A

Useful distributions

Bernoulli distribution

A Bernoulli random variable is the indicator function of an event or, in other words, a discrete random variable whose only possible values are zero and one. If $X \sim \mathcal{B}e(p)$,

$$P(X = 1) = 1 - P(X = 0) = p.$$

The probability mass function is

$$\mathcal{B}e(x; p) = \begin{cases} 1 - p & \text{if } x = 0, \\ p & \text{if } x = 1. \end{cases}$$

Normal distribution

Arguably the most used (and abused) probability distribution. Its density is

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

with expected value μ and variance σ^2.

Beta distribution

The support of a Beta distribution is the interval (0,1). For this reason it is often used as prior distribution for an unknown probability. The distribution is parametrized in terms of two positive parameters, a and b, and is denoted by $\mathcal{B}(a, b)$. Its density is

$$\mathcal{B}(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \qquad 0 < x < 1,$$

For a random variable $X \sim \mathcal{B}(a, b)$ we have

$$\mathrm{E}(X) = \frac{a}{a+b}, \qquad \mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

A multivariate generalization is provided by the Dirichlet distribution.

Gamma distribution

A random variable X has a Gamma distribution, with parameters (a, b), if it has density

$$\mathcal{G}(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \qquad x > 0$$

where a and b are positive parameters. We find that

$$\mathrm{E}(X) = \frac{a}{b}, \qquad \mathrm{Var}(X) = \frac{a}{b^2}.$$

If $a > 1$, there is a unique mode at $(a-1)/b$. For $a = 1$, the density reduces to the (negative) exponential distribution with parameter b. For $(a = k/2, b = 1/2)$ it is a Chi-square distribution with k degrees of freedom, $\chi^2(k)$.

If $X \sim \mathcal{G}(a, b)$, the density of $Y = 1/X$ is called Inverse-Gamma, with parameters (a, b), and we have $\mathrm{E}(Y) = b/(a-1)$ if $a > 1$ and $\mathrm{Var}(Y) = b^2/((a-1)^2(a-2))$ if $a > 2$.

Student-t distribution

If $Z \sim \mathcal{N}(0, 1), U \sim \chi^2(k), k > 0$ and Z and U are independent, then the random variable $T = Z/\sqrt{U/k}$ has a (central) *Student-t distribution* with k degrees of freedom, with density

$$f(t; k) = c \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}},$$

where $c = \Gamma((k+1)/2)/(\Gamma(k/2)\sqrt{k\pi})$. We write $T \sim \mathcal{T}(0, 1, k)$ or simply $T \sim \mathcal{T}_k$.

It is clear from the definition that the density is positive on the whole real line and symmetric around the origin. It can be shown that, as k increases to infinity, the density converges to a standard Normal density at any point. We have

$$\mathrm{E}(X) = 0 \qquad \text{if } k > 1,$$
$$\mathrm{Var}(X) = \frac{k}{k-2} \qquad \text{if } k > 2.$$

If $T \sim \mathcal{T}(0, 1, k)$, then $X = \mu + \sigma T$ has a Student-t distribution, with parameters (μ, σ^2) and k degrees of freedom; we write $X \sim \mathcal{T}(\mu, \sigma^2, k)$. Clearly $\mathrm{E}(X) = \mu$ if $k > 1$ and $\mathrm{Var}(X) = \sigma^2 \frac{k}{k-2}$ if $k > 2$.

Normal-Gamma distribution

Let (X, Y) be a bivariate random vector. If $X|Y = y \sim \mathcal{N}(\mu, (n_0 y)^{-1})$, and $Y \sim \mathcal{G}(a, b)$, then we say that (X, Y) has a Normal-Gamma density with parameters (μ, n_0^{-1}, a, b) (where of course $\mu \in \mathbb{R}$, $n_0, a, b \in \mathbb{R}^+$). We write $(X, Y) \sim \mathcal{NG}(\mu, n_0^{-1}, a, b)$ The marginal density of X is a Student-t, $X \sim \mathcal{T}(\mu, (n_0 \frac{a}{b})^{-1}, 2a)$.

Multivariate Normal distribution

A continuous random vector $Y = (Y_1, \ldots, Y_k)'$ has a k-variate Normal distribution with parameters $\mu = (\mu_1, \ldots, \mu_k)'$ and Σ, where $\mu \in \mathbb{R}^k$ and Σ is a symmetric positive-definite matrix, if it has density

$$\mathcal{N}_k(y; \mu, \Sigma) = |\Sigma|^{-1/2}(2\pi)^{-k/2} \exp\left\{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right\}, \qquad y \in \mathbb{R}^k$$

where $|\Sigma|$ denotes the determinant of the matrix Σ. We write

$$Y \sim \mathcal{N}_k(\mu, \Sigma).$$

Clearly, if $k = 1$, so that Σ is a scalar, the $\mathcal{N}_k(\mu, \Sigma)$ reduces to the univariate Normal density.

