

Maximum Entropy Approximation of Small Sample Distributions

Ximing Wu* and Suojin Wang[†]

February 23, 2006

Abstract

In this paper, we propose a simple approach for approximating small sample distributions based on the method of Maximum Entropy (maxent) density. Unlike previous studies that use the maxent method as a curve-fitting tool, we focus on the “small sample asymptotics” of the approximation. We show that this method obtains the same asymptotic order of accuracy as that of the classical Edgeworth expansion and the exponential Edgeworth expansion by Field and Hampel (1982). In addition, the maxent approximation attains the minimum Kullback-Leibler distance among all approximations of the same exponential family. For symmetric fat-tailed distributions, we adopt the dampening approach in Wang (1992) to ensure the integrability of maxent densities and effectively extend the maxent method to accommodate all admissible skewness and kurtosis space. Our numerical experiments show that the proposed method compares competitively with and often outperforms the Edgeworth and exponential Edgeworth expansions, especially when the sample size is small.

1 Introduction

Asymptotic approximation of density or distribution functions of estimators or test statistics is commonly used for statistics inference, such as approximate confidence limits for a parameter of interest or p -values for a hypothesis test. Use of the leading terms of the approximation leads to inference based on the limiting form of the distribution of the statistic. Under certain regularity conditions, the statistic is often distributed as a normal distribution asymptotically according to the central limit theorem. However, it is well known that the normal approximation of the statistic can be inadequate for small sample problems. Fortunately, use of further, usually higher-order, terms can often improve the accuracy of the approximation.

*Department of Agricultural Economics, Texas A&M University; email: xwu@ag.tamu.edu.

[†]Department of Statistics, Texas A&M University; email: sjwang@stat.tamu.edu.

The most commonly used high-order approximation method is the Edgeworth expansion. The Edgeworth expansion has its root in the Charlier differential series, which approximates a distribution $F(x)$ by

$$F(x) = \exp \left\{ \sum_{r=1}^{\infty} (\kappa_{F,r} - \kappa_{G,r}) \frac{(-D)^r}{r!} \right\} G(x), \quad (1)$$

where $G(x)$ is a developing function, $\kappa_{F,r}$ and $\kappa_{G,r}$ are the r^{th} cumulants of $F(x)$ and $G(x)$ respectively, and D denotes the differential operator. If $F(x)$ is asymptotically normal, a natural candidate for the developing function $G(x)$ is the normal distribution. In this case $G(x) = \Phi(x)$ and $(-D)^r \phi(x) = H_r(x) \phi(x)$, where $\phi(x)$ and $\Phi(x)$ are standard normal density and cumulative distribution function respectively, and $H_r(x)$ is the Hermite polynomial of degree r . The Gram-Charlier expansion uses the normal distribution and collects terms according to the order of derivatives. The seminal work by Edgeworth (1905) applies this expansion to the distribution of the sum of n independently and identically distributed random variables and expands (1) using the normal distribution but by collecting terms according to the powers of n , leading to the so-called Edgeworth expansion.

Suppose Y_1, \dots, Y_n is an independently and identically distributed sample from a distribution with mean zero for simplicity and variance σ^2 . Denote $X_n = \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n Y_i$ and $f_n(x)$ the density function of X_n . A two-term Edgeworth expansion takes the form

$$g_{n,4}(x) = \phi(x) \left[1 + \frac{\kappa_3}{3!} \frac{H_3(x)}{\sqrt{n}} + \frac{\kappa_4}{4!} \frac{H_4(x)}{n} + \frac{10\kappa_3^2}{6!} \frac{H_6(x)}{n} \right],$$

where κ_j is the j^{th} cumulant of $F(y)$; the first subscript of $g(x)$ indicates the sample size and the second indicates the highest order cumulant used in the approximation. The Edgeworth expansion is known to work well in the central part of a distribution, but the tail approximation can be rather inaccurate. Sometimes it produces negative densities.

The Edgeworth expansion is a high-order expansion around 0, which explains its somewhat unsatisfactory tail performance. Hampel (1973) introduces an alternative ‘‘small sample asymptotic’’ method, which is essentially a local Edgeworth expansion. This method uses a low-order expansion at each point separately and then integrates the results rather than use a high-order expansion around a single point. As accurate as it is, it requires the knowledge of the true distribution, which limits its applicability in practice.¹ Field and Hampel (1982) proposes an alternative expansion which does not require the knowledge of the underlying distribution. This expansion is based on Hampel (1973)’s ‘‘small sample asymptotic’’ approximation to the distribution of $T_n = \frac{1}{n} \sum_{i=1}^n Y_i$. Instead of approximating the density directly, Hampel’s method approximates the derivative of the logarithm of $f_n(x)$,

$$-K_n(t) = f'_n/f_n(t) = -n\alpha(t) - \beta(t) + o(n^{-1}) \quad (2)$$

with some functions $\alpha(t)$ and $\beta(t)$.

¹Another closely-related alternative is the saddlepoint approximation, which requires the knowledge of cumulant generating function.

Taking the Taylor expansion of $\alpha(t)$ and $\beta(t)$ and then integrating both sides of (2) yields an approximation to $\log f_n(t)$. Finally, substituting in $x = t\sqrt{n}/\sigma$ and exponentiating both sides yields the alternative approximation to $f_n(x)$ for $X_n = \frac{1}{\sqrt{n\sigma}} \sum_{i=1}^n Y_i$

$$h_{n,4}(x) = \phi(x) \exp\left(\frac{3\kappa_4 - 5\kappa_3^2}{24n} - \frac{\kappa_3}{2\sqrt{n}}x + \frac{2\kappa_3^2 - \kappa_4}{4n}x^2 + \frac{\kappa_3}{6\sqrt{n}}x^3 + \frac{\kappa_4 - 3\kappa_3^2}{24n}x^4\right). \quad (3)$$

One can easily verify that the Edgeworth expansion $g_{n,4}$ can be derived directly from $h_{n,4}$ by taking the Taylor expansion of its exponent and collecting terms to the order of $O(n^{-1})$. In the rest of the paper, we call this approximation ‘‘exponential Edgeworth’’ expansion. Field and Hampel (1982) notes that ‘‘..., (the Edgeworth expansion) is quite good but can apparently still be improved by putting the expansion back into the exponent where it arrives naturally by integration of the expansion for $K_n(t)$ around $t = 0$.’’ This observation is confirmed by their examples. However, for a distribution defined on the entire real line, $h_{n,4}$ may not be integrable for some cumulants. For example, $h_{n,4}$ will explode for large $|x|$ if $\kappa_4 - 3\kappa_3^2 > 0$. This restriction limits the domain of ‘‘admissible cumulants’’ considerably.

