Sums of Random Variables via Brute Force

Note Title 18/10/2007

Let $X_1, X_2, \dots$ be independent random variables identically distributed with PDF

$$p(t) = \begin{cases} e^{-s} & , \quad s \geq 0 \\ 0 & , \quad s < 0 ; \end{cases}$$

we say that the X_i are EXPONENTIALLY DISTRIBUTED.

The random vector $\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^n$ has

PDF $$q(\underline{s}) = \begin{cases} e^{-(s_1 + s_2 + \dots + s_n)} & , \text{ for } s_1, \dots, s_n \geq 0, \\ 0 & , \text{ otherwise.} \end{cases}$$

Thus

$$\mathbb{P}(a \leq X_1 + \dots + X_n \leq b)$$

$$= \int_{a \leq s_1 + \dots + s_n \leq b} q(\underline{s}) \, d\underline{s}$$

$$= \int_{\substack{a \leq s_1 + \dots + s_n \leq b \\ s_1, \dots, s_n \geq 0}} e^{-(s_1 + \dots + s_n)} \, d\underline{s}$$

$$= \int_a^b du \; e^{-u} \, \mathrm{vol}_{n-1}(T_u),$$

where $T_u = \{ \underline{s} \in \mathbb{R}^n : s_1 + \dots + s_n = u \text{ and } s_1, \dots, s_n \geq 0 \}$.

Thus $\mathrm{vol}_{n-1}(T_u) = u^{n-1} \mathrm{vol}_{n-1} T_1$

and

$\mathbb{P}(a \leq X_1 + \cdots + X_n \leq b)$

$$= (\mathrm{vol}_{n-1} T_1) \int_a^b du \quad e^{-u} u^{n-1} \quad du .$$

Setting $a = 0$, $b = \infty$, we obtain

$$1 = (\mathrm{vol}_{n-1} T_1) \int_0^\infty e^{-u} u^{n-1} du,$$

or

$$\mathrm{vol}_{n-1} T = \frac{1}{(n-1)!} .$$

Hence

$\mathbb{P}(a \leq X_1 + \cdots + X_n \leq b)$

$$= \int_a^b \frac{e^{-u} u^{n-1}}{(n-1)!} du .$$

Sums of Gaussians

Now suppose we have a pair of independent $N(0,1)$ random variables Z_1 and Z_2. The Gaussian random vector $\underline{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \in \mathbb{R}^2$ has PDF

$$p(\underline{s}) = (2\pi)^{-1} e^{-\|\underline{s}\|^2/2}, \quad \text{for } \underline{s} \in \mathbb{R}^2.$$

Thus

$$\mathbb{P}(a \le Z_1 + Z_2 \le b)$$

$$= \iint_{a \le s_1 + s_2 \le b} (2\pi)^{-1} e^{-(s_1^2 + s_2^2)/2} \, ds_1 \, ds_2$$

We shall now change coordinates, using

$$u_1 = s_1 + s_2,$$

$$u_2 = s_1 - s_2,$$

or

$$\underline{u} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \underline{s} .$$

Then

$$d\underline{u} = \left| \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right| d\underline{s} = 2 \, d\underline{s}$$

and

$$s_1 = \tfrac{1}{2}(u_1 + u_2), \qquad s_2 = \tfrac{1}{2}(u_1 - u_2).$$

Therefore

$$s_1^2 + s_2^2 = \frac{1}{4}(u_1 + u_2)^2 + \frac{1}{4}(u_1 - u_2)^2$$

$$= \tfrac{1}{2}(u_1^2 + u_2^2),$$

which implies

$$\mathbb{P}(a \le Z_1 + Z_2 \le b) = \frac{1}{2} \int_a^b du_1 \int_{-\infty}^{\infty} du_2 \, (2\pi)^{-1} e^{-(u_1^2 + u_2^2)/4}$$

$$= \int_a^b du_1 \, (2\pi \cdot 2)^{-1/2} e^{-u_1^2/2 \cdot 2},$$

i.e. $Z_1 + Z_2 \sim N(0, 2)$.