We have $\mathrm{E}(Y_i) = \mu_i$ and, denoting by $\sigma_{i,j}$ the elements of Σ, $\mathrm{Var}(Y_i) = \sigma_{i,i}$ and $\mathrm{Cov}(Y_i, Y_j) = \sigma_{i,j}$. The inverse of the covariance matrix Σ, $\Phi = \Sigma^{-1}$ is the precision matrix of Y.

Several results are of interest; their proof can be found in any multivariate analysis textbook (see, e.g. Barra and Herbach; 1981, pp.92,96).

1. If $Y \sim \mathcal{N}_k(\mu, \Sigma)$ and X is a linear transformation of Y, that is $X = AY$ where A is a $n \times k$ matrix, then $X \sim \mathcal{N}_k(A\mu, A\Sigma A')$.
2. Let X and Y be two random vectors, with covariance matrices Σ_X and Σ_Y, respectively. Let Σ_{YX} be the covariance between Y and X, i.e. $\Sigma_{YX} = \mathrm{E}((Y - \mathrm{E}(Y))(X - \mathrm{E}(X))')$. The covariance between X and Y is then $\Sigma_{XY} = \Sigma'_{YX}$. Suppose that Σ_X is nonsingular. Then it can be proved that the joint distribution of (X, Y) is Gaussian if and only if the following conditions are satisfied:
   (i) X has a Gaussian distribution;
   (ii) the conditional distribution of Y given $X = x$ is a Gaussian distribution whose mean is
   $$\mathrm{E}(Y|X = x) = \mathrm{E}(Y) + \Sigma_{YX}\Sigma_X^{-1}(x - \mathrm{E}(X))$$
   and whose covariance matrix is
   $$\Sigma_{Y|X} = \Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}.$$

Multinomial distribution

Consider a set of n independent and identically distributed observations taking values in a finite label set $\{L_1, L_2, \ldots, L_k\}$. Denote by p_i the probability of an observation being equal to L_i, $i = 1, \ldots, k$. The vector of label counts $X = (X_1, \ldots, X_k)$, where X_i is the number of observations equal to L_i ($i = 1, \ldots, k$) has a Multinomial distribution, whose probability mass function is

$$\mathcal{M}ult(x_1, \ldots, x_k; n, p) = \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k},$$

where $p = (p_1, \ldots, p_k)$ and the counts $x_1, \ldots, x_k$ satisfy the constraint $\sum x_i = n$.

Dirichlet distribution

The Dirichlet distribution is a multivariate generalization of the Beta distribution. Consider a parameter vector $a = (a_1, \ldots, a_k)$. The Dirichlet distribution $\mathcal{D}ir(a)$ has $k-1$-dimensional density

$$\mathcal{D}ir(x_1, \ldots, x_{k-1}; a) = \frac{\Gamma(a_1 + \cdots + a_k)}{\Gamma(a_1) \ldots \Gamma(a_k)} x_1^{a_1 - 1} \ldots x_{k-1}^{a_{k-1}-1} \left(1 - \sum_{i=1}^{k-1} x_i\right)^{a_k - 1},$$

$$\text{for} \qquad \sum_{i=1}^{k-1} x_i < 1, \quad x_i > 0, \quad i = 1 \ldots, k-1.$$

Wishart distribution

Let W be a symmetric positive-definite matrix of random variables $w_{i,j}$, $i, j = 1, \ldots, k$. The distribution of W is the joint distribution of its entries (in fact, the distribution of the $k(k+1)/2$-dimensional vector of the distinct entries). We say that W has a *Wishart* distribution with parameters α and B ($\alpha > (k-1)/2$ and B a symmetric, positive-definite matrix), if it has density

$$\mathcal{W}_k(W; \alpha, B) = c|W|^{\alpha - (k+1)/2} \, \exp\big(-\operatorname{tr}(BW)\big),$$

where $c = |B|^\alpha / \Gamma_k(\alpha)$, $\Gamma_k(\alpha) = \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma((2\alpha + 1 - i)/2)$ is the *generalized gamma function* and $\operatorname{tr}(\cdot)$ denotes the trace of a matrix argument. We write $W \sim \mathcal{W}_k(\alpha, B)$ or just $W \sim \mathcal{W}(\alpha, B)$. We have

$$\mathrm{E}(W) = \alpha \, B^{-1}.$$

The Wishart distribution arises in sampling from a multivariate Gaussian distribution. If $(Y_1, \ldots, Y_n)$, $n > 1$, is a random sample from a multivariate normal distribution $\mathcal{N}_k(\mu, \Sigma)$ and $\bar{Y} = \sum_{i=1}^n Y_i/n$, then $\bar{Y} \sim \mathcal{N}_k(\mu, \Sigma/n)$ and

$$S = \sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})'$$

is independent of $\bar{Y}$ and has a Wishart distribution $\mathcal{W}_k((n-1)/2, \Sigma^{-1}/2)$. In particular, if $\mu = 0$, then

$$W = \sum_{i=1}^n Y_i Y_i' \sim \mathcal{W}_k\left(\frac{n}{2}, \frac{1}{2}\Sigma^{-1}\right),$$

whose density (for $n > k-1$) is

$$f(w; n, \Sigma) \propto |W|^{\frac{n-k-1}{2}} \, \exp\left\{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1} W)\right\}.$$

