MAP 433 : Introduction aux méthodes statistiques. Cours 6

2 Octobre 2015

Aujourd'hui

1. Tests statistiques
   - Notion de test et d'erreur de test
   - Lemme de Neyman-Pearson
2. p-valeur
3. Tests gaussiens
   - Tests sur la moyenne
4. Compléments : p-valeur et liens entre tests et régions de confiance

Exemple introductif

- On observe 10 lancers d'une pièce de monnaie et on obtient le résultat suivant :

$$(P, P, F, F, P, F, P, P, F, P).$$

La pièce est-elle équilibrée ?

- Répondre à cette question revient à construire une procédure de décision :

$$\varphi = \varphi(P, P, F, F, P, F, P, P, F, P)$$
$$= \begin{cases} 0 & \text{on accepte l'hypothèse la pièce est équilibrée} \\ 1 & \text{on rejette l'hypothèse la pièce est équilibrée} \end{cases}$$

Résolution

- On associie l'expérience statistique (par exemple)

$$\mathcal{E}^{10} = \big(\{0,1\}^{10}, \text{parties de}(\{0,1\}^{10}), \{\mathbb{P}_\theta^{10}, \theta \in \Theta = [0,1]\}\big),$$

avec $(P = 0,\ F = 1)$

$$\mathbb{P}_\theta^{10} = \big(\theta\delta_0(dx) + (1-\theta)\delta_1(dx)\big)^{\otimes 10}.$$

- Hypothèse nulle : la pièce est équilibrée

$$\boxed{H_0 : \theta \in \Theta_0 = \{1/2\}}$$

- Hypothèse alternative : la pièce est truquée

$$\boxed{H_1 : \theta \in \Theta_1 = \Theta \setminus \{1/2\}}$$

Résolution

- Θ_0= ensemble des paramètres sous laquelle l'hypothèse nulle est satifaite
- Θ_1= ensemble des paramètres sous laquelle l'hypothèse nulle n'est pas satisfaite = alternative
- $\Theta = \Theta_0 \cup \Theta_1$.

Règle de décision

- On note Z l'observation.
- On construit une règle de décision simple :

$$\varphi(Z) = \mathbb{1}_{\mathcal{R}}(Z) = \begin{cases} 0 & \text{on accepte l'hypothèse} \\ 1 & \text{on rejette l'hypothèse.} \end{cases}$$

- $\mathcal{R} \subset \mathfrak{Z}$ (espace des observables) : zone de rejet ou région critique.
- Exemple[1]

$$\mathcal{R} = \left\{ \left| \widehat{\theta}(z) - 1/2 \right| > t_0 \right\}, \;\; \widehat{\theta}(Z) = \widehat{\theta}_n^{\,\mathtt{mv}} \left(\overset{exemple}{=} 0,6 \right)$$

où t_0 est un seuil à choisir... Comment ?

[1] léger abus de notation...

Terminologie

- Une règle de décision (non-randomisée) assigne à chaque réalisation z de l'observation Z une décision.
- Si $z \in \mathcal{A}$, l'hypothèse nulle est acceptée; autrement, l'hypothèse est rejetée.
- Terminologie:
  - $\mathcal{A}=$ zone d'acceptation,
  - $\mathcal{R}=$ zone de rejet ou région critique.

Erreur de décision

- Lorsque l'on prend la décision $\varphi(Z)$, on peut se tromper de deux manières :

$$\text{Rejeter } H_0 \;\; (\varphi(Z) = 1) \;\; \text{alors que} \;\; \theta = \frac{1}{2}$$

ou encore

$$\text{Accepter } H_0 \;\; (\varphi(Z) = 0) \;\; \text{alors que} \;\; \theta \neq \frac{1}{2}.$$

- Erreur de première espèce (=rejeter à tort)

$$\mathbb{P}^{10}_{\frac{1}{2}}\left[\varphi(Z) = 1\right]$$

- Erreur de seconde espèce (=accepter à tort)

$$\left(\mathbb{P}^{10}_{\theta}\left[\varphi(Z) = 0\right], \;\; \theta \neq \frac{1}{2}\right).$$

