Ecole Santé Sciences
Probas & Stats

Hadrien Lorenzo, hadrien.lorenzo@u-bordeaux.fr

Université de Bordeaux

2018-2019

Note aux étudiants

Pourrez-vous amener vos ordis pour les cours des 14, 15 et 16 novembre ?

Les probabilités

Raisonnement statistique :

- Formulation d'une **hypothèse**,
- Délimitation d'un **échantillon** : ensemble d'individus avec des caractéristiques communes et d'autres différentes représentant la population générale et permettant donc de tester l'hypothèse formulée,
- Test de l'hypothèse : calcul de **statistiques** décrivant l'échantillon.

Il faut un **modèle statistique** pour comparer les données de l'échantillon. Une variation **significative** des statistiques décrivant l'échantillon permettent d'assurer que l'hypothèse est **vraisemblable**. Ou pas. La conclusion serait la même, avec un seuil d'assurance suffisant, sur un nouvel échantillon.

Vocabulaire

- **Expérience aléatoire** : Expérience dont on ne connaît pas le résultat,
- **Évènement** : Différents résultats accessibles via une expérience aléatoire,
- Ω **: Ensemble des évènements** : Ensemble des évènements possibles,

Mesure de probabilité

Mesure de probabilité

Une mesure de probabilité, ou *probabilité*, notée $\mathbb{P}$ est une application $\mathbb{P} : \mathcal{F} \to [0, 1]$ telle que

- $\mathcal{F}$ soit une famille de sous-ensembles de Ω telle que $\Omega \in \mathcal{F}$, $\forall A \in \mathcal{F},\ A^c \in \mathcal{F}$ l'évènement complémentaire à A,
- $\mathbb{P}(\Omega) = 1$ et $\mathbb{P}(\emptyset) = 0$,
- $\forall (A_i)_{i=1..n} \in \mathcal{F}^n$, **incompatibles** ou **disjoints** (= *leur réalisation simultanée est impossible*)

$$\mathbb{P}(\bigcup_{i=0}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

Exemples

- Soit Ω l'ensemble des valeurs prenables par un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$, on peut définir $\mathcal{F}_\Omega = \mathcal{P}(\Omega)$, l'ensemble des **parties** de Ω :

$$\mathcal{F}_\Omega = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \cdots, \{1, 5\}, \{3, 4, 6\}, \cdots, \Omega\},$$

  de cardinalité $2^6 = 64$.
- Soit $\Omega = \mathbb{R}$, $\mathcal{F}$ est l'ensemble contenant tous les intervalles de la forme $]a, b]$ avec $a < b$, appelée *tribu des Boréliens notée* $\mathcal{B}$, avec $F : \mathbb{R} \to [0, 1]$ une fonction de répartition et $\mathbb{P} : \mathcal{B} \to [0, 1]$ telle que $\forall]a, b] \in \mathcal{B}$

$$\mathbb{P}(]a, b]) = F(b) - F(a),$$

  est une mesure de probabilité.

Probabilité

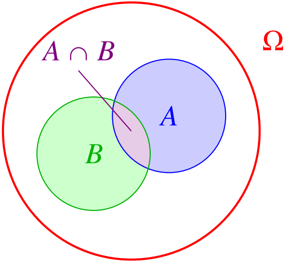
La **probabilité** d'un évènement est un réel positif inférieur à 1 qui représente la vraisemblance de cet évènement. $\forall \omega \in \Omega$, on note $\mathbb{P}(\omega) \in [0, 1]$ ce nombre et $\mathbb{P}(\omega^c) = 1 - \mathbb{P}(\omega)$, la probabilité de l'évènement **complémentaire** à ω.

Exemple : Le dé équilibré : $\Omega = \{1, 2, 3, 4, 5, 6\}$

$\mathbb{P}(\Omega) = 1,\ \mathbb{P}(\{1\}) = \frac{1}{6},\ \mathbb{P}(\{1, 2\}) = \frac{1}{3},\ \mathbb{P}(\{\text{pairs}\}) = \mathbb{P}(\{2, 4, 6\}) = \frac{1}{2},\ \mathbb{P}(\{\text{non multiples de } 3\}) = \mathbb{P}(\{3, 6\}^c) = \frac{2}{3}$.

