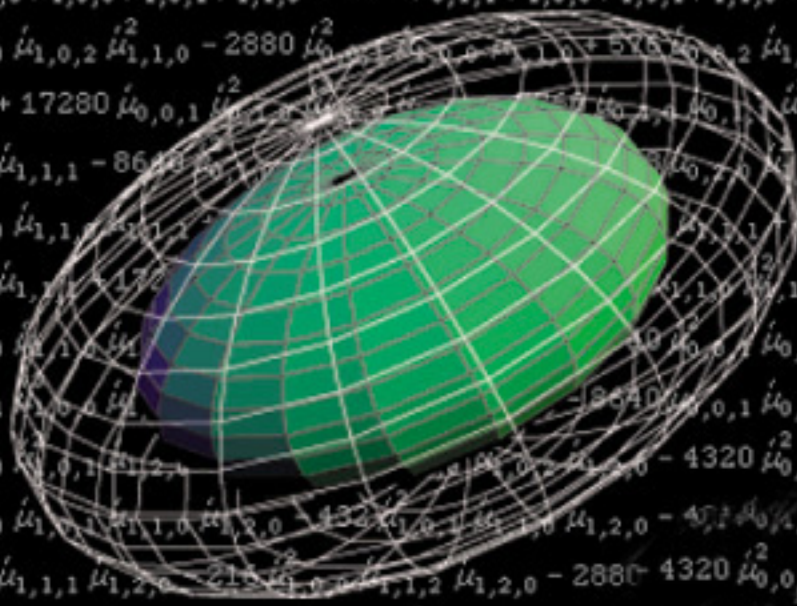
SPRINGER TEXTS IN STATISTICS

MATHEMATICAL STATISTICS *with* *Mathematica*®



COLIN ROSE
MURRAY D. SMITH

Mathematical Statistics with *Mathematica*

Chapter 6 – Multivariate Distributions

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 6

Multivariate Distributions

6.1 Introduction

Thus far, we have considered the distribution of a single random variable. This chapter extends the analysis to a collection of random variables $\vec{X} = (X_1, X_2, \ldots, X_m)$. When $m = 2$, we have a bivariate setting; when $m = 3$, a trivariate … and so on. Although the transition from univariate to multivariate analysis is 'natural', it does introduce some new concepts, in particular: joint densities §6.1 A, non-rectangular domains §6.1 B, joint distribution functions §6.1 C, marginal distributions §6.1 D, and conditional distributions §6.1 E. Multivariate expectations, product moments, generating functions and multivariate moment conversion functions are discussed in §6.2. Next, §6.3 examines the properties of independence and dependence. §6.4 is devoted to the multivariate Normal, §6.5 discusses the multivariate t and the multivariate Cauchy, while §6.6 looks at the Multinomial distribution and the bivariate Poisson distribution.

6.1 A Joint Density Functions

○ ***Continuous Random Variables***

Let $\vec{X} = (X_1, \ldots, X_m)$ denote a collection of m random variables defined on a domain of support $\Lambda \subset \mathbb{R}^m$, where we assume Λ is an open set in $\mathbb{R}^m$. Then a function $f : \Lambda \to \mathbb{R}_+$ is a joint *probability density function* (pdf) if it has the following properties:

$$\begin{aligned} f(x_1, \ldots, x_m) &> 0, \quad \text{for } (x_1, \ldots, x_m) \in \Lambda \\ \int \cdots \int_{\Lambda} f(x_1, \ldots, x_m)\, dx_1 \cdots dx_m &= 1 \end{aligned} \tag{6.1}$$

⊕ ***Example 1:*** Joint pdf

Consider the function $f(x, y)$ with domain of support $\Lambda = \{(x, y) : 0 < x < \infty, 0 < y < \infty\}$:

$$\mathbf{f} = \frac{e^{\frac{-1-x}{y}}\, x}{y^4}; \qquad \mathbf{domain[f] = \{\{x, 0, \infty\}, \{y, 0, \infty\}\};}$$

Clearly, f is positive over its domain, and it integrates to unity over the domain:

```
Integrate[f, {x,0,∞}, {y,0,∞}]
```

```
1
```

Thus, $f(x, y)$ may represent the joint pdf of a pair of random variables. Figure 1 plots $f(x, y)$ over part of its support.

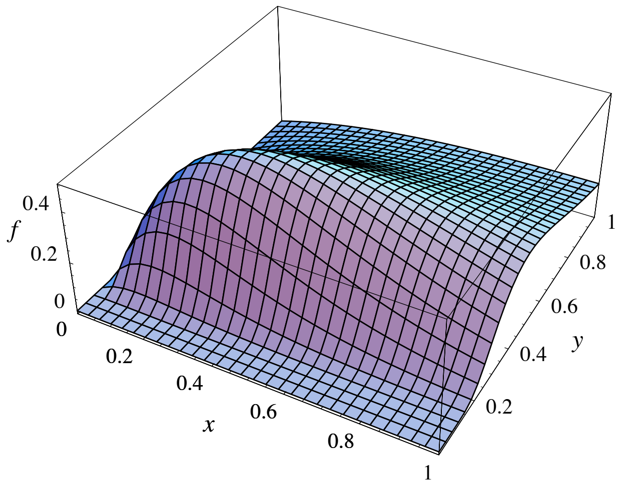


Fig. 1: The joint pdf $f(x, y)$

A contour plot allows one to pick out specific contours along which $z = f(x, y)$ is constant. That is, each contour joins points on the surface that have the same height z. Figure 2 plots all combinations of x and y such that $f(x, y) = \frac{1}{30}$. The edge of the dark-shaded region is the contour line.

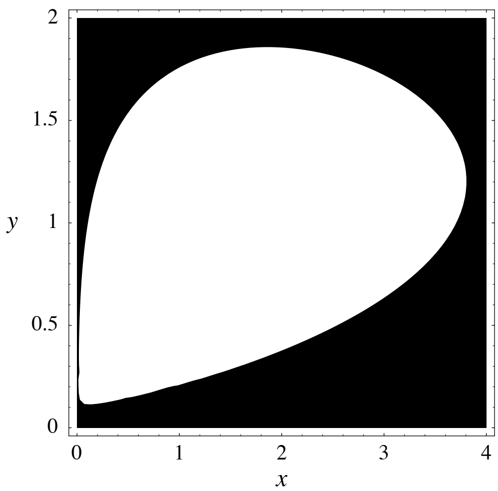


Fig. 2: The contour $f(x, y) = \frac{1}{30}$

○ ***Discrete Random Variables***

Let $\vec{X} = (X_1, \ldots, X_m)$ denote a collection of m random variables defined on a domain of support $\Lambda \subset \mathbb{R}^m$. Then a function $f : \Lambda \to \mathbb{R}_+$ is a joint *probability mass function* (pmf) if it has the following properties:

$$f(x_1, \ldots, x_m) = P(X_1 = x_1, \ldots, X_m = x_m) > 0, \text{ for } (x_1, \ldots, x_m) \in \Lambda$$

$$\sum \cdots \sum_{\Lambda} f(x_1, \ldots, x_m) = 1 \tag{6.2}$$

⊕ ***Example 2:*** Joint pmf

Let random variables X and Y have joint pmf $h(x, y) = \frac{x+1-y}{54}$ with domain of support $\Lambda = \{(x, y) : x \in \{3, 5, 7\},\ y \in \{0, 1, 2, 3\}\}$, as per Table 1.

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$
$X = 3$	$\frac{4}{54}$	$\frac{3}{54}$	$\frac{2}{54}$	$\frac{1}{54}$
$X = 5$	$\frac{6}{54}$	$\frac{5}{54}$	$\frac{4}{54}$	$\frac{3}{54}$
$X = 7$	$\frac{8}{54}$	$\frac{7}{54}$	$\frac{6}{54}$	$\frac{5}{54}$

Table 1: Joint pmf of $h(x, y) = \frac{x+1-y}{54}$

In *Mathematica*, this pmf may be entered as:

```
pmf = Table[(x + 1 - y)/54, {x, 3, 7, 2}, {y, 0, 3}]
```

$$\begin{pmatrix} \frac{2}{27} & \frac{1}{18} & \frac{1}{27} & \frac{1}{54} \\ \frac{1}{9} & \frac{5}{54} & \frac{2}{27} & \frac{1}{18} \\ \frac{4}{27} & \frac{7}{54} & \frac{1}{9} & \frac{5}{54} \end{pmatrix}$$

This is a well-defined pmf since all the probabilities are positive, and they sum to 1:

```
Plus @@ Plus @@ pmf
```

1

The latter can also be evaluated with:

```
Plus @@ (pmf // Flatten)
```

1

Figure 3 interprets the joint pmf in the form of a three-dimensional bar chart.

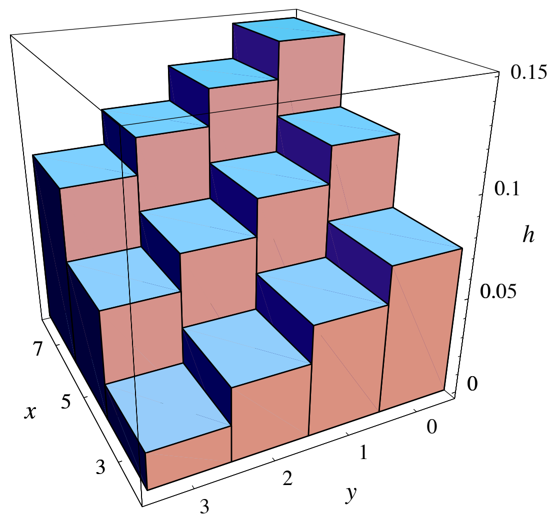


Fig. 3: Joint pmf of $h(x, y) = \frac{x+1-y}{54}$

6.1 B Non-Rectangular Domains

If the domain of a joint pdf does not depend on any of its constituent random variables, then we say the domain defines an independent product space. For instance, the domain $\{(x, y): \frac{1}{2} < x < 3,\ 1 < y < 4\}$ is an independent product space, because the domain of X does not depend on the domain of Y, and vice versa. We enter such domains into **mathStatica** as:

$$\texttt{domain[f]} = \left\{\left\{\texttt{x}, \frac{1}{2}, 3\right\}, \{\texttt{y}, 1, 4\}\right\}$$

If plotted, this domain would appear rectangular, as Fig. 4 illustrates. In this vein, we refer to such domains as being *rectangular*.

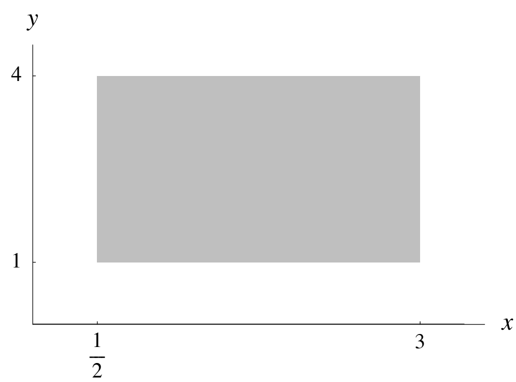


Fig. 4: A rectangular domain

Sometimes, the domain itself may depend on random variables. We refer to such domains as being *non-rectangular*. Examples include:

(i) $\{(x, y): 0 < x < y < \infty\}$. This would appear triangular in the two-dimensional plane. We can enter this domain into **mathStatica** as:

```
domain[f] = {{x, 0, y}, {y, x, ∞}}
```

(ii) $\{(x, y): x^2 + y^2 < 1\}$. This would appear circular in the two-dimensional plane. At present, **mathStatica** does not support such domains. However, this feature is planned for a future version of **mathStatica**, once *Mathematica* itself can support multiple integration over inequality defined regions.

6.1 C Probability and `Prob`

○ ***Continuous Random Variables***

Given some joint pdf $f(x_1, \ldots, x_m)$, the joint *cumulative distribution function* (cdf) is given by:

$$P(X_1 \le x_1, \ldots, X_m \le x_m) = \int_{-\infty}^{x_m} \cdots \int_{-\infty}^{x_1} f(w_1, \ldots, w_m)\, dw_1 \cdots dw_m. \tag{6.3}$$

The **mathStatica** function `Prob[{`$x_1, \ldots, x_m$`}, `f`]` calculates $P(X_1 \le x_1, \ldots, X_m \le x_m)$. The position of each element $\{x_1, x_2, \ldots\}$ in `Prob[{`$x_1, \ldots, x_m$`}, `f`]` is important, and must correspond to the ordering specified in the `domain` statement.

⊕ ***Example 3:*** Joint cdf

Consider again the joint pdf given in *Example 1*:

$$\texttt{f} = \frac{e^{\frac{-1-x}{y}}\, x}{y^4}; \qquad \texttt{domain[f] = \{\{x, 0, }\infty\texttt{\}, \{y, 0, }\infty\texttt{\}\};}$$

Here is the cdf $F(x, y) = P(X \le x, Y \le y)$:

```
F = Prob[{x, y}, f]
```

$$e^{-1/y}\left(1 - \frac{e^{-\frac{x}{y}}\,(x + x^2 + y + 2\,x\,y)}{(1 + x)^2\, y}\right)$$

Since $F(x, y)$ may be viewed as the anti-derivative of $f(x, y)$, differentiating `F` yields the original joint pdf $f(x, y)$:

```
D[F, x, y] // Simplify
```

$$\frac{e^{-\frac{1+x}{y}}\, x}{y^4}$$

Figure 5 plots the joint cdf.

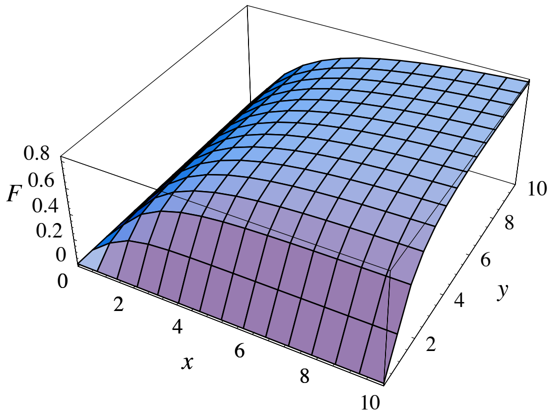


Fig. 5: The joint cdf $F(x, y)$

The surface approaches 1 asymptotically, which it reaches in the limit:

```
Prob[{∞, ∞}, f]
```

```
1
```

⊕ ***Example 4:*** Probability Content of a Region — Introducing `MrSpeedy`

Let $\vec{X} = (X_1, X_2, X_3)$ have joint pdf $g(x_1, x_2, x_3)$:

$$g = k\, e^{x_1}\, x_1\, (x_2 + 1) / x_3^2;$$
$$\text{domain}[g] = \{\{x_1, 0, 1\}, \{x_2, 2, 4\}, \{x_3, 3, 5\}\};$$

where the constant $k > 0$ is defined such that g integrates to unity over its domain. The cdf of g is:

```
Clear[G];
G[x1_, x2_, x3_] = Prob[{x1, x2, x3}, g]
```

$$\frac{k\,(1 + e^{x1}\,(-1 + x1))\,(-2 + x2)\,(4 + x2)\,(-3 + x3)}{6\,x3}$$

Note that we have set up `G` as a *Mathematica* function of `x1` through `x3`, and can thus apply it as a function in the standard way. Here, we find k by evaluating `G` at the upper boundary of the domain:

```
G[1, 4, 5]
```

$$\frac{16\,k}{15}$$

This requires $k = \frac{15}{16}$ in order for g to be a well-defined pdf. If we require the probability content of a region within the domain, we could just type in the whole integral. For instance, the probability of being within the region

$$S = \{(x_1, x_2, x_3): \quad 0 < x_1 < \tfrac{1}{2}, \quad 3 < x_2 < \tfrac{7}{2}, \quad 4 < x_3 < \tfrac{9}{2}\}$$

is given by:

$$\int_4^{\frac{9}{2}} \int_3^{\frac{7}{2}} \int_0^{\frac{1}{2}} \mathbf{g} \, d\mathbf{x_1} \, d\mathbf{x_2} \, d\mathbf{x_3}$$

$$\frac{17}{288} \left(1 - \frac{\sqrt{e}}{2}\right) k$$

While this is straightforward, it is by no means the fastest solution. In particular, the probability content of a region within the domain can be found purely by using the function G[] (which we have already found) *and* the boundaries of that region, without any need for further integration. *Note:* the solution is *not* $G[\frac{1}{2}, \frac{7}{2}, \frac{9}{2}] - G[0, 3, 4]$. Rather, one must evaluate the cdf at every possible extremum defined by set S. The **mathStatica** function MrSpeedy[cdf, S] does this.

? MrSpeedy

```
MrSpeedy[cdf, S] calculates the probability content
  of a region defined by set S, by making use of the
  known distribution function cdf[x1, x2, ..., xm].
```

For our example:

$$\mathbf{S = \left\{\left\{0, \frac{1}{2}\right\}, \left\{3, \frac{7}{2}\right\}, \left\{4, \frac{9}{2}\right\}\right\};}$$

MrSpeedy[G, S]

$$\frac{17}{288} \left(1 - \frac{\sqrt{e}}{2}\right) k$$

MrSpeedy typically provides at least a 20-fold speed increase over direct integration. To see the calculations MrSpeedy performs, replace G with say Φ:

MrSpeedy[Φ, S]

$$-\Phi[0, 3, 4] + \Phi\left[0, 3, \frac{9}{2}\right] + \Phi\left[0, \frac{7}{2}, 4\right] - \Phi\left[0, \frac{7}{2}, \frac{9}{2}\right] + \Phi\left[\frac{1}{2}, 3, 4\right] - \Phi\left[\frac{1}{2}, 3, \frac{9}{2}\right] - \Phi\left[\frac{1}{2}, \frac{7}{2}, 4\right] + \Phi\left[\frac{1}{2}, \frac{7}{2}, \frac{9}{2}\right]$$

MrSpeedy evaluates the cdf at each of these points. Note that this approach applies to any m-variate distribution. ■

○ ***Discrete Random Variables***

Given some joint pmf $f(x_1, \ldots, x_m)$, the joint cdf is

$$P(X_1 \le x_1, \ldots, X_m \le x_m) \;=\; \sum_{w_1 \le x_1} \cdots \sum_{w_m \le x_m} f(w_1, \ldots, w_m). \tag{6.4}$$

Note that the `Prob` function does not operate on multivariate *discrete* domains.

⊕ ***Example 5:*** Joint cdf

In *Example 2*, we considered the bivariate pmf $h(x, y) = \frac{x+1-y}{54}$ with domain of support $\Lambda = \{(x, y) : x \in \{3, 5, 7\},\ y \in \{0, 1, 2, 3\}\}$. The cdf, $H(x, y) = P(X \le x, Y \le y)$, can be defined in *Mathematica* as follows:

```
H[x_, y_] = Sum[(w1 + 1 - w2)/54, {w1, 3, x, 2}, {w2, 0, y}]
```

$$\frac{1}{108}\left(8 + 7\,y - y^2 + 10\,\text{Floor}\left[\frac{1}{2}\,(-3 + x)\right] + 9\,y\,\text{Floor}\left[\frac{1}{2}\,(-3 + x)\right] - y^2\,\text{Floor}\left[\frac{1}{2}\,(-3 + x)\right] + 2\,\text{Floor}\left[\frac{1}{2}\,(-3 + x)\right]^2 + 2\,y\,\text{Floor}\left[\frac{1}{2}\,(-3 + x)\right]^2\right)$$

Then, for instance, $P(X \le 5, Y \le 3)$ is:

```
H[5, 3]
```

$$\frac{14}{27}$$

Figure 6 plots the joint cdf as a three-dimensional bar chart.

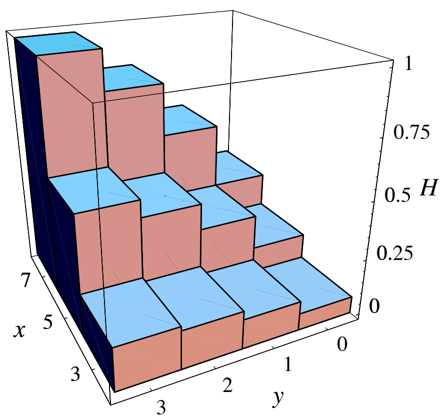


Fig. 6: The joint cdf $H(x, y)$

6.1 D Marginal Distributions

○ ***Continuous Random Variables***

Let the *continuous* random variables X_1 and X_2 have joint pdf $f(x_1, x_2)$. Then the *marginal pdf* of X_1 is $f_1(x_1)$, where

$$f_1(x_1) = \int_{x_2} f(x_1, x_2)\, dx_2. \tag{6.5}$$

More generally, if $(X_1, \ldots, X_m)$ have joint pdf $f(x_1, \ldots, x_m)$, then the marginal pdf of a group $r < m$ of these random variables is obtained by 'integrating out' the $(m - r)$ variables that are not of interest. The **mathStatica** function, `Marginal[`$\vec{x}_r, f$`]`, derives the marginal joint pdf of the variable(s) specified in $\vec{x}_r$. If there is more than one variable in $\vec{x}_r$, then it must take the form of a list. The ordering of the variables in this list does not matter.

