6 Alguns models de probabilitat.

6.1 Model de probabilitat uniforme.

Uniforme discreta. Una variable aleatòria discreta X que pren n valors diferents ($n < \infty$), $S = \{x_1, x_2, \ldots, x_n\}$, es diu que segueix una **distribució uniforme discreta** si tots els possibles valors tenen la mateixa probabilitat $1/n$, és a dir, la funció de probabilitat és

$$f(x) = P(X = x) = \begin{cases} 1/n, & \text{si } x \in S, \\ 0, & x \notin S. \end{cases}$$

Exemple 21 *Si X té llei uniforme discreta amb $f(x) = 1/n$, per $x = 1, 2, \ldots, n$, demostreu que:*

a) $E(X) = \frac{n+1}{2}$,

b) $\mathrm{var}(X) = \frac{n^2-1}{2}$,

c) com a aplicació, trobeu l'esperança i variància de la puntuació obtinguda en llençar un dau no trucat.

a) L'esperança de X és:

$$E(X) = \sum_{x=1}^{n} x\, f(x) = \sum_{x=1}^{n} x\, \frac{1}{n} = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$$

b) La variància de X és: $\mathrm{var}(X) = E(X^2) - (E(X))^2$,

$$E(X^2) = \sum_{x=1}^{n} x^2\, f(x) = \sum_{x=1}^{n} x^2\, \frac{1}{n} = \frac{1}{n} \frac{n(n+1)(2n+1)}{6},$$

$$\begin{aligned} \mathrm{var}(X) &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n+1}{2}\left(\frac{2n+1}{3} - \frac{n+1}{2}\right) \\ &= \frac{n+1}{12}(4n+2-3n-3) = \frac{n+1}{12}(n-1) = \frac{n^2-1}{12} \end{aligned}$$

c) La variable aleatòria X que enregistra les puntuacions obtingudes en tirar un dau no trucat pren valors sobre el conjunt $S = \{1, 2, 3, 4, 5, 6\}$ i té distribució uniforme discreta amb

$$f(x) = 1/6, \quad \forall x \in S,$$

per tant, $E(X) = (6+1)/2 = 7/2$ *i* $\mathrm{var}(X) = (6^2 - 1)/12 = 35/12$.

Uniforme contínua. Donat un interval $[a,b]$ finit, diem que X té distribució **uniforme en l'interval** $[a,b]$, i escrivim $X \sim U[a,b]$, si X pren valors en l'interval $[a,b]$ distribuïts segons la funció de densitat:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{si } x \in [a,b], \\ 0, & \text{si } x \notin [a,b]. \end{cases}$$

La probabilitat d'un interval $(c,d) \subseteq [a,b]$ és:

$$P(c < X < d) = P(X \in (c,d)) = \int_c^d \frac{1}{b-a}\, dx = \frac{d-c}{b-a},$$

és a dir, és l'àrea del rectangle de base $d-c$ i alçada $1/(b-a)$. La funció de distribució de la llei uniforme és:

$$F(x) = P(X \le x) = \begin{cases} 0, & \text{si } x \le a, \\ \frac{x-a}{b-a}, & \text{si } a \le x \le b, \\ 1, & \text{si } x \ge b. \end{cases}$$

L'esperança i variància són:

$$E(X) = \frac{a+b}{2}, \quad \mathrm{var}(X) = \frac{(b-a)^2}{12}.$$

Proposició 6.1 *Si $X \sim U[a,b]$, aleshores el moment central d'ordre k, μ_k, és*

$$\mu_k = \begin{cases} 0, & \text{si } k \text{ és senar}, \\ \frac{1}{k+1}\left(\frac{b-a}{2}\right)^k, & \text{si } k \text{ és parell}. \end{cases}$$

Demostració:

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_a^b \frac{x}{b-a}\, dx = \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

$$\forall k > 1, \quad \mu_k = E((X-E(X))^k) = \int_a^b \left(x - \frac{b+a}{2}\right)^k \frac{1}{b-a}\, dx$$

fent el canvi de variable $u = x - (b+a)/2$, obtenim:

$$= \int_{(a-b)/2}^{(b-a)/2} \frac{1}{b-a} u^k\, du = \frac{1}{b-a} \frac{u^{k+1}}{k+1}\Big|_{(a-b)/2}^{(b-a)/2}$$

$$= \frac{1}{b-a} \frac{1}{k+1} \left[\left(\frac{b-a}{2}\right)^{k+1} - \left(\frac{a-b}{2}\right)^{k+1} \right]$$

$$= \frac{1}{b-a} \frac{1}{k+1} \frac{(b-a)^{k+1}}{2^{k+1}} (1-(-1)^{k+1}) = \frac{1}{k+1} \frac{(b-a)^k}{2^{k+1}} (1-(-1)^{k+1}).$$

