

Asymptotically minimax Bayesian predictive densities for multinomial models

Fumiyasu Komaki

*Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
e-mail: komaki@mist.i.u-tokyo.ac.jp*

*RIKEN Brain Science Institute
2-1 Hirosawa, Wako City, Saitama 351-0198, Japan*

Abstract: One-step ahead prediction for the multinomial model is considered. The performance of a predictive density is evaluated by the average Kullback-Leibler divergence from the true density to the predictive density. Asymptotic approximations of risk functions of Bayesian predictive densities based on Dirichlet priors are obtained. It is shown that a Bayesian predictive density based on a specific Dirichlet prior is asymptotically minimax. The asymptotically minimax prior is different from known objective priors such as the Jeffreys prior or the uniform prior.

AMS 2000 subject classifications: Primary 62C10, 62C20; secondary 62F15.

Keywords and phrases: Dirichlet prior, Jeffreys prior, Kullback-Leibler divergence, latent information prior, reference prior.

Received March 2012.

1. Introduction

We consider one step ahead prediction for the multinomial model. Suppose that we observe a random variable $x = (x_1, x_2, \dots, x_{k-1})$ distributed according to the multinomial distribution

$$p(x|\theta) = \binom{N}{x_1, x_2, \dots, x_k} \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_k^{x_k}$$

where $x_k := N - \sum_{i=1}^{k-1} x_i$, $\theta = (\theta_1, \dots, \theta_{k-1})$, $\theta_k := 1 - \sum_{i=1}^{k-1} \theta_i$, and

$$\binom{N}{x_1, x_2, \dots, x_k} := \frac{N!}{x_1! x_2! \cdots x_k!}.$$

The parameter space is

$$\Delta := \left\{ \theta = (\theta_1, \theta_2, \dots, \theta_{k-1}) \mid \theta_i \geq 0 \ (i = 1, \dots, k), \ \theta_k := 1 - \sum_{i=1}^{k-1} \theta_i \right\}.$$

The objective is to predict y distributed according to the multinomial distribution

$$p(y|\theta) = \theta_1^{y_1} \theta_2^{y_2} \cdots \theta_k^{y_k}$$

with index 1, where $y = (y_1, \dots, y_{k-1})$ and $y_k := 1 - \sum_{i=1}^{k-1} y_i$, by using a predictive density $q(y; x)$.

The performance of a predictive density $q(y; x)$ is evaluated by the risk function

$$R(\theta, q(y; x)) = \sum_y \sum_x p(x, y|\theta) \log \frac{p(y|\theta)}{q(y; x)}, \quad (1.1)$$

which is the average Kullback-Leibler divergence from the true density $p(y|\theta)$ to the predictive density $q(y; x)$.

When a Dirichlet prior

$$\pi_a(\theta) d\theta_1 \cdots d\theta_{k-1} = \frac{\Gamma(A)}{\Gamma(a_1) \cdots \Gamma(a_k)} \theta_1^{a_1-1} \cdots \theta_k^{a_k-1} d\theta_1 \cdots d\theta_{k-1}, \quad (1.2)$$

where $A := \sum_{i=1}^k a_i$, $a = (a_1, \dots, a_k)$ and $a_i > 0$ for every i , is adopted, the posterior density and the Bayesian predictive density are given by

$$\begin{aligned} p_{\pi_a}(\theta|x) d\theta_1 \cdots d\theta_{k-1} \\ = \frac{\Gamma(N+A)}{\Gamma(x_1+a_1) \cdots \Gamma(x_k+a_k)} \theta_1^{x_1+a_1-1} \cdots \theta_k^{x_k+a_k-1} d\theta_1 \cdots d\theta_{k-1}, \end{aligned}$$

and

$$\begin{aligned} p_{\pi_a}(y|x) &= \int p(y|\theta) p_{\pi_a}(\theta|x) d\theta_1 \cdots d\theta_{k-1} = \frac{B(x_1+y_1+a_1, \dots, x_k+y_k+a_k)}{B(x_1+a_1, \dots, x_k+a_k)} \\ &= \frac{\sum_{i=1}^k (x_i+a_i) y_i}{N+A}, \end{aligned}$$

respectively, where

$$B(x_1, \dots, x_k) := \frac{\Gamma(x_1) \cdots \Gamma(x_k)}{\Gamma(\sum_{i=1}^k x_i)}.$$

We define

$$\bar{\pi}_\alpha(\theta) d\theta_1 \cdots d\theta_{k-1} = \frac{\Gamma(k\alpha)}{\{\Gamma(\alpha)\}^k} \theta_1^{\alpha-1} \cdots \theta_k^{\alpha-1} d\theta_1 \cdots d\theta_{k-1},$$

which is π_a with $a_1 = \cdots = a_k = \alpha$.

In the present paper, we consider the asymptotics as the sample size N goes to infinity, and construct a Bayesian predictive density based on a Dirichlet prior that is asymptotically minimax in the sense described below. It is known that a minimax predictive density for one step ahead prediction for the multinomial

model can be constructed by using a latent information prior defined as a prior maximizing the conditional mutual information between y and θ given x ; see [11]. The latent information priors for one step ahead prediction is very different from minimax priors in the reference prior formulation. However, the explicit form of such a prior is difficult to obtain, and we need to develop asymptotic methods.

We consider a sequence of parameter subspaces

$$\Delta_{\varepsilon_N} := \left\{ \theta = (\theta_1, \theta_2, \dots, \theta_{k-1}) \mid \theta_i \geq \varepsilon_N \ (i = 1, \dots, k), \ \theta_k := 1 - \sum_{i=1}^{k-1} \theta_i \right\},$$

where $\{\varepsilon_N\}$ is a decreasing sequence of real numbers such that $\lim_{N \rightarrow \infty} \varepsilon_N = 0$ and $0 < \varepsilon_N < 1/k$ for every N , to avoid singularity problems concerning the boundary of the original parameter space Δ . Then, $\Delta_{\varepsilon_N} \subset \Delta_{\varepsilon_{N+1}}$, $\lim_{n \rightarrow \infty} \Delta_{\varepsilon_n} = \Delta$, and $\theta_i \in [\varepsilon_N, 1 - (k - 1)\varepsilon_N]$. Increasing sequences of parameter subspaces converging to the original parameter space are often considered to construct asymptotic objective priors; see e. g. [2, 6], and [4].

Let $\pi_*^{(N)}$ be a prior on Δ_{ε_N} such that the corresponding Bayesian predictive density $p_{\pi_*^{(N)}}(y \mid x)$ is minimax with respect to the parameter space Δ_{ε_N} . Thus,

$$\sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, p_{\pi_*^{(N)}}(y \mid x)) = \inf_q \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, q(y; x)).$$

The existence of such a prior is guaranteed by Theorem 2 in [11], since $p_\pi(x) > 0$ for every x if $\pi \in \mathcal{P}(\Delta_{\varepsilon_N})$. Here, $\mathcal{P}(\Delta_{\varepsilon_N})$ is the set of all probability measures on Δ_{ε_N} .

We show that the Bayesian predictive density based on a Dirichlet prior $\bar{\pi}_{\hat{\alpha}}$ with $\hat{\alpha} := 1 + 1/\sqrt{6}$ is asymptotically minimax in the sense that

$$\left| \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, p_{\pi_*^{(N)}}(y \mid x)) - \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, p_{\bar{\pi}_{\hat{\alpha}}}(y \mid x)) \right| = o(N^{-2}) \quad (1.3)$$

if $\{\varepsilon_N\}$ satisfies appropriate conditions.

For example, when the model is binomial ($k = 2$), the minimax prior is $\theta^{1/\sqrt{6}}(1 - \theta)^{1/\sqrt{6}}/B(1 + 1/\sqrt{6}, 1 + 1/\sqrt{6})$ and is different from the Jeffreys prior $\theta^{-1/2}(1 - \theta)^{-1/2}/B(1/2, 1/2)$ or the uniform prior.

Although the multinomial model is relatively simple, the results in the present paper could be a prototype for further development of theories on other models.

Closely related but essentially different prediction problems have been extensively studied in the framework of reference prior and Bayes coding; see e. g. [8, 3, 6], and [4]. In this setting, the objective is to predict a large number of future observables without using data at hand. Roughly speaking, the Jeffreys prior is asymptotically minimax under suitable regularity conditions.

In contrast, we consider here one step ahead prediction by using N observed data at hand and consider the asymptotics as N goes to infinity. The priors attaining minimax prediction in these two settings are quite different; see [10]

and [11] for discussion on the relation between the two settings, and see [5] for various related approaches.

In Section 2, we obtain an asymptotic approximation of risk functions of Bayesian predictive densities based on Dirichlet priors. The approximation is uniform on Δ_{ε_N} . In Section 3, we prove that the Bayesian predictive density based on the Dirichlet prior $\bar{\pi}_{\hat{\alpha}}$ with $\hat{\alpha} := 1 + 1/\sqrt{6}$ is asymptotically minimax if $\{\varepsilon_N\}$ satisfies appropriate conditions. In Section 4, some discussion is given.

2. Asymptotic evaluation of the risk function

In this section, we obtain an asymptotic approximation for the risk function of a Bayesian predictive density based on a Dirichlet prior. This approximation is uniform on the set Δ_{ε_N} .

Observe that the risk function (1.1) of $p_{\pi_a}(y|x)$ based on π_a defined by (1.2) is given by

$$\begin{aligned} R(\theta, p_{\pi_a}(y|x)) &= \sum_{i=1}^k \theta_i \sum_{x_i=0}^N \binom{N}{x_i} \theta_i^{x_i} (1-\theta_i)^{N-x_i} \log \left(\frac{\theta_i}{\frac{x_i+a_i}{N+A}} \right) \\ &= \sum_i \theta_i \sum_{x_i=0}^N \binom{N}{x_i} \theta_i^{x_i} (1-\theta_i)^{N-x_i} \\ &\quad \times \left\{ -\log \left(\frac{N\theta_i + a_i}{N\theta_i + A\theta_i} \right) - \log \left(\frac{x_i + a_i}{N\theta_i + a_i} - 1 + 1 \right) \right\} \\ &= -\sum_i \theta_i \log(1 + s_i) - \sum_i \theta_i \sum_{x_i=0}^N \binom{N}{x_i} \theta_i^{x_i} (1-\theta_i)^{N-x_i} \log(w_i + 1), \quad (2.1) \end{aligned}$$

where

$$s_i := \frac{a_i - A\theta_i}{N\theta_i + A\theta_i} \quad \text{and} \quad w_i := \frac{x_i + a_i}{N\theta_i + a_i} - 1 = \frac{x_i - N\theta_i}{N\theta_i + a_i}.$$

Here, w_i ($i = 1, 2, \dots, k$) are random variables with $E_{\theta_i}(w_i) = 0$, and $\sum_{i=1}^k \theta_i s_i = 0$.

If we fix a true parameter value θ satisfying $\theta_i \in (0, 1)$ for all $i = 1, \dots, k$, then it is easy to verify that

$$R(\theta, p_{\pi_a}(y|x)) = \frac{k-1}{2N} + O(N^{-2}).$$

A higher order pointwise approximation of the risk function has been studied; see [9].

Here, instead of the pointwise approximation, we obtain an asymptotic approximation that is uniform for $\theta \in \Delta_{\varepsilon_N}$.

We use the following Lemmas 2.1–2.3 to prove Theorem 2.1. The proofs of the lemmas are given in the appendix.

Lemma 2.1. For every nonnegative integer m and every $x > -1$,

$$\sum_{i=1}^{2m} (-1)^{i-1} \frac{1}{i} x^i + \frac{1}{2m+1} \frac{x^{2m+1}}{1+x} \leq \log(1+x) \leq \sum_{i=1}^{2m+1} (-1)^{i-1} \frac{1}{i} x^i.$$

Lemma 2.2. Let $\mu_m(N, \theta)$ be the m -th central moments of the binomial distribution $\text{Bi}(N, \theta)$ with index N and parameter θ . Suppose that $\{\varepsilon_N\}$ be a decreasing sequence of real numbers such that $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, $\lim_{N \rightarrow \infty} N\varepsilon_N = \infty$, and $0 < \varepsilon_N < 1$ for every N .

