Compte-rendu en *Principes et méthodes statistiques*

Analyse de signaux oculométriques

Aurélien PEPIN, Léo DESBUREAUX, Julien LABOURÉ (Ensimag)

5 mai 2017

1 Analyse d'échantillons de loi binomiale négative

QUESTION 1. On suppose dans un premier temps r connu. L'estimateur obtenu par la méthode des moments est noté $\tilde{p}_n$.

> Comme on suppose l'échantillon $X_1, X2, ..., X_n$ indépendantes et de même loi binomiale négative $\mathcal{BN}(r,p)$, on a $E[X] = \frac{r}{p}$. Alors l'estimateur des moments est :
>
> $$\tilde{p}_n = \frac{r}{\overline{X}_n}$$
>
> où $\overline{X}_n$ désigne la moyenne empirique de l'échantillon.

Une deuxième méthode d'estimation ponctuelle est l'estimation par maximum de vraisemblance. On note maintenant l'estimateur trouvé $\hat{p}_n$.

> $$\begin{aligned} \mathcal{L}(p; x_1, ..., x_n) &= P(X_1 = x_1, ..., X_n = x_n; p) \\ &= \prod_{i=1}^{n} P(X = x_i; p) \end{aligned}$$
>
> Plutôt que de dériver directement ce produit, on préfère maximiser le logarithme de la fonction de vraisemblance $\mathcal{L}$, c'est la **log-vraisemblance** :
>
> $$\begin{aligned} \ln \mathcal{L}(p; x_1, ..., x_n) &= \ln \prod_{i=1}^{n} P(X = x_i; p) = \sum_{i=1}^{n} \ln P(X = x_i; p) \\ &= \sum_{i=1}^{n} \ln \binom{x_i - 1}{r - 1} (1-p)^{x_i - r} p^r \end{aligned}$$
>
> On cherche désormais à obtenir $\hat{p}_n$, valeur qui maximise cette log-vraisemblance. On dérive pour cela l'expression précédente :

$$
\begin{aligned}
\frac{\partial}{\partial p}\ln \mathcal{L}(p; x_1, ..., x_n) &= \frac{\partial}{\partial p}(\sum_{i=1}^{n} \ln \binom{x_i - 1}{r - 1}) + \sum_{i=1}^{n}(x_i - r)\frac{\partial}{\partial p}(\ln(1-p)) + \sum_{i=1}^{n} r\frac{\partial}{\partial p}(\ln(p)) \\
&= 0 - \sum_{i=1}^{n}(\frac{x_i - r}{1-p}) + \sum_{i=1}^{n}\frac{r}{p}
\end{aligned}
$$

Cette expression s'annule sous les conditions suivantes :

$$
\begin{aligned}
-\sum_{i=1}^{n}(\frac{x_i - r}{1-p}) + \sum_{i=1}^{n}\frac{r}{p} = 0 &\iff \sum_{i=1}^{n}\frac{-px_i + pr + (1-p)r}{p(1-p)} = 0 \\
&\iff \sum_{i=1}^{n} -px_i + r = 0 \\
&\iff \sum_{i=1}^{n}\frac{r}{p} = \sum_{i=1}^{n} x_i \\
&\iff n\frac{r}{p} = \frac{1}{n}\sum_{i=1}^{n} x_i \\
&\iff \frac{r}{p} = \overline{X}_n
\end{aligned}
$$

donc

$$
\hat{p}_n = \frac{r}{\overline{X}_n} = \tilde{p}_n
$$

Le cas où $p = 0$ ou $p = 1$ correspond à des situations triviales où tous les X_i sont identiques, il n'y a aucune part d'aléatoire.