In fact, the Wishart distribution is usually parametrized in n and Σ, as in the expression above; then the parameter n is called *degrees of freedom*. Note that $\mathrm{E}(W) = n\Sigma$. We used the parametrization in α and B for analogy with the Gamma distribution; indeed, if $k = 1$, so that B is a scalar, then $\mathcal{W}_1(\alpha, B)$ reduces to the Gamma density $\mathcal{G}(\cdot; \alpha, B)$.

The following properties of the Wishart distribution can be proved. Let $W \sim \mathcal{W}_k(\alpha = n/2, B = \Sigma^{-1}/2)$ and $Y = AWA'$, where A is an $(m \times k)$ matrix of real numbers ($m \leq k$). Then Y has a Wishart distribution of dimension m with parameters α and $\frac{1}{2}(A\Sigma A)^{-1}$, if the latter exists. In particular, if W and Σ conformably partition into

$$W = \begin{bmatrix} W_{1,1} & W_{1,2} \\ W_{2,1} & W_{2,2} \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix},$$

where $W_{1,1}$ and $\Sigma_{1,1}$ are $h \times h$ matrices $(1 \leq h < k)$, then

$$W_{1,1} \sim \mathcal{W}_h \left(\alpha = \frac{n}{2}, \frac{1}{2} \Sigma_{1,1}^{-1} \right).$$

This property allows to compute the marginal distribution of the elements on the diagonal of W; for example, if $k = 2$ and $A = (1, 0)$, then $Y = w_{1,1} \sim \mathcal{G}(\alpha = n/2, \sigma_{1,1}^{-1}/2)$, where $\sigma_{1,1}$ is the first element of the diagonal of Σ. It follows that $w_{1,1}/\sigma_{1,1} \sim \chi^2(n)$. Then,

$$\mathrm{E}(w_{1,1}) = n\sigma_{1,1}, \quad \mathrm{Var}(w_{1,1}) = 2n\sigma_{1,1}^2.$$

More generally, it can be proved that

$$\mathrm{Var}(w_{i,j}) = n(\sigma_{i,j}^2 + \sigma_{i,i}\sigma_{j,j}), \quad \mathrm{Cov}(w_{i,j}, w_{l,m}) = n(\sigma_{i,l}\sigma_{j,m} + \sigma_{i,m}\sigma_{j,l}).$$

If $W \sim \mathcal{W}_k(\alpha = n/2, B = \Sigma^{-1}/2)$, then $V = W^{-1}$ has an *Inverse-Wishart* distribution and

$$\mathrm{E}(V) = \mathrm{E}(W^{-1}) = \left(\alpha - \frac{k+1}{2} \right)^{-1} B = \frac{1}{n-k-1} \Sigma^{-1}.$$

Multivariate Student-t distribution

If Y is a p-variate random vector with $Y \sim \mathcal{N}_p(0, \Sigma)$ and $U \sim \chi^2(k)$, with Y and U independent, then $X = \frac{Y}{\sqrt{U/k}} + \mu$ has a p-variate Student-t distribution, with parameters (μ, Σ) and $k > 0$ degrees of freedom, with density

$$f(x) = c\, [1 - \frac{1}{k}\, (x - \mu)' \Sigma^{-1} (x - \mu)]^{-(k+p)/2} \quad , x \in \mathbb{R}^p \, ,$$

where $c = \Gamma((k+p)/2)/(\Gamma(k/2)\pi^{p/2}k^{p/2}|\Sigma|^{1/2}$. We write $X \sim \mathcal{T}(\mu, \Sigma, k)$. For $p = 1$ it reduces to the univariate Student-t distribution. We have

$$\mathrm{E}(X) = \mu \text{ if } k > 1$$
$$\mathrm{Var}(X) = \Sigma \frac{k}{k-2} \text{ if } k > 2.$$

Multivariate Normal-Gamma distribution

Let (X, Y) be a random vector, with $X|Y = y \sim \mathcal{N}_m(\mu, (N_0 y)^{-1})$, and $Y \sim Ga(a, b)$. Then we say that (X, Y) has a Normal-Gamma density with parameters (μ, N_0^{-1}, a, b), denoted as $(X, Y) \sim \mathcal{NG}(\mu, N_0^{-1}, a, b)$.

