Notes de cours Statististique inférencielle 2 : Test d'hypothèses

Pr. Mokhtar KOUKI

Université de Carthage (ESSAI)
mokhtar.kouki@essai.ucar.tn

Mars - Mai 2023

Contents

Contenu

Généralités

On considère un échantillon $(X_1, X_2, \cdots, X_n)$ un échantillon issu d'une loi P_θ et de densité de probabilité $f(x, \theta)$. Dans le processus de test d'hypothèse $H_0 : \theta \in \Theta_0$ contre une hypothèse alternative notée $H_1 : \theta \in \Theta_1$ (ou Θ_0 et Θ_1 des sous ensemble de $\mathbb{R}$, peut être résumé comme suit :

		Décision	
		H_0	H_1
Réalité	H_0	Accepter H_0 sachant que H_0 vraie	Rejeter H_0 sachant que H_0 vraie
	H_1	Accepter H_0 sachant que H_1 vraie	Rejeter H_0 sachant que H_1 vraie

Les cellules en surbrillance jaune constituent les cas de décision erronée.

Généralités

Région critique

La région de rejet de l'hypothèse H_0 est appelée région critique. On la notera W. $\overline{W}$ est appelée région d'acceptation de H_0

Généralités

Région critique

La région de rejet de l'hypothèse H_0 est appelée région critique. On la notera W. $\overline{W}$ est appelée région d'acceptation de H_0

Erreur de 1ère espèce, errruer de 2ème espèce et puissance d'un test

L'erreur (risque) de première espèce corresspond à la probabiité de ne pas accepter H_0 sachant que H_0 est vraie. On le note α

$$\alpha = P(\text{ rejeter } H_0/H_0 \text{ est vraie}) = P(W/\theta \in \Theta_0)$$

Généralités

Région critique

La région de rejet de l'hypothèse H_0 est appelée région critique. On la notera W. $\overline{W}$ est appelée région d'acceptation de H_0

Erreur de 1ère espèce, errruer de 2ème espèce et puissance d'un test

L'erreur (risque) de première espèce corresspond à la probabiité de ne pas accepter H_0 sachant que H_0 est vraie. On le note α

$$\alpha = P(\text{ rejeter } H_0/H_0 \text{ est vraie}) = P(W/\theta \in \Theta_0)$$

L'erreur (risque) de 2ème espèce corresspond à la probabiité d'acceper H_0 sachant que H_1 est vraie. On la note β

$$\beta = P(\text{ accepter } H_0/H_1 \text{ est vraie}) = P(\overline{W}/\theta \in \Theta_1)$$

Généralités

Région critique

La région de rejet de l'hypothèse H_0 est appelée région critique. On la notera W. $\overline{W}$ est appelée région d'acceptation de H_0

Erreur de 1ère espèce, errruer de 2ème espèce et puissance d'un test

L'erreur (risque) de première espèce corresspond à la probabiité de ne pas accepter H_0 sachant que H_0 est vraie. On le note α

$$\alpha = P(\text{ rejeter } H_0/H_0 \text{ est vraie}) = P(W/\theta \in \Theta_0)$$

L'erreur (risque) de 2ème espèce corresspond à la probabiité d'acceper H_0 sachant que H_1 est vraie. On la note β

$$\beta = P(\text{ accepter } H_0/H_1 \text{ est vraie}) = P(\overline{W}/\theta \in \Theta_1)$$

La puissance d'un test corresspond à la probabiité de ne pas accepter H_0 sachant que H_1 est vraie. On la note $\eta = 1 - \beta$

$$1 - \beta = P(\text{ rejeter } H_0/H_1 \text{ est vraie}) = P(W/\theta \in \Theta_1)$$

Test d'hypotèse simple : $H_0 : \theta = \theta_0$ vs $H_0 : \theta = \theta_1$

Région critique du test $H_0 : \theta = \theta_0$ vs $H_0 : \theta = \theta_1$: principe de Neyman-Pearson

Pour un risque de 1ère espèce α donné, la région critique de rejet de l'hypothèse H_0 est définie par la relation suivante :

$$W = \left\{ (X_1, X_2, \cdots, X_n) \text{ tq } \frac{L(X_1, X_2, \cdots, X_n, \theta_0)}{L(X_1, X_2, \cdots, X_n, \theta_1)} \leq k \right\}$$

La constante k es déterminée par la relation $p(W/\theta = \theta_0) = \alpha$.
$L(X_1, X_2, \cdots, X_n, \theta)$ étant la vraisemblance de léchantillon.

