Chapter 11

The Central Limit Theorem

Printed: April 13, 2020

1. Sums of independent identically distributed random variables

Denote by Z the "standard normal random variable" with density $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

Lemma 11.1. $Ee^{itZ} = e^{-t^2/2}$

Proof. We use the same calculation as for the moment generating function:

$$\int_{-\infty}^{\infty} \exp(itx - \frac{1}{2}x^2)dx = e^{-t^2/2}\int_{-\infty}^{\infty} \exp(-\frac{1}{2}(x-it)^2)dx = \sqrt{2\pi}$$

Note that $e^{-z^2/2}$ is an analytic function so $\oint_\gamma e^{-z^2/2}dz = 0$ over any closed path. So

$$\int_{-A}^{A} \exp -(x-it)^2/2dx - \int_{-A}^{A} e^{-x^2/2}dx + \int_0^{it} \exp(-(A-is)^2/2)ds - \int_0^{it} \exp(-(-A-is)^2/2)ds = 0$$

□

Theorem 11.2 (CLT for i.i.d.). *Suppose $\{X_n\}$ is i.i.d. with mean m and variance $0 < \sigma^2 < \infty$. Let $S_n = X_1 + \cdots + X_n$. Then*

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} Z$$

This is one of the special cases of the Lindeberg theorem and the proof uses characteristic functions. Note that $\varphi_{S_n/\sqrt{n}}(t) = e^{-t^2/2}$ when X_j are independent $N(0,1)$.

In general, $\varphi_{S_n/\sqrt{n}}(t)$ is a complex number. For example, when X_n are exponential with parameter $\lambda = 1$, the conclusion says that

$$\varphi_{S_n/\sqrt{n}}(t) = \frac{e^{-it\sqrt{n}}}{\left(1 - i\frac{t}{\sqrt{n}}\right)^n} \to e^{-t^2/2}$$

which is not so obvious to see. On the other hand, characteristic function in Exercise 10.5 on page 119 is real and the limit can be found using calculus:

$$\varphi_{S_n/\sqrt{n}}(t) = \cos^n(t/\sqrt{n}) \to e^{-t^2/2}.$$

Here is a simple inequality that will suffice for the proof in the general case.

Lemma 11.3. *If $z_1, \dots, z_m$ and $w_1, \dots, w_m$ are complex numbers of modulus at most 1 then*

$$|z_1 \dots z_m - w_1 \dots w_m| \leq \sum_{k=1}^{m} |z_k - w_k| \tag{11.1}$$

Proof. Write the left hand side of (11.1) as a telescoping sum:

$$z_1 \dots z_m - w_1 \dots w_m = \sum_{k=1}^{m} z_1 \dots z_{k-1}(z_k - w_k)w_{k+1} \dots w_m$$

□

(Omitted in 2020)

Example 11.1. We show how to complete the proof for the exponential distribution.

$$\left|\frac{e^{-it\sqrt{n}}}{\left(1 - i\frac{t}{\sqrt{n}}\right)^n} - e^{-t^2/2}\right| = \left|\left(\frac{e^{-it/\sqrt{n}}}{1 - i\frac{t}{\sqrt{n}}}\right)^n - (e^{-t^2/(2n)})^n\right| \leq n\left|\frac{e^{-it/\sqrt{n}}}{1 - i\frac{t}{\sqrt{n}}} - e^{-t^2/(2n)}\right|$$

$$= n\left|\frac{1 - it/\sqrt{n} + t^2/(2n) + it^3/(6n\sqrt{n}) - \dots}{1 - i\frac{t}{\sqrt{n}}} - 1 + t^2/(2n) - t^4/(6n^2) + \dots\right|$$

$$= n\left|\left(1 - \frac{it}{\sqrt{n}} - \frac{t^2}{2n} - \frac{it^3}{6n\sqrt{n}} + \dots\right)\left(1 + i\frac{t}{\sqrt{n}} - \frac{t^2}{n} + \dots\right) - 1 + \frac{t^2}{2n} - \frac{t^4}{6n^2} + \dots\right|$$

$$= n\left|(1 - \frac{t^2}{n} + \frac{t^2}{2n} + i\frac{t^3}{6n\sqrt{n}} - \dots - 1 + \frac{t^2}{2n} - \frac{t^4}{6n^2} + \dots\right| \leq n\frac{C(t)}{n\sqrt{n}} \to 0.$$

Proof of Theorem 11.2. Without loss of generality we may assume $m = 0$ and $\sigma = 1$. We have $\varphi_{S_n/\sqrt{n}}(t) = \varphi_X(t/\sqrt{n})^n$. For a fixed $t \in \mathbb{R}$ choose n large enough so that $1 - \frac{t^2}{2n} > -1$. For such n, we can apply (11.1) with $z_k = \varphi_X(t/\sqrt{n})$ and $w_k = 1 - \frac{t^2}{2n}$. We get

$$\left|\varphi_{S_n/\sqrt{n}}(t) - \left(1 - \tfrac{t^2}{2n}\right)^n\right| \leq n\left|\varphi_X(t/\sqrt{n}) - 1 - \frac{t^2}{2n}\right| \leq t^2 E \min\left\{\frac{|t||X|^3}{\sqrt{n}}, X^2\right\}$$

Noting that $\lim_{n\to\infty} \min\{|t||X|^3/\sqrt{n}, X^2\} = 0$, by dominated convergence theorem (the integrand is dominated by the integrable function X^2) we have $E \min\left\{\frac{|t||X|^3}{\sqrt{n}}, X^2\right\} \to 0$ as $n \to \infty$. So

$$\lim_{n\to\infty}\left|\varphi_{S_n/\sqrt{n}}(t) - \left(1 - \frac{t^2}{2n}\right)^n\right| = 0.$$

It remains to notice that $(1-\frac{t^2}{2n})^n \to e^{-t^2/2}$. ☐

Remark 11.4. If $X_n \xrightarrow{\mathcal{D}} Z$ then the cumulative distribution functions converge uniformly: $\sup_n |P(X_n \leq x) - P(Z \leq x)| \to 0$.

Example 11.2 (Normal approximation to Binomial)**.** If X_n is $Bin(n,p)$ and p is fixed then $P(\frac{1}{n}X_n < p + x/\sqrt{n}) \to P(Z \leq x\sqrt{p(1-p)}$ as $n \to \infty$.

Example 11.3 (Normal approximation to Poisson)**.** If X_λ is $Poiss$ and p is fixed then $(X_\lambda - \lambda)/\sqrt{\lambda} \xrightarrow{\mathcal{D}} Z$ as $\lambda \to \infty$. (Strictly speaking, the CLT gives only convergence of $(X_{\lambda n} - \lambda n)/\sqrt{n\lambda} \xrightarrow{\mathcal{D}} Z$ as $n \to \infty$.)

(Omitted in 2020)

2. General form of a limit theorem

The general problem of convergence in distribution can be stated as follows: Given a sequence Z_n of random variables, find normalizing constants a_n, b_n and a limiting distribution/random variable Z such that $(Z_n - b_n)/a_n \to Z$.

In Example 9.1, Z_n is a maximum, $a_n = 1$, $b_n = \log n$.

