Probabilitat i Estadística 1
Problemes Tema 4. Sumes de variables aleatòries. Normal bivariant. Llei dels Grans Nombres. Teorema del Límit Central.

(Professors: Pedro Delicado i Oriol Serra)

1 Mitjana i variància mostral

1. Siguin $X_1, \ldots, X_n$ v.a.i.i.d. amb esperança μ i variància σ^2. Es defineix la variància mostral

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Proveu que $\mathbb{E}(S^2) = \sigma^2$.

Solució:

$$\begin{aligned}
\mathbb{E}(S^2) =& \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) \\
=& \frac{n}{n-1}\mathbb{E}\left((X_1 - \bar{X})^2\right) \\
& (\mathbb{E}(X_1 - \bar{X}) = 0) \\
=& \frac{n}{n-1}\mathrm{Var}\left((X_1 - \bar{X})^2\right) \\
=& \frac{n}{n-1}\left(\mathrm{Var}(X_1) + \mathrm{Var}(\bar{X}) - 2\mathrm{Cov}(X_1, \bar{X})\right) \\
& \left(\mathrm{Cov}(X_1, \bar{X}) = \frac{1}{n}\mathrm{Cov}(X_1, \sum_{j=1}^{n} X_j) = \frac{1}{n}\mathrm{Cov}(X_1, X_1) = \frac{\sigma^2}{n}\right) \\
=& \frac{n}{n-1}\left(\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\right) \\
=& \frac{n}{n-1}\frac{(n-1)\sigma^2}{n} \\
=& \sigma^2.
\end{aligned}$$

2 Distribució normal bivariada.

2. Siguin (X_1, X_2) dues v.a. amb distribució conjunta normal bivariada i tals que $\mathrm{Var}(X_1) = \mathrm{Var}(X_2)$. Demostreu que les v.a $S = X_1 + X_2$ i $D = X_1 - X_2$ són independents.

Solució:
Com que sabem que conjuntament (S, D) tenen distribució normal bivariant, només cal veure que la seva covariància és 0:

$$\begin{aligned}
Cov(S, D) &= Cov(X_1 + X_2, X_1 - X_2) \\
&= V(X_1) - Cov(X_1, X_2) + Cov(X_2, X_1) - V(X_2) \\
&= 0.
\end{aligned}$$

3. Sigui $X \sim N(0, 1)$ i $a > 0$. Sigui

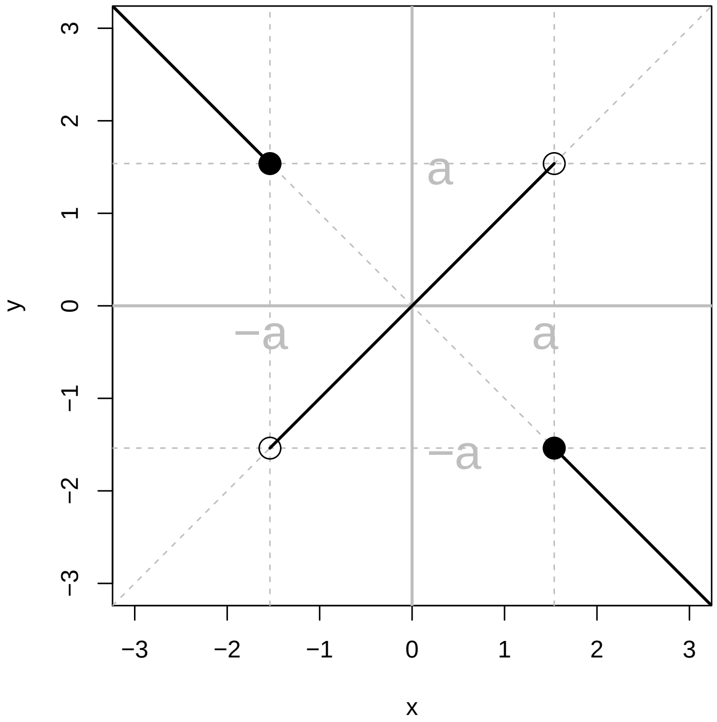
$$Y = \begin{cases} X & \text{si } |X| < a, \\ -X & \text{si } |X| \geq a. \end{cases}$$

(a) Proveu que $Y \sim N(0,1)$.
(b) La distribució conjunta de (X,Y), és normal bivariant? Són independents?
(c) Doneu l'expressió de $\rho(a) = \text{cor}(X,Y)$ en termes de la funció de densitat $\phi(x)$ de X.
(d) Deduïu-ne que existeix un valor a^* pel qual $\text{Corr}(X,Y) = 0$ tot i que X i Y no siguin independents.

Solució:
Per a $a > 0$. Sigui la funció

$$g(x) = \begin{cases} x & \text{si} \quad |x| < a, \\ -x & \text{si} \quad |x| \geq a. \end{cases}$$

Aleshores $Y = g(X)$.



Siguin $\phi(x)$ i $\Phi(x)$ les funcions de densitat i de distribució, respectivament, de una $N(0,1)$. Per simetria de la distribució $N(0,1)$ respecte al 0 sabem que

$$\phi(x) = \phi(-x), \;\; \Phi(x) = 1 - \Phi(-x) \text{ per a tot } x \in \mathbb{R}.$$

(a) Calculem la funció de distribució de Y. Sigui $y \leq -a$,

$$F_Y(y) = \Pr(Y \leq y) = \Pr(X \leq -y) = 1 - \Phi(-y) = \Phi(y).$$

Sigui ara $-a < y < a$,

$$F_Y(y) = \Pr(Y \leq y) = \Pr(X \in (-a, y)) + \Pr(X \geq a) =$$

$$\Phi(y) - \Phi(-a) + (1 - \Phi(a)) = \Phi(y) - \Phi(-a) + \Phi(-a) = \Phi(y).$$

Finalment, sigui $y \geq a$,

$$F_Y(y) = \Pr(Y \leq y) = \Pr(X \geq -y) = 1 - \Phi(-y) = \Phi(y).$$

Aleshores la funció de distribució de Y coincideix amb la d'una $N(0,1)$. Per tant $Y \sim N(0,1)$.

(b) La distribució conjunta de (X, Y) no és normal bivariant perquè posa tota la probabilitat en el conjunt
$$\{(x, y) \in \mathbb{R}_+^2 : y = g(x)\},$$
anomenat *graf* de la funció $g(x)$.
X i Y no són independents: si ho fossin (X, Y) seria normal bivariant.

(c) Calculem $\rho(a) = \mathrm{Corr}(X, Y)$. Observeu primer que $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X, Y)$ perquè $E(X) = E(Y) = 0$. Per altra banda $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$. Per tant
$$\begin{aligned}
\rho(a) &= \mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y) = E(XY) = E(Xg(X)) \\
&= \int_{\mathbb{R}} xg(x)\phi(x)dx \\
&= \int_{-\infty}^{-a} (-x^2)\phi(x)dx + \int_{-a}^{a} x^2\phi(x)dx + \int_{a}^{\infty} (-x^2)\phi(x)dx \\
&= \int_{-\infty}^{\infty} x^2\phi(x)dx - 4\int_{a}^{\infty} x^2\phi(x)dx \\
&= \mathrm{Var}(X) - 4\int_{a}^{\infty} x^2\phi(x)dx \\
&= 1 - 4\int_{a}^{\infty} x^2\phi(x)dx.
\end{aligned}$$
Aleshores
$$\rho(0) = 1 - 4\int_{0}^{\infty} x^2\phi(x)dx1 - 2\int_{-\infty}^{\infty} x^2\phi(x)dx = -1,$$
i
$$\lim_{a\to\infty} \rho(a) = 1.$$

(d) Com que $\rho(a)$ és funció contínua de a, existeix un valor $a^* \in [0, \infty)$ tal que $\rho(a^*) = 0$. Aquest valor és
$$a^* = 1.5382$$
Sigui Y_{a^*} la variable Y definida pel valor a^*. Tenim que la variable aleatòria bivariant (X, Y_{a^*}) verifica:
- Les seves marginals són $N(0, 1)$.
- No són independents.
- Són incorrelades.

Aquests fets no contradiuen el resultat que ens diu que sota normalitat conjunta hi ha independència si i només si hi ha incorrelació, perquè (X, Y_{a^*}) no tenen distribució conjunta normal bivariant.

4. Siguin (X_1, X_2) dues v.a. amb distribució conjunta normal bivariada amb mitjanes μ_1 i μ_2, variàncies σ_1^2 i σ_2^2 i correlació ρ. Determineu la distribució de $X_1 - 3X_2$.

Solució:
Definim $Y_1 = X_1 - 3X_2$. Sabem que Y_1 és normal univariant, per ser combinació lineal de (X_1, X_2), una normal bivariada.

Calculem l'esperança i la variància d'Y_1:

$$\begin{aligned} E(Y_1) &= E(X_1 - 3X_2) \\ &= E(X_1) - 3E(X_2) \\ &= \mu_1 - 3\mu_2, \end{aligned}$$

$$\begin{aligned} Var(Y_1) &= E((Y_1)^2) - (E(Y_1))^2 \\ &= E((X_1 - 3X_2)) - (E(X_1) - 3E(X_2))^2 \\ &= E(X_1^2) - 6E(X_1X_2) + 9E(X_2^2) \\ &\quad -E(X_1^2) + 6E(X_1)E(X_2) - 9E(X_2)^2 \\ &= Var(X_1) + 9Var(X_2) - 6Cov(X_1X_2) \\ &= \sigma_1^2 + 9\sigma_2^2 - 6\rho\sigma_1\sigma 2. \end{aligned}$$

Per tant $Y_1 = X_1 - 3X_2 \sim N(\mu_1 - 3\mu_2, \sigma_1^2 + 9\sigma_2^2 - 6\rho\sigma_1\sigma 2)$.

5. Sigui (X, Y) un vector aleatori que segueix una distribució normal bivariada de paràmetres $\mu_X = 40$, $\mu_Y = 20$, $\sigma_X^2 = 9$, $\sigma_Y^2 = 4$ i $\rho = 0.6$. Trobeu l'interval més curt ($[a, b]$) tal que la probabilitat condicionada de que Y pertanyi a l'interval ($[a, b]$) donat que $X = 22$ sigui igual a 0.90.

   Solució:
   Sabem que si
   $$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$$
   aleshores la distribució condicional de Y, donat $X = x$, és la distribució normal de mitjana $\mu_* = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ i de variància $\sigma_*^2 = \sigma_Y^2(1 - \rho^2)$. És a dir,
   $$Y|_{X=x} \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right).$$

   Aleshores $Y|_{X=22} \sim N(12.8, (1.6)^2)$. Hem de trobar un interval $[a, b]$ tal que $b - a$ sigui mínim de manera que $P[a \le U \le b] = 0.9$ per
   $U \sim N(12.8, 1.6^2)$. $Z = \frac{U - 12.8}{1.6} \sim N(0, 1)$.
   Per tant
   $$P[a \le U \le b] = P[\frac{a - 12.8}{1.6} \le Z \le \frac{b - 12.8}{1.6}] = 0.9$$
   Per tal que l'interval sigui mínim prenem un interval centrat en 12.8.
   $$12.8 - a = b - 12.8 \Rightarrow b + a = 25.6 \Rightarrow a = 25.6 - b$$
   $$P[\tfrac{12.8 - b}{1.6} \le Z \le \tfrac{b - 12.8}{1.6}] = 0.9$$
   $$\tfrac{b - 12.8}{1.6} = 1.645 \Rightarrow b = 1.645 \cdot 1.6 + 12.8$$
   $$b = 15.432 \Rightarrow a = 10.168.$$

6. Sigui $X \sim N(0, 1)$, i $(Y|X = x) \sim N(2x - 3, 12)$. Determineu la distribució marginal de Y i la correlació entre X i Y.

   Solució:
   Sabem que si
   $$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$$
   aleshores la distribució condicional de Y, donat $X = x$, és la distribució normal de mitjana $\mu_* = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ i de variància $\sigma_*^2 = \sigma_Y^2(1 - \rho^2)$. És a dir,
   $$Y|_{X=x} \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right).$$

En aquest cas tenim que $\mu_X = 0$, $\sigma_X^2 = 1$,

$$\mu^* = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X) = 2X - 3, \;\; \sigma_*^2 = \sigma_Y^2(1-\rho^2) = 12.$$

Aleshores,

$$\left.\begin{array}{rcl} \mu_Y + \rho\sigma_Y X & = & 2X - 3 \\ \sigma_Y^2(1-\rho^2) & = & 12 \end{array}\right\} \Rightarrow \quad \left.\begin{array}{rcl} \mu_Y & = & -3 \\ \rho\sigma_Y & = & 2 \\ \sigma_Y^2(1-\rho^2) & = & 12 \end{array}\right\}$$

Resolem aquest sistema d'eqüacions:

$$\left.\begin{array}{rcl} \rho & = & 2/\sigma_Y \\ \sigma_Y^2(1-\rho^2) & = & 12 \end{array}\right\} \Rightarrow \frac{2^2}{\rho^2}(1-\rho^2) = 12 \Rightarrow \frac{1}{\rho^2} - 1 = 3 \Rightarrow \rho = \frac{1}{2} \Rightarrow \sigma_Y = 4.$$

Aleshores $\rho(X,Y) = \rho = 0.5$ i $Y \sim N(\mu_Y = -3, \sigma_Y^2 = 4^2)$.

3 Convergència en distribució

7. Sigui $\{X_n\}_{n=1}^{\infty}$ una successió de variables aleatòries tals que la distribució de X_n ve donada per la funció de probabilitat següent:

$$P\left(X_n = \frac{k}{n}\right) = \frac{1}{n}, \text{ per a } k = 1, \ldots, n.$$

Determineu a quina distribució convergeix X_n en distribució i proveu que efectivament es té aquesta convergència.

