The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 225-233

Bayesian Analysis under Heavy-Tailed Priors in Finite Population Sampling[1]

Dal Ho Kim[2], In Suk Lee[3], Joong Kweon Sohn[4], Jang Sik Cho[5]

Abstract

In this paper, we propose Bayes estimators of the finite population mean based on heavy-tailed prior distributions using scale mixtures of normals. Also, the asymptotic optimality property of the proposed Bayes estimators is proved. A numerical example is provided to illustrate the results.

1. Introduction

Consider a finite population U with units labeled $1,2,\ldots,N$. Let y_i denote the value of a single characteristic attached to the unit i. The vector $y=(y_1,\ldots,y_N)^T$ is the unknown state of nature, and is assumed to belong to $\Theta=R^N$. A subset s of $\{1,\ldots,N\}$ is called a sample. Let $n(s)$ denote the number of elements belonging to s. The set of all possible samples is denoted by S. A design is a function p on S such that $p(s)\in[0,1]$ for all $s\in S$ and $\sum_{s\in S}p(s)=1$. Given $y\in\Theta$ and $s=\{i_1,\ldots,i_{n(s)}\}$ with $1\le i_1<\ldots<i_{n(s)}\le N$, let $y(s)=\{y_{i_1},\ldots,y_{i_{n(s)}}\}$. One of the main objectives in sample surveys is to draw inference about y or some function (real or vector valued) $\gamma(y)$ of y on the basis of s and $y(s)$.

A unified and elegant formulation of Bayes estimation in finite population sampling was given by Ericson(1969). Since then, there are many papers in the area of Bayes estimation in finite population sampling. However, Ericson(1969) and others assumed the conjugate normal priors for normal superpopulation models.

---

1) The Present Studies were Supported (in part) by the Basic Research Institute Program, Ministry of Education, 1995, Project No. BSRI-95-1403.
2) Full-time Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.
3) Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.
4) Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.
5) Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

Substantial evidence has been presented to the effect that priors with tails that are flatter than those of the likelihood function tend to be fairly robust (e.g., Box and Tiao(1968, 1973), Dawid(1973), O'Hagan(1979, 1989) and West(1985)). Priors which are scale mixtures of normal have flatter tails than those of the normal automatically by construction. This class of priors includes the Student t family, double exponential, logistic, and the exponential power family of Box and Tiao(1973) among others. Andrews and Mallows(1974) and West(1987) used such distributions for the simulation and the analysis of outlier models.

The price to be paid for utilization of such heavy-tailed priors is computational; closed form calculation is no longer possible. Recently, however, the Markov chain Monte Carlo integration techniques, in particular the Gibbs sampling(Geman and Geman(1984), Gelfand and Smith(1990), and Gelfand et al.(1990)) has proved to be a simple yet powerful tool for performing Bayes computations.

The outline of the remaining sections is as follows. In Section 2, we provide the Bayes estimators of the finite population mean based on heavy-tailed priors using scale mixtures of normals in the normal superpopulation model. Also, the asymptotic optimality(A.O.) in the sense of Robbins(1964) of proposed Bayes estimators is proved. In Section 3, a numerical example is provided to illustrate the results of the Section 2.

For simplicity, in the subsequent sections, only the case when $n(s) \neq n \Rightarrow p(s) = 0$ is considered. This amounts to considering only fixed samples of size n. Also, throughout the loss is assumed to be squared error.

2. Bayes Estimation Under Heavy-Tailed Priors

Ericson(1969) considered the superpopulation model $y_i = \theta + \varepsilon_i$, $(i = 1, \cdots, N)$, where θ, $\varepsilon_1, \cdots, \varepsilon_N$ are independently distributed with $\theta \sim N(\mu_0, \tau_0^2)$ and ε_i's are iid $N(0, \sigma^2)$. Write $f = n/N$, $M_0 = \sigma^2/\tau_0^2$, $B_0 = M_0/(M_0 + n)$, $\bar{y}(s) = n^{-1} \sum_{i \in s} y_i$, and $y(\bar{s}) = \{ y_i : i \notin s \}$, the suffixes in $y(\bar{s})$ being arranged in ascending order. Under the $N(\mu_0, \tau_0^2)$ prior, the Bayes estimator of $\gamma(\mathbf{y}) = N^{-1} \sum_{i=1}^{N} y_i$ is

$$\delta^0(s, y(s)) = f\bar{y}(s) + (1-f)\{(1-B_0)\bar{y}(s) + B_0\mu_0\}. \tag{2.1}$$

