MATH 235A – Probability Theory
Lecture Notes, Fall 2009
Part III: The central limit theorem

Dan Romik
Department of Mathematics, UC Davis

Draft (December 4, 2009)

Lecture 10: Stirling's formula and the de Moivre-Laplace theorem

We start with a motivating example that was also historically the first instance in which the phenomenon that came to be known as the Central Limit Theorem was observed. Let $X_1, X_2, \ldots$ be an i.i.d. of $\mathrm{Binom}(1,p)$ random variables, and let $S_n = \sum_{k=1}^n X_k$, a r.v. with distribution $\mathrm{Binom}(n,p)$.

Theorem 1 (The de Moivre-Laplace theorem)**.** *For any* $t \in \mathbb{R}$,

$$\mathbf{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le t\right) \xrightarrow[n\to\infty]{} \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2}\, dx.$$

Since this is such a concrete example, the proof will simply require us to estimate a sum of the form $\sum_{0\le k\le t} \binom{n}{k} p^k (1-p)^{n-k}$. Knowing how to estimate such sums is a useful skill in its own right. Since the binomial coefficients are involved, we also need some preparation related to Stirling's formula.

Lemma 2. *The limit* $C = \lim_{n\to\infty} \frac{n!}{\sqrt{n}(n/e)^n}$ *exists.*

Proof.

$$\begin{aligned}
\log n! &= \sum_{k=1}^{n} \log k = \sum_{=1}^{n} \int_1^k \frac{dx}{x} = \int_1^n \frac{n - \lfloor x \rfloor}{x}\, dx \\
&= \int_1^n \frac{n + \frac{1}{2} + (\{x\} - \frac{1}{2}) - x}{x}\, dx = (n + 1/2)\log n - n + 1 + \int_1^n \frac{\{x\} - \frac{1}{2}}{x}\, dx \\
&= (n + 1/2)\log n - n + 1 + \int_1^\infty \frac{\{x\} - \frac{1}{2}}{x}\, dx + o(1),
\end{aligned}$$

where the last integral converges because $\int_1^t (\{x\} - \frac{1}{2}) dx$ is bounded and $1/x$ goes to 0 as $x \to \infty$. ☐

Note that an easy consequence of Lemma 2 is that $\binom{2n}{n} = (1 + o(1))2^{2n}/C\sqrt{n/2}$. We shall now use this to find the value of C.

Lemma 3. *Let $f : \mathbb{R} \to \mathbb{R}$ be an $n + 1$ times continuously-differentiable function. Then for all $x \in \mathbb{R}$, we have*

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + R_n(x),$$

where

$$R_n(x) = \frac{1}{n!}\int_0^x f^{(n+1)}(t)(x - t)^n\, dt.$$

Proof. This follows by induction on n, using integration by parts. ☐

Lemma 4. *$C = \sqrt{2\pi}$.*

Proof. Apply Lemma 3 with $f(x) = (1 + x)^{2n+1}$ to compute $R_n(1)$:

$$\begin{aligned}
\frac{1}{2^{2n+1}} R_n(1) &= \frac{1}{2^{2n+1}} \cdot \frac{1}{n!} \int_0^1 (2n+1)(2n)\cdots(n+1)(1+t)^n(1-t)^n\, dt \\
&= \frac{2\binom{2n}{n}}{2^{2n+1}}(n + \frac{1}{2}) \int_0^1 (1 - t^2)^n\, dt = \frac{\binom{2n}{n}\sqrt{n}}{2^{2n}}(1 + \frac{1}{2n}) \int_0^{\sqrt{n}} \left(1 - \frac{u^2}{n}\right)^n du \\
&\xrightarrow[n\to\infty]{} \frac{\sqrt{2}}{C} \int_0^\infty e^{-u^2}\, du = \frac{\sqrt{2}}{C} \cdot \frac{\sqrt{\pi}}{2}.
\end{aligned}$$

The convergence of the integrals is justified by the fact that $(1-u^2/n)^n \leq e^{-u^2}$ for all $0 \leq u \leq \sqrt{n}$, and $(1-u^2/n)^n \to e^{-u^2}$ as $n \to \infty$, uniformly on compact intervals. To finish the proof, note that

$$\frac{1}{2^{2n+1}} R_n(1) = \sum_{n<k\leq 2n+1} \frac{\binom{2n+1}{k}}{2^{2n+1}} = \frac{1}{2}$$

(this is the probability that a $\mathrm{Binom}(2n+1,1/2)$ random variable takes a value $> n$). Therefore $C = \sqrt{2\pi}$, as claimed. $\square$

Corollary 5 (Stirling's formula). $\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}(n/e)^n} = 1$.

Note that the proof is based on computing $\mathbf{P}(S_{2n+1} > n)$ in two different ways, when $S_{2n+1} \sim \mathrm{Binom}(2n+1,1/2)$. This is just the special case $p = 1/2, t = 0$ of Theorem 1! In this very special case, by symmetry the probability is equal to 1/2; on the other hand, Lemma 3 enables us to relate this to the asymptotic behavior of $n!$ and to (half of) the gaussian integral $\int_{-\infty}^{\infty} e^{-x^2}\,dx$. The evaluation of the constant C in Stirling's formula is the part that is attributed to James Stirling. The form that appears in Lemma 2 is due to Abraham de Moivre (1733).

With this preparation, it is now possible to apply the same technique to prove Theorem 1. Instead of the function $f(x) = (1+x)^{2n+1}$, take the function $g(x) = ((1-p)+px)^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} x^k$, and compute the remainder $R_k(1)$ of the Taylor expansion of g, where $k \approx np + t\sqrt{np(1-p)}$. This should converge to $1 - \Phi(t)$, and indeed, this follows without too much difficulty from Lemma 3. The computation is left as an exercise. We also sketch another way of proving Theorem 1 by directly approximating the probabilities $\binom{n}{k} p^k (1-p)^{n-k}$ by Gaussian densities.

Sketch of Proof of Theorem 1. Denote $q = 1-p$. For a large n, let k be approximately equal to $np + t\sqrt{npq}$, and use Stirling's formula to estimate the probability $\mathbf{P}(S_n = k)$, as follows:

$$\begin{aligned}
\mathbf{P}(S_n = k) &= \binom{n}{k} p^k q^{n-k} = (1+o(1)) \frac{\sqrt{2\pi n}(n/e)^n p^k q^{n-k}}{\sqrt{2\pi k}(k/e)^k \sqrt{2\pi(n-k)}((n-k)/e)^{n-k}} \\
&= \frac{1+o(1)}{\sqrt{2\pi npq}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \\
&= \frac{1+o(1)}{\sqrt{2\pi npq}} \left(1 + \frac{t\sqrt{q}}{\sqrt{np}}\right)^{-k} \left(1 - \frac{t\sqrt{p}}{\sqrt{nq}}\right)^{-(n-k)}.
\end{aligned}$$

Taking the logarithm of the product of the last two factors, using the facts that $k \approx np + t\sqrt{npq}$, $n - k \approx nq - t\sqrt{npq}$, and that $\log(1+x) = x - x^2/2 + O(x^3)$ when $x \to 0$, we see that

$$\begin{aligned}
&\log\left[\left(1+\frac{t\sqrt{q}}{\sqrt{np}}\right)^{-k}\left(1-\frac{t\sqrt{p}}{\sqrt{nq}}\right)^{-(n-k)}\right] \\
&= -(np+t\sqrt{npq})\log\left(1+\frac{t\sqrt{q}}{\sqrt{np}}\right)-(nq-t\sqrt{npq})\log\left(1-\frac{t\sqrt{p}}{nq}\right) \\
&= -(np+t\sqrt{npq})\left(\frac{t\sqrt{q}}{\sqrt{np}}-\frac{t^2q}{2np}\right)-(nq-t\sqrt{npq})\left(-\frac{t\sqrt{p}}{\sqrt{nq}}-\frac{t^2p}{2nq}\right)+O\left(\frac{t^3}{\sqrt{n}}\right) \\
&= -t\sqrt{npq}-t^2q+\frac{t^2q}{2}+t\sqrt{npq}-t^2p+\frac{t^2p}{2}+O\left(\frac{t^3}{\sqrt{n}}\right) \\
&= -\frac{t^2}{2}+O\left(\frac{t^3}{\sqrt{n}}\right).
\end{aligned}$$

It follows that

$$\mathbf{P}(S_n = k) = \frac{1+o(1)}{\sqrt{2\pi npq}}e^{-t^2/2}$$

In other words, the individual probabilities for S_n approximate a normal density! From here, it is not too hard to show that the probability

$$\mathbf{P}\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = \sum_{np+a\sqrt{npq}\le k\le np+b\sqrt{npq}} \mathbf{P}(S_n = k)$$

is approximately a Riemann sum for the integral $(2\pi)^{-1/2}\int_a^b e^{-x^2/2}\,dx = \Phi(b) - \Phi(a)$. In fact, this is true since for a, b fixed and k ranging between $np + a\sqrt{npq}$ and $np + b\sqrt{npq}$, the error concealed by the $o(1)$ term is uniformly small (smaller than any $\epsilon > 0$, say, when n is sufficiently large), since this error term originates with three applications of Stirling's approximation formula (for $n!$, for $k!$ and for $(n-k)!$) followed by the log function second-order Taylor expansion above. $\square$

One lesson that can be learned from this proof is that doing computations for specific distributions can be *messy*! So we might be better off looking for more general, and therefore more conceptual, techniques for proving convergence to the normal distribution, that require less explicit computations; fortunately such techniques exist, and will lead us to the much more general central limit theorem.

Lecture 11: Convergence in distribution

11.1 Definition

Since we will be talking about convergence of the distribution of random variables to the normal distribution, it makes sense to develop the general theory of convergence of distributions to a limiting distribution.

Definition 6. *Let $(F_n)_{n=1}^{\infty}$ be a sequence of distribution functions. We say that F_n* ***converges to a limiting distribution function*** *F, and denote this by $F_n \implies F$, if $F_n(x) \to F(x)$ as $n \to \infty$ for any $x \in \mathbb{R}$ which is a continuity point of F. If $X, (X_n)_{n=1}^{\infty}$ are random variables, we say that X_n* ***converges in distribution to*** *X (or, interchangeably,* ***converges in distribution to*** *F_X) if $F_{X_n} \implies F_X$.*

This definition, which may seem unnatural at first sight, will become more reasonable after we prove the following lemma.

Lemma 7. *The following are equivalent:*

1. *$X_n \implies X$.*
2. *$\mathbf{E}f(X_n) \xrightarrow[n\to\infty]{} \mathbf{E}f(X)$ for any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$.*
3. *There exists a r.v. Y and a sequence $(Y_n)_{n=1}^{\infty}$ of r.v.'s, all defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $Y_n \to Y$ a.s., Y is equal in distribution to X, and each Y_n is equal in distribution to the respective X_n.*

Proof. Proof that $2 \implies 1$: Assume that $\mathbf{E}f(X_n) \xrightarrow[n\to\infty]{} \mathbf{E}f(X)$ for any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$, and fix $x \in \mathbb{R}$. For any $t \in \mathbb{R}$ and $\epsilon > 0$, define a function $g_{t,\epsilon} : \mathbb{R} \to \mathbb{R}$ by

$$g_{t,\epsilon}(u) = \begin{cases} 1 & u < t, \\ \frac{t-u+\epsilon}{\epsilon} & u \le t \le t+\epsilon, \\ 0 & u > t+\epsilon. \end{cases}$$

Then we have that

$$\mathbf{E}(g_{x-\epsilon,\epsilon}(X_n)) \le F_{X_n}(x) = \mathbf{E}(\mathbf{1}_{(-\infty,x]}(X_n)) \le \mathbf{E}(g_{x,\epsilon}(X_n))$$

Letting $n \to \infty$ gives the chain of inequalities

$$F_X(x-\epsilon) \le \mathbf{E}(g_{x-\epsilon,x}(X)) \le \liminf_{n\to\infty} F_{X_n}(x) \le \limsup_{n\to\infty} F_{X_n}(x) \le \mathbf{E}(g_{x,\epsilon}(X)) \le F_X(x+\epsilon).$$

Now if x is a point of continuity of F_X, letting $\epsilon \downarrow 0$ gives that $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$.

Proof that $3 \implies 2$: this follows immediately by applying the bounded convergence theorem to the sequence $g(Y_n)$.

Proof that $1 \implies 3$: Take $(\Omega, \mathcal{F}, \mathbf{P}) = ((0,1), \mathcal{B}(0,1), \text{Leb})$. For each $n \ge 1$, let $Y_n(x) = \sup\{y : F_{X_n}(y) < x\}$ be the **lower percentile function** of X_n, as discussed in a previous lecture, and similarly let $Y(x) = \sup\{y : F_X(y) < x\}$ be the lower percentile function of X. Then as we previously showed, we have $F_Y \equiv F_X$ and $F_{Y_n} \equiv F_{X_n}$ for all n. It remains to show that $Y_n(x) \to Y(x)$ for almost all $x \in (0,1)$. In fact, we show that this is true for all but a countable set of x's. Denote $Y^*(x) = \inf\{y : F_X(y) > x\}$ (the **upper percentile function** of X). As we have seen, we always have $Y(x) \le Y^*(x)$, and $Y(x) = Y^*(x)$ for all $x \in (0,1)$ except on a countable set of x's (the exceptional x's correspond to intervals where F_X is constant; these intervals are disjoint and each one contains a rational point).

