Original Research Article

Variational Bayesian Method: Ritz Method in Stochastic system

.

ABSTRACT

Bayesian inference is to find posterior probabilities, but since it is difficult to find analytical solutions, it is often the case that approximate solutions are found. The variational Bayesian method is a powerful method for finding an approximate solution. It is a variational method in a stochastic system. Variational methods have been developed for the deterministic system since ancient times, and are one of the most powerful foundations for numerical solutions of a wide range of problems defined by partial differential equations. A detailed comparison and explanation of the classical variational principle and the variational Bayesian method are given, and the basic application examples of the variational Bayesian method are also given. Programming codes written in C are also shown to aid the readers' understanding.

Keywords: Variational Method, Bayesian Inference, Posterior Probability, Stochastic System, Ritz Method

1. INTRODUCTION

Since the advent of deep learning, the neural network has brought a big innovation in the world. However, deep learning might be far from perfect, because of "the inference is a black box", "unexpected answer due to the overfitting", and "large scale of the network and long time learning". The earliest answer to them should be given. Among them, the black box nature would be a fundamental problem.

Bayesian inference performs inference similar to neural networks. Bayesian inference, like neural networks, learns and infers based on data, but it is more deductive and less data-dependent than neural networks. The difference appears as a difference in the number of unknown variables. Neural networks have a much larger number of unknown variables. Although it can be said that the flexibility is high, overlearning is likely to occur and the basis of reasoning becomes incomprehensible.

On the other hand, the learning of Bayesian inference is to find the likelihood function from the frequency distribution, and to find the posterior probability from the likelihood function and the prior distribution. Therefore, it can be said that the learning and reasoning of Bayesian inference are highly deductive. However, it is powerless when the relationship between cause and effect cannot be clearly mathematically modeled, such as when discriminating figures. Humans can easily discriminate figures and understand languages. However, it is extremely difficult to clearly mathematically model the process of judgment and understanding for these. Human judgment and understanding in this regard is a black box.

Neural networks make learning decisions while remaining in the black box. In other words, neural networks are extremely flexible because they can be learned simply by giving input data and teacher data. In comparison, Bayesian inference requires some modeling, so it must be said that it is less flexible. From another point of view, it can be said that the mechanism of neural networks is closer to human learning and reasoning than to Bayesian reasoning. The price is that human learning, like neural network learning, takes a lot of time.

Bayesian inference is to find posterior probabilities, but since it is difficult to find analytical solutions, it is often the case that approximate solutions are found. The variational Bayesian method is a powerful method for finding an approximate solution. The variational Bayesian method is a variational method in a stocastic system. Variational methods have been developed for the deterministic system since old times, and are one of the most powerful foundations for numerical solutions of a wide range of problems defined by partial differential equations. It is none other than the basis of the finite element method, which is one of the most influential numerical solutions that have made great progress in recent years.

In this paper, we compare the variational principle of the deterministic case and the stocastistic case, and describe in detail the variational Bayesian method of the stocastistic case, which is comparable to the Ritz method of the deterministic case.

Recently, the variational Bayesian method has been attracting attention in the field of information processing of the cerebrum, which is called the principle of minimum free energy, strictly speaking, the principle of minimum variational free energy [1-4]. Information processing in the brain is considered to be information processing that infers recognition from sensory information that is a stimulus from the outside world by sensory organs. Then, the idea is to regard it as a mathematical model for finding posterior probabilities by Bayesian inference. Friston introduces the variational Bayesian method into this reasoning process. Friston pointed out that the variational Bayesian method corresponds to Helmholz's principle of minimum free energy in thermodynamics, and that improving the accuracy of the sensory organs greatly contributes to improving the accuracy of inference. He calls it active reasoning.

2. VARIATIONAL PRINCIPLE IN DETERMINISTIC SYSTEM

The variational principle has long been used for deterministic systems. First, the variational principle of a deterministic system will be explained by taking the Dirichlet problem, which is a boundary value problem of partial differential equations well known in fluid mechanics, and the Dirichlet principle, which is an equivalent variational principle, as an example.

The Dirichlet problem is defined as follows:

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\phi=0 \quad \text{in } \Omega, \quad \phi=g \quad \text{on } \Gamma. \tag{1}$$

The Dirichlet principle is given by

$$F[\phi]=\iiint_\Omega\left[\left(\frac{\partial\phi}{\partial x}\right)^2+\left(\frac{\partial\phi}{\partial y}\right)^2+\left(\frac{\partial\phi}{\partial z}\right)^2\right]dV=\min, \quad \text{under} \quad \phi=g \quad \text{on } \Gamma. \tag{2}$$

Dirichlet's principle is obtained as follows:

$$0=\iiint_{\Omega}\left(\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+\frac{\partial^2\phi}{\partial z^2}\right)\delta\phi\, dV=-\iiint_{\Omega}\left(\frac{\partial\phi}{\partial x}\frac{\partial\delta\phi}{\partial x}+\frac{\partial\phi}{\partial y}\frac{\partial\delta\phi}{\partial y}+\frac{\partial\phi}{\partial z}\frac{\partial\delta\phi}{\partial z}\right)\delta\phi\, dV \quad (3)$$
$$\text{under}\quad \delta\phi=0 \quad \text{on}\ \Gamma, \quad \text{for}\ \forall\delta\phi.$$

This is proved by

$$\begin{aligned}
0&=\iiint_{\Omega}\left(\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+\frac{\partial^2\phi}{\partial z^2}\right)\delta\phi\, dV\\
&=\iiint_{\Omega}\left[\frac{\partial}{\partial x}\left(\frac{\partial\phi}{\partial x}\delta\phi\right)+\frac{\partial}{\partial y}\left(\frac{\partial\phi}{\partial y}\delta\phi\right)+\frac{\partial}{\partial z}\left(\frac{\partial\phi}{\partial z}\delta\phi\right)-\frac{\partial\phi}{\partial x}\frac{\partial\delta\phi}{\partial x}-\frac{\partial\phi}{\partial y}\frac{\partial\delta\phi}{\partial y}-\frac{\partial\phi}{\partial z}\frac{\partial\delta\phi}{\partial z}\right]dV\\
&=\iint_{S}\left(\frac{\partial\phi}{\partial x}n_x+\frac{\partial\phi}{\partial y}n_x+\frac{\partial\phi}{\partial z}n_z\right)\delta\phi\, dS-\frac{1}{2}\delta\iiint_{\Omega}\left[\left(\frac{\partial\phi}{\partial x}\right)^2+\left(\frac{\partial\phi}{\partial y}\right)^2+\left(\frac{\partial\phi}{\partial z}\right)^2\right]dV\\
&=-\frac{1}{2}\delta\iiint_{\Omega}\left[\left(\frac{\partial\phi}{\partial x}\right)^2+\left(\frac{\partial\phi}{\partial y}\right)^2+\left(\frac{\partial\phi}{\partial z}\right)^2\right]dV.
\end{aligned} \quad (4)$$

The key to moving from a deterministic system to a stochastic system is what to think instead of Eq. (3).

