Information Theoretic Asymptotic Approximations for Distributions of Statistics

<table>
<thead>
<tr>
<th>Ximing Wu</th>
<th>Suojin Wang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Agricultural Economics</td>
<td>Department of Statistics</td>
</tr>
<tr>
<td>Texas A&amp;M University</td>
<td>Texas A&amp;M University</td>
</tr>
</tbody>
</table>

Running headline: Information Theoretic Asymptotic Approximations
Abstract

We propose an information theoretic approach to approximating asymptotic distributions of statistics using the maximum entropy densities. Conventional maximum entropy densities are typically defined on a bounded support. For distributions defined on unbounded supports, we propose to use an asymptotically negligible dampening function for the maximum entropy approximation such that it is well defined on the real line. We establish order $n^{-1}$ asymptotic equivalence between the proposed method and the classical Edgeworth expansion for general statistics that are smooth functions of sample means. Numerical examples are provided to demonstrate the efficacy of the proposed method.

Keywords: asymptotic approximation; Edgeworth approximation; maximum entropy density

1 Introduction

Asymptotic approximations to the distributions of estimators or test statistics are commonly used in statistical inferences, such as approximate confidence limits for a parameter of interest or $p$-values for a hypothesis test. According to the central limit theorem, a statistic often has a limiting normal distribution under mild regularity conditions. The normal approximation, however, can be inadequate, especially when sample size is small. In this study, we are concerned with using higher order approximation methods to approximate the distributions of statistics with limiting normal distributions.

The most commonly used higher order approximation method is the Edgeworth expansion, which takes a polynomial form weighted by a normal distribution or density function. Suppose $\{Y_i\}_{i=1}^n$ is an independently and identically distributed random sample with mean zero and variance one. Let $X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$ and $f_n$ the density of $X_n$. The two-term Edgeworth approximation to $f_n$ takes the form

$$g_n(x) = \phi(x) \left\{ 1 + \frac{\kappa_3 \mathcal{H}_3(x)}{6\sqrt{n}} + \frac{\kappa_4 \mathcal{H}_4(x)}{24n} + \frac{\kappa_5^2 \mathcal{H}_6(x)}{72n} \right\},$$

where $\phi(x)$ is the standard normal density, $\kappa_j$ is the $j^{th}$ cumulant of $Y_1$, and $\mathcal{H}_j$ is the $j^{th}$ Hermite polynomial of $x$. The Edgeworth expansion is known to be quite accurate near the mean of a distribution. However due to its polynomial form, it can sometimes produce spurious oscillations and give poor fits to the exact distributions in parts of the tails. Moreover density and tail approximations can produce negative values.

Aiming to restrain the tail difficulties due to polynomial oscillations, Niki and Konishi (1985) proposed a transformation method for the Edgeworth approximation. Noting that the degree of
Information Theoretic Asymptotic Approximations

polynomial in (1) is reduced to 4 if $\kappa_3 = 0$, they suggested transforming the statistic $x$ to $\varphi(x)$ such that the third cumulant of $\varphi(x)$ is zero. This transformation, however, is rather difficult to derive and does not guarantee the positiveness of density approximations.

Field and Hampel (1982) discussed a variant of the Edgeworth expansion, which takes the form $h_n(x) = \exp(\sum_{j=0}^{K} \gamma_{n,j} x^j)$. Collecting coefficients of polynomial in the exponent up to order $n^{-1}$, they obtained

$$h_n(x) = \phi(x) \exp\left\{ \frac{3\kappa_4 - 5\kappa_3}{24n} - \frac{\kappa_3}{2\sqrt{n}}x + \frac{2\kappa_3^2 - \kappa_4 - \kappa_2}{4n}x^2 + \frac{\kappa_3}{6\sqrt{n}}x^3 + \frac{\kappa_4 - 3\kappa_3^2}{24n}x^4 \right\}. \tag{2}$$

In this study, we call (2) the Exponential Edgeworth (E.E.) approximation due to its close relation to the Edgeworth expansion. Clearly, the E.E. approximation to densities is always positive. In addition, compared with the Edgeworth expansion (1), the degree of polynomial (in the exponent) is reduced to 4. Thus, the E.E. approximation is more likely to restrain the oscillation associated with polynomials. This approximation, however, is generally not integrable. In addition, we note that it has an undesirable information distortion effect. Like the Edgeworth expansion, $h_n$ is based solely on the first four cumulants of the underlying distribution. The Edgeworth expansion maintains the information in the sense that its first four cumulants match those of the underlying distribution. In contrast, the first four cumulants of the E.E. approximation, if exist, generally differ from the true cumulants.

In this study, we propose an alternative higher order approximation which is similar to the E.E. approximation in terms of functional form but free of the above-mentioned difficulties. We recognize the advantages of the E.E. approximation in its positiveness and reduced order of polynomials. We then ask if we can adjust the approximation by choosing the polynomial coefficients in the exponent of (2) to restore information such that its first four moments match those of $f_n$. Furthermore, does this adjustment, if feasible, improve the approximation in any way?

We show that the answers to both questions are positive. In particular, we propose an approximation with the form

$$f^*_n(x) = \exp\left\{ \sum_{j=1}^{4} \gamma^*_n x^j - \psi(\gamma^*_n) \right\}, \tag{3}$$

where $\gamma^*_n = (\gamma^*_{n,1}, \ldots, \gamma^*_{n,4})$ are the coefficients of the canonical exponential density, and $\psi(\gamma^*_n)$ is a normalization term such that $f^*_n$ integrates to unity. The proposed approximation has an appealing information theoretic interpretation as the Maximum Entropy (M.E.) density, which is obtained by maximizing the information entropy subject to given moment conditions. The M.E. approxima-
tion is always positive and well-defined on a bounded support. When the underlying distribution has an unbounded support, the M.E. approximation in its conventional form is not necessarily integrable. Instead of assuming a bounded support, we propose a modified M.E. approximation with a dampening function which ensures its integrability and is at the same time asymptotically negligible.

Existing studies on the M.E. density focus on its usage as a density estimator. The degree of the polynomial in the exponent is often prescribed by the number of known moments (Jaynes, 1957), or is allowed to increase with sample size (Barron and Sheu, 1991). This study is conceptually different in that it uses the M.E. density to approximate asymptotic distributions of statistics. For some given number of polynomial terms, we investigate the performance of the approximation as $n \to \infty$. To the best of our knowledge, this is the first study to use the M.E. density for asymptotic inference. We establish in the general framework of smooth functions of sample means that the M.E. approximations based on the first four moments of $f_n$ are asymptotically equivalent to the two-term Edgeworth expansion up to order $n^{-1}$. In this study we focus on approximations up to order $n^{-1}$ since expansions beyond order $n^{-1}$ are rarely used in practice. Higher order results can be obtained in a similar manner. We provide some numerical examples, whose results suggest that the proposed method is a useful complement to the Edgeworth expansion, especially when the sample size is small.

The rest of the paper is organized as follows. Section 2 reviews the Edgeworth expansion and some relevant variants. Section 3 introduces the Maximum Entropy density estimation and generalizes it to distributions with unbounded supports. Section 4 establishes the asymptotic properties of the proposed M.E. approximation. Section 5 provides some numerical examples. Some concluding remarks are given in Section 6. Proofs of theorems and some technical details are collected in the Appendices.

2 Brief Review of Edgeworth Expansion

The Edgeworth expansion has its root in the Gram-Charlier differential series. In his seminal work, Edgeworth (1905) applied the Gram-Charlier expansion to the distribution of a sample mean and collected terms in powers of $n^{-1/2}$, which led to an asymptotic expansion. It has been extensively studied in the literature. See for example, Barndorff-Nielsen and Cox (1989), Hall (1992), and references therein. The Edgeworth expansion applies to not only sample means, but also more general statistics. In this section, we focus on the Edgeworth expansion of smooth functions of
sample means as in Hall (1992), and derive corresponding E.E. expansion.

