8 Lecture 8: Higher order asymptotics

8.1 Motivation

General results in estimation and in other inference procedures can usually be defended on asymptotic grounds only (i.e., when the sample size n tends to infinity). An inference procedure that is optimal for a finite sample size is very rarely possible to construct and it depends heavily on the specific distributional assumptions about the sample that has given rise to the data. Hence such a procedure is too much individually tailored and can not be offered as a general methodological tool to be used in situations where the distributional assumptions have been changed. General inference procedures could be offered on asymptotic grounds only and we have seen how useful they can be when discussing the MLE, for example.

Most useful in statistical practice are the so-called **first-order** asymptotic results. They state that under some regularity conditions asymptotic normality of certain estimator holds (or that under regularity conditions, the null distribution of a certain test-statistic is asymptotically normal or chi-squared etc.) We have already discussed many such results in Lectures 5 and 6.

The techniques used to show such results are usually a combination of central limit theorem (CLT) and Taylor expansions. Again, in our Lecture 5 about properties of MLE we have seen these techniques demonstrated at work. The final product of such results is the statement about asymptotic normality of a suitably normalized statistic. However, sometimes, the sample sizes used in practice are not large enough to warrant that the accuracy achieved by the asymptotic normal approximation is precise enough. Then it is worth trying to include **higher order expansions** for the distribution of the statistic of interest hoping that these more complex expressions will bring about a better approximation for the distribution when the sample size is not as large.

To take an example to illustrate our point, assume that we are dealing with the distribution of the sample mean $\bar{X}$ taken from a (not necessarily normal) population with finite mean μ and variance σ^2. Let $Z_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ and $F_n(z) = P(Z_n \leq z)$. Then the CLT says that $F_n(z) \to_{n\to\infty} \Phi(z)$ for every $z \in R$. If we also assume that a third finite moment γ exists then the famous Berry-Esseen theorem states that

$$|F_n(z) - \Phi(z)| = O(\frac{1}{\sqrt{n}})$$

uniformly in z where the bound on the right hand side depends on the absolute third moment. More precise statement can be obtained in the form

$$|F_n(z) - \Phi(z) - \frac{C_1(F)p_1(z)\phi(z)}{\sqrt{n}}| = O(\frac{1}{n}) = o(\frac{1}{\sqrt{n}})$$

uniformly in z where $\phi(z)$ is the density of the standard normal, C_1 is a suitably chosen constant and $p_1(z)$ is certain first degree polynomial. Expansions in the form

$$F_n(z) = \Phi(z) + \sum_{s=1}^{k} \frac{q_s(z)}{n^{s/2}} + o(n^{-k/2})$$

can often be obtained for the distribution $F_n(.)$ of suitably normalized statistics of interest and are called **Edgeworth expansions.** Here $q_s(z)$ are some polynomials multiplied by the standard normal density. The coefficients of the polynomials depend on the cumulants of $F(z)$ and the cumulants are related to the moments of the distribution $F(.)$. In such a way, going beyond the normal expansion given by the CLT, the effects of skewness, kurtosis and of higher order moments on the approximation of the statistic Z_n can be captured.

8.2 Moments and cumulants

As is known, the moment generating function (**MGF**) of a random variable X is defined as $M_X(t) = E[\exp(tX)]$ whenever the expectation is finite. It is obvious that always $M_X(0) = 1$ holds. The MGF may only be defined in a small neigbourhood of 0 (or even only for $t = 0$). If it exists in an open interval around 0 then all moments of the random variable exist and can be obtained via $E(X^r) = M_X^{(r)}(t)_{|t=0} = \mu'_r$. We then have the Taylor expansion

$$M_X(t) = 1 + \mu'_1 t + \mu'_2 \frac{t^2}{2!} + \cdots + \mu'_r \frac{t^r}{r!} + O(t^{r+1})$$

when $t \to 0$. In addition, we define the cumulant generating function (**CGF**) of the random variable X:

$$K_X(t) = \log[M_X(t)],$$

(defined on the same interval as the MGF). The Taylor series expansion

$$K_X(t) = \kappa_1 t + \kappa_2 \frac{t^2}{2!} + \cdots + \kappa_r \frac{t^r}{r!} + O(t^{r+1})$$

when $t \to 0$ holds in this interval and defines the **cumulants** $k_s, s = 1, 2, \ldots$ of the distribution of X. Clearly $\kappa_r = K_X^{(r)}(t)_{|t=0}$ holds.

Moreover, by equating the coefficients in the expansions of $\exp[K_X(t)]$ and of $M_X(t)$, the relationships between moments and cumulants can be established. In particular, the relationships $\mu'_1 = \kappa_1 = E(X), \kappa_2 = Var(X) = \mu'_2 - (\mu'_1)^2$ and $\kappa_3 = 2\mu'^3_1 - 3\mu'_1\mu'_2 + \mu'_3$ hold. The `mathStatica` functions `CumulantToRaw` and `RawToCumulant` can be used to express analytically the conversions between moments and cumulants and vice versa (consult Section 2.4G of the book).

Per **definition** the third cumulant is the skewness and the fourth cumulant is the curtosis of the distribution of X. To summarise:

- the first cumulant is the first moment
- the second cumulant is the variance
- the third cumulant is the skewness
- the fourth cumulant is the kurtosis

of the distribution.

The reason we introduce the cumulants in our discussion is that although the moments are a more familiar concept from your introductory statistics lectures, the higher order asymptotic expansions we are about to discuss in this lecture, can in fact be presented in more compact form by using the cumulants.

Exercise 1. Show that if $X_1, X_2, \ldots, X_n$ are i.i.d. then $K_{\sum_{i=1}^n X_i}(t) = nK_{X_1}(t)$ holds. Also, for any constants a, b we have $K_{aX_1+b}(t) = bt + K_{X_1}(at)$.

8.3 Asymptotic expansions

We should explicitly stress here that the Edgeworth type expansions are representative of what we call *asymptotic expansions* in statistics. They arise in the following general way. We want to represent a set of functions $f_n(z)$ (indexed by $n = 1, 2, \ldots$) in the form

$$f_n(z) = \gamma_0(z)b_{0,n} + \gamma_1(z)b_{1,n} + \cdots + \gamma_k(z)b_{k,n} + o(b_{k,n}) \tag{13}$$

The expansion is considered as $n \to \infty$. Typical choices of the doubly indexed coefficients $b_{k,n}$ are $\{1, n^{-1/2}, n^{-1}, \ldots, n^{-k/2}\}$ or $\{1, n^{-1}, n^{-2}, \ldots, n^{-k}\}$. The idea is to have contributions progressively falling in order with sample size and essentially requesting for this effect the condition $b_{r+1,n} = o(b_{r,n})$ as $n \to \infty$ for each $r = 0, 1, 2, \ldots, k-1$. In particular examples of asymptotic expansions, the function $f_n(z)$ may be a density or a CDF of a statistic calculated for a sample size of n at a point z on the real axis. The $\gamma_0(z)$ would be the density of the CDF of the standard normal depending on the context.

It is important to notice that the above expansion **is not necessarily a convergent series** for $f_n(z)$ for any fixed z! This means that if n is fixed, just increasing the number of terms k does not necessarily lead to convergence and to improved approximation of $f_n(z)$. One has to increase the sample size, too.

8.3.1 Edgeworth expansion for cdf

A major condition to deriving Edgeworth expansions of the form (13) is the **Cramér condition**. A cdf $F(.)$ on the real line is said to satisfy Cramér's condition if for the characteristic function $\chi_F(t) = E_F(e^{itX})$

$$\limsup_{t\to\infty} |E_F(e^{itX})| < 1 \tag{14}$$

holds. This condition is needed to deal with some more singular cases. It can be shown that *all* continuous distributions do satisfy Cramér's condition. On the other hand, the lattice distributions (whose whole mass is concentrated on a lattice in the form

$$a + bn, n \text{ integer}; a, b \text{ constants})$$

do *not*. An easy intuitive explanation for this phenomenon is that for lattice distributions, the cdf of $F_n(.)$ would have jump discontinuities for any fixed sample size n whereas the Edgeworth expansion is a smooth function and accuracy of the type shown in (13) can not be expected to hold. To give more explicit expression for the terms involved in the Edgeworth expansion, we will define it here for the particular case of the CDF (not the density) and $k = 2$.

