8 Higher order asymptotics

8.1 Motivation

The main topic of this chapter is higher order asymptotics. We will expand gradually our knowledge about the first order asymptotic expansions that we learned about in Chapter 5.

As we have realised already, an inference procedure that is optimal for a finite sample size is very rarely possible to construct and is too specialised to the specific data distribution.

Hence such a procedure is too much individually tailored and cannot be offered as a general methodological tool. The first order asymptotic expansions are the fundamental tool used to get the asymptotic distribution of estimators and tests.

Using these, we have seen that for large families of (not necessarily normal) distributions $f(x, \theta)$ the maximum likelihood estimator $\hat{\theta}$ is consistent and asymptotically normally distributed.

The consistency is useful for the construction of the point estimator and the asymptotic normality is useful for constructing asymptotic confidence intervals and for testing hypotheses about the parameter θ.

The null distribution of a certain test statistic is asymptotically normal or chi-squared in the case of the generalised likelihood ratio tests.

χ^2_n

The main instruments used in the first order asymptotics are the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). The final product of their application is a statement about asymptotic normality of the statistic of interest.

However, if sample sizes are not large enough to warrant that the asymptotic normal approximation is precise enough, it is worth trying to include higher order asymptotic expansions.

By doing so, we hope to bring about a better approximation of the distribution when the sample size is not as large.

$X_1, \ldots, X_n$ f $\bar{X}$

Lets take an example to illustrate the 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. That is, i.i.d. observations $X_1, X_2, \ldots, X_n$ are given with the same distribution and with

$$\mathbb{E}(X_i) = \mu \quad \text{and} \quad \text{Var}(X_i) = \sigma^2, \quad i = 1, 2, \ldots, n.$$

$\bar{X} \overset{a}{\sim} N(\mu, \frac{\sigma^2}{n})$

- only requires the first two moments of X.

Let

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \quad \text{and} \quad F_n(z) = P(Z_n \leq z).$$

Then the **CLT** says that

$$F_n(z) \xrightarrow[n \to \infty]{} \Phi(z)$$

for every $z \in \mathbb{R}$ where $\Phi(z)$ denotes the cdf of the standard normal distribution.

→ Higher order moments might produce more precise approximations.

If we also assume that a finite absolute third moment

$$E(|X_i|^3) = \gamma, i = 1, 2, \ldots, n$$

exists then the famous Berry–Esseen theorem states that

$\sup_z |F_n(z) - \Phi(z)|$

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

depends on γ

uniformly in z where the bound on the right-hand side depends on the third absolute moment.

The notation $O(\frac{1}{\sqrt{n}})$ means that the ratio:

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

when we "blow up" the error by $\sqrt{n}$ it is still bounded.

is bounded from above by a certain positive constant M for any value of n.

A more precise statement can be obtained in the form

$\longrightarrow 0$

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

$n \qquad\qquad \le M$

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 a certain first degree polynomial.

Now we have the left-hand side (LHS) being $O(\frac{1}{n})$, which indicates that it tends to zero more quickly than

$$n\,|F_n(z) - \Phi(z)| \longrightarrow \infty$$

does, since even when blown up with the factor of n, this LHS still remains bounded for any n.

This also implies that if we blow up this LHS by just $\sqrt{n}$ instead of n, it will not only stay bounded but will even tend to zero as $n \to \infty$. This fact is denoted as

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

It is clear then, from the above formula, that

$$\Phi(z) + \frac{C_1(F)p_1(z)\phi(z)}{\sqrt{n}}$$

$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$

$\parallel$

delivers a more precise approximation to the distribution of Z_n than the approximation $\Phi(z)$ does.

CLT

The above expansion can be extended using expansions in the form

$$n^{k/2} \left| F_n(z) - \Phi(z) - \sum_{s=1}^{k} \frac{q_s(z)}{n^{s/2}} \right| \to 0$$

$$\frac{1}{n^{1/2}}, \frac{1}{n}, \frac{1}{n^{3/2}}, \frac{1}{n^2} \dots$$

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

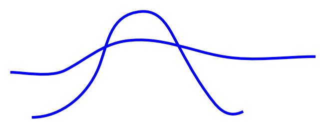
The expansions of this type can often be obtained for the distribution $F_n(.)$ of suitably normalised statistics of interest and are called Edgeworth expansions.

$\phi(z)$

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 $F(z)$. 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.

κ_r

μ_r



8.2 Moments and cumulants

How do we determine the summands in the last k-th order expansion in the previous illustration?

We know from the definition of Z_n that for the CLT we only need the mean μ and the variance σ^2 of X_i.

If we need more precise approximation (namely the Edgeworth approximation), we need to involve more knowledge about the distribution of X_i (i.e., we need to "pay more" to obtain a better approximation). This knowledge involves the higher order moments of the distribution of X_i.

Moment generating function

The moment generating function (MGF) of a random variable X is defined as

$$M_X(t) = E[\exp(tX)]$$

$E e^{0} = E(1) = 1$

whenever the expectation is finite. It is obvious that always $M_X(0) = 1$ holds.