Also, the associated posterior variance is given by

$$V(\gamma(\mathbf{y}) \mid s, y(s)) = N^{-2}(N-n)\sigma^2(M_0+N)/(M_0+n). \tag{2.2}$$

Now, we consider a refinement based on heavy tailed priors on θ using scale mixtures of normals. That is, consider the case when (i) $y_i \mid \theta \overset{iid}{\sim} N(\theta, \sigma^2)$ $(i=1,\cdots,N)$ and (ii) $\theta \sim \frac{1}{\tau_0} p\left(\frac{\theta-\mu_0}{\tau_0}\right)$ where $p(x) = \int_0^\infty \lambda^{1/2}\phi(x\lambda^{1/2})g(\lambda)d\lambda$ and $\phi(x)$ denotes the standard normal pdf. Here $p(\cdot)$ is a scale mixture of the normal distribution with mixing distribution $g(\cdot)$. Note that we can write (ii) in the following two step; (iia) $\theta \mid \lambda \sim N(\mu_0, \lambda^{-1})$ and (iib) $\lambda \sim \tau_0^2 g(\tau_0^2\lambda)$ where $\int_0^\infty g(x)dx = 1$. The following list identifies the necessary functional form for $g(\lambda)$ to obtain a wide range of densities which represent departures from normality:

t-priors: If $k\lambda \sim \chi_k^2$ then θ is Student t with k degrees of freedom, location parameter μ_0, and scale parameter τ_0.

double exponential priors: If $1/\lambda$ has exponential distribution with mean 2 then θ is double exponential with location parameter μ_0 and scale parameter τ_0.

exponential power family priors: If λ has positive stable distribution with index $\alpha/2$ then θ has exponential power distribution with location parameter μ_0 and scale parameter τ_0.

logistic priors: If $\sqrt{\lambda}$ has the asymptotic Kolmogorov distance distribution then θ is logistic with location parameter μ_0 and scale parameter τ_0. [A random variable Z is said to have an asymptotic Kolmogorov distance distribution if it has a pdf of the form $f(z) = 8z\sum_{j=1}^{\infty}(-1)^{j-1}j^2\exp(-2j^2z^2)I_{(0,\infty)}(z)$].

Then the posterior distribution of $y(\bar{s})$ given by s and $y(s)$ is obtained as follows:

(i) conditional on s, $y(s)$ and λ, $y(\bar{s})$ has

$$N\left((B(\lambda)\mu_0 + (1-B(\lambda))\bar{y}(s))\ \mathbf{1}_{N-n},\ \sigma^2\left(\ I_{N-n} + \frac{1}{\lambda\sigma^2+n}\ \mathbf{1}_{N-n}\ \ \mathbf{1}^T_{N-n}\right)\right), \tag{2.3}$$

where $B(\lambda) = \lambda\sigma^2/(\lambda\sigma^2+n)$;

(ii) the conditional distribution of λ given s and $y(s)$ has pdf

$$f(\lambda \mid s, y(s)) \propto (\sigma^2 + n\lambda^{-1})^{-1/2}\exp\left[-\frac{n(\bar{y}(s)-\mu_0)^2}{2(\sigma^2+n\lambda^{-1})}\right]g(\tau_0^2\lambda). \tag{2.4}$$

Note that under the posterior distribution given in (2.3) and (2.4), the Bayes estimator of $\gamma(y) = N^{-1}\sum_{i=1}^{N} y_i$ is given by

$$\delta^{SM}(s, y(s)) = E[\gamma(\mathbf{y}) \mid s, y(s)] = E[E[\gamma(\mathbf{y}) \mid s, y(s), \lambda] \mid s, y(s)]$$
$$= f\bar{y}(s) + (1-f)\{E[B(\lambda) \mid s, y(s)]\mu_0 + (1 - E[B(\lambda) \mid s, y(s)])\bar{y}(s)\} \qquad (2.5)$$

