应用概率统计 第七卷
第二期 1991年5月

Chinese Journal of Applied Probability
and Statistics Vol. 7. No. 2. May. 1991

SOME MAXIMAL INFORMATION AND GENERALIZED MAXIMAL ENTROPY PRIORS

ZHANG YAOTING　　YAO QIWEI
(*Wuhan University*)　　(*Southeast University*)

Abstract

This paper is concerned with there procedures to derive prior distributions with knowledge of parameters in terms. By means of the entropy inequality, we simplify the proof of Zellner (1984) and get the uniqueness for the maximal data information prior. A generalized maximal entropy prior is proposed, which improves the classical maximal entropy principle in some respects. An intermediate solution for the maximal relative prior is developed, from which the maximal relative prior densities for a great number of distribution families are presented.

§1. Introduction

In Bayesian statistics, it is often desirable to have a posterior distribution to reflect the information in a given set of data. To achieve this objective, it is necessary to employ a prior distribution that adds little information from the sample. Much work has been done to provide procedures for formulating such prior distributions. For example, Zellner (1984) presents the maximal data information prior (MDIP) which bears the idea that it provides maximal prior average data information, with the information being represented by Shannon's entropy. In the present paper, Zellner's [1] main result is more simply proved by means of the entropy inequality and we find that if MDIP exists, it must be unique. Furthermore, we present two tentative rules for the choice of priors——the generalized maximal entropy prior (GMEP) and the maximal relative prior (MRP). In comparison with the classical maximal entropy principle (cf. Kullback 1959), GMEP pays specioial attention to the obvious connection between the sample and the parameters in every specific statistical problem. Very similar to MDIP, GMEP can be derived easily. MRP is just Lindley's rule (cf. Lindley 1961). It seems to us that only result on it in the earlier literature is that its asymptotic approximation was Jeffrey's rule (cf. [3],[4]). Since in Bayesian approach, the

本文1989年5月24日收到.

large sample problem is not sensitive to the choice of prior distribution (cf [3]; [4]), it is of significance to study MRP for the small size sample. Although it is hard to derive a MRP distribution for a general model, we develop an intermediate solution, from which we get the MRP's for a great number of distribution families, for example, location parameters family, scale parameters family, etc. Much different from MDIP and GMEP, MRP is invariant to any differentiable $1-1$ transformation either for samples or for parameters, which also leads in an indirect way to derive MRP for some special distribution families, especially we get the MRP for Weibull distributions. In deriving MDIP and GMEP, only proper distribution densities can be chosen (cf. Theorem 1, 2). For this reason, Zellner (1984) introduces an artificial restriction which keeps the parameter space bounded. However the MRP densities may be improper (cf. Theorem 3). Since improper prior seems necessary for some common distribution families (for example, normal distribution), this restriction would bring some convenience in practice.

The plan of the paper is as follows. Section 2 discusses MDIP. Some results on GMEP are reported in Section 3. Section 4 deals with the MRP. Finally, we list the above three kinds of prior densities for some wellknown distribution families in a table.

Before ending this section, we explain some notations to be used in this paper Suppose $\{p(y|\theta),\ \theta\in\Theta\}$ is a distribution family, y is a sample random variable (or vecotr) which distributed, for some $\theta\in\Theta$, according to the probability density $p(y|\theta)$. with respect to a Lebesgue measure, and Θ is the parameter space which is a subset of some Euclidean space. We always assume that $\pi^*(\theta)$, $\pi_*(\theta)$, and $\pi_r(\theta)$ indicate the MDIP, GMEP, and MRP densities (with respect to Lebesgue measure) on Θ respectively, and $\pi(\theta)$ indicate any prior density. Consequently, $\pi_*(\theta|y)$, $\pi_*(\theta|y)$, $\pi_r(\theta|y)$ or $\pi(\theta|y)$ and $p^*(y)$, $p_*(y)$, $p_r(y)$ or $p(y)$ indicate corresponding posterior densities and marginal densities. For simplicity in notation, we shall not generally attempt to be specific in describing the density functions. Thus, $p(x)$ denotes the density for random quantity x and $p(y)$ that of y without any suggestion that x and y have the same distributions. All densities we mention are proper (i. e. its integration on total space equals 1) unless declared specifically.

§ 2. Maximal Data Information Priors.