$M_X(t) = E\, e^{tX} \Rightarrow M_X'(t) = E\left[X e^{tX}\right]$ or $M_X'(0) = E\left[X e^{0}\right] = E(X) = \mu_1'$

The MGF may only be defined in a small neighbourhood 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:

$$E(X^r) = M_X^{(r)}(t)\Big|_{t=0} = \mu_r'.$$

with μ_r' being short-hand notation for the raw moment $E(X^r)$ (as opposed to $\mu_r = E(X - \mu_1')^r)$ used to denote the central moments).

This is the reason that $M_X(t)$ has the name moment generating function as it helps to generate the moments.

$f(x)$ around a: $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$

$a=0$

We then have the Taylor expansion

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

when $t \to 0$.

$M_X(t) = \frac{M_X(0)}{0!}(t-0)^0 + \frac{M'_X(0)}{1!}(t-0)^1 + \frac{M''_X(0)}{2!}(t-0)^2 + \cdots$

$= 1 + \mu'_1 t + \frac{\mu'_2}{2!} t^2 + \cdots$

Cumulant generating function

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.

We note also that $\exp[K_X(t)] = M_X(t)$ holds.

$K_X(0) = \log M_X(0)$

$= \log(1) = 0$

The Taylor series expansion is:

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

when $t \to 0$ holds in this interval and defines the cumulants

$$k_s, \quad s = 1, 2, \ldots$$

of the distribution of X where

$$\kappa_r = K_X^{(r)}(t)\Big|_{t=0}$$

$$M_X(t) = e^{K_X(t)} \iff K_X(t) = \log M_X(t)$$

By equating the coefficients in the Taylor expansions of $\exp[K_X(t)]$ and of $M_X(t)$, relationships between moments and cumulants can be established. By using this approach, one can show that:

i) $\kappa_1 = \mu_1' = \mathbb{E}(X)$

ii) $\kappa_2 = \text{Var}(X) = \mu_2' - (\mu_1')^2$

iii) $\kappa_3 = 2\mu_1'^3 - 3\mu_1'\mu_2' + \mu_3'$

iv) $\kappa_4 = -6(\mu_1')^4 + 12(\mu_1')^2\mu_2' - 3(\mu_2')^2 - 4\mu_1'\mu_3' + \mu_4'$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

Solution

Since $K_X(0) = 0$ always, we get the Taylor expansion

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

By substituting into $\exp[K_X(t)] = M_X(t)$ we get:

$$e^{\kappa_1 t} e^{\kappa_2 \frac{t^2}{2!}} e^{\kappa_3 \frac{t^3}{3!}} \cdots = 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \ldots$$

Expanding the exponent on the LHS as

$$e^{\kappa_1 t} = 1 + \kappa_1 t + \kappa_1^2 \frac{t^2}{2!} + \dots,$$

and similarly for the remaining exponents, multiplying and combining the coefficients in front of each power of t and finally equating these to coefficients in front of the respective power of t on the RHS we get:

$$\kappa_1 = \mu_1', \qquad \kappa_2 + \kappa_1^2 = \mu_2'$$

from which we get

$$\kappa_2 = \mu_2' - (\mu_1')^2 = \mathrm{Var}(X) = \sigma^2$$

and similarly:

$$\kappa_3 = 2\mu_1'^3 - 3\mu_1'\mu_2' + \mu_3',$$

and

$$\kappa_4 = -6(\mu_1')^4 + 12(\mu_1')^2\mu_2' - 3(\mu_2')^2 - 4\mu_1'\mu_3' + \mu_4'.$$

Per **definition** the third cumulant is the skewness and the fourth cumulant is the kurtosis 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, the higher order asymptotic expansions we are about to discuss can in fact, be presented in more compact form by using the cumulants.

Exercise 8.38 (at lecture)

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

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

$$f_n(z) = \gamma_0(z)b_{0,n} + \gamma_1(z)b_{1,n} + \cdots + \gamma_k(z)b_{k,n} + o(b_{k,n}) \qquad (12)$$

The expansion is considered as $n \to \infty$. Typical choices of the doubly indexed coefficients $b_{k,n}$ are

$$\{1, n^{-1/2}, n^{-1}, \ldots, n^{-k/2}\} \quad \text{or} \quad \{1, n^{-1}, n^{-2}, \ldots, n^{-k}\}.$$

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 or the CDF of the standard normal depending on the context.

ϕ

Φ

Remark 8.21

Important: 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 (12) is the Cramér condition. A cdf $F(.)$ on the real line is said to satisfy Cram'er's condition if for the characteristic function $\chi_F(t) = E_F(e^{itX})$:

$$\limsup_{t \to \infty} \left| E_F(e^{itX}) \right| < 1 \tag{13}$$

holds. This is needed to deal with some singular cases. It can be shown that *all* continuous distributions do satisfy Cram'er condition.