Also, one gets

$$V(\gamma(\mathbf{y}) \mid s, y(s))$$
$$= E[V(\gamma(\mathbf{y}) \mid s, y(s), \lambda) \mid s, y(s)] + V[E(\gamma(\mathbf{y}) \mid s, y(s), \lambda) \mid s, y(s)]$$
$$= N^{-2}\sigma^2\{(N-n) + (N-n)^2 E((\lambda\sigma^2 + n)^{-1} \mid s, y(s))\}$$
$$+ (1-f)^2 V(B(\lambda)\mu_0 + (1-B(\lambda))\bar{y}(s) \mid s, y(s)). \qquad (2.6)$$

The calculations in (2.5) and (2.6) can be performed using one-dimensional numerical integration. Alternatively, one can use Monte Carlo numerical integration techniques to generate the posterior distribution and the associated means and variances. More specifically, in this section, we use Gibbs sampling originally introduced in Geman and Geman(1984), and more recently popularized by Gelfand and Smith(1990) and Gelfand et al.(1990).

Using Gibbs sampling, the posterior distribution of $y(\bar{s})$ is approximated by

$$q^{-1}\sum_{j=1}^{q}[y(\bar{s}) \mid s, y(s), \theta = \theta_j, \lambda = \lambda_j]. \qquad (2.7)$$

To estimate the posterior moments, we use Rao-Blackwellized estimates as in Gelfand and Smith(1991). Note that $E[\gamma(\mathbf{y}) \mid s, y(s)]$is approximated by

$$f\bar{y}(s) + (1-f)q^{-1}\sum_{j=1}^{q}(B(\lambda_j)\mu_0 + (1-B(\lambda_j))\bar{y}(s)). \qquad (2.8)$$

Next one approximates $V(\gamma(\mathbf{y}) \mid s, y(s))$ by

$$N^{-2}\sigma^2\{(N-n) + (N-n)^2 q^{-1}\sum_{j=1}^{q}(\lambda_j\sigma^2 + n)^{-1}\}$$
$$+ (1-f)^2\Big[q^{-1}\sum_{j=1}^{q}(B(\lambda_j)\mu_0 + (1-B(\lambda_j))\bar{y}(s))^2$$
$$- \{q^{-1}\sum_{j=1}^{q}(B(\lambda_j)\mu_0 + (1-B(\lambda_j))\bar{y}(s)\}^2\Big]. \qquad (2.9)$$

The Gibbs sampling analysis is based on the following posterior distributions:

(i) $\theta \mid s, y(s), y(\bar{s}), \lambda \sim N\Big[(\lambda\mu_0 + \sum_{i=1}^{N} y_i/\sigma^2)/(\lambda + N/\sigma^2),\ (\lambda + N/\sigma^2)^{-1}\Big]$;

(ii) $f(\lambda \mid s, y(s), y(\bar{s}), \theta) \propto \sqrt{\lambda}\exp[-\frac{\lambda}{2}(\theta - \mu_0)^2]g(\tau_0^2\lambda)$;

(iii) $y(\bar{s}) \mid s, y(s), \theta, \lambda \sim N[\theta\, \mathbf{1}_{N-n},\ \sigma^2\, \mathbf{I}_{N-n}]$.

Note that if $k\lambda \sim \chi^2{}_k$ then $f(\lambda \mid s, y(s), y(\bar{s}), \theta)$ reduces to a Gamma $\left(\frac{1}{2}\{\tau_0^2 k+(\theta-\mu_0)^2\},\ \frac{1}{2}(k+1)\right)$ density. [A random variable W is said to have a Gamma (α, β) distribution if it has a pdf of the form $f(w) \propto \exp(-\alpha w) w^{\beta-1} I_{(0,\infty)}(w)$, where I denotes the usual indicator function]. Also, if $1/\lambda$ has exponential distribution with mean 2 then $f(\lambda \mid s, y(s), y(\bar{s}), \theta)$ reduces to a IGN $(1/\sqrt{\tau_0^2(\theta-\mu_0)^2},\ 1/\tau_0^2)$ density. [A random variable V is said to have a IGN (η_1, η_2) distribution if it has a pdf of the form $f(v)=\sqrt{\frac{\eta_2}{2\pi}}\, v^{-3/2} \exp\left(-\frac{\eta_2(v-\eta_1)^2}{2\eta_1^2 v}\right) I_{(0,\infty)}(v)$].

