Department of Mathematics, IST — Probability and Statistics Unit

Probability Theory

2nd. Test 1st. Semester — **2013/14**

Duration: **1h30m** **2014/01/10** — **8:00 AM**, Room V1.10

- Please justify your answers.
- This test has TWO PAGES and THREE GROUPS. The total of points is 20.0.

Group I — Independence, Poisson processes and expectation 3.0 points

Let $S(t)$ denote the price of a security[1] at time t. According to a popular model for the stochastic process $\{S(t),\, t \geq 0\}$, the price remains unchanged until a *shock* occurs, at which time the price is multiplied by a random factor. This model supposes that

$$S(t) = S(0) \prod_{i=1}^{N(t)} X_i,$$

where:

- $S(0)$ is the initial and constant price;
- $N(t)$ denotes the number of *shocks* up to time t;
- $\{N(t),\, t \geq 0\}$ is a Poisson process with rate λ;
- X_i represents the i^{th} multiplicative factor;
- X_i are independent exponential r.v. with rate μ;[2]
- $\{N(t),\, t \geq 0\}$ is independent of the X_i;
- $\prod_{i=1}^{N(t)} X_i = 1$ when $N(t) = 0$.

Find:

(a) $E[S(t)]$; [3] (1.5)

- **Random shocks**
  $X_i \stackrel{i.i.d.}{\sim} X,\, i \in \mathbb{N}$
  $X \sim \text{Exponential}(\mu)$
  $E(X) = \frac{1}{\mu}$
  $E(X^2) = V(X) + E^2(X) = \frac{1}{\mu^2} + \frac{1}{\mu^2} = \frac{2}{\mu^2}$
- **Number of random shocks**
  $N(t) =$ number of random shocks up to time t

[1]A certificate attesting credit, the ownership of stocks or bonds, or the right to ownership connected with tradable derivatives.

[2]Admit that $\mu \geq \sqrt{2}$.

[3]**Hint**: It might be of some use to know that the r.v. $Y \sim \text{Poisson}(\xi)$ has m.g.f. and p.g.f. given by $M_Y(t) = E(e^{tY}) = e^{\xi\,(e^t-1)}$ and $P_Y(z) = E(z^Y) = e^{\xi\,(z-1)}$ ($|z| \leq 1$), respectively.

  $\{N(t),\, t \geq 0\} \sim PP(\lambda)$ and independent of the X_i
  $N(t) \sim \text{Poisson}(\lambda\, t)$

- **Requested expectation**
  Due to the fact that $X_i \stackrel{i.i.d.}{\sim} X$ and $\{N(t),\, t \geq 0\}$ is independent of the X_i, we can successively write:

$$\begin{aligned}
E[S(t)] &= E\{E[S(t)|\, N(t)]\} \\
&= S(0) \times E\left\{ E\left[\prod_{i=1}^{N(t)} X_i \right| N(t) \right]\right\} \\
&= S(0) \times E\left[\prod_{i=1}^{N(t)} E(X_i) \right| N(t) \Bigg] \\
&= S(0) \times E\left\{ [E(X)]^{N(t)} \right\} \\
&= S(0) \times E\left[(1/\mu)^{N(t)} \right] \\
&= S(0) \times P_{N(t)}\,(1/\mu)\,,
\end{aligned}$$

where $P_{N(t)}(z) = E\left[z^{N(t)}\right] = e^{\lambda t\,(z-1)}$ ($|z| \leq 1$) represents the p.g.f. of $N(t)$. Consequently, if we admit that $1/\mu \leq 1$ then we get:

$$E[S(t)] = S(0) \times e^{\lambda t\,(1/\mu - 1)}.$$

(b) $E[S^2(t)]$ and $V[S(t)]$. (1.5)

- **Requested 2nd. moment**
  $E[S^2(t)]$ can be obtained in a similar fashion:

$$\begin{aligned}
E[S^2(t)] &= S^2(0) \times E\left\{ E\left[\prod_{i=1}^{N(t)} X_i^2 \right| N(t) \right]\right\} \\
&= S^2(0) \times E\left[\prod_{i=1}^{N(t)} E(X_i^2) \right| N(t) \Bigg] \\
&= S^2(0) \times E\left\{ \left[E(X^2)\right]^{N(t)} \right\} \\
&= S^2(0) \times E\left[\left(2/\mu^2\right)^{N(t)} \right] \\
&= S^2(0) \times P_{N(t)}\left(2/\mu^2\right) \\
&\stackrel{2/\mu^2 \leq 1}{=} S^2(0) \times e^{\lambda t\,(2/\mu^2 - 1)}.
\end{aligned}$$

- **Requested variance**

$$\begin{aligned}
V[S(t)] &= E[S^2(t)] - E^2[S(t)] \\
&= S^2(0) \times e^{\lambda t\,(2/\mu^2-1)} - \left[S(0) \times e^{\lambda t\,(1/\mu - 1)}\right]^2 \\
&= S^2(0) \times \left[e^{\lambda t\,(2/\mu^2-1)} - e^{\lambda t\,(2/\mu - 2)}\right].
\end{aligned}$$

Group II — Expectation **9.0 points**

1. Let X and Y be two r.v. in L^2 such that $E(X) = E(Y) = 0$, $V(X) = V(Y) = 1$ and $\rho = corr(X,Y) \in (0,1)$. (2.0)

   Determine all values of a and b for which $(X - a\,Y)$ and $(Y - b\,X)$ are uncorrelated, and plot them for $\rho = 0.5$.

   - **Random vector**

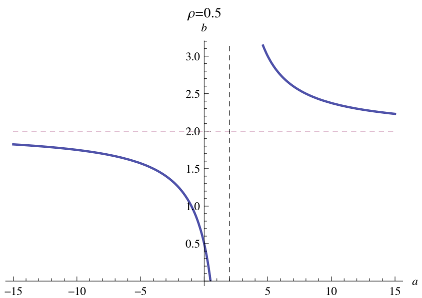
     $$(X,Y): \begin{cases} E(X) = E(Y) = 0 \\ V(X) = V(Y) = 1 \\ corr(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)\times V(Y)}} = cov(X,Y) = \rho \in (0,1) \end{cases}$$

   - **Determining a and b**

     Taking advantage of the fact that the covariance is a bilinear operator (Proposition 4.159), we get:

     $$\begin{aligned} (a,b) \quad : \quad & corr(X - a\,Y, Y - b\,X) = 0 \\ & cov(X - a\,Y, Y - b\,X) = 0 \\ & cov(X,Y) - b\,V(X) - a\,V(Y) + a\,b\,cov(X,Y) = 0 \\ & \rho - b - a + a\,b\,\rho = 0 \\ & b = \frac{a-\rho}{a\,\rho - 1} \quad (a \neq \rho^{-1}) \\ & a = \frac{b-\rho}{b\,\rho - 1} \quad (b \neq \rho^{-1}) \end{aligned}$$

     The requested values of a and b belong to $\{(a,b) \in \mathbb{R}^2 : \rho - b - a + a\,b\,\rho = 0$ and $((a,b) \neq (0,0)$ or $a \neq \rho^{-1}$ or $b \neq \rho^{-1})\}$.



2. Suppose that the total amount earned by a company in a year (X) is a r.v. with p.d.f. $\frac{x^m\,e^{-x}}{m!}$, $x > 0$, where $m \in \mathbb{N}_0$. Moreover, the utility of this amount is equal to $\sqrt{X}$.