More generally, let's prove that, if Z_1, Z_2 are independent $N(0,1)$ random variables, then

$w_1 Z_1 + w_2 Z_2 \sim N(0, w_1^2 + w_2^2)$, for any $w_1, w_2 \in \mathbb{R}$.

To calculate this PDF, we proceed as before:

$$p := \mathbb{P}(a \le w_1 Z_1 + w_2 Z_2 \le b)$$

$$= \iint_{a \le w_1 s_1 + w_2 s_2 \le b} (2\pi)^{-1} e^{-(s_1^2 + s_2^2)/2} \, ds_1 \, ds_2 .$$

The problem here is to rotate our coordinate system so that the integral becomes simple. If we let

$$\underline{u}_1 = \frac{1}{\sqrt{w_1^2 + w_2^2}} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \quad \text{and} \quad \underline{u}_2 = \frac{1}{\sqrt{w_1^2 + w_2^2}} \begin{pmatrix} -w_2 \\ w_1 \end{pmatrix},$$

then these are orthogonal unit vectors. Our new coordinates t_1, t_2 are defined by

$$\underline{s} = t_1 \underline{u}_1 + t_2 \underline{u}_2$$

or

$$\underline{s} = (\underline{u}_1 \ \ \underline{u}_2) \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}.$$

Then $\begin{pmatrix} \underline{u}_1^T \\ \underline{u}_2^T \end{pmatrix} (\underline{u}_1 \ \ \underline{u}_2) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, so

$$\underline{t} = \begin{pmatrix} \underline{u}_1^T \\ \underline{u}_2^T \end{pmatrix} \underline{s}$$

and $\det(\underline{u}_1 \ \ \underline{u}_2) = 1$.

Finally

$$\{ \underline{s} \in \mathbb{R}^2 : \quad a \le w_1 s_1 + w_2 s_2 \le b$$

$$= \{ \underline{s} \in \mathbb{R}^2 : \frac{a}{\|\underline{w}\|} \le \frac{w_1 s_1 + w_2 s_2}{\|\underline{w}\|} \le \frac{b}{\|\underline{w}\|} \}$$

$$= \{ \underline{t} \in \mathbb{R}^2 : \frac{a}{\|\underline{w}\|} \le t_1 \le \frac{b}{\|\underline{w}\|} \}.$$

Further, $s_1^2 + s_2^2 = t_1^2 + t_2^2$, so that

$$P = \int_{a/\|\underline{w}\|}^{b/\|\underline{w}\|} dt_1 \int_{-\infty}^{\infty} dt_2 \, (2\pi)^{-1} e^{-(t_1^2 + t_2^2)/2}$$

$$= \int_{a/\|\underline{w}\|}^{b/\|\underline{w}\|} dt_1 \, (2\pi)^{-1/2} e^{-t_1^2/2} \overbrace{\left\{ \int_{-\infty}^{\infty} dt_2 \, (2\pi)^{-1/2} e^{-t_2^2/2} \right\}}^{= 1}$$

$$= \int_a^b dv_1 \, (2\pi \|\underline{w}\|^2)^{-1/2} e^{-v_1^2/(2\|\underline{w}\|^2)}.$$

$(v_1 = t_1 \|\underline{w}\|)$

Hence $w_1 Z_1 + w_2 Z_2 \sim N(0, \|\underline{w}\|^2)$.

Corollary If Z_1, Z_2 are independent $N(0,1)$,

then $V_1 = \cos\theta\, Z_1 - \sin\theta\, Z_2$,

$V_2 = \sin\theta\, Z_1 + \cos\theta\, Z_2$

are also $N(0,1)$; furthermore, $\mathbb{E}(V_1 V_2) = 0$.

