Estimation of a Continuous Distribution on the Real Line by Discretization Methods

Yo Sheena*

Jun. 2018

For an unknown continuous distribution on the real line, we consider the approximate estimation by discretization. There are two methods for discretization. The first method is to divide the real line into several intervals before taking samples ("fixed interval method"). The second method is to divide the real line using the estimated percentiles after taking samples ("moving interval method"). In either method, we arrive at the estimation problem of a multinomial distribution. We use (symmetrized) f-divergence to measure the discrepancy between the true distribution and the estimated distribution. Our main result is the asymptotic expansion of the risk (i.e., expected divergence) up to the second-order term in the sample size. We prove theoretically that the moving interval method is asymptotically superior to the fixed interval method. We also observe how the presupposed intervals (fixed interval method) or percentiles (moving interval method) affect the asymptotic risk.

MSC(2010) *Subject Classification*: Primary 60F99; Secondary 62F12
Key words and phrases: f-divergence, alpha-divergence, asymptotic risk, asymptotic expansion, multinomial distribution.

1 Introduction

Consider a probability distribution on the real line that is absolutely continuous with respect to the Lebesgue measure. (We do not assume it has full support $(-\infty, \infty)$.) We call this distribution the "mother distribution." Let $P(a, b)$ denote the probability

*Faculty of Economics, Shinshu University, Japan; Faculty of Data Science, Shiga University, Japan. On behalf of all authors, the corresponding author states that there is no conflict of interest.

of the mother distribution for the interval (a, b). We discretize the mother distribution and obtain the corresponding multinomial distribution as follows: let

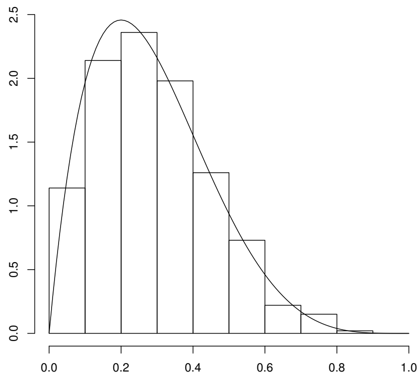
$$-\infty(\triangleq a_0) < a_1 < a_2 < \ldots < a_p < \infty(\triangleq a_{p+1}). \tag{1}$$

Consider the multinomial distribution with possible results C_i $(i = 0, \ldots, p)$ each of which having a probability $P(a_i, a_{i+1})$. This multinomial distribution is an approximation of the mother distribution and convey a certain amount of information on the mother distribution. In many practical cases, this information could be enough for a statistical analysis with an appropriate selection of a_i. (See e.g., Barbiero [?], Drezner and Zerom [?], and English et al. [?].) This approximation has the merit that it can be applicable to any mother distribution even when the mother distribution cannot be approximated well by a parametric family.

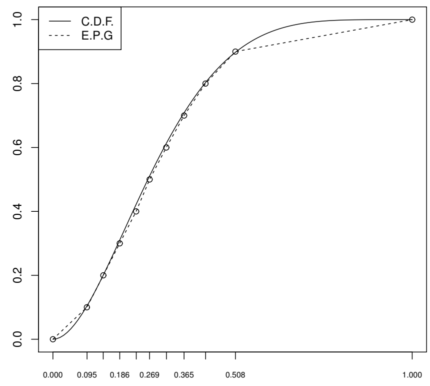
In this paper, we consider the estimation of the unknown mother distribution through the approximation by discretization. There are two theoretically contrasting methods on how to choose the a_i. One is the "fixed interval method." The a_i are given before collecting the sample. In other words, we choose the intervals independently of the sample from the mother distribution. The other method is the "moving interval method." First choose the percentiles to be estimated, $\xi_1 < \ldots < \xi_p$, and estimate them from the sample of the mother distribution. The estimated percentiles $\hat{\xi}_i$ $(i = 1, \ldots, p)$ are used as the endpoints of the intervals, that is, $a_i = \hat{\xi}_i$ $(i = 1, \ldots, p)$. The difference between the two methods lies in "intervals first" or "percentiles first."

Let us introduce some aspects of the two methods from a practical point of view. First, we note that in some practical situations, we are only allowed to choose the fixed interval method. Although the mother distribution is continuous, we are only able to observe samples within some significant digits. If that number of significant digits is not large enough to recognize small individual differences, we obtain numbers with many ties as samples, where we are forced to choose the fixed intervals. For example, if we have to weigh adult males using kilogram units without accuracy for decimal places, the above points a_i would be numbers such as 49.5, 50.5, 51.5, …, 120.5 (kg), which are prefixed independently of the samples.

Second the fixed intervals and the moving intervals correspond to a histogram and an empirical percentile graph (E.P.G.), respectively, in view of summarizing the data. Figure **??** shows the density function of $Beta(2, 5)$ as the mother distribution and the histogram (relative frequency) made from 10^3 randomly chosen samples, where the intervals are prefixed as $a_i = i/10,\ i = 1, \ldots, 9$. Figure **??** shows the cumulative distribution function (C.D.F.) of the mother distribution and the E.P.G. made from the same sample as the histogram. Percentiles are chosen as $\xi_i = i/10,\ i = 1, \ldots, 9$ and they are estimated from the order statistics $x_{(1)} < x_{(2)} < \cdots < x_{(1000)}$, that is, $\hat{\xi}_i = x_{(100i)},\ i = 1, \ldots, 9$. Both the histogram and the E.P.G. are the approximation of the mother distribution and used for the estimation (prediction) of the mother distribution. The information of the whole sample is summarized into a set of nine numbers for both approximations: the relative frequencies for 10 intervals for the histogram (because the total area of the histogram is fixed to be 1, there are 9 degrees of freedom) and 9 empirical percentiles



(a) Density and Histogram



(b) C.D.F. and E.P.G.

Figure 1: $Beta(2, 5)$ and its approximation

for the E.P.G. The question then arises, "Which estimate is more efficient?" This is the main motivation for our research.

Finally, though the two methods are theoretically contrasting, we often use one method as an auxiliary for another. When we make a histogram, we usually select the width or number of the intervals using the empirical percentiles. For example, we often choose the first endpoint a_1 and the last one a_p for the histogram after observing the minimum and maximum values of the sample. On the interval width (say h) of a histogram, we might use the "Freedman–Diaoconis rule," which recommends choosing h as

$$h = 2(\text{the third quantile} - \text{the first quantile})/n^{1/3}.$$

In contrast, when we know the mother distribution is symmetric, we set the central point as the median for the moving interval methods regardless of the value of the sample median.

Once the intervals a_i are given, we have the estimation problem of the parameters of the multinomial distribution. If we use the fixed interval method, the true (unknown) parameters are $P(a_i, a_{i+1})$ ($i = 0, \ldots, p$) and we need to estimate these parameters based on the frequencies of the sample. On the other hand, for the moving interval method, the true parameter is $P(\hat{\xi}_i, \hat{\xi}_{i+1})$ ($\hat{\xi}_0 \triangleq -\infty$ and $\hat{\xi}_{p+1} \triangleq \infty$), whereas the estimand is the probability given by the presupposed percentiles; if ξ_i is the lower $100\lambda_i\%$ percentile for $1 \leq i \leq p$, then the estimated probability for each interval is given by $\lambda_{i+1} - \lambda_i$ ($i = 0, \ldots, p$) with $\lambda_0 \triangleq 0$, $\lambda_{p+1} \triangleq 1$. In the above example, where 10^3 samples are taken from $Beta(2, 5)$, we have two sets of a multinomial distribution and its estimand (see Tables **??** and **??**).

