A BAYES APPROACH TO A GROUPING OF SMALL EVENTS IN THE MULTINOMIAL DISTRIBUTION

ETSUO KUMAGAI * AND NOBUO INAGAKI †

Received November 18, 1999

Abstract. It is known that, in the chi-square test of the goodness of fit, the expected frequency of each cells should be greater than 5 for the large sample theory to hold, and thus that, if there are cells with small observed frequencies, one should group them so that the new grouped cell has the observed frequency greater than 5. In the present paper, we treat the problem of grouping of small cells in the test of goodness of fit from the viewpoint of a Bayes approach to the decision theoretic framework of model fitting proposed by Inagaki(1977b). Then we have two errors, one of which is caused by the estimation of probabilities of cells and the other by the grouping of small cells, and obtain the exact and asymptotic representations of two errors explicitly. By using them, we compare a new Bayes grouping rule of this model fitting to the usual grouping rule of small cells.

1 Introduction Let us consider the test of goodness of fit for an assumed distribution function where we usually go through the following four steps. (1) We adequately divide the total space into several disjoint subsets which are called cells, (2) count the observed frequency and the corresponding expected frequency of each cell under the assumed distribution, (3) calculate the value of the chi-square test statistic or the log-likelihood ratio statistic, (4) decide to accept or reject the hypothesis of assumed distribution by comparing its value to the critical limit value. In this procedure, we notice that the used model is not the very distribution assumed above but the multinomial distribution, that is, the used information is not the very observed values but the observed frequencies fallen in cells, and furthermore, that the critical point is approximately determined by using the fact that the test statistics asymptotically distribute to the chi-square.

For the large sample theory to hold in the test of goodness of fit, it is necessary that we group small events into a new big event so as to have the observed frequency greater than 5 of every event. For example, see Rao(1973, page 396) and Azzalini(1996, page 137). We could interpret grouping observations of small events as equalizing their probabilities in the multinomial distribution model, because the observation appears as the exponent of the probability of event in the probability function. Equalizing probabilities of grouped cells leads to decreasing the number of parameters to be estimated (the dimension of parameter space) and thus reducing the error of estimation, while that causes to increase the error of modeling.

1991 *Mathematics Subject Classification.* 62B10, 62F20.

Key words and phrases. test of goodness of fit, multinomial distribution, grouping of small events, Bayes grouping rule, two errors in modeling, Dirichlet prior distribution, AIC statistic.

*This research was partially supported by the Ministry of Education, Science, Sports, and Culture, Grant-in-Aid for Encouragement of Young Scientists, 10740052, 1998.

†This research was partially supported by the Ministry of Education, Science, Sports, and Culture, Grant-in-Aid for Scientific Research(C), 11640117, 1999.

The main aim of the present paper is to apply the decision theoretic framework of model fitting proposed by Inagaki(1977b) to grouping small cells in the test of goodness of fit, to obtain the exact and asymptotic representations of two errors explicitly, and further, to compare a new Bayes grouping rule of this model fitting to the usual grouping rule of small cells.

In section 2, we discuss the error of model fitting consists of the sum of the modeling error and the estimation error. We derive a Bayes grouping rule by using a Dirichlet distribution as a prior distribution in this decision framework of the model fitting.

In section 3, we show that a criterion by the Bayes grouping rule converges to the AIC statistic (Akaike(1973)) of the model selection as the sample increases to be large. Furthermore, we obtain the asymptotic second order term of the difference of these two statistics, which depends on the hyperparameters of the prior distribution.

In section 4, we carry out a simulation for a simple situation. It is very interesting that both the criterion by the Bayes grouping rule and the AIC statistic give so similar results as the usual grouping rule of small cells in large sample cases, while these rules give different decisions in not so large sample cases.

2 Bayes grouping rule for small events For the large sample theory to hold in the test of goodness of fit, it is necessary that we group several (two or three or four) small events located side by side into one new large event so as to have the observed frequency greater than 5 of every event.

Let a whole event with observed frequency n be divided into $k+1$ events : $E_1, \ldots, E_r$, $E_{r+1}, \ldots, E_{k+1}$, $(r \leq k)$ with their observed frequencies : $n_1, \ldots, n_r$, $n_{r+1}, \ldots, n_{k+1}$, respectively, where $n_1 + \cdots + n_{k+1} = n$ for n_i positive integers.

We assume that the first r events are located not to be small but the last $k-r$ to be small $(n_1, \ldots, n_r \geq 5, \quad n_{r+1}, \ldots, n_{k+1} \leq 5)$. The readers may consider that the assumption is very artificial, but this is *a* setting in our article and it is often the case with practical data by sequencing $k+1$ events. Thus, we leave the first r events as they are but group the last $k+1-r$ events and make the new $r+1$-th event with the summation of their frequencies : $n_{r+1} + \cdots + n_{k+1}$. We could interpret grouping observations of small events as equalizing their probabilities in the multinomial distribution model, because the observation appears as the exponent of the probability of event in the probability function. Our interpretation is based on the property of the exponential family of the multinomial distribution. See the table 1.

	Original Refinement		Grouping of Events
Event	$E_1, \ldots, E_r,\ E_{r+1}, \ldots, E_{k+1}$		$E_1, \ldots, E_r,\ (E_{r+1}, \ldots, E_{k+1})$
Probability	$p_1, \ldots, p_r,\ p_{r+1}, \ldots, p_{k+1}$	$\Rightarrow$	$q_1, \ldots, q_r,\ q_{r+1} = \cdots = q_{k+1}$
Observation	$n_1, \ldots, n_r,\ n_{r+1}, \ldots, n_{k+1}$		$n_1, \ldots, n_r,\ (n_{r+1}, \ldots, n_{k+1})$

Table 1: Grouping of Small Events

Then, the observation $\boldsymbol{n} = (n_1, \ldots, n_{k+1})$ has the probability functions of $(k+1)$-variate multinomial distribution for the probability $\boldsymbol{p} = (p_1, \ldots, p_{k+1})$ with $p_i > 0$ and $p_1 + \cdots + p_{k+1} = 1$:

$$f(\boldsymbol{n} \,|\, \boldsymbol{p}) \;=\; \frac{n!}{n_1! \cdots n_{k+1}!}\, p_1^{n_1} \cdots p_{k+1}^{n_{k+1}} \;,$$

and for the grouped probability $\boldsymbol{q}_r = (q_1, \ldots, q_r, \underbrace{q_{r+1}, \ldots, q_{r+1}}_{k+1-r})$ with $q_i > 0$ and $q_1 + \cdots +$

$q_r + (k+1-r)q_{r+1} = 1$:

$$f_r(\boldsymbol{n} \mid \boldsymbol{q}_r) \;=\; \frac{n!}{n_1! \cdots n_{k+1}!}\, q_1^{n_1} \cdots q_r^{n_r} q_{r+1}^{n_{r+1}+\cdots+n_{k+1}} \;, \tag{2.1}$$

respectively. Equalizing probabilities of grouped cells leads to decreasing the number of parameters to be estimated (the dimension of parameter space) and thus reducing the error of estimation, while that causes to increase the error of modeling. Thus we apply the decision theoretic framework of model fitting proposed by Inagaki to grouping small cells in the test of goodness of fit.

