



Departament d'Estadística
i Investigació Operativa

UNIVERSITAT POLITÈCNICA DE CATALUNYA



Models de VA i simulació

Conceptes

Bloc B – Probabilitat i Estadística

2023

Índex

Models de Variables Aleatòries Discretes (VAD)

Bernoulli

Binomial

Geomètrica

Binomial Negativa

Poisson

Models de Variables Aleatòries Contínues (VAC)

Exponencial

Uniforme

Normal

Teorema Central del Límit

Models de VAD i VAC

- **A Wikipedia** (https://en.wikipedia.org/wiki/List_of_probability_distributions): “Many probability distributions that are important in theory or applications have been given specific names.”
- **VAD – VA Discretes**: Binomial, Poisson, Bernoulli, Geomètrica, Binomial Negativa
- **VAC – VA Contínues**: Exponencial, Normal, Uniforme
- A partir dels paràmetres de cada model es calculen **indicadors**
 - Esperança $\rightarrow E(X) = \mu_X$
 - Variància $\rightarrow V(X) = \sigma_X^2$
- A partir de les **funcions de probabilitat i distribució** de probabilitat es calculen probabilitats:

VAD	VAC
$P(X=k) = p_X(k)$	$P(X=k) = 0$
$P(X \leq k) = F_X(k) = S_{j \leq k} p_X(j)$	$P(X \leq k) = F_X(k)$
$P(X < k) = P(X \leq k-1) = F_X(k-1)$	$P(X < k) = P(X \leq k) = F_X(k)$
$P(a < X \leq b) = F_X(b) - F_X(a)$	$P(a \leq X \leq b) = F_X(b) - F_X(a)$

- A més, es calcularan inverses (donada una probabilitat α , calcular el **quantil α** , o **percentil α en %**):
 x_α és el quantil α de X si es compleix: $F_X(x_\alpha) = \alpha$ ($0 \leq \alpha \leq 1$)

Model de Bernoulli

- **Definició:** Número d'èxits en la realització d'un únic experiment amb 2 possibles resultats*: **0** ("no èxit") i **1** ("èxit")
- **Notació:** $X \sim \text{Bern}(p)$
- **Paràmetres:** p (probabilitat d'èxit)
- **Funció de probabilitat:**

* Parlem de respostes binàries o dicotòmiques

K	$P_X(k)$
0	$1-p = q$
1	p

Els valors "0" i "1" poden tenir un sentit ampli:

- "1" significa "èxit" en l'opció d'interès. En un sentit ampli pot significar *encert, positiu, ...*;

- "0" significa "no èxit" en l'opció d'interès. Representa el complementari: *error, fracàs, negatiu, ...*

- És el model teòric general més senzill aplicable a una variable aleatòria. El cas més habitual són experiències aleatòries que impliquen repeticions de proves Bernoulli. És necessari que unes proves siguin **independents** d'altres i que la **probabilitat d'èxit sigui constant** i igual a p

Models associats a la Bernoulli

- En una experiència aleatòria que implica repetició de proves Bernoulli independents, es plantegen com a distribucions interessants les següents **VAD**:
 - **Binomial**: sobre n repeticions, número “d’èxits” totals
 - **Geomètrica**: número de repeticions fins observar el primer “èxit”
 - **Binomial negativa**: número de repeticions fins observar el r -èssim “èxit”
- En experiències aleatòries on el número de repeticions n és molt gran i p és un valor petit (fenòmens *estranyos*), pot ser més fàcil identificar la mitjana (np) d’“èxits” (en l’interval-unitat) que explícitament el valor de n i p . En aquest cas es plantegen com distribucions interessants:
 - **Poisson**: número “d’èxits” en l’interval → **VAD**
 - **Exponencial**: temps entre “èxits” → **VAC**

En aquests darrers casos parlem d’un **procés de Poisson** en el qual la VAD Poisson i la VAC Exponencial comparteixen un paràmetre o taxa que relaciona la mitjana d’èxits i el temps entre “èxits”

(http://en.wikipedia.org/wiki/Poisson_point_process)

Model Binomial

- **Definició:** Número d'èxits en la repetició de n proves de *Bernoulli* independents amb probabilitat constant p
- **Notació:** $X \sim B(n, p)$ Notació de funcions en **R** pel model: `dbinom()`, `pbinom()`, `qbinom()`
- **Paràmetres:** n (nombre de repeticions), p (probabilitat d'observar 1 èxit)

- **Funció de probabilitat:**

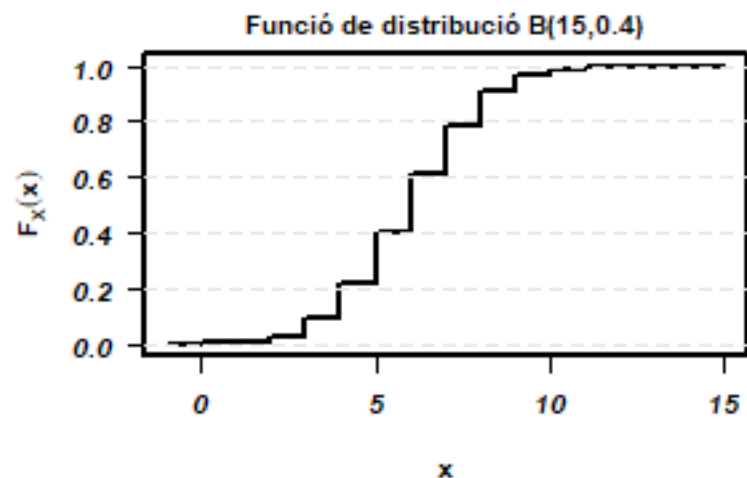
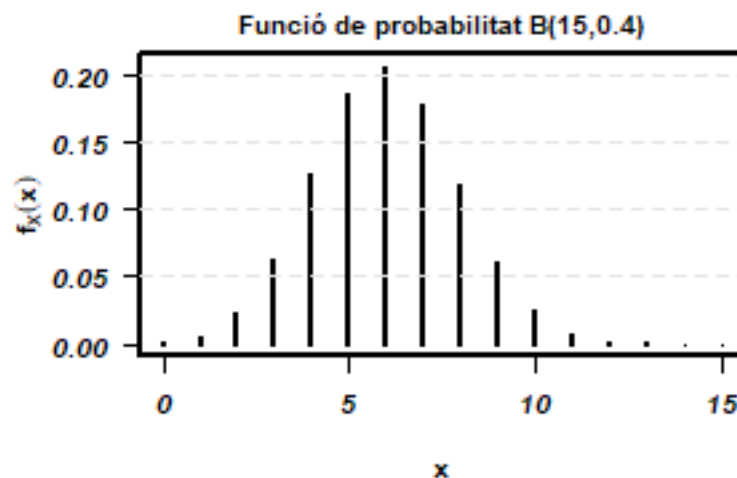
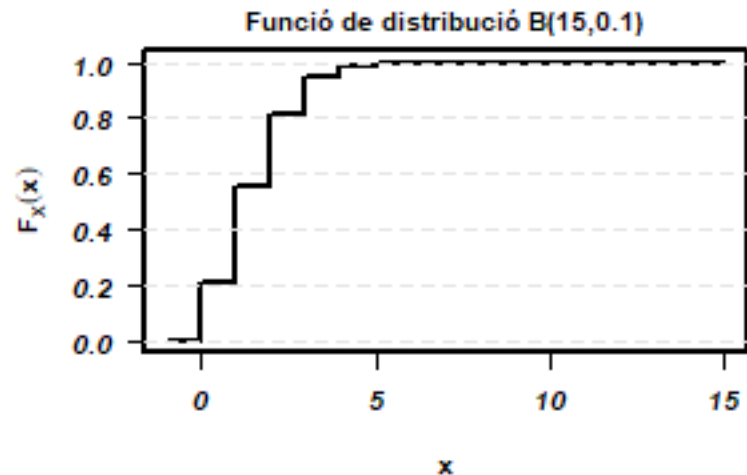
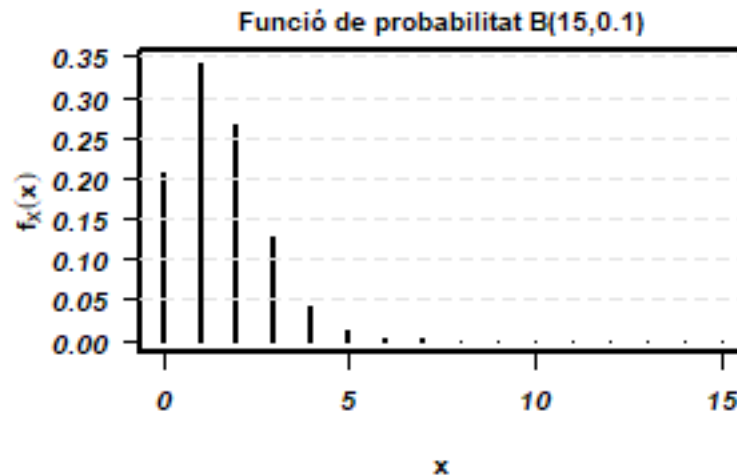
$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \quad \text{amb } k = 0, 1, \dots, n$$

on $[q = 1 - p]$

- No té **funció de distribució** analítica explícita → És el sumatori de probabilitats puntuals
- **Indicadors:**
 - $E(X) = n \cdot p$
 - $V(X) = n \cdot p \cdot q$

Model Binomial. Representació gràfica

Ex: Com es distribueix el número de correus *spam* entre els 15 primers rebuts al dia segons si la probabilitat de que un correu sigui *spam* és 0.1 o 0.4?