- (1) For every positive integer l , there exists a positive constant C_{2l-1} such that $\frac{|\mu_{2l-1}(N, \theta)|}{(N\theta)^{l-1}} \leq C_{2l-1}$ for all $\theta \in [\varepsilon_N, 1]$ and N .
- (2) For every positive integer l , there exists a positive constant C_{2l} such that $\frac{|\mu_{2l}(N, \theta)|}{(N\theta)^l} \leq C_{2l}$ for all $\theta \in [\varepsilon_N, 1]$ and N .

Lemma 2.3. Let x be a random variable distributed according to the binomial distribution $\text{Bi}(N, \theta)$. Define

$$w = \frac{x - N\theta}{N\theta + a},$$

where a is a positive real number. Suppose that $\{\varepsilon_N\}$ be a decreasing sequence of real numbers such that $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, $\lim_{N \rightarrow \infty} N\varepsilon_N = \infty$, and $0 < \varepsilon_N < 1$ for every N . Then, for every nonnegative integer l , there exists a constant $C_{2l+1}^{(w)}$ such that

$$E_\theta \left(-\frac{w^{2l+1}}{1+w} \right) \leq \frac{1}{(N\theta)^l} C_{2l+1}^{(w)}$$

for all $\theta \in [\varepsilon_N, 1]$ and N .

By using the lemmas, we prove Theorem 2.1.

Theorem 2.1. Let $p_{\pi_a}(y | x)$ be a Bayesian predictive density based on a Dirichlet prior π_a defined by (1.2). Suppose that $\{\varepsilon_N\}$ be a decreasing sequence of real numbers such that $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, $\lim_{N \rightarrow \infty} N\varepsilon_N = \infty$, and $0 < \varepsilon_N < 1/k$ for every N . Then, the risk function $R(\theta, p_{\pi_a}(y | x))$ satisfies

$$\begin{aligned} & \sup_{\theta \in \Delta_{\varepsilon_N}} \left| R(\theta, p_{\pi_a}(y | x)) - \frac{k-1}{2N} - \frac{1}{N^2} \left\{ \sum_{i=1}^k \frac{1}{12\theta_i} (6a_i^2 - 12a_i + 5) \right. \right. & (2.2) \\ & \left. - \frac{1}{2}A^2 + A - \frac{1}{2}k + \frac{1}{12} \right\} - \frac{1}{N^3} \left\{ \sum_{i=1}^k \frac{1}{12\theta_i^2} (-4a_i^3 + 18a_i^2 - 24a_i + 9) \right. \\ & \left. + \sum_{i=1}^k \frac{1}{4\theta_i} (-6a_i^2 + 12a_i - 5) + \frac{1}{3}A^3 - A + \frac{1}{2}k \right\} \\ & \left. - \frac{1}{N^4} \left\{ \sum_{i=1}^k \frac{1}{120\theta_i^3} (30a_i^4 - 240a_i^3 + 660a_i^2 - 720a_i + 251) \right. \right. \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^k \frac{1}{2\theta_i^2} (4a_i^3 - 18a_i^2 + 24a_i - 9) + \sum_{i=1}^k \frac{1}{12\theta_i} (42a_i^2 - 84a_i + 35) \\
& - \left. \frac{1}{4}A^4 + A - \frac{1}{2}k - \frac{1}{120} \right\} \Big| = O(N^{-5}\varepsilon_N^{-4}).
\end{aligned}$$

Proof. From (2.1) and Lemma 2.1, we have

$$\begin{aligned}
R(\theta, p_{\pi_a}(y | x)) & \leq \sum_{i=1}^k \theta_i E_{\theta} \left\{ \sum_{l=1}^{10} \frac{1}{l} (-w_i)^l - \frac{1}{11} \frac{w_i^{11}}{1+w_i} \right\} \\
& + \sum_{i=1}^k \theta_i \left\{ \sum_{l=1}^4 \frac{1}{l} (-s_i)^l - \frac{1}{5} \frac{s_i^5}{1+s_i} \right\},
\end{aligned} \tag{2.3}$$

and

$$R(\theta, p_{\pi_a}(y | x)) \geq \sum_{i=1}^k \theta_i E_{\theta} \left\{ \sum_{l=1}^9 \frac{1}{l} (-w_i)^l \right\} + \sum_{i=1}^k \theta_i \left\{ \sum_{l=1}^5 \frac{1}{l} (-s_i)^l \right\}. \tag{2.4}$$

From Lemma 2.2, we have

$$|E_{\theta}(w_i^{2l-1})| = \frac{|\mu_{2l-1}(N, \theta_i)|}{(N\theta_i + a_i)^{2l-1}} \leq \frac{|\mu_{2l-1}(N, \theta_i)|}{(N\theta_i)^{2l-1}} \leq \frac{C_{2l-1}}{(N\theta_i)^l}, \tag{2.5}$$

and

$$|E_{\theta}(w_i^{2l})| = \frac{|\mu_{2l}(N, \theta_i)|}{(N\theta_i + a_i)^{2l}} \leq \frac{|\mu_{2l}(N, \theta_i)|}{(N\theta_i)^{2l}} \leq \frac{C_{2l}}{(N\theta_i)^l}, \tag{2.6}$$

for every $a_i > 0$.

Obviously, the inequality

$$\frac{1}{1+s_i} = \frac{N\theta_i + A\theta_i}{N\theta_i + a_i} \leq \frac{A}{a_i} \tag{2.7}$$

holds since $0 < \theta_i < 1$ and $0 < a_i < A$.

From (2.3), (2.5), (2.6), (2.7), $E_{\theta_i}(w_i) = 0$, $\sum_{i=1}^k \theta_i s_i = 0$, and Lemma 2.3, we have

$$\begin{aligned}
R(\theta, p_{\pi_a}(y | x)) & \leq \sum_{i=1}^k \theta_i E_{\theta} \left\{ \sum_{l=2}^8 \frac{1}{l} (-w_i)^l + \sum_{l=9}^{10} \frac{1}{l} (-w_i)^l \right\} \\
& + \sum_{i=1}^k \frac{C_{11}^{(w)}}{N^5 \theta_i^4} + \sum_{i=1}^k \theta_i \sum_{l=2}^4 \frac{1}{l} (-s_i)^l + \frac{1}{5} \sum_{i=1}^k \theta_i \frac{A}{a_i} \left| \frac{a_i - A\theta_i}{N\theta_i + A\theta_i} \right|^5 \\
& \leq \sum_{i=1}^k \theta_i E_{\theta} \left\{ \sum_{l=2}^8 \frac{1}{l} (-w_i)^l \right\} + \sum_{i=1}^k \theta_i \sum_{l=2}^4 \frac{1}{l} (-s_i)^l + \frac{C'}{N^5 \varepsilon_N^4},
\end{aligned} \tag{2.8}$$

where C' is a positive constant not depending on N or θ .

In a similar way, from (2.4) and (2.5), we have

$$R(\theta, p_{\pi_a}(y | x)) \geq \sum_{i=1}^k \theta_i E_{\theta} \left\{ \sum_{l=2}^8 \frac{1}{l} (-w_i)^l \right\} + \sum_{i=1}^k \theta_i \sum_{l=2}^4 \frac{1}{l} (-s_i)^l - \frac{C'''}{N^5 \varepsilon_N^4}, \tag{2.9}$$

where C''' is a positive constant not depending on N or θ .

The first to eighth central moments of the binomial distribution $\text{Bi}(N, \theta)$ are given by

$$\begin{aligned} \mu_1(N, \theta) &= 0, & \mu_2(N, \theta) &= N\theta(1 - \theta), & \mu_3(N, \theta) &= N\theta(1 - \theta)(1 - 2\theta), \\ \mu_4(N, \theta) &= 3N^2\theta^2(1 - \theta)^2 + N\theta(1 - \theta)(1 - 6\theta + 6\theta^2), \\ \mu_5(N, \theta) &= 10N^2\theta^2(1 - \theta)^2(1 - 2\theta) + N\theta(1 - \theta)(1 - 2\theta)(1 - 12\theta + 12\theta^2), \\ \mu_6(N, \theta) &= 15N^3\theta^3(1 - \theta)^3 + 5N^2\theta^2(1 - \theta)^2(5 - 26\theta + 26\theta^2) + N\theta\phi_{6,1}(\theta), \\ \mu_7(N, \theta) &= 105N^3\theta^3(1 - \theta)^3(1 - 2\theta) + N^2\theta^2\phi_{7,2}(\theta) + N\theta\phi_{7,1}(\theta), \\ \mu_8(N, \theta) &= 105N^4\theta^4(1 - \theta)^4 + N^3\theta^3\phi_{8,3}(\theta) + N^2\theta^2\phi_{8,2}(\theta) + N\theta\phi_{8,1}(\theta), \end{aligned} \tag{2.10}$$

where $\phi_{i,j}(\theta)$ ($(i, j) = (6, 1), (7, 1), (7, 2), (8, 1), (8, 2), (8, 3)$) are polynomials of θ .

Therefore, by using (2.8), (2.9), (2.10), and the inequalities

$$\begin{aligned} \frac{1}{N\theta_i} \sum_{l=1}^{2m-1} \left(-\frac{a_i}{N\theta_i}\right)^l &= \frac{1}{N\theta_i} \frac{1 - \left(\frac{a_i}{N\theta_i}\right)^{2m}}{1 + \frac{a_i}{N\theta_i}} \leq \frac{1}{N\theta_i + a_i} \\ &\leq \frac{1}{N\theta_i} \sum_{l=1}^{2m} \left(-\frac{a_i}{N\theta_i}\right)^l = \frac{1}{N\theta_i} \frac{1 + \left(\frac{a_i}{N\theta_i}\right)^{2m+1}}{1 + \frac{a_i}{N\theta_i}}, \end{aligned}$$

we obtain (2.2) by a straightforward but lengthy calculation. In addition to the calculation by hand, the result is verified by using a computer algebra software. \square

From Theorem 2.1, we obtain the following corollaries.

Corollary 2.1. *Suppose that $\{\varepsilon_N\}$ be a decreasing sequence of real numbers such that $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, $\lim_{N \rightarrow \infty} N^{3/4} \varepsilon_N = \infty$, and $0 < \varepsilon_N < 1/k$ for every N . Then,*

$$\begin{aligned} \sup_{\theta \in \Delta_{\varepsilon_N}} \left| R(\theta, p_{\pi_a}) - \frac{k-1}{2N} - \frac{1}{N^2} \left\{ \sum_{i=1}^k \frac{1}{12\theta_i} (6a_i^2 - 12a_i + 5) \right. \right. & \tag{2.11} \\ \left. \left. - \frac{1}{2}A^2 + A - \frac{1}{2}k + \frac{1}{12} \right\} - \frac{1}{N^3} \sum_{i=1}^k \frac{1}{12\theta_i^2} (-4a_i^3 + 18a_i^2 - 24a_i + 9) \right. \\ \left. - \frac{1}{N^4} \sum_{i=1}^k \frac{1}{120\theta_i^3} (30a_i^4 - 240a_i^3 + 660a_i^2 - 720a_i + 251) \right| &= o(N^{-2}). \end{aligned}$$

Proof. Since $\lim_{N \rightarrow \infty} N^{3/4} \varepsilon_N = \infty$, $|N^{-3} \theta_i^{-1}| \leq N^{-3} \varepsilon_N^{-1} = o(N^{-2})$, $|N^{-4} \theta_i^{-1}| \leq N^{-4} \varepsilon_N^{-1} = o(N^{-3})$, $|N^{-4} \theta_i^{-2}| \leq N^{-4} \varepsilon_N^{-2} = o(N^{-2})$, and $N^{-5} \varepsilon_N^{-4} = N^{-2} (N^{-3/4} \varepsilon_N^{-1})^4 = o(N^{-2})$, we obtain (2.11) from Theorem 2.1. \square