First, we consider the modeling error of accepting the grouped model $\boldsymbol{q}_r$ against the original model $\boldsymbol{p}$, which is denoted by $K^M(r\,|\,\boldsymbol{p})$ and is defined by the infimum of the Kullback-Leibler information of $f_r(\boldsymbol{n} \mid \boldsymbol{q}_r)$ under $f(\boldsymbol{n} \mid \boldsymbol{p})$:

$$K^M(r\,|\,\boldsymbol{p}) \;\equiv\; \inf_{\boldsymbol{q}_r} \left\{ \sum_{\boldsymbol{n}} \log \frac{f(\boldsymbol{n} \mid \boldsymbol{p})}{f_r(\boldsymbol{n} \mid \boldsymbol{q}_r)} \cdot f(\boldsymbol{n} \mid \boldsymbol{p}) \right\} \;. \tag{2.2}$$

It is easy to see that the infimum is achieved at $\boldsymbol{q}_r = \boldsymbol{q}_r(\boldsymbol{p})$:

$$\boldsymbol{q}_r(\boldsymbol{p}) \;=\; \left(p_1, \ldots, p_r, \underbrace{\frac{p_{r+1}+\cdots+p_{k+1}}{k+1-r}, \ldots, \frac{p_{r+1}+\cdots+p_{k+1}}{k+1-r}}_{k+1-r} \right),$$

and thus, that the modeling error (2.2) is represented as follows :

$$\begin{aligned} K^M(r\,|\,\boldsymbol{p}) \;&=\; \sum_{\boldsymbol{n}} \log \frac{f(\boldsymbol{n} \mid \boldsymbol{p})}{f_r(\boldsymbol{n} \mid \boldsymbol{q}_r(\boldsymbol{p}))} \cdot f(\boldsymbol{n} \mid \boldsymbol{p}) \\ &=\; n \left\{ \sum_{j=r+1}^{k+1} p_j \log p_j \;-\; \left(\sum_{j=r+1}^{k+1} p_j \right) \log \frac{\sum_{j=r+1}^{k+1} p_j}{k+1-r} \right\}. \end{aligned} \tag{2.3}$$

For the observation $\boldsymbol{n}$, let $\widehat{\boldsymbol{p}} = \widehat{\boldsymbol{p}}(\boldsymbol{n})$ and $\widehat{\boldsymbol{q}}_r = \widehat{\boldsymbol{q}}_r(\boldsymbol{n})$ be maximum likelihood estimators of $\boldsymbol{p}$ and $\boldsymbol{q}_r$ in the original probability function $f(\boldsymbol{n} \mid \boldsymbol{p})$ and the grouped one $f_r(\boldsymbol{n} \mid \boldsymbol{q}_r)$, respectively. Then, we obtain

$$\widehat{\boldsymbol{p}} \;=\; \left(\frac{n_1}{n}, \ldots, \frac{n_r}{n}, \frac{n_{r+1}}{n}, \ldots, \frac{n_{k+1}}{n} \right), \tag{2.4}$$

$$\widehat{\boldsymbol{q}}_r \;=\; \left(\frac{n_1}{n}, \ldots, \frac{n_r}{n}, \underbrace{\frac{n_{r+1}+\cdots+n_{k+1}}{n(k+1-r)}, \ldots, \frac{n_{r+1}+\cdots+n_{k+1}}{n(k+1-r)}}_{k+1-r} \right). \tag{2.5}$$

This shows the invariance of maximum likelihood estimators for the dimension of parameter space : $\widehat{\boldsymbol{q}}_r = \boldsymbol{q}_r(\widehat{\boldsymbol{p}})$.

Second, we consider the estimation error $K^E(\boldsymbol{T}_r,\ \boldsymbol{p} \mid r)$ which is defined by the Kullback-Leibler information between an estimator $\boldsymbol{T}_r = \boldsymbol{T}_r(\boldsymbol{n})$ for the original observation $\boldsymbol{n} = (n_1, \ldots, n_{k+1})$ and a parameter $\boldsymbol{q}_r(\boldsymbol{p})$ under the selected model $f_r(\boldsymbol{n} \mid \boldsymbol{q}_r(\boldsymbol{p}))$:

$$K^E(\boldsymbol{T}_r,\ \boldsymbol{p} \mid r) \;\equiv\; \sum_{\boldsymbol{n}} \left[\sum_{\boldsymbol{n}'} \log \frac{f_r(\boldsymbol{n}' \mid \boldsymbol{q}_r(\boldsymbol{p}))}{f_r(\boldsymbol{n}' \mid \boldsymbol{T}_r)} \cdot f_r(\boldsymbol{n}' \mid \boldsymbol{q}_r(\boldsymbol{p})) \right] f(\boldsymbol{n} \mid \boldsymbol{p}) \;, \tag{2.6}$$

where $\boldsymbol{n}' = (n'_1, \dots, n'_{k+1})$ with $n = \sum_{j=1}^{k+1} n'_j$ is independent of the original observation $\boldsymbol{n}$.

Inagaki(1977b) defined the risk function $R(r, \boldsymbol{T}_r \mid \boldsymbol{p})$ of statistical model fitting by the sum of the modeling error (2.3) and the estimation error (2.6) :

$$R(r, \boldsymbol{T}_r \mid \boldsymbol{p}) \;\equiv\; K^M(r \mid \boldsymbol{p}) \;+\; K^E(\boldsymbol{T}_r,\ \boldsymbol{p} \mid r)\ .$$

Similarly, the loss function $W(r, \boldsymbol{T}_r \mid \boldsymbol{p})$ is defined by the sum of the modeling loss $W^M(r \mid \boldsymbol{p})$ and the estimation loss $W^E(\boldsymbol{T}_r,\ \boldsymbol{p} \mid r)$:

$$W(r, \boldsymbol{T}_r \mid \boldsymbol{p}) \;\equiv\; W^M(r \mid \boldsymbol{p}) \;+\; W^E(\boldsymbol{T}_r,\ \boldsymbol{p} \mid r)\ , \tag{2.7}$$

which are represented in the following lemma :

Lemma 2.1

$$\begin{aligned}
W^M(r \mid \boldsymbol{p}) &= \log \frac{f(\boldsymbol{n} \mid \boldsymbol{p})}{f_r(\boldsymbol{n} \mid \boldsymbol{q}_r(\boldsymbol{p}))}, \\
&= \sum_{j=r+1}^{k+1} n_j \log \frac{p_j}{(\sum_{i=r+1}^{k+1} p_i)/(k+1-r)}, \\
W^E(\boldsymbol{T}_r,\ \boldsymbol{p} \mid r) &= \sum_{\boldsymbol{n}'} \log \frac{f_r(\boldsymbol{n}' \mid \boldsymbol{q}_r(\boldsymbol{p}))}{f_r(\boldsymbol{n}' \mid \boldsymbol{T}_r)} \cdot f_r(\boldsymbol{n}' \mid \boldsymbol{q}_r(\boldsymbol{p})) \\
&= n \left[\sum_{j=1}^{r} p_j \log \frac{p_j}{T_j} \;+\; \left(\sum_{j=r+1}^{k+1} p_j \right) \log \frac{(\sum_{j=r+1}^{k+1} p_j)/(k+1-r)}{T_{r+1}} \right],
\end{aligned}$$

where the estimator $\boldsymbol{T}_r = (T_1, \dots, T_{r+1})$.

