# Cours 5 – Estimateur du Maximum de Vraisemblance.

*Eya ZOUGAR* *

***Institut National des Sciences appliqués-INSA***

Génie mathématiques GM3
Thursday 16th February, 2023

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN NORMANDIE

*Basé sur le cours de Bruno PORTIER

# 1. Introduction.

La méthode du maximum de vraisemblance est une méthode d'estimation.

Elle consiste à :

- construire des estimateurs performants ;
- construire des intervalles de confiance précis ;
- mettre en oeuvre des tests statistiques puissants.

C'est une méthode très performante mais qui nécessite, à la différence d'autres méthodes comme la méthode des moments ou la méthode des moindres carrés, de connaitre la loi des variables aléatoires mises en jeu.

# 2. Notions de vraisemblance.
## 2.1. Cas des variables discrètes.

Soient $x_1, x_2, \ldots, x_n$ les mesures d'une variable quantitative à valeurs discrètes.

On suppose que ces données sont les réalisations de $n$ variables aléatoires $X_1, X_2, \ldots, X_n$ de lois discrète, dépendant d'un paramètre $\theta$.

**Vraissemblance (Likelihood en anglais)**

On appelle vraisemblance de l'échantillon $x_1, x_2, \ldots, x_n$, la probabilité d'observer cet échantillon. On la note généralement $L(x_1, x_2, \ldots, x_n, \theta)$ et on a:

$$L(x_1, \ldots, x_n, \theta) = \mathbb{P}_\theta \left[X_1 = x_1, \ldots, X_n = x_n\right]$$

# 2.2. Cas des variables discrètes indépendants.

**Vraissemblance (Likelihood en anglais)**

Si les variables $(X_j)$ sont indépendantes, la vraisemblance dans le cas discréts, peut alors se mettre sous la forme:

$$L(x_1, \ldots, x_n, \theta) = \prod_{j=1}^{n} \mathbb{P}_\theta \left[X_j = x_j\right]$$

**Exemples:** On suppose que les variables aléatoires $(X_i)_i$ sont i.i.d.

1. Loi de Bernouilli: $X_i \sim \mathcal{B}(p)$, $\forall 1 \leq i \leq n$

# 2.2. Cas des variables discrètes indépendants.

### Vraissemblance (Likelihood en anglais)

Si les variables $(X_j)$ sont indépendantes, la vraisemblance dans le cas discréts, peut alors se mettre sous la forme:

$$L(x_1, \ldots, x_n, \theta) = \prod_{j=1}^{n} \mathbb{P}_\theta \left[X_j = x_j\right]$$

**Exemples:** On suppose que les variables aléatoires $(X_i)_i$ sont i.i.d.

1. Loi de Bernouilli: $X_i \sim \mathcal{B}(p)$, $\forall 1 \leq i \leq n$

$$L(x_1, \ldots, x_n, p) = \prod_{i=1}^{n} (1-p)^{1-x_i}\, p^{x_i} = (1-p)^{n-n\bar{x}}\, p^{n\bar{x}}$$

avec $\bar{x} = \dfrac{1}{n} \sum_{i=1}^{n} x_i$.

# Exemples.

2. Loi de Binomiale: $X_i \sim \mathcal{B}(N, p),\ \forall 1 \leq i \leq n$

# Exemples.

2. Loi de Binomiale: $X_i \sim \mathcal{B}(N, p), \ \forall 1 \leq i \leq n$

$$
\begin{aligned}
L(x_1, \ldots, x_n, p) &= \prod_{i=1}^{n} C_N^{x_i} (1-p)^{N-x_i} p^{x_i} \\
&= \left( \prod_{i=1}^{n} C_N^{x_i} \right) (1-p)^{nN - n\bar{x}} p^{n\bar{x}}
\end{aligned}
$$

