STT 1000 - STATISTIQUES

Arthur Charpentier

Average, mean, median, mode, etc

Moyenne (empirique)

Given a sample $\boldsymbol{y} = \{y_1, \cdots, y_n\}$, the average is $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

```
> import statistics
> y = [1, 2, 3, 4, 5, 6]
> print(statistics.mean(y))
3.5
```

```
> y = 1:6
> mean(y)
[1] 3.5
```

Average, mean, median, mode, etc

On peut montrer que $\overline{y}$ est solution de $\overline{y} = \underset{m\in\mathbb{R}}{\text{argmin}}\left\{\sum_{i=1}^n (y_i - m)^2\right\}$

Preuve: soit $\boldsymbol{y} = \{y_1, \cdots, y_n\}$, posons

$$h(m) = \sum_{i=1}^n (y_i - m)^2$$

$$\frac{\partial h(m)}{\partial m} = \frac{\partial}{\partial m}\sum_{i=1}^n (y_i - m)^2 = \sum_{i=1}^n \frac{\partial}{\partial m}(y_i - m)^2 = \sum_{i=1}^n -2(y_i - m)$$

Condition du premier ordre $\left.\frac{\partial h(m)}{\partial m}\right|_{m=m^\star} = 0$,

$$\sum_{i=1}^n (y_i - m^\star) = 0 \text{ si et seulement si } \sum_{i=1}^n y_i = nm^\star$$

i.e. $y^\star = \overline{y}$.

Average, mean, median, mode, etc

Espérance mathématique

The average (mean) is the empirical version of the expected value of a random variable,

$$\mathbb{E}(X) = \sum_x x\mathbb{P}[X = x] \text{ or } \int xf(x)dx$$

Example: a coin has *heads* with probability p. Let $x = \mathbf{1}(heads)$,

$$\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1 - p) = p$$

Linearity:

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b,\ \forall a, b \in \mathbb{R}, X$$

$$\mathbb{E}(X_1 + \cdots + X_k) = \mathbb{E}(X_1) + \cdots \mathbb{E}(X_k),\ \forall X_1, \cdots, X_k$$

Example: toss n coins, of bias p, X is the number of heads

$$\mathbb{E}(X) = \mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots \mathbb{E}(X_n) = np$$

Average, mean, median, mode, etc

$$\mathbb{E}(X) = \sum_x x\mathbb{P}[X = x] \text{ ou } \int xf(x)dx$$

$$\mathbb{E}(h(X)) = \sum_x h(x)\mathbb{P}[X = x] \text{ ou } \int h(x)f(x)dx$$

Exemple: pour une loi $\mathcal{N}(0,1)$, que vaut $\mathbb{E}[\cos[X]]$?

$$\mathbb{E}\big[\cos[X]\big] = \int_{-\infty}^{+\infty} \cos(x)\varphi(x)dx$$

```
> f = function(x) cos(x)*dnorm(x,0,1)
> integrate(f,-Inf,Inf)
0.6065307 with absolute error < 7.2e-08
> log(integrate(f,-Inf,Inf)$value)
[1] -0.5
```

Nonlinear transformation & Jensen Inequality

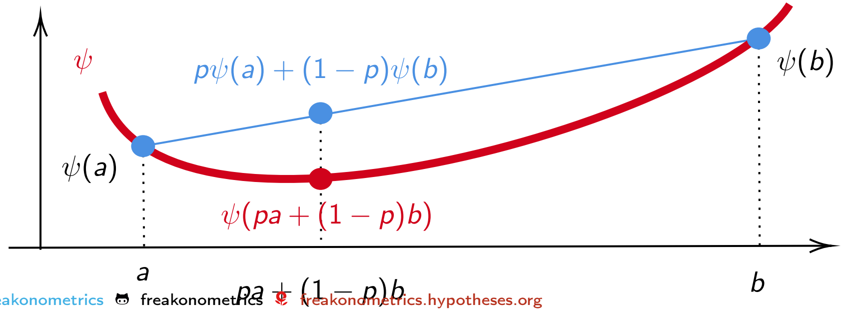
Let $\psi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}(\psi(X)) = \sum_x \psi(x)\mathbb{P}[X = x] \text{ or } \int \psi(x)f(x)dx \neq \psi\big(\mathbb{E}(X)\big)$$