Retour à l'exemple

- Sous $\mathbb{P}_\theta$, $n\hat{\theta}(Z) = \sum_{i=1}^n Z_i$ suit une loi binomiale de paramètre de succès θ. En notant $\mathrm{Bin}_{n,\theta}$ la fonction de répartition de la loi binomiale de paramètre θ,

$$\mathbb{P}_\theta(|\hat{\theta}(Z) - 1/2| > t_0) \\ = 1 - \{\mathrm{Bin}_{n,\theta}(n(1+t_0)/2) - \mathrm{Bin}_{n,\theta}(n(1-t_0)/2)\}\,.$$

- Erreur de première espèce:

$$\mathbb{P}_{1/2}(|\hat{\theta}(Z) - 1/2| > t_0) \\ = 1 - \{\mathrm{Bin}_{n,1/2}(n(1+t_0)/2) - \mathrm{Bin}_{n,1/2}(n(1-t_0)/2)\}$$

Erreur de première espèce

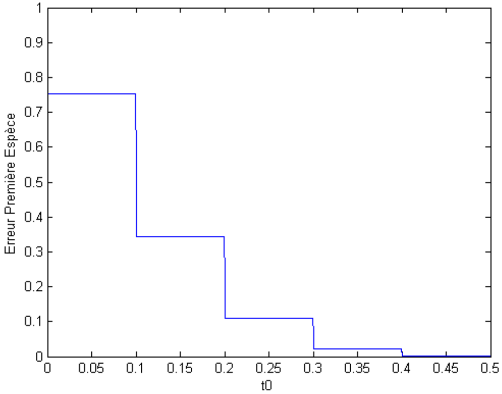

Figure: Erreur de Première Espèce en fonction de la valeur critique: fonction décroissante de t_0 !

Retour à l'exemple

- Sous $\mathbb{P}_\theta$, $n\hat{\theta}(Z) = \sum_{i=1}^n Z_i$ suit une loi binomiale de paramètre de succès θ. En notant $\mathrm{Bin}_{n,\theta}$ la fonction de répartition de la loi binomiale,

$$\mathbb{P}_\theta(|\hat{\theta}(Z) - 1/2| \leq t_0) = \{\mathrm{Bin}_{n,\theta}(n(1+t_0)/2) - \mathrm{Bin}_{n,\theta}(n(1-t_0)/2)\}\,.$$

- Erreur de seconde espèce: pour une valeur critique fixée:

$$\theta \mapsto \mathbb{P}_\theta(|\hat{\theta}(Z) - 1/2| \leq t_0) = \mathrm{Bin}_{n,1/2}(n(1+t_0)/2) - \mathrm{Bin}_{n,1/2}(n(1-t_0)/2)$$

Erreur de seconde espèce

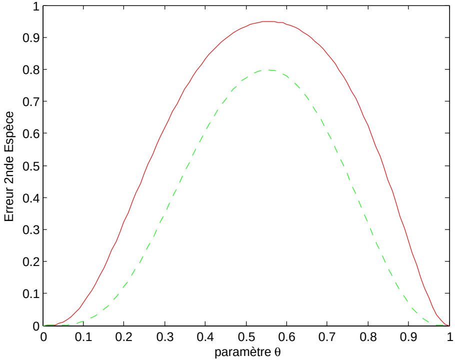

Figure: Erreur de deuxième Espèce en fonction du paramètre $\theta \in \Theta_1 = [0,1] \setminus \{1/2\}$ pour deux valeurs du seuil critique: $t_0 = 0.3$, Erreur 1ere espèce 0.0654 et $t_0 = 0.2$, Erreur 1ère espèce 0.22

Puissance du test

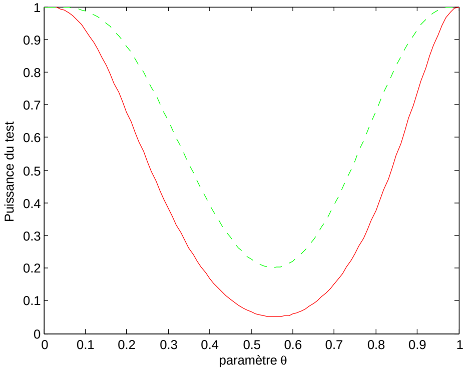

Figure: Puisance du test en fonction du paramètre $\theta \in \Theta_1 = [0,1] \setminus \{1/2\}$ pour deux valeurs du seuil critique: $t_0 = 0.3$. Erreur 1ere espèce: 0.0654 et $t_0 = 0.2$, Erreur 1ère espèce: 0.22

Erreur 1ère espèce / Puissance

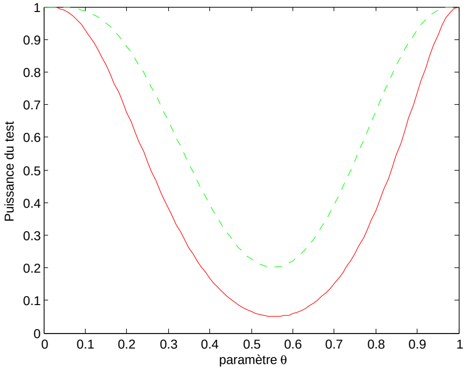

Figure: Erreur de première espèce / puisance pour une contre-alternative fixée $\theta_1 = 0.75$ pour différentes valeurs du seuil critique t_0.

Conclusion provisoire

- Un bon test φ devrait garantir simultanément des erreurs de première et seconde espèce petites.
- Mais il faut réaliser un compromis entre erreur de 1ère espèce et erreur de 2nde espèce (ou de façon équivalente entre erreur de 1ère espèce et puissance).
- Question: comment aborder la notion d'optimalité et comment construire une procédure de test satisfaisante ?

Définition formelle

- Situation : $\mathcal{E} = \big(\mathcal{Z}, \mathfrak{Z}, \{\mathbb{P}_\theta, \theta \in \Theta\}\big)$ engendrée par l'observation Z.
- Hypothèse nulle et alternative : $\Theta_0 \subset \Theta$ et $\Theta_1 = \Theta \setminus \Theta_0$

Definition (Test simple)

Un test (simple) de l'hypothèse nulle $H_0 : \theta \in \Theta_0$ contre l'alternative $H_1 : \theta \in \Theta_1$ est une statistique $\varphi = \varphi(Z) \in \{0, 1\}$. (Fonction d') erreur de première espèce :

$$\theta \in \Theta_0 \rightsquigarrow \mathbb{P}_\theta\left[\varphi(Z) = 1\right]$$

(Fonction d') erreur de seconde espèce

$$\theta \in \Theta_1 \rightsquigarrow \mathbb{P}_\theta\left[\varphi(Z) = 0\right] = 1 - \mathit{puissance}_\varphi(\theta).$$

Principe de Neyman

- On disymétrise les hypothèses H_0 et H_1 : H_0 est plus importante que H_1 dans le sens suivant : on impose une erreur de première espèce prescrite.

Definition

Pour $\alpha \in [0,1]$, *un test* $\varphi = \varphi_\alpha$ *de l'hypothèse nulle* $H_0 : \theta \in \Theta_0$ *contre une alternative* H_1 *est de niveau* α *si*

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta \big[\varphi_\alpha = 1\big] \leq \alpha.$$

- Un test de niveau α ne dit rien sur l'erreur de seconde espèce (comportement sur l'alternative).

Principe de Neyman (cont.)

- Choix de la disymétrisation = choix de modélisation.
- Principe de Neyman : $\alpha \in (0,1)$, parmi les test de niveau α, chercher celui (ou ceux) ayant une erreur de seconde espèce minimale.

Definition

Un test de niveau α est dit Uniformément Plus Puissant (UPP) si son erreur de seconde espèce est minimale parmi celles des tests de niveau α.

- Pour le cas d'une hypothèse simple contre une alternative simple, un test UPP existe.