For the measurement of the performance of the estimators, we use the f-divergence. The f-divergence between the two multinomial distributions (say M_1 and M_2) is defined

Table 1: True distribution and estimand for the fixed intervals

Intervals	0–0.1	0.1–0.2	$\cdots$	0.8–0.9	0.9–1.0
True Distribution	0.114265	0.230375	$\cdots$	0.001545	0.000055
Estimand	0.114	0.214	$\cdots$	0.002	0.000

Table 2: True distribution and estimand for the moving intervals

Intervals	0–0.095	0.095–0.140	$\cdots$	0.425–0.508	0.508–1.0
True Distribution	0.103679	0.097323	$\cdots$	0.094771	0.102116
Estimand	0.1	0.1	$\cdots$	0.1	0.1

as

$$D_f[M_1 : M_2] \triangleq \sum_{i=0}^{p} p1_i\, f\left(\frac{p2_i}{p1_i}\right), \tag{2}$$

where $p1_i, p2_i,\ i = 0, \ldots, p$, are the probabilities of each result for M_1 and M_2, respectively, and f is a smooth convex function such that $f(1) = 0,\ f'(1) = 0,\ f''(1) = 1$. The f-divergence is natural in view of the sufficiency of the sample information. If we use the dual function of f defined by $f^*(x) = xf(1/x)$, we have

$$D_{f^*}[M_1 : M_2] = D_f[M_2 : M_1]. \tag{3}$$

(See Amari [?] and Vajda [?] for the property of f-divergence.)

When the f-divergence is too abstract for us to gain some concrete result, we use the α-divergence. This is a one-parameter (α) family given by (??) with $f_\alpha(x)$ such that

$$f_\alpha(x) \triangleq \begin{cases} \frac{4}{1-\alpha^2}\left(1 - x^{(1+\alpha)/2}\right) + \frac{2}{1-\alpha}(x-1) & \text{if } \alpha \neq \pm 1, \\ x \log x + 1 - x & \text{if } \alpha = 1, \\ -\log x + x - 1 & \text{if } \alpha = -1. \end{cases} \tag{4}$$

We use the notation $\overset{\alpha}{D}[M_1 : M_2]$ instead of $D_{f_\alpha}[M_1 : M_2]$. The α-divergence is the subclass of the f-divergence, but still a broad class that contains the frequently used divergences such as the Kullback–Leibler divergence ($\alpha = -1$), Hellinger distance ($\alpha = 0$), and χ^2-divergence ($\alpha = 3$). Note that f_α^*, the dual funcion of f_α, equals $f_{-\alpha}$; hence,

$$\overset{-\alpha}{D}[M_1 : M_2] = \overset{\alpha}{D}[M_2 : M_1]. \tag{5}$$

In general, divergence $D[M_1 : M_2]$ satisfies the condition

$$D[M_1 : M_2] \geq 0, \qquad D[M_1 : M_2] = 0 \text{ if and only if } M_1 \overset{d}{=} M_2. \tag{6}$$

However, the triangle inequality and symmetry do not hold true. In this paper, we adopt the mean of the dual divergences to satisfy the symmetry (see Amari and Cichocki [?]):

$$\overset{|\alpha|}{D}[M_1 : M_2] \triangleq \frac{1}{2}\left\{\overset{\alpha}{D}[M_1 : M_2] + \overset{-\alpha}{D}[M_1 : M_2]\right\}. \tag{7}$$

We take the expectation of the divergence between the estimated multinomial distribution $\hat{M}$ and the true distribution M:

$$ED \triangleq E\big[D_f[M : \hat{M}]\big]. \tag{8}$$

This is the risk of $\hat{M}$ and we use it to describe the quality of the estimation. In this paper, we only consider the basic estimators, that is, the maximum likelihood estimator (m.l.e.) for the fixed interval and the ordered sample for the moving interval.

As it is difficult to analyze the risk theoretically under small sample sizes, we focus on asymptotic risk under large sample sizes. In Section **??**, as the main result, we show the asymptotic expansion of the risk for both methods: the fixed interval and the moving interval (Theorems 1 and 2). Using this result, first we observe how the asymptotic risk is affected by the presupposed intervals (the fixed intervals) or percentiles (the moving intervals). Second, we compare the asymptotic risk between the two methods and report the superiority of the moving interval methods when the percentiles are given with equiprobable intervals.

2 Main Result

We state the asymptotic expansion of the risk (**??**) up to second order with respect to the sample size, n, for both methods, that is, the fixed interval method (Section **??**) and the moving interval method (Section **??**). In each subsection, we analyze how the asymptotic risk is determined with respect to the sample size, the dimension of the multinomial distribution and the prefixed intervals (fixed intervals) or percentiles (moving intervals). In Section **??**, we compare both methods and show the superiority of the moving intervals when the percentiles are given with equiprobable intervals.

2.1 Fixed Intervals

We prefix the intervals with the endpoints (**??**) *before* taking the sample from the mother distribution. In other words, we choose the endpoints (**??**) independently of the sample.

We consider the multinomial distribution with the possible results $C_i,\ i = 0, \ldots, p$. If a sample from the mother distribution takes the value within the interval (a_i, a_{i+1}) for $i = 0, \ldots, p$, we count it as the sample with the result C_i. Then this multinomial distribution is an approximation of the mother distribution by a discretization. The probability for C_i is given by

$$m_i \triangleq P(a_i, a_{i+1}), \quad i = 0, \ldots, p,$$

where $P(a_i, a_{i+1})$ is the probability of the mother distribution for the interval (a_i, a_{i+1}).

We estimate this multinomial distribution through the m.l.e. Let $X_i,\ i = 1, \ldots, n$, be the independent and identically distributed (i.i.d.) sample from the mother distribution. Then the m.l.e. of $m \triangleq (m_0, \ldots, m_p)$ is given by $\hat{m} \triangleq (\hat{m}_0, \ldots, \hat{m}_p)$, where

$$\hat{m}_i \triangleq \#\{X_i | X_i \in (a_i, a_{i+1})\}/n, \quad i = 0, \ldots, p. \tag{9}$$

Its discrepancy from the true m is given by

$$D_f[m:\hat{m}] \triangleq \sum_{i=0}^{p} m_i\, f\left(\frac{\hat{m}_i}{m_i}\right). \tag{10}$$

The performance of $\hat{m}$ is measured by the risk,

$$ED_I \triangleq E\big[D_f[m:\hat{m}]\big]. \tag{11}$$

Let $f^{(i)},\ i=1,2,\ldots$, be the ith derivative of f in (**??**). For a general multinomial distribution, which is not necessarily given by a mother distribution as above, the following result holds.