3. Loi Géométrique: $X_i \sim \mathcal{G}(p), \ \forall 1 \leq i \leq n$

## Exemples.

2. Loi de Binomiale: $X_i \sim \mathcal{B}(N, p), \ \forall 1 \leq i \leq n$

$$L(x_1, \ldots, x_n, p) = \prod_{i=1}^{n} C_N^{x_i} (1-p)^{N-x_i} p^{x_i} = \left(\prod_{i=1}^{n} C_N^{x_i}\right) (1-p)^{nN-n\bar{x}} p^{n\bar{x}}$$

3. Loi Géométrique: $X_i \sim \mathcal{G}(p), \ \forall 1 \leq i \leq n$

$$L(x_1, \ldots, x_n, p) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p = (1-p)^{n\bar{x}-n} p^n$$

4. Loi de Poisson: $X_i \sim \mathcal{P}(\lambda), \ \forall 1 \leq i \leq n$

# Exemples.

2. Loi de Binomiale: $X_i \sim \mathcal{B}(N,p),\ \forall 1 \leq i \leq n$

$$\begin{aligned} L(x_1,\dots,x_n,p) &= \prod_{i=1}^{n} C_N^{x_i}\,(1-p)^{N-x_i}\,p^{x_i} \\ &= \left(\prod_{i=1}^{n} C_N^{x_i}\right)\,(1-p)^{nN-n\bar{x}}\,p^{n\bar{x}} \end{aligned}$$

3. Loi Géométrique: $X_i \sim \mathcal{G}(p),\ \forall 1 \leq i \leq n$

$$L(x_1,\dots,x_n,p) = \prod_{i=1}^{n}(1-p)^{x_i-1}\,p = (1-p)^{n\bar{x}-n}\,p^n$$

4. Loi de Poisson: $X_i \sim \mathcal{P}(\lambda),\ \forall 1 \leq i \leq n$

$$L(x_1,\dots,x_n,\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}\,e^{-\lambda}}{x_i!} = \frac{\lambda^{n\bar{x}}\,e^{-n\lambda}}{\prod_{i=1}^{n}(x_i!)}$$

# 2.3. Vraisemblance dans le cas de variables continues.

On considére un échantillon $\{x_1, \ldots, x_n\}$ de $n$ réalisations de variables aléatoires $X_1, X_2, \ldots, X_n$ de lois continus de fonction de densité $f$, dépendant d'un paramètre $\theta$.

**Vraissemblance (Likelihood en anglais)**

On appelle vraisemblance de l'échantillon $x_1, x_2, \ldots, x_n$, la probabilité d'observer cet échantillon. On la note généralement $L(x_1, x_2, \ldots, x_n, \theta)$ et on a:

$$L(x_1, \ldots, x_n, \theta) = f_\theta(x_1, \ldots, x_n)$$

Si les variables $(X_j)$ sont i.i.d, la vraisemblance dans le cas continus, peut alors se mettre sous la forme:

$$L(x_1, \ldots, x_n, \theta) = \prod_{j=1}^{n} f_\theta(x_j)$$

# Exemples.

On suppose que les variables aléatoires $(X_i)_i$ sont i.i.d.

1. Loi de Uniform: $X_i \sim \mathcal{U}([0, a]), \ \forall 1 \leq i \leq n$

# Exemples.

On suppose que les variables aléatoires $(X_i)_i$ sont i.i.d.

1. Loi de Uniform: $X_i \sim \mathcal{U}([0, a]), \ \forall 1 \leq i \leq n$
$$L(x_1, \ldots, x_n, a) = \prod_{i=1}^{n} \frac{1}{a} \mathbf{1}_{[0,a]}(x_i) = \left(\frac{1}{a}\right)^n, \ \forall\, 0 \leq x_i \leq a$$
2. Loi de Exponentielle: $X_i \sim \mathcal{E}(\lambda), \ \forall 1 \leq i \leq n$

## Exemples.

On suppose que les variables aléatoires $(X_i)_i$ sont i.i.d.