Additional notations are required for more general statistics than sample means. We use bold face to denote a $d$-dimensional vector $v$ and $v_{(j)}$ its $j^{th}$ element. Let $S_n$ be a statistic with a limiting standard normal distribution, and $X_1, X_2, \ldots$ be an iid random column $d$-vector with mean $\mu$, and put $\bar{X} = n^{-1} \sum X_i$. Suppose we have a smooth function $B : \mathbb{R}^d \to \mathbb{R}$ satisfying $B(\mu) = 0$, and $S_n = n^{1/2} B(\bar{X})$. For example, $B(x) = \{\tau(x) - \tau(\mu)\} / \delta(\mu)$, where $\tau(\mu)$ is estimated by $\tau(\bar{X})$ and $\delta(\mu)^2$ is the asymptotic variance of $n^{1/2} \tau(\bar{X})$; or $B(x) = \{\tau(x) - \tau(\mu)\} / \delta(x)$, where $\tau(\cdot)$ and $\delta(\cdot)$ are smooth functions, and $\delta(\bar{X})$ is an estimate of $\delta(\mu)$.

Since the exact moments for $S_n$ are not available in general, we construct instead approximate moments. Define $Z = n^{1/2} (\bar{X} - \mu)$ and

$$b_{i_1 \ldots i_d} = \left( \frac{\partial^d}{\partial x_{(i_1)} \cdots \partial x_{(i_d)}} B(x) \right)_{x=\mu}.$$ 

By a Taylor expansion, and since $B(\mu) = 0$,

$$S_n = n^{1/2} B(\bar{X}) = \sum_{i_1 = 1}^d b_{i_1} Z_{(i_1)} + n^{-1/2} \frac{1}{2} \sum_{i_1 = 1}^d \sum_{i_2 = 1}^d b_{i_1 i_2} Z_{(i_1)} Z_{(i_2)}$$

$$+ n^{-1} \frac{1}{6} \sum_{i_1 = 1}^d \sum_{i_2 = 1}^d \sum_{i_3 = 1}^d b_{i_1 i_2 i_3} Z_{(i_1)} Z_{(i_2)} Z_{(i_3)} + O_p\left(n^{-3/2}\right).$$

The (approximate) moments of $S_n$ can then be computed from the above expansion. In principle, one can compute its moments to an arbitrary degree of accuracy by taking a sufficiently large number of terms. Below we give general formulae that allow approximations accurate to order $n^{-1}$.

In particular, we have

$$\tilde{\mu}_{n,1} = \frac{1}{\sqrt{n}} s_{1,2}, \quad \tilde{\mu}_{n,2} = s_{2,1} + \frac{1}{n} s_{2,2}, \quad \tilde{\mu}_{n,3} = \frac{1}{\sqrt{n}} s_{3,1}, \quad \tilde{\mu}_{n,4} = s_{4,1} + \frac{1}{n} s_{4,2},$$

where $s_{1,2}, \ldots, s_{4,2}$ are defined in the Appendix. It is seen that $\mathbb{E}[S_n^j] = \tilde{\mu}_{n,j} + o(n^{-1})$, $j = 1, \ldots, 4$.

Denote the distribution and the density functions of $S_n$ by $F_n$ and $f_n$ respectively. The corresponding Edgeworth expansion can be constructed using its approximate cumulants. See for example, Bhattacharya and Ghosh (1978) and Hall (1992). In particular, we have

$$\tilde{\kappa}_{n,1} = \frac{1}{\sqrt{n}} k_{1,2}, \quad \tilde{\kappa}_{n,2} = k_{2,1} + \frac{1}{n} k_{2,2}, \quad \tilde{\kappa}_{n,3} = \frac{1}{\sqrt{n}} k_{3,1}, \quad \tilde{\kappa}_{n,4} = \frac{1}{n} k_{4,1},$$

where $k_{1,2}, \ldots, k_{4,1}$ are functions of interest.
where \( k_{1,2}, \ldots, k_{4,1} \) are defined in the Appendix. Let \( \kappa_{n,j} \) be the \( j \)th cumulant of \( S_n \). One can show that \( \kappa_{n,j} = \tilde{\kappa}_{n,j} + o(n^{-1}), j = 1, \ldots, 4 \). We can then construct the two-term Edgeworth approximation to \( f_n \), which takes the form

\[
\tilde{g}_n(x) = \phi(x) \left\{ 1 + \frac{Q_1(x)}{\sqrt{n}} + \frac{Q_2(x)}{n} \right\},
\]

(6)

\[
Q_1(x) = \frac{k_{1,2} \mathcal{H}_1(x)}{\sigma} + \frac{k_{3,1} \mathcal{H}_3(x)}{6\sigma^3},
\]

\[
Q_2(x) = \frac{(k_{2,2} + k_{1,2}^2) \mathcal{H}_2(x)}{2\sigma^2} + \frac{(k_{4,1} + 4k_{1,2}k_{3,1}) \mathcal{H}_4(x)}{24\sigma^4} + \frac{k_{3,1}^2 \mathcal{H}_6(x)}{72\sigma^6}.
\]

Note that \( \tilde{g}_n(x) \) is a polynomial of degree 6, which might cause the so-called ‘tail difficulties’ discussed in Wallace (1958): The tail of the Edgeworth approximation exhibits increasing oscillation as \( |x| \) increases, due to the high degree polynomial. Although it always integrates to unity, the Edgeworth expansion is not necessarily a bona fide density, because it sometimes produces negative density approximations at the tails of distributions.

Let a general statistic \( S_n = \sqrt{n}(T_n - \mu_T)/\sigma_T \), where \( \mu_T \) and \( \sigma_T^2 \) are the mean and variance of \( T_n \). Niki and Konishi (1985) noted that the tail difficulties can be mitigated through a transformation \( \varphi(T_n) \) such that the third cumulant of standardized \( \varphi(T_n) \) is zero, in which case the Edgeworth expansion is reduced to a degree 4 polynomial. The transformed statistic takes the form

\[
z_n = \sqrt{n} \frac{\varphi(T_n) - \varphi(\mu_T)}{\sigma_T \varphi'(\mu_T)},
\]

where \( \varphi' \) is the first derivative of \( \varphi \). For example, let \( \chi_k^2 \) be the \( \chi^2 \) variate with \( k \) degrees of freedom, then \( \varphi(x) = x^{1/3} \). It is, however, generally rather difficult to derive the proposed transformation except for some special cases. Moreover, there is no guarantee that the Edgeworth expansion associated with the transformed statistic is always positive.

Being a higher order expansion around the mean, the Edgeworth expansion is known to work well near the mean of a distribution, but the performance deteriorates when one moves away from the mean (towards the tails). Hampel (1973) introduces the so-called “small sample asymptotic” method, which is essentially a local Edgeworth expansion. Although generally more accurate than the Edgeworth expansion, Hampel’s method requires the knowledge of \( f_n \), which limits its applicability in practice. Field and Hampel (1982) discussed an approximation to Hampel’s expansion that only requires the first few cumulants of \( S_n \). When \( S_n \) is the sample mean, their method yields the E.E. expansion (2). Note that this approximation can also be obtained by “putting the Edgeworth expansion back to the exponent” and collecting terms such that its first order Taylor expansion is equivalent to the Edgeworth expansion.

We now derive the E.E. expansion associated with the general statistic \( S_n \). Denote the E.E.
expansion by $\phi(x) \exp(K(x))$. It can be constructed such that Taylor expansion of $\exp(K(x))$ at $K(x) = 0$ up to order $n^{-1}$ matches the second factor of $\tilde{g}_n(x)$. We then have:

$$
\tilde{h}_n(x) = \phi(x) \exp \left\{ \frac{Q'_1(x)}{\sqrt{n}} + \frac{Q'_{2a}(x) + Q'_{2b}(x)}{n} \right\},
$$

(7)

$$
Q'_1(x) = \frac{k_{1,2} \mathcal{H}_1(x)}{\sigma} + \frac{k_{3,1} \mathcal{H}_3(x)}{6\sigma^3},
$$

$$
Q'_{2a}(x) = \frac{k_{2,2} \mathcal{H}_2(x)}{2\sigma^2} + \frac{k_{4,1} \mathcal{H}_4(x)}{24\sigma^4},
$$

$$
Q'_{2b}(x) = \frac{k_{3,2}^2 (\mathcal{H}_2(x) - \mathcal{H}_1^2(x))}{2\sigma^2} + \frac{k_{1,2} k_{3,1} (\mathcal{H}_4(x) - \mathcal{H}_1(x) \mathcal{H}_3(x))}{6\sigma^4} + \frac{k_{3,3}^2 (\mathcal{H}_6(x) - \mathcal{H}_3^2(x))}{72\sigma^6}.
$$

Note that the exponent of $\tilde{h}_n(x)$ is a polynomial of degree 4 since $\mathcal{H}_6(x) - \mathcal{H}_3^2(x)$ is a degree 4 polynomial. For a given $x \in (-\infty, \infty)$, $\tilde{h}_n(x)$ is equivalent to the Edgeworth expansion $\tilde{g}_n(x)$ up to order $n^{-1}$. Unlike the Edgeworth expansion, the E.E. expansion is always positive.