Règle de décision randomisée

- Pour toute valeur de l'observation z, la règle de décision choisit alternative avec une probabilité $\varphi(z)$ et l'hypothèse nulle avec la probabilité $1-\varphi(z)$.
- Une procédure de test randomisée est entièrement spécifiée par la donnée de la fonction critique du test $\varphi : z \to \varphi(z) \in [0,1]$. Si φ prend simplement les valeurs 0 et 1, on obtient un test non randomisé.
- La probabilité de rejet est donnée, pour tout $\theta \in \Theta$, par $\mathbb{E}_\theta[\varphi(Z)]$.

Principe de Neyman-Pearson

Le problème revient donc à maximiser la puissance du test

$$\pi(\theta) = \mathbb{E}_\theta[\varphi(Z)]\,, \theta \in \Theta_1$$

sous la contrainte que le niveau du test soit inférieure à α

$$\mathbb{E}_\theta[\varphi(Z)] \leq \alpha$$

Un cas élémentaire

- Supposons que $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$.
- On note $p_0(z) = f(\theta_0, z)$ et $p_1(z) = f(\theta_1, z)$ les densités des lois $\mathbb{P}_{\theta_0}$ et $\mathbb{P}_{\theta_1}$ par rapport à une mesure de domination μ (existe toujours)

Théorème (Existence d'un test de niveau α)

Soit $\alpha \in [0,1]$. Pour tester $H_0 : \{\theta = \theta_0\}$ contre l'alternative $H_1 : \{\theta = \theta_1\}$, il existe un test φ et une constante c_α telle que

$$\mathbb{E}_{\theta_0}[\phi(Z)] = \alpha$$

et

$$\varphi(z) = \begin{cases} 1 & \textit{quand } p_1(z) > c_\alpha p_0(z) \\ 0 & \textit{quand } p_1(z) < c_\alpha p_0(z) \end{cases}$$

Preuve Existence-1

- Pour $\alpha = 0$ et $\alpha = 1$, le résultat est élémentaire en posant $c_0 = \infty$ et $c_1 = 0$.
- On suppose $\alpha \in (0,1)$ et considère la fonction

$$c \mapsto A(c) = \mathbb{P}_0(p_1(Z) > cp_0(Z)) = \mathbb{P}_0(p_1(Z)/p_0(Z) > c, p_0(Z) > 0)\,,$$

- La fonction A est décroissante, continue à droite et admet des limites à gauche ($c \mapsto 1 - A(c)$ est la fonction de répartition de la variable aléatoire $p_1(Z)/p_0(Z)$ qui est définie $\mathbb{P}_0$-p.s.):

$$A(c^-) - A(c) = \mathbb{P}_0(p_1(Z)/p_0(Z) = c)\,, A(-\infty) = 1, A(+\infty) = 0$$

Preuve Existence-2

- Pour tout $\alpha \in (0,1)$, il existe c_α tel que $A(c_\alpha) \leq \alpha \leq A(c_\alpha^-)$.
- On considère le test φ_α définit par

$$\varphi_\alpha(z) = \begin{cases} 1 & \text{quand } p_1(z) > c_\alpha p_0(z) \\ \frac{\alpha - A(c_\alpha)}{A(c_\alpha^-) - A(c_\alpha)} & \text{quand } p_1(z) = c_\alpha p_0(z) \\ 0 & \text{quand } p_1(z) < c_\alpha p_0(z) \end{cases}$$

  Si $A(c_\alpha^-) = A(c_\alpha)$, alors $\mathbb{P}_0(p_1(Z) = c_\alpha p_0(Z)) = 0$ et il n'y pas lieu de spécifier la valeur du test sur cet événement.
- Le niveau de ϕ est donné par

$$\begin{aligned} \mathbb{E}_0[\varphi_\alpha(Z)] &= \mathbb{P}_0\left(\frac{p_1(Z)}{p_0(Z)} > c_\alpha\right) + \frac{\alpha - A(c_\alpha)}{A(c_\alpha^-) - A(c_\alpha)} \mathbb{P}_0\left(\frac{p_1(Z)}{p_0(Z)} = c_\alpha\right) \\ &= \alpha \end{aligned}$$

Test Uniformément Plus Puissant

Théorème

Un test φ *vérifiant*

$$\mathbb{E}_{\theta_0}[\phi(Z)] = \alpha$$

et

$$\varphi(z) = \begin{cases} 1 & \text{quand } p_1(z) > c_\alpha p_0(z) \\ 0 & \text{quand } p_1(z) < c_\alpha p_0(z) \end{cases}$$

est uniformément le plus puissant pour tester l'hypothèse nulle $H_0 : \{\theta = \theta_0\}$ *contre l'alternative* $H_1 : \{\theta = \theta_1\}$.