⊕ ***Example 6:*** `Marginal`

Let the continuous random variables $\vec{X} = (X_1, X_2, X_3, X_4)$ have joint pdf $f(x_1, x_2, x_3, x_4)$:

$$\mathbf{f = k\, e^{x_1}\, x_1\, (x_2 + 1)\, (x_3 - 3)^2 / x_4^2;}$$
$$\mathbf{domain[f] = \{\{x_1, 0, 1\}, \{x_2, 1, 2\}, \{x_3, 2, 3\}, \{x_4, 3, 4\}\};}$$

where k is a constant. The marginal bivariate distribution of X_2 *and* X_4 is given by:

$$\mathbf{Marginal[\{x_2, x_4\}, f]}$$

$$\frac{k\,(1 + x_2)}{3\, x_4^2}$$

The resulting marginal density depends only on values of X_2 and X_4, since X_1 and X_3 have been integrated out. Similarly, the marginal distribution of X_4 does not depend on values of X_1, X_2 or X_3:

$$\mathbf{Marginal[x_4, f]}$$

$$\frac{5\,k}{6\, x_4^2}$$

We can use `Marginal` to determine k, by letting $\vec{x}_r$ be an empty set. Then all the random variables are 'integrated out':

```
Marginal[{}, f]
```

$$\frac{5\,k}{72}$$

Thus, in order for f to be a well-defined density function, k must equal $\frac{72}{5}$. ■

○ *Discrete Random Variables*

In a discrete world, the ∫ symbol in (6.5) is replaced by the summation symbol Σ. Thus, if the discrete random variables X_1 and X_2 have joint pmf $f(x_1, x_2)$, then the *marginal pmf* of X_1 is $f_1(x_1)$, where

$$f_1(x_1) = \sum_{x_2} f(x_1, x_2). \tag{6.6}$$

The `Marginal` function only operates on continuous domains; it is not currently implemented for discrete domains.

⊕ ***Example 7:*** Discrete Marginal

Recall, from *Example 2*, the joint pmf $h(x, y) = \frac{x+1-y}{54}$ with domain of support $\{(x, y) : x \in \{3, 5, 7\},\ y \in \{0, 1, 2, 3\}\}$:

```
pmf = Table[(x + 1 - y)/54, {x, 3, 7, 2}, {y, 0, 3}];
```

By (6.6), the marginal pmf of Y is:

```
pmfY = Sum[(x + 1 - y)/54, {x, 3, 7, 2}] // Simplify
```

$\frac{6-y}{18}$

where Y may take values of 0, 1, 2 or 3. That is:

```
pmfY /. y → {0, 1, 2, 3}
```

$\left\{\frac{1}{3}, \frac{5}{18}, \frac{2}{9}, \frac{1}{6}\right\}$

Alternatively, we can derive the same result directly, by finding the sum of each column of Table 1:

```
Plus @@ pmf
```

$\left\{\frac{1}{3}, \frac{5}{18}, \frac{2}{9}, \frac{1}{6}\right\}$

The sum of each row can be found with:

```
Plus @@ Transpose[pmf]
```

$\left\{\frac{5}{27}, \frac{1}{3}, \frac{13}{27}\right\}$

Further examples of discrete multivariate distributions are given in §6.6. ■

6.1 E Conditional Distributions

○ ***Continuous Random Variables***

Let the continuous random variables X_1 and X_2 have joint pdf $f(x_1, x_2)$. Then the *conditional pdf* of X_1 given $X_2 = x_2$ is denoted by $f(x_1 \mid X_2 = x_2)$ or, for short, $f(x_1 \mid x_2)$. It is defined by

$$f(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}, \qquad \text{provided } f_2(x_2) > 0 \tag{6.7}$$

where $f_2(x_2)$ denotes the marginal pdf of X_2 evaluated at $X_2 = x_2$. More generally, if $(X_1, \ldots, X_m)$ have joint pdf $f(x_1, \ldots, x_m)$, the joint conditional pdf of a group of r of these random variables (given that the remaining $m - r$ variables are fixed) is the joint pdf of the m variables divided by the joint marginal pdf of the $m - r$ fixed variables.

Since the conditional pdf $f(x_1 \mid x_2)$ is a well-defined pdf, we can use it to calculate probabilities and expectations. For instance, if $u(X_1)$ is a function of X_1, then the *conditional expectation* $E[u(X_1) \mid X_2 = x_2]$ is given by

$$E[u(X_1) \mid x_2] = \int_{x_1} u(x_1)\, f(x_1 \mid x_2)\, dx_1. \tag{6.8}$$

With **mathStatica**, conditional expectations are easily calculated by first deriving the conditional density, say $f_{\text{con}}(x_1) = f(x_1 \mid x_2)$ and `domain[`f_{con}`]`. The desired conditional expectation is then given by `Expect[`u, f_{con}`]`. Two particular examples of conditional expectations are the conditional mean $E[X_1 \mid x_2]$, which is known as the *regression function* of X_1 on X_2, and the conditional variance $\text{Var}(X_1 \mid x_2)$, which is known as the *scedastic function*.

⊕ ***Example 8:*** `Conditional`

The **mathStatica** function, `Conditional[`$\vec{x}_r$, f`]`, derives the conditional pdf of $\vec{x}_r$ variable(s), given that the remaining variables are fixed. As above, if there is more than one variable in $\vec{x}_r$, then it must take the form of a list; it does not matter how the variables in this list are sorted. To eliminate any confusion, a message clarifies what is (and what is not) being conditioned on. For density $f(x_1, x_2, x_3, x_4)$, defined in *Example 6*, the joint conditional pdf of X_2 and X_4, given $X_1 = x_1$ and $X_3 = x_3$ is:

```
Conditional[{x₂, x₄}, f]
```

– Here is the conditional pdf $f(x_2, x_4 \mid x_1, x_3)$:

$$\frac{24\,(1 + x_2)}{5\, x_4^2}$$

Note that this output is the same as the first `Marginal` example above (given $k = \frac{72}{5}$). This is because (X_1, X_2, X_3, X_4) are mutually stochastically independent (see §6.3 A). ■

⊕ ***Example 9:*** Conditional Expectation (Continuous)

Let X_1 and X_2 have joint pdf $f(x_1, x_2) = x_1 + x_2$, supported on the unit rectangle $\{(x_1, x_2) : 0 < x_1 < 1, 0 < x_2 < 1\}$:

```
f = x₁ + x₂;     domain[f] = {{x₁, 0, 1}, {x₂, 0, 1}};
```

as illustrated below in Fig. 7. Derive the conditional mean and conditional variance of X_1, given $X_2 = x_2$.

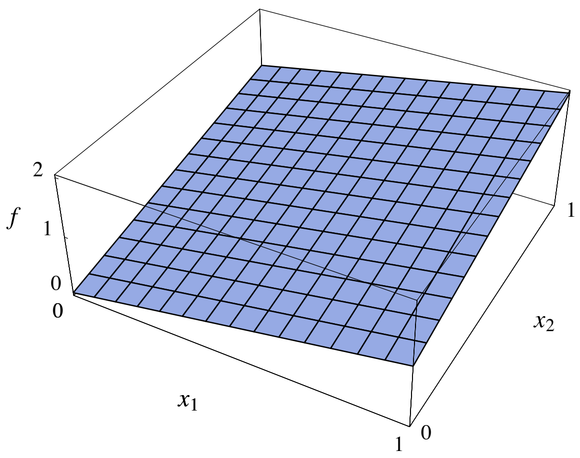


Fig. 7: The joint pdf $f(x_1, x_2) = x_1 + x_2$

Solution: The conditional pdf $f(x_1 \mid x_2)$, denoted f_{con}, is:[1]

```
fcon = Conditional[x₁, f]
```

— Here is the conditional pdf $f(x_1 \mid x_2)$:

$$\frac{x_1 + x_2}{\frac{1}{2} + x_2}$$

In order to apply **mathStatica** functions to the conditional pdf f_{con}, we need to declare the domain over which it is defined. This is because **mathStatica** will only recognise f_{con} as a pdf if its domain has been specified. Since random variable X_2 is now fixed at x_2, the domain of f_{con} is:

```
domain[fcon] = {x₁, 0, 1};
```

The required conditional mean is:

```
Expect[x₁, fcon]
```

$$\frac{2 + 3\,x_2}{3 + 6\,x_2}$$

The conditional variance is:

```
Var[x1, fcon]
```

$$\frac{1+6\,x_2+6\,x_2^2}{18\,(1+2\,x_2)^2}$$

As this result depends on X_2, the conditional variance is heteroscedastic. ■

○ *Discrete Random Variables*

The transition to a discrete world is once again straightforward: if the discrete random variables, X_1 and X_2, have joint pmf $f(x_1, x_2)$, then the *conditional pmf* of X_2 given $X_1 = x_1$ is denoted by $f(x_2 \mid X_1 = x_1)$ or, for short, $f(x_2 \mid x_1)$. It is defined by

$$f(x_2 \mid x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}, \qquad \text{provided } f_1(x_1) > 0 \tag{6.9}$$

where $f_1(x_1)$ denotes the marginal pmf of X_1, evaluated at $X_1 = x_1$, as defined in (6.6). Note that **mathStatica**'s `Conditional` function only operates on continuous domains; it is not implemented for discrete domains. As above, the conditional pmf $f(x_2 \mid x_1)$ can be used to calculate probabilities and expectations. Thus, if $u(X_2)$ is a function of X_2, the *conditional expectation* $E[u(X_2) \mid X_1 = x_1]$ is given by

$$E[u(X_2) \mid x_1] = \sum_{x_2} u(x_2)\, f(x_2 \mid x_1). \tag{6.10}$$

⊕ ***Example 10:*** Conditional Mean (Discrete)

Find the conditional mean of X, given $Y = y$, for the pmf $h(x, y) = \frac{x+1-y}{54}$ with domain of support $\{(x, y) : x \in \{3, 5, 7\},\ y \in \{0, 1, 2, 3\}\}$.

Solution: We require $E[X \mid Y = y] = \sum_x x\, h(x \mid y) = \sum_x x\, \frac{h(x, y)}{h_y(y)}$. In *Example 7*, we found that the marginal pmf of Y was $h_y(y) = \frac{6-y}{18}$. Hence, the solution is:

```
sol = Sum[x (x + 1 - y)/54 / (6 - y)/18, {x, 3, 7, 2}] // Simplify
```

$$\frac{98-15\,y}{18-3\,y}$$

This depends, of course, on $Y = y$. Since we can assign four possible values to y, the four possible conditional expectations $E[X \mid Y = y]$ are:

```
sol /. y → {0, 1, 2, 3}
```

$$\left\{\frac{49}{9}, \frac{83}{15}, \frac{17}{3}, \frac{53}{9}\right\}$$

6.2 Expectations, Moments, Generating Functions

6.2 A Expectations

Let the collection of m random variables $(X_1, \ldots, X_m)$ have joint density function $f(x_1, \ldots, x_m)$. Then the *expectation* of some function u of the random variables, $u(X_1, \ldots, X_m)$, is

$$E\big[u(X_1, \ldots, X_m)\big] = \begin{cases} \int_{x_m} \cdots \int_{x_1} u(x_1, \ldots, x_m)\, f(x_1, \ldots, x_m)\, dx_1 \cdots dx_m \\ \sum_{x_1} \cdots \sum_{x_m} u(x_1, \ldots, x_m)\, f(x_1, \ldots, x_m) \end{cases} \tag{6.11}$$

corresponding to the continuous and discrete cases, respectively. **mathStatica**'s `Expect` function generalises neatly to a multivariate continuous setting. For instance, in §6.1 D, we considered the following pdf $g(x_1, x_2, x_3, x_4)$:

```
g = 72/5 ⅇ^x1 x1 (x2 + 1) (x3 - 3)^2 / x4^2;
domain[g] = {{x1, 0, 1}, {x2, 1, 2}, {x3, 2, 3}, {x4, 3, 4}};
```

We now find both $E[X_1(X_4^2 - X_2)]$ and $E[X_4]$:

```
Expect[x1 (x4^2 - x2), g]
```

$\frac{157}{15}(-2 + e)$

```
Expect[x4, g]
```

$12 \,\text{Log}\left[\frac{4}{3}\right]$

6.2 B Product Moments, Covariance and Correlation

Multivariate moments are a special type of multivariate expectation. To illustrate, let X_1 and X_2 have joint bivariate pdf $f(x_1, x_2)$. Then, the bivariate *raw moment* $\acute{\mu}_{r,s}$ is

$$\acute{\mu}_{r,s} = E[X_1^r X_2^s]\,. \tag{6.12}$$

With $s = 0$, $\acute{\mu}_{r,0}$ denotes the r^{th} raw moment of X_1. Similarly, with $r = 0$, $\acute{\mu}_{0,s}$ denotes the s^{th} raw moment of X_2. More generally, $\acute{\mu}_{r,s}$ is known as a *product* raw moment or joint raw moment. These definitions extend in the obvious way to higher numbers of variables.

The bivariate *central moment* $\mu_{r,s}$ is defined as

$$\mu_{r,s} = E\big[(X_1 - E[X_1])^r (X_2 - E[X_2])^s\big]. \tag{6.13}$$

The *covariance* of X_i and X_j, denoted $\mathrm{Cov}(X_i, X_j)$, is defined by

$$\mathrm{Cov}(X_i, X_j) = E\big[(X_i - E[X_i])(X_j - E[X_j])\big]. \tag{6.14}$$

When $i = j$, $\mathrm{Cov}(X_i, X_j)$ is equivalent to $\mathrm{Var}(X_i)$. More generally, the *variance-covariance* matrix of $\vec{X} = (X_1, X_2, \ldots, X_m)$ is the $(m \times m)$ symmetric matrix:

$$\mathrm{Varcov}(\vec{X}) = E\left[\big(\vec{X} - E[\vec{X}]\big)\big(\vec{X} - E[\vec{X}]\big)^{\mathrm{T}}\right]$$

$$= E\left[\begin{pmatrix} X_1 - EX_1 \\ X_2 - EX_2 \\ \vdots \\ X_m - EX_m \end{pmatrix} \Big((X_1 - EX_1),\ (X_2 - EX_2),\ \ldots,\ (X_m - EX_m) \Big)\right]$$

$$= E\left[\begin{pmatrix} (X_1 - EX_1)^2 & (X_1 - EX_1)(X_2 - EX_2) & \cdots & (X_1 - EX_1)(X_m - EX_m) \\ (X_2 - EX_2)(X_1 - EX_1) & (X_2 - EX_2)^2 & \cdots & (X_2 - EX_2)(X_m - EX_m) \\ \vdots & \vdots & \ddots & \vdots \\ (X_m - EX_m)(X_1 - EX_1) & (X_m - EX_m)(X_2 - EX_2) & \cdots & (X_m - EX_m)^2 \end{pmatrix}\right]$$

$$= \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_m) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_m) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_m, X_1) & \mathrm{Cov}(X_m, X_2) & \cdots & \mathrm{Var}(X_m) \end{pmatrix}$$

It follows from (6.14) that $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$, and thus that the variance-covariance matrix is symmetric. In the notation of (6.13), one could alternatively express $\mathrm{Varcov}(\vec{X})$ as follows:

$$\mathrm{Varcov}(\vec{X}) = \begin{pmatrix} \mu_{2,0,0,\ldots,0} & \mu_{1,1,0,\ldots,0} & \cdots & \mu_{1,0,\ldots,0,1} \\ \mu_{1,1,0,\ldots,0} & \mu_{0,2,0,\ldots,0} & \cdots & \mu_{0,1,\ldots,0,1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{1,0,\ldots,0,1} & \mu_{0,1,0,\ldots,1} & \cdots & \mu_{0,0,\ldots,0,2} \end{pmatrix} \tag{6.15}$$

which again highlights its symmetry.

Finally, the *correlation* between X_i and X_j is defined as

$$\rho(X_i, X_j) = \rho_{ij} = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i)\,\mathrm{Var}(X_j)}} \tag{6.16}$$

where it can be shown that $-1 \le \rho_{ij} \le 1$. If X_i and X_j are mutually stochastically independent (§6.3 A), then $\rho_{ij} = 0$; the converse does not always hold (see *Example 16*).

⊕ ***Example 11:*** Product Moments, `Cov`, `Varcov`, `Corr`

Let the continuous random variables X, Y and Z have joint pdf $f(x, y, z)$:

```
f = 1/Sqrt[2 π λ] e^(-x^2/2 - z/λ) (1 + α (2 y - 1) Erf[x/Sqrt[2]]);
domain[f] = {{x, -∞, ∞}, {y, 0, 1}, {z, 0, ∞}}
              && {-1 < α < 1, λ > 0};
```

$$f = \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{x^2}{2}-\frac{z}{\lambda}} \left(1+\alpha\,(2y-1)\,\mathrm{Erf}\left[\frac{x}{\sqrt{2}}\right]\right);$$

The mean vector is $\vec{\mu} = E[(X, Y, Z)]$:

```
Expect[{x, y, z}, f]
```

$\left\{0, \frac{1}{2}, \lambda\right\}$

Here is the product raw moment $\acute{\mu}_{3,2,1} = E[X^3\, Y^2\, Z]$:

```
Expect[x^3 y^2 z, f]
```

$$\frac{5\,\alpha\,\lambda}{12\,\sqrt{\pi}}$$

Here is the product central moment $\mu_{2,0,2} = E\left[(X - E[X])^2\,(Z - E[Z])^2\right]$:

```
Expect[ (x - Expect[x, f])^2 (z - Expect[z, f])^2, f]
```

λ^2

Cov(X, Y) is given by:

```
Cov[{x, y}, f]
```

$$\frac{\alpha}{6\,\sqrt{\pi}}$$

More generally, the variance-covariance matrix is:

```
Varcov[f]
```

$$\begin{pmatrix} 1 & \frac{\alpha}{6\sqrt{\pi}} & 0 \\ \frac{\alpha}{6\sqrt{\pi}} & \frac{1}{12} & 0 \\ 0 & 0 & \lambda^2 \end{pmatrix}$$

The correlation between X and Y is:

```
Corr[{x, y}, f]
```

$$\frac{\alpha}{\sqrt{3\,\pi}}$$

6.2 C Generating Functions

The multivariate *moment generating function* (mgf) is a natural extension to the univariate case defined in Chapter 2. Let $\vec{X} = (X_1, \ldots, X_m)$ denote an m-variate random variable, and let $\vec{t} = (t_1, \ldots, t_m) \in \mathbb{R}^m$ denote a vector of dummy variables. Then the mgf $M_{\vec{X}}(\vec{t})$ is a function of $\vec{t}$; when no confusion is possible, we denote $M_{\vec{X}}(\vec{t})$ by $M(\vec{t})$. It is defined by

$$M(\vec{t}) = E\left[e^{\vec{t}.\vec{X}}\right] = E[e^{t_1 X_1 + \cdots + t_m X_m}] \tag{6.17}$$

provided the expectation exists for all $t_i \in (-c, c)$, for some constant $c > 0$, $i = 1, \ldots, m$. If it exists, the mgf can be used to generate the product raw moments. In, say, a bivariate setting, the product raw moment $\acute{\mu}_{r,s} = E[X_1^r X_2^s]$ may be obtained from $M(\vec{t})$ as follows:

$$\acute{\mu}_{r,s} = E[X_1^r X_2^s] = \left.\frac{\partial^{r+s} M(\vec{t})}{\partial t_1^r \, \partial t_2^s}\right|_{\vec{t}=\vec{0}} . \tag{6.18}$$

The *central moment generating function* may be obtained from the mgf (6.17) as follows:

$$E\left[e^{\vec{t}.(\vec{X}-\vec{\mu})}\right] = e^{-\vec{t}.\vec{\mu}} M(\vec{t}), \qquad \text{where } \vec{\mu} = E[\vec{X}] . \tag{6.19}$$

The *cumulant generating function* is the natural logarithm of the mgf. The multivariate *characteristic function* is similar to (6.17) and given by

$$C(\vec{t}) = E\left[\exp\left(i\,\vec{t}.\vec{X}\right)\right] = E\left[\exp\left(i\,(t_1 X_1 + t_2 X_2 + \cdots + t_m X_m)\right)\right] \tag{6.20}$$

where i denotes the unit imaginary number.

Given discrete random variables defined on subsets of the non-negative integers $\{0, 1, 2, \ldots\}$, the multivariate *probability generating function* (pgf) is

$$\Pi(\vec{t}) = E[t_1^{X_1} t_2^{X_2} \cdots t_m^{X_m}] . \tag{6.21}$$

The pgf provides a way to determine the probabilities. For instance, in the bivariate case,

$$P(X_1 = r, X_2 = s) = \frac{1}{r!\,s!} \left.\frac{\partial^{r+s} \Pi(\vec{t})}{\partial t_1^r \, \partial t_2^s}\right|_{\vec{t}=\vec{0}} . \tag{6.22}$$

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

$$\begin{aligned} \acute{\mu}[r, s] &= E[X_1^{[r]} X_2^{[s]}] \\ &= E[X_1(X_1 - 1) \cdots (X_1 - r + 1) \times X_2(X_2 - 1) \cdots (X_2 - s + 1)] \end{aligned} \tag{6.23}$$

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

$$\acute{\mu}[r, s] = E[X_1^{[r]} X_2^{[s]}] = \left.\frac{\partial^{r+s} \Pi(\vec{t})}{\partial t_1^r \, \partial t_2^s}\right|_{\vec{t}=\vec{1}} . \tag{6.24}$$

Note that $\vec{t}$ is set here to $\vec{1}$ and not $\vec{0}$. To then convert from factorial moments to product raw moments, see the `FactorialToRaw` function of §6.2 D.

⊕ ***Example 12:*** Working with Generating Functions

Gumbel (1960) considered a bivariate Exponential distribution with cdf given by:

$$F = 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta x y)};$$

for $0 \leq \theta \leq 1$. Because X and Y are continuous random variables, the joint pdf $f(x, y)$ may be obtained by differentiation:

```
      f  = D[F, x, y] // Simplify
domain[f] = {{x, 0, ∞}, {y, 0, ∞}} && {0 < θ < 1};
```

$$e^{-x-y-x y \theta} (1 + (-1 + x + y) \theta + x y \theta^2)$$

This is termed a bivariate Exponential distribution because its marginal distributions are standard Exponential. For instance:

```
Marginal[x, f]
```

$$e^{-x}$$

Here is the mgf (this takes about 100 seconds on our reference machine):

```
t⃗ = {t₁, t₂};   V⃗ = {x, y};   mgf = Expect[e^(t⃗.V⃗), f]
```

— This further assumes that: $\{t_1 < 1, \text{Arg}[\frac{-1+t_2}{\theta}] \neq 0\}$

$$-\frac{t_1}{-1+t_1} + \frac{1}{1-t_2} + \frac{1}{\theta^2} \left(e^{\frac{(-1+t_1)(-1+t_2)}{\theta}} \left(\text{MeijerG}\left[\{\{\}, \{1\}\}, \{\{0, 0\}, \{\}\}, \frac{(-1+t_1)(-1+t_2)}{\theta}\right] (-1+t_1)(1+(-1+\theta) t_2) + \text{ExpIntegralE}\left[1, \frac{(-1+t_1)(-1+t_2)}{\theta}\right] (1 - t_1 + (-1+\theta+t_1) t_2)\right)\right)$$

where the condition $\text{Arg}[\frac{-1+t_2}{\theta}] \neq 0$ is just *Mathematica*'s way of saying $t_2 < 1$. We can now obtain any product raw moment $\acute{\mu}_{r,s} = E[X_1^r X_2^s]$ from the mgf, as per (6.18). For instance, $\acute{\mu}_{3,4} = E[X_1^3 X_2^4]$ is given by:

```
D[ mgf, {t₁, 3}, {t₂, 4}] /. t_ → 0   // FullSimplify
```

$$\frac{12 \theta (1 + \theta (5 + 2 \theta)) - 12 e^{\frac{1}{\theta}} (1 + 6 \theta (1 + \theta)) \text{Gamma}[0, \frac{1}{\theta}]}{\theta^6}$$

If we plan to do many of these calculations, it is convenient to write a little *Mathematica* function, Moment[r, s] $= E[X^r\, Y^s]$, to automate this calculation:

```
Moment[r_, s_] :=
 D[ mgf, {t₁, r}, {t₂, s}] /. t_ → 0  // FullSimplify
```

Then $\acute{\mu}_{3,4}$ is now given by:

```
Moment[3, 4]
```

$$\frac{12\,\theta\,(1+\theta\,(5+2\,\theta)) - 12\,e^{\frac{1}{\theta}}\,(1+6\,\theta\,(1+\theta))\,\text{Gamma}[0, \frac{1}{\theta}]}{\theta^6}$$

Just as we derived the 'mgf about the origin' above, we can also derive the 'mgf about the mean' (*i.e.* the central mgf). To do so, we first need the mean vector $\vec{\mu} = (E[X], E[Y])$, given by:

```
μ⃗ = {Moment[1, 0], Moment[0, 1]}
```

{1, 1}

Then, by (6.19), the centralised mgf is:

```
mgfc = e^(-t⃗.μ⃗) mgf;
```

Just as differentiating the mgf yields raw moments, differentiating the centralised mgf yields central moments. In particular, the variances and the covariance of X and Y can be obtained using the following function:

```
MyCov[i_, j_] := D[mgfc, tᵢ, tⱼ] /. t_ → 0 // FullSimplify
```

which we apply as follows:

```
Array[MyCov, {2, 2}]
```

$$\begin{pmatrix} 1 & -1 + \frac{e^{\frac{1}{\theta}}\,\text{Gamma}[0, \frac{1}{\theta}]}{\theta} \\ -1 + \frac{e^{\frac{1}{\theta}}\,\text{Gamma}[0, \frac{1}{\theta}]}{\theta} & 1 \end{pmatrix}$$

To see how this works, evaluate:

```
Array[σ, {2, 2}]
```

$$\begin{pmatrix} \sigma[1, 1] & \sigma[1, 2] \\ \sigma[2, 1] & \sigma[2, 2] \end{pmatrix}$$

We could, of course, alternatively derive the variance-covariance matrix directly with Varcov[f], which takes roughly 6 seconds to evaluate on our reference machine. ■

6.2 D Moment Conversion Formulae

The moment converter functions introduced in Chapter 2 extend naturally to a multivariate setting. Using these functions, one can express any multivariate moment ($\acute{\mu}$, μ or κ) in terms of any other moment ($\acute{\mu}$, μ or κ). The supported conversions are:

function	*description*
RawToCentral[{r, s, …}]	not implemented
RawToCumulant[{r, s, …}]	$\acute{\mu}_{r,s,\ldots}$ in terms of $\kappa_{i,j,\ldots}$
CentralToRaw[{r, s, …}]	$\mu_{r,s,\ldots}$ in terms of $\acute{\mu}_{i,j,\ldots}$
CentralToCumulant[{r, s, …}]	$\mu_{r,s,\ldots}$ in terms of $\kappa_{i,j,\ldots}$
CumulantToRaw[{r, s, …}]	$\kappa_{r,s,\ldots}$ in terms of $\acute{\mu}_{i,j,\ldots}$
CumulantToCentral[{r, s, …}]	$\kappa_{r,s,\ldots}$ in terms of $\mu_{i,j,\ldots}$
and	
RawToFactorial[{r, s, …}]	$\acute{\mu}_{r,s,\ldots}$ in terms of $\acute{\mu}[i, j, \ldots]$
FactorialToRaw[{r, s, …}]	$\acute{\mu}[r, s]$ in terms of $\acute{\mu}_{i,j}$

Table 2: Multivariate moment conversion functions

⊕ ***Example 13:*** Express Cov(X, Y) in terms of Raw Moments

Solution: By (6.13), the covariance between X and Y is the central moment $\mu_{1,1}(X, Y)$. Thus, to express the covariance in terms of raw moments, we use the function `CentralToRaw[{1,1}]`:

CentralToRaw[{1, 1}]

$$\mu_{1,1} \to -\acute{\mu}_{0,1}\,\acute{\mu}_{1,0} + \acute{\mu}_{1,1}$$

This is just the well-known result that $\mu_{1,1} = E[X\,Y] - E[Y]\,E[X]$. ■

Cook (1951) gives *raw → cumulant* conversions and *central → cumulant* conversions, as well as the inverse relations *cumulant → raw* and *cumulant → central*, all in a bivariate world with $r + s \leq 6$; see also Stuart and Ord (1994, Section 3.29). With **mathStatica**, we can derive these relations on the fly. Here is the bivariate raw moment $\acute{\mu}_{3,2}$ expressed in terms of bivariate cumulants:

RawToCumulant[{3, 2}]

$$\acute{\mu}_{3,2} \to \kappa_{0,1}^2\,\kappa_{1,0}^3 + \kappa_{0,2}\,\kappa_{1,0}^3 + 6\,\kappa_{0,1}\,\kappa_{1,0}^2\,\kappa_{1,1} + 6\,\kappa_{1,0}\,\kappa_{1,1}^2 + 3\,\kappa_{1,0}^2\,\kappa_{1,2} + 3\,\kappa_{0,1}^2\,\kappa_{1,0}\,\kappa_{2,0} + 3\,\kappa_{0,2}\,\kappa_{1,0}\,\kappa_{2,0} + 6\,\kappa_{0,1}\,\kappa_{1,1}\,\kappa_{2,0} + 3\,\kappa_{1,2}\,\kappa_{2,0} + 6\,\kappa_{0,1}\,\kappa_{1,0}\,\kappa_{2,1} + 6\,\kappa_{1,1}\,\kappa_{2,1} + 3\,\kappa_{1,0}\,\kappa_{2,2} + \kappa_{0,1}^2\,\kappa_{3,0} + \kappa_{0,2}\,\kappa_{3,0} + 2\,\kappa_{0,1}\,\kappa_{3,1} + \kappa_{3,2}$$

Working 'about the mean' (*i.e.* set $\kappa_{1,0} = \kappa_{0,1} = 0$) yields the `CentralToCumulant` conversions. Here is:

```
CentralToCumulant[{3, 2}]
```

$$\mu_{3,2} \to 3\,\kappa_{1,2}\,\kappa_{2,0} + 6\,\kappa_{1,1}\,\kappa_{2,1} + \kappa_{0,2}\,\kappa_{3,0} + \kappa_{3,2}$$

The inverse relations are given by `CumulantToRaw` and `CumulantToCentral`. Here, for instance, is the trivariate cumulant $\kappa_{2,1,1}$ expressed in terms of trivariate raw moments:

```
CumulantToRaw[{2, 1, 1}]
```

$$\kappa_{2,1,1} \to -6\,\acute{\mu}_{0,0,1}\,\acute{\mu}_{0,1,0}\,\acute{\mu}_{1,0,0}^2 + 2\,\acute{\mu}_{0,1,1}\,\acute{\mu}_{1,0,0}^2 + 4\,\acute{\mu}_{0,1,0}\,\acute{\mu}_{1,0,0}\,\acute{\mu}_{1,0,1} + 4\,\acute{\mu}_{0,0,1}\,\acute{\mu}_{1,0,0}\,\acute{\mu}_{1,1,0} - 2\,\acute{\mu}_{1,0,1}\,\acute{\mu}_{1,1,0} - 2\,\acute{\mu}_{1,0,0}\,\acute{\mu}_{1,1,1} + 2\,\acute{\mu}_{0,0,1}\,\acute{\mu}_{0,1,0}\,\acute{\mu}_{2,0,0} - \acute{\mu}_{0,1,1}\,\acute{\mu}_{2,0,0} - \acute{\mu}_{0,1,0}\,\acute{\mu}_{2,0,1} - \acute{\mu}_{0,0,1}\,\acute{\mu}_{2,1,0} + \acute{\mu}_{2,1,1}$$

The converter functions extend to any arbitrarily large variate system, of any weight. Here is the input for a 4-variate cumulant $\kappa_{3,1,3,1}$ of weight 8 expressed in terms of central moments:

```
CumulantToCentral[{3, 1, 3, 1}]
```

The same expression in raw moments is about 5 times longer and contains 444 different terms. It takes less than a second to evaluate:

```
Length[CumulantToRaw[{3, 1, 3, 1}][[2]]] // Timing
```

```
{0.383333 Second, 444}
```

Factorial moments were discussed in §6.2 C, and are applied in §6.6 B. David and Barton (1957, p. 144) list multivariate *factorial* → *raw* conversions up to weight 4, along with the inverse relation *raw* → *factorial*. With **mathStatica**, we can again derive these relations on the fly. Here is the bivariate factorial moment $\acute{\mu}[3, 2]$ expressed in terms of bivariate raw moments:

```
FactorialToRaw[{3, 2}]
```

$$\acute{\mu}[3,2] \to -2\,\acute{\mu}_{1,1} + 2\,\acute{\mu}_{1,2} + 3\,\acute{\mu}_{2,1} - 3\,\acute{\mu}_{2,2} - \acute{\mu}_{3,1} + \acute{\mu}_{3,2}$$

and here is a trivariate `RawToFactorial` conversion of weight 7:

```
RawToFactorial[{4, 1, 2}]
```

$$\acute{\mu}_{4,1,2} \to \acute{\mu}[1,1,1] + \acute{\mu}[1,1,2] + 7\,\acute{\mu}[2,1,1] + 7\,\acute{\mu}[2,1,2] + 6\,\acute{\mu}[3,1,1] + 6\,\acute{\mu}[3,1,2] + \acute{\mu}[4,1,1] + \acute{\mu}[4,1,2]$$

○ *The Converter Functions in Practice*

Sometimes, one might know how to derive one class of moments (say raw moments) but not another (say cumulants), or vice versa. In such situations, the converter functions come to the rescue, for they enable one to derive any moment ($\acute{\mu}$, μ or κ), provided one class of moments can be calculated. This section illustrates how this can be done. The general approach is as follows: first, we express the desired moment (say $\kappa_{2,1}$) in terms of moments that we can calculate (say raw moments):

```
CumulantToRaw[{2, 1}]
```

$$\kappa_{2,1} \to 2\,\acute{\mu}_{0,1}\,\acute{\mu}_{1,0}^{2} - 2\,\acute{\mu}_{1,0}\,\acute{\mu}_{1,1} - \acute{\mu}_{0,1}\,\acute{\mu}_{2,0} + \acute{\mu}_{2,1}$$

and then we evaluate each raw moment $\acute{\mu}_{\_}$ for the relevant distribution. This can be done in two ways:

Method (i): derive $\acute{\mu}_{\_}$ from a known mgf
Method (ii): derive $\acute{\mu}_{\_}$ directly using the `Expect` function.

Examples 14 and *15* illustrate the two approaches, respectively.

⊕ ***Example 14:*** Method (i)

Find $\mu_{2,1,2}$ for Cheriyan and Ramabhadran's multivariate Gamma distribution.

Solution: Kotz *et al.* (2000, p. 456) give the joint mgf of Cheriyan and Ramabhadran's *m*-variate Gamma distribution as follows:

$$\texttt{GammaMGF[m\_] :=} \left(1 - \sum_{j=1}^{m} t_j\right)^{-\theta_0} \prod_{j=1}^{m} (1 - t_j)^{-\theta_j}$$

So, for a trivariate system, the mgf is:

```
mgf = GammaMGF[3]
```

$$(1 - t_1)^{-\theta_1}\,(1 - t_2)^{-\theta_2}\,(1 - t_3)^{-\theta_3}\,(1 - t_1 - t_2 - t_3)^{-\theta_0}$$

The desired central moment $\mu_{2,1,2}$ can be expressed in terms of raw moments:

```
sol = CentralToRaw[{2, 1, 2}]
```

$$\begin{aligned}\mu_{2,1,2} \to\ & 4\,\acute{\mu}_{0,0,1}^{2}\,\acute{\mu}_{0,1,0}\,\acute{\mu}_{1,0,0}^{2} - \\ & \acute{\mu}_{0,0,2}\,\acute{\mu}_{0,1,0}\,\acute{\mu}_{1,0,0}^{2} - 2\,\acute{\mu}_{0,0,1}\,\acute{\mu}_{0,1,1}\,\acute{\mu}_{1,0,0}^{2} + \acute{\mu}_{0,1,2}\,\acute{\mu}_{1,0,0}^{2} - \\ & 4\,\acute{\mu}_{0,0,1}\,\acute{\mu}_{0,1,0}\,\acute{\mu}_{1,0,0}\,\acute{\mu}_{1,0,1} + 2\,\acute{\mu}_{0,1,0}\,\acute{\mu}_{1,0,0}\,\acute{\mu}_{1,0,2} - \\ & 2\,\acute{\mu}_{0,0,1}^{2}\,\acute{\mu}_{1,0,0}\,\acute{\mu}_{1,1,0} + 4\,\acute{\mu}_{0,0,1}\,\acute{\mu}_{1,0,0}\,\acute{\mu}_{1,1,1} - \\ & 2\,\acute{\mu}_{1,0,0}\,\acute{\mu}_{1,1,2} - \acute{\mu}_{0,0,1}^{2}\,\acute{\mu}_{0,1,0}\,\acute{\mu}_{2,0,0} + 2\,\acute{\mu}_{0,0,1}\,\acute{\mu}_{0,1,0}\,\acute{\mu}_{2,0,1} - \\ & \acute{\mu}_{0,1,0}\,\acute{\mu}_{2,0,2} + \acute{\mu}_{0,0,1}^{2}\,\acute{\mu}_{2,1,0} - 2\,\acute{\mu}_{0,0,1}\,\acute{\mu}_{2,1,1} + \acute{\mu}_{2,1,2}\end{aligned}$$

Here, each term $\acute{\mu}_{r,s,v}$ denotes $\acute{\mu}_{r,s,v}(X, Y, Z) = E[X^r\, Y^s\, Z^v]$, which we can, in turn, find by differentiating the mgf. Since we wish to do this many times, let us write a little *Mathematica* function, `Moment[`r, s, v`]` $= E[X^r\, Y^s\, Z^v]$, to automate this calculation:

```
Moment[r_, s_, v_] :=
 D[mgf, {t₁, r}, {t₂, s}, {t₃, v}] /. t_ → 0
```

Then, the solution is:

```
sol /. μ́k__ :→ Moment[k] // Simplify
```

$\mu_{2,1,2} \to 2\,\theta_0\ (12 + 10\,\theta_0 + \theta_1 + \theta_3)$

An alternative solution to this particular problem, without using the converter functions, is to first find the mean vector $\acute{\mu} = \{E[X], E[Y], E[Z]\}$:

```
μ⃗ = {Moment[1, 0, 0], Moment[0, 1, 0], Moment[0, 0, 1]}
```

$\{\theta_0 + \theta_1,\ \theta_0 + \theta_2,\ \theta_0 + \theta_3\}$

Second, find the central mgf, by (6.19):

```
t⃗ = {t₁, t₂, t₃};   mgfc = ⅇ^(-t⃗.μ⃗) mgf
```

$$e^{-t_1(\theta_0+\theta_1)-t_2(\theta_0+\theta_2)-t_3(\theta_0+\theta_3)}\,(1-t_1)^{-\theta_1}\,(1-t_2)^{-\theta_2}\,(1-t_3)^{-\theta_3}\,(1-t_1-t_2-t_3)^{-\theta_0}$$

Then, differentiating the central mgf yields the desired central moment $\mu_{2,1,2}$ again:

```
D[mgfc, {t₁, 2}, {t₂, 1}, {t₃, 2}] /. t_ → 0 // Simplify
```

$2\,\theta_0\ (12 + 10\,\theta_0 + \theta_1 + \theta_3)$

⊕ ***Example 15:*** Method (ii)

Let random variables X and Y have joint density $f(x, y)$:

$$f = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}-2y}\left(e^y + \alpha\,(e^y - 2)\,\text{Erf}\left[\frac{x}{\sqrt{2}}\right]\right);$$

```
domain[f] = {{x, -∞, ∞}, {y, 0, ∞}} && {-1 < α < 1};
```

For the given density, find the product cumulant $\kappa_{2,2}$.

Solution: If we knew the mgf, we could immediately derive the cumulant generating function. Unfortunately, *Mathematica* Version 4 can not derive the mgf; nor is it likely to be listed in any textbook, because this is not a common distribution. To resolve this problem, we will make use of the moment conversion formulae. The desired solution, $\kappa_{2,2}$, expressed in terms of raw moments, is:

```
sol = CumulantToRaw[{2, 2}]
```

$$\kappa_{2,2} \to -6\,\acute{\mu}_{0,1}^2\,\acute{\mu}_{1,0}^2 + 2\,\acute{\mu}_{0,2}\,\acute{\mu}_{1,0}^2 + 8\,\acute{\mu}_{0,1}\,\acute{\mu}_{1,0}\,\acute{\mu}_{1,1} - 2\,\acute{\mu}_{1,1}^2 - 2\,\acute{\mu}_{1,0}\,\acute{\mu}_{1,2} + 2\,\acute{\mu}_{0,1}^2\,\acute{\mu}_{2,0} - \acute{\mu}_{0,2}\,\acute{\mu}_{2,0} - 2\,\acute{\mu}_{0,1}\,\acute{\mu}_{2,1} + \acute{\mu}_{2,2}$$

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

```
sol /. μ́r_,s_ :→ Expect[x^r y^s, f] // Simplify
```

$$\kappa_{2,2} \to -\frac{\alpha^2}{2\,\pi}$$

The calculation takes about 6 seconds on our reference machine. ■

6.3 Independence and Dependence

6.3 A Stochastic Independence

Let random variables $\vec{X} = (X_1, \ldots, X_m)$ have joint pdf $f(x_1, \ldots, x_m)$, with marginal density functions $f_1(x_1), \ldots, f_m(x_m)$. Then $(X_1, \ldots, X_m)$ are said to be *mutually stochastically independent* if and only if

$$f(x_1, \ldots, x_m) \;=\; f_1(x_1) \times \cdots \times f_m(x_m)\,. \tag{6.25}$$

That is, the joint pdf is equal to the product of the marginal pdf's. A number of well-known theorems apply to mutually stochastically independent random variables, which we state here without proof. In particular:

	If $(X_1, \ldots, X_m)$ are mutually stochastically independent, then:
(i)	$P(a \le X_1 \le b, \ldots, c \le X_m \le d) \;=\; P(a \le X_1 \le b) \times \cdots \times P(c \le X_m \le d)$
(ii)	$E[u_1(X_1) \cdots u_m(X_m)] \;=\; E[u_1(X_1)] \times \cdots \times E[u_m(X_m)]$ for arbitrary functions $u_i(\cdot)$
(iii)	$M(t_1, \ldots, t_m) \;=\; M(t_1) \times \cdots \times M(t_m)$ mgf of the joint distribution = product of the mgf's of the marginal distributions
(iv)	$\text{Cov}(X_i, X_j) \;=\; 0$ for all $i \ne j$ However, zero covariance does *not* imply independence.

Table 3: Properties of mutually stochastic independent random variables

⊕ ***Example 16:*** Stochastic Dependence and Correlation

Let the random variables X, Y and Z have joint pdf $h(x, y, z)$:

```
h = Exp[-1/2 (x^2 + y^2 + z^2)] (1 + x y z Exp[-1/2 (x^2 + y^2 + z^2)]) / (2 π)^(3/2);
domain[h] = {{x, -∞, ∞}, {y, -∞, ∞}, {z, -∞, ∞}};
```

Since the product of the marginal pdf's:

```
Marginal[x, h] Marginal[y, h] Marginal[z, h]
```

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

… is *not* equal to the joint pdf $h(x, y, z)$, it follows by (6.25) that X, Y and Z are mutually stochastically *dependent*. Even though X, Y and Z are mutually dependent, their correlations ρ_{ij} $(i \neq j)$ are all zero:

```
Varcov[h]
```

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Clearly, zero correlation does not imply independence. ■

6.3 B Copulae

Copulae provide a method for constructing multivariate distributions from known marginal distributions. We shall only consider the bivariate case here. For more detail, see Joe (1997) and Nelsen (1999).

Let the continuous random variable X have pdf $f(x)$ and cdf $F(x)$; similarly, let the continuous random variable Y have pdf $g(y)$ and cdf $G(y)$. We wish to create a bivariate distribution $H(x, y)$ from these marginals. The joint distribution function $H(x, y)$ is given by

$$H(x, y) = C(F, G) \tag{6.26}$$

where C denotes the copula function. Then, the joint pdf $h(x, y)$ is given by

$$h(x, y) = \frac{\partial^2 H(x, y)}{\partial x\, \partial y}. \tag{6.27}$$

Table 4 lists some examples of copulae.

copula	formula	restrictions
Independent	$C = F\,G$	
Morgenstern	$C = F\,G\,(1 + \alpha(1 - F)\,(1 - G))$	$-1 < \alpha < 1$
Ali–Mikhail–Haq	$C = \dfrac{F\,G}{1 - \alpha(1 - F)\,(1 - G)}$	$-1 \le \alpha \le 1$
Frank	$C = -\frac{1}{\alpha}\log\left[1 + \dfrac{(e^{-\alpha F} - 1)\,(e^{-\alpha G} - 1)}{e^{-\alpha} - 1}\right]$	$\alpha \neq 0$

Table 4: Copulae

With the exception of the independent case, each copula in Table 4 includes parameter α. This term induces a new parameter into the joint bivariate distribution $h(x, y)$, which gives added flexibility. In each case, setting parameter $\alpha = 0$ (or taking the limit $\alpha \to 0$, in the Frank case) yields the independent copula $C = F\,G$ as a special case. When $\alpha = 1$, the Ali–Mikhail–Haq copula simplifies to $C = \frac{F\,G}{F + G - F\,G}$, as used in Exercise 8.

In the following two examples, we shall work with the Morgenstern (1956) copula.[2] We enter it as follows:

```
ClearAll[F, G]
Copula := F G (1 + α (1 - F) (1 - G))
```

⊕ ***Example 17:*** Bivariate Uniform (à la Morgenstern)

Let $X \sim$ Uniform(0, 1) with pdf $f(x)$ and cdf $F(x)$, and let $Y \sim$ Uniform(0, 1) with pdf $g(y)$ and cdf $G(y)$:

```
f = 1;   domain[f] = {x, 0, 1};    F = Prob[x, f];
g = 1;   domain[g] = {y, 0, 1};    G = Prob[y, g];
```

Let $h(x, y)$ denote the bivariate Uniform obtained via a Morgenstern copula. Then:

```
h = D[Copula, x, y] // Simplify
```

```
1 + (-1 + 2 x) (-1 + 2 y) α
```

with domain of support:

```
domain[h] = {{x, 0, 1}, {y, 0, 1}} && {-1 < α < 1};
```

Figure 8 plots the joint pdf $h(x, y)$ when $\alpha = \frac{1}{2}$. Clicking the 'View Animation' button in the electronic notebook brings up an animation of $h(x, y)$, allowing parameter α to vary from -1 to 1 in step sizes of $\frac{1}{10}$. This provides a rather neat way to visualise positive and negative correlation.

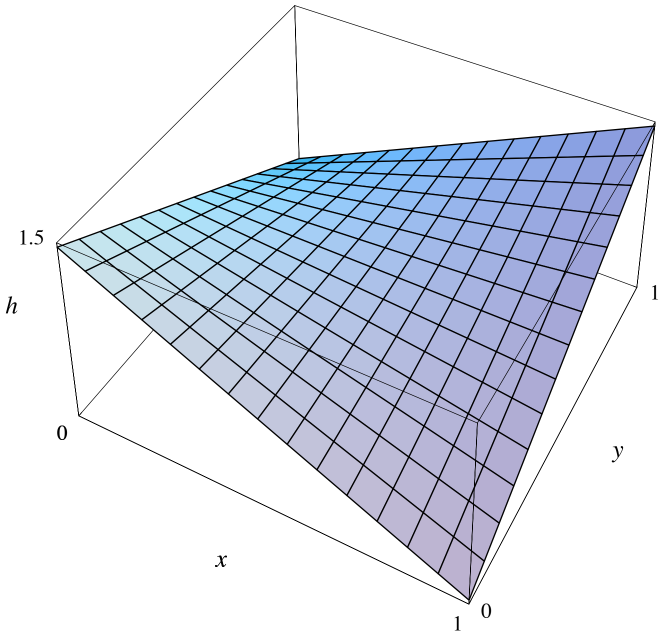


Fig. 8: Bivariate Uniform joint pdf $h(x, y)$ when $\alpha = \frac{1}{2}$

We already know the joint cdf $H(x, y) = P(X \le x, Y \le y)$, which is just the copula function:

```
Copula
```

$x\,y\,(1 + (1 - x)\,(1 - y)\,\alpha)$

The variance-covariance matrix is given by:

```
Varcov[h]
```

$$\begin{pmatrix} \frac{1}{12} & \frac{\alpha}{36} \\ \frac{\alpha}{36} & \frac{1}{12} \end{pmatrix}$$

⊕ ***Example 18:*** Normal–Uniform Bivariate Distribution (à la Morgenstern)

Let $X \sim N(0, 1)$ with pdf $f(x)$ and cdf $F(x)$, and let $Y \sim \text{Uniform}(0, 1)$ with pdf $g(y)$ and cdf $G(y)$:

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

```
g = 1;          domain[g] = {y, 0, 1};     G = Prob[y, g];
```

Let $h(x, y)$ denote the bivariate distribution obtained via a Morgenstern copula. Then:

```
h = D[Copula, x, y]  // Simplify
```

$$\frac{e^{-\frac{x^2}{2}}\left(1+(-1+2\,y)\,\alpha\,\mathrm{Erf}\left[\frac{x}{\sqrt{2}}\right]\right)}{\sqrt{2\,\pi}}$$

with domain of support:

```
domain[h] = {{x, -∞, ∞}, {y, 0, 1}} && {-1 ≤ α ≤ 1};
```

Figure 9 plots the joint pdf $h(x, y)$ when $\alpha = 0$.

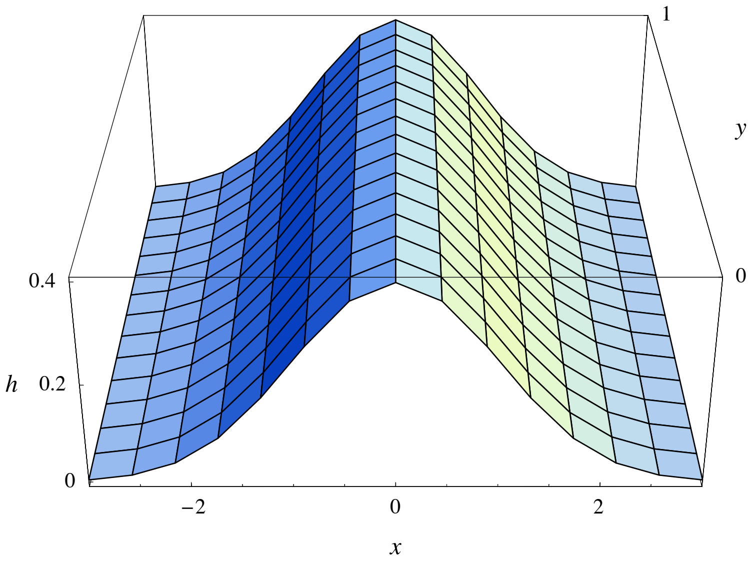


Fig. 9: Normal–Uniform joint pdf $h(x, y)$ when $\alpha = 0$

The joint cdf $H(x, y) = P(X \leq x, Y \leq y)$ is the copula function:

```
Copula // Simplify
```

$$\frac{1}{2}\,y\left(1+\frac{1}{2}\,(-1+y)\,\alpha\left(-1+\mathrm{Erf}\left[\frac{x}{\sqrt{2}}\right]\right)\right)\left(1+\mathrm{Erf}\left[\frac{x}{\sqrt{2}}\right]\right)$$

We can confirm that the marginal distributions are in fact Normal and Uniform, respectively:

```
Marginal[x, h]
Marginal[y, h]
```

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

$$1$$

The variance-covariance matrix is:

```
Varcov[h]
```

$$\begin{pmatrix} 1 & \frac{\alpha}{6\sqrt{\pi}} \\ \frac{\alpha}{6\sqrt{\pi}} & \frac{1}{12} \end{pmatrix}$$

Let $h_c(y)$ denote the conditional density function of Y, given $X = x$:

```
hc = Conditional[y, h]
```

— Here is the conditional pdf $h(y \mid x)$:

$$1 + (-1 + 2\,y)\,\alpha\,\text{Erf}\left[\frac{x}{\sqrt{2}}\right]$$

with domain:

```
domain[hc] = {y, 0, 1} && {-1 ≤ α ≤ 1};
```

Then, the conditional mean $E[Y \mid X = x]$ is:

```
Expect[y, hc]
```

$$\frac{1}{6}\left(3 + \alpha\,\text{Erf}\left[\frac{x}{\sqrt{2}}\right]\right)$$

and the conditional variance $\text{Var}(Y \mid X = x)$ is:

```
Var[y, hc]
```

$$\frac{1}{36}\left(3 - \alpha^2\,\text{Erf}\left[\frac{x}{\sqrt{2}}\right]^2\right)$$

Figure 10 plots the conditional mean and the conditional variance, when X and Y are correlated ($\alpha = 1$) and uncorrelated ($\alpha = 0$).

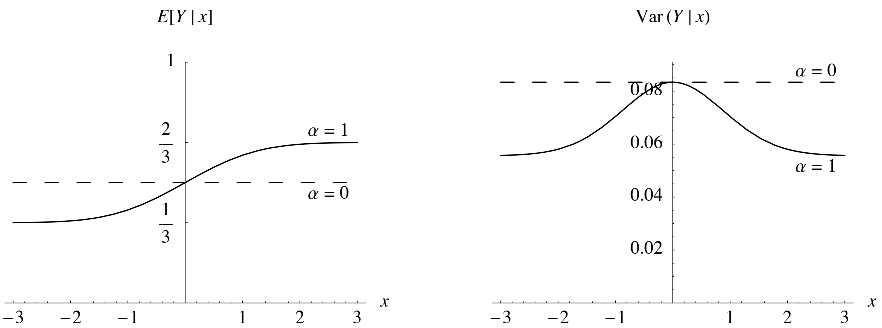


Fig. 10: Conditional mean and variance

6.4 The Multivariate Normal Distribution

The *Mathematica* package, `Statistics`MultinormalDistribution``, has several functions that are helpful throughout this section. We load this package as follows:

```
<< Statistics`
```

The multivariate Normal distribution is pervasive throughout statistics, so we devote an entire section to it and to some of its properties. Given $\vec{X} = (X_1, \ldots, X_m)$, we denote the m-variate *multivariate Normal distribution* by $N(\vec{\mu}, \Sigma)$, with mean vector $\vec{\mu} = (\mu_1, \ldots, \mu_m) \in \mathbb{R}^m$, variance-covariance matrix Σ, and joint pdf

$$f(\vec{x}) = (2\pi)^{-m/2} \left|\Sigma\right|^{-1/2} \exp\left(-\tfrac{1}{2} (\vec{x} - \vec{\mu})^{\mathrm{T}} \Sigma^{-1} (\vec{x} - \vec{\mu})\right) \tag{6.28}$$

where $\vec{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$, and Σ is a symmetric, positive definite $(m \times m)$ matrix. When $m = 1$, (6.28) simplifies to the univariate Normal pdf.

6.4 A The Bivariate Normal

Let random variables X_1 and X_2 have a bivariate Normal distribution, with zero mean vector, and variance-covariance matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Here, ρ denotes the correlation coefficient between X_1 and X_2. That is:

$\vec{X} = \{x_1, x_2\}; \quad \vec{\mu} = \{0, 0\}; \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix};$

```
dist2 = MultinormalDistribution[μ⃗, Σ];
```

Then, we enter our bivariate Normal pdf $f(x_1, x_2)$ as:

```
       f  = PDF[dist2, X⃗] // Simplify
domain[f] = Thread[{X⃗, -∞, ∞}] && {-1 < ρ < 1}
```

$$\frac{e^{\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{-2+2\rho^2}}}{2\pi\sqrt{1-\rho^2}}$$

$\{\{x_1, -\infty, \infty\}, \{x_2, -\infty, \infty\}\}\ \&\&\ \{-1 < \rho < 1\}$

where the `PDF` and `MultinormalDistribution` functions are defined in *Mathematica*'s `Statistics` package.

When $\rho = 0$, the cdf can be expressed in terms of the built-in error function as:[3]

```
F0 = Prob[{x1, x2}, f /. ρ → 0]
```

$$\frac{1}{4}\left(1 + \mathrm{Erf}\left[\frac{x_1}{\sqrt{2}}\right]\right)\left(1 + \mathrm{Erf}\left[\frac{x_2}{\sqrt{2}}\right]\right)$$

○ ***Diagrams***

Figure 11 plots the zero correlation pdf and cdf.

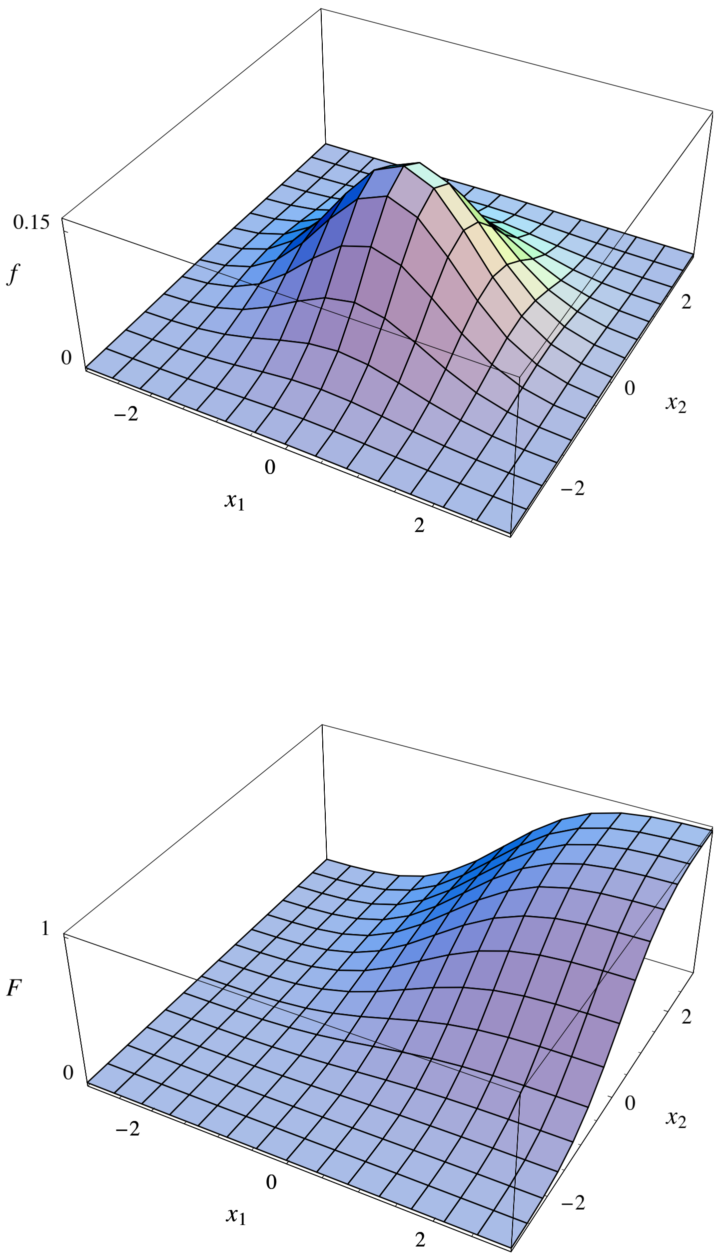


Fig. 11: The bivariate Normal joint pdf f (top) and joint cdf F (bottom), when $\rho = 0$

The shape of the contours of $f(x_1, x_2)$ depends on ρ, as Fig. 12 illustrates with a set of contour plots.

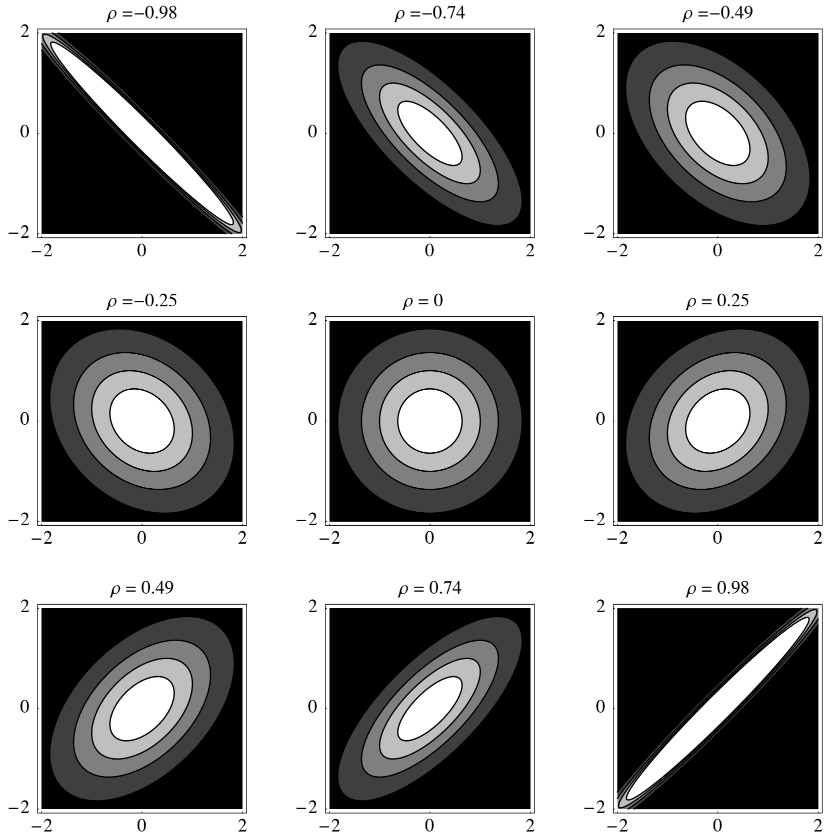


Fig. 12: Contour plots of the bivariate Normal pdf, for different values of ρ

Each plot corresponds to a specific value of ρ. In the top left corner, $\rho = -0.98$ (almost perfect negative correlation), whereas in the bottom right corner, $\rho = 0.98$ (almost perfect positive correlation). The middle plot corresponds to the case of zero correlation. In any given plot, the edge of each shaded region represents the contour line, and each contour is a two-dimensional ellipse along which f is constant. The ellipses are aligned along the $x_1 = x_2$ line when $\rho > 0$, or the $x_1 = -x_2$ line when $\rho < 0$.

We can even plot the specific ellipse that encloses q% of the distribution by using the `EllipsoidQuantile[`*dist*, *q*`]` function in *Mathematica*'s `Statistics` package. This is illustrated in Fig. 13, which plots the ellipses that enclose 15% (bold), 90% (dashed) and 99% (plain) of the distribution, respectively, when ρ is 0.6. Figure 14 superimposes 1000 pseudo-random drawings from this distribution on top of Fig. 13. On average, we would expect around 1% of the simulated data to lie outside the 99% quantile. For this particular set of simulated data, there are 11 such points (the large dots in Fig. 14).