In Theorem 11.2, Z_n is the sum, the normalizing constants are $b_n = E(S_n)$ and $a_n = \sqrt{Var(S_n)}$, and we will make the same choice for sums of independent random variables in the next section. However, finding an appropriate normalization for CLT may be not obvious or easy, see Section 5.

One may wonder how much flexibility do we have in the choice of the normalizing constants a_n, b_n

Theorem 11.5 (Convergence of types)**.** *Suppose $X_n \xrightarrow{\mathcal{D}} X$ and $a_nX_n + b_n \xrightarrow{\mathcal{D}} Y$ for some $a_n > 0$, $b_n \in \mathbb{R}$, and both X, Y are non-degenerate. Then $a_n \to a > 0$ and $b_n \to b$ and in particular Y has the same law as $aX + b$.*

So if $(Z_n - b_n)/a_n \to Z$ and $(Z_n - b'_n)/a'_n \to Z'$ then $(Z_n - b'_n)/a'_n = \frac{a_n}{a'_n}((Z_n - b_n)/a_n) + (b_n - b'_n)/a'_n$, which means that $a_n/a'_n \to a > 0$ and $(b_n - b'_n)/a'_n \to b$. So $a'_n = a_n/a$, $b'_n = b_n - \frac{b}{a}a_n$ and $Z' = aZ + b$.

(Omitted in 2020)

Proof. To be written... ☐

It is clear that independence alone is not sufficient for the CLT.

3. Lindeberg's theorem

The setting is of sums of triangular arrays: For each n we have a family of independent random variables

$$X_{n,1}, \dots, X_{n,r_n}$$

and we set $S_n = X_{n,1} + \cdots + X_{n,r_n}$.

For Theorem 11.2, the triangular array can be $X_{n,k} = \frac{X_k - m}{\sigma\sqrt{n}}$. Or one can take $X_{n,k} = \frac{X_k - m}{\sigma}$...

Through this section we assume that random variables are square-integrable with mean zero, and we use the notation

$$E(X_{n,k}) = 0, \ \sigma^2_{nk} = E(X^2_{n,k}), \ s^2_n = \sum_{k=1}^{r_n} \sigma^2_{nk} \tag{11.2}$$

Definion 11.1 (The Lindeberg condition)**.** We say that the *Lindeberg condition* holds if

$$\lim_{n\to\infty} \frac{1}{s_n^2}\sum_{k=1}^{r_n}\int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP = 0 \text{ for all } \varepsilon > 0 \tag{11.3}$$

(Note that strict inequality $\int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP$ can be replaced by $\int_{|X_{nk}|\geq\varepsilon s_n} X_{nk}^2 dP$ and the resulting condition is the same.)

Remark 11.6. Under the Lindeberg condition, we have

$$\lim_{n\to\infty}\max_{k\leq r_n}\frac{\sigma_{nk}^2}{s_n^2} = 0 \tag{11.4}$$

Indeed,

$$\sigma_{nk}^2 = \int_{|X_{nk}|\leq\varepsilon s_n} X_{nk}^2 dP + \int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP \leq \varepsilon s_n^2 + \int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP$$

So

$$\max_{k\leq r_n}\frac{\sigma_{nk}^2}{s_n^2} \leq \varepsilon + \frac{1}{s_n^2}\max_{k\leq r_n}\int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP \leq \varepsilon + \frac{1}{s_n^2}\sum_{k=1}^{r_n}\int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP$$

Theorem 11.7 (Lindeberg CLT)**.** *Suppose that for each n the sequence $X_{n1}\dots X_{n,r_n}$ is independent with mean zero. If the Lindeberg condition holds for all $\varepsilon > 0$ then $S_n/s_n \xrightarrow{\mathcal{D}} Z$.*

Example 11.4 (Proof of Theorem 11.2)**.** In the setting of Theorem 11.2, we have $X_{n,k} = \frac{X_k - m}{\sigma}$ and $s_n = \sqrt{n}$. The Lindeberg condition is

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\int_{|X_k-m|>\varepsilon\sigma\sqrt{n}}\frac{(X_k-m)^2}{\sigma^2}dP = \lim_{n\to\infty}\frac{1}{\sigma^2}\int_{|X_1-m|>\varepsilon\sigma\sqrt{n}}(X_1-m)^2 = 0$$

by Lebesgue dominated convergence theorem, say. (Or by Corollary 6.12 on page 71.)

Proof. Without loss of generality we may assume that $s_n^2 = 1$ so that $\sum_{k=1}^{r_n}\sigma_{nk}^2 = 1$. Denote $\varphi_{nk} = E(e^{itX_{nk}})$. From (10.13) we have

$$\left|\varphi_{nk}(t) - (1 - \frac{1}{2}t^2\sigma_{nk}^2)\right| \leq E\left(\min\{|tX_{nk}|^2, |tX_{nk}|^3\}\right) \tag{11.5}$$

$$\leq \int_{|X_{nk}|<\varepsilon}|tX_{nk}|^3 dP + \int_{|X_{nk}|\geq\varepsilon}|tX_{nk}|^2 dP \leq t^3\varepsilon\sigma_{nk}^2 + t^2\int_{|X_{nk}|\geq\varepsilon}X_{nk}^2 dP$$

Using (11.1), we see that

$$\left|\varphi_{S_n}(t) - \prod_{k=1}^{n}(1 - \frac{1}{2}t^2\sigma_{nk}^2)\right| \leq \varepsilon t^3\sum_{k=1}^{n}\sigma_{nk}^2 + t^2\sum_{k=1}^{n}\int_{|X_{nk}|>\varepsilon}|X_{nk}^2 dP \tag{11.6}$$

This shows that

$$\lim_{n\to\infty}\left|\varphi_{S_n}(t) - \prod_{k=1}^{n}(1 - \tfrac{1}{2}t^2\sigma_{nk}^2)\right| = 0$$

It remains to verify that $\lim_{n\to\infty}\left|e^{-t^2/2}-\prod_{k=1}^{n}(1-\frac{1}{2}t^2\sigma_{nk}^2)\right|=0$.

To do so, we apply the previous proof to the triangular array $\sigma_{n,k}Z_k$ of independent normal random variables. Note that

$$\varphi_{\sum Z_{nk}}(t)=\prod_{k=1}^{r_n}e^{-t^2\sigma_{nk}^2/2}=e^{-t^2/2}$$

We only need to verify the Lindeberg condition for $\{Z_{nk}\}$:

$$\int_{Z_{nk}>\varepsilon}Z_{nk}^2dP=\sigma_{nk}^2\int_{|x|>\varepsilon/\sigma_{nk}}x^2f(x)dx$$

So

$$\sum_{k=1}^{r_n}\int_{Z_{nk}>\varepsilon}Z_{nk}^2dP\le\sum_{k=1}^{r_n}\sigma_{nk}^2\int_{|x|>\varepsilon/\sigma_{nk}}x^2f(x)dx\le\max_{1\le k\le r_n}\int_{|x|>\varepsilon/\sigma_{nk}}x^2f(x)dx\le\int_{|x|>\varepsilon/\max_k\sigma_{nk}}x^2f(x)dx$$

The right hand side goes to zero as $n\to\infty$, because by $\max_{1\le k\le r_n}\sigma_{nk}\to 0$ by (11.4). □

