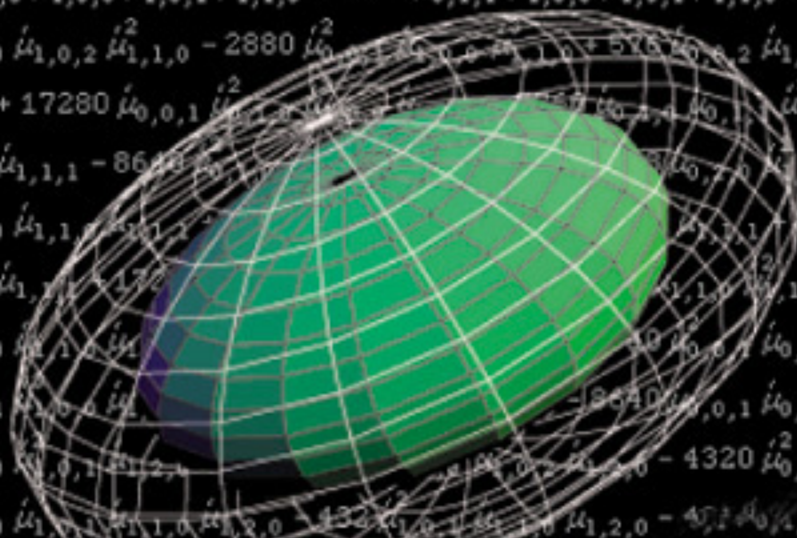
SPRINGER TEXTS IN STATISTICS

MATHEMATICAL STATISTICS *with* *Mathematica*®



COLIN ROSE
MURRAY D. SMITH

Mathematical Statistics with *Mathematica*

Chapter 2 – Continuous Random Variables

Please reference this 2002 edition as:

Rose, C. and Smith, M.D. (2002)
Mathematical Statistics with Mathematica, Springer-Verlag, New York.

Latest edition

For the latest up-to-date edition, please visit: www.mathStatica.com

Chapter 2

Continuous Random Variables

2.1 Introduction

Let the continuous random variable X be defined on a domain of support $\Lambda \subset \mathbb{R}$. Then a function $f : \Lambda \to \mathbb{R}_+$ is a *probability density function* (pdf) if it has the following properties:

$$f(x) > 0 \text{ for all } x \in \Lambda$$

$$\int_\Lambda f(x)\, dx = 1 \tag{2.1}$$

$$P(X \in S) = \int_S f(x)\, dx, \quad \text{for } S \subset \Lambda$$

The *cumulative distribution function* (cdf) of X, denoted $F(x)$, is defined by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(w)\, dw, \qquad -\infty < x < \infty. \tag{2.2}$$

The **mathStatica** function `Prob[x, f]` calculates $P(X \le x)$. Random variable X is said to be a *continuous random variable* if $F(x)$ is continuous. In fact, although our starting point in **mathStatica** is typically to enter a pdf, it should be noted that the fundamental statistical concept is really the cdf, not the pdf. Table 1 summarises some properties of the cdf for a continuous random variable (a and b are constants).

(i)	$0 \le F(x) \le 1$
(ii)	$F(x)$ is a non-decreasing function of x
(iii)	$F(-\infty) = 0,\ F(\infty) = 1$
(iv)	$P(a < X \le b) = F(b) - F(a)$, for $a < b$
(v)	$P(X = x) = 0$
(vi)	$\dfrac{dF(x)}{dx} = f(x)$

Table 1: Properties of the cdf $F(x)$ for a continuous random variable

The *expectation* of a function $u(X)$ is defined to be:

$$E\big[u(X)\big] \;=\; \int_x u(x)\, f(x)\, dx \tag{2.3}$$

The **mathStatica** function `Expect[u, f]` calculates $E[u]$, where $u = u(X)$. Table 2 summarises some properties of the expectation operator, where a and b are again constants.

(i)	$E[a]$	$=$	a
(ii)	$E[a\,u(X)]$	$=$	$a\,E[u(X)]$
(iii)	$E[u(X)+b]$	$=$	$b + E[u(X)]$
(iv)	$E\big[\sum_{i=1}^{n} a_i\, X_i\big]$	$=$	$\sum_{i=1}^{n} a_i\, E[X_i]$

Table 2: Basic properties of the expectation operator

⊕ ***Example 1:*** Maxwell–Boltzmann: The Distribution of Molecular Speed in a Gas

The Maxwell–Boltzmann speed distribution describes the distribution of the velocity X of a random molecule of gas in a closed container. The pdf can be entered directly from **mathStatica**'s *Continuous* palette:

$$\mathbf{f} = \frac{\sqrt{2/\pi}}{\sigma^3}\, \mathbf{x}^2\, e^{-\frac{x^2}{2\sigma^2}}; \qquad \mathbf{domain[f]} = \{\mathbf{x}, 0, \infty\}\ \&\&\ \{\sigma > 0\};$$

From a statistical point of view, the distribution depends on just a single parameter $\sigma > 0$. Formally though, in physics, $\sigma = \sqrt{T\, k_B / m}$ where k_B denotes Boltzmann's constant, T denotes temperature in Kelvin, and m is the mass of the molecule. The cdf $F(x)$ is $P(X \le x)$:

```
F = Prob[x, f]
```

$$-\frac{e^{-\frac{x^2}{2\sigma^2}}\sqrt{\frac{2}{\pi}}\; x}{\sigma} + \mathrm{Erf}\Big[\frac{x}{\sqrt{2}\,\sigma}\Big]$$

Figure 1 plots the pdf (left panel) and cdf (right panel) at three different values of σ.

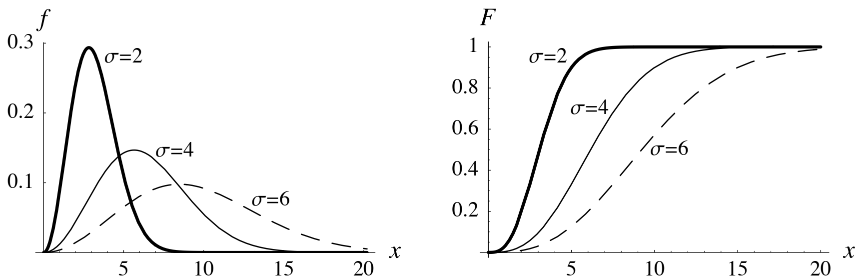


Fig. 1: The Maxwell–Boltzmann pdf (left) and cdf (right), when σ = 2, 4, 6

The average molecular speed is $E[X]$:

```
Expect[x, f]
```

$$2\sqrt{\frac{2}{\pi}}\ \sigma$$

The average kinetic energy per molecule is $E\left[\frac{1}{2}\, m\, X^2\right]$:

```
Expect[1/2 m x², f] /. σ → √(T k_B / m)
```

$$\frac{3\ \mathrm{T}\ \mathrm{k_B}}{2}$$

⊕ ***Example 2:*** The Reflected Gamma Distribution

Some density functions take a piecewise form, such as:

$$f(x) = \begin{cases} f_1(x) & \text{if } x < \alpha \\ f_2(x) & \text{if } x \geq \alpha \end{cases}$$

Such functions are often not smooth, with a kink at the point $x = \alpha$. In *Mathematica*, the natural way to enter such expressions is with the `If[`condition is true, then f_1, else f_2`]` function. That is,

```
f = If[x<α, f1, f2];      domain[f] = {x,-∞,∞}
```

where `f1` and `f2` must still be stated. **mathStatica** has been designed to seamlessly handle `If` statements, without the need for any extra thought or work. In fact, by using this structure, **mathStatica** can solve many integrals that *Mathematica* could not normally solve by itself. To illustrate, let us suppose X is a continuous random variable such that $X = x \in \mathbb{R}$ with pdf

$$f(x) = \begin{cases} \dfrac{(-x)^{\alpha-1}\, e^{x}}{2\,\Gamma[\alpha]} & \text{if } x < 0 \\ \dfrac{x^{\alpha-1}\, e^{-x}}{2\,\Gamma[\alpha]} & \text{if } x \geq 0 \end{cases}$$

where $0 < \alpha < 1$. This is known as a Reflected Gamma distribution, and it nests the standard Laplace distribution as a special case when $\alpha = 1$. We enter $f(x)$ as follows:

```
f = If[x < 0, ((-x)^(α-1) e^x)/(2 Γ[α]), (x^(α-1) e^-x)/(2 Γ[α])];
domain[f] = {x, -∞, ∞} && {α > 0};
```

Here is a plot of $f(x)$ when $\alpha = 1$ and 3:

```
PlotDensity[f /. α → {1, 3}];
```

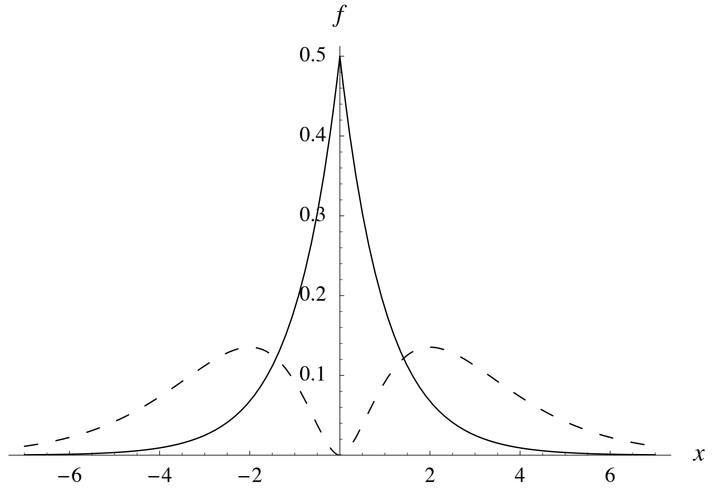


Fig. 2: The pdf of the Reflected Gamma Distribution, when $\alpha = 1$ (——) and 3 (– – –)

Here is the cdf, $P(X \le x)$:

```
cdf = Prob[x, f]
```

$$\text{If}\left[x < 0,\ \frac{\text{Gamma}[\alpha, -x]}{2\,\Gamma[\alpha]},\ 1 - \frac{\text{Gamma}[\alpha, x]}{2\,\Gamma[\alpha]}\right]$$

Figure 3 plots the cdf when $\alpha = 1$ and 3.

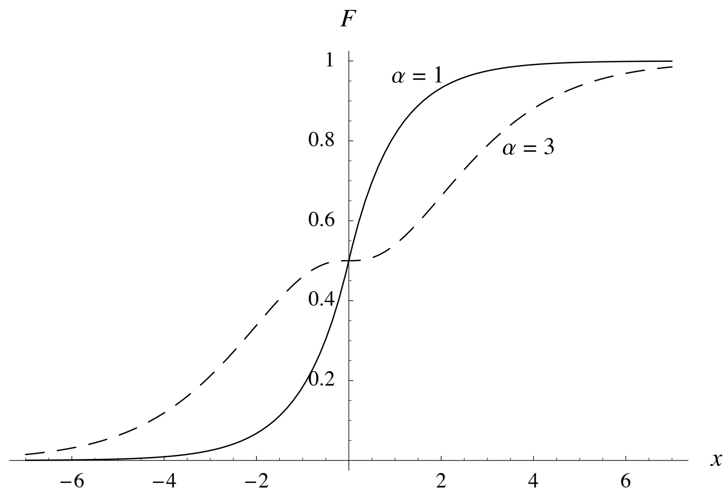


Fig. 3: The cdf of the Reflected Gamma Distribution ($\alpha = 1$ and 3)

2.2 Measures of Location

2.2 A Mean

Let the continuous random variable X have pdf $f(x)$. Then the population mean, or *mean* for short, notated by μ or $\acute{\mu}_1$, is defined by

$$\acute{\mu}_1 = E[X] = \int_x x\, f(x)\, dx \tag{2.4}$$

if the integral converges.

⊕ ***Example 3:*** The Mean for Sinc^2 and Cauchy Random Variables

Let random variable X have a Sinc^2 distribution with pdf $f(x)$, and let Y have a Cauchy distribution with pdf $g(y)$:

$$f = \frac{1}{\pi}\,\frac{\text{Sin}[x]^2}{x^2}; \qquad \text{domain}[f] = \{x, -\infty, \infty\};$$

$$g = \frac{1}{\pi\,(1+y^2)}; \qquad \text{domain}[g] = \{y, -\infty, \infty\};$$

Figure 4 compares the pdf's of the two distributions.

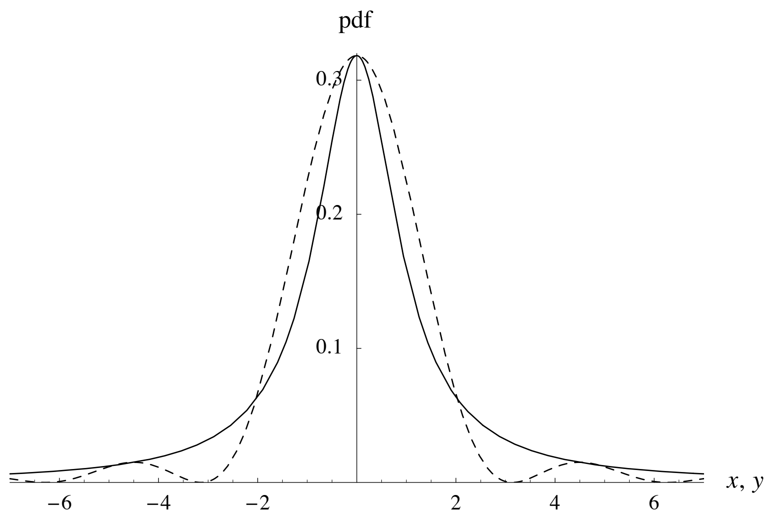


Fig. 4: Cauchy pdf (——) and Sinc^2 pdf (– – –)

The tails of the Sinc^2 pdf are snake-like, and they contact the axis repeatedly at non-zero integer multiples of π.

The mean of the Sinc^2 random variable is $E[X]$:

```
Expect[x, f]
```

0

By contrast, the mean of the Cauchy random variable, $E[Y]$, does not exist:

```
Expect[y, g]
```

— Integrate::idiv :
Integral of $\frac{y}{1+y^2}$ does not converge on {-∞, ∞}.

— Integrate::idiv :
Integral of $\frac{y}{1+y^2}$ does not converge on {-∞, ∞}.

$$\frac{\int_{-\infty}^{\infty} \frac{y}{1+y^2}\, dy}{\pi}$$

2.2 B Mode

Let random variable X have pdf $f(x)$. If $f(x)$ has a local maximum at value x_m, then we say there is a *mode* at x_m. If there is only one mode, then the distribution is said to be unimodal. If the pdf is everywhere continuous and twice differentiable, and there is no corner solution, then a mode is the solution to

$$\frac{d f(x)}{dx} = 0, \qquad \frac{d^2 f(x)}{dx^2} < 0. \tag{2.5}$$

Care should always be taken to check for corner solutions.

⊕ ***Example 4:*** The Mode for a Chi-squared Distribution

Let random variable $X \sim$ Chi-squared(n) with pdf $f(x)$:

$$\mathbf{f} = \frac{x^{n/2-1}\, e^{-x/2}}{2^{n/2}\, \Gamma[\frac{n}{2}]}; \qquad \mathbf{domain[f]} = \{x, 0, \infty\}\ \&\&\ \{n > 0\};$$

The first-order condition for a maximum is obtained via:

```
FOC = D[f, x] // Simplify;     Solve[FOC == 0, x]
```

— Solve::ifun : Inverse functions are being
used by Solve, so some solutions may not be found.

$$\left\{\left\{x \to 0^{\frac{2}{-4+n}}\right\}, \{x \to -2+n\}\right\}$$

Consider the interior solution, $x_m = n - 2$, for $n > 2$. The second-order condition for a maximum, at $x_m = n - 2$, is:

```
SOC = D[f, {x, 2}] /. x → n - 2 // Simplify
```

$$-\frac{2^{-1-\frac{n}{2}}\, e^{1-\frac{n}{2}}\, (-2+n)^{\frac{1}{2}(-4+n)}}{\Gamma[\frac{n}{2}]}$$

which is negative for $n > 2$. Hence, we conclude that x_m is indeed a mode, when $n > 2$. If $n \leq 2$, the mode is the corner solution $x_m = 0$. Figure 5 illustrates the two scenarios by plotting the pdf when $n = 1.98$ and $n = 3$.

