Prior Distributions Using the Entropy Principles

Jung Jin Lee *, Wan Seon Shin **

〈Abstract〉

Several practical prior distributions are derived using the maximum entropy principle. Also, an interactive method for estimating a prior distribution which uses the minimum cross-entropy principle is proposed when there are many prior informations. The consistency of the prior distributions obtained by the entropy principles is discussed.

I. INTRODUCTION

In Bayesian decision analysis, a prior distribution should be estimated before we collect any sample. Typically, we choose one family of distirbutions and then select a particular member within the chosen family [6], [7]. The choice of the family of distributions and the selection of a particular member are wholly determined by the decision maker's expressed beliefs or betting odds. This subjective method for estimating a prior distribution has been the main criticism of the Bayesian approach to statistical decision making.

In order to remove 'subjectiveness' in estimating a prior distribution, Janes [4] proposed to use the maximum entropy principle (MEP). He said that the prior distributions obtained by the MEP may be as 'objective' as the sampling distributions and correspond to the traditional frequency distributions.

However, Janes [4] considered only a simple expectation as a prior information in his MEP formulation for estimating a prior distribution. In this paper, explicit solutions of the MEP formulations using various prior informations are derived. A consistency property of the prior distribution obtained by the MEP is discussed.

If there are many prior informations, the MEP formulation does not usually have an explicit solution and we have to solve a system of complicated nonlinear equations. In this paper, an interactive method which uses the minimum cross-entropy principle

* Soong Sil University, Dept. of Statistics, Sangdo-dong, Dongjak-Ku, Seoul, Korea 156 743
** Mississippi State University, Dept. of Industrial Engineering, Mississippi State, MS 39762, U.S.A.

(MCEP) is proposed to find a prior distribution when there are many prior informations. A consistency property of the prior distri-bution obtained by the MCEP is discussed.

II. THE ENTROPY PRINCIPLES

Consider a random variable X with the state space Ω and the probability density function $p(x)$. The Shannon's measure of entropy [10], or uncertainty, is defined as follows :

$$-\int_{\Omega} p(x) \ln p(x)\, dx \tag{1}$$

The MEP by Janes [2], [3] based on the entropy measure has been used with great success in many areas (for a list of references, see [5], [13]). The MEP is a useful method for finding a probability distribution whatever knowledge we have about the outcomes of the distribution. In this method, the probability distribution is taken to be the one that maximizes the entropy subject to the constraints that reflect the knowledge about the outcomes. The MEP states that, of all the distributions which satisfy the constraints, one should choose "the most unprejudiced" probability distribution.

The MEP formulation to estimate the $p(x)$ when the knowledge of outcomes are expectations is as follows :

$$\begin{aligned} &\text{Max} \quad -\int_{\Omega} p(x) \ln p(x)\, dx \\ &\text{Subject to} \\ &\qquad \int_{\Omega} p(x)\, dx = 1 \\ &\qquad \int_{\Omega} g_j(x)\, p(x)\, dx = G_j \qquad j = 1, \cdots, m \\ &\qquad\quad p(x) \geqq 0, \text{ for } x \in \Omega \end{aligned} \tag{2}$$

where $g_j(x)$, $j=1, \cdots, m$, is an arbitrary function whose expectation exists. The well known solution of the MEP formulation (2) has the form :

$$p(x) = \frac{1}{Z(\lambda_1, \cdots, \lambda_m)} \exp\left[\lambda_1 g_1(x) + \cdots + \lambda_m g_m(x)\right] \tag{3}$$

where

$$Z(\lambda_1, \cdots, \lambda_m) = \int_{\Omega} \exp[\lambda_1 g_1(x) + \cdots + \lambda_m g_m(x)]\, dx \tag{4}$$

and the real constants, $\lambda_1, \cdots, \lambda_m$, are to be determined from the constraints which reduce to the relations

$$G_i = \frac{\partial}{\partial \lambda_i} \ln Z(\lambda_1, \cdots, \lambda_m) \tag{5}$$

The cross-entropy (called also 'discrimination information' or 'directed divergence') between probability density function $q(x)$ and $p(x)$ is defined as follows :

$$H(q, p) = \int q(x) \ln[q(x)/p(x)]\, dx \tag{6}$$

This measure by Kullback [8] has also a long history of applications in many areas (for a list of references, see [11], [12]).

