# A

# Appendix

## A.1 Summary of Probability Distributions

Here we briefly summarize all probability distributions used in this book. For an exhaustive review of probability distributions see Johnson et al. (1993, 1994, 1995). The parameterization of the densities closely follows Bernardo and Smith (1994).

*Notation*

$Y$ denotes a univariate random variable; $y$ refers to a realization of $Y$. $f(y)$ refers to the density of the probability distribution of $Y$ with respect to an appropriate measure (Lebesgue measure or counting measure, depending on the context). $\mathrm{E}(Y)$ denotes the expectation of $Y$, whereas $\mathrm{Var}(Y)$ denotes the variance of $Y$. For multivariate random variables $\mathbf{Y}$ and $\mathbf{y}$ are used, whereas $Y_j$ and $y_j$ refer to a certain element of $\mathbf{Y}$ and $\mathbf{y}$, respectively. $\mathrm{E}(\mathbf{Y})$ denotes the mean vector; $\mathrm{Var}(Y)$ denotes the variance–covariance matrix of $\mathbf{Y}$.

### A.1.1 The Beta Distribution

The Beta distribution $Y \sim \mathcal{B}(\alpha, \beta)$ with $\alpha, \beta \in \Re^+$, is a univariate distribution defined on the unit interval $y \in [0, 1]$. For mixture models the Beta distribution appears mainly as a posterior distribution of an unknown probability. Density, mean, and variance are given by

$$f_B(y; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1-y)^{\beta-1}, \tag{A.1}$$

$$\mathrm{E}(Y) = \frac{\alpha}{\alpha+\beta}, \qquad \mathrm{Var}(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)},$$

where $B(\alpha, \beta)$ is the Beta function:

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{A.2}$$

For a proper density both $\alpha$ and $\beta$ need to be positive. If $0 < \alpha < 1$, the density is unbounded at 0; if $0 < \beta < 1$, the density is unbounded at 1. If $\alpha = 1$ and $\beta = 1$, the density is equal to the density of a uniform distribution. If $\alpha = 1$ and $\beta > 1$, the mode of the density lies at 0; if $\alpha > 1$ and $\beta = 1$, the mode of the density lies at 1. For $\alpha, \beta > 1$ the mode of the density lies in the interior of $[0, 1]$ at $(\alpha - 1)/(\alpha + \beta - 2)$.

### A.1.2 The Binomial Distribution

The binomial distribution, $Y \sim \text{BiNom}\,(n, p)$, with $n = 1, 2, \ldots$ and $p \in [0, 1]$, is frequently chosen to model the outcome of repeated measurements. The density is defined for $y \in \{0, 1, 2, \ldots, n\}$,

$$f_{BN}(y; n, p) = \binom{n}{y} p^y (1-p)^{n-y}, \tag{A.3}$$

and mean and variance are given by

$$\mathrm{E}(Y) = np, \qquad \mathrm{Var}(Y) = np(1-p). \tag{A.4}$$

### A.1.3 The Dirichlet Distribution

The Dirichlet distribution, $\mathbf{Y} \sim \mathcal{D}\,(\alpha_1, \ldots, \alpha_K)$, is a standard choice in the context of modeling an unknown discrete probability distribution $\mathbf{Y} = (Y_1, \ldots, Y_K)$ where $\sum_{j=1}^{K} Y_j = 1$ and therefore is of great importance for mixture and switching models. The Dirichlet distribution is a distribution on the unit simplex $\mathcal{E}_K \subset (\Re^+)^K$, defined by the following constraint,

$$\mathcal{E}_K = \left\{ \mathbf{y} = (y_1, \ldots, y_K) \in (\Re^+)^K : \sum_{j=1}^{K} y_j = 1 \right\}.$$

The density is given by

$$f_D(\mathbf{y}; \alpha_1, \ldots, \alpha_K) = y_1^{\alpha_1 - 1} \cdots y_K^{\alpha_K - 1} c, \qquad c = \frac{\Gamma(\Sigma_\alpha)}{\prod_{i=1}^{K} \Gamma(\alpha_i)}, \tag{A.5}$$

where $\Sigma_\alpha = \sum_{j=1}^{K} \alpha_j$. As each $y_j$ is completely determined given the remaining elements of $\mathbf{y}$, one element $y_j$ has to be substituted in (A.5) by $1 - \sum_{k \neq j} y_k$ to obtain a proper density. The density is proper if all $\alpha_j$ are positive; the density is bounded if all $\alpha_j$ are greater than or equal to 1. The density of a $\mathcal{D}\,(\alpha_1, \ldots, \alpha_K)$-distribution is improper whenever any $\alpha_j$ is equal to 0.

For $\alpha_1 = \cdots = \alpha_K = 1$ the uniform distribution over the unit simplex results. For $K = 2$ the Dirichlet distribution is equal to $Y_1 \sim \mathcal{B}\,(\alpha_1, \alpha_2)$.

The marginal distribution of $Y_j$ is a $\mathcal{B}(\alpha_j, \varSigma_\alpha - \alpha_j)$-distribution, therefore mean and variance of $Y_j$ are given by

$$\mathrm{E}(Y_j) = \frac{\alpha_j}{\varSigma_\alpha}, \qquad \mathrm{Var}(Y_j) = \frac{\alpha_j(\varSigma_\alpha - \alpha_j)}{\varSigma_\alpha^2(\varSigma_\alpha + 1)}. \tag{A.6}$$

The mode of the marginal density of $Y_j$ is given by $(\alpha_j - 1)/(\varSigma_\alpha - K)$.

The easiest way to sample a random variable $\mathbf{Y} = (Y_1, \ldots, Y_K)$ from the $\mathcal{D}(\alpha_1, \ldots, \alpha_K)$-distribution is to sample $K$ independent random variables $Y_1^\star$, $\ldots$, $Y_K^\star$ from the following Gamma distributions, $Y_j^\star \sim \mathcal{G}(\alpha_j, 1)$, $j = 1, \ldots, K$, and to normalize: $Y_j = Y_j^\star / \sum_{k=1}^K Y_k^\star$.

### A.1.4 The Exponential Distribution

The exponential distribution, $Y \sim \mathcal{E}(\beta)$ with $\beta > 0$, is a univariate distribution defined on the positive real line $y \geq 0$ and is mainly used as a sampling distribution in the context of finite mixture models. There exist different ways to parameterize this distribution and we follow Bernardo and Smith (1994), by defining the density as

$$f_E(y; \beta) = \beta e^{-\beta y}. \tag{A.7}$$

Mean and variance are given by

$$\mathrm{E}(Y) = \mathrm{Var}(Y) = 1/\beta. \tag{A.8}$$

In this parameterization, the exponential distribution is equal to the $\mathcal{G}(1, \beta)$-distribution.

### A.1.5 The F-Distribution

The F-distribution, $Y \sim \mathrm{F}(\alpha_1, \alpha_2)$, is a univariate distribution defined on the positive real line $y \geq 0$. For finite mixture models based on the normal distribution, it appears as part of modeling the prior of heterogeneous variances. For $\alpha_1 > 0, \alpha_2 > 0$ the density is given by

$$f_F(y; \alpha_1, \alpha_2) = \frac{\varGamma((\alpha_1 + \alpha_2)/2)\alpha_1}{\varGamma(\alpha_1/2)\varGamma(\alpha_2/2)\alpha_2} \left(\frac{\alpha_1}{\alpha_2} y\right)^{\alpha_1/2-1} \left(1 + \frac{\alpha_1}{\alpha_2} y\right)^{-(\alpha_1+\alpha_2)/2} \tag{A.9}$$

For $y \to \infty$, the density behaves as $y^{-\alpha_2/2-1}$, whereas for $y \to 0$, the density behaves as $y^{\alpha_1/2-1}$. The density is an improper density, whenever $\alpha_1$ or $\alpha_2$ are 0. If $\alpha_1 > 0$ and $\alpha_2 > 0$, the density is proper, but unbounded at $y = 0$ for $0 < \alpha_1 < 2$.

### Relation to the $\chi^2$-Distribution

Assume that $Y_1$ and $Y_2$ are independent random variables, each following a $\chi^2$-distribution: $Y_1 \sim \chi^2_{\alpha_1}$ and $Y_2 \sim \chi^2_{\alpha_2}$. Then

$$Y = \frac{Y_1/\alpha_1}{Y_2/\alpha_2} \sim \mathrm{F}\left(\alpha_1, \alpha_2\right). \tag{A.10}$$

## A.1.6 The Gamma Distribution

The Gamma distribution, $Y \sim \mathcal{G}\left(\alpha, \beta\right)$, is a univariate distribution defined on the positive real line $y \geq 0$. It is encountered in finite mixture models as a posterior density within a Bayesian analysis for certain nonnormal models, in particular for observations from the Poisson and the exponential distribution. There exist various ways to parameterize this distribution and we follow Bernardo and Smith (1994), where the density is given by

$$f_G(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}. \tag{A.11}$$

The density is an improper density whenever $\beta = 0$. If $\beta > 0$, the density is proper, but unbounded at $y = 0$ for $0 < \alpha < 1$. For $\alpha > 0, \beta > 0$, mean and variance are given by

$$\mathrm{E}(Y) = \frac{\alpha}{\beta}, \qquad \mathrm{Var}(Y) = \frac{\alpha}{\beta^2}. \tag{A.12}$$

The mode is given by $(\alpha - 1)/\beta$. If $Y \sim \mathcal{G}\left(\alpha, \beta\right)$, then $\omega Y \sim \mathcal{G}\left(\alpha, \beta/\omega\right)$.

### The $\chi^2$-Distribution

The $\mathcal{G}\left(\nu/2, 1/2\right)$-distribution with $\nu = 1, 2, \ldots$ is called the $\chi^2_\nu$-distribution, which is the distribution of the sum of squares of $\nu$ independent standard normal random variables, $Y = \sum_{j=1}^{\nu} Y_j^2$, $Y_j \sim \mathcal{N}\left(0, 1\right)$. Therefore $\mathrm{E}(Y) = \nu$, $\mathrm{Var}(Y) = 2\nu$.

### Inverted Gamma Distribution

A random variable $Y$ follows an inverted Gamma distribution, $Y \sim \mathcal{G}^{-1}\left(\alpha, \beta\right)$, if $Y^{-1}$ has a Gamma distribution: $Y^{-1} \sim \mathcal{G}\left(\alpha, \beta\right)$. The inverted Gamma distribution often appears as a posterior distribution of an unknown variance within a Bayesian analysis of finite mixture models based on the normal distribution. The density is given by

$$f_{IG}(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{y}\right)^{\alpha+1} e^{-\beta/y}. \tag{A.13}$$

$Y$ has finite expectation, iff $\alpha > 1$, and finite variance, iff $\alpha > 2$:

$$\mathrm{E}(Y) = \frac{\beta}{\alpha - 1}, \qquad \mathrm{Var}(Y) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}. \tag{A.14}$$

The mode is given by $\beta/(\alpha+1)$. $\mathcal{G}^{-1}(0,0)$ has an improper density proportional to $1/y$. If $Y \sim \mathcal{G}^{-1}(\alpha,\beta)$, then $\omega Y \sim \mathcal{G}^{-1}(\alpha,\omega\beta)$.

### A.1.7 The Geometric Distribution

The geometric distribution is defined for $y \in \{1,2,\ldots\}$. Density, mean and variance are given by

$$f(y;p) = p(1-p)^{y-1}, \tag{A.15}$$
$$\mathrm{E}(Y) = \frac{1}{p}, \qquad \mathrm{Var}(Y) = \frac{1-p}{p^2}\ .$$

### A.1.8 The Multinomial Distribution

Let $\mathbf{Y}$ be a categorical random variable with $D$ categories coded as 1 to $D$. $\mathbf{Y} \sim \mathrm{MulNom}\,(n, p_1, \ldots, p_D)$, iff for all $\mathbf{y} = (y_1, \ldots, y_D)$, where $\sum_{j=1}^D y_j = n$ and $y_j \geq 0$,

$$\Pr(\mathbf{Y} = \mathbf{y}) = f_{MN}(\mathbf{y}; n, p_1, \ldots, p_D) = \binom{n}{y_1 \ldots y_D} p_1^{y_1} \ldots p_D^{y_D}, \tag{A.16}$$

with the general binomial coefficient being defined as:

$$\binom{n}{y_1 \ldots y_D} = \frac{n!}{y_1! \ldots y_D!}$$

The mean and variance of $Y_j$ are given by

$$\mathrm{E}(Y_j) = np_j, \qquad \mathrm{Var}(Y_j) = np_j(1-p_j). \tag{A.17}$$

The binomial distribution is that special case of the multinomial distribution where $D = 2$.

### A.1.9 The Negative Binomial Distribution

The negative binomial distribution is defined for $y \in \{0,1,2,\ldots\}$: $Y \sim \mathrm{NegBin}\,(\alpha,\beta)$. For $\alpha > 0, \beta > 0$ density, mean, and variance of $Y$ are given by

$$f_{NB}(y;\alpha,\beta) = \binom{\alpha+y-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{y},$$
$$\mathrm{E}(Y) = \frac{\alpha}{\beta}, \qquad \mathrm{Var}(Y) = \frac{\alpha}{\beta^2}(\beta+1). \tag{A.18}$$

The $\mathrm{NegBin}\,(\alpha,\beta)$-distribution is an infinite mixture of $\mathcal{P}\,(\mu)$-distributions, where $\mu \sim \mathcal{G}\,(\alpha,\beta)$:

$$f_{NB}(y;\alpha,\beta) = \int f_P(y;\mu) f_G(\mu;\alpha,\beta)\,d\mu.$$

### A.1.10 The Normal Distribution

The normal distribution is the most important density in finite mixture modeling, both as a common choice as sampling density as well as posterior density in a Bayesian analysis of such models.

