Estimation et tests d’hypothèses

Thierry Dhorne

14 septembre 2016

Rappels

État des lieux

- deux situations étudiées
  - ➤ test sur une probabilité (proportion)
  - ➤ test sur une espérance (moyenne)

Test sur une probabilité

★ problèmes dûs

— au caractère discret de la règle de décision
— au type d'échantillonnage

➤ nous reviendrons sur ces points

Test sur une espérance pour une variance connue

- pas de problème particulier
- ➤ il existe plusieurs types de couples d'hypothèses

— $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1 (\mu_1 > \mu_0)$

— $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$

— $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$

— $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$

★ les 3 premiers couples (et leur symétrique) conduisent à des régions de rejet unilatérales

★ la dernière conduit à une région de rejet bilatérale

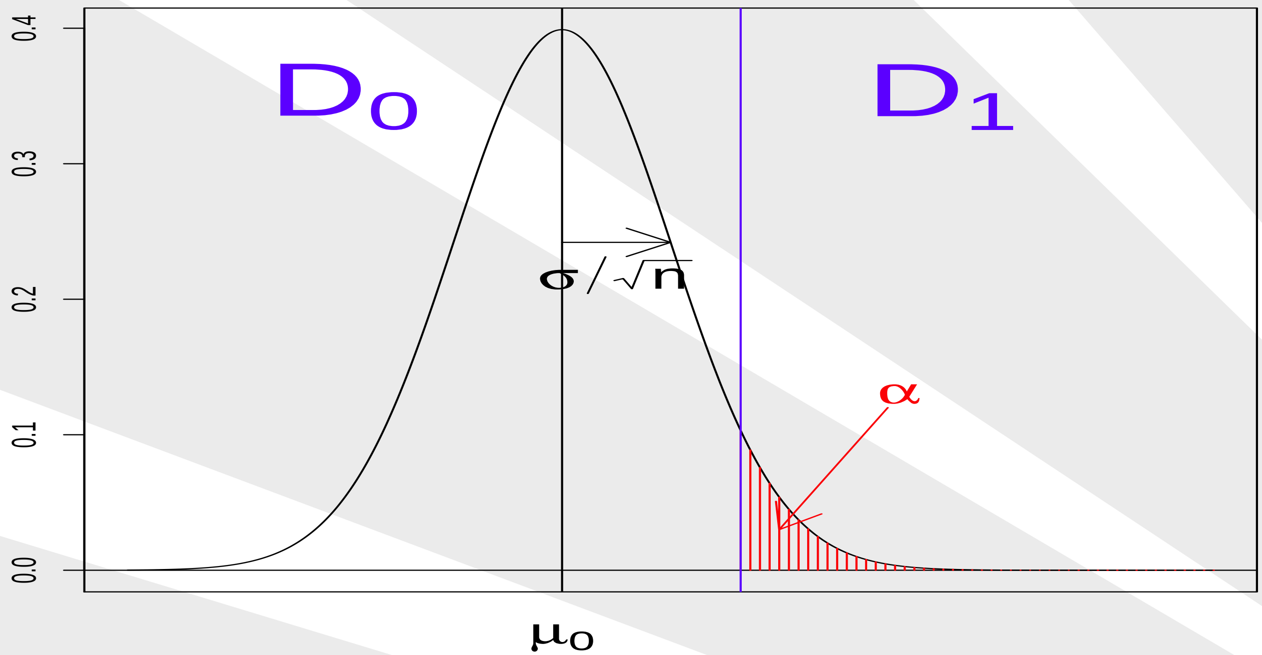
Normalisation de la région de rejet

- la statistique de test est toujours $\bar{X}$, moyenne d'un échantillon de taille n
- la loi de $\bar{X}$ est complètement déterminée