Model Binomial. Ex. de càlcul de probabilitats

Sigui $X \sim B(n=20, p=0.5)$

- **Probabilitat puntual.** Quina és la probabilitat de 14?
 - Amb fórmules $\rightarrow P(X = 14) = \binom{20}{14} \cdot 0.5^{14} \cdot 0.5^6 = \mathbf{0.037}$
 - Amb R $\rightarrow P(X = 14) = \text{dbinom}(x = 14, size = 20, prob = 0.5) = \mathbf{0.03696442}$
- **Probabilitat acumulada.** Quina és la probabilitat de 14 o menys?
 - Amb fórmules $\rightarrow P(X \leq 14) = \binom{20}{0} \cdot 0.5^0 \cdot 0.5^{20} + \dots + \binom{20}{14} \cdot 0.5^{14} \cdot 0.5^6 = \mathbf{0.979}$
 - Amb R $\rightarrow P(X \leq 14) = \text{pbinom}(q = 14, size = 20, prob = 0.5) = \mathbf{0.9793053}$
- **Quantils.** Quin és el valor tq la probabilitat de quedar per sota d'ell és almenys 0.95?
 - Amb fórmules $\rightarrow P(X \leq x_{0.95}) = 0.95 \rightarrow$ *Molt complicat!!!*
 - Amb R $\rightarrow P(X \leq x_{0.95}) = 0.95 \rightarrow \text{qbinom}(p = 0.95, size = 20, prob = 0.5) = \mathbf{14}$

dbinom, pbinom, qbinom són funcions en R, Rstudio, i R online (<https://rdr.io/snippets/>)

El model Binomial permet relacionar probabilitats de variables que segueixen models amb paràmetre p o bé $1-p$:

$P(X \leq k) = 1 - P(Y \leq n-k-1)$ on $X \sim B(n, p)$ i $Y \sim B(n, 1-p)$ (o bé $\text{pbinom}(k, n, p) = 1 - \text{pbinom}(n-k-1, n, 1-p)$)

Model Geomètric

- **Definició:** nombre d'intents (k) d'un experiment de Bernoulli fins observar el **primer èxit**
- **Notació:** $X \sim \text{Geom}(p)$ Notació de funcions en R pel model: `dgeom()`, `pgeom()`, `qgeom()`

Les funcions en R són pel número de fracassos enlloc d'intents (*k intents són k-1 fracassos*)

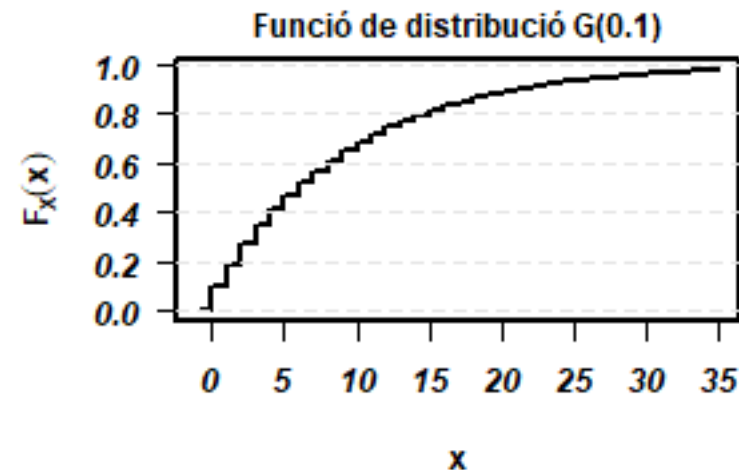
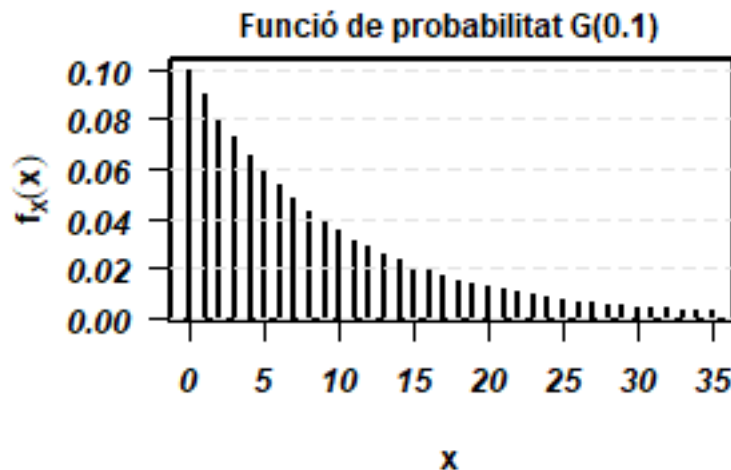
- **Paràmetres:** p (probabilitat d'observar 1 èxit)
- **Funció de probabilitat :**

$$P(X = k) = q^{k-1} \cdot p \quad k \geq 1$$

Indicadors:

$$E(X) = 1/p$$

$$V(X) = (1-p)/p^2$$



No està acotada, qualsevol valor enter > 0 és possible [Ex. “tirar el dau moltes vegades fins que surti el primer 1”]. Encara que el més probable és que el número d'intents no sigui molt alt: quan p augmenta, $P_X(k)$ es trasllada a valors més baixos, i $F_X(k)$ creix més ràpid.

Model Binomial negativa

- **Definició:** nombre d'intents (k) d'un experiment de Bernoulli fins observar r èxits

- **Notació:** $X \sim \text{BN}(r, p)$ Notació de funcions en **R** pel model: `dnbinom()`, `pnbinom()`, `qnbinom()`

Les funcions en **R** són pel número de fracassos enlloc d'intents (*k intents són k-r fracassos*)

- **Paràmetres:** r (nombre d'èxits), p (probabilitat d'observar 1 èxit)

- **Funció de probabilitat:**

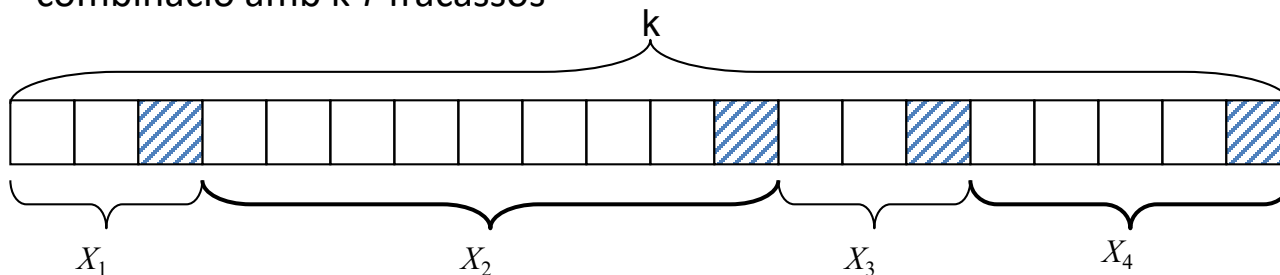
Indicadors:

$$P(X = k) = \binom{k-1}{r-1} \cdot p^r \cdot q^{k-r}$$

$$E(X) = r/p$$

$$V(X) = r(1-p)/p^2$$

Sense comptar l'últim intent (que ha de ser un èxit), són $r-1$ èxits barrejats en qualsevol combinació amb $k-r$ fracassos



Nota: En general, podem pensar que una BN és una suma de r geomètriques independents
Si $r=1$ tenim la distribució geomètrica

Model Poisson

- **Definició:** Número d'ocurrències favorables en un determinat interval de temps o espai

Al igual que en el model binomial pot haver-hi pròpiament una repetició d'experiències idèntiques tipus Bernoulli, però també pot correspondre a fenòmens que ocorren inesperadament

(Ex: trucades a una centraleta es poden representar amb una variable Poisson de “nombre de trucades per hora”, sabent la mitjana de trucades rebudes per hora)

- **Notació:** $X \sim P(\lambda)$ Notació de funcions en **R** pel model: dpois(), ppois(), qpois()

- **Paràmetres:** λ (taxa de l'esdeveniment)

- **Funció de probabilitat:**
$$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!} \quad \text{amb } k = 0, 1, \dots$$

- No té **funció de distribució** analítica → És el sumatori de probabilitats puntuals

- **Indicadors:**

- $E(X) = \lambda$

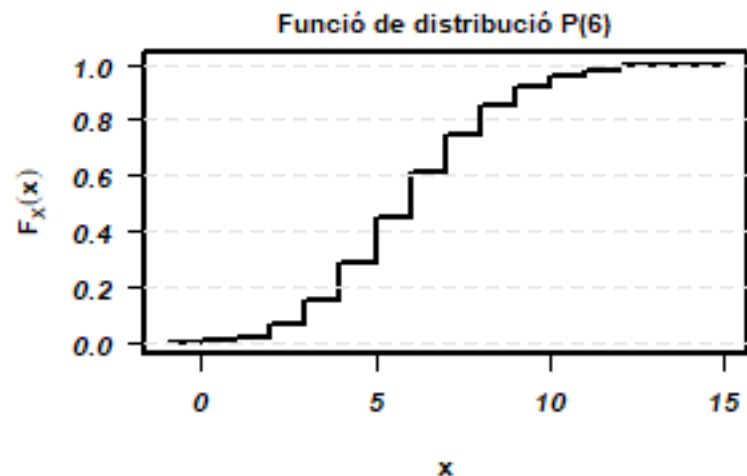
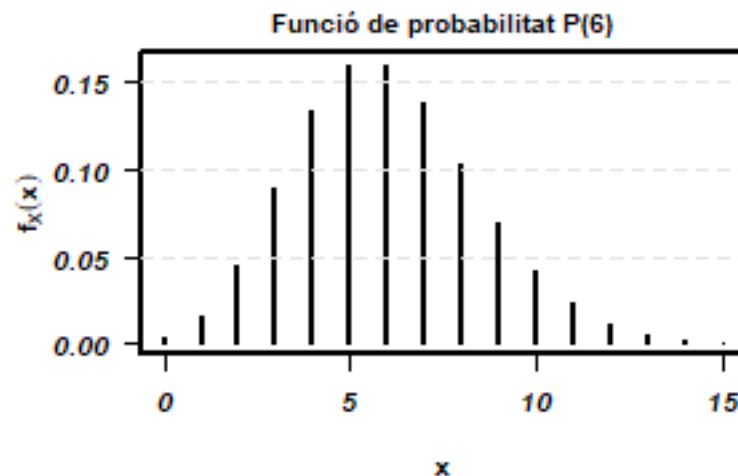
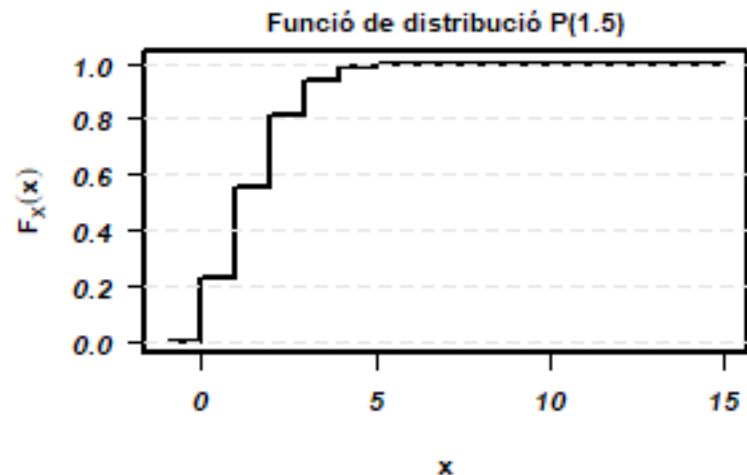
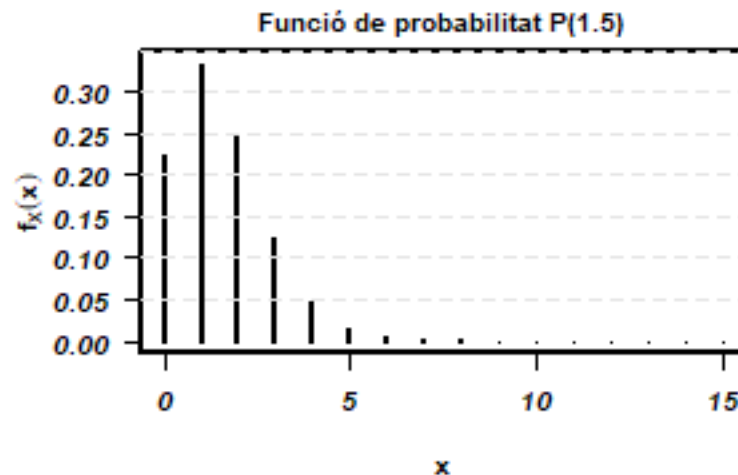
- $V(X) = \lambda$

(λ és un número real positiu que representa la taxa mitjana d'ocurrències per unitat considerada)

Nota: una variable de Poisson pot agafar qualsevol valor enter $k \geq 0$, encara que en la pràctica sols els que estan relativament a prop de λ tenen probabilitats rellevants

Model Poisson. Representació gràfica

EXEMPLE. Com es distribueix el número de correus *spam* rebuts al dia segons si el valor de la mitjana és 1.5 o 6?



Model Poisson. Ex. de càlcul de probabilitats

Sigui $X \sim P(\lambda = 2)$:

- **Probabilitat puntual.** Quina és la probabilitat de 3?
 - Amb fórmules $\rightarrow P(X = 3) = \frac{e^{-2} \cdot 2^3}{3!} = \mathbf{0.1804}$
 - Amb R $\rightarrow P(X = 3) = \text{dpois}(x = 3, \text{lambda} = 2) = \mathbf{0.1804470}$
- **Probabilitat acumulada.** Quina és la probabilitat de 3 o menys?
 - Amb fórmules $\rightarrow P(X \leq 3) = \frac{e^{-2} \cdot 2^0}{0!} + \dots + \frac{e^{-2} \cdot 2^3}{3!} = \mathbf{0.857}$
 - Amb R $\rightarrow P(X \leq 3) = \text{ppois}(q = 3, \text{lambda} = 2) = \mathbf{0.8571235}$
- **Quantils.** Quin és el valor tq la probabilitat de quedar per sota d'ell és almenys 0.95?
 - Amb fórmules $\rightarrow P(X \leq x_{0.95}) = 0.95 \rightarrow \text{Molt complicat!!!}$
 - Amb R $\rightarrow P(X \leq x_{0.95}) = 0.95 \rightarrow \text{qpois}(p = 0.95, \text{lambda} = 2) = \mathbf{5}$

Nota: dpois, ppois, qpois són funcions en R, Rstudio, i R online (<https://rdr.io/snippets/>)

Nota: La suma de VAD Poisson és també una VAD Poisson amb paràmetre λ igual a la suma dels paràmetres:

$$X \sim P(\lambda_1) \quad Y \sim P(\lambda_2) \quad \rightarrow \quad X+Y \sim P(\lambda_1 + \lambda_2)$$

A partir d'una VAD Poisson, podem definir altres VAD aplicant proporcionalment al paràmetre el canvi en l'interval:

$$X = \text{"...en interval } t" \sim P(\lambda) \quad \rightarrow \quad Y = \text{"...en interval } kt" \sim P(k\lambda)$$