The minimum cross-entropy principle (MCEP) estimates an unknown probability density $q(x)$ when there exists a prior estimate $p(x)$ of $q(x)$ and new information about $q(x)$ in the form of constraints on expected values. The MCEP states that, of all the densities that satisfy the constraints, one should choose $q(x)$ which is "the most difficult to discriminate (or the least cross-entropy)" from $p(x)$ as follows :

$$\begin{aligned} &\text{Min} \quad H(q, p) = \int q(x) \ln[q(x)/(x)]\, dx \\ &\text{Subject to} \\ &\quad \int_{\Omega} q(x)\, dx = 1 \\ &\quad \int_{\Omega} g_j(x)\, q(x)\, dx = G_j \qquad j = 1, \cdots, m \\ &\quad q(x) \geqq 0, \text{ for } x \in \Omega \end{aligned} \tag{7}$$

where $g_j(x)$, $j=1, \cdots, m$, is an arbitrary function whose expectation exists. The solution has the form

$$q(x) = p(x) \exp[\mu + \lambda_1 g_1(x) + \cdots + \lambda_m g_m(x)] \tag{8}$$

where the real constants μ and λ_k, $k=1, \cdots, m$, are to be determined from the constraints

Note that the MEP is equivalent to the MCEP in the special case of uniform prior estimate $p(x)$.

III. PRIOR DISTRIBUTIONS USING THE MEP

As Janes [4] explained, the prior distributions obtained by the MEP may be as 'objective' as the sampling distributions and correspond to the traditional frequency distributions. However, certain prior information in our real world decision making seems too vague to be used in the MEP formulation. For example, if we were told that "Dow-Jones index would rise tomorrow", then it is less clear how it could lead to any unique prior probability assignment. In order to use the MEP for estimatikng a prior distribbution, each prior information can be translatable into probabilistic terms. The following examples of prior information are commonly encountered in our real world decision and can be used in the MEP formulation. Let X be a random variable with state space Ω and A, B, A^1, A^2, $\cdots$, A_n be disjoint events of Ω.

I_1 : Expectation of X

I_2 : Expectation and variance of X

I_3 : Difference of $P(A)$ and $P(B)$

I_4 : Ratio of $P(A)$ and $P(B)$

I_5 : Information about $P(A_1)$, $P(A_2)$, $\cdots$, $P(A_n)$

The well known solution of the MEP formulation using I_1 is an exponential distribution. The solution of the MEP formulation using I_2 is a normal distribution. The explicit solutions using the prior information I_3, I_4 and I_5 are derived in the following propositions. The proofs are in the appendix.

Proposition 1. Consider a continuous random variable, X, with the state space $\Omega = \{(L, U)\}$. Let $A = \{(a, b)\}$ and $B = \{(c, d)\}$ be two disjoint subsets of Ω, Denote $P(A)$ and $P(B)$ are the probabilities of A and B respectively. Suppose we know the difference d of $P(A)$ and $P(B)$ as the prior information I_3. Then the solution of the MEP formulation with the constraint

$$\int_A p(x)\,\mathrm{dx} - \int_B p(x)\mathrm{dx} + d = 0 \qquad (9)$$

is

$$\begin{aligned} p(x) &= 1/C_1C_2 && \text{if } x \in A \\ &= C_1/C_2 && \text{if } x \in B \\ &= 1/C_2 && \text{otherwise} \end{aligned} \qquad (10)$$

where

$$C_1 = \frac{d(n - n_A - n_B) + \sqrt{d^2(n - n_A - n_B)^2 + 4(1-d^2)n_A n_B}}{2n_B(1-d)},$$

$$C_2 = \frac{(n - n_A - n_B) + \sqrt{d^2(n - n_A - n_B)^2 + 4(1-d^2)n_A n_B}}{(1-d^2)},$$

$n = U - L$, $n_A = b - a$, and $n_B = d - c$.