We note that one can rewrite the exponential Edgeworth expansion (3) into the canonical form

$$f_{n,K}(x) = \exp\left(\sum_{k=1}^K \gamma_k x^k - \Psi(\gamma)\right), \quad (4)$$

where $\Psi(\gamma) = \log\left[\int \exp\left(\sum_{k=1}^K \gamma_k x^k\right) dx\right]$ is the rescaling factor that ensures that $f_{n,K}(x)$ integrates to one. Suppose that the logarithm of true density $\log f_n(x)$ can be approximated by

$$\log f_n(x) = \sum_{k=0}^{\infty} \gamma_k x^k.$$

It follows that

$$f_n(x) = \exp\left(\sum_{k=1}^{\infty} \gamma_k x^k - \Psi(\gamma)\right). \quad (5)$$

Assuming that $\gamma_k \neq 0$ for some $k > K$, we have

$$\mu_j(f_{n,K}) = \int x^j f_{n,K}(x) dx \neq \int x^j f_n(x) dx = \mu_j(f_n), 1 \leq j \leq K.$$

Therefore, the first K moments of $f_{n,K}$, $\mu(f_{n,K}) = [\mu_1(f_{n,K}), \dots, \mu_K(f_{n,K})]$, are different from those of $f_n(x)$ due to the truncation of the polynomial in the exponent of Equation (5). One might anticipate to improve the approximation by carefully choosing γ in $f_{n,K}(x)$ such that its first K moments match those of the true density $f_n(x)$. Since $f_{n,K}$ belongs to the exponential family, this method of moment approach is equivalent to the maximum likelihood method.

Following this line of thinking, in this study, we use a member of exponential family to approximate $f_n(x)$. The approximation takes the form

$$f_{n,2K}^*(x) = \exp \left(\sum_{k=1}^{2K} \gamma_k^* x^k - \Psi(\gamma^*) \right),$$

where γ^* is the maximum likelihood estimates (MLE) of the canonical exponential family. We show that this approximation attains the same order of accuracy as that of comparable Edgeworth expansion asymptotically. In addition, this approximation enjoys an information-theoretic interpretation as the maximum entropy (maxent) density. It is also a special case of the minimum relative entropy density with an uniform prior/reference density. This characterization of the minimum relative entropy density facilitates incorporation of prior information. We also show that the maxent approximation has the minimum approximation error in terms of the Kullback-Leibler distance compared to all approximations within the family (4).

Like the exponential Edgeworth approximation, the maxent approximation is strictly non-negative. Although not all moment combinations are admissible under this approach, we show that the set of admissible moment/cumulant combinations is considerably larger than that of the exponential Edgeworth method. For example, for normalized densities defined on the real line, the only restriction is that if $\mu_3 = 0$, $\mu_4 \leq 3$ when $K = 2$. If $K > 2$, there is no known restriction for the first four moments. Thus, the admissible moment space is substantially larger than that of the exponential Edgeworth approximation. Nonetheless, to tackle situations where the classical maximum entropy method fails to work, we use the dampening method in Wang (1992) to ensure the integrability of the maxent density and extend the maxent approximation to the skewness and kurtosis space satisfying $\mu_4 > 1 + \mu_3^2$.²

To our knowledge, this is the first study to apply the method of maxent density to approximating distribution of general statistics and investigate its asymptotic properties. Previous studies that examine the maxent density use this method as a density fitting tool, wherein the order of polynomial in the exponent increases to infinity at a slower rate than the sample size. This amounts to a series approximation of the logarithm of the density, and the choice of the order of polynomial is essentially equivalent to the bandwidth selection in the nonparametric density fitting. Instead, our study focuses on the approximation of the distributions known to be asymptotically normal. The coefficients of for high-order terms in the polynomials is negligible as $n \rightarrow \infty$. We show that all approximation methods considered in this study obtain the same order of asymptotic accuracy but note some differences in their small sample performances.

The rest of the paper is organized as follows: we briefly review the literature on the maximum entropy density in next section. In Section 3, we establish some asymptotic properties of the maxent approximation. We describe the dampening method to accommodate inadmissible moments in Section 4. We investigate the small sample performance of the proposed method in Section 5. Some concluding remarks are given in Section 6.

²This restriction is required to ensure the positive definiteness of the Hankel matrix.

2 Background

The estimation of a probability density function $f(x)$ by sequences of exponential families, which is equivalent to approximating the logarithm of a density, has long been studied. Earlier studies on the approximation of log-densities using polynomials include Neyman (1937) and Good (1963).³ This approximation yields generalized exponential family $f_K(x) = \exp\left(\sum_{k=1}^K \gamma_k x^k - \Psi(\gamma)\right)$. The maximum likelihood method gives efficient estimates of this canonical exponential family. Crain (1974, 1977) establishes the existence and consistency of the maximum likelihood estimator.

This method of density estimation arises naturally according to the principle of maximum entropy or minimum relative entropy. Shannon's information entropy, a measure of uncertainty or disorder, is defined as

$$W(f) = - \int f(x) \log f(x) dx. \quad (6)$$

Jaynes (1957)'s Principle of Maximum Entropy states that among all distributions that satisfy a given set of empirical constraints, one should pick the one that maximizes entropy. This resulting maxent density "is uniquely determined as the one which is maximally non-committal with regard to missing information, and that it agrees with what is known, but expresses maximum uncertainty with respect to all other matters."

The maximum entropy density is obtained by maximizing Equation (6) subject to empirical constraints

$$\int \psi_k(x) f(x) dx = \hat{\psi}_k(x), \quad k = 1, \dots, K, \quad (7)$$

where $\psi_k(x)$ is continuous function and $\hat{\psi}_k(x) = \frac{1}{n} \sum \psi_k(x_i)$. The resulting density takes the form

$$f_K^*(x) = \exp\left(\sum_{k=1}^K \gamma_k^* \psi_k(x) - \Psi(\gamma^*)\right).$$

If $\psi_k(x) = x^k$, one obtains the classical maximum entropy density $f_K^*(x) = \exp(\sum_{k=1}^K \gamma_k^* x^k - \Psi(\gamma^*))$.

Generally there is no analytical solution for this problem and numerical optimization methods are used (see for example, Zeller and Highfield (1988) and Wu (2003)). Denote H the Hessian matrix of the entropy function in the optimization problem. It is defined as $H_{ij} = \mu_{i+j} - \mu_i \mu_j$. The positive-definitiveness of the Hessian ensures the existence and

³Barron and Sheu (1991) also considers splines and trigonometric series as the basis function in the approximation of log-densities.

uniqueness of the solution.⁴

There is a duality between the maximum entropy problem and MLE for exponential families. Denote γ^* the coefficients of a maximum entropy density subject to empirical moment constraints $\frac{1}{n} \sum_{i=1}^n \psi_k(X_i)$, $k = 1, \dots, K$. The maximized entropy multiplied by the sample size n is equal to the maximized log-likelihood of the MLE. Hence, the estimated coefficients are asymptotically normal and efficient.

A closely related concept is the relative entropy, or Kullback-Leibler distance, between two densities $f(x)$ and $g(x)$,

$$D(f||g) = \int f(x) \log \frac{f(x)}{g(x)} dx. \quad (8)$$

The relative entropy measures the degree of discrepancy between two densities; $D(f||g) \geq 0$ and $D(f||g) = 0$ if and only if $f(x) = g(x)$ everywhere. The minimum relative entropy density is obtained by minimizing Equation (8) subject to empirical constraints (7). The resulting density

$$f_K^*(x; g) = g(x) \exp \left(\sum_{k=1}^K \gamma_k^* \psi(x) - \Psi(\gamma^*) \right).$$

One can verify that the maximum density is a special case of the minimum relative entropy density by setting the reference density $g(x)$ to be uniform.

