Chapter 6

Edgeworth expansions

Date: January 24 2006

6.1 Hermite polynomials and Hermite measure

In the following we will be discuss a number of constructions related to the standard normal distribution. We refer to the density and the distribution function of $N(0, 1)$ as $\phi$ and $\Phi$ respectively. That is,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \Phi(x) = \int_{-\infty}^{x} \phi(y) dy.$$ 

Observe that if $p(x)$ is any differentiable function, then

$$\frac{d}{dx} \left( p(x) \phi(x) \right) = \left( p'(x) - xp(x) \right) \phi(x). \quad (6.1)$$

If $p(x)$ is a polynomial, then $p'(x) - xp(x)$ is again a polynomial, and so the class

$$\{ p(x)\phi(x) \mid p(x) \text{is a polynomial} \}$$

is stable under differentiation.
Definition 6.1 The k\textsuperscript{th} Hermite polynomial $H_n$ satisfies the equation

$$\frac{d^k}{dx^k} \phi(x) = (-1)^k H_k(x) \phi(x)$$

(6.2)

The above arguments shows that there is a unique polynomial solution to (6.2. We have the equation

$$(-1)^{k+1} H_{k+1}(x) \phi(x) = \frac{d^{k+1}}{dx^{k+1}} \phi(x) = \frac{d}{dx} \left( \frac{d^k}{dx^k} \phi(x) \right) = \frac{d}{dx} \left( (-1)^k H_k(x) \phi(x) \right)$$

$$= \left( (-1)^k H'_k(x) - (-1)^k x H_k(x) \right) \phi(x)$$

Hence we have the recursive relation

$$H_{k+1}(x) = x H_k(x) - H'_k(x)$$

(6.3)

Starting from $H_0(x) = 1$, this recursive relation lets us compute the first Hermite polynomials easily, and we obtain

$$H_0(x) = 1 \quad H_1(x) = x \quad H_2(x) = x^2 - 1$$

$$H_3(x) = x^3 - 3x \quad H_4(x) = x^4 - 6x^2 + 3 \quad H_5(x) = x^5 - 10x^3 + 15x$$

and so on. Various properties can be observed in this sequence of polynomials (rigorous proofs easily obtained using (6.3). The polynomial $H_k(x)$ has degree $k$ and it is normalized in the sense that the coefficient for the $x^k$-term is 1 (this is the point of the $-1$-factor in definition 6.1, if this factor was omitted, the leading term of the polynomials would alternate sign).

And very important: if $n$ is odd, then $H_n(x)$ only contains odd terms, implying that is an odd function. And if $n$ is even, then $H_n(x)$ only contains even terms, implying that it is an even function. A number of insights obtained from Edgeworth expansions follows from this odd/even observation, while the particular form of the Hermite polynomial is less relevant.

There are other classical relations involving Hermite polynomials, for instance

$$H'_{k+1}(x) = (k+1) H_k(x)$$

(6.4)
which we probably will no be needing. The origin of the concept is \( L_2 \)-theory: if we consider \((\mathbb{R}, \mathcal{B})\) equipped with \(N(0, 1)\), then the Hermite polynomials are orthogonal in the \(L_2\)-sense. In fact, except for normalisation issues, they are what comes out of applying the Gram-Schmidt procedure to the usual monomials \(1, x, x^2, \ldots\). We are not interested in \(L_2\)-theory, and we have a completely different application in mind. And so we will not prove the orthogonality - though it is not very difficult to do.

**Definition 6.2** We define the \(k\)'the Hermite measure on \((\mathbb{R}, \mathcal{B})\) as

\[
\mu_n(A) = \int_A H_n(x) \phi(x) \, dx \quad \text{for all } A \in \mathcal{A}.
\]

As \(N(0, 1)\) has moments of all orders, we see that \(H_n \phi\) is integrable, so \(\mu_n\) is a well-defined signed measure. We recognise \(\mu_0\) to be \(N(0, 1)\), so we start at a probability measure, but the remaining Hermite measures are not positive.

**Lemma 6.3** The Hermite measure \(\mu_n\) has distribution function \(\Phi_n\) given as

\[
\Phi_n(x) = -H_{n-1}(x) \phi(x) \quad \text{for } n = 1, 2, \ldots
\]  

(6.5)

**Proof:** It follows from (6.1) that

\[
\frac{d}{dx}(H_n(x) \phi(x)) = \left(H'_n(x) - x H_n(x)\right) \phi(x) = -H_{n+1}(x) \phi(x).
\]

(6.6)

Hence for \(n \geq 1\) we have that

\[
\Phi_n(x) = \int_{-\infty}^{\infty} H_n(y) \phi(y) \, dy = -\int_{-\infty}^{x} \frac{d}{dy}(H_{n-1}(y) \phi(y)) \, dy
\]

\[
= -\left[H_{n-1}(y) \phi(y)\right]_{-\infty}^{x} = -H_{n-1}(x) \phi(x)
\]

using that \(x^k \phi(x) \to 0\) for \(x \to \pm \infty\) for any \(k\).

\[\square\]

Note how the distribution function for \(\mu_n\) is really the density of \(\mu_{n-1}\). As a simple application of this lemma, it follows that except for \(\mu_0\), all the Hermite measures have total mass 0, as

\[
\mu_n(\mathbb{R}) = \lim_{x \to \infty} \Phi_n(x) = \lim_{x \to \infty} -H_{n-1}(x) \phi(x) = 0 \quad \text{for } n = 1, 2, \ldots
\]
Our interest in Hermite measures stems from the fact that they have very simple characteristic functions:

\[ \xi_n(t) = (it)^n e^{-t^2/2} \quad \text{for all } t \in \mathbb{R} \]  

for \( n = 0, 1, 2, \ldots \).

**Theorem 6.4** The \( n \)'the Hermite measure has characteristic function

\[ \xi_n(t) = (it)^n e^{-t^2/2} \quad \text{for all } t \in \mathbb{R} \]  

**Proof:** We establish the recursive formula

\[ \xi_n(t) = it \xi_{n-1}(t) \quad \text{for } n = 1, 2, \ldots \]  

Combining with the well-known result that \( \xi_0 = e^{-t^2/2} \), the result (6.7) follows by induction.

Partial integration and (6.6) shows that

\[ \xi_n(t) = \int_{-\infty}^{\infty} e^{itx} H_n(x) \phi(x) \, dx = - \int_{-\infty}^{\infty} e^{itx} \frac{d}{dx} \left( H_{n-1}(x) \phi(x) \right) \, dx \]

\[ = - \left[ e^{itx} H_{n-1}(x) \phi(x) \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} it e^{itx} H_{n-1}(x) \phi(x) \, dx \]

\[ = 0 + it \int_{-\infty}^{\infty} e^{itx} H_{n-1}(x) \phi(x) \, dx \]

\[ = it \int_{-\infty}^{\infty} e^{itx} d\mu_{n-1}(x) \]

We will make the following use of this result: Let \( p(x) \) be an arbitrary polynomial of degree \( n \), say \( p(x) = \sum_{k=0}^{n} a_k x^k \). Then the signed measure \( \nu = \sum_{k=0}^{n} a_k \mu_k \) has characteristic function

\[ \xi(t) = \sum_{k=0}^{n} a_k \xi_k(t) = \sum_{k=0}^{n} a_k (it)^k e^{-t^2/2} \]

This is a very flexible class of characteristic functions, that can be used to mimic almost every other characteristic functions, at least on a compact interval. Combined
6.2. What are Edgeworth expansions?

with the Esseen smoothing lemma, it follows that combinations of Hermite measures are close to about any finite, positive measure on the real line. The details are not that easy, as the Esseen lemma require the approximating measure must have a density which is bounded. Well, the combination of Hermite measures do have a bounded density, but the bound varies - in particular it grows with the order of \( p \). So care has to be taken. But nonetheless, this is the fundamental idea behind the Edgeworth expansions.

6.2 What are Edgeworth expansions?

Let \((\nu_n)_{n \in \mathbb{N}}\) be a sequence of probability measures on \( \mathbb{R} \). An **Edgeworth expansion** of order \( N \) to this sequence is a collection of signed measures \( \lambda_0, \lambda_1, \ldots, \lambda_N \) such that the sequence

\[
\nu_n = \lambda_0 + \sum_{j=1}^{N} \frac{1}{n^{j/2}} \lambda_j \quad \text{for } n = 1, 2, \ldots
\]

satisfies a bound of the form

\[
\|\nu_n - \nu_n^*\| \leq \frac{C}{n^{(N+1)/2}} \quad \text{for all } n \in \mathbb{R}
\]

for a suitable constant \( C > 0 \). The norm occurring in this bound is the sup-norm for the corresponding distribution functions. Note that the number of terms in the expansion is fixed - the only things that are changing as \( n \) is going to infinity are the coefficients of the \( \lambda_j \)-measures. If such an expansion exists, it is unique. It is easy to see if all the \( \nu_n \)'s are probability measures, then \( \lambda_0 \) must be too - indeed in almost all cases of interest \( \lambda_0 \) will be the standard normal distribution. And one can check that the subsequent measures \( \lambda_1, \ldots, \lambda_N \) must all have total mass zero.

The idea behind such an Edgeworth expansion is that the approximating measure \( \nu_n^* \) are extremely close to \( \nu_n \), at least for large values of \( n \), but still they have a form where the role of \( n \) is clearly exhibited. There is a price to pay for the close approximation, though, and that is that the approximating measures \( \nu_n^* \) are not necessarily probability measures - the modifications of \( \lambda_0 \) involve signed measures, and even though these signed measures has total mass zero, they may be dominant in certain regions of the real line. So there may be sets of negative measure for \( \nu_n^* \) - this is in fact typical.

The Berry-Esseen theorem can be coined in this language as it represents an Edgeworth expansion of order 0. It is a special case, though, in the sense that the main
effort in the work around the Berry-Esseen theorem is devoted to controlling the constant $C$ in (6.10). In general not much attention is paid to this constant - it will usually depend strongly on the specific sequence $(\nu_n)_{n \in \mathbb{N}}$, but in a non-transparent way.

One may wonder why the expansion is in integral powers of $n^{1/2}$ - integral powers of $n$ seems more aesthetic to most people. But this is a question of applicability: it turns out that expansions in powers of $n$ seldom exists, while expansions in powers of $n^{1/2}$ can be constructed in a broad array of situations.

If we have an Edgeworth expansion of order $N$, it is easy to obtain an expansion of order $N - 1$ - we simply delete the $N$-term from the expansion. Combined with the uniqueness, we can go the other way and try to create expansions of high order by adding terms to an expansion of lower order. Sometimes this can be achieved for all values of $N$. In that case one may be tempted to conclude that we get a close approximation to a fixed $\nu_n$ by adding as many terms to the expansion as possible. It is important to understand that this line of thinking is highly erroneous. It may well be that as more and more terms are added to the expansion, we get progressively worse approximations to $\nu_n$. The point is that the good approximations only occur for 'large' values of $n$. And what 'large' means, may change as we increase the order of the expansion.

Edgeworth expansions belong to a mathematical domain known as **divergent series**. Such series have appeared in mathematics ever since the invention of infinite series, but they have always been a mystery. They were useful in the hands of the semigods of analysis in the 18'th century (in particular Euler), but lesser minds made horrible errors relying on them. The most famous early divergent series is no doubt Stirlings formula,

$$
\log n! = \left(n + \frac{1}{2}\right) \log n - n + \log \sqrt{2\pi n} + \frac{c_1}{n} + \frac{c_3}{n^3} + \frac{c_5}{n^5} + \ldots
$$

where Stirling himself in 1730 published the first 5 coefficients. The dots represent the diverging phenomenon just mentioned. For fixed $n$ the series will diverge if we keep adding terms. But for a fixed number of terms, the agreement between $\log n!$ and the expansion becomes extremely good as $n \to \infty$.

Divergent series were quite explicitly banned when Cauchy and his contemporaries installed modern rigour in mathematics in the early 19’th century. But nobody could deny their usefulness in high precision numeric work, and they were an indespensable tool for constructing tables of various functions. In particular they continued to
play a prominent role in astronomy, where there was a need for very close agreement between measurements and theory - one of the key questions in astronomy has always been if the solar system is stable under the actions of gravity, and this is not a question that can be answered by using two digits of precision in the computations. . . . Edgeworth himself was an astronomer, and his introduction of expansions in probability theory around World War I actually followed the astronomers tradition quite closely, for instance by not paying attention to explicit error bounds like (6.10). It was only later, in the hands of Esseen, that rigour was installed in the subject.

Not every sequence of measures will have an Edgeworth expansion of a specified order, and it is a non-trivial task to identify the sequences that have. Sometimes candidates for \(\lambda_1, \ldots, \lambda_N\) can be found by formal manipulations, where all the small approximation errors made in the process are simply disregarded - this is what happens when we leave part of the expressions as 'dots'. Such a candidate expansions are frequently referred to as formal Edgeworth expansion. Proving that the formal expansion possesses the relevant approximation properties is considered one of the hardest tasks in classical probability theory.

On the other hand, we have made things look harder than they are, as we have pretended that we would be interested in expansions of high order. The expansions are not used for numerical work anymore, they are used for qualitative arguments. And nobody has ever figured out how to make use of more than two correction terms to the limit measure \(\lambda_0\) in such qualitative arguments.

Furthermore, in the case of sums of iid variables, it turns out that good candidates for the \(\lambda_j\)'s can be constructed as linear combinations of Hermite measures. So we are not starting out from scratch.

**Example 6.5** To illustrate these ideas, we will find a formal Edgeworth expansion of order 2 for (the distribution of) \(Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\) where \(X_1, X_2, \ldots\) are independent and identically distributed real stochastic variables with \(E X_i = 0\) and \(E X_i^2 = 1\). If you have sound mathematical instincts, you may get airsick during the ride - but please keep along. We will justify the whole exercise afterwards, when we give a rigorous proof that the formal expansion we find, indeed satisfies a bound of the form (6.10).

Let \(K(t)\) be the cumulant generating function for \(X_i\). We know that

\[
K(t) = -\frac{t^2}{2} + \kappa_3 \frac{(it)^3}{3!} + \kappa_4 \frac{(it)^4}{4!} + \ldots
\]
where \( \kappa_3 \) and \( \kappa_4 \) are the third and fourth cumulant. The cumulant generating function \( K_n \) for \( Y_n \) satisfies that

\[
K_n(t) = n K \left( \frac{t}{\sqrt{n}} \right)
\]

and combining these expressions we get

\[
K_n(t) = -\frac{t^2}{2} + \kappa_3 \frac{(it)^3}{3!} \frac{1}{\sqrt{n}} + \kappa_4 \frac{(it)^4}{4!} \frac{1}{n} + \ldots
\]

where the dots now represents terms of order \( n^{-3/2} \) or higher - and we do not want these terms in the final second order expansion, so we keep them semi-invisible. The characteristic function for \( Y_n \) is

\[
\phi_n(t) = \exp \left( -\frac{t^2}{2} \right) \exp \left( \kappa_3 \frac{(it)^3}{3!} \frac{1}{\sqrt{n}} + \kappa_4 \frac{(it)^4}{4!} \frac{1}{n} + \ldots \right)
\]

Using the power series for the exponential function, the last factor equals

\[
1 + \left( \kappa_3 \frac{(it)^3}{3!} \frac{1}{\sqrt{n}} + \kappa_4 \frac{(it)^4}{4!} \frac{1}{n} + \ldots \right) + \frac{1}{2} \left( \kappa_3 \frac{(it)^3}{3!} \frac{1}{\sqrt{n}} + \kappa_4 \frac{(it)^4}{4!} \frac{1}{n} + \ldots \right)^2 + \ldots
\]

where the last dots represents the observation that the argument to the exponential functions is itself of order \( n^{-1/2} \), so raising it to a power 3 or above gives a term of the non-desired order \( n^{-3/2} \). Multiplying out the square term, and again leaving out all terms of order \( n^{-3/2} \) or higher, we obtain that

\[
\phi_n(t) = \exp \left( -\frac{t^2}{2} \right) \left( 1 + \frac{\kappa_3}{3!} \frac{(it)^3}{\sqrt{n}} + \frac{\kappa_4}{4!} \frac{(it)^4}{n} + \frac{\kappa_3^2}{2 3!^2} \frac{1}{n} (it)^6 + \ldots \right)
\]

Except for the dots, this is exactly the characteristic function for

\[
\nu_n = \mu_0 + \frac{\kappa_3}{3!} \frac{\mu_3}{\sqrt{n}} + \frac{\kappa_4}{4!} \frac{\mu_4}{n} + \frac{\kappa_3^2}{2 3!^2} \frac{1}{n} \mu_6
\]

Hence we conjecture that the second order Edgeworth expansion - if it exists - is given by

\[
\lambda_0 = \mu_0 \, , \quad \lambda_1 = \frac{\kappa_3}{3!} \mu_3 \, , \quad \lambda_2 = \frac{\kappa_4}{4!} \mu_4 + \frac{\kappa_3^2}{2 3!^2} \mu_6
\]

This turns out to be right, as we shall see shortly, at least if we add certain regularity conditions. But the computations so far has probably not convinced anybody. To the contrary, we see that \( \lambda_2 \) has contributions from several terms in the power series for
6.2. What are Edgeworth expansions?

the exponential function, which shows how hard it is to treat the dots right. It is scary how easily a term could be forgotten, even though it had an order that should make it included.