One problem with the E.E. approximation is that it is not integrable on the real line. To see this, note that we can rewrite $\tilde{h}_n(x)$ in the canonical exponential form $\exp(\sum_{j=0}^4 \gamma_n x^j)$. Obviously, $\tilde{h}_n(x)$ goes to infinity when $|x| \to \infty$ if $\gamma_{n,4} > 0$. A less obvious problem is about the preservation of information in this approximation. Like the two-term Edgeworth expansion, the E.E. expansion $\tilde{h}_n(x)$ is based only on the cumulants $\{\tilde{\kappa}_{n,j}\}_{j=1}^4$. However unlike the Edgeworth expansion, the first four cumulants of the E.E. approximation, if exist, generally differ from $\{\tilde{\kappa}_{n,j}\}_{j=1}^4$. Below we propose an approximation that has the same functional form as (7) and at the same time preserves information such that the first four moments of the proposed approximation match $\{\tilde{\mu}_{n,j}\}_{j=1}^4$. In fact, given the canonical exponential form of (7), the notion of matching moments coincides with the classical maximum likelihood estimation. As we show below, it also arises naturally from the framework of maximizing the information entropy subject to given moment restrictions.

3 Maximum entropy approximation

In this section, we first provide a brief review of the classical maximum entropy density. We then generalize the M.E. density to approximate asymptotic distributions of statistics defined on possibly unbounded supports.
3.1 Maximum entropy density

The approximation of a probability density function by sequences of exponential families arises naturally according to the principle of maximum entropy or minimum relative entropy. Given a continuous random variable $X$ with a density function $f$, the differential entropy is defined as

$$W(f) = -\int f(x) \log f(x) \, dx.$$  \hspace{1cm} (8)

The principle of maximum entropy provides a method of constructing a density from known moment conditions. Suppose the density function $f$ is unknown but a small number of moments of $X$ are given. There might exist an infinite number of densities satisfying given moment conditions. Jaynes (1957)'s Principle of Maximum Entropy suggests selecting the density, among all candidates satisfying the moment conditions, that maximizes the entropy. The resulting maximum entropy density “is uniquely determined as the one which is maximally noncommittal with regard to missing information, and that it agrees with what is known, but expresses maximum uncertainty with respect to all other matters.”

Formally, the maximum entropy density is defined as

$$f^* = \arg \max_f W(f)$$

subject to the moment constraints

$$\int f(x) \, dx = 1, \int x^j f(x) \, dx = \mu_j, \ j = 1, \ldots, K,$$  \hspace{1cm} (9)

where $\mu_j = EX^j$. The problem can be solved using calculus of variations. Denote the Lagrangian

$$L = \int f(x) \log f(x) \, dx - \gamma_0 \left\{ \int f(x) \, dx - 1 \right\} - \sum_{j=1}^K \gamma_j \left\{ \int x^j f(x) \, dx - \mu_j \right\}.$$  

The necessary conditions for a stationary value are given by the Euler-Lagrange equation

$$\log f^*(x) + 1 - \gamma_0^* - \sum_{j=1}^K \gamma_j^* x^j = 0,$$  \hspace{1cm} (10)

plus the constraints given in (9). Thus from (10), the maximum entropy density takes the general
Information Theoretic Asymptotic Approximations

form

\[ f^* (x) = \exp \left\{ \sum_{j=1}^{K} \gamma_j^* x^j - 1 + \gamma_0^* \right\} = \exp \left\{ \sum_{j=1}^{K} \gamma_j^* x^j - \psi (\gamma^*) \right\}, \] (11)

where \( \gamma^* = (\gamma_1^*, \ldots, \gamma_K^*) \), and \( \psi (\gamma^*) = \log \left\{ \int \exp \left( \sum_{j=1}^{K} \gamma_j^* x^j \right) dx \right\} \) is the normalizing factor under the assumption \( \int \exp \left( \sum_{j=1}^{K} \gamma_j^* x^j \right) dx < \infty \). The Lagrangian multiplier \( \gamma^* \) can be solved from the \( K \) moment conditions in (9) by plugging \( f = f^* \) into (9). For distributions defined on \( \mathbb{R} \), necessary conditions for the integrability of \( f^* \) are that \( K \) is even and \( \gamma_K^* < 0 \).

A closely related concept is the relative entropy, also called cross entropy or Kullback-Leibler distance, between densities \( f \) and \( f_0 \) defined on a common support:

\[ D (f || f_0) = \int f (x) \log \frac{f (x)}{f_0 (x)} dx. \] (12)

It is known that \( D (f || f_0) \geq 0 \) and \( D (f || f_0) = 0 \) if and only if \( f (x) = f_0 (x) \) almost everywhere. The minimum relative entropy density is obtained by minimizing (12) subject to the same moment constraints (9). The resulting density takes the form

\[ f^* (x; f_0) = f_0 (x) \exp \left\{ \sum_{j=1}^{K} \gamma_j^* x^j - \psi (\gamma^*) \right\}, \] (13)

which is the unique density that is closest to the reference density \( f_0 \) in the sense of the relative entropy and at the same time satisfies given moment conditions. Shore and Johnson (1981) provide an axiomatic justification of this approach. The maximum entropy density is a special case of the minimum relative entropy density, wherein the reference density \( f_0 \) is set to be constant.

Existing studies on the M.E. densities mostly focus on its usage as a density estimator. The M.E. approximation is equivalent to approximating the logarithm of a density by polynomials, which has long been studied. Earlier studies on the approximation of log-densities using polynomials include Neyman (1937) and Good (1963). When sample moments rather than population moments are used, 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. Barron and Sheu (1991) allow the number of moments to increase with the sample size and establish formally the M.E. approximation as a nonparametric density estimator. Their results are further extended to multivariate densities in Wu (2007).

The current study is conceptually different from existing studies. The M.E. densities are used
to approximate distributions of statistics that are asymptotically normal. Instead of letting $K$, the degree of exponential polynomial, increase with sample size, we study the accuracy of the approximation when the sample size $n \to \infty$ under some given $K$. Moreover, below we propose a modified M.E. density that is integrable in $\mathbb{R}$. This is particularly important for asymptotic approximations to distributions of statistics that are defined on unbounded supports.

### 3.2 Approximation for distributions on an unbounded support

The existence of an integrable M.E. approximation on an unbounded support depends on the moment conditions. For example, consider the case where the first four moments, with $\mu_1 = 0$ and $\mu_2 = 1$, are used to construct an M.E. density. Rockinger and Jondeau (2002) compute numerically a boundary on the $(\mu_3, \mu_4)$ space beyond which an M.E. approximation is not obtainable. Under the same condition, Tagniali (2003) shows that an M.E. approximation is not integrable if $\mu_3 = 0$ and $\mu_4 > 0$.