We shall now evaluate the performance of the Bayes estimator δ^{SM} of θ for large n under the $N(\mu_0, \tau_0^2)$ prior, say π_0. The Bayes estimator of θ under this prior is δ^0 which is given by (2.1). Let $r(\pi_0, \delta)$ denote the Bayes risk of an estimator δ of θ under the prior π_0. Our aim is to show that $r(\pi_0, \delta^{SM}) - r(\pi_0, \delta^0) \to 0$ as $n \to \infty$.

Lemma 2.1 Assume $E(\lambda^{3/2}) < \infty$. Then $E[B(\lambda) \mid s, y(s)] \xrightarrow{p} 0$ as $n \to \infty$.

Proof Note that

$$E[B(\lambda) \mid s, y(s)] = \sigma^2 \int_0^\infty (\sigma^2 + n\lambda^{-1})^{-3/2} \exp[-\frac{n(\bar{y}(s)-\mu_0)^2}{2(\sigma^2+n\lambda^{-1})}] g(\tau_0^2\lambda) d\lambda$$

$$\div \int_0^\infty (\sigma^2 + n\lambda^{-1})^{-1/2} \exp[-\frac{n(\bar{y}(s)-\mu_0)^2}{2(\sigma^2+n\lambda^{-1})}] g(\tau_0^2\lambda) d\lambda$$

$$\le E\left[\left(\frac{1}{1+n(\sigma^2)^{-1}\lambda^{-1}}\right)^{1/2} \frac{1}{1+n(\sigma^2)^{-1}\lambda^{-1}}\right]$$

$$\div E\left[\left(\frac{1}{1+n(\sigma^2)^{-1}\lambda^{-1}}\right)^{1/2} \exp(-\frac{\lambda}{2}(\bar{y}(s)-\mu_0)^2)\right]$$

$$= E\left[\left(\frac{n(\sigma^2)^{-1}\lambda^{-1}}{1+n(\sigma^2)^{-1}\lambda^{-1}}\right)^{1/2} \lambda^{1/2} \frac{1}{1+n(\sigma^2)^{-1}\lambda^{-1}}\right]$$

$$\div E\left[\left(\frac{n(\sigma^2)^{-1}\lambda^{-1}}{1+n(\sigma^2)^{-1}\lambda^{-1}}\right)^{1/2} \lambda^{1/2} \exp(-\frac{\lambda}{2}(\bar{y}(s)-\mu_0)^2)\right]$$

$$= \frac{P_n}{Q_n} (say). \tag{2.10}$$

Now, $P_n \le n^{-1}\sigma^2 E(\lambda^{3/2}) \xrightarrow{p} 0$, if $E(\lambda^{3/2}) < \infty$. Also,

$$Q_n = E[(n^{-1}\sigma^2\lambda+1)^{-1/2}\lambda^{1/2}\exp(-\frac{\lambda}{2}(\bar{y}(s)-\mu_0)^2)]. \qquad (2.11)$$

Note that $\bar{y}(s)-\mu_0$ is the centered mean of an exchangeable sequence of random variables, and hence, is a centered backward martingale. Hence, $(\bar{y}(s)-\mu_0)^2$ is a backward submartingale. Since $\limsup_{n\to\infty} E(\bar{y}(s)-\mu_0)^2 = \tau_0^2 < +\infty$, by the submartingale convergence theorem, $(\bar{y}(s)-\mu_0)^2$ converges a.s. to a rv, say Y_0. Hence, using Fatou's lemma

$$\liminf_{n\to\infty} Q_n \ge E[\lambda^{1/2}\exp(-\frac{\lambda}{2}Y_0)], \qquad (2.12)$$

where the lower bound is bounded away from zero a.s.. Hence, $P_n/Q_n \xrightarrow{p} 0$ as $n\to\infty$.