Theorem 8.1. *Suppose that $F(.)$ satisfies Cramér's condition and $E_F(X^4) < \infty$ holds. Then*

$$F_n(z) = \Phi(z) - \frac{C_1(F)p_1(z)\phi(z)}{\sqrt{n}} - \frac{C_2(F)p_2(z) + C_3(F)p_3(z)}{n}\phi(z) + o(n^{-1}). \tag{15}$$

Here $C_1(F) = \frac{E(X-\mu)^3}{6\sigma^3}$ (skewness correction), $C_2(F) = \frac{\frac{E(X-\mu)^4}{\sigma^4} - 3}{24}$ (kurtosis correction), $C_3(F) = \frac{C_1^2(F)}{2}$, $p_1(z) = z^2 - 1$, $p_2(z) = z^3 - 3z$, $p_3(z) = z^5 + 15z - 10z^3$. The polynomials p_1, p_2 and p_3 are the second, third and fifth Hermite polynomials $H_j, j = 2, 3, 5$. Their general definition is given by $H_j(z) = \frac{(-1)^j \phi^{(j)}(z)}{\phi(z)}, j = 0, 1, 2, \ldots$ where $\phi^{(j)}(z)$ is the j-th order derivative of the standard normal density.

About the proof: The proof is subtle and here we only sketch the idea behind it. The main ingredient in proving such result is *Esseen's smoothing lemma*. This lemma bounds the closeness of a cdf and of its approximation by closeness of their Fourier transforms. For a cdf F and a differentiable function G with G' uniformly bounded and such that $\int |F(t) - G(t)|dt < \infty$, Esseen's lemma tells us that for each $T > 0$

$$\sup_z |F(z) - G(z)| \leq \frac{1}{\pi}\int_{-T}^{T} |\frac{\chi_F(t) - \chi_G(t)}{t}|dt + \frac{A}{T}\sup_t |G'(t)| \tag{16}$$

where A is an absolute constant and $\chi_F(t) = \int_{-\infty}^{\infty} e^{itx}dF(x), \chi_G(t) = \int_{-\infty}^{\infty} e^{itx}dG(x)$.

If we set

$$G(z) = \Phi(z) - \frac{C_1(F)p_1(z)\phi(z)}{\sqrt{n}} - \frac{C_2(F)p_2(z) + C_3(F)p_3(z)}{n}\phi(z)$$

and $F = F_n$ to be the CDF of Z_n then it is possible to control from above the difference $\chi_F(t) - \chi_G(t)$ on the right hand side of (16) thus getting control from above of the left hand side.

The control on the difference $\chi_F(t) - \chi_G(t)$ is possible to achieve for the following reason. Since $EZ_n = 0$ and $Var(Z_n) = 1$, one can show that starting with the expansion

$$E(e^{itZ_n}) = 1 + remainder \tag{17}$$

with remainder involving the moments of Z_n, using the fact that $\log(1+z)$ can be approximated by z for small z, we can first take logs of both sides in (17), apply the approximation and then exponentiate back to get

$$E(e^{itZ_n}) = \exp\{n\sum_{j=2}^{4} \frac{(it/\sqrt{n})^j \rho_j}{j!} + \frac{t^4}{n}\theta(t, n)\} \tag{18}$$

where $\sup_{n\geq 1}\sup_{|t|\leq \epsilon\sqrt{n}} |\theta(t, n)| \to 0$ as $\epsilon \to 0$ and $\rho_j = \frac{\kappa_j}{\kappa_2^{j/2}}$ are the *standardised cumulants*. We see that the leading term in the summation in the exponent (for $j = 2$) is $-t^2/2$. Hence the RHS of (18) is in the form

$$\exp(-t^2/2)\{1 + (...)/\sqrt{n} + (..)/n + \text{ remainder }\}$$

and the remainder is "under control". When we calculate $\chi_G(t) = \int e^{itx} dG(x)$ we get again

$$\exp(-t^2/2)\{1 + (...)/\sqrt{n} + (..)/n\}$$

with the same expressions in the brackets (..) in front of the $n^{-1/2}$ and n^{-1} powers.

At this point we can apply the smoothing lemma. We apply it with $T = n^4$, however when evaluating the RHS we subdivide the integral part into two regions. More precisely, we take $T_1 = \epsilon\sqrt{n}$ for a suitably small but fixed $\epsilon > 0$ and evaluate the integral separately for regions $t \in [-T_1, T_1]$ and for $T_1 \leq |t| \leq T$. For the evaluation in $T_1 \leq |t| \leq T$ we need Cramér's condition. The details of these evaluations are subtle and are omitted. The final evaluation from above of the right hand side helps us to bound the left hand side:

$$\sup_z |F(z) - G(z)| = o(n^{-1}).$$

8.3.2 Edgeworth expansion for the density of Z_n

Formally, by differentiating both sides in the the expansion from Theorem 8.1 we can get the two term expansion for the density of $f_{Z_n}(z)$ of Z_n. The final result is:

$$\phi(z)\{1 + \frac{\rho_3}{6\sqrt{n}} H_3(z) + \frac{1}{n}[\frac{\rho_4 H_4(z)}{24} + \frac{\rho_3^2 H_6(z)}{72}]\} + o(n^{-1}) \tag{19}$$

and the expansion (19) holds uniformly in $z \in R^1$. Here we have denoted by $\rho_r = \kappa_r/\kappa_2^{r/2}$ the *standardised cumulants*. As you can easily convince yourself, $H_4(z) = z^4 - 6z^2 + 3, H_6(z) = z^6 - 15z^4 + 45z^2 - 15$ holds.

The interpretation of (19) is that the leading term is the standard normal density in lieu with the Central Limit Theorem. Then there are higher order correction terms whose relevance becomes important for small to moderate sample sizes. The correction terms represent simultaneous adjustment of the normal approximation by using the information for the standardised skewness and kurtosis. It is also interesting to note that if the distribution of X is symmetric (hence $\rho_3 = 0$) the normal approximation is more accurate in the sense that the correction term is then of order n^{-1} rather than the $n^{-1/2}$ when $\rho_3 \neq 0$. Other important observation to be made is that the accuracy of the approximation depends on the position of the argument z. When we are in the tails (i.e., for $|z|$ large enough, the Edgeworth approximation (19) may worsen quite a lot and may become even negative. The same observation holds also for (15) which may become either negative or bigger than one in the tails. The saddlepoint approximation to be discussed next is meant to improve upon Edgeworth especially in the tails.

8.3.3 Cornish-Fisher expansions for quantiles

One major application of the Edgeworth expansion in practice is in fact in "reversing" it to construct better confidence intervals for small to moderate sample sizes. When studying coverage probabilities of confidence intervals, we need to solve, given confidence level α, equations of the type $F_{Z_n}(z_\alpha) = 1 - \alpha$ as accurately as possible. We know that Z_n is "close to standard normally distributed" and hence believe, not unreasonably, that z_α is in a vicinity of $u_\alpha : \Phi(u_\alpha) = 1 - \alpha$. If an Edgeworth expansion of the cdf $F_{Z_n}(.)$ has been

obtained already, one can of course try to replace $F_{Z_n}(z_\alpha)$ by its Edgeworth expansion and then equate this Edgeworth expansion to $\Phi(u_\alpha)$. Solving the resulting equation w.r. to the argument (that is, "inverting" it), gives an approximation for z_α. This inversion will involve the powers of u_α and is called *Cornish-Fisher* expansion η_α. The precise result (again if we contain ourselves to order $o(n^{-1})$ only) is given as follows:

Theorem 8.2. *Let $F(.)$ be the cdf of a single observation having a finite moment generating function in a neighbourhood of 0 and satisfying Cramér's condition. Then*

$$\eta_\alpha = u_\alpha + \frac{(u_\alpha^2-1)\rho_3}{6\sqrt{n}} + \frac{(u_\alpha^3-3u_\alpha)\rho_4}{24n} - \frac{(2u_\alpha^3-5u_\alpha)\rho_3^2}{36n} + o(n^{-1})$$

holds.

The accuracy of the above two-term Cornish-Fisher expansion is usually quite impressive! One numerical illustration follows now.

Example . Assume that $W_n \sim \chi_n^2$ is chi-squared with n d.f. Then we know that per definition. W_n/n can be represented as an average $\bar{X^2}$ of n i.i.d. squared standard normals $X_i, i = 1, 2, \ldots, n$. You can calculate the leading cumulants of the square of a standard normal directly (or you can help yourself with `mathStatica` as illustrated in the handout). You get for the standardised cumulants: $\rho_3 = 2\sqrt{2}, \rho_4 = 12$. Since $\sqrt{n}(W_n/n - 1)/\sqrt{2} = (W_n - n)/\sqrt{2n}$ is about standard normal by the central limit theorem, the first order approximation of the α -quantile of W_n would be just $n + u_\alpha\sqrt{2n}$. The second order approximation using Theorem above will be

$$n + \sqrt{2n}[u_\alpha + \frac{(u_\alpha^2-1)2\sqrt{2}}{6\sqrt{n}} + \frac{(u_\alpha^3-3u_\alpha)12}{24n} - \frac{(2u_\alpha^3-5u_\alpha)8}{36n}] =$$

$$n + \sqrt{2n}u_\alpha + \frac{(u_\alpha^2-1)2}{3} + \frac{(u_\alpha^3-7u_\alpha)}{9\sqrt{2n}}$$

For $\alpha = 0.01$ we have $u_\alpha = 2.326$. Applying the Central Limit Theorem-based approximation $n+u_\alpha\sqrt{2n}$ for $n = 5$ gives 12.36. The first order approximation $n+\sqrt{2n}[u_\alpha+\frac{(u_\alpha^2-1)2\sqrt{2}}{6\sqrt{n}}]$ gives 15.296 whereas the second order approximation delivers 15.07. The true 99th percentile of W_n for $n = 5$ is known to be 15.09. Thus clearly in this case, the CLT-based approximation is much poorer and the two Cornsih-Fisher approximations are quite good, with the second order approximation "almost hitting" the true value.

8.3.4 Edgeworth expansions for other statistics

Our discussion was related to Edgeworth expansion for the normalised distribution of the sample mean because this is the easiest case to discuss. However, from practical point of view, Edgeworth expansions are of interest also for many other statistics. In particular, Edgeworth expansions for the MLE are of particular interest and can be obtained.

8.3.5 Saddlepoint density approximation for the mean

As mentioned already, the Edgeworth expansion of the density of Z_n may not be very accurate particularly in the tails where it could even become negative. One explanation

for this phenomenon is that the error approximation in the Edgeworth expansion is only *absolute* instead of *relative* and in the tails, where the true density is small to start with, even a small absolute error may turn out to be quite large relative to the true density's value and can cause serious error of the approximation. This is the reason to look for alternative expansions where the error of approximation is actually relative. The *Saddlepoint approximation method* offers such alternatives. It provides very accurate numerical approximations for densities and tail areas of statistics of interest down to surprisingly small sample sizes such as 5 or 10 far out in the tails.

There are several approaches to introduce the saddlepoint approximation. The historically first one is linked to the original paper by Henry Daniels in *The Annals of Mathematical Statistics* in 1954. This approach also explains the name of the method. It is based on Fourier inversion of the moment generating function and trying to approximate this inversion (represented as integral over the imaginary axis) by exploiting the biggest contribution to the integral coming form a small region around the *saddlepoint* of the integrand.

However, it is our feeling that the above approach is more difficult to comprehend. The second approach seems to be easier and leads to the same result so we only outline this second approach here. We start with the observation that since $Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$, if we use the density transformation formula, we can rewrite (19) in terms of the variable $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ as follows:

$$f_{\bar{X}}(\bar{x}) = \frac{\sqrt{n}}{\sigma}\phi(z)\{1 + \frac{\rho_3}{6\sqrt{n}}H_3(z) + \frac{1}{n}[\frac{\rho_4 H_4(z)}{24} + \frac{\rho_3^2 H_6(z)}{72}] + o(n^{-1})\} \tag{20}$$

where $z = \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}$. Because at the centre $z = 0$ the Hermite polynomials are $H_3(0) = 0, H_4(0) = 3, H_6(0) = -15$ we can see that the approximation (20) is in fact **more accurate** when $\bar{x} = \mu$, i.e., at the mean, since then the $O(\frac{1}{\sqrt{n}})$ correction term disappears. The explicit expression for the approximation there is:

$$f_{\bar{X}}(\bar{x}) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\{1 + \frac{1}{n}(\frac{1}{8}\rho_4 - \frac{5}{24}\rho_3^2) + O(n^{-2})\}. \tag{21}$$

The idea is then, if we want similar better accuracy for other values of $\bar{x} \neq \mu$, to *tilt* the distribution (apply the so-called *exponential tilting*≡ Esscher transform) to shift the mean so that $\bar{x}$ coincides with the new mean.

More specifically, we notice first of all that for any random variable X with a density $f(x)$ and a cumulant generating function $K_X(t)$ that converges on an open neighbourhood A of the point 0, we can define the tilted density family $f(x;t) = e^{tx-K_X(t)}f(x), t \in A$. This means that in fact we are embedding the original density $f(x)$ in a one-parameter exponential family parameterized (indexed) by t. Of course, we can also write

$$f(x) = e^{K_X(t)-tx}f(x;t) \text{ for any } t \in A \tag{22}$$

thus allowing us a choice of a suitable $t \in A$ when representing $f(x)$.

We can also easily see (DO IT (!)) that if X_t denotes a random variable with the density $f(x;t)$ then for the its mean, variance and standardised cumulants we have:

$$E[X_t] = K'_X(t), Var[X_t] = K''_X(t), \rho_i(t) = \frac{K_X^{(i)}(t)}{K''_X(t)^{i/2}}, i \geq 3. \tag{23}$$

Our strategy to derive the saddlepoint approximation then is as follows:

Step one: Given x, choose a suitable $\hat{t}$ such that the Edgeworth expansion for $f(x;\hat{t})$ is most accurate (and this means, as seen, that it is applied at the mean of $X_{\hat{t}}$).

Step two: Transform back to calculate $f(x)$ using the formula (22) by replacing t by $\hat{t}$ there.

To implement Step one, means that the fixed value of the argument x has to be the mean of X_t, that is,

$$E[X_t] = K'_X(\hat{t}) = x \tag{24}$$

The equation (24) is called the saddlepoint equation. Then we apply the Edgeworth expansion at the mean as given in (21). This gives

$$f(x;\hat{t}) \approx \frac{1}{\sqrt{2\pi K''_X(\hat{t})}}\{1 + [\frac{1}{8}\rho_4(\hat{t}) - \frac{5}{24}\rho_3^2(\hat{t})]\}.$$

Then in the second step we substitute the above approximation in the right hand side of (22) to get the approximation:

$$f(x) = e^{\{-\hat{t}x + K_X(\hat{t})\}} f(x;\hat{t}) \approx \frac{1}{\sqrt{2\pi K''_X(\hat{t})}} e^{[K_X(\hat{t}) - \hat{t}x]}\{1 + [\frac{1}{8}\rho_4(\hat{t}) - \frac{5}{24}\rho_3^2(\hat{t})]\}. \tag{25}$$