To ensure its integrability, we propose to use a dampening function in the exponent of an M.E. density. A similar approach is used in Wang (1992) to ensure numerical stability of general saddlepoint approximations. Define the dampening function

$$\theta_n(x) = \frac{c \exp(|x|)}{n^2}, \quad (14)$$

where $c$ is a positive constant. The modified M.E. density takes the form

$$f^*_n(x) = \exp \left( \sum_{j=1}^{4} \gamma^*_n x^j - \theta_n(x) - \psi(\gamma^*_n) \right), \quad (15)$$

where the normalization term is defined as

$$\psi(\gamma^*_n) = \log \left\{ \int \exp \left( \sum_{j=1}^{4} \gamma^*_n x^j - \theta_n(x) \right) dx \right\}.$$

Alternatively, this M.E. density can be obtained by

$$\min_f \int f(x) \log \frac{f(x)}{\exp(-\theta_n(x))} dx$$
such that
\[ \int f(x) \, dx = 1, \int x^j f(x) \, dx = \mu_j(f_n), j = 1, \ldots, 4. \]

The modified M.E. approximation is essentially a minimum cross entropy density based on the same moment conditions, but with a nonconstant reference density \( \exp(-\theta_n(x)) \). This dampening function ensures that the exponent of the M.E. density is bounded above and goes to \(-\infty\) as \(|x| \to \infty\) such that the M.E. density is integrable in \( \mathbb{R} \). One may be tempted to use a term like \( \tilde{\theta}_n(x) = cx^r/n^2 \) as the dampening function, where \( r \) is an even integer larger than 4. However, integrability of the M.E. approximation is not warranted with this choice. As is shown below, the coefficient for \( x^4, \gamma_{n,4}^* \), in the exponent is \( O(n^{-1}) \). Suppose \( r = 6 \). For \( x = O(n^{1/3}) \), it follows that \( \gamma_{n,4}^* x^4 = O(n^{1/3}) \) and \( \tilde{\theta}_n(x) = O(1) \). Then for \( \gamma_{n,4}^* > 0 \), \( f_n^*(x) \) goes to infinity as \(|x| \to \infty\) and thus not integrable on \( \mathbb{R} \).

Below we use the same dampening function to ensure the integrability of the E.E. approximations in \( \mathbb{R} \). The modified E.E. expansion takes the form

\[ h_n(x) = \phi(x) \exp\left\{ \frac{3\kappa_4 - 5\kappa_3^2}{24n} x + \frac{2\kappa_3^2 - \kappa_4}{4n} x^2 + \frac{\kappa_3}{6\sqrt{n}} x^3 + \frac{\kappa_4 - 3\kappa_3^2}{24n^2} x^4 - \theta_n(x) \right\}, \quad (16) \]

where \( \theta_n(x) \) is given in (14). We note that the M.E. approximation adapts to the presence of the dampening function in the sense that its coefficients are determined through an optimization process in which the dampening function is incorporated as the reference density. In contrast, the E.E. expansion does not have this appealing property because its coefficients are invariant to the introduction of a dampening function. This noteworthy difference is explored numerically in Section 5.

Although it does not affect order \( n^{-1} \) asymptotic result, the dampening function might affect small sample performance of the M.E. and the E.E. approximations. In practice, we use the following rule: the constant \( c \) in the dampening function is chosen such that it is the smallest positive number that ensures both tails of the approximation decrease monotonically. In particular, let \( f_n^*(x; c) \) be the M.E. approximation with a dampening function. We select \( c \) according to

\[ c = \inf \left\{ a > 0 : \frac{df_n^*(x; a)}{dx} > 0 \text{ for } x < -2 \text{ and } \frac{df_n^*(x; a)}{dx} < 0 \text{ for } x > 2 \right\}. \quad (17) \]

In other words, \( c \) is chosen such that the approximation decays monotonically for \(|x| > 2\) as \(|x| \to \infty\), which corresponds to approximately 5% tail regions of the normal approximation. We use the lower bound \( c = 10^{-6} \) for \( c \) arbitrarily close to zero. The same rule is used to select the
dampening function for the E.E. approximation (16).

4 Asymptotic Properties

In this section, we derive asymptotic properties of the M.E. approximation. As noted by Wallace (1958), asymptotic orders of several alternative higher order approximations are established through their relationship with the Edgeworth approximation. The same approach is employed in our analysis of the M.E. approximation. In particular, we use the E.E. expansion to facilitate our analysis, noting that the E.E. expansion is closely related to the Edgeworth expansion and at the same time shares a common canonical exponential form with the M.E. approximation.

The asymptotic properties of the Edgeworth approximation have been extensively studied. Recall that we are interested in approximating the distribution of $S_n$ with a density function $f_n$.

We start with the following result for the Edgeworth approximation $\tilde{g}_n$ to $f_n$.

Lemma 1 (Hall (1992), p.78) Assume that the function $B$ has four continuous derivatives in a neighborhood of $\mu$ and satisfies

$$\sup_{n \geq 1} \int_{S(n,x)} (1 + \|y\|)^{-4} dy < \infty,$$

where $S(n,x) = \{y \in \mathbb{R}^d : n^{1/2} B \left( \mu + n^{-1/2} y \right) / \sigma = x \}$. Assume that $E\|X\|^4 < \infty$ and that the characteristic function $\varphi$ of $X$ satisfies

$$\int_{\mathbb{R}^d} |\varphi(t)|^\lambda dt < \infty$$

for some $\lambda \geq 1$. Then as $n \to \infty$, $|f_n(x) - \tilde{g}_n(x)| = o\left(n^{-1}\right)$ uniformly in $x$.

The above condition is a generalization of the Cramer condition for the Edgeworth expansion of distribution of sample means (Hall, 1992). We next show that the E.E. approximation (16), with $\kappa_j$ replaced by $\tilde{\kappa}_{n,j}, j = 1, \ldots, 4$, is equivalent to the two-term Edgeworth expansion at order $O(n^{-1})$.

Theorem 1 Let $\tilde{h}_n$ be the (modified) exponential Edgeworth approximation (16) to $f_n$. Under the conditions of Lemma 1, $|f_n(x) - \tilde{h}_n(x)| = o\left(n^{-1}\right)$ uniformly in $x$ as $n \to \infty$.

A proof of Theorem 1 and other proofs are all given in the Appendix.
Next let \( \hat{f}_n^* \) be the M.E. approximation obtained by maximizing the entropy subject to the approximate moments \( \{ \tilde{\mu}_{n,j} \}_{j=1}^4 \) in (4). To make it comparable with the Edgeworth and the E.E. expansions, we factor out \( \phi(x) \). Using the following slightly different form, we replace \( \tilde{\gamma}_{n,2}^* + \frac{1}{2} \) in (15) with \( \tilde{\gamma}_{n,2}^* \):

\[
\hat{f}_n^* = \phi(x) \exp \left( \sum_{j=1}^4 \tilde{\gamma}_{n,j}^* x^j - \theta_n(x) - \psi(\tilde{\gamma}_n^*) \right),
\]

with

\[
\psi(\tilde{\gamma}_n^*) = \log \left\{ \int \phi(x) \exp \left( \sum_{j=1}^4 \tilde{\gamma}_{n,j}^* x^j - \theta_n(x) \right) dx \right\}.
\]

The asymptotic properties of \( \hat{f}_n^* \) are established with the aid of the E.E. expansion (7). Heuristically, using Theorem 1, one can show that \( \mu_j(\hat{h}_n) = \tilde{\mu}_{n,j} + o(n^{-1}) \) for \( j = 1, \ldots, 4 \). Let \( \mu(f) = (\mu_1(f), \ldots, \mu_4(f)) \) and \( \|\mu(f)\| \) be the Euclidean norm. Since \( \mu(f_n) = \mu(\hat{f}_n^*) + o(n^{-1}) \), it follows that \( \|\mu(\hat{h}_n) - \mu(\hat{f}_n^*)\| = o(n^{-1}) \). Under this condition, we are then able to establish order \( n^{-1} \) equivalence between \( \hat{h}_n \) and \( \hat{f}_n^* \), who share a common exponential polynomial form. By Lemma 1 and Theorem 1, we demonstrate formally the following theorem.

**Theorem 2** Let \( \hat{f}_n^* \) be the M.E. approximation (18) to \( f_n \). Under the conditions of Lemma 1, \( |f_n(x) - \hat{f}_n^*(x)| = o(n^{-1}) \) uniformly in \( x \) as \( n \to \infty \).

The asymptotic properties of approximations to distribution functions follow directly from the above results. Let \( F_n(x) = P(S_n \leq x) \). Denote \( \hat{G}_n(x) \) the two-term Edgeworth expansion of \( F_n \). It is known that \( |F_n(x) - \hat{G}_n(x)| = o(n^{-1}) \); see e.g., Feller (1971), p.539. Define \( \hat{H}_n(x) = \int_{-\infty}^{x} \hat{h}_n(t) dt \) and \( \hat{F}_n^*(x) = \int_{-\infty}^{x} \hat{f}_n^*(t) dt \) respectively. Below we show that the same asymptotic order applies to \( \hat{H}_n \) and \( \hat{F}_n^* \).