Previous studies focus on using the maxent density as a smoothing tool. Barron and Sheu (1991) shows that if the logarithm of the density has r square-integrable derivatives, $\int |D^r \log f|^2 < \infty$, then the sequence of density estimator $f_{n,m}^*(x)$ converges to $f(x)$ in the sense of Kullback-Leibler distance $\int f \log(f/f_{n,m}^*)$ at rate $O_p(1/m^{2r} + m/n)$ if $m \rightarrow \infty$ and $m^3/n \rightarrow 0$ as $n \rightarrow \infty$, where m is the order of polynomial and n is the sample size. If there is no estimation errors due to sampling variation of the empirical moments or inefficient estimators, the convergence rate will be $O(1/m^{2r})$. Practically, the choice of m amounts to the bandwidth selection in nonparametric density estimation.

In this study, we use maxent densities to approximate the density of mean of random variables or more general statistics. This study is conceptually different from previous studies in that the distributions to approximate are known to be asymptotically normal under certain regularity conditions. Hence, the problem of interest is the ‘‘small sample asymptotics’’, or, approximation of small sample distributions, where the normal approximation is often inadequate. Instead of letting m goes to infinity with n , we approximate the underlying

⁴Let $\beta' = [\beta_0, \beta_1, \dots, \beta_K]$ be a non-zero vector and $\phi_0(x) = 1$, we have

$$\begin{aligned} \beta' H \beta &= \sum_{k=0}^K \sum_{j=0}^K \beta_k \beta_j \int \phi_k(x) \phi_j(x) f(x, \gamma) dx \\ &= \int \left(\sum_{k=0}^K \gamma_k \phi_k(x) \right)^2 f(x, \gamma) dx > 0. \end{aligned}$$

Hence, H is positive-definite.

distribution holding m fixed and study the performance of the approximation up to certain order, say, n^{-1} , anticipating higher order terms be negligible for large n .

We note that both exponential Edgeworth and maxent approximations, sharing a common functional form, have a limited range of admissible moments for density defined on the real line.⁵ For example, when $K = 4$, they share a common lower bound $\mu_4 \geq \mu_3^2 + 1$. On the other hand, the exponential Edgeworth will explode for large $|x|$ if $\mu_4 - 3 \geq 3\mu_3^2$.⁶ The maxent density is substantially less restrictive in terms of admissible moments. According to Tagniali (2003), when $\mu_3 \neq 0$, there is no upper bound for μ_4 . The only restriction is $\mu_4 \leq 3$ if $\mu_3 = 0$. Figure 1 shows the admissible moments in the space of μ_3 and μ_4 for $|\mu_3| \leq 3$. The area between the two solid lines is the admissible moment space for the exponential Edgeworth approximation, whereas the area above the lower solid line, excluding the vertical dashed line, is the admissible moment space for the maxent approximation. One can see that the admissible moment space for the maxent approximation encompasses that of exponential Edgeworth and is substantially larger. For the maxent approximation, there is no known upper bounds for μ_4 when $K \geq 3$.

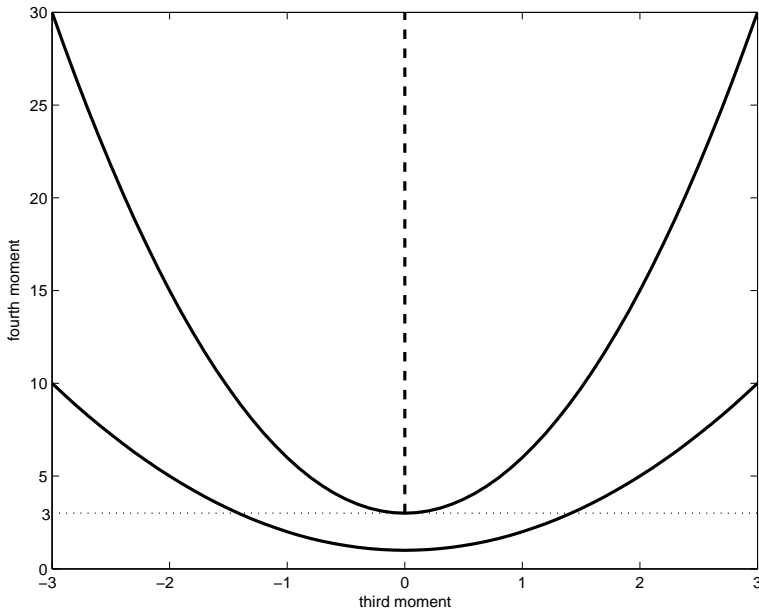


Figure 1: Admissible moments for maxent density

Previous studies using the maxent densities as a density estimator typically assume a bounded support to ensure the integrability of the estimated density. However, this assump-

⁵The classical Edgeworth approximation often exhibits negative densities when the moments/cumulants deviate substantially from those of the normal. Eriksson et al. (2004) calculates the skewness and kurtosis space wherein the Edgeworth expansion will yield non-negative densities.

⁶To ensure integrability, the coefficient for the dominant term x^4 in the exponent of Equation (3) has to be negative, implying $\kappa_4 - 3\kappa_3^2 \leq 0$. The corresponding moment condition can be obtained easily by setting $\mu_4 = \kappa_4/n + 3$ and $\mu_3 = \kappa_3/\sqrt{n}$.

tion is not innocuous for the approximation of the density of general statistics, where the underlying density is unbounded and often the interest is on the tail of the density. Two conditions are needed to ensure the integrability of a maxent density defined for the real line: (i) the order K of the polynomial in the exponent is even; (ii) the coefficient of the highest order term, γ_K , is negative. Condition (i) prescribes that the maxent approximation expands in the order of n^{-1} for standardized statistics, different from the $n^{-1/2}$ of the Edgeworth expansion. Under condition (i) and (ii), the maxent densities for the sample statistics, which is asymptotically normal, vanish to 0 as $|x| \rightarrow \infty$.

For $K = 4$, if $\mu_3 = 0$ and $\mu_4 > 3$, one can show that γ_K is positive, which renders the approximation not integrable on the real line. This constraint severely restricts the applicability of the maxent method to approximate fat-tailed distributions. We employ the dampening method in Wang (1992) to ensure the integrability of the maxent density. The details is given in Section 4.

3 Asymptotic properties

In this section, we investigate the asymptotic properties of maxent approximation. We compare the maxent approximation to the Edgeworth and exponential Edgeworth approximation. The exponential Edgeworth approximation shares the same exponential functional form with the maxent approximation; on the other hand, the Edgeworth approximation can be derived directly from the exponential Edgeworth. Hence, the exponential Edgeworth approximation is not only of interest in its own right, but also bridges the comparison between the maxent approximation and the more commonly used Edgeworth approximation in this study.

