Modèles et estimation

Jean-Marc.Vincent@imag.fr

ID-IMAG Laboratory

http://www-id.imag.fr/ jvincent

Modélisation

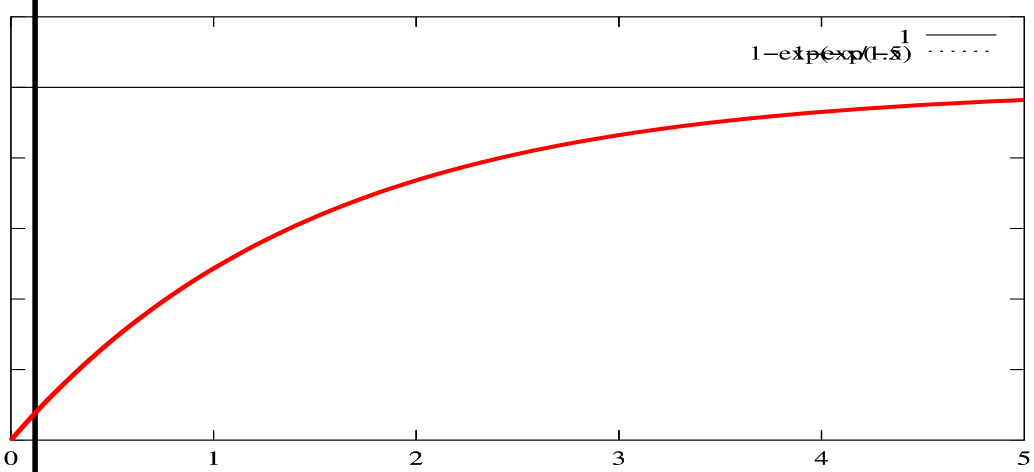
Hypothèse : les données sont considérées comme une séquence de valeurs aléatoires indépendantes et de même loi statistique
Danger : vérifier que les conditions expérimentales garantissent celà.
Représentation par une variable aléatoire X.
Loi de X :
Fonction de répartition : probability distribution function (pdf)

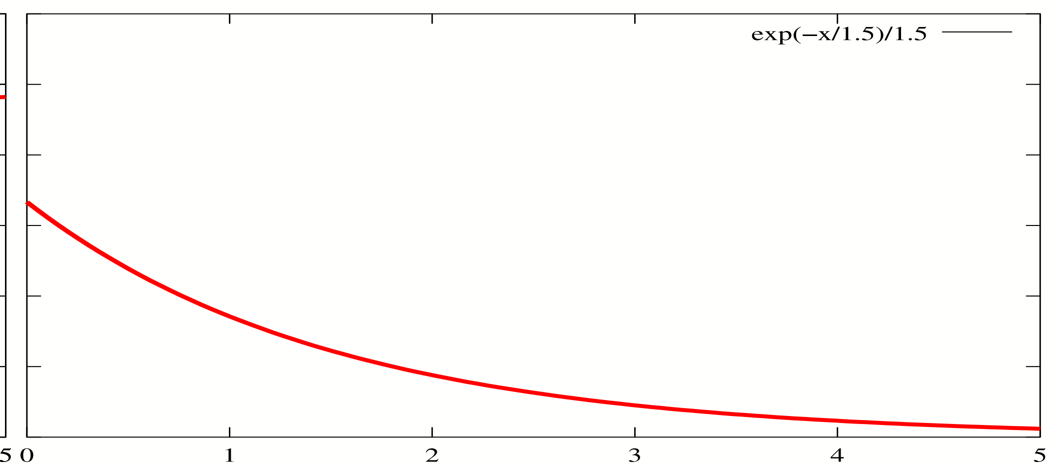
$$F_X(x) = \mathbb{P}(X \leq x);$$

Densité : density function

$$f_X(x) = F'_X(x) = \mathbb{P}(x \leq X \leq x + dx)/dx.$$

Exemple : loi exponentielle $f_X(x) = \frac{2}{3}e^{-\frac{2}{3}x}$





$$Mode = f_X^{-1}(\max_x f_X(x)) = 0;$$

$$Médiane = F_X^{-1}(0.5) \simeq 1.04;$$

$$Moyenne = \int x f_X(x)dx = \mathbb{E}X = 1.5.$$

Loi des grands nombres

Theorem 1 (Loi forte des grands nombres)
Soit $\{X_n\}_{n\in\mathbb{N}}$ *suite de variables aléatoires iid telles que* $\mathbb{E}X^2 < +\infty$ *alors*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}X, \quad P-ps \text{ et dans } L^1.$$

$\to$ loi de convergence des fréquences empiriques
- pour toute expérimentation on a le résultat
- convergence en loi

$$\text{Notation : } \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Théorème de la Limite Centrale