#### The Univariate Normal Distribution

The univariate normal distribution, $Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$, with $\mu \in \Re$ and $\sigma > 0$, is defined on $\Re$. Density, mean, and variance of $Y$ are given by

$$\begin{aligned} f_N(y;\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/(2\sigma^2)}, \\ \mathrm{E}(Y) = \mu, &\qquad \mathrm{Var}(Y) = \sigma^2. \end{aligned} \tag{A.19}$$

$1/\sigma^2$ is often called precision. A normal distribution with zero mean and zero precision has an improper density proportional to a constant. Higher-order moments $\mathrm{E}((Y-\mu)^m)$ are zero for $m$ odd, otherwise:

$$\mathrm{E}((Y-\mu)^m) = \sigma^{2m} \prod_{n=1}^{m/2} (2n-1). \tag{A.20}$$

#### The Multivariate Normal Distribution

The multivariate normal distribution, $\mathbf{Y} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \Re^d$ and $\boldsymbol{\Sigma} \in \Re^{d \times d}$ is a symmetric, positive definite matrix, is defined for $\mathbf{y} \in \Re^d$. Density, mean, and covariance matrix of $\mathbf{Y}$ are given by

$$\begin{aligned} f_N(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) &= (2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\tfrac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{'}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right), \\ \mathrm{E}(Y) = \boldsymbol{\mu}, &\quad \mathrm{Var}(\mathbf{Y}) = \boldsymbol{\Sigma}. \end{aligned} \tag{A.21}$$

$\boldsymbol{\Sigma}^{-1}$ is the information matrix. A normal distribution with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma}^{-1} = \mathbf{O}$ has an improper density proportional to a constant.

To sample from a multivariate normal distribution the following result is useful. Let $\mathbf{Z} = (Z_1 \dots Z_d)^{'}$ be $d$ independent copies of an $\mathcal{N}(0,1)$-distribution and let $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^{'}$ be the Cholesky decomposition of $\boldsymbol{\Sigma}$. Then $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

#### Marginal and Conditional Distributions

The quadratic form $(\mathbf{Y}-\boldsymbol{\mu})^{'}\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu}) \sim \chi^2_d$. Assume that $\mathbf{Y}$ is divided into two blocks, $\mathbf{Y}_1$ containing the first $d_1$ components of $\mathbf{Y}$, and $\mathbf{Y}_2$ containing the remaining $d_2 = d - d_1$ components. Apply a similar partition on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Then, marginally, $\mathbf{Y}_2 \sim \mathcal{N}_{d_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ and the conditional distribution of $\mathbf{Y}_1|\mathbf{Y}_2$ is $\mathcal{N}_{d_1}\left(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{11|2}\right)$, where

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{Y}_2 - \boldsymbol{\mu}_2),$$
$$\boldsymbol{\Sigma}_{11|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}'.$$

### The Normal Gamma Distribution

A pair $(\mathbf{Y}_1, Y_2)$ of random variables $\mathbf{Y}_1$ and $Y_2$ with $\mathbf{Y}_1|Y_2 \sim \mathcal{N}_d(\boldsymbol{\mu}, Y_2\boldsymbol{\Sigma})$ and $1/Y_2 \sim \mathcal{G}(\nu/2, \nu/2)$ follows the so-called normal Gamma distribution. The marginal distribution of $\mathbf{Y}_1$ is the $d$-variate $t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distribution, whereas the marginal distribution of $Y_2$ is $\mathcal{G}^{-1}(\nu/2, \nu/2)$.

## A.1.11 The Poisson Distribution

The Poisson distribution, $Y \sim \mathcal{P}(\mu)$ with $\mu > 0$, is defined for $y \in \{0, 1, 2, \ldots\}$ making it a standard distribution to model a random count variable. Density, mean, and variance are given by

$$f_P(y;\mu) = \frac{\mu^y}{y!}e^{-\mu}, \tag{A.22}$$
$$\mathrm{E}(Y) = \mathrm{Var}(Y) = \mu.$$

## A.1.12 The Student-*t* Distribution

The Student-$t$ distribution is useful as a sampling distribution for robust modeling.

### The Univariate Student-*t* Distribution

The univariate Student-$t$ distribution, $Y \sim t_\nu\left(\mu, \sigma^2\right)$ with $\mu \in \Re$, $\sigma > 0$, and $\nu > 0$, is defined for $y \in \Re$. The density is given by

$$f_{t_\nu}(y;\mu,\sigma^2) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi\sigma^2}}\left(1 + \frac{(y-\mu)^2}{\nu\sigma^2}\right)^{-(\nu+1)/2}.$$

If $\nu > 1$, then $\mathrm{E}(Y) = \mu$; if $\nu > 2$, then

$$\mathrm{Var}(Y) = \sigma^2\frac{\nu}{\nu-2}.$$

The univariate $t_\nu\left(\mu, \sigma^2\right)$-distribution is an infinite mixture of $\mathcal{N}\left(\mu, \sigma^2/\omega\right)$-distributions, where $\omega \sim \mathcal{G}(\nu/2, \nu/2)$:

$$f_{t_\nu}(y;\mu,\sigma^2) = \int f_N(y;\mu,\sigma^2/\omega) f_G(\omega;\nu/2,\nu/2)d\omega.$$

#### The Multivariate Student-$t$ Distribution

The multivariate Student-$t$ distribution, $\mathbf{Y} \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \Re^d$ and $\boldsymbol{\Sigma} \in \Re^{d \times d}$ is a symmetric, positive definite matrix, is defined for $\mathbf{y} \in \Re^d$. The density is given by

$$f_{t_\nu}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \tag{A.23}$$
$$\frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)}(\nu\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}\left(1 + \frac{1}{\nu}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)^{-(\nu+d)/2}.$$

If $\nu > 1$, then $\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu}$; if $\nu > 2$, then

$$\mathrm{Var}(\mathbf{Y}) = \frac{\nu}{\nu-2}\boldsymbol{\Sigma}.$$

The multivariate $t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distribution is an infinite mixture of $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\omega)$-distributions, where $\omega \sim \mathcal{G}(\nu/2, \nu/2)$:

$$f_{t_\nu}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int f_N(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}/\omega) f_G(\omega; \nu/2, \nu/2) d\omega.$$

### A.1.13 The Uniform Distribution

The uniform distribution, $Y \sim \mathcal{U}[0,1]$, is defined over the unit interval $y \in [0,1]$. Density, mean, and variance are given by

$$f_U(y) = 1, \qquad \mathrm{E}(Y) = 1/2, \qquad \mathrm{Var}(Y) = 1/12. \tag{A.24}$$

### A.1.14 The Wishart Distribution

The Wishart distribution, $\mathbf{Y} \sim \mathcal{W}_d(\alpha, \mathbf{S})$, with $\mathbf{S}$ being a $(d \times d)$ symmetric, nonsingular matrix ($|\mathbf{S}| > 0$), is a standard distribution law used for a random $(d \times d)$ symmetric, positive-definite matrix $\mathbf{Y}$. Many parameterizations are in use for this density; the following one, introduced by Bernardo and Smith (1994), is used in this book, because in this parameterization the $\mathcal{W}_1(\alpha, S)$ distribution reduces the $\mathcal{G}(\alpha, S)$ distribution.

For $\alpha > (d-1)/2$ the density of the $d(d+1)/2$-dimensional random vector of distinct elements of $\mathbf{Y}$ is given by

$$f_W(\mathbf{Y}; \alpha, \mathbf{S}) = \frac{|\mathbf{S}|^\alpha}{\Gamma_d(\alpha)}|\mathbf{Y}|^{\alpha-(d+1)/2}\exp\{-\mathrm{tr}\,(\mathbf{SY})\}, \tag{A.25}$$

where

$$\Gamma_d(\alpha) = \pi^{d(d-1)/4}\prod_{j=1}^{d}\Gamma\left(\frac{2\alpha+1-j}{2}\right) \tag{A.26}$$

is the generalized Gamma function. The mean of $\mathbf{Y}$ is given by

$$\mathrm{E}(\mathbf{Y}) = \alpha \mathbf{S}^{-1};$$

the mode reads:

$$(\alpha - \frac{d+1}{2})\mathbf{S}^{-1}.$$

The variance of the elements $Y_{ij}$ of $\mathbf{Y}$ is equal to

$$\mathrm{Var}(Y_{ij}) = \alpha \left( (S_{ij}^{-1})^2 + S_{ii}^{-1} S_{jj}^{-1} \right),$$

where $S_{ij}$ is the $(i, j)$th element of $\mathbf{S}$. If $\mathbf{Y} \sim \mathcal{W}_d(\alpha, \mathbf{S})$ and $\mathbf{A}$ is an $(m \times d)$ matrix with $m \leq d$, then

$$\mathbf{AYA}' \sim \mathcal{W}_m \left( \alpha, (\mathbf{AS}^{-1}\mathbf{A}')^{-1} \right),$$

provided that the scale matrix exists.

### Inverted Wishart Distribution

A random $(d \times d)$ symmetric, positive-definite matrix $\mathbf{Y}$ follows an inverted Wishart distribution, $\mathbf{Y} \sim \mathcal{W}_d^{-1}(\alpha, \mathbf{S})$, if $\mathbf{Y}^{-1} \sim \mathcal{W}_d(\alpha, \mathbf{S})$. In the parameterization used for the Wishart density in (A.25), the $\mathcal{W}_1^{-1}(\alpha, S)$ distribution reduces to the $\mathcal{G}^{-1}(\alpha, S)$ distribution. The density of the inverted Wishart distribution is given by

$$f_{IW}(\mathbf{Y}; \alpha, \mathbf{S}) = \frac{|\mathbf{S}|^{\alpha}}{\Gamma_d(\alpha)} |\mathbf{Y}^{-1}|^{\alpha+(d+1)/2} \exp\{-\mathrm{tr}\left(\mathbf{SY}^{-1}\right)\}, \qquad \text{(A.27)}$$

with $\Gamma_d(\alpha)$ being the same as in (A.26). $\mathbf{Y}$ has finite expectation, iff $\alpha > (d+1)/2$:

$$\mathrm{E}(\mathbf{Y}) = \mathbf{S}/(\alpha - (d+1)/2);$$

the mode is given by

$$\mathbf{S}/(\alpha + (d+1)/2).$$

## A.2 Software

A toolbox of MATLAB version 6 scripts and functions has been written by the author with the purpose of carrying out practical statistical modeling based on finite mixture and Markov switching models. The package is available at `http://www.ifas.jku.at/personal/fruehwirth/fruehwirthspringer06`.

You will need a valid MATLAB license including the STATISTICS toolbox to run functions from this package. The MATLAB software is available from MathWorks, Inc. (`http://www.mathworks.com/`).

# References

Abraham, B. and G. E. P. Box (1978). Linear models and spurious observations. *Applied Statistics 27*, 131–138.

Abraham, B. and J. Ledolter (1986). Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures. *International Statistical Review 54*, 51–66.

Ackerson, G. A. and K. S. Fu (1970). On state estimation in switching environments. *IEEE Transactions on Automatic Control 15*, 10–17.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing 6*, 251–262.

Aitkin, M. and I. Aitkin (1996). A hybrid EM/Gauss-Newton algorithm for maximum likelihood in mixture distributions. *Statistics and Computing 6*, 127–130.

Aitkin, M., D. Anderson, and J. Hinde (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, Ser. A 144*, 419–461.

Aitkin, M. and D. B. Rubin (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society, Ser. B 47*, 67–75.

Aitkin, M. and G. T. Wilson (1980). Mixture models, outliers, and the EM algorithm. *Technometrics 22*, 325–331.

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control 19*, 716–723.

Akashi, H. and H. Kumamoto (1977). Random sampling approach to state estimation in switching environments. *Automatica 13*, 429–434.

Albert, J. H. and S. Chib (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics 11*, 1–15.

Albert, P. S. (1991). A two-stage Markov mixture model for a time series of epileptic seizure counts. *Biometrics 47*, 1371–1381.

Albert, P. S. and L. E. Dodd (2004). A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics 60*, 427–435.

Allenby, G. M., N. Arora, and J. L. Ginter (1998). On the heterogeneity of demand. *Journal of Marketing Research 35*, 384–389.

Alspach, D. L. and H. W. Sorenson (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control 17*, 439–448.

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

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (2 ed.). Chichester: Wiley.

Andrews, D. W. K., I. Lee, and W. Ploberger (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics 70*, 9–38.

Ang, A. and G. Bekaert (2002). Regime switches in interest rates. *Journal of Business & Economic Statistics 20*, 163–182.

Antelman, G. (1997). *Elementary Bayesian Statistics*. Cheltenham: Edward Elgar.

Aoki, M. (1990). *State Space Modeling of Time Series*. New York: Springer.

Asea, P. K. and B. Blomberg (1998). Lending cycles. *Journal of Econometrics 83*, 89–128.

Atkinson, A. C. (1978). Posterior probabilities for choosing a regression model. *Biometrika 65*, 39–48.

Bacher, J. (2000). A probabilistic clustering model for variables of mixed type. *Quality & Quantity 34*, 223–235.

Baltagi, B. H. (1995). *Econometric Analysis of Panel Data*. Chichester: Wiley.

Banfield, J. D. and A. E. Raftery (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics 49*, 803–821.

Banjeree, T. and S. R. Paul (1999). An extension of Morel-Nagaraj's finite mixture distribution for modelling multinomial clustered data. *Biometrika 86*, 723–727.

Bar-Shalom, Y. and E. Tse (1975). Tracking in a cluttered environment with probabilistic data association. *Automatica 11*, 451–460.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. Chichester: Wiley.

Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society, Ser. B 42*, 293–321.

Baum, L. E. and T. Petrie (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics 37*, 1554–1563.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics 41*, 164–171.

Bauwens, L., C. S. Bos, H. K. van Dijk, and R. D. van Oest (2006). Adaptive radial-based direction sampling. *Journal of Econometrics*. Forthcoming.

Bauwens, L. and M. Lubrano (1998). Bayesian inference on GARCH models using the Gibbs sampler. *The Econometrics Journal 1*, C23–C46.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. Reprinted in *Biometrika 45*, 1958, 296–315.

Behboodian, J. (1970). On the modes of a mixture of two normal distributions. *Technometrics 12*, 131–139.

Bekaert, G. and C. R. Harvey (1995). Time-varying world market integration. *The Journal of Finance 50*, 403–444.

Bensmail, H. and G. Celeux (1996). Regularized Gaussian discriminant analysis through eigenvalue decomposition. *Journal of the American Statistical Association 91*, 1743–1748.

