

Article

A Two-Stage Maximum Entropy Prior of Location Parameter with a Stochastic Multivariate Interval Constraint and Its Properties

Hea-Jung Kim

Department of Statistics, Dongguk University-Seoul, Seoul 100-715, Korea; kim3hj@dongguk.edu;
Tel.: +82-2-2260-3221

Academic Editor: Raúl Alcaraz Martínez

Received: 25 January 2016; Accepted: 10 May 2016; Published: 20 May 2016

Abstract: This paper proposes a two-stage maximum entropy prior to elicit uncertainty regarding a multivariate interval constraint of the location parameter of a scale mixture of normal model. Using Shannon's entropy, this study demonstrates how the prior, obtained by using two stages of a prior hierarchy, appropriately accounts for the information regarding the stochastic constraint and suggests an objective measure of the degree of belief in the stochastic constraint. The study also verifies that the proposed prior plays the role of bridging the gap between the canonical maximum entropy prior of the parameter with no interval constraint and that with a certain multivariate interval constraint. It is shown that the two-stage maximum entropy prior belongs to the family of rectangle screened normal distributions that is conjugate for samples from a normal distribution. Some properties of the prior density, useful for developing a Bayesian inference of the parameter with the stochastic constraint, are provided. We also propose a *hierarchical constrained scale mixture of normal model* (HCSMN), which uses the prior density to estimate the constrained location parameter of a scale mixture of normal model and demonstrates the scope of its applicability.

Keywords: hierarchical constrained scale mixture of normal model; rectangle-screened normal distribution; two-stage maximum entropy prior; uncertain constraint

MSC: 62H30; 62F15

1. Introduction

Suppose y_i 's are independent observations from a scale mixture of a p -variate normal distribution with the $p \times 1$ location parameter θ and known scale matrix. Then, a simple location model for the p -variate observations with $y_i \in \mathbb{R}^p$ is:

$$y_i = \theta + \epsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where the distribution of the $p \times 1$ vector variable ϵ_i is $F \in \mathcal{F}$ with:

$$\mathcal{F} = \left\{ F : N_p(\mathbf{0}, \kappa(\eta)\Lambda), \eta \sim G(\eta) \text{ with } \kappa(\eta) > 0, \text{ and } \eta > 0 \right\}, \quad (2)$$

where η is a mixing variable with the cdf $G(\eta)$ and $\kappa(\eta)$ is a suitably-chosen weight function.

Bayesian analysis of the model (1) begins with the specification of a prior distribution, which represents the information about the uncertain parameter θ that is combined with the joint probability distribution of y_i 's to yield the posterior distribution. When there are no constraints on the location parameter, then usual priors (e.g., Jeffreys invariant prior or an informative normal conjugate prior) can be used, and posterior inference can be performed without any difficulty. In some practical situations,

however, we may have prior information that θ have a multivariate interval constraint, and thus, the value of θ needs to be located in a restricted space $\mathbf{C} \subset \mathbb{R}^p$, where $\mathbf{C} = (\mathbf{a}, \mathbf{b})$ is a p -variate interval with $\mathbf{a} = (a_1, \dots, a_p)^\top$ and $\mathbf{b} = (b_1, \dots, b_p)^\top$. For the remainder of this paper, we use $\theta \in \mathbf{C}$ to denote the multivariate interval constraint:

$$\{\theta; a_i \leq \theta_i \leq b_i, i = 1, \dots, p\}, \text{ where } \theta = (\theta_1, \dots, \theta_p)^\top. \quad (3)$$

When we have sufficient evidence that the constraint condition on the model (1) is true, then a suitable restriction on the parameter space, such as using a truncated prior distribution, is expected. See, e.g., [1–4], for various applications of the truncated prior distribution in Bayesian inference. However, it is often the case that prior information about the constraint is not certain for Bayesian inference. Further, even the observations from the assumed model (1) often do not provide strong evidence that the constraint is true and, therefore, may appear to contradict the assumption of the model associated with the constraint. In this case, it is expected that the uncertainty about the constraint is taken into account in eliciting a prior distribution of θ . When the parameter constraint is not certain for Bayesian estimation in the univariate normal location model, the seminal work by [5] proposed the use of a two-stage hierarchical prior distribution by constructing a family of skew densities based on the positively-truncated normal prior distribution. Generalizing the framework of the prior hierarchy proposed by [5–10], among others, various priors were considered for the Bayesian estimation of normal and scale mixture of normal models with uncertain interval constraints. In particular, [7] obtained the prior of θ as the normal selection distribution (see, e.g., [11]) and, thus, exploited the class of weighted normal distribution by [12] for reflecting the uncertain prior belief on θ . On the other hand, there are situations to set up a prior density of θ on the basis of information regarding the moments of the density, such as the mean and covariance matrix. A useful method of dealing with this situation is through the concept of entropy by [13,14]. Other general references where moment inequality constraints have been considered include [15,16]. To the best of our knowledge, however, a formal method to set up a prior density of θ , consistent with information regarding the moments of the density, as well as the uncertain prior belief on the location parameter, has not previously been investigated in the literature. Thus, such practical considerations motivate us to develop a prior density of θ , which is tackled in this paper.

As discussed by [17–20], the entropy has a direct relationship to information theory and measures the amount of uncertainty inherent in the probability distribution. Using this property of the entropy, we propose a two-stage hierarchical method for setting up the two-stage maximum entropy prior density of θ . The method will enable us to elicit information regarding the moments of the prior distribution, as well as the degree of belief in the constraint $\theta \in \mathbf{C}$. Furthermore, this paper also suggests an objective method to measure the degree of belief regarding the multivariate interval constraint accounted for by using the prior. We also propose a simple way of controlling the degree of belief regarding the constraint of θ in Bayesian inference. This is done by investigating the relation between the degree of belief and the enrichment of the hyper-parameters of the prior density. In this respect, the study concerning the two-stage maximum entropy prior is interesting both from a theoretical and an applied point of view. On the theoretical side, it develops yet another conjugate prior of constrained θ based on the maximum entropy approach. The study provides several properties of the proposed prior, which advocate the idea of two stages of a prior hierarchy to elicit information regarding the moments of the prior and the stochastic constraint of θ . From the applied view point, the prior is especially useful for a Bayesian subjective methodology for inequality constrained multivariate linear models.

The remainder of this paper is arranged as follows. In Section 2, we propose the two-stage maximum entropy prior of θ by applying Boltzmann's maximum entropy theorem (see, e.g., [21,22]) to the frame of the two-stage prior hierarchy by [5]. We also suggest an objective measure of uncertainty regarding the stochastic constraint of θ that is accounted for by the two-stage maximum entropy prior. In Section 3, we briefly discuss the properties of the proposed prior of θ , which will be useful for

the Bayesian analysis of θ subject to uncertainty regarding the multivariate interval constraint $\theta \in \mathbf{C}$. Section 4 provides a hierarchical scale mixture of normal model of Equation (1) using the two-stage prior, referred to as the *hierarchical constrained scale mixture of normal model* (HCSMN). Section 4 explores the Bayesian estimation model (1) by deriving the posterior distributions of the unknown parameters under the HCSMN and discusses the properties of the proposed measure of uncertainty that can be explained in the context of the HCSMN. In Section 5, we compare the empirical performance of the proposed prior based on synthetic data and real data applications with the HCSMN models for the estimation of θ with a stochastic multivariate interval constraint. Finally, the concluding remarks along with a discussion are provided in Section 6.

2. Two-Stage Maximum Entropy Prior

2.1. Maximum Entropy Prior

Sometimes, we have a situation in which partial prior information is available, outside of which it is desirable to use a prior that is as non-informative as possible. Assume that we can specify the partial information concerning θ in Equation (1) with continuous space Θ . That is:

$$E[t_j(\theta)] = \int_{\Theta} t_j(\theta) \pi(\theta) d\theta = t_j, \quad j = 1, \dots, k. \quad (4)$$

The maximum entropy prior can be obtained by choosing $\pi(\theta)$ that maximizes the entropy:

$$\zeta(\pi) = - \int_{\Theta} \pi(\theta) \log \pi(\theta) d\theta,$$

in the presence of the partial information in the form of Equation (4). A straightforward application of the calculus of variation leads us to the following theorem.

Lemma 1. (Boltzmann's maximum entropy theorem): *The density $\pi(\theta)$ that maximizes $\zeta(\pi)$, subject to the constraints $E[t_j(\theta)] = t_j, j = 1, \dots, k$, takes the k -parameter exponential family form:*

$$\pi_{max}(\theta) \propto \exp \{ \lambda_1 t_1(\theta) + \lambda_2 t_2(\theta) \cdots + \lambda_k t_k(\theta) \}, \quad \theta \in \Theta, \quad (5)$$

where $\lambda_1, \lambda_2, \dots, \lambda_k$ can be determined, via the k -constraints, in terms of t_1, \dots, t_k .

Proof. See, [22] for the proof. \square

When the partial information is about the mean and covariance matrix of θ , outside of which it is desired to use a prior that is as non-informative as possible, then the theorem yields the following result.

Corollary 1. *As partial prior information, let the parameter $\theta = (\theta_1, \dots, \theta_p)^\top$ have a probability distribution on \mathbb{R}^p with mean vector $\theta_0 = (\theta_{01}, \dots, \theta_{0p})^\top$ and covariance matrix Σ , then the maximum entropy prior of θ is:*

$$\pi_{max}(\theta) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\theta - \theta_0)^\top \Sigma^{-1} (\theta - \theta_0) \right\}, \quad \theta \in \mathbb{R}^p, \quad (6)$$

a density of the $N_p(\theta_0, \Sigma)$ distribution.

Proof. According to Lemma 1, the partial information gives $t_j(\theta) = \theta_j$ and $t_j = \theta_{0j}$ for $j = 1, \dots, p$, $t_{p+1}(\theta) = \text{tr}[\Sigma^{-1}(\theta - \theta_0)(\theta - \theta_0)^\top]$ and $t_{p+1} = p$. $\int_{\mathbb{R}^p} \pi_{max}(\theta) d\theta = 1$ requires $\lambda_1 = \dots = \lambda_p = 0$ and $\lambda_{p+1} < 0$. Thus, the density $\pi_{max}(\theta)$ is proportional to $\exp \{ -\lambda_{p+1} \text{tr}[\Sigma^{-1}(\theta - \theta_0)(\theta - \theta_0)^\top] \}$. Setting $\lambda_{p+1} = -t_{p+1}/2p$ and obtaining the normalizing constant, then we see that the maximum entropy prior of the parameter in Equation (1) is Equation (6). \square

In practical situations, we sometimes have partial information about a multivariate interval constraint (i.e., $\theta \in \mathbf{C}$) in addition to the first two moments as given in Corollary 1.

Corollary 2. Assume that the prior distribution of $\theta = (\theta_1, \dots, \theta_p)^\top$ has the mean vector $\theta_0 = (\theta_{01}, \dots, \theta_{0p})^\top$ and covariance matrix Σ . Further assume, a priori, that the space of θ is constrained to a multivariate interval, $\{\theta; \theta \in \mathbf{C}\}$ given in Equation (3). Then, a constrained maximum entropy prior of θ is given by:

$$\pi_{const}(\theta) = \frac{(2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\theta - \theta_0)^\top \Sigma^{-1} (\theta - \theta_0) \right\}}{\Pr(\theta \in \mathbf{C})}, \theta \in \mathbf{C}, \tag{7}$$

a density of the $N_p(\theta_0, \Sigma)I(\theta \in \mathbf{C})$ distribution, which is a p -dimensional truncated $N_p(\theta_0, \Sigma)$ distribution with the space \mathbf{C} .