Corollary 2.2. *Suppose that $\{\varepsilon_N\}$ be a decreasing sequence of real numbers such that $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, $\lim_{N \rightarrow \infty} N^{3/4} \varepsilon_N = \infty$, and $0 < \varepsilon_N < 1/k$ for every N . Then, the risk function of the Bayesian predictive density based on a Dirichlet prior $\bar{\pi}_{\hat{\alpha}}(\theta)$, where $\hat{\alpha} := 1 + 1/\sqrt{6}$, satisfies*

$$\begin{aligned} \sup_{\theta \in \Delta_{\varepsilon_N}} \left| R(\theta, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) - \frac{k-1}{2N} + \frac{1}{N^2} \frac{k-1}{12} \left\{ 1 + (7 + 2\sqrt{6})k \right\} \right. \\ \left. + \frac{1}{N^3} \frac{1}{18\sqrt{6}} \sum_{i=1}^k \frac{1}{\theta_i^2} + \frac{1}{N^4} \left(\frac{1}{6\sqrt{6}} - \frac{11}{720} \right) \sum_{i=1}^k \frac{1}{\theta_i^3} \right| = o(N^{-2}) \end{aligned} \quad (2.12)$$

and

$$\sup_{\theta \in \Delta_{\varepsilon_N}} \left(R(\theta, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) - \frac{k-1}{2N} + \frac{1}{N^2} \frac{k-1}{12} \left\{ 1 + (7 + 2\sqrt{6})k \right\} \right) = o(N^{-2}). \quad (2.13)$$

Proof. We have (2.12) from Corollary 2.1 because $6\hat{\alpha}^2 - 12\hat{\alpha} + 5 = 0$, $-4\hat{\alpha}^3 + 18\hat{\alpha}^2 - 24\hat{\alpha} + 9 = -\sqrt{6}/9$, $30\hat{\alpha}^4 - 240\hat{\alpha}^3 + 660\hat{\alpha}^2 - 720\hat{\alpha} + 251 = -(20\sqrt{6} - 11)/6$, and $-\hat{A}^2/2 + \hat{A} - k/2 + 1/12 = -(k-1)\{1 + (7 + 2\sqrt{6})k\}/12$, where $\hat{A} := k\hat{\alpha}$.

The equality (2.13) is directly obtained from (2.12) because $1/(6\sqrt{6}) - 11/720 > 0$. \square

We see that the Bayesian predictive density $p_{\pi_J}(y | x)$ based on the Jeffreys prior π_J is not asymptotically minimax in a predictive sense. The Jeffreys prior π_J is a Dirichlet prior $\bar{\pi}_{\alpha}$ with $\alpha = 1/2$. Thus, $6\alpha^2 - 12\alpha + 5 = 1/2$, and $-A^2/2 + A - k/2 + 1/12 = -(3k^2 - 2)/24$, where $A = k\alpha = k/2$. Thus, from Theorem 2.1, we have

$$\begin{aligned} \sup_{\theta \in \Delta_{\varepsilon_N}} \left| R(\theta, p_{\pi_J}(y | x)) - \frac{k-1}{2N} - \frac{1}{N^2} \left\{ \sum_{i=1}^k \frac{1}{24\theta_i} - \frac{1}{24}(3k^2 - 2) \right\} \right| \\ = o(N^{-2} \varepsilon_N^{-1}). \end{aligned}$$

By putting $\theta_1 = \varepsilon_N$ and $\theta_i = (1 - \varepsilon_N)/(k-1)$ ($i = 2, \dots, k$), we have

$$\sup_{\theta \in \Delta_{\varepsilon_N}} \left\{ R(\theta, p_{\pi_J}(y | x)) - \frac{k-1}{2N} \right\} \geq \frac{1}{24} \frac{1}{N^2 \varepsilon_N} + o(N^{-2} \varepsilon_N^{-1}).$$

Therefore, $p_{\pi_J}(y | x)$ is not asymptotically minimax.

From Corollary 2.2, we obtain Corollary 2.3, which is used to prove Theorem 3.2 in the next section.

We define

$$\pi_\alpha^{(N)}(\theta) = \begin{cases} \frac{\pi_\alpha(\theta)}{\int_{\Delta_{\varepsilon_N}} \pi_\alpha(\theta) d\theta_1 \cdots d\theta_{k-1}}, & \theta \in \Delta_{\varepsilon_N} \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\bar{\pi}_\alpha^{(N)}(\theta) = \begin{cases} \frac{\bar{\pi}_\alpha(\theta)}{\int_{\Delta_{\varepsilon_N}} \bar{\pi}_\alpha(\theta) d\theta_1 \cdots d\theta_{k-1}}, & \theta \in \Delta_{\varepsilon_N} \\ 0, & \text{otherwise.} \end{cases}$$

The Bayes risk of a predictive density $q(y; x)$ with respect to a prior π is denoted by

$$R(\pi, q(y; x)) := \int \pi(\theta) R(\theta, q(y; x)) d\theta.$$

Corollary 2.3. *Suppose that $\{\varepsilon_N\}$ is a decreasing real number sequence such that $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, $\lim_{N \rightarrow \infty} N^{3/4} \varepsilon_N = \infty$, and $0 < \varepsilon_N < 1/k$ for every N . Then,*

$$\begin{aligned} R(\bar{\pi}_{\hat{\alpha}}^{(N)}, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) &= \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) + o(N^{-2}) \\ &= \frac{k-1}{2N} - \frac{k-1}{12N^2} \{1 + (7 + 2\sqrt{6})k\} + o(N^{-2}). \end{aligned}$$

Proof. From (2.12), we obtain

$$\begin{aligned} R(\bar{\pi}_{\hat{\alpha}}^{(N)}, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) &= \int_{\Delta_{\varepsilon_N}} \bar{\pi}_{\hat{\alpha}}^{(N)}(\theta) R(\theta, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) d\theta \\ &= \frac{k-1}{2N} - \frac{k-1}{12N^2} \{1 + (7 + 2\sqrt{6})k\} \\ &\quad + \int_{\Delta_{\varepsilon_N}} \bar{\pi}_{\hat{\alpha}}^{(N)}(\theta) \left\{ -\frac{1}{N^3} \frac{1}{18\sqrt{6}} \sum_{i=1}^k \frac{1}{\theta_i^2} - \frac{1}{N^4} \left(\frac{1}{6\sqrt{6}} - \frac{11}{720} \right) \sum_{i=1}^k \frac{1}{\theta_i^3} \right\} d\theta. \end{aligned}$$

Here, we have

$$\begin{aligned} &\left| \int_{\Delta_{\varepsilon_N}} \bar{\pi}_{\hat{\alpha}}^{(N)}(\theta) \left\{ -\frac{1}{N^3} \frac{1}{18\sqrt{6}} \sum_{i=1}^k \frac{1}{\theta_i^2} - \frac{1}{N^4} \left(\frac{1}{6\sqrt{6}} - \frac{11}{720} \right) \sum_{i=1}^k \frac{1}{\theta_i^3} \right\} d\theta \right| \\ &= C \int_{\Delta_{\varepsilon_N}} \bar{\pi}_{\hat{\alpha}}(\theta) \left\{ \frac{1}{N^3} \frac{1}{18\sqrt{6}} \sum_{i=1}^k \frac{1}{\theta_i^2} + \frac{1}{N^4} \left(\frac{1}{6\sqrt{6}} - \frac{11}{720} \right) \sum_{i=1}^k \frac{1}{\theta_i^3} \right\} d\theta \\ &\leq \frac{kC}{18\sqrt{6}N^3} \int_{\theta_1 \geq \varepsilon_N} \bar{\pi}_{\hat{\alpha}}(\theta) \frac{1}{\theta_1^2} d\theta_1 + \frac{kC}{N^4} \left(\frac{1}{6\sqrt{6}} - \frac{11}{720} \right) \int_{\theta_1 \geq \varepsilon_N} \bar{\pi}_{\hat{\alpha}}(\theta) \frac{1}{\theta_1^3} d\theta_1, \end{aligned}$$

where $C = 1 / \int_{\Delta_{\varepsilon_N}} \bar{\pi}_{\hat{\alpha}}(\theta) d\theta_1 \cdots d\theta_{k-1}$.

Since the marginal density of θ_1 of the Dirichlet prior $\bar{\pi}_{\hat{\alpha}}$ is the beta density $\theta_1^{\hat{\alpha}-1}(1-\theta_1)^{(k-1)\hat{\alpha}-1}/B(\hat{\alpha}, (k-1)\hat{\alpha})$, we have

$$\begin{aligned} & \left| \int_{\Delta_{\varepsilon_N}} \bar{\pi}_{\hat{\alpha}}^{(N)}(\theta) \left\{ -\frac{1}{N^3} \frac{1}{18\sqrt{6}} \sum_{i=1}^k \frac{1}{\theta_i^2} - \frac{1}{N^4} \left(\frac{1}{6\sqrt{6}} - \frac{11}{720} \right) \sum_{i=1}^k \frac{1}{\theta_i^3} \right\} d\theta \right| \\ & \leq \frac{kC}{N^3} \frac{1}{18\sqrt{6}} \int_{\theta_1 \geq \varepsilon_N} \frac{\theta_1^{\hat{\alpha}-1}(1-\theta_1)^{(k-1)\hat{\alpha}-1}}{B(\hat{\alpha}, (k-1)\hat{\alpha})} \frac{1}{\theta_1^2} d\theta_1 \\ & \quad + \frac{kC}{N^4} \left(\frac{1}{6\sqrt{6}} - \frac{11}{720} \right) \int_{\theta_1 \geq \varepsilon_N} \frac{\theta_1^{\hat{\alpha}-1}(1-\theta_1)^{(k-1)\hat{\alpha}-1}}{B(\hat{\alpha}, (k-1)\hat{\alpha})} \frac{1}{\theta_1^3} d\theta_1 \\ & \leq \frac{kC}{N^3} \frac{1}{18\sqrt{6}} \frac{1}{2-\hat{\alpha}} \frac{1}{\varepsilon_N^{2-\hat{\alpha}}} \frac{1}{B(\hat{\alpha}, (k-1)\hat{\alpha})} \\ & \quad + \frac{kC}{N^4} \left(\frac{1}{6\sqrt{6}} - \frac{11}{720} \right) \frac{1}{3-\hat{\alpha}} \frac{1}{\varepsilon_N^{3-\hat{\alpha}}} \frac{1}{B(\hat{\alpha}, (k-1)\hat{\alpha})} \\ & = o(N^{-2}) \end{aligned}$$

because $O(N^3\varepsilon_N^{2-\hat{\alpha}}) \geq O(N^2(N\varepsilon_N)) \geq O(N^2)$ and $O(N^4\varepsilon_N^{3-\hat{\alpha}}) = O((N^3\varepsilon_N^{2-\hat{\alpha}})(N\varepsilon_N)) \geq O(N^2)$. \square

3. Minimax predictive densities

In this section, we prove that the Bayesian predictive density based on a Dirichlet prior $\bar{\pi}_{\hat{\alpha}}$, where $\hat{\alpha} := 1 + 1/\sqrt{6}$, is asymptotically minimax in the sense of (1.3) if $\{\varepsilon_N\}$ satisfies appropriate conditions.

