Introduction aux statistiques
MA3-4GE

E.Bretin, Z. Smida

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON
Pole Math

1. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison

2. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

1. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison

2. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Échantillon

Définition 1

On appelle **échantillon de taille** *n de X une suite de* **variables aléatoires** $\{X_i\}_{1\leq i\leq n}$ *de même loi que X.*
On parle d'échantillon

- **Exhaustif** *: les variables X_i ne sont pas indépendantes les unes des autres*
- **Non exhaustif** *: les variables X_i sont 2 à 2 indépendantes. On parle alors de variables aléatoires indépendantes identiquement distribuées (v.a. i.i.d).*

On note aussi $\{x_i\}_{1\leq i\leq n}$**, la réalisation** *de cet échantillon*

Motivation

Soit X une variable aléatoire de loi $\mathcal{L}(\theta)$ paramétrée par un paramètre $\theta \in \mathrm{R}^d$.

L'objectif est d'estimer une valeur de θ à partir de la réalisation $\{x_i\}_{1 \leq i \leq n}$ d'un échantillon indépendant et identiquement distribué $\{X_i\}_{1 \leq i \leq n}$ de la loi de X,

Définition 2

Un estimateur de θ est une statistique

$$\Theta_n = h(X_1, X_2, \cdots, X_n)$$

dont une réalisation

$$\theta_n = h(x_1, x_2, \cdots, x_n)$$

permet de donner une estimation "plausible" de θ.

Propriétés d'un estimateur

Définition 3

- *Un estimateur est dit sans biais si*

$$\mathbb{E}[\Theta_n] = \theta.$$

- *Un estimateur est dit convergent si*

$$\lim_{n\to\infty} \mathbb{V}[\Theta_n] = 0.$$

Propriété 1

Si $\mathbb{E}[\Theta_n] = \theta$ *et* $\lim_{n\to\infty} \mathbb{V}[\Theta_n] = 0$, *alors*

$$\lim_{n\to\infty} P(|\Theta_n - \theta| > \varepsilon) = 0.$$

Erreur quadratique d'un estimateur

Définition 4

Soit Θ_n un estimateur de θ. L'erreur quadratique de Θ_n est définie comme

$$\mathbb{E}[(\Theta_n - \theta)^2] = (\mathbb{E}[\Theta_n] - \theta)^2 + \mathbb{V}[\Theta_n].$$

Si $\mathbb{E}[(\Theta_n - \theta)^2] \to 0$ quand $n \to \infty$, on dit que Θ_n est un estimateur de θ convergeant en moyenne quadratique.

Pour un estimateur sans biais, la convergence en moyenne quadratique équivaut à la convergence vers 0 de la variance. L'idée est alors de chercher à construire de **estimateurs sans biais** et **de plus petite variance possible**.

Exemple avec une loi de Bernoulli

Soit $\{X_i\}_{1 \leq i \leq n}$ un échantillon indépendant et identiquement distribué de loi de Bernoulli $X_i \sim \mathcal{B}(p)$.

Quel choix pour l'estimateur du paramètre $p \in [0, 1]$?

- Espérance et variance :

$$\mathbb{E}(X_i) = p \quad \text{et} \quad \mathbb{V}(X_i) = p(1-p).$$

- **Choix de l'estimateur:** moyenne de l'échantillon

$$\mathbf{P_n = \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i}$$

- **Un estimateur sans biais**

$$\mathbb{E}[P_n] = \mathbb{E}[X_i] = p,$$

- **Un estimateur convergeant:**

$$\mathbb{V}[P_n] = \frac{p(1-p)}{n} \to 0 \text{ quand } n \to \infty.$$

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON
Pole Math

Estimation de l'espérance et de la variance

Soit $\{X_i\}_{1 \leq i \leq n}$ un échantillon indépendant et identiquement distribué de loi d'espérance $\mathbb{E}(X_i) = \mu$ et de variance $\mathbb{V}(X_i) = \sigma^2$.

Quel choix pour l'estimateur du paramètre $\mu \in \mathbb{R}$?

- **Moyenne de l'échantillon**

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- **Un estimateur sans biais et convergeant:**

$$\mathbb{E}[\overline{X}_n] = \mu \text{ et } \mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}.$$

Quel choix pour l'estimateur du paramètre $\sigma^2 \in \mathbb{R}^+$?

Estimation de l'espérance et de la variance

- **Variance empirique**

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2.$$

- **Un estimateur biaisé !**

$$\begin{aligned}
\mathbb{E}[S_n^2] &= \mathbb{E}[(X_1 - \overline{X}_n)^2] = \frac{(n-1)^2}{n^2}\mathbb{E}[(X_1 - \frac{1}{n-1}\sum_{i=2}^{n} X_i)^2] \\
&= \frac{(n-1)^2}{n^2}\mathbb{E}[((X_1 - \mu) - (\frac{1}{n-1}\sum_{i=2}^{n} X_i - \mu))^2] \\
&= \frac{(n-1)^2}{n^2}\left(\mathbb{E}[(X_1 - \mu)^2)] + \mathbb{E}[(\frac{1}{n-1}\sum_{i=2}^{n} X_i - \mu))^2] - 2\mathbb{E}[(X_1 - \mu)(\frac{1}{n-1}\sum_{i=2}^{n} X_i - \mu)]\right) \\
&= \frac{(n-1)^2}{n^2}(\sigma^2(1 + 1/(n-1))) = \frac{n-1}{n}\sigma^2
\end{aligned}$$

- **Variance empirique débaisée**

$$S_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \text{ avec } \mathbb{E}[S_n'^2] = \sigma^2$$

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON
Pole Math

Méthode des moments

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué de loi $\mathcal{L}(\theta)$

- **Exprimer les moments de X_i en fonction de θ**

$$\mathbb{E}[X_i] = g(\theta), \quad \mathbb{V}[X_i] = h(\theta).$$

- **Exprimer θ en fonction des moments de X_i**

$$\theta = m(\mathbb{E}[X_i], \mathbb{V}[X_i]),$$

- **Considérer l'estimateur Θ_n suivant :**

$$\Theta_n = m(\overline{X_n}, S_n'^2).$$

Exemple avec une loi géométrique

Soit $\{X_i\}_{1 \leq i \leq n}$ un échantillon indépendant et identiquement distribué de loi géométrique $\mathcal{G}(p)$ avec $p \in [0, 1]$.

- **Loi de** X_i

$$P(X_i = k) = (1 - p)^{k-1} p$$

- **Moments** de X_i

$$\mathbb{E}[X_i] = \sum_{k=1}^{\infty} k(1 - p)^{k-1} p = \frac{1}{p}$$

- **Estimateur** de la méthode des moments pour p:

$$P_n = \frac{1}{\overline{X}_n}$$

- **Etude des propriétés** de ces 2 estimateurs ?

Exemple avec une loi uniforme

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué de loi uniforme $\mathcal{U}([a,b])$ avec $b > a$.

- Estimation de ces 2 premiers moments:

$$\mathbb{E}[X_i] = \frac{a+b}{2} \text{ et } \mathbb{V}[X_i] = (b-a)^2/12$$

- Estimation de a et b en fonction de $\mathbb{E}[X_i]$ et $\mathbb{V}[X_i]$:

$$a = \mathbb{E}[X_i] - \sqrt{3\mathbb{V}[X_i]} \text{ et } b = \mathbb{E}[X_i] + \sqrt{3\mathbb{V}[X_i]}$$

- Exemple d'estimateur par la méthode des moments

$$A_n = \overline{X}_n - \sqrt{3S_n'^2} \text{ et } B_n = \overline{X}_n + \sqrt{3S_n'^2}.$$

- Etude des propriétés de ces 2 estimateurs ?

