Variables aleatòries

Variable aleatòria

Una variable aleatòria (v.a.) és una variable els valors de la qual són resultats numèrics d'un fenomen aleatori.

- Exemples: Suma dels valors d'una tirada de dos daus de sis cares, temps de durada d'una bombeta, nombre de llampecs durant una tempesta, pes d'un pot d'olives, …

El domini d'una v.a. és el conjunt de valors numèrics que pot prendre.

- Exemples: $\mathbb{R}^+, \mathbb{N}_0, \{2, 3, \ldots, 12\}, \ldots$

La funció de distribució

La *funció de distribució* és una funció que caracteritza una v.a. X a través de la seva probabilitat acumulada, i.e.

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

Propietats:

- F_X és **no decreixent**,
- F_X és contínua per la dreta,
- $\forall x \in \mathbb{R}, 0 \leq F_X(x) \leq 1$ i
- $\lim_{x \to -\infty} F_X(x) = 0$ i $\lim_{x \to +\infty} F_X(x) = 1$

Exemple: Suma de dos daus

Si X és la v.a. "suma del valor de dos daus", amb la seva **funció de distribució** podem respondre preguntes sobre la probabilitat d'esdeveniments associats.

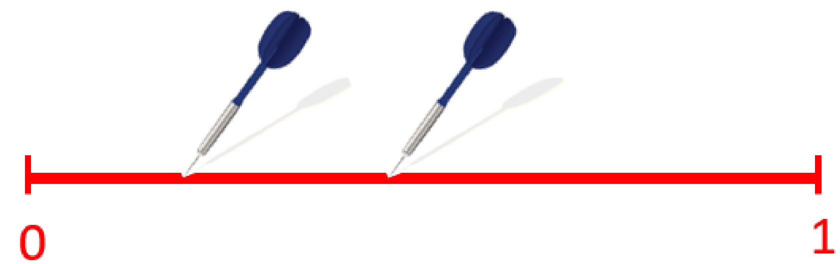
- Quina és la probabilitat d'obtenir un 7?, o sigui quant val $P(X = 7)$?
- Quina és la probabilitat d'obtenir un valor superior a 10, o sigui, quant val $P(X > 10)$?
- Quin és el llindar ℓ que ens garanteix que la probabilitat que la nostra tirada sigui inferior a ℓ és com a mínim 0.5?, o sigui, quin ℓ garanteix $P(X < \ell) \geq 0.5$?

x	F(x)
2	0.0278
3	0.0833
4	0.1667
5	0.2778
6	0.4167
7	0.5833
8	0.7222
9	0.8333
10	0.9167
11	0.9722
12	1.0000

Exemple: Tirada de dos dards

Denotem per X la v.a. “distància entre dos dards”. La seva **funció de distribució** és

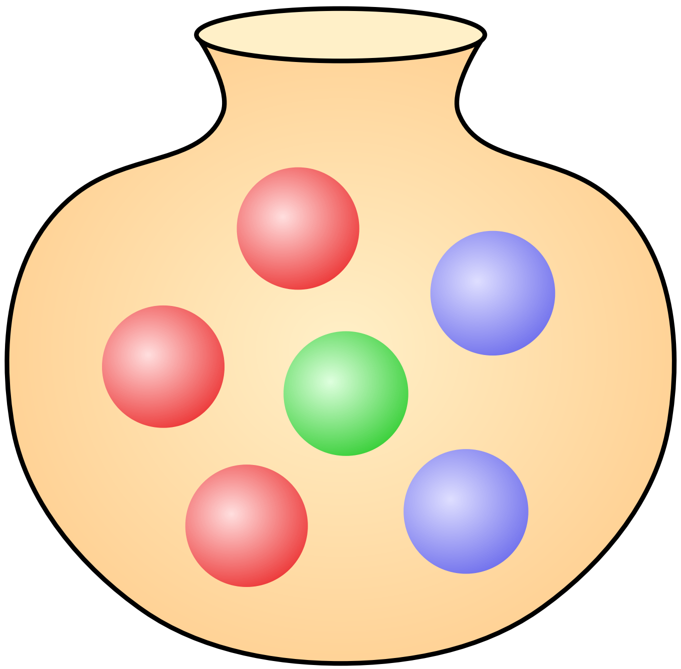
$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ -x^2 + 2x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$



- Quina és la probabilitat que la distància entre els dos dards sigui més petita que 0.1?
- Quina és la probabilitat que la distància entre els dos dards sigui 0.1?
- Quin és el llindar que ens garantirà que el 50% de les distàncies entre dards no superarà aquest llindar?

