Espérance, variance, quantiles

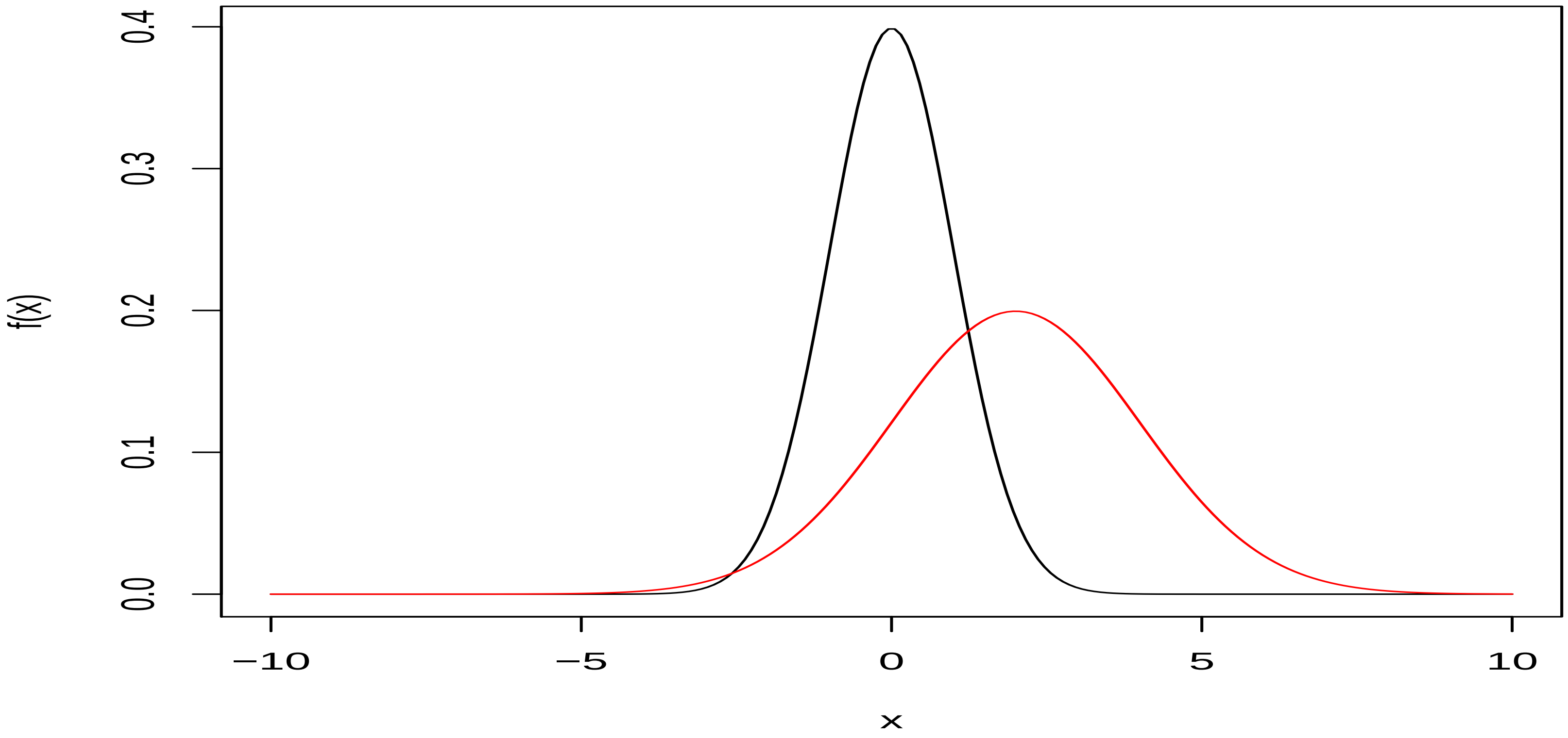
Mathématiques Générales B
Université de Genève

Sylvain Sardy

22 mai 2008

0. Motivation

Mesures de centralité (ex. espérance) et de dispersion (ex. variance)



1. Espérance

Définition. L'espérance d'une v.a. X (si elle existe) est :

– Cas discret : $\mathrm{E}(X) = \sum_x x\mathrm{P}(\mathrm{X} = \mathrm{x})$.

– Cas continu : $\mathrm{E}(X) = \int xf(x)dx$.

C'est une mesure du centre de la distribution. On l'a note souvent $\mu = \mathrm{E}(X)$.

Exemple : un jeu de hasard est programmé selon la table suivante

Gain	Probabilité	Gain $\times$ Proba
10	.01	0.1
50	.001	0.05
1'000	1.0 $\times 10^{-5}$	0.01
10'000	5 $\times 10^{-7}$	5.0 $\times 10^{-3}$
somme	0.0110105	0.165

Quelle est l'espérance de gain X d'un joueur ?

Distributions discrètes :

Si $X \sim$ **Bernoulli**(p), alors son espérance est $\mathrm{E}(X) =$

Si $X \sim$ **Binomiale(n,p)**, alors son espérance est $\mathrm{E}(X) =$

Si $X \sim$ **Poisson(**λ**)**, alors son espérance est $\mathrm{E}(X) =$

Distributions continues :

Si $X \sim$ **Uniforme(a,b)**, alors son espérance est $\mathrm{E}(X) =$

Si $X \sim$ **Gaussienne N(0,1)**, alors son espérance est $\mathrm{E}(X) =$

Si $X \sim$ **Gaussienne N(**μ**,** σ^2**)**, alors son espérance est $\mathrm{E}(X) =$

Exemple : Roger Federer joue ses matchs en 3 sets 60% du temps, 4 sets 25% et 5 sets 15%. Quelle est l'espérance du nombre de sets joués par le numéro 1 ?

Exemple : Le temps d'attente en minute au standard téléphonique d'un cinéma est bien modélisé par une distribution exponentielle $\mathrm{Exp}(0.5)$ (de densité $f(t) = \lambda \exp(-\lambda t)$, $t > 0$ avec $\lambda = 0.5$). Quel est le temps moyen d'attente ?

Espérance d'une fonction de v.a.

Propriété 1.1 (sans démonstration) :

– Cas discret : Si X a une distribution discrète et $Y = r(X)$, alors

$$\mathrm{E}(r(X)) = \mathrm{E}(Y) = \sum_y y\mathrm{P}(\mathrm{Y} = \mathrm{y}) = \sum_{\mathrm{x}} \mathrm{r}(\mathrm{x})\mathrm{P}(\mathrm{X} = \mathrm{x}).$$

– Cas continu : Si X a pour densité f_X et $Y = r(X)$, alors

$$\mathrm{E}(r(X)) = \mathrm{E}(Y) = \int y f_Y(y)dy = \int r(x) f_X(x)dx.$$

Définition : (utilisant $r(x) = x^2$) le **moment d'ordre** k d'une v.a. X est

$$\mathrm{E}(X^k) =$$

Exemple : Si $X \sim$ **Bernoulli**(p), alors son moment d'ordre 2 est

$$\mathrm{E}(X^2) = 0^2\mathrm{P}(\mathrm{X} = 0) + 1^2\mathrm{P}(\mathrm{X} = 1) = \mathrm{p}.$$

Exemple : Si $X \sim$ **Poisson**(λ), alors son moment d'ordre 2 est

$$\mathrm{E}(X^2) = \lambda(\lambda + 1).$$

Exemple : Si $X \sim$ **Exp**(λ), alors son moment d'ordre 2 est

$$\mathrm{E}(X^2) = \int_0^\infty x^2 \lambda \exp(-\lambda x) dx = \frac{2}{\lambda^2}.$$

Pour deux v.a. X et Y et une fonction $r(x, y)$, on a

– Cas discret :

$$\mathrm{E}(r(X, Y)) = \sum_x \sum_y r(x, y) \mathrm{P}(\mathrm{X} = \mathrm{x}, \mathrm{Y} = \mathrm{y})$$

– Cas continu :

$$\mathrm{E}(r(X, Y)) = \int \int r(x, y) f(x, y) dx\ dy$$

Propriété 1.2 : si a et b sont des constantes, alors

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b.$$

Démonstration :

Propriété 1.3 : $\mathrm{E}\{X - \mathrm{E}(X)\} = 0.$

Propriété 1.4 : Soit X et Y deux v.a. de densité conjointe $f(x, y)$. Alors

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y).$$

Cette propriété est également vraie dans le cas discret.

Cette propriété est vraie que les v.a. soient indépendantes ou non.

Démonstration : on utilise la fonction $r(x, y) = x + y$:

$$\begin{aligned} \mathrm{E}(X+Y) &= \int\int (x+y) f(x, y) dx\ dy \\ &= \end{aligned}$$

Exemple : Quelle est l'espérance d'une **Binomiale(n,p)** ?

Exemple : n hommes et n femmes dansent par deux et forment une paire aléatoirement. Quelle est l'espérance du nombre de couples D_n qui dansent ensemble parmi les n couples présents ? Cette espérance augmente-t-elle avec n ?

Soit X_i la variable binaire telle que $X_i = 1$ si le $i$ème homme danse avec sa femme ($X_i = 0$ sinon). Par la combinatoire on sait que $\mathrm{P(X_i = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}}$, donc $\mathrm{E}(X_i) =$

De plus $D_n = X_1 + \ldots + X_n$, donc

$$\mathrm{E}(D_n) =$$

Propriété 1.5 : Si X et Y sont deux v.a. indépendantes, alors

$$\mathrm{E}(XY) = \mathrm{E}(X)\ \mathrm{E}(Y).$$

Démonstration :

Exemple : Soit X et Y de distribution conjointe

X	Y=1	0
1	0	.3
0	.5	.2

$\mathrm{E}(XY) = (1)(0) + (0)(.3) + (0)(.5) + (0)(.2) = 0$.

$\mathrm{E}(X) = (1)(.3) + (0)(.7) = .3$ et $\mathrm{E}(Y) = (1)(.5) + (0)(.5) = .5$.

Que se passe-t-il ?

2. Variance

Définition : La **variance** d'une v.a. X (si elle existe) est

$$\mathrm{var}(X) = \mathrm{E}\{X - \mathrm{E}(X)\}^2.$$

C'est une mesure de dispersion autour de l'espérance. On la note souvent $\sigma^2 = \mathrm{var}(X)$.

Définition équivalente : $\mathrm{var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2$

Démonstration :

Exemple : Si $X \sim$ **Bernoulli**(p), alors sa variance est

$$\mathrm{var}(X) =$$

Exemple : Si $X \sim$ **Poisson**(λ), alors sa variance est

$$\mathrm{var}(X) =$$

Exemple : Si $X \sim$ **Exp**(λ), alors sa variance est

$$\mathrm{var}(X) =$$

Exemple : Si $X \sim$ **N**(0,1), alors sa variance est

$$\mathrm{var}(X) = 1$$

Propriété 2.1 : si a et b sont des constantes, alors

$$\mathrm{var}(aX + b) = a^2 \mathrm{var}(X).$$

Démonstration :

Conséquence (Propriétés 1.2 et 2.1) : si X est une variable aléatoire telle que $\mathrm{E}(X) = 0$ et $\mathrm{var}(X) = 1$, alors $Y = \mu + \sigma X$ est telle que $\mathrm{E}(X) =$ et $\mathrm{var}(Y) =$.

Exemple : Si $X \sim N(0,1)$, alors $Y = \mu + \sigma X \sim \mathrm{N}(\mu, \sigma^2)$. Pratique : la distribution Gaussienne est paramétrisée avec son espérance μ et sa variance σ^2. Sa fonction de densité est

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2).$$

Exemple : $X \sim \mathrm{Poisson}(\lambda)$. Que peut-on dire de $Y = aX$?

Définition : l'**écart type** d'une v.a. X est $\sigma = \sqrt{\mathrm{var}(X)}$.

L'écart type a l'avantage d'être dans la même unité que la variable. Par exemple si X est en m, alors $\mathrm{var}(X)$ est en m^2, mais $\sigma(X)$ est en m.

Exemple : on lance un dé et notons X la valeur. Quelles sont son espérance et sa variance ?

$$\begin{aligned} \mathrm{E}(X) &= \\ \mathrm{var}(X) &= \end{aligned}$$

Donc $\mathrm{var}(X) = \frac{105}{36} = 2.9166$ et $\sigma(X) = \sqrt{\mathrm{var}(X)} = 1.7078$.

Une mesure de déviation plus naturelle est $\mathrm{E}(|X - \mathrm{E}(X)|) = 1.5$. En effet

$$|X - \mu| = \begin{cases} 0.5 & \text{when } X = 3, 4 \\ 1.5 & \text{when } X = 2, 5 \\ 2.5 & \text{when } X = 1, 6 \end{cases}$$

Donc $\mathrm{E}(|X - \mu|) = (0.5)(\frac{1}{6} + \frac{1}{6}) + (1.5)(\frac{1}{6} + \frac{1}{6}) + (2.5)(\frac{1}{6} + \frac{1}{6}) = 1.5$.

