Échantillons Aléatoires et lois d'échantillonnage

Wissem Maazoun

École Polytechnique de Montréal,
Département de Mathématiques et de génie industriel

Introduction

L'inférence statistique sur les paramètres (moyenne, variance, etc.) d'une variable X définie dans une population se fait à l'aide d'un échantillon de la population étudiée. Un tel échantillon doit être tiré au hasard (de la population) pour être représentatif de celle-ci. On parle alors d'échantillon aléatoire.

Échantillons aléatoires

Soit X une v.a. (i.e. une population). Un échantillon aléatoire de taille n de la v.a. X est une suite de n v.a. $X_1, X_2, \ldots, X_n$ indépendantes et identiquement distribuées (i.i.d.) selon la loi de probabilité de X. lorsque les données $x_1, \ldots, x_n$ sont obtenues, on dit qu'on a une réalisation de l'échantillon.

Définition

Soit $X_1, \ldots, X_n$ un échantillon aléatoire d'une v.a. X. Toute fonction $h(X_1, \ldots, X_n)$ qui ne dépend que de $X_1, X_2, \ldots, X_n$ constitue une statistique.

1. La moyenne $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ est une statistique.
2. La variance $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ est une statistique.
3. La médiane de l'échantillon est une statistique.

Une statistique $h(X_1, X_2, \ldots, X_n)$ étant une fonction de v.a. est aussi une v.a. Sa distribution est qualifiée de distribution échantillonnale. Chaque statistique possède une distribution échantillonnale ; cette distribution dépend de celle de la v.a. X, et de la taille n de l'échantillon.

Exemple

Soit une population de 3 unités pour lesquelles X prend les valeurs $0, 1, 2$. Si on extrait de cette population un échantillon aléatoire de taille 2 (avec remise), déterminons la distribution échantillonnale de $\bar{X}$.

Remarque

En pratique les populations étudiées sont nombreuses, voire infinies ; le nombre d'échantillons possibles est également très grand, voire infini. La distribution échantillonnale de la moyenne $\bar{X}$ est donc en général de type continu (loi normale).

Théorème

Soit X une v.a. de moyenne $\mu = E(X)$ et de variance $\sigma^2 = V(X)$, et soit $X_1, \ldots, X_n$ un échantillon aléatoire de taille n de la v.a. X. Si la v.a. X est de loi normale ou si la taile de l'échantillon est grande, alors

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \text{i.e.} \quad \bar{X} \sim N(\mu, \frac{\sigma^2}{n}).$$

Remarque

D'autres lois de probabilité permettent de caractériser les distributions échantillonnage de certaines statistiques usuelles.

1. La loi du Khi-deux (χ^2).
2. La loi T de Student.
3. La loi F de Fisher.

Définition

Soient $Z_1, Z_2, \ldots, Z_k$ des v.a. i.i.d de loi $N(0, 1)$. La v.a. définie par

$$\chi_k^2 = Z_1^2 + Z_2^2 + \ldots + Z_k^2$$

est dite de loi khi-deux à k degrés de liberté. Sa densité celle d'une Gamma avec $r = k/2$ et $\lambda = 1/2$ est

$$f(x) = \begin{cases} \frac{x^{k/2-1}}{2^{k/2}\Gamma(k/2)} e^{-x/2} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

On peut montrer (propriétés de loi Gamma) que $E(\chi_k^2) = k$ et $V(\chi_k^2) = 2k$.
Pour une valeur α, telle que $0 \leq \alpha \leq 1$ on définit le centile $\chi_{\alpha;k}^2$ par la relation

$$P(\chi_k^2 > \chi_{\alpha;k}^2) = \alpha.$$

Échantillons Aléatoires et lois d'échantillonnage 6 / 15
La loi du khi-deux (ou khi-carré)

Wissem Maazoun Échantillons Aléatoires et lois d'échantillonnage 6 / 15

Exemples

$P(\chi^2_{10} > 18,31) = 0,05$ i.e. $\chi^2_{0,05;10} = 18,31$.
$P(\chi^2_4 \leq 0,48) = 1 - P(\chi^2_4 > 0,48) = 1 - 0,975 = 0,025$.

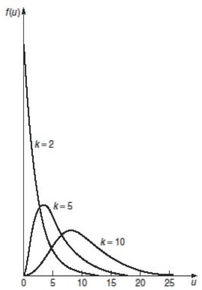

Figure 9.1 La fonction de densité de différentes lois du khi-carré.

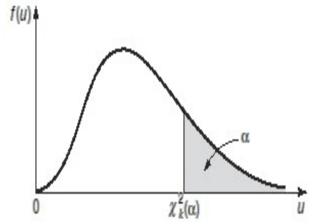

Figure 9.2 Le centile $\chi^2_k(\alpha)$ d'une loi du khi-carré.

Théorème : Additivité de la loi khi-deux

Soient $\chi^2_{k_1}, \chi^2_{k_2}, \ldots, \chi^2_{k_p}$ des v.a indépendantes distribuées selon des lois Khi-deux, alors la v.a $Y = \chi^2_{k_1} + \chi^2_{k_2} + \ldots + \chi^2_{k_p} = \sum_{i=1}^{p} \chi^2_{k_i}$ suit une loi χ^2_k avec $k = k_1 + k_2 + \ldots + k_p$ degrés de liberté.

Exemple

Soient X_1 et X_2 deux v.a indépendantes telles que $X_1 \sim \chi^2_2$ et $X_2 \sim \chi^2_1$. Calculer $P(X_1 + X_2 \leq 7,81)$.