For the ease of exposition and comparison with the Edgeworth expansion, we present our theorems for the density of standardized sample means. The extension to general statistics can be modeled in a straightforward manner. Consider an iid sample Y_1, \dots, Y_n from a continuous distribution with mean zero, variance σ^2 and $E(|Y_1|^{2K+1}) < \infty$. We are interested in $f_n(x)$, the density of $X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i/\sigma$ with mean zero and unit variance. It is also assumed that the Edgeworth, exponential Edgeworth and maxent approximations to $f_n(x)$ are defined and integrable over the real line.

The following lemmas are useful for the main results of this section.

Lemma 1. *Let $g_{n,2K}(x)$ be the Edgeworth approximation to $f_n(x)$. For $n > 2K$, X_n has a continuous bounded density function $f_n(x)$ such that*

$$\sup \{|f_n(x) - g_{n,2K}(x)| : x \in \mathbb{R}\} = o(n^{1-K})$$

as $n \rightarrow \infty$.

Proof. All proofs are given in Appendix A. ■

Since the Edgeworth approximation can be obtained directly from the exponential Edgeworth by expanding its exponent using the Taylor expansion, it follows directly that the

exponential Edgeworth has the same order of asymptotic accuracy as that of the Edgeworth expansion pointwisely. Below we prove that this relationship holds true uniformly, which leads to the following result.

Lemma 2. *Let $h_{n,2K}(x)$ be the exponential Edgeworth approximation to $f_n(x)$. For $n > 2K$, X_n has a continuous bounded density function $f_n(x)$ such that*

$$\sup \{|f_n(x) - h_{n,2K}(x)| : x \in \mathbb{R}\} = o(n^{1-K})$$

as $n \rightarrow \infty$.

Next we show that maxent approximation also attains the same order of asymptotic accuracy.

Theorem 1. *Let $f_{n,2K}^*(x)$ be the maxent approximation to $f_n(x)$. For $n > 2K$, X_n has a continuous bounded density function $f_n(x)$ such that*

$$\sup \{|f_n(x) - f_{n,2K}^*(x)| : x \in \mathbb{R}\} = o(n^{1-K})$$

as $n \rightarrow \infty$

Denote $F_n(x) = \int_{-\infty}^x f_n(v) dv$ and $F_{n,2K}^*(x) = \int_{-\infty}^x f_{n,2K}^*(v) dv$, we establish the following result for the approximation of distribution function.

Theorem 2. *X_n has a continuous cumulative distribution function such that*

$$\sup \{|F_n(x) - F_{n,2K}^*(x)| : x \in \mathbb{R}\} = o(n^{1-K})$$

as $n \rightarrow \infty$.

Next we establish some properties of $f_{n,2K}^*(x)$ based on its characterization as the maximum entropy density. The convergence results are examined in terms of the Kullback-Leibler distance.

Theorem 3. *For all approximations within the exponential family*

$$f_{n,2K}(x) = \exp \left(\sum_{k=1}^{2K} \gamma_k x^k - \Psi(\gamma) \right),$$

the maxent approximation $f_{n,2K}^*(x)$ attains the minimum Kullback-Leibler distance; that is, the Kullback-Leibler distance

$$D(f_n || f_{n,2K}^*) \leq D(f_n || f_{n,2K}).$$

The equality holds if and only if $f_{n,2K}^* = f_{n,2K}$.

This theorem suggests that $f_{n,2K}$ generally approximates f_n better than the exponential Edgeworth in terms of the Kullback-Leibler distance, with the exception that the approximation $h_{n,2K}(x)$ is exact, wherein $f_{n,2K}^*(x) = h_{n,2K}(x)$.

The Kullback-Leibler distance $D(f_n || f_{n,2K}^*)$ also provides some bounds on the error of distribution approximation. Define the total variation

$$V(f, g) = \int |f(x) - g(x)| dx.$$

It is known that $D(f || g) \geq V(f, g)^2 / 2$. Kullback (1967) shows the bound can be improved to

$$D(f || g) \geq \frac{V(f, g)^2}{2} + \frac{V(f, g)^4}{36}.$$

The following result is immediately available from the above inequality.

Theorem 4. *The error of cumulative distribution function approximation by the maxent density satisfies*

$$|F_n(x) - F_{n,2K}^*(x)| \leq 3 \left[-1 + \left\{ 1 + \frac{4}{9} D(f_n || f_{n,2K}^*) \right\}^{1/2} \right]^{1/2}.$$

Note that the inequality can be further improved by the inequality

$$D(f || g) \geq \frac{V(f, g)^2}{2} + \frac{V(f, g)^4}{36} + \frac{V(f, g)^6}{288},$$

which leads to a rather cumbersome expression and is therefore not pursued here.

4 Dampening adjustment

Although the maxent approximation is considerably more flexible in terms of admissible moments compared with the exponential Edgeworth approximation, it does not accommodate all moment conditions. When $K = 2$, if $\mu_3 = 0$, $\mu_4 \leq 3$ is needed to ensure the integrability of the maxent density on the real line. One possible method to relax this restriction is to increase the order of the polynomials in the exponent. However, increasing the order of the polynomial is equivalent to increasing the order of moment conditions for the maxent density. It is well known that higher order moments are rather sensitive to outliers. In practice, moments higher than μ_4 are rarely used. In addition to the robustness consideration, Delén (1987) shows that sample moments are bounded by sample size, which makes higher order moments unsuitable for small sample asymptotics studies.

Instead, we propose to use a dampening method to deal with this problem. Rewrite the maxent density

$$f_{n,2K}^*(x) = c(\gamma^*) e^{-x^2/2} \exp\left(\sum_{k=1}^{2K} \gamma_k^* x^k + \frac{x^2}{2}\right),$$

where $c(\gamma^*)$ is the normalization term. In the spirit of Wang (1992)'s modification of the general saddlepoint approximation, we introduce a dampening term for the second exponent. We define the alternative density

$$\tilde{f}_{n,2K}(x; b) = \tilde{c}(\gamma^*) e^{-x^2/2} \exp \left\{ \left(\sum_{k=1}^{2K} \gamma_k^* x^k + \frac{x^2}{2} \right) w_b(x) \right\},$$

where $w_b(x) = \exp(-bx^2/2)$ and $b > 0$ is a properly chosen constant. Note that $w_b(0) = 1$ and $w_b(x)$ goes rapidly to 0 as $|x| \rightarrow \infty$. Therefore, the regulation serves to properly control the influence of higher-order terms in the exponent as $|x|$ increases.

Asymptotically, b can be any fixed constant. In practice, we define b as follows:

$$b = \max \left(0, \inf \left\{ a \mid a \geq 0 \text{ and } x \frac{\partial \tilde{f}_{n,2K}(x; a)}{\partial x} \leq 0 \text{ for } f(x) < \varepsilon \right\} \right),$$

where ε is a small positive number, say, 10^{-6} . Under certain regularity conditions, $f_n(x) \rightarrow N(0, 1)$ as $n \rightarrow \infty$, implying that $f(x) < \varepsilon$ as $|x| \rightarrow \infty$. The condition $x \frac{\partial \tilde{f}_{n,2K}(x; a)}{\partial x} \leq 0$ for $f(x) < \varepsilon$ prescribes that the density vanishes at either end as $|x| \rightarrow \infty$. The constant b can be obtained numerically. When $\gamma_{2K}^* < 0$, the density $f_{n,2K}^*$ vanishes at either end by itself. Correspondingly, $b = 0$ and no dampening is imposed.