Définition : La **covariance** entre deux v.a. X et Y est

$$\operatorname{cov}(X,Y) = \mathrm{E}\{(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\}.$$

Il est facile de montrer la définition équivalente

$$\operatorname{cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

Conséquence (Propriété 1.5) : Si X et Y sont deux v.a. indépendantes, alors

$$\operatorname{cov}(X,Y) = \quad .$$

Propriété 2.2 : $\operatorname{var}(X+Y) =$

Donc si X et Y sont deux v.a. indépendantes, alors $\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y)$.

Définition (plus faible que l'indépendance) : deux v.a. X et Y sont non-corrélées si $\text{cov}(X, Y) = 0$.

Il suffit donc que X et Y soient non-corrélées pour que $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

Exemple : Quelle est la variance de la Binomiale(n,p) ?

Exemple : Soit X une v.a. qui prend pour valeur $-2, -1, 0, 1, 2$ avec probabilité $1/5$, et soit $Y = X^2$.

$\text{cov}(X, Y) =$

Les v.a. X et Y sont-elles indépendantes ?

Définition : le **coefficient de corrélation** entre X et Y

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)} \in [-1,1].$$

mesure la force d'associaton linéaire entre deux v.a.

IMPORTANT : faites toujours la distinction entre indépendant et non-corrélé. Ne concluez jamais pour l'indépendance après avoir calculer un coefficient de corrélation nul !

3. Quantiles

Définition : La **médiane** m d'une v.a. X satisfait

$$\mathrm{P}(\mathrm{X} \leqslant \mathrm{m}) \geqslant 1/2 \quad \text{et} \quad \mathrm{P}(\mathrm{X} \geqslant \mathrm{m}) \geqslant 1/2.$$

Propriété 3.1 : Si X est une v.a. avec une fonction de distribution F continue et strictement croissante, alors

$$m = F^{-1}(0.5).$$

Propriété 3.2 : Si X est une v.a. avec une fonction de densité symmétrique, alors l'espérance de X est égale à sa médiane.

Exemple : si $X \sim \mathrm{U}(0,1)$, alors $m =$.

Exemple : si $X \sim \mathrm{N}(\mu, \sigma^2)$, alors $m =$.

Exemple : si $X \sim \mathrm{Exp}(\lambda)$, alors $m = \quad < \mathrm{E}(X) = \frac{1}{\lambda}$.

Modélisation statistique : chercher une distribution F qui ressemble le plus aux données.

Exemple : $x_1, \ldots, x_{108}$ de la taille des étudiants en Math Générale 2007.

On peut ensuite :

- Déterminer la taille médiane $m = T_{\text{Médiane}}$ telle que 50% de la population soit en dessous/dessus, c'est-à-dire :