4. Lyapunov's theorem

Theorem 11.8. *Suppose that for each n the sequence $X_{n1}\dots X_{n,r_n}$ is independent with mean zero. If the* Lyapunov's condition

$$\lim_{n\to\infty}\frac{1}{s_n^{2+\delta}}\sum_{k=1}^{n}E|X_{nk}|^{2+\delta}=0 \tag{11.7}$$

holds for some $\delta>0$, then $S_n/s_n\xrightarrow{\mathcal{D}}Z$

Proof. We use the following bound to verify Lindeberg's condition:

$$\frac{1}{s_n^2}\sum_{k=1}^{r_n}\int_{|X_{nk}|>\varepsilon s_n}X_{nk}^2dP\le\frac{1}{\varepsilon^\delta s_n^{2+\delta}}\sum_{k=1}^{r_n}\int_{|X_{nk}|>\varepsilon s_n}|X_{nk}|^{2+\delta}dP\le\frac{1}{\varepsilon^\delta s_n^{2+\delta}}\sum_{k=1}^{n}E|X_{nk}|^{2+\delta}$$

□

Corollary 11.9. *Suppose X_k are independent with mean zero, variance σ^2 and that $\sup_k E|X_k|^{2+\delta}<\infty$. Then $S_n/\sqrt{n}\xrightarrow{\mathcal{D}}\sigma Z$.*

Proof. Let $C=\sup_k E|X_k|^{2+\delta}$ Then $s_n=\sqrt{n}$ and $\frac{1}{s_n^{2+\delta}}\sum_{k=1}^{n}E(|X_k|^{2+\delta})\le C/n^{\delta/2}\to 0$, so Lyapunov's condition is satisfied. □

Corollary 11.10. *Suppose X_k are independent, uniformly bounded, and have mean zero. If $\sum_n Var(X_n)=\infty$, then $S_n/\sqrt{Var(S_n)}\xrightarrow{\mathcal{D}}N(0,1)$.*

Proof. Suppose $|X_n| \le C$ for a constant C. Then

$$\frac{1}{s_n^3}\sum_{k=1}^{n} E|X_n|^3 \le C\frac{s_n^2}{s_n^3} = \frac{C}{s_n} \to 0$$

□

5. Normal approximation without Lindeberg condition

One basic idea is truncation: $X_n = X_n I_{|X_n|\le a_n} + X_n I_{|X_n|>a_n}$. One wants to show that $\frac{1}{s_n}\sum X_k I_{|X_k|\le a_n} \to Z$ and that $\frac{1}{s_n}\sum X_k I_{|X_k|>a_n} \xrightarrow{P} 0$. Then S_n/s_n is asymptotically normal by Slutski's theorem.

Example 11.5. Let $X_1, X_2, \ldots$ be independent random variables with the distribution ($k \ge 1$)

$$\begin{aligned}
\Pr(X_k &= \pm 1) = 1/4,\\
\Pr(X_k &= k^k) = 1/4^k,\\
\Pr(X_k &= 0) = 1/2 - 1/4^k.
\end{aligned}$$

Then $\sigma_k^2 = \frac{1}{2} + \left(\frac{k}{4}\right)^k$ and $s_n \ge n^n/4^n$. But $S_n/s_n \xrightarrow{\mathcal{D}} 0$ and in fact we have $S_n/\sqrt{n} \xrightarrow{\mathcal{D}} Z/\sqrt{2}$. To see this, note that $Y_k = X_k I_{|X|_k} \le 1$ are independent with mean 0, variance $\frac{1}{2}$ and $P(Y_k \ne X_k) = 1/4^k$ so by the first Borel Cantelli Lemma (Theorem 3.8) $|\frac{1}{\sqrt{n}}\sum_{k=1}^n (Y_k - X_k)| \le \frac{U}{\sqrt{n}} \to 0$ with probability one.

It is sometimes convenient to use Corollary 9.5 (Exercise 9.2) combined with the law of large numbers. This is how one needs to proceed in Exercise 11.2.

Example 11.6. Suppose $X_1, X_2, \ldots,$ are i.i.d. with mean 0 and variance $\sigma^2 > 0$. Then

$$\frac{\sum_{k=1}^n X_k}{\sqrt{\sum_{k=1}^n X_k^2}}$$

converges in distribution to $N(0,1)$. To see this, write

$$\frac{\sum_{k=1}^n X_k}{\sqrt{\sum_{k=1}^n X_k^2}} = \frac{\sigma}{\sqrt{\frac{1}{n}\sum_{k=1}^n X_k^2}} \times \frac{\sum_{k=1}^n X_k}{\sigma\sqrt{n}}$$

and note that the first factor converges to 1 with probability one.

Required Exercises

Exercise 11.1. Suppose a_{nk} is an array of numbers such that $\sum_{k=1}^n a_{nk}^2 = 1$ and $\max_{1\le k\le n}|a_{nk}| \to 0$. Let X_j be i.i.d. with mean zero and variance 1. Show that $\sum_{k=1}^n a_{nk}X_k \xrightarrow{\mathcal{D}} Z$.

Exercise 11.2. Suppose that $X_1, X_2, \dots$ are i.i.d., $\mathbb{E}(X_1) = 1$, $\mathrm{Var}(X_1^2) = \sigma^2 < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$. Show that for all $k > 0$

$$\sqrt{n}\left(\bar{X}_n^k - 1\right) \xrightarrow{\mathcal{D}} N(0, k\sigma)$$

as $n \to \infty$.

Exercise 11.3. Suppose $X_1, X_2, \dots$ are independent, $X_k = \pm 1$ with probability $\frac{1}{2}(1 - k^{-2})$ and $X_k = \pm k$ with probability $\frac{1}{2}k^{-2}$. Let $S_n = \sum_{k=1}^n X_k$

(i) Show that $S_n/\sqrt{n} \xrightarrow{\mathcal{D}} N(0, 1)$

(ii) Is the Lindeberg condition satisfied?

Exercise 11.4. Suppose $X_1, X_2, \dots$ are independent random variables with distribution $\Pr(X_k = 1) = p_k$ and $\Pr(X_k = 0) = 1 - p_k$. Prove that if $\sum Var(X_k) = \infty$ then

$$\frac{\sum_{k=1}^n (X_k - p_k)}{\sqrt{\sum_{k=1}^n p_k(1 - p_k)}} \xrightarrow{\mathcal{D}} N(0, 1).$$

Exercise 11.5. Suppose X_k are independent and have density $\frac{1}{|x|^3}$ for $|x| > 1$. Show that $\frac{S_n}{\sqrt{n\log n}} \to N(0, 1)$.

Hint: Verify that Lyapunov's condition (11.7) holds with $\delta = 1$ for truncated random variables. Several different truncations can be used, but technical details differ:

- $Y_k = X_k I_{|X_k|\le\sqrt{k}}$ is a solution in [Billingsley]. To show that $\frac{1}{\sqrt{n\log n}}\sum_{k=1}^n (X_k - Y_k) \xrightarrow{P} 0$ use L_1-convergence.
- Triangular array $Y_{nk} = X_k I_{|X_k|\le\sqrt{n}}$ is simpler computationally
- Truncation $Y_k = X_k I_{|X_k|\le\sqrt{k}\log k}$ leads to "asymptotically equivalent" sequences.

