Statistique Mathématique

Salim Lardjane

Université Bretagne Sud

Partie VI

Tests d'Hypothèses Statistiques

Tests d'hypothèses statistiques

Dans le cadre d'études statistiques, on souhaite souvent tester des hypothèses sur la valeur des paramètres inconnus d'un modèle probabiliste.

On désigne une hypothèse par la lettre $\mathcal{H}$. L'objectif est de vérifier si l'hypothèse $\mathcal{H}$ n'est pas *contredite* par les données empiriques.

Ces dernières prennent la forme d'un échantillon i.i.d. associé à une variable aléatoire X. On le note $X_1, \ldots, X_n$.

La confrontation de l'hypothèse émise avec les données empiriques est faite à l'aide d'un *test d'hypothèse statistique*.

Tests d'hypothèses statistiques

Le résultat de cette confrontation peut être :

— *négatif* : les données empiriques *contredisent* l'hypothèse avancée et il faut alors y renoncer

— *positif* : les données empiriques *confirment* l'hypothèse avancée et celle-ci peut être retenue comme base d'autres développements

Notons que si le résultat est positif, cela ne signifie pas que notre hypothèse est la meilleure et la seule possible ; tout ce qu'on peut affirmer, c'est qu'elle n'est pas contredite par le données empiriques. Il peut très bien exister d'autres hypothèses ayant cette propriété.

Tests d'hypothèses statistiques

Une hypothèse statistique, même statistiquement confirmée, n'est pas un fait absolu acquis une fois pour toute, mais seulement une proposition vraisemblable *ne contredisant pas les données de l'expérience*.

Examinons à présent quelques *tests d'hypothèses* utilisés en pratique.

Tests d'hypothèses statistiques

Tout d'abord, quand on analyse un échantillon i.i.d. $X_1, \ldots, X_n$ associé à une v.a. X, il est important de sélectionner et de justifier un *modèle* probabiliste pour X.

Celui-ci prend usuellement la forme d'une *fonction de répartition* $F_{\text{mod}}(x)$ ou d'une densité de probabilité $f_{\text{mod}}(x)$.

Tests d'hypothèses statistiques

On doit donc s'intéresser à des hypothèses du type :

$$\mathcal{H} : F_X \equiv F_{\text{mod}}$$

où la fonction modèle hypothétique peut être définie :

1. *de façon unique* , auquel cas $F_{\text{mod}} = F_0$, où F_0 est une fonction entièrement connue
2. comme une fonction *appartenant à une famille paramétrique*, auquel cas $F_{\text{mod}} = F_\theta$, où θ est un paramètre inconnu mais susceptible d'être estimé à partir des données..

La vérification d'hypothèses du type précédent est faite à l'aide de *tests d'ajustement*, dont les plus connus sont le *test du Khi-deux* et le *test de Kolmogorov-Smirnov*.

Tests d’hypothèses statistiques

Supposons à présent que les observations $X_1, \ldots, X_n$ sont les valeurs d’une caractéristique d’une unité statistique, obtenues par des mesures effectuées sur n produits tirés au hasard dans la production d’une machine par exemple.

Notons a_0 la *valeur nominale* (la *spécification*) de cette caractéristique.

Chaque observation X_i peut naturellement différer de la valeur nominale a_0.

Pour s’assurer que la machine est correctement réglée, il faut s’assurer que la *valeur moyenne* de la caractéristique sur la production entière correspond à la valeur nominale.

Tests d'hypothèses statistiques

Autrement dit, on doit tester une hypothèse du type :

$$\mathcal{H} : \mathbb{E}(X) = a_0$$

Dans le cas général, les hypothèse de cette nature sont de la forme :

$$\mathcal{H} : \theta \in \Theta_0$$

où Θ_0 est un domaine de valeurs hypothétiques, qui peut très bien être réduit à un point.

Principe général d’un test statistique

Les test statistiques diffèrent beaucoup quant à leur *finalité* mais ils sont tous construits selon le même principe logique, que l’on va décrire dans la suite.

Principe général d'un test statistique

Etape 1. On avance une hypothèse $\mathcal{H}_0$

Etape 2. On définit le *seuil de signification* ou *niveau* α du test.

Toute décision statistique, c'est-à-dire toute décision prise sur la base d'un nombre fini d'observations aléatoires, est prise avec un certain *risque* d'erreur dans un sens comme dans l'autre.

Avec une probabilité α, l'hypothèse $\mathcal{H}_0$ *peut être rejetée alors qu'elle est vraie* (risque de *première espèce*) et, inversement, avec une probabilité β, l'hypothèse $\mathcal{H}_0$ *peut être acceptée alors qu'elle est fausse* (risque de *deuxième espèce*).

Principe général d'un test statistique

A taille d'échantillon n fixée, on peut choisir à notre convenance la probabilité de l'un seulement de ces risques.

Si l'on peut faire croître n autant que l'on veut, on peut rendre arbitrairement petites les probabilités α et β, pour toute *hypothèse alternative* $\mathcal{H}_1$, retenue lorsque $\mathcal{H}_0$ est rejetée.

Si la taille de l'échantillon est fixée, on se donne en général la probabilité α de rejet à tort de l'hypothèse $\mathcal{H}_0$. Cette dernière est souvent appelée dans ce contexte *hypothèse nulle*.

La probabilité α est appelée *seuil de signification* ou *niveau* du test.

Principe général d'un test statistique

La valeur du seuil de signification α la *plus répandue en pratique* est $\alpha = 0.05$.

Elle exprime que, si $\mathcal{H}_0$ est vraie, seulement 5% des échantillons d'observations susceptibles d'être obtenus conduisent à rejeter $\mathcal{H}_0$.

Principe général d'un test statistique

Etape 3. On se donne une fonction des observations

$$\eta_n = \eta(X_1, \ldots, X_n)$$

appelée *statistique critique*.

La statistique critique η_n est une variable aléatoire qui, si $\mathcal{H}_0$ est vraie, suit une loi de probabilité bien déterminée, de densité $f_n(x)$.

La statistique critique sert à mesurer l'écart entre les caractéristiques empiriques de l'échantillon et leurs valeurs théoriques sous $\mathcal{H}_0$.

Principe général d'un test statistique

Etape 4. Dans les tables de la loi de densité $f_n(x)$ (ou à l'aide d'un logiciel statistique ou d'un tableur), on détermine le quantile

$$\eta_{\alpha/2}^{\text{max}}$$

de niveau $1 - \alpha/2$ et le quantile

$$\eta_{\alpha/2}^{\text{min}}$$

de niveau $\alpha/2$.

Principe général d'un test statistique

Ces quantiles subdivisent le domaine des valeurs possibles de la variable aléatoire η_n en trois régions :

1. La région des valeurs invraisemblablement petites
2. La région des valeurs invraisemblablement grandes
3. La région des valeurs vraisemblables sous $\mathcal{H}_0$.

On est là dans le cadre d'un *test bilatère*, c'est-à-dire qu'on envisage que les valeurs de la statistique critique puissent être trop grandes ou trops petites.

Principe général d'un test statistique

Si on envisage seulement que les valeurs de la statistique critique puissent être trop petites ou seulement qu'elles puissent être trop grandes, on ne cherche *qu'un seul quantile*.

Dans le premier cas, on détermine le quantile η_α^{min} de niveau α qui partage le domaine des valeurs possibles de η_n en deux régions :

1. une région de valeurs invraisemblablement petites,
2. une région de valeurs vraisemblables.

Principe général d'un test statistique

Dans le deuxième cas, on détermine le quantile $\eta_\alpha^{\max}$,de niveau $1-\alpha$, qui partage le domaine des valeurs possibles de η_n en deux régions :

1. une région de valeurs invraisemblablement grandes,
2. une région de valeurs vraisemblables.

Dans les deux cas, on dit qu'on effectue un *test unilatère*.

Principe général d'un test statistique

Etape 5. Enfin, on injecte les données empiriques $x_1, \ldots, x_n$ dans la fonction η_n et on calcule la *valeur prise* par η_n sur l'échantillon observé : $\eta_n(x_1, \ldots, x_n)$.

Si cette valeur appartient à la région de valeurs vraisemblables, les données empiriques ne contredisent pas l'hypothèse $\mathcal{H}_0$.

Dans le cas contraire, c'est-à-dire si $\eta_n(x_1, \ldots, x_n)$ est trop petite ou trop grande, on conclut que η_n ne suit pas la loi de densité $f_n(x)$ postulée sous $\mathcal{H}_0$, ce qui implique que l'hypothèse $\mathcal{H}_0$ est fausse et qu'il faut y renoncer.

Principe général d'un test statistique

La décision suggérée par un test statistique peut être fausse aussi bien dans le cas où $\mathcal{H}_0$ est rejetée à tort (ce qui arrive avec une probabilité α) que dans le cas où elle est acceptée à tort (ce qui arrive avec une probabilité β).

Les probabilités d'erreur α et β sont appelées respectivement *risque de première espèce* et *risque de deuxième espèce*.

La quantité $1 - \beta$ est appelée *puissance du test*. C'est la probabilité que $\mathcal{H}_0$ soit rejetée avec raison.

Entre deux tests caractérisés par un même risque de première espèce α, on préfère celui dont la puissance est la plus grande.

Principe général d'un test statistique

Si $\mathcal{H}_0$ consiste à conjecturer que la valeur d'un paramètre θ est très exactement égale à une valeur donnée θ_0, on dit que $\mathcal{H}_0$ est une *hypothèse simple*

Dans tous les autres cas, elle est dite *multiple*.

Exemples

Exemple 1. Soit $X_1, \ldots, X_n$ un échantillon i.i.d. associé à une v.a.r. X de loi $\mathcal{N}(\theta, \sigma^2)$ où σ est supposé connue.

On veut tester l'hypothèse nulle :

$$\mathcal{H}_0 : \theta \leq \theta_0$$

contre l'hypothèse alternative :

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

Exemples

Sous $\mathcal{H}_1$, $\bar{X}$ aura tendance à être plus grand que sous $\mathcal{H}_0$.

On décide donc d'effectuer un *test unilatère* et de rejeter $\mathcal{H}_0$ si $\bar{X} > \theta_0 + C$.

Plus précisément, on adopte pour *statistique critique* :

$$\eta_n = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim_{\mathcal{H}_0} \mathcal{N}(0,1).$$

La suite des opérations s'ensuit aisément.

Exemples

Exemple 2. Considérons le *test d'ajustement* d'hypothèse nulle :

$$\mathcal{H}_0 : X \sim \mathcal{U}_{[0,1]}$$

c'est-à-dire $f_X(x) = \mathbb{I}_{[0,1]}(x)$, contre l'hypothèse alternative :

$$\mathcal{H}_1 : f_X(x) = 3x^2\mathbb{I}_{[0,1]}(x)$$

Exemples

On suppose qu'on dispose *d'une seule observation* X_1 ($n = 1$) et qu'on adopte pour statistique critique :

$$\eta_1(X_1) = X_1.$$

On décide d'effectuer un test unilatère.

Plus précisément, on décide de rejeter $\mathcal{H}_0$ si $X_1 > c$, où $0 < c < 1$.

Exemples

Alors, le risque de première espèce est donné par :

$$\alpha = \mathbb{P}_{\mathcal{H}_0}(X_1 > c) = 1 - c$$

et le risque de deuxième espèce par :

$$\beta = \mathbb{P}_{\mathcal{H}_1}(X_1 \leq c) = \int_0^c 3x^2 dx = c^3.$$

On voit donc que si on fixe le niveau du test α à 0.05, on a nécessairement $\beta = (0.95)^3 \approx 0.86$.

Ainsi, le risque de deuxième espèce est élevé (la puissance est faible) mais on pouvait s'y attendre car on n'a qu'une seule observation.

P-value

Remarque. Pour de nombreux tests, il est possible de définir une quantité p, *fonction de l'échantillon observé* et appelée *p-value*, qui s'utilise de la façon suivante :

1. Si $p < \alpha$, on rejette $\mathcal{H}_0$ au niveau α : les données empiriques contredisent l'hypothèse nulle.
2. Si $p > \alpha$, on ne rejette pas $\mathcal{H}_0$ au niveau α : les données empiriques ne contredisent pas l'hypothèse nulle.
3. Si $p = \alpha$, cela dépend du test effectué.

P-value

Exemple. Reprenons l'exemple 2 précédent.

On souhaite tester l'hypothèse nulle

$$\mathcal{H}_0 : f_X(x) = \mathbb{I}_{[0,1]}(x)$$

contre l'hypothèse alternative :

$$\mathcal{H}_1 : f_X(x) = 3x^2\mathbb{I}_{[0,1]}(x)$$

en se basant sur une seule observation X_1.

P-value

On retient comme statistique critique :

$$\eta_1(X_1) = X_1$$

et on effectue un test unilatère.

Plus précisément, on rejette $\mathcal{H}_0$ si $X_1 > c$ $(0 < c < 1)$ et on ne rejette pas $\mathcal{H}_0$ sinon.

Le risque de première espèce (niveau du test) est donné par :

$$\alpha = \mathbb{P}_{\mathcal{H}_0}(X_1 > c) = 1 - c$$

d'où $c = 1 + \alpha$.

P-value

Notons x_1 la réalisation observée de X_1.

On définit la p-value par :

$$p = \mathbb{P}_{\mathcal{H}_0}(X_1 > x_1) = 1 - x_1$$

Si $p = 1 - x_1 < \alpha$, c'est-à-dire si $x_1 > 1 + \alpha = c$, on rejette $\mathcal{H}_0$ au niveau α.

Si $p = 1 - x_1 > \alpha$, c'est-à-dire si $x_1 < 1 + \alpha = c$, on ne rejette pas $\mathcal{H}_0$ au niveau α.

Si $p = 1 - x_1 = \alpha$, c'est-à-dire si $x_1 = 1 + \alpha$, on ne rejette pas $\mathcal{H}_0$ au niveau α vu la définition du test.

Tests obtenus à partir d'IC

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. associé à une v.a.r. X de loi $\mathcal{N}(\mu_0, \sigma^2)$.

Nous avons obtenu précédemment un intervalle de confiance pour μ_0 lorsque σ^2 est inconnue, en écrivant :

$$\mathbb{P}\left(-b \leq \frac{\bar{X} - \mu_0}{S/\sqrt{n-1}} \leq b\right) = 2F_T(b) - 1$$

où

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n-1}} \sim St_{n-1}.$$

Tests obtenus à partir d'IC

Supposons que $2F_T(b) - 1 = 0.95$, ce qui implique :

$$\mathbb{P}(|T| > b) = 0.05.$$

Supposons. à présent que $X \sim \mathcal{N}(\mu, \sigma^2)$.

Si $\mu = \mu_0$, on est dans la situation précédente et $\{|T| > b\}$ est un événement de probabilité faible.

Tests obtenus à partir d'IC

Il est donc naturel de tester l'hypothèse :

$$\mathcal{H}_0 : \mu = \mu_0$$

contre

$$\mathcal{H}_1 : \mu \neq \mu_0$$

en rejetant $\mathcal{H}_0$ si

$$|T| > b,$$

c'est-à-dire si μ_0 *n'appartient pas à l'intervalle de confiance de niveau* $1 - \alpha$ *pour* μ.

Tests obtenus à partir d'IC

La fonction puissance

$$K(\mu) = \mathbb{P}_\mu(|T| > b)$$

est difficile à calculer si $\mu \neq \mu_0$, car $(\bar{X} - \mu_0)/(\sigma/\sqrt{n})$ n'est plus d'espérance nulle.

On a par conséquent recours à une *loi de Student supérieurement non centrée*.

Tests obtenus à partir d'IC

Remarque (test unilatère). Si on souhaite tester :

$$\mathcal{H}_0 : \mu = \mu_0$$

contre

$$\mathcal{H}_1 : \mu > \mu_0$$

on rejette $\mathcal{H}_0$ si

$$T > b.$$

Test du Rapport de Vraisemblance

Dans le cas du test d'un hypothèse simple contre une hypothèse simple, si on impose que la probabilité d'erreur de type 1 soit au plus égale à α et qu'on cherche à minimiser la probabilité d'erreur de type 2, le test *optimal* s'avère être un *test du rapport de vraisemblance*.

Test du Rapport de Vraisemblance

Soit λ une constante et soit

$$L(x) = \frac{f_1(x)}{f_0(x)}$$

le *rapport de vraisemblance*, où $f_0(x)$ est la vraisemblance des données sous l'hypothèse nulle et $f_1(x)$ leur vraisemblance sous l'hypothèse alternative.

Un *test du rapport de vraisemblance* a la forme suivante :

$$\begin{array}{rcl} \text{Si } L(x) > \lambda & : & \text{rejeter } \mathcal{H}_0 \\ \text{Si } L(x) < \lambda & : & \text{accepter } \mathcal{H}_0 \\ \text{Si } L(x) = \lambda & : & \text{prendre une décision arbitraire} \end{array}$$

Test du Rapport de Vraisemblance

Intuitivement, l'idée est que si ce qu'on observe est significativement plus vraisemblable sous $\mathcal{H}_1$, on rejette $\mathcal{H}_0$.

Le résultat d'optimalité est l'objet du *Théorème de Neyman-Pearson* qui sera vu dans la suite.

Si $\mathcal{H}_0$ ou $\mathcal{H}_1$ est une hypothèse composée, il n'y a plus de résultat d'optimalité.

Pour de nombreux modèles usuels (normal, Poisson, binomial, exponentiel), $L(x) = L(x_1, \ldots, x_n)$ est une fonction de la somme des observations ou, de façon équivalente, de la moyenne arithmétique.

Cela justifie la construction de tests basés sur $\sum_{i=1}^{n} X_i$ ou $\bar{X}$.

Théorème de Neyman-Pearson

On teste l'hypothèse simple que la v.a.r. X est de densité f_0 contre l'hypothèse simple que X est de densité f_1.

Soit φ_λ un test du rapport de vraisemblance de paramètre λ (une constante positive).

En d'autre termes, $\varphi_\lambda(x)$ est la probabilité de rejeter $\mathcal{H}_0$ lorsque x est observée et, plus précisément :

$$\varphi_\lambda(x) = \begin{cases} 1 & \text{si} \quad L(x) > \lambda \\ 0 & \text{si} \quad L(x) < \lambda \\ \text{valeur quelconque} & \text{si} \quad L(x) = \lambda \end{cases}$$

Théorème de Neyman-Pearson

Notons α_λ la probabilité d'erreur de type 1 (rejet à tort de $\mathcal{H}_0$) pour le test ϕ_λ et β_λ la probabilité d'erreur de type 2 (acceptation à tort de $\mathcal{H}_0$).

Théorème de Neyman-Pearson. *Avec les notations précédentes, soit φ un test statistique arbitraire de probabilités d'erreur α et β. Si $\alpha \leq \alpha_\lambda$, alors $\beta \geq \beta_\lambda$.*

Autrement dit, le test du rapport de vraisemblance est celui de puissance maximale à niveau fixé.

Ce résultat sera admis. Pour des éléments de preuve, voir Tassi 2004. Pour une preuve détaillée, voir Monfort 1997.