Bensmail, H., G. Celeux, A. E. Raftery, and C. P. Robert (1997). Inference in model-based cluster analysis. *Statistics and Computing 7*, 1–10.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics 5*, 445–463.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2 ed.). New York/Berlin/Heidelberg: Springer.

Berger, J. O. and M. Delampady (1987). Testing precise hypotheses. *Statistical Science 2*, 317–352.

Berger, J. O. and T. Sellke (1987). Testing a point null hypothesis: The irreconcilability of $p$ values and evidence. *Journal of the American Statistical Association 82*, 112–139.

Berkhof, J., I. van Mechelen, and A. Gelman (2003). A Bayesian approach to the selection and testing of mixture models. *Statistica Sinica 13*, 423–442.

Bernardinelli, L., D. Clayton, and C. Montomoli (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine 14*, 2411–2431.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society, Ser. B 41*, 113–147.

Bernardo, J. M. and F. G. Girón (1988). A Bayesian analysis of simple mixture problems. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 3*, pp. 67–78. New York: Clarendon.

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

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Ser. B 36*, 192–236.

Besag, J. (1989). A candidate's formula: A curious result in Bayesian prediction. *Biometrika 76*, 183–183.

Bhar, R. and S. Hamori (2004). *Hidden Markov Models. Applications to Financial Economics.* Dordrecht: Kluwer Academic.

Bickel, P. J., Y. Ritov, and T. Rydén (1998). Asymptotic normality of the maximum likelhood estimator for general hidden Markov models. *The Annals of Statistics 26*, 1614–1635.

Biernacki, C., G. Celeux, and G. Govaert (2000). Assessing a mixture model for clustering with the integrated classification likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 719–725.

Biernacki, C., G. Celeux, and G. Govaert (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statisics and Data Analysis 41*, 561–575.

Biernacki, C. and G. Govaert (1997). Using the classification likelihood to choose the number of clusters. *Computing Science and Statistics 29*, 451–457.

Billio, M. and A. Monfort (1998). Switching state-space models: Likelihood function, filtering and smoothing. *Journal of Statistical Planning and Inference 68*, 65–103.

Billio, M., A. Monfort, and C. P. Robert (1999). Bayesian estimation of switching ARMA models. *Journal of Econometrics 93*, 229–255.

Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika 65*, 31–38.

Binder, D. A. (1981). Approximations to Bayesian clustering rules. *Biometrika 68*, 275–285.

Blackwell, D. and L. Koopmans (1957). On the identifiability problem for functions of finite Markov chains. *Annals of Mathematical Statistics 28*, 1011–1015.

Blom, H. and Y. Bar-Shalom (1988). The interacting multiple model algorithm for systems with Markovian switching coefficients. *IEEE Transactions on Automatic Control 33*, 780–783.

Bock, H. H. (1996). Probabilistic models in cluster analysis. *Computational Statisics and Data Analysis 23*, 5–28.

Böhning, D. (1998). Zero-inflated Poisson models and C.A.MAN: A tutorial collection of evidence. *Biometrical Journal 40*, 833–843.

Böhning, D. (2000). *Computer-Assisted Analysis of Mixtures and Applications*. London: Chapman & Hall.

Böhning, D., E. Dietz, and P. Schlattmann (1998). Recent developments in computer-assisted analysis of mixtures. *Biometrics 54*, 525–536.

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

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992). ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics 52*, 5–39.

Bottolo, L., G. Consonni, P. Dellaportas, and A. Lijoi (2003). Bayesian analysis of extreme values by mixture modeling. *Extremes 6*, 25–47.

Bougerol, P. and N. Picard (1992a). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics 52*, 115–127.

Bougerol, P. and N. Picard (1992b). Strict stationarity of generalized autoregressive processes. *The Annals of Probability 20*, 1714–1730.

Box, G. E. P. and G. M. Jenkins (1970). *Time Series Analysis — Forecasting and Control*. San Francisco: Holden-Day.

Box, G. E. P. and G. C. Tiao (1968). A Bayesian approach to some outlier problems. *Biometrika 55*, 119–129.

Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The generel theory and its analytical extensions. *Psychometrika 52*, 345–370.

Bozdogan, H. and S. L. Sclove (1984). Multi-sample cluster analysis using Akaike's information criterion. *Annals of the Institute of Statistical Mathematics 36*, 163–180.

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association 88*, 9–25.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (2 ed.). New York: Springer.

Brooks, S. P. (2001). On Bayesian analyses and finite mixtures for proportions. *Statistics and Computing 11*, 179–190.

Brooks, S. P., P. Giudici, and G. O. Roberts (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society, Ser. B 65*, 3–55.

Brooks, S. P., B. J. T. Morgan, M. S. Ridout, and S. E. Pack (1997). Finite mixture models for proportions. *Biometrics 53*, 1097–1115.

Brusco, M. J. and J. D. Cradit (2001). A variable selection heuristic for $k$-means clustering. *Psychometrika 66*, 249–270.

Bryant, P. G. and J. A. Williamson (1978). Asymptotic behaviour of classification maximum likelihood estimates. *Biometrika 65*, 273–281.

Burmeister, E., K. D. Wall, and J. D. Hamilton (1986). Estimation of unobserved expected monthly inflation using Kalman filtering. *Journal of Business & Economic Statistics 4*, 147–160.

Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics 12*, 309–316.

Calinski, T. and J. Harabasz (1974). A dendrite method for cluster analysis. *Communications in Statistics 3*, 1–27.

Cameron, A. and P. Trivedi (1998). *Regression Analysis of Count Data.* Cambridge: Cambridge University Press.

Canova, F. (2004). Testing for convergence clubs in income per-capita: A predictive density approach. *International Economic Review 45*, 49–77.

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

Cappé, O., C. P. Robert, and T. Rydén (2003). Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *Journal of the Royal Statistical Society, Ser. B 65*, 679–700.

Carlin, B. P. and S. Chib (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Ser. B 57*, 473–484.

Carlin, B. P., N. G. Polson, and D. S. Stoffer (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association 87*, 493–500.

Carreira-Perpiñán, M. A. and S. Renals (2000). Practical identifiability of finite mixtures of multivariate Bernoulli distributions. *Neural Computation 12*, 141–152.

Carreira-Perpiñán, M. A. and C. K. I. Williams (2003). On the number of modes in a Gaussian mixture. In L. D. Griffin and M. Lillholm (Eds.), *Scale-Space Methods in Computer Vision*, Number 2695 in Lecture Notes in Computer Science, pp. 625–640. Berlin/Heidelberg: Springer.

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

Carter, C. K. and R. Kohn (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika 83*, 589–601.

Casella, G. and R. L. Berger (2002). *Statistical Inference* (2 ed.). Pacific Grove, CA: Wadsworth Group.

Casella, G., M. Lavine, and C. P. Robert (2001). Explaining the perfect sampler. *The American Statistician 55*, 299–305.

Casella, G., K. L. Mengersen, C. P. Robert, and D. T. Titterington (2002). Perfect slice samplers for mixtures of distributions. *Journal of the Royal Statistical Society, Ser. B 64*, 777–790.

Casella, G., C. P. Robert, and M. T. Wells (2000). Mixture models, latent variables and partitioned importance sampling. Technical Report DT-2000-03, INSEE, CREST, Paris.

Catchpole, E. and B. J. T. Morgan (1997). Detecting parameter redundancy. *Biometrika 84*, 187–196.

Cecchetti, S. G., P. Lam, and N. C. Mark (1990). Mean reversion in equilibrium asset prices. *The American Economic Review 80*, 398–418.

Celeux, G. (1998). Bayesian inference for mixture: The label switching problem. In P. J. Green and R. Rayne (Eds.), *COMPSTAT 98*, pp. 227–232. Heidelberg: Physica.

Celeux, G. and J. Diebolt (1985). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly 2*, 73–82.

Celeux, G. and G. Govaert (1991). Clustering criteria for discrete data and latent class models. *Journal of Classification 8*, 157–176.

Celeux, G. and G. Govaert (1992). A classification EM algorithm and two stochastic versions. *Computational Statisics and Data Analysis 14*, 315–332.

Celeux, G. and G. Govaert (1993). Comparison of the mixture and the classification maximum likelihood in cluster analysis. *Journal of Statistical Computation and Simulation 47*, 127–146.

Celeux, G. and G. Govaert (1995). Gaussian parsimonious clustering models. *Pattern Recognition 28*, 781–793.

Celeux, G., M. Hurn, and C. P. Robert (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association 95*, 957–970.

Celeux, G. and G. Soromenho (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification 13*, 195–212.

Chan, K. S. and J. Ledolter (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association 90*, 242–252.

Chandra, S. (1977). On the mixtures of probability distributions. *Scandinavian Journal of Statistics 4*, 105–112.

Chang, W. C. (1983). On using principal components before separating a mixture of two normal distributions. *Applied Statistics 32*, 267–275.

Charlier, C. V. L. and S. D. Wicksell (1924). On the dissection of frequency functions. *Arkiv för Matematik, Astronomi och Fysik 18*, 1–64.

Chen, C. and G. C. Tiao (1990). Random level-shift time series models, ARIMA approximations, and level-shift detection. *Journal of Business & Economic Statistics 8*, 83–97.

Chen, H., J. Chen, and J. D. Kalbfleisch (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society, Ser. B 63*, 19–29.

Chen, H., J. Chen, and J. D. Kalbfleisch (2004). Testing for a finite mixture model with two components. *Journal of the Royal Statistical Society, Ser. B 66*, 95–115.

Chen, J. and J. D. Kalbfleisch (1996). Penalized minimum-distance estimates in finite mixture models. *The Canadian Journal of Statistics 24*, 167–175.

Chen, M.-H., Q.-M. Shao, and J. G. Ibrahim (2000). *Monte Carlo Methods in Bayesian Computation*. Springer Series in Statistics. New York: Springer.

Chen, R. and J. S. Liu (1996). Predictive updating methods with application to Bayesian classification. *Journal of the Royal Statistical Society, Ser. B 58*, 397–415.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association 90*, 1313–1321.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics 75*, 79–97.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*, 221–241.

Chib, S. (2001). Markov chain Monte Carlo methods: Computation and inference. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, pp. 3569–3649. Amsterdam: North Holland.

Chib, S. and B. P. Carlin (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing 9*, 17–26.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *Journal of the American Statistical Association 49*, 327–335.

Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association 96*, 270–281.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics 108*, 281–316.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika 89*, 539–551.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear linear regressions (Ref: 74V42 p601-608). *Econometrica 28*, 591–605.

Clements, M. P. and H.-M. Krolzig (1998). A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. *The Econometrics Journal 1*, 47–75.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, and M. E. Sobel (Eds.), *Handbook of Statistical Modelling for the Social and Behavioral Sciences*, pp. 311–359. New York: Plenum.

Clogg, C. C. and L. A. Goodman (1984). Latent structure analysis of a set of multidimensional contincency tables. *Journal of the American Statistical Association 79*, 762–771.

Cohen, A. C. (1960). An extension of a truncated Poisson distribution. *Biometrics 16*, 446–450.

Cohen, A. C. (1966). A note on certain discrete mixed distributions. *Biometrics 22*, 566–572.

Congdon, P. (2001). *Bayesian Statistical Modelling*. Chichester: Wiley.

Congdon, P. (2005). *Bayes Models for Categorical Data*. Chichester: Wiley.

Copas, J. B. (1988). Binary regression models for contaminated data. *Journal of the Royal Statistical Society, Ser. B 50*, 225–265.

Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society, Ser. A 134*, 321–367.

Cosslett, S. R. and L.-F. Lee (1985). Serial correlation in latent discrete variable models. *Journal of Econometrics 27*, 79–97.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Crawford, S. L. (1994). An application of the Laplace method to finite mixture distributions. *Journal of the American Statistical Association 89*, 259–267.

Crowder, M. J. and D. J. Hand (1990). *Analysis of Repeated Measures*. London: Chapman & Hall.

Cruz-Medina, I. R., T. P. Hettmansperger, and H. Thomas (2004). Semiparametric mixture models and repeated measures: The multinomial cut point model. *Applied Statistics 53*, 463–474.

Cutler, A. and O. Cordero-Braña (1996). Minimum Hellinger distance estimation for finite mixture models. *Journal of the American Statistical Association 91*, 1716–1723.

Dacunha-Castelle, D. and E. Gassiat (1997). The estimation of the order of a mixture model. *Bernoulli 3*, 279–299.

Dalal, S. R. and W. J. Hall (1983). Approximating priors by mixtures of natural conjugate priors. *Journal of the Royal Statistical Society, Ser. B 45*, 278–286.

Dalrymple, M. L., I. L. Hudson, and R. P. K. Ford (2003). Finite mixture, zero-inflated Poisson and Hurdle models with application to SIDS. *Computational Statisics and Data Analysis 41*, 491–504.

Damien, P., J. Wakefield, and S. Walker (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society, Ser. B 61*, 331–344.

Dasgupta, A. and A. E. Raftery (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association 93*, 294–302.

Davidian, M. and D. M. Giltinan (1998). *Nonlinear Models for Repeated Measurement Data*. London: Chapman & Hall.

Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika 56*, 463–474.

De Jong, P. and S. Chu-Chun-Lin (1994). Stationary and non-stationary state space models. *Journal of Time Series Analysis 15*, 151–166.

De Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–350.

De Soete, G. and W. S. DeSarbo (1991). A latent class probit model for analyzing pick any/N data. *Journal of Classification 8*, 45–63.

Dellaportas, P., J. J. Forster, and I. Ntzoufras (2002). On Baysian model and variable selection using MCMC. *Statistics and Computing 12*, 27–36.

Dellaportas, P. and I. Papageorgiou (2006). Multivariate mixtures of normals with unknown number of components. *Statistics and Computing 16*, 57–68.

Delmar, P., S. Robin, D. Tronik-Le Roux, and J. J. Daudin (2005). Mixture model on the variance for the differential analysis of gene expression data. *Applied Statistics 54*, 31–50.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B 39*, 1–37.

DeSarbo, W. S. and W. L. Cron (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification 5*, 248–282.

DeSarbo, W. S., M. Wedel, M. Vriens, and V. Ramaswamy (1992). Latent class metric conjoint analysis. *Marketing Letters 3*, 273–288.