$$\mathrm{P(T \leqslant m) = 0.5 = F(m)}$$

ou

$$m = F^{-1}(0.5).$$

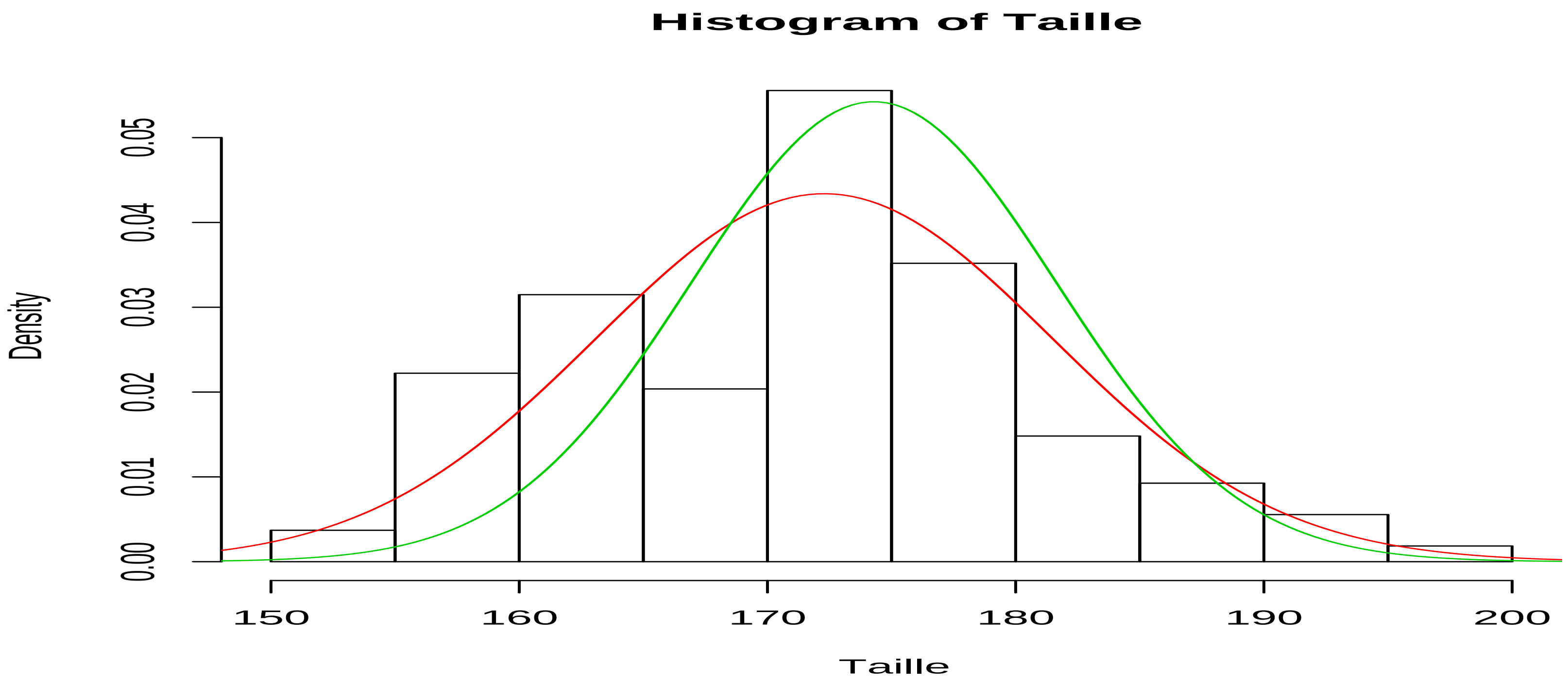
C'est le quantile à 50%.

– Déterminer 3 tailles de pantalon T_S, T_M, T_L pour obtenir 4 groupes équiprobables

$$\mathrm{P}(\mathrm{T} \leqslant \mathrm{T_S}) = \mathrm{P}(\mathrm{T_S} < \mathrm{T} \leqslant \mathrm{T_M}) = \mathrm{P}(\mathrm{T_M} < \mathrm{T} \leqslant \mathrm{T_L}) = \mathrm{P}(\mathrm{T_L} < \mathrm{T}) = \frac{1}{4}.$$

Comment trouver une bonne fonction f ou F : l'analyse exploratoire (EDA).

Première approche : considérer des fonctions de densité $f(x)$ et les comparer à l'histogramme.



Deuxième approche : considérer des fonctions de répartition F et comparer la médiane théorique à la médiane empirique

$$F^{-1}(0.5) = m \quad \overset{?}{\approx} \quad x_{.5}.$$

Rappel : les statistiques d'ordre empiriques $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ sont les données ordonnées, c'est-à-dire $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$.

La médiane empirique : Valeur telle que 50% des données sont plus petites/grandes

$$x_{.5} = \begin{cases} x_{(\frac{n+1}{2})} & \text{si } n \text{ est impaire,} \\ \frac{1}{2}(x_{(n/2)} + x_{(1+n/2)}) & \text{si } n \text{ est paire.} \end{cases}$$

On peut aussi comparer les quartiles inférieur et supérieur théoriques aux empiriques

$$F^{-1}(0.25) \quad \stackrel{?}{\approx} \quad x_{.25} \qquad \text{et} \qquad \mathrm{F}^{-1}(0.75) \quad \stackrel{?}{\approx} \quad \mathrm{x}_{.75}.$$

Tous les quantiles : prendre 108 valeur de probabilités équiréparties $\frac{1}{109}, \frac{2}{109}, \ldots, \frac{108}{109}$ et comparer

$$F^{-1}(\frac{k}{109}) \quad \stackrel{?}{\approx} \quad x_{\frac{k}{109}}.$$

On obtient un **graphique quantile-quantile** (q-q plot). Si F est une bonne fonction de répartition pour nos données, alors les points devraient être grosso-modo alignés le long de la droite $y = x$ (ligne pointillée).

La distribution F de référence pour un q-q plot est souvent celle de $\mathrm{N}(\mu, \sigma^2)$: