The Annals of Statistics
2020, Vol. 48, No. 4, 1940–1958
https://doi.org/10.1214/19-AOS1871

ON THE VALIDITY OF THE FORMAL EDGEWORTH EXPANSION FOR POSTERIOR DENSITIES

BY JOHN E. KOLASSA[1] AND TODD A. KUFFNER[2]

[1]*Department of Statistics and Biostatistics, Rutgers University, kolassa@stat.rutgers.edu*
[2]*Department of Mathematics and Statistics, Washington University in St. Louis, kuffner@wustl.edu*

We consider a fundamental open problem in parametric Bayesian theory, namely the validity of the formal Edgeworth expansion of the posterior density. While the study of valid asymptotic expansions for posterior distributions constitutes a rich literature, the validity of the formal Edgeworth expansion has not been rigorously established. Several authors have claimed connections of various posterior expansions with the classical Edgeworth expansion, or have simply assumed its validity. Our main result settles this open problem. We also prove a lemma concerning the order of posterior cumulants which is of independent interest in Bayesian parametric theory. The most relevant literature is synthesized and compared to the newly-derived Edgeworth expansions. Numerical investigations illustrate that our expansion has the behavior expected of an Edgeworth expansion, and that it has better performance than the other existing expansion which was previously claimed to be of Edgeworth type.

1. Introduction. The Edgeworth series expansion of a density function is a fundamental tool in classical asymptotic theory for parametric inference. Such expansions are natural refinements to first-order asymptotic Gaussian approximations to large-sample distributions of suitably centered and normalized functionals of sequences of random variables, $X_1, \ldots, X_n$. Here, n is the available sample size, asymptotic means $n \to \infty$, and first-order means that the approximation using only the standard Gaussian distribution incurs an absolute approximation error of order $O(n^{-1/2})$. The term formal in conjunction with Edgeworth expansions means that derivation of the expansion begins by expanding the log characteristic function, and then utilizes Fourier inversion to obtain the corresponding density ([1], page 280). The coefficients of such expansions are expressed in terms of cumulants of the underlying density, together with a set of orthogonal basis functions for a suitably general hypothesized function space for the density being approximated. Asymptotic expansions are said to be valid if the absolute approximation error incurred, as an order of magnitude in n, by truncating the series expansion after a finite number of terms, is of the same asymptotic order as the first omitted term. For a distribution standardized to zero mean and unit variance, the Edgeworth approximation to a density using the first four moments is given by

$$e_{4,n}(\vartheta) = \phi(\vartheta)\{1 + h_3(\vartheta)\kappa_3/6 + \kappa_4 h_4(\vartheta)/24 - \kappa_3^2 h_6(\vartheta)/72\}. \tag{1}$$

Here, ϑ represents a potential value for the random variable, $h_2(\vartheta) = \vartheta^2 - 1$, $h_3(\vartheta) = \vartheta^3 - 3\vartheta$, $h_4(\vartheta) = \vartheta^4 - 6\vartheta^2 + 3$ and $h_6(\vartheta) = \vartheta^6 - 15\vartheta^4 + 45\vartheta^2 - 15$. These polynomials are called Hermite polynomials. The quantities multiplying the Hermite polynomials are calculated from the difference between the cumulants of the distribution being approximated, and the distribution forming the basis of the approximation (in this case, the normal). These cumulant differences are converted using standard formulas for calculating moments from cumulants to form what are called in this manuscript pseudo-moments; these pseudo-moments

Received October 2017; revised March 2019.
MSC2020 subject classifications. Primary 62E20, 62F15; secondary 62F99.
Key words and phrases. Edgeworth expansion, higher-order asymptotics, posterior, cumulant expansion.

are functions of the original cumulants, which, in turn, can be expressed in terms of the original moments. These pseudo-moments are examined termwise to retain only those terms larger than the order of the expected error. We refer to approximations precisely of the form (1), with these specific Hermite polynomials, and with the coefficients (that is, the pseudo-moments) calculated from the exact cumulants, which are, in turn, calculated from the exact moments, as a true Edgeworth approximation.

The validity of the formal Edgeworth expansion is of foundational importance, in the sense that this implies a certain degree of regularity of the statistical model, and the expansion itself offers deeper insights into the finite sample performance of many frequentist inference procedures, such as those based on the likelihood. Applying standard arguments to justify term-by-term integration of the truncated Edgeworth expansion for a density yields the corresponding Edgeworth approximation for the cumulative distribution function. Such expansions are essential to studying the coverage accuracy of approximate confidence sets, as well as establishing higher-order relationships between different methods for constructing such approximate confidence sets. Our understanding of the bootstrap has also been greatly enhanced by studying connections with Edgeworth expansions [21].

In contrast to central limit theorems in the frequentist context, the large-sample Gaussian approximation of the posterior distribution of a suitably centered and normalized parameter is typically justified using a Bernstein–von Mises theorem. Such theorems establish stochastic convergence of the total variation distance between the sequence of posterior distributions and an appropriate Gaussian distribution, where stochastic convergence is with respect to the true distribution from which samples are independently drawn. The exact form depends on the centering statistic and its corresponding variance estimate. For lucid discussions, see van der Vaart ([46], Chapter 10), and Ventura and Reid [47]. Reid [39] gives background on asymptotic techniques for various types of statistical inference, including Bayesian inference.

Somewhat surprisingly, the validity of formal Edgeworth expansion for the Bayesian posterior density has not previously been established. This is less surprising when the challenging nature of this problem is understood. The term valid in the context of posterior expansions means that, when approximating the posterior by truncating the series, the absolute error is uniformly of the proper order on a set of parameter values whose posterior probability does not go to zero. While some authors have studied related expansions, or made claims about the similarity of such expansions to classical Edgeworth expansions, to our knowledge there is no existing proof of the validity of the formal Edgeworth series expansion for posterior distributions. Apart from formal Edgeworth expansion validity being of foundational importance, we note that approximate posterior inference through higher-order asymptotics remains of interest in parametric Bayesian theory; see, for example, Ruli, Sartori and Ventura [40] or Kharroubi and Sweeting [30]. Other approximate Bayesian inference procedures, such as variational Bayes, or approximate Bayesian computation, have become popular due to their ability to circumvent computationally expensive Markov chain Monte Carlo procedures. Higher-order asymptotics offers another route to approximate Bayesian inference which, in many settings of practical interest, can be extremely accurate and inexpensive to implement.

To establish validity of the formal Edgeworth expansion for a density, it is required to show that the coefficients in the expansion, that is, the cumulants of the statistical functional which is being approximated, are of the proper asymptotic order to ensure that the terms of the expansion have the claimed orders as powers of $n^{-1/2}$. This entails formally proving that power series expansions for those cumulants are valid.

The existing results concerning validity of cumulant expansions are all within the sampling distribution framework, and as such are not applicable to the Bayesian setting. Numerous authors have studied expansions for posterior moments, but such results do not actually imply that the corresponding cumulants are of the proper asymptotic order. Edgeworth expansion

relies on proper order for cumulants of the variable under investigation, after dividing by its standard deviation. These cumulants for the standardized variable are known as invariant cumulants, and demonstrating their proper order is more delicate than demonstrating the proper order for the underlying moments. As an example, consider the relationship between the posterior variance σ^2 and the fourth central moment μ_4'. In order for the formal Edgeworth expansion to have the proper asymptotic behavior, $\mu_4'\sigma^{-4} - 3 = O(1/n)$, and so $\mu_4' = \sigma^4 O(1)$. However, the converse—that $\mu_4' = \sigma^4 O(1)$ implies $\mu_4'\sigma^{-4} - 3 = O(1/n)$—does not hold. Hence bounds on the moments (even for central moments) cannot guarantee proper size of the invariant cumulants. Results exploring the parameter centered at something only approximating the posterior mean (e.g., the maximum likelihood estimator), and standardized by something other than the exact posterior standard deviation (e.g., an approximation based on the Fisher information) will not ensure proper order for the invariant cumulants.

In this paper, we make several novel contributions. First, we prove the validity of the formal Edgeworth series expansion of the posterior density and distribution function. This requires us to prove a lemma concerning the asymptotic order of posterior cumulants, which is of independent interest, and appears to be the first rigorously established general result of this type. We also synthesize the relevant literature on posterior expansions, giving rigorous explanations of how existing Edgeworth-type expansions are not actually formal Edgeworth expansions. Finally, we provide a numerical illustration of our results.

2. Background.

2.1. *Posterior expansions.* Before proceeding, we note that one could consider either analytic or stochastic expansions for posterior densities. For an analytic expansion, the observed data sequence is viewed as a subsequence of a given, fixed infinite sequence of realizations. Deriving analytic expansions amounts to showing that the posterior has certain asymptotic properties for a given, well-behaved infinite sequence of observations. The stochastic expansion viewpoint asserts that such well-behaved sequences occur with probability tending to one, with respect to the true data generating probability distribution. In this paper, we consider analytic approximations for a given well-behaved infinite sequence of observations, though it would be possible to give analogous stochastic versions where $O(\cdot)$ terms are replaced by corresponding $O_p(\cdot)$ terms; see Sweeting [43] and Kass, Tierney and Kadane [29].