Test d'hypotèse simple : $H_0 : \theta = \theta_0$ vs $H_0 : \theta = \theta_1$

Région critique du test $H_0 : \theta = \theta_0$ vs $H_0 : \theta = \theta_1$: principe de Neyman-Pearson

Pour un risque de 1ère espèce α donné, la région critique de rejet de l'hypothèse H_0 est définie par la relation suivante :

$$W = \left\{ (X_1, X_2, \cdots, X_n) \text{ tq } \frac{L(X_1, X_2, \cdots, X_n, \theta_0)}{L(X_1, X_2, \cdots, X_n, \theta_1)} \leq k \right\}$$

La constante k es déterminée par la relation $p(W/\theta = \theta_0) = \alpha$.
$L(X_1, X_2, \cdots, X_n, \theta)$ étant la vraisemblance de léchantillon.

Exemple 1 : test sur l'espérance mathématique d'une loi normale On considère $(X_1, X_2, \cdots, X_n)$ un échantillon iid suivant une loi normale $N(m, \sigma^2)$ et on veut réaliser le test suivant :

$$\begin{cases} H_0 : & m = m_0 \\ H_1 : & m = m_1 \end{cases}$$

Test d'hypotèse simple : $H_0 : \theta = \theta_0$ vs $H_0 : \theta = \theta_1$

$$
\begin{aligned}
\frac{L(X_1, X_2, \cdots, X_n, \theta_0)}{L(X_1, X_2, \cdots, X_n, \theta_1)} &= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(X_i - m_0)^2 - \sum_{i=1}^{n}(X_i - m_1)^2\right)\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\left(-2(m_0 - m_1)\sum_{i=1}^{n} X_i + n(m_0^2 - m_1^2)\right)\right) \\
&= \exp\left(-\frac{-n}{2\sigma^2}(m_0^2 - m_1^2)\right)\exp\left(\frac{m_0 - m_1}{\sigma^2}\sum_{i=1}^{n} X_i\right)
\end{aligned}
$$

Test d'hypotèse simple : $H_0 : \theta = \theta_0$ vs $H_0 : \theta = \theta_1$

Et Pour $m_0 > m_1$, on a :

$$\begin{aligned}
P(W/m = m_0) &= P\left(\frac{m_0 - m_1}{\sigma^2}\sum_{i=1}^{n} X_i \leq k_2/m = m_0\right) \\
&= P\left(\sum_{i=1}^{n} X_i \leq c/m = m_0\right) \\
&= P(\sqrt{n}\frac{\overline{X} - m_0}{\sigma} \leq \sqrt{n}\frac{c - m_0}{\sigma}) = \alpha
\end{aligned}$$

Conclusion : pour une variance σ^2 connue, la région critique est définie par :

$$W = \left\{(X_1, X_2, \cdots, X_n) \text{ tq } \sum_{i=1}^{n} X_i \leq nm_0 + \sqrt{n}\sigma u_\alpha\right\}$$

u_α est tel que $\Phi(u_\alpha) = \alpha$ Par analogie, si $m_0 < m_1$, la région critique est définie par :

$$W = \left\{(X_1, X_2, \cdots, X_n) \text{ tq } \sum_{i=1}^{n} X_i \geq nm_0 + \sqrt{n}\sigma u_{1-\alpha}\right\}$$

test d'hypothèse $H_0 : \theta \leq \theta_0$ vs $\theta > \theta_0$

Définition : Famille à rapport de vraisemblance monotone (RVM)

On considère X une variable aléatoire réelle de loi de probabilité P_θ et de densité de probabilité $f(x, \theta)$, avec $\theta \in \Theta \subset \mathbb{R}$. La loi P_θ est dite à rapport e vraisemblance monothène (**RVM**) s'il existe une statistique S telle pour un échantillon $(X_1, X_2, \cdots, X_n)$ le rapport de vraisemblance

$$\frac{L(X_1, X_2, \cdots, X_n, \theta_2)}{L(X_1, X_2, \cdots, X_n, \theta_1)}, \theta_2 > \theta_1$$

est une fonction croissante de S.