**Theorem 3** Under the conditions of Lemma 1, (a) the E.E. approximations to distribution function satisfies that \( |F_n(x) - \hat{H}_n(x)| = o(n^{-1}) \) uniformly in \( x \) as \( n \to \infty \); (b) the M.E. approximation to the distribution function satisfies that \( |F_n(x) - \hat{F}_n^*(x)| = o(n^{-1}) \) uniformly in \( x \) as \( n \to \infty \).

### 5 Numerical examples

In this section, we provide some numerical examples to illustrate finite sample performance of the proposed method. Our experiments compare the Edgeworth, the E.E. and the M.E. approximations.
We also include the results of the normal approximation, which is nested in all three higher-order approximations. Alternative higher order asymptotic methods, such as the bootstrap and the saddlepoint approximation, are not considered. We focus our comparison between the M.E. approximation and the Edgeworth expansion because they share the same information requirement. For example, only the first four moments or cumulants are required for approximations up to order $n^{-1}$. The other methods are conceptually different: the bootstrap uses re-sampling, and the saddlepoint approximation requires the knowledge of the characteristic function. We leave the comparison with these alternative methods for future study.

Since both the Edgeworth and the M.E. approximations integrate to unity by construction, we normalize the E.E. expansion such that it is directly comparable with the other two. We are interested in $F_n$, the distribution of $S_n$ as a smooth function of sample means. Define $s_\alpha$ by

$$
\Pr (S_n \leq s_\alpha) = \alpha.
$$

Let $\Psi_n$ be any approximation to $F_n$. In each experiment, we report the one-sided and two-sided tail probabilities defined by

$$
\alpha_1 (\Psi_n) = \int_{-\infty}^{s_\alpha} d\Psi_n (s), \quad \alpha_2 (\Psi_n) = 1 - \int_{s_{\alpha/2}}^{s_{1-\alpha/2}} d\Psi_n (s).
$$

We also report the relative error defined as $|\alpha_i (\Psi_n) - \alpha| / \alpha, i = 1, 2$.

In the first example, suppose we are interested in the distribution of the standardized mean

$$
S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - n) / \sqrt{2n},
$$

where $\{Y_i\}$ follows a $\chi^2$ distribution with one degree of freedom. The exact density of $S_n$ is given by

$$
f_n (s) = \frac{\sqrt{2n} \left( \sqrt{2ns} + n \right)^{(n-2)/2}}{\Gamma \left( \frac{n}{2} \right) 2^{n/2}} e^{-\left( \sqrt{2ns} + n \right)/2}, \text{for } s > -\sqrt{n/2}.
$$

We set $n = 10$. The constant $c$ in the dampening function is chosen to be $10^{-6}$ for both the M.E. and the E.E. approximations. The results are reported in Table 1. The accuracy of all approximations generally improves as $\alpha$ increases. The three higher-order approximations provide comparable performance and dominate the normal approximation. Although the magnitudes of relative errors are similar, it is noted that both the one-sided and two-sided tail probabilities of the Edgeworth expansion are negative at extreme tails.

[Table 1 about here.]

We next examine a case where $f_n$ is symmetric and fat-tailed. Suppose $S_n$ is the standardized mean of random variables from the $t$-distribution with five degrees of freedom, where $n = 10$. Al-
though the exact distribution of $S_n$ is unknown, we can calculate its moments analytically ($\mu_{n,3} = 0$ and $\mu_{n,4} = 3.6$). As is discussed in the previous section, for symmetric fat-tailed distributions, both the M.E. and the E.E. approximations, if without a proper dampening function, are not integrable on the real line. In this experiment, the constant $c$ in the dampening function is chosen to be 0.56 and 4.05 for the M.E. and the E.E. approximations respectively, using the automatic selection rule (17). We used one million simulations to obtain the critical values of the distribution of $X_n$. The approximation results are reported in Table 2. The overall performance of the M.E. approximation is considerably better than that of the Edgeworth expansion, which in turn is better than the normal approximation. Unlike the previous examples, the performance of the E.E. approximation is essentially dominated by the other three. Note that a non-trivial $c$ is chosen for both the M.E. and the E.E. approximations in this case. A relatively heavy dampening function is chosen for the E.E. approximation to ensure its integrability. Its performance is clearly affected by the dampening function. In contrast, the dampening function required for the M.E. approximation is considerably weaker.

Generally, the selection of the constant in the dampening function has little effects on the M.E. approximation because it adapts to the dampening function by adjusting its coefficients through an optimization process. On the other hand, it has an nonnegligible effect on the E.E. approximation, whose coefficients are invariant to the introduction of the dampening function. To further investigate how sensitive the reported results are with respect to the specification of the dampening function, we conducted some further experiments. We used a grid search to locate the optimal $c$ for the M.E. and the E.E. approximations. The optimal $c$ is defined as the minimizer of the sum of the 10 relative errors reported in Table 2. The optimal $c$ for the M.E. approximation is 0.6, which is obtained through a grid search on $[0.1,5]$ with an increment of 0.1. The optimal $c$ for the E.E. approximation, which is 5, is obtained in a similar fashion through a grid search on $[1,20]$ with an increment of 1. Thus, the constant $c$ selected by the automatic rule, 0.56 and 4.05 respectively for the M.E. and the E.E. approximation, are rather close to their optimal values, which are generally infeasible in practice.

In the third experiment, we used sample moments rather than population moments in the approximation of the distribution of a sample mean. Suppose the variance is unknown in this case.
We have
\[ S_n = \sqrt{n} \frac{\bar{Y} - EY_1}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (Y_t - \bar{Y})^2}}, \]

where \( \bar{Y} = \frac{1}{n} \sum_{t=1}^{n} Y_t \). The studentised statistic is a smooth function of sample means. Without loss of generality, suppose that \( \{Y_t\} \) follows a distribution with mean zero and variance one. Using the general formulae given in the previous section, we obtain the following approximate moments for \( S_n \):

\[ \tilde{\mu}_{n,1} = -\frac{1}{2} \frac{\mu_3}{\sqrt{n}}, \quad \tilde{\mu}_{n,2} = 1 + \frac{3 + 2 \mu_3^2}{n}, \]
\[ \tilde{\mu}_{n,3} = -\frac{7}{2} \frac{\mu_3}{\sqrt{n}}, \quad \tilde{\mu}_{n,4} = 3 + \frac{30 + 28 \mu_3^2 - 2 \mu_4}{n}, \]

where \( \mu_j = E(Y_1^j), j = 1, \ldots, 4. \)

The distribution of \( S_n \) was obtained by one million Monte Carlo simulations. In our experiment, we set \( n = 20 \), and generated \( \{Y_t\} \) from the standard normal distribution. The population moments \( \mu_3 \) and \( \mu_4 \) are replaced by sample moments in the approximations. The dampening function is selected automatically according to the rule in (17). The experiment is repeated 500 times. Table 3 reports the average tail probabilities and corresponding standard errors. The results are qualitatively similar to those in Table 2. The M.E. and the Edgeworth approximations dominate the E.E. and the normal approximations. The overall performance of the M.E. approximation is better than that of the Edgeworth approximations, especially for tail probability at 0.01 and 0.05. In addition, the standard errors of the M.E. approximations are smaller than those of the Edgeworth approximations in most cases.

Table 3 about here.

6 Conclusion

In this paper, we have proposed a maximum entropy approach for approximating the asymptotic distributions of a smooth function of sample means. Conventional maximum entropy densities are typically defined on a bounded support. The proposed maximum entropy approximation, equipped with an asymptotically negligible dampening function, is well defined on the real line. We have established order \( n^{-1} \) asymptotic equivalence between the proposed method and the classical Edgeworth expansion. Our numerical experiments show that the proposed method is comparable
with and sometimes outperforms the Edgeworth expansion. Finally, we note that the proposed method can be generalized to approximating bootstrap distributions of statistics and those of high dimensional statistics in future research.