Proof: V_1, V_2 are $N(0,1)$ because $\cos^2\theta + \sin^2\theta = 1$.

$$\mathbb{E}(V_1 V_2) = \mathbb{E}\left(\cos\theta\, Z_1 - \sin\theta\, Z_2\right)\left(\sin\theta\, Z_1 + \cos\theta\, Z_2\right)$$

$$= \cos\theta\sin\theta - \sin\theta\cos\theta = 0. \qquad \square$$

In fact, if $\underline{Z} \in \mathbb{R}^n$ is a normalized Gaussian random vector and if $\underline{a}, \underline{b} \in \mathbb{R}^n$, then

$$\mathbb{E}\left[(\underline{a}^T \underline{Z})(\underline{b}^T \underline{Z})\right]$$

$$= \sum_{k=1}^{n}\sum_{l=1}^{n} a_k b_l \underbrace{\mathbb{E}(Z_k Z_l)}_{=\ \delta_{kl}\ =\ \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}}$$

$$= \sum_{k=1}^{n} a_k b_k$$

$$= \underline{a}^T \underline{b}.$$

In fact, as we shall see later, Z_1, Z_2 are independent.

General Sums of Random Variables

Let X_1, X_2 be independent continuous random variables with PDFs f_1, f_2. Then the PDF of $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ is $p(\underline{s}) = f_1(s_1) f_2(s_2)$ and

$$\mathbb{P}(a \le X_1 + X_2 \le b) = \iint_{a \le s_1 + s_2 \le b} f_1(s_1) f_2(s_2)\, ds_1 ds_2$$

Using the same linear change of variable as before, i.e. $\begin{cases} u_1 = s_1 + s_2 \\ u_2 = s_1 - s_2 \end{cases}$, or $\underline{u} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \underline{s}$,

we get $d\underline{u} = \left| \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right| d\underline{s} = 2\, d\underline{s}$ and

$$\mathbb{P}(a \le X_1 + X_2 \le b) = \frac{1}{2} \int_a^b du_1 \int_{-\infty}^{\infty} du_2\, f_1\left(\frac{u_1+u_2}{2}\right) f_2\left(\frac{u_1-u_2}{2}\right)$$

Thus the PDF for $X_1 + X_2$ is given by

$$p(u_1) = \frac{1}{2} \int_{-\infty}^{\infty} f_1\left(\frac{u_1+u_2}{2}\right) f_2\left(\frac{u_1-u_2}{2}\right) du_2 .$$

If we now let $y = \frac{u_1+u_2}{2}$, then $\frac{u_1-u_2}{2} = u_1 - y$, $dy = \frac{1}{2} du_2$, and the integral becomes

$$p(u_1) = \int_{-\infty}^{\infty} f_1(y) f_2(u_1 - y)\, dy .$$

This is called the CONVOLUTION of f_1 and f_2, written $f_1 * f_2(u_1)$.

If we let $z = u_1 - y$, then the integral becomes
$$f_1 * f_2 (u_1) = \int_{-\infty}^{\infty} f_1(u_1 - z) f_2(z) \, dz.$$
In other words,
$$f_1 * f_2 (u_1) = f_2 * f_1 (u_1).$$
We can also deduce $f_1 * f_2 = f_2 * f_1$ by observing that
$$f_1 * f_2 = \text{PDF of } X_1 + X_2 = \text{PDF of } X_2 + X_1 = f_2 * f_1.$$

We now summarize these important results:

<u>Theorem</u> If X_1, X_2 are independent random variables with PDFs f_1, f_2, then their sum $X_1 + X_2$ has the convolution of f_1, f_2 as its PDF, i.e.
$$f_1 * f_2 (s) = \int_{-\infty}^{\infty} f_1(s-t) f_2(t) \, dt,$$
for $s \in \mathbb{R}$.

<u>Corollary</u> If $X_1, X_2, \dots, X_n$ are independent r.v.s with PDFs $f_1, f_2, \dots, f_n$, then $X_1 + X_2 + \cdots + X_n$ has PDF $f_1 * f_2 * \cdots * f_n$.

[Here $f_1 * \cdots * f_n = (f_1 * \cdots * f_{n-1}) * f_n$.

