SYS865 Inférence statistique avec programmation R

Ornwipa Thamsuwan

17 janvier 2024

Plan de la séance

Plan de la séance

- Une variable aléatoire (suite)
  - Fonctions de probabilité (suite)
  - Espérance et variance
- Plusieurs variables aléatoires
  - Propiétés
  - Indépendance vs. covariance

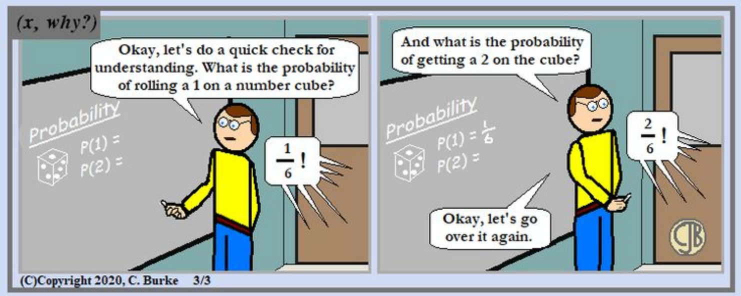


Figure 1: Probabilité

Recap

Recap

Le dernier cours...

- Fonction de masse de probabilité : Variables aléatoires discrètes
  - Lancer une pièce de monaire, quelle est la probabilité d'obtenir une pile ? et une face ?
  - Lancer un dé équilibré, quelle est la probabilité d'obtenir "1"... "6" ?
- Fonction de densité de probabilité : Variables aléatoires continues
  - Quelle est la probabilité qu'un élève de la classe ait 23 ans ?

Recap

Le dernier cours. . .

- Fonction de masse de probabilité : Variables aléatoires discrètes
  - Lancer une pièce de monaire, quelle est la probabilité d'obtenir une pile ? et une face ?
  - Lancer un dé équilibré, quelle est la probabilité d'obtenir "1". . . "6" ?
- Fonction de densité de probabilité : Variables aléatoires continues
  - Quelle est la probabilité qu'un élève de la classe ait 23 ans ?

Dans ce cours. . .

- Fonction de répartition
  - Lancer un dé équilibré, quelle est la probabilité d'obtenir un numéro moins de 2 ?
  - Quelle est la probabilité qu'un élève de la classe ait plus de 23 ans ?

Fonction de répartition de probabilité

Variables aléatoires discrètes

Résultats possibles d'un lancer de dé

```
outcomes <- 1:6
```

Variables aléatoires discrètes

Résultats possibles d'un lancer de dé

```
outcomes <- 1:6
```

Probabilité pour chaque résultat

```
probabilities <- rep(1/6, 6)
```

Variables aléatoires discrètes

Résultats possibles d'un lancer de dé

```r
outcomes <- 1:6
```

Probabilité pour chaque résultat

```r
probabilities <- rep(1/6, 6)
```

Vecteur représentant la fonction de masse de probabilité

```
##      1      2      3      4      5      6
## 0.1667 0.1667 0.1667 0.1667 0.1667 0.1667
```

Variables aléatoires discrètes

Résultats possibles d'un lancer de dé

```
outcomes <- 1:6
```

Probabilité pour chaque résultat

```
probabilities <- rep(1/6, 6)
```

Vecteur représentant la fonction de masse de probabilité

```
##      1      2      3      4      5      6
## 0.1667 0.1667 0.1667 0.1667 0.1667 0.1667
```

Quelle est la probabilité d'obtenir un numéro moins de 3 ?

Variables aléatoires discrètes (suite)

La **fonction de répartition** donne la probabilité que la variable aléatoire soit **inférieure ou égale** à cette valeur.

Si X est une variable aléatoire, sa fonction de répartition $F(x)$ est définie par $F(x) = P(X \leq x)$.

Variables aléatoires discrètes (suite)

La **fonction de répartition** donne la probabilité que la variable aléatoire soit **inférieure ou égale** à cette valeur.

Si X est une variable aléatoire, sa fonction de répartition $F(x)$ est définie par $F(x) = P(X \leq x)$.

Pour un dé à six faces équilibré, considérons la probabilité d'obtenir un résultat inférieur ou égal à 2, c'est-à-dire $F(2)$.

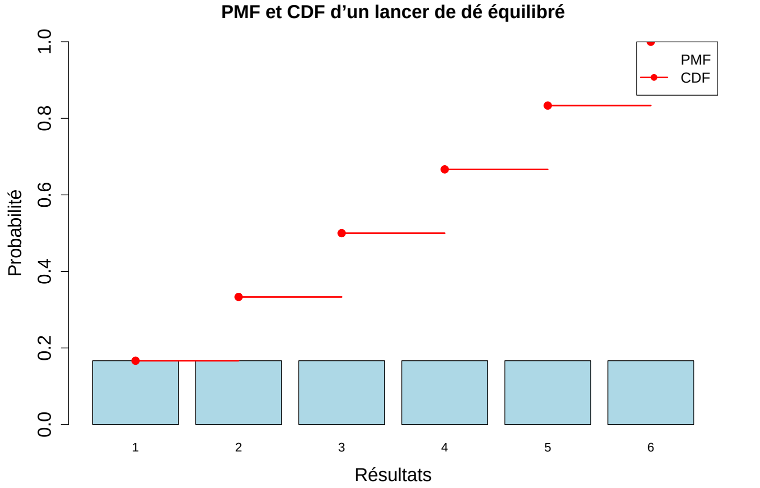
Les résultats possibles du dé sont 1, 2, 3, 4, 5 et 6, chacun avec une probabilité de $\frac{1}{6}$.

Donc,
$F(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.
Ainsi, la probabilité d'obtenir un résultat de 2 ou moins lors du lancer d'un dé est $\frac{1}{3}$.

Variables aléatoires discrètes (suite)

Fonction de répartition : Somme cumulée des probabilités

```r
cdf <- cumsum(probabilities)
```



Variables aléatoires continues

La fonction de répartition cumulative (CDF) d'une variable aléatoire continue X, notée $F(x)$, est l'intégrale de sa fonction de densité de probabilité (PDF) $f(t)$ de $-\infty$ à x :

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

Elle varie de 0 à 1 et est utile pour calculer les probabilités cumulatives.

Variables aléatoires continues

La fonction de répartition cumulative (CDF) d'une variable aléatoire continue X, notée $F(x)$, est l'intégrale de sa fonction de densité de probabilité (PDF) $f(t)$ de $-\infty$ à x :

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

Elle varie de 0 à 1 et est utile pour calculer les probabilités cumulatives.

Quelle est la probabilité qu'un élève de la classe ait plus de 23 ans ?

Variables aléatoires continues

La fonction de répartition cumulative (CDF) d'une variable aléatoire continue X, notée $F(x)$, est l'intégrale de sa fonction de densité de probabilité (PDF) $f(t)$ de $-\infty$ à x :

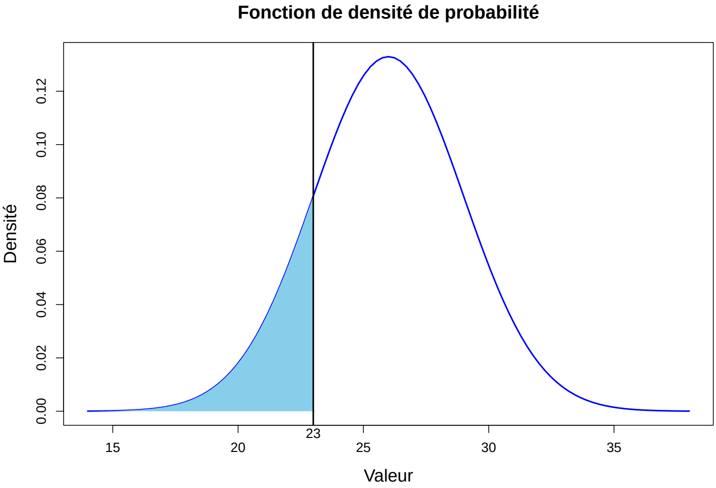
$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

Elle varie de 0 à 1 et est utile pour calculer les probabilités cumulatives.