Assume θ has the density $\pi(\theta)$, its Shannon's entropy is defined to be

$$H(\theta)=-\int\pi(\theta)\log\pi(\theta)d\theta.$$

Similarly, the conditional entropy of y relative to θ is

$$H(y|\theta)=-\int p(y|\theta)\log p(y|\theta)dy \tag{1}$$

Zellner (1984) set $-H(\theta)$ as the information measure in $\pi(\theta)$, and $-H(y|\theta)$ as the information in the data distribution $p(y|\theta)$. Then the prior average information in $p(y|\theta)$ is

$$\bar{I}=-\int H(y|\theta)\pi(\theta)d\theta=\int\pi(\theta)d\theta\int p(y|\theta)\log p(y|\theta)dy.$$

Consequently, this average information minus the information in the prior $\pi(\theta)$ is

$$\begin{aligned}G_1(\pi)&=-E[H(y|\theta)]+H(\theta)\\&=\int\pi(\theta)d\theta\int p(y|\theta)\log p(y|\theta)\,dy-\int\pi(\theta)\log\pi(\theta)d\theta.\end{aligned}\tag{2}$$

Definition 1. (Zellner 1984). A MDIP is a density function on Θ which maximizes $G_1(\pi)$, defined in (2).

Theorem 1. (Zellner 1984). If there exists a constant c such that

$$\pi^*(\theta)=c\cdot\exp\{-H(y|\theta)\}\tag{3}$$

is a probability density function, $H(y|\theta)$ is defined in (1), then $\pi^*(\theta)$ is the unique MDIP.

Zellner [1] proved Theorem 1, however didn't get the uniqueness. By means of the well-known entropy inequality (Lemma 1, its proof is on page 15 of [2]), Zellner's proof can be simplified and the uniqueness follows subsequently.

Lemma 1. (Entropy Inequality). Assume $p_1(x)$ and $p_2(x)$ are two probability densities on R^n. Then

$$\int p_1(x)\log\frac{p_1(x)}{p_2(x)}\,dx\geqslant 0$$

with equality if and only if $p_1=p_2$ a.s..

Proof of Theorem 1. From (3) $\log\pi^*(\theta)=\log c-H(y|\theta)$. For any density $\pi(\theta)$, via Lemma 1,

$$\begin{aligned}G_1(\pi)&=\int\pi(\theta)[\log\pi^*(\theta)-\log c-\log\pi(\theta)]d\theta\\&=-\log c-\int\pi(\theta)\log\frac{\pi(\theta)}{\pi^*(\theta)}d\theta\leqslant-\log c\end{aligned}$$

with equality if and only if $\pi=\pi^*$ a.s.. Hence

$$G_1(\pi)\leqslant-\log c=G_1(\pi^*)$$

i.e. π^* is the unique MDIP.

Some MDIP densities are presented in Table 1, which can be got via (3) easily.

§ 3. Generalized Maximal Entropy Prirs

Edwin Jaynes said in his letter to A. Zellner: MDIP can be recognized as arestricted maximum point, i. e. under the restriction $E[H(y|\theta)]=c$, choose $\pi(\theta)$ such that $H(\theta)$ attains its maximum.In this view, MDIP procedure also can be understood as a special kind of maximal entropy principle. In fact, MDIP is only ruled by $p(y|\theta)$, and dosen't depend on the sample y, and $p(y|\theta)$ represents the relationship between y and θ, which is explict for any parameter statistical problem. The maximal entropy principle chooes $\pi(\theta)$ which maximizes $H(\theta)$ (i. e. θ has the maximal uncertainty), yet pays no attention to the $p(y|\theta)$. We believe it is reasonable that for different $p(y|\theta)$, θ has different rules and effects. Hence it is more reasonable we should choose $\pi(\theta)$ such that contingent entropy $H(y, \theta)$ attains its maximum. This $\pi(\theta)$ is called GMEP.

Now the joint density of (y, θ) is $\pi(\theta)p(y|\theta)$, so

$$H(y, \theta)=H(\theta, y)=-\iint\pi(\theta)p(y|\theta)\log[\pi(\theta)p(y|\theta)]dy\,d\theta$$
$$=-\int\pi(\theta)d\theta\int p(y|\theta)\log p(y|\theta)dy-\int\pi(\theta)\log\pi(\theta)d\theta$$
$$=E[H(y|\theta)]+H(\theta).$$

To stress the effect of prior, let

$$G_2(\pi)=E[H(y|\theta)]+H(\theta). \tag{4}$$

Definition 2. A GMEP is a density function on Θ which maximizes $G_2(\pi)$, defined in (4).

To compare (4) and (2), only difference between $G_2(\pi)$ and $G_1(\pi)$ is the sign before $E[H(y|\theta)]$. From the proof of Theorem 1, Theorem 2 follows immediately.

Theorem 2. If there exists a constant c such that

$$\pi_*(\theta)=c\cdot\exp\{H(y|\theta)\} \tag{5}$$

is a probability density function, where $H(y|\theta)$ is in (1), then $\pi_*(\theta)$ is the unique GMEP.

Corollary 1. If $z=Ay+b$, A is a non-random non-singular matrix and b is a constant vector. Then the GMEP for θ associated with $p(y|\theta)$ is identical to that with $p(z|\theta)$.

Corollary 1 indicates that GMEP dosen't depend on the unit of measurement for samples. This property also holds for MDIP (cf. [1] Theorem 2). The proof of Corollary 1 is trivial. (5) tells us that GMEP can be got from MIDP without any more calculation (cf. Table 1).

§ 4. Maximal Relative Priors

According to Zellner's (1984) view, $G_1(\pi)=-E[H(y|\theta)]-[-H(\theta)]$ which is the difference of the two information quantities: $-H(\theta)$ is the information for θ in $\pi(\theta)$, and $-E[H(y|\theta)]$ is the prior average information for y in $p(y|\theta)$. In other words, $G_1(\pi)=H(\theta)-E[H(y|\theta)]$, which is the decrement in entropy, but $H(\theta)$ is the entropy for θ and $E[H(y|\theta)]$ is the prior average entropy for y. Naturally, perhaps we hould consider the decrement in the entropy for θ after sampling y, which is

$$G_3(\pi)=H(\theta)-E[H(\theta|y)]. \tag{6}$$

Sometimes $G_3(\pi)$ is called the information for parameter θ in sample y (cf. [2]). Now $p(y)=\int p(y|\theta)\pi(\theta)d\theta$, $\pi(\theta|y)=\pi(\theta)p(y|\theta)/p(y)$, thus

$$G_3(\pi)=\int p(y)dy\int\pi(|\theta|y)\log\pi(\theta|y)d\theta-\int\pi(\theta)\log\pi(\theta)d\theta$$
$$=\int\pi\ (\theta)de\int p(y|\theta)\log p(y|\theta)dy-\int p(y)\log p(y)dy$$
$$=\iint\pi(\theta)p(y|\theta)\log\frac{p(y|\theta)}{p(y)}dy\,d\theta$$

which is just Lindley information (cf. [3]) and it is a suitable measurement for the relativity between y and θ.

Definition 3. A MRP is a density function on Θ which maximizes $G_3(\pi)$, defined in (6).

It seems very hard to get an explict expression for MRP like Theorem 1 or Theorem 2. Theorem 3 presents an intermediate solution for MRP. On the other hand, the MRP may be improper, i.e., ite integrand on parameter may be infinite, this is much different from MDIP or GMEP.

Theorem 3. If there exists a prior $\pi_r(\theta)$ (may be improper) for which

$$\int p(y|\theta)\log\frac{p(y|\theta)}{p_r(y)}dy=c \tag{7}$$

where $p_r(y)=\int\pi_r(\theta)p(y|\theta)d\theta$, c is a positive constant. Then for any proper prior density π,

$$G_3(\pi_r)\geqslant G_3(\pi) \tag{8}$$

with equality if and only if

$$p_r(y)=p(y)=\int\pi(\theta)p(y|\theta)d\theta\quad\text{a.s.}.$$

Proof. If π_r is improper, $G_3(\pi_r)=c\int\pi_r(\theta)d\theta=\infty$, (8) holds obviously. For proper π_r, $G_3(\pi_r)=c$, hence if $\pi(\theta)$ is a proper density,

$$G_3(\pi_r)-G_3(\pi)=\int c\pi(\theta)d\theta-G_3(\pi)$$
$$=\iint\pi(\theta)p(y|\theta)\log\frac{p(y|\theta)}{p_r(y)}dy\,d\theta-\iint\pi(\theta)p(y|\theta)\log\frac{p(y|\theta)}{p(y)}dy\,d\theta$$
$$=\iint\pi(\theta)p(y|\theta)\log\frac{p(y)}{p_r(y)}dy\,d\theta$$
$$=\int p(y)\log\frac{p(y)}{p_r(y)}dy$$

Lemma 1 insures the right hand side of above expression non negative and equals zero if and only if $p_r(y)=p(y)$ a.s., Hence, it completes the proof.

Example 1. Let $p(y|\theta)=p(y-\theta)$, $-\infty<y,\ \theta<\infty$. Then $\pi_r(\theta)\propto 1$. Now

$$p_r(y)=\int\pi_r(\theta)p(y-\theta)d\theta\propto 1,$$

and $\int p(y-\theta)\log p(y-\theta)dy=c$, hence (7) holds.

Example 2. For

$$p(y|\theta)=\frac{1}{\theta}p\left(\frac{y}{\theta}\right),\ \theta>0,\ \ \pi_r(\theta)\propto\frac{1}{\theta}.$$

simple integration shows $p_r(y)\propto 1/|y|$, hence

$$\int p(y|\theta)\log\frac{p(y|\theta)}{p_r(y)}dy=\int\frac{1}{\theta}p\left(\frac{y}{\theta}\right)\log\frac{p(y/\theta)}{\theta/|y|}dy$$
$$=\int p(t)\log[|t|p(t)]dt=c.$$

Now we treat Example 2 in another way. For simplicity in argument, assume $y>0$. Let $y=e^z$, $\theta=e^u$, then

$$p(z|u)=e^{-u}p(z^{z-u})\left|\frac{dy}{dz}\right|=p(z^{z-u})\triangleq p(z-u).$$

From Example 1, $\pi_r(u)\propto 1$. Thus

$$\pi_r(\theta)=\pi_r(u(\theta))\left|\frac{du}{d\theta}\right|\propto\frac{1}{\theta},$$

which is the same as that we derive directly for θ just now. It implies some invariance. In fact, such invariance holds for MRP generally, which is presented in Corollary 2.

Corollary 2. Assume that $\{p(y|\theta),\ \theta\in\Theta\}$ is a distribution family, $\pi_r(\theta)$ is its MRP. If $y=y(z)$ and $\theta=\theta(u)$ are two differentiable $1-1$ transformations for samples and parameters respectively, and $p(z|u)$, $u\in U=\{u:\ \theta(u)\in\Theta\}$, is the induced distribution density, $\pi_r(u)$ is the MRP for u associated with $p(z|u)$. Then

$$\pi_r(u)=\pi_r(\theta(u))J(u) \tag{9}$$

whene $J(u)$ is the Jacobian of the transformation $\theta=\theta(u)$.

Proof. We only need to show G_3 is invariant, i.e., for any density $\pi(\theta)$

$$\iint p(z|u)\pi(u)\log\frac{p(z|u)}{p(z)}dz\,du=\iint p(y|\theta)\pi(\theta)\log\frac{p(y|\theta)}{p(y)}dy\,d\theta. \tag{10}$$

Now $p(z|u)=p(y(z)|\theta(u))J(z)$, $\pi(u)=\pi(\theta(u))J(u)$, and $p(z)=\int p(z|u)\pi(u)du=\int p(y(z)|\theta(u))J(z)\pi(\theta(u))J(u))du=J(z)\int p(y(z)|\theta)\pi(\theta)d\theta=J(z)p(y(z))$, hence the left hand side of (10)

$$=\iint p(y(z)|\theta(u))\pi(\theta(u))\log\left[\frac{p(y(z)|\theta(u))J(z)}{p(y(z))J(z)}\right]J(z)J(u)dz\,du$$
$$=\iint p(y|\theta)\pi(\theta)\log\frac{p(y|\theta)}{p(y)}dy\,d\theta=\text{The right hand side of (10)}.$$

In fact, we needn't confine specifically to the densities $p(y|\theta)$ and $\pi(\theta)$, with respect to Lebesgue measure in the discussion for MRP since G_3 dosen't depend on the dominating measures. It can be shown easily that MRP is also invariant for change of the dominating measures. Unfortunately, this property fails for MDIP and GMEP. Moreover the invariance presented in Corollary 2 also fails for MDIP and GMEP in general.

Example 3. $p(y|\theta)=p(y-g(\theta))$, g is a differentiable 1−1 function. For this distribution, it is not explicit how to derive $\pi_r(\theta)$ by (7) directly. Howerer, (9) offers an indirect way to do it. Let $u=g(\theta)$, then $p(y|u)=p(y-u)$, hence $\pi_r(u)\propto 1$. From (9), $\pi_r(\theta)\propto \left|\frac{dg}{d\theta}\right|$.