Fonction de vraisemblance

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué d'une loi $\mathcal{L}(\theta)$.

Définition 5

On appelle **la vraisemblance** *de la réalisation* $\{x_i\}_{1\leq i\leq n}$ *d'un échantillon* $\{x_i\}_{1\leq i\leq n}$ *la quantité*

$$\mathcal{L}(\theta, \{x_1, x_2, \cdots x_n\}) = \begin{cases} \Pi_{i=1}^{n} P_\theta(X_i = x_i) & \text{si la loi est discrète} \\ \Pi_{i=1}^{n} f_\theta(x_i) & \text{si la loi est continue} \end{cases}$$

Estimateur de maximum de vraisemblance

Définition 6

La méthode du maximum de vraisemblance *consiste alors à déterminer la réalisation* $\hat{\theta}_n$ *de l'estimateur* $\hat{\Theta}_n$ *qui maximise la vraisemblance:*

$$\hat{\theta}_n = \arg\max_{\theta} \{\mathcal{L}(\theta, \{x_1, x_2, \cdots x_n\})\}$$

Propriété 2

Si Θ_n *est l'estimateur du maximum de vraisemblance pour le paramètre* θ*, alors* $h(\Theta_n)$ *est l'estimateur du maximum de vraisemblance pour le paramètre* $h(\theta)$*.*

Exemple avec une loi géométrique

Soit $\{X_i\}_{1 \leq i \leq n}$ un échantillon indépendant et identiquement distribué de loi géométrique $\mathcal{G}(p)$ avec $p \in [0, 1]$.

- **Expression de la vraisemblance $\mathcal{L}$**

$$\mathcal{L}(p, \{x_1, x_2, \cdots x_n\}) = \Pi_{i=1}^{n} (1-p)^{x_i - 1} p = \frac{p^n}{(1-p)^n} (1-p)^{\sum_{i=1}^{n} x_i}$$

- **Calcul du log de $\mathcal{L}$:**

$$log(\mathcal{L}(p, \{x_1, x_2, \cdots x_n\})) = n\, log(p) - n\, log(1-p) + \sum_{i=1}^{n} x_i\, log(1-p)$$

- **Dérivation par rapport à p**

$$\partial_p [log(\mathcal{L}(p, \{x_1, x_2, \cdots x_n\}))] = n/p + n/(1-p) + \sum_{i=1}^{n} x_i/(1-p)$$

Exemple avec une loi géométrique

- **Résolution de l'équation** $\partial_p[log(\mathcal{L}(p, \{x_1, x_2, \cdots x_n\}))] = 0,$

$$n/\hat{p_n} + n/(1-\hat{p_n}) + \sum_{i=1}^{n} x_i/(1-\hat{p_n}) = 0 \Leftrightarrow \hat{p_n} = \frac{n}{\sum_{i=1}^{n} x_i}$$

- **Déduction de l'estimateur de maximum de vraisemblance**

$$\hat{P}_n = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}_n}$$

Exemple avec une loi uniforme

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué de loi uniforme $\mathcal{U}([a,b])$ avec $b > a$.

- **Expression de la vraisemblance** $\mathcal{L}$

$$\mathcal{L}(a,b,\{x_1,x_2,\cdots x_n\}) = \frac{1}{(b-a)^n} 1_{[a,b]}(x_i)$$

En particulier

$$\mathcal{L}(a,b,\{x_1,x_2,\cdots x_n\}) = \begin{cases} 0 & \text{si } a > \min x_i \text{ ou } b < \max x_i \\ \frac{1}{(b-a)^n} & \text{si } a \leq \min x_i \text{ et } b \geq \max x_i \end{cases}$$

On en déduit que $\hat{a}_n = \min_{i=1:n} x_i$ et $\hat{b}_n = \max_{i=1:n} x_i$

- **Estimateur de maximum de vraisemblance** de a et b

$$\hat{A}_n = \min_{i=1:n} X_i \text{ et } \hat{B}_n = \max_{i=1:n} X_i$$

Exercice 1

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué d'une loi normale $\mathcal{N}(\mu, \sigma^2)$.

1) *Calculer l'estimateur du maximum de vraisemblance pour μ.*
2) *Calculez l'estimateur du maximum de vraisemblance pour vraisemblance pour σ^2. Quel est l'estimation du maximum de vraisemblance pour σ ?*

Exercice 2

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué d'une loi uniforme $\mathcal{U}([0,a])$ avec $a>0$.

1) *Montrer que la méthode des moments conduit à un estimateur $\overline{A}_n = 2\overline{X_n}$ qui est non biaisé et de variance $\mathbb{V}[\overline{A}_n] = a^2/(3n)$*
2) *Montrer que la méthode de maximum de vraisemblance conduit à l'estimateur $\hat{A}_n = \max_{i=1:n} X_i$ dont les moments s'identifient à*

$$\mathbb{E}[\hat{A}_n] = a\frac{n}{n+1}, \text{ et } \mathbb{V}[\hat{A}_n] = a^2\frac{n}{(n+1)^2(n+2)}.$$

3) *En déduire le meilleur estimateur du paramètre a ?*

Exercice 3

On sait qu'une durée T vérifie qu'il existe $d > 0$ tel que $X = T - d$ suit une loi exponentielle $\mathcal{E}(\lambda)$. On considère un échantillon $\{T_i\}_{1 \leq i \leq n}$ de taille n de la variable aléatoire T.

1) *Soit $A_n = \min_{1 \leq i \leq n}(T_i)$. Donner le domaine de A_n, sa densité, son espérance et sa variance.*
2) *On note $B_n = \overline{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_i$. Donner une approximation de la loi de B_n pour n grand. Calculer son espérance et son écart type.*
3) *Déterminer la vraisemblance de l'échantillon en fonction de λ et d.*
4) *Si λ est fixé, pour quelle valeur $\hat{d}_n$ de d cette vraisemblance est-elle maximale ?*
5) *Calculer en fonction de d l'estimation du maximum de vraisemblance pour le paramètre λ.*
6) *En déduire l'estimation du maximum de vraisemblance pour le couple (λ, d) et déterminer le biais de l'estimateur pour d.*

1. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison

2. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Motivation

Soit $\{X_i\}_{1 \leq i \leq n}$ un échantillon indépendant et identiquement distribué de loi $\mathcal{L}(\theta)$.

Nous avons vu précédemment comment obtenir un **estimateur** Θ_n du paramètre θ dont la réalisation à n fixé permet d'obtenir une valeur plausible du paramètre θ avec **une marge d'erreur non contôlée**. L'objectif de cette partie est de construire un **intervalle** dans lequel le paramètre θ a une **forte probabilité de se trouver**.

Définition 7

Soit θ un paramètre donné. On appelle **intervalle de confiance** *de niveau β un intervalle aléatoire $[T_1, T_2]$ tel que*

$$P(\theta \in [T_1, T_2]) = \beta.$$

Exemple pour une proportion

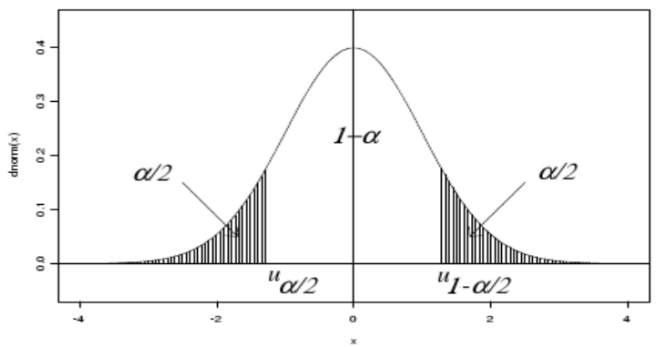
Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué de loi de Bernoulli $X_i \sim \mathcal{B}(p)$.

