MATH60604
Modélisation statistique
§ 3 Vraisemblance

HEC Montréal
Département de sciences de la décision

Vraisemblance

- La vraisemblance $L(\boldsymbol{\theta})$ est une fonction des paramètres de la loi des observations, $\boldsymbol{\theta}$.
  - La vraisemblance est la probabilité d'observer l'échantillon sous le modèle avec $\boldsymbol{\theta}$.
  - Les observations sont fixées à leurs valeurs réalisées.
- L'estimateur du maximum de vraisemblance $\widehat{\boldsymbol{\theta}}$ est la valeur de $\boldsymbol{\theta}$ qui maximise la vraisemblance.
  - soit la valeur pour laquelle l'échantillon observé est le plus probable ou susceptible.
  - raisonnement scientifique: on s'attend à voir ce qu'on observe.

Échantillon de loi Bernoulli

- Supposons que nous voulons estimer la probabilité qu'un événement survienne, sans aucune variable explicative.
- Par exemple, est-ce que le client achète un certain produit ou non, est-ce qu'un participant à une étude réussit la tâche demandée, etc.
- On a à disposition un échantillon aléatoire de taille n, où X_i suit une loi Bernoulli de paramètre $p \in (0,1)$,

$$\mathrm{P}\left(X_i = 1\right) = p, \qquad \mathrm{P}\left(X_i = 0\right) = 1 - p.$$

- Par convention, on désigne le résultat « 1 » par un succès et « 0 » par un échec.

Probabilité conjointe d'événements binaires

Une manière compacte d'écrire la fonction de masse de l'observation X_i est

$$P(X_i = x_i \mid p) = p^{x_i}(1-p)^{(1-x_i)}, \qquad x_i = 0, 1.$$

Comme les observations sont indépendantes, la probabilité conjointe d'avoir un résultat donné est le produit des probabilités pour chaque observation,

$$\begin{aligned} P(X_1 = x_1, \ldots, X_n = x_n \mid p) &= \prod_{i=1}^{n} \mathrm{P}\left(X_i = x_i \mid p\right) \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{(1-x_i)}. \end{aligned}$$

Vraisemblance du modèle Bernoulli

La fonction de vraisemblance pour un échantillon aléatoire est

$$\begin{aligned} L(p;X) \equiv L(p;X) &= \prod_{i=1}^{n} p^{X_i}(1-p)^{(1-X_i)} \\ &= p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i}. \end{aligned}$$

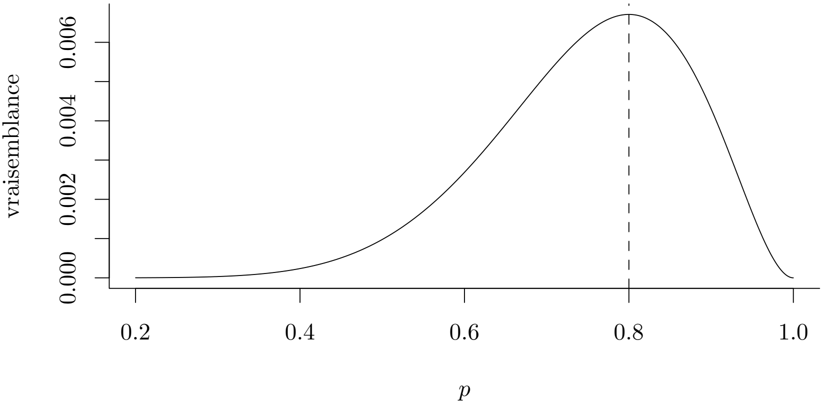
La vraisemblance est, à une constante de normalisation près, la fonction de masse d'un échantillon de loi binomiale avec n essais et probabilité de succès p.

- La vraisemblance ne dépend que du nombre de succès observés, peu importe l'ordre des observations.
- L'estimateur du maximum de vraisemblance est la proportion de succès dans l'échantillon.

Supposons qu'on a $n = 10$ observations avec huit succès.

- La fonction de vraisemblance est $L(p) = p^8(1-p)^2$.

Graphe de la vraisemblance



Log-vraisemblance d'un échantillon Bernoulli

- Dans notre exemple, la fonction de log-vraisemblance est

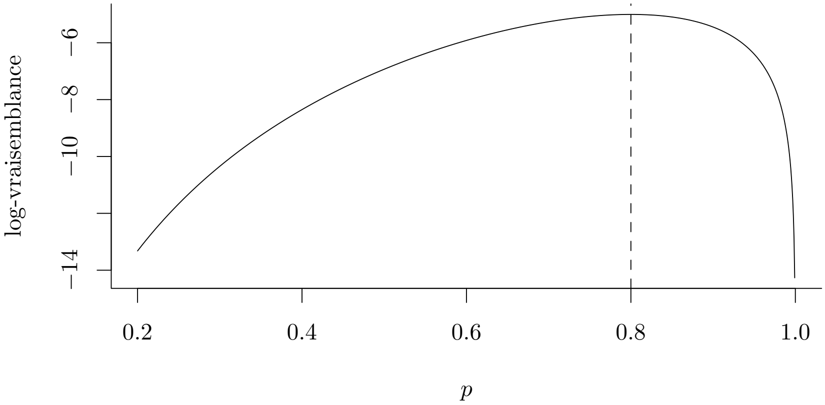
$$\ell(p) = \sum_{i=1}^{n} \log \left\{ p^{x_i} (1-p)^{1-xi} \right\}$$

- En utilisant la propriété $\log(a^b) = b\log(a)$, on obtient

$$\ell(p) = \log(p) \sum_{i=1}^{n} x_i + \log(1-p) \left(n - \sum_{i=1}^{n} x_i \right).$$

- Dans notre exemple numérique, avec huit succès et deux échecs, la log-vraisemblance est $\ell(p) = 8\log(p) + 2\log(1-p)$.

