LE TEST STATISTIQUE PAR L'EXEMPLE [1]

PRÉPARATION À L'AGRÉGATION EXTERNE DE MATHÉMATIQUES - ENS RENNES

COMPLÉMENTS DE COURS - ANNÉE 2016/2017

1. LE PROBLÈME

Un industriel, responsable d'une machine qui produit des pièces classées soit « apte au service », codé par 0, soit « défectueuses », codé par 1. Le nombre de pièces fabriquées étant gigantesque et l'examen de chaque pièce étant relativement coûteux, il ne peut évaluer la qualité de sa production que sur un lot de taille n faible au regard de la production, par exemple $n = 100$. Soit $x = (x_1, \cdots, x_n)$ la suite de 0 et de 1 observée, et $\bar{x}_n = 0.22$ la moyenne observée.

Pour l'industriel, l'enjeu est de déterminer si la proportion $\theta_0 \in]0,1[$ de pièces défectueuses est plus grande que 0.2, valeur calculée en fonction des contraintes économiques de l'entreprise. Si c'est le cas, il décidera d'arrêter la production pour réparer ou changer la machine. Cependant, une telle opération est très coûteuse, et il ne veut en arriver à une telle extrémité qu'en cas de nécessité absolue.

2. PRÉLIMINAIRE

Avant de construire le modèle statistique, il y a lieu de se demander où se situe l'intervention du hasard. De manière stéréotypée, on la retrouve dans

1. la *variablilité intrinsèque du phénomène* : de petites variations (position...) influent sur la qualité ;
2. l'*échantillonage* : seule la qualité de l'échantillon tiré au hasard est examinée.

On considère que $x_1, \cdots, x_n$ sont des réalisations de v.a.i.i.d. Au regard des 2 manifestations du hasard, ceci est justifié par

1. l'observation de la machine : la production est stable (donc v.a. de même loi), et les petites variations ne se répercutent pas d'une pièce à l'autre (donc v.a. indépendantes) ;
2. l'échantillonage, qui a été effectué par un tirage uniforme avec remise dans la population de toutes les pièces produites.

Le modèle statistique considéré est $(\{0,1\}^n, \{P_\theta\}_{\theta \in]0,1[})$ avec $P_\theta = \mathscr{B}(\theta)^{\otimes n}$, sur l'espace des observations $\{0,1\}^n$, de sorte qu'un *échantillon* $(X_1, \cdots, X_n)$ de la loi P_θ est une suite i.i.d. de v.a. de loi $\mathscr{B}(\theta)$. Ce modèle est *paramétrique* et *identifiable*.

Il est important de comprendre la démarche : la seule information est apportée par l'observation $(x_1, \cdots, x_n)$. En étudiant les propriétés générales du modèle, i.e. les propriétés d'un échantillon de chaque loi P_θ, on est en mesure de décider quelle est la valeur de θ qui correspond le mieux à l'observation, autrement dit, on est en mesure de construire une fonction $\theta(x_1, \cdots, x_n)$ proche de θ_0. *La démarche peut être résumée ainsi : on construit le modèle statistique en examinant le phénomène, puis on ajuste la loi du modèle à l'aide des observations.*

Calculons tout d'abord un intervalle de confiance pour θ_0. Comme $\bar{X}_n(1-\bar{X}_n) \to \theta(1-\theta)$ en P_θ-probabilité, on déduit du TCL et du lemme de Slutsky la convergence en loi sous P_θ suivante :

$$S_n := \sqrt{n}\frac{\bar{X}_n - \theta}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{\mathscr{L}/P_\theta} \mathscr{N}(0,1).$$

Si $g_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi $\mathscr{N}(0,1)$,

$$P_\theta\left(\theta \in \left[\bar{X}_n - g_{1-\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + g_{1-\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right]\right) = P_\theta\left(S_n \in \left[-g_{1-\alpha/2}, g_{1-\alpha/2}\right]\right),$$

converge vers $1-\alpha$ lorsque $n \to \infty$. L'intervalle de confiance asymptotique pour θ_0 au niveau $1-\alpha$ est donc :

$$I = \left[\bar{x}_n - g_{1-\alpha/2}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}, \bar{x}_n + g_{1-\alpha/2}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}\right].$$

On calcule alors, pour $n = 100$, $\bar{x}_{100} = 0.22$, ce qui donne avec $\alpha = 5\%$: $I = [0.14, 0.30]$. Il n'est donc pas possible de décider par ce biais si θ_0 est plus grand ou plus petit que la valeur seuil 0.2 : l'intervalle de confiance n'a donné à l'industriel aucune information susceptible d'être exploitée.

1. Benoît Cadre - ENS Rennes

2. Une procédure de décision : le test statistique

2.1 Principes généraux

Pour l'industriel, la vraie question n'est pas tant de connaître une valeur approchée pour θ_0, mais plutôt de savoir si il faut poursuivre ou arrêter la production, i.e. si $\theta_0 \leq 0.2$ ou $\theta_0 > 0.2$. Il faut donc construire une stratégie de décision afin de répondre à cette question.

Définition [Problème de test] *Soit $(\mathscr{H}^n, \{P_\theta\}_{\theta\in\Theta})$ un modèle statistique et Θ_0, Θ_1 des sous-ensembles disjoints de Θ. Un problème de test est la confrontation d'une hypothèse nulle $H_0 : \theta \in \Theta_0$ contre une hypothèse alternative $H_1 : \theta \in \Theta_1$.*

Il n'est pas souhaitable de faire jouer le même rôle aux 2 hypothèses H_0 et H_1, et ceci pour deux raisons :

- ▷ il faut en effet tenir compte de l'intérêt de l'acteur qui pose la question. C'est ce que nous faisons dans la vie courante :
  - Un procès aux assises est un problème de test avec H_0 : « le prévenu est innocent » présumée vraie. Ce principe est nécessaire, sinon les prisons seraient remplies d'innocents. Néanmoins, il déséquilibre le procès.
  - Le *principe de précaution* en est une autre illustration. Par exemple, si on est amené à se poser la question « une bombe va-t-elle tomber ? », n'importe quelle personne censée privilégiera l'affirmative car les conséquences seront moins graves.
- ▷ il n'est en général pas possible de placer au même niveau simultanément les 2 erreurs qui peuvent être commises, i.e. H_0 est rejetée alors qu'elle est vraie et H_0 est acceptée alors qu'elle est fausse. Dans ce cas, on prend le risque de se retrouver avec ces deux erreurs simultanément grandes.

Par convention, H_0 est l'hypothèse privilégiée, i.e. l'hypothèse qui tient compte de l'intérêt de l'acteur qui pose la question.

Exemple Dans notre problème :

- ▷ il est très coûteux pour l'industriel d'arrêter sa production, donc il ne veut en arriver à cette extrémité qu'en cas de nécessité absolue. De ce fait, l'hypothèse $\theta \leq 0.2$ est celle qui lui est la moins préjudiciable. Donc, de son point de vue, le problème de test est $H_0 : \theta \leq 0.2$ contre $H_1 : \theta > 0.2$.
- ▷ un organisme de contrôle de la qualité, à l'inverse, privilégie $\theta > 0.2$. Son problème de test serait donc $H_0 : \theta > 0.2$ contre $H_1 : \theta \leq 0.2$.

Définition [Test statistique] *Soit $(\mathscr{H}^n, \{P_\theta\}_{\theta\in\Theta})$ un modèle statistique et le problème de test de H_0 contre H_1. Une zone de rejet (ou région critique) R est un sous-ensemble de l'espace des observations $\mathscr{H}^n$. Pour l'observation $x \in \mathscr{H}^n$, la stratégie de décision est la suivante :*

- ▷ *Si $x \notin R$, H_0 n'est pas rejetée ;*
- ▷ *Si $x \in R$, H_0 est rejetée.*

Le test statistique de H_0 contre H_1 de zone de rejet R est la stratégie de décision ci-dessus.

On ne parle donc dans cette définition que de « rejet » ou « non rejet » de H_0. Ceci illustre à nouveau la dissymétrie des 2 hypothèses et le point de vue très négatif adopté dans la construction d'un test, qui n'est au fond mis en oeuvre que pour rejeter H_0 ! (de même qu'un procès aux assises n'est organisé que pour éprouver la présomption d'innocence). Ainsi,

- ▷ un test confronte une hypothèse à la réalité d'une observation : si, au regard de celle-ci, il y a désaccord grave, H_0 est rejetée ; sans quoi, et peut-être faute de mieux, H_0 ne l'est pas.
- ▷ la réponse donnée à l'issue d'un test est « peut-être » (H_0 non rejetée), ou « non » (H_0 rejetée).

Cela dit, dans un soucis de simplicité, on se permettra dans la suite de dire que H_0 est acceptée, plutôt que non rejetée.

Terminologie

1. H_1 est *fantôme* si $\Theta_1 = \Theta_0^c$;
2. une hypothèse est *simple* si l'ensemble associé est un singleton. Sinon, elle est *composite* ;
3. si $\Theta \subset \mathbb{R}$, un test dont l'hypothèse alternative est fantôme est *unilatère* (resp. *bilatère*) lorsque l'une des deux hypothèses s'écrit $\theta \leq \theta_0$ ou $\theta < \theta_0$ (resp. $\theta = \theta_0$).

De quelle manière lier la zone de rejet et l'hypothèse nulle ? Il convient d'introduire un premier type d'erreur.

Définition [Erreur de 1ère espèce] *Soit $(\mathscr{H}^n, \{P_\theta\}_{\theta\in\Theta})$ un modèle statistique et R une zone de rejet pour H_0 vs H_1.*

1. *L'erreur de 1ère espèce est la fonction définie sur Θ_0 par $\theta \mapsto P_\theta(R)$: elle consiste à rejeter H_0 à tort.*
2. *Le test est de niveau α si l'erreur de 1ère espèce maximale est égale à α, i.e.*

$$\sup_{\theta\in\Theta_0} P_\theta(R) = \alpha.$$

Puisque H_0 est l'hypothèse privilégiée, rejeter H_0 à tort est plus préjudiciable qu'accepter H_0 à tort (de même, dans un procès, accuser le prévenu à tort est plus préjudiciable que de conforter à tort sa présomption d'innocence, au moins pour le prévenu !). De ce fait, on adopte le principe de Neyman, qui conduit à à minimiser en priorité le niveau du test.

On choisit souvent le niveau $\alpha = 5\%$. La notion de niveau d'un test n'est manifestement pas suffisante : il est possible de construire la zone de rejet de manière à ce que H_0 soit toujours acceptée (prendre une zone de rejet vide) et, dans ce cas, le niveau du test est nul. Cependant, on accepte toujours H_0, même si elle n'est pas vraie ! Il convient donc de définir un autre type d'erreur, afin d'éviter ces tests qui n'enseignent rien.

Définition [ERREUR DE 2ÈME ESPÈCE] *Soit* $(\mathscr{H}^n, \{P_\theta\}_{\theta\in\Theta})$ *un modèle statistique et* R *une zone de rejet pour* H_0 *vs* H_1.

1. *L'erreur de 2ème espèce est la fonction définie sur* Θ_1 *par* $\theta \mapsto P_\theta(R^c)$ *: elle consiste à accepter* H_0 *à tort.*
2. *La puissance du test est la fonction définie sur* Θ_1 *par* $\theta \mapsto P_\theta(R)$.

Ayant retenu le principe de Neyman, l'objectif est de rechercher ensuite, parmi les tests de niveau α, ceux dont la puissance est maximale. La mécanique est simple, mais en pratique, la construction du test est délicate :
- ▷ **choix du niveau** : plus il est faible, plus la zone de rejet à tendance à être petite, ce qui diminue la puissance du test.
- ▷ **choix des hypothèses** : plus Θ_1 est vaste, plus la puissance diminue.

La puissance d'un test est en général inaccessible, sauf dans quelques cas simples. Si on ne dispose pas de théorème limite précis, on peut utiliser les inégalités classiques, telles que Bienaymé-Tchebytchev, Hoeffding, Berry-Esséèn,...

2.2 TEST POUR UNE TAILLE D'ÉCHANTILLON FINIE

Dorénavant, $\theta_s = 0.2$. Les principes généraux étant établis, on se penche maintenant sur le problème de la construction de la zone de rejet, pour le test de $H_0 : \theta \leq \theta_s$ contre $H_1 : \theta > \theta_s$. On fixe un niveau α pour ce test. Soit $(X_1, \cdots, X_n) \sim P_\theta$. D'après la LGN, $\bar{X}_n$ prend des valeurs proches de θ : il est alors naturel de rejeter H_0 lorsque la moyenne empirique prend des valeurs anormalement grandes. La zone de rejet est donc de la forme :

$$R^1 := \{y \in \{0,1\}^n : \bar{y}_n > a\},$$

où a est une constante à déterminer en fonction de α et n. Pour une telle zone de rejet, le niveau du test est

$$\sup_{\theta\in\Theta_0} P_\theta(R^1) = \sup_{\theta\leq\theta_s} P_\theta(\bar{X}_n > a).$$

La fonction $\theta \mapsto P_\theta(\bar{X}_n > a)$ est croissante d'où, en désignant par B une v.a. de loi $\mathscr{B}(n, \theta_s)$:

$$\sup_{\theta\in\Theta_0} P_\theta(R^1) = P_{\theta_s}(\bar{X}_n > a) = \mathbb{P}(B > na) = 1 - \mathbb{P}(B \leq na).$$

La loi de la statistique de test étant discrète, on s'oriente vers un test de niveau inférieur à α. Soit $a \geq q_{1-\alpha}/n$, avec $q_{1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi $\mathscr{B}(n, \theta_s)$. Parmis tous les tests construits pour $a \geq q_{1-\alpha}/n$, on cherche maintenant celui dont la puissance est maximale. La puissance est ici donnée par la fonction définie sur $]\theta_s, 1[$ par $\theta \mapsto P_\theta(\bar{X}_n > a)$. Elle est maximale pour la plus petite valeur possible de a, soit $a = q_{1-\alpha}/n$. Par conséquent, la zone de rejet est :

$$R^1 = \left\{y \in \{0,1\}^n : \bar{y}_n > \frac{q_{1-\alpha}}{n}\right\}.$$

Avec les valeurs $\theta_s = 0.2$, $n = 100$ et un seuil $\alpha = 5\%$, on trouve $q_{1-\alpha}/n = 0.23$ d'où $R^1 = \{y \in \{0,1\}^n : \bar{y}_n > 0.23\}$. Comme $(x_1, \cdots, x_n) \notin R^1$, on est amené à accepter H_0 au niveau inférieur à 5%.

Une manière équivalente de décider est d'utiliser la notion de *p-valeur* :

la p-valeur traduit la probabilité, sous H_0, d'observer pire que les observations.

Dans ce cas, il s'agit de la quantité

$$p\text{-valeur} := \sup_{\theta\in\Theta_0} P_\theta(\bar{X}_n > \bar{x}_n).$$

Si la p-valeur est faible, ce que l'on observe est rare sous H_0, ce qui remet en cause H_0. En particulier, H_0 est rejetée au niveau inférieur à α dès que la p-valeur est plus petite que α. Mais si la p-valeur n'est pas faible, cela signifie qu'il est raisonnable d'observer cette valeur sous H_0. Ici, la p-valeur vaut environ 40%, donc on accepte H_0 au niveau inférieur à 5%.

2.3 TEST ASYMPTOTIQUE

Dans la mesure où n est relativement grand, on peut utiliser l'approximation gaussienne de $\bar{X}_n$, ce qui donne une zone de rejet très proche de R^1. Le test est alors *asymptotique*.

Définition [TEST ASYMPTOTIQUE] *Soit* $(\mathscr{H}^n, \{P_\theta\}_{\theta\in\Theta})$ *un modèle statistique et* R_n *une zone de rejet pour* H_0 *vs* H_1*. Le test est*

1. *de niveau asymptotique* α *si*
$$\limsup_{n\to\infty} \sup_{\theta\in\Theta_0} P_\theta(R_n) = \alpha.$$
2. *convergent ou consistant si, pour tout* $\theta \in \Theta_1$,
$$\lim_{n\to\infty} P_\theta(R_n) = 1.$$

La zone de rejet asymptotique est encore de la forme $R_n^1 = \{y \in \{0,1\}^n : \bar{y}_n > a\}$. De plus, comme précédemment :

$$\sup_{\theta\in\Theta_0} P_\theta(R_n^1) = \sup_{\theta\leq\theta_s}(R_n^1) = P_{\theta_s}(R_n^1).$$

Notons $g_{1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi $\mathcal{N}(0,1)$. D'aprés le théorème de la limite centrale,

$$\lim_n P_{\theta_s}\left(\bar{X}_n > \sqrt{\frac{\theta_s(1-\theta_s)}{n}}\, g_{1-\alpha} + \theta_s\right) = \alpha.$$

Par suite, pour toute valeur de a telle que

$$a \geq \sqrt{\frac{\theta_s(1-\theta_s)}{n}}\, g_{1-\alpha} + \theta_s, \text{ on a } \limsup_n \sup_{\theta\in\Theta_0} P_\theta(R_n^1) \leq \alpha.$$

Par ailleurs, la plus petite de ces valeurs de a donne une puissance maximale. Ainsi, la zone de rejet asymptotique est

$$R_n^1 = \left\{ y \in \{0,1\}^n : \bar{y}_n > \sqrt{\frac{\theta_s(1-\theta_s)}{n}}\, g_{1-\alpha} + \theta_s \right\}.$$

Cette zone de rejet définit un test de niveau asymptotique α, qui est de plus convergent car sous H_1, i.e. si $\theta > \theta_s$:

$$\lim_n P_\theta\left(\bar{X}_n > \sqrt{\frac{\theta_s(1-\theta_s)}{n}}\, g_{1-\alpha} + \theta_s\right) = P_\theta(\theta > \theta_s) = 1.$$

Pour les valeurs $\alpha = 5\%$, $n = 100$ et $\theta_s = 0.2$, on a $g_{1-\alpha} = 1.64$ d'où $R_n^1 = \{y \in \{0,1\}^n : \bar{y}_n > 0.24\} \approx R^1$. Comme $\bar{x}_n = 0.22$, on accepte une nouvelle fois H_0 au niveau $\alpha = 5\%$.

On aurait aussi pu calculer la p-valeur approchée du test. Comme $\bar{x}_n = 0.22$, elle vaut

$$\sup_{\theta\leq\theta_s} P_\theta(\bar{X}_n > 0.22) = P_{\theta_s}(\bar{X}_n > 0.22) \approx \mathbb{P}\left(G > \sqrt{\frac{n}{\theta_s(1-\theta_s)}}(0.22-\theta_s)\right) = \mathbb{P}(G > 0.5) = 40\%,$$

G désignant une v.a. de loi $\mathcal{N}(0,1)$, ce qui nous amène à accepter H_0 au niveau $\alpha = 5\%$ (l'inégalité de Berry-Esséèn permet de quantifier l'erreur commise dans cette approximation).

2.4 POINT DE VUE D'UN ORGANISME DE CONTRÔLE DE LA QUALITÉ

Dans un élan d'objectivité, acceptons l'éventualité de mettre en péril notre décision en nous placant du point de vue d'un organisme de contrôle de la qualité. Le test confronte alors $H_0 : \theta > \theta_s$ contre $H_1 : \theta \leq \theta_s$. La zone de rejet au niveau α est

$$R^2 = \left\{ y \in \{0,1\}^n : \bar{y}_n \leq \frac{q_\alpha}{n} \right\},$$

q_α désignant le quantile d'ordre α de la loi $\mathscr{B}(n,\theta_s)$. On calcule alors $R^2 = \{y \in \{0,1\}^n : \bar{y}_n \leq 0.13\}$, ce qui nous amène à accepter H_0 au niveau 5%. Du point de vue d'un organisme de contrôle de la qualité, il faut arrêter la production !

L'occasion est idéale pour rappeler qu'un test statistique ne rend pas une décision absolue, et que la question posée oriente déjà la réponse ! De la même manière que, dans un procès d'assise, il y a présomption d'innocence, en théorie des tests, il y a présomption de H_0 : il faut donc montrer que H_0 est peu probable pour le rejeter, et H_0 peut être acceptée à tort faute de preuves plaidant en faveur de son rejet (par exemple à cause d'un nombre d'observations trop réduit).

Il est clair ici que, si les 2 tests statistiques ont rendu des décisions différentes, c'est parce que les valeurs de θ_s et de $\bar{x}_n$ sont très proches. Si les contraintes de l'entreprise avaient menées à une valeur $\theta_s = 0.3$ par exemple, on aurait obtenu pour le 1er test la zone de rejet $\{y \in \{0,1\}^n : \bar{y}_n > 0.37\}$, et pour le 2nd test $\{y \in \{0,1\}^n : \bar{y}_n \leq 0.23\}$. Comme $\bar{x}_n = 0.22$, on est donc amené, dans les 2 cas, à décider que $\theta \leq \theta_s$, ce qui n'a rien d'étonnant au vu des observations...