- **Estimateur ponctuel** de p: $P_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.
- **Théorème central limite**

$$\frac{\sum_{i=1}^n X_i - n\mathbb{E}[X]}{\sqrt{n\mathbb{V}[X]}} = \frac{\overline{X_n} - p}{\sqrt{p(1-p)/n}} \to U \sim \mathcal{N}(0,1)$$

- **Quantile de la loi normale:**

$$P(U \leq u_\alpha) = \alpha$$



- **Propriété de symétrie** : $u_\alpha = -u_{1-\alpha}$.

Table de loi :

Quantile de loi normale $\alpha \mapsto u_\alpha$.

	[0.000]	[0.005]	[0.010]	[0.015]	[0.020]	[0.025]	[0.030]	[0.035]	[0.040]	[0.045]
[0.50]	0.0000	0.0125	0.0251	0.0376	0.0502	0.0627	0.0753	0.0878	0.1004	0.1130
[0.55]	0.1257	0.1383	0.1510	0.1637	0.1764	0.1891	0.2019	0.2147	0.2275	0.2404
[0.60]	0.2533	0.2663	0.2793	0.2924	0.3055	0.3186	0.3319	0.3451	0.3585	0.3719
[0.65]	0.3853	0.3989	0.4125	0.4261	0.4399	0.4538	0.4677	0.4817	0.4959	0.5101
[0.70]	0.5244	0.5388	0.5534	0.5681	0.5828	0.5978	0.6128	0.6280	0.6433	0.6588
[0.75]	0.6745	0.6903	0.7063	0.7225	0.7388	0.7554	0.7722	0.7892	0.8064	0.8239
[0.80]	0.8416	0.8596	0.8779	0.8965	0.9154	0.9346	0.9542	0.9741	0.9945	1.0152
[0.85]	1.0364	1.0581	1.0803	1.1031	1.1264	1.1503	1.1750	1.2004	1.2265	1.2536
[0.90]	1.2816	1.3106	1.3408	1.3722	1.4051	1.4395	1.4758	1.5141	1.5548	1.5982
[0.95]	1.6449	1.6954	1.7507	1.8119	1.8808	1.9600	2.0537	2.1701	2.3263	2.5758

Exemple pour une proportion

- **Intervalle centrée en** 0 **pour** U:

$$P\left(U^2 \leq u_{\alpha/2}^2\right) = 1 - \alpha, \text{ avec } U \simeq \frac{\overline{X_n} - p}{\sqrt{p(1-p)/n}}$$

- **équivalence**

$$\begin{aligned} U^2 \leq u_{\alpha/2}^2 &\Leftrightarrow (\overline{X_n} - p)^2 \leq \frac{p(1-p)}{n} u_{\alpha/2}^2 \\ &\Leftrightarrow (1 + \frac{u_{\alpha/2}^2}{n})p^2 - (2\overline{X_n} + \frac{u_{\alpha/2}^2}{n})p + \overline{X_n}^2 \leq 0 \\ &\Leftrightarrow p \in [\Lambda_n^-, \Lambda_n^+] \text{ où } \Lambda_n^{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \end{aligned}$$

- **Intervalle de confiance pour** p**:**

$$P(p \in [\Lambda_n^-, \Lambda_n^+]) \simeq 1 - \alpha$$

Exemple pour une proportion

- **Intervalle de confiance** en utilisant l'approximation $p \simeq \overline{X_n}$:

$$P(p \in [\Lambda_n^-, \Lambda_n^+]) \simeq 1 - \alpha$$

avec

$$\Lambda_n^{\pm} = \overline{X_n} \pm \frac{|u_{\alpha/2}|}{\sqrt{n}} \sqrt{\overline{X_n}(1 - \overline{X_n})}$$

Remarque 1

La taille caractéristique de cet intervalle est de l'ordre de $|u_{\alpha/2}|/\sqrt{n}$.

- *L'intervalle s'agrandit lorsque* α *décroît (et donc que la probabilité d'erreur diminue)*
- *L'intervalle se rétrécit lorsque n augmente avec une taille caractéristique en* $1/\sqrt{n}$

Exercice 4

Une entreprise hésite à changer son logo et lance un sondage parmi ses salariés pour savoir quel logo plaît le plus. Le taux de participation au sondage est de 67%. *Parmi les* 1011 *réponses,* 51% *des personnes interrogées déclaraient préférer le deuxième logo.*

1) *Donner un intervalle de confiance asymptotique de niveau* 95% *de la proportion de salariés qui préfèrent le second logo. Commenter.*
2) *Combien de personnes faudrait-il pour avoir une largeur d'intervalle* 0.02 *avec des données similaires.*
3) *Reprendre la question* 1 *pour un intervalle de confiance de niveau* 99%.

Cas de la loi normale

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué de loi normale $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

1er cas: intervalle de confiance pour μ avec σ connu

- **Moyenne empirique**

$$\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \Rightarrow \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = U \sim \mathcal{N}(0,1).$$

- **Intervalle pour une loi normale centrée réduite**

$$P(U \in [u_{\alpha/2}, u_{1-\alpha/2}]) = 1-\alpha$$

- **Intervalle de confiance pour μ**

$$U \in [u_{\alpha/2}, u_{1-\alpha/2}] \Leftrightarrow \mu \in [\overline{X}_n + \frac{u_{\alpha/2}}{\sqrt{n}}\sigma, \overline{X}_n + \frac{u_{1-\alpha/2}}{\sqrt{n}}\sigma]$$

Cas de la loi normale

2ème cas: intervalle de confiance pour σ^2 avec σ inconnu

- **Moyenne empirique**

$$\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \Rightarrow \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = U \sim \mathcal{N}(0,1).$$

- **Utiliser la variance empirique** $S_n'^2$ à la place de σ^2

$$S_n'^2 = \frac{1}{n-1}\sum (X_i - \overline{X_n})^2 \simeq \sigma^2$$

- **Quelle loi est la loi de** T_{n-1} **?**

$$T_{n-1} = \frac{\overline{X}_n - \mu}{\sqrt{S_n'^2/n}}$$

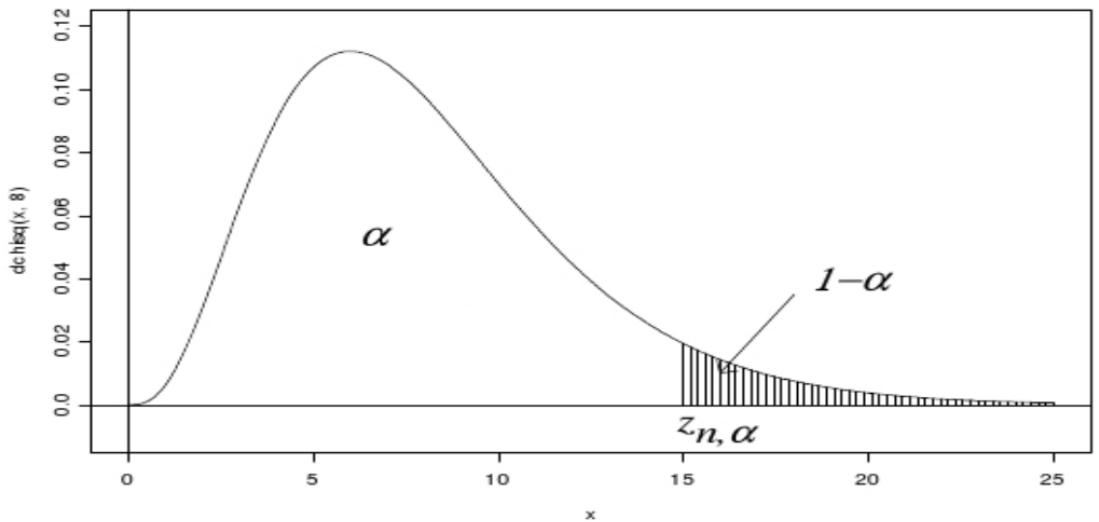
Loi du χ^2

Définition 8

Soient $X_1, \cdots, X_n$ indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$, alors $Z = \sum_{i=1}^{n} X_i^2 \sim \chi^2_n$ suit une loi du χ^2 à n degré de liberté.