Penser en patates

$$\forall (A, B) \in \Omega^2,\ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$



Incompatibilité

Deux évènements sont **incompatibles ssi** au plus un de ces deux évènements peut arriver simultanément. Généralisation à une famille quelconque d'évènements.

Conséquence directe :

La probabilité d'une **famille d'évènements incompatibles** est égale à la somme de leurs probabilités.

$$(\omega_i)_{i\in[\![1,n]\!]}, \text{ incompatibles, alors } \mathbb{P}\left(\bigcup_{i=0}^{n}\omega_i\right) = \sum_{i=1}^{n}\mathbb{P}(\omega_i)$$

Exemple

$\forall \omega \in \Omega,\ \mathbb{P}(\omega^c) = 1 - \mathbb{P}(\omega)$,

Si $\Omega = \{\omega_1, \cdots, \omega_n\} |\ \forall (i,j) \in [\![1,n]\!]^2, \mathbb{P}(\omega_i) = \mathbb{P}(\omega_j)$. Alors si tous les évènements sont incompatibles, leur probabilité est $\mathbb{P}(\omega_i) = 1/n$.

Composition d'évènements

Multiplication de résultats

Soient deux expériences aléatoires alors l'expérience aléatoire du produit des expériences a un nombre de possibilités égal au produit des nombres de possibilités de chaque expérience.

Tirages successifs avec remise

On réitère N fois la même expérience aléatoire à k états possibles en remettant en jeu à chaque fois l'évènement tiré. Alors le nombre total d'arrangements est k^N.

Exemple : Cadenas à 4 positions de 10 chiffres : 10^4 possibilités. *Quid* de si l'on connaît les chiffres mais pas les positions ?

Dénombrement

Permutations

Il existe $n!$, lire "*factoriel n*", façons différentes d'ordonner n objets. $n! = 1 * 2 * 3 * ... * (n-1) * n$

Tirages successifs sans remise (arrangements)

Il existe A_n^p, lire "*a n p*", façons différentes d'ordonner n objets en p groupes. $A_n^p = n * (n-1) * ... * (n-(p-1)) = \frac{n!}{(n-p)!}$

Tirages simultannés

Il existe $\binom{n}{p}$, lire "*p parmi n*", façons différentes d'ordonner n objets en p groupes sans intérêt pour l'ordre. $\binom{n}{p} = \frac{n!}{p!(n-p)!}$

Probabilité conditionnelle

Loi de probabilité conditionnelle

$\forall (A, B) \in \Omega^2 | \ \mathbb{A} \neq 0$, alors la probabilité conditionnelle de B sachant A est

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

Loi de Bayes

Loi de Bayes

$\forall (A, B) \in \Omega^2$

$$\mathbb{P}(B \cap A) = \mathbb{P}(B|A)\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B)$$

Indépendance d'évènements

$\forall (A, B) \in \Omega^2$, A et B sont dits indépendants **ssi**

$$\mathbb{P}(B \cap A) = \mathbb{P}(B)\mathbb{P}(A),$$

et par la loi de Bayes il vient

$$\mathbb{P}(B|A) = \mathbb{P}(B),\ \mathbb{P}(A|B) = \mathbb{P}(A)$$

Évènements indépendants, exemples

On tire un dé équilibré. Dire si les évènements suivants sont **indépendants**

1. $A\{$Le résultat est pair$\}$ et
   $B\{$Le résultat supérieur ou égal à 5$\}$
2. $A\{$Le résultat est pair$\}$ et
   $B\{$Le résultat supérieur ou égal à 4$\}$

Dire maintenant s'ils sont **incompatibles**.

Loi des causes totales

Loi des causes totales

Soit une famille d'évènements $(C_i)_{i=1..n}$ incompatibles deux à deux et vérifiant $\Omega = \bigcup_{i=0}^{n} C_i$, alors

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap C_i) = \sum_{i=1}^{n} \mathbb{P}(A|C_i)\mathbb{P}(C_i),$$

où la deuxième égalité est une implication de la loi de Bayes.