Within the existing literature on posterior expansions, there have been several dominant approaches to obtaining approximations of posterior quantities. A common starting point is to express the posterior mean or density as a ratio of two integrals. Expansions for the numerator and denominator separately, which may be truncated and integrated to yield integral approximations, can yield valid expansions for the posterior quantity through formal division of the numerator and denominator expansions. One approach which heavily emphasizes Taylor expansion is found in Johnson [27, 28] and Ghosh, Sinha and Joshi [20]. The expansions given by Bertail and Lo [3] are similar to these earlier papers, though the latter authors consider centering at the posterior mean or posterior mode rather than the maximum likelihood estimator, and find that this can be advantageous from a second-order efficiency standpoint. Bertail and Lo [3] also claim that their expansions are Edgeworth expansions, but this is not correct.

The second and more popular approach to deriving posterior expansions is to apply Laplace's method to approximate the respective integrals, and then use the ratio of these approximations; see Lindley [33, 34], Davison [13], Tierney and Kadane [44]. Validity of Laplace expansions for posterior densities is considered in Kass, Tierney and Kadane [29]. We explore the Edgeworth alternative because the resulting approximation is more easily interpreted in terms of the first few moments of the true posterior distribution, and because,

conceptually, exploration of the Edgeworth expansion fills a hole in the range of rigorously-demonstrated applications.

Another well-established method of posterior expansions utilizes Stein's identity; see, for instance, Woodroofe [54, 55], Weng [49], Weng and Tsai [51] and Weng [50]. In this paradigm, Weng [50] claimed to have established an Edgeworth expansion for the posterior density. Compared to other expansions, Weng's approach most closely resembles the final form of an Edgeworth expansion in that it is expressed in terms of moments, but it is not a formal Edgeworth expansion, and its structure is actually quite different from an Edgeworth expansion, as we show below.

From a formal Edgeworth series perspective, existing posterior expansions are centered at the wrong place, typically either the maximum likelihood estimator or true parameter value, instead of the posterior mean or something which is approximating the posterior mean. As noted by DasGupta ([10], Section 20.8), the maximum likelihood estimator and posterior mean are closely related. Suppose that one observes a sequence of observations $X^{(n)} = (X_1, \ldots, X_n)$, each of which are identical copies of a random variable X whose distribution P_θ depends on a scalar parameter θ. Write P_{θ_0} for the distribution corresponding to the true value θ_0 under which each component of $X^{(n)}$ is generated. Let $E(\theta|X^{(n)})$ denote the posterior mean under some prior density, and let $\hat{\theta}_n$ be the maximum likelihood estimator for θ based on $X^{(n)}$. Under standard regularity conditions, $E(\theta|X^{(n)}) - \hat{\theta}_n$, and also $n^{1/2}(E(\theta|X^{(n)}) - \hat{\theta}_n)$ converge in P_{θ_0}-probability to zero. Therefore,

$$E\big(\theta|X^{(n)}\big) - \hat{\theta}_n = o_p\big(n^{-1/2}\big) \tag{2}$$

under P_{θ_0}. The effects of centering in the wrong place are examined in Section 4. In particular, we discuss the expansions given by Weng [50] and Hartigan [22]. These two expansions are not Edgeworth series expansions, but for reasons explained in Section 4, these can be considered to have the closest relationship to our formal Edgeworth expansions.

It may appear strange to the reader that we are claiming the Edgeworth expansion for the posterior has not been established as valid, even under regularity conditions common in the literature. After all, there is the celebrated posterior Bartlett correction of Bickel and Ghosh [6], and the conventional derivation of the validity of the Bartlett correction requires a valid Edgeworth expansion. Some authors, for example, Chang and Mukerjee [9], refer to the Bickel and Ghosh regularity conditions on the Bayesian model specification ([6], page 1078), as Edgeworth assumptions. Indeed, Bickel and Ghosh [6] simply assume that an Edgeworth expansion exists, but not only do not provide a proof of validity of the Edgeworth approximation, but, in fact, do not make explicit whether the approximation they conjecture is of the standard Edgeworth form of a Gaussian approximation modified by the appropriate expression involving invariant cumulants, times the appropriate Hermite polynomial. Related to the approach of Bickel and Ghosh [6], some authors use an expansion for the log likelihood at the maximum likelihood estimate to obtain a posterior expansion which has some structure resembling an Edgeworth expansion; see Datta and Mukerjee ([11], equation (2.2.19)). As with other expansions mentioned above, this is not an Edgeworth expansion for several reasons. First, it is not derived by formal expansion of the posterior characteristic function. Second, the centering is at the maximum likelihood estimate, not the posterior mean. Third, the first correction term is a linear one, which vanishes in an Edgeworth expansion. Moreover, the coefficients are not cumulants of the posterior. We further note that in Bickel and Ghosh [6] and Datta and Mukerjee ([11], Lemma 4.2.1), an approximation is given for the posterior characteristic function of the log likelihood ratio statistic. DiCiccio and Stern [14] consider approximation of the posterior moment generating function of this statistic. All of these expansions rely on regularity conditions which amount to assuming the validity of the Edgeworth expansion, though none of these papers contain proofs, nor do they actually

use formal Edgeworth expansions in their arguments. A lucid discussion of frequentist and Bayesian Bartlett correction, and where the assumption of validity of Edgeworth expansions is essential, is found in DiCiccio and Stern [15].

2.2. *Validity and formal Edgeworth expansions for sampling distributions.* In the frequentist context, asymptotic expansions and their components are studied with respect to the sampling distribution of the relevant statistical functional, under hypothetical repeated sampling. Wallace [48] provided the conventional notion of validity for an asymptotic expansion. Suppose each element in a sequence of functions $\{g_n\}_{n\geq 1}$ is approximated by any partial sum of a series $\sum_{j=0}^{\infty} n^{-j/2}A_j(y)$, where the $A_j(\cdot)$ do not depend on n. If for some function $C_r(y)$, not depending on n, the absolute errors satisfy

$$\left| g_n(y) - \sum_{j=0}^{r} n^{-j/2}A_j(y) \right| \leq n^{-(r+1)/2}C_r(y),$$

then the asymptotic expansion is said to be valid to r terms. If the constant $C_r(y)$ does not depend on y, then the asymptotic expansion is called uniformly valid in y. Hence, validity requires that the absolute error in the approximation, using any partial sum, is of the same order of magnitude as the first neglected term.

Consider a scalar random variable X with characteristic function $\gamma_X(t) = E\exp(itX)$, and denote by $X^{(n)} = (X_1, \ldots, X_n)$ a sequence of independent and identically distributed copies of X. Suppose that it is required to approximate the density $g_n(y)$ of $Y_n = n^{1/2}s(X_1, \ldots, X_n)$ for some scalar-valued function $s(\cdot)$, such that Y_n is a centered and scaled statistic possessing an asymptotically standard normal distribution to first order. The formal Edgeworth expansion of the density $g_n(y)$ is derived according to the following steps; see, for example, Jensen ([26], Section 1.5), McCullagh ([35], Chapter 5), Hall ([21], Chapter 2) or Kolassa ([31], Chapter 3). First, Taylor expands the cumulant generating function of Y_n, $\log\gamma_{Y_n}(t)$, in a neighborhood of zero, $|t| < cn^{1/2}$ for some $c > 0$. Next, expand the Fourier inversion integral over the region $|t| < cn^{1/2}$. Then obtain a bound on the inversion integral over the region $|t| > cn^{1/2}$. If Y_n satisfies the assumptions of the smooth function model ([21], Section 2.4), then one can follow the program in the references above to rigorously establish the validity of the Edgeworth expansion for $g_n(y)$. Other standard references for Edgeworth series expansions include Feller [16], Bhattacharya and Ghosh [4], Bhattacharya and Rao [5] and Ghosh [17].

To ensure that the formal Edgeworth series expansion for the density of Y_n is valid in the sense of Wallace [48], it is required that the jth cumulant of Y_n, denoted by $\kappa_{j,n}$, is of order $n^{-(j-2)/2}$, and may expanded in a power series in n^{-1}:

$$\kappa_{j,n} = n^{-(j-2)/2}\left(c_{j,1} + n^{-1}c_{j,2} + n^{-2}c_{j,3} + \cdots\right), \quad j \geq 1. \tag{3}$$

Since Y_n is centered and scaled so that $\kappa_{1,n} = E(Y_n) \to 0$ and $\kappa_{2,n} = \mathrm{var}(Y_n) \to 1$, then $c_{1,1} = 0$ and $c_{2,1} = 1$. When the function s is the sum of its arguments, the cumulants of Y_n are the invariant cumulants of X; that is, they are the cumulants of X adjusted for rescaling of Y_n to unit standard deviation; the cumulants are adjusted by dividing by the appropriate power of the second cumulant. The origins of this result in the frequentist, repeated sampling setting can be traced to the combinatorial arguments of James [23, 24], James and Mayne [25] and Leonov and Shiryaev [32]. The interested reader is referred to Withers [52, 53], McCullagh ([35], Chapter 2), Hall ([21], Chapter 2), Mykland [36], Kolassa [31] and Stuart and Ord ([42], Chapters 12 and 13), for more details about cumulant expansions.

It is also of interest to integrate the Edgeworth series expansion of the density of Y_n to obtain an Edgeworth expansion for its corresponding distribution function. Analogously to

the density setting, this expansion is desired to be valid for fixed j as $n \to \infty$, and the remainder should be of the stated order uniformly in y. Sufficient regularity conditions ([21], Section 2.2), for the validity of this expansion to order j are that $E(|X|^{j+2}) < \infty$ and

$$\limsup_{|t|\to\infty} |\gamma_X(t)| < 1. \tag{4}$$

The latter condition is known as Cramér's condition.

2.3. *Conditional expansions and posterior expansions.* Since the posterior is a conditional density, it is natural to ask if an expansion for the posterior could be obtained by writing the posterior as a ratio of the joint density of $(\theta, X^{(n)})$ and the marginal density of $X^{(n)}$, and then taking the ratio of expansions for the numerator and denominator. This approach can be used in the frequentist sampling distribution framework concerning expansions for conditional densities. Specifically, after deriving Edgeworth expansions for the numerator and denominator, formal division of these series expansions yields what is referred to as a direct-direct Edgeworth expansion for the conditional density ([2], Chapter 7). Proving validity for these direct-direct expansions requires the analogous proofs of validity for expansions of conditional cumulants, which are in general very difficult. Such direct-direct expansions are not the same as a direct expansion of a conditional density, but the bigger obstacle to their utility is that they are non-Bayesian in nature. Standard sampling distribution arguments do not apply when deriving an Edgeworth expansion for the posterior density of $\vartheta_n = (\theta - \theta_0)/\sigma$, where θ_0 and σ are the posterior expectation and standard deviation. Both of these quantities depend on the data, and hence on n. In particular, in this posterior setting, one does not have independent and identically distributed θs, but rather a single θ. Furthermore, posterior inference is conditional on a single data set, without appealing to repeated sampling arguments. An obvious point worth emphasizing is that the consideration of increasingly larger sample sizes is not the same as considering hypothetical repeated sampling. As discussed above, we are assuming that the data represent a subsequence from a fixed infinite sequence, rather than repeated random samples from a probability distribution.

A major obstacle to proving validity of posterior Edgeworth expansions is the issue of the cumulant orders. In the sampling distribution framework, there are well-known results concerning the relationship between conditional and unconditional cumulants, but these results are unfortunately of no use in the posterior framework. In particular, Brillinger [8] established a theorem which permits computation of unconditional cumulants from conditional cumulants; see also Speed [41] and McCullagh ([35], Section 2.9 and Section 5.6). However, there is no converse to Brillinger's theorem, and even if there were, one must still overcome the issue that θ is a single random variable, rather than a sequence of independent and identically distributed random variables.

Due to the challenges just mentioned, there are no general results about posterior cumulants available in the literature. Pericchi, Sansó and Smith [37] give some specific results regarding the form of the cumulant generating function only relevant to exponential families. Hartigan ([22], page 1145) alluded to the order of posterior cumulants, but was not precise about how such orders could be shown.

3. Main results. We consider expansions for the posterior distribution of the scalar-valued quantity $\vartheta_n = (\theta - \theta_0)/\sigma$, for θ_0 and σ the posterior expectation and standard deviation. It is assumed throughout that an appropriate Bernstein–von Mises theorem holds for the sequence of posterior distributions of ϑ_n. Many asymptotic normality results appear in the literature; cf. van der Vaart ([46], Chapter 10), or Ghosh and Ramamoorthi ([19], Chapter 1). Since our goal is to prove validity of the Edgeworth expansion in some generality, we do not

discuss all of the conditions needed for the various specific Bernstein–von Mises theorems to hold. Our treatment of regularity conditions focuses on those conditions of particular relevance to the validity of the formal Edgeworth expansion. Our results assume that a suitable asymptotic normality result holds for the posterior distribution of ϑ_n.

The first step in our analysis is to prove that the invariant posterior cumulants admit a valid power series expansion, establishing that the coefficients in the Edgeworth series expansion will have the correct asymptotic order. We then prove validity of the expansion for the posterior density and distribution function, respectively.

3.1. *The order of posterior cumulants.* Given a sequence $X^{(n)} = (X_1, \ldots, X_n)$ of n random variables which, conditional on the value of a scalar random parameter θ taking values in a set $\Theta \subset \mathbb{R}$, are independent and identically distributed according to density $f(x|\theta)$, define $L_n(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$ to be the likelihood function. Here, $\mathbf{x}$ represents the observations of the sequence $X^{(n)}$. Denote the log likelihood function by $\ell_n(\theta) \equiv \ell_n(\theta; \mathbf{x})$. Assume that, prior to observing the data, the uncertainty about θ is described by a prior density function $\pi(\theta)$. The posterior density of θ is defined as

$$f(\theta|\mathbf{x}) = \frac{\pi(\theta)\prod_{i=1}^n f(x_i|\theta)}{\int_\Theta \pi(\theta)\prod_{i=1}^n f(x_i|\theta)\,d\theta} = \frac{L_n(\theta;\mathbf{x})\pi(\theta)}{\int_\Theta L_n(\theta;\mathbf{x})\pi(\theta)\,d\theta}.$$

Throughout, we suppress the dependence on $\mathbf{x}$ when there is no chance of confusion.

LEMMA 1. *Assume that the likelihood function has a single global maximizer $\hat{\theta}_n$. Define the average log likelihood $\bar{\ell}_n(\theta) = \ell_n(\theta)/n$, and assume that $\bar{\ell}_n$ and the log prior density have six continuous derivatives in a neighborhood of the form $\hat{\theta}_n \pm \epsilon$ for ϵ independent of n, and such that $\bar{\ell}_n(\theta) < \bar{\ell}_n(\hat{\theta}_n) - \delta$ for $\theta \notin (\hat{\theta}_n - \epsilon, \hat{\theta}_n + \epsilon)$, and assume that the second derivative of the average log likelihood is bounded away from zero on this neighborhood. Assume further that, for a sufficient number of observations, the posterior distribution is proper, and has six finite moments. Then the invariant cumulant of θ of order j, defined to be the cumulant of order j of $(\theta - \theta_0)/\sigma$, and denoted by κ_j, is $O(n^{(2-j)/2})$ for $j \in \{3, 4, 5\}$. Here, θ_0 and σ are the expectation and standard deviation of the posterior distribution, respectively.*

PROOF. Arguments in the first half of this proof hold for log likelihood functions and log prior densities with varying numbers of derivatives; denote this number by k, and until specified otherwise, it might, but need not be, 5. We have phrased this assumption in slightly more generality than is necessary, in order to facilitate future derivations of higher-order expansions. Our regularity conditions imply (2); hence there exists N (potentially dependent on the sample) so that for $n \geq N$, $|\theta_0 - \hat{\theta}_n| < \epsilon/2$, and so continuous derivatives to order k exist for $\bar{\ell}_n(\theta)$ and the log prior density at θ_0. Hence the log prior has an expansion

$$-\sum_{j=0}^{k} h_j(\theta - \theta_0)^j/j! + Q^\dagger(\theta^\dagger)(\theta - \theta_0)^{k+1}/(k+1)!,$$

and the log likelihood has an expansion

$$-n\left[\sum_{j=0}^{k} g_j(\theta - \theta_0)^j/j! + Q^*(\theta^*)(\theta - \theta_0)^{k+1}/(k+1)!\right],$$

where the coefficients g_j and h_j may be calculated from derivatives of the log likelihood and log prior, respectively. Here, $Q^\dagger(\theta)$ and $Q^*(\theta)$ are the standard Taylor series remainder terms, calculated from the derivatives of order $k+1$ of the log prior density and average

log likelihood, respectively, evaluated at parameter values $\theta^\dagger$ and θ^* intermediate between θ_0 and θ.

The log posterior can be expressed as

$$\sum_{j=0}^{k} p_j(\theta-\theta_0)^j/j! + \bigl[nQ^*(\theta)+Q(\theta)\bigr], \tag{5}$$

where $p_j = -h_j - ng_j$. The first term p_0 may be chosen to make the approximate posterior integrate to 1. Since $\theta^\dagger$ and θ^* are functions of θ, the error terms $Q^\dagger(\theta^\dagger)$ and $Q^*(\theta^*)$ may be taken as bounded for $\theta \in (\hat\theta_n - \epsilon/2, \hat\theta_n + \epsilon/2)$. The choice of θ_0 ensures that $p_1 = O(n^{1/2})$.

Let $\omega(\theta)$ be the polynomial resulting from retaining only terms of order k and smaller in the power series for $\exp(\sum_{j=3}^k p_j(\theta-\theta_0)^j)$. Let μ^* represent the extended Laplace approximation to the posterior moments; that is,

$$\mu_j^* = \int_{-\infty}^{\infty} \exp(p_0)\exp\biggl(\frac{1}{2}(\theta-\theta_0)^2 p_2\biggr)\theta^j\omega(\theta)\,d\theta.$$

By "extended Laplace," we refer to the approximation as implemented by [45].

Let $\mu_j = E(\theta^j|\mathbf{x})$ denote the true values of these moments. Choose $\epsilon > 0$ to satisfy the conditions of the lemma such that both

$$\bigl|\log\bigl(\pi(\theta_0)\bigr) + n\bar\ell_n(\theta_0) - \log\bigl(\pi(\theta)\bigr) - n\bar\ell_n(\theta)\bigr| \le p_2(\theta-\theta_0)^2/4$$

and

$$\biggl|\sum_{j=3}^{k} p_j(\theta-\theta_0)^j\biggr| \le p_2(\theta-\theta_0)^2/2,$$

for $|\theta-\theta_0| < \epsilon$. Since the difference between the maximum likelihood estimator and the posterior expectation is $O_p(n^{-1/2})$, then there exists m such that the posterior based on the first m observations $\mathbf{x}^*$ gives a proper posterior. Let $\mathbf{x}^\dagger$ represent the final $n-m$ observations. There exists $\delta > 0$ such that $\bar\ell_{n-m}(\theta;\mathbf{x}^\dagger) < \bar\ell_{n-m}(\theta_0;\mathbf{x}^\dagger) - \delta$ for $\theta \in (-\epsilon/2,\epsilon/2)^c$, and the contribution from outside of the interval to the absolute approximation error $|\mu_j^* - \mu_j|$ is bounded by $\exp(-(n-m)\delta)$. Inside $(-\epsilon/2,\epsilon/2)$, the contribution to $|\mu_j - \mu_j^*|$ is bounded by

$$\exp\biggl(-p_0 - \frac{n}{4}(\theta-\theta_0)^2 p_2\biggr)\biggl(\bigl|nQ^*(\theta)+Q(\theta)\bigr| + \frac{(\sum_{j=3}^k p_j(\theta-\theta_0)^j)^{k+1}}{(k+1)!}\biggr),$$

by Kolassa ([31], Theorem 2.5.3), which is $O(n^{-(k+1)/2})$.

Take $k=5$. In this case,

$$\begin{aligned}\omega(\theta) &= 1 - p_1(\theta-\theta_0) + p_2^2(\theta-\theta_0)^2/2 - (p_1^3+p_3)(\theta-\theta_0)^3/6\\ &\quad + (p_1^4 + 4p_3p_1 - p_4)(\theta-\theta_0)^4/24\\ &\quad - (\theta-\theta_0)^5(p_1^5 + 10p_3p_1^2 - 5p_4p_1 + p_5)/120.\end{aligned}$$

Moments approximated in this way are accurate to $O(n^{-7/2})$, and cumulants approximated using standard formulas for producing cumulants from moments ([31], page 10) are accurate to the same order. Denote the cumulant of θ of order j by β_j. The first cumulant is $\beta_1 = -g_1g_2^{-1} + (g_1h_2g_2^{-2} - h_1g_2^{-1})n^{-1} + O(1/n^2)$. Recall that the g_j terms are the coefficients in the expansion of the average log likelihood about the posterior mean. They are all $O(1)$ except g_1. Since the posterior mean is within $O(n^{-1/2})$ of the maximum likelihood estimate, then the choice of θ_0 as the posterior mean forces $g_1 = O(n^{-1/2})$. The second cumulant

is $\beta_2 = \frac{1}{2}g_2^{-1}n^{-1} + O(n^{-2})$, the third cumulant is $\beta_3 = -\frac{1}{6}g_3 g_2^{-3}n^{-2} + O(n^{-3})$ and the fourth cumulant is $\beta_4 = \frac{1}{24}(3g_3^2 g_2^{-5} - g_4 g_2^{-4})n^{-3} + O(n^{-7/2})$. The most delicate part of the preceding argument is the calculation of these cumulants, and ensuring that larger terms cancel to leave a remainder of the proper order. Further technical details used to establish the orders are provided in the Appendix. The proof is then completed by dividing the cumulants by the proper power of the second cumulant, β_2, which is $O(n^{-1})$, to see that the quotient is of the proper order. Specifically, the invariant cumulant of order j, $j \geq 3$, is $\kappa_j = \beta_j/\beta_2^{j/2}$. Since g_2 is bounded away from zero, the invariant cumulants of order 3 and 4 are $O(n^{-1/2})$ and $O(n^{-1})$, respectively. Similar calculations show that the invariant cumulant of order 5 is $O(n^{-3/2})$. □

REMARK 1. The argument above holds to provide bounds on moments of all orders. Orders of cumulants are more delicate, since cumulants are expressed in terms of differences of products of moments, and proper order for cumulants requires that leading terms of the representation properly cancel. At present, we know of no way to do this except on a case-by-case basis. This difficulty extends to bounds on invariant cumulants.

REMARK 2. In Section 1, we argued that one cannot bound the cumulants based on the order of the moments. We are instead bounding cumulants by getting a Laplace expansion for the moments, and observing that enough leading terms cancel to show that the cumulants are of the proper order.

REMARK 3. As one would expect, the h_j terms, corresponding to coefficients in the expansion of the log prior about the posterior mean, appear only in terms of order n^{-1} and smaller.

REMARK 4. Lemma 1 can easily be extended to multivariate cumulants. In this case, the condition placing a lower bound on the second derivative of the average log likelihood should be replaced by requiring that the smallest eigenvalue of the second derivative of the average log likelihood be bounded away from zero, and the condition on higher derivatives of the log likelihood is applied componentwise. Theorem 1 extends directly in the multivariate case. The resulting density approximation can be integrated termwise to give a multivariate version of Theorem 2.

3.2. *Validity of Edgeworth expansion for the posterior density.* In this section, we establish the validity of the formal Edgeworth approximation to the distribution of the standardized posterior of θ. Define $\vartheta_n = (\theta - \theta_0)/\sigma$, with θ_0 and σ respectively the posterior mean and standard deviation of θ. We use the notation ϑ to denote a generic value of ϑ_n. Let $\pi(\theta|\mathbf{x})$ represent the posterior density for θ. The density for the standardized parameter ϑ_n is $\pi(\theta_0 + \sigma\vartheta|\mathbf{x})\sigma$. Let

$$e_n(\vartheta) = \phi(\vartheta)\{1 + h_3(\vartheta)\kappa_3/6 + \kappa_4 h_4(\vartheta)/24 - \kappa_3^2 h_6(\vartheta)/72 + \cdots\} \tag{6}$$

represent the formal series formed by dividing the characteristic function $\varphi(\tau)$ of ϑ_n by $\exp(\vartheta^2/2)$, replacing powers of $-\tau$ by Hermite polynomials, and multiplying by the standard normal density. Here, κ_j are cumulants of ϑ_n, and hence the invariant cumulants of θ; by Lemma 1, κ_j satisfies

$$\kappa_j = O\big(n^{-(j-2)/2}\big), \quad 3 \leq j \leq 5. \tag{7}$$

THEOREM 1. *Let $\mathcal{A} = \mathcal{I}^n \subseteq \mathbb{R}^n$ be a subset of the sample space formed as the product of n copies of the same real interval. Choose $k \in \{2, 3, 4\}$. Suppose that the following assumptions hold*:

1. *The prior is absolutely continuous with respect to Lebesgue measure on the parameter space.*
2. *For any $\mathbf{x} \in \mathcal{A}$, the likelihood function $L_n(\theta; \mathbf{x})$ is a measurable function of θ.*
3. *The posterior is proper, and, for sufficiently large n, has a bounded density and moments to order 6.*
4. *For any $\mathbf{x} \in \mathcal{A}$, the likelihood function has a unique global maximizer $\hat{\theta}_n(\mathbf{x})$.*
5. *For any $\mathbf{x} \in \mathcal{A}$, the log likelihood function $\ell_n(\theta)$ is k-times differentiable in a neighborhood of $\hat{\theta}_n(\mathbf{x})$, with average second derivative $\ell_n''(\theta, \mathbf{x})/n$ bounded away from zero, and average jth derivative*

$$\ell_n^{(j)}(\theta, \mathbf{x})/n$$

bounded, for $2 \le j \le k$.

Let $e_{k,n}$ represent (6) truncated to contain only terms with products of invariant cumulants of the form $\prod_{m=1}^{j} \kappa_{r_m}$, such that $\sum_{m=1}^{j}(r_m - 2) < k$. Then

$$\sup_{\mathbf{x}\in\mathcal{A}} \left| e_{k,n}(\vartheta) - \pi(\theta_0 + \sigma\vartheta\,|\mathbf{x})\sigma \right| = O\left(n^{-k/2}\right),$$

uniformly in ϑ and uniformly in $\mathbf{x}$ in a compact subset of the sample space, but not relatively.

PROOF. Let $\lambda(\tau) = \int_{-\infty}^{\infty} L_n(\theta)\pi(\theta)\exp(i\theta\tau)\,d\theta$. The characteristic function for θ is then $\lambda(\tau)/\lambda(0)$, and the characteristic function for $\vartheta_n = (\theta - \theta_0)/\sigma$ is

$$\varphi(\tau) = \lambda\left(n^{1/2}\tau\right)\exp\left(-n^{1/2}\theta_0\tau i\right)/\lambda(0). \tag{8}$$

The Riemann–Lebesgue theorem, using Assumption 3 above, indicates that $|\lambda(\tau)| \le C_n'/|\tau|$ for $C_n' = 2\int_{-\infty}^{\infty} L_n(\theta)\pi(\theta)\,d\theta$; see Billingsley ([7], Theorem 26.1). Furthermore, by Assumption 3, there exist m and C_1 and C_2 such that

$$\int_{-\infty}^{\infty} |\lambda(\tau)|^m\,d\tau \le C_1 \quad \text{and hence} \quad |\varphi(\tau)| \le C_2/\left(n^{1/2}|\tau|\right); \tag{9}$$

C_1 and C_2 are calculated from the minimal sample yielding a bounded density, and are not dependent on n. Then the Fourier inversion of φ to obtain $e_{k,n}(\vartheta)$ as the posterior density of ϑ_n is performed by first expanding φ in τ near 0. The Fourier inversion integral results approximately in $e_{k,n}$.

More formally, let $\gamma_n(\zeta) = \int_{-\infty}^{\infty}\exp(i\zeta\vartheta)e_{k,n}(\vartheta)\,d\zeta$, for $e_{k,n}(\vartheta)$ defined in (6). In parallel with the development of Feller ([16], Section XVI.2), the posterior density for ϑ_n is given by

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-i\zeta\vartheta)\varphi(\zeta)\,d\zeta, \tag{10}$$

and the difference between the true posterior density and the Edgeworth series approximation of (6) is bounded by

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-i\zeta\vartheta)\left|\varphi(\zeta) - \gamma_n(\zeta)\right|d\zeta \\
&\quad= \frac{1}{2\pi}\int_{(-\delta n^{1/2},\delta n^{1/2})}\exp(-i\zeta\vartheta)\left|\varphi(\zeta) - \gamma_n(\zeta)\right|d\zeta \\
&\qquad+ \frac{1}{2\pi}\int_{(-\delta n^{1/2},\delta n^{1/2})^c}\exp(-i\zeta\vartheta)\left|\varphi(\zeta) - \gamma_n(\zeta)\right|d\zeta.
\end{aligned} \tag{11}$$

In (11), we bound the difference between the target and the approximation, on the characteristic function scale, as represented by the left of (11), and break the range of integration into two parts, as on the right. As discussed by Kolassa ([31], Section 3.7), the first of these integrals is bounded by $\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-i\zeta\vartheta)p(\zeta,n)/n^{k/2}\,d\zeta$, for $p(\zeta,n)$ a polynomial in ζ and $1/n^{1/2}$. This polynomial has coefficients that depend on derivatives of the log likelihood, and so the error is uniformly of the proper order. This polynomial is constructed in even powers of ζ and with positive coefficients, so as to make absolute values unnecessary. Note that result (9) applies to $\gamma_n(\zeta)$ as well as to $\varphi(\zeta)$; choose the resulting constants $m^*\geq m$, $C_1^*\geq C_1$ and $C_2^*\geq C_2$. These together show that the second integral in (11) is bounded by $(C_2^*/(\delta n^{1/2}))^{n-m^*}C_1^* n/\delta$, which is geometrically small. □

REMARK 5. To elaborate on which terms are retained for various values of k, when $k=2$, the term involving κ_3 times a Hermite polynomial and no other invariant cumulant is retained, because $j=1$ and $r_1=3$, and when $k=3$, the term involving κ_4 and a Hermite polynomial and no other invariant cumulant is retained, because $j=1$ and $r_1=4$. Also when $k=3$, the term involving κ_3^2 times a Hermite polynomial and no other invariant cumulant is retained, because $j=2$ and $r_1=r_2=3$.

REMARK 6. Assumption 3 requires that the posterior is proper, but not necessarily that the prior is proper. As is well known ([18], page 106), the Bernstein–von Mises theorem holds even if the prior is improper, provided that there exists an n_0 such that the posterior of θ given $(x_1,\ldots,x_{n_0})$ is proper for almost every realization of $(x_1,x_2,\ldots,x_{n_0})$.

3.3. *Validity of Edgeworth expansion for the posterior distribution function.* By assuming that the prior is a density, and that the likelihood is continuously differentiable, we actually have more smoothness than is required for Cramér's condition (4) to hold. However, as noted above, the extra smoothness implied by these assumptions is necessary to prove the validity of the cumulant expansions in Lemma 1.

THEOREM 2. *Under the assumptions of Theorem 1, define the Edgeworth approximation to the posterior cumulative distribution function of ϑ as*

$$E_{k,n}(\vartheta)=\Phi(\vartheta)-\phi(\vartheta)\{h_2(\vartheta)\kappa_3/6+\kappa_4 h_3(\vartheta)/24-\kappa_3^2 h_5(\vartheta)/72+\cdots\}. \tag{12}$$

The absolute error incurred in using (12) to approximate the distribution function of ϑ_n is of order $O(n^{-3/2})$, uniformly in ϑ and uniformly in $\mathbf{x}$ in a compact subset of the sample space, but not relatively.

PROOF. As noted by Kolassa ([31], equation (48)), the error in applying (12) to approximate the posterior distribution function is given by the left-hand side of (11), modified by dividing the integrand by $|\zeta|$, in this case with a density. By the previous application of the Riemann–Lebesgue theorem, the modified integral representing error converges absolutely, and is multiplied by the proper power of the sample size. □

4. Relationship to existing Edgeworth-type expansions. Weng [50] provides an asymptotic expansion for the posterior distribution of a parameter indexing a statistical model, given n observations which are independent and identically distributed conditionally on the parameter value, and assuming certain regularity conditions, by centering the distribution at the maximum likelihood estimate, and scaling the difference between a potential parameter value and the estimate by the second derivative of the log likelihood evaluated at the maximum likelihood. This produces an asymptotic expansion valid to $O(n^{-(s+1)/2})$,

and uses $3s-1$ terms. For example, when approximating the posterior CDF, the approximation with error $O(n^{-3/2})$ uses $s=2$, and hence uses five terms, including the leading term represented by the normal cumulative distribution function. This expansion includes Hermite polynomials to order 5, as is found in the standard Edgeworth expansion presented by, for example, McCullagh [35]. However, since the maximum likelihood estimate is not the same as the posterior expectation, the leading term in the Weng approximation does not match the target distribution as well as one centered at the true posterior expectation. This lack of match leads to a more complicated expansion. Furthermore, the example Weng presents provides finite sample performance that is inferior to that generally expected from an approximation with asymptotic error $O(n^{-3/2})$, as we illustrate in Section 5.

We now demonstrate that Weng's approximation has an error that is equivalent to that of our approximation; these calculations also demonstrate the differences between the two approximations. Weng's approximation for the posterior distribution function of $\tilde{\vartheta}_n = (\theta - \hat{\theta}_n)/\hat{\sigma}$, where $\hat{\theta}_n$ is the maximum likelihood estimate and $\hat{\sigma}$ is the square root of the observed information evaluated at $\hat{\theta}_n$, is of the form

$$P\big[(\theta-\hat{\theta})/\hat{\sigma} < \tilde{\vartheta}_n|\mathbf{x}\big] = \Phi(\tilde{\vartheta}_n) - \sum_{i=1}^{3s} q_{i-1}(\tilde{\vartheta}_n)\phi(\tilde{\vartheta}_n)c_i,$$

for q_i Hermite polynomials, and c_i constants given by Stein's lemma. Weng [50] shows that this approximation holds uniformly for $\tilde{\vartheta}_n \in \mathbb{R}$.

Hartigan [22] also provides an approximation to the posterior, in this case to the density, and obtains an approximation of a similar form. This approximation is also about a center other than the posterior expectation; in this case, the expansion is in the neighborhood of a true parameter value. We use the notation of Hartigan [22] to further clarify the distinction between our approximation and that of Weng [50].

Suppose data $X_1, \ldots, X_n$ is observed, with observations independent and identically distributed, conditional on a scalar parameter θ, with common log density $g(x_i|\theta)$ and log prior density $h(\theta)$. Let $\omega_j = \int_\Theta \theta^j \exp(h(\theta) + \sum_{i=1}^n g(x_i|\theta))\,d\theta$; these quantities depend on the data. The posterior expectation is then ω_1/ω_0. Take $s=2$; then Weng's approximation to the posterior distribution function is

$$\begin{aligned}&\Phi(\tilde{\vartheta}_n) + \phi(\tilde{\vartheta}_n)\big\{c_1 + \tilde{\vartheta}_n c_2 + (\tilde{\vartheta}_n^2 - 1)c_3 + (\tilde{\vartheta}_n^3 - \tilde{\vartheta}_n)c_4\\ &\quad + (\tilde{\vartheta}_n^5 - 10\tilde{\vartheta}_n^3 + 15\tilde{\vartheta}_n)c_6\big\},\end{aligned}$$

and to the density is

$$\begin{aligned}w(\tilde{\vartheta}_n) = \phi(\tilde{\vartheta}_n) - \phi(\tilde{\vartheta}_n)\big\{&\tilde{\vartheta}_n c_1 + (\tilde{\vartheta}_n^2 - 1)c_2 + (\tilde{\vartheta}_n^3 - \tilde{\vartheta}_n)c_3 + (\tilde{\vartheta}_n^4 - 6\tilde{\vartheta}_n^2 + 3)c_4\\ &+ (\tilde{\vartheta}_n^6 - 15\tilde{\vartheta}_n^4 + 45\tilde{\vartheta}_n^2 - 15)c_6\big\},\end{aligned}$$

uniformly (in θ) to $O(n^{-3/2})$. Then the expectation associated with this density approximation is $-c_1$, the variance associated with this approximation is $1 - c_1^2 - 2c_2$, and the approximation to the density $d(\rho)$ of $\rho = (\theta - \hat{\theta}_n)/\hat{\sigma} - c_1$ satisfies

$$d(\rho) = w\big(\rho(1-c_1^2-2c_2)^{1/2} - c_1\big)(1-c_1^2-2c_2)^{1/2} + O\big(n^{-3/2}\big). \tag{13}$$

Weng [50] notes that

$$c_1, c_3 = O\big(n^{-1/2}\big), \qquad c_2, c_4, c_6 = O\big(n^{-1}\big). \tag{14}$$

Expanding $w(\rho(1-c_1^2-2c_2)^{1/2} - c_1)(1-c_1^2-2c_2)^{1/2}$ in terms of powers of $n^{1/2}$, and bounding errors using, for example, Kolassa ([31], Theorem 2.5.3), one can exhibit $w(\rho(1-$

$c_1^2 - 2c_2)^{1/2} - c_1)(1 - c_1^2 - 2c_2)^{1/2}$ of the form

$$\begin{aligned}(15)\qquad &\phi(\rho) - \phi(\rho)\{\rho c_1^* + (\rho^2 - 1)c_2^* + (\rho^3 - \rho)c_3^* + (\rho^4 - 6\rho^2 + 3)c_4^* \\ &\quad + (\rho^6 - 15\rho^4 + 45\rho^2 - 15)c_6^*\},\end{aligned}$$

for constants c_j^* satisfying (14), and furthermore, $c_1^* = c_2^* = 0$. Hence (13) is an Edgeworth expansion to $O(n^{-3/2})$.

REMARK 7. We have given two proofs of the validity of Edgeworth expansion of the posterior. The first is a direct proof for the formal Edgeworth expansion (Theorem 1), while the second is not a formal Edgeworth expansion, but rather shows how to correct Weng's expansion due to using the wrong center. As the above arguments demonstrate, one can obtain an Edgeworth-type expansion by centering at the maximum likelihood estimate $\hat{\theta}_n$ and correcting. Such an expansion could have the same form as an Edgeworth series, but would not be a formal Edgeworth expansion, and would require additional work to compute the correction factors $c_1^*, \ldots, c_6^*$ in (15).

REMARK 8. We have used the Laplace approximation of the cumulants only to show that they are of the correct asymptotic order. It is not necessary to use the Laplace approximation for implementation of the expansion. In practice, any sufficiently accurate estimator of the posterior moments could be used to implement the Edgeworth expansion. One approach is given by Hartigan [22]. Another is to use the constants given by Weng and adjust them accordingly.

REMARK 9. Weng's expansion is actually more similar to a Gram–Charlier expansion of the posterior, not an Edgeworth expansion. Weng ensures that the pseudo-moments are of the correct order, but not the cumulants. For example, in her displayed equation (45), the set J_2 (corresponding to the n^{-1} term in the expansion) includes the sixth Hermite polynomial and its multiplier. For an Edgeworth expansion, this term is discarded.

REMARK 10. The proof of Weng [50] applies to distributions of manifest variables with no conditions limiting discreteness. Continuity in this case is ensured by the continuity of the prior. This application in the case of highly discrete manifest distributions carries over to our result. Contrast this with frequentist expansions, which require control of or correction for discreteness. Note further that we do not require the priors to be proper; we require instead that for sufficiently large data sets, the posterior be proper.

REMARK 11. Furthermore, standard frequentist Edgeworth approximations have the undesirable property that the density approximation need not be positive, and consequently the distribution approximation need not be nondecreasing, or even confined to the interval [0, 1]. Our new application to posteriors shares this drawback. This drawback is related to the inability of the Edgeworth expansion to achieve uniform relative error: the first omitted term is a multiple of a Hermite polynomial, and this polynomial is unbounded for unbounded values of the ordinate.

5. Examples. Consider a random variable having a binomial distribution, $X \sim \text{Bin}(\theta, n)$ with a beta prior, $\theta \sim \text{Beta}(a, b)$. Suppose that $a = 0.5$, $b = 4.0$, $n = 5$ and $x = 2$. This example was previously considered by Weng [50] who, using Stein's identity, derived an asymptotic expansion using up to 40 moments. This is an ideal example for illustrating the

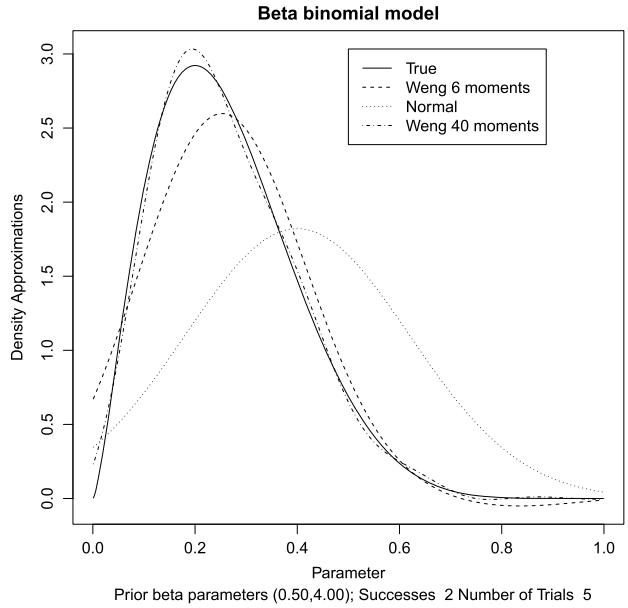

FIG. 1. *Density approximation of Weng* [50].

performance of posterior expansions, because the sample size is small, and the normal approximation is inaccurate, due to the skewness in the posterior. As pointed out by a referee, in this example the renormalized Laplace approximation is exact.

Weng's expansion, for the density of θ, is given in Figure 1. This figure should be compared with Figure 2, showing our posterior Edgeworth approximation for the posterior density of θ. Note that the standard Edgeworth approximation, with four moments, behaves better than the Stein's identity approach using 40 moments. To understand this phenomenon, consider the formal Edgeworth derivation due to Davis [12] and presented by McCullagh [35]. When constructing an approximation of a density f around a baseline density g, obtain the