TAULA resum de models de VAD

Distribució	Declaració	Domini	Esperança $E(X) = \mu_x$	Variància $V(X) = \sigma_x^2$
Bernoulli	Bern(p)	0, 1	p	p·q
Binomial	B(n,p)	0,1,...,n	n·p	n·p·q
Geomètrica	Geom(p)	1,2,3,...	1/p	q/p ²
Binomial negativa	BN(r,p)	r, r+1,...	r/p	q·r/p ²
Poisson	P(λ)	0, 1, 2,...	λ	λ

$$0 < p < 1$$

$$q = 1 - p$$

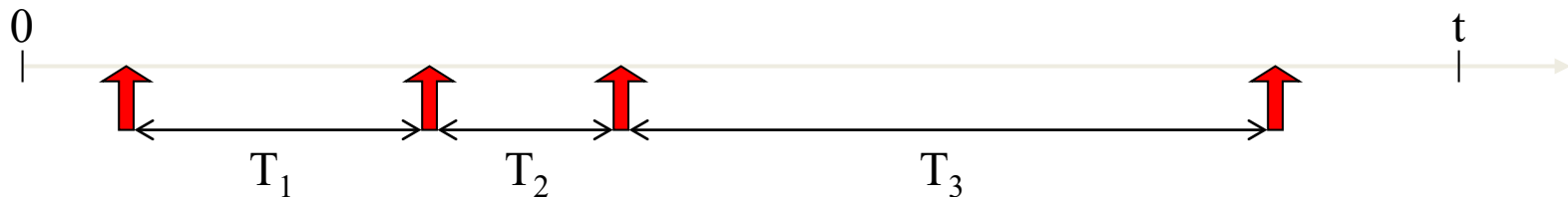
$$n \in \mathbb{N}$$

$$r \in \mathbb{N}$$

$$\lambda \in \mathbb{R}^+$$

Model Exponencial

- Definició:** Distribució del temps entre arribades (ocurrències) en un procés de Poisson. És a dir, si a l'interval $[0, t]$ les arribades al sistema (N_t) segueixen una distribució de Poisson, amb taxa $\lambda \cdot t$ (la taxa per unitat de temps és λ), llavors el temps entre dues arribades consecutives és una magnitud continua i indeterminista que es distribueix exponencialment. [Ex. d'aplicació: vida útil d'un component electrònic]



- Notació:** $X \sim \text{Exp}(\lambda)$ Notació de funcions en **R** pel model: `dexp()`, `pexp()`, `qexp()`
- Paràmetres:** λ (taxa d'aparició de l'esdeveniment per unitat de temps)
- Funció de densitat i de distribució:**

$$f_X(x) = \lambda \cdot e^{-\lambda x} \quad \text{amb } x > 0$$

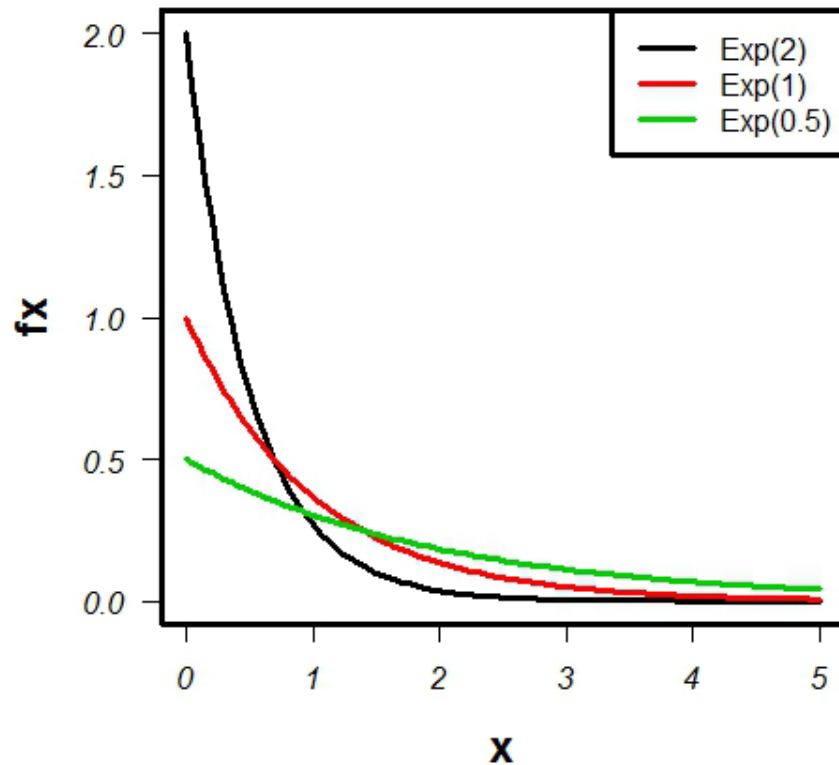
$$F_X(x) = 1 - e^{-\lambda x} \quad \text{amb } x > 0$$

- Indicadors:**
 - $E(X) = 1/\lambda$
 - $V(X) = 1/\lambda^2$

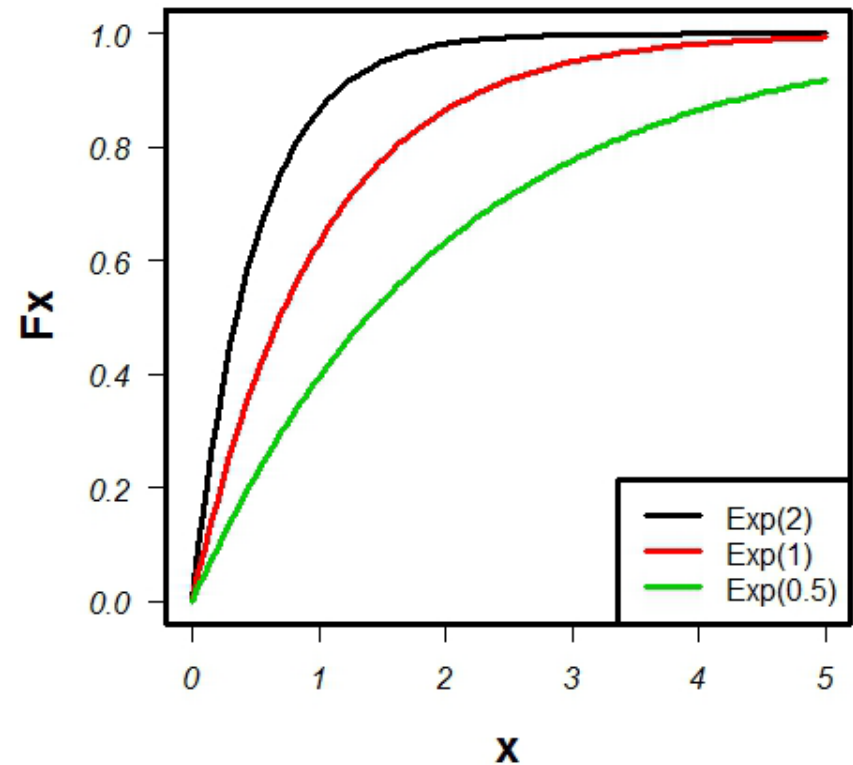
Model Exponencial. Representació gràfica

Ex: Com es distribueix el temps entre correus *spam* si rebo una mitjana de 2 per hora? I si rebo 1 per hora? I si rebo 1 cada dues hores?

Funció de densitat



Funció de distribució



Model Exponencial. Ex. de càlcul de probabilitats

Sigui $X \sim \text{Exp}(\lambda = 2)$:

- **Probabilitat puntual.** \rightarrow Recordeu que $P(X=x) = 0$ per qualsevol x , ja que és una VAC
- **Probabilitat acumulada.** Quina és la probabilitat de 2 o menys?
 - Amb fórmules $\rightarrow P(X \leq 2) = 1 - e^{-2 \cdot 2} = 1 - e^{-4} = \mathbf{0.9817}$
 - Amb R $\rightarrow P(X \leq 2) = \text{pexp}(q = 2, \text{rate} = 2) = \mathbf{0.9816844}$
- **Quantils.** Quin és el valor tq la probabilitat de quedar per sota d'ell és almenys 0.95?
 - Amb fórmules $\rightarrow P(X \leq x_{0.95}) = 0.95 \rightarrow 1 - e^{-2 \cdot x_{0.95}} = 0.95 \rightarrow x_{0.95} = \mathbf{1.4979}$
 - Amb R $\rightarrow P(X \leq x_{0.95}) = 0.95 \rightarrow \text{qexp}(p = 0.95, \text{rate} = 2) = \mathbf{1.497866}$

Nota: pexp, qexp són funcions en R, Rstudio, i R online (<https://rdr.io/snippets/>)

Nota: La distribució Exponencial té funció de distribució amb expressió analítica

Model Exponencial (algunes propietats)

- $f_x(x)$ no és $P(X=x)$ [$P(X=x) = 0$ per definició] $\rightarrow f_x(x)$ **no** és una probabilitat, a diferència de la $p_x(x)$ de les VAD

- Recordem que en una VAC:

$$P(a \leq X) = P(a < X) \quad \text{i} \quad P(a \leq X \leq b) = P(a < X < b) = F_x(b) - F_x(a)$$

Recordem també que en el model exponencial les probabilitats acumulades es calculen directament amb la funció de distribució de probabilitat: $F_x(x) = 1 - e^{-\lambda \cdot x}$

- **Propietat de Markov (o de NO memòria):** La distribució de probabilitat d'una variable aleatòria Exponencial no depèn del que hagi passat amb anterioritat al moment present:

$$P(T > t_1 | T > t_0) = P(T > t_1 - t_0) \quad \text{per } t_1 > t_0$$

Atenció: $P(T > t_1 | T > t_0) \neq P(T > t_1)$

- Ex: En el servidor de BBDD, en un instant donat fa 10'' que no arriben peticions. Que és més probable: (A) rebre en els 10'' següents, o (B) rebre 10'' després d'una arribada?

Solució: Igual

Model Uniforme (continu)

- **Definició:** VAC amb funció de densitat constant en un determinat rang [la probabilitat de pertànyer a un interval concret en aquest rang només depèn de la longitud de l'interval]
- **Notació:** $X \sim U(a, b)$ Notació de funcions en R pel model: `dunif()`, `punif()`, `qunif()`
- **Paràmetres:** a (valor mínim del rang de X), b (valor màxim del rang de X)
- **Funció de densitat i distribució:**

Constant!!! $\rightarrow f_X(x) = \frac{1}{b-a}$ amb $a < x < b$

$F_X(x) = 0$ si $x < a$
 $F_X(x) = 1$ si $x > b$ $\rightarrow F_X(x) = \frac{x-a}{b-a}$ amb $a < x < b$

- **Indicadors:**

- $E(X) = (b+a)/2$
- $V(X) = (b-a)^2/12$

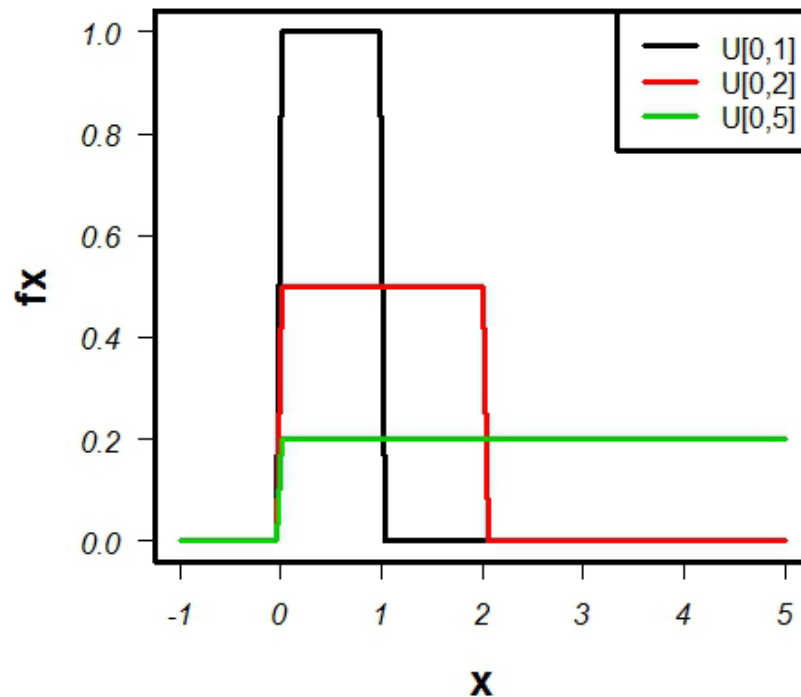
$$\begin{aligned}
 E(X) = \mu_X &= \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \\
 &= \frac{1}{b-a} \cdot \left[\frac{x^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}
 \end{aligned}$$

(intuïtivament ja es veu que la mitjana ha de ser el centre de l'interval)

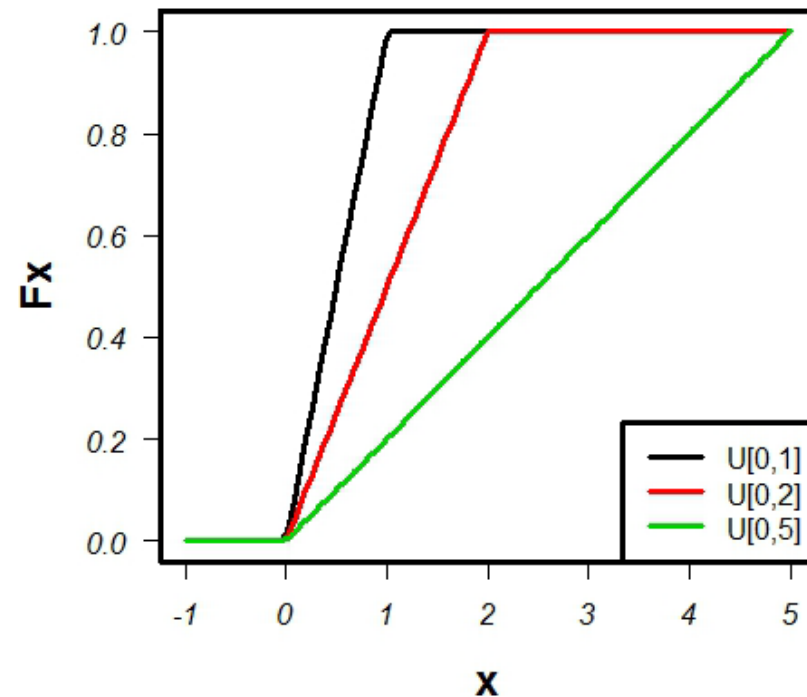
Model Uniforme. Representació gràfica

Ex: Com es distribueixen el nombres reals aleatoris entre 0 i 1? I entre 0 i 2? I entre 0 i 5?

Funció de densitat



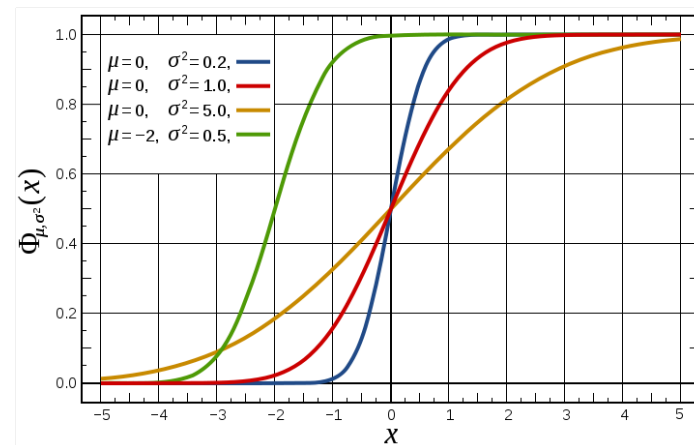
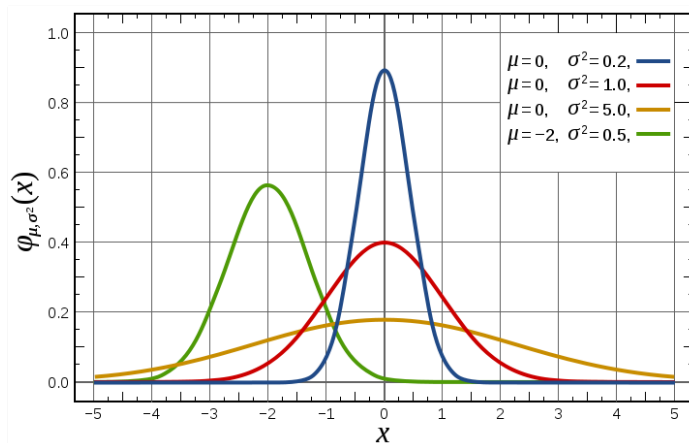
Funció de distribució



Model Normal (o de Gauss). Introducció

(Wikipedia.org) *Normal Distribution*:

- “the **normal** (or **Gaussian**) **distribution**, is a continuous probability distribution that is often used as a first approximation to describe real-valued random variables that tend to cluster around a single mean value”
- “the normal distribution is **commonly encountered in practice**, and is used throughout statistics, natural sciences, and social sciences”