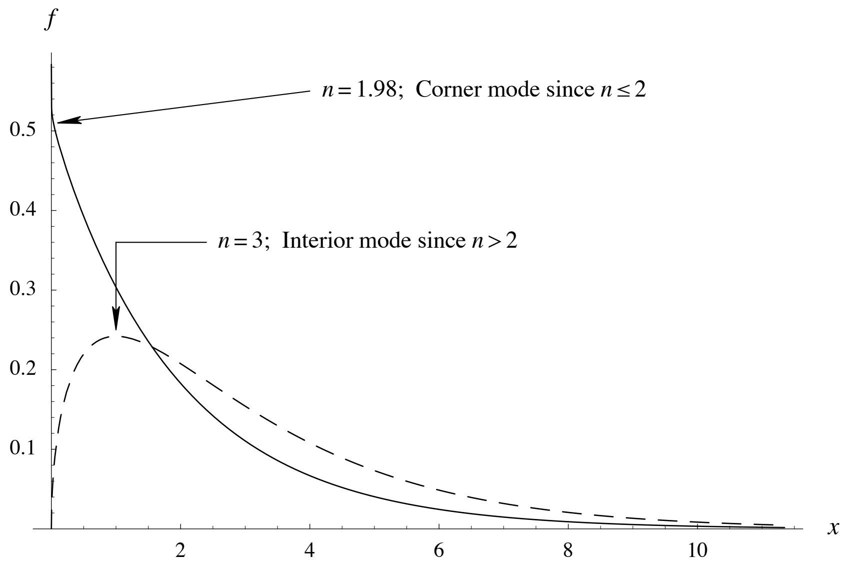


Fig. 5: Corner mode (when $n \leq 2$) and interior mode (when $n > 2$)

2.2 C Median and Quantiles

Let the continuous random variable X have pdf $f(x)$ and cdf $F(x) = P(X \leq x)$. Then, the *median* is the value of X that divides the total probability into two equal halves; *i.e.* the value x at which $F(x) = \frac{1}{2}$. More generally, the p^{th} quantile is the value of X, say x_p, at which $F(x_p) = p$, for $0 < p < 1$. Quantiles are calculated by deriving the inverse cdf, $x_p = F^{-1}(p)$. Ideally, inversion should be done symbolically (algebraically). Unfortunately, for many distributions, symbolic inversion can be difficult, either because the cdf can not be found symbolically and/or because the inverse cdf can not be found. In such cases, one can often resort to numerical methods. Symbolic and numerical inversion are also discussed in §2.6 B and §2.6 C, respectively.

⊕ ***Example 5:*** Symbolic Inversion: The Median for the Pareto Distribution

Let random variable $X \sim$ Pareto(a, b) with pdf $f(x)$:

```
f = a b^a x^-(a+1);    domain[f] = {x, b, ∞} && {a > 0, b > 0};
```

and cdf $F(x)$:

```
F = Prob[x, f]
```

$$1 - \left(\frac{b}{x}\right)^a$$

The median is the value of X at which $F(x) = \frac{1}{2}$:

```
Solve[F == 1/2, x]
```

— Solve::ifun : Inverse functions are being used by Solve, so some solutions may not be found.

$$\left\{\left\{x \to 2^{\frac{1}{a}}\, b\right\}\right\}$$

More generally, if *Mathematica* can find the inverse cdf, the p^{th} quantile is given by:

```
Solve[F == p, x]
```

— Solve::ifun : Inverse functions are being used by Solve, so some solutions may not be found.

$$\{\{x \to b\,(1-p)^{-1/a}\}\}$$

Figure 6 plots the cdf and inverse cdf, when $a = 4$ and $b = 2$.

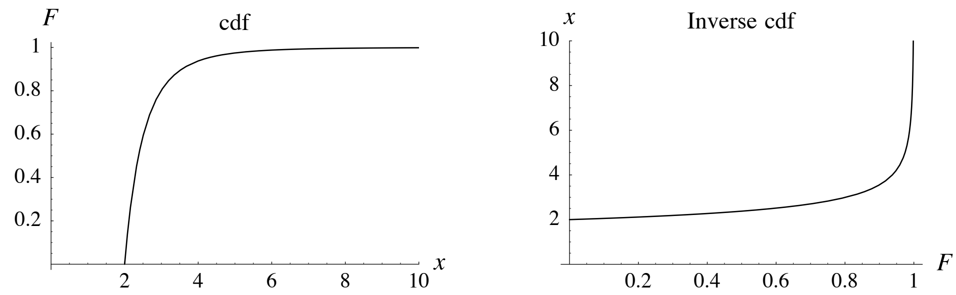


Fig. 6: cdf and inverse cdf

⊕ ***Example 6:*** Numerical Inversion: Quantiles for a Birnbaum–Saunders Distribution

Let $f(x)$ denote the pdf of a Birnbaum–Saunders distribution, with parameters $\alpha = \frac{1}{2}$ and $\beta = 4$:

$$\mathbf{f} \;=\; \frac{e^{-\frac{(x-\beta)^2}{2\alpha^2\beta x}}\,(x+\beta)}{2\,\alpha\,\sqrt{2\,\pi\,\beta}\;x^{3/2}}\;\;/.\;\left\{\alpha \to \frac{1}{2},\;\beta \to 4\right\};$$

```
domain[f] = {x, 0, ∞} && {α > 0, β > 0};
```

Mathematica cannot find the cdf symbolically; that is, `Prob[x, f]` fails. Instead, we can construct a numerical cdf function `NProb`:

```
NProb[w_] := NIntegrate[f, {x, 0, w}]
```

For example, $F(8) = P(X \leq 8)$ is given by:

```
NProb[8]
```

```
0.92135
```

which means that $X = 8$ is approximately the 0.92 quantile. Suppose we want to find the 0.7 quantile: one approach would be to manually try different values of X. As a first guess, how about $X = 6$?

```
NProb[6]
```

```
0.792892
```

Too big. So, try $X = 5$:

```
NProb[5]
```

```
0.67264
```

Too small. And so on. Instead of doing this iterative search manually, we can use *Mathematica*'s `FindRoot` function to automate the search for us. Here, we ask *Mathematica* to search for the value of X at which $F(x) = 0.7$, starting the search by trying $X = 1$ and $X = 10$:

```
sol = FindRoot[NProb[x] == 0.7, {x, {1, 10}}]
```

```
{x → 5.19527}
```

This tells us that $X = 5.19527$ ... is the 0.7 quantile, as we can check by substituting it back into our numerical $F(x)$ function:

```
NProb[x /. sol]
```

```
0.7
```

Care is always required with numerical methods, in part because they are not exact, and in part because different starting points can sometimes lead to different 'solutions'. Finally, note that numerical methods can only be used if the pdf itself is numerical. Thus, numerical methods cannot be used to find quantiles as a function of parameters α and β—the method can only work given numerical values for α and β. ■

2.3 Measures of Dispersion

A number of methods exist to measure the dispersion of the distribution of a random variable X. The most well known is the *variance* of X, defined as the second central moment

$$\text{Var}(X) = \mu_2 = E[(X-\mu)^2] \tag{2.6}$$

where μ denotes the mean $E[X]$. The **mathStatica** function `Var[x, f]` calculates Var(X). The *standard deviation* is the (positive) square root of the variance, and is often denoted σ.[1] Another measure is the *mean deviation* of X, defined as the first absolute central moment

$$E\left[\,|X-\mu|\,\right]. \tag{2.7}$$

The above measures of dispersion are all expressed in terms of the units of X. This can make it difficult to compare the dispersion of one population with another. By contrast, the following statistics are independent of the variable's units of measurement. The *coefficient of variation* is defined by

$$\sigma/\mu\,. \tag{2.8}$$

Gini's coefficient lies within the unit interval; it is discussed in *Example 9*. Alternatively, one can often compare the dispersion of two distributions by standardising them. A *standardised* random variable Z has zero mean and unit variance:

$$Z = \frac{X-\mu}{\sigma}. \tag{2.9}$$

Related measures are $\sqrt{\beta_1}$ and β_2, where

$$\begin{aligned}\sqrt{\beta_1} &= \frac{\mu_3}{\mu_2^{3/2}} = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]\\ \beta_2 &= \frac{\mu_4}{\mu_2^2} = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]\end{aligned} \tag{2.10}$$

Here, the μ_i terms denote central moments, which are introduced in §2.4 A. If a density is not symmetric about μ, it is said to be skewed. A common measure of *skewness* is $\sqrt{\beta_1}$. If the distribution of X is symmetric about μ, then $\mu_3 = E[(X-\mu)^3] = 0$ (assuming μ_3 exists). However, $\mu_3 = 0$ does not guarantee symmetry; Ord (1968) provides examples. Densities with long tails to the right are called *skewed to the right* and they tend to have $\mu_3 > 0$, while densities with long tails to the left are called *skewed to the left* and tend to have $\mu_3 < 0$. *Kurtosis* is commonly said to measure the peakedness of a distribution. More correctly, kurtosis is a measure of both the peakedness (near the centre) and the tail weight of a distribution. Balanda and MacGillivray (1988, p. 116) define kurtosis as "the location- and scale-free movement of probability mass from the shoulders of a distribution into its centre and tails. In particular, this definition implies that peakedness and tail weight are best viewed as components of kurtosis, since any movement of mass from the shoulders into the tails must be accompanied by a movement of mass into the centre if the scale is to be left unchanged." The expression β_2 is Pearson's measure of the *kurtosis* of a distribution. For the Normal distribution, $\beta_2 = 3$, and so the value 3 is often used as a reference point.

⊕ ***Example 7:*** Mean Deviation for the Chi-squared(n) Distribution

Let $X \sim$ Chi-squared(n) with pdf $f(x)$:

$$\mathbf{f = \frac{x^{n/2-1}\, e^{-x/2}}{2^{n/2}\, \Gamma[\frac{n}{2}]};\qquad domain[f] = \{x,\ 0,\ \infty\}\ \&\&\ \{n > 0\};}$$

The mean μ is:

```
μ = Expect[x, f]
```

```
n
```

The mean deviation is $E\big[\,|X - \mu|\,\big]$. Evaluating this directly using `Abs[]` fails to yield a solution:

```
Expect[Abs[x - μ], f]
```

$$\frac{2^{-n/2}\int_0^\infty e^{-x/2}\, x^{-1+\frac{n}{2}}\, \text{Abs}[n - x]\, dx}{\Gamma[\frac{n}{2}]}$$

In fact, quite generally, *Mathematica* Version 4 is not very successful at integrating expressions containing absolute values. Fortunately, **mathStatica**'s support for `If[`*a*, *b*, *c*`]` statements provides a backdoor way of handling absolute values—to see this, express $y = |x - \mu|$ as:

```
y = If[x < μ, μ - x, x - μ];
```

Then the mean deviation $E\big[\,|X - \mu|\,\big]$ is given by:[2]

```
Expect[y, f]
```

$$\frac{4\,\text{Gamma}[1 + \frac{n}{2}, \frac{n}{2}] - 2\,n\,\text{Gamma}[\frac{n}{2}, \frac{n}{2}]}{\Gamma[\frac{n}{2}]}$$

⊕ ***Example 8:*** β_1 and β_2 for the Weibull Distribution

Let $X \sim$ Weibull(a, b) with pdf $f(x)$:

$$\mathbf{f} = \frac{\mathbf{a\,x^{a-1}}}{\mathbf{b^a}\, e^{(\frac{x}{b})^a}}; \quad \mathbf{domain[f] = \{x,\ 0,\ \infty\}\ \&\&\ \{a > 0,\ b > 0\};}$$

Here, a is termed the shape parameter, and b is termed the scale parameter. The mean μ is:

$$\mu = \mathbf{Expect[x,\ f]}$$

$$\mathrm{b}\,\Gamma\left[1 + \frac{1}{a}\right]$$

while the second, third and fourth central moments are:

$$\{\mu_2,\ \mu_3,\ \mu_4\} = \mathbf{Expect[(x-\mu)^{\{2,\,3,\,4\}},\ f];}$$

Then, β_1 and β_2 are given by:

$$\{\beta 1,\ \beta 2\} = \left\{\frac{\mu_3^2}{\mu_2^3},\ \frac{\mu_4}{\mu_2^2}\right\}$$

$$\left\{\frac{\left(2\,\Gamma[1+\frac{1}{a}]^3 - \frac{6\,\Gamma[\frac{1}{a}]\,\Gamma[\frac{2}{a}]}{a^2} + \Gamma[\frac{3+a}{a}]\right)^2}{\left(-\Gamma[1+\frac{1}{a}]^2 + \Gamma[\frac{2+a}{a}]\right)^3},\ \frac{-\frac{3\,\Gamma[\frac{1}{a}]\,\left(\Gamma[\frac{1}{a}]^3 - 4\,a\,\Gamma[\frac{1}{a}]\,\Gamma[\frac{2}{a}] + 4\,a^2\,\Gamma[\frac{3}{a}]\right)}{a^4} + \Gamma[\frac{4+a}{a}]}{\left(-\Gamma[1+\frac{1}{a}]^2 + \Gamma[\frac{2+a}{a}]\right)^2}\right\}$$

Note that both β_1 and β_2 only depend on the shape parameter a; the scale parameter b has disappeared, as per intuition. Figure 7 plots β_1 and β_2 for different values of parameter a.

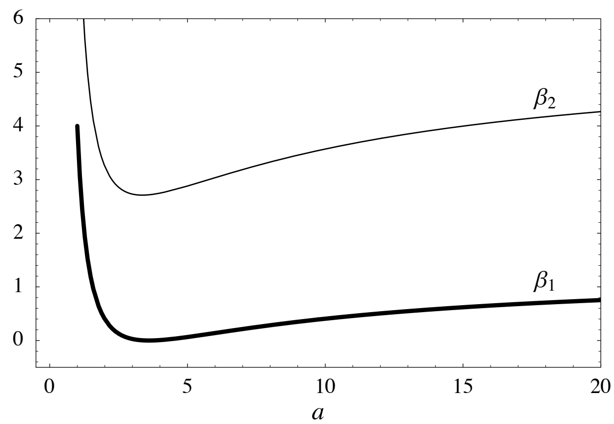


Fig. 7: β_1 and β_2 for the Weibull distribution (plotted as a function of parameter a)

Note that the symbols μ_2, μ_3 and μ_4 are 'reserved' for use by **mathStatica**'s moment converter functions. To avoid any confusion, it is best to `Unset` them:

```
μ = .;   μ2 =.;   μ3 =.;   μ4 =.;
```

prior to leaving this example. ■

⊕ ***Example 9:*** The Lorenz Curve and the Gini Coefficient

```
ClearAll[a, b, p, x, u, f, F]
```

Let X be a positive random variable with pdf $f(x)$ and cdf $F(x)$, and let $p = F(x)$. The *Lorenz curve* is the graph of $L(p)$ against p, where

$$L(p) = \frac{1}{E[X]} \int_0^p F^{-1}(u)\, du \tag{2.11}$$

and where $F^{-1}(\cdot)$ denotes the inverse cdf. In economics, the Lorenz curve is often used to measure the extent of inequality in the distribution of income. To illustrate, suppose income X is Pareto distributed with pdf $f(x)$:

```
f = a b^a x^-(a+1);   domain[f] = {x, b, ∞} && {a > 0, b > 0};
```

and cdf $F(x)$:

```
F = Prob[x, f]
```

$$1 - \left(\frac{b}{x}\right)^a$$

The inverse cdf is found by solving the equation $p = F(x)$ in terms of x:

```
Solve[p == F, x]
```

```
— Solve::ifun : Inverse functions are being
    used by Solve, so some solutions may not be found.
```

$\{\{x \to b\,(1-p)^{-1/a}\}\}$

Equation (2.11) requires that the mean of X exists:

```
mean = Expect[x, f]
```

```
— This further assumes that: {a > 1}
```

$$\frac{a\,b}{-1+a}$$

… so we shall impose the tighter restriction $a > 1$. We can now evaluate (2.11):

```
LC = (1/mean) Integrate[b (1 - u)^(-1/a), {u, 0, p}]
```

$$\frac{(-1+a)\left(\frac{a}{-1+a}+\frac{a\,(1-p)^{-1/a}\,(-1+p)}{-1+a}\right)}{a}$$

Note that the solution does not depend on the location parameter *b*. The solution can be simplified further:

```
LC = FullSimplify[ LC, {0 < p < 1, a > 1}]
```

$$1-(1-p)^{1-\frac{1}{a}}$$

The Lorenz curve is a plot of `LC` as a function of *p*, as illustrated in Fig. 8. The horizontal axis (p) measures quantiles of the population sorted by income; that is, $p = 0.25$ denotes the poorest 25% of the population. The vertical axis, $L(p)$, measures what proportion of society's total income accrues to the poorest *p* people. In the case of Fig. 8, where $a = 2$, the poorest 50% of the population earn only 29% of the total income:

```
LC /. {a → 2, p → .50}

0.292893
```

The 45° line, $L(p) = p$, represents a society with absolute income equality. By contrast, the line $L(p) = 0$ represents a society with absolute income inequality: here, all the income accrues to just one person.

