Inférence statistique

Séance 6

Tests statistiques d'hypothèse:
introduction, tests paramétriques

Frédéric Sur

Mines Nancy

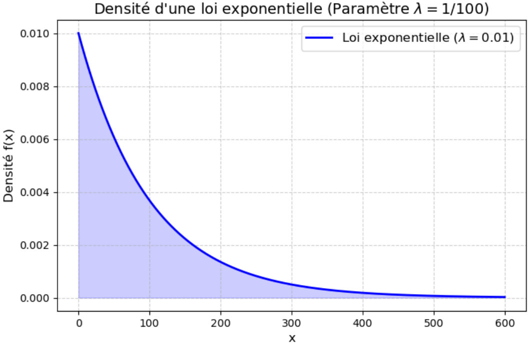
`https://members.loria.fr/FSur/`

Plan

1. Exemple introductif, vocabulaire
   - Hypothèse nulle, hypothèse alternative
   - Risques d'erreur
   - Région de rejet, région d'acceptation
2. Test statistique d'hypothèse
   - Démarche générale
   - *p*-valeur
   - Exemple de mise en œuvre
3. Conclusion

Cas d'étude : *Mean time between failures* (MTBF) (1)

Exemple : temps de défaillance d'une machine industrielle
$\rightarrow$ des tables statistiques (observations sur un grand temps d'un grand nombre de machines identiques) permettent de modéliser la durée de temps entre deux pannes par une *loi exponentielle de paramètre* λ



Rappel : $1/\lambda$ est le temps moyen entre deux pannes (**100 heures**)

Exemples de réalisations : 39,0 - 191,1 - 156,7 - 5.2 - 157.7 - 47.2 ...

Cas d'étude : *Mean time between failures* (MTBF) (2)

Scénario : un prestataire propose une « amélioration » de la machine pour 10k€ d'investissement
$\rightarrow$ au bout de quelques semaines, on observe des pannes après 99, 150, 120, 35, 38, 230 heures.
(moyenne : 112 heures)

Question : l'amélioration est-elle réelle ?

Plus précisément : de telles durées de fonctionnement sont-elles encore compatibles avec une loi exponentielle de moyenne $1/\lambda = 100$ heures ?

Notion d'hypothèse

Hypothèse : énoncé sur une caractéristique d'un problème

Les données expérimentales permettent-elles

- d'accepter l'hypothèse $\mathcal{H}_0$: la nouvelle durée moyenne de fonctionnement reste à 100 heures ?
- d'accepter l'hypothèse $\mathcal{H}_1$: la nouvelle durée moyenne de fonctionnement est supérieure à 100 heures ?

$\rightarrow$ décision prise selon une observation (*réalisation d'un n-échantillon*)

$\rightarrow$ la décision est prise avec une *incertitude*
(on ne peut pas être certain à cause des *fluctuations d'échantillonnage*)

Test d'hypothèse : démarche construisant une règle de décision sur la réalisation d'une statistique basée sur un échantillon, permettant de faire un choix entre deux hypothèses avec risque d'erreur

Hypothèses et modélisation statistique

$\rightarrow$ phénomène modélisé par X une variable aléatoire dont la loi dépend d'un paramètre θ **inconnu**.

Exemple : durée de fonctionnement, θ durée moyenne entre deux pannes

$\rightarrow$ des connaissances sur le problème permettent de faire une hypothèse $\mathcal{H}_0$ sur θ

Exemple : $\theta = \theta_0$, où $\theta_0 = 100$ heures

$\mathcal{H}_0$ est appelée hypothèse nulle (de référence)

$\rightarrow$ on construit une hypothèse alternative $\mathcal{H}_1$ à accepter si $\mathcal{H}_0$ est réfutée.

Exemple : $\mathcal{H}_1 : \theta > \theta_0$

$\rightarrow$ à partir d'un *n*-échantillon de loi X, on calcule une statistique de test S estimant θ

Exemple : S moyenne empirique des temps de fonctionnement sans panne

$\rightarrow$ on calcule la valeur s de S sur les observations dont on dispose

Exemple : $s = \overline{\theta} = 112$ heures

Test et risque

Problème : la différence entre θ_0 (100h) et $\overline{\theta}$ (112h) peut s'expliquer simplement par les fluctuations d'échantillonnage (et pas par l'amélioration de la machine)

$\rightarrow$ L'écart $|\theta_0 - \overline{\theta}|$ est-il trop grand pour être dû aux fluctuations d'échantillonnage (cas 1), ou au contraire n'est pas incompatible avec la loi de X (cas 2) ?

Premier cas : on rejette $\mathcal{H}_0$. On dit que l'écart est *significatif*
Deuxième cas : on ne rejette pas $\mathcal{H}_0$, car cette hypothèse n'est pas incohérente avec les données observées

$\rightarrow$ la décision est prise avec risque de se tromper
une autre réalisation de l'échantilon (une autre série d'observations) aurait peut-être conduit à inverser la décision

$\rightarrow$ on introduit la probabilité α de rejeter $\mathcal{H}_0$ alors qu'elle est vraie.
Valeur typique : $\alpha = 0.05$
on a 5 chances sur 100 d'observer un échantillon selon la loi de X pour lequel l'écart n'aurait pas été assez grand pour rejeter $\mathcal{H}_0$

Vocabulaire

Hypothèse statistique : affirmation sur une caractéristique d'une population

Test d'hypothèse : démarche construisant une règle de décision permettant de faire un choix entre deux hypothèses statistiques

Hypothèse nulle $\mathcal{H}_0$ vs *hypothèse alternative* $\mathcal{H}_1$
$\rightarrow$ c'est $\mathcal{H}_0$ qui est soumise au test (on la considère vraie, et on voit si on peut l'invalider)

Test paramétrique : on teste la valeur d'un paramètre θ
$\rightarrow$ on verra aussi des tests non-paramétriques

Dans les tests paramétriques :

- hypothèse simple : $H : \theta = \theta_0$
- hypothèse composite : $H : \theta \in A$ (A : ensemble)
(généralement $\theta > \theta_0$, $\theta < \theta_0$, $\theta \neq \theta_0$)

Risques de première et deuxième espèce

Le test permet de décider entre deux hypothèses selon la réalisation d'un échantillon
→ il y a un risque d'erreur à choisir $\mathcal{H}$ alors que $\mathcal{H}'$ est vraie : $P(\text{choisir } \mathcal{H}|\mathcal{H}')$

Décision	Vérité $\mathcal{H}_0$	Vérité $\mathcal{H}_1$
$\mathcal{H}_0$	$1-\alpha$	β
$\mathcal{H}_1$	α	$1-\beta$

α : *risque de première espèce*
β : *risque de deuxième espèce*
$1-\beta$: *puissance du test*
α et β sont antagonistes

Principe fondamental : on cherche à maîtriser le risque de première espèce α, car $\mathcal{H}_0$ est critique
→ on veut limiter la probabilité de se tromper en rejetant $\mathcal{H}_0$ alors qu'elle est vraie

Région de rejet, région d'acceptation

s : valeur de la statistique de test S sur les observations

On cherche une région de rejet W telle que :
si $s \in W$, alors on rejette $\mathcal{H}_0$ (on accepte $\mathcal{H}_1$)

ou : si $s \in \overline{W}$, alors on accepte $\mathcal{H}_0$ (on rejette $\mathcal{H}_1$)

W déterminée telle que $P(S \in W | H_0) = \alpha$
(ou « plus grande » région telle que $P(S \in W | H_0) < \alpha$)

En pratique : W choisie « intuitivement » en fonction de H_1
voir TD pour intro théorie de Neyman-Pearson

Remarque : léger abus de langage ici, W est normalement définie dans $\mathbb{R}^n$, voir polycopié