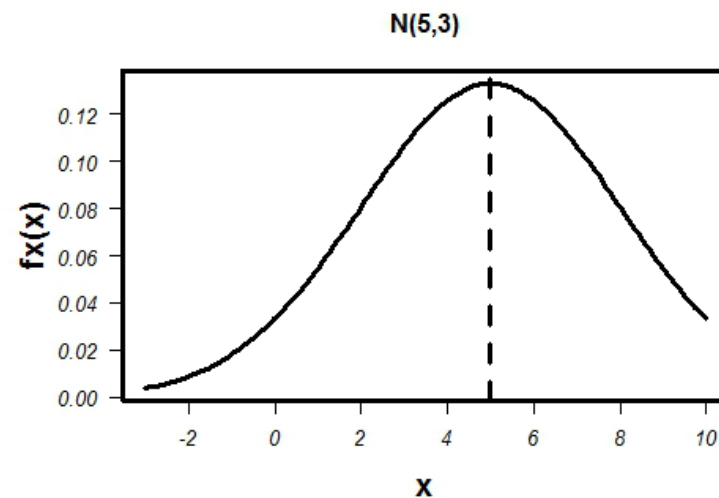
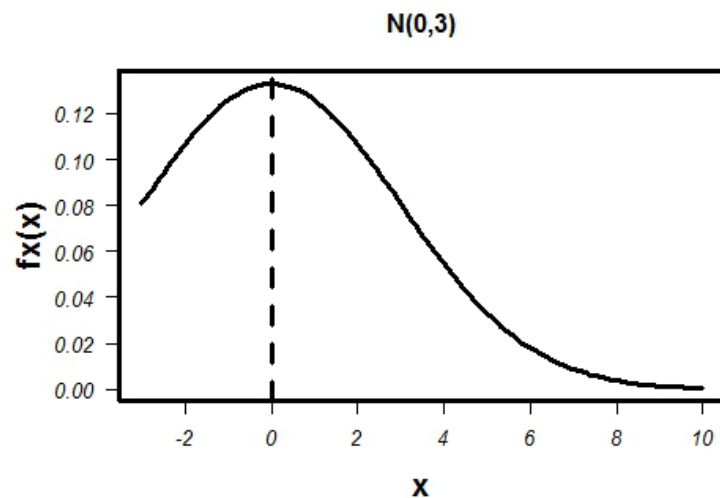
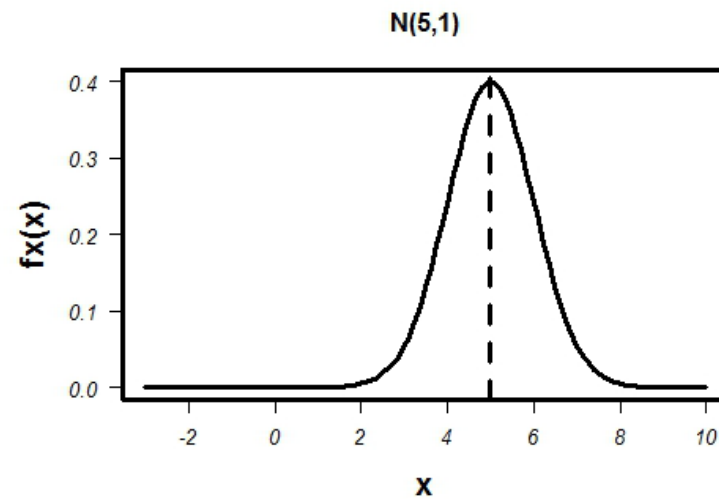
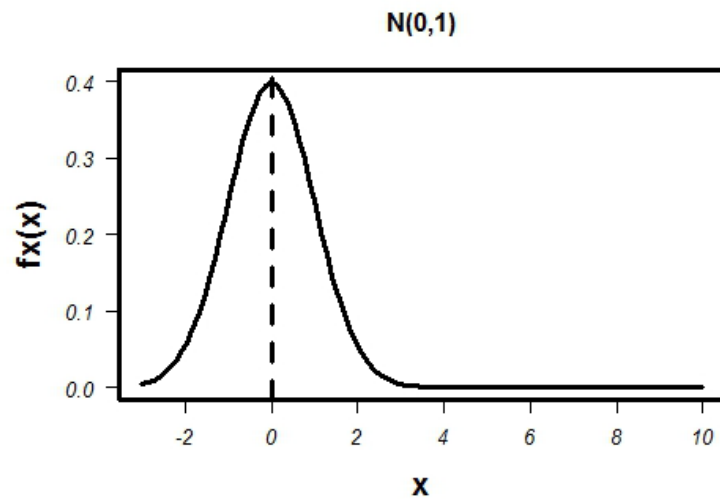
Model Normal

- **Definició:** Model que serveix per representar els valors provinents de múltiples fenòmens trobats en diferents disciplines científiques [Ex: alçades de persones, efecte d'un fàrmac, soroll en telecomunicacions...]
- **Notació:** $X \sim N(\mu, \sigma)$ Notació de funcions en R pel model: `dnorm()`, `pnorm()`, `qnorm()`
- **Paràmetres:** μ (esperança), σ (desviació estàndard) [vigilar si s'usa σ^2 i no σ com a paràmetre]
- **Funció de densitat:**
$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{amb } x \in R$$
- **La funció de distribució** no té expressió analítica explícita $\rightarrow R$
- **Indicadors:**
 - $E(X) = \mu$
 - $V(X) = \sigma^2$

Nota: la Normal amb paràmetres $\mu = 0$ i $\sigma=1$ s'anomena **Normal estàndard**

Model Normal. Representació gràfica

Ex: Com són les funcions de densitat de diferents Normals segons els valors de μ i σ ?



Model Normal. Exemple de càlcul de probabilitats

Sigui $X \sim N(\mu=0, \sigma=1)$:

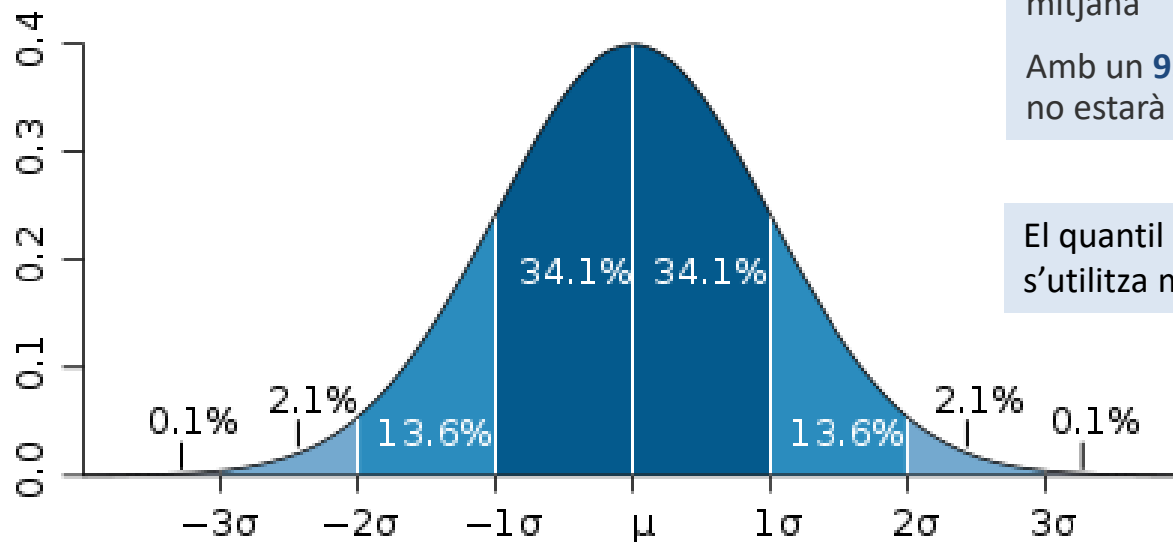
- **Probabilitat puntual.** \rightarrow Recordeu que $P(X=x) = 0$ per qualsevol x ja que és una VAC
- **Probabilitat acumulada.** Quina és la probabilitat de 2 o menys?
 - Amb fórmules \rightarrow **No es pot fer!!!**
 - Amb R $\rightarrow P(X \leq 2) = pnorm(q = 2, mean = 0, sd = 1) = \mathbf{0.9772499}$
($pnorm(2) = 0.9772499$)
- **Quantils.** Quin és el valor tq la probabilitat de quedar per sota d'ell és almenys 0.95?
 - Amb fórmules \rightarrow **No es pot fer!!!**
 - Amb R $\rightarrow P(X \leq x_{0.95}) = 0.95 \rightarrow qnorm(p = 0.95, mean = 0, sd = 1) = \mathbf{1.645}$
($qnorm(0.95) = 1.64$)

Nota: pnorm, qnorm són funcions en R, Rstudio, i R online (<https://rdr.io/snippets/>)

Nota: les anteriors funcions en R serveixen per a qualsevol Normal, però és habitual calcular-ho amb l'estàndard (**N(0,1)**) amb la transformació que veurem d'estandarització. De fet, mean=0 i sd=1 són els valors per defecte en les funcions

Model Normal. Propietats i quantils

- La funció de densitat $f(x)$ és simètrica respecte al punt $x = \mu$, que és a la vegada, la mitjana i la mediana de la distribució.
- Els punts d'inflexió es troben a una desviació estàndard de la μ ($x=\mu-\sigma$ i $x=\mu+\sigma$)
- Els quantils de la Normal estàndard $Z \sim N(0,1)$, normalment, es denoten amb z_p . El quantil z_p representa aquell valor tal que en una Normal estàndard té una probabilitat p de caure en l'interval $(-\infty, z_p]$



A la pràctica, X es concentra molt a prop de la mitjana

Amb un **95.4%** de probabilitat, un valor al atzar no estarà més lluny de **2σ** de la mitjana μ .

El quantil més emprat és el **$Z_{0.975} = 1.96$** ja que s'utilitza molt en inferència estadística

Model Normal. Estandardització

- La **combinació lineal de variables Normals** és Normal:
 - Sigui a i b , dos escalars i $X \sim N(\mu_X, \sigma_X) \rightarrow Y = a \cdot X + b \sim N(\mu_Y = a \cdot \mu_X + b, \sigma_Y = a \cdot \sigma_X)$
 - Sigui a i b , dos escalars, $X \sim N(\mu_X, \sigma_X)$ i $Y \sim N(\mu_Y, \sigma_Y) \rightarrow$
 $\rightarrow W = a \cdot X + b \cdot Y \sim N(\mu_W = a \cdot \mu_X + b \cdot \mu_Y, \sigma_W = \sqrt{a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2 \cdot a \cdot b \cdot \rho_{XY} \cdot \sigma_X \cdot \sigma_Y})$
- Aquesta propietat permet relacionar distribucions Normals a base de translacions i escalars. En particular, transformar a la Normal estàndard $Z \sim N(0,1)$, **estandarditzar**, permet buscar en les taules de Z , probabilitats de qualsevol Normal
- Amb $X \sim N(\mu, \sigma)$ i $Z \sim N(0, 1)$ podem relacionar:
 $Z = X/\sigma - \mu/\sigma = (X - \mu) / \sigma \sim N(0, 1)$ ($a = 1/\sigma$, $b = -\mu/\sigma$ són escalars). És a dir:

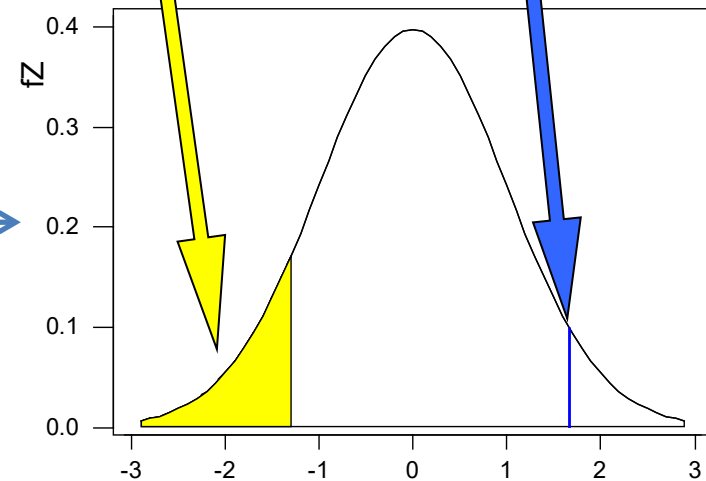
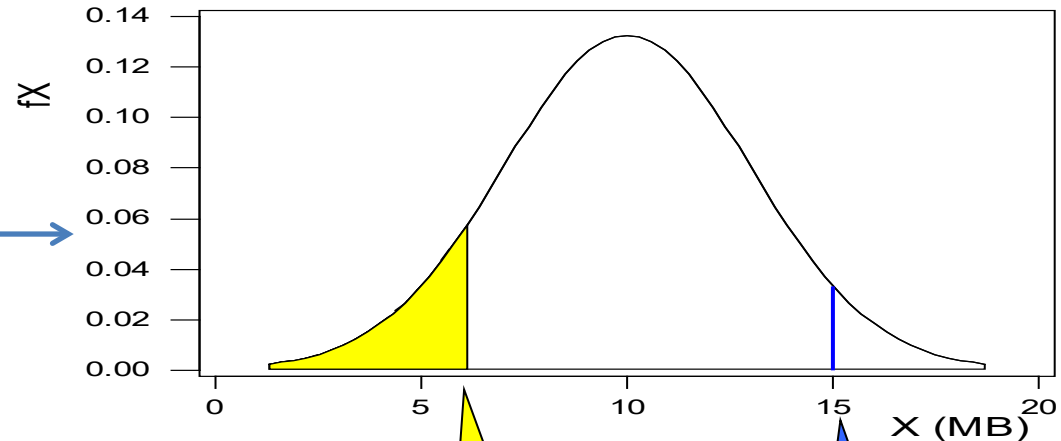
$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1) \rightarrow X = \mu + \sigma \cdot Z$$

Model Normal. Estandardització

Variable X: situació real (per exemple, MB d'un fitxer)

$$Z = \frac{X - \mu}{\sigma}$$

Variable Z: situació estandarditzada, sense unitats, centrada en 0, dispersió tipificada (igual a la unitat)

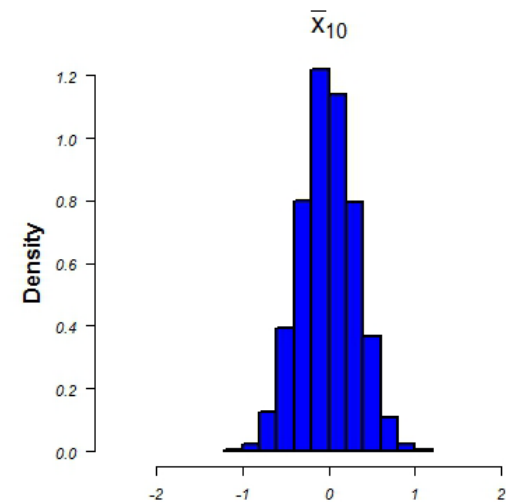
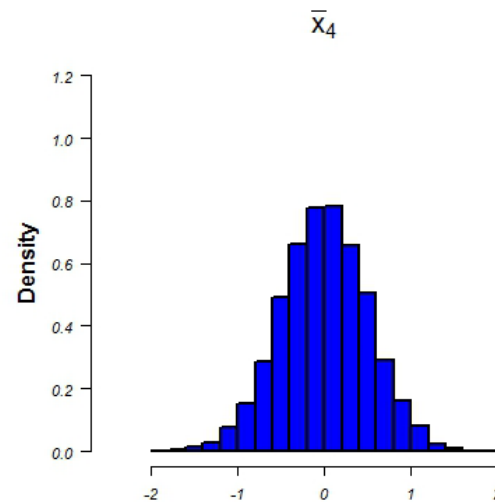
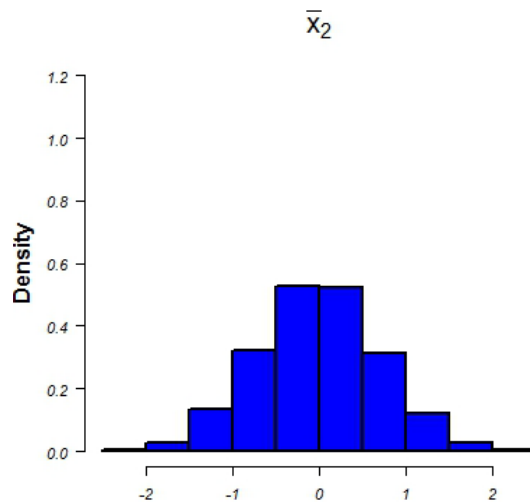


Distribució de la mitjana de v.a.