Example if X takes values in $\{a, b\}$, with probability p and $1 - p$,

$$\mathbb{E}(\psi(X)) = \psi(a)p + \psi(b)(1 - p)$$

If ψ is a convex function, $\mathbb{E}(\psi(X)) \geq \psi\big(\mathbb{E}(X)\big)$



Nonlinear transformation & Jensen Inequality

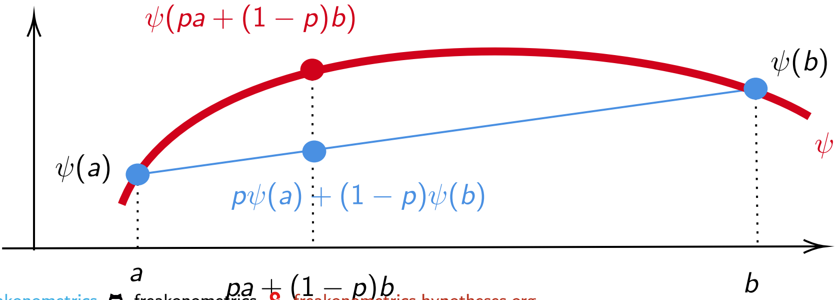
Let $\psi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}(\psi(X)) = \sum_x \psi(x)\mathbb{P}[X = x] \text{ or } \int \psi(x)f(x)dx \neq \psi\big(\mathbb{E}(X)\big)$$

Example if X takes values in $\{a, b\}$, with probability p and $1 - p$,

$$\mathbb{E}(\psi(X)) = \psi(a)p + \psi(b)(1 - p)$$

If ψ is a concave function, $\mathbb{E}(\psi(X)) \leq \psi\big(\mathbb{E}(X)\big)$



St Petersburg's Paradox

As we will see (law of large numbers) if x_i are realizations of random variables X_i (with identical expected value μ), $\overline{x} \to \mu$ as $n \to \infty$.

A fair coin is tossed at each stage. The initial stake begins at 2 dollars and is doubled every time heads appears. The first time tails appears, the game ends and the player wins whatever is in the pot. Let X denote the gain.

$$\mathbb{E}(X) = \frac{1}{2}\cdot 2 + \frac{1}{4}\cdot 4 + \frac{1}{8}\cdot 8 + \frac{1}{16}\cdot 16 + \cdots = 1+1+1+1+\cdots = +\infty$$

the expected value is infinite (but the average always exists)

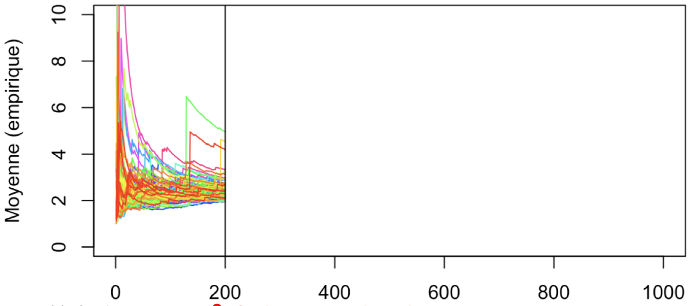
Average, mean, median, mode, etc

Example Loi de Pareto, $F(x) = 1 - x^{-\alpha}$ pour $x \geq 1$

La densité est $f(x) = \dfrac{\alpha}{x^{\alpha+1}}$ et l'espérance

$$\mathbb{E}[X] = \int_{x_m}^{\infty} \frac{x\alpha}{x^{\alpha+1}} dx = \begin{cases} \dfrac{\alpha}{\alpha - 1} & \text{si } \alpha > 1 \\ \infty & \text{si } \alpha \leq 1 \end{cases}$$

mais $\overline{x} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i$ est toujours fini ! Exemple pour $\alpha = 1.7$



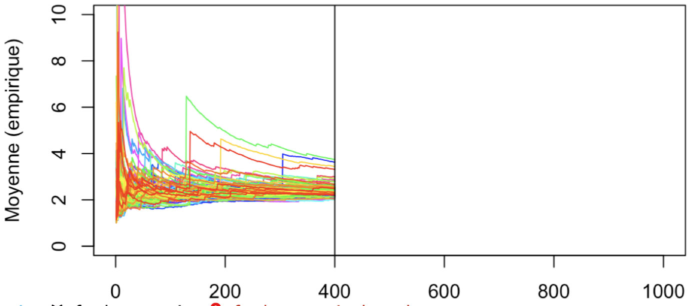
Average, mean, median, mode, etc

Example Loi de Pareto, $F(x) = 1 - x^{-\alpha}$ pour $x \geq 1$

La densité est $f(x) = \dfrac{\alpha}{x^{\alpha+1}}$ et l'espérance

$$\mathbb{E}[X] = \int_{x_m}^{\infty} \frac{x\alpha}{x^{\alpha+1}} dx = \begin{cases} \dfrac{\alpha}{\alpha - 1} & \text{si } \alpha > 1 \\ \infty & \text{si } \alpha \leq 1 \end{cases}$$

mais $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ est toujours fini ! Exemple pour $\alpha = 1.7$



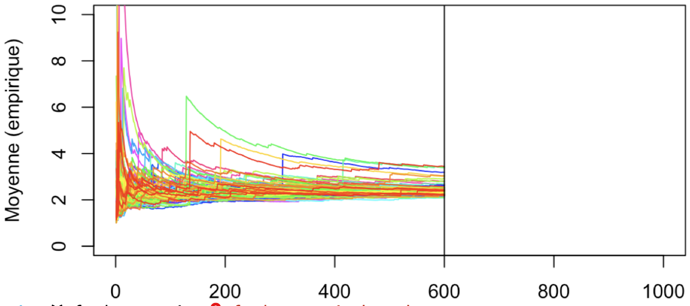
Average, mean, median, mode, etc

Example Loi de Pareto, $F(x) = 1 - x^{-\alpha}$ pour $x \geq 1$

La densité est $f(x) = \dfrac{\alpha}{x^{\alpha+1}}$ et l'espérance

$$\mathbb{E}[X] = \int_{x_m}^{\infty} \frac{x\alpha}{x^{\alpha+1}} dx = \begin{cases} \dfrac{\alpha}{\alpha - 1} & \text{si } \alpha > 1 \\ \infty & \text{si } \alpha \leq 1 \end{cases}$$

mais $\overline{x} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i$ est toujours fini ! Exemple pour $\alpha = 1.7$



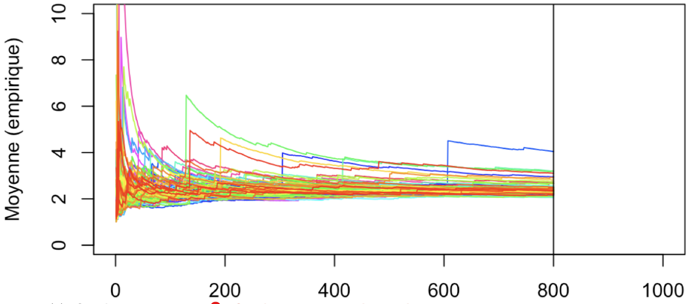
Average, mean, median, mode, etc

Example Loi de Pareto, $F(x) = 1 - x^{-\alpha}$ pour $x \geq 1$

La densité est $f(x) = \dfrac{\alpha}{x^{\alpha+1}}$ et l'espérance

$$\mathbb{E}[X] = \int_{x_m}^{\infty} \frac{x\alpha}{x^{\alpha+1}} dx = \begin{cases} \dfrac{\alpha}{\alpha - 1} & \text{si } \alpha > 1 \\ \infty & \text{si } \alpha \leq 1 \end{cases}$$

mais $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ est toujours fini ! Exemple pour $\alpha = 1.7$



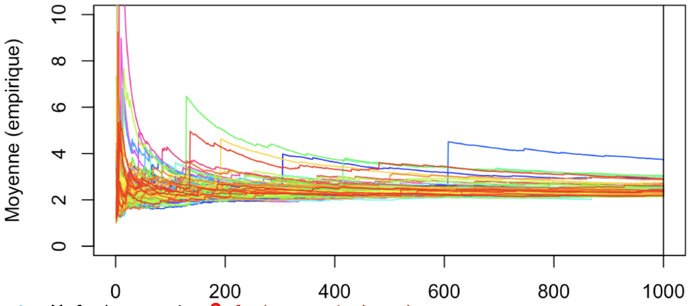
Average, mean, median, mode, etc

Example Loi de Pareto, $F(x) = 1 - x^{-\alpha}$ pour $x \geq 1$

La densité est $f(x) = \dfrac{\alpha}{x^{\alpha+1}}$ et l'espérance

$$\mathbb{E}[X] = \int_{x_m}^{\infty} \frac{x\alpha}{x^{\alpha+1}} dx = \begin{cases} \dfrac{\alpha}{\alpha - 1} & \text{si } \alpha > 1 \\ \infty & \text{si } \alpha \leq 1 \end{cases}$$

mais $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ est toujours fini ! Exemple pour $\alpha = 1.7$