Exemple 1 : Loi de poisson

Pour un échantillon $(X_1, X_2, \cdots, X_n)$ issue d'une loi de poisson de paramètre θ. Le rappot de vraisemblance s'écrit :

$$\frac{L(X_1, X_2, \cdots, X_n, \theta_2)}{L(X_1, X_2, \cdots, X_n, \theta_1)} = \exp(\theta_2 - \theta_1) \left(\frac{\theta_2}{\theta_1}\right)^{\sum_{i=1}^n X_i}$$

Pour $S = \sum_{i=1}^n X_i$ la loi de poisson est à **RVM**

Région critique du test $H_0 : \theta \leq \theta_0$ vs $\theta > \theta_0$

Théorème de Lehman

On considère la variable aléatoire réelle X à rapport de vraisemblance monotone (**RVM**) et $(X_1, X_2, \cdots, X_n)$ un échantillon iid de même loi que X. La région critique de rejet de l'hypothèse $H_0 : \theta \leq \theta_0$ de risque de 1ère espèce α est définie par :

$$W = \{(X_1, X_2, \cdots, X_n) \text{ tq } S(X_1, X_2, \cdots, X_n) > k\}$$

et $P(W/\theta = \theta_0) = \alpha$

Exemple 1 : Espérance d'une loi normale Pour un échantillon d'une loi normale $(X_1, X_2, \cdots, X_n)$ d'espérnce mathématique m et de variance σ^2, le rapport de vraisemblance pour $m_2 > m_1$:

$$\frac{L(X_1, X_2, \cdots, X_n, m_2)}{L(X_1, X_2, \cdots, X_n, m_1)} = \exp\left[\frac{1}{\sigma^2}(m_2 - m_1)\sum_{i=1}^{n} X_i\right] \exp\left[-\frac{n}{2\sigma^2}(m_2^2 - m_1^2)\right]$$

Ainsi, pour $S = \sum_{i=1}^{n} X_i$, la loi normale est à **RVM**. Et la région critique du test est définie par :

$$W = \left\{(X_1, X_2, \cdots, X_n) \text{ tq } \sum_{i=1}^{n} X_i > k\right\}$$

et

$$\begin{aligned} P(W/m = m_0) &= P(\sum_{i=1}^{n} X_i > k/m = m_0) \\ &= P\left(\frac{\sqrt{n}}{\sigma}(\overline{X} - m_0) > \frac{\sqrt{n}}{\sigma}(\frac{k}{n} - m_0)\right) \\ &= P\left(N(0,1) > \frac{\sqrt{n}}{\sigma}(\frac{k}{n} - m_0)\right) = \alpha \end{aligned}$$

test d'hypothèse $H_0 : \theta_1 \leq \theta \leq \theta_2$ vs $H_1 : \theta < \theta_1$ ou $\theta > \theta_2$

Définition : Famille exponentielle

On considère X une variable aléatoire réelle de loi de probabilité P_θ et de densité $f(x,\theta)$. La loi P_θ est dite apparetenant à la famille exponentielle si pour un échantillon iid $(X_1, X_2, \cdots, X_n)$ il existe une statistique S et $h()$ et $g()$ deux fonctions, telles que la vraisemblance peut s'écrire :

$$L(X_1, X_2, \cdots, X_n, \theta) = h(\theta) \exp\left(g(\theta) S(X_1, X_2, \cdots, X_n)\right)$$

Exemple : La loi normale La vraisemblance d'un échatillon iid issu d'une loi normale despérance mathémaique m et de variance σ^2 peut être écrite sous la forme :

$$f(X_1, X_2, \cdots, X_n, m, \sigma^2) = (2\pi)^{-2/n} \sigma^{-n} \exp\left(\frac{-n}{2\sigma^2} m^2\right) \exp\left[\frac{1}{\sigma^2} m \sum_{i=1}^{n} X_i\right] \exp\left\{\frac{-1}{\sigma^2} \sum_{i=1}^{n} X_i^2\right\}$$

Région critique du test $H_0 : \theta_1 \leq \theta \leq \theta_2$ vs $H_1 : \theta < \theta_1$ ou $\theta > \theta_2$

Ainsi, pour

$$S_1 = \sum_{i=1}^{n} X_i \text{ et } S_2 = \sum_{i=1}^{n} X_i^2$$

respectivement des statistiques pour m et σ^2, la loi normale appartient à la famille des lois exponenetielles