Solució:
Observeu que X_n té distribució uniforme discreta als nombres

$$\{1/n, 2/n, \ldots, (n-1)/n, 1\}.$$

Aleshores sembla lògic pensar que la distribució límit serà uniforme contínua a l'interval $[0,1]$, que té funció de distribució $F(x) = x$ per a $x \in [0,1]$, $F(x) = 0$ si $x < 0$ i $F(x) = 1$ si $x > 1$. Aquesta funció de distribució és contínua en tota la recta real.

Veiem que efectivament

$$F_n(x) \to F(x) \text{ quan } n \to \infty$$

per a tot $x \in \mathbb{R}$, on $F_n(x)$ és la funció de distribució de X_n. Observeu que $F_n(x) = F(x) = 0$ si $x < 0$ i $F_n(x) = F(x) = 1$ si $x > 1$. Considerem ara $x \in [0,1]$. Aleshores

$$F_n(x) = P(X_n \leq x) = P(X_n \leq \lfloor nx \rfloor/n) = \lfloor nx \rfloor/n.$$

Com que

$$(nx-1)/n \leq \lfloor nx \rfloor/n \leq (nx+1)/n$$

i

$$(nx-1)/n \to_{n\to\infty} x, \; (nx+1)/n \to_{n\to\infty} x,$$

es segueix que

$$F_n(x) = \lfloor nx \rfloor/n \to_{n\to\infty} x = F(x)$$

i això prova la convergència en distribució.

8. Sigui $\{U_n, n \in \mathbb{N}\}$ una successió de v.a. i.i.d. definides per

$$P(U_n = 1) = p, \quad P(U_n = -1) = q, \; 0 \leq p \leq 1, \; p + q = 1.$$

Definim la successió: $V_1 = U_1, V_2 = U_1U_2, \ldots, V_n = \prod_{k=1}^n U_k$.

(a) Calculeu $E[V_n]$.

(b) Trobeu p_n i q_n on $p_n = P(V_n = 1)$ i $q_n = P(V_n = -1)$.

(c) Estudieu la convergència en llei de la successió $\{V_n\}_{n\geq 1}$.

Solució:

(a)

$$E[V_n] = E\left[\prod_{k=1}^{n} U_k\right] = \prod_{k=1}^{n} E[U_k] = (p-q)^n.$$

(b) Hem definit $V_n = \prod_{k=1}^{n} U_k$. Sigui X_n el nombre de U_i's que valen -1. Observeu que $X_n \sim B(n,q)$ i que $V_n = 1$ si i només si X_n és parell. Per tant

$$p_n = \sum_{\substack{k=0 \\ k \text{ parell}}}^{n} \binom{n}{k} q^k p^{n-k}, \quad q_n = \sum_{\substack{k=0 \\ k \text{ senar}}}^{n} \binom{n}{k} q^k p^{n-k}.$$

Observeu que

$$E[V_n] = p_n - q_n = \sum_{\substack{k=0 \\ k \text{ parell}}}^{n} \binom{n}{k} (-1)^k q^k p^{n-k} + \sum_{\substack{k=0 \\ k \text{ senar}}}^{n} \binom{n}{k} (-1)^k q^k p^{n-k} =$$

$$(-q+p)^n = (p-q)^n,$$

que és l'expressió obtinguda a l'apartat anterior.

(c) Observeu que podem expressar p_n en funció de p_{n-1}:

$$p_n = P(V_n = 1) =$$

$$P(V_n = 1|V_{n-1} = 1)P(V_{n-1} = 1) + P(V_n = 1|V_{n-1} = -1)P(V_{n-1} = -1) =$$

$$p \cdot p_{n-1} + (1-p)(1-p_{n-1}).$$

Si prenem límits quan $n \to \infty$ i anomenem $x = \lim_n p_n$, que coincideix amb $\lim_n p_{n-1}$, tenim que

$$x = px + (1-p)(1-x) \iff x(1-p) = (1-p)(1-x) \iff$$

$$x = 1 - x \iff x = \frac{1}{2}.$$

Per tant $\lim_n p_n = \lim_n q_n = 1/2$ i V_n convergeix en distribució a la variable aleatòria que pren els valors -1 i 1 amb la mateixa probabilitat.

3.1 Altres tipus de convergències de succecions de variables aleatòries

9. Siguin $X_1, X_2, \ldots, X_n, \ldots$ variables aleatòries amb distribucions discretes donades per les funcions de probabilitat següents:

$$P(X_n = 0) = 1 - \frac{1}{n}, \; P(X_n = -1) = P(X_n = 1) = \frac{1}{2n}.$$

(a) Proveu que $X_n \xrightarrow{p} 0$ quan $n \to \infty$.

(b) Proveu que $X_n \xrightarrow{r} 0$ quan $n \to \infty$ per a qualsevol $r > 0$.

Solució:

(a) Sigui $\varepsilon > 0$. Si $\varepsilon \geq 1$ aleshores $P(|X_n - 0| > \varepsilon) = 0$. Si $0 < \varepsilon < 1$,

$$P(|X_n - 0| > \varepsilon) = P(X_n \neq 0) = \frac{1}{n} \xrightarrow[n]{} 0.$$

(b)

$$E(|X_n - 0|^r) = E(|X_n|^r) = 0 \cdot \left(1 - \frac{1}{n}\right) + 1 \cdot \frac{1}{n} = \frac{1}{n} \xrightarrow[n]{} 0.$$

10. Siguin $X_1, X_2, \ldots, X_n, \ldots$ variables aleatòries independents amb distribució $U([0,1])$.
    (a) Proveu que $X_{(n),n} = \max\{X_1, \ldots, X_n\} \xrightarrow{p} 1$ quan $n \to \infty$.
    (b) Proveu que $X_{(1),n} = \min\{X_1, \ldots, X_n\} \xrightarrow{p} 0$ quan $n \to \infty$.

    Solució:
    Recordem que la funció de distribució de la $U([0,1])$ val $F(x) = x$ si $x \in [0,1]$. Calculem la funció de distribució de $X_{(1),n}$ i $X_{(n),n}$. Sigui $x \in [0,1]$,

$$F_{X_{(n),n}}(x) = P(\max\{X_1, \ldots, X_n\} \leq x) = P\left(\cap_{i=1}^n \{X_i \leq x\}\right) =$$

$$\prod_{i=1}^n P(X_i \leq x) = x^n,$$

$$1 - F_{X_{(1),n}}(x) = P(\min\{X_1, \ldots, X_n\} > x) = P\left(\cap_{i=1}^n \{X_i > x\}\right) =$$

$$\prod_{i=1}^n P(X_i > x) = (1-x)^n,$$

i per tant $F_{X_{(1),n}}(x) = 1 - (1-x)^n$.

(a) Sigui $\varepsilon > 0$,

$$P(|X_{(n),n} - 1| > \varepsilon) = P(X_{(n),n} \leq 1 - \varepsilon)$$
$$= F_{X_{(n),n}}(1-\varepsilon) = (1-\varepsilon)^n \to 0 \text{ quan } n \to \infty,$$

i per tant $X_{(n),n} \xrightarrow{p} 1$.

(b) Sigui $\varepsilon > 0$,

$$P(|X_{(1),n} - 0| > \varepsilon) = P(X_{(1),n} > \varepsilon)$$
$$= 1 - F_{X_{(1),n}}(\varepsilon) = (1-\varepsilon)^n \to 0 \text{ quan } n \to \infty,$$

i per tant $X_{(1),n} \xrightarrow{p} 0$.

11. Siguin $X_1, X_2, \ldots, X_n, \ldots$ variables aleatòries independents amb distribució $U([0,1])$. Sigui $X_{(1),n} = \min\{X_1, \ldots, X_n\}$. Proveu que $Y_n = nX_{(1),n} \xrightarrow{d} \text{Exp}(1)$ quan $n \to \infty$.

    Solució:
    El conjunt de punts de continuïtat de $Y \sim \text{Exp}(1)$ és $\mathbb{R} - \{0\}$. Si $x < 0$ aleshores $F_{Y_n}(x) = 0 = F_Y(x)$. Per tant només cal veure si $F_{Y_n}(x) \longrightarrow_n F_Y(x)$ per $x > 0$. Calculem la funció de distribució del mínim entre les primeres n observacions, $X_{(1),n}$:

$$1 - F_{X_{(1),n}}(x) = P(\min\{X_1, \ldots, X_n\} > x) = P\left(\cap_{i=1}^n \{X_i > x\}\right) =$$

$$\prod_{i=1}^n P(X_i > x) = (1-x)^n,$$

i per tant $F_{X_{(1),n}}(x) = 1 - (1-x)^n$. Com a conseqüència, si $x > 0$

$$F_{Y_n}(x) = P(Y_n \leq x) = P(X_{(1),n} \leq x/n) = 1 - \left(1 - \frac{x}{n}\right)^n \xrightarrow[n]{} 1 - e^{-x} = F_Y(x)$$

com volíem demostrar.