1. Loi de Uniform: $X_i \sim \mathcal{U}([0,a]),\ \forall 1 \leq i \leq n$
$$L(x_1,\dots,x_n,a) = \prod_{i=1}^{n} \frac{1}{a}\mathbf{1}_{[0,a]}(x_i) = \left(\frac{1}{a}\right)^n,\ \forall\, 0 \leq x_i \leq a$$
2. Loi de Exponentielle: $X_i \sim \mathcal{E}(\lambda),\ \forall 1 \leq i \leq n$
$$L(x_1,\dots,x_n,\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda}\, \mathbf{1}_{\mathrm{R}_+}(x_i) = \lambda^n\, e^{-n\lambda} \quad \forall x_i \in \mathrm{R}_+$$
3. Loi de Gauss (Normal): $X_i \sim \mathcal{N}(\mu,\sigma^2),\ \forall 1 \leq i \leq n$

# Exemples.

On suppose que les variables aléatoires $(X_i)_i$ sont i.i.d.

1. Loi de Uniform: $X_i \sim \mathcal{U}([0,a]),\ \forall 1 \le i \le n$

$$L(x_1,\ldots,x_n,a) = \prod_{i=1}^{n} \frac{1}{a}\mathbf{1}_{[0,a]}(x_i) = \left(\frac{1}{a}\right)^n,\ \forall\, 0 \le x_i \le a$$

2. Loi de Exponentielle: $X_i \sim \mathcal{E}(\lambda),\ \forall 1 \le i \le n$

$$L(x_1,\ldots,x_n,\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda}\,\mathbf{1}_{\mathrm{R}_+}(x_i) = \lambda^n\, e^{-n\lambda} \quad \forall x_i \in \mathrm{R}_+$$

3. Loi de Gauss (Normal): $X_i \sim \mathcal{N}(\mu,\sigma^2),\ \forall 1 \le i \le n$

$$L(x_1,\ldots,x_n,\mu,\sigma^2) = \prod_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_j-\mu)^2}{2\sigma^2}\right)$$
$$= \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{\sum_{j=1}^{n}(x_j-\mu)^2}{2\sigma^2}\right)$$

# 3. Estimateur du maximum de vraisemblance.
## 3.1. Le cadre et l'objectif.

**Hpothéses:**

1. On suppose désormais que les variables aléatoires $X_1, X_2, \ldots, X_n$ sont indépendantes et de même loi.
2. On suppose que cette loi dépend d'un paramètre $\theta \in \mathbb{R}^d$ inconnu.

**But:** Estimer le paramètre $\theta$ à partir des données observées $x_1, x_2, \ldots, x_n$, par la méthode du maximum de vraisemblance.

**Principe de la méthode:** Il consiste à trouver la valeur de $\theta$ qui maximise la vraisemblance.

L'idée de cette méthode, est que si l'on a observé $x_1, x_2, \ldots, x_n$ plutôt que d'autres valeurs, ce n'est pas "pour rien". Il faut donc chercher la valeur de $\theta$ qui rend le plus probable l'observation des valeurs $x_1, x_2, \ldots, x_n$.

# 3.2. L'estimation du maximum de vraisemblance.

On note la vraisemblance de l'échantillon $\{x_1, \ldots, x_n\}$ par $L(\mathbf{x}, \theta)$, avec $\mathbf{x}(x_1, \ldots, x_n)$.

**Estimation du maximum de Vraissemblance**

On appelle estimation du maximum de vraisemblance la valeur $t_n$ définie par :

$$t_n = \arg\max_{\theta \in \mathbb{R}^d} L(\mathbf{x}, \theta)$$

Puisqu'il est plus facile de maximiser une somme qu'un produit, et que maximiser la vraisemblance ou bien son logarithme est équivalent, en pratique, on maximisera plutôt le logarithme de la vraisemblance, qu'on appellera log-vraisemblance.
Autrement:

$$t_n = \arg\max_{\theta \in \mathbb{R}^d} L(\mathbf{x}, \theta) = \arg\max_{\theta \in \mathbb{R}^d} \log\left(L(\mathbf{x}, \theta)\right).$$

# 3.3. L'estimateur du maximum de vraisemblance.

Pour la plupart des lois de probabilité usuelles, l'estimation du maximum de vraisemblance est défini de façon unique, et se calcule explicitement.
Dans ce cas $t_n = T(x_1, x_2, \ldots, x_n)$ et l'estimateur $T_n$ du maximum de vraisemblance est donné par :

$$T_n = T(X_1, X_2, \ldots, X_n)$$

Quand une détermination explicite de $t_n$ est impossible, il faut avoir recours à une détermination numérique, par **un algorithme d'optimisation de type gradient ou Newton**.