Preuve test U.P.P.-1

- Soit φ un test satisfaisant les conditions

$$\mathbb{E}_{\theta_0}[\phi(Z)] = \alpha$$

et

$$\varphi(z) = \begin{cases} 1 & \text{quand } p_1(z) > c_\alpha p_0(z) \\ 0 & \text{quand } p_1(z) < c_\alpha p_0(z) \end{cases}$$

- Soit φ^* un test de niveau $\mathbb{E}_0[\varphi^*(Z)] \leq \alpha$

Preuve test U.P.P.-1

- On note:

$$S^+ = \{z : \varphi(z) - \varphi^*(z) > 0\}$$
$$S^- = \{z : \varphi(z) - \varphi^*(z) < 0\}.$$

- Pour $z \in S^+$, $\phi(z) > 0$ et donc $p_1(z) \geq c_\alpha p_0(z)$ (car $\varphi(z) = 0$ si $p_1(z) < c_\alpha p_0(z)$).
- Pour $z \in S^-$, $\phi(z) < 1$ et donc $p_1(z) \leq c_\alpha p_0(z)$ (car $\varphi(z) = 1$ si $p_1(z) > c_\alpha p_0(z)$).
- Par conséquent:

$$\int (\varphi - \varphi^*)(p_1 - c_\alpha p_0)\mathrm{d}\mu$$
$$= \int_{S^+ \cup S^-} (\varphi - \varphi^*)(p_1 - c_\alpha p_0)\mathrm{d}\mu \geq 0$$

Preuve test U.P.P.-1

Conclusion

$$\int (\varphi - \varphi^*)(p_1 - c_\alpha p_0)\mathrm{d}\mu \geq 0$$

ce qui implique

$$\begin{aligned}
\int (\varphi - \varphi^*) p_1 \mathrm{d}\mu &\geq c_\alpha \int (\varphi - \varphi^*) p_0 \mathrm{d}\mu \\
&\geq c_\alpha \left\{\mathbb{E}_0[\varphi(Z)] - \mathbb{E}_0[\varphi^*(Z)]\right\} \\
&= c_\alpha \{\alpha - \mathbb{E}_0[\varphi^*(Z)]\} \geq 0\,.
\end{aligned}$$

Puissance d'un test U.P.P

Lemme

Notons π la puissance du test U.P.P de niveau α du test $H_0 = \{\theta = \theta_0\}$ contre l'alternative $H_1 = \{\theta = \theta_1\}$. Alors $\alpha \leq \pi$ avec égalité si $\mathbb{P}_{\theta_0} = \mathbb{P}_{\theta_1}$.

Proof.

Comme le test $\varphi(z) \equiv \alpha$ a un niveau α, nous avons $\alpha \leq \pi$. $\square$

Exemple de mise en oeuvre

- On observe

$$Z = (X_1, \ldots, X_n) \sim_{\text{i.i.d.}} \mathcal{N}(\theta, \sigma^2) \quad \sigma^2 \text{ connu.}$$

,

- Construction du test de N-P. de $H_0 : \theta = \theta_0 = 0$ contre $H_1 : \theta = \theta_1$, avec $0 < \theta_1$.
- Mesure dominante $\mu^n =$ mesure de Lebesgue sur $\mathbb{R}^n$ et

$$f(\theta, Z) = \tfrac{1}{(2\pi)^{n/2}} \exp\big(- \frac{1}{2} \sum_{i=1}^n X_i^2 + n\theta \overline{X}_n - \frac{n\theta^2}{2} \big).$$

- Rapport de vraisemblance

$$\frac{f(\theta_1, Z)}{f(\theta_0, Z)} = \exp\left(\frac{n\theta_1}{\sigma^2} \overline{X}_n - \frac{n\theta_1^2}{2\sigma^2} \right).$$

Exemple (cont.)

