A.2 Inferential processes

**Multinomial Model**

\[
\begin{align*}
z &= \{r_1, \ldots, r_k, n\}, \quad r_i = 0, 1, 2, \ldots, \quad \sum_{i=1}^{k} r_i \leq n \\
p(z \mid \theta) &= \text{Mult}(z \mid \theta, n), \quad 0 < \theta_1 < 1, \quad \sum_{i=1}^{k} \theta_i \leq 1 \\
t(z) &= (r, n), \quad \theta = (r_1, \ldots, r_k) \\
p(r \mid \theta) &= \text{Mult}(r \mid \theta, n) \\
p(\theta) &= \text{Dir}(\theta \mid \alpha), \quad \alpha = \{\alpha_1, \ldots, \alpha_{k+1}\} \\
p(r \mid \alpha) &= \text{Mult}(r \mid \alpha, n) \\
p(\theta \mid z) &= \text{Dir}(\theta \mid \alpha_1 + r_1, \ldots, \alpha_k + r_k, \alpha_{k+1} + n - \sum_{i=1}^{k} r_i) \\
p(z \mid \theta) &= \text{Mult}(z \mid \alpha_1 + r_1, \ldots, \alpha_k + r_k, \alpha_{k+1} + n - \sum_{i=1}^{k} r_i, n)
\end{align*}
\]

**Multivariate Normal Model**

\[
\begin{align*}
z &= \{x_1, \ldots, x_n\}, \quad x_i \in \mathbb{R}^k \\
p(x_i \mid \mu, \lambda) &= \mathcal{N}(x_i \mid \mu, \lambda), \quad \mu \in \mathbb{R}^k, \quad \lambda_{k \times k \text{ positive-definite}} \\
t(z) &= (\bar{x}, S), \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad S = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^t \\
p(\bar{x} \mid \mu, \lambda) &= \mathcal{N}(\bar{x} \mid \mu, n\lambda) \\
p(S \mid \lambda) &= \mathcal{W}(S \mid \frac{1}{2}(n-1), \frac{1}{2}\lambda) \\
p(\mu, \lambda) &= \mathcal{N}_k(\mu \mid \mu_0, n_0, \alpha, \beta) = \mathcal{N}_k(\mu \mid \mu_0, n_0\lambda) \mathcal{W}(\lambda \mid \alpha, \beta) \\
p(x) &= \text{St}_k(\mu \mid \mu_0, (n_0 + 1)^{-1}n_0(\alpha = \frac{1}{2}(k - 1))\beta^{-1}, 2\alpha - k + 1) \\
p(\mu \mid z) &= \text{St}_k(\mu \mid \mu_n, (n + n_0)\alpha_n\beta_n^{-1}, 2\alpha_n), \\
\quad \mu_n = (n_0 + n)^{-1}(n_0\mu_0 + n\bar{x}), \\
\quad \beta_n = \beta + \frac{1}{2}S + \frac{1}{2}(n + n_0)^{-1}n_0(n\mu_0 - \bar{x})(\mu_0 - \bar{x})^t \\
p(\lambda \mid z) &= \mathcal{W}(\lambda \mid \alpha + \frac{1}{2}n, \beta_n) \\
p(x \mid z) &= \text{St}_k(x \mid \mu_n, (n_0 + n + 1)^{-1}(n_0 + n)\alpha_n\beta_n^{-1}, 2\alpha_n), \\
\quad \alpha_n = \alpha + \frac{1}{2}n - \frac{1}{2}(k - 1)
\end{align*}
\]
3 Generalisations

The mean vector and covariance matrix are given by

$$E[x_i] = np_i, \quad p_i = \frac{\alpha_i}{\sum_{j=1}^{k+1} \alpha_j}$$

$$V[x_i] = \frac{n + \sum_{j=1}^{k+1} \alpha_j}{1 + \sum_{j=1}^{k+1} \alpha_j} np_i(1 - p_i)$$

$$C[x_i, x_j] = \frac{n + \sum_{j=1}^{k+1} \alpha_j}{1 + \sum_{j=1}^{k+1} \alpha_j} np_i p_j$$

The marginal distribution of the subset \(\{x_1, \ldots, x_s\}\) is a multinomial-Dirichlet with parameters \(\{\alpha_1, \ldots, \alpha_s, \sum_{j=1}^{s} \alpha_j - \sum_{j=1}^{s} \alpha_j\}\) and \(n\). In particular, the marginal distribution of \(x_i\) is the binomial-beta \(B_b(x_i | \alpha_i, \sum_{j=1}^{k+1} \alpha_j - \alpha_i)\). Moreover, the conditional distribution of \(\{x_{s+1}, \ldots, x_k\}\) given \(\{x_1, \ldots, x_s\}\) is also multinomial-Dirichlet, with parameters \(\{\alpha_{s+1}, \ldots, \alpha_k, \sum_{j=1}^{s+1} \alpha_j - \sum_{j=s+1}^{k+1} \alpha_j\}\) and \(n - \sum_{j=1}^{s} x_j\). For an interesting characterization of this distribution, see Basu and Pereira (1983).