As a special case, if $\Omega = A \cup B$, the solution becomes as follows ;

$$\begin{aligned} p(x) &= (1-d)/[2(b-a)] && \text{if } x \in A \\ &= (1+d)/[2(d-c)] && \text{if } x \in B, \end{aligned} \tag{11}$$

Proposition 2. Suppose a situation is similar to the Proposition 1, but we know the ratio of $P(A)$ and $P(B)$ as the prior information I_4. Then the solution of the MEP formulation with the constraint

$$\int_A p(x)\mathrm{dx} - r\int_B p(x)\mathrm{dx} = 0 \tag{12}$$

$$\begin{aligned} p(x) &= 1/C_1C_2 && \text{if } x \in A \\ &= (C_1)^r/C_2 && \text{if } x \in B \\ &= 1/C_2 && \text{otherwise} \end{aligned} \tag{13}$$

where

$$C_1 = [n_A/(rn_B)]^{1/(1+r)},$$

$$C_2 = n_A(rn_B/n_A)^{1/(1+r)} + n_B[n_A/(rn_B)]^{r/(1+r)} + n - n_A - n_B,$$

$n = U - L$, $n_A = b - a$, and $n_B = d - c$.

Proposition 3. Consider a continuous random variable, X, with the state space $\Omega = \{[L, U)\}$. Let $L = b_0 < b_1 < b_2 < \cdots < b_n = U$, Suppose we know the probability e_k of each sub-interval $[b_{k-1}, b_k,]$ $k = 1, 2, \cdots, n$, as the prior information I_5. Then the solution of the MEP formulation with the constraints

$$\int_{b_{k-1}}^{b_k} p(x)\mathrm{dx} = e_k \qquad k = 1, \cdots, n \tag{14}$$

is

$$p(x) = e_k/(b_k - b_{k-1}) \qquad \text{if } x \in [b_{k-1}, b_k),\ k = 1, \cdots, n. \tag{15}$$

In Propositions 1 and 2, we used only a single prior information. In these cases, if the prior information comes from an actual probability density, the prior distribution obtained by the MEP formulation does not always coincide with the actual probability density. However, we assumed many prior informations in Proposition 3. Hence, intuitively, if the prior information comes from an actual probability density in Proposition 3, then the prior distribution (15) obtained by the MEP should be close to the actual probability density. The following Theorem 1 shows this consistency property. The proof is in the appendix.

Theorem 1. Let $F(x)$ be a distribution function with the state space $\Omega = \{[L, U]\}$ and density $p(x)$. Let $b_0 = L$, $b_k = L + K(U-L)/n$, $k = 1, 2, \cdots, n$. Let $F_n(x)$ be a sequence of distribution functions with density :

$$\begin{aligned} p_n(x) &= (F(b_k) - F(b_{k-1}))/(b_k - b_{k-1}) \qquad \text{if } x \in [b_k, b_{k-1}) \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{16}$$

The $F_n(x)$ converges to $F(x)$ in the wide sense.

In general, if there are many prior informations, the solution of the MEP formulation does not have an explicit form as the above example. We discuss this difficulty in the next section.

Ⅳ. PRIOR DISTRIBUTIONS USING THE MCEP

If there are many prior informations, then the solution of the MEP formulation, in general, does not have an explicit functional form. In this case, we have to solve a system of complicated nonlinear equations. Except for special cases, direct methods for solving a system of mulit-dimensional nonlinear equations are generally not feasible and some iteration processes are used. However, the iteration processes do not always provide a satisfactory solution [9].