Dey, D. K., L. Kuo, and S. K. Sahu (1995). A Bayesian predictive approach to determining the number of components in a mixture distribution. *Statistics and Computing 5*, 297–305.

Diaconis, P. and D. Ylvisaker (1979). Conjugate priors for exponential families. *The Annals of Statistics 7*, 269–281.

Dias, J. G. and M. Wedel (2004). An empirical comparison of EM, SEM and MCMC performance for problematical Gaussian mixture likelihoods. *Statistics and Computing 14*, 323–332.

DiCiccio, T. J., R. E. Kass, A. Raftery, and L. Wasserman (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association 92*, 903–915.

Dickey, D. A. and W. A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica 49*, 1057–1072.

Dickey, J. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics 42*, 204–23.

Diebold, F. X., J.-H. Lee, and G. C. Weinbach (1994). Regime switching with time-varying transition probabilities. In C. P. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302. Oxford: Oxford University Press.

Diebold, F. X. and G. D. Rudebusch (1996). Measuring business cycles: A modern perspective. *The Review of Economics and Statistics 78*, 67–77.

Diebolt, J. and C. P. Robert (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Ser. B 56*, 363–375.

Diggle, P. J., P. Heagerty, K.-Y. Liang, and S. L. Zeger (2002). *Analysis of Longitudinal Data* (2 ed.). Oxford: Oxford University Press.

Douc, R., E. Moulines, and T. Rydén (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *The Annals of Statistics 32*, 2254–2304.

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

Dueker, M. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business & Economic Statistics 15*, 26–34.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society, Ser. B 62*, 3–56.

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

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*, 603–615.

Durland, J. M. and T. H. McCurdy (1994). Duration-dependent transitions in a Markov model of U.S. GNP growth. *Journal of Business & Economic Statistics 12*, 279–288.

Edwards, W., H. Lindman, and L. J. Savage (1963). Bayesian statistical inference for psychological research. *Psychological Review 70*, 193–242.

Efron, B. and C. Morris (1977). Stein's paradox in statistics. *Scientific American 236*, 119–127.

Engel, C. (1994). Can the Markov switching model forecast exchange rates? *Journal of International Economics 36*, 151–165.

Engel, C. and J. D. Hamilton (1990). Long swings in the Dollar: Are they in the data and do markets know it? *The American Economic Review 80*, 689–713.

Engel, C. and C.-J. Kim (1999). The long-run U.S./U.K. real exchange rate. *Journal of Money, Credit, and Banking 31*, 335–356.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric Reviews 5*, 1–87.

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

Engle, R. F. and V. K. Ng (1995). Measuring and testing the impact of news on volatility. In R. Engle (Ed.), *ARCH: Selected Readings*, pp. 145–175. Oxford: Oxford University Press.

Engle, R. F. and M. Watson (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association 76*, 774–781.

Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association 90*, 577–588.

Escobar, M. D. and M. West (1998). Computing nonparametric hierarchical models. In D. Dey, P. Müller, and D. Sinha (Eds.), *Practical Nonparametric and Semiparametric Bayesian Statistics*, Number 133 in Lecture Notes in Statistics, pp. 1–22. Berlin: Springer.

Evans, M., Z. Gilula, and I. Guttman (1989). Latent class analysis of two-way contingency tables by Bayesian methods. *Biometrika 76*, 557–563.

Evans, M., I. Guttman, and I. Olkin (1992). Numerical aspects in estimating the parameters of a mixture of normal distributions. *Journal of Computational and Graphical Statistics 1*, 351–365.

Everitt, B. S. (1979). Unresolved problems in cluster analysis. *Biometrics 35*, 169–181.

Everitt, B. S. (1985). Mixture distributions. In S. Kotz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, Volume 5, pp. 559–569. New York: Wiley.

Everitt, B. S. (1988). A finite mixture model for the clustering of mixed mode data. *Statistics & Probability Letters 6*, 305–309.

Everitt, B. S. and D. J. Hand (1981). *Finite Mixture Distributions*. London: Chapman & Hall.

Everitt, B. S., S. Landau, and M. Leese (2001). *Cluster Analysis* (4 ed.). London: Edward Arnold.

Everitt, B. S. and C. Merette (1990). The clustering of mixed mode data: A comparison of possible approaches. *Journal of Applied Statistics 17*, 283–297.

Fahrmeir, L. and L. Knorr-Held (2000). Dynamic and semiparametric models. In M. Schimek (Ed.), *Smoothing and Regression: Approaches, Computation and Application*, pp. 417–452. New York: Wiley.

Fair, R. C. and D. M. Jaffee (1972). Methods of estimation for markets in disequilibrium. *Econometrica 40*, 497–514.

Falk, B. (1986). Further evidence on the asymmetric behaviour of economic time series over the business cycle. *Journal of Political Economy 94*, 1096–1109.

Fama, E. (1965). The behavior of stock market prices. *Journal of Business 38*, 34–105.

Fama, E. and M. Gibbons (1982). Inflation, real returns and capital investment. *Journal of Monetary Economics 9*, 297–323.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics 38*, 1041–1046.

Farewell, V. T. and D. A. Sprott (1988). The use of a mixture model in the analysis of count data. *Biometrics 44*, 1191–1194.

Fearnhead, P. and P. Clifford (2003). On-line inference for hidden Markov models via particle filters. *Journal of the Royal Statistical Society, Ser. B 65*, 887–899.

Feller, W. (1943). On a general class of contagious distributions. *Annals of Mathematical Statistics 14*, 389–400.

Feng, Z. and C. McCulloch (1996). Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society, Ser. B 58*, 609–617.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics 1*, 209–230.

Fernández, C. and P. J. Green (2002). Modeling spatially correlated data via mixtures: A Bayesian approach. *Journal of the Royal Statistical Society, Ser. B 64*, 805–826.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics 12*, 299–308.

Filardo, A. J. and S. F. Gordon (1998). Business cycle duration. *Journal of Econometrics 85*, 99–123.

Finch, S. J., N. R. Mendell, and H. C. Thode (1989). Probabilistic measures of adequacy of a numerical search for global maximum. *Journal of the American Statistical Association 84*, 1020–1023.

Fokoué, E. and D. M. Titterington (2003). Mixture of factor analysers. Bayesian estimation and inference by stochastic simulation. *Machine Learning 50*, 73–94.

Follmann, D. A. and D. Lambert (1989). Generalizing logistic regression by nonparametric mixing. *Journal of the American Statistical Association 84*, 295–300.

Follmann, D. A. and D. Lambert (1991). Identifiability of finite mixtures of logistic regression models. *Journal of Statistical Planning and Inference 27*, 375–381.

Fong, W. (1997). Volatility persistence and switching ARCH in Japanese stock return. *Financial Engeneering and the Japanese Market 4*, 37–57.

Formann, A. K. (1982). Linear logistic latent class analysis. *Biometrical Journal 24*, 171–190.

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association 87*, 476–486.

Formann, A. K. (1993). Fixed-distance latent class models for the analysis of sets of two-way contingency tables. *Biometrics 49*, 511–521.

Formann, A. K. (1994a). Measurement errors in caries diagnosis: Some further latent class models. *Biometrics 50*, 865–871.

Formann, A. K. (1994b). Measuring change in latent subgroups using dichotomous data: Unconditional, conditional, and semiparametric maximum likelihood estimation. *Journal of the American Statistical Association 89*, 1027–1034.

Formann, A. K. and T. Kohlmann (1996). Latent class analysis in medical research. *Statistical Methods in Medical Research 5*, 179–211.

Fowlkes, E. B. (1979). Some methods for studying the mixture of two normal (lognormal) distributions. *Journal of the American Statistical Association 74*, 561–575.

Fowlkes, E. B., R. Gnanadesikan, and J. R. Kettering (1988). Variable selection in clustering. *Journal of Classification 5*, 205–228.

Fraley, C. and A. E. Raftery (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *Computer Journal 41*, 578–588.

Fraley, C. and A. E. Raftery (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association 97*, 611–631.

Francq, C. and M. Roussignol (1997). On white noises driven by hidden Markov chains. *Journal of Time Series Analysis 18*, 553–578.

Francq, C. and M. Roussignol (1998). Ergodicity of autoregressive processes with Markov-switching and consistency of the maximimum-likelihood estimator. *Statistics 32*, 151–173.

Francq, C., M. Roussignol, and J. Zakoian (2001). Conditional heteroscedasticity driven by hidden Markov chains. *Journal of Time Series Analysis 22*, 197–220.

Francq, C. and J. Zakoian (1999). Linear-representations based estimation of switching-regime GARCH models. Proceedings of the NBER/NSF Time Series Conference.

Francq, C. and J. Zakoian (2001). Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics 102*, 339–364.

Franses, P. H. and D. van Dijk (2000). *Nonlinear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press.

Friedman, H. P. and J. Rubin (1967). On some invariant criteria for grouping data. *Journal of the American Statistical Association 63*, 1159–1178.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association 84*, 165–175.

Frühwirth, R. (1987). Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research A262*, 444–450.

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

Frühwirth-Schnatter, S. (1995). Bayesian model discrimination and Bayes factors for linear Gaussian state space models. *Journal of the Royal Statistical Society, Ser. B 57*, 237–246.

Frühwirth-Schnatter, S. (2001a). Fully Bayesian analysis of switching Gaussian state space models. *Annals of the Institute of Statistical Mathematics 53*, 31–49.

Frühwirth-Schnatter, S. (2001b). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*, 194–209.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *The Econometrics Journal 7*, 143–167.

Frühwirth-Schnatter, S. and R. Frühwirth (2006). Auxiliary mixture sampling with applications to logistic models. Research Report IFAS, http://www.ifas.jku.at/.

Frühwirth-Schnatter, S. and S. Kaufmann (2006a). How do changes in monetary policy affect bank lending? An analysis of Austrian bank data. *Journal of Applied Econometrics 21*, 275–305.

Frühwirth-Schnatter, S. and S. Kaufmann (2006b). Model-based clustering of multiple time series. Research Report IFAS, http://www.ifas.jku.at/.

Frühwirth-Schnatter, S. and T. Otter (1999). Conjoint-analysis using mixed effect models. In H. Friedl, A. Berghold, and G. Kauermann (Eds.), *Statistical Modelling. Proceedings of the Fourteenth International Workshop on Statistical Modelling*, pp. 181–191. Graz.

Frühwirth-Schnatter, S., R. Tüchler, and T. Otter (2004). Bayesian analysis of the heterogeneity model. *Journal of Business & Economic Statistics 22*, 2–15.

Frühwirth-Schnatter, S., R. Tüchler, and T. Otter (2005). Capturing unobserved consumer heterogeneity using the Bayesian heterogeneity model. In A. Taudes (Ed.), *Adaptive Information Systems and Modelling in Economics and Management Science*, pp. 57–70. Wien: Springer.

Frühwirth-Schnatter, S. and H. Wagner (2006). Auxiliary mixture sampling for parameter-driven models of time series of small counts with applications to state space modelling. Research Report IFAS 11, http://www.ifas.jku.at/.

Gamerman, D. (1997). *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference*. London: Chapman & Hall.

Gamerman, D. and A. F. M. Smith (1996). Bayesian analysis of longitudinal data studies. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M.

Smith (Eds.), *Bayesian Statistics 5*, pp. 587–597. Oxford: Oxford University Press.

Garcia, R. and P. Perron (1996). An analysis of real interest rate under regime shift. *The Review of Economics and Statistics 78*, 111–125.

Geisser, S. (1964). Estimation in the uniform covariance case. *Journal of the Royal Statistical Society, Ser. B 26*, 477–483.

Geisser, S. and J. Cornfield (1963). Posterior distributions for multivariate normal parameters. *Journal of the Royal Statistical Society, Ser. B 25*, 368–376.

Gelfand, A., S. Sahu, and B. Carlin (1995). Efficient parametrisations for normal linear mixed models. *Biometrika 82*, 479–488.

Gelfand, A. E. and D. K. Dey (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Ser. B 56*, 501–514.

Gelfand, A. E., D. K. Dey, and H. Chang (1992). Model determination using predictive distributions with implementation via sampling-based methods. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 147–167. Oxford: Oxford University Press.

Gelfand, A. E. and A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis* (2 ed.). London: Chapman & Hall/CRC.

Gelman, A., X.-L. Meng, and H. Stern (1996). Posterior predictive assessment of model fitness via realized discrepancies (Disc: p760-807). *Statistica Sinica 6*, 733–760.

George, E. I. and R. McCulloch (1997). Approaches for Bayesian variable selection. *Statistica Sinica 7*, 339–373.

Gerlach, R., C. C. and R. Kohn (2000). Efficient Bayesian inference for dynamic mixture models. *Journal of the American Statistical Association 95*, 819–828.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–1339.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Disc: p189-193). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 169–188. Oxford: Oxford University Press.

Geweke, J. (1993). Bayesian treatment of the independent Student-$t$ linear model. *Journal of Applied Econometrics 8(Supplement)*, 19–40.

Geweke, J. (1999). Using simulation methods for Bayesian econometric models: Inference, development, and communication (Disc: p75-126). *Econometric Reviews 18*, 1–73.

Geweke, J. and M. Keane (1999). Mixture of normals probit models. In C. Hsiao, M. H. Pesaran, K. L. Lahiri, and L. F. Lee (Eds.), *Analysis of*

*Panels and Limited Dependent Variables*, pp. 49–78. Cambridge: Cambridge University Press.

Geyer, C. and J. Møller (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics 21*, 359–373.

Ghosh, D. and A. M. Chinnaiyan (2002). Mixture modelling of gene expression data from microarray experiments. *Bioinformatics 18*, 275–286.

Gilula, Z. (1979). Singular value decomposition of probability matrices: Probabilistic aspects of latent dichotomous variables. *Biometrika 66*, 339–44.

Gnanadesikan, R., J. R. Kettering, and S. L. Tao (1995). Weighting and selection of variables for cluster analysis. *Journal of Classification 12*, 113–136.

Godsill, S. J. (1997). Bayesian enhancement of speech and audio signals which can be modelled as ARMA processes. *International Statistical Review 65*, 1–21.

Godsill, S. J. (2001). On the relation between MCMC model uncertainty methods. *Journal of Computational and Graphical Statistics 10*, 230–248.

