NOTE ON THE BAYES APPROACH TO A GROUPING OF SMALL EVENTS

Etsuo Kumagai and Nobuo Inagaki

Received February 21, 2003

Abstract. By using the digamma function we extend the hyper-parameter of the prior Dirichlet distribution from positive integers to positive real numbers in the Bayes approach to a grouping of small events in the multinomial distribution by Kumagai and Inagaki(2000).

1 Introduction Lindley(1964) studied an approximation of the posterior distribution for a particular prior distribution in the multinomial distribution, and Block and Watson(1967) suggested an improved correction procedure for it. Connor and Mosimann(1969) and Antelman(1972) studied several kinds of Dirichlet distributions as the prior distribution. For the details of the Dirichlet distribution, see Wilks(1962, page 177).

In the goodness-of-fit test in the multinomial distribution, it is well known that we have to combine small events with less than five frequencies of observations into a grouped event with greater than or equal to five in order that the chi-square statistic has better approximation to the chi-square distribution. For the grouping of small events, Kumagai and Inagaki(2000) investigated the exact formulae of two errors, that is, the modeling error and the estimation error (see Inagaki(1977) and Gruber(1998, page 48)), from the viewpoint of the decision theoretical framework by the Bayes approach that the prior distribution is the Dirichlet distribution $D_{iri}(\nu_1, \cdots, \nu_k\,;\,\nu_{k+1})$ whose the parameters $\{\nu_j\}$ are all positive integers.

Our aim is to replace integers with positive real numbers in the above assumption with respect to the hyper-parameters , and to obtain some generalized results with respect to the decision theoretical framework.

2 Extension from integer to real positive number Preliminary to the extension of the results by Kumagai and Inagaki(2000), we show several notations and their extensions.

For any positive integer n, let $L(n)$ be a partial sum of harmonic series, that is, $L(n) = \sum_{j=1}^{n} 1/j$, and $\gamma(n)$ the difference $L(n) - \log(n)$, that is, $\gamma(n) = \sum_{j=1}^{n} 1/j - \log n$, which is called the Euler's sequence whose properties are known:

(1) $\gamma(n)$ is a monotone decreasing series and converges to the Euler constant γ, that is,

$$\lim_{n\to\infty} \gamma(n) \;=\; \gamma \;=\; 0.577215\cdots,$$

(2) For a large positive integer n, the Euler's sequence $\gamma(n)$ has the following expansion:

$$\gamma(n) \;=\; \gamma + 1/(2n) - 1/(12n^2) + O(n^{-4}).$$

2000 *Mathematics Subject Classification.* 62F15 .

Key words and phrases. Bayes approach, Dirichlet distribution, prior, posterior, digamma function, multinomial distribution.

In the extension of $L(n)$ from an integer n to a real positive number x, a simple exchange of n for a real positive number x is not always adequate because the definition of $L(n)$ is a summation by the inverse of integer. Here we need the properties of the digamma function for a positive real number s:

$$\psi(s) \;=\; \frac{d}{ds}\log(\Gamma(s)) \;=\; \frac{\Gamma'(s)}{\Gamma(s)},$$

where $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}dx$ is the gamma function and $\Gamma'(s)$ is its differentiation. For arbitrary positive real numbers s and t, it is known that

$$\Gamma'(s)\,\Gamma(t) = \Gamma'(s+t)\,B(s,t) + \Gamma(s+t)\cdot\int_0^1 \log x\cdot x^{s-1}(1-x)^{t-1}\,dx,$$

where $B(s,t) = \Gamma(s)\Gamma(t)/\Gamma(s+t)$ is the beta function, so that we obtain the relationship

$$\psi(s) - \psi(s+t) \;=\; \int_0^1 \frac{1}{B(s,t)}\log x\cdot x^{s-1}(1-x)^{t-1}\,dx\ . \tag{2.1}$$

Let $L^*(x)$ be an extension of $L(n)$ as follows:

$$L^*(x) \;=\; \sum_{j=0}^{[x]-1}\frac{1}{x-j} + \psi(x-[x]+1) - \psi(1), \qquad (\forall x\in(0,\infty)), \tag{2.2}$$

where the symbol $[x]$ means the maximum integer less than or equal to x. And an extension $\gamma^*(x)$ is defined by

$$\gamma^*(x) \;=\; L^*(x) - \log(x), \qquad (\forall x\in(0,\infty)). \tag{2.3}$$

LEMMA 2.1 *The extensions $L^*(x)$ and $\gamma^*(x)$ are represented by*

$$L^*(x) \;=\; \psi(x+1) + \gamma, \tag{2.4}$$

$$\gamma^*(x) \;=\; \gamma + \frac{1}{2x} - \frac{1}{12x^2} + O(x^{-4}). \tag{2.5}$$

And $\gamma^*(x)$ satisfies the first property of $\gamma(n)$.

