Research Article

A Measure of Uncertainty regarding the Interval Constraint of Normal Mean Elicited by Two Stages of a Prior Hierarchy

Hea-Jung Kim

Department of Statistics, Dongguk University, Seoul Campus, Pil-Dong 3Ga, Chung-Gu, Seoul 100-715, Republic of Korea

Correspondence should be addressed to Hea-Jung Kim; kim3hj@dongguk.edu

Received 3 March 2014; Revised 18 June 2014; Accepted 20 June 2014; Published 6 July 2014

Academic Editor: Montserrat Guillén

Copyright © 2014 Hea-Jung Kim. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper considers a *hierarchical screened Gaussian model* (HSGM) for Bayesian inference of normal models when an interval constraint in the mean parameter space needs to be incorporated in the modeling but when such a restriction is uncertain. An objective measure of the uncertainty, regarding the interval constraint, accounted for by using the HSGM is proposed for the Bayesian inference. For this purpose, we drive a maximum entropy prior of the normal mean, eliciting the uncertainty regarding the interval constraint, and then obtain the uncertainty measure by considering the relationship between the maximum entropy prior and the marginal prior of the normal mean in HSGM. Bayesian estimation procedure of HSGM is developed and two numerical illustrations pertaining to the properties of the uncertainty measure are provided.

1. Introduction

Consider the following model for normally distributed data:

$$y_i = \theta + \epsilon_i, \quad \epsilon_i \sim N(0, \tau^2), \quad i = 1, \dots, n. \quad (1)$$

Bayesian analysis of the model (1) begins with the specification of prior distributions for unknown parameters θ and the noise variance τ^2 . Specifically, we assign a normal prior distribution for θ and an inverse gamma (IG) prior for τ^2 , that is, $\theta \sim N(\mu, \sigma^2)$ and $\tau^2 \sim \text{IG}(c, d)$, which are commonly used in a normal model as conjugate priors, where all the hyperparameters μ and σ^2 , c , and d are assumed to be known in the first place.

On the other hand, when we are completely sure about a functional constraint of θ a priori; a suitable restriction on the parameter space $\Theta = \mathbb{R}$, such as using a truncated normal distribution, is expected. However, it is often the case that the actual observations of (1) may violate the constraint on account of the measurement error or due to some other reasons. Further, the data may provide strong evidence that the constraint is inappropriate and therefore may appear to contradict the theory associated with the constraint. In this respect, it is expected that the uncertainty

about the constraint is taken into account in the estimation procedure. O'Hagan and Leonard [1] indeed proposed two stages of a prior hierarchy based on the truncated prior distribution, which reflects the uncertainty about the parameter constraint. Liseo and Loperfido [2], Kim [3], and Kim and Choi [4] among others considered the Bayesian estimation of normal models with uncertain interval constraints using the idea of two stages of a prior hierarchy. In particular, Kim [3] obtained the marginal prior of θ as the normal selection distribution (e.g., [5]) and thus exploited the class of weighted normal distribution by Kim [6] for reflecting the uncertain prior belief on θ .

Although the two-stage prior is applied by many investigators, there is no objective method to measure (or control) the degree of uncertainty regarding the interval constraint of θ accounted for by using the two stages of a prior hierarchy. This is a major hindrance factor in developing the idea of the two stages of a prior hierarchy which is advocated by O'Hagan and Leonard [1]. Thus, such practical problem motivates us to develop a formal measure of uncertainty about the constraint in order to show how the uncertainty of the prior information regarding interval constrained parameter space Θ of θ can be reflected by utilizing the two stages of a prior hierarchy. This topic is tackled in this paper.

To propose the uncertainty measure, we consider the Bayesian inference of the normal mean θ in (1), but subject to an uncertain interval constraint. Because the maximum entropy prior by Jaynes [7, 8] is useful for describing (or measuring) the relative levels of uncertainty about the distribution of the prior parameter, our investigation focuses on the theoretical relationship between the two stages of a prior hierarchy of θ by O'Hagan and Leonard [1] and the maximum entropy prior subject to an uncertain interval constraint. The remainder of this paper is organized as follows. In Section 2, we briefly discuss the two-stage prior of θ , which will be used for the Bayesian analysis of θ subject to uncertainty regarding the interval constraint. Accordingly, influenced by the seminal work of O'Hagan and Leonard [1], we provide a normal model based on the two-stage prior distribution of θ , referred to as *hierarchical screened Gaussian model (HSGM)*. In Section 3, we explore the theoretical properties of the two-stage prior of θ by using Boltzmann's maximum entropy theorem [9, 10]. Based on the properties, we propose an objective measure of uncertainty regarding the interval constraint of θ that is accounted for by the two-stage prior. In Section 4, we explore Bayesian estimation procedure by analytically deriving the posterior distribution of the unknown parameters under HSGM, and we discuss the properties of the proposed measure of uncertainty that can be explained in the context of HSGM. Finally, the concluding remarks along with a discussion are made in Section 5.