The marginal density of X is a multivariate Student-t, $X \sim \mathcal{T}(\mu, (N_0 \frac{a}{b})^{-1}, 2a)$, so that $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = N_0^{-1} b/(a-1)$.

B

Matrix algebra: Singular Value Decomposition

Let M be a $p \times q$ matrix and let $r = \min\{p, q\}$. The singular value decomposition (SVD) of M consists in a triple of matrices (U, D, V) with the following properties:

(i) U is a $p \times p$ orthogonal matrix;
(ii) V is a $q \times q$ orthogonal matrix;
(iii) D is a $p \times q$ matrix with entries $D_{ij} = 0$ for $i \neq j$;
(iv) $UDV' = M$.

If M is a square matrix, D is a diagonal matrix. If, in addition, M is nonnegative definite, then $D_{ii} \geq 0$ for every i. In this case one can define a diagonal matrix S by setting $S_{ii} = \sqrt{D_{ii}}$, so that $M = US^2V'$. It can be shown that if M is also symmetric, such as, for example, a variance matrix, then $M = US^2U'$. M is invertible if and only if $S_{ii} > 0$ for every i. The SVD has many applications in numerical linear algebra. For example, it can be used to compute a square root[1]of a variance matrix M, i.e., a square matrix N such that $M = N'N$. In fact, if $M = US^2U'$, it is enough to set $N = SU'$. The inverse of M can also be easily computed from its SVD, provided M is invertible. In fact, it is immediate to verify that $M^{-1} = US^{-2}U'$. Note also that $S^{-1}U'$ is a square root of M^{-1}. More generally, for a noninvertible M, a generalized inverse M^- is a matrix with the property that $MM^-M = M$. The generalized inverse of a variance matrix can be found by defining the diagonal matrix S^-,

$$S^-_{ii} = \begin{cases} S^{-1}_{ii} & \text{if } S_{ii} > 0, \\ 0 & \text{if } S_{ii} = 0, \end{cases}$$

and setting $M^- = U(S^-)^2V'$.

In package dlm the SVD is used extensively to compute filtering and smoothing variances in a numerically stable way, see Wang et al. (1992) and

[1] Note that our definition of matrix square root differs slightly from the most common one based on the relation $M = M^{\frac{1}{2}}M^{\frac{1}{2}}$.

Zhang and Li (1996) for a complete discussion of the algorithms used. For example, consider the filtering recursion used to compute C_t. The calculation can be broken down in three steps:

(i) compute $R_t = G_t C_{t-1} G_t' + W_t$;
(ii) compute[2] $C_t^{-1} = F_t' V_t^{-1} F_t + R_t^{-1}$;
(iii) invert C_t^{-1}.

Suppose $(U_{C,t-1}, S_{C,t-1})$ are the components of the SVD of C_{t-1}, so that $C_{t-1} = U_{C,t-1} S_{C,t-1}^2 U_{C,t-1}'$ and $N_{W,t}$ is a square root of W_t. Define the $2p \times p$ partitioned matrix

$$M = \begin{bmatrix} S_{C,t-1} U_{C,t-1}' G_t' \\ N_{W,t} \end{bmatrix}$$

and let (U, D, V) be its SVD. The crossproduct of M is

$$\begin{aligned} M'M &= V D U' U D V' = V D^2 V' \\ &= G_t U_{C,t-1} S_{C,t-1}^2 U_{C,t-1}' G_t' + N_{W,t}' N_{W,t} \\ &= G_t C_{t-1} G_t' + W_t = R_t. \end{aligned}$$

Since V is orthogonal and D^2 is diagonal, (V, D^2, V) is the SVD of R_t and we can set $U_{R,t} = V$ and $S_{R,t} = D$. Now consider a square root $N_{V,t}$ of V_t^{-1}, define the $(m+p) \times p$ partitioned matrix

$$M = \begin{bmatrix} N_{V,t} F_t U_{R,t} \\ S_{R,t}^{-1} \end{bmatrix}$$

and let (U, D, V) be its SVD. The crossproduct of M is

$$\begin{aligned} M'M &= V D U' U D V' = V D^2 V' \\ &= U_{R,t}' F_t' N_{V,t}' N_{V,t} F_t U_{R,t} + S_{R,t}^{-2}. \end{aligned}$$

Premultiplying by $U_{R,t}$ and postmultiplying by $U_{R,t}'$ we obtain

$$\begin{aligned} U_{R,t} M' M U_{R,t}' &= U_{R,t} V D^2 V' U_{R,t}' \\ &= U_{R,t} U_{R,t}' F_t' N_{V,t}' N_{V,t} F_t U_{R,t} U_{R,t}' + U_{R,t} S_{R,t}^{-2} U_{R,t}' \\ &= F_t' V_t^{-1} F_t + R_t^{-1} = C_t^{-1}. \end{aligned}$$

[2] The expression for C_t^{-1} follows from the formulas in Proposition 2.2 by applying the general matrix equality

$$(A + BEB')^{-1} = A^{-1} - A^{-1} B (B' A^{-1} B + E^{-1})^{-1} B' A^{-1},$$

in which A and E are nonsingular matrices of orders m and n and B is a $m \times n$ matrix.