Proof. The certain multivariate interval constraint, $\Pr(\theta \in \mathbf{C}) = 1$, can be expressed in terms of moment, $E[I(\theta \in \mathbf{C})] = 1$. Upon applying Lemma 1 with $t_j(\theta) = \theta_j$ and $t_j = \theta_{0j}$ for $j = 1, \dots, p$, $t_{p+1}(\theta) = \text{tr}[\Sigma^{-1}(\theta - \theta_0)(\theta - \theta_0)^\top]$ and $t_{p+1} = p$, $t_{p+2}(\theta) = I(\theta \in \mathbf{C})$ and $t_{p+2} = 1$, and $\int_{\mathbf{C}} \pi_{const}(\theta) d\theta = 1$, we see that $\lambda_1 = \dots = \lambda_p = 0$, $\lambda_{p+1} < 0$, $\lambda_{p+2} = 1$, and $\pi_{const}(\theta) \propto \exp \left\{ -\lambda_{p+1} \text{tr}[\Sigma^{-1}(\theta - \theta_0)(\theta - \theta_0)^\top] \right\}$. Setting $\lambda_{p+1} = -1/2$, and obtaining the normalizing constant, we obtain Equation (7). \square

2.2. Two-Stage Maximum Entropy Prior

This subsection considers the case where the maximum entropy prior of θ has stochastic constraint in the form of a multivariate interval, i.e., $\Pr(\theta \in \mathbf{C}) = \gamma$, where \mathbf{C} is defined by Equation (3) and $\gamma \in [\gamma_{max}, 1]$. Here, γ_{max} is $\Pr(\theta \in \mathbf{C})$ calculated by using the maximum entropy prior distribution in Equation (6). We develop a two-stage prior of θ , denoted by $\pi_{two}(\theta)$, which has a different formula according to the degree of belief, γ , regarding the constraint.

Suppose we have only partial information about the covariance matrix, Ω_2 , of the parameter θ , in the first stage of a prior elicitation. Then, for a given mean vector μ_0 , we may construct the maximum entropy prior, Equation (6), so that the first stage maximum entropy prior will be $\pi_{max}(\theta|\mu_0)$, which is the density of the $N_p(\mu_0, \Omega_2)$ distribution. In addition to the information, suppose we have collected prior information about the unknown μ_0 , which gives a value of the mean vector θ_0 and covariance matrix Ω_1 , as well as a stochastic (or certain) constraint, indicating $\Pr(\mu_0 \in \mathbf{C}) = 1$. Then, in the second stage of the prior elicitation, one can elicit the additional prior partial information by using the constrained maximum entropy prior in Equation (7).

Analogous to the work of [5], we can specify all of the partial information about θ by following two stages of the maximum entropy prior hierarchy over $\theta \in \mathbb{R}^p$:

$$\pi_{max}(\theta|\mu_0) = \phi_p(\mu_0, \Omega_2), \tag{8}$$

$$\pi_{const}(\mu_0) = \phi_p(\theta_0, \Omega_1)I(\mu_0 \in \mathbf{C}), \tag{9}$$

where $\phi_p(\theta_0, \Omega_1)I(\mu_0 \in \mathbf{C})$ is a truncated normal density, i.e., the density of the $N_p(\theta_0, \Omega_1)I(\mu_0 \in \mathbf{C})$ variate, and $\Omega_1 + \Omega_2 = \Sigma$. Thus, the two stages of prior hierarchy are as follows. In the first stage, given μ_0 , θ has a maximum entropy prior that is the $N_p(\mu_0, \Omega_2)$ distribution as in Equation (6). In the second stage, μ_0 has a distribution obtained by truncating the maximum entropy prior distribution to elicit uncertainty about the prior information that $\theta \in \mathbf{C}$. It may be sensible to assume that the value of θ_0 is located in the multivariate interval \mathbf{C} or in the centroid of the interval.

Definition 1. The marginal prior density of θ , obtained from the two stages of the maximum entropy prior hierarchy Equations (8) and (9), is called as a two-stage maximum entropy prior of θ .

Since $\Omega_1 + \Omega_2 = \Sigma$, if the constraint is completely certain (i.e., $\gamma = 1$), we may set $\Omega_2 \rightarrow \mathbf{O}$ to get the $\pi_{const}(\theta)$ from the two stages of maximum entropy prior, while the two-stage prior yields $\pi_{max}(\theta)$ with $\gamma = \gamma_{max}$ for the case where $\Omega_1 = \mathbf{O}$. Thus, the hyper-parameters Ω_1 and Ω_2 may need to be assessed to achieve the degree of belief γ about the stochastic constraint. When $\Omega_1 \neq \mathbf{O}$ and $\Omega_2 \neq \mathbf{O}$, the above hierarchy of priors yields the following marginal prior of θ .

Lemma 2. *The two stages of the prior hierarchy of Equations (8) and (9) yield the two-stage maximum entropy prior distribution of θ given by:*

$$\pi_{two}(\theta) = \frac{\phi_p(\theta; \theta_0, \Sigma) \bar{\Phi}_p(\mathbf{C}; \mu, Q)}{\bar{\Phi}_p(\mathbf{C}; \theta_0, \Omega_1)}, \theta \in \mathbb{R}^p \tag{10}$$

where $\phi_p(\mathbf{x}; \mathbf{c}, \mathbf{A})$ denotes the pdf of $\mathbf{X} \sim N_p(\mathbf{c}, \mathbf{A})$ and $\bar{\Phi}_p(\mathbf{C}; \mathbf{c}, \mathbf{A})$ denotes a p -dimensional rectangle probability of the distribution of \mathbf{X} , i.e., $P(\mathbf{X} \in \mathbf{C})$, $\mu = \theta_0 + \Omega_1 \Sigma^{-1}(\theta - \theta_0)$, $\Sigma = \Omega_1 + \Omega_2$ and $Q = (\Omega_1^{-1} + \Omega_2^{-1})^{-1}$.

Proof.

$$\begin{aligned} \pi_{two}(\theta) &= \frac{\int_{\mu_0 \in \mathbf{C}} \phi_p(\theta; \mu_0, \Omega_2) \phi_p(\mu_0; \theta_0, \Omega_1) d\mu_0}{Pr(\mu_0 \in \mathbf{C})}, \\ &= \frac{\phi_p(\theta; \theta_0, \Sigma) \int_{\mu_0 \in \mathbf{C}} \phi_p(\mu_0; \mu, Q) d\mu_0}{\bar{\Phi}_p(\mathbf{C}; \theta_0, \Omega_1)}, \end{aligned}$$

because $\mu = \theta_0 + \Omega_1 \Sigma^{-1}(\theta - \theta_0) = \theta + \Omega_2 \Sigma^{-1}(\theta_0 - \theta)$ and $\theta^\top \Omega_2^{-1} \theta + \theta_0^\top \Omega_1^{-1} \theta_0 - \mu^\top (\Omega_2^{-1} \theta + \Omega_1^{-1} \theta_0) = (\theta - \theta_0)^\top \Sigma^{-1}(\theta - \theta_0)$. \square

In fact, the density $\pi_{two}(\theta)$ belongs to the family of rectangle screened multivariate normal (RSN) distributions studied by [23].

Corollary 3. *The distribution law of θ with the density in Equation (10) is:*

$$\theta \stackrel{d}{=} [\mathbf{X}_2 | \mathbf{X}_1 \in \mathbf{C}] \sim RSN_p(\mathbf{C}; \tau, \Psi), \tag{11}$$

which is a p -dimensional RSN distribution with respective location and scale parameters τ and Ψ and the rectangle screening space \mathbf{C} . Here, the joint distribution of \mathbf{X}_1 and \mathbf{X}_2 is $N_{2p}(\tau, \Psi)$, where $\tau = (\theta_0^\top, \theta_0^\top)^\top$ and

$$\Psi = \begin{pmatrix} \Omega_1 & \Omega_1 \\ \Omega_1 & \Sigma \end{pmatrix}.$$

Proof. The density of $[\mathbf{X}_2 | \mathbf{X}_1 \in \mathbf{C}]$ is:

$$\begin{aligned} \pi_{two}(\mathbf{x}_2) &= \frac{\phi_p(\mathbf{x}_2; \theta_0, \Sigma) \int_{\mathbf{C}} \phi_p(\mathbf{x}_1; \mu_{\mathbf{x}_1 | \mathbf{x}_2}, \Omega_{\mathbf{x}_1 | \mathbf{x}_2}) d\mathbf{x}_1}{\int_{\mathbf{C}} \phi_p(\mathbf{x}_1; \theta_0, \Omega_1) d\mathbf{x}_1} \\ &= \frac{\phi_p(\mathbf{x}_2; \theta_0, \Sigma) \bar{\Phi}_p(\mathbf{C}; \mu_{\mathbf{x}_1 | \mathbf{x}_2}, \Omega_{\mathbf{x}_1 | \mathbf{x}_2})}{\bar{\Phi}_p(\mathbf{C}; \theta_0, \Omega_1)} \end{aligned}$$

where $\mu_{\mathbf{x}_1 | \mathbf{x}_2} = \theta_0 + \Omega_1 \Sigma^{-1}(\mathbf{x}_2 - \theta_0)$ and $\Omega_{\mathbf{x}_1 | \mathbf{x}_2} = \Omega_1 - \Omega_1 \Sigma^{-1} \Omega_1$. By use of the binomial inverse theorem (see, e.g., [24] p. 23), one can easily see that $\mu_{\mathbf{x}_1 | \mathbf{x}_2}$ and $\Omega_{\mathbf{x}_1 | \mathbf{x}_2}$ are respectively equivalent to μ and Q , in Equation (10), provided that \mathbf{x}_2 is changed to θ . \square

According to [23], we see that the stochastic representation for the RSN vector $\boldsymbol{\theta} \sim RSN_p(\mathbf{C}; \boldsymbol{\tau}, \Psi)$ is:

$$\boldsymbol{\theta} \stackrel{d}{=} \boldsymbol{\theta}_0 + \mathbf{Y}_1^{(\boldsymbol{\alpha}, \boldsymbol{\beta})} + \mathbf{Y}_2, \quad (12)$$

where $\mathbf{Y}_1 \sim N_p(\mathbf{0}, \Omega_1)$ and $\mathbf{Y}_2 \sim N_p(\mathbf{0}, \Omega_2)$ are independent random vectors. Here, $\mathbf{Y}_1^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}$ denotes a doubly-truncated multivariate normal random vector whose distribution is defined by $\mathbf{Y}_1^{(\boldsymbol{\alpha}, \boldsymbol{\beta})} \stackrel{d}{=} [\mathbf{Y}_1 | \mathbf{Y}_1 \in (\boldsymbol{\alpha}, \boldsymbol{\beta})]$ with $\boldsymbol{\alpha} = \mathbf{a} - \boldsymbol{\theta}_0$ and $\boldsymbol{\beta} = \mathbf{b} - \boldsymbol{\theta}_0$. This representation enables us to implement a one-for-one method for generating a random vector with the $RSN_p(\mathbf{C}, \boldsymbol{\tau}, \Psi)$ distribution. For generating the doubly-truncated multivariate normal vector $\mathbf{Y}_1^{(\boldsymbol{\alpha}, \boldsymbol{\beta})}$, the R package `tmvtnorm` by [25] can be used, where R is a computer language and an environment for statistical computing and graphics.

2.3. Entropy of a Maximum Entropy Prior

Suppose we have partial *a priori* information that we can specify values for the covariance matrices Ω_1 and Ω_2 , where $\Sigma = \Omega_1 + \Omega_2$.

2.3.1. Case 1: Two-stage Maximum Entropy Prior

When the two-stage maximum entropy prior $\pi_{two}(\boldsymbol{\theta})$ is assumed for the prior distribution of $\boldsymbol{\theta}$, its entropy is given by:

$$\begin{aligned} Ent(\pi_{two}(\boldsymbol{\theta})) &= - \int_{\mathbb{R}^p} \pi_{two}(\boldsymbol{\theta}) \log \pi_{two}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \frac{p}{2} \log(2\pi) + \log \bar{\Phi}_p(\mathbf{C}; \boldsymbol{\theta}_0, \Omega_1) + \frac{1}{2} \text{tr} \left[\Sigma^{-1} E_{two} [(\boldsymbol{\theta} - \boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top] \right] \\ &+ \frac{1}{2} \log |\Sigma| - E_{two} [\log h(\boldsymbol{\theta})], \end{aligned} \quad (13)$$

where $\Sigma = \Omega_1 + \Omega_2$, $h(\boldsymbol{\theta}) = \bar{\Phi}_p(\mathbf{C}; \boldsymbol{\theta}_0 + \Omega_1^\top \Sigma^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \Omega_1 - \Omega_1^\top \Sigma^{-1} \Omega_1)$, and the E_{two} denotes the expectation with respect to the RSN distribution with the density $\pi_{two}(\boldsymbol{\theta})$. Equation (12) shows that $E[\boldsymbol{\theta}] = \boldsymbol{\theta}_0 + \boldsymbol{\zeta}$, and $Cov(\boldsymbol{\theta}) = \Omega_2 + H$. Here, $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_p)^\top$ and $H = \{h_{ij}\}$, $i, j = 1, \dots, p$, are the mean vector and covariance matrix of the doubly-truncated multivariate normal random vector, $\mathbf{Y}_1 \sim N_p(\mathbf{0}, \Omega_1) I(\mathbf{y}_1 \in (\boldsymbol{\alpha}, \boldsymbol{\beta}))$. Readers are referred to [25] with the R package `tmvtnorm` and [26] with the R package `mvtnorm` for implementing the respective calculations of doubly-truncated moments and integrations. As seen in Equation (13), an analytic calculation of $E[\log h(\boldsymbol{\theta})]$ involves a complicated integration. Instead, by using a Monte Carlo integration, we may calculate it approximately. According to Equation (12), it follows that the stochastic representation of the prior distribution $\boldsymbol{\theta} \sim RSN_p(\mathbf{C}; \boldsymbol{\tau}, \Psi)$ with density $\pi_{two}(\boldsymbol{\theta})$ is useful for generating $\boldsymbol{\theta}'$ s from the prior distribution $\boldsymbol{\theta}$ by using the R packages `mvtnorm` and `tmvtnorm` and, hence, implementing the Monte Carlo integration.

2.3.2. Case 2: Constrained Maximum Entropy Prior

When the constrained maximum entropy prior $\pi_{const}(\boldsymbol{\theta})$ in Equation (7) is assumed for the prior distribution of $\boldsymbol{\theta}$, its entropy is given by:

$$\begin{aligned} Ent(\pi_{const}(\boldsymbol{\theta})) &= - \int_{\mathbf{C}} \pi_{const}(\boldsymbol{\theta}) \log \pi_{const}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \frac{p}{2} \log(2\pi) + \log \bar{\Phi}_p(\mathbf{C}; \boldsymbol{\theta}_0, \Sigma) + \frac{1}{2} \log |\Sigma| \\ &+ \frac{1}{2} \text{tr} \left(\Sigma^{-1} E_{const} [(\boldsymbol{\theta} - \boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top] \right). \end{aligned}$$

The E_{const} denotes the expectation with respect to the doubly-truncated multivariate normal distribution with the density $\pi_{const}(\boldsymbol{\theta})$, and its analytic calculation is not possible. Instead, the R

packages tmvtnorm and mvtnorm are available for calculating the respective moment and integration in the expression of $Ent(\pi_{const})$.

2.3.3. Case 3: Maximum Entropy Prior

On the other hand, if the maximum entropy prior $\pi_{max}(\theta)$ is assumed for the prior distribution of the location parameter θ , its entropy is given by:

$$\begin{aligned} Ent(\pi_{max}(\theta)) &= - \int_{\mathbb{R}^p} \pi_{max}(\theta) \log \pi_{max}(\theta) d\theta \\ &= \frac{p}{2} + \frac{p}{2} \log(2\pi) + \frac{1}{2} \log |\Sigma|. \end{aligned}$$

The following theorem asserts the relationship among the degrees of belief, accounted for by the three priors, about the *a priori* uncertain constraint $\{\theta; \theta \in \mathbf{C}\}$.

Theorem 1. *The degrees of belief γ_{max} , γ_{two} , and γ_{const} about the *a priori* constraint $\{\theta; \theta \in \mathbf{C}\}$, accounted for by $\pi_{max}(\theta)$, $\pi_{two}(\theta)$ and $\pi_{const}(\theta)$, have the following relation:*

$$\gamma_{max} \leq \gamma_{two} \leq \gamma_{const}, \tag{14}$$

provided that the parameters of $\pi_{two}(\theta)$ in Equation (10) satisfy:

$$\bar{\Phi}_{2p}(\mathbf{C}^*; \tau, \Psi) \geq \bar{\Phi}_p(\mathbf{C}; \theta_0, \Omega_1) \bar{\Phi}_p(\mathbf{C}; \theta_0, \Sigma),$$

where $\mathbf{C}^* = \{\mathbf{x}; \mathbf{x}_1 \in \mathbf{C}, \mathbf{x}_2 \in \mathbf{C}\}$ denotes the $2p$ -variate interval of random vector $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$, the equality $\gamma_{max} = \gamma_{two}$ holds for $\Omega_1 = \mathbf{O}$, $\gamma_{two} = \gamma_{const}$ holds for $\Omega_2 = \mathbf{O}$ and $\gamma_{max} = \gamma_{two} = \gamma_{const}$ holds for $\mathbf{C} = \mathbb{R}^p$.

Proof. The conditions for equalities are straightforward from the stochastic representation in Equation (12). Under the $\pi_{max}(\theta)$ in Equation (6),

$$\gamma_{max} = Pr(\theta \in \mathbf{C}) = \bar{\Phi}_p(\mathbf{C}; \theta_0, \Sigma), \quad \gamma_{two} = \int_{\theta \in \mathbf{C}} \pi_{two}(\theta) d\theta = \frac{\bar{\Phi}_{2p}(\mathbf{C}^*; \tau, \Psi)}{\bar{\Phi}_p(\mathbf{C}; \theta_0, \Omega_1)},$$

because $\pi_{two}(\theta)$ is the density of $\theta \sim RSN_p(\mathbf{C}; \tau, \Psi)$, and $\gamma_{const} = \int_{\theta \in \mathbf{C}} \pi_{const}(\theta) d\theta = 1$. Therefore, the condition $\bar{\Phi}_{2p}(\mathbf{C}^*; \tau, \Psi) \geq \bar{\Phi}_p(\mathbf{C}; \theta_0, \Omega_1) \bar{\Phi}_p(\mathbf{C}; \theta_0, \Sigma)$ gives the inequality relation. \square

3. Properties

3.1. Objective Measure of Uncertainty

In constructing the two stages of prior hierarchy over $\theta \in \mathbb{R}^p$, the usual practice is to set the value of θ_0 as the centroid of the uncertain constrained multivariate interval $\mathbf{C} = (\mathbf{a}, \mathbf{b})$. In this case, we have the following result.

Corollary 4. *In the case where the value of θ_0 in $\pi_{two}(\theta)$ is the centroid of the multivariate interval \mathbf{C} ,*

$$\gamma_{max} \leq \gamma_{two} \leq \gamma_{const}. \tag{15}$$

Proof. Equation (12) indicates that:

$$\gamma_{two} = Pr(\mathbf{Y}_1 + \mathbf{Y}_2 \in (\alpha, \beta) | \mathbf{Y}_1 \in (\alpha, \beta)) \text{ and } \gamma_{max} = Pr(\mathbf{Y}_1 + \mathbf{Y}_2 \in (\alpha, \beta)),$$

where $\mathbf{Y}_1 \sim N_p(\mathbf{0}, \Omega_1)$ and $\mathbf{Y}_2 \sim N_p(\mathbf{0}, \Omega_2)$ are independent random vectors, $\boldsymbol{\alpha} = \mathbf{a} - \boldsymbol{\theta}_0$, and $\boldsymbol{\beta} = \mathbf{b} - \boldsymbol{\theta}_0$. When $\boldsymbol{\theta}_0$ is the centroid of \mathbf{C} , $\boldsymbol{\alpha} = -\boldsymbol{\beta}$, and hence:

$$Pr(\mathbf{Y}_1 + \mathbf{Y}_2 \in (\boldsymbol{\alpha}, \boldsymbol{\beta}), \mathbf{Y}_1 \in (\boldsymbol{\alpha}, \boldsymbol{\beta})) \geq Pr(\mathbf{Y}_1 + \mathbf{Y}_2 \in (\boldsymbol{\alpha}, \boldsymbol{\beta}))Pr(\mathbf{Y}_1 \in (\boldsymbol{\alpha}, \boldsymbol{\beta}))$$

by the theorem of [27]. This leads to the first inequality, $\gamma_{max} \leq \gamma_{two}$. Since $\gamma_{const} = 1$, we see that the second inequality in Equation (15) holds. \square

The following are immediate from Theorem 1 and Corollary 4 : (i) The two-stage maximum entropy prior achieves γ_{two} for the degree of belief about the uncertain multivariate interval constraint $\{\boldsymbol{\theta}; \boldsymbol{\theta} \in \mathbf{C}\}$, and its value satisfies $\gamma_{two} \in [\gamma_{max}, 1]$ if the condition in the theorem is satisfied. Note that the equality $\gamma_{two} = 1$ holds for $\Omega_2 = \mathbf{O}$; (ii) The degree of belief about the multivariate interval constraint is a function of the covariance matrices Ω_1 and Ω_2 . Thus, if we have the partial *a priori* information that specifies values of the covariance matrices Ω_1 and Ω_2 , the degree of belief γ_{two} , associated with $\pi_{two}(\boldsymbol{\theta})$, can be assessed.

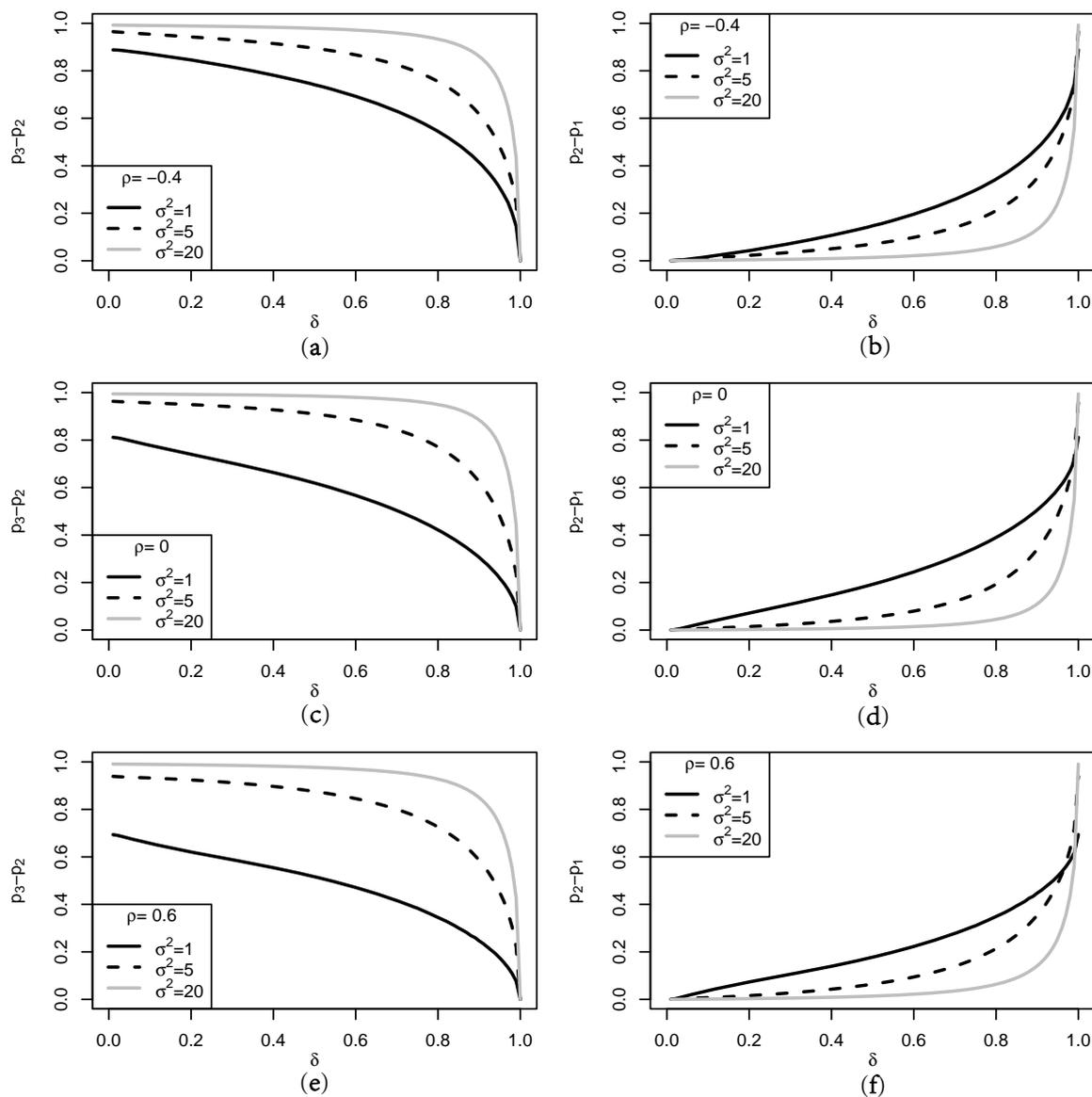


Figure 1. Graphs of the difference between $p_1 = \gamma_{max}$, $p_2 = \gamma_{two}$ and $p_3 = \gamma_{const}$. (a), (c), and (e) for the difference between p_3 and p_2 ; (b), (d), and (f) for the difference between p_2 and p_1 .