   (a) Obtain an upper limit for the expected utility $E(\sqrt{X})$. (1.0)

   - **R.v.**

     X = total amount earned by a company in a year

   - **P.d.f.**

     $f_X(x) = \frac{x^m\,e^{-x}}{m!}$, $x > 0$

   - **Distribution**

     $X \sim \text{Gamma}(\alpha = m+1, \beta = 1)$, $m \in \mathbb{N}_0$

   - **Requested upper limit**

     Since $E(X) \overset{form.}{=} \frac{\alpha}{\beta} = m+1$ and $\sqrt{X}$ is a concave function, we can apply Jensen's inequality to obtain:

     $$\begin{aligned} E(\sqrt{X}) &\leq \sqrt{E(X)} \\ &= \sqrt{m+1}. \end{aligned}$$

   (b) Compare this upper limit with the true value of $E(\sqrt{X})$ when $m = 2$.[4] (1.5)

   - **Requested expected utility**

     $$\begin{aligned} E(\sqrt{X}) &= \int_0^{+\infty} x^{\frac{1}{2}} \times \frac{x^{(m+1)-1}\,e^{-x}}{\Gamma(m+1)}\,dx \\ &= \frac{\Gamma(m+\frac{3}{2})}{\Gamma(m+1)} \times \int_0^{+\infty} \frac{1}{\Gamma(m+\frac{3}{2})} x^{(m+\frac{3}{2})-1}\,e^{-x}\,dx \\ &= \frac{\Gamma(m+\frac{3}{2})}{\Gamma(m+1)} \times \int_0^{+\infty} f_{Gamma(m+\frac{3}{2},1)}(x)\,dx \\ &= \frac{\Gamma(m+\frac{3}{2})}{m!}. \end{aligned}$$

     Since $m = 2$, $\Gamma(0.5) = \sqrt{\pi}$ and $\Gamma(\alpha+1) = \alpha \times \Gamma(\alpha)$, $\alpha > 0$, we obtain:

     $$\begin{aligned} E(\sqrt{X}) &\overset{m=2}{=} \frac{\Gamma(3.5)}{2!} \\ &= \frac{2.5 \times 1.5 \times 0.5 \times \Gamma(0.5)}{2} \\ &= \frac{2.5 \times 1.5 \times 0.5 \times \sqrt{\pi}}{2} \\ &\simeq 1.661675. \end{aligned}$$

   - **Requested comparison**

     As for the upper limit, it is equal to $\sqrt{2+1} \simeq 1.732051$; the associated relative error is approximately equal to

     $$\frac{1.732051 - 1.661675}{1.732051} \times 100\% \simeq 4.063\%,$$

     a quite small relative error, suggesting that the Jensen's inequality provides a tight bound in this case.

3. A known personal trainer is promoting a weight reduction program. After a preliminary study about the efficiency of that program, she admitted that the initial weight X (in kg) and the weight Y (in kg) after a person has been submitted to the program constitute a random vector with a bivariate normal distribution with parameters $\mu_X = 64\,\text{kg}$, $\mu_Y = 59\,\text{kg}$, $\sigma^2_X = 9\,\text{kg}^2$, $\sigma^2_Y = 16\,\text{kg}^2$ and correlation coefficient ρ.

[4] Recall that $\Gamma(0.5) = \sqrt{\pi}$.

(a) Determine the value of ρ when we know that $E(Y|X > 64) \simeq 57.404231$. (2.5)

- **Random vector** (X, Y)
  X = initial weight
  Y = final weight
  $(X,Y) \sim \text{Normal}_2(\underline{\mu}, \mathbf{\Sigma})$, where

$$\underline{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 64 \\ 59 \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_X^2 & cov(X,Y) \\ cov(X,Y) & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho\,\sigma_X\,\sigma_Y \\ \rho\,\sigma_X\,\sigma_Y & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} 9 & \rho\sqrt{9\times 16} \\ \rho\sqrt{9\times 16} & 16 \end{bmatrix}$$

- **Determining** ρ
  Let:

$$Z_X = \frac{X-\mu_X}{\sigma_X}; \quad z_X = \frac{x-\mu_X}{\sigma_X}; \quad Z_Y = \frac{Y-\mu_Y}{\sigma_Y}.$$