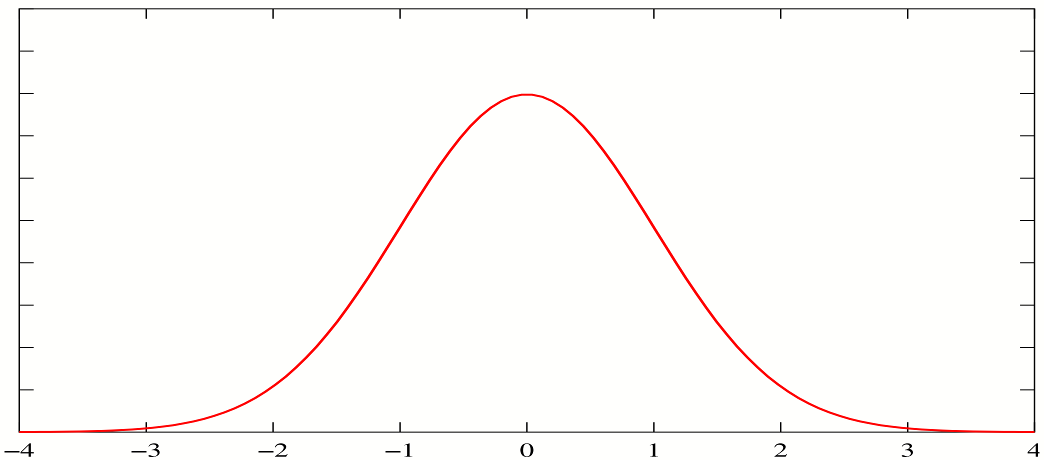
Theorem 2
Soit $\{X_n\}_{n\in\mathbb{N}}$ *suite de variables aléatoires iid telles que* $\sigma^2 = VarX = \mathbb{E}X^2 - (\mathbb{E}X)^2 < +\infty$ *alors*

$$\lim_{n\to+\infty} \frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mathbb{E}X\right) \stackrel{L}{=} N(0,1).$$

$\longrightarrow$ loi des erreurs (loi de Gauss, loi Normale,...)

$\longrightarrow$ moyenne 0, variance 1...

Loi normale $N(0,1)$



Distribution :

$$\mathbb{P}(X \in [-1,1]) = 0.68;$$

$$\mathbb{P}(X \in [-2,2]) = 0.95;$$

$$\mathbb{P}(X \in [-3,3]) < 0.01.$$

Quantiles

$$\mathbb{P}(X \notin [-1.65, 1.65]) = 0.10;$$

$$\mathbb{P}(X \notin [-1.96, 1.96]) = 0.05;$$

$$\mathbb{P}(X \notin [-2.57, 2.57]) = 0.01.$$

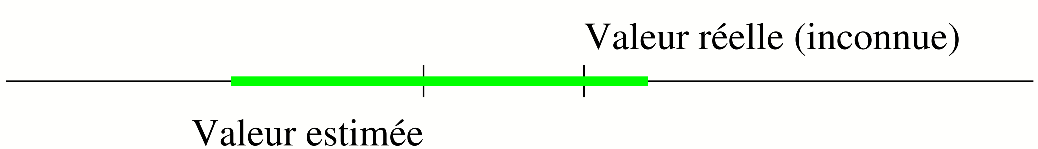
Intervalles de confiance

Niveau de confiance $\alpha \rightarrow$ on calcule ϕ_α tel que

$$\mathbb{P}(X \in [-\phi_\alpha, \phi_\alpha]) = \alpha$$

Pour n assez grand ($n > 50$)

$$\mathbb{P}\left([\overline{X}_n - \frac{\phi_\alpha \sigma}{\sqrt{n}}, \overline{X}_n + \frac{\phi_\alpha \sigma}{\sqrt{n}}] \ldots . \mathbb{E}X \right) = \alpha.$$



Intervalles de confiance (2)

Estimateur de la variance:

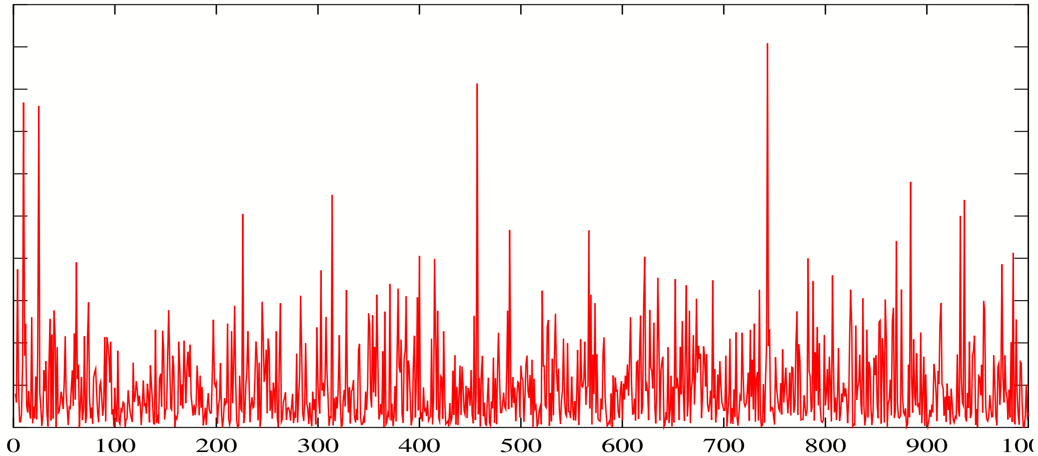
$$\hat{\sigma}_n = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2 .$$

Danger : si n petit (< 30) $\Rightarrow$ faire hypothèse de normalité de X puis approximation IC par Student.