Variables aleatòries discretes

Exemples de variables aleatòries discretes



Donada una urna com la de l'esquerra, podríem definir les següents v.a. discretes:

- Nombre de boles blaves en treure dues boles de l'urna.
- Producte del nombre de boles vermelles obtingudes pel nombre de boles blaves obtingudes en treure quatre boles de l'urna.
- Nombre de boles extretes abans de treure la bola verda.

Funció de probabilitat

Per una v.a. discreta X que pot prendre els valors $D_X = \{x_i; i \in I\}$ amb $I \subseteq \mathbb{N}$. La *funció de probabilitat* és la funció p definida de D_X a $[0, 1]$ tal que

$$p_X(x_i) = P(X = x_i).$$

Propietats:

- Tenim que $\sum_{x \in D_X} p_X(x) = 1$.
- La probabilitat de B, $B \subseteq D_X$, serà $P(B) = \sum_{x \in B} p_X(x)$.
- La **funció de distribució** d'una v.a. discreta en un valor x la podem calcular com la suma de tots els valors més petits o iguals que x,

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p_X(x_i).$$

Exemple: Suma de dos daus (2)

Si X és la v.a. "suma del valor de dos daus", amb la seva **funció de probabilitat** podem respondre preguntes sobre la probabilitat d'esdeveniments associats.

- Quina és la probabilitat d'obtenir un 7?, o sigui quant val $P(X=7)$?
- Quina és la probabilitat d'obtenir un valor superior a 10, o sigui, quant val $P(X>10)$?
- Quin és el llindar ℓ que ens garanteix que la probabilitat que la nostra tirada sigui inferior a ℓ és com a mínim 0.5?, o sigui, quin ℓ garanteix $P(X<\ell)\geq 0.5$?

x	p(x)
2	0.0278
3	0.0556
4	0.0833
5	0.1111
6	0.1389
7	0.1667
8	0.1389
9	0.1111
10	0.0833
11	0.0556
12	0.0278

Variables aleatòries contínues

Exemples de variables aleatòries contínues

- Distància al centre d’una diana de diàmetre de 450 mm després de llançar un dard.
- Temps d’espera d’un servidor.
- Durada d’una bombeta.
- Kilograms suportats per una barra d’acer fins que es trenca.

Funció de densitat

Per una v.a. continua X que pot prendre els valors $D_X \subseteq \mathbb{R}$. La *funció de densitat* (pdf) és la funció f_X definida de D_X a $\mathbb{R}^+$ tal que per a i b reals, tenim que $P(a \leq X \leq b) = \int_a^b f_X(y)dy$.

> $f_X(x)$ mesura quantes realitzacions d'X esperaríem trobar al voltant d'un valor x.

Propietats:

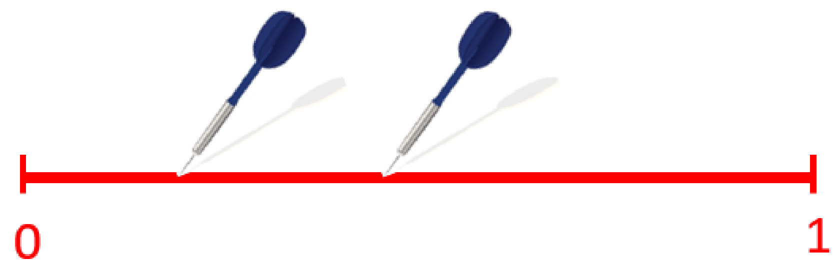
- Tenim que $\int_{-\infty}^{\infty} f_X(y)dy = 1$.
- La probabilitat d'un esdeveniment B, $B \subseteq D_X$, és $P(B) = \int_{x \in B} f_X(y)dy$.
- En particular, la **funció de distribució** d'una v.a. continua en un valor x es defineix com la integral impròpia $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)dy$.

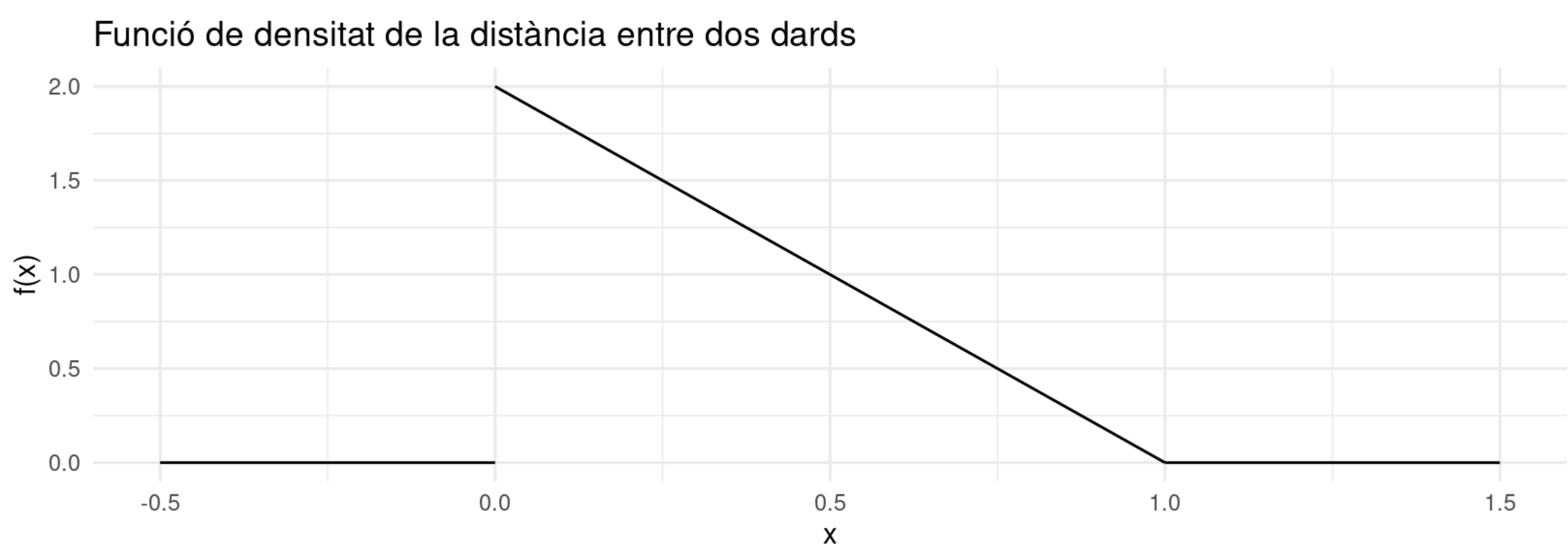
> Una pdf no retorna una probabilitat, retorna una densitat.

Exemple: Tirada de dos dards (2)

Denotem per X la v.a. “distància entre dos dards”. La seva **funció de densitat** és

$$f_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ -2x + 2 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{si } x > 1 \end{cases}$$





Quantitats importants d’una v.a.

Els p-quantils

> El p-quantil és el valor q més petit pel que la probabilitat de que la v.a. sigui més petita o igual que ell és com a mínim p.

Formalment, el p-quantil d'una variable aleatòria, és el mínim valor $q \in D_x$ tal que

$$p \leq F_X(q)$$

Casos especials:

- El 0.5-quantil normalment s'anomena *la mediana*.
- El 0.25-quantil i el 0.75-quantil s'anomenen *el primer* i *el tercer quartil* respectivament. El 0.5-quantil també se l'anomena *el segon quartil*.
- De manera semblant als quartils tenim els quintils (5 parts), els decils (10 parts), els percentils (100 parts), …

El valor esperat

El *valor esperat* o el *primer moment* d'una variable aleatòria és el valor mitjà esperat d'una variable aleatòria.

v.a. discreta | v.a. continua

El valor esperat d'una v.a. ve definit per

$$\mathbb{E}\left[X\right] = \sum_{x \in D_X} x\, p(x)$$

En general el moment k-èssim d'una variable aleatòria és

$$\mathbb{E}\left[X^k\right] = \sum_{x \in D_X} x^k\, p(x).$$

La variància

La *variància* o el *moment central de segon ordre* és el valor esperat del quadrat de la diferència entre l'observació i el valor central de la variable aleatòria.

v.a. discreta | v.a. continua

La variància d'una v.a. ve definida per

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \sum_{x \in D_X} (x-\mu)^2\, p(x)$$

on $\mu = \mathbb{E}\left[X\right]$.

En general, el moment central d'ordre k-èssim és

$$\mathbb{E}\left[(X-\mu)^k\right] = \sum_{x \in D_X} (x-\mu)^k\, p(x).$$

Exemple: Suma de dos daus (3)

```
x = 2:12
Fx = c('2' = 0.0278, '3' = 0.0833, '4' = 0.1667, '5' = 0.2778, '6' = 0.4167, '7' = 0.5833,
       '8' = 0.7222, '9' = 0.8333, '10' = 0.9167, '11' = 0.9722, '12' = 1.0000)
px = c('2' = 0.0278, '3' = 0.0556, '4' = 0.0833, '5' = 0.1111, '6' = 0.1389, '7' = 0.1666,
       '8' = 0.1389, '9' = 0.1111, '10' = 0.0833, '11' = 0.0556, '12' = 0.0278)
```

- La mediana serà

```
x[min(which(0.5 <= Fx))]
## [1] 7
```

- El valor esperat serà $\mathbb{E}\left[X\right] = \sum_{x \in D_X} x\, p(x)$,

```
sum(x * px)
## [1] 7
```

- La variància serà $\mathrm{Var}\left(X\right) = \sum_{x \in D_X} (x-7)^2\, p(x)$,

```
sum( (x-7)^2 * px)
## [1] 5.8352
```

Exemple: Suma de dos daus (4)

Anem a fer una simulació: fem 1000 tirades d'un dau d1 i 1000 tirades d'un dau d2.

```
d1 = sample(x = 1:6, size = 1000, replace = TRUE)
d2 = sample(x = 1:6, size = 1000, replace = TRUE)
```

Si calculem els estadístics bàsics de la suma de les simulacions de d1 i d2 hauríem d'obtenir resultats semblants a les quantitats calculades:

```
median(x = d1+d2)
## [1] 7

mean(d1+d2)
## [1] 6.994

var(d1+d2)
## [1] 5.573538
```

Exemple: Tirada de dos dards (3)

```r
Fx = function(x) ifelse(x < 0, 0, ifelse(x < 1, -x^2 + 2*x, 1))
fx = function(x) ifelse(x < 0, 0, ifelse(x < 1, -2*x + 2, 0))
```

- La mediana serà

```r
# Busquem el valor d'F(x) que acumula 0.5
optimise(function(x) (0.5 - Fx(x))^2, interval = c(0, 1))
## $minimum
## [1] 0.2929073
##
## $objective
## [1] 3.979888e-10
# Si resolem l'equació de 2n grau manualment podem veure que la mediana és exactament 1-sqrt(2)/2
```

- El valor esperat serà $\mathbb{E}\left[X\right] = \int_0^1 x\, f(x)\mathrm{d}x$,

```r
integrate(function(x) x * fx(x), lower = 0, upper = 1)
## 0.3333333 with absolute error < 3.7e-15
```

- La variància serà $\mathrm{Var}\left(X\right) = \int_0^1 (x-(1-\sqrt{2}/2))^2\, f(x)\mathrm{d}x$,

```r
integrate(function(x) (x - (1-sqrt(2)/2))^2 * fx(x), lower = 0, upper = 1)
## 0.05719096 with absolute error < 6.3e-16
# Si resolem l'integral manualment, el resultat exacte és 1-(2 sqrt(2)/3)
```