- Zone de rejet du test de N-P. :

$$\{f(\theta_1, Z) > cf(\theta_0, Z)\} = \left\{\frac{n\theta_1}{\sigma^2}\overline{X}_n - \frac{n\theta_1^2}{2\sigma^2} > \log c\right\}$$
$$= \left\{\overline{X}_n > \frac{\theta_1}{2} + \frac{\sigma^2 \log c}{n\theta_1}\right\}.$$

La région de rejet est donc de la forme

$$\{\overline{X}_n > t_{n,\alpha}\}$$

- Choix de $t_{n,\alpha}$. On choisit $t_{n,\alpha}$ pour ajuster le niveau du test

$$\mathbb{P}_{\theta_0}\left[\overline{X}_n > t_{n,\alpha}\right] = \alpha.$$

Exemple (fin)

- On note $z_{1-\alpha}$ le quantile $1-\alpha$ d'une loi Gaussienne standardisée: $\Phi(z_{1-\alpha}) = 1-\alpha$.
- Sous $\mathbb{P}_{\theta_0}$, $\sqrt{n}\bar{X}_n/\sigma \sim \mathcal{N}(0,1)$, par conséquent

$$\begin{aligned}\mathbb{P}_{\theta_0}\left(\bar{X}_n > t_{n,\alpha}\right) &= \mathbb{P}_{\theta_0}\left(\frac{\sqrt{n}\bar{X}_n}{\sigma} > \frac{\sqrt{n}}{\sigma}t_{n,\alpha}\right)\\ &= 1-\Phi\left(\frac{\sqrt{n}}{\sigma}t_{n,\alpha}\right)\end{aligned}$$

- Conclusion

$$t_{n,\alpha} = \frac{\sigma z_{1-\alpha}}{\sqrt{n}}$$

On remarque d'ailleurs que le seuil critique ne dépend pas de θ_1 et que le même test est U.P.P. contre toutes les contre-alternatives $\theta_1 > 0$.

p-valeur

- Supposons que sous $\mathbb{P}_0$, la distribution du rapport de vraisemblance $p_1(Z)/p_0(Z)$ est continue.
- Le test U.P.P. de niveau α est non-randomisé et rejette l'hypothèse nulle si $p_1(Z)/p_0(Z) > c_\alpha$, où la constante c_α est choisie de façon à ce que
$$\mathbb{P}_0\big(p_1(Z)/p_0(Z) \geq c_\alpha\big) = \alpha.$$
- Lorsque l'on fait fait varier le niveau α, on obtient ainsi une famille de régions de réjections, $\{\mathcal{R}_\alpha\}_{\alpha\in[0,1]}$ qui sont, dans de nombreux cas, d'intérêt emboitées
$$\mathcal{R}_\alpha \subset \mathcal{R}_{\alpha'} \quad \text{si } \alpha < \alpha'.$$

p-valeur

- Lorsque l'on fait fait varier le niveau α, on obtient ainsi une famille de régions de réjections, $\{\mathcal{R}_\alpha\}_{\alpha \in [0,1]}$, qui sont dans de nombreux cas d'intérêt emboitées

$$\mathcal{R}_\alpha \subset \mathcal{R}_{\alpha'} \quad \text{si } \alpha < \alpha'.$$

- Lorsque cette condition est satisfaite, il est intéressant de déterminer
  - non seulement si le test est accepté ou rejeté à un niveau de signification donné...
  - mais aussi de déterminer le plus petit niveau de signification auquel l'hypothèse serait rejetée.

p-valeur

Definition (p-valeur)

La p-valeur d'un test pour une observation Z donnée est le plus petit niveau de signification auquel le test serait rejeté

$$\hat{p}(Z) = \inf\{\alpha : Z \in \mathcal{R}_\alpha\}.$$

- Une petite p-valeur suggère que l'observation contredit l'hypothèse.
- Une grande p-valeur s'interprète en faveur de ne pas vouloir rejeter l'hypothèse de base.