The Normal-Gamma Distribution

A continuous bivariate random vector \((x, y)\) has a normal-gamma distribution, with parameters \(\mu, \lambda, \alpha, \beta\), \((\mu \in \mathbb{R}, \lambda > 0, \alpha > 0, \beta > 0)\) if its density

$$Ng(x, y | \mu, \lambda, \alpha, \beta) = N(x | \mu, \lambda y)Ga(y | \alpha, \beta), \quad x \in \mathbb{R}, y > 0,$$

where the normal and gamma densities are defined in Section 3.2.2. It is clear from the definition, that the conditional density of \(x\) given \(y\) is \(N(x | \mu, \lambda y)\) and that the marginal density of \(y\) is \(Ga(y | \alpha, \beta)\). Moreover, the marginal density of \(x\) is \(St(x | \mu, \alpha \lambda / \beta, 2 \alpha)\).

The shape of a normal-gamma distribution is illustrated in Figure 3.1, where the probability density of \(Ng(x, y | 0, 1, 5, 5)\) is displayed both as a surface and in terms of equal density contours.

The Multivariate Normal Distribution

A continuous random vector \(x = (x_1, \ldots, x_k)\) has a multivariate normal distribution of dimension \(k\), with parameters \(\mu = (\mu_1, \ldots, \mu_k)\) and \(\lambda\), where \(\mu \in \mathbb{R}^k\) and \(\lambda\) is a \(k \times k\) symmetric positive-definite matrix, if its probability density \(N_k(x | \mu, \lambda)\) is

$$N_k(x | \mu, \lambda) = c \exp\{-\frac{1}{2}(x - \mu)^t \lambda (x - \mu)\}, \quad x \in \mathbb{R}^k,$$
where $c = (2\pi)^{-k/2} |\Lambda|^{1/2}$.

If $k = 1$, so that $\Lambda$ is a scalar, $\lambda$, $N_k(x \mid \mu, \lambda)$ reduces to the univariate normal density $N(x \mid \mu, \lambda)$.

In the general case, $E[x_i] = \mu_i$, and, with $\Sigma = \lambda^{-1}$ of general element $\sigma_{ij}$, $E[x_i^2] = \sigma_{ii}$ and $C[x_i, x_j] = \sigma_{ij}$, so that $V[x] = \lambda^{-1}$. The parameter $\mu$ therefore is the mean vector and the parameter $\lambda$ the precision matrix (the inverse of the
covariance matrix, $\Sigma$). If $y = A x$, where $A$ is an $m \times k$ matrix of real numbers such that $A \Sigma A^T$ is non-singular, then $y$ has density $N_m(y \mid A \mu, (A \Sigma A^T)^{-1})$.

In particular, the marginal density for any subvector of $x$ is (multivariate) normal, of appropriate dimension, with mean vector and covariance matrix given by the corresponding subvector of $\mu$ and submatrix of $\Lambda^{-1}$.

Moreover, if $x = (x_1, x_2)$ is a partition of $x$, with $x_1$ having dimension $k_1$ and $k_1 + k_2 = k$, and if the corresponding partitions of $\mu$ and $\lambda$ are

$$
\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix},
$$

then the conditional density of $x_1$ given $x_2$ is also (multivariate) normal, of dimension $k_1$ with mean vector and precision matrix given, respectively, by

$$
\mu_1 - \lambda_{11}^{-1}\lambda_{12}(x_2 - \mu_2) \quad \text{and} \quad \lambda_{11},
$$

The random quantity $y = (x - \mu)^T \lambda (x - \mu)$ has a $\chi^2(y \mid k)$ density.