Theorem 1. *Suppose that $f^{(5)}$ exists on $(0,\infty)$ and is bounded on $[\epsilon,\infty)$ for any $\epsilon(>0)$. For a multinomial distribution with the probability $m \triangleq (m_0,\ldots,m_p)$ and its m.l.e. $\hat{m}$, the risk of m.l.e. (**??**) based on an i.i.d. sample of size n is given as follows:*

$$ED_I = \frac{p}{2n} + \frac{1}{24n^2}\Big[4f^{(3)}(1)\Big(-3p-1+M\Big) + 3f^{(4)}(1)\Big(-2p-1+M\Big)\Big] + o(n^{-2}), \tag{12}$$

where

$$M \triangleq \sum_{i=0}^{p} m_i^{-1}.$$

–Proof–
Let

$$R_i \triangleq \frac{\hat{m}_i - m_i}{m_i}.$$

Note that

$$\Big(\sqrt{n}(\hat{m}_1-m_1),\ldots,\sqrt{n}(\hat{m}_p-m_p)\Big) \xrightarrow{d} N_p(0,\Sigma),$$

where

$$\Sigma \triangleq (\sigma_{ij}), \qquad \sigma_{ij} \triangleq \begin{cases} m_i(1-m_i) & \text{if } i=j, \\ -m_i m_j & \text{if } i\neq j. \end{cases}$$

(See e.g., (5.4.13) of [**?**].) Using this fact and $f(1)=0,\ f'(1)=0,\ f''(1)=1$, we have the following expansion $D_f[m:\hat{m}]$ with respect to n:

$$\begin{aligned}
&D_f[m:\hat{m}] \\
&= \sum_{i=0}^{p} m_i f(1+R_i) \\
&= \sum_{i=0}^{p} m_i\Big(f(1) + f'(1)R_i + \frac{1}{2}f''(1)R_i^2 + \frac{1}{6}f^{(3)}(1)R_i^3 + \frac{1}{24}f^{(4)}(1)R_i^4 + \frac{1}{120}f^{(5)}(1+R_i^*)R_i^5\Big) \\
&= \frac{1}{2}\sum_{i=0}^{p} m_i R_i^2 + \frac{1}{6}f^{(3)}(1)\sum_{i=0}^{p} m_i R_i^3 + \frac{1}{24}f^{(4)}(1)\sum_{i=0}^{p} m_i R_i^4 + \frac{1}{120}\sum_{i=0}^{p} f^{(5)}(1+R_i^*)m_i R_i^5
\end{aligned}$$

$$= \frac{1}{2}\sum_{i=0}^{p} m_i^{-1}(\hat{m}_i - m_i)^2 + \frac{1}{6}f^{(3)}(1)\sum_{i=0}^{p} m_i^{-2}(\hat{m}_i - m_i)^3$$
$$+ \frac{1}{24}f^{(4)}(1)\sum_{i=0}^{p} m_i^{-3}(\hat{m}_i - m_i)^4 + \frac{1}{120}\sum_{i=0}^{p} f^{(5)}(1+R_i^*)m_i^{-4}(\hat{m}_i - m_i)^5, \tag{13}$$

where R_i^* is a smooth function of R_i, hence of $\hat{m}_i$ such that

$$0 < R_i^*(\hat{m}_i) < R_i(\hat{m}_i), \text{ if } R_i(\hat{m}_i) > 0, \qquad 0 > R_i^*(\hat{m}_i) > R_i(\hat{m}_i) \geq -1, \text{ if } R_i(\hat{m}_i) < 0. \tag{14}$$

Let

$$\epsilon \triangleq \min_{0 \leq \hat{m}_i \leq 1} R_i^*(\hat{m}_i).$$

From (**??**), we have

$$R_i^*(\hat{m}_i) > -1,$$

hence, $\epsilon > -1$, which means that

$$R_i^* + 1 \geq \epsilon + 1 > 0.$$

Owing to the condition of the boundedness of $f^{(5)}(x)$ on $(\epsilon+1, \infty)$, we have some B_i (> 0) such that

$$|f^{(5)}(1+R_i^*)| \leq B_i.$$

From the central moments of the standardized multinomial distribution, we have

$E[\hat{m}_i - m_i] = 0, \qquad E[(\hat{m}_i - m_i)^2] = n^{-1}(m_i - m_i^2),$
$E[(\hat{m}_i - m_i)^3] = n^{-2}(m_i - 3m_i^2 + 2m_i^3), \qquad E[(\hat{m}_i - m_i)^4] = 3n^{-2}(m_i - m_i^2)^2 + o(n^{-2}),$
$E[(\hat{m}_i - m_i)^6] = O(n^{-3}),$

By noticing that

$$\begin{aligned}
&\left|E\left[f^{(5)}(1+R_i^*)m_i^{-4}(\hat{m}_i - m_i)^5\right]\right| \\
&\leq E\left[\left|f^{(5)}(1+R_i^*)m_i^{-4}(\hat{m}_i - m_i)^5\right|\right] \\
&\leq m_i^{-4}B_iE\left[\left|\hat{m}_i - m_i\right|^5\right] \\
&= m_i^{-4}B_iE\left[\left((\hat{m}_i - m_i)^6\right)^{5/6}\right] \\
&\leq m_i^{-4}B_i\left(E[(\hat{m}_i - m_i)^6]\right)^{5/6} \\
&= o(n^{-2}),
\end{aligned} \tag{15}$$

we have

$$ED_I = \frac{1}{2n}\sum_{i=0}^{p}(1-m_i) + \frac{1}{6n^2}f^{(3)}(1)\sum_{i=0}^{p}(m_i^{-1}-3+2m_i) + \frac{1}{8n^2}f^{(4)}(1)\sum_{i=0}^{p}(m_i^{-1}-2+m_i) + o(n^{-2}),$$

which is equivalent to the result (**??**) because $\sum_{i=0}^{p} m_i = 1$. *Q.E.D.*

In particular, for the α-divergence,

$$\overset{\alpha}{D}[m : \hat{m}] \triangleq D_{f_\alpha}[m : \hat{m}], \qquad \overset{|\alpha|}{D}[m : \hat{m}] \triangleq \frac{1}{2}\Big\{D_{f_\alpha}[m : \hat{m}] + D_{f_{-\alpha}}[m : \hat{m}]\Big\},$$

where f_α is given by (**??**), the following results hold. (Sheena [**?**] gained this result as an example of the asymptotic risk of the m.l.e. for a general parametric model.)

Corollary 1. *Suppose that $\alpha \le 9$, then*

$$\overset{\alpha}{ED}_I \triangleq E\big[\overset{\alpha}{D}[m : \hat{m}]\big] = \frac{p}{2n} + \frac{1}{96n^2}\Big\{(\alpha-3)(3\alpha-7)(M-1) - 6(\alpha-3)(\alpha-1)p\Big\} + o(n^{-2}), \tag{16}$$

$$\overset{|\alpha|}{ED}_I \triangleq E\big[\overset{|\alpha|}{D}[m : \hat{m}]\big] = \frac{p}{2n} + \frac{1}{32n^2}\Big\{(\alpha^2+7)(M-1) - 2(\alpha^2+3)p\Big\} + o(n^{-2}). \tag{17}$$

–Proof–

From (**??**), we find that

$$f_\alpha^{(3)}(x) = \frac{(\alpha-3)}{2}x^{(\alpha-5)/2}, \qquad f_\alpha^{(4)}(x) = \frac{(\alpha-3)(\alpha-5)}{2^2}x^{(\alpha-7)/2},$$
$$f_\alpha^{(5)}(x) = \frac{(\alpha-3)(\alpha-5)(\alpha-7)}{2^3}x^{(\alpha-9)/2},$$

hence,

$$f_\alpha^{(3)}(1) = (\alpha-3)/2 \qquad f_\alpha^{(4)}(1) = (\alpha-3)(\alpha-5)/4,$$

and if $\alpha \le 9$, then $f_\alpha^{(5)}(x)$ is bounded on the interval $[\epsilon, \infty)$ for any ϵ (> 0). *Q.E.D.*

We observe the following points from (**??**), (**??**), and (**??**).