An interactive method using the minimum cross-entropy principle (MCEP) can be used in this situation. Let $p_0(x)$ be an uniform distribution which implies we don't have any information. Suppose a decision maker provides the first prior information. Then find a probability distribution $p_1(x)$ using the MCEP which is "the most difficult to discriminate" from $p_0(x)$ given the first prior information. Next, the decision maker examines $p_1(x)$ whether it corresponds to his beliefs or not. He may not be satisfied with this $p_1(x)$ and provide the second prior information. Then find $p_2(x)$ using the MCEP which is "the most difficult to discriminate" from $p_1(x)$, given the second prior information. The decision maker will continue this process until he finds a satisfactory prior distribution. The interactive method for estimating a prior distribution using the MCEP can be summarized as follows :

For $j = 1, 2, \cdots$ find $p_j(x)$ which

$$\begin{aligned} &\text{minimizes} \quad \int_{\Omega} p_j(x) \ln [p_j(x) / p_{j-1}(x)] \, dx \\ &\text{subject to} \\ &\qquad \int_{\Omega} p_j(x) \, dx = 1 \\ &\qquad \text{the } j^{th} \text{ prior information} \\ &\qquad p_j(x) \geqq 0, \text{ for } x \in \Omega. \end{aligned} \tag{17}$$

The above nonlinear programming problem can be solved easily by solving a single nonlinear equation rather than solving a system of nonlinear equations. Proposition 4 and 5 illustrate the interactive method using the several different prior informations. The proofs are in the apprndix.

Proposition 4. Consider a continuous random variable, X, with state space $\Omega = \{[L, U)\}$. Suppose the j^{th} prior information is the difference of two probabilities as follows :

$$\int_{a_j}^{b_j} p_j(x)\,dx - \int_{b_j}^{c_j} p_j(x)\,dx + d_j = 0 \tag{18}$$

where $\{(a_j, b_j)\}$ and $\{(b_j, c_j)\}$ are disjoint subsets of Ω such that $L \leqq a_j \leqq b_j \leqq c_j \leqq U$, $j=1, \cdots, n$. Then the solution of the MCEP formulation is

$$\begin{aligned} p_j(x) &= p_{j-1}(x) \frac{1}{E\,F} && \text{if } x \in [a_j, b_j) \\ &= p_{j-1}(x) \frac{E}{F} && \text{if } x \in [b_j, c_j) \\ &= p_{j-1}(x) \frac{1}{F} && \text{otherwise} \end{aligned} \tag{19}$$

where $E = \dfrac{d_j(A+D) + \sqrt{d_j^2(A+D)^2 + 4(1-d_j^2)BC}}{2C(1-d_j)}$,

$F = \dfrac{(A+D) + \sqrt{d_j^2(A+D)^2 + 4(1-d_j^2)BC}}{(1-d_j^2)}$,

$A = \int_L^{a_j} p_{j-1}(x)dx, \qquad B = \int_{a_j}^{b_j} p_{j-1}(x)\,dx,$

$C = \int_{b_j}^{c_j} p_{j-1}(x)\,dx, \text{ and } D = \int_{c_j}^{U} p_{j-1}(x)\,dx.$

Proposition 5. Consider a continuous random variable, X, with state space $\Omega = \{[L, U)\}$. Let $L = b_0 < b_1 < b_2 \cdots < b_n = U$. Suppose the j^{th} prior information is the probability of sub-interval $[b_{j-1}, b_j)$. Then the solution of the MCEP formulation with constraints

$$\int_{b_{j-1}}^{b_j} p_j(x)\,dx = e_j \tag{20}$$

is a histogram distribution, i.e.,

$$p_j(x) = p_{j-1}(x)\,\frac{e_j}{A_j} \qquad \text{if } x \in [b_{j-1},\ b_j) \tag{21}$$

$$\quad = p_{j-1}(x)\,\frac{1-e_j}{1-A_j} \qquad \text{otherwise}$$

where

$$A_j = \int_{b_{j-1}}^{b_j} p_{j-1}(x)\,\mathrm{dx}.$$

Note that, in general, the solution of the interactive method using the MCEP is not the same as the solution of the MEP formulation with the same prior informations. However, under appropriate conditions, the prior distribution obtained by the interactive method converges to the actual probability density. For example, Theorem 2 shows that the solution in proposition 5 converges to the actual probability density under a sepcial condition. The proof is in the appendix.