2. Hierarchical Screened Gaussian Model

Let us assume the normal model (1) and consider the two stages of a prior hierarchy in the following way:

$$\begin{aligned} y_i &= \theta + \epsilon_i, \quad \epsilon_i \sim N(0, \tau^2), \quad i = 1, \dots, n, \\ \tau^2 &\sim \text{IG}(c, d), \\ \theta | \theta_0 &\sim N(\theta_0, \sigma_0^2), \quad \text{independent of } (\epsilon_1, \dots, \epsilon_n)^\top, \\ \theta_0 &\sim N_{(a \leq \theta_0 \leq b)}(\mu, \sigma_1^2), \end{aligned} \quad (2)$$

where $N_{(a \leq \mu \leq b)}(\mu, \sigma_1^2)$ denotes a doubly truncated $N(\mu, \sigma_1^2)$ distribution with the lower truncation point a and upper truncation point b . In practice, there are certain cases in which we have a priori information that θ is highly likely to have an interval constraint, and thus the value of θ needs to be located with uncertainty in a restricted space Θ ,

$$\Theta = \{\theta; a \leq \theta \leq b\} \quad \text{or, simply, } \Theta = [a, b]. \quad (3)$$

In order to elicit the prior distribution on the uncertain interval constraint, we utilize the two-stage hierarchical model as in (2), which was initially advocated by O'Hagan and Leonard [1] in which values of θ that do not belong to Θ are penalized to a lesser extent. In this respect, the normal model structure of (2) is referred to as HSGM in the remainder of the paper, because the two stages of a prior hierarchy on θ are considered and the resulting marginal prior distribution of θ becomes the weighted normal (or interval screened normal)

distribution studied by Kim [6]. This is shown as follows. Since $\theta_0 \sim N_{(a,b)}(\mu, \sigma_1^2)$, the marginal prior of θ under HSGM is

$$\begin{aligned} p(\theta) &= \left(\phi(\theta; \mu, \omega^2) \left\{ \Phi(\xi v_1(b) - \eta(\theta - \mu)) \right. \right. \\ &\quad \left. \left. - \Phi(\xi v_1(a) - \eta(\theta - \mu)) \right\} \right) \\ &\quad \times (\Phi(v_1(b)) - \Phi(v_1(a)))^{-1}, \quad \theta \in \mathbb{R}, \end{aligned} \quad (4)$$

where $\phi(\cdot; \mu, \omega^2)$ and $\Phi(\cdot)$, respectively, denote the density of $N(\mu, \omega^2)$ and the distribution function of $N(0, 1)$, $\omega^2 = \sigma_0^2 + \sigma_1^2$, $v_1(a) = (a - \mu)/\sigma_1$, $v_1(b) = (b - \mu)/\sigma_1$, $\xi = \omega/\sigma_0$ and $\eta = \xi\sigma_1/\omega^2$.

We see that (4) is the density of a weighted normal (WN) distribution by Kim [6]. This leads to the following assertion in Lemma 1.

Lemma 1. *The two-stage prior of θ in HSGM of (2), eliciting the uncertain interval constraint (3), is marginally distributed as a weighted normal distribution,*

$$\theta \sim \text{WN}_{(a,b)}(\mu, \Omega) \quad (5)$$

with its density (4), where

$$\mu = \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \quad \Omega = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_0^2 + \sigma_1^2 \end{pmatrix}. \quad (6)$$

Note that $\theta \stackrel{d}{=} [V \mid a \leq U \leq b]$, where $(U, V)^\top \sim N_2(\mu, \Omega)$, a bivariate normal distribution. See Kim [6] for various properties of the $\text{WN}_{(a,b)}(\mu, \Omega)$ distribution. According to O'Hagan and Leonard [1], the first stage variance, σ_0^2 , of the mean may measure the degree of uncertainty in the constraint. However, there is no systematic method to assign the values of σ_0^2 (or σ_1^2), according to a priori specified degree (say, $(1 - \alpha) \times 100\%$) of uncertainty about the interval constraint (3). This is a major hindrance factor in developing the HSGM.

3. The Measure of Uncertainty

3.1. A Maximum Entropy Prior. Sometimes we have a situation where partial prior information is available, outside of which it is desired to use a priori that is as noninformative as possible. A useful method of dealing with this problem is through the concept of entropy by Jaynes [7, 8]. As discussed by Rosenkrantz [11], entropy has a direct relationship to information theory and in a sense measures the amount of uncertainty inherent in the probability distribution.

Assume now that we can specify the partial information concerning a location parameter θ (including the normal mean) with continuous space Θ of the form

$$E[t_j(\theta)] = \int_{\Theta} t_j(\theta) \pi(\theta) d\theta = t_j, \quad j = 1, \dots, k, \quad (7)$$

but with nothing else about our prior distribution $\pi(\theta)$. Then the maximum entropy prior can be obtained by choosing $\pi(\theta)$ that maximizes the entropy

$$\xi(\pi) = - \int_{\Theta} \pi(\theta) \log \pi(\theta) d\theta, \quad (8)$$

in the presence of the partial information in the form of (7). A straightforward application of the calculus of variation leads us to Boltzmann's maximum entropy theorem. This tells us that the density $\pi(\theta)$ that maximizes $\xi(\pi)$, subject to the constraints $E[t_j(\theta)] = t_j$, $j = 1, \dots, k$, takes the k -parameter exponential family form

$$\pi(\theta) \propto \exp \{ \lambda_1 t_1(\theta) + \lambda_2 t_2(\theta) + \dots + \lambda_k t_k(\theta) \}, \quad \theta \in \Theta, \quad (9)$$

where $\lambda_1, \lambda_2, \dots, \lambda_k$ can be determined, via the k -constraints, in terms of $t_1, \dots, t_k$. See Leonard and Hsu [12] for the proof.