Dans ce cas, nous n'avons qu'une approximation de $t_n$ et l'estimateur $T_n$ du maximum de vraisemblance est défini par :

$$T_n = \arg\max_{z \in \mathbb{R}^d} L(X_1, X_2, \ldots, X_n; z)$$

# 3.4. Propriétés.

Cette méthode fournit un estimateur qui, sur le plan théorique, à de nombreux avantages. Sous des hypothèses vérifiées par de nombreux modèles courants, on démontre qu'il est asymptotiquement sans biais et convergent. On démontre de plus que sa variance est minimale.

Par exemple, supposons le paramètre $\theta$ réel et notons $G$ la fonction de $\mathbb{R}$ dans $\mathbb{R}$ définie pour tout réel $z$ par :

$$G_n(z) = -\log(L(\mathbf{x}, z)).$$

Alors, si la fonction $G_n$ est strictement convexe ($g_n'' > 0$), on a le TLC suivant :

$$\sqrt{g''(T_n)}\,(T_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

# 4. Un exemple de loi discrète : Loi de Bernouilli.
## 4.1. Le cadre et le problème.

**Expérience:** On dispose de $n$ observations $x_1, x_2, \ldots, x_n$ d'une variable qualitative à 2 modalités : 0 (échec) et 1 (succès).

**Hypothéses:**

1. On suppose que ces données sont les réalisations de $n$ variables aléatoires $X_1, X_2, \ldots, X_n$ indépendantes et de même loi de Bernouilli de paramètre $p$, c'est à dire que pour tout $j = 1, 2, \ldots, n$,

$$\mathbb{P}\left[X_j = x\right] = p^x(1-p)^{(1-x)} = \begin{cases} p & \text{si } x = 1 \\ 1-p & \text{si } x = 0 \end{cases}$$

2. On suppose le paramètre $p$ inconnu et on s'intéresse à son estimation par la méthode du maximum de vraisemblance.

# 4.2. Vraisemblance et Log-vraisemblance.

1. On commence donc par expliciter la vraisemblance de l'échantillon $\{x_1, \ldots, x_n\}$.

   Puisque les variables aléatoires $X_1, X_2, \ldots, X_n$ sont indépendantes et de loi discrète, la vraisemblance s'écrit ici:

   $$L(\mathbf{x}, p) = \prod_{j=1}^{n} p^{x_j} (1-p)^{1-x_j} = p^{s_n} (1-p)^{n-s_n}$$

   avec $s_n = \sum_{j=1}^{n} x_j = n\bar{x}$.

2. La fonction log-vraisemblance est alors égale à:

   $$LL(\mathbf{x}, p) = \log(L(\mathbf{x}, p)) = s_n \log(p) + (n - s_n) \log(1-p).$$

3. L'estimation $p_n$ du maximum de vraisemblance est alors donnée par:

   $$p_n = \arg \max_{p \in ]0,1[} L(\mathbf{x}, p) = \arg \max_{p \in ]0,1[} LL(\mathbf{x}, p) = \arg \min_{p \in ]0,1[} -LL(X, p)$$

# 4.3. Résolution du problème de maximisation.

Pour tout $z \in ]0,1[$, On note

$$g(z) = -LL(\mathbf{x}, z) = -s_n \log(z) - (n - s_n) \log(1-z).$$

La valeur de $z$ qui maximise la log-vraisemblance, et qui minimise la fonction $g$, est solution de l'équation d'Euler $g'(z) = 0$, avec:

$$g'(z) = -\frac{s_n}{z} + \frac{n-s_n}{1-z} = \frac{-s_n(1-z) + z(n-s_n)}{z(1-z)} = \frac{nz - s_n}{z(1-z)}.$$

Ainsi, $g'(z) = 0 \Longleftrightarrow z = \frac{s_n}{n} = \overline{x}_n$.