Exemple

Maladie à 3 formes lié à un gène

Une maladie a 3 formes, par gravité croissante : $B_1,\ B_2,\ B_3$. Le gène A a deux variations, protecteur ou pas, notés A_1 et A_2. $3/4$ des gens présentent A_1 et les autres présentent A_2. On sait aussi que parmi les gens présentant A_1 90% sont légèrement malades (B_1) alors que 10% sont sérieusement malades (B_2).
On sait aussi qu'un individu ne présentant pas le gène sous la forme A_1 a autant de chance de décéder (B_3) que d'être sérieusement malade (B_2) mais ne peut pas être légèrement malade (B_1).
Quelle est la probabilité qu'une personne rencontrée et ayant la forme sérieuse mais sans pour autant mortelle de la maladie présente la forme protectrice du gène ?

Variable aléatoire

Définition

Une variable aléatoire est une application de $\Omega \to \mathbb{R}$. Si les valeurs prises par cette variable sont finies ou dénombrables, on dit que c'est une **variable aléatoire discrère**.

loi d'une variable aléatoire discrète

Soit une variable aléatoire discrète X.
La loi de X est l'ensemble des probabilités associées aux valeurs de $X(\Omega) = \{x_i\}_{i=1..n}$, on note alors

$$\forall i = 1..n,\ p_i = \mathbb{P}(X = x_i),$$

avec $\sum_{i}^{n} p_i = 1$ car $\mathbb{P}$ est une mesure de probabilité sur Ω.

Exemple

3 tirages d'une pièce équilibrée

On tire successivement 3 fois une pièce équilibrée. A chaque fois que l'on tombe sur face on gagne s, un pile fait inversement perdre s. On note X la somme totale en fin d'expérience. Décrire la loi de X via un tableau.

Espérance

Définition dans le cas fini et dénombrable

Soit une variable aléatoire X avec pour valeurs $\{x_i\}_{i=1..n}$ et pour probabilités associées $\{p_i\}_{i=1..n}$. On définit alors l'espérance mathématique de X, notée $\mathbb{E}(X)$, via

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i p_i$$

L'espérance mathématique d'une variable aléatoire est une application linéaire. Donc $\forall(\lambda, \mu) \in \mathbb{R}^2$ et X et Y deux variables aléatoires, on a $\mathbb{E}(\lambda X + \mu Y) = \lambda\mathbb{E}(X) + \mu\mathbb{E}(Y)$. De plus, l'espérance d'une constante est elle même : $\mathbb{E}(\lambda) = \lambda$.

Espérance

Dans les cas non finis mais

- **Dénombrables** : $\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p_i$
- **Indénombrables** : ... on verra plus tard

Variance

Définition

Soit une variable aléatoire X, on appelle variance de la variable X la grandeur

$$Var(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

- La variance est toujours positive ou nulle.
- On appelle **écart-type** de la variable X la grandeur $\sigma(X) = \sqrt{Var(X)}$

Variance, propriétés

$\forall(\lambda, \mu) \in \mathbb{R}^2$ et X et Y deux variables aléatoires, alors :

- $Var(\lambda X) = \lambda^2 Var(X)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$, où Cov est la **covariance** et vérifie

$$Cov(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big)$$

- La **corrélation** est définie par

$$\rho(X, Y) = \frac{Cov(X, Y)}{Var(X)Var(Y)},$$

il faut donc $Var(X)Var(Y) \neq 0$

Lois et distributions classiques

Parmi les expériences aléatoires observées et modélisées il y a certaines lois et distributions qu'il est bon de connaître.
Par connaître nous entendons : connaître l'espérance mathématique et la variance associée ainsi que les cas d'applications usuels.
Soit une loi $\mathcal{L}$ de paramètre θ, potentiellement vectoriel. Si X est une variable aléatoire qui se comporte comme la loi $\mathcal{L}$ de paramètre θ, on dit que X suit une loi $\mathcal{L}$ de paramètre θ et on note

$$X \sim \mathcal{L}(\theta).$$

Loi uniforme

Définition

Une variable aléatoire discrète sur $\{1, \cdots, n\}$ X suit une loi uniforme **ssi**

$$\forall i \in \{1, \cdots, n\}, \ \mathbb{P}(X = i) = 1/n.$$

- $\mathbb{E}(X) = (n+1)/2$
- $Var(X) = (n^2 - 1)/12$

Application à un dé équilibré à 6 faces.

Démonstrations

$$\mathbb{E}(X) = \sum_{i=1}^{n} p_i x_i = \sum_{i=1}^{n} \frac{1}{n} i = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n}\frac{1}{2} n(n+1) = \frac{n+1}{2}$$

$$\begin{aligned}
Var(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \text{ avec } \mathbb{E}(X)^2 = \frac{(n+1)^2}{4} \\
\mathbb{E}(X^2) &= \sum_{i=1}^{n} p_i x_i^2 = \sum_{i=1}^{n} \frac{1}{n} i^2 = \frac{1}{n}\frac{1}{6} n(n+1)(2n+1) = \frac{(n+1)(2n+1)}{6} \\
Var(X) &= \frac{n^2-1}{12}
\end{aligned}$$

Loi de Bernoulli

Définition

On dit que X suit une loi de Bernoulli si X n'a que deux issues possibles notées 1 et 0 et de probabilités :

$$\mathbb{P}(X = 1) = p, \ \mathbb{P}(X = 0) = q = 1 - p$$

- $\mathbb{E}(X) = p$
- $Var(X) = p(1 - p)$

C'est ce qui se passe lorsqu'une pièce est non équilibrée avec $p \neq 0.5$.

Démonstrations

A vous de faire

Loi binomiale

Définition

Une variable aléatoire X suit une loi Binomiale $\mathcal{B}(n,p)$ si elle s'écrit comme la somme de n variables de Bernoulli indépendantes et de même paramètre p. X est à valeurs dans $\{0,\cdots,n\}$. Elle compte donc le nombre de succès. On a alors

$$\forall i \in \{0,\cdots,n\},\ \mathbb{P}(X=i) = \binom{n}{i} p^i (1-p)^{n-i}$$

Pourquoi le $\begin{pmatrix} n \\ i \end{pmatrix}$? Par-ce que nous regardons un tirage simultanée à i succès parmi n tirages donc sans intérêt pour l'ordre.

Propriétés

- $\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np$, par linéarité de l'espérance mathématique.
- $Var(X) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = np(1-p)$, par indépendance des variables aléatoires.

Exercice

Démonter que la loi de Bernoulli est bien une loi de probabilité

Exercice

Les vaccins

Une entreprise pharmaceutique fabrique des vaccins dont 5% sont efficaces sur la population. On considère 100 vaccins et on se demande la probabilité que :

1. Le vaccin fonctionne sur tout le monde,
2. Le vaccin ne fonctionne pas sur au plus 2 personnes.

Le vaccin est-il efficace ?

Exercice, suite

X_i la variable aléatoire qui affiche 1 si le vaccin ne fonctionne pas. On lui associe naturellement une loi de Bernoulli de paramètre $p = 0.05$. On suppose les variables X_i indépendantes les unes des autres et on définit la variable aléatoire $X = \sum_{i=1}^{n} X_i$ qui suit donc une loi Binomiale de paramètres $n = 100$ et $p = 0.05$. Ainsi

1. $P(X = 0) = 1 * 0.05^0 * 0.95^{100} \approx 5.9 * 10^{-3}$
2. $P(X \leqslant 2) = 1 * 0.05^0 * 0.95^{100} + \binom{100}{1} * 0.05^1 * 0.95^{99} + \binom{100}{2} * 0.05^2 * 0.95^{98} \approx 0.12$

L'efficacité dépend du seuil fixé par les autorités de santé ☺

Loi de Poisson

Définition

Une variable aléatoire X suit une loi de Poisson $\mathcal{P}(\lambda)$ si elle prend ses valeurs dans l'ensemble des entiers naturels, $\mathbb{N}$. On a alors

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

- $\mathbb{E}(X) = \lambda$
- $Var(X) = \lambda$

Loi utilisée pour modéliser le nombre d'arrivées à un péage pour un intervalle de temps donné.
En premier approximation : le nombre d'ARN d'un gène produit par un groupe de cellules

Les lois continues

On passe au continue, c'est pareil sauf que les $\sum$ deviennent des $\int$ et la probabilité en un point est nul.
En vrai on va tout de même introduire deux trois objets mathématiques.

Fasten your seatbelts !

Nouvelle expérience

Concentration de magnésium dans une bouteille d'eau

- **Expérience aléatoire** : Ouvrir une bouteille d'eau et titrer la teneur en magnésium,
- **Variable aléatoire** X : La teneur en magnésium de la bouteille ouverte et titrée,

$\Omega \subset \mathbb{R}$, mais devra valoir $\approx$ la valeur affichée par la bouteille.

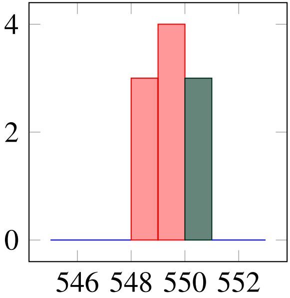
Comment visualiser X : l'histogramme

L'histogramme : En abscisse les valeurs prises par X et en ordonnée le nombre de fois que la valeur est prise.

Histogrammes

Expérience

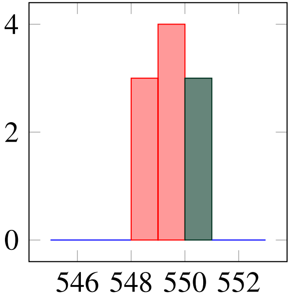
On teste ≈ 10 bouteilles.



Histogrammes

Expérience

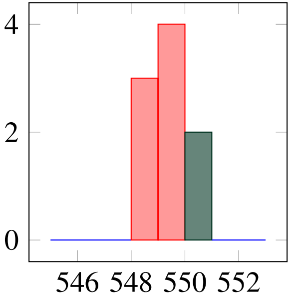
En rouge, les bouteilles pour lesquelles la valeur est comprise entre $548mg/L$ et $550mg/L$.
En vert, les bouteilles pour lesquelles la valeur est égale à $\approx 550mg/L$.



Histogrammes

Probabilités et histogrammes

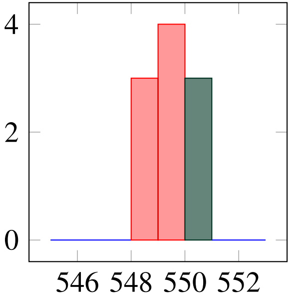
Le calcul de la probabilité revient à calculer l'aire sous la courbe histogramme.



Histogrammes

Probabilités et histogrammes

Pour un segment de valeurs : indépendant de la taille des **bins**.
Pour une valeur particulière : dépendant de la taille du **bin** considéré.

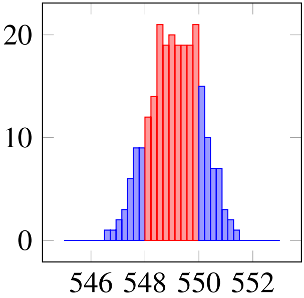


Histogrammes

Expérience

On teste plusieurs centaines de bouteilles.

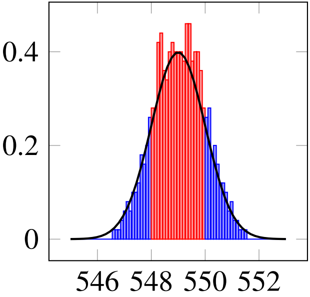
- La taille des bins tend vers 0 pour que le graphique reste joli.
- Les probabilités ponctuelles tendent vers 0.
- Les probabilités sur des segments de $\mathbb{R}$ se stabilisent.



Histogrammes

Modification de l'histogramme

Diviser le nombre de chaque barre par le nombre total de bouteilles testées.
L'enveloppe de l'histogramme nous rappelle quelque chose ? ! ?