3.2. Maximum Entropy Prior for Constrained Normal Mean. Let the location parameter θ of our interest be the normal mean in (1). Then the results of the previous subsection can be applied to the prior for the normal mean θ . This subsection considers the case where the partial priori information of θ is in the form of an interval, that is, $\theta \in \Theta$, $\Theta = [a, b]$ with $a < b$, and examines how the maximum entropy prior of θ , that is, $\pi(\theta)$, has different formula according to the degree of uncertainty regarding the interval constraint.

Case 1. $\Theta = \mathbb{R}$, and $\theta \in \Theta$ with certainty.

When we have partial priori information about θ that we can specify values for both mean μ and variance σ^2 . Then the $N(\mu, \sigma^2)$ prior specification is the maximum entropy prior for θ (e.g., [12]). Thus the prior density and its entropy for the Case 1 are

$$\begin{aligned} \pi_1(\theta) &= \phi(\theta; \mu, \sigma^2), \quad \theta \in \mathbb{R}, \\ \text{Ent}(\pi_1(\theta)) &= - \int_{\Theta} \phi(\theta; \mu, \sigma^2) \ln \phi(\theta; \mu, \sigma^2) d\theta \\ &= \frac{1}{2} (1 + \ln(2\pi) + \ln \sigma^2), \end{aligned} \quad (10)$$

respectively.

Case 2. $\Theta = [a, b]$, and $\theta \in \Theta$ with certainty.

On the other hand, when we are completely sure about the priori interval constraint of θ , a suitable restriction on the parameter space Θ such as using a truncated distribution is expected. This case supposes that we can specify values for both $E[\theta] = \mu$ and $E[(\theta - \mu)^2] = \sigma^2$ on the space $\theta \in \mathbb{R}$ by a priori information. Further suppose that we have certain prior information that the parameter θ is concentrated on the region $[a, b]$, that is,

$$p(a \leq \theta \leq b) = 1 \quad (11)$$

but nothing else about our prior distribution $\pi(\theta)$. The last condition is equivalent to $E[I(a \leq \theta \leq b)] = 1$. Therefore, by Boltzmann's maximum entropy theorem, the prior density of θ for the Case 2 is

$$\begin{aligned} \pi_2(\theta) &\propto \exp \{ \lambda_1 t_1(\theta) + \lambda_2 t_2(\theta) + \lambda_3 t_3(\theta) \} \\ &\propto \exp \left\{ \frac{[\theta - (\mu - \lambda_1/2\lambda_2)]^2}{\lambda_2^{-1}} \right\} \\ &\quad \times \exp \{ \lambda_3 I(a \leq \theta \leq b) \} \end{aligned} \quad (12)$$

by (9), because $t_1(\theta) = \theta$, $t_1 = \mu$, $t_2(\theta) = (\theta - \mu)^2$, $t_2 = \sigma^2$, and $t_3(\theta) = I(a \leq \theta \leq b)$. Since $t_3 = 1$, the maximum entropy prior, subject to these three restrictions, is thus

$$\pi_2(\theta) = \frac{\phi(\theta; \mu, \sigma^2)}{\Phi(v(b)) - \Phi(v(a))}, \quad a \leq \theta \leq b, \quad (13)$$

provided that we choose $\lambda_1 = 0$, $\lambda_2 = -1/2\sigma^2$, $\lambda_3 = -\ln\{\Phi(v(b)) - \Phi(v(a))\}$, where $v(a) = (a - \mu)/\sigma$ and $v(b) = (b - \mu)/\sigma$.

Some algebra using the moments of the $N_{(a,b)}(\mu, \sigma^2)$ distribution in Johnson et al. [13] yields the entropy of $\pi_2(\theta)$ which is

$$\begin{aligned} \text{Ent}(\pi_2(\theta)) &= \frac{1}{2} \left(1 + \ln(2\pi\sigma^2) + \frac{v(a)\phi(v(a)) - v(b)\phi(v(b))}{\Phi(v(b)) - \Phi(v(a))} \right) \\ &\quad + \ln(\Phi(v(b)) - \Phi(v(a))). \end{aligned} \quad (14)$$

Case 3. $\Theta = [a, b]$, and $\theta \in \Theta$ with $(1 - \alpha) \times 100\%$ uncertain.

Now suppose that we have partial priori information that we can specify values for both mean μ and variance σ^2 of θ for $\theta \in \mathbb{R}$. In addition, suppose that we have priori uncertain interval constraint information that $\theta \in [a, b]$. Then along with the priori moment conditions $E[\theta] = \mu$ and $E[(\theta - \mu)^2] = \sigma^2$, the additional uncertain prior information about the interval constraint can be expressed by

$$p(a \leq \theta \leq b) = \alpha, \quad 0 \leq \alpha \leq 1, \quad (15)$$

where $1 - \alpha$ (or $(1 - \alpha) \times 100\%$) denotes the degree of uncertainty. Thus the priori uncertain interval constraint is equivalent to

$$E[I(a \leq \theta \leq b)] = \alpha. \quad (16)$$

Applying this partial prior information to Boltzmann's maximum entropy theorem, we have the following lemma.