Exemple

- On reprend l'exemple du test $H_0 = \{\theta = \theta_0 = 0\}$ contre $H_1 = \{\theta = \theta_1 > 0\}$.
- Dans ce cas, pour un niveau α, la région de réjection est

$$\mathcal{R}_\alpha = \left\{ \bar{X}_n > \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

- Si $0 < \alpha < \alpha' < 1$, on a $z_{1-\alpha} \geq z_{1-\alpha'}$ et donc

$$\mathcal{R}_\alpha = \left\{ \bar{X}_n \geq \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\} \subset \mathcal{R}_{\alpha'} = \left\{ \bar{X}_n \geq \frac{\sigma}{\sqrt{n}} z_{1-\alpha'} \right\}$$

Exemple: calcul explicite de la p-valeur

- En notant Φ^{-1} la fonction des quantiles, $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$.
- Par conséquent, nous avons

$$\left\{\bar{X}_n > \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\right\} = \left\{\bar{X}_n > \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha)\right\}$$
$$= \left\{\frac{\sqrt{n}\bar{X}_n}{\sigma} > \Phi^{-1}(1-\alpha)\right\} = \left\{\alpha > 1 - \Phi\left(\frac{\sqrt{n}\bar{X}_n}{\sigma}\right)\right\}$$

- Pour une valeur donnée de la moyenne empirique $\bar{X}_n$, l'infimum par rapport au niveau α est

$$\hat{p} = 1 - \Phi\left(\frac{\sqrt{n}\bar{X}_n}{\sigma}\right).$$

Tests gaussiens incontournables

- On observe
$$\boldsymbol{Y} = (Y_1, \ldots, Y_n) \sim \mathcal{N}(\mu, \sigma^2 \mathrm{Id}_n).$$
- Test sur la moyenne, variance connue
$$H_0 : \mu \leq \mu_0 \quad \text{contre} \quad H_1 : \mu > \mu_0$$
- Principe on estime μ et on rejette H_0 si l'estimateur est plus grand que μ_0.
$$\mathcal{R}(c_\alpha) = \left\{\overline{Y}_n - \mu_0 \geq c_\alpha\right\}, \quad c_\alpha \text{ à déterminer.}$$
- On choisit c_α de sorte que
$$\sup_{\mu \leq \mu_0} \mathbb{P}_\mu \left[\mathcal{R}(c_\alpha)\right] \leq \alpha.$$

Calcul de c_α

- Majoration de l'erreur de première espèce. Si $\mu \leq \mu_0$, on a

$$\begin{aligned}\mathbb{P}_\mu\left[\overline{Y}_n - \mu_0 \geq c_\alpha\right] &= \mathbb{P}_\mu\left[\left(\mu + \tfrac{\sigma}{\sqrt{n}}\xi^{n,\mu}\right) - \mu_0 \geq c_\alpha\right] \\ &= \mathbb{P}_\mu\left[\tfrac{\sigma}{\sqrt{n}}\xi^{n,\mu} \geq c_\alpha + (\mu_0 - \mu)\right] \\ &\leq \mathbb{P}_\mu\left[\tfrac{\sigma}{\sqrt{n}}\xi^{n,\mu} \geq c_\alpha\right].\end{aligned}$$

  où $\xi^{n,\mu}$ est, en loi sous $\mathbb{P}_\mu$, une gaussienne standard.
- Petit miracle : la loi de $\xi^{n,\mu}$ sous $\mathbb{P}_\mu$ ne dépend pas de μ Donc

$$\mathbb{P}_\mu\left[\tfrac{\sigma}{\sqrt{n}}\xi^{n,\mu} \geq c_\alpha\right] = 1 - \Phi\left(\tfrac{\sqrt{n}}{\sigma}c_\alpha\right) \overset{\text{on veut}}{\leq} \alpha.$$

- Le choix $c_{\alpha,n} = \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)$ conduit à la zone de rejet $\mathcal{R}(c_\alpha)$ maximale.

Contrôle de l'erreur de seconde espèce

- On a construit un test de niveau α parmi une classe donnée a priori de tests basés sur un estimateur raisonnable, de sorte que l'on ait une zone de rejet maximale. Désormais, $c_{\alpha,n}$ est fixé.
- On évalue à la main l'erreur de seconde espèce ou la fonction de puissance

$$\begin{aligned}\mu \in (\mu_0, +\infty) \rightsquigarrow \mathbb{P}_\mu\left[\overline{Y}_n - \mu_0 < c_{\alpha,n}\right] \\ = 1 - \text{puissance du test au point } \mu\end{aligned}$$

- Montrer que pour tout $\mu > \mu_0$, on a $\mathbb{P}_\mu\left[\overline{Y}_n - \mu_0 < c_{\alpha,n}\right] \to 0$ lorsque $n \to \infty$.
- Pour l'optimalité dans un sens plus fort, il faut d'autres outils.

Autres tests classiques gaussiens

- Ingrédient principal :

$$s_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2 = \frac{n}{n-1}\left(\widehat{\sigma}_n^2\right)^{\texttt{mv}}$$

  alors

$$(n-1)\frac{s_n^2}{\sigma^2} \sim \chi^2(n-1)$$

  et

$$\frac{\sqrt{n}(\overline{Y}_n - \mu)}{s_n} \sim \text{Student}(n-1)$$

  et ces variables sont pivotales : leur loi ne dépend pas de μ, σ^2 sous $\mathbb{P}_{\mu,\sigma^2}$.
- Les lois du χ^2 et de Student (à k degrés de liberté) sont classiques et s'étudient indépendamment.

p-valeurs

- Exemple : on observe

$$X_1, \ldots, X_n \sim_{\text{i.i.d.}} \mathcal{N}(\mu, \sigma^2), \quad \sigma^2 \text{ connu.}$$

- Objectif: tester $H_0 : \mu = 0$ contre $H_1 : \mu \neq 0$.
- Au niveau $\alpha = 5\%$, on rejette si

$$\left|\overline{X}_n\right| > \frac{\phi^{-1}(1-\alpha/2)}{\sqrt{n}}$$

- Application numérique : $n = 100$, $\overline{X}_{100} = 0.307$. On a $\frac{\phi^{-1}(1-0.05/2)}{\sqrt{100}} \approx 0.196$. on rejette l'hypothèse....

p-valeur (cont.)

- Et pour un autre choix de α ?. Pour $\alpha = 0.01$, on a $\frac{\phi^{-1}(1-0.05/2)}{\sqrt{100}} \approx 0.256$. On rejette toujours... Pour $\alpha = 0.001$, on a $\frac{\phi^{-1}(1-0.05/2)}{\sqrt{100}} \approx 0.329$. On accepte H_0 !
- Que penser de cette petite expérience ?
  - En pratique, on a une observation une bonne fois pour toute (ici 0.307) et on choisit α... comment ?
  - On ne veut pas α trop grand (trop de risque), mais en prenant α de plus en plus petit... on va fatalement finir par accepter H_0 !
- Défaut de méthodologie inhérent au principe de Neyman (contrôle de l'erreur de première espèce).

p-valeur

- Quantité significative : non par le niveau α, mais le seuil de basculement de décision : c'est la p-valeur (p-value) du test.

Definition

Soit $\mathcal{R}_\alpha$ une famille de zones de rejet d'un test de niveau α pour une hypothèse H_0 contre une alternative H_1. Soit Z l'observation associée à l'expérience. On a $Z \in \mathfrak{Z}$ et $\mathcal{R}_0 = \mathfrak{Z}$. On appelle p-valeur du test la quantité

$$p - valeur(Z) = \inf\{\alpha, Z \in \mathcal{R}_\alpha\}.$$

Interprétation de la p-valeur

- Une grande valeur de la p-valeur s'interprète en faveur de ne pas vouloir rejeter l'hypothèse.
- Ne pas vouloir rejeter l'hypothèse peut signifier deux choses :
  - L'hypothèse est vraie
  - L'hypothèse est fausse mais le test n'est pas puissant (erreur de seconde espèce grande).
- Souvent : la p-valeur est la probabilité (sous H_0) que la statistique de test d'une expérience copie soit $\geq$ à la statistique de test observée.
- Exemple du test du χ^2 et de l'expérience de Mendel (à suivre)