Proof: With respect to an extension of $L(n)$, we consider the descending order of summation for n in $L(n)$, that is,

$$L(n) = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + 1.$$

Since the digamma function has a property $\psi(x+1) = 1/x + \psi(x)$ for $x > 0$, it holds that

$$\psi(x+1) \;=\; \frac{1}{x} + \frac{1}{x-1} + \cdots + \frac{1}{x-[x]+1} + \psi(x-[x]+1).$$

When x is an integer n, it holds that $\psi(n+1) - \psi(1) = L(n)$. By the relation $\psi(1) = -\gamma$ and the definition (2.2), we obtain (2.4). Since the digamma function has the following expansion

$$\psi(x+1) \;=\; \log(x) + \frac{1}{2x} - \frac{1}{12x^2} + O(x^{-4}),$$

the definition (2.3) implies (2.5), which we have the last required. □

Now we study extensions of the results by Kumagai and Inagaki(2000) under the above preliminaries. Let $(p_1, \ldots, p_{k+1})$ be the parameters whose p_j is the probability of j-th cell in the multinomial distribution with $k+1$ cells, and let the k-variate Dirichlet distribution $D_{iri}(\nu_1, \ldots, \nu_k\,;\, \nu_{k+1})$ a prior distribution, where $\{\nu_j\}$ are positive real numbers. Remark that these $\{\nu_j\}$ were assumed to be integers in Kumagai and Inagaki(2000). Then the probability density function of prior distribution is

$$\pi_\nu(\boldsymbol{p}) \;=\; \frac{1}{D(\nu_1, \ldots, \nu_k; \nu_{k+1})} \prod_{j=1}^{k+1} p_j^{\nu_j - 1}, \qquad \boldsymbol{p} = (p_1, \ldots, p_{k+1}),$$

where

$$D(\nu_1, \ldots, \nu_k; \nu_{k+1}) \;=\; \frac{\prod_{j=1}^{k+1} \Gamma(\nu_j)}{\Gamma(\nu)}, \qquad \nu = \sum_{j=1}^{k+1} \nu_j.$$

We would call $D(\nu_1, \ldots, \nu_k; \nu_{k+1})$ the Dirichlet function because of the similarity with respect to the relation between the beta distribution and the beta function. Let $\boldsymbol{n} = (n_1, \ldots, n_{k+1})$ be an observation distributed with the multinomial distribution. Then the posterior probability density function $\pi_\nu(\boldsymbol{p} \,|\, \boldsymbol{n})$ is

$$\pi_\nu(\boldsymbol{p} \mid \boldsymbol{n}) \;=\; \frac{1}{D(n_1 + \nu_1, \ldots, n_k + \nu_k; n_{k+1} + \nu_{k+1})} \prod_{j=1}^{k+1} p_j^{n_j + \nu_j - 1}.$$

This posterior also belongs to the k-variate Dirichlet distribution. When we regard $\boldsymbol{p}$ as a random variable, we use the symbol $\boldsymbol{P} = (P_1, \ldots, P_{k+1})$. By properties of the Dirichlet distribution, each P_j is distributed with the beta distribution, and it holds that

$$\begin{aligned} P_j &\sim B_E(\, n_j + \nu_j,\ \sum_{i \neq j} (n_i + \nu_i)\,), \\ \sum_{j=r+1}^{k+1} P_j &\sim B_E(\, \sum_{j=r+1}^{k+1} (n_j + \nu_j),\ \sum_{j=1}^{r} (n_j + \nu_j)\,). \end{aligned} \tag{2.6}$$

Lemma 2.2 *The posterior expectations in several forms of P_j are represented as follows:*

$$\begin{aligned} E_P[\, P_j\,] &= \frac{n_j + \nu_j}{n + \nu}, \\ E_P[\, \log P_j\,] &= \psi(n_j + \nu_j) - \psi(n + \nu), \\ E_P\left[\log(\sum_{j=r+1}^{k+1} P_j)\right] &= \psi(\sum_{j=r+1}^{k+1} (n_j + \nu_j)) - \psi(n + \nu), \\ E_P[\, P_j\, \log P_j\,] &= \frac{n_j + \nu_j}{n + \nu}\ \{\psi(n_j + \nu_j + 1) - \psi(n + \nu + 1)\}, \\ E_P\left[(\sum_{j=r+1}^{k+1} P_j) \log(\sum_{j=r+1}^{k+1} P_j)\right] &= \frac{\sum_{j=r+1}^{k+1} (n_j + \nu_j)}{n + \nu} \left\{\psi(\sum_{j=r+1}^{k+1} (n_j + \nu_j) + 1) - \psi(n + \nu + 1)\right\}, \end{aligned}$$

where $n = \sum_{j=1}^{k+1} n_j$.