We describe the Ritz method, which is an important idea for applying the classical variational principle to numerical calculations. We approximate ϕ in Eq. (2) is as follows:

$$\phi(x,y,z)=\Phi_0(x,y,z)+\sum_{i=1}^{N}A_i\Phi_i(x,y,z)\,, \quad (5a)$$

where

$$\Phi_0(x,y,z)=g(x,y,z) \quad on\ \Gamma, \quad \Phi_i(x,y,z)=0,\ i=1,2,\cdots,N \quad on\ \Gamma\,. \quad (5b)$$

Substituting Eq. (5a) into F defined in Eq. (2) yields:

From this, we obtain a system of linear algebraic equations for the following unknown coefficients $A_i \quad i=1,2,\cdots,N$:

$$\begin{aligned}
0=\frac{\partial F}{\partial A_j}&=2\iiint_{\Omega}\left[\left(\frac{\partial\Phi_0}{\partial x}+\sum_{i=1}^{N}A_i\frac{\partial\Phi_i}{\partial x}\right)\frac{\partial\Phi_j}{\partial x}+\left(\frac{\partial\Phi_0}{\partial y}+\sum_{i=1}^{N}A_i\frac{\partial\Phi_i}{\partial y}\right)\frac{\partial\Phi_j}{\partial y}+\left(\frac{\partial\Phi_0}{\partial z}+\sum_{i=1}^{N}A_i\frac{\partial\Phi_i}{\partial z}\right)\frac{\partial\Phi_j}{\partial z}\right]dV\\
&=2\iiint_{\Omega}\left[\left(\frac{\partial\Phi_0}{\partial x}\frac{\partial\Phi_j}{\partial x}+\frac{\partial\Phi_0}{\partial y}\frac{\partial\Phi_j}{\partial y}+\frac{\partial\Phi_0}{\partial z}\frac{\partial\Phi_j}{\partial z}\right)\right]dV+2\sum_{i=1}^{N}\iiint_{\Omega}\left(\frac{\partial\Phi_i}{\partial x}\frac{\partial\Phi_j}{\partial x}+\frac{\partial\Phi_i}{\partial y}\frac{\partial\Phi_j}{\partial y}+\frac{\partial\Phi_i}{\partial z}\frac{\partial\Phi_j}{\partial z}\right)dV A_i, \quad (7)\\
&\qquad \text{for}\quad j=12,\cdots,N.
\end{aligned}$$

Swapping the subscripts i and j yields the following equation:

$$\sum_{j=1}^{N}\iiint_{\Omega}\left(\frac{\partial\Phi_i}{\partial x}\frac{\partial\Phi_j}{\partial x}+\frac{\partial\Phi_i}{\partial y}\frac{\partial\Phi_j}{\partial y}+\frac{\partial\Phi_i}{\partial z}\frac{\partial\Phi_j}{\partial z}\right)dV A_j=-\iiint_{\Omega}\left(\frac{\partial\Phi_0}{\partial x}\frac{\partial\Phi_i}{\partial x}+\frac{\partial\Phi_0}{\partial y}\frac{\partial\Phi_i}{\partial y}+\frac{\partial\Phi_0}{\partial z}\frac{\partial\Phi_i}{\partial z}\right)dV, \quad (8)$$
$$\text{for}\quad j=12,\cdots,N.$$

By solving this, an approximate solution of ϕ can be obtained.

Next, we describe the least squares method that spans both deterministic and stocastistic systems. Suppose the observation equation is given below

$$u_i = \mu + \varepsilon_i, \quad i = 1, 2, \cdots, N, \tag{9a}$$

where

$$p(\varepsilon_i) = \frac{1}{\sqrt{2\pi\Sigma}} \exp\left(-\frac{\varepsilon_i^{\ 2}}{2\Sigma} \right). \tag{9b}$$

The least squares method yields

$$S = \frac{1}{2\Sigma} \sum_{i=1}^{N} (u_i - \mu)^2 = \min \to 0 = \frac{\partial S}{\partial \mu} = -\frac{1}{\Sigma} \sum_{i=1}^{N} (u_i - \mu) \to \mu = \frac{1}{N} \sum_{i=1}^{N} u_i \ . \tag{10}$$

Since the exponential function is a monotonically increasing function, the magnitude relationship does not change even if the exponential of S is taken. If you take a minus, the minimum becomes the maximum:

In summary, the following relationship can be obtained:

$$\begin{aligned} S = \frac{1}{2\Sigma} \sum_{i=1}^{N} (u_i - \mu)^2 = \min \leftrightarrow F' = \prod_{i=1}^{N} \exp\left(-\frac{1}{2\Sigma} (u_i - \mu)^2 \right) = \max \\ \leftrightarrow F = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\Sigma}} \exp\left(-\frac{1}{2\Sigma} (u_i - \mu)^2 \right) = \prod_{i=1}^{N} p(u_i \mid \mu) = \max. \end{aligned} \tag{12}$$

This is nothing but the maximum likelihood method.

3. VARIATIONAL PRINCIPLE IN STOCHASTIC SYSTEM: VARIATIONAL BAYESIAN PRINCIPLE

3.1 Derivation of variational Bayesian principle

The unknown variable in the stochastic case is the probability distribution, and whether the two probability distributions $q_1(\vartheta)$ and $q_2(\vartheta)$ are equal or not is measured by Kullback-Leibler divergence (KLD) $KL(q_1(\vartheta) \| q_2(\vartheta))$ of $q_1(\vartheta)$ and $q_2(\vartheta)$:

$$KL(q_1(\vartheta) \| q_2(\vartheta)) = \int q_1(\vartheta) \ln \frac{q_1(\vartheta)}{q_2(\vartheta)} d\vartheta \, . \tag{13}$$

KLD has the following properties:

$$KL(q_1(\vartheta) \| q_2(\vartheta)) \begin{cases} = 0 & \text{when} \quad q_1(\vartheta) = q_2(\vartheta) \\ > 0 & \text{otherwise} \end{cases}. \tag{14}$$

Appendix A shows the proof of Eq. (14). However, since it is not symmetric, it is not a distance. It is a scale that shows the difference between the probability distributions $q_1(\vartheta)$ and $q_2(\vartheta)$.

The equation corresponding to Eq. (3) in the case of the deterministic is:

$$KL(q(\vartheta)\,\|\,p(\vartheta\,|\,\varphi)) = \int q(\vartheta)\ln\frac{q(\vartheta)}{p(\vartheta\,|\,\varphi)}d\vartheta\,. \tag{15}$$

Considering the non-negativeness of KLD, KLD is minimized when $q(\vartheta) = p(\vartheta\,|\,\varphi)$. However, even if Eq. (15) itself is used as the minimum value problem for finding $p(\vartheta\,|\,\varphi)$, a meaningful answer cannot be found. Variational Bayesian Principle (VBP) can be obtained by rewriting Eq. (15). Firstly, we have

$$\begin{aligned} KL(q(\vartheta)\,\|\,p(\vartheta\,|\,\varphi)) &= \int q(\vartheta)\log\frac{q(\vartheta)p(\varphi)}{p(\vartheta,\varphi)}d\vartheta = \int q(\vartheta)\log\frac{q(\vartheta)}{p(\vartheta,\varphi)}d\vartheta + \int q(\vartheta)\log p(\varphi)d\vartheta \\ &= \int q(\vartheta)\log\frac{q(\vartheta)}{p(\vartheta,\varphi)}d\vartheta + \log p(\varphi) = KL(q(\vartheta)\,\|\,p(\vartheta,\varphi)) + \log p(\varphi), \end{aligned} \tag{16}$$

and rewrting further, we finally derive the functional F of the the variational Bayesian principle:

$$F \equiv KL(q(\vartheta)\,\|\,p(\vartheta,\varphi)) = \int q(\vartheta)\ln\frac{q(\vartheta)}{p(\vartheta,\varphi)}d\vartheta = KL(q(\vartheta)\,\|\,p(\vartheta\,|\,\varphi)) - \ln p(\vartheta)\,. \tag{17}$$

At first glance, the unknown variable of the minimum value problem of F seems to be only ϑ, but φ is also an unknown variable. Since both of the two terms on the right-hand side of Eq. (17) are non-negative (Appendix A), q that minimizes F is $p(\vartheta\,|\,\varphi)$, but it can be seen that F also has the effect of minimizing $p(\vartheta)$.

