions, the m.g.f. of X, $\mathbb{M}_X(s) = \mathbb{E}\left[e^{tX}\right]$, **characterizes** the distribution of X.
- There are many situations in which the derivation of the m.g.f. of a r.v. is (often far) easier than that of its distribution.
- The best illustration of this principle is the m.g.f. of a sum of **independent** r.v.s X and Y, given by

$$\mathbb{M}_{X+Y}(t) = \mathbb{E}\left[e^{t(X+Y)}\right] = \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right] = \mathbb{M}_X(t)\,\mathbb{M}_Y(t)\,,$$

  using the fact that they are independent.
- This result extends in an obvious way to n independent variables $X_1, \ldots, X_n$.

Random Variables Closed Under Addition

- If X and Y are Poisson distributed with parameters λ_1 and λ_2, then

$$\begin{aligned} \mathbb{M}_{X+Y}(t) &= \mathbb{M}_X(t)\,\mathbb{M}_Y(t) = \exp\left\{\lambda_1\left(\mathrm{e}^t - 1\right)\right\}\exp\left\{\lambda_2\left(\mathrm{e}^t - 1\right)\right\} \\ &= \exp\left\{(\lambda_1 + \lambda_2)\left(\mathrm{e}^t - 1\right)\right\} \end{aligned}$$

  is the m.g.f. of a Poisson random variable with parameter $\lambda_1 + \lambda_2$, which determines the distribution of $X + Y$.
- Inspection of the m.g.f.s of binomial (with constant p) and negative binomial (with constant p) r.v.s show that they are also **closed under addition**, meaning that the sum belongs to the same distribution family as its components.
- This also follows directly by expressing the binomial as a sum of i.i.d. Bernoulli r.v.s (and likewise for the negative binomial with i.i.d. geometric r.v.s.).

Random Variables Closed Under Addition (2)

- Recalling the results on linear combinations of normal r.v.s, we see that normal r.v.s are also **closed under addition**.
- Recall that, if $X \sim \mathrm{Gam}(\alpha, \beta)$, then $\mathbb{M}_X(t) = \left(\frac{\beta}{\beta - t}\right)^{\alpha}$.
- It follows immediately that, if $X_i \overset{\text{ind}}{\sim} \mathrm{Gam}(\alpha_i, \beta)$, then $\sum_i X_i \sim \mathrm{Gam}(\sum_i \alpha_i, \beta)$, i.e., a sum of independent gamma random variables **each with the same scale parameter** also follows a gamma distribution.

Special Cases

- Two special cases are of great interest:
  1. If $X_i \overset{\text{iid}}{\sim} \mathrm{Exp}(\lambda)$, then $\sum_i^n X_i \sim \mathrm{Gam}(n, \lambda)$
  2. If $Y_i \overset{\text{ind}}{\sim} \chi^2(\nu_i)$, then $\sum_i^n Y_i \sim \chi^2\left(\sum_i^n \nu_i\right)$.
- Fact: if $X \sim \mathrm{N}(0,1)$ and $Y = X^2$, then $Y \sim \chi_1^2$.
- Combine these: For $Z_i \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$, $\sum_{i=1}^n Z_i^2 \sim \chi^2(n)$.
- More generally:

Important Result

$$\text{If } X_i \overset{\text{iid}}{\sim} \mathrm{N}\left(\mu_i, \sigma_i^2\right), \quad \text{then} \quad \sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2(n).$$

Not Recognizable m.g.f.

Question

What to do when, under independence, the resulting m.g.f. is not recognizable?

Answer

The inversion formulae and numeric integration can be used to compute the p.d.f. and c.d.f..

Example One: Exponential

The m.g.f. of $X \sim \mathrm{Exp}(\lambda)$, with $f_X(x) = \lambda e^{-\lambda x} \mathbb{I}_{(0,\infty)}(x)$ is

$$\mathbb{M}_X(t) = \lambda / (\lambda - t), \quad t < \lambda,$$

so

$$\varphi_X(t) = \mathbb{M}_X(it) = \lambda / (\lambda - it).$$

Now let $X_j \overset{\text{ind}}{\sim} \mathrm{Exp}(\lambda_j)$, $j = 1, \ldots, n$ and $S = \sum_{j=1}^n X_j$. Then

$$\varphi_S(t) = \prod_{j=1}^n \varphi_{X_j}(t) = \prod_{j=1}^n \frac{\lambda_j}{\lambda_j - it},$$

and the inversion formula can be applied.
It turns out that, in this case, a simple, computable formula exists for the p.m.f. when the λ_j are distinct; see below.

Example Two: Binomial

Let $X_j \overset{\text{ind}}{\sim} \mathrm{Bin}(n_j, p_j)$ and $S = \sum_{j=1}^{k} X_j$.

For $k = 2$, the p.m.f. is given using the discrete convolution formula

$$\begin{aligned} \Pr(X_1 + X_2 = x) &= \sum_{i=0}^{x} \Pr(X_1 = i) \Pr(X_2 = x - i) \\ &= \sum_{i=0}^{x} \Pr(X_1 = x - i) \Pr(X_2 = i). \end{aligned}$$

This makes sense because X_1 and X_2 are independent and, in order for X_1 and X_2 to sum to x, it must be the case that one of the events $\{X_1 = 0, X_2 = x\}$, $\{X_1 = 1, X_2 = x - 1\}$, ..., $\{X_1 = x, X_2 = 0\}$ must have occurred. These events partition the event $\{X = x\}$.

Example Two: Binomial (cont.)

It is straightforwardly generalized for $k > 2$, but the 'curse of dimensionality' appears.

Instead, from

$$\mathbb{M}_S(t) = \prod_{j=1}^{k} \left(p_j \mathrm{e}^t + q_j\right)^{n_j}$$

and

$$\varphi_S(t) = \mathbb{M}_S(it)\,,$$

the inversion formula can be calculated to obtain the p.m.f..

Notice that k could be replaced by $\sum_{j=1}^{k} n_j$ and the n_j set to one, without loss of generality.

Example Three: Geometric

Imagine that boxes of a certain breakfast cereal each contain one of r different prizes, and that

1. equal numbers of the r prizes were distributed out,
2. randomly so, and
3. the number of cereal boxes manufactured is so large that purchases (sampling) can be adequately assumed to be conducted with replacement.

Let Y_k be the number of cereal boxes necessary to purchase in order to get at least one of k different prizes, $k = 2, \ldots, r$.

Example Three: Geometric (cont.)

- For $k = 1$, Y_1 is degenerate, with $Y_1 = 1$.
- For $n \leq k - 1$, $\Pr(Y_k = n) = 0$.
- For $n \geq k$, combinatoric arguments can be used to show (see Paolella, 2006, page 137 and Section 2.3.2) that, with $m = r - k + 1$,

$$\Pr(Y_k > n) = r^{-n} \sum_{i=m}^{r-1} (-1)^{i-m} \binom{i-1}{i-m} \binom{r}{i} (r-i)^n,$$

for $k = 2, \ldots, r$.

Example Three: Geometric (cont.)

A program to calculate the c.d.f. of Y_k is given below:

```
function cdf = occupancycdf(nvec,r,k)
nl=length(nvec); cdf=zeros(nl,1); m=r-k+1;
for nloop=1:nl
  n=nvec(nloop);
  s=0;
  i=m:(r-1); sgn=(-1).^(i-m);
  c1=C(i-1,i-m); c2=C(r,i); frc=( (r-i)/r ).^n;
  s=sum(sgn.*c1.*c2.*frc);
  cdf(nloop)=1-s;
end
```

where function `C(a,b)` computes $\binom{a}{b}$.

Example Three: Geometric (cont.)

- Define G_i to be the number of required purchases to get a new prize so far not in the collection—in which there are currently i different prizes, $i = 0, 1, \ldots, k-1$.
- Clearly, $G_0 = 1$.
- Recalling the geometric distribution and the nature of the sampling scheme, we see that $G_j \overset{\text{ind}}{\sim} \mathrm{Geo}(p_j)$ with density

$$f_{G_j}(g; p_j) = p_j (1 - p_j)^{g-1} \mathbb{I}_{\{1,2,\ldots\}}(g), \qquad p_j = \frac{r-j}{r}$$

- Thus,

$$Y_k = \sum_{j=0}^{k-1} G_j, \quad G_j \overset{\text{ind}}{\sim} \mathrm{Geo}(p_j), \quad p_j = \frac{r-j}{r}.$$

Example Three: Geometric (cont.)

- The m.g.f. of G_j is $p_j e^t \left(1 - q_j e^t\right)^{-1}$, $q_j = 1 - p_j$, so that

$$\mathbb{M}_{Y_k}(t) = \prod_{j=0}^{k-1} \mathbb{M}_{G_j}(t) = \mathrm{e}^{kt} \prod_{j=0}^{k-1} \frac{p_j}{1 - q_j \mathrm{e}^t},$$

  and $\varphi_{Y_k}(t) = \mathbb{M}_{Y_k}(it)$.
- A program in the textbook, `occpmf`, computes the p.m.f. of Y_k by applying the inversion formula to the characteristic function φ_{Y_k}.

Example Three: Geometric (cont.)

Thus, for example, with $r = 50$ and $k = 40$, $\Pr(71 \leq Y_k \leq 80)$ can be computed in the following two ways:

```
r=50; k=40;
occupancycdf(80,r,k)-occupancycdf(70,r,k)
sum( occpmf(71:80,r,k) )
```

This yields the same answer to about 9 significant digits,

```
ans = 0.377144288496197
ans = 0.377144288676300
```

Example Three: Geometric (cont.)

- Because of the nature of Y_k, a combinatoric derivation of its c.d.f. is possible, as given above, and thus for a sum of independent geometric random variables with the precise structure on the p_j, $p_j = (r - j)/r$.
- It turns out that a simple c.d.f. expression exists for the general case of $S_{n,\mathbf{p}} = \sum_{j=1}^{n} G_j$, where $G_j \overset{\text{ind}}{\sim} \mathrm{Geo}(p_j)$, $\mathbf{p} = (p_1, \ldots, p_n)$, and the p_j are unrestricted, *except that they need to be distinct*.

Example Three: Geometric (cont.)

- The result is:

$$\Pr\left(S_{n,\mathbf{p}} \leq s\right) = 1 - \sum_{i=1}^{n} q_i^s \prod_{i \neq j}^{n} \frac{p_j}{p_j - p_i}, \quad q_i = 1 - p_i.$$

- The proof by induction is easy; derivation of the expression itself could presumably be done with contour integration...
- Before showing the induction proof, let's check it numerically.
- Recall that, for Y_k, $G_0 = 1$, so, to evaluate the c.d.f. of Y_k at y for a given k and y, we want

$$\begin{aligned} \Pr\left(G_0 + G_1 + \cdots + G_{k-1} \leq y\right) &= \Pr\left(G_1 + \cdots + G_{k-1} \leq y - 1\right) \\ &= \Pr\left(S_{k-1,\mathbf{p}} \leq y - 1\right), \end{aligned}$$

with

$$\mathbf{p} = \left(\frac{r-1}{r}, \ldots, \frac{r-(k-1)}{r}\right).$$

Example Three: Geometric (cont.)

A program to compute $\Pr\left(S_{n,\mathbf{p}} \leq s\right)$ is as follows:

```
function cdf=rosspekoz(p,s)
n=length(p); q=1-p; total=0;
for i=1:n
  prd=1;
  for j=1:n
    if i ~= j, prd=prd*p(j)/(p(j)-p(i)); end
  end
  total=total + q(i)^s * prd;
end
cdf=1-total;
```

Example Three: Geometric (cont.)

Then running the code:

```
r=50; k=40; y=80; pvec=( (r-1):-1:(r-k+1) )/r;
rosspekoz(pvec,y-1)
occupancycdf(y,r,k)
```

gives

```
ans = 0.607477726039690
ans = 0.607477726040767
```

which are the same, up to round-off error.

It is important to note that, if some of the p_j are the same, the expression for $\Pr(S_{n,\mathbf{p}} \leq s)$ above will not work, but the method using the inversion formula still does.

Example Three: Geometric (cont.)

The result can be proven by induction. As in Ross and Peköz (2007, page 23), let $A_{s,n} = \Pr(S_{n,\mathbf{p}} > s)$ and note that

$$A_{1,1} = \Pr(G_1 > 1) = 1 - \Pr(G_1 = 1) = 1 - p_1 = q_1.$$

Then, assume $A_{i,j}$ is as stated above for all $i + j < s + n$. Note that (writing S_n for $S_{n,\mathbf{p}}$),

$$\begin{aligned} \Pr(S_n > s \mid G_n > 1) &= \Pr(G_1 + \cdots + G_{n-1} + G_n - 1 > s - 1 \mid G_n > 1) \\ &= \Pr(G_1 + \cdots + G_{n-1} + G_n > s - 1) \\ &= A_{s-1,n} \end{aligned}$$

because, given that $G_n > 1$ and the independence of the Bernoulli trials in a geometric r.v., $(G_n - 1) \mid G_n > 1$ has the same distribution as G_n.

Example Three: Geometric (cont.)

Thus, conditioning on the event $\{G_n > 1\}$, we have

$$\begin{aligned}
A_{s,n} &= \Pr(G_n > 1) \Pr(S_n > s \mid G_n > 1) + \Pr(G_n = 1) \Pr(S_n > s \mid G_n = 1) \\
&= q_n A_{s-1,n} + p_n \Pr(G_1 + \cdots + G_{n-1} + 1 > s) \\
&= q_n A_{s-1,n} + p_n A_{s-1,n-1} \\
&= q_n \sum_{i=1}^{n} q_i^{s-1} \prod_{i \neq j}^{n} \frac{p_j}{p_j - p_i} + p_n \sum_{i=1}^{n-1} q_i^{s-1} \prod_{i \neq j}^{n-1} \frac{p_j}{p_j - p_i} \\
&= q_n \sum_{i=1}^{n} q_i^{s-1} \prod_{i \neq j}^{n} \frac{p_j}{p_j - p_i} + p_n \sum_{i=1}^{n-1} q_i^{s-1} \frac{p_n - p_i}{p_n} \prod_{i \neq j}^{n} \frac{p_j}{p_j - p_i}
\end{aligned}$$

(continued on next page)

Example Three: Geometric (cont.)

Letting P_i be the term involving the product, with $q_i = 1 - p_i$,

$$
\begin{aligned}
A_{s,n} &= q_n \sum_{i=1}^{n} q_i^{s-1} P_i + p_n \sum_{i=1}^{n-1} q_i^{s-1} \left(1 - \frac{p_i}{p_n}\right) P_i \\
&= q_n \sum_{i=1}^{n} q_i^{s-1} P_i + p_n \sum_{i=1}^{n} q_i^{s-1} \left(1 - \frac{p_i}{p_n}\right) P_i \\
&= q_n \sum_{i=1}^{n} q_i^{s-1} P_i + p_n \sum_{i=1}^{n} q_i^{s-1} P_i - p_n \sum_{i=1}^{n} q_i^{s-1} \frac{p_i}{p_n} P_i \\
&= \sum_{i=1}^{n} q_i^{s-1} P_i - \sum_{i=1}^{n} q_i^{s-1} (1 - q_i) P_i \\
&= \sum_{i=1}^{n} q_i^{s} P_i
\end{aligned}
$$

as was to be shown.

Density of Simple Functions of Two r.v.s

- Notice that the "m.g.f. method" described above is limited to the sums of independent r.v.s.
- Here, integral formulae are introduced for the density of sums, differences, products and quotients of two, possibly dependent, r.v.s.
- Let X and Y are jointly distributed continuous random variables with density $f_{X,Y}(x, y)$, then the densities of $S = X + Y$, $D = X - Y$, $P = XY$ and $R = X/Y$ can be respectively expressed as follows:

Density of Simple Functions of two r.v.s (2)

$$f_S(s) = \int_{-\infty}^{\infty} f_{X,Y}(x, s-x)\, \mathrm{d}x = \int_{-\infty}^{\infty} f_{X,Y}(s-y, y)\, \mathrm{d}y$$

$$f_D(d) = \int_{-\infty}^{\infty} f_{X,Y}(x, x-d)\, \mathrm{d}x = \int_{-\infty}^{\infty} f_{X,Y}(d+y, y)\, \mathrm{d}y$$

$$f_P(p) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{p}{x}\right) \mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{X,Y}\left(\frac{p}{y}, y\right) \mathrm{d}y$$

$$f_R(r) = \int_{-\infty}^{\infty} \frac{|x|}{r^2} f_{X,Y}\left(x, \frac{x}{r}\right) \mathrm{d}x = \int_{-\infty}^{\infty} |y|\, f_{X,Y}(ry, y)\, \mathrm{d}y$$

Derivation for Sum: Double Integral

- Let X and Y be continuous r.v.s with joint distribution $f_{X,Y}$ and define $S = X + Y$.
- With $u = x + y$, we have $y = u - x$ and $\mathrm{d}y = \mathrm{d}u$, so that

$$
\begin{aligned}
\Pr(S \le s) &= \iint_{x+y \le s} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_{-\infty}^{\infty}\int_{-\infty}^{s-x} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{s} f_{X,Y}(x,u-x)\,\mathrm{d}u\,\mathrm{d}x \\
&= \int_{-\infty}^{s}\int_{-\infty}^{\infty} f_{X,Y}(x,u-x)\,\mathrm{d}x\,\mathrm{d}u.
\end{aligned}
$$

- Thus,

$$
f_S(s) = \frac{\mathrm{d}}{\mathrm{d}s} F_S(s) = \int_{-\infty}^{\infty} f_{X,Y}(x, s-x)\,\mathrm{d}x.
$$

Exercise for Product with Double Integral

Use the double integral approach to derive the expression for the density of the product.

Exercise for Product: Solution

For $P = XY$, note that $xy \leq p \Rightarrow y \leq p/x$ for $x > 0$ and $xy \leq p \Rightarrow y \geq p/x$ for $x < 0$. Then, with $u = xy$, $y = u/x$, $\mathrm{d}y = \mathrm{d}u/x$,

$$
\begin{aligned}
F_P(p) &= \Pr(XY \leq p) = \iint_{xy \leq p} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{0} \int_{p/x}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_{0}^{\infty} \int_{-\infty}^{p/x} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{0} \int_{p}^{-\infty} f_{X,Y}\left(x, \frac{u}{x}\right) \frac{1}{x}\,\mathrm{d}u\,\mathrm{d}x + \int_{0}^{\infty} \int_{-\infty}^{p} f_{X,Y}\left(x, \frac{u}{x}\right) \frac{1}{x}\,\mathrm{d}u\,\mathrm{d}x \\
&= \int_{-\infty}^{0} \int_{-\infty}^{p} f_{X,Y}\left(x, \frac{u}{x}\right) \left(-\frac{1}{x}\right) \mathrm{d}u\,\mathrm{d}x + \int_{0}^{\infty} \int_{-\infty}^{p} f_{X,Y}\left(x, \frac{u}{x}\right) \frac{1}{x}\,\mathrm{d}u\,\mathrm{d}x \\
&= \int_{-\infty}^{0} \int_{-\infty}^{p} f_{X,Y}\left(x, \frac{u}{x}\right) \frac{1}{|x|}\,\mathrm{d}u\,\mathrm{d}x + \int_{0}^{\infty} \int_{-\infty}^{p} f_{X,Y}\left(x, \frac{u}{x}\right) \frac{1}{|x|}\,\mathrm{d}u\,\mathrm{d}x \\
&= \int_{-\infty}^{p} \int_{-\infty}^{\infty} f_{X,Y}\left(x, \frac{u}{x}\right) \frac{1}{|x|}\,\mathrm{d}x\,\mathrm{d}u.
\end{aligned}
$$

Exercise for Product: Solution

Applying Leibniz' rule to

$$F_P(p) = \int_{-\infty}^{p} \int_{-\infty}^{\infty} f_{X,Y}\left(x, \frac{u}{x}\right) \frac{1}{|x|} \, \mathrm{d}x \, \mathrm{d}u.$$

gives

$$f_P(p) = \frac{\mathrm{d}}{\mathrm{d}p} F_P(p) = \int_{-\infty}^{\infty} f_{X,Y}\left(x, \frac{p}{x}\right) \frac{1}{|x|} \, \mathrm{d}x.$$

Derivation for Sum: Jacobian Transformation

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be an n-dimensional continuous r.v. and $\mathbf{g} = (g_1(\mathbf{x}), \ldots, g_n(\mathbf{x}))$ a continuous bijection which maps $\mathcal{S}_{\mathbf{X}}$, the support of $\mathbf{X}$, onto $\mathcal{S}_{\mathbf{Y}}$, a subset of $\mathbb{R}^n$.

Then the p.d.f. of $\mathbf{Y} = (Y_1, \ldots, Y_n) = \mathbf{g}(\mathbf{X})$ is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) \left|\det \mathbf{J}\right|,$$

where $\mathbf{x} = \left(g_1^{-1}(\mathbf{y}), \ldots, g_n^{-1}(\mathbf{y})\right)$ and

$$\mathbf{J} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial g_1^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial g_1^{-1}(\mathbf{y})}{\partial y_n} \\ \frac{\partial g_2^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial g_2^{-1}(\mathbf{y})}{\partial y_2} & & \frac{\partial g_2^{-1}(\mathbf{y})}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial g_n^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial g_n^{-1}(\mathbf{y})}{\partial y_n} \end{pmatrix}$$

is the Jacobian of $\mathbf{g}$.

Derivation for Sum: Jacobian Transformation

- To derive f_S using a bivariate transformation, a second, "dummy" variable is required.
- We take $T = Y$, which is both simple and such that $(s, t) = (g_1(x, y), g_2(x, y)) = (x + y, y)$ is a bijection.
- The inverse transformation is $(x, y) = \left(g_1^{-1}(s, t), g_2^{-1}(s, t)\right) = (s - t, t)$, so that

$$f_{S,T}(s, t) = |\det \mathbf{J}|\, f_{X,Y}(x, y),$$

where

$$\mathbf{J} = \begin{bmatrix} \partial x/\partial s & \partial x/\partial t \\ \partial y/\partial s & \partial y/\partial t \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \quad |\det \mathbf{J}| = 1,$$

or $f_{S,T}(s, t) = f_{X,Y}(s - t, t)$.
- Thus,

$$f_S(s) = \int_{-\infty}^{\infty} f_{X,Y}(s - t, t)\, \mathrm{d}t.$$

Exercise for Product with Jacobian

Use the Jacobian transformation approach to derive the expression for the density of the product.

Exercise for Product: Solution

Let $Q = Y$ so that the inverse transformation of $\{p = xy,\ q = y\}$ is $\{x = p/q,\ y = q\}$, and $f_{P,Q}(p,q) = |\det \mathbf{J}|\, f_{X,Y}(x,y)$, where

$$\mathbf{J} = \begin{bmatrix} \partial x/\partial p & \partial x/\partial q \\ \partial y/\partial p & \partial y/\partial q \end{bmatrix} = \begin{bmatrix} 1/q & -pq^{-2} \\ 0 & 1 \end{bmatrix}, \quad |\det \mathbf{J}| = \frac{1}{|q|},$$

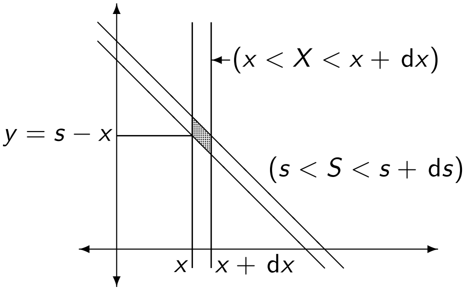
or $f_{P,Q}(p,q) = |q|^{-1} f_{X,Y}(p/q,\ q)$. Thus,

$$f_P(p) = \int_{-\infty}^{\infty} |q|^{-1} f_{X,Y}\left(\frac{p}{q}, q\right) \mathrm{d}q.$$

Derivation for Sum: Graphical

- Recall that, if X is a continuous r.v. with p.d.f. f_X and support $\mathcal{S}_X$, then, for all $x \in \mathcal{S}_X$, $\Pr(X = x) = 0$, while, for small $\epsilon > 0$, $\Pr(x < X < x + \epsilon) \approx f_X(x)\epsilon$.
- A similar statement holds for bivariate r.v.s via $\Pr(x < X < x + \epsilon_1,\, y < Y < y + \epsilon_2) \approx f_{X,Y}(x, y)\epsilon_1\epsilon_2$.
- The following figure corresponds to sum $S = X + Y$ depicted for a constant value $S = s$.

Derivation for Sum: Graphical (cont.)



For any x, the probability of event $(x < X < x + \mathrm{d}x,\ s < S < s + \mathrm{d}s)$ is approximated by the area of the shaded parallelogram times the density $f_{X,Y}$ evaluated at x and $s - x$, i.e., $f_{X,Y}(x, s - x)\, \mathrm{d}x\, \mathrm{d}s$.

The probability that $S = s$ is then obtained by 'summing' over all possible x, i.e., $\Pr(s < S < s + \mathrm{d}s) \approx \left[\int f_{X,Y}(x, s - x)\, \mathrm{d}x \right] \mathrm{d}s$.

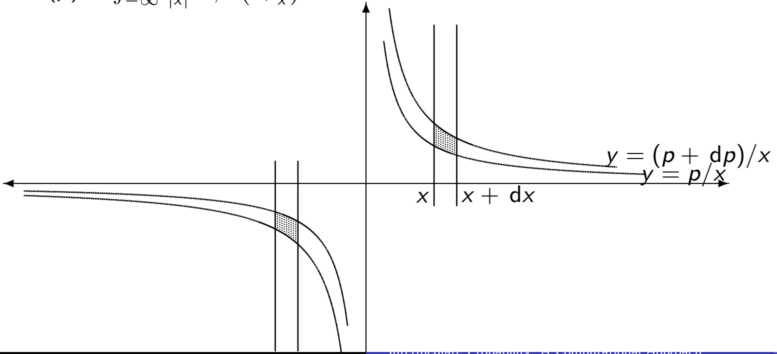
Graphical Derivation for Product

For the product $P = XY$, the area of the shaded regions, when treated as rectangles, is given by $\mathrm{d}x$ times the absolute value of $(p + \mathrm{d}p)/x - p/x$ or $\mathrm{d}x\,\mathrm{d}p/|x|$.

Integrating out x from the expression

$$\Pr(x < X < x + \mathrm{d}x,\ p < P < p + \mathrm{d}p) \approx f_{X,Y}(x, p/x)\,|x|^{-1}\,\mathrm{d}x\,\mathrm{d}p$$

yields $f_P(p) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{p}{x}\right)\,\mathrm{d}x$.



Example with Sums of Independent Exponential

- As above, let $X_i \overset{\text{ind}}{\sim} \mathrm{Exp}(\lambda_i)$, $i = 1, \ldots, n$ and $S = \sum_{i=1}^{n} X_i$.
- If all the λ_i are the same, then $S \sim \mathrm{Gam}(n, \lambda)$. If not, we saw that

$$\varphi_S(t) = \prod_{j=1}^{n} \varphi_{X_j}(t) = \prod_{j=1}^{n} \frac{\lambda_j}{\lambda_j - it},$$

and the inversion formula can be applied to get the p.d.f. and c.d.f..

Example with Sums of Independent Exponential (cont.)

- Another way is as follows. First let $n = 2$. Then, from the integral convolution formula,

$$\begin{aligned} f_S(s) &= \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(s-x)\,\mathrm{d}x = \int_0^s \lambda_1 \mathrm{e}^{-\lambda_1 x} \lambda_2 \mathrm{e}^{-\lambda_2(s-x)}\,\mathrm{d}x \\ &= \lambda_1\lambda_2 \mathrm{e}^{-\lambda_2 s} \int_0^s \mathrm{e}^{-(\lambda_1-\lambda_2)x}\,\mathrm{d}x \\ &= \lambda_1\lambda_2 \mathrm{e}^{-\lambda_2 s} \frac{1}{\lambda_1-\lambda_2}\left(1-\mathrm{e}^{-(\lambda_1-\lambda_2)s}\right) \\ &= \frac{\lambda_1}{\lambda_1-\lambda_2}\lambda_2 \mathrm{e}^{-\lambda_2 s} + \frac{\lambda_2}{\lambda_2-\lambda_1}\lambda_1 \mathrm{e}^{-\lambda_1 s}. \end{aligned}$$

- This obviously holds only for $\lambda_1 \neq \lambda_2$.

Example with Sums of Independent Exponential (cont.)

- Similar to the case shown above with geometric random variables, it can be proven by induction that

$$f_S(s) = \sum_{i=1}^{n} C_{i,n} \lambda_i \mathrm{e}^{-\lambda_i s}, \qquad C_{i,n} = \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i}.$$

- From this, the c.d.f. is clearly

$$F_S(s) = \int_0^s f_S(t)\, \mathrm{d}t = \sum_{i=1}^{n} C_{i,n} \int_0^s \lambda_i \mathrm{e}^{-\lambda_i t}\, \mathrm{d}t = \sum_{i=1}^{n} C_{i,n} \left(1 - \mathrm{e}^{-\lambda_i s}\right).$$

- See the text for an extension to $Q = \sum_{i=1}^{n} X_i$, where $X \overset{\text{ind}}{\sim} \mathrm{Gam}(a_i, b_i)$ and $a_i \in \mathbb{N}$.

Example with Sums of Uniform

Let $X_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}\,(0,1)$ and $S = X_1 + X_2$.

From the convolution formula and independence,

$$f_S\,(s) = \int_{-\infty}^{\infty} f_{X_1}\,(x_1)\, f_{X_2}\,(s - x_1)\; \mathrm{d}x_1 = \int_{-\infty}^{\infty} \mathbb{I}_{(0,1)}\,(x_1)\, \mathbb{I}_{(0,1)}\,(s - x_1)\; \mathrm{d}x_1.$$

The latter indicator function is $0 < s - x_1 < 1 \Leftrightarrow s - 1 < x_1 < s$, so that both conditions

$$\text{(a):}\quad 0 < x_1 < 1 \quad \text{and} \quad \text{(b):}\quad s - 1 < x_1 < s$$

must be satisfied. If $0 < s < 1$, the lower bound of (a) and the upper bound of (b) are binding, while, for $1 \le s < 2$, the opposite is true.

As there is no overlap between the two cases,

$$f_S\,(s) = \int_0^s \mathbb{I}_{(0,1)}\,(s)\; \mathrm{d}x_1 + \int_{s-1}^1 \mathbb{I}_{[1,2)}\,(s)\; \mathrm{d}x_1 = s\mathbb{I}_{(0,1)}\,(s) + (2-s)\,\mathbb{I}_{[1,2)}\,(s)\,,$$

which is a triangle when plotted.

Example with Sums of Uniform (cont.)

- Thus,

$$\Pr(X_1 + X_2 > 1) = \int_1^2 (2 - s)\, ds = \frac{1}{2} = \Pr(X_1 + X_2 \leq 1).$$

- Now let $S_n = \sum_{i=1}^{n} X_i$. Problem 2.11 shows that

$$f_{S_3}(s) = \frac{1}{2} s^2 \mathbb{I}_{(0,1)}(s) + \left(3s - s^2 - \frac{3}{2}\right) \mathbb{I}_{[1,2)}(s) + \frac{(s-3)^2}{2} \mathbb{I}_{[2,3)}(s),$$

so that

$$\Pr(S_3 \leq 1) = \int_0^1 f_{S_3}(s)\, ds = \int_0^1 \frac{1}{2} s^2\, ds = \frac{1}{6}.$$

Example with Sums of Uniform (cont.)

In general,

$$f_{S_n}(s) = \sum_{k=0}^{n} \binom{n}{k} (-1)^k \frac{(s-k)^{n-1} \mathbb{I}(s \geq k)}{(n-1)!},$$

and, to compute $\Pr(S_n \leq 1)$ (or its complement), just the term for $k = 0$ is needed, i.e.,

$$\Pr(S_n \leq 1) = \int_0^1 \binom{n}{0} (-1)^0 \frac{(s-0)^{n-1}}{(n-1)!} \mathrm{d}s = \frac{1}{n!},$$

which holds for $n \geq 0$.

Example with Sums of Uniform (cont.)

- Returning to the $n = 2$ case, we can also compute $\Pr(X_1 + X_2 \leq 1)$ as

$$\begin{aligned}\Pr(X_1 + X_2 \leq 1) &= \iint\limits_{x_1+x_2\leq 1} f_{X_1}(x_1)\, f_{X_2}(x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\ &= \int_0^1 \int_0^{1-x_2} \mathrm{d}x_1\, \mathrm{d}x_2 = \frac{1}{2} = \frac{1}{n!}.\end{aligned}$$

- For $n = 3$, $\Pr(X_1 + X_2 + X_3 \leq 1)$ is given by

$$\begin{aligned}&\iiint_{x_1+x_2+x_3\leq 1} f_{X_1}(x_1)\, f_{X_2}(x_2)\, f_{X_3}(x_3)\, \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}x_3 \\ &= \int_0^1 \int_0^{1-x_3} \int_0^{1-x_2-x_3} \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}x_3 = \frac{1}{6} = \frac{1}{n!}.\end{aligned}$$

Example with Sums of Uniform (cont.)

- Now let $N = \min\{n : S_n > 1\}$, i.e., N is the smallest number of U_i which need to be summed such that the sum exceeds one.
- To compute $\mathbb{E}[N]$, use the result (shown below)

$$\mathbb{E}[N] = \sum_{n=1}^{\infty} \Pr(N \geq n) = \sum_{n=0}^{\infty} \Pr(N > n).$$

- Observe that event $\{N > n\} = \{S_n \leq 1\}$ and above we saw that $\Pr(S_n \leq 1) = 1/n!$ so that

$$\mathbb{E}[N] = \sum_{n=0}^{\infty} \Pr(N > n) = \sum_{n=0}^{\infty} 1/n! = \mathrm{e}.$$

- More generally, with $N(x) = \min\{n : S_n > x\}$, $\mathbb{E}[N(x)] = \mathrm{e}^x$; see, e.g., Ross (2006, p. 375, 416) [A First Course in Probability] for two other methods of proof.

Expectation Result: Discrete

To see that $\mathbb{E}[X] = \sum_{i=1}^{\infty} \Pr(X \geq i)$ for X a nonnegative integer r.v., write

$$X = \sum_{i=1}^{\infty} \mathbb{I}(X \geq i) \qquad \text{so that} \qquad \mathbb{E}[X] = \sum_{i=1}^{\infty} \Pr(X \geq i),$$

where the interchange of $\mathbb{E}$ and $\sum_{i=1}^{\infty}$ is justified via the monotone convergence theorem; see Gut (2005, Corollary 5.2, page 56).

Expectation Result: Continuous

Analogous to the discrete case, if X is continuous and non-negative, then $\mathbb{E}[X] = \int_0^\infty \Pr(X > x)\, \mathrm{d}x$. To see this, write

$$\begin{aligned}\int_0^\infty (1 - F_X(x))\, \mathrm{d}x &= \int_0^\infty \int_x^\infty f_X(t)\, \mathrm{d}t\, \mathrm{d}x \\ &= \int_0^\infty \left(\int_0^t \mathrm{d}x \right) f_X(t)\, \mathrm{d}t = \int_0^\infty t f_X(t)\, \mathrm{d}t.\end{aligned}$$

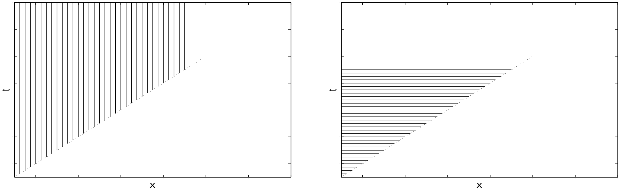


Figure: Left: the x, t surface over which we integrate f; outer integral is w.r.t. x and for a fixed x (signified by the vertical lines) the inner integral over t goes from x to ∞. Right: same but with the outer integral w.r.t. t: for a fixed t (horizontal lines), the inner integral over x goes from 0 to t.

Student's t Distribution

- The ratio of a standard normal r.v. to the square root of an independent chi-square r.v. divided by its degrees of freedom follows a Student's t distribution.
- That is, with $Z \sim \mathrm{N}(0,1)$ independent of $C \sim \chi^2_\nu$, and defining $X = \sqrt{C/\nu}$, we have $T = Z/X \sim t_\nu$.
- Recall: If X is a continuous random variable with p.d.f. f_X and g is a continuous differentiable function with domain contained in the range of X and $\mathrm{d}g/\mathrm{d}x \neq 0 \; \forall x \in \mathcal{S}_X$, then f_Y, the pdf of $Y = g(X)$, can be calculated by

$$f_Y(y) = f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|$$

where $x = g^{-1}(y)$ is the inverse function of Y.

Student's t Distribution (cont.)

Then, as $C \sim \chi^2_\nu$ has p.d.f.

$$f_C(c;\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} c^{\nu/2-1} e^{-c/2} \mathbb{I}_{(0,\infty)}(c),$$

using the substitution $c = \nu x^2,\ \mathrm{d}c/\,\mathrm{d}x = 2\nu x$ gives

$$\begin{aligned} f_X(x;\nu) &= \left|\frac{\mathrm{d}c}{\mathrm{d}x}\right| f_C(c) = 2\nu x \cdot \frac{2^{-\nu/2}}{\Gamma(\nu/2)} \left(\nu x^2\right)^{\nu/2-1} e^{-\left(\nu x^2\right)/2} \mathbb{I}_{(0,\infty)}\left(\nu x^2\right) \\ &= \frac{2^{-\nu/2+1}\nu^{\nu/2}}{\Gamma(\nu/2)} x^{\nu-1} e^{-\left(\nu x^2\right)/2} \mathbb{I}_{(0,\infty)}(x). \end{aligned}$$

Student's t Distribution (cont.)

So, with

$$f_X(x;\nu) = \frac{2^{-\nu/2+1}\nu^{\nu/2}}{\Gamma(\nu/2)} x^{\nu-1} e^{-\left(\nu x^2\right)/2} \mathbb{I}_{(0,\infty)}(x)$$

and

$$f_Z(z;0,1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\},$$

the integral formula for ratios gives

$$\begin{aligned} f_T(t;\nu) &= \int_{-\infty}^{\infty} |x|\, f_Z(tx)\, f_X(x)\, \mathrm{d}x \\ &= K \int_0^{\infty} x \exp\left[-\frac{1}{2}(tx)^2\right] x^{\nu-1} e^{-\left(\nu x^2\right)/2}\, \mathrm{d}x \\ &= K \int_0^{\infty} x^{\nu} \exp\left(-\frac{1}{2}x^2\left(t^2+\nu\right)\right) \mathrm{d}x, \end{aligned}$$

where $K := \left[2^{-\nu/2+1}\nu^{\nu/2}\right] / \left[\sqrt{2\pi}\,\Gamma(\nu/2)\right]$.

Student's t Distribution (cont.)

Substituting $y = x^2$, $\mathrm{d}x/\mathrm{d}y = y^{-1/2}/2$,

$$\begin{aligned} f_T(t;\nu) &= K \int_0^\infty x^{2\frac{\nu}{2}} \mathrm{e}^{-\left(x^2/2\right)\left(\nu+t^2\right)} \,\mathrm{d}x = K \int_0^\infty y^{\frac{\nu}{2}} \mathrm{e}^{-y\frac{1}{2}\left(\nu+t^2\right)} \frac{1}{2} y^{-1/2} \,\mathrm{d}y \\ &= \frac{K}{2} \int_0^\infty y^{\frac{\nu-1}{2}} \mathrm{e}^{-y\frac{1}{2}\left(\nu+t^2\right)} \,\mathrm{d}y = \frac{K}{2} \int_0^\infty y^{\frac{\nu+1}{2}-1} \mathrm{e}^{-y\frac{1}{2}\left(\nu+t^2\right)} \,\mathrm{d}y \\ &= \frac{K}{2} \Gamma\left(\frac{\nu+1}{2}\right) \left[\frac{1}{2}\left(\nu+t^2\right)\right]^{-(\nu+1)/2}, \end{aligned}$$

obtained by setting $z = (\nu + t^2)/2$ and $h = (\nu + 1)/2$ and using the gamma p.d.f. relation

$$\int_0^\infty \mathrm{e}^{-zx} x^{h-1} \,\mathrm{d}x = \Gamma(h)\, z^{-h}.$$

Student's t Distribution (cont.)

Simplifying expression

$$f_T(t;\nu) = \frac{K}{2}\Gamma\left(\frac{\nu+1}{2}\right)\left[\frac{1}{2}\left(\nu+t^2\right)\right]^{-(\nu+1)/2}$$

with $K = \left[2^{-\nu/2+1}\nu^{\nu/2}\right] / \left[\sqrt{2\pi}\Gamma\left(\nu/2\right)\right]$ gives

$$f_T\left(t;\nu\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\nu^{\nu/2}}{\sqrt{\pi}\Gamma\left(\nu/2\right)}\left[\nu+t^2\right]^{-(\nu+1)/2} = \frac{\nu^{-1/2}}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}\left[1+t^2/\nu\right]^{-(\nu+1)/2},$$

which is defined to be the Student's t density with ν degrees of freedom.

Student's t Distribution (cont.)

Another way of deriving the t is by using the conditioning result which states that, for some event E,

$$\Pr(E) = \int \Pr(E \mid X = x)\, f_X(x)\, \mathrm{d}x.$$

For $T = Z/X$,

$$\begin{aligned} f_T(t;\nu) &= F_T'(t) = \frac{\mathrm{d}}{\mathrm{d}t} \Pr\left(Z < t\sqrt{\chi^2_\nu/\nu}\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \int \Pr(Z < tX \mid X = x)\, f_X(x)\, \mathrm{d}x \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \int \Pr(Z < tx)\, f_X(x)\, \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}t} \int \int_{-\infty}^{tx} f_Z(z)\, \mathrm{d}z \cdot f_X(x)\, \mathrm{d}x. \end{aligned}$$

Student's t Distribution (cont.)

Substituting,

$$\begin{aligned} f_T(t;\nu) &= K\frac{\mathrm{d}}{\mathrm{d}t}\int_0^\infty \int_{-\infty}^{tx} e^{-\frac{1}{2}z^2}\,\mathrm{d}z \cdot x^{\nu-1}\mathrm{e}^{-\left(\nu x^2\right)/2}\,\mathrm{d}x \\ &= K\int_0^\infty x^{\nu-1}\mathrm{e}^{-\left(\nu x^2\right)/2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{tx}\mathrm{e}^{-\frac{1}{2}z^2}\,\mathrm{d}z\,\mathrm{d}x \\ &= K\int_0^\infty x^{\nu-1}\mathrm{e}^{-\left(\nu x^2\right)/2}\cdot\left[\mathrm{e}^{-\frac{1}{2}(tx)^2}x - \mathrm{e}^{-\infty}\cdot 0\right]\mathrm{d}x \\ &= K\int_0^\infty x^{\nu}\mathrm{e}^{-\left(x^2/2\right)\left(\nu+t^2\right)}\,\mathrm{d}x \end{aligned}$$

from Leibniz' rule. This is the same integral as above.
Leibniz' rule states that, if $I = \int_{\ell(z)}^{h(z)} f(s,z)\,\mathrm{d}s$, then

$$\frac{\partial I}{\partial z} = \int_{\ell(z)}^{h(z)}\frac{\partial f}{\partial z}\,\mathrm{d}s + f\left(h(z),z\right)\frac{\mathrm{d}h}{\mathrm{d}z} - f\left(\ell(z),z\right)\frac{\mathrm{d}\ell}{\mathrm{d}z}.$$

Exercise: Cauchy Distribution

- Let $X, Y \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0,1)$.
- Consider the distribution of the ratio $R = X/Y$.
- Use the quotient formula

$$f_R(r) = \int_{-\infty}^{\infty} \frac{|x|}{r^2} f_{X,Y}\left(x, \frac{x}{r}\right) \mathrm{d}x$$

to derive the p.d.f. of R.

- Hint: Let $k = r^2/(1+r^2)$ and then $u = \frac{1}{2}(x^2/k)$. This leads to

$$f_R(r) = \frac{1}{\pi} \frac{1}{1+r^2},$$

so that R is a Cauchy random variable.

Approximating the Mean and Variance of Functions of $\mathbf{X}$

- Let g be a function of $X_1, \ldots, X_n$.
- There are many situations in which the exact mean and variance of g is difficult to derive.
- Thus, it is desirable to have an approximation to them.
- These could for example be used with a normal distribution.
- Method: Use first few terms of a Taylor series expansion of g.

Approximating the Mean

- Let $g(\mathbf{x}) := g(x_1, x_2)$ be a bivariate continuous function which is defined on an open neighborhood of $\mathbf{x}^0 = \left(x_1^0, x_2^0\right) \in \mathbb{R}^2$ and whose first and second derivatives exist.
- For convenience, for $i = 1, 2$, let

$$\triangle_i = \left(x_i - x_i^0\right), \qquad \dot{g}_i(\mathbf{x}) = \frac{\partial g(\mathbf{x})}{\partial x_i}, \qquad \ddot{g}_{ij}(\mathbf{x}) = \frac{\partial^2 g(\mathbf{x})}{\partial x_i \partial x_j}.$$

- The two-term Taylor series approximation for $g(\mathbf{x})$ is

$$\begin{aligned} g(\mathbf{x}) \quad \approx \quad & g\left(\mathbf{x}^0\right) + \triangle_1 \dot{g}_1\left(\mathbf{x}^0\right) + \triangle_2 \dot{g}_2\left(\mathbf{x}^0\right) \\ & + \frac{1}{2} \triangle_1^2 \, \ddot{g}_{11}\left(\mathbf{x}^0\right) + \frac{1}{2} \triangle_2^2 \, \ddot{g}_{22}\left(\mathbf{x}^0\right) + \triangle_1 \triangle_2 \, \ddot{g}_{12}\left(\mathbf{x}^0\right). \end{aligned}$$

Approximating the Mean (2)

- Now let $\mathbf{X} = (X_1, X_2)$ be a bivariate r.v. with mean $\mathbf{x}^0 = (\mu_{X_1}, \mu_{X_2})$. Taking expectations gives

$$\begin{aligned}
\mathbb{E}\left[\triangle_i \dot{g}_i\left(\mathbf{x}^0\right)\right] &= \dot{g}_i\left(\mathbf{x}^0\right)\mathbb{E}\left[\left(X_i - \mu_{X_i}\right)\right] = 0, \\
\mathbb{E}\left[\triangle_i^2 \ddot{g}_{ii}\left(\mathbf{x}^0\right)\right] &= \ddot{g}_{ii}\left(\mathbf{x}^0\right)\mathbb{E}\left[\left(X_i - \mu_{X_i}\right)^2\right] = \ddot{g}_{ii}\left(\mathbf{x}^0\right)\mathbb{V}\left(X_i\right), \\
\mathbb{E}\left[\triangle_i \triangle_j \ddot{g}_{ij}\left(\mathbf{x}^0\right)\right] &= \ddot{g}_{ij}\left(\mathbf{x}^0\right)\mathbb{E}\left[\left(X_i - \mu_{X_i}\right)\left(X_j - \mu_{X_j}\right)\right] \\
&= \ddot{g}_{ij}\left(\mathbf{x}^0\right)\mathrm{Cov}\left(X_i, X_j\right), \quad i, j = 1, 2.
\end{aligned}$$

- Thus, the second–order Taylor series approximation of $\mathbb{E}\left[g\left(\mathbf{X}\right)\right]$ is

$$\begin{aligned}
\mathbb{E}\left[g\left(\mathbf{X}\right)\right] \approx\ & g\left(\mathbf{x}^0\right) + \frac{1}{2}\ddot{g}_{11}\left(\mathbf{x}^0\right)\mathbb{V}\left(X_1\right) + \frac{1}{2}\ddot{g}_{22}\left(\mathbf{x}^0\right)\mathbb{V}\left(X_2\right) \\
& + \ddot{g}_{12}\left(\mathbf{x}^0\right)\mathrm{Cov}\left(X_1, X_2\right).
\end{aligned}$$

Approximating the Variance

- To approximate the variance $\mathbb{V}(g(\mathbf{X})) = \mathbb{E}\left[(g(\mathbf{X}) - \mathbb{E}[g(\mathbf{X})])^2\right]$ first replace $\mathbb{E}[g(\mathbf{X})]$ with just the zero–order term $g(\mathbf{x}^0)$ from its Taylor series expansion, to get

$$\mathbb{E}\left[(g(\mathbf{X}) - \mathbb{E}[g(\mathbf{X})])^2\right] \approx \mathbb{E}\left[\left(g(\mathbf{X}) - g\left(\mathbf{x}^0\right)\right)^2\right].$$

- Use $g(\mathbf{X}) \approx g\left(\mathbf{x}^0\right) + \triangle_1 \dot{g}_1\left(\mathbf{x}^0\right) + \triangle_2 \dot{g}_2\left(\mathbf{x}^0\right)$ to get

$$\begin{aligned}\mathbb{E}\left[\left(g(\mathbf{X}) - g\left(\mathbf{x}^0\right)\right)^2\right] &\approx \mathbb{E}\left[\left(\triangle_1 \dot{g}_1\left(\mathbf{x}^0\right) + \triangle_2 \dot{g}_2\left(\mathbf{x}^0\right)\right)^2\right] \\ &= \dot{g}_1^2\left(\mathbf{x}^0\right)\mathbb{E}\left[\triangle_1^2\right] + \dot{g}_2^2\left(\mathbf{x}^0\right)\mathbb{E}\left[\triangle_2^2\right] + 2\dot{g}_1\left(\mathbf{x}^0\right)\dot{g}_2\left(\mathbf{x}^0\right)\mathbb{E}\left[\triangle_1\triangle_2\right].\end{aligned}$$

$$\mathbb{V}(g(\mathbf{X})) \approx \dot{g}_1^2\left(\mathbf{x}^0\right)\mathbb{V}(X_1) + \dot{g}_2^2\left(\mathbf{x}^0\right)\mathbb{V}(X_2) + 2\dot{g}_1\left(\mathbf{x}^0\right)\dot{g}_2\left(\mathbf{x}^0\right)\mathrm{Cov}(X_1, X_2).$$

Approximations for Ratios

- Most prominent application is for $g(x,y) = x/y$.
- With $\dot{g}_1(x,y) = y^{-1}$, $\ddot{g}_{11}(x,y) = 0$, $\ddot{g}_{12}(x,y) = -y^{-2}$, $\dot{g}_2(x,y) = -xy^{-2}$, $\ddot{g}_{22}(x,y) = 2xy^{-3}$, and letting $\sigma_X^2 = \mathbb{V}(X)$, $\sigma_Y^2 = \mathbb{V}(Y)$ and $\sigma_{X,Y} = \mathrm{Cov}(X,Y)$, the approximations reduce to

$$\mathbb{E}\left[\frac{X}{Y}\right] \approx \frac{\mu_X}{\mu_Y}\left(1 + \frac{\sigma_Y^2}{\mu_Y^2} - \frac{\sigma_{X,Y}}{\mu_X \mu_Y}\right)$$

and $\mathbb{V}(X/Y) \approx \mu_Y^{-2}\sigma_X^2 + \mu_X^2\mu_Y^{-4}\sigma_Y^2 - 2\mu_X\mu_Y^{-3}\sigma_{X,Y}$, or

$$\mathbb{V}\left(\frac{X}{Y}\right) \approx \left(\frac{\mu_X}{\mu_Y}\right)^2 \left(\frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2} - \frac{2\sigma_{X,Y}}{\mu_X \mu_Y}\right).$$

Approximations for Reciprocals

- It is easy to derive moment approximations of $1/Y$ by setting $\mu_X = 1$ and $\sigma_X^2 = 0$, i.e., X is the degenerate r.v. $X = 1$ and $\sigma_{X,Y} = 0$.
- In particular,

$$\mathbb{E}\left[\frac{1}{Y}\right] \approx \frac{1}{\mu_Y} + \frac{\sigma_Y^2}{\mu_Y^3} \quad \text{and} \quad \mathbb{V}\left(\frac{1}{Y}\right) \approx \frac{\sigma_Y^2}{\mu_Y^4}.$$

- Thus, while $\mathbb{E}[1/Y] \neq 1/\mathbb{E}[Y]$, we see that it is the zero–order approximation.

Approximations for Products

- Another popular application is with $g(x, y) = xy$, in which case

$$\mathbb{E}[XY] \approx \mu_X \mu_Y + \sigma_{X,Y}$$

and

$$\mathbb{V}(XY) \approx \mu_Y^2 \sigma_X^2 + \mu_X^2 \sigma_Y^2 + 2\mu_Y \mu_X \sigma_{X,Y}.$$

- In this case, exact results are easy to obtain: Taking expectation of the identity

$$XY = \mu_X \mu_Y + (X - \mu_X)\mu_Y + (Y - \mu_Y)\mu_X + (X - \mu_X)(Y - \mu_Y)$$

immediately gives

$$\mathbb{E}[XY] = \mu_X \mu_Y + \mathrm{Cov}(X, Y).$$

Approximations for Products (cont.)

- For the variance, let $c_{ij} := \mathbb{E}\left[(X-\mu_X)^i (Y-\mu_Y)^j \right]$. Then

  $$\mathbb{V}(XY) = \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + 2\mu_X\mu_Y\, c_{11} - c_{11}^2 + c_{22} + 2\mu_Y c_{21} + 2\mu_X c_{12}$$

  provided that $\mathbb{V}(XY)$ exists.
- Thus, the approximation to the expected value is exact, while the approximate variance

  $$\mathbb{V}(XY) \approx \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + 2\mu_Y\mu_X\sigma_{X,Y}$$

  is missing several higher order terms.
- If X and Y are independent, then $c_{11} = c_{12} = c_{21} = 0$ and $c_{22} = \mathbb{V}(X)\,\mathbb{V}(Y)$, so that

  $$X \perp Y \quad \Rightarrow \quad \mathbb{E}[XY] = \mu_X\mu_Y,\ \mathbb{V}(XY) = \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + \sigma_X^2\sigma_Y^2.$$

Example with Student's t

- Let $C \sim \chi^2_k$ and $Y = 1/\sqrt{C/k} = k^{1/2}C^{-1/2}$.
- Then $Y^2 = k/C$ and $\mathbb{E}\left[Y^2\right] = k\mathbb{E}\left[C^{-1}\right] = k/\left(k-2\right)$ (shown below) for $k > 2$.
- Now, with $X \sim \mathrm{N}\left(0,1\right)$ independent of C, we know that

$$T = \frac{X}{\sqrt{C/k}} = XY$$

  is student's t with k degrees of freedom.
- Above we saw that

$$X \perp Y \quad \Rightarrow \quad \mathbb{E}\left[XY\right] = \mu_X\mu_Y, \ \mathbb{V}\left(XY\right) = \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + \sigma_X^2\sigma_Y^2$$

  so, with $\mu_X = 0$ and $\sigma_X^2 = 1$, it follows that $\mathbb{E}\left[T\right] = 0$ and

$$\mathbb{V}\left(T\right) = \mathbb{V}\left(XY\right) = \mu_Y^2 + 0 + \sigma_Y^2 = \mathbb{E}\left[Y^2\right] = \frac{k}{k-2}, \ k > 2.$$

Moments of χ^2

- Let $X \sim \chi^2_k$ with p.d.f.

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} \mathbb{I}_{(0,\infty)}(x)$$

- To compute $\mathbb{E}[X^s]$ for $X \sim \chi^2_k$, let $u = x/2$, so for $k/2 + s > 0$,

$$\int_0^\infty x^{k/2+s-1} e^{-x/2}\, dx = 2^{k/2+s} \int_0^\infty u^{k/2+s-1} e^{-u}\, du = 2^{k/2+s}\Gamma(k/2+s)$$

and

$$\mathbb{E}[X^s] = 2^s \frac{\Gamma(k/2+s)}{\Gamma(k/2)}, \quad s > -\frac{k}{2}.$$

- Then $\mathbb{E}[X] = k$ and $\mathbb{E}\left[X^2\right] = (k+2)\,k$, so that $\mathbb{V}(X) = 2k$.
- Also, as needed above,

$$\mathbb{E}\left[X^{-1}\right] = 1/(k-2), \qquad k > 2.$$

Extension to *n* Random Variables

Recall for $\mathbf{X} = (X_1, X_2)$ the approximations

$$\begin{aligned}\mathbb{E}\left[g\left(\mathbf{X}\right)\right] \approx\ & g\left(\mathbf{x}^0\right) + \frac{1}{2}\ddot{g}_{11}\left(\mathbf{x}^0\right)\mathbb{V}\left(X_1\right) + \frac{1}{2}\ddot{g}_{22}\left(\mathbf{x}^0\right)\mathbb{V}\left(X_2\right) \\ & + \ddot{g}_{12}\left(\mathbf{x}^0\right)\mathrm{Cov}\left(X_1, X_2\right).\end{aligned}$$

and

$$\mathbb{V}\left(g\left(\mathbf{X}\right)\right) \approx \dot{g}_1^2\left(\mathbf{x}^0\right)\mathbb{V}\left(X_1\right) + \dot{g}_2^2\left(\mathbf{x}^0\right)\mathbb{V}\left(X_2\right) + 2\dot{g}_1\left(\mathbf{x}^0\right)\dot{g}_2\left(\mathbf{x}^0\right)\mathrm{Cov}\left(X_1, X_2\right).$$

These are easily extended to the case of *n* random variables.

Extension to n Random Variables

With $\mathbf{x}^0 = \mathbb{E}[\mathbf{X}] = (\mu_{X_1}, \ldots, \mu_{X_n})$,

$$\mathbb{E}\left[g\left(\mathbf{X}\right)\right] \approx g\left(\mathbf{x}^0\right) + \frac{1}{2}\sum_{i=1}^{n} \ddot{g}_{ii}\left(\mathbf{x}^0\right)\mathbb{V}\left(X_i\right) + \sum_{i=1}^{n}\sum_{j=i+1}^{n} \ddot{g}_{ij}\left(\mathbf{x}^0\right)\mathrm{Cov}\left(X_i, X_j\right)$$

and

$$\mathbb{V}\left(g\left(\mathbf{X}\right)\right) \approx \sum_{i=1}^{n} \dot{g}_i^2\left(\mathbf{x}^0\right)\mathbb{V}\left(X_i\right) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \dot{g}_i\left(\mathbf{x}^0\right)\dot{g}_j\left(\mathbf{x}^0\right)\mathrm{Cov}\left(X_i, X_j\right).$$

Example for Ratios

- Let X_i be i.i.d. r.v.s with finite mean. Define $S := \sum_{i=1}^{n} X_i$ and $R_i := X_i / S$, $i = 1, \ldots, n$. It follows that

$$\sum_{i=1}^{n} R_i = 1 = \mathbb{E}[1] = \mathbb{E}\left[\sum_{i=1}^{n} R_i\right] = n\mathbb{E}\left[R_1\right], \quad \text{so} \quad \mathbb{E}\left[R_i\right] = n^{-1}.$$

- If the X_i are positive r.v.s (or negative r.v.s), then $0 < R_i < 1$, and the expectation exists.
- Now let the X_i be i.i.d. positive r.v.s, and let λ_i, $i = 1, \ldots, n$, be a set of known constants. The expectation of

$$R := \frac{\sum_{i=1}^{n} \lambda_i X_i}{\sum_{i=1}^{n} X_i} = \frac{\sum_{i=1}^{n} \lambda_i X_i}{S} = \sum_{i=1}^{n} \lambda_i R_i$$

is

$$\mathbb{E}\left[R\right] = \sum_{i=1}^{n} \lambda_i \mathbb{E}\left[R_i\right] = n^{-1} \sum_{i=1}^{n} \lambda_i =: \bar{\lambda}.$$

Example for Ratios (cont.)

A very useful special case is when $X_i \overset{\text{i.i.d.}}{\sim} \chi^2_1$, which arises in the study of ratios of quadratic forms. As $\mathbb{E}\left[X_i\right] = 1$,

$$\mathbb{E}\left[R\right] = \frac{\sum_{i=1}^{n} \lambda_i}{n} = \frac{\mathbb{E}\left[\sum_{i=1}^{n} \lambda_i X_i\right]}{\mathbb{E}\left[\sum_{i=1}^{n} X_i\right]},$$

and the expectation of the ratio R is, exceptionally, the ratio of expectations!

This clearly holds for any set of positive i.i.d. r.v.s with finite mean.

Example for Ratios (cont.)

For example, with $\mathrm{Unif}(0, 5)$ r.v.s, we can simulate the ratio with the following code and confirm the result.

```
lam=-3:4; n=length(lam); lambar=mean(lam)
sim=10000;
rmeanvec=zeros(sim,1); numvec=rmeanvec; denvec=numvec;
for s=1:sim
   %X=randn(n,1).^2; % for chi2
   X=unifrnd(0,5,n,1);
   nn=dot(lam,X); dd=sum(X); rr=nn/dd;
   rmeanvec(s)=rr; numvec(s)=nn; denvec(s)=dd;
end
mean(rmeanvec), mean(numvec)/mean(denvec)
```

Running this yields:

```
lambar =  0.500000000000000
ans = 0.503105917206093
ans = 0.502629397239206
```

Example for Ratios (cont.)

- Let's compute the approximations to the mean and variance.
- Use $g(\mathbf{X}) = R$ with $X_i \overset{\text{i.i.d.}}{\sim} \chi^2_1$. Verify that

$$\dot{g}_i(\mathbf{x}) = \frac{S\lambda_i - \sum_{i=1}^n \lambda_i x_i}{S^2}, \quad \ddot{g}_{ii}(\mathbf{x}) = 2\frac{\left(\sum_{i=1}^n \lambda_i x_i\right) - S\lambda_i}{S^3}.$$

- Noting that S evaluated at $\mathbf{x}_0 = (1, \ldots, 1)$ is n, these yield $\dot{g}_i(\mathbf{x}_0) = (\lambda_i - \bar{\lambda})/n$ and $\ddot{g}_{ii}(\mathbf{x}_0) = 2(\bar{\lambda} - \lambda_i)/n^2$.
- Thus, with $\mathbb{V}(X_i) = 2$, we get

$$\mathbb{E}[R] \approx \bar{\lambda} + \frac{1}{n^2}\sum_{i=1}^n \left(\bar{\lambda} - \lambda_i\right) = \bar{\lambda} = \mathbb{E}[R],$$

  showing that the approximation in this case is exact.
- For the variance, we get $\mathbb{V}(R) \approx 2n^{-2}\sum_{i=1}^n \left(\lambda_i - \bar{\lambda}\right)^2$.
- Exact value just replaces n^2 by $n(n+2)$.

Intermediate Probability: A Computational Approach

Matrix Algebra

We assume that the reader is already familiar with vector and matrix notation, transpose, symmetric matrices, matrix addition and multiplication, diagonal matrices, computation of determinant and matrix inverses, rank, span, and solutions of systems of linear equations. A very useful resource is Abadir and Magnus' book: Matrix algebra.

If $\mathbf{A}$ is an $n \times n$ symmetric real matrix, then $\mathbf{A}$ is said to be *positive definite*, denoted $\mathbf{A} > 0$, if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n \setminus \mathbf{0}$. If $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$, then $\mathbf{A}$ is said to be *positive semi-definite*, denoted $\mathbf{A} \geq 0$.

If $\mathbf{A}$ is a square matrix, then its trace, $\mathrm{tr}(\mathbf{A})$, is the sum of its diagonal elements. A useful fact is that, for $\mathbf{B}$ an $n \times m$ matrix $\mathbf{C}$ an $m \times n$ matrix,

$$\mathrm{tr}(\mathbf{BC}) = \mathrm{tr}(\mathbf{CB}).$$

Now let $\mathbf{B}$ and $\mathbf{C}$ be $n \times n$ matrices. The determinant of the product is

$$\det(\mathbf{BC}) = \det(\mathbf{B})\det(\mathbf{C}) = \det(\mathbf{CB}).$$

Matrix Algebra

Let $\mathbf{A}$ be an $n \times n$ matrix. The *principal minors* of order k, $1 \leq k \leq n$, are determinants of all the $k \times k$ matrices obtained by deleting the same $n - k$ rows and columns of $\mathbf{A}$.

The *leading principal minor* of order k, $1 \leq k \leq n$, is the principal minor obtained by deleting the first k rows and columns of $\mathbf{A}$.

If $\mathbf{A} > 0$ then all the leading principal minors are positive.

This result gives a method for determining if a matrix is positive definite without having to compute the eigenvalues.

Matrix Algebra

For matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, there are at most n distinct roots of the *characteristic equation* $|\lambda \mathbf{I}_n - \mathbf{A}| = 0$, and these values are referred to as the *eigenvalues* of $\mathbf{A}$, the set of which is denoted $\mathrm{Eig}(\mathbf{A})$.

If two or more roots coincide, then we say that the eigenvalues have multiplicities.

Each matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ has exactly n eigenvalues, counting multiplicities. For example, the identity matrix $\mathbf{I}_n$ has one unique eigenvalue (unity), but has n eigenvalues, counting multiplicities.

Denote the n eigenvalues, counting multiplicities, of matrix $\mathbf{A}$ as $\lambda_1, \ldots, \lambda_n$. Vector $\mathbf{x}$ is a (column) *eigenvector* of eigenvalue λ if it satisfies $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$.

Eigenvectors are usually normalized to have norm one, i.e., $\mathbf{x}'\mathbf{x} = 1$.

Matrix Algebra

Of great importance is the fact that eigenvalues of a symmetric matrix are real. If, in addition, $\mathbf{A} > 0$, then all its eigenvalues are positive. This follows because $\mathbf{Ax} = \lambda\mathbf{x} \Rightarrow \mathbf{x}'\mathbf{Ax} = \lambda\mathbf{x}'\mathbf{x}$ and both $\mathbf{x}'\mathbf{Ax}$ and $\mathbf{x}'\mathbf{x}$ are positive.

The square matrix $\mathbf{U}$ with columns $\mathbf{u}_1, \dots, \mathbf{u}_n$ is *orthogonal* if $\mathbf{U}'\mathbf{U} = \mathbf{I}$.

If the $\mathbf{u}_i$ are orthonormal, then it is obvious that $\mathbf{U} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \mathbf{u}_n)$ satisfies $\mathbf{U}'\mathbf{U} = \mathbf{I}$, but not immediately clear that $\mathbf{UU}' = \mathbf{I}$. To prove this:

$$\mathbf{U}'\mathbf{U} = \mathbf{I} \Rightarrow \det(\mathbf{U}')\det(\mathbf{U}) = \det(\mathbf{U}'\mathbf{U}) = \det(\mathbf{I}) = 1 \Rightarrow \det(\mathbf{U}) \neq 0 \Rightarrow \exists \mathbf{U}^{-1}.$$

Then

$$\mathbf{U}' = \mathbf{U}'\left(\mathbf{UU}^{-1}\right) = (\mathbf{U}'\mathbf{U})\,\mathbf{U}^{-1} = \mathbf{IU}^{-1} = \mathbf{U}^{-1},$$

and in particular, $\mathbf{UU}' = \mathbf{I}$.

Matrix Algebra

As a special case of *Schur's Decomposition Theorem*, if $\mathbf{A}$ is an $n \times n$ symmetric matrix, then there exists an orthogonal matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ such that $\mathbf{A} = \mathbf{UDU}'$. We refer to this as the *spectral decomposition*.

It implies that $\mathbf{AU} = \mathbf{UD}$ or $\mathbf{Au}_i = \lambda_i \mathbf{u}_i$, $i = 1, \ldots, n$, i.e., $\mathbf{u}_i$ is an eigenvector of $\mathbf{A}$ corresponding to eigenvalue λ_i. The theorem is easy to verify if the n eigenvalues have no multiplicities, i.e., they are mutually distinct:

Let λ_i be an eigenvalue of $\mathbf{A}$ associated with normed eigenvector $\mathbf{u}_i$, so $\mathbf{Au}_i = \lambda_i \mathbf{u}_i$. Then, for $i \neq j$, as $\mathbf{u}_i' \mathbf{Au}_j$ is a scalar and $\mathbf{A}$ is symmetric, $\mathbf{u}_i' \mathbf{Au}_j = \mathbf{u}_j' \mathbf{Au}_i$, but $\mathbf{u}_i' \mathbf{Au}_j = \lambda_j \mathbf{u}_i' \mathbf{u}_j$ and $\mathbf{u}_j' \mathbf{Au}_i = \lambda_i \mathbf{u}_j' \mathbf{u}_i$, i.e., $\lambda_i \mathbf{u}_i' \mathbf{u}_j = \lambda_j \mathbf{u}_i' \mathbf{u}_j$.

Thus $\mathbf{u}_i' \mathbf{u}_j = 0$ because $\lambda_i \neq \lambda_j$, i.e., the eigenvectors are orthogonal to one another and $\mathbf{U} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \mathbf{u}_n)$ is an orthogonal matrix.

Matrix Algebra

Let the $n \times n$ matrix $\mathbf{A}$ have spectral decomposition $\mathbf{UDU}'$ (with $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and λ_i are the eigenvalues of $\mathbf{A}$). Then

$$\mathrm{tr}\,(\mathbf{A}) = \mathrm{tr}\,(\mathbf{UDU}') = \mathrm{tr}\,(\mathbf{U}'\mathbf{UD}) = \mathrm{tr}\,(\mathbf{D}) = \sum_{i=1}^{n} \lambda_i.$$

and

$$\det\,(\mathbf{A}) = \det\,(\mathbf{UDU}') = \det\,(\mathbf{U}'\mathbf{UD}) = \det\,(\mathbf{D}) = \prod_{i=1}^{n} \lambda_i,$$

recalling that the determinant of a diagonal matrix is the product of the diagonal elements.
Thus, if $\mathbf{A}$ is not full rank, i.e., is *singular*, then $|\mathbf{A}| = 0$, implying that at least one eigenvalue is zero.

Matrix Algebra

For $n \times n$ matrix $\mathbf{A} \geq 0$ with spectral decomposition $\mathbf{A} = \mathbf{UDU}'$, the *Cholesky decomposition* of $\mathbf{A}$, denoted $\mathbf{A}^{1/2}$, is a matrix such that $\mathbf{A} = \mathbf{A}^{1/2}\mathbf{A}^{1/2}$. It can be computed as $\mathbf{A}^{1/2} = \mathbf{UD}^{1/2}\mathbf{U}'$, where $\mathbf{D}^r := \operatorname{diag}(\lambda_1^r, \ldots, \lambda_n^r)$ for $r \in \mathbb{R}_{>0}$, and for $r = 1/2$, $\lambda_i^{1/2}$ is the nonnegative square root of λ_i. Indeed,

$$\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{UD}^{1/2}\mathbf{U}'\,\mathbf{UD}^{1/2}\mathbf{U}' = \mathbf{UD}^{1/2}\mathbf{D}^{1/2}\mathbf{U}' = \mathbf{UDU}'.$$

Notice that, if $\mathbf{A} > 0$, then $\mathbf{A}^{1/2} > 0$.

If $\mathbf{A} > 0$, then $\min(\lambda_i) > 0$ and r can be any real number. In particular, for $r = -1$, $\mathbf{A}^{-1} = \mathbf{UD}^{-1}\mathbf{U}'$. The notation $\mathbf{A}^{-1/2}$ refers to the inverse of $\mathbf{A}^{1/2}$.

Matrix Algebra

The *rank* of $m \times n$ matrix $\mathbf{A}$ is the number of linearly independent columns, which is equivalent to the number of linearly independent rows. Clearly, $\mathrm{rank}(\mathbf{A}) \leq \min(m, n)$; if $\mathrm{rank}(\mathbf{A}) = \min(m, n)$, then $\mathbf{A}$ is said to be *full rank*. It can be shown that

$$\mathrm{rank}\left(\mathbf{A}\right) = \mathrm{rank}\left(\mathbf{A}'\right) = \mathrm{rank}\left(\mathbf{A}\mathbf{A}'\right) = \mathrm{rank}\left(\mathbf{A}'\mathbf{A}\right).$$

Let $\mathbf{A}$ be an $m \times n$ matrix, $\mathbf{B}$ be an $m \times m$ matrix, and $\mathbf{C}$ be an $n \times n$ matrix. Then if $\mathbf{B}$ and $\mathbf{C}$ are nonsingular, $\mathrm{rank}\left(\mathbf{A}\right) = \mathrm{rank}\left(\mathbf{BAC}\right)$.