Since $U_{R,t}V$ is the product of two orthogonal matrices of order p, it is itself an orthogonal matrix. Therefore $U_{R,t}VD^2V'U'_{R,t}$ gives the SVD of C_t^{-1}: the "U" matrix of the SVD of C_t^{-1} is $U_{R,t}V$ and the "S" matrix is D. It follows that for the SVD of C_t we have $U_{C,t} = U_{R,t}V$ and $S_{C,t} = D^{-1}$.

Index

References

Akaike, H. (1974a). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes, *Annals of the Institute of Statistical Mathematics* **26**: 363–387.

Akaike, H. (1974b). Stochastic theory of minimal realization, *IEEE Trans. on Automatic Control* **19**: 667–674.

Amisano, G. and Giannini, C. (1997). *Topics in Structural VAR Econometrics*, 2nd edn, Springer, Berlin.

Anderson, B. and Moore, J. (1979). *Optimal Filtering*, Prentice-Hall, Englewood Cliffs.

Aoki, M. (1987). *State Space Modeling of Time Series*, Springer Verlag, New York.

Barndorff-Nielsen, O., Cox, D. and Klüppelberg, C. (eds) (2001). *Complex Stochastic Systems*, Chapman & Hall, London.

Barra, J. and Herbach, L. H. (1981). *Mathematical Basis of Statistics*, Academic Press, New York.

Bawens, L., Lubrano, M. and Richard, J.-F. (1999). *Bayesian inference in Dynamic Econometric Models*, Oxford University Press, New York.

Bayes, T. (1763). *An essay towards solving a problem in the doctrine of chances.* Published posthumously in *Phil. Trans. Roy. Stat. Soc. London*, **53**, 370–418 and **54**, 296–325. Reprinted in *Biometrika* **45** (1958), 293–315, with a biographical note by G.A. Barnard. Reproduced in Press (1989), 185–217.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer, Berlin.

Bernardo, J. (1979a). Expected information as expected utility, *Annals of Statistics* pp. 686–690.

Bernardo, J. (1979b). Reference posterior distributions for Bayesian inference (with discussion), *Journal of the Royal Statistical Society, Series B* pp. 113–147.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory*, Wiley, Chichester.

Berndt, R. (1991). *The Practice of Econometrics*, Addison-Wesley.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**: 307–327.

Box, G., Jenkins, G. and Reinsel, G. (2008). *Time Series Analysis: Forecasting and Control*, 4th edn, Wiley, New York.

Brandt, P. (2008). *MSBVAR: Markov-Switching Bayesian Vector Autoregression Models*. R package version 0.3.2.
URL: *http://www.utdallas.edu/~pbrandt/*

Brown, P., Le, N. and Zidek, J. (1994). Inference for a covariance matrix, *in* P. Freeman and e. A.F.M. Smith (eds), *Aspects of Uncertainty: A Tribute to D. V. Lindley*, Wiley, Chichester, pp. 77–92.

Caines, P. (1988). *Linear Stochastic Systems*, Wiley, New York.

Campbell, J., Lo, A. and MacKinley, A. (1996). *The Econometrics of Financial Markets*, Princeton University Press, Princeton.

Canova, F. (2007). *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton.

Cappé, O., Godsill, S. and Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo, *Proceedings of the IEEE* **95**: 899–924.

Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*, Springer, New York.

Carmona, R. A. (2004). *Statistical analysis of financial data in S-plus*, Springer-Verlag, New York.

Caron, F., Davy, M., A., D., Duflos, E. and Vanheeghe, P. (2008). Bayesian inference for linear dynamic models with dirichlet process mixtures, *IEEE Transactions on Signal Processing* **56**: 71–84.

Carter, C. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika* **81**: 541–553.

Chang, Y., Miller, J. and Park, J. (2005). Extracting a common stochastic trend: Theories with some applications, *Technical report*, Rice University.
URL: *http://ideas.repec.org/p/ecl/riceco/2005-06.html.*

Chatfield, C. (2004). *The Analysis of Time Series*, 6th edn, CRC-Chapman & Hall, London.

Cifarelli, D. and Muliere, P. (1989). *Statistica Bayesiana*, Iuculano Editore, Pavia. (In Italian).

Consonni, G. and Veronese, P. (2001). Conditionally reducible natural exponential families and enriched conjugate priors, *Scandinavian Journal of Statistics* **28**: 377–406.

Consonni, G. and Veronese, P. (2003). Enriched conjugate and reference priors for the wishart family on symmetric cones, *Annals of Statistics* **31**: 1491–1516.