References


Information Theoretic Asymptotic Approximations


Corresponding Author : Ximing Wu, Department of Agricultural Economics, Texas A&M University, College Station, TX 77843-2124. Email: xwu@ag.tamu.edu

Appendix 1: Approximate Moments and Cumulants

We first define \( \mu_{i_1 \ldots i_j} = E[(X - \mu)_{(i_1)} \cdots (X - \mu)_{(i_j)}], j \geq 1 \), and

\[
\begin{align*}
\nu_1 &= \mu_{i_1}, \\
\nu_2 &= \mu_{i_1 i_2}, \\
\nu_3 &= \mu_{i_1 i_2 i_3}, \\
\nu_4 &= \mu_{i_1 i_2 i_3 i_4}, \\
\nu_{2,2} &= \mu_{i_1 i_2}, \\
\nu_{2,3} &= \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_3 i_4}, \\
\nu_{2,4} &= \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4}, \\
\nu_{2,5} &= \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4}, \\
\nu_{2,6} &= \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4} + \mu_{i_1 i_2 i_3 i_4}
\end{align*}
\]

where the dependence of \( \nu' \)'s on \( i_1, \ldots, i_6 \) is suppressed for simplicity.
One can then show that

\[ E[Z_{(i_1)}] = \nu_1 = 0, \]
\[ E[Z_{(i_1)}Z_{(i_2)}] = \nu_2, \]
\[ E[Z_{(i_1)}Z_{(i_2)}Z_{(i_3)}] = n^{-1/2}\nu_3, \]
\[ E[Z_{(i_1)}Z_{(i_2)}Z_{(i_3)}Z_{(i_4)}] = \nu_{2,2} + n^{-1}(\nu_4 - \nu_{2,2}) + o\left(n^{-1}\right), \]
\[ E[Z_{(i_1)}\cdots Z_{(i_5)}] = n^{-1/2}\nu_{2,3} + o\left(n^{-1}\right), \]
\[ E[Z_{(i_1)}\cdots Z_{(i_6)}] = \nu_{2,2,2} + O\left(n^{-1}\right). \]

Expanding \( E[S_i^j], j = 1, \ldots, 4 \), and collecting terms by order of \( n^{-1/2} \) then yields the first four approximate moments of \( S_n \). Denoting the \( j \)-fold summation \( \sum_{i_1=1}^{d} \cdots \sum_{i_j=1}^{d} \) by \( \sum_{i_1,\ldots,i_j} \) to ease notation, we obtain approximate moments of \( S_n \) as follows:

\[
\bar{\mu}_{n,1} = \frac{1}{\sqrt{n}} \frac{1}{2} \sum_{i_1,i_2} b_{i_1i_2} \nu_2 = \frac{1}{\sqrt{n}} s_{1,2},
\]
\[
\bar{\mu}_{n,2} = \sum_{i_1,i_2} b_{i_1} b_{i_2} \nu_2 + \frac{1}{n} \left\{ \sum_{i_1,i_2,i_3} b_{i_1} b_{i_2} b_{i_3} \nu_3 + \sum_{i_1,\ldots,i_4} \frac{1}{3} b_{i_1} b_{i_2} b_{i_3} b_{i_4} + \frac{1}{4} b_{i_1} b_{i_2} b_{i_3} \nu_{2,2} \right\}
= s_{2,1} + \frac{1}{n} s_{2,2},
\]
\[
\bar{\mu}_{n,3} = \frac{1}{\sqrt{n}} \left( \sum_{i_1,i_2,i_3} b_{i_1} b_{i_2} b_{i_3} \nu_3 + \sum_{i_1,\ldots,i_4} \frac{3}{2} b_{i_1} b_{i_2} b_{i_3} b_{i_4} \nu_{2,2} \right) = \frac{1}{\sqrt{n}} s_{3,1},
\]
\[
\bar{\mu}_{n,4} = \sum_{i_1,i_2,i_3} b_{i_1} b_{i_2} b_{i_3} b_{i_4} \nu_{2,2} + \frac{1}{n} \left\{ \sum_{i_1,\ldots,i_4} b_{i_1} b_{i_2} b_{i_3} b_{i_4} (\nu_4 - \nu_{2,2}) + \sum_{i_1,\ldots,i_5} 2 b_{i_1} b_{i_2} b_{i_3} b_{i_4} b_{i_5} \nu_{2,2} \right\}
+ \sum_{i_1,\ldots,i_6} \left( \frac{2}{3} b_{i_1} b_{i_2} b_{i_3} b_{i_4} b_{i_5} b_{i_6} + \frac{3}{2} b_{i_1} b_{i_2} b_{i_3} b_{i_4} b_{i_5} b_{i_6} \right) \nu_{2,2,2}
= s_{4,1} + \frac{1}{n} s_{4,2},
\]

where the definitions of \( s_{1,2}, \ldots, s_{4,2} \) are self-evident. Since usually only a few \( b_{i_1,\ldots,i_j} \) terms are nonzero, the actual formulae for the approximate moments of \( S_n \) are often simpler than as appeared above.

To construct approximate cumulants, we first write the cumulants as a power series of \( n^{-1/2} \)
such that

\[
\begin{align*}
\kappa_{n,1} &= E[S_n] = n^{-1/2}s_{1,2} + o(n^{-1}) = n^{-1/2}k_{1,2} + o(n^{-1}), \\
\kappa_{n,2} &= E\left[S_n^2\right] - E[S_n]^2 = s_{2,1} + n^{-1}(s_{2,2} - s_{1,2}^2) + o(n^{-1}) = k_{2,1} + n^{-1}k_{2,2} + o(n^{-1}), \\
\kappa_{n,3} &= E\left[S_n^3\right] - 3E\left[S_n^2\right]E[S_n] + 2E[S_n]^3 \\
&= n^{-1/2}(s_{3,1} - 3s_{1,2}s_{2,1}) + o(n^{-1}) = n^{-1/2}k_{3,1} + o(n^{-1}), \\
\kappa_{n,4} &= E\left[S_n^4\right] - 4E\left[S_n^3\right]E[S_n] + 3E\left[S_n^2\right]^2 + 12E\left[S_n^2\right]E[S_n]^2 - 6E[S_n]^4 \\
&= n^{-1}(12s_{1,2}^2s_{2,1} - 6s_{2,2}s_{2,1} - 4s_{3,1}s_{1,2} + 4s_{4,2}) + o(n^{-1}) \\
&= n^{-1}k_{4,1} + o(n^{-1}),
\end{align*}
\]

where the definition of \(k_{1,2}, \ldots, k_{4,1}\) are self-evident. The last equality holds because \(s_{4,1} - 3s_{2,1}^2 = 0\). When \(S_n\) is a sample mean, its cumulants can be easily calculated by \(\kappa_{n,j} = \kappa_j/n^{j/2-1}\). For general statistics, however, this simple identity does not hold, and the cumulants have to be constructed from the moments constructed above.

**Appendix 2: Proof of Theorems**

**Proof of Theorem 1.** Denote the exponent of (16) by

\[
K(x) = \frac{P_3(x)}{6\sqrt{n}} + \frac{P_4(x)}{24n} - \frac{c\exp(|x|)}{n^2},
\]

where \(P_3(x)\) and \(P_4(x)\) are polynomials of degree 3 and 4 respectively, whose coefficients are constants that do not depend on \(n\).

Without loss of generality, assume that \(K(x) > 0\). By Taylor’s theorem, there exists a \(z\) such that \(0 < K(z) < K(x)\), and

\[
h_n(x) = \phi(x)\left\{1 + K(x) + \frac{K^2(x)}{2!} + \frac{K^3(x)}{3!}\exp(K(z))\right\} \\
= \phi(x)\left\{1 + \frac{P_3(x)}{6\sqrt{n}} + \frac{P_4(x)}{24n} + \frac{P_3^2(x)}{72n}\right\} + \phi(x)Q(x) + \phi(x)\frac{K^3(x)}{3!}\exp(K(z)), \quad (A.1)
\]
where
\[
Q(x) = \frac{P_4(x)}{2(24n)^2} + \frac{c^2 \exp(|x|)}{2n^4} + \frac{P_3(x) P_4(x)}{144n^{3/2}} - \frac{P_3(x) c \exp(|x|)}{6n^{5/2}} - \frac{P_4(x) c \exp(|x|)}{24n^3} - \frac{c \exp(|x|)}{n^2}.
\]

Note that the first term of (A.1) equals the Edgeworth expansion \( \tilde{g}_n(x) \).