Let r be the number of nonzero eigenvalues of $\mathbf{A}$, counting multiplicities. Then $r \leq \mathrm{rank}\left(\mathbf{A}\right)$, with equality holding when $\mathbf{A}$ is symmetric.

Matrix Algebra

If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times m$ matrix, $n \geq m$, then the nonzero eigenvalues of $\mathbf{AB}$ and $\mathbf{BA}$ are the same, and $\mathbf{BA}$ will have at least $n - m$ zeros (see, e.g., Abadir and Magnus, 2005, page 167, for proof).

For $n \times n$ symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, all $\mathrm{Eig}(\mathbf{AB})$ are real if either $\mathbf{A}$ or $\mathbf{B}$ is positive semi-definite.

To see this, let $\mathbf{A} \geq 0$ so that it admits a Cholesky decomposition $\mathbf{A}^{1/2}$. Then $\mathrm{Eig}(\mathbf{AB}) = \mathrm{Eig}(\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2})$, but the latter matrix is symmetric, so its eigenvalues are real.

See Graybill (1983, Thm. 12.2.11) for another proof.

Vector Expectation

- Let $\mathbf{X} = (X_1, \ldots, X_n)'$ be a r.v. such that $\mathbb{E}[X_i] = \mu_i$, $\mathbb{V}(X_i) = \sigma_i^2$ and $\mathrm{Cov}(X_i, X_j) = \sigma_{ij}$.
- Then, for function $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ expressible as

$$\mathbf{g}(\mathbf{X}) = (g_1(\mathbf{X}), \ldots, g_m(\mathbf{X}))',$$

the expected value of vector $\mathbf{g}(\mathbf{X})$ is defined by

$$\mathbb{E}[\mathbf{g}(\mathbf{X})] := (\mathbb{E}[g_1(\mathbf{X})], \ldots, \mathbb{E}[g_m(\mathbf{X})])'.$$

- The most prominent case is the elementwise identity function, i.e.,

$$\mathbf{g}(\mathbf{X}) = (g_1(\mathbf{X}), \ldots, g_n(\mathbf{X}))' = (X_1, \ldots, X_n)' = \mathbf{X},$$

so that

$$\mathbb{E}[\mathbf{X}] := \mathbb{E}\left[(X_1, \ldots, X_n)'\right] = (\mu_1, \ldots, \mu_n)',$$

usually denoted $\boldsymbol{\mu}_{\mathbf{X}}$ or just $\boldsymbol{\mu}$.

Variance-Covariance Matrix

- Although one could analogously define the vector of variances as, say, $\mathbf{d}(\mathbf{X}) = \left(\sigma_1^2, \ldots, \sigma_n^2\right)$, more common is to denote by $\mathbb{V}(\mathbf{X})$ the matrix of covariances,

$$\mathbb{V}(\mathbf{X}) := \mathbb{E}\left[(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})'\right] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & & \sigma_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & & \sigma_n^2 \end{bmatrix},$$

  which is symmetric and often designated by $\boldsymbol{\Sigma}_{\mathbf{X}}$ or just $\boldsymbol{\Sigma}$. A particular element of $\boldsymbol{\Sigma}$ is given by $\sigma_{ij} = \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right]$.
- Note that the vector of variances $\mathbf{d}$ is $\operatorname{diag}(\boldsymbol{\Sigma})$.

Variance-Covariance Matrix

- For any matrix $\mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^n$ and vector $\mathbf{b} \in \mathbb{R}^n$, it follows from the properties of univariate expected value that

$$\mathbb{E}\left[\mathbf{AX} + \mathbf{b}\right] = \mathbf{A}\boldsymbol{\mu}_{\mathbf{X}} + \mathbf{b},$$

while the matrix of covariances is given by

$$\begin{aligned}
&\mathbb{V}\left(\mathbf{AX} + \mathbf{b}\right) \\
&\quad = \mathbb{E}\left[\left(\left(\mathbf{AX} + \mathbf{b}\right) - \left(\mathbf{A}\boldsymbol{\mu}_{\mathbf{X}} + \mathbf{b}\right)\right)\left(\left(\mathbf{AX} + \mathbf{b}\right) - \left(\mathbf{A}\boldsymbol{\mu}_{\mathbf{X}} + \mathbf{b}\right)\right)'\right] \\
&\quad = \mathbb{E}\left[\left(\mathbf{A}\left(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}\right)\right)\left(\mathbf{A}\left(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}\right)\right)'\right] \\
&\quad = \mathbf{A}\mathbb{V}\left(\mathbf{X}\right)\mathbf{A}' = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'.
\end{aligned}$$

Variance-Covariance Matrix

- As a special case, if $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$, then

$$\mathbb{V}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = \sum_{i=1}^{n} a_i^2 \mathbb{V}(X_i) + \sum\sum_{i \neq j} a_i a_j \operatorname{Cov}(X_i, X_j).$$

- The matrix $\boldsymbol{\Sigma} = \mathbb{V}(\mathbf{X})$ is positive semi-definite: Let $Y = \mathbf{a}'\mathbf{X}$ for any vector $\mathbf{a} \in \mathbb{R}^n$ and note that

$$0 \leq \mathbb{V}(Y) = \mathbb{V}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\mathbb{V}(\mathbf{X})\,\mathbf{a}.$$

- Matrix $\boldsymbol{\Sigma}$ is positive definite, or $\boldsymbol{\Sigma} > 0$, if $0 < \mathbb{V}(Y)$ for all nonzero $\mathbf{a}$, i.e., $\mathbf{a} \in \mathbb{R}^n \setminus \mathbf{0}$.

Covariance of Vector r.v.s

- As in the univariate case, it also makes sense to speak of the covariance of two r.v.s $\mathbf{X} = (X_1, \ldots, X_n)'$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)'$, given by $\boldsymbol{\Sigma}_{\mathbf{X},\mathbf{Y}} := \mathrm{Cov}\left(\mathbf{X}, \mathbf{Y}\right) := \mathbb{E}\left[\left(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}\right)\left(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}\right)'\right]$

$$= \begin{bmatrix} \sigma_{X_1,Y_1} & \sigma_{X_1,Y_2} & \cdots & \cdots & \sigma_{X_1,Y_m} \\ \sigma_{X_2,Y_1} & \sigma_{X_2,Y_2} & & & \\ \vdots & & & & \\ \sigma_{X_n,Y_1} & \sigma_{X_n,Y_2} & \cdots & \cdots & \sigma_{X_n,Y_m} \end{bmatrix},$$

  an $n \times m$ matrix with $(ij)^{\text{th}}$ element $\sigma_{X_i,Y_j} = \mathrm{Cov}\left(X_i, Y_j\right)$.
- From symmetry, $\mathrm{Cov}\left(\mathbf{X}, \mathbf{Y}\right) = \mathrm{Cov}\left(\mathbf{Y}, \mathbf{X}\right)'$.
- More generally,

$$\mathrm{Cov}\left(\mathbf{AX}, \mathbf{BY}\right) = \mathbb{E}\left[\mathbf{A}\left(\mathbf{X} - \mu_{\mathbf{X}}\right)\left(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}\right)' \mathbf{B}'\right] = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{X},\mathbf{Y}}\mathbf{B}'.$$

Covariance of Vector r.v.s

- An important special case is $\mathrm{Cov}\left(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y}\right) =$

$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \mathbf{a}'\mathbf{\Sigma}_{\mathbf{X},\mathbf{Y}}\mathbf{b} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \,\mathrm{Cov}\left(X_i, Y_j\right)$$

  for vectors $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and $\mathbf{b} = (b_1, \ldots, b_m) \in \mathbb{R}^m$.

- If sizes of $\mathbf{AX}$, $\mathbf{BY}$ and $\mathbf{b}$ are the same, then $\mathbb{V}\left(\mathbf{AX} + \mathbf{BY} + \mathbf{b}\right)$ is given by

$$\mathbf{A}\mathbb{V}\left(\mathbf{X}\right)\mathbf{A}' + \mathbf{B}\mathbb{V}\left(\mathbf{Y}\right)\mathbf{B}' + \mathbf{A}\,\mathrm{Cov}\left(\mathbf{X},\mathbf{Y}\right)\mathbf{B}' + \mathbf{B}\,\mathrm{Cov}\left(\mathbf{Y},\mathbf{X}\right)\mathbf{A}'.$$

  Observe how this generalizes the univariate case, $\mathbb{V}(aX + bY + c)$.

The Multivariate Normal Distribution

- The **multivariate normal** distribution is one of the most important multivariate distributions.
- We first mention its most important properties without proving them.
- The joint density of vector $\mathbf{Z} = (Z_1, \ldots Z_n)'$, where $Z_i \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ is

$$f_{\mathbf{Z}}(z) = \prod_{i=1}^{n} f_{Z_i}(z_i) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right\} = (2\pi)^{-n/2} \mathrm{e}^{-\mathbf{z}'\mathbf{z}/2},$$

  and is referred to as **the standard (*n*-variate) multivariate normal** and denoted $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_n)$ or $\mathbf{Z} \sim \mathrm{N}_n(\mathbf{0}, \mathbf{I}_n)$.
- From the iid-ness of the components of $\mathbf{Z}$, it follows that $\mathbb{E}[\mathbf{Z}] = \mathbf{0}$ and $\mathbb{V}(\mathbf{Z}) = \mathbf{I}_n$.

The Multivariate Normal Distribution

In general, $\mathbf{Y}$ is an (n-variate) multivariate normal r.v. if its density is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2}(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\left((\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)\right\},$$

denoted $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$ and $\boldsymbol{\Sigma}$ is a real, symmetric, positive definite matrix with $(ij)^{\text{th}}$ element σ_{ij}, $\sigma_i^2 := \sigma_{ii}$. This is derived below.

Example of Bivariate Case

The standard bivariate normal distribution takes $n = 2$, $\mu_1 = \mu_2 = 0$, and $\sigma_1^2 = \sigma_2^2 = 1$, with σ_{12} such that $\mathbf{\Sigma}$ is positive definite, or $|\sigma_{12}| < 1$.

The figure illustrates this for $\sigma_{12} = 0.6$ with a contour plot of the density (left), and a scatterplot of 1,000 simulated values (right).

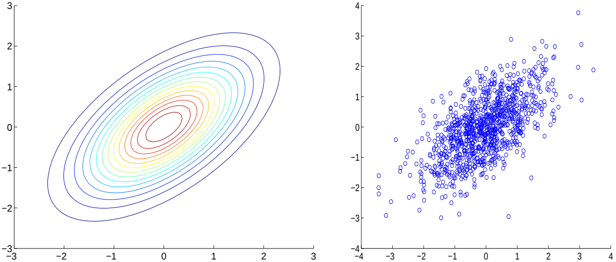


Figure: Bivariate example

Important Facts

1. The mean and variance are

$$\mathbb{E}[\mathbf{Y}] = \boldsymbol{\mu} \qquad \text{and} \qquad \mathbb{V}(\mathbf{Y}) = \boldsymbol{\Sigma}.$$

The parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ completely determine or characterize the distribution. Thus, if $\mathbf{X}$ and $\mathbf{Y}$ are both multivariate normal with the same mean and variance, then they have the same distribution.

2. All $2^n - 2$ possible marginal distributions are normally distributed with mean and variance given appropriately from $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$; e.g., $Y_i \sim \mathrm{N}\left(\mu_i, \sigma_i^2\right)$ and, for $i \neq j$,

$$\left(\begin{array}{c} Y_i \\ Y_j \end{array}\right) \sim \mathrm{N}\left(\left(\begin{array}{c} \mu_i \\ \mu_j \end{array}\right), \left(\begin{array}{cc} \sigma_i^2 & \sigma_{ij} \\ \sigma_{ij} & \sigma_j^2 \end{array}\right)\right).$$

Important Facts

3. An important special case is the bivariate normal,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right),$$

with $\mathrm{Corr}\,(Y_1, Y_2) = \rho$ and marginals $Y_i \sim \mathrm{N}\left(\mu_i, \sigma_i^2\right)$, $i = 1, 2$.

4. If Y_i and Y_j are jointly normally distributed, then they are independent iff $\mathrm{Cov}\,(Y_i, Y_j) = 0$.

   For the bivariate normal above, Y_1 and Y_2 are independent iff $\rho = 0$.

   For the general multivariate case, this extends to non-overlapping subsets $\mathbf{Y}_{(i)}$ and $\mathbf{Y}_{(j)}$, i.e.,

$$\mathbf{Y}_{(i)} \text{ and } \mathbf{Y}_{(j)} \text{ are independent iff } \mathrm{Cov}\left(\mathbf{Y}_{(i)}, \mathbf{Y}_{(j)}\right) =: \mathbf{\Sigma}_{ij} = \mathbf{0}.$$

Exercise I

In the previous result, it is important to emphasize that $\mathbf{Y}_{(i)}$ and $\mathbf{Y}_{(j)}$ need to be jointly normally distributed. If X, Y have a joint normal distribution, then $\mathrm{Cov}(X, Y) = 0 \Leftrightarrow X$ and Y are independent. However, if X and Y each are normally distributed without being jointly normally distributed, the result is not true.

Exercise: Let $X \sim \mathrm{N}(0, 1)$ and let $Y = X$ if $|X| \leq c$ and $Y = -X$ otherwise, where c is such that $\mathbb{E}[XY] = 0$. As both X and $-X$ are $\mathrm{N}(0, 1)$, so is Y. By construction, $\mathbb{E}[XY] = \mathbb{E}[(X - 0)(Y - 0)] = 0$, so that they are uncorrelated, but they are not independent.
Show that c is given as the solution to the equation

$$\frac{3}{4} = -c\phi(c) + \Phi(c).$$

Exercise I (solution)

Solution: To find c, we need to solve the equation

$$
\begin{aligned}
0 = \mathbb{E}[XY] &= \mathbb{E}\left[X^2\mathbb{I}(|X| \leq c)\right] - \mathbb{E}\left[X^2\mathbb{I}(|X| > c)\right] \\
&= \int_{-c}^{c} x^2\phi(x)\ \mathrm{d}x - \int_{c}^{\infty} x^2\phi(x)\ \mathrm{d}x - \int_{-\infty}^{-c} x^2\phi(x)\ \mathrm{d}x \\
&= 2\int_0^c x^2\phi(x)\,\mathrm{d}x - 2\int_c^\infty x^2\phi(x)\,\mathrm{d}x \quad - 2\int_0^c x^2\phi(x)\,\mathrm{d}x + 2\int_0^c x^2\phi(x)\,\mathrm{d}x \\
&= 4\int_0^c x^2\phi(x)\ \mathrm{d}x - 2\int_0^\infty x^2\phi(x)\ \mathrm{d}x \\
&= 4\int_0^c x^2\phi(x)\ \mathrm{d}x - \frac{1}{2}\cdot 2\int_{-\infty}^\infty x^2\phi(x)\ \mathrm{d}x \\
&= 4\int_0^c x^2\phi(x)\ \mathrm{d}x - 1.
\end{aligned}
$$

Exercise I (solution, cont.)

Let $u = x$ and $\mathrm{d}v = x\phi(x)\,\mathrm{d}x$. Below we show $v = -\phi(x)$, so:

$$\int_0^c x^2\phi(x)\,\mathrm{d}x = -x\phi(x)\big|_0^c + \int_0^c \phi(x)\,\mathrm{d}x = -c\phi(c) + \Phi(c) - \frac{1}{2}.$$

Thus, we wish to solve the equation:

$$0 = 4\left(-c\phi(c) + \Phi(c) - \frac{1}{2}\right) - 1 \quad \text{or} \quad \frac{3}{4} = -c\phi(c) + \Phi(c).$$

[It is very easy to verify that $\mathrm{d}/\mathrm{d}x\big[-\phi(x)\big] = x\phi(x)$, but to derive it...]

To see $v = -\phi(x)$, let $I = (2\pi)^{-1/2}\int x\exp\left\{-\frac{1}{2}x^2\right\}\mathrm{d}x$, assume $x \geq 0$, and let $u = -x^2/2$, so $x = \sqrt{-2u}$, $\mathrm{d}x = \frac{1}{2}(-2u)^{-1/2}(-2)\,\mathrm{d}u$, and

$$I = -\frac{1}{\sqrt{2\pi}}\int \exp\{u\}\,\mathrm{d}u = -\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}x^2\right\} = -\phi(x).$$

A similar calculation shows that the result holds for $x < 0$.

Important Facts (Cont.)

5. For nonoverlapping subsets $\mathbf{Y}_{(i)}$ and $\mathbf{Y}_{(j)}$ of $\mathbf{Y}$, the conditional distribution of $\mathbf{Y}_{(i)} \mid \mathbf{Y}_{(j)}$ is also normally distributed.

   The general case is given later. In the bivariate normal case, the conditionals are

$$Y_1 \mid Y_2 \sim \mathrm{N}\left(\mu_1 + \rho\sigma_1\sigma_2^{-1}\left(y_2 - \mu_2\right),\ \sigma_1^2\left(1-\rho^2\right)\right),$$
$$Y_2 \mid Y_1 \sim \mathrm{N}\left(\mu_2 + \rho\sigma_2\sigma_1^{-1}\left(y_1 - \mu_1\right),\ \sigma_2^2\left(1-\rho^2\right)\right).$$

6. The linear combination $L = \mathbf{a}'\mathbf{Y} = \sum_{i=1}^n a_i Y_i$ is normally distributed with mean $\mathbb{E}[L] = \sum_{i=1}^n a_i\mu_i = \mathbf{a}'\boldsymbol{\mu}$ and variance $\mathbb{V}(L)$.

   More generally, the set of linear combinations

$$\mathbf{L} = (L_1, \ldots, L_m)' = \mathbf{AY}$$

   is jointly normally distributed with mean $\mathbf{A}\boldsymbol{\mu}$ and variance $\mathbb{V}(\mathbf{AL}) = \mathbf{A\Sigma A}'$.

Examples

1. Let $X_i \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$. Derive the joint density of $S = X_1 + X_2$ and $D = X_1 - X_2$.

   With $\mathbf{X} = (X_1, X_2) \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_2)$ being standard bivariate normal and $\mathbf{A} = \left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$, from property 6 above,

$$\mathbf{Y} = (S, D) = \mathbf{AX} \sim \mathrm{N}\left(\mathbf{A0}, \mathbf{AIA}'\right)$$

   or $\mathbf{Y} \sim \mathrm{N}(\mathbf{0}, 2\mathbf{I}_2)$, i.e., $S \sim \mathrm{N}(0,2)$, $D \sim \mathrm{N}(0,2)$ and that S and D are independent.

Examples

2. Let X and Y be bivariate normal with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and correlation coefficient $\rho = 0.5$.

   To evaluate $\Pr(X > Y + 1)$, observe that, from fact 6 above and the formula for the variance of a weighted sum, $X - Y \sim \mathrm{N}(0 - 0, 1 + 1 - 2(0.5)) = \mathrm{N}(0, 1)$, so that

   $$\Pr(X > Y + 1) = 1 - \Phi(1) \approx 0.16.$$

   To evaluate $\mathrm{Corr}(X - Y + 1, X + Y - 2)$, note that

   $$\begin{aligned} \mathrm{Cov}(X - Y + 1, X + Y - 2) &= \mathbb{E}[(X - Y)(X + Y)] \\ &= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[Y^2\right] = 0 \end{aligned}$$

   and thus the correlation is zero. From facts 4 and 6, we see that $X + Y$ and $X - Y$ are independent.

Examples

3. We wish to calculate the density of $X_1 \mid S$, where $X_i \overset{\text{iid}}{\sim} \mathrm{N}\left(\mu, \sigma^2\right)$, $i = 1, \ldots, n$, and $S = \sum_{i=1}^n X_i$.
   Let $L_1 = X_1 = \mathbf{a}_1' \mathbf{X}$ and $L_2 = \sum_{i=1}^n X_i = \mathbf{a}_2' \mathbf{X}$ for $\mathbf{a}_1 = (1, 0, \ldots, 0)'$ and $\mathbf{a}_2 = (1, 1, \ldots, 1)'$, so that, from property 6,

$$\begin{pmatrix} L_1 \\ L_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu \\ \sum_{i=1}^n \mu \end{pmatrix}, \begin{pmatrix} 1 & 0 & \vdots & 0 \\ 1 & 1 & \vdots & 1 \end{pmatrix} \mathbf{\Sigma} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \right)$$

$$= \mathrm{N}\left(\begin{pmatrix} \mu \\ n\mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & n\sigma^2 \end{pmatrix} \right),$$

   with $\rho = n^{-1/2}$.
   From property 5, $(X_1 \mid S = s) \sim \mathrm{N}\left(s/n,\ \sigma^2\left(1 - n^{-1}\right)\right)$.
   A direct approach to solving this is also given in the text.

Exercise II (setup)

- Let X, Y have a joint bivariate normal distribution

$$\left(\begin{array}{c} X \\ Y \end{array} \right) \sim \mathrm{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \quad \text{where } \boldsymbol{\mu} = \left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right], \ \boldsymbol{\Sigma} = \left[\begin{array}{cc} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{array} \right].$$

From the general p.d.f. expression for the multivariate normal,

$$f_{\mathbf{Y}}\left(\mathbf{y}\right) = \frac{1}{\left|\boldsymbol{\Sigma}\right|^{1/2}\left(2\pi\right)^{n/2}} \exp\left\{-\frac{1}{2}\left(\left(\mathbf{y}-\boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)\right)\right\},$$

so, with $\tilde{x} = \left(x-\mu_1\right)/\sigma_1$ and $\tilde{y} = \left(y-\mu_2\right)/\sigma_2$, $f_{X,Y}\left(x,y\right)$ is

$$K \exp\left\{-\frac{1}{2}\left(\begin{array}{c} x-\mu_1 \\ y-\mu_2 \end{array} \right)' \left[\begin{array}{cc} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{array} \right]^{-1} \left(\begin{array}{c} x-\mu_1 \\ y-\mu_2 \end{array} \right)\right\}$$

$$= K \exp\left\{-\frac{\tilde{x}^2 - 2\rho\tilde{x}\tilde{y} + \tilde{y}^2}{2\left(1-\rho^2\right)}\right\}, \quad \text{with} \quad K = \frac{1}{2\pi\sigma_1\sigma_2\left(1-\rho^2\right)^{1/2}}.$$

Exercise II (questions)

1. Let $U = (X - \mu_1)/\sigma_1$ and $V = (Y - \mu_2)/\sigma_2$. Show that

$$f_{U,V}(u,v) = \frac{1}{2\pi\left(1-\rho^2\right)^{1/2}} \exp\left\{-\frac{u^2 - 2\rho uv + v^2}{2\left(1-\rho^2\right)}\right\}.$$

2. The m.g.f.s of (U, V) and (X, Y) are

$$\mathbb{M}_{U,V}(s,t) = \exp\left\{\frac{1}{2}\left(s^2 + 2\rho st + t^2\right)\right\}, \quad \text{and}$$

$$\mathbb{M}_{X,Y}(s,t) = \exp\left\{\frac{1}{2}\left[\sigma_1^2 s^2 + 2\rho\sigma_1\sigma_2 st + \sigma_2^2 t^2\right] + \mu_1 s + \mu_2 t\right\}.$$

Explain how you would use this to show that $\rho = \mathrm{Corr}(X, Y)$.

3. Show that X and $Z = Y - \rho\sigma_2 X/\sigma_1$ are independent.

Exercise II (Solutions)

1. Let $g_1 : \mathbb{R}^2 \to \mathbb{R}$ be given by $g_1(x, y) = (x - \mu_1)/\sigma_1$ and, similarly, $g_2(x, y) = (y - \mu_2)/\sigma_2$. Then $U = g_1(X, Y)$, with $X = g_1^{-1}(U, V) = \sigma_1 U + \mu_1$, and $V = g_2(X, Y)$, with $Y = g_2^{-1}(U, V) = \sigma_2 V + \mu_2$. The Jacobian involved is

$$\mathbf{J} = \begin{bmatrix} \frac{\partial g_1^{-1}(u,v)}{\partial u} & \frac{\partial g_1^{-1}(u,v)}{\partial v} \\ \frac{\partial g_2^{-1}(u,v)}{\partial u} & \frac{\partial g_2^{-1}(u,v)}{\partial v} \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix},$$

   with $|\det \mathbf{J}| = \sigma_1 \sigma_2$ and $f_{U,V}(u, v) = f_{X,Y}(x, y) \ |\det \mathbf{J}|$, which easily leads to the stated result.

2. The m.g.f. is derived in the chapter exercises. We have (Sec. 1.1.4)

$$\mathbb{E}[XY] = \left. \frac{\partial^2 \mathbb{M}_{X,Y}(s,t)}{\partial s \partial t} \right|_{s=t=0} = \mu_1 \mu_2 + \rho \sigma_1 \sigma_2,$$

   so that $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \rho \sigma_1 \sigma_2$, from which the correlation follows.

Exercise II (Solutions, cont.)

3. (proof I) From the linearity of the expectation operator,

$$\mathbb{E}[Z] = \mathbb{E}\left[Y - \frac{\rho\sigma_2}{\sigma_1}X\right] = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1, \qquad \text{so that}$$

$$\begin{aligned}
\mathrm{Cov}(X, Z) &= \mathbb{E}\left[(X - \mu_1)\left(\left(Y - \frac{\rho\sigma_2}{\sigma_1}X\right) - \left(\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1\right)\right)\right] \\
&= \mathbb{E}\left[(X - \mu_1)\left(Y - \mu_2 - \frac{\rho\sigma_2}{\sigma_1}(X - \mu_1)\right)\right] \\
&= \mathbb{E}\left[(X - \mu_1)(Y - \mu_2) - \frac{\rho\sigma_2}{\sigma_1}(X - \mu_1)^2\right] \\
&= \mathrm{Cov}(X, Y) - \frac{\rho\sigma_2}{\sigma_1}\sigma_1^2 = 0.
\end{aligned}$$

As X and Z are a linear combinations of X and Y, and X and Y follow a joint normal distribution, X and Z also follow a joint normal distribution (fact 6). Then zero covariance implies independence (fact 4).

Exercise II (Solutions, cont)

3. (proof II) They are independent if $\mathbb{M}_{X,Z}(s,t) = \mathbb{M}_X(s)\,\mathbb{M}_Z(t)$. Using the above expression for $\mathbb{M}_{X,Y}(s,t)$, $\mathbb{M}_{X,Z}(s,t)$ is

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{sX + t\left(Y - \frac{\rho\sigma_2}{\sigma_1}X\right)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{tY + \left(s - t\frac{\rho\sigma_2}{\sigma_1}\right)X\right\}\right] = \mathbb{M}_{X,Y}\left(s - t\frac{\rho\sigma_2}{\sigma_1}, t\right) \\
&= \exp\left\{\begin{array}{c} \frac{1}{2}\left[\sigma_1^2\left(s - t\frac{\rho\sigma_2}{\sigma_1}\right)^2 + 2\rho\sigma_1\sigma_2\left(s - t\frac{\rho\sigma_2}{\sigma_1}\right)t + \sigma_2^2 t^2\right] \\ +\mu_1\left(s - t\frac{\rho\sigma_2}{\sigma_1}\right) + \mu_2 t \end{array}\right\}.
\end{aligned}$$

Next...

Exercise II (Solutions, cont.)

...recall that the m.g.f. of a $\mathrm{N}\left(\mu, \sigma^2\right)$ r.v. is $\exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}$. As $Z = Y - \rho\sigma_2 X/\sigma_1 \sim \mathrm{N}\left(\left(\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1\right), \left(1-\rho^2\right)\sigma_2^2\right)$, its m.g.f. is

$$\mathbb{M}_Z(t) = \exp\left\{\left(\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1\right)t + \frac{1}{2}\left(1-\rho^2\right)\sigma_2^2 t^2\right\}.$$

Factoring the term in the exponent in the above expression for $\mathbb{M}_{X,Z}(s,t)$ yields

$$\begin{aligned}\mathbb{M}_{X,Z}(s,t) &= \exp\left\{\mu_1 s + \frac{1}{2}\sigma_1^2 s^2\right\} \\ &\quad \times \exp\left\{\left(\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1\right)t + \frac{1}{2}\left(1-\rho^2\right)\sigma_2^2 t^2\right\},\end{aligned}$$

so that $\mathbb{M}_{X,Z}(s,t) = \mathbb{M}_X(s)\,\mathbb{M}_Z(t)$, as was to be shown.

Derivation of p.d.f. of General Multivariate Normal

- Recall that $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (n-variate) with density

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2}(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\left((\mathbf{y}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y}-\boldsymbol{\mu})\right)\right\}.$$

- Let $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_n)$ so that $f_Z(z) = (2\pi)^{-n/2} \mathrm{e}^{-\mathbf{z}'\mathbf{z}/2}$. Its m.g.f. is

$$\mathbb{M}_{\mathbf{Z}}(\mathbf{t}) = \mathbb{E}\left[\mathrm{e}^{\mathbf{t}'\mathbf{z}}\right] = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \exp\{\mathbf{t}'\mathbf{z} - \mathbf{z}'\mathbf{z}/2\}\, \mathrm{d}\mathbf{z}$$

$$= \prod_{i=1}^{n} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{t_i z_i - \frac{1}{2} z_i^2\right\} \mathrm{d}z_i = \prod_{i=1}^{n} \mathbb{M}_{Z_i}(t_i) = \mathrm{e}^{\mathbf{t}'\mathbf{t}/2}.$$

- Let $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{Z}$, where $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{B} \in \mathbb{R}^{m \times n}$ and $\operatorname{rank}(\mathbf{B}) = m \leq n$. Then $\boldsymbol{\mu} := \mathbb{E}[\mathbf{Y}] = \mathbf{a}$, $\boldsymbol{\Sigma} := \mathbb{V}(\mathbf{Y}) = \mathbf{B}\mathbf{B}' \in \mathbb{R}^{m \times m}$ and

$$\mathbb{M}_{\mathbf{Y}}(\mathbf{t}) = \mathbb{E}\left[\mathrm{e}^{\mathbf{t}'\mathbf{Y}}\right] = \mathbb{E}\left[\exp\{\mathbf{t}'\mathbf{a} + \mathbf{t}'\mathbf{B}\mathbf{Z}\}\right] = \mathrm{e}^{\mathbf{t}'\boldsymbol{\mu}} \mathbb{M}_{\mathbf{Z}}(\mathbf{B}'\mathbf{t})$$

$$= \mathrm{e}^{\mathbf{t}'\boldsymbol{\mu}} \exp\{\mathbf{t}'\mathbf{B}\,\mathbf{B}'\mathbf{t}/2\} = \exp\{\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\boldsymbol{\Sigma}\,\mathbf{t}/2\}.$$

Derivation of p.d.f. of General Multivariate Normal (cont.)

- To derive the density of $\mathbf{Y}$, first observe that

$$m = \operatorname{rank}(\mathbf{B}) = \operatorname{rank}\left(\mathbf{B}\mathbf{B}'\right) = \operatorname{rank}(\boldsymbol{\Sigma}),$$

so that $\boldsymbol{\Sigma} > 0$ and $\boldsymbol{\Sigma}^{1/2} > 0$, where $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$.

- Defining $\mathbf{L} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})$,

$$\begin{aligned}
\mathbb{M}_{\mathbf{L}}(\mathbf{t}) &= \mathbb{E}\left[\exp\left\{\mathbf{t}'\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})\right\}\right] \\
&= \exp\left\{-\mathbf{t}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}\right\}\mathbb{M}_{\mathbf{Y}}\left(\boldsymbol{\Sigma}^{-1/2}\mathbf{t}\right) \\
&= \exp\left\{-\mathbf{t}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}\right\}\exp\left\{\mathbf{t}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} + \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\mathbf{t}\right\} \\
&= e^{\mathbf{t}'\mathbf{t}/2},
\end{aligned}$$

so that, from the characterization property of m.g.f.s, $\mathbf{L}$ is standard multivariate normal (of size m).

Derivation of p.d.f. of General Multivariate Normal (cont.)

- As $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{L}$ and $\partial\, \boldsymbol{\Sigma}^{-1/2} (\mathbf{Y} - \boldsymbol{\mu}) / \partial \mathbf{Y} = \boldsymbol{\Sigma}^{-1/2}$, transforming and using a basic property of determinants yields

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \left| \boldsymbol{\Sigma}^{-1/2} \right| f_{\mathbf{L}} \left(\boldsymbol{\Sigma}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}) \right) \\ &= \frac{1}{\left| \boldsymbol{\Sigma}^{1/2} \right| (2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\}, \end{aligned}$$

  which is the multivariate normal density $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- From $\mathbb{M}_{\mathbf{Y}}$, it is clear that the distribution is completely determined by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, i.e., its mean and variance-covariance matrix.
- If $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then, for appropriately sized $\mathbf{a}$ and $\mathbf{B}$,

$$\mathbf{X} = \mathbf{a} + \mathbf{B}\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\nu}, \boldsymbol{\Omega}), \quad \boldsymbol{\nu} = \mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \quad \boldsymbol{\Omega} = \mathbf{B}\, \boldsymbol{\Sigma}\, \mathbf{B}'.$$

Simulation

- Generating a sample of observations from r.v. $\mathbf{Z} \sim \mathrm{N}\left(\mathbf{0}, \mathbf{I}_n\right)$ is easy: Each of the n components of vector $\mathbf{Z}$ is i.i.d. standard univariate normal, for which simulation methods are well known.
- As $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$ follows a $\mathrm{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ distribution, realizations of $\mathbf{Y}$ can be obtained from the computed samples $\mathbf{z}$ as $\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}$, where $\boldsymbol{\Sigma}^{1/2}$ can be computed via the Cholesky decomposition, or use of the spectral decomposition.

Simulation (cont.)

- To simulate a pair of mean–zero bivariate normal r.v.s with

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

as a covariance matrix, first generate two i.i.d. standard normal r.v.s, say $\mathbf{z} = (z_1, z_2)'$ and then set $\mathbf{y} = \boldsymbol{\Sigma}^{1/2}\mathbf{z}$.

- The following code accomplishes this in Matlab for $\rho = 0.6$:

```
rho=0.6; [V,D]=eig([1 rho; rho 1]); C=V*sqrt(D)*V';
z=randn(2,1); y=C*z;
```

To generate T realizations of $\mathbf{y}$ replace the last line above with

```
T=100; z=randn(2,T); y=C*z;
```

Calculation of the c.d.f.

- With the ability to straightforwardly simulate r.v.s, the c.d.f. $\Pr(\mathbf{Y} \leq \mathbf{y})$ can be calculated by generating a large sample of observations and computing the fraction which satisfy $\mathbf{Y} \leq \mathbf{y}$.
- Continuing the previous calculation, the c.d.f. of $\mathbf{Y}$ at $(1.5, 2.5)$ can be approximated by

  ```
  length( find(y(1,:)<1.5 & y(2,:)<2.5) ) / length(y)
  ```

  which, using $T = 10^6$, yields about three to four digit accuracy.
- The exact answer, to seven digits, is 0.4459117.

Calculation of the c.d.f. (cont.)

- The simulation technique can be used with more complicated regions.
- For example, to calculate the region given by $Y_2 > 0$ and $Y_1 > Y_2$, i.e.,
$$\int_0^\infty \int_y^\infty f_{Y_1,Y_2}(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$
one simulates the values of y as before, and then computes:

```
length(find(y(2,:)>0 & y(1,:)>y(2,:)))/length(y)
```

Calculation of the c.d.f. (cont.)

- In the bivariate case, numeric integration of the density is still a feasible alternative to simulation, though hardly any faster; see the text for a program to do this.
- The evaluation of the bivariate normal c.d.f. arises frequently in practice (such as in financial option pricing); it is sufficient to construct an algorithm for the c.d.f. of the standard bivariate normal.
- Fast way is to sum enough terms of a convergent series, the most popular of which is the so–called **tetrachoric series**, developed by Pearson in 1901. The convergence becomes slower as $|\rho| \to 1$, prompting the work of Vasicek (1998), who gives a different series expression which converges fast for large ρ. See the text for the program.

Marginals of Multivariate Normal

- Let $\mathbf{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We saw above: $\mathbb{M}_{\mathbf{X}}(\mathbf{t}) = \exp\left\{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\}$.
- For $1 \leq k \leq n$, the m.g.f. of the marginal distribution of the k–size subset of $\mathbf{X}$, $\{X_{i_1}, \ldots, X_{i_k}\}$, is given by $\mathbb{M}_{\mathbf{X}}(\mathbf{t})$ with the $n-k$ elements $(1, \ldots, n) \setminus (i_1, \ldots, i_k)$ of $\mathbf{t}$ set to zero.
- From the form of the normal m.g.f., the m.g.f. of the subset is that of a k–dimensional normal r.v. with the appropriate elements of $\boldsymbol{\mu}$ and the appropriate rows and columns of $\boldsymbol{\Sigma}$ deleted.
- For example, if $n = 3$ and $\mathbf{X} = (X_1, X_2, X_3)'$, then $\mathbb{M}_{\mathbf{X}}(\mathbf{t})$ is

$$\exp\left\{\sum_{i=1}^{3} t_i\mu_i + \frac{1}{2}\begin{pmatrix} t_1 & t_2 & t_3 \end{pmatrix}\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix}\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}\right\}$$

and, for the three univariate marginals,
$\mathbb{M}_{X_j}(t) = \exp\left\{\mu_j\, t + \sigma_j^2\, t^2/2\right\}$, $j = 1, 2, 3$.

Zero Covariance and Independence

- Suppose that $\mathbf{Y} = (Y_1, \ldots, Y_n)' \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is partitioned into two subvectors $\mathbf{Y} = \left(\mathbf{Y}'_{(1)}, \mathbf{Y}'_{(2)}\right)'$, where:

$$\mathbf{Y}_{(1)} = (Y_1, \ldots, Y_p)' \quad \text{and} \quad \mathbf{Y}_{(2)} = (Y_{p+1}, \ldots, Y_n)'$$

  with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ partitioned accordingly such that $\mathbb{E}\left[\mathbf{Y}_{(i)}\right] = \boldsymbol{\mu}_{(i)}$, $\mathbb{V}\left(\mathbf{Y}_{(i)}\right) = \boldsymbol{\Sigma}_{ii}$, $i = 1, 2$, and $\operatorname{Cov}\left(\mathbf{Y}_{(1)}, \mathbf{Y}_{(2)}\right) = \boldsymbol{\Sigma}_{12}$, i.e., $\boldsymbol{\mu} = \left(\boldsymbol{\mu}'_{(1)}, \boldsymbol{\mu}'_{(2)}\right)'$ and

$$\boldsymbol{\Sigma} = \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right], \quad \boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}.$$

- One of the most useful properties of the multivariate normal is that zero correlation implies independence, i.e., $\mathbf{Y}_{(1)}$ and $\mathbf{Y}_{(2)}$ are independent iff $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Zero Covariance and Independence

- Recall that the covariance of two independent r.v.s is always zero, but the opposite need not be true.
- For the normal, use the m.g.f. to show that it factors iff $\mathbf{\Sigma}_{12} = \mathbf{0}$.
- With $\mathbf{t} = \left(\mathbf{t}'_{(1)}, \mathbf{t}'_{(2)}\right)'$,

$$\begin{aligned}
\mathbb{M}_{\mathbf{Y}}(\mathbf{t}) &= \exp\left\{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\begin{pmatrix} \mathbf{t}'_{(1)} & \mathbf{t}'_{(2)} \end{pmatrix}\begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{pmatrix}\begin{pmatrix} \mathbf{t}_{(1)} \\ \mathbf{t}_{(2)} \end{pmatrix}\right\} \\
&= \exp\left\{\mathbf{t}'_{(1)}\boldsymbol{\mu}_{(1)} + \mathbf{t}'_{(2)}\boldsymbol{\mu}_{(2)} + \frac{1}{2}\left(\mathbf{t}'_{(1)}\mathbf{\Sigma}_{11}\mathbf{t}_{(1)} + \mathbf{t}'_{(2)}\mathbf{\Sigma}_{22}\mathbf{t}_{(2)}\right)\right\} \\
&= \exp\left\{\mathbf{t}'_{(1)}\boldsymbol{\mu}_{(1)} + \frac{1}{2}\mathbf{t}'_{(1)}\mathbf{\Sigma}_{11}\mathbf{t}_{(1)}\right\}\exp\left\{\mathbf{t}'_{(2)}\boldsymbol{\mu}_{(2)} + \frac{1}{2}\mathbf{t}'_{(2)}\mathbf{\Sigma}_{22}\mathbf{t}_{(2)}\right\} \\
&= \mathbb{M}_{\mathbf{Y}_{(1)}}\left(\mathbf{t}_{(1)}\right)\mathbb{M}_{\mathbf{Y}_{(2)}}\left(\mathbf{t}_{(2)}\right).
\end{aligned}$$

- It can be proven (see the text for references) that:

$$\mathbb{M}_{\mathbf{Y}}(\mathbf{t}) = \mathbb{M}_{\mathbf{Y}_{(1)}}\left(\mathbf{t}_{(1)}\right)\mathbb{M}_{\mathbf{Y}_{(2)}}\left(\mathbf{t}_{(2)}\right) \quad \text{iff} \quad \mathbf{Y}_{(1)} \perp \mathbf{Y}_{(2)}.$$

Conditional Normal Distribution

- If $\mathbf{\Sigma}_{22} > 0$ (which is true if $\mathbf{\Sigma} > 0$), then (proved in the text)

$$\left(\mathbf{Y}_{(1)} \mid \mathbf{Y}_{(2)} = \mathbf{y}_{(2)}\right) \sim \mathrm{N}\left(\boldsymbol{\mu}_{(1)} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left(\mathbf{y}_{(2)} - \boldsymbol{\mu}_{(2)}\right),\ \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\right).$$

- A special case of great importance is the partition as $Y = Y_{(1)}$ and $\mathbf{X} = \mathbf{Y}_{(2)}$. With $\mu_Y = \mathbb{E}[Y]$ and $\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{X}]$,

$$\left(Y \mid \mathbf{X} = \mathbf{x}\right) \sim \mathrm{N}\left(\mu_Y + \mathbf{b}\left(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}\right),\ \sigma^2\right),$$

  where $\mathbf{b} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$ and $\sigma^2 = \mathbb{V}(Y) = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$.
- The conditional mean $\mu_Y + \mathbf{b}\left(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}\right)$ is referred to as the *regression function* of Y on $\mathbf{X}$ with $\mathbf{b}$ the *regression coefficient*. Notice that the regression function is linear in $\mathbf{x}$ and that σ^2 does not depend on $\mathbf{x}$. This latter property is referred to as *homoscedasticity*.

Exercise (Example 3.9 in the text)

- Let

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \sim \mathrm{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \quad \boldsymbol{\mu} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

- Show $\det\left(\boldsymbol{\Sigma}\right) = 2 \neq 0$, so $\mathbf{Y}$ is not degenerate.
- Calculate the six marginal distributions.
- Show that $Y_2 \mid (Y_1, Y_3) \sim \mathrm{N}\left(y_1 - y_3 - 1\,,\ 2\right)$.
- Show that the distribution of (X_1, X_2), where $X_1 = \sum_{i=1}^{3} Y_i$ and $X_2 = Y_1 - Y_3$, is given by

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 10 & 2 \\ 2 & 1 \end{bmatrix} \right).$$

Partial Correlation

- Let $\mathbf{Y} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- The covariance between two of the univariate random variables in $\mathbf{Y}$, say Y_i and Y_j, is determined from the (i, j)th entry of $\boldsymbol{\Sigma}$.
- Now we wish to consider the correlation of Y_i and Y_j *when conditioning on a set of other variables in* $\mathbf{Y}$.
- This structure has various uses in statistical analysis, one of which is the study of autoregressive models in time series analysis.

Partial Correlation

- Let $\mathbf{Y}_{(1)} = (Y_i, Y_j)'$ and $\mathbf{Y}_{(2)} = \mathbf{Y} \setminus \mathbf{Y}_{(1)}$, i.e., $\mathbf{Y}_{(2)}$ is $\mathbf{Y}$ but with the elements Y_i and Y_j removed.
- Let $\boldsymbol{\Sigma}_{11} = \mathbb{V}\left(\mathbf{Y}_{(1)}\right)$, $\boldsymbol{\Sigma}_{22} = \mathbb{V}\left(\mathbf{Y}_{(2)}\right)$, and $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = \operatorname{Cov}\left(\mathbf{Y}_{(1)}, \mathbf{Y}_{(2)}\right)$, so that

$$\mathbb{V}(\mathbf{Y}) = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

- Let $\mathbf{C}$ be the 2×2 conditional covariance matrix given by

$$\mathbf{C} = \begin{bmatrix} \sigma_{11|\mathbf{Y}_{(2)}} & \sigma_{12|\mathbf{Y}_{(2)}} \\ \sigma_{21|\mathbf{Y}_{(2)}} & \sigma_{22|\mathbf{Y}_{(2)}} \end{bmatrix} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \ .$$

- Motivated by the conditional variance in (3.22), the *partial correlation of Y_i and Y_j, given* $\mathbf{Y}_{(2)}$ is defined by

$$\rho_{ij|\left(\mathbf{Y}_{(2)}\right)} = \rho_{ij|(\{1,2,\ldots,n\}\setminus\{i,j\})} = \frac{\sigma_{12|\mathbf{Y}_{(2)}}}{\sqrt{\sigma_{11|\mathbf{Y}_{(2)}}\sigma_{22|\mathbf{Y}_{(2)}}}}.$$

Example

As above, let

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \sim \mathrm{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \quad \boldsymbol{\mu} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

To compute $\rho_{13|(2)}$, first write

$$\begin{bmatrix} Y_1 \\ Y_3 \\ Y_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{(1)} \\ \mu_{(2)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \Sigma_{22} \end{bmatrix}\right), \qquad \text{where}$$

$$\begin{bmatrix} \boldsymbol{\mu}_{(1)} \\ \mu_{(2)} \end{bmatrix} := \mathbb{E}\begin{bmatrix} Y_1 \\ Y_3 \\ \cdots \\ Y_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_3 \\ \cdots \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ \cdots \\ 1 \end{bmatrix} \qquad \text{and}$$

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \Sigma_{22} \end{bmatrix} := \mathbb{V}\left(\begin{bmatrix} Y_1 \\ Y_3 \\ \cdots \\ Y_2 \end{bmatrix}\right) = \begin{bmatrix} \sigma_{11} & \sigma_{13} & \sigma_{12} \\ \sigma_{31} & \sigma_{33} & \sigma_{32} \\ \sigma_{21} & \sigma_{23} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix}.$$

Example (cont'd)

Thus, we have

$$\mathbf{C} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 5/3 & 1 \\ 1 & 1 \end{bmatrix}$$

and $\rho_{13|(2)} = \frac{1}{\sqrt{5/3 \cdot 1}} = \sqrt{3/5}$. In general terms,

$$\begin{aligned} \mathbf{C} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix} - \begin{bmatrix} \sigma_{12} \\ \sigma_{32} \end{bmatrix} \begin{bmatrix} \sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{12} & \sigma_{32} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} - \sigma_{12}^2/\sigma_{22} & \sigma_{13} - \sigma_{12}\sigma_{32}/\sigma_{22} \\ \sigma_{31} - \sigma_{32}\sigma_{12}/\sigma_{22} & \sigma_{33} - \sigma_{32}^2/\sigma_{22} \end{bmatrix} \end{aligned}$$

Substituting and simplifying (see the text), we get

$$\rho_{13|(2)} = \frac{\rho_{13} - \rho_{12}\rho_{23}}{\sqrt{\left(1 - \rho_{12}^2\right)\left(1 - \rho_{23}^2\right)}}.$$

Substituting $\rho_{13} = 1/\sqrt{2}$, $\rho_{12} = 1/\sqrt{6}$ and $\rho_{23} = 0$ yields $\rho_{13|(2)} = \sqrt{3/5}$.

Exercise (Example 3.11 in the text)

If the partial correlation of several pairs of the elements of $\mathbf{Y}$, all conditional on the same set $\mathbf{Y}_{(2)}$, are to be computed, the $\mathbf{C}$ matrix can be appropriately extended. In this example, $\mathbf{C}$ is 3×3.
Let $\mathbf{Y} = (Y_1, \ldots Y_4)' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with, for $|a| < 1$,

$$\boldsymbol{\Sigma} = \frac{1}{1-a^2} \begin{bmatrix} 1 & a & a^2 & a^3 \\ a & 1 & a & a^2 \\ a^2 & a & 1 & a \\ a^3 & a^2 & a & 1 \end{bmatrix}.$$

Determine $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ so that $(Y_1, Y_3, Y_4, Y_2)' \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$. Partition them so that

$$\left(Y_1, Y_3, Y_4 \mid Y_2\right)' \sim \mathrm{N}(\boldsymbol{\nu}, \ \mathbf{C}),$$

where $\boldsymbol{\nu} = \boldsymbol{\mu}_{(1)} + \boldsymbol{\Omega}_{12}\Omega_{22}^{-1}\left(\mathbf{y}_{(2)} - \mu_{(2)}\right)$ and $\mathbf{C} = \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\Omega_{22}^{-1}\boldsymbol{\Omega}_{21}$. Show:

$$\rho_{13|(2)} = \frac{\sigma_{13|(2)}}{\sqrt{\sigma_{11|(2)}\ \sigma_{33|(2)}}} = \frac{0}{1} = 0, \quad \rho_{14|(2)} = \frac{\sigma_{14|(2)}}{\sqrt{\sigma_{11|(2)}\ \sigma_{44|(2)}}} = \frac{0}{\sqrt{1+a^2}} = 0$$

and $\rho_{34|(2)} = a/\sqrt{1+a^2}$.

$\bar{X}$ and S^2

- A **statistic** is a function of the data (and known constants such as the sample size). Thus, before the data are collected, **a statistic is a random variable**.
- For example, the **sample mean**, defined by $\bar{X}_n = \bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ and the **sample variance**, defined by

$$S_n^2(X) = S^2 = (n-1)^{-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2,$$

  are examples of statistics. Together, they have a bivariate density and distribution function, $f_{\bar{X},S^2}$.
- If the (scalar) data, say X_1 through X_n, are normally distributed, then $\bar{X}$, being a linear combination of the X_i, is also normally distributed with parameters $\mathbb{E}\left[\bar{X}\right] = n^{-1}\mathbb{E}\left[\sum X_i\right]$ and $\mathbb{V}\left(\bar{X}\right) = n^{-2}\mathbb{V}\left(\sum X_i\right)$, formulas for which we know.

Joint Distribution of $\bar{X}$ and S^2 for iid Normal Samples

- An interesting and very useful property of the normal distribution is that the sample mean and the sample variance for an iid sample **are independent**, which we denote as $\bar{X} \perp S^2$ for short.
- This is most easily demonstrated by showing the stronger result that $\bar{X} \perp \left(X_i - \bar{X}\right)$ for all i or, as $\bar{X}$ and $X_i - \bar{X}$ are both normally distributed, $\mathrm{Cov}\left(\bar{X}, X_i - \bar{X}\right) = 0$.
- In particular, as $\mathbb{V}\left(\bar{X}\right) = \sigma^2/n$ and using the covariance of two weighted sums,

$$\mathrm{Cov}\left(\bar{X}, X_i - \bar{X}\right) = \mathrm{Cov}\left(\bar{X}, X_i\right) - \mathbb{V}\left(\bar{X}\right) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{Cov}\left(X_j X_i\right) - \frac{\sigma^2}{n} = 0,$$

  as $\mathrm{Cov}\left(X_j, X_i\right) = 0$ for $i \neq j$ and σ^2 for $i = j$.
- Because S^2 can be expressed in terms of a function strictly of the $X_i - \bar{X}$, it follows that $\bar{X} \perp S^2$.

Marginal Distribution of S^2

- The $\{X_i - \bar{X}\}$ are themselves **not** independent: for $i \neq j$,

$$\begin{aligned}\mathrm{Cov}\left(X_i - \bar{X}, X_j - \bar{X}\right) &= \mathrm{Cov}\left(X_i, X_j\right) - 2\,\mathrm{Cov}\left(\bar{X}, X_i\right) + \mathbb{V}\left(\bar{X}\right) \\ &= -2\frac{\sigma^2}{n} + \frac{\sigma^2}{n} = -\frac{\sigma^2}{n}.\end{aligned}$$

- That this covariance is negative is intuitive: as the $X_i - \bar{X}$ are deviations from their mean, a positive $X_i - \bar{X}$ implies at least one negative $X_j - \bar{X}$.
- As n grows, the covariance weakens.
- Notice that $\mathrm{Cov}\left(\bar{X}, X_i - \bar{X}\right) = 0$ holds even without the normality assumption, but to conclude independence of $\bar{X}$ and S^2 **does** require normality.
- With independence established, the joint density of $\bar{X}$ and S^2 can be factored. The density of $\bar{X}$ is simply $\mathrm{N}\left(\mu, \sigma^2/n\right)$ in the iid case.

Marginal Distribution of S^2

- With $X_i \overset{\text{iid}}{\sim} \mathrm{N}\left(\mu, \sigma^2\right)$, recall that $\sum_{i=1}^{n} \left(X_i - \mu\right)^2 / \sigma^2 \sim \chi_n^2$.
- Define A, B and C as

$$
\begin{aligned}
A &:= \sum_{i=1}^{n} \sigma^{-2} \left(X_i - \mu\right)^2 = \sigma^{-2} \sum_{i=1}^{n} \left(X_i - \bar{X} + \bar{X} - \mu\right)^2 \\
&= \sigma^{-2} \sum_{i=1}^{n} \left[\left(X_i - \bar{X}\right)^2 + \left(\bar{X} - \mu\right)^2 + 2\left(X_i - \bar{X}\right)\left(\bar{X} - \mu\right)\right] \\
&= \sigma^{-2} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 + \left(\frac{\bar{X} - \mu}{\sigma\sqrt{n}}\right)^2 =: B + C,
\end{aligned}
$$

so that $A \sim \chi_n^2$ and $C \sim \chi_1^2$.

- Note that $B \perp C$ because $\left(X_i - \bar{X}\right) \perp \bar{X}$.

Marginal Distribution of S^2

- Therefore,

$$\left(\frac{1}{1-2t}\right)^{n/2} = \mathbb{M}_A(t) = \mathbb{M}_B(t)\,\mathbb{M}_C(t) = \mathbb{M}_B(t)\left(\frac{1}{1-2t}\right)^{1/2},$$

implying that $\mathbb{M}_B(t) = (1-2t)^{-(n-1)/2}$ and, hence, $B \sim \chi^2_{n-1}$.
- In terms of S^2, $B = (n-1)\,S^2/\sigma^2$, i.e., $S^2 \sim \sigma^2\chi^2_{n-1}/\,(n-1)$, a scaled chi-square random variable.
- From this and recalling how a Student's t r.v. is formed from a normal and χ^2 r.v., we see that, with $Z := \left(\bar{X}-\mu\right)/\left(\sigma/\sqrt{n}\right)$ standard normal, the random variable

$$\frac{Z}{\sqrt{B/\,(n-1)}} = \frac{\left(\bar{X}-\mu\right)/\left(\sigma/\sqrt{n}\right)}{\sqrt{\sigma^{-2}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2/\,(n-1)}} = \frac{\bar{X}-\mu}{S_n/\sqrt{n}}$$

has a Student's t distribution with $n-1$ degrees of freedom (and does not depend on σ).

S as an Estimate of σ

- The density of $S = \sqrt{S^2}$ could be derived by transformation, from which we could, among other things, investigate the extent to which S is biased for σ.
- That it is not unbiased follows because the square root is not a linear function.
- We know that the expected value of S can be obtained directly without having to transform. This gives

$$\mathbb{E}[S] = \mathbb{E}\left[\sqrt{S^2}\right] = \mathbb{E}\left[\frac{\sigma}{\sqrt{n-1}}\sqrt{\frac{(n-1)\,S^2}{\sigma^2}}\right] = \frac{\sigma}{\sqrt{n-1}}\mathbb{E}\left[U^{1/2}\right],$$

where $U \sim \chi^2_{n-1}$.

S as an Estimate of σ

- From previous results on expected powers of a chi–square r.v.,

$$\mathbb{E}\left[S\right] = K\sigma, \qquad K = \frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

- Thus, a **bias-corrected** estimate of σ would be given by $\sqrt{S^2}/K$.
- A plot of K shows that S is **downward biased** for σ, i.e., $K < 1$ for $n \geq 2$.
- In fact, the direction of the bias could have been determined without calculation using **Jensen's inequality** and the fact that $x^{1/2}$ is a **concave** function.

Intermediate Probability: A Computational Approach

1. Sums of R.V.s
   - Generating Functions
   - Sums and Other Functions
   - The Multivariate Normal Distribution
2. Asymptotics and Other Approximations
   - Convergence Concepts
   - Saddle Point Approximations
   - Order Statistics
3. More Flexible and Advanced Random Variables
   - Generalizing and Mixing
   - The Stable Paretian Distribution
   - GIG and GHyp Distributions
   - Noncentral Distributions

Asymptotics and Convergence

- The study of *asymptotic* properties of random variables ("what happens as the sample size goes to infinity"?) is of both great theoretical and practical relevance.
- Some important examples are already discussed in introductory statistics classes, a classic and important one being: when the normal distribution can be used to approximate the c.d.f. of a $\mathrm{Bin}(n, p)$ random variable, which is a special case of the Central Limit Theorem.
- In contrast to some other chapters, the material emphasizes mathematical proofs instead of mechanical operations from calculus, and so is a bit more abstract.
- We begin with some useful inequalities for real numbers and r.v.s.

Mathematical Inequalities: Triangle

- For $a \in \mathbb{R}$, $|a|$ is a if $a \geq 0$ and $-a$ if $a < 0$.
- Clearly, $a \leq |a|$ and, $\forall\ a, b \in \mathbb{R}$, $|ab| = |a|\,|b|$.
- For $b \geq 0$, $-b \leq a \leq b \quad \Leftrightarrow \quad |a| \leq b$.
- The *triangle inequality* states that

$$|x + y| \leq |x| + |y|\,, \quad \forall\ x, y \in \mathbb{R},$$

seen by squaring both sides to get

$$|x + y|^2 = (x + y)^2 = x^2 + 2xy + y^2$$

and

$$(|x| + |y|)^2 = x^2 + 2\,|x|\,|y| + y^2.$$

- The result follows by noting that $xy \leq |xy| = |x|\,|y|$.

Mathematical Inequalities: Cauchy-Schwarz

- Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, $n \in \mathbb{N}$. The *Cauchy-Schwarz inequality* is

$$|x_1 y_1 + \cdots + x_n y_n| \leq (x_1^2 + \cdots x_n^2)^{1/2}(y_1^2 + \cdots y_n^2)^{1/2}.$$

- Proof: Let $f(r) = \sum_{i=1}^n (rx_i + y_i)^2 = Ar^2 + Br + C$, where

$$A = \sum_{i=1}^n x_i^2, \quad B = 2\sum_{i=1}^n x_i y_i \quad C = \sum_{i=1}^n y_i^2.$$

- As $f(r) \geq 0$, the quadratic $Ar^2 + Br + C$ has one or no real roots, so that its discriminant $B^2 - 4AC$ is less than or equal to 0, i.e., $B^2 \leq 4AC$.
- Substituting gives $\left(\sum_{i=1}^n x_i y_i\right)^2 \leq \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right)$.

Mathematical Inequalities: Vector Triangle

- The Cauchy-Schwarz inequality is used to show the vector triangle inequality

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|,$$

where $\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}$ is the *norm* of $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. In particular, using the previous notation for A, B and C,

$$\|\mathbf{x}+\mathbf{y}\|^2 = \sum_{i=1}^n (x_i + y_i)^2 = \sum_{i=1}^n x_i^2 + 2\sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 = A + B + C$$

and, as $B^2 \leq 4AC$,

$$A + B + C \leq A + 2\sqrt{AC} + C = \left(\sqrt{A} + \sqrt{C}\right)^2.$$

- Taking square roots gives

$$\|\mathbf{x}+\mathbf{y}\| = \sqrt{A+B+C} \leq \sqrt{A} + \sqrt{C} = \|\mathbf{x}\| + \|\mathbf{y}\|.$$

Inequalities for Random Variables: Cauchy–Schwarz

- In order for a particular inequality to hold, it is often necessary that certain (absolute) moments of the r.v.s are finite.
- It is thus convenient to let

$$L_r = \{\text{r.v.s } X : \mathbb{E}[|X|^r] < \infty\}.$$

- For example, the Cauchy–Schwarz inequality (for r.v.s) states that, for any two r.v.s $U, V \in L_2$ (i.e., with finite variance),

$$|\mathbb{E}[UV]| \leq +\sqrt{\mathbb{E}[U^2]\,\mathbb{E}[V^2]}.$$

  This is proven the same way as above, using $\mathbb{E}\left[(rU + V)^2\right]$.
- Exercise: Do so!

Solution to Cauchy–Schwarz Derivation

We wish to show

$$|\,\mathbb{E}[UV]\,| \leq +\sqrt{\mathbb{E}[U^2]\,\mathbb{E}[V^2]} \tag{3}$$

Using the hint, $\forall\ r \in \mathbb{R}$,

$$0 \leq \mathbb{E}\left[(rU+V)^2\right] = \mathbb{E}\left[U^2\right] r^2 + 2\mathbb{E}[UV]\,r + \mathbb{E}\left[V^2\right] = ar^2 + br + c, \tag{4}$$

where $a := \mathbb{E}\left[U^2\right]$, $b := 2\mathbb{E}[UV]$ and $c = \mathbb{E}\left[V^2\right]$.
First let $V = -rU$, so that $0 = \mathbb{E}[0] = \mathbb{E}\left[(rU+V)^2\right]$, and the lhs of (3) is $|\mathbb{E}[UV]| = \left|-r\mathbb{E}\left[U^2\right]\right| = |r|\,\mathbb{E}\left[U^2\right]$. Likewise, the rhs of (3) is

$$\sqrt{\mathbb{E}[U^2]\,\mathbb{E}[V^2]} = \sqrt{\mathbb{E}[U^2]\,\mathbb{E}\left[(-rU)^2\right]} = \sqrt{r^2\mathbb{E}[U^2]\,\mathbb{E}[U^2]} = |r|\,\mathbb{E}\left[U^2\right],$$

and (3) holds with equality.
Now assume $V \neq -rU$, so that the inequality in (4) is strict. This implies that the quadratic $ar^2 + br + c$ has no real roots, or that its discriminant $b^2 - 4ac < 0$.
Substituting gives

$$4\left(\mathbb{E}[UV]\right)^2 - 4\mathbb{E}\left[U^2\right]\mathbb{E}\left[V^2\right] < 0$$

or $\left(\mathbb{E}[UV]\right)^2 - \mathbb{E}\left[U^2\right]\mathbb{E}\left[V^2\right] < 0$, i.e., $\left(\mathbb{E}[UV]\right)^2 < \mathbb{E}\left[U^2\right]\mathbb{E}\left[V^2\right]$, which gives (3) after taking square roots of both sides.

Cauchy–Schwarz

- The Cauchy–Schwarz inequality was given as

$$|\mathbb{E}[UV]| \leq \sqrt{\mathbb{E}[U^2]\,\mathbb{E}[V^2]},$$

  but a sharper version exists, which is

$$|\mathbb{E}[UV]| \leq \mathbb{E}\big[\,|UV|\,\big] \leq \sqrt{\mathbb{E}[U^2]\,\mathbb{E}[V^2]}.$$

- The first inequality follows because, in general, $|\mathbb{E}[X]| \leq \mathbb{E}\big[\,|X|\,\big]$.
- The second inequality is a result of Hölder's inequality, given below.

Jensen's Inequality

- A function f is (strictly) concave on $[a, b]$ if

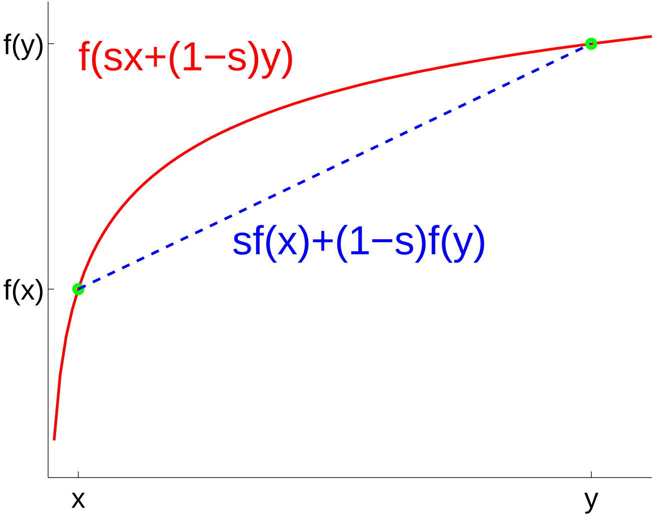
$$\forall x, y \in [a, b] \text{ and } \forall s \in [0, 1] \text{ with } t = 1 - s,$$

$$f(sx + ty) > sf(x) + tf(y).$$

- Function f is convex on $[a, b]$ iff $-f$ is concave on $[a, b]$.
- In particular, a (possibly piecewise) differentiable function f is concave on an interval if its derivative f' is non-increasing on that interval; a twice-differential function f is concave on an interval if $f'' \leq 0$ on that interval.
- *Jensen's inequality* states that, for any r.v. X with finite mean,

$$\begin{aligned} \mathbb{E}\left[g\left(X\right)\right] &\geq g(\mathbb{E}[X]), \qquad g(\cdot) \text{ convex}, \\ \mathbb{E}\left[g\left(X\right)\right] &\leq g(\mathbb{E}[X]), \qquad g(\cdot) \text{ concave}. \end{aligned}$$

Jensen's Inequality: Graphic Illustration

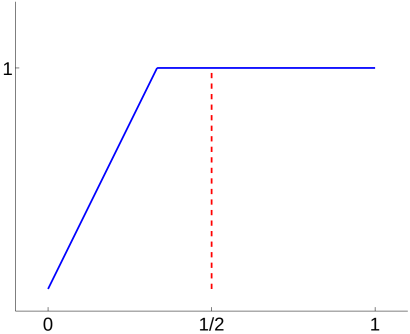


Jensen's Inequality: Intuition from the Graphic

Let $X \sim \mathrm{Unif}(0,1)$ and consider the concave function

$$g(x) = \begin{cases} 3x, & \text{if} \quad x \leq 1/3, \\ 1, & \text{if} \quad x > 1/3. \end{cases}$$

Then $g(\mathbb{E}[X]) = 1$ and $\mathbb{E}[g(X)] = \int_0^{1/3} 3x \, \mathrm{d}x + \int_{1/3}^{1} \mathrm{d}x = 5/6$, i.e., $\mathbb{E}[g(X)] < g(\mathbb{E}[X])$.



Jensen's Inequality (cont.)

- Let's verify Jensen's inequality for g convex and assuming $g''(x)$ exists with $g''(x) \geq 0$ for all x.
- We wish to show $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$.
- Let X be a r.v. with finite mean μ. Then, for g a twice differentiable, convex function, there exists a value ξ such that

$$g(x) = g(\mu) + g'(\mu)(x - \mu) + \frac{1}{2} g''(\xi)(x - \xi)^2 .$$

- That is, $g(x) \geq g(\mu) + g'(\mu)(x - \mu)$ for all x.
- Thus, $g(X) \geq g(\mu) + g'(\mu)(X - \mu)$.
- Take expectations of both sides.

Jensen's Inequality (cont.)

- **Example**: Assume X is a r.v. with finite mean μ. Let $g(x) = x^2$ with $g''(x) = 2$, so that g is convex. Then $\mathbb{E}\left[X^2\right] \geq \mu^2$. Note: if $\mathbb{E}\left[X^2\right] < \infty$, then $\mathbb{E}\left[X^2\right] = \mathbb{V}(X) + \mu^2$, which shows the result immediately.
- **Example**: Let X be a nonnegative r.v., and take $g(x) = \sqrt{x}$. As $g''(x)$ is negative for $x > 0$, g is concave and $\mathbb{E}[\sqrt{X}] \leq \sqrt{\mu}$. Note: As

$$0 \leq \mathbb{V}(\sqrt{X}) = \mathbb{E}[|X|] - (\mathbb{E}[\sqrt{X}])^2 = \mathbb{E}[X] - (\mathbb{E}[\sqrt{X}])^2,$$

  the result follows immediately.
- **Example**: Let $g(x) = \ln(x)$ for $x > 0$. Then g is concave because $g''(x) = -x^{-2}$ and $\mathbb{E}[\ln X] \leq \ln \mu$.

The c_r Inequality

- Let $U, V \in L_r$ for $r > 0$. Then

$$\mathbb{E}\left[|U+V|^r\right] \leq \mathbb{E}\left[(|U|+|V|)^r\right] \leq 2^r \left(\mathbb{E}\left[|U|^r\right] + \mathbb{E}\left[|V|^r\right]\right).$$

- Proof: For $a, b \in \mathbb{R}$, and using the triangle inequality,

$$\begin{aligned} |a+b|^r \leq \big(|a|+|b|\big)^r &\leq \big(2\max(|a|,|b|)\big)^r \\ &= 2^r \max\left(|a|^r, |b|^r\right) \leq 2^r\left(|a|^r + |b|^r\right). \end{aligned}$$

- Because this holds for any $a, b \in \mathbb{R}$, it also holds for all possible realizations of r.v.s U and V. Replacing a with U, b with V, and taking expectations (which is inequality preserving) yields the result.
- The above bound can be sharpened to

$$\mathbb{E}\left[|U+V|^r\right] \leq c_r \left(\mathbb{E}\left[|U|^r\right] + \mathbb{E}\left[|V|^r\right]\right), \quad c_r = \begin{cases} 1, & \text{if } 0 < r \leq 1, \\ 2^{r-1}, & \text{if } r \geq 1. \end{cases}$$

Triangle Inequality for Random Variables

From the triangle inequality

$$|x + y| \leq |x| + |y|, \quad \forall\, x, y \in \mathbb{R},$$

it follows (from the inequality preserving nature of expectation, and that expectation of sums is the sum of expectations) that, for r.v.s $U, V \in L_1$,

$$\mathbb{E}[|U + V|] \leq \mathbb{E}[|U|] + \mathbb{E}[|V|].$$

Observe that this is a special case of the c_r inequality.

Moment Inequalities

- The *k-norm* of r.v. X is defined to be

$$\|X\|_k = \left(\mathbb{E}[|X|^k]\right)^{1/k}, \quad \text{for } k \geq 1.$$

- *Hölder's inequality* generalizes Cauchy–Schwarz to

$$\|UV\|_1 \leq \|U\|_p \|V\|_q, \qquad p, q > 1, \quad p^{-1} + q^{-1} = 1,$$

  for r.v.s $U \in L_p$ and $V \in L_q$.
- *Lyapunov's inequality*: If $X \in L_s$, then

$$\|X\|_r \leq \|X\|_s, \qquad 1 \leq r \leq s.$$

- *Minkowski's inequality* generalizes the triangle inequality to

$$\|U + V\|_p \leq \|U\|_p + \|V\|_p, \qquad p \geq 1,$$

  for r.v.s $U, V \in L_p$.