The Bayesian predictive density with respect to the prior $\bar{\pi}_{\alpha}$ is given by

$$p_{\bar{\pi}_{\alpha}}(y | x) = \frac{B(x_1 + y_1 + \alpha, \dots, x_k + y_k + \alpha)}{B(x_1 + \alpha, \dots, x_k + \alpha)} = \frac{\sum_{i=1}^k x_i y_i + \alpha}{N + k\alpha}, \quad (3.1)$$

and that with respect to the prior $\bar{\pi}_{\alpha}^{(N)}$ is given by

$$\begin{aligned} p_{\bar{\pi}_{\alpha}^{(N)}}(y|x) &= \frac{B_{\Delta_{\varepsilon_N}}(x_1 + y_1 + \alpha, \dots, x_k + y_k + \alpha)}{B_{\Delta_{\varepsilon_N}}(x_1 + \alpha, \dots, x_k + \alpha)} \\ &= \frac{\sum_{i=1}^k x_i y_i + \alpha}{N + k\alpha} \frac{I_{\Delta_{\varepsilon_N}}(x_1 + y_1 + \alpha, \dots, x_k + y_k + \alpha)}{I_{\Delta_{\varepsilon_N}}(x_1 + \alpha, \dots, x_k + \alpha)}, \end{aligned} \quad (3.2)$$

where we define

$$B_{\Delta_{\varepsilon}}(\alpha_1, \dots, \alpha_k) := \int_{\Delta_{\varepsilon}} \theta_1^{\alpha_1-1} \dots \theta_k^{\alpha_k-1} d\theta_1 \dots d\theta_{k-1}$$

and

$$I_{\Delta_{\varepsilon}}(\alpha_1, \dots, \alpha_k) := \frac{B_{\Delta_{\varepsilon}}(\alpha_1, \dots, \alpha_k)}{B(\alpha_1, \dots, \alpha_k)}$$

for $\alpha_i > 0$ ($i = 1, \dots, k$) and $0 < \varepsilon < 1/k$. If $k = 2$, $I_{\Delta_\varepsilon}(\alpha_1, \alpha_2) = \{\int_\varepsilon^{1-\varepsilon} \theta^{\alpha_1-1}(1-\theta)^{\alpha_2-1}d\theta\}/B(\alpha_1, \alpha_2)$.

In the proof of minimaxity of prediction, the inequalities

$$\begin{aligned} \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, p_\pi(y | x)) &\geq \inf_q \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, q(y; x)) = \inf_q \sup_{\pi' \in \mathcal{P}(\Delta_{\varepsilon_N})} R(\pi', q(y; x)) \\ &\geq \sup_{\pi' \in \mathcal{P}(\Delta_{\varepsilon_N})} \inf_q R(\pi', q(y; x)) \geq \inf_q R(\pi^*, q(y; x)) = R(\pi^*, p_{\pi^*}(y | x)), \end{aligned} \tag{3.3}$$

which hold for every $\pi \in \mathcal{P}(\Delta)$ and $\pi^* \in \mathcal{P}(\Delta_{\varepsilon_N})$, play an essential role; see [7] for related inequalities in a very general setting. Each inequality in (3.3) is easy to verify. The last inequality in (3.3) is due to the fact, proved by Aitchison [1], that the Bayes risk of a predictive density with respect to a prior π^* is minimized when it is the Bayesian predictive density $p_{\pi^*}(y | x)$ based on π^* . Thus, by putting $\pi^* = \bar{\pi}_\alpha^{(N)}$ in (3.3), we have

$$R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x)) \geq \inf_q R(\bar{\pi}_\alpha^{(N)}, q(y; x)) = R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x)).$$

In the following, we first prove Theorem 3.1 that shows that the difference $R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x)) - R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x))$ is $O(N^{-1}\varepsilon_N^\alpha)$ if $\{\varepsilon_N\}$ satisfies appropriate conditions. Next, combining Corollary 2.3 and Theorem 3.1, we prove Theorem 3.2 showing that $p_{\bar{\pi}_\alpha^{(N)}}(y | x)$ is asymptotically minimax under suitable conditions.

Theorem 3.1. *Let $p_{\bar{\pi}_\alpha^{(N)}}(y | x)$ and $p_{\bar{\pi}_\alpha}(y | x)$ be predictive densities (3.1) and (3.2), respectively. Suppose that $\{\varepsilon_N\}$ is a decreasing sequence of real numbers such that $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, $\lim_{N \rightarrow \infty} N\varepsilon_N = \infty$, and $0 < \varepsilon_N < 1/k$ for every N . Then the difference of the Bayes risks of $p_{\bar{\pi}_\alpha^{(N)}}(y | x)$ and $p_{\bar{\pi}_\alpha}(y | x)$ with respect to $\bar{\pi}_\alpha^{(N)}$ satisfies*

$$R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x)) - R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x)) = O(N^{-1}\varepsilon_N^\alpha).$$

Theorem 3.1 means that the disadvantage of adopting a prior $\bar{\pi}_\alpha$ that does not satisfy $\int_{\Delta_{\varepsilon_N}} \bar{\pi}_\alpha(\theta)d\theta_1 \cdots d\theta_{k-1} = 1$ is asymptotically small.

We use Lemmas 3.1–3.5 below to prove Theorem 3.1. The proofs of the lemmas are given in the Appendix.

Lemma 3.1. *For every $\alpha_1 > 0, \dots, \alpha_k > 0$ and $0 < \varepsilon < 1/k$,*

$$\begin{aligned} I_{\Delta_\varepsilon}(\alpha_1 + 1, \alpha_2, \dots, \alpha_k) - I_{\Delta_\varepsilon}(\alpha_1, \dots, \alpha_k) \\ \leq \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\Gamma(\alpha_1 + 1)\Gamma(\sum_{i=2}^k \alpha_i)} \varepsilon^{\alpha_1} (1 - \varepsilon)^{\alpha_2 + \dots + \alpha_k}. \end{aligned}$$

Lemma 3.2. *If $0 \leq s < t \leq 1$, $0 \leq u < v \leq 1$, $s \leq u$, and $t \leq v$, then for all $\alpha > 0$ and $\beta > 0$,*

$$\frac{B_{[s,t]}(\alpha + 1, \beta)}{B_{[s,t]}(\alpha, \beta)} \leq \frac{B_{[u,v]}(\alpha + 1, \beta)}{B_{[u,v]}(\alpha, \beta)},$$

where

$$B_{[s,t]}(\alpha, \beta) := \int_s^t \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta.$$

Lemma 3.3. For every $\alpha_1 > 0, \dots, \alpha_k > 0$, and $0 < \varepsilon < 1/k$, the inequality

$$\frac{B_{\Delta_\varepsilon}(\alpha_1 + 1, \alpha_2, \dots, \alpha_k)}{B_{\Delta_\varepsilon}(\alpha_1, \alpha_2, \dots, \alpha_k)} \leq \frac{B_{[\varepsilon,1]}(\alpha_1 + 1, \alpha_2 + \dots + \alpha_k)}{B_{[\varepsilon,1]}(\alpha_1, \alpha_2 + \dots + \alpha_k)}$$

holds.

Lemma 3.4. For every $\alpha_1 > 0, \dots, \alpha_k > 0$, the equality

$$\begin{aligned} \sum_{x_2, \dots, x_k: \sum_{i=2}^k x_i = N - x_1} \frac{B(x_1 + \alpha_1, x_2 + \alpha_2, \dots, x_k + \alpha_k)}{B(\alpha_1, \alpha_2, \dots, \alpha_k)} \binom{N}{x_1, \dots, x_k} \\ = \frac{B(x_1 + \alpha_1, N - x_1 + \sum_{i=2}^k \alpha_i)}{B(\alpha_1, \sum_{i=2}^k \alpha_i)} \binom{N}{x_1} \end{aligned}$$

holds.

Lemma 3.5. For every $\alpha > 0, \beta > 0$, and $\varepsilon \in [0, 1)$, the inequality

$$\begin{aligned} \frac{B_{[\varepsilon,1]}(\alpha + 1, \beta)}{B_{[\varepsilon,1]}(\alpha, \beta)} &= \frac{\int_\varepsilon^1 \theta^\alpha(1-\theta)^{\beta-1}d\theta}{\int_\varepsilon^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta} = \frac{\alpha}{\alpha + \beta} + \frac{1}{\alpha + \beta} \frac{\varepsilon^\alpha(1-\varepsilon)^\beta}{B_{[\varepsilon,1]}(\alpha, \beta)} \\ &\leq \frac{(1-\varepsilon)\alpha}{\alpha + \beta} + \varepsilon \end{aligned}$$

holds.

By using the lemmas, we prove Theorem 3.1.

Proof of Theorem 3.1. From (3.1) and (3.2), the difference between the risk functions of $R(\theta, p_{\bar{\pi}_\alpha}(y | x))$ and $R(\theta, p_{\bar{\pi}_\alpha^{(N)}}(y | x))$ is given by

$$\begin{aligned} R(\theta, p_{\bar{\pi}_\alpha}(y | x)) - R(\theta, p_{\bar{\pi}_\alpha^{(N)}}(y | x)) &= \sum_x \sum_y p(x, y|\theta) \log \frac{p_{\bar{\pi}_\alpha^{(N)}}(y|x)}{p_{\bar{\pi}_\alpha}(y|x)} \\ &= \sum_x \sum_y \theta_1^{y_1} \dots \theta_k^{y_k} \binom{N}{x_1, \dots, x_k} \theta_1^{x_1} \theta_2^{x_2} \dots \theta_k^{x_k} \\ &\quad \times \log \frac{I_{\Delta_{\varepsilon N}}(x_1 + y_1 + \alpha, x_2 + y_2 + \alpha, \dots, x_k + y_k + \alpha)}{I_{\Delta_{\varepsilon N}}(x_1 + \alpha, \dots, x_k + \alpha)}. \end{aligned}$$

To evaluate the difference between the Bayes risks $R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x))$ and $R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x))$, it is sufficient to consider the case $y_1 = 1$ because of the symmetry of the index i . Thus,

$$\begin{aligned} 0 \leq R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x)) - R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x)) \\ = k \int_{\Delta_{\varepsilon N}} \frac{\theta_1^{\alpha-1} \dots \theta_k^{\alpha-1}}{B_{\Delta_{\varepsilon N}}(\alpha, \alpha, \dots, \alpha)} \theta_1 \sum_x \binom{N}{x_1, \dots, x_k} \theta_1^{x_1} \theta_2^{x_2} \dots \theta_k^{x_k} \end{aligned}$$

$$\begin{aligned} & \times \log \frac{I_{\Delta_{\varepsilon_N}}(x_1 + 1 + \alpha, x_2 + \alpha, \dots, x_k + \alpha)}{I_{\Delta_{\varepsilon_N}}(x_1 + \alpha, \dots, x_k + \alpha)} d\theta_1 \cdots d\theta_{k-1} \\ &= \frac{k}{B_{\Delta_{\varepsilon_N}}(\alpha, \alpha, \dots, \alpha)} \sum_x B_{\Delta_{\varepsilon_N}}(x_1 + 1 + \alpha, x_2 + \alpha, \dots, x_k + \alpha) \binom{N}{x_1, \dots, x_k} \\ & \times \log \frac{I_{\Delta_{\varepsilon_N}}(x_1 + 1 + \alpha, x_2 + \alpha, \dots, x_k + \alpha)}{I_{\Delta_{\varepsilon_N}}(x_1 + \alpha, \dots, x_k + \alpha)}. \end{aligned}$$

Because $\log(x + 1) \leq x$ for $x > -1$, we have

$$\begin{aligned} & R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x)) - R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x)) \\ & \leq \frac{k}{B_{\Delta_{\varepsilon_N}}(\alpha, \alpha, \dots, \alpha)} \sum_x B_{\Delta_{\varepsilon_N}}(x_1 + 1 + \alpha, x_2 + \alpha, \dots, x_k + \alpha) \binom{N}{x_1, \dots, x_k} \\ & \times \left\{ \frac{I_{\Delta_{\varepsilon_N}}(x_1 + 1 + \alpha, x_2 + \alpha, \dots, x_k + \alpha)}{I_{\Delta_{\varepsilon_N}}(x_1 + \alpha, \dots, x_k + \alpha)} - 1 \right\} \\ &= \frac{k}{B_{\Delta_{\varepsilon_N}}(\alpha, \alpha, \dots, \alpha)} \sum_x B(x_1 + \alpha, \dots, x_k + \alpha) \\ & \times \frac{B_{\Delta_{\varepsilon_N}}(x_1 + 1 + \alpha, x_2 + \alpha, \dots, x_k + \alpha)}{B_{\Delta_{\varepsilon_N}}(x_1 + \alpha, \dots, x_k + \alpha)} \binom{N}{x_1, \dots, x_k} \\ & \times \{ I_{\Delta_{\varepsilon_N}}(x_1 + 1 + \alpha, x_2 + \alpha, \dots, x_k + \alpha) - I_{\Delta_{\varepsilon_N}}(x_1 + \alpha, \dots, x_k + \alpha) \}. \end{aligned}$$