➤ $\bar{X} \rightsquigarrow \mathcal{N}(\mu_0, \frac{\sigma^2}{n})$

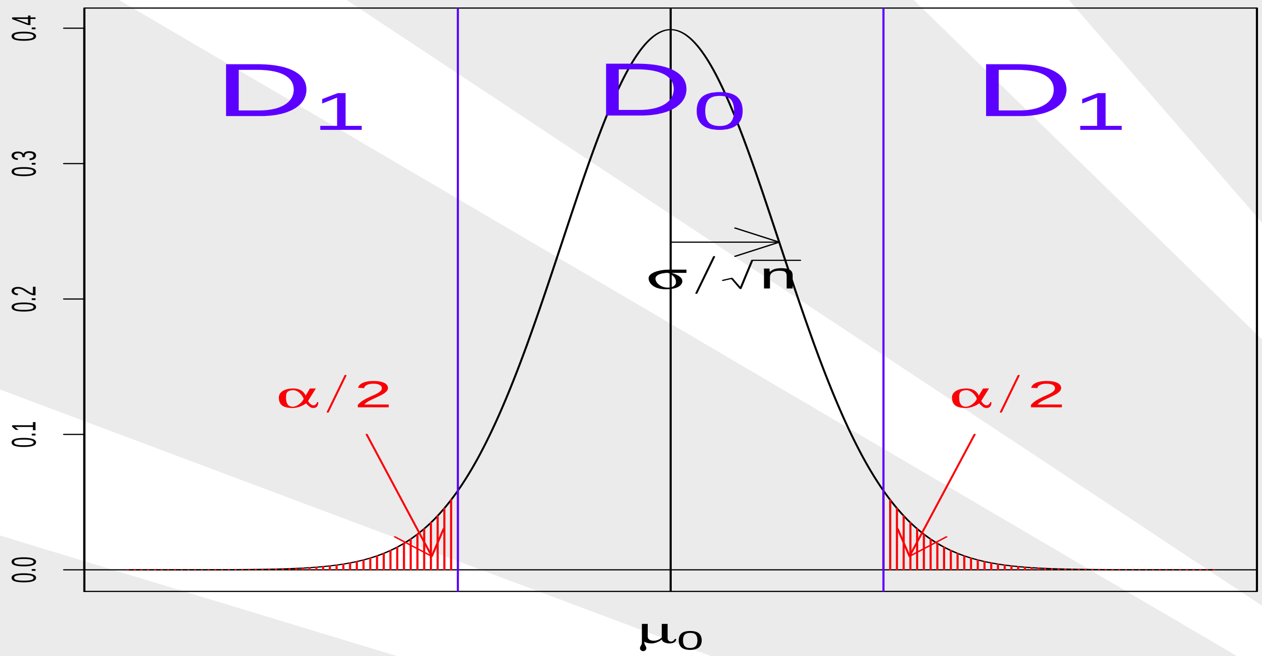
★ la normalisation ne dépend donc que du caractère uni ou bi-latéral précédent

Normalisation de la région de rejet (cas unilatéral)



- en noir la loi de la statistique de test sous H_0
- en rouge la normalisation par le risque de première espèce
- en bleu les régions de décisions et leur(s) frontières

Normalisation de la région de rejet (cas bilatéral)



- en noir la loi de la statistique de test sous H_0
- en rouge la normalisation par le risque de première espèce
- en bleu les régions de décisions et leur(s) frontières

Qui décide ?

- Peut-être ceci n'a pas été clair dans les amphis précédents
  - c'est la statistique de test T qui décide
  - en se réalisant

— si elle « tombe » dans D_0, on maintient H_O
— si elle « tombe » dans D_1, on décide H_1

  - donc par exemple quand on écrit $P(D1)$ on veut dire $P(T \in D_1)$

Conclusions

- bien que la situation soit légèrement irréaliste
- en général la variance est inconnue
- cet exemple est un exemple générique
- qui est généralisable à tout test paramétrique
- nous allons l'appliquer au test sur une variance