Lemma 2. If $\alpha = \bar{\Phi}_2((a, b); \mu, \Sigma) / (\Phi(v^*(b)) - \Phi(v^*(a)))$, the maximum entropy prior distribution of θ , reflecting the uncertain interval constraint in (15) is

$$\theta \sim WN_{(a,b)}(\mu, \Sigma), \quad (17)$$

where $v^*(a) = (a - \mu)/(\sigma\sqrt{\delta})$, $v^*(b) = (b - \mu)/(\sigma\sqrt{\delta})$, $\mathbf{a} = (a, a)^\top$, $\mathbf{b} = (b, b)^\top$, $\Phi_2((\mathbf{a}, \mathbf{b}); \boldsymbol{\mu}, \Sigma)$ denote a rectangle probability $p(a \leq X_2 \leq b, a \leq X_1 \leq b)$ of X_1 and X_2 whose joint distribution is a bivariate normal $N_2(\boldsymbol{\mu}, \Sigma)$,

$$\boldsymbol{\mu} = \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \quad \Sigma = \sigma^2 \begin{pmatrix} \delta & \delta \\ \delta & 1 \end{pmatrix}, \quad 0 < \delta < 1. \quad (18)$$

Proof. Taking $t_1(\theta) = \theta$, $t_1 = \mu$, $t_2(\theta) = (\theta - \mu)^2$, $t_2 = \sigma^2$, $t_3(\theta) = I(a \leq \theta \leq b)$, and $t_3 = \alpha$, and setting $\lambda_1 = 0$ and $\lambda_2 = -1/2\sigma^2$, the maximum entropy prior $\pi(\theta)$ in (9) reduces to

$$\pi_3(\theta) \propto \exp \left\{ \frac{(\theta - \mu)^2}{2\sigma^2} \right\} \exp \{ \lambda_3 t_3(\theta) \}, \quad \theta \in \mathbb{R}, \quad (19)$$

by Boltzmann's maximum entropy theorem. Now the second exponential term in the right hand side of (19) can be determined by using the condition (16) as follows. Among all the possible proper prior densities of the form (19), the choice of

$$\begin{aligned} & \lambda_3 t_3(\theta) \\ &= \ln \left\{ \frac{\Phi(\xi^* v^*(b) - \eta^*(\theta - \mu)) - \Phi(\xi^* v^*(a) - \eta^*(\theta - \mu))}{\Phi(v^*(b)) - \Phi(v^*(a))} \right\} \end{aligned} \quad (20)$$

yields the proper prior density (22). Further, this choice leads to the only proper prior that satisfies

$$\begin{aligned} \alpha &= p(a \leq X_2 \leq b \mid a \leq X_1 \leq b) \\ &= \frac{\Phi_2((\mathbf{a}, \mathbf{b}); \boldsymbol{\mu}, \Sigma)}{(\Phi(v^*(b)) - \Phi(v^*(a)))} \end{aligned} \quad (21)$$

in the condition (16), because (22) is the density of $\theta \stackrel{d}{=} [X_2 \mid a \leq X_1 \leq b] \sim \text{WN}_{(a,b)}(\boldsymbol{\mu}, \Sigma)$. $\square$

Thus the maximum entropy prior density for the Case 3 is given by

$$\begin{aligned} \pi_3(\theta) &= \phi(\theta; \mu, \sigma^2) \\ &\times \frac{\Phi(\xi^* v^*(b) - \eta^*(\theta - \mu)) - \Phi(\xi^* v^*(a) - \eta^*(\theta - \mu))}{\Phi(v^*(b)) - \Phi(v^*(a))}, \\ &\theta \in \mathbb{R}, \end{aligned} \quad (22)$$

where $\xi^* = (1 - \delta)^{-1/2}$ and $\eta^* = \xi^* \sqrt{\delta}/\sigma$.

Note, from Lemma 2, that $\pi_1(\theta) = \pi_2(\theta) = \pi_3(\theta)$ for $a = -\infty$ and $b = \infty$. This is consistent with our partial priori information that $E[\theta] = \mu$ and $\text{Var}(\theta) = \sigma^2$ for $\theta \in \mathbb{R}$. Some algebra using the moments of the $\text{WN}_{(a,b)}(\boldsymbol{\mu}, \Sigma)$ distribution

given in Kim [6] yields the entropy of $\pi_3(\theta)$ (i.e., $\text{Ent}(\pi_3(\theta))$) given by

$$\begin{aligned} & \frac{1}{2} \left\{ 1 + \ln(2\pi\sigma^2) + \frac{v^*(a)\phi(v^*(a)) - v^*(b)\phi(v^*(b))}{\Phi(v^*(b)) - \Phi(v^*(a))} \delta \right\} \\ & + \ln(\Phi(v^*(b)) - \Phi(v^*(a))) \\ & - \int_{\mathbb{R}} \ln \{ \Phi(\xi^* v^*(b) - \eta^*(\theta - \mu)) \\ & - \Phi(\xi^* v^*(a) - \eta^*(\theta - \mu)) \} \pi_3(\theta) d\theta. \end{aligned} \quad (23)$$

As seen in (23), the calculation of $\text{Ent}(\pi_3(\theta))$ involves a complicated integration. Instead, by using a Monte Carlo integration, we may calculate it approximately. According to Kim [6], it follows that the stochastic representation of the prior distribution $\theta \sim \text{WN}_{(a,b)}(\boldsymbol{\mu}, \Sigma)$ with density $\pi_3(\theta)$ is

$$\theta \stackrel{d}{=} \mu + \sigma\sqrt{\delta} Z_{(v^*(a), v^*(b))} + \sigma(1 - \delta)^{1/2} Z, \quad (24)$$

where $Z_{(v^*(a), v^*(b))} \sim N_{(v^*(a), v^*(b))}(0, 1)$ and $Z \sim N(0, 1)$, and they are independent. This stochastic representation is useful for generating θ 's from the prior distribution θ , and hence implementing the Monte Carlo integration. The following corollary asserts the relationship among $\pi_j(\theta)$, $j = 1, 2, 3$, in terms of $\delta \in (0, 1)$.