EXAMPLE Suppose $X_1, X_2, \ldots$ are independent identically distributed random variables with PDF

$$p_1(s) = \begin{cases} e^{-s}, & s \geq 0 \\ 0, & s < 0. \end{cases} \qquad (9.1)$$

Let's calculate the PDF $p_n(s)$ for $X_1 + X_2 + \cdots + X_n$.

Now

$$p_2(s) = p_1 * p_1(s)$$

$$= \int_{-\infty}^{\infty} \underbrace{p_1(z)}_{\substack{= \\ 0 \\ \text{for } z<0}} \underbrace{p_2(s-z)}_{\substack{= \\ 0 \\ \text{for } s-z<0, \\ \text{i.e. } z>s}} dz$$

.

Thus

$$p_2(s) = \int_0^s p_1(z)\, p_2(s-z)\, dz$$

$$= \int_0^s e^{-z}\, e^{-(s-z)}\, dz$$

$$= e^{-s} \int_0^s dz,$$

or $$p_2(s) = s e^{-s}. \qquad (9.2)$$

Furthermore,

$$p_n(s) = p_1 * p_{n-1}(s)$$

$$= \int_{-\infty}^{\infty} p_1(z)\, p_{n-1}(s-z)\, dz,$$

or $$p_n(z) = \int_0^s p_1(z)\, p_{n-1}(s-z)\, dz \qquad (9.3)$$

EXERCISE: Use (9.1), (9.2) and (9.3) to show that

$$p_n(s) = \begin{cases} \dfrac{s^{n-1}}{(n-1)!} e^{-s}, & \text{for } s \geq 0 \\ 0, & \text{for } s < 0, \end{cases}$$

for $n = 1, 2, \ldots$.

It's extremely tedious computing convolutions via brute force, as we've being doing, although it's an excellent way to learn probability theory. Fortunately, there's an extremely clever way due to Fourier.

DEFINITION Let X be a continuous random variable with PDF $p(s)$. Then the CHARACTERISTIC FUNCTION, or CF, of X is defined by

$$\phi_X(z) = \mathbb{E}\left(e^{izX}\right), \quad \text{for } z \in \mathbb{R}.$$

<u>EXAMPLE</u> Suppose $X \sim U[-\frac{1}{2}, \frac{1}{2}]$, i.e. X is uniformly distributed on $[-1/2, 1/2]$. Its PDF is given by

$$p(s) = \begin{cases} 1, & -1/2 \leq s \leq 1/2 \\ 0, & |s| > 1/2. \end{cases}$$

Then

$$\begin{aligned} \phi_X(z) &= \mathbb{E}(e^{izX}) \\ &= \int_{-1/2}^{1/2} e^{izs} \, ds \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \underbrace{\cos(zs)}_{\text{EVEN}} + i \underbrace{\sin(zs)}_{\text{ODD}} \, ds \\ &= 2 \int_0^{1/2} \cos(zs) \, ds \end{aligned}$$

$$\phi_X(z) = \frac{\sin(z/2)}{(z/2)}. \qquad \square$$

[If you've met Fourier transforms before, then you have already recognized ϕ_X as the Fourier transform of the PDF $p(s)$.]

You may be wondering why CFs have been introduced. Their importance rests on the following result:

<u>Theorem</u> Let X_1, X_2 be independent continuous random variables with CFs $\phi_{X_1}(z)$, $\phi_{X_2}(z)$. Then

$$\phi_{X_1+X_2}(z) = \phi_{X_1}(z)\,\phi_2(z).$$

<u>Proof:</u>

$$\begin{aligned}
\phi_{X_1+X_2}(z) &= \mathbb{E}\, e^{iz(X_1+X_2)} \\
&= \mathbb{E}\, e^{izX_1} e^{iz X_2} \\
&= \left(\mathbb{E}\, e^{izX_1}\right)\left(\mathbb{E}\, e^{izX_2}\right) \\
&\qquad \text{(by independence of } X_1, X_2\text{)} \\
&= \phi_{X_1}(z)\,\phi_{X_2}(z).
\end{aligned}$$