Quantile de la loi du χ^2:

$$P(Z \leq z_{n,\alpha}) = \alpha$$



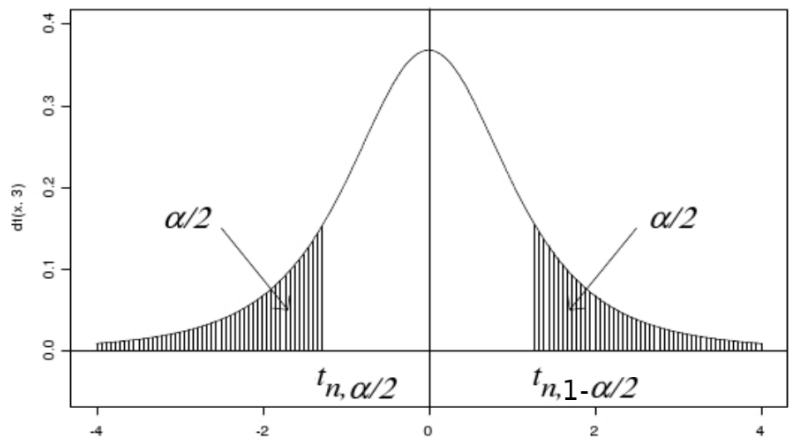
Loi de student

Définition 9

Si $U \sim \mathcal{N}(0,1)$ et $Z \sim \chi^2_n$, et si U et Z sont indépendants, alors $T = \frac{U}{\sqrt{Z/n}} \sim St(n)$ suit une loi de Student à à n degré de liberté.

Quantile de la loi du $St(n)$:

$$P(T \leq t_{n,\alpha}) = \alpha$$



Propriété de symétrie: $t_{n,\alpha} = t_{n,1-\alpha}$

Théorème de Fisher

Propriété 3

Soient $X_1, \cdots, X_n$ indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$.

- $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *et* $S_n'^2 = \frac{1}{n-1}\sum(X_i - \overline{X_n})^2$ *sont indépendants*
- $U = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
- $Z_n = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2_n$
- $Z_{n-1} = \frac{1}{\sigma^2}(n-1)S_n'^2 \sim \chi^2_{n-1}$
- $\frac{\overline{X_n} - \mu}{\sqrt{S_n'^2/n}} = \frac{U}{\sqrt{Z_{n-1}/(n-1)}} \sim St(n-1)$

Cas de la loi normale

Retour au 2ème cas: intervalle de confiance pour σ^2 avec σ inconnu

- **Moyenne empirique:**

$$\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \Rightarrow \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = U \sim \mathcal{N}(0,1).$$

- **Quelle loi est la loi de T_{n-1} ?**

$$T_{n-1} = \frac{\overline{X}_n - \mu}{\sqrt{S_n'^2/n}} \sim St(n-1)$$

- **Intervalle** pour T_{n-1}

$$P(T_{n-1} \in [t_{n-1,\alpha/2}, t_{n-1,1-\alpha/2}]) = 1 - \alpha$$

- **Intervalle de confiance pour μ**

$$T_{n-1} \in [t_{n-1,\alpha/2}, t_{n-1,1-\alpha/2}] \Leftrightarrow \mu \in [\overline{X}_n + \frac{t_{n-1,\alpha/2}}{\sqrt{n}} S_n', \overline{X}_n + \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}} S_n']$$

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON
Pole Math

Cas de la loi normale

3ème cas: intervalle de confiance pour σ^2 avec μ inconnu

- **Variance empirique**

$$S_n'^2 = \frac{1}{n-1}\sum (X_i - \overline{X_n})^2 \text{ et } \frac{1}{\sigma^2}(n-1)S_n'^2 = Z_{n-1} \sim \chi^2_{n-1}$$

- **Intervalle** pour Z_{n-1}

$$P(Z_{n-1} \in [z_{n-1,\alpha/2}, z_{n-1,1-\alpha/2}]) = 1-\alpha$$

- **Intervalle de confiance pour** σ^2

$$P\left(\sigma^2 \in \left[\frac{S_n'^2}{z_{n-1,1-\alpha/2}/(n-1)}, \frac{S_n'^2}{z_{n-1,\alpha/2}/(n-1)}\right]\right) = 1-\alpha$$

Exercice 5

Une chaîne de fabrication est dédiée à la production de yaourts. Une doseuse remplit les pots et la masse ainsi versée suit une loi normale d'espérance μ et d'écart type σ. Chaque heure, un contrôle portant sur 16 yaourts est effectué par le fabricant pour contrôler la valeur de μ. Les valeurs trouvées sont les suivantes :

$$124.5, 121.6, 125.5, 125.3, 125.6, 124.6, 116.8, 124.3,$$
$$127.9, 124.8, 128.2, 118.9, 130.0, 121.2, 123.8, 130.6$$

Au vu de cet échantillon.

1) *Calculer un intervalle de confiance pour μ de niveau $\alpha = 10\%$*
2) *Calculer un intervalle de confiance pour σ^2 de niveau $\alpha = 10\%$*

Exercice 6

Deux étudiants désirent estimer le paramètre a d'une loi uniforme sur l'intervalle $[0, a]$. *Ils disposent pour cela d'un échantillon* $\{X_1, X_2, \cdot X_n\}$ *de taille* $n = 50$.

- *Le premier propose d'utiliser l'estimateur* $A_n^1 = 2\overline{X}_n$
- *Le second propose d'utiliser l'estimateur* $A_n^2 = \max_{i=1:n} X_i$

1) *Donner l'expression des deux intervalles de confiances sur a associés à chacune de ces statistique pour un niveau de confiance* $1 - \alpha$.
2) *En déduire les 2 intervalles correspondants pour* $\alpha = 10\%$ *et dans le cas où* $\overline{x}_n = 18.7$ *et* $m_n = 35.0$.

1. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison

2. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Objectif et principe

Un test d'hypothèse a pour **objectif** de décider sur la base d'un **échantillon** si la loi mère vérifie ou non une **hypothèse** sur un de ses **paramètres** caractéristiques.

Démarche:

- Choix de **l'hypothèse** H_0 à tester (et de l'hypothèse alternative H_1)
- Choix d'une **statistique** sur un échantillon qui permet de tester H_0
- Choix d'un niveau de signification : **le risque** α de rejeter H_0 alors qu'elle est vraie
- Définition de **la zone d'acceptation** de H_0 et de rejet de H_0 (intervalle unilatéral ou bilatéral)

Hypothèses H_0, H_1 et risques

		Réalité	
		H_0 vraie	H_0 fausse
Conclusion du test	Non rejet de H_0	Ok	Risque β
	Rejet de H_0	Risque α	Ok

- **Risque α:** risque de rejeter à tort H_0. On rejette H_0 si l'échantillon a une très faible probabilité de se produire
- **Risque β:** risque de ne pas rejeter H_0 alors que H_1 est vrai.
- **la puissance du test** $1 - \beta$: pour un risque α donné, on essaie de choisir un test avec un risque β le plus petit possible en jouant sur **la statistique de test** ou encore sur la forme de **l'intervalle d'acceptation** de H_0.

Test d'hypothèse sur une proportion

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué de loi de Bernoulli $X_i \sim \mathcal{B}(p)$.