Let $x \in (0,1)$ be such that $Y(x) = Y^*(x)$. This means that for any $y < Y(x)$ we have $F_X(y) < x$, and for any $z > Y(x)$ we have $F_X(z) > x$. Now, take a $y < Y(x)$ which is a continuity point of F_X. Then $F_{X_n}(y) \to F_X(y)$ as $n \to \infty$, so also $F_{X_n}(y) < x$ for sufficiently large n, which means (by the definition of Y_n) that $Y_n(x) \ge y$ for such large n. This establishes that $\liminf_{n\to\infty} Y_n(x) \ge Y(x)$, since we have continuity points $y < Y(x)$ that are arbitrarily close to $Y(x)$.

Similarly, take a $z > Y(x)$ which is a continuity point of F_X. Then $F_{X_n}(z) \to F_x(z)$ as $n \to \infty$, so also $F_{X_n}(z) > x$ for large n, which implies that $Y_n(x) \le z$. Again, by taking continuity points $z > Y(x)$ that are arbitrarily close to $Y(x)$ we get that $\limsup_{n\to\infty} Y_n(x) \le Y(x)$. Combining these last two results shows that $Y_n(x) \to Y(x)$ which was what we wanted. $\square$

11.2 Examples

1. **Normal convergence:** We showed that if $X_1, X_2, \ldots$ are i.i.d. $\text{Binom}(1,p)$ r.v.'s and $S_n = \sum_{k=1}^n X_k$, then

$$\frac{S_n - n\mathbf{E}(X_1)}{\sqrt{n}\sigma(X_1)} \implies N(0,1).$$

Similarly, using explicit computations (see the homework) it is not too difficult to see that this is also true when $X_1 \sim \mathrm{Poisson}(1)$, $X_1 \sim \mathrm{Exp}(1)$, and in other specific examples. The central limit theorem generalizes this claim to any i.i.d. sequence with finite variance.

2. **Waiting for rare events:** If for each $0 < p < 1$ we have a r.v. $X_p \sim \mathrm{Geom}_0(p)$, then $\mathbf{P}(X_p \ge n) = (1-p)^{n-1}$. It follows that

$$\mathbf{P}(pX_p > x) = (1-p)^{\lfloor x/p \rfloor} \xrightarrow[p \downarrow 0]{} e^{-x}, \qquad (x > 0),$$

so

$$pX_p \implies \mathrm{Exp}(1) \qquad \text{as } p \downarrow 0.$$

3. **Pólya's urn:** Let X_n be the number of white balls in the Pólya urn experiment after starting with one white ball and one black ball and performing the experiment for n steps (so that there are $n+2$ balls). In a homework exercise we showed that X_n is a discrete uniform r.v. on $\{1, 2, \ldots, n+1\}$. It follows easily that the proportion of white balls in the urn converges in distribution:

$$\frac{X_n}{n+2} \implies U(0,1).$$

4. **Gumbel distribution:** If $X_1, X_2, \ldots$ are i.i.d. Exp(1) random variables, and $M_n = \max(X_1, \ldots, X_n)$, we showed in a homework exercise that

$$\mathbf{P}(M_n - \log n \le x) \xrightarrow[n\to\infty]{} e^{-e^{-x}}, \qquad x \to \infty$$

It follows that

$$M_n - \log n \implies F$$

where $F(x) = \exp\left(-e^{-x}\right)$ is called the Gumbel distribution.

11.3 Compactness and tightness

Theorem 8 (Helly's selection theorem). *If $(F_n)_{n=1}^{\infty}$ is a sequence of distribution functions, then there is a subsequence F_{n_k} and a right-continuous, nondecreasing function $H : \mathbb{R} \to [0,1]$ such that*

$$F_{n_k}(x) \xrightarrow[n\to\infty]{} H(x)$$

holds for any $x \in \mathbb{R}$ which is a continuity point of H.

Note. The subsequential limit H need not be a distribution function, since it may not satisfy the properties $\lim_{x\to-\infty} H(x) = 0$ or $\lim_{x\to\infty} H(x) = 1$. For example, taking $F_n = F_{X_n}$, where $X_n \sim U[-n, n]$, we see that $F_n(x) \to 1/2$ for all $x \in \mathbb{R}$. For a more interesting example, take $G_n = (F_n + F_{Z_n})/2$ where F_n are as in the previous example, and Z_n is some sequence of r.v.'s that converges in distribution.

Proof. First, note that we can find a subsequence $(n_k)_{k=1}^{\infty}$ such that $F_{n_k}(r)$ converges to a limit $G(r)$ at least for any *rational* number r. This is done by combining the compactness of the interval $[0, 1]$ (which implies that for any specific $a \in \mathbb{R}$ we can always take a subsequence to make the sequence of numbers $F_n(a)$ converge to a limit) with a diagonal argument (for some enumeration $r_1, r_2, r_3, \ldots$ of the rationals, first take a subsequence to force convergence at r_1; then take a subsequence of that subsequence to force convergence at r_2, etc.; now form a subsequence whose k-th term is the k-th term of the k-th subsequence in this series).

Now, use $G(\cdot)$, which is defined only on the rationals and not necessarily right-continuous (but is nondecreasing), to define a function $H : \mathbb{R} \to \mathbb{R}$ by

$$H(x) = \inf\{G(r) : r \in \mathbb{Q}, r > x\}.$$

This function is clearly nondecreasing, and is also right-continuous, since we have

$$\lim_{x_n \downarrow x} H(x_n) = \inf\{G(r) : r \in \mathbb{Q}, r > x_n \text{ for some } n\} = \inf\{G(r) : r \in \mathbb{Q}, r > x\} = H(x).$$

Finally, let x be a continuity point of H. To show that $F_{n_k}(x) \to H(x)$, fix some $\epsilon > 0$ and let r_1, r_2, s be rationals such that $r_1 < r_2 < x < s$ and

$$H(x) - \epsilon < H(r_1) \le H(r_2) \le H(x) \le H(s) < H(x) + \epsilon.$$

Then since $F_{n_k}(r_2) \to G(r_2) \ge H(r_1)$, and $F_n(s) \to G(s) \le H(s)$, it follows that for sufficiently large k we have

$$H(x) - \epsilon < F_{n_k}(r_2) \le F_{n_k}(x) \le F_{n_k}(s) < H(x) + \epsilon.$$

Therefore

$$H(x) - \epsilon \le \liminf_{n\to\infty} F_{n_k}(x) \le \limsup_{n\to\infty} F_{n_k}(x) \le H(x) + \epsilon,$$

and since ϵ was arbitrary this proves the claim. $\square$

Theorem 8 can be thought of as a kind of compactness property for probability distributions, except that the subsequential limit guaranteed to exist by the theorem is not a distribution function. To ensure that we get a distribution function, it turns out that a certain property called **tightness** has to hold.