From Lemma 3.1, we obtain

$$\begin{aligned} & R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x)) - R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x)) \\ & \leq \frac{k}{B_{\Delta_{\varepsilon_N}}(\alpha, \alpha, \dots, \alpha)} \\ & \times \sum_x B(x_1 + \alpha, \dots, x_k + \alpha) \frac{B_{\Delta_{\varepsilon_N}}(x_1 + 1 + \alpha, x_2 + \alpha, \dots, x_k + \alpha)}{B_{\Delta_{\varepsilon_N}}(x_1 + \alpha, \dots, x_k + \alpha)} \\ & \times \binom{N}{x_1, \dots, x_k} \frac{\Gamma(N + k\alpha)}{\Gamma(x_1 + 1 + \alpha)\Gamma(N - x_1 + (k - 1)\alpha)} \\ & \times \varepsilon_N^{x_1 + \alpha} (1 - \varepsilon_N)^{N - x_1 + (k - 1)\alpha}. \end{aligned}$$

From Lemmas 3.3 and 3.4, we have

$$\begin{aligned} & R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x)) - R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x)) \\ & \leq \frac{k}{B_{\Delta_{\varepsilon_N}}(\alpha, \alpha, \dots, \alpha)} \\ & \times \sum_x B(x_1 + \alpha, \dots, x_k + \alpha) \frac{B_{[\varepsilon_N, 1]}(x_1 + 1 + \alpha, N - x_1 + (k - 1)\alpha)}{B_{[\varepsilon_N, 1]}(x_1 + \alpha, N - x_1 + (k - 1)\alpha)} \\ & \times \binom{N}{x_1, \dots, x_k} \frac{\Gamma(N + k\alpha)}{\Gamma(x_1 + 1 + \alpha)\Gamma(N - x_1 + (k - 1)\alpha)} \end{aligned}$$

$$\begin{aligned}
& \times \varepsilon_N^{x_1+\alpha} (1 - \varepsilon_N)^{N-x_1+(k-1)\alpha} \\
& = \frac{kB(\alpha, \alpha, \dots, \alpha)}{B_{\Delta_{\varepsilon_N}}(\alpha, \alpha, \dots, \alpha)B(\alpha, (k-1)\alpha)} \\
& \times \sum_{x_1=0}^N \frac{1}{x_1 + \alpha} \frac{B_{[\varepsilon_N, 1]}(x_1 + 1 + \alpha, N - x_1 + (k-1)\alpha)}{B_{[\varepsilon_N, 1]}(x_1 + \alpha, N - x_1 + (k-1)\alpha)} \\
& \times \binom{N}{x_1} \varepsilon_N^{x_1+\alpha} (1 - \varepsilon_N)^{N-x_1+(k-1)\alpha}.
\end{aligned}$$

Since

$$\frac{B_{[\varepsilon_N, 1]}(x_1 + 1 + \alpha, N - x_1 + (k-1)\alpha)}{B_{[\varepsilon_N, 1]}(x_1 + \alpha, N - x_1 + (k-1)\alpha)} \leq \frac{(1 - \varepsilon_N)(x_1 + \alpha)}{N + k\alpha} + \varepsilon_N$$

because of Lemma 3.5, we have

$$\begin{aligned}
& R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x)) - R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x)) \\
& \leq \frac{k\varepsilon_N^\alpha (1 - \varepsilon_N)^{(k-1)\alpha}}{I_{\Delta_{\varepsilon_N}}(\alpha, \dots, \alpha)B(\alpha, (k-1)\alpha)} \\
& \times \left\{ \frac{1 - \varepsilon_N}{N + k\alpha} + \varepsilon_N \sum_{x_1=0}^N \frac{1}{x_1 + \alpha} \binom{N}{x_1} \varepsilon_N^{x_1} (1 - \varepsilon_N)^{N-x_1} \right\} \\
& = \frac{k\varepsilon_N^\alpha (1 - \varepsilon_N)^{(k-1)\alpha}}{I_{\Delta_{\varepsilon_N}}(\alpha, \dots, \alpha)B(\alpha, (k-1)\alpha)} \left\{ \frac{1 - \varepsilon_N}{N + k\alpha} \right. \\
& \left. + \sum_{x_1=0}^N \frac{x_1 + 1}{x_1 + \alpha} \frac{1}{N + 1} \frac{(N + 1)!}{(x_1 + 1)!(N + 1 - x_1 - 1)!} \varepsilon_N^{x_1+1} (1 - \varepsilon_N)^{N+1-x_1-1} \right\} \\
& = \frac{k\varepsilon_N^\alpha (1 - \varepsilon_N)^{(k-1)\alpha}}{I_{\Delta_{\varepsilon_N}}(\alpha, \dots, \alpha)B(\alpha, (k-1)\alpha)} \left\{ \frac{1 - \varepsilon_N}{N + k\alpha} \right. \\
& \left. + \sum_{z=0}^{N+1} \frac{z}{z + \alpha - 1} \frac{1}{N + 1} \binom{N + 1}{z} \varepsilon_N^z (1 - \varepsilon_N)^{N+1-z} \right\},
\end{aligned}$$

where we define $z/(z + \alpha - 1) = 0$ if $\alpha = 1$ and $z = 0$.

Since there exists a constant $C_\alpha > 0$ such that $|z/(z + \alpha - 1)| < C_\alpha$ for every z , we have

$$\begin{aligned}
& R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha}(y | x)) - R(\bar{\pi}_\alpha^{(N)}, p_{\bar{\pi}_\alpha^{(N)}}(y | x)) \\
& \leq \frac{k\varepsilon_N^\alpha (1 - \varepsilon_N)^{(k-1)\alpha}}{I_{\Delta_{\varepsilon_N}}(\alpha, \dots, \alpha)B(\alpha, (k-1)\alpha)} \left(\frac{1 - \varepsilon_N}{N + k\alpha} + \frac{C_\alpha}{N + 1} \right) = O(N^{-1}\varepsilon_N^\alpha).
\end{aligned}$$

□

Now we prove Theorem 3.2 that shows $p_{\hat{\pi}_\alpha}(y | x)$, where $\hat{\alpha} = 1 + 1/\sqrt{6}$, is asymptotically minimax. The constant $1/\hat{\alpha} = \sqrt{6}/(\sqrt{6} + 1)$ in the theorem is approximately 0.7101.

Theorem 3.2. Let $p_{\bar{\pi}_{\hat{\alpha}}}(y | x)$ be the predictive density based on the prior

$$\bar{\pi}_{\hat{\alpha}}(\theta)d\theta_1 \cdots d\theta_{k-1} \propto \theta_1^{1/\sqrt{6}} \cdots \theta_{k-1}^{1/\sqrt{6}} (1 - \sum_{i=1}^{k-1} \theta_i)^{1/\sqrt{6}} d\theta_1 \cdots d\theta_{k-1}.$$

Suppose that $\{\varepsilon_N\}$ be a decreasing sequence of real numbers such that $\lim_{N \rightarrow \infty} N^{3/4} \varepsilon_N = \infty$, $\lim_{N \rightarrow \infty} N^{1/\hat{\alpha}} \varepsilon_N = 0$, and $0 < \varepsilon_N < 1/k$ for every N . Then,

$$\sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) = \inf_q \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, q(y; x)) + o(N^{-2}).$$

Proof. By setting $\pi = \bar{\pi}_{\hat{\alpha}}$ and $\pi^* = \bar{\pi}_{\hat{\alpha}}^{(N)}$ in (3.3), we obtain

$$\sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) \geq \inf_q \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, q(y; x)) \geq R(\bar{\pi}_{\hat{\alpha}}^{(N)}, p_{\bar{\pi}_{\hat{\alpha}}^{(N)}}(y | x)). \tag{3.4}$$

From Theorem 3.1, we have

$$R(\bar{\pi}_{\hat{\alpha}}^{(N)}, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) - R(\bar{\pi}_{\hat{\alpha}}^{(N)}, p_{\bar{\pi}_{\hat{\alpha}}^{(N)}}(y | x)) = O(N^{-1} \varepsilon_N^{\hat{\alpha}}) = o(N^{-2}), \tag{3.5}$$

because $\varepsilon_N = o(N^{-1/\hat{\alpha}})$. From (3.4) and (3.5), we have

$$\begin{aligned} \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) &\geq \inf_q \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, q(y; x)) \\ &\geq R(\bar{\pi}_{\hat{\alpha}}^{(N)}, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) + o(N^{-2}). \end{aligned} \tag{3.6}$$

Here, from Corollary 2.3,

$$R(\bar{\pi}_{\hat{\alpha}}^{(N)}, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) = \sup_{\theta \in \Delta_{\varepsilon_N}} R(\theta, p_{\bar{\pi}_{\hat{\alpha}}}(y | x)) + o(N^{-2}). \tag{3.7}$$

From (3.6) and (3.7), we obtain the desired equality. □

4. Discussion

The results in the present paper indicate that $\bar{\pi}_{\hat{\alpha}}(\theta) \propto \theta_1^{1/\sqrt{6}} \cdots \theta_{k-1}^{1/\sqrt{6}} (1 - \sum_{i=1}^{k-1} \theta_i)^{1/\sqrt{6}}$ could be a reasonable objective prior for one-step ahead prediction. The prior $\bar{\pi}_{\hat{\alpha}}(\theta)$ can be regarded as an asymptotic approximation to the latent information prior, based on which a minimax predictive density is constructed, and it seems to consistent with some numerical results in [11]. Bayesian predictive densities based on commonly used objective priors, such as the Jeffreys priors π_J on Δ , $\pi_J^{(N)}$ on Δ_{ε_N} , the uniform priors π_U on Δ , or $\pi_U^{(N)}$ on Δ_{ε_N} are not asymptotically minimax in the predictive sense.

The constant $\hat{\alpha} = 1 + 1/\sqrt{6}$ appears as a solution of the equation $6a_i^2 - 12a_i + 5 = 0$. Note that the term of order N^{-2} in (2.2) is $(1/N^2) \{ \sum_{i=1}^k (6a_i^2 - 12a_i + 5) / (12\theta_i) - A^2/2 + A - k/2 + 1/12 \}$ if the parameter value θ is fixed.

The conditions $\lim_{N \rightarrow \infty} N^{3/4} \varepsilon_N = \infty$ and $\lim_{N \rightarrow \infty} N^{1/\hat{\alpha}} \varepsilon_N = 0$ assumed in Theorem 3.2 are sufficient conditions. If ε_N converges to 0 very rapidly, then the condition $\lim_{N \rightarrow \infty} N^{3/4} \varepsilon_N = \infty$ is not satisfied and we need to take into consideration the singularity at the boundary of the parameter space Δ . In this case, we need to consider further modification of the prior. This is similar to the situation, where the boundary points of the simplex Δ are assigned mass by the minimax prior, studied by Xie and Barron [13, 14] in the reference prior formulation. In the present paper, we do not use such a modification because we consider a sequence of subsets Δ_{ε_N} converging to the original parameter space Δ at a rate satisfying the conditions.

If ε_N converges to 0 very slowly, then the condition $\lim_{N \rightarrow \infty} N^{1/\hat{\alpha}} \varepsilon_N = 0$ is not satisfied and we cannot neglect the difference between $\bar{\pi}_{\hat{\alpha}}$ and $\bar{\pi}_{\hat{\alpha}}^{(N)}$. The constant $1/\hat{\alpha} = \sqrt{6}/(\sqrt{6} + 1) \simeq 0.7101$ is not much smaller than $3/4$. It may be possible to weaken the condition $\lim_{N \rightarrow \infty} N^{3/4} \varepsilon_N = \infty$ by using an expansion of the risk function with terms higher order than those in the formula in Theorem 2.1.

