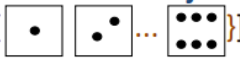
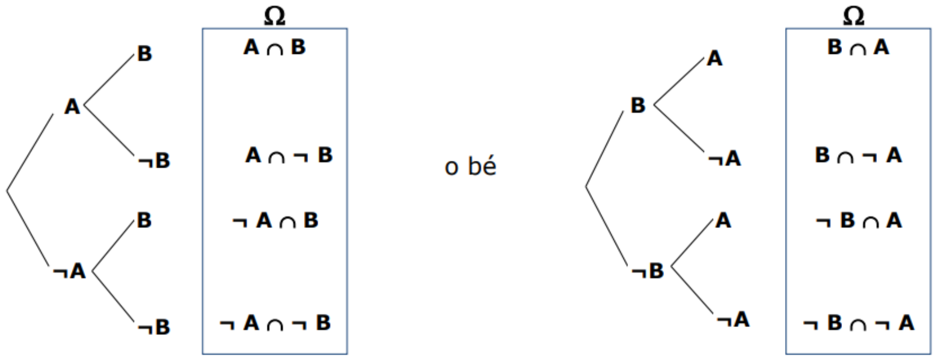
1.a. Experiència aleatòria. Definicions

- Els **fenòmens deterministes** porten a uns mateixos resultats a partir d'unes mateixes condicions inicials. [Ex: si poso la mà al foc, em cremaré]
- Els **fenòmens aleatoris** tenen una certa incertesa en el resultat d'una propera realització de l'experiència aleatòria. [Ex: Si llenço un dau, no sé quin número sortirà]
- Tota experiència aleatòria té associat un **conjunt de resultats possibles** (Ω ={w1,w2,...}) [Ex: en un dau, Ω ={  }]
- Qualsevol subconjunt de Ω és un **esdeveniment o succés** (A,B,...). [Ex: Ω (segur) o $\varnothing$ (impossible)]
- Una **partició** és un conjunt d'esdeveniments $A_i \neq \varnothing$, disjunts i que la seva unió és Ω. [Ex: en un dau, A_1 = parell ; A_2 = senar → $A_1 \cup A_2 = \Omega$ i $A_1 \cap A_2 = \varnothing$]

1.a. Experiència aleatòria. Representacions

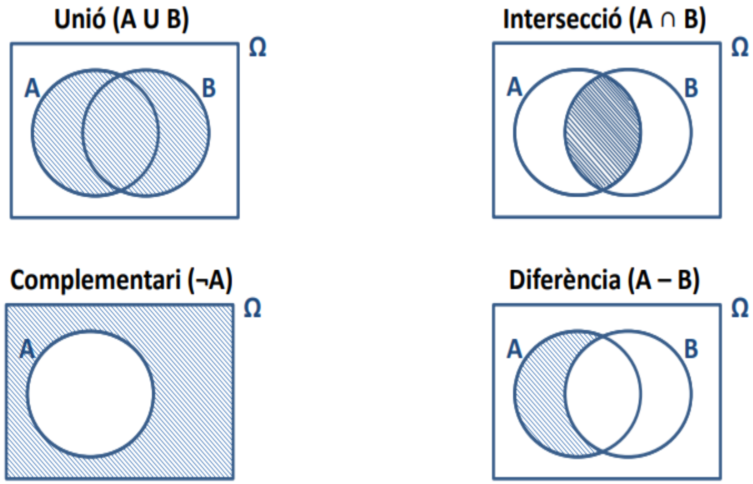
- Conjunts/diagrames de Venn

- Arbres



1.a. Experiència aleatòria. Operacions amb conjunts

Com que els esdeveniments són conjunts, totes les operacions dels conjunts es poden aplicar, i el resultat és un altre esdeveniment.



Definicions: Dos conjunts, A i B són **complementaris** (o formen una **partició**) si $A \cap B = \varnothing$ i $A \cup B = \Omega$
Dos conjunts, A i B són **disjunts** si $A \cap B = \varnothing$

1.b. Probabilitat. Definició i propietats

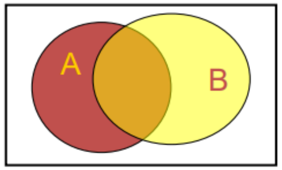
- Per quantificar la incertesa podem definir **una aplicació** que **a cada succés li fa correspondre un valor entre 0 i 1** que anomenem **probabilitat**
- Propietats per definició:
  - $0 \leq P(A) \leq 1$
  - $P(A_0 \cup A_1 \cup ... \cup A_n) = P(A_0) + P(A_1) + ... + P(A_n)$ si $A_i \cap A_j = \varnothing$ per $i \neq j$
  - $P(\Omega) = 1$
- Propietats deduïdes:
  - $P(\neg A)=1 - P(A)$
  - $P(\varnothing) = 0$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

NOTA: El cas particular en que la probabilitat s'obté de casos favorables/casos totals es presenta en condicions d'*equiprobabilitat* (tots els successos elementals tenen la mateixa probabilitat). No es pot generalitzar a qualsevol experiència! [Ex: es podria aplicar a un llançament d'una moneda però no als possibles resultats d'una travessa]

1.b. Probabilitat condicionada (Veure aquest *vídeo*)

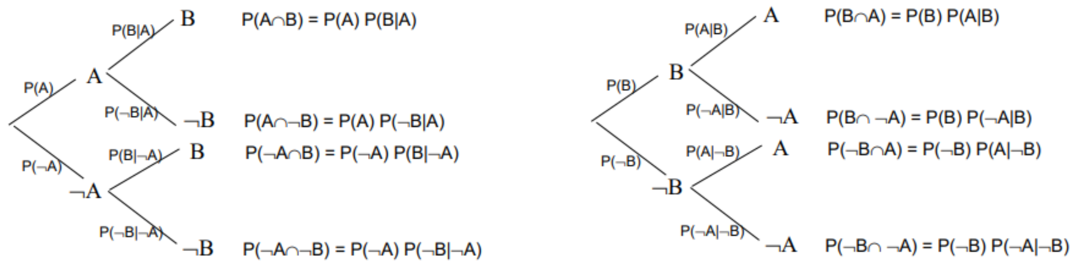
- Si l'expectativa d'un esdeveniment es representa en funció d'un altre parlem de **probabilitat condicionada P(A|B)** (es llegeix com a "probabilitat d'observar A tenint en compte que s'ha realitzat B" o "probabilitat de A condicionada per B") [Ex: A="Ploure"/B="Estar ennuvolat"]
- Per definició, si P(B)>0, llavors **P(A|B)** és:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$



- A la pràctica, **condicionar per B significa reduir a B el conjunt de resultats observables**, i les probabilitats han de recalcular-se
- En considerar P(A|B), **cada esdeveniment juga un paper diferent**: A és incert però B és conegut
- En general, **P(A|B) ≠ P(B|A) ≠ P(A ∩ B)**
[Ex: A="Fumar"/B="Tenir càncer de pulmó". La probabilitat de ser fumador si tens càncer de pulmó és més alta que no la inversa]

1.b. Probabilitat condicionada. Representació



probabilitats conjuntes

	B	¬B	
A	P(A∩B)	P(A∩¬B)	P(A)
¬A	P(¬A∩B)	P(¬A∩¬B)	P(¬A)
	P(B)	P(¬B)	1

probabilitats marginals

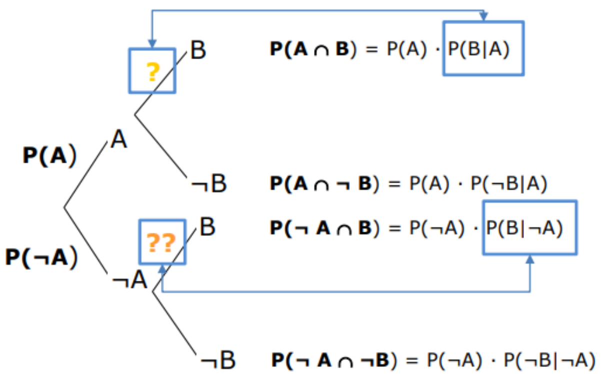
	B	¬B
A	$P(A\mid B)$	$P(A\mid \neg B)$
¬A	$P(\neg A\mid B)$	$P(\neg A\mid \neg B)$
	1	1

probabilitats condicionades

	B	¬B	
A	$P(B\mid A)$	$P(\neg B\mid A)$	1
¬A	$P(B\mid \neg A)$	$P(\neg B\mid \neg A)$	1

1.b. Probabilitat condicionada. Representació

Els arbres d'esdeveniments incorporen probabilitats condicionades.



La branca que va de 'A' a 'B' porta una probabilitat condicionada P(B|A) (?): estem suposant que 'A' ha passat.

P(A∩B) es pot obtener com el producte de les probabilitats en el camí des del node arrel fins al node A∩B:

P(A∩B) = P(A) P(B|A)

1.b. Probabilitat a posteriori. Fórmula de Bayes

A partir de la definició de probabilitat condicionada:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

i de la probabilitat de la intersecció aïllada de la condicionada complementària:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \quad \rightarrow \quad P(B \cap A) = P(A \cap B) = P(B|A) \cdot P(A)$$

es dedueix la fórmula de Bayes:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

que, coneixent P(A) i P(B), permet passar de P(A|B) a P(B|A)

i viceversa (usualment en l'enunciat del cas, les probabilitats condicionades en un sentit són conegudes i interessa calcular les altres condicionades). [Exemple: si conec la probabilitat de pluja si està ennuvolat i vull conèixer la probabilitat d'ennuvolat si plou]

1.b. Probabilitat a posteriori. Llei Probabilitats totals

Podem calcular la probabilitat d'un succés B_k a partir de les probabilitats de les interseccions d'aquest amb una partició $A_1, A_2, \ldots A_j$ de W:

$$P(B_k) = P(B_k \cap A_1) + P(B_k \cap A_2) + \ldots + P(B_k \cap A_j) =$$
$$= P(B_k|A_1)\cdot P(A_1) + P(B_k|A_2)\cdot P(A_2) + \ldots + P(B_k|A_j)\cdot P(A_j)$$

Llei de probabilitats totals (LPT). S'aplica quan disposem d'una partició, i la probabilitat del succés d'interès és senzilla d'obtenir si està condicionat per un element qualsevol de la partició.

Combinant la fórmula de Bayes amb la llei de probabilitats totals (i una partició $\{A_i\}$ adequada) s'obté el teorema de Bayes:

$$P(A_i \mid B_k) = \frac{P(B_k \mid A_i)\cdot P(A_i)}{\sum_j P(B_k \mid A_j)\cdot P(A_j)}$$

1.c. Independència i probabilitat condicionada

- Si A i B són independents

Independents

$$\mathbf{P(A|B) = P(A \cap B)/P(B) = P(A)\cdot P(B)/P(B) = P(A)}$$

- Que passi B no canvia l'expectativa de A i viceversa, que passi A no canvia l'expectativa de B. Si A i B són independents, llavors:

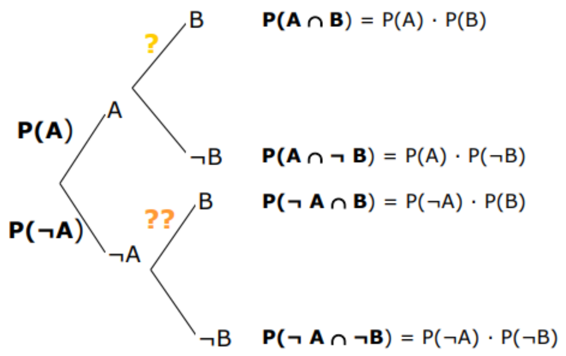
$$\mathbf{P(B|A) = P(B|\neg A) = P(B)}$$

- La idea d'esdeveniments independents està lligada a la de la *informació* que un aporta sobre l'altre: A i B són independents quan la probabilitat d'A és la mateixa, indiferentment del que passi amb B.

1.c. Independència

Independència aplicat a 2 (o més) esdeveniments és defineix com:

A i B són independents $\Leftrightarrow$ $\boldsymbol{P(A \cap B) = P(A)\cdot P(B)}$



Quan A i B són independents,

a) P(B | A) = "?" és P(B)

b) P(B | ¬A) = "??" és P(B)

c) Per tant, P(B | A) = P(B | ¬A) = P(B)

Exercici Aeroport (probs. facturació i embarcament)

Per estudiar l'eficiència en un aeroport, una primera aproximació ens porta a estudiar les probabilitats d'haver-se d'esperar a l'hora de facturar i a l'hora d'embarcar. Considerem el cas d'un aeroport on s'ha comprovat que per a un viatger que arriba, la probabilitat de trobar cua a facturació és 0.4; i de trobar cua a l'embarcament és 0.6 si va trobar cua a facturació, i 0.2 si no en va trobar.

Calculeu les següents probabilitats:

a) de trobar cua a la facturació i a l'embarcament

b) de trobar cua a l'embarcament

c) de trobar cua a l'embarcament si s'ha trobat cua a la facturació

d) d'haver trobat cua a la facturació si no ha trobat cua a l'embarcament

A- 0,4*0,6 B- 0,36 C- 0,6 D- 0,25

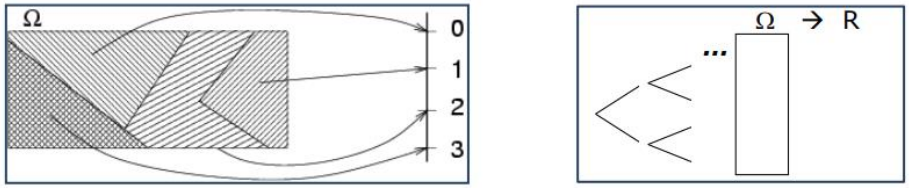
2.a. Variable aleatòria. Definició

- La majoria d'experiències aleatòries porten a resultats interpretables com un número. Ens interessa l'estudi de l'experiment des del **punt de vista numèric**.
- Una **variable aleatòria** X és una aplicació entre el conjunt Ω i la recta real:

$$X: \ \Omega \longrightarrow \mathbf{R}$$

Notació: Les variables aleatòries les denotarem amb un símbol tal com *X*, *Y*, *Z*, ...(lletra llatina majúscula). I els seus i possibles valors amb minúscula x_i, y_i, z_i

- Una variable X indueix una partició de Ω amb els valors x_i que adopta:



- Les probabilitats definides en Ω es transfereixen als valors de la VA *X*, definint unes **funcions de probabilitat** (o de densitat) i de **distribució de probabilitat**:

$$\Omega \xrightarrow{X} \mathbf{R} \xrightarrow{\text{prob.}} [0,1]$$

2.a. Variable aleatòria. Funció de prob. i distr. en VAD

Les funcions de **probabilitat i distribució** es defineixen segons si són VAC o VAD

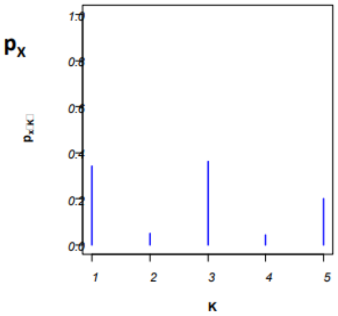
- La **funció de probabilitat** (p_X) en una VA DISCRETA (VAD) defineix la probabilitat puntual de cada un dels possibles valors k

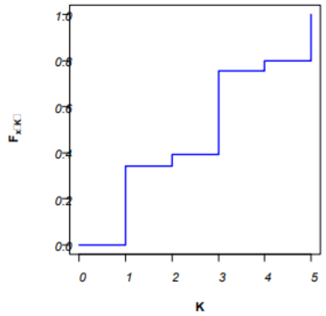
$$p_X(k) = P(X = k) \text{ (complint } \textstyle\sum_k p_X(k) = 1)$$

- La **funció de distribució** (F_X) de probabilitat en una VAD defineix la probabilitat acumulada, és a dir:

$$F_{X(k)} = P(X \leq k) = \sum_{j \leq k} p_X(j)$$

Funció de probabilitat p_X





Funció de distribució de probabilitat F_X

app

2.a. Variable aleatòria : discreta (VAD) i contínua (VAC)

Distingim dos tipus de VA:

- Si el conjunt de valors que poden agafar és <u>enumerable</u> (per exemple: un conjunt de valors enters (0..n), el conjunt dels naturals N,...), la VA és **Discreta (VAD)**

  Per exemple, en l'experiència de llençar una moneda 3 vegades:

  - VAD X = "cara o creu en l'últim llançament" (possibles valors: 0,1; probabilitats: ½, ½)
  - VAD Y = "número de cares" (possibles valors: 0,1,2,3; probabilitats: ?)

  O en un servidor durant un període de temps:

  - VAD Z = "número de caigudes del sistema" (possibles valors: 0,1,2,3,4,5,6,...; probabilitats:?)

- Si agafa valors d'un conjunt no discret (és a dir, normalment la recta real o un segment d'ella) la VA és **Contínua (VAC)**
  - VAC X = "temps entre caigudes del sistema" (possibles valors: R^+)
  - En general, les VAC són mesures físiques de temps, longituds,....

2.a. Variable aleatòria. Funció de prob. i distr. en VAC

- La **funció de densitat** de probabilitat (f_X) d'una VA CONTINUA és la funció que cobreix l'espai on està definida la variable, complint:

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

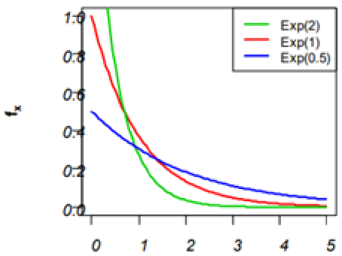
Observem que $f_X(k)$ és el valor de la funció en k, però no és una probabilitat puntual, $f_X(k) \neq P(X=k)$, i P(X=k) val 0

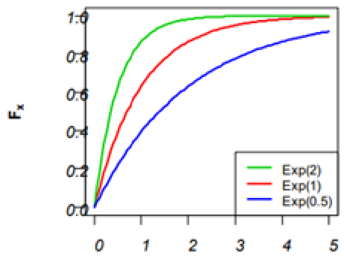
- La **funció de distribució** de probabilitat (F_X) d'una VA CONTINUA defineix la probabilitat acumulada, és a dir:

$$F_X(k) = P(X \leq k) = \int_{-\infty}^{k} f_X(x)\,dx$$

Observem que $f_X(x) = \frac{dF_X(x)}{dx}$

Funció de densitat f_X





Funció de distribució F_X

2.a. Variable aleatòria. Funció de prob. i distr. en VAC

És a dir, en el cas VA CONTINUA, una **funció positiva**, $f_X(x) \geq 0$, que compleix:

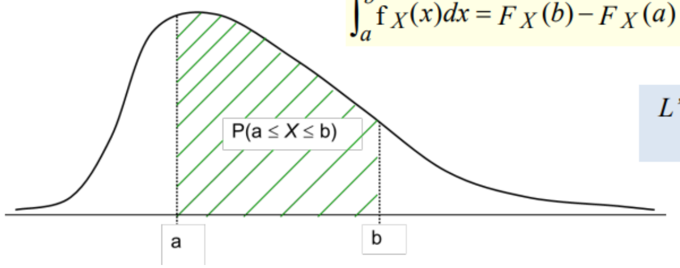
$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

és una <u>funció de densitat</u> vàlida, és a dir, caracteritza la variable. La <u>funció de distribució</u> s'obté amb:

$$F_X(u) = \int_{-\infty}^{u} f_X(x)dx$$

Tota l'àrea sota la funció de densitat és sempre igual a 1. En particular, l'àrea que hi ha sota la funció de densitat entre els límits *a* per l'esquerra i *b* per la dreta és:

$$\int_a^b f_X(x)dx = F_X(b) - F_X(a)$$



L'àrea sota f_X equival a una probabilitat!

app

2.b. Probabilitats en variables aleatòries

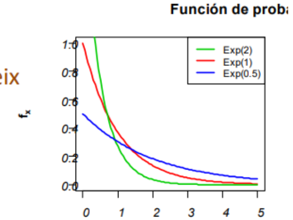
Siguin *X* i *Y* variables aleatòries, i siguin *k*, *a* i *b* escalars:

- **Probabilitats en VAD**
  - $P(X = k) = p_X(k)$
  - $P(X \leq k) = F_X(k) = \sum_j^k p_X(j)$
  - $P(X < k) = P(X \leq k\text{-}1) = F_X(k\text{-}1)$
  - $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$
  - $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F_X(b) - F_X(a\text{-}1)$
  - $P(a < X < b) = P(X < b) - P(X \leq a) = F_X(b\text{-}1) - F_X(a)$

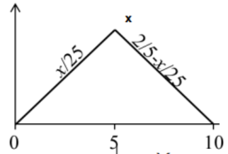
  Quan el valor previ a k és k-1
- **Probabilitats en VAC**
  - $P(X = k) = 0 \quad (\neq f_X(k))$
  - $P(X \leq k) = F_X(k)$
  - $P(X<k) = P(X \leq k) = F_X(k)$
  - $P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b) = F_X(b) - F_X(a)$
- **Probabilitats condicionades i independència**
  - $P(X \leq a \mid X \leq b) = P(X \leq a \cap X \leq b) / P(X \leq b)$
  - If X,Y are independent: $P(X \leq a \cap X \leq b) = P(X \leq a) \cdot P(X \leq b)$ and $P(X \leq a \mid X \leq b) = P(X \leq a)$)

2.a. Variable aleatòria. Exemples de VAC

Exemple 1. La "vida útil d'un transistor en anys" segueix una distribució que decreix exponencialment.

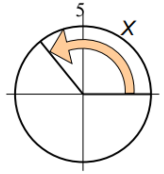


Exemple 2. "L'esforç requerit per desenvolupar un projecte" es pot mesurar en persones/mes, per exemple segons la figura.



Exemple 3. "L'angle X senyalat per la agulla d'una ruleta al parar-se", $0 \leq X \leq 2\pi$.

$F_X(k) = P(X \leq k) = k/(2\pi)$.



a) Escriure l'expressió analítica de la funció de densitat

$$f(x) = \begin{cases} \frac{x}{25} & 0 < x < 5 \\ \frac{2}{5} - \frac{x}{25} & 5 < x < 10 \\ 0 & altrament \end{cases}$$

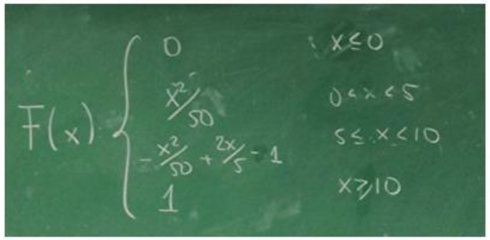
b) Comprovar que es una funció de densitat

$$\int_0^5 \frac{x}{25}dx + \int_5^{10} \left(\frac{2}{5} - \frac{x}{25}\right)dx = 1$$

c) Trobar l'expressió analítica de la funció de distribució.

$$\int_0^x \frac{k}{25}dk = \frac{x^2}{50}$$

$$\int_0^5 \frac{x}{25}dx + \int_5^x \left(\frac{2}{5} - \frac{k}{25}\right)dk = \frac{1}{2}$$



d) Calcular el percentil 90

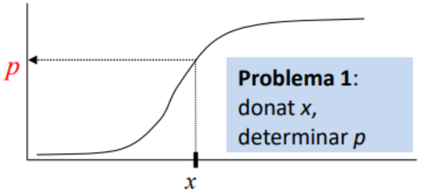
$P(x <= X\alpha) = 0{,}9$

$$-\frac{x^2}{50} + \frac{2}{5}x - 1 = 0{,}9$$

$$x = 7{,}76$$

2.b. Quantils

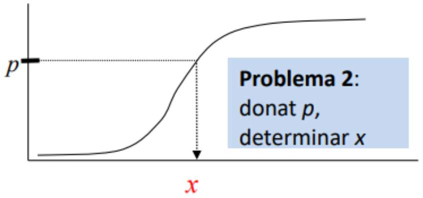
- Sigui X una variable aleatòria, i α un valor real $(0 \leq \alpha \leq 1)$ diem que x_α és el **quantil α** de X si es compleix: $F_X(x_\alpha) = \alpha$
- Calcular un quantil és el problema invers al càlcul de **probabilitats acumulades**. La funció inversa de la funció de distribució ens retorna x_α
- En el cas de VAC, és habitual plantejar problemes en els dos sentits:



Donat ***x*** calcular la probabilitat ***p*** tq:

$$p = F_X(x) = P(X \leq x)$$

Exemple: si els llits dels hotels mesuren 200 cm, quina proporció de congressistes poden dormir ben estirats?



Donada una probabilitat ***p*** calcular ***x*** tq:

$$x = F_X^{-1}(p)$$

Exemple: si desitgem que pugin dormir ben estirats el 98% dels congressistes, quina longitud han de tenir els llits?

2.c. Indicadors en variables aleatòries

- <u>Indicadors en una variable aleatòria</u>:
  - De <u>tendència central</u>, usem el **valor esperat o esperança**
    - **Notació: E(X) o μ_X**
  - De <u>dispersió</u>, usem la **variància** o la seva arrel quadrada, la **desviació estàndard, o *tipus***
    - **Notació per la variància: V(X) o σ_X^2**
    - **Notació per la desviació estàndard: σ_X**
- <u>Indicadors en una mostra</u>: Una mostra de valors expressa amb *n* observacions la variabilitat d'una experiència; si volem resumir aquestes dades utilitzarem la **mitjana mostral** $\bar{x}$ i la **desviació estàndard mostral** s_x [tal com es veu a l'Estadística descriptiva]

Propiedades y métodos de calcular

- <u>En la suma/resta</u>: La integral de la suma es la suma de las integrales: $\int (u+v)\,dx = \int u\,dx + \int v\,dx$

2.c. Indicadors en variables aleatòries

- **Esperança de X**

$$\textbf{VAD} \rightarrow E(X) = \mu_X = \sum_{\forall k} (k \cdot p_X(k))$$

$$\textbf{VAC} \rightarrow E(X) = \mu_X = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx$$

Si existe $\int_{-\infty}^{+\infty} f(x)\,dx$ entonces su valor coincide con

$$\lim_{b\to\infty} \int_{-b}^{b} f(x)\,dx \,,$$

- **Variància de X**

$$\textbf{VAD} \rightarrow V(X) = \sigma_X^2 = \sum_{\forall k} \left[(k - E(X))^2 \cdot p_X(k) \right] \rightarrow \sigma_X = \sqrt{\sigma_X^2} = \sqrt{V(X)}$$

$$\textbf{VAC} \rightarrow V(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} (x - E(X))^2 \cdot f_X(x) dx \rightarrow \sigma_X = \sqrt{\sigma_X^2} = \sqrt{V(X)}$$

- **Relació entre Esperança i Variància (en VAD i VAC):**

$$V(X) = E\left[(X - E(X))^2 \right] = E(X^2) - E(X)^2$$

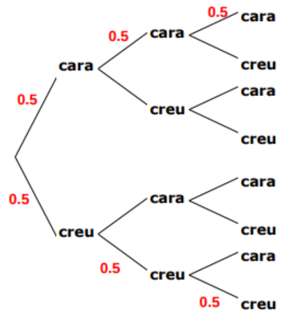
Inmediatas	*Cuasi inmediatas (con funciones)*
$\int x^n dx = \frac{x^{n+1}}{n+1} + k$ se suma 1 al expo y se divide por lo mismo	$\int f^n(x) \cdot f'(x) dx = \frac{f^{n+1}(x)}{n+1} + k$
$\int \frac{1}{x} dx = \ln\|x\| + k$	$\int \frac{f'(x)}{f(x)} dx = \ln\|f(x)\| + k$
$\int e^x dx = e^x + k$ se queda igual	$\int e^{f(x)} f'(x) dx = e^{f(x)} + k$
$\int a^x dx = \frac{a^x}{\ln a} + k$	$\int a^{f(x)} f'(x) dx = \frac{a^{f(x)}}{\ln a} + k$
$\int \text{sen}x dx = -\cos x + k$	$\int \text{sen}(f(x)) f'(x) dx = -\cos(f(x)) + k$
$\int \cos x dx = \text{sen}x + k$	$\int \cos(f(x)) f'(x) dx = \text{sen}(f(x)) + k$
$\int \frac{1}{\cos^2 x} dx = \int (1 + \text{tg}^2 x) dx = \text{tg}x + k$	$\int \frac{f'(x)}{\cos^2(f(x))} dx = \int (1 + \text{tg}^2(f(x))) f'(x) dx = \text{tg}(f(x)) + k$
$\int \frac{1}{\text{sen}^2 x} dx = \int (1 + \text{cotg}^2 x) dx = -\text{ctg}x + k$	$\int \frac{f'(x)}{\text{sen}^2(f(x))} dx = \int (1 + \text{ctg}^2(f(x))) f'(x) dx = -\text{ctg}(f(x)) + k$
$\int \frac{1}{\sqrt{1-x^2}} dx = \text{arcsen}x + k$	$\int \frac{f'(x)}{\sqrt{1-f^2(x)}} dx = \text{arcsen}(f(x)) + k$
$\int \frac{1}{1+x^2} dx = \text{arctg}x + k$	$\int \frac{f'(x)}{1+f^2(x)} dx = \text{arctg}(f(x)) + k$

Exemple Moneda. Variable aleatòria discreta

Estudiarem l'experiència aleatòria de llençar una moneda equilibrada tres vegades. Calcularem:

- P(A) sent A = "obtenir 2 cares"
- P(B) sent B = "obtenir almenys 2 cares"

Al bloc 1 s'han calculat les probabilitats a partir de l'arbre i dels successos; ara ho calcularem **definint la variable aleatòria X = "número de cares"**



(Cara,Cara,Cara) = w1	P(w1) = 1/8	X = 3
(Cara,Cara,Creu) = w2	P(w2) = 1/8	X = 2
(Cara,Creu,Cara) = w3	P(w3) = 1/8	X = 2
(Cara,Creu,Creu) = w4	P(w4) = 1/8	X = 1
(Creu,Cara,Cara) = w5	P(w5) = 1/8	X = 2
(Creu,Cara,Creu) = w6	P(w6) = 1/8	X = 1
(Creu,Creu,Cara) = w7	P(w7) = 1/8	X = 1
(Creu,Creu,Creu) = w8	P(w8) = 1/8	X = 0

X / Px(x)

0 / 1/8

1 / 3/8

2 / 3/8

3 / 1/8

E(x)= 0* 1/8 + 1* 3/8 + 2*3/8 + 3* 1/8 = **1,5**

E(x²)= 0²* 1/8 + 1²* 3/8 + 2²*3/8 + 3²* 1/8 = **3**

Var(x) = 3- 1,5² = **0,75**

Exercici Aeroport (VAD i VAC de viatgers i temps)

D'una altra banda, quan s'ha estudiat el temps que un viatger roman en el taulell de facturació, s'ha trobat que la següent funció:

$$f_X(k) = 0.2 \cdot e^{-0.2k} \quad \text{per } x > 0$$

és un model adequat per a representar la variable aleatòria T de "temps (en minuts)".