PROOF. We have easily the result by the direct calculations. □

Let us consider a Bayes grouping rule by using a Dirichlet distribution as a prior distribution in this decision framework of the model fitting. We assume that parameters $(p_1, \dots, p_{k+1})$ are distributed with k-variate Dirichlet distribution $D(\nu_1, \dots, \nu_k\ ;\ \nu_{k+1})$ as a prior distribution. Then the prior density is

$$\pi(\boldsymbol{p}) \;=\; \frac{\Gamma(\nu_1 + \cdots + \nu_{k+1})}{\Gamma(\nu_1) \cdots \Gamma(\nu_{k+1})} \prod_{j=1}^{k+1} p_j^{\nu_j - 1}, \tag{2.8}$$

where $\{\nu_j\}$ are positive real numbers and $\Gamma(\cdot)$ is the gamma function. About the Dirichlet distribution, for instance, see Wilks(1962, page 179). Then, the joint probability function $f(\boldsymbol{n}, \boldsymbol{p})$ is

$$f(\boldsymbol{n}\,, \boldsymbol{p}) \;=\; \pi(\boldsymbol{p})\, f(\boldsymbol{n} \mid \boldsymbol{p}) \;=\; \frac{n!}{n_1! \cdots n_{k+1}!} \cdot \frac{\Gamma(\nu_1 + \cdots + \nu_{k+1})}{\Gamma(\nu_1) \cdots \Gamma(\nu_{k+1})} \prod_{j=1}^{k+1} p_j^{n_j + \nu_j - 1}.$$

Since the marginal probability function $f(\boldsymbol{n})$ of $\boldsymbol{n}$ is one of Dirichlet Multinomial, that is,

$$f(\boldsymbol{n}) = \int f(\boldsymbol{n},\ \boldsymbol{p})\, d\boldsymbol{p} = \frac{n!}{n_1! \cdots n_{k+1}!} \cdot \frac{\Gamma(\sum_{j=1}^{k+1} \nu_j)}{\prod_{j=1}^{k+1} \Gamma(\nu_j)} \cdot \frac{\prod_{j=1}^{k+1} \Gamma(n_j + \nu_j)}{\Gamma(\sum_{j=1}^{k+1} (n_j + \nu_j))}\ ,$$

the posterior density function $\pi(\boldsymbol{p} \mid \boldsymbol{n})$ is

$$\pi(\boldsymbol{p} \mid \boldsymbol{n}) \;=\; \frac{f(\boldsymbol{n},\ \boldsymbol{p})}{f(\boldsymbol{n})} \;=\; \frac{\Gamma(\sum_{j=1}^{k+1}(n_j+\nu_j))}{\prod_{j=1}^{k+1}\Gamma(n_j+\nu_j)} \prod_{j=1}^{k+1} p_j^{n_j+\nu_j-1}, \tag{2.9}$$

which is of the k-variate Dirichlet distribution $D(n_1+\nu_1, \ldots, n_k+\nu_k\ ;\ n_{k+1}+\nu_{k+1})$.

By the loss function (2.7) and the posterior distribution (2.9), the Bayes risk $R_B(\boldsymbol{T}_r)$ is represented by

$$\begin{aligned} R_B(\boldsymbol{T}_r) &= \sum_{\boldsymbol{n}} \left\{ \int W(r, \boldsymbol{T}_r \mid \boldsymbol{p})\ \pi(\boldsymbol{p} \mid \boldsymbol{n})\ d\boldsymbol{p} \right\}\ f(\boldsymbol{n}) \\ &= \sum_{\boldsymbol{n}} \left\{ \int W^M(r \mid \boldsymbol{p})\ \pi(\boldsymbol{p} \mid \boldsymbol{n})\ d\boldsymbol{p} \right\}\ f(\boldsymbol{n}) \\ &\quad + \sum_{\boldsymbol{n}} \left\{ \int W^E(r, \boldsymbol{T}_r \mid \boldsymbol{p})\ \pi(\boldsymbol{p} \mid \boldsymbol{n})\ d\boldsymbol{p} \right\}\ f(\boldsymbol{n}) \\ &= R_B^M(r) + R_B^E(\boldsymbol{T}_r) \qquad \text{(say)}\ . \end{aligned} \tag{2.10}$$

The integrated part of the right-hand side of the last equation is the posterior risk which is denoted by $\rho(r,\ \boldsymbol{T}_r)$ and into two parts ρ^M, ρ^E corresponding to two loss function (2.7) :

$$\begin{aligned} \rho(r,\ \boldsymbol{T}_r) &\equiv \int W(r, \boldsymbol{T}_r \mid \boldsymbol{p})\ \pi(\boldsymbol{p} \mid \boldsymbol{n})\, d\boldsymbol{p} \\ &= \int W^M(r \mid \boldsymbol{p})\ \pi(\boldsymbol{p} \mid \boldsymbol{n})\, d\boldsymbol{p}\ +\ \int W^E(\boldsymbol{T}_r, \boldsymbol{p} \mid r)\ \pi(\boldsymbol{p} \mid \boldsymbol{n})\, d\boldsymbol{p} \\ &= \rho^M(r) + \rho^E(r, \boldsymbol{T}_r) \quad \text{(say)}. \end{aligned} \tag{2.11}$$

Let us denote *a partial sum of harmonic series* by $L(n)$ for a positive integer n :

$$L(n) \;\equiv\; 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n},$$

and furthermore, the difference function of it and $\log n$ by $\gamma(n)$:

$$\gamma(n) = L(n) - \log n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n, \tag{2.12}$$

which we call *Euler's sequence*. Then, the following two lemmas are known (see Courant and John(1965, page 526)) :

Lemma 2.2 *The sequence $\gamma(n)$ is monotone decreasing and converges to the Euler's constant γ :*

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

Lemma 2.3 *The Euler's sequence $\gamma(n)$ has an asymptotic expansion for all large numbers n :*

$$\gamma(n) \;=\; \gamma + \frac{1}{2n} - \frac{1}{12n^2} + O(n^{-4})\ . \tag{2.13}$$

Lemma 2.4 *For any positive large integer n, $L(n)$ has the asymptotic representation :*

$$L(n) \;=\; \log(n) + \gamma + \frac{1}{2\,n} - \frac{1}{12\,n^2} + O(n^{-4}).$$

PROOF. The result is obtained by using the two representations (2.12) and (2.13) in Euler's sequence. □

We assume that $\{\nu_j\}$ in $\pi(\boldsymbol{p})$ are all positive integers in the sequel of this paper.

When the $(k+1)$-dimensional random vector $\boldsymbol{P} = (P_1, \cdots, P_{k+1})$ is distributed to Dirichlet Distribution $D(n_1+\nu_1, \cdots, n_k+\nu_k; n_{k+1}+\nu_{k+1})$, let the notation $E_P[\cdot]$ denote the expectation due to the posterior density $\pi(\boldsymbol{p} \mid \boldsymbol{n})$ (2.9).