Figure 1 compares the degrees of belief about the uncertain multivariate interval constraint $\{\theta; \theta \in \mathbf{C}\}$, accounted for by the three priors of θ . The figure is obtained in terms of $\delta \in [0, 1]$ with $\Omega_1 = \delta\Sigma$ and $\Omega_2 = (1 - \delta)\Sigma$, $p = 3$, $\mathbf{C} = (\mathbf{a}, 2\mathbf{1}_p + \mathbf{a})$ and $\theta_0 = \mathbf{0}$, where $\mathbf{a} = (-0.1 \times p)\mathbf{1}_p$, $\Sigma = \sigma^2(1 - \rho)\mathbf{I}_p + \sigma^2\rho\mathbf{1}_p\mathbf{1}_p^\top$ is an intra-class covariance matrix, and $\mathbf{1}_p$ denotes a $p \times 1$ summing vector whose every element is unity. When the constraint is changed to $\mathbf{C} = (-(2\mathbf{1}_p + \mathbf{a}), -\mathbf{a})$ in this comparison, one can easily check that the degrees of belief do not change and give the same results seen in Figure 1. The figure depicts exactly the same inequality relationship given in Theorem 1. In comparison with γ_{two} and $\gamma_{const} = 1$, we see that the degree of belief in the uncertain constraint, accounted for by using $\pi_{two}(\theta)$, becomes large as $\Omega_2 \rightarrow \mathbf{O}$ (or equivalently $\Omega_1 \rightarrow \Sigma$). In particular, this tendency is more evident for small σ^2 and large ρ values. Third, the difference in γ_{two} and γ_{max} in the right panel suggests that the difference becomes large as Ω_2 tends to \mathbf{O} . In particular, for fixed values of δ and ρ , the figure shows that the difference increases as the value of σ^2 decreases, while it decreases as the value of ρ increases for fixed values of δ and σ^2 . Therefore, the figure confirms that the two-stage maximum entropy prior $\pi_{two}(\theta)$ accounts for the *a priori* uncertain constraint $\{\theta; \theta \in \mathbf{C}\}$ with the degree of belief $\gamma_{two} \in [\gamma_{max}, 1]$. The figure also notes that the magnitude of γ_{two} depends on both the first stage covariance Ω_2 and the second stage covariance Ω_1 in the two stages of prior hierarchy in Equations (8) and (9). All other choices of the values of p, ρ , and \mathbf{C} , satisfying the condition in Theorem 1, produced similar graphics depicted in Figure 1, with the exception of the magnitude of the differences among the degrees of belief.

3.2. Properties of the Entropy

The expected uncertainty in the multivariate interval constraint of the location parameter θ , $\{\theta; \theta \in \mathbf{C}\}$, accounted for by the two-stage prior $\pi_{two}(\theta)$, is measured by its entropy $Ent(\pi_{two}(\theta))$, and information about the constraint is defined by $-Ent(\pi_{two}(\theta))$. Thus, as considered by [20,28], the difference between the Shannon measures of information, before and after applying the uncertain constraint $\{\theta; \theta \in \mathbf{C}\}$, can be explained by the following property.

Corollary 5. *When θ_0 is the centroid of the multivariate interval \mathbf{C} ,*

$$Ent(\pi_{max}(\theta)) \geq Ent(\pi_{two}(\theta)) \geq Ent(\pi_{const}(\theta)), \tag{16}$$

where $Ent(\pi_{two}(\theta))$ reduces to $Ent(\pi_{max}(\theta))$ for $\Omega_1 = \mathbf{O}$, while $Ent(\pi_{two}(\theta))$ is equal to $Ent(\pi_{const}(\theta))$ for $\Omega_2 = \mathbf{O}$. All of the equalities hold for $\mathbf{C} = \mathbb{R}^p$.

Proof. It is straightforward to check the equalities by using the stochastic representation in Equation (12). Since $\pi_{max}(\theta)$ is the maximum entropy prior, it is sufficient to show that $Ent(\pi_{two}(\theta)) \geq Ent(\pi_{const}(\theta))$. First, $\Sigma - \Omega_1 = \Omega_2 > 0$ implies that $\Phi_p(\mathbf{C}; \theta_0, \Omega_1) \geq \Phi_p(\mathbf{C}; \theta_0, \Sigma)$ by the lemma of [27]. Second, $\gamma_{const} = Pr(\mathbf{Y}_1 + \mathbf{Y}_2 \in (\alpha, \beta) | \mathbf{Y}_1 + \mathbf{Y}_2 \in (\alpha, \beta)) \geq Pr(\mathbf{Y}_1 + \mathbf{Y}_2 \in (\alpha, \beta) | \mathbf{Y}_1 \in (\alpha, \beta)) = \gamma_{two}$ by Corollary 4. This and the lemma of [27] indicate that $Cov(\mathbf{Y}_1 + \mathbf{Y}_2 | \mathbf{Y}_1 \in (\alpha, \beta)) - Cov(\mathbf{Y}_1 + \mathbf{Y}_2 | \mathbf{Y}_1 + \mathbf{Y}_2 \in (\alpha, \beta))$ is a positive-semi-definite, and hence, $tr\left(\Sigma^{-1}E_{two}\left[(\theta - \theta_0)(\theta - \theta_0)^\top\right]\right) \geq tr\left(\Sigma^{-1}E_{const}\left[(\theta - \theta_0)(\theta - \theta_0)^\top\right]\right)$ by ([29], p. 54), where $E_{two}\left[(\theta - \theta_0)(\theta - \theta_0)^\top\right] = Cov(\mathbf{Y}_1 + \mathbf{Y}_2 | \mathbf{Y}_1 \in (\alpha, \beta))$ and $E_{const}\left[(\theta - \theta_0)(\theta - \theta_0)^\top\right] = Cov(\mathbf{Y}_1 + \mathbf{Y}_2 | \mathbf{Y}_1 + \mathbf{Y}_2 \in (\alpha, \beta))$ for $\alpha = -\beta$. These two results give the inequality $Ent(\pi_{two}(\theta)) \geq Ent(\pi_{const}(\theta))$, because $E_{two}[\log h(\theta)] \leq 0$. \square

Figure 2 depicts the difference between $Ent(\pi_{max}(\theta))$, $Ent(\pi_{two}(\theta))$ and $Ent(\pi_{const}(\theta))$ using the same parameter values used in constructing Figure 1. Figure 2 coincides with the inequality relation given in Corollary 5 and indicates the following consequences: (i) Even though θ_0 is not the centroid of the multivariate interval \mathbf{C} , we see that $Ent(\pi_{max}(\theta)) > Ent(\pi_{two}(\theta)) > Ent(\pi_{const}(\theta))$ for $\delta \in (0, 1)$. (ii) The difference $Ent(\pi_{two}(\theta)) - Ent(\pi_{const}(\theta))$ is a monotone decreasing function of δ , while $Ent(\pi_{max}(\theta)) - Ent(\pi_{two}(\theta))$ is a monotone increasing function. (iii) The differences get bigger

the larger σ^2 becomes for $\delta \in (0, 1)$. This indicates that the entropy of $\pi_{two}(\theta)$ is associated not only with the covariance of the first stage prior Ω_2 , but that of the second stage prior Ω_1 in Equations (8) and (9), respectively. (iv) Upon comparing Figures 1 and 2, the entropy $Ent(\pi_{two}(\theta))$ is closely related to the degree of belief γ_{two} , such that:

$$Ent(\pi_{two}(\theta)) = c_{two}(1 - \gamma_{two}),$$

where $c_{two} > 0$ is obtained by using Equations (13) and (16) and $1 - \gamma_{two}$ denotes the degree of uncertainty in a priori information regarding the multivariate interval constraint $\{\theta; \theta \in \mathbf{C}\}$ elicited by $\pi_{two}(\theta)$. These consequences and Corollary 5 indicate that $1 - \gamma_{two}$ stands between $1 - \gamma_{const}$ and $1 - \gamma_{max}$. Thus, the two-stage prior $\pi_{two}(\theta)$ is useful for eliciting uncertain information about the multivariate interval constraint. Theorem 1 and the above statements produce an objective method for eliciting the stochastic constraint $\{\theta; \theta \in \mathbf{C}\}$ via $\pi_{two}(\theta)$.

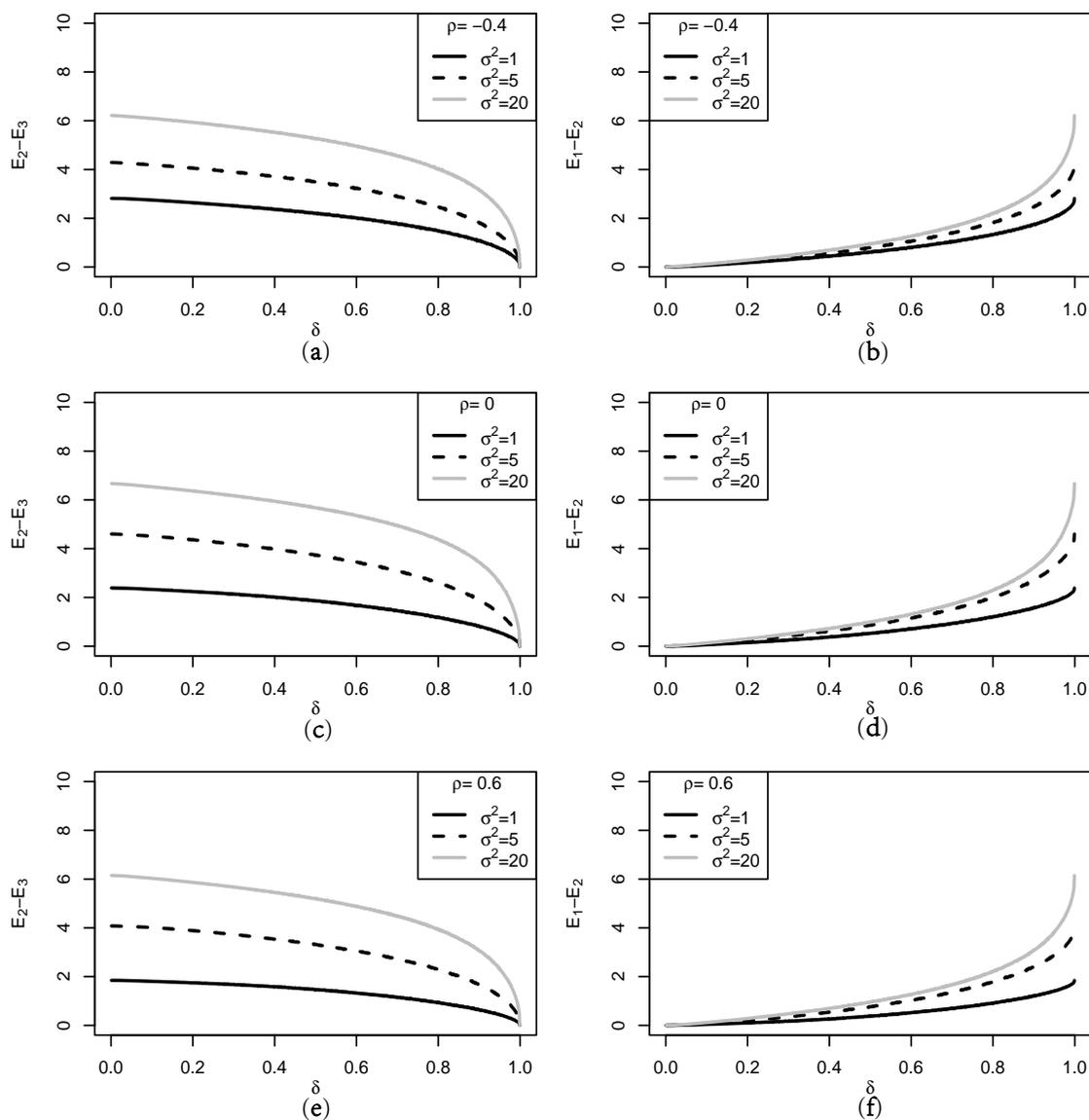


Figure 2. Graphs of the entropy difference between $E_1 = Ent(\pi_{max}(\theta))$, $E_2 = Ent(\pi_{two}(\theta))$ and $E_3 = Ent(\pi_{const}(\theta))$ for different values of $\delta \in [0, 1]$. (a), (c), and (e) for the difference between E_2 and E_3 ; (b), (d), and (f) for the difference between E_1 and E_2 .

Corollary 6. Suppose the degree $(1 - \gamma_{two})$ of uncertainty associated with the stochastic constraint $\{\theta; \theta \in \mathbf{C}\}$ is given. An objective way of eliciting the prior information by using $\pi_{two}(\theta)$ is to choose the covariance matrices Ω_1 and Ω_2 in $\pi_{two}(\theta)$, such that $\gamma_{two} = \bar{\Phi}_{2p}(\mathbf{C}^*; \tau, \Psi) / \bar{\Phi}_p(\mathbf{C}; \theta_0, \Omega_1)$, where $\Sigma = \Omega_1 + \Omega_2$ is known and $\Omega_1 = \delta\Sigma$ with $\delta \in [0, 1]$.

Since $\gamma_{const} = 1$, the degree of uncertainty $(1 - \gamma_{two})$ is equal to $\gamma_{const} - \gamma_{two}$. The left panel of Figure 1 plots a graph of $1 - \gamma_{two}$ against δ . The graph indicates that a δ value for $\pi_{two}(\theta)$ can be easily determined for given Σ , and the value is in inverse proportion to the degree of uncertainty regardless of Σ .

3.3. Posterior Distribution

Suppose the distribution of the error vector in the model (1) belongs to the family of scale mixture of normal distributions defined in Equation (2); then, the conditional distribution of the data information from $n = 1$ is $[\mathbf{y}|\eta] \sim N_p(\theta, \kappa(\eta)\Lambda)$. It is well known that the priors $\pi_{max}(\theta)$ and $\pi_{const}(\theta)$ are conjugate priors for the location vector θ , provided that η and Λ are known. That is, conditional on η , each prior satisfies the conjugate property that the prior and the posterior distributions of θ belong to the same family of distributions. The following corollary provides that the conditional conjugate property also applies to $\pi_{two}(\theta)$.

Corollary 7. Let $[\mathbf{y}|\eta] \sim N_p(\theta, \kappa(\eta)\Lambda)$ with known Λ . Then, the two-stage maximum entropy prior $\pi_{two}(\theta)$ in Equation (10) yields the conditional posterior distribution of θ given by:

$$[\theta|\mathbf{y}, \eta] \sim RSN_p(\mathbf{C}; \tau_\eta^*, \Psi_\eta^*), \tag{17}$$

where $\Omega_1 = \delta\Sigma$, $\delta \in (0, 1)$, $\Sigma = \Omega_1 + \Omega_2$, $\Sigma_{1\eta}^* = \delta(1 - \delta)\Sigma + \delta^2\Sigma_\eta^*$, $\Sigma_\eta^* = (\kappa(\eta)^{-1}\Lambda^{-1} + \Sigma^{-1})^{-1}$,

$$\tau_\eta^* = \begin{pmatrix} \theta_{0\eta}^* \\ \theta_\eta^* \end{pmatrix}, \Psi_\eta^* = \begin{pmatrix} \Sigma_{1\eta}^* & \delta\Sigma_\eta^* \\ \delta\Sigma_\eta^* & \Sigma_\eta^* \end{pmatrix}, \theta_{0\eta}^* = (1 - \delta)\theta_0 + \delta\theta_\eta^*, \text{ and } \theta_\eta^* = \Sigma_\eta^*(\kappa(\eta)^{-1}\Lambda^{-1}\mathbf{y} + \Sigma^{-1}\theta_0).$$

Proof. When the two-stage prior $\pi_{two}(\theta)$ in Equation (10) is used, the conditional posterior density of θ given η is:

$$\begin{aligned} p(\theta|\mathbf{y}, \eta) &\propto \phi_p(\mathbf{y}; \theta, \kappa(\eta)\Lambda)\phi_p(\theta; \theta_0, \Sigma)\bar{\Phi}_p(\mathbf{C}; \mu, Q) / \bar{\Phi}_p(\mathbf{C}; \theta_0, \Omega_1), \\ &\propto \phi_p(\theta; \theta_\eta^*, \Sigma_\eta^*)\bar{\Phi}_p(\mathbf{C}; \mu_\eta^*, Q_\eta^*), \end{aligned}$$

in that $\bar{\Phi}_p(\mathbf{C}; \mu, Q) = \bar{\Phi}_p(\mathbf{C}; \mu_\eta^*, Q_\eta^*)$, where $\mu = \theta_0 + \Omega_1\Sigma^{-1}(\theta - \theta_0)$, $\mu_\eta^* = \theta_{0\eta}^* + \delta(\theta - \theta_\eta^*)$ and $Q_\eta^* = \Sigma_{1\eta}^* - \delta^2\Sigma_\eta^*$. The last term of the proportional relations is a kernel of the $RSN_p(\mathbf{C}; \tau_\eta^*, \Psi_\eta^*)$ density defined by Corollary 3. \square

Corollaries 3 and 7 establish the conditional conjugate property of $\pi_{two}(\theta)$: Suppose the location parameter θ is the normal mean vector, then the RSN prior distribution, i.e., $\pi_{two}(\theta)$, yields the conditional posterior distribution, which belongs to the class of RSN distributions as given in Corollary 7. In the particular case where the distribution of η degenerates at $\kappa(\eta) = 1$, i.e., the model (1) is a normal model, then the conditional conjugate property of $\pi_{two}(\theta)$ reduces to the unconditional conjugate property.

Using the relation between the distribution of Equation (11) and that of Equation (12), we can obtain the stochastic representation for the conditional posterior RSN distribution in Equation (17) as follows.

Corollary 8. Conditional on the mixing variable η , the stochastic representation of $[\theta|\mathbf{y}, \eta] \sim RSN_p(\mathbf{C}; \tau_\eta^*, \Psi_\eta^*)$ is:

$$[\theta|\mathbf{y}, \eta] \stackrel{d}{=} \theta_\eta^* + \delta\Sigma_\eta^*\Sigma_{1\eta}^{*-1}\mathbf{W}_1^{(\alpha_\eta^*, \beta_\eta^*)} + (\Sigma_\eta^* - \delta^2\Sigma_\eta^*\Sigma_{1\eta}^{*-1}\Sigma_\eta^*)^{1/2}\mathbf{W}_2, \tag{18}$$

where $\mathbf{W}_1 \sim N_p(\mathbf{0}, \Sigma_{1\eta}^*)$ and $\mathbf{W}_2 \sim N_p(\mathbf{0}, I_p)$ are independent and $\mathbf{W}_1^{(\alpha_\eta^*, \beta_\eta^*)} \stackrel{d}{=} [\mathbf{W}_1 | \mathbf{W}_1 \in (\alpha_\eta^*, \beta_\eta^*)]$, where $\alpha^* = \mathbf{a} - \theta_{0\eta}^*$ and $\beta^* = \mathbf{b} - \theta_{0\eta}^*$.

Proof. Suppose the distributions of \mathbf{X}_1 and \mathbf{X}_2 in Equation (11) changed to $\mathbf{X}_1 \stackrel{d}{=} \theta_{0\eta}^* + \mathbf{W}_1$ and $\mathbf{X}_2 \stackrel{d}{=} \theta_\eta^* + \delta \Sigma_\eta^* \Sigma_{1\eta}^{*-1} \mathbf{W}_1 + (\Sigma_\eta^* - \delta^2 \Sigma_\eta^* \Sigma_{1\eta}^{*-1} \Sigma_\eta^*)^{1/2} \mathbf{W}_2$. Then, the stochastic representation in Equation (12) associated with the distribution $[\mathbf{X}_2 | \mathbf{X}_1 \in \mathbf{C}]$ in Equation (11) gives the result. \square

4. Hierarchical Constrained Scale Mixture of Normal Model

For the model (1), if we are completely sure about a multivariate interval constraint on θ , a suitable restriction on the parameter space $\theta \in \mathbb{R}^p$, such as using a truncated normal prior distribution, is expected for eliciting the information. However, there are certain cases where we have *a priori* information that the location parameter θ is highly likely to have a multivariate interval constraint, and thus, the value of θ needs to be located with uncertainty in a restricted space $\{\theta \in \mathbf{C}\}$ with $\mathbf{C} = (\mathbf{a}, \mathbf{b})$. Then, we cannot be sure about the constraint, and then, the constraint becomes stochastic (or uncertain), as in our problem of interest. In this case, the uncertainty about the constraint must be taken into account in the estimation procedure of the model (1). This section considers a hierarchical Bayesian estimation of the scale mixture of normal models reflecting the uncertain prior belief on θ .

4.1. The Hierarchical Model

Let us consider a hierarchical constrained scale mixture of normal model (HCSMN) that uses the hierarchy of the scale mixture of normal model (1) and includes the two stages of a prior hierarchy in the following way:

$$\begin{aligned} [\mathbf{y}_i | \eta_i, \Lambda] &= \theta + \epsilon_i, \epsilon_i \sim N_p(0, \kappa(\eta_i)\Lambda), \quad i = 1, \dots, n, \\ \theta | \mu_0 &\sim N_p(\mu_0, \Omega_2), \quad \text{independent of } (\epsilon_1, \dots, \epsilon_n)^\top, \\ \mu_0 &\sim N_p(\theta_0, \Omega_1) I(\mu_0 \in \mathbf{C}), \\ \Lambda &\sim W_p^{-1}(D, d), \quad d > 2p, \\ \eta_i &\stackrel{iid}{\sim} g(\eta), \quad i = 1, \dots, n, \end{aligned} \tag{19}$$

where $W_p^{-1}(D, d)$ denotes the inverted Wishart distribution with positive definite scale matrix D and d degrees of freedom whose pdf $W_p^{-1}(\Lambda; D, d)$ is:

$$W_p^{-1}(\Lambda; D, d) \propto |\Lambda|^{-d/2} \exp \left\{ -\frac{1}{2} \text{tr}(\Lambda^{-1}D) \right\},$$

$\Omega_1 + \Omega_2 = \Sigma$ with $\Omega_1 = \delta \Sigma$, and $\delta \in [0, 1]$.

4.2. The Gibbs Sampler

Based on the HCSMN model structure with the likelihood and the prior distributions in Equation (19), the joint posterior distribution of θ, Λ and $\eta = (\eta_1, \dots, \eta_n)^\top$ given the data $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ is:

$$\begin{aligned} p(\theta, \Lambda, \eta | \text{Data}) &\propto \prod_{i=1}^n |\kappa(\eta_i)\Lambda|^{-1/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\Lambda^{-1} \kappa(\eta_i)^{-1} (\mathbf{y}_i - \theta)(\mathbf{y}_i - \theta)^\top \right] \right\} \\ &\times \phi_p(\theta; \mu_0, \Omega_2) \phi_p(\mu_0; \theta_0, \Omega_1) I(\mu_0 \in \mathbf{C}) \\ &\times |\Lambda|^{-d/2} \exp \left\{ -\frac{1}{2} \text{tr}(\Lambda^{-1}D) \right\} \prod_{i=1}^n g_i(\eta_i), \end{aligned} \tag{20}$$

where $g(\eta_i)$'s denote the densities of the mixing variables η_i 's. Note that the joint posterior of Equation (20) is not simplified in an analytic form of the known density and, thus, intractable for the posterior inference. Instead, we use the Gibbs sampler for the posterior inference. See [30] for a reference. To run the Gibbs sampler, we need the following full conditional posterior distributions:

- (i) The full conditional posterior densities of η_i 's are given by:

$$p(\eta_i | \boldsymbol{\theta}, \Lambda, \mathbf{y}_i) \propto \kappa(\eta_i)^{-\frac{p}{2}} \exp \left\{ -\frac{(\mathbf{y}_i - \boldsymbol{\theta})^\top \Lambda^{-1} (\mathbf{y}_i - \boldsymbol{\theta})}{2\kappa(\eta_i)} \right\} g(\eta_i), \quad i = 1, \dots, n, \quad (21)$$

- (ii) The full conditional distribution of $\boldsymbol{\theta}$ is obtained by using the way analogous to the proof of Corollary 7. It is:

$$[\boldsymbol{\theta} | \Lambda, \boldsymbol{\eta}, Data] \sim RSN_p(\mathbf{C}; \boldsymbol{\tau}_{pos}, \boldsymbol{\Psi}_{pos}), \quad (22)$$

where:

$$\boldsymbol{\tau}_{pos} = \begin{pmatrix} \boldsymbol{\tau}_0 \\ \boldsymbol{\tau}_1 \end{pmatrix}, \quad \boldsymbol{\Psi}_{pos} = \begin{pmatrix} \Omega_0^* & \delta\Omega^* \\ \delta\Omega^* & \Omega^* \end{pmatrix}, \quad \boldsymbol{\tau}_0 = (1 - \delta)\boldsymbol{\theta}_0 + \delta\boldsymbol{\tau}_1,$$

$\boldsymbol{\tau}_1 = \Omega^* (\boldsymbol{\Sigma}^{-1} \boldsymbol{\theta}_0 + \sum_{i=1}^n \kappa(\eta_i) \Lambda^{-1} \mathbf{y}_i)$, $\Omega^* = (\boldsymbol{\Sigma}^{-1} + \sum_{i=1}^n \kappa(\eta_i) \Lambda^{-1})^{-1}$, and $\Omega_0^* = \delta(1 - \delta)\boldsymbol{\Sigma} + \delta^2 \Omega^*$.

- (iii) The full conditional posterior distribution of Λ is an inverse-Wishart distribution:

$$[\Lambda | \boldsymbol{\theta}, \boldsymbol{\eta}, Data] \sim W_p^{-1}(V, m), \quad m > 2p, \quad (23)$$

where $V = D + \sum_{i=1}^n \kappa(\eta_i)^{-1} (\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})^\top$ and $m = n + d$.

4.3. Markov Chain Monte Carlo Sampling Scheme

When conducting a posterior inference of the HCSMN model, using the Gibbs sampling algorithm with the full conditional posterior distributions of η_i 's, $\boldsymbol{\theta}$ and Λ , the following points should be noted.

note 1: Variable η_i at $\kappa(\eta_i) = 1$, i.e., the HCN (hierarchical constrained normal) model with $\boldsymbol{\epsilon}_i \stackrel{iid}{\sim} N_p(\mathbf{0}, \Lambda)$, $i = 1, \dots, n$, the Gibbs sampler consists of two conditional distributions $[\boldsymbol{\theta} | \Lambda, Data]$ and $[\Lambda | \boldsymbol{\theta}, Data]$. To sample from the first full conditional posterior distribution, we can utilize the stochastic representations of the RSN distribution in Corollary 8. The R package `tmvtnorm` and the R package `mvtnorm` can be used to sample from the RSN distribution in Equation (22).

note 2: According to choice of the distribution η_i and the mixing function $\kappa(\eta_i)$, the HCSMN model may produce a different model other than the HCN model, such as hierarchical constrained multivariate t_v (HC t_v), hierarchical constrained multivariate *logit*, hierarchical constrained multivariate *stable* and hierarchical constrained multivariate *exponential power* models. See, e.g., [31,32], for various distributions of η_i and corresponding function $\kappa(\eta_i)$, which can be used to construct the HCSMN model.

note 3: When the hierarchical constrained multivariate t_v (HC t_v) model is considered, the hierarchy of the model in Equation (19) consists of $\boldsymbol{\epsilon}_i \stackrel{iid}{\sim} N_p(\mathbf{0}, \kappa(\eta_i) \Lambda)$ with $\kappa(\eta_i) = \eta_i^{-1}$ and $\eta_i \sim \text{Gamma}(v/2, v/2)$, $i = 1, \dots, n$. Thus, the Gibbs sampler comprises the conditional posterior Equations (21)–(23). Under the HC t_v model, the distribution of Equation (21) reduces to:

$$[\eta_i | \boldsymbol{\theta}, \Lambda, \mathbf{y}_i] \sim \text{Gamma}(v^*/2, h/2),$$

where $v^* = p + v$ and $h = v + (\mathbf{y}_i - \boldsymbol{\theta})^\top \Lambda^{-1} (\mathbf{y}_i - \boldsymbol{\theta})$. To limit model complexity, we consider only fixed v , so that we can investigate different HC t_v models. As suggested by [32], a uniform prior on $1/v$ ($0 < 1/v < 1$) can be considered. However, this will bring additional computational burden.

note 4: Except for the HCN and HCT_v models, the Metropolis–Hastings algorithm within the Gibbs sampler is used for estimating the HCSMN models, because the conditional posterior densities Equation (20) do not have explicit forms of known distributions as in Equations (21) and (22). See, e.g., [22], for the algorithm for sampling η_i from various mixing distributions, $g_i(\eta_i)$. A general procedure for the algorithm is as follows: Given the current values $\Theta = \{\boldsymbol{\eta}, \boldsymbol{\theta}, \Lambda\}$, we independently generate a candidate η_i from a proposal density $q(\eta_i^*|\eta_i) = g_i(\eta_i^*)$, as suggested by [33], which is used for a Metropolis–Hastings algorithm. Then, accept the candidate value with the acceptance rate:

$$\alpha(\eta_i, \eta_i^*) = \min \left\{ \frac{p(\Theta|\eta_i^*)}{p(\Theta|\eta_i)}, 1 \right\}$$

$i = 1, \dots, n$. Because the target density is proportional to $p(\Theta|\eta_i)g_i(\eta_i)$ and $p(\Theta|\eta_i) = \phi_p(\mathbf{y}_i; \boldsymbol{\theta}, \kappa(\eta_i)\Lambda)$ is uniformly bounded for $\eta_i > 0$.

note 5: As noted from Equations (8) and (9), the second and third stage priors of the HCSMN model in Equation (19) reduce to the two-stage prior $\pi_{two}(\boldsymbol{\theta})$, eliciting the stochastic multivariate interval constraint with degree of uncertainty $1 - \gamma_{two}$. Instead, if the maximum entropy prior $\pi_{max}(\boldsymbol{\theta})$ and the constrained maximum entropy prior $\pi_{const}(\boldsymbol{\theta})$ are used for the HCSMN, then the respective full conditional distributions of $\boldsymbol{\theta}$ of the Gibbs sampler change from Equation (22) to:

$$\left[\boldsymbol{\theta} \mid \Lambda, \boldsymbol{\eta}, Data \right] \sim N_p(\boldsymbol{\tau}_1, \Omega^*) \text{ for } \boldsymbol{\theta} \in \mathbb{R}^p \text{ and } \left[\boldsymbol{\theta} \mid \Lambda, \boldsymbol{\eta}, Data \right] \sim N_p(\boldsymbol{\tau}_1, \Omega^*) I(\boldsymbol{\theta} \in \mathbf{C}),$$

where $\boldsymbol{\tau}_1$ and Ω^* are the same as given in Equation (22).

4.4. Bayes Estimation

For a simple example, let us consider the HCN model with known Λ . When we assume a stochastic constraint $\{\boldsymbol{\theta}; \boldsymbol{\theta} \in \mathbf{C}\}$ obtained from *a priori* information, we may use the two-stage maximum entropy prior $\pi_{two}(\boldsymbol{\theta})$ defined by the second and third stages of the HCSMN model (19) with $\delta \in (0, 1)$, where the value of δ is determined by using Corollary 6. This yields a Bayes estimate based on the two-stage maximum entropy prior. Corollary 8 yields:

$$\hat{\boldsymbol{\theta}}_{two} = \boldsymbol{\tau}_1 + \delta \Omega^* \Omega_0^{*-1} E[\boldsymbol{\theta}_{tn}^*] = \boldsymbol{\tau}_1 + \delta \Omega^* \Omega_0^{*-1} \boldsymbol{\zeta} \tag{24}$$

and:

$$\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_p)^\top \text{ with } \zeta_i = w_{0i} \frac{\phi(u_i/w_{0i}) - \phi(v_i/w_{0i})}{\Phi(v_i/w_{0i}) - \Phi(u_i/w_{0i})}, \quad i = 1, \dots, p,$$

where $\boldsymbol{\theta}_{tn}^* \sim N_p(\mathbf{0}, \Omega_0^*) I(\boldsymbol{\theta}_{tn}^* \in (\mathbf{u}, \mathbf{v}))$, a truncated normal distribution with $\mathbf{u} = \mathbf{a} - \boldsymbol{\tau}_0 = (u_1, \dots, u_p)^\top$ and $\mathbf{v} = \mathbf{b} - \boldsymbol{\tau}_0 = (v_1, \dots, v_p)^\top$ and w_{0i} denotes i -th diagonal element of Ω_0^* . Here, $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_0$ are the same as those in Equation (22), and $\phi(\cdot)$ denotes the univariate standard normal density function. See [25,34] for the first moment of the truncated multivariate normal distribution and for a numerical calculation of the posterior covariance matrix $Cov(\boldsymbol{\theta}_{tn}^*)$, respectively.

On the other hand, when we have certainty about the constraint $\{\boldsymbol{\theta}; \boldsymbol{\theta} \in \mathbf{C}\}$, we may use the HCSMN model with $\delta = 1$, which uses the constrained maximum entropy prior $\pi_{const}(\boldsymbol{\theta})$ instead of $\pi_{two}(\boldsymbol{\theta})$ in its hierarchy. This case gives the Bayes estimate:

$$\hat{\boldsymbol{\theta}}_{const} = E[\boldsymbol{\theta}_{tn}] = \boldsymbol{\tau}_1 + \boldsymbol{\zeta}^* \tag{25}$$

and:

$$\boldsymbol{\zeta}^* = (\zeta_1^*, \dots, \zeta_p^*)^\top \text{ with } \zeta_i^* = w_i \frac{\phi(a_i/w_i) - \phi(b_i/w_i)}{\Phi(b_i/w_i) - \Phi(a_i/w_i)}, \quad i = 1, \dots, p,$$

where $\boldsymbol{\theta}_{tn} \sim N_p(\boldsymbol{\tau}_1, \Omega^*) I(\boldsymbol{\theta}_{tn} \in \mathbf{C})$ and w_i denotes i -th diagonal element of Ω^* .