Région critique du test $H_0 : \theta_1 \leq \theta \leq \theta_2$ vs $H_1 : \theta < \theta_1$ ou $\theta > \theta_2$

On considère la variable aléatoire réelle X de loi de probabillité appartenant à la famille exponentielle et de densité de probabilité $f(x, \theta)$ et $(X_1, X_2, \cdots, X_n)$ un échantillon iid de même loi que X. La région critique de rejet de l'hypothèse $H_0 : \theta_1 \leq \theta \leq \theta_2$ de risque de 1ère espèce θ est définie par :

$$\begin{aligned} W &= \{(X_1, X_2, \cdots, X_n) \text{ tq } S(X_1, X_2, \cdots, X_n) < k_1 \\ &\quad \text{ou } S(X_1, X_2, \cdots, X_n) > k_2\} \end{aligned}$$

et les deux constantes k_1 et k_2 sont déterminées par $P(W/\theta = \theta_1) = \alpha$ et $P(W/\theta = \theta_2) = \alpha$.

Exemple : Espérance mathématique d'une loi normale $\rightarrow (m, \sigma^2)$

La région critique du test de l'hypothèse $H_0 : m_1 \leq m \leq m_2$ vs $H_1 : m < m_1$ ou $m > m_2$ est éfinie par :

$$W = \left\{ (X_1, X_2, \cdots, X_n) \text{ tq } \sum_{i=1}^{n} X_i < k_1 \text{ ou } \sum_{i=1}^{n} X_i > k_2 \right\}$$

$$\begin{aligned} P(W/H_0) &= P\left(\sum_{i=1}^{n} X_i < k_1 \text{ ou } \sum_{i=1}^{n} X_i > k_2 / H_0 \right) \\ &= P\left(\sum_{i=1}^{n} X_i < k_1 / m_1 < m < m_2 \right) \\ &\quad + P\left(\sum_{i=1}^{n} X_i > k_2 / m_1 < m < m_2 \right) \end{aligned}$$

$$\Phi\left(\frac{\sigma}{\sqrt{n}} \left(\frac{k_1}{n} - m_1 \right) \right) + 1 - \Phi\left(\frac{\sigma}{\sqrt{n}} \left(\frac{k_2}{n} - m_2 \right) \right) = \alpha$$

La résolution du système précédent (une équation et deux inconnues) nécessite le partage de α comme suit :

$$\Phi\left(\frac{\sigma}{\sqrt{n}}\left(\frac{k_1}{n} - m_1\right)\right) = \alpha_1 \text{ et } 1 - \Phi\left(\frac{\sigma}{\sqrt{n}}\left(\frac{k_2}{n} - m_2\right)\right) = \alpha_2$$

et $\alpha_1 + \alpha_2 = \alpha$. Ce qui permet de donner :

$$k_1 = nm_1 + \sqrt{n}\sigma u_{\alpha_1} = nm_1 - \sqrt{n}\sigma u_{1-\alpha_1} \text{ et } k_2 = nm_2 + \sqrt{n}\sigma u_{1-\alpha_2}$$

Dans le cas où $\alpha_1 = \alpha_2 = \alpha/2$, la région critique du test est défiie par :

$$W = \left\{(X_1, X_2, \cdots, X_n) \quad \text{tq} \quad \sum_{i=1}^{n} X_i < nm_1 - \sqrt{n}\sigma u_{1-\alpha/2} \right.$$
$$\left. ou \quad \sum_{i=1}^{n} X_i > nm_2 + \sqrt{n}\sigma u_{1-\alpha/2} \right\}$$

ou bien

$$W = \left\{(X_1, X_2, \cdots, X_n) \text{ tq } \overline{X} < m_1 - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} \text{ ou } \overline{X} > m_2 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right\}$$

Remaqrue : cas où la variance, σ^2, est inconnue

Dans le cas où la vriance est inconnue, on remplace σ^2 par son estimateur :

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

et les quantiles sont extraites à partir d'une loi de student avec de degré liberté $n-1$.

Cas particulier : $m_1 = m_2 = m_0$

Dans le cas où $m_1 = m_2 = m_0$, le test dvient $H_0 : m = m_0$ vs $H_1 : m \neq m_0$. Et la région critique est définie par :

$$W = \left\{(X_1, X_2, \cdots, X_n) \text{ tq } |\overline{X} - m_0| > \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right\}$$

Test d'égalité des variances de deux lois normales

On considère $(X_1, X_2, \cdots, X_{n_1})$ un échatillon issu d'une loi normale $N(m_1, \sigma_1^2)$ et $(Y_1, Y_2, \cdots, Y_{n_2})$ un échatillon issu d'une loi normale $N(m_2, \sigma_2^2)$. On se propose de réaliser le test suivant, pour un risque de 1ère espéèce α :

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Les deux estimateurs des variances, dans le cas où m_1 et m_2 sont inconnus :

$$S_x^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(X_i - \overline{X}\right)^2 \text{ et } S_y^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} \left(Y_j - \overline{Y}\right)^2$$

Ce qui implique :

$$\frac{(n_1 - 1)S_x^2}{\sigma_1^2} \to \chi^2_{(n_1-1)} \text{ et } \frac{(n_2 - 1)S_y^2}{\sigma_2^2} \to \chi^2_{(n_2-1)}$$

Et le rapport des deux quantités donne :

$$\frac{S_x^2}{S_y^2} \frac{\sigma_2^2}{\sigma_1^2} \to F(n_1 - 1, n_2 - 1)$$

Test d'égalité des variances de deux lois normales

Sous H_0, $\frac{\sigma_2^2}{\sigma_1^2} = 1$ et la région critique est définie par :

$$W = \left\{ (X_1, X_2, \cdots, X_{n_1}), (Y_1, Y_2, \cdots, Y_{n_2}), \text{ tq } \frac{S_x^2}{S_y^2} < k_1 \text{ ou } \frac{S_x^2}{S_y^2} >_k 2 \right\}$$

et

$$P(W/H_0) = \alpha$$

Pour une partition $\alpha_1 + \alpha_2 = \alpha$, $k_1 = F_{\alpha_1}(n_1 - 1, n_2 - 1)$ et $k_2 = F_{1-\alpha_2}(n_1 - 1, n_2 - 1)$

Test d'égalité de l'espérance de deux lois normales

On considère $(X_1, X_2, \cdots, X_{n_1})$ un échatillon issu d'une loi normale $N(m_1, \sigma_1^2)$ et $(Y_1, Y_2, \cdots, Y_{n_2})$ un échatillon issu d'une loi normale $N(m_2, \sigma_2^2)$. On se propose de réaliser le test suivant, pour un risque de 1ère espéèce α :

$$\begin{cases} H_0 : & m_1 = m_2 \\ H_1 : & m_1 \neq m_2 \end{cases}$$

Sous l'hypothèse de la normalité des X_i et y_j, les deux estimateurs de m_1 et m_2 suivent des lois normales :

$$\hat{m}_1 = \overline{X} \curvearrowright N\left(m_1, \frac{\sigma_1^2}{n_1}\right) \text{ et } \hat{m}_2 = \overline{Y} \curvearrowright N\left(m_2, \frac{\sigma_2^2}{n_2}\right)$$

Et sous H_0

$$U = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \curvearrowright N(0, 1)$$

Test d'égalité de l'espérance de deux lois normales

- Cas où deux variances sont connues : La région critique du test d'égalité des espérances mathématiques est définie par :

$$W = \left\{ (X_1, X_2, \cdots, X_{n_1}), (Y_1, Y_2, \cdots, Y_{n_2}), \quad \text{tq} \quad |\overline{X} - \overline{Y}| > \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} u_{1-\alpha/2} \right\}$$

où $\Phi(u_{1-\alpha/2}) = \alpha/2$, $\Phi()$ étant la fonction de répartition de la loi normale centrée-réduite.