Since \( \sup_{x \in \mathbb{R}} |\phi(x) x^r| < \infty \) for all non-negative integers \( r \), we have \( \sup_{x \in \mathbb{R}} |\phi(x) P_r(x)| < \infty \). Similarly one can show that \( \sup_{x \in \mathbb{R}} |P_r(x) \phi(x) \exp(|x|)| < \infty \). It follows that
\[
\phi(x) Q(x) = O \left( n^{-3/2} \right).
\]

Using a similar argument, we have
\[
K^3(x) = \left\{ \frac{P_3(x)}{6\sqrt{n}} + \frac{P_4(x)}{24n} \right\}^3 - 3 \left\{ \frac{P_3(x)}{6\sqrt{n}} + \frac{P_4(x)}{24n} \right\}^2 \frac{c \exp(|x|)}{n^2} + 3 \left\{ \frac{P_3(x)}{6\sqrt{n}} + \frac{P_4(x)}{24n} \right\} \left\{ \frac{c \exp(|x|)}{n^2} \right\}^2 - \left\{ \frac{c \exp(|x|)}{n^2} \right\}^3.
\]

Using the fact that
\[
\sup_{x \in \mathbb{R}} \left| \phi(x) P_r(x) \exp(|x|)^j \right| < \infty, j = 0, \ldots, 3, r = 0, 1, \ldots,
\]
we have
\[
\phi(x) K^3(x) = O \left( n^{-3/2} \right).
\]

It follows that
\[
\tilde{h}_n = \tilde{g}_n + O \left( n^{-3/2} \right) + O \left( n^{-3/2} \right) \exp(K(z)). \quad (A.3)
\]

It now suffices to show that \( \exp(K(z)) < \infty \) such that \( h_n = g_n + o(1) \). Note that when \( z = o \left( n^{1/6} \right) \), we have \( P_3(z) / \sqrt{n} = o(1), P_4(z) / n = o(1) \), thus \( K(z) \) is bounded above because its third term is negative. When \( z = O \left( n^{1/6} \right) \), \( P_3(z) / \sqrt{n} = O(1), P_4(z) / n = o(1) \). But \( \exp(|x|) / n^2 \) is increasing at a faster rate than any polynomial of \( z \). Thus, the third term of \( K(z) \) dominates and \( K(z) \) is bounded above. Lastly, for \( z = O \left( n^{1/6+\varepsilon} \right) , \varepsilon > 0 \), the third term of \( K(z) \) dominates by the same argument. Therefore, \( \frac{P_3(z)}{\sqrt{n}} + \frac{P_4(z)}{n} - \frac{c \exp(|z|)}{n^2} \) is bounded above for \( z \in \mathbb{R} \). It follows...
We then have non-negative integers \( r \) and \( r_1 \). Because for any arbitrary non-negative integer \( r \),

\[
\| \gamma_n (x) - h_n (x) \| = o (n^{-1}) .
\]

Thus \( \exp (K (z)) < \infty \) and from (A.3), \( | \tilde{g}_n (x) - \tilde{h}_n (x) | = o (n^{-1}) \). It then follows that

\[
| f_n (x) - h_n (x) | \leq | f_n (x) - \tilde{g}_n (x) | + | \tilde{g}_n (x) - \tilde{h}_n (x) | = o (n^{-1}) .
\]

Since the above result does not depend on \( x \), it holds uniformly in \( x \).

**Proof. of Theorem 2.** Using an argument similar to the proof of Theorem 1, we have for an arbitrary non-negative integer \( r \),

\[
| \mu_r (\tilde{g}_n) - \mu_r (\tilde{h}_n) | \leq \int | x^r | | \tilde{g}_n (x) - \tilde{h}_n (x) | \, dx
\]

\[
= \int | x^r | \phi (x) | Q (x) + \frac{K^3 (x)}{3!} \exp (K (z)) | \, dx
\]

\[
= o (n^{-1}) ,
\]

(A.4)

where \( Q (x) \) is given by (A.2). The last equality uses the fact that \( \int | x^r | \phi (x) \, dx < \infty \) for all non-negative integers \( r \) and the boundedness of \( \exp (K (z)) \) established in the proof of Theorem 1. Because for \( r = 1, \ldots, 4 \), \( \mu_r (g_n) = \mu_r (f_n^*) \), it follows that \( | \mu_r (f_n^*) - \mu_r (h_n) | = o (n^{-1}) \).

We then have \( \| \mu (f_n^*) - \mu (h_n) \| = o (n^{-1}) \), where \( \| v \| = \sqrt{\sum_{j=1}^{d} v_j^2} \) is the Euclidean norm of a \( d \)-dimensional vector \( v \).

Let \( f (x) = \phi (x) \exp \left( \sum_{j=0}^{4} \gamma_j x^j - \frac{\exp (|x|)}{n^2} \right) \) and \( \mu (f) \) be its first four moments. There is a unique correspondence between \( \mu (f) \) and \( \{ \gamma_j \}_{j=0}^{4} \). Denote the coefficients \( \{ \gamma_j \}_{j=0}^{4} \) of \( f \) by \( \gamma (\mu (f)) \).

According to Lemma 5 of Barron and Sheu (1991), \( \| \gamma (\mu (f)) - \gamma (\mu (g)) \| \) and \( \| \mu (f) - \mu (g) \| \) are of the same asymptotic order for \( \mu (f) \) and \( \mu (g) \) sufficiently close. Denote the coefficients of the M.E. and the E.E. approximation by \( \tilde{\gamma}_n \) and \( \gamma_n' \) respectively. Since \( \| \mu (f_n^*) - \mu (h_n) \| = o (n^{-1}) \), we have \( \| \tilde{\gamma}_n - \gamma_n' \| = o (n^{-1}) \).

Define \( \tilde{K}^* (x) = \sum_{j=0}^{4} \tilde{\gamma}_{n,j} x^j - \frac{\exp (|x|)}{n^2} \) and \( K' (x) \) similarly. Without loss of generality, assume
that $K'(x) > 0$. There exists a $z$ such that $0 < K'(z) < K'(x)$,

$$\left| \tilde{f}_n^*(x) - \tilde{h}_n(x) \right| = \phi(x) \left| \left\{ \sum_{j=0}^{4} (\tilde{\gamma}_{n,j} - \gamma'_{n,j}) x^j \right\} \exp (K'(z)) \right|$$

$$\leq \exp (K'(z)) \sum_{j=0}^{4} \phi(x) |x^j| |\tilde{\gamma}_{n,j} - \gamma'_{n,j}| .$$

From the proof of Theorem 1, we have $\exp (K'(z)) < \infty$ and $\sup_{x \in \mathbb{R}} |\phi(x) x^j| < \infty$ for all non-negative integers $j$. It follows that $\left| \tilde{f}_n^*(x) - \tilde{h}_n(x) \right| = o (n^{-1})$.

Lastly, we have

$$\left| f_n(x) - \tilde{f}_n^*(x) \right| \leq \left| f_n(x) - \tilde{h}_n(x) \right| + \left| \tilde{h}_n(x) - \tilde{f}_n^*(x) \right| = o (n^{-1}) .$$

Since the result does not depend on $x$, it holds uniformly in $x$. ■

**Proof of Theorem 3.** Note that

$$\left| F_n(x) - H_n(x) \right| = \left| \int_{-\infty}^{x} \left\{ f_n(t) - \tilde{h}_n(t) \right\} dt \right|$$

$$\leq \int_{-\infty}^{x} |f_n(t) - \tilde{h}_n(t)| dt$$

$$\leq \int |f_n(t) - \tilde{g}_n(t)| dt + \int |\tilde{g}_n(t) - \tilde{h}_n(t)| dt .$$

According to Hall (1992, p.51), the first integral is of order $o (n^{-1})$. The second integral is of order $o (n^{-1})$ according to (A.4) in the previous proof. It follows that $\left| F_n(x) - H_n(x) \right| = o (n^{-1})$ uniformly in $x$.