Godsill, S. J. and P. Rayner (1998). Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler. *IEEE Transactions on Speech and Audio Processing 6*, 352–372.

Goldfeld, S. and R. Quandt (1973). A Markov model for switching regression. *Journal of Econometrics 1*, 3–16.

Goodman, L. A. (1974a). The analysis of systems of quantitative variables when some of the variables are unobservable. Part I - a modified latent structure approach. *American Journal of Sociology 79*, 1179–1259.

Goodman, L. A. (1974b). Exploratory latent structure analysis using both identiable and unidentifiable models. *Biometrika 61*, 215–231.

Goodman, L. A. (1978). *Analyzing Qualitative/Categorical Data: Log-linear Models and Latent Structure Analysis.* Reading, MA: Addison-Wesley.

Goodwin, T. H. (1993). Business cycle analysis with a Markov switching model. *Journal of Business & Economic Statistics 11*, 331–339.

Gordon, A. (1999). *Classification* (2 ed.). London: Chapman & Hall.

Gordon, K. and A. F. M. Smith (1990). Modeling and monitoring biomedical time series. *Journal of the American Statistical Association 85*, 328–337.

Gosh, J. K. and P. K. Sen (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In L. LeCam and A. Olshen (Eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Volume II, pp. 789–806. Monterey, CA: Wadsworth Advanced Books.

Granger, C. W. J. and D. Orr (1972). Infinite variance and research strategy in time series analysis. *Journal of the American Statistical Association 67*, 275–285.

Granger, C. W. J. and T. Teräsvirta (1993). *Modelling Nonlinear Economic Relationships.* Oxford: Oxford University Press.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime switching process. *Journal of Financial Economics 42*, 27–62.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 711–732.

Green, P. J. (2003). Trans-dimensional Markov chain Monte Carlo. In P. J. Green, N. L. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 179–198. Oxford: Oxford University Press.

Green, P. J. and T. O'Hagan (2000). Model choice with MCMC on product spaces without using pseudo-priors. University of Nottingham Research Report.

Green, P. J. and S. Richardson (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics 28*, 355–375.

Green, P. J. and S. Richardson (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Association 97*, 1–16.

Grilli, V. and G. Kaminsky (1991). Nominal exchange rate regimes and the real exchange rate: Evidence from the United States and Great Britain, 1885-1986. *Journal of Monetary Economics 27*, 191–212.

Grim, J. and M. Haindl (2003). Texture modelling by discrete distribution mixtures. *Computational Statisics and Data Analysis 41*, 603–615.

Gruet, M.-A., A. Philippe, and C. P. Robert (1999). MCMC control spreadsheets for exponential mixture estimation. *Journal of Computational and Graphical Statistics 8*, 298–317.

Grün, B. (2002). Identifizierbarkeit von multinomialen Mischmodellen. Masters thesis, Technische Universität Wien, Vienna, Austria.

Grün, B. and F. Leisch (2004). Bootstrapping finite mixture models. In J. Antoch (Ed.), *COMPSTAT 2004. Proceedings in Computational Statistics*, pp. 1115–1122. Heidelberg: Physica-Verlag/Springer.

Guttman, I. (1973). Care and handling of univariate or multivariate outliers in detecting spuriousity – A Bayesian approach. *Technometrics 15*, 723–738.

Guttman, I., R. Dutter, and P. R. Freeman (1978). Care and handling of univariate outliers in the general linear model to detect spuriousity – A Bayesian approach. *Technometrics 20*, 187–193.

Hall, S. G., Z. Psaradakis, and M. Sola (1997). Cointegration and changes in regime: The Japanese consumption function. *Journal of Applied Econometrics 12*, 151–168.

Hall, S. G., Z. Psaradakis, and M. Sola (1999). Detecting periodically collapsing bubbles: A Markov-switching unit root test. *Journal of Applied Econometrics 14*, 143–154.

Hamilton, J. (1994a). State-space models. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 3039–3080. Amsterdam: North-Holland.

Hamilton, J. (1994b). *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hamilton, J. D. (1988). Rational expectations econometric analysis of changes in regime: An investigation on the term structure of interest rates. *Journal of Economic Dynamics and Control 12*, 385–423.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics 45*, 39–70.

Hamilton, J. D. and G. Lin (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics 11*, 573–593.

Hamilton, J. D. and B. Raj (2002). New directions in business cycle research and financial analysis. *Empirical Economics 27*, 149–162.

Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics 64*, 307–333.

Han, C. and B. P. Carlin (2001). Markov chain Monte Carlo methods for computing Bayes factors: A comparative review. *Journal of the American Statistical Association 96*, 1122–1132.

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

Hansen, B. (1992). The likelihood ratio test under non-standard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics 7*, S61 –S82.

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

Hartigan, J. A. (1975). *Clustering Algorithm.* Chichester: Wiley.

Hartigan, J. A. (1977). Distribution problems in clustering. In *Classification and Clustering*, pp. 45–72. San Diego: Academic.

Hartigan, J. A. (1985). A failure of likelihood ratio asymptotics for normal mixtures. In L. LeCam and A. Olshen (Eds.), *Proceeding of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Volume II, pp. 807–810. Monterey, CA: Wadsworth Advanced Books.

Harvey, A., E. Ruiz, and E. Sentana (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics 52*, 129–157.

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

Harvey, A. C. (1993). *Time Series Models* (2 ed.). Harvester-Wheatsheaf.

Harvey, A. C. and C. Fernandes (1989). Time series models for count or qualitative observations. *Journal of Business* & *Economic Statistics 7*, 407–417.

Harvey, A. C., S. J. Koopman, and J. Penzer (1998). Messy time series: A unified approach. *Advances in Econometrics 13*, 103–143.

Harvey, A. C. and P. H. J. Todd (1983). Forecasting economic time series with structural and Box-Jenkins models: A case study. *Journal of Business & Economic Statistics 1*, 299–315.

Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics 8*, 431–444.

Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association 64*, 1459–1471.

Hastie, T. and R. Tibshirani (1996). Discriminant analysis by Gaussian mixtures. *Journal of the Royal Statistical Society, Ser. B 58*, 155–176.

Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics 13*, 795–800.

Hathaway, R. J. (1986). Another interpretation of the EM algorithm for mixture distributions. *Statistics & Probability Letters 4*, 53–56.

Hawkins, D. S., D. M. Allen, and A. J. Stromberg (2001). Determining the number of components in mixtures of linear models. *Computational Statisics and Data Analysis 38*, 14–48.

Heckman, J. J., R. Robb, and J. R. Walker (1990). Testing the mixture of exponentials hypothesis and estimating the mixing distribution by the method of moments. *Journal of the American Statistical Association 85*, 582–589.

Heckman, J. J. and B. Singer (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica 52*, 271–320.

Heller, A. (1965). On stochastic processes derived from Markov chains. *Annals of Mathematical Statistics 36*, 1286–1291.

Hennig, C. (2000). Identifiability of models for clusterwise linear regression. *Journal of Classification 17*, 273–296.

Hettmansperger, T. P. and H. Thomas (2000). Almost nonparametric inference for repeated measures in mixture models. *Journal of the Royal Statistical Society, Ser. B 62*, 811–825.

Hinton, G. E., P. Dayan, and M. Revow (1997). Modeling the manifolds of images of handwritten digits. *IEEE Transactions on Neural Networks 8*, 65–73.

Hobert, J. P. and G. Casella (1998). Functional compatibility, Markov chains, and Gibbs sampling with improper posteriors. *Journal of Computational and Graphical Statistics 7*, 42–60.

Hobert, J. P., C. P. Robert, and D. M. Titterington (1999). On perfect simulation for some mixtures of distributions. *Statistics and Computing 9*, 287–298.

Hoek, H., A. Lucas, and H. K. van Dijk (1995). Classical and Bayesian aspects of robust unit root inference. *Journal of Econometrics 69*, 27–59.

Holst, U., G. Lindgren, J. Holst, and M. Thuvesholmen (1994). Recursive estimation in switching autoregressions with a Markov regime. *Journal of Time Series Analysis 15*, 489–506.

Hosmer, D. W. (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three types of sample. *Biometrics 29*, 761–770.

Hosmer, D. W. (1974). Maximum likelihood estimates of the parameters of a mixture of two regression lines. *Communications in Statistics, Part A – Theory and Methods 3*, 995–1006.

Hunt, L. and M. Jorgensen (2003). Mixture model clustering for mixed data with missing information. *Computational Statisics and Data Analysis 41*, 429–440.

Hurn, M., A. Justel, and C. P. Robert (2003). Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics 12*, 55–79.

Jacobs, R. A., M. I. Jordan, S. J. Nowlan, and G. E. Hinton (1991). Adaptive mixtures of local experts. *Neural Computation 3*, 79–87.

Jacobs, R. A., F. Peng, and M. A. Tanner (1997). A Bayesian approach to model selection in hierarchical mixtures-of-experts architectures. *Neural Networks 10*, 231–241.

Jacobs, R. A., M. A. Tanner, and F. Peng (1996). Bayesian inference for hierarchical mixtures-of-experts with applications to regression and classification. *Statistical Methods in Medical Research 5*, 375–390.

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

Jalali, A. and J. Pemberton (1995). Mixture models for time series. *Journal of Applied Probability 32*, 123–138.

Jansen, R. C. (1993). Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. *Biometrics 49*, 227–231.

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

Jeffreys, S. H. (1948). *Theory of Probability* (2 ed.). Oxford: Clarendon.

Jensen, J. L. and N. V. Petersen (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *The Annals of Statistics 27*, 514–535.

Jiang, W. and M. A. Tanner (1999). On the identifiability of mixtures-of-experts. *Neural Network 12*, 1253–1258.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions* (2 ed.), Volume 1. New York: Wiley.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions* (2 ed.), Volume 2. New York: Wiley.

Johnson, N. L., S. Kotz, and A. Kemp (1993). *Univariate Discrete Distributions* (2 ed.). New York: Wiley.

Jordan, M. I. and R. A. Jacobs (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation 6*, 181–214.

Jorgensen, M. A. (2004). Using multinomial mixture models to cluster Internet traffic. *Australian & New Zealand Journal of Statistics 46*, 205–218.

Juang, B. H. and L. R. Rabiner (1985). Mixture autoregressive hidden Markov models for speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing 33*, 1404–1413.

Justel, A. and D. Peña (1996). Gibbs sampling will fail in outlier problems with strong masking. *Journal of Computational and Graphical Statistics 5*, 176–189.

Justel, A. and D. Peña (2001). Bayesian unmasking in linear models. *Computational Statistics and Data Analysis 36*, 69–84.

Kadane, J. B. and N. Lazar (2004). Methods and criteria for model selection. *Journal of the American Statistical Association 99*, 279–290.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions ASME Journal of Basic Engeneering 82*, 35–45.

Kalman, R. E. (1961). New results in filtering and prediction theory. *Transactions ASME Journal of Basic Engeneering 81*, 95–108.

Kamakura, W. A. (1991). Estimating flexible distributions of ideal-points with external analysis of preferences. *Psychometrika 56*, 419–48.

Kamakura, W. A. and G. J. Russell (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research 26*, 379–90.

Karawatzki, R., J. Leydold, and K. Pötzelberger (2005). Automatic Markov chain Monte Carlo procedures for sampling from multivariate distributions. Electronic publication available at `http://epub.wu-wien.ac.at`, Vienna University of Economics and Business Administration.

Karlin, S. and H. M. Taylor (1975). *A First Course in Stochastic Processes* (2 ed.). San Diego: Academic.

Karlis, D. and E. Xekalaki (1998). Minimum Hellinger distance estimation for Poisson mixtures. *Computational Statisics and Data Analysis 29*, 81–103.

Karlis, D. and E. Xekalaki (2001). Robust inference for finite Poisson mixtures. *Journal of Statistical Planning and Inference 93*, 93–115.

Karlis, D. and E. Xekalaki (2003). Choosing initial values for the EM algorithm for finite mixtures. *Computational Statisics and Data Analysis 41*, 577–590.

Karlis, D. and E. Xekalaki (2005). Mixed Poisson distributions. *International Statistical Review 73*, 35–58.

Karlsen, H. A. (1990). Existence of moments in a stationary stochastic difference equation. *Advances in Applied Probability 22*, 129–146.

Karlsen, H. A. and D. Tjøstheim (1990). Autoregressive segmentation of signal traces with application to geological dipmeter measurements. *IEEE Transactions on Geoscience and Remote Sensing 28*, 171–181.

Kashyap, R. L. (1970). Maximum likelihood identification of stochastic linear systems. *IEEE Transactions on Automatic Control 15*, 25–34.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*, 773–795.

Kass, R. E., L. Tierney, and J. B. Kadane (1988). Asymptotics in Bayesian computation. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 3*, pp. 261–278. New York: Clarendon.

Kass, R. E., L. Tierney, and J. B. Kadane (1990). The validity of posterior expansions based on Laplace's method. In S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner (Eds.), *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard*, pp. 473–488. New York: Elsevier.

Kass, R. E. and S. K. Vaidyanathan (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statistical Society, Ser. B 54*, 129–144.

Kaufman, L. and P. J. Rousseeuw (1990). *Finding Groups in Data: An Introduction to Cluster Analysis.* New York: Wiley.

Kaufmann, S. (2000). Measuring business cycles with a dynamic Markov switching factor model: An assessment using Bayesian simulation methods. *The Econometrics Journal 3*, 39–65.

Kaufmann, S. (2002). Is there an asymmetric effect of monetary policy over time. *Empirical Economics 27*, 277–297.

Kaufmann, S. and S. Frühwirth-Schnatter (2002). Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis 23*, 425–458.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhya A 62*, 49–66.

Kiefer, N. M. (1978). Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica 46*, 427–434.

Kiefer, N. M. and J. Wolfowitz (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics 27*, 887–906.

Kim, B. S. and B. H. Margolin (1992). Testing goodness of fit of a multinomial model against overdispersed alternatives. *Biometrics 48*, 711–719.

Kim, C.-J. (1993a). Sources of monetary growth uncertainty and economic activity: The time-varying parameter model with heteroscedastic disturbances. *The Review of Economics and Statistics 75*, 483–492.