Comme on peut montrer que $g''(z) > 0$ pour tout $z$, la valeur $\overline{x}_n$ est l'unique solution du problème de minimisation.

En effet, on a pour tout $z \in ]0,1[$ :

$$\begin{aligned} g''(z) &= \frac{n\,z(1-z) - (nz - s_n)(1-2z)}{z^2(1-z)^2} = \frac{nz^2 - 2zs_n + s_n}{z^2(1-z)^2} \\ &= \frac{n\left((z-\overline{x}_n)^2 + \overline{x}_n(1-\overline{x}_n)\right)}{z^2(1-z)^2} > 0 \end{aligned}$$

# 4.4. L'estimateur du maximum de vraisemblance.

Donc, l'estimateur du maximum de vraisemblance du paramètre $p$ est:

$$\widehat{p}_n = \overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j$$

**Remarques:**

1. On retrouve l'estimateur classique de l'espérance, puisque $p = \mathbb{E}[X_1]$.
2. On sait que cet estimateur est sans biais et convergent. Par LGN, on a

$$\widehat{p}_n \xrightarrow[n\to\infty]{p.s.} p$$

on retrouve le TLC

$$\sqrt{n}\,(\widehat{p}_n - p) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, p(1-p)).$$

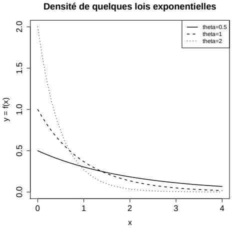
# 5. Loi exponentielle.
## 5.1. Densité.

On considère $n$ données réelles positives $x_1, x_2, \ldots, x_n$. On suppose que ces données sont les réalisations de $n$ variables aléatoires $X_1, X_2, \ldots, X_n$ i.i.d de loi exponentielle de paramètre $\theta > 0$.

La densité de la loi exponentielle est définie pour tout $x \geq 0$ par:

$$f(x) \;=\; \theta \exp(-\theta x)$$

Sa fonction de répartition $F$ est donnée par: $F(x) = 1 - \exp(-\theta x)$. On trouvera dans le graphique ci-dessous quelques courbes de densité avec différents paramétres $(\theta)$.



## 5.2. Simulation d'une réalisation.

Soit $X$ une var. a. de loi exponentielle de paramètre $\theta$ donné.

**Comment simuler une réalisation $x$ de $X$ ?**

Puisque la fonction de répartition de $X$ est explicite, strictement croissante sur $[0, +\infty[$, et qu'elle admet une fonction réciproque explicite, on peut utiliser le théorème d'inversion pour simuler une réalisation $x$ de $X$.

En effet, on a pour tout $u \in ]0, 1[$,

$$F(x) = u \Longleftrightarrow 1 - \exp\left(-\theta\, x\right) = u \Longleftrightarrow x = F^{-1}(u) = \frac{-\log(1-u)}{\theta}$$

Ainsi, pour simuler une réalisation $x$ de $X$,

- on simule une réalisation $u$ de $U \sim \mathcal{U}[0, 1]$,
- et on pose $x = \dfrac{-\log(1-u)}{\theta}$.

Avec le logiciel R, on utilise la fonction `rexp(1,theta)`, ou bien `rexp(n,theta)` pour simuler $n$ valeurs.

# 5.3. Vraisemblance et log-vraisemblance.

On souhaite estimer le paramètre $\theta$ par la méthode du maximum de vraisemblance.

1. On commence donc par expliciter la vraisemblance de l'échantillon.

   Puisque les variables aléatoires $X_1, X_2, \ldots, X_n$ sont indépendantes et de loi continue, la vraisemblance s'écrit ici :

$$L(x_1, x_2, \ldots, x_n, \theta) = \prod_{j=1}^{n} (\theta \exp(-\theta x_j)) = \theta^n \exp\left(-\theta \sum_{j=1}^{n} x_j\right).$$

2. La log-vraisemblance est alors égale à: $\forall \mathbf{x} = (x_1, x_2, \ldots, x_n)$.

$$\mathcal{L}(\mathbf{x}, \theta) = \log\left(L(\mathbf{x}, \theta)\right) = n \log(\theta) - \theta\, s_n.$$

Avec $s_n = \sum_{j=1}^{n} x_j$.