Calculeu la probabilitat que el temps sigui 7 minuts;
menys de 7 minuts;
més de 7 minuts;
entre 7 i 8 minuts

Com podem saber quin és el temps que, en mitjana*, un viatger roman en facturació?

* Nota: la paraula "mitjana" tant pot referir-se a l'esperança com a la mitjana mostral comú. En quin contexte s'està aplicant aquí?

$$P(T=7) = 0$$

$$Ft(t) = \int_0^t 0,2 \cdot e^{-0,2k} dk = 1 - e^{-0,2t}$$

$$P(T < 7) = Ft(7) = 0,7534$$

$$P(T > 7) = 1 - Ft(7) = 0,2466$$

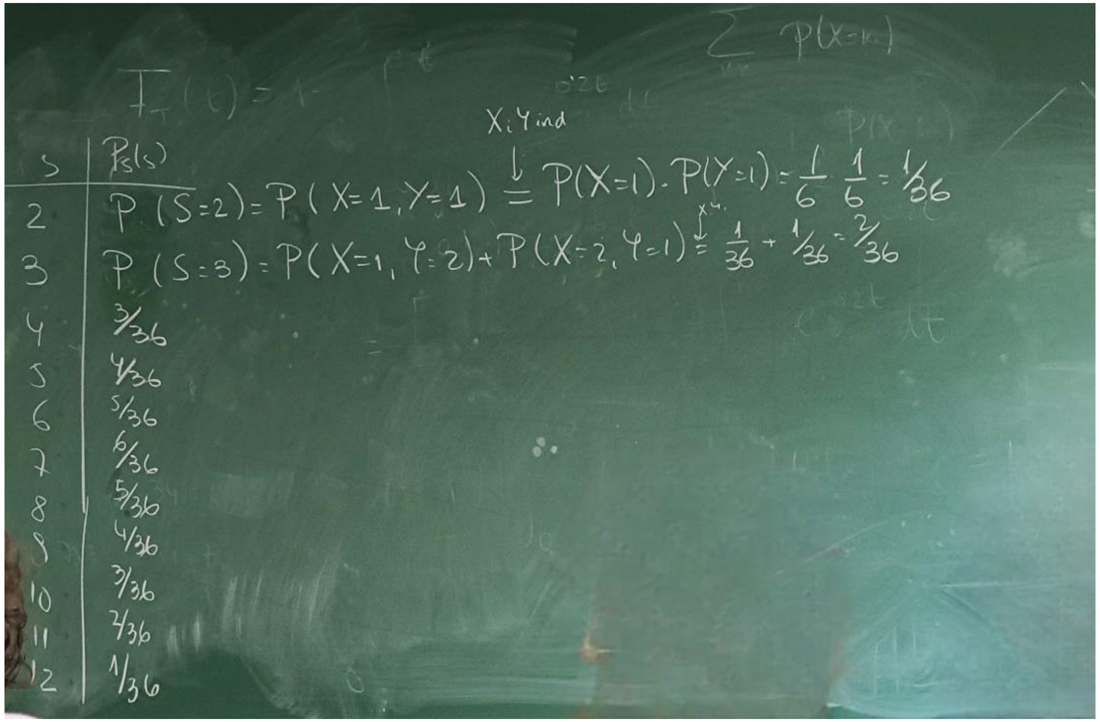
$$P(7 < T < 8) = Ft(8) - Ft(7)$$

$$E(t) = \int_0^{+\infty} t \cdot 0,2 \cdot e^{-0,2t} dt = 5$$

2.c. Indicadors en variables aleatòries. Propietats

Siguin *X* i *Y* dues variables aleatòries, i *a* i *b* dos escalars

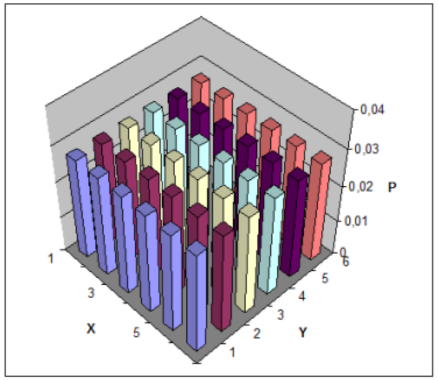
Propietats de l'Esperança	Propietats de la Variància
$E(a+X) = a + E(X)$	$V(a+X) = V(X)$
$E(bX) = b \cdot E(X)$	$V(bX) = b^2 \cdot V(X)$
$E(a+bX) = a + b\, E(X)$	$V(a+bX) = b^2 \cdot V(X)$
$E(X+Y) = E(X) + E(Y)$	$V(X \pm Y) = V(X) + V(Y)$ si X, Y són independents
$E(X \cdot Y) = E(X) \cdot E(Y)$ si X, Y són independents	$V(X \cdot Y) = ?$



2.d. Parell de variables. Parell de VAD

- Quan d'una experiència s'obtenen dues variables *X* i *Y*, quines relacions aleatòries es duen a terme entre elles?
- Per exemple, llancem dues vegades un dau equilibrat, i anomenem *X* al "primer resultat" i *Y* al "segon resultat".
- Raonablement, els dos llançaments són independents, llavors:

  $P(X=x \cap Y=y) = P(X=x) \cdot P(Y=y) = 1/6 \cdot 1/6 = 1/36$, per $x, y = 1, 2, \ldots, 6$.



2.d. Parell de variables. Funcions de prob. en parell de VAD

- Definim **funció de probabilitat conjunta** de *X* i *Y*:

  $$P_{X,Y}(x,y) = P(X=x \cap Y=y)$$

- Definim la **funció de probabilitat de X condicionada per Y**:

  $$P_{X/Y=y}(x) = P_{X,Y}(x,y) \,/\, P_Y(y)$$

- *X* i *Y* són **V.A. independents si:**

  $$P_{X,Y}(x,y) = P_X(x) \cdot P_Y(y) \iff P_{X/Y=y}(x) = P_X(x) \iff P_{Y/X=x}(y) = P_Y(y)$$

Càlculs de probabilitats condicionades i independència amb dues VAD

$P(X=x \mid Y=y) = P(X=x \cap Y=y) \,/\, P(Y=y) = P_{X|Y=y}(x) = P_{X,Y}(x,y) \,/\, P_Y(y)$

si X,Y són independents, $P(X=x \cap Y=y) = P(X=x) \cdot P(Y=y)$

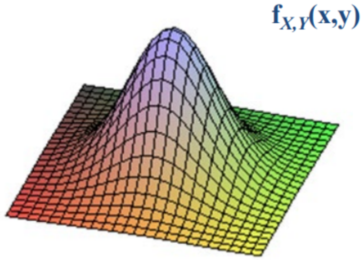
o bé $P(X=x \mid Y=y) = P(X=x)$)

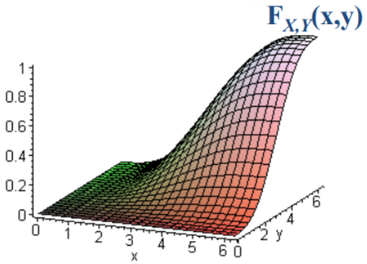
o bé $P(Y=y \mid X=x) = P(Y=y)$)

2.d. Parell de variables. Parell de VAC

En cas de que existeixin dos variables contínues *X* i *Y* en la mateixa experiència, la relació comuna es reflecteix a través de la **funció de densitat conjunta, $f_{X,Y}(x,y)$.**

- Sigui $F_{X,Y}(x,y) = P(X \leq x \cap Y \leq y)$, *funció de distribució conjunta* de les variables aleatòries
- Si derivem $F_{X,Y}(x,y)$ respecte a les variables (x i y), obtenim $f_{X,Y}(x,y)$
- La definició de funcions condicionades és idèntica que per a V.A.D: $f_{X|Y=y}(x) = f_{X,Y}(x,y) / f_Y(y)$
- El volum total tancat sota $f_{X,Y}(x,y)$ és igual a 1, i una porció d'ell equival a una probabilitat





2.d. Parell de variables. Indicadors

- A partir d'un parell de variables *X* i *Y* definim indicadors de la seva relació bivariant (equivalents als mostrals que es veuen a estadística descriptiva)
- La covariància indica si existeix relació lineal o no, a partir del producte, per cada parell de valors, de la diferència respecte al seu valor esperat

$$\boldsymbol{VAD} \rightarrow Cov(X,Y) = \sum_{\forall x} \sum_{\forall y} (x - E(X))(y - E(Y)) \cdot p_{XY}(x,y)$$

$$\boldsymbol{VAC} \rightarrow Cov(X,Y) = \iint_{\forall x,y} (x - E(X))(y - E(Y)) \cdot f_{XY}(x,y)\, \mathrm{dxdy}$$

- La correlació indica si existeix relació lineal o no, relativitzant-ho a valors entre -1 i 1 (a partir de la covariància i dividint per les desviacions corresponents)

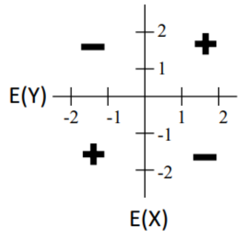
$$corr(X,Y) = \rho_{X,Y} = \rho = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Per qualsevol parell de variables *X* i *Y* → $-1 \leq \rho_{X,Y} \leq 1$

La correlació és més interpretable per estar estandaritzada

2.d. Parell de variables. Indicadors

Si una variable té $\mu = 0$ i $\sigma = 1$ es diu que és una variable *centrada* i *reduïda* (estandaritzada). Igualment, diem que la correlació ve estandaritzada perquè pren valors entre -1 i 1.



Aquests quadrants indiquen el signe del producte resultant. Si la probabilitat es reparteix preferentment entre els quadrants positius, la relació entre *X* i *Y* és *directa*. En cas contrari la relació és *inversa* (si *X* augmenta, *Y* tendeix a disminuir).

app

- Si $|\rho_{X,Y}|=1$, la relació és total i lineal: $\boldsymbol{Y = a+b \cdot X}$ (signe $\rho_{X,Y}$ = signe *b*)
- $|\rho_{X,Y}|$ a prop de 1 ⇒ *X* i *Y* estan molt relacionades linealment
- *X* i *Y* independents ⇒ $\rho_{X,Y}=0$, però a la inversa no és cert
- La magnitud de la covariància depen de l'escala agafada per les variables [Per exemple, si canvio les unitats de metres a quilòmetres, la covariància canviarà però la correlació, no]

2.d. Parell de variables. Indicadors

Siguin *X* i *Y* dues variables aleatòries, i *a* i *b* dos escalars:

- **Esperança**
  - E(*X*·*Y*) = E(*X*)·E(*Y*) + Cov(*X*,*Y*) (Si *X* i *Y* són **independents**, llavors E(*X*·*Y*) = E(*X*)·E(*Y*))
- **Variància**
  - V(*X*+*Y*) = V(*X*) + V(*Y*) + 2 Cov(*X*,*Y*)
  - V(*X*–*Y*) = V(*X*) + V(*Y*) – 2 Cov(*X*,*Y*)

  (Si *X* i *Y* són **independents**, llavors V(*X*±*Y*) = V(*X*) + V(*Y*)
  → ATENCIÓ: a l'expressió de la dreta sempre és un **"+"**)
- **Covariància**
  - Cov(a*X*, b*Y*) = a·b·Cov(*X*,*Y*)
  - Cov(*X*,*X*) = V(*X*)

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 → $E(X) = \mu_X$
  - Variància → $V(X) = \sigma_X^2$
- A partir de les **funcions de probabilitat i distribució** de probabilitat es calculen probabilitats:

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

- 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: $F_X(x_\alpha) = \alpha \quad (0 \le \alpha \le 1)$

Model de Bernoulli

- **Definició**: Número d'èxits en la realització d'un únic experiment amb 2 possibles resultats*: **0** ("no èxit") i **1** ("èxit")

* Parlem de respostes binàries o dicotòmiques

- **Notació**: X~Bern(p)
- **Paràmetres**: p (probabilitat d'èxit)
- **Funció de probabilitat**:

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*

$$E(x) = 0 \cdot q + 1 \cdot p = p \qquad Var(x) = (0-p)^2 \cdot q + (1-p)^2 \cdot p = p \cdot q$$

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 *estranys*), 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~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**:

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

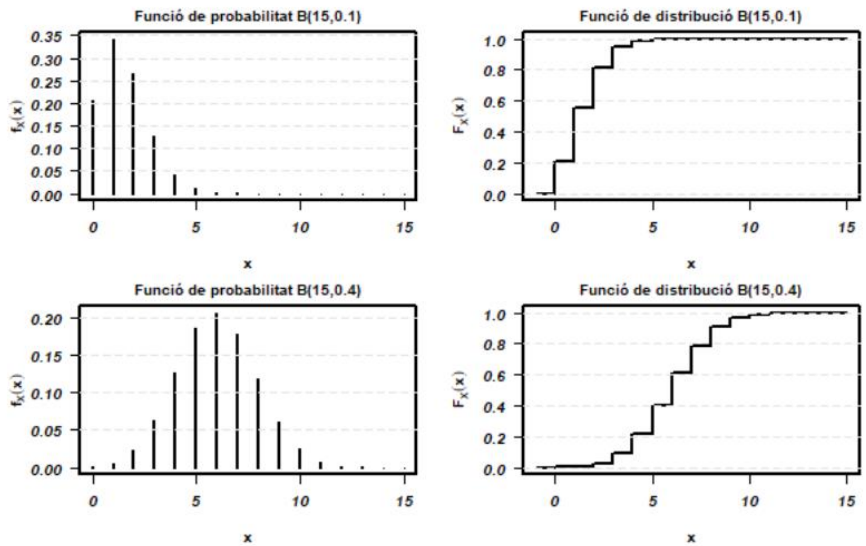
on [q = 1 – p]

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

$$C_n^r = \binom{n}{r} = \frac{n!}{r! \cdot (n-r)!}$$

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

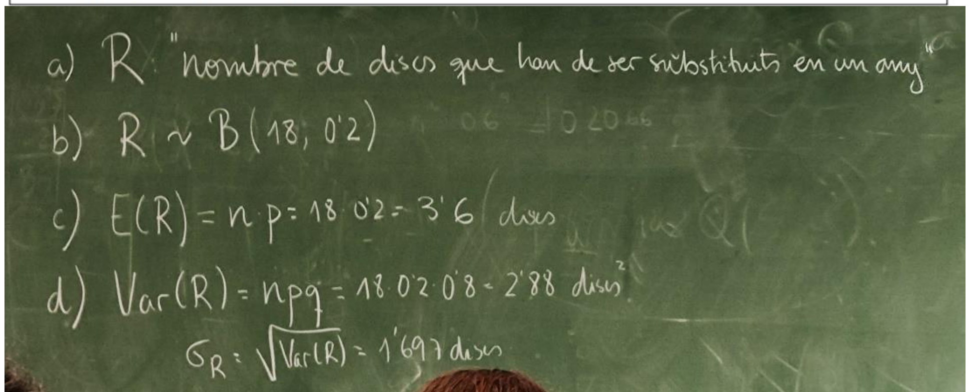
dbinom, pbinom, qbinom són funcions en R, Rstudio, i R online (https://rdrr.io/snippets/)

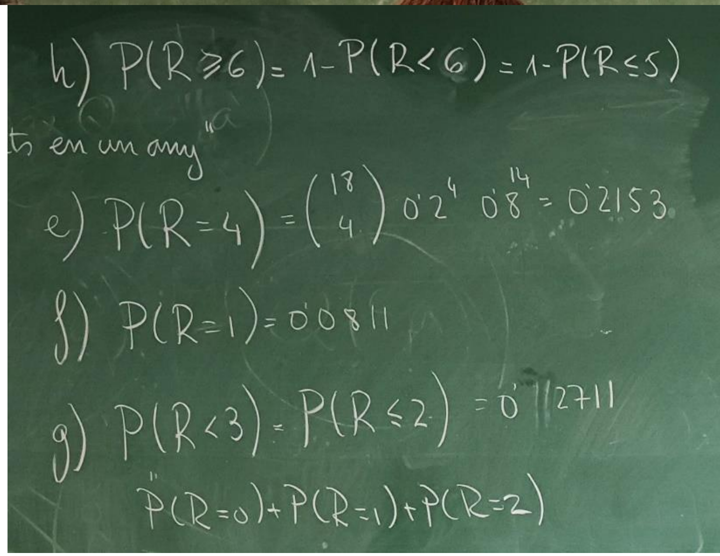
El model Binomial permet relacionar probabilitats de variables que segueixen models amb paràmetre p o bé 1-p:
P (X ≤ k) = 1 - P (Y ≤ n-k-1) on *X ~* B(*n, p*) i *Y ~*B(*n*, 1-*p*) (o bé pbinom(k,n,p)=1-pbinom(n-k-1,n,1-p))

Exercici Model Binomial

La cabina de discos d'un servidor conté 18 discs idèntics. Un disc de cada 5 necessita ser substituït al cap d'un any per ser reparat

VAD que compta substitucions:	
R: "número de discs que han de ser substituïts al cap de l'any"	
Model per a la VAD:	R ~ B(n=18, p=0.2)
Número esperat de substitucions anuals:	E(R)=3.6
Variància i desviació típica:	V(R)=2.88 → σ_x = 1.697
Prob. d'observar 4 casos:	P(R=4)=0.2153
Prob. sols una substitució:	P(R=1) =0.0811
Prob. de tenir-ne menys de 3:	P(R<3)=0.2713
Prob. que almenys 6 discs siguin substituïts:	P(R≥6)=0.1329





Model Geomètric

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

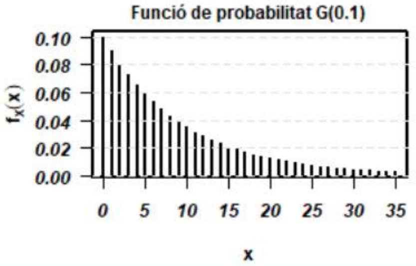
  Les funcions en **R** són pel número de fracassos enlloc d'intents *(k intents són k-1 fracassos)*
- **Paràmetres:** p (probabilitat d'observar 1 èxit)
- **Funció de probabilitat** :

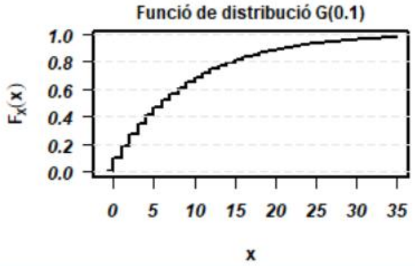
$$P(X = k) = q^{k-1} \cdot p \qquad k \geq 1$$

Indicadors:

E(X) = 1/p

V(X) = (1-p)/p²





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~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**:

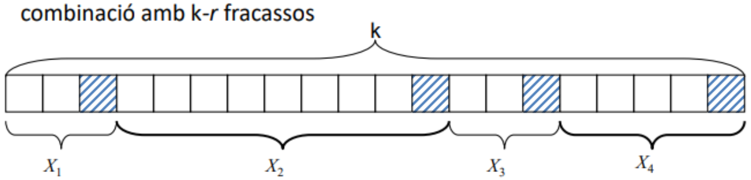
$$P(X = k) = \binom{k-1}{r-1} \cdot p^r \cdot q^{k-r}$$

Indicadors:

E(X) = r/p

V(X) = r(1-p)/p²

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

Exercici Model Geomètric

La probabilitat d'un disc dur defectuós d'una determinada marca és 0.05. Es durà a terme una inspecció sobre un lot de discs durs.

Quina distribució segueix la VAD X: "Nombre de discs durs a revisar fins a trobar el primer defectuós"?:

X~Geom(p=0.05)

Quina és la probabilitat d'haver-ne de revisar només 1? P(X=1) = 0.05

Quina és la probabilitat d'haver-ne de revisar 10? P(X=10) = 0.959 · 0.05 = 0.032

Quina és la probabilitat d'haver-ne de revisar més de 2?

P(X>2) = 1 – P(X=1) – P(X=2) = 1 – 0.05 – 0-95*0.05 = 0.9025

Quin és el nombre esperat de discos que s'hauran de revisar? E(X) = 1/0.05 = 20

Amb quin nombre de discos en el lot podrem estar segurs que trobarem algun defectuós amb una probabilitat almenys del 50% P(X 0.50 → x= 13

qgeom(0.50,0.05)

Exercici Model Binomial negatiu

Per aprovar un determinada assignatura, es requereix treure més d'un 5 en un problema d'estatus almenys 3 vegades. Un alumne concret té una probabilitat invariable de 0.6 de treure més d' un 5.

Quina distribució segueix la VAD X: "Nombre d'execucions per treure 3 notes superiors a 5"?:

X~BN(r = 3, p=0.6)

Quina és la probabilitat d'haver-ne de fer 1 execució?

P(X=1) = 0

Quina és la probabilitat d'haver-ne de fer 10 execucions?

P(X = 10) = $\binom{9}{2}$ = $0.6^3 * 0{,}4^7$ = 0.013

Quin és el nombre esperat de execucions que s'hauran de fer per aprovar?

E(X) = 3/0.6 = 5

Model Poisson

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

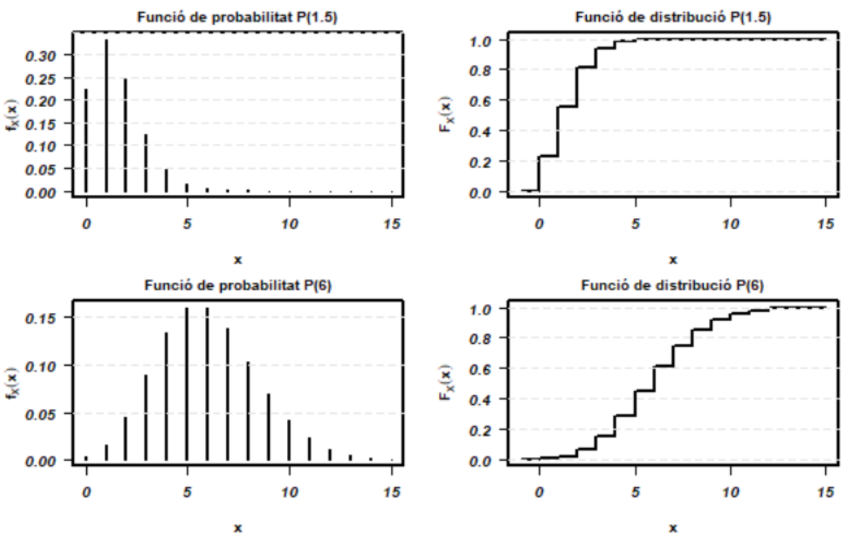
  Al igual que en el model binomial pot haver-hi pròpiament una repetició d'experiències idèntiques tipus Bernoulli, però també pot correspondre a fenòmens que ocorren inesperadament
  (Ex: trucades a una centraleta es poden representar amb una variable Poisson de "nombre de trucades per hora", sabent la mitjana de trucades rebudes per hora)
- **Notació**: X~P(λ) — Notació de funcions en **R** pel model: dpois(), ppois(), qpois()
- **Paràmetres:** λ (taxa de l'esdeveniment)
- **Funció de probabilitat:** $P(X=k)=\frac{e^{-\lambda}\cdot\lambda^k}{k!} \quad amb \quad k=0,1,\ldots$
- No té **funció de distribució** analítica → És el sumatori de probabilitats puntuals
- **Indicadors:**
  - E(X) = λ
  - V(X) = λ

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

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

Nota: La suma de VAD Poisson és també una VAD Poisson amb paràmetre λ igual a la suma dels paràmetres:
X ~ P(λ1) Y ~ P(λ2) → X+Y ~ P(λ1+ λ2)
A partir d'una VAD Poisson, podem definir altres VAD aplicant proporcionalment al paràmetre el canvi en l'interval:
X="...en interval t" ~ P(λ) → *Y="...en interval kt"* ~ P(kλ)

Exercici Model Poisson

El centre de càlcul d'una empresa atén les incidències que sorgeixen als treballadors. S'ha observat que aquestes apareixen esporàdicament, encara que l'elevat número d'usuaris implica que el volum de problemes a tractar diàriament sigui considerable (s'ha suposat una mitjana de 2.35 incidències/dia).

Quina és la distribució de la v.a "número d'incidències diàries"?

X: "número d'incidències al dia" → X~ P(λ=2.35)

Quina és la seva Esperança? I la seva desviació típus?

E(X) = 2.35 incidències i σ_X=1.53

Quina és la distribució de la v.a "número d'incidències en 5 dies"? I en 7?

X_5 : "número d'incidències 5 dies" →X_5~ P(λ =11.75) ; X_7~ P(λ =16.45)

Probabilitat que en un dia es produeixi 3 incidències

P(X=3) = 0.2063

Probabilitat d'observar menys de 3 incidències en un dia:

P(X<3) = P(X≤2) = 0.5828

Entre el dilluns i el dimarts s'han rebut sis incidències. Quina és la probabilitat que de dilluns a divendres es tractin no més de 15?

$P(X_3 \le 15-6) = P(X_3 \le 9) = 0.8254$

Quina és la probabilitat que cap dia de la setmana presenti incidències?

$P(X_7=0) = 0.0000079$

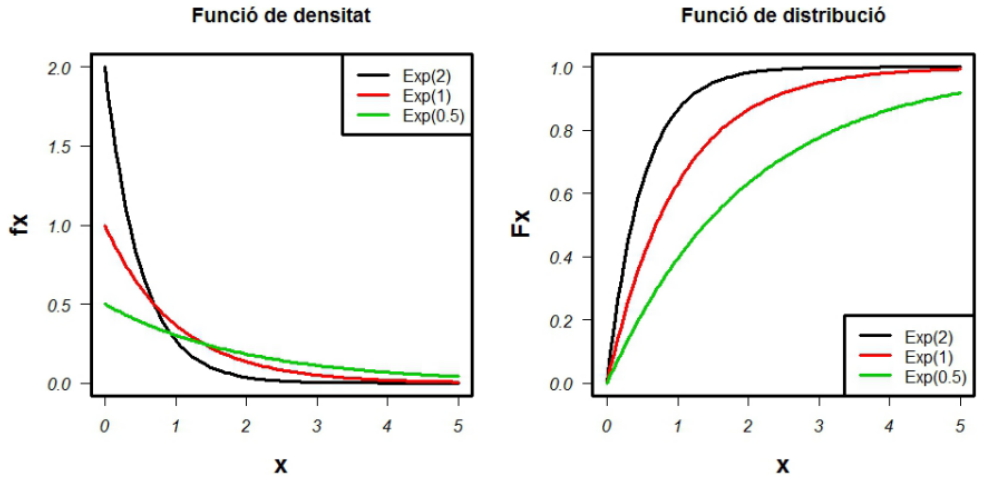
TAULA resum de models de VAD

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	$q \cdot r/p^2$
Poisson	P(λ)	0, 1, 2,…	λ	λ

$0 < p < 1$ $q = 1 - p$ $n \in N$ $r \in N$ $\lambda \in R^+$

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?



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~Exp(λ) 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**:

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

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

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

Model Exponencial. Ex. de càlcul de probabilitats

Sigui $X \sim$ Exp(λ =2):

- **Probabilitat puntual**. → 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 → $P(X \le 2) = 1 - e^{-2 \cdot 2} = 1 - e^{-4} = \mathbf{0.9817}$
  - Amb R → $P(X \le 2) = pexp(q = 2, \text{rate} = 2) = \mathbf{0.9816844}$
- **Quantils**. Quin és el valor tq la probabilitat de quedar per sota d'ell és almenys 0.95?
  - Amb fórmules → $P(X \le x_{0.95}) = 0.95 \to 1 - e^{-2 \cdot x_{0.95}} = 0.95 \to x_{0.95} = \mathbf{1.4979}$
  - Amb R → $P(X \le x_{0.95}) = 0.95 \to qexp(p = 0.95, rate = 2) = \mathbf{1.497866}$

Nota: pexp, qexp són funcions en R, Rstudio, i R online (https://rdrr.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] → $f_X(x)$ no és una probabilitat, a diferència de la $p_X(x)$ de les VAD
- Recordem que en una VAC:

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

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:

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

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 en els 10'' següents, o (B) rebre 10'' després d'una arribada?

  Solució: Igual

Aplicacions del Model Exponencial

- **"Failure Rate and Reliability"** (Jane Horgan lecture 17)

  $X \sim Exp(\lambda)$ $\quad F_X(x)=P(X \le x)=1 - e^{-\lambda \cdot x}$ $\quad R_X(x)= P(X > x) = e^{-\lambda \cdot x}$

  - R és la funció de "reliability" o de fiabilitat (probabilitat de durar més de..)
  - λ és "failure rate" o taxa d'error
  - E(X) = 1/λ és "MTTF" o "Mean Time To Failure"

- **"Modelling Response Times: M/M/1"** (Jane Horgan lecture 17)
  - La teoria de cues permet estudiar sistemes en que interaccionen més d'una variable (per exemple temps de resposta en un sistema d'espera amb cues).

El model M/M/1 indica: 1 per contemplar una sola cua no finita, i dos M's pels dos temps d'arribada i servei exponencials (*Exp*(λ) i *Exp*(μ) o número d'arribades **P**(λ) i serveis **P**(μ)), que compleixen la propietat de Markov de no tenir memòria.

Amb l i m es defineix un indicador del sistema com factor de càrrega o "traffic intensity" (λ / μ):
- si λ /μ > 1 (λ > μ) → la taxa d'arribades és superior a la de sortides (sobreutilització del sistema)
- si λ /μ = 1 (λ = μ) →la taxa d'arribades s'iguala amb la taxa de sortides
- si λ /μ < 1 (λ < μ) → la taxa d'arribades és inferior a la de sortides (infrautilització del sistema)

(el model M/M/1 permet un tractament per teoria de cues o per simulació. Models més complexos poden permetre sols el tractament per simulació)

Exercici Model Exponencial i Poisson

El centre de càlcul d'una important empresa atén les incidències que els sorgeixen als treballadors. Se suposa una mitjana de 4 incidències/dia, i 8 hores laborables al dia.

Quina és la distribució de la v.a "número d'incidències diàries"?

1--> X: "número d'incidències al dia" → X~ P(λ=4)

Quina és la distribució de la v.a "número d'incidències per hora"?

2 --> Xh: "número d'incidències per hora" → Xh~ P(λ=4/8) = P(λ=1/2)

Probabilitat de rebre 0 incidències en un dia:

3 --> P(X = 0) = 0.0183

Probabilitat de rebre 0 incidències en una hora:

4 --> P(Xh = 0) = 0.607

Quina és l'esperança de la variable temps (en hores) entre incidències?

5 --> T: "Temps en hores entre incidències" ~ Exp(λ = 1/2) → E(T) = 2 hores

I la desviació tipus?

6 --> V(X) = 1/λ2 → V(X) = 4 → σx = 2 hores

Probabilitat d'estar 8 o més hores sense rebre incidències:

7 --> P(T > 8) = 0.0183

Exercici Model Exponencial i Poisson. Fiabilitat

El centre de càlcul d'una important empresa garanteix treballar amb una mitjana de 2 hores entre incidències.

Quina és la distribució de la v.a "temps en hores entre incidències"?

1 --> T: "temps en hores entre incidències" → E(T) = 1/λ = 2 →T ~ Exp(λ=0.5)

Quina és la funció de distribució?

2 --> $F_T(t) = P(T \le t) = 1 - e^{-t/2}$

Quina és la funció de fiabilitat?

3 --> $R_T(t) = P(T > t) = e^{-t/2}$

Quina és la taxa d'errors?

4 --> λ = 0.5

Quin és el MTTF?

5 --> MTTF = E(T) = 2

Quin és el valor d'hores entre incidències que podem garantir que es superarà amb una fiabilitat del 78%?

6 --> $P(T > t_{0.78})=0.78 \rightarrow t_{0.78} =$

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~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!!! → $f_X(x) = \dfrac{1}{b-a} \qquad amb \quad a < x < b$

$F_X(x) = 0$ si $x < a$
$F_X(x) = 1$ si $x > b$ → $F_X(x) = \dfrac{x-a}{b-a} \qquad amb \quad a < x < b$

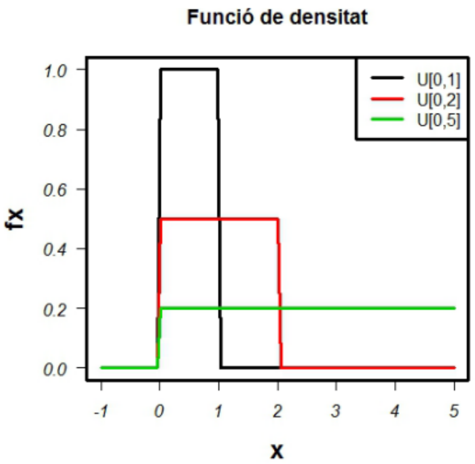
- **Indicadors:**
  - E(X) = (b + a)/2
  - V(X) = $(b-a)^2/12$

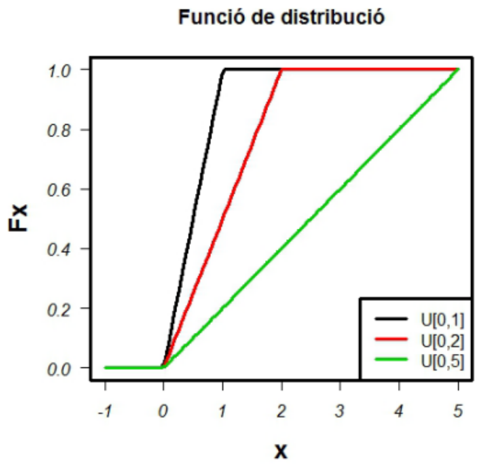
$$E(X) = \mu_X = \int_{-\infty}^{+\infty} x \cdot f_X(x)dx = \int_a^b x \cdot \frac{1}{b-a}dx = = \frac{1}{b-a} \cdot \left.\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

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

Model Uniforme. Representació gràfica

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

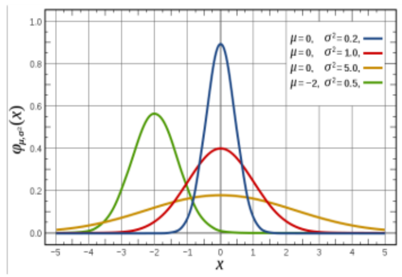


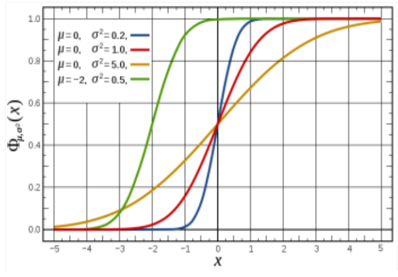


Model Normal (o de Gauss). Introducció

(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"





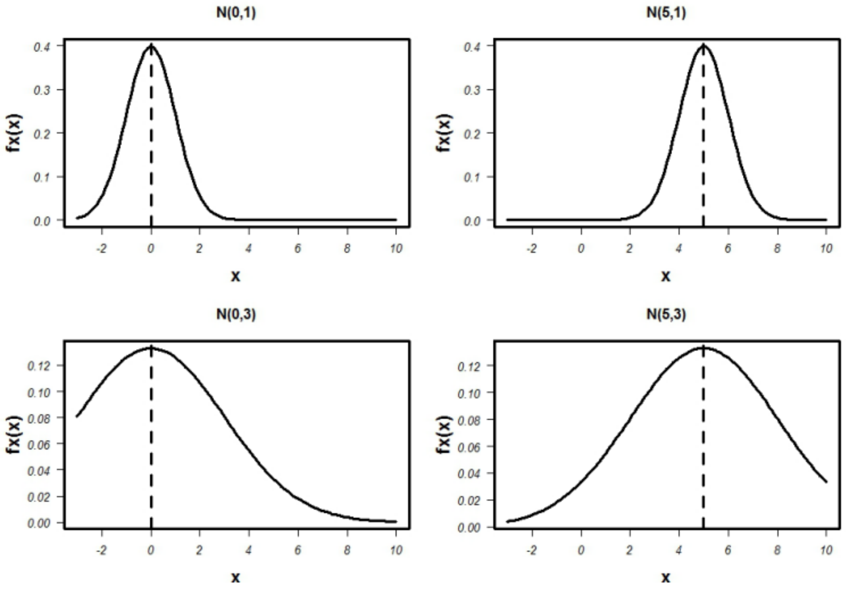
Model Normal

- **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~N(μ , σ) | 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**: $f_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad amb \quad x \in R$
- **La funció de distribució** no té expressió analítica explícita → R
- **Indicadors:**
  - E(X) = μ
  - V(X) = σ^2

Nota: la Normal amb paràmetres μ = 0 i σ=1 s'anomena **Normal estàndard**

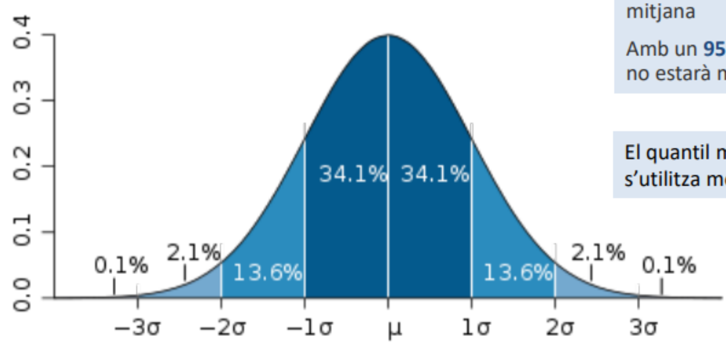
Model Normal. Representació gràfica

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



Model Normal. 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~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 $\mathbf{Z_{0.975} = 1.96}$ ja que s'utilitza molt en inferència estadística

Model Normal. Exemple de càlcul de probabilitats

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

- **Probabilitat puntual**. → 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 →**No es pot fer!!!**
  - Amb R → $P(X \le 2) = pnorm(q = 2, \text{mean} = 0, sd = 1) = \mathbf{0.9772499}$
    $(pnorm(2) = 0.9772499)$
- **Quantils**. Quin és el valor tq la probabilitat de quedar per sota d'ell és almenys 0.95?
  - Amb fórmules →**No es pot fer!!!**
  - Amb R →$P(X \le x_{0.95}) = 0.95 \rightarrow qnorm(p = 0.95, mean = 0, sd = 1) = \mathbf{1.645}$
    $(qnorm(0.95) = 1.64)$

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

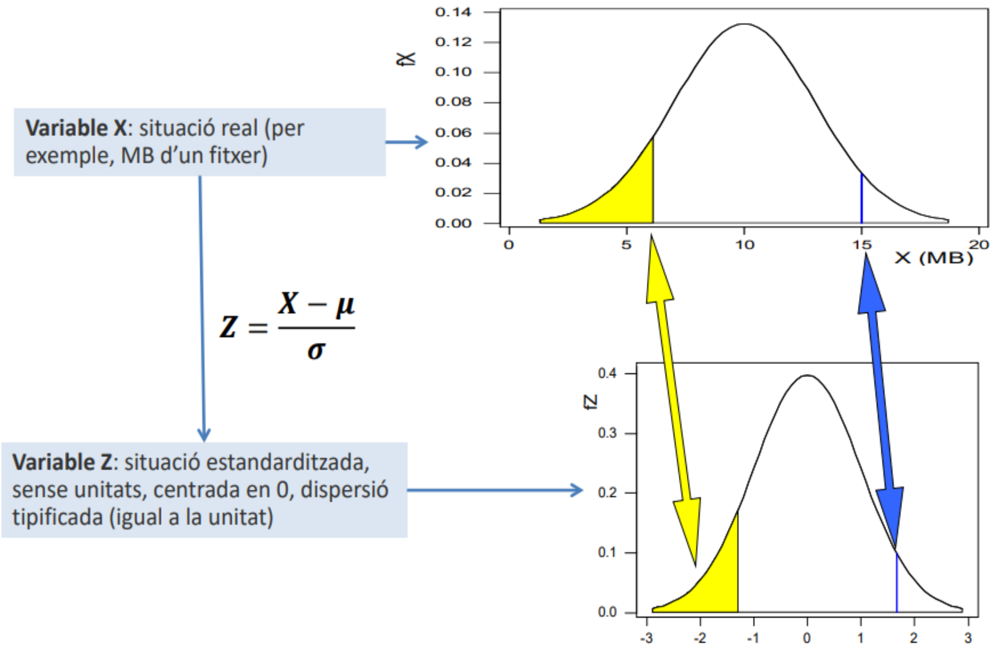
Nota: les anteriors funcions en R serveixen per a qualsevol Normal, però és habitual calcular-ho amb l'estàndard (**N(0,1)**) amb la transformació que veurem d'**estandarització**. De fet, mean=0 i sd=1 són els valors per defecte en les funcions