But if this anxiety can be overcome, the technique is in fact applicable to produce expansions of arbitrary high order - as long as we do not have to prove anything about their approximation properties. The reader is challenged to see that if we include a \( r^5 \)-term in the cumulant function \( K(t) \), and if we include a cube term in the power series expansion of the exponential function, a third order expansion for \( Y_n \) can be produced with

\[
\lambda_3 = \frac{k_5}{5!}\mu_5 + \frac{k_3k_4}{3!4!}\mu_7 + \frac{k_3^3}{3!^3}\mu_9 .
\]

(6.13)

And keeping a further term in all the computations gives a fourth order expansion for \( Y_n \) with

\[
\lambda_4 = \frac{k_6}{6!}\mu_6 + \left( \frac{k_4^2}{2\cdot4!^2} + \frac{k_3k_5}{3!5!} \right)\mu_8 + \frac{k_3^2k_4}{2\cdot3!^2\cdot4!}\mu_{10} + \frac{k_3^4}{4!^3\cdot3!}\mu_{12} .
\]

(6.14)

After that, combinatorial problems tends to complicate things too much for hand-computations. The problem is in the expansion of the exponential function, where we have to take a sum of many terms and raise it to a high power - it is difficult to see what happens.

Example 6.6 Let \( X_1, X_2, \ldots \) be independent and exponentially distributed with mean 1. We consider the standardized averages

\[
Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - 1)
\]

In most examples it is not possible to find the distribution of standardized averages explicitly, but in this particular case we see that it is \( \Gamma \)-distributed with form \( n \) and scale 1, and with a position parameter of \( -\sqrt{n} \). We will compare this distribution to the Edgeworth approximations of order 0, 1 and 2 - the 0’th order approximation being simply the standard normal.

The Edgeworth approximations are are extremely easy to compute from 6.12 and the table of Hermite polynomials. The difficulties with this kind of theory is in understanding what the whole thing means, and in developing general formulas for the approximations. To apply the approximations, on the other hand, is a breeze. We just
Figure 6.1: Comparison of standardized means of exponentially distributed variables with the standard normal. The thin line is the density of the approximation.

need to plug in the values of the cumulants, \( \kappa_3 = 2 \) and \( \kappa_4 = 6 \). The comparison is done on figure 6.1, 6.2 and 6.3.

We observe that quality of the normal approximations are not impressive for these small values of \( n \). This is as expected, since the exponential distribution is very far from normal, so it will take a while for the forces behind CLT to take over. But already the first Edgeworth approximation makes the problem go away, and the second Edgeworth approximation is almost indistinguishable from the true distribution for \( n \) as low as 7. Well, there is a minor visible difference around \( x = -2 \), where the approximate density is slightly negative. Such negative densities are the price we pay for the astounding quality of the approximation.

The explicit formulas for the Edgeworth approximations to \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \) for iid summands are too tedious to work with in many applications. The main thing is usually the existence, and a few collected qualitative facts. We may note that all the measures in the expansion have the form

\[ \lambda_j = p_j(x) \phi(x) \cdot m \]

where \( p_j \) is a polynomial. It is indeed a linear combination of Hermite polynomials,
6.2. What are Edgeworth expansion?

and the only thing specific to the $X_i$’s are the coefficients of the linear combinations. A general feature which can be read off from the formulas (and a little handwaving, as to how things will develop for higher order) is that $p_j$ involves the Hermite polynomials $H_{j+2}, H_{j+4}, \ldots, H_{3j}$. Two facts emerge:

1) The degree of $p_j$ is at most $3j$, and it is exactly $3j$ unless a great many cumulants happen to be 0.

2) If $j$ is even, only even Hermite polynomials enter in $p_j$, and so $p_j$ itself is an even function. Similarly, if $j$ is odd, only odd Hermite polynomials enter in $p_j$, and so $p_j$ itself is an odd function.

Edgeworth expansions can be shown to exist in many other situations than averages of iid variables. Notably there is a delta formula for Edgeworth expansion. One often encounters variables

$$Y_n = \sqrt{n} g \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)$$

where $X_1, X_2, \ldots$ are iid variables with mean 0 and variance 1, and where $g : \mathbb{R} \to \mathbb{R}$ is a $C^1$-map with $g(0) = 0$ and $g'(0) \neq 0$. In such situations one can derive Edge-
Chapter 6. Edgeworth expansions

Figure 6.3: Comparison of standardized means of exponentially distributed variables with the second order Edgeworth expansion. The thin line is the density of the approximation.

Edgeworth expansions for $Y_n$ from the Edgeworth expansion of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$. The measure involved will also in this general case have the form $\lambda_j = p_j(x) \phi(x) \cdot m$ where $p_j$ is a polynomial, and the polynomials can again be written as linear combinations of Hermite polynomials, but the coefficients will now be complicated: they will mix cumulants of the underlying $X$'es with the higher order derivatives of $g$ in zero. There is no reason why the coefficient should relate to the true cumulants of $Y_n$, and in many cases they will not. They are known as formal cumulants, and do in general not reflect the moments of $Y_n$.

Actually, this is well known already for the normal approximations, where the asymptotic mean and variance occurring in the delta method normal approximation in many cases are unrelated to the true moments of $Y_n$ - indeed there are many situations where the delta method can be used to prove asymptotic normality for variables that do not even have first moment! Be that as it may, it will still hold true that $p_j(x)$ is odd or even as $j$ is odd or even.

We will now give a rigorous proof that Edgeworth expansion found by formal manipulations in example 6.5 does indeed satisfy an approximation inequality of the form (6.10) if we add sufficiently many regularity conditions. The proof is lengthy and somewhat technical, but the details are important. With sufficient grasp of these,
6.2. What are Edgeworth expansion?

it is possible to adapt the proof to new situations, in particular to the stochastic Edge-
worth expansions we will encounter in the analysis of the finer details of the boot-
strap. We start by covering some ground with a number of lemmas on the behaviour
of characteristic functions and of the complex exponential function.

**Definition 6.7** A real-valued stochastic variable $X$ is said to have a **lattice distribution** if some affine transformation $Y = a + bX$ (with $b \neq 0$) satisfies that $P(Y \in \mathbb{Z}) = 1$.

If the distribution of $X$ is not of this lattice-type, then $X$ is said to be **non-lattice**. It is a frequently occurring condition in asymptotic theory that certain distributions are non-lattice. This is because lattice distributions violates the Cramér condition in a quite spectacular way, as we shall see.

If $P(Y \in \mathbb{Z}) = 1$, then the characteristic function satisfies

$$
\phi_Y(t + 2\pi) = \sum_{k=-\infty}^{\infty} e^{i(t+2\pi)k} P(Y = k) = \sum_{k=-\infty}^{\infty} e^{i2\pi k} e^{ikt} P(Y = k)
$$

$$
= \sum_{k=-\infty}^{\infty} e^{ikt} P(Y = k) = \phi_Y(t)
$$

So $\phi_Y$ is periodic with period $2\pi$. In particular $\phi_Y(k \cdot 2\pi) = \phi_Y(0) = 1$ for all $k \in \mathbb{Z}$.

If $X$ is of lattice type, we may write $X = c + dY$ for a nonzero $d$, where $Y$ is concen-
trated on the integers. Hence $X$ has characteristic function

$$
\phi_X(t) = e^{ic} \phi_Y(dt)
$$

which implies that

$$
\phi_X \left( \frac{k \cdot 2\pi}{d} \right) = e^{ic \cdot 2\pi/d} \phi_Y(k \cdot 2\pi) = e^{ic \cdot 2\pi/d} \quad \text{for all } k \in \mathbb{Z}.
$$

In particular it holds that

$$
\left| \phi_X \left( \frac{k \cdot 2\pi}{d} \right) \right| = 1 \quad \text{for all } k \in \mathbb{Z}
$$

So $\phi_X(t)$ hits the unit circle over and over again, on each point of an equidistant lattice.
Violation of Cramér’s condition are usually much more subtle than banging into the unit circle again again. Indeed, we show now that if the characteristic function $\phi(t)$ hits the unit circle anywhere (except in the trivial point $t = 0$) then the distribution is of lattice type.

**Lemma 6.8** A stochastic variable $X$ on $(\mathbb{R}, \mathcal{B})$ has a lattice distribution if and only if the characteristic function satisfies that

$$|\phi(t_0)| = 1 \quad \text{for some } t_0 \neq 0$$

(6.15)

*Proof:* We have already established that the characteristic function of a distribution of lattice type satisfies (6.15). Indeed, the characteristic function of such a distribution hits the unit circle not only once, but on each point of an equidistant lattice.

To prove the opposite implication, suppose that $X$ is a stochastic variable with distribution function $\phi_X(t)$, such that $|\phi_X(t_0)| = 1$ for some $t_0 \neq 0$. The affine transformation $Y = a + bX$ will have a characteristic function satisfying

$$\phi_Y(1) = e^{ia} \phi_X(b)$$

Choosing $b = t_0$ and $a$ such that $e^{-ia} = \phi_X(t_0)$ we see that $\phi_Y(1) = 1$. Considering only the real part of the characteristic function, this amounts to

$$\int \cos Y \, dP = 1.$$ 

But $\cos Y \leq 1$, so the only way the integral can be equal to 1 is if $\cos Y = 1$ almost surely. We conclude that

$$P(Y \in 2\pi \mathbb{Z}) = 1.$$ 

So $Y$ - and thus $X$ - is of lattice type.

\[ \square \]

**Lemma 6.9** For any two complex numbers $z$ and $w$ and any $r > 0$ it holds that

$$|e^z - e^w| \leq e^r |z - w| \quad \text{if } |z| \leq r, \ |w| \leq r.$$
6.2. What are Edgeworth expansions?

Proof: For any $k \in \mathbb{N}$ it follows by telescoping that

$$z^k - w^k = (z - w) \sum_{j=0}^{k-1} z^j w^{k-1-j}$$

When $|z|, |w| \leq r$ this leads to the estimate

$$|z^k - w^k| \leq |z - w| k r^{k-1}$$

From the power series of the exponential function we see that

$$|e^z - e^w| = \left| \left( 1 + \sum_{k=1}^{\infty} \frac{z^k}{k!} \right) - \left( 1 + \sum_{k=1}^{\infty} \frac{w^k}{k!} \right) \right| \leq \sum_{k=1}^{\infty} \frac{|z-w| k r^{k-1}}{k!} = |z-w| \sum_{k=1}^{\infty} \frac{r^{k-1}}{(k-1)!}$$

and the result follows by shifting the index of the sum.

\[ \square \]

Lemma 6.10 For every complex number $z$ and every $k \in \mathbb{N}$ it holds that

$$\left| e^z - \sum_{j=0}^{k} \frac{z^j}{j!} \right| \leq e^{|z|} \frac{|z|^{k+1}}{(k+1)!}.$$  

Proof: We see that

$$\left| e^z - \sum_{j=0}^{k} \frac{z^j}{j!} \right| = \left| \sum_{j=k+1}^{\infty} \frac{z^j}{j!} \right| = \frac{|z|^{k+1}}{(k+1)!} \sum_{j=0}^{\infty} \frac{(k+1)!}{(k+1+j)!} |z|^j$$

The result follows from observing that

$$(k+1+j)! = (k+1)! \prod_{\ell=1}^{j} (k+1+\ell) \geq (k+1)! \prod_{\ell=1}^{j} \ell = (k+1)! j!$$

so that

$$\left| e^z - \sum_{j=0}^{k} \frac{z^j}{j!} \right| \leq \frac{|z|^{k+1}}{(k+1)!} \sum_{j=0}^{\infty} \frac{|z|^j}{j!}$$

\[ \square \]
Theorem 6.11  Let $X_1, X_2, \ldots$ be independent and identically distributed real stochastic variables. Assume that $E|X|^5 < \infty$, and that the distribution satisfies Crâmer’s condition. Assume also that $E X_i = 0$ and that $E X_i^2 = 1$. Let $\nu_n$ be the distribution of