On the other hand, the lattice distributions, whose whole mass is concentrated on a lattice in the form

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

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 (12) 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 8.31

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

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

Here we have the coefficients:

$$z = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$$

- $C_1(F) = \dfrac{E(X-\mu)^3}{6\sigma^3}$ (skewness correction),
- $C_2(F) = \dfrac{\frac{E(X-\mu)^4}{\sigma^4} - 3}{24}$ (kurtosis correction),
- $C_3(F) = \dfrac{C_1^2(F)}{2}$,

More to come . . .

$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$\phi'(x) = \frac{-2x}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$= -x\phi(x)$

And the polynomials are given by:

- $p_1(z) = z^2 - 1$,
- $p_2(z) = z^3 - 3z$,
- $p_3(z) = z^5 + 15z - 10z^3$.

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

$$H_j(z) = \frac{(-1)^j \phi^{(j)}(z)}{\phi(z)}, j = 0, 1, 2, \ldots.$$

and $\phi^{(j)}(z)$ is the j-th order derivative of the standard normal density.

The first six Hermite polynomials are explicitly:

$$
\begin{aligned}
H_1(z) &= z \\
H_2(z) &= z^2 - 1 \\
H_3(z) &= z^3 - 3z \\
H_4(z) &= z^4 - 6z^2 + 3 \\
H_5(z) &= z^5 - 10z^3 + 15z \\
H_6(z) &= z^6 - 15z^4 + 45z^2 - 15
\end{aligned}
$$

Edgeworth expansions are very often formulated by using the so-called standardised cumulants. The definiton of the rth order standardised cumulant ρ_r is

$$\rho_r = \frac{\kappa_r}{\kappa_2^{r/2}}.$$

$r=1$

$$\rho_1 = \frac{\kappa_1}{\kappa_2^{1/2}} = \frac{\mu_1'}{\sqrt{\sigma^2}} = \frac{\mu_1'}{\sigma}$$

In particular,

- $\rho_3 = \frac{\kappa_3}{\kappa_2^{3/2}}$ is called the standardised skewness;
- $\rho_4 = \frac{\kappa_4}{\kappa_2^2}$ is called the standardised kurtosis.

Using the standardised cumulants, we can see that from the formula in the previous slide we have the following relations for the constants $C_i(F), i = 1, 2, 3:$

$$C_1(F) = \frac{1}{6}\frac{\kappa_3}{\sigma^3} = \frac{1}{6}\rho_3,$$
$$C_2(F) = \frac{1}{24}\rho_4,$$
$$C_3(F) = \frac{C_1^2}{2} = \frac{\rho_3^2}{72}.$$

8.3.2 Edgeworth expansion for the density of Z_n

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

$$\hat{f}_{Z_n}(z) = \phi(z)\left\{1 + \frac{\rho_3}{6\sqrt{n}}H_3(z) + \frac{1}{n}\left[\frac{\rho_4 H_4(z)}{24} + \frac{\rho_3^2 H_6(z)}{72}\right]\right\} + o(n^{-1}) \quad (14)$$

and this expansion holds uniformly in $z \in \mathbb{R}$.

The interpretation of the previous formula is that:

1) The leading term is the standard normal density in lieu of the CLT. Then there are higher order correction terms whose relevance becomes important for small to moderate sample sizes.
2) The correction terms represent simultaneous adjustment of the normal approximation by using the information for the standardised skewness and kurtosis.
3) 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 $\rho_3 \neq 0$.

$\kappa_3 = 0$

Z_n is centered around zero $Z_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$

4) Another 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 may worsen quite a lot and may even become negative. The same observation holds also for $F_n(z)$, which may become either negative or greater than one in the tails. The saddlepoint approximation to be discussed later this week is meant to improve upon Edgeworth, especially in the tails.

At $z=0 \Rightarrow H_3(0) = 0^3 - 3\times 0 = 0$

No $\frac{1}{\sqrt{n}}$ term and so its most accurate at the center and worsens at the tail!

8.3.3 Cornish-Fisher expansions for quantiles

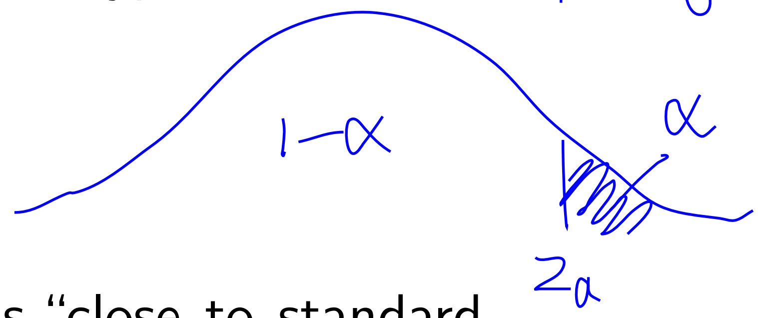
One major application of the Edgeworth expansion in practice is, in fact, in "reversing" it to construct better confidence intervals for small to moderate sample sizes.