Kim, C.-J. (1993b). Unobserved-component time series models with Markov-switching heteroscedasticity: Changes in regime and the link between inflation rates and inflation uncertainty. *Journal of Business & Economic Statistics 11*, 341–349.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics 60*, 1–22.

Kim, C.-J. and M.-J. Kim (1996). Transient fads and the crash of '87. *Journal of Applied Econometrics 11*, 41–58.

Kim, C.-J. and C. R. Nelson (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime-switching. *The Review of Economics and Statistics 80*, 188–201.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications.* Cambridge, MA: MIT Press.

Kim, C.-J. and C. R. Nelson (2001). A Bayesian approach to testing for Markov switching in univariate and dynamic factor models. *International Economic Review 42*, 989–1013.

Kim, I.-M. (1993c). A dynamic programming approach to the estimation of Markov switching regression models. *Journal of Statistical Computation and Simulation 45*, 61–76.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Kirnbauer, R., S. Schnatter, and D. Gutknecht (1987). Bayesian estimation of design floods under regional and subjective prior information. In R. Viertl (Ed.), *Probability and Bayesian Statistics*, pp. 285–294. New York: Plenum.

Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive filter. *Journal of Time Series Analysis 2*, 103–116.

Kitagawa, G. and W. Gersch (1984). A smoothness priors state space modeling of time series with trend and seasonality. *Journal of the American Statistical Association 79*, 378–389.

Klaasen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics 27*, 363–394.

Kleibergen, F. and R. Paap (2002). Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics 111*, 223–249.

Kohn, R. and C. F. Ansley (1986). Estimation, prediction, and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association 81*, 751–761.

Kohn, R. and C. F. Ansley (1987). A new algorithm for spline smoothing based on smoothing a stochastic process. *SIAM Journal on Scientific and Statistical Computing 8*, 33–48.

Kon, S. J. (1984). Models of stock returns – A comparison. *The Journal of Finance 39*, 147–165.

Koop, G. (2003). *Bayesian Econometrics*. Chichester: Wiley.

Koop, G. and H. K. van Dijk (2000). Testing for integration using evolving trend and seasonals models: A Bayesian approach. *Journal of Econometrics 97*, 261–291.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika 80*, 117–126.

Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association 92*, 1630–1638.

Koopman, S. J. and J. Durbin (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis 21*, 281–296.

Koski, T. (2001). *Hidden Markov Models for Bioinformatics*. Dordrecht: Kluwer Academic.

Krishnamurthy, V. and T. Rydén (1998). Consistent estimation of linear and non-linear autoregressive models with Markov regime. *Journal of Time Series Analysis 19*, 291–307.

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis.* Lecture Notes in Economics and Mathematical Systems. New York/Berlin/Heidelberg: Springer.

Krolzig, H.-M. (2001). Business cycle measurement in the presence of structural change: International evidence. *International Journal of Forecasting 17*, 349–368.

Krolzig, H.-M. and M. Sensier (2000). A disaggregated Markov-switching model of the business cycle in UK manufacturing. *The Manchester School 68*, 442–460.

Krzanowski, W. J. and Y. T. Lai (1985). A criterion for determining the number of clusters in a data set. *Biometrics 44*, 23–34.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics 22*, 79–86.

Kuttner, K. N. (1994). Estimating potential output as a latent variable. *Journal of Business & Economic Statistics 12*, 361–368.

Laird, N. M. and J. H. Ware (1982). Random-effects model for longitudinal data. *Biometrics 38*, 963–974.

Lam, P.-S. (1990). The Hamilton model with a general autoregressive component: Estimation and comparison with other models of economic time series. *Journal of Monetary Economics 26*, 409–432.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics 34*, 1–14.

Lamoureux, C. G. and W. D. Lastrapes (1990). Persistence in variance, structural change and the GARCH model. *Journal of Business & Economic Statistics 8*, 225–234.

Lavine, M. and M. West (1992). A Bayesian method for classification and discrimination. *The Canadian Journal of Statistics 20*, 451–461.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics 15*, 209–225.

Lawrance, A. J. and P. A. W. Lewis (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables (with discussions). *Journal of the Royal Statistical Society, Ser. B 47*, 165–202.

Lawrence, C. J. and W. J. Krzanowski (1996). Mixture separation for mixed-mode data. *Statistics and Computing 6*, 85–92.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, and J. A. Clausen (Eds.), *Studies in Social Psycholgy in World War II, Vol. IV: Measurement and Prediction*, pp. 362–412. Princeton, NJ: Princeton University.

Lazarsfeld, P. F. and N. W. Henry (1968). *Latent Structure Analysis.* New York: Houghton Mifflin.

Le, N. D., B. G. Leroux, and M. L. Puterman (1992). Exact likelihood evaluation in a Markov mixture model for time series of seizure counts. *Biometrics 48*, 317–323.

Lehmann, E. L. (1983). *Theory of Point Estimation.* New York: Wiley.

Leisch, F. (2004). Exploring the structure of mixture model components. In J. Antoch (Ed.), *COMPSTAT 2004. Proceedings in Computational Statistics*, pp. 1405–1412. Heidelberg: Physica-Verlag/Springer.

Lenk, P. J. and W. S. DeSarbo (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika 65*, 93–119.

Leroux, B. G. (1992a). Consistent estimation of a mixing distribution. *The Annals of Statistics 20*, 1350–1360.

Leroux, B. G. (1992b). Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and Their Applications 40*, 127–143.

Leroux, B. G. and M. L. Puterman (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics 48*, 545–558.

Levison, S. E., L. R. Rabiner, and M. M. Sondhi (1983). An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *The Bell System Technical Journal 62*, 1035–1074.

Lewis, S. M. and A. E. Raftery (1997). Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association 92*, 648–655.

Li, J. and R. M. Gray (2000). *Image Segmentation and Compression Using Hidden Markov Models.* Dordrecht: Kluwer Academic.

Li, J. Q. and A. R. Barron (2000). Mixture density estimation. Technical report, Department of Statistics, Yale University, New Haven, CT.

Li, L. A. and N. Sedransk (1988). Mixtures of distributions: A topological approach. *The Annals of Statistics 16*, 1623–1634.

Liang, Z., R. J. Jaszczak, and R. E. Coleman (1992). Parameter estimation of finite mixtures using the EM algorithm and information criteria with applications to medical image processing. *IEEE Transactions on Nuclear Science 39*, 1126–1133.

Liao, T. F. (2004). Estimating household structure in ancient China by using historical data: A latent class analysis of partially missing patterns. *Journal of the Royal Statistical Society, Ser. A 167*, 125–139.

Lindgren, G. (1978). Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics 5*, 81–91.

Lindley, D. V. (1957). A statistical paradoxon. *Biometrika 44*, 187–192.

Lindley, D. V. and A. F. M. Smith (1972). Bayes' estimates for the linear model. *Journal of the Royal Statistical Society, Ser. B 34*, 1–41.

Lindsay, B. G. (1989). Moment matrices: Applications in mixtures. *The Annals of Statistics 17*, 722–740.

Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *The Annals of Statistics 22*, 1081–1114.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry, and Applications.* Institute of Mathematical Statistics.

Lindsay, B. G. and P. Basak (1993). Multivariate normal mixtures: A fast consistent method of moments. *Journal of the American Statistical Association 88*, 468–476.

Lindsay, B. G. and K. Roeder (1992). Residual diagnostics for mixture models. *Journal of the American Statistical Association 87*, 785–794.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association 89*, 958–966.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing.* Springer Series in Statistics. New York/Berlin/Heidelberg: Springer.

Liu, J. S., W. H. Wong, and A. Kong (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika 81*, 27–40.

Liu, J. S., J. L. Zhang, M. J. Palumbo, and C. L. Lawrence (2003). Bayesian clustering with variable and transformation selection. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 7*, pp. 249–275. Oxford: Oxford University Press.

Lopes, H. F., P. Müller, and G. L. Rosner (2003). Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. *Biometrics 59*, 66–75.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Ser. B 44*, 226–233.

Lu, Z., Y. V. Hui, and A. H. Le (2003). Minimum Hellinger distance estimation for finite mixtures of Poisson regression models and its application. *Biometrics 59*, 1016–1026.

Luginbuhl, R. and A. de Vos (1999). Bayesian analysis of an unobserved-component time series model of GDP with Markov-switching and time-varying growths. *Journal of Business & Economic Statistics 17*, 456–465.

Lwin, T. and P. J. Martin (1989). Probits of mixtures. *Biometrics 45*, 721–732.

MacDonald, I. and W. Zucchini (1997). *Hidden Markov and Other Models for Discrete-valued Time Series.* Monographs on Statistics and Applied Probability. London: Chapman & Hall.

MacKay Altman, R. (2004). Assessing the goodness-of-fit of hidden Markov models. *Biometrics 60*, 444–450.

MacRae Keenan, D. (1982). A time series analysis of binary data. *Journal of the American Statistical Association 77*, 816–821.

Madigan, D. and J. York (1995). Bayesian graphical models for discrete data. *International Statistical Review 63*, 215–232.

Magill, D. T. (1965). Optimal adaptive estimation of sampled stochastic processes. *IEEE Transactions on Automatic Control 10*, 434–439.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business 36*, 394–419.

Manton, K. G., M. A. Woodbury, and E. Stallard (1981). A variance components approach to categorical data models with heterogeneous cell populations: Analysis of spatial gradients in lung cancer mortality rates in North Carolina counties. *Biometrics 37*, 259–269.

Marrs, A. D. (1998). An application of reversible-jump MCMC to multivariate spherical Gaussian mixtures. In M. I. Jordan, M. J. Kearns, and S. A. Solla (Eds.), *Advances in Neural Information Processing Systems*, Volume 10, pp. 577–583. Cambridge, MA: MIT Press.

Masreliez, C. J. (1975). Approximate non-Gaussian filtering with linear state and observation relations. *IEEE Transactions on Automatic Control 20*, 107–110.

Masreliez, C. J. and R. D. Martin (1975). Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Transactions on Automatic Control 20*, 361–371.

Mayer, J. (1999). Bayesian-type count data models with varying coefficients: estimation and testing against a parametric alternative. In H. Friedl, A. Berghold, and G. Kauermann (Eds.), *Statistical Modelling. Proceedings of the Fourteenth International Workshop on Statistical Modelling*, pp. 273–280. Graz.

McCullagh, P. and J. A. Nelder (1999). *Generalized Linear Models*. London: Chapman & Hall.

McCulloch, R. E. and P. E. Rossi (1992). Bayes factors for nonlinear hypotheses and likelihood distributions. *Biometrika 79*, 663–676.

McCulloch, R. E. and R. S. Tsay (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association 88*, 968–978.

McCulloch, R. E. and R. S. Tsay (1994a). Bayesian inference of trend- and difference-stationarity. *Econometric Theory 10*, 596–608.

McCulloch, R. E. and R. S. Tsay (1994b). Statistical analysis of economic time series via Markov switching models. *Journal of Time Series Analysis 15*, 523–539.

McLachlan, G. J. and K. E. Basford (1988). *Mixture Models: Inference and Applications to Clustering*. New York/ Basel: Marcel Dekker.

McLachlan, G. J., R. W. Bean, and D. Peel (2002). A mixture-model based approach to the clustering of microarray expression data. *Bioinformatics 18*, 413–422.

McLachlan, G. J. and D. Peel (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics. New York: Wiley.

McLachlan, G. J., D. Peel, and R. W. Bean (2003). Modelling high-dimensional data by mixtures of factor analyzers. *Computational Statisics and Data Analysis 41*, 379–388.

McQueen, G. and S. Thorely (1991). Are stock returns predictable? A test using Markov chains. *The Journal of Finance 46*, 239–263.

Meinhold, R. J. and N. D. Singpurwalla (1983). Understanding the Kalman filter. *The American Statistician 37*, 123–127.

Meinhold, R. J. and N. D. Singpurwalla (1989). Robustification of Kalman filter models. *Journal of the American Statistical Association 84*, 479–486.

Meng, X.-L. (1994). Posterior predictive $p$-values. *The Annals of Statistics 22*, 1142–1160.

Meng, X.-L. (1997). The EM algorithm and medical studies: A historical link. *Statistical Methods in Medical Research 6*, 3–23.

Meng, X.-L. and D. B. Rubin (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association 86*, 899–909.

Meng, X.-L. and S. Schilling (2002). Warp bridge sampling. *Journal of Computational and Graphical Statistics 11*, 552–586.

Meng, X.-L. and D. Van Dyk (1997). The EM algorithm — An old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society, Ser. B 59*, 511–567.

Meng, X.-L. and D. Van Dyk (1999). Seeking efficient data augmentation schemes via conditional and marginal data augmentation. *Biometrika 86*, 301–320.

Meng, X.-L. and W. H. Wong (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica 6*, 831–860.

Mengersen, K. L. and C. P. Robert (1996). Testing for mixtures: A Bayesian entropic approach. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 5 — Proceedings of the Fifth Valencia International Meeting*, pp. 255–276. Oxford: Oxford University Press.

Milligan, G. W. and M. C. Cooper (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika 50*, 159–179.

Miloslavsky, M. and M. J. van der Laan (2003). Fitting of mixtures with unspecified number of components using cross validation distance estimate. *Computational Statisics and Data Analysis 41*, 413–428.

Morbiducci, M., A. Nardi, and C. Rossi (2003). Classification of "cured" individuals in survival analysis: The mixture approach to the diagnostic-prognostic problem. *Computational Statisics and Data Analysis 41*, 515–529.

Morel, J. G. and N. K. Nagaraj (1993). A finite mixture distribution for modelling multinomial extra variation. *Biometrika 80*, 363–371.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association 78*, 47–55.

Müller, P., A. Erkanli, and M. West (1996). Bayesian curve fitting using multivarite normal mixtures. *Biometrika 83*, 67–79.

Murtagh, F. and A. E. Raftery (1984). Fitting straight lines to point patterns. *Pattern Recognition 17*, 479–483.

Muthén, B. and K. Shedden (1999). Finite mixture modelling with mixture outcomes using the EM algorithm. *Biometrics 55*, 463–469.

Nahi, N. E. (1969). Optimal recursive estimation with uncertain observations. *IEEE Transactions on Information Theory 15*, 457–462.

Nakatsuma, T. (2000). Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach. *Journal of Econometrics 95*, 57–69.