# 5.4. Calcul de l'estimateur du maximum de vraisemblance.

L'estimation du maximum de vraisemblance du paramètre $\theta$ est donné par:

$$t_n = \arg\max_{z \in \mathbb{R}_+} L(\mathbf{x}, z) = \arg\max_{z \in \mathbb{R}_+} \mathcal{L}(\mathbf{x}, z)$$

• On pose $g(z) = \mathcal{L}(\mathbf{x}, z) = n \log(z) - z\, s_n$.
On montre facilement que :

$$g'(z) = \frac{n}{z} - s_n \quad \text{et} \quad g''(z) = -\frac{n}{z^2}$$

Ainsi, puisque $g'(z) = 0 \iff z = n/s_n$ et $g''(z) < 0$, la valeur $t_n$ qui maximise la vraisemblance est égale à $t_n = 1/\overline{x}_n$.

L'estimateur $T_n$ du maximum de vraisemblance est, alors, défini par :

$$T_n = \frac{1}{\overline{X}_n} \quad \text{avec} \quad \overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j.$$

# 5.5. Propriétés.

On démontre facilement que l'estimateur $T_n$ du maximum de vraisemblance est fortement consistant :

$$T_n \xrightarrow[n\to\infty]{p.s.} \theta$$

En effet, grâce à la loi forte des grands nombres :

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow[n\to\infty]{p.s.} \mathbb{E}\left[X_1\right] = \frac{1}{\theta}$$

De la même manière, puisque grâce au théorème de limite centrale, on a :

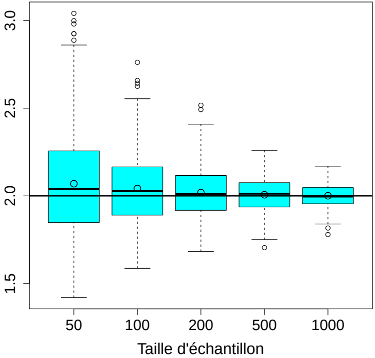
$$\sqrt{n}\left(\overline{X}_n - \theta^{-1}\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \mathbb{V}\mathrm{ar}(X_1) = \frac{1}{\theta^2}\right)$$

on déduit par la delta-méthode que :

$$\sqrt{n}\left(T_n - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \theta^2\right)$$

# 5.6. Etude par simulations du biais et de la variance.

On simule 400 échantillons de 1000 réalisations indépendantes d'une loi exponentielle de paramètre 2. On construit pour chaque échantillon la suite des valeurs successives de l'EMV. On trouvera dans le graphique ci-dessous les boites à moustaches des 400 estimations pour les tailles d'échantillon $n = 50, 100, 200, 500$ et 1000. Qu'observe-t-on et qu'illustre-t-on ?



## 5.7. Commentaires.

L'examen du graphique précédent conduit aux remarques suivantes:

- On constate des fluctuations d'échantillonnage importantes. En effet, pour $n = 50$, l'estimation du paramètre varie entre 1,4 et 3. L'estimation du paramètre peut donc être très mauvaise pour des petites tailles d'échantillon. Cependant, ces fluctuations se réduisent avec l'augmentation de la taille de l'échantillon.
- On constate que la taille des boites à moustaches se réduit avec l'augmentation de la taille de l'échantillon, ce qui montre que la variabilité de l'estimation diminue avec l'augmentation de la taille de l'échantillon. On illustre ainsi le fait que la variance de l'estimateur tend vers 0 lorsque $n$ tend vers l'infini.
- On constate que l'estimation est légèrement biaisée pour les petites tailles d'échantillons, la moyenne des 400 estimations, qui est une estimation de $\mathbb{E}\left[\widehat{\theta}_n\right]$, est en effet plus grande que 2. On illustre ainsi le fait que l'estimateur n'est pas sans biais, mais asymptotiquement sans biais.