When studying coverage probabilities of confidence intervals, we need to solve, given confidence level α, equations of the type

$Z_n = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$

$P(Z_n > z_\alpha) = \alpha$

$$P(Z_n \le z_\alpha) = F_{Z_n}(z_\alpha) = 1 - \alpha$$



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

$$u_\alpha : \Phi(u_\alpha) = 1 - \alpha.$$

u_α, η_α to approximate z_α

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 with respect 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 the Cornish–Fisher expansion η_α.

The precise result (again if we contain ourselves to order $o(n^{-1})$ only) is given as follows:

Approximate z_α : $F_{Z_n}(z_\alpha) = 1-\alpha$ by η_α below.

Theorem 8.32

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

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

holds.

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

Assume that $W_n \sim \chi^2_n$ is chi-squared with n degrees of freedom (d.f.). Per definition this means that W_n is a sum $\sum_{i=1}^n X_i^2$ of n i.i.d. squared standard normal random variables.

$X_i \sim N(0,1)$

Each X_i^2 is chi-squared with one degree of freedom with $\mathbb{E}(X_i^2) = 1$ and $\text{Var}(X_i^2) = 2$ holds.

Then W_n/n can be represented as an average $\frac{1}{n}\sum_{i=1}^n X_i^2$ of n i.i.d. squared standard normals X_i, $i = 1, 2, \ldots, n$.

$$\frac{W_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \overline{X^2}$$

$K_X(t)$

You can calculate the first four cumulants of the square of a standard normal directly. You can then get for the standardised cumulants:

$$\rho_3 = 2\sqrt{2} \quad \text{and} \quad \rho_4 = 12.$$

Exercise 8.39 (at lecture)

Confirm that $\rho_3 = 2\sqrt{2}$ and $\rho_4 = 12$ for the chi-square distribution with $n = 1$ degrees of freedom from the previous example.

Since CLT : $\frac{W_n}{n} \overset{a}{\sim} N(1, \frac{2}{n})$

$$\sqrt{n}\frac{(\frac{W_n}{n}-1)}{\sqrt{2}} = \frac{(W_n - n)}{\sqrt{2n}} \quad \sim N(0,1)$$

is about standard normal by the CLT. Then since,

$$\alpha = \quad P\Big(W_n \geq z_\alpha\Big) = P\Big(\frac{(W_n - n)}{\sqrt{2n}} \geq \frac{z_\alpha - n}{\sqrt{2n}}\Big) \quad \approx P\Big(Z \geq \frac{z_\alpha - n}{\sqrt{2n}}\Big)$$

we know that $(z_\alpha - n)/\sqrt{2n}$ should be "close" to u_α so then

$P(Z \geq u_\alpha) = \alpha$

Solve $u_\alpha = \frac{z_\alpha - n}{\sqrt{2n}}$

$$n + u_\alpha\sqrt{2n}$$

is the first order approximation of the $(1-\alpha)\%$ quantile of W_n.

Recall that u_α is the upper $\alpha \times 100\%$ point of the standard normal distribution, i.e., $u_\alpha : \Phi(u_\alpha) = 1 - \alpha$ holds. For example, if:

$$\alpha = 0.025 \to u_\alpha = 1.96$$
$$\alpha = 0.05 \to u_\alpha = 1.645$$
$$\alpha = 0.01 \to u_\alpha = 2.326$$

qnorm(1-α) ⟺ qnorm(α, lower.tail = FALSE)

$\rho_3 = 2\sqrt{2} \qquad \rho_4 = 12$

The higher order approximation using the Cornish-Fisher expansion (Theorem 8.32) will be

$$\eta_\alpha = n + \sqrt{2n}\left[u_\alpha + \frac{(u_\alpha^2 - 1)2\sqrt{2}}{6\sqrt{n}} + \frac{(u_\alpha^3 - 3u_\alpha)12}{24n} - \frac{(2u_\alpha^3 - 5u_\alpha)8}{36n}\right] \quad (15)$$

which leads us to

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

which should be more precise than $n + u_\alpha\sqrt{2n}$

$\alpha = 0.01 \quad n = 5 \qquad \chi^2_5$

For $\alpha = 0.01$ we have $u_\alpha = 2.326$. By applying the CLT-based approximation $\underline{n + u_\alpha \sqrt{2n}}$ for $n = 5$ gives 12.36.

$\chi^2_{5,0.01} = 15.09$

Including the next term:

$$n + \sqrt{2n}\left[u_\alpha + \frac{(u_\alpha^2 - 1)2\sqrt{2}}{6\sqrt{n}}\right] = n + \sqrt{2n}u_\alpha + \frac{(u_\alpha^2 - 1)2}{3}$$

gives 15.296 whereas using all the terms

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

delivers 15.16.