On the contrary, when we have completely no *a priori* information about the constraint in the space of θ , the HCSMN model with the maximum entropy prior $\pi_{max}(\theta)$ (equivalently, the HCSMN model with $\delta = 0$) may be used for the posterior inference. In this model, the Bayes estimate of the location parameter is given by:

$$\hat{\theta}_{max} = \tau_1. \quad (26)$$

Comparing Equations (24) and (25) to Equation (26), we see that Equations (24) and (25) are the same for $\delta = 1$, and the last term in Equation (24) vanishes when we assume that there is no *a priori* information about the stochastic constraint, $\{\theta; \theta \in \mathbf{C}\}$. In this sense, the last term in Equation (24) can be interpreted as a *shrinkage effect* of the HCSMN model with $\delta \neq 0$. This effect makes the Bayes estimator of θ shrink toward the stochastic constraint. In addition, we can calculate the difference between the estimates in Equations (24) and (25):

$$Diff = \hat{\theta}_{const} - \hat{\theta}_{two} = \zeta^* - \delta \Omega^* \Omega_0^{*-1} \zeta.$$

This difference vector is a function of the degree of belief γ_{two} or $\delta \in (0, 1)$ for Equation (25) is based on $\gamma_{const} = 1$ and $\delta = 1$ and $Diff = \mathbf{0}$ for $\delta = 1$. Thus, the difference represents a *stochastic effect* of the multivariate interval constraint.

5. Numerical Illustrations

This section presents an empirical analysis of the proposed approach (using the HCSMN model) to the stochastic multivariate interval constraint on the location model. We provide numerical simulation results and a real data application comparing the proposed approach to the hierarchical Bayesian approaches, which use usual priors, $\pi_{max}(\theta)$ and $\pi_{const}(\theta)$. For numerical implementations, we develop our program written in R, which is available from the author upon request.

5.1. Simulation Study

To examine the performance of the HCSMN model for estimating the location parameter with a stochastic multivariate interval constraint, we conduct a simulation study. The study is based up 200 synthetic datasets for different sample sizes $n = 20, 200$ generated from each distribution of $N_4(\theta, \Lambda)$ and $t_4(\theta, \Lambda, \nu)$, a four-dimensional t with the location parameter θ , scale matrix Λ and degrees of freedom $\nu = 5$. For the simulation, we used the following choice of parameter values: $\theta = (\theta, \theta, m\theta, m\theta)^\top$ and $\Lambda = (1 - \rho)I_4 + \rho \mathbf{1}_4 \mathbf{1}_4^\top$, where $m = (-1)^{\theta+1}$, $\rho = 0.5$, and $\theta = 1, 2$.

To fit each of the 200 synthetic datasets (Dataset I) generated from the $N_4(\theta, \Lambda)$ distribution, we implemented the Markov chain Monte Carlo (MCMC) posterior simulation with the three different HCN models with the multivariate interval constraint $\mathbf{C} = (\mathbf{a}, \mathbf{b})^\top$: the HCN models that use $\pi_{two}(\theta)$, $\pi_{max}(\theta)$, and $\pi_{const}(\theta)$. We denote these models by $\text{HCN}(\pi_{two})$, $\text{HCN}(\pi_{max})$ and $\text{HCN}(\pi_{const})$. For each dataset, MCMC posterior sampling was based on the first 10,000 posterior samples as the burn-in, followed by a further 100,000 posterior samples with a thinning size of 10. Thus, the final MCMC posterior samples with a size of 10,000 were obtained for each of the three HCN models. Exactly the same MCMC posterior sampling scheme is applied to each of the 200 synthetic datasets (Dataset II) from the $t_4(\theta, \Lambda, \nu)$ distribution based on the three $\text{HC}t_\nu$ models, $\text{HC}t_\nu(\pi_{two})$, $\text{HC}t_\nu(\pi_{max})$ and $\text{HC}t_\nu(\pi_{const})$. To satisfy a subjective perspective of the hierarchical models, we set $\theta_0 = \mathbf{0}$, $\Sigma = \Omega_1 + \Omega_2 = 0.5\theta I_4$, and $\Omega_1 = 0.85\Sigma$ to specify our information about the parameter θ , while we set $D = 10^{-2}I_4$ and $d = 2p + 5$ to elicit no information about Λ (see, e.g., [32]). For the stochastic multivariate interval constraint, we set $\mathbf{a} = -0.5\theta \mathbf{1}_4$ and $\mathbf{b} = 0.5\theta \mathbf{1}_4$, and this constraint gives the degree of belief $\gamma_{max} = 0.073$ (or 0.217) and $\gamma_{two} = 0.394$ (or 0.571) for $\theta = 1$ (or 2). Note that the degree of belief in the constraint, accounted for by $\pi_{const}(\theta)$, is $\gamma_{const} = 1$ for all of the values of θ .

Summary statistics of the posterior samples of the location parameters (the mean and the standard deviation of 200 posterior means of each parameter) along with the degrees of belief about the

constraint \mathbf{C} (γ_{max} , γ_{two} and γ_{const}) are listed in Table 1. For the sake of saving a space, we omit the summary statistics regarding Λ from the table. The table indicates the followings: (i) The MCMC method performs well in estimating the location parameters of all of the models considered. This can be justified by the estimation results of the $\text{HCN}(\pi_{max})$ and $\text{HC}t_V(\pi_{max})$ models. Specifically, in the posterior estimation of θ , the data information tends to dominate the prior information about θ for the large sample case (*i.e.*, $n = 200$), while the latter tends to dominate the former for the small sample case of $n = 20$. Furthermore, the convergence of the MCMC sampling algorithm was evident, and a discussion about the convergence will be given in Subsection 5.2; (ii) The estimates of θ obtained from the $\text{HCN}(\pi_{two})$ and $\text{HC}t_V(\pi_{two})$ models are uniformly closer to the stochastic constraint $\theta \in \mathbf{C}$ than those from the $\text{HCN}(\pi_{max})$ and $\text{HC}t_V(\pi_{max})$ models. This confirms that $\pi_{two}(\theta)$ induces an obvious shrinkage effect in Bayesian estimation of the location parameter with a stochastic multivariate interval constraint; (iii) Comparing the estimates of θ obtained from the $\text{HCN}(\pi_{two})$ (or $\text{HC}t_V(\pi_{two})$) model to those from the $\text{HCN}(\pi_{const})$ (or $\text{HC}t_V(\pi_{const})$) model, we see that the difference between their vector values is significant. Thus, we can expect an apparent stochastic effect if we use $\pi_{two}(\theta)$ in Bayesian estimation of the location parameter with a stochastic multivariate interval constraint.

Table 1. Summaries of posterior samples of $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^\top$ obtained by using three different priors; $\pi_{two}(\theta)$, $\pi_{max}(\theta)$ and $\pi_{const}(\theta)$. HCN, hierarchical constrained normal.

Dataset I	HCN(π_{max})				HCN(π_{two})				HCN(π_{const})			
	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
<i>n</i> = 20												
true	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
mean	0.853	0.862	0.869	0.857	0.525	0.530	0.542	0.549	0.141	0.112	0.139	0.121
s.d.	0.203	0.212	0.196	0.194	0.189	0.187	0.176	0.167	0.099	0.111	0.096	0.055
<i>n</i> = 200												
true	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
mean	0.978	0.979	0.981	0.977	0.912	0.914	0.916	0.912	0.342	0.291	0.294	0.299
s.d.	0.073	0.072	0.074	0.068	0.060	0.067	0.069	0.064	0.037	0.058	0.077	0.064
<i>n</i> = 20												
true	2.000	2.000	-2.000	-2.000	2.000	2.000	-2.000	-2.000	2.000	2.000	-2.000	-2.000
mean	1.837	1.846	-1.853	-1.842	1.306	1.314	-1.331	-1.349	0.599	0.501	-0.483	-0.493
s.d.	0.215	0.223	0.208	0.205	0.295	0.271	0.262	0.238	0.232	0.263	0.291	0.141
<i>n</i> = 200												
true	2.000	2.000	-2.000	-2.000	2.000	2.000	-2.000	-2.000	2.000	2.000	-2.000	-2.000
mean	1.977	1.979	-1.980	-1.977	1.782	1.784	-1.785	-1.782	0.884	0.856	-0.795	-0.733
s.d.	0.074	0.072	0.074	0.068	0.068	0.065	0.068	0.062	0.029	0.051	0.068	0.062
Dataset II	HCt ₅ (π_{two})				HCt ₅ (π_{max})				HCt ₅ (π_{const})			
	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
<i>n</i> = 20												
true	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
mean	0.681	0.691	0.688	0.774	0.332	0.368	0.403	0.361	0.155	0.157	0.158	0.161
s.d.	0.172	0.182	0.201	0.205	0.198	0.190	0.218	0.186	0.108	0.112	0.120	0.099
<i>n</i> = 200												
true	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
mean	1.003	0.974	1.033	1.006	0.802	0.818	0.814	0.806	0.362	0.411	0.341	0.351
s.d.	0.175	0.178	0.172	0.199	0.061	0.065	0.070	0.081	0.048	0.052	0.039	0.047
<i>n</i> = 20												
true	2.000	2.000	-2.000	-2.000	2.000	2.000	-2.000	-2.000	2.000	2.000	-2.000	-2.000
mean	1.724	1.763	-1.764	-1.675	0.886	0.874	-0.857	-0.924	0.415	0.428	-0.496	-0.489
s.d.	0.239	0.230	0.245	0.231	0.324	0.319	0.343	0.313	0.201	0.196	0.193	0.204
<i>n</i> = 200												
true	2.000	2.000	-2.000	-2.000	2.000	2.000	-2.000	-2.000	2.000	2.000	-2.000	-2.000
mean	2.018	1.943	-1.991	-2.103	1.702	1.699	-1.715	-1.689	0.967	0.986	-0.965	-0.959
s.d.	0.079	0.082	0.081	0.073	0.096	0.099	0.091	0.089	0.053	0.049	0.047	0.045

5.2. Car Body Assembly Data Example

John and Wichern consider car body assembly data (accessible through www.prenhall.com/statistics, [35]) obtained from a study of its sheet metal assembly process. A major automobile manufacturer uses sensors that record the deviation from the nominal thickness (millimeters $\times 10^{-1}$) at

a specific location on a car, which has the following levels: the deviation of the car body at the final stage of assembly (Y_1) and that at an early stage of assembly (Y_2). The data consist of 50 pairs of observations of (Y_1, Y_2) , and they provide summary statistics as listed in Table 2. The tests given by ([36], p. 148), using the measures of multivariate skewness and kurtosis, accept the bivariate normality of the joint distribution of $Y = (Y_1, Y_2)^T$. The respective skewness and kurtosis are $b_{1p} = 0.074$ and $b_{2p} = 7.337$, which give respective p -values of 0.954 (chi-square test for the skewness) and 0.721 (normal test for the kurtosis), indicating the observation model for the dataset is:

$$y_i = \theta + \epsilon_i, \quad i = 1, \dots, 50,$$

where $\epsilon_i \stackrel{iid}{\sim} N_2(\mathbf{0}, \Lambda)$, $\Lambda = \{\lambda_{ij}\}$. The Shapiro–Wilk (S-W) test is also implemented to see the marginal normality of each $Y_i, i = 1, 2$. The test statistic values and corresponding p -values of the S-W test are listed in Table 2.

Table 2. Summary statistics for the car body assembly data. S-W, Shapiro–Wilk.