12. Sigui $\{X_n, n \in \mathbb{N}\}$ una successió de v.a.i.i.d. amb esperança μ i variància finita σ^2. Definim

$$Y_n = \frac{2}{n(n+1)} \sum_{i=1}^{n} iX_i.$$

Demostreu que Y_n convergeix a μ en probabilitat.
Nota: Recordeu que $\sum_{i=1}^{n} i = n(n+1)/2$, aleshores

$$Y_n = \frac{\sum_{i=1}^{n} iX_i}{\sum_{i=1}^{n} i}.$$

Per tant Y_n és una mitjana ponderada de les variables aleatòries $X_1, \ldots, X_n$, on el pes de cada variable aleatòria X_i és proporcional al seu índex i.

Solució:
Calculem la esperança i la variància d'Y_n:

$$E(Y_n) = \frac{2}{n(n+1)} \sum_{i=1}^{n} iE(X_i) = \frac{2\mu}{n(n+1)} \sum_{i=1}^{n} i = \frac{2\mu}{n(n+1)} \frac{n(n+1)}{2} = \mu,$$

$$V(Y_n) = \frac{2^2}{n^2(n+1)^2} \sum_{i=1}^{n} i^2 V(X_i) = \frac{2^2\sigma^2}{n^2(n+1)^2} \sum_{i=1}^{n} i^2 =$$

$$\frac{2^2\sigma^2}{n^2(n+1)^2} \frac{n(n+1)(2n+1)}{6} = \sigma^2 \frac{2(2n+1)}{3n(n+1)}.$$

Observeu que

$$E((Y_n - \mu)^2) = V(Y_n) \underset{n}{\longrightarrow} 0.$$

Aleshores tenim que $Y_n \xrightarrow{2} \mu$ i d'aquí es dedueix la convergència en probabilitat.
Si ho volem provar directament, aplicarem la desigualtat de Tchebychev. Donat $\varepsilon > 0$,

$$P(|Y_n - \mu| > \varepsilon) = P(|Y_n - E(Y_n)| > \varepsilon) \leq \frac{V(Y_n)}{\varepsilon^2} \underset{n}{\longrightarrow} 0,$$

i tenim la convergència en probabilitat directament.

4 Teorema del Límit Central i Lleis dels Grans Nombres

13. Sigui $X_r \sim \text{BinNeg}(r, p)$, $r \in \mathbb{N}$, $p \in (0,1)$. X_r compte el nombre d'experiments independents de Bernoulli amb probabilitat d'èxit p que s'han de fer fins a obtenir r èxits. Recordem que $E(X_r) = r/p$, $V(X_r) = r(1-p)/p^2$. Feu servir el TCL per provar que

$$\frac{pX_r - r}{\sqrt{r(1-p)}} \xrightarrow{d} N(0,1) \text{ quan } r \to \infty.$$

Solució:
Sabem que X_r és la suma d'r distribucions Y_i, $i = 1, \ldots, r$, geomètriques independents de paràmetre p: $X_r = \sum_{i=1}^{r} Y_i$. Aplicant el TCL a les variables Y_i, tenim que

$$\frac{\sqrt{r}(\bar{Y}_r - 1/p)}{\sqrt{(1-p)/p^2}} \xrightarrow{d} N(0,1).$$

Si arreglem el terme de l'esquerra tenim que

$$\frac{\sqrt{r}(\bar{Y}_r - 1/p)}{\sqrt{(1-p)/p^2}} = \sqrt{\frac{r}{1-p}}(p\bar{Y}_r - 1) = \sqrt{\frac{r}{1-p}}(pX_r/r - 1) = \frac{pX_r - r}{\sqrt{r(1-p)}}$$

i per tant

$$\frac{pX_r - r}{\sqrt{r(1-p)}} \xrightarrow{d} N(0,1) \text{ quan } r \to \infty.$$

14. Sigui $h(x)$ una funció real definida en $[0,1]$ tal que se sap que la seva integral

$$I(h) = \int_0^1 h(x)dx$$

existeix y és finita. Es vol calcular $I(h)$ i no és possible fer-ho explícitament. Una forma de resoldre numèricament aquesta integral és el mètode de Monte Carlo, que consisteix en

(i) generar n v.a.i.i.d. $X_1, \ldots, X_n$, amb $X_i \sim U([0,1])$, i

(ii) estimar $I(h)$ mitjançant

$$\hat{I}(h) = \frac{1}{n}\sum_{i=1}^{n} h(X_i).$$

(a) Demostreu que $E(\hat{I}(h)) = I(h)$.

(b) Demostreu que $\hat{I}(h) \xrightarrow{p} I(h)$ quan $n \to \infty$.