We also note that, from the form of the multivariate normal density, we can deduce the integral formula

$$
\int_{\mathbb{R}^k} \exp\left\{-\frac{1}{2}(x - \mu)^T \lambda (x - \mu)\right\} dx = \frac{(2\pi)^{k/2}}{|\lambda|^{1/2}}.
$$

The Wishart Distribution

A symmetric, positive-definite matrix $x$ of random quantities $x_{ij} = x_{ji}$, for $i = 1, \ldots, k, j = 1, \ldots, k$, has a Wishart distribution of dimension $k$, with parameters $\alpha$ and $\beta$ (with $2\alpha > k - 1$ and $\beta$ a $k \times k$ symmetric, nonsingular matrix), if the density $W_k(x \mid \alpha, \beta)$ of the $k(k+1)/2$ dimensional random vector of the distinct entries of $x$ is

$$
W_k(x \mid \alpha, \beta) = c |x|^{(\alpha-(k+1)/2)} \exp\{-\operatorname{tr} (\beta x)\},
$$

where $c = |\beta|^{\alpha/2} \Gamma_k(\alpha),$

$$
\Gamma_k(\alpha) = \pi^{(k-1)/4} \prod_{i=1}^{k-1} \Gamma\left(\frac{2\alpha + 1 - i}{2}\right)
$$

is the generalized gamma function and $\operatorname{tr}()$, as before, denotes the trace of a matrix argument. If $k = 1$, so that $\beta$ is a scalar $\beta$, then $W_k(x \mid \alpha, \beta)$ reduces to the gamma density $Ga(x \mid \alpha, \beta)$.

$\Sigma_k(x \mid \mu, \lambda, \alpha)$ is independent of $\bar{x}$.

The following $f(x) = \alpha \beta^{-1}$ and $\Sigma$ an $m \times k$ matrix $m$-dimension $m$ with parameters $\alpha, \beta$, $\mathcal{N}(x \mid \mu, \lambda)$, and $\bar{x}$ is

$$
x =
$$

where $x_{ii}, \sigma_{ii}$ are $s$ distribution are independent $k \times$ parameters $\alpha, \beta$, with parameters $\alpha_1 +$ parameters $\alpha_1 +$

We note that, from the formula

$$
\int \ldots
$$

the integration being elements of the matrix

The Multivariate Student

A continuous random distribution of dimension $k$ a symmetric, positive

$$
St_k(x \mid \mu, \lambda, \alpha)
$$

where

If $k = 1$, so that $\lambda$ is the Student density $St_1(\lambda^{-1}(\alpha/(\alpha - 2)))$. Although the parameter $\lambda$ is
3.2 Review of Probability Theory

If \( \{x_1, \ldots, x_n\} \) is a random sample of size \( n > 1 \) from a multivariate normal \( N_k(x_i | \mu, \lambda) \), and \( \bar{x} = n^{-1} \sum x_i \), then \( \bar{x} \) is \( N_k(\bar{x} | \mu, n\lambda) \), and

\[
S = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^t
\]

is independent of \( \bar{x} \), and has a Wishart distribution \( W_k(S | \frac{1}{2}(n - 1), \frac{1}{2}\lambda) \).

The following properties of the Wishart distribution are easily established: \( E[x] = \alpha\beta^{-1} \) and \( E[x^{-1}] = (\alpha - (k + 1)/2)^{-1} \beta \); if \( y = AxA^t \) where \( A \) is an \( m \times k \) matrix \( (m \leq k) \) of real numbers, then \( y \) has a Wishart distribution of dimension \( m \) with parameters \( \alpha \) and \( (A\beta^{-1}A^t)^{-1} \), if the latter exists; in particular, if \( x \) and \( \beta^{-1} \) conformably partition into

\[
\begin{pmatrix}
\mathbf{x}_{11} & \mathbf{x}_{12} \\
\mathbf{x}_{21} & \mathbf{x}_{22}
\end{pmatrix}, \quad \beta^{-1} = \begin{pmatrix}
\mathbf{\sigma}_{11} & \mathbf{\sigma}_{12} \\
\mathbf{\sigma}_{21} & \mathbf{\sigma}_{22}
\end{pmatrix},
\]

where \( \mathbf{x}_{11}, \mathbf{\sigma}_{11} \) are square \( h \times h \) matrices \((1 \leq h < k)\), then \( \mathbf{x}_{11} \) has a Wishart distribution of dimension \( h \) with parameters \( \alpha \) and \( (\mathbf{\sigma}_{11})^{-1} \). Moreover, if \( x_1, \ldots, x_s \) are independent \( k \times k \) random matrices, each with a Wishart distribution, with parameters \( \alpha_i, \beta, i = 1, \ldots, s \), then \( x_1 + \cdots + x_s \), also has a Wishart distribution, with parameters \( \alpha_1 + \cdots + \alpha_s \) and \( \beta \).

We note that, from the form of the Wishart density, we can deduce the integral formula

\[
\int |x|^{n-(k+1)/2} \exp\{ -\text{tr}(\beta x) \} \, dx = c^{-1},
\]

the integration being understood to be with respect to the \( k(k+1)/2 \) distinct elements of the matrix \( x \).

The Multivariate Student Distribution

A continuous random vector \( x = (x_1, \ldots, x_k) \) has a multivariate Student distribution of dimension \( k \), with parameters \( \mu = (\mu_1, \ldots, \mu_k) \), \( \lambda \) and \( \alpha \), if \( \mu \in \mathbb{R}^k \), \( \lambda \) is symmetric, positive-definite \( k \times k \) matrix, \( \alpha > 0 \) if its probability density

\[
S_k(x | \mu, \lambda, \alpha) = c \left[ 1 + \frac{1}{\alpha}(x - \mu)^t\lambda(x - \mu) \right]^{-(\alpha+k)/2}, \quad x \in \mathbb{R}^k,
\]

where

\[
c = \frac{\Gamma((\alpha + k)/2)}{\Gamma(\alpha/2)(\alpha \pi)^{k/2}|\lambda|^{1/2}}.
\]

If \( \lambda = 1 \), so that \( \lambda \) is a scalar, \( \lambda \), then \( S_k(x | \mu, \lambda, \alpha) \) reduces to the univariate Student density \( S(x | \mu, \lambda, \alpha) \). In the general case, \( E[x] = \mu \) and \( V[x] = \lambda^{-1}(\alpha/(\alpha - 2)). \) Although not exactly equal to the inverse of the covariance matrix, the parameter \( \lambda \) is often referred to as the precision matrix of the distribution.
If \( y = Ax \), where \( A \) is an \( m \times k \) matrix (\( m \leq k \)) of real numbers such that \( A\lambda^{-1}A^t \) is non-singular, then \( y \) has density \( \text{St}_m(y \mid A\mu, (A\lambda^{-1}A^t)^{-1}, \alpha) \). In particular, the marginal density for any subvector of \( x \) is (multivariate) Student, of appropriate dimension, with mean vector and inverse of the precision matrix given by the corresponding subvector of \( \mu \) and submatrix of \( \lambda^{-1} \). Moreover, if \( x = (x_1, x_2) \) is a partition of \( x \) and the corresponding partitions of \( \mu \) and \( \lambda \) are given by

\[
\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix},
\]

then the conditional density of \( x_1 \), given \( x_2 \) is also (multivariate) Student, of dimension \( k_1 \), with \( \alpha + k_2 \) degrees of freedom, and mean vector and precision matrix, respectively, given by

\[
\mu_1 - \lambda_{11}^{-1}\lambda_{12}(x_2 - \mu_2), \quad \alpha + k_2, \quad \lambda_{11}^{-1} - \lambda_{22}^{-1}\lambda_{12}^{-1} \lambda_{12} (x_2 - \mu_2) \lambda_{22}^{-1}.
\]

The random quantity \( y = (x - \mu)^t\lambda(x - \mu) \) has an \( F_k(y \mid k, \alpha) \) density.

### The Multivariate Normal-Gamma Distribution

A continuous random vector \( x = (x_1, \ldots, x_k) \) and a random quantity \( y \) have a joint multivariate normal-gamma distribution of dimension \( k \), with parameters \( \mu, \lambda, \alpha, \beta \) (\( \mu \in \mathbb{R}^k \), \( \lambda \) a \( k \times k \) symmetric, positive-definite matrix, \( \alpha > 0 \) and \( \beta > 0 \)) if the joint probability density of \( x \) and \( y \), \( \text{Ng}_k(x, y \mid \mu, \lambda, \alpha, \beta) \), is

\[
\text{Ng}_k(x, y \mid \mu, \lambda, \alpha, \beta) = \text{N}_k(x \mid \mu, \lambda y) \text{Ga}(y \mid \alpha, \beta),
\]

where the multivariate normal and gamma densities have already been defined.

From the definition, the conditional density of \( x \) given \( y \) is \( \text{N}_k(x \mid \mu, \lambda y) \) and the marginal density of \( y \) is \( \text{Ga}(y \mid \alpha, \beta) \). Moreover, the marginal density of \( x \) is \( \text{St}_k(x \mid \mu, \alpha^{-1}\beta\lambda, 2\alpha) \).