Theorem 2.1 *The posterior risk (2.11) is explicitly represented by*

$$\rho(r,\ \boldsymbol{T}_r) \ = \ E_P[\,W^M(r \mid \boldsymbol{P})\,] + E_P[\,W^E(\boldsymbol{T}_r,\ \boldsymbol{P} \mid r)\,] \ = \ \rho^M(r) + \rho^E(r,\ \boldsymbol{T}_r),$$

where

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

PROOF. We have the following moments of Dirichlet distribution :

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

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

where $j = 1, \ldots, k+1$ in the above first, second, and fourth equations. Thus, for two loss functions in Lemma 2.1, $\rho^M(r)$ is obtained by applying the above second and third

equations to the expectation of the modeling loss $W^M(r \mid \boldsymbol{p})$, similarly $\rho^E(r, \boldsymbol{T}_r)$ is obtained by applying the above first, fourth, and fifth equations to the expectation of the estimation loss $W^E(\boldsymbol{T}_r, \boldsymbol{p} \mid r)$. Hence the proof is completed. □

Theorem 2.2 *The elements of the Bayes solution $\boldsymbol{q}_r^*$ for the Bayes risk (2.10) are obtained as*

$$q_j^* = \begin{cases} (n_j+\nu_j)/(\sum_{i=1}^{k+1}(n_i+\nu_i)) & (j \le r), \\ (\sum_{i=r+1}^{k+1}(n_i+\nu_i))/((k+1-r)(\sum_{i=1}^{k+1}(n_i+\nu_i))) & (j = r+1). \end{cases}$$

Thus the minimized posterior risk $\rho(r,\ \boldsymbol{q}_r^*)$ by the Bayes solution is represented by the sum of the modeling risk $\rho^M(r)$ and the estimation risk $\rho^E(r, \boldsymbol{q}_r^*)$:

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

where the modeling risk is

$$\rho^M(r) = \sum_{j=r+1}^{k+1} n_j\, L(n_j+\nu_j-1) - \left(\sum_{j=r+1}^{k+1} n_j\right)\left\{L(\sum_{j=r+1}^{k+1}(n_j+\nu_j)\ -1) - \log(k+1-r)\right\}, \tag{2.14}$$

and the estimation risk is

$$\begin{aligned}\rho^E(r) &= \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\sum_{j=1}^{r}(n_j+\nu_j)\ \gamma(n_j+\nu_j) \\ &\quad + \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\left(\sum_{j=r+1}^{k+1}(n_j+\nu_j)\right)\ \gamma(\sum_{j=r+1}^{k+1}(n_j+\nu_j)) \\ &\quad - n\,\gamma(\sum_{j=1}^{k+1}(n_j+\nu_j)).\end{aligned} \tag{2.15}$$

Proof. Since to minimize the Bayes risk (2.10) with respect to $\boldsymbol{T}_r$ equals to minimize $\rho^E(r, \boldsymbol{T}_r)$ in the posterior risk of Theorem 2.1, the differentiations of $\rho^E(r, \boldsymbol{T}_r)$ by the elements of $\boldsymbol{T}_r$ imply the Bayes solution $\boldsymbol{q}_r^*$. By substituting $\boldsymbol{q}_r^*$ for $\boldsymbol{T}_r$ in the posterior risk (2.11), Theorem 2.1 and (2.12) imply that the minimized posterior risk $\rho(r,\ \boldsymbol{q}_r^*)$ is represented by the sum of the modeling risk $\rho^M(r)$ and the estimation risk $\rho^E(r)$ explicitly. The proof is completed. □

Set $\boldsymbol{p}^* = \boldsymbol{q}_k^*$, that is,

$$p_j^* = \frac{n_j+\nu_j}{\sum_{i=1}^{k+1}(n_i+\nu_i)} \qquad j = 1, \cdots, k+1.$$

Then, we also obtain the invariance of Bayes solutions : $\boldsymbol{q}_r^* = \boldsymbol{q}_r(\boldsymbol{p}^*)$.

We shall propose the minimized posterior risk $\rho(r,\ \boldsymbol{q}_r^*)$ in Theorem 2.2 as a criterion for the grouping of small events in the multinomial distribution, and let us give the name "Bayes grouping rule" to our application of the decision theoretic framework for the model fitting. If every $\nu_j = 1$ in the Dirichlet prior density (2.8), then the prior is the uniform distribution. Thus, we have the following corollary :

Corollary 2.1 *Let $\rho_U(r)$ be the minimized posterior risk when the prior is uniform. Then*

$$\rho_U(r) \;=\; \rho_U^M(r) + \rho_U^E(r),$$

where the modeling risk and the estimation risk are

$$\begin{aligned}
\rho_U^M(r) &= \sum_{j=r+1}^{k+1} n_j\, L(n_j) - \left(\sum_{j=r+1}^{k+1} n_j\right)\left\{L(\sum_{j=r+1}^{k+1} n_j + k - r) - \log(k+1-r)\right\},\\
\rho_U^E(r) &= \frac{n}{n+k+1}\sum_{j=1}^{r}(n_j+1)\;\gamma(n_j+1)\\
&\quad + \frac{n}{n+k+1}\left(\sum_{j=r+1}^{k+1}(n_j+1)\right)\;\gamma(\sum_{j=r+1}^{k+1}(n_j+1)) - n\,\gamma(n+k+1),
\end{aligned}$$

respectively.

Theorem 2.3 *The minimized posterior risk $\rho(r, \boldsymbol{q}_r^*)$ has the asymptotic representations for the modeling risk (2.14) and the estimation risk (2.15) as follows : the modeling risk is*

$$\rho^M(r) \;=\; BR^M(r) + O(n^{-3}) \quad \textit{(say)}, \tag{2.16}$$

where the term $BR^M(r)$ is

$$\begin{aligned}
&BR^M(r)\\
&= \sum_{j=r+1}^{k+1} n_j\,\log(n_j+\nu_j-1)\\
&\quad - \left(\sum_{i=r+1}^{k+1} n_i\right)\left\{\log(\sum_{i=r+1}^{k+1}(n_i+\nu_j)\;-1) - \log(k+1-r)\right\}\\
&\quad + \left\{\sum_{j=r+1}^{k+1}\frac{n_j}{2\,(n_j+\nu_j-1)} - \frac{\sum_{j=r+1}^{k+1} n_j}{2\,(\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1)}\right\}\\
&\quad + \left\{\frac{\sum_{j=r+1}^{k+1} n_j}{12\,(\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1)^2} - \sum_{j=r+1}^{k+1}\frac{n_j}{12\,(n_j+\nu_j-1)^2}\right\}.
\end{aligned}$$

And the estimation risk is

$$\rho^E(r) \;=\; BR^E(r) + O(n^{-3}) \quad \textit{(say)}, \tag{2.17}$$

where the term $BR^E(r)$ is

$$\begin{aligned}
BR^E(r) \;&=\; \frac{r}{2}\,\frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\\
&\quad - \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\left\{\sum_{j=1}^{r}\frac{1}{12\,(n_j+\nu_j)} + \frac{1}{12\,\sum_{j=r+1}^{k+1}(n_j+\nu_j)} - \frac{1}{12\,\sum_{j=1}^{k+1}(n_j+\nu_j)}\right\}.
\end{aligned}$$