Si k és parell, aleshores $(-1)^{k+1} = -1$ i, per tant:

$$\mu_k = \frac{1}{k+1} \frac{(b-a)^k}{2^{k+1}} (1+1) = \frac{1}{k+1} \left(\frac{b-a}{2}\right)^k.$$

Si k és senar, aleshores $(-1)^{k+1} = 1$ i, per tant:

$$\mu_k = \frac{1}{k+1} \frac{(b-a)^k}{2^{k+1}} (1-1) = 0.$$

□

Observem que

$$\text{var}(X) = \mu_2 = \frac{1}{3} \left(\frac{b-a}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

Exemple 22 *Suposem que el* temps que tarda en donar resposta a un malalt *el personal sanitari d'un hospital és uniforme entre 0 i 8 minuts.*

1. *Determineu la funció de densitat i la funció de distribució i dibuixeu-les.*
2. *Calculeu la probabilitat que una resposta excedeixi els 7 minuts. Dibuixeu l'àrea que representa aquesta probabilitat en la funció de densitat. Com representaríeu aquesta probabilitat en la funció de distribució?*
3. *Quin creieu que és el temps mig de resposta?*

1. La v.a. X té una llei uniforme contínua $U(0,8)$. Per tant, la funció de densitat i funció de distribució de X són:

$$f(x) = \begin{cases} 1/8, & si\ x \in (0,8), \\ 0, & si\ x \notin (0,8), \end{cases} \qquad F(x) = \begin{cases} 0, & si\ x \leq 0, \\ x/8, & si\ 0 \leq x \leq 8, \\ 1, & si\ x \geq 8. \end{cases}$$

2. $P(X > 7) = 1 - P(X \leq 7) = 1 - F(7) = 1 - 7/8 = 1/8$.
3. El temps mig de resposta és $E(X) = (0+8)/2 = 4$ minuts.

6.2 Model binomial.

Bernoulli. És una forma de modelar estadísticament qualsevol experiment aleatori que tingui només **dos resultats possibles**, mútuament excloents, normalment anomenats èxit i fracàs, amb la condició que la probabilitat d'aquests dos resultats es mantingui constant en cada realització de l'experiment (experiments de Bernoulli). Si la probabilitat d'èxit és p i, per tant, la probabilitat de fracàs és $q = 1 - p$, definim la variable aleatòria **de Bernoulli** com

$$X(\omega) = \begin{cases} 1, & \text{si } \omega \text{ és un èxit,} \\ 0, & \text{si } \omega \text{ és un fracàs.} \end{cases}$$

La seva funció de probabilitat és

$$f(x) = P(X = x) = \begin{cases} p, & \text{si } x = 1, \\ 1 - p = q, & \text{si } x = 0. \end{cases}$$

Observem que la v.a. X pren valors sobre el conjunt $S = \{0, 1\}$. L'esperança i variància de X són:

$$E(X) = 0\, f(0) + 1\, f(1) = p,$$

$$var(X) = E(X^2) - (E(X))^2 = 0^2\, f(0) + 1^2\, f(1) - p^2 = p - p^2 = p\,(1-p) = p\,q.$$

Per dir que X té una llei o distribució de Bernoulli escriurem $X \sim Bernoulli(p)$. Observem que si p és coneguda podem trobar la funció de probabilitat de X, per tant, p determina la v.a. X.

Exemple 23 *Tenim una moneda trucada de manera que la probabilitat d'obtenir una cara és 2/3. La variable aleatòria* $X =$*"obtenir una cara en una tirada" té distribució de Bernoulli de paràmetre* $p = 2/3$*, és a dir, si assignem el valor* 1 *al resultat "cara" i* 0 *al resultat "creu",* $P(X = 1) = 2/3$ *i* $P(X = 0) = 1 - 2/3 = 1/3$.

Binomial. Es realitzen n experiments de Bernoulli independents amb la mateixa probabilitat d'èxit p. La variable aleatòria X que compta el *nombre d'èxits observats en aquests n experiments* es diu que té **llei** o **distribució binomial** de paràmetres n i p. En aquest cas, escriurem $X \sim B(n, p)$. La funció de probabilitat de X és:

$$f(x) = P(X = x) = \binom{n}{x} p^x\,(1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n, \quad 0 \leq p \leq 1,$$

on

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \quad 0 \leq x \leq n,$$

recordeu que, per conveni, $0! = 1$.
La v.a. X pren valors sobre el conjunt $S = \{0, 1, 2, \ldots, n\}$.
Aquest model s'aplica a poblacions finites de les quals prenem elements a l'atzar *amb devolució*, i també en poblacions conceptualment infinites, com per exemple, les peces que produeix una màquina, sempre que la proporció d'èxits (o de fracassos) sigui constant a llarg termini, i el resultat en cada moment sigui independent del que hagi succeït anteriorment.

Proposició 6.2 *Si* $X \sim B(n, p)$*, aleshores:*

$$E(X) = n\,p, \quad \text{var}(X) = n\,p\,(1-p).$$

Demostració:

$$E(X) = \sum_{k=0}^{n} k\, P(X=k) = \sum_{k=0}^{n} k \binom{n}{k} p^k\,(1-p)^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k\,(1-p)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{k\, n!}{k!\,(n-k)!} p^k\,(1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!\,(n-k)!} p^k\,(1-p)^{n-k}$$

$$= n\, p \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}\,(1-p)^{n-k}$$

fent el canvi $j = k-1$ en el sumatori i anomenant $m = n-1$, obtenim:

$$= n\, p \sum_{j=0}^{m} \frac{m!}{j!(m-j)!} p^j\,(1-p)^{m-j} = n\, p \sum_{j=0}^{m} \binom{m}{j} p^j\,(1-p)^{m-j}$$

$$= n\, p \sum_{j=0}^{m} P(Y=j) = n\, p,$$

on $Y \sim B(m,p)$ i, per tant, $\sum_{j=0}^{m} P(Y=j) = 1$.
Per a calcular $E(X^2)$ utilitzarem que $X^2 = X(X-1) + X$ i, per tant, $E(X^2) = E(X(X-1)) + E(X)$.

$$E(X(X-1)) = \sum_{k=0}^{n} k\,(k-1)\, P(X=k) = \sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k\,(1-p)^{n-k}$$

$$= \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k\,(1-p)^{n-k} = \sum_{k=2}^{n} k(k-1) \frac{n!}{k!(n-k)!} p^k\,(1-p)^{n-k}$$

$$= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k\,(1-p)^{n-k}$$

$$= n\,(n-1)\, p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2}\,(1-p)^{n-k}$$

fent el canvi $j = k-2$ i anomenant $m = n-2$, obtenim:

$$= n\,(n-1)\, p^2 \sum_{k=0}^{m} \frac{m!}{j!(m-j)!} p^j\,(1-p)^{m-j}$$

$$= n\,(n-1)\, p^2 \sum_{k=0}^{m} \binom{m}{j} p^j\,(1-p)^{m-j} = n\,(n-1)\, p^2,$$

i, per tant,

$$\mathit{var}(X) = n\,(n-1)\, p^2 + n\, p - (n\, p)^2 = -n\, p^2 + n\, p = n\, p\,(1-p).$$

□

Exemple 24 *Tirem 10 vegades la moneda de l'exemple 23 i volem calcular la probabilitat d'obtenir 6 cares. Si anomenem X a la variable aleatòria que compta el nombre de cares que s'obtenen en tirar 10 vegades la moneda anterior, hem de calcular la probabilitat $P(X=6)$ amb $X \sim B(10, 2/3)$, és a dir,*

$$P(X=6) = \binom{10}{6}\left(\frac{2}{3}\right)^6\left(1-\frac{2}{3}\right)^{10-6} = 0.227608.$$

Volem calcular també la probabilitat d'obtenir menys de 3 cares.

$$P(X<3) = P(X=0) + P(X=1) + P(X=2)$$

$$= 1.69 \cdot 10^{-5} + 3.39 \cdot 10^{-4} + 9.14 \cdot 10^{-3} = 9.50 \cdot 10^{-3}.$$

Exemple 25 *El 25% dels arbres d'un cert tipus tenen una malaltia a les fulles. Triem 5 arbres a l'atzar. Quina és la probabilitat que tots tinguin la malaltia? Quina és la probabilitat que menys de 2 la pateixin? I la que 2 o més la pateixin?*
Considerem la v.a. $X =$"nombre d'arbres malalts d'un total de 5", que té una llei $B(5, 0.25)$. Ens demanen calcular les següents probabilitats:

$$P(X=5) = \binom{5}{5} 0.25^5(1-0.25)^0 = 0.25^5 = 0.00098.$$

$$P(X<2) = P(X=0) + P(X=1)$$

$$= \binom{5}{0} 0.25^0\, 0.75^5 + \binom{5}{1} 0.25^1\, 0.75^4 = 0.6328.$$

$$P(X \geq 2) = 1 - P(X<2) = 1 - 0.6328 = 0.3672.$$

Exercici 9 *Se sap que certa llavor germina amb probabilitat 0.9. Es fan 20 parcel·les i s'hi planta una llavor a cada parcel·la. Quina és la probabilitat que en germinin menys de 15? Quin és el nombre esperat de germinacions?*

Exercici 10 *Un de cada 10^9 bacteris d'*Escherichia coli *és mutant i es fa resistent a l'anti-biòtic. Es té un cultiu amb 2×10^9 bacteris. Quina és la probabilitat que cap d'ells faci mutació? Quin és el nombre esperat de mutacions?*

Proporció d'èxits en n experiments. Considerem la v.a. $Y = X/n$, on $X \sim B(n, p)$, que representa la *proporció d'èxits en n experiments.*

Proposició 6.3 *Si $X \sim B(n,p)$ i $Y = X/n$, aleshores:*

a) $E(Y) = p$ *i* $\mathrm{var}(Y) = p(1-p)/n$,
b) $P(|Y-p| < c) \geq 1 - p(1-p)/(nc^2)$, *per a tota constant positiva c,*
c) $\lim_{n\to\infty} P(|Y-p| < c) = 1$.

Demostració: Com que $X \sim B(n,p)$, aleshores $E(X) = np$ i $var(X) = np(1-p)$. Vegem a):

$$E(Y) = E\left(\frac{X}{n}\right) = \frac{1}{n}np = p,$$

$$var(Y) = var\left(\frac{X}{n}\right) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}.$$

Vegem b):

$$P(|Y-p| < c) = 1 - P(|Y-p| \geq c) \geq 1 - \frac{var(Y)}{c^2} = 1 - \frac{p(1-p)}{nc^2},$$

resultat que s'obté després d'aplicar la desigualtat de Txebixev.
Vegem c):

$$\lim_{n\to\infty} P(|Y-p| < c) \geq \lim_{n\to\infty}\left(1 - \frac{p(1-p)}{nc^2}\right) = 1,$$

i com que $0 \leq P(|Y-p| < c) \leq 1$, en ser una probabilitat, tenim que $\lim_{n\to\infty} P(|Y-p| < c) = 1$. □
Observacions: c) ens diu que quan n tendeix a infinit, és a dir, a mesura que augmentem el nombre d'experiments, la probabilitat que la proporció d'èxits estigui entre $p-c$ i $p+c$ s'aproxima a 1, per a tota constant positiva c. Aquest resultat s'anomena *llei dels grans nombres* i observeu que s'aplica a la proporció d'èxits, no al nombre absolut d'èxits, és a dir, no és cert que quan n sigui gran el nombre d'èxits estigui necessàriament a prop de np.

6.3 Model Normal.

La distribució normal és molt important en estadística, per sí mateixa, perquè modela bé moltes situacions reals, perquè lleis com la Binomial poden aproximar-se per la llei normal (sota certes condicions dels paràmetres) i perquè nombroses distribucions que s'utilitzen en inferència estadística com la t d'Student o la χ^2, es deriven d'ella.
Es va investigar per primera vegada en el segle XIX quan els científics observaren un grau sorprenent de regularitat en els errors de mesura. Van veure que els patrons (les distribucions) que observaven podien aproximar-se per corbes contínues, a les que anomenaven *corbes normals d'errors* i les atribuïen a les lleis de l'atzar. Abraham de Moivre (1667-1745), Pierre Laplace (1749-1827) i Karl Gauss (1777-1855) van estudiar per primera vegada les propietats matemàtiques d'aquestes corbes normals.
Exemples clàssics de distribució normal són:

1. Mesures biomèdiques com: els pesos, les alçades, les talles, longituds, el coeficient d'intel·ligència, etc.

2. Els errors aleatoris de mesura en observacions astronòmiques, pesades d'una balança, mesures realitzades mitjançant algun aparell.

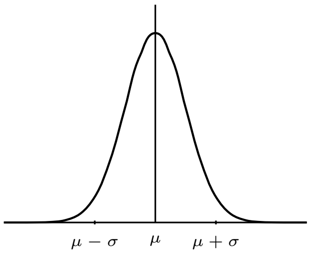
Definició: Direm que una variable aleatòria contínua X té una **distribució normal** de paràmetres μ i σ^2, i escriurem $X \sim N(\mu, \sigma^2)$, si la funció de densitat de X és

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\,\sigma^2}\right), \quad x \in \mathbb{R},\ \sigma > 0. \tag{6}$$

La representació gràfica de $f(x)$ és coneguda com la *campana de Gauss.* L'esperança i variància són:

$$E(X) = \mu, \quad \mathrm{var}(X) = \sigma^2.$$

Figura 1: Exemple de distribució normal $N(\mu, \sigma^2)$.



Propietats:

- La v.a. X pot prendre qualsevol valor en $(-\infty, +\infty)$.
- La funció de densitat X, $f(x)$, és simètrica respecte de l'esperança de X, μ, té dos punts d'inflexió en $x = \mu + \sigma$ i $x = \mu - \sigma$, i una asímptota horitzontal en la recta $y = 0$, és a dir, $\lim_{x\to+\infty} f(x) = 0$, $\lim_{x\to-\infty} f(x) = 0$.
- El 99.7% de l'àrea de $f(x)$ està concentrada en l'interval $[\mu-3\sigma, \mu+3\sigma]$.
- És una distribució unimodal: els intervals que tenen màxima probabilitat són els centrats entorn de l'esperança.
- Si $X \sim N(\mu, \sigma^2)$ i $a \neq 0$, aleshores $a\,X \sim N(a\,\mu, a^2\,\sigma^2)$.
- Si $X_1 \sim N(\mu_1, \sigma_1^2)$ i $X_2 \sim N(\mu_2, \sigma_2^2)$ independents, aleshores $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Quan $\mu = 0$ i $\sigma = 1$, s'obté una cas molt particular de distribució normal: **la distribució normal tipificada o estàndard**, que s'escriu $N(0, 1)$. Les variables aleatòries que segueixen aquesta distribució solen denotar-se per la lletra Z, per remarcar aquest caràcter especial. La funció de densitat d'una variable aleatòria $Z \sim N(0,1)$ és

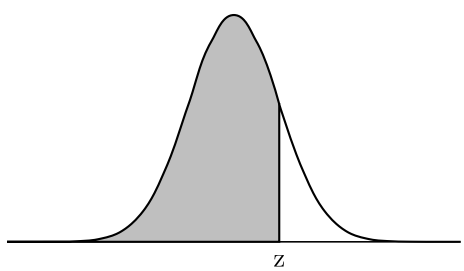
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < +\infty,$$

que és una funció simètrica, centrada en el 0, i d'àrea total igual a 1. El 99.7% de l'àrea està concentrada en l'interval $[-3, 3]$, però això no vol dir que la variable aleatòria Z no pugui prendre valors més grans que 3 o inferiors a -3, sinó que *aquests valors són poc probables*.
Les variables aleatòries normals satisfan, entre d'altres, les següents propietats:

- si $X \sim N(\mu, \sigma^2)$, aleshores la **tipificada** de X és
$$Z = \frac{X - \mu}{\sigma} \tag{7}$$
i $Z \sim N(0,1)$.
- Recíprocament, si $Z \sim N(0,1)$, aleshores la inversa de la tipificada és
$$X = \mu + \sigma Z \tag{8}$$
i $X \sim N(\mu, \sigma^2)$.

Càlcul de probabilitats amb la distribució $N(0,1)$. Sigui F la funció de distribució d'una variable aleatòria $Z \sim N(0,1)$, és a dir, la probabilitat acumulada fins al punt z, $F(z) = P(Z \leq z)$.
La funció F no té una expressió analítica exacta i els seus valors s'obtenen per mètodes aproximats. Aquesta funció es troba tabulada en la majoria de textos d'Estadística.
Gràficament, $F(z)$ és l'àrea de la funció de densitat (6) situada a l'esquerra del punt z. Vegeu la figura 2.

Figura 2: Probabilitat acumulada sota la corba $N(0,1)$.



Exercici 11 *Utilitzant les taules calculeu:*

1. $P(Z \leq 2)$, $P(Z \leq -2)$,
2. $P(Z \leq 1.03)$, $P(Z \leq -1.03)$,
3. $P(Z \leq -0.748)$, $P(Z \leq 0.748)$,
4. $P(Z \leq 4)$, $P(Z \leq -4)$.

A partir de la gràfica es veu que la probabilitat que Z prengui un valor concret qualsevol és zero, perquè un punt no determina cap àrea. És a dir,

$$P(Z = z) = 0.$$

Això no vol dir que la variable no pugui prendre un valor concret, sinó que hi ha tants valors, *una infinitat no numerable*, que la probabilitat d'un valor concret és zero. **Només tenen probabilitat no nul·la els intervals.**
A partir de la figura anterior també resulta obvi que

$$P(Z > z) = 1 - F(z), \quad P(Z \geq z) = 1 - F(z).$$

Exercici 12 *Utilitzant les taules calculeu:*

1. $P(Z \geq 2.1)$, $P(Z \geq -2.22)$,
2. $P(Z \geq 1.13)$, $P(Z \geq -1.335)$,
3. $P(Z \geq 1.96)$, $P(Z \geq -2.33)$,
4. $P(Z \geq 2.576)$, $P(Z \geq -1.642)$.

La probabilitat que Z prengui valors en un interval (sense distingir si és obert o tancat) es calcula fent la diferència d'àrees de l'extrem superior menys l'extrem inferior, és a dir

$$P(a \leq Z \leq b) = F(b) - F(a).$$

Exercici 13 *Utilitzant les taules calculeu:*

1. $P(1 \leq Z \leq 2)$,
2. $P(0.5 \leq Z \leq 1.5)$,
3. $P(-1.96 \leq Z \leq 1.96)$,
4. $P(-2.58 \leq Z \leq 2.58)$.

Càlcul de probabilitats amb la distribució $N(\mu, \sigma^2)$. A partir de la relació (7) entre la normal general i la normal estàndard podem obtenir les probabilitats d'una variable aleatòria $X \sim N(\mu, \sigma^2)$ a partir de la funció de distribució F de la normal estàndard. Per exemple, si volem calcular

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = F\left(\frac{x-\mu}{\sigma}\right).$$

Per tant, tindrem que

$$P(X > x) = 1 - F\left(\frac{x-\mu}{\sigma}\right),$$

$$P(x \le X \le x') = F\left(\frac{x'-\mu}{\sigma}\right) - F\left(\frac{x-\mu}{\sigma}\right).$$

Les mateixes propietats es tenen substituint els símbols $\le$ per $<$ i $>$ per $\ge$, ja que la probabilitat d'un valor concret és zero.

Exercici 14 *Suposeu que l'alçada de les dones d'una certa població és una variable aleatòria amb distribució normal de mitjana 163 cm i variància 49. Calculeu:*

1. *La probabilitat que una dona escollida a l'atzar faci menys de 172 cm.*
2. *La probabilitat que una dona escollida a l'atzar faci menys de 160 cm.*
3. *La probabilitat que una dona escollida a l'atzar faci més de 155 cm.*
4. *La probabilitat que una dona escollida a l'atzar faci més de 175 cm.*
5. *La probabilitat que una dona escollida a l'atzar faci entre 158 cm i 166 cm.*

Càlcul dels percentils de la distribució $N(0, 1)$. Si α és un nombre real entre 0 i 1, el *percentil* $\alpha \times 100$ *de la distribució* $N(0, 1)$ és el valor

$$z_\alpha \text{ tal que } P(Z \le z_\alpha) = \alpha.$$

Aquest valor el trobarem (aproximadament) a les taules de la Distribució Normal estàndard, però utilitzant aquestes taules de forma inversa, és a dir, partint de l'àrea que més s'aproxima a α mirarem quin valor z li correspon.

Exercici 15 *Suposant que $Z \sim N(0, 1)$, calculeu utilitzant les taules:*

1. *El percentil 25 (o primer quartil) és* $z_{0.25} = 0.67$
2. *El percentil 50 (o mediana) és*
3. *El percentil 75 (o tercer quartil) és*
4. *El percentil 95 és*
5. *El percentil 99 és*

6. *Entre quins dos valors de Z simètrics respecte del 0 es troben el 95% de les observacions? Això vol dir quins dos percentils deixen en la part central de la distribució una àrea de 0.95 i a les cues una àrea de 0.025. Aquests percentils són* $z_{0.025} = -1.96$ *i* $z_{0.975} = 1.96$.
7. *Entre quins dos valors de Z simètrics respecte del 0 es troben el 90% de les observacions?*
8. *Entre quins dos valors de Z simètrics respecte del 0 es troben el 99% de les observacions?*

Càlcul dels percentils de la distribució $N(\mu, \sigma^2)$. El *percentil* $\alpha \times 100$ *de la distribució* $N(\mu, \sigma^2)$ és el valor

$$x_\alpha \text{ tal que } P(X \leq x_\alpha) = \alpha.$$

De la relació (8) deduïm que el percentil $\alpha \times 100$ de la distribució $N(\mu, \sigma^2)$ és el valor

$$x_\alpha = z_\alpha \, \sigma + \mu,$$

on z_α és el percentil $\alpha \times 100$ de la normal estàndard.

Exercici 16 *Considereu la variable alçada de les dones* $X \sim N(163, 49)$ *i calculeu:*

1. *El percentil 25 (o primer quartil) és* $x_{0.25} = z_{0.25} \times 7 + 163 = 158.31$. *Què vol dir?*
2. *El percentil 50 (o mediana) és*
3. *El percentil 75 (o tercer quartil) és*
4. *El percentil 95 és*
5. *El percentil 99 és*
6. *Entre quins dos valors de X simètrics respecte de* $\mu = 163$ *es troben el 95% de les observacions? Això vol dir quins dos percentils deixen en la part central de la distribució una àrea de 0.95 i a les cues una àrea de 0.025.*
7. *Entre quins dos valors de X simètrics respecte de* $\mu = 163$ *es troben el 90% de les observacions?*
8. *Entre quins dos valors de X simètrics respecte de* $\mu = 163$ *es troben el 99% de les observacions?*

6.4 Aproximació de la llei Binomial per la llei Normal

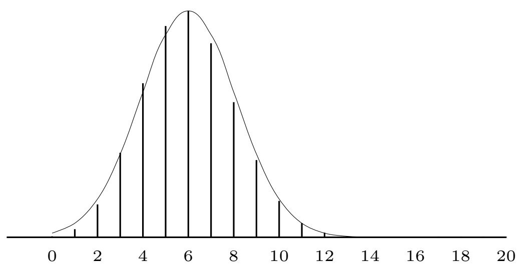
De vegades, s'introdueix la distribució normal com una distribució contínua que proporciona una aproximació propera a la distribució binomial quan el nombre d'experiments n és molt gran i la probabilitat d'èxit p és propera a 1/2.

El **teorema de De Moivre-Laplace** assegura que la distribució Binomial $B(n, p)$ es pot aproximar per la distribució normal quan n és prou gran, és a dir,

$$X \sim B(n, p) \approx N(np, np(1 - p)),$$

sempre que $n \geq 30$ i $0.1 < p < 0.9$. Vegeu les figures 3 i 4. Sovint, però, la distribució normal s'utilitza per aproximar probabilitats binomials malgrat que n sigui relativament petita. Una *bona regla empírica* és utilitzar l'aproximació només quan $np > 5$ i $n(1 - p) > 5$.

Figura 3: Funció de probabilitat d'una distribució $B(20, 0.3)$ i funció de densitat d'una distribució normal $N(6, 4.2)$.



Exemple 26 *Tirem una moneda equilibrada 1000 vegades. Calculeu la probabilitat que el nombre de cares estigui entre 490 i 510.*
Considerem la v.a. $X =$*"nombre de cares obtingudes en tirar una moneda 1000 vegades", que té distribució binomial* $B(1000, 1/2)$. *com que* $np = 500 > 5$ *i* $n(1 - p) = 500 > 5$, *podem aproximar la llei de* X *per una llei normal* $N(500, 250)$. *Aleshores:*

$$P(490 < X < 510) = P\left(\frac{490 - 500}{\sqrt{250}} < \frac{X - 500}{\sqrt{250}} < \frac{510 - 500}{\sqrt{250}}\right)$$

$$= P(-0.63 < Z < 0.63), \quad on\ Z \sim N(0, 1).$$

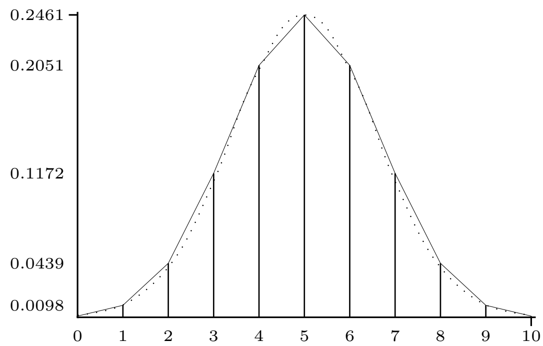
Finalment, utilitzant les taules de la llei normal estàndard,

$$P(-0.63 < Z < 0.63) = F(0.63) - F(-0.63) = 2F(0.63) - 1 = 0.471306.$$

Exemple 27 *El 5% dels termòstats fabricats per una empresa no satisfan las especificacions tècniques. S'extreu un lot de 2000 termòstats.*

a) Utilitzant l'aproximació a la llei normal, calcular la probabilitat que el lot contingui més de 120 termòstats defectuosos.

Figura 4: Funció de probabilitat i polígon de freqüències d'una distribució $B(10, 0.5)$ i funció de densitat d'una distribució $N(5, 2.5)$



b) S'extreu del lot una mostra de 200 termòstats. Si el nombre de termòstats defectuosos en la mostra és inferior a 12, s'accepta el lote, i es rebutja en cas contrari. Quina és la probabilitat que s'accepti el lot?

a) La variable aleatòria X que compta el nombre de termòstats defectuosos en un lot de mida 2000 *té una distribució binomial de paràmetres $n = 2000$ i $p = 0.05$. Com que n és gran, és convenient aproximar la distribució binomial per la distribució normal de paràmetres $\mu = n\,p = 100$ i $\sigma^2 = n\,p\,(1-p) = 95$ i, per tant, $\sigma = \sqrt{95}$. Aleshores, tipificat la variable aleatòria X i utilitzant la taula de la distribució normal estàndard, s'obté:*

$$P(X > 120) = P\left(\frac{X - 100}{\sqrt{95}} > \frac{120 - 100}{\sqrt{95}}\right) \approx P(Z > 2.052) = 0.0200.$$

b) Ara, considerem la variable aleatòria Y que compta el nombre de termòstats defectuosos en la mostra de mida 200. La distribució de Y és binomial de paràmetres $n = 200$ i $p = 0.05$, que, com abans, l'aproximarem per una distribució normal de paràmetres $\mu = n\,p = 10$ i $\sigma^2 = n\,p\,(1-p) = 9.5$. En aquest cas, $\sigma = \sqrt{9.5}$. La probabilitat d'acceptar el lot és:

$$P(Y < 12) = P\left(\frac{Y - 10}{\sqrt{9.5}} < \frac{12 - 10}{\sqrt{9.5}}\right) \approx P(Z < 0.649) = 0.7418.$$

Exercici 17 *En l'exemple 26 trobeu aquell valor a tal que $P(500 - a < X < 500 + a) = 0.95$. (Solució: $a = 30.99$).*

Exercici 18 *Una empresa produeix paquets d'arròs. El pes dels paquets es distribueix segons una normal de mitjana 1 kg i desviació 50 g. L'empresa té un mecanisme de control que retira automàticament tots els paquets que* *pesen menys de 0.9 kg i els que pesen més de 1.1 kg. La resta de la producció es posa a la venda.*

a) Quin percentatge dels paquets produïts es retiren pel mecanisme de control?

b) Si un paquet passa el control, quina és la probabilitat que tingui un pes superior a 1.05 kg?

Exercici 19 *S'estima que el pes en el moment del naixement dels nadons d'un cert país segueix una distribució normal de mitjana 2.6 kg i desviació 0.5 kg.*

a) Si els nadons de menys de 1.7 kg necessiten passar per un període d'incubació artificial, quin percentatge dels nadons necessiten passar per aquest període?

b) Quin percentatge dels nadons pesa més de 3.5 kg?

c) A partir de quin pes es troben el 10% de nadons que més pesen?

d) Quin és l'interval centrat en la mitjana que conté el pes del 90% dels nadons?

Exercici 20 *La* fiabilitat *d'un cert tipus de fusible elèctric és la probabilitat que un fusible, agafat a l'atzar entre els fusibles acabats de produir, funcioni correctament sota les condicions per a les quals va ser dissenyat. En una comprovació rutinària, es van provar 1000 fusibles i se'n van trobar 27 de defectuosos. Suposant que la fiabilitat del fusible és de 0.98, quina és la probabilitat d'observar 27 o més fusibles defectuosos?*