Test d'égalité de l'espérance de deux lois normales

- Cas où les deux variances sont inconnues et égales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) : La région critique du test d'égalité des espérances mathématiques est définie par :

$$W = \left\{ (X_1, X_2, \cdots, X_{n_1}), (Y_1, Y_2, \cdots, Y_{n_2}), \quad \text{tq} \quad |\overline{X} - \overline{Y}| > \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{2}{n_2}} z_{1-\alpha/2} \right\}$$

avec

$$\hat{\sigma}^2 = \frac{1}{n_1 + n_2 - 2}\left\{ \sum_{i=1}^{n_1} \left(X_i - \overline{X}\right)^2 + \sum_{j=1}^{n_2} \left(Y_j - \overline{Y}\right)^2 \right\} = \frac{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}{n_1 + n_2 - 2}$$

Test d'égalité de l'espérance de deux lois normales

- (suite) et $z_{1-\alpha/2}$ est le quantile d'une loi de student de degré de liberté $n_1 + n_2 - 2$. Car

$$T = \frac{\overline{X} - \overline{Y}}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \curvearrowright ST(n_1 + n_2 - 2)$$

Procédure de test

Pour le test d'égalité des espérances de deux lois normales, il est indiqué de de procéder en 1er lieu à un test d'égalité des variances

Test d'égalité de l'espérance de deux lois normales

- Cas où les deux variances sont inconnues et différentes ($\sigma_1^2 \neq \sigma_2^2$) : La région critique est définie par

$$W = \left\{ (X_1, X_2, \cdots, X_{n_1}), (Y_1, Y_2, \cdots, Y_{n_2}), \quad \text{tq} \quad |\overline{X} - \overline{Y}| > \sqrt{\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}} z_{1-\alpha/2} \right\}$$

et

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}}} \curvearrowright ST(\eta)$$

$z_{1-\alpha/2}$ est le quantile d'une loi de student de degré de liberté η, approximé par :

$$\eta = \frac{\left\{ \frac{S_x^2}{n_1} + \frac{S_x^2}{n_2} \right\}^2}{\left(\frac{S_x^2}{n_1} \right)^2 / (n_1 - 1) + \left(\frac{S_y^2}{n_2} \right)^2 / (n_2 - 1)}$$

Test d'égalité de deux proportions

On considère $(X_1, X_2, \cdots, X_{n_1})$ un échatillon issu d'une loi de Bernouil de paramètre p_1 et $(Y_1, Y_2, \cdots, Y_{n_2})$ un échatillon issu d'une loi de Bernouilli de paramètre p_2. On se propose de réaliser le test suivant, pour un risque de 1ère espéèce α :

$$\begin{cases} H_0 : & p_1 = p_2 \\ H_1 : & p_1 \neq p_2 \end{cases}$$

La vraisemblance d'un échantillon $(X_1, X_2, \cdots, X_n)$ issu d'une loi de Bernoulli de paramètre p s'écrit sous la forme :

$$L(X_1, X_2, \cdots, X_n) = p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} = p^n \exp\left(\sum_{i=1}^n X_i \ln(\frac{p}{1-p})\right)$$

Conclusion : La loi de Bernoulli appartient à la famille exponentielle.

Pour n_1 et n_2 assez grand, les deux estimateurs de p_1 et p_2, notés respectivement $\overline{X} = F_x$ et $\overline{Y} = F_y$ (F pour fréquence), convergent en loi vers des lois normales :

$$F_x \longrightarrow N(p_1, \frac{p_1(1-p_1)}{n_1}) \text{ et } F_y \longrightarrow N(p_2, \frac{p_2(1-p_2)}{n_2})$$

et

$$F_x - F_y = \overline{X} - \overline{Y} \longrightarrow N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_1}\right)$$

Région critique de comparaison de deux proportions

La région de test de comparaison de deux proportions p_1 et p_2 est défnie par :

$$\begin{aligned} W &= \left\{ |F_x - F_y| > \sqrt{F(1-F)\left\{\frac{1}{n_1} + \frac{1}{n_2}\right\}} u_{1-\alpha/2} \right\} \\ F &= (n_1 F_x + n_2 F_y)/(n_1 + n_2) \end{aligned}$$

Exercices

Exercice 1

Un commerçant a observé pendant une longue période que son bénéfice mensuel B suit une loi normale $N(m, \sigma^2)$. A la suite de difficultés de trésorerie, il est conduit à s'intéresser aux variations de ces bénéfices. Pendant les 9 derniers mois on a

$$\sum_{i=1}^{9}(b_i - \overline{b})^2 = 28800$$

1. Par quelle quantité le commerçant doit-il estimer (sans biais) la variance de son bénéfice ?
2. Construire un inervalle de confiance unilateral de la forme $[0, a[$ pour σ^2 au niveau 95%