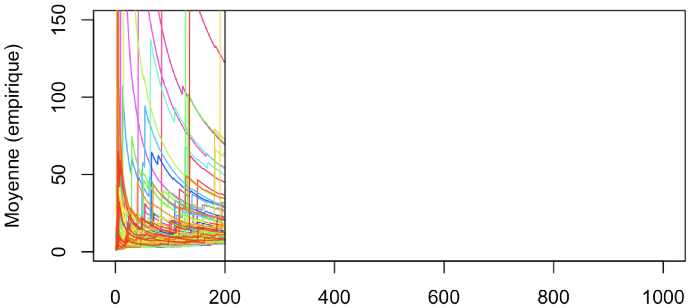
Average, mean, median, mode, etc

Example Loi de Pareto, $F(x) = 1 - x^{-\alpha}$ pour $x \geq 1$

La densité est $f(x) = \dfrac{\alpha}{x^{\alpha+1}}$ et l'espérance

$$\mathbb{E}[X] = \int_{x_m}^{\infty} \frac{x\alpha}{x^{\alpha+1}} dx = \begin{cases} \dfrac{\alpha}{\alpha - 1} & \text{si } \alpha > 1 \\ \infty & \text{si } \alpha \leq 1 \end{cases}$$

mais $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ est toujours fini ! Exemple pour $\alpha = 0.9$



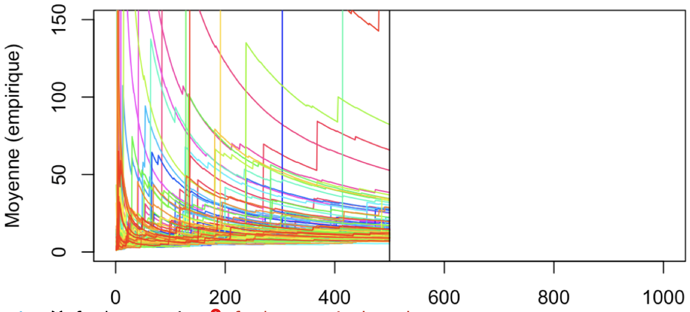
Average, mean, median, mode, etc

Example Loi de Pareto, $F(x) = 1 - x^{-\alpha}$ pour $x \geq 1$

La densité est $f(x) = \dfrac{\alpha}{x^{\alpha+1}}$ et l'espérance

$$\mathbb{E}[X] = \int_{x_m}^{\infty} \frac{x\alpha}{x^{\alpha+1}} dx = \begin{cases} \dfrac{\alpha}{\alpha - 1} & \text{si } \alpha > 1 \\ \infty & \text{si } \alpha \leq 1 \end{cases}$$

mais $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ est toujours fini ! Exemple pour $\alpha = 0.9$



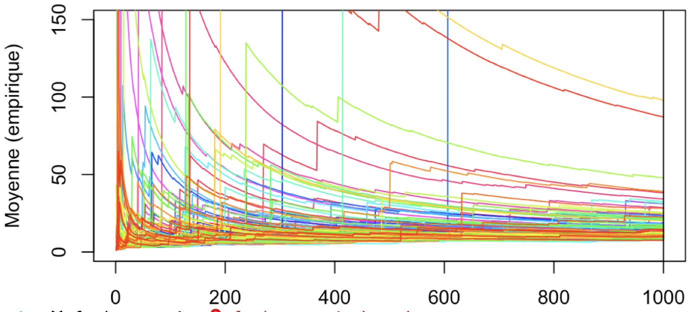
Average, mean, median, mode, etc

Example Loi de Pareto, $F(x) = 1 - x^{-\alpha}$ pour $x \geq 1$

La densité est $f(x) = \dfrac{\alpha}{x^{\alpha+1}}$ et l'espérance

$$\mathbb{E}[X] = \int_{x_m}^{\infty} \frac{x\alpha}{x^{\alpha+1}} dx = \begin{cases} \dfrac{\alpha}{\alpha-1} & \text{si } \alpha > 1 \\ \infty & \text{si } \alpha \leq 1 \end{cases}$$

mais $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ est toujours fini ! Exemple pour $\alpha = 0.9$