- **Exemple de choix d'hypothèses**:

$$H_0 = "p = p_0" \text{ contre } H_1 = "p \neq p_0"$$

- **Choix d'une statistique** de test: $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- En supposant H_0 vraie, **une zone d'admissibilité** correspond à un $[\ell_\alpha^-, \ell_\alpha^+]$ telle que

$$P(\overline{X}_n \in [\ell_\alpha^-, \ell_\alpha^+]) = 1 - \alpha.$$

En particulier, si H_0 est vraie, et pour n suffisamment grand,

$$\frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \simeq U \sim \mathcal{N}(0,1).$$

- Utilisation **des quantiles de la loi normale**

$$P(u_{\alpha/2} \leq U \leq u_{1-\alpha/2}) = 1-\alpha \Leftrightarrow P(\overline{X}_n \in [\ell_\alpha^-, \ell_\alpha^+]) \simeq 1-\alpha,$$

avec

$$\ell_\alpha^- = p_0 + u_{\alpha/2}\sqrt{p_0(1-p_0)/n} \text{ et } \ell_\alpha^- = p_0 + u_{1-\alpha/2}\sqrt{p_0(1-p_0)/n}$$

- **La construction du test est terminé**. Il suffit maintenant de vérifier si la réalisation $\overline{x}_n$ de la statistique $\overline{X}_n$ appartient ou pas à l'intervalle $[\ell_\alpha^-, \ell_\alpha^+]$ pour ne pas rejeter ou rejeter H_0.

Test d'hypothèse sur une proportion

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué de loi de Bernoulli $X_i \sim \mathcal{B}(p)$.

- Dans certaine application, **il peut aussi être intéressant de tester**

$$H_0 = "p = p_0" \text{ contre } H_1 = "p = p_1 > p_0"$$

- On peut alors garder la statistique $\overline{X}_n$.
- Il est dans ce cas plus intéressant de considérer une **zone d'admissibilité** de la forme $[-\infty, \ell_\alpha^+]$ telle que

$$P(\overline{X}_n \leq \ell_\alpha^+) = 1 - \alpha.$$

Comme précédemment, sous l'hypothèseH_0, et pour n suffisamment grand,

$$\frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \simeq U \sim \mathcal{N}(0,1).$$

- Utilisation des **quantiles de la loi normale**

$$P(U \leq u_{1-\alpha}) = 1-\alpha \Leftrightarrow P(\overline{X}_n \leq \ell_\alpha^+) \simeq 1-\alpha,$$

avec

$$\ell_\alpha^+ = p_0 + u_{1-\alpha}\sqrt{p_0(1-p_0)/n}$$

- Il est alors possible de déterminer **le risque** β. En effet, sous l'hypothèse H_1,

$$\frac{\overline{X}_n - p_1}{\sqrt{p_1(1-p_1)/n}} \simeq U \sim \mathcal{N}(0,1),$$

et alors

$$\begin{aligned}\beta &= P(\overline{X}_n \leq p_0 + u_{1-\alpha}\sqrt{p_0(1-p_0)/n}) \\ &\simeq P\left(U \leq \frac{p_0 - p_1 + u_{1-\alpha}\sqrt{p_0(1-p_0)/n)}}{\sqrt{p_1(1-p_1)/n}}\right)\end{aligned}$$

Exercice 7

Pièce équilibré

On désire vérifier qu'une pièce n'avantage pas face avec un nombre minimum de lancers. Pour ce faire, on souhaite réaliser un test avec :

- $H_0 = "p = 0.5"$ *contre* $H_1 = "p = 0.51"$
- $\alpha = 0.2$ *et* $\beta = 0.4$

Construire un test simple répondant à ces critères.

Pièce défectueuse

On désire vérifier la proportion p de pièces défectueuses d'une production. Pour ce faire, on souhaite réaliser un test avec :

- $H_0 = "p = 0.1"$ *contre* $H_1 = "p = 0.2"$
- $\alpha = 0.2$ *et* $\beta = 0.4$

Construire un test simple répondant à ces critères et calculez β*. Répondre aux mêmes questions si* $n = 60$*.*

Test sur l'espérance d'une loi normale

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué de loi normale $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

- Exemple de **choix d'hypothèses:**

$$H_0 = "\mu = \mu_0" \text{ contre } H_1 = "\mu \neq \mu_0"$$

- Choix **d'une statistique de test**: $\overline{X}_n$ et S'^2_n.
- Si H_0 est vraie,

$$\frac{\overline{X}_n - p_0}{\sqrt{S'^2_n/n}} = T_{n-1} \sim St(n-1).$$

- **Quantile pour la loi de Student**

$$P\left(t_{n-1,\alpha/2} \leq T_{n-1} \leq t_{n-1,1-\alpha/2}\right) = 1-\alpha.$$

- **Zone d'admissibilité** de H_0 :

$$\frac{\overline{X}_n - p_0}{\sqrt{S'^2_n/n}} \in \left[t_{n-1,\alpha/2}, t_{n-1,1-\alpha/2}\right].$$

Test sur la variance d'une loi normale

Soit $\{X_i\}_{1\leq i\leq n}$ un échantillon indépendant et identiquement distribué de loi normale $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

- Exemple **de choix d'hypothèses**:

$$H_0 = "\sigma = \sigma_0" \text{ contre } H_1 = "\sigma \neq \sigma_0"$$

- **Choix d'une statistique** de test: $S_n'^2$.
- Si H_0 est vraie,

$$\frac{(n-1)S_n'^2}{\sigma_0^2} = Z_{n-1} \sim \chi^2_{n-1}.$$

- **Quantile pour la loi de Chi-deux:**

$$P\left(z_{n-1,\alpha/2} \leq Z_{n-1} \leq z_{n-1,1-\alpha/2}\right) = 1-\alpha.$$

- **Zone d'admissibilité** de H_0:

$$S_n'^2 \in \left[\frac{\sigma_0^2\ t_{n-1,\alpha/2}}{n-1}, \frac{\sigma_0^2\ t_{n-1,1-\alpha/2}}{n-1}\right].$$

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON
Pole Math

Réussite du test et *p*-value

Comme nous l'avons vu précédemment, plus le risque α est supposé faible, plus la zone d'admissibilité de H_0 est grande. Il peut être intéressant d'introduire aussi **une notion de niveau de réussite du test**.

Définition 10

La *p*-valeur *d'un test correspond au seuil critique α_c pour laquelle le conclusion du test change:*

$$\begin{cases} \text{si } \alpha \leq \alpha_c, \text{ on ne rejette pas } H_0 \\ \text{si } \alpha > \alpha_c, \text{ on rejette } H_0 \end{cases}$$

En général, on cherchera donc à avoir **une *p*-valeur petite**, de sorte à pouvoir valider l'hypothèse (H_1) avec un faible risque de se tromper.

Exercice 8

Une chaîne de fabrication est dédiée à la production de yaourts. Une doseuse remplit les pots et la masse ainsi versée suit une loi normale $\mathcal{N}(\mu, \sigma^2)$ où la précision est a priori donnée par $\sigma = 3g$.

1) *Chaque jour, le fabricant contrôle σ en prélevant un échantillon de 30 pots. Construire un test correspondant. Vérifier que la réalisation $s'_n = 3.13g$ permet bien d'accepter $\sigma = 3g$.*

2) *On suppose ici que $\sigma = 3g$. Chaque heure, un contrôle portant sur 16 yaourts est effectué par le fabricant pour contrôler la valeur de μ. Les valeurs trouvées sont les suivantes :*

$$124.5, 121.6, 125.5, 125.3, 125.6, 124.6, 116.8, 124.3,$$
$$127.9, 124.8, 128.2, 118.9, 130.0, 121.2, 123.8, 130.6$$

Que pensez-vous de l'hypothèse $H_0 =$ "$\mu = 125$" contre l'hypothèse $H_1 =$ "$\mu < 125$" pour $\alpha = 0.05$ au vu de cet échantillon ?

3) *Même question si σ est inconnu ?*

Exercice 9

On désire tester la durée de vie de composants censée suivre une loi exponentielle d'espérance $1/\lambda$ *au moins égale à* $800h$*. Pour cela sont testés en simultané* $n = 100$ *composants. L'hypothèse alternative* $H_1 = "1/\lambda < 800h"$ *et le risque de première espèce est fixé à* $\alpha = 0.1$*.*

1) *Construire le test vérifiant que l'espérance n'est pas inférieure à* $800h$*.*
2) *Calculer l'espérance de la durée du test de la question précédente. Qu'en pensez-vous ?*

Pour accélérer le test, les composants sont regroupés par dispositifs de 10 *composants mis en série. Pour chaque série de* 10 *composants, une horloge mémorise l'instant I de la première panne d'un composant de la série.*

3) *Quelle est la loi de I ? Calculer l'espérance de la durée du test accéléré.*
4) *En utilisant les résultats sur les lois de chi-deux et les lois d'Erlang, construire le nouveau test.*

Exercice 10

Pour analyser l'accueil téléphonique d'un service après-vente, une centaine d'appels ont été analysés. La durée de la phase d'orientation de l'appel semble suivre une loi exponentielle. Une précédente étude avait estimé cette durée à $d = 150$ *secondes. Certains pensent que cette durée a augmenté, ce que l'étude doit infirmer ou confirmer. La moyenne des durées observées étant égale à* $\overline{x}_n = 175.1$, *l'hypothèse d'une durée de* 150 *secondes est-elle acceptable ?*

1. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison

2. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Objectif

L'objectif d'un test d'ajustement est de savoir si une **variable aléatoire** X suit ou non **une loi donnée**. Il s'agit d'un test où les hypothèses sont de la forme :

$$H_0 = \text{"}X \text{ suit la loi } \cdots\text{"} \quad \text{contre} \quad H_1 = \text{"}X \text{ ne suit pas la loi "}$$

L'idée est donc de prélever **un échantillon**, et d'essayer de **mesurer une "distance"** entre les valeurs observées, et les valeurs attendues si H_0 est vérifiée.

Exemple avec dé:

On effectue 72 lancers d'un dé à 6 faces. Voici les résultats obtenus.

i	1	2	3	4	5	6
n_i	4	10	16	10	12	20

On souhaite tester l'hypothèse

$$H_0 = \text{"le dé n'est pas pipé"}$$

contre l'hypothèse

$$H_1 = \text{"le dé est pipé"}$$

Question: sous l'hypothèse H_0, quelle est la loi vérifiée par ces lancers de dé ?

Loi multinomiale

Définition 11 (Loi multinomiale)

La loi multinomiale *généralise la loi binomiale en considérant m variables aléatoires N_i couplées de loi binomiale $\mathcal{B}(p_i, n)$ pour $i \in \{1, 2, \cdots m\}$:*

$$P(N_i = n_i) = \frac{n!}{n_i!(n - n_i!)} p_i^{n_i} (1 - p_i)^{n - n_i},$$

suivant la loi

$$P(N_1 = n_1, N_2 = n_2, \cdots N_m = n_m) = \frac{n!}{n_1! n_2! \cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

Propriété 4

Si $(N_1, N_2, \cdots, N_m)$ suit une loi multinomiale de paramètres $(p_1, p_2, \cdot p_m, n)$, alors

$$\lim_{n \to \infty} \sum_{i=1}^{m} \frac{(N_i - np_i)^2}{np_i} = Z_{m-1} \sim \chi^2_{m-1}.$$

Retour à l'exemple avec un dé:

On suppose que l'hypothèse H_0 est vérifiée.

- Le vecteur aléatoire $(N_1, N_2, N_3, N_4, N_5, N_6)$ suit une **loi multinomiale** de paramètres $(1/6, 1/6, 1/6, 1/6, 1/6, 1/6, 72)$.
- **Statistique de test:**

$$\Delta^2 = \sum_{i=1}^{6} \frac{(N_i - \frac{72}{6})^2}{\frac{72}{6}} \simeq Z_5 \sim \chi^2_5$$

- **Zone d'acceptation**

$$I = [0, z_{5,1-\alpha}].$$

- **Application** avec $\alpha = 0.05$:

$$\delta^2 = \sum_{i=1}^{6} \frac{(n_i - \frac{72}{6})^2}{\frac{72}{6}} = 12.67 > 11.0705.$$

$\Rightarrow$ on ne peut pas accepter H_0.

Test du chi-deux, cas variables discrètes

Soient $X_1, X_2, \cdots X_n$, n répétitions indépendantes d'une variables aléatoire X, à valeurs dans $e_1, e_2, \cdots e_k$.
Nous souhaitons tester **l'hypothèse**
$H_0 =$ "X suit la loi discrète $P(X = e_j) = p_j$" contre
$H_1 =$ "X ne suit pas la loi discrète $P(X = e_j) = p_j$"
Soient N_j la variable aléatoire correspondant au nombre de variables X_i ayant pour réalisation e_j.

- **Sous l'hypothèse** H_0:
  le vecteur $(N_1, \cdots, N_k)$ suit une loi binomiale de paramètre $(p_1, \cdots, p_k, n)$"
- **Statistique de test**

$$\Delta^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \simeq Z_{k-1} \sim \chi^2_{k-1}$$

- **Zone d'acceptation de** H_0

$$I = [0, z_{k-1,1-\alpha}]$$

Test du chi-deux, cas variables continues

Soient $X_1, X_2, \cdots X_n$, n répétitions indépendantes d'une variables aléatoire X continues dans un ensemble E.
Nous souhaitons tester **l'hypothèse**
$H_0 =$ " X suit la loi continue de densité f" contre
$H_1 =$ " X ne suit pas la loi continue de densité f ".

L'idée consiste alors à

- **Introduire une partition** de l'ensemble E en k **intervalles**

$$E = \cup_{j=1}^{k} e_j \text{ avec } e_j =]a_{j-1}, a_j]$$

- **Se ramener au cas discret** avec
  1) la loi
$$p_j = P(X \in e_j) = \int_{a_{j-1}}^{a_j} f(x)dx.$$
  2) La variable aléatoire N_j correspondant au nombre de variables X_i dans l'intervalle e_j.

Exercice 11

Nous désirons tester l'équiprobabilité en France des naissances de filles et de garçons dans les familles de 6 enfants. Pour cela, les données correspondant à un échantillon de $n = 1000$ familles ont été collectées et l'observation du nombre de garçons a donné :

0	1	2	3	4	5	6
17	90	231	278	255	104	24

On admet que les naissances des filles et des garçons sont équiprobables. Peut-on conclure de cette étude que les sexes des différents enfants d'une même famille sont indépendants ?