The condition $N^{3/4} \varepsilon_N \rightarrow \infty$ in Corollaries 2.1–2.3 and Theorem 3.2 is stronger than $N \varepsilon_N \rightarrow \infty$ in Theorem 2.1. The rate $N^{3/4} \varepsilon_N \rightarrow \infty$ is required because the right-hand side $O(N^{-5} \varepsilon_N^{-4})$ of (2.2) is $o(N^{-2})$ under this condition. If we use an expansion up to order N^{-k} ($k = 5, 6, \dots$) with terms higher order than those in (2.2), we need to assume $N^{k-1/k} \varepsilon_N \rightarrow \infty$ to assure that the remaining term is $o(N^{-2})$. The condition $N^{k-1/k} \varepsilon_N \rightarrow \infty$ becomes closer to $N \varepsilon_N \rightarrow \infty$ in Theorem 2.1 as k increases.

Appendix A: Proofs of lemmas

Proof of Lemma 2.1. (1) Let

$$f(x) := \log(1+x) - \sum_{i=1}^{2m+1} (-1)^{i-1} \frac{1}{i} x^i.$$

Then, $f(0) = 0$, and

$$f'(x) = \frac{1}{1+x} - \sum_{i=0}^{2m} (-1)^i x^i = \frac{1}{1+x} - \left(\frac{1}{1+x} + \frac{x^{2m+1}}{1+x} \right) = -\frac{x^{2m+1}}{1+x}.$$

Thus, $f'(x) > 0$ for $-1 < x < 0$, $f'(x) = 0$ for $x = 0$, and $f'(x) < 0$ for $x > 0$. Therefore, $f(x) \leq 0$ for $x > -1$, and the equality holds only when $x = 0$.

(2) Let

$$f(x) := \log(1+x) - \sum_{i=1}^{2m} (-1)^{i-1} \frac{1}{i} x^i - \frac{1}{2m+1} \frac{x^{2m+1}}{1+x}.$$

Then, $f(0) = 0$, and

$$f'(x) = \frac{1}{1+x} - \sum_{i=0}^{2m-1} (-1)^i x^i - \frac{x^{2m}}{1+x} + \frac{1}{2m+1} \frac{x^{2m+1}}{(1+x)^2}$$

$$\begin{aligned}
 &= \frac{1}{1+x} - \left(\frac{1}{1+x} - \frac{x^{2m}}{1+x} \right) - \frac{x^{2m}}{1+x} + \frac{1}{2m+1} \frac{x^{2m+1}}{(1+x)^2} \\
 &= \frac{1}{2m+1} \frac{x^{2m+1}}{(1+x)^2}.
 \end{aligned}$$

Thus, $f'(x) < 0$ for $-1 < x < 0$, $f'(x) = 0$ for $x = 0$, and $f'(x) > 0$ for $x > 0$. Therefore, $f(x) \geq 0$ for $-1 < x$, and the equality holds only when $x = 0$. \square

Proof of Lemma 2.2. We prove the desired results by induction. Assume that $\mu_{2l-1}(N, \theta)$ and $\mu_{2l}(N, \theta)$, where l is a positive integer, are represented as

$$\mu_{2l-1}(N, \theta) = \sum_{i=1}^{l-1} f_{2l-1,i}(\theta)(N\theta)^i \quad \text{and} \quad \mu_{2l}(N, \theta) = \sum_{i=1}^l f_{2l,i}(\theta)(N\theta)^i \quad (\text{A.1})$$

where $f_{2l-1,i}(\theta)$ ($i = 1, 2, \dots, l-1$) and $f_{2l,i}(\theta)$ ($i = 1, 2, \dots, l$) are polynomials with integer coefficients. Then, by using the recurrence equation

$$\mu_{m+1}(N, \theta) = \theta(1-\theta) \left\{ Nm\mu_{m-1}(N, \theta) + \frac{d\mu_m(N, \theta)}{d\theta} \right\} \quad (m = 2, 3, 4, \dots)$$

by Romanovsky [12], we have

$$\mu_{2l+1}(N, \theta) = \theta(1-\theta) \left\{ 2Nl \sum_{i=1}^{l-1} f_{2l-1,i}(\theta)(N\theta)^i + \frac{d}{d\theta} \sum_{i=1}^l f_{2l,i}(\theta)(N\theta)^i \right\}$$

and

$$\mu_{2l+2}(N, \theta) = \theta(1-\theta) \left\{ 2N(l+1) \sum_{i=1}^l f_{2l,i}(\theta)(N\theta)^i + \frac{d}{d\theta} \sum_{i=1}^l f_{2l+1,i}(\theta)(N\theta)^i \right\}.$$

Thus, $\mu_{2l+1}(N, \theta)$ and $\mu_{2l+2}(N, \theta)$ are represented as

$$\mu_{2l+1}(N, \theta) = \sum_{i=1}^l f_{2l+1,i}(\theta)(N\theta)^i \quad \text{and} \quad \mu_{2l+2}(N, \theta) = \sum_{i=1}^{l+1} f_{2l+2,i}(\theta)(N\theta)^i,$$

where $f_{2l+1,i}(\theta)$ and $f_{2l+2,i}(\theta)$ are polynomials of θ with integer coefficients.

Since $\mu_1(N, \theta) = 0$ and $\mu_2(N, \theta) = N\theta(1-\theta)$, the equation (A.1) holds for every positive integer l .

Therefore, because $N\varepsilon_N$ goes to infinity, there exist constants C_{2l-1} and C_{2l} not depending on N or θ such that

$$\frac{|\mu_{2l-1}(N, \theta)|}{(N\theta)^{l-1}} \leq \sum_{i=1}^{l-1} \frac{|f_{2l-1,i}(\theta)|}{(N\theta)^{l-1-i}} \leq \sum_{i=1}^{l-1} \frac{\max_{\theta \in [0,1]} |f_{2l-1,i}(\theta)|}{(N\varepsilon_N)^{l-1-i}} \leq C_{2l-1}$$

and

$$\frac{|\mu_{2l}(N, \theta)|}{(N\theta)^l} \leq \sum_{i=1}^l \frac{|f_{2l,i}(\theta)|}{(N\theta)^{l-i}} \leq \sum_{i=1}^l \frac{\max_{\theta \in [0,1]} |f_{2l,i}(\theta)|}{(N\varepsilon_N)^{l-i}} \leq C_{2l},$$

respectively. \square

Proof of Lemma 2.3. We have

$$\begin{aligned} \mathbb{E}\left(\frac{w^{2l}}{1+w}\right) &= \sum_{x=0}^N \binom{N}{x} \theta^x (1-\theta)^{N-x} \frac{N\theta+a}{x+a} \left(\frac{x-N\theta}{N\theta+a}\right)^{2l} \\ &= \frac{1}{(N\theta+a)^{2l-1}} \sum_{x=0}^N \frac{(N+1)!}{(x+1)! \{(N+1)-(x+1)\}!} \\ &\quad \times \frac{1}{N+1} \frac{1}{\theta} \theta^{x+1} (1-\theta)^{N+1-(x+1)} \frac{x+1}{x+a} (x-N\theta)^{2l}. \end{aligned}$$

Here, for every $x \geq 0$,

$$\frac{x+1}{x+a} \leq \begin{cases} 1, & \text{if } a \geq 1, \\ \frac{1}{a}, & \text{if } 0 < a < 1. \end{cases}$$

Thus,

$$\begin{aligned} \mathbb{E}\left(\frac{w^{2l}}{1+w}\right) &\leq \frac{\max\left(1, \frac{1}{a}\right)}{(N\theta+a)^{2l-1} (N+1)\theta} \sum_{z=0}^{N+1} \frac{(N+1)!}{z! (N+1-z)!} \theta^z (1-\theta)^{N+1-z} \\ &\quad \times \{z - (N+1)\theta - (1-\theta)\}^{2l} \\ &= \frac{\max\left(1, \frac{1}{a}\right)}{(N\theta+a)^{2l-1} (N+1)\theta} \sum_{z=0}^{N+1} \binom{N+1}{z} \theta^z (1-\theta)^{N+1-z} \\ &\quad \times \sum_{j=0}^{2l} \binom{2l}{j} \{z - (N+1)\theta\}^j \{-(1-\theta)\}^{2l-j} \\ &\leq \frac{\max\left(1, \frac{1}{a}\right)}{(N\theta+a)^{2l-1} (N+1)\theta} \sum_{j=0}^{2l} \binom{2l}{j} |\mu_j(N+1, \theta)|, \end{aligned}$$

where we define $\mu_0(N+1, \theta) := 1$. By Lemma 2.2, there exist positive constants \tilde{C}_i ($i = 0, \dots, 2l$) such that

$$\begin{aligned} \mathbb{E}\left(\frac{w^{2l}}{1+w}\right) &\leq \frac{\max\left(1, \frac{1}{a}\right)}{(N\theta+a)^{2l-1} (N+1)\theta} \left\{ \sum_{j=0}^l \tilde{C}_{2j} \{(N+1)\theta\}^j + \sum_{j=0}^{l-1} \tilde{C}_{2j+1} \{(N+1)\theta\}^j \right\} \\ &\leq \max\left(1, \frac{1}{a}\right) \left\{ \sum_{j=0}^l \left(\frac{N+1}{N}\right)^{j-1} \frac{\tilde{C}_{2j}}{(N\theta)^{2l-j}} + \sum_{j=0}^{l-1} \left(\frac{N+1}{N}\right)^{j-1} \frac{\tilde{C}_{2j+1}}{(N\theta)^{2l-j}} \right\} \\ &\leq 2^{l-1} \max\left(1, \frac{1}{a}\right) \frac{1}{(N\theta)^l} \left\{ \tilde{C}_{2l} + \sum_{j=0}^{l-1} \frac{\tilde{C}_{2j}}{(N\varepsilon_N)^{l-j}} + \sum_{j=0}^{l-1} \frac{\tilde{C}_{2j+1}}{(N\varepsilon_N)^{l-j}} \right\}. \end{aligned}$$

Since $N\varepsilon_N$ goes to infinity, there exists a constant C such that

$$\mathbb{E}\left(\frac{w^{2l}}{1+w}\right) \leq \frac{C}{(N\theta)^l}.$$

Therefore,

$$\mathbb{E}\left(-\frac{w^{2l+1}}{1+w}\right) = -\mathbb{E}(w^{2l}) + \mathbb{E}\left(\frac{w^{2l}}{1+w}\right) \leq \frac{C}{(N\theta)^l}.$$

□

Proof of Lemma 3.1. The desired inequality is equivalent to

$$\begin{aligned} & (\alpha_1 + \dots + \alpha_k)B_{\Delta_\varepsilon}(\alpha_1 + 1, \alpha_2, \dots, \alpha_k) - \alpha_1 B_{\Delta_\varepsilon}(\alpha_1, \dots, \alpha_k) \quad (\text{A.2}) \\ & \leq \frac{\Gamma(\alpha_2) \dots \Gamma(\alpha_k)}{\Gamma(\alpha_2 + \dots + \alpha_k)} \varepsilon^{\alpha_1} (1 - \varepsilon)^{\alpha_2 + \dots + \alpha_k}. \end{aligned}$$

Let

$$w_i := \frac{\theta_i}{1 - \theta_1} \quad (i = 2, \dots, k).$$

Then,

$$\begin{aligned} & \theta_1^{\alpha_1 - 1} \dots \theta_{k-1}^{\alpha_{k-1} - 1} (1 - \theta_1 - \dots - \theta_{k-1})^{\alpha_k - 1} d\theta_1 \dots d\theta_{k-1} \\ & = \theta_1^{\alpha_1 - 1} (1 - \theta_1)^{\alpha_2 + \dots + \alpha_k - 1} w_2^{\alpha_2 - 1} \dots w_{k-1}^{\alpha_{k-1} - 1} (1 - w_2 - \dots - w_{k-1})^{\alpha_k - 1} \\ & \quad \times d\theta_1 dw_2 \dots dw_{k-1}. \end{aligned}$$

We define

$$\Delta'_{\varepsilon/(1-\varepsilon)} := \left\{ w = (w_2, \dots, w_{k-1}) \mid w_i \geq \varepsilon/(1-\varepsilon) \ (i=2, \dots, k), w_k := 1 - \sum_{i=2}^{k-1} w_i \right\}.$$

If $\theta \in \Delta_\varepsilon$, then $(\theta_2/(1 - \theta_1), \dots, \theta_{k-1}/(1 - \theta_1)) \in \Delta'_{\varepsilon/(1-\varepsilon)}$.

If $w = (w_2, \dots, w_{k-1}) \in \Delta'_{\varepsilon/(1-\varepsilon)}$ is fixed, then $\{\theta = (\theta_1, \dots, \theta_{k-1}) \mid \theta \in \Delta_\varepsilon, \theta_i = (1 - \theta_1)w_i \ (i = 2, \dots, k-1)\}$, which is a subset of Δ_ε , is represented as $\{\theta \mid (\theta_1, (1 - \theta_1)w_2, \dots, (1 - \theta_1)w_{k-1}) \mid L(w_2, \dots, w_{k-1}) \leq \theta_1 \leq U(w_2, \dots, w_{k-1})\}$ by using appropriate functions $L(w_2, \dots, w_{k-1})$ and $U(w_2, \dots, w_{k-1})$ because Δ_ε is a bounded closed convex set.

If $(\theta_1, (1 - \theta_1)w_2, \dots, (1 - \theta_1)w_{k-1}) \in \Delta_\varepsilon$, then $(\varepsilon, (1 - \varepsilon)w_2, \dots, (1 - \varepsilon)w_{k-1}) \in \Delta_\varepsilon$ because $(1 - \varepsilon)w_i \geq (1 - \theta_1)w_i \geq \varepsilon$ for $i = 2, \dots, k$ and $\varepsilon + \sum_{i=2}^k (1 - \varepsilon)w_i = 1$. Thus, $L(w_2, \dots, w_{k-1}) \geq \varepsilon$. Obviously, $L(w_2, \dots, w_{k-1}) \leq \varepsilon$ because $\theta \notin \Delta_\varepsilon$ if $\theta_1 < \varepsilon$. Hence, $L(w_2, \dots, w_{k-1}) = \varepsilon$.

Since $\theta_1 = 1 - \theta_2 - \dots - \theta_k \leq 1 - (k - 1)\varepsilon$, $U(w_2, \dots, w_{k-1}) \leq 1 - (k - 1)\varepsilon$.

Therefore, we have

$$\begin{aligned} & B_{\Delta_\varepsilon}(\alpha_1 + 1, \alpha_2, \dots, \alpha_k) \\ & = \int_{\Delta_\varepsilon} \theta_1^{\alpha_1} \theta_2^{\alpha_2 - 1} \dots \theta_{k-1}^{\alpha_{k-1} - 1} (1 - \theta_1 - \dots - \theta_{k-1})^{\alpha_k - 1} d\theta_1 \dots d\theta_{k-1} \end{aligned}$$

$$\begin{aligned}
&= \int_{\Delta'_{\varepsilon/(1-\varepsilon)}} \left\{ \int_{\varepsilon}^{U(w_2, \dots, w_{k-1})} \theta_1^{\alpha_1} (1 - \theta_1)^{\sum_{i=2}^k \alpha_i - 1} d\theta_1 \right\} \\
&\quad \times w_2^{\alpha_2 - 1} \cdots w_{k-1}^{\alpha_{k-1} - 1} (1 - w_2 - \cdots - w_{k-1})^{\alpha_k - 1} dw_2 \cdots dw_{k-1} \\
&= \int_{\Delta'_{\varepsilon/(1-\varepsilon)}} \left\{ \left[\theta_1^{\alpha_1} \frac{-1}{\sum_{i=2}^k \alpha_i} (1 - \theta_1)^{\sum_{i=2}^k \alpha_i} \right]_{\varepsilon}^{U(w_2, \dots, w_{k-1})} \right. \\
&\quad \left. + \int_{\varepsilon}^{U(w_2, \dots, w_{k-1})} \frac{\alpha_1}{\sum_{i=2}^k \alpha_i} \theta_1^{\alpha_1 - 1} (1 - \theta_1)^{\sum_{i=2}^k \alpha_i} d\theta_1 \right\} \\
&\quad \times w_2^{\alpha_2 - 1} \cdots w_{k-1}^{\alpha_{k-1} - 1} (1 - w_2 - \cdots - w_{k-1})^{\alpha_k - 1} dw_2 \cdots dw_{k-1} \\
&= \int_{\Delta'_{\varepsilon/(1-\varepsilon)}} \frac{1}{\sum_{i=2}^k \alpha_i} \left[\varepsilon^{\alpha_1} (1 - \varepsilon)^{\sum_{i=2}^k \alpha_i} \right. \\
&\quad \left. - \{U(w_2, \dots, w_{k-1})\}^{\alpha_1} \{1 - U(w_2, \dots, w_{k-1})\}^{\sum_{i=2}^k \alpha_i} \right] \\
&\quad \times w_2^{\alpha_2 - 1} \cdots w_{k-1}^{\alpha_{k-1} - 1} (1 - w_2 - \cdots - w_{k-1})^{\alpha_k - 1} dw_2 \cdots dw_{k-1} \\
&\quad + \int_{\Delta'_{\varepsilon/(1-\varepsilon)}} \left\{ \int_{\varepsilon}^{U(w_2, \dots, w_{k-1})} \frac{\alpha_1}{\sum_{i=2}^k \alpha_i} \theta_1^{\alpha_1 - 1} (1 - \theta_1)^{\sum_{i=2}^k \alpha_i} d\theta_1 \right\} \\
&\quad \times w_2^{\alpha_2 - 1} \cdots w_{k-1}^{\alpha_{k-1} - 1} (1 - w_2 - \cdots - w_{k-1})^{\alpha_k - 1} dw_2 \cdots dw_{k-1} \\
&\leq \frac{1}{\sum_{i=2}^k \alpha_i} \varepsilon^{\alpha_1} (1 - \varepsilon)^{\sum_{i=2}^k \alpha_i} \\
&\quad \times \int_{\Delta'_{\varepsilon/(1-\varepsilon)}} w_2^{\alpha_2 - 1} \cdots w_{k-1}^{\alpha_{k-1} - 1} (1 - w_2 - \cdots - w_{k-1})^{\alpha_k - 1} dw_2 \cdots dw_{k-1} \\
&\quad + \frac{\alpha_1}{\sum_{i=2}^k \alpha_i} B_{\Delta'_{\varepsilon/(1-\varepsilon)}}(a_1, a_2, \dots, a_k) \\
&\leq \frac{B(\alpha_2, \dots, \alpha_k)}{\sum_{i=2}^k \alpha_i} \varepsilon^{\alpha_1} (1 - \varepsilon)^{\alpha_2 + \dots + \alpha_k} + \frac{\alpha_1}{\sum_{i=2}^k \alpha_i} B_{\Delta_{\varepsilon}}(a_1, a_2, \dots, a_k).
\end{aligned}$$

Thus, (A.2) is obtained. \square

Proof of Lemma 3.2. We obtain the desired inequality from

$$\begin{aligned}
&\frac{\partial}{\partial t} \frac{B_{[s,t]}(\alpha + 1, \beta)}{B_{[s,t]}(\alpha, \beta)} \\
&= \frac{\left\{ \frac{\partial}{\partial t} B_{[s,t]}(\alpha + 1, \beta) \right\} B_{[s,t]}(\alpha, \beta) - B_{[s,t]}(\alpha + 1, \beta) \left\{ \frac{\partial}{\partial t} B_{[s,t]}(\alpha, \beta) \right\}}{\{B_{[s,t]}(\alpha, \beta)\}^2} \\
&= \frac{1}{\{B_{[s,t]}(\alpha, \beta)\}^2} t^{\alpha-1} (1-t)^{\beta-1} \int_s^t (t-\theta) \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta \geq 0
\end{aligned}$$

and

$$\begin{aligned} & \frac{\partial}{\partial s} \frac{B_{[s,t]}(\alpha + 1, \beta)}{B_{[s,t]}(\alpha, \beta)} \\ &= \frac{\left\{ \frac{\partial}{\partial s} B_{[s,t]}(\alpha + 1, \beta) \right\} B_{[s,t]}(\alpha, \beta) - B_{[s,t]}(\alpha + 1, \beta) \left\{ \frac{\partial}{\partial s} B_{[s,t]}(\alpha, \beta) \right\}}{\{B_{[s,t]}(\alpha, \beta)\}^2} \\ &= \frac{1}{\{B_{[s,t]}(\alpha, \beta)\}^2} s^{\alpha-1} (1-s)^{\beta-1} \int_s^t (\theta-s)\theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta \geq 0. \end{aligned}$$

□

Proof of Lemma 3.3. Define w_i ($i = 2, \dots, k$), $\Delta'_{\varepsilon/(1-\varepsilon)}$, and $U(w_2, \dots, w_{k-1})$ as in the proof of Lemma 3.1. Let

$$\begin{aligned} & \tilde{p}(\theta_1, w_2, \dots, w_{k-1}) d\theta_1 dw_2 \cdots dw_{k-1} \\ &:= \theta_1^{\alpha_1-1} \cdots \theta_{k-1}^{\alpha_{k-1}-1} (1-\theta_1 - \cdots - \theta_{k-1})^{\alpha_k-1} d\theta_1 \cdots d\theta_{k-1} \\ &= \theta_1^{\alpha_1-1} (1-\theta_1)^{\alpha_2+\cdots+\alpha_k-1} w_2^{\alpha_2-1} \cdots w_{k-1}^{\alpha_{k-1}-1} (1-w_2 - \cdots - w_{k-1})^{\alpha_k-1} \\ &\quad \times d\theta_1 dw_2 \cdots dw_{k-1} \end{aligned}$$

and

$$p(\theta_1, w_2, \dots, w_{k-1}) := \frac{\tilde{p}(\theta_1, w_2, \dots, w_{k-1})}{B_{\Delta_\varepsilon}(\alpha_1, \alpha_2, \dots, \alpha_k)}.$$

Since

$$\begin{aligned} & B_{\Delta_\varepsilon}(\alpha_1, \alpha_2, \dots, \alpha_k) \\ &:= \int_{\Delta'_{\varepsilon/(1-\varepsilon)}} \left\{ \int_\varepsilon^{U(w_2, \dots, w_{k-1})} \tilde{p}(\theta_1, w_2, \dots, w_{k-1}) d\theta_1 \right\} dw_2 \cdots dw_{k-1}, \end{aligned}$$

$p(\theta_1, w_2, \dots, w_{k-1})$ is a probability density. The marginal density of (w_2, \dots, w_{k-1}) is

$$p(w_2, \dots, w_k) = \int_\varepsilon^{U(w_2, \dots, w_k)} p(\theta_1, w_2, \dots, w_k) d\theta_1.$$

The conditional density of θ_1 given (w_2, \dots, w_{k-1}) is

$$\begin{aligned} p(\theta_1 | w_2, \dots, w_{k-1}) &= \frac{p(\theta_1, w_2, \dots, w_{k-1})}{p(w_2, \dots, w_{k-1})} \\ &= \begin{cases} \frac{\theta_1^{\alpha_1-1} (1-\theta_1)^{\alpha_2+\cdots+\alpha_k-1}}{\int_\varepsilon^{U(w_2, \dots, w_{k-1})} \theta_1^{\alpha_1-1} (1-\theta_1)^{\alpha_2+\cdots+\alpha_k-1} d\theta_1}, & \varepsilon \leq \theta_1 \leq U(w_2, \dots, w_{k-1}), \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Then, from Lemma 3.2 and $U(w_2, \dots, w_{k-1}) \leq 1 - (k - 1)\varepsilon \leq 1$,

$$\begin{aligned} & \frac{B_{\Delta_\varepsilon}(\alpha_1 + 1, \alpha_2, \dots, \alpha_k)}{B_{\Delta_\varepsilon}(\alpha_1, \alpha_2, \dots, \alpha_k)} \\ &= \int_{\Delta'_{\varepsilon/(1-\varepsilon)}} \int_{\varepsilon}^{U(w_2, \dots, w_{k-1})} \theta_1 p(\theta_1, w_2, \dots, w_{k-1}) d\theta_1 dw_2 \cdots dw_{k-1} \\ &= \int_{\Delta'_{\varepsilon/(1-\varepsilon)}} \left\{ \int_{\varepsilon}^{U(w_2, \dots, w_{k-1})} \theta_1 p(\theta_1 \mid w_2, \dots, w_{k-1}) d\theta_1 \right\} \\ & \quad \times p(w_2, \dots, w_{k-1}) dw_2 \cdots dw_{k-1} \\ &= \int_{\Delta'_{\varepsilon/(1-\varepsilon)}} \left\{ \frac{\int_{\varepsilon}^{U(w_2, \dots, w_{k-1})} \theta_1^{\alpha_1} (1 - \theta_1)^{\alpha_2 + \dots + \alpha_k - 1} d\theta_1}{\int_{\varepsilon}^{U(w_2, \dots, w_{k-1})} \theta_1^{\alpha_1 - 1} (1 - \theta_1)^{\alpha_2 + \dots + \alpha_k - 1} d\theta_1} \right\} \\ & \quad \times p(w_2, \dots, w_{k-1}) dw_2 \cdots dw_{k-1} \\ &\leq \frac{\int_{\varepsilon}^{1-(k-1)\varepsilon} \theta_1^{\alpha_1} (1 - \theta_1)^{\alpha_2 + \dots + \alpha_k - 1} d\theta_1}{\int_{\varepsilon}^{1-(k-1)\varepsilon} \theta_1^{\alpha_1 - 1} (1 - \theta_1)^{\alpha_2 + \dots + \alpha_k - 1} d\theta_1} \\ &= \frac{B_{[\varepsilon, 1-(k-1)\varepsilon]}(\alpha_1 + 1, \alpha_2 + \dots + \alpha_k)}{B_{[\varepsilon, 1-(k-1)\varepsilon]}(\alpha_1, \alpha_2 + \dots + \alpha_k)} \\ &\leq \frac{B_{[\varepsilon, 1]}(\alpha_1 + 1, \alpha_2 + \dots + \alpha_k)}{B_{[\varepsilon, 1]}(\alpha_1, \alpha_2 + \dots + \alpha_k)}. \end{aligned}$$

□

Proof of Lemma 3.4. The right hand side of the equation is represented by

$$\begin{aligned} & \sum_{x_2, \dots, x_k: \sum_{i=2}^k x_i = N - x_1} \frac{B(x_1 + \alpha_1, x_2 + \alpha_2, \dots, x_k + \alpha_k)}{B(\alpha_1, \alpha_2, \dots, \alpha_k)} \binom{N}{x_1, \dots, x_k} \\ &= \sum_{x_2, \dots, x_k: \sum_{i=2}^k x_i = N - x_1} \int_{\Delta} \frac{\theta_1^{\alpha_1 - 1} \cdots \theta_{k-1}^{\alpha_{k-1} - 1} (1 - \sum_{i=1}^{k-1} \theta_i)^{\alpha_k - 1}}{B(\alpha_1, \dots, \alpha_k)} \\ & \quad \times \binom{N}{x_1, \dots, x_k} \theta_1^{x_1} \cdots \theta_{k-1}^{x_{k-1}} (1 - \sum_{i=1}^{k-1} \theta_i)^{N - \sum_{i=1}^{k-1} x_i} d\theta_1 \cdots d\theta_{k-1} \\ &= \int_{\Delta} \frac{\theta_1^{\alpha_1 - 1} \cdots \theta_{k-1}^{\alpha_{k-1} - 1} (1 - \sum_{i=1}^{k-1} \theta_i)^{\alpha_k - 1}}{B(\alpha_1, \dots, \alpha_k)} \\ & \quad \times \binom{N}{x_1} \theta_1^{x_1} (1 - \theta_1)^{N - x_1} d\theta_1 \cdots d\theta_{k-1}. \end{aligned}$$

From the relation

$$\begin{aligned}
& \int_{\Delta} \theta_1^{\alpha_1-1} \cdots \theta_{k-1}^{\alpha_{k-1}-1} (1 - \sum_{i=1}^{k-1} \theta_i)^{\alpha_k-1} \theta_1^{x_1} (1 - \theta_1)^{N-x_1} d\theta_1 \cdots d\theta_{k-1} \\
&= \int_0^1 \int_0^{1-\theta_1} \cdots \int_0^{1-\sum_{i=1}^{k-2} \theta_i} \theta_1^{\alpha_1-1} \cdots \theta_{k-1}^{\alpha_{k-1}-1} (1 - \sum_{i=1}^{k-1} \theta_i)^{\alpha_k-1} \theta_1^{x_1} (1 - \theta_1)^{N-x_1} \\
&\quad d\theta_{k-1} d\theta_{k-2} \cdots d\theta_1 \\
&= \int_0^1 \int_0^{1-\theta_1} \cdots \int_0^{1-\sum_{i=1}^{k-3} \theta_i} \int_0^1 \theta_1^{\alpha_1+x_1-1} \theta_2^{\alpha_2-1} \cdots \theta_{k-2}^{\alpha_{k-2}-1} \\
&\quad \times (1 - \sum_{i=1}^{k-2} \theta_i)^{\alpha_{k-1}+\alpha_k-1} \bar{\theta}_{k-1}^{\alpha_{k-1}-1} (1 - \bar{\theta}_{k-1})^{\alpha_k-1} (1 - \theta_1)^{N-x_1} \\
&\quad d\bar{\theta}_{k-1} d\theta_{k-2} \cdots d\theta_1 \\
&= B(\alpha_{k-1}, \alpha_k) \int_0^1 \int_0^{1-\theta_1} \cdots \int_0^{1-\sum_{i=1}^{k-3} \theta_i} \theta_1^{\alpha_1+x_1-1} \theta_2^{\alpha_2-1} \cdots \theta_{k-2}^{\alpha_{k-2}-1} \\
&\quad \times (1 - \sum_{i=1}^{k-2} \theta_i)^{\alpha_{k-1}+\alpha_k-1} (1 - \theta_1)^{N-x_1} d\theta_{k-2} \cdots d\theta_1 \\
&= B(\alpha_{k-1}, \alpha_k) B(\alpha_{k-2}, \alpha_{k-1} + \alpha_k) \cdots B(\alpha_2, \sum_{i=3}^k \alpha_i) \\
&\quad \times \int_0^1 \theta_1^{\alpha_1+x_1-1} (1 - \theta_1)^{\sum_{i=2}^k \alpha_i + N - x_1 - 1} d\theta_1 \\
&= \frac{B(\alpha_1, \alpha_2, \dots, \alpha_k)}{B(\alpha_1, \sum_{i=2}^k \alpha_i)} B(\alpha_1 + x_1, \sum_{i=2}^k \alpha_i + N - x_1),
\end{aligned}$$

where $\bar{\theta}_{k-1} = \theta_{k-1} / \sum_{j=1}^{k-2} \theta_j$, we obtain the desired result. \square

Proof of Lemma 3.5. Since

$$\begin{aligned}
\int_{\varepsilon}^1 \theta^{\alpha} (1 - \theta)^{\beta-1} d\theta &= \left[-\frac{1}{\beta} \theta^{\alpha} (1 - \theta)^{\beta} \right]_{\varepsilon}^1 + \frac{\alpha}{\beta} \int_{\varepsilon}^1 \theta^{\alpha-1} (1 - \theta)^{\beta} d\theta \\
&= \frac{1}{\beta} \varepsilon^{\alpha} (1 - \varepsilon)^{\beta} - \frac{\alpha}{\beta} \int_{\varepsilon}^1 \theta^{\alpha} (1 - \theta)^{\beta-1} d\theta + \frac{\alpha}{\beta} \int_{\varepsilon}^1 \theta^{\alpha-1} (1 - \theta)^{\beta-1} d\theta,
\end{aligned}$$

we have

$$\begin{aligned}
(\alpha + \beta) \frac{\int_{\varepsilon}^1 \theta^{\alpha} (1 - \theta)^{\beta-1} d\theta}{\int_{\varepsilon}^1 \theta^{\alpha-1} (1 - \theta)^{\beta-1} d\theta} &= \alpha + \frac{\varepsilon^{\alpha} (1 - \varepsilon)^{\beta}}{\int_{\varepsilon}^1 \theta^{\alpha-1} (1 - \theta)^{\beta-1} d\theta} \\
&\leq \alpha + \frac{\varepsilon^{\alpha} (1 - \varepsilon)^{\beta}}{\varepsilon^{\alpha-1} \int_{\varepsilon}^1 (1 - \theta)^{\beta-1} d\theta} = \alpha + \frac{\varepsilon (1 - \varepsilon)^{\beta}}{\frac{1}{\beta} [(1 - \theta)^{\beta}]_{\varepsilon}^1} = \alpha + \beta \varepsilon.
\end{aligned}$$

\square

Acknowledgements

The author is grateful to the referee for constructive comments. This research was partially supported by Grant-in-Aid for Scientific Research (23300104, 23650144) and by the Aihara Project, the FIRST program from JSPS, initiated by CSTP.

References

- [1] AITCHISON, J. (1975). Goodness of prediction fit. *Biometrika*, **62**, 547–554. [MR0391353](#)
- [2] BERGER, J. O. AND BERNARDO, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *Journal of the American Statistical Association*, **84**, 200–207. [MR0999679](#)
- [3] BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *Journal of Royal Statistical Society B*, **41**, 113–147. [MR0547240](#)
- [4] BERNARDO, J. M. (2005). Reference analysis. *Handbook of Statistics*, **25**, Dey, K. K. and Rao C. R. eds., Elsevier, Amsterdam 17–90. [MR2490522](#)
- [5] CLARKE, B. (2007). Information optimality and Bayesian modelling. *Journal of Econometrics*, **138**, 405–429. [MR2391317](#)
- [6] CLARKE, B. AND BARRON, A. R. (1994). Jeffreys' prior is asymptotically least favorable under entropy risk. *Journal of Statistical Planning and Inference*, **41**, 36–60. [MR1292146](#)
- [7] GRÜNWARD, P. D. AND DAWID, A. P. (2004). Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics*, **32**, 1367–1433. [MR2089128](#)
- [8] IBRAGIMOV, I. A. AND HASMINSKII, R. Z. (1973). On the information contained in a sample about a parameter. In *2nd Intl. Symp. on Information Theory*, Akademiai, Kiado, Budapest 295–309. [MR0356948](#)
- [9] KOMAKI, F. (1996). On asymptotic properties of predictive distributions. *Biometrika*, **83**, 299–313. [MR1439785](#)
- [10] KOMAKI, F. (2004). Simultaneous prediction of independent Poisson observables. *Annals of Statistics*, **32**, 1744–1769. [MR2089141](#)
- [11] KOMAKI, F. (2011). Bayesian predictive densities based on latent information priors. *Journal of Statistical Planning and Inference*, **141**, 3705–3715. [MR2823641](#)
- [12] ROMANOVSKY, V. (1923). Note on the moments of a binomial $(p + q)^n$ about its mean, *Biometrika*, **15**, 410–412.
- [13] XIE, Q. AND BARRON, A. R. (1997). Minimax redundancy for the class of memoryless sources. *IEEE Transactions on Information Theory*, **43**, 646–657.
- [14] XIE, Q. AND BARRON, A. R. (2000). Asymptotic minimax regret for data compression, gambling, and prediction. *IEEE Transaction on Information Theory*, **46**, 431–445. [MR1748979](#)