Average, mean, median, mode, quantiles etc

Quantile

Pour une loi F,

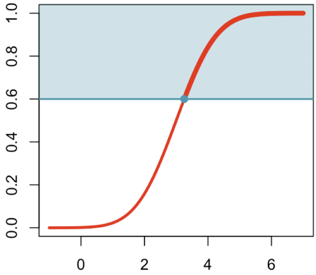
$$Q(p) = \inf\{x \in \mathbb{R} : p \leq F(x)\}$$

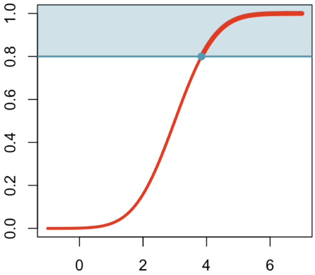
Le quantile est la seule fonction telle que

$$Q(p) \leq x \text{ si et seulement si } p \leq F(x)$$

Si F est continue et strictement croissante $Q(p) = F^{-1}(p)$.
Q est l'inverse à gauche: presque sûrement

$$Q(F(X)) \overset{p.s.}{=} X$$





Average, mean, median, mode, quantiles etc

Intégrale de la fonction quantile

Si l'espérance d'une variable de loi F existe,

$$\int_0^1 Q(p)dp = \int_{-\infty}^{\infty} xf(x)dx = \mathbb{E}[X]$$

En effet, par changement de variable $p = F(x)$,
$dp = F'(x)dx = f(x)dx$,

$$\int_0^1 F^{-1}(p)dp = \int_{-\infty}^{\infty} xf(x)dx = \mathbb{E}[X]$$

On peut aussi écrire

$$\mathbb{E}[X] = \mathbb{E}\big[(F^{-1}(U)\big] = \int_0^1 F^{-1}(u)du$$

où U suit une loi uniforme.

Average, mean, median, mode, quantiles etc

Exemple: pour une loi $\mathcal{N}(\mu, 1)$,

```
> mu = 7.3
> f = function(x) x*dnorm(x,mu,1)
> integrate(f,-Inf,Inf)
7.3 with absolute error < 3.3e-06
> g = function(p) qnorm(p,mu,1)
> integrate(g,0,1)
7.3 with absolute error < 8.1e-14
```

Définir les quantiles empiriques est plus compliqué

```
> ?quantile
> quantile(0:10,.95,type=7)
95%
9.5
> quantile(0:10,.95,type=3)
95%
  9
```

Average, mean, median, mode, quantiles etc

Quantile empirique (1)

Notons $\{x_{(i)}\}$ une version ordonnée de $\{x_i\}$, $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$.Le quantile empirique de niveau $p \in (0,1)$ est

$$\widehat{q}_p = (1-f)x_{(k)} + fx_{(k+1)}$$

où $k = \lceil np \rceil$ et $f = n\alpha - \lfloor np \rfloor$.

Quantile empirique (2)

Étant donné $\{x_i\}$, si $\widehat{F}$ est la fonction de répartition empirique associée, on peut poser

$$\widetilde{q}_p = \widehat{F}^{-1}(p) = x_{(k)} \text{ où } k = \lceil np \rceil$$

Average, mean, median, mode, etc

The average is very sensitive to outliers and extremal values.

Médiane

Notons $\{x_{(i)}\}$ une version ordonnée de $\{x_i\}$, $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. On appelle médiane $Q(1/2)$, et sa version empirique est

$$md(x) = \begin{cases} x_{((n+1)/2)} & \text{si } n \text{ pair} \\ \dfrac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right) & \text{si } n \text{ impair} \end{cases}$$

(50% observations are smaller/larger)

Note that $md(x) \in \underset{m\in\mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^{n} |x_i - m| \right\}$

Average & Paradoxes

See also Will Rogers phenomenon,
"*When the Okies left Oklahoma and moved to California, they raised the average intelligence level in both states*"

$$\{1,2,3,4,5\}\ \{6,7,8,9,10\}$$

$$\{1,2,3,4,5,6\}\ \{7,8,9,10\}$$

See The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer for real implications

Variance

Given a sample $\boldsymbol{x} = \{x_1, \cdots, x_n\}$, $s^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2$