### The Multivariate Normal-Wishart Distribution

A continuous random vector \( x \) and a symmetric, positive-definite matrix of random quantities \( y \) have a joint Normal-Wishart distribution of dimension \( k \), with parameters \( \mu, \lambda, \alpha, \beta \) (\( \mu \in \mathbb{R}^k \), \( \lambda > 0 \), integer \( 2\alpha > k - 1 \), and \( \beta \) a \( k \times k \) symmetric, non-singular matrix), if the probability density of \( x \) and the \( k(k + 1)/2 \) distinct elements of \( y \), \( \text{Nw}_k(x, y \mid \mu, \lambda, \alpha, \beta) \), is

\[
\text{Nw}_k(x, y \mid \mu, \lambda, \alpha, \beta) = \text{N}_k(x \mid \mu, \lambda y) \text{W}_k(y \mid \alpha, \beta),
\]

where the multivariate normal and Wishart densities are as defined above.

From the definition, the conditional density of \( x \) given \( y \) is \( \text{N}_k(x \mid \mu, \lambda y) \) and the marginal density of \( y \) is \( \text{W}_k(y \mid \alpha, \beta) \). Moreover, the marginal density of \( x \) is \( \text{St}_k(x \mid \mu, \alpha^{-1}\beta\lambda, 2\alpha) \).
3.3 Generalised Options and Utilities

The Bilateral Pareto Distribution
A continuous bivariate random vector \((x, y)\) has a bilateral Pareto distribution with parameters \(\beta_0, \beta_1\), and \(\alpha\) \(\{\beta_0, \beta_1\} \in \mathbb{R}^2, \beta_0 < \beta_1, \alpha > 0\) if its density function
\[
Pa_2(x, y | \alpha, \beta_0, \beta_1) = c (y - x)^{-(\alpha+2)}, \quad x \leq \beta_0, \ y \geq \beta_1,
\]
where \(c = \alpha(\alpha + 1)(\beta_1 - \beta_0)^\alpha\). The mean and variance are given by
\[
E[x] = \frac{\alpha \beta_0 - \beta_1}{\alpha - 1}, \quad E[y] = \frac{\alpha \beta_1 - \beta_0}{\alpha - 1}, \quad \text{if } \alpha > 1,
\]
\[
V[x] = V[y] = \frac{\alpha(\beta_1 - \beta_0)^2}{(\alpha - 1)^2(\alpha - 2)}, \quad \text{if } \alpha > 2,
\]
and the correlation between \(x\) and \(y\) is \(-\alpha^{-1}\). The marginal distributions of \(t_1 = \beta_1 - x\) and \(t_2 = y - \beta_0\) are both \(Pa(t \mid \beta_1 - \beta_0, \alpha)\).

3.3 GENERALISED OPTIONS AND UTILITIES

3.3.1 Motivation and Preliminaries

For reasons of mathematical or descriptive convenience, it is common in statistical decision problems to consider sets of options which consist of part or all of the real line (as in problems of point estimation) or are part of some more general space. It is therefore desirable to extend the concepts and results of Chapter 2 to a much more general mathematical setting, going beyond finite, or even countable, frameworks, first by taking \(\mathcal{E}\) to be a \(\sigma\)-algebra and then suitably extending the fundamental notion of an option.

In the finite case, an option was denoted by \(a = \{c_j \mid E_j, j \in J\}\), with the straightforward interpretation that, if option \(a\) is chosen, \(c_j\) is the consequence of the occurrence of the event \(E_j\). The extension of this function definition to infinite settings clearly requires some form of constructive limit process, analogous to that used in Lebesgue measure and integration theory in passing from simple (i.e., "step") functions to more general functions. Since the development given in Chapter 2 led to the assessment of options in terms of their expected utilities, the "natural" definition of limit that suggests itself is one based fundamentally on the expected utility idea (Bernardo, Ferrándiz and Smith, 1985).

Let us therefore consider a decision problem \(\{\mathcal{A}, \mathcal{E}, \mathcal{C}, \leq\}\), which is described by a probability space \(\{\Omega, \mathcal{F}, P\}\) and utility function \(u : \mathcal{C} \rightarrow \mathbb{R}\), and let
\[
D = \{d : \Omega \rightarrow \mathcal{C}; \quad u(d) = \int_{\Omega} u(d(\omega)) \, dP(\omega) < \infty\}.
\]