First, recall that

$$Z_X \sim Z_Y \sim \text{Normal}(0,1).$$

Secondly, note that $E(Y|X > x)$ can be written in terms of the inverse Mills' ratio referring to a bivariate normal random vector with zero means, unit variances and correlation coefficient (see Definition 4.202):

$$\begin{aligned} E(Y|X>x) &= E\left(\mu_Y + \sigma_Y \times Z_Y | Z_X > z_X\right) \\ &= \mu_Y + \sigma_Y \times E\left(Z_Y | Z_X > z_X\right) \\ &= \mu_Y + \sigma_Y \times \frac{\rho \times \phi(z_X)}{\Phi(-z_X)}. \end{aligned}$$

Finally, we apply this last formula to find the correlation coefficient:

$$\rho \; : \; E(Y|X>64) \simeq 57.404231$$

$$59 + \sqrt{16} \times \frac{\rho \times \phi\left(\frac{64-64}{\sqrt{9}}\right)}{\Phi\left(-\frac{64-64}{\sqrt{9}}\right)} \simeq 57.404231$$

$$59 + 4 \times \frac{\rho \times \phi(0)}{\Phi(0)} \simeq 57.404231$$

$$59 + 4 \times \frac{\rho \times \frac{1}{\sqrt{2\pi}}}{0.5} \simeq 57.404231$$

$$\rho \simeq (57.404231 - 59) \times \sqrt{\frac{\pi}{32}}$$

$$\rho \simeq -0.5.$$

(b) What is the probability that the final weight after that program represents less than 85% of the initial weight?[5] (2.0)

[5]Consider $\rho = -0.5$ in case you have not solved the previous question.

- **Requested probability**
  $P(Y < 0.85\,X) = P(-0.85\,X + Y < 0)$
- **New r.v.**

$$W = -0.85\,X + Y = \mathbf{C} \times \begin{bmatrix} X \\ Y \end{bmatrix} + \underline{b}, \qquad \text{where } \mathbf{C} = [-0.85 \quad 1] \text{ and } \underline{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

- **Distribution of** W
  According to Theorem 4.190, $W = Y - 0.85\,X \sim \text{Normal}(E(W), V(W))$, where:

$$\begin{aligned} E(W) &= \mathbf{C}\,\underline{\mu} + \underline{b} \\ &= -0.85 \times \mu_X + \mu_Y \\ &= -0.85 \times 64 + 59 \\ &= 4.6; \\ V(W) &= \mathbf{C}\,\mathbf{\Sigma}\,\mathbf{C}^\top \\ &= (-0.85)^2 \times V(X) + 2 \times (-0.85) \times cov(X,Y) + V(Y) \\ &= 0.85^2 \times \sigma_X^2 - 2 \times 0.85 \times \rho\,\sigma_X\,\sigma_Y + \sigma_Y^2 \\ &= 0.85^2 \times 9 - 2 \times 0.85 \times (-0.5) \times \sqrt{9} \times \sqrt{16} + 16 \\ &= 32.7025. \end{aligned}$$

- **Requested probability** (cont'd)

$$\begin{aligned} P(W<0) &= \Phi\left[\frac{0-E(W)}{\sqrt{V(W)}}\right] \\ &= \Phi\left[\frac{0-4.6}{\sqrt{32.7025}}\right] \\ &\simeq \Phi(-0.80) \\ &= 1 - \Phi(0.80) \\ &\overset{tabela}{=} 1 - 0.7881 \\ &= 0.2119. \end{aligned}$$

Group III — Convergence of sequences of r.v. **8.0 points**

1. The Exponential distribution has been extensively used to model the duration of components which do not deteriorate (or improve their performance) with time. Let:
   - $\{X_1, X_2, \ldots\}$ be a sequence of i.i.d. durations with an Exponential(1) distribution;
   - $X_{(n)} = \max_{i=1,\ldots,n} X_i$ be the maximum duration of the n first components.
   - $Y_n = X_{(n)} - \ln(n)$ be the scaled maximum duration of the n first components.