Graphe de la log-vraisemblance $\ell(p)$



Estimateur du maximum de vraisemblance pour la probabilité de succès d'une variable Bernoulli

En dérivant la fonction $\ell(p)$ en fonction de p, on obtient

$$\frac{\mathrm{d}}{\mathrm{d}p}\ell(p) = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{(1-p)}\left(n - \sum_{i=1}^{n} x_i\right).$$

Si on résoud l'équation $\mathrm{d}\ell(p)/\mathrm{d}p = 0$, on trouve

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

La dérivée seconde,

$$\frac{\mathrm{d}^2\ell(p)}{\mathrm{d}p^2} = -\frac{1}{p^2}\sum_{i=1}^{n} x_i - \frac{1}{(1-p)^2}\left(n - \sum_{i=1}^{n} x_i\right),$$

est négative, donc $L(p)$ atteint un maximum à $\widehat{p}$. L'estimateur du maximum de vraisemblance de p est la **proportion de succès**.

Information

La courbure de la vraisemblance renseigne sur l'incertitude des estimateurs.
L'information observée est $j(p) = -\mathrm{d}^2\ell(p)/\mathrm{d}p^2$ et pour notre exemple

$$j(\widehat{p}) = \frac{n}{\overline{x}} + \frac{n}{(1-\overline{x})} = \frac{n}{\overline{x}(1-\overline{x})}.$$

La variance de $\widehat{p}$ est $j^{-1}(\widehat{p}) = 0.016$ est l'erreur-type 0.1265.
L'information de Fisher est

$$i(\theta) = \frac{n}{p(1-p)}.$$

- Pour données indépendantes et identiquement distribuées, l'information cumulative d'un échantillon de taille n est la somme des n contributions individuelles identiques (l'information s'accumule de manière linéaire).

Procédures de tests

Suppose qu'on veut tester l'hypothèse bilatérale

$$\mathscr{H}_0 : p_0 = 0.5 \qquad \text{versus} \qquad \mathscr{H}_a : p_0 \neq 0.5.$$

Trois statistiques basés sur la vraisemblance peuvent être employés pour tester cette hypothèse:

- la statistique de Wald

$$W(p_0) = \frac{(\widehat{p} - p_0)^2}{\mathrm{Var}\,(\widehat{p})} = \frac{(\widehat{p} - p_0)^2}{\widehat{p}(1-\widehat{p})/n}$$

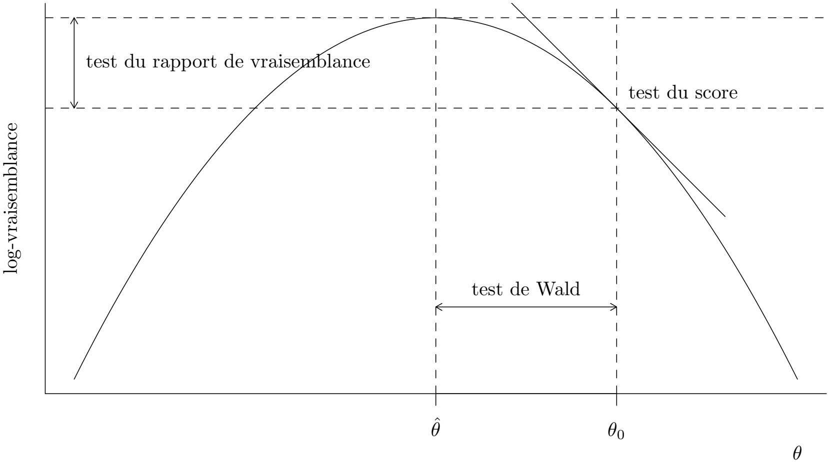
- la statistique du score (ou multiplicateurs de Lagrange)

$$S(p_0) = \frac{U^2(p_0)}{i(p_0)} = \frac{(\widehat{p} - p_0)^2}{p_0(1-p_0)/n}$$

- la statistique du rapport de vraisemblance

$$\begin{aligned} R(p_0) &= 2\{\ell(\widehat{p}) - \ell(p_0)\} \\ &= 2\left\{y \ln\left(\frac{\widehat{p}}{p_0}\right) + (n-y)\ln\left(\frac{1-\widehat{p}}{1-p_0}\right)\right\} \end{aligned}$$

Illustrations des tests basés sur la vraisemblance



Résultats numériques et intervalles de confiance

- Avec 8 succès sur 10 essais, les statistiques valent $W = 5.62$, $S = 3.6$, $R = 3.855$;
- on compare ces valeurs au 0.95 quantile de la loi χ^2_1, 3.84.

- Si la taille de l'échantillon est petite ou que la distribution d'échantillonage de l'estimateur est asymmétrique, le test basé sur la statistique de Wald n'est pas fiable.
- On peut inverser la statistique de Wald pour obtenir un intervalle de confiance à 95%

$$\widehat{p} \pm \mathfrak{z}_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

- Ce dernier vaut $0.8 \pm 1.96 \cdot 0.1265 = [0.55, 1.048]$!
- Comparez avec les intervalles de confiance à 95%
  - du rapport de vraisemblance, $[0.5005, 0.964]$.
  - du score, $[0.49, 0.943]$.

Résoudre les équations $\{p : S(p) \leq 3.84\}$ et $\{p : R(p) \leq 3.84\}$ numériquement.