Corollary 3. As $\delta \rightarrow 1$, the maximum entropy prior $\pi_3(\theta)$ approximates to $\pi_2(\theta)$ in (13), while $\pi_3(\theta)$ is equivalent to $\pi_1(\theta)$ for $\delta = 0$.

Proof. Note from (22) that $\xi v^*(b) - \eta(\theta - \mu) \rightarrow \infty$ and $\xi v^*(a) - \eta(\theta - \mu) \rightarrow -\infty$ as $\delta \rightarrow 1$ for $\theta \in [a, b]$. Also note, from the conditional property of $N_2(\boldsymbol{\mu}, \Sigma)$, that $\alpha = \Phi_2((\mathbf{a}, \mathbf{b}); \boldsymbol{\mu}, \Sigma)/(\Phi(v^*(b)) - \Phi(v^*(a))) \rightarrow 1$ is equivalent to $\delta \rightarrow 1$ for finite values of μ, σ, a , and b with $b > a$. Applying these two results to $\pi_3(\theta)$, we see that it approximates to $\pi_2(\theta)$ as $\delta \rightarrow 1$. It is straightforward to see from Lemma 2 that $\pi_3(\theta) = \pi_1(\theta)$ for $\delta = 0$, because the $\text{WN}_{(a,b)}(\boldsymbol{\mu}, \Sigma)$ distribution is equivalent to $N(\mu, \sigma^2)$ for $\delta = 0$. $\square$

Each graph of Figure 1 depicts the difference between $\text{Ent}(\pi_3(\theta))$ and $\text{Ent}(\pi_2(\theta))$ as a function of $\delta \in (0, 1)$ for three values of σ^2 , two cases of $\Theta = [a, b]$, and $\mu = 0$. In Figure 1, the difference is denoted by Diff_{Ent} . Since each graph coincides with the results of Corollary 3, we can obtain the following implications from the figure. (i) As expected, we see that $\text{Ent}(\pi_1(\theta)) > \text{Ent}(\pi_3(\theta)) > \text{Ent}(\pi_2(\theta))$ for $\delta \in (0, 1)$. (ii) The entropy of $\pi_3(\theta)$ is a monotone decreasing function of δ . (iii) Each entropy of the three priors increases as σ^2 becomes large. (iv) Diff_{Ent} is closely related with degree of uncertainty (i.e., $(1 - \alpha) \times 100\%$) for it is a monotone decreasing function of $\delta \in (0, 1)$ and α is a function of δ . (v) Diff_{Ent} is a monotone increasing function of σ^2 for the case where a value of δ is given.

3.3. Objective Measure of Uncertainty. From a relationship between Lemmas 1 and 2, we can propose an objective

FIGURE 1: Plots of $\text{Diff}_{\text{Ent}} = \text{Ent}(\pi_3(\theta)) - \text{Ent}(\pi_2(\theta))$ for different values of $\delta \in (0, 1)$ and σ^2 for the cases where (a) $\Theta = [-1.0, 0]$ and (b) $\Theta = [-1.0, 1.5]$.

measure for the degree of uncertainty regarding a prior interval constraint accounted for by HSGM. The following theorem proposes the objective measure using the same notations as used in Lemmas 1 and 2.

Theorem 4. Let $\sigma_0^2 = (1 - \delta)\sigma^2$, $\sigma_1^2 = \delta\sigma^2$, and $\delta \in (0, 1)$. Then the two-stage prior of θ defined by HSGM in (2) is equivalent to $\pi_3(\theta)$ and the degree of uncertainty regarding the interval constraint, $\theta \in [a, b]$, accounted for by the HSGM is

$$\left\{ 1 - \frac{\bar{\Phi}_2((\mathbf{a}, \mathbf{b}); \boldsymbol{\mu}, \Sigma)}{(\Phi(v^*(b)) - \Phi(v^*(a)))} \right\} \times 100\%. \quad (25)$$

Proof. If $\sigma_0^2 = (1 - \delta)\sigma^2$, $\sigma_1^2 = \delta\sigma^2$, and $\delta \in (0, 1)$, the marginal prior distribution of θ in Lemma 1 is equivalent to the density of $\text{WN}_{(a,b)}(\boldsymbol{\mu}, \Sigma)$. Thus the prior density $p(\theta)$ in (4) is equal to the maximum entropy prior $\pi_3(\theta)$ in (22). Now Lemma 2 asserts that $\pi_3(\theta)$ reflects uncertainty about the interval constraint $\theta \in [a, b]$ by the degree of $(1 - \alpha) \times 100\%$ with $\alpha = \bar{\Phi}_2((\mathbf{a}, \mathbf{b}); \boldsymbol{\mu}, \Sigma) / (\Phi(v^*(b)) - \Phi(v^*(a)))$. $\square$

Theorem 4 provides an exact measure of the uncertainty about the priori interval constraint on θ accounted for by HSGM, and it shows that the uncertainty measure is different from that of O'Hagan and Leonard [1] which mainly depends on the first stage variance, σ_0^2 , of θ in HSGM. Theorem 4 also indicates that HSGM in (2) can be used to elicit the priori uncertain interval information associated with $\pi_3(\theta)$. Further, the entropy of the two-stage prior θ defined by the HSGM (i.e., $p(\theta)$ in (4)) can be calculated by using the

formula of $\text{Ent}(\pi_3(\theta))$ in (23). We can visualize the degree of uncertainty about the priori interval constraint, $\theta \in [a, b]$, by plotting $1 - \alpha$ for different values of $\delta \in (0, 1)$ in Figure 2.

From Figure 2, we can see, for fixed value of σ^2 , that HSGM with small values of δ tends to increase the uncertainty regarding the priori constraint. This coincides with the result which is asserted by Corollary 3. Further, we see from the figure that, for a fixed value of $\delta \in (0, 1)$, we can enlarge (or reduce) the uncertainty about the priori constraint by increasing (or decreasing) the amount of σ^2 (or equivalently σ_0^2 and σ_1^2) in the two stages of a prior hierarchy of θ .

When the data information is $y \sim N(\theta, \tau^2)$, it is well known that the maximum entropy priors (4) and (9) are conjugate priors for the normal mean θ when τ^2 is known. This is obtained from the following posteriors:

$$\begin{aligned} p_1(\theta | y) &\propto \phi(y; \theta, \tau^2) \pi_1(\theta) \propto \phi(\theta; \mu^*, \sigma^{*2}) \\ &\text{for } \theta \in \mathbb{R}, \\ p_2(\theta | y) &\propto \phi(y; \theta, \tau^2) \pi_2(\theta) \propto \phi(\theta; \mu^*, \sigma^{*2}) \\ &\text{for } \theta \in [a, b], \end{aligned} \quad (26)$$

where $\mu^* = (\sigma^2 y + \tau^2 \mu) / (\sigma^2 + \tau^2)$ and $\sigma^{*2} = \sigma^2 \tau^2 / (\sigma^2 + \tau^2)$. Thus, we see that each prior satisfies the conjugate property that $\pi_j(\theta)$ and $p_j(\theta | y)$ belong to the same family of distributions for $j = 1, 2$. The following corollary provides this conjugate property which also applies to $\pi_3(\theta)$ in (22).

FIGURE 2: Plots of $(1 - \alpha)$ for different values of $\delta \in (0, 1)$ and σ^2 for the cases where (a) $\Theta = [-1.0, 0]$ and (b) $\Theta = [-1.0, 1.5]$.

Corollary 5. Let $y \sim N(\theta, \tau^2)$ with known τ^2 . Then the prior $\pi_3(\theta)$ yields the posterior distribution of θ given by

$$(\theta | y) \sim WN_{(a,b)}(\mu^*, \Sigma^*), \quad (27)$$

where

$$\begin{aligned} \mu^* &= \begin{pmatrix} (1 - \delta)\mu + \delta\mu^* \\ \mu^* \end{pmatrix}, \\ \Sigma^* &= \begin{pmatrix} \delta(1 - \delta)\sigma^2 + \delta^2\sigma^{*2} & \delta\sigma^{*2} \\ \delta\sigma^{*2} & \sigma^{*2} \end{pmatrix}, \end{aligned} \quad (28)$$

$\mu^* = (\sigma^2 y + \tau^2 \mu) / (\sigma^2 + \tau^2)$, and $\sigma^{*2} = \sigma^2 \tau^2 / (\sigma^2 + \tau^2)$.

Proof. Since $y \sim N(\theta, \tau^2)$ with known τ^2 , under the prior $\pi_3(\theta)$ in (22), the posterior density of θ is given by

$$\begin{aligned} p(\theta | y) &\propto \phi(y; \theta, \tau^2) \phi(\theta; \mu, \sigma^2) \\ &\quad \times \{\Phi(\xi^* v^*(b) - \eta^*(\theta - \mu)) \\ &\quad - \Phi(\xi^* v^*(a) - \eta^*(\theta - \mu))\} \\ &\propto \phi(\theta; \mu^*, \sigma^{*2}) \{\Phi(\xi^* \gamma(b) - \eta^*(\theta - \mu^*)) \\ &\quad - \Phi(\xi^* \gamma(a) - \eta^*(\theta - \mu^*))\} \end{aligned} \quad (29)$$

in that $\{\Phi(\xi^* v^*(b) - \eta^*(\theta - \mu)) - \Phi(\xi^* v^*(a) - \eta^*(\theta - \mu))\} = \{\Phi(\xi^* \gamma(b) - \eta^*(\theta - \mu^*)) - \Phi(\xi^* \gamma(a) - \eta^*(\theta - \mu^*))\}$, where $\gamma(b) = \{b - ((1 - \delta)\mu + \delta\mu^*) / (\sigma\sqrt{\delta})\}$ and $\gamma(a) = \{a - ((1 - \delta)\mu + \delta\mu^*) / (\sigma\sqrt{\delta})\}$. This is a kernel of $WN_{(a,b)}(\mu^*, \Sigma^*)$ density. $\square$

4. Posterior Estimation

Let us revisit the HSGM with the following two stages of a prior hierarchy of θ :

$$y_i = \theta + \epsilon_i, \quad \epsilon_i \sim N(0, \tau^2), \quad i = 1, \dots, n,$$

$$\tau^2 \sim \text{IG}(c, d),$$

$$\theta | \theta_0 \sim N(\theta_0, (1 - \delta)\sigma^2), \quad \text{independent of } (\epsilon_1, \dots, \epsilon_n)^\top,$$

$$\theta_0 \sim N_{(a \leq \theta_0 \leq b)}(\mu, \delta\sigma^2), \quad \delta \in (0, 1).$$

(30)

According to Theorem 4, we see that the two-stage prior of θ is essentially the same as the maximum entropy prior which properly elicits the partial prior information of an uncertain interval constraint, $\{\theta; \theta \in [a, b]\}$, with $(1 - \alpha) \times 100\%$ degree of uncertainty, where $\alpha = \Phi_2((a, b); \mu, \Sigma) / (\Phi(v^*(b)) - \Phi(v^*(a)))$. Here μ and σ^2 are true prior mean and variance of the parameter θ when the uncertain priori interval condition does not exist.

Based on the marginal prior distribution of θ in Lemma 1, we have the joint posterior distribution of θ and τ^2 proportional to the product of likelihood and the prior distribution,

$$p(\theta, \tau^2 | D_n) \propto \prod_{i=1}^n \phi(y_i; \theta, \tau^2) \pi_3(\theta) \text{IG}(\tau^2; c, d), \quad (31)$$

where $D_n = \{y_1, \dots, y_n\}$, $\text{IG}(\cdot; c, d)$ is the inverse-gamma density with parameters c and d , and $\pi_3(\theta)$ is the density (22), that is, the density of $\theta \sim WN_{(a,b)}(\mu, \Sigma)$. Note that the joint posterior is not simplified in an analytic form of the known

density and thus intractable for posterior inference. Instead, we derive each of the conditional posterior distributions of θ and τ^2 in an explicit form, which will be useful for posterior inference such as Gibbs sampling (e.g., [14]).

Corollary 6. *Given the joint posterior distribution (31), we have the following.*

- (i) *The full conditional posterior distribution of θ is given by*

$$\theta \mid (\tau^2, D_n) \sim WN_{(a,b)}(\theta^*, \Omega^*), \quad (32)$$

where

$$\begin{aligned} \theta^* &= \left(\frac{(1-\delta)\mu + \delta\theta^*}{\theta^*} \right), \\ \Omega^* &= \left(\frac{\delta(1-\delta)\sigma^2 + \delta^2\omega^{*2}}{\delta\omega^{*2}}, \frac{\delta\omega^{*2}}{\omega^{*2}} \right), \end{aligned} \quad (33)$$

$\theta^* = (\bar{y}\sigma^2 + \mu\tau^2/n)/(\sigma^2 + \tau^2/n)$, and $\omega^{*2} = \sigma^2\tau^2/(n\sigma^2 + \tau^2)$.

- (ii) *The full conditional posterior distribution of τ^2 is the inverse-Gamma distribution*

$$\tau^2 \mid (\theta, D_n) \sim IG\left(c + \frac{2}{n}, d + \sum_{i=1}^n (y_i - \theta)^2\right). \quad (34)$$

Proof. (i) The unnormalized conditional density of θ given that τ^2 and D_n is proportional to

$$p(\theta \mid \tau^2, D_n) \propto \phi\left(\bar{y}; \theta, \frac{\tau^2}{n}\right) \pi_3(\theta), \quad (35)$$

where $\bar{y} = \sum_{i=1}^n y_i/n$. Thus direct application of Corollary 5 yields the result.

(ii) It is straightforward to see from (31) that the full conditional posterior distribution of τ^2 is

$$\begin{aligned} f(\tau^2 \mid \theta, D_n) \\ \propto (\tau^2)^{-(c+n/2+1)} \exp\left\{-\frac{(d + \sum_{i=1}^n (y_i - \theta)^2)}{\tau^2}\right\}, \end{aligned} \quad (36)$$

which is a kernel of $IG(c+2/n, d + \sum_{i=1}^n (y_i - \theta)^2)$ distribution. $\square$

It is not difficult to construct the Gibbs sampling scheme working with (θ, τ^2) because their full conditional distributions are given in Corollary 6. A routine Gibbs sampler would work to generate posterior samples of (θ, τ^2) . The only difficulty would lie in generating random samples from WN distribution, that is, $\theta \mid (\tau^2, D_n) \sim WN_{(a,b)}(\mu^*, \Omega^*)$ as given in (32). In order to generate random samples from the WN distribution, we can make use of the stochastic representation given in (24). Also we can note that the stochastic representation (24) of the full conditional posterior

distribution (32) provides the posterior mean and variance given by

$$\begin{aligned} E[\theta \mid (\tau^2, D_n)] &= \theta^* - \beta_1 \kappa, \\ \text{Var}(\theta \mid (\tau^2, D_n)) &= \omega^{*2} - \beta_2 \kappa^2, \end{aligned} \quad (37)$$

where $\kappa = \delta\omega^{*2}/\omega_{11}^* - \beta_1$, and $1 - \beta_2$ are respective mean and variance of the truncated standard normal distribution, $N_{(\gamma^*(a), \gamma^*(b))}(0, 1)$ with $\gamma^*(a) = (a - \theta_1^*)/\omega_{11}^*$, $\gamma^*(b) = (b - \theta_1^*)/\omega_{11}^*$, $\theta_1^* = (1 - \delta)\mu + \delta\theta^*$, and $\omega_{11}^{*2} = \delta(1 - \delta)\sigma^2 + \delta^2\omega^{*2}$. Johnson et al. [13] shows that they are