The variational Bayesian principle (VBP) is called the minimum principle of free energy in the fields of information processing in the brain. Strictly speaking, the free energy is the variational free energy (VFE). The principle of minimum free energy, which is proposed by Friston, concerns sensation and cognition. Since the sensation and the physical movement (action) are directly connected, the action causes a change in the sensation φ, which improves the accuracy of reasoning. Improving the accuracy of reasoning through actions is called active reasoning. Active reasoning was demonstrated first by the principle of minimal free energy by Friston [1] clearly.

For the reader's understanding, the model of information processing in the brain is shown in Fig. 1 and Fig. 2. In the following, φ may be called sensation and ϑ may be called recognition.

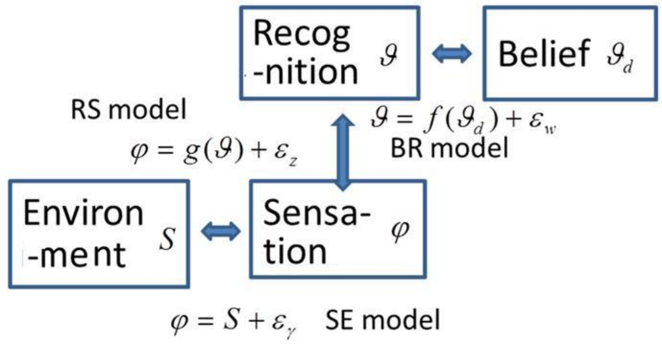

Fig. 1. Information flow in environment and brain

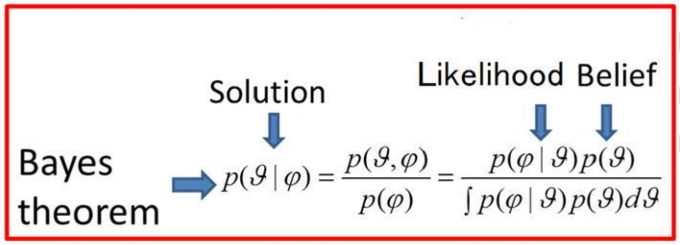

Fig. 2. Bayes theorem

4. APPROXIMATION OF VARIATIONAL FREE ENERDY

4.1 The simplest approximation (Estimation of the mean value of the posterior distribution)

Replacing Bayesian inference with the Variational Bayesian Principle (VBP), and using Gaussian distribution approximation and the steepest descent method, can lead to calculations also feasible in the brain. For simplicity, we consider the case of one variable. First, we consider a Gaussian approximation of the likelihood function and prior distribution:

$$p(\varphi \mid \vartheta) = \frac{1}{\sqrt{2\pi\Sigma_\varphi}} \exp\left(-\frac{(\varphi - g(\vartheta))^2}{2\Sigma_\varphi} \right). \quad p(\vartheta) = \frac{1}{\sqrt{2\pi\Sigma_\vartheta}} \exp\left(-\frac{(\vartheta - \bar{\mu})^2}{2\Sigma_{\bar{\vartheta}}} \right). \tag{18}$$

From this, in the VBP of Eq. (17), the joint distribution is approximated as follows:

$$p(\varphi, \vartheta) = \frac{1}{\sqrt{2\pi\Sigma_\varphi}} \frac{1}{\sqrt{2\pi\Sigma_\vartheta}} \exp\left(-\frac{(\varphi - g(\vartheta))^2}{2\Sigma_\varphi} - \frac{(\vartheta - \bar{\mu})^2}{2\Sigma_{\bar{\vartheta}}} \right). \tag{19}$$

An approximation of the posterior distribution $p(\vartheta \mid \varphi)$ can be obtained by approximating $q(\vartheta)$ under the condition of minimum VBP. Therefore, let us assume q in the following form, which means a normal distribution with a mean μ variance of 0:

$$q(\vartheta) = \delta(\vartheta - \mu), \tag{20}$$

where, $\delta(\vartheta)$ is a Dirac delta function. The unknown variable is μ. Substituting Eqs. (21) and (22) into Eq. (17) yields the following equation:

$$
\begin{aligned}
F[\mu] &= \int \delta(\vartheta-\bar{\mu})\ln\delta(\vartheta-\bar{\mu}) - \int \delta(\vartheta-\bar{\mu})\ln p(\vartheta,\varphi) \\
&= \ln\delta(0) + \frac{1}{2}\int \delta(\vartheta-\mu)\left(\frac{(\varphi-g(\vartheta))^2}{\Sigma_\varphi} + \frac{(\vartheta-\bar{\mu})^2}{\Sigma_\vartheta} + \ln 4\pi^2\sqrt{\Sigma_\varphi\Sigma_\vartheta}\right)d\vartheta \\
&\sim \frac{1}{2}\left(\frac{(\varphi-g(\mu))^2}{\Sigma_\varphi} + \frac{(\mu-\bar{\mu})^2}{\Sigma_\vartheta} + \ln 4\pi^2\sqrt{\Sigma_\varphi\Sigma_\vartheta}\right).
\end{aligned} \tag{21}
$$

The term $\ln\delta(0)$ is a constant term and can be omitted. The integral for ϑ is done in $-\infty < \vartheta < \infty$. Also, $F[\mu]$ is used to indicate that μ is an unknown variable. Therefore, the minimum VBP should be simplified and the following minimum value problem should be solved:

$$
\mu = \arg\min_{\mu}\left[\ln p(\mu,\varphi)\right] = \arg\min_{\mu}\frac{1}{2}\left(\frac{(\varphi-g(\mu))^2}{\Sigma_\varphi} + \frac{(\mu-\bar{\mu})^2}{\Sigma_\vartheta} + \log 4\pi^2\sqrt{\Sigma_\varphi\Sigma_\vartheta}\right). \tag{22}
$$

What should not be forgotten here is that the effect of the action on improving the accuracy of the sensations is not forgotten. Eq. (22) includes effect of actions:

$$
(\mu,a) = \arg\min_{(\mu,a)} F = \arg\min_{(\mu,a)}\left[\ln p(\mu,\varphi)\right] = \arg\min_{(\mu,a)}\frac{1}{2}\left(\frac{(\varphi-g(\mu))^2}{\Sigma_\varphi} + \frac{(\mu-\bar{\mu})^2}{\Sigma_\vartheta} + \ln 4\pi^2\sqrt{\Sigma_\varphi\Sigma_\vartheta}\right). \tag{23}
$$

To solve Eq. (23) by the steepest descent method, it becomes:

$$
(\mu,a)_{t+\delta t} = (\mu,a)_t + \left(\frac{\partial F}{\partial\mu}, \frac{\partial F}{\partial\varphi}\frac{d\varphi}{da}\right)\cdot k\delta t. \tag{24}
$$

Here, t is a sequential step, δt is a sequential unit, and k is a sequential parameter. Since the algorithms of Eqs. (9.25) and (9.26) are simple, it is considered that they can be implemented in vivo.

4.2 A second simplest approximation (estimation of mean value μ and variance ζ of posterior distribution ... with the assumption of a sharp peak in joint distribution)

We consider the Gaussian approximation for the likelihood function $p(\varphi\,|\,\vartheta)$ and the prior distribution $p(\vartheta)$ are given by Eq. (18), so the joint distribution remains $p(\vartheta\,|\,\varphi)$ in Eq. (19). And $q(\vartheta)$, which is an approximation of the posterior distribution $p(\vartheta\,|\,\varphi)$, is replaced by Eq. (26) with mean μ and variance ζ. Let us consider the case of approximation by the Gaussian distribution [2]. That is,

$$
p(\varphi,\vartheta) = \frac{1}{\sqrt{2\pi\Sigma_\varphi}}\frac{1}{\sqrt{2\pi\Sigma_\vartheta}}\exp\left(-\frac{(\varphi-g(\vartheta))^2}{2\Sigma_\varphi} - \frac{(\vartheta-\bar{\mu})^2}{2\Sigma_\vartheta}\right) \tag{25}
$$

and

$$q(\vartheta)=\frac{1}{\sqrt{2\pi\zeta}}\exp\left(-\frac{(\vartheta-\mu)^2}{2\zeta}\right). \tag{26}$$

This time, μ and ζ are unknown variables.