Exercice 12

Une compagnie ferroviaire cherche à analyser la rentabilité de ses lignes locales. Elle étudie donc X_i, le nombre moyen de passagers par jour de $i = 1, ..., n$ lignes de train.

nb moyen de voya.	100 – 150	150 – 200	200 – 250	250 – 300	300 – 350	350 – 400	400 – 450	450 – 500
nb de lignes	462	192	141	72	56	39	22	16

On souhaite tester une modélisation avec une loi de Pareto $\mathcal{P}(a, r)$ avec $a > r = 100$ de densité

$$f(x) = \begin{cases} \frac{ar^a}{x^{1+a}} \text{ si } x > r \\ 0 \text{ sinon} \end{cases}$$

1) *En supposant que $r = 100$, appliquer la méthode des moments pour trouver l'estimateur $\overline{A}_n^1 = \frac{\overline{X}_n}{\overline{X}_n - r}$ du paramètre a.*
2) *Réaliser le test d'adéquation $H_0 =$ " X_i suit une loi de Pareto $\mathcal{P}(2, 100)$" contre $H_1 =$ " X_i ne suit pas une loi de Pareto $\mathcal{P}(2, 100)$"*

1. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison

2. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Motivation

Nous souhaitons maintenant comparer un paramètre non pas à une valeur donnée mais à **une valeur inconnue** que nous avons estimée. Par exemple, nous souhaitons comparer l'effet de deux traitements médicaux sur des patients

Nous nous focalisons par la suite uniquement dans le cas **d'échantillons Gaussiens indépendants**:

- Soient $X_1, \cdots X_{n_X}$ i.i.d. selon une loi $\mathcal{N}(\mu_X, \sigma_X^2)$
- Soient $Y_1, \cdots Y_{n_Y}$ i.i.d selon une loi $\mathcal{N}(\mu_Y, \sigma_Y^2)$.

Nous supposons que les (X_i), (Y_j) sont également mutuellement indépendants. L'idée est alors de faire un test d'égalité :

$$H_0 = "\sigma_X^2 = \sigma_Y^2 \text{ et } \mu_X = \mu_Y"$$

Loi de Fisher-Snedecor

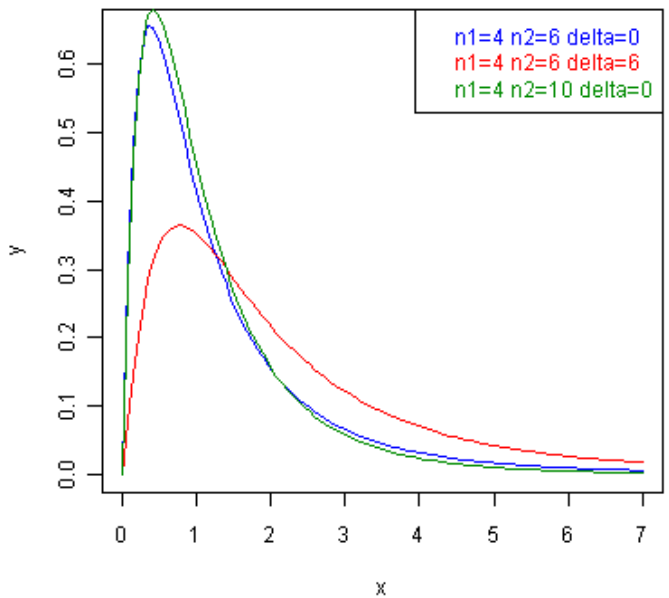
Définition 12

Si $X \sim \chi^2_n$ *et* $Y \sim \chi^2_m$*, et si* X *et* Y *sont indépendantes, alors* $F_{n,m} = \frac{X/n}{Y/m} \sim \mathcal{F}(n,m)$ *suit une loi de Fisher de paramètres* n *et* m*.*

- **Quantile de la loi** $\mathcal{F}(n,m)$: $f_{n,m,\alpha}$ est défini par

$$P(F_{n,m} \leq f_{n,m,\alpha}) = \alpha$$

- **Hypothèse de symétrie:** $f_{n,m,\alpha} = f^{-1}_{m,n,1-\alpha}$



Test d'égalité des variances

- **Hypothèse** : $H_0 = "\sigma_X^2 = \sigma_Y^2"$ contre $H_1 = "\sigma_X^2 \neq \sigma_Y^2"$
- **Théorème de Fisher**

$$\frac{(n_X - 1)S'^2_{n_X}}{\sigma_X^2} \sim \chi^2_{n_X - 1} \quad \text{et} \quad \frac{(n_Y - 1)S'^2_{n_Y}}{\sigma_Y^2} \sim \chi^2_{n_Y - 1}$$

- **Statistique de test** sous l'hypothèse $H_0 = "\sigma_X^2 = \sigma_Y^2"$

$$F = S'^2_X / S'^2_Y \sim \mathcal{F}(n_X - 1, n_Y - 1).$$

- **Zone d'acceptation** de H_0 pour S'^2_X / S'^2_Y

$$I = [f_{n_X - 1, n_Y - 1, \alpha/2}, f_{n_X - 1, n_Y - 1, 1 - \alpha/2}]$$

Comparaison des espérances avec σ_X^2 et σ_Y^2 connues

- **Hypothèses:** $H_0 = "\mu_X = \mu_Y"$ contre $H_1 = "\mu_X \neq \mu_Y"$
- **Estimateurs** : $\overline{X}_{n_X} \sim \mathcal{N}(\mu_X, \sigma_X^2/n_X)$ et $\overline{Y}_{n_Y} \sim \mathcal{N}(\mu_Y, \sigma_Y^2/n_Y)$ avec

$$\overline{X}_{n_X} - \overline{X}_{n_Y} \sim \mathcal{N}(\mu_X - \mu_Y, \sigma_X^2/n_X + \sigma_Y^2/n_Y)$$

- **Statistique de test** sous l'hypothèse $H_0 = "\sigma_X^2 = \sigma_Y^2"$

$$U = \frac{\overline{X}_{n_X} - \overline{X}_{n_Y}}{\sqrt{\sigma_X^2/n_X + \sigma_Y^2/n_Y}} \sim \mathcal{N}(0,1).$$

- **Zone d'acceptation** de H_0 pour U

$$I = [u_{\alpha/2}, u_{1-\alpha/2}]$$

Test de Student : cas $\sigma_X^2 = \sigma_Y^2$ inconnues

L'objectif est d'utiliser les estimateurs $S'^2_{n_X,X}$ et $S'^2_{n_Y,Y}$ de σ_X^2 et σ_Y^2:

- **Théorème de Fisher**

$$Z = \frac{(n_X - 1)S'^2_{n_X,X}}{\sigma_X^2} + \frac{(n_Y - 1)S'^2_{n_Y,Y}}{\sigma_Y^2} \sim \chi^2_{n_X+n_Y-2}$$

et

$$T = \sqrt{n_X + n_Y - 2}\frac{U}{\sqrt{Z}} \sim St(n_X + n_Y - 2)$$

- **Statistique de test:** sous l'hypothèse $H_0 = "\sigma_X^2 = \sigma_Y^2"$

$$T = \frac{\overline{X}_{n_X} - \overline{X}_{n_Y}}{S'_{X,Y}\sqrt{n_X^{-1} + n_Y^{-1}}} \text{ avec } S'^2_{X,Y} = \frac{(n_X - 1)S'^2_{n_X,X} + (n_Y - 1)S'^2_{n_Y,Y}}{n_X + n_Y - 2}$$

- **Zone d'acceptation** de H_0 pour T

$$I = [t_{n_X+n_Y-2,\alpha/2}, t_{n_X+n_Y-2,1-\alpha/2}]$$

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON
Pole Math

Exercice 13

Une expérience a été conduite dont le but est d'évaluer l'effet de l'inoculation d'un mycorrhize sur la croissance en hauteur de plantules de Pinus kesiya. Dans l'expérience, 10 *plantules, formant le Groupe 1, ont été inoculées, et* 10 *autres (Groupe 2) ont été laissées telles quelles. Le tableau ci-dessous donne les hauteurs obtenues dans les deux groupes de plantules.*

Parcelles	*Groupe 1*	*Groupe 2*
1	23.0	8.5
2	17.4	9.6
3	17.0	7.7
4	20.5	10.1
5	20.5	10.1
6	24.0	13.2
7	22.5	10.3
8	22.7	9.1
9	19.4	10.5
10	18.8	7.4

Vérifier si l'inoculation améliore significativement la croissance en hauteur de plantules.

1. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison

2. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Loi de Bernoulli (L.D)

Définition 13

On considère une expérience qui se traduit par un ***succès*** *avec une probabilité p ou par un* ***échec*** *avec la probabilité* $1 - p$. *On note X la v.a. définie par :*

$$X = \begin{cases} 1 & \text{si c'est un succès} \\ 0 & \text{si c'est un échec} \end{cases}$$

On dit que X suit une ***loi de Bernoulli*** *de paramètre p et on note :*

$$X \sim \mathcal{B}(p).$$

Sa ***loi de probabilité*** *est définie par :*

$$P(X = 0) = 1 - p \text{ et } P(X = 1) = p.$$

Propriété 5

L'espérance *et la* ***variance*** *de* $X \sim \mathcal{B}(p)$ *sont données par :*

$$\mathbb{E}(X) = p \text{ et } \mathbb{V}(X) = p(1 - p).$$

Loi Binomiale (L.D)

Définition 14

On répète de façon indépendante n fois une expérience de Bernoulli dont la probabilité de succès est p. La loi de la v.a. X qui compte le ***nombre de succès*** *sur ces n expériences indépendantes est appelée* ***loi Binomiale*** *de paramètres n et p, ce que l'on note :*

$$X \sim \mathcal{B}in(n, p).$$

Elle prend ses valeurs dans $\{0, 1, \ldots, n\}$ *et sa* ***loi de probabilité*** *est définie par :*

$$\forall k \in \{0, 1, \ldots, n\}, \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Propriété 6

L'espérance *et la* ***variance*** *de* $X \sim \mathcal{B}in(n, p)$ *sont données par :*

$$\mathbb{E}(X) = np \text{ et } \mathbb{V}(X) = np(1-p).$$

Loi de Poisson (L.D)

Définition 15

*La **loi de Poisson** de paramètre λ apparaît comme la limite lorsque $n \to +\infty$, de la loi binomiale $\mathcal{B}in(n, \lambda/n)$, où λ est un réel strictement positif. Une v.a. X suivant la loi de Poisson prend ses valeurs dans $\mathbb{N}$. On dit que X suit une **loi de Poisson** de paramètre λ et l'on note :*

$$X \sim \mathcal{P}(\lambda).$$

*Sa **loi de probabilité** est définie par :*

$$\forall k \in \mathbb{N}, \quad P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

.

Propriété 7

***L'espérance** et la **variance** de $X \sim \mathcal{P}(\lambda)$ sont données par :*

$$\mathbb{E}(X) = \lambda \text{ et } \mathbb{V}(X) = \lambda.$$

Loi géométrique (L.D)

Définition 16

*Une v.a. X qui suit une **loi géométrique** de paramètre p prend ses valeurs dans $\mathbb{N}^* = \{1, 2, \ldots, n, \ldots\}$ et on note :*

$$X \sim \mathcal{G}(p).$$

*Sa **loi de probabilité** est donnée par :*

$$\forall k \in \mathbb{N}, \quad P(X = k) = p(1-p)^{k-1}.$$

C'est la loi du temps d'attente du premier succès dans une succession d'épreuves de Bernoulli indépendantes.

Propriété 8

***L'espérance** et la **variance** de $X \sim \mathcal{G}(p)$ valent respectivement*

$$\mathbb{E}(X) = \frac{1}{p} \text{ et } \mathbb{V}(X) = \frac{1-p}{p^2}.$$

1. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison

2. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Loi uniforme (L.C)

Définition 17

Une v.a. X qui suit la ***loi uniforme*** *sur l'intervalle $I = [a, b]$, notée*

$$X \sim U[a, b],$$

prend ses valeurs dans tout l'intervalle I et sa loi de probabilité est définie par la ***densité de probabilité*** *:*

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

C'est la loi exacte des phénomènes continus uniformément répartis sur un intervalle.

Propriété 9

Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = \frac{a+b}{2} \text{ et } \mathbb{V}(X) = \frac{(a-b)^2}{12}.$$

Loi exponentielle (L.C)

Définition 18

Soit $\lambda \in \mathbb{R}_+$, une v.a. X qui suit la ***loi exponentielle*** *de paramètre λ, notée*

$$X \sim \mathcal{E}(\lambda),$$

est une v.a. continue à valeurs dans $\mathbb{R}_+$. Sa ***densité de probabilité*** *est donnée par :*

$$f(x) = \begin{cases} 0 & si\ x < 0 \\ \lambda e^{-\lambda x} & si\ x \geq 0 \end{cases}$$

C'est la loi du temps d'attente d'un phénomène poissonien de taux λ : temps d'attente du premier évènement ou intervalle de temps entre deux évènements consécutifs, dans certaines situations on parle de durée de vie.

Propriété 10

Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = \frac{1}{\lambda} \text{ et } \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

Loi de Erlang

Définition 19

Soit $k \in \mathbb{N}^*$ et $\lambda \in \mathbb{R}_+$. Une v.a. X qui suit la ***loi de Rayleigh*** *de paramètres λ, notée*

$$X \sim Erl(k, \lambda),$$

est une v.a. continue à valeurs dans $\mathbb{R}_+$. Sa ***densité de probabilité*** *est donnée par :*

$$f(x) = \frac{\lambda^k x^{k-1} \exp(-\lambda x)}{(k-1)!} \text{ pour } x \geq 0.$$

Propriété 11

La distribution d'Erlang est la distribution de la somme de k variables aléatoires indépendantes et identiquement distribuées selon une loi exponentielle de paramètre λ Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = k/\lambda \text{ et } \mathbb{V}(X) = k\lambda^2$$

Loi normale ou loi Gaussienne (L.C)

Définition 20

Soit $\mu \in \mathbb{R}, \sigma \in \mathbb{R}_+$. Une v.a. X qui suit la ***loi normale*** *de paramètres μ et σ, notée*

$$X \sim \mathcal{N}(\mu, \sigma),$$

est une v.a. continue à valeurs dans $\mathbb{R}$. Sa ***densité de probabilité*** *est donnée par :*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Propriété 12

Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = \mu \text{ et } \mathbb{V}(X) = \sigma^2.$$

De très nombreuses distributions naturelles sont approximées par la loi normale.

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON
Pole Math

Loi de Rayleigh

Définition 21

Soit $\sigma \in \mathbb{R}_+$. Une v.a. X qui suit la ***loi de Rayleigh*** *de paramètres σ, notée*

$$X \sim \mathcal{R}(\sigma),$$

est une v.a. continue à valeurs dans $\mathbb{R}_+$. Sa ***densité de probabilité*** *est donnée par :*

$$f(x) = \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right) \text{ pour } x \geq 0.$$

Propriété 13

Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = \sigma\sqrt{\pi/2} \text{ et } \mathbb{V}(X) = \frac{4-\pi}{2}\sigma^2$$

Loi de Weibull

Définition 22

Soient $\beta > 0$ et $\Sigma > 0$. Une v.a. X qui suit la ***loi de Weibull*** *de paramètres β et σ, notée*

$$X \sim Weib(\beta, \sigma),$$

est une v.a. continue à valeurs dans $\mathbb{R}_+$. Sa ***densité de probabilité*** *est donnée par :*

$$f(x) = \frac{\beta}{\sigma}\left(\frac{x}{\sigma}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\sigma}\right)^{\beta}\right) \text{ pour } x \geq 0.$$

Propriété 14

Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = \sigma\Gamma(1 + 1/\beta) \text{ et } \mathbb{V}(X) = \sigma^2(\Gamma(1 + 2/\beta) - \Gamma(1 + 1/\beta)^2)$$