The discussion until now involved essentially *any* random variable X. The asymptotic point of view is brought forward when (25) is applied to the random variable $\bar{X}$. The density $f(\bar{x})$ of $\bar{X}$ is tilted in the exponential family

$$f(\bar{x};t) = e^{nt\bar{x} - nK_X(t)} f(\bar{x})$$

with $\bar{X}_t \sim f(\bar{x};t)$. The mean of $\bar{X}_t$ is, of course, just $K'_X(t)$ and the optimal tilt is obtained by solving $K'_X(\hat{t}) = \bar{x}$. It is easily seen that the standardised cumulants for $\bar{X}_{\hat{t}}$ are just $\hat{\rho}_i = n^{1-i/2}\rho_i(\hat{t})$, where $\rho_i(t)$ are given in (23). Hence we get

$$\hat{f}(\bar{x}) \approx \sqrt{\frac{n}{2\pi K''_X(\hat{t})}} e^{\{nK_X(\hat{t}) - n\hat{t}\bar{x}\}}\{1 + [\frac{1}{8n}\hat{\rho}_4 - \frac{5}{24n}\hat{\rho}_3^2]\}. \tag{26}$$

The above approximation (26) is called the *second order saddlepoint density approximation* and is *extremely accurate*! Sometimes, for the sake of numerical simplicity, even just the *first order saddlepoint approximation*

$$\hat{f}(\bar{x}) \approx \sqrt{\frac{n}{2\pi K''_X(\hat{t})}} e^{\{nK_X(\hat{t}) - n\hat{t}\bar{x}\}} \tag{27}$$

is used and it is still very accurate!

Final note: The saddlepoint approximation (26) is quite different from the Edgeworth expansion. It is an asymptotic expansion on powers of n^{-1} instead of $n^{-1/2}$, as in the Edgeworth. This implies that already the simple approximation (27) has already absorbed the skewness correction. The leading term (27) is clearly *not* the normal (and in fact any other) density; it may not integrate to one (although it can be renormalised to integrate to one). The saddlepoint approximation is more accurate than Edgeworth, especially in

the tails. These advantages are achieved due to the fact that for the derivation of the saddlepoint, we used the *whole cumulant generating function* of the distribution of X_1 whereas for Edgeworth expansion we only need the four leading cumulants. Also, the saddlepoint method is computationally more intensive: the saddlepoint equation $K'_X(\hat{t}) = \bar{x}$ has to be solved for each value of the argument $\bar{x}$. To summarise, via the saddlepoint approximation "by requiring more we achieve more".

8.3.6 Saddlepoint for CDF and examples

First, we note that a very accurate asymptotic approximation exists also for the cumulative distribution function (CDF) of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ where $K_X(t)$ is the cumulant generating function of X_1. This approximation is derived via two layers of approximation: first the density is approximated via (27). The CDF is approximated as integral of this approximated density then. Finally, the resulting integral is approximated using the *Temme approximation*. The latter represents an approximate integration by parts formula. We omit the details and only give the final formula. It is called the *Lugannani-Rice* formula:

$$P(\bar{X} \leq \bar{x}) = \Phi(\hat{w}_n) + \phi(\hat{w}_n)[\frac{1}{\hat{w}_n} - \frac{1}{\hat{u}_n}] + O(\frac{1}{n})(\text{ for } \bar{x} \neq E(\bar{X}).$$

Here $\hat{w}_n = sgn(\hat{t})\sqrt{2n[\hat{t}\bar{x} - K_X(\hat{t})]}, \hat{u}_n = \hat{t}\sqrt{nK''_X(\hat{t})}$ and $\hat{t}$ is the saddlepoint, that is, the solution to the equation: $K_X(\hat{t})' = \bar{x}$. Like in the density case, the term of order $O(\frac{1}{n})$ can also be calculated but is more complicated and we do not give its explicit form here.

Examples Typically, it is difficult to give closed form formulae for the saddlepoint approximations in particular cases since the saddlepoint equation is typically non-linear and it is solved using iterative numerical methods. Some very simple cases can be dealt with.

Example 1. The saddlepoint approximation for the sample mean is exact for the standard normal distribution. $f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Indeed, we have $K_X(t) = \frac{t^2}{2}, \hat{t} = K'_X(\hat{t}) = \bar{x}, K''_X(t) \equiv 1$ hence (27) gives

$$\hat{f}(\bar{x}) = \sqrt{\frac{n}{2\pi}}e^{-\frac{n\bar{x}^2}{2}}$$

which is the density of $N(0, \frac{1}{n})$. Also, the Lugannani-Rice formula gives $P(\bar{X} \leq \bar{x}) = \Phi(\sqrt{n}\bar{x})$ which is the CDF of $N(0, \frac{1}{n})$.

Example 2. Saddlepoint approximation for the density of the sample mean of Gamma(α, 1) density. Here $f(x) = \frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}, x > 0$.

We have $K_X(t) = -\alpha\log(1-t), t < 1, K''_X(t) = \frac{\alpha}{(1-t)^2}$ and the saddlepoint equation $\frac{\alpha}{1-t} = \bar{x}$ has a root $\hat{t} = 1 - \frac{\alpha}{\bar{x}}$. This implies $K''_X(\hat{t}) = \bar{x}^2/\alpha$. Hence we get from (27)

$$\hat{f}(\bar{x}) = \sqrt{\frac{n\alpha}{2\pi\bar{x}^2}}exp[-n\alpha\log(1-\hat{t}) - n\hat{t}\bar{x}] = (\sqrt{\frac{2\pi}{n\alpha}}(n\alpha)^{n\alpha}e^{-n\alpha})^{-1}(n\bar{x})^{n\alpha-1}e^{-n\bar{x}}n$$

The expression $\frac{1}{\Gamma(n\alpha)}(n\bar{x})^{n\alpha-1}e^{-n\bar{x}}n$ is the exact density of $\bar{X}$ and we see that the difference between the exact and the saddlepoint approximation is just that the exact normalising constant $\frac{1}{\Gamma(n\alpha)}$ has been replaced by the constant $(\sqrt{\frac{2\pi}{n\alpha}}(n\alpha)^{n\alpha}e^{-n\alpha})^{-1}$. However,

the famous *Stirling approximation* of the Gamma function states precisely the fact that $\sqrt{\frac{2\pi}{n\alpha}}(n\alpha)^{n\alpha}e^{-n\alpha} \approx \Gamma(n\alpha)$ hence the ratio of the two normalising constants tends to one when sample size is increased. In fact if we renormalised the saddlepoint approximation, we would again recover the exact density!

8.4 Extensions of the saddlepoint method

If it was only about calculating precise approximations for the density of the sample mean **only**, the saddlepoint method would not have got very widespread. However, it turns out that the saddlepoint approximation idea can be extend and can be applied for approximating densities of maximum likelihood estimators in exponential families, of likelihood ratio or score statistics, of Bayes estimators etc. Many statistics can be approximated by sums of i.i.d. random variables (as seen from the representation

$$T(F_n) \approx T(F) + \frac{1}{n}\sum_{i=1}^{n} a(X_i) + remainder$$

in our discussion about estimating statistical functionals. For these, again the saddlepoint idea can be applied. There are also multivariate extensions of the method to approximate the joint distribution of vector-statistics.

We do not discuss these because of lack of time. We will finish with a formula (called *Barndorff-Nielsen's formula*) about the saddlepoint approximation of the density of the MLE in a k-parameter exponential family:

If $F_X(x;\theta) = e^{\theta' t(x) - \psi(\theta) - d(x)}$ is the density of a single observation, $L(\theta)$ is the joint likelihood of the sample, if $J(\theta) = -\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}$ then for the density of the MLE of $\theta \in R^k$, the following approximation can be derived:

$$\hat{f}(\theta) = (2\pi)^{-k}|J(\hat{\theta})|^{1/2}\{L(\theta)/L(\hat{\theta})\}[1 + O(n^{-1})] \tag{28}$$

(We note that the MLE $\hat{\theta}$ is a one-to-one transformation of the statistic $T = \sum_{i=1}^{n} t(X_i)$ in this case).

If in fact the constant $(2\pi)^{-k}$ is replaced by the true renormalisation constant then the order of the approximation in (28) improves to $O(n^{-3/2})$. The accuracy of formula (28) for approximating the density of the MLE, especially after renormalisation, is so impressive that it has been termed "the magic formula" by prominent statisticians (in an overview paper with discussions: "R. A. Fisher in the 21st Century" by B. Efron (*Statistical Science*, 1998, Vol. 13, No. 2, 95-122). Applications of this formula outside the exponential family, have been also been investigated by Barndorff-Nielsen.