Theorem 2. Let $F(x)$ be a distribution function with state space $\Omega = \{[L,\ U)\}$ and density $p(x)$. Let $b_0 = L,\ b_k = L + k(U-L)/n,\ e_k = F(b_k) - F(b_{k-1}),\ k = 1,\ 2,\ \cdots,\ n$, Let $F_n(x)$ be a sequence of distribution functions with density $p_n(x)$ in proposition 5. If

$$\frac{(1-e_{k+1})\,(1-e_{k+2})\cdots(1-e_n)}{(1-A_{k+1})\,(1-A_{k+2})\cdots(1-A_n)} \leqq 1 \qquad \text{for } k = 1,\ 2,\ \cdots,\ n-1, \tag{22}$$

then $F_n(x)$ converges to $F(x)$ in the wide sense.

V. CONCLUSION

The logical foundation of Bayesian decision analysis cannot be fully satisfactory until the problem of 'subjectiveness' in estimating a prior distribution is resolved. The estimation using the MEP represents one step to solve this problem. However, when there are many prior informations, the MEP formulation for estimating a prior distribution may not be solved easily. In this situation, an interactive method using the MCEP can help a decision maker to find a prior distribution which corresponds to his belief.

APPENDIX

〈Proof of Proposition 1〉

The Lagrangian equation of the problem is

$$L[p(x)] = -\int_{\Omega} p(x)\ln[p(x)]\mathrm{dx} + (\mu-1)\left(\int_{\Omega} p(x)\mathrm{dx} - 1\right) + \lambda\left(\int_{B} p(x)\mathrm{dx} - \int_{B} p(x)\mathrm{dx} + d\right). \quad \text{(A1)}$$

$\Delta L(p) = 0$ gives

$$\begin{aligned} p(x) &= \exp(-\mu-\lambda) && \text{if } x\in A \quad \text{(A2)}\\ &= \exp(-\mu+\lambda) && \text{if } x\in B\\ &= \exp(-\mu) && \text{otherwise.}\end{aligned}$$

Hence the constraints become

$$(b-a)\exp(-\mu-\lambda)+(d-c)\exp(-\mu+\lambda)+[(U-L)-(b-a)-(d-c)]\exp(-\mu)=1 \quad \text{(A3)}$$

$$(b-a)\exp(-\mu-\lambda)-(d-c)\exp(-\mu+\lambda)+d=0 \quad \text{(A4)}$$

Solving this system of nonlinear equations gives

$$\begin{aligned} p(x) &= 1/C_1C_2 && \text{if } x\in A \quad \text{(A5)}\\ &= C_1/C_2 && \text{if } x\in B\\ &= 1/C_2 && \text{otherwise}\end{aligned}$$

where

$$C_1 = \frac{d(n-n_A-n_B)+\sqrt{d^2(n-n_A-n_B)^2+4(1-d^2)n_An_B}}{2n_B(1-d)},$$

$$C_2 = \frac{(n-n_A-n_B)+\sqrt{d^2(n-n_A-n_B)^2+4(1-d^2)n_An_B}}{(1-d^2)},$$

$n=U-L$, $n_A=b-a$, and $n_B=d-c$.

〈Proof of Proposition 2〉

The Lagrangian equation of the problem is

$$L[p(x)] = -\int_{\Omega} p(x)\ln[p(x)]\mathrm{dx} + (\mu-1)\left(\int_{\Omega} p(x)\mathrm{dx} - 1\right)$$

$$+ \lambda \left(\int_{A} p(x)\mathrm{dx} - r \int_{B} p(x)\mathrm{dx} \right). \quad \text{(A6)}$$

$\Delta L(p) = 0$ gives

$$\begin{aligned} p(x) &= \exp(-\mu - \lambda) && \text{if } x \in A && \text{(A6)} \\ &= \exp(-\mu + r\lambda) && \text{if } x \in B \\ &= \exp(-\mu) && \text{otherwise.} \end{aligned}$$

Hence the constraints become

$$(b-a)\exp(-\mu-\lambda)+(d-c)\exp(-\mu+r\lambda)+[(U-L)-(b-a)-(d-c)]\exp(-\mu)=1 \quad \text{(A8)}$$
$$(b-a)\exp(-\mu-\lambda)-(d-c)\exp(-\mu+r\lambda)=0$$