The situation is more interesting, however, if the parameters can be derived into two parts, namely the conditional density of sample is $p(y|\theta, \varphi)$. In this case, $p(y)$ may equal to infinite when π is improper, and (7) is impossible consequently. We try to untilize Lindley's (1961) ideas to treat this question. Assume $\pi(\theta, \varphi)$ is a prior, then

$$\pi(\theta)=\int \pi(\theta, \varphi)d\varphi$$

$$\pi(\theta|y)=\int \pi(\theta, \varphi|y)\,d\varphi=\int \frac{p(y|\theta, \varphi)\pi(\theta, \varphi)}{p(y)}d\varphi$$

If we can say that y gives no information about θ, it should imply $\pi(\theta)=\pi(\theta|y)$. Hence

$$p(y)=\int \frac{p(y|\theta, \varphi)\pi(\theta, \varphi)}{\pi(\theta|y)}\,d\varphi=\int \frac{p(y|\theta, \varphi)\pi(\theta)\pi(\varphi|\theta)}{\pi(\theta)}\,d\varphi$$

$$=\int p(y|\theta, \varphi)\pi(\varphi|\theta)d\varphi. \tag{11}$$

In this case, if there exists $\pi(\theta, \varphi)=\pi(\theta)\pi(\varphi|\theta)$ for which

$$\int p(y|\theta, \varphi)\log \frac{p(y|\theta, \varphi)}{p(y)}\,dy=c \tag{12}$$

where $p(y)$ is defined in (11), then Theorem 3 implies that this $\pi(\theta, \varphi)$ maximizes $G_3(\pi)$ over the set $\mathscr{P}=\{\pi(\theta, \varphi)\colon \pi(\theta)=\pi(\theta|y)\}$, we call it a generalized MRP (GMRP).

Example 4. For normal distribution $p(y|u, \sigma^2)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2\sigma^2}(y-u)^2\right\}$ almost all objective Bayesian procedures approves the marginal prior for mean should be $\pi(u)\propto 1$ and that for variance be $\pi(\sigma)\propto\frac{1}{\sigma}$. From the intuition, a single normal variable gives no information on variance if the mean is uniformly distributed over the whole real line. Thus we may assume $\pi(\sigma)=\pi(\sigma|y)$. Furthermore, it is easy to check that $\pi(u|\sigma)\propto 1$ satisfies (12). It means that GMRP is $\pi(u, \sigma)\propto\frac{1}{\sigma}$.

More general for $p(y|u, \sigma)=p\left(\frac{y-u}{\sigma}\right)$, similar argument leads to GMRP$\propto\frac{1}{\sigma}$, which seems not reasonable, but coincides with MDIP.

Example 5. For Weibull distribution $p(y|\alpha, \eta)=\frac{\alpha}{\eta^\alpha}y^{\alpha-1}\exp\left(-\left(\frac{y}{\eta}\right)^\alpha\right)$ for $y>0$, with parameters $\alpha>0$, $\eta>0$, let $y=e^z$, $\alpha=\frac{1}{\sigma}$ and $\eta=e^u$, then