5 Numerical Examples

In this section, we provide several numerical examples to illustrate the small sample performance of the proposed maxent density. For comparison, we also report the results for the Edgeworth expansion and exponential Edgeworth expansion.

Our first example reproduces an experiment reported in Field and Hampel (1982). In this experiment, we approximate the density of mean of four independent exponentially distributed random variables with mean one. The results are reported in Table 1, in the same format as Field and Hampel (1982). We note that the Edgeworth expansion produces negative density for $x = -0.25$. In contrast, densities from both exponential Edgeworth and maxent approximation are always non-negative. As expected, the Edgeworth and exponential Edgeworth produce accurate density estimates when x is close to 1, the mean of the distribution. On the other hand, the maxent approximation offers more accurate approximation at the tails, especially at the lower end of the distribution.

In the second experiment, suppose Y_1, \dots, Y_n is an iid sample from the standard normal distribution. We are interested the density of $X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i^2 - n) / \sqrt{2n}$. Figure 2 reports the absolute errors of approximations to the theoretical density with $n = 10$ for either tail. For the left tail, the figure reports the errors for $x \leq -\sqrt{n}/2$, where the theoretical density is zero. One can see that the Edgeworth expansion, represented by the dash-dotted line, exhibits the typical oscillation of polynomials and has a portion of negative densities. In contrast, the performance of maxent and exponential Edgeworth approximation is considerably better. The errors of the maxent approximation are smaller than that of the exponential

Table 1: Approximation for the density of mean of four standard exponential random variables

| x | Exact | Edg | % err. | E.Edg | % err. | Maxent | % err. |
|-------|--------|---------|--------|--------|--------|--------|--------|
| -0.25 | 0 | -0.0392 | | 0.0017 | | 0.0009 | |
| 0 | 0 | 0.0217 | | 0.0460 | | 0.0266 | |
| 0.1 | 0.0286 | 0.1040 | 2.64 | 0.1119 | 2.91 | 0.0707 | 1.47 |
| 0.25 | 0.2453 | 0.2977 | 21.37 | 0.2954 | 20.46 | 0.2161 | -11.88 |
| 0.5 | 0.7218 | 0.6923 | -4.08 | 0.7041 | -2.45 | 0.6464 | -10.45 |
| 0.75 | 0.8962 | 0.8886 | -0.85 | 0.8913 | -0.55 | 0.9419 | 5.10 |
| 1 | 0.7815 | 0.7813 | -0.03 | 0.7814 | -0.01 | 0.8464 | 8.31 |
| 1.25 | 0.5615 | 0.5658 | 0.77 | 0.5636 | 0.37 | 0.5682 | 1.20 |
| 1.5 | 0.3569 | 0.3697 | 3.57 | 0.3615 | 1.28 | 0.3304 | -7.45 |
| 1.75 | 0.2085 | 0.2005 | -3.84 | 0.2031 | -2.62 | 0.1846 | -11.49 |
| 2 | 0.1145 | 0.0937 | -18.14 | 0.0895 | -21.82 | 0.1052 | -8.15 |
| 2.25 | 0.0600 | 0.0557 | -0.07 | 0.0253 | -0.59 | 0.0621 | 0.04 |

Edgeworth. For the right tail, all three approximations produce similar results and converge to the true density rapidly (note that the scale of error is smaller in the right panel).

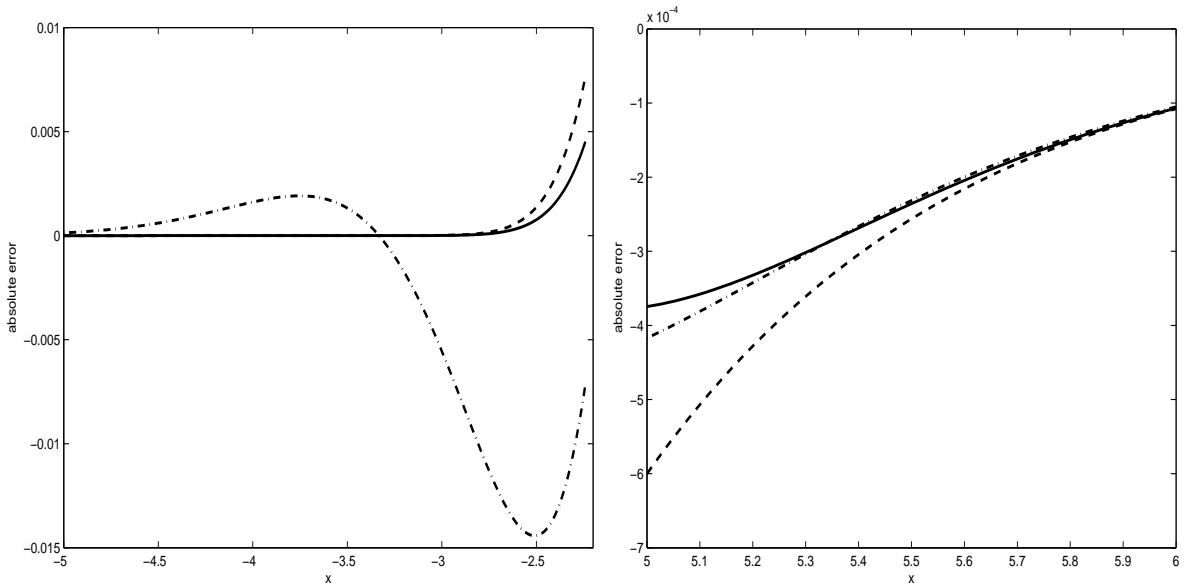


Figure 2: Left: absolute error of the left tail; Right: absolute error of the right tail. (solid: maxent; dash-dot: Edgeworth; dash: exponential Edgeworth)

We also investigate the approximation to the tail of the distribution function, which is crucial to statistic inference, especially for small sample problem where the normal approximation may be inadequate. Since all three approximations provide comparable performance

at the right tail, we focus on the left tail. Table 2 reports the predicted CDF and its relative errors. The results suggest that all three approximations improve with sample size. The Edgeworth expansion produces negative densities for sample size as large as 30. The exponential Edgeworth and maxent approximation perform considerably better. Between these two, the maxent approximation is always more accurate.