Variable	Mean	s.d.	S-W	p -Value
Y_1	−1.996	2.781	0.959	0.083
Y_2	7.426	5.347	0.989	0.926

In practical situations, we may have information about the mean vector of the observation model (*i.e.*, mean deviation from the nominal thickness) from a past study of the sheet metal assembly process or a quality control report of the automobile manufacturer. Suppose that the information about the centroid of the mean deviation vector, $\theta = (\theta_1, \theta_2)^T$, is $(-1, 4)^T$ with $Cov(\theta) = diag\{1, 4\}$. Furthermore, there is uncertain information that $\theta \in (\mathbf{a}, \mathbf{b})$, where $\mathbf{a} = (-1.5, 3)^T$ and $\mathbf{b} = (-0.5, 5)^T$. This paper has proposed the two-stage maximum entropy prior $\pi_{two}(\theta)$ to represent all of the information, which is not available with the other priors, such as $\pi_{max}(\theta)$ and $\pi_{const}(\theta)$.

Using the three hierarchical models (*i.e.*, the HCN(π_{max}), HCN(π_{two}) and HCN(π_{const}) models), we obtain 10,000 posterior samples from the MCMC sampling scheme based on each of the three models with a 10 thinning period after a burn-in period of 10,000 samples. In estimating the Monte Carlo (MC) error, we used the batch mean method with 50 batches; see, e.g., [37] (pp. 39–40). For a formal test for the convergence of the MCMC algorithm, we applied the Heidelberger–Welch diagnostic test of [38] to single-chain MCMC runs and calculated the p -values of the test. For the posterior simulation, we used the following choice of hyper-parameter values: $\theta_0 = (-1, 4)^T$, $\Sigma = \Omega_1 + \Omega_2 = 10I_2$, $\Omega_1 = \delta\Sigma$, $\Omega_2 = (1 - \delta)\Sigma$, $\delta \in (0, 1)$, $D = 10^{-2}I_2$ and $d = 10^2 + 2p + 1$. The posterior estimation and the convergence test results are shown in Table 3. Note that Columns 7–9 of the table list the values obtained from implementing the MCMC sampling for the posterior estimation of HCN(π_{two}).

Table 3. The posterior estimates and the convergence test results.

δ	γ_{two}	Parameter	HCN(π_{max})	HCN(π_{const})	HCN(π_{two})	s.d.	MC Error	p -Value
0.8	0.423	θ_1	−1.874	−1.321	−1.665	0.581	0.005	0.483
		θ_2	7.045	4.814	6.332	1.089	0.008	0.354
		λ_{11}	7.549	7.863	7.682	1.231	0.007	0.551
		λ_{12}	−4.632	−4.905	−4.819	1.631	0.005	0.671
		λ_{22}	26.872	25.021	27.351	4.926	0.013	0.352
0.9	0.567	θ_1	−1.874	−1.321	−1.557	0.496	0.004	0.434
		θ_2	7.045	4.814	5.905	1.112	0.006	0.298
		λ_{11}	7.526	7.781	7.959	1.317	0.008	0.635
		λ_{12}	−4.726	−5.347	−4.989	1.546	0.006	0.712
		λ_{22}	27.587	25.347	28.012	4.836	0.012	0.384

The small MC error values listed in Table 3 convince us of the convergence of the MCMC algorithm. Furthermore, the p -values of the Heidelberger–Welch test for the stationarity of the single MCMC run are larger than 0.1. Thus, both of the diagnostic checking methods advocate the convergence of the proposed MCMC sampling scheme. Similar to Table 1, this table also shows that $\pi_{two}(\boldsymbol{\theta})$ induces the shrinkage and stochastic effects in the Bayesian estimation of $\boldsymbol{\theta}$ with the uncertain multivariate interval constraint: (i) From the comparison of the posterior estimates obtained from $\text{HCN}(\pi_{two})$ with those from $\text{HCN}(\pi_{max})$, we see that the estimates of θ_1 and θ_2 , obtained from $\text{HCN}(\pi_{two})$, shrink toward the stochastic interval \mathbf{C} . The magnitude of shrinkage effect induced by using the proposed prior $\pi_{two}(\boldsymbol{\theta})$ becomes more evident as the degree of belief in the interval constraint γ_{two} (or δ) gets larger; (ii) On the other hand, we can see the stochastic effect of the prior $\pi_{two}(\boldsymbol{\theta})$ by comparing the posterior estimate of $\boldsymbol{\theta}$ obtained from $\text{HCN}(\pi_{two})$ with that from $\text{HCN}(\pi_{const})$. The stochastic effect can be measured by the difference between the estimates, and we see that the difference becomes smaller as γ_{two} (or δ) gets larger.

6. Conclusions

In this paper, we have proposed a two-stage maximum entropy prior $\pi_{two}(\boldsymbol{\theta})$ of the location parameter of a scale mixture of normal model. The prior is derived by using the two stages of a prior hierarchy advocated by [5] to elicit a stochastic multivariate interval constraint, $\{\boldsymbol{\theta}; \boldsymbol{\theta} \in \mathbf{C}\}$. With regard to eliciting the stochastic constraint, the two-stage maximum entropy prior has the following properties. (i) Theorem 1 and Corollary 4 indicate that the two-stage prior is flexible enough to elicit all of the degrees of belief in the stochastic constraint; (ii) Corollary 4 confirms that the entropy of the two-stage prior is commensurate with the uncertainty about the constraint $\{\boldsymbol{\theta}; \boldsymbol{\theta} \in \mathbf{C}\}$; (iii) As given in Corollary 6, the preceding two properties enable us to propose an objective way of eliciting the uncertain prior information by using $\pi_{two}(\boldsymbol{\theta})$. From the inferential view point: (i) the two-stage prior for the normal mean vector has the conjugate property that the prior and posterior distributions belong to the same family of the *RSN* distributions by [23]; (ii) the conjugate property enables us to construct an analytically simple Gibbs sampler for the posterior inference of the model (1) with unknown covariance matrix Λ ; (iii) this paper also provides the HCSMN model, which is flexible enough to elicit all of the types of stochastic constraints and the scale mixture for Bayesian inference of the model (1). Based on the HCSMN model, the full conditional posterior distributions of unknown parameters were derived, and the calculation of posterior summary was discussed by using the Gibbs sampler and two numerical applications.

The methodological results of the Bayesian estimation procedure proposed in the paper can be extended to other multivariate models that incorporate functional means, such as linear and nonlinear regression models. For example, the seemingly unrelated regression (SUR) model and the factor analysis model (see, e.g., [24]) can be explained in the same framework of the proposed HCSMN in Equation (1). We hope to address these issues in the near future.

Acknowledgments: The research of Hea-Jung Kim was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01057106).

Conflicts of Interest: The author declares no conflict of interest.

References

1. O'Hagan, A. Bayes estimation of a convex quadratic. *Biometrika* **1973**, *60*, 565–572.
2. Steiger, J. When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychol. Methods* **2002**, *7*, 210–227.
3. Lopes, H.F.; West, M. Bayesian model assessment in factor analysis. *Stat. Sin.* **2004**, *14*, 41–67.
4. Loken, E. Identification constraints and inference in factor models. *Struct. Equ. Model.* **2005**, *12*, 232–244.

5. O'Hagan, A.; Leonard, T. Bayes estimation subject to uncertainty about parameter constraints. *Biometrika* **1976**, *63*, 201–203.
6. Liseo, B.; Loperfido, N. A Bayesian interpretation of the multivariate skew-normal distribution. *Stat. Probab. Lett.* **2003**, *49*, 395–401.
7. Kim, H.J. On a class of multivariate normal selection priors and its applications in Bayesian inference. *J. Korean Stat. Soc.* **2011**, *40*, 63–73.
8. Kim, H.J. A measure of uncertainty regarding the interval constraint of normal mean elicited by two stages of a prior hierarchy. *Sic. World J.* **2014**, *2014*, 676545, doi:10.1155/2014/676545.
9. Kim, H.J.; Choi, T. On Bayesian estimation of regression models subject to uncertainty about functional constraints. *J. Korean Stat. Soc.* **2015**, *43*, 133–147.
10. Kim, H.J.; Choi, T.; Lee, S. A hierarchical Bayesian regression model for the uncertain functional constraint using screened scale mixture of Gaussian distributions. *Statistics* **2016**, *50*, 350–376.
11. Arellano-Valle, R.B.; Branco, M.D.; Genton, M.G. A unified view on skewed distributions arising from selection. *Can. J. Stat.* **2006**, *34*, 581–601.
12. Kim, H.J. A class of weighted multivariate normal distributions and its properties. *J. Multivar. Anal.* **2008**, *99*, 1758–1771.
13. Jaynes, E.T. Prior probabilities. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 227–241.
14. Jaynes, E.T. *Papers on Probability, Statistics, and Statistical Physics*; Rosenkrantz, R.D., Ed.; Reidel: Boston, MA, USA, 1983.
15. Smith, S.R.; Grandy, W. *Maximum-Entropy and Bayesian Methods in Inverse Problems*; Reidel: Boston, MA, USA, 2013.
16. Ishwar, P.; Moulin, P. On the existence of characterization of the Maxent distribution under general moment inequality constraints. *IEEE Trans. Inf. Theory* **2005**, *51*, 3322–3333.
17. Rosenkrantz, R.D. *Inference, Method, and Decision: Towards a Bayesian Philosophy and Science*; Reidel: Boston, MA, USA, 1977.
18. Rosenkrantz, R.D. (Ed.) *E.T. Jaynes: Papers on Probability, Statistics, and Statistical Physics*; Kluwer Academic: Dordrecht, The Netherlands, 1989.
19. Yuen, K.V. *Bayesian Methods for Structural Dynamics and Civil Engineering*; John Wiley & Sons: Singapore, Singapore, 2010.
20. Wu, N. *The Maximum Entropy Method*; Springer: New York, NY, USA, 2012.
21. Cercignani, C. *The Boltzman Equation and Its Applications*; Springer: Berlin/Heidelberg, Germany, 1988.
22. Leonard, T.; Hsu, J.S.J. *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*; Cambridge University Press: New York, NY, USA, 1999.
23. Kim, H.J.; Kim, H.-M. A Class of Rectangle-Screened Multivariate Normal Distributions and Its Applications. *Statistics* **2015**, *49*, 878–899.
24. Press, S.J. *Applied Multivariate Analysis*, 2nd ed.; Dover Publications, Inc.: New York, NY, USA, 2005.
25. Wilhelm, S.; Manjunath, B.G. *tmvtnorm: Truncated Multivariate Normal Distribution and Student t Distribution*. Available online: <http://CRAN.R-project.org/package=tmvtnorm> (accessed on 17 May 2016).
26. Genz, A.; Bretz, F. *Computation of Multivariate Normal and t Probabilities*; Springer: New York, NY, USA, 2009.
27. Gupta, S.D. A note on some inequalities for multivariate normal distribution. *Bull. Calcutta Stat. Assoc.* **1969**, *18*, 179–180.
28. Lindly, D.V. *Bayesian Statistics: A Review*; SIAM: Philadelphia, PA, USA, 1970.
29. Khuri, A.I. *Advanced Calculus with Applications in Statistics*; John Wiley & Son: New York, NY, USA, 2003.
30. Gamerman, D.; Lopes, H.F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed.; Chapman and Hall: New York, NY, USA, 2006.
31. Branco, M.D. A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113.
32. Chen, M.-H.; Dey, D.K. Bayesian modeling of correlated binary response via scale mixture of multivariate normal link functions. *Sankhyā* **1998**, *60*, 322–343.
33. Chib, S.; Greenberg, E. Understanding the Metropolis-Hastings algorithm. *Am. Stat.* **1995**, *49*, 327–335.
34. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Distribution in Statistics: Continuous Univariate Distributions*, 2nd ed.; John Wiley & Son: New York, NY, USA, 1994; Volume 1.
35. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*, 6th ed.; Pearson Prentice Hall: London, UK, 2007.

36. Mardia, K.V.; Kent, J.T.; Bibby, J.M. *Multivariate Analysis*; Academic Press: London, UK, 1979.
37. Ntzoufras, I. *Bayesian Modeling Using WinBUGS*; John Wiley & Son: New York, NY, USA, 2009.
38. Heidelberger, P.; Welch, P. Simulation run length control in the presence of an initial transient. *Oper. Res.* **1992**, *31*, 1109–1144.



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).