(c) Sabem que

$$\pi = 4\int_0^1 \sqrt{1-x^2}dx$$

donat que aquesta integral és l'àrea d'un quart del cercle de centre $(0,0)$ i radi unitat. Feu servir el Teorema del Límit Central per calcular quantes observacions aleatòries X_i s'han de generar per tal que, amb una probabilitat de 0.95, l'estimació de π mitjançant el mètode de Monte Carlo tingui una precisió de 0.01? Dit d'una altra manera, calculeu n tal que

$$P(|\hat{\pi}_{MC} - \pi| \leq 0.01) \geq 0.95.$$

Solució:

(a) Observeu que $\hat{I}(h)$ és la mitjana mostral de les variables $h(X_i)$, $i = 1, \ldots, n$. Aleshores el valor esperat de $\hat{I}(h)$ és $E(h(X_i))$:

$$E(\hat{I}(h)) = E(h(X_i)) = \int_0^1 h(x)f_U(x)dx = \int_0^1 h(x)dx = I(h),$$

on hem fet servir que la densitat de la uniforme és $f_U(x) = 1$ per $x \in [0,1]$.

(b) La convergència en probabilitat segueix del fet que $\hat{I}(h)$ és la mitjana mostral de variables aleatòries amb esperança $I(h)$. S'aplica llavors la LDGN.

(c) En aquest cas $h(x) = \sqrt{1-x^2}$ i $\pi = 4I(h)$. Pel TCL sabem que

$$\frac{\sqrt{n}(\hat{I}(h) - I(h))}{\sigma_h} \xrightarrow{d} N(0,1),$$

on

$$\sigma_h^2 = V(h(X_i)) = E(X_i^2) - I(h)^2 = \int_0^1 h(x)^2 dx - I(h)^2.$$

En aquest cas

$$\sigma_h^2 = \int_0^1 (1-x^2)dx - (\pi/4)^2 = (x - x^3/3)|_0^1 - (\pi/4)^2 = 2/3 - (\pi/4)^2 \approx 0.0498.$$

Ens demanen n que garanteixi una precisió de 0.01 amb una probabilitat de 0.95 a l'estimació de π. Aleshores ha de ser

$$0.95 = P(|4\hat{I}(h) - \pi| \leq 0.01) = P(|\hat{I}(h) - \pi/4| \leq 0.0025) =$$

$$P\left(\sqrt{n}|\hat{I}(h) - I(h)|/\sigma_h \leq \sqrt{n} \cdot 0.0025/\sigma_h\right) \approx P(|Z| \leq \sqrt{n} \cdot 0.0025/\sigma_h),$$

on $Z \sim N(0,1)$. Mirant les taules de la normal sabem que

$$P(|Z| \leq 1.96) = 0.95$$

i per tant fem

$$\sqrt{n} \cdot 0.0025/\sigma_h = 1.96 \iff n = (1.96\sigma_h/0.0025)^2 =$$
$$(1.96 \cdot \sqrt{0.0498}/0.0025)^2 = 30609.87.$$

Per tant faríem $n = 30610$.

15. Sigui $\{X_n\}$ una successió de v.a.i.i.d. $U(0,1)$. Estudieu la convergència en probabilitat de la successió $Z_n = \prod_{i=1}^n X_i^{1/n}$. ($Z_n$ és la mitjana geomètrica de les n primeres variables X_i).

Solució:
Sigui $V_n = \log Z_n = \frac{1}{n}\sum \log X_i$. Observeu que V_n és la mitjana mostral de les variables $Y_i = \log X_i$, $i = 1, \ldots, n$. Per la Llei dels Gran Nombres sabem que

$$V_n = \bar{Y}_n \xrightarrow{p} E(Y_1)$$

Calculem $E(Y_1)$:

$$E(Y_1) = E(\log X_1) = \int_0^1 \log(x)dx =$$
$$\left\{ \begin{array}{c} x = e^{-t} \\ dx = -e^{-t}dx \end{array} \right\} = -\int_0^\infty te^{-t}dt = -\Gamma(2) = -1.$$

Per tant $V_n =\xrightarrow{p} -1$ i com que $g(t) = \exp(t)$ és una funció contínua tenim que

$$Z_n = \exp(V_n) \xrightarrow{p} \exp(-1) = \frac{1}{e}.$$

16. Suposeu que una companyia envia paquets de pesos variables, amb un pes mig de 9 Kg i una desviació estàndard de 4 Kg. Els paquets provenen d'un gran nombre de clients i per tant és raonable el pensar que llurs pesos són independents. Trobeu la probabilitat que el pes total corresponent a 100 paquets sobrepassi una tona.