$\square$

<u>EXAMPLE</u> If $X_1, X_2, \ldots$ are independent exponentially distributed random variables with PDF

$$p(s) = \begin{cases} e^{-s}, & s \geq 0 \\ 0, & s < 0. \end{cases}$$

Then the CF for each X_i is the function

$$\phi(z) = \mathbb{E}\, e^{izX}$$

$$= \int_0^\infty e^{izs}\, e^{-s}\, ds$$

$$= \int_0^\infty e^{s(-1+iz)}\, ds$$

$$= \left[\frac{e^{s(-1+iz)}}{-1+iz} \right]_{s=0}^{\infty}$$

$$= \frac{-1}{-1+iz}$$

$$\phi(z) = \frac{1}{1-iz}.$$

Their sum $S_n = X_1 + \cdots + X_n$ therefore has PDF

$$\phi_{S_n}(z) = (1-iz)^{-n}. \qquad \square$$

<u>Lemma</u> If $c \in \mathbb{R}$ and $Z \sim N(0,1)$, then

$$\mathbb{E}\, e^{cZ} = e^{c^2/2}.$$

<u>Proof:</u> $\mathbb{E}\, e^{cZ} = \int_{-\infty}^{\infty} e^{cs} (2\pi)^{-1/2} e^{-s^2/2}\, ds$

$$= (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(s^2 - 2cs)}\, ds$$

$$= (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(s-c)^2 - c^2]}\, ds$$

$$= e^{c^2/2} \int_{-\infty}^{\infty} (2\pi)^{-1/2} e^{-\frac{1}{2}(s-c)^2} \, ds$$

$$\underset{(t = s-c)}{=} e^{c^2/2} \int_{-\infty}^{\infty} (2\pi)^{-1/2} e^{-t^2/2} \, dt$$

$$= e^{c^2/2}. \qquad \square$$

<u>Theorem</u> The Lemma is true for all complex $c \in \mathbb{C}$.

<u>Corollary</u> If $W \sim N(0,1)$, then

$$\phi_W(z) = e^{-z^2/2}, \qquad \text{for } z \in \mathbb{R}.$$

<u>Proof:</u>

$$\phi_W(z) = \mathbb{E}\, e^{izW}$$

$$= e^{(iz)^2/2} \qquad \text{(by the Lemma)}$$

$$= e^{-z^2/2}. \qquad \square$$

<u>Corollary</u> If $V \sim N(\mu, \sigma^2)$, then

$$\phi_V(z) = e^{i\mu z - \sigma^2 z^2/2}.$$

<u>Proof:</u> We write $V = \mu + \sigma W$. Then $W \sim N(0,1)$ and

$$\mathbb{E}\, e^{iVz} = \mathbb{E}\, e^{i(\mu + \sigma W)z}$$

$$= e^{i\mu z}\, \mathbb{E}\, e^{i\sigma z W}$$

$$= e^{i\mu z}\, e^{-\sigma^2 z^2/2} \qquad \text{(by Lemma)}.$$

$\square$

<u>Corollary</u> If $Z_k \sim N(\mu_k, \sigma_k^2)$, $k = 1, 2, \ldots,$ are independent Gaussian random variables, then

$S_n = Z_1 + \cdots + Z_n \sim N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2)$.

<u>Proof:</u> Let $\phi_k(z)$ be the CF of Z_k; by the previous result

$$\phi_k(z) = e^{i\mu_k z} e^{-\sigma_k^2 z^2/2}.$$

Thus

$$\phi_{S_n}(z) = \phi_1(z)\phi_2(z)\cdots\phi_n(z)$$
$$= e^{i(\mu_1 + \cdots + \mu_n)} e^{-\frac{z^2}{2}(\sigma_1^2 + \cdots + \sigma_n^2)}.$$

$\square$

<u>Scaling</u> If X is a continuous random variable with PDF $p(s)$ and CF $\phi_X(z)$, what are the PDF/CF of $Y = cX$, where $c > 0$?