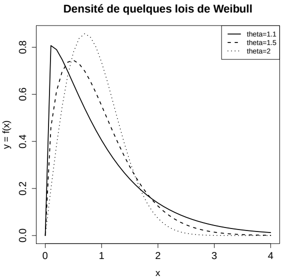
# 6. Loi de Weibull standard (fiabilité).
## 6.1. Densité.

On considère ici $n$ données réelles positives $x_1, \ldots, x_n$. On suppose que ces données sont les réalisations de $n$ variables aléatoires i.i.d de loi de Weibull standard.

La densité de la loi de Weibull standard est définie par:

$$f(x) = \theta\, x^{\theta-1} \exp\left(-x^{\theta}\right), \quad \forall x \geq 0$$

avec $\theta > 1$. Notons que $F(x) = 1 - \exp\left(-x^{\theta}\right)$ pour tout $x \geq 0$.



# 6.2. Vraisemblance et log-vraisemblance.

On souhaite estimer le paramètre $\theta$ en utilisant la méthode du maximum de vraisemblance.

1. La fonction de Vraissemblance:

   Puisque les variables aléatoires $X_1, X_2, \ldots, X_n$ sont indépendantes et de loi continue, la vraisemblance s'écrit ici:

$$L(\mathbf{x}, \theta) = \theta^n \left( \prod_{j=1}^{n} x_j^{\theta-1} \right) \exp\left( -\sum_{j=1}^{n} x_j^{\theta} \right)$$

2. La log-vraisemblance est alors égale à:

$$\mathcal{L}(\mathbf{x}, \theta) = \log(L(\mathbf{x}, \theta)) = n \log(\theta) + (\theta - 1) \sum_{j=1}^{n} \log(x_j) - \sum_{j=1}^{n} x_j^{\theta}$$

# 6.3. Résolution du problème de maximisation.

L'estimation du maximum de vraisemblance du paramètre $\theta$ est donnée par:

$$t_n = \arg\max_{z\in\mathbb{R}^+} L(\mathbf{x}, z) = \arg\max_{z\in\mathbb{R}^+} \mathcal{L}(\mathbf{x}, z) = \arg\min_{z\in\mathbb{R}^+} (-\mathcal{L}(\mathbf{x}, z))$$

Pour tout réel $z$, on pose $g(z) = -\mathcal{L}(\mathbf{x}, z)$.
Ici, l'équation

$$g'(z) = 0 \text{ n'admet pas de solution explicite.}$$

En effet, on a : $g'(z) = -\dfrac{n}{z} - \displaystyle\sum_{j=1}^{n} \log(x_j) + \sum_{j=1}^{n} \log(x_j)\, x_j^z.$

Il faut donc recourir à un algorithme d'approximation (Algorithme de Minimisation) itératif pour obtenir une estimation du paramètre $\theta$.

# 6.4. Algorithme de Newton.

Pour trouver une solution approchée au problème de minimisation de la fonction $g$, on peut utiliser **l'algorithme de Newton** suivant:

- On pose $Z_0 = 1$ et $K = 0$.
- Répéter

$$Z_{K+1} = Z_K - \frac{g'(Z_K)}{g''(Z_K)}, \quad K = K + 1,$$

  jusqu'à convergence ($\|Z_{K+1} - Z_K\| < \varepsilon$ par exemple).
- On pose ensuite $t_n = Z_K$.

avec $g''(z) = \dfrac{n}{z^2} + \sum_{j=1}^{n} (\log(x_j))^2 \, x_j^z > 0$.