Tail Inequalities: Markov

- *Markov's inequality*: if $X \in L_r$ for some $r > 0$, then, for all $a > 0$,

$$\Pr(|X| \geq a) \leq \frac{\mathbb{E}\left[|X|^r\right]}{a^r}.$$

- Common special case: if $X \in L_1$ is nonnegative, then for all $a > 0$,

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof:

$$\begin{aligned}\mathbb{E}[X] = \int_0^\infty x \,\mathrm{d}F_X &= \int_0^a x \,\mathrm{d}F_X + \int_a^\infty x \,\mathrm{d}F_X \\ &\geq \int_a^\infty x \,\mathrm{d}F_X \geq \int_a^\infty a \,\mathrm{d}F_X = a \int_a^\infty \mathrm{d}F_X = a \Pr(X \geq a).\end{aligned}$$

Tail Inequalities: Chebyshev

- *Chebyshev's inequality*: For $X \in L_2$ with mean μ and variance σ^2, and any $b > 0$,
$$\Pr\left(|X - \mu| \geq b\right) \leq \frac{\sigma^2}{b^2}.$$
- Proof: Use Markov's inequality $\Pr(|X| \geq a) \leq \mathbb{E}\big[|X|^r\big]/a^r$ with $r = 2$ and $X - \mu$ in place of X.
- **Example**: Let $X \in L_2$ with $\mathbb{V}(X) = 0$.
We wish to prove that $\mathbb{V}(X) = 0$ implies $\Pr(X = \mu) = 1$.
For $n \geq 1$, Chebyshev implies $\Pr\left(|X - \mu| > n^{-1}\right) = 0$.
Taking limits of both sides yields

$$0 = \lim_{n\to\infty} \Pr\left(|X - \mu| > n^{-1}\right) = \Pr\left(\lim_{n\to\infty} \left\{|X - \mu| > n^{-1}\right\}\right) = \Pr(X \neq \mu),$$

where we used the fundamental fact (proven below) that, if $A_1, A_2, \ldots$ is a monotone sequence of events, then

$$\lim_{i\to\infty} \Pr\left(A_i\right) = \Pr\left(\lim_{i\to\infty} A_i\right).$$

Tail Inequalities: Chernoff

- For r.v. X and $c > 0$, *Chernoff's Inequality* states

$$\Pr(X \geq c) \leq \inf_{t>0} \mathbb{E}\left[e^{t(X-c)}\right].$$

- Proof: Markov's inequality states: If $X \in L_1$ is nonnegative, then for all $a > 0$, $\Pr(X \geq a) \leq \mathbb{E}[X]/a$. So, for $t > 0$,

$$\Pr(X \geq c) = \Pr\left(e^{t(X-c)} \geq 1\right) \leq \mathbb{E}\left[e^{t(X-c)}\right].$$

- As this holds for any $t > 0$, one would take the infimum of the right hand side with respect to t.

Tail Inequalities: Chernoff (cont.)

- From Chernoff's Inequality $\Pr(X \geq c) \leq \inf_{t>0} \mathbb{E}\left[e^{t(X-c)}\right]$, it is easy to show the *Chernoff Bound*

$$\Pr\left(\bar{X}_n \geq c\right) \leq \inf_{t>0} \exp\left(n \log \mathbb{M}\left(\frac{t}{n}\right) - tc\right),$$

where $\bar{X}_n := n^{-1} \sum_{i=1}^{n} X_i$, $\mathbb{M}$ is the moment generating function of each of the X_i, and the X_i are i.i.d. r.v.s.

- Proof:

$$\begin{aligned} \mathbb{E}\left[e^{t\left(\bar{X}_n - c\right)}\right] &= \mathbb{E}\left[e^{-tc} e^{\frac{t}{n}\sum_{i=1}^{n} X_i}\right] \\ &= \mathbb{M}^n\left(\frac{t}{n}\right) e^{-tc} = \exp\left(n \log \mathbb{M}\left(\frac{t}{n}\right) - tc\right). \end{aligned}$$

Tail Inequalities: Kolmogorov

For $X_i \stackrel{\text{ind}}{\sim} (0, \sigma_i^2)$, let $S_n = \sum_{i=1}^n X_i$. Then $\mathbb{E}[S_n] = 0$, $\mathbb{V}(S_n) = \sum_{i=1}^n \sigma_i^2$. Chebyshev's inequality implies $\Pr\left(|S_n| \geq a\right) \leq \mathbb{V}(S_n)/a^2$, i.e.,

$$\Pr\left(|X_1 + \cdots + X_n| \geq a\right) \leq \frac{1}{a^2} \sum_{i=1}^n \sigma_i^2.$$

It turns out that this bound applies to the larger set

$$A_{a,n} := \bigcup_{j=1}^n \left\{|S_j| \geq a\right\} = \left\{\max_{1 \leq j \leq n} |S_j| \geq a\right\},$$

instead of just $\left\{|S_n| \geq a\right\}$. This is *Kolmogorov's inequality*:

$$\Pr\left(A_{a,n}\right) \leq \frac{1}{a^2} \sum_{i=1}^n \sigma_i^2.$$

Sequences of Sets

- Let Ω denote the sample space of a random experiment.
- Let $\{A_n\}$ be the infinite sequence $A_1, A_2, \ldots$ of subsets of Ω.
- Recall that the union and intersection of $\{A_n\}$ are given by

$$\bigcup_{n=1}^{\infty} A_n = \{\omega : \omega \in A_n \text{ for some } n \in \mathbb{N}\}$$

and

$$\bigcap_{n=1}^{\infty} A_n = \{\omega : \omega \in A_n \text{ for all } n \in \mathbb{N}\}.$$

Sequences of Sets

- Sequence $\{A_n\}$ is monotone increasing if $A_1 \subset A_2 \subset \cdots$ and monotone decreasing if $A_1 \supset A_2 \supset \cdots$.
- If $\{A_n\}$ is monotone increasing, then

$$\lim_{n\to\infty} A_n = A := \bigcup_{n=1}^{\infty} A_n.$$

  We denote this as $A_n \uparrow A$.
- Similarly, if the A_i are monotone decreasing, then

$$\lim_{n\to\infty} A_n = A := \bigcap_{n=1}^{\infty} A_n,$$

  denoted $A_n \downarrow A$.

Sequences of Sets

- Boole's inequality: For sequence $\{A_n\}$,

$$\Pr\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \Pr\left(A_n\right).$$

- Continuity property of $\Pr(\cdot)$: For **monotone** sequence $\{A_n\}$,

$$\lim_{n\to\infty} \Pr\left(A_n\right) = \Pr\left(\lim_{n\to\infty} A_n\right).$$

  We prove this next.

- After that we will prove: Let $\{A_n\}$ be a sequence of events which is **not necessarily monotone**. Then

$$\text{If } A_n \to A \text{, then } \lim_{n\to\infty} \Pr(A_n) \text{ exists, and } \lim_{n\to\infty} \Pr(A_n) = \Pr(A).$$

Continuity of Probability Measure

- Recall that a probability measure is a set function which assigns a real number $\Pr(A)$ to each event $A \in \mathcal{A}$ such that (i) $\Pr(A) \geq 0$, (ii) $\Pr(\Omega) = 1$, and (iii) for *mutually exclusive* A_i,

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i),$$

  where the latter requirement is known as countable additivity.
- The property of countable additivity is a crucial assumption for showing the following important result:
  If $A_1, A_2, \ldots$ is a monotone sequence of (measurable) events, then

$$\lim_{i \to \infty} \Pr(A_i) = \Pr\left(\lim_{i \to \infty} A_i\right).$$

Continuity of Probability: Increasing Events

Let $A_1, A_2, \ldots$ be a seq. of increasing events, i.e., $A_1 \subset A_2 \subset \cdots$.

Let $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1} = A_n A_{n-1}^c$, $n = 2, 3, \ldots,$. Thus, B_2 is the part of A_2 which is "new", i.e., not already in A_1. Thus,

$$\begin{aligned} A_n &= A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots \cup (A_n \setminus A_{n-1}) \\ &= B_1 \cup B_2 \cup \cdots \cup B_n = \bigcup_{i=1}^{n} B_i, \end{aligned}$$

and, by construction, the B_i are disjoint, so that $\Pr(A_n) = \sum_{i=1}^n \Pr(B_i)$. Then

$$\Pr\left(\lim_{n\to\infty} A_n\right) = \Pr\left(\lim_{n\to\infty} \bigcup_{i=1}^{n} B_i\right) = \Pr\left(\bigcup_{i=1}^{\infty} B_i\right)$$

and, from countable additivity,

$$\Pr\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \Pr(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \Pr(B_i) = \lim_{n\to\infty} \Pr(A_n).$$

Continuity of Probability: Decreasing Events

Now consider the case for monotone decreasing A_i.

Recall (see Appendix A.1 on sets), if $A_1 \subset A_2 \subset \cdots$, so that the A_n are monotone increasing, then

$$\lim_{i\to\infty} A_i = \bigcup_{i=1}^{\infty} A_i.$$

Similarly, if $A_1 \supset A_2 \supset \cdots$, (monotone decreasing), then

$$\lim_{i\to\infty} A_i = \bigcap_{i=1}^{\infty} A_i.$$

Recall also De Morgan's laws,

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c \quad \text{and} \quad \left(\bigcap_{n=1}^{\infty} A_n\right)^c = \bigcup_{n=1}^{\infty} A_n^c.$$

Continuity of Probability: Decreasing Events

For monotone decreasing A_i, $A_1 \supset A_2 \supset \cdots$, and $A_1^c \subset A_2^c \subset \cdots$, so that

$$\lim_{i\to\infty} \Pr(A_i^c) = \Pr\left(\lim_{i\to\infty} A_i^c\right)$$

from the previous result. Then

$$\lim_{i\to\infty} \Pr(A_i^c) = \lim_{i\to\infty} (1 - \Pr(A_i)) = 1 - \lim_{i\to\infty} \Pr(A_i)$$

and, from the above results,

$$\Pr\left(\lim_{i\to\infty} A_i^c\right) = \Pr\left(\bigcup_{i=1}^{\infty} A_i^c\right) = 1 - \Pr\left(\bigcap_{i=1}^{\infty} A_i\right) = 1 - \Pr\left(\lim_{i\to\infty} A_i\right),$$

so that

$$1 - \lim_{i\to\infty} \Pr(A_i) = 1 - \Pr\left(\lim_{i\to\infty} A_i\right).$$

Exercise with Continuity I

Let $A_n = \left[0, 1 + n^{-1}\right]$, $n = 1, 2, \ldots$.

Show that $\{A_n\}$ is monotone and compute $L := \lim_{n\to\infty} A_n$.

Let $B_n := A_n \setminus A_{n+1}$, $n = 1, 2, \ldots$.

Express B_n as an interval and express A_n in terms of the B_i and L.

Exercise with Continuity I: **Solution**

As $1/(n+1) < 1/n$, $A_n = [0, 1+1/n] \supset [0, 1+1/(n+1)] = A_{n+1}$ and so $\{A_n\}$ is monotone decreasing for $n = 1, 2, \ldots$.

So, $L = \lim_{n\to\infty} A_n = \cap_{i=1}^{\infty} A_i$ which, as $\lim_{n\to\infty} n^{-1} = 0$, is $[0,1]$.

We have $B_1 = [0,2] \setminus [0,1.5] = (1.5, 2]$; $B_2 = (1+1/3, 1+1/2]$; and

$$B_n = [0, 1+1/n] \setminus [0, 1+1/(n+1)] = (1+1/(n+1), 1+1/n].$$

Also,

$$\begin{aligned} A_n &= \left[0, 1+\frac{1}{n}\right] \\ &= [0,1] \cup \left(1+\frac{1}{n+1}, 1+\frac{1}{n}\right] \cup \left(1+\frac{1}{n+2}, 1+\frac{1}{n+1}\right] \cup \cdots \\ &= L \cup \bigcup_{i=n}^{\infty} B_i. \end{aligned}$$

It should be clear that L and the B_i are mutually exclusive.

Exercise with Continuity II

Let $\{A_n\}$ be a monotone decreasing sequence of events. Show that event A_n can be expressed in terms of the $B_n := A_n \setminus A_{n+1}$ and $L := \lim_{n\to\infty} A_n$.

Exercise with Continuity II: **Solution**

We claim that $A_n = L \cup \bigcup_{i=n}^{\infty} B_i$. We first show $A_n \subset L \cup \bigcup_{i=n}^{\infty} B_i$.

Fix an $n \in \mathbb{N}$ and assume $\omega \in A_n$. In general, the two events $A_n A_{n+1}$ and $A_n A_{n+1}^c$ are mutually exclusive and partition A_n, so either $\omega \in A_n A_{n+1}$ or $\omega \in A_n A_{n+1}^c$. This says, in our case with $A_n \supset A_{n+1}$ for all n, that either $\omega \in A_n A_{n+1} = A_{n+1}$ or $\omega \in A_n A_{n+1}^c = A_n \setminus A_{n+1} = B_n$.

If the latter is true, i.e., $\omega \in B_n$, then obviously $\omega \in L \cup \bigcup_{i=n}^{\infty} B_i$. If the former is true, i.e., $\omega \in A_{n+1}$, then we repeat the argument; either $\omega \in A_{n+1} A_{n+2} = A_{n+2}$ or $\omega \in A_{n+1} A_{n+2}^c = A_{n+1} \setminus A_{n+2} = B_{n+1}$.

Continuing, we see that it must be the case that either $\omega \in B_n$ or $\omega \in B_{n+1}$ or ..., or $\nexists m \geq n$ such that $\omega \in B_m = A_m \setminus A_{m+1}$, i.e., $\nexists m \geq n$ such that $\omega \notin A_{m+1}$, in which case it is in $A_n, A_{n+1}, \ldots$, and in $A_1, \ldots, A_{n-1}$ because the A_n are monotone decreasing, i.e., $\omega \in \cap_{i=1}^{\infty} A_i = \lim_{n\to\infty} A_n = L$. Thus, $\omega \in L \cup \bigcup_{i=n}^{\infty} B_i$.

To prove $A_n \supset L \cup \bigcup_{i=n}^{\infty} B_i$, simply note that, by definition, $L \subset A_n$, and as $A_n \supset A_{n+1}$, $B_i \subset A_n$ for $i = n, n+1, \ldots$.

Exercise with Continuity III

Let $\{A_n\}$ be a monotone decreasing sequence of events. Show that $\lim_{n\to\infty} \Pr(A_n) = \Pr(\lim_{n\to\infty} A_n)$ using the property of countable additivity.

Exercise with Continuity III: **Solution**

With $B_n := A_n \setminus A_{n+1}$, from countable additivity and the result in the previous problem,

$$\Pr(A_n) = \Pr(L \cup \cup_{i=n}^{\infty} B_i) = \Pr(L) + \sum_{i=n}^{\infty} \Pr(B_i) = \Pr(L) + \lim_{k\to\infty} \sum_{i=n}^{k} \Pr(B_i).$$

Taking limits of both sides,

$$\lim_{n\to\infty} \Pr(A_n) = \Pr(L) + \lim_{n\to\infty} \lim_{k\to\infty} \sum_{i=n}^{k} \Pr(B_i) = \Pr(L) = \Pr\left(\lim_{n\to\infty} A_n\right),$$

which follows because $\sum \Pr(B_i)$ is convergent (and the Cauchy criterion applies; see, e.g., Appendix A.2, page 384).

Exercise with Continuity IV

We wish to prove that the continuity of probability and the property of countable additivity are equivalent:

Prove: For any sequence of increasing measurable events $A_1 \subset A_2 \subset \cdots$, if $\lim_{i\to\infty} \Pr(A_i) = \Pr(\lim_{i\to\infty} A_i)$, then countable additivity holds, i.e., for an arbitrary sequence of mutually exclusive measurable events B_i, $\Pr\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \Pr(B_i)$.

Exercise with Continuity IV: **Solution**

Let $\{B_i\}$ denote an arbitrary sequence of disjoint measurable events, and define $A_n = \cup_{i=1}^n B_i$, $n = 1, 2, \ldots$, so that $\{A_n\}$ is an increasing sequence of measurable events.

Then

$$\bigcup_{i=1}^{\infty} B_i = \lim_{n\to\infty} \bigcup_{i=1}^{n} B_i = \lim_{n\to\infty} A_n$$

and, as $\lim_{i\to\infty} \Pr(A_i) = \Pr(\lim_{i\to\infty} A_i)$,

$$\begin{aligned}\Pr\left(\bigcup_{i=1}^{\infty} B_i\right) = \Pr\left(\lim_{n\to\infty} A_n\right) &= \lim_{n\to\infty} \Pr\left(\cup_{i=1}^n B_i\right) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \Pr(B_i) = \sum_{i=1}^{\infty} \Pr(B_i),\end{aligned}$$

which is countable additivity.

Limit Superior and Limit Inferior

- The events *limit supremum* (or *limit superior*) of $\{A_n\}$, and the *limit infimum* (or *limit inferior*) of A_n, are denoted and defined as

$$A^* = \limsup_{i \to \infty} A_i = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \ , \qquad A_* = \liminf_{i \to \infty} A_i = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n \ .$$

- First consider A^*:
- For an $\omega \in \Omega$, if $\omega \in A^*$, then $\omega \in \bigcup_{n=k}^{\infty} A_n$ for <u>*every*</u> k.
- In other words, for any k, no matter how large, there exists an $n \geq k$ with $\omega \in A_n$.
- This means that $\omega \in A_n$ for infinitely many values of n.

Limit Superior and Limit Inferior

- For

$$A_* = \liminf_{i \to \infty} A_i = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n \ ,$$

  if $\omega \in \Omega$ belongs to A_*, then it belongs to $\bigcap_{n=k}^{\infty} A_n$ for *some* k.
- That is, there exists a k such that $\omega \in A_n$ for all $n \geq k$.
- Thus, the two definitions are equivalent to

$$A^* = \{\omega : \omega \in A_n \text{ for infinitely many } n \in \mathbb{N}\} \ ,$$
$$A_* = \{\omega : \omega \in A_n \text{ for all but finitely many } n \in \mathbb{N}\} \ ,$$

  and are thus sometimes abbreviated as $A^* = \{A_n \text{ i.o.}\}$ and $A_* = \{A_n \text{ ult.}\}$, where "i.o." stands for "infinitely often" and "ult." stands for "ultimately".

Practice with liminf and limsup (Example 4.4)

- Recall: For sequence $\{A_n\}$, De Morgan's rule states

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c \quad \text{and} \quad \left(\bigcap_{n=1}^{\infty} A_n\right)^c = \bigcup_{n=1}^{\infty} A_n^c.$$

- Let $B_k = \bigcup_{n=k}^{\infty} A_n$. Then $B_k^c = \bigcap_{n=k}^{\infty} A_n^c$ and

$$(A^*)^c = \left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right)^c = \left(\bigcap_{k=1}^{\infty} B_k\right)^c = \bigcup_{k=1}^{\infty} B_k^c = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n^c = (A^c)_* ,$$

  where $A^c := \{A_n^c\}$. Similarly, $(A_*)^c = (A^c)^*$.
- One can also argue as follows. To show $(A_*)^c = (A^c)^*$,

$$\begin{aligned} (A_*)^c &= \{\omega : \omega \text{ ultimately in all } A_n\}^c \\ &= \{\omega : \omega \text{ occasionally (inf. often) not in } A_n\} \\ &= \{\omega : \omega \text{ inf. often in } A_n^c\} = (A^c)^* . \end{aligned}$$

Practice with liminf and limsup (Example 4.5)

- Let $\{A_n\}$ be arbitrary (not necessarily monotone).
- Let $B_k := \bigcup_{n=k}^{\infty} A_n$, $k = 1, 2, \ldots$. Then $\{B_k\}$ is a monotone decreasing sequence of events, so that

$$\bigcup_{n=k}^{\infty} A_n = B_k \downarrow \bigcap_{k=1}^{\infty} B_k = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = A^*.$$

- That is, as $k \to \infty$, $\bigcup_{n=k}^{\infty} A_n \downarrow A^*$.
- Thus,

$$\lim_{k\to\infty} \Pr\left(\bigcup_{n=k}^{\infty} A_n\right) = \Pr\left(\lim_{k\to\infty} \bigcup_{n=k}^{\infty} A_n\right) = \Pr(A^*).$$

- Similarly,

$$\Pr(A_*) = \lim_{k\to\infty} \Pr\left(\bigcap_{n=k}^{\infty} A_n\right). \tag{5}$$

Convergence of Sequences of Sets

- Let $\{A_n\}$ be an arbitrary sequence of sets. We wish to show

$$\Pr(A_*) \leq \liminf \Pr(A_n).$$

- Let $B_k = \cap_{n=k}^{\infty} A_n$. Then $B_k \subset A_k$ and $\Pr(B_k) \leq \Pr(A_k)$.
- B_k is monotone increasing, so $B_k \uparrow \cup_{k=1}^{\infty} B_k$, and

$$\cup_{k=1}^{\infty} B_k = \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} A_n = \liminf A_n.$$

- From equation (5) above,

$$\Pr(A_*) = \Pr(\liminf A_k) = \lim_{k\to\infty} \Pr(B_k).$$

- Now just need to prove: $\lim_{k\to\infty} \Pr(B_k) \leq \liminf \Pr(A_k)$.

Convergence of Sequences of Sets

To show $\lim_{k\to\infty} \Pr(B_k) \leq \liminf \Pr(A_k)$, note that, from real analysis,

1. if sequences b_k and a_k are such that $b_k \leq a_k$ for all k, then $\lim_{k\to\infty} b_k \leq \lim_{k\to\infty} a_k$, and
2. while $\lim_{k\to\infty} a_k$ may not exist, $\liminf_k a_k$ always does, so that $\lim_{k\to\infty} b_k \leq \liminf_{k\to\infty} a_k$.

The result to show, $\Pr(A_*) \leq \liminf \Pr(A_n)$, now follows because $\Pr(B_k) \leq \Pr(A_k)$ and $\Pr(A_*) = \Pr(\liminf A_k) = \lim_{k\to\infty} \Pr(B_k)$.

The proof that

$$\limsup \Pr(A_n) \leq \Pr(A^*)$$

is similar, and starts by letting $B_k = \cup_{n=k}^{\infty} A_k$.

Convergence of Sequences of Sets

As a definition, the sequence $\{A_n\}$ converges to A, written $A_n \to A$, iff $A = A^* = A_*$, i.e.,

$$A_n \to A \quad \text{iff} \quad A = \limsup A_n = \liminf A_n.$$

Fundamental result: Let $\{A_n\}$ be a sequence of events which is not necessarily monotone. Then

$$\text{if } A_n \to A, \text{ then } \lim_{n\to\infty} \Pr(A_n) \text{ exists, and } \lim_{n\to\infty} \Pr(A_n) = \Pr(A).$$

This is easy to prove using the above results, as we now show.

(Observe that the converse is not true. Construct a simple example to demonstrate this.)

Convergence of Sequences of Sets

Recall: If s_n is a deterministic sequence of real numbers, then

$$U = \limsup s_n \text{ and } L = \liminf s_n \text{ exist, and } L \leq U$$

and

$$\lim s_n \text{ exists iff } U = L, \text{ in which case } \lim s_n = U = L.$$

Thus, if A_n is a sequence of events, and $s_n = \Pr(A_n)$, then

$$\liminf \Pr(A_n) \leq \limsup \Pr(A_n).$$

From this and the results

$$\Pr(A_*) \leq \liminf \Pr(A_n) \quad \text{and} \quad \limsup \Pr(A_n) \leq \Pr(A^*),$$

we have

$$\Pr(\liminf A_n) \leq \liminf \Pr(A_n) \leq \limsup \Pr(A_n) \leq \Pr(\limsup A_n).$$

Convergence of Sequences of Sets

Again, we have shown:

$$\Pr(\liminf A_n) \leq \liminf \Pr(A_n) \leq \limsup \Pr(A_n) \leq \Pr(\limsup A_n) . \quad (6)$$

If $A_n \to A$, then by definition, $A = A^* = A_*$, so that (6) implies

$$\Pr(A) \leq \liminf \Pr(A_n) \leq \limsup \Pr(A_n) \leq \Pr(A) ,$$

i.e., $p := \liminf \Pr(A_n) = \limsup \Pr(A_n)$.

Thus, $\lim_n \Pr(A_n)$ exists and $\lim_n \Pr(A_n) = p$.

Again from (6), we have $\Pr(A) \leq p \leq \Pr(A)$, or $\lim_{n\to\infty} \Pr(A_n) = \Pr(A)$, as was to be shown.

Borel-Cantelli

- The two *Borel-Cantelli* lemmas are fundamental results:
- First, for a sequence $\{A_n\}$ of arbitrary events,

$$\sum_{n=1}^{\infty} \Pr(A_n) < \infty \quad \Rightarrow \quad \Pr(A_n \text{ i.o.}) = 0.$$

- Second, for a sequence $\{A_n\}$ of *independent* events,

$$\sum_{n=1}^{\infty} \Pr(A_n) = \infty \quad \Rightarrow \quad \Pr(A_n \text{ i.o.}) = 1.$$

Borel-Cantelli

- To prove the first one, use (i) the result from Example 4.5,

$$\Pr(A^*) = \lim_{k\to\infty} \Pr\left(\bigcup_{n=k}^{\infty} A_n\right),$$

(ii) Boole's inequality

$$\Pr\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \Pr(A_n),$$

and (iii) the Cauchy criterion for convergent sums: For $\epsilon > 0$,

$$\sum_{k=1}^{\infty} f_k \textbf{ converges } \Leftrightarrow \exists N \in \mathbb{N} \textbf{ such that, } \forall\ n, m \geq N\textbf{,} \left|\sum_{k=n+1}^{m} f_k\right| < \epsilon.$$

- The result then follows immediately:

$$\Pr(A_n \text{ i.o.}) = \lim_{k\to\infty} \Pr\left(\bigcup_{n=k}^{\infty} A_n\right) \leq \lim_{k\to\infty} \sum_{n=k}^{\infty} \Pr(A_n) = 0.$$

Convergence of Sequences of R.V.s

- There are several different notions of convergence when it comes to random variables, and some imply one or more of the others.
- Important to understand the difference between two r.v.s, say X and Y, "being close" (meaning that, when both are observed, their values coincide with probability one) and their *distributions* being close.
- Let X and Y be r.v.s defined on the same probability space $\{\mathbb{R}, \mathcal{B}, \Pr(\cdot)\}$. If $\Pr(X \in A) = \Pr(Y \in A)$ for all $A \in \mathcal{B}$, then X and Y are said to be *equal in distribution*, written $X \stackrel{d}{=} Y$.
- If the set $\{\omega : X(\omega) \neq Y(\omega)\}$ is an event in $\mathcal{B}$ having probability zero (termed a *null event*), then X and Y are said to be *equal almost surely*, written $X \stackrel{a.s.}{=} Y$.
- To emphasize the difference, let X and Y be i.i.d. standard normal. They then have *exactly* the same distribution, but, as they are independent, they are equal with probability zero.

Convergence in Probability

- The sequence of (univariate) random variables $\{X_n\}$ is said to *converge in probability* to random variable X iff, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0,$$

  and denoted $X_n \xrightarrow{p} X$.

- More commonly, this is written without reference to ω as

$$\lim_{n\to\infty} \Pr\left(|X_n - X| > \epsilon\right) = 0$$

  or, equivalently, as

$$\lim_{n\to\infty} \Pr\left(|X_n - X| < \epsilon\right) = 1,$$

  for all $\epsilon > 0$.

Convergence in Probability

- We can also say: $X_n \xrightarrow{p} X$ iff $\forall \epsilon > 0$ and $\delta > 0$, $\exists N \in \mathbb{N}$ such that $\Pr(|X_n - X| > \epsilon) < \delta$ for all $n \geq N$.
- In the following, we will write "Assume $X_n \xrightarrow{p} X$" to mean "Let $\{X_n\}$ be a sequence of r.v.s which converges in probability to X".
- Another common notation for $X_n \xrightarrow{p} X$, is $\operatorname{plim} X_n = X$ (read: the *probability limit of X_n is X*), as introduced into the literature by the influential paper of Mann and Wald (1943).

Weak Law of Large Numbers, or WLLN

- Let $\{X_n\}$ be a sequence of *uncorrelated* r.v.s in L_2, each with mean μ and variance σ^2, and let $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$, the average of the first n elements of the sequence.
- The WLLN states that $\bar{X}_n \xrightarrow{p} \mu$.
- To prove this, as $\bar{X}_n$ has mean μ and variance σ^2/n, it follows immediately from Chebychev's inequality that, for any $\epsilon > 0$, $\Pr\left(\left|\bar{X}_n - \mu\right| \geq \epsilon\right) \leq \sigma^2 / \left(n\epsilon^2\right)$.
- Thus, recalling the definition of convergence in probability, $\lim_{n\to\infty} \Pr\left(|X_n - X| > \epsilon\right) = 0$, we have $\bar{X}_n \xrightarrow{p} \mu$.
- The finite-variance assumption of the WLLN can be relaxed, but then we require that $\{X_n\}$ be i.i.d..

Convergence in Probability of Functions of R.V.s

- Assume $X_n \xrightarrow{p} a$, let $A \subset \mathbb{R}$, and let $g : A \to \mathbb{R}$ be a function continuous at point a with $a \in A$.
- We wish to show that $g(X_n) \xrightarrow{p} g(a)$.
- Recall: Function g is continuous at a if, for a given $\epsilon > 0$, $\exists\ \delta > 0$ such that, if $|x - a| < \delta$ and $x \in A$, then $|g(x) - g(a)| < \epsilon$.
- The contrapositive of this is: if g is continuous at a, then, for a given $\epsilon > 0$, $\exists\ \delta > 0$ such that, if $|g(x) - g(a)| \geq \epsilon$ then $\{|x - a| \geq \delta\}$.
- This implies

$$\{\omega : |g(X_n(\omega)) - g(a)| \geq \epsilon\} \subset \{\omega : |X_n(\omega) - a| \geq \delta\}.$$

Convergence in Probability of Functions of R.V.s

- Again: $\{\omega : |g(X_n(\omega)) - g(a)| \geq \epsilon\} \subset \{\omega : |X_n(\omega) - a| \geq \delta\}$.
- Recall: For sets $A, B \subset \Omega$, $A \subset B \Rightarrow \Pr(A) \leq \Pr(B)$.
- Thus, for a given $\epsilon > 0$, $\exists\ \delta > 0$ such that

$$\Pr\{|g(X_n) - g(a)| \geq \epsilon\} \leq \Pr\{|X_n - a| \geq \delta\}.$$

- For any $\delta > 0$, taking the limit of both sides

$$\lim_{n\to\infty} \Pr\{|g(X_n) - g(a)| \geq \epsilon\} = 0,$$

i.e., $g(X_n) \xrightarrow{p} g(a)$.

- The previous result holds if the constant value of a is replaced by a (nondegenerate) random variable:

$$X_n \xrightarrow{p} X,\ g \in \mathcal{C}^0 \quad \Rightarrow \quad g(X_n) \xrightarrow{p} g(X)$$

Convergence of Sums of Sequences

- Assume $X_n \xrightarrow{p} X$, $Y_n \xrightarrow{p} Y$, and $\epsilon > 0$.
- Let $d_X = X_n - X$, $d_Y = Y_n - Y$, $S_n = X_n + Y_n$, and $S = X + Y$.
- From the triangle inequality, $|d_X + d_Y| \leq |d_X| + |d_Y|$, so

$$\{|S_n - S| > \epsilon\} = \{|d_X + d_Y| > \epsilon\} \subset \{|d_X| + |d_Y| > \epsilon\} =: C.$$

- Let $A = \{|d_X| > \epsilon/2\}$ and $B = \{|d_Y| > \epsilon/2\}$. Show: $C \subset \{A \cup B\}$.
- The contrapositive is $A^c \cap B^c$ implies C^c: If $|d_X| \leq \epsilon/2$ $(= A^c)$ and $|d_Y| \leq \epsilon/2$ $(= B^c)$, then $|d_X| + |d_Y| \leq \epsilon$ $(= C^c)$.
- Thus,

$$\Pr(|S_n - S| > \epsilon) \leq \Pr(C) \leq \Pr(A \cup B) \leq \Pr(A) + \Pr(B) \to 0,$$

implying

$$X_n \xrightarrow{p} X, \ \ Y_n \xrightarrow{p} Y \quad \Rightarrow \quad X_n + Y_n \xrightarrow{p} X + Y.$$

Convergence in Probability of Functions of R.V.s

- Combining the previous results, we have:
- If $X_n \xrightarrow{p} X$, $Y_n \xrightarrow{p} Y$, and $g, h \in \mathcal{C}^0$, then

$$g(X_n) + h(Y_n) \xrightarrow{p} g(X) + g(Y).$$

- For example,

$$X_n Y_n = \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n + Y_n)^2 \xrightarrow{p} \frac{1}{2}X^2 + \frac{1}{2}Y^2 - \frac{1}{2}(X+Y)^2 = XY,$$

i.e.,

$$\text{if } X_n \xrightarrow{p} X \text{ and } Y_n \xrightarrow{p} Y, \text{ then } X_n Y_n \xrightarrow{p} XY.$$

Exercise with Convergence in Probability I

Assume $X_n \xrightarrow{p} X$ and $X_n \xrightarrow{p} Y$. Show that

1. $\Pr(X = Y) = 1$
2. As $n, m \to \infty$, $X_n - X_m \xrightarrow{p} 0$.

Solution

- We showed above that, for random variables d_X and d_Y,

$$\Pr\left(|d_X \pm d_Y| > \epsilon\right) \leq \Pr\left(|d_X| > \frac{\epsilon}{2}\right) + \Pr\left(|d_Y| > \frac{\epsilon}{2}\right). \qquad (7)$$

- To show $\Pr(X = Y) = 1$: Letting $d_X = X - X_n$ and $d_Y = Y - X_n$, (7) implies, in the limit as $n \to \infty$,

$$\Pr\left(|X - Y| > \epsilon\right) \leq \Pr\left(|X - X_n| > \frac{\epsilon}{2}\right) + \Pr\left(|Y - X_n| > \frac{\epsilon}{2}\right) \to 0,$$

  so that $\Pr(X = Y) = 1$.

- To show that, as $n, m \to \infty$, $X_n - X_m \xrightarrow{p} 0$: Let $d_X = X_n - X$ and $d_Y = X_m - X$. Then (7) implies, in the limit as $n, m \to \infty$,

$$\Pr\left(|X_n - X_m| > \epsilon\right) \leq \Pr\left(|X_n - X| > \frac{\epsilon}{2}\right) + \Pr\left(|X_m - X| > \frac{\epsilon}{2}\right) \to 0.$$

Convergence in Probability: Multivariate R.V.s

- The concept of convergence in probability is easily extended to sequences of multivariate r.v.s.
- Recall that the norm of vector $\mathbf{z} = (z_1, \ldots, z_m) \in \mathbb{R}^m$ is $\|\mathbf{z}\| = \sqrt{z_1^2 + \cdots + z_m^2}$.
- The sequence $\{\mathbf{X}_n\}$ of k-dimensional r.v.s converges in probability to k-dimensional r.v. $\mathbf{X}$ iff

$$\lim_{n\to\infty} \Pr\left(\|\mathbf{X}_n - \mathbf{X}\| > \epsilon\right) = 0,$$

and we write $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$.

Small Exercise

Show that, for a set of nonnegative r.v.s $Y_1, \ldots, Y_k$, for any $\epsilon > 0$,

$$\Pr\left(\sum_{j=1}^{k} Y_j \geq \epsilon\right) \leq \sum_{j=1}^{k} \Pr\left(Y_j \geq \frac{\epsilon}{k}\right). \tag{8}$$

Small Solution

Observe that $\sum_{j=1}^{k} Y_j \geq \epsilon$ implies $Y_j \geq \epsilon/k$ for at least one $j \in \{1, \ldots, k\}$.

Recall that $A \subset B \Rightarrow \Pr(A) \leq \Pr(B)$ and Boole's inequality: For sequence $\{A_n\}$,

$$\Pr\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \Pr(A_n).$$

It follows that

$$\Pr\left(\sum_{j=1}^{k} Y_j \geq \epsilon\right) \leq \Pr\left(\exists j \text{ such that } Y_j \geq \frac{\epsilon}{k}\right) \leq \sum_{j=1}^{k} \Pr\left(Y_j \geq \frac{\epsilon}{k}\right),$$

as was to be shown.

Exercise with Convergence in Probability II

- Let X_{nj} denote the jth marginal random variable in the vector $\mathbf{X}_n$ and let X_j denote the jth element of vector $\mathbf{X}$.
- Verify that

$$\mathbf{X}_n \xrightarrow{p} \mathbf{X} \quad \Leftrightarrow \quad X_{nj} \xrightarrow{p} X_j, \quad j = 1, 2, \ldots, k.$$

- Hint: First confirm that, for vector $\mathbf{z} = (z_1, \ldots, z_m) \in \mathbb{R}^m$,

$$|z_i| \leq \|\mathbf{z}\| \leq |z_1| + \cdots + |z_m|, \quad i = 1, \ldots, m.$$

Solution

- For the hint, square the terms. Thus, for any $\mathbf{z} \in \mathbb{R}^m$,

$$|z_i| \leq \|\mathbf{z}\| \leq |z_1| + \cdots + |z_m|, \quad i = 1, \ldots, m. \tag{9}$$

- We wish to show that $\mathbf{X}_n \xrightarrow{p} \mathbf{X} \Leftrightarrow X_{nj} \xrightarrow{p} X_j$, $j = 1, 2, \ldots, k$.
- ($\Rightarrow$) Assume $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$. Then, from the first inequality in (9),

$$|X_{nj} - X_j| \leq \|\mathbf{X}_n - \mathbf{X}\|,$$

so that, for any $\epsilon > 0$,

$$\Pr(|X_{nj} - X_j| \geq \epsilon) \leq \Pr(\|\mathbf{X}_n - \mathbf{X}\| \geq \epsilon),$$

and taking limits,

$$\lim_{n\to\infty} \Pr(|X_{nj} - X_j| \geq \epsilon) \leq \lim_{n\to\infty} \Pr(\|\mathbf{X}_n - \mathbf{X}\| \geq \epsilon) = 0.$$

Solution (cont.)

- ($\Leftarrow$) Assume $X_{nj} \xrightarrow{p} X_j$, $j = 1, 2, \ldots, k$. From the second inequality in (9), $\|\mathbf{X}_n - \mathbf{X}\| \leq \sum_{j=1}^{k} |X_{nj} - X_j|$, so that, for any $\epsilon > 0$,

$$\Pr(\|\mathbf{X}_n - \mathbf{X}\| \geq \epsilon) \leq \Pr\left(\sum_{j=1}^{k} |X_{nj} - X_j| \geq \epsilon\right).$$

- From (8),

$$\Pr\left(\sum_{j=1}^{k} |X_{nj} - X_j| \geq \epsilon\right) \leq \sum_{j=1}^{k} \Pr\left(|X_{nj} - X_j| \geq \frac{\epsilon}{k}\right).$$

- Thus, taking limits, and using the assumed fact that $X_{nj} \xrightarrow{p} X_j$, $j = 1, 2, \ldots, k$, gives, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(\|\mathbf{X}_n - \mathbf{X}\| \geq \epsilon) \leq \sum_{j=1}^{k} \lim_{n\to\infty} \Pr\left(|X_{nj} - X_j| \geq \frac{\epsilon}{k}\right) = 0.$$

Almost Sure Convergence

- First recall the pointwise convergence of functions:
  Let $\{f_n\}$ be a sequence of functions with the same domain, say D.
  $\{f_n\}$ converges pointwise to f, if

$$\forall\, x \in D,\ \lim_{n\to\infty} f_n(x) = f(x).$$

  That is, $\forall\, x \in D$ and for every given $\epsilon > 0$, $\exists\, N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \epsilon$, $\forall\, n > N$. We write $f_n \to f$.
- The sequence of r.v.s $\{X_n\}$ is said to converge *almost surely* to r.v. X iff

$$\Pr\left(\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1,$$

  written $X_n \overset{a.s.}{\to} X$.
- Almost sure convergence is similar to pointwise convergence of functions, but ...

Almost Sure Convergence

- ...but does not impose that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for *all* $\omega \in \Omega$, but rather only on a set of ω with probability one.
- In particular, $X_n \stackrel{a.s.}{\to} X$ iff there exists an event or *exception set* $E \in \mathcal{A}$ with $\Pr(E) = 0$ and $\forall \omega \in E^c$, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$.
- Observe that the definition allows E to be empty.
- Example: Let ω be a uniformly distributed r.v. on the interval $[0,1]$. Let $X_n(\omega) = n\mathbb{I}_{[0,1/n]}(\omega)$ and let $E = \{0\}$.

  For $\omega \in E^c = (0,1]$, $\lim_{n\to\infty} X_n(\omega) \to 0$, but as $X_n(0) = n$, $\lim_{n\to\infty} X_n(\omega) \nrightarrow 0$ for all $\omega \in \Omega$. Thus, the sequence $\{X_n\}$ converges almost surely to zero.

Almost Sure Convergence: Uniqueness of X

Assume $X_n \stackrel{a.s.}{\to} X$, $X_n \stackrel{a.s.}{\to} Y$. Show $\Pr(X = Y) = 1$

- Use the notation $N(\{Z_n\}, Z)$ to denote the exception set with respect to sequence $\{Z_n\}$ and random variable Z, i.e., $N(\{Z_n\}, Z) = \{\omega : \lim_{n\to\infty} Z_n(\omega) \neq Z(\omega)\}$.
- Let $N_X = N(\{X_n\}, X)$ and $N_Y = N(\{X_n\}, Y)$.
- Choose an $\omega \in (N_X \cup N_Y)^c$, so that, using the triangle inequality and taking the limit as $n \to \infty$,

$$|X(\omega) - Y(\omega)| \leq |X(\omega) - X_n(\omega)| + |X_n(\omega) - Y(\omega)| \to 0.$$

- Thus, $(N_X \cup N_Y)^c \subset \{\omega : X(\omega) = Y(\omega)\}$.
- Using the fact that $A \subset B \Rightarrow \Pr(B^c) \leq \Pr(A^c)$ then gives

$$\Pr(X \neq Y) \leq \Pr(N_X \cup N_Y) \leq \Pr(N_X) + \Pr(N_Y) = 0.$$

Addition and Continuous Transformation

$$\text{If } X_n \stackrel{a.s.}{\to} X \text{ and } Y_n \stackrel{a.s.}{\to} Y, \text{ then } X_n + Y_n \stackrel{a.s.}{\to} X + Y.$$

$$\text{if } X_n \stackrel{a.s.}{\to} X \text{ and } g \in \mathcal{C}^0, \text{ then } g(X_n) \stackrel{a.s.}{\to} g(X).$$

a.s. conv., limsup, liminf

It can be shown that sequence $X_n \stackrel{a.s.}{\to} X$ iff, for every $\epsilon > 0$,

$$\lim_{m\to\infty} \Pr\left(\bigcup_{n=m}^{\infty} \{\, |X_n - X| > \epsilon\}\right) = \lim_{m\to\infty} \Pr\left(\sup_{n\geq m} |X_n - X| > \epsilon\right) = 0,$$

or, equivalently, $\lim_{m\to\infty} \Pr\left(\bigcap_{n\geq m} \{\, |X_n - X| \leq \epsilon\}\right) = 1$.

Recall $\Pr(A^*) = \lim_{k\to\infty} \Pr\left(\bigcup_{n=k}^{\infty} A_n\right)$. With $A_n = |X_n - X| > \epsilon$, $\forall \epsilon > 0$,

$$X_n \stackrel{a.s.}{\to} X \quad \Leftrightarrow \quad \Pr(A^*) = \lim_{m\to\infty} \Pr\left(\bigcup_{n=m}^{\infty} A_n\right) = 0. \tag{10}$$

Recall $\Pr(A_*) = \lim_{k\to\infty} \Pr\left(\bigcap_{n=k}^{\infty} A_n\right)$. With $A_n^c = |X_n - X| \leq \epsilon$, $\forall \epsilon > 0$,

$$X_n \stackrel{a.s.}{\to} X \quad \Leftrightarrow \quad \Pr\left((A^c)_*\right) = \lim_{m\to\infty} \Pr\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 1. \tag{11}$$

Interpretation of Almost Sure Convergence

Recall that the limit inferior is

$$A_* = \{\omega : \omega \in A_n \text{ for all but finitely many } n \in \mathbb{N}\},$$

and is the set of ω which is "ultimately" obtained.

Thus, the result (11) above, namely:

$$X_n \overset{a.s.}{\to} X \quad \Leftrightarrow \quad \Pr\left((A^c)_*\right) = \lim_{m\to\infty} \Pr\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 1,$$

can be loosely interpreted as: with probability one, an ω occurs such that, for any $\epsilon > 0$, ultimately (for all n sufficiently large), $|X_n - X| \leq \epsilon$.

A.S. Convergence and Borel-Cantelli

Let $\{A_n\}$ be a sequence of arbitrary events, with

$$A^* = \limsup_{n\to\infty} A_n = \{\omega : \omega \in A_n \text{ for infinitely many } n \in \mathbb{N}\}.$$

The first Borel-Cantelli lemma states

$$\sum\nolimits_{n=1}^{\infty} \Pr(A_n) < \infty \quad \Rightarrow \quad \Pr(A^*) = 0.$$

This and the result (10) above, namely:

$$X_n \stackrel{a.s.}{\to} X \quad \Leftrightarrow \quad \Pr(A^*) = \lim_{m\to\infty} \Pr\left(\bigcup_{n=m}^{\infty} A_n\right) = 0,$$

imply, again with $A_n = |X_n - X| > \epsilon$, that

$$\sum\nolimits_{n=1}^{\infty} \Pr(|X_n - X| > \epsilon) < \infty \quad \Rightarrow \quad \Pr(A^*) = 0 \quad \Leftrightarrow \quad X_n \stackrel{a.s.}{\to} X.$$

A.S. Convergence and Borel-Cantelli

The previous result showed that, for $\{A_n\}$ arbitrary,

$$\sum_{n=1}^{\infty} \Pr(|X_n - X| > \epsilon) < \infty \quad \Rightarrow \quad X_n \stackrel{a.s.}{\to} X. \tag{12}$$

For the **converse**, assume the $\{A_n\}$ are independent. Then, the two BC lemmas form a *zero-one law*, i.e., $\Pr(A^*)$ can be either 0 or 1:

If events $\{A_n\}$ are independent, then, with $S = \sum_{n=1}^{\infty} \Pr(A_n)$,

$$\Pr(A^*) = \begin{cases} 0, & \text{if } S < \infty, \\ 1, & \text{if } S = \infty. \end{cases}$$

Example: From the contrapositive of BC2, for independent $\{A_n\}$, $\Pr(A^*) \neq 1 \Rightarrow S \neq \infty \Leftrightarrow S < \infty \Rightarrow \Pr(A^*) = 0$. Thus,

$$\Pr(A^*) = 0 \Leftrightarrow \Pr(A^*) \neq 1 \Rightarrow S < \infty. \tag{13}$$

For independent $\{A_n\}$, the converse of (12) now follows from (13) and result (10) above (i.e., $X_n \stackrel{a.s.}{\to} X \Leftrightarrow \Pr(A^*) = 0$).

A.S. Convergence and the Strong Law of Large Numbers

- Let $\{X_n\}$ be a sequence of i.i.d. r.v.s in L_4 with expected value μ.
- The Strong Law of Large Numbers, or SLLN, states that $\bar{X}_n \stackrel{a.s.}{\to} \mu$.
- See the text for straightforward proof.

A.S. Convergence and Convergence in Probability

- Let $X_n \stackrel{a.s.}{\to} X$ and $A_n = \{|X_n - X| > \epsilon\}$.
- For all $n \in \mathbb{N}$, $A_n \subset \bigcup_{k=n}^{\infty} A_k$ and $\Pr(A_n) \leq \Pr(\bigcup_{k=n}^{\infty} A_k)$.
- Taking limits of this last expression and recalling that $X_n \stackrel{a.s.}{\to} X$ iff, for every $\epsilon > 0$, $\lim_{m\to\infty} \Pr\left(\bigcup_{n=m}^{\infty}\{A_n\}\right) = 0$, we have

$$\lim_{n\to\infty} \Pr\left(|X_n - X| > \epsilon\right) = \lim_{n\to\infty} \Pr\left(A_n\right) \leq \lim_{n\to\infty} \Pr\left(\bigcup_{k=n}^{\infty} A_k\right) = 0.$$

- This shows that

$$X_n \stackrel{a.s.}{\to} X \quad \Rightarrow \quad X_n \stackrel{p}{\to} X.$$

- The converse does not hold; see Example 4.15 in the text for an example.

Exercise

(Problem 4.10) Sequence $\{X_n\}$ is said to *converge completely* to r.v. X iff, for all $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \Pr(|X_n - X| > \epsilon) < \infty,$$

and denoted $X_n \overset{c.c.}{\to} X$.

1. Prove that, for sequence $\{X_n\}$ with the X_n not necessarily independent,
$$X_n \overset{c.c.}{\to} X \quad \Rightarrow \quad X_n \overset{a.s.}{\to} X. \tag{14}$$
2. Prove, for constant c and X_n independent r.v.s,
$$X_n \overset{c.c.}{\to} c \quad \Leftrightarrow \quad X_n \overset{a.s.}{\to} c \qquad \text{if the } X_n \text{ are independent.} \tag{15}$$
3. Why is it not possible to extend (15) to $X_n \overset{c.c.}{\to} X$ instead of using a degenerate random variable?

Solution

To show $X_n \stackrel{c.c.}{\to} X \;\Rightarrow\; X_n \stackrel{a.s.}{\to} X$, use the first Borel-Cantelli lemma, namely: for a sequence $\{A_n\}$ of arbitrary events,

$$\sum_{n=1}^{\infty} \Pr(A_n) < \infty \quad \Rightarrow \quad \Pr(A_n \text{ i.o.}) = 0,$$

and (10), which states that, with $A_n = |X_n - X| > \epsilon$,

$$X_n \stackrel{a.s.}{\to} X \quad \Leftrightarrow \quad \Pr(A^*) = \lim_{m\to\infty} \Pr\left(\bigcup_{n=m}^{\infty} A_n\right) = 0. \qquad (16)$$

The result follows from these and the definition of complete convergence.

Solution, cont.

To show $X_n \stackrel{c.c.}{\to} c \Leftrightarrow X_n \stackrel{a.s.}{\to} c$ for X_n independent, note that, in light of the previous result, we only need to confirm that

$$X_n \stackrel{c.c.}{\to} c \quad \Leftarrow \quad X_n \stackrel{a.s.}{\to} c \qquad \text{if the } X_n \text{ are independent.}$$

From (16), $X_n \stackrel{a.s.}{\to} c$ implies that $\Pr(A^*) = 0$.
Recall the second Borel-Cantelli lemma: for a sequence $\{A_n\}$ of independent events,

$$\sum_{n=1}^{\infty} \Pr(A_n) = \infty \quad \Rightarrow \quad \Pr(A_n \text{ i.o.}) = 1.$$

The contrapositive of this is $\Pr(A_n \text{ i.o.}) \neq 1 \Rightarrow \sum_{n=1}^{\infty} \Pr(A_n) < \infty$, which, with $A_n = |X_n - X| > \epsilon$, is the definition of complete convergence.

Note that, even if the X_n are independent r.v.s, the $X_n - X$ are *not* independent, which is why a constant c is used.

Convergence in r–Mean

- Let $\{X_n\}$ be a sequence of random variables in L_r.
- We say $\{X_n\}$ converges in r-mean to $X \in L_r$ iff

$$\lim_{n\to\infty} \mathbb{E}\left[\,|X_n - X|^r\,\right] = 0,$$

  and write $X_n \xrightarrow{L_r} X$ or $X_n \xrightarrow{L^r} X$ or $X_n \xrightarrow{r} X$.
- The cases $r = 1$ and $r = 2$ arise frequently: If $X_n \xrightarrow{L_1} X$, we say X_n converges in mean to X, and if $X_n \xrightarrow{L_2} X$, we say X_n converges in mean square to X.

Convergence in r–Mean

- Let $\{X_n\}$ be an i.i.d. sequence of random variables with $\mathbb{E}[X_n] = \mu$ and $\mathbb{V}(X_n) = \sigma^2 < \infty$.
- As usual, let $\bar{X} = n^{-1} S_n$, $S_n = \sum_{i=1}^{n} X_i$.
- Observe that

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\bar{X}_n - \mu\right|^2\right] = \lim_{n\to\infty} \mathbb{V}(\bar{X}_n) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0,$$

  showing that

$$\bar{X} \xrightarrow{L_2} \mu.$$

- Recall the WLLN and the SLLN: $\bar{X}_n \xrightarrow{p} \mu$ and $\bar{X}_n \xrightarrow{a.s.} \mu$, respectively. However, in general, $X_n \xrightarrow{a.s.} X \nRightarrow X_n \xrightarrow{L_r} X$.

Some r-Mean Convergence Results

- Recall Markov's inequality:
  If $X \in L_1$ is nonnegative, then for all $a > 0$, $\Pr(X \geq a) \leq \mathbb{E}[X]/a$.
- Thus, for any $\epsilon > 0$,

$$\Pr\left(|X_n - X| \geq \epsilon\right) \leq \frac{\mathbb{E}[|X_n - X|]}{\epsilon},$$

  from which it follows that $X_n \xrightarrow{L_1} X \Rightarrow X_n \xrightarrow{p} X$.
- Recall the Cauchy-Schwarz (CS) inequality:
  For r.v.s $U, V \in L_2$, $\mathbb{E}\left[\,|UV|\,\right] \leq \sqrt{\mathbb{E}\left[U^2\right]\mathbb{E}\left[V^2\right]}$.
- Assume $X_n \xrightarrow{L_2} X$, and let $U = |X_n - X|$ and $V = 1$. Then CS implies $\mathbb{E}\left[|X_n - X|\right] \leq \sqrt{\mathbb{E}\left[|X_n - X|^2\right]}$, so that

$$X_n \xrightarrow{L_2} X \Rightarrow X_n \xrightarrow{L_1} X.$$

Some r-Mean Convergence Results, cont.

- From $\left|\mathbb{E}\left[Y\right]\right| \le \mathbb{E}\left[|Y|\right]$, we have $\left|\mathbb{E}\left[X_n - X\right]\right| \le \mathbb{E}\left[|X_n - X|\right]$, from which it follows that

$$X_n \xrightarrow{L_1} X \Rightarrow \mathbb{E}[X_n] \to \mathbb{E}[X].$$

- We wish to also show that $X_n \xrightarrow{L_2} X \Rightarrow \mathbb{E}[X_n^2] \to \mathbb{E}[X^2]$. By writing

$$\mathbb{E}\left[X_n^2\right] = \mathbb{E}\left[(X_n - X)^2\right] + \mathbb{E}\left[X^2\right] + 2\mathbb{E}\left[X(X_n - X)\right]$$

and applying the Cauchy-Schwarz inequality, we have

$$\left|\mathbb{E}\left[X(X_n - X)\right]\right| \le \mathbb{E}\left[|X(X_n - X)|\right] \le \sqrt{\mathbb{E}\left[X^2\right]\mathbb{E}\left[(X_n - X)^2\right]}.$$

Taking limits shows

$$X_n \xrightarrow{L_2} X \Rightarrow \mathbb{E}[X_n^2] \to \mathbb{E}[X^2].$$

General r-Mean Convergence Results

- The previous results can be generalized as follows, with proofs in the book:

$$X_n \xrightarrow{L_r} X \quad \Rightarrow \quad X_n \xrightarrow{p} X, \quad r > 0,$$

$$X_n \xrightarrow{L_s} X \quad \Rightarrow \quad X_n \xrightarrow{L_r} X, \qquad s \geq r \geq 1,$$

$$X_n \xrightarrow{L_r} X \quad \Rightarrow \quad \mathbb{E}\left[|X_n|^r\right] \to \mathbb{E}\left[|X|^r\right], \quad r > 0.$$

- Examples 4.20 and 4.21 demonstrate that $X_n \xrightarrow{a.s.} X \not\Rightarrow X_n \xrightarrow{L_r} X$. See the text for references with examples showing that $X_n \xrightarrow{p} X \not\Rightarrow X_n \xrightarrow{L_r} X$ and $X_n \xrightarrow{L_r} X \not\Rightarrow X_n \xrightarrow{a.s.} X$.

Example 4.22

To show that

$$X_n \xrightarrow{L_2} X, \quad Y_n \xrightarrow{L_2} Y \quad \Rightarrow \quad X_n Y_n \xrightarrow{L_1} XY,$$

use the triangle inequality ($\mathbb{E}[|U+V|] \leq \mathbb{E}[|U|] + \mathbb{E}[|V|]$) and the Cauchy-Schwarz inequality twice to write

$$\begin{aligned}
\mathbb{E}\big[|X_n Y_n - XY|\big] &= \mathbb{E}\big[|X_n Y_n - X_n Y + X_n Y - XY|\big] \\
&\leq \mathbb{E}\big[|X_n Y_n - X_n Y|\big] + \mathbb{E}\big[|X_n Y - XY|\big] \\
&\leq \sqrt{\mathbb{E}\big[X_n^2\big]\mathbb{E}\big[(Y_n - Y)^2\big]} + \sqrt{\mathbb{E}\big[Y^2\big]\mathbb{E}\big[(X_n - X)^2\big]}.
\end{aligned}$$

The first term goes to zero in the limit because $Y_n \xrightarrow{L_2} Y$ and because $X_n \xrightarrow{L_2} X \Rightarrow \mathbb{E}[X_n^2] \to \mathbb{E}[X^2]$.
The second term converges to zero in the limit because $X_n \xrightarrow{L_2} X$.

Exercise: Problem 4.9

- First, show that, if $\{X_n\}$ converges in r-mean, then the limiting random variable is unique. That is, assume $X_n \xrightarrow{r} X$ and $X_n \xrightarrow{r} Y$ and show that $\Pr(X = Y) = 1$.
  Hint: Use the inequality from Section 4.1 which states that, for r.v.s U and V, both in L_r,
  $$\mathbb{E}\left[|U+V|^r\right] \leq 2^r \left(\mathbb{E}\left[|U|^r\right] + \mathbb{E}\left[|V|^r\right]\right).$$
- Second, show that, if $X_n \xrightarrow{r} X$ and $Y_n \xrightarrow{r} Y$, then $X_n + Y_n \xrightarrow{r} X + Y$.

Solution

From the hint, $\mathbb{E}\big[|U+V|^r\big] \leq 2^r \left(\mathbb{E}\big[|U|^r\big] + \mathbb{E}\big[|V|^r\big]\right)$ for r.v.s U and V, and with $U = X - X_n$ and $V = X_n - Y$, this reads

$$\mathbb{E}\big[|X-Y|^r\big] \leq 2^r \left(\mathbb{E}\big[|X-X_n|^r\big] + \mathbb{E}\big[|X_n-Y|^r\big]\right).$$

By assumption, the r.h.s. converges to zero in the limit, so that $\mathbb{E}\big[|X-Y|^r\big] = 0$. If X and Y have joint p.d.f. $f_{X,Y}$ and support $\mathcal{S}_{X,Y}$, this implies that

$$0 = \mathbb{E}\big[|X-Y|^r\big] = \iint_{\mathcal{S}_{X,Y}} |x-y|^r f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

For $(x,y) \in \mathcal{S}_{X,Y}$ with $x \neq y$, as $|x-y|^r > 0$, it must be the case that (except on a set in $\mathbb{R}^2$ of measure zero), $f_{X,Y}(x,y) = 0$.

Thus, $f_{X,Y}(x,y)$ is positive only for $(x,y) \in \mathcal{S}_{X,Y}$ and $x = y$. That is, $\Pr(X \neq Y) = 0$, or $\Pr(X = Y) = 1$.

Solution, cont.

We wish to show that $X_n + Y_n \xrightarrow{r} X + Y$ if $X_n \xrightarrow{r} X$ and $Y_n \xrightarrow{r} Y$.

Again from the inequality $\mathbb{E}\left[|U+V|^r\right] \leq 2^r \left(\mathbb{E}\left[|U|^r\right] + \mathbb{E}\left[|V|^r\right]\right)$ for r.v.s U and V, with $U = X_n - X$ and $V = Y_n - Y$,

$$\mathbb{E}\left[|(X_n + Y_n) - (X + Y)|^r\right] \leq 2^r \left(\mathbb{E}\left[|X_n - X|^r\right] + \mathbb{E}\left[|Y_n - Y|^r\right]\right),$$

and the r.h.s. converges to zero in the limit.

Convergence in Distribution

- For a given c.d.f. F, define the set:

$$C(F) = \{x : F(x) \text{ is continuous at } x\}.$$

- Sequence $\{X_n\}$ is said to *converge in distribution* to X iff

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \forall x \in C(F_X).$$

- We write $X_n \xrightarrow{d} X$.
- Also written *convergence in law* or *weak convergence*.
- This is the weakest form of convergence.

Convergence in Distribution: Example 4.23

We illustrate that $X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{L_r} X$.

- Let $\{X_n\}$ be a sequence of discrete r.v.s such that $\Pr(X_n = 0) = 1 - n^{-1}$ and $\Pr(X_n = n) = n^{-1}$, so that

$$F_{X_n}(x) = \left(1 - n^{-1}\right)\mathbb{I}_{[0,n)}(x) + \mathbb{I}_{[n,\infty)}(x).$$

- Then $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all points of continuity of F_n, where $F_X(x) = \mathbb{I}_{[0,\infty)}(x)$, and F_X is the c.d.f. of the r.v. X which is degenerate at zero, i.e., $\Pr(X = 0) = 1$.
- For all $k \in \mathbb{N}$, $\mathbb{E}[X^k] = 0$ but $\mathbb{E}[X_n^k] = n^{k-1}$, so that, for $r > 0$, $\mathbb{E}\left[|X_n|^r\right] \nrightarrow \mathbb{E}\left[|X|^r\right]$.
- Thus, from the contrapositive of the result

$$X_n \xrightarrow{L_r} X \quad \Rightarrow \quad \mathbb{E}\left[|X_n|^r\right] \to \mathbb{E}\left[|X|^r\right], \quad r > 0,$$

it follows that X_n does not converge in r-mean to X.

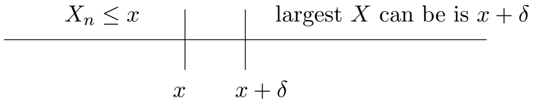
Example 4.24: $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$

- Let F_n and F be the c.d.f.s of X_n and X, and $x \in C(F)$.
- Let $\delta > 0$ and event $B = \{|X_n - X| \leq \delta\}$.
- Split $F_{X_n}(x)$ into the two disjoint events

$$\Pr(X_n \leq x) = \Pr(\{X_n \leq x\} \cap B) + \Pr(\{X_n \leq x\} \cap B^c).$$

- Observe from the diagram below that

$$\{X_n \leq x\} \cap B \;\subset\; \{X \leq x + \delta\} \cap B \;\subset\; \{X \leq x + \delta\}.$$



Example 4.24: $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$

- Again, $\{X_n \leq x\} \cap B \ \subset \ \{X \leq x + \delta\}$.
- Also, trivially, $\{X_n \leq x\} \cap B^c \subset B^c$. Thus,

$$\begin{aligned} \Pr(X_n \leq x) &= \Pr(\{X_n \leq x\} \cap B) + \Pr(\{X_n \leq x\} \cap B^c) \\ &\leq \Pr(X \leq x + \delta) + \Pr(B^c). \end{aligned}$$

- For any $\epsilon > 0$, choose a $\delta > 0$ such that $F(x+\delta) - F(x-\delta) < \epsilon/2$, and choose an $N \in \mathbb{N}$ such that, for $n \geq N$, $\Pr(B^c) = \Pr(|X_n - X| > \delta) < \epsilon/2$. Then

$$\begin{aligned} \Pr(X_n \leq x) &\leq \Pr(X \leq x + \delta) + \Pr(B^c) \\ &< \left[F(x-\delta) + \epsilon/2\right] \quad + \epsilon/2 \\ &\leq F(x) + \epsilon/2 + \epsilon/2. \end{aligned}$$

Example 4.24: $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$, cont.

- We showed that $\Pr(X_n \leq x) \leq \Pr(X \leq x + \delta) + \Pr(B^c)$. In the derivation, we could switch the roles of X_n and X. This gives

$$\Pr(X \leq x) \leq \Pr(X_n \leq x + \delta) + \Pr(B^c)$$

or, replacing x by $x - \delta$,

$$\Pr(X \leq x - \delta) \leq \Pr(X_n \leq x) + \Pr(B^c).$$

- Recall $F(x + \delta) - F(x - \delta) < \epsilon/2$.
- Thus,

$$\begin{aligned} F(x) &\leq F(x + \delta) \\ &\leq \epsilon/2 + \Pr(X \leq x - \delta) \\ &\leq \epsilon/2 + \Pr(X_n \leq x) + \Pr(B^c) \leq \epsilon/2 + F_n(x) + \epsilon/2. \end{aligned}$$

Converse of $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$

- We have shown that, with F_n and F the c.d.f.s of X_n and X, respectively, and $x \in C(F)$, if $X_n \xrightarrow{p} X$, then, for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that, for $n \geq N$,

$$F_n(x) < F(x) + \epsilon$$

and

$$F(x) - \epsilon < F_n(x).$$

- As $\epsilon > 0$ is arbitrary, it follows that $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$.

Converse of $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$

If $\{X_n\}$ is a sequence of r.v.s and X is a degenerate r.v. with $\Pr(X = c) = 1$ for some $c \in \mathbb{R}$, then $X_n \xrightarrow{d} X \Rightarrow X_n \xrightarrow{p} X$.

To prove this, for any $\epsilon > 0$, as $c \pm \epsilon \in C(F_X) = \{x : x \neq c\}$, we have

$$\begin{aligned}
\Pr(|X_n - c| > \epsilon) &= 1 - \Pr(c - \epsilon \leq X_n \leq c + \epsilon) \\
&= 1 - F_{X_n}(c + \epsilon) + F_{X_n}(c - \epsilon) - \Pr(X_n = c - \epsilon) \\
&\leq 1 - F_{X_n}(c + \epsilon) + F_{X_n}(c - \epsilon) \\
&\to 1 - 1 + 0 = 0,
\end{aligned}$$

so that $\lim_{n\to\infty} \Pr(|X_n - c| > \epsilon) = 0$.

Slutsky's Theorem

- Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$ for some $a \in \mathbb{R}$. Then

$$X_n \pm Y_n \xrightarrow{d} X \pm a, \tag{17}$$

$$X_n \cdot Y_n \xrightarrow{d} X \cdot a, \quad \text{and} \quad X_n / Y_n \xrightarrow{d} X/a, \quad a \neq 0. \tag{18}$$

- Example 4.26: Recall that if $G \sim N(0,1)$ independent of $C_n \sim \chi^2_n$, then $T_n := G/\sqrt{C_n/n} \sim t_n$. Also, $C_n \overset{d}{=} \sum_{i=1}^n X_i$, where $X_i \overset{\text{i.i.d.}}{\sim} \chi^2_1$.
- The WLLN implies that $C_n/n \xrightarrow{p} \mathbb{E}[X_1] = 1$.
- We showed above: If $X_n \xrightarrow{p} a$, $A \subset \mathbb{R}$, and $g : A \to \mathbb{R}$ is a function continuous at point a with $a \in A$, then $g(X_n) \xrightarrow{p} g(a)$.
- Thus, $\sqrt{C_n/n} \xrightarrow{p} \sqrt{1} = 1$.
- Apply (18) to T (trivially, $G \xrightarrow{d} \mathrm{N}(0,1)$), to get $T_n \xrightarrow{d} \mathrm{N}(0,1)$.

Relation between Convergence Concepts

All the relationships discussed above (or in the Problems) between the various methods of convergence are embodied in the following diagram, for a sequence $\{X_n\}$ and constant $c \in \mathbb{R}$.

$$\begin{array}{ccccccc}
\boxed{X_n \overset{c.c.}{\to} X} & \begin{array}{c} \overset{X=c,\, X_n \text{ ind}}{\Leftarrow} \\ \Rightarrow \end{array} & \boxed{X_n \overset{a.s.}{\to} X} & \begin{array}{c} \overset{X_n \text{ mon}}{\Leftarrow} \\ \Rightarrow \end{array} & \boxed{X_n \overset{p}{\to} X} & \begin{array}{c} \overset{X=c}{\Leftarrow} \\ \Rightarrow \end{array} & \boxed{X_n \overset{d}{\to} X} \\
 & & & & \Uparrow & & \\
 & & & & \boxed{X_n \overset{L_r}{\to} X} & \overset{0<s<r}{\Rightarrow} & \boxed{X_n \overset{L_s}{\to} X}
\end{array}$$

The Central Limit Theorem: Continuity Theorem

- (The Continuity Theorem for m.g.f.s) Let X_n be a sequence of r.v.s such that the corresponding m.g.f.s $\mathbb{M}_{X_n}(t)$ exist for $|t| < h$, for some $h > 0$, and all $n \in \mathbb{N}$.
- If X is a random variable whose m.g.f. $\mathbb{M}_X(t)$ exists for $|t| \leq h_1 < h$ for some $h_1 > 0$, then

$$\lim_{n\to\infty} \mathbb{M}_{X_n}(t) = \mathbb{M}_X(t) \text{ for } |t| < h_1 \quad \Rightarrow \quad X_n \xrightarrow{d} X.$$

- See Section 1.1.3 for discussion and Section 4.3.4 for references for proof.

The Central Limit Theorem: Example 1.10(a)

- For $X \sim \mathrm{Gam}(a, b)$, we showed earlier that the m.g.f. is $\left(\frac{b}{b-t}\right)^a$, $t < b$, from which we get $\mathbb{E}[X] = a/b$ and $\mathbb{V}(X) = a/b^2$.
- Recall: If $Y = \mu + \sigma X$, then $\mathbb{M}_X(t) = \mathrm{e}^{t\mu}\mathbb{M}_X(t\sigma)$.
- Let $b > 0$ be a fixed value and, for any $a > 0$, let $X_a \sim \mathrm{Gam}(a, b)$.
- Define $Y_a = (X_a - a/b) / \sqrt{a/b^2} = (b/\sqrt{a})X_a - \sqrt{a}$.
- For $t < a^{1/2}$,

$$\mathbb{M}_{Y_a}(t) = \mathrm{e}^{-t\sqrt{a}}\mathbb{M}_{X_a}\left(\frac{b}{\sqrt{a}}t\right) = \mathrm{e}^{-t\sqrt{a}}\left(\frac{b}{b - ba^{-1/2}t}\right)^a$$
$$= \mathrm{e}^{-t\sqrt{a}}\left(\frac{1}{1 - a^{-1/2}t}\right)^a$$

or $\mathbb{K}_{Y_a}(t) = -t\sqrt{a} - a\log\left(1 - a^{-1/2}t\right)$.

The Central Limit Theorem: Example 1.10(a)

We now showed that

$$\mathbb{K}_{Y_a}(t) = -t\sqrt{a} - a\log\left(1 - a^{-1/2}t\right).$$

From the result

$$\log(1+x) = \sum_{i=1}^{\infty} (-1)^{i+1}\frac{x^i}{i},$$

we have

$$\log\left(1 - a^{-1/2}t\right) = -\frac{t}{a^{1/2}} - \frac{t^2}{2a} - \frac{t^3}{3a^{3/2}} - \cdots.$$

Thus,

$$\lim_{a\to\infty} \mathbb{K}_{Y_a}(t) = t^2/2,$$

which is the cumulant generating function of a standard normal distribution.

The Central Limit Theorem: Example 1.10(a)

- So, as $a \to \infty$, $Y_a \xrightarrow{d} \mathrm{N}(0,1)$, or, for large a, $X_a \overset{\text{app}}{\sim} \mathrm{N}\left(a/b, a/b^2\right)$.
- Recall that (see Example 2.3):
  If $X_i \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(\lambda)$, then $\sum_i^n X_i \sim \mathrm{Gam}(n, \lambda)$.
- This implies: If $X_i \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(\lambda)$, then, as $n \to \infty$, the distribution of the sum $\sum_{i=1}^n X_i$ behaves like that of a random variable with distribution $\mathrm{N}\left(n/\lambda, n/\lambda^2\right)$.
- (In our "slang" notation, $\sum_{i=1}^n X_i \xrightarrow{d} \mathrm{N}\left(n/\lambda, n/\lambda^2\right)$.)