$$\mathbb{P}(a \le cX \le b) = \mathbb{P}\left(\frac{a}{c} \le X \le \frac{b}{c}\right)$$
$$= \int_{a/c}^{b/c} p(s)\, ds \underset{(t = cs)}{=} \int_a^b p\left(\frac{t}{c}\right) c^{-1}\, dt.$$

Thus $$p_{cX}(t) = \frac{p_X(t/c)}{c}.$$

Now

$$\phi_{cX}(z) = \mathbb{E}\, e^{icXz}$$

$$= \mathbb{E}\, e^{iczX}$$

or

$$\boxed{\phi_{cX}(z) = \phi_X(cz).}$$

<u>Corollary</u> Let $X_1, X_2, \dots, X_n$ be continuous i.i.d. random variables with common CF $\phi(z)$. Let $A_n = S_n/n$, where $S_n = X_1 + \dots + X_n$. Then

$$\phi_{A_n}(z) = \phi(z/n)^n.$$

<u>Proof:</u>

$$\phi_{A_n}(z) = \phi_{\frac{S_n}{n}}(z) = \phi_{S_n}\left(\frac{z}{n}\right)$$

$$= \phi\left(\frac{z}{n}\right)^n. \qquad \square$$

<u>Corollary</u> Let $\hat{A}_n = S_n/\sqrt{n}$. Then

$$\phi_{\hat{A}_n}(z) = \phi\left(\frac{z}{\sqrt{n}}\right)^n.$$

<u>Proof:</u> Exercise. $\square$

Now suppose $Z_1, Z_2, \ldots$ are independent $N(0,1)$. Their average

$$A_n = \frac{Z_1 + \cdots + Z_n}{n}$$

has, by the above corollaries, the CF

$$\phi_{A_n}(z) = \phi(z/n)^n,$$

where $\phi(z) = e^{-z^2/2}$. Thus

$$\phi_{A_n}(z) = \left(e^{-z^2/2n^2}\right)^n$$

$$= e^{-\frac{z^2}{2n}}.$$

Hence $A_n \sim N(0, \frac{1}{n})$.

In contrast, $\hat{A}_n = (Z_1 + \cdots + Z_n)/\sqrt{n}$ has

$$\phi_{\hat{A}_n}(z) = \phi\left(\frac{z}{\sqrt{n}}\right)^n$$

$$= \left(e^{-z^2/2n}\right)^n = e^{-z^2/2},$$

i.e. $\hat{A}_n \sim N(0,1)$, $\forall n$.

Averages of the Cauchy distribution

If X has PDF

$$p(s) = \frac{1}{\pi(1+s^2)}, \qquad s \in \mathbb{R},$$

then the CF can be shown to be

$$\phi(z) = e^{-|z|}, \qquad \text{for } z \in \mathbb{R}.$$

Corollary If $X_1, X_2, \ldots, X_n$ are independent Cauchy random variables, then so to is their average $A_n = \frac{X_1 + \cdots + X_n}{n}$.

Proof:

$$\phi_{A_n}(z) = \phi\left(\frac{z}{n}\right)^n = \left(e^{-|\frac{z}{n}|}\right)^n = e^{-|z|}.$$

$\square$

Thus averages of Cauchy distributions don't behave nicely: the CLT does not apply because its PDF does not have a well-defined average, in the sense that $\mathbb{E}|X| = \int_{-\infty}^{\infty} \frac{|s|}{\pi(1+s^2)} ds = \infty$.

The Bernstein–Azuma–Hoeffding Inequality

Let $X_1, X_2, \ldots$ be IID Bernoulli with $\mathbb{P}(X_k = \pm) = \frac{1}{2}$.

Then

$$\mathbb{P}\left(\left|\sum_{k=1}^{n} a_k X_k\right| \geq t\right) \leq 2e^{-t^2/2},$$

for any real sequence $a_1, a_2, \ldots, a_n$ satisfying $\sum_{k=1}^{n} a_k^2 = 1$. Proof: See later. $\square$

Application 1: Suppose $a_k = \frac{1}{\sqrt{n}}$, for $1 \leq k \leq n$.