Substituting Eqs. (25) and (26) for the functional F of the VBP gives the following equation:

$$\begin{aligned}F[\mu,\zeta]&=\int q(\vartheta)\ln\frac{q(\vartheta)}{p(\vartheta,\varphi)}d\vartheta=\int q(\vartheta)\left(-\ln\sqrt{2\pi\zeta}-\frac{(\vartheta-\mu)^2}{2\zeta}\right)d\vartheta+\int q(\vartheta)E(\vartheta,\varphi)d\vartheta\\&=-\ln\sqrt{2\pi\zeta}-\frac{1}{2}+\int q(\vartheta)E(\vartheta,\varphi)d\vartheta,\end{aligned} \tag{27}$$

where

$$E(\vartheta,\varphi)=-\ln p(\vartheta,\varphi)=\frac{1}{2}\left(\ln(\Sigma_\varphi\Sigma_\vartheta)+\frac{(\varphi-g(\vartheta))^2}{\Sigma_\varphi}+\frac{(\vartheta-\bar{\mu})^2}{\Sigma_\vartheta}\right)+\ln(2\pi)\,. \tag{28}$$

$E(\vartheta,\varphi)$ is called energy. In the following discussion, the last constant on the right side is omitted.

Now, for the sake of simple results, assume that $E(\vartheta,\varphi)$ has a sharp peak at $\vartheta=\mu$:

$$E(\vartheta,\varphi)\approx E(\mu,\varphi)+\left[\frac{\partial E}{\partial\vartheta}\right]_{\vartheta=\mu}(\vartheta-\mu)+\frac{1}{2}\left[\frac{\partial^2E}{\partial\vartheta^2}\right]_{\vartheta=\mu}(\vartheta-\mu)^2. \tag{29}$$

Approximating the last term on the right-hand side of Eq. (27) using Eq. (26), we obtain

$$\int q(\vartheta)E(\vartheta,\varphi)d\vartheta\approx E(\mu,\varphi)+\frac{1}{2}\left[\frac{\partial^2E}{\partial\vartheta^2}\right]_{\vartheta=\mu}\int q(\vartheta)(\vartheta-\mu)^2d\vartheta=E(\mu,\varphi)+\frac{1}{2}\left[\frac{\partial^2E}{\partial\vartheta^2}\right]_{\vartheta=\mu}\zeta\,. \tag{30}$$

Substituting Eq. (30) into Eq. (27) yields the following equation:

$$F[\mu,\zeta]=-\ln\sqrt{2\pi\zeta}-\frac{1}{2}+E(\mu,\varphi)+\frac{1}{2}\left[\frac{\partial^2E}{\partial\vartheta^2}\right]_{\vartheta=\mu}\zeta=E(\mu,\varphi)+\frac{1}{2}\left(\left[\frac{\partial^2E}{\partial\vartheta^2}\right]_{\vartheta=\mu}\zeta-\ln(2\pi\zeta)-1\right). \tag{31}$$

To minimize F, the derivative of F by ζ should also be 0, so let $\zeta*$ be ζ that satisfies this condition, then $\zeta*$ is:

$$\left[\frac{\partial^2E}{\partial\vartheta^2}\right]_{\vartheta=\mu}-\frac{1}{\zeta*}=0\quad\rightarrow\quad\zeta*=\left(\left[\frac{\partial^2E}{\partial\vartheta^2}\right]_{\vartheta=\mu}\right)^{-1}. \tag{32}$$

Substituting Eq. (9.34) into Eq. (9.33) yields the following equation:

$$F = E(\mu,\varphi) - \frac{1}{2}\ln(2\pi\zeta^*)\,. \tag{33}$$

In the calculation of the mean μ, the second term on the right side can be omitted in the calculation of Eq. (33):

$$F[\mu] = E(\mu,\varphi)\,. \tag{34}$$

After all, the mean of $q(\vartheta)$, which is an approximation of the posterior distribution $p(\vartheta\,|\,\varphi)$, can be calculated by minimizing Eq. (34), and the variance can be calculated by Eq. (32).

4.3 The most accurate approximation (Estimation of mean value μ and variance ζ of posterior distribution ... no assumption of sharp peak in joint distribution)

Without assuming the assumption of Eq. (29) that the joint distribution $p(\vartheta,\varphi)$ has a sharp peak, we consider the functional F of the VBP without introducing any approximation other than the Gaussian approximation:

$$\begin{aligned} F[\mu,\zeta] &= \int q(\vartheta)\ln\frac{q(\vartheta)}{p(\vartheta,\varphi)}d\vartheta \\ &= \int \frac{1}{\sqrt{2\pi\zeta}}exp\left(-\frac{(\vartheta-\mu)^2}{2\zeta}\right)\begin{pmatrix} -\ln\sqrt{2\pi\zeta} - \frac{(\vartheta-\mu)^2}{2\zeta} \\ +\ln(2\pi) + \frac{1}{2}\ln(\Sigma_\varphi\Sigma_\vartheta) + \frac{(\varphi-g(\vartheta))^2}{2\Sigma_\varphi} + \frac{(\vartheta-\bar{\mu})^2}{2\Sigma_\vartheta} \end{pmatrix} d\vartheta. \end{aligned} \tag{35}$$

If we consider the minimum value problem by the steepest descent method, the mean μ and variance ζ of the posterior distribution $p(\vartheta\,|\,\varphi)$ are the solutions to the following problem:

$$(\mu,\zeta) = \underset{(\mu,\zeta)}{\arg\min}\, F(\mu,\zeta)\,. \tag{36}$$

Specifically, it becomes the following equation:

$$(\mu,\zeta)_{t+\delta t} = (\mu,\zeta)_{t+\delta t} + \left(\frac{\partial F}{\partial \mu},\frac{\partial F}{\partial \zeta}\right)_t k\delta t\,. \tag{37}$$

As can be seen from Eq. (35), $\partial F/\partial\mu$ and $\partial F/\partial\zeta$ appear to be too complex to be implemented in vivo.

Now, we consider a calculation that uses the VBP itself given by Eq. (35) from a mathematical interest. In preparation for that, the posterior distribution $p(\vartheta\,|\,\varphi)$ is derived from the joint distribution $p(\varphi,\vartheta)$:

$$p(\varphi,\vartheta)=\frac{1}{\sqrt{2\pi\Sigma_\varphi}}\frac{1}{\sqrt{2\pi\Sigma_\vartheta}}\exp\left(-\frac{(\varphi-g(\vartheta))^2}{2\Sigma_\varphi}-\frac{(\vartheta-\bar{\mu})^2}{2\Sigma_\vartheta}\right). \tag{19}$$

Integrating $p(\varphi,\vartheta)$ with ϑ gives the following equation:

$$p(\varphi)=\int_{-\infty}^{\infty}p(\varphi,\vartheta)\,d\vartheta=\frac{1}{\sqrt{2\pi\Sigma_\varphi}}\frac{1}{\sqrt{2\pi\Sigma_\vartheta}}\int_{-\infty}^{\infty}\exp\left(-\frac{(\varphi-g(\vartheta))^2}{2\Sigma_\varphi}-\frac{(\vartheta-\bar{\mu})^2}{2\Sigma_\vartheta}\right)d\vartheta. \tag{38}$$