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i,$$

and let

$$\lambda_0 = \mu_0, \quad \lambda_1 = \frac{\kappa_3}{3!} \mu_3, \quad \lambda_2 = \frac{\kappa_4}{4!} \mu_4 + \frac{\kappa_3^2}{2 \cdot 3!} \mu_6$$

where the $\mu_m$’s are the Hermite measures and where $\kappa_3$ and $\kappa_4$ are the third and fourth cumulant for $X_i$. Let

$$\nu_n^* = \lambda_0 + \frac{\lambda_1}{\sqrt{n}} + \frac{\lambda_2}{n}.$$

There exists a $C > 0$ such that

$$\|\nu_n - \nu_n^*\| \leq \frac{C}{n^{3/2}} \quad \text{for all } n \in \mathbb{N}. \quad (6.16)$$

Proof: Let $\phi_n$ be the characteristic function associated to $\nu_n$, and let $\gamma_n$ be the characteristic function associated to $\nu_n^*$. The proof progresses by controlling $|\phi_n(t) - \gamma_n(t)|$. For this to work, we divide the real axis into five regions, as shown on figure 6.4.

<table>
<thead>
<tr>
<th>Esseen</th>
<th>Cramér</th>
<th>Analysis</th>
<th>Cramér</th>
<th>Esseen</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-n^{3/2}$</td>
<td>$-c \sqrt{n}$</td>
<td>0</td>
<td>$c \sqrt{n}$</td>
<td>$n^{3/2}$</td>
</tr>
</tbody>
</table>

Figure 6.4: An illustration of the strategy behind the proof of theorem 6.11. We control the difference between the two characteristic functions $\phi_n$ and $\gamma_n$ in each of the five intervals above. We do this by analytic arguments, by Esseen’s smoothing lemma and by arguments based on Cramér’s condition.

It is no real surprise that the formal calculations of example 6.5 can be made into rigorous bounds for $|\phi_n(t) - \gamma_n(t)|$ for $t$ close to zero. The details are hard core analysis, but it would be very surprising if the idea could not be made to work at all. However, it turns out that the analytic arguments are only really successful in an interval of the form $(-c \sqrt{n}, c \sqrt{n})$, and only so if $c > 0$ is small enough.
6.2. What are Edgeworth expansion?

The two outer domains, corresponding to $|t| > n^{3/2}$, are dealt with by the Esseen smoothing lemma, meaning that we don’t have to control $|\phi_n(t) - \gamma_n(t)|$ in these domains at all. This leaves the two intermediate regions, where $c \sqrt{n} \leq |t| \leq n^{3/2}$. This bit requires ingenuity, and it is for this purpose that we require the Cramér condition.

We start by making sure that Esseens smoothing lemma really handles the two outer domains. Observe that $n$ is positive, and that $\gamma_n$ has a density with respect to Lebesgue measure. If we let $j = \sup_{x \in \mathbb{R}} H(x) H(-x)$ for $j = 0, 1, \ldots$ then we see that the density for $\gamma_n$ is bounded by

$$C_1 = \alpha_0 + \frac{|\kappa_3| \alpha_3}{3!} + \frac{|\kappa_4| \alpha_4}{4!} + \frac{\kappa_3^2 \alpha_6}{24!}$$

for all $n$. Hence Esseens lemma tells us that

$$\|\gamma_n - \gamma_n^d\| \leq \frac{1}{\pi} \int_{-n^{3/2}}^{n^{3/2}} |\phi_n(t) - \gamma_n(t)| \frac{dt}{t} \leq \frac{48 C_1}{\pi n^{3/2}}$$

for all $n \in \mathbb{N}$. (6.17)

The last term on the right hand side has the desired $O(n^{-3/2})$-magnitude, so all we have to do is to bound the integral with a bound of the right order of magnitude.

For the moment we leave the intermediate intervals alone, and focus on $|t| < c \sqrt{n}$. We will show that if $c$ is chosen small enough, then

$$|\phi_n(t) - \gamma_n(t)| \leq \frac{r^2 p(t) e^{-r^2/4}}{n^{3/2}}$$

when $|t| \leq c \sqrt{n}$ (6.18)

where $p(t)$ is a non-negative polynomial. If (6.18) holds, then

$$\int_{-c \sqrt{n}}^{c \sqrt{n}} \frac{|\phi_n(t) - \gamma_n(t)|}{t} dt \leq \int_{-c \sqrt{n}}^{c \sqrt{n}} |t| p(t) e^{-r^2/4} dt \leq \frac{1}{n^{3/2}} \int_{-\infty}^{\infty} |t| p(t) e^{-r^2/4} dt$$

(6.19)

and since the last integral is finite, this is again a bound of the desired $O(n^{-3/2})$-magnitude.

So let us establish (6.18). Since $X_i$ has 5th moment, its characteristic function and hence its cumulant generating function $K(t)$ is $C^5$. Taylors theorem shows that we can find $\delta > 0$ and $C_2 > 0$ such that

$$\left| K(t) - \left( - \frac{t^2}{2} + \frac{\kappa_3 (it)^3}{3!} + \frac{\kappa_4 (it)^4}{4!} \right) \right| \leq C_2 |t^5|$$

for $|t| < \delta$. (6.20)
Chapter 6. Edgeworth expansions

Applying this rigorous bound to the cumulant generating function $K_n(t) = nK(t/\sqrt{n})$ for $Y_n$, and introducing the abbreviation

$$\beta_n(t) = \kappa_3 \frac{(it)^3}{3!} \frac{1}{\sqrt{n}} + \kappa_4 \frac{(it)^4}{4!} \frac{1}{n},$$