(a) Determine $\lim_{n\to+\infty} F_{X_{(n)}}(x)$. (0.5)

- **R.v.**
  $X_i \overset{i.i.d.}{\sim} X,\ i \in I\!N$
  $X \sim \text{Exponential}(1)$
  $f_X(x) = \begin{cases} 0, & x < 0 \\ e^{-x}, & x \geq 0 \end{cases}$
  $F_X(x) = \begin{cases} 0, & x < 0 \\ \int_0^x e^{-t}\,dt = 1 - e^{-x}, & x \geq 0 \end{cases}$
- **Another r.v.**
  $X_{(n)} = \max_{i=1,\dots,n} X_i$
  $F_{X_{(n)}}(x) = [F_X(x)]^n$ (see Example 3.67)
- **Obtaining** $\lim_{n\to+\infty} F_{X_{(n)}}(x)$
  Note that

$$\begin{aligned}\lim_{n\to+\infty} F_{X_{(n)}}(x) &= \lim_{n\to+\infty} [F_X(x)]^n \\ &= \begin{cases} 0, & 0 \leq F_X(x) < 1 \\ 1, & F_X(x) = 1. \end{cases}\end{aligned}$$

  However, since in this case $F_X(x) \in [0,1),\ x \in \mathbb{R}$,

$$\lim_{n\to+\infty} F_{X_{(n)}}(x) = 0,\ x \in \mathbb{R}.$$

  [This is result suggests that $X_{(n)} = +\infty$, a degenerate r.v. in $\overline{\mathbb{R}}$.]

(b) Show that $\{Y_1, Y_2, \dots\}$ converges in distribution and determine the limiting c.d.f.[6] (2.0)

- **R.v.**
  $X_i \overset{i.i.d.}{\sim} X,\ i \in I\!N$
  $X \sim \text{Exponential}(1)$
  $F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x \geq 0 \end{cases}$
- **Another r.v.**
  $Y_n = X_{(n)} - \ln(n)$
- **C.d.f. of** Y_n

$$\begin{aligned} F_{Y_n}(y) &= P\left[X_{(n)} - \ln(n) \leq y\right] \\ &= \{F_X[y + \ln(n)]\}^n \\ &= \left[1 - e^{-(y+\ln(n))}\right]^n \\ &= \left(1 - \frac{e^{-y}}{n}\right)^n,\ y + \ln(n) > 0 \end{aligned}$$

[6] This result illustrates the fact that the Exponential distribution belongs to the domain of attraction of the Gumbel distribution, when it comes to the asymptotic behaviour of the maximum.

- **Checking the convergence in distribution**
  Since
  (i) $\lim_{n\to+\infty} F_{Y_n}(y) = \lim_{n\to+\infty}\left(1 - \frac{e^{-y}}{n}\right)^n = e^{-e^{-y}},\ y \in \mathbb{R}$,
  (ii) $F_Y(y) = e^{-e^{-y}},\ y \in \mathbb{R}$, is the c.d.f. of the absolutely continuous r.v. Y,[7]
  (iii) $\lim_{n\to+\infty} F_{Y_n}(y) = F_Y(y)$, for all the continuity points of the c.d.f. of Y,
  we can conclude that

$$Y_n \overset{d}{\to} Y.$$

2. Let $\{X_1, X_2, \dots\}$ be a sequence of i.i.d. r.v. to $X \in L^k$, for some positive integer k, and common k^{th} moment $E(X^k)$.