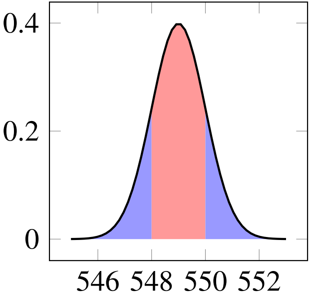
Densités de probabilité

Cas limite

En ouvrant une infinité de bouteilles on récupère une figure d'histogrammes avec des bins infiniment fins et une courbe que l'on peut associer à une courbe bien connue, la courbe de Gauss, de moyenne $\mu = 549$ et de variance $\sigma^2 = 1$

$$f_X(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

la **densité de probabilité.**



Densité de probabilité

Définition

Soit $(a, b) \in \mathbb{R}^2 | [a, b] \subset \Omega$, Ω associée à la variable aléatoire X. On dit que X est une variable aléatoire **à densité** lorsqu'il existe une fonction positive ou nulle, notée $f_X : \Omega \to \mathbb{R}^+$ telle que

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_X(x)dx,$$

c'est l'aire sous la courbe. On doit avoir $\mathbb{P}(\Omega) = \int_\Omega f_X(x)dx = 1$ par définition. Souvent $\Omega = \mathbb{R}$.

Application aux bouteilles

Énoncé

Sur les bouteilles, on veut connaître la probabilité qu'une bouteille ait une concentration en magnésium comprise entre $548mg/L$ et $550mg/L$. Quel calcul devons-nous poser ?

Application aux bouteilles

Énoncé

Sur les bouteilles, on veut connaître la probabilité qu'une bouteille ait une concentration en magnésium comprise entre $548mg/L$ et $550mg/L$. Quel calcul devons-nous poser ?

Solution

$$\begin{aligned} \mathbb{P}(548 \leqslant X \leqslant 550) &= \int_{548}^{550} f_X(x, \mu = 549, \sigma^2 = 1)dx \\ &= \int_{-1}^{1} f_X(x, \mu = 0, \sigma^2 = 1)dx \end{aligned}$$

C'est l'aire sous les rectangles à droite !

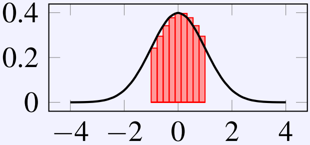
Application aux bouteilles

Version approchée

On fix N le nombre de bins dans l'intervalle $[548, 550]$, alors

$$\mathbb{P}(548 \leqslant X \leqslant 550) \quad \approx \frac{2}{N} \sum_{i=0}^{N-1} f_X(-1 + \frac{i}{N} * 2, \mu = 0, \sigma^2 = 1)$$

Ici un exemple avec $N = 9$.



Fonction de répartition

Définition

Soit X une variable aléatoire à densité de densité de probabilités f_X définie sur $\Omega \subset \mathbb{R}$, on appelle **fonction de répartition** la fonction F_X définie $\forall t \in \Omega$ par

$$F_X(t) = \mathbb{P}(X \leqslant t),$$

par positivité de f_X, F_X est croissante et positive. Elle représente la probabilité que X soit inférieur ou égal à t.

Espérance

Définition de l'espérance

Soit X une variable aléatoire à densité de densité de probabilités f_X définie sur $\Omega \subset \mathbb{R}$, on appelle **espérance** de X le réel

$$\mathbb{E}(X) = \int_{\Omega} x f_X(x) dx,$$

définie si l'intégrale précédente a un sens...

Cette espérance a les même propriétés de linéarité que celles définies dans le cas discret.

Variance

Variance

Soit X une variable aléatoire à densité de densité de probabilités f_X définie sur $\Omega \subset \mathbb{R}$, on appelle **espérance** de X le réel positif

$$Var(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right),$$

qui doit aussi avoir un sens...

Qui a les même propriétés de bi-linéarité que celles définies dans le cas discret.

Loi uniforme sur $[a,b]$

Définition

La loi uniforme sur $[a,b]$, notée $\mathcal{U}([a,b])$ est une loi à densité de fonction de densité

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{sinon} \end{cases},$$

naturellement la fonction de répartition s'écrit

$$F_X(x) = \begin{cases} \dfrac{x-a}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{si } x < a \\ 1 & \text{si } x > b \end{cases},$$

Loi uniforme

Calculs de l'espérance et de la variance

Loi uniforme

Calculs de l'espérance et de la variance

L'espérance

$$\begin{aligned}\mathbb{E}(X) &= \int_{\mathbb{R}} x f_X(x)dx = \int_a^b \frac{x}{b-a}dx \\ &= \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}(b^2-a^2)/2 \\ &= \frac{b+a}{2}\end{aligned}$$