Solving this system of nonlinear equations gives

$$\begin{aligned} p(x) &= 1/C_1C_2 && \text{if } x \in A && \text{(A10)} \\ &= (C_1)^r/C_2 && \text{if } x \in B \\ &= 1/C_2 && \text{otherwise} \end{aligned}$$

where

$$C_1 = [n_A/(rn_B)]^{1/(1+r)},$$
$$C_2 = n_A(rn_B/n_A)^{1/(1+r)} + n_B[n_A/(rn_B)]^{r/(1+r)} + n - n_A - n_B,$$
$$n = U - L,\ n_A = b - a, \text{ and } n_B = d - c.$$

〈Proof of Proposition 3〉

The Lagrangian equation of the problem is

$$L[p(x)] = -\int_{\Omega} p(x)\ln[p(x)]\,\mathrm{dx} = (\mu-1)\left(\int_{\Omega} p(x)\mathrm{dx} - 1\right)$$
$$+ \sum_{k=1}^{n} \lambda_k \left(\int_{b_{k-1}}^{b_k} p(x)\mathrm{dx} - e_k \right) \quad \text{(A11)}$$

$\Delta L(p) = 0$ gives

$$p(x) = \exp(-\mu-\lambda_k) \qquad \text{if } x \in (b_{k-1}, b_k),\ k=1, \cdots, n. \quad \text{(A12)}$$

Hence the constraints become

$$(b_k - b_{k-1})\exp(-\mu-\lambda_k) = e_k \qquad k=1, \cdots, n \quad \text{(A13)}$$

i. e.,

$$p(x) = e_k/(b_k - b_{k-1}) \qquad \text{if } x \in (b_{k-1}, b_k),\ k=1, \cdots, n. \quad \text{(A14)}$$

⟨Proof of Theorem 1⟩

Let $\varepsilon > 0$ and $x \in \text{cont } F = \{x : F(x) \text{ is continuous at x}\}$. Since $F(x)$ is continuous from the right, there exist $\delta > 0$ such that $y - x < \delta$ implies $F(y) - F(x) < \varepsilon$ for $y > x$. Select N such that $(U-L)/N < \delta$. For this N, $x \in [b_{k-1}, b_k)$ for some k and $b_k - b_{k-1} = (U-L)/N < \delta$. Then

$$\begin{aligned} |F_N(x) - F(x)| &= \left| F(b_{k-1}) + (x - b_{k-1}) \frac{F(b_k) - F(b_{k-1})}{b_k - b_{k-1}} - F(x) \right| \\ &\leqq \left| F(b_{k-1}) + (b_k - b_{k-1}) \frac{F(b_k) - F(b_{k-1})}{b_k - b_{k-1}} - F(x) \right| \\ &= |F(b_k) - F(x)| < \varepsilon. \end{aligned}$$

⟨Proof of Proposition 4⟩

The Lagrangian equation of the problem is

$$\begin{aligned} L[p_j(x)] = &- \int_\Omega p_j(x) \ln[p_j(x)/p_{j-1}(x)] \text{dx} + (\mu - 1) \left(\int_\Omega p_j(x) \text{dx} - 1 \right) \\ &+ \lambda \left[\int_{a_j}^{b_j} pj(x) \text{dx} - \int_{b_j}^{c_j} p_j(x) \text{dx} + d_j \right]. \end{aligned} \tag{A15}$$

$\varDelta L(p_j) = 0$ gives

$$\begin{aligned} p_j(x) &= p_{j-1}(x) \exp(-\mu - \lambda) && \text{if } x \in [a_j, b_j) \\ &= p_{j-1}(x) \exp(-\mu + \lambda) && \text{if } x \in [b_j, c_j) \\ &= p_{j-1}(x) \exp(-\mu) && \text{otherwise.} \end{aligned} \tag{A16}$$