Definition 9. *A sequence $(\mu_n)_{n=1}^\infty$ of probability measures on $(\mathbb{R}, \mathcal{B})$ is called **tight** if for any $\epsilon > 0$ there exists an $M > 0$ such that*

$$\liminf_{n\to\infty} \mu_n([-M, M]) \geq 1 - \epsilon.$$

A sequence of distribution functions $(F_n)_{n=1}^\infty$ is called tight if the associated probability measures determined by F_n form a tight sequence, or, more explicitly, if for any $\epsilon > 0$ there exists an $M > 0$ such that

$$\limsup_{n\to\infty}(1 - F_n(M) + F_n(-M)) < \epsilon.$$

A sequence of random variables is called tight if the sequence of their distribution functions is tight.

Theorem 10. *If $(F_n)_{n=1}^\infty$ is a tight sequence of distribution functions, then there exists a subsequence $(F_{n_k})_{k=1}^\infty$ and a distribution function F such that $F_{n_k} \implies F$. In fact, any subsequential limit H as guaranteed to exist in the previous theorem is a distribution function.*

Exercise 11. *Prove that the converse is also true, i.e., if a sequence is not tight then it must have at least one subsequential limit H (in the sense of the subsequence converging to H at any continuity point of H) that is not a proper distribution function. In particular, it is worth noting that a sequence that converges in distribution is tight.*

Proof. Let H be a nondecreasing, right-continuous function that arises as a subsequential limit-in-distribution of a subsequence F_{n_k}, that we know exists by Theorem 8. To show that H is a distribution function, fix $\epsilon > 0$, and let $M > 0$ be the constant guaranteed to exist in the definition of tightness. Let $x < -M$ be a continuity point of H. We have

$$H(x) = \lim_{k\to\infty} F_{n_k}(x) \leq \limsup_{k\to\infty} F_{n_k}(-M) \leq \limsup_{k\to\infty}(F_{n_k}(-M) + (1 - F_{n_k}(M))) < \epsilon,$$

so this shows that $\lim_{x\to-\infty} H(x) = 0$. Similarly, let $x > M$ be a continuity point of H. Then

$$H(x) = \lim_{k\to\infty} F_{n_k}(x) \geq \liminf_{k\to\infty} F_{n_k}(M) \geq \liminf_{k\to\infty}(F_{n_k}(M)) - F_{n_k}(-M)) > 1 - \epsilon,$$

which shows that $\lim_{x\to\infty} H(x) = 1$. $\square$

The condition of tightness is not very restrictive, and in practical situations it is usually quite easy to verify. The following lemma gives an example that is relevant for our purposes.

Lemma 12. *If* $X_1, X_2, \dots$ *are r.v.'s such that* $\mathbf{E}X_n = 0$ *and* $\mathbf{V}(X_n) < C$ *for all* n*, then* $(X_n)_n$ *is a tight sequence.*

Proof. Use Chebyshev's inequality:

$$\mathbf{P}(|X_n| > M) \le \frac{\mathbf{V}(X_n)}{M^2} \le \frac{C}{M^2},$$

so, if $\epsilon > 0$ is given, taking $M = \sqrt{C/\epsilon}$ ensures that the left-hand side is bounded by ϵ. ☐

Lecture 12: Characteristic functions

12.1 Definition and basic properties

A main tool in our proof of the central limit theorem will be that of characteristic functions. The basic idea will be to show that

$$\mathbf{E}\left[g\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right)\right] \xrightarrow[n\to\infty]{} \mathbf{E}g(N(0,1))$$

for a sufficiently large family of functions g. It turns out that the family of functions of the form

$$g_t(x) = e^{itx}, \qquad (t \in \mathbb{R}),$$

is ideally suited for this purpose. (Here and throughout, $i = \sqrt{-1}$).

Definition 13. *The characteristic function of a r.v.* X*, denoted* φ_X*, is defined by*

$$\varphi_X(t) = \mathbf{E}\left(e^{itX}\right) = \mathbf{E}(\cos(tX)) + i\mathbf{E}(\sin(tX)), \qquad (t \in \mathbb{R}).$$

Note that we are taking the expectation of a *complex-valued random variable* (which is a kind of two-dimensional random vector, really). However, the main properties of the expectation operator (linearity, the triangle inequality etc.) that hold for real-valued random variables also hold for complex-valued ones, so this will not pose too much of a problem.

Here are some simple properties of characteristic functions. For simplicity we denote $\varphi = \varphi_X$ where there is no risk of confusion.

1. $\varphi(o) = \mathbf{E}e^{i\cdot 0\cdot X} = 1$.
2. $\varphi(-t) = \mathbf{E}e^{-itX} = \mathbf{E}\left(\overline{e^{itX}}\right) = \overline{\varphi(t)}$ (where $\overline{z}$ denotes the complex conjugate of a complex number z).
3. $|\varphi(t)| \le \mathbf{E}\left|e^{itX}\right| = 1$ by the triangle inequality.
4. $|\varphi(t) - \varphi(s)| \le \mathbf{E}\left|e^{itX} - e^{isX}\right| = \mathbf{E}\left|e^{isX}\left(e^{i(t-s)X} - 1\right)\right| = \mathbf{E}\left|e^{i(t-s)X} - 1\right|$. Note also that $\mathbf{E}\left|e^{iuX} - 1\right| \to 0$ as $u \downarrow 0$ by the bounded convergence theorem. It follows that φ is a uniformly continuous function on $\mathbb{R}$.
5. $\varphi_{aX}(t) = \mathbf{E}e^{iatX} = \varphi_X(at), \ (a \in \mathbb{R})$.
6. $\varphi_{X+b}(t) = \mathbf{E}e^{it(X+b)} = e^{ibt}\varphi_X(t), \ (b \in \mathbb{R})$.
7. **Important:** If X, Y are independent then

   $$\varphi_{X+Y}(t) = \mathbf{E}\left(e^{it(X+Y)}\right) = \mathbf{E}\left(e^{itX}e^{itY}\right) = \mathbf{E}\left(e^{itX}\right)\mathbf{E}\left(e^{itY}\right) = \varphi_X(t)\varphi_Y(t).$$

   Note that this is the main reason why characteristic functions are such a useful tool for studying the distribution of a sum of independent random variables.

A note on terminology. If X has a density function f, then the characteristic function can be computed as

$$\varphi_X(t) = \int_{-\infty}^{\infty} f_X(x)e^{itx}\,dx.$$

In all other branches of mathematics, this would be called the **Fourier transform**[1] **of** f. So the concept of a characteristic function generalizes the Fourier transform. If μ is the distribution measure of X, some authors write

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx}d\mu(x)$$

(which is an example of a **Lebesgue-Stieltjes integral**) and call this the **Fourier-Stieltjes transform** (or just the Fourier transform) of the measure μ.