PROOF. By applying Lemma 2.4 to the terms $L(\cdot)$ in the modeling risk (2.14), $\rho^M(r)$ becomes

$$
\begin{aligned}
&\rho^M(r) \\
&= \sum_{j=r+1}^{k+1} n_j\, L(n_j+\nu_j-1) - \left(\sum_{j=r+1}^{k+1} n_j\right)\left\{L(\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1) - \log(k+1-r)\right\} \\
&= \sum_{j=r+1}^{k+1} n_j \left\{\log(n_j+\nu_j-1)+\gamma+\frac{1}{2(n_j+\nu_j-1)} - \frac{1}{12(n_j+\nu_j-1)^2} + O(n^{-4})\right\} \\
&\quad - \left(\sum_{j=r+1}^{k+1} n_j\right)\left\{\log(\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1)+\gamma\right. \\
&\qquad\qquad \left. + \frac{1}{2(\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1)} - \frac{1}{12(\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1)^2} + O(n^{-4})\right\} \\
&\quad + \left(\sum_{j=r+1}^{k+1} n_j\right)\log(k+1-r),
\end{aligned}
$$

so that the modeling risk has the asymptotic representation (2.16). By applying the asymptotic expansion (2.13) to the terms $\gamma(\cdot)$ in the estimation risk (2.15), $\rho^E(r)$ becomes

$$
\begin{aligned}
\rho^E(r) &= \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)} \sum_{j=1}^{r}(n_j+\nu_j)\ \gamma(n_j+\nu_j) \\
&\quad + \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\left(\sum_{j=r+1}^{k+1}(n_j+\nu_j)\right)\ \gamma(\sum_{j=r+1}^{k+1}(n_j+\nu_j)) - n\,\gamma(\sum_{j=1}^{k+1}(n_j+\nu_j)) \\
&= \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)} \sum_{j=1}^{r}(n_j+\nu_j)\left\{\gamma+\frac{1}{2(n_j+\nu_j)} - \frac{1}{12(n_j+\nu_j)^2} + O(n^{-4})\right\} \\
&\quad + \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\left(\sum_{j=r+1}^{k+1}(n_j+\nu_j)\right)\left\{\gamma+\frac{1}{2\sum_{j=r+1}^{k+1}(n_j+\nu_j)}\right. \\
&\qquad\qquad \left. - \frac{1}{12(\sum_{j=r+1}^{k+1}(n_j+\nu_j))^2} + O(n^{-4})\right\} \\
&\quad - n\left\{\gamma+\frac{1}{2\sum_{j=1}^{k+1}(n_j+\nu_j)} - \frac{1}{12(\sum_{j=1}^{k+1}(n_j+\nu_j))^2} + O(n^{-4})\right\},
\end{aligned}
$$

and

$$\begin{aligned}
\rho^E(r) &= \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\left\{\sum_{j=1}^{r}(n_j+\nu_j)+\sum_{j=r+1}^{k+1}(n_j+\nu_j)\right\}\gamma - n\,\gamma \\
&\quad + \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\left\{\sum_{j=1}^{r}\frac{1}{2}+\frac{1}{2}\right\} - \frac{n}{2(\sum_{j=1}^{k+1}(n_j+\nu_j))} \\
&\quad - \frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\left\{\sum_{j=1}^{r}\frac{1}{12(n_j+\nu_j)}+\frac{1}{12\sum_{j=r+1}^{k+1}(n_j+\nu_j)}\right\} \\
&\quad + \frac{n}{12(\sum_{j=1}^{k+1}(n_j+\nu_j))^2} + O(n^{-3}),
\end{aligned}$$

so that the estimation risk has the asymptotic representation (2.17). Therefore the proof is completed. □

Corollary 2.2 *The modeling risk $\rho_U^M(r)$ has the asymptotic representation :*

$$\rho_U^M(r) \;=\; UBR^M(r) + O(n^{-3}) \quad \textit{(say)},$$

where the term $UBR^M(r)$ is

$$\begin{aligned}
UBR^M(r) &= \sum_{j=r+1}^{k+1} n_j \log n_j - \left(\sum_{i=r+1}^{k+1} n_i\right)\left\{\log(\sum_{i=r+1}^{k+1} n_i \;+k-r) - \log(k+1-r)\right\} \\
&\quad + \left\{\frac{k+1-r}{2} - \frac{\sum_{j=r+1}^{k+1} n_j}{2(\sum_{j=r+1}^{k+1} n_j \;+k-r)}\right\} \\
&\quad + \left\{\frac{\sum_{j=r+1}^{k+1} n_j}{12(\sum_{j=r+1}^{k+1} n_j \;+k-r)^2} - \sum_{j=r+1}^{k+1}\frac{1}{12\,n_j}\right\}.
\end{aligned}$$

And the estimation risk $\rho_U^E(r)$ has the asymptotic representation :

$$\rho_U^E(r) \;=\; UBR^E(r) + O(n^{-3}) \quad \textit{(say)},$$

where the term $UBR^E(r)$ is

$$\begin{aligned}
UBR^E(r) &= \frac{r}{2}\,\frac{n}{n+k+1} \\
&\quad - \frac{n}{n+k+1}\left\{\sum_{j=1}^{r}\frac{1}{12\,(n_j+1)} + \frac{1}{12\,\sum_{j=r+1}^{k+1}(n_j+1)} - \frac{1}{12\,(n+k+1)}\right\}.
\end{aligned}$$

3 Comparison between the Bayes grouping rule and the AIC rule We shall compare the Bayes grouping rule of model fitting to the AIC rule for small cells. It is enough that we actually compare the minimized posterior risk in Theorem 2.3 to 1/2 times the AIC statistic. For our grouped probability (2.1), the AIC statistic is denoted by

$$AIC(r) \;=\; 2\,\log\frac{f(\boldsymbol{n}\mid\widehat{\boldsymbol{p}})}{f_r(\boldsymbol{n}\mid\widehat{\boldsymbol{q}}_r)} \;+\; 2r-k\;,$$

where $\widehat{\boldsymbol{p}}$, $\widehat{\boldsymbol{q}}_r$ are maximum likelihood estimators (2.4), (2.5) respectively. The half of $AIC(r)$ has an explicit representation by calculating it directly :

$$\begin{aligned}
&\frac{1}{2}AIC(r) \\
&= \left[\sum_{j=r+1}^{k+1} n_j \log n_j + \left(\sum_{i=r+1}^{k+1} n_i\right)\left\{\log(k+1-r) - \log(\sum_{i=r+1}^{k+1} n_i)\right\} + \frac{r-k}{2}\right] + \frac{r}{2} \\
(3.1) = &\ \frac{1}{2}AIC^M(r) \ + \ \frac{1}{2}AIC^E(r) \quad \text{(say)}.
\end{aligned}$$

For $r \le k$, the exact difference between the minimized posterior risk $\rho(r,\ \boldsymbol{q}_r^*)$ and $AIC(r)/2$ is easy to obtain by Theorem 2.2. Note that, in the case $r = k$, it holds that $AIC^M(k)/2 = 0$ and $AIC^E(k)/2 = k/2$.

Thus, we have the following asymptotic difference and its limit as n goes to infinity :

Theorem 3.1 *The asymptotic difference between $\rho(r,\ \boldsymbol{q}_r^*)$ and $AIC(r)/2$ is represented by*

$$\frac{1}{2}AIC(r) - \rho(r,\ \boldsymbol{q}_r^*) \ = \ \frac{D_n^M(r) + D_n^E(r)}{n} + O(n^{-2}), \tag{3.2}$$

where the terms $D_n^M(r), D_n^E(r)$ are described by

$$\begin{aligned}
D_n^M(r) \ = \ & \frac{n}{2}\left\{\sum_{j=r+1}^{k+1} \frac{(\nu_j - 1)^2}{n_j} - \frac{(\sum_{j=r+1}^{k+1} \nu_j \ - 1)^2}{\sum_{j=r+1}^{k+1} n_j}\right\} \\
& + \frac{n}{2}\left\{\sum_{i=r+1}^{k+1} \frac{\nu_j - 1}{n_j + \nu_j - 1} - \frac{\sum_{j=r+1}^{k+1} \nu_j \ - 1}{\sum_{i=r+1}^{k+1}(n_j + \nu_j) - 1}\right\} \\
& - \frac{n}{12}\left\{\frac{\sum_{j=r+1}^{k+1} n_j}{(\sum_{j=r+1}^{k+1}(n_j + \nu_j) - 1)^2} - \sum_{j=r+1}^{k+1} \frac{n_j}{(n_j + \nu_j - 1)^2}\right\}, \\
D_n^E(r) \ = \ & n\left[\sum_{j=1}^{r} \frac{1}{12(n_j + \nu_j)} + \frac{1}{12\sum_{j=r+1}^{k+1}(n_j + \nu_j)}\right. \\
& \left. - \frac{1}{12\sum_{j=1}^{k+1}(n_j + \nu_j)} + \left(\frac{r}{2}\right)\frac{\sum_{j=1}^{k+1} \nu_j}{\sum_{j=1}^{k+1}(n_j + \nu_j)}\right].
\end{aligned}$$

And the limit of the difference (3.2) is zero as n goes to infinity.

PROOF. By the asymptotic description (2.16) of the modeling risk $\rho^M(r)$ in Theorem 2.3,

the asymptotic difference between $AIC^M(r)/2$ and $\rho^M(r)$ is

$$\begin{aligned}
\frac{1}{2}AIC^M(r) - \rho^M(r) &= \frac{1}{2}AIC^M(r) - BR^M(r) + O(n^{-3}) \\
&= -\sum_{j=r+1}^{k+1} n_j \log \frac{n_j+\nu_j-1}{n_j} + \left(\sum_{j=r+1}^{k+1} n_j\right) \log \frac{\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1}{\sum_{j=r+1}^{k+1} n_j} \\
&\quad + \frac{r-k}{2} - \left\{ \sum_{j=r+1}^{k+1} \frac{n_j}{2(n_j+\nu_j-1)} - \frac{\sum_{j=r+1}^{k+1} n_j}{2(\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1)} \right\} \\
&\quad - \left\{ \frac{\sum_{j=r+1}^{k+1} n_j}{12(\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1)^2} - \sum_{j=r+1}^{k+1} \frac{n_j}{12(n_j+\nu_j-1)^2} \right\} + O(n^{-3}).
\end{aligned}$$

By the Taylor expansion

$$\log(1+x) = x - \frac{x^2}{2} + O(x^3),$$

it holds that

$$\begin{aligned}
\log \frac{n_j+\nu_j-1}{n_j} &= \frac{\nu_j-1}{n_j} - \frac{1}{2}\left(\frac{\nu_j-1}{n_j}\right)^2 + O(n^{-3}), \\
\log \frac{\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1}{\sum_{j=r+1}^{k+1} n_j} &= \frac{\sum_{j=r+1}^{k+1}\nu_j - 1}{\sum_{j=r+1}^{k+1} n_j} - \frac{1}{2}\left(\frac{\sum_{j=r+1}^{k+1}\nu_j - 1}{\sum_{j=r+1}^{k+1} n_j}\right)^2 + O(n^{-3}),
\end{aligned}$$

and the transformations

$$\frac{n_j}{n_j+\nu_j-1} = 1 - \frac{\nu_j-1}{n_j+\nu_j-1}, \quad \frac{\sum_{j=r+1}^{k+1} n_j}{\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1} = 1 - \frac{\sum_{j=r+1}^{k+1}\nu_j - 1}{\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1}$$

hold, so that

$$\begin{aligned}
&\frac{1}{2}AIC^M(r) - \rho^M(r) \\
&= \frac{1}{2}\left\{ \sum_{j=r+1}^{k+1} \frac{(\nu_j-1)^2}{n_j} - \frac{(\sum_{j=r+1}^{k+1}\nu_j - 1)^2}{\sum_{j=r+1}^{k+1} n_j} \right\} \\
&\quad + \frac{1}{2}\left\{ \sum_{i=r+1}^{k+1} \frac{\nu_j-1}{n_j+\nu_j-1} - \frac{\sum_{j=r+1}^{k+1}\nu_j - 1}{\sum_{i=r+1}^{k+1}(n_j+\nu_j)-1} \right\} \\
&\quad - \frac{1}{12}\left\{ \frac{\sum_{j=r+1}^{k+1} n_j}{(\sum_{j=r+1}^{k+1}(n_j+\nu_j)-1)^2} - \sum_{j=r+1}^{k+1} \frac{n_j}{(n_j+\nu_j-1)^2} \right\} + O(n^{-2}).
\end{aligned}$$

Then the term $D_n^M(r)$ with respect to $AIC^M(r)/2 - \rho^M(r)$ is obtained. Similarly, by the asymptotic description (2.17) of the estimation risk $\rho^E(r)$ in Theorem 2.3 and by the transformation

$$\frac{n}{\sum_{j=1}^{k+1}(n_j+\nu_j)} = 1 - \frac{\sum_{j=1}^{k+1}\nu_j}{\sum_{j=1}^{k+1}(n_j+\nu_j)},$$

the asymptotic difference between $AIC^E(r)/2$ and $\rho^E(r)$ is

$$\begin{aligned}
\frac{1}{2}AIC^E(r) - \rho^E(r) &= \frac{1}{2}AIC^E(r) - BR^E(r) + O(n^{-3}) \\
&= \frac{r}{2} - \frac{r}{2}\left\{1 - \frac{\sum_{j=1}^{k+1}\nu_j}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\right\} \\
&\quad + \left\{\sum_{j=1}^{r}\frac{1}{12\,(n_j+\nu_j)} + \frac{1}{12\,\sum_{j=r+1}^{k+1}(n_j+\nu_j)} - \frac{1}{12\,\sum_{j=1}^{k+1}(n_j+\nu_j)}\right\} \\
&\quad - \frac{\sum_{j=1}^{k+1}\nu_j}{\sum_{j=1}^{k+1}(n_j+\nu_j)}\left\{\sum_{j=1}^{r}\frac{1}{12\,(n_j+\nu_j)} + \frac{1}{12\,\sum_{j=r+1}^{k+1}(n_j+\nu_j)} - \frac{1}{12\,\sum_{j=1}^{k+1}(n_j+\nu_j)}\right\} \\
&\quad + O(n^{-3}).
\end{aligned}$$

Then the term $D_n^E(r)$ with respect to $AIC^E(r)/2 - \rho^E(r)$ is obtained. Since it is easy to check that $D_n^M(r) = O(1)$ and $D_n^E(r) = O(1)$, the limit of the difference (3.2) is zero as n goes to infinity. Thus the proof is completed. □

Corollary 3.1

$$\frac{1}{2}AIC(r) - \rho_U(r) = \frac{UD_n^M(r) + UD_n^E(r)}{n} + O(n^{-2}),$$

where the terms $UD_n^M(r), UD_n^E(r)$ are described by

$$\begin{aligned}
&UD_n^M(r) \\
&= -\frac{n}{2}\left[\frac{(k-r)^2}{\sum_{j=r+1}^{k+1} n_j} + \frac{k-r}{\sum_{j=r+1}^{k+1} n_j + k - r}\right] - \frac{n}{12}\left[\frac{\sum_{j=r+1}^{k+1} n_j}{(\sum_{j=r+1}^{k+1} n_j + k - r)^2} - \sum_{j=r+1}^{k+1}\frac{1}{n_j}\right], \\
&UD_n^E(r) \\
&= n\left[\sum_{j=1}^{r}\frac{1}{12(n_j+1)} + \frac{1}{12\sum_{j=r+1}^{k+1}(n_j+1)} - \frac{1}{12\sum_{j=1}^{k+1}(n_j+1)} + \left(\frac{r}{2}\right)\frac{k+1}{n+k+1}\right].
\end{aligned}$$

And its limit is same as one in Theorem 3.1.