(6.21)

we see that

$$\left| K_n(t) - \left( -\frac{t^2}{2} + \beta_n(t) \right) \right| \leq C_2 \frac{|t|^5}{n^{3/2}} \quad \text{when } |t| \leq \delta \sqrt{n}$$

Let us consider the three magnitudes

$$\begin{align*}
\text{(1)} & = \left| \phi_n(t) - e^{-t^2/2} e^{\beta_n(t)} \right| \\
\text{(2)} & = \left| e^{-t^2/2} e^{\beta_n(t)} - e^{-t^2/2} \left( 1 + \beta_n(t) + \frac{\beta_n(t)^2}{2} \right) \right| \\
\text{(3)} & = \left| e^{-t^2/2} \left( 1 + \beta_n(t) + \frac{\beta_n(t)^2}{2} \right) - \gamma_n(t) \right|
\end{align*}$$

The triangle inequality trivially shows that

$$|\phi_n(t) - \gamma_n(t)| \leq \text{(1)} + \text{(2)} + \text{(3)}$$

so if we can estimate each of these three terms with a bound of the form of the right hand side of (6.18), such a bound will also apply to $|\phi_n(t) - \gamma_n(t)|$.

The term $\text{(1)}$ is essentially controlled by lemma 6.9. We see that if $|t| \leq c \sqrt{n}$ then

$$|\beta_n(t)| \leq \frac{|\kappa_3| |t|^3}{3! \sqrt{n}} + \frac{|\kappa_4| |t|^4}{4! n} \leq \left( \frac{\kappa_3}{3!} + \frac{c |\kappa_4|}{4!} \right) \frac{|t|^3}{\sqrt{n}}$$

(6.22)

If $c$ is so small that

$$\left( \frac{\kappa_3}{3!} + \frac{c |\kappa_4|}{4!} \right) c \leq \frac{1}{8}$$

(6.23)

it holds that

$$|\beta_n(t)| \leq \frac{t^2}{8} \quad \text{for } |t| < c \sqrt{n}$$

(6.24)

We also have that

$$\left| K_n(t) + \frac{t^2}{2} - \beta_n(t) \right| \leq C_2 \frac{|t|^5}{n^{3/2}} \leq C_2 c^3 \frac{t^2}{8} \leq \frac{t^2}{8} \quad \text{for } |t| < c \sqrt{n}$$
6.2. What are Edgeworth expansion?

if \( c \) is chosen so small that \( C_2 c^3 \leq \frac{1}{8} \). Combining these two informations, we see that

\[
\left| K_n(t) + \frac{t^2}{2} \right| \leq \frac{t^2}{4}, \quad |\beta_n(t)| \leq \frac{t^2}{4} \quad \text{for } |t| < c \sqrt{n}
\]

where the last inequality is a trivial weakening of (6.24). So lemma 6.9 shows that

\[
\begin{align*}
\| & \beta_n(t) - e^{-r^2/2} e^{r^2/2} \left| e^{K_n(t) + r^2/2} - e^{\beta_n(t)} \right| \\
& \leq e^{-r^2/2} e^{r^2/4} \left| K_n(t) + \frac{t^2}{2} - \beta_n(t) \right| \\
& \leq e^{-r^2/4} C_2 \frac{|t|^3}{n^{3/2}}
\end{align*}
\]

for all \( |t| \leq c \sqrt{n} \). For aesthetic reasons we might replace \( |t|^5 \) with \( t^2 (1 + t^4) \) to have a positive polynomium in \( t \), not in \( |t| \), but this does not really affect the argument.

As for \( \circledast \), we observe from lemma 6.10 and (6.22) that

\[
\begin{align*}
\| & e^{-r^2/2} e^{r^2/8} e^{\beta_n(t)} - e^{-r^2/2} \left( 1 + \beta_n(t) + \frac{\beta_n(t)^2}{2} \right) \| \\
& \leq e^{-r^2/2} e^{r^2/8} \left( \frac{k_3}{3!} + \frac{c |k_4|}{4!} \right)^3 \frac{|t|^9}{n^{3/2}}
\end{align*}
\]

for all \( |t| \leq c \sqrt{n} \), which again is a bound of the desired kind - perhaps after replacing \( e^{-r^2/8} \) by \( e^{-r^2/4} \) and \( |t|^9 \) by \( t^2 (1 + t^4) \).

As for \( \circledast \), we simply calculate the square of \( \beta_n(t) \) and plug in. This gives

\[
1 + \beta_n(t) + \frac{\beta_n(t)^2}{2} = 1 + \frac{k_3 (it)^3}{3! \sqrt{n}} + \frac{k_4 (it)^4}{4! n} + \frac{k_3^2 (it)^6}{2 3! n^2} + \frac{k_3 k_4 (it)^7}{3! 4! n^{3/2}} + \frac{k_4^2 (it)^8}{2 4!^2 n^2}
\]

which implies that

\[
\begin{align*}
\| & e^{-r^2/2} \left( 1 + \beta_n(t) + \frac{\beta_n(t)^2}{2} \right) - \left( 1 + \frac{k_3 (it)^3}{3! \sqrt{n}} + \frac{k_4 (it)^4}{4! n} + \frac{k_3^2 (it)^6}{2 3! n^2} \right) \| \\
& \leq e^{-r^2/2} \left( \frac{k_3 k_4 (it)^7}{3! 4! n^{3/2}} + \frac{k_4^2 (it)^8}{2 4!^2 n^2} \right) \frac{|t|^7}{n^{3/2}}
\end{align*}
\]
for $|t| \leq c \sqrt{n}$. This is again of the desires form, perhaps after replacing $e^{-t^2/2}$ by $e^{-t^2/4}$ and $|t|^7$ by $t^2(1 + t^6)$.