χ^2_5

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 Cornish–Fisher approximations are quite good, with the second more complex approximation almost hitting the true value.

qchisq(0.99, 5) = 15.08627

8.3.4 Edgeworth expansions for other statistics

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$$

Our discussion was related to Edgeworth expansion for the normalised distribution of the sample mean (since it was the easiest case).

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 in the tails. One explanation for this phenomenon is that the error approximation in the Edgeworth expansion is only absolute instead of relative.

Hence, 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 a serious error of the approximation.

This is the reason to look for alternative expansions where the error of approximation is 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:

In the original paper by Henry Daniels in *The Annals of Mathematical Statistics* in 1954. This 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 imaginery axis) by exploiting the biggest contribution to the integral coming form a small region around the saddlepoint of the integrand.

However this is difficult to comprehend. The second approach is easier and leads to the same result.

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 Eq. (14) in terms of the variable $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ to obtain the density for $f_{\bar{X}}(\bar{x})$ as follows:

$$\frac{\sqrt{n}}{\sigma}\phi(z)\left\{1 + \frac{\rho_3}{6\sqrt{n}}H_3(z) + \frac{1}{n}\left[\frac{\rho_4 H_4(z)}{24} + \frac{\rho_3^2 H_6(z)}{72}\right] + o(n^{-1})\right\} \quad (16)$$

$\frac{dz}{d\bar{x}} = \frac{\sqrt{n}}{\sigma}$ $\qquad$ $f_{\bar{x}}(\bar{x}) = f(z(\bar{x}))\cdot\left|\frac{\sqrt{n}}{\sigma}\right|$

where $z = \dfrac{\sqrt{n}(\bar{x}-\mu)}{\sigma}$.

$z=0$ when $\bar{x}=\mu$

Because, at the centre $z=0$, the Hermite polynomials are

$$H_3(0)=0, \quad H_4(0)=3, \quad H_6(0)=-15$$

we can see that the approximation in Eq. (16) 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:

$$f_{\bar{X}}(\bar{x}) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\left\{1 + \frac{1}{n}\left(\frac{1}{8}\rho_4 - \frac{5}{24}\rho_3^2\right) + O(n^{-2})\right\}. \tag{17}$$

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 $\equiv$ 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 CGF $K_X(t)$ that converges on an open neighbourhood A of the point 0, we can define the tilted density family

- Multiply by e^{tx}

- then $e^{-K_X(t)}$ is the normalizing constant

$f(x)\, e^{-K_X(t)}$ e^{tx}

$$f(x;t) = e^{tx - K_X(t)} f(x), \quad t \in A.$$

This means that, in fact, we are embedding the original density $f(x)$ in a one-parameter exponential family parameterised (indexed) by t. Of course, we can also write

$$f(x) = e^{K_X(t) - tx} f(x;t) \quad \text{for any } t \in A \tag{18}$$

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

$K''_X(0) = \kappa_2$

It can then be seen that if X_t denotes a random variable with the density $f(x;t)$ then for its mean, variance and standardised cumulants we have the relationship:

$$\mathbb{E}[X_t] = K'_X(t), \quad \mathrm{Var}[X_t] = K''_X(t), \quad \rho_i(t) = \frac{K_X^{(i)}(t)}{K''_X(t)^{i/2}}, \quad i \geq 3. \tag{19}$$

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 $X_{\hat{t}}$). Since $\mu_3(0) = 0$ then no $\sqrt{n}$ term //

Step two: Transform back to calculate $f(x)$ using the formula Eq. (18) 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,

$$E[X_t] = K'_X(\hat{t}) = x. \tag{20}$$

Eq. (20) is called the saddlepoint equation. Then we apply the Edgeworth expansion at the mean as given in Eq. (17). This gives

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

Then, in the second step, we substitute the above approximation in the right-hand side of Eq. (18) to get the approximation:

$$f(x) = e^{\{-\hat{t}x + K_X(\hat{t})\}} f(x;\hat{t})$$
$$\approx \frac{1}{\sqrt{2\pi K_X''(\hat{t})}} e^{[K_X(\hat{t}) - \hat{t}x]} \left\{1 + \left[\frac{1}{8}\rho_4(\hat{t}) - \frac{5}{24}\rho_3^2(\hat{t})\right]\right\}. \qquad (21)$$

$\hat{t} : K_X'(\hat{t}) = x$

The discussion until now involved essentially any random variable X. The asymptotic point of view is brought forward when Eq. (21) 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 just $K'_X(t)$ and the optimal tilt is obtained by solving $K'_X(\hat{t}) = \bar{x}$.