Table 2: Approximation for the lower tail of the distribution of standardized mean of n squared standard normal random variables (R.E.: relative errors)

| F_n | Edg | R.E. | E.Edg | R.E. | Maxent | R.E. |
|----------|---------|--------|--------|-------|--------|-------|
| $n = 10$ | | | | | | |
| 0.0500 | 0.0553 | 0.10 | 0.0681 | 0.36 | 0.0513 | 0.02 |
| 0.0100 | 0.0105 | 0.05 | 0.0234 | 1.34 | 0.0159 | 0.59 |
| 0.0050 | 0.0032 | -0.35 | 0.0161 | 2.22 | 0.0106 | 1.12 |
| 0.0010 | -0.0044 | -5.37 | 0.0080 | 6.96 | 0.0050 | 3.99 |
| 0.0005 | -0.0058 | -12.57 | 0.0062 | 11.48 | 0.0039 | 6.72 |
| 0.0001 | -0.0074 | -74.56 | 0.0040 | 38.62 | 0.0024 | 22.92 |
| $n = 20$ | | | | | | |
| 0.0500 | 0.0515 | 0.03 | 0.0572 | 0.14 | 0.0495 | -0.01 |
| 0.0100 | 0.0093 | -0.07 | 0.0148 | 0.48 | 0.0119 | 0.18 |
| 0.0050 | 0.0034 | -0.31 | 0.0087 | 0.74 | 0.0068 | 0.36 |
| 0.0010 | -0.0015 | -2.45 | 0.0029 | 1.86 | 0.0021 | 1.12 |
| 0.0005 | -0.0020 | -4.96 | 0.0019 | 2.71 | 0.0014 | 1.71 |
| 0.0001 | -0.0021 | -21.65 | 0.0008 | 6.52 | 0.0005 | 4.33 |
| $n = 30$ | | | | | | |
| 0.0500 | 0.0508 | 0.01 | 0.0542 | 0.08 | 0.0494 | -0.01 |
| 0.0100 | 0.0094 | -0.05 | 0.0127 | 0.27 | 0.0109 | 0.09 |
| 0.0050 | 0.0040 | -0.19 | 0.0071 | 0.41 | 0.0059 | 0.18 |
| 0.0010 | -0.0003 | -1.25 | 0.0020 | 0.95 | 0.0016 | 0.57 |
| 0.0005 | -0.0007 | -2.32 | 0.0012 | 1.33 | 0.0009 | 0.84 |
| 0.0001 | -0.0007 | -7.90 | 0.0004 | 2.79 | 0.0003 | 1.91 |

In our last experiment, we approximate the distribution of mean of Student's t random variables. The exponential Edgeworth expansion is not integrable on the real line for this distribution as $\kappa_4 - 3\kappa_3^2 > 0$ and is therefore not reported. Since $\mu_3 = 0$ and $\mu_4 > 3$, the maxent is not integrable on the real line as well. Correspondingly, we apply the dampening method described in the previous section. The results for distribution approximation are reported in Table 3. In this example, we set both the sample size and the degrees of freedom of the t distribution to be 10. Since the distribution is symmetric, all odd terms in the Edgeworth expansion is zero. Hence, the Edgeworth expansion is competitive in this case. Nonetheless, one can see that the performances of the Edgeworth and maxent approximation are quite comparable. Since the distribution converges to the standard normal rapidly as n increases, we do not report results for larger sample size.

Table 3: Approximation for the lower tail of the distribution of standardized mean of 10 random variables from Student’s t distribution with 10 degrees of freedom (R.E.: relative errors)

| F_n | Edg | R.E. | Maxent | R.E. |
|--------|--------|-------|--------|-------|
| 0.0500 | 0.0491 | -0.02 | 0.0491 | -0.02 |
| 0.0100 | 0.0096 | -0.04 | 0.0095 | -0.05 |
| 0.0050 | 0.0046 | -0.09 | 0.0045 | -0.10 |
| 0.0010 | 0.0008 | -0.21 | 0.0008 | -0.23 |
| 0.0005 | 0.0003 | -0.33 | 0.0003 | -0.36 |
| 0.0001 | 0.0001 | -0.45 | 0.0001 | -0.48 |

6 Conclusion

In this paper, we propose a simple approach for approximating small sample distributions based on the method of Maximum Entropy (maxent) density. Unlike previous studies that use the maxent method as a curve-fitting tool, we focus on the “small sample asymptotics” of the approximation. We show that this method obtains the same asymptotic order of accuracy as that of the classical Edgeworth expansion and the exponential Edgeworth expansion by Field and Hampel (1982). In addition, the maxent approximation attains the minimum Kullback-Leibler distance among all approximations of the same exponential family. For symmetric fat-tailed distributions, we adopt the dampening approach in Wang (1992) to ensure the integrability of maxent densities and effectively extend the maxent method to accommodate all admissible skewness and kurtosis space. Our numerical experiments show that the proposed method compares competitively with and often outperforms the Edgeworth and exponential Edgeworth expansions, especially when the sample size is small.

Appendix A

Proof of Lemma 1. See Chapter XVI of Feller (1971). ■

Proof of Lemma 2. For expository ease, we first look at the mostly commonly used case with $K = 2$. Rewrite

$$\begin{aligned} h_{n,4}(x) &= \phi(x) \exp\left(\frac{3\kappa_4 - 5\kappa_3^2}{24n}\right) \exp\left(-\frac{\kappa_3 x}{2\sqrt{n}} + \frac{2\kappa_3^2 - \kappa_4}{4n}x^2 + \frac{\kappa_3}{6\sqrt{n}}x^3 + \frac{\kappa_4 - 3\kappa_3^2}{24n}x^4\right) \\ &= \phi(x) \exp\left(\frac{3\kappa_4 - 5\kappa_3^2}{24n}\right) \exp(M(x)). \end{aligned} \quad (9)$$

We have

$$\begin{aligned} M(x)^2 &= \frac{\kappa_3^2}{4n}x^2 - \frac{\kappa_3^2}{6n}x^4 + \frac{\kappa_3^2}{36n}x^6 \\ &\quad + \frac{\kappa_3\kappa_4}{4n^{\frac{3}{2}}}x^3 - \frac{\kappa_3\kappa_4}{8n^{\frac{3}{2}}}x^5 + \frac{\kappa_3\kappa_4}{72n^{\frac{3}{2}}}x^7 - \frac{\kappa_3^3}{2n^{\frac{3}{2}}}x^3 + \frac{7\kappa_3^3}{24n^{\frac{3}{2}}}x^5 - \frac{\kappa_3^3}{24n^{\frac{3}{2}}}x^7 \\ &\quad + \frac{\kappa_4^2}{16n^2}x^4 + \frac{\kappa_3^2}{4n^2}x^4 - \frac{\kappa_4^2}{48n^2}x^6 - \frac{\kappa_3^2}{8n^2}x^6 + \frac{\kappa_4^2}{576n^2}x^8 \\ &\quad + \frac{\kappa_3^4}{64n^2}x^8 - \frac{\kappa_3^2\kappa_4}{4n^2}x^4 + \frac{5\kappa_3^2\kappa_4}{48n^2}x^6 - \frac{\kappa_3^2\kappa_4}{96n^2}x^8 \\ &= \frac{\kappa_3^2}{4n}x^2 - \frac{\kappa_3^2}{6n}x^4 + \frac{\kappa_3^2}{36n}x^6 + \frac{P_1(x)}{n^{3/2}} + \frac{P_2(x)}{n^2}, \end{aligned} \quad (10)$$

where $P_1(x)$ and $P_2(x)$ are polynomials with constant coefficients. Similarly, one can show that