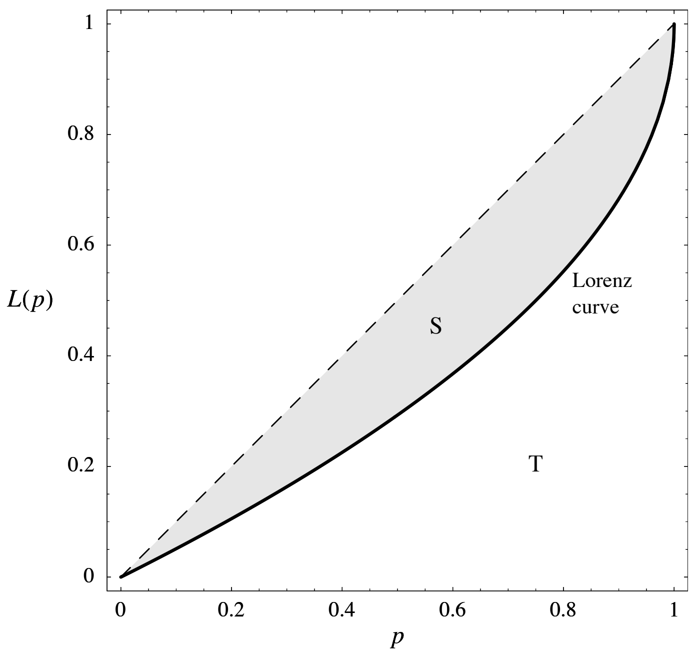


Fig. 8: The Lorenz Curve for a Pareto distribution ($a = 2$)

The *Gini coefficient* is often used in economics to quantify the extent of inequality in the distribution of income. The advantage of the Gini coefficient over the variance as a measure of dispersion is that the Gini coefficient is unitless and lies within the unit interval. Let S denote the shaded area in Fig. 8, and let T denote the area below the Lorenz curve. The Gini coefficient (GC) is defined by the ratio $\text{GC} = \frac{S}{S+T} = \frac{S}{1/2} = 2\,S$. That is, GC = twice the shaded area. Since it is easy to compute area T, and since $S = \frac{1}{2} - T$, we use GC = 2 S = 1 − 2 T. Then, for our Pareto example, the Gini coefficient is:

```
1 - 2 Integrate[ LC, {p, 0, 1}, Assumptions → a > 1] //
 Simplify
```

```
    1
---------
-1 + 2 a
```

This corresponds to a Gini coefficient of $\frac{1}{3}$ for Fig. 8 where $a = 2$. If $a = 1$, then GC = 1 denoting absolute income inequality. As parameter a increases, the Lorenz curve shifts toward the 45° line, and the Gini coefficient tends to 0, denoting absolute income equality. ∎

2.4 Moments and Generating Functions

2.4 A Moments

The r^{th} *raw moment* of the random variable X is denoted by $\acute{\mu}_r(X)$, or $\acute{\mu}_r$ for short, and is defined by

$$\acute{\mu}_r = E[X^r]. \tag{2.12}$$

Note that $\acute{\mu}_0 = 1$, since $E[X^0] = E[1] = 1$. The first moment, $\acute{\mu}_1 = E[X]$, is the *mean* of X, and it is also denoted μ.

The r^{th} *central moment* μ_r is defined by

$$\mu_r = E\left[(X - \mu)^r\right] \tag{2.13}$$

where $\mu = E[X]$. This is also known as the r^{th} *moment about the mean*. Note that $\mu_0 = 1$, since $E\left[(X-\mu)^0\right] = E[1]$. Similarly, $\mu_1 = 0$, since $E\left[(X-\mu)^1\right] = E[X] - \mu$. The second central moment, $\mu_2 = E\left[(X-\mu)^2\right]$, is known as the *variance* of X, and is denoted $\text{Var}(X)$. The *standard deviation* of X is the (positive) square root of the variance, and is often denoted σ. Moments can also be obtained via generating functions; see §2.4 B. Further, the various types of moments can be expressed in terms of one another; this is discussed in §2.4 G.

⊕ ***Example 10:*** Raw Moments for a Standard Normal Random Variable

Let $X \sim N(0, 1)$ with pdf $f(x)$:

```
f = e^(-x²/2)/Sqrt[2 π];      domain[f] = {x, -∞, ∞};
```

The r^{th} raw moment $E[X^r]$ is given by:

```
sol = Expect[x^r, f]

— This further assumes that: {r > -1}

2^(1/2 (-2+r)) (1 + (-1)^r) Γ[(1+r)/2] / Sqrt[π]
```

Then, the first 15 raw moments are given by:

```
sol /. r → Range[15]

{0, 1, 0, 3, 0, 15, 0, 105, 0, 945, 0, 10395, 0, 135135, 0}
```

The odd moments are all zero, because the standard Normal distribution is symmetric about zero. ■

2.4 B The Moment Generating Function

The *moment generating function* (mgf) of a random variable X is a function that may be used to generate the moments of X. In particular, the mgf $M_X(t)$ is a function of a real-valued dummy variable t. When no confusion is possible, we denote $M_X(t)$ by $M(t)$. We first consider whether or not the mgf exists, and then show how moments may be derived from it, if it exists.

Existence: Let X be a random variable, and $t \in \mathbb{R}$ denote a dummy variable. Let $\underline{t}$ and $\bar{t}$ denote any two real-valued constants such that $\underline{t} < 0$ and $\bar{t} > 0$; thus, the open interval $(\underline{t}, \bar{t})$ includes zero in its interior. Then, the mgf is given by

$$M(t) = E\left[e^{tX}\right] \tag{2.14}$$

provided $E[e^{tX}] \in \mathbb{R}_+$ for all t in the chosen interval $\underline{t} < t < \bar{t}$. The condition that $M(t)$ be positive real for all $t \in (\underline{t}, \bar{t})$ ensures that $M(t)$ is differentiable with respect to t at zero. Note that when $t = 0$, $M(0)$ is always equal to 1. However, $M(t)$ may fail to exist for $t \neq 0$.

Generating moments: Let X be a random variable for which the mgf $M(t)$ exists. Then, the r^{th} raw moment of X is obtained by differentiating the mgf r times with respect to t, followed by setting $t = 0$ in the resulting formula:

$$\acute{\mu}_r = \left. \frac{d^r M(t)}{d t^r} \right|_{t=0} . \tag{2.15}$$

Proof: If $M(t)$ exists, then $M(t)$ is 'r-times' differentiable at $t = 0$ (for integer $r > 0$) and $\frac{d E[e^{tX}]}{d t} = E\left[\frac{d e^{tX}}{d t}\right]$ for all $t \in (\underline{t}, \bar{t})$ (Mittelhammer (1996, p. 142)). Hence,

$$\left. \frac{d^r E[e^{tX}]}{d t^r} \right|_{t=0} = \left. E\left[\frac{d^r e^{tX}}{d t^r}\right] \right|_{t=0} = \left. E\left[X^r e^{tX}\right] \right|_{t=0} = E\left[X^r\right] \quad \square$$

Using **mathStatica**, the expectation $E[e^{tX}]$ can be found in the usual way with `Expect`. However, before using the obtained solution as the mgf of X, one must check that the mgf definition (2.14) is satisfied; *i.e.* that $M(t)$ is positive real for all $t \in (\underline{t}, \bar{t})$.

⊕ ***Example 11:*** The mgf of the Normal Distribution

Let $X \sim \text{Normal}(\mu, \sigma^2)$. Derive the mgf of X, and derive the first 4 raw moments from it.

Solution: Input the pdf of X:

$$\texttt{f} = \frac{1}{\sigma\sqrt{2\pi}} \texttt{Exp}\left[-\frac{(x-\mu)^2}{2\sigma^2}\right];$$

```
domain[f] = {x, -∞, ∞} && {μ ∈ Reals, σ > 0};
```

Evaluating (2.14), we find:

```
M = Expect[e^(t x), f]
```

$$e^{t\mu + \frac{t^2\sigma^2}{2}}$$

By inspection, `M` $\in \mathbb{R}_+$ for all $t \in \mathbb{R}$, and `M` = 1 when $t = 0$. Thus, `M` corresponds to the mgf of X. Then, to determine say $\acute{\mu}_2$ from `M`, we apply (2.15) as follows:

```
D[M, {t, 2}] /. t → 0
```

$\mu^2 + \sigma^2$

More generally, to determine $\acute{\mu}_r$, $r = 1, \ldots, 4$, from `M`:

```
Table[ D[M, {t, r}] /. t → 0, {r, 4}]
```

$\{\mu,\ \mu^2 + \sigma^2,\ \mu^3 + 3\mu\sigma^2,\ \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4\}$

⊕ ***Example 12:*** The mgf of the Uniform Distribution

Let $X \sim$ Uniform(0, 1). Derive the mgf of X, and derive the first 4 raw moments from it.

Solution: Input the pdf of X, and derive `M`:

```
f = 1;    domain[f] = {x, 0, 1} ;    M = Expect[ⅇ^(t x), f]
```

$$\frac{-1 + e^{t}}{t}$$

Figure 9 plots `M` in the neighbourhood of $t = 0$.

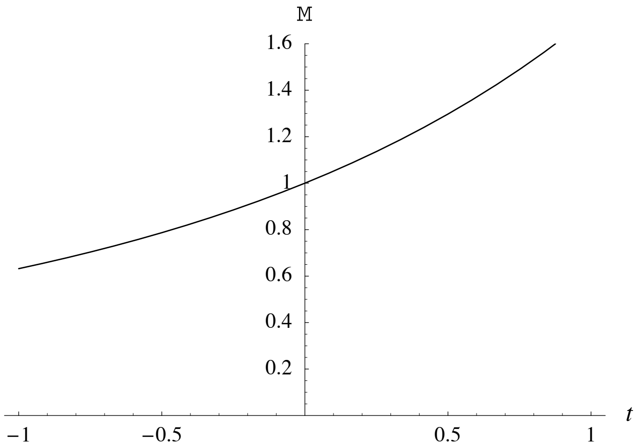


Fig. 9: Function `M` for $-1 < t < 1$

Clearly, $\text{M} \in \mathbb{R}_+$ in a neighbourhood of values about $t = 0$. At the particular value $t = 0$, the plot seems to indicate that $\text{M} = 1$. If we input `M/.t→0`, *Mathematica* replaces t with 0 to yield 0/0:

```
M /. t → 0
```

```
— Power::infy : Infinite expression 1/0 encountered.
— ∞::indet :
   Indeterminate expression 0 ComplexInfinity encountered.
```

```
Indeterminate
```

To correctly determine the value of `M` at $t = 0$, L'Hôpital's rule should be applied. This rule is incorporated into *Mathematica*'s `Limit` function:

```
Limit[M, t → 0]
```

```
1
```

Thus, $\text{M} = 1$ when $t = 0$, as required. Since all requirements of the mgf definition are now satisfied, `M` is the mgf of X.

To determine the first 4 raw moments of X, we again apply (2.15), but this time in tandem with the `Limit` function:

```
Table[ Limit[ D[M, {t, r}] , t → 0],  {r, 4}]
```

$$\left\{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}\right\}$$

More generally, $E[X^r] = \frac{1}{1+r}$, as we can verify with `Expect[x^r, f]`. ■

⊕ ***Example 13:*** The mgf of the Pareto Distribution?

Let X be Pareto distributed with shape parameter $a > 0$ and location parameter $b > 0$. Does the mgf of X exist?

Solution: Input the pdf of X via the **mathStatica** palette:

```
f = a b^a x^-(a+1);   domain[f] = {x, b, ∞}  &&  {a > 0, b > 0};
```

The solution to $M(t) = E[e^{tX}]$ is given by **mathStatica** as:

```
M = Expect[e^(t x), f]
```

```
a ExpIntegralE[1 + a, -b t]
```

If we consult *Mathematica*'s on-line help system on `ExpIntegralE`, we see that the `ExpIntegralE` function is complex-valued if the value of its second argument, $-b\,t$, is negative. Since $b > 0$, `M` will be complex-valued for any positive value assigned to t. To illustrate, suppose parameters a and b are given specific values, and `M` is evaluated for various values of $t > 0$:

```
M /. {a → 5, b → 1} /. t → {.2, .4, .6, .8}
```

```
{1.28704 - 0.0000418879 i, 1.66642 - 0.00134041 i,
 2.17384 - 0.0101788 i, 2.85641 - 0.0428932 i}
```

Hence, the requirement that `M` must be positive real in an open interval that includes the origin is not satisfied. Therefore, the mgf of X does not exist. The non-existence of the mgf does not necessarily mean that the moments do not exist. The Pareto is a case in point, for from:

```
Expect[x^r, f]
```

```
— This further assumes that: {a > r}
```

$$\frac{a\,b^r}{a-r}$$

… we see that the raw moment $\acute{\mu}_r$ exists, under the given conditions. ■

2.4 C The Characteristic Function

As *Example 13* illustrated, the mgf of a random variable does not have to exist. This may occur if e^{tx} is unbounded (see (2.14)). However, the function e^{itx}, where i denotes the unit imaginary number, does not suffer from unboundedness. On an Argand diagram, for any $t \in \mathbb{R}$, e^{itx} takes values on the unit circle. This leads to the so-called *characteristic function* (cf) of random variable X, which is defined as

$$C(t) = E[e^{itX}]. \tag{2.16}$$

The cf of a random variable exists for any choice of $t \in \mathbb{R}$ that we may wish to make; note $C(0) = 1$. If the mgf of a random variable exists, the relationship between the cf and the mgf is simply $C(t) = M(i t)$. Analogous to (2.15), raw moments can be obtained from the cf via

$$\acute{\mu}_r = i^{-r} \left. \frac{d^r C(t)}{d t^r} \right|_{t=0} \tag{2.17}$$

provided $\acute{\mu}_r$ exists.

⊕ ***Example 14:*** The cf of the Normal Distribution

Let $X \sim N(\mu, \sigma^2)$. Determine the cf of X.

Solution: Input the pdf of X:

$$f = \frac{1}{\sigma \sqrt{2\pi}} \operatorname{Exp}\left[-\frac{(x-\mu)^2}{2\sigma^2}\right];$$

```
domain[f] = {x, -∞, ∞} && {μ ∈ Reals, σ > 0};
```

Since we know from *Example 11* that the mgf exists, the cf of X can be obtained via $C(t) = M(i t)$. This sometimes works better in *Mathematica* than trying to evaluate `Expect[ⅇ^(ⅈ t x), f]` directly:

```
cf = Expect[ⅇ^(t x), f] /. t → ⅈ t
```

$$e^{i t \mu - \frac{t^2 \sigma^2}{2}}$$

Then, the first 4 moments are given by:

```
Table[ⅈ^-r D[cf, {t, r}] /. t → 0, {r, 4}] // Simplify
```

$$\{\mu,\ \mu^2 + \sigma^2,\ \mu^3 + 3\mu\sigma^2,\ \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4\}$$

⊕ ***Example 15:*** The cf of the Lindley Distribution

Let the random variable X be Lindley distributed with parameter $\delta > 0$. Derive the cf, and derive the first 4 raw moments from it.

Solution: Input the pdf of X from the **mathStatica** palette:

$$\mathbf{f = \frac{\delta^2}{\delta + 1} (x + 1)\, e^{-\delta x}; \quad domain[f] = \{x, 0, \infty\} \&\& \{\delta > 0\};}$$

The cf is given by

```
cf = Expect[e^(i t x), f]
```

— This further assumes that: {Im[t] == 0}

$$\frac{\delta^2 (1 - i t + \delta)}{(1 + \delta) (-i t + \delta)^2}$$

The condition on t output by **mathStatica** is not relevant here, for we restrict the dummy variable t to the real number line. The first 4 raw moments of X are given by:

```
Table[i^-r D[cf, {t, r}] /. t → 0, {r, 4}] // Simplify
```

$$\left\{\frac{2 + \delta}{\delta + \delta^2}, \frac{2 (3 + \delta)}{\delta^2 (1 + \delta)}, \frac{6 (4 + \delta)}{\delta^3 (1 + \delta)}, \frac{24 (5 + \delta)}{\delta^4 (1 + \delta)}\right\}$$

⊕ ***Example 16:*** The cf of the Pareto Distribution

Let X be Pareto distributed with shape parameter $a = 4$ and location parameter $b = 1$. Derive the cf, and from it, derive those raw moments which exist.