Exercices

Exercice 2

Des statistiques antérieures ont permis d'établir que l'écart-type des montants des ventes d'un produit, par magasin, est estimé à $\hat{\sigma} = 200$ dinars. Suppposons que la population des montants des ventes suivent une loi normale. Quelle est la taille minimale de léchantilllon des ventes pour estimer les ventes moyennes, par magasin, à 100 dinars prés avec un niveau de confiance de 95%

Exercice 3

On considère $(X_1, X_2, \cdots, X_n)$ un échantillon de taille n extrait d'une loi de poisson $P(\lambda)$. Déterminerl'estimateur de λ par la méthode du maximum de vraisemblance, noté $\hat{\lambda}$.Quelle est la loi asymptotique de $\sqrt{n}\left(\hat{\lambda} - \lambda\right)$. Pour n assez grand, on utilise cette distribution pour construire un intervalle de confiance pour λ. Application numérique : $n = 64$, $\overline{X} = 1.25$ et $1 - \alpha = 95\%$.

Exercices

Exercice 4

Le directeur d'un magasin veut estimer déterminer le pourcentage des personnes qui ne quittent le magasin qu'après avoir acheté au moins un article.

1) Quelle doit être la taille de l'échantillon à observer si on veut déterminer ce pourcentage avec une précision absolue de 3% et avec un niveau de confiance de 0.95%
2) On a observé 1100 personnes quittant le magasin, 750 d'entre elles avaient acheté au moins un article. Donner une estimation par intervalle de confiance du pourcentage d'individus qui sont réellement acheteurs, à 95%, en utilisant :
   - l'approximation de p par $\hat{p}$
   - la borne supérieure de la variance

Exercice 5

On considère $(X_1, X_2, \cdots, X_{n_1})$ un échatillon issu d'une loi normale $N(m_1, \sigma_1^2)$ et $(Y_1, Y_2, \cdots, Y_{n_2})$ un échatillon issu d'une loi normale $N(m_2, \sigma_2^2)$. On se propose de réaliser le test suivant, pour un risque de 1ère espéèce α :

$$\begin{cases} H_0 : & \sigma_1^2 = \sigma_2^2 \\ H_1 : & \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

- Déterminer la région critique du test
- Quelle est la décision du test pour $n_1 = 10$, $n_2 = 11$, $\alpha_1 = \alpha_2 = 2.5\%$, $S_x^2 = 1.5$ et S_y^2. En déduire la puissance du test.

Exercices

Exercie 6

Pour un échantillon $(X_1, X_2, \cdots, X_n)$ issu d'une loi exponentielle de paramètre $\xi(\frac{1}{\theta})$. Pour un risque de 1ère espèce α, déterminer la région critique de test suivant :

$$\begin{cases} H_0 : & \theta = \theta_0 \\ H_1 : & \theta \neq \theta_0 \end{cases}$$

Exercice 7

Un sondage réalisé sur 1000 personnes visant émettre un indicateur sur la notorité d'une marque, a conduit au choix d la marque A par 600 personnes.

- Déterminer l'intervalle de confiance de la notorité de chaque marque à 95%.
- Construire un intervalle de confiance à 95% pour la différence de notorité entre les deux marques

Exercices

Exercice 8

On considère deux échantillons $(X_1, X_2, \cdots, X_n1)$ et $(Y_1, Y_2, \cdots, Y_n2)$ suivant respectivement la loi normale $N(m_1, \sigma_1^2)$ et la loi normale $N(m_2, \sigma_2^2)$.

- Déterminer la région critique pour le test $H_0 : m_1 = m_2$ vs $H_1 : m_1 \neq M_2$.
- Pour $\overline{X} = 154$, $n1 = 15$, $S_x^2 = 35$, $\overline{Y} = 145$, $n2 = 25$, $S_y^2 = 12$, et $\alpha = 5\%$ réaliser test. En déduire la fonction puissance de ce test.
- Déterminer la taille minimale de l'échantillon pour un écrat $m_1 - m_2 = 6$, un risque de 2ème espèce égal à 10% et un coeffcient de parité égal à 1.
- Reprendre les deux premières questions pour le test de l'hypothèse $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_0 : \sigma_1^2 \neq \sigma_2^2$

MERCI POUR VOTRE ATTENTION

Pr. Mokhtar KOUKI
mokhtar.kouki@essai.ucar.tn