Solució:
Sigui X_i, $i = 1, \ldots, n = 100$, el pes del paquet i-èsim. Ens demanen aproximar

$$p = P\left(\sum_{i=i}^{100} X_i > 1000\right) = P\left((1/100)\sum_{i=i}^{100} X_i > 1000/100\right) =$$
$$P(\bar{X}_{100} > 10) = P(\sqrt{100}(\bar{X}_{100} - 9)/4 > \sqrt{100}(10-9)/4 = 2.5)$$

Pel TCL, si $Z \sim N(0,1)$

$$p \approx P(Z > 2.5) = 0.0062.$$

17. Es tira 1000 vegades una moneda perfecta i es demana:

(a) Probabilitat que el nombre de cares estigui comprés entre 490 i 510.

(b) L'interval $[a,b]$, centrat en 500, que compleixi $P([a,b]) = 0.95$.

Solució:
Sigui X el nombre de cares obtingudes:

$$X \sim B(1000, 1/2) \approx Y \sim N\left(1000 \cdot \frac{1}{2}, 1000 \cdot \frac{1}{2}\left(1 - \frac{1}{2}\right)\right) \equiv N(500, 250).$$

(a)

$$P(490 \leq X \leq 510) \approx P(490 \leq Y \leq 510) =$$
$$P\left(\frac{490-500}{\sqrt{250}} \leq \frac{Y-500}{\sqrt{250}} \leq \frac{510-500}{\sqrt{250}}\right) =$$
$$P(-0.6324555 \leq Z \leq 0.6324555) = 0.4729107,$$

on $Z \sim N(0,1)$.
Si fem correcció per continuitat,

$$P(490 \leq X \leq 510) \approx P(489.5 \leq Y \leq 510.5) =$$
$$P\left(\frac{489.5-500}{\sqrt{250}} \leq \frac{Y-500}{\sqrt{250}} \leq \frac{510.5-500}{\sqrt{250}}\right) =$$
$$P(-0.6640783 \leq Z \leq 0.6640783) = 0.4933598.$$

La probabilitat exacta és 0.49334, i es pot calcular fent servir taules de la binomial o bé amb la funció `pbinom` de R:
`pbinom(510,1000,.5)-pbinom(489,1000,.5)`
La correcció per continuitat millora l'aproximació.

(b) Volem que

$$0.95 = P(a \leq X \leq b) \approx P(a \leq Y \leq b) =$$
$$P\left(\frac{a-500}{\sqrt{250}} \leq \frac{Y-500}{\sqrt{250}} \leq \frac{b-500}{\sqrt{250}}\right) = P\left(\frac{a-500}{\sqrt{250}} \leq Z \leq \frac{b-500}{\sqrt{250}}\right)$$

i sabem que

$$0.95 = P(-1.96 \leq Z \leq 1.96).$$

Per tant fem

$$\frac{a-500}{\sqrt{250}} = -1.96, \ \frac{b-500}{\sqrt{250}} = 1.96,$$

aleshores

$$a = 500 - 1.96\sqrt{250} = 469.0097, \ b = 500 + 1.96\sqrt{250} = 530.9903.$$

18. Sigui $X_1, X_2, \ldots, X_n, \ldots$ una successió de variables aleatòries independents tals que $E[X_i] = \mu$ per a tot i i $Var(X_i) = \sigma_i^2$. Demostreu que si $\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \longrightarrow 0$ quan $n \to \infty$, aleshores $\overline{X}_n \longrightarrow \mu$ en probabilitat quan $n \to \infty$.

Solució:
Per definició: X_n convergeix en probabilitat a X si per a tot $\epsilon \geq 0$, $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$.
Volem veure que donat $\epsilon \geq 0$,

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \epsilon) = 0.$$

Fent servir la desigualtat de Tchebychev tenim que

$$0 \leq P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\epsilon^2} = \frac{1}{\epsilon^2 n^2}\sum_{i=1}^n Var(X_i) = \frac{1}{\epsilon^2 n^2}\sum_{i=1}^n \sigma_i^2 \equiv b_n.$$

Per hipòtesi sabem que

$$\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \longrightarrow 0,$$

i per tant

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2}{\epsilon^2} = 0.$$

Podem emprar el següent Teorema:

$$0 \leq \ a_n \leq \ b_n \quad \text{i} \quad b_n \to 0, \quad \text{aleshores} \quad a_n \to 0.$$

Definint

$$a_n = P(|\overline{X}_n - \mu| \geq \epsilon),$$

concloem que

$$P(|\overline{X}_n - \mu| \geq \epsilon) \longrightarrow 0$$

i per tant que

$$\overline{X}_n \longrightarrow \mu,$$

en probabilitat.