Collecting the estimates of $\hat{\theta}$, $\hat{\tau}$ and $\hat{\sigma}$ we see that we indeed have established (6.18). What remains to be done is to control in intermediate regions, in form of an estimate
\[
\int_{c \sqrt{n}}^{n^{3/2}} \frac{\phi_n(t) - \gamma_n(t)}{t} dt \leq \frac{C_3}{n^{3/2}}
\]
and similarly with the integral over $[-n^{3/2}, -c \sqrt{n}]$. This is done by appealing to Cramér's condition. There is some $K$ and some $\eta \in (0, 1)$ such that the characteristic function $\phi$ for $X_i$ satisfies that
\[
|\phi(t)| \leq 1 - \eta \quad \text{if } |t| \geq K.
\]
We would like this inequality to hold for $|t| \geq c$. Probably $K$ is larger that $c$ - very likely it is a whole lot larger. But should this be the case, we can get around the difficulty by observing that Cramér's condition also implies that $X_i$ is not of lattice type, and so
\[
|\phi(t)| < 1 \quad \text{for all } |t| \in [c, K]
\]
Appealing to the continuity of $\phi$, we have that $|\phi(t)|$ has a maximum over the union of the two compact sets, and this maximum is smaller than one. And so we see that there is a $\epsilon \in (0, 1)$ such that
\[
|\phi(t)| \leq 1 - \epsilon \quad \text{if } |t| \geq c.
\]
Applying this to $\phi_n(t) = \phi(t/ \sqrt{n})^n$ we see that
\[
|\phi_n(t)| \leq (1 - \epsilon)^n \quad \text{if } |t| \geq c \sqrt{n}.
\]
The same argument does not apply to $\gamma_n$, but simpler estimates will do. We see that
\[
|\gamma_n(t)| = \left| e^{-t^2/2} \left( 1 + \frac{\kappa_3}{3! \sqrt{n}} (i t)^3 + \frac{\kappa_4}{4!} (i t)^4 + \frac{\kappa_5^2}{2 \cdot 3!^2} (i t)^6 \right) \right|
\leq e^{-t^2/4} \frac{\kappa_3}{3!} \frac{1}{\sqrt{n}} (i t)^3 + \frac{\kappa_4}{4!} (i t)^4
\leq C_4 e^{-t^2/4}
\]
for a suitably chosen $C_4$ dominating the product of $e^{-t^2/4}$ and the polynomiun. This inequality holds for every $t$. But if $\epsilon$ is chosen so small that $e^{-t^2/4} < 1 - \epsilon$, the we in particular obtain that
\[
|\gamma_n(t)| \leq C_4 e^{-t^2/4} \leq C_4 (1 - \epsilon)^n \quad \text{for } |t| \geq c \sqrt{n}
\]
6.2. What are Edgeworth expansion?

Hence (6.25) can be obtained as follows:

\[ \int_{c \sqrt{n}}^{\sqrt{n}/2} \left| \frac{\phi_n(t) - \gamma_n(t)}{t} \right| dt \leq \int_{c \sqrt{n}}^{\sqrt{n}/2} \frac{(1 + C_4) (1 - \varepsilon)^n}{c \sqrt{n}} dt \]

\[ \leq \frac{1 + C_4}{c} (1 - \varepsilon)^n n \]

And clearly

\[ (1 - \varepsilon)^n n \leq \frac{C_5}{n^{1/2}} \text{ for } n \in \mathbb{N} \]

for some sufficiently large \( C_5 \), since the exponential goes faster to zero than any power function.

\[ \square \]

**Corollary 6.12** Under the same assumptions as theorem 6.16, the standard normal approximation to \( \nu_n \) satisfies

\[ \sqrt{n} ||\nu_n - \lambda_0|| \to \frac{|k_3|}{3! \sqrt{2\pi}} \text{ for } n \to \infty. \]

**Comment:** The proof below is based on a strong second order Edgeworth expansion, and as this is way too much for the present purpose, we have in a way imposed too strong conditions. A weak first order Edgeworth expansion will do, and that can be derived under milder assumptions on the distribution of \( X_i \). The existence of 5th moment can be weakened to the existence of 3rd moment, and the Cramér condition can be weakened to that the distribution of \( X_i \) should be non-lattice.

**Proof:** Using the same notation as in the theorem, we may write

\[ \nu_n = \lambda_0 + \frac{\lambda_1}{\sqrt{n}} + \frac{\lambda_2}{n} + \xi_n \]

where the remainder term \( \xi_n \) satisfies that \( n^{3/2} ||\xi_n|| \) is bounded. This implies that

\[ \sqrt{n} (\nu_n - \lambda_0) \to \lambda_1 \text{ for } n \to \infty, \]

and in particular

\[ \sqrt{n} ||\nu_n - \lambda_0|| \to ||\lambda_1|| = \frac{|k_3| ||\mu_3||}{3!} \text{ for } n \to \infty. \]
What is left, is to identify \( \| \mu_3 \| \). But using lemma 6.3 we see that

\[
\| \mu_3 \| = \sup_{x \in \mathbb{R}} |H_2(x)\phi(x)|
\]

The function we want to maximize goes to 0 at infinity, so standard calculus tells us that the above supremum is attained in one of the stationary points. That is, in one of the zeroes of \( H_3 \). These are easily identified as 0 and \( \pm \sqrt{3} \). Plugging these points in, we obtain that

\[
\| \mu_3 \| = \| H_2(0)\phi(0) \| = \frac{1}{\sqrt{2\pi}}.
\]

This result shows that if there is skewness present in the distribution of \( X_i \), meaning that the third cumulant is non-zero, then the speed at which \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \) converges to a normal distribution, is exactly of the order \( \frac{1}{\sqrt{n}} \). This is so to speak the rate at which skewness disappears in the averages.

If \( \kappa_3 = 0 \), faster convergence may appear. If \( \kappa_4 \neq 0 \), the rate of convergence is essentially \( \frac{|\kappa_4|}{n} \), and so we may say that the kurtosis disappears at a rate of \( \frac{1}{n} \). But we may note that the \( \frac{1}{n} \)-term in the Edgeworth-expansion not only consists of a term controlled by the kurtosis, but also has a second-order effect of the skewness.