Using this, $p(\vartheta\,|\,\varphi)$ can be found as follows:

$$p(\vartheta\,|\,\varphi)=\frac{p(\vartheta,\varphi)}{p(\varphi)}=\frac{\exp\left(-\frac{(\varphi-g(\vartheta))^2}{2\Sigma_\varphi}-\frac{(\vartheta-\bar{\mu})^2}{2\Sigma_\vartheta}\right)}{\int_{-\infty}^{\infty}\exp\left(-\frac{(\varphi-g(\vartheta))^2}{2\Sigma_\varphi}-\frac{(\vartheta-\bar{\mu})^2}{2\Sigma_\vartheta}\right)d\vartheta}. \tag{39}$$

In the case of linear, that is, $g(\vartheta)=\vartheta$, Eqs. (38) and (39) are as follows:

$$p(\varphi)=\frac{1}{\sqrt{(2\pi)^2\Sigma_\varphi\Sigma_\vartheta}}\sqrt{2\pi\frac{\Sigma_\varphi\Sigma_\vartheta}{\Sigma_\vartheta+\Sigma_\varphi}}\exp\left(-\frac{(\bar{\mu}-\varphi)^2}{2(\Sigma_\vartheta+\Sigma_\varphi)}\right)=\frac{1}{\sqrt{2\pi(\Sigma_\vartheta+\Sigma_\varphi)}}\exp\left(-\frac{(\varphi-\bar{\mu})^2}{2(\Sigma_\vartheta+\Sigma_\varphi)}\right), \tag{40}$$

$$p(\vartheta\,|\,\varphi)=\frac{\exp\left(-\frac{(\varphi-\vartheta)^2}{2\Sigma_\varphi}-\frac{(\vartheta-\bar{\mu})^2}{2\Sigma_\vartheta}\right)}{\int_{-\infty}^{\infty}\exp\left(-\frac{(\varphi-\vartheta)^2}{2\Sigma_\varphi}-\frac{(\vartheta-\bar{\mu})^2}{2\Sigma_\vartheta}\right)d\vartheta}=\sqrt{\frac{\Sigma_\vartheta+\Sigma_\varphi}{2\pi\Sigma_\varphi\Sigma_\vartheta}}\exp\left(-\frac{\left(\vartheta-\frac{\Sigma_\varphi\bar{\mu}+\Sigma_\vartheta\varphi}{\Sigma_\vartheta+\Sigma_\varphi}\right)^2}{\frac{2\Sigma_\varphi\Sigma_\vartheta}{\Sigma_\vartheta+\Sigma_\varphi}}\right). \tag{41}$$

This is the correct answer for posterior probability $p(\vartheta\,|\,\varphi)$ when the joint probability is given by Eq. (25) and $g(\vartheta)=\vartheta$. The mean μ and variance ζ are given by:

$$\mu=\frac{\Sigma_\varphi\bar{\mu}+\Sigma_\vartheta\varphi}{\Sigma_\vartheta+\Sigma_\varphi},\quad \zeta=\frac{\Sigma_\varphi\Sigma_\vartheta}{\Sigma_\vartheta+\Sigma_\varphi}. \tag{42}$$

Next, consider a method for calculating $p(\vartheta\,|\,\varphi)$ numerically from Eq. (35). first, we use

$$\int_{-\infty}^{\infty}exp\left(-\frac{(\vartheta-\mu)^2}{2\zeta}\right)d\vartheta=\sqrt{2\zeta\pi},\quad \int_{-\infty}^{\infty}(\vartheta-\mu)^2 exp\left(-\frac{(\vartheta-\mu)^2}{2\zeta}\right)d\vartheta=(2\zeta)^{3/2}\Gamma\left(\frac{3}{2}\right)=(2\zeta)^{3/2}\frac{\sqrt{\pi}}{2}, \tag{43}$$

and simplify Eq. (35):

$$F(\mu,\zeta)=\left(-\frac{1}{2}\ln\zeta+\frac{1}{2}\ln(\Sigma_\varphi\Sigma_\vartheta)+\frac{1}{2}\ln(2\pi)+\frac{(\mu-\bar{\mu})^2}{2\Sigma_\vartheta}\right) + \zeta\left(-\frac{1}{2\zeta}+\frac{1}{2\Sigma_\vartheta}\right)+\frac{1}{\sqrt{2\pi\zeta}}\int_{-\infty}^{\infty} exp\left(-\frac{(\vartheta-\mu)^2}{2\zeta}\right)\frac{(\varphi-g(\vartheta))^2}{2\Sigma_\varphi}d\vartheta. \tag{44}$$

From Eq. (44), $M=\partial F/\partial\mu=0$ and $Z=\partial F/\partial\zeta=0$ are given by the following equation using numerical differentiation:

$$M(\mu,\zeta)\equiv\frac{\partial F(\mu,\zeta)}{\partial\mu}=\frac{F(\mu+d\mu,\zeta)-F(\mu,\zeta)}{d\mu}=0, \tag{45a}$$

$$Z(\mu,\zeta)\equiv\frac{\partial F(\mu,\zeta)}{\partial\zeta}=\frac{F(\mu,\zeta+d\zeta)-F(\mu,\zeta)}{d\zeta}=0. \tag{45b}$$

Using Newton-Raphson's method to solve Eq. (45) gives a two-dimensional simultaneous algebraic equation for $d\mu$ and $d\zeta$:

$$\frac{\partial M}{\partial\mu}d\mu+\frac{\partial M}{\partial\zeta}d\zeta=-M, \tag{46a}$$

$$\frac{\partial Z}{\partial\mu}d\mu+\frac{\partial Z}{\partial\zeta}d\zeta=-Z. \tag{46b}$$

The derivative of M and Z is also calculated by numerical differentiation. Solving this, we can find $d\mu$ and $d\zeta$:

$$d\mu=-\frac{M\frac{\partial Z}{\partial\zeta}-Z\frac{\partial M}{\partial\zeta}}{\frac{\partial M}{\partial\mu}\frac{\partial Z}{\partial\zeta}-\frac{\partial Z}{\partial\mu}\frac{\partial M}{\partial\zeta}},\quad d\zeta=-\frac{\frac{\partial M}{\partial\mu}Z-\frac{\partial Z}{\partial\mu}M}{\frac{\partial M}{\partial\mu}\frac{\partial Z}{\partial\zeta}-\frac{\partial Z}{\partial\mu}\frac{\partial M}{\partial\zeta}}. \tag{47}$$

Therefore, the convergence value of the iterative calculation that updates μ and ζ according to the following equation is the solution:

$$\mu_{new}=\mu_{old}+d\mu,\quad \zeta_{new}=\zeta_{old}+d\zeta \tag{48}$$

By substituting this into Eq. (37), the functional F of the VBP can also be obtained.

The calculation code in C language is shown in Appendix B. Numerical calculations were actually performed, and the result by Eq. (41) and the result by the minimum VBP were compared. The following equation was assumed in the calculation:

$$g(\vartheta)=\vartheta. \tag{49}$$

The parameter values used in the calculation are shown below:

$$\bar{\mu} = 1, \quad \Sigma_{\varphi} = 1.5, \quad \Sigma_{\vartheta} = 2.5, \quad d\mu = d\zeta = 0.000001. \tag{50}$$

In Fig. 3, the initial values of μ and ζ are set only for the minimum value of φ, and φ is sequentially increased while the previous convergence value is used as the initial value of the next calculation. For the initial value at the start, refer to Eq. (42). Figures 3 and 4 show a comparison between the correct answer given in Eqs. (42) and (41) and the approximate solution by solving Eq. (45) numerically, and sufficient accuracy is obtained.

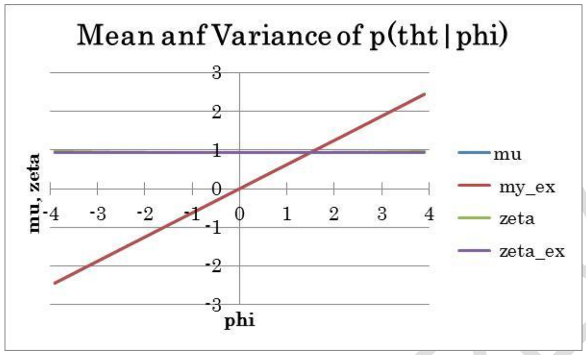

Fig. 3. Mean and variance of $p(\vartheta | \varphi)$

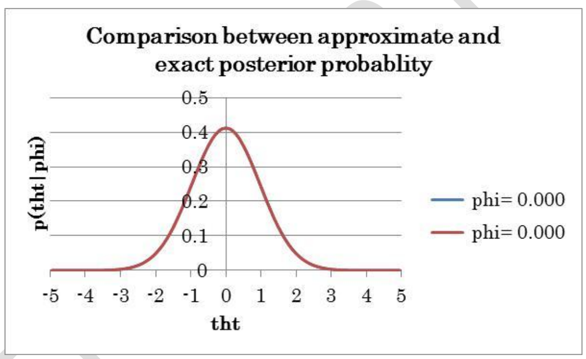

Fig. 4. Posterior probability $p(\vartheta | \varphi)$ at $\varphi = 0$

5. CONCLUSION

The variational Bayesian method is a variational principle in a stocastistic system. Using the variational Bayesian method, it is possible to obtain an approximation of unknown parameters contained in posterior probabilities, similar to the Ritz method in the classical variational principle applied to a deterministic system.

In this paper, we gave a detailed explanation of the classical variational principle and the variational Bayesian method, and the basic application examples of the variational Bayesian method.

C language code is also included to deepen the reader's understanding. Numerical calculations would be very useful, when the stochastic system is nonlinear.

Recently, the variational Bayesian method has been attracting attention in the field of information processing in the cerebrum, which is called the principle of minimum free energy.

REFERENCES

1. Friston K. Nature Reviews | Neuroscie. 2010;11.

2. Buckley CL et al. Journal of Mathematical Psychology. 2017;81, 55–79.

3. Bogacz R. Journal of Mathematical Psychology. 76, 2017, 198-211.

4. Friston K. Hierarchical Models in the Brain. PLoS Computational Biology. 2008;4(11), 1-24.

APPENDIX A. NON=NEGATIVITY OF KULLBACL-LEIBLER DIBERGENCE

The natural logarithm has the following properties:

$$\ln x \le x - 1 . \tag{A.1}$$

If the probability distribution is continuous, then if Θ is the set of all ϑ where $q(\vartheta)$ is not zero, then the following equation holds:

$$-\int_\Theta q \ln \frac{p}{q} d\vartheta \ge -\int_\Theta q\left(\frac{p}{q} - 1\right) d\vartheta = -\int_\Theta p d\vartheta + \int_\Theta q d\vartheta = -\int_\Theta p d\vartheta + 1 \ge 0 . \tag{A.2}$$

Therefore, we obtain:

$$-\int_\Theta q \ln p d\vartheta \ge -\int_\Theta q \ln q d\vartheta . \tag{A.3}$$

Even if 0 is added to both sides, the magnitude relationship does not change, so the following equation holds:

$$\int_{-\infty}^{\infty} q \ln p d\vartheta \le \int_{-\infty}^{\infty} q \ln q d\vartheta . \tag{A.4}$$

Therefore, in the case of continuous probabilities, the non-negativeness of Kullback-Leibler divergence is derived:

$$\int_{-\infty}^{\infty} q \ln \frac{p}{q} d\vartheta \le 0 \to \int_{-\infty}^{\infty} q \ln \frac{q}{p} d\vartheta \ge 0 . \tag{A.5}$$

If the probability distribution is discrete, let I be the set of all i whose q_i is not zero.

$$-\sum_{i\in I} q_i \ln \frac{p_i}{q_i} \ge -\sum_{i\in I} q_i \left(\frac{p_i}{q_i} - 1\right) = -\sum_{i\in I} p_i + \sum_{i\in I} q_i = -\sum_{i\in I} p_i + 1 \ge 0 . \tag{A.6}$$

Therefore, we obtain:

$$-\sum_{i\in I} q_i \ln p_i \ge -\sum_{i\in I} q_i \ln q_i . \tag{A.7}$$

Even if 0 is added to both sides, the magnitude relationship does not change, so the following equation holds:

$$\sum_{i=1}^{n} q_i \ln p_i \le \sum_{i=1}^{n} q_i \ln q_i \,. \tag{A.8}$$

Therefore, in the case of discrete probabilities, the non-negativeness of Kullback-Leibler divergence is also derived:

$$\sum_{i=1}^{n} q_i \ln \frac{p_i}{q_i} \le 0 \rightarrow \sum_{i=1}^{n} q_i \ln \frac{q_i}{p_i} \ge 0 \,. \tag{A.9}$$

APPENDIX B. C-LANGUAGE CODE FOR NUMERICAL CALCULATION OF POSTERIOR PROBABILITY USING VARIATIONAL BAYESIAN METHOD

(1) Code file: PostProbNCalByVFE.c

```
// ---------------------------------------------------------------------------------- //
//                                                                                    //
// File Name: PostProbNCal.c                              2022.04.25-2022.04.28 //
// File Name: PostProbNCalByVFE.c                          2022.04.28-2022.05.08 //
//                                                                                    //
//    Numerical Cal. of Posteriro Prob.                                         //
//                                                                                    //
// ---------------------------------------------------------------------------------- //

#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>

#define PI     3.14159265358979323846    // 円周率の定義

// ----- function ---------------------------------------------------------- //

void main();
void pushKey();

double F(double, double, int);                  // F(myu,zeta)
double M(double, double, int);                  // M(myu,zeta)
double Z(double, double, int);                  // Z(myu,zeta)

double DMmyu(double, double, int);              // DM/Dmyu(myu,zeta,m)
double DMzeta(double, double, int);             // DM/Dzeta(myu,zeta,m)
double DZmyu(double, double, int);              // DZ/Dmyu(myu,zeta,m)
double DZzeta(double, double, int);             // DZ/Dzeta(myu,zeta,m)

// ----- 変数 ------------------------------------------------------ //

char title_memo[5000];
```

```
int Ntht;                      // number of divisiob of theta
int Nphi;                      // number of divisiob of phi

int ITR;                       // number of iteration in Newton-Raphson method

double Tht;                    // -Tht <= tht <= Tht
double tht[2001];              // theta
double dtht;                   // dtheta
double Phi;                    // -Phi <= tht <= Phi
double phi[101];               // phi
double dphi;                   // dphi

double myu;                    // mean of probability q(tht)
double zeta;                   // variance of probability q(tht)
double myu_ini;                // initial value of myu
double zeta_ini;               // initial value of zeta
double dmyu;                   // dmyu
double dzeta;                  // dzeta
double myub;                   // myyu_bar
double Sphi;                   // variance of liklihood functio p(phi|tht)
double Stht;                   // variance of prior probability p(tht)
double delm;                   // delm
double delz;                   // delz
double alp;                    // parameter for iteration

double myuRec[101];            // mean of p(tht|phi);
double zetaRec[101];           // variance of p(tht|phi);

double p_phiBtht[20001][101];      // p(tht|phi): liklihood function
double p_tht[20001];               // p(tht)    : prior probability
double p_tht_phi[2001][101];       // p(tht,phi): joint probability
double p_phi[2001];                // p(phi)    : marginal distribution
double p_thtBphi[2001][101];       // p(tht|phi): posterior probability

double p_thtBphi_VFE[2001][101];  // p(tht|phi): posterior probability by VFE

FILE *fp_inp;                  // file pointer of input file
FILE *fp_out;                  // file pointer of output file

char InputDataFile[80];        // input file name
char OutputDataFile[80];       // output file name

char buf[5000];                // buffer

int prt_ctrl;                  // print control; if prt_ctrl = 1, then print

// ------------------------------------------------------------ //

void main()
{
   int i, m, itr;
   double sum;
```

```
//// open input file for parameters
sprintf(InputDataFile, "PostProbNCalByVFE_inp.dat");

if ((fp_inp = fopen(InputDataFile, "r")) == NULL) {
   printf("Failed in Reading Input Data File! ... %s\n", InputDataFile);
   exit(1);
}

//// open output file
sprintf(OutputDataFile, "PostProbNCalByVFE_out.csv");

if ((fp_out = fopen(OutputDataFile, "w")) == NULL) {
   printf("Failed in Reading Output Data File! ... %s\n", OutputDataFile);
   exit(1);
}

//// input from file
fscanf(fp_inp, "%s", title_memo);

fscanf(fp_inp, "%s %d", buf, &Ntht);
fscanf(fp_inp, "%s %d", buf, &Nphi);

fscanf(fp_inp, "%s %d", buf, &ITR);

fscanf(fp_inp, "%s %lf", buf, &Tht);
fscanf(fp_inp, "%s %lf", buf, &Phi);

fscanf(fp_inp, "%s %lf", buf, &myub);
fscanf(fp_inp, "%s %lf", buf, &Sphi);
fscanf(fp_inp, "%s %lf", buf, &Stht);

fscanf(fp_inp, "%s %lf", buf, &myu_ini);
fscanf(fp_inp, "%s %lf", buf, &zeta_ini);

fscanf(fp_inp, "%s %lf", buf, &delm);
fscanf(fp_inp, "%s %lf", buf, &delz);
fscanf(fp_inp, "%s %lf", buf, &alp);

fscanf(fp_inp, "%s %d", buf, &prt_ctrl);
fclose(fp_inp);

//// output to display
printf("memo: %s\n", title_memo);

printf("Ntht  = %d\n", Ntht);
printf("Nphi  = %d\n", Nphi);

printf("ITR  = %d\n", ITR);

printf("Tht   = %12.6f\n", Tht);
printf("Phi   = %12.6f\n", Phi);

printf("myub  = %12.6f\n", myub);
```

```
printf("Sphi  = %12.6f\n", Sphi);
printf("Stht  = %12.6f\n", Stht);

printf("myu_ini  = %12.6f\n", myu_ini);
printf("zeta_ini = %12.6f\n", zeta_ini);

printf("delm     = %12.6f\n", delm);
printf("delz     = %12.6f\n", delz);
printf("alp      = %12.6f\n", alp);

printf("prt_ctrl = %d\n", prt_ctrl);
printf("\n");

//// output to file
fprintf(fp_out, "memo: %s\n", title_memo);
fprintf(fp_out, "\n");

fprintf(fp_out, "Ntht =, %d\n", Ntht);
fprintf(fp_out, "Nphi =, %d\n", Nphi);

fprintf(fp_out, "ITR =, %d\n", ITR);

fprintf(fp_out, "Tht =, %12.6f\n", Tht);
fprintf(fp_out, "Phi =, %12.6f\n", Phi);

fprintf(fp_out, "myub =, %12.6f\n", myub);
fprintf(fp_out, "Sphi =, %12.6f\n", Sphi);
fprintf(fp_out, "Stht =, %12.6f\n", Stht);

fprintf(fp_out, "myu_ini =, %12.6f\n", myu_ini);
fprintf(fp_out, "zeta_ini =, %12.6f\n", zeta_ini);

fprintf(fp_out, "delm =, %12.6f\n", delm);
fprintf(fp_out, "delz =, %12.6f\n", delz);
fprintf(fp_out, "alp =, %12.6f\n", alp);

fprintf(fp_out, "prt_ctrl =, %d\n", prt_ctrl);
fprintf(fp_out, "\n");

pushKey();


///////////////////////////
//                       //
//   Analytical Solution   //
//                       //
///////////////////////////

dtht = 2.0*Tht/(Ntht+0.0);
for (i = 0; i < Ntht; i++)
    tht[i] = -Tht+(i+0.5)*dtht;

dphi = 2.0*Phi/(Nphi+0.0);
```

```
for (m = 0; m < Nphi; m++)
    phi[m] = -Phi+(m+0.5)*dphi;


// p(phi|tht): liklihood function
for (m = 0; m < Nphi; m++)
    for (i = 0; i < Ntht; i++)
        p_phiBtht[i][m] = 1.0/sqrt(2.0*PI*Sphi)*exp(-(phi[m]-tht[i])*(phi[m]-tht[i])/2.0/Sphi);


fprintf(fp_out, "**********\n");
i = Ntht/2;
fprintf(fp_out, "i, m, tht[Ntht/2], phi[m], p_phiBtht[Ntht/2][m]\n");
for (m = 0; m < Nphi; m++)
    fprintf(fp_out, "%d, %d, %lf, %lf, %lf\n",
            i, m, tht[i], phi[m], p_phiBtht[i][m]);
fprintf(fp_out, "\n");



fprintf(fp_out, "p(phi|tht):, liklihood function\n");