Natarajan, R. and C. E. McCulloch (1998). Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics 7*, 267–277.

Neal, R. N. (1998). Erroneous results in "Marginal likelihood from the Gibbs output.". http://www.cs.utoronto.ca/radford/radford@stat.utoronto.ca.

Neftçi, S. N. (1984). Are economic time series asymmetric over the business cycle? *Journal of Political Economy 92*, 307–328.

Nelder, J. A. and R. W. M. Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society, Ser. A 135*, 370–384.

Nelson, C. R. (1988). Spurious trend and cycle in the state space decomposition of a time series with a unit root. *Journal of Economic Dynamics and Control 12*, 475–488.

Nelson, C. R., J. Piger, and E. Zivot (2001). Markov regime switching and unit-root tests. *Journal of Business & Economic Statistics 19*, 404–415.

Nelson, C. R. and C. I. Plosser (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics 10*, 139–162.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory 6*, 318–334.

Neuhaus, J. M., W. W. Hauck, and J. D. Kalbfleisch (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrica 79*, 755–762.

Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics 8*, 343–366.

Newton, M. A. and A. E. Raftery (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (Disc: p26-48). *Journal of the Royal Statistical Society, Ser. B 56*, 3–26.

Nicholls, D. F. and A. R. Pagan (1985). Varying coefficient regression. In E. J. Hannan, P. R. Krishnaiah, and M. M. Rao (Eds.), *Handbook of Statistics*, Volume 5, Chapter 16, pp. 413–449. New York: Elsevier.

Nobile, A. (2004). On the posterior distribution of the number of components in a finite mixture. *The Annals of Statistics 32*, 2044–2073.

Nobile, A. and P. J. Green (2000). Bayesian analysis of factorial experiments by mixture modelling. *Biometrika 87*, 15–35.

Oberhofer, W. (1980). Die Nichtkonsistenz der ML Schätzer im Switching Regression Problem. *Metrika 27*, 1–13.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2004). Stochastic volatility with leverage: Fast likelihood inference. Research Report.

Oskrochi, G. R. and R. B. Davies (1997). An EM-type algorithm for multivariate mixture models. *Statistics and Computing 7*, 145–151.

Otter, T., R. Tüchler, and S. Frühwirth-Schnatter (2002). Bayesian latent class metric conjoint analysis — A case study from the Austrian mineral water market. In O. Opitz and M. Schwaiger (Eds.), *Exploratory Data Analysis in Empirical Research*, Studies in Classification, Data Analysis and Knowledge Organization, pp. 157–169. New York/Berlin/Heidelberg: Springer.

Otter, T., R. Tüchler, and S. Frühwirth-Schnatter (2004). Capturing consumer heterogeneity in metric conjoint analysis using Bayesian mixture models. *International Journal of Marketing Research 21*, 285–297.

Paap, R. and H. van Dyck (2003). Bayes estimates of Markov trends in possibly cointegrated series: An application to U.S. consumption and income. *Journal of Business & Economic Statistics 21*, 547–563.

Pagan, A. R. and G. W. Schwert (1990). Alternative models for conditional stock volatility. *Journal of Econometrics 45*, 267–290.

Paul, S. R., K. Y. Liang, and S. G. Self (1989). On testing departure from the binomial and multinomial assumptions. *Biometrics 45*, 231–236.

Pauler, D. K., M. D. Escobar, J. A. Sweeney, and J. Greenhouse (1996). Mixture models for eye-tracking data: A case study. *Statistics in Medicine 15*, 1365–1376.

Peña, D. and I. Guttman (1988). Bayesian approach to robustifying the Kalman filter. In J. C. Spall (Ed.), *Bayesian Analysis of Time Series and Dynamic Linear Models*, pp. 227–253. New York: Marcel Dekker.

Peña, D. and I. Guttman (1993). Comparing probabilistic methods for outlier detection in linear models. *Biometrika 80*, 603–610.

Peña, D. and G. C. Tiao (1992). Bayesian robustness functions for linear models. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*, pp. 365–388. Oxford: Oxford University Press.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London A 185*, 71–110.

Pearson, K. (1915). On certain types of compound frequency distributions in which the components can be individually described by binomial series. *Biometrika 11*, 139–144.

Peel, D. and G. McLachlan (2000). Robust mixture modelling using the $t$ distribution. *Statistics and Computing 10*, 339–348.

Peng, F., R. A. Jacobs, and M. A. Tanner (1996). Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition. *Journal of the American Statistical Association 91*, 953–960.

Peria, M. S. M. (2002). A regime-switching aproach to the study of speculative attacks: A focus on the EMS crisis. *Empirical Economics 27*, 299–334.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica 57*, 1361–1401.

Perron, P. (1990). Testing for a unit root in a time series with a changing mean. *Journal of Business & Economic Statistics 8*, 153–162.

Peskun, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika 60*, 607–612.

Petrie, T. (1969). Probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics 40*, 97–115.

Pfeiffer, R. and N. Jeffries (1999). A mixture model for the probability distribution of rain rate. In H. Friedl, A. Berghold, and G. Kauermann (Eds.), *Statistical Modelling. Proceedings of the Fourteenth International Workshop on Statistical Modelling*, pp. 610–615. Graz.

Phillips, D. B. and A. F. M. Smith (1996). Bayesian model comparison via jump diffusions. In W. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 215–239. London: Chapman & Hall.

Phillips, K. L. (1991). A two-country model of stochastic output with changes in regime. *Journal of International Economics 31*, 121–142.

Pitt, M. K. and N. Shephard (1999). Analytic convergence rates and parameterization issues for the Gibbs sampler applied to state space models. *Journal of Time Series Analysis 20*, 63–85.

Ploberger, W., W. Kramer, and K. Kontrus (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics 40*, 307–318.

Poskitt, D. S. and S.-H. Chung (1996). Markov chain models, time series analysis and extreme value theory. *Advances in Applied Probability 28*, 405–425.

Posse, C. (2001). Hierarchical model-based clustering for large data sets. *Journal of Computational and Graphical Statistics 10*, 464–486.

Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics 9*, 705–724.

Press, S. J. (2003). *Subjective and Objective Bayesian Statistics* (2 ed.). New York: Wiley.

Psaradakis, Z. (2001). Markov-level shifts and the unit-root hypothesis. *The Econometrics Journal 4*, 225–241.

Psaradakis, Z. (2002). On the asymptotic behaviour of unit root tests in the presence of a Markov trend. *Statistics & Probability Letters 57*, 101–109.

Qian, W. and D. M. Titterington (1991). Estimation of parameters in hidden Markov models. *Philosophical Transactions of the Royal Society of London, Series A, Math and Physical Science 337*, 407–428.

Quandt, R. E. (1958). The estimation of the parameter of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association 53*, 873–880.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association 55*, 324–330.

Quandt, R. E. (1972). A new approach to estimating switching regressions. *Journal of the American Statistical Association 67*, 306–310.

Quandt, R. E. and J. B. Ramsey (1978). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association 73*, 730–752.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*, 257–286.

Rabiner, L. R. and B. H. Juang (1986). An introduction to hidden Markov models. *IEEE Transactions on Acoustics, Speech, and Signal Processing 34*, 4–16.

Raftery, A., D. Madigan, and J. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association 92*, 179–191.

Raftery, A. E. (1996a). Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika 83*, 251–266.

Raftery, A. E. (1996b). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 163–188. London: Chapman & Hall.

Raj, B. (2002). Asymmetries of business cycle: The Markov switching approach. In A. Ullah, A. Wan, and A. Chaturvedi (Eds.), *Handbook of Applied Econometrics and Statistical Inference*, pp. 687–710. New York: Marcel Dekker.

Ramaswamy, V., E. W. Anderson, and W. S. DeSarbo (1994). Negative binomial models for count data. *Management Science 40*, 405–412.

Ramaswamy, V., W. S. DeSarbo, D. Reibstein, and W. T. Robinson (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science 12*, 103–124.

Rao, C. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Ser. B 10*, 159–203.

Rao, C. R. (1975). Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics 31*, 545–554.

Ray, S. and B. Lindsay (2005). The topography of multivariate normal mixtures. *The Annals of Statistics 33*, 2042–2065.

Redner, R. A. and H. Walker (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review 26*, 195–239.

Richardson, S. and P. J. Green (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Ser. B 59*, 731–792.

Ripley, B. D. (1977). Modelling spatial patterns (C/R: P192-212). *Journal of the Royal Statistical Society, Ser. B 39*, 172–192.

Robert, C. P. (1996). Mixtures of distributions: Inference and estimation. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 441–464. London: Chapman & Hall.

Robert, C. P. and G. Casella (1999). *Monte Carlo Statistical Methods.* Springer Series in Statistics. New York/Berlin/Heidelberg: Springer.

Robert, C. P., G. Celeux, and J. Diebolt (1993). Bayesian estimation of hidden Markov chains: A stochastic implementation. *Statistics & Probability Letters 16*, 77–83.

Robert, C. P. and K. L. Mengersen (1999). Reparameterisation issues in mixture modelling and their bearing on MCMC algorithms. *Computational Statistics and Data Analysis 29*, 325–343.

Robert, C. P., T. Rydén, and D. M. Titterington (1999). Convergence controls for MCMC algorithms, with applications to hidden Markov chains. *Journal of Statistical Computation and Simulation 64*, 327–355.

Robert, C. P., T. Rydén, and D. M. Titterington (2000). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society, Ser. B 62*, 57–75.

Robert, C. P. and D. M. Titterington (1998). Reparameterization strategies for hidden Markov models and Bayesian approaches to maximum likelihood estimation. *Statistics and Computing 8*, 145–158.

Robertson, C. A. and J. G. Fryer (1969). Some descriptive properties of normal mixtures. *Skand. Aktuarietidskr. 52*, 137–146.

Roeder, K. and L. Wasserman (1997a). Discussion of the paper by Richardson and Green "On Bayesian analysis of mixtures with an unknown number of components.". *Journal of the Royal Statistical Society, Ser. B 59*, 782.

Roeder, K. and L. Wasserman (1997b). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association 92*, 894–902.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2005). *Bayesian Statistics and Marketing.* Chichester: Wiley.

Rothenberg, T. (1971). Identification in parametric models. *Econometrica 39*, 577–591.

Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics 9*, 130–134.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics 12*, 1151–1172.

Rue, H. (1995). New loss functions in Bayesian imaging. *Journal of the American Statistical Association 90*, 900–908.

Rydén, T., T. Teräsvirta, and S. Åsbrink (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics 13*, 217–244.

Rydén, T. and D. Titterington (1998). Computational Bayesian analysis of hidden Markov models. *Journal of Computational and Graphical Statistics 7*, 194–211.

Sahu, S. K. and R. C. H. Cheng (2003). A fast distance based approach for determining the number of components in mixtures. *The Canadian Journal of Statistics 31*, 3–22.

Satagopan, J. M., M. A. Newton, and A. E. Raftery (2000). Easy estimation of normalizing constants and Bayes factors from posterior simulations: Stabilizing the harmonic mean estimator. Technical Report 382, Department of Statistics, University of Washington.

Scaccia, L. and P. J. Green (2003). Bayesian growth curves using normal mixtures with nonparametric weights. *Journal of Computational and Graphical Statistics 12*, 308–331.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika 78*, 719–727.

Schaller, H. and S. van Norden (2002). Fads or bubbles? *Empirical Economics 27*, 335–362.

Schervish, M. J. and R. Tsay (1988). Bayesian modeling and forecasting in autoregressive models. In J. C. Spall (Ed.), *Bayesian Analysis of Time Series and Dynamic Linear Models*, pp. 23–53. New York: Marcel Dekker.

Schlattmann, P. and D. Böhning (1993). Mixture models and disease mapping. *Statistics in Medicine 12*, 1943–1950.

Schlattmann, P., E. Dietz, and D. Böhning (1996). Covariate adjusted mixture models and disease mapping with the program DismapWin. *Statistics in Medicine 15*, 919–929.

Schnatter, S. (1988a). Bayesian forecasting of time series by Gaussian sum approximation. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 3*, pp. 757–764. London: Clarendon.

Schnatter, S. (1988b). *Dynamic Bayesian Models with Applications to Hydrological Short-Term Forecasting (in German).* Ph. D. thesis, Vienna University of Technology, Austria.

Schnatter, S., D. Gutknecht, and R. Kirnbauer (1987). A Bayesian approach to estimating the parameters of a hydrological forecasting system. In R. Viertl (Ed.), *Probability and Bayesian Statistics*, pp. 415–422. New York: Plenum.

Schneider, W. (1988). Analytical uses of Kalman filtering in econometrics: A survey. *Statistical Papers 29*, 3–33.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*, 461–464.

Schweppe, F. C. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory 11*, 294–305.

Sclove, S. L. (1983). Time series segmentation: A model and a method. *Information Science 29*, 7–25.

Scott, A. J. and M. Symons (1971). Clustering methods based on likelihood ratio criteria. *Biometrics 27*, 387–397.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* New York: Wiley.

Scott, D. W. (2004). Outlier detection and clustering by partial mixture modelling. In J. Antoch (Ed.), *COMPSTAT 2004. Proceedings in Computational Statistics*, pp. 453–464. Heidelberg: Physica-Verlag/Springer.

Scott, D. W. and W. F. Szewczyk (2001). From kernels to mixtures. *Technometrics 43*, 323–335.

Scott, S. L. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association 97*, 337–351.

Scott, S. L., G. M. James, and C. A. Sugar (2005). Hidden Markov models for longitudinal comparisons. *Journal of the American Statistical Association 100*, 359–369.

Shaked, M. (1980). On mixtures from exponential families. *Journal of the Royal Statistical Society, Ser. B 42*, 192–198.

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

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (Eds.), *Time Series Models: In Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman & Hall.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–667.

Shively, T. S. and R. Kohn (1997). A Bayesian approach to model selection in stochastic coefficient regression models and structural time series models. *Journal of Econometrics 76*, 39–52.

Shumway, R. H. and D. S. Stoffer (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis 3*, 253–264.

Shumway, R. H. and D. S. Stoffer (1991). Dynamic linear models with switching. *Journal of the American Statistical Association 86*, 763–769.