Théorème : Utilisation en inférence statistique

Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire d'une v.a X de loi $N(\mu, \sigma^2)$. Alors

$$(n-1)\frac{S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1},$$

$$\text{où } S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}.$$

Exemple

Soit $X_1, \ldots, X_{10}$ un échantillon aléatoire d'une v.a X de loi $N(\mu, 4)$. Calculer $P(S^2 > 7,52)$.

Échantillons Aléatoires et lois d'échantillonnage 9 / 15
└La loi du khi-deux (ou khi-carré)

Wissem Maazoun Échantillons Aléatoires et lois d'échantillonnage 9 / 15

Définition

Soit Z une v.a de loi $N(0,1)$, et soit χ^2_k une v.a de loi khi-deux à k degré de liberté, telle que Z et χ^2_k sont indépendantes. La v.a $T_k = \frac{Z}{\sqrt{\chi^2_k/k}}$ est dite de loi T_k de Student avec k degré de liberté.
Sa fonction de densité est

$$f(t) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi k}\Gamma(k/2)} \frac{1}{(\frac{t^2}{k}+1)^{\frac{k+1}{2}}}, \qquad -\infty < t < \infty$$

On peut montrer que $E(T_k) = 0$ et $V(T_k) = \frac{k}{k-2}$ si $k > 2$.
Pour une valeur α, telle que $0 \leq \alpha \leq 1$ on définit le centile $t_{\alpha;k}$ par la relation $P(T_k > t_{\alpha;k}) = \alpha$.
La distribution d'une v.a de loi T_k est symétrique par rapport à sa moyenne 0. On a donc $-t_{\alpha;k} = t_{1-\alpha;k}$.

Exemples

1. $t_{0,25;4} = 0,741$ i.e $P(T_4 > 0,741) = 0,25$.
2. On a aussi $t_{0,75;4} = -t_{0,25;4} = -0,741$.

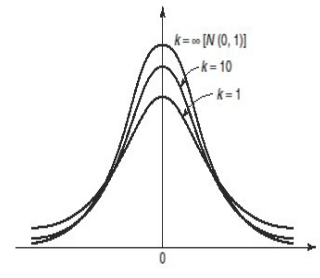

Figure 9.3 La fonction de densité de différentes lois de t.

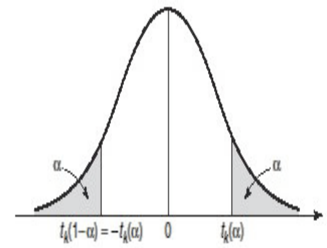

Figure 9.4 Les centiles de la loi de t.

Théorème : Utilisation en inférence statistique

Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire d'une v.a X de loi $N(\mu, \sigma^2)$. Alors

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim T_{n-1},$$

$$\text{où } S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad \bar{X} = \frac{\sum_{i=1}^n X_i}{n},$$

de plus $\bar{X}$ et S^2 sont des v.a indépendantes.

Exemple

Soit X_1, X_2, X_3, X_4 un échantillon aléatoire d'une v.a X de loi $N(0, \sigma^2)$. Calculer $P(\frac{\bar{X}}{S} > 0,819)$.

Échantillons Aléatoires et lois d'échantillonnage 12 / 15
La loi T de Student

Wissem Maazoun Échantillons Aléatoires et lois d'échantillonnage 12 / 15

Définition

Soit χ^2_u et χ^2_v deux v.a indépendantes de lois khi-deux à u et v degrés de liberté respectivement. Alors, la v.a $F = \frac{\chi^2_u/u}{\chi^2_v/v}$ est dite de loi F de Fisher à u et v degré de liberté. On la note $F_{u;v}$. Sa fonction de densité est

$$f(x) = \begin{cases} \frac{\Gamma(\frac{u+v}{2})}{\Gamma(u/2)\Gamma(v/2)} \frac{\left(\frac{u}{v}\right)^{u/2} x^{u/2-1}}{\left[\left(\frac{u}{v}\right)x+1\right]^{\frac{u+v}{2}}} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Il est démontré que $E(F_{u;v}) = \frac{v}{v-2}$ si $v > 2$ et
$V(F_{u;v}) = \frac{2v^2(u+v-2)}{u(v-2)^2(v-4)}$ si $v > 4$.
Pour une valeur α, telle que $0 \leq \alpha \leq 1$, on définit le centile $f_{\alpha;u,v}$ par la relation

$$P(F_{u;v} > f_{\alpha;u,v}) = \alpha.$$

Exemples

$f_{0,05;7,9} = 3,29$ i.e $P(F_{7;9} > 3,29) = 0,05$.

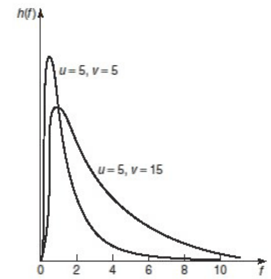

Figure 9.5 La fonction de densité de deux lois de F.

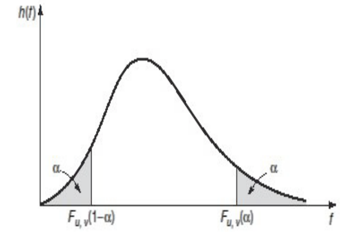

Figure 9.6 Les centiles inférieurs et supérieurs d'une loi de F.

Théorème : Utilisation en inférence statistique

Soient $X_{11}, X_{12}, \ldots, X_{1n_1}$ et $X_{21}, X_{22}, \ldots, X_{2n_2}$ deux échantillons aléatoires indépendants provenant de deux populations (ou v.a X_1 et X_2) de lois $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$. On pose

$$\bar{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i} \quad ; \quad S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)^2$$

$$\bar{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i} \quad ; \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(X_{2i} - \bar{X}_2)^2,$$

alors la v.a. $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1;n_2-1}$.