Hence the constraints become

$$\begin{aligned} &\int_\Omega p_j(x) \text{dx} = 1 \\ &\int_{a_j}^{b_j} p_{j-1}(x) \exp(-\mu - \lambda) \text{dx} - \int_{b_j}^{c_j} p_{j-1}(x) \exp(-\mu + \lambda) \text{dx} + d_j = 0. \end{aligned} \tag{A17}$$

Solving this nonlinear system of equations gives

$$\begin{aligned} p_j(x) &= p_{j-1}(x) \frac{1}{E\ F} && \text{if } x \in [a_j, b_j) \\ &= p_{j-1}(x) \frac{E}{F} && \text{if } x \in [b_j, c_j) \\ &= p_{j-1}(x) \frac{1}{F} && \text{otherwise} \end{aligned} \tag{A18}$$

where

$$E = \frac{d_j(A-D) + \sqrt{d_j^2(A+D)^2 - 4(1-d_j^2)BC}}{2C(1-d_j)},$$

$$F = \frac{(A+D) - \sqrt{d_j^2(A-D)^2 - 4(1-d_j^2)BC}}{(1-d_j^2)},$$

$$A = \int_{L}^{a_j} p_{j-1}(x)\mathrm{dx}, \quad B = \int_{a_j}^{b_j} p_{j-1}(x)\,\mathrm{dx},$$

$$C = \int_{b_j}^{c_j} p_{j-1}(x)\,\mathrm{dx}, \text{ and } \quad D = \int_{c_j}^{U} p_{j-1}(x)\,\mathrm{dx}.$$

〈Proof of Proposition 5〉

The Lagrangian equation of the problem is

$$L[p_j(x)] = -\int_{\Omega} p_j(x)\ \ln[p_j(x)/p_{j-1}(x)]\,\mathrm{dx} + (\mu - 1)\left(\int_{\Omega} p_j(x)\,\mathrm{dx} - 1\right) + \lambda\left[\int_{b_{j-1}}^{b_j} p_j(x)\,\mathrm{dx} - e_j\right] \tag{A19}$$

$\Delta L(p_j) = 0$ gives

$$\begin{aligned} p_j(x) &= p_{j-1}(x)\exp(-\mu-\lambda) && \text{if } x \in [b_{j-1},\ b_j) \\ &= p_{j-1}(x)\exp(-\mu) && \text{otherwise} \end{aligned}$$

Hence the constraints become

$$\int_{\Omega} p_j(x)\mathrm{dx} = 1 \tag{A20}$$

$$\int_{b_{j-1}}^{b_j} p_{j-1}(x)\exp(-\mu-\lambda)\mathrm{dx} = e_j.$$

Solving this nonlinear system of equations gives

$$\begin{aligned} p_j(x) &= p_{j-1}(x)\frac{e_j}{A_j} && \text{if } x \in [b_{j-1},\ b_j) \\ &= p_{j-1}(x)\frac{1-e_j}{1-A_j} && \text{otherwise} \end{aligned} \tag{A21}$$

where

$$A_j = \int_{b_{j-1}}^{b_j} p_{j-1}(x)\,\mathrm{dx}.$$

⟨Proof of Theorem 2⟩

Let $\varepsilon > 0$ and $x \in \text{cont } F = \{x : F(x) \text{ is continuous at } x\}$. Since $F(x)$ is continuous from the right, there exist $\delta > 0$ such that $y - x < \delta$ implies $F(y) - F(x) < \varepsilon$ for $y > x$. Select N such thhat $d = (U - L)/N < \delta$. For this N, $x \in [b_{k-1}, b_k)$ for some k and $b_k - b_{k-1} = (U - L)/N < \delta$. Note that $p_N(x)$ can also written as

$$p_N(x) = \frac{e_k(1 - e_{k+1})(1 - e_{k+2}) \cdots (1 - e_N)}{d(1 - A_{k+1})(1 - A_{k+2}) \cdots (1 - A_N)} \quad \text{if } x \in [b_{k-1}, b_k] \; k = 1, \cdots, n \quad \text{(A22)}$$