It can be 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 Eq.(19). Hence we get

$$\hat{f}(\bar{x}) \approx \sqrt{\frac{n}{2\pi K''_X(\hat{t})}} e^{\{nK_X(\hat{t})-n\hat{t}\bar{x}\}} \left\{1 + \left[\frac{1}{8n}\hat{\rho}_4 - \frac{5}{24n}\hat{\rho}_3^2\right]\right\} \tag{22}$$

$\rightarrow$ No powers of $n^{-1/2}, n^{-3/2}, \ldots$ as in the Edgeworth expansions

The above approximation Eq.(22) is called the second order saddle-point density approximation and is extremely accurate! Sometimes, for the sake of numerical simplicity, even just the first order saddlepoint approximation

$$\hat{f}(\bar{x}) \approx \sqrt{\frac{n}{2\pi K_X''(\hat{t})}} e^{\{nK_X(\hat{t}) - n\hat{t}\bar{x}\}} \tag{23}$$

is used and it is still very accurate!

$\hat{t}$

$\hat{F}(\bar{x}) = \int \hat{f}(\bar{x}) d\bar{x}$

$K_X(t)$

Example 8.64

Describe the similarities and the differences between the saddlepoint approximation Eq. (22) and the Edgeworth expansion for densities of $\bar{X}$.

1) The saddlepoint approximation is an asymptotic expansion on powers of n^{-1} instead of $n^{-1/2}$, as in the Edgeworth.
2) The latter fact implies that the simple approximation (Eq. (23)) has already absorbed the skewness correction.
3) The leading term (Eq. (22)) is clearly *not* the normal (nor, in fact, any other) density.
4) Both the saddlepoint and the Edgeworth expansion may not integrate to one.

5) 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. In other words, "for the saddlepoint approximation by requiring more we achieve more".

$\kappa_1, \kappa_2, \kappa_3 \ldots$

6) 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 $\bar{x}$.

8.3.6 Saddlepoint for CDF and examples

A very accurate asymptotic approximation also exists for the cumulative distribution function of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ where $K_X(t)$ is the cgf of X_i.

This approximation is derived via two layers of approximation: first, the density is approximated via Eq. (23). The cdf is then approximated as an approximation (the so-called Temme approximation) to the integral of this approximated density. The latter represents an approximate integration by parts formula.

We omit the details and only give the final formula.

Φ cdf $N(0,1)$
ϕ pdf $N(0,1)$

It is called the Lugannani-Rice formula:

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

where

$\hat{t}$ $K_X(t)$

$$\hat{w}_n = sgn(\hat{t})\sqrt{2n[\hat{t}\bar{x} - K_X(\hat{t})]} \qquad \text{and} \qquad \hat{u}_n = \hat{t}\sqrt{nK_X''(\hat{t})}$$

and $\hat{t}$ is the saddlepoint, that is, the solution to the equation:

$$\boxed{K_X'(\hat{t}) = \bar{x}.}$$

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.

Typically, it is difficult to give closed form formula for the saddlepoint approximations in particular cases since the saddlepoint equation is typically non-linear and it is solved using iterative numerical methods. Next, we show that some very simple cases can be dealt with explicitly.

Example 8.65

The saddlepoint approximation for the sample mean is exact for the standard normal distribution with density

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

We have

$$K_X(t) = \frac{t^2}{2}, \quad \hat{t} = K'_X(\hat{t}) = \bar{x}, \quad \text{and} \quad K''_X(t) \equiv 1$$

hence Eq. (23) 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 $P(\bar{X} \le \bar{x}) = \Phi(\sqrt{n}\bar{x})$ which is the cdf of $N(0, \frac{1}{n})$.

Example 8.66

Find the saddlepoint approximation for the density of the sample mean of Gamma(α, 1) density:

$$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}, \qquad x > 0$$

Solution:

First notice that for $t < 1$ we have that

$$= \int_0^\infty e^{tx} \cdot \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx$$

$$\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-(1-t)x} x^{\alpha-1} \mathrm{d}x
\end{aligned}$$

$x: 0 \to \infty$

$y: 0 \to \infty$

Now let $(1-t)x = y$ so that $\mathrm{d}x = \mathrm{d}y/(1-t)$

$$\begin{aligned}
&= \frac{1}{\Gamma(\alpha)(1-t)^\alpha} \underbrace{\int_0^\infty e^{-y} y^{\alpha-1} \mathrm{d}y}_{\Gamma(\alpha)} \\
&= \frac{\cancel{\Gamma(\alpha)}}{\cancel{\Gamma(\alpha)}(1-t)^\alpha} \\
&= (1-t)^{-\alpha}
\end{aligned}$$

$t < 1$

where we have used the fact that the Gamma$(\alpha, 1)$ density must integrate to one:

$$\int_0^\infty e^{-y} y^{\alpha-1} \mathrm{d}y = \Gamma(\alpha).$$

Hence, we have

$K_x(t) = \log (1-t)^{-\alpha} = -\alpha \log (1-t)$

$$K_X(t) = -\alpha \log(1-t), \qquad t < 1,$$
$$K'_X(t) = \frac{\alpha}{1-t},$$
$$K''_X(t) = \frac{\alpha}{(1-t)^2}.$$

The saddlepoint equation $\boxed{K'_x(\hat{t}) = \bar{x}}$ implies that

$$\Rightarrow \quad \frac{\alpha}{1-\hat{t}} = \bar{x} \qquad \text{that is} \qquad \boxed{\hat{t} = 1 - \frac{\alpha}{\bar{x}}.}$$