Loi uniforme

Calculs de l'espérance et de la variance

La variance

$$\begin{aligned}
Var(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_{\mathbb{R}} x^2 f_X(x)dx - \mathbb{E}(X)^2 \\
&= \frac{1}{(b-a)}\int_a^b x^2 dx - \mathbb{E}(X)^2 = \frac{1}{(b-a)}(b^3 - a^3)/3 - \mathbb{E}(X)^2 \\
&= \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} - (\frac{b+a}{2})^2 \\
&= \frac{4b^2+4ab+4a^2}{12} - \frac{3b^2+6ab+3a^2}{12} = \frac{(b-a)^2}{12}
\end{aligned}$$

Loi Normale de paramètres μ et σ^2

Définition

La loi normale de paramètres μ et σ^2, notée $\mathcal{N}(\mu, \sigma^2)$, est une loi à densité de fonction de densité

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

sans formule simple de la fonction de répartition.

- $\mathbb{E}(X) = \mu$
- $Var(X) = \sigma^2$

Loi Normale de paramètres μ et σ^2

- $\approx 95\%$ de la masse est concentrée dans un rayon 2σ autour de μ

$$\mathbb{P}(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) \approx 0.95$$

.

- On a deja vu

$$\mathbb{P}(\mu - \sigma \leqslant X \leqslant \mu + \sigma) \approx 0.68$$

- C'est la loi la plus répandue : bruits des capteurs de poids, détecteurs d'ADN, répartition des photons dans un faisceau laser...

Loi Normale de paramètres μ et σ^2

Définition

Soient deux variables aléatoirs réelle X_1 et X_2 de moyenne μ_1 et μ_2, et de variance σ_1^2 et σ_2^2 indépendantes alors $X = X_1 + X_2$ suit une loi normale telle que

$$X \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^1 + \sigma_1^2)$$

Loi exponentielle de paramètre a

Proposition

La loi exponentielle de paramètre $a > 0$, est une loi à densité de fonction de densité

$$f_X(x) = \begin{cases} ae^{-ax} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases},$$

naturellement la fonction de répartition s'écrit

$$F_X(x) = \begin{cases} 1 - e^{-ax} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases},$$

A démontrer...

Loi exponentielle de paramètre a

Cas d'utilisation

Très utilisé pour simuler la survenue d'un décès chez un patient. Il est à la base des **modèles de survie.**
Voir par exemple **modèles de Cox**.

- $\mathbb{E}(X) = 1/a$
- $Var(X) = 1/a^2$

Autres lois importantes

Loi du χ^2 à n degrés de libertés

Soit n variables aléatoires $X_1, \cdots, X_n$ indépendantes et identiquement distribuées qui suivent une $\mathcal{N}(0,1)$ alors la variable aléatoire $Z = X_1^2 + \cdots + X_n^2$ suit une loi dite du χ^2 à n degrés de libertés. $Z \sim \chi^2(n)$

Très utilisé dans le test du χ^2 : tester la pertinence d'un modèle pour des données.

Loi de Student à n degrés de libertés

Si $X \sim \mathcal{N}(0,1)$ et $Y \sim \chi^2(n)$, indépndantes, alors $Z = \dfrac{X}{\sqrt{Y/n}}$ suit une loi de Student à n degrés de libertés. $Z \sim Student(n)$.

Pourquoi ces lois ?

Supposons que les X_i suivent tous une $\mathcal{N}(\mu, \sigma^2)$ et soient indépendants.
Estimateur de l'espérance : la moyenne empirique

$$\hat{X}_n = \sum_{i=1}^{n} X_i / n,$$

alors on a

$$\hat{X}_n \sim \mathcal{N}(\mu, \sigma^2 / n)$$

Pourquoi ces lois ?

L'écart-type empirique $\hat{\sigma}_n = \sqrt{\sum_{i=1}^{n}(X_i - \hat{X}_n)^2/n}$, alors la variable aléatoire $Z = \dfrac{\hat{X}_n - \mu}{\hat{\sigma}_n/\sqrt{n-1}}$, suit une loi de Student à $n-1$ degrés de liberté.

Les tests

Définition

En pratique on s pose des questions : *Les données suivent-elle bien telle ou telle loi ?*
On y répond en testant une hypothèse $\mathcal{H}_0$

$$\mathcal{H}_0 : \text{Le médicament est sans effet,}$$

ainsi ce serait "bien" de contredire cette hypothèse et de donc valider l'hypothèse inverse $\mathcal{H}_1$

$$\mathcal{H}_1 : \text{Le médicament a un effet.}$$

On ramène les data à ce qui devrait ressembler à une distribution connue et on compare l'empirique et le théorique.

Le test de Student

Le cas de la comparaison de deux échantillons indépendants.
Le plus simple : passer par la moyenne en prenant en compte la variance.
On suppose données deux échantillons A et B indépendants et formés de n_A et n_B observations correspondant à deux conditions expérimentales que l'on souhaiterait tester.
Exemple : traitement fonctionne contre traitement fonctionne pas...

Hypothèse

$\mathcal{H}_0$: Les deux échantillons suivent deux lois de même espérance μ.

Modèles statistiques

On pose les modèles statistiques sur X_A et X_B

$$\begin{aligned} X_A &\sim \mathcal{N}(\mu_A, \sigma^2) \\ X_B &\sim \mathcal{N}(\mu_B, \sigma^2) \end{aligned}$$

Le test de Student

On peut estimer les moyennes des échantillons :

$$\begin{aligned}\hat{X}_A &= \sum_{i_A=1}^{n_A} X_{i_A}/n_A \\ \hat{X}_B &= \sum_{i_B=1}^{n_B} X_{i_B}/n_B\end{aligned}$$

Si l'hypothèse de distribution est vérifiée

$$\begin{aligned}\hat{X}_A &\sim \mathcal{N}(\mu_A, \sigma^2/n_A) \\ \hat{X}_B &\sim \mathcal{N}(\mu_B, \sigma^2/n_B)\end{aligned}$$

Le test de Student

Grâce à ce que l'on a vu

$$\hat{X}_A - \hat{X}_B \sim \mathcal{N}(\mu_A - \mu_B, \sigma^2(1/n_A + 1/n_B))$$

On reformule alors les hypothèses

$$\begin{aligned} \mathcal{H}_0 &: \mu_A = \mu_B \\ \mathcal{H}_1 &: \mu_A \neq \mu_B \end{aligned}$$

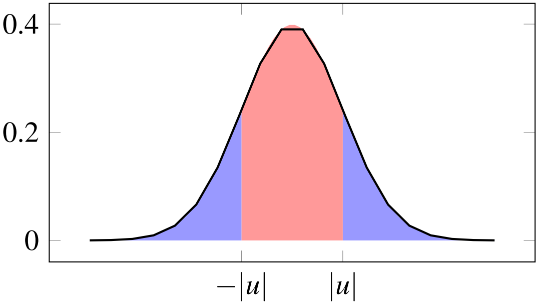
Cas 1 : la variance est connue

Sous $\mathcal{H}_0$: $U = \dfrac{\hat{X}_A - \hat{X}_B}{\sigma\sqrt{1/n_A + 1/n_B}} \sim \mathcal{N}(0,1)$, On calcule alors la statistique et on estime la **p-value** grâce à la table de la loi normale centrée réduite, en notant u la valeur estimée pour U on a

$$p - value = \mathbb{P}(U \leqslant -|u|) + \mathbb{P}(U \geqslant |u|)$$

Classiquement on pose une $p - value$ à 0.05 ou 0.01 ou 0.001 en fonction de pas mal de paramètres empiriques.

Loi normale



Cas 2 : la variance est inconnue

Il faut passer par l'estimateur de la variance $\hat{\sigma}^2$ et alors $T_{n_A+n_B-2} = \dfrac{\hat{X}_A - \hat{X}_B}{\sigma\sqrt{1/n_A + 1/n_B}} \sim \mathcal{T}(n_A + n_B - 2)$, suit une loi de Student à $n_A + n_B - 2$ degrés de liberté. On utilise pareillement la table de la loi de Student pour estimer la *plausibilité* du résultat via la $p - value$ fixée en début d'expérience.