Retour à l'exemple MTBF (1)

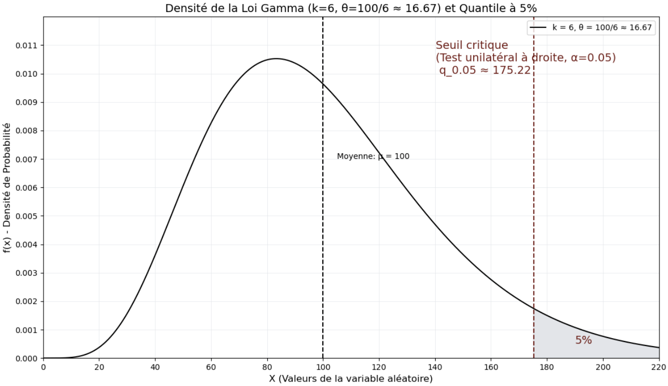
$\mathcal{H}_0 : 1/\lambda = 100$ heures
$\mathcal{H}_1 : 1/\lambda > 100$ heures

Statistique de test : il faut choisir une quantité capturant l'information pertinente, dont on sait établir la loi en fonction de celle de X.
ici : $S = \frac{1}{6}(X_1 + \cdots + X_6) \sim \Gamma(6, 1/(6\lambda))$

Question : quelle région de rejet W ?
$\rightarrow$ vue l'hypothèse $\mathcal{H}_1$, on va chercher W sous la forme $[s_0, +\infty[$, avec s_0 à déterminer. C'est un *test unilatéral à droite*.

Ici, on cherche à répondre à la question : quelle valeur s_0 est telle qu'observer une réalisation de S plus grande que s_0 a une probabilité α sous $\mathcal{H}_0$?

Retour à l'exemple MTBF (2)



s_0 t.q. $P(S > s_0|\mathcal{H}_0) = 0.05$, soit $P(S < s_0|\mathcal{H}_0) = 0.95$
D'où $s_0 = F_\lambda^{-1}(0.95)$ avec F_λ fonction de repartition de $\Gamma(6, 1/(6\lambda))$
On trouve numériquement (ici $\lambda = 1/100$) : $s_0 = 175,2$

Observation : $(99 + 150 + 120 + 35 + 38 + 230)/6 = 112$

Conclusion : on ne peut pas rejeter l'hypothèse $\mathcal{H}_0$

Et la puissance du test ?

Rappel : β est la proba d'accepter $\mathcal{H}_0$ alors que $\mathcal{H}_1$ est vraie.
$\beta = P(S < s_0 | \mathcal{H}_1)$ dépend de λ t.q. $1/\lambda > 100$: on ne peut pas calculer *une* puissance de test
$\rightarrow$ on trace la *courbe de puissance*
$\pi(\lambda) = 1 - \beta(\lambda) = 1 - F_\lambda(175, 2)$
(voir discussion en TD)

Remarque importante : sans connaître β, il n'est pas vraiment raisonnable de conclure qu'on accepte $\mathcal{H}_0$ car on ne maîtrise pas le risque d'erreur (qui peut être grand).
$\rightarrow$ on se contente de dire qu'on rejette $\mathcal{H}_0$ si $s \in W$; et qu'on ne peut pas rejeter $\mathcal{H}_0$ si $s \in \overline{W}$.

Plan

Démarche de test : résumé

1. choix des deux hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$;
2. choix de la statistique de test S ;
3. définition de l'allure de la région critique W en fonction de $\mathcal{H}_1$
   $\rightarrow$ revient à choisir un intervalle I
   (test unliatéral à gauche, à droite, ou test bilatéral
4. calcul de la région critique en fonction de α ;
5. calcul (si possible) de la puissance $1-\beta$ du test ;
6. calcul de la valeur s prise par la statistique de test ;
7. décision : rejet de $\mathcal{H}_0$ si $s \in W$.

Remarque : seules les étapes 6 et 7 nécessitent les observations.

Point de vue de Fisher

Test basé sur une *statistique de test* S et sur W

Différents choix possibles pour S ou pour l'allure de W
$\rightarrow$ ces choix peuvent conduire à des tests de puissance différente pour les mêmes hypothèses $\mathcal{H}_0$ / $\mathcal{H}_1$

La loi sous $\mathcal{H}_0$ doit être connue (pour trouver W), mais la loi sous $\mathcal{H}_1$ (donc la puissance) est souvent inconnue.

D'où le **point de vue de Fisher** sur les tests : on se contente de calculer $P(\text{rejet de } \mathcal{H}_0 | \mathcal{H}_0)$, qui quantifie la proba d'erreur de rejeter $\mathcal{H}_0$ à tort.
C'est la ***p*-valeur**.

p-valeur ou valeur p (p-value)

Exemple d'un test à droite

p-valeur : $P(S > s|\mathcal{H}_0)$

Remarque :

s_0 : seuil de décision au risque α

alors :

$s > s_0 \;\Leftrightarrow\; P(S > s|\mathcal{H}_0) < \alpha$

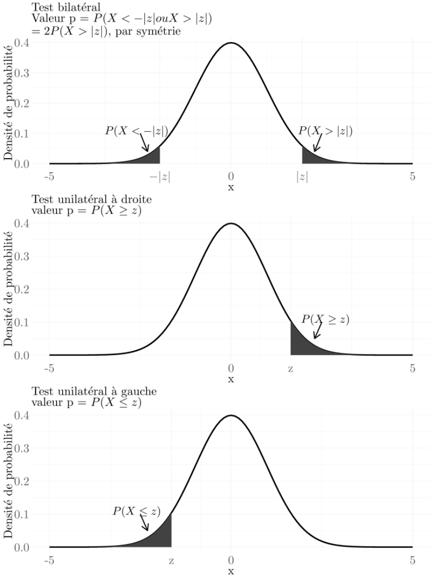
car $P(S > s_0|\mathcal{H}_0) = \alpha$

$\rightarrow$ les logiciels calculent la p-valeur, qui doit être inférieure à α pour rejeter $\mathcal{H}_0$.

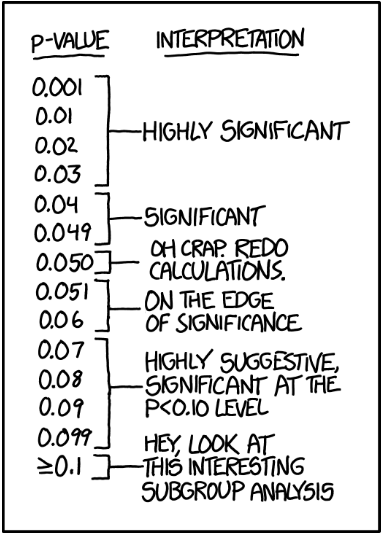
Exemple MTBF

p-valeur $= P(S > 112|\mathcal{H}_0) = 0,157$

on ne peut pas rejeter $\mathcal{H}_0$



Un peu d'humour https://xkcd.com/1478/



If all else fails, use "significant at a $p > 0.05$ *level" and hope no one notices.*

Un dernier exemple

Vous travaillez au sein du service achat d'une entreprise d'électronique. Un fournisseur vous livre des composantes de durées de vie égale à 1000 heures selon le contrat.

Pour vérifier la durée de vie des composantes, vous en prélevez 10 et testez leur durée de vie. Vous trouvez un nombre d'heures de : 950, 1100, 1070, 920, 1010, 880, 910, 1010, 910, 880.

Devez-vous accepter le lot ?

Quelle modélisation statistique ?

Durée de vie : v.a. X de loi normale de moyenne m et d'écart-type σ.
On va faire un test sur la moyenne d'une loi normale de variance inconnue (voir poly)
On choisit le risque de première espèce : $\alpha = 0,05$.