Similarly, we have

$$\left| F_n(x) - \tilde{F}_n^*(x) \right| = \left| \int_{-\infty}^{x} \left\{ f_n(t) - \tilde{f}_n^*(t) \right\} dt \right|$$

$$\leq \int_{-\infty}^{x} |f_n(t) - \tilde{f}_n^*(t)| dt$$

$$\leq \int |f_n(t) - \tilde{h}_n(t)| dt + \int |\tilde{h}_n(t) - \tilde{f}_n^*(t)| dt .$$

Taking integration of the right hand side of inequality (A.4) yields that $\int |\tilde{h}_n(t) - \tilde{f}_n^*(t)| dt = o (n^{-1})$. Note that the presence of $\phi(x)$ in (A.4) ensures the $n^{-1}$ order is preserved in the integration. Therefore, $\left| F_n(x) - \tilde{F}_n^*(x) \right| = o (n^{-1})$ uniformly in $x$. ■
Table 1: Tail probabilities of approximated distributions of the standardized mean of $\chi^2$ variables (NORM: normal; ED: Edgeworth; E.E.: exponential Edgeworth; M.E: maximum entropy; r.e.: relative errors)

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>NORM</th>
<th>r.e.</th>
<th>ED</th>
<th>r.e.</th>
<th>E.E.</th>
<th>r.e.</th>
<th>M.E.</th>
<th>r.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.0284</td>
<td>27.36</td>
<td>-0.0044</td>
<td>5.37</td>
<td>0.0080</td>
<td>7.03</td>
<td>0.0050</td>
<td>4.00</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0397</td>
<td>6.94</td>
<td>0.0032</td>
<td>0.35</td>
<td>0.0163</td>
<td>2.25</td>
<td>0.0106</td>
<td>1.12</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0481</td>
<td>3.81</td>
<td>0.0105</td>
<td>0.05</td>
<td>0.0236</td>
<td>1.36</td>
<td>0.0159</td>
<td>0.59</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0877</td>
<td>0.75</td>
<td>0.0553</td>
<td>0.11</td>
<td>0.0687</td>
<td>0.37</td>
<td>0.0513</td>
<td>0.03</td>
</tr>
</tbody>
</table>

\[
\alpha_1 (\Psi_n) = \int_{-\infty}^{s_\alpha} d\Psi_n (s)
\]

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>NORM</th>
<th>r.e.</th>
<th>ED</th>
<th>r.e.</th>
<th>E.E.</th>
<th>r.e.</th>
<th>M.E.</th>
<th>r.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.0254</td>
<td>24.40</td>
<td>-0.0057</td>
<td>6.69</td>
<td>0.0063</td>
<td>5.29</td>
<td>0.0040</td>
<td>2.98</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0339</td>
<td>5.78</td>
<td>0.0010</td>
<td>0.80</td>
<td>0.0117</td>
<td>1.34</td>
<td>0.0098</td>
<td>0.96</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0401</td>
<td>3.01</td>
<td>0.0089</td>
<td>0.11</td>
<td>0.0163</td>
<td>0.63</td>
<td>0.0163</td>
<td>0.63</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0751</td>
<td>0.50</td>
<td>0.0563</td>
<td>0.13</td>
<td>0.0493</td>
<td>0.01</td>
<td>0.0579</td>
<td>0.16</td>
</tr>
<tr>
<td>α</td>
<td>NORM</td>
<td>r.e.</td>
<td>ED</td>
<td>r.e.</td>
<td>E.E.</td>
<td>r.e.</td>
<td>M.E.</td>
<td>r.e.</td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>0.001</td>
<td>0.0003</td>
<td>0.69</td>
<td>0.0012</td>
<td>0.16</td>
<td>0.0022</td>
<td>1.15</td>
<td>0.0009</td>
<td>0.05</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0035</td>
<td>0.30</td>
<td>0.0066</td>
<td>0.31</td>
<td>0.0054</td>
<td>0.09</td>
<td>0.0048</td>
<td>0.03</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0084</td>
<td>0.16</td>
<td>0.0121</td>
<td>0.21</td>
<td>0.0091</td>
<td>0.09</td>
<td>0.0097</td>
<td>0.03</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0513</td>
<td>0.03</td>
<td>0.0498</td>
<td>0.00</td>
<td>0.0409</td>
<td>0.18</td>
<td>0.0497</td>
<td>0.01</td>
</tr>
</tbody>
</table>

\[
\alpha_1 (\Psi_n) = \int_{-\infty}^{s_n} d\Psi_n (s)
\]

<table>
<thead>
<tr>
<th>α</th>
<th>NORM</th>
<th>r.e.</th>
<th>ED</th>
<th>r.e.</th>
<th>E.E.</th>
<th>r.e.</th>
<th>M.E.</th>
<th>r.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.0002</td>
<td>0.81</td>
<td>0.0010</td>
<td>0.05</td>
<td>0.0033</td>
<td>2.27</td>
<td>0.0010</td>
<td>0.02</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0027</td>
<td>0.46</td>
<td>0.0067</td>
<td>0.33</td>
<td>0.0070</td>
<td>0.39</td>
<td>0.0048</td>
<td>0.04</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0070</td>
<td>0.30</td>
<td>0.0131</td>
<td>0.31</td>
<td>0.0109</td>
<td>0.09</td>
<td>0.0097</td>
<td>0.03</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0484</td>
<td>0.03</td>
<td>0.0534</td>
<td>0.07</td>
<td>0.0411</td>
<td>0.18</td>
<td>0.0492</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\[
\alpha_2 (\Psi_n) = 1 - \int_{s_{\alpha/2}}^{s_1} d\Psi_n (s)
\]
Table 3: Empirical tail probabilities of approximated distribution of studentised mean of 20 standard normal random variables (NORM: normal; ED: Edgeworth; E.E.: exponential Edgeworth; M.E: maximum entropy; s.e.: standard error)

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>NORM</th>
<th>s.e.</th>
<th>ED</th>
<th>s.e.</th>
<th>E.E.</th>
<th>s.e.</th>
<th>M.E.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.0001</td>
<td>(0.0045)</td>
<td>0.0012</td>
<td>(0.0013)</td>
<td>0.0074</td>
<td>(0.0105)</td>
<td>0.0011</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0017</td>
<td>(0.0100)</td>
<td>0.0070</td>
<td>(0.0059)</td>
<td>0.0130</td>
<td>(0.0163)</td>
<td>0.0053</td>
<td>(0.0042)</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0046</td>
<td>(0.0135)</td>
<td>0.0132</td>
<td>(0.0093)</td>
<td>0.0177</td>
<td>(0.0195)</td>
<td>0.0105</td>
<td>(0.0065)</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0478</td>
<td>(0.0240)</td>
<td>0.0535</td>
<td>(0.0208)</td>
<td>0.0511</td>
<td>(0.0296)</td>
<td>0.0530</td>
<td>(0.0156)</td>
</tr>
</tbody>
</table>

$$\alpha_1 \left( \hat{\Psi}_n \right) = \int_{-\infty}^{s_\alpha} d\hat{\Psi}_n (s)$$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>NORM</th>
<th>s.e.</th>
<th>ED</th>
<th>s.e.</th>
<th>E.E.</th>
<th>s.e.</th>
<th>M.E.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.0001</td>
<td>(0.0074)</td>
<td>0.0010</td>
<td>(0.0012)</td>
<td>0.0123</td>
<td>(0.0097)</td>
<td>0.0012</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0011</td>
<td>(0.0148)</td>
<td>0.0070</td>
<td>(0.0064)</td>
<td>0.0204</td>
<td>(0.0145)</td>
<td>0.0053</td>
<td>(0.0036)</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0033</td>
<td>(0.0197)</td>
<td>0.0141</td>
<td>(0.0105)</td>
<td>0.0261</td>
<td>(0.0164)</td>
<td>0.0105</td>
<td>(0.0056)</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0314</td>
<td>(0.0351)</td>
<td>0.0579</td>
<td>(0.0177)</td>
<td>0.0600</td>
<td>(0.0181)</td>
<td>0.0525</td>
<td>(0.0131)</td>
</tr>
</tbody>
</table>