Cowell, R., Dawid, P., Lauritzen, S. and Spiegelhalter, D. (1999). *Probabilistic networks and expert systems*, Springer-Verlag, New York.

D'Agostino, R. and Stephens, M. (eds) (1986). *Goodness-of-fit Techniques*, Dekker, New York.

Dalal, S. and Hall, W. (1983). Approximating priors by mixtures of conjugate priors, *J. Roy. Statist. Soc. Ser. B* **45**: 278–286.

Dawid, A. (1981). Some matrix-variate distribution theory: Notational considerations and a bayesian application, *Biometrika* **68**: 265–274.

Dawid, A. and Lauritzen, S. (1993). Hyper-markov laws in the statistical analysis of decomposable graphical models, *Ann. Statist.* **21**: 1272–1317.

de Finetti, B. (1970a). *Teoria della probabilità I*, Einaudi, Torino. English translation as *Theory of Probability I* in 1974, Wiley, Chichester.

de Finetti, B. (1970b). *Teoria della probabilità II*, Einaudi, Torino. English translation as *Theory of Probability II* in 1975, Wiley, Chichester.

De Finetti, B. (1972). *Probability, Induction and Statistics*, Wiley, Chichester.

DeGroot, M. (1970). *Optimal Statistical Decisions*, McGraw Hill, New York.

Del Moral, P. (2004). *Feyman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Springer-Verlag, New York.

Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion, *in* J. Bernardo, M. deGroot, D. Lindley and A. Smith (eds), *Bayesian Statistics 2*, Elsevier Science Publishers B.V. (North Holland), pp. 133–156.

Diebold, F. and Li, C. (2006). Forecasting the term structure of government bond yields, *Journal of Econometrics* **130**: 337–364.

Diebold, F., Rudebuschb, G. and Aruoba, S. (2006). The macroeconomy and the yield curve: A dynamic latent factor approach, *Journal of Econometrics* **131**: 309–338.

Doan, T., Litterman, R. and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions, *Econometric Reviews* **3**: 1–144.

Doucet, A., De Freitas, N. and Gordon, N. (eds) (2001). *Sequential Monte Carlo Methods in Practice*, Springer, New York.

Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, *Econometrica* **50**: 987–1008.

Engle, R. and Granger, C. (1987). Co-integration and error correction: Representation, estimation, and testing, *Econometrica* **55**: 251–276.

Fearnhead, P. (2002). Markov chain Monte Carlo, sufficient statistics, and particle filter, *Journal of Computational and Graphical Statistics* **11**: 848–862.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation, *Review of Economics and Statistics* **82**: 540–552.

Frühwirth-Schnatter, S. and Kaufmann, S. (2008). Model-based clustering of multiple time series, *Journal of Business and Economic Statistics* **26**: 78–89.

Früwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *Journal of Time Series Analysis* **15**: 183–202.

Früwirth-Schnatter, S. and Wagner, H. (2008). Stochastic model specification search for Gaussian and non-Gaussian state space models, *IFAS Research Papers, 2008-36* .

Gamerman, D. and Migon, H. (1993). Dynamic hierarchical models, *Journal of the Royal Statistical Society, Series B* **55**: 629–642.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2004). *Bayesian Data Analysis*, 2nd edn, Chapman & Hall/CRC, Boca Raton.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*, Wiley, Hoboken.

Gilbert, P. (2008). *Brief User's Guide: Dynamic Systems Estimation.*
URL: `http:www.bank-banque-canada.ca/pgilbert/.`

Gilks, W. and Berzuini, C. (2001). Following a moving target – Monte Carlo inference for dynamic Bayesian models, *Journal of the Royal Statistical Society, Series B* **63**: 127–146.

Gilks, W., Best, N. and Tan, K. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling (Corr: 97V46 p541-542 with R.M. Neal), *Applied Statistics* **44**: 455–472.

Gilks, W. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling, *Applied Statistics* **41**: 337–348.

Gourieroux, C. and Monfort, A. (1997). *Time Series and Dynamic Models*, Cambridge University Press, Cambridge.

Granger, C. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* **16**: 150–161.

Gross, J. (n.d.). *Nortest: Tests for Normality.* R package version 1.0.

Hannan, E. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*, Wiley, New York.

Harrison, P. and Stevens, C. (1976). Bayesian forecasting (with discussion), *Journal of the Royal Statistical Society, Series B* **38**: 205–247.

Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman filter*, Cambridge University Press, Cambridge.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**: 97–109.

Hutchinson, C. (1984). The Kalman filter applied to aerospace and electronic systems, *Aerospace and Electronic Systems, IEEE Transactions on Aerospace and Electronic Systems* **AES-20**: 500–504.

Hyndman, R. (2008). *Forecast: Forecasting Functions for Time Series.* R package version 1.14.
URL: *http://www.robhyndman.com/Rlibrary/forecast/.*

Hyndman, R. (n.d.). Time Series Data Library.
URL: *http://www.robjhyndman.com/TSDL.*

Hyndman, R., Koehler, A., Ord, J. and Snyder, R. (2008). *Forecasting with Exponential smoothing*, Springer, Berlin.

Jacquier, E., Polson, N. and Rossi, P. (1994). Bayesian analysis of stochastic volatility models (with discussion), *Journal of Business and Economic Statistics* **12**: 371–417.

Jazwinski, A. (1970). *Stochastic Processes and Filtering Theory*, Academic Press, New York.

Jeffreys, H. (1998). *Theory of Probability*, 3rd edn, Oxford University Press, New York.

Johannes, M. and Polson, N. (2009). MCMC methods for continuous-time financial econometrics, *in* Y. Ait-Sahalia and L. Hansen (eds), *Handbook of Financial Econometrics*, Elsevier. (To appear).

Kalman, R. (1960). A new approach to linear filtering and prediction problems, *Trans. of the AMSE - Journal of Basic Engineering (Series D)* **82**: 35–45.

Kalman, R. (1961). On the general theory of control systems, *Proc. IFAC Congr Ist.* **1**: 481–491.

Kalman, R. (1968). Contributions to the theory of optimal control, *Bol. Soc. Mat. Mexicana* **5**: 558–563.

Kalman, R. and Bucy, R. (1963). New results in linear filtering and prediction theory, *Trans. of the AMSE – Journal of Basic Engineering (Series D)* **83**: 95–108.

Kalman, R., Ho, Y. and Narenda, K. (1963). Controllability of linear dynamical systems, *in* J. Lasalle and J. Diaz (eds), *Contributions to Differential Equations*, Vol. 1, Wiley Interscience.

Kim, C.-J. and Nelson, C. (1999). *State Space Models with Regime Switching*, MIT Press, Cambridge.

Kolmogorov, A. (1941). Interpolation and extrapolation of stationary random sequences, *Bull. Moscow University, Ser. Math.* **5**.

Künsch, H. (2001). State space and hidden Markov models, *in* O. Barndorff-Nielsen, D. Cox and C. Klüppelberg (eds), *Complex stochastic systems*, Chapman & Hall/CRC, Boca Raton, pp. 109–173.

Kuttner, K. (1994). Estimating potential output as a latent variable, *Journal of Business and Economic Statistics* **12**: 361–68.

Landim, F. and Gamerman, D. (2000). Dynamic hierarchical models; an extension to matrix-variate observations, *Computational Statistics and Data Analysis* **35**: 11–42.

Laplace, P. (1814). *Essai Philosophique sur les Probabilitiès*, Courcier, Paris. The 5th edn (1825) was the last revised by Laplace. English translation in 1952 as *Philosophical Essay on Probabilities*, Dover, New York.

Lau, J. and So, M. (2008). Bayesian mixture of autoregressive models, *Computational Statistics and Data Analysis* **53**: 38–60.

Lauritzen, S. (1981). Time series analysis in 1880: A discussion of contributions made by T.N. Thiele, *International Statist. Review* **49**: 319–331.

Lauritzen, S. (1996). *Graphical Models*, Oxford University Press, Oxford.

Lindley, D. (1978). The Bayesian approach (with discussion), *Scandinavian Journal of Statistics* **5**: 1–26.

Lindley, D. and Smith, A. (1972). Bayes estimates for the linear model, *Journal of the Royal Statistical Society, Series B* **34**: 1–41.

Lipster, R. and Shiryayev, A. (1972). Statistics of conditionally Gaussian random sequences, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Univ. California Press, Berkeley.

Litterman, R. (1986). Forecasting with Bayesian vector autoregressions – five years of experience, *Journal of Business and Economic Statistics* **4**: 25–38.

Liu, J. (2001). *Monte Carlo Strategies in Scientific Computing*, Springer, New York.

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering, *in* A. Doucet, N. De Freitas and N. Gordon (eds), *Sequential Monte Carlo Methods in Practice*, Springer, New York.

Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models, *Biometrika* **65**: 297–303.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Maybeck, P. (1979). *Stochastic Models, Estimation and Control*, Vol. 1 and 2, Academic Press, New York.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1091.

Migon, H., Gamerman, D., Lopez, H. and Ferreira, M. (2005). Bayesian dynamic models, *in* D. Day and C. Rao (eds), *Handbook of Statistics*, Vol. 25, Elsevier B.V., chapter 19, pp. 553–588.

Morf, M. and Kailath, T. (1975). Square-root algorithms for least-squares estimation, *IEEE Trans. Automatic Control* **AC-20**: 487–497.

Muliere, P. (1984). Modelli lineari dinamici, *Studi Statistici (8)*, Istituto di Metodi Quantitativi, Bocconi University, Milan. (In Italian).