Hence,

$K_x(\hat{t}) = -\alpha \log\left(1-\left(1-\frac{\alpha}{\bar{x}}\right)\right)$

$= -\alpha \log \frac{\alpha}{\bar{x}}$

$$K''_X(\hat{t}) = \frac{\bar{x}^2}{\alpha}.$$

$= \frac{\alpha}{\left(\alpha - \left(\alpha - \frac{\alpha}{\bar{x}}\right)\right)^2} = \frac{\alpha}{\left(\frac{\alpha}{\bar{x}}\right)^2} = \frac{x^2}{\alpha}$

$$\hat{f}(\bar{x}) = \sqrt{\frac{n}{2\pi K_x''(\hat{t})}}\, e^{nK_x(\hat{t}) - n\hat{t}\bar{x}}$$

Hence by substituting these expressions into the first-order saddlepoint approximation formula in (Eq. (23)) we get:

$$n\left(1 - \frac{\alpha}{\bar{x}}\right)\bar{x} = n(\bar{x} - \alpha)$$

$$\begin{aligned}
\hat{f}(\bar{x}) &= \sqrt{\frac{n\alpha}{2\pi\bar{x}^2}} \exp\left[-n\alpha\log(1-\hat{t}) - n\hat{t}\bar{x}\right] \\
&= \sqrt{\frac{n\alpha}{2\pi\bar{x}^2}} \exp\left[-n\alpha\log\left(\frac{\alpha}{\bar{x}}\right) - n(\bar{x}-\alpha)\right] \\
&= \boxed{\left(\sqrt{\frac{2\pi}{n\alpha}}(n\alpha)^{n\alpha}e^{-n\alpha}\right)^{-1}} (n\bar{x})^{n\alpha-1}e^{-n\bar{x}}n
\end{aligned}$$

$$e^{-n\bar{x}} \quad e^{-n\alpha}$$

$$\exp\log\left(\frac{\bar{x}}{\alpha}\right)^{n\alpha} = \left(\frac{\bar{x}}{\alpha}\right)^{n\alpha}$$

$$\propto (n\bar{x})^{n\alpha-1} e^{-n\bar{x}}$$

$X_i \sim Gamm(\alpha, 1)$

The expression

$$\boxed{\frac{1}{\Gamma(n\alpha)}}\underbrace{(n\bar{x})^{n\alpha-1}e^{-n\bar{x}}n}$$

$= f_{gamma}(n\bar{x}) \cdot n$

is the exact density of $\bar{X}$ since

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \sum_{i=1}^{n} X_i \sim Gamma(n\alpha, 1)$$

we can just apply the density transformation formula. 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

$$\left(\sqrt{\frac{2\pi}{n\alpha}}(n\alpha)^{n\alpha}e^{-n\alpha}\right)^{-1}.$$

However, the famous Stirling approximation of the Gamma function states precisely the fact that

Approx

$$\sqrt{\frac{2\pi}{n\alpha}}(n\alpha)^{n\alpha}e^{-n\alpha} \approx \Gamma(n\alpha)$$