$$\begin{aligned} \beta_1 &= \frac{\phi(\gamma^*(b)) - \phi(\gamma^*(a))}{\Phi(\gamma^*(b)) - \Phi(\gamma^*(a))}, \\ \beta_2 &= \frac{\gamma^*(b)\phi(\gamma^*(b)) - \gamma^*(a)\phi(\gamma^*(a))}{\Phi(\gamma^*(b)) - \Phi(\gamma^*(a))} + \beta_1^2, \end{aligned} \quad (38)$$

where $\phi(\cdot)$ denotes the density of $N(0, 1)$. It is seen that each of the first terms in (37), that is, θ^* and ω^{*2} , is the posterior mean and variance of θ when the prior $\pi_1(\theta)$ is put on θ (e.g. [15, 16]), instead of the two-stage prior of Theorem 4, that is, $\pi_3(\theta)$. That is each of the second terms in the posterior mean and covariance vanishes when θ is assumed to be $N(\mu, \sigma^2)$ a priori without any interval constraint. In this sense, the second terms in (37) can be interpreted as a *constraining effect* obtained by adopting the HSGM. See, for example, Kim [3], Kim and Choi [4], and H. j. Kim and H. M. Kim [17] for practical applications of HSGM methodology to get the *constraining effect* in various constrained parameter problems.

5. Conclusion

This paper considered the normal models which include the two-stage prior of the normal mean, referred to as *hierarchical screened Gaussian normal model* (HSGM). The HSGM is based on the two stages of a prior hierarchy advocated by O'Hagan and Leonard [1] and elicits partial priori information obtained for the case where an interval constraint of the normal mean needs to be incorporated in the modeling but such a restriction is uncertain. Then we proposed an objective method to measure (or control) the uncertainty accounted for by HSGM. For this purpose, we derived the maximum entropy prior, reflecting the uncertainty about an interval constraint on the normal mean, by using Boltzmann's maximum entropy theorem. As a result, we found that both the maximum entropy prior and the two-stage prior in HSGM belong to the family of weighted normal distributions considered by Kim [6]. By exploring the distributional relationship between the two priors, we proposed the objective measure of uncertainty. This paper also proposed the Bayesian estimation procedure of HSGM and investigated some properties of the procedure by deriving the full conditional posterior distributions of unknown parameters under HSGM in analytic forms.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning (2013R1A2A2A01004790).

References

- [1] A. O'Hagan and T. Leonard, "Bayes estimation subject to uncertainty about parameter constraints," *Biometrika*, vol. 63, no. 1, pp. 201–203, 1976.
- [2] B. Liseo and N. Loperfido, "A Bayesian interpretation of the multivariate skew-normal distribution," *Statistics & Probability Letters*, vol. 61, no. 4, pp. 395–401, 2003.
- [3] H. J. Kim, "On a class of multivariate normal selection priors and its applications in Bayesian inference," *Journal of the Korean Statistical Society*, vol. 40, no. 1, pp. 63–73, 2011.
- [4] H. J. Kim and T. Choi, "On Bayesian estimation of regression models subject to uncertainty about functional constraints," *Journal of the Korean Statistical Society*, vol. 43, no. 1, pp. 133–147, 2014.
- [5] R. B. Arellano-Valle, M. D. Branco, and M. G. Genton, "A unified view on skewed distributions arising from selections," *The Canadian Journal of Statistics*, vol. 34, no. 4, pp. 581–601, 2006.
- [6] H. J. Kim, "A class of weighted multivariate normal distributions and its properties," *Journal of Multivariate Analysis*, vol. 99, no. 8, pp. 1758–1771, 2008.
- [7] E. T. Jaynes, "Prior probabilities," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 3, pp. 227–241, 1968.
- [8] R. D. Rosenkrantz, Ed., *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Reidel, Dordrecht, The Netherlands, 1983.
- [9] C. Cercignani, *The Boltzman Equation and Its Applications*, Springer, Berlin, Germany, 1988.
- [10] R. D. Rosenkrantz, *E. T. Jaynes: Papers on Probability, Statistics, and Statistical Physics*, Kluwer Academic, Norwell, Mass, USA, 1989.
- [11] R. D. Rosenkrantz, *Inference, Method, and Decision: Towards a Bayesian Philosophy and Science*, Reidel, Boston, Mass, USA, 1977.
- [12] T. Leonard and J. S. J. Hsu, *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK, 1999.
- [13] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1, John Wiley & Sons, New York, NY, USA, 1994.
- [14] A. E. Gelfand, A. F. M. Smith, and T. Lee, "Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling," *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 523–532, 1992.
- [15] C. E. Rasmussen and C. K. I. Williams, *Gaussian Process for Machine Learning*, The MIT Press, Boston, Mass, USA, 2006.
- [16] J. Q. Shi and T. Choi, *Gaussian Process Regression Analysis for Functional Data*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, UK, 2011.
- [17] H. J. Kim and H.-M. Kim, "A class of rectangle-screened multivariate normal distributions and its applications," *Statistics: A Journal of Theoretical and Applied Statistics*, 2014.