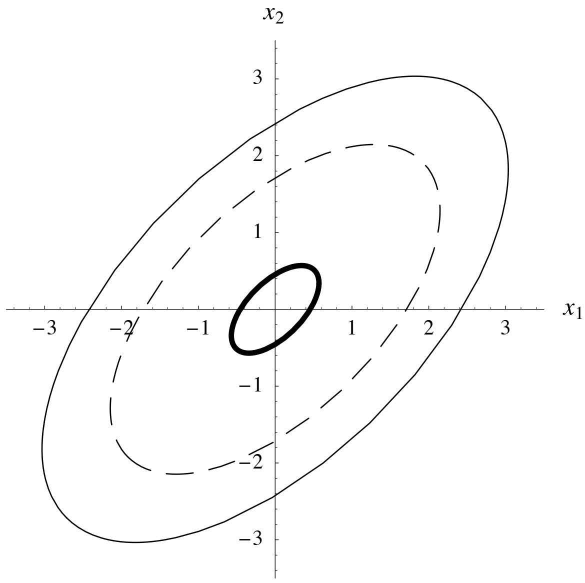


Fig. 13: Quantiles: 15% (bold), 90% (dashed) and 99% (plain)

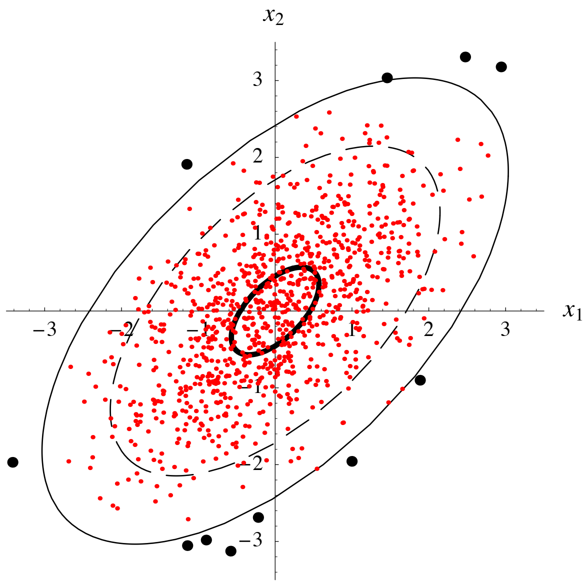


Fig. 14: Quantiles plotted with 1000 pseudo-random drawings

○ *Applying the* **mathStatica** *Toolset*

We can try out the **mathStatica** toolset on density f. The marginal distribution of X_1 is well known to be $N(0, 1)$, as we confirm with:

```
Marginal[x₁, f]
```

$$\frac{e^{-\frac{x_1^2}{2}}}{\sqrt{2\pi}}$$

The variance-covariance matrix is, of course, equal to Σ:

```
Varcov[f]
```

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

The conditional distribution of X_1 given $X_2 = x_2$ is $N(\rho\, x_2,\ 1 - \rho^2)$, as we confirm with:

```
Conditional[x₁, f]
```

— Here is the conditional pdf f (x1 | x2):

$$\frac{e^{\frac{(x_1-\rho\, x_2)^2}{2\,(-1+\rho^2)}}}{\sqrt{2\pi}\ \sqrt{1-\rho^2}}$$

Here is the product moment $E[X_1^2\ X_2^2]$:

```
Expect[x₁² x₂², f]
```

$$1 + 2\,\rho^2$$

The moment generating function is given by:

```
t⃗ = {t₁, t₂};  mgf = Expect[e^(t⃗.x⃗), f]
```

$$e^{\frac{1}{2}\,(t_1^2+2\,\rho\, t_1\, t_2+t_2^2)}$$

Here, again, is the product moment $E[X_1^2\ X_2^2]$, but now derived from the mgf:

```
D[mgf, {t₁, 2}, {t₂, 2}] /. t_ → 0
```

$$1 + 2\,\rho^2$$

If the mgf is known, this approach to deriving moments is much faster than the direct `Expect` approach. However, in higher variate (or more general) examples, *Mathematica* may not always be able to find the mgf, nor the cf. In the special case of the multivariate Normal distribution, this is not necessarily a problem since *Mathematica*'s Statistics package 'knows' the solution. Of course, this concept of 'knowledge' is somewhat artificial — *Mathematica*'s Statistics package does not derive the solution, but rather regurgitates the answer just like a textbook appendix does. In this vein, the Statistics package and a textbook appendix both work the same way: someone typed the answer in. For instance, for our example, the cf is immediately outputted (*not* derived) by the Statistics package as:

```
CharacteristicFunction[dist2, {t₁, t₂}]
```

$$e^{\frac{1}{2}(-t_2(\rho t_1+t_2)-t_1(t_1+\rho t_2))}$$

While this works well here, the regurgitation approach unfortunately breaks down as soon as one veers from the chosen path, as we shall see in *Example 21*.

⊕ ***Example 19:*** The Normal Linear Regression Model

Let us suppose that the random variables Y and X are jointly distributed, and that the conditional mean of Y given $X = x$ can be expressed as

$$E\left[Y \mid X = x\right] = \alpha_1 + \alpha_2 x \tag{6.29}$$

where α_1 and α_2 are unknown but fixed parameters. The conditional mean, being linear in the parameters, is called a *linear regression function*. We may write

$$Y = \alpha_1 + \alpha_2 x + U \tag{6.30}$$

where the random variable $U = Y - E\left[Y \mid X = x\right]$ is referred to as the *disturbance*, and has, by construction, a conditional mean equal to zero; that is, $E\left[U \mid X = x\right] = 0$. If Y is conditionally Normally distributed, then by linearity so too is U conditionally Normal, in which case we have the ***Normal*** *linear regression model*. This model can arise from a setting in which (Y, X) are jointly Normally distributed. To see this, let (Y, X) have joint bivariate pdf $N(\vec{\mu}, \Sigma)$ where:

```
μ⃗ = {μY, μX};     Σ = ( σY²         σY σX ρ )
                       ( σY σX ρ     σX²     );
cond = {σY > 0, σX > 0, -1 < ρ < 1};
dist = MultinormalDistribution[μ⃗, Σ];
```

Let $f(y, x)$ denote the joint pdf:

```
       f = Simplify[PDF[dist, {y, x}], cond]
domain[f] = {{y, -∞, ∞}, {x, -∞, ∞}} && cond
```

$$\frac{e^{\frac{(y-\mu_Y)^2\sigma_X^2-2\rho(x-\mu_X)(y-\mu_Y)\sigma_X\sigma_Y+(x-\mu_X)^2\sigma_Y^2}{2(-1+\rho^2)\sigma_X^2\sigma_Y^2}}}{2\pi\sqrt{1-\rho^2}\,\sigma_X\sigma_Y}$$

{{y, -∞, ∞}, {x, -∞, ∞}} && {$\sigma_Y > 0$, $\sigma_X > 0$, $-1 < \rho < 1$}

The regression function $E\left[Y \mid X=x\right]$ can be derived in two steps (as per *Example 9*):

(i) We first determine the conditional pdf of Y given $X = x$:

```
fcon = Conditional[y, f]
```

– Here is the conditional pdf f(y | x):

$$\frac{e^{\frac{((y-\mu_Y)\sigma_X+\rho(-x+\mu_X)\sigma_Y)^2}{2(-1+\rho^2)\sigma_X^2\sigma_Y^2}}}{\sqrt{2\pi}\sqrt{1-\rho^2}\,\sigma_Y}$$

where the domain of the conditional distribution is:

```
domain[fcon] = {y, -∞, ∞} && cond;
```

(ii) We can now find $E\left[Y \mid X=x\right]$:

```
regf = Expect[y, fcon]
```

$$\mu_Y + \frac{\rho\,(x-\mu_X)\,\sigma_Y}{\sigma_X}$$

This expression is of form $\alpha_1 + \alpha_2\, x$. To see this, we can use the `CoefficientList` function to obtain the parameters α_1 and α_2:

```
CoefficientList[regf, x]
```

$$\left\{\mu_Y - \frac{\rho\,\mu_X\,\sigma_Y}{\sigma_X},\ \frac{\rho\,\sigma_Y}{\sigma_X}\right\}$$

In summary, if (Y, X) are jointly bivariate Normal, then the regression function $E\left[Y \mid X=x\right]$ is linear in the parameters, of form $\alpha_1 + \alpha_2\, x$, where $\alpha_1 = \mu_Y - \alpha_2\,\mu_X$ and $\alpha_2 = \frac{\rho\,\sigma_Y}{\sigma_X}$, which is what we set out to show. Finally, inspection of `fcon` reveals that the conditional distribution of $Y \mid (X = x)$ is Normal. Joint Normality therefore determines a Normal linear regression model. ■

⊕ ***Example 20:*** Robin Hood

Robin Hood has entered the coveted Nottingham Forest Archery competition, where contestants shoot arrows at a vertical target. For Mr Hood, it is known that the distribution of horizontal and vertical deviations from the centre of the target is bivariate Normal, with zero means, equal variances σ^2 and correlation ρ. What is the probability that he gets a bull's-eye, if the latter has unit radius?

Solution: We begin by setting up the appropriate bivariate Normal distribution:

```
X⃗ = {x1, x2};   μ⃗ = {0, 0};   Σ = σ² (1 ρ
                                      ρ 1);
dist = MultinormalDistribution[μ⃗, Σ];
cond = {σ > 0, -1 < ρ < 1, r > 0, 0 < θ < 2 π};
```

Let $f(x_1, x_2)$ denote the joint pdf:

```
         f  =  Simplify[PDF[dist, x⃗], cond]
domain[f] = {{x₁, -∞, ∞}, {x₂, -∞, ∞}} && cond;
```

$$\frac{e^{\frac{x_1^2-2\rho x_1 x_2+x_2^2}{2(-1+\rho^2)\sigma^2}}}{2\pi\sqrt{1-\rho^2}\ \sigma^2}$$

The solution requires a transformation to polar co-ordinates. Thus:

```
Ω = {x₁ → r Cos[θ], x₂ → r Sin[θ]};
```

Here, $R = \sqrt{X_1^2 + X_2^2}$ represents the distance of (X_1, X_2) from the origin, while $\Theta = \arctan(X_2 / X_1)$ represents the angle of (X_1, X_2) with respect to the X_1 axis. Thus, $R = r \in \mathbb{R}_+$ and $\Theta = \theta \in \{\theta : 0 < \theta < 2\pi\}$. We seek the joint pdf of R and Θ. We thus apply the transformation method (Chapter 4). We do so manually (see §4.2 C), because there are two solutions, differing only in respect to sign. The desired joint density is $g(r, \theta)$:

```
         g  = Simplify[(f /. Ω) Jacob[x⃗ /. Ω, {r, θ}], cond]
domain[g]  = {{r, 0, ∞},  {θ, 0, 2π}} && cond;
```

$$\frac{e^{-\frac{r^2(-1+\rho\,\mathrm{Sin}[2\theta])}{2(-1+\rho^2)\sigma^2}}\ r}{2\pi\sqrt{1-\rho^2}\ \sigma^2}$$

The probability of hitting the bull's-eye is given by $P(R \le 1)$. In the simple case of zero correlation ($\rho = 0$), this is:

```
pr  = Prob[{1, 2π},  g /. ρ → 0]
```

$$1 - e^{-\frac{1}{2\sigma^2}}$$

As expected, this probability is decreasing in the standard deviation σ, as Fig. 15 illustrates.

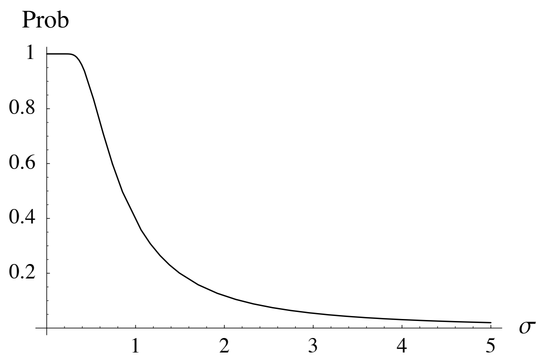


Fig. 15: Probability that Robin Hood hits a bull's-eye, as a function of σ

More generally, in the case of non-zero correlation ($\rho \neq 0$), *Mathematica* cannot determine this probability exactly. This is not surprising as the solution does not have a convenient closed form. Nevertheless, given values of the parameters σ and ρ, one can use numerical integration. For instance, if $\sigma = 2$, and $\rho = 0.7$, the probability of a bull's-eye is:

```
NIntegrate[g /. {σ → 2, ρ → 0.7}, {r, 0, 1}, {θ, 0, 2 π}]
```

```
0.155593
```

which contrasts with a probability of 0.117503 when $\rho = 0$. More generally, it appears that a contestant whose shooting is 'elliptical' ($\rho \neq 0$) will hit the bull's-eye more often than an 'uncorrelated' ($\rho = 0$) contestant! ■

⊕ ***Example 21:*** Truncated Bivariate Normal

Let $(X, Y) \sim N(\vec{0}, \Sigma)$ with joint pdf $f(x, y)$ and cdf $F(x, y)$, with $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where we shall assume that $0 < \rho < 1$. Corresponding to $f(x, y)$, let $g(x, y)$ denote the pdf of a truncated distribution with Y restricted to the positive real line ($Y > 0$). We wish to find the pdf of the truncated distribution $g(x, y)$, the marginal distributions $g_X(x)$ and $g_Y(y)$, and the new variance-covariance matrix.

Solution: Since the truncated distribution is not a 'textbook' Normal distribution, *Mathematica*'s `Multinormal` package is not designed to answer such questions. By contrast, **mathStatica** adopts a general approach and so can solve such problems. Given:

```
V = {x, y};   μ⃗ = {0, 0};   Σ = (1 ρ; ρ 1);   cond = {0 < ρ < 1};
```

Then, the parent pdf $f(x, y)$ is:

```
f = Simplify[PDF[MultinormalDistribution[μ⃗, Σ], V ], cond];
domain[f] = {{x, -∞, ∞}, {y, -∞, ∞}} && cond;
```

By familiar truncation arguments (§2.5 A):

$$g(x, y) \;=\; \frac{f(x, y)}{1 - F(\infty, 0)} \;=\; 2\, f(x, y), \qquad \text{for } x \in \mathbb{R},\, y \in \mathbb{R}_+$$

which we enter as:

```
      g  = 2 f;
domain[g] = {{x, -∞, ∞}, {y, 0, ∞}} && cond;
```

The marginal pdf of Y, when Y is truncated below at zero, is $g_Y(y)$:

```
gY = Marginal[y, g]
```

$$e^{-\frac{y^2}{2}} \sqrt{\frac{2}{\pi}}$$

This is the pdf of a half-Normal random variable, as illustrated in Fig. 16.

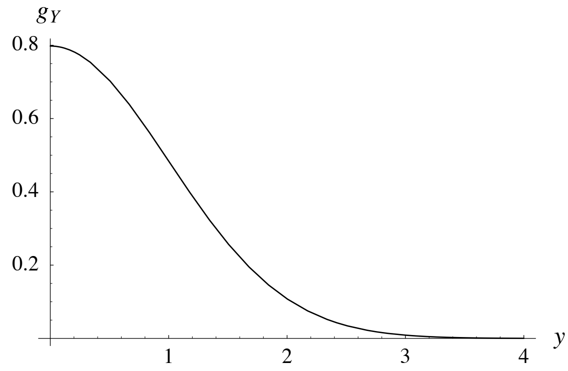


Fig. 16: The marginal pdf of Y, when Y is truncated below at zero

By contrast, the marginal pdf of X, when Y is truncated below at 0, is given by $g_X(x)$:

gX = Marginal[x, g]

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

which is Azzalini's skew-Normal(λ) pdf with $\lambda = \rho \big/ \sqrt{1-\rho^2}$ (see Chapter 2, Exercise 2). Even though X is not itself truncated, $g_X(x)$ *is* affected by the truncation of Y, because X is correlated with Y. Now consider the two extremes: if $\rho = 0$, X and Y are uncorrelated, so $g_X(\cdot) = f_X(\cdot)$, and we obtain a standard Normal pdf; at the other extreme, if $\rho = 1$, X and Y are perfectly correlated, so $g_X(\cdot) = g_Y(\cdot)$, and we obtain a half-Normal pdf. For $0 < \rho < 1$, we obtain a result between these two extremes. This can be seen from Fig. 17, which plots both extremes, and three cases in between.

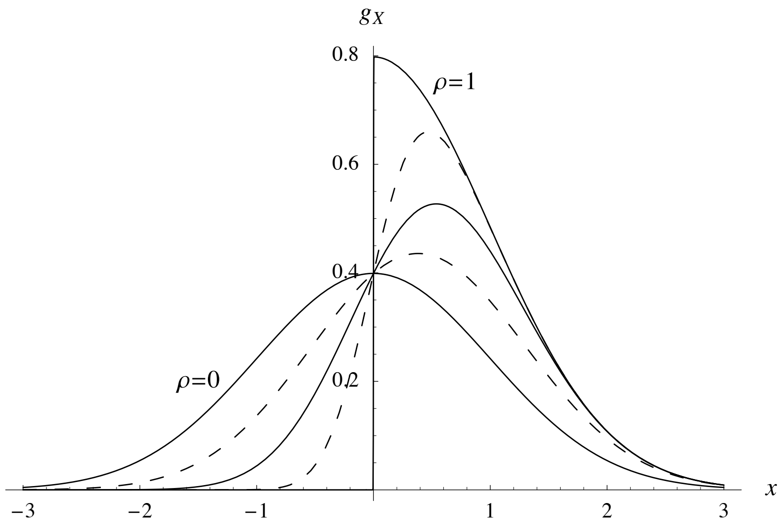


Fig. 17: The marginal pdf of X, when Y is truncated below at zero.

The mean vector, when Y is truncated below at zero, is:

```
Expect[{x, y}, g]
```

$$\left\{\sqrt{\frac{2}{\pi}}\,\rho,\ \sqrt{\frac{2}{\pi}}\right\}$$

The variance-covariance matrix for (X, Y), when Y is truncated below at zero, is:

```
Varcov[g]
```

$$\begin{pmatrix} 1-\frac{2\rho^2}{\pi} & \frac{(-2+\pi)\,\rho}{\pi} \\ \frac{(-2+\pi)\,\rho}{\pi} & \frac{-2+\pi}{\pi} \end{pmatrix}$$

This illustrates that, in a mutually dependent setting, the truncation of one random variable effects all the random variables (not just the truncated variable). ■

6.4 B The Trivariate Normal

The trivariate Normal distribution for (X, Y, Z) is fully specified by the (3×1) vector of means and the (3×3) variance-covariance matrix. When the mean vector is $\vec{0}$ and the variances are all equal to unity, we have:

$$\texttt{V = \{x, y, z\};}\quad \vec{\mu} = \{0, 0, 0\};\quad \Sigma = \begin{pmatrix} 1 & \rho_{xy} & \rho_{xz} \\ \rho_{xy} & 1 & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & 1 \end{pmatrix};$$

```
dist3 = MultinormalDistribution[μ⃗, Σ];

cond = {-1 < ρxy < 1, -1 < ρxz < 1, -1 < ρyz < 1, Det[Σ] > 0};
```

where ρ_{ij} denotes the correlation between variable i and variable j, and the condition `Det[Σ] > 0` reflects the fact that the variance-covariance matrix is positive definite. Let $g(x, y, z)$ denote the joint pdf:

```
g = PDF[dist3, V] // Simplify
```

$$\frac{e^{\frac{x^2+y^2+z^2-z^2\rho_{xy}^2-y^2\rho_{xz}^2-2yz\rho_{yz}-x^2\rho_{yz}^2-2x\rho_{xz}(z-y\rho_{yz})+2\rho_{xy}(-xy+yz\rho_{xz}+xz\rho_{yz})}{2(-1+\rho_{xy}^2+\rho_{xz}^2-2\rho_{xy}\rho_{xz}\rho_{yz}+\rho_{yz}^2)}}}{2\sqrt{2}\,\pi^{3/2}\sqrt{1-\rho_{xy}^2-\rho_{xz}^2+2\rho_{xy}\rho_{xz}\rho_{yz}-\rho_{yz}^2}}$$

with domain:

```
domain[g] = Thread[{V, -∞, ∞}] && cond
```

$$\{\{x,-\infty,\infty\},\{y,-\infty,\infty\},\{z,-\infty,\infty\}\}\ \&\&\ \{-1<\rho_{xy}<1,\ -1<\rho_{xz}<1,\ -1<\rho_{yz}<1,\ 1-\rho_{xy}^2-\rho_{xz}^2+2\rho_{xy}\rho_{xz}\rho_{yz}-\rho_{yz}^2>0\}$$

Here, for example, is $E[X\,Y\,e^Z]$; the calculation takes about 70 seconds on our reference computer:

```
Expect[x y ⅇ^z, g]
```

$$\sqrt{e}\;(\rho_{xy} + \rho_{xz}\,\rho_{yz})$$

Figure 12, above, illustrated that a contour plot of a bivariate Normal pdf yields an ellipse, or a circle given zero correlation. Figure 18 illustrates a specific contour of the trivariate pdf $g(x, y, z)$, when $\rho_{xy} \to 0.2$, $\rho_{yz} \to 0.3$, $\rho_{xz} \to 0.4$, and $g(x, y, z) = 0.05$. Once again, the symmetry of the plot will be altered by the choice of correlation coefficients. Whereas the bivariate Normal yields elliptical contours (or a circle given zero correlation), the trivariate case yields the intuitive 3D equivalent, namely the surface of an ellipsoid (or that of a sphere given zero correlations). Here, parameter ρ_{xy} alters the 'orientation' of the ellipsoid in the x-y plane, just as ρ_{yz} does in the y-z plane, and ρ_{xz} does in the x-z plane.