O'Hagan, A. (1994). *Bayesian Inference*, Kendall's Advanced Theory of Statistics, 2B, Edward Arnold, London.

Oshman, Y. and Bar-Itzhack, I. (1986). Square root filtering via covariance and information eigenfactors, *Automatica* **22**: 599–604.

Petris, G. and Tardella, L. (2003). A geometric approach to transdimensional Markov chain Monte Carlo, *Canadian Journal of Statistics* **31**: 469–482.

Pfaff, B. (2008a). *Analysis of Integrated and Cointegrated Time Series with R*, 2nd edn, Springer, New York.

Pfaff, B. (2008b). Var, svar and svec models: Implementation within R package vars, *Journal of Statistical Software* **27**.
URL: *http://www.jstatsoft.org/v27/i04/.*

Pitt, M. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association* **94**: 590–599.

Plackett, R. (1950). Some theorems in least squares, *Biometrika* **37**: 149–157.

Poirier, D. (1995). *Intermediate Statistics and Econometrics: a Comparative Approach*, MIT Press, Cambridge.

Pole, A., West, M. and Harrison, J. (n.d.). *Applied Bayesian forecasting and time series analysis*, Chapman & Hall, New York.

Prakasa Rao, B. (1999). *Statistical Inference for Diffusion Type Processes*, Oxford University Press, New York.

Rabiner, L. and Juang, B. (1993). *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs.

Rajaratnam, B., Massam, H. and Carvalho, C. (2008). Flexible covariance estimation in graphical Gaussian models, *Annals of Statistics* **36**: 2818–2849.

Reinsel, G. (1997). *Elements of Multivariate Time Series Analysis*, 2nd edn, Springer-Verlag, New York.

Robert, C. (2001). *The Bayesian Choice*, 2nd edn, Springer-Verlag, New York.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd edn, Springer, New York.

Rydén, T. and Titterington, D. (1998). Computational Bayesian analysis of hidden Markov models, *J. Comput. Graph. Statist.* **7**: 194–211.

Savage, L. (1954). *The Foundations of Statistics*, Wiley, New York.

Schervish, M. (1995). *Theory of Statistics*, Springer-Verlag, New York.

Shephard, N. (1994). Partial non-Gaussian state space models, *Biometrika* **81**: 115–131.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility, *in* D. Cox, D. Hinkley and O. Barndorff-Nielsen (eds), *Time Series Models with Econometric, Finance and other Applications*, Chapman and Hall, London, pp. 1–67.

Shumway, R. and Stoffer, D. (2000). *Time Series Analysis and its Applications*, Springer-Verlag, New York.

Sokal, A. (1989). *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*, Cours de Troisiéme Cycle de la Physique en Suisse Romande, Lausanne.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *in* J. Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*, University of California Press, Berkeley, pp. 197–206.

Storvik, G. (2002). Particle filters for State-Space models with the presence of unknown static parameters, *IEEE Transactions on Signal Processing* **50**: 281–289.

Theil, H. (1966). *Applied Economic Forecasting*, North Holland, Amsterdam.

Thiele, T. (1880). Om anvendelse af mindste kvadraters methode i nogle tilflde, hvor en komplikation af visse slags uensartede tilfldige fejlkilder giver fejlene en “systematisk” karakter, *Det Kongelige Danske Videnskabernes Selskabs Skrifter – Naturvidenskabelig og Mathematisk Afdeling* pp. 381–408. English Transalation in: Thiele: Pioneer in Statistics, S. L. Lauritzen, Oxford University Press (2002).

Tierney, L. (1994). Markov chain for exploring posterior distributions (with discussion), *Annals of Statistics* **22**: 1701–1786.

Uhlig, H. (1994). On singular Wishart and singular multivariate beta distributions, *Annals of Statistics* **22**: 395–405.

Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S*, 4th edn, Springer-Verlag, New York.

Wang, L., Liber, G. and Manneback, P. (1992). Kalman filter algorithm based on singular value decomposition, *Proc. of the 31st Conf. on Decision and Control*, pp. 1224–1229.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd edn, Springer, New York.

West, M., Harrison, J. and Migon, H. (1985). Dynamic generalized linear models and Bayesian forecasting, *Journal of the American Statistical Association* **80**: 73–83.

Wiener, N. (1949). *The Extrapolation, Intepolation and Smoothing of Stationary Time Series*, Wiley, New York.

Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*, Almquist and Wiksell, Uppsala.

Wuertz, D. (2008). *fBasics: Rmetrics – Markets and Basic Statistics*. R package version 280.74.
URL: *http://www.rmetrics.org*.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Zhang, Y. and Li, R. (1996). Fixed-interval smoothing algorithm based on singular value decomposition, *Proceedings of the 1996 IEEE International Conference on Control Applications*, pp. 916–921.