Exemple: Tirada de dos dards (4)

Anem a fer una simulació: fem 1000 tirades de dard d1 i 1000 tirades de dard d2.

```
d1 = runif(1000, min = 0, max = 1)
d2 = runif(1000, min = 0, max = 1)
```

Si calculem els estadístics bàsics de la diferència (en valor absolut) de les simulacions de d1 i d2 hauríem d'obtenir resultats semblants a les quantitats calculades:

```
median(x = abs(d1-d2))
## [1] 0.3085164

mean(abs(d1-d2))
## [1] 0.342715

var(abs(d1-d2))
## [1] 0.05836114
```

Variables aleatòries conjuntes

Independència de variables aleatòries

Si tenim dues variables aleatòries X i Y, parlem de la *funció de distribució conjunta d'X i Y*, i la denotem per $F_{X,Y}$, la funció bivariada $F_{X,Y}(x, y) = P(X \leq x \cap Y \leq y)$[1].

- Diem que dues variables aleatòries X i Y són independents si

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

  per tots els possibles valors d'x i y.

La covariància entre dues v.a.'s

Si tenim dues variables aleatòries X i Y, la covariància entre X i Y es calcula com

$$\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

> Per calcular la covariancia entre X i Y seria necessari disposar de la funció de probabilitat conjunta $p(X, Y)$ (v.a. discretes) o la funció de densitat conjunt $f(x, y)$ (v.a. continues).

- Si dues variables són independents, tindrem que $\mathrm{cov}(X, Y) = 0$. El contrari no sempre és cert.
- Si $\mathrm{cov}(X, Y) \neq 0$, aleshores les dues variables no són independents.
  - $\mathrm{cov}(X, Y) > 0$ es diu que les dues variables estan associades positivament.
  - $\mathrm{cov}(X, Y) < 0$ es diu que les dues variables estan associades negativament.

Relacions entre dues v.a.'s

Si tenim dues variables aleatòries X i Y, aleshores per qualssevol valors reals α i β tenim que

- $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$, i
- $\mathrm{Var}(\alpha X + \beta Y) = \alpha^2 \mathrm{Var}(X) + \beta^2 \mathrm{Var}(Y) + 2\alpha\beta \mathrm{Cov}(X, Y)$.

Si a més tenim que X i Y són independents, aleshores $\mathrm{cov}(X, Y) = 0$, i per tant,

- $\mathrm{Var}(\alpha X + \beta Y) = \alpha^2 \mathrm{Var}(X) + \beta^2 \mathrm{Var}(Y)$.

Exemple: Dos daus i dos dards

Sigui X la suma de dos daus i Y la distància entre dos dards, definim una nova variable aleatòria $Z = \frac{X}{12} + Y$.

- El valor esperat de $Z, \mathbb{E}[Z]$, serà

$$\mathbb{E}[\frac{X}{12} + Y] = \frac{1}{12}\mathbb{E}[X] + \mathbb{E}[Y] = \frac{1}{12} \times 7 + (1 - \frac{\sqrt{2}}{2}) \approx 0.8762.$$

- Com que X i Y són independents, la variància de $Z, \mathrm{Var}(Z)$, serà

$$\mathrm{Var}(\frac{X}{12} + Y) = (\frac{1}{12})^2 \mathrm{Var}(X) + \mathrm{Var}(Y) = \frac{1}{12^2} \times 5.8352 + 0.0572 \approx 0.0977$$

Exemple: Dos daus i dos dards (2)

Anem a fer una simulació: fem 1000 tirades de dos daus i dos dards

```
d1 = sample(x = 1:6, size = 1000, replace = TRUE)
d2 = sample(x = 1:6, size = 1000, replace = TRUE)
X = d1+d2

d1 = runif(1000, min = 0, max = 1)
d2 = runif(1000, min = 0, max = 1)
Y = abs(d1-d2)

Z = X/12+Y
```

Si calculem els estadístics bàsics hauríem d’obtenir resultats semblants a les quantitats calculades:

```
mean(Z)
## [1] 0.9249409

var(Z)
## [1] 0.09548045
```