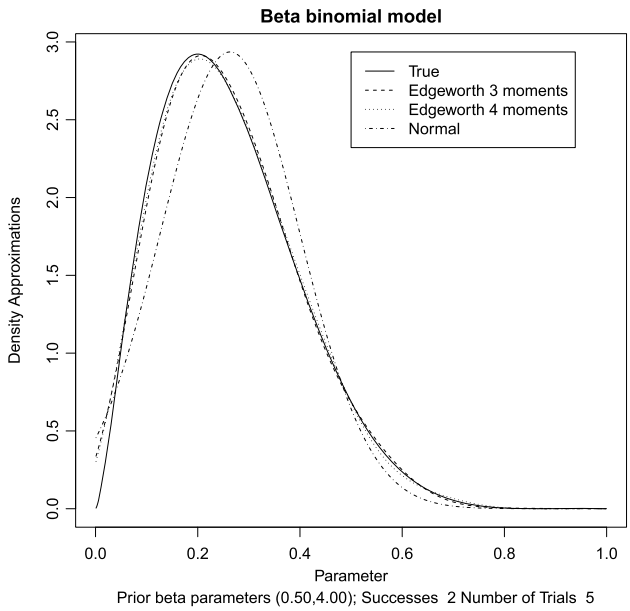

FIG. 2. *Edgeworth approximation of the posterior density.*

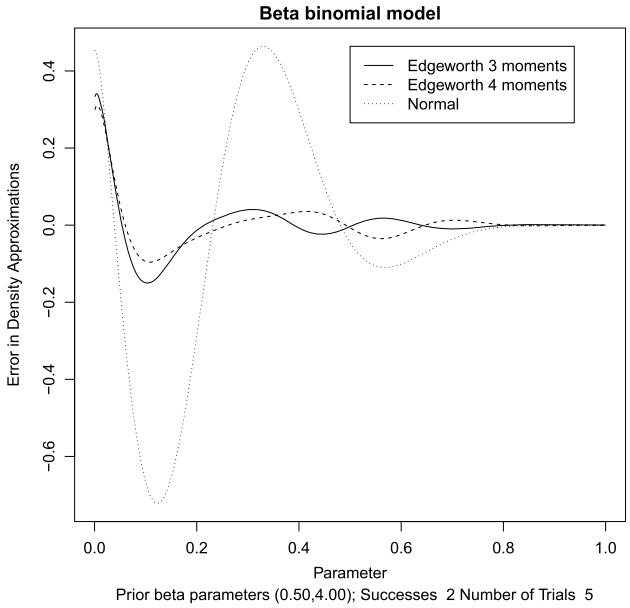

FIG. 3. *Absolute error of density approximations.*

formal series

$$f(x) = g(x) \sum_{j=0}^{\infty} h_j(x) \mu_j^* / j!,$$

where the functions h_j are ratios of derivatives of g to g itself. The coefficients μ_j^* represent the results of calculating differences in cumulants between f and g, and applying the standard relationship giving moments from cumulants to these cumulant differences. In this manuscript, we refer to such coefficients as pseudo-moments. Standard Edgeworth approximation techniques and the method of Weng [50] use as g a normal density; standard Edgeworth approximations use g with mean and variance matching f. When applying Edgeworth techniques to a posterior, then the standard approach is to match the mean and variance. Weng [50] uses instead the maximum likelihood estimate and its usual standard error, and so convergence is slower.

Figure 3 displays the absolute error the normal approximation, and the Edgeworth approximation of the posterior density.

As a second example, we consider inference on failure rate. Proschan [38] presents data on times between failures of airplane air conditioner data, which he argues is well approximated by an exponential distribution with a rate depending on the aircraft. We examine data from plane 1. These six failure times had the sum 493. Below are highest posterior density regions, and normal and Edgeworth approximations to them, for the failure rate (in inverse hours), using a gamma prior with expectation 100 hours and standard deviation 50 hours (see Table 1).

APPENDIX

Here, we provide additional technical details used in the proof of Lemma 1. The exponentiated log posterior of (5) is

$$C \exp\Big(-\delta p_1 - \frac{1}{2}\delta^2 p_2 - \frac{1}{6}\delta^3 p_3 - \frac{1}{24}\delta^4 p_4 - \frac{1}{120}\delta^5 p_5 - \frac{1}{720}\delta^6 p_6 - \frac{1}{5040}\delta^7 p_7 + O(\delta^8) \Big), \tag{16}$$

TABLE 1
Approximations for the failure rate example

Method	Highest posterior density interval	True posterior probability
Normal	(0.01290, 0.02766)	0.7612
Edgeworth	(0.01208, 0.02630)	0.7511
True	(0.01207, 0.02626)	0.7500

for $\delta = \theta - \theta_0$. Expanding the terms in the exponent of order 3 and higher,

$$\begin{aligned}\tilde{p}(\theta) = Ce^{-\frac{1}{2}\delta^2 p_2}\Big(1 &- \delta p_1 + \frac{\delta^2}{2}p_1^2 - \frac{\delta^3}{6}(p_1^3 + p_3)\\ &+ \frac{\delta^4}{24}(p_1^4 + 4p_3p_1 - p_4) - \frac{\delta^5}{120}(p_1^5 + 10p_3p_1^2 - 5p_4p_1 + p_5)\\ &+ \frac{\delta^6}{720}(p_1^6 + 20p_3p_1^3 - 15p_4p_1^2 + 6p_5p_1 + 10p_3^2 - p_6)\\ &- \frac{\delta^7}{5040}(p_1^7 + 35p_3p_1^4 - 35p_4p_1^3 + 21p_5p_1^2\\ &+ (70p_3^2 - 7p_6)p_1 - 35p_3p_4 + p_7)\Big) + E(\delta).\end{aligned} \tag{17}$$

Forcing $\tilde{p}$ to integrate to 1 yields

$$\begin{aligned}C = 24\sqrt{\frac{2}{\pi}}p_2^{7/2}&(p_1^6 + 6p_2p_1^4 + 20p_3p_1^3 + 24p_2^2p_1^2 - 15p_4p_1^2 + 24p_2p_3p_1\\ &+ 6p_5p_1 + 48p_2^3 + 10p_3^2 - 6p_2p_4 - p_6)^{-1}.\end{aligned}$$

The quantity $E(\delta)$ satisfies $|E(\delta)| \le K_1\delta^8$, for K_1 independent of δ. This bound follows from the order of error in (17), and Theorem 2.5.3 of Kolassa [31].

Integrals of (17) times powers of $\sqrt{n}(\theta - \theta_0)$ give moments accurate to $O(n^{-4})$. Let $\tilde{\mu}_j = \int \theta^j \tilde{p}(\theta)\,d\theta$. Hence $\tilde{\mu}_1$, $\tilde{\mu}_2$, $\tilde{\mu}_3$ and $\tilde{\mu}_4$ approximate the first four posterior moments of θ to $O(n^{-2})$.

Integration yields

$$\begin{aligned}\tilde{\mu}_1 = -\frac{C\sqrt{\frac{\pi}{2}}}{24p_2^{9/2}}&(p_1^7 + 6p_2p_1^5 + 35p_3p_1^4 + (24p_2^2 - 35p_4)p_1^3\\ &+ 3(20p_2p_3 + 7p_5)p_1^2 + (48p_2^3 - 30p_4p_2 + 70p_3^2 - 7p_6)p_1\\ &+ 24p_2^2p_3 - 35p_3p_4 + 6p_2p_5 + p_7),\\ \tilde{\mu}_2 = \frac{C\sqrt{\frac{\pi}{2}}}{24p_2^{9/2}}&(7p_1^6 + 30p_2p_1^4 + 140p_3p_1^3 + 3(24p_2^2 - 35p_4)p_1^2\\ &+ 6(20p_2p_3 + 7p_5)p_1 + 48p_2^3 + 70p_3^2 - 30p_2p_4 - 7p_6),\\ \tilde{\mu}_3 = -\frac{C\sqrt{\frac{\pi}{2}}}{8p_2^{11/2}}&(3p_1^7 + 14p_2p_1^5 + 105p_3p_1^4 + 5(8p_2^2 - 21p_4)p_1^3\end{aligned}$$

$$+7(20p_2p_3+9p_5)p_1^2+(48p_2^3-70p_4p_2+210p_3^2-21p_6)p_1$$
$$+5p_3(8p_2^2-21p_4)+14p_2p_5+3p_7),$$
$$\tilde{\mu}_4=\frac{C\sqrt{\frac{\pi}{2}}}{8p_2^{11/2}}(21p_1^6+70p_2p_1^4+420p_3p_1^3+15(8p_2^2-21p_4)p_1^2$$
$$+14(20p_2p_3+9p_5)p_1+48p_2^3+210p_3^2-70p_2p_4-21p_6).$$

Substituting $p_k=ng_k+h_1$, using the rules giving cumulants from moments $\beta_1=\tilde{\mu}_1$, $\beta_2=\tilde{\mu}_2-\tilde{\mu}_1{}^2$ and $\beta_3=\tilde{\mu}_3-2\tilde{\mu}_1\tilde{\mu}_2+2\tilde{\mu}_1{}^3$ and $\beta_4=\tilde{\mu}_4-3\tilde{\mu}_2^2-6\tilde{\mu}_1^4-4\tilde{\mu}_3\tilde{\mu}_1+12\tilde{\mu}_1^2\tilde{\mu}_2$, give

$$\beta_1=-\frac{g_1}{g_2}+\frac{g_1h_2-g_2h_1}{g_2^2n}+O\left(\frac{1}{n^2}\right).$$

The choice of θ_0 as the posterior mean implies that $g_1=0$ and, therefore,

$$\beta_2=\frac{1}{2ng_2}+O\left(\frac{1}{n^2}\right),$$
$$\beta_3=-\frac{g_3}{6g_2^3n^2}+O\left(\frac{1}{n^3}\right),$$
$$\beta_4=\frac{3g_3^2-g_2g_4}{24g_2^5n^3}+O\left(\frac{1}{n^4}\right).$$

Acknowledgments. The second author is grateful to J. K. Ghosh, Jens Jensen, Trevor Sweeting and Alastair Young for helpful correspondence and discussion related to this topic during 2010–2011.

The first author was supported in part by National Science Foundation Grant DMS-1712839.

The second author was supported in part by National Science Foundation Grant DMS-1712940.

REFERENCES

[1] BARNDORFF-NIELSEN, O. and COX, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications. *J. Roy. Statist. Soc. Ser. B* **41** 279–312. MR0557595

[2] BARNDORFF-NIELSEN, O. E. and COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR1010226 https://doi.org/10.1007/978-1-4899-3424-6

[3] BERTAIL, P. and LO, A. Y. (2002). On Johnson's asymptotic expansion for a posterior distribution. Technical Report No. 2002-39, Institut National de la Statistique et des Estudes Economiques, Série des Documents de Travail du CREST (Centre de Recherche en Economie et Statistique).

[4] BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. MR0471142

[5] BHATTACHARYA, R. N. and RAO, R. R. (2010). *Normal Approximation and Asymptotic Expansions*. *Classics in Applied Mathematics* **64**. SIAM, Philadelphia, PA. MR3396213 https://doi.org/10.1137/1.9780898719895.ch1

[6] BICKEL, P. J. and GHOSH, J. K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction—A Bayesian argument. *Ann. Statist.* **18** 1070–1090. MR1062699 https://doi.org/10.1214/aos/1176347740

[7] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR1324786

[8] BRILLINGER, D. R. (1969). The calculation of cumulants via conditioning. *Ann. Inst. Statist. Math.* **21** 215–218.
[9] CHANG, I. H. and MUKERJEE, R. (2006). Probability matching property of adjusted likelihoods. *Statist. Probab. Lett.* **76** 838–842. MR2266098 https://doi.org/10.1016/j.spl.2005.10.015
[10] DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability*. *Springer Texts in Statistics*. Springer, New York. MR2664452
[11] DATTA, G. S. and MUKERJEE, R. (2004). *Probability Matching Priors*: *Higher Order Asymptotics*. *Lecture Notes in Statistics* **178**. Springer, New York. MR2053794 https://doi.org/10.1007/978-1-4612-2036-7
[12] DAVIS, A. W. (1976). Statistical distributions in univariate and multivariate Edgeworth populations. *Biometrika* **63** 661–670. MR0518636 https://doi.org/10.1093/biomet/63.3.661
[13] DAVISON, A. C. (1986). Approximate predictive likelihood. *Biometrika* **73** 323–332. MR0855892 https://doi.org/10.1093/biomet/73.2.323
[14] DICICCIO, T. J. and STERN, S. E. (1993). On Bartlett adjustments for approximate Bayesian inference. *Biometrika* **80** 731–740. MR1282782 https://doi.org/10.1093/biomet/80.4.731
[15] DICICCIO, T. J. and STERN, S. E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *J. Roy. Statist. Soc. Ser. B* **56** 397–408. MR1281943
[16] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, *Vol. II*, 2nd ed. Wiley, New York. MR0270403
[17] GHOSH, J. K. (1994). *Higher-Order Asymptotics*. IMS, Hayward.
[18] GHOSH, J. K., DELAMPADY, M. and SAMANTA, T. (2006). *An Introduction to Bayesian Analysis*: *Theory and Methods*. *Springer Texts in Statistics*. Springer, New York. MR2247439
[19] GHOSH, J. K. and RAMAMOORTHI, R. V. (2003). *Bayesian Nonparametrics*. *Springer Series in Statistics*. Springer, New York. MR1992245
[20] GHOSH, J. K., SINHA, B. K. and JOSHI, S. N. (1982). Expansions for posterior probability and integrated Bayes risk. In *Statistical Decision Theory and Related Topics*, *III*, *Vol.* 1 (*West Lafayette*, *Ind.*, 1981) (S. Gupta and J. Berger, eds.) 403–456. Academic Press, New York. MR0705299
[21] HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7
[22] HARTIGAN, J. A. (1965). The asymptotically unbiased prior distribution. *Ann. Math. Stat.* **36** 1137–1152. MR0176539 https://doi.org/10.1214/aoms/1177699988
[23] JAMES, G. S. (1955). Cumulants of a transformed variate. *Biometrika* **42** 529–531. MR0071681 https://doi.org/10.1093/biomet/42.3-4.529
[24] JAMES, G. S. (1958). On moments and cumulants of systems of statistics. *Sankhyā* **20** 1–30. MR0103550
[25] JAMES, G. S. and MAYNE, A. J. (1962). Cumulants of functions of random variables. *Sankhyā*, *Ser. A* **24** 47–54. MR0148183
[26] JENSEN, J. L. (1995). *Saddlepoint Approximations*. *Oxford Statistical Science Series* **16**. Oxford University Press, New York. MR1354837
[27] JOHNSON, R. A. (1967). An asymptotic expansion for posterior distributions. *Ann. Math. Stat.* **38** 1899–1906. MR0219161 https://doi.org/10.1214/aoms/1177698624
[28] JOHNSON, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Stat.* **41** 851–864. MR0263198 https://doi.org/10.1214/aoms/1177696963
[29] KASS, R. E., TIERNEY, L. and KADANE, J. P. (1990). The validity of posterior expansions based on Laplace's method. In *Bayesian and Likelihood Methods in Statistics and Econometrics* (S. Geisser, J. S. Hodges, S. J. Press and A. Zellner, eds.) 473–488. Elsevier Science Publishers B.V., Amsterdam.
[30] KHARROUBI, S. A. and SWEETING, T. J. (2016). Exponential tilting in Bayesian asymptotics. *Biometrika* **103** 337–349. MR3509890 https://doi.org/10.1093/biomet/asw018
[31] KOLASSA, J. E. (2006). *Series Approximation Methods in Statistics*, 3rd ed. *Lecture Notes in Statistics* **88**. Springer, New York. MR1295242 https://doi.org/10.1007/978-1-4757-4275-6
[32] LEONOV, V. P. and SIRJAEV, A. N. (1959). On a method of semi-invariants. *Theory Probab. Appl.* **4** 319–329. MR0123345 https://doi.org/10.1137/1104031
[33] LINDLEY, D. V. (1961). The use of prior probability distributions in statistical inference and decisions. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, *Vol. I* 453–468. Univ. California Press, Berkeley, CA. MR0156437
[34] LINDLEY, D. V. (1980). Approximate Bayesian methods. *Trab. Estad. Investig. Oper.* **31** 223–237.
[35] MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR0907286
[36] MYKLAND, P. A. (1999). Bartlett identities and large deviations in likelihood theory. *Ann. Statist.* **27** 1105–1117. MR1724043 https://doi.org/10.1214/aos/1018031270
[37] PERICCHI, L. R., SANSÓ, B. and SMITH, A. F. M. (1993). Posterior cumulant relationships in Bayesian inference involving the exponential family. *J. Amer. Statist. Assoc.* **88** 1419–1426. MR1245379

[38] PROSCHAN, F. (1963). Theoretical explanation of bbserved decreasing failure rate. *Technometrics* **5** 375–383.
[39] REID, N. (2003). Asymptotics and the theory of inference. *Ann. Statist.* **31** 1695–1731. MR2036388 https://doi.org/10.1214/aos/1074290325
[40] RULI, E., SARTORI, N. and VENTURA, L. (2014). Marginal posterior simulation via higher-order tail area approximations. *Bayesian Anal.* **9** 129–145. MR3188302 https://doi.org/10.1214/13-BA851
[41] SPEED, T. P. (1983). Cumulants and partition lattices. *Aust. J. Stat.* **25** 378–388. MR0725217
[42] STUART, A. and ORD, J. K. (1994). *Kendall's Advanced Theory of Statistics*, *Vol.* 1, 6th ed. Wiley, New York. MR1280717
[43] SWEETING, T. J. (1995). A framework for Bayesian and likelihood approximations in statistics. *Biometrika* **82** 1–23. MR1332836 https://doi.org/10.1093/biomet/82.1.1
[44] TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86. MR0830567
[45] TIERNEY, L., KASS, R. E. and KADANE, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Statist. Assoc.* **84** 710–716. MR1132586
[46] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256
[47] VENTURA, L. and REID, N. (2014). Approximate Bayesian computation with modified log-likelihood ratios. *Metron* **72** 231–245. MR3233151 https://doi.org/10.1007/s40300-014-0041-4
[48] WALLACE, D. L. (1958). Asymptotic approximations to distributions. *Ann. Math. Stat.* **29** 635–654. MR0097109 https://doi.org/10.1214/aoms/1177706528
[49] WENG, R. C. (2003). On Stein's identity for posterior normality. *Statist. Sinica* **13** 495–506. MR1977739
[50] WENG, R. C. (2010). A Bayesian Edgeworth expansion by Stein's identity. *Bayesian Anal.* **5** 741–763. MR2740155 https://doi.org/10.1214/10-BA526
[51] WENG, R. C. and TSAI, W.-C. (2008). Asymptotic posterior normality for multiparameter problems. *J. Statist. Plann. Inference* **138** 4068–4080. MR2455988 https://doi.org/10.1016/j.jspi.2008.03.034
[52] WITHERS, C. S. (1982). Second order inference for asymptotically normal random variables. *Sankhyā*, *Ser. B* **44** 19–27. MR0692891
[53] WITHERS, C. S. (1984). Asymptotic expansions for distributions and quantiles with power series cumulants. *J. Roy. Statist. Soc. Ser. B* **46** 389–396. MR0790623
[54] WOODROOFE, M. (1989). Very weak expansions for sequentially designed experiments: Linear models. *Ann. Statist.* **17** 1087–1102. MR1015139 https://doi.org/10.1214/aos/1176347257
[55] WOODROOFE, M. (1992). Integrable expansions for posterior distributions for one-parameter exponential families. *Statist. Sinica* **2** 91–111. MR1152299