Exercise 11.6 (stat)**.** A real estate aggent wishes to estimate the unknown mean sale price of a house μ which she believes is well described by the distribution which has finite second moment. She estimates μ by the sample mean $\bar{X}_n$ of the i.i.d. sample $X_1, \dots, X_n$, and she estimates the variance by the expression

$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

She then uses a formula $\bar{X}_n \pm z_\alpha S_n/\sqrt{n}$ from Wikpiedia to produce the large sample confidence interval for μ. To understand why this procedure works, she would like to know that

$$(\bar{X}_n - \mu)/S_n \xrightarrow{\mathcal{D}} N(0, 1).$$

Please supply the proof.

Exercise 11.7 (stat)**.** A psychologist wishes to estimate parameter $\lambda > 0$ of the exponential distribution, see Example 2.4, by taking the average $\bar{X}_n$ of the i.i.d. sample $X_1, \dots, X_n$, and defining $\hat{\lambda}_n = 1/\bar{X}_n$. Show that $\hat{\lambda}_n$ is asymptotically normal, i.e. determine $a_n(\lambda)$ such that the α-confidence interval for λ is

$$\hat{\lambda}_n \pm a_n(\lambda) z_{\alpha/2}$$

where $z_{\alpha/2}$ comes from the normal table $P(Z > z_{\alpha/2}) = \alpha/2$.

Some previous prelim problems

Exercise 11.8 (May 2018)**.** Suppose that $X_1, X_2, \ldots$ are independent random variables with distributions

$$P(X_k = \pm 1) = \frac{1}{2k} \text{ and } P(X_k = 0) = \frac{1-k}{k}.$$

Prove that

$$\frac{1}{\sqrt{\ln n}} \sum_{k=1}^{n} X_k \xrightarrow{\mathcal{D}} N(0,1).$$

Exercise 11.9 (Aug 2017)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be a collection of independent random variables with

$$\mathbb{P}(X_n = \pm n^2) = \frac{1}{2n^\beta} \text{ and } \mathbb{P}(X_n = 0) = 1 - \frac{1}{n^\beta}, n \in \mathbb{N},$$

where $\beta \in (0,1)$ is fixed for all $n \in \mathbb{N}$. Consider $S_n := X_1 + \cdots + X_n$. Show that

$$\frac{S_n}{n^\gamma} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

for some $\sigma > 0, \gamma > 0$. Identify σ and γ as functions of β. You may use the formula

$$\sum_{k=1}^{n} k^\theta \sim \frac{n^{\theta+1}}{\theta+1}$$

for $\theta > 0$, and recall that by $a_n \sim b_n$ we mean $\lim_{n\to\infty} a_n/b_n = 1$.

Exercise 11.10 (May 2017)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be independent random variables with $\mathbb{P}(X_n = 1) = 1/n = 1 - \mathbb{P}(X_n = 0)$. Let $S_n := X_1 + \cdots + X_n$ be the partial sum.

(i) Show that

$$\lim_{n\to\infty} \frac{\mathbb{E}S_n}{\log n} = 1 \quad \text{and} \quad \lim_{n\to\infty} \frac{\operatorname{Var}(S_n)}{\log n} = 1.$$

(ii) Prove that

$$\frac{S_n - \log n}{\sqrt{\log n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

as $n \to \infty$. Explain which central limit theorem you use. State and verify all the conditions clearly.

Hint: recall the relation $\lim_{n\to\infty} \frac{\sum_{k=1}^{n} 1/k}{\log n} = 1$.

Exercise 11.11 (May 2016)**.**(a) State Lindeberg–Feller central limit theorem.

(b) Use Lindeberg–Feller central limit theorem to prove the following. Consider a triangular array of random variables $\{Y_{n,k}\}_{n\in\mathbb{N},k=1,\ldots,n}$ such that for each n, $\mathbb{E}Y_{n,k}=0, k=1,\ldots,n$, and $\{Y_{n,k}\}_{k=1,\ldots,n}$ are independent. In addition, with $\sigma_n := (\sum_{k=1}^n \mathbb{E}Y_{n,k}^2)^{1/2}$, assume that

$$\lim_{n\to\infty}\frac{1}{\sigma_n^4}\sum_{k=1}^n \mathbb{E}Y_{n,k}^4 = 0.$$

Show that

$$\frac{Y_{n,1}+\cdots+Y_{n,n}}{\sigma_n}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

Exercise 11.12 (Aug 2015)**.** Let $\{U_n\}_{n\in\mathbb{N}}$ be a collection of i.i.d. random variables with $\mathbb{E}U_n=0$ and $\mathbb{E}U_n^2=\sigma^2\in(0,\infty)$. Consider random variables $\{X_n\}_{n\in\mathbb{N}}$ defined by $X_n=U_n+U_{2n}, n\in\mathbb{N}$, and the partial sum $S_n=X_1+\cdots+X_n$. Find appropriate constants $\{a_n,b_n\}_{n\in\mathbb{N}}$ such that

$$\frac{S_n-b_n}{a_n}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

Exercise 11.13 (May 2015)**.** Let $\{U_n\}_{n\in\mathbb{N}}$ be a collection of i.i.d. random variables distributed uniformly on interval $(0,1)$. Consider a triangular array of random variables $\{X_{n,k}\}_{k=1,\ldots,n,n\in\mathbb{N}}$ defined as

$$X_{n,k}=\mathbf{1}_{\{\sqrt{n}U_k\le 1\}}-\frac{1}{\sqrt{n}}.$$

Find constants $\{a_n,b_n\}_{n\in\mathbb{N}}$ such that

$$\frac{X_{n,1}+\cdots+X_{n,n}-b_n}{a_n}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

Exercise 11.14 (Aug 2014)**.** Let $X_1,X_2,\ldots$ be independent and identically distributed random variables with

$$P(X_i=1)=P(X_i=-1)=1/2.$$

Prove that

$$\frac{\sqrt{3}}{\sqrt{n^3}}\sum_{k=1}^n kX_k\xrightarrow{\mathcal{D}}N(0,1)$$

(You may use formulas $\sum_{j=1}^n j^2=\frac{1}{6}n(n+1)(2n+1)$ and $\sum_{j=1}^n j^3=\frac{1}{4}n^2(n+1)^2$ without proof.)

Exercise 11.15 (May 2014)**.** Let $\{X_{nk}:k=1,\ldots,n,n\in\mathbb{N}\}$ be a family of independent random variables satisfying

$$P\left(X_{nk}=\frac{k}{\sqrt{n}}\right)=P\left(X_{nk}=-\frac{k}{\sqrt{n}}\right)=P(X_{nk}=0)=1/3$$

Let $S_n=X_{n1}+\cdots+X_{nn}$. Prove that S_n/s_n converges in distribution to a standard normal random variable for a suitable sequence of real numbers s_n.