The Central Limit Theorem: Example 1.10(b)

Now let $S_n \sim \mathrm{Gam}(n, 1)$ for $n \in \mathbb{N}$, so that, for large n, $S_n \overset{\text{app}}{\sim} \mathrm{N}(n, n)$. The definition of convergence in distribution, and the continuity of the c.d.f. of S_n and that of its limiting distribution informally suggests the limiting behavior of the p.d.f. of S_n, i.e.,

$$f_{S_n}(s) = \frac{1}{\Gamma(n)} s^{n-1} \exp(-s) \approx \frac{1}{\sqrt{2\pi n}} \exp\left(-\frac{(s-n)^2}{2n^2}\right).$$

Choosing $s = n$ leads to $\Gamma(n+1) = n! \approx \sqrt{2\pi}\,(n+1)^{n+1/2} \exp(-n-1)$. As $\lim_{n\to\infty}(1+\lambda/n)^n = \mathrm{e}^{\lambda}$,

$$(n+1)^{n+1/2} = n^{n+1/2}\left(1+\frac{1}{n}\right)^{n+1/2} \approx n^{n+1/2}\mathrm{e},$$

and substituting this into the previous expression for $n!$ yields Stirling's approximation

$$n! \approx \sqrt{2\pi}\, n^{n+1/2}\, \mathrm{e}^{-n}.$$

The Central Limit Theorem

The previous example is a special case of the *Central Limit Theorem*, or CLT:

$$\text{if} \quad X_i \overset{\text{i.i.d.}}{\sim} \left(\mu, \sigma^2\right), \quad \text{then} \quad \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathrm{N}\left(0, 1\right).$$

This can also be expressed informally as $\sum_{i=1}^{n} X_i \xrightarrow{d} \mathrm{N}\left(n\mu, n\sigma^2\right)$ as $n \to \infty$, but with the understanding that the above statement is correct.

The Central Limit Theorem: Proof

- Let $X_i \overset{\text{i.i.d.}}{\sim} (\mu, \sigma^2)$.
- This means $\{X_i\}$ is a sequence of i.i.d. r.v.s with common mean μ and variance σ^2.
- Additionally assume the m.g.f. of X_i exists on a neighborhood of zero.
- Let $Z_i = (X_i - \mu)/\sigma$ with common m.g.f. $\mathbb{M}_Z(t)$.
- Let $S_n = n^{-1/2} \sum_{i=1}^{n} Z_i$.
- The m.g.f. of $n^{-1/2} Z_i$ is clearly $\mathbb{M}_Z\left(tn^{-1/2}\right)$.
- As the Z_i are also i.i.d., $\mathbb{M}_{S_n}(t) = \left(\mathbb{M}_Z\left(tn^{-1/2}\right)\right)^n$.

The Central Limit Theorem: Proof

Recalling that

$$\mathbb{M}_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}\left[X^k\right],$$

we have

$$\mathbb{M}_Z\Big(tn^{-1/2}\Big) = \sum_{k=0}^{\infty} \frac{\mu'_k}{k!} \left(tn^{-1/2}\right)^k = 1 + \mu'_1 \frac{t}{n^{1/2}} + \frac{\mu'_2}{2} \frac{t^2}{n} + \frac{\mu'_3}{3!} \frac{t^3}{n^{3/2}} + \cdots,$$

where $\mu'_k = \mathbb{M}_Z^{(k)}(t)\Big|_{t=0}$ with $\mu'_0 = 1$, $\mu'_1 = \mathbb{E}[Z_i] = 0$ and $\mu'_2 = \mathbb{V}(Z_i) + (\mathbb{E}[Z_i])^2 = 1$.

The Central Limit Theorem: Proof

As

$$\lim_{n\to\infty}\left(1+\frac{k_1}{n}+\frac{k_2}{n^p}\right)^n = e^{k_1}$$

for $p > 1$,

$$\lim_{n\to\infty}\mathbb{M}_{S_n}(t) = \lim_{n\to\infty}\left(\mathbb{M}_Z\left(tn^{-1/2}\right)\right)^n = \lim_{n\to\infty}\left(1+\frac{t^2}{2n}+\cdots\right)^n = \exp\left(\frac{t^2}{2}\right).$$

Then, from the characterizing property of moment generating functions, the result follows.

Another Proof of the CLT: (Problem 4.15).

- The cumulant generating function of S_n is

$$\begin{aligned}\mathbb{K}_{S_n}(t) &= \ln \mathbb{M}_{S_n}(t) \\ &= \ln \left(\mathbb{M}_Z\left(tn^{-1/2}\right)\right)^n \\ &= n\mathbb{K}_Z\left(tn^{-1/2}\right).\end{aligned}$$

- From the continuity of log and assuming validity of exchange of limit and integral,

$$\lim_{n\to\infty} \mathbb{K}_Z\left(tn^{-1/2}\right) = \ln \lim_{n\to\infty} \mathbb{M}_Z\left(tn^{-1/2}\right) = \ln \mathbb{M}_Z(0) = \ln 1.$$

- Use l'Hôpital's rule to show that $\lim_{n\to\infty} \mathbb{K}_{S_n}(t) = t^2/2$.

Another Proof of the CLT: Solution

We have $\mathbb{K}_{S_n}(t) = n\mathbb{K}_Z\left(tn^{-1/2}\right)$, so that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{K}_{S_n}(t) &= \lim_{n\to\infty} \frac{\mathbb{K}_Z\left(tn^{-1/2}\right)}{1/n} = \lim_{n\to\infty} \frac{\mathbb{K}_Z'\left(tn^{-1/2}\right) t\,(-1/2)\,n^{-3/2}}{-n^{-2}} \\
&= \frac{t}{2} \lim_{n\to\infty} \frac{\mathbb{K}_Z'\left(tn^{-1/2}\right)}{1/n^{1/2}} = \frac{t}{2} \lim_{n\to\infty} \frac{\mathbb{K}_Z''\left(tn^{-1/2}\right) t\,(-1/2)\,n^{-3/2}}{(-1/2)n^{-3/2}} \\
&= \frac{t^2}{2} \lim_{n\to\infty} \mathbb{K}_Z''\left(tn^{-1/2}\right),
\end{aligned}$$

having applied L'Hopitals rule twice.
Now, assuming exchange of limit and derivative,

$$\lim_{n\to\infty} \mathbb{K}_Z''\left(tn^{-1/2}\right) = \mathbb{K}_Z''\left(\lim_{n\to\infty} tn^{-1/2}\right) = \mathbb{K}_Z''(0) = \mathbb{V}(Z) = 1,$$

so that $\lim_{n\to\infty} \mathbb{K}_{S_n}(t) = t^2/2$.

Intermediate Probability: A Computational Approach

Saddlepoint Approximations

The method has been extensively used in statistics and also insurance. A selected list of the use of saddlepoint approximations in finance include:

1. Rogers and Zane (1999): Saddlepoint approximations to option prices. **Annals of Applied Probability** 9, 493-503.
2. Martin, Thompson, Browne (2001); Taking to the saddle. **Risk** 14 (June), 91-94.
3. Duffie and Pan (2001): Analytical value-at-risk with jumps and credit risk. **Finance and Stochastics** 5, 155-180.
4. Gordy (2002), Saddlepoint approximation of Credit Risk. **Journal of Banking and Finance** 26, 1335-1353.
5. Dembo and Deuschel and Duffie (2004): Large portfolio losses. **Finance and Stochastics** 8, 3-16.
6. Glasserman (2004): Large portfolio losses (2004). **Journal of Derivatives** 12, 24–42.

Saddlepoint Approximations

7. Xiong, Wong Salopek (2005): Saddlepoint approximations to option price in a general equilibrium model, **Statistics & Probability Letters** 71, 361-369.
8. Yang, Hurd, Zhang (2006): Saddlepoint approximation method for pricing CDOs. **Journal of Computational Finance** 10, 1-20.
9. Aït-Sahalia and Yu (2006): Saddlepoint approximations for continuous-time Markov processes. **Journal of Econometrics** 134, 507-551.
10. Martin (2006), The Saddlepoint Method and Portfolio Optionalities, **Risk Magazine** 19, 93–95.
11. Wong (2008: Backtesting trading risk of commercial banks using expected shortfall, **Journal of Banking & Finance** 32, 1404-1415.
12. Glasserman, Kyoung-Kuk Kim (2009) Saddlepoint approximations for affine jump-diffusion models, **Journal of Economic Dynamics & Control** 33, 15–36.

Saddlepoint Approximations: Introduction

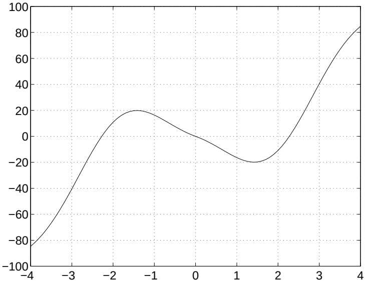
- For many situations occurring in statistics, the density of random variable X arising from a function of other random variables, or as a **test statistic** or **parameter estimate** is either unknown or is algebraically intractable.
- Recalling the **Central Limit Theorem**, if X can be represented as a sum of n (usually independent) r.v.s with finite variance, then, as n increases, the normal distribution is a valid candidate for approximating f_X and F_X.
- The normal approximation tend to be adequate near the mean of X, but typically **breaks down as one moves into the tails**.
- To see this, consider the following two examples:

Example 1: The Student's t Distribution

- Let $X \sim t(\nu)$. Approximate with $Y \sim \mathrm{N}(0, \nu/(\nu-2))$.
- Graph shows the **(relative) percentage error** of the c.d.f. approximation for $\nu = 6$, i.e.,

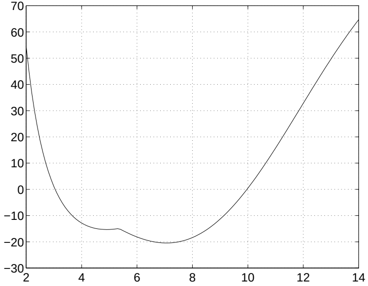
$$100\,(\text{Approx} - \text{True})\,/\min(\text{True}, 1 - \text{True})$$

- It increases without bound as $|x|$ increases!



Example 2: The χ^2 Distribution

- Now let $X \sim \chi^2(6)$, a chi-square r.v. with 6 degrees of freedom.
- Approximated by $Y \sim \mathrm{N}(\nu, 2\nu)$, $\nu = 6$.
- As Y is not symmetric, the error is not zero at the mean.
- Also, because of the truncated true left tail, the percentage error will blow up as x moves towards zero.
- It also increases without bound as x increases.



Derivation

- The basis of the **saddlepoint approximation**, or **SPA** is to overcome the inadequacy of the normal approximation in the tails by "**tilting**" the random variable of interest in such a way that the normal approximation **is evaluated at a point near the mean**.
- Let X be a r.v. with density $f_X(x)$, finite variance, and m.g.f. $\mathbb{M}_X(s)$ existing in a neighborhood U of zero.
- Recall that the mean and variance of X can be expressed as $\mathbb{K}'_X(0)$ and $\mathbb{K}''_X(0)$, respectively, where $\mathbb{K}_X(s) = \ln \mathbb{M}_X(s)$ is the c.g.f. of X.
- For an $s \in U$, define the p.m.f. or p.d.f. of T_s as

$$f_{T_s}(x; s) = \frac{e^{xs} f_X(x)}{\mathbb{M}_X(s)},$$

  and T_s is referred to as an **exponentially tilted** random variable.
- Notice that its density integrates to one!

Derivation

- Its m.g.f. and c.g.f. are easily seen to be

$$\mathbb{M}_{T_s}(t) = \frac{\mathbb{M}_X(t+s)}{\mathbb{M}_X(s)}, \quad \mathbb{K}_{T_s}(t) = \mathbb{K}_X(t+s) - \mathbb{K}_X(s),$$

  so that $\mathbb{E}[T_s] = \mathbb{K}'_{T_s}(0) = \mathbb{K}'_X(s)$ and $\mathrm{Var}(T_s) = \mathbb{K}''_X(s)$.
- Let $s_0 \in U$. Now consider using the normal distribution to approximate the true distribution of T_{s_0};
- It must have mean $x_0 := \mathbb{K}'_X(s_0)$ and variance $v_0 := \mathbb{K}''_X(s_0)$, and is thus given by $x \mapsto \phi(x; x_0, v_0)$, where ϕ is the normal p.d.f..
- The expression for $f_{T_s}(x; s)$ then yields an approximation for f_X as

$$\begin{aligned} x &\mapsto \phi(x; x_0, v_0)\, \mathbb{M}_X(s_0)\, \mathrm{e}^{-s_0 x} \\ &= \frac{1}{\sqrt{2\pi\, v_0}} \exp\left\{-\frac{1}{2v_0}(x - x_0)^2\right\} \mathbb{M}_X(s_0)\, \mathrm{e}^{-s_0 x}. \end{aligned}$$

Derivation

- The accuracy of this approximation to f_X, for a fixed x, depends crucially on the choice of s_0.
- We know that, in general, the normal approximation to the distribution of a random variable X is accurate near the mean of X, but degrades in tails. As such, we are motivated to choose an s_0 such that x is close to the mean of the tilted distribution.
- In particular, we would like to find a value $\hat{s}$ such that

$$\mathbb{K}_X'(\hat{s}) = x,$$

for which it can be shown that a unique solution exists when $\hat{s}$ is restricted to U.

Derivation

The normal density approximation to the tilted r.v. with mean x at the point x is $\phi\left(x; \mathbb{K}_X'(\hat{s}), \mathbb{K}_X''(\hat{s})\right)$, and the approximation for f_X becomes

$$\hat{f}_X(x) = \frac{1}{\sqrt{2\pi\,\mathbb{K}_X''(\hat{s})}} \exp\left\{-\frac{1}{2\mathbb{K}_X''(\hat{s})}(x - x)\right\} \mathbb{M}_X(\hat{s})\, \mathrm{e}^{-\hat{s}x}$$
$$= \frac{1}{\sqrt{2\pi\,\mathbb{K}_X''(\hat{s})}} \exp\left\{\mathbb{K}_X(\hat{s}) - x\hat{s}\right\}.$$

The density SPA

- Summarizing,

$$\hat{f}_X(x) = \frac{1}{\sqrt{2\pi\, \mathbb{K}''_X(\hat{s})}} \exp\left\{\mathbb{K}_X(\hat{s}) - x\hat{s}\right\}, \quad x = \mathbb{K}'_X(\hat{s}) \ .$$

- Approximation $\hat{f}$ is referred to as the **(first order) saddlepoint density approximation** to f, abbreviated **SPA**, where $\hat{s} = \hat{s}(x)$ is the solution to the **saddlepoint equation** and is referred to as the **saddlepoint** at x.

The density SPA

- It is valid for all values of x in the interior of the support of X; for example, if X follows a gamma distribution, then the s.p.a. is valid only for $x > 0$, and if $X \sim \mathrm{Bin}(n, p)$, then the s.p.a. is valid for $x = 1, 2, \ldots, n-1$.
- There always exists a unique root to the saddlepoint equation when $\hat{s}$ is restricted to the convergence strip of the m.g.f. of X.
- Numerically speaking, only very close to the borders of the support of X do numerical problems arise in solving the saddlepoint equation.
- Of course, no problems arise if a closed–form expression for $\hat{s}$ is obtainable, as occurs in simple examples, but also in some very useful and nontrivial cases, such as the important context of the noncentral distributions.

Example: Gamma Distribution

- Let $X \sim \mathrm{Gam}\,(a, b)$ with p.d.f. and m.g.f.

$$f_X(x; a, b) = b^a x^{a-1} \mathrm{e}^{-xb} \mathbb{I}_{(0,\infty)}\,(x) \,/\, \Gamma\,(a)\,, \quad \mathbb{M}_X(s) = \left(\frac{b}{b-s}\right)^a .$$

- The saddlepoint equation is $x = \mathbb{K}'_X(\hat{s}) = a / \,(b - \hat{s})$, with closed–form solution $\hat{s} = b - a/x$.
- Substituting this into the SPA density formula and simplifying yields

$$\hat{f}_X\,(x) = b^a \frac{a^{-a+1/2} \mathrm{e}^a}{\sqrt{2\pi}} x^{a-1} \mathrm{e}^{-xb} \mathbb{I}_{(0,\infty)}\,(x)\,.$$

- Amazingly, approximation $\hat{f}_X$ has exactly the same **kernel** as f_X itself.

Example: Gamma Distribution

- Recalling Stirling's approximation

$$n! \approx (2\pi)^{1/2}\, n^{n+1/2} \mathrm{e}^{-n},$$

  we see that

$$\frac{a^{-a+1/2}\mathrm{e}^{a}}{\sqrt{2\pi}} \approx \frac{1}{\Gamma(a)}.$$

- Thus, the saddlepoint approximation to the gamma density will be extremely accurate for all x and values of a not too small.
- As a increases, the relative accuracy increases. This is quite plausible, recalling that $X = \sum_{i=1}^{a} E_i \sim \mathrm{Gam}(a, b)$, where $E_i \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(b)$; as a increases, the accuracy of the normal approximation to X via the Central Limit Theorem increases.

Renormalization

- The density SPA will not, in general, integrate to one, although it will usually not be far off. It will often be possible to (re)normalize it, i.e.,

$$\bar{\hat{f}}_X(x) = \frac{\hat{f}_X(x)}{\int \hat{f}_X(x)\,\mathrm{d}x},$$

which is a proper density.

- For the gamma example, the normalized density $\bar{\hat{f}}_X$ is exactly equal to f_X.

Example: Sums of two independent Gamma

- Let $X_i \overset{\text{ind}}{\sim} \mathrm{Gam}(a_i, b_i)$, $i = 1, 2$, and $Q = X_1 + X_2$.
- If $b_1 = b_2$, we know the sum is itself gamma distributed. But what if not?
- If the a_i are integer values and $b_1 \neq b_2$, Coelho (1998) gave an easily computable expression for the p.d.f.; see Example 2.12.
- Otherwise, the usual convolution formula can be used, i.e.,

$$f_Q(q) = \int_0^q f_{X_1}(x)\, f_{X_2}(q - x)\, \mathrm{d}x.$$

  This will be relatively slow, as it involves numeric integration for each point of the density.
- Consider the saddlepoint approximation:

Example: Sums of two independent Gamma

- For the s.p.a., the m.g.f. is

$$\mathbb{M}_Q(t) = \mathbb{M}_{X_1}(t)\,\mathbb{M}_{X_2}(t) = \left(\frac{b_1}{b_1 - t}\right)^{a_1}\left(\frac{b_2}{b_2 - t}\right)^{a_2},$$

with $t \in U := (-\infty, \min(b_1, b_2))$, and straightforward computation yields

$$\mathbb{K}_Q'(t) = \frac{a_1}{b_1 - t} + \frac{a_2}{b_2 - t} \quad \text{and} \quad \mathbb{K}_Q''(t) = \frac{a_1}{(b_1 - t)^2} + \frac{a_2}{(b_2 - t)^2}.$$

Example: Sums of two independent Gamma

- The saddlepoint equation is, for a fixed q, $\mathbb{K}_Q'(t) = q$, or $t^2 A + tB + C = 0$, where

$$\begin{aligned} A &= q, \\ B &= (a_1 + a_2) - q(b_1 + b_2) \quad \text{and} \\ C &= b_1 b_2 q - (a_2 b_1 + a_1 b_2). \end{aligned}$$

- The two solutions are $\left(-B \pm \sqrt{B^2 - 4AC}\right)/(2A)$, denoted t_- and t_+, both of which are real. (That both roots are real follows because, alternatively, both are complex, which would rule out the fact that the saddlepoint exists.)
- For a given set of parameter values q, a_1, b_1, a_2, b_2, only one of the two roots t_- and t_+ can be in U, the convergence strip of Q, because we know that, in general, there exists a unique solution to the saddlepoint equation which lies in U.

Example: Sums of two independent Gamma

- If $b_1 = b_2 =: b$, then it is easy to verify that $t_+ = b \notin U$, so that t_- must be the root.
- Computation with several sets of parameter values suggests that t_- is always the correct root (the reader is invited to algebraically prove this), so that

$$\hat{t} = \frac{1}{2A}\left(-B - \sqrt{B^2 - 4AC}\right),$$

and the s.p.a. for f_Q is expressible in closed–form.

Example: Sums of two independent Gamma

- Consider the weighted sum

$$Q = cX_1 + (1-c)\,X_2, \qquad 0 < c < 1, \tag{19}$$

with m.g.f. $\mathbb{M}_Q(t) = \mathbb{M}_{X_1}(ct)\,\mathbb{M}_{X_2}((1-c)\,t)$. The s.p.a. to (19) can similarly be computed. However, it is important to realize that (19) is not a generalization of $Q = X_1 + X_2$ because the parameters c, b_1 and b_2 are not jointly *identified*, i.e., different combinations of them can give rise to the same distribution.

- For example, write Q as $k_1 X_1 + k_2 X_2$, with $X_i \overset{\text{ind}}{\sim} \text{Gam}\,(a_i, 1)$, $k_1 = c/b_1$ and $k_2 = (1-c)\,/b_2$. Then the two values of k are jointly identified. Solving gives $c = 1 - b_2 k_2$, $b_1 = (1 - b_2 k_2)\,/k_1$ and b_2 is free (except zero); the requirement $0 < c < 1$ can be written as

$$0 < c < 1 \Leftrightarrow 0 < 1 - b_2 k_2 < 1 \Leftrightarrow 0 < b_2 < 1/k_2,$$

giving the allowed range of b_2.

Example: Sums of Geometric

- Recall from Chapter 2 the example of collecting prizes from cereal boxes (occupancy problem). In particular, let Y_k be the number of cereal boxes necessary to purchase in order to get at least one of k different prizes, $k = 2, \ldots, r$.
- Define G_i to be the number of required purchases to get a new prize so far not in the collection—in which there are currently i different prizes, $i = 0, 1, \ldots, k-1$.
- Thus, $G_0 = 1$, and

$$Y_k = \sum_{j=0}^{k-1} G_j, \quad G_j \overset{\text{ind}}{\sim} \mathrm{Geo}\left(p_j\right), \quad p_j = \frac{r-j}{r},$$

with $f_{G_j}\left(g; p_j\right) = p_j \left(1 - p_j\right)^{g-1} \mathbb{I}_{\{1,2,\ldots\}}\left(g\right)$.

Example: Sums of Geometric

- We showed earlier that the m.g.f. of G_j is $p_j e^t \left(1 - q_j e^t\right)^{-1}$, $q_j = 1 - p_j$, so that

$$\mathbb{M}_{Y_k}(t) = \prod_{j=0}^{k-1} \mathbb{M}_{G_j}(t) = \mathrm{e}^{kt} \prod_{j=0}^{k-1} \frac{p_j}{1 - q_j \mathrm{e}^t}$$

and

$$\mathbb{K}_{Y_k}(t) = kt + \sum_{i=0}^{k-1} \ln p_i - \sum_{i=0}^{k-1} \ln\left(1 - q_i e^t\right).$$

Example: Sums of Geometric

- The saddlepoint equation is

$$\mathbb{K}'_{Y_k}(t) = k + \sum_{i=0}^{k-1} \frac{q_i \mathrm{e}^t}{1 - q_i \mathrm{e}^t} = y,$$

which needs to be numerically solved in general.

- Convergence of the m.g.f. requires $1 - q_i e^t > 0$ for each i or $1 - \max(q_i e^t) > 0$ or

$$t < -\ln(\max(q_i)) = -\ln(1 - \min(p_i)).$$

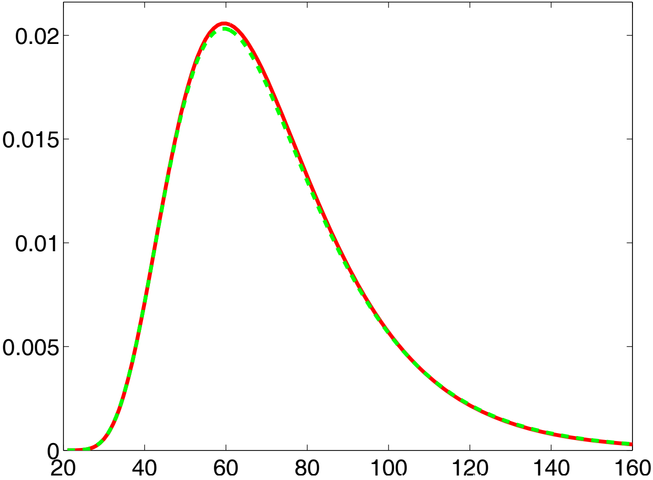
- Next, a simple computation reveals that

$$\mathbb{K}''_{Y_k}(t) = \sum_{i=0}^{k-1} \frac{q_i \mathrm{e}^t}{(1 - q_i \mathrm{e}^t)^2},$$

so that the s.p.a. is straightforward to calculate for ordinates in the interior of the support, i.e., $y = k+1, k+2, \ldots$.

Example: Sums of Geometric: Accuracy

Exact (solid) and s.p.a. (dashed) of the p.m.f. of Y_k with $k = r = 20$.
See Figure 5.3 on page 175 for related graphs.



Example: Sums of Geometric: Code

```
function pdf = occupancyspa(yvec,r,k)
ivec=0:(k-1); pvec=(r-ivec)/r; yl=length(yvec); pdf=zeros(yl,1);
for loop=1:yl
  y=yvec(loop); shat = occspe(y,pvec);  qet=(1-pvec)*exp(shat);
  Ks=k*shat+sum(log(pvec)) - sum(log(1-qet));
  Kpp = sum(qet./(1-qet).^2);
  pdf(loop) = exp(Ks-y*shat) / sqrt(2*pi*Kpp);
end

function spe=occspe(y,pvec)
q=1-pvec; opt=optimset('Display','none','TolX',1e-6);
uplim = -log(max(q)); lolim = -1e1;
spe=fzero(@speeq,[lolim,0.9999*uplim],opt,q,y);

function dd=speeq(t,q,y)
k=length(q); et=exp(t); kp = k+et*sum(q./(1-q*et)); dd = y-kp;
```

CDF Evaluation

- The approximate cdf of X could be obtained by numerically integrating $\hat{f}$.
- However, in a celebrated paper, Lugannani and Rice (1980) derived a simple expression for the c.d.f. as

$$\hat{F}_X(x) = \Pr(X < x) = \Phi(\hat{w}) + \phi(\hat{w})\left\{\frac{1}{\hat{w}} - \frac{1}{\hat{u}}\right\}, \quad x \neq \mathbb{E}[X],$$

- Φ and ϕ are the cdf and pdf of the standard normal, and

$$\hat{w} = \operatorname{sgn}(\hat{s})\sqrt{2\hat{s}x - 2\mathbb{K}_X(\hat{s})} \quad \hat{u} = \begin{cases} \hat{s}\sqrt{\mathbb{K}_X''(\hat{s})}, & \text{if } x \text{ is continuous,} \\ \left(1 - \mathrm{e}^{-\hat{s}}\right)\sqrt{\mathbb{K}_X''(\hat{s})}, & \text{if } x \text{ is discrete.} \end{cases}$$

CDF Evaluation: Some Remarks

The LR expression is an approximation to $\Pr(X < x)$ and not $F_X(x) = \Pr(X \leq x)$.

This is, of course, only relevant when X is discrete.

There are other expressions for the c.d.f. approximation in the discrete case, and these exhibit different accuracies depending on the distribution of X and choice of x; see Butler (2007) for a very detailed and accessible discussion.

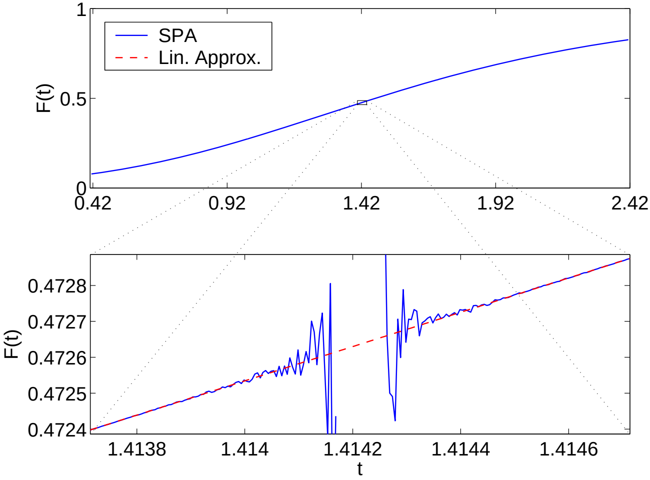
CDF Evaluation Near the Mean

- Recall that, in general, when the m.g.f. exists for r.v. X, $\mathbb{E}[X] = \mathbb{K}'_X(0)$. Thus, if $x = \mathbb{E}[X]$, then $\hat{s} = 0$ is the saddlepoint for $\mathbb{E}[X]$. Clearly, $\mathbb{K}_X(0) = \ln \mathbb{E}[e^{0X}] = 0$, so that $\hat{w} = \operatorname{sgn}(\hat{s}) \sqrt{2\hat{s}x - 2\mathbb{K}_X(\hat{s})} = 0$, and the LR formula cannot be computed.
- This singularity is removable however, and can be shown to be

$$\hat{F}_X(\mathbb{E}[X]) = \frac{1}{2} + \frac{\mathbb{K}'''_X(0)}{6\sqrt{2\pi}\mathbb{K}''_X(0)^{3/2}}.$$

- For practical use however, it is numerically wiser to use linear interpolation based on the s.p.a. to $\mathbb{E}[X] \pm \epsilon$, where ϵ is chosen small enough to ensure high accuracy, but large enough to ensure numerical stability.
- See the next slide for an illustration from Broda and Paolella (2007) in the context of the doubly noncentral t distribution.

CDF Evaluation Near the Mean



CDF Evaluation: 2nd Order Expansion

- The second order p.d.f. approximation is given by

$$\tilde{f}(x) = \hat{f}(x)\left(1 + \frac{\hat{\kappa}_4}{8} - \frac{5}{24}\hat{\kappa}_3^2\right),$$

where $\hat{\kappa}_i = \mathbb{K}_X^{(i)}(\hat{s}) / \left[\mathbb{K}_X''(\hat{s})\right]^{i/2}$.

- Similarly, for the c.d.f.,

$$\tilde{F}(x) = \hat{F}(x) - \phi(\hat{w})\left\{\hat{u}^{-1}\left(\frac{\hat{\kappa}_4}{8} - \frac{5}{24}\hat{\kappa}_3^2\right) - \hat{u}^{-3} - \frac{\hat{\kappa}_3}{2\hat{u}^2} + \hat{w}^{-3}\right\},$$

for $x \neq \mathbb{E}[X]$.

Example: Sum of Normal and Gamma

- Let $Z \sim \mathrm{N}\left(0, \sigma^2\right)$ independent of $G \sim \mathrm{Gam}\,(a, c)$, for $\sigma, a, c \in \mathbb{R}_{>0}$, and let $X = Z + G$.
- We wish to derive the s.p.a. to $f_X(x)$ for $x > 0$.
- The m.g.f. of X is given by

$$\mathbb{M}_X(s;\, \sigma, a, c) = \left(\frac{c}{c-s}\right)^a \mathrm{e}^{(\sigma^2 s^2)/2}, \quad -\infty < s < c,$$

which easily leads to

$$\mathbb{K}_X'(s;\, \sigma, a, c) = \frac{a}{c-s} + \sigma^2 s,$$

implying that the saddlepoint $\hat{s}$ is given by the single real root of the quadratic

$$\sigma^2 s^2 - \left(\sigma^2 c + x\right) s + (xc - a) = 0, \quad \text{such that} \quad s < c.$$

Example: Sum of Normal and Gamma

- Both roots are given by

$$s_{\pm} = \frac{\sigma^2 c + x \pm \sqrt{C}}{2\sigma^2}, \quad C = \left(\sigma^2 c - x\right)^2 + 4\sigma^2 a.$$

  Clearly, $C > 0$, and $\sqrt{C} \geq \sigma^2 c - x$.

- From the constraint $s < c$, this rules out use of s_+, i.e., $\hat{s} = s_- = \left(\sigma^2 c + x - \sqrt{C}\right)/2\sigma^2$, and the saddlepoint equation gives rise to a closed-form solution.
- Higher order derivatives of $\mathbb{K}_X(s)$ are easily obtained; in particular, $\mathbb{K}_X''(s) = a\left(s - c\right)^{-2} + \sigma^2$.

Example: Differences of i.i.d. Gamma

- Let $Z = X_1 - X_2$ for $X_i \overset{\text{i.i.d.}}{\sim} \text{Gam}\,(a)$.
- For $X \sim \text{Gam}\,(a)$, $\mathbb{M}_X(s) = (1-s)^{-a}$, $s < 1$.
- As in Example 2.10,

$$\mathbb{M}_{(-X)}(s) = \mathbb{E}\left[\mathrm{e}^{s(-X)}\right] = \mathbb{M}_X(-s)\,,$$

and it follows that the m.g.f. of Z is given by

$$\begin{aligned} \mathbb{M}_Z(s) &= \mathbb{E}\left[\mathrm{e}^{sZ}\right] = \mathbb{E}\left[\mathrm{e}^{s(X_1 - X_2)}\right] \\ &= \mathbb{M}_X(s)\,\mathbb{M}_X(-s) \\ &= (1-s)^{-a}\,(1+s)^{-a} \\ &= \left(1-s^2\right)^{-a}, \quad |s| < 1. \end{aligned}$$

Example: Differences of i.i.d. Gamma

- Thus,

$$\begin{aligned}\mathbb{K}_Z(s) &= -a\log\left(1-s^2\right), \quad |s|<1,\\ \mathbb{K}'_Z(s) &= \frac{2as}{1-s^2},\\ \mathbb{K}''_Z(s) &= 2a\left(1-s^2\right)^{-1} + 4as^2\left(1-s^2\right)^{-2},\end{aligned}$$

  and the saddlepoint $\hat{s}$ is given by the solution of the quadratic

$$xs^2 + 2as - x = 0,$$

  i.e., $\hat{s} = \left(-a \pm \sqrt{a^2+x^2}\right)/x$ such that $|\hat{s}| < 1$.

- It is noteworthy that the s.p.a. has no numeric trouble at all when a is close to one or for density values in the tails, as does the inversion of the c.f., as discussed in Example 1.23.

Example: Differences of i.i.d. Gamma

- To see which of the two roots is correct, use the fact that $a > 0$ and the requirement that $-1 < s < 1$.
- To confirm that the root with the $+$ is the correct one, we need to show

$$-1 < \frac{-a + \sqrt{a^2 + x^2}}{x} < 1.$$

- For $x > 0$, simple manipulations leads to

$$a^2 - 2ax + x^2 = (a - x)^2 < a^2 + x^2 < (x + a)^2 = 2ax + a^2 + x^2$$

  or $-2ax + x^2 < x^2 < 2ax + x^2$, which is true.
- Similarly, for $x < 0$, we get $-2ax + x^2 > x^2 > 2ax + x^2$, which is true.

Example: Differences of i.i.d. Gamma

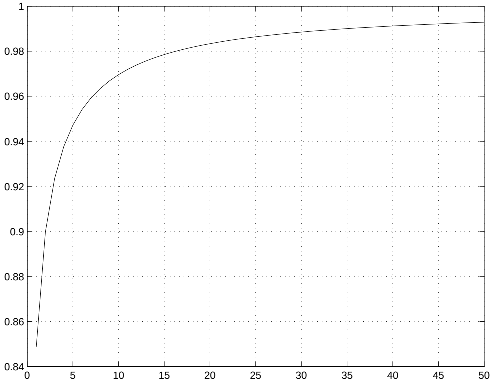
- To normalize the saddlepoint density of Z, use the fact that $x = \mathbb{K}_Z'(\hat{s})$ to give $\mathrm{d}x = \mathbb{K}_Z''(\hat{s})\,\mathrm{d}\hat{s}$.
- Then, using the above expressions for $\mathbb{K}_Z(\hat{s})$, $\mathbb{K}_Z'(\hat{s})$ and $\mathbb{K}_Z''(\hat{s})$, we easily get

$$\begin{aligned} S(a) &= \int_{-\infty}^{\infty} \hat{f}_Z(x)\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} \sqrt{\mathbb{K}_Z''(\hat{s})} \exp\left\{\mathbb{K}_Z(\hat{s}) - \hat{s}\mathbb{K}_Z'(\hat{s})\right\} \mathrm{d}\hat{s} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} \sqrt{2a\left(1-t^2\right)^{-1-2a} + 4at^2\left(1-t^2\right)^{-2(1+a)}} \\ &\quad \times \exp\left\{\frac{-2at^2}{1-t^2}\right\} \mathrm{d}t, \end{aligned}$$

which can be numerically approximated for each value $a > 0$.

Example: Differences of i.i.d. Gamma

The figure shows a plot of $S(a)$ for $a = 1, 2, \ldots, 50$.
Notice that, as $a \to \infty$, $S(a) \uparrow 1$, but rather slowly.



Example: Differences of i.i.d. Gamma

- Instead of integrating each time, one could attempt to fit a low order polynomial or some suitable function in a to the points in the figure.
- Some trial and error leads to the specification

$$S(a) \approx 0.57801 + 0.0036746a - 0.18594a^{1/2} + 0.39464a^{1/3} + 0.092409\left(1 - e^{-a}\right),$$

which exhibits a maximal absolute error of about 0.00045.

Multivariate Saddlepoint Approximations

The p.m.f. or p.d.f. saddlepoint approximation generalizes naturally to the multivariate case:

For a d-dimensional random vector $\mathbf{X}$ having joint c.g.f. $\mathbb{K}$ with gradient $\mathbb{K}'$ and Hessian $\mathbb{K}''$, the approximation is given by

$$\hat{f}_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-d/2} |\mathbb{K}''(\hat{\mathbf{s}})|^{-1/2} \exp\left(\mathbb{K}(\hat{\mathbf{s}}) - \hat{\mathbf{s}}'\mathbf{x}\right),$$

where the multivariate saddlepoint satisfies $\mathbb{K}'(\hat{\mathbf{s}}) = \mathbf{x}$ for $\hat{\mathbf{s}}$ in the convergence region of the m.g.f. of $\mathbf{X}$.

Multivariate Saddlepoint Approximations

The text states the saddlepoint approximations, without proof, and illustrates their use, for the cases of continuous distributions for:

- The c.d.f. of the conditional distribution of $X \mid Y$,
- The bivariate c.d.f. of (X, Y),
- The p.d.f. and c.d.f. of Y_1, where (Y_1, Y_2) is a bijective function of (X_1, X_2).

The latter can be used to derive a s.p.a. for the Student's t distribution (which itself does not have an m.g.f.), the p.d.f. of which is, upon renormalization, exact! More usefully, it can be used for the singly and doubly noncentral t distribution, whose exact p.d.f. and c.d.f. are not trivial to compute.

Appendix: Hypergeometric Functions

Some exercises in this and other chapters require knowledge of the ${}_1F_1$ and ${}_2F_1$ functions. They also arise prominently in the study of noncentral distributions.

The *generalized hypergeometric function* is denoted ${}_jF_k$, with the low order cases ${}_1F_1$ and ${}_2F_1$ being the most popular. For $a, b, c, z \in \mathbb{R}$, they are given by

$$
{}_1F_1(a, b; z) = \sum_{n=0}^{\infty} \frac{a^{[n]}}{b^{[n]}} \frac{z^n}{n!} \quad \text{and} \quad {}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{a^{[n]} b^{[n]}}{c^{[n]}} \frac{z^n}{n!}, \qquad (20)
$$

respectively, where

$$
a^{[j]} = \begin{cases} a(a+1)\cdots(a+j-1), & \text{if } j \geq 1, \\ 1, & \text{if } j = 0, \end{cases} \qquad (21)
$$

is the ascending factorial.

Function ${}_1F_1$ is also referred to as the *confluent hypergeometric function*.

Hypergeometric Functions

Valuable both analytically and for numerical evaluation are the integral equations

$$
{}_1F_1\left(a, b; z\right) = \frac{1}{B\left(a, b-a\right)} \int_0^1 y^{a-1} \left(1-y\right)^{b-a-1} \mathrm{e}^{zy} \, \mathrm{d}y, \quad \text{for} \quad a > 0, \; b-a > 0,
$$

and

$$
{}_2F_1\left(a, b; c; z\right) = \frac{1}{B\left(a, c-a\right)} \int_0^1 y^{a-1}\left(1-y\right)^{c-a-1} \left(1-zy\right)^{-b} \, \mathrm{d}y, \quad \text{for} \quad a > 0, \; c-a > 0, \; z < 1,
$$

but note their parameter restrictions.

See the text for derivation.

Hypergeometric Functions

Example 5.11 shows that

$$ {}_2F_1\left(a, b; c; z\right) = \left(1 - z\right)^{-b} \; {}_2F_1\left(c - a, b; c; \frac{z}{z - 1}\right), $$

while *Kummer's (first) transformation* is given by

$$ {}_1F_1\left(a, b, x\right) = \mathrm{e}^{x}\, {}_1F_1\left(b - a, b, -x\right), $$

confirmed by expanding both sides and comparing coefficients of powers of x.

Also,

$$ \frac{\mathrm{d}^j}{\mathrm{d}z^j}\, {}_1F_1\left(a, b; z\right) = \frac{a^{[j]}}{b^{[j]}}\, {}_1F_1\left(a + j, b + j; z\right) $$

and

$$ \frac{\mathrm{d}^j}{\mathrm{d}z^j}\, {}_2F_1\left(a, b; c; z\right) = \frac{a^{[j]}\, b^{[j]}}{c^{[j]}}\, {}_2F_1\left(a + j, b + j; c + j; z\right). $$

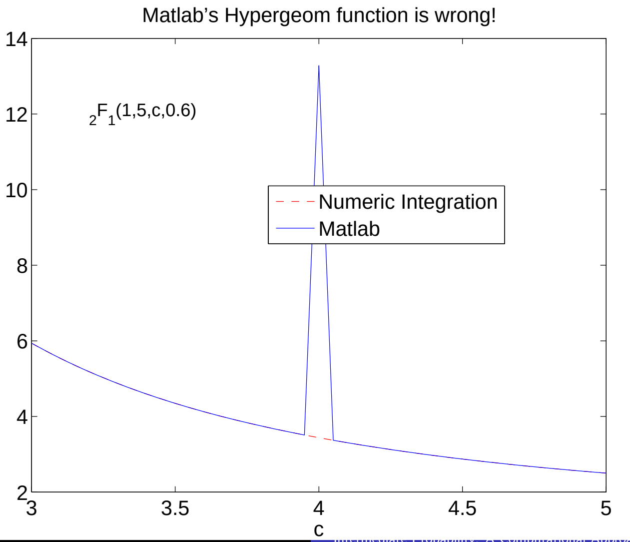
Hypergeometric Functions

Functions ${}_1F_1$ and ${}_2F_1$ can be approximated by truncating the infinite sum or applying numerical integration to the above expressions.

The Maple toolbox in Matlab can be used to compute them by calling `mfun('Hypergeom',a,b,x)` and `mfun('Hypergeom',[a,b],c,x)` for ${}_1F_1(a, b; x)$ and ${}_2F_1(a, b; c; x)$, respectively. Calling `Hypergeom` also works.

This is much faster than 'doing it yourself' and the results are near machine precision. However, it pays to check the results. While it works now, in the 2008 version of Matlab, we got...

Hypergeometric Functions



Hypergeometric Functions

Example 5.12 shows that

$$\Gamma_x(a) = {}_1F_1(a, a+1; -x).$$

It is straightforward to show that

$$B_x(p, q) = \frac{x^p}{p} \, {}_2F_1(p, 1-q; 1+p; x).$$

Example 5.13 shows that

$$\Phi(z) = \frac{1}{2} + z\,\phi(z)\, {}_1F_1\left(1, \frac{3}{2}; \frac{z^2}{2}\right),$$

where $\phi(z)$ is the standard normal p.d.f..

Exercises 1/4

Prove the relation between the incomplete beta function and the confluent hypergeometric function,

$$B_x\left(p,q\right) = \frac{x^p}{p}\ {}_2F_1\left(p, 1-q; 1+p; x\right), \tag{22}$$

as found in, e.g., Gradshteyn and Ryzhik (2007, p. 910).

Exercises 2/4

Using (22), prove that

$$B_x\left(p,q\right)=\sum_{j=0}^{\infty}(-1)^j\binom{q-1}{j}\frac{x^{p+j}}{p+j}. \tag{23}$$

Recall that the c.d.f. of the Student's t is given by

$$F_T\left(t\right)=\frac{1}{2}\bar{B}_L\left(\frac{n}{2},\frac{1}{2}\right),\quad L=\frac{n}{n+t^2},\quad t<0. \tag{24}$$

Use (23) for computing $F_T\left(t;n\right)$ and compare the results to the exact values.

Explain when the formula will work well (meaning, fast convergence), and when not.

Exercises 3/4

Another relation is

$$B_x(p,q) = \frac{x^p (1-x)^q}{p} \, {}_2F_1(1, p+q; 1+p; x). \tag{25}$$

This can be proven directly with the right substitution which, after some trial and error, is $u = x(1-y)/(1-xy)$, where y is the variable of integration, and, solving, $y = (x-u)/(x(1-u))$. Use this to verify (25). Another way of proving (25) is to use a transformation formula of the confluent hypergeometric function. In particular, Example 5.11 showed that

$${}_2F_1(a,b;c;z) = (1-z)^{-b} \, {}_2F_1\left(c-a, b; c; \frac{z}{z-1}\right). \tag{26}$$

Use this to show

$$\begin{aligned} {}_2F_1(a,b;c;x) &= (1-z)^{c-a-b} \, {}_2F_1(c-b, c-a; c; z) \\ &= (1-z)^{c-a-b} \, {}_2F_1(c-a, c-b; c; z). \end{aligned} \tag{27}$$

Result (25) then immediately follows from (22) and (27).

Exercises 4/4

A different derivation of the c.d.f. (see Problem 7.8) yields that, for all t,

$$F_T(t; n) = \bar{B}_y\left(\frac{n}{2}, \frac{n}{2}\right), \quad y = \frac{1}{2} + \frac{t}{2\sqrt{t^2 + n}}. \tag{28}$$

Construct an expression like (23) for evaluating $F_T(t; n)$ as given in (28) and compare its numeric performance to the use of (23) and (42).

Solutions 1/4

For (22), with $u = xy$ (recalling the incomplete beta integral, we want to choose a substitution such that, when $y = 0$, $u = 0$ and when $y = 1$, $u = x$) and that $B(p, 1) = 1/p$,

$$\begin{aligned}
& \frac{x^p}{p} \, {}_2F_1(p, 1-q; 1+p; x) \\
= \; & \frac{x^p}{p} \frac{1}{B(p,1)} \int_0^1 y^{p-1} (1-y)^0 (1-xy)^{q-1} \, \mathrm{d}y \\
= \; & x^p \int_0^1 y^{p-1} (1-xy)^{q-1} \, \mathrm{d}y = \int_0^x u^{p-1} (1-u)^{q-1} \, \mathrm{d}u \\
= \; & B_x(p, q).
\end{aligned}$$

Solutions 2/4

For (23), recalling from (I.1.11) the expansion

$$(1-x)^t = \sum_{j=0}^{\infty} \binom{t}{j} (-x)^j, \quad |x| < 1,$$

we have

$$\begin{aligned}
\frac{x^p}{p}\ {}_2F_1(p, 1-q; 1+p; x) &= x^p \int_0^1 y^{p-1} (1-xy)^{q-1}\, \mathrm{d}y \\
&= x^p \int_0^1 y^{p-1} \sum_{j=0}^{\infty} \binom{q-1}{j} (-xy)^j\, \mathrm{d}y \\
&= \sum_{j=0}^{\infty} (-1)^j \binom{q-1}{j} \frac{x^{p+j}}{p+j}.
\end{aligned}$$

Solutions 2/4

For the Student's t, taking $n = 5$ and $t = -3.36493$, the exact c.d.f. is $F_T(t; 5) = 0.01$ to 8 significant digits. With

$$B\left(\frac{n}{2}, \frac{1}{2}\right) = \frac{\Gamma(5/2)\,\Gamma(1/2)}{\Gamma(3)} = \frac{3\pi}{8}$$

and truncating (23) at $j_{\max} = 5$ gives 4 digit accuracy; with $j_{\max} = 10$, we get 8 digit accuracy. The sum converges faster for smaller x, which corresponds to values of t further in the tail.

As $t \to 0$, $x \to 1$ and the series is useless for computation—using $t = 0$ and 1001 terms in the sum yields 0.485, the true answer being 0.5.

Solutions 3/4

For (25), with the suggested substitution, we see

$$y = \frac{x-u}{x(1-u)}, \quad \mathrm{d}y = \frac{(x-1)}{x(1-u)^2}\,\mathrm{d}u, \quad 1-y = \frac{u(1-x)}{x(1-u)}, \quad 1-xy = \frac{1-x}{1-u}$$

and

$$\begin{aligned}
& \frac{x^p(1-x)^q}{p}\frac{1}{B(1,p)}\int_0^1 y^0(1-y)^{p-1}(1-xy)^{-(p+q)}\,\mathrm{d}y \\
= \quad & x^p(1-x)^q\int_0^x\left(\frac{u(1-x)}{x(1-u)}\right)^{p-1}\left(\frac{1-x}{1-u}\right)^{-(p+q)}\frac{1-x}{x(1-u)^2}\,\mathrm{d}u \\
= \quad & \int_0^x u^{p-1}(1-u)^{q-1}\,\mathrm{d}u \\
= \quad & B_x(p,q).
\end{aligned}$$

Solutions 3/4

Applying (26) to itself yields the original ${}_2F_1\left(a, b; c; z\right)$. But, as ${}_2F_1\left(a, b; c; z\right) = {}_2F_1\left(b, a; c; z\right)$ (obvious from its definition as an infinite sum), we apply (26) to itself but reversing a and b, or

$$
\begin{aligned}
&{}_2F_1\left(a, b; c; z\right) = (1-z)^{-b}\ {}_2F_1\left(c-a, b; c; \frac{z}{z-1}\right) \\
= \quad & (1-z)^{-b}\ {}_2F_1\left(b, c-a; c; \frac{z}{z-1}\right) \\
= \quad & (1-z)^{-b}\left(1-\frac{z}{z-1}\right)^{a-c}\ {}_2F_1\left(c-b, c-a; c; \frac{\frac{z}{z-1}}{\frac{z}{z-1}-1}\right) \\
= \quad & (1-z)^{c-a-b}\ {}_2F_1\left(c-b, c-a; c; z\right).
\end{aligned}
$$

Solutions 4/4

From (22), $B_x(p,q) = \frac{x^p}{p}\ {}_2F_1(p, 1-q; 1+p; x)$ so that, as before, with $(1-x)^t = \sum_{j=0}^{\infty} \binom{t}{j} (-x)^j$,

$$
\begin{aligned}
F_T(t;n) &= \bar{B}_y\left(\frac{n}{2},\frac{n}{2}\right), \quad y = \frac{1}{2} + \frac{t}{2\sqrt{t^2+n}}, \\
&= \frac{1}{B(n/2,n/2)} \frac{y^{n/2}}{n/2}\ {}_2F_1\left(\frac{n}{2}, 1-\frac{n}{2}; 1+\frac{n}{2}; y\right) \\
&= \frac{1}{B(n/2,n/2)} \frac{y^{n/2}}{n/2} \frac{1}{B(n/2,1)} \int_0^1 w^{n/2-1}(1-w)^0 (1-yw)^{n/2-1}\, \mathrm{d}w \\
&= \frac{1}{B(n/2,n/2)} \sum_{j=0}^{\infty} \binom{n/2-1}{j} (-1)^j \frac{y^{n/2+j}}{n/2+j}. \qquad (29)
\end{aligned}
$$

Because of the way y is defined, we expect (29) to converge quickly for all values of t in $(-\infty, 0)$.

Solutions 4/4

Indeed, for $n = 5$ and $t = -3.36493$, and with $B\left(\frac{n}{2}, \frac{n}{2}\right) = 3\pi/128$, computing (29) with just $j_{\max} = 4$ yields the answer of 0.01 to 7 significant digits.

The worst case is for $t = 0$, in which $y = 1/2$; then, with $j_{\max} = 8$, (29) is correct (0.5) with 5 significant digits.

As a last case, with $n = 1$ and $t = -1$, the exact c.d.f. is $1/4$. With $j_{\max} = 4$, (29) yields 6 significant digits.

Intermediate Probability: A Computational Approach

Order Statistics: Motivation

- Statistics such as $\bar{X}_n$ and S_n^2 are of great importance for ascertaining basic properties about the underlying distribution from which an iid sample has been observed.
- However, there are many occasions in which the **extremes** of the sample, or the smallest and largest observations, are more relevant.
- The choice of height of a water dam, the size of a military force, or an optimal financial strategy to avoid bankruptcy will be at least influenced, if not dictated, not by average or typical behavior, but by consideration of maximal occurrences.
- Similar arguments can be made for the behavior of the minimum; the saying about a chain's strength being only that of its weakest link comes to mind.

Order Statistics: Motivation (2)

- There are other sorted values besides the extremes which are useful for inference. For example, another **measure of central tendency** besides the sample mean is the sample **median**, or the middle observation of the sorted vector.
- The median is useful when working with heavy tailed data, such as Cauchy, and plays a prominent role in **robust** statistical inference, which is designed to minimize the influence of contaminated or suspect data.

Order Statistics: Definition

- The **order statistics** of a random iid sample X_i, $i = 1, \ldots, n$, are the n values arranged in **ascending** order and denoted

$$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$$

or

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

or

$$Y_1 \leq Y_2 \leq \ldots \leq Y_n.$$

- The best notation depends on the context.

Order Statistics: Distribution

- The i^{th} order statistic of iid sample X_i, $i = 1, \ldots, n$ from distribution $F = F_X$ (and density $f = f_X$) is distributed with c.d.f.

$$F_{Y_i}(y) = \Pr(Y_i \leq y) = \sum_{j=i}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}.$$

- To see this, let $Z_i = \mathbb{I}_{(-\infty, y]}(X_i)$ so that the Z_i are i.i.d. Bernoulli and $S_y = \sum_{i=1}^{n} Z_i \sim \mathrm{Bin}(n, F(y))$ is the number of X_i which are less than or equal to y. Then

$$\Pr(Y_i \leq y) = \Pr(\text{at least } i \text{ of the } X\text{'s are} \leq y) = \Pr(S_y \geq i).$$

- Special cases of interest are the sample minimum

$$F_{Y_1}(y) = 1 - [1 - F(y)]^n$$

and sample maximum $F_{Y_n}(y) = [F(y)]^n$.

Example of Min and Max

Let $X_i \overset{\text{i.i.d.}}{\sim} \mathrm{Par\,II}\,(1)$, $i = 1, \ldots, n$, each with c.d.f.

$$F(x) = [1 - (1+x)^{-1}]\mathbb{I}_{(0,\infty)}(y).$$

Then, with $S_n = nX_{1:n}$ and $F_{X_{1:n}}(y) = 1 - [1 - F(y)]^n$,

$$\begin{aligned} F_{S_n}(y) &= \Pr(S_n \le y) = \Pr\left(X_{1:n} \le \frac{y}{n}\right) = F_{X_{1:n}}\left(\frac{y}{n}\right) \\ &= 1 - \left[1 - F\left(\frac{y}{n}\right)\right]^n = 1 - \left(\frac{1}{1 + y/n}\right)^n \\ &= \left[1 - \left(1 + \frac{y}{n}\right)^{-n}\right]\mathbb{I}_{(0,\infty)}(y). \end{aligned}$$

Thus, $\lim_{n\to\infty} F_{S_n}(y) = (1 - e^{-y})\,\mathbb{I}_{(0,\infty)}(y)$, and $S_n \xrightarrow{d} \mathrm{Exp}(1)$.

Example of Min and Max

Now let $L_n = X_{n:n}/n$, so that, using $F_{X_{n:n}}(y) = [F(y)]^n$,

$$F_{L_n}(y) = F_{X_{n:n}}(ny) = [F(ny)]^n = \left(1 + \frac{1}{ny}\right)^{-n} \mathbb{I}_{(0,\infty)}(y)$$

and $\lim_{n\to\infty} F_{L_n}(y) = \exp(-1/y)\, \mathbb{I}_{(0,\infty)}(y)$.

Example of Gumbel Distribution

- Recall that the Gumbel, or extreme value, distribution, $\mathrm{Gum}(\lambda)$ has c.d.f. given by

$$F_{\mathrm{Gum}}(x; a, b) = \exp\left(-\exp\left(-\frac{x-a}{b}\right)\right).$$

- Let $X_i \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(\lambda)$, with c.d.f. $F(x) = \left(1 - e^{-\lambda x}\right) \mathbb{I}_{(0,\infty)}(x)$.
- Let $L_n = \lambda X_{n:n} - \ln n$. From the result $F_{X_{n:n}}(y) = [F(y)]^n$,

$$\begin{aligned} F_{L_n}(y) &= \Pr\left(X_{n:n} \leq \frac{y + \ln n}{\lambda}\right) = \left[F\left(\frac{y+\ln n}{\lambda}\right)\right]^n \\ &= \left(1 - \frac{e^{-y}}{n}\right)^n \mathbb{I}_{(-\ln n,\infty)}(y). \end{aligned}$$

- Thus, $\lim_{n\to\infty} F_{L_n}(y) = \exp\left(-e^{-y}\right)$, or $L_n \xrightarrow{d} \mathrm{Gum}(0,1)$.

Example of Gumbel Distribution: Simulation

- The construction of L_n suggests a way of simulating a $\mathrm{Gum}(0,1)$ random variable: take the maximum of a set of (say, $n = 20,000$) i.i.d. $\mathrm{Exp}(1)$ r.v.s and subtract $\ln n$.
- Alternatively, the use of the probability integral transform for simulation is applicable and vastly simpler; we solve $x = \exp\left(-e^{-y}\right)$ to get $y = -\ln(-\ln x)$.
- Both methods were used and produced the histograms shown in the book. They look indistinguishable.

Mass Function of Order Statistics

For discrete F_X with k support points $\{x_1, \ldots, x_k\}$, the mass function f_{Y_i} can be evaluated in the obvious way as

$$f_{Y_i}(y) = \begin{cases} F_{Y_i}(x_1), & \text{if } y = x_1, \\ F_{Y_i}(x_i) - F_{Y_i}(x_{i-1}), & \text{if } y = x_i \in \{x_2, \ldots, x_k\}, \\ 0, & \text{otherwise.} \end{cases}$$

See the text for a more detailed discussion of this.

Mass Function of Order Statistics: Example

Recall that random variable X follows a geometric distribution with parameter $p \in (0,1)$, written $X \sim \mathrm{Geo}(p)$, if it has p.m.f.

$$f_X(x) = f_{\mathrm{Geo}}(x;p) = p(1-p)^x \, \mathbb{I}_{\{0,1,\dots\}}(x),$$

and represents the number of "failed" Bernoulli trials required until (and not including) a "success" is observed.

A straightforward calculation reveals that, with $q = 1 - p$,

$$F_X(x) = \sum_{k=0}^{x} f_X(k) = p \sum_{k=0}^{x} q^k = \left(1 - q^{x+1}\right) \mathbb{I}_{\{0,1,\dots\}}(x), \qquad (30)$$

with $F_X(0) = p$.

Mass Function of Order Statistics: Example

Let $X_i \overset{\text{i.i.d.}}{\sim} \mathrm{Geo}(p)$, $i = 1, \ldots, n$, and $Y_1 = \min X_i$. As

$$F_{Y_1}(y) = 1 - [1 - F_X(y)]^n = 1 - (1-p)^{n(y+1)},$$

we have $f_{Y_1}(0) = F_{Y_1}(0) = 1 - (1-p)^n$. This makes sense, because $f_{Y_1}(0)$ is the probability that the smallest X_i is zero, meaning at least one X_i was zero, or, one minus the probability that all X_i are at least one, or

$$1 - \Pr(\text{all } X_i \geq 1) = 1 - [1 - \Pr(X_i = 0)]^n = 1 - [1-p]^n.$$

For $y \in \{1, 2, \ldots\}$,

$$\begin{aligned} f_{Y_1}(y) &= F_{Y_1}(y) - F_{Y_1}(y-1) \\ &= 1 - (1-p)^{n(y+1)} - \left[1 - (1-p)^{ny}\right] = q^{ny}(1 - q^n). \end{aligned} \quad (31)$$

For $n = 2$, this is $f_{Y_1}(y) = q^{2y}\left(1 - q^2\right)$.

Mass Function of Order Statistics: Example

Now let $X_i \overset{\text{indep}}{\sim} \mathrm{Geo}(p_i)$, $q_i = 1 - p_i$, $M = \min(X_1, X_2)$. Then $\Pr(M \leq m) = \Pr(X_1 \leq m \text{ or } X_2 \leq m)$, or easier, from (30), $\Pr(M > m)$ is

$$\Pr(X_1 \leq m \text{ and } X_2 \leq m) = \Pr(X_1 \leq m)\Pr(X_2 \leq m) = (q_1 q_2)^{m+1},$$

so that $F_M(m) = 1 - \Pr(M > m) = 1 - (q_1 q_2)^{m+1}$.

We have

$$\begin{aligned} f_M(m) &= \Pr(M = m) = \Pr(M > m-1) - \Pr(M > m) \\ &= (q_1 q_2)^m - (q_1 q_2)^{m+1} = (q_1 q_2)^m (1 - q_1 q_2). \end{aligned}$$

If $q_1 = q_2 =: q$, then $f_M(m) = q^{2m}\left(1 - q^2\right)$, which agrees with (31) for $n = 2$.

Density Function of Order Statistics

For continuous F_X, an easy way to derive the pdf of Y_i is by using a limiting argument based on the multinomial distribution.

That is, $f_{Y_i}(y)$ is given by

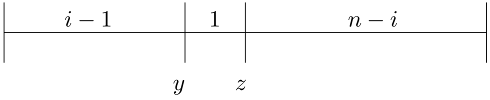
$$\lim_{\Delta y \to 0} \frac{\Pr\left[i-1\ X_i \in (-\infty, y]\,;\ \text{one } X_i \in (y, y+\Delta y)\,; n-i\ X_i \in (Y+\Delta y, \infty)\right]}{\Delta y}$$

which implies

$$f_{Y_i}(y) = \frac{n!}{(i-1)!\,(n-i)!} F(y)^{i-1} \left[1 - F(y)\right]^{n-i} f(y)\,.$$

It is easy to verify that

$$\binom{n}{i-1, 1, n-i} = \frac{n!}{(i-1)!\,(n-i)!} = i\binom{n}{i}.$$



Density Function of Order Statistics

Notice the two special cases, **min** and **max**, follow from the general expression

$$f_{Y_i}(y) = \frac{n!}{(i-1)!\,(n-i)!} F(y)^{i-1} [1 - F(y)]^{n-i} f(y).$$

or as derivatives of the corresponding c.d.f. expressions

$$F_{Y_1}(y) = 1 - [1 - F(y)]^n, \quad F_{Y_n}(y) = [F(y)]^n,$$

namely:

$$f_{Y_1}(y) = n[1 - F(y)]^{n-1} f(y) \quad \text{and} \quad f_{Y_n}(y) = n[F(y)]^{n-1} f(y).$$

Density and Distribution Function

It is easy to see that differentiating

$$F_{Y_i}(y) = \frac{n!}{(n-i)!\,(i-1)!} \int_0^{F(y)} t^{i-1} (1-t)^{n-i} \, \mathrm{d}t \tag{32}$$

yields the p.d.f. expression

$$f_{Y_i}(y) = \frac{n!}{(i-1)!\,(n-i)!} F(y)^{i-1} [1 - F(y)]^{n-i} f(y),$$

so that (32) is another expression for the c.d.f. of Y_i in the continuous case.

Repeated integration by parts shows that (32) is equivalent to the previous expression

$$F_{Y_i}(y) = \Pr(Y_i \leq y) = \sum_{j=i}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}.$$

Example with Uniform

Let $X_i \stackrel{\text{iid}}{\sim} \mathrm{Unif}(0,1)$. From the general p.d.f. expression

$$f_{Y_i}(y) = \frac{n!}{(i-1)!\,(n-i)!} F(y)^{i-1} \left[1 - F(y)\right]^{n-i} f(y),$$

we have, with $f(y) = \mathbb{I}_{(0,1)}(y)$ and $F(y) = y\mathbb{I}_{(0,1)}(y) + \mathbb{I}_{[1,\infty)}(y)$,

$$\begin{aligned} f_{Y_i}(y) &= \frac{n!}{(i-1)!\,(n-i)!} y^{i-1} (1-y)^{n-i} \\ &= \frac{\Gamma(i)\,\Gamma(n-i+1)}{\Gamma(n+1)} y^{i-1} (1-y)^{n-i} \\ &= \frac{1}{B(i, n-i+1)} y^{i-1} (1-y)^{n-i} \end{aligned}$$

so that $Y_i \sim \mathrm{Beta}(i, n-i+1)$.

Example with Uniform

The c.d.f. of Y_i can be expressed using

$$F_{Y_i}(y) = \Pr(Y_i \leq y) = \sum_{j=i}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}$$

or integrating the expression for the p.d.f. $f_{Y_i}(y)$.

Doing so, and equating the two yields, for $0 \leq y \leq 1$, the identity

$$F_{Y_i}(y) = \sum_{j=i}^{n} \binom{n}{j} y^j (1-y)^{n-j} = \frac{n!}{(i-1)!\,(n-i)!} \int_0^y x^{i-1} (1-x)^{n-i}\, \mathrm{d}x. \tag{33}$$

This gives rise to an interesting identity as well as a computational method for evaluating the incomplete beta function.

Nonparametric Confidence Interval for a Quantile

Recall that the quantile ξ_p of continuous r.v. X is that value such that $F_X(\xi_p) = p$ for given probability $0 < p < 1$.

To calculate

$$\Pr(Y_i \leq \xi_p \leq Y_j)$$

for $1 \leq i < j \leq n$, let $U_i = F_X(Y_i)$ and use a similar derivation as for the probability integral transform, i.e.,

$$\begin{aligned} F_{U_i}(u) = \Pr(U_i \leq u) &= \Pr(F_X(Y_i) \leq u) \\ &= \Pr\left(Y_i \leq F_X^{-1}(u)\right) = F_{Y_i}\left(F_X^{-1}(u)\right). \end{aligned}$$

Nonparametric Confidence Interval for a Quantile

From the general formula

$$F_{Y_i}(y) = \Pr(Y_i \le y) = \sum_{j=i}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j},$$

$F_{U_i}(u)$ is given by

$$\begin{aligned} F_{Y_i}\left(F_X^{-1}(u)\right) &= \sum_{j=i}^{n} \binom{n}{j} \left[F_X\left(F_X^{-1}(u)\right)\right]^j \left[1 - F_X\left(F_X^{-1}(u)\right)\right]^{n-j} \\ &= \sum_{j=i}^{n} \binom{n}{j} u^j (1-u)^{n-j}. \end{aligned}$$

Now, we see from the left side of (33) that the $F_X(Y_1), \ldots, F_X(Y_n)$ have the same distribution as order statistics from an i.i.d. uniform sample.

Nonparametric Confidence Interval for a Quantile

Thus, with $\xi_p = F_X^{-1}(p)$, $\Pr(Y_i \leq \xi_p \leq Y_j)$ is given by

$$\begin{aligned}
& \Pr\left(F_X(Y_i) \leq p \leq F_X(Y_j)\right) \\
= \;& \Pr\left(U_i \leq p \leq U_j\right) \\
= \;& \Pr\left(U_i \leq p \;\cap\; p \leq U_j\right) \\
= \;& \Pr\left(U_i \leq p\right) + \left(1 - \Pr\left(U_j < p\right)\right) - \Pr\left(U_i \leq p \;\cup\; p \leq U_j\right) \\
= \;& \sum_{k=i}^{n} \binom{n}{k} p^k (1-p)^{n-k} + \left(1 - \sum_{k=j}^{n} \binom{n}{k} p^k (1-p)^{n-k}\right) - 1
\end{aligned}$$

because, with $i < j$, $\Pr\left(U_i \leq p \;\cup\; p \leq U_j\right) = 1$, i.e.,

$$\Pr\left(Y_i \leq \xi_p \leq Y_j\right) = \sum_{k=i}^{j-1} \binom{n}{k} p^k (1-p)^{n-k} = F_B(j-1, n, p) - F_B(i, n, p) \tag{34}$$

where $B \sim \mathrm{Bin}(n, p)$ and F_B is the c.d.f. of B.

Nonparametric Confidence Interval for a Quantile

A natural (and popular) use of (34) is to construct a *nonparametric confidence interval* for ξ_p. This is attractive because f_X and F_X are not needed (and in reality, F_X is not known!)
For example, to construct a 95% confidence interval for the population median based on $n = 100$ i.i.d. samples of data, we require values i and j such that

$$F_B\left(j-1, 100, 0.5\right) - F_B\left(i, 100, 0.5\right) \approx 0.95,$$

which need to be numerically found. With

$$F_B^{-1}\left(0.025, 100, 0.5\right) = 40 \quad \text{and} \quad F_B^{-1}\left(1-0.025, 100, 0.5\right) = 60$$

(computed using the `binoinv` function in Matlab), we calculate

$$F_B\left(59, 100, 0.5\right) - F_B\left(40, 100, 0.5\right) = 0.943.$$

Thus, $\Pr\left(Y_{40} < \xi_{0.5} < Y_{59}\right) = 0.94$ based on $n = 100$.

Exercise: Example 6.8

- Let X and Y be independent, continuous r.v.s, but not necessarily identically distributed.
- Let $M = \min(X, Y)$.
- Derive an expression for f_M.
- Calculate and simplify the formula for f_M when $X \sim \mathrm{Exp}(\theta_1)$ independent of $Y \sim \mathrm{Exp}(\theta_2)$, where $X \sim \mathrm{Exp}(\theta)$ has density $f_X(x) = \theta e^{-\theta x} \mathbb{I}_{(0,\infty)}(x)$, and compare the result to a simulation.

Exercise: Example 6.8, Solution

The survivor function of $M = \min(X, Y)$ is

$$\begin{aligned}\Pr(M > m) &= \Pr(\min(X, Y) > m) = \Pr(X > m \text{ and } Y > m) \\ &= \Pr(X > m)\Pr(Y > m) = \big(1 - F_X(m)\big)\big(1 - F_Y(m)\big),\end{aligned}$$

and the p.d.f. of M is

$$f_M(m) = f_X(m)\big(1 - F_Y(m)\big) + f_Y(m)\big(1 - F_X(m)\big),$$

as $f_M(m) = \mathrm{d}F_M(m)/\mathrm{d}m = -\mathrm{d}\Pr(M > m)/\mathrm{d}m$.

If X and Y are i.i.d. with p.d.f. and c.d.f. f and F, respectively, then this simplifies to $f_M(m) = 2f(m)[1 - F(m)]$, which of course agrees with the the general formula for minimum of n i.i.d. r.v.s, $f_{Y_1}(y) = n[1 - F(y)]^{n-1} f(y)$, for $n = 2$.