We now turn to the theorem which proves the A.O. property of δ^{SM} obtained in (2.5).

Theorem 2.1 Assume $E(\lambda^{3/2}) < \infty$. Then $r(\pi_0, \delta^{SM}) - r(\pi_0, \delta^0) \to 0$ as $n\to\infty$.

Proof. Standard Bayesian calculations yield

$$\begin{aligned} r(\pi_0, \delta^{SM}) - r(\pi_0, \delta^0) &= E(\delta^{SM}-\delta^0)^2 \\ &= (1-f)^2 E[(E(B(\lambda)\mid s, y(s)) - B_0)^2(\bar{y}(s)-\mu_o)^2]. \end{aligned} \qquad (2.13)$$

By lemma 2.1, $E[B(\lambda)\mid s, y(s)] \xrightarrow{p} 0$ as $n\to\infty$. Also, $B_0 \to 0$ as $n\to\infty$. Hence, $(E(B(\lambda)\mid s, y(s)) - B_0)^2 \xrightarrow{p} 0$ as $n\to\infty$. Also, $\mid E(B(\lambda)\mid s, y(s)) - B_0 \mid \le 1$ and $(\bar{y}(s)-\mu_0)^2$ being a backward submartingale is uniformly integrable. Hence, the rhs of (2.13) $\to 0$ as $n\to\infty$. This completes the proof of the theorem.

3. An Example

We illustrate the results of Section 2 with an analysis of data in Cochran(1977). The data set consists of the 1920 and 1930 number of inhabitants, in thousands, of 64 large cities in the United States. The data were obtained by taking the cities which ranked fifth to sixty-eighth in the United States in total number of inhabitants in 1920. The cities are arranged in two strata, the first containing the 16 largest cities and the second the remaining 48 cities. But for our purpose, we just use the second stratum only. For the complete population, we find the population mean to be 197.875 and the population variance 5580.92. We use the 1920 data to elicit the prior in our setting so that $\mu_0 = 165.438$ and $\tau_0^2 = 71.424$. We want to estimate the average(or total) number of inhabitants in all 48 cities in 1930 based on a sample of size 16 (i.e. 1/3 sample). For illustrative purposes, we report our analysis for one sample.

To implement and monitor the convergence of the Gibbs sampling, we follow the basic approach of Gelman and Rubin(1992). We consider 10 independent sequences, each with a sample of size 5000 with a burn in sample of another 5000. We sample the θ initially from a t distribution with 2 degree of freedom. The justification for t distributions as well as the choice the specific parameters of this distribution are given below.

First note that from the posterior distribution of λ given s and $y(s)$ as given in (2.4), we find the posterior mode, say $\hat{\lambda}$ by using the Newton-Raphson algorithm. Also, we use $\bar{y}(s)$ for y_i, $i \in \bar{s}$ based on sample. We can now very well use the $N[(\hat{\lambda}\mu_0 + N\bar{y}(s)/\sigma^2)/(\hat{\lambda} + N/\sigma^2),\ (\hat{\lambda} + N/\sigma^2)^{-1}]$ as the starting posterior distribution for θ. But in order to start with an overdispersed distribution as recommended by Gelman and Rubin(1992), we take t distribution with 2 degree of freedom. Also, note that once the initial θ value have been generated, the rest of the procedure uses the posterior distributions as given in (i)-(iii) in Section 2.

Table 3.1 provides the Bayes estimates of $\gamma(\mathbf{y})$ and the associated posterior standard deviations for the normal, double exponential and t prior with degree of freedom 1, 3, 5, 10 and 15. Note that naive estimate, that is, the sample mean is 207.69.