Some useful identities:

$$\begin{aligned}\sum_{k=1}^n k &= \frac{1}{2}n(n+1)\\ \sum_{k=1}^n k^2 &= \frac{1}{6}n(n+1)(2n+1)\\ \sum_{k=1}^n k^3 &= \frac{1}{4}n^2(n+1)^2\end{aligned}$$

Exercise 11.16 (Aug 2013)**.** Suppose $X_1, Y_1, X_2, Y_2, \ldots,$ are independent identically distributed with mean zero and variance 1. For integer n, let

$$U_n = \frac{1}{n}\left(\sum_{j=1}^{n} X_j\right)^2 + \frac{1}{n}\left(\sum_{j=1}^{n} Y_j\right)^2 .$$

Prove that $\lim_{n\to\infty} P(U_n \le u) = 1 - e^{-u/2}$ for $u > 0$.

Exercise 11.17 (May 2013)**.** Suppose $X_{n,1}, X_{n,2}, ...$ are independent random variables centered at expectations (mean 0) and set $s_n^2 = \sum_{k=1}^{n} E\left((X_{n,k})^2\right)$. Assume for all k that $|X_{n,k}| \le M_n$ with probability 1 and that $M_n/s_n \to 0$. Let $Y_{n,i} = 3X_{n,i} + X_{n,i+1}$. Show that

$$\frac{Y_{n,1} + Y_{n,2} + ... + Y_{n,n}}{s_n}$$

converges in distribution and find the limiting distribution.

Limit Theorems in $\mathbb{R}^k$

This is based on [Billingsley, Section 29]. Printed: April 13, 2020

1. The basic theorems

If $\mathbf{X} : \Omega \to \mathbb{R}^k$ is measurable, then $\mathbf{X}$ is called a random vector. $\mathbf{X}$ is also called a k-variate random variable, as $\mathbf{X} = (X_1, \dots, X_k)$.

Recall that a probability distribution of $\mathbf{X}$ is a probability measure μ on Borel subsets of $\mathbb{R}^k$ defined by $\mu(U) = P(\{\omega : \mathbf{X}(\omega) \in U\})$.

Recall that a (joint) cumulative distribution function of $\mathbf{X} = (X_1, \dots, X_n)$ is a function $F : \mathbb{R}^n \to [0, 1]$ such that

$$F(x_1, \dots, x_k) = P(X_1 \le x_1, \dots, X_k \le x_k)$$

From $\pi - \lambda$ theorem we know that F determines uniquely μ. In particular, if

$$F(x_1, \dots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(y_1, \dots, y_k) dy_1 \dots dy_k$$

then $\mu(U) = \int_U f(y_1, \dots, y_k) dy_1 \dots dy_k$.

Let $X_n : \Omega \to \mathbb{R}^k$ be a sequence of random vectors.

Definion 12.1. We say that $\mathbf{X}_n$ *converges in distribution* to $\mathbf{X}$ if for every bounded continuous function $f : \mathbb{R}^k \to \mathbb{R}$ the sequence of numbers $E(f(\mathbf{X}_n)$ converges to $Ef(\mathbf{X})$.

We will write $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}$; if μ_n is the law of $\mathbf{X}_n$ we will also write $\mu_n \to D$; the same notation in the language of cumulative distribution functions is $F_n \xrightarrow{\mathcal{D}} F$; the latter can be defined as $F_n(\mathbf{x}) \xrightarrow{\mathcal{D}} F(\mathbf{x})$ for all points of continuity of F, but it is simpler to use Definition 12.1.

Proposition 12.1. *If $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}$ and $g : \mathbb{R}^k \to \mathbb{R}^m$ is a continuous function then $g(\mathbf{X}_n) \xrightarrow{\mathcal{D}} g(\mathbf{X})$*

For example, if $(X_n, Y_n) \xrightarrow{\mathcal{D}} (Z_1, Z_2)$ then $X_n^2 + Y_n^2 \xrightarrow{\mathcal{D}} Z_1^2 + Z_2^2$.

Proof. Denoting by $\mathbf{Y}_n = g(\mathbf{X}_n)$, we see that for any bounded continuous function $f : \mathbb{R}^m \to \mathbb{R}$, $f(bY_n)$ is a bounded continuous function $f \circ g$ of $\mathbf{X}_n$. □

(Omitted in 2020) The following is a k-dimensional version of Portmanteau Theorem 9.7

Theorem 12.2. *For a sequence μ_n of probability measures on the Borel sets of $\mathbb{R}^k$, the following are equivalent:*

(i) $\mu_n \xrightarrow{\mathcal{D}} \mu$

(ii) $\limsup_{n\to\infty} \mu_n(C) \le \mu(C)$ *for all closed sets* $C \subset \mathbb{R}^k$.

(iii) $\liminf_{n\to\infty} \mu_n(G) \le \mu(G)$ *for all open sets* $G \subset \mathbb{R}^k$.

(iv) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for all sets* $A \subset \mathbb{R}^k$ *such that* $\mu(\partial A) = 0$

Proof. The detailed proof is omitted. Here are some steps:

- By passing to complements, it is clear that (2) and (3) are equivalent.
- Since the interior A° of a set A is its subset, $A^\circ \subset A \subset \bar{A}$. So $\mu_n(A^\circ) \le \mu_n(A) \le \mu_n(\bar{A})$ and we get
$$\mu(A^\circ) \le \liminf \mu_n(A) \le \limsup \mu_n(A) \le \mu(\bar{A})$$
Since $\partial A = \bar{A} \setminus A^\circ$, we have $\mu(A^\circ) = \mu(\bar{A}) = \mu(A)$ so it is clear that (2)+(3) imply (4).
- To see how (1) implies (2), fix closed set $C \subset \mathbb{R}^k$ and consider a bounded continuous function f such that $f = 1$ on C and $f = 0$ on $C^\varepsilon = \{\mathbf{x} \in \mathbb{R}^k : d(\mathbf{x}, C) \le \varepsilon\}$ Then $\mu_n(C) \le \int f(\mathbf{x})\mu_n(d\mathbf{x}) \to \int f\mu(d\mathbf{x}) \le \mu(C^\varepsilon)$. Since $\lim_{\varepsilon\to 0} \mu(C^\varepsilon) = \mu(\bigcap_{\varepsilon>0} C_\varepsilon) = \mu(C)$, we get the conclusion.

□

Definion 12.2. The sequence of measures μ_n on $\mathbb{R}^k$ is tight if for every $\varepsilon > 0$ there exists a compact set $K \subset \mathbb{R}^k$ such that $\mu_n(K) \ge 1 - \varepsilon$ for all n.

Theorem 12.3. *If μ_n is a tight sequence of probability measures then there exists μ and a subsequence n_k such that $\mu_{n_k} \xrightarrow{\mathcal{D}} \mu$*

Proof. The detailed proof is omitted.

Here are the main steps in the proof: □

Corollary 12.4. *If $\{\mu_n\}$ is a tight sequence of probability measures on Borel subsets of $\mathbb{R}^k$ and if each convergent subsequence has the same limit μ, then $\mu_n \xrightarrow{\mathcal{D}} \mu$*

2. Multivariate characteristic function

Recall the dot-product $\mathbf{x}\cdot\mathbf{y} := \mathbf{x}'\mathbf{y} \sum_{j=1}^k x_j y_j$. The multivariate characteristic function $\varphi : \mathbb{R}^k \to \mathbb{C}$ is

$$\varphi(\mathbf{t}) = \mathbb{E}\exp(i\mathbf{t}\cdot\mathbf{X}) \tag{12.1}$$

This is also written as $\varphi(t_1, \dots, t_k) = E\exp(\sum_{j=1}^k t_j X_j)$.