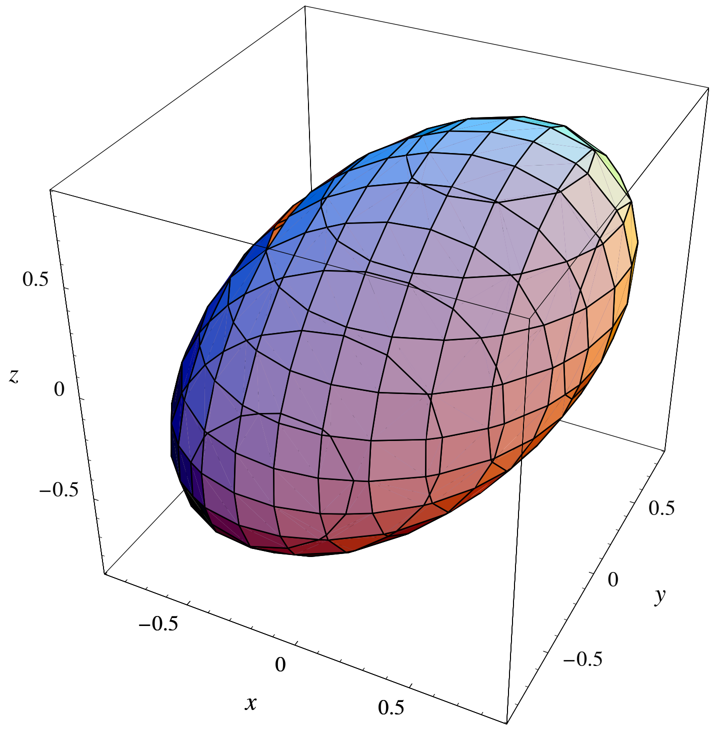


Fig. 18: The contour $g(x, y, z) = 0.05$ for the trivariate Normal pdf

Just as in the 2D case, we can plot the specific ellipsoid that encloses q% of the distribution by using the function `EllipsoidQuantile[`*dist*, *q*`]`. This is illustrated in Fig. 19 below, which plots the ellipsoids that enclose 60% (solid) and 90% (wireframe) of the distribution, respectively, given $\rho_{xy} \to 0.01$, $\rho_{yz} \to 0.01$, $\rho_{xz} \to 0.4$. Ideally, one would plot the 90% ellipsoid using translucent graphics. Unfortunately, *Mathematica* Version 4 does not support translucent graphics, so we use a `WireFrame` instead.

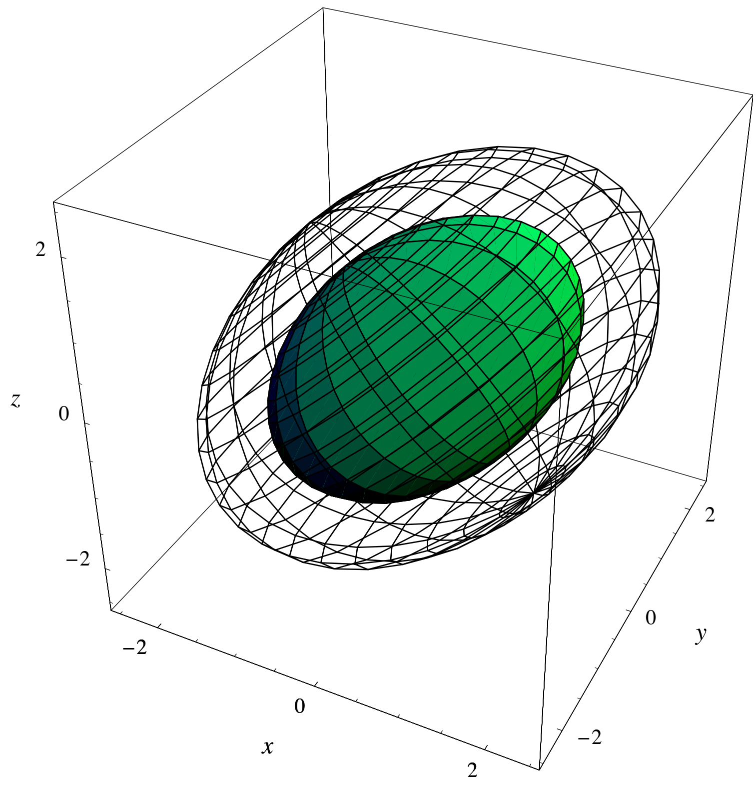


Fig. 19: Quantiles: 60% (solid) and 90% (wireframe)

⊕ ***Example 22:*** Correlation and Positive Definite Matrix

Let X, Y and Z follow a standardised trivariate Normal distribution. It is known that $\rho_{xy} = 0.9$ and $\rho_{xz} = -0.8$, but ρ_{yz} is not known. What can we say, if anything, about the correlation ρ_{yz}?

Solution: Although there is not enough information to uniquely determine the value of ρ_{yz}, there *is* enough information to specify a range of values for it (of course, $-1 < \rho_{yz} < 1$ must always hold). This is achieved by using the property that Σ must be a positive definite matrix, which implies that the determinant of Σ must be positive:

```
dd = Det[Σ] /. {ρxy → .9, ρxz → -.8}
```

$$-0.45 - 1.44\,\rho_{yz} - \rho_{yz}^2$$

This expression is positive when ρ_{yz} lies in the following interval:

```
<< Algebra`
InequalitySolve[dd > 0, ρyz]
```

$$-0.981534 < \rho_{yz} < -0.458466$$

6.4 C CDF, Probability Calculations and Numerics

While it is generally straightforward to find numerical values for any multivariate Normal pdf, it is not quite as easy to do so for the cdf. To illustrate, we use the trivariate Normal pdf $g(x, y, z)$ = `PDF[dist3, {x, y, z}]` defined at the start of §6.4 B. We distinguish between two possible scenarios: (i) zero correlation, and (ii) non-zero correlation.

○ ***Zero Correlation***

Under zero correlation, it is possible to find an exact *symbolic* solution using **mathStatica** in the usual way.[3,4] Let $G(x, y, z)$ denote the cdf $P(X \leq x, Y \leq y, Z \leq z)$ under zero correlation:

```
Clear[G];    G[x_, y_, z_] = Prob[{x, y, z}, g /. ρ_ → 0]
```

$$\frac{1}{8}\left(1+\text{Erf}\left[\frac{x}{\sqrt{2}}\right]\right)\left(1+\text{Erf}\left[\frac{y}{\sqrt{2}}\right]\right)\left(1+\text{Erf}\left[\frac{z}{\sqrt{2}}\right]\right)$$

This solution is virtuous in two respects: first, it is an exact symbolic expression; second, because the solution is already 'evaluated', it will be computationally efficient in application. Here, for instance, is the exact symbolic solution to $P(X \leq -2, Y \leq 0, Z \leq 2)$:

```
G[-2, 0, 2]
```

$$\frac{1}{8}\left(1-\text{Erf}\left[\sqrt{2}\right]\right)\left(1+\text{Erf}\left[\sqrt{2}\right]\right)$$

Because the solution is an exact symbolic expression, we can use *Mathematica*'s arbitrary precision numerical engine to express it as a numerical expression, to any desired number of digits of precision. Here is `G[-2,0,2]` calculated to 40 digits of precision:

```
N[G[-2, 0, 2] , 40]
```

```
0.01111628172225982147533684086722435761304
```

If we require the probability content of a region within the domain, we could just type in the whole integral. For instance, the probability of being within the region

$$S = \{(x, y, z): \quad 1 < x < 2, \quad 3 < y < 4, \quad 5 < z < 6\}$$

is given by:

```
Integrate[g /. ρ_ → 0,
    {x, 1, 2}, {y, 3, 4}, {z, 5, 6}] // N // Timing
```

$\{0.27\ \text{Second},\ 5.1178 \times 10^{-11}\}$

Alternatively, we can use the **mathStatica** function `MrSpeedy` (*Example 4*). `MrSpeedy` finds the probability content of a region within the domain just by using the known cdf `G[]` (which we have already found) and the boundaries of the region, without any need for further integration:

```
S = {{1, 2}, {3, 4}, {5, 6}};  MrSpeedy[G, S] // N // Timing
```

$\{0.\ \text{Second},\ 5.1178\times 10^{-11}\}$

MrSpeedy often provides enormous speed increases over direct integration.

○ *Non-Zero Correlation*

In the case of non-zero correlation, a closed form solution to the cdf does not generally exist, so that numerical integration is required. Even if we use the CDF function in *Mathematica*'s Multinormal statistics package, ultimately, in the background, we are still resorting to numerical integration. This, in turn, raises the two interrelated motifs of accuracy and computational efficiency, which run throughout this section.

Consider, again, the trivariate Normal pdf $g(x, y, z)$ = PDF[dist3, $\{x, y, z\}$] defined in §6.4 B. If $\rho_{xy} = \rho_{xz} = \rho_{yz} = \frac{1}{2}$, the cdf is:

```
Clear[G];   G[var__] := CDF[dist3 /. ρ_ → 1/2, {var}]
```

Hence, $P(X \leq 1, Y \leq -7, Z \leq 3)$ evaluates to:[5]

```
G[1, -7, 3]
```

1.27981×10^{-12}

If we require the probability content of a region within the domain, we can again use MrSpeedy. The probability of being within the region

$$S = \{(x, y, z):\ \ 1 < x < \infty,\quad -3 < y < 4,\quad 5 < z < 6\}$$

is then given by:

```
S = {{1, ∞}, {-3, 4}, {5, 6}};  MrSpeedy[G, S] // Timing
```

$\{0.55\ \text{Second},\ 2.61015\times 10^{-7}\}$

This is a significant improvement over using numerical integration directly, since the latter is both less accurate (at default settings) and *far* more resource hungry:

```
NIntegrate[ g /. ρ_ → 1/2,
    {x, 1, ∞}, {y, -3, 4}, {z, 5, 6}] // Timing
```

```
— NIntegrate::slwcon :
   Numerical integration converging too slowly; suspect one
     of the following: singularity, value of the integration
     being 0, oscillatory integrand, or insufficient
     WorkingPrecision. If your integrand is oscillatory
     try using the option Method->Oscillatory in NIntegrate.
```

$\{77.39\ \text{Second},\ 2.61013\times 10^{-7}\}$

The direct numerical integration approach can be 'sped up' by sacrificing some accuracy. This can be done by altering the PrecisionGoal option; see Rose and Smith (1996a or 1996b). This can be useful when working with a distribution whose cdf is not known (or cannot be derived), such that one has no alternative but to use direct numerical integration.

Finally, it is worth stressing that since the CDF function in *Mathematica*'s Multinormal statistics package is using numerical integration in the background, the numerical answer that is printed on screen is not exact. Rather, the answer will be correct to several decimal places, and incorrect beyond that; only symbolic entities are exact. To assess the accuracy of the CDF function, we can compare the answer it gives with symbolic solutions that are known for special cases. For example, Stuart and Ord (1994, Section 15.10) report symbolic solutions for the standardised bivariate Normal orthant probability $P(X \le 0,\ Y \le 0)$ as:

$$\textbf{P2} = \frac{1}{4} + \frac{\textbf{ArcSin}[\rho]}{2\,\pi};$$

while the standardised trivariate Normal orthant probability $P(X \le 0,\ Y \le 0,\ Z \le 0)$ is:

$$\textbf{P3} = \frac{1}{8} + \frac{1}{4\,\pi}\,(\textbf{ArcSin}[\rho_{xy}] + \textbf{ArcSin}[\rho_{xz}] + \textbf{ArcSin}[\rho_{yz}]);$$

We choose some values for ρ_{xy}, ρ_{xz}, ρ_{yz}:

$$\textbf{lis} = \left\{\rho_{xy} \to \frac{1}{17},\ \rho_{xz} \to \frac{1}{12},\ \rho_{yz} \to \frac{2}{5}\right\};$$

Because P3 is a symbolic entity, we can express it numerically to any desired precision. Here is the correct answer to 30 digits of precision:

```
N[P3 /. lis, 30]
```

```
0.169070356956715121611195785538
```

By contrast, the CDF function yields:

```
CDF[dist3 /. lis, {0, 0, 0}] // InputForm
```

```
0.1690703504574683
```

In this instance, the CDF function has only 8 digits of precision. In other cases, it may offer 12 digits of precision. Even so, 8 digits of precision is better than most competing packages. For more detail on numerical precision in *Mathematica*, see Appendix A.1.

In summary, *Mathematica*'s CDF function and **mathStatica**'s MrSpeedy function make an excellent team; together, they are more accurate and faster than using numerical integration directly. How then does *Mathematica* compare with highly specialised multivariate Normal computer programs (see Schervish (1984)) such as Bohrer–Schervish, MULNOR, and MVNORM? For zero-correlation, *Mathematica* can easily outperform such programs in both accuracy and speed, due to its symbolic engine. For non-zero correlation, *Mathematica* performs well on accuracy grounds.

6.4 D Random Number Generation for the Multivariate Normal

○ ***Introducing*** `MVNRandom`

The **mathStatica** function `MVNRandom[`n, $\vec{\mu}$, Σ`]` generates n pseudo-random m-dimensional drawings from the multivariate Normal distribution with mean vector $\vec{\mu}$, and $(m \times m)$ variance-covariance matrix Σ; the function assumes dimension m is an integer larger than 1. Once again, Σ is required to be symmetric and positive definite. The function has been optimised for speed. To demonstrate its application, we generate 6 drawings from a trivariate Normal with mean vector and variance-covariance matrix given by:

$$\vec{\mu} = \{10, 0, -20\}; \quad \Sigma = \begin{pmatrix} 1 & 0.2 & 0.4 \\ 0.2 & 2 & 0.3 \\ 0.4 & 0.3 & 3 \end{pmatrix}; \quad \text{MVNRandom}[6, \vec{\mu}, \Sigma]$$

$$\begin{pmatrix} 10.1802 & 0.792264 & -20.7549 \\ 9.61446 & 0.936577 & -20.3007 \\ 9.00878 & 1.51215 & -17.9076 \\ 10.0042 & -0.749123 & -23.6165 \\ 12.2513 & -1.28886 & -19.8166 \\ 10.7216 & -0.626802 & -15.847 \end{pmatrix}$$

The output from `MVNRandom` is a set of n lists (here $n = 6$). Each list represents a single pseudo-random drawing from the distribution and so has the dimension of the random variable ($m = 3$). In this way, `MVNRandom` has recorded 6 pseudo-random drawings from the 3-dimensional $N(\vec{\mu}, \Sigma)$ distribution.

Instead of using **mathStatica**'s `MVNRandom` function, one can alternatively use the `RandomArray` function in *Mathematica*'s `Multinormal` Statistics package. To demonstrate, we generate 20000 drawings using both approaches:

```
MVNRandom[20000, μ⃗, Σ]; // Timing
```

```
{0.22 Second, Null}
```

```
RandomArray[
   MultinormalDistribution[μ⃗, Σ], 20000] ; // Timing
```

```
{2.53 Second, Null}
```

In addition to its obvious efficiency, `MVNRandom` has other advantages. For instance, it advises the user if the variance-covariance matrix is not symmetric and/or if it is not positive definite.

○ ***How*** `MVNRandom` ***Works***

`MVNRandom` works by transforming a pseudo-random drawing from an m-dimensional $N(\vec{0}, I_m)$ distribution into a $N(\vec{\mu}, \Sigma)$ drawing: the transformation is essentially the multivariate equivalent of a location shift plus a scale change. The transformation relies upon the spectral decomposition (using `Eigensystem`) of the variance-covariance matrix; that is, the decomposition of $\Sigma = H D H^{\mathrm{T}}$ into its spectral components H and D. The columns of the $(m \times m)$ matrix H are the eigenvectors of Σ, and the $(m \times m)$ diagonal matrix D contains the eigenvalues of Σ. Then, for a random vector $\vec{Y} \sim N(\vec{0}, I_m)$, a linear transformation from $\vec{Y}$ to a new random vector $\vec{X}$, according to the rule

$$\vec{X} = \vec{\mu} + H D^{1/2} \vec{Y} \tag{6.31}$$

finds $\vec{X} \sim N(\vec{\mu}, \Sigma)$. By examining the mean vector and variance-covariance matrix, it is easy to see why this transformation works:

$$E[\vec{X}] = E\left[\vec{\mu} + H D^{1/2} \vec{Y}\right] = \vec{\mu}, \quad \text{because } E[\vec{Y}] = \vec{0}$$

and

$$\begin{aligned} \text{Varcov}(\vec{X}) &= \text{Varcov}\left(\vec{\mu} + H D^{1/2} \vec{Y}\right) \\ &= \text{Varcov}\left(H D^{1/2} \vec{Y}\right) \\ &= H D^{1/2} \, \text{Varcov}(\vec{Y}) \, D^{1/2} H^{\mathrm{T}} \\ &= H D H^{\mathrm{T}} \\ &= \Sigma \end{aligned}$$

because $\text{Varcov}(\vec{Y}) = I_m$. We wish to sample the distribution of $\vec{X}$, which requires that we generate a pseudo-random drawing of $\vec{Y}$ and apply (6.31) to it. So, all that remains is to do the very first step — generate $\vec{Y}$ — but that is the easiest bit! Since the components of $\vec{Y}$ are independent, it suffices to combine together m pseudo-random drawings from the univariate standard Normal distribution $N(0, 1)$ into a single column.

○ *Visualising Random Data in 2D and 3D Space*

With *Mathematica*, we can easily visualise random data that has been generated in two or three dimensions. We will use the functions `D2` and `D3` to plot the data in two-dimensional and three-dimensional space, respectively:

```
D2[x_] := ListPlot[x, PlotStyle → Hue[1],
   AspectRatio → 1, DisplayFunction → Identity];

D3[x_] :=
 Graphics3D[ {Hue[1], Map[Point, x]}, Axes → True]
```

Not only can we plot the data in its appropriate space, but we can also view the data projected onto a hypersphere; for example, two-dimensional data can be projected onto a circle, while three-dimensional data can be projected onto a sphere. This is achieved by normalising the data by using the `norm` function defined below. Finally, the function `MVNPlot` provides a neat way of generating our desired diagrams:

```
norm[x_] := Map[ # / √(#.#) &, x];

MVNPlot[DD_, w_] := Show[GraphicsArray[
    {DD[w], DD[norm[w]]}, GraphicsSpacing → .3]];
```

The Two-Dimensional Case

(i) *Zero correlation:* Fig. 20 shows two plots: the left panel illustrates the generated data in two-dimensional space; the right panel projects this data onto the unit circle. A random vector $\vec{X}$ is said to be *spherically distributed* if its pdf is equivalent to that of $\vec{Y} = H\,\vec{X}$, for all orthogonal matrices H. Spherically distributed random variables have the property that they are uniformly distributed on the unit circle / sphere / hypersphere. The zero correlation bivariate Normal is a member of the spherical class.[6] This explains why the generated data appears uniform on the circle.

$$\vec{\mu} = \{0,\ 0\}; \qquad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \qquad \texttt{w = MVNRandom[1500, } \vec{\mu}\texttt{, } \Sigma\texttt{];}$$

```
MVNPlot[D2, w];
```

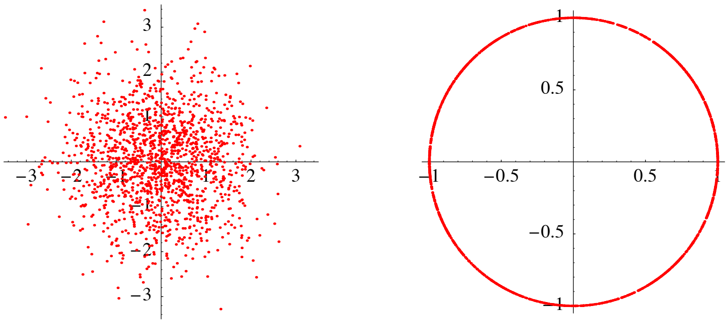


Fig. 20: Zero correlation bivariate Normal: random data

(ii) *Non-zero correlation:* Fig. 21 again shows two plots, but now in the case of non-zero correlation. The left panel shows that the data has high positive correlation. The right panel shows that the distribution is no longer uniform on the unit circle, for there are relatively few points projected onto it in the north-west and south-east quadrants. This is because the correlated bivariate Normal does not belong to the spherical class; instead, it belongs to the elliptical class of distributions. For further details on elliptical distributions, see Muirhead (1982).

$$\vec{\mu} = \{0,\ 0\}; \qquad \Sigma = \begin{pmatrix} 1 & .95 \\ .95 & 1 \end{pmatrix}; \qquad \texttt{w = MVNRandom[1500, } \vec{\mu}\texttt{, } \Sigma\texttt{];}$$

```
MVNPlot[D2, w];
```

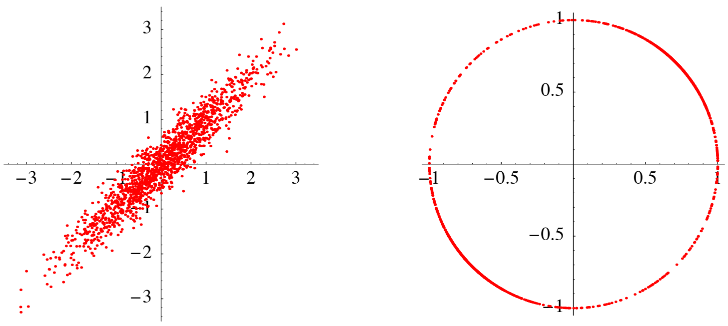


Fig. 21: Correlated bivariate Normal: random data

The Three-Dimensional Case

(i) *Zero correlation:* Fig. 22 again shows two plots. The left panel illustrates the generated data in three-dimensional space. The right panel projects this data onto the unit sphere. The distribution appears uniform on the sphere, as indeed it should, because this particular trivariate Normal is a member of the spherical class.

$$\vec{\mu} = \{0, 0, 0\}; \quad \Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \text{w = MVNRandom[2000, } \vec{\mu}, \Sigma];$$

```
MVNPlot[D3, w];
```

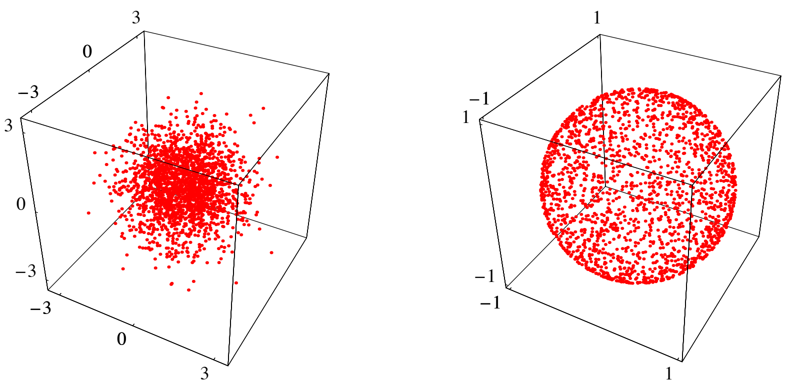


Fig. 22: Zero correlation trivariate Normal: random data

(ii) *Non-zero correlation:* see Fig. 23 below. The three-dimensional plot on the left illustrates that the data is now highly correlated, while the projection onto the unit sphere (on the right) provides ample evidence that this particular trivariate Normal distribution is no longer spherical.

$$\vec{\mu} = \{0, 0, 0\}; \quad \Sigma = \begin{pmatrix} 1 & .95 & .95 \\ .95 & 1 & .95 \\ .95 & .95 & 1 \end{pmatrix}; \quad \text{w = MVNRandom[2000, } \vec{\mu}, \Sigma];$$

```
MVNPlot[D3, w];
```

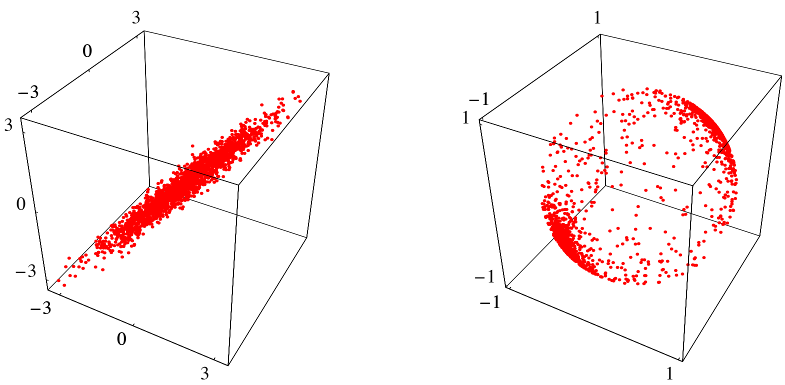


Fig. 23: Correlated trivariate Normal: random data

6.5 The Multivariate *t* and Multivariate Cauchy

Let $(X_1, \ldots, X_m)$ have a joint *standardised* multivariate Normal distribution with correlation matrix R, and let $Y \sim$ Chi-squared(v) be independent of $(X_1, \ldots, X_m)$. Then the joint pdf of

$$T_j = \frac{X_j}{\sqrt{Y/v}}, \qquad (j = 1, \ldots, m) \tag{6.32}$$

defines the multivariate t distribution with v degrees of freedom and correlation matrix R, denoted $t(\mathrm{R}, v)$. The multivariate Cauchy distribution is obtained when $\mathrm{R} = I_m$ and $v = 1$. The multivariate t is included in *Mathematica*'s `Multinormal` Statistics package, so our discussion here will be brief. First, we ensure the appropriate package is loaded:

```
<< Statistics`
```

Let random variables W_1 and W_2 have joint pdf $t(\mathrm{R}, v)$ where $\mathrm{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, and ρ denotes the correlation coefficient between W_1 and W_2. So:

```
W⃗ = {w₁, w₂};   R = (1 ρ
                     ρ 1) ; cond = {-1 < ρ < 1, v > 0};

dist2 = MultivariateTDistribution[R, v];
```

Then our bivariate t pdf $f(w_1, w_2)$ is given by:

```
f  = FullSimplify[PDF[dist2, W⃗], cond]
```

$$\frac{v^{\frac{2+v}{2}}\ (1-\rho^2)^{\frac{1+v}{2}}\ (v - v\,\rho^2 + w_1^2 - 2\,\rho\,w_1\,w_2 + w_2^2)^{-1-\frac{v}{2}}}{2\,\pi}$$

with domain of support:

```
domain[f] = Thread[{W⃗, -∞, ∞}] && cond

{{w₁, -∞, ∞}, {w₂, -∞, ∞}} && {-1 < ρ < 1, v > 0}
```

Example 23 below derives this pdf from first principles. The shape of the contours of $f(w_1, w_2)$ depend on ρ. We can plot the specific ellipse that encloses $q\%$ of the distribution by using the function `EllipsoidQuantile[`*dist*, *q*`]`. This is illustrated in Fig. 24 which plots the ellipses that enclose 15% (bold), 90% (dashed) and 99% (plain) of the distribution, respectively, with $\rho = 0.4$ and $v = 2$ degrees of freedom. The long-tailed nature of the t distribution is apparent, especially when this diagram is compared with Fig. 13.

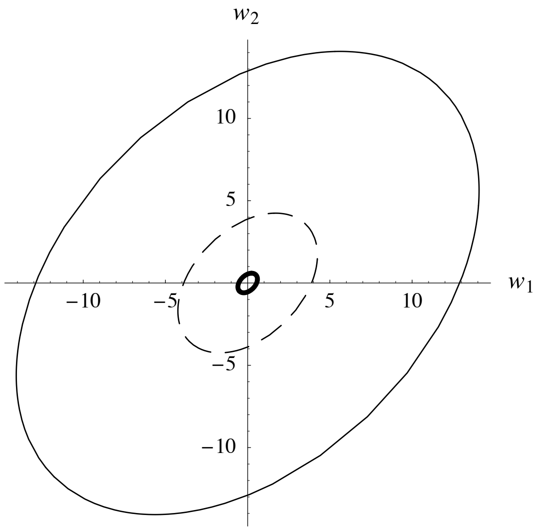


Fig. 24: Quantiles: 15% (bold), 90% (dashed) and 99% (plain)

The bivariate Cauchy distribution is obtained when $\mathrm{R} = I_2$ and $v = 1$:

```
f /. {ρ → 0, v → 1}
```

$$\frac{1}{2\,\pi\,(1 + w_1^2 + w_2^2)^{3/2}}$$

Under these conditions, the marginal distribution of W_1 is the familiar (univariate) Cauchy distribution:

```
Marginal[w₁, f /. {ρ → 0, v → 1}]
```

$$\frac{1}{\pi + \pi\,w_1^2}$$

As in §6.4 C, one can use functions like `MrSpeedy` in conjunction with *Mathematica*'s `CDF` function to find probabilities, and `RandomArray` to generate pseudo-random drawings.

⊕ ***Example 23:*** Deriving the pdf of the Bivariate t

Find the joint pdf of:

$$T_j = \frac{X_j}{\sqrt{Y/v}}, \qquad (j = 1, 2)$$

from first principles, where (X_1, X_2) have a joint *standardised* multivariate Normal distribution, and $Y \sim$ Chi-squared(v) is independent of (X_1, X_2).

Solution: Due to independence, the joint pdf of (X_1, X_2, Y), say $\varphi(x_1, x_2, y)$, is just the pdf of (X_1, X_2) multiplied by the pdf of Y:

```
φ = ( e^((x1^2 - 2 ρ x1 x2 + x2^2)/(-2 + 2 ρ^2)) / (2 π Sqrt[1 - ρ^2]) ) * ( e^(-y/2) y^(v/2 - 1) / (2^(v/2) Γ[v/2]) );
cond = {v > 0, -1 < ρ < 1};
domain[φ] = {{x1, -∞, ∞}, {x2, -∞, ∞}, {y, 0, ∞}} && cond;
```

$$\varphi = \left(\frac{e^{\frac{x_1^2-2\rho x_1 x_2+x_2^2}{-2+2\rho^2}}}{2\pi\sqrt{1-\rho^2}}\right) * \left(\frac{e^{-\frac{y}{2}}\, y^{\frac{v}{2}-1}}{2^{v/2}\,\Gamma[\frac{v}{2}]}\right);$$

Let $U = Y$. Then, using **mathStatica**'s `Transform` function, the joint pdf of (T_1, T_2, U) is:

```
f = Transform[{t1 == x1/Sqrt[y/v], t2 == x2/Sqrt[y/v], u == y}, φ]
```

$$\frac{2^{-1-\frac{v}{2}}\, e^{\frac{u\,(v - v\rho^2 + t_1^2 - 2\rho\, t_1\, t_2 + t_2^2)}{2\,v\,(-1+\rho^2)}}\, u^{v/2}}{\pi\, v\,\sqrt{1-\rho^2}\;\Gamma[\frac{v}{2}]}$$

with domain:

```
domain[f] = {{t1, -∞, ∞}, {t2, -∞, ∞}, {u, 0, ∞}} && cond;
```

Then, the marginal joint pdf of random variables T_1 and T_2 is:

```
Marginal[{t1, t2}, f]
```

$$\frac{v\,\sqrt{1-\rho^2}\,\left(\frac{v - v\rho^2 + t_1^2 - 2\rho\, t_1\, t_2 + t_2^2}{v - v\rho^2}\right)^{-1-\frac{v}{2}}}{2\,\pi\,(v - v\,\rho^2)}$$

which is the desired pdf. Note that this output is identical to the answer given to `PDF[dist2, {t1, t2}] // FullSimplify`. ■

6.6 Multinomial and Bivariate Poisson

This section discusses two discrete multivariate distributions, namely the Multinomial and the bivariate Poisson. Both of these distributions are also discussed in *Mathematica*'s `Statistics`MultiDiscreteDistributions`` package.

6.6 A The Multinomial Distribution

The Binomial distribution was discussed in Chapter 3. Here, we present it in its degenerate form: consider an experiment with n independent trials, with two mutually exclusive outcomes per trial ($\mathbb{E}_1$ or $\mathbb{E}_2$). Let p_i $(i = 1, 2)$ denote the probability of outcome $\mathbb{E}_i$ (subject to $p_1 + p_2 = 1$, and $0 \le p_i \le 1$), with p_i remaining the same from trial to trial. Let the 'random variables' of interest be X_1 and X_2, where X_i is the number of trials in which outcome $\mathbb{E}_i$ occurs ($x_1 + x_2 = n$). The joint pmf of X_1 and X_2 is

$$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2) = \frac{n!}{x_1!\, x_2!}\, p_1^{x_1}\, p_2^{x_2}, \qquad x_i \in \{0, 1, \ldots, n\}. \tag{6.33}$$

Since $X_1 + X_2 = n$, one of these 'random variables' is of course degenerate, so that the Binomial is actually a univariate distribution, as in Chapter 3. This framework can easily be generalised into a *Trinomial* distribution, where instead of having just two possible outcomes, we now have three ($\mathbb{E}_1$, $\mathbb{E}_2$ or $\mathbb{E}_3$), subject to $p_1 + p_2 + p_3 = 1$:

$$f(x_1, x_2, x_3) = P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!\, x_2!\, x_3!}\, p_1^{x_1}\, p_2^{x_2}\, p_3^{x_3}. \tag{6.34}$$

More generally, the m-variate *Multinomial* distribution has pmf

$$f(x_1, \ldots, x_m) = P(X_1 = x_1, \ldots, X_m = x_m) = \frac{n!}{x_1! \cdots x_m!}\, p_1^{x_1} \cdots p_m^{x_m} \tag{6.35}$$

$$\text{subject to } \sum_{i=1}^{m} p_i = 1, \text{ and } \sum_{i=1}^{m} x_i = n.$$

Since $\sum_{i=1}^{m} x_i = n$, it follows, for example, that $x_m = n - \sum_{i=1}^{m-1} x_i$. This implies that, given n, the m-variate multinomial can be fully described using only $m - 1$ variables; see also Johnson *et al*. (1997).[7] We enter (6.35) into *Mathematica* as:

$$\texttt{Clear[f];} \quad \texttt{f[X\_List, p\_List, n\_] := n!} \prod_{\texttt{i=1}}^{\texttt{Length[X]}} \frac{\texttt{p[[i]]}^{\texttt{X[[i]]}}}{\texttt{X[[i]]!}}$$

The multinomial moment generating function is derived in *Example 26* below, where we show that

$$M(\vec{t}) = \left(\sum_{i=1}^{m} p_i\, e^{t_i}\right)^n. \tag{6.36}$$

⊕ ***Example 24:*** Age Profile

Table 5 gives the age profile of people living in Australia (Australian Bureau of Statistics, 1996 Census). The data is divided into five age classes.

class	*age*	*proportion*
I	0 – 14	21.6 %
II	15 – 24	14.5 %
III	25 – 44	30.8 %
IV	45 – 64	21.0 %
V	65 +	12.1 %

Table 5: Age profile of people living in Australia

Let $\vec{p}$ denote the probability vector $(p_1, p_2, p_3, p_4, p_5)$:

```
p⃗ = {0.216, 0.145, 0.308, 0.210, 0.121};
```

(a) If we randomly select 10 people from the population, what is the probability they all come from Class I?

Solution:

```
x⃗ = {10, 0, 0, 0, 0};   f[x⃗ , p⃗, 10]
```

2.21074×10^{-7}

(b) If we again randomly select 10 people, what is the probability that 3 people will be from Class I, 1 person from Class II, 2 from Class III, 4 from Class IV, and 0 from Class V?

Solution:

```
x⃗ = {3, 1, 2, 4, 0};   f[x⃗ , p⃗, 10]
```

```
0.00339687
```

(c) If we again randomly select 10 people, what is the probability that Class III will contain exactly 1 person?

Solution: If Class III contains 1 person, then the remaining classes must contain 9 people. Thus, we need to calculate every possible way of splitting 9 people over the remaining four classes, then calculate the probability for each case, and then add it all up. The composition of 9 into 4 parts can be obtained using the `Compositions` function in the ``DiscreteMath`Combinatorica` `` package, which we load as follows:

```
<< DiscreteMath`
```

Here are the compositions of 9 into 4 parts. The list is very long, so we just display the first few compositions:

```
lis = Compositions[9, 4];  lis // Shallow
```

```
{{0, 0, 0, 9}, {0, 0, 1, 8}, {0, 0, 2, 7},
 {0, 0, 3, 6}, {0, 0, 4, 5}, {0, 0, 5, 4}, {0, 0, 6, 3},
 {0, 0, 7, 2}, {0, 0, 8, 1}, {0, 0, 9, 0}, <<210>>}
```

Since Class III must contain 1 person in our example, we need to insert a ‘1’ at position 3 of each of these lists, so that, for instance, {0, 0, 0, 9} becomes {0, 0, 1, 0, 9}:

```
lis2 = Map[Insert[#, 1, 3] &, lis];     lis2 // Shallow
```

```
{{0, 0, 1, 0, 9}, {0, 0, 1, 1, 8},
 {0, 0, 1, 2, 7}, {0, 0, 1, 3, 6}, {0, 0, 1, 4, 5},
 {0, 0, 1, 5, 4}, {0, 0, 1, 6, 3}, {0, 0, 1, 7, 2},
 {0, 0, 1, 8, 1}, {0, 0, 1, 9, 0}, <<210>>}
```

We can now compute the pmf at each of these cases, and add them all up:

```
Plus @@ Map[f[#, p⃗, 10] &, lis2]
```

```
0.112074
```

So, the probability that a random sample of 10 Australians will contain exactly 1 person aged 25–44 is 11.2%. For the 15–24 age group, this probability rises to 35.4%.

An alternative (more automated, but less flexible) approach to solving (c) is to use the summation operator, taking great care to ensure that the summation iterators satisfy the constraint $\sum_{i=1}^{5} x_i = 10$. So, if Class III is fixed at $x_3 = 1$, then x_1 can take values from 0 to 9; x_2 may take values from 0 to $(9 - x_1)$; and x_4 may take values from 0 to $(9 - x_1 - x_2)$. That leaves x_5 which is degenerate: that is, given $x_1, x_2, x_3 = 1$, and x_4, we know that x_5 must equal $9 - x_1 - x_2 - x_4$. Then the required probability is:

```
Sum[f[{x1, x2, 1, x4, x5}, p⃗, 10],
            {x1, 0, 9},
            {x2, 0, 9 - x1},
            {x4, 0, 9 - x1 - x2},
            {x5, 9 - x1 - x2 - x4, 9 - x1 - x2 - x4}]
```

```
0.112074
```

Example 26 provides another illustration of this summation approach. ■

⊕ ***Example 25:*** Working with the mgf

In the case of the Trinomial, the mgf is:

$$\texttt{mgf} = \left(\sum_{i=1}^{3} p_i\, e^{t_i}\right)^n$$

$$(e^{t_1}\, p_1 + e^{t_2}\, p_2 + e^{t_3}\, p_3)^n$$

The product raw moments $E[X_1^a X_2^b X_3^c]$ can now be obtained from the mgf in the usual fashion. To keep things neat, we write a little *Mathematica* function `Moment[`*a*, *b*, *c*`]` function to calculate $E[X_1^a X_2^b X_3^c]$ from the mgf, now noting that $\sum_{i=1}^{m} p_i = 1$:

$$\texttt{Moment[a\_, b\_, c\_] :=}$$
$$\texttt{D[mgf, \{t}_1\texttt{, a\}, \{t}_2\texttt{, b\}, \{t}_3\texttt{, c\}] /. t\_} \to 0 \;\texttt{/.}\; \sum_{i=1}^{3} p_i \to 1$$

The moments are now easy to obtain. Here is the first moment of X_2, namely $\acute{\mu}_{0,1,0}$:

```
Moment[0, 1, 0]
```

$n\, p_2$

Here is the second moment of X_2, namely $\acute{\mu}_{0,2,0}$:

```
Moment[0, 2, 0]
```

$n\, p_2 + (-1 + n)\, n\, p_2^2$

By symmetry, we then have the more general result that $E[X_i] = n\, p_i$ and $E[X_i^2] = n\, p_i + (n-1)\, n\, p_i^2$. Here is the product raw moment $E[X_1^2\, X_2\, X_3] = \acute{\mu}_{2,1,1}$:

```
Moment[2, 1, 1] // Simplify
```

$(-2 + n)\,(-1 + n)\, n\, p_1\,(1 + (-3 + n)\, p_1)\, p_2\, p_3$

The covariance between X_1 and X_3 is given by $\mu_{1,0,1}$, which can be expressed in raw moments as:

```
cov = CentralToRaw[{1, 0, 1}]
```

$\mu_{1,0,1} \to -\acute{\mu}_{0,0,1}\, \acute{\mu}_{1,0,0} + \acute{\mu}_{1,0,1}$

Evaluating each $\acute{\mu}_{\_}$ term with the `Moment` function then yields this covariance:

```
cov /. μ́r__ :→ Moment[r] // Simplify
```

$\mu_{1,0,1} \to -n\, p_1\, p_3$

Similarly, the product cumulant $\kappa_{3,1,2}$ is given by:

```
CumulantToRaw[{3, 1, 2}] /. μ́x__ :→ Moment[x] // Simplify
```