Solution: The Pareto pdf is:

$$\mathbf{f = a\, b^a\, x^{-(a+1)}; \quad domain[f] = \{x, b, \infty\} \&\& \{a > 0, b > 0\};}$$

When $a = 4$ and $b = 1$, the solution to the cf of X is:

```
cf = Expect[e^(i t x), f /. {a → 4, b → 1}]
```

$$\frac{1}{6} \left(e^{i t} (6 - i t (-2 + t (-i + t))) + t^4 \,\text{Gamma}[0, -i t]\right)$$

From *Example 13*, we know that the mgf of X does not exist. However, the moments of X up to order $r < a = 4$ do exist, which we obtain from the cf by applying (2.17):

```
Table[ Limit[i^-r D[cf, {t, r}], t → 0], {r, 4}]
```

$$\left\{\frac{4}{3}, 2, 4, \infty\right\}$$

Notice that we have utilised `Limit` to obtain the moments here, so as to avoid the 0/0 problem discussed in *Example 12*. ■

2.4 D Properties of Characteristic Functions (and mgf's)

§2.4 B and §2.4 C illustrated how the mgf and cf can be used to generate the moments of a random variable. A second (and more important) application of the mgf and cf is to prove that a random variable has a specific distribution. This methodology rests on the Uniqueness Theorem, which we present here using characteristic functions; of course, the theorem also applies to moment generating functions, provided the mgf exists, since then $C(t) = M(i\,t)$.

Uniqueness Theorem: There is a one-to-one correspondence between the cf and the pdf of a random variable.

Proof: The pdf determines the cf via (2.16). The cf determines the pdf via the Inversion Theorem below.

The Uniqueness Theorem means that if two random variables X and Y have the same distribution, then X and Y must have the same mgf. Conversely, if they have the same mgf, then they must have the same distribution. The following results can be especially useful when applying the Uniqueness Theorem. We present these results as the MGF Theorem, which holds provided the mgf exists. A similar result holds, of course, for any cf, with t replaced by $i\,t$.

MGF Theorem: Let random variable X have mgf $M_X(t)$, and let a and b denote constants. Then

$M_{X+a}(t) = e^{t\,a}\,M_X(t)$ *Proof:* $M_{X+a}(t) = E[e^{t(X+a)}] = e^{t\,a}\,M_X(t)$

$M_{b\,X}(t) = M_X(b\,t)$ *Proof:* $M_{b\,X}(t) = E[e^{t(b\,X)}] = E[e^{(t\,b)X}] = M_X(t\,b)$

$M_{a+b\,X}(t) = e^{t\,a}\,M_X(b\,t)$ *Proof:* via above.

Further, let $(X_1, \ldots, X_n)$ be independent random variables with mgf's $M_{X_i}(t)$, $i = 1, \ldots, n$, and let $Y = \sum_{i=1}^{n} X_i$. Then

$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$ *Proof:* via independence (see Table 3 of Chapter 6).

If we can match the functional form of $M_Y(t)$ with a well-known moment generating function, then we know the distribution of Y. This matching is usually done using a textbook that lists the mgf's for well-known distributions. Unfortunately, the matching process is often neither easy nor obvious. Moreover, if the pdf of Y is not well-known (or not listed in the textbook), the matching may not be possible. Instead of trying to match $M_Y(t)$ in a textbook appendix, we can (in theory) derive the pdf that is associated with it by means of the Inversion Theorem. This is particularly important if the derived cf is not of a standard (or common) form. Recall that the characteristic function (cf) is defined by

$$C(t) \;=\; \int_{-\infty}^{\infty} e^{i t x} f(x)\, dt\,. \tag{2.18}$$

Then, the Inversion Theorem is given by:

Inversion Theorem: The characteristic function $C(t)$ uniquely determines the pdf $f(x)$ via

$$f(x) \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i t x}\, C(t)\, dt \tag{2.19}$$

Proof: See Roussas (1997, p. 142) or Stuart and Ord (1994, p. 126).

If the mgf exists, one can replace $C(t)$ with $M(i\,t)$ in (2.19). Inverting a characteristic function is often computationally difficult. With *Mathematica*, one can take two approaches: symbolic inversion and numerical inversion.

Symbolic inversion: If we think of (2.18) as the Fourier transform $f(x) \to C(t)$, then (2.19) is the inverse Fourier transform $C(t) \to f(x)$ which can be implemented in *Mathematica* via:

InverseFourierTransform[cf, *t*, *x*, FourierParameters→{1,1}]

To further automate this mapping, we shall create a function InvertCF[$t \to x$, cf]. Moreover, we shall allow this function to take an optional third argument, InvertCF[$t \to x$, cf, *assume*], which we can use to make assumptions about x, such as $x > 0$, or $x \in$ Reals. Here is the code for InvertCF:

```
InvertCF[t_ → x_, cf_, Assum_:{}] :=
  Module[{sol},
    sol = InverseFourierTransform[cf, t, x,
                                  FourierParameters→{1,1}];
    If[Assum === {}, sol, FullSimplify[sol, Assum]]]
```

Numerical inversion: There are many characteristic functions that *Mathematica* cannot invert symbolically. In such cases, we can resort to numerical methods. We can automate the inversion (2.19) $C(t) \to f(x)$ using numerical integration, by constructing a function NInvertCF[$t \to x$, cf]:

```
NInvertCF[t_ → x_, cf_] :=
   1
  ─── NIntegrate[ ⅇ^(-ⅈ t x) cf, {t, -∞, 0, ∞},
  2 π
                      Method → DoubleExponential]
```

The syntax {t,-∞,0,∞} tells *Mathematica* to check for singularities at 0.

⊕ ***Example 17:*** Linnik Distribution

The distribution whose characteristic function is

$$C(t) = \frac{1}{1 + |t|^{\alpha}}, \qquad t \in \mathbb{R}, \quad 0 < \alpha \le 2 \tag{2.20}$$

is known as a Linnik distribution; this is also known as an α-Laplace distribution. The standard Laplace distribution is obtained when $\alpha = 2$. Consider the case $\alpha = \frac{3}{2}$:

```
cf = 1 / (1 + Abs[t]^(3/2));
```

Inverting the cf symbolically yields the pdf $f(x)$:

```
f = InvertCF[t → x, cf]
```

$$\frac{1}{4\sqrt{3}\,\pi^{7/2}} \text{MeijerG}\left[\left\{\left\{\frac{1}{12}, \frac{1}{3}, \frac{7}{12}\right\}, \{\}\right\}, \left\{\left\{0, \frac{1}{12}, \frac{1}{3}, \frac{1}{3}, \frac{7}{12}, \frac{2}{3}, \frac{5}{6}\right\}, \left\{\frac{1}{6}, \frac{1}{2}\right\}\right\}, \frac{x^6}{46656}\right]$$

where `domain[f]={x,-∞,∞}`. Figure 10 compares the $\alpha = \frac{3}{2}$ pdf to the $\alpha = 2$ pdf.

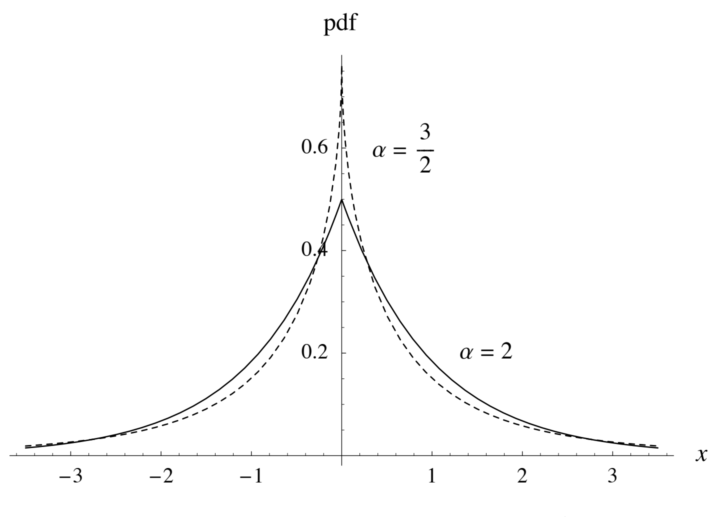


Fig. 10: The pdf of the Linnik distribution, when $\alpha = \frac{3}{2}$ and 2

⊕ ***Example 18:*** The Sum of Uniform Random Variables

Let $(X_1, \ldots, X_n)$ be independent Uniform(0, 1) random variables, each with characteristic function $C(t) = \frac{e^{it}-1}{it}$. It follows from the MGF Theorem that the cf of $Y = \sum_{i=1}^{n} X_i$ is:

```
cf = ((e^(i t) - 1) / (i t))^n;
```

The pdf of Y is known as the Irwin–Hall distribution, and it can be obtained in *Mathematica*, for a given value of n, by inverting the characteristic function cf. For instance, when $n = 1, 2, 3$, the pdf's are, respectively, f1, f2, f3:

```
{f1, f2, f3} = InvertCF[t → y, cf /. n → {1, 2, 3}, y > 0]
```

$$\left\{\frac{1}{2}\,(1 + \text{Sign}[1 - y]),\ \frac{1}{2}\,(y + \text{Abs}[-2 + y] - 2\,\text{Abs}[-1 + y]),\ \frac{1}{4}\,(y^2 + 3\,(-1 + y)^2\,\text{Sign}[1 - y] + (-3 + y)^2\,\text{Sign}[3 - y] + 3\,(-2 + y)^2\,\text{Sign}[-2 + y])\right\}$$

Figure 11 plots the three pdf's. When $n = 1$, we obtain the Uniform(0, 1) distribution, $n = 2$ yields a Triangular distribution, while $n = 3$ already looks somewhat bell-shaped.

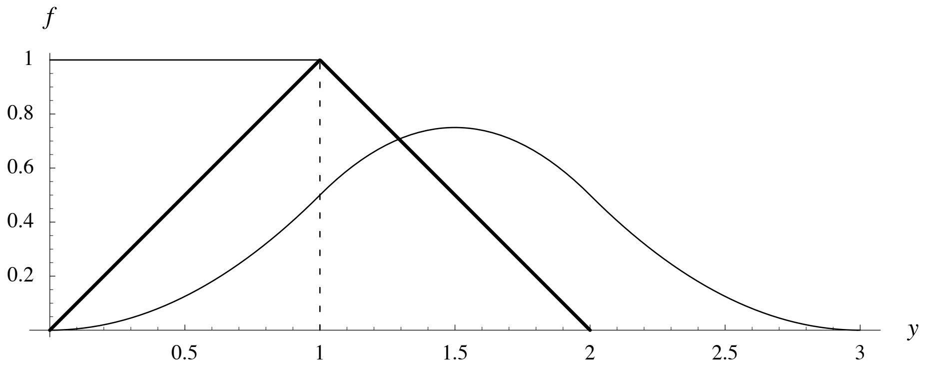


Fig. 11: The pdf of the sum of n Uniform(0, 1) random variables, when $n = 1, 2, 3$

⊕ ***Example 19:*** Numerical Inversion

Consider the distribution whose characteristic function is:

$$\text{cf} = e^{-\frac{t^2}{2}} + \sqrt{\frac{\pi}{2}}\; t\left(\text{Erf}\left[\frac{t}{\sqrt{2}}\right] - \text{Sign}[t]\right);$$

Alas, *Mathematica* Version 4 cannot invert this cf symbolically; that is, InvertCF[t→x, cf] fails. However, by using the NInvertCF function defined above, we can numerically invert the cf at a specific point such as $x = 2.9$, which yields the pdf evaluated at $x = 2.9$:

```
NInvertCF[t → 2.9, cf]
```

0.0467289 + 0. i

By doing this at many points, we can plot the pdf:

```
Plot[ NInvertCF[t → x, cf], {x, -10, 10},
 AxesLabel → {"x", "pdf"}, PlotRange → {0, .21}];
```

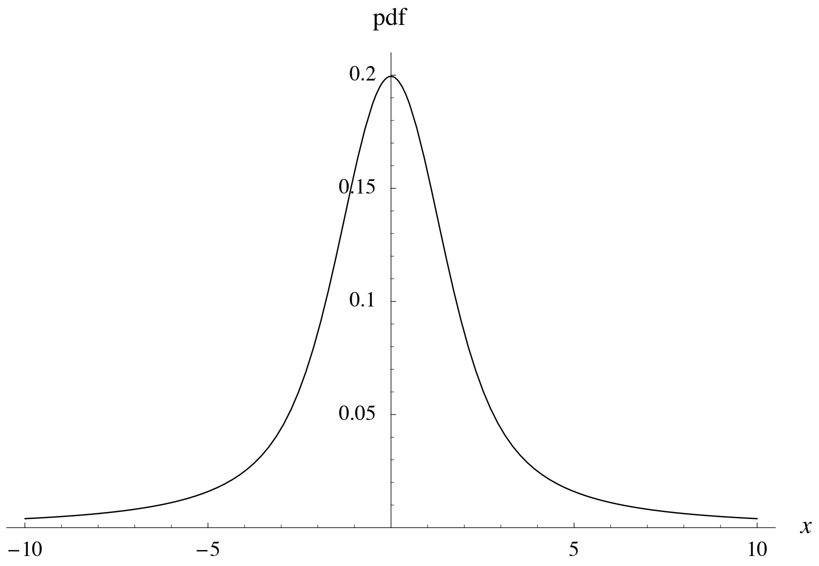


Fig. 12: The pdf, now obtained by numerically inverting the cf

2.4 E Stable Distributions

According to the Central Limit Theorem, the sum of a large number of iid random variables with finite variance converges to a Normal distribution (which is itself a special member of the stable family) when suitably standardised. If the finite variance assumption is dropped, one obtains a Generalised Central Limit Theorem, which states that the resulting limiting distribution must be a member of the stable class. The word 'stable' is used because, informally speaking, when iid members of a stable family are added together, the shape of the distribution does not change. Stable distributions are becoming increasingly important in empirical work. For example, in finance, financial returns are the sum of an enormous number of separate trades that arrive continuously in time. Yet, the distribution of financial returns often has fatter tails and more skewness than is consistent with Normality; by contrast, non-Gaussian stable distributions can often provide a better description of the data. For more detail on stable distributions, see Uchaikin and Zolotarev (1999), Nolan (2001), and McCulloch (1996).