Then $\sum_{k=1}^{n} a_k^2 = n\cdot\left(\frac{1}{\sqrt{n}}\right)^2 = 1$. So

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}(X_1+\dots+X_n)\right| \geq t\right) \leq 2e^{-t^2/2}.$$

Now choose $\epsilon > 0$ and let $t = \epsilon\sqrt{n}$. Then

$$\mathbb{P}\left(\left|\frac{X_1+\dots+X_n}{n}\right| \geq \epsilon\right) \leq 2e^{-\epsilon^2 n^2/2}. \qquad (1)$$

Compare this with Chebyshev's inequality, which states that

$$\mathbb{P}\left(\left|\frac{X_1+\dots+X_n}{n}\right| \geq \epsilon\right) \leq \frac{1}{n\epsilon^2}. \qquad (2)$$

If $\epsilon = 10^{-2}$ and $n = 10^4$, then $RHS(1) = 2e^{-5000}$, $RHS(2) = 1$.

We are now <u>very</u> close to the Central Limit Theorem. Suppose we have any sequence of identically distributed random variables: $X_1, X_2, \ldots$. Let their common CF be $\phi(z)$. Further, suppose $\mathbb{E}X_i = 0$ and $\mathbb{E}(X_i^2) = \sigma^2$, for all i. Define

$$\hat{A}_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}.$$

Then

$$\phi_{\hat{A}_n}(z) = \phi\left(\frac{z}{\sqrt{n}}\right)^n.$$

Now

$$\begin{aligned}\phi(w) &= \mathbb{E}\, e^{iwX} \\ &= \mathbb{E}\left(1 + iwX - \tfrac{1}{2}w^2X^2 + \cdots\right) \\ &= 1 + iw\,\mathbb{E}X - \tfrac{1}{2}w^2\,\mathbb{E}X^2 + \cdots \\ &= 1 - \tfrac{1}{2}\sigma^2 w^2 + \cdots, \quad \text{for small } w.\end{aligned}$$

Thus

$$\begin{aligned}\phi\left(\frac{z}{\sqrt{n}}\right) &= 1 - \tfrac{1}{2}\sigma^2 (z/\sqrt{n})^2 + \cdots \\ &= 1 - \frac{\sigma^2 z^2}{2n} + \cdots, \quad \text{for large } n;\end{aligned}$$

which implies

$$\phi_{\hat{A}_n}(z) = \phi\left(\frac{z}{\sqrt{n}}\right)^n = \left(1 - \frac{\sigma^2 z^2}{2n} + \cdots\right)^n \longrightarrow e^{-\sigma^2 z^2/2}.$$

We say $\hat{A}_n$ converges to $N(0, \sigma^2)$ in probability.

An unusual application

Let W be any continuous random variable with CF $\phi(z) = \mathbb{E}\left(e^{izW}\right)$, for $z \in \mathbb{R}$.

Suppose now that we pick distinct points $z_1, z_2, \ldots, z_n \in \mathbb{R}$ and form the $n \times n$ matrix A whose elements are given by

$$A_{jk} = \phi\,(z_j - z_k), \qquad 1 \leq j, k \leq n.$$

Then

$$\begin{aligned}
\underline{c}^T A\, \underline{c} &= \sum_{j=1}^{n} \sum_{k=1}^{n} c_j c_k\, \phi\,(z_j - z_k) \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} c_j c_k\, \mathbb{E}\, e^{i(z_j - z_k)W} \\
&= \mathbb{E}\left(\sum_{j=1}^{n} \sum_{k=1}^{n} c_j c_k\, e^{i(z_j - z_k)W} \right) \\
&= \mathbb{E}\left(\left| \sum_{k=1}^{n} c_k\, e^{iz_k W} \right|^2 \right) \\
&\geq 0.
\end{aligned}$$