À partir de l'échantillon de taille $n = 10$, on calcule :

- Moyenne de l'échantillon ($\bar{x}$) :

$$\bar{x} = \frac{\sum x_i}{n} = 964 \text{ heures}$$

- Écart-type de l'échantillon ($\bar{s}$) :

$$\bar{s} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \approx 78,9 \text{ heures}$$

Hypothèses et région critique

- H_0 : $\mu = \mu_0 = 1000$ (le lot est conforme)
- H_1 : $\mu < 1000$ (le lot est non conforme)

$\rightarrow$ nous réalisons un test de Student (t-test) unilatéral à gauche pour vérifier si la moyenne du lot est significativement inférieure au seuil contractuel.

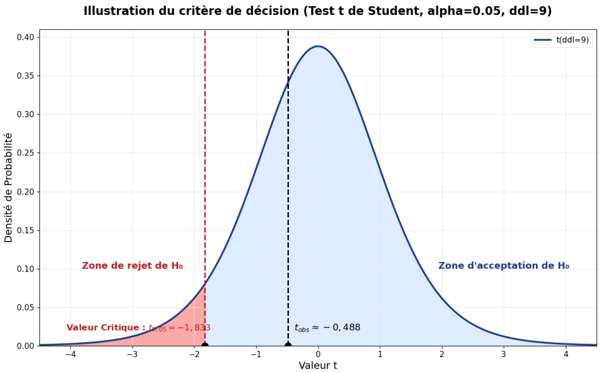
Choix de la forme de la région de rejet de $\mathcal{H}_0$: $]-\infty, \bar{x}_0]$

$\bar{x}_0$ déterminé par $P(\overline{X} < \bar{x}_0 | \mathcal{H}_0) = \alpha$
ou, de manière équivalente : on cherche t_{critique} t.q.

$$P\left(\frac{\overline{X} - \mu_0}{\bar{s}/\sqrt{n}} < t_{\text{critique}}\right) = \alpha$$

où $T_9 = (\overline{X} - \mu_0)(\bar{s}/\sqrt{n})$ (statistique de test) suit la loi de Student à 9 d.d.l.

Statistique de test



Critère de décision :

Pour un seuil de signification $\alpha = 0,05$ et un nombre de degrés de liberté $n - 1 = 9$, la valeur critique dans la table de Student est :

$$t_{\text{critique}} = -1,833$$

Décision

La valeur t_{obs} observée est calculée par la formule :

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{964 - 1000}{78,9/\sqrt{10}} \approx -1,443$$

Conclusion

$t_{\text{obs}} > t_{\text{critique}}$ $(-1,433 > -1,833)$, donc nous **ne pouvons pas rejeter l'hypothèse nulle** H_0.

$\rightarrow$ il n'y a pas assez de preuves pour affirmer que le lot est non conforme, malgré une moyenne inférieure à 1000 heures

p-valeur : $P(T_9 < t_{\text{obs}}) = 0,0915\ldots$

exemple de calculateur en ligne : `https://www.statpowers.com/`

Plan

1. Exemple introductif, vocabulaire
    - Hypothèse nulle, hypothèse alternative
    - Risques d'erreur
    - Région de rejet, région d'acceptation
2. Test statistique d'hypothèse
    - Démarche générale
    - p-valeur
    - Exemple de mise en œuvre
3. Conclusion

Conclusion

Test statistique d'hypothèse : démarche rigoreuse permettant de prendre une décision, à un certain niveau de confiance, à partir d'observations, sous une modélisation statistique.

Aujourd'hui en TD : tests d'hypothèse paramétriques

Deux séances suivantes : quelques problèmes statistiques et les tests d'hypothèse adaptés

Message de service 1 : notes de cours + sujets TD dans les casiers, pensez à les apporter à chaque séance.

Message de service 2 : séance de soutien 17h-18h et 18h-19h mercredi 6 mai.