The inversion formula shows how to determine $\mu(U)$ for a rectangle $U = (a_1, b_1] \times (a_2, b_2] \times \dots \times (a_k, b_k]$ such that $\mu(\partial U) = 0$:

$$\mu(U) = \lim_{T\to\infty} \frac{1}{(2\pi)^k} \int_{-T}^{T} \cdots \int_{-T}^{T} \prod_{j=1}^{k} \frac{e^{-ia_kjt_j} - e^{-ib_jt_j}}{it_j} \varphi(t_1, \dots, t_k) dt_1 \dots dt_k \tag{12.2}$$

Thus the characteristic function determines the probability measure μ uniquely.

Corollary 12.5 (Cramer-Wold devise)**.** *The law of* $\mathbf{X}$ *is uniquely determined by the univariate laws* $\mathbf{t} \cdot \mathbf{X} = \sum_{j=1}^{k} t_j X_j$.

Corollary 12.6. X, Y *are independent iff* $\varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t)$

Theorem 12.7. $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{Y}$ *iff* $\varphi_n(\mathbf{t}) \to \varphi(\mathbf{t})$ *for all* $\mathbf{t} \in \mathbb{R}^k$.

Note that this means that $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{Y}$ iff $\sum t_j X_j(n) \xrightarrow{\mathcal{D}} \sum t_j Y_j$ for all $t_1, \dots, t_k$

Example 12.1. If X, Y are independent normal then $X + Y$ and $X - Y$ are independent normal. Indeed, $\varphi_{X+Y,X-Y}(s,t) = \varphi_X(s+t)\varphi_Y(s-t) = \exp((t+s)^2/2 + (s-t)^2/2) = e^{s^2}e^{t^2}$, and $\varphi_{X\pm Y}(s) = e^{s^2/2}e^{s^2/2} = e^{s^2}$.

Corollary 12.8. *If* $Z_1, \dots Z_m$ *are independent normal and* $\mathbf{X} = \mathbf{AZ}$ *then* $\sum_j t_j X_j$ *is (univariate) normal.*

Proof. Lets simplify the calculations by assuming Z_j are standard normal. The characteristic function of $S = \sum_j t_j X_j$ is

$$\varphi(s) = E\exp(is\mathbf{t} \cdot \mathbf{X}) = E\exp(is\mathbf{t} \cdot \mathbf{AZ}) = E\exp(is(\mathbf{A}^T\mathbf{t}) \cdot \mathbf{Z}) = \prod_{i=1}^{m} e^{-s^2[\mathbf{A}^T\mathbf{t}]_i^2/2} = e^{-s^2\|\mathbf{A}^T\mathbf{t}\|^2/2}$$

□

The generalization of this property is the simplest definition of the mutlivariate normal distribution. Note that

$$\|\mathbf{A}^T\mathbf{t}\|^2 = (\mathbf{A}^T\mathbf{t}) \cdot (\mathbf{A}^T\mathbf{t}) = \mathbf{t}'\mathbf{A}\mathbf{A}^T\mathbf{t} = \mathbf{t}'\Sigma\mathbf{t}$$

3. Multivariate normal distribution

Two equivalent approaches

Definion 12.3. **X** is *multivariate normal* if there is a vector **m** and a positive-definite matrix Σ such that its characteristic function is

$$\varphi(\mathbf{t}) = \exp\left(i\mathbf{m}'\mathbf{t} - \tfrac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right)$$

Notation: $N(\mathbf{m}, \Sigma)$. (How do we know that this is a characteristic function? See the proof of Corollary 12.8!)

We need to show that this is indeed a characteristic function! But if it is, then by differentiation the parameters have interpretation: $\mathbb{E}\mathbf{X} = \mathbf{m}$ and $\Sigma_{i,j} = \mathrm{cov}(\mathrm{X_i}, \mathrm{X_j})$.

Remark 12.9. If **X** is normal $N(\mathbf{m}, \Sigma)$, then $\mathbf{X} - \mathbf{m}$ is centered normal $N(0, \Sigma)$. In the sequel, to simplify notation we only discuss centered case.

The simplest way to define the univariate distribution is to start with a standard normal random variable Z with density $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, and then define the general normal as the linear function $X = \mu + \sigma Z$. It is then easy to work out the density of X and the characteristic function, which is $\varphi_X(t) = e^{it\mu + \frac{1}{2}\sigma^2 t^2}$.

Exercise 12.1 is worth doing in two ways - using both definitions.

In $\mathbb{R}^k$ the role of the standard normal distribution is played by the distribution of $\mathbf{Z} = (Z_1, \ldots, Z_k)$ of i.i.d. $N(0,1)$ r.v.. Their density is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}\|\mathbf{x}\|^2} \tag{12.3}$$

The characteristic function $Ee^{i\mathbf{t}'\mathbf{Z}}$ is just the product of the individual characteristic functions $\prod_{j=1}^n e^{-t_j^2/2}$ which in vector notation is

$$\varphi_{\mathbf{Z}}(\mathbf{y}) = e^{-\|\mathbf{t}\|^2/2}$$

Definion 12.4. We will say that **X**, written as a column vector, has *multivariate normal distribution* if $\mathbf{X} = \mathbf{m} + \mathbf{A}\mathbf{Z}$.

Clearly, $E(\mathbf{X}) = \mathbf{m}$. In the sequel we will only consider centered multivariate normal distribution with $E(\mathbf{X}) = 0$.

Remark 12.10. Denoting by $\mathbf{a}_k$ the columns of **A**, we have $\mathbf{X} = \sum_{j=1}^k Z_j \mathbf{a}_j$. This is the universal feature of Gaussian vectors, even in infinite-dimensional vector spaces – they all can be written as linear combinations of deterministic vectors with independent real-valued "noises" as coefficients. For example, the random "vector" $(W_t)_{0\leq t\leq 1}$ with values in the vector space $C[0,1]$ of continuous functions on $[0,1]$ can be written as $W_t = \sum_{k=1}^\infty Z_j g_j(t)$ with deterministic functions $g_j(t) = \frac{1}{2j+1}\sin((2j+1)\pi t)$.

Proposition 12.11. *The characteristic function of the centered normal distribution is*

$$\varphi(\mathbf{t}) = \exp\left(-\tfrac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right) \tag{12.4}$$

where Σ is a $k \times k$ positive definite matrix.

Proof. This is just a calculation:

$$Ee^{i\mathbf{t}'\mathbf{X}} = Ee^{i\mathbf{t}'\mathbf{AZ}} = Ee^{i(\mathbf{A}'\mathbf{t})'\mathbf{Z}} = e^{-\|\mathbf{A}'\mathbf{t}\|^2/2} = \exp\left(-\tfrac{1}{2}(\mathbf{A}'\mathbf{t})'\mathbf{A}'\mathbf{t}\right) = \exp\left(-\tfrac{1}{2}\mathbf{t}'\mathbf{AA}'\mathbf{t}\right) = \exp\left(-\tfrac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right)$$

□

Remark 12.12. Notice that $E(\mathbf{XX}') = E(\mathbf{AZZ}'\mathbf{A}') = \mathbf{A}E(\mathbf{ZZ}')\mathbf{A}' = \mathbf{A}I\mathbf{A}' = \Sigma$ is the *covariance matrix* of $\mathbf{X}$.