1. The main term, i.e., n^{-1}th-order term, is determined by p/n, that is the ratio of the dimension of the multinomial distribution model (the number of the free parameters) to the sample size. We call this the "p–n ratio" hereafter. The p–n ratio shows the complexity of the model to be estimated relative to the sample size. The main term is independent of f or α, and m_i $(i = 0, \ldots, p)$.

2. The second term, i.e., n^{-2}th-order term, depends on the parameter of the multinomial distribution through
$$M \triangleq \sum_{i=0}^{p} m_i^{-1}.$$
Here M attains the minimum value $(p+1)^2$ when $m_0 = m_1 = \cdots = m_p$. It increases rapidly if one of the m_i is near to zero. The effect of M on the risk depends on the choice of f or α. If you choose f such that $4f^{(3)}(1) + 3f^{(4)}(1)$ is non-positive or α such that $7/3 \le \alpha \le 3$, (**??**) and (**??**), respectively, decreases or is constant as M increases. This is rather unnatural because it contradicts our *beliefs* that the existence of a result with a small probability makes estimation

more difficult for a multinomial distribution. In this sense, the χ^2-distance with $\alpha = 3$ seems inappropriate, because it is asymptotically insensitive to the difference in the parameters m_i $(i = 0, \dots, p)$. (See Sheena [?], who reported that the α-divergence seems statistically unnatural when $|\alpha|$ is large for a regression model.) The α-divergence is a distance if and only if $\alpha = 0$, and the pair of α- and $-\alpha$-divergences work dually in the same manner as a distance. (For the "generalized Pythagorean theorem," see [?] or [?].) In this respect, the divergence $\overset{|\alpha|}{D}$ seems natural. In fact, (??) shows that the risk is a monotonically increasing function of M for any α.

3. The n^{-2}th-order term of (??) or (??) can be negative for some $f(\text{or } \alpha), p$, while that of (??) is always positive as

$$(\alpha^2+7)(M-1)-2(\alpha^2+3)p \geq (\alpha^2+7)((p+1)^2-1)-2(\alpha^2+3)p = p^2\alpha^2+7p^2+8p > 0.$$

2.2 Moving Intervals

First, we choose points $\lambda_i (1 \leq i \leq p)$ in the interval $(0, 1)$:

$$\lambda_0 (\triangleq 0) < \lambda_1 < \lambda_2 < \cdots < \lambda_p < \lambda_{p+1} (\triangleq 1). \tag{18}$$

Let

$$\xi_i \triangleq F^{-1}(\lambda_i),\ 1 \leq i \leq p, \quad \xi_0 \equiv -\infty, \quad \xi_{p+1} \equiv \infty,$$

where F^{-1} is the inverse function of the C.D.F., F, of the mother distribution. We call ξ the percentiles of the mother distribution.

In the moving intervals method, we estimate the percentiles of the mother distribution from the sample of the mother distribution, and use them as the endpoints of (??):

$$a_i = \hat{\xi}_i, \quad 1 \leq i \leq p, \tag{19}$$

where $\hat{\xi}_i$ is the estimator of ξ_i for $i = 1, \dots p$ and $\hat{\xi}_0 \equiv -\infty$ and $\hat{\xi}_{p+1} \equiv \infty$. In this case, the multinomial distribution that approximates the mother distribution has unknown parameters

$$\hat{m} \triangleq (\hat{m}_0, \dots, \hat{m}_p), \qquad \hat{m}_i \triangleq P(a_i, a_{i+1}) \equiv P(\hat{\xi}_i, \hat{\xi}_{i+1}) \quad 0 \leq i \leq p,$$

whereas it is estimated as

$$m \triangleq (m_0, \dots, m_p), \qquad m_i \triangleq \lambda_{i+1} - \lambda_i \quad 0 \leq i \leq p.$$

Although there are several ways to estimate the percentile ξ, we focus here on the simple estimator using the order statistic itself. Take an i.i.d. sample of size n from the mother distribution and let the ordered sample be denoted by

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

We estimate ξ_i by

$$\hat{\xi}_i \triangleq X_{(n_i)} \quad 1 \le i \le p, \tag{20}$$

where n_i is a function of n with the values in $\{1, 2, \ldots, n\}$. Let r_i denote the gap between n_i and $n\lambda_i$, namely

$$r_i \triangleq n_i - n\lambda_i \quad 1 \le i \le p, \qquad r_0 \triangleq 0, \qquad r_{p+1} \triangleq 1. \tag{21}$$

We measure the discrepancy between m and $\hat{m}$ by f-divergence,

$$D_f[m : \hat{m}] \triangleq \sum_{i=0}^{p} m_i\, f\left(\frac{\hat{m}_i}{m_i}\right). \tag{22}$$

If one might think it is natural to consider $D_f[\hat{m} : m]$ in the sense that the true parameter should come first, it is satisfied by using the dual function f^* (see (**??**)). Hence, we proceed with (**??**).

The risk for the moving interval method is given by

$$ED_P \triangleq E\big[D_f[m : \hat{m}]\big],$$

and the following result holds.

Theorem 2. *Suppose that $f^{(5)}$ exists on $(0, \infty)$ and is bounded on $[\epsilon, \infty)$ for any $\epsilon(> 0)$. If $r_i/n = o(n_i^{-1/2})$, then we have*

$$\begin{aligned} ED_P = \frac{p}{2n} + \frac{1}{24n^2}\Big[&-24 - 36p + 12\sum_{i=0}^{p}(r_{i+1} - r_i)(r_{i+1} - r_i + 1)m_i^{-1} \\ &+ 4f^{(3)}(1)\Big\{-5 - 9p + \sum_{i=0}^{p}\big(3(r_{i+1} - r_i) + 2\big)m_i^{-1}\Big\} \\ &+ f^{(4)}(1)\Big\{-3 - 6p + 3\sum_{i=0}^{p} m_i^{-1}\Big\}\Big] + o(n^{-2}). \end{aligned} \tag{23}$$

–Proof–
The whole process of the proof is lengthy; hence, we only state the outline of the proof here. All the details can be found in the Appendix of [**?**]. Let

$$U_{(n_i)} \triangleq F(X_{(n_i)}), \qquad \Delta_i \triangleq \sqrt{n}(U_{(n_i)} - \lambda_i) \quad 1 \le i \le p$$

and $\Delta_0 \triangleq 0,\ \Delta_{p+1} \triangleq 0$. The following relationship holds for $0 \le i \le p$:

$$\begin{aligned} \hat{m}_i &= F(\hat{\xi}_{i+1}) - F(\hat{\xi}_i) \\ &= F(X_{(n_{i+1})}) - F(X_{(n_i)}) \\ &= U_{(n_{i+1})} - U_{(n_i)} \\ &= \lambda_{i+1} - \lambda_i + n^{-1/2}(\Delta_{i+1} - \Delta_i) \end{aligned}$$

$$= m_i + n^{-1/2}(\Delta_{i+1} - \Delta_i). \tag{24}$$

Note that

$$(\Delta_1, \dots, \Delta_p) \xrightarrow{d} N_p(0, \Sigma),$$

where

$$\Sigma = (\sigma_{ij}) = \lambda_i(1-\lambda_j) \quad 1 \le i \le j \le p$$

(see e.g., Theorem 5.4.5 of [?]). Similarly to (**??**), the following equation holds:

$$D_f[m : \hat{m}] = \frac{1}{2}\sum_{i=0}^{p} m_i R_i^2 + \frac{1}{6} f^{(3)}(1) \sum_{i=0}^{p} m_i R_i^3 + \frac{1}{24} f^{(4)}(1) \sum_{i=0}^{p} m_i R_i^4$$
$$+ \frac{1}{120}\sum_{i=0}^{p} f^{(5)}(1+R_i^*) m_i R_i^5, \tag{25}$$

where R_i^* is a function of $\hat{m}_i$ and there is some positive constant B_i such that

$$|f^{(5)}(1+R_i^*)| \le B_i.$$

Using this boundedness and the fact $E[R_i^6] = O(n^{-3})$, completely similarly to (**??**) , we have

$$E[f^{(5)}(1+R_i^*) m_i R_i^5] = o(n^{-2}),$$

hence,

$$ED_P = \frac{1}{2}\sum_{i=0}^{p} m_i E[R_i^2] + \frac{1}{6} f^{(3)}(1) \sum_{i=0}^{p} m_i E[R_i^3] + \frac{1}{24} f^{(4)}(1) \sum_{i=0}^{p} m_i E[R_i^4] + o(n^{-2}). \tag{26}$$