---

[1] Well, more or less – it is really the inverse Fourier transform; but it will be the Fourier transform if we replace t by $-t$, so that is almost the same thing

12.2 Examples

No study of characteristic functions is complete without "dirtying your hands" a little to compute the characteristic function for some important cases. The following exercise is highly recommended

Exercise 14. *Compute the characteristic functions for the following distributions.*

1. ***Coin flips:*** *Compute* φ_X *when* $\mathbf{P}(X = -1) = \mathbf{P}(X = 1) = 1/2$ *(this comes out slightly more symmetrical than the usual Bernoulli r.v. for which* $\mathbf{P}(X = 0) = \mathbf{P}(X = 1) = 1/2$*).*
2. ***Symmetric random walk:*** *Compute* φ_{S_n} *where* $S_n = \sum_{k=1}^n X_k$ *is the sum of* n *i.i.d. copies of the coin flip distribution above.*
3. ***Poisson distribution:*** $X \sim Poisson(\lambda)$*.*
4. ***Uniform distribution:*** $X \sim U[a, b]$*, and in particular* $X \sim [-1, 1]$ *which is especially symmetric and useful in applications.*
5. ***Exponential distribution:*** $X \sim Exp(\lambda)$*.*
6. ***Symmetrized exponential:*** *A r.v.* Z *with density function* $f_Z(x) = e^{-|x|}$*. Note that this is the distribution of the exponential distribution after being "symmetrized" in either of two ways: (i) We showed that if* $X, Y \sim Exp(1)$ *are independent then* $X - Y$ *has density* $e^{-|x|}$*; (ii) alternatively, it is the distribution of an "exponential variable with random sign", namely* $\varepsilon \cdot X$ *where* $X \sim Exp(1)$ *and* ε *is a random sign (same as the coin flip distribution mentioned above) that is independent of* X*.*

The normal distribution has the nice property that its characteristic function is equal, up to a constant, to its density function.

Lemma 15. *If* $Z \sim N(0, 1)$ *then*

$$\varphi_Z(t) = e^{-t^2/2}.$$

Proof.

$$\begin{aligned}\varphi_Z(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx}e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-t^2/2}e^{(x-it)^2/2}\,dx \\ &= e^{-t^2/2}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{(x-it)^2/2}\,dx\right).\end{aligned}$$

As Durrett suggests in his "physics proof" (p. 91 in 3rd ed.), the expression in parentheses is 1, since it is the integral of a normal density with mean it and variance 1. This is a nonsensical argument, of course (it being an imaginary number), but the claim is true, easy and is proved in any complex analysis course using contour integration.

Alternatively, let $S_n = \sum_{k=1}^n X_k$ where $X_1, X_2, \ldots$ are i.i.d. coin flips with $\mathbf{P}(X_k) = -1 = \mathbf{P}(X_k) = 1 = 1/2$. We know from the de Moivre-Laplace theorem (Theorem 1) that

$$S_n/\sqrt{n} \implies N(0,1),$$

so that

$$\varphi_{S_n/\sqrt{n}}(t) = \mathbf{E}\left(e^{itS_n/\sqrt{n}}\right) \xrightarrow[n\to\infty]{} \varphi_Z(t), \qquad (t \in \mathbb{R}),$$

since the function $x \to e^{itx}$ is bounded and continuous. On the other hand, from the exercise above it is easy to compute that $\varphi_{S_n}(t) = \cos^n(t)$, which implies that

$$\varphi_{S_n/\sqrt{n}}(t) = \cos^n\left(\frac{t}{\sqrt{n}}\right) = \left(1 - \frac{t^2}{2n} + O(t^4)\right)^n \xrightarrow[n\to\infty]{} e^{-t^2/2}.$$

□

As a consequence, let $X \sim N(0,\sigma_1^2)$ and $Y \sim N(0,\sigma_2^2)$ be independent, and let $Z = X+Y$. Then

$$\varphi_X(t) = e^{-\sigma_1^2 t^2/2}, \qquad \varphi_Y(t) = e^{-\sigma_2^2 t^2/2},$$

so $\varphi_Z(t) = e^{-(\sigma_1^2+\sigma^2)/2}$. This is the same as $\varphi_W(t)$, where $W \sim N(0,\sigma_1^2+\sigma_2^2)$. It would be nice if we could deduce from this that $Z \sim N(0,\sigma_1^2+\sigma_2^2)$ (we already proved this fact in a homework exercise, but it's always nice to have several proofs of a result, especially an important one like this one). This naturally leads us to an important question about characteristic functions, which we consider in the next section.

12.3 The inversion formula

A fundamental question about characteristic functions is whether they contain all the information about a distribution, or in other words whether knowing the characteristic function determines the distribution uniquely. This question is answered (affirmatively) by the following theorem, which is a close cousin of the standard inversion formula from analysis for the Fourier transform.

Theorem 16 (The inversion formula)**.** *If X is a r.v. with distribution μ_X, then for any $a<b$ we have*

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}\varphi_X(t)\,dt &= \mu_X((a,b))+\frac{1}{2}\mu_X(\{a,b\}) \\ &= \mathbf{P}(a<X<b)+\frac{1}{2}\mathbf{P}(X=a)+\frac{1}{2}\mathbf{P}(X=b).\end{aligned}$$

Proof. Throughout the proof, denote $\varphi(t)=\varphi_X(t)$ and $\mu=\mu_X$. For convenience, we use the notation of Lebesgue-Stieltjes integration with respect to the measure μ, remembering that this really means taking the expectation of some function of the r.v. X. Denote

$$I_T=\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}\varphi(t)\,dt=\int_{-T}^{T}\int_{-\infty}^{\infty}\frac{e^{-iat}-e^{-ibt}}{it}e^{itx}d\mu(x)\,dt. \tag{1}$$

Since $\frac{e^{-iat}-e^{-ibt}}{it}=\int_a^b e^{-ity}\,dy$ is a bounded function of t (it is bounded in absolute value by $b-a$), it follows by Fubini's theorem that we can change the order of integration, so

$$\begin{aligned}I_T &= \int_{-\infty}^{\infty}\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}e^{itx}dt\,d\mu(x) \\ &= \int_{-\infty}^{\infty}\left[\int_{-T}^{T}\frac{\sin(t(x-a))}{t}\,dt-\int_{-T}^{T}\frac{\sin(t(x-b))}{t}\,dt\right]d\mu(x) \\ &= \int_{-\infty}^{\infty}\left(R(x-a,T)-R(x-b,T)\right)d\mu(x),\end{aligned}$$

where we denote $R(\theta,T)=\int_{-T}^{T}\sin(\theta t)/t\,dt$. Note that in the notation of expectations this can be written as $I_T=\mathbf{E}\left(R(X-a,T)-R(X-b,T)\ \right)$. This can be simplified somewhat; in fact, observe also that

$$R(\theta,T)=2\mathrm{sgn}(\theta)\int_0^{|\theta|T}\frac{\sin x}{x}\,dx=2\mathrm{sgn}(\theta)S(|\theta|T),$$

where we denote $S(x) = \int_0^x \frac{\sin(u)}{u}\,du$ and $\mathrm{sgn}(\theta)$ is 1 if $\theta > 0$, -1 if $\theta < 0$ and 0 if $\theta = 0$. By a standard convergence test for integrals, the improper integral $\int_0^\infty \frac{\sin u}{u}\,du = \lim_{x\to\infty} S(x)$ converges; denote its value by $C/4$. Thus, we have shown that $R(\theta, T) \to 2\mathrm{sgn}(\theta)C$ as $T \to \infty$, hence that

$$R(x-a,T) - R(x-b,T) \xrightarrow[T\to\infty]{} \begin{cases} C & a < x < b, \\ C/2 & x = a \text{ or } x = b, \\ 0 & x < a \text{ or } x > b. \end{cases}$$

Furthermore, the function $R(x-a,T)-R(x-b,T)$ is bounded in absolute value by $2\sup_{x\ge 0} S(x)$. It follows that we can apply the bounded convergence theorem in (1) to get that

$$I_T \xrightarrow[T\to\infty]{} C\mathbf{E}(1_{a<X<b}) + (C/2)\mathbf{E}(\mathbf{1}_{\{X=a\}} + \mathbf{1}_{\{X=b\}}) = C\mu((a,b)) + (C/2)\mu(\{a,b\}). \tag{2}$$

This is just what we claimed, minus the fact that $C = 2\pi$. This fact is a well-known integral evaluation from complex analysis. We can also deduce it in a self-contained manner, by applying what we proved to a specific measure μ and specific values of a and b for which we can evaluate the limit in (1) directly. This is not entirely easy to do, but one possibility, involving an additional limiting argument, is outlined in the next exercise; see also Exercise 6.6 in Appendix A.6 in [Durrett] (p. 470 in 3rd ed.) for a different approach to finding the value of C. □

Exercise 17. ***(Recommended for aspiring analysts...)*** *For each $\sigma > 0$, let X_σ be a r.v. with distribution $N(0,\sigma^2)$ and therefore with density $f_X(x) = (\sqrt{2\pi}\sigma)^{-1}e^{-x^2/2\sigma^2}$ and characteristic function $\varphi_X(t) = e^{-\sigma^2t^2/2}$. For fixed σ, apply Theorem 16 in its weak form given by* (2) *(that is, without the knowledge of the value of C), with parameters $X = X_\sigma$, $a = -1$ and $b = 1$, to deduce the identity*

$$\frac{C}{\sqrt{2\pi}\sigma}\int_{-1}^{1} e^{-x^2/2\sigma^2}\,dx = \int_{-\infty}^{\infty} \frac{2\sin t}{t} e^{-\sigma^2t^2/2}\,dt.$$

Now multiply both sides by σ and take the limit as $\sigma \to \infty$. For the left-hand side this should give in the limit (why?) the value $(2C)/\sqrt{2\pi}$. For the right-hand side this should give $2\sqrt{2\pi}$. Justify these claims and compare the two numbers to deduce that $C = 2\pi$.

The following theorem shows that the inversion formula can be written as a simpler connection between the characteristic function and the density function of a random variable, in the case when the characteristic function is integrable.

Theorem 18. *If* $\int_{-\infty}^{\infty} |\varphi_X(t)|\,dt < \infty$*, then* X *has a bounded and continuous density function* f_X*, and the density and characteristic function are related by*

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{\infty} f_X(x) e^{itx}\,dx, \\ f_X(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_X(t) e^{-itx}\,dt. \end{aligned}$$

In the lingo of Fourier analysis, this is known as the **inversion formula for Fourier transforms**.

Proof. This is a straightforward corollary of Theorem 16. See [Durrett, 3rd ed., p. 95]. □

12.4 The continuity theorem

Theorem 19. *Let* $(X_n)_{n=1}^{\infty}$ *be r.v.'s. Then:*

(i) If $X_n \implies X$ *for some r.v.* X*, then* $\varphi_{X_n}(t) \to \varphi_X(t)$ *for all* $t \in \mathbb{R}$*.*

(ii) If the limit $\varphi(t) = \lim_{n\to\infty} \varphi_{X_n}(t)$ *exists for all* $t \in \mathbb{R}$*, and* φ *is continuous at 0, then* $\varphi \equiv \varphi_X$ *for some r.v.* X*, and* $X_n \implies X$*.*

Proof. Part (i) follows immediately from the fact that convergence in distribution implies that $\mathbf{E}g(X_n) \to \mathbf{E}g(X)$ for any bounded continuous function. It remains to prove the less trivial claim in part (ii). Assume that $\varphi_{X_n}(t) \to \varphi(t)$ for all $t \in \mathbb{R}$ and that φ is continuous at 0. First, we show that the sequence $(X_n)_{n=1}^{\infty}$ is tight. Fixing an $M > 0$, we can bound the probability $\mathbf{P}(|X_n| > M)$, as follows:

$$\begin{aligned} \mathbf{P}(|X_n| > M) &= \mathbf{E}\left(\mathbf{1}_{\{|X_n|>M\}}\right) \le \mathbf{E}\left[2\left(1 - \frac{M}{2X_n}\right)\mathbf{1}_{\{|X_n|>M\}}\right] \\ &\le \mathbf{E}\left[2\left(1 - \frac{\sin(2X_n/M)}{2X_n/M}\right)\mathbf{1}_{\{|X_n|>M\}}\right]. \end{aligned}$$

But this last expression can be related to the behavior of the characteristic function near 0. Denote $\delta = 2/M$. Reverting again to the Lebesgue-Stieltjes integral notation, we have

$$\begin{aligned}\mathbf{E}\left[2\left(1-\frac{\sin(2X_n/M)}{2X_n/M}\right)\mathbf{1}_{\{|X_n|>M\}}\right] &= 2\int_{|x|>2/\delta}\left(1-\frac{\sin(\delta x)}{\delta x}\right)d\mu_{X_n}(x) \\ \leq\ 2\int_{-\infty}^{\infty}\left(1-\frac{\sin(\delta x)}{\delta x}\right)d\mu_{X_n}(x) &= \int_{-\infty}^{\infty}\frac{1}{\delta}\left(\int_{-\delta}^{\delta}(1-e^{itx})\,dt\right)\,d\mu_{X_n}(x).\end{aligned}$$

Now use Fubini's theorem to get that this bound can be written as

$$\frac{1}{\delta}\int_{-\delta}^{\delta}\int_{-\infty}^{\infty}(1-e^{itx})\,d\mu_{X_n}(x)\,dt = \frac{1}{\delta}\int_{-\delta}^{\delta}(1-\varphi_{X_n}(t))\,dt \xrightarrow[n\to\infty]{} \frac{1}{\delta}\int_{-\delta}^{\delta}(1-\varphi(t))\,dt$$

(the convergence follows from the bounded convergence theorem). So we have shown that

$$\limsup_{n\to\infty}\mathbf{P}(|X_n|>M) \leq \frac{1}{\delta}\int_{-\delta}^{\delta}(1-\varphi(t))\,dt.$$

But, because of the assumption that $\varphi(t) \to \varphi(0) = 1$ as $t \to 0$, it follows that if δ is sufficiently small then $\delta^{-1}\int_{-\delta}^{\delta}(1-\varphi(t))\,dt < \epsilon$, where $\epsilon > 0$ is arbitrary; so this establishes the tightness claim.