Remark 12.13. From linear algebra, any positive definite matrix $\Sigma = U\Lambda U'$ so each such matrix can be written as $\Sigma = \mathbf{AA}'$ with $\mathbf{A} = U\Lambda^{1/2}U'$. So $\varphi(\mathbf{t}) = \exp(-\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t})$ is a characteristic function of $\mathbf{X} = \mathbf{AZ}$.

Remark 12.14. If $\det(\Sigma) > 0$ then $\det \mathbf{A} \neq 0$ and (by linear algebra) the inverse $\mathbf{A}^{-1}$ exists. The density of $\mathbf{X}$ is recalculated from (12.3) as follows

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}\det(\mathbf{A})}e^{-\frac{1}{2}\|\mathbf{A}^{-1}\mathbf{x}\|^2} = \frac{1}{(2\pi)^{k/2}\det(\Sigma)^{1/2}}e^{-\frac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x}}$$

Remark 12.15. Matrix $\mathbf{A}$ in the representation $\mathbf{X} = \mathbf{AZ}$ is not unique, but the covariance matrix $\Sigma = \mathbf{AA}'$ is unique. For example

$$\mathbf{X} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

can also be represented as

$$\mathbf{X} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

(Exercise 12.6)

Example 12.2. Suppose $\varphi(s,t) = e^{-s^2/2 - t^2/2 - \rho st}$. Then

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

is non-negative definite for any $|\rho| \leq 1$ and this is a characteristic function of a random variable $\mathbf{X} = (X_1, X_2)$ with univariate N(0,1) laws, with correlation $E(X_1X_2) = -\frac{\partial^2}{\partial s \partial t}\varphi(s,t)|_{s=t=0} = \rho$. If Z_1, Z_2 are independent $N(0,1)$ then

$$X_1 = Z_1,\ X_2 = \rho Z_1 + \sqrt{1-\rho^2}Z_2 \tag{12.5}$$

will have exactly the same second moments, and the same characteristic function.

Since $\det \Sigma = 1 - \rho^2$, when $\rho^2 \neq 1$ the matrix is invertible and the resulting bivariate normal density is

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)}\right)$$

From (12.5) we also see that $X_2 - \rho X_1$ is independent of X_1 and has variance $1 - \rho^2$

4. The CLT

Theorem 12.16. *Let $\mathbf{X}_n = (X_{n1}, \dots, X_{nk})$ be independent random vectors with the same distribution and finite second moments. Denote $\mathbf{m} = E\mathbf{X}_k$ and $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$. Then*

$$(\mathbf{S}_n - n\mathbf{m})/\sqrt{n} \xrightarrow{\mathcal{D}} \mathbf{Y}$$

where $\mathbf{Y}$ is a centered normal distribution with the covariance matrix $\Sigma = E(\mathbf{X}_n\mathbf{X}_n') - \mathbf{m}\mathbf{m}'$.

The notation is $N(0, \Sigma)$. Note that this is inconsistent with the univariate notation $N(\mu, \sigma)$ which for consistency with the multivariate case should be replaced by $N(\mu, \sigma^2)$.

Proof. Without loss of generality we can assume $\mathbf{m} = 0$. Let $\mathbf{t} \in \mathbb{R}^k$. Then $X_n := \mathbf{t}'\mathbf{X}_n$ are independent random variables with mean zero and variance $\sigma^2 = \mathbf{t}'\Sigma\mathbf{t}$. By Theorem 11.2, we have $S_n/\sqrt{n} \xrightarrow{\mathcal{D}} \sigma Z$.

If $\mathbf{Y} = (Y_1, \dots, Y_k)$ has multivariate normal distribution with covariance Σ, then $\mathbf{t}'\mathbf{Y}$ is univariate normal with variance σ^2. So we showed that $\mathbf{t}'\mathbf{S}_n/\sqrt{n} \xrightarrow{\mathcal{D}} \mathbf{t}'\mathbf{Y}$ for all $\mathbf{t} \in \mathbb{R}^k$. This ends the proof by Theorem 12.7.

□

Example 12.3. Suppose ξ_k, η_k are i.i.d with mean zero variance one. Then $(\sum_{k=1}^n \eta_k, \sum_{k=1}^n (\eta_k + \xi_k) \xrightarrow{\mathcal{D}} (Z_1, Z_1 + Z_2)$.

Indeed, random vectors $\mathbf{X}_k = \begin{bmatrix} \xi_k \\ \xi_k + \eta_k \end{bmatrix}$ has covariance matrix $\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} Z_1 \\ Z_1 + Z_2 \end{bmatrix}$ has the same covariance matrix.

4.1. Application: Chi-Squared test for multinomial distribution. A multinomial experiment has k outcomes with probabilities $p_1, \dots, p_k$. A multinomial random variable $\mathbf{S}_n = (S_1(n), \dots, S_k(n))$ lists observed counts per category in n repeats of the multinomial experiment.

The following result is behind the use of the chi-squared statistics in tests of consistency.

Theorem 12.17. $\sum_{j=1}^k \frac{(S_j(n) - np_j)^2}{np_j} \xrightarrow{\mathcal{D}} Z_1^2 + \cdots + Z_{k-1}^2$

Proof. Lets prove this for $k = 3$. Consider independent random vectors $\mathbf{X}_k$ that take three values $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ with probabilities p_1, p_2, p_3. Then $\mathbf{S}_n$ is the sum of n independent identically distributed vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$.

Clearly, $E\mathbf{X}_k = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$. To compute the covariance matrix, write $\mathbf{X}$ for $\mathbf{X}_k$. For non-centered vectors, the covariance is $E(\mathbf{X}\mathbf{X}') - E(\mathbf{X})E(\mathbf{X}')$. We have

$$E(\mathbf{X}\mathbf{X}') = p_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} + p_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} + p_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{bmatrix}$$

So

$$\Sigma = E(\mathbf{X}\mathbf{X}') - E(\mathbf{X})E(\mathbf{X}') = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 \\ -p_1p_2 & p_2(1-p_2) & -p_2p_3 \\ -p_1p_3 & -p_2p_3 & p_3(1-p_3) \end{bmatrix}$$

Then $\mathbf{S}_n$ is the sum of n independent vectors, and the central limit theorem implies that $\frac{1}{\sqrt{n}}\left(\mathbf{S}_n - n\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}\right) \xrightarrow{\mathcal{D}} \mathbf{X}$.

In particular, by Proposition 12.1 we have

$$\sum_{j=1}^{3} \frac{(S_j(n) - np_j)^2}{np_j} \xrightarrow{\mathcal{D}} \sum_{j=1}^{3} \frac{X_j^2}{p_j}$$

where (X_1, X_2, X_3) is multivariate normal with covariance matrix Σ.

Note that since $\sum_{j=1}^{k} S_j(n) = n$, the gaussian distribution is degenerate: $X_1 + X_2 + X_3 = 0$.