After a long but straightforward calculation (see the Appendix of [?]), we have

$$\sum_{i=0}^{p} m_i E[R_i^2] = n^{-1}p + n^{-2}[-2 - 3p + \sum_{i=0}^{p}(r_{i+1} - r_i)(r_{i+1} - r_i + 1)m_i^{-1}],$$
$$\sum_{i=0}^{p} m_i E[R_i^3] = n^{-2}[-5 - 9p + \sum_{i=0}^{p}\big(3(r_{i+1} - r_i) + 2\big)m_i^{-1}],$$
$$\sum_{i=0}^{p} m_i E[R_i^4] = n^{-2}[-3 - 6p + 3\sum_{i=0}^{p} m_i^{-1}].$$

If we insert these results into (**??**), we have the result. *Q.E.D.*

We also have the following formulas for the α-divergence.

Corollary 2. *Suppose that $\alpha \le 9$ and $r_i/n = o(n_i^{-1/2})$, then*

$$\overset{\alpha}{E}D_P = \frac{p}{2n} + \frac{1}{96n^2}\Big[-\alpha^2(3+6p) - \alpha(16+24p) - 18p - 21$$

$$+\sum_{i=0}^{p}\Big\{48(r_{i+1}-r_i)^2+24(\alpha-1)(r_{i+1}-r_i)+3\alpha^2-8\alpha-3\Big\}m_i^{-1}\Big]$$

$$+o(n^{-2}), \tag{27}$$

$$\overset{|\alpha|}{ED}_P=\frac{p}{2n}+\frac{1}{96n^2}\Big[-\alpha^2(3+6p)-18p-21$$

$$+\sum_{i=0}^{p}\Big\{48(r_{i+1}-r_i)^2-24(r_{i+1}-r_i)+3\alpha^2-3\Big\}m_i^{-1}\Big]+o(n^{-2}) \tag{28}$$

–Proof–

We can derive the result from (**??**) in the same way as Corollary **??**. *Q.E.D.*

We make some comments on ED_P, $\overset{\alpha}{ED}_P$, and $\overset{|\alpha|}{ED}_P$.

1. The main term is half the p–n ratio just like ED_I. It is independent of f or α, and $m_i (i=0,\dots,p)$.
2. The risk is independent of the mother distribution (it is due to the fact (**??**)). It is determined by our choice of m_i or equivalently λ_i in (**??**).
3. The choice of n_i or equivalently r_i $(i=1,\dots,p)$ effects the n^{-2}th-order term. It is possible that the coefficient of m_i^{-1} could be negative for some r_i and f (or α). In this case, small m_i could reduce the risk.

The most natural selection of n_i is $[n\lambda_i]$ or $[n\lambda_i]+1$, where $[\ \cdot\ ]$ is the Gauss symbol. Let

$$\bar{r}_i \triangleq [n\lambda_i]-n\lambda_i. \tag{29}$$

In this paper, we consider the following randomized choice between $[n\lambda_i]$ and $[n\lambda_i]+1$:

$$P(r_i=\bar{r}_i)\Big(=P(n_i=[n\lambda_i])\Big)=1+\bar{r}_i,\quad P(r_i=1+\bar{r}_i)\Big(=P(n_i=[n\lambda_i]+1)\Big)=-\bar{r}_i \tag{30}$$

for $1\le i\le p$, whereas $r_0\equiv 0$ and $r_{p+1}\equiv 1$ as in (**??**). This is natural in that n_i is chosen to be $[n\lambda_i]$ and $[n\lambda_i]+1$, respectively, with the probabilities proportional to the closeness to both points. Locating $\hat{\xi}_i$ between $X_{([n\lambda_i])}$ and $X_{([n\lambda_i]+1)}$ according to $\bar{r}_i$ is another appealing idea. However, if we adopt this estimation of ξ_i, then the risk depends on the mother distribution. We do not study this case here.

Let

$$ED_P^*\triangleq E[ED_P],\quad \overset{\alpha}{ED}{}_P^*\triangleq E[\overset{\alpha}{ED}_P],\quad \overset{|\alpha|}{ED}{}_P^*\triangleq E[\overset{|\alpha|}{ED}_P],$$

where the entire expectation is taken with respect to the distribution (**??**). The following results hold for the randomized choice of r_i (**??**).

Proposition 1. *We have*

$$ED_P^*=\frac{p}{2n}+\frac{1}{48n^2}\Big[-48-72p+24\Big\{-\bar{r}_1(1+\bar{r}_1)m_0^{-1}+\big(2-\bar{r}_p(1+\bar{r}_p)\big)m_p^{-1}$$

$$- \sum_{i=1}^{p-1} \big(\bar{r}_i(1+\bar{r}_i) + \bar{r}_{i+1}(1+\bar{r}_{i+1})\big) m_i^{-1} \Big\}$$

$$+ 8f^{(3)}(1) \Big\{ -5 - 9p + 2\sum_{i=0}^{p} m_i^{-1} + 3m_p^{-1} \Big\}$$

$$+ 2f^{(4)}(1) \Big\{ -3 - 6p + 3\sum_{i=0}^{p} m_i^{-1} \Big\} \Big] + o(n^{-2}),$$

For $\alpha \le 9$,

$$\overset{\alpha}{E}D_P^* = \frac{p}{2n} + \frac{1}{96n^2} \Big[-\alpha^2(3+6p) - \alpha(16+24p) - 18p - 21$$

$$- 48\bar{r}_1(1+\bar{r}_1)m_0^{-1} + \big(-48\bar{r}_p(1+\bar{r}_p) + 24(\alpha+1)\big)m_p^{-1}$$

$$- 48\sum_{i=1}^{p-1} \big(\bar{r}_i(1+\bar{r}_i) + \bar{r}_{i+1}(1+\bar{r}_{i+1})\big) m_i^{-1}$$

$$+ (3\alpha^2 - 8\alpha - 3)\sum_{i=0}^{p} m_i^{-1} \Big] + o(n^{-2}),$$

For $\alpha \le 9$,

$$\overset{|\alpha|}{E}D_P^* = \frac{p}{2n} + \frac{1}{96n^2} \Big[-\alpha^2(3+6p) - 18p - 21$$

$$- 48\bar{r}_1(1+\bar{r}_1)m_0^{-1} + \big(24 - 48\bar{r}_p(1+\bar{r}_p)\big)m_p^{-1}$$

$$- 48\sum_{i=1}^{p-1} \big(\bar{r}_i(1+\bar{r}_i) + \bar{r}_{i+1}(1+\bar{r}_{i+1})\big) m_i^{-1}$$

$$+ (3\alpha^2 - 3)\sum_{i=0}^{p} m_i^{-1} \Big] + o(n^{-2}).$$

–Proof–
As was proved in the Appendix of [**?**], the following results hold:

$$E[r_i] = 0 \text{ for } 0 \le i \le p, \qquad E[r_{p+1}] = 1$$
$$E[r_i^2] = -\bar{r}_i(1+\bar{r}_i) \text{ for } 1 \le i \le p, \qquad E[r_0^2] = 0, \qquad E[r_{p+1}^2] = 1$$
$$E[r_i r_{i+1}] = 0 \text{ for } 0 \le i \le p.$$