Note that $s^2 = \displaystyle\min_{m\in\mathbb{R}}\left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2\right\}$

```
> import statistics
> x = [1, 2, 3, 4, 5, 6]
> statistics.variance(x)
3.5
> print(statistics.stdev(x))
1.8708286933869707
```

```
> x = 1:6
> var(x)
[1] 3.5
> sd(x)
[1] 1.870829
```

$s = \sqrt{s^2}$ is stdev($\boldsymbol{x}$) (standard deviation)

Dispersion, variance, standard deviation

Variance empirique

On appelle variance empirique $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$

Note C'est celle calculée par la plupart des logiciels

```
> x = 1:6
> var(x)
[1] 3.5
> sum((x-mean(x))^2)/5
[1] 3.5
> sum((x-mean(x))^2)/6
[1] 2.916667
```

C'est la version empirique de la variance

$$\text{Var}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \mathbb{E}\big(X^2\big) - \mathbb{E}(X)^2$$

Dispersion, variance, standard deviation

Note

$$\frac{1}{n-1}\sum_{i=1}^n (x_i-\overline{x})^2 = \frac{1}{n-1}\sum_{i=1}^n x_i^2 - \overline{x}^2 \neq \frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2$$

```
> x = 1:6
> var(x)
[1] 3.5
> mean(x^2)-mean(x)^2
[1] 2.916667
```

Example: toss a coin of bias p, with outcome $X \in \{0,1\}$,

$$\mathbb{E}(X)=p,\ \mathbb{E}(X^2)=p,\ \text{Var}(X)=p-p^2=p(1-p).$$

$$\text{Var}(aX+b)=a^2 Var(X),\ \forall a,b\in\mathbb{R},X$$

$\text{Var}(X_1+\cdots+X_k)=\text{Var}(X_1)+\cdots+\text{Var}(X_k)$, if X_i's are not correlated

See symmetric random walk, $X_i\in\{-1,+1\}$, $X=X_1+\cdots+X_n$, then

$$\mathbb{E}(X)=0,\ \text{Var}(X)=n \text{ and } \text{stdev}(X)=\sqrt{n}$$

Dispersion, variance, standard deviation

Example: The outcome of a (fair) six-sided die has expected value

$$\mathbb{E}[Y]=\sum_{i=1}^{6}\frac{1}{6}i=\frac{1+2+3+4+5+6}{6}=\frac{21}{6}=\frac{7}{2}$$

and variance

$$\text{Var}[Y]=\frac{1}{5}\sum_{i=1}^{6}\left(i-\frac{7}{2}\right)^2=\frac{1}{5}\left[\left(\frac{2-7}{2}\right)^2+\cdots+\left(\frac{12-7}{2}\right)^2\right]=\frac{7}{2}$$

```
> x = 1:6
> var(x)
[1] 3.5
> sd(x)
[1] 1.870829
```

Dispersion, variance, standard deviation

Écart-type (*standard deviation*)

L'écart-type empirique est $s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[x_i - \overline{x}]^2}$

Variance empirique (2)

On peut définir $\tilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$

Moments & Generating Function

1. Define the moment generating function, $M(t) = \mathbb{E}[e^{tX}]$,

$$M(t) = \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx$$

One can prove that

$$\frac{d^k M(t)}{dt^k} = \frac{d^k}{dt^k} \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx = \int_{-\infty}^{+\infty} \frac{d^k}{dt^k} e^{tx} f(x)\, dx$$

and $\left.\dfrac{d^k M(t)}{dt^k}\right|_{t=0} = \mathbb{E}[X^k]$, $\forall k \in \mathbb{N}$.

Example: si $X \sim \mathcal{E}(\lambda)$, $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$, pour $t < \lambda$

$$M(t) = \mathbb{E}[e^{tX}] = \int_0^{\infty} \lambda e^{-\lambda x} e^{tx} dx = \left[-\frac{\lambda}{\lambda - t} e^{-(\lambda - t)x} \right]_0^{\infty} = \frac{\lambda}{\lambda - t}$$

$$\frac{d^k M(t)}{dt^k} = \frac{\lambda k!}{(\lambda - t)^{k+1}} \text{ et } \mathbb{E}[X^k] = \frac{k!}{\lambda^k}$$