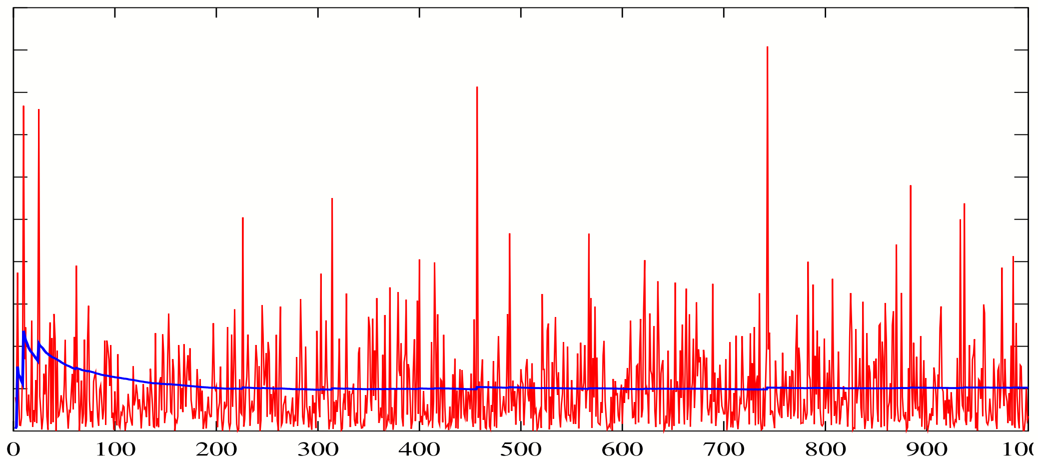
On peut faire des tests d'hypothèse (comparatifs)

Méthode

Valeurs

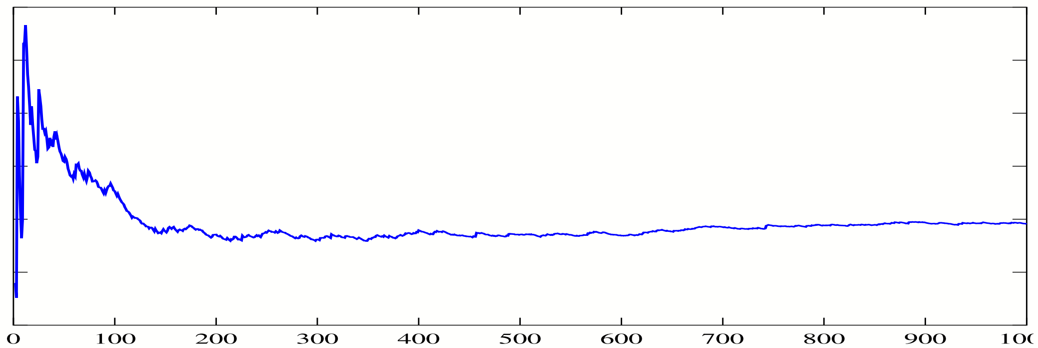


Valeurs et estimateur de la moyenne

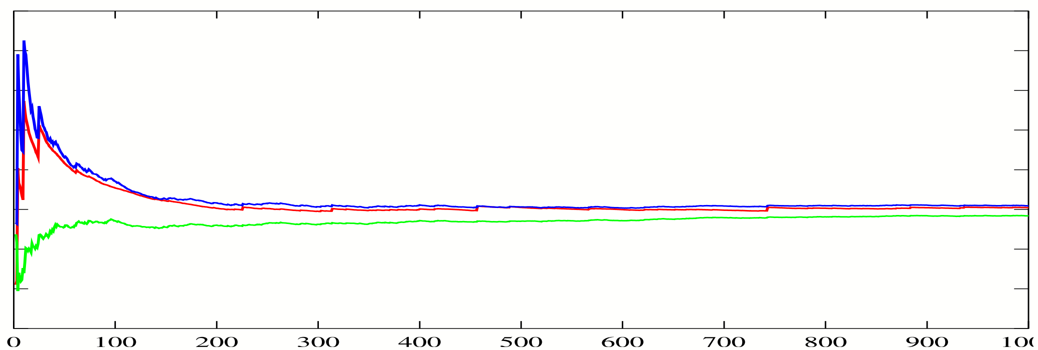


Estimation

Estimation de l'écart-type



Intervalles de confiance



Stabilisation de $\hat{\sigma}_n$ puis calcul d'intervalle de confiance