Applying these results to $E[(r_{i+1} - r_i)^2] = E[r_i^2] + E[r_{i+1}^2] - 2E[r_i r_{i+1}]$ and $E[r_{i+1}] - E[r_i]$ in (**??**), (**??**), and (**??**), we have the results. *Q.E.D.*

2.3 Comparison of the two methods

We compare the risks between the fixed interval method and the moving interval method. For both methods, the main term (n^{-1}th-order term) are common, but we can see some difference in the second term (n^{-2}th-order term). The biggest difference between the two methods lies in m_i. In the fixed interval method, m_i depend on the unknown mother distribution; hence, we are unable to control them. As we observed in Section **??**, if they include even one small m_i near to zero, then the (asymptotic) risk becomes extremely high through M. The more intervals (endpoints) we use for discretization, the more likely we are to have small m_i. Even if we have a large sample set, we have to be cautious about increasing the dimension of the multinomial distribution. In contrast, for the moving interval method, m_i are controllable. We can choose m_i so that the risk attains as small a value as possible.

To make the comparison between the two methods, we need to evaluate the upper bounds for ED_P^*, $\overset{\alpha}{ED}{}_P^*$, and $\overset{|\alpha|}{ED}{}_P^*$. From the inequalities

$$-1 \le \bar{r}_i \le 0,\ 0 \le -\bar{r}_i(1+\bar{r}_i) \le 1/4, \quad \text{for } i = 1, \dots, p,$$

we have the following relations:

$$\begin{aligned}
ED_P^* \le \frac{p}{2n} + \frac{1}{48n^2}\Big[&-48 - 72p + 6\Big\{m_0^{-1} + 9m_p^{-1} + 2\sum_{i=1}^{p-1} m_i^{-1}\Big\} \\
&+ 8f^{(3)}(1)\Big\{-5 - 9p + 2\sum_{i=0}^{p} m_i^{-1} + 3m_p^{-1}\Big\} \\
&+ 2f^{(4)}(1)\Big\{-3 - 6p + 3\sum_{i=0}^{p} m_i^{-1}\Big\}\Big] + o(n^{-2}) \quad \Big(\mathit{say}\ \overline{ED}{}_P^*\Big), \\
\overset{\alpha}{ED}{}_P^* \le \frac{p}{2n} + \frac{1}{96n^2}\Big[&-\alpha^2(3+6p) - \alpha(16+24p) - 18p - 21 \\
&+ 12m_0^{-1} + (24\alpha + 36)m_p^{-1} \\
&+ 24\sum_{i=1}^{p-1} m_i^{-1} + (3\alpha^2 - 8\alpha - 3)\sum_{i=0}^{p} m_i^{-1}\Big] + o(n^{-2}) \\
= \frac{p}{2n} + \frac{1}{96n^2}\Big[&-\alpha^2(3+6p) - \alpha(16+24p) - 18p - 21 \\
&+ (3\alpha^2 - 8\alpha + 9)m_0^{-1} + (3\alpha^2 + 16\alpha + 33)m_p^{-1} \\
&+ (3\alpha^2 - 8\alpha + 21)\sum_{i=1}^{p-1} m_i^{-1}\Big] + o(n^{-2}) \quad \Big(\mathit{say}\ \overline{\overset{\alpha}{ED}}{}_P^*\Big), \\
\overset{|\alpha|}{ED}{}_P^* \le \frac{p}{2n} + \frac{1}{32n^2}\Big[&-\alpha^2(1+2p) - 6p - 7 + (\alpha^2+3)m_0^{-1} + (\alpha^2 + 11)m_p^{-1} \\
&+ (\alpha^2 + 7)\sum_{i=1}^{p-1} m_i^{-1}\Big] + o(n^{-2}) \quad \Big(\mathit{say}\ \overline{\overset{|\alpha|}{ED}}{}_P^*\Big).
\end{aligned}$$

If we choose the equal right-end and left-end probabilities, i.e., $m_0 = m_p$,

$$\overline{\overset{|\alpha|}{ED}}{}_P^* = \frac{p}{2n} + \frac{1}{32n^2}\Big[-\alpha^2(1+2p) - 6p - 7 + (\alpha^2+7)M\Big] + o(n^{-2}). \tag{31}$$

This upper bound for $\overset{|\alpha|}{ED}{}^*_P$ is affected by m_i only through M just like (??). This indicates that the choice of equally valued m_i, that is, $m_i = 1/(p+1),\ i = 1, \ldots, p$, is reasonable for the estimation of the mother distribution. It is needless to say that the percentiles with a common increment ("quantiles") are most often used in a practical situation. If we choose "quantiles" for the moving interval method, we have the following result for $\alpha \le 9$,

Theorem 3. *Suppose that $\alpha \le 9$. Set λ_i in (??) so that $m_i = 1/(p+1),\ i = 0, \ldots, p$, then asymptotically (exactly speaking, as for the comparison up to the n^{-2}th-order term), the following inequality holds:*

$$\overset{|\alpha|}{ED}{}_I \ge \overset{|\alpha|}{ED}{}^*_P. \tag{32}$$

–Proof–
As $M \ge (p+1)^2$, from (??), we have

$$\overset{|\alpha|}{ED}{}_I \ge \frac{p}{2n} + \frac{1}{32n^2}\Big\{(\alpha^2+7)(p^2+2p) - 2(\alpha^2+3)p\Big\} + o(n^{-2}) \quad \Big(\text{say } \underline{\overset{|\alpha|}{ED}}{}_I\Big),$$

whereas when $m_i = 1/(p+1),\ i = 0, \ldots, p$, $\overline{\overset{|\alpha|}{ED}}{}^*_P$ equals

$$\frac{p}{2n} + \frac{1}{32n^2}\Big[-\alpha^2(1+2p) - 6p - 7 + (\alpha^2+7)(p^2+2p+1)\Big] + o(n^{-2}).$$

Up to the n^{-2}th-order term, we have

$$\overset{|\alpha|}{ED}{}_I - \overset{|\alpha|}{ED}{}^*_P \ge \underline{\overset{|\alpha|}{ED}}{}_I - \overline{\overset{|\alpha|}{ED}}{}^*_P = 0. \tag{33}$$

Q.E.D.

The above theorem states that even if we are lucky enough to choose the best intervals (that is, equiprobable intervals) for the fixed interval method, it is asymptotically dominated by the moving interval method with "quantiles." We can conclude that if we can choose both methods, it is better, at least asymptotically, to use the moving interval method.

We also present a numerical comparison between the methods.

– Example 1 –
First we take the example given in the introduction again where the mother distribution is $Beta(2,5)$. Suppose that we have prior knowledge that the mother distribution has the support $[0,1]$, and set a_i as $a_i = i/10\ (1 \le i \le 9)$ for the fixed intervals. The m_i for the moving interval method with "quantiles" are $m_i = 1/10 (0 \le i \le 9)$. The corresponding probabilities for the fixed intervals are given by

$$(m_0, m_1, \ldots, m_9) \doteqdot (0.114, 0.230, 0.235, 0.187, 0.124, 0.068, 0.030, 0.009, 0.002, 5.5\times10^{-5}).$$

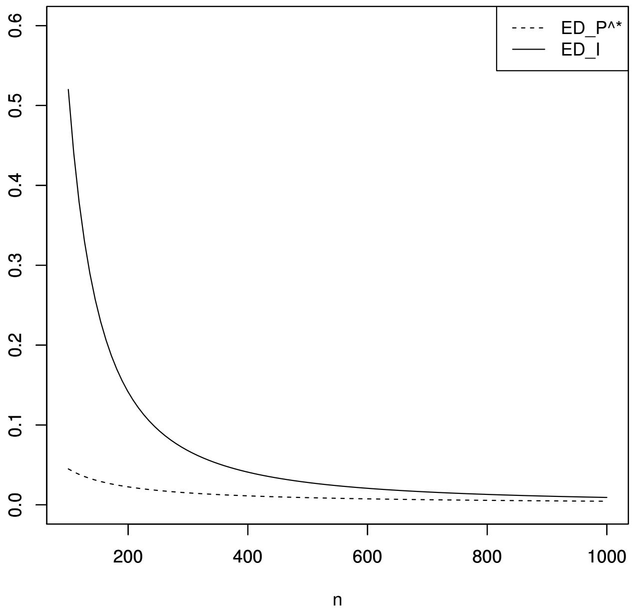


Figure 2: Approximated $\overset{|1|}{ED}{}_P^*$ and $\overset{|1|}{ED}_I$ for $Beta(2, 5)$

Note that even if we have the information on the compact support, the fixed intervals made by equally dividing the support could contain that with a very small probability such as 5.5×10^{-5}.

Let us skip the $o(n^{-2})$ part of $\overset{|\alpha|}{ED}_I$ ($\overset{|\alpha|}{ED}{}_P^*$), and call it the approximated $\overset{|\alpha|}{ED}_I$ ($\overset{|\alpha|}{ED}{}_P^*$). The graphs of the approximated risks as n varies for both methods with $\alpha = 1$ are shown in Figure **??**. (Note that $|\alpha|$ are skipped in the legend.)

We propose one indicator for the comparison between the two methods. Let us consider the approximated $\overset{|\alpha|}{ED}_I$ and $\overset{|\alpha|}{ED}{}_P^*$ as the functions of n and we have the equation

$$(\text{The approximated } \overset{|\alpha|}{ED}_I)(n) = \ (\text{The approximated } \overset{|\alpha|}{ED}{}_P^*)(100). \tag{34}$$

The solution of this equation indicates how large a sample is required for the approximated $\overset{|\alpha|}{ED}_I$ to attain the same risk as that of the approximated $\overset{|\alpha|}{ED}{}_P^*$ with $n = 100$. The solution under the present setting when $\alpha = 1$ is given by $n \doteqdot 379$. Even if we are lucky enough to know the finite support of the mother distribution, the fixed interval method is still quite inefficient to the moving interval method.

We carried out a more general comparison between the two methods. We calculated the solution of Equation (**??**) with $\alpha = 1$ for various values of the two shape parameters ν_1 and ν_2 of $Beta(\nu_1, \nu_2)$. (The a_i and m_i are the same as mentioned previously.)

The result is shown in Table **??**. We observe the sample size n significantly changes as the mother distribution varies. For example, with $\nu_1 = 0.5$ fixed, when ν_2 is 0.5 or

Table 3: Required Sample Size $-Beta(\nu_1, \nu_2)-$

ν_1	ν_2	n	ν_1	ν_2	n	ν_1	μ_2	n	ν_1	μ_2	n	ν_1	μ_2	n
0.5	0.5	106	1.5	1.5	106	2.5	2.5	110	3.5	3.5	123	4.5	4.5	152
0.5	1	107	1.5	2	108	2.5	3	115	3.5	4	133	4.5	5	172
0.5	1.5	111	1.5	2.5	113	2.5	3.5	127	3.5	4.5	157	5	5	176
0.5	2	122	1.5	3	128	2.5	4	153	3.5	5	205			
0.5	2.5	148	1.5	3.5	159	2.5	4.5	203	4	4	135			
0.5	3	204	1.5	4	219	2.5	5	294	4	4.5	149			
0.5	3.5	312	1.5	4.5	330	3	3	115	4	5	182			
0.5	4	513	1.5	5	526	3	3.5	122						
0.5	4.5	881	2	2	107	3	4	139						
0.5	5	1548	2	2.5	110	3	4.5	175						
1	1	105	2	3	119	3	5	240						
1	1.5	106	2	3.5	138									
1	2	111	2	4	177									
1	2.5	122	2	4.5	250									
1	3	148	2	5	379									
1	3.5	201												
1	4	302												
1	4.5	485												
1	5	812												

1.0, it is very close to 100, the benchmark sample size of the moving interval method, whereas it jumps up to 1548 when the mother distribution is $Beta(0.5, 5)$.

– *Example 2* –

Now we take as the mother distributions Pearson type IV distributions (in fact, the Beta distribution in the first example is a Pearson type I distribution), which consists of skewed-t-type distributions. Its density $f(x|m, \nu)$ is given by

$$f(x|m,\nu) \propto (1+x^2)^{-m} \exp\Big(-\nu \arctan(x)\Big), \quad -\infty < x < \infty,\ m > 1/2, -\infty < \nu < \infty.$$

We calculated the solution for (**??**) with $\alpha = 1$ under various values of m and ν. See Table **??**, where two cases (a) and (b) are treated:

$$\begin{aligned} &\text{Case (a)} \quad a_i = -1.0 + (i-1) * 0.2 \ (1 \le i \le 11) \text{ for the fixed intervals} \\ &\qquad\qquad\ m_i = 1/12 \ (0 \le i \le 11) \text{ for the moving intervals} \qquad (35) \\ &\text{Case (b)} \quad a_i = -2.0 + (i-1) * 0.5 \ (1 \le i \le 9) \text{ for the fixed intervals} \\ &\qquad\qquad\ m_i = 1/10 \ (0 \le i \le 9) \text{ for the moving intervals} \qquad (36) \end{aligned}$$

We observe that the relative inefficiency of the fixed interval method largely depends on the determination of the intervals and the mother distribution. More specifically, we note that when m is fixed and ν increases (i.e., the skewness increases), the required sample size rapidly increases. This is more remarkable for the case (b). For example,

Table 4: Required Sample Size $-PearsonIV(m,\nu)-$

(a) $a = (-1, -0.8, \ldots, 1.0)$

m	ν	n	m	ν	n	m	ν	n
1.5	0.5	104	2.5	0.5	105	3.5	0.5	108
1.5	1	107	2.5	1	107	3.5	1	110
1.5	1.5	112	2.5	1.5	111	3.5	1.5	115
1.5	2	123	2.5	2	119	3.5	2	124
1.5	2.5	142	2.5	2.5	132	3.5	2.5	139
1.5	3	177	2.5	3	155	3.5	3	162
1.5	3.5	236	2.5	3.5	191	3.5	3.5	196
1.5	4	336	2.5	4	250	3.5	4	249
1.5	4.5	501	2.5	4.5	342	3.5	4.5	329
1.5	5	775	2.5	5	487	3.5	5	449
1.5	5.5	1232	2.5	5.5	715	3.5	5.5	630
1.5	6	1999	2.5	6	1079	3.5	6	906
2	0.5	104	3	0.5	106	4	0.5	110
2	1	106	3	1	108	4	1	113
2	1.5	111	3	1.5	113	4	1.5	119
2	2	119	3	2	121	4	2	130
2	2.5	134	3	2.5	134	4	2.5	146
2	3	159	3	3	156	4	3	170
2	3.5	202	3	3.5	191	4	3.5	207
2	4	272	3	4	244	4	4	261
2	4.5	386	3	4.5	327	4	4.5	342
2	5	569	3	5	454	4	5	461
2	5.5	866	3	5.5	651	4	5.5	637
2	6	1350	3	6	955	4	6	901

(b) $a = (-2, -1.5, \ldots, 2.0)$

m	ν	n	m	ν	n	m	ν	n
1.5	0.5	110	2.5	0.5	126	3.5	0.5	179
1.5	1	115	2.5	1	137	3.5	1	206
1.5	1.5	127	2.5	1.5	160	3.5	1.5	255
1.5	2	153	2.5	2	200	3.5	2	334
1.5	2.5	202	2.5	2.5	267	3.5	2.5	457
1.5	3	292	2.5	3	377	3.5	3	648
1.5	3.5	453	2.5	3.5	559	3.5	3.5	946
1.5	4	739	2.5	4	861	3.5	4	1416
1.5	4.5	1252	2.5	4.5	1365	3.5	4.5	2165
1.5	5	2174	2.5	5	2216	3.5	5	3368
1.5	5.5	3845	2.5	5.5	3665	3.5	5.5	5323
1.5	6	6890	2.5	6	6152	3.5	6	8528
2	0.5	115	3	0.5	145	4	0.5	232
2	1	123	3	1	163	4	1	273
2	1.5	138	3	1.5	197	4	1.5	346
2	2	169	3	2	252	4	2	460
2	2.5	222	3	2.5	342	4	2.5	634
2	3	315	3	3	484	4	3	898
2	3.5	474	3	3.5	710	4	3.5	1303
2	4	746	3	4	1076	4	4	1930
2	4.5	1215	3	4.5	1672	4	4.5	2910
2	5	2032	3	5	2651	4	5	4459
2	5.5	3464	3	5.5	4277	4	5.5	6928
2	6	5992	3	6	6999	4	6	10899

in the case (b) with $m = 1.5,\ \nu = 0.5$, required n equals 110, whereas a sample size as large as 6890 is required when $m = 1.5,\ \nu = 6.0$. In fact, the probability of the mother distribution for each fixed interval in (**??**) is given by

$$(m_0, m_1, \ldots, m_9) \doteqdot (0.094, 0.050, 0.092, 0.167, 0.237, 0.189, 0.090, 0.038, 0.017, 0.027)$$

when $m = 1.5,\ \nu = 0.5$, and

$$\begin{aligned}(m_0, m_1, \ldots, m_9) \doteqdot &(7.784 \times 10^{-1}, 9.938 \times 10^{-2}, 7.769 \times 10^{-2}, 3.689 \times 10^{-2}, 7.092 \times 10^{-3} \\ &4.596 \times 10^{-4}, 2.202 \times 10^{-5}, 2.009 \times 10^{-6}, 3.655 \times 10^{-7}, 1.946 \times 10^{-7})\end{aligned}$$

when $m = 1.5,\ \nu = 6.0$. As an example, the graphs of the approximated $\overset{|1|}{ED}_I$ and $\overset{|1|}{ED}{}^*_P$ as n varies are drawn in Figure **??** when $m = 1.5,\ \nu = 3$ in case (b).

We saw that the moving interval method is superior theoretically and numerically to the fixed interval method as an estimation of the mother distribution. Now recall the question raised in the introduction: "Which is more efficient as an estimator of the approximated mother distribution, histogram or E.P.G.?" We predict E.P.G. would be more efficient because histogram is a mixture of both methods as referenced in the introduction, whereas E.P.G. is genuinely based on the moving interval method.

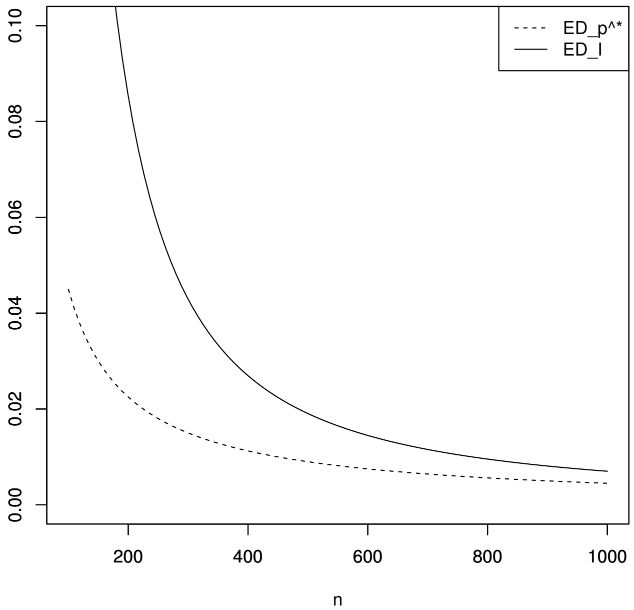


Figure 3: Approximated $\overset{|1|}{ED}{}^*_P$ and $\overset{|1|}{ED}_I$ for a Pearson Type IV distribution (m = 1.5, ν=3)

Acknowledgment

We would like to express our deepest gratitude to the anonymous referees and Hajo Holzmann, Editor in Chief of *Metrika*, for their useful advice and comments, which improved the quality of the paper.

References

[1] S. Amari. *Information Geometry and Its Applications*. Springer, 2016.

[2] S. Amari and A. Cichocki. Information geometry of divergence functions. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 58:183-195, 2010.

[3] S. Amari and H. Nagaoka. *Methods of Information Geometry*. Translations of Mathematical Monographs 191. American Mathematical Society, 2000.

[4] A. Barbiero. A general discretization procedure for reliability computation in complex stress–strength models *Mathematics and Computers in Simulation*, 82: 1667-1676, 2012,

[5] Z. Drezner and D. Zerom. A simple and effective discretization of a continuous

random variable *Communications in Statistics – Simulation and Computation*, 45: 3798-3810, 2016.

[6] J.R. English, T. Sargent, and T.L. Landers. A discretizing approach for stress/strength analysis *IEEE Transactions on Reliability*, 45: 84-89, 1996.

[7] E. L. Lehmann. *Elements of Large Sample Theory.* Springer, 1999.

[8] Y. Sheena. *Asymptotic expansion of the risk of maximum likelihood estimator with respect to α-divergence as a measure of the difficulty of specifying a parametric model, to be published in Communications in Statistics – Theory and Methods.*

[9] Y. Sheena. Asymptotic Expansion of Risk for a Regression Model with respect to α-Divergence with an Application to the Sample Size Problem. *Far East Journal of Theoretical Statistics*, 53, 187-230, 2017.

[10] Y. Sheena. Estimation of a Continuous Distribution on the real line by Discretization Methods –Complete Version–. arXiv:1709.09520.

[11] I. Vajda. *Theory of Statistical Inference and Information*, Kluwer Academic Publishers, 1989.