Shumway, R. H. and D. S. Stoffer (2000). *Time Series Analysis and Its Applications*. New York: Springer.

Sichel, D. E. (1994). Inventories and the three phases of the business cycle. *Journal of Business & Economic Statistics 12*, 269–277.

Silverman, B. W. (1999). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *The Annals of Statistics 4*, 1200–1209.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*, 1–48.

Sims, F. L. and D. G. Lainiotis (1969). Recursive algorithm for the calculation of the adaptive Kalman filter weighting coefficients. *IEEE Transactions on Automatic Control 14*, 215–218.

Slud, E. (1997). Testing for imperfect debugging in software reliability. *Scandinavian Journal of Statistics 24*, 555–572.

Smith, A. F. M. and G. O. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (Disc: p53-102). *Journal of the Royal Statistical Society, Ser. B 55*, 3–23.

Smith, A. F. M. and D. J. Spiegelhalter (1980). Bayes factors and choice criteria for linear models. *Journal of the Royal Statistical Society, Ser. B 42*, 213–220.

Smith, A. F. M. and M. West (1983). Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics 39*, 867–878.

Smith, D. R. (2002). Markov switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business & Economic Statistics 20*, 183–197.

Solka, J. L., E. J. Wegman, C. E. Priebe, W. L. Poston, and G. W. Rogers (1998). Mixture structure analysis using the Akaike information criterion and the bootstrap. *Statistics and Computing 8*, 177–188.

Sorenson, H. W. and D. L. Alspach (1971). Recursive Bayesian estimation using Gaussian sums. *Automatica 6*, 465–479.

Spiegelhalter, D. J., K. R. Abrams, and J. P. Myles (2003). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* Chichester: Wiley.

Stephens, M. (1997a). *Bayesian Methods for Mixtures of Normal Distributions.* Ph. D. thesis, University of Oxford.

Stephens, M. (1997b). Discussion of the paper by Richardson and Green. *Journal of the Royal Statistical Society, Ser. B 59*, 768–769.

Stephens, M. (2000a). Bayesian analysis of mixture models with an unknown number of components – An alternative to reversible jump methods. *The Annals of Statistics 28*, 40–74.

Stephens, M. (2000b). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Ser. B 62*, 795–809.

Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*, 147–162.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Ser. B 36*, 111–147.

Sugar, C. A. and G. M. James (2003). Finding the number of clusters in a data set: An information-theoretic approach. *Journal of the American Statistical Association 98*, 750–763.

Symons, M. J. (1981). Clustering criteria and multivariate normal mixtures. *Biometrics 37*, 35–43.

Tadesse, M. G., N. Sha, and M. Vannucci (2005). Bayesian variable selection in clustering high-dimensional data. *Journal of the American Statistical Association 100*, 602–617.

Tan, W. Y. and W. C. Chang (1972). Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities. *Journal of the American Statistical Association 67*, 702–708.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association 82*, 528–540.

Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics 51*, 899–907.

Teicher, H. (1960). On the mixture of distributions. *The Annals of Mathematical Statistics 31*, 55–73.

Teicher, H. (1961). Identifiability of mixtures. *The Annals of Mathematical Statistics 32*, 244–248.

Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics 34*, 1265–1269.

Teicher, H. (1967). Identifiability of mixtures of product measures. *The Annals of Mathematical Statistics 38*, 1300–1302.

Teräsvirta, T. and H. M. Anderson (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics 7*, S119–S136.

Thompson, P. A. and R. B. Miller (1986). Sampling the future: A Bayesian approach to forecasting from univariate time series models. *Journal of Business & Economic Statistics 4*, 427–436.

Thyer, M. and G. Kuczera (2000). Modelling long-term persistence in hydro-climatic time series using a hidden state Markov model. *Water Resources Research 36*, 3301–3310.

Tibshirani, R., G. Walther, and T. Hastie (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, Ser. B 63*, 411–423.

Tierney, L. (1994). Markov chain for exploring posterior distributions. *The Annals of Statistics 22*, 1701–1762.

Tierney, L., R. E. Kass, and J. B. Kadane (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association 84*, 710–716.

Timmermann, A. (2000). Moments of Markov switching models. *Journal of Econometrics 96*, 75–111.

Tipping, M. E. and C. M. Bishop (1999). Mixtures of probabilistic principal component analysers. *Neural Computation 11*, 443–482.

Titterington, D. M. (1990). Some recent research in the analysis of mixture distributions. *Statistics 21*, 619–641.

Titterington, D. M., A. F. M. Smith, and U. E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley Series in Probability and Statistics. New York: Wiley.

Tjøstheim, D. (1986). Some doubly stochastic time series models. *Journal of Time Series Analysis 7*, 51–72.

Tong, H. (1990). *Non-linear Time Series. A Dynamical System Approach.* Oxford: Oxford University Press.

Tsai, C. and L. Kurz (1983). An adaptive robustizing approach to Kalman filtering. *Automatica 19*, 279–288.

Tsay, R. S. (1987). Conditional heteroscedastic time series models. *Journal of the American Statistical Association 82*, 590–604.

Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting 7*, 1–20.

Tucker, A. (1992). A reexamination of finite- and infinite-variance distributions as models of daily stock returns. *Journal of Business & Economic Statistics 10*, 73–81.

Tugnait, J. K. (1982). Detection and estimation for abruptly changing systems. *Automatica 18*, 607–615.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, and H. Mann (Eds.), *Contributions to Probability and Statistics*, pp. 448–485. Stanford, CA: Stanford University Press.

Turner, C. M., R. Startz, and C. R. Nelson (1989). A Markov model of heteroscedasticity, risk, and learning in the stock market. *Journal of Financial Economics 25*, 3–22.

Utsugi, A. and T. Kumagai (2001). Bayesian analysis of mixtures of factor analyzers. *Neural Computation 13*, 993–1002.

van Dyk, D. and X.-L. Meng (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics 10*, 1–50.

Verbeke, G. and E. Lesaffre (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association 91*, 217–221.

Verbeke, G. and E. Lesaffre (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis 23*, 541–556.

Verbeke, G. and G. Molenberghs (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. New York/Berlin/Heidelberg: Springer.

Verdinelli, I. and L. Wasserman (1991). Bayesian analysis of outlier problems using the Gibbs sampler. *Statistics and Computing 1*, 105–117.

Verdinelli, I. and L. Wasserman (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association 90*, 614–618.

Viallefont, V., S. Richardson, and P. J. Green (2002). Bayesian analysis of Poisson mixtures. *Journal of Nonparametric Statistics 14*, 181–202.

Waagepetersen, R. and D. Sorensen (2001). A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping. *International Statistical Review 69*, 49–61.

Wang, J. and E. Zivot (2000). A Bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business & Economic Statistics 18*, 374–386.

Wang, P. and M. L. Puterman (1998). Mixed logistic regression models. *Journal of Agricultural, Biological, and Environmental Statistics 3*, 175–200.

Wang, P. and M. L. Puterman (1999). Markov Poisson regression models for discrete time series. Part I: Methodology. *Journal of Applied Statistics 26*, 855–869.

Wang, P., M. L. Puterman, I. Cockburn, and N. Le (1996). Mixed Poisson regression models with covariate dependent rates. *Biometrics 52*, 381–400.

Wasserman, L. (2000). Asymptotic inference for mixture models using data-dependent priors. *Journal of the Royal Statistical Society, Ser. B 62*, 159–180.

Watier, L., S. Richardson, and P. J. Green (1999). Using Gaussian mixtures with unknown number of components for mixed model estimation. In H. Friedl, A. Berghold, and G. Kauermann (Eds.), *Statistical Modelling. Proceedings of the Fourteenth International Workshop on Statistical Modelling*, pp. 394–401. Graz.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics 23*, 385–400.

Wecker, W. E. and C. Ansley (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association 78*, 81–89.

Wedel, M. and W. S. DeSarbo (1993a). A latent class binomial logit methodology for the analysis of paired comparison data. *Decision Sciences 24*, 1157–1170.

Wedel, M. and W. S. DeSarbo (1993b). A review of recent developments in latent class regression models. In R. P. Bagozzi (Ed.), *Advanced Methods of Marketing Research*, pp. 352–388. Oxford: Blackwell.

Wedel, M. and W. S. DeSarbo (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification 12*, 21–55.

Wedel, M., W. S. DeSarbo, J. R. Bult, and V. Ramaswamy (1993). A latent class Poisson regression model for heterogeneous count data. *Journal of Applied Econometrics 8*, 397–411.

Wedel, M. and J.-B. Steenkamp (1991). A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research 28*, 385–96.

West, M. (1981). Robust sequential approximate Bayesian estimation. *Journal of the Royal Statistical Society, Ser. B 43*, 157–166.

West, M. (1984). Outlier models and prior distribution in Bayesian linear regression. *Journal of the Royal Statistical Society, Ser. B 46*, 431–439.

West, M. (1985). Generalized linear models: Scale parameters, outlier accommodation and prior distributions. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 2*, pp. 531–558. New York: Elsevier.

West, M. (1992). Modelling with mixtures. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*, pp. 503–524. Oxford: Oxford University Press.

West, M. (1993). Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society, Ser. B 55*, 409–422.

West, M. and P. J. Harrison (1997). *Bayesian Forecasting and Dynamic Models* (2 ed.). New York: Springer.

Whittaker, J. and S. Frühwirth-Schnatter (1994). A dynamic changepoint model for detecting the onset of growth in bacteriological infections. *Applied Statistics 43*, 625–640.

Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics 9*, 60–62.

Willse, A. and R. J. Boik (1999). Identifiable finite mixtures of location models for clustering mixed-mode data. *Statistics and Computing 9*, 111–121.

Windham, M. P. and A. Cutler (1992). Information ratios for validating mixture analyses. *Journal of the American Statistical Association 87*, 1188–1192.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research 5*, 329–350.

Wong, C. S. and W. K. Li (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society, Ser. B 62*, 95–115.

Wong, C. S. and W. K. Li (2001). On a mixture autoregressive conditional heteroscedastic model. *Journal of the American Statistical Association 96*, 982–995.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics 11*, 95–103.

Yakowitz, S. J. and J. D. Spragins (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics 39*, 209–214.

Yao, J. and J. Attali (2000). On stability of nonlinear AR processes with Markov switching. *Advances in Applied Probability 32*, 394–407.

Yau, K. K. W., A. H. Lee, and A. S. K. Ng (2003). Finite mixture regression model with random effects: Application to neonatal hospital lenght of stay. *Computational Statisics and Data Analysis 41*, 359–366.

Yeung, K. Y., C. Fraley, A. Murua, A. E. Raftery, and W. L. Ruzzo (2001). Model-based clustering and data transformations for gene expression data. *Bioinformatics 17*, 977–987.

Yeung, K. Y. and W. L. Ruzzo (2001). Principal component analysis for clustering gene expression data. *Bioinformatics 17*, 763–774.

Zeger, S. L. and B. Qaqish (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics 44*, 1019–1031.

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

Zhang, H. P. and K. Merikangas (2000). A frailty model of segregation analysis: Understanding the familial transmission of alcoholism. *Biometrics 56*, 815–823.

Zhu, H.-T. and H. Zhang (2004). Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society, Ser. B 66*, 3–16.

Zucchini, W. and P. Guttorp (1991). A hidden Markov model for space-time precipitation. *Water Resources Research 27*, 1917–1923.

# Index

# Springer Series in Statistics

*(Continued from page ii)*

*Heyde:* Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation.
*Huet/Bouvier/Poursat/Jolivet:* Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples, 2nd edition.
*Ibrahim/Chen/Sinha:* Bayesian Survival Analysis.
*Jolliffe:* Principal Component Analysis, 2nd edition.
*Knottnerus:* Sample Survey Theory: Some Pythagorean Perspectives.
*Küchler/Sørensen:* Exponential Families of Stochastic Processes.
*Kutoyants:* Statistical Inference for Ergodic Diffusion Processes.
*Lahiri:* Resampling Methods for Dependent Data.
*Le Cam:* Asymptotic Methods in Statistical Decision Theory
*Le Cam/Yang:* Asymptotics in Statistics: Some Basic Concepts, 2nd edition.
*Le/Zidek:* Statistical Analysis of Environmental Space-time Processes
*Liu:* Monte Carlo Strategies in Scientific Computing.
*Manski:* Partial Identification of Probability Distributions.
*Mielke/Berry:* Permutation Methods: A Distance Function Approach.
*Molenberghs/Verbeke:* Models for Discrete Longitudinal Data.
*Mukerjee/Wu:* A Modern Theory of Factorial Designs.
*Nelsen:* An Introduction to Copulas, 2nd edition.
*Pan/Fang:* Growth Curve Models and Statistical Diagnostics.
*Politis/Romano/Wolf:* Subsampling.
*Ramsay/Silverman:* Applied Functional Data Analysis: Methods and Case Studies.
*Ramsay/Silverman:* Functional Data Analysis, 2nd edition.
*Reinsel:* Elements of Multivariate Time Series Analysis, 2nd edition.
*Rosenbaum:* Observational Studies, 2nd edition.
*Rosenblatt:* Gaussian and Non-Gaussian Linear Time Series and Random Fields.
*Särndal/Swensson/Wretman:* Model Assisted Survey Sampling.
*Santner/Williams/Notz:* The Design and Analysis of Computer Experiments.
*Schervish:* Theory of Statistics
*Shaked/Shanthikumar:* Stochastic Orders.
*Shao/Tu:* The Jackknife and Bootstrap.
*Simonoff:* Smoothing Methods in Statistics
*Sprott:* Statistical Inference in Science.
*Stein:* Interpolation of Spatial Data: Some Theory for Kriging.
*Taniguchi/Kakizawa:* Asymptotic Theory for Statistical Inference for Time Series.
*Tanner:* Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, 3rd edition.
*Tillé:* Sampling Algorithms.
*Tsaitis:* Semiparametric Theory and Missing Data.

*van der Laan/Robins:* Unified Methods for Censored Longitudinal Data and Causality.
*van der Vaart/Wellner:* Weak Convergence and Empirical Processes: With Applications to Statistics.
*Verbeke/Molenberghs:* Linear Mixed Models for Longitudinal Data.
*Weerahandi:* Exact Statistical Methods for Data Analysis.