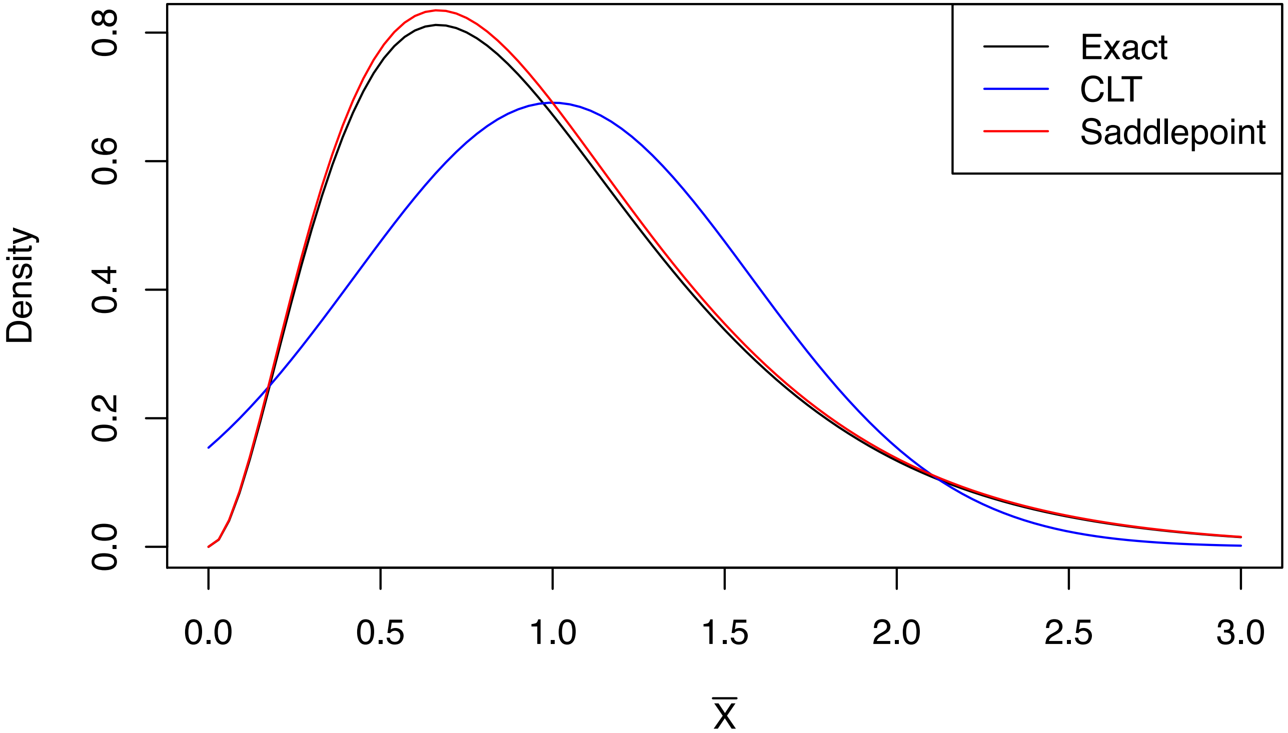
Exact

$n \to \infty$

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!

$\bar{X}$ $n=3$ $\alpha=1$

With $n = 3$, $\alpha = 1$ we have



Exercise 8.40 (at lecture)

Assume that X_i, $i = 1, 2, \ldots, n$ are $N(\mu, \sigma^2)$ (instead of just $N(0,1)$ and consider the saddlepoint density approximation for $\bar{X}$.

i) Show that the cgf of X_i is $K(t) = \mu t + \frac{\sigma^2 t^2}{2}$.

ii) Show that the solution to the saddlepoint equation is $\hat{t} = \frac{\bar{x}-\mu}{\sigma^2}$ and that $K''(\hat{t}) = \sigma^2$ holds.

iii) Show that the saddlepoint density approximation for $\bar{x}$ coincides with the true density of $\bar{x}$, i.e., it is exact with no error.

Exercise 8.41 (at lecture)

Assume that X_i, $i = 1, 2, \ldots, n$ and $\lambda > 0$ is a parameter. Consider the exponentially distributed random variable with a density

$$f(x, \lambda) = \lambda e^{-x\lambda}, \quad x > 0$$

(and zero else).

i) Find the cumulant generating function $K_X(t)$.

ii) Find the solution $\hat{t}$ of the saddlepoint equation $K'_X(\hat{t}) = \bar{x}$ and $K''_X(\hat{t})$.

iii) Find the saddlepoint approximation for the density of $\bar{X}$.

iv) Compare the saddlepoint approximation formula with the true density formula of $\bar{X}$. What is the discrepancy between the two expressions, if any?

Tutorial Problem 5

Exercise 8.42 (at lecture)

Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are i.i.d. observations from Poission(λ) distribution.

i) Show that the cgf for a single observation X_i is $K(t) = \lambda(e^t - 1)$.

ii) Show that the one-term saddlepoint approximation for the density of the sum of n i.i.d. obervations $Y = \sum_{i=1}^n X_i$ from the distribution is given by

$$\hat{g}(y) = \frac{1}{\sqrt{2\pi}} \frac{e^{-n\lambda} e^y (n\lambda)^y}{y^{y+1/2}}, \qquad y = 0, 1, 2, \ldots$$

Note: Stirling's approximation of

$$n! \approx \sqrt{2\pi n} e^{-n} n^n$$

for large n, the above approximation coincides up to a normalizing constant the exact density (which is Poisson$(n\lambda)$) and the consant tends to one when $n \to \infty$.

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 widespread.

However, the saddlepoint approximation idea can be extended and 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. Suppose:

$$F_X(x;\theta) = e^{\theta' t(x) - \psi(\theta) - d(x)}$$

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 $\theta \in \mathbb{R}^k$, the following approximation can be derived:

$$\hat{f}(\theta) = (2\pi)^{-k} \left|J(\hat{\theta})\right|^{1/2} \left\{\frac{L(\theta)}{L(\hat{\theta})}\right\} \left[1 + O(n^{-1})\right]. \tag{24}$$

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 $(2\pi)^{-k}$ is replaced by the true renormalisation constant then the order of the approximation in (24) improves to $O(n^{-3/2})$.

The accuracy of formula (24) for approximating the denisty 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.

Exercise 8.43 (at lecture)

Assume that $X_1, X_2, \ldots, X_n$ are a sample of exponential distribution with mean:

$$f(x,\theta) = \begin{cases} \theta e^{-\theta x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Here $\theta > 0$ is unknown parameter.

i) Show that the MLE is $\hat{\theta} = \frac{1}{\bar{X}}$.

ii) Derive the magic formula for the density of the MLE.