Proof: Since P_j satisfies (2.6), we have

$$\begin{aligned} E_P[P_j] &= \int_0^1 \frac{1}{B(n_j+\nu_j, \sum_{i\neq j}(n_i+\nu_i))} p_j^{n_j+\nu_j}(1-p_j)^{\sum_{i\neq j}(n_i+\nu_i)-1}\, dp_j \\ &= \frac{B(n_j+\nu_j+1, \sum_{i\neq j}(n_i+\nu_i))}{B(n_j+\nu_j, \sum_{i\neq j}(n_i+\nu_i))} = \frac{n_j+\nu_j}{n+\nu} \end{aligned}$$

By the relations (2.1) and (2.4), it holds that

$$E_P[\log P_j] = \psi(n_j+\nu_j) - \psi(n+\nu) = L^*(n_j+\nu_j-1) - L^*(n+\nu-1).$$

We obtain the other three equations by similar calculations. □

Inagaki(1977) showed the relationship that the risk R is decomposed by the two errors, that is, the error of modeling K^M and the errors of estimation K^E as follows: when the parameter θ is true,

$$R(\alpha, T\,|\,\theta) = K^M(\alpha\,|\,\theta) + K^E(T, \theta\,|\,\alpha),$$

where α is an index and T an estimator. Since Lemma 2.2 is the real-valued version of several moments by Kumagai and Inagaki(2000), the real-valued version of the posterior risk function and the minimum posterior risk in the two errors, are consequently obtained by the following theorems:

THEOREM 2.1 *The posterior risk function $\rho(r,\ \boldsymbol{T}_r)$ is exactly represented as follows :*

$$\rho(r,\ \boldsymbol{T}_r) = \rho^M(r) + \rho^E(r,\ \boldsymbol{T}_r),$$

where

$$\rho^M(r) = \sum_{j=r+1}^{k+1} n_j\, L^*(n_j+\nu_j-1) - \left(\sum_{j=r+1}^{k+1} n_j\right)\left\{L^*\Big(\sum_{j=r+1}^{k+1}(n_j+\nu_j) - 1\Big) - \log(k+1-r)\right\},$$

$$\begin{aligned} \rho^E(r, \boldsymbol{T}_r) &= -n\, L^*(n+\nu) + n\sum_{j=1}^{r}\frac{n_j+\nu_j}{n+\nu}\{L^*(n_j+\nu_j) - \log T_j\} \\ &\quad + n\,\frac{\sum_{j=r+1}^{k+1}(n_j+\nu_j)}{n+\nu}\left\{L^*\Big(\sum_{j=r+1}^{k+1}(n_j+\nu_j)\Big) - \log((k+1-r)\,T_{r+1})\right\}. \end{aligned}$$

□

THEOREM 2.2 *By the Bayes solution $\boldsymbol{q}_r^*$, the minimum posterior risk $\rho(r,\ \boldsymbol{q}_r^*)$ is represented as the sum of the modeling risk $\rho^M(r)$ and the estimation risk $\rho^E(r,\ \boldsymbol{q}_r^*)$:*

$$\rho(r,\ \boldsymbol{q}_r^*) = \rho^M(r) + \rho^E(r,\ \boldsymbol{q}_r^*) = \rho^M(r) + \rho^E(r),$$

where the modeling risk $\rho^M(r)$ is the same in Theorem 2.1 and the estimation risk is

$$\begin{aligned} \rho^E(r) &= \frac{n}{n+\nu}\sum_{j=1}^{r}(n_j+\nu_j)\ \gamma^*(n_j+\nu_j) \\ &\quad + \frac{n}{n+\nu}\left(\sum_{j=r+1}^{k+1}(n_j+\nu_j)\right)\ \gamma^*\Big(\sum_{j=r+1}^{k+1}(n_j+\nu_j)\Big) - n\,\gamma^*(n+\nu). \end{aligned}$$

□

In the exact formulae of Theorem 2.1 and 2.2, we obtain true-extended representations for the posterior risk and the minimum posterior risk. On the other hand, it is easily checked that the asymptotic extended representations are equivalent to *Theorem 2.3* in Kumagai and Inagaki(2000) by the followings; (2.5), Theorem 2.1, and Theorem 2.2.

3 Conclusion By using the digamma function we extend the hyper-parameter of the prior Dirichlet distribution, and obtained the extended real-valued version for the results in Kumagai and Inagaki(2000).

Acknowledgement

This research was partially supported by the following grants-in-aid from Japan: grant 13740065 in the first author and grant 14540115 in the second one.

References

Antelman, G.R., 1972. Interrelated Bernoulli Process. J. American Statist. Assoc. 67(340), 831–841.

Bloch, D.A. and Watson, G.S., 1967. A Bayesian Study of the Multinomial Distribution. Ann. Math. Stat. 38(5), 1423–1435.

Connor, R.J. and Mosimann, J.E., 1969. Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution. J. American Statist. Assoc. 64(324), 194–206.

Gruber, M.H.J., 1998. Improving Efficiency by Shrinkage : The James-Stein and Ridge Regression Estimators. Marcel Dekker, Inc.

Inagaki, N., 1977. Two Errors in Statistical Model Fitting. Ann. Inst. Stat. math. 29(A), 131–152.

Kumagai,E. and Inagaki,N., 2000. A Bayes Approach to a Grouping of Small Events in the Multinomial Distribution. Scientiae Mathematicae (Electronic Journal) 3(1), 1–17.

Lindley, D.V., 1964. The Bayesian Analysis of Contingency Tables. Ann. Math. Stat. 35(4), 1622–1643.

Wilks,S.S., 1962. Mathematical Statistics. John Wiley & Sons.

Department of Financial Statistics, Graduate School of Engineering Science, Osaka University,
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan