

# Bayesian Inference from Symplectic Geometric Viewpoint

Tomonori Noda<sup>1</sup>, Hinako Matsuyama<sup>2</sup>

<sup>1</sup>General Education and Research Center, Meiji Pharmaceutical University, Kiyose, Japan

<sup>2</sup>Meiji Pharmaceutical University, Kiyose, Japan

Email: noda@my-pharm.ac.jp

**How to cite this paper:** Noda, T. and Matsuyama, H. (2019) Bayesian Inference from Symplectic Geometric Viewpoint. *Advances in Pure Mathematics*, 9, 827-831. <https://doi.org/10.4236/apm.2019.910039>

**Received:** August 30, 2019

**Accepted:** September 24, 2019

**Published:** September 27, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The purpose of this article is to formulate Bayesian updating from dynamical viewpoint. We prove that Bayesian updating for population mean vectors of multivariate normal distributions can be expressed as an affine symplectic transformation on a phase space with the canonical symplectic structure.

## Keywords

Bayesian Inference, Hamiltonian, Symplectic, Canonical Transformation

---

## 1. Introduction

In Bayesian statistics, we consider parameters in models as random variables and its probability distributions, and we calculate the posterior distribution by using Bayes' theorem.

In Hamiltonian dynamical system, any time evolution is defined by Hamiltonian equations and expressed by canonical transformations (or symplectic diffeomorphisms) on phase spaces. Phase spaces and equations of motion are abstract symplectic manifolds and Hamiltonian vector fields respectively. Under time evolution for Hamiltonian dynamical system, Hamiltonian functions and the phase volume are preserved. These are direct consequences of skew-symmetry of symplectic structures. In the case where the dimension of a phase space is greater than or equal to 4, there are other conserved quantities that are called symplectic capacities. Symplectic capacities are far from trivial and are deep result in symplectic geometry. For detail, see [1] [2] [3] [4].

In this paper we prove that Bayesian updating for multivariate normal population mean vector can be expressed by an affine symplectic diffeomorphism (affine canonical transformation). The main result is the following.

**Theorem 1.** *There exists a linear symplectic diffeomorphism  $\Phi_1$  on  $\mathbb{R}^{2n}$  such that the first component of the composition  $T(\mu_n, 0) \circ \Phi_1 \circ T(-\mu_0, 0)$  maps a prior distribution to the posterior, where  $T(a, b)$  is the parallel translation  $T : (x, y) \mapsto (x + a, y + b)$  on  $\mathbb{R}^{2n}$ .*

In this theorem to reformulate Bayesian updating from symplectic geometric viewpoint, we consider the cotangent space of which its base space contains population mean  $\mu$ . The reason why we use the cotangent space is as follows. If we assume  $\mu$  is a point in  $\mathbb{R}^{2n}$ , to express Bayes' theorem by an affine symplectic transformation on  $\mathbb{R}^{2n}$ , in the case where variance is known, we have to find an element  $\Phi$  in  $\text{Sp}(2n, \mathbb{R})$  such that  $\Phi^T \Lambda_0^{-1} \Phi = \Lambda_0^{-1} + n\Sigma^{-1}$ . (Note that  $\Phi, \Sigma, \Lambda$  are all  $2n \times 2n$ .) However usually in Bayesian updating, we fix a type of the posterior which is a section of density function, and then we normalize it:

$$p(\theta | y) = \frac{l(\theta | y)p(\theta)}{p(y)} \propto l(\theta | y)p(\theta). \tag{1}$$

It is well known that any canonical transformations are volume-preserving. Hence we cannot expect the existence of desired transformation. Moreover, in this case the population mean  $\mu$  is in  $\mathbb{R}^{2n}$ , so we can only treat even dimensional case. The key to getting rid of this drawback is to use of Lagrangian submanifolds. Consider a symplectic manifold which has a Lagrangian submanifold containing  $\mu$ , and construct desired canonical transformation on the total space. Canonical transformations on the total space may change a measure on Lagrangian submanifolds.

There is another approach to Bayesian inference from symplectic-contact geometric viewpoint due to Mori [5] [6]. Mori considers the square of the parameter space of normal distributions and its Lagrangian submanifold to describe Bayes' theorem by Hamiltonian follows, and he simultaneously gives Bayesian updating for mean and variance in univariate case. Taking account of Mori's considerations, we should use the Poincaré type symplectic form to express a Bayesian updating for covariant matrices while we use the canonical symplectic structure on  $\mathbb{R}^{2n}$  for mean vectors. For information geometry and a relation between information geometry and symplectic geometry, see [7] [8] [9].

## 2. Bayesian Updating

In this section we review Bayes' theorem for multivariate normal distributions. For detail, see e.g., [10].

Consider a posterior distribution of mean vector for a multivariate normal distribution with covariant matrix. Fix a positive definite symmetric matrix  $\Sigma$ . First we treat the case of covariance matrix  $\Sigma$  is known. Let a prior distribution  $p(\mu)$  of  $\mu$  is distributed  $N(\mu_0, \Lambda_0)$ :

$$p(\mu) \propto \exp\left[-\frac{1}{2}(\mu - \mu_0)^T \Lambda_0^{-1} (\mu - \mu_0)\right]. \tag{2}$$

The posterior distribution with sample  $y$  is  $\mu | y, \Sigma \sim N(\mu_n, \Lambda_n)$  and

$$p(\mu | y, \Sigma) \propto \exp \left[ -\frac{1}{2} (\mu - \mu_n)^T \Lambda_n^{-1} (\mu - \mu_n) \right], \tag{3}$$

where

$$\mu_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1} (\Lambda_0^{-1}\mu_0 + \Sigma^{-1}\bar{y}), \quad \Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}. \tag{4}$$

Next we consider the case of the variance  $\Sigma$  is unknown. If we denote a priori distribution of  $\mu$  by  $\mu | \Sigma \sim N(\mu_0, \Sigma/k_0)$ ,  $\Sigma \sim IW(\nu_0, \Lambda_0)$ , then the posterior is  $\mu | y, \Sigma \sim N(\mu_n, \Sigma/k_n)$ ,  $\Sigma | y \sim IW(\nu_n, \Lambda_n)$ , where

$$\nu_n = \frac{k_0}{k_0 + n} \nu_0 + \frac{n}{k_0 + n} \bar{y}, \quad k_n = k_0 + n, \quad \nu_n = \nu_0 + n, \quad \Lambda_n = \Lambda_0 + S + \frac{k_0 n}{k_0 + n} S_0. \tag{5}$$

### 3. Symplectic Group and Affine Canonical Transformation

In this section we review properties of the symplectic group and Hamiltonian flows.

Denote the set of all linear symplectic transformations on  $\mathbb{R}^{2n}$  by

$$\text{Sp}(2n, \mathbb{R}) = \left\{ S \in \text{GL}(2n, \mathbb{R}); S^T J S = J \right\} \tag{6}$$

and call the *symplectic group*, where  $J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$ .

Let  $z$  be a vector in  $\mathbb{R}^{2n}$  and  $\omega_0$  be the canonical symplectic structure on  $2n$  dimensional vector space  $\mathbb{R}^{2n}$ , then a necessary and sufficient condition for  $\omega_0(z, z') = \omega_0(Sz, Sz')$  is  $S^T J S = J$ . For any  $S \in \text{Sp}(2n, \mathbb{R})$  we have  $\det S = 1$ , where  $\det S$  denotes the determinant of matrix  $S$ . We also have  $\text{Sp}(2n, \mathbb{R}) = \text{SL}(2, \mathbb{R})$  for  $n = 1$ . In general  $\text{Sp}(2n, \mathbb{R})$  is a connected Lie group of dimension  $n(2n + 1)$ , and the Lie algebra is given by

$\mathfrak{sp}(n, \mathbb{R}) = \{ M \in M(2n, \mathbb{R}); JM + M^T J = 0 \}$ . If we write  $S \in \text{Sp}(2n, \mathbb{R})$  in

terms of  $n \times n$  block matrices by  $S = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ , then

$$A^T C = C^T A, \quad B^T D = D^T B, \quad A^T D - C^T B = I_n. \tag{7}$$

Hence the inverse matrix of  $S$  is given by  $S^{-1} = \begin{bmatrix} D^T & -B^T \\ -C^T & A^T \end{bmatrix}$ . For details, see

Abraham-Marsden [1] and de Gosson [2].

If we consider time evolutions of  $\rho_0 \sim N(0, \Sigma)$  by a Hamiltonian flow, the resulting function is distributed multivariate normal.

**Lemma 1.** *If we evolve a density function such that  $\rho_0 \sim N(0, \Sigma)$  by a linear Hamiltonian system with transition matrix  $\Phi_t \in \text{Sp}(2n, \mathbb{R})$ , then we have  $\rho_t \sim N(0, \Sigma(t))$ , where  $\Sigma(t) = \Phi_t \Sigma \Phi_t^T$ .*

For any Hamiltonian equation

$$\begin{cases} \dot{q}_j = \frac{\partial H}{\partial p_j} \\ \dot{p}_j = -\frac{\partial H}{\partial q_j} \end{cases} \tag{8}$$

the transition matrix is given by  $\Phi_t = \exp \left( t \begin{bmatrix} 0 & \frac{\partial H}{\partial p_j} \\ -\frac{\partial H}{\partial q_j} & 0 \end{bmatrix} \right)$  and satisfies

$\Phi_t \in \text{Sp}(2n, \mathbb{R})$  for any  $t$ .

Lemma 1 shows that if we evolve a density function distributed a normal distribution by Hamiltonian equations, then the result is also distributed a normal distribution whose variance is obtained to original variance by multiplying the transition matrix from left and its transpose from right. The proof is straightforward as follows. By  $\Phi_{-t} = \Phi_t^{-1}$  and  $\det \Phi_t = 1$ , we have

$$\begin{aligned} \rho_t &= \rho_0(\Phi_{-t}x) = \left(\frac{1}{2\pi}\right)^n (\det \Sigma)^{-1/2} \exp\left[-\frac{1}{2}x^T \Phi_{-t}^T \Sigma^{-1} \Phi_{-t} x\right] \\ &= \left(\frac{1}{2\pi}\right)^n (\det(\Phi_t \Sigma \Phi_t^T))^{-1/2} \exp\left[-\frac{1}{2}x^T (\Phi_t \Sigma \Phi_t^T)^{-1} x\right]. \end{aligned}$$

### 4. Proof of the Theorem

To prove the theorem, we explicitly construct affine symplectic diffeomorphisms.

First we consider known variance case. Let  $\omega_0$  be the canonical symplectic structure on  $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$ . We consider  $\mu$  is in the first factor of  $\mathbb{R}^{2n}$ , and take the matrix  $\begin{bmatrix} \Lambda_0^{-1} & 0 \\ 0 & I_n \end{bmatrix}$  which corresponds to apriori distribution. Let

$$\Phi_1 = S^T = \begin{bmatrix} I_n & -(\Lambda_0^{-1} + n\Sigma^{-1})^{-1} (n\Sigma^{-1})^{1/2} \\ (n\Sigma^{-1})^{1/2} & I_n - (n\Sigma^{-1})^{1/2} (\Lambda_0^{-1} + n\Sigma^{-1})^{-1} (n\Sigma^{-1})^{1/2} \end{bmatrix} \tag{9}$$

then we have  $\Phi_1 \in \text{Sp}(2n, \mathbb{R})$  and

$$S \begin{bmatrix} \Lambda_0^{-1} & 0 \\ 0 & I_n \end{bmatrix} S^T = \begin{bmatrix} \Lambda_n^{-1} & 0 \\ 0 & I_n - (n\Sigma^{-1})^{1/2} (\Lambda_0^{-1} + n\Sigma^{-1})^{-1} (n\Sigma^{-1})^{1/2} \end{bmatrix}. \tag{10}$$

Hence the Bayesian updating can be expressed as

$$T(\mu_n, 0) \circ \tilde{\Phi}_1 \circ T(-\mu_0, 0), \tag{11}$$

where  $T(a, b)$  denotes the parallel translation  $T : (x, y) \mapsto (x + a, y + b)$  on  $\mathbb{R}^{2n}$ , and  $\circ$  denotes the composition of maps.

In the case where unknown variance, we take a matrix  $\begin{bmatrix} k_0 \Sigma^{-1} & 0 \\ 0 & I_n \end{bmatrix}$  as apriori distribution and set

$$\Phi_1 = S^T = \begin{bmatrix} I_n & n^{1/2} \Sigma^{-1/2} \\ -(2k_0)^{-1} n^{1/2} \Sigma^{1/2} & \left(1 - \frac{n}{2k_0}\right) \Sigma \end{bmatrix} \in \text{Sp}(2n, \mathbb{R}), \tag{12}$$

then the desired transformation is given by

$$T(\mu_n, 0) \circ \tilde{\Phi}_1 \circ T(-\mu_0, 0). \quad (13)$$

## 5. Conclusion

In this paper we show that Bayesian updating can be expressed by an affine symplectic diffeomorphism on  $\mathbb{R}^{2n}$  whose base space contains a population mean vector. Bayesian updating is widely used in several areas, and recently it is usual to use computers to determine the posterior, implicitly. However our theorem expresses the posterior explicitly and concretely, and gives a dynamical interpretation of Bayesian updating.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Abraham, R. and Marsden, J.E. (1987) Foundations of Mechanics. Second Edition, Addison-Wesley Publishing Company, Inc., Boston.
- [2] De Gosson, M.A. (2006) Symplectic Geometry and Quantum Mechanics. Birkhäuser, Basel. <https://doi.org/10.1007/3-7643-7575-2>
- [3] Hofer, H. and Zehnder, E. (1994) Symplectic Invariants and Hamiltonian Dynamics. Birkhäuser, Basel. <https://doi.org/10.1007/978-3-0348-8540-9>
- [4] Weinstein, A. (1981) Symplectic Geometry. *Bulletin of the American Mathematical Society*, **5**, 1-13. <https://doi.org/10.1090/S0273-0979-1981-14911-9>
- [5] Mori, A. (2018) Information Geometry in a Global Setting. *Hiroshima Mathematical Journal*, **48**, 291-305. <https://doi.org/10.32917/hmj/1544238029>
- [6] Mori, A. A Congruence Theorem for Alpha-Connections on the Space of T-Distributions and Its Application.
- [7] Amari, S. and Nagaoka, H. (2000) Methods of Information Geometry. Translations of Mathematical Monographs, Vol. 191, American Mathematical Society, Providence, Oxford University Press, Oxford, Translated from the 1993 Japanese Original by Daishi Harada.
- [8] Boumuki, N. and Noda, T. (2016) On Gradient and Hamiltonian Flows on Even Dimensional Dually Flat Spaces. *Fundamental Journal of Mathematics and Mathematical Sciences*, **6**, 51-66.
- [9] Noda, T. (2011) Symplectic Structures on Statistical Manifolds. *Journal of the Australian Mathematical Society*, **90**, 371-384. <https://doi.org/10.1017/S1446788711001285>
- [10] Lesaffre, E. and Lawson, A.B. (2012) Bayesian Biostatistics (Statistics in Practice). Wiley, Hoboken. <https://doi.org/10.1002/9781119942412>