Test sur une variance

Test d'homogénéité

- problème opérationnel

➤ tester si la variabilité d'un process est sous contrôle ou hors contrôle

- le modèle est un modèle gaussien

➤ $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$,

- μ connu ou non (peu de différences)

➤ on a donc naturellement les deux hypothèses :

— $H_0 : \sigma^2 \leq \sigma_0^2$

— $H_1 : \sigma^2 > \sigma_0^2$

Statistique de test

- la statistique de test est naturellement l'estimateur du paramètre concerné
  - $\frac{1}{n-1}\sum_{r=1}^{n}(X_r - \bar{X})^2$

Régions de décision

- la région de rejet est naturellement unilatérale supérieure

➤ on rejette H_0 si le bruit estimé est au-dessus d'un certain seuil

Loi de la statistique de test

- $\frac{1}{n-1}\sum_{r=1}^{n}(X_r - \bar{X})^2$

➤ il faut donc connaître la loi de sommes de carrés de gaussiennes

Loi du carré d'une gaussienne

- on commence par une gaussienne centrée réduite
- $X \rightsquigarrow \mathcal{N}(0,1)$

➤ on cherche

$$\begin{aligned} P(X^2 \leq x) &= P(X \in [-\sqrt{x}; \sqrt{x}]) \\ &= F(\sqrt{x}) - F(-\sqrt{x}) \\ &= F(\sqrt{x}) - (1 - F(\sqrt{x})) \\ &= 2.F(\sqrt{x}) - 1 \end{aligned}$$

où F est la fonction de répartition de la loi normale

Densité

- pour avoir la densité on dérive

$$
\begin{aligned}
2.F(\sqrt{x})' &= 2F'(\sqrt{x}) \times \sqrt{x}' \\
&= 2f(\sqrt{x}) \times \frac{1}{2\sqrt{x}} \\
&= \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \\
&= \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}
\end{aligned}
$$

- ➤ on trouve la densité d'un χ^2 à 1 degré de liberté
- ce sont des cas particulier de loi Γ
- ➤ ils possèdent leur propriété d'additivité
- ★ la somme d'un χ^2 à n_1 ddl et d'un χ^2 à n_2 ddl
- ➤ est un χ^2 à $n_1 + n_2$ ddl

Loi de la statistique de test

- à la place de $\frac{1}{n-1}\sum_{r=1}^{n}(X_r - \bar{X})^2$
- ➤ on peut prendre $\sum_{r=1}^{n}(X_r - \bar{X})^2$ (c'est plus simple)
- ➤ cette quantité est