An inspection of Table 3.1 reveals that there can be significant improvement in the estimate of $\gamma(\mathbf{y})$ by using heavy-tailed prior distributions rather than using the normal prior distribution in the sense of the closeness to $\gamma(\mathbf{y})$. For instance, using the double exponential and the t(1), t(3), t(5), t(10) and t(15) priors, the percentage improvements over the normal are given respectively by 45.78%, 89.05%, 52.06%, 30.68%, 15.53% and 9.06%. Here the percentage improvement of e_1 over e_2 is calculated by

$$((e_2 - \text{truth})^2 - (e_1 - \text{truth})^2)/(e_2 - \text{truth})^2,$$

where e_1 is the Bayes estimate based on heavy-tailed prior distributions and e_2 is the Bayes estimate using the normal prior. Also as one might expect, flatter the prior, closer is the Bayes estimates to the sample mean. In general, for most cases we have considered, the Cauchy prior (i.e., t prior with 1 degree of freedom) leads to an estimate which is closest to the population mean.

To monitor the convergence of the Gibbs sampler for θ, we follow Gelman and Rubin(1992). We find $\hat{R}$ value (the potential scale reduction factors) corresponding to θ. If $\hat{R}$ is near 1, it is reasonable to assume that the desired convergence is achieved in the Gibbs sampling algorithm (see Gelman and Rubin(1992) for the complete discussion).

The second column of Table 3.2 provides the $\hat{R}$ values corresponding to the estimand θ using Cauchy and double exponential priors based on $10 \times 5000 = 50000$ simulated values. The third column provides the corresponding 97.5% quantiles which are also equal to 1. The rightmost five columns of Table 3.2 show the simulated quantiles of the target posterior distribution of θ for each one of the 2 estimates based on 50000 simulated values.

Table 3.1. Bayes Estimates and Associated Posterior Standard Deviations

Priors	Bayes Estimates	Posterior SD
Normal	184.31	10.19
DE	187.89	12.16
t(1)	193.38	15.01
t(3)	188.48	12.63
t(5)	186.58	11.51
t(10)	185.41	10.75
t(15)	184.94	10.50

Table 3.2. Potential Scale Reduction and simulated Quantiles

Priors	Potential scale reduction		Simulated quantiles				
	$\hat{R}$	97.5 %	2.5 %	25.0%	50.0%	75.0%	97.5%
Chaucy	1.00	1.00	160.00	171.98	183.09	198.17	226.89
DE	1.00	1.00	158.68	168.54	176.00	185.55	206.82

References

[1] Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *The Journal of the Royal Statistical Society*, Ser. B, 36, 99-102.

[2] Box, G. E. P. and Tiao, G. C. (1968). A Bayesian approach to some outlier problems. *Biometrika*, 51, 153-167.

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

[4] Cochran, W. G. (1977). *Sampling Techniques*. (3rd edn.). Wiley, New York.

[5] Dawid, A. P. (1973). Posterior expectations for large observations. *Biometrika*, 60, 664-667.

[6] Ericson, W. A. (1969). Subjective Bayesian models in sampling finite populations (with discussion), *The Journal of the Royal Statistical Society*, Ser. B, 31, 195-233.

[7] Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85, 972-985.

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

[9] Gelfand, A. E. and Smith, A. F. M. (1991). Gibbs sampling for marginal posterior expectations. *Communications in Statistics-Theory and Methods*, 20(5 & 6), 1747-1766.

[10] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7, 457-511.

[11] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

[12] O'Hagan, A. (1979). On outlier rejection phenomena in Bayes inference. *The Journal of the Royal Statistical Society*, Ser, B, 41, 358-367.

[13] O'Hagan, A. (1989). Modeling with heavy tails. *Bayesian Statistics* 3, Eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindely and A. F. M. Smith. Oxford University Press, Oxford, 349-359.

[14] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *The Annals of Mathematical Statistics*, 35, 1-20.

[15] Sen, P. K. (1970). The Hajek-Renyi inequality for sampling from a finite population. *Sankhya*, Ser. A, 32, 181-188.

[16] West, M. (1985). Generalized linear models: Scale parameters, outliers accommodation and prior distributions. Bayesian Statistics 2, Eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith. North-Holland, Amsterdam, 531-557.

[17] West, M. (1987). On scale mixtures of normal distributions, *Biometrika*, 74, 646-648.