Exercise: Example 6.8, Solution

In the special case of exponential r.v.s., $F_X(x;\theta) = 1 - e^{-\theta x}$ so

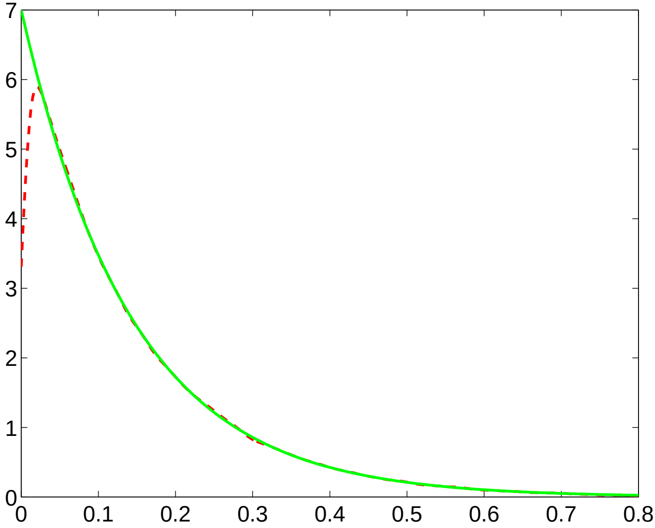
$$\begin{aligned} f_M(m) &= \theta_1 e^{-\theta_1 x}\left(e^{-\theta_2 x}\right) + \theta_2 e^{-\theta_2 x}\left(e^{-\theta_1 x}\right) \\ &= (\theta_1 + \theta_2)\, e^{-x(\theta_1+\theta_2)}. \end{aligned}$$

Matlab code to compute $n = 100,000$ replications of the minimum, and its kernel density, and plot it against the p.d.f. of an exponential with parameter $\theta_1 + \theta_2$ is given as follows, using $\theta_1 = 2$ and $\theta_2 = 5$.

```
n=100000; t1=2; t2=5;
S=min([exprnd(1,n,1)/t1 , exprnd(1,n,1)/t2]');
grd=0:0.0025:0.8;
[pdf]=ksdensity(S,grd);
xx=0:0.01:0.8; den=(t1+t2)*exp(-(t1+t2)*xx);
plot(grd,pdf,'r--',xx,den,'g-','linewidth',2)
axis([0 0.8 0 7]), set(gca,'fontsize',16)
```

The graphic is shown on the next page, with red dashes being the kernel density from the simulation, and green the true p.d.f..

Exercise: Example 6.8, Solution



Exercise: Problem 6.2

Let $X_i \overset{\text{ind}}{\sim} \text{Exp}(\lambda_i)$, $i = 1, \ldots, n$, and let $S = \min(X_i)$.

Derive a simple expression for $\Pr(X_i = S)$.

Hint: Recall that, for event A,

$$\Pr(A) = \int_{-\infty}^{\infty} \Pr(A \mid X = x) \, \mathrm{d}F_X(x)$$

Exercise: Problem 6.2, Solution

If all the λ_i are equal, then, because the X_i are i.i.d., the probability is just $1/n$. Otherwise, use (i) $\Pr(A) = \int_{-\infty}^{\infty} \Pr(A \mid X = x) \, \mathrm{d}F_X(x)$, (ii) independence, and (iii) $F_{X_i}(x) = 1 - e^{-\lambda_i x}$, to get

$$\begin{aligned}
\Pr(X_i = S) &= \int_0^\infty \Pr(X_i = S \mid X_i = x) \, \lambda_i e^{-\lambda_i x} \, \mathrm{d}x \\
&= \int_0^\infty \Pr(X_j > x, \ j \neq i) \, \lambda_i e^{-\lambda_i x} \, \mathrm{d}x \\
&= \int_0^\infty \prod_{j \neq i} \left(1 - F_{X_j}(x)\right) \lambda_i e^{-\lambda_i x} \, \mathrm{d}x \\
&= \lambda_i \int_0^\infty e^{-(\lambda_1 + \cdots + \lambda_n)x} \, \mathrm{d}x = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}.
\end{aligned}$$

This clearly reduces to $1/n$ when all the λ_i are equal.

Exercise: Problem 6.5

Let X be a r.v. with pdf $f_X(x) = 2\theta^{-2} x \mathbb{I}_{(0,\theta)}(x)$, for $\theta > 0$.

a. Derive $\mathbb{E}[X]$ and the c.d.f. of X.
b. Let $X_1, \ldots, X_n$ be i.i.d. r.v.s each with p.d.f. f_X. Calculate f_M and $\mathbb{E}[M]$, where $M = \min X_i$.
c. Now repeat for $M = \max X_i$.

Exercise: Problem 6.5, Solution

a. The expected value is $2\theta^{-2}\int_0^\theta x^2\,\mathrm{d}x = 2\theta/3$ and the c.d.f. is

$$F_X(x) = 2\theta^{-2}\int_0^x t\,\mathrm{d}t = x^2\theta^{-2}\mathbb{I}_{(0,\theta)}(x).$$

b. Recall that $f_{Y_1}(y) = n\left[1 - F(y)\right]^{n-1} f(y)$.
From this and F_X, f_M for $M = \min X_i$ simplifies to

$$f_M(m) = 2n\theta^{-2n} m\left(\theta^2 - m^2\right)^{n-1}\mathbb{I}_{(0,\theta)}(m).$$

With $u = \theta^2 - m^2$, $m = \sqrt{\theta^2 - u}$ and
$\mathrm{d}m = -(1/2)\left(\theta^2 - u\right)^{-1/2}\mathrm{d}u$, f_M integrates to one because

$$2n\theta^{-2n}\int_0^\theta m\left(\theta^2 - m^2\right)^{n-1}\mathrm{d}m = n\theta^{-2n}\int_0^{\theta^2} u^{n-1}\,\mathrm{d}u = 1.$$

Exercise: Problem 6.5, Solution

b. (cont) For $\mathbb{E}[M]$, with $u = \theta^2 - m^2$, $m = \sqrt{\theta^2 - u}$ and $\mathrm{d}m = -(1/2)\left(\theta^2 - u\right)^{-1/2} \mathrm{d}u$,

$$\mathbb{E}[M] = 2n\theta^{-2n} \int_0^{\theta} m^2 \left(\theta^2 - m^2\right)^{n-1} \mathrm{d}m = n\theta^{-2n} \int_0^{\theta^2} \left(\theta^2 - u\right)^{1/2} u^{n-1} \mathrm{d}u.$$

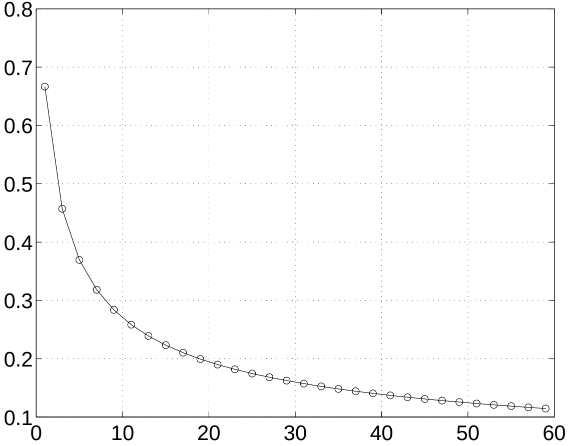
Next, with $w = \left(\theta^2 - u\right)/\theta^2$, $u = (1-w)\,\theta^2$, and $\mathrm{d}u = -\theta^2\,\mathrm{d}w$,

$$n\theta^{-2n} \int_0^{\theta^2} \left(\theta^2 - u\right)^{1/2} u^{n-1} \mathrm{d}u = n\theta \int_0^1 w^{1/2} (1-w)^{n-1} \mathrm{d}w$$
$$= n\theta B\left(\frac{3}{2}, n\right) = \frac{n\theta}{2} \sqrt{\pi} \frac{\Gamma(n)}{\Gamma(3/2+n)}.$$

With $\Gamma(3/2+1) = (3/2)\,\Gamma(3/2)$, for $n = 1$, $\mathbb{E}[M]$ reduces to $(2/3)\,\theta = \mathbb{E}[X]$. A plot of $\mathbb{E}[M]$ versus n for $\theta = 1$ is shown below.

Exercise: Problem 6.5, Solution

Expected minimum from n iid obs. with $f_X(x) = 2\theta^{-2}x\mathbb{I}_{(0,\theta)}(x)$.



Exercise: Problem 6.5, Solution

c. For $M = \max X_i$, recall $f_{Y_n}(y) = n\left[F(y)\right]^{n-1} f(y)$.
The density of f_M simplifies to $f_M(m) = 2n\theta^{-2n} m^{2n-1} \mathbb{I}_{(0,\theta)}(m)$,
which clearly integrates to one. Thus,

$$\mathbb{E}[M] = 2n\theta^{-2n} \int_0^\theta m^{2n} \, \mathrm{d}m = \frac{2n}{2n+1}\theta$$

and, for $n = 1$, $\mathbb{E}[M] = \mathbb{E}[X]$.

Joint Distribution of Several Order Statistics

Let's first consider the bivariate case, Y_i and Y_j with $i < j$. There are two cases to examine:

Case I: $x \geq y$.
The assumption that $i < j$ implies $Y_i \leq Y_j$. So, the three inequalities $y \leq x$, $Y_i \leq Y_j$ and $Y_j \leq y$ imply (just string them together) that

$$Y_i \leq Y_j \leq y \leq x,$$

implying that $Y_i \leq x$.

Thus, we obviously always have $\{Y_i \leq x,\ Y_j \leq y\} \subset \{Y_j \leq y\}$ and the above inequalities show that $\{Y_i \leq x,\ Y_j \leq y\} \supset \{Y_j \leq y\}$, i.e., the two events are equal, and

$$\Pr(Y_i \leq x, Y_j \leq y) = \Pr(Y_j \leq y)$$

and the c.d.f. for case I is $F_{Y_i,Y_j}(x,y) = F_{Y_j}(y)$.

Joint Distribution of Several Order Statistics

Case II: $x < y$. $F_{Y_i,Y_j}(x,y)$ is

$$\Pr\left(\text{at least } i \text{ of the } X\text{'s} \le x \cap \text{at least } j \text{ of the } X\text{'s} \le y\right)$$
$$= \sum_{a=j}^{n} \sum_{b=i}^{a} \Pr\left(\text{exactly } b \text{ of the } X\text{'s} \le x \cap \text{exactly } a \text{ of the } X\text{'s} \le y\right).$$

The inner sum stops at a because, for a fixed value of a from the outer sum, if exactly a of the X's $\le y$, then at most a can also be less than x (for $x < y$, as assumed).

Joint Distribution: Example and Exercise

(Example 6.9). Let $n = 2$, $i = 1$, $j = 2$, $V = \min(X_1, X_2)$ and $W = \max(X_1, X_2)$. For $v < w$, $F_{V,W}(v, w)$ simplifies to

$$2F(v)F(w) - [F(v)]^2,$$

which can also be written as

$$[F(w)]^2 - [F(w) - F(v)]^2. \tag{35}$$

Now consider Problem 6.9(a): Let X and Y be independent (but not necessarily identical) random variables and set $V = \min(X, Y)$ and $W = \max(X, Y)$. Derive the joint c.d.f. of (V, W).

Joint Distribution: Solution

Consider the value of the joint distribution function of V and W at all real values v and w. We distinguish the cases $v \leq w$ and $v \geq w$. First, let $v \leq w$: we then have $V \leq v$ and $W \leq w$ precisely if $X, Y \leq w$, but not both X and Y have their value in $(v, w]$. This yields

$$\begin{aligned} F_{V,W}(v,w) &= \Pr(X \leq w \wedge Y \leq w) - \Pr(v < X \leq w \wedge v \leq Y \leq w) \\ &= F_X(w)F_Y(w) - (F_X(w) - F_X(v))(F_Y(w) - F_Y(v)). \end{aligned}$$

This reduces precisely to (35) when X and Y are i.i.d..

Now if $v \geq w$, then $W \leq w$ implies $V \leq W \leq w \leq v$, so $V \leq v$ is automatically satisfied and

$$F_{V,W}(v,w) = \Pr(W \leq w) = \Pr(X \leq w \wedge Y \leq w) = F_X(w)F_Y(w). \tag{36}$$

We observe with satisfaction that both formulae coincide if $v = w$.

Joint Distribution: Solution

Observe that, for the case $v \leq w$,

$$\begin{aligned} F_V(v) &= \lim_{w\to\infty} F_{V,W}(v,w) && (37) \\ &= \lim_{w\to\infty} \quad F_X(w)F_Y(w) - (F_X(w) - F_X(v))(F_Y(w) - F_Y(v)) \\ &= 1 - (1 - F_X(v))(1 - F_Y(v)), && (38) \end{aligned}$$

and for $v \geq w$, either note that

$F_V(v) = \lim_{w\to\infty} F_{V,W}(v,w) = F_{V,W}(\infty,\infty) = 1$, or from (36), $F_V(v) = \lim_{w\to\infty} F_X(w)F_Y(w) = 1$.

Returning to the case with $X_i \overset{\text{indep}}{\sim} \mathrm{Geo}(p_i)$, $i = 1, 2$, $q_i = 1 - p_i$, and $M = \min(X_1, X_2)$, with $F_X(x) = \left(1 - q^{x+1}\right)\mathbb{I}_{\{0,1,\dots\}}(x)$, (38) implies

$$F_M(m) = 1 - (1 - F_{X_1}(m))(1 - F_{X_2}(m)) = 1 - \left(q_1^{m+1}\right)\left(q_2^{m+1}\right),$$

which is the same as the expression obtained earlier.

Joint Distribution of Several Order Statistics

The expression

$$\Pr\left(\text{at least } i \text{ of the } X\text{'s} \leq x \cap \text{at least } j \text{ of the } X\text{'s} \leq y\right)$$
$$= \sum_{a=j}^{n} \sum_{b=i}^{a} \Pr\left(\text{exactly } b \text{ of the } X\text{'s} \leq x \cap \text{exactly } a \text{ of the } X\text{'s} \leq y\right)$$

can be written as, using a multinomial argument like in the univariate case,

$$F_{Y_i, Y_j}(x, y) = \sum_{a=j}^{n} \sum_{b=i}^{a} \frac{n! \left[F(x)\right]^{b} \left[F(y) - F(x)\right]^{a-b} \left[1 - F(y)\right]^{n-a}}{b! \, (a-b)! \, (n-a)!}.$$

The extension of this to the general kth order case, i.e.,

$$F_{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}}, \quad 1 \leq i_1 < i_2 < \cdots < i_k \leq n,$$

follows along similar lines.

Joint Distribution of Several Order Statistics

For the joint density when F_X is discrete (continuous), appropriate differencing (differentiation) of the c.d.f. will yield the density.

For the continuous case, use of the multinomial method considered above proves to be easiest. In particular, for two order statistics Y_i and Y_j with $i < j$ and $x < y$, we obtain

$$\begin{aligned} f_{Y_i,Y_j}(x,y) = & \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!} \\ & \times [F(x)]^{i-1}\,[F(y)-F(x)]^{j-i-1}\,[1-F(y)]^{n-j} \\ & \times f(x)\,f(y)\,\mathbb{I}_{(x,\infty)}(y)\,. \end{aligned} \tag{39}$$

Joint Distribution of Several Order Statistics

Taking $i = 1$ and $j = 2$ in (39) gives the joint density of the first two order statistics as

$$f_{Y_1,Y_2}(x,y) = \frac{n!}{(n-2)!}\left[1 - F(y)\right]^{n-2} f(x)\, f(y)\, \mathbb{I}_{(x,\infty)}(y),$$

and generalizing this to the first k order statistics gives

$$\begin{aligned} f_{Y_1,\ldots,Y_k}(y_1,\ldots,y_k) &= \frac{n!}{(n-k)!}\left[1 - F(y_k)\right]^{n-k} \prod_{i=1}^{k} f(y_i) \\ &\quad \times \mathbb{I}\{y_1 < y_2 < \cdots < y_k\}. \end{aligned}$$

Observe that

$$\binom{n}{n-k, 1, \ldots, 1} = \frac{n!}{(n-k)!} = k!\binom{n}{n-k}.$$

Joint Distribution of Several Order Statistics

For the last two order statistics, with $i = n-1$ and $j = n$, (39) gives

$$f_{Y_{n-1},Y_n}(x,y) = \frac{n!}{(n-2)!}\left[F(x)\right]^{n-2} f(x)\, f(y)\, \mathbb{I}_{(x,\infty)}(y),$$

and, generalizing to the case of the last k order statistics, $1 \le k \le n$,

$$\begin{aligned} f_{Y_{n-k+1},\ldots,Y_n}(y_{n-k+1},\ldots,y_n) &= \frac{n!}{(n-k)!}\left[F(y_{n-k+1})\right]^{n-k} \prod_{i=1}^{k} f(y_i) \\ &\quad \times \mathbb{I}\{y_{n-k+1} < \cdots < y_n\}. \end{aligned}$$

Note that, for $k=1$, this reduces to the p.d.f. expression obtained earlier for the largest order statistic, i.e., $f_{Y_n}(y) = n\left[F(y)\right]^{n-1} f(y)$.
For $k = n$ (in either the p.d.f. expression for the first k, or last k, order statistics) we get the p.d.f. of the whole sample of order statistics,

$$f_{Y_1,\ldots,Y_n}(y_1,y_2,\ldots,y_n) = n! \prod_{i=1}^{n} f(y_i), \quad y_1 < y_2 < \cdots < y_n.$$

Joint Distribution of All Order Statistics

Again, the p.d.f. of the whole sample of order statistics is

$$f_{Y_1,\ldots,Y_n}\left(y_1, y_2, \ldots, y_n\right) = n! \prod_{i=1}^{n} f\left(y_i\right), \quad y_1 < y_2 < \cdots < y_n.$$

It is important to keep in mind that $\int \cdots \int f_{\mathbf{Y}}\left(\mathbf{y}\right) d\mathbf{y} = 1$, and not $n!$. This is because $\mathbf{Y}$ has support

$$\mathcal{S} := \{\mathbf{y} \in \mathbb{R}^n : y_1 < \cdots < y_n\} \in \mathbb{R}^n,$$

i.e.,

$$1 = \int \cdots \int_{\mathcal{S}} f_{\mathbf{Y}}\left(\mathbf{y}\right) \mathrm{d}\mathbf{y} = \int_{-\infty}^{\infty} \int_{y_1}^{\infty} \int_{y_2}^{\infty} \cdots \int_{y_{n-1}}^{\infty} f_{\mathbf{Y}}\left(\mathbf{y}\right) \, \mathrm{d}y_n \cdots \mathrm{d}y_1.$$

Joint Distribution of All Order Statistics

As an illustration, let $X_i \overset{\text{i.i.d.}}{\sim} \text{Exp}\,(\lambda)$, $i = 1, \ldots, n$. Then the joint density of the corresponding order statistics $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is, with $s = \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$, given by

$$f_{\mathbf{Y}}(\mathbf{y}) = n! \prod_{i=1}^{n} \lambda \mathrm{e}^{-\lambda y_i} = n! \lambda^n \mathrm{e}^{-\lambda s}, \quad 0 < y_1 < y_2 < \cdots < y_n.$$

To verify that $f_{\mathbf{Y}}$ indeed integrates to one, note that, for $n = 3$,

$$\begin{aligned}
\int \cdots \int f_{\mathbf{Y}}(\mathbf{y})\, \mathrm{d}\mathbf{y} &= 3! \int_0^\infty \lambda \mathrm{e}^{-\lambda y_1} \int_{y_1}^\infty \lambda \mathrm{e}^{-\lambda y_2} \int_{y_2}^\infty \lambda \mathrm{e}^{-\lambda y_3}\, \mathrm{d}y_3\, \mathrm{d}y_2\, \mathrm{d}y_1 \\
&= 3! \int_0^\infty \lambda \mathrm{e}^{-\lambda y_1} \int_{y_1}^\infty \lambda \mathrm{e}^{-\lambda y_2} \left(\mathrm{e}^{-\lambda y_2}\right) \mathrm{d}y_2\, \mathrm{d}y_1 \\
&= 3! \int_0^\infty \lambda \mathrm{e}^{-\lambda y_1} \left(\frac{1}{2} \mathrm{e}^{-2\lambda y_1}\right) \mathrm{d}y_1 = 3! \left(\frac{1}{2}\right) \left(\frac{1}{3}\right) = 1,
\end{aligned}$$

with the pattern for higher n being clear.

Sample Range and Midrange

- Let X_i, $i = 1, \ldots, n$, be an i.i.d. sample from a population with density f and c.d.f. F, and denote the order statistics as $Y_1, \ldots, Y_n$.
- The **sample range** is defined to be $R = Y_n - Y_1$, and provides a measure of the length of the support, or range, of the underlying distribution.
- The **sample midrange** is defined to be $T = (Y_1 + Y_n)/2$, and is a measure of central tendency.

Their joint distribution, $f_{R,T}(r, t)$, is given by

$$n(n-1)\left[F\left(t+\frac{r}{2}\right) - F\left(t-\frac{r}{2}\right)\right]^{n-2} f\left(t-\frac{r}{2}\right) f\left(t+\frac{r}{2}\right) \mathbb{I}_{(0,\infty)}(r),$$

as derived in Problem 6.4, and the marginals are computed in the usual way.

Sample Range and Midrange

- The mth raw moment of R can be expressed as

$$\mathbb{E}[R^m] = \int_0^\infty r^m f_R(r)\,\mathrm{d}r = \int_0^\infty \int_{-\infty}^\infty r^m f_{R,T}(r,t)\,\mathrm{d}t\,\mathrm{d}r$$

or, for $m = 1$,

$$\mathbb{E}[R] = \mathbb{E}[Y_n] - \mathbb{E}[Y_1] = \int_0^\infty y f_{Y_n}(y)\,\mathrm{d}y - \int_0^\infty y f_{Y_1}(y)\,\mathrm{d}y.$$

- A third way is as follows. Recall that, for continuous r.v. X,

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty (1 - F_X(x))\,\mathrm{d}x - \int_{-\infty}^0 F_X(x)\,\mathrm{d}x \\ &= \int_0^\infty \bar{F}_X(x)\,\mathrm{d}x - \int_{-\infty}^0 F_X(x)\,\mathrm{d}x.\end{aligned}$$

Sample Range and Midrange

Then, with

$$F_{Y_1}(y) = 1 - [1 - F(y)]^n =: 1 - \bar{F}^n(y) \quad \text{and} \quad F_{Y_n}(y) =: F^n(y),$$

it is straightforward to show (see the text) that

$$\mathbb{E}[R] = \int_{-\infty}^{\infty} \left[1 - F^n(y) - \bar{F}^n(y)\right] \mathrm{d}y$$

first reported in Tippett (1925) using other methods of proof (see the text).

Sample Range and Midrange

The last expression for $\mathbb{E}[R]$ can be generalized to

$$\mathbb{E}\left[R^m\right] = m! \int_{-\infty<y_1<\cdots<y_m<\infty} \cdots \int G(y_1, y_m)\, \mathrm{d}y_1 \cdots \mathrm{d}y_m, \qquad (40)$$

for $m = 1, 2, \ldots$, where

$$G(y_1, y_m) = \left[1 - F^n\left(y_m\right) - \bar{F}^n\left(y_1\right) + \left\{F\left(y_m\right) - F\left(y_1\right)\right\}^n\right].$$

This can further be generalized to the independent but not (necessarily) identical case, by replacing $G(y_1, y_m)$ in (40) by

$$1 - \prod_{i=1}^{n} F_i\left(y_m\right) - \prod_{i=1}^{n} \bar{F}_i\left(y_1\right) + \prod_{i=1}^{n} \left\{F_i\left(y_m\right) - F_i\left(y_1\right)\right\},$$

as shown by Jones and Balakrishnan (2002).

Example

Based on Kamara and Siegel (1987), in the context of optimal hedging strategies in futures markets. Let

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \right),$$

and define $T = Y - X$, which is normally distributed with mean $\mu_T = \mu_Y - \mu_X$ and variance $\sigma_T^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$.
As detailed in the text: With $t := \mu_T / \sigma_T$ and $M := \min(X, Y)$,

$$\mathbb{E}[M] = \mu_X - \sigma_T \phi(t) + \mu_T \Phi(-t),$$

$$\mathbb{V}(M) = \sigma_X^2 + (\sigma_Y^2 - \sigma_X^2)\Phi(-t) + \frac{\mu_T^2}{4} - \left[\frac{\mu_T}{2} - \mu_T \Phi(-t) + \sigma_T \phi(t) \right]^2,$$

and

$$\mathrm{Cov}(X, M) = \sigma_X^2 - (\sigma_X^2 - \sigma_{XY})\Phi(-t),$$

where Φ and ϕ are the standard normal c.d.f. and p.d.f., respectively.

Intermediate Probability: A Computational Approach

Warmup: Student's t

Recall that the gamma function is

$$\Gamma(a) := \int_0^\infty x^{a-1} e^{-x} \, \mathrm{d}x, \quad a \in \mathbb{R}_{>0},$$

with $\Gamma(a) = (a-1)\,\Gamma(a-1)$, $a \in \mathbb{R}_{>1}$, and, in particular, $\Gamma(n) = (n-1)!$, $n \in \mathbb{N}$.

The beta function is

$$B(a, b) := \int_0^1 x^{a-1} (1-x)^{b-1} \, \mathrm{d}x, \quad a, b \in \mathbb{R}_{>0}.$$

They are related by the identity

$$B(a, b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}.$$

Warmup: Student's t

The Student's t p.d.f. with n degrees of freedom, abbreviated $t\left(n\right)$ or t_n, $n \in \mathbb{R}_{>0}$, is given by

$$f_t\left(x; n\right) = \frac{\Gamma\left(\frac{n+1}{2}\right) n^{\frac{n}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(n + x^2\right)^{-\frac{n+1}{2}} = \frac{n^{-\frac{1}{2}}}{B\left(\frac{n}{2}, \frac{1}{2}\right)} \left(1 + x^2/n\right)^{-\frac{n+1}{2}}.$$

If $n = 1$, then the Student's t distribution reduces to the Cauchy distribution while, as $n \to \infty$, it converges in distribution to the normal.

Let $T \sim t_n$. The mean of T is zero for $n > 1$, but does not otherwise exist. For the variance,

$$\mathbb{V}(T) = \frac{n}{n-2}, \qquad \text{for } n > 2. \tag{41}$$

It is easy to see from the p.d.f. that $f_t(x; n) \propto |x|^{-(n+1)}$, which is similar to the type I Pareto, showing that the maximally existing moment of the Student's t is bounded above by n.

Warmup: Student's t

Recall that the incomplete beta function is

$$B_x(p,q) = \mathbb{I}_{[0,1]}(x) \int_0^x t^{p-1} (1-t)^{q-1} \, \mathrm{d}t$$

and the normalized function $B_x(p,q)/B(p,q)$ is the incomplete beta ratio, denoted by $\bar{B}_x(p,q)$.

For $t < 0$, the c.d.f. of the Student's t is given by

$$F_T(t) = \frac{1}{2} \bar{B}_L \left(\frac{n}{2}, \frac{1}{2} \right), \quad L = \frac{n}{n+t^2}, \quad t < 0. \tag{42}$$

For $t > 0$, the symmetry of the t density about zero implies that $F_T(t) = 1 - F_T(-t)$.

Warmup: Student's t

Let $T \sim t_n$.

1. Show that $\mathbb{E}[T] = 0$ if $n > 1$ and does not otherwise exist.
2. Calculate $\mathbb{E}\left[\,|T|^k\,\right]$.
3. Determine which moments exist.
4. Show (41).
5. Show (42).

Warmup: Student's t: Solutions to Exercise

Show that $\mathbb{E}[T] = 0$ if $n > 1$ and does not otherwise exist.
We want

$$\mathbb{E}[T] = \int_{-\infty}^{\infty} t f_T(t; n)\, \mathrm{d}t = K_n \int_{-\infty}^{\infty} t \left(1 + \frac{t^2}{n}\right)^{-k} \mathrm{d}t,$$

where $k = (n+1)/2$.
Let's ignore K_n for now, and just look at the integral. Split it at zero (because we have t^2 in the integrand) and let $u = t^2$. Then, for $t < 0$, the solution is $t = -\sqrt{u}$, with $\mathrm{d}t = -\frac{1}{2}u^{-1/2}\,\mathrm{d}u$. Also, when $t = -\infty$ (the lower bound in the integral) $u = \infty$, and when $t = 0$ (the upper bound), $u = 0$. So,

$$\begin{aligned}
\int_{-\infty}^{0} t \left(1 + \frac{t^2}{n}\right)^{-k} \mathrm{d}t &= \int_{+\infty}^{0} (-\sqrt{u}) \left(1 + \frac{u}{n}\right)^{-k} \left(-\frac{1}{2}u^{-1/2}\,\mathrm{d}u\right) \\
&= \frac{1}{2}\int_{\infty}^{0} \left(1 + \frac{u}{n}\right)^{-k} \mathrm{d}u \\
&= -\frac{1}{2}\int_{0}^{\infty} \left(1 + \frac{u}{n}\right)^{-k} \mathrm{d}u
\end{aligned}$$

Warmup: Student's t: Solutions to Exercise

Similarly, for $t > 0$, the solution is $t = +\sqrt{u}$, with $\mathrm{d}t = \frac{1}{2}u^{-1/2}\,\mathrm{d}u$, and the integral is

$$\begin{aligned}\int_0^\infty t\left(1+\frac{t^2}{n}\right)^{-k}\mathrm{d}t &= \int_0^\infty u^{1/2}\left(1+\frac{u}{n}\right)^{-k}\frac{1}{2}u^{-1/2}\,\mathrm{d}u\\ &= \frac{1}{2}\int_0^\infty\left(1+\frac{u}{n}\right)^{-k}\mathrm{d}u.\end{aligned}$$

Thus, adding the two pieces, we get that $\mathbb{E}[T] = 0$, **if the integral**

$$I = \int_0^\infty\left(1+\frac{u}{n}\right)^{-k}\mathrm{d}u$$

exists. Now, let $v = 1 + u/n$, so that $u = n(v-1)$ and $\mathrm{d}u = n\,\mathrm{d}v$, and

$$I = n\int_1^\infty v^{-k}\,\mathrm{d}v = \frac{n}{1-k}\,v^{1-k}\Big|_{v=1}^{\infty} = \begin{cases}\dfrac{n}{k-1}, & \text{if} \quad k > 1,\\ \infty, & \text{if} \quad k \le 1.\end{cases}$$

Thus, $\mathbb{E}[T] = 0$ if $k > 1$, which is the same as $k = (n+1)/2 > 1$, or $n > 1$.

Warmup: Student's t: Solutions to Exercise

Calculate $\mathbb{E}\left[\,|T|^k\,\right]$.

First observe that, because of symmetry, the density of $|T|$ is just

$$f_{|T|}(t) = 2 f_T(t; n)\, \mathbb{I}_{(0,\infty)}(t).$$

Let

$$u = \frac{t^2/n}{1+t^2/n}, \quad t = +\sqrt{n \frac{u}{1-u}}, \quad \mathrm{d}t = \frac{n^{1/2}}{2} u^{-1/2} (1-u)^{-3/2}\, \mathrm{d}u,$$

so that, after some simplifying,

$$\begin{aligned}
\mathbb{E}\left[\,|T|^k\,\right] &= 2 \frac{n^{-1/2}}{B\left(\frac{n}{2}, \frac{1}{2}\right)} \int_0^\infty t^k \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \mathrm{d}t \\
&= \frac{n^{k/2}}{B\left(\frac{n}{2}, \frac{1}{2}\right)} \int_0^1 u^{(k-1)/2} (1-u)^{(n-k-2)/2}\, \mathrm{d}u \\
&= \frac{n^{k/2}}{B\left(\frac{n}{2}, \frac{1}{2}\right)} B\left(\frac{k+1}{2}, \frac{n-k}{2}\right).
\end{aligned}$$

Warmup: Student's t: Solutions to Exercise

This simplifies further to

$$\mathbb{E}\left[\,|T|^k\,\right] = n^{k/2}\frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{n-k}{2}\right)}{\Gamma\left(\frac{k+1}{2}+\frac{n-k}{2}\right)}\frac{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)} = n^{k/2}\frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{n-k}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)}. \tag{43}$$

Warmup: Student's t: Solutions to Exercise

Determine which moments exist.
From the argument of the second gamma term in the numerator of (43), we see that $\mathbb{E}\left[\,|T|^k\,\right]$ exists only if $n > k$.

Show (41).
For $k = 2$, we have, for $n > 2$, that $\mathbb{E}\left[|T|^2\right] = \mathbb{E}\left[T^2\right]$, and $\mathbb{E}\left[T^2\right] = \mathbb{V}(T)$ because $\mathbb{E}[T] = 0$. Thus, from (43),

$$\mathbb{V}(T) = n\frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = n\frac{\frac{1}{2}\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\frac{\Gamma\left(\frac{n-2}{2}\right)}{\left(\frac{n}{2}-1\right)\Gamma\left(\frac{n}{2}-1\right)} = \frac{n}{2}\frac{1}{\frac{n}{2}-1} = \frac{n}{n-2}.$$

Warmup: Student's t: Solutions to Exercise

Show (42).

Use the substitution $u = 1 + x^2/n$ (so that $x = -n^{1/2}(u-1)^{1/2}$ and $\mathrm{d}x = -n^{1/2}(1/2)(u-1)^{-1/2}\,\mathrm{d}u$), followed by $v = (u-1)/u$ (so that $u = (1-v)^{-1}$ and $\mathrm{d}u = (1-v)^{-2}\,\mathrm{d}v$) to get

$$\begin{aligned} F_T(t) &= \frac{n^{-1/2}}{B(n/2,1/2)} \int_{-\infty}^{t} \left(1 + x^2/n\right)^{-(n+1)/2} \mathrm{d}x \\ &= \frac{1}{2B(n/2,1/2)} \int_{1+t^2/n}^{\infty} u^{-(n+1)/2} (u-1)^{-1/2}\,\mathrm{d}u \\ &= \frac{1}{2B(n/2,1/2)} \int_{1-L}^{1} v^{-1/2} (1-v)^{(n-2)/2}\,\mathrm{d}v, \end{aligned}$$

where

$$1 - L = \frac{t^2/n}{1 + t^2/n} = \frac{t^2}{n+t^2} = 1 - \frac{n}{n+t^2}.$$

Warmup: Student's t: Solutions to Exercise

Thus, as

$$\int_g^1 v^{a-1}(1-v)^{b-1}\,\mathrm{d}v = \int_0^{1-g} y^{b-1}(1-y)^{a-1}\,\mathrm{d}y,$$

we have

$$F_T(t) = \frac{1}{2B(n/2,1/2)}\int_0^L y^{(n-2)/2}(1-y)^{-1/2}\,\mathrm{d}y = \frac{1}{2}\bar{B}_L(n/2,1/2).$$

Overview

We wish to examine a selection of univariate distributions which extend, generalize and/or nest some of the more common, or basic ones.
This is organized as follows:

1. Rudimentary ways, such as
   - nesting,
   - generalizing constants
   - extensions to allow for asymmetry
   - extending positive r.v.s to the whole real line
   - transformations and "inventions"
2. Weighted sums of r.v.s
3. Mixture distributions (discrete and continuous)

Nesting and Generalizing Constants

Oftentimes, two or more families of distributions can be *nested* by constructing a probability density function with one or more parameters such that, when these assume specific values, or p.d.f. reduces to one of the special, nested cases.

Discrete Example

Recall the binomial and hypergeometric distributions. They have completely different sampling schemes and p.m.f.s, so it seems they cannot be nested. But:

- they can both be nested under the following sampling scheme:
- from an urn with w white and b black balls, a ball is drawn, and then replaced, *along with s balls of the same color*
- This is repeated n times.
- Let X be the number of black balls drawn.
- Taking $s = 0$ yields the binomial, while $s = -1$ gives the hypergeometric.
- The p.m.f. of the general case is straightforward to derive:

$$f_X(k) = \binom{n}{k} \frac{B\left(\frac{b}{s} + k, \frac{w}{s} + n - k\right)}{B\left(\frac{b}{s}, \frac{w}{s}\right)} \mathbb{I}_{\{0,1,\dots,n\}}(k),$$

and is referred to as the Pólya–Eggenberger distribution.

Discrete Example

To see that the Pólya–Eggenberger distribution sums to one, we need to show that

$$B(a_1, a_2) = \sum_{i=0}^{k} \binom{k}{i} B(a_1 + i, a_2 + k - i).$$

Let $X \sim \mathrm{Beta}(a_1, a_2)$. Then

$$\begin{aligned} 1 &= \int_0^1 f_X(x)\,\mathrm{d}x = \frac{1}{B(a_1, a_2)} \int_0^1 (x + 1 - x)^k x^{a_1 - 1} (1 - x)^{a_2 - 1}\,\mathrm{d}x \\ &= \frac{1}{B(a_1, a_2)} \sum_{i=0}^{k} \binom{k}{i} \int_0^1 x^{a_1 + i - 1} (1 - x)^{a_2 - 1 + k - i}\,\mathrm{d}x \\ &= \frac{1}{B(a_1, a_2)} \sum_{i=0}^{k} \binom{k}{i} B(a_1 + i, a_2 + k - i), \end{aligned}$$

using the binomial theorem.

Example: The Normal and Laplace Distributions

- Recall that the location zero, scale one kernels of the normal and Laplace distributions are given by $\exp\left\{-x^2/2\right\}$ and $\exp\left\{-|x|\right\}/2$, respectively.
- By replacing the fixed exponent values (2 and 1) with power p, $p \in \mathbb{R}_{>0}$, the two become nested.
- The resulting distribution is known as the *generalized exponential distribution*, or GED, with density

$$f_X(x;p) = \frac{p}{2\Gamma\left(p^{-1}\right)} \exp\left\{-|x|^p\right\}, \quad p > 0,$$

- The choice of name is unfortunate, but standard; it refers to the generalization of the exponent, *not* the exponential distribution!
- The GED is commonly used in applications where model residuals have excess kurtosis relative to the normal distribution (or fat-tails).

Further Examples

The text illustrates several other examples, which include:

- Generalized gamma distribution; it nests the gamma and Weibull (both of which are generalizations of the exponential distribution).
- The generalized Student's t, or GT (still symmetric, but replaces the exponent of 2 with a positive real number):

$$f_{\mathrm{GT}}(z; d, \nu) = K_{d,\nu} \left(1 + \frac{|z|^d}{\nu}\right)^{-(\nu+1/d)}, \quad d, \nu \in \mathbb{R}_{>0}, \qquad (44)$$

  with $K_{d,\nu}^{-1} = 2d^{-1}\nu^{1/d}\mathrm{B}\left(d^{-1}, \nu\right)$. When a scale parameter is introduced, the Student's t becomes a special case, or is nested by, the GT (take $d = 2$ and $\nu = n/2$). First proposed by McDonald and Newey (1988), and used in several papers since then.
- Type II Pareto to a generalized Pareto and type III Pareto.
- The chapter exercises contain numerous other examples.

Generalizing the Characteristic Function

The c.f. of a standard Laplace random variable X is

$$\varphi_X(t) = \mathbb{M}_X(it) = \frac{1}{1+t^2}.$$

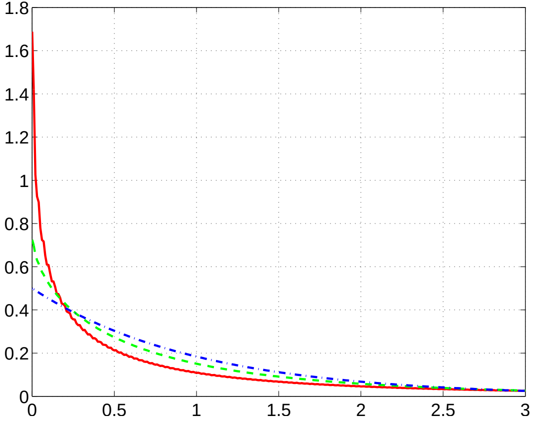
A natural generalization is to relax the quadratic exponent, giving $\varphi_Y(t) = (1 + |\,t|^\alpha)^{-1}$. It has been shown that this is a valid c.f. for $0 < \alpha \leq 2$, and that the corresponding p.d.f. is unimodal.

Recall that the p.d.f. must also be symmetric, because the c.f. is real.

Use of the p.d.f. inversion formula allows the density to be computed. The figure shows (the positive half of) the p.d.f. for several values of α.

As α decreases from two, the center becomes more peaked and the tails become more heavy.

Generalizing the Characteristic Function



The density on the half line corresponding to $\varphi_Y(t) = (1 + |t|^{\alpha})^{-1}$, $0 < \alpha \leq 2$, for $\alpha = 1$ (solid), $\alpha = 1.5$ (dashed) and $\alpha = 2$ (dash-dot). The latter corresponds to the Laplace.

Generalizing the Characteristic Function

- Example 1.21 showed the use of the density inversion formula applied to the characteristic function (c.f.) $\varphi_Z(t) = \mathrm{e}^{-t^2/2}$ to verify that $Z \sim \mathrm{N}(0,1)$.
- Similarly, Example 1.22 inverted the c.f. $\varphi_X(t) = \mathrm{e}^{-c|t|}$, for $c > 0$, to show that $X \sim \mathrm{Cau}(c)$.
- The two c.f.s are easily nested as (omitting the scale parameter) $\varphi(t;\alpha) = \exp\{-|t|^{\alpha}\}$.
- In this general case, however, the inversion integral applied to this c.f. cannot be simplified, and numerical techniques must be applied.
- It can be shown that this is a valid c.f. for $0 < \alpha \leq 2$, and is the c.f. of a theoretically and practically very important random variable, the (symmetric) stable Paretian. This topic, and its extension to the asymmetric stable Paretian, is covered in a later chapter.

Asymmetric Extensions

- Many stochastic phenomena exhibit an asymmetric probability structure, i.e., their densities are skewed.
- A variety of ways of introducing asymmetry into an existing density exist, several of which are now discussed.
- Note that some methods are more general than others.

The Skew Normal Distribution

- O'Hagan and Leonard (1976) and, independently and analyzed in more detail, Azzalini (1985), consider the asymmetric generalization of the normal distribution given by

$$f_{\mathrm{SN}}(z;\lambda) = 2\phi(z)\Phi(\lambda z), \qquad \lambda \in \mathbb{R},$$

- Adopting the notation in Azzalini (1985), SN stands for 'skew normal', and ϕ and Φ are the standard normal p.d.f. and c.d.f., respectively.
- Clearly, for $\lambda = 0$, f_{SN} reduces to the standard normal density, while for $\lambda \neq 0$, it is skewed.
- We wish to prove that is a proper density function.

The SN Distribution, proof

We first prove a simple and intuitive result: Let X and Y be independent, continuous r.v.s with p.d.f.s symmetric about zero.

Then (irrespective of the symmetry about zero), from the continuous extension of the basic law of total probability, $\Pr(X \leq Y)$ can be computed by "summing" $\Pr(X \leq y)$ weighted by the p.d.f. of Y:

$$P := \Pr(X \leq Y) = \int_{-\infty}^{\infty} F_X(y)\, f_Y(y)\, \mathrm{d}y. \tag{45}$$

From the symmetry constraint, it follows that, for any $z \in \mathbb{R}$,

$$F_X(-z) = 1 - F_X(z) \quad \text{and} \quad f_Y(y) = f_Y(-y).$$

Setting $z = -y$ yields the following, showing that $P = 1/2$:

$$P = -\int_{\infty}^{-\infty} F_X(-z)\, f_Y(-z)\, \mathrm{d}z = \int_{-\infty}^{\infty} (1 - F_X(z))\, f_Y(z)\, \mathrm{d}z = 1 - P.$$

The SN Distribution, proof

Now express $\Pr(X < \lambda Y)$ also as the sum of $\Pr(X < \lambda Y)$ weighted by the p.d.f. of Y:

$$\Pr(X < \lambda Y) = \int_{-\infty}^{\infty} F_X(\lambda y) f_Y(y)\, \mathrm{d}y.$$

As a check, this can also be seen as follows: Let $Z = \lambda Y$, implying that $f_Z(z) = (1/\lambda) f_Y(z/\lambda)$, so that, from (45),

$$\begin{aligned}\Pr(X < \lambda Y) = \Pr(X \leq Z) &= \int F_X(h)\, f_Z(h)\, \mathrm{d}h \\ &= \int F_X(h) \frac{1}{\lambda} f_Y\left(\frac{h}{\lambda}\right) \mathrm{d}h = \int F_X(\lambda y)\, f_Y(y)\, \mathrm{d}y,\end{aligned}$$

where in the last equality, we used the substitution $y = h/\lambda$ (and $h = \lambda y$, $\mathrm{d}h = \lambda\, \mathrm{d}y$).

The SN Distribution, proof

Finally, as the density of $Z = \lambda Y$ is also symmetric about zero,

$$\frac{1}{2} = \Pr(X < Z) = \int_{-\infty}^{\infty} F_X(\lambda y) f_Y(y) \, \mathrm{d}y,$$

and, from the symmetry of the standard normal distribution, the integral over $\mathbb{R}$ of $f_{\mathrm{SN}}(z; \lambda) = 2\phi(z)\Phi(\lambda z)$ is unity.

The SN Distribution: Relation to χ^2

An appealing property of the SN distribution not shared by other, more *ad hoc* methods for introducing skewness into a normal density is that, if $X \sim \mathrm{SN}(\lambda)$, then $X^2 \sim \chi^2_1$.

This is proven in Azzalini (1985), though follows immediately from the result, as noted in Gupta et al. (2004),[3] due to Roberts and S. Geisser (1966):[4]

$W^2 \sim \chi^2_1$ if and only if the p.d.f. of W has the form $f(w) = h(w)\exp(-w^2/2)$, where $h(w) + h(-w) = \sqrt{2/\pi}$.

[3] Arjun K. Gupta, Truc T. Nguyen and Jose Almer T. Sanqui (2004), *Characterization of the Skew-Normal Distribution*, Ann. Inst. Statist. Math. Vol. 56, No. 2, 351–360 .

[4] C. Roberts and S. Geisser (1966), *A Necessary and Sufficient Condition for the Square of a Random Variable to be Gamma*, Biometrika, 53, 275–277.

The SN Distribution: Available Software

Code is available for Matlab, R, and Splus for the p.d.f., c.d.f., quantiles, random number generators, and other things for the skew normal, as well as for multivariate extensions, extensions to the skewed Student's t, etc.

Follow the links from the Azzalini web page on the skew normal distribution.

Fernández and Steel (1998)

- Fernández and Steel (1998) proposed and investigated a simple and *ad hoc*, but very effective, method of introducing asymmetry into a symmetric, continuous density:

$$f(z;\theta) = \frac{2}{\theta + 1/\theta} \left\{ f\left(\frac{z}{\theta}\right) \mathbb{I}_{[0,\infty)}(z) + f(z\theta)\, \mathbb{I}_{(-\infty,0)}(z) \right\},$$

  where $f(z) = f(|z|)$ and $\theta > 0$.

- For $\theta = 1$, the density is symmetric, while for $0 < \theta < 1$, the distribution is skewed to the left and otherwise to the right.

Fernández and Steel (1998)

- This can be applied to the Student's t distribution, resulting in, say, the Fernández Steel t distribution $f_{\mathrm{FS-t}}(z;\nu,\theta)$. This was applied in Fernández and Steel (1998), and earlier in Hansen (1994).
- Another popular case is application to the GED, giving

$$f_{\mathrm{FS-GED}}(z;d,\theta)=K_{d,\theta}\begin{cases}\exp\left(-\left(-\theta z\right)^{d}\right), & \text{if} \quad z<0,\\ \exp\left(-\left(z/\theta\right)^{d}\right), & \text{if} \quad z\geq 0,\end{cases} \tag{46}$$

$\theta, d \in \mathbb{R}_{>0}$, and where $K_{d,\theta}^{-1}=\left(\theta+\theta^{-1}\right)d^{-1}\Gamma\left(d^{-1}\right)$.
This distribution is applied in Fernández, Osiewalski and Steel (1995) and similar constructions are examined in great detail in Ayebo and Kozubowski (2003) and Komunjer (2006).

Generalized Asymmetric t

The GT in (44) and the $f_{\mathrm{FS-GED}}$ in (46) are easily nested to yield the generalized asymmetric t (GAt), with p.d.f.,

$$f_{\mathrm{GAt}}(z; d, \nu, \theta) = K \times \begin{cases} \left(1 + \dfrac{(-z \cdot \theta)^d}{\nu}\right)^{-\left(\nu + \frac{1}{d}\right)}, & \text{if } z < 0, \\ \left(1 + \dfrac{(z/\theta)^d}{\nu}\right)^{-\left(\nu + \frac{1}{d}\right)}, & \text{if } z \geq 0, \end{cases}$$

$d, \nu, \theta \in \mathbb{R}_{>0}$.

It nests many distributions used for modeling financial returns, such as the normal, Laplace, GED, Student's t, GT, Cauchy, and offers an asymmetric version of each.

Problem 7.7 discusses various properties of this distribution.

Further Asymmetric Student's t Distributions

Based on a class of distributions proposed by Lye and Martin (1993), a variant is

$$f_{\mathrm{LyMd}}(z; d, \nu, \theta) = K \exp\left(\theta \arctan\left(\frac{z}{\nu^{1/d}}\right) - \left(\nu + \frac{1}{d}\right) \log\left(1 + \frac{|z|^d}{\nu}\right)\right),$$

$d \in \mathbb{R}_{>0}$, which is asymmetric for $\theta \neq 0$, and nests the GT density.

Jones and Faddy (2003) suggests an asymmetric generalization of Student's t, with density

$$f_{\mathrm{JoF}}(t; a, b) = C \left(1 + \frac{t}{(a+b+t^2)^{1/2}}\right)^{a+1/2} \left(1 - \frac{t}{(a+b+t^2)^{1/2}}\right)^{b+1/2},$$

where $a, b \in \mathbb{R}_{>0}$ and $C^{-1} = B(a,b)(a+b)^{1/2} 2^{a+b-1}$. If $a < b$ ($a > b$), then S is negatively (positively) skewed, while $S \sim t(2a)$ if $a = b$. Problem 7.8 discusses various properties of this distribution.

Further Asymmetric Student's t Distributions

Hyperbolic Asymmetric t, as discussed in Chapter 9,

$$f_{\mathrm{HA}t}(x; n, \beta, \mu, \delta) = \frac{2^{\frac{-n+1}{2}} \delta^n}{\sqrt{\pi}\Gamma(n/2)} \left(\frac{y_x}{|\beta|} \right)^{-\frac{n+1}{2}} K_{-\frac{n+1}{2}} \left(|\beta| \, y_x \right) \mathrm{e}^{\beta(x-\mu)},$$

where $y_x = \sqrt{\delta^2 + (x-\mu)^2}$, for $n > 0$, $\beta, \mu \in \mathbb{R}$, $\beta \neq 0$, $\delta > 0$. Function K is the modified Bessel function of the third kind.

Let $X \sim \mathrm{N}(\mu, 1)$ independent of $Y \sim \chi^2(k, \theta)$. Then $T = X/\sqrt{Y/k} \sim t''(k, \mu, \theta)$ is a doubly noncentral t distribution with k degrees of freedom, numerator noncentrality parameter μ and denominator noncentrality parameter θ.

If $\theta = 0$, then T is singly noncentral t with noncentrality parameter μ, and we write $T \sim t'(k, \mu)$.

Noncentral Student's t Distributions (Chapter 10)

For $T \sim t'(k, \mu)$,

$$f_T(t; k, \mu) = e^{-\mu^2/2} \frac{\Gamma((k+1)/2) k^{k/2}}{\sqrt{\pi}\Gamma(k/2)} \left(\frac{1}{k+t^2}\right)^{\frac{k+1}{2}} \times \left(\sum_{i=0}^{\infty} \frac{(t\mu)^i}{i!} \left(\frac{2}{t^2+k}\right)^{i/2} \frac{\Gamma((k+i+1)/2)}{\Gamma((k+1)/2)} \right).$$

For $T \sim t''(k, \mu, \theta)$,

$$f_T(t; k, \mu, \theta) = \frac{e^{-(\theta+\mu^2)/2}}{\sqrt{\pi k}} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(\theta/2)^i}{i!j!} \frac{\Gamma((k+2i+j+1)/2)}{\Gamma(i+k/2)} \times \left(t\mu\sqrt{2/k}\right)^j \left(1+t^2/k\right)^{-(k+2i+j+1)/2}.$$

Extension to the Real Line

- Let X be a continuous r.v. with p.d.f. f_X and support on $\mathbb{R}_{>0}$ and define Y to be a r.v. with p.d.f.

$$f_Y(y) = \frac{1}{2} f_X(|y|) .$$

- Then the distribution of Y is obtained by reflecting the p.d.f. of X onto $(-\infty, 0)$ and rescaling.
- The Laplace is probably the most prominent example, whereby $X \sim \mathrm{Exp}(\lambda)$ with $f_X(x) = \lambda \mathrm{e}^{-\lambda x}$ and $Y \sim \mathrm{Lap}(0, \sigma)$ with $\sigma = 1/\lambda$ and $f_Y(y) = \exp\{-|y|/\sigma\}/(2\sigma)$.

Example: the Double Weibull

- If Y is a r.v. which follows the double Weibull distribution with shape $\beta > 0$ and scale $\sigma > 0$, we write $Y \sim \mathrm{DWeib}(\beta, \sigma)$, where

$$f_{\mathrm{DWeib}}(y; \beta, \sigma) = f_Y(y; \beta, \sigma) = \frac{\beta}{2\sigma} \left| \frac{y}{\sigma} \right|^{\beta - 1} \exp\left(- \left| \frac{y}{\sigma} \right|^{\beta} \right).$$

- When $\beta = 1$, Y reduces to a Laplace random variable.
- From the symmetry of the density, the mean of Y is clearly zero, though a location parameter μ could also be introduced in the usual way.

Asymmetric Double Weibull

One simple method of introducing asymmetry into the double Weibull is to allow for two different shape parameters, depending on the sign of z:

We define the *Asymmetric Double Weibull* density as

$$f_{\mathrm{ADWeib}}\left(z;\beta^{-},\beta^{+},\sigma\right)=\begin{cases} f_{\mathrm{DWeib}}\left(z;\beta^{-},\sigma\right), & \text{if} \quad z<0, \\ f_{\mathrm{DWeib}}\left(z;\beta^{+},\sigma\right), & \text{if} \quad z\geq 0, \end{cases}$$

where β^{-} and β^{+} denote the shape parameters on the left ($z<0$) and right support ($z\geq 0$), respectively.

Let $Z\sim\mathrm{ADWeib}\left(\beta^{-},\beta^{+},1\right)$.

Exercise: Show that the c.d.f. is $F_Z(x)=\exp\left(-(-x)^{\beta^{-}}\right)/2$ for $x\leq 0$ and $F_Z(x)=1-\exp\left(-x^{\beta^{+}}\right)/2$ for $x>0$. The expected shortfall of Z is also easily derived: For $x<0$ and $b=\beta^{-}$, show that

$$\mathbb{E}\left[Z\mid Z\leq x\right]=\frac{-1}{2F_Z(x)}\left[\Gamma\left(1+1/b\right)-\Gamma_{(-x)^b}\left(1+1/b\right)\right].$$

Asymmetric Double Weibull: Solution

For the c.d.f., for $x < 0$, substitute $r = (-z)^b$ to get

$$z = -\left(r^{1/b}\right), \quad \mathrm{d}z = -\frac{1}{b} r^{\frac{1}{b}-1} \mathrm{d}r$$

and

$$\begin{aligned}
F_Z(x) &= \int_{-\infty}^{x} f_Z\left(z; b, \beta^+, 1\right) \mathrm{d}z = \int_{-\infty}^{x} \frac{b}{2} (-z)^{b-1} \exp\left(-(-z)^b\right) \mathrm{d}z \\
&= \int_{\infty}^{(-x)^b} \frac{b}{2} r^{(b-1)/b} \exp(-r) \left(-\frac{1}{b} r^{\frac{1}{b}-1} \mathrm{d}r\right) = \frac{1}{2} \int_{(-x)^b}^{\infty} \exp(-r)\, \mathrm{d}r \\
&= \frac{1}{2} \exp\left(-(-x)^b\right),
\end{aligned}$$

and $F_Z(0) = 1/2$.

Asymmetric Double Weibull: Solution

For $x \geq 0$, similar to the previous calculation with $r = z^b$,

$$\begin{aligned} F_Z(x) &= \frac{1}{2} + \int_0^x f_Z\left(z; \beta^-, b, 1\right) \mathrm{d}z = \frac{1}{2} + \frac{1}{2}\int_0^x bz^{b-1} \exp\left(-z^b\right) \mathrm{d}z \\ &= \frac{1}{2} + \frac{1}{2}\int_0^{x^b} \exp\left(-r\right) \mathrm{d}r \\ &= \frac{1}{2} + \frac{1}{2}\left(1 - \exp\left(-x^b\right)\right) \\ &= 1 - \frac{1}{2}\exp\left(-x^b\right). \end{aligned}$$

Asymmetric Double Weibull: Solution

For the ES, use the incomplete gamma function $\Gamma_v(a) = \int_0^v x^{a-1} e^{-x} \mathrm{d}x$ and, for $x < 0$, with $b = \beta^-$, and substitute $r = (-z)^b$ to get

$$\begin{aligned}
\int_{-\infty}^{x} z f_Z(z)\, \mathrm{d}z &= \int_{-\infty}^{x} z\, \frac{b}{2} (-z)^{b-1} \exp\left(-(-z)^b\right) \mathrm{d}z \\
&= -\frac{1}{2} \int_{(-x)^b}^{\infty} r^{1/b} \exp(-r)\, \mathrm{d}r \\
&= -\frac{1}{2} \left(\Gamma\left(1 + \frac{1}{b}\right) - \Gamma_{(-x)^b}\left(1 + \frac{1}{b}\right) \right).
\end{aligned}$$

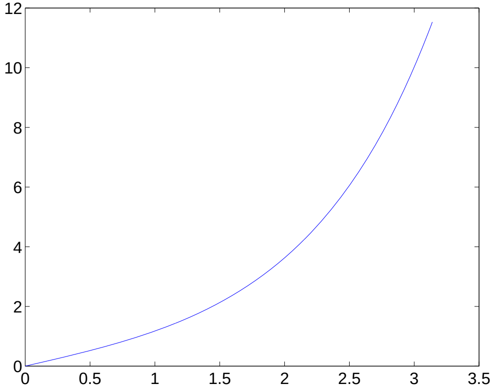
Nonlinear Transformations: IHS

- The *inverse hyperbolic sine*, or IHS, distribution, was introduced by Johnson (1949), and is also referred to as the S_U distribution.
- It is a flexible, asymmetric and leptokurtic distribution whose inverse c.d.f. is straightforwardly calculated, and has found recent use in empirical finance studies.
- We need the following: The functions hyperbolic cosine (cosh) and hyperbolic sine (sinh), are defined by

$$\cosh(\theta) = \frac{e^{\theta} + e^{-\theta}}{2} \quad \text{and} \quad \sinh(\theta) = \frac{e^{\theta} - e^{-\theta}}{2}.$$

Nonlinear Transformations: IHS

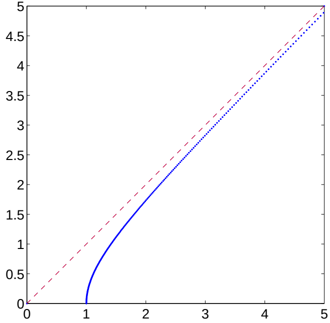
Plot of θ and $\sinh(\theta)$



Nonlinear Transformations: IHS

A simple calculation reveals that $\cosh^2(\theta) - \sinh^2(\theta) = 1$.
Recall that the circle $x^2 + y^2 = 1$ can be defined by the pair of functions $x = \cos(\theta)$ and $y = \sin(\theta)$. Similarly, the hyperbola $x^2 - y^2 = 1$ can be defined by the pair of functions $x = \cosh(\theta)$ and $y = \sinh(\theta)$.

Plot of $x = \cosh(\theta)$ and $y = \sinh(\theta)$ for set of values of θ starting at zero



IHS

- Let $\sinh^{-1}$ denote the inverse function of sinh, so that, if $y = \sinh(x)$, then $x = \sinh^{-1}(y)$.
- Function $\sinh^{-1}$ takes on the simple form $\sinh^{-1}(y) = \ln\left(y + \sqrt{1+y^2}\right)$, which is easily verified by simplifying $y = \sinh\left(\sinh^{-1}(y)\right)$.
- It is also easily derived: Let $y = \sinh(x)$ and note that $\sinh(x) + \cosh(x) = e^x$ or

$$e^x = y + \cosh(x) = y + \sqrt{1 + \sinh^2(x)} = y + \sqrt{1+y^2}.$$

- Similarly, $\cosh^{-1}(y) = \ln\left(y + \sqrt{y^2-1}\right)$.

IHS

Random variable $Y \sim \mathrm{IHS}(\lambda, \theta)$ if $\sinh^{-1}(Y) \sim \mathrm{N}(\lambda, \theta^2)$ for $\theta > 0$, or, with $Z \sim \mathrm{N}(0,1)$, $Y = \sinh(\lambda + \theta Z)$. As $\sinh^{-1}$ is a nondecreasing function, the p.d.f. is obtained by transformation:

$$f_Y(y) = f_Z(z) \left| \frac{dz}{dy} \right|, \quad z = \frac{\sinh^{-1}(y) - \lambda}{\theta},$$

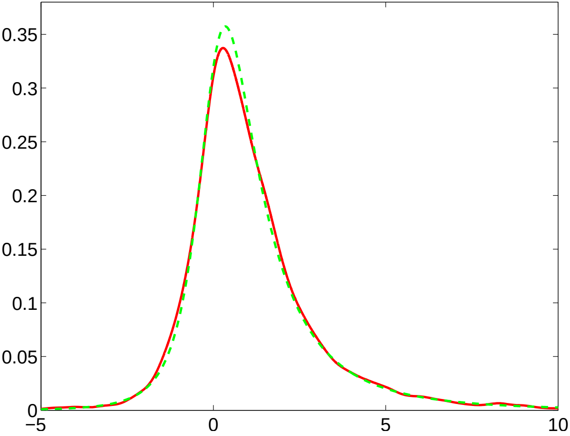
$$\frac{dz}{dy} = \frac{1}{\theta} \frac{1}{y + \sqrt{1+y^2}} \left(1 + \frac{1}{2} \left(1+y^2\right)^{-1/2} 2y \right) = \frac{1}{\theta\sqrt{1+y^2}},$$

giving

$$f_Y(y; \lambda, \theta) = \frac{1}{\theta\sqrt{2\pi\left(1+y^2\right)}} \exp\left\{ -\frac{1}{2} \left(\frac{\sinh^{-1}(y) - \lambda}{\theta} \right)^2 \right\}.$$

IHS

Kernel density (solid) based on 5,000 simulated IHS r.v.s, for $\lambda = 2/3$ and $\theta = 1$, computed via relation $\sinh^{-1}(Y) \sim \mathrm{N}(\lambda, \theta^2)$ and the p.d.f. expression $f_Y(y; \lambda, \theta)$ given above (dashed).



IHS

The c.d.f. is most easily derived by just writing

$$\begin{aligned} F_Y(y) &= \Pr\left(\sinh(\lambda + \theta Z) \le y\right) = \Pr\left(Z \le \frac{\sinh^{-1}(y) - \lambda}{\theta}\right) \\ &= \Phi\left(\frac{\sinh^{-1}(y) - \lambda}{\theta}\right). \end{aligned}$$

Another way (the one in the textbook) is shown on the next slide.

IHS

Note that $F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t$, so letting

$$w = \frac{\sinh^{-1}(t) - \lambda}{\theta} = \frac{\ln\left(t + \sqrt{1+t^2}\right) - \lambda}{\theta}, \quad \mathrm{d}w = \frac{1}{\theta\sqrt{1+t^2}}\,\mathrm{d}t\ ,$$

as calculated above, also gives $F_Y(y)$:

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^{\frac{1}{\theta}\left(\sinh^{-1}(y)-\lambda\right)} \frac{1}{\theta\sqrt{2\pi\left(1+t^2\right)}} \exp\left\{-\frac{1}{2}w^2\right\} \theta\sqrt{1+t^2}\,\mathrm{d}w \\ &= \int_{-\infty}^{\frac{1}{\theta}\left(\ln\left(y+\sqrt{1+y^2}\right)-\lambda\right)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}w^2\right\} \mathrm{d}w \\ &= \Phi\left(\frac{1}{\theta}\left(\sinh^{-1}(y) - \lambda\right)\right). \end{aligned}$$

IHS

- The median of Y solves $0.5 = \Phi\left(\frac{1}{\theta}\left(\sinh^{-1}(y) - \lambda\right)\right)$ or $y = \sinh(\lambda)$.
- For the inverse c.d.f., let $y_q = F_Y^{-1}(q)$ be the qth quantile of Y, $0 < q < 1$, so that

$$q = F_Y(y_q) = \Phi\left(\left(\sinh^{-1}(y_q) - \lambda\right)/\theta\right)$$

  or, solving, $y_q = \sinh\left(\lambda + \theta\Phi^{-1}(q)\right)$.
- Expressions for the first four moments are given in the text. These include the 3rd moment, which is never zero for $\theta > 0$, so that the distribution has some constraints with respect to its range of skewness and kurtosis.

Countable Mixtures

- There are many situations in which the random variable of interest, X, is actually the realization of one of k random variables, say $X_1, \ldots, X_k$, but from which one it came is unknown.
- The resulting r.v. is said to follow a *finite mixture distribution*.
- Denote the p.m.f. or p.d.f. of X_i as $f_{X|i}$, $i = 1, \ldots, k$, to make the dependence on i explicit.
- The p.m.f. or p.d.f. of X is given by

$$f_X(x) = \sum_{i=1}^{k} \lambda_i f_{X|i}(x), \quad \lambda_i \in (0,1), \quad \sum_{i=1}^{k} \lambda_i = 1.$$

- The λ_i are referred to as the *mixture component weights*.

Mixed Normal

- A very popular model is the k-component mixed normal. It obviously generalizes the Gaussian model and usually a small value of k (2 or 3) is enough to capture the deviations from normality in the data such as skewness and excess kurtosis.
- A r.v. X with this distribution is designated as $X \sim \mathrm{MixN}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$. With ϕ denoting the normal p.d.f., the density is

$$f_{\mathrm{MixN}}\left(x; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}\right) = \sum_{i=1}^{k} \lambda_i \phi\left(x; \mu_i, \sigma_i^2\right), \quad \lambda_i \in (0,1), \quad \sum_{i=1}^{k} \lambda_i = 1.$$

- It is crucial to realize that X is **not** the weighted sum of two normal random variables (and thus itself normally distributed) but rather a random variable whose p.d.f. is a weighted sum of two p.d.f.s.

Mixed Normal Density

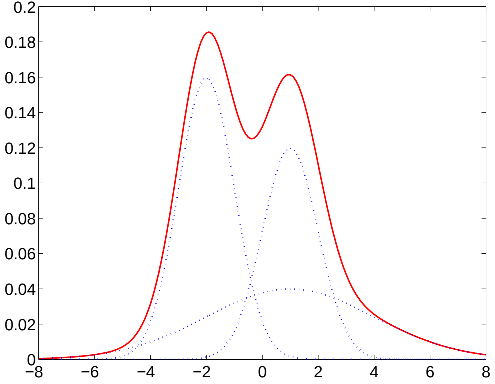


Figure: Mixed normal density with $\boldsymbol{\mu} = (-2, 1, 1)$, $\boldsymbol{\sigma} = (1, 1, 3)$ and $\boldsymbol{\lambda} = (0.4, 0.3, 0.3)$ (solid) and the individual weighted components (dotted)

Finite Mixtures

- Introduce a discrete random variable, C, with support $\mathcal{C} = \{1, \ldots, k\}$, and p.m.f. $f_C(c) = \Pr(C = c) = \lambda_c$, $c \in \mathcal{C}$, $\lambda_c \in (0,1)$, $\sum_{c=1}^{k} \lambda_c = 1$.
- The realization of C indicates the component p.m.f. or p.d.f. from which X is to be drawn.
- From the total probability formula, the p.m.f. or p.d.f. of X is,

$$f_X(x) = \int_{\mathcal{C}} f_{X|C}(x \mid c)\, \mathrm{d}F_C(c) = \sum_{c=1}^{k} f_{X|C}(x \mid c)\, f_C(c) = \sum_{c=1}^{k} \lambda_c f_{X|C}(x \mid c).$$

  (Later, we let C be continuous, which is why we use here the general notation $\mathrm{d}F_C$.)
- Observe that X is a marginal distribution computed from the joint distribution of X and C, $f_{X,C} = f_{X|C} f_C$.

Finite Mixtures: C.D.F.

The c.d.f. of X is

$$F_X(x) = \int_{\mathcal{C}} F_{X|C}(x \mid c)\, f_C(c)\, \mathrm{d}c = \sum_{c=1}^{k} \lambda_c F_{X|C}(x \mid c).$$

For $X \sim \mathrm{MixN}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$, with $f_{\mathrm{MixN}}(x; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) = \sum_{c=1}^{k} \lambda_c \phi\left(x; \mu_c, \sigma_c^2\right)$, the c.d.f. is

$$F_X(x; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) = \sum_{c=1}^{k} \lambda_c \Phi\left(x; \mu_c, \sigma_c^2\right) = \sum_{c=1}^{k} \lambda_c \Phi\left(\frac{x - \mu_c}{\sigma_c}, 0, 1\right)$$

where $\Phi(x; \mu, \sigma)$ denotes the c.d.f. of a normal r.v. with mean μ and variance σ^2.

Finite Mixtures: Generation

A realization from a finite mixture can be generated by taking a realization of X_c with probability λ_c, $c = 1, \ldots, k$. This is true, because when sampled this way,

$$\begin{aligned}
\Pr(X \in (x + \triangle x)) &= \sum_{c=1}^{k} \Pr(X \in (x + \triangle x) \mid C = c) \Pr(C = c) \\
&= \sum_{c=1}^{k} \lambda_c \Pr(X_c \in (x + \triangle x)) \\
&\approx \sum_{c=1}^{k} \lambda_c f_{X|c}(x) \triangle x = f_X(x) \triangle x.
\end{aligned}$$

Finite Mixtures: Generation

We first need the value of C. This is a draw from a multinomial distribution with vector of probabilities $p = (\lambda_1, \ldots, \lambda_k)$ and the "n" parameter as one:

```
function C=randmultinomial(p)
pp=cumsum(p); u=rand; C=1+sum(u>pp);
```

This is fast, and can be repeated N times.

In Matlab, R, etc., it is far more efficient because of the "vectorized functions" to generate the set of k vectors of X_i, $i = 1, \ldots, k$. In particular, let $\mathbf{X}$ be the $N \times k$ matrix with columns consisting of vector realizations of X_1, X_2, up to X_k.

So, we should elementwise (or entrywise) multiply (Hadamard product) the matrix `pick` generated as:

```
N=100; k=3; pick=zeros(N,k); lambda=[0.4,0.3,0.3];
for i=1:N, c=randmultinomial(lambda); pick(i,c)=1; end
```

with the $\mathbf{X}$ matrix. For example, using the values in the previous mixed normal plot, $\boldsymbol{\mu} = (-2, 1, 1)$, $\boldsymbol{\sigma} = (1, 1, 3)$ and $\boldsymbol{\lambda} = (0.4, 0.3, 0.3)$, we would generate $\mathbf{X}$ as

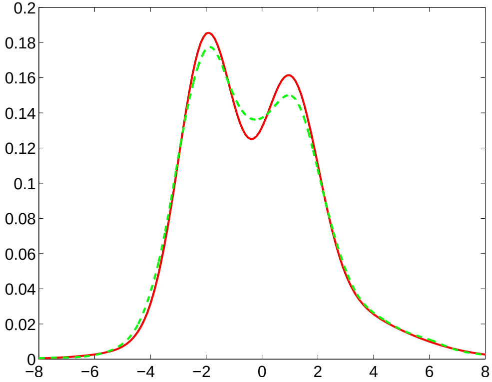
```
k=3; mu=[-2,1,1]; sigma=[1,1,3]; X=[];
for c=1:k
  X=[X mu(c)+sigma(c)*randn(N,1)];
end
```

and finally compute the realized normal mixture vector as

```
X=pick.*X; X=sum(X')';
% in matlab, sum of a matrix sums columnwise
```

Finite Mixtures: Generation

Running the above code with $N = 10,000$ yields a kernel density estimate as shown as the dashed line, with true as solid.