$\kappa_{3,1,2} \to 2\, n\, p_1\, p_2\, p_3\, (1 + p_1^2\,(12 - 60\, p_3) - 3\, p_3 + 9\, p_1\,(-1 + 4\, p_3))$

⊕ ***Example 26:*** Deriving the Multinomial mgf

Consider a model with $m = 4$ classes. The pmf is:

```
x⃗ = {x1, x2, x3, x4};
p⃗ = {p1, p2, p3, p4};
pmf = f[x⃗, p⃗, n]
```

$$\frac{n!\, p_1^{x_1}\, p_2^{x_2}\, p_3^{x_3}\, p_4^{x_4}}{x_1!\, x_2!\, x_3!\, x_4!}$$

Recall that the moment generating function for a discrete distribution is:

$$E\left[e^{\vec{t}.\vec{X}}\right] = \sum_{x_1} \cdots \sum_{x_m} \exp\left(\sum_{i=1}^{m} t_i\, x_i\right) f(x_1, \ldots, x_m).$$

Some care must be taken here to ensure the summation iterators satisfy the constraint $\sum_{i=1}^{m} x_i = n$; thus, if we let x_1 take values from 0 to n, then x_2 may take values from 0 to $n - x_1$, and then x_3 may take values from 0 to $n - x_1 - x_2$. That leaves x_4 which is degenerate; that is, given x_1, x_2 and x_3, we know that x_4 must be equal to $n - x_1 - x_2 - x_3$. Then the mgf is:

```
t⃗ = {t₁, t₂, t₃, t₄};

mgf = FullSimplify[
        Σ_{x₁=0}^{n} Σ_{x₂=0}^{n-x₁} Σ_{x₃=0}^{n-x₁-x₂} Σ_{x₄=n-x₁-x₂-x₃}^{n-x₁-x₂-x₃} Evaluate[e^{t⃗.x⃗} pmf],
      n ∈ Integers] // PowerExpand
```

$$(e^{t_1}\, p_1 + e^{t_2}\, p_2 + e^{t_3}\, p_3 + e^{t_4}\, p_4)^n$$

It follows by symmetry that the general solution is $M(\vec{t}) = \left(\sum_{i=1}^{m} p_i\, e^{t_i}\right)^n$, where $\sum_{i=1}^{m} p_i = 1$. ■

6.6 B The Bivariate Poisson

```
Clear[g]
```

Let Y_0, Y_1 and Y_2 be mutually stochastically independent Poisson random variables, with non-negative parameters λ_0, λ_1 and λ_2, respectively, and pmf's $g_i(y_i)$ for $i \in \{0, 1, 2\}$:

$$g_{i\_} = \frac{e^{-\lambda_i}\, \lambda_i^{y_i}}{y_i\,!};$$

defined on $y_i \in \{0, 1, 2, \ldots\}$. Due to independence, the joint pmf of (Y_0, Y_1, Y_2) is:

```
g = g₀ g₁ g₂
```

$$\frac{e^{-\lambda_0 - \lambda_1 - \lambda_2}\, \lambda_0^{y_0}\, \lambda_1^{y_1}\, \lambda_2^{y_2}}{y_0\,!\; y_1\,!\; y_2\,!}$$

with domain:

```
domain[g] = {{y₀, 0, ∞}, {y₁, 0, ∞}, {y₂, 0, ∞}}
              && {λ₀ > 0, λ₁ > 0, λ₂ > 0} && {Discrete};
```

A non-trivial *bivariate Poisson* distribution is the joint distribution of X_1 and X_2 where

$$X_1 = Y_1 + Y_0 \qquad \text{and} \qquad X_2 = Y_2 + Y_0\,. \tag{6.37}$$

○ ***Probability Mass Function***

We shall consider four approaches for deriving the joint pmf of X_1 and X_2, namely: (i) the transformation method, (ii) the probability generating function (pgf) approach, (iii) limiting forms, and (iv) *Mathematica*'s Statistics package.

(i) *Transformation method*

We wish to find the joint pmf of X_1 and X_2, as defined in (6.37). Let $X_0 = Y_0$ so that the number of new variables X_i is equal to the number of old variables Y_i. Then, the desired transformation here is:

```
eqn = {x1 == y1 + y0, x2 == y2 + y0, x0 == y0};
```

Then, the joint pmf of (X_0, X_1, X_2), say $\psi(x_0, x_1, x_2)$, is:

```
ψ = Transform[eqn, g]
```

$$\frac{e^{-\lambda_0-\lambda_1-\lambda_2}\,\lambda_0^{x_0}\,\lambda_1^{-x_0+x_1}\,\lambda_2^{-x_0+x_2}}{x_0!\,(-x_0+x_1)!\,(-x_0+x_2)!}$$

We desire the joint marginal pmf of X_1 and X_2, so we now need to 'sum out' X_0. Since Y_1 is non-negative, it follows that $X_0 \le X_1$:

$$\text{pmf} = \sum_{x_0=0}^{x_1} \text{Evaluate}[\psi]$$

$$\left(e^{-\lambda_0-\lambda_1-\lambda_2}\,\text{HypergeometricU}\left[-x_1,\,1-x_1+x_2,\,-\frac{\lambda_1\,\lambda_2}{\lambda_0}\right]\lambda_1^{x_1}\,\lambda_2^{x_2}\left(-\frac{\lambda_1\,\lambda_2}{\lambda_0}\right)^{-x_1}\right)\Big/\left(\Gamma[1+x_1]\,\Gamma[1+x_2]\right)$$

Mathematica, ever the show-off, has found the pmf in terms of the confluent hypergeometric function. Here, for instance, is $P(X_1 = 3, X_2 = 2)$:

```
pmf /. {x1 → 3, x2 → 2} // Simplify
```

$$\frac{1}{12}\,e^{-\lambda_0-\lambda_1-\lambda_2}\,\lambda_1\,(6\,\lambda_0^2 + 6\,\lambda_0\,\lambda_1\,\lambda_2 + \lambda_1^2\,\lambda_2^2)$$

(ii) *Probability generating function approach*

By (6.21), the joint pgf is $E\left[t_1^{X_1}\,t_2^{X_2}\,\cdots\,t_m^{X_m}\right]$:

$$\text{pgf} = \sum_{y_0=0}^{\infty}\sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty} \text{Evaluate}[t_1^{y_1+y_0}\,t_2^{y_2+y_0}\,g]$$

$$e^{-\lambda_0+t_1\,t_2\,\lambda_0-\lambda_1+t_1\,\lambda_1-\lambda_2+t_2\,\lambda_2}$$

The pgf, in turn, determines the probabilities by (6.22). Then, $P(X_1 = r, X_2 = s)$ is:

```
Clear[P];
P[r_, s_] := D[pgf, {t1, r}, {t2, s}] / (r! s!) /. {t_ → 0} // Simplify
```

For instance, $P(X_1 = 3, X_2 = 2)$ is:

```
P[3, 2]
```

$$\frac{1}{12} e^{-\lambda_0 - \lambda_1 - \lambda_2} \lambda_1 (6 \lambda_0^2 + 6 \lambda_0 \lambda_1 \lambda_2 + \lambda_1^2 \lambda_2^2)$$

as per our earlier result.

(iii) *Limiting forms*
Just as the univariate Poisson can be obtained as a limiting form of the Binomial, the bivariate Poisson can similarly be obtained as a limiting form of the Multinomial. Hamdan and Al-Bayyati (1969) discuss this approach, while Johnson *et al.* (1997, p. 125) provide an overview.

(iv) *Mathematica's statistics package*
The bivariate Poisson pmf can also be obtained by using *Mathematica*'s `Statistics`MultiDiscreteDistributions`` package, as follows:

```
<< Statistics`
dist = MultiPoissonDistribution[λ0, {λ1, λ2}];
```

Then, the package gives the joint pmf of (X_1, X_2) as:

```
MmaPMF = PDF[dist, {x1, x2}] // Simplify
```

$$e^{-\lambda_0-\lambda_1-\lambda_2} \lambda_1^{x_1} \lambda_2^{x_2} \left(-\left(\text{HypergeometricPFQ}\left[\{1, 1+\text{Min}[x_1, x_2]-x_1, 1+\text{Min}[x_1, x_2]-x_2\}, \{2+\text{Min}[x_1, x_2]\}, \frac{\lambda_0}{\lambda_1 \lambda_2}\right] \left(\frac{\lambda_0}{\lambda_1 \lambda_2}\right)^{1+\text{Min}[x_1, x_2]}\right) \Big/ (\Gamma[2+\text{Min}[x_1, x_2]]\, \Gamma[-\text{Min}[x_1, x_2]+x_1]\, \Gamma[-\text{Min}[x_1, x_2]+x_2]) + \frac{\text{HypergeometricU}[-x_1, 1-x_1+x_2, -\frac{\lambda_1 \lambda_2}{\lambda_0}] \left(-\frac{\lambda_1 \lambda_2}{\lambda_0}\right)^{-x_1}}{\Gamma[1+x_1]\, \Gamma[1+x_2]}\right)$$

While this is not as neat as the result obtained above via the transformation method (i), it nevertheless gives the same results. Here, again, is $P(X_1 = 3, X_2 = 2)$:

```
MmaPMF /. {x1 → 3, x2 → 2} // Simplify
```

$$\frac{1}{12} e^{-\lambda_0 - \lambda_1 - \lambda_2} \lambda_1 (6 \lambda_0^2 + 6 \lambda_0 \lambda_1 \lambda_2 + \lambda_1^2 \lambda_2^2)$$

○ *Moments*

We shall consider three approaches for deriving moments, namely: (i) the direct approach, (ii) the mgf approach, and (iii) moment conversion formulae.

(i) *Direct approach*

Even though we know the joint pmf of X_1 and X_2, it is simpler to work with the underlying Y_i random variables. For instance, suppose we wish to find the product moment $\acute{\mu}_{1,1}$ for the bivariate Poisson. This can be expressed as:

$$\acute{\mu}_{1,1} = E[X_1 X_2] = E[(Y_1 + Y_0)(Y_2 + Y_0)]$$

which is then evaluated as:

$$\sum_{y_2=0}^{\infty} \sum_{y_1=0}^{\infty} \sum_{y_0=0}^{\infty} \textbf{Evaluate}[(y_1 + y_0)\ (y_2 + y_0)\ g]\ //\ \textbf{Expand}$$

$$\lambda_0 + \lambda_0^2 + \lambda_0\lambda_1 + \lambda_0\lambda_2 + \lambda_1\lambda_2$$

(ii) *MGF approach*

The joint mgf of X_1 and X_2 is:

$$E\big[\exp\big(t_1 X_1 + t_2 X_2\big)\big] = E\big[\exp\big(t_1 Y_1 + t_2 Y_2 + (t_1 + t_2) Y_0\big)\big]$$

which is then evaluated as:[8]

$$\textbf{mgf = Simplify}\Big[\sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} \sum_{y_0=0}^{\infty} \textbf{Evaluate}[e^{t_1 y_1 + t_2 y_2 + (t_1+t_2) y_0}\ g]\Big]$$

$$e^{(-1+e^{t_1+t_2})\lambda_0 + (-1+e^{t_1})\lambda_1 + (-1+e^{t_2})\lambda_2}$$

Differentiating the mgf yields the raw product moments, as per (6.18).

```
Moment[r_, s_] := D[mgf, {t1, r}, {t2, s}] /. t_ → 0
```

Then, $\acute{\mu}_{1,1} = E[X_1 X_2]$ is now obtained by:

```
Moment[1, 1] // Expand
```

$$\lambda_0 + \lambda_0^2 + \lambda_0\lambda_1 + \lambda_0\lambda_2 + \lambda_1\lambda_2$$

which is the same result we obtained using the direct method. Here is $\acute{\mu}_{3,1} = E[X_1^3 X_2^1]$:

```
Moment[3, 1]
```

$$\lambda_0 + 6\lambda_0(\lambda_0 + \lambda_1) + 3\lambda_0(\lambda_0 + \lambda_1)^2 + (\lambda_0 + \lambda_1)(\lambda_0 + \lambda_2) + 3(\lambda_0 + \lambda_1)^2(\lambda_0 + \lambda_2) + (\lambda_0 + \lambda_1)^3(\lambda_0 + \lambda_2)$$

The mean vector $\vec{\mu} = (E[X_1], E[X_2])$ is:

```
μ⃗ = {Moment[1, 0], Moment[0, 1]}
```

$\{\lambda_0 + \lambda_1, \ \lambda_0 + \lambda_2\}$

By (6.19), the central mgf is given by:

```
t⃗ = {t₁, t₂};     mgfc = e^(-t⃗.μ⃗) mgf   // Simplify
```

$e^{(-1+e^{t_1+t_2}-t_1-t_2)\lambda_0+(-1+e^{t_1}-t_1)\lambda_1+(-1+e^{t_2}-t_2)\lambda_2}$

Then, $\mu_{1,1} = \text{Cov}(X_1, X_2)$ is:

```
D[mgfc, {t₁, 1}, {t₂, 1}] /. t_ → 0
```

λ_0

while the variances of X_1 and X_2 are, respectively:

```
D[mgfc, {t₁, 2}] /. t_ → 0
```

$\lambda_0 + \lambda_1$

```
D[mgfc, {t₂, 2}] /. t_ → 0
```

$\lambda_0 + \lambda_2$

(iii) *Conversion formulae*

The pgf (derived above) can be used as a factorial moment generating function, as follows:

```
Fac[r_, s_] := D[pgf, {t₁, r}, {t₂, s}] /. t_ → 1
```

Thus, the factorial moment $\mu[1, 2] = E[X_1^{[1]} X_2^{[2]}]$ is given by:

```
Fac[1, 2]
```

$2\lambda_0 (\lambda_0 + \lambda_2) + (\lambda_0 + \lambda_1)(\lambda_0 + \lambda_2)^2$

In part (ii), we found $\acute{\mu}_{3,1} = E[X_1^3 X_2^1]$ using the mgf approach. We now find the same expression, but this time do so using factorial moments. The solution, in terms of *factorial* moments, is:

```
sol = RawToFactorial[{3, 1}]
```

$\acute{\mu}_{3,1} \to \acute{\mu}[1, 1] + 3\,\acute{\mu}[2, 1] + \acute{\mu}[3, 1]$

so $\acute{\mu}_{3,1}$ can be obtained as:

```
sol /. μ́[r__] :→ Fac[r]
```

$$\acute{\mu}_{3,1} \to \lambda_0 + 3\,\lambda_0\,(\lambda_0+\lambda_1)^2 + (\lambda_0+\lambda_1)\,(\lambda_0+\lambda_2) + (\lambda_0+\lambda_1)^3\,(\lambda_0+\lambda_2) + 3\,(2\,\lambda_0\,(\lambda_0+\lambda_1) + (\lambda_0+\lambda_1)^2\,(\lambda_0+\lambda_2))$$

It is easy to show that this is equal to `Moment[3, 1]`, as derived above.

6.7 Exercises

1. Let random variables X and Y have Gumbel's bivariate Logistic distribution with joint pdf

$$f(x,\,y) = \frac{2\,e^{-y-x}}{(1+e^{-y}+e^{-x})^3}, \quad (x,\,y) \in \mathbb{R}^2.$$

(i) Plot the joint pdf; (ii) plot the contours of the joint pdf; (iii) find the joint cdf; (iv) show that the marginal pdf's are Logistic; (v) find the conditional pdf $f(Y \mid X = x)$.

2. Let random variables X and Y have joint pdf

$$f(x,\,y) = \frac{1}{\lambda\,\mu}\exp\left[-\left(\frac{x}{\lambda}+\frac{y}{\mu}\right)\right], \quad \text{defined on} \quad x > 0,\ y > 0$$

with parameters $\lambda > 0$ and $\mu > 0$. Find the bivariate mgf. Use the mgf to find (i) $E[X]$, (ii) $E[Y]$, (iii) $\acute{\mu}_{3,4} = E[X^3\,Y^4]$, (iv) $\mu_{3,4}$. Verify by deriving each expectation directly.

3. Let random variables X and Y have McKay's bivariate Gamma distribution, with joint pdf

$$f(x,\,y) = \frac{c^{a+b}}{\Gamma[a]\,\Gamma[b]}\,x^{a-1}(y-x)^{b-1}\,e^{-c\,y}, \quad \text{defined on} \quad 0 < x < y < \infty$$

with parameters a, b, $c > 0$. Hint: use `domain[`f`]` $= \{\{x, 0, y\},\ \{y, x, \infty\}\}$ *etc.*

(i) Show that the marginal pdf of X is Gamma.

(ii) Find the correlation between X and Y.

(iii) Derive the bivariate mgf. Use it to find $\acute{\mu}_{3,2} = E[X^3 Y^2]$.

(iv) Plot $f(x,\,y)$ when $a = 3$, $b = 2$ and $c = 2$. Hint: use an `If` statement, as per `Plot3D[If[0 < x < y, f, 0], {x, 0, 4}, {y, 0, 4}, etc.]`

(v) Create an animation showing how the pdf plot changes as parameter a increases from 2 to 5 — the animation should look similar to the solution given here:

4. Let random variable $X \sim N(0,\,1)$ and let $Y = X^2 - 2$. Show that $\text{Cov}(X,\,Y) = 0$, even though X and Y are clearly dependent.

5. Let random variables X and Y have a Gumbel (1960) bivariate Exponential distribution (see *Example 12*). Find the regression function $E[Y \mid X = x]$ and the scedastic function $\text{Var}(Y \mid X = x)$. Plot both when $\theta = 0,\ \frac{1}{2},\ 1$.

6. Find a Normal–Exponential bivariate distribution (*i.e.* a distribution whose marginal pdf's are standard Normal and standard Exponential) using the Morgenstern copula method. Find the joint cdf and the variance-covariance matrix.

7. Find a bivariate distribution whose marginal distributions are both standard Exponential, using Frank's copula method. Plot the joint pdf $h(x, y)$ when $\alpha = -10$. Find the conditional pdf $h(x \mid Y = y)$.

8. Gumbel's bivariate Logistic distribution (defined in Exercise 1) has no parameters. While this is virtuous in being simple, it can also be restrictive.

(i) Construct a more general bivariate distribution $h(x, y; \alpha)$ whose marginal distributions are both standard Logistic, using the Ali–Mikhail–Haq copula, with parameter α.

(ii) Show that Gumbel's bivariate Logistic distribution is obtained as the special case $h(x, y; \alpha = 1)$.

(iii) Plot the joint pdf $h(x, y)$ when $\alpha = \frac{1}{2}$.

(iv) Find the conditional pdf $h(x \mid Y = y)$.

9. Let $f(x, y; \vec{\mu}, \Sigma)$ denote the joint pdf of a bivariate Normal distribution $N(\vec{\mu}, \Sigma)$. For $0 < \omega < 1$, define a bivariate Normal component-mixture density by:

$$\tilde{f}(x, y) = \omega f(x, y; \vec{\mu}_1, \Sigma_1) + (1 - \omega) f(x, y; \vec{\mu}_2, \Sigma_2)$$

Let $\vec{\mu}_1 = (2, 2)$, $\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\vec{\mu}_2 = (0, 0)$ and $\Sigma_2 = \begin{pmatrix} 1 & \frac{3}{4} \\ \frac{3}{4} & 1 \end{pmatrix}$.

(i) Find the functional form for $\tilde{f}(x, y)$.

(ii) Plot $\tilde{f}(x, y)$ when $\omega = \frac{7}{10}$. Construct contour plots of $\tilde{f}(x, y)$ when $\omega = 0$ and when $\omega = 1$.

(iii) Create an animation showing how the contour plot changes as ω increases from 0 to 1 in step sizes of 0.025 — the animation should look something like the solution given here:

(iv) Find the marginal pdf of X, namely $\tilde{f}_x(x)$. Find the mean and variance of the latter.

(v) Plot the marginal pdf derived in (iv) when $\omega = 0$, $\frac{1}{2}$ and 1.

10. Let random variables (W, X, Y, Z) have a multivariate Normal distribution $N(\vec{\mu}, \Sigma)$, with:

$$\vec{\mu} = (0, 0, 0, 0), \qquad \Sigma = \begin{pmatrix} 1 & \frac{2}{3} & \frac{3}{4} & \frac{4}{5} \\ \frac{2}{3} & 1 & \frac{1}{2} & \frac{8}{15} \\ \frac{3}{4} & \frac{1}{2} & 1 & \frac{3}{5} \\ \frac{4}{5} & \frac{8}{15} & \frac{3}{5} & 1 \end{pmatrix}$$

(i) Find the joint pdf $f(w, x, y, z)$.

(ii) Use the multivariate Normal mgf, $\exp(\vec{t}.\vec{\mu} + \frac{1}{2}\,\vec{t}.\Sigma.\vec{t})$, to find $E[W X Y Z]$ and $E[W X^2 Y Z^2]$.

(iii) Find $E[W \exp(X + Y + Z)]$.

(iv) Use Monte Carlo methods (not numerical integration) to check whether the solution to (iii) seems 'correct'.

(v) Find $P(-3 < W < 3, -2 < X < \infty, -7 < Y < 2, -1 < Z < 1)$.