## 6.5. Remarques sur les propriétés.

- On notera que $g''(z) > 0$ pour tout $z > 0$, garantissant ainsi que l'estimateur du maximum de vraisemblance est unique.
- On notera que la valeur $t_n$ n'est pas ici une réalisation de l'estimateur $T_n$, mais une approximation de cette réalisation. Elle reste cependant une estimation du paramètre $\theta$.
- L'estimateur du maximum de vraisemblance $T_n$ n'a pas d'expression explicite, il est seulement défini par :

$$T_n = \arg\min_{z>0} \left( -n\log(z) - (z-1)\sum_{j=1}^{n} \log(X_j) + \sum_{j=1}^{n} X_j^z \right).$$

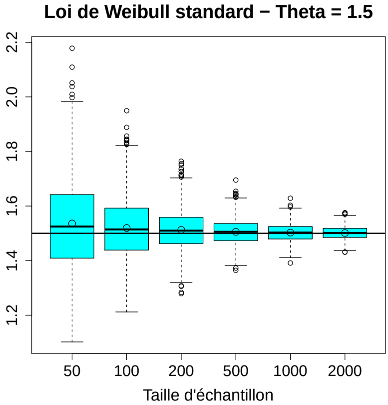
- On démontre que $T_n$ est un estimateur asymptotiquement sans biais et convergent. On démontre aussi que

$$\sqrt{G''(T_n)}(T_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

avec $G''(z) = \frac{n}{z^2} + \sum_{j=1}^{n} (\log(X_j))^2 X_j^z$.

# 6.6. Etude par simulations.

On simule 400 échantillons de taille 2000 d'une loi de Weibull de paramètre $\theta = 1.5$. On s'intéresse au comportement de l'estimateur du maximum de vraisemblance pour différentes tailles d'échantillon. On notera que le calcul de l'estimation est obtenu en 5 itérations, pour une précision à $10^{-8}$.



# 6.6. Commentaires.

On constate que la taille des boîtes à moustache se réduit avec l'augmentation de la taille de l'échantillon, ce qui illustre bien que la variance de l'estimateur diminue avec la taille de l'échantillon. On peut identifier un biais léger de l'estimateur pour des petites tailles d'échantillon, diminuant avec l'augmentation de la taille de l'échantillon.

# 7. Un exemple de loi continue : Loi Gaussienne.
## 7.1. Introduction.

On considère ici $n$ données réelles $x_1, \ldots, x_n$. On suppose que ces données sont les réalisations de $n$ variables aléatoires indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$.

On souhaite estimer les paramètres $\mu$ et $\sigma^2$ par la méthode du maximum de vraisemblance.

**La vraisemblance** s'écrit ici:

$$L(\mathbf{x}, \mu, \sigma^2) = \prod_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_j-\mu)^2}{2\sigma^2}\right) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{\displaystyle\sum_{j=1}^{n}(x_j-\mu)^2}{2\sigma^2}\right)$$

**La log-vraisemblance** s'écrit:

$$g(\mu, \sigma^2) = -n\log(\sqrt{2\pi}) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2$$

## 7.2. Calcul de l'estimateur.

La solution est obtenue en résolvant le système d'équations :

$$\begin{cases} \dfrac{\partial g}{\partial \mu}(\mu, \sigma^2) = 0 \\ \dfrac{\partial g}{\partial \sigma^2}(\mu, \sigma^2) = 0 \end{cases} \iff \begin{cases} -\dfrac{1}{\sigma^2} \sum_{j=1}^{n} (\mu - x_j) = 0 \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4} \sum_{j=1}^{n} (x_j - \mu)^2 = 0 \end{cases}$$

On obtient alors les solutions :

$$\begin{cases} \mu &= \dfrac{1}{n} \sum_{j=1}^{n} x_j \\ \sigma^2 &= \dfrac{1}{n} \sum_{j=1}^{n} (x_j - \mu)^2 \end{cases} \iff \begin{cases} \mu &= \overline{x}_n \\ \sigma^2 &= \dfrac{1}{n} \sum_{j=1}^{n} (x_j - \overline{x}_n)^2 \end{cases}$$

ce qui conduit aux estimateurs du maximum de vraisemblance des paramètres $\mu$ et $\sigma^2$ :

$$\widehat{\mu} = \overline{X}_n \text{ et } \widehat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} \left(X_j - \overline{X}_n\right)^2$$

On retrouve les estimateurs classiques de l'espérance et de la variance.