fprintf(fp_out, "phi, ");
for (i = 0; i < Ntht; i++)
    fprintf(fp_out, "tht=%6.3f, ", tht[i]);
fprintf(fp_out, "\n");
for (m = 0; m < Nphi; m++) {
    fprintf(fp_out, "%lf, ", phi[m]);
    for (i = 0; i < Ntht; i++)
        fprintf(fp_out, "%lf, ",  p_phiBtht[i][m]);
    fprintf(fp_out, "\n");
}
fprintf(fp_out, "\n");

// p(tht): prior probability
fprintf(fp_out, "p(tht):, prior probability\n");
fprintf(fp_out, " i, tht, p_tht\n");
for (i = 0; i < Ntht; i++) {
    p_tht[i] = 1.0/sqrt(2.0*PI*Stht)*exp(-(tht[i]-myub)*(tht[i]-myub)/2.0/Stht);
    fprintf(fp_out, "%d, %lf, %lf\n", i, tht[i], p_tht[i]);
}
fprintf(fp_out, "\n");

// p(tht,phi): joint probability
for (i = 0; i < Ntht; i++)
    for (m = 0; m < Nphi; m++)
        p_tht_phi[i][m] = 1.0/sqrt(2.0*PI*Stht)*1.0/sqrt(2.0*PI*Sphi)
                          *exp(-(phi[m]-tht[i])*(phi[m]-tht[i])/2.0/Sphi
                             -(tht[i]-myub)*(tht[i]-myub)/2.0/Stht);

fprintf(fp_out, "p(thtCphi):, joint probability\n");
fprintf(fp_out, "tht, ");
for (m = 0; m < Nphi; m++)
    fprintf(fp_out, "phi=%6.3f, ", phi[m]);
```

```
fprintf(fp_out, "\n");
for (i = 0; i < Ntht; i++) {
    fprintf(fp_out, "%lf, ", tht[i]);
    for (m = 0; m < Nphi; m++)
        fprintf(fp_out, "%lf, ",  p_tht_phi[i][m]);
    fprintf(fp_out, "\n");
}
fprintf(fp_out, "\n");

// p(phi): marginal distribution
fprintf(fp_out, "p(phi):, marginal probability\n");
fprintf(fp_out, "m, phi, p(phi)\n");
for (m = 0; m < Nphi; m++) {
    p_phi[m] = 1.0/sqrt(2.0*PI*(Stht+Sphi))*exp(-(phi[m]-myub)*(phi[m]-myub)
                          /2.0/(Stht+Sphi));
    fprintf(fp_out, "%d, %lf, %lf\n", m, phi[m], p_phi[m]);
}
fprintf(fp_out, "\n");

// p(tht|phi): posterior probability...analytical
for (i = 0; i < Ntht; i++)
    for (m = 0; m < Nphi; m++)
        p_thtBphi[i][m] = p_tht_phi[i][m]/p_phi[m];

fprintf(fp_out, "p(tht|phi):, posterior probabilityy...analytical\n");
fprintf(fp_out, "tht, ");
for (m = 0; m < Nphi; m++)
    fprintf(fp_out, "phi=%6.3f, ", phi[m]);
fprintf(fp_out, "\n");
for (i = 0; i < Ntht; i++) {
    fprintf(fp_out, "%lf, ", tht[i]);
    for (m = 0; m < Nphi; m++)
        fprintf(fp_out, "%lf, ",  p_thtBphi[i][m]);
    fprintf(fp_out, "\n");
}
fprintf(fp_out, "\n");

fprintf(fp_out, "p(tht|phi):, posterior probabilityy...analytical\n");
fprintf(fp_out, "phi, ");
for (i = 0; i < Ntht; i++)
    fprintf(fp_out, "tht=%6.3f, ", tht[i]);
fprintf(fp_out, "\n");
for (m = 0; m < Nphi; m++) {
    fprintf(fp_out, "%lf, ", phi[m]);
    for (i = 0; i < Ntht; i++)
        fprintf(fp_out, "%lf, ",  p_thtBphi[i][m]);
    fprintf(fp_out, "\n");
}
fprintf(fp_out, "\n");

////////////////////////////////
//                            //
// Numerical Solution by      //
```

```
// Variational Basian Principle //
//                              //
//////////////////////////////////

// Newton Raphson

myu = myu_ini;
zeta = zeta_ini;

for (m = 0; m < Nphi; m++) {
   if (prt_ctrl == 1) {
      fprintf(fp_out, "m =, %d, phi =, %lf\n", m, phi[m]);
      fprintf(fp_out, "itr, dmyu, dzeta, myu, zeta, F\n");
      fprintf(fp_out, "%d, %lf, %lf, %lf, %lf\n", 0, 0.0, 0.0, myu, zeta);
   }
   for (itr = 1; itr <= ITR; itr++) {
      dmyu = -(M(myu,zeta,m)*DZzeta(myu,zeta,m) - Z(myu,zeta,m)*DMzeta(myu,zeta,m))
            /(DMmyu(myu,zeta,m)*DZzeta(myu,zeta,m)
             - DZmyu(myu,zeta,m)*DMzeta(myu,zeta,m));
      dzeta = -(DMmyu(myu,zeta,m)*Z(myu,zeta,m) - DZmyu(myu,zeta,m)*M(myu,zeta,m))
            /(DMmyu(myu,zeta,m)*DZzeta(myu,zeta,m)
             - DZmyu(myu,zeta,m)*DMzeta(myu,zeta,m));

      myu += alp*dmyu;
      zeta += alp*dzeta;

      if (prt_ctrl == 1)
         fprintf(fp_out, "%d, %lf, %lf, %lf, %lf, %lf\n",
               itr, dmyu, dzeta, myu, zeta, F(myu,zeta,m));
   }
   if (prt_ctrl == 1)
      fprintf(fp_out, "\n");

   myuRec[m] = myu;
   zetaRec[m] = zeta;
}

fprintf(fp_out, "mean and variance of p(tht|phi)\n");
fprintf(fp_out, "m, phi,myuRec, my_ex, zetaRec, zeta_ex\n");
for (m = 0; m < Nphi; m++)
   fprintf(fp_out, "%d, %lf, %lf, %lf, %lf, %lf\n",
         m, phi[m], myuRec[m], (Sphi*myub+Stht*phi[m])/(Stht+Sphi),
         zetaRec[m], Sphi*Stht/(Stht+Sphi));
fprintf(fp_out, "\n");

// p(tht|phi) by VFE)
for (i = 0; i < Ntht; i++)
   for (m = 0; m < Nphi; m++)
      p_thtBphi_VFE[i][m] = 1.0/sqrt(2.0*PI*zetaRec[m])*exp(-(tht[i]-myuRec[m])
                     *(tht[i]-myuRec[m])/2.0/zetaRec[m]);

fprintf(fp_out, "p(tht|phi):, posterior probability by VFE...numerical\n");
fprintf(fp_out, "tht, ");
```

```
    for (m = 0; m < Nphi; m++)
        fprintf(fp_out, "phi=%6.3f, ", phi[m]);
    fprintf(fp_out, "\n");
    for (i = 0; i < Ntht; i++) {
        fprintf(fp_out, "%lf, ", tht[i]);
        for (m = 0; m < Nphi; m++)
            fprintf(fp_out, "%lf, ",  p_thtBphi_VFE[i][m]);
        fprintf(fp_out, "\n");
    }
    fprintf(fp_out, "\n");


    fclose(fp_out);

    pushKey();
}

// ---------------------------------------------------------------- //

 void pushKey()
{
    printf("\n      Push Return Key! ");
    getchar();
    getchar();
}

// ---------------------------------------------------------------- //

double F(double myu, double zeta, int m)
{
    int i;
    double sum;

    sum = 0.0;
    for (i = 0; i < Ntht; i++)
        sum += exp(-(tht[i]-myu)*(tht[i]-myu)/2.0/zeta)*(phi[m]-tht[i])
                   *(phi[m]-tht[i])/2.0/Sphi;
    sum *= dtht;
    sum *= 1.0/sqrt(2.0*PI*zeta);

    return (-0.5*log(zeta)+0.5*log(Sphi*Stht)+0.5*log(2.0*PI)
          +(myu-myub)*(myu-myub)/2.0/Stht)
          + zeta*(-1.0/2.0/zeta+1.0/2.0/Stht) + sum;
}

// ---------------------------------------------------------------- //

double M(double myu, double zeta, int m)
{
    return (F(myu+delm,zeta,m)-F(myu,zeta,m))/delm;
}

// ---------------------------------------------------------------- //
```

```
double Z(double myu, double zeta, int m)
{
   return (F(myu,zeta+delz,m)-F(myu,zeta,m))/delz;
}

// ---------------------------------------------------------------- //

double DMmyu(double myu, double zeta, int m)
{
   return (M(myu+delm,zeta,m)-M(myu,zeta,m))/delm;
}

// ---------------------------------------------------------------- //

double DMzeta(double myu, double zeta, int m)
{
   return (M(myu,zeta+delz,m)-M(myu,zeta,m))/delz;
}

// ---------------------------------------------------------------- //

double DZmyu(double myu, double zeta, int m)
{
   return (Z(myu+delm,zeta,m)-Z(myu,zeta,m))/delm;
}

// ---------------------------------------------------------------- //

double DZzeta(double myu, double zeta, int m)
{
   return (Z(myu,zeta+delz,m)-Z(myu,zeta,m))/delz;
}

// ---------------------------------------------------------------- //
```

(2) Input file: PostProbNCalByVFE_inp.dat

```
2022.05.08

Ntht        101
Nphi         41

ITR          20

Tht         5.0
Phi         4.0

myub         0.0
Sphi        1.5
Stht        2.5
```

myu_ini -2.5
zeta_ini 1.0

delm 0.000001
delz 0.000001
alp 0.25

prt_ctrl 0