Moments & Generating Function

2. Define the moment generating function, $M(t)=\mathbb{E}[e^{tX}]$,

$$M(t)=\int_{-\infty}^{+\infty}e^{tx}f(x)\,dx=\int_{-\infty}^{+\infty}\left(1+tx+\frac{t^2x^2}{2}+\cdots\right)f(x)\,dx$$

$$M(t)=\int_{-\infty}^{+\infty}1f(x)\,dx+\int_{-\infty}^{+\infty}txf(x)\,dx+\int_{-\infty}^{+\infty}\frac{t^2x^2}{2}f(x)\,dx+\cdots$$

$$M(t)=\underbrace{\int_{-\infty}^{+\infty}f(x)\,dx}_{=1}+t\underbrace{\int_{-\infty}^{+\infty}xf(x)\,dx}_{=\mathbb{E}[X]}+\frac{t^2}{2}\underbrace{\int_{-\infty}^{+\infty}\frac{t^2x^2}{2}f(x)\,dx}_{=\mathbb{E}[X^2]}+\cdots$$

$$M(t)=1+t\mathbb{E}X+\frac{t^2}{2}\mathbb{E}X^2+\frac{t^3}{2\cdot 3}\mathbb{E}X^3+\cdots$$

so that $\left.\dfrac{d^kM(t)}{dt^k}\right|_{t=0}=\mathbb{E}[X^k],\ \forall k\in\mathbb{N}.$

Moments & Generating Function

Example: $X \sim \mathcal{P}(\lambda)$,

1. Espérance,

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k\, \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k\, \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}$$

$$= \lambda\, \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda\, \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda\, \mathrm{e}^{-\lambda}\, \mathrm{e}^{\lambda} = \lambda.$$

Note La fonction génératrice des moments s'écrit

$$M(t) = \sum_{k=0}^{\infty} \mathrm{e}^{tk} \mathbb{P}(X = k) = \sum_{k=0}^{\infty} \mathrm{e}^{tk} \frac{\lambda^k}{k!}\, \mathrm{e}^{-\lambda}$$

$$= \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda\, \mathrm{e}^t)^k}{k!} = \mathrm{e}^{-\lambda} \mathrm{e}^{\lambda\, \mathrm{e}^t} = \mathrm{e}^{\lambda(\mathrm{e}^t - 1)}.$$

Moments & Generating Function

$$\frac{dM(t)}{dt} = \lambda e^{\lambda(-1+e^t)+t} \text{ et } \left.\frac{dM(t)}{dt}\right|_{t=0} = \lambda$$

2. Variance,

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \sum_{k=1}^{\infty} k^2\, \mathbb{P}(X=k) - \lambda^2 = \sum_{k=1}^{\infty} k^2\, \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} - \lambda^2$$

$$= \lambda\, \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^{k-1}}{(k-1)!} - \lambda^2 = \lambda\, \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{d}{d\lambda} \frac{\lambda^k}{(k-1)!} - \lambda^2$$

$$= \lambda\, \mathrm{e}^{-\lambda} \frac{d}{d\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} - \lambda^2 = \lambda\, \mathrm{e}^{-\lambda} \frac{d}{d\lambda} \left[\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}\right] - \lambda^2$$

$$= \lambda\, \mathrm{e}^{-\lambda} \frac{d}{d\lambda} [\lambda\, \mathrm{e}^{\lambda}] - \lambda^2 = \lambda\, \mathrm{e}^{-\lambda} (\lambda+1)\, \mathrm{e}^{\lambda} - \lambda^2$$

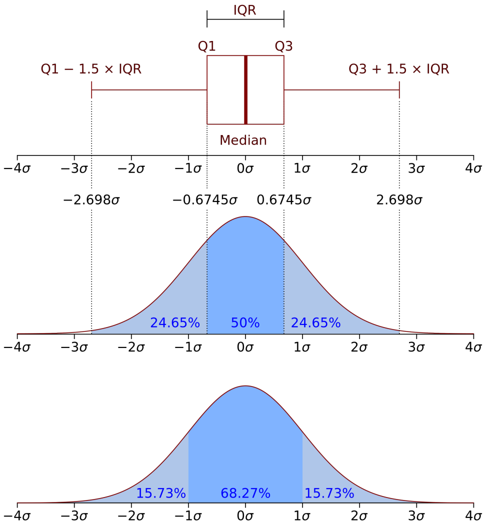
$$= \lambda\,(\lambda+1) - \lambda^2 = \lambda.$$

Intervalle interquartile

Interquartile range (IQR)

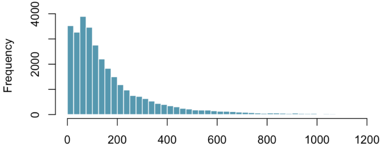
$$\text{IQR} = Q(3/4) - Q(1/4)$$

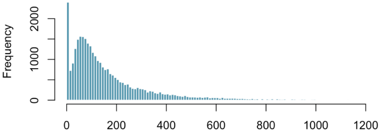
cf wikipedia



31,492 appels au service à la clientèle d'une banque

```
> hist(bankcall$Time[bankcall$Time<1200], breaks=seq
    (0,1200,by=10))
```






```
> mean(bankcall$Time/60)
[1] 3.143204
```

soit 3 minutes et 8 secondes pour la moyenne

```
> median(bankcall$Time/60)
[1] 1.916667
```

soit 1 minute et 55 secondes pour la médiane (115 sec.)

Statistiques descriptives

La moyenne tronquée (trimmed mean en anglais) à α (5% ou 10%) est la moyenne du jeu de données obtenu en supprimant une proportion α des plus petites valeurs et α plus grandes valeurs :

```
> mean(bankcall$Time/60,trim=.1)
[1] 2.357288
```

```
> mean(bankcall$Time/60)
[1] 3.143204
```

```
> quantile(bankcall$Time)
   0%    25%    50%    75%   100%
    1     57    115    225  28739
```

Calculs Formels

Exemple: $X \sim \mathcal{N}(1, 3^2)$ et $Y \sim \mathcal{E}(2)$, indépendantes. Que vaut $E_1 = \mathbb{E}[(X^2 - 1)Y]$? $E_1 = \dfrac{9}{2}$

Exemple: $\mathbb{P}[X \leq x] = \dfrac{x^2 - 2x + 2}{2}$ sur $[1, 2]$, 0 avant et 1 après. que vaut $E_2 = \mathbb{E}[X]$?

Exemple: $\mathbb{P}[X \leq x] = \dfrac{x}{8}$ sur $[0, 2)$, $\dfrac{x^2}{16}$ sur $[2, 4)$, 0 avant et 1 après. Que vaut $E_3 = \text{Var}[X - 1] + 1$? $E_3 = \dfrac{311}{144}$

Exemple:

$$f_X(x) = \begin{cases} 2(3 - 2x)/5 & \text{pour } 0 \leq x \leq 1 \\ 2(2 - x)/5 & \text{pour } 1 \leq x \leq 2 \\ 0 & \text{sinon} \end{cases}$$

Que vaut m_1 la médiane de X? $m_1 = \frac{1}{2}$

Calculs Numériques

Exemple: $X \sim \mathcal{P}(10)$, $p_1 = \mathbb{P}[X > 20]$? $p_1 \approx 0.001588$
Exemple: $X \sim \mathcal{B}(.1, 50)$, $p_2 = \mathbb{P}\big[X \leq \mathbb{E}(X)\big]$? $p_2 \approx 0.6161$
Exemple: $X \sim \mathcal{N}(10, 5)$, $p_3 = \mathbb{P}\big[X \leq 0\big]$? $p_3 \approx 0.02275$
Exemple: $X \sim \mathcal{B}eta(2, 3)$ et $Y \sim \mathcal{B}eta(5, 4)$, indépendantes, $p_4 = \mathbb{P}\big[X_1 + X_2 \leq 1\big]$? $p_4 \approx 0.5757$

Exemple: $X \sim \mathcal{N}(10, 5)$, q_1 tel que $\mathbb{P}\big[X > q_1\big] = 20\%$? $q_1 \approx 14.208$
Exemple: $X \sim \mathcal{B}eta(2, 3)$ et $Y \sim \mathcal{B}eta(5, 4)$, indépendantes, q_2 tel que $\mathbb{P}\big[X_1 + X_2 \leq q_2\big] = 50\%$? $q_2 \approx 0.9487$
Exemple: $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 1)$, indépendantes, q_3 tel que $\mathbb{P}\left[\dfrac{X_1}{X_2^2} > q_3\right] = 10\%$? $q_3 \approx 10.4079$
Exemple: $U_1, \cdots, U_4 \sim \mathcal{U}_{[0,1]}$, indépendantes, q_4 tel que $\mathbb{P}\left[X_1 + X_2 + X_3 + X_4 \leq q_4\right] = 10\%$? $q_4 \approx 1.2465$