Formally, a *stable distribution* $S(\alpha, \beta, c, a)$ is a 4-parameter distribution with characteristic function $C(t)$ given by

$$C(t) = \begin{cases} \exp\Big(a\,i\,t - c\,|t|^{\alpha}\,\{1 + i\,\beta\,\text{sign}(t)\,\tan(\tfrac{\pi}{2}\,\alpha)\}\Big) & \text{if } \alpha \neq 1 \\ \exp\Big(a\,i\,t - c\,|t|\,\{1 + i\,\beta\,\text{sign}(t)\,\tfrac{2}{\pi}\log|t|\}\Big) & \text{if } \alpha = 1 \end{cases} \tag{2.21}$$

where $0 < \alpha \le 2$, $-1 \le \beta \le 1$, $c > 0$ and $a \in \mathbb{R}$. Parameter α is known as the 'characteristic exponent' and controls tail behaviour, β is a skewness parameter, c is a scale parameter, and a is a location parameter. Since the shape parameters α and β are of primary interest, we will let $S(\alpha, \beta)$ denote $S(\alpha, \beta, 1, 0)$. Then $C(t)$ reduces to

$$C(t) = \begin{cases} \exp\left(-\left| t \right|^{\alpha} \{ 1 + i\, \beta \operatorname{sign}(t) \tan(\frac{\pi}{2}\, \alpha) \} \right) & \text{if } \alpha \neq 1 \\ \exp\left(-\left| t \right| \{ 1 + i\, \beta \operatorname{sign}(t)\, \frac{2}{\pi} \log \left| t \right| \} \right) & \text{if } \alpha = 1 \end{cases} \tag{2.22}$$

with support,

$$\text{support } f(x) = \begin{cases} \mathbb{R}_+ & \text{if } \alpha < 1 \text{ and } \beta = -1 \\ \mathbb{R}_- & \text{if } \alpha < 1 \text{ and } \beta = 1 \\ \mathbb{R} & \text{otherwise} \end{cases} \tag{2.23}$$

If $\alpha \le 1$, the mean does not exist; if $1 < \alpha < 2$, the mean exists, but the variance does not; if $\alpha = 2$ (the Normal distribution), both the mean and the variance exist. A symmetry property is that $f(x;\ \alpha, \beta) = f(-x;\ \alpha, -\beta)$. Thus, if the skewness parameter $\beta = 0$, we have $f(x;\ \alpha, 0) = f(-x;\ \alpha, 0)$, so that the pdf is symmetrical about zero. In *Mathematica*, we shall stress the dependence of the cf on its parameters α and β by defining the cf (2.22) as a *Mathematica* function of α and β, namely `cf[α,β]`:

```
Clear[cf]

cf[α_, β_] := Exp[-Abs[t]^α (1 + ⅈ β Sign[t] *
                If[α == 1, 2/π Log[Abs[t]], Tan[π/2 α]] )]
```

In the usual fashion, inverting the cf yields the pdf. Surprisingly, there are only three known stable pdf's that can be expressed in terms of *elementary* functions, and they are:

(i) *The Normal Distribution:* Let $\alpha = 2$; then the cf is:

```
cf[2, β]
```

$e^{-\text{Abs}[t]^2}$

which simplifies to e^{-t^2} for $t \in \mathbb{R}$. Inverting the cf yields a Normal pdf (the `InvertCF` function was defined in §2.4 D above):

```
f = InvertCF[t → x, cf[2, β]]
```

$$\frac{e^{-\frac{x^2}{4}}}{2\sqrt{\pi}}$$

(ii) *The Cauchy Distribution:* Let $\alpha = 1$ and $\beta = 0$; then the cf and pdf are:

```
cf[1, 0]
```

$e^{-\text{Abs}[t]}$

```
f = InvertCF[t → x, cf[1, 0]]
```

$$\frac{1}{\pi + \pi x^2}$$

(iii) *The Levy Distribution:* Let $\alpha = \frac{1}{2}$, $\beta = -1$; then the cf is:

```
cf[1/2, -1]
```

$e^{-\sqrt{\text{Abs}[t]}\,(1 - i\,\text{Sign}[t])}$

which, when inverted, yields the Levy pdf:

```
f = InvertCF[t → x, cf[1/2, -1], x > 0]
domain[f] = {x, 0, ∞};
```

$$\frac{e^{-\frac{1}{2x}}}{\sqrt{2\pi}\, x^{3/2}}$$

Here is a plot of the Levy pdf:

```
PlotDensity[f, {x, 0, 6}];
```

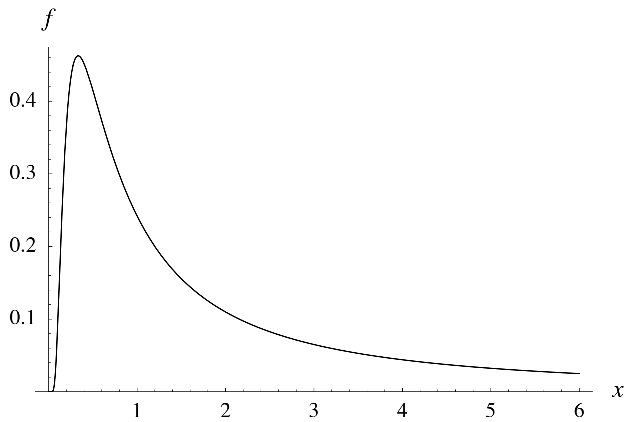


Fig. 13: The Levy pdf

The Levy distribution may also be obtained as a special case of the InverseGamma(γ, b) distribution with $\gamma = \frac{1}{2}$ and $b = 2$.

○ ***Only Three Known pdf's?***

It is often claimed that, aside from the Normal, Cauchy and Levy, no other stable pdf can be expressed in terms of known functions. This is not quite true: it depends on which functions are *known*. Hoffman-Jørgenson (1993) showed that some stable densities can be expressed in terms of hypergeometric ${}_pF_q$ functions, while Zolotarev (1995) showed more generally that some stable pdf's can be expressed in terms of MeijerG functions. Quite remarkably, *Mathematica* can often derive symbolic stable pdf's in terms of ${}_pF_q$ functions, without any extra help! To illustrate, suppose we wish to find the pdf of $S(\frac{1}{2}, 0)$. Inverting the cf in the standard way yields:

```
ff = InvertCF[t → x, cf[1/2, 0], x ∈ Reals]
```

$$\frac{1}{4\,\pi\,\text{Abs}[x]^{7/2}}\left(-2\,\text{Abs}[x]^{3/2}\,\text{HypergeometricPFQ}\left[\{1\},\left\{\frac{3}{4},\frac{5}{4}\right\},-\frac{1}{64\,x^2}\right]+\sqrt{2\,\pi}\,x^2\left(\text{Cos}\left[\frac{1}{4\,x}\right]+\text{Sign}[x]\,\text{Sin}\left[\frac{1}{4\,x}\right]\right)\right)$$

Since *Mathematica* does not handle densities containing `Abs[x]` very well, we shall eliminate the absolute value term by considering the $x < 0$ and $x > 0$ cases separately:

```
f₋ = Simplify[ ff /. Abs[x] → -x,  x < 0];
f₊ = Simplify[ ff, x > 0];
```

and then re-express the $S(\frac{1}{2}, 0)$ stable density as:

```
f = If[x < 0, f₋, f₊];   domain[f] = {x, -∞, ∞};
```

Note that we are now working with a stable pdf in symbolic form that is *neither* Normal, Cauchy, nor Levy. Further, because it is a symbolic entity, we can apply standard **mathStatica** functions in the usual way. For instance, `Expect[x, f]` correctly finds that the integral does not converge, while the cdf $F(x) = P(X \leq x)$ is obtained in the familiar way, as a symbolic entity!

```
F = Prob[x, f]
```

$$\text{If}\Big[x<0,\ i\left(\text{FresnelC}\left[\frac{1}{\sqrt{2\,\pi}\,\sqrt{x}}\right]-\text{FresnelS}\left[\frac{1}{\sqrt{2\,\pi}\,\sqrt{x}}\right]\right)+\frac{\text{HypergeometricPFQ}\left[\left\{\frac{1}{2},1\right\},\left\{\frac{3}{4},\frac{5}{4},\frac{3}{2}\right\},-\frac{1}{64\,x^2}\right]}{2\,\pi\,x},$$
$$1-\text{FresnelC}\left[\frac{1}{\sqrt{2\,\pi}\,\sqrt{x}}\right]-\text{FresnelS}\left[\frac{1}{\sqrt{2\,\pi}\,\sqrt{x}}\right]+\frac{\text{HypergeometricPFQ}\left[\left\{\frac{1}{2},1\right\},\left\{\frac{3}{4},\frac{5}{4},\frac{3}{2}\right\},-\frac{1}{64\,x^2}\right]}{2\,\pi\,x}\Big]$$

Figure 14 plots the pdf and cdf.

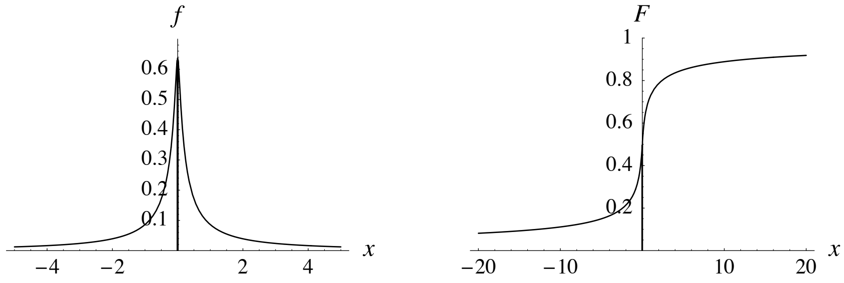


Fig. 14: The $S(\frac{1}{2}, 0)$ pdf and cdf

More generally, examples fall into two classes: those that can be inverted symbolically, and those that can only be inverted numerically. To illustrate, we shall consider $S(\frac{1}{2}, \beta)$ using symbolic methods; then $S(1, \beta)$ via numerical methods, and finally $S(\frac{3}{2}, \beta)$ with both numerical and symbolic methods, all plotted when $\beta = 0, \frac{1}{2}, 1$. Figures 15–17 illustrate these cases: as usual, the code to generate these diagrams is given in the electronic version of the text, along with some discussion.

2.4 F Cumulants and Probability Generating Functions

The *cumulant generating function* is the natural logarithm of the mgf. The r^{th} *cumulant*, κ_r, is given by

$$\kappa_r = \left. \frac{d^r \log(M(t))}{d t^r} \right|_{t=0} \tag{2.24}$$

provided $M(t)$ exists. Unlike the raw and central moments, cumulants can not generally be obtained by direct integration. To find them, one must either derive them from the cumulant generating function, or use the moment conversion functions of §2.4 G.

The *probability generating function* (pgf) is

$$\Pi(t) = E\left[t^X\right] \tag{2.25}$$

and is mostly used when working with discrete random variables defined on the set of non-negative integers {0, 1, 2, …}. The pgf provides a way to determine the probabilities. For instance:

$$P(X = r) = \frac{1}{r!} \left. \frac{d^r \Pi(t)}{d t^r} \right|_{t=0}, \qquad r = 0, 1, 2, \dots . \tag{2.26}$$

The pgf can also be used as a *factorial moment generating function*. For instance, the *factorial moment*

$$\acute{\mu}[r] = E\left[X^{[r]}\right] = E\left[X(X-1)\cdots(X-r+1)\right]$$

may be obtained from $\Pi(t)$ as follows:

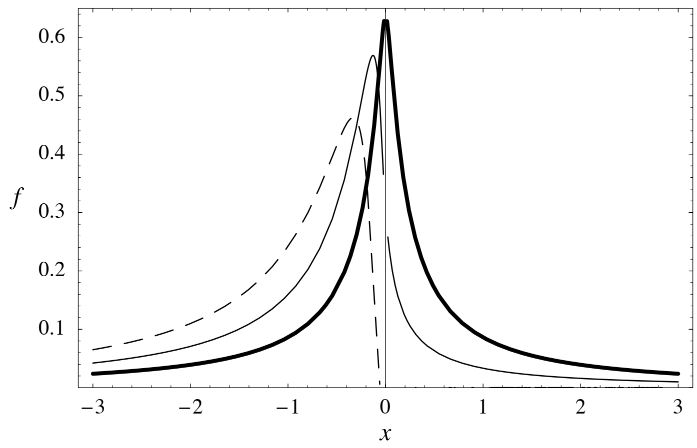


Fig. 15: $S(\frac{1}{2}, \beta)$ with $\beta = 0, \frac{1}{2}, 1$ (bold, plain, dashed)

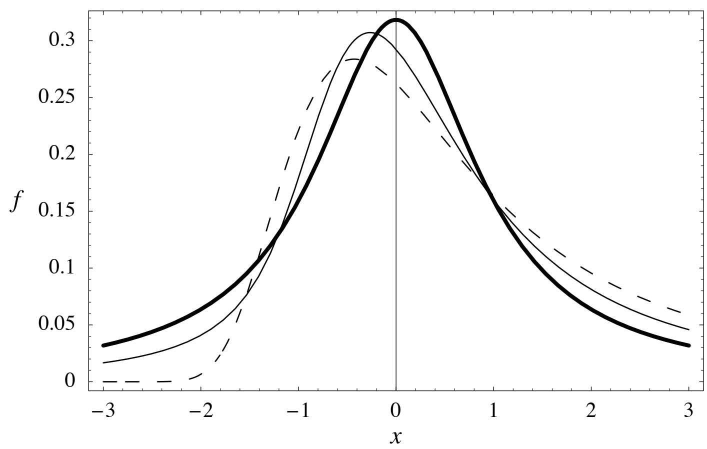


Fig. 16: $S(1, \beta)$ with $\beta = 0, \frac{1}{2}, 1$ (bold, plain, dashed)

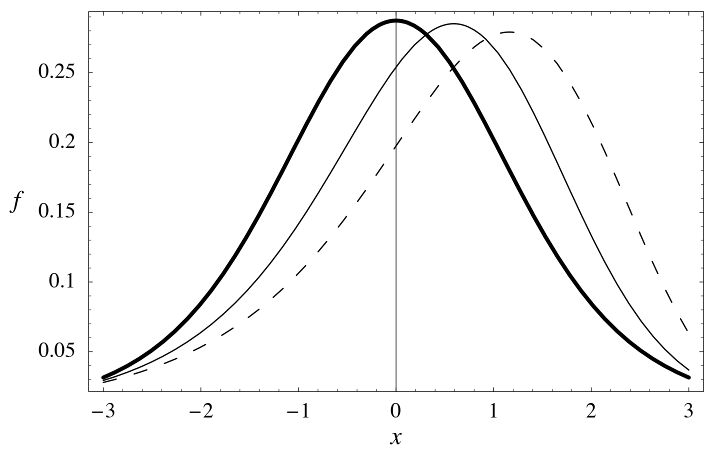


Fig. 17: $S(\frac{3}{2}, \beta)$ with $\beta = 0, \frac{1}{2}, 1$ (bold, plain, dashed)

$$\acute{\mu}[r] \;=\; E\big[X^{[r]}\big] \;=\; \left.\frac{d^r\,\Pi(t)}{d\,t^r}\right|_{t=1} \qquad (2.27)$$

where we note that t is set to 1 and not 0. To convert from factorial moments to raw moments, see the `FactorialToRaw` function of §2.4 G.

2.4 G Moment Conversion Formulae

One can express any moment ($\acute{\mu}$, μ, or κ) in terms of any other moment ($\acute{\mu}$, μ, or κ). To this end, **mathStatica** provides a suite of functions to automate such conversions. The supported conversions are:

function	*description*
`RawToCentral[r]`	$\acute{\mu}_r$ in terms of μ_i
`RawToCumulant[r]`	$\acute{\mu}_r$ in terms of κ_i
`CentralToRaw[r]`	μ_r in terms of $\acute{\mu}_i$
`CentralToCumulant[r]`	μ_r in terms of κ_i
`CumulantToRaw[r]`	κ_r in terms of $\acute{\mu}_i$
`CumulantToCentral[r]`	κ_r in terms of μ_i
and	
`RawToFactorial[r]`	$\acute{\mu}_r$ in terms of $\acute{\mu}[i]$
`FactorialToRaw[r]`	$\acute{\mu}[r]$ in terms of $\acute{\mu}_i$

Table 3: Univariate moment conversion functions

For instance, to express the 2nd central moment (the variance) $\mu_2 = E\big[(X-\mu)^2\big]$ in terms of raw moments $\acute{\mu}_i$, we enter:

```
CentralToRaw[2]
```

$\mu_2 \to -\acute{\mu}_1^2 + \acute{\mu}_2$

This is just the well-known result that $\mu_2 = E[X^2] - (E[X])^2$. Here are the first 6 central moments in terms of raw moments:

```
Table[CentralToRaw[i], {i, 6}] // TableForm
```

$\mu_1 \to 0$

$\mu_2 \to -\acute{\mu}_1^2 + \acute{\mu}_2$

$\mu_3 \to 2\,\acute{\mu}_1^3 - 3\,\acute{\mu}_1\,\acute{\mu}_2 + \acute{\mu}_3$

$\mu_4 \to -3\,\acute{\mu}_1^4 + 6\,\acute{\mu}_1^2\,\acute{\mu}_2 - 4\,\acute{\mu}_1\,\acute{\mu}_3 + \acute{\mu}_4$

$\mu_5 \to 4\,\acute{\mu}_1^5 - 10\,\acute{\mu}_1^3\,\acute{\mu}_2 + 10\,\acute{\mu}_1^2\,\acute{\mu}_3 - 5\,\acute{\mu}_1\,\acute{\mu}_4 + \acute{\mu}_5$

$\mu_6 \to -5\,\acute{\mu}_1^6 + 15\,\acute{\mu}_1^4\,\acute{\mu}_2 - 20\,\acute{\mu}_1^3\,\acute{\mu}_3 + 15\,\acute{\mu}_1^2\,\acute{\mu}_4 - 6\,\acute{\mu}_1\,\acute{\mu}_5 + \acute{\mu}_6$