— une somme de carrés de variables gaussiennes

— centrées

— mais pas réduites

— il y a n gaussiennes

— mais elles ne sont pas indépendantes

★ il faut donc réduire les gaussiennes

★ il faut trouver la dimension des n gaussiennes non indépendantes

Loi de la statistique de test

Loi de $\frac{1}{\sigma^2}\sum_{r=1}^{n}(X_r - \bar{X})^2$

Après réduction et normalisation

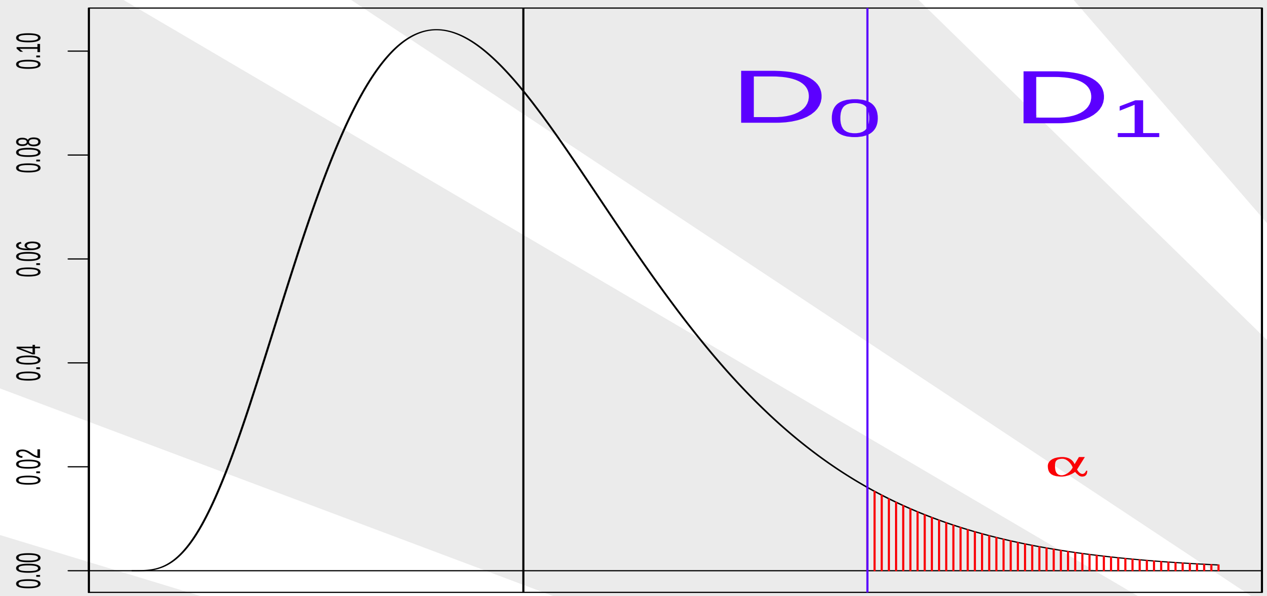
$$\frac{1}{\sigma^2}\sum_{r=1}^{n}(X_r - \bar{X})^2 \rightsquigarrow \chi^2_{n-1}$$

- on en déduit que

$$\sum_{r=1}^{n}(X_r - \bar{X})^2 \rightsquigarrow \sigma^2\chi^2_{n-1}$$

- on dispose donc de la loi de l'estimateur sous H_0 $(\sigma^2 = \sigma_0^2)$
- on peut donc normaliser les régions de décision

Normalisation de la région de rejet



- en noir la loi de la statistique de test sous H_0
- en rouge la normalisation par le risque de première espèce
- en bleu les régions de décisions et leur(s) frontières

Test sur une espérance (variance inconnue)

Test sur une espérance (remarques)

- mêmes conditions que pour les barquettes

★ mais postulat plus réaliste (et opérationnelle) de la variance inconnue

➢ on va s'inspirer de l'estimation de σ^2

- rappel du cas déjà vu

➢ la statistique de test est $\bar{X}$:

$$\bar{X} \rightsquigarrow \mathcal{N}(\mu_0, \frac{\sigma}{\sqrt{n}})$$

sous H_0

- on peut donc utiliser aussi $\bar{X} - \mu_0$ qui est centrée

➢ ou mieux $\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ qui est centrée réduite

Nouvelle statistique de test

- $\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$

★ cela ne change pas le « dessin » car la loi est homothétique

- la nouvelle statistique est telle que

$$\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0,1)$$

- on peut s'en inspirer et remplacer σ inconnu par son estimateur S

$$\frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$$

où

$$S = \sqrt{\frac{1}{n-1}\sum_{r=1}^{n}(X_r-\bar{X})^2}$$

Loi de la statistique de test

- $\frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}} \rightsquigarrow T_{n-1}$
- T a été tabulée par Student

➢ même allure que la loi normale

➢ donc pas plus difficile pour vous

★ en revanche la région D_0 est plus large à α fixé (perte de précision)