- Hem simulat $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ tal que X_i siguin i.i.d. Observem que:
 - tendeix a concentrar-se al voltant de μ quan n augmenta
 - tendeix a assemblar-se a una Normal a mesura que n es fa gran.

$$E(\bar{X}_n) = \frac{E(\sum X_i)}{n} = \frac{\sum E(X_i)}{n} = \frac{n\mu}{n} = \mu \quad V(\bar{X}_n) = \frac{V(\sum X_i)}{n^2} = \frac{\sum V(X_i)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

- Per qualsevol n , l'esperança de la mitjana és μ i la variància decreix amb n : amb una mostra gran, utilitzant la mitjana mostral ens aproximem més a μ .



Teorema Central del Límit (TCL)

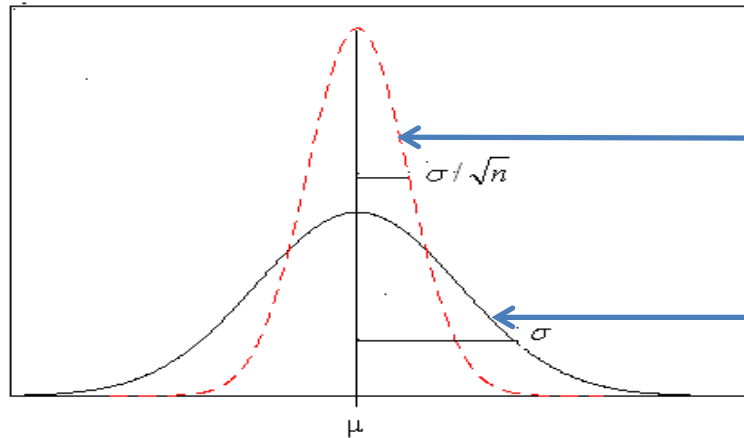
- Siguin X_1, X_2, \dots, X_n independents i idènticament distribuïdes amb esperança μ i desviació típica σ . Llavors:

$$S_n = \sum X_i \xrightarrow{n \text{ gran}} N(n\mu, \sigma\sqrt{n}) \Rightarrow \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{n \text{ gran}} N(0,1)$$

$$\bar{X}_n = \frac{\sum X_i}{n} \xrightarrow{n \text{ gran}} N(\mu, \sigma/\sqrt{n}) \Rightarrow \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{n \text{ gran}} N(0,1)$$

- És a dir, amb n gran, la funció de distribució de la variable Suma (S_n) i mitjana (\bar{X}_n) tendeix a una Normal amb uns determinats paràmetres **independentment de la distribució de les X_i !**
- Veure [app](#)

Teorema Central del Límit (TCL)



$$\bar{X}_n = \frac{\sum X_i}{n} \xrightarrow{n \text{ gran}} N(\mu, \sigma/\sqrt{n})$$

X_i

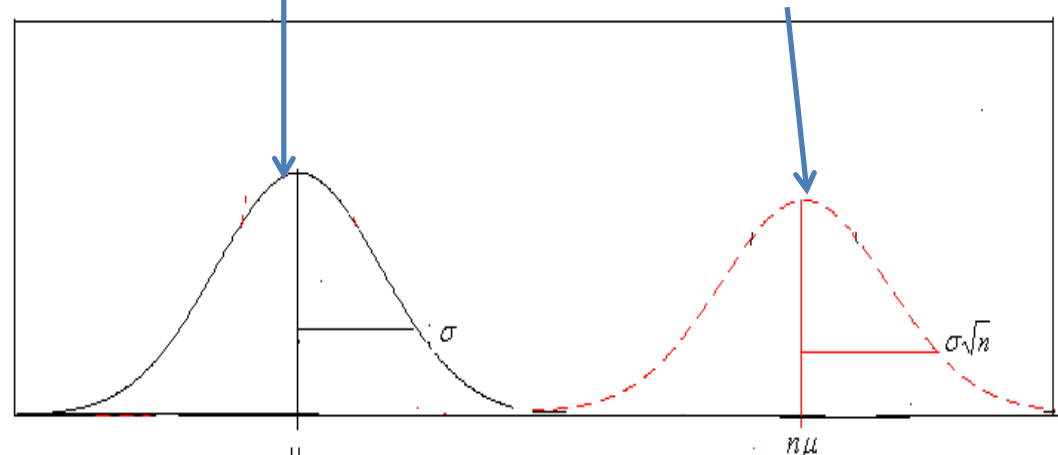
Els X_i no necessàriament han de ser Normals!!!!

Només han de complir:

$$E(X) = \mu$$

$$V(X) = \sigma^2$$

$$S_n = \sum X_i \xrightarrow{n \text{ gran}} N(n\mu, \sigma\sqrt{n})$$

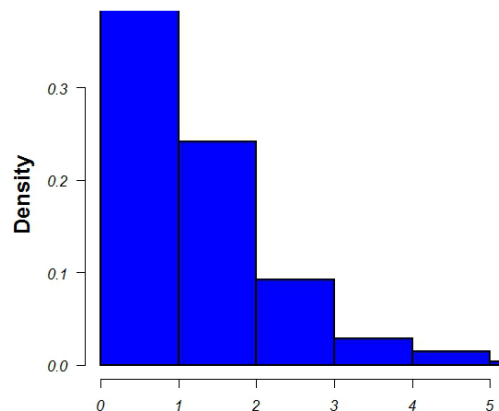


Teorema Central del Límit (TCL). "n"

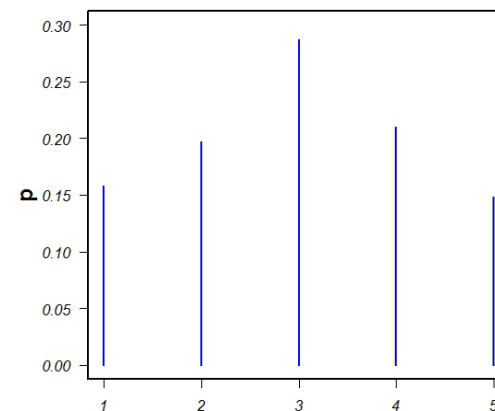
Quan n és *suficientment gran* per aplicar el TCL?

- Depèn de com sigui la distribució original i de que es desitgi calcular.
- La convergència a la normal és més lenta si la distribució de les X_i és **poc simètrica** o són **variables discretes** (especialment si només pot prendre pocs valors):

Distribució asimètrica



Distribució discreta



- Aplicacions del TCL: la normal aproxima bé certes distribucions. [Exemple: variable de Poisson, si λ és gran. La t-Student, i la χ^2 son derivades de la Normal que es veuran més endavant]

TCL. Relacions entre distribucions

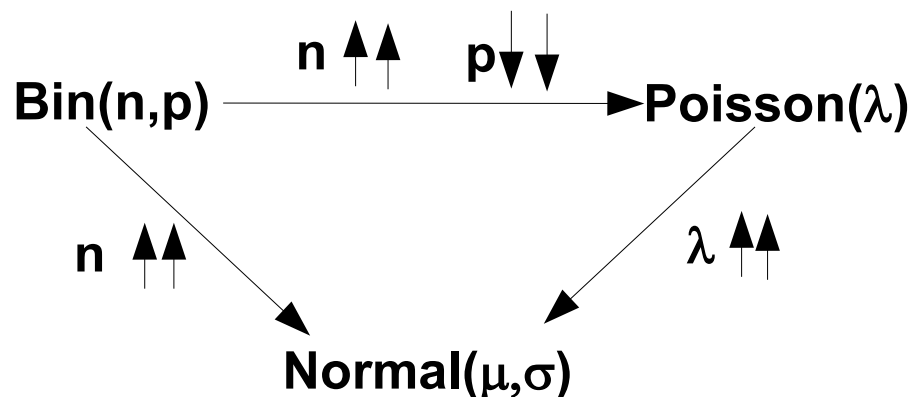
Una de les aplicacions pràctiques del TCL és que la distribució Normal es pot emprar com a aproximació d'altres distribucions:

- La **distribució Binomial** (suma de Bernoullis) amb paràmetres n i p es pot aproximar per una Normal quan n és gran i la p no massa extrema (ni molt a prop de 0 ni de 1). Llavors, els paràmetres de la Normal són

$$\mu = n \cdot p \quad \text{i} \quad \sigma^2 = n \cdot p \cdot (1-p)$$

- La **distribució de Poisson** amb paràmetre λ es pot aproximar per una Normal quan la λ és prou gran. Llavors els paràmetres de la Normal són:

$$\mu = \lambda \quad \text{i} \quad \sigma^2 = \lambda$$



Nota: la Binomial es pot aproximar a una Poisson quan la n és prou gran i la p prou petita. Llavors $\lambda=np$