Next, we express the 5th raw moment in terms of cumulants:

```
sol = RawToCumulant[5]
```

$\acute{\mu}_5 \to \kappa_1^5 + 10\,\kappa_1^3\,\kappa_2 + 15\,\kappa_1\,\kappa_2^2 + 10\,\kappa_1^2\,\kappa_3 + 10\,\kappa_2\,\kappa_3 + 5\,\kappa_1\,\kappa_4 + \kappa_5$

which is an expression in κ_i, for $i = 1, \ldots, 5$. Here are the inverse relations:

```
inv = Table[CumulantToRaw[i], {i, 5}]; inv // TableForm
```

$\kappa_1 \to \acute{\mu}_1$

$\kappa_2 \to -\acute{\mu}_1^2 + \acute{\mu}_2$

$\kappa_3 \to 2\,\acute{\mu}_1^3 - 3\,\acute{\mu}_1\,\acute{\mu}_2 + \acute{\mu}_3$

$\kappa_4 \to -6\,\acute{\mu}_1^4 + 12\,\acute{\mu}_1^2\,\acute{\mu}_2 - 3\,\acute{\mu}_2^2 - 4\,\acute{\mu}_1\,\acute{\mu}_3 + \acute{\mu}_4$

$\kappa_5 \to 24\,\acute{\mu}_1^5 - 60\,\acute{\mu}_1^3\,\acute{\mu}_2 + 30\,\acute{\mu}_1\,\acute{\mu}_2^2 + 20\,\acute{\mu}_1^2\,\acute{\mu}_3 - 10\,\acute{\mu}_2\,\acute{\mu}_3 - 5\,\acute{\mu}_1\,\acute{\mu}_4 + \acute{\mu}_5$

Substituting the inverse relations back into `sol` yields $\acute{\mu}_5$ again:

```
sol /. inv // Simplify
```

$\acute{\mu}_5 \to \acute{\mu}_5$

Working 'about the mean' (*i.e.* taking $\kappa_1 = 0$) yields the `CentralToCumulant` conversions:

```
Table[ CentralToCumulant[r], {r, 5}]
```

$\{\mu_1 \to 0,\ \mu_2 \to \kappa_2,\ \mu_3 \to \kappa_3,\ \mu_4 \to 3\,\kappa_2^2 + \kappa_4,\ \mu_5 \to 10\,\kappa_2\,\kappa_3 + \kappa_5\}$

The inverse relations are given by `CumulantToCentral`. Here is the 5th factorial moment $\acute{\mu}[5] = E\left[X(X-1)(X-2)(X-3)(X-4)\right]$ expressed in terms of raw moments:

```
FactorialToRaw[5]
```

$\acute{\mu}[5] \to 24\,\acute{\mu}_1 - 50\,\acute{\mu}_2 + 35\,\acute{\mu}_3 - 10\,\acute{\mu}_4 + \acute{\mu}_5$

This is easy to confirm by noting that:

```
x (x - 1) (x - 2) (x - 3) (x - 4) // Expand
```

$24\,x - 50\,x^2 + 35\,x^3 - 10\,x^4 + x^5$

The inverse relations are given by `RawToFactorial`:

```
RawToFactorial[5]
```

$\acute{\mu}_5 \to \acute{\mu}[1] + 15\,\acute{\mu}[2] + 25\,\acute{\mu}[3] + 10\,\acute{\mu}[4] + \acute{\mu}[5]$

○ ***The Converter Functions in Practice***

Sometimes, we know how to derive one class of moments (say raw moments), but not another (say cumulants). In these situations, the converter functions come to the rescue, for they enable us to derive the unknown moments in terms of the moments that can be calculated. This section illustrates how this can be done. The general approach is: First, express the desired moment (say κ_5) in terms of moments that we can calculate (say raw moments). Then, evaluate each raw moment $\acute{\mu}_i$ for the relevant distribution.

⊕ ***Example 20:*** Cumulants of $X \sim$ Beta(a, b)

Let random variable $X \sim$ Beta(a, b) with pdf $f(x)$:

$$f = \frac{x^{a-1}\,(1-x)^{b-1}}{\text{Beta}[a,\ b]}; \quad \text{domain}[f] = \{x,\ 0,\ 1\}\ \&\&\ \{a > 0,\ b > 0\};$$

We wish to find the fourth cumulant. To do so, we can use the cumulant generating function approach, or the moment conversion approach.

(i) The cumulant generating function is:

```
cgf = Log[Expect[ⅇ^(t x), f]]
```

```
Log[Hypergeometric1F1[a, a + b, t]]
```

Then, the fourth cumulant is given by (2.24) as:

```
D[cgf, {t, 4}] /. t → 0 // FullSimplify
```

$$\frac{6\,a\,b\,(a^3 + a^2\,(1-2\,b) + b^2\,(1+b) - 2\,a\,b\,(2+b))}{(a+b)^4\,(1+a+b)^2\,(2+a+b)\,(3+a+b)}$$

(ii) Moment conversion approach: Express the fourth cumulant in terms of raw moments:

```
sol = CumulantToRaw[4]
```

$$\kappa_4 \to -6\,\acute{\mu}_1^4 + 12\,\acute{\mu}_1^2\,\acute{\mu}_2 - 3\,\acute{\mu}_2^2 - 4\,\acute{\mu}_1\,\acute{\mu}_3 + \acute{\mu}_4$$

Here, each term $\acute{\mu}_r$ denotes $\acute{\mu}_r(X) = E[X^r]$, and hence can be evaluated with the `Expect` function. In the next input, we calculate each of the expectations that we require:

```
sol /. μ́_r_ :→ Expect[x^r, f] // FullSimplify
```

$$\kappa_4 \to \frac{6\,a\,b\,(a^3 + a^2\,(1-2\,b) + b^2\,(1+b) - 2\,a\,b\,(2+b))}{(a+b)^4\,(1+a+b)^2\,(2+a+b)\,(3+a+b)}$$

which is the same answer. ■

2.5 Conditioning, Truncation and Censoring

2.5 A Conditional/Truncated Distributions

Let random variable X have pdf $f(x)$, with cdf $F(x) = P(X \le x)$. Further, let a and b be constants lying within the support of the domain. Then, the conditional density is

$$f(x \mid a < X \le b) = \frac{f(x)}{F(b) - F(a)} \qquad \text{Doubly truncated} \qquad (2.28)$$

$$f(x \mid X > a) = \frac{f(x)}{1 - F(a)} \quad (\text{let } b = \infty) \qquad \text{Truncated below} \qquad (2.29)$$

$$f(x \mid X \le b) = \frac{f(x)}{F(b)} \quad (\text{let } a = -\infty) \qquad \text{Truncated above} \qquad (2.30)$$

These conditional distributions are also sometimes known as *truncated distributions*. In each case, the conditional density on the left-hand side is expressed in terms of the unconditional (parent) pdf $f(x)$ on the right-hand side, which is adjusted by a scaling constant in the denominator so that the density still integrates to unity.

Proof of (2.30): The conditional probability that event Ω_1 occurs, given event Ω_2, is

$$P(\Omega_1 \mid \Omega_2) = \frac{P(\Omega_1 \cap \Omega_2)}{P(\Omega_2)} \quad \text{provided } P(\Omega_2) \ne 0.$$

$$\therefore\ P(X \le x \mid X \le b) = \frac{P(X \le x \cap X \le b)}{P(X \le b)} = \frac{P(X \le x)}{P(X \le b)} \quad \text{provided } x \le b.$$

$\therefore\ F(x \mid X \le b) = \dfrac{F(x)}{F(b)}$. Differentiating both sides with respect to x yields (2.30). □

⊕ ***Example 21:*** A 'Truncated Above' Standard Normal Distribution

```
ClearAll[f, F, g, b]
```

Let $X \sim N(0, 1)$ with pdf $f(x)$:

$$\texttt{f} = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}; \qquad \texttt{domain[f] = \{x, -}\infty\texttt{, }\infty\texttt{\};}$$

and cdf $F(x)$:

```
F[x_] = Prob[x, f];
```

Let $g(x) = f(x \mid X \leq b) = \dfrac{f(x)}{F(b)}$ denote a standard Normal pdf truncated above at b:

```
g = f/F[b];    domain[g] = {x, -∞, b} && {b ∈ Reals};
```

Figure 18 plots $g(x)$ at three different values of b.

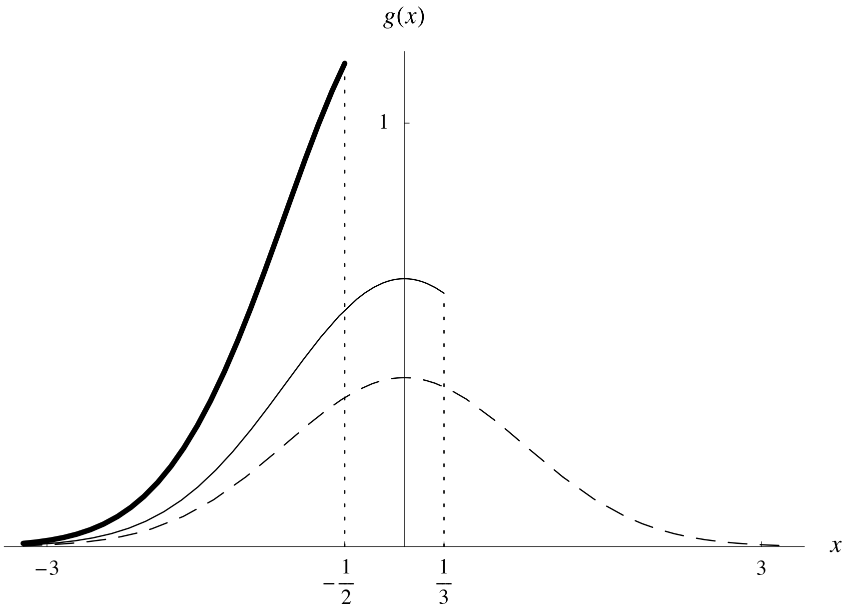


Fig. 18: A standard Normal pdf truncated above at $b = -\frac{1}{2}, \frac{1}{3}, \infty$

2.5 B Conditional Expectations

Let X have pdf $f(x)$. We wish to find the *conditional expectation* $E_f[u(X) \mid a < X \leq b]$, where the notation $E_f[\cdot]$ indicates that the expectation is taken with respect to the random variable X whose pdf is $f(x)$. From (2.28), it follows that

$$E_f[u(X) \mid a < X \leq b] = \frac{\int_a^b u(x)\, f(x)\, dx}{F(b) - F(a)}. \tag{2.31}$$

With **mathStatica**, an easier method is to first derive the conditional density via (2.28), say $g(x) = f(x \mid a < X \leq b)$ with `domain[g] = {x, a, b}`. Then,

$$E_f[u(X) \mid a < X \leq b] = E_g[u(X)]. \tag{2.32}$$

⊕ ***Example 22:*** Mean and Variance of a 'Truncated Above' Normal

Continuing *Example 21*, we have $X \sim N(0, 1)$ with pdf $f(x)$ (the parent distribution), and $g(x) = f(x \mid X \le b)$ (a *truncated above* distribution). We wish to find $E_f[X \mid X \le b]$. The solution is $E_g[X]$:

```
Expect[x, g]
```

$$-\frac{e^{-\frac{b^2}{2}}\sqrt{\frac{2}{\pi}}}{1 + \text{Erf}\left[\frac{b}{\sqrt{2}}\right]}$$

Because $g(x)$ is 'truncated above' while $f(x)$ is not, it must always be the case that $E_g[X] < E_f[X]$. As b becomes 'large', the truncation becomes less severe, so $E_g[X] \to E_f[X]$. Thus, for our example, as $b \to \infty$, $E_g[X] \to 0$ from below, as per Fig. 19 (left panel). At the other extreme, as $b \to -\infty$, the 45° line forms an upper bound, since $E_g[X] \le b$, if $X \le b$.

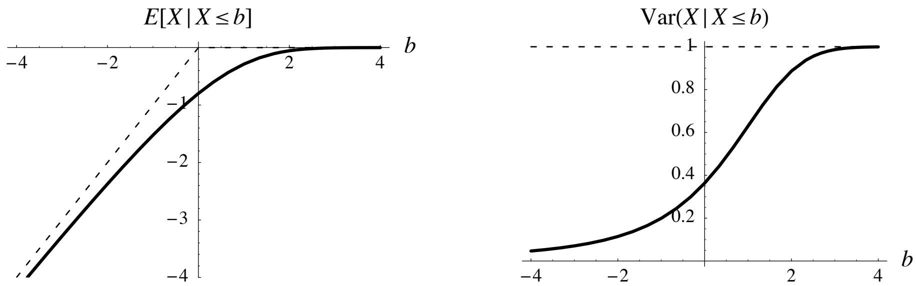


Fig. 19: Conditional mean (left) and variance (right) as a function of b

Similarly, the variance of a truncated distribution must always be smaller than the variance of its parent distribution, because the truncated distribution is a constrained version of the parent. As b becomes 'large', this constraint becomes insignificant, and so $\text{Var}_g(X) \to \text{Var}_f(X)$ from below. By contrast, as b tends toward the lower bound of the domain, truncation becomes more and more binding, causing the conditional variance to tend to 0, as per Fig. 19 (right panel). The conditional variance $\text{Var}(X \mid X \le b)$ is:

```
Var[x, g]
```

$$1 - \frac{2\, e^{-b^2}}{\pi\left(1 + \text{Erf}\left[\frac{b}{\sqrt{2}}\right]\right)^2} - \frac{b\, e^{-\frac{b^2}{2}}\sqrt{\frac{2}{\pi}}}{1 + \text{Erf}\left[\frac{b}{\sqrt{2}}\right]}$$

Finally, we `Clear` some symbols:

```
ClearAll[f, F, g]
```

… to prevent notational conflicts in future examples. ■

2.5 C Censored Distributions

Consider the following examples:

(i) The demand for tickets to a concert is a random variable. Actual ticket sales, however, are bounded by the fixed capacity of the concert hall.

(ii) Similarly, electricity consumption (a random variable) is constrained above by the capacity of the grid.

(iii) The water level in a dam fluctuates randomly, but it can not exceed the physical capacity of the dam.

(iv) In some countries, foreign exchange rates are allowed to fluctuate freely within a band, but if they reach the edge of the band, the monetary authority intervenes to prevent the exchange rate from leaving the band.

Examples (i) and (ii) draw the distinction between *observed* data (*e.g.* ticket sales, electricity supply) and *unobserved* demand (some people may have been unable to purchase tickets). Examples (iii) and (iv) fall into the general class of stochastic processes that are bounded by reflecting (sticky) barriers; see Rose (1995). All of these examples (i–iv) can be modelled using censored distributions.

Let random variable X have pdf $f(x)$ and cdf $F(x)$, and let c denote a constant lying within the support of the domain. Then, Y has a *censored distribution*, censored below at point c, if

$$Y = \begin{cases} c & \text{if } X \le c \\ X & \text{if } X > c \end{cases} \tag{2.33}$$

Figure 20 compares the pdf of X (the parent distribution) with the pdf of Y (the censored distribution). While X has a continuous pdf, the density of Y has both a discrete part and a continuous part. Here, all values of X smaller than c get compacted onto a single point c: thus, the point c occurs with positive probability $F(c)$.

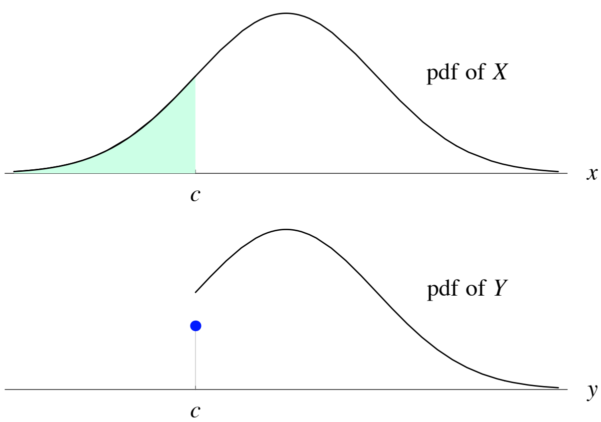


Fig. 20: Parent pdf (top) and censored pdf (bottom)

The definitions for a 'censored above' distribution, and a 'doubly censored' distribution (censored above and below) follow in similar fashion.