Famille des lois de Student

Rappels

Test sur une variance

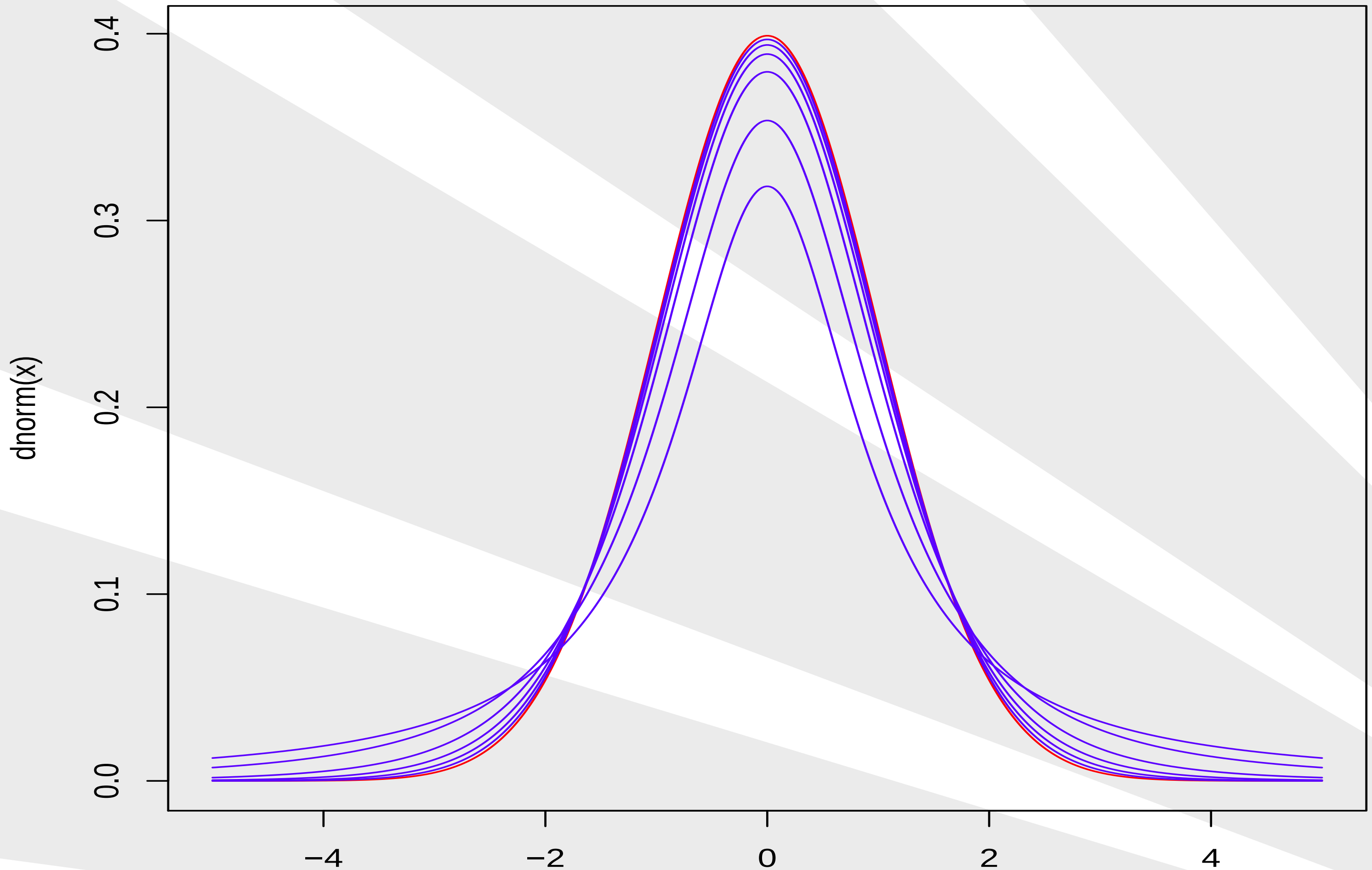
Test sur une espérance (variance inconnue)

❖ Test sur une espérance (remarques)

❖ Nouvelle statistique de test

❖ Loi de la statistique de test

❖ Famille des lois de Student



- Student à 1, 2, 5, 10, 20, 50 degrés de liberté
- ➤ la gaussienne est en rouge