Finally, to finish the proof, let $(n_k)_{k=1}^{\infty}$ be a subsequence (guaranteed to exist by tightness) such that $X_{n_k} \implies Y$ for some r.v. Y. Then $\varphi_{X_{n_k}}(t) \to \varphi_Y(t) = \varphi(t)$ as $k \to \infty$ for all $t \in \mathbb{R}$, so $\varphi \equiv \varphi_Y$. This determines the distribution of Y, which means that the limit in distribution is the same no matter what convergent in distribution subsequence of the sequence $(X_n)_n$ we take. But this implies that $X_n \implies Y$ (why? The reader is invited to verify this last claim; it is best to use the definition of convergence in distribution in terms of expectations of bounded continuous functions). $\square$

12.5 Moments

The final step in our lengthy preparation for the proof of the central limit theorem will be to tie the behavior of the characteristic function $\varphi_X(t)$ near $t = 0$ to the moments of X. Note that, computing formally without regards to rigor, we can write

$$\varphi_X(t) = \mathbf{E}(e^{itX}) = \mathbf{E}\left[\sum_{n=0}^{\infty}\frac{i^n t^n X^n}{n!}\right] = \sum_{n=0}^{\infty}\frac{i^n\mathbf{E}X^n}{n!}t^n.$$

So it appears that the moments of X appear as (roughly) the coefficients in the Taylor expansion of φ_X around $t = 0$. However, for CLT we don't want to assume anything beyond the existence of the second moment, so a (slightly) more delicate estimate is required.

Lemma 20. $\left| e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!} \right| \le \min \left(\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!} \right)$.

Proof. Start with the identity

$$R_n(x) := e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!} = \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is}\, ds,$$

which follows from Lemma 3 that we used in the proof of Stirling's formula. Taking the absolute value and using the fact that $|e^{is}| = 1$ gives

$$|R_n(x)| \le \frac{1}{n!} \left| \int_0^x |x-s|^n \, ds \right| = \frac{|x|^{n+1}}{n!}. \tag{3}$$

To get a bound that is better-behaved for large x, note that

$$\begin{aligned} R_n(x) &= R_{n-1}(x) - \frac{(ix)^n}{n!} = R_{n-1}(x) - \frac{i^n}{(n-1)!} \int_0^x (x-s)^{n-1}\, ds \\ &= \frac{i^n}{(n-1)!} \int_0^x (x-s)^{n-1} (e^{is} - 1)\, ds. \end{aligned}$$

So, since $|e^{is} - 1| \le 2$, we get that

$$|R_n(x)| \le \frac{2}{(n-1)!} \left| \int_0^x |x-s|^{n-1} \, ds \right| = \frac{2|x|^n}{(n-1)!}. \tag{4}$$

Combining (3) and (4) gives the claim. □

Now let X be a r.v. with $\mathbf{E}|X|^n < \infty$. Letting $x = tX$ in Lemma 20, taking expectations and using the triangle inequality, we get that

$$\left| \varphi_X(t) - \sum_{m=0}^{n} \frac{i^m \mathbf{E} X^m}{m!} t^m \right| \le \mathbf{E} \left[\min \left(\frac{|t|^{n+1} |X|^{n+1}}{(n+1)!}, \frac{2|t|^n |X|^n}{n!} \right) \right]. \tag{5}$$

Note that in this minimum of two terms, when t is very small the first term gives a better bound, but when taking expectations we need the second term to ensure that the expectation is finite if X is only assumed to have a finite n-th moment.

Theorem 21. *If X is a r.v. with mean $\mu = \mathbf{E}X$ and $\mathbf{V}(X) < \infty$ then*

$$\varphi_X(t) = 1 + i\mu t - \frac{\mathbf{E}X^2}{2}t^2 + o(t^2) \qquad \text{as } t \to 0.$$

Proof. By (5) above, we have

$$\frac{1}{t^2}\left|\varphi_X(t) - \left(1 + i\mu t - \frac{\mathbf{E}X^2}{2}t^2\right)\right| \le \mathbf{E}\left[\min\left(|t| \cdot |X|^3/6, X^2\right)\right].$$

As $t \to 0$, the right-hand side converges to 0 by the dominated convergence theorem. $\square$

Lecture 13: Central limit theorems

13.1 The case of i.i.d. r.v.'s

We are now ready to prove:

Theorem 22 (The central limit theorem)**.** *Let $X_1, X_2, \ldots$ be an i.i.d. sequence of r.v.'s with finite variance. Denote $\mu = \mathbf{E}X_1$, $\sigma = \sigma(X_1)$ and $S_n = \sum_{k=0}^n X_k$. Then as $n \to \infty$ we have the convergence in distribution*

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \implies N(0,1).$$

Proof. For convenience, denote $\hat{X}_k = (X_k - \mu)/\sigma$ and $\hat{S}_n = \sum_{k=0}^n \hat{X}_k$. Then

$$\varphi_{\hat{S}_n/\sqrt{n}}(t) = \varphi_{\hat{S}_n}(t/\sqrt{n}) = \prod_{k=1}^n \varphi_{\hat{X}_k}(t/\sqrt{n}) = \left(\varphi_{\hat{X}_1}(t/\sqrt{n})\right)^n.$$

Note that $\mathbf{E}\hat{X}_1 = 0$ and $\mathbf{V}(\hat{X}_1) = \mathbf{E}\hat{X}_1^2 = 1$. Therefore by Theorem 21, $\varphi_{\hat{X}_1}$ satisfies

$$\varphi_{\hat{X}_1}(u) = 1 - \frac{u^2}{2} + o(u^2)$$

as $u \to 0$. It follows that

$$\varphi_{\hat{S}_n/\sqrt{n}}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \xrightarrow[n\to\infty]{} e^{-t^2/2}$$

for any $t \in \mathbb{R}$. Using the continuity theorem (Theorem 19) and our previous computations, it follows that $\hat{S}_n \implies N(0,1)$, as claimed. $\square$

13.2 Generalizations

The CLT can be generalized in many ways. None of the assumptions (independence, identical distributions, even finite variance) are entirely necessary. A central paradigm of probability theory is that any random quantity that arises as a sum of many small contributions that are either independent or not too strongly dependent, will converge to the normal distribution in some asymptotic limit. Thousands of examples exist, but there is no single all-encompassing theorem that includes all of them as a special case. Rather, probabilists have a toolbox of tricks and techniques that they try to apply in order to prove normal convergence in any given situation. characteristic functions are among the more useful techniques. Another important technique, the so-called **moment method**, involves the direct use of moments: If we can show that $\mathbf{E}(W_n^k) \to \mathbf{E}(Z^k)$, where $(W_n)_{n=1}^\infty$ is the (normalized) sequence being studied, and $Z \sim N(0,1)$, then by Theorem (3.12) in [Durrett, 3rd ed., p. 109], that implies $W_n \implies N(0,1)$.