Inagaki(1977a) demonstrated the above corollary in the comparison between AIC and Bayes risk function whose the prior is the uniform distribution.

The following theorem is easily derived from Theorem 3.1, where we remark

$$\widehat{p}_j = \frac{n_j}{n} \longrightarrow p_j \qquad \text{as} \quad n \to \infty,$$

for $j = 1, \ldots, k+1$:

Theorem 3.2 *The second order term of the difference of $AIC(r)/2 - \rho(r,\ \boldsymbol{q}_r^*)$ is*

$$n\left[\frac{1}{2}AIC(r) - \rho(r,\ \boldsymbol{q}_r^*)\right] = \{D_n^M(r) + D_n^E(r)\} + O(n^{-1}),$$

and

$$D_n^M(r) \;\rightarrow\; \sum_{j=r+1}^{k+1} \frac{6\,\nu_j(\nu_j-1)+1}{12\,p_j} - \frac{6(\sum_{j=r+1}^{k+1}\nu_j)(\sum_{j=r+1}^{k+1}\nu_j\;-1)+1}{12\,\sum_{i=r+1}^{k+1}p_i},$$

$$D_n^E(r) \;\rightarrow\; \sum_{j=1}^{r}\frac{1}{12\,p_j} + \frac{1}{12\sum_{j=r+1}^{k+1}p_j} - \frac{1}{12} + \frac{r}{2}\,(\sum_{i=1}^{k+1}\nu_i),$$

as $n \rightarrow \infty$.

Corollary 3.2

$$n\left[\frac{1}{2}AIC(r) - \rho_U(r)\right] \;=\; \{UD_n^M(r) + UD_n^E(r)\} + O(n^{-1}),$$

and

$$UD_n^M(r) \;\rightarrow\; \sum_{j=r+1}^{k+1}\frac{1}{12p_j} - \frac{6(k+1-r)(k-r)+1}{12\,\sum_{i=r+1}^{k+1}p_i},$$

$$UD_n^E(r) \;\rightarrow\; \sum_{j=1}^{r}\frac{1}{12\,p_j} + \frac{1}{12\sum_{j=r+1}^{k+1}p_j} - \frac{1}{12} + \frac{r(k+1)}{2}$$

as $n \rightarrow \infty$.

Note that the asymptotic descriptions $BR^M(r)$ and $BR^E(r)$ in Theorem 2.3 are similar to $AIC^M(r)/2$ and $AIC^E(r)$ in (3.1), respectively, but the former retains the information by the prior density. It is interesting that the influence of prior does not appear in the first order term of the difference between $\rho(r,\, \boldsymbol{q}_r^*)$ and $AIC(r)/2$ in Theorem 3.1, but it does appear in the second order term of its difference in Theorem 3.2.

4 Simulation Our simulation is carried out by the software *Mathematica* (©Wolfram Research, Inc.) for Corollary 2.1 and equation (3.1). In *Mathematica*, we set that the value of random seed is 1997, that the cases of (n, k) are $(50, 5), (100, 10), (150, 15), (200, 20)$, and that the patterns of (n_{k-1}, n_k, n_{k+1}) are in the table 2. Remark that n means a number of sample and k a maximum number of parameter and that, with respect to a random number, we generates uniformly random numbers for the terms from 1 to $k-2$. Additionally, since an important point in these criteria should use not the values themselves but the relative values, when $r < k$, both $AIC/2$ and ρ_U in Table 2 are assumed to be relative differences of $AIC/2(r)$ and $\rho_U(r)$ from $AIC/2(k)$ and $\rho_U(k)$ respectively, that is, $AIC/2 = AIC/2(r) - AIC/2(k)$, $\rho_U = \rho_U(r) - \rho_U(k)$, and, when $r = k$, both $AIC/2$ and ρ_U are assumed to be zero, that is, $AIC/2 = \rho_U = 0$. Also let $Diff$ in Table 2 mean the exact difference of $AIC/2(r) - \rho_U(r)$. Note that the values of simulation are rounded.

n_i			n	k	$AIC/2$		ρ_U		$Diff$		
k-1	k	k+1			k-2	k-1	k-2	k-1	k-2	k-1	k
2	2	1	50	5	-1.782	-0.830	-1.288	-0.571	-0.128	0.107	0.366
			100	10	-1.782	-0.830	-1.295	-0.575	0.145	0.376	0.631
			150	15	-1.782	-0.830	-1.298	-0.576	0.409	0.638	0.893
			200	20	-1.782	-0.830	-1.299	-0.576	0.679	0.908	1.161
3	2	1	50	5	-1.477	-0.830	-1.037	-0.571	-0.079	0.101	0.361
			100	10	-1.477	-0.830	-1.045	-0.575	0.195	0.372	0.627

n_i			n	k	$AIC/2$		ρ_U		$Diff$			
k-1	k	k+1			k-2	k-1	k-2	k-1	k-2	k-1	k	
			150	15	-1.477	-0.830	-1.047	-0.576	0.461	0.636	0.891	
			200	20	-1.477	-0.830	-1.048	-0.576	0.724	0.898	1.152	
4	2	1	50	5	-1.000	-0.830	-0.604	-0.571	-0.038	0.097	0.358	
			100	10	-1.000	-0.830	-0.611	-0.575	0.234	0.367	0.623	
			150	15	-1.000	-0.830	-0.614	-0.576	0.502	0.634	0.888	
			200	20	-1.000	-0.830	-0.615	-0.576	0.764	0.895	1.149	
5	2	1	50	5	-0.413	-0.830	-0.053	-0.571	-0.006	0.096	0.355	
			100	10	-0.413	-0.830	-0.060	-0.575	0.266	0.363	0.619	
			150	15	-0.413	-0.830	-0.062	-0.576	0.535	0.631	0.886	
			200	20	-0.413	-0.830	-0.064	-0.576	0.796	0.892	1.146	
6	2	1	50	5	0.249	-0.830	0.581	-0.571	0.023	0.096	0.355	
			100	10	0.249	-0.830	0.574	-0.575	0.293	0.361	0.617	
			150	15	0.249	-0.830	0.571	-0.576	0.564	0.631	0.885	
			200	20	0.249	-0.830	0.570	-0.576	0.822	0.889	1.142	
7	2	1	50	5	0.968	-0.830	1.276	-0.571	0.045	0.093	0.352	
			100	10	0.968	-0.830	1.268	-0.575	0.320	0.364	0.620	
			150	15	0.968	-0.830	1.266	-0.576	0.583	0.626	0.881	
			200	20	0.968	-0.830	1.264	-0.576	0.848	0.890	1.144	
3	3	1	50	5	-1.339	-0.477	-0.939	-0.265	-0.045	0.143	0.355	
			100	10	-1.339	-0.477	-0.947	-0.268	0.227	0.411	0.619	
			150	15	-1.339	-0.477	-0.949	-0.270	0.502	0.684	0.892	
			200	20	-1.339	-0.477	-0.951	-0.270	0.761	0.943	1.150	
4	3	1	50	5	-1.006	-0.477	-0.639	-0.265	-0.014	0.140	0.353	
			100	10	-1.006	-0.477	-0.647	-0.268	0.256	0.406	0.615	
			150	15	-1.006	-0.477	-0.649	-0.270	0.527	0.676	0.883	
			200	20	-1.006	-0.477	-0.651	-0.270	0.791	0.940	1.146	
5	3	1	50	5	-0.544	-0.477	-0.207	-0.265	0.011	0.137	0.349	*
			100	10	-0.544	-0.477	-0.214	-0.268	0.287	0.409	0.618	*
			150	15	-0.544	-0.477	-0.217	-0.270	0.551	0.672	0.879	*
			200	20	-0.544	-0.477	-0.218	-0.270	0.818	0.938	1.144	*
6	3	1	50	5	0.007	-0.477	0.321	-0.265	0.036	0.138	0.350	
			100	10	0.007	-0.477	0.313	-0.268	0.311	0.409	0.617	
			150	15	0.007	-0.477	0.311	-0.270	0.571	0.668	0.876	
			200	20	0.007	-0.477	0.310	-0.270	0.843	0.939	1.145	
7	3	1	50	5	0.625	-0.477	0.919	-0.265	0.054	0.136	0.348	
			100	10	0.625	-0.477	0.912	-0.268	0.327	0.405	0.613	
			150	15	0.625	-0.477	0.909	-0.270	0.591	0.668	0.876	
			200	20	0.625	-0.477	0.908	-0.270	0.858	0.934	1.141	
4	4	1	50	5	-0.797	-0.036	-0.458	0.143	0.009	0.169	0.348	
			100	10	-0.797	-0.036	-0.465	0.140	0.284	0.440	0.616	
			150	15	-0.797	-0.036	-0.468	0.138	0.550	0.705	0.879	
			200	20	-0.797	-0.036	-0.469	0.138	0.815	0.969	1.143	
5	4	1	50	5	-0.447	-0.036	-0.131	0.143	0.029	0.166	0.346	
			100	10	-0.447	-0.036	-0.139	0.140	0.301	0.434	0.609	
			150	15	-0.447	-0.036	-0.141	0.138	0.570	0.701	0.876	
			200	20	-0.447	-0.036	-0.142	0.138	0.832	0.963	1.137	
6	4	1	50	5	0.004	-0.036	0.300	0.143	0.049	0.166	0.345	*
			100	10	0.004	-0.036	0.293	0.140	0.326	0.439	0.615	*
			150	15	0.004	-0.036	0.290	0.138	0.597	0.709	0.884	*
			200	20	0.004	-0.036	0.289	0.138	0.852	0.963	1.137	*
7	4	1	50	5	0.531	-0.036	0.811	0.143	0.064	0.164	0.343	*
			100	10	0.531	-0.037	0.803	0.140	0.336	0.432	0.608	*
			150	15	0.531	-0.036	0.801	0.138	0.608	0.703	0.878	*
			200	20	0.531	-0.036	0.799	0.138	0.870	0.964	1.138	*
8	4	1	50	5	1.118	-0.036	1.383	0.143	0.079	0.165	0.344	*
			100	10	1.118	-0.036	1.376	0.140	0.350	0.432	0.607	*
			150	15	1.118	-0.036	1.373	0.138	0.620	0.700	0.875	*
			200	20	1.118	-0.036	1.372	0.138	0.884	0.963	1.137	*
5	5	1	50	5	-0.198	0.456	0.100	0.611	0.047	0.189	0.344	*
			100	10	-0.198	0.456	0.092	0.608	0.324	0.461	0.613	*
			150	15	-0.198	0.456	0.089	0.607	0.589	0.723	0.874	*
			200	20	-0.198	0.456	0.088	0.606	0.854	0.990	1.140	*
6	5	1	50	5	0.162	0.456	0.443	0.611	0.062	0.187	0.343	
			100	10	0.162	0.456	0.435	0.608	0.335	0.456	0.608	
			150	15	0.162	0.456	0.433	0.607	0.602	0.722	0.873	

n_i			n	k	$AIC/2$		ρ_U		$Diff$			
k-1	k	k+1			k-2	k-1	k-2	k-1	k-2	k-1	k	
			200	20	0.162	0.456	0.431	0.606	0.876	0.994	1.145	
7	5	1	50	5	0.606	0.456	0.872	0.611	0.077	0.187	0.343	
			100	10	0.606	0.456	0.865	0.608	0.351	0.457	0.609	
			150	15	0.606	0.456	0.862	0.607	0.616	0.721	0.872	
			200	20	0.606	0.456	0.861	0.606	0.884	0.988	1.138	
7	6	1	50	5	0.806	0.981	1.059	1.119	0.086	0.202	0.340	
			100	10	0.806	0.981	1.052	1.116	0.361	0.472	0.607	
			150	15	0.806	0.981	1.049	1.114	0.626	0.736	0.869	
			200	20	0.806	0.981	1.048	1.114	0.897	1.006	1.139	
8	7	1	50	5	1.473	1.531	1.706	1.655	0.111	0.219	0.343	
			100	10	1.473	1.531	1.698	1.652	0.379	0.483	0.604	
			150	15	1.473	1.531	1.696	1.650	0.645	0.748	0.868	
			200	20	1.473	1.531	1.695	1.650	0.913	1.016	1.134	
9	8	1	50	5	2.159	2.099	2.375	2.212	0.127	0.230	0.343	
			100	10	2.159	2.099	2.367	2.208	0.396	0.496	0.605	
			150	15	2.159	2.099	2.365	2.207	0.671	0.768	0.877	
			200	20	2.159	2.099	2.363	2.206	0.934	1.031	1.138	

Table 2: Result of Simulation

Each $AIC/2$ in the same cell pattern (n_{k-1}, n_k, n_{k+1}) is constant in spite of changes of sample number n and cell number k, because only the cell pattern determines the value of $AIC/2$ by the formulation (3.1) of the AIC statistic. On the other hand, all ρ_U is affected by all of the elements n_j $(j = 1, \ldots, k+1)$, because of the representations in Corollary 2.1. In the cell patterns $(4, 4, 1)$, $(5, 4, 1)$, $(8, 7, 1)$, ρ_U shows not the same behavior but the same smallest value with $AIC/2$. In the following patterns

$$(5, 3, 1),\ (6, 4, 1),\ (7, 4, 1),\ (8, 4, 1),\ (5, 5, 1),$$

which are marked by $*$ in the rightmost column of table 2, ρ_U has the smallest value which is different from $AIC/2$, and we find that these patterns have a kind of difficulty when the grouping of cells would be decided. In other cell patterns, ρ_U shows the same behavior and smallest value with $AIC/2$.

Thus we could mention that both the Bayes grouping rule and the AIC rule give so similar results as the usual grouping rule of small cells in large sample cases (see Theorem 3.1), while these rules give different decisions in not so large sample cases.

References

Akaike, H.(1973). Information Theory and an Extension of the Maximum Likelihood Principle, *2nd International Symposium on Information Theory*, 267–281, Akademiai Kiadó, Budapest.

Azzalini, A.(1996). *Statistical Inference*, Chapman & Hall.

Courant, R. and John, F.(1965). *Introduction to Calculus and Analysis, Volume 1*, Wiley International Edition.

Inagaki, N.(1977a). Manuscript for lecture in 45th annual meeting of Japan Statistical Society.

Inagaki, N.(1977b). Two Errors in Statistical Model Fitting, *Ann. Inst. Statist. Math.*, **29**, 131–152.

Rao, C.R.(1973). *Linear Statistical Inference and Its Applications, Second Edition*, John Wiley & Sons.

Wilks, S.S.(1962). *Mathematical Statistics*, John Wiley & Sons.

Department of Informatics and Mathematical Science,
Graduate School of Engineering Science, Osaka University,

1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan.

e-mail: kumagai@sigmath.es.osaka-u.ac.jp, inagaki@sigmath.es.osaka-u.ac.jp