Then

$$\begin{aligned} |F_N(x) - F(x)| &= \left| F(b_{k-1}) + (x - b_{k-1}) \frac{e_k(1 - e_{k+1})(1 - e_{k+2}) \cdots (1 - e_N)}{d(1 - A_{k+1})(1 - A_{k+2}) \cdots (1 - A_N)} - F(x) \right| \\ &\leqq \left| F(b_{k-1}) + d \frac{F(b_k) - F(b_{k-1})}{d} - F(x) \right| \\ &= |F(b_k) - F(x)| < \varepsilon. \end{aligned}$$

〈REFERENCES〉

[1] J. M. Cozzolino, and M. J. Zahner, "The maximum-entropy distribution of the future market price of a stock", *Operations research,* vol. 21, pp. 1200−1211, 1973.

[2] E. T. Jaynes, "Information theory and statistical mechanics I", *Phys. Rev.,* vol. 106, pp. 620−630, 1957.

[3] E. T. Jaynes, "Information theory and statistical mechanics II", *Phys. Rev.,* vol. 108, pp. 171−190, 1957.

[4] E. T. Jaynes, "Prior probabilities", *IEEE Trans. On Systems Science and Cybernetics,,* SSC-4, vol. 3, pp. 227−241, 1968.

[5] E. T. Jaynes, "Where do we go from here", *Maximum Entropy and Bayesian Methods in Inverse Problems,* pp. 21−58, C. R. Smith and W. T. Grandy (eds), D. Reidel Publishing Company, 1985.

[6] R. W. Johnson, "Comments on 'Prior probability and uncertainty'", *IEEE Trans. Inform. Theory,* vol. IT-25, pp. 129−132, 1979.

[7] R. L. Kashyap, "Prior probability and uncertainty," *IEEE Trans. Inform. Theory,* vol. IT -17, pp. 641−650, 1971.

[8] S. Kullback, *Information Theory and Statistics,* New York : Wiley, 1969.

[9] W. C. Rheinboldt, *Methods for Solving Systems of Nonlinear Equations,* Society for Industrial and Applied Mathematics, 1987.

[10] C. E. Shannon, "A mathematical theory of communication", *Bell System Technical Journal,* vol. 27, pp. 379−432, pp. 623−656, 1948.

[11] J. E. Shore, and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy", *IEEE Trans. Inform. Theory,* vol. IT-26, pp. 26−37, 1980.

[12] J. E. Shore, and R. W. Johnson, "Properties of cross-entropy minimization", *IEEE Trans. Inform. Theory,* vol. IT-27, pp. 472−482, 1981.

[13] M. Tribus, "Thirty years of information theory", *The Maximum Entropy Formalism,* pp. 1−14, R. D. Levine and M. Tribus (eds). MIT Press, 1978.

엔트로피 이론을 이용한 사전 확률 분포함수의 추정

이 정진*, 신 완선**

〈요 약〉

베이시안 결정론에서 사전 확률 분포함수는 표본을 추출하기 이전에 추정하여야 한다. 대개는 분포함수군을 먼저 선택한 후, 그 중 하나를 결정자의 경험을 통하여 선택한다. 이러한 주관적인 사전 확률 분포함수의 선택방법이 베이시안 결정론에 대한 주요비판이 항상 되어왔다.

본 논문에서는 최대 엔트로피 이론을 이용하여 우리 주변의 의사결정에 많이 이용되는 정보들에 관한 객관적인 사전 확률 분포함수들을 구하였다. 그 결과는 히스토그램 형태의 분포함수가 된다. 그러나 사전 정보가 많을 경우에는 최대 엔트로피 모형의 해를 구하기 위하여 복잡한 비선형 연립방정식을 풀어야 하는데, 구체적인 형태의 함수를 구하지 못하는 경우가 대부분이다. 이 때에는 최소의 크로스 엔트로피 모형을 이용하여 사전확률 분포함수를 구하는 것이 편리하다. 그 밖에 엔트로피 이론으로 구한 사전확률 분포함수의 확률적 수렴성을 증명하였다.

* 숭실대학교 통계학과, 156 743 서울시 동작구 상도 1동 1 1.
** 미시시피 주립대학 산업공학과, 미국