It remains to show that $\sum_{j=1}^{3} \frac{X_j^2}{p_j}$ has the same law as $Z_1^2 + Z_2^2$ i.e. that it is exponential. To do so, we first note that the covariance of $(Y_1, Y_2, Y_3,) := (X_1/\sqrt{p_1}, X_2/\sqrt{p_2}, X_3/\sqrt{p_3})$ is

$$\Sigma_{\mathbf{Y}} = \begin{bmatrix} 1-p_1 & -\sqrt{p_1p_2} & -\sqrt{p_1p_3} \\ -\sqrt{p_1p_2} & 1-p_2 & -\sqrt{p_2p_3} \\ -\sqrt{p_1p_3} & -\sqrt{p_2p_3} & 1-p_3 \end{bmatrix} = I - \begin{bmatrix} \sqrt{p_1} \\ \sqrt{p_2} \\ \sqrt{p_3} \end{bmatrix} \times \begin{bmatrix} \sqrt{p_1} & \sqrt{p_2} & \sqrt{p_3} \end{bmatrix}$$

Since $\mathbf{v}_1 = \begin{bmatrix} \sqrt{p_1} \\ \sqrt{p_2} \\ \sqrt{p_3} \end{bmatrix}$ is a unit vector, we can complete it with two additional vectors $\mathbf{v}_2 = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$ and $\mathbf{v}_3 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$ to form an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ of $\mathbb{R}^3$. This can be done in many ways, for example by the Gram-Schmidt orthogonalization to $\mathbf{v}_1, [100]', [010]'$. However, the specific form of $\mathbf{v}_2, \mathbf{v}_3$ does not enter the calculation - we only need to know that $\mathbf{v}_2, \mathbf{v}_3$ are orthonormal.

The point is that $I = \mathbf{v}_1\mathbf{v}_1' + \mathbf{v}_2\mathbf{v}_2' + \mathbf{v}_3\mathbf{v}_3'$ as these are orthogonal eigenvectors of I with $\lambda = 1$. (Or, because $\mathbf{x} = \mathbf{v}_1\mathbf{v}_1'\mathbf{x} + \mathbf{v}_2\mathbf{v}_2'\mathbf{v}_2 + \mathbf{v}_3\mathbf{v}_3'\mathbf{x}$ as $\mathbf{v}_j'\mathbf{x} = \mathbf{x} \cdot \mathbf{v}_j$ are the coefficients of expansion of $\mathbf{x}$ in orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ of $\mathbb{R}^3$.)

Therefore,

$$\Sigma_{\mathbf{Y}} = \mathbf{v}_2\mathbf{v}_2' + \mathbf{v}_3\mathbf{v}_3'$$

We now notice that $\Sigma_{\mathbf{Y}}$ is the covariance of the multivariate normal random variable $\mathbf{Z} = \mathbf{v}_2 Z_2 + \mathbf{v}_3 Z_3$ where Z_2, Z_3 are independent real-valued $N(0,1)$. Indeed,

$$E\mathbf{Z}\mathbf{Z}' = \sum_{i,j=2}^{3} \mathbf{v}_i\mathbf{v}_j' E(Z_iZ_j) = \sum_{i=2}^{3} \mathbf{v}_i\mathbf{v}_i'$$

Therefore, vector $[Y_1Y_2Y_3]'$ has the same distribution as $\mathbf{Z}$, and $Y_1^2 + Y_2^2 + Y_3^2$ has the same distribution as

$$\|\mathbf{Z}\|^2 = \|\mathbf{v}_2 Z_2 + \mathbf{v}_3 Z_3\|^2 = \|\mathbf{v}_2 Z_2\|^2 + \|\mathbf{v}_3 Z_3\|^2 = Z_2^2 + Z_3^2$$

(recall that $\mathbf{v}_2$ and $\mathbf{v}_3$ are orthogonal unit vectors).

□

Remark 12.18. It is clear that this proof generalizes to all k.

We note that the distribution of $Z_1^2 + \cdots + Z_{k-1}^2$ is Gamma with parameters $\alpha = (k-1)/2$ and $\beta = 2$, which is known under the name of chi-squared distribution with $k-1$ degrees of freedom. To see that $Z_2^2 + Z_3^2$ is indeed chi-squared with two-degrees of freedom (i.e., exponential), we can determine the cumulative distribution function by computing $1 - F(u)$:

$$P(Z_2^2 + Z_3^2 > u) = \frac{1}{2\pi} \int_{x^2+y^2>u} e^{-(x^2+y^2)/2} dxdy = \frac{1}{2\pi} \int_0^{2\pi} \int_{r>\sqrt{u}} e^{-r^2/2} r dr d\theta = e^{-u/2}$$

Required Exercises

Exercise 12.1. Suppose X, Y are independent univariate normal random variables. Use Definition 12.3 to verify that each of the following is bivariate normal: $\mathbf{X} = (X, X)$, $\mathbf{X} = (X, Y)$, $\mathbf{X} = (X + \varepsilon Y, X - \varepsilon Y)$.

Exercise 12.2. Suppose that $\mathbb{R}^{2k}$-valued random variables $(\mathbf{X}_n, \mathbf{Y}_n)$ are such that $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}$ and $\mathbf{Y}_n \xrightarrow{P} 0$ (that is, $\lim_{n\to\infty} P(\|\mathbf{Y}_n\| > \varepsilon) = 0$ for all $\varepsilon > 0$).

Prove that $\mathbf{X}_n + \mathbf{Y}_n \xrightarrow{\mathcal{D}} \mathbf{X}$

Exercise 12.3. Suppose (X_n, Y_n) are pairs of independent random variables and $X_n \xrightarrow{\mathcal{D}} X$, $Y_n \xrightarrow{\mathcal{D}} Y$. Show that $(X_n, Y_n) \xrightarrow{\mathcal{D}} \mu$ where μ is the product of the laws of X and Y.

Exercise 12.4. Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables such that $E(\xi_1) = 0$, $E(\xi_1^2) = 1$. For $i = 1, 2, \ldots$, define $\mathbb{R}^2$-valued random variables $X_i = \begin{bmatrix} \xi_i \\ \xi_{i+1} \end{bmatrix}$ and let $S_n = \sum_{i=1}^n X_i$. Show that

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{\mathcal{D}} N(0, \Sigma)$$

for a suitable 2×2 covariance matrix Σ.

Exercise 12.5. Suppose ξ_j, η_j, γ_j are i.i.d. mean zero variance 1. Construct the following vectors:

$$\mathbf{X}_j = \begin{bmatrix} \xi_j - \eta_j \\ \eta_j - \gamma_j \\ \gamma_j - \xi_j \end{bmatrix}$$

Let $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$. Show that $\frac{1}{n}\|\mathbf{S}_n\|^2 \xrightarrow{\mathcal{D}} Y$. (In fact, Y has gamma density.)

Exercise 12.6. Use the characteristic function to verify that Remark 12.15 indeed gives two representations of the same normal law.

Bibliography

[Billingsley] P. Billingsley, Probability and Measure IIIrd edition

[Durrett] R. Durrett, Probability: Theory and Examples, Edition 4.1 (online)

[Gut] A. Gut, Probability: a graduate course

[Resnik] S. Resnik, A Probability Path, Birkhause 1998

[Proschan-Shaw] S M. Proschan and P. Shaw, Essential of Probability Theory for Statistitcians, CRC Press 2016

[Varadhan] S.R.S. Varadhan, Probability Theory, (online pdf from 2000)

Index