We now discuss several examples of interesting generalizations of CLT.

13.2.1 Triangular arrays

Theorem 23 (Lindeberg-Feller CLT for triangular arrays)**.** *Let* $(X_{n,k})_{1\le k\le n<\infty}$ *be a triangular array of r.v.'s. Denote* $S_n = \sum_{k=1}^n X_{n,k}$ *(the sum of the n-th row). Assume that:*

1. *For each* n*, the r.v.'s* $(X_{n,k})_{k=1}^n$ *are independent.*
2. $\mathbf{E}X_{n,k} = 0$ *for all* n, k*.*
3. $\mathbf{V}(S_n) = \sigma_n^2 \to \sigma < \infty$ *as* $n \to \infty$*.*
4. *For all* $\epsilon > 0$*,* $\lim_{n\to\infty} \sum_{k=1}^n \mathbf{E}\left(X_{n,k}^2 \mathbf{1}_{\{|X_{n,k}|>\epsilon\}}\right) = 0$*.*

Then $S_n \implies N(0, \sigma^2)$*.*

Proof. See [Durrett, p. 115–116]. The proof uses the characteristic function technique and is a straightforward extension of the proof of the i.i.d. case. □

Example 13.1 Record times and cycles in permutations. Let $X_1, X_2, \ldots$ be i.i.d. $U(0,1)$ r.v.'s. Let A_n be the event that $\{X_n = \max(X_1, \ldots, X_n)\}$ (in this case, we say that n is a **record time**). Let $S_n = \sum_{k=1}^n \mathbf{1}_{A_k}$ be the number of record times up to time n. We saw in a homework exercise that the A_k's are independent events and $\mathbf{P}(A_k) = 1/k$. This implies that $\mathbf{E}(S_n) = \sum_{k=1}^n \frac{1}{k} = H_n$ (the n-th **harmonic number**) and $\mathbf{V}(S_n) = \sum_{k=1}^n \frac{k-1}{k^2}$. Note that both $\mathbf{E}(S_n)$ and $\mathbf{V}(S_n)$ are approximately equal to $\log n$, with an error term that is $O(1)$. Now taking $X_{n,k} = (\mathbf{1}_{A_k} - k^{-1})/\sqrt{\mathbf{V}(S_n)}$ in Theorem 23, it is easy to check that the assumptions of the theorem hold. It follows that

$$\frac{S_n - H_n}{\sigma(S_n)} \implies N(0,1).$$

Equivalently, because of the asymptotic behavior of $\mathbf{E}(S_n)$ and $\mathbf{V}(S_n)$ it is also true that

$$\frac{S_n - \log n}{\sqrt{\log n}} \implies N(0,1).$$

Note: S_n describes the distribution of another interesting statistic on random permutations. It is not too difficult to show by induction (using an amusing construction often referred to as the **Chinese restaurant process**) that if $\sigma \in S_n$ is a uniformly random permutation on n elements, then the number of cycles in σ is a random variable which is equal in distribution to S_n.

13.2.2 Erdös-Kac theorem

Theorem 24 (Erdös-Kac theorem (1940))**.** *Let $g(m)$ denote the number of prime divisors of an integer k (for example, $g(28) = 2$). For each $n \geq 1$, let X_n be a uniformly random integer chosen in $\{1, 2, \ldots, n\}$, and let $Y_n = g(X_n)$ be the number of prime divisors of X_n. Then we have*

$$\frac{Y_n - \log\log n}{\sqrt{\log\log n}} \implies N(0,1).$$

In other words, for any $x \in \mathbb{R}$ we have

$$\frac{1}{n}\#\left\{1 \leq k \leq n : g(k) \leq \log\log n + k\sqrt{\log\log n}\right\} \xrightarrow[n\to\infty]{} \Phi(x).$$

Proof. See [Durrett, p. 119-124]. The proof uses the moment method. □

Note that Y_n can be written in the form $\sum_{p\leq n} \mathbf{1}_{\{p|X_n\}}$, namely the sum over all primes $p \leq n$ of the indicator of the event that X_n is divisible by p. The probability that X_n is divisible by p is roughly $1/p$, at least if p is significantly smaller than n. Therefore we can expect Y_n to be on the average around

$$\sum_{\text{prime } p\le n} \frac{1}{p},$$

a sum that is known (thanks to Euler) to behave roughly like $\log\log n$. The Erdös-Kac theorem is intuitively related to the observation that these indicators $\mathbf{1}_{\{p|X_n\}}$ for different p's are also close to being independent (a fact which follows from the Chinese remainder theorem). Of course, they are only approximately independent, and making these observations precise is the challenge to proving the theorem. In fact, many famous open problems in number theory (even the Riemann Hypothesis, widely considered to be the most important open problem in mathematics) can be formulated in terms of a statement about approximate independence (in some loose sense) of some arithmetic sequence relating to the prime numbers.

13.2.3 The Euclidean algorithm

As a final example from number theory, consider the following problem: For some $n \ge 1$, choose X_n and Y_n independently and uniformly at random in $\{1, 2, \dots, n\}$, and compute their greatest common divisor (g.c.d.) using the Euclidean algorithm. Let N_n be the number of division (with remainder) steps that were required. For example, if $X = 58$ and $Y = 24$ then the application of the Euclidean algorithm would result in the sequence of steps

$$(58, 24) \to (24, 10) \to (10, 4) \to (4, 2) \to (2, 0),$$

so 4 division operations were required (and the g.c.d. is 2).

Theorem 25 (CLT for the number of steps in the Euclidean algorithm; D. Hensley (1992))**.** *There exists a constant σ_∞ (which has a very complicated definition) such that*

$$\frac{N_n - \frac{12\log 2}{\pi^2}\log n}{\sigma_\infty\sqrt{\log n}} \implies N(0,1).$$

Hensley's theorem was in recent years significantly generalized and the techniques extended by Brigitte Vallée, a French mathematician. The fact that the average value of N_n is approximately $(12\log 2/\pi^2)\log n$ was previously known from work of Heilbronn and Dixon in 1969–1970, using ideas dating back to Gauss, who discovered the probability distribution now called the "Gauss measure". This is the probability distribution on $(0, 1)$ with density $\frac{1}{\log 2(1+x)}$, which Gauss found (but did not prove!) describes the limiting distribution of the ratio of a pair of independent $U(0, 1)$ random variables after many iterations of the division-with-remainder step in the Euclidean algorithm.

End of Part III