⊕ ***Example 23:*** A 'Censored Below' Normal Distribution

```
ClearAll[f, c]
```

Let $X \sim N(0, 1)$ with pdf $f(x)$, and let $Y = \begin{cases} c & \text{if } X \le c \\ X & \text{if } X > c \end{cases}$. We enter all this as:

```
f = E^(-x^2/2)/Sqrt[2 π];   domain[f] = {x, -∞, ∞};   y = If[x ≤ c, c, x];
```

Then, $E[Y]$ is:

```
Expect[y, f]
```

$$\frac{e^{-\frac{c^2}{2}}}{\sqrt{2\pi}} + \frac{1}{2}\, c \left(1 + \text{Erf}\left[\frac{c}{\sqrt{2}}\right]\right)$$

Note that this expression is equal to $f(c) + c\,F(c)$, where $F(c)$ is the cdf of X evaluated at the censoring point c. Similarly, $\text{Var}(Y)$ is:

```
Var[y, f]
```

$$\frac{1}{4}\left(2 + c^2 - \frac{2\, e^{-c^2}}{\pi} + \left(-2 - 2\, c\, e^{-\frac{c^2}{2}} \sqrt{\frac{2}{\pi}}\right) \text{Erf}\left[\frac{c}{\sqrt{2}}\right] - c^2\, \text{Erf}\left[\frac{c}{\sqrt{2}}\right]^2\right)$$

Figure 21 plots $E[Y]$ and $\text{Var}(Y)$ as a function of the censoring point c.

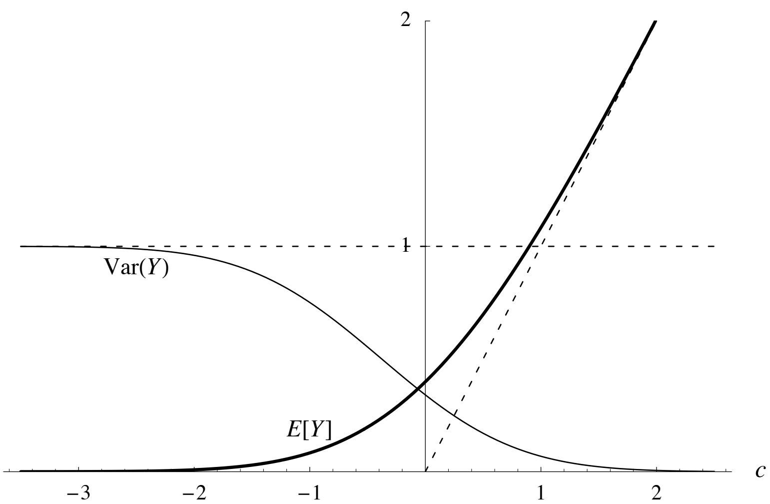


Fig. 21: The mean and variance of Y, plotted at different values of c

2.5 D Option Pricing

Financial options are an interesting application of censored distributions. To illustrate, let time $t = 0$ denote today, let $\{S(t),\ t \geq 0\}$ denote the price of a stock at time t, and let S_T denote the stock price at a fixed future 'expiry' date $T > 0$. A European *call option* is a financial asset that gives its owner the right (but not the obligation) to buy stock at time T at a fixed price k (called the *strike price*). For example, if you own an Apple call option expiring on 19 July with strike $k = \$100$, it means you have the right to buy one share in Apple Computer at a price of \$100 on July 19. If, on July 19, the stock price S_T is greater than $k = \$100$, the value of your option on the expiry date is $S_T - k$; however, if S_T is less than \$100, it would not be worthwhile to purchase at \$100, and so your option would have zero value. Thus, the value of a call option *at expiry* T is:

$$V_T = \begin{cases} S_T - k & \text{if } S_T > k \\ 0 & \text{if } S_T \leq k \end{cases} \tag{2.34}$$

We now know the value of an option at expiry — what then is the value of this option *today*, at $t = 0$, prior to expiry? At $t = 0$, the current stock price $S(0)$ is always known, while the future is of course unknown. That is, the future price S_T is a random variable whose pdf $f(s_T)$ is assumed known. Then, the value $V = V(0)$ of the option at $t = 0$ is simply the expected value of V_T, discounted for the time value of money between expiry ($t = T$) and today ($t = 0$):

$$V = V(0) = e^{-rT} E[V_T] \tag{2.35}$$

where r denotes the risk-free interest rate. This is the essence of option pricing, and we see that it rests crucially on censoring the distribution of future stock prices, $f(s_T)$.

⊕ ***Example 24:*** Black–Scholes Option Pricing (via Censored Distributions)

The Black–Scholes (1973) option pricing model is now quite famous, as acknowledged by the 1997 Nobel Memorial Prize in economics.[3] For our purposes, we just require the pdf of future stock prices $f(s_T)$. This, in turn, requires some stochastic calculus; readers unfamiliar with stochastic calculus can jump directly to (2.38) where $f(s_T)$ is stated, and proceed from there.

If investors are risk neutral,[4] and stock prices follow a geometric Brownian motion, then

$$\frac{dS}{S} = r\,dt + \sigma\,dz \tag{2.36}$$

with drift r and instantaneous standard deviation σ, where z is a Wiener process. By Ito's Lemma, this can be expressed as the ordinary Brownian motion

$$d\log(S) = \left(r - \frac{\sigma^2}{2}\right)dt + \sigma\,dz \tag{2.37}$$

so that $d\log(S_T) \sim N\left(\left(r - \frac{\sigma^2}{2}\right) T,\ \sigma^2\, T\right)$. Expressing $d\log(S_T)$ as $\log(S_T) - \log(S(0))$, it then follows that

$$\log(S_T) \sim N(a, b^2) \quad \text{where} \quad \begin{cases} a = \log(S(0)) + \left(r - \frac{\sigma^2}{2}\right) T \\ b = \sigma\sqrt{T} \end{cases} \tag{2.38}$$

That is, $S_T \sim \text{Lognormal}(a, b^2)$, with pdf $f(s_T)$:

$$\mathbf{f} = \frac{1}{s_T\, b\sqrt{2\pi}} \mathbf{Exp}\left[-\frac{(\mathbf{Log}[s_T] - a)^2}{2\, b^2}\right];$$

```
domain[f] = {sT, 0, ∞} && {a ∈ Reals, b > 0};
```

The value of the option at expiry, V_T, may be entered via (2.34) as:

```
VT = If[sT > k, sT - k, 0];
```

while the value $V = V(0)$ of a call option today is given by (2.35):

```
V = ⅇ^(-r T) Expect[VT , f]
```

$$\frac{1}{2} e^{-r T} \left(-k \left(1 + \mathrm{Erf}\left[\frac{a - \mathrm{Log}[k]}{\sqrt{2}\, b}\right]\right) + e^{a+\frac{b^2}{2}} \left(1 + \mathrm{Erf}\left[\frac{a + b^2 - \mathrm{Log}[k]}{\sqrt{2}\, b}\right]\right)\right)$$

where a and b were defined in (2.38). This result is, in fact, identical to the Black–Scholes solution, though our derivation here via expectations is quite different (and much simpler) than the solution via partial differential equations used by Black and Scholes. Substituting in for a and b, and denoting today's stock price $S(0)$ by p, we have:

$$\mathbf{Value} = \mathbf{V} \,/.\, \left\{a \to \mathbf{Log}[p] + \left(r - \frac{\sigma^2}{2}\right) T,\quad b \to \sigma\sqrt{T}\right\};$$

For example, if the current price of Apple stock is $p = S(0) = \$104$, the strike price is $k = \$100$, the interest rate is 5%, the volatility is 44% per annum ($\sigma = .44$), and there are 66 days left to expiry ($T = \frac{66}{365}$), then the value today (in \$) of the call option is:

$$\mathbf{Value} \,/.\, \left\{p \to 104,\ k \to 100,\ r \to .05,\ \sigma \to .44,\ T \to \frac{66}{365}\right\}$$

```
10.2686
```

More generally, we can plot the value of our call option as a function of the current stock price p, as shown in Fig. 22.

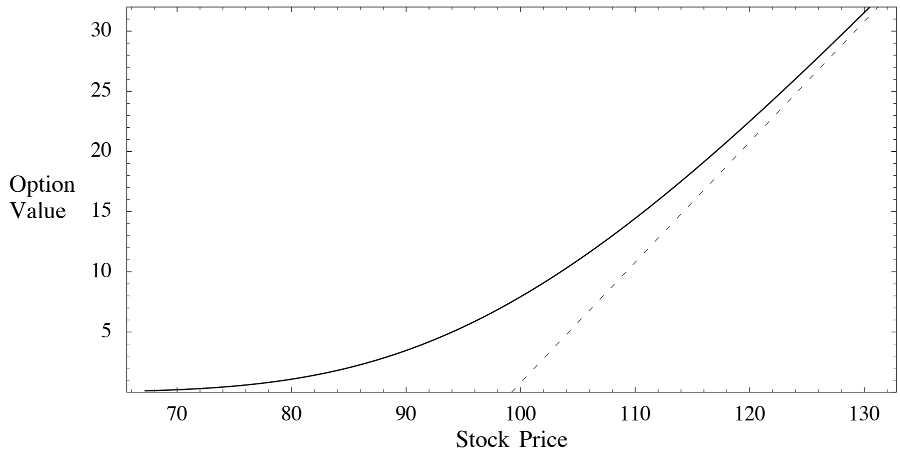


Fig. 22: Value of a call option as a function of today's stock price

As $p \to 0$, we become certain that $S_T < k$. Referring to (2.34), this means that as $p \to 0$, $V \to 0$, as Fig. 22 shows. By contrast, as $p \to \infty$, $P(S_T > k) \to 1$, so we become certain that $S_T > k$, and thus $V \to e^{-rT} E[S_T - k]$. The latter is equal to $p - e^{-rT} k$, as the reader can verify with `Expect[sT - k, f]` and then substituting in for `a` and `b`. This explains the asymptotes in Fig. 22.

Many interesting comparative static calculations are now easily obtainable with *Mathematica*; for example, we can find the rate of change of option value with respect to σ as a symbolic entity with `D[Value,σ]//Simplify`. ■

2.6 Pseudo-Random Number Generation

This section discusses different ways to generate pseudo-random drawings from a given distribution. If the distribution happens to be included in *Mathematica*'s Statistics package, the easiest approach is often to use the `Random[`*distribution*`]` function included in that package (§2.6 A). Of course, this is not a general solution, and it breaks down as soon as one encounters a distribution that is not in that package.

In the remaining parts of this section (§2.6 B – D), we discuss procedures that allow, in principle, any distribution to be sampled. We first consider the Inverse Method, which requires that both the cdf and inverse cdf can be computed, using either symbolic (§2.6 B) or numerical (§2.6 C) methods. Finally, §2.6 D discusses the Rejection Method, where neither the cdf nor the inverse cdf is required. Random number generation for discrete random variables is discussed in Chapter 3.

2.6 A *Mathematica*'s Statistics Package

The *Mathematica* statistics packages, `ContinuousDistributions`` and `NormalDistribution``, provide built-in pseudo-random number generation for well-known distributions such as the Normal, Gamma, and Cauchy. If we want to generate pseudo-random numbers from one of these well-known distributions, the simplest solution is to use these packages. They can be loaded as follows:

```
<< Statistics`
```

Suppose we want to generate pseudo-random drawings from a Gamma(a, b) distribution:

$$f = \frac{x^{a-1}\, e^{-x/b}}{\Gamma[a]\, b^{a}}; \qquad \text{domain}[f] = \{x, 0, \infty\} \,\&\&\, \{a > 0, b > 0\};$$

If $a = 2$ and $b = 3$, a single pseudo-random drawing is obtained as follows:

```
dist = GammaDistribution[2, 3];    Random[dist]
8.61505
```

while 10000 pseudo-random values can be generated with:

```
data = RandomArray[dist, 10000];
```

The **mathStatica** function, `FrequencyPlot`, can be used to compare this 'empirical' data with the true pdf $f(x)$:

```
FrequencyPlot[data, f /. {a → 2, b → 3}];
```

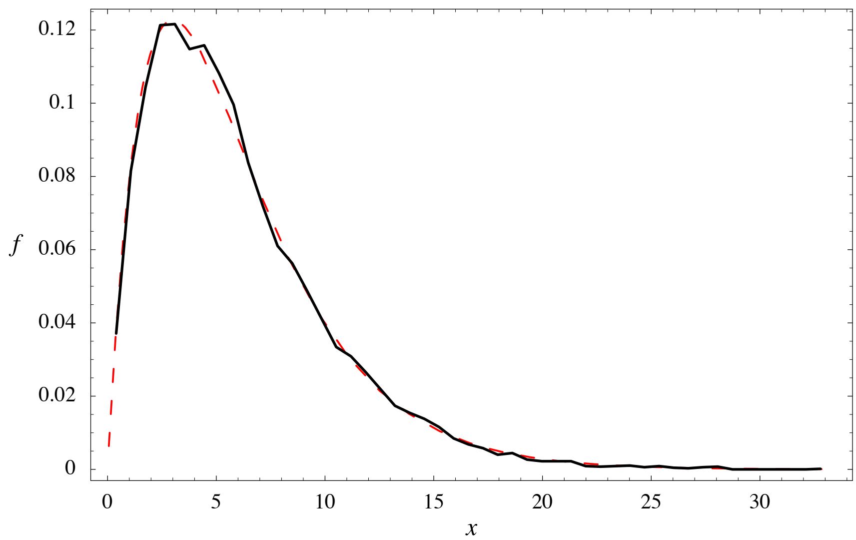


Fig. 23: The empirical pdf (——) and true pdf (– – –)

While it is certainly convenient to have pre-written code for special well-known distributions, this approach must, of course, break down as soon as we consider a distribution that is not in the package. Thus, more general methods are needed.

2.6 B Inverse Method (Symbolic)

Let random variable X have pdf $f(x)$, cdf $p = F(x)$ and inverse cdf $x = F^{-1}(p)$, and let u be a pseudo-random drawing from Uniform(0, 1). Then a pseudo-random drawing from $f(x)$ is given by

$$x = F^{-1}(u) \tag{2.39}$$

In order for the *Inverse Method* to work efficiently, the inverse function $F^{-1}(\cdot)$ should be computationally tractable. Here is an example with the Levy distribution, with pdf $f(x)$:

$$\mathbf{f} = \frac{e^{-\frac{1}{2x}}}{\sqrt{2\pi}\, x^{3/2}}; \qquad \mathbf{domain[f] = \{x, 0, \infty\};}$$

The cdf $F(x)$ is given by:

```
F = Prob[x, f]
```

$$1 - \text{Erf}\left[\frac{1}{\sqrt{2}\,\sqrt{x}}\right]$$

while the inverse cdf is:

```
inv = Solve[u == F, x] // Flatten

— Solve::ifun : Inverse functions are being
    used by Solve, so some solutions may not be found.
```

$$\left\{x \to \frac{1}{2\,\text{InverseErf}[0, 1-u]^2}\right\}$$

When `u = Random[]`, this rule generates a pseudo-random Levy drawing x. More generally, if the inverse yields more than one possible solution, we would have to select the appropriate solution before proceeding. We now generate 10000 pseudo-random numbers from the Levy pdf, by replacing `u` with `Random[]`:

$$\mathbf{data = Table}\left[\frac{1}{2\,\text{InverseErf}[0, 1-\text{Random}[\,]]^2}, \{10000\}\right]; \text{ // Timing}$$

```
{2.36 Second, Null}
```

It is always a good idea to check the data set before continuing. The output here should only consist of positive real numbers. To check, here are the last 10 values:

```
Take[data, -10]

{6.48433, 0.229415, 3.70733, 4.53735, 0.356657,
 0.646354, 1.09913, 0.443604, 1.17306, 0.532637}
```

These numbers seem fine. We use the **mathStatica** function `FrequencyPlot` to inspect fit, and superimpose the parent density $f(x)$ on top:

```
FrequencyPlot[data, {0, 10, .1}, f];
```

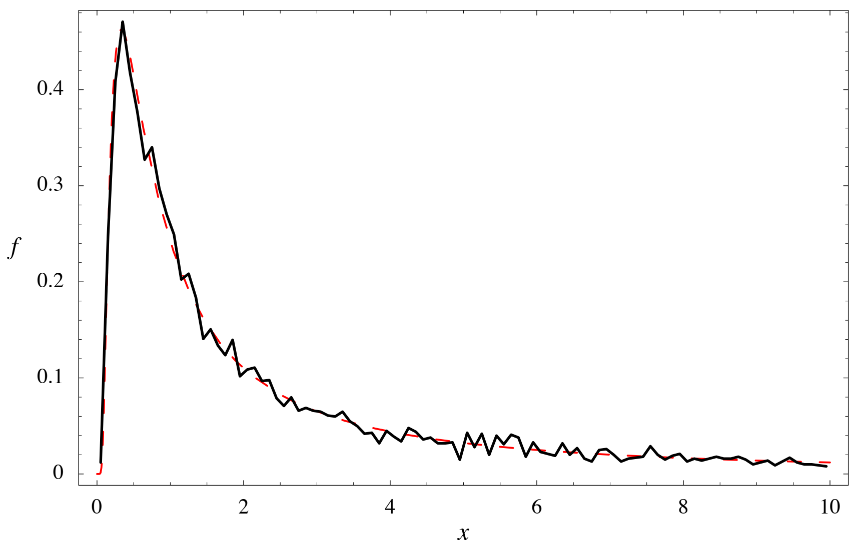


Fig. 24: The empirical pdf (——) and true pdf (– – –)

Some caveats: The Inverse Method can only work if we can determine both the cdf and its inverse. Inverse functions are tricky, and *Mathematica* may occasionally experience some difficulty in this regard. Also, since one ultimately has to work with a numerical density (*i.e.* numerical parameter values) when generating pseudo-random numbers, it is often best to specify parameter values at the very start — this makes it easier to calculate both the cdf and the inverse cdf.

2.6 C Inverse Method (Numerical)

If it is difficult or impossible to find the inverse cdf symbolically, we can resort to doing so numerically. To illustrate, let random variable X have a half-Halo distribution with pdf $f(x)$:

$$f = \frac{2}{\pi}\sqrt{1-(x-2)^2}; \quad \text{domain}[f] = \{x, 1, 3\};$$

with cdf $F(x)$:

```
F = Prob[x, f]
```

$$\frac{(-2+x)\sqrt{-(-3+x)(-1+x)} + \text{ArcCos}[2-x]}{\pi}$$

Mathematica cannot invert this cdf symbolically; that is, `Solve[u==F,x]` fails. Nevertheless, we can derive the inverse cdf using numerical methods. We do so by evaluating (F, x) at a finite number of different values of x, and then use interpolation to fill in the gaps in between these known points. How then do we decide at which values of x we should evaluate (F, x)? This is the same type of problem that *Mathematica*'s `Plot` function has to solve each time it makes a plot. So, following Abbott (1996), we use the `Plot` function to automatically select the values of x at which $(F(x), x)$ is to be constructed, and then record these values in a list called `lis`. The larger the number of `PlotPoints`, the more accurate will be the end result:

```
lis = {};
Plot[ (ss = F; AppendTo[lis, {ss, x}]; ss),  {x, 1, 3},
  PlotPoints → 2000,
  PlotRange → All, AxesLabel → {"x", "F"}];
```

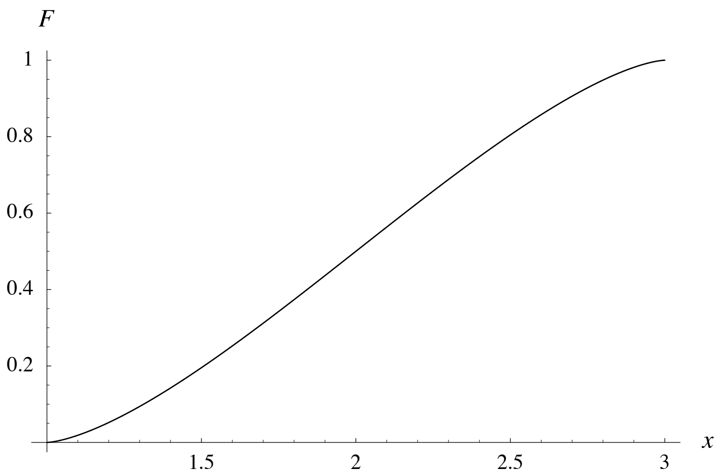


Fig. 25: The cdf $F(x)$ plotted as a function of x

Mathematica's `Interpolation` function is now used to fill in the gaps between the chosen points. We shall take the `Union` of `lis` so as to eliminate duplicate values that the `Plot` function can sometimes generate. Here, then, is our numerical inverse cdf function:

```
InverseCDF = Interpolation[Union[lis]]
```

```
InterpolatingFunction[{{1.89946×10^-14, 1.}}, <>]
```

Here are 60000 pseudo-random drawings from the half-Halo distribution:

```
data = Table[ InverseCDF[ Random[] ], {60000}];
 // Timing
```

```
{1.1 Second, Null}
```

Figure 26 compares this pseudo-random data with the true pdf $f(x)$:

```
FrequencyPlot[data, {1, 3, .02}, f];
```

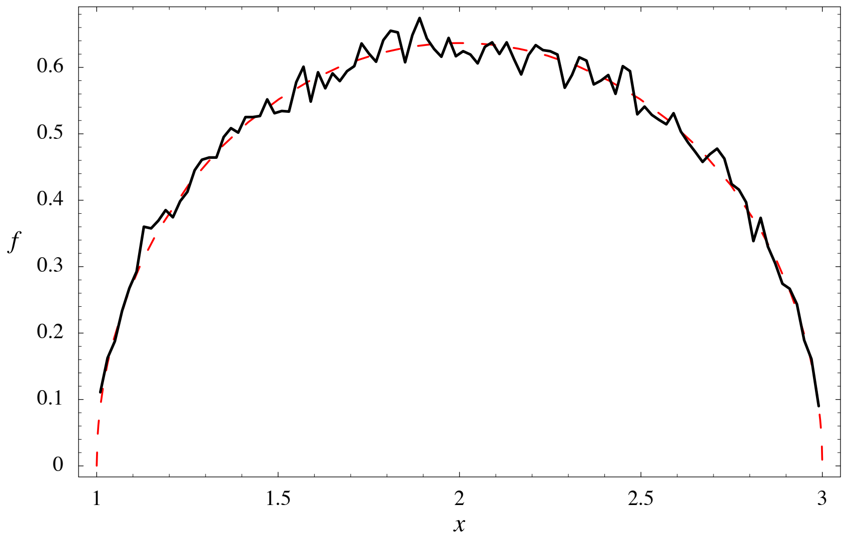


Fig. 26: The empirical pdf (——) and true half-Halo pdf (– – –)

2.6 D Rejection Method

Our objective is to generate pseudo-random numbers from some pdf $f(x)$. Sometimes, the Inverse Method may fail: typically, this happens because the cdf or the inverse cdf has an intractable functional form. In such cases, the Rejection Method can be very helpful — it provides a way to generate pseudo-random numbers from $f(x)$ (which we do *not* know how to do) by generating pseudo-random numbers from a density $g(x)$ (which we *do* know how to generate). Density $g(x)$ should have the following properties:

- $g(x)$ is defined over the same domain as $f(x)$, and
- there exists a constant $c > 0$ such that $\frac{f(x)}{g(x)} \leq c$ for all x. That is, $c = \sup\left(\frac{f(x)}{g(x)}\right)$.

Let x_g denote a pseudo-random drawing from $g(x)$, and let u denote a pseudo-random drawing from the Unifom(0, 1) distribution. Then, the *Rejection Method* generates pseudo-random drawings from $f(x)$ in three steps:

The Rejection Method	
(1)	Generate x_g and u.
(2)	If $u \leq \frac{1}{c} \frac{f(x_g)}{g(x_g)}$, accept x_g as a random selection from $f(x)$.
(3)	Else, return to step (1).

To illustrate, let $f(x)$ denote the pdf of a Birnbaum–Saunders distribution, with parameters α and β. This distribution has been used to represent the lifetime of components. We wish to generate pseudo-random drawings from $f(x)$ when say $\alpha = \frac{1}{2}$, $\beta = 4$:

$$\mathbf{f} = \frac{e^{-\frac{(x-\beta)^2}{2\alpha^2 \beta x}} (x+\beta)}{2\alpha\sqrt{2\pi\beta}\ x^{3/2}} \ /.\ \left\{\alpha \to \frac{1}{2},\ \beta \to 4\right\};$$

$$\mathbf{domain[f]} = \{x, 0, \infty\}\ \&\&\ \{\alpha > 0, \beta > 0\};$$

The Inverse Method will be of little help to us here, because *Mathematica* Version 4 cannot find the cdf of this distribution. Instead, we try the Rejection Method. We start by choosing a density $g(x)$. Suitable choices for $g(x)$ might include the Lognormal or the Levy (§2.6 B) or the Chi-squared(n), because each of these distributions has a similar shape to $f(x)$; this is easy to verify with a plot. We use Chi-squared(n) here, with $n = 4$:

$$\mathbf{g} = \frac{x^{n/2-1}\, e^{-x/2}}{2^{n/2}\,\Gamma[\frac{n}{2}]} \ /.\ n \to 4; \qquad \mathbf{domain[g]} = \{x, 0, \infty\};$$

Note that $g(x)$ is defined over the same domain as $f(x)$. Moreover, we can easily check whether $c = \sup\left(\frac{f(x)}{g(x)}\right)$ exists, by doing a quick plot of $\frac{f(x)}{g(x)}$.

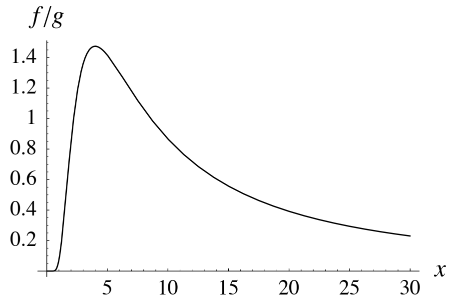


Fig. 27: $\frac{f(x)}{g(x)}$ plotted as a function of x

This suggests that c is roughly equal to 1.45. We can find the value of c more accurately using numerical methods:

$$\mathbf{c} = \mathbf{FindMaximum}\left[\frac{f}{g}, \{x, 3, 6\}\right][[1]]$$

```
1.4739
```

We can easily generate pseudo-random drawings x_g from $g(x)$ using *Mathematica*'s `Statistics` package:

```
<< Statistics`
dist = ChiSquareDistribution[4];    xg = Random[dist]
```

```
18.8847
```

By step (2) of the Rejection Method, we accept x_g as a random selection from $f(x)$ if $u \le Q(x_g)$, where $Q(x_g) = \frac{1}{c}\frac{f(x_g)}{g(x_g)}$. We enter $Q(x)$ into *Mathematica* as follows:

$$Q[x\_] = \frac{1}{c}\frac{f}{g} \text{ // Simplify}$$

$$\frac{29.5562\, e^{-8/x}\,(4+x)}{x^{5/2}}$$

Steps (1) – (3) can now be modelled in just one line, by setting up a recursive function. In the following input, note how x_g (a pseudo-random Chi-squared drawing) is used to generate x_f (a pseudo-random Birnbaum–Saunders drawing):

```
xf  :=
  (xg = Random[dist];  u = Random[];  If[u ≤ Q[xg], xg, xf])
```

So, let us try it out … here are 10000 pseudo-random Birnbaum–Saunders drawings:

```
data = Table[xf, {10000}];
```

Check the fit:

```
FrequencyPlot[data, f];
```

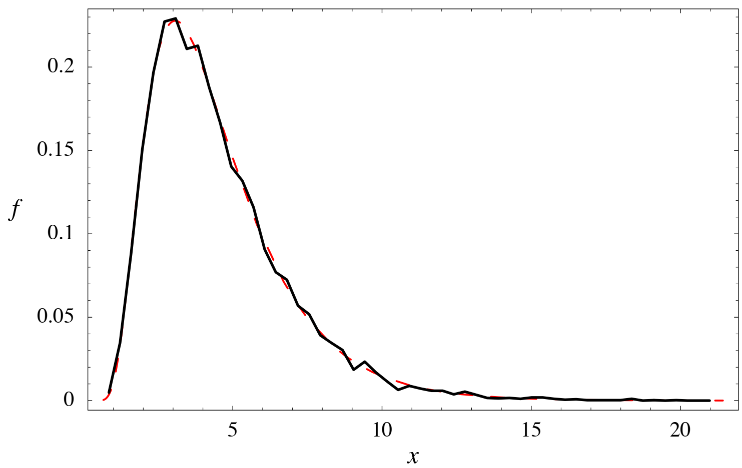


Fig. 28: The empirical pdf (——) and true pdf (– – –)

The Rejection Method is most useful when working with densities $f(x)$ that are not covered by *Mathematica*'s Statistics package, and for which the symbolic Inverse Method does not work. When using the Rejection Method, density $g(x)$ should be chosen so that it is easy to generate from, and is as similar in shape to $f(x)$ as possible. It is also worth checking that, at each stage of the process, output is numerical.

2.7 Exercises

1. Let continuous random variable X have a semi-Circular (half-Halo) distribution with pdf $f(x) = 2\sqrt{1-x^2}\,/\pi$ and domain of support $x \in (-1, 1)$. Plot the density $f(x)$. Find the cdf $P(X \leq x)$ and plot it. Find the mean and the variance of X.

2. Azzalini (1985) showed that if random variable X has a pdf $f(x)$ that is symmetric about zero, with cdf $F(x)$, then $2\, f(x)\, F(\lambda\, x)$ is also a pdf, for parameter $\lambda \in \mathbb{R}$. In particular, when X is $N(0, 1)$, the density $g(x) = 2\, f(x)\, F(\lambda\, x)$ is known as Azzalini's skew-Normal distribution. Find $g(x)$. Plot density $g(x)$ when $\lambda = 0$, 1 and 2. Find the mean and variance. Find upper and lower bounds on the variance.

3. Let $X \sim \text{Lognormal}(\mu, \sigma)$. Find the r^{th} raw moment, the cdf, p^{th} quantile, and mode.

4. Let $f(x)$ denote a standard Normal pdf; further, let pdf $g(x) = (2\pi)^{-1}(1 + \cos(x))$, with domain of support $x \in (-\pi, \pi)$. Compare $f(x)$ with $g(x)$ by plotting both on a diagram. From the plot, which distribution has greater kurtosis? Verify your choice by calculating Pearson's measure of kurtosis.

5. Find the y^{th} quantile for a standard Triangular distribution. Hence, find the median.

6. Let $X \sim \text{InverseGaussian}(\mu, \sigma)$ with pdf $f(x)$. Find the first 3 negative moments (*i.e.* $E[X^{-1}]$, $E[X^{-2}]$, $E[X^{-3}]$). Find the mgf, if it exists.

7. Let X have pdf $f(x) = \text{Sech}[x]/\pi$, $x \in \mathbb{R}$, which is known as the Hyperbolic Secant distribution. Derive the cf, and then the first 12 raw moments. Why are the odd-order moments zero?

8. Find the characteristic function of X^2, if $X \sim N(\mu, \sigma^2)$.

9. Find the cdf of the stable distribution $S(\frac{2}{3}, -1)$ as an exact symbolic entity.

10. The distribution of IQ in Class E2 at Rondebosch Boys High School is $X \sim N(\mu, \sigma^2)$. Mr Broster, the class teacher, decides to break the class into two streams: Stream 1 for those with IQ $> \omega$, and Stream 2 for those with IQ $\leq \omega$.

(i) Find the average (expected) IQ in each stream, for any chosen value of ω.
(ii) If $\mu = 100$ and $\sigma = 16$, plot (on one diagram) the average IQ in each stream as a function of ω.
(iii) If $\mu = 100$ and $\sigma = 16$, how should Mr Broster choose ω if he wants:
(a) the same number of students in each stream?
(b) the average IQ of Stream 1 to be twice the average of Stream 2?
For each case (a)–(b), find the average IQ in each stream.

11. Apple Computer is planning to host a live webcast of the next Macworld Conference. Let random variable X denote the number of people (measured in thousands) wanting to watch the live webcast, with pdf $f(x) = \frac{1}{144}\, e^{-x/12}\, x$, for $x > 0$. Find the expected number of people who want to watch the webcast. If Apple's web server can handle at most c simultaneous live streaming connections (measured in thousands), find the expected number of people who will be able to watch the webcast as a function of c. Plot the solution as a function of c.

12. Generate 20000 pseudo-random drawings from Azzalini's ($\lambda = 1$) skew-Normal distribution (see Exercise 2), using the exact inverse method (symbolic).