Expected Shortfall: Introduction

- An important measure in financial risk management is the *expected shortfall*. Before defining it, we state a general result:
- Let the p.d.f. and c.d.f. of r.v. R be f_R and F_R. Then the expected value of measurable function $g(R)$, given that $R \leq c$, is

$$\mathbb{E}[g(R) \mid R \leq c] = \frac{\int_{-\infty}^{c} g(r) f_R(r) \, \mathrm{d}r}{F_R(c)}.$$

- **Exercise I** Show that, for $R \sim \mathrm{N}(0,1)$ with p.d.f. ϕ and c.d.f. Φ, and a fixed $c < 0$, $\mathbb{E}[R \mid R \leq c] = -\phi(c)/\Phi(c)$.

Expected Shortfall: Introduction

Solution to Exercise I

Let $u = -r^2/2$. Then

$$\begin{aligned}
\mathbb{E}\left[R \mid R \leq c\right] &= \frac{1}{\Phi(c)} \int_{-\infty}^{c} r\phi(r)\,\mathrm{d}r \\
&= \frac{1}{\Phi(c)} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} r\exp\left\{-\frac{1}{2}r^2\right\}\mathrm{d}r \\
&= \frac{1}{\Phi(c)} \frac{1}{\sqrt{2\pi}} \left(-\exp\left\{-\frac{1}{2}c^2\right\}\right) = -\frac{\phi(c)}{\Phi(c)}.
\end{aligned}$$

Expected Shortfall: Definition and Properties

- The expected shortfall is defined as

$$\mathrm{ES}_\theta\left(R\right) = \mathbb{E}\left[R \mid R \leq q_{R,\theta}\right] = \frac{1}{\theta}\int_{-\infty}^{q_{R,\theta}} r\, f_R\left(r\right) \mathrm{d}r,$$

  where R is a future period financial return and $q_{R,\theta}$ is the θ-quantile such that $\Pr\left(R \leq q_{R,\theta}\right) = \theta$ and θ is small, typically 1%.

- **Exercise II** Let Z be a location zero, scale one r.v., and let $Y = \sigma Z + \mu$ for $\sigma > 0$. Show that

$$\mathrm{ES}_\theta\left(Y\right) = \mu + \sigma \mathrm{ES}_\theta\left(Z\right),$$

  i.e., that ES preserves location-scale transformations.
  Hint: First show that $q_{Y,\theta} = \sigma q_{Z,\theta} + \mu$.

Expected Shortfall: Properties

Solution to Exercise II

First, we have

$$\begin{aligned}\Pr\left(Z \le q_{Z,\theta}\right) = \theta \quad &\Leftrightarrow \quad \Pr\left(\sigma Z + \mu \le \sigma q_{Z,\theta} + \mu\right) = \theta \\ &\Leftrightarrow \quad q_{Y,\theta} = \sigma q_{Z,\theta} + \mu.\end{aligned}$$

Then,

$$\begin{aligned}\mathrm{ES}_\theta\left(Y\right) &= \mathbb{E}\left[Y \mid Y \le q_{Y,\theta}\right] \\ &= \mathbb{E}\left[\sigma Z + \mu \mid \sigma Z + \mu \le \sigma q_{Z,\theta} + \mu\right] \\ &= \sigma \mathbb{E}\left[Z \mid Z \le q_{Z,\theta}\right] + \mu = \sigma\, \mathrm{ES}_\theta\left(Z\right) + \mu.\end{aligned}$$

Expected Shortfall: Properties

- **Exercise III** Let Q_X be the quantile function of continuous r.v. X, i.e., $Q_X : (0,1) \to \mathbb{R}$ with $p \mapsto F_X^{-1}(p)$. Show that $\mathrm{ES}_\theta(X)$ can be expressed as

$$\mathrm{ES}_\theta(X) = \frac{1}{\theta} \int_0^\theta Q_X(p)\, \mathrm{d}p.$$

- This is a common form of expressing ES because a weighting function (called the risk spectrum or risk-aversion function) can be incorporated into the integral to form the so-called spectral risk measure.

Expected Shortfall: Properties

Solution to Exercise III

Let $u = Q_X(p)$, so that $p = F_X(u)$ and $\mathrm{d}p = f_X(u)\,\mathrm{d}u$. Then, with $q_\theta = Q_X(\theta)$,

$$\int_0^\theta Q_X(p)\,\mathrm{d}p = \int_{-\infty}^{q_\theta} u\, f_X(u)\,\mathrm{d}u.$$

To verify this in Matlab, we use the $\mathrm{N}(0,1)$ case and run:

```
alpha=0.01; c=norminv(alpha);
ES1 = -normpdf(c)/normcdf(c)
ES2 = quadl(@norminv, 1e-7, alpha, 1e-7, 0) / alpha
```

Expected Shortfall: Properties

Exercise IV

We defined the expected shortfall to be

$$\mathrm{ES}_\theta\left(R\right) = \frac{1}{\theta}\int_{-\infty}^{q_{R,\theta}} r\, f_R\left(r\right)\,\mathrm{d}r.$$

Show via integration by parts that, if $\mathbb{E}\left[R\right]$ exists, then we can also write

$$\mathrm{ES}_\theta\left(R\right) = q_{R,\theta} - \frac{1}{\theta}\int_{-\infty}^{q_{R,\theta}} F_R\left(r\right)\,\mathrm{d}r. \tag{47}$$

Recalling that $q_{R,\theta}$ is the θ-level VaR, this shows that, in absolute terms, $\mathrm{ES}_\theta\left(R\right)$ will be more extreme than the VaR.

Expected Shortfall: Properties

Solution to Exercise IV

For the integral in (47), with $u = F_R(r)$ and $\mathrm{d}v = \mathrm{d}r$,

$$\int_{-\infty}^{q_{R,\theta}} F_R(r)\,\mathrm{d}r = rF_R(r)\Big|_{-\infty}^{q_{R,\theta}} - \int_{-\infty}^{q_{R,\theta}} rf_R(r)\,\mathrm{d}r.$$

The result then follows if we can show that the first term is $q_{R,\theta}\ \theta$, i.e., if $\lim_{r\to-\infty} rF_R(r) = 0$.

In particular: Let X be a continuous random variable with finite expected value. Show that $\lim_{x\to-\infty} xF_X(x) = 0$ and, if X is non-positive, then $\mathbb{E}[X] = -\int_{-\infty}^{0} F_X(x)\,\mathrm{d}x$.

Solution to Exercise IV, cont.

To show $\lim_{x\to-\infty} xF_X(x) = 0$, note that, as $0 \le F_X(x) \le 1$ for all x,

$$
\begin{aligned}
0 \ge \lim_{x\to-\infty} xF_X(x) &= \lim_{x\to-\infty} x \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \\
&= \lim_{x\to-\infty} \int_{-\infty}^{x} xf_X(t)\,\mathrm{d}t \\
&\overset{x>t}{\ge} \lim_{x\to-\infty} \int_{-\infty}^{x} tf_X(t)\,\mathrm{d}t = 0,
\end{aligned}
$$

where the last equality follows because we assumed that $\mathbb{E}[X]$ exists.

So, $\lim_{x\to-\infty} xF_X(x)$ is bound above and below by zero and, thus, is zero.

Solution to Exercise IV, cont.

Another way is to note that, if $t < x < 0$, then $|x| < |t|$, so that

$$\begin{aligned} 0 \leq \left| x F_X(x) \right| &= |x| \int_{-\infty}^{x} f_X(t)\, \mathrm{d}t \\ &\leq \int_{-\infty}^{x} |t|\, f_X(t)\, \mathrm{d}t = \int_{-\infty}^{x} (-t)\, f_X(t)\, \mathrm{d}t = -\int_{-\infty}^{x} t f_X(t)\, \mathrm{d}t, \end{aligned}$$

and taking limits shows that

$$0 \leq \lim_{x \to -\infty} |x F_X(x)| \leq -\lim_{x \to -\infty} \int_{-\infty}^{x} t f_X(t)\, \mathrm{d}t = 0,$$

where, as before, the last equality follows because we assumed that $\mathbb{E}[X]$ exists.

Thus, $\lim_{x \to -\infty} \left| x F_X(x) \right| = 0$, which implies $\lim_{x \to -\infty} x F_X(x) = 0$.

Expected Shortfall: Properties

Exercise V

Recall that the Student's t p.d.f. with v degrees of freedom is given by

$$f_t(x; v) = \frac{\Gamma\left(\frac{v+1}{2}\right) v^{\frac{v}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{v}{2}\right)} \left(v + x^2\right)^{-\frac{v+1}{2}} = \frac{v^{-\frac{1}{2}}}{B\left(\frac{v}{2}, \frac{1}{2}\right)} \left(1 + x^2/v\right)^{-\frac{v+1}{2}}.$$

Let ϕ_v and Φ_v denote the p.d.f. and c.d.f. of the Student's t distribution with v degrees of freedom, respectively.

Calculate the ES for the Student's t.

Expected Shortfall: Properties

Solution to Exercise V

With $K = v^{-1/2} \,/\, B\left(v/2, 1/2\right)$ and $u = 1 + r^2/v$,
$\mathbb{E}\left[R \mid R < c\right] = \frac{1}{\Phi_v(c)} \int_{-\infty}^{c} r\phi_v(r)\,\mathrm{d}r$ is

$$
\begin{aligned}
& \frac{K}{\Phi_v\left(c\right)} \int_{-\infty}^{c} r\left(1 + r^2/v\right)^{-(v+1)/2}\,\mathrm{d}r \\
= \quad & \frac{K}{\Phi_v\left(c\right)} \frac{v}{2} \int_{\infty}^{1+c^2/v} u^{-(v+1)/2}\,\mathrm{d}u \\
= \quad & -\frac{K}{\Phi_v\left(c\right)} \frac{v}{1-v}\, u^{1-(v+1)/2}\Big|_{1+c^2/v}^{\infty} = \frac{K}{\Phi_v\left(c\right)} \frac{v}{1-v} \left(1 + c^2/v\right)^{1-(v+1)/2} \\
= \quad & -\frac{\phi_v(c)}{\Phi_v(c)} \times \left[\frac{v+c^2}{v-1}\right],
\end{aligned}
$$

from which it is clear that, as $v \to \infty$, the expression approaches that based on the normal distribution.

Expected Shortfall: Properties

This is very easy to confirm via simulation with all the built-in functions of Matlab:

```
N=10^6; df=3; gama=0.01; q=tinv(gama,df);
X=trnd(df,N,1); use=X(X<q);
ES = -tpdf(q,df)/tcdf(q,df) * (df+q^2)/(df-1);
empiricalES_and_trueES = [mean(use) ES]
```

Simulation of Expected Shortfall

Invoking the (weak) law of large numbers, simulation can be used to compute, at least approximately, the ES. Taking the standard normal case to illustrate, for some $\gamma \in (0,1)$ and $q = q_{Z,\gamma}$ the γ-quantile of Z,

$$\mathrm{ES}_\gamma(Z) = \mathbb{E}[Z \mid Z < q] = \frac{1}{\gamma}\int_{-\infty}^{q} z f_Z(z)\,\mathrm{d}z.$$

Let $Z_i \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$, $i = 1, \ldots, n$.
The integral in ES can be written as $\mathbb{E}[g(Z)]$ with $g(Z) = Z\mathbb{I}(Z < q)$.
So, defining $Y_i = Z_i\mathbb{I}(Z_i < q)$ and $\bar{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i$,

$$\frac{1}{\gamma}\bar{Y}_n \xrightarrow{p} \mathrm{ES}_\gamma(Z).$$

This can be computed in Matlab as

```
gama=0.05; q=norminv(0.05); Z=randn(1e6,1);
I=(Z<q); mean(Z.*I)/gama
```

Simulation of Expected Shortfall

An alternative way of computing this is to approximate $\mathbb{E}[Z \mid Z < q]$ directly by taking the average of just those values of Z which are less than q. In Matlab,

```
gama=0.05; q=norminv(0.05); Z=randn(1e6,1);
W=Z(Z<q); mean(W)
```

The two methods are asymptotically equivalent, and are numerically identical if we take the first one to be

```
I=(Z<q); mean(Z.*I)/mean(I)
```

Computation of ES via SPA

Recall the saddlepoint approximation (spa) to the distribution of a random variable Z, based on its m.g.f.: With $\hat{s}$ the saddlepoint, formulae for the p.d.f. and c.d.f. are available. Martin (2006)[5] has shown that the integral in $\mathrm{ES}_\gamma(Z)$ can be approximated as

$$\int_{-\infty}^{q} z f_Z(z)\,\mathrm{d}z \approx \mu_Z F_Z(q) - f_Z(q)\frac{q-\mu_Z}{\hat{s}},$$

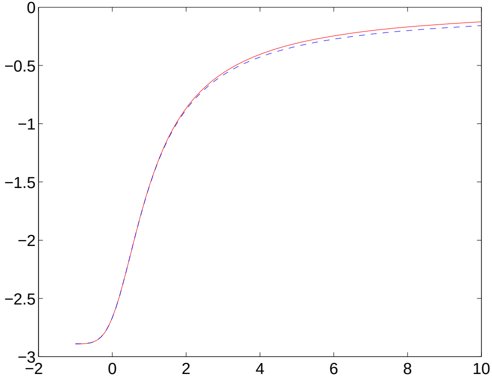
where $\mu_Z = \mathbb{E}[Z]$. Values $f_Z(q)$ and $F_Z(q)$ can of course be replaced by their s.p.a. counterparts.

As an example, the next figure shows the true (solid) expected shortfall and the s.p.a. (dashed) for $\gamma = 0.01$ for the skew normal distribution. The exact values were calculated via numeric integration.

[5] Richard Martin (2006), *The Saddlepoint Method and Portfolio Optionalities*, Risk Magazine, 19(12), 93-95.

Computation of ES via SPA

The approximation is exact for the normal case ($\lambda = 0$), stays extremely accurate for $\lambda < 0$, and worsens as λ increases from zero.



Expected Shortfall for Normal Mixtures

Let $X \sim \mathrm{MixN}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$, with $f_{\mathrm{MixN}}\left(x; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}\right) = \sum_{c=1}^{k} \lambda_c \phi\left(x; \mu_c, \sigma_c^2\right)$.

The γ–quantile of X, $q_{X,\gamma}$, can be found numerically by solving

$$\gamma - F_X\left(q_{X,\gamma}; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}\right) = 0,$$

where F_X is given above.

The ES can be computed directly from the definition, using numeric integration (and replacing $-\infty$ with, say, -100). This is easy to implement and fast to compute.

However, a bit of algebra shows that the ES can be expressed in other forms which are more convenient for numerical calculation and also interpretation. We consider two such forms.

Expected Shortfall for Normal Mixtures: 1st Form

Let $X_j \sim \mathrm{N}\left(\mu_j, \sigma_j^2\right)$ be the jth component in the mixture with density $f_{X_j}\left(x; \mu_j, \sigma_j^2\right) = \phi\left(x; \mu_j, \sigma_j^2\right)$.

Now, (i) use the fact that, if $Z \sim \mathrm{N}(0,1)$ and $X_j = \mu_j + \sigma_j Z$, then $f_{X_j}(x) = \sigma_j^{-1} f_Z(z)$ where $z = (x - \mu_j)/\sigma_j$, (ii) substitute $z = (x - \mu_j)/\sigma_j$, and (iii) recall that $\int_{-\infty}^{c} z f_Z(z) \mathrm{d}z = -\phi(c)$ to get...

Expected Shortfall for Normal Mixtures: 1st Form

$$\begin{aligned}
\mathrm{ES}_{\gamma}\left(X;\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\lambda}\right) &= \frac{1}{\gamma}\int_{-\infty}^{q_{X,\gamma}} x\, f_X(x)\ \mathrm{d}x \\
&= \frac{1}{\gamma}\sum_{j=1}^{k}\lambda_j\int_{-\infty}^{q_{X,\gamma}} x\, f_{X_j}\left(x;\mu_j,\sigma_j^2\right)\ \mathrm{d}x \\
&= \frac{1}{\gamma}\sum_{j=1}^{k}\lambda_j\int_{-\infty}^{q_{X,\gamma}} x\ \sigma_j^{-1} f_Z\left(\frac{x-\mu_j}{\sigma_j}\right)\ \mathrm{d}x \\
&= \frac{1}{\gamma}\sum_{j=1}^{k}\lambda_j\int_{-\infty}^{\frac{q_{X,\gamma}-\mu_j}{\sigma_j}} \left(\sigma_j z+\mu_j\right)\sigma_j^{-1} f_Z(z)\ \sigma_j\ \mathrm{d}z \\
&= \frac{1}{\gamma}\sum_{j=1}^{k}\lambda_j\left[-\sigma_j\phi\left(\frac{q_{X,\gamma}-\mu_j}{\sigma_j}\right)+\mu_j\Phi\left(\frac{q_{X,\gamma}-\mu_j}{\sigma_j}\right)\right],
\end{aligned}$$

which is easily numerically calculated.

Expected Shortfall for Normal Mixtures: 1st Form

Further, letting $c_j := (q_{X,\gamma} - \mu_j)/\sigma_j$ and factoring out $\Phi(c_j)$ gives

$$\begin{aligned}
\mathrm{ES}_\gamma(X; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) &= \frac{1}{\gamma} \sum_{j=1}^{k} \lambda_j \Phi(c_j) \left[\mu_j - \sigma_j \frac{\phi(c_j)}{\Phi(c_j)}\right] \\
&= \sum_{j=1}^{k} \frac{\lambda_j \Phi(c_j)}{\gamma} \left[\mu_j - \sigma_j \frac{\phi(c_j)}{\Phi(c_j)}\right],
\end{aligned}$$

which has the appearance of a weighted sum of the component ESs, but notice that $\mu_j - \sigma_j \phi(c_j)/\Phi(c_j)$ is not $\mathrm{ES}_\gamma(X_j)$ because

$$c_j = (q_{X,\gamma} - \mu_j)/\sigma_j \neq (q_{X_j,\gamma} - \mu_j)/\sigma_j = q_{Z,\gamma} = \Phi^{-1}(\gamma).$$

Expected Shortfall for Normal Mixtures: 1st Form

We *could* write

$$\mathrm{ES}_\gamma\left(X; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}\right) = \sum_{j=1}^{k} \omega_j \, \mathrm{ES}_\gamma\left(X_j\right)$$

for

$$\omega_j := \frac{\lambda_j \Phi\left(c_j\right)}{\gamma} \frac{\mu_j - \sigma_j \phi\left(c_j\right) / \Phi\left(c_j\right)}{\mu_j - \sigma_j \phi\left(q_{Z,\gamma}\right) / \Phi\left(q_{Z,\gamma}\right)},$$

and let $\omega_j^* = \omega_j / \sum_{j=1}^{k} \omega_j$, and the ω_j^* can be interpreted as the fraction of the ES attributed to component j.

Expected Shortfall for Normal Mixtures: 2nd Form

It is also straightforward to express the ES as

$$\mathrm{ES}_\gamma\left(X;\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\lambda}\right)=\sum_{j=1}^{k}\lambda_j\,\mathrm{ES}_\gamma\left(X_j\right)+\text{ something remaining,}$$

which is also easy to compute, and explicitly shows the how the ES is a function of the weighted individual ES components. Writing

$$\begin{aligned}\mathrm{ES}_\gamma\left(X;\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\lambda}\right) &= \frac{1}{\gamma}\int_{-\infty}^{q_{X,\gamma}} x\,f_X\left(x\right)\mathrm{d}x\\ &= \sum_{j=1}^{k}\lambda_j\frac{1}{\gamma}\int_{-\infty}^{q_{X,\gamma}} x\,f_{X_j}\left(x;\mu_j,\sigma_j^2\right)\mathrm{d}x\end{aligned}$$

or, splitting up the integral into two pieces, and letting

$$q_j := q_{X_j,\gamma}=\Phi^{-1}\left(\gamma;\mu_j,\sigma_j^2\right)=\mu_j+\sigma_j\Phi^{-1}\left(\gamma\right)$$

being the γ-quantile of X_j, we have...

Expected Shortfall for Normal Mixtures: 2nd Form

$$
\begin{aligned}
&\mathrm{ES}_\gamma\left(X;\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\lambda}\right)\\
&= \sum_{j=1}^{k}\lambda_j\left[\frac{1}{\gamma}\int_{-\infty}^{q_j} x\, f_{X_j}\left(x;\mu_j,\sigma_j^2\right)\,\mathrm{d}x + \frac{1}{\gamma}\int_{q_j}^{q_{X,\gamma}} x\, f_{X_j}\left(x;\mu_j,\sigma_j^2\right)\,\mathrm{d}x\right]\\
&= \sum_{j=1}^{k}\lambda_j\,\mathrm{ES}_\gamma\left(X_j\right) + \frac{1}{\gamma}\sum_{j=1}^{k}\lambda_j g_j\left(\gamma\right),
\end{aligned}
$$

where

$$
g_j\left(\gamma\right) := \int_{q_j}^{q_{X,\gamma}} x f_{X_j}\left(x;\mu_j,\sigma_j^2\right)\,\mathrm{d}x
$$

and, with $Z \sim \mathrm{N}\left(0,1\right)$ and recalling that ES preserves location-scale transformations,

$$
\mathrm{ES}_\gamma\left(X_j\right) = \mu_j + \sigma_j\,\mathrm{ES}_\gamma\left(Z\right) = \mu_j - \sigma_j\phi\left(\Phi^{-1}\left(\gamma\right)\right)/\gamma.
$$

Lower Partial Moments

If it exists, the nth order *lower partial moment* with respect to reference point c is, for $n \in \mathbb{N}$,

$$\mathrm{LPM}_{n,c}(X) = \int_{-\infty}^{c} (c - x)^n f_X(x)\,\mathrm{d}x.$$

This is an important measure for financial portfolio risk with many advantages over the traditional measure (variance). It is related to the Expected Shortfall.

The LPM can be computed with numeric integration, though for both the normal and for fat-tailed distributions, choosing the lower bound on the integral can be problematic.

Lower Partial Moments

Applying the binomial theorem to $(c-x)^n$, we can write

$$\mathrm{LPM}_{n,c}(X) = \sum_{h=0}^{n} K_{h,c} T_{h,c}(X), \tag{48}$$

where we define

$$K_{h,c} = K_{h,c}(n) = \binom{n}{h} c^{n-h} (-1)^h$$

and

$$T_{h,c}(X) = \int_{-\infty}^{c} x^h f_X(x)\, \mathrm{d}x.$$

Now we just need "closed form" expressions for $T_{h,c}(X)$. With them, (48) can be quickly and accurately evaluated without the aforementioned numeric integration problem.

Exercise: Lower Partial Moments for Normal

For $Z \sim \mathrm{N}(0,1)$ and $c < 0$, calculation shows (let $u = z^2/2$ for $z < 0$) that, for $h \in \mathbb{N}$,

$$T_{h,c}(Z) = \frac{(-1)^h \, 2^{h/2-1}}{\sqrt{\pi}} \left[\Gamma\left(\frac{h+1}{2}\right) - \Gamma_{c^2/2}\left(\frac{h+1}{2}\right) \right], \qquad (49)$$

where $\Gamma_x(a)$ is the incomplete gamma function.

In particular,

$$T_{0,c}(Z) = \Phi(c) \text{ and } T_{1,c}(Z) = -\phi(c).$$

First show $T_{0,c}(Z)$, then $T_{1,c}(Z)$, and finally the general expression for $T_{h,c}(Z)$ above.

Exercise: Lower Partial Moments for Student's t

For $X \sim T(v)$, with density

$$f_T(x; v) = K_v \left(1 + x^2/v\right)^{-\frac{v+1}{2}}, \quad K_v = \frac{v^{-\frac{1}{2}}}{B\left(\frac{v}{2}, \frac{1}{2}\right)},$$

(substitute $u = 1 + x^2/v$ for $x < 0$ and then $x = (u-1)/u$), for $h < v$, $T_{h,c}(X; v)$ is

$$\frac{(-1)^h v^{h/2}}{2B\left(\frac{v}{2}, \frac{1}{2}\right)} \left[B\left(\frac{h+1}{2}, \frac{v-h}{2}\right) - B_w\left(\frac{h+1}{2}, \frac{v-h}{2}\right)\right], \tag{50}$$

where $w = \frac{c^2/v}{1+c^2/v}$ and B_w is the incomplete beta function. Show this.

In particular,

$$T_{0,c}(X; v) = F_X(c; v) = \Phi_v(c) \text{ and } T_{1,c}(X; v) = \phi_v(c)\left(v + c^2\right)/(1-v).$$

Simulation of LPM

Similar to the simulation of Expected Shortfall, for

$$\mathrm{LPM}_{n,c}(Z) = \int_{-\infty}^{c} (c-z)^n f_Z(z)\,\mathrm{d}z,$$

we have $g(Z) = (c-Z)^n \mathbb{I}(Z<c)$, and $\bar{Y}_n \xrightarrow{p} \mathrm{LPM}_{n,c}(Z)$ for $Y_i = (c-Z_i)^n \mathbb{I}(Z_i<c)$, computed for $n=2$ as

```
gama=0.05; q=norminv(0.05); Z=randn(1e6,1);
I=(Z<q); Y=(q-Z).^2; mean(Y.*I)
```

Or we can compute the conditional version as

$$\frac{\Pr(Z<c)}{\Pr(Z<c)} \int_{-\infty}^{c} (c-z)^n f_Z(z)\,\mathrm{d}z = \Pr(Z<c)\,\mathbb{E}\left[(c-Z)^n \mid Z<c\right]$$

using

```
gama=0.05; q=norminv(0.05); Z=randn(1e6,1);
W=Z(Z<q); mean((q-W).^2)*gama
```

LPM for Mixtures of Normals and Student's t

If $X \sim \mathrm{MixN}\left(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}\right)$ with k components, then $\mathrm{LPM}_{n,c}\left(X\right)$ can be computed from (48) and (49), where $T_{h,c}\left(X\right)$ is, similar to the derivation of ES for mixtures, given by

$$T_{h,c}\left(X\right) = \sum_{j=1}^{k} \lambda_j \sum_{m=0}^{h} \binom{h}{m} \sigma_j^m \mu_j^{h-m} T_{m,(c-\mu_j)/\sigma_j}\left(Z\right), \qquad (51)$$

for $Z \sim \mathrm{N}(0,1)$. An important special case is for $k = 1$, which corresponds to $X \sim \mathrm{N}(\mu, \sigma^2)$. Similarly, if $X \sim \mathrm{MixT}\left(\mathbf{v}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}\right)$ and $h < \min v_j$, the same result holds, just replace $T_{m,(c-\mu_j)/\sigma_j}\left(Z\right)$ in (51) with $T_{m,(c-\mu_j)/\sigma_j}\left(Y; v_j\right)$ from (50), with $Y \sim t_{v_j}$.

The $\mathrm{LPM}_{n,c}\left(X\right)$ can also be easily computed with simulation (though is far slower and less accurate), and serves as a check that (51) is programmed correctly.

Countable Mixtures

- For any $k \in \mathbb{N}$, X is a finite mixture (e.g., mixed normal).
- With $k = \infty$, $\mathcal{C}$ becomes countably infinite (e.g., Poisson), and X is then said to follow *countable mixture distribution*.
- An important example which we will study in more detail later is the noncentral χ^2 distribution: Let $X_i \overset{\text{ind}}{\sim} \mathrm{N}(\mu_i, 1)$, $\mu_i \in \mathbb{R}$, $i = 1, \ldots, n$, or $\mathbf{X} = (X_1, \ldots, X_n) \sim \mathrm{N}_n(\boldsymbol{\mu}, \mathbf{I})$, with $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$.
- Then $X = \sum_{i=1}^n X_i^2$ follows a noncentral χ^2 distribution with *noncentrality parameter* $\boldsymbol{\mu}'\boldsymbol{\mu} = \sum_{i=1}^n \mu_i^2$.
- We show later that the p.d.f. of X can be expressed as

$$f_X(x; n, \theta) = \sum_{i=0}^{\infty} \omega_{i,\theta} \; g_{n+2i}(x),$$

where g_v denotes the χ_v^2 density and $\omega_{i,\theta} = \mathrm{e}^{-\theta/2} (\theta/2)^i \,/\, i!$ are weights corresponding to a Poisson distribution.

Moments of Mixtures

- An important consequence of X being the marginal distribution from $f_{X,C} = f_{X|C} f_C$ is that the iterated expectation and conditional variance formula are applicable, i.e.,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid C]]$$

  and

$$\mathbb{V}(X) = \mathbb{E}[\mathbb{V}(X \mid C)] + \mathbb{V}(\mathbb{E}[X \mid C]).$$

- For example, let $X \sim \mathrm{MixN}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$. Then, from the moments of the normal distribution,

$$\mathbb{E}[X] = \sum_{i=1}^{k} \lambda_i \mu_i, \quad \mathbb{E}\left[X^2\right] = \sum_{i=1}^{k} \lambda_i \left(\mu_i^2 + \sigma_i^2\right),$$

  from which $\mathbb{V}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$ can be computed.

Moments of Mixtures

More generally, to compute $\mu_r'(X)$, the rth raw moment of X, recall

$$\mathbb{E}\left[g(X) \mid Y = y\right] = \int_{x\in\mathbb{R}} g(x)\, \mathrm{d}F_{X|Y}(x \mid y);$$

in particular, for a fixed $r \in \mathbb{N}$ and $c \in \mathcal{C}$,

$$\mathbb{E}\left[X^r \mid C = c\right] = \mu_r'(X \mid C = c) = \int_{x\in\mathbb{R}} x^r\, \mathrm{d}F_{X|C}(x \mid c).$$

Taking expectations of both sides with respect to Y led to the law of the iterated expectation $\mathbb{E}\,\mathbb{E}\left[g(X) \mid Y\right] = \mathbb{E}\left[g(X)\right]$, i.e.,

$$\mathbb{E}\left[X^r\right] = \mu_r'(X) = \int_{\mathcal{C}} \mu_r'(X \mid C = c)\, \mathrm{d}F_C(c) = \sum_{c=1}^{k} \lambda_c \mu_r'(X \mid C = c).$$

C.F. of Mixture of Normals

Similarly, the m.g.f. of X is

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}_C\left[\mathbb{E}\left[e^{tX} \mid C = c\right]\right] = \int_{\mathcal{C}} \mathbb{M}_{X \mid C=c}(t)\, \mathrm{d}F_C(c) = \sum_{c=1}^{k} \lambda_c \mathbb{M}_{X \mid C=c}(t).$$

Let $X_j \sim \mathrm{N}\left(\mu_j, \sigma_j^2\right)$, and recall that the c.f. of each X_j is

$$\varphi_{X_j}(t) = \mathbb{E}\left[e^{itX_j}\right] = \exp\left\{it\mu_j - \frac{\sigma_j^2 t^2}{2}\right\}.$$

With $M \sim \mathrm{MixN}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$, the c.f. follows from the same calculation as was used for the m.g.f. above, i.e.,

$$\varphi_M(t) = \sum_{j=1}^{k} \lambda_j \varphi_{M \mid C=j}(t) = \sum_{j=1}^{k} \lambda_j \exp\left\{it\mu_j - \frac{\sigma_j^2 t^2}{2}\right\}.$$

C.F. of Mixture of Normals

Just to check, let $k = 2$, so that

$$f_M(x; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) = \lambda_1 f_{X_1}\left(x; \mu_1, \sigma_1^2\right) + \lambda_2 f_{X_2}\left(x; \mu_2, \sigma_2^2\right).$$

The c.f. of M is then, from the definition,

$$\begin{aligned}
\varphi_M(t) &= \mathbb{E}\left[\mathrm{e}^{itM}\right] \\
&= \int_{-\infty}^{\infty} \mathrm{e}^{itx}\left[\lambda_1 f_{X_1}\left(x; \mu_1, \sigma_1^2\right) + \lambda_2 f_{X_2}\left(x; \mu_2, \sigma_2^2\right)\right] \mathrm{d}x \\
&= \lambda_1 \varphi_{X_1}(t) + \lambda_2 \varphi_{X_2}(t) \\
&= \lambda_1 \exp\left\{it\mu_1 - \frac{\sigma_1^2 t^2}{2}\right\} + \lambda_2 \exp\left\{it\mu_2 - \frac{\sigma_2^2 t^2}{2}\right\}.
\end{aligned}$$

C.F. of Mixture of Normals

As a final check, we wish to calculate the c.f. of M and verify numerically that the p.d.f. calculated via the inversion formula yields the same values as using the pdf expression for f_M directly.

The inversion of the c.f. to get the p.d.f. and c.d.f. is programmed in `mixnormcf.m` (next slide), using equations (1.60) and (1.71) from the text.

We verify that the c.f. is correct by using the code:

```
x=-5:0.1:5; mu=[-0.2 0.3]; std=[2 1]; lam=[0.4 0.6];
f1=lam(1)*normpdf(x,mu(1),std(1)) + lam(2)*normpdf(x,mu(2),std(2));
[f2,F2] = mixnormcf(x,mu,std.^2,lam); plot(x,f1'-f2)
```

C.F. of Mixture of Normals

```
function [f,F] = mixnormcf(xvec,mu,s2,lam)
bordertol=1e-8; lo=bordertol; hi=1-bordertol; tol=1e-8;
xl=length(xvec); F=zeros(xl,1); f=F;
for loop=1:length(xvec),   x=xvec(loop); dopdf=1;
  f(loop)= quadl(@fff,lo,hi,tol,[],x,mu,s2,lam,1) / pi;
  if nargout>1
    F(loop)=0.5-(1/pi)* quadl(@fff,lo,hi,tol,[],x,mu,s2,lam,0);
  end
end;

function I=fff(uvec,x,mu,s2,lam,dopdf);
for ii=1:length(uvec),   u=uvec(ii);  t  = (1-u)/u;
  cf = lam(1)*exp( i*t*mu(1) - (s2(1)*t.^2)/2 ) ...
      + lam(2)*exp( i*t*mu(2) - (s2(2)*t.^2)/2 );
  z  = exp(-i*t*x) .* cf;
  if dopdf==1, g=real(z); else g=imag(z)./t; end
  I(ii)  = g / u^2;
end
```

Sums of Independent Mixture of Normals

Now let $M_j \overset{\text{ind}}{\sim} \text{MixN}\left(\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j, \boldsymbol{\lambda}_j\right)$, $j = 1, \ldots, d$, and M_j has k_j mixture components, $j = 1, \ldots, d$.

With $\varphi_{M_j}(t; \boldsymbol{\mu}_j, \boldsymbol{\sigma}_j, \boldsymbol{\lambda}_j)$ denoting the c.f. of each M_j, the c.f. of the weighted sum $S = \sum_{j=1}^{d} w_j M_j$ is just

$$\begin{aligned} \varphi_S(t) &= \prod_{j=1}^{d} \varphi_{M_j}(w_j t; \boldsymbol{\mu}_j, \boldsymbol{\sigma}_j, \boldsymbol{\lambda}_j) \\ &= \prod_{j=1}^{d} \sum_{c=1}^{k_j} \lambda_{j,c} \exp\left\{ i t w_j \mu_{j,c} - \frac{\sigma_{j,c}^2 t^2 w_j^2}{2} \right\} \end{aligned}$$

where $\lambda_{j,c}$ denotes the cth element of $\boldsymbol{\lambda}_j$, and similarly for $\mu_{j,c}$ and $\sigma_{j,c}$.

The inversion theorems or FFT can be applied in the usual way to this expression.

Continuous Mixtures

The random variable X is said to follow a *continuous mixture distribution* if its p.d.f. can be expressed as

$$f_X(x) = \int_{\mathcal{C}} f_{X|C}(x \mid c) \, f_C(c) \, \mathrm{d}c.$$

or, equivalently, if its c.d.f. is $F_X(x) = \int_{\mathcal{C}} F_{X|C}(x \mid c) \, f_C(c) \, \mathrm{d}c$.

Continuous Mixtures

The mean and variance are as in the discrete case,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid C]] \quad \text{and} \quad \mathbb{V}(X) = \mathbb{E}[\mathbb{V}(X \mid C)] + \mathbb{V}(\mathbb{E}[X \mid C]).$$

Similarly, the raw moments are given by

$$\mu_r'(X) = \int_{\mathcal{C}} \mu_r'(X \mid C = c)\, \mathrm{d}F_C(c),$$

and the m.g.f. of X is

$$\mathbb{E}\left[\mathrm{e}^{tX}\right] = \int_{\mathcal{C}} \mathbb{M}_{X \mid C=c}(t)\, f_C(c)\, \mathrm{d}c.$$

Example (negative binomial)

Let $(X \mid R = r) \sim \mathrm{Poi}(r)$, $r \in \mathbb{R}_{>0}$, and $R \sim \mathrm{Gam}(a, b)$. Then

$$\mathbb{M}_{X \mid R=r}(t) = \exp\left\{ r \left(\mathrm{e}^t - 1 \right) \right\},$$

and with $k = b + 1 - \mathrm{e}^t$,

$$\begin{aligned}
\mathbb{M}_X(t) &= \int_0^\infty \exp\left\{ r \left(\mathrm{e}^t - 1 \right) \right\} \; \frac{b^a}{\Gamma(a)} r^{a-1} \exp\{-br\} \, \mathrm{d}r \\
&= \frac{b^a}{\Gamma(a)} \int_0^\infty \exp\{-kr\} \, r^{a-1} \, \mathrm{d}r = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a)}{k^a} = \left(\frac{b}{b + 1 - \mathrm{e}^t} \right)^a .
\end{aligned}$$

Setting $p = b/(b+1)$, $b = p/(1-p)$ and simplifying gives

$$\mathbb{M}_X(t) = \left(\frac{p}{1 - (1-p)\,\mathrm{e}^t} \right)^a, \quad \text{or} \quad X \sim \mathrm{NBin}\left(a, \frac{b}{b+1} \right).$$

Example (Type II Generalized Pareto)

Let $\Theta \sim \mathrm{Gam}(b, c)$, and $(X \mid \Theta = \theta) \sim \mathrm{Gam}(a, \theta)$, so

$$\begin{aligned} f_X(x; a, b, c) &= \int_{-\infty}^{\infty} f_{X|\Theta}(x \mid a, \theta)\, f_\Theta(\theta; b, c)\, \mathrm{d}\theta \\ &= \frac{1}{\Gamma(a)} \frac{c^b}{\Gamma(b)} x^{a-1} \mathbb{I}_{(0,\infty)}(x) \int_0^{\infty} \exp\{-\theta(x+c)\} \theta^{a+b-1}\, \mathrm{d}\theta. \end{aligned}$$

From the gamma p.d.f., the integral is $\Gamma(a+b)/(x+c)^{a+b}$, so

$$f_X(x; a, b, c) = \frac{c^b}{B(a, b)} \frac{x^{a-1}}{(x+c)^{a+b}} \mathbb{I}_{(0,\infty)}(x),$$

which is the type II generalized Pareto density.

(With $a = 1$, it is the type II Pareto.)

Example (Laplace)

Let $(X \mid V = v) \sim \mathrm{N}(\mu, v)$ and $V \sim \mathrm{Exp}(\lambda)$.
The m.g.f. of X is, with $T = \lambda - t^2/2$ and $T > 0$ (i.e., $|t| < \sqrt{2\lambda}$),

$$\begin{aligned}
\mathbb{M}_X(t) &= \int_0^\infty \exp\left\{\mu t + \frac{1}{2} v t^2\right\} \lambda \exp\{-\lambda v\}\, \mathrm{d}v \\
&= \lambda \mathrm{e}^{\mu t} \int_0^\infty \exp\{-vT\}\, \mathrm{d}v \\
&= \frac{\lambda \mathrm{e}^{\mu t}}{\lambda - t^2/2} = \mathrm{e}^{\mu t} \frac{1}{1 - t^2/(2\lambda)}, \quad |t| < \sqrt{2\lambda}.
\end{aligned}$$

From the form of the m.g.f., we see that μ is just a location parameter: we set it to zero without loss of generality; and λ serves as a scale parameter (the scaling parameter is $1/\sqrt{2\lambda}$), so it suffices to examine $\mathbb{M}_X(t) = 1/(1 - t^2)$ for $|t| < 1$.

This is the m.g.f. of a Laplace random variable!

Example (Laplace generalization)

Generalize the previous example as follows:
Let $(X \mid \Theta = \theta) \sim \mathrm{N}(0, \theta)$ and $\Theta \sim \mathrm{Gam}(b, c)$. Then, with $H = c - t^2/2$ and $H > 0$ (i.e., $|t| < \sqrt{2c}$),

$$\begin{aligned}
\mathbb{M}_X(t) &= \int_0^\infty \exp\left\{\frac{1}{2}\theta t^2\right\} \frac{c^b}{\Gamma(b)} \theta^{b-1} \exp\{-c\theta\} \,\mathrm{d}\theta \\
&= \frac{c^b}{\Gamma(b)} \int_0^\infty \theta^{b-1} \exp\{-\theta H\} \,\mathrm{d}\theta = \frac{c^b}{\Gamma(b)} \frac{\Gamma(b)}{H^b} \\
&= \left(1 - t^2/(2c)\right)^{-b}, \quad |t| < \sqrt{2c}.
\end{aligned}$$

As before, c serves as a scale parameter, and $1/\sqrt{2c}$ could be set to unity without loss of generality.
This m.g.f. is not recognizable, but Example 2.10 showed that it coincides with the m.g.f. of the difference of two i.i.d. scaled gamma r.v.s.

Example (Laplace generalization)

That f_X is symmetric about zero follows immediately from its construction as a mixture of symmetric–about–zero densities.

It also follows from the fact that t enters the m.g.f. only as a square, so that its c.f. is real, i.e.,

$$\varphi_X(t) = \mathbb{M}_X(it) = \left(1 - (it)^2 / (2c)\right)^{-b} = \left(1 + t^2 / (2c)\right)^{-b}.$$

Example 1.23 discussed the use of the inversion formulae to calculate its p.d.f. and c.d.f..

Example (Laplace generalization)

The s.p.a. is applicable: For $x \neq 0$,

$$\mathbb{K}_X'(t) = \frac{2bt}{2c - t^2}, \quad \mathbb{K}_X'(t) = x \Longrightarrow t = \frac{-b \pm \sqrt{b^2 + 2cx^2}}{x},$$

and simplifying $t^2 < 2c$ with Maple easily shows that the value of t with the plus sign is the correct one. Similarly,

$$\mathbb{K}_X''(t) = 2b\frac{2c + t^2}{(2c - t^2)^2},$$

with expressions for $\mathbb{K}_X^{(3)}(t)$ and $\mathbb{K}_X^{(4)}(t)$ given in the text, for use with the 2nd order s.p.a.. An application of L'Hopital's rule shows that

$$\lim_{x \to 0} \frac{-b + \sqrt{b^2 + 2cx^2}}{x} = \lim_{x \to 0} \frac{2cx}{\sqrt{b^2 + 2cx^2}} = 0.$$

This implies that, at $x = 0$, the 1st order density s.p.a. is

$$\hat{f}_X(0) = \frac{1}{\sqrt{2\pi\, \mathbb{K}_X''(0)}} \exp\left\{\mathbb{K}_X(0)\right\} = \frac{1}{\sqrt{2\pi\, b/c}} \quad \approx 0.40 \text{ for } b = c.$$

Example (Laplace generalization)

The symmetry of f_X and the existence of the m.g.f. implies that all odd positive moments exist and are zero.

For the even moments, we use the result (proven below) that

$$(1-x)^{-n} = \sum_{j=0}^{\infty} \binom{n+j-1}{j} x^j, \quad |x| < 1. \tag{52}$$

This implies

$$\begin{aligned}
\mathbb{M}_X(t) &= \left(1 - \frac{t^2}{2c}\right)^{-b} \\
&= 1 + \binom{b}{1}\left(\frac{t^2}{2c}\right) + \binom{b+1}{2}\left(\frac{t^2}{2c}\right)^2 + \binom{b+2}{3}\left(\frac{t^2}{2c}\right)^3 + \cdots \\
&= 1 + \frac{t^2}{2!}\frac{b}{c} + \frac{3b\,(b+1)}{c^2}\frac{t^4}{4!} + \cdots .
\end{aligned}$$

Example (Laplace generalization)

Again

$$\mathbb{M}_X(t) = 1 + \frac{t^2}{2!}\frac{b}{c} + \frac{3b\,(b+1)}{c^2}\frac{t^4}{4!} + \cdots$$

so that, recalling 1.3, i.e.,

$$\mathbb{M}_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty}\frac{(tX)^k}{k!}\right] = \sum_{k=0}^{\infty}\frac{t^k}{k!}\mathbb{E}\left[X^k\right],$$

we have

$$\mu_2' = \mu_2 = \mathrm{Var}\,(X) = \frac{b}{c}, \quad \mu_4' = \mu_4 = \frac{3b\,(b+1)}{c^2}$$

and, thus,

$$\mathrm{kurt}\,(X) = \frac{\mu_4}{\mu_2^2} = 3\left(1 + \frac{1}{b}\right).$$

Example (Laplace generalization): Proof of (52)

First, recall that (Example 1.7 of Fundamental Probability) for n a positive integer,

$$\begin{aligned}\binom{-n}{k} &= \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!} \\ &= (-1)^k \frac{(n)(n+1)\cdots(n+k-1)}{k!} \\ &= (-1)^k \binom{n+k-1}{k}. \qquad (53)\end{aligned}$$

For $n = 1$, the result is just $(-1)^k$.

Example (Laplace generalization): Proof of (52)

To prove (52), (Example 1.9 of Fundamental Probability), let $f(x) = (1-x)^t$, $t \in \mathbb{R}$. With

$$f'(x) = -t(1-x)^{t-1}, \quad f''(x) = t(t-1)(1-x)^{t-2}$$

and, in general, $f^{(j)}(x) = (-1)^j t_{[j]} (1-x)^{t-j}$, the Taylor series expansion of $f(x)$ around zero is given by

$$\sum_{j=0}^{\infty} (-1)^j t_{[j]} \frac{x^j}{j!} = \sum_{j=0}^{\infty} \binom{t}{j} (-x)^j, \quad |x| < 1. \tag{54}$$

or $(1+x)^t = \sum_{j=0}^{\infty} \binom{t}{j} x^j$, $|x| < 1$.

For $t = -1$, (54) and (53) yield the familiar $(1-x)^{-1} = \sum_{j=0}^{\infty} x^j$, while for $t = -n$, they imply (52), i.e.,

$$(1-x)^{-n} = \sum_{j=0}^{\infty} \binom{-n}{j} (-x)^j = \sum_{j=0}^{\infty} \binom{n+j-1}{j} x^j, \quad |x| < 1.$$

Example (Laplace generalization)

A closed-form expression for the density of X is possible. The *modified Bessel function of the third kind* is denoted $K_v(z)$, with real argument $z > 0$ and $v \in \mathbb{R}$, and can be expressed as

$$K_v(z) = \frac{1}{2}\int_0^\infty u^{v-1} \exp\left(-\frac{z}{2}\left(\frac{1}{u}+u\right)\right) \mathrm{d}u.$$

It can be shown that $K_v(z) = K_{-v}(z)$, and that

$$\lim_{z \downarrow 0} K_v(z) = \Gamma(v) 2^{v-1} z^{-v}, \quad v > 0. \tag{55}$$

In Matlab, $K_v(z)$ is best computed with the built–in function `besselk(v,z)`.

Example (Laplace generalization)

Using the p.d.f. expression for the mixture,

$$f_X(x) = \int_{\mathcal{C}} f_{X|C}(x \mid c) \, f_C(c) \, \mathrm{d}c,$$

we get, for $x > 0$, with $\theta = ux/\sqrt{2c}$,

$$f_X(x) = \frac{c^b}{\sqrt{2\pi}\Gamma(b)} \int_0^\infty \theta^{b-3/2} \exp\left\{-c\theta - \frac{x^2}{2\theta}\right\} \mathrm{d}\theta \tag{56}$$

$$= \frac{2c^b}{\sqrt{2\pi}\Gamma(b)} \left(\frac{x}{\sqrt{2c}}\right)^{b-1/2} \frac{1}{2} \int_0^\infty u^{b-3/2} \exp\left\{-\frac{x\sqrt{2c}}{2}\left(u + \frac{1}{u}\right)\right\} \mathrm{d}u$$

$$= \frac{2c^b}{\sqrt{2\pi}\Gamma(b)} \left(\frac{x}{\sqrt{2c}}\right)^{b-1/2} K_{b-1/2}\left(x\sqrt{2c}\right), \quad x > 0, \tag{57}$$

as first given by Teichroew (1957).

Example (Laplace generalization)

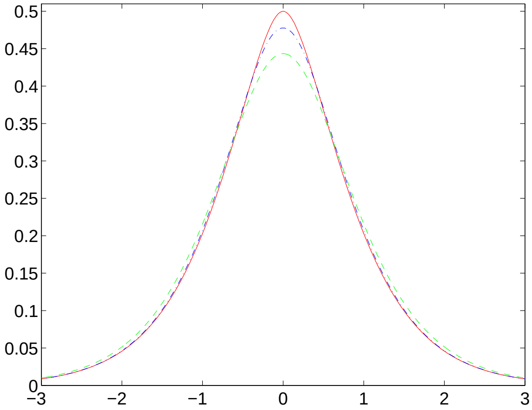
An expression for $x = 0$ is easily obtained from (56) for $b > 1/2$.

Otherwise, use of the relation $K_v(z) = K_{-v}(z)$ and (55) show that $\lim_{x \downarrow 0} f_X(x) = \infty$.

Recall that the s.p.a. is $\hat{f}_X(0) = \frac{1}{\sqrt{2\pi\, b/c}}$. It follows that, as b decreases to $1/2$, and as x approaches zero, the s.p.a. will worsen, as the next figures show.

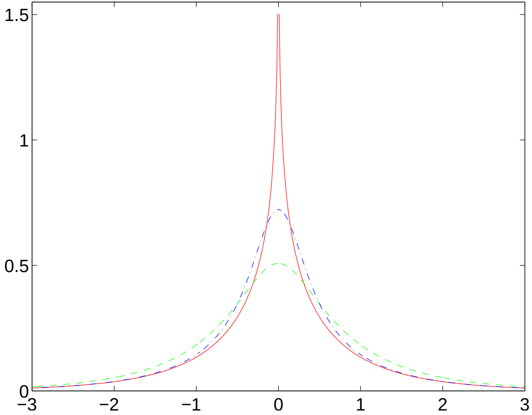
Example (Laplace generalization)

True (solid), 1st order renormalized s.p.a. (dashed) and 2nd order renormalized s.p.a. (dash-dot) for $b = c = 2$



Example (Laplace generalization)

True (solid), 1st order renormalized s.p.a. (dashed) and 2nd order renormalized s.p.a. (dash-dot) for $b = c = 0.5$



Example (Variance–gamma)

A natural way of introducing asymmetry into the previously studied normal–gamma mixture is to take

$$(X \mid \Theta = \theta) \sim \mathrm{N}(m\theta, \theta), \quad \Theta \sim \mathrm{Gam}(b, c), \quad m \in \mathbb{R}.$$

The m.g.f. simplifies to

$$\mathbb{M}_X(t) = \left(1 - \frac{mt}{c} - \frac{t^2}{2c}\right)^{-b}.$$

Example (Variance–gamma)

We can then show that

$$\mu = \mathbb{E}[X] = \frac{bm}{c}, \quad \mu_2 = \mathrm{Var}(X) = \frac{b\left(c+m^2\right)}{c^2},$$

$$\mathrm{Skew}(X) = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3' - 3\mu_2'\mu + 2\mu^3}{\mu_2^{3/2}} = \frac{m\left(3c+2m^2\right)}{b^{1/2}\left(c+m^2\right)^{3/2}},$$

and $\mathrm{Kurt}(X) = \frac{\mu_4}{\mu_2^2}$ given by

$$\frac{\mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4}{\mu_2^2} = \frac{3}{b}\frac{(b+1)\,c^2 + m^2\left(2c+m^2\right)(b+2)}{\left(c+m^2\right)^2}.$$

Is it true now that $\mathrm{Kurt}(X) > 3$?

Example (Variance–gamma)

Observe that, as $b > 0$,

$$\begin{aligned} \mathrm{Kurt}(X) &= \frac{3}{b}\frac{(b+1)\,c^2 + m^2\left(2c+m^2\right)(b+2)}{\left(c+m^2\right)^2} \\ &> 3\frac{c^2 + m^2\left(2c+m^2\right)}{\left(c+m^2\right)^2} = 3, \end{aligned}$$

so that, as in the $m = 0$ case, $\mathrm{Kurt}(X) > 3$.

Example (Variance–gamma)

The derivation of the density is straightforward and yields, for $x > 0$,

$$f_X(x; m, b, c) = \frac{2c^b e^{xm}}{\sqrt{2\pi}\Gamma(b)} \left(\frac{x}{\sqrt{m^2 + 2c}} \right)^{b-1/2} K_{b-1/2}\left(x\sqrt{m^2 + 2c} \right).$$

For $x < 0$ we use the result

$$f_X(x; m, b, c) = f_X(-x; -m, b, c).$$

Random variable X is said to follow the *variance–gamma* distribution, and was popularized by its use for modeling financial returns data by Madan and Seneta (1990).

It is a special case of the generalized hyperbolic distribution which we study in detail in Chapter 9.

Example (Noncentral Student's t)

Let $(X \mid V = v) \sim \mathrm{N}\left(\mu v^{-1/2}, v^{-1}\right)$ and $kV \sim \chi^2_k$. Then $V \sim \mathrm{Gam}(k/2, k/2)$, i.e.,

$$f_V(v; k) = \frac{k^{k/2}}{2^{k/2}\Gamma(k/2)} v^{k/2-1} \mathrm{e}^{-vk/2} \mathbb{I}_{(0,\infty)}(v)$$

and $f_X(x; \mu) = \int_{-\infty}^{\infty} f_{X|V}(x \mid v) f_V(v; k)\, \mathrm{d}v$ simplifies to

$$f_X(x; \mu) = \frac{1}{\sqrt{2\pi}} \frac{k^{k/2}}{2^{k/2-1}\Gamma(k/2)} \int_0^{\infty} z^k \exp\left\{-\frac{1}{2}\left((xz-\mu)^2 + kz^2\right)\right\} \mathrm{d}z.$$

Random variable X is said to follow the so–called *(singly) noncentral t* distribution with k degrees of freedom and *noncentrality parameter* μ.

Chapter 10 will derive alternative expressions for the p.d.f. and c.d.f..

Example (Noncentral Student's t)

Imagine how a realization x from the unconditional distribution of X would be generated: As $kV \sim \chi^2_k$, let v be drawn from a χ^2_k r.v., then divided by k.

Conditionally, $(X \mid V = v) \sim \mathrm{N}\left(\mu v^{-1/2}, v^{-1}\right)$, so we would set

$$x = \frac{1}{v^{1/2}}\left(\mu + z\right),$$

where z is a realization from a standard normal distribution.

We can write this procedure in terms of the r.v.s $kV \sim \chi^2_k$ and $Z \sim \mathrm{N}(0,1)$ as

$$X = \frac{\mu + Z}{\sqrt{V}},$$

and note that V is independent of Z.

This is precisely how a (singly) noncentral Student's t random variable is constructed, and observe that when $\mu = 0$, this reduces to the usual Student's t distribution.

Example (Noncentral Student's t)

It is straightforward to show (see the text)

$$\mathbb{E}[X] = \int_0^\infty \mu (X \mid V = v) f_V(v) \, \mathrm{d}v = \mu \sqrt{\frac{k}{2}} \frac{\Gamma\left(\frac{k-1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}.$$

Similarly, with $\mu_2'(X \mid V = v) = v^{-1}\left(1 + \mu^2\right)$,

$$\mathbb{E}\left[X^2\right] = \int_0^\infty \mu_2'(X \mid V = v) f_V(v) \, \mathrm{d}v = \left(1 + \mu^2\right) \frac{k}{k-2},$$

from which the variance can be computed.

Intermediate Probability: A Computational Approach

Introduction

Many of the problems faced by the Hill and related estimators of the tail decay parameter α can be overcome if one is prepared to adopt a more parametric model and assume, for example, stable innovations.
Robert J. Adler (1997) in "Discussion: Heavy Tail Modeling and Teletraffic Data", The Annals of Statistics, 25(5):1849–1852

The stable distribution exhibits heavy–tails and possible skewness, and so makes it a useful candidate for modeling a variety of data which exhibit such characteristics.

However, there is another property of the stable which separates its use from that of other, *ad hoc* distributions which also possess fat tails and skewness. This is the result of the *Generalized Central Limit Theorem*.

Finally, it is closed under summation, making it useful in contexts when sums of random variables are required.

Symmetric Stable

- From Examples 1.21 and 1.22, by algebraically inverting the c.f. $\varphi_Z(t) = \mathrm{e}^{-t^2/2}$, we get that $Z \sim \mathrm{N}(0,1)$. Similarly, inverting $\varphi_X(t) = \mathrm{e}^{-c|t|}$ for $c > 0$ shows that $X \sim \mathrm{Cau}(c)$. The two are easily nested as

$$\varphi_X(t;\alpha) = \exp\left\{-|t|^{\alpha}\right\}, \quad 0 < \alpha \leq 2.$$

- For $\alpha = 2$, $\varphi_X(t;\alpha) = \exp\left\{-t^2\right\}$, which is the same as $\exp\left\{-t^2\sigma^2/2\right\}$ with $\sigma^2 = 2$, i.e., as $\alpha \to 2$, X approaches a $\mathrm{N}(0,2)$ distribution.
- With $0 < \alpha \leq 2$, this defines a class of distributions called the **symmetric stable Paretian distribution with tail index** α, or, in short, just the α-stable distribution.

Symmetric Stable

- Location and scale parameters are incorporated as usual by setting $Y = cX + \mu$, $c > 0$. Then $\varphi_Y(t) = \exp\left(i\mu t - c^{\alpha}|t|^{\alpha}\right)$ and $f_Y(y; \alpha, \mu, c) = c^{-1} f_X((y-\mu)/c)$.
- We write $Y \sim S_{\alpha}(\mu, c)$. The notation S$\alpha$S ('symmetric alpha stable') is also popular.
- It can be shown that only values of α such that $0 < \alpha \leq 2$ give rise to a valid c.f..
- For $0 < \alpha \leq 2$, it can be shown that the p.d.f. is unimodal (this also holds for the asymmetric case discussed below).
- As $\varphi_X(t; \alpha)$ is real, f_X must be symmetric about zero.

Density

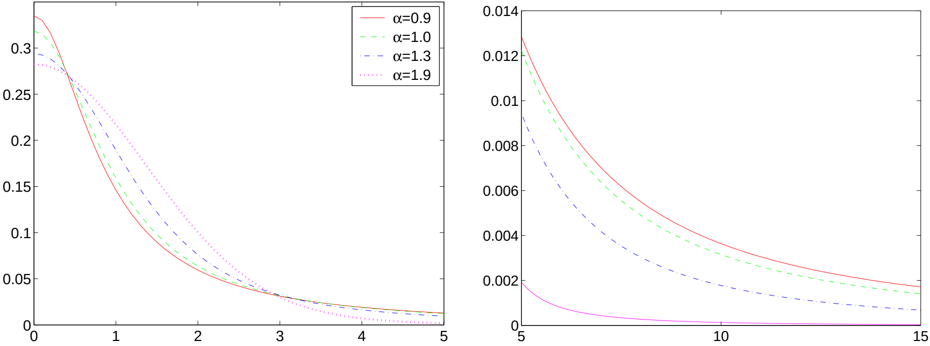


Figure: Density functions of symmetric stable Paretian r.v.s

Symmetric Stable: Density Computation

Applying the p.d.f. inversion formula to the c.f. $\varphi_X(t;\alpha) = \exp\{-|t|^{\alpha}\}$ gives

$$
\begin{aligned}
2\pi f_X(x) &= \int_{-\infty}^{\infty} \mathrm{e}^{-ixt}\mathrm{e}^{-|t|^{\alpha}}\,\mathrm{d}t \\
&= \int_{-\infty}^{0} \cos(tx)\,\mathrm{e}^{-(-t)^{\alpha}}\,\mathrm{d}t - i\int_{-\infty}^{0} \sin(tx)\,\mathrm{e}^{-(-t)^{\alpha}}\,\mathrm{d}t \\
&\quad + \int_{0}^{\infty} \cos(tx)\,\mathrm{e}^{-t^{\alpha}}\,\mathrm{d}t - i\int_{0}^{\infty} \sin(tx)\,\mathrm{e}^{-t^{\alpha}}\,\mathrm{d}t
\end{aligned}
$$

or

$$
f_X(x) = \frac{1}{\pi}\int_{0}^{\infty} \cos(tx)\,\mathrm{e}^{-t^{\alpha}}\,\mathrm{d}t. \tag{58}
$$

Symmetric Stable: Density Computation

From (58) and the definition of cosine as $\cos(z) = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k}}{(2k)!}$,

$$f_X(x;\alpha) = \frac{1}{\pi}\int_0^\infty \cos(tx)\,\mathrm{e}^{-t^\alpha}\,\mathrm{d}t = \frac{1}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k x^{2k}}{(2k)!}\int_0^\infty t^{2k}\mathrm{e}^{-t^\alpha}\,\mathrm{d}t$$

and with $u = t^\alpha$ ($t = u^{1/\alpha}$, $\mathrm{d}t = (1/\alpha)\,u^{1/\alpha-1}\mathrm{d}u$),

$$\begin{aligned} f_X(x;\alpha) &= \frac{1}{\pi\alpha}\sum_{k=0}^{\infty}\frac{(-1)^k x^{2k}}{(2k)!}\int_0^\infty u^{(2k+1)/\alpha\;-1}\,\mathrm{e}^{-u}\,\mathrm{d}u \\ &= \frac{1}{\pi\alpha}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k)!}\Gamma\left(\frac{2k+1}{\alpha}\right)x^{2k}. \end{aligned} \quad (59)$$

In particular, $f_X(0;\alpha) = \Gamma(1/\alpha)/(\pi\alpha)$ and, it can be shown that, for $1 < \alpha < 2$, the series converges for all x; and diverges for all $x \neq 0$ when $0 < \alpha < 1$.

Clearly, use of (59) will be problematic as $|x|$ grows.

Symmetric Stable: Density Computation

Expression (59) can be evaluated with a truncated sum - the number of terms, say T, required to obtain 10 digit accuracy, is an increasing function of $|x|$ and α. For $\alpha = 1.5$ and $0 \leq x \leq 4$, a conservative value is $T = \mathrm{round}(1 + 15x)$. Using this, the series method is extremely fast: about 100 times faster than numeric integration of (58).

For $|x| > 4$, we use (58).

For computing a large number of p.d.f. values (especially for $|x| > 4$), use of the FFT-based method can be much faster.

Symmetric Stable: Density Computation

Numeric calculation of the density with (58) is unfortunately not as simple as it appears. As α moves from two towards zero, and as $|x|$ increases, the integrand in the inversion formula

$$f_X(x) = \frac{1}{\pi} \int_0^1 h_x \left(\frac{1-u}{u} \right) u^{-2} \, \mathrm{d}u, \qquad h_x(t) = \mathrm{Re}\left[\mathrm{e}^{-itx} \varphi_X(t) \right]$$

begins to exhibit increasingly problematic oscillatory behavior.

Thus, in the worst case, the tail of the density for small values of α is practically impossible to numerically determine. The use of the FFT to obtain the density will also suffer accordingly.

Fortunately, for most applied work, interest centers on values of α larger than one, and observations (x–values) 'excessively' far into the tail do not occur. Moreover, the simple limiting approximation for the p.d.f., $f_X(x) \approx \alpha k(\alpha) x^{-\alpha-1}$, becomes increasingly accurate as one moves further into the tails.

Symmetric Stable: Density Computation

From Zolotarev (1986, Equation 2.2.18), for $\alpha \neq 1$ and $x \neq 0$,

$$f_S(x;\alpha) = \frac{\alpha}{\pi\,|\alpha-1|\,|x|}\int_0^{\pi/2} g(t;\alpha,x)\exp\{-g(t;\alpha,x)\}\,\mathrm{d}t, \qquad (60)$$

for

$$g(t;\alpha,x) = \left(\frac{x\cos t}{\sin(\alpha t)}\right)^{\frac{\alpha}{\alpha-1}} \frac{\cos(\alpha-1)t}{\cos t}.$$

There are (also) some numeric problems associated with this expression. For example, unsurprisingly, for x very close to zero (more so for $\alpha \leq 1.2$), the method fails. Calculation shows that it also fails as $|x|$ grows; already for $|x| > 9$ there are problems in a Matlab (quadl) implementation.[6]

[6] For further details, see Matsui and Takemura (2004), Some Improvements in Numerical Evaluation of Symmetric Stable Density and its Derivatives, Working paper No. CIRJE-F-292, University of Tokyo.

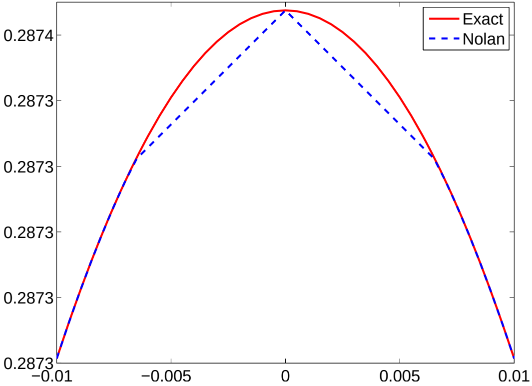
Symmetric Stable: Computation

Perhaps the best solution is to buy the software provided by John Nolan, which highly accurate and is far faster than Matlab's numeric integration and also the FFT approach.

However, no one is perfect: The series approximation (for $|x| < 4$) is still (slightly) faster than Nolan's "exact" routine `stablepdf`. And, very close to zero, Nolan apparently uses linear interpolation—due presumably to use of (60). The figure below shows the true p.d.f. and Nolan's values for $\alpha = 1.5$.

It should be mentioned that Nolan's "Quick stable" routine `stableqkpdf` (using a spline approximation) is extraordinarily fast (far more so than any of the exact methods so far discussed), and (because it uses splines) does not have the problem for x very close to zero.

Symmetric Stable: Nolan Near Zero



C.D.F. Calculation

The c.d.f. can be calculated via the Gil-Peleaz inversion formula. See Program 8.1 in the text for the p.d.f. and c.d.f. calculation.

A specific integral expression (due to Zolotarev) not involving complex numbers is as follows. For $\alpha > 1$, let

$$J(x;\alpha) = \int_0^{1/2} \exp\left\{ -\left| \frac{x}{t_\alpha(y)} \right|^{\alpha/(\alpha-1)} \right\} \mathrm{d}y,$$

where

$$t_\alpha(y) = \frac{\sin(\pi\alpha y)}{\cos(\pi y)} \left(\frac{\cos(\pi y)}{\cos(\pi(\alpha-1)y)} \right)^{(\alpha-1)/\alpha}.$$

The c.d.f. of $X \sim S_\alpha$ for $x < 0$ is $F_X(x;\alpha) = J(x;\alpha)$. As usual, from the symmetry of f_X, $F_X(x;\alpha) = 1 - F_X(-x;\alpha)$ for all x.

Observe that $J(x;\alpha) = J(|x|;\alpha)$ so in this case, for $x > 0$, $F_X(x;\alpha) = 1 - J(x;\alpha)$.

C.D.F. Calculation

If many values of the c.d.f. are to be computed, the FFT output for the equally-spaced grid of p.d.f. values, $f(x_i)$, can be used.

That is,

$$F_X(x_i) \approx \begin{cases} \frac{N-1}{N\widetilde{F}(x_{n-1})}\widetilde{F}(x_i) & \text{if} \quad 1 \le i < N, \\ 1, & \text{if} \quad i = N, \end{cases}$$

where $\widetilde{F}(x_i)$ are computed from the trapezoidal rule (use the `cumsum` command in Matlab for speed) as

$$\widetilde{F}(x_i) = \widetilde{F}(x_{i-1}) + \frac{s}{2}\left(f(x_i) + f(x_{i+1})\right), \quad i = 1, \ldots, N-1,$$

$\widetilde{F}(x_0) = 0$ and s is the grid width of the x_i.

The c.d.f. at the desired points is then approximated by linearly interpolating between the $F_X(x_i)$.

C.D.F. Calculation

The method using the FFT will be much faster when many points are desired, but less accurate, especially as α decreases towards one, and as $|x|$ increases.

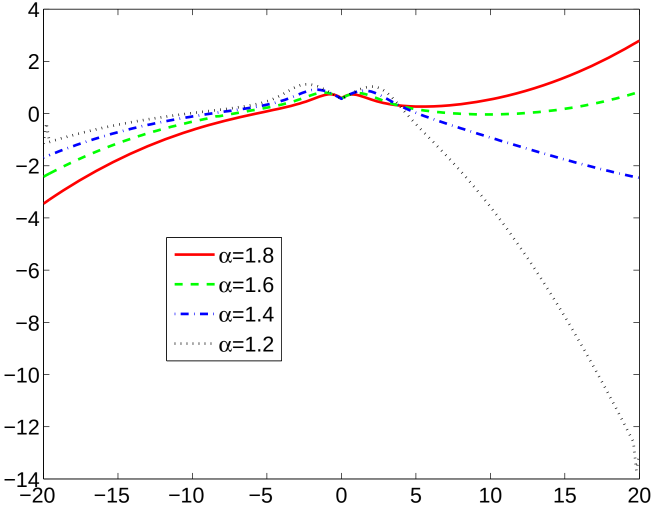
To assess its accuracy, we plot its relative percentage error, computed as

$$\text{r.p.e.} = 100 \frac{\text{approx} - \text{exact}}{\min(\text{exact}, 1 - \text{exact})}$$

and "approx" is the above method (via use of the FFT, the trapezoidal rule, and finally linear interpolation), and "exact" refers to the computation using either the Zolotarev expression or Gil-Peleaz inversion formula.

The best option for overall speed and accuracy is to use Nolan's "exact" (`stablecdf`) or quick (`stableqkcdf`) routines. For the grid of 40,001 values from 0 to 20 in increments of 0.0005, Nolan's exact routine takes my PC 11.1 seconds; quick takes 0.016 seconds; and the FFT takes 17 seconds.

C.D.F. Calculation: Accuracy of FFT Approximation



Symmetric Stable: Properties

- One of the most important properties of the stable distribution is summability (or **stability**): With $X_i \overset{\text{ind}}{\sim} S_\alpha(\mu_i, c_i)$ and $S = \sum_{i=1}^n X_i$,

$$\varphi_S(t) = \prod_{i=1}^n \varphi_{X_i}(t) = \exp\left(i\mu_1 t - c_1^\alpha |t|^\alpha\right) \cdots \exp\left(i\mu_n t - c_n^\alpha |t|^\alpha\right)$$
$$= \exp\left(i\mu t - c^\alpha |t|^\alpha\right), \qquad \text{i.e.,}$$

$$S \sim S_\alpha(\mu, c), \quad \text{where} \quad \mu = \sum_{t=1}^n \mu_i \quad \text{and} \quad c = \left(c_1^\alpha + \cdots + c_n^\alpha\right)^{1/\alpha}.$$

Symmetric Stable: Properties

- The word 'Paretian' in the name reflects the fact that the asymptotic tail behavior of the S_α distribution is the same as that of the Pareto distribution, i.e., S_α has power tails, for $0 < \alpha < 2$. It can be shown that, for $X \sim S_\alpha(0,1)$, $0 < \alpha < 2$, as $x \to \infty$,

$$\bar{F}_X(x) = \Pr(X > x) \approx k(\alpha)\, x^{-\alpha}, \quad k(\alpha) = \pi^{-1} \sin(\pi\alpha/2)\, \Gamma(\alpha),$$

  and $a \approx b$ means that a/b converges to one as $x \to \infty$.
- Informally, differentiating the limiting value of $1 - \bar{F}_X(x)$ gives the asymptotic density in the right tail,

$$f_X(x) \approx \alpha k(\alpha)\, x^{-\alpha-1}.$$

  Expressions for the asymptotic left–tail behavior follows from the symmetry of the density about zero, i.e., $f_X(x) = f_X(-x)$ and $F_X(x) = \bar{F}_X(-x)$.

Symmetric Stable: Properties

- The term Paretian goes back to Mandelbrot (1960), who pointed out that Vilfredo Pareto had suggested the use of distributions with power tails at the end of the nineteenth century (far before the stable class was characterized) for modeling a variety of economic phenomena.
- From the moments of the usual Pareto distribution, it follows that, for $0 < \alpha < 2$, the (fractional absolute) moments of $X \sim S_\alpha$ of order α and higher do not exist, i.e., $\mathbb{E}\left[|X|^r\right]$ is finite for $r < \alpha$, and infinite otherwise. For $-1 < r < \alpha$, we have

$$\mathbb{E}\left[\,|X|^r\,\right] = \frac{1}{\kappa}\Gamma\left(1 - \frac{r}{\alpha}\right), \quad \kappa = \begin{cases} \Gamma(1-r)\cos(\pi r/2), & \text{if} \quad r \neq 1, \\ \pi/2, & \text{if} \quad r = 1. \end{cases}$$

Comment on Tail Behavior

The fact that moments of $X \sim S_\alpha$ of order α and higher do not exist for $0 < \alpha < 2$, while for $\alpha = 2$, all positive moments exist, implies quite different tail behavior than the Student's t distribution, which also nests the Cauchy and, as the degrees of freedom parameter $v \to \infty$, the normal.

For the t, moments of order v and higher do not exist, and as $v \to \infty$, the tail behavior gradually moves from power tails to exponential tails. For the stable distribution, there is a 'knife–edge' change from $\alpha < 2$ (power tails) to $\alpha = 2$ (normal, with exponential tails).

The transition is still smooth however, in the sense that the p.d.f. of $X \sim S_\alpha$ for $\alpha = 2 - \epsilon$, $\epsilon > 0$, can be made arbitrarily close to that of the normal p.d.f.. This is just a consequence of the continuity of the c.f. and the fact that, like parameter $v \in \mathbb{R}_{>0}$, the set of numbers $[2 - \epsilon, 2]$ is uncountably infinite.

Expected Shortfall: Definition and Properties

- Recall that the expected shortfall is defined as

$$\mathrm{ES}_\theta(R) = \mathbb{E}\left[R \mid R \le q_{R,\theta}\right] = \frac{1}{\theta}\int_{-\infty}^{q_{R,\theta}} r\, f_R(r)\,\mathrm{d}r,$$

  where R is a future period financial return and $q_{R,\theta}$ is the θ-quantile such that $\Pr(R \le q_{R,\theta}) = \theta$ and θ is small, typically 1%.

- Recall also that, if Z is a location zero, scale one r.v., and $Y = \sigma Z + \mu$ for $\sigma > 0$, then

$$\mathrm{ES}_\theta(Y) = \mu + \sigma \mathrm{ES}_\theta(Z).$$

Expected Shortfall: Symmetric Stable

Stoyanov, Samorodnitsky, Rachev and Ortobelli (2006)[7] give an integral expression not explicitly involving the stable density to compute the ES of the (symmetric and asymmetric) stable distribution.

For $\alpha > 1$, $S \sim S_\alpha(0,1)$, $0 < \gamma < 1$ and $q_{S,\gamma} = F_S^{-1}(\gamma)$,

$$\mathrm{ES}_\gamma(S; \alpha, 0, 1) = \frac{1}{\gamma} \mathrm{Stoy}(q_{S,\gamma}, \alpha),$$

where the tail component $\int_{-\infty}^{c} x f_S(x;\alpha) \mathrm{d}x$ is

$$\mathrm{Stoy}(c,\alpha) = \frac{\alpha}{\alpha-1} \frac{|c|}{\pi} \int_0^{\pi/2} g(\theta) \exp\left(-|c|^{\alpha/(\alpha-1)} v(\theta)\right) \mathrm{d}\theta,$$

and

$$g(\theta) = \frac{\sin(\alpha-2)\theta}{\sin\alpha\theta} - \frac{\alpha\cos^2\theta}{\sin^2\alpha\theta}, \qquad v(\theta) = \left(\frac{\cos\theta}{\sin\alpha\theta}\right)^{\alpha/(\alpha-1)} \frac{\cos(\alpha-1)\theta}{\cos\theta}.$$

[7] Computing the Portfolio Conditional Value-at-Risk in the alpha-Stable Case, *Probability and Mathematical Statistics*, 26:1–22.

Expected Shortfall: Symmetric Stable

To confirm that $\mathrm{Stoy}(c,\alpha)$ works, we can perform the integration $I(\gamma,\alpha)=\int_{-\infty}^{q_{S,\gamma}} x f_S(x;\alpha)\mathrm{d}x$, but we know that calculation of f_S for $|x|$ very large can be problematic.

Instead, use the asymptotic tail behavior of $S \sim S_{\alpha,0}(0,1)$; as $x \to -\infty$, $f_S(x;\alpha) \approx K_\alpha(-x)^{-\alpha-1}$, where $K_\alpha := \alpha\pi^{-1}\sin(\pi\alpha/2)\,\Gamma(\alpha)$ and $a \approx b$ means that a/b converges to one in the limit.

Then, for some cutoff value ℓ, the integral can be approximated by

$$\begin{aligned} I(\gamma,\alpha) &\approx K_\alpha \int_{-\infty}^{\ell} x\,(-x)^{-\alpha-1}\,\mathrm{d}x + \int_{\ell}^{q_{S,\gamma}} x f_S(x;\alpha)\,\mathrm{d}x \\ &= K_\alpha \frac{(-\ell)^{1-\alpha}}{1-\alpha} + \int_{\ell}^{q_{S,\gamma}} x f_S(x;\alpha)\,\mathrm{d}x. \end{aligned}$$

Some trial and error shows that a value of $\ell = -120$ results in very good performance.

Mixtures of Symmetric Stable

Recall the flexibility of the normal mixture, which can be skewed, multimodal, and also replicate fat-tailed behavior over most of the support where observations lie.

A mixture of symmetric stable distributions will have these features as well, but might require less components to capture the fat-tails, and the tails will be genuinely Pareto.

We denote such a distribution as $\mathrm{MixStab}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$.

Mixtures of Symmetric Stable

For $X \sim \mathrm{MixStab}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$, the p.d.f. at real value x is given by

$$f_X(x; \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) = \sum_{j=1}^{k} \lambda_j f_S(x; \alpha_j, \mu_j, \sigma_j),$$

where:

- $f_S(x; \alpha, \mu, c)$ is the location-μ, scale-c, symmetric stable Paretian density function with tail index α,
- $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$ is the set of tail indices corresponding to the k symmetric stable distributional components,
- $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)'$ is the set of nonnegative weights which sum to one,
- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ is the set of component location parameters,
- $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_k) \in \mathbb{R}^k_{>0}$ is the set of scale parameters.

If, as in most applications in finance, all the $\alpha_i > 1$, then the μ_i are the individual means, and $\mathbb{E}[X] = \sum_{i=1}^{k} \lambda_i \mu_i$.

Mixtures of Symmetric Stable: C.D.F. and Quantiles

Similar to the mixed normal case, the c.d.f. of $X \sim \mathrm{MixStab}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$ is

$$F_X(x) = \Pr\left(X \leq x\right) = \sum_{j=1}^{k} \lambda_j F_S\left(\frac{x - \mu_j}{\sigma_j}; \alpha_j, 0, 1\right),$$

where $F_S\left(\cdot; \alpha, 0, 1\right)$ is the c.d.f. of the standard symmetric stable distribution, evaluated using any of the methods discussed above.

The γ-quantile of X, $q_{X,\gamma}$, can be found by numerically solving $F_X(q_{X,\gamma}) - \gamma = 0$.

Mixtures of Symmetric Stable: Expected Shortfall

Let $X \sim \mathrm{MixStab}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$ with $q_{X,\gamma}$ the γ-quantile of X for some $\gamma \in (0, 1)$.

Let $X_j \sim S_{\alpha_j}(\mu_j, \sigma_j)$ be the j th component in the mixture with density $f_{X_j}(\cdot; \alpha_j, \mu_j, \sigma_j)$ and let

$$q_j := q_{X_j,\gamma} = F_{X_j}^{-1}(\gamma; \alpha_j, \mu_j, \sigma_j) = \mu_j + \sigma_j F_S^{-1}(\gamma; \alpha_j, 0, 1)$$

be the γ-quantile of X_j, with $S \sim S_\alpha(0, 1)$.

Mixtures of Symmetric Stable: Expected Shortfall

The same simple calculations used for the 1st way of ES decomposition for the normal mixture give, with $S \sim S_\alpha(0,1)$, an expression for

$$\mathrm{ES}_\gamma(X; \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) = \frac{1}{\gamma} \int_{-\infty}^{q_{X,\gamma}} x\, f_X(x)\ \mathrm{d}x$$

as

$$\frac{1}{\gamma} \sum_{j=1}^{k} \lambda_j \left[\sigma_j \int_{-\infty}^{\frac{q_{X,\gamma}-\mu_j}{\sigma_j}} z f_S(z; \alpha_j)\ \mathrm{d}z + \mu_j \int_{-\infty}^{\frac{q_{X,\gamma}-\mu_j}{\sigma_j}} f_S(z; \alpha_j)\ \mathrm{d}z \right],$$

Mixtures of Symmetric Stable: Expected Shortfall

or, with $c_j := (q_{X,\gamma} - \mu_j) / \sigma_j$,

$$\mathrm{ES}_\gamma (X; \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) = \frac{1}{\gamma} \sum_{j=1}^{k} \lambda_j [\sigma_j \, \mathrm{Stoy}(c_j, \alpha_j) + \mu_j F_S (c_j; \alpha_j)],$$

which is easily numerically calculated because both the Stoyanov et al. integral and the Zolotarev integral expression for the stable c.d.f. do not involve integration into the tail of the density.

Mixtures of Symmetric Stable: Expected Shortfall

As in the normal mixture case, we can write

$$\mathrm{ES}_\gamma\left(X;\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\lambda}\right)=\sum_{j=1}^{k}\omega_j\,\mathrm{ES}_\gamma\left(X_j;\alpha\right)$$

for

$$\omega_j=\frac{\lambda_j}{\gamma}\frac{\sigma_j\,\mathrm{Stoy}\left(c_j,\alpha_j\right)+\mu_j F_S\left(c_j;\alpha_j\right)}{\mathrm{ES}_\gamma\left(X_j\right)},$$

and interpret $\omega_j^*=\omega_j/\sum_{j=1}^{k}\omega_j$ as the fraction of the ES attributed to component j.

Mixtures of Symmetric Stable: Expected Shortfall

Also, similar to the 2nd way of ES decomposition for the normal mixture, the expected shortfall of X is

$$\begin{aligned}
\mathrm{ES}_\gamma\left(X;\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\lambda}\right) &= \frac{1}{\gamma}\int_{-\infty}^{q_{X,\gamma}} x\ f_X\left(x\right)\ \mathrm{d}x \\
&= \sum_{j=1}^{k}\lambda_j\,\mathrm{ES}_\gamma\left(X_j;\alpha_j,\mu_j,\sigma_j\right)+\frac{1}{\gamma}\sum_{j=1}^{k}\lambda_j g_j\left(\gamma\right),
\end{aligned}$$

where

$$g_j\left(\gamma\right):=\int_{q_j}^{q_{X,\gamma}} x\ f_{X_j}\left(x;\alpha_j,\mu_j,\sigma_j\right)\mathrm{d}x$$

and

$$\mathrm{ES}_\gamma\left(X_j;\alpha_j,\mu_j,\sigma_j\right)=\mu_j+\sigma_j\,\mathrm{ES}_\gamma\left(S;\alpha_j,0,1\right)=\mu_j+\sigma_j\frac{1}{\gamma}\,\mathrm{Stoy}\left(q_j,\alpha_j\right).$$

Mixtures of Symmetric Stable: Expected Shortfall

The following code empirically verifies the above ES expression, computed via program `stablemixtureES.m` (homework!) and 5 million replications using program `stabgen.m` (see below).

```
gama=0.01; alpha=[1.5 1.6 1.7]; mu=[-3 0 0];
scale=[1 2 1]; lambda=[0.4 0.3 0.3];
[ES, q]=stablemixtureES(gama,alpha,mu,scale,lambda)

N=5000000; k=length(alpha); pick=zeros(N,k); X=[];
for i=1:N, c=randmultinomial(lambda); pick(i,c)=1; end
for c=1:k
  add = STABGEN(N,alpha(c))'; X=[X  mu(c) + scale(c) * add];
end
X=pick.*X; X=sum(X')'; clear pick
use=X(X<q); proportion = length(use)/N
empiricalES_and_trueES = [mean(use) ES]
```

The output is: `-24.1694 -23.9985`

Asymmetric Stable

A skewness parameter β, $-1 \leq \beta \leq 1$, can be introduced, to get the **asymmetric stable Paretian distribution**, with tail index α and skewness parameter β. We write $X \sim S_{\alpha,\beta}(\mu, c)$.

Its c.f. is given by

$$\ln \varphi_X(t) = \begin{cases} -c^\alpha |t|^\alpha \left[1 - i\beta \operatorname{sgn}(t) \tan \frac{\pi\alpha}{2}\right] + i\mu t, & \text{if } \alpha \neq 1, \\ -c|t| \left[1 + i\beta \frac{2}{\pi} \operatorname{sgn}(t) \ln |t|\right] + i\mu t, & \text{if } \alpha = 1. \end{cases}$$

The form of the c.f. does not lend itself to much intuition. It arises from characterizing the class of distributions for which summability holds, and is quite an involved derivation.

As $\alpha \to 2$, the effect of β diminishes because $\tan(\pi\alpha/2) \to 0$; and when $\alpha = 2$, $\tan(\pi) = 0$, and β has no effect. Thus, there is no 'skewed normal' distribution within the stable family (though there is a skewed Cauchy).

Asymmetric Stable: P.D.F. Calculation

Instead of using the complex-integral inversion formula, one can use the real integral expression from Zolotarev for computing the p.d.f. of the asymmetric stable distribution: For $x > 0$,

$$f_X(x;\alpha,\beta) = \frac{\alpha x^{1/(\alpha-1)}}{\pi|\alpha-1|}\int_{-\theta_0}^{\pi/2} v(\theta;\alpha,\beta)\exp\left\{-x^{\alpha/(\alpha-1)}v(x;\alpha,\beta)\right\}\, d\theta,$$

where $\theta_0 = \arctan(\beta\tan(\pi\alpha/2))/\alpha$, and $v(\theta;\alpha,\beta)$ is

$$(\cos(\alpha\theta_0))^{1/(\alpha-1)}\left(\frac{\cos(\theta)}{\sin(\alpha(\theta_0+\theta))}\right)^{\alpha/(\alpha-1)}\frac{\cos(\alpha(\theta_0+\theta)-\theta)}{\cos(\theta)}.$$

Asymmetric Stable: P.D.F. Calculation

For $x < 0$, use the fact that $f_X(x; \alpha, \beta) = f_X(-x; \alpha, -\beta)$, and for $x = 0$,

$$f_X(0; \alpha, \beta) = \frac{\Gamma(1 + 1/\alpha)\cos(\theta_0)}{\pi\left(1 + (\beta \tan(\pi\alpha/2))^2\right)^{1/(2\alpha)}}.$$

Obviously, around a neighborhood of zero, there will be potential problems.

Some very limited trial and error for $\alpha = 1.7$ and $\beta = -0.7$ (depends on x, α and β, the numeric integration routine used, and the specified accuracy tolerance) suggests that this integral is more difficult to compute than the inversion formula, is lower (and more variable) in accuracy, and can fail for large x.

The numeric integration associated with the inversion formula yields nearly identical results to the software provided by John Nolan.

Asymmetric Stable: Expected Shortfall

Stoyanov et al. (2006) also provide the tail component $\int_{-\infty}^{c} x f_S(x; \alpha, \beta)\,\mathrm{d}x$ in this case. Similar to the symmetric case, for $\alpha > 1$, $S \sim S_{\alpha,\beta}(0,1)$, $0 < \gamma < 1$ and $q_{S,\gamma} = F_S^{-1}(\gamma)$,

$$\mathrm{ES}_\gamma(S; \alpha, \beta, 0, 1) = \frac{1}{\gamma} \mathrm{Stoy}(q_{S,\gamma}, \alpha, \beta),$$

where the tail component $\int_{-\infty}^{c} x f_S(x; \alpha, \beta)\,\mathrm{d}x$ is

$$\mathrm{Stoy}(c, \alpha, \beta) = \frac{\alpha}{\alpha - 1} \frac{|c|}{\pi} \int_{-\bar{\theta}_0}^{\pi/2} g(\theta) \exp\left(-|c|^{\alpha/(\alpha-1)} v(\theta)\right) \mathrm{d}\theta,$$

with ...

Asymmetric Stable: Expected Shortfall

$$g(\theta) = \frac{\sin\left(\alpha\left(\bar{\theta}_0 + \theta\right) - 2\theta\right)}{\sin\left(\alpha\left(\bar{\theta}_0 + \theta\right)\right)} - \frac{\alpha \cos^2(\theta)}{\sin^2\left(\alpha\left(\bar{\theta}_0 + \theta\right)\right)},$$

$$v(\theta) = \left(\cos\left(\alpha\bar{\theta}_0\right)\right)^{1/(\alpha-1)} \left(\frac{\cos(\theta)}{\sin\left(\alpha\left(\bar{\theta}_0 + \theta\right)\right)}\right)^{\alpha/(\alpha-1)} \frac{\cos\left(\alpha\left(\bar{\theta}_0 + \theta\right) - \theta\right)}{\cos(\theta)},$$

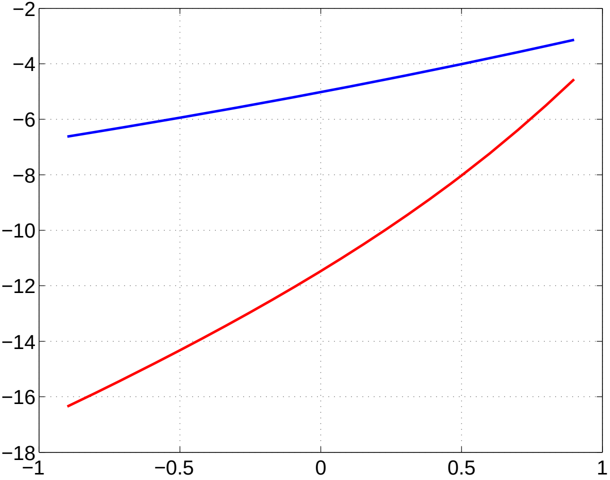
and[8]

$$\bar{\theta}_0 = \frac{1}{\alpha} \arctan\left(\bar{\beta} \tan\left(\frac{\pi\alpha}{2}\right)\right) \quad \text{and} \quad \bar{\beta} = \operatorname{sign}(c)\,\beta.$$

[8] The formulae for $\bar{\beta}$ and $\mathrm{Stoy}(c, \alpha, \beta)$ have a minus sign in front in Stoyanov et al. (2006) because they use the positive sign convention for the VaR and ES.

Asymmetric Stable: Expected Shortfall

ES for $\gamma = 0.01$ and $\gamma = 0.05$ for $\alpha = 1.7$ as a function of β.



Asymmetric Stable: Summability

- As in the symmetric case, summability means that the sum of independent stable r.v.s, **each with the same tail index** α, also follows a stable distribution. In particular,

$$\text{if} \quad X_i \overset{\text{ind}}{\sim} S_{\alpha,\beta_i}\left(\mu_i, c_i\right), \quad \text{then} \quad S = \sum_{i=1}^{n} X_i \sim S_{\alpha,\beta}\left(\mu, c\right),$$

$$\mu = \sum_{t=1}^{n} \mu_i, \quad c = \left(c_1^{\alpha} + \cdots + c_n^{\alpha}\right)^{1/\alpha} \quad \text{and} \quad \beta = \frac{\beta_1 c_1^{\alpha} + \cdots + \beta_n c_n^{\alpha}}{c_1^{\alpha} + \cdots + c_n^{\alpha}}.$$

- The class of stable distributions is the only one which has the property of summability (or stability).

Asymmetric Stable

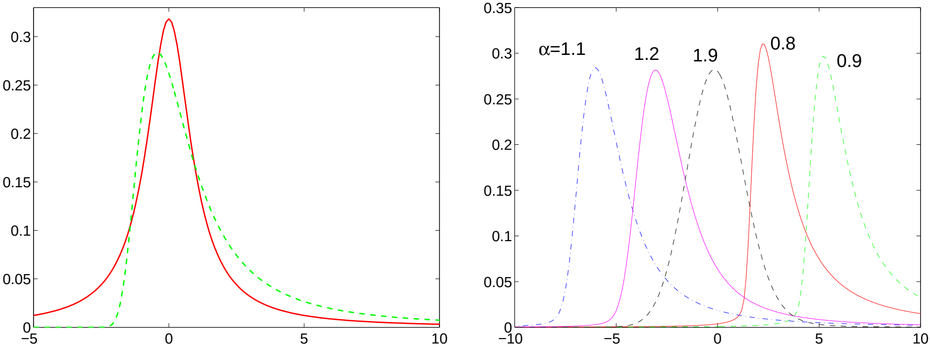


Figure: Cauchy (solid) and 'totally skewed Cauchy' (dashed) computed by inverting the c.f. with $\beta = 1$ (left). Densities of $S_{\alpha,0.9}(0,1)$ for various α (right)

Asymmetric Stable

- For positive β, as $\alpha \uparrow 1$, the mode drifts to $+\infty$, while for $\alpha \downarrow 1$, the mode drifts to $-\infty$. At $\alpha = 1$, the mode is near zero (and depends on β).
- There are other parameterizations of the $S_{\alpha,\beta}$ c.f. which are more advantageous depending on the purpose; the so–called Zolotarev's (M) parameterization avoids the 'erratic mode' problem. Nolan's book offers a detailed discussion of the various parameterizations and their respective advantages.

Asymmetric Stable

- Similar tail behavior for the asymmetric case, but parameter β dictates the 'relative heights' of the two tails: For $X \sim S_{\alpha,\beta}(0,1)$, as $x \to \infty$,

$$\bar{F}_X(x) = \Pr(X > x) \approx k(\alpha)(1+\beta)\, x^{-\alpha}$$

and

$$f_X(x) \approx \alpha k(\alpha)(1+\beta)\, x^{-\alpha-1}.$$

- For the left tail, the symmetry relations $f_X(x;\alpha,\beta) = f_X(-x;\alpha,-\beta)$ and $F_X(x;\alpha,\beta) = \bar{F}_X(-x;\alpha,-\beta)$ can be used.

Asymmetric Stable

- From the asymptotic tail behavior, we can ascribe a particular meaning to parameter β. Let $\alpha < 2$ and $X \sim S_{\alpha,\beta}(0,1)$. Then, informally, as $x \to \infty$,

$$\frac{P(X > x) - P(X < -x)}{P(X > x) + P(X < -x)} \approx \frac{k(\alpha)(1+\beta)x^{-\alpha} - k(\alpha)(1-\beta)x^{-\alpha}}{k(\alpha)(1+\beta)x^{-\alpha} + k(\alpha)(1-\beta)x^{-\alpha}} = \frac{2k(\alpha)\beta x^{-\alpha}}{2k(\alpha)x^{-\alpha}} = \beta,$$

  i.e., β measures the asymptotic difference in the two tail masses, scaled by the sum of the two tail areas.

- With $\tau = \tan\theta$ and $\theta = \arctan\{\beta \tan(\pi\alpha/2)\}$, for $-1 < r < \alpha$

$$\mathbb{E}\left[\,|X|^r\,\right] = \kappa^{-1}\Gamma\left(1 - \frac{r}{\alpha}\right)\left(1+\tau^2\right)^{r/2\alpha}\cos\left(\frac{r}{\alpha}\arctan\tau\right),$$

  where

$$\kappa = \begin{cases} \Gamma(1-r)\cos(\pi r/2), & \text{if} \quad r \neq 1, \\ \pi/2, & \text{if} \quad r = 1. \end{cases}$$

Simulation

A method for simulating stable Paretian r.v.s was developed in Chambers, Mallows and Stuck (1976): Let $U \sim \mathrm{Unif}\,(-\pi/2, \pi/2)$ independent of $E \sim \mathrm{Exp}\,(1)$. Then $Z \sim S_{\alpha,\beta}\,(0,1)$, where

$$Z = \begin{cases} \dfrac{\sin\alpha\,(\theta + U)}{(\cos\alpha\theta\ \cos U)^{1/\alpha}} \left[\dfrac{\cos\left(\alpha\theta + (\alpha-1)\,U\right)}{E}\right]^{(1-\alpha)/\alpha}, & \text{if} \quad \alpha \neq 1, \\ \dfrac{2}{\pi}\left[\left(\dfrac{\pi}{2} + \beta U\right)\tan U - \beta \ln\left(\dfrac{(\pi/2)\,E\cos U}{(\pi/2) + \beta U}\right)\right], & \text{if} \quad \alpha = 1, \end{cases}$$

and $\theta = \arctan\left(\beta \tan\left(\pi\alpha/2\right)\right)/\alpha$ for $\alpha \neq 1$.

Incorporation of location and scale changes is done as usual when $\alpha \neq 1$, i.e., $X = \mu + cZ \sim S_{\alpha,\beta}\,(\mu, c)$, but for $\alpha = 1$, $X = \mu + cZ + (2\beta c \ln c)/\pi \sim S_{1,\beta}\,(\mu, c)$ (note that the latter term is zero for $\beta = 0$ or $c = 1$).

Simulation

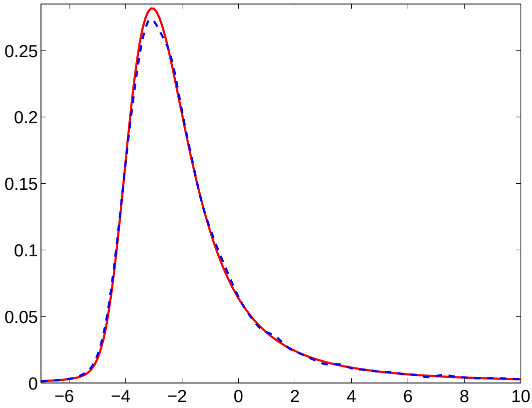


Figure: True p.d.f. (solid) and kernel density estimate (dashed) (based on 30,000 draws) of the $S_{1.2,0.9}(0,1)$ distribution

Generalized Central Limit Theorem

- Let $X_1, X_2, \ldots$ be an i.i.d. sequence of r.v.s (with or without finite means and variances). The generalized CLT, or GCLT, states that there exist real constants $a_n > 0$ and b_n, and a non-degenerate r.v. Z such that, as $n \to \infty$,

$$a_n \sum_{j=1}^{n} X_j - b_n \xrightarrow{d} Z,$$

  iff $Z \sim S_{\alpha,\beta}(\mu, c)$. In the standard CLT, $a_n = n^{-1/2}$. In the GCLT, a_n is of the form $n^{-1/\alpha}$, which nests the $\alpha = 2$ case.
- Recall that the normal distribution is justified in statistical and econometric applications as the appropriate error distribution, with the reasoning that the discrepancy between the model and the observations are the cumulative result of many independent random factors (too numerous and/or difficult to capture in the model).

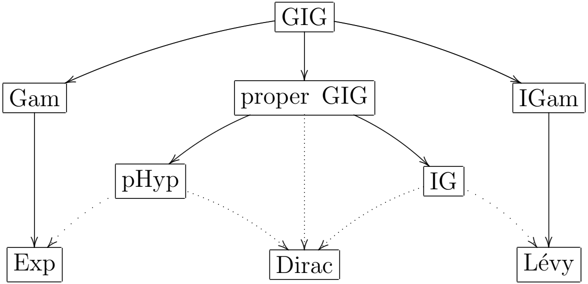
Generalized Central Limit Theorem

- The GCLT states that the asymmetric stable Paretian distribution (with special case the normal) is the only possible candidate for modeling the error term.
- Many data sets exist, from a variety of disciplines, which exhibit much fatter tails than the normal, and sometimes considerable skewness as well. The most notable example is in the field of finance, but also the motion picture business, telecommunications, biology and geology.

Intermediate Probability: A Computational Approach

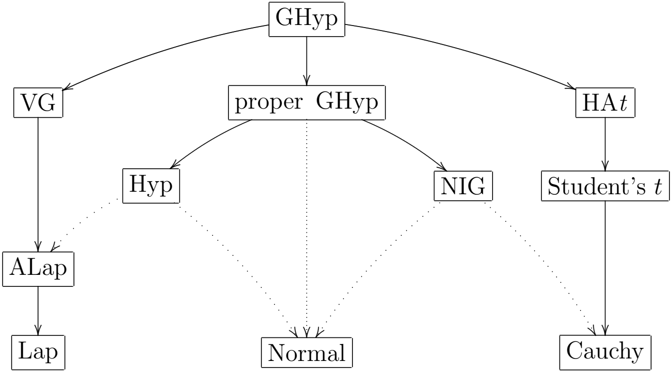
Introduction to GIG and GHyp

The generalized inverse Gaussian distributions, or GIG, is a very flexible distribution which nests some well-known distributions.



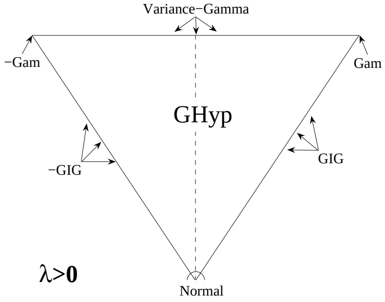
Introduction to GIG and GHyp

The generalized hyperbolic distributions, or GHyp, arise as mean-variance mixtures of normals, using the GIG as mixing weights.



Shape Triangle of the Generalized Hyperbolic

The GHyp distributions are fat-tailed and skewed, but with a finite variance, and also possess other interesting and desirable properties.



The Bessel Function: Definition and Computation

To understand and work with these distributions, we need a Bessel function.

- For every $\nu \in \mathbb{R}$, the *modified Bessel function of the third kind*, hereafter simply called the Bessel function, with index ν is defined as

$$K_\nu(x) := \frac{1}{2}\int_0^\infty t^{\nu-1} \mathrm{e}^{-\frac{1}{2}x(t+t^{-1})}\,\mathrm{d}t, \quad x > 0.$$

- It is related to the gamma function, and, in general, numerical methods are necessary for its evaluation.
- In Matlab, $K_v(x)$ is computed with the built–in function `besselk(v,x)`.

The Bessel Function: Identities

Of great use are the following:

$$K_\nu(x) = K_{-\nu}(x), \qquad \nu \in \mathbb{R},\ x \in \mathbb{R}_{>0},$$

as easily shown (in the text).

The Bessel function can be stated explicitly for $\nu = 1/2$, with

$$K_{1/2}(x) = K_{-1/2}(x) = \sqrt{\frac{\pi}{2x}}\mathrm{e}^{-x}, \qquad x \in \mathbb{R}_{>0}.$$

The Bessel Function: Identities

A recursive formula is given by

$$K_{\nu+1}(x) = \frac{2\nu}{x} K_\nu(x) + K_{\nu-1}(x), \quad \nu \in \mathbb{R},\ x \in \mathbb{R}_{>0}.$$

See the text for derivation, and Problem 9.2.

For $x = 0$, the Bessel function has a singularity. For $\nu = 0$,

$$K_0(x) \simeq -\ln(x) \quad \text{for} \quad x \downarrow 0,\ \nu = 0.$$

For $\nu \neq 0$,

$$K_\nu(x) \simeq \Gamma(|\nu|) 2^{|\nu|-1} x^{-|\nu|} \quad \text{for} \quad x \downarrow 0,\ \nu \neq 0.$$

Important Integral

Throughout this chapter, we will frequently meet the following integral which is closely related to the Bessel function:

$$\mathrm{k}_\lambda(\chi, \psi) := \int_0^\infty x^{\lambda-1} \exp\left[-\frac{1}{2}(\chi x^{-1} + \psi x)\right] \mathrm{d}x.$$

It converges for arbitrary $\lambda \in \mathbb{R}$ and $\chi, \psi > 0$.

Using the notation $\eta := \sqrt{\chi/\psi}$ and $\omega := \sqrt{\chi\psi}$, we show in the text that

$$\mathrm{k}_\lambda\left(\chi,\, \psi\right) = 2\eta^\lambda K_\lambda(\omega) = 2\left(\frac{\chi}{\psi}\right)^{\lambda/2} K_\lambda\left(\sqrt{\chi\psi}\right)$$

Important Integral

The integral also converges in two 'boundary cases': If $\chi = 0$ and $\psi > 0$, it converges if and only if (iff) $\lambda > 0$. On the other hand, if $\chi > 0$ and $\psi = 0$, it converges iff $\lambda < 0$. We show in the text:

$$\mathrm{k}_\lambda(0, \psi) = \left(\frac{\psi}{2}\right)^{-\lambda} \Gamma(\lambda), \quad \text{and} \quad \mathrm{k}_\lambda(\chi, 0) = \left(\frac{\chi}{2}\right)^{\lambda} \Gamma(-\lambda).$$

Function $\mathrm{k}_\lambda\,(\chi,\,\psi)$ possesses some symmetries, namely

$$\mathrm{k}_\lambda(\chi, \psi) = \mathrm{k}_{-\lambda}(\psi, \chi)$$

and, for all $r > 0$,

$$\mathrm{k}_\lambda(\chi, \psi) = r^\lambda \mathrm{k}_\lambda(r^{-1}\chi, r\psi).$$

The Generalized Inverse Gaussian Distribution

The generalized inverse Gaussian p.d.f. is given by

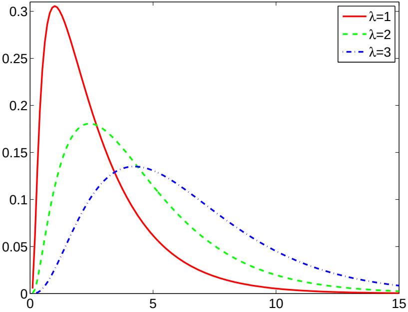
$$f_{\mathrm{GIG}}(x;\lambda,\chi,\psi) = \frac{1}{\mathrm{k}_\lambda(\chi,\psi)} x^{\lambda-1} \exp\left[-\frac{1}{2}(\chi x^{-1} + \psi x)\right] \mathbb{I}_{(0,\infty)}(x).$$

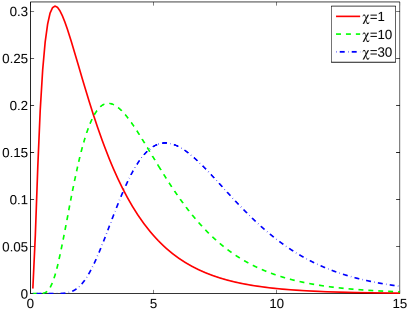
We denote the parameter space by Θ_{GIG}, which consists of three cases given by

$$\begin{aligned}\Theta_{\mathrm{GIG}} := \Big\{(\lambda,\chi,\psi)\in\mathbb{R}^3 : \quad & \lambda\in\mathbb{R},\ \chi>0,\ \psi>0 && \textbf{(normal case),}\\ \text{or} \quad & \lambda>0,\ \chi=0,\ \psi>0 && \textbf{(boundary case I),}\\ \text{or} \quad & \lambda<0,\ \chi>0,\ \psi=0\} && \textbf{(boundary case II).}\end{aligned}$$

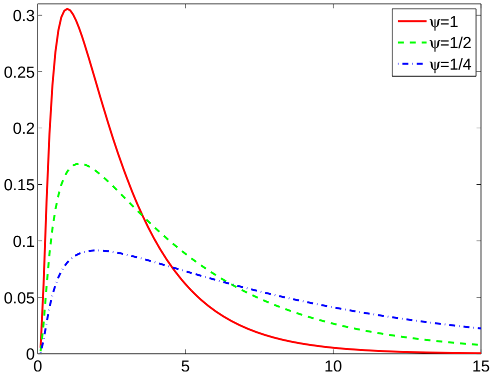
Note that the GIG reduces to the gamma distribution in boundary case I.

GIG Density: Non-specified parameters = 1





GIG Density: Non-specified parameters = 1



GIG Moments: General Case

Assume that $(\lambda, \chi, \psi) \in \Theta_{\text{GIG}}$ and $X \sim \text{GIG}(\lambda, \chi, \psi)$. Then the rth raw moment of X is

$$\begin{aligned}
\mathbb{E}[X^r] &= \int_0^\infty x^r f_{\text{GIG}}(x; \lambda, \chi, \psi)\, \mathrm{d}x \\
&= \frac{1}{\mathrm{k}_\lambda(\chi, \psi)} \int_0^\infty x^{\lambda+r-1} \exp\left[-\frac{1}{2}(\chi x^{-1} + \psi x)\right] \mathrm{d}x \\
&= \frac{\mathrm{k}_{\lambda+r}(\chi, \psi)}{\mathrm{k}_\lambda(\chi, \psi)}.
\end{aligned}$$

The moment generating function of X is

$$\begin{aligned}
\mathbb{M}_{\text{GIG}}(t; \lambda, \chi, \psi) &= \int_0^\infty \mathrm{e}^{tx} f_{\text{GIG}}(x; \lambda, \chi, \psi)\, \mathrm{d}x \\
&= \frac{1}{\mathrm{k}_\lambda(\chi, \psi)} \int_0^\infty x^{\lambda-1} \exp\left[-\frac{1}{2}(\chi x^{-1} + (\psi - 2t)x)\right] \mathrm{d}x \\
&= \frac{\mathrm{k}_\lambda(\chi, \psi - 2t)}{\mathrm{k}_\lambda(\chi, \psi)}.
\end{aligned}$$

GIG Simulation

A method for simulating a GIG random variable was given by Dagpunar (1989).[9]

There is an implementation in R (package `HyperbolicDist`) by David Scott, Richard Trendall and Melanie Luen, and one in Matlab (function `randraw`) by Alex Bar Guy and Alexander Podgaetsky. Both are well-documented and available in internet.

[9] Dagpunar, J. S. (1989). An Easily Implemented Generalised Inverse Gaussian Generator, Commun. Statist. -Simula., 18(2), 703-710. See also Dagpunar, J. S. (1988) Principles of Random Variate Generation, Clarendon Press, Oxford.

Proper GIG [if $\lambda \in \mathbb{R}$, $\chi > 0$, $\psi > 0$]

With $\eta := \sqrt{\chi/\psi}$ and $\omega := \sqrt{\chi\psi}$, for the proper GIG, we get

$$f_{\mathrm{GIG}}(x; \lambda, \chi, \psi) = \frac{1}{2\eta K_\lambda(\omega)} \left(\frac{x}{\eta}\right)^{\lambda-1} \exp\left[-\frac{1}{2}\omega\left((x/\eta)^{-1} + x/\eta\right)\right] \mathbb{I}_{(0,\infty)}(x),$$

showing that $\eta = \sqrt{\chi/\psi}$ is a scale parameter.

For $r \in \mathbb{R}$, the rth raw moment of $X \sim \mathrm{GIG}(\lambda, \chi, \psi)$ is

$$\mathbb{E}[X^r] = \int_0^\infty x^r f_{\mathrm{GIG}}(x; \lambda, \chi, \psi)\, \mathrm{d}x = \eta^r \frac{K_{\lambda+r}(\omega)}{K_\lambda(\omega)},$$

and

$$\mathbb{M}_{\mathrm{GIG}}(t; \lambda, \chi, \psi) = \frac{K_\lambda\left(\omega\sqrt{1-2t/\psi}\right)}{K_\lambda(\omega)\,(1-2t/\psi)^{\lambda/2}}, \quad -\infty < t < \psi/2.$$

Inverse Gaussian IG [if $\lambda = -1/2$, $\chi > 0$, $\psi > 0$]

If $\lambda = -1/2$, then the GIG distribution in the normal case reduces to the inverse Gaussian distribution,

$$\mathrm{IG}_1(\chi, \psi) := \mathrm{GIG}(-1/2, \chi, \psi), \quad \chi, \psi > 0,$$

with p.d.f. simplifying to

$$f_{\mathrm{IG}_1}(x; \chi, \psi) = \left(\frac{\chi}{2\pi}\right)^{1/2} \mathrm{e}^{\sqrt{\chi\psi}} x^{-3/2} \exp\left[-\frac{1}{2}(\chi x^{-1} + \psi x)\right] \mathbb{I}_{(0,\infty)}(x).$$

The m.g.f. simplifies to

$$\mathbb{M}_{\mathrm{IG}_1}(t, \chi, \psi) = \exp\left[\sqrt{\chi}\left(\sqrt{\psi} - \sqrt{\psi - 2t}\right)\right], \quad -\infty < t \leq \psi/2.$$

Sums of Independent IG

Let $X_i \stackrel{\text{ind}}{\sim} \mathrm{IG}_1(\chi_i, \psi)$. The m.g.f. of $S = X_1 + X_2 + \cdots + X_n$ is then

$$
\begin{aligned}
& \prod_{i=1}^{n} \exp\left[\sqrt{\chi_i}\left(\sqrt{\psi} - \sqrt{\psi - 2t}\right)\right] \\
= \; & \exp\left[\left(\sum_{i=1}^{n} \sqrt{\chi_i}\right)\left(\sqrt{\psi} - \sqrt{\psi - 2t}\right)\right] \\
= \; & \exp\left[\sqrt{\chi}\left(\sqrt{\psi} - \sqrt{\psi - 2t}\right)\right],
\end{aligned}
$$

for all $t \in (-\infty, \psi/2)$, where $\chi = \left(\chi_1^{1/2} + \chi_2^{1/2} + \cdots + \chi_n^{1/2}\right)^2$.

Thus, $S \sim \mathrm{IG}_1(\chi, \psi)$.

Reciprocal Inverse Gaussian: Exercise

Let $X \sim \text{GIG}(\lambda, \chi, \psi)$ and define $Y = 1/X$. Show that $Y \sim \text{GIG}(-\lambda, \psi, \chi)$.

Now let $X \sim \mathrm{IG}_1(\chi, \psi) = \text{GIG}(-1/2, \chi, \psi)$. Then $Y = 1/X \sim \text{GIG}(1/2, \psi, \chi)$ and Y is referred to as a *reciprocal inverse Gaussian*, written $Y \sim \mathrm{RIG}(\psi, \chi)$.

Compute the mean and variance of Y.

Derive the m.g.f. of Y.

Finally, show the convolution result

$$\mathrm{IG}_1(\chi_1, \psi) \star \mathrm{RIG}(\chi_2, \psi) = \mathrm{RIG}(\chi, \psi), \quad \chi = \left(\sqrt{\chi_1} + \sqrt{\chi_2}\right)^2.$$

Reciprocal Inverse Gaussian: Solution

Let $X \sim \mathsf{GIG}(\lambda, \chi, \psi)$ with

$$f_X(x; \lambda, \chi, \psi) = \frac{1}{\mathrm{k}_\lambda(\chi, \psi)} x^{\lambda - 1} \exp\left[-\frac{1}{2}(\chi x^{-1} + \psi x)\right] \mathbb{I}_{(0,\infty)}(x)$$

and define $Y = 1/X$. Recalling (9.11), i.e., $\mathrm{k}_\lambda(\chi, \psi) = \mathrm{k}_{-\lambda}(\psi, \chi)$, we have $\mathrm{d}x/\mathrm{d}y = -y^{-2}$ and

$$\begin{aligned} f_Y(y) &= f_X(x) \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = y^{-2} \frac{1}{\mathrm{k}_\lambda(\chi, \psi)} \left(\frac{1}{y}\right)^{\lambda - 1} \exp\left[-\frac{1}{2}(\chi y + \psi y^{-1})\right] \mathbb{I}_{(0,\infty)}(y) \\ &= \frac{1}{\mathrm{k}_{-\lambda}(\psi, \chi)} y^{-\lambda - 1} \exp\left\{-\frac{1}{2}\left(\psi y^{-1} + \chi y\right)\right\} \mathbb{I}_{(0,\infty)}(y), \end{aligned}$$

so that $Y \sim \mathsf{GIG}(-\lambda, \psi, \chi)$.

Now let $X \sim \mathrm{IG}_1(\chi, \psi) = \mathsf{GIG}(-1/2, \chi, \psi)$. Then $Y = 1/X \sim \mathsf{GIG}(1/2, \psi, \chi)$ and Y is referred to as a *reciprocal inverse Gaussian*, written $Y \sim \mathrm{RIG}(\psi, \chi)$, with density

$$f_Y(y; \psi, \chi) = \frac{1}{\mathrm{k}_{1/2}(\psi, \chi)} y^{-1/2} \exp\left\{-\frac{1}{2}\left(\psi y^{-1} + \chi y\right)\right\} \mathbb{I}_{(0,\infty)}(y),$$

where, from (9.11) and the simplification of $\mathrm{k}_{-1/2}(\chi, \psi)$ shown on page 313,

$$\frac{1}{\mathrm{k}_{1/2}(\psi, \chi)} = \frac{1}{\mathrm{k}_{-1/2}(\chi, \psi)} = \left(\frac{\chi}{2\pi}\right)^{1/2} e^{\sqrt{\chi\psi}}.$$

Reciprocal Inverse Gaussian: Solution

Problem 9.2 shows that

$$K_{3/2}(x) = \left(\frac{1}{x} + 1\right)\sqrt{\frac{\pi}{2x}}e^{-x}, \quad K_{5/2}(x) = \left(\frac{3}{x^2} + \frac{3}{x} + 1\right)\sqrt{\frac{\pi}{2x}}e^{-x}.$$

Using this, (9.8) and (9.3),

$$\mathbb{E}[Y] = \frac{\mathrm{k}_{3/2}(\psi,\chi)}{\mathrm{k}_{1/2}(\psi,\chi)} = \frac{2\left(\frac{\psi}{\chi}\right)^{3/4} K_{3/2}\left(\sqrt{\psi\chi}\right)}{2\left(\frac{\psi}{\chi}\right)^{1/4} K_{1/2}\left(\sqrt{\psi\chi}\right)} = \left(\frac{\psi}{\chi}\right)^{1/2}\left(\frac{1}{\sqrt{\psi\chi}} + 1\right) = \frac{1}{\chi} + \sqrt{\frac{\psi}{\chi}}.$$

Similarly,

$$\mathbb{E}\left[Y^2\right] = \frac{\mathrm{k}_{5/2}(\psi,\chi)}{\mathrm{k}_{1/2}(\psi,\chi)} = \left(\frac{\psi}{\chi}\right)\left(\frac{3}{\psi\chi} + \frac{3}{\sqrt{\psi\chi}} + 1\right)$$

so that

$$\mathbb{V}(Y) = \frac{1}{\chi}\left(\frac{2}{\chi} + \sqrt{\frac{\psi}{\chi}}\right).$$

Reciprocal Inverse Gaussian: Solution

From (9.21), (9.8) and (9.3), $\mathbb{M}_{\mathrm{RIG}}(t;\psi,\chi)=\mathbb{M}_{\mathrm{GIG}}(t;1/2,\psi,\chi)$ is

$$\begin{aligned}\frac{\mathrm{k}_{1/2}(\psi,\chi-2t)}{\mathrm{k}_{1/2}(\psi,\chi)} &= \frac{2\left(\psi/(\chi-2t)\right)^{1/4}}{2\left(\psi/\chi\right)^{1/4}}\frac{K_{1/2}\left(\sqrt{\psi(\chi-2t)}\right)}{K_{1/2}\left(\sqrt{\psi\chi}\right)} \\ &= \left(\frac{\chi}{\chi-2t}\right)^{1/4}\frac{\sqrt{\frac{\pi}{2\sqrt{\psi(\chi-2t)}}}e^{-\sqrt{\psi(\chi-2t)}}}{\sqrt{\frac{\pi}{2\sqrt{\psi\chi}}}e^{-\sqrt{\psi\chi}}} \\ &= \left(\frac{\chi}{\chi-2t}\right)^{1/2}\exp\left(\sqrt{\psi}\left(\sqrt{\chi}-\sqrt{\chi-2t}\right)\right).\end{aligned}$$

Then, with

$$\mathbb{M}_{\mathrm{IG}_1}(t;\chi_1,\psi)=\exp\left(\sqrt{\chi_1}\left(\sqrt{\psi}-\sqrt{\psi-2t}\right)\right)$$

(as derived on page 314) and

$$\mathbb{M}_{\mathrm{RIG}}(t;\chi_2,\psi)=\left(\frac{\psi}{\psi-2t}\right)^{1/2}\exp\left(\sqrt{\chi_2}\left(\sqrt{\psi}-\sqrt{\psi-2t}\right)\right),$$

inspection of $\mathbb{M}_{\mathrm{RIG}}(t;\chi_2,\psi)\,\mathbb{M}_{\mathrm{IG}_1}(t;\chi_1,\psi)$ shows that

$$\mathrm{IG}_1(\chi_1,\psi)\star\mathrm{RIG}(\chi_2,\psi)=\mathrm{RIG}(\chi,\psi),\quad \chi=\left(\sqrt{\chi_1}+\sqrt{\chi_2}\right)^2.$$

Normal Mixtures

- We consider a *variance-mean-mixture of normals*.
- Suppose Z is a continuous, positive random variable with p.d.f. f_Z, and μ and β are constants. If X is a random variable satisfying

$$X \mid Z \sim \mathrm{N}(\mu + \beta Z, Z),$$

we can calculate the density function f_X of X as follows:

$$f_X(x) = \int_0^\infty f_{\mathrm{N}}(x; \mu + \beta z, z) f_Z(z)\,\mathrm{d}z.$$

- We denote the distribution of X by $\mathrm{Mix}_\pi(\mu, \beta)$, where π is the distribution of Z, subsequently called the *weight*.
- If β is non-zero, the distribution will be skewed.

Example

Let $Z \sim \mathrm{Exp}(\lambda)$ and $\mu = \beta = 0$. Then $f_{\mathrm{Mix}_{\mathrm{Exp}(\lambda)}(0,0)}(x)$ is given by

$$\int_0^\infty f_{\mathrm{N}}(x;0,z) f_{\mathrm{Exp}}(z;\lambda)\,\mathrm{d}z = \int_0^\infty \frac{1}{\sqrt{2\pi z}} \exp\left[-\frac{x^2}{2z}\right] \lambda \mathrm{e}^{-\lambda z}\,\mathrm{d}z$$

$$= \frac{\lambda}{\sqrt{2\pi}} \int_0^\infty z^{-1/2} \exp\left[-\frac{1}{2}\left(x^2 z^{-1} + 2\lambda z\right)\right] \mathrm{d}z$$

$$\overset{(9.7)}{=} \frac{\lambda}{\sqrt{2\pi}} \mathrm{k}_{1/2}\left(x^2,\ 2\lambda\right)$$

$$\overset{(9.8)}{=} \frac{\lambda}{\sqrt{2\pi}} 2 \left(\frac{|x|}{\sqrt{2\lambda}}\right)^{1/2} K_{1/2}\left(\sqrt{2\lambda}|x|\right)$$

$$\overset{(9.3)}{=} \frac{\lambda}{\sqrt{2\pi}} 2 \left(\frac{|x|}{\sqrt{2\lambda}}\right)^{1/2} \sqrt{\frac{\pi}{2\sqrt{2\lambda}|x|}} \mathrm{e}^{-\sqrt{2\lambda}|x|} = \frac{\sqrt{2\lambda}}{2} \mathrm{e}^{-\sqrt{2\lambda}|x|}.$$

This is the Laplace density with scale parameter $1/\sqrt{2\lambda}$.

Moments and Generating Functions

For $Z \sim \pi$ and $X \sim \text{Mix}_\pi(\mu, \beta)$, we have

$$\mathbb{E}[X] = \mu + \beta\,\mathbb{E}[Z], \quad \mathbb{V}(X) = \mathbb{E}[Z] + \beta^2\mathbb{V}(Z),$$

third central moment

$$\mu_3(X) = 3\beta\mathbb{V}(Z) + \beta^3\mu_3(Z),$$

and m.g.f. and c.f.

$$\mathbb{M}_X(t) = e^{\mu t}\mathbb{M}_Z\left(\beta t + t^2/2\right), \quad \varphi_X(v) = e^{i\mu v}\varphi_Z\left(\beta v + iv^2/2\right).$$

The mean and variance results follow immediately from the iterated expectation and conditional variance formula. In particular, as $(X \mid Z) \sim \text{N}\left(\mu + \beta Z, Z\right)$,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Z]] = \mathbb{E}[\mu + \beta Z] = \mu + \beta\mathbb{E}[Z]$$

and

$$\mathbb{V}(X) = \mathbb{E}[\mathbb{V}(X \mid Z)] + \mathbb{V}(\mathbb{E}[X \mid Z]) = \mathbb{E}[Z] + \mathbb{V}(\mu + \beta Z) = \mathbb{E}[Z] + \beta^2\mathbb{V}(Z).$$

Direct Proof for the Mean

As in the textbook, for Z continuous, first note that, for every $z > 0$, the integral

$$\int_{-\infty}^{\infty} x f_{\mathrm{N}}(x; \mu + \beta z, z) \, \mathrm{d}x$$

is just the mean of a normal distribution with parameters $\mu + \beta z$ and z.

An application of Fubini's theorem gives

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) \, \mathrm{d}x = \int_{-\infty}^{\infty} x \int_0^{\infty} f_{\mathrm{N}}(x; \mu + \beta z, z) f_Z(z) \, \mathrm{d}z \, \mathrm{d}x \\
&= \int_0^{\infty} \int_{-\infty}^{\infty} x f_{\mathrm{N}}(x; \mu + \beta z, z) \, \mathrm{d}x f_Z(z) \, \mathrm{d}z = \int_0^{\infty} (\mu + \beta z) f_Z(z) \, \mathrm{d}z \\
&= \int_0^{\infty} \mu f_Z(z) \, \mathrm{d}z + \beta \int_0^{\infty} z f_Z(z) \, \mathrm{d}z = \mu + \beta \, \mathbb{E}[Z].
\end{aligned}$$

Proof for the MGF

For the m.g.f., use the equation

$$\int_{-\infty}^{\infty} \mathrm{e}^{tx} f_{\mathrm{N}}(x; \mu + \beta z, z) \, \mathrm{d}x = \exp\left[(\mu + \beta z)t + \frac{z}{2} t^2 \right],$$

which is just the moment generating function for a normal distribution with parameters $\mu + \beta z$ and z, where $z > 0$ is arbitrary. Thus

$$\begin{aligned} \mathbb{M}_X(t) &= \int_{-\infty}^{+\infty} \mathrm{e}^{tx} f_X(x) \, \mathrm{d}x = \int_{-\infty}^{+\infty} \mathrm{e}^{tx} \int_0^{\infty} f_{\mathrm{N}}(x; \mu + \beta z, z) f_Z(z) \, \mathrm{d}z \, \mathrm{d}x \\ &= \int_0^{\infty} \int_{-\infty}^{+\infty} \mathrm{e}^{tx} f_{\mathrm{N}}(x; \mu + \beta z, z) \, \mathrm{d}x f_Z(z) \, \mathrm{d}z \\ &= \int_0^{\infty} \exp\left[(\mu + \beta z)t + \frac{z}{2} t^2 \right] f_Z(z) \, \mathrm{d}z = \mathrm{e}^{\mu t} \mathbb{M}_Z(\beta t + t^2/2). \end{aligned}$$

The Generalized Hyperbolic Distribution

- For given parameters α, β and δ, set $\chi := \delta^2$ and $\psi := \alpha^2 - \beta^2$.
- Let $\alpha, \delta \in \mathbb{R}_{\geq 0}$, $\lambda, \mu \in \mathbb{R}$ and $\beta \in [-\alpha, \alpha]$.
- For $(\lambda, \chi, \psi) \in \Theta_{\mathrm{GIG}}$, we define the generalized hyperbolic distribution $\mathrm{GHyp}(\lambda, \alpha, \beta, \delta, \mu)$ as a $\mathrm{GIG}(\lambda, \chi, \psi)$-variance-mean-mixture of normal distributions, i.e.,

$$\mathrm{GHyp}(\lambda, \alpha, \beta, \delta, \mu) := \mathrm{Mix}_{\mathrm{GIG}(\lambda, \delta^2, \alpha^2 - \beta^2)}(\mu, \beta)$$

- The domain of variation of the parameters of the GHyp distribution is $\lambda, \mu \in \mathbb{R}$, $\alpha, \delta \geq 0$, $\beta \in [-\alpha, \alpha]$ such that the following conditions are satisfied:

$$\begin{array}{llllll} |\beta| < \alpha & \text{and} & \delta > 0 & \text{if} & \lambda \in \mathbb{R}, \\ |\beta| < \alpha & \text{and} & \delta \geq 0 & \text{if} & \lambda > 0, \\ |\beta| = \alpha & \text{and} & \delta > 0 & \text{if} & \lambda < 0. \end{array}$$

GHyp: Density

Let $\lambda, \mu \in \mathbb{R}$, $\alpha, \delta \geq 0$, $\beta \in [-\alpha, \alpha]$ such that $(\lambda, \delta^2, \alpha^2 - \beta^2) \in \Theta_{\text{GIG}}$.

Then, as shown in the text, a straightforward calculation, eased with the convenient function k, gives

$$\begin{aligned} & f_{\text{GHyp}}(x; \lambda, \alpha, \beta, \delta, \mu) = f_{\text{Mix}_{\text{GIG}(\lambda, \delta^2 \alpha^2 - \beta^2)}}(x; \mu, \beta) \\ = & \int_0^\infty f_{\text{N}}(x; \mu + \beta z, z) f_{\text{GIG}}(z; \lambda, \delta^2, \alpha^2 - \beta^2)\, \mathrm{d}z \\ = & \frac{\mathrm{k}_{\lambda - \frac{1}{2}}\left((x-\mu)^2 + \delta^2,\, \alpha^2\right)}{\sqrt{2\pi}\, \mathrm{k}_{\lambda}\left(\delta^2,\, \alpha^2 - \beta^2\right)} \mathrm{e}^{\beta(x-\mu)}. \end{aligned}$$

GHyp: Moments

- Suppose $X \sim \text{GHyp}(\lambda, \alpha, \beta, \delta, \mu)$. Let $\chi := \delta^2$ and $\psi := \alpha^2 - \beta^2$.
- For $Z \sim \text{GIG}(\lambda, \chi, \psi)$, as

$$\mathbb{E}[Z] = \frac{\mathrm{k}_{\lambda+1}(\chi, \psi)}{\mathrm{k}_{\lambda}(\chi, \psi)},$$

it follows that

$$\mathbb{E}[X] = \mu + \beta\mathbb{E}[Z] = \mu + \beta\frac{\mathrm{k}_{\lambda+1}(\chi, \psi)}{\mathrm{k}_{\lambda}(\chi, \psi)}.$$

- Similarly, use $\mathbb{V}(X) = \mathbb{E}[Z] + \beta^2\mathbb{V}(Z)$ with the result that

$$\mathbb{V}(Z) = \frac{\mathrm{k}_{\lambda}(\chi, \psi)\,\mathrm{k}_{\lambda+2}(\chi, \psi) - (\mathrm{k}_{\lambda+1}(\chi, \psi))^2}{(\mathrm{k}_{\lambda}(\chi, \psi))^2}.$$

GHyp: MGF

- From the m.g.f. of the GIG distribution,

$$\mathbb{M}_X(t) = \mathrm{e}^{\mu t}\mathbb{M}_Z\left(\beta t + t^2/2\right) = \mathrm{e}^{\mu t}\frac{\mathrm{k}_\lambda\left(\chi,\, \psi - 2(\beta t + t^2/2)\right)}{\mathrm{k}_\lambda\left(\chi,\, \psi\right)}.$$

- As

$$\psi - 2(\beta t + t^2/2) = \alpha^2 - \beta^2 - 2\beta t - t^2 = \alpha^2 - (\beta + t)^2,$$

  we have

$$\mathbb{M}_X(t) = \mathrm{e}^{\mu t}\frac{\mathrm{k}_\lambda\left(\delta^2,\, \alpha^2 - (\beta + t)^2\right)}{\mathrm{k}_\lambda\left(\delta^2,\, \alpha^2 - \beta^2\right)}.$$

- As the k function requires nonnegative arguments, solving the quadratic $\alpha^2 - \beta^2 - 2\beta t - t^2 = 0$ shows that the m.g.f. is valid for all t such that $-\alpha - \beta \leq t \leq \alpha - \beta$.

GHyp Simulation

Recall: For $X \mid Z \sim \mathrm{N}(\mu + \beta Z, Z)$, the distribution of X is $\mathrm{Mix}_\pi(\mu, \beta)$, where π is the distribution of Z.

Then, from the mixture relation

$$\mathrm{GHyp}(\lambda, \alpha, \beta, \delta, \mu) := \mathrm{Mix}_{\mathrm{GIG}(\lambda, \delta^2, \alpha^2 - \beta^2)}(\mu, \beta),$$

we immediately see that, to generate $X \sim \mathrm{GHyp}(\lambda, \alpha, \beta, \delta, \mu)$, set $\chi := \delta^2$ and $\psi := \alpha^2 - \beta^2$, let $Z \sim \mathrm{GIG}(\lambda, \chi, \psi)$ and set

$$X = \mu + \beta Z + \sqrt{Z} Y,$$

where $Y \sim \mathrm{N}(0, 1)$.

GHyp Subfamilies

- The GHyp distribution family is quite flexible, but this comes at the price of complexity.
- As GHyp is constructed as a GIG-mixture of normals, it follows that, to every subfamily of GIG, there corresponds a subfamily of GHyp, and, moreover, this GHyp subfamily inherits properties from the mixing weights used.

Proper GHyp

This takes: $\lambda \in \mathbb{R}$, $\alpha > 0$, $\beta \in (-\alpha, \alpha)$, $\delta > 0$, $\mu \in \mathbb{R}$.

It follows that $\chi := \delta^2 > 0$ and $\psi := \alpha^2 - \beta^2 > 0$, so we have a mixture of proper GIG distributions.

Simplifying the general expression (see the text) yields, with $y_x := \sqrt{\delta^2 + (x-\mu)^2}$, the p.d.f. as

$$f_{\mathrm{GHyp}}(x; \lambda, \alpha, \beta, \delta, \mu) = \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}} y_x^{\lambda - \frac{1}{2}}}{\sqrt{2\pi}\alpha^{\lambda - \frac{1}{2}} \delta^\lambda K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} K_{\lambda - \frac{1}{2}}(\alpha y_x) \, \mathrm{e}^{\beta(x-\mu)}.$$

Proper GHyp

In the normal case, we use the following abbreviations:

$$\eta := \sqrt{\chi/\psi} = \sqrt{\frac{\delta^2}{\alpha^2 - \beta^2}}$$

and

$$\omega := \sqrt{\chi\psi} = \sqrt{\delta^2\left(\alpha^2 - \beta^2\right)} = \delta\sqrt{\alpha^2 - \beta^2}.$$

Then, for $X \sim \mathrm{GHyp}(\lambda, \alpha, \beta, \delta, \mu)$ and $Z \sim \mathrm{GIG}(\lambda, \chi, \psi)$,

$$\mathbb{E}[X] = \mu + \beta\mathbb{E}[Z] = \mu + \beta\eta\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)},$$

and, similarly,

$$\mathbb{V}(X) = \mathbb{E}[Z] + \beta^2\mathbb{V}(Z) = \eta\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)} + \beta^2\eta^2\frac{K_\lambda(\omega)K_{\lambda+2}(\omega) - K_{\lambda+1}(\omega)^2}{K_\lambda(\omega)^2}.$$

Proper GHyp MGF

The m.g.f. of the GHyp distribution in the normal case can be derived from (9.16) and (9.25) or simplified from the general GHyp m.g.f. expression. We get

$$\mathbb{M}_X(t) = e^{\mu t} \frac{K_\lambda\left(\delta\sqrt{\alpha^2-(\beta+t)^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)\left(\frac{\alpha^2-(\beta+t)^2}{\alpha^2-\beta^2}\right)^{\lambda/2}}, \tag{61}$$

with convergence strip $-\alpha-\beta < t < \alpha-\beta$.

Using the abbreviation $\psi_t := \alpha^2-(\beta+t)^2$, and recalling $\chi = \delta^2 > 0$ and $\psi = \alpha^2-\beta^2 > 0$,

$$\mathbb{M}_X(t) = \mathrm{e}^{\mu t} \frac{K_\lambda\left(\sqrt{\chi\psi_t}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}\left(\frac{\psi}{\psi_t}\right)^{\lambda/2}.$$

Proper GHyp Density for $\delta = 1$ and $\mu = 0$

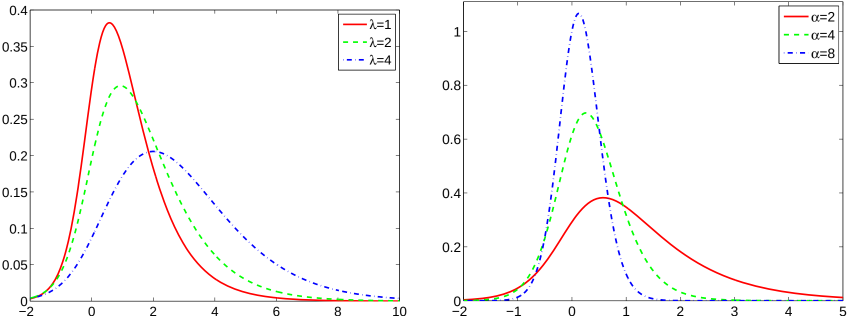


Figure: Left uses $\alpha = 2$ and $\beta = 1$; right uses $\lambda = 1$ and $\beta = 1$

Proper GHyp Density for $\delta = 1$ and $\mu = 0$

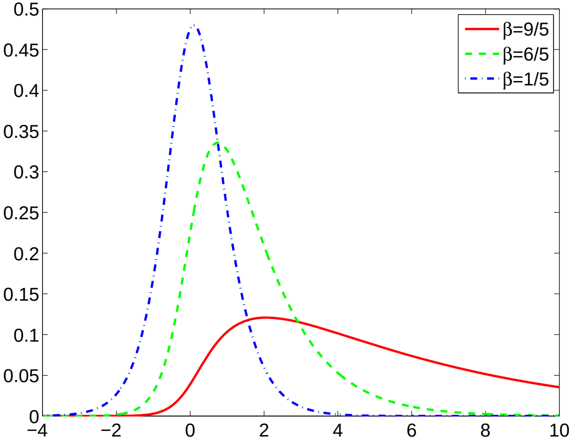


Figure: Uses $\lambda = 1$ and $\alpha = 2$

Variance–Gamma VG

This takes $\lambda > 0$, $\alpha > 0$, $\beta \in (-\alpha, \alpha)$, $\delta = 0$, $\mu \in \mathbb{R}$, so that $\chi = \delta^2 = 0$ and $\psi = \alpha^2 - \beta^2 > 0$. This implies we have a gamma mixture of normals, hence the name.
The p.d.f. reads

$$f_{\mathrm{GHyp}}(x; \lambda, \alpha, \beta, 0, \mu) = \frac{2\left(\frac{\alpha^2-\beta^2}{2}\right)^{\lambda}}{\sqrt{2\pi}\Gamma(\lambda)} \left(\frac{|x-\mu|}{\alpha}\right)^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}(\alpha|x-\mu|)e^{\beta(x-\mu)}.$$

It was popularized by Madan and Seneta (1990) in their study of financial returns data, and continues to receive attention in this context; see Seneta (2004) and the references therein.

Hyperbolic asymmetric t HAt

This takes $\lambda < 0$, $\beta \in \mathbb{R}$, $\alpha = |\beta|$, $\delta > 0$, $\mu \in \mathbb{R}$, so $\chi = \delta^2 > 0$ and $\psi = \alpha^2 - \beta^2 = 0$, and we have an inverse gamma mixture of normals.

- We coin this the *hyperbolic asymmetric* (Student's) t, or HAt,

$$\mathrm{HA}t(n, \beta, \mu, \delta) := \mathrm{GHyp}(\lambda, |\beta|, \beta, \delta, \mu),$$

  where $n = -2\lambda$, $n > 0$, $\beta, \mu \in \mathbb{R}$, and $\delta > 0$ and $\lambda = -n/2$.
- With $\alpha = |\beta| > 0$, it can be interpreted as (another) skewed t, because, if $\beta = 0$, it reduces to the Student's t p.d.f..
- Bibby and Sørensen (2003) refer to it as the *asymmetric scaled t-distribution*, whereas Aas and Haff (2005, 2006) name it the *(generalized hyperbolic) skew Student's t*.
- For $\beta \neq 0$ and $y_x = \sqrt{\delta^2 + (x-\mu)^2}$, the p.d.f. is

$$f_{\mathrm{HA}t}(x; n, \beta, \mu, \delta) = \frac{2^{\frac{-n+1}{2}} \delta^n}{\sqrt{\pi}\Gamma(n/2)} \left(\frac{y_x}{|\beta|}\right)^{-\frac{n+1}{2}} K_{-\frac{n+1}{2}}\left(|\beta|\, y_x\right) \mathrm{e}^{\beta(x-\mu)}.$$

Hyperbolic Hyp

This takes $\lambda = 1$, $\alpha > 0$, $\beta \in (-\alpha, \alpha)$, $\delta > 0$, $\mu \in \mathbb{R}$, so $\lambda = 1$, $\chi > 0$, $\psi > 0$, and we have a pHyp mixture of normals. We take the hyperbolic (there are several parameterizations) to be

$$\text{Hyp}_3(\alpha, \beta, \delta, \mu) := \text{GHyp}(1, \alpha, \beta, \delta, \mu).$$

with density $f_{\text{Hyp}_3}(x; \alpha, \beta, \delta, \mu)$ given by

$$\frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} \exp\left[-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)\right].$$

The name originates from the log-density shape: They are hyperbolae, even in the skewed case ($\beta \neq 0$), while the log-density of the normal distribution is a parabola.

Exactly as parabolae are limits of hyperbolae, the normal distribution is a limit of hyperbolic distributions.

Normal Inverse Gaussian NIG

This takes $\lambda = -1/2$, $\alpha > 0$, $\beta \in (-\alpha, \alpha)$, $\delta > 0$, $\mu \in \mathbb{R}$, and

$$\mathrm{NIG}(\alpha, \beta, \delta, \mu) := \mathrm{GHyp}(-1/2, \alpha, \beta, \delta, \mu),$$

with density $f_{\mathrm{NIG}}(x; \alpha, \beta, \delta, \mu)$ given by

$$e^{\delta\sqrt{\alpha^2-\beta^2}} \frac{\alpha\delta}{\pi\sqrt{\delta^2 + (x-\mu)^2}} K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) e^{\beta(x-\mu)}.$$

If $X \sim \mathrm{NIG}(\alpha, \beta, \delta, \mu)$, then

$$\mathbb{E}[X] = \mu + \beta\eta, \quad \mathbb{V}(X) = \eta + \beta^2 \frac{\eta^2}{\omega}, \quad \mu_3(X) = 3\beta\frac{\eta^2}{\omega} + 3\beta^3\frac{\eta^3}{\omega^2},$$

where $\eta = \delta/\sqrt{\alpha^2 - \beta^2}$ and $\omega = \delta\sqrt{\alpha^2 - \beta^2}$, and

$$\mathbb{M}_X(t) = e^{\mu t} e^{\delta\left(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+t)^2}\right)}, \quad -\alpha - \beta \le t \le \alpha - \beta. \qquad (62)$$

GHyp Properties: Location-Scale I

If $X \sim \text{GHyp}(\lambda, \alpha, \beta, \delta, \mu)$ and $a, b \in \mathbb{R}$ with $a \neq 0$, then

$$aX + b \sim \text{GHyp}(\lambda,\ \alpha/|a|,\ \beta/a,\ \delta|a|,\ a\mu + b). \tag{63}$$

Proof is direct: With $\alpha' = \alpha/|a|$, $\beta' = \beta/a$, $\delta' = \delta|a|$ and $\mu' = a\mu + b$, use general GHyp density expression. Then $f_{aX+b}(x)$ is

$$\begin{aligned}
& a^{-1} f_{\text{GHyp}}((x-b)/a; \lambda, \alpha, \beta, \delta, \mu) \\
= \quad & a^{-1} \frac{\mathrm{k}_{\lambda-\frac{1}{2}}\left(((x-b)/a-\mu)^2+\delta^2,\ \alpha^2\right)}{\sqrt{2\pi}\,\mathrm{k}_\lambda\left(\delta^2,\ \alpha^2-\beta^2\right)}\,\mathrm{e}^{\beta((x-b)/a-\mu)} \\
= \quad & a^{-1} \frac{\mathrm{k}_{\lambda-\frac{1}{2}}\left((x-b-a\mu)^2/a^2+(\delta|a|)^2/a^2,\ a^2(\alpha/|a|)^2\right)}{\sqrt{2\pi}\,\mathrm{k}_\lambda\left((\delta|a|)^2/a^2,\ a^2\left((\alpha/|a|)^2-(\beta/a)^2\right)\right)}\,\mathrm{e}^{(\beta/a)(x-b-a\mu)} \\
= \quad & a^{-1} \frac{\mathrm{k}_{\lambda-\frac{1}{2}}\left(\left((x-\mu')^2+\delta'^2\right)/a^2,\ a^2\alpha'^2\right)}{\sqrt{2\pi}\,\mathrm{k}_\lambda\left(\delta'^2/a^2,\ a^2\left(\alpha'^2-\beta'^2\right)\right)}\,\mathrm{e}^{\beta'(x-\mu')}.
\end{aligned}$$

Now apply $\mathrm{k}_\lambda(\chi, \psi) = r^\lambda \mathrm{k}_\lambda(r^{-1}\chi, r\psi)$. See the text...

GHyp Properties: Location-Scale II (1/2)

A second, more general way is as follows. Recall that, by construction, GHyp is a mixture with mixing weight GIG. Let π be a positive (continuous) distribution and $a, b \in \mathbb{R}$ such that $a \neq 0$. Then for all $\beta, \mu \in \mathbb{R}$,

$$a\,(\mathrm{Mix}_{\pi}(\mu, \beta)) + b = \mathrm{Mix}_{a^2\pi}(a\mu + b, \beta/a).$$

To see this, let X and Z be random variables such that $Z \sim \pi$ and $(X \mid Z) \sim \mathrm{N}(\mu + \beta Z, Z)$. Then, by definition, $X \sim \mathrm{Mix}_{\pi}(\mu, \beta)$. Thus,

$$\begin{aligned}
(aX + b \mid Z) &\sim a\,\mathrm{N}\,(\mu + \beta Z, Z) + b \\
&= \mathrm{N}\left(a\mu + a\beta Z + b, a^2 Z\right) \\
&= \mathrm{N}\left((a\mu + b) + \frac{\beta}{a}(a^2 Z), a^2 Z\right),
\end{aligned}$$

so that

$$aX + b \sim \mathrm{Mix}_{a^2\pi}\left(a\mu + b, \frac{\beta}{a}\right).$$

GHyp Properties: Location-Scale II (2/2)

Now use $\pi = \text{GIG}(\lambda, \delta^2, \alpha^2 - \beta^2)$ as mixing weight, and formula

$$r\,\text{GIG}(\lambda, \chi, \psi) = \text{GIG}(\lambda, r\chi, r^{-1}\psi),$$

(proved in the Problems) to get

$$\begin{aligned} a\,\text{GHyp}(\lambda, \alpha, \beta, \delta, \mu) + b &= a\left(\text{Mix}_{\text{GIG}(\lambda,\delta^2,\alpha^2-\beta^2)}(\mu, \beta)\right) + b \\ &= \text{Mix}_{a^2\,\text{GIG}(\lambda,\delta^2,\alpha^2-\beta^2)}(a\mu + b, \frac{\beta}{a}) \\ &= \text{Mix}_{\text{GIG}(\lambda,(a\delta)^2,(\alpha/|a|)^2-(\beta/a)^2)}(a\mu + b, \frac{\beta}{a}) \\ &= \text{GHyp}\left(\lambda, \frac{\alpha}{|a|}, \frac{\beta}{a}, a\delta, a\mu + b\right), \end{aligned}$$

providing a conceptual proof of the above result.

GHyp Properties: Location Parameter

We showed (63) above, namely: If $X \sim \text{GHyp}(\lambda, \alpha, \beta, \delta, \mu)$ and $a, b \in \mathbb{R}$ with $a \neq 0$, then

$$aX + b \sim \text{GHyp}(\lambda,\ \alpha/|a|,\ \beta/a,\ \delta|a|,\ a\mu + b).$$

In particular, if $X \sim \text{GHyp}(\lambda, \alpha, \beta, \delta, 0)$, then, for any $\mu \in \mathbb{R}$, $Y = X + \mu \sim \text{GHyp}(\lambda, \alpha, \beta, \delta, \mu)$, so that μ is a location parameter.

This is also seen as follows: μ is a location parameter iff $f_Y(y) = f_X(y - \mu)$. This is true, seen from (9.41) because all occurrences of x are of the form $x - \mu$.

GHyp Properties: Tail Behavior

α is the tail parameter in the sense that it dictates their fatness. The larger α is, the lighter are the tails.

$\beta \in [-\alpha, \alpha]$ is the skewness parameter: As $|\beta|$ grows (compared to α), so does the amount of skewness; the distribution is symmetric for $\beta = 0$.

β also influences the fatness of the tails of the GHyp distribution in that it can shift mass from one tail to the other. More precisely,

$$f_{\mathrm{GHyp}}(x; \lambda, \alpha, \beta, \delta, \mu) \propto |x|^{\lambda-1} e^{(\mp\alpha+\beta)x} \quad \text{as} \quad x \to \pm\infty,$$

showing that the tails are "semi-heavy".

Recall that the boundary case $|\beta| = \alpha$ is for the hyperbolic asymmetric t distribution, HAt. If $|\beta| = \alpha > 0$, then one of the tails is a (heavy) power-tail and the other is semi-heavy, while, if $\alpha = \beta = 0$, then both tails are heavy.

GHyp Properties: Alternative Parameterizations

Recall (63), i.e., if $X \sim \text{GHyp}(\lambda, \alpha, \beta, \delta, \mu)$ and $a, b \in \mathbb{R}$ with $a \neq 0$, then

$$aX + b \sim \text{GHyp}(\lambda,\ \alpha/|a|,\ \beta/a,\ \delta|a|,\ a\mu + b).$$

Then, restricting attention to the proper GHyp (so that $\delta > 0$), this can be used to show that $(\lambda, \overline{\alpha}, \overline{\beta}, \delta, \mu)$ with $\overline{\alpha} := \alpha\delta$, $\overline{\beta} := \beta\delta$ is a parameterization such that δ is a scale parameter and $\lambda, \overline{\alpha}, \overline{\beta}$ are location-scale invariant.

In particular, let $\text{GHyp}^{(2)}$ denote usage of the alternative parameter set. Then, for for the standard case $\delta = 1$ and $\mu = 0$,

$$X \sim \text{GHyp}(\lambda, \alpha, \beta, 1, 0) = \text{GHyp}^{(2)}(\lambda, \alpha, \beta, 1, 0)$$

and from (63), for $a > 0$,

$$aX + b \sim \text{GHyp}(\lambda, \frac{\alpha}{a}, \frac{\beta}{a}, a, b) = \text{GHyp}^{(2)}(\lambda, \alpha, \beta, a, b),$$

showing that $\lambda, \overline{\alpha}, \overline{\beta}$ are location-scale invariant and a is a scale parameter.

GHyp Properties: Alternative Parameterizations

This is of course trivially implemented:

```
function f=ghyp2pdf(x,lambda,alphabar,betabar,delta,mu)
if nargin<6, mu=0; end
if nargin<5, delta=1; end
a=alphabar/delta; b=betabar/delta;
f=ghyppdf(x,lambda,a,b,delta,mu); % page 318, Listing 9.2
```

Just as a check:

For $Z \sim \text{GHyp}(\lambda, \alpha, \beta, 1, 0) = \text{GHyp}^{(2)}(\lambda, \alpha, \beta, 1, 0)$ and $Y = aX + b$, $f_Y(y)$ can be computed as $a^{-1} f_Z((y-b)/a)$ using `ghyppdf`, or using that $Y \sim \text{GHyp}^{(2)}(\lambda, \alpha, \beta, a, b)$, using `ghyp2pdf`.

GHyp Properties: Alternative Parameterizations

In finance applications, it might be desirable to work with a mean-zero version, to which a (possibly time-varying) scale term is applied.

Let $X \sim \text{GHyp}^{(2)}(\lambda, \alpha, \beta, 1, 0)$ ($= \text{GHyp}(\lambda, \alpha, \beta, 1, 0)$) and, with $\delta = 1$, from the expression for the mean of the proper GHyp (top of page 320),

$$\mu = \mathbb{E}[X] = \beta\eta \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}, \quad \eta = \frac{\delta}{\sqrt{\alpha^2 - \beta^2}}, \omega = \delta\sqrt{\alpha^2 - \beta^2},$$

so that $Z = X - \mu$ has mean zero, and $Z \sim \text{GHyp}^{(2)}(\lambda, \alpha, \beta, 1, -\mu)$.

Then, for $c > 0$, we take

$$R = cZ = cX - c\mu \sim \text{GHyp}^{(2)}(\lambda, \alpha, \beta, c, -c\mu),$$

and $\mathbb{E}[R] = c\mathbb{E}[Z] = 0$.

GHyp Properties: Alternative Parameterizations

Another common parameterization of the proper GHyp is

$$(\lambda, \zeta, \rho, \delta, \mu) \quad \text{with} \quad \zeta := \delta\sqrt{\alpha^2 - \beta^2} = \omega, \ \rho := \frac{\beta}{\alpha}.$$

Again, μ and δ are location and scale parameters, respectively, and λ, ζ, ρ are location-scale invariant.

This can be deduced from the fact that ζ and ρ can be expressed in terms of $\overline{\alpha}$ and $\overline{\beta}$, namely $\zeta = \sqrt{\overline{\alpha}^2 - \overline{\beta}^2}$ and $\rho = \overline{\beta}/\overline{\alpha}$.

As $\overline{\alpha}$ and $\overline{\beta}$ are both location-scale invariant, so are ζ and ρ.

GHyp Properties: Alternative Parameterizations

Again with $\zeta := \delta\sqrt{\alpha^2 - \beta^2}$ and $\rho := \beta/\alpha$, a useful parameterization of the proper GHyp distribution is

$$(\lambda, p, q, \delta, \mu) \quad \text{with} \quad p := (1+\zeta)^{-1/2}, \ q := \rho p,$$

Because p and q are formed out of the location-scale invariant parameters ζ and ρ, they (and λ as well) are again location-scale invariant.

The parameter ζ is allowed to be any positive number, so $0 < p < 1$. As $-1 < \rho < 1$, we get $-p < q < p$. So the parameters p and q vary in the so-called *shape triangle* defined by

$$\left\{(q, p) \in \mathbb{R}^2 \mid 0 \leq |q| < p < 1\right\},$$

which was shown graphically in the beginning of the slides, and discussed in detail in the text.

Convolution and Infinite Divisibility

Suppose X and Y are independent random variables with distributions π and ρ, respectively. The distribution of sum $X+Y$ is obtained using the convolution formula.

Note that the distribution of $X+Y$ depends on X and Y only through their distributions π and ρ. As such, a popular notation to designate the convolution of two r.v.s is $\pi \star \rho$, read the *convolution of* π *and* ρ.

If π and ρ have probability density functions f_π and f_ρ, then the p.d.f. of $\pi \star \rho$ is

$$f_{\pi\star\rho}(t) = \int_{-\infty}^{+\infty} f_\pi(t-s) f_\rho(s)\,\mathrm{d}s.$$

Convolution and Infinite Divisibility

If π, ρ and σ are distributions, then the following rules easily carry over from the addition of random variables:

$$\pi \star \rho = \rho \star \pi \qquad \text{and} \qquad (\pi \star \rho) \star \sigma = \pi \star (\rho \star \sigma).$$

The n-fold convolution of π with itself is abbreviated by $\pi^{\star n}$, i.e.,

$$\pi^{\star n} = \underbrace{\pi \star \pi \star \cdots \star \pi}_{n \text{ times}}.$$

A distribution ρ is called an nth *convolution root of* π if $\rho^{\star n} = \pi$.

If π is a distribution on $(0, \infty)$, i.e., π is positive with probability one, then we want ρ to be positive as well.

Convolution and Infinite Divisibility

There are distributions which do not possess convolution roots. A distribution π is called *infinitely divisible* if, for every n, an nth convolution root of π exists (see also Problem 1.13).

As a simple example, suppose $\mu \in \mathbb{R}$ and $\sigma > 0$. Then the normal distribution $\mathrm{N}(\mu, \sigma^2)$ is infinitely divisible.

To see this, let $n \in \mathbb{N}$. If $X_i \overset{\text{ind}}{\sim} \mathrm{N}(\mu/n, \sigma^2/n)$, then

$$S = X_1 + \cdots + X_n \sim \mathrm{N}(n\mu/n, n\sigma^2/n) = \mathrm{N}(\mu, \sigma^2).$$

So $\mathrm{N}(\mu/n, \sigma^2/n)$ is an nth convolution root of $\mathrm{N}(\mu, \sigma^2)$.

Convolution and Infinite Divisibility: Mixtures

A mixture of normals behaves well under convolution. In particular, suppose π and ρ are distributions on $(0, \infty)$ and $\mu_\pi, \mu_\rho, \beta \in \mathbb{R}$. Then

$$\mathrm{Mix}_{\pi\star\rho}(\mu_\pi + \mu_\rho, \beta) = \mathrm{Mix}_\pi(\mu_\pi, \beta) \star \mathrm{Mix}_\rho(\mu_\rho, \beta).$$

In words, it does not matter whether you take the convolution product first and then mix, or mix first and subsequently build the convolution product.

See the text for proof of this important result.

Convolution and Infinite Divisibility: Mixtures

Barndorff-Nielsen and Halgreen (1977) show that the GIG distribution is infinitely divisible.

In the special case of the IG distribution, we showed above via the m.g.f. that the convolution of two IG distributions is again IG:

$$\mathrm{IG}_1(\chi_1, \psi) \star \mathrm{IG}_1(\chi_2, \psi) = \mathrm{IG}_1((\sqrt{\chi_1} + \sqrt{\chi_2})^2, \psi). \tag{64}$$

This applies also to the limiting case $\psi \to 0$, i.e., the Lévy distribution is also invariant under convolution.

We have shown this to also be true for the Cauchy (Example I.8.16) and normal. All three cases are nested in the stable Paretian family, which is invariant under convolution.

From the Barndorff-Nielsen and Halgreen (1977) result, it follows directly that the GHyp distribution is infinitely divisible (Problem 9.13).

Convolution and Infinite Divisibility: NIG

For the special case of the GHyp distribution with $\lambda = -1/2$, i.e., for the NIG distribution, we have an explicit formula:

$$\mathrm{NIG}(\alpha, \beta, \delta_1, \mu_1) * \mathrm{NIG}(\alpha, \beta, \delta_2, \mu_2) = \mathrm{NIG}(\alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2). \quad (65)$$

Exercise (Problem 9.14):

1. Show the convolution (65) formula for the NIG distribution using the convolution formula (64), i.e.,

   $$\mathrm{IG}_1(\chi_1, \psi) \star \mathrm{IG}_1(\chi_2, \psi) = \mathrm{IG}_1((\sqrt{\chi_1} + \sqrt{\chi_2})^2, \psi)$$

   and our knowledge of mixtures.
2. Show (65) using the formula (62) for the m.g.f. of the NIG distribution, i.e.,

   $$\mathbb{M}_X(t) = \mathrm{e}^{\mu t} \mathrm{e}^{\delta\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + t)^2}\right)}, \quad -\alpha - \beta \le t \le \alpha - \beta.$$

Convolution and Infinite Divisibility: NIG: Solution

1. Let $\psi = \alpha^2 - \beta^2 > 0$, $\chi_1 = \delta_1^2 > 0$ and $\chi_2 = \delta_2^2 > 0$. Then

$$
\begin{aligned}
\mathrm{NIG}(\alpha,\beta,\delta_1,\mu_1) * \mathrm{NIG}(\alpha,\beta,\delta_2,\mu_2) &= \mathrm{Mix}_{\mathrm{IG}_1(\chi_1,\psi)}(\mu_1,\beta) \star \mathrm{Mix}_{\mathrm{IG}_1(\chi_2,\psi)}(\mu_2,\beta) \\
&= \mathrm{Mix}_{\mathrm{IG}_1(\chi_1,\psi)\star\mathrm{IG}_1(\chi_2,\psi)}(\mu_1+\mu_2,\beta) \\
&= \mathrm{Mix}_{\mathrm{IG}_1\left(\left(\sqrt{\chi_1}+\sqrt{\chi_2}\right)^2,\psi\right)}(\mu_1+\mu_2,\beta) \\
&= \mathrm{NIG}\left(\alpha,\beta,\sqrt{\chi_1}+\sqrt{\chi_2},\mu_1+\mu_2\right) \\
&= \mathrm{NIG}\left(\alpha,\beta,\delta_1+\delta_2,\mu_1+\mu_2\right).
\end{aligned}
$$

2. We have that $\mathbb{M}_{\mathrm{NIG}(\alpha,\beta,\delta_1,\mu_1)*\mathrm{NIG}(\alpha,\beta,\delta_2,\mu_2)}(t)$ is

$$
\begin{aligned}
& \mathbb{M}_{\mathrm{NIG}(\alpha,\beta,\delta_1,\mu_1)}(t)\ \mathbb{M}_{\mathrm{NIG}(\alpha,\beta,\delta_2,\mu_2)}(t) \\
= \ & e^{\mu_1 t} e^{\delta_1\left(\sqrt{\alpha^2-\beta^2}-\sqrt{\alpha^2-(\beta+t)^2}\right)} e^{\mu_2 t} e^{\delta_2\left(\sqrt{\alpha^2-\beta^2}-\sqrt{\alpha^2-(\beta+t)^2}\right)} \\
= \ & e^{(\mu_1+\mu_2)t} e^{(\delta_1+\delta_2)\left(\sqrt{\alpha^2-\beta^2}-\sqrt{\alpha^2-(\beta+t)^2}\right)} = \mathbb{M}_{\mathrm{NIG}(\alpha,\beta,\delta_1+\delta_2,\mu_1+\mu_2)}(t).
\end{aligned}
$$

This is valid for all $t \in [-\alpha-\beta, \alpha-\beta]$. Therefore, the distributions to which these mgfs belong must be equal.

Distribution of Sums of Independent Proper GHyp

Let $S = \sum_{j=1}^{d} a_j X_j$, where $X_j \overset{\text{ind}}{\sim} \text{GHyp}^{(2)}\left(\lambda_j, \bar{\alpha}_j, \bar{\beta}_j, \delta_j, \mu_j\right)$, and $a_j > 0$ are fixed constants, $j = 1 \ldots, d$. Then, from the scaling result, $S = \sum_{j=1}^{d} Y_j$, where

$$\begin{aligned} Y_j := a_j X_j &\sim \text{GHyp}^{(2)}\left(\lambda_j, \bar{\alpha}_j, \bar{\beta}_j, a_j\delta_j, a_j\mu_j\right) \\ &= \text{GHyp}\left(\lambda_j, \frac{\bar{\alpha}_j}{a_j\delta_j}, \frac{\bar{\beta}_j}{a_j\delta_j}, a_j\delta_j, a_j\mu_j\right) \\ &=: \text{GHyp}\left(\lambda_j, \alpha_j, \beta_j, c_j, d_j\right), \end{aligned}$$

where $\alpha_j, \beta_j, c_j, d_j$ are so-defined.

The c.f. of Y_j is indeed $\varphi_{Y_j}(t) = \mathbb{M}_{Y_j}(it)$, with the bessel function $K_\lambda(\cdot)$ defined for complex arguments.

Thus, the c.f. of S is tractable and the usual inversion formulae or FFT approach can be successfully applied to get the density of S.

Distribution of Sums of Independent Proper GHyp

Alternatively, consider a saddlepoint approximation (s.p.a.).

For this we require the c.g.f. $\mathbb{K}_{Y_i}(t) = \ln \mathbb{M}_{Y_i}(t)$, and its first two derivatives, from which we can compute $\mathbb{K}_S(t) = \sum_{i=1}^{d} \mathbb{K}_{Y_i}(t)$, with similar expressions for $\mathbb{K}'_S(t)$ and $\mathbb{K}''_S(t)$.

For notational convenience, let $Y \sim \text{GHyp}(\lambda, \alpha, \beta, c, \mu)$ (drop the subscript i). Then, with $\omega = \sqrt{\chi\psi} = \delta\sqrt{\alpha^2 - \beta^2}$, and defining $Q_t = Q(t) := \sqrt{1 - (2\beta t + t^2)/\psi}$, it is easy to verify that the m.g.f. can be written as

$$\mathbb{M}_Y(t) = e^{\mu t} \frac{K_\lambda(\omega Q_t)}{K_\lambda(\omega)} Q_t^{-\lambda},$$

so that $\mathbb{K}_Y(t) = \mu t + \ln K_\lambda(\omega Q_t) - \ln K_\lambda(\omega) - \lambda \ln(Q_t)$.

Distribution of Sums of Independent Proper GHyp

Of critical use is the result (Problem 9.1) that

$$-2K'_{\nu}(x) = K_{\nu-1}(x) + K_{\nu+1}(x), \quad \nu \in \mathbb{R},\ x \in \mathbb{R}_{>0}. \tag{66}$$

Then $\mathrm{d}Q(t)/\mathrm{d}t = -(\beta + t)/Q_t\psi$ and, via (66),

$$\mathbb{K}'_Y(t) = \mu + \frac{\beta + t}{Q_t\psi}\left(\frac{\omega}{2}\frac{K_{\lambda-1}(\omega Q_t) + K_{\lambda+1}(\omega Q_t)}{K_\lambda(\omega Q_t)} + \frac{\lambda}{Q_t}\right).$$

The saddlepoint $\hat{t}$ needs to be numerically determined by solving $\mathbb{K}'_S(t) = x$. It will have one, and only one, solution in the convergence strip of the m.g.f. of S, $\max(-\alpha_j - \beta_j) < t < \min(\alpha_j - \beta_j)$.

Distribution of Sums of Independent Proper GHyp

Finally, a bit of straightforward work then shows that

$$\mathbb{K}_Y''(t) = \frac{\beta+t}{Q_t\psi} \times P_1 + \left[\frac{\omega}{2}\frac{K_{\lambda-1}(\omega Q_t)+K_{\lambda+1}(\omega Q_t)}{K_\lambda(\omega Q_t)} + \frac{\lambda}{Q_t}\right] \times \left[\frac{1}{Q_t\psi}\left(1+\frac{(\beta+t)^2}{Q_t^2\psi}\right)\right],$$

where

$$P_1 = \frac{\omega}{2}P_2 + \frac{\lambda(\beta+t)}{Q_t^3\psi}$$

and for P_2, $K_\lambda^2(\omega Q_t)\,P_2$ is

$$K_\lambda(\omega Q_t) \times \left[\frac{\omega}{2}\left(\frac{\beta+t}{Q_t\psi}\right)\left(K_{\lambda-2}(\omega Q_t)+2K_\lambda(\omega Q_t)+K_{\lambda+2}(\omega Q_t)\right)\right]$$
$$-\left[K_{\lambda-1}(\omega Q_t)+K_{\lambda+1}(\omega Q_t)\right] \times \left[\frac{\omega}{2}\left(K_{\lambda-1}(\omega Q_t)+K_{\lambda+1}(\omega Q_t)\right)\left(\frac{\beta+t}{Q_t\psi}\right)\right].$$

Distribution of Sums of Independent NIG

For the special case of Normal Inverse Gaussian, with $\psi_t = \alpha^2 - (\beta + t)^2$ and $\psi = \alpha^2 - \beta^2$,

$$\mathbb{K}_Y(t) = \mu t + \delta\left(\sqrt{\psi} - \sqrt{\psi_t}\right), \quad -\alpha - \beta \leq t \leq \alpha - \beta,$$

and we easily get

$$\mathbb{K}'_Y(t) = \mu + \delta \psi_t^{-1/2}(\beta + t), \mathbb{K}''_Y(t) = \delta \psi_t^{-1/2}\left(1 + \psi_t^{-1}(\beta + t)^2\right).$$

For the 2nd order SPA, we require $\mathbb{K}_Y^{(3)}(t)$ and $\mathbb{K}_Y^{(4)}(t)$. Standard calculation yields

$$\begin{aligned}
\mathbb{K}_Y^{(3)}(t) &= 3\delta \psi_t^{-3/2}(\beta + t)\left(1 + \psi_t^{-1}(\beta + t)^2\right), \\
\mathbb{K}_Y^{(4)}(t) &= 3\delta \psi_t^{-3/2}\left(1 + \psi_t^{-1}(\beta + t)^2\right)\left(1 + 5\psi_t^{-1}(\beta + t)^2\right).
\end{aligned}$$

Distribution of Sums of Independent NIG

In the $d = 1$ case, we get a simple, closed-form solution to the saddlepoint equation: With $z = (x - \mu)/\delta$ and $h = \beta + t$, the saddlepoint equation satisfies $z^2 = h^2 / \left(\alpha^2 - h^2\right)$, which leads to t being given by $\pm\alpha k_z - \beta$, where $k_z = z\left(1 + z^2\right)^{-1/2}$. From the convergence strip and that $|k_z| < 1$, the positive root must be the correct one, i.e., the saddlepoint is

$$\hat{t} = \frac{\alpha z}{\sqrt{1 + z^2}} - \beta, \quad z = \frac{x - \mu}{\delta}.$$

Furthermore, for $d = 1$, with $\kappa_j(t) = \mathbb{K}_Y^{(j)}(t) \,/\, \left[\mathbb{K}_Y^{(2)}(t)\right]^{j/2}$ for $j = 3, 4$, we arrive at

$$\kappa_3\left(\hat{t}\right) = 3\left(\delta\alpha\right)^{-1/2} z\left(1 + z^2\right)^{-1/4}, \; \kappa_4\left(\hat{t}\right) = 3\left(\delta\alpha\right)^{-1}\left(1 + z^2\right)^{-1/2}\left(1 + 5z^2\right).$$

2nd Order SPA Accuracy to NIG

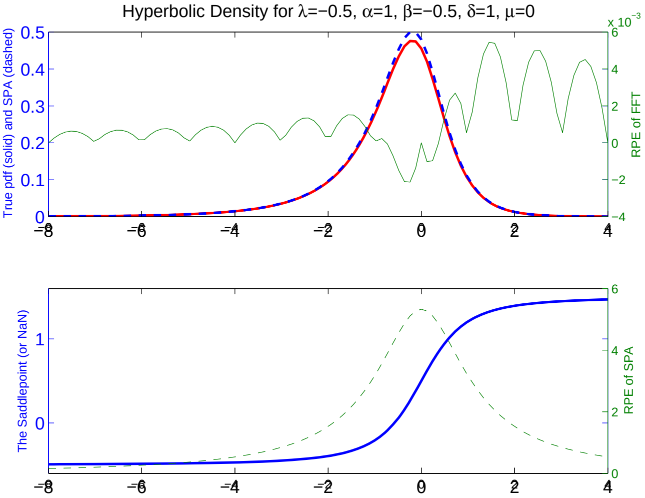
The next two slides show the accuracy of the 2nd order SPA, in terms of relative percentage error, `rpe=100*(fspa-ftrue)./ftrue`, for the NIG distribution. Also shown is the rpe for the FFT inversion, which should be nearly exact. The saddlepoint is also shown: As we go into the tails, the saddlepoint approaches the borders of the convergence strip, and eventually, numerical problems can arise.

The first slide is for $\alpha = 1$, $\beta = -1/2$, and $\delta = 1$. The second slide is the same but $\delta = 1/10$. We see that, as δ moves towards zero, the SPA breaks down. In addition, even the FFT suffers in accuracy.

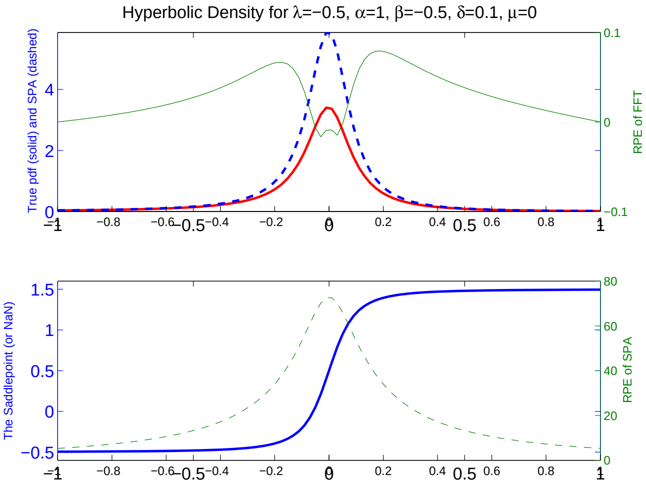
The third slide is for the GHyp with $\lambda = -2$ (same α and β, $\delta = 1$), for which we only have the 1st order SPA. As λ decreases, the SPA breaks down. It is obvious that even rescaling the SPA will not help. Most alarmingly, the saddlepoint equation can be solved only for values close to the mode.

Matters improve as λ increases; the fourth slide shows the case for $\lambda = 6$. Here, the SPA is extremely accurate, and the saddlepoint is well within the convergence strip.

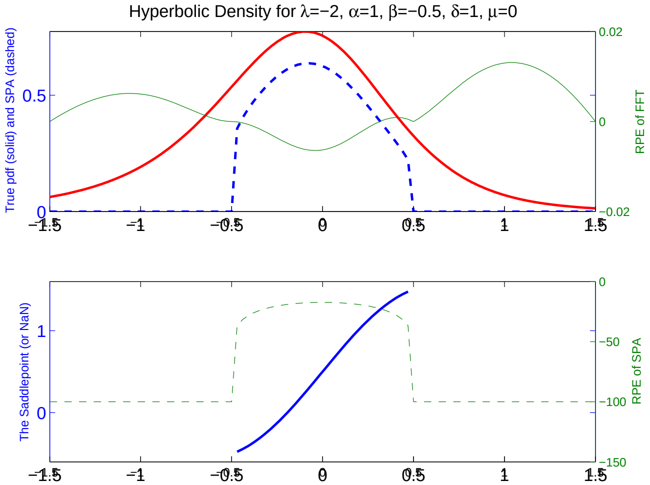
2nd Order SPA Accuracy to NIG



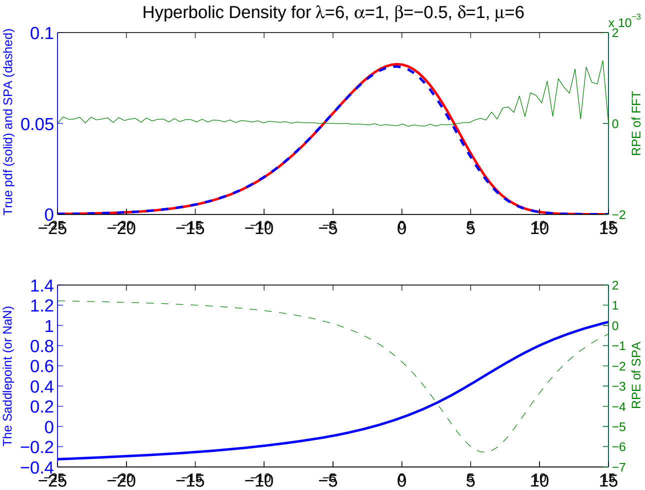
2nd Order SPA Accuracy to NIG



SPA Accuracy to GHyp with $\lambda = -2$



SPA Accuracy to GHyp with $\lambda = 6$ (and $\mu = 6$)



Intermediate Probability: A Computational Approach

Noncentral Distributions: The Noncentral Chi Square

The noncentral χ^2 distribution arises in many statistical applications such as goodness of fit, contingency tables and likelihood ratio tests. It also has a number of uses in finance and engineering, such as its role in the Cox-Ingersoll-Ross Process.[10]

Let $X_i \overset{\text{ind}}{\sim} \mathrm{N}(\mu_i, 1)$, $i = 1, \ldots, n$, with at least one $\mu_i \neq 0$. Interest centers on the distribution of $X = \sum_{i=1}^{n} X_i^2$.

Let $\mathbf{X} = (X_1, \ldots, X_n)'$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$ so that $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{I})$ and define $\theta = \boldsymbol{\mu}'\boldsymbol{\mu} = \sum_{i=1}^{n} \mu_i^2$, which is referred to as the **noncentrality parameter**.

We will use the notation $X \sim \chi^2(n, \theta)$ to denote a χ^2 r.v. with n degrees of freedom and noncentrality parameter θ.

[10] See, for example, S. Dyrting (2004), *Evaluating the Noncentral Chi-Square Distribution for the Cox-Ingersoll-Ross Process*, Computational Economics, Vol. 24(1), and the references therein.

Noncentral Chi Square: Derivation

As in Rao (1973) and Stuart, Ord and Arnold (1999), let $\mathbf{B}$ be an orthogonal $n \times n$ matrix with its first row given by $\boldsymbol{\mu}'\theta^{-1/2}$ for $\theta > 0$. Then $\mathbf{B}\boldsymbol{\mu} = (\theta^{1/2}, 0, \ldots, 0)'$.

To see this, let $\mathbf{b}_j$ denote the jth row of $\mathbf{B}$. Then, the condition on the first row of $\mathbf{B}$ is $\mathbf{b}_1 = \boldsymbol{\mu}'\theta^{-1/2}$, or $\boldsymbol{\mu} = \mathbf{b}_1'\theta^{1/2}$. From the orthogonality of $\mathbf{B}$, $\mathbf{b}_1\boldsymbol{\mu} = \mathbf{b}_1\mathbf{b}_1'\theta^{1/2} = \theta^{1/2}$ and, for $j = 2, \ldots, n$, $\mathbf{b}_j\,\boldsymbol{\mu} = \mathbf{b}_j\mathbf{b}_1'\theta^{1/2} = 0$, i.e., $\mathbf{B}\boldsymbol{\mu} = (\theta^{1/2}, 0, \ldots, 0)'$.

For example, with $n = 2$ and $\theta > 0$,

$$\mathbf{B} = \theta^{-1/2}\begin{pmatrix} \mu_1 & \mu_2 \\ \mu_2 & -\mu_1 \end{pmatrix}$$

Noncentral Chi Square: Derivation

Recall $\mathbf{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \mathbf{I})$ and $\mathbf{B}\boldsymbol{\mu} = (\theta^{1/2}, 0, \ldots, 0)'$.

Let $\mathbf{Y} = \mathbf{B}\mathbf{X}$, with $\mathbb{E}[\mathbf{Y}] = \mathbf{B}\boldsymbol{\mu}$ and $(\mathbf{B}\boldsymbol{\mu})' \mathbf{B}\boldsymbol{\mu} = \boldsymbol{\mu}'\boldsymbol{\mu} = \theta$.

Then $\mathbf{Y} \sim \mathrm{N}(\mathbf{B}\boldsymbol{\mu}, \mathbf{I})$, i.e., the Y_i are independent, unit variance, normal r.v.s with $\mathbb{E}[Y_1] = \theta^{1/2}$, and $\mathbb{E}[Y_i] = 0$, $i = 2, \ldots, n$. Thus,

$$X := \mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X} = \mathbf{Y}'\mathbf{Y} = Y_1^2 + \left(Y_2^2 + \cdots + Y_n^2\right) = Y_1^2 + Z,$$

where $Z := \left(Y_2^2 + \cdots + Y_n^2\right) \sim \chi^2_{n-1}$ (Example 2.3) with density

$$f_Z(z) = \frac{1}{2^{(n-1)/2}\Gamma\left((n-1)/2\right)} z^{(n-1)/2-1} e^{-z/2} \mathbb{I}_{(0,\infty)}(z).$$

Noncentral Chi Square: Derivation

With $Y_1 \sim \mathrm{N}(\sqrt{\theta}, 1)$, $F_{Y_1^2}(y) = \Pr(Y_1^2 \leq y)$ is

$$\Pr\left(-\sqrt{y} \leq Y_1 \leq \sqrt{y}\right) = \Phi(\sqrt{y} - \sqrt{\theta}) - \Phi(-\sqrt{y} - \sqrt{\theta}).$$

Differentiating, for $y > 0$,

$$\begin{aligned} f_{Y_1^2}(y) &= \frac{1}{2} y^{-1/2} \left(\phi\left(\sqrt{y} - \sqrt{\theta}\right) + \phi\left(-\sqrt{y} - \sqrt{\theta}\right)\right) \\ &= \frac{y^{-1/2}}{2\sqrt{2\pi}} \left(\exp\left\{-\frac{1}{2}\left(\sqrt{y} - \sqrt{\theta}\right)^2\right\} + \exp\left\{-\frac{1}{2}\left(-\sqrt{y} - \sqrt{\theta}\right)^2\right\}\right) \\ &= \frac{y^{-1/2}}{2\sqrt{2\pi}} \exp\left(-\frac{y+\theta}{2}\right)\left(\exp\left(\sqrt{y\theta}\right) + \exp\left(-\sqrt{y\theta}\right)\right). \end{aligned}$$

Then, as

$$\frac{e^z + e^{-z}}{2} = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \frac{z^6}{6!} + \cdots = \cosh(z),$$

$$f_{Y_1^2}(y) = \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y+\theta}{2}\right) \sum_{i=0}^{\infty} \frac{(y\theta)^i}{(2i)!} \mathbb{I}_{(0,\infty)}(y).$$

Noncentral Chi Square: Derivation

Now compute $f_X(x) = \int_{-\infty}^{\infty} f_{Y_1^2}(y)\, f_Z(x-y)\, \mathrm{d}y$. We get

$$
\begin{aligned}
f_X(x) &= e^{-x/2} \sum_{i=0}^{\infty} \frac{e^{-\theta/2}(\theta/2)^i}{i!} \frac{x^{n/2+i-1}}{2^{n/2+i}\Gamma(i+n/2)} \mathbb{I}_{(0,\infty)}(x) \\
&= \sum_{i=0}^{\infty} \omega_{i,\theta}\; g_{n+2i}(x), \qquad (67)
\end{aligned}
$$

where g_v denotes the (central) χ_v^2 density and $\omega_{i,\theta}$ are Poisson weights. The c.d.f. of X can thus be expressed similarly, in terms of the c.d.f.s of central χ_v^2 r.v.s

$$
\int_0^x f_X(x)\, \mathrm{d}x = \sum_{i=0}^{\infty} \omega_{i,\theta} \int_0^x g_{n+2i}(x)\, \mathrm{d}x = \sum_{i=0}^{\infty} \omega_{i,\theta}\; G_{n+2i}(x).
$$

Noncentral χ^2: Derivation

Simple manipulations show that the p.d.f. of $X \sim \chi^2(n, \theta)$ can be represented as

$$\frac{1}{2} e^{-(x+\theta)/2} x^{(n-2)/4} \theta^{-(n-2)/4} I_{(n-2)/2}(\sqrt{\theta x}) \mathbb{I}_{(0,\infty)}(x),$$

where

$$I_v(z) = \sum_{i=0}^{\infty} \frac{(z/2)^{v+2i}}{i!\,\Gamma(v+i+1)}$$

is the modified Bessel function of the first kind. In Matlab,

```
n=4; t=20; x=0:0.25:60; f1=ncx2pdf(x,n,t);
f2 = 0.5*exp(-(x+t)/2) .* x.^((n-2)/4) * t^(-(n-2)/4) ...
      .* besseli((n-2)/2,sqrt(t*x));
```

The Matlab function `ncx2pdf` (wisely) uses “outward summing” applied to (67) to get the exact p.d.f., but this is about 9 times slower than use of the Bessel function expression.

Noncentral χ^2: Moments

From

$$f_X(x) = \sum_{i=0}^{\infty} \omega_{i,\theta} \; g_{n+2i}(x), \qquad \omega_{i,\theta} = e^{-\theta/2} (\theta/2)^i \; / \, i!,$$

we have, with $P \sim \mathrm{Poi}(\theta/2)$,

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=0}^{\infty} \omega_{i,\theta} \int_0^{\infty} x g_{n+2i}(x) \, \mathrm{d}x \\ &= \sum_{i=0}^{\infty} \frac{e^{-\theta/2} (\theta/2)^i}{i!} (n+2i) = \mathbb{E}[n+2P] = n+\theta, \end{aligned}$$

using the expected value of central χ^2 and Poisson random variables.

Noncentral χ^2: Moments

Similarly, $\mathrm{Var}(X) = 2n + 4\theta$.

More generally, for $s \in \mathbb{R}$ with $s > -n/2$,

$$\mathbb{E}[X^s] = \frac{2^s}{e^{\theta/2}} \frac{\Gamma(n/2+s)}{\Gamma(n/2)} \, {}_1F_1(n/2+s, n/2; \theta/2),$$

where ${}_1F_1$ is the confluent hypergeometric function; see Section 5.3.

For $s \in \mathbb{N}$, it can be shown that

$$\mathbb{E}[X^s] = 2^s \Gamma\left(s + \frac{n}{2}\right) \sum_{i=0}^{s} \binom{s}{i} \frac{(\theta/2)^i}{\Gamma(i+n/2)}, \quad s \in \mathbb{N}.$$

Noncentral χ^2: Exercise

Let X be a r.v. with finite mean μ. How do $\mathbb{E}[|X|]$ and $|\mathbb{E}[X]|$ compare? Compare them explicitly in the case that $X \sim \mathrm{N}(\mu, 1)$. In doing so, show that

$$\lim_{|\mu|\to\infty} \frac{1}{|\mu|} \, {}_1F_1\left(-\frac{1}{2}, \frac{1}{2}; -\frac{\mu^2}{2}\right) = \sqrt{\frac{\pi}{2}}.$$

Noncentral χ^2: Solution

Clearly, from properties of the Riemann integral, for any r.v. X with finite mean μ,

$$\mathbb{E}\left[|X|\right] = \int_{-\infty}^{\infty} |x| \, \mathrm{d}F_X(x) = \int_{-\infty}^{\infty} |x \mathrm{d}F_X(x)| \geq \left| \int_{-\infty}^{\infty} x \mathrm{d}F_X(x) \right| = \left|\mathbb{E}\left[X\right]\right|.$$

For the case with $X \sim \mathrm{N}(\mu, 1)$, $\mathbb{E}\left[X^2\right] = 1^2 + \mu^2$ and from Jensen's inequality and square root being concave,

$$\sqrt{1+\mu^2} = \sqrt{\mathbb{E}\left[X^2\right]} \geq \mathbb{E}\left[|X|\right] \geq \left|\mathbb{E}\left[X\right]\right| = |\mu|. \tag{68}$$

Noncentral χ^2: Solution

For calculating $\mathbb{E}[|X|]$: For $\mu = 0$, (I.7.39) easily shows that $\mathbb{E}[|X|] = \sqrt{2/\pi}$. For $\mu \neq 0$, we need (II.10.9) with $s = 1/2$, $n = 1$ and $\theta = \mu^2$; this and Kummer's transformation (II.5.29) give

$$\mathbb{E}[|X|] = \exp\left(-\mu^2/2\right)\sqrt{\frac{2}{\pi}}\,{}_1F_1\left(1, \frac{1}{2}; \frac{\mu^2}{2}\right) = \sqrt{\frac{2}{\pi}}\,{}_1F_1\left(-\frac{1}{2}, \frac{1}{2}; -\frac{\mu^2}{2}\right).$$

All we can (easily) then say is that, since $\mathbb{E}[|X|] \geq |\mathbb{E}[X]|$,

$${}_1F_1\left(-\frac{1}{2}, \frac{1}{2}; -\frac{\mu^2}{2}\right) \geq |\mu|\sqrt{\frac{\pi}{2}}$$

and that, from the upper bound on $\mathbb{E}[|X|]$ from (68), that $\mathbb{E}[|X|] \to |\mu|$ as $|\mu|$ increases, so that

$$\lim_{|\mu|\to\infty} \frac{1}{|\mu|}\,{}_1F_1\left(-\frac{1}{2}, \frac{1}{2}; -\frac{\mu^2}{2}\right) = \sqrt{\frac{\pi}{2}}.$$

Noncentral χ^2: MGF

Problem 10.6 shows that the m.g.f. of X is

$$\mathbb{M}_X(t) = (1-2t)^{-n/2} \exp\left\{ \frac{t\theta}{1-2t} \right\}, \quad t < 1/2. \tag{69}$$

With c.g.f. $\mathbb{K}_X(t) = \ln \mathbb{M}_X(t)$, it is easy to verify that

$$\mathbb{K}_X'(t) = n\,(1-2t)^{-1} + \theta\,(1-2t)^{-2},$$

with higher order terms easily computed.

From these, we obtain

$$\kappa_i = \mathbb{K}_X^{(i)}(0) = 2^{i-1}\,(i-1)!\,(n+i\theta).$$

For $i=1$ and $i=2$, this immediately yields the mean and variance. For the 3rd central moment, $\mu_3 = \kappa_3 = 8\,(n+3\theta)$.

Noncentral χ^2: SPA

The simple form of the cumulative generating function obtained implies that the saddlepoint approximation (s.p.a.) can be easily implemented.

In particular, we require the solution to $x = \mathbb{K}_X'(t)$, or the zeros of $4xt^2 - 2t\,(2x - n) - n - \theta + x$, given by

$$t_\pm = \frac{1}{4x}\left(2x - n \pm \sqrt{n^2 + 4x\theta}\right).$$

Rearranging and using the facts that

1. constraint $t < 1/2$ from (69),
2. $\theta \geq 0$,
3. the interior of the support of X is $\mathbb{R}_{>0}$ (i.e., $x > 0$),

easily shows that t_- is always the correct solution.

This is another example of when a closed–form solution to the saddlepoint equation and, thus, the approximate p.d.f. exists.

Weighted Sums of Independent Central χ^2

There are many examples of test statistics (in particular, when working with linear regression models and time series analysis) in which the distribution of a weighted sum of independent χ^2 r.v.s arises.

The special case with central χ^2 r.v.s is of most importance, so we consider it separately now; the general case is discussed below.

Let $X_i \overset{\text{ind}}{\sim} \chi^2(n_i)$ and define $X = \sum_{i=1}^{k} a_i X_i$, $a_i \neq 0$.
The m.g.f. is

$$\mathbb{M}_X(s) = \prod_{i=1}^{k} \mathbb{M}_{a_i X_i}(s) = \prod_{i=1}^{k} \mathbb{M}_{X_i}(a_i s) = \prod_{i=1}^{k} (1 - 2a_i s)^{-n_i/2},$$

which is valid for s such that $1 - 2a_i s > 0$, $i = 1, \ldots, k$.

Weighted Sums of Independent Central χ^2

From the m.g.f., the saddlepoint approximation is clearly applicable for the p.d.f. and c.d.f.; see the text for details.

The inversion formulae can be used: The characteristic function is

$$\varphi_X(t) = \mathbb{M}_X(it) = \prod_{j=1}^{k} (1 - 2a_j it)^{-n_j/2}.$$

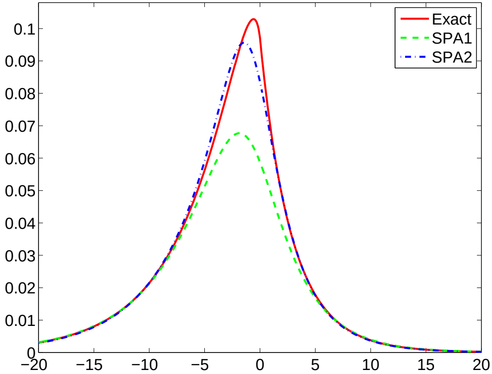
A specific method for the c.d.f. has been developed by Grad and Solomon (1955) and Pan (1968) for $X = \sum_{i=1}^{k} a_i X_i$ for $X_i \overset{\text{i.i.d.}}{\sim} \chi^2(1)$ and $a_i \neq a_j$. See the text for details.

The p.d.f. can then be accurately approximated as $(F_X(x+\delta) - F_X(x-\delta))/(2\delta)$ for, say, $\delta = 10^{-7}$.

The Pan method is about 15 times faster than use of the inversion formula (and 3 times slower than the SPA), but the inversion formula for the p.d.f. will be more accurate than use of the Pan c.d.f. procedure.

Weighted Sums of Independent Central χ^2

As in Figure 10.1 page 348, the figure below shows density of $\sum_{i=1}^{5} a_i X_i$, where $\mathbf{a} = (-3, -2, -1, 1, 2)$ and $X_i \overset{\text{i.i.d.}}{\sim} \chi^2(1)$. Solid line is exact (compute with either inversion formula or from the Pan c.d.f.), dashed and dash-dot are the first and second order saddlepoint approximations (programs given in the book).



Weighted Sums of Independent Noncentral χ^2: SPA

Now let $X_i \overset{\text{ind}}{\sim} \chi^2(n_i, \theta_i)$ and $X = \sum_{i=1}^k a_i X_i$, $a_i \neq 0$.

The s.p.a. is easy to implement. From (69),

$$\mathbb{M}_{X_i}(a_i s) = (1 - 2a_i s)^{-n_i/2} \exp\left\{\frac{a_i s \theta_i}{1 - 2a_i s}\right\}, \quad 1 - 2a_i s > 0,$$

so that $\mathbb{M}_X(s) = \prod_{i=1}^k \mathbb{M}_{X_i}(a_i s)$ for s in the convergence strip (see the text for details).

Weighted Sums of Independent Noncentral χ^2: SPA

For convenience, let $\vartheta_i = \vartheta_i(s) = (1 - 2sa_i)^{-1}$. Then straightforward calculation yields

$$\mathbb{K}_X(s) = \frac{1}{2}\sum_{i=1}^{k} n_i \ln \vartheta_i + s\sum_{i=1}^{k} a_i\theta_i\vartheta_i, \quad \mathbb{K}_X'(s) = \sum_{i=1}^{k} a_i\vartheta_i\left(n_i + \theta_i\vartheta_i\right)$$

and

$$\mathbb{K}_X''(s) = 2\sum_{i=1}^{k} a_i^2\vartheta_i^2\left(n_i + 2\theta_i\vartheta_i\right), \quad \mathbb{K}_X'''(s) = 8\sum_{i=1}^{k} a_i^3\vartheta_i^3\left(n_i + 3\theta_i\vartheta_i\right),$$

and $\mathbb{K}_X^{(4)}(s) = 48\sum_{i=1}^{k} a_i^4\vartheta_i^4\left(n_i + 4\theta_i\vartheta_i\right)$, from which the s.p.a. can be calculated once $\hat{s}$ is determined (see the text for details).

Weighted Sums of Independent Noncentral χ^2

The exact c.d.f. can be computed from the inversion formula. In this case, it can be simplified to yield an integral which does not involve complex numbers (and so is faster to evaluate). See the text for all the details.

The resulting expression, as given in the text, is due to Imhof (1961), which is still an enormously cited article because of the vast number of statistical applications which require evaluation of the c.d.f. of X (such as the Durbin Watson statistic, and many others).

The "Imhof method" (again, it is just the inversion theorem simplified to get a non-complex expression) is much more general than the Pan method, but also much slower.

The SPA is very accurate in this context; see the text for an example.

Noncentral F

- The **singly noncentral** F distribution plays a central role in the **analysis of variance**, where it determines the **power for tests of linear hypotheses**.
- The **doubly noncentral** F also arises in such contexts.
- Several applications of the doubly noncentral F in **econometrics**.
- It is also of use in **signal processing** and **pattern recognition** applications.
- Let $X_i \overset{\text{ind}}{\sim} \chi^2(n_i, \theta_i)$, $i = 1, 2$, and define $F = (X_1/n_1)/(X_2/n_2)$ and $\omega_{i,\theta} = e^{-\theta/2}(\theta/2)^i / i!$.

Noncentral F: PDF

It is straightforward to show (Problem 10.8) that the density of F,

$$f_F\left(x; n_1, n_2, \theta_1, \theta_2\right) = f_F\left(x\right),$$

is given by

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \omega_{i,\theta_1}\omega_{j,\theta_2} \frac{n_1^{n_1/2+i}}{n_2^{-n_2/2-j}} \frac{x^{n_1/2+i-1}\left(xn_1+n_2\right)^{-(n_1+n_2)/2-i-j}}{B\left(i+n_1/2, j+n_2/2\right)},$$

which is referred to as the **doubly noncentral** F distribution, $\mathrm{F}\left(n_1, n_2, \theta_1, \theta_2\right)$.

If $\theta_2 = 0$, this reduces to

$$f_F\left(x\right) = \sum_{i=0}^{\infty} \omega_{i,\theta_1} \frac{n_1^{n_1/2+i}}{n_2^{-n_2/2}} \frac{x^{n_1/2+i-1}\left(xn_1+n_2\right)^{-(n_1+n_2)/2-i}}{B\left(i+n_1/2, n_2/2\right)},$$

(taking $\left(\theta_2/2\right)^0 = 1$), which is the **(singly) noncentral** F distribution.

If the noncentrality is desired only in the denominator, recall that $\Pr\left(F < x\right) = \Pr\left(1/F > 1/x\right)$.

Noncentral F: CDF

The c.d.f. $F_F(x; n_1, n_2, \theta_1, \theta_2) = F_F(x)$ can be expressed as

$$F_F(x) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \omega_{i,\theta_1} \omega_{j,\theta_2} \bar{B}_y(i + n_1/2, j + n_2/2), \quad y = \frac{n_1 x}{n_1 x + n_2},$$

where $\bar{B}$ is the incomplete beta function ratio.

For $\theta_2 = 0$,

$$F_F(x) = \sum_{i=0}^{\infty} \omega_{i,\theta_1} \bar{B}_y(i + n_1/2, \ n_2/2), \quad y = \frac{n_1 x}{n_1 x + n_2},$$

and reduces to $F_F(x) = \bar{B}_y(n_1/2, n_2/2)$ in the central case.

Inspection of the terms in the single and double sums of the p.d.f. and c.d.f. suggest a smart way to evaluate them to machine precision; see the text for details and programs of this “outer summing” method, and remarks of how many software packages get it wrong...

Noncentral F: Alternative Computation

The c.d.f. of $F = (X_1/n_1)\,/\,(X_2/n_2)$ can also be computed by writing

$$\Pr(F \le x) = \Pr\left(\frac{n_2}{n_1} X_1 - x X_2 \le 0\right) = \Pr(Y_x \le 0),$$

where Y_x is the so–defined linear combination of $k = 2$ independent noncentral χ^2 r.v.s, and then using the inversion formula (Imhof method).

This has the advantage that the computation time involved is not a function of the noncentrality terms as it is in the outward summing method, and is thus of particular use in the doubly noncentral case with large noncentrality parameters.

An exact method for the p.d.f. based on the Geary (1944) result, and an equivalent integral expression without complex numbers is derived in Broda and Paolella (2008).

Moments: Singly Noncentral F

In the singly noncentral case, we show that

$$\mu = \mathbb{E}[X] = \frac{n_2}{n_1} \frac{n_1 + \theta_1}{n_2 - 2}, \quad n_2 > 2$$

and

$$\mathrm{Var}(X) = 2 \frac{n_2^2}{n_1^2} \frac{(n_1 + \theta_1)^2 + (n_1 + 2\theta_1)(n_2 - 2)}{(n_2 - 2)^2 (n_2 - 4)}, \quad n_2 > 4,$$

both of which reduce to the well-known expressions in the central case.

Moments: Doubly Noncentral F

Some work shows that the rth raw moment of the doubly noncentral F can be expressed in terms of the rth raw moment of the singly noncentral F, and the confluent hypergeometric function. In particular,

$$_2\mu_r' = {}_1\mu_r' \, {}_1F_1 \left(r, n_2/2, -\theta_2/2 \right),$$

Section 5.3 gives a highly accurate and trivially computed approximation to the ${}_1F_1$ function, so that, unless high accuracy is required, the raw moments of the doubly noncentral F are easily computed, from which the central moments can be obtained in the usual way.

Noncentral (Student's) t Distribution

Similar to the noncentral F, the noncentral t distribution arises in power calculations for statistical hypothesis testing and also in construction of certain confidence intervals.

It has also found use in a large variety of other contexts; see the examples and references in Johnson *et al*. (1994, p. 512). Some more recent applications in finance include its use in modeling the dynamics of variance and skewness from daily and monthly equity returns data,[11] in studying the behavior of tests of asset pricing,[12] and in real estate asset allocation models.[13]

[11] Campbell R. Harvey and Akhtar Siddique (1999), *Autoregressive Conditional Skewness*, Journal of Financial and Quantitative Analysis, Vol. 34(4), pp. 465-487.

[12] Raymond Kan and Chu Zhang (1999), *Two-Pass Tests of Asset Pricing Models with Useless Factors*, The Journal of Finance, Vol. 54(1), pp. 203-235.

[13] Mark S. Coleman and Asieh Mansour (2005), *Real Estate in the Real World: Dealing with Non-Normality and Risk in an Asset Allocation Model*, Journal of Real Estate Portfolio Management, Vol. 11(1) pp. 37-54.

Noncentral (Student's) t Distribution

Let $X \sim \mathrm{N}(\mu, 1)$ independent of $Y \sim \chi^2(k, \theta)$. Random variable $T = X/\sqrt{Y/k}$ is said to follow a doubly noncentral t distribution with numerator noncentrality parameter μ and denominator noncentrality parameter θ.

If $\theta = 0$, is **singly noncentral** t with noncentrality parameter μ. Consider this special case first:

A simple transformation shows that the density of $Z = \sqrt{Y/k}$ is given by

$$f_Z(z) = \frac{2^{-k/2+1} k^{k/2}}{\Gamma(k/2)} z^{k-1} e^{-\left(kz^2\right)/2} \mathbb{I}_{(0,\infty)}(z).$$

Singly Noncentral t: CDF

Recall from Chapter 5: If X and Y are continuous random variables and event $A = \{X < aY\}$, then, conditioning on Y,

$$\begin{aligned}\Pr(A) &= \Pr(X < aY) = \int_{-\infty}^{\infty} \Pr(X < aY \mid Y = y)\, f_Y(y)\, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} F_{X|Y}(ay)\, f_Y(y)\, \mathrm{d}y.\end{aligned}$$

Using this with X and Z independent, the c.d.f. of T is

$$\begin{aligned}F_T(t;k,\mu) &= \Pr(X \le tZ) = \int_{-\infty}^{\infty} F_{X|Z}(tz)\, f_Z(z)\, \mathrm{d}z = \int_0^{\infty} F_X(tz)\, f_Z(z)\, \mathrm{d}z \\ &= \frac{2^{-k/2+1} k^{k/2}}{\Gamma(k/2)} \int_0^{\infty} \Phi(tz;\mu,1)\; z^{k-1} \exp\left\{-\frac{1}{2}kz^2\right\} \mathrm{d}z,\end{aligned}$$

where $\Phi(tz;\mu,1) = (2\pi)^{-1/2} \int_{-\infty}^{tz} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} \mathrm{d}x$.

See the text for a program which implements this.

Singly Noncentral t: PDF

Differentiating the above c.d.f. expression yields an integral expression for the p.d.f.. Further manipulations (see the text and Problem 10.10) yield

$$
\begin{aligned}
f_T(t; k, \mu) &= e^{-\mu^2/2} \frac{\Gamma\left((k+1)/2\right) k^{k/2}}{\sqrt{\pi}\,\Gamma\left(k/2\right)} \left(\frac{1}{k+t^2}\right)^{\frac{k+1}{2}} \\
&\quad \times \left(\sum_{i=0}^{\infty} \frac{(t\mu)^i}{i!} \left(\frac{2}{t^2+k}\right)^{i/2} \frac{\Gamma\left((k+i+1)/2\right)}{\Gamma\left((k+1)/2\right)} \right),
\end{aligned}
$$

which lends itself well to numeric computation.

Further work, as detailed in the text and exercises, yields an infinite sum representation for the c.d.f. as well, which is well-suited for numeric calculation.

Doubly Noncentral t

Now, $T = X/\sqrt{Y/k}$, where $X \sim \mathrm{N}(\mu, 1)$, independent of $Y \sim \chi^2(k, \theta)$, so that $T \sim t''(k, \mu, \theta)$. Some work shows that

$$f_T(t; k, \mu, \theta) = \frac{\mathrm{e}^{-\left(\theta+\mu^2\right)/2}}{\sqrt{\pi k}} \sum_{j=0}^{\infty} A_j(t),$$

where

$$A_j(t) = \frac{1}{j!} \frac{\left(t\mu\sqrt{2/k}\right)^j}{\left(1+t^2/k\right)^{(k+j+1)/2}} \frac{\Gamma\left(\frac{k+j+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \, {}_1F_1\left(\frac{k+j+1}{2}, \frac{k}{2}, \frac{\theta}{2\left(1+t^2/k\right)}\right).$$

For the c.d.f., Kocherlakota and Kocherlakota (1991) derive (Problem 10.13) the following representation, an infinite Poisson weighted sum of *singly* noncentral t c.d.f. values, which is convenient for numeric computation:

$$F_T(t; k, \mu, \theta) = \sum_{i=0}^{\infty} \omega_{i,\theta} \, F_{t'}\left(t\left(\frac{k+2i}{k}\right)^{1/2}; k+2i, \mu\right).$$

Doubly Noncentral t: Saddlepoint Approximation

The Daniels and Young (1991) marginal s.p.a. approximation is applicable in this case. Broda and Paolella (2007) derive the closed-form solution to the saddlepoint equation, so that the p.d.f. and c.d.f. of the doubly noncentral t can be virtually instantaneously approximated. See the text for the details and programs.

The s.p.a. simplifies further in the singly noncentral case with $\theta = 0$, and agrees with the results of DiCiccio and Martin (1991).

Due to the computational burden of the density calculation of the (singly) noncentral t involved in the model of Harvey and Siddique (1999), Leon, Rubio and Serna (2005)[14] advocated use of distributions which are easier to calculate. Observe that the saddlepoint approximation to the singly (and doubly) noncentral t solves this problem.

[14] *Autoregressive Conditional Volatility, Skewness and Kurtosis*, The Quarterly Review of Economics and Finance, Vol. 45, pp. 599-618.

Singly Noncentral t Moments

We show in the text that the raw moments of the singly noncentral t r.v. $T = X/\sqrt{Y/k}$ are, for $k > s$ and $s \in \mathbb{N}$,

$$\mathbb{E}\left[T^s\right] = \left(\frac{k}{2}\right)^{s/2} \frac{\Gamma\left(k/2 - s/2\right)}{\Gamma\left(k/2\right)} \sum_{i=0}^{\lfloor s/2 \rfloor} \binom{s}{2i} \frac{(2i)!}{2^i i!} \mu^{s-2i}.$$

In particular, with $T \sim t'(k, \mu)$, for $k > 1$ and $k > 2$, respectively,

$$\mathbb{E}\left[T\right] = \left(\frac{k}{2}\right)^{1/2} \frac{\Gamma\left(k/2 - 1/2\right)}{\Gamma\left(k/2\right)} \mu, \qquad \mathbb{E}\left[T^2\right] = \frac{k}{k-2}\left(1 + \mu^2\right).$$

Doubly Noncentral t Moments

For the doubly noncentral case, with $k > s$ and $s \in \mathbb{N}$,

$$\mathbb{E}\left[T^s\right] = \left(\frac{k}{2}\right)^{s/2} \frac{\Gamma\left(k/2 - s/2\right)}{\Gamma\left(k/2\right)} \, {}_1F_1\left(s/2, k/2, -\theta/2\right) \sum_{i=0}^{\lfloor s/2 \rfloor} \binom{s}{2i} \frac{(2i)!}{2^i i!} \mu^{s-2i},$$

as was derived in Krishnan (1967).

For $s = 1$,

$$\mathbb{E}\left[T\right] = \mu \left(\frac{k}{2}\right)^{1/2} \frac{\Gamma\left(k/2 - 1/2\right)}{\Gamma\left(k/2\right)} \, {}_1F_1\left(1/2, n/2, -\theta/2\right), \quad k > 1,$$

while for $s = 2$,

$$\mathbb{E}\left[T^2\right] = \left(1 + \mu^2\right) \frac{k}{(k-2)} \, {}_1F_1\left(1, k/2, -\theta/2\right) \quad k > 2.$$

Doubly Noncentral t Absolute Moments

Finally, some work shows that

$$\mathbb{E}\left[|T|^m\right] = k^{m/2} \frac{\Gamma\left((k-m)/2\right)\Gamma\left((1+m)/2\right)}{\Gamma\left(k/2\right)\sqrt{\pi}} \, {}_1F_1\left(\frac{m}{2}, \frac{k}{2}, -\frac{\theta}{2}\right) \, {}_1F_1\left(-\frac{m}{2}, \frac{1}{2}; \frac{\mu^2}{2}\right),$$

for $0 < m < k$.

Exercise

As in Ellison (1964, p. 92),[15] let $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$ independent of $Y \sim \sqrt{\chi^2_m/m}$. Show that

$$\mathbb{E}\left[\Phi\left(X + cY\right)\right] = F_T\left(\frac{c}{\sqrt{1+\sigma^2}}; m, \delta\right), \quad \delta = -\frac{\mu}{\sqrt{1+\sigma^2}},$$

where c is any real number, Φ is the standard normal c.d.f., and $T \sim t'\left(m, \delta\right)$. Consider the special case with $c = 0$.

[15] *Two Theorems for Inferences about the Normal Distribution with Applications in Acceptance Sampling*, JASA, Vol. 59, pp. 89-95.

Solution

With $Z \sim \mathrm{N}(0,1)$ independent of X and Y, and recalling the law of total probability, which states that, for event A,

$$\Pr(A) = \int \Pr(A \mid B = b)\, f_B(b)\, \mathrm{d}b = \mathbb{E}_B\left[\Pr(A \mid B = b)\right],$$

$$\begin{aligned}
\mathbb{E}\left[\Phi(X + cY)\right] &= \int_{-\infty}^{\infty}\int_0^{\infty} \Phi(x + cy)\, f_{X,Y}(x,y)\, \mathrm{d}x\mathrm{d}y \\
&= \mathbb{E}_{X,Y}\left[\Pr(Z \le X + cY \mid X = x, Y = y)\right] \\
&= \Pr(Z \le X + cY) = \Pr\left(\frac{Z - X}{Y} \le c\right),
\end{aligned}$$

and since $Z - X =: D \sim \mathrm{N}\left(-\mu, 1 + \sigma^2\right)$ and $D/\sqrt{1+\sigma^2} \sim \mathrm{N}(\delta, 1)$,

$$\Pr\left(\frac{Z - X}{Y} \le c\right) = \Pr\left(T \le \frac{c}{\sqrt{1+\sigma^2}}\right).$$

For $c = 0$,

$$\begin{aligned}
\mathbb{E}\left[\Phi(X)\right] &= \Pr\left(\frac{Z - X}{Y} \le c\right) = \Pr(D \le 0) \\
&= \Pr\left(\frac{D + \mu}{\sqrt{1+\sigma^2}} \le \frac{0 + \mu}{\sqrt{1+\sigma^2}}\right) = \Phi\left(\frac{\mu}{\sqrt{1+\sigma^2}}\right).
\end{aligned}$$

Exercise: Problem 10.7

Let $X_i \stackrel{\text{ind}}{\sim} \mathrm{N}\left(\mu_i, \sigma_i^2\right)$, $i = 1, 2$. Compute the density of $R = \left(X_1/X_2\right)^2$. Then, for $\sigma_1^2 = \sigma_2^2 = 1$ and $\mu_2 = 0$, show that

$$f_R\left(x\right) = \frac{\mathrm{e}^{-\mu_1/2}}{\pi} \frac{1}{\sqrt{x}\left(1+x\right)} \, {}_1F_1\left(1, \frac{1}{2}, \frac{\mu_1}{2} \frac{x}{1+x}\right).$$

Solution

With $X_i \stackrel{\text{ind}}{\sim} \mathrm{N}\left(\mu_i, \sigma_i^2\right)$, $i = 1, 2$, and $R = (X_1/X_2)^2$, let

$$S = \frac{\sigma_2^2}{\sigma_1^2} R = \left(\frac{X_1/\sigma_1}{X_2/\sigma_2}\right)^2,$$

so $S \sim \mathrm{F}(1, 1, \theta_1, \theta_2)$, $\theta_i = \mu_i/\sigma_i$, with $f_S(x)$ given by (10.23) as

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{e^{-\theta_1/2}(\theta_1/2)^i}{i!}\frac{e^{-\theta_2/2}(\theta_2/2)^j}{j!}\frac{x^{i-1/2}}{(1+x)^{1+i+j}}\frac{\Gamma(i+j+1)}{\Gamma(i+1/2)\Gamma(j+1/2)},$$

from which that for R, being just a scale transform, can be easily given.

With $\sigma_1^2 = \sigma_2^2 = 1$ and $\mu_2 = 0$, this reduces to

$$\begin{aligned} f_R(x) &= \frac{1}{\sqrt{\pi}} \sum_{i=0}^{\infty} \frac{e^{-\theta_1/2}(\theta_1/2)^i}{i!} \frac{x^{i-1/2}}{(1+x)^{1+i}} \frac{\Gamma(i+1)}{\Gamma(i+1/2)} \\ &= \frac{e^{-\mu_1/2}}{\sqrt{\pi}} \frac{x^{-1/2}}{1+x} \sum_{i=0}^{\infty} \left(\frac{\mu_1}{2}\frac{x}{1+x}\right)^i \frac{1}{\Gamma(i+1/2)}. \end{aligned}$$

Solution

Then, from (10.2), i.e.,

$$\Gamma\left(i+1/2\right)=\frac{1\cdot 3\cdot 5\cdot\cdots\cdot(2i-1)}{2^i}\sqrt{\pi},$$

$$\begin{aligned}
f_R\left(x\right) &= \frac{e^{-\mu_1/2}}{\pi}\frac{x^{-1/2}}{1+x}\sum_{i=0}^{\infty}\left(\frac{\mu_1}{2}\frac{x}{1+x}\right)^i\frac{2^i}{1\cdot 3\cdot 5\cdot\cdots\cdot(2i-1)}\\
&= \frac{e^{-\mu_1/2}}{\pi}\frac{x^{-1/2}}{1+x}\sum_{i=0}^{\infty}\frac{2^i}{1\cdot 3\cdot 5\cdot\cdots\cdot(2i-1)}\left(\frac{\mu_1}{2}\frac{x}{1+x}\right)^i\\
&= \frac{e^{-\mu_1/2}}{\pi}\frac{x^{-1/2}}{1+x}\sum_{i=0}^{\infty}\frac{i!}{\frac{1}{2}\left(\frac{1}{2}+1\right)\left(\frac{1}{2}+2\right)\cdots\left(\frac{1}{2}+i-1\right)}\frac{1}{i!}\left(\frac{\mu_1}{2}\frac{x}{1+x}\right)^i\\
&= \frac{e^{-\mu_1/2}}{\pi}\frac{1}{\sqrt{x}\left(1+x\right)}\,{}_1F_1\left(1,\frac{1}{2},\frac{\mu_1}{2}\frac{x}{1+x}\right),
\end{aligned}$$

recalling expression (5.25) for the confluent hypergeometric function.