$$M(x)^3 = \sum_{j=3}^6 \frac{P_j(x)}{n^{j/2}}, \quad (11)$$

where $P_j(x)$ is a polynomial with constant coefficients for $j = 3, \dots, 6$. Without loss of generality, assume that $x > 0$. Applying the Mean Value Theorem to Equation (9) yields

$$h_{n,4} = \phi(x) \left[1 + \frac{\kappa_4}{8n} - \frac{5\kappa_3^2}{24n} + o(n^{-1})\right] \left[1 + M(x) + \frac{M(x)^2}{2} + \frac{\exp(M(\xi)) M(x)^3}{6}\right],$$

for some $\xi \in (0, \infty)$. Using results from (10) and (11), we have

$$\begin{aligned} h_{n,4} &= \phi(x) \left[1 + \frac{\kappa_4}{8n} - \frac{5\kappa_3^2}{24n} + o(n^{-1})\right] \\ &\quad \left[1 + M(x) + \frac{\kappa_3^2}{4n}x^2 - \frac{\kappa_3^2}{6n}x^4 + \frac{\kappa_3^2}{36n}x^6 + \frac{P_1(x)}{2n^{3/2}} + \frac{P_2(x)}{2n^2} + \frac{\exp(M(\xi))}{6} \sum_{j=3}^6 \frac{P_j(x)}{n^{j/2}}\right]. \end{aligned}$$

Collecting terms by order of $n^{-1/2}$ yields

$$\begin{aligned}
h_{n,4}(x) &= \phi(x) \left\{ 1 + \frac{\kappa_3 H_3(x)}{3! \sqrt{n}} + \frac{\kappa_4 H_4(x)}{4! n} + \frac{\kappa_3^2 H_6(x)}{6! n} \right. \\
&\quad \left. + \frac{P_1(x)}{2n^{3/2}} + \frac{P_2(x)}{2n^2} + \frac{\exp(M(\xi))}{6} \sum_{j=3}^6 \frac{P_j(x)}{n^{j/2}} + o(n^{-1}) \right\} \\
&= g_{n,4}(x) + \phi(x) \left[\frac{P_1(x)}{2n^{3/2}} + \frac{P_2(x)}{2n^2} \right] + \phi(x) \frac{\exp(M(\xi))}{6} \sum_{j=3}^6 \frac{P_j(x)}{n^{j/2}} + o(n^{-1}) \quad (12)
\end{aligned}$$

Since for arbitrary polynomial $P(x)$, $\phi(x)P(x) = O(1)$ for all x , the second term in (12) has order $o(n^{-1})$. As for the third term, since

$$\exp(M(\xi)) = \frac{h_{n,4}(\xi)}{\phi(\xi) \exp\left(\frac{3\kappa_4 - 5\kappa_3^2}{24n}\right)} < \infty,$$

it also has order $o(n^{-1})$. Therefore, we have

$$\sup \{|h_{n,4}(x) - g_{n,4}(x)|, x \in \mathbb{R}\} = o(n^{-1}).$$

Using Lemma 1, we then have

$$\sup \{|f_n(x) - h_{n,4}(x)|, x \in \mathbb{R}\} = o(n^{-1}).$$

Generalization of the proof to higher-order cases, although notationally complex, is straightforward. \blacksquare

Proof of Theorem 1. Let $\mu_k = \int x^k f_n(x) dx$ and $\mu = [\mu_1, \dots, \mu_{2K}]$. We have

$$\begin{aligned}
\mu_{h,k} &= \int x^k h_{n,2K}(x) dx \\
&= \int x^k f_n(1 + o(n^{1-K})) dx \\
&= \mu_k (1 + o(n^{1-K})).
\end{aligned}$$

Let $\gamma = [\gamma_1, \dots, \gamma_{2K}]$. Denote the continuous mapping $\gamma = M(\mu)$, and H the gradient of the mapping, with $H_{ij} = \mu_{i+j} - \mu_i \mu_j$, $1 \leq i, j \leq 2K$. Since H is positive definite and $\mu_f = \mu$, we have

$$\begin{aligned}
\gamma^* &= \gamma' + H^{-1}(\mu_h - \mu_f) + o(n^{1-K})^2 \Rightarrow \\
\gamma^* &= \gamma' + H^{-1}(\mu_h - \mu) + o(n^{1-K})^2 \Rightarrow \\
\gamma^* &= \gamma' [1 + o(n^{1-K})].
\end{aligned}$$

We then have for all x ,

$$\begin{aligned}
f_{n,2k}^*(x) &= \exp\left(\sum_{k=1}^{2K} \gamma_k^* x^k - \Psi(\gamma^*)\right) \\
&= \exp\left(\sum_{k=1}^{2K} \gamma'_k [1 + o(n^{1-K})] x^k - \Psi(\gamma' [1 + o(n^{1-K})])\right) \\
&= \exp\left(\sum_{k=1}^{2K} \gamma'_k x^k + o(n^{1-K}) \sum_{k=1}^{2K} \gamma'_k x^k - \Psi(\gamma') + o(n^{1-K})\right) \\
&= \exp\left(\sum_{k=1}^{2K} \gamma'_k x^k - \Psi(\gamma')\right) \exp\left(o(n^{1-K}) \sum_{k=1}^{2K} \gamma'_k x^k + o(n^{1-K})\right) \\
&= h_{n,2K}(x) \exp\left(o(n^{1-K}) \left[\sum_{k=1}^{2K} \gamma'_k x^k - \Psi(\gamma')\right] + o(n^{1-K})\right) \\
&= h_{n,2K}(x) [1 + o(n^{1-K})] \exp\left(o(n^{1-K}) \left[\sum_{k=1}^{2K} \gamma'_k x^k - \Psi(\gamma')\right]\right) \\
&= h_{n,2K}(x) [1 + o(n^{1-K})] [h_{n,2K}(x)]^{o(n^{1-K})} \\
&= h_{n,2K}(x) [1 + o(n^{1-K})].
\end{aligned}$$

The last inequality holds since $h_{n,2K}(x)$ is bounded. Using the boundedness of $h_{n,2K}(x)$ again, we have

$$\sup \{|h_{n,2K}(x) - f_{n,2K}^*(x)| : x \in \mathbb{R}\} = o(n^{1-K}).$$

Using Lemma 1 and Lemma 2, we then have

$$\sup \{|f_n(x) - f_{n,2K}^*(x)| : x \in \mathbb{R}\} = o(n^{1-K}). \quad \blacksquare$$

Proof of Theorem 2. Using Theorem 1, for all x , we have

$$\begin{aligned}
&|F_n(x) - F_{n,2K}^*(x)| \\
&= \left| \int_{-\infty}^x [f_n(v) - f_{n,2K}^*(v)] dv \right| \\
&= \left| \int_{-\infty}^x f_n(v) - f_n(v) (1 + o(n^{1-K})) dv \right| \\
&= \left| \int_{-\infty}^x f_n(v) o(n^{1-K}) dv \right| \\
&= o(n^{1-K}) \int_{-\infty}^x f_n(v) dv \\
&= o(n^{1-K}). \quad \blacksquare
\end{aligned}$$