(a) Prove that $\frac{1}{n}\sum_{i=1}^n X_i^k \overset{P}{\to} E(X^k)$; (1.5)

- **Sequence of r.v.**
  $\{X_1, X_2, \dots\}$, where $X_i \overset{i.i.d.}{\sim} X$ and $X \in L^k$
- **Another sequence of r.v.**
  $\{Y_1, Y_2, \dots\}$, where $Y_i = X_i^k \overset{i.i.d.}{\sim} Y = X^k$ and $Y \in L^1$
- **To prove**
  $\frac{1}{n}\sum_{i=1}^n X_i^k \overset{P}{\to} E(X^k)$, i.e., $\frac{1}{n}\sum_{i=1}^n Y_i \overset{P}{\to} E(Y)$
- **Proof**
  The sequence of r.v. $\{Y_1, Y_2, \dots\}$ is under the conditions of Theorem 5.143 (WLLN for i.i.d. r.v. in L^1), therefore we conclude that

$$\frac{1}{n}\sum_{i=1}^n Y_i \overset{P}{\to} E(Y) = E(X^k).$$

(b) Discuss the pertinence of the previous result. (0.5)

- **Pertinence of the convergence result** $\frac{1}{n}\sum_{i=1}^n X_i^k \overset{P}{\to} E(X^k)$
  This means that the k^{th} sample moment, $\frac{1}{n}\sum_{i=1}^n X_i^k$, is a consistent estimator of the k^{th} moment of X, $E(X^k)$, if the X_i are i.i.d. r.v. in L^k [that is, if we are dealing with a random sample of size n, $(X_1, \dots, X_n)$, drawn from the population $X \in L^k$].

3. (a) State and prove the Lindeberg-Lévy Central Limit Theorem. (2.5)

- **Statement of the Lindeberg-Lévy CLT**
  Let:
  (i) $\{X_1, X_2, \dots\}$ be a sequence of i.i.d. r.v. such that $E(X_i) = \mu$ and $V(X_i) = \sigma^2 \in I\!R^+$, for $i \in \mathbb{N}$;
  (ii) $S_n = \sum_{i=1}^n X_i$ be the sum of the first n terms of that sequence of i.i.d. r.v.;

[7] After all this function is absolutely continuous, $\lim_{y\to-\infty} F_Y(y) = 0$ and $\lim_{y\to+\infty} F_Y(y) = 1$. Moreover, the footnote leads us to believe that Y has a Gumbel distribution (this is indeed the case!).

(iii) $\{Z_1, Z_2, \dots\}$ be the sequence of the standardized partial sums, where

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}.$$

Then

$$Z_n \xrightarrow{d} \text{Normal}(0,1).$$

- **Proof of the Lindeberg-Lévy CLT**

Let

$$X_i^\star = \frac{X_i - \mu}{\sigma}, \; i = 1, 2, \dots$$

then

$$\begin{aligned} X_i^\star &\overset{i.i.d.}{\sim} X^\star \\ E(X^\star) &= 0 \\ V(X^\star) &= 1 \\ E[(X^\star)^2] &= 1 \\ X^\star &\in L^2. \end{aligned}$$

Moreover, if we consider $S_n^\star = \sum_i^n X_i^\star$ and $a_n = \sqrt{n}$ then

$$Z_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = a_n^{-1} S_n^\star$$

and the characteristic function of Z_n is given by

$$\begin{aligned} \varphi_{Z_n}(t) &= E\left(e^{it\,Z_n}\right) \\ &= \varphi_{a_n^{-1} S_n^\star}(t) \\ &\overset{Prop. 5.161}{=} \left[\varphi_{X^\star}(a_n^{-1}t)\right]^n \\ &= \left[\varphi_{X^\star}(t/\sqrt{n})\right]^n. \end{aligned}$$

Furthermore, capitalizing on the fact that $X^\star \in L^k$, where $k = 2$, we can write

$$\begin{aligned} \varphi_{Z_n}(t) &\overset{Prop. 5.177}{=} \left[\sum_{j=0}^{k} \frac{(\frac{it}{\sqrt{n}})^j}{j!} E[(X^\star)^j] + o\left(\left|\frac{t}{\sqrt{n}}\right|^k\right)\right]^n \\ &\overset{k=2}{=} \left[1 + \frac{it}{\sqrt{n}} E(X^\star) + \frac{(\frac{it}{\sqrt{n}})^2}{2!} E[(X^\star)^2] + o(t^2/n)\right]^n \\ &\overset{E(X^\star)=0, E[(X^\star)^2]=1}{=} \left[1 - \frac{t^2}{2n} + o(1/n)\right]^n. \end{aligned}$$

Consequently,

$$\begin{aligned} \lim_{n\to+\infty} \varphi_{Z_n}(t) &= \lim_{n\to+\infty} \left[1 + \frac{\frac{-t^2}{2}}{n}\right]^n \\ &= e^{-\frac{t^2}{2}} \\ &= \varphi_{N(0,1)}(t), \; \forall t. \end{aligned}$$

Finally, by invoking the Theorem 5.181 (Continuity Theorem!), we can conclude that

$$Z_n \xrightarrow{d} \text{Normal}(0,1).$$

(b) Elaborate on two applications of this theorem. (1.0)

- **Two applications of the Lindeberg-Lévy Central Limit Theorem**

[This is probably the most notable case of convergence in distribution. Two of its applications:]

  – The Lindeberg-Lévy CLT allows us to add that, when we deal with a sufficiently large number n of i.i.d. r.v. $X_1, \dots, X_n$, with common mean μ and common positive and finite variance σ^2, the c.d.f. of the sum of these r.v. can be approximate as follows:

$$\begin{aligned} P(S_n \leq s) &= P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{s - n\mu}{\sqrt{n\sigma^2}}\right) \\ &\overset{CLT}{\simeq} \Phi\left(\frac{s - n\mu}{\sqrt{n\sigma^2}}\right). \end{aligned}$$

  This result is particularly important because, unlike the Binomial, Poisson and Normal distributions, most distributions are not closed under convolution and it is crucial to provide an approximate distribution for sums (or means) of r.v.

  – Under certain conditions (see Theorem 5.200) and the finiteness of the *Fisher information*,

$$I(\theta) = E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial\theta}\right)^2\right],$$

  we get the asymptotic normality of the standardised estimation error associated with the maximum likelihood estimator (MLE) of θ, $\hat{\theta}_n$:

$$\sqrt{n\,I(\theta)} \times (\hat{\theta}_n - \theta) \xrightarrow{d} \text{Normal}(0,1).$$

  Moreover, if we replace $I(\theta)$ with its MLE, $I(\hat{\theta}_n)$, then the asymptotic normality is still valid[8] and we end up with the pivotal quantity

$$\sqrt{n\,I(\hat{\theta}_n)} \times (\hat{\theta}_n - \theta) \xrightarrow{d} \text{Normal}(0,1),$$

  which is essential to derive asymptotic confidence intervals for θ.
  This result is extremely relevant because in most cases the exact distribution of the MLE of θ is unknown or very hard to derive.

[8] According to Theorem 5.200, $\hat{\theta}_n \xrightarrow{P} \theta$. If we add to this the fact that $I(\theta)$ is continuous (see conditions of Theorem 5.200) and the closure of the convergence in probability under continuous mappings, then we conclude that $\sqrt{I(\hat{\theta}_n)/I(\theta)} \xrightarrow{P} 1$. Finally, we have to invoke Theorem 5.95 (Slutsky's theorem or preservation of convergence in distribution under (restricted) product) to conclude that $\sqrt{n\,I(\hat{\theta}_n)} \times (\hat{\theta}_n - \theta) = \left[\sqrt{n\,I(\theta)} \times (\hat{\theta}_n - \theta)\right] \times \sqrt{I(\hat{\theta}_n)/I(\theta)} \xrightarrow{d} \text{Normal}(0,1)$.