Model Normal. Estandardització

- La **combinació lineal de variables Normals** és Normal:
  - Sigui a i b, dos escalars i $X \sim N(\mu_X, \sigma_X)$ → $Y = a\cdot X + b \sim N(\boldsymbol{\mu_Y} = a\cdot\mu_X + b, \boldsymbol{\sigma_Y} = a\cdot\sigma_1)$
  - Sigui a i b, dos escalars, $X \sim N(\mu_X, \sigma_X)$ i $Y \sim N(\mu_Y, \sigma_Y)$ →
    → $W = a\cdot X + b\cdot Y \sim N\left(\boldsymbol{\mu_W} = a\cdot\mu_X + b\cdot\mu_Y, \boldsymbol{\sigma_W} = \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2 + 2\cdot ab\cdot\rho_{XY}\cdot\sigma_1\cdot\sigma_2}\right)$
- Aquesta propietat permet relacionar distribucions Normals a base de translacions i escalars. En particular, transformar a la Normal estàndard Z~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/σ, b= −μ/σ són escalars). És a dir:

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

Model Normal. Estandardització

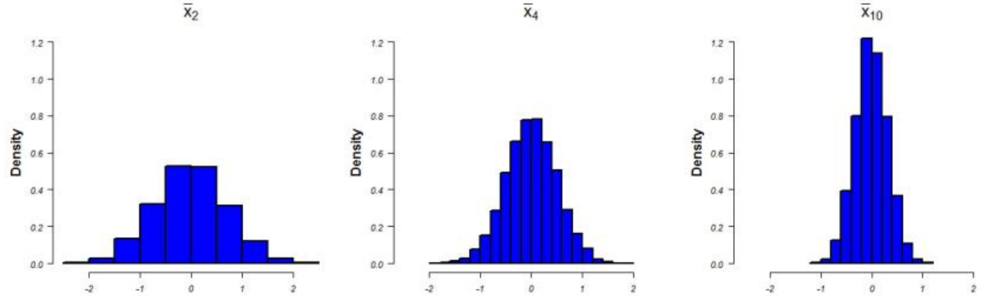


Distribució de la mitjana de v.a.

- Hem simulat $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$ 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.

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

- 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.



Exemples model Normal

Exemple 1: [**probabilitats inferiors a 0.5**] . Sigui Z ~ N(0, 1). Quin és el valor *z* que deixa una probabilitat per sota de p = 0.25?

Com 0.25 < 0.5, no surt a les taules; s'ha d'utilitzar simetries:

$$P(Z < -|z|) = P(Z > |z|) \rightarrow F_Z(0.6745) = 0.75 \rightarrow z = -0.6745$$

Exemple 2: [**Estandardització**] Sigui X: "Increment diari espai disc" ~ N(10 MB, 3 MB)

1) Quant val P(*X*>15)?

$$Z = \frac{X-10}{3} \sim N(0,1) \rightarrow P(X > 15) = P\left(Z > \frac{15-10}{3}\right) = P\left(Z > \frac{5}{3}\right) = 1 - P\left(Z \leq \frac{5}{3}\right)$$

$$= 1 - F_Z\left(\frac{5}{3}\right) = 1 - 0.9522 = 0.0478$$

2) 1 de cada 10 dies, l'augment és inferior a quant? (Quant val *t* tal que P(*X* < *t*) = 0.1?)

$$P(Z < t') = 0.1 \rightarrow t' = F_Z^{-1}(0.1) = -1.2816 \rightarrow t = 10 + 3 \cdot t' = 6.15\ MB$$

Teorema Central del Límit (TCL)

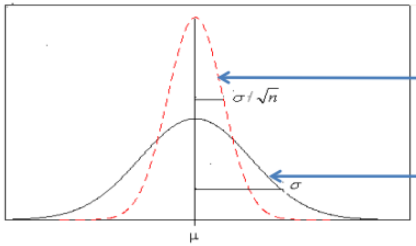
- Siguin X_1, X_2, ..., X_n independents i idènticament distribuïdes amb esperança μ i desviació típica σ. Llavors:

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

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

- 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!**

Teorema Central del Límit (TCL)



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

X_i

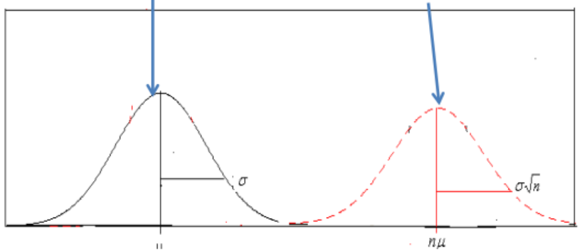
X_i $\qquad S_n = \sum X_i \xrightarrow[n\ gran]{} N(n\mu, \sigma\sqrt{n})$

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

Només han de complir:

E(X) = μ

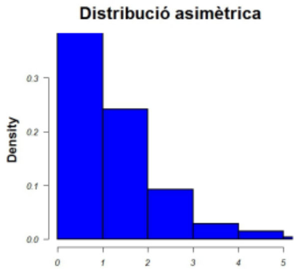
V(X) = σ^2

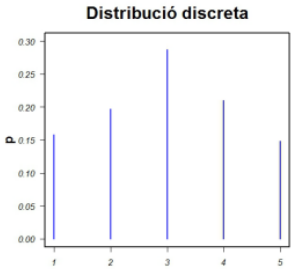


Teorema Central del Límit (TCL). "n"

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):





- 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

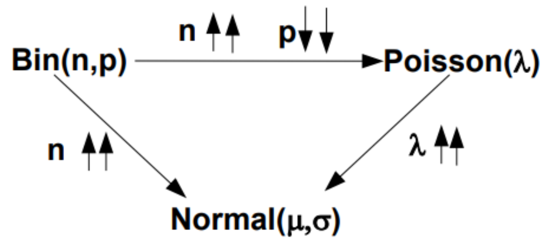
Una de les aplicacions pràctiques del TCL és que la distribució Normal es pot emprar com a aproximació d'altres distribucions:

- 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

  μ = n·p $\qquad$ i $\qquad$ σ^2 = n · p ·(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:

  μ = λ $\qquad$ i $\qquad$ σ^2 = λ



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

Exemple Teorema Central del Límit (TCL)

El treball de CPU per fer un *backup* presenta unes característiques diàries de mitjana 30'/dia i una desviació de 15'/dia. Si volem calcular probabilitats sobre el consum de CPU total mensual (suposant independència entre els 30 dies del mes), haurem de plantejar-nos la variable:

$S_n = X_1 + X_2 + \ldots + X_{30}$ (no ens donen la distribució de X_i, sols μ i σ)

1) Quina és la distribució de S_n?

$$S_n = \sum X_i \xrightarrow[n\ gran]{} N(n\mu, \sigma\sqrt{n}) = N(900\,, 82.15)\ min = N(15{,}1.37)\ hores$$

2) Quina és la probabilitat d'un consum total mensual de més de 18 hores de CPU?

$$P(S_n > 18) = P\left(\frac{S_n - 15}{1.37} > \frac{18-15}{1.37}\right) = P(Z > 2.19) = 1 - p(Z < 2.19) = 1 - 0.9857 = 0.0143$$

[un de cada 70 mesos]

3) I la probabilitat de, en un mes, un consum mitjà diari inferior a 36'?

$$\bar{X}_{30} \sim N\left(\mu, \frac{\sigma}{\sqrt{30}}\right) = N(30\,, 2.74)\ min \rightarrow P(\bar{X}_{30} < 36) = P\left(Z < \frac{36-30}{2.74}\right) = P(Z < 2.19) = 0.9857$$

Exercici Teorema Central del Límit (TCL)

Suposem que s'ha establert que el pes del equipatge d'un viatger segueix una distribució Normal amb mitjana 18.9 Kg. i desviació 4 Kg. Habitualment, si el equipatge d'un viatger sobrepassa els 20 Kg., llavors té un sobrepreu que depèn de l'excés de pes. Contesta les següents qüestions:

1. Troba la probabilitat que un viatger hagi de pagar sobrepreu per excedir el seu equipatge els 20 Kg. de pes. 0.392 [1-pnorm(20,18.9,4)]

2. Quin és el pes que podem assegurar , amb un 95% de probabilitat , que un equipatge no superarà? 25.479 [qnorm (0.95,18.9,4)]
3. Indica els paràmetres (μ , σ) de la variable pes <u>total</u> de 10 equipatges

   $N(\mu = 189, \sigma = 4\sqrt{10} = 12.649)$
4. Indica els paràmetres (μ , σ) de la variable pes <u>promig</u> de 10 equipatges

   $N(\mu = 18.9, \sigma = 4/\sqrt{10} = 1.265)$

Exercici diversos models

Continuem analitzant el cas de l'aeroport, donant importància al procés d'arribades de passatgers als punts de facturació. Respon a les següents preguntes.

1. Un determinat punt de facturació es caracteritza perquè el número de viatgers que arriben per minut es distribueix segons una Poisson amb mitjana de 9.5. Calcula la probabilitat que aquest número sigui menor que 7 [0.165]
2. En aquest punt de facturació, quina és la probabilitat d'observar exactament 10 arribades en un minut? [0.123]
3. En el mateix punt de facturació, quina és l'esperança de la variable temps (*en segons*) entre dues arribades? [6.316 s.]
4. Al punt de facturació, quina és la probabilitat d'estar menys de 4 segons sense arribades? [0.469]
5. Considerant 18 punts de facturació caracteritzats per una probabilitat 0.8 d'observar exactament 0 arribades en un minut, quina és la probabilitat de tenir més de 14 punts amb 0 arribades? [0.501]