Proof of Theorem 3. Since the densities are positive, we have

$$\log \frac{f_n(x)}{f_{n,2K}(x)} = \log \frac{f_n(x)}{f_{n,2K}^*(x)} + \log \frac{f_{n,2K}^*(x)}{f_{n,2K}(x)}.$$

Taking expectation on both sides of yields

$$\begin{aligned} \int f_n(x) \log \frac{f_n(x)}{f_{n,2K}(x)} dx &= \int f_n(x) \log \frac{f_n(x)}{f_{n,2K}^*(x)} dx + \int f_n(x) \log \frac{f_{n,2K}^*(x)}{f_{n,2K}(x)} dx \Rightarrow \\ D(f_n || f_{n,2K}) &= D(f_n || f_{n,2K}^*) + \int f_n(x) \log \frac{f_{n,2K}^*(x)}{f_{n,2K}(x)} dx. \end{aligned}$$

Note that

$$\begin{aligned} &\int f_n(x) \log \frac{f_{n,2K}^*(x)}{f_{n,2K}(x)} dx \\ &= \int f_n(x) \log f_{n,2K}^*(x) dx - \int f_n(x) \log f_{n,2K}(x) dx \\ &= \int f_n(x) \left[\sum_{k=1}^{2K} \gamma_k^* x^k - \Psi(\gamma^*) \right] dx - \int f_n(x) \left[\sum_{k=1}^{2K} \gamma_k x^k - \Psi(\gamma) \right] dx \\ &= \int f_{n,2K}^*(x) \left[\sum_{k=1}^{2K} \gamma_k^* x^k - \Psi(\gamma^*) \right] dx - \int f_{n,2K}^*(x) \left[\sum_{k=1}^{2K} \gamma_k x^k - \Psi(\gamma) \right] dx \\ &= \int f_{n,2K}^*(x) \log f_{n,2K}^*(x) dx - \int f_{n,2K}^*(x) \log f_{n,2K}(x) dx \\ &= D(f_{n,2K}^* || f_{n,2K}). \end{aligned}$$

The third equality holds because $f_{n,2K}^*(x)$ shares the same first $2K$ moments as those of $f_n(x)$. We then have the Pythagorean-like identity

$$D(f_n || f_{n,2K}) = D(f_n || f_{n,2K}^*) + D(f_{n,2K}^* || f_{n,2K}).$$

Since $D(f_{n,2K}^* || f_{n,2K}) \geq 0$, it follows

$$D(f_n || f_{n,2K}^*) \leq D(f_n || f_{n,2K}),$$

where the equality holds if and only if $f_{n,2K}^*(x) = f_{n,2K}(x)$ everywhere. ■

Proof of Theorem 4. See Tagniali (2003). ■

Appendix B

Let Y_1, \dots, Y_n be an iid sample from a distribution $F(y)$ with mean μ and variance σ^2 and $X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \mu) / \sigma$. The moments and density function $f_n(x)$ for X_n is given below for the three distributions considered in this study.

- $F(y)$ is exponential with mean one. Then $\sum_{i=1}^n Y_i$ has a gamma distribution with shape parameter n and scale parameter one; X_n has the following properties:

- Skewness: $2/\sqrt{n}$
- Kurtosis: $6/n$
- Density function: for $x \geq -\sqrt{n}$

$$f_n(x) = f_n(x) = \frac{(n + \sqrt{n}x)^{n-1} e^{-n - \sqrt{n}x} \sqrt{n}}{\Gamma(n)},$$

$f_n(x) = 0$ otherwise.

- $F(y)$ is normal with mean zero. Then $\sum_{i=1}^n Y_i^2$ has a χ^2 distribution with n degrees of freedom; X_n has the following properties:

- Skewness: $2\sqrt{2}/\sqrt{n}$
- Kurtosis: $12/n$
- Density: for $x \geq -\sqrt{n/2}$

$$f_n(x) = \frac{\sqrt{2n} (\sqrt{2n}x + n)^{(n-2)/2} e^{-(\sqrt{2n}x + n)/2}}{\Gamma(\frac{n}{2}) 2^{n/2}},$$

$f_n(x) = 0$ otherwise.

- $F(y)$ is Student's t distribution with v degrees of freedom. Then $X_n(x)$ has the following properties:

- Skewness: 0
- Kurtosis: $6/(v - 4)/n$
- Density function: no closed form for $f_n(x)$; obtained by Monte Carlo simulations.

7 References

- Barron, A.R., and C. Sheu. "Approximation of Density Functions by Sequences of Exponential Families." *Annals of Statistics* 19 (1991): 1347-1369.
- Crain, B.R. "Estimation of Distributions using Orthogonal Expansions," *Annals of Statistics*, 16 (1974): 454-463.
- Crain, B.R. "An Information Theoretic Approach to Approximation a Probability Distribution," *SIAM Journal of Applied Mathematics*, 32 (1977): 339-346.
- Dalén, J. "Algebraic Bounds on Standardized Sample Moments." *Statistics and Probability Letters* 5 (1987): 329-331.
- Edgeworth, F.Y. "The Law of Error." *Transactions of the Cambridge Philosophical Society* 20 (1905): 36-65.
- Eriksson, A., L. Forsberg, and E. Ghysels. "Approximating the Probability Distribution of Functions of Random Variables: A New Approach." Working paper, 2004.
- Feller, W. *An Introduction to Probability Theory and its Application, II*. New York: Wiley, 1971.
- Field, C.A. and F.R. Hampel. "Small-sample Asymptotic Distributions of M-estimators of Location." *Biometrika* 69 (1982): 29-46.
- Field, C.A., and E. Ronchetti. *Small Sample Asymptotics*. Hayward: Institute of Mathematical Statistics, 1990.
- Good, I.J. "Maximum Entropy for Hypothesis Formulation, especially for Multidimensional Contingency Tables." *Annals of Mathematical Statistics*, 34 (1963): 911-934.
- Hampel, F.R. "Some Small-sample Asymptotics." in *Proceedings of the Prague Symposium on Asymptotic Statistics*, ed. J. Hajek, Prague: Charles University, 1973.
- Jaynes, E.T. "Information Theory and Statistical Mechanics." *Physics Review* 106 (1957): 620-30.
- Kullback, S. "A Lower Bound for Discrimination Information in Terms of Variation." *IEEE Transactions on Information Theory*, 13 (1967): 126-127.
- Neyman, J. "Smooth Test for Goodness of Fit." *Scandinavian Aktuarial*, 20 (1937): 149-199.
- Tagliani, A. "A Note on Proximity of Distributions in Terms of Coinciding Moments." *Applied Mathematics and Computation* 145 (2003): 195-203.
- Wang, S. "General Saddlepoint Approximations in the Bootstrap." *Statistics and Probability Letters*, 13 (1992): 61-66.
- Wu, X. "Calculation of Maximum Entropy Densities with Application to Income Distribution." *Journal of Econometrics* 115 (2003): 347-354.
- Zellner, A., and R.A. Highfield. "Calculation of Maximum Entropy Distribution and Approximation of Marginal Posterior Distributions." *Journal of Econometrics* 37 (1988): 195-209.