Table 1

$p(y\|\theta)$	MDIP $\pi^*(\theta)$	GMEP $\pi_*(\theta)$	MRP $\pi_r(\theta)$
1. Normal			
1a. $N(u, 1)$	$\propto 1$	$\propto 1$	$\propto 1$
1b. $N(u, \sigma^2)$	$\propto 1/\sigma$	$\propto \sigma$	$\propto 1/\sigma$
1c. $N(u, \sigma^2)$	$\propto 1/\sigma$	$\propto \sigma$	$(\propto 1/\sigma)$
2. Location parameter			
2a. $p(y-\theta)$	$\propto 1$	$\propto 1$	$\propto 1$
2b. $p(y-g(\theta))$			$\propto \left\|\frac{dg}{d\theta}\right\|$
3. Scale parameter			
3a. $\frac{1}{\theta} p(y\|\theta)$	$\propto 1/\theta$	$\propto \theta$	$\propto 1/\theta$
3b. $\left(\prod_{i=1}^{n} \theta_i\right)^{-1} p\left(\frac{y_1}{\theta_1}, \cdots, \frac{y_n}{\theta_n}\right)$	$1/\prod_{i=1}^{n} \theta_i$	$\propto \prod_{i=1}^{n} \theta_i$	$\propto 1/\prod_{i=1}^{n} \theta_i$
3c. $\frac{1}{\|\Sigma\|} p(\Sigma^{-1} y)$	$\propto \|\Sigma\|^{-1}$	$\propto \|\Sigma\|$	
3d. $\frac{1}{g(\theta)} p(y/g(\theta))$			$\propto \frac{1}{g(\theta)} \left\|\frac{dg}{d\theta}\right\|$
4. Location-Scale parameter			
4a. $\frac{1}{\sigma} p\left(\frac{y-u}{\sigma}\right)$	$\propto 1/\sigma$	$\propto \sigma$	$(\propto 1/\sigma)$
4b. $\frac{1}{\|\Sigma\|} p(\Sigma^{-1}(y-u))$	$\propto \|\Sigma\|^{-1}$	$\|\Sigma\|$	
5. Uniform distribution			
$U(0, \theta)$	$\propto 1/\theta$	$\propto \theta$	$\propto 1/\theta$
6. Exponential			
$\frac{1}{\theta} \exp(-y/\theta)$	$\propto 1/\theta$	$\propto \theta$	$\propto 1/\theta$
7. Log-normal			
$\frac{1}{\sigma y \sqrt{2\pi}} \exp\{-(\log y - \log\theta)^2/(2\sigma^2)\}$	$\propto \frac{1}{\theta\sigma}$	$\propto \theta\sigma$	$\left(\propto \frac{1}{\theta\sigma}\right)$
8. Exponential family $\exp\{f(\theta)k(y)+s(y)+q(\theta)\}$	$\propto \exp(f(\theta)\bar{k}+\bar{s}+q(\theta))$	$\propto \exp[-\{f(\theta)\bar{k}+\bar{s}+q(\theta)\}]$	
9. Weibull distribution			
9a. $\alpha y^{\alpha-1} \exp(-y^\alpha)$			$\propto 1/\alpha$
9b. $\frac{a}{\eta^a} y^{a-1} \exp[-(y/\eta)]^a$ (a is a constant)			$\propto 1/\eta$
9c $\frac{\alpha}{\eta^\alpha} y^{\alpha-1} \exp(-(y/\eta)^\alpha))$			$\left(\propto \frac{1}{\alpha\eta}\right)$

$$p(z|u, \sigma) = \frac{1}{\sigma} \exp\left(\frac{x-u}{\sigma}\right) \exp(-e^{(x-u)/\sigma}),$$

which is a location-scale parameters family. It follows Corollary 2 and Example 1, 2 that $\pi_r(\alpha)\propto\frac{1}{\alpha}$ if η is given, and $\pi_r(\eta)\propto\frac{1}{\eta}$ if α is known. Furthermore, if both α and η are unknown, the GMRP $\propto\frac{1}{\alpha\eta}$.

§ 5. The Priors For Some Distribution Families

We list some MDIP, GMEP, MRP (or GMRP) densities in Table 1.

Because both of MDIP and GMEP must be proper, we have to put some restriction in calculation. For example, for normal distribution in line 1a of Table 1, we should confine $u\in[-a, b]$ for some $0<a, b<\infty$. In practice, one can choose a and b sufficiently large empirically, then $\pi^*(u)=\pi_*(u)=\frac{1}{a+b}I^{(-a\leq u\leq b)}$.

In Table 1, the column of MRP contains some parentheses. These parenthetical entries are not real MRP. They are GMRP merely, which can be derived by arguments similar to that in Example 4.

All g appeared in Table 1 are differentiable $1-1$ functions. Furthermore, it is positive in line 3d.

In the Exponential family line of Table 1,

$$\bar{k}=\int k(y)p(y|\theta)dy, \quad \bar{s}=\int s(y)p(y|\theta)dy.$$

References

[1] Zellner, A., Maximal data information prior distributions. *Basic Issues in Econometrics*, 201—215. University of Chicago Press. 1984.

[2] Kullback, S., Information Theorey and Statistics. John Wiley & Sons, New York. 1959.

[3] Lindley, D. V., The use of prior probability distributions in statistical inference and decisions. *Proc. 4th Berkeley Symposium on Math. Statist. and Probab.*, 1 (1961), 436—468.

[4] Box, G. E. P., Tiao, G. C., *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading. 1973.

一些最大信息先验分布与广义最大熵先验分布

张尧庭　　姚琦伟

(武汉大学)　(东南大学)

本文讨论了三种确定无信息先验分布的方法. 利用熵不等式, 简化了 Zellner(1984)关于最大数据信息先验分布的一条定理的证明, 并得到了其唯一性. 在此基础上提出了广义最大熵先验分布, 它在某些方面改进了经典的最大熵原理, 对于最大相关先验分布(即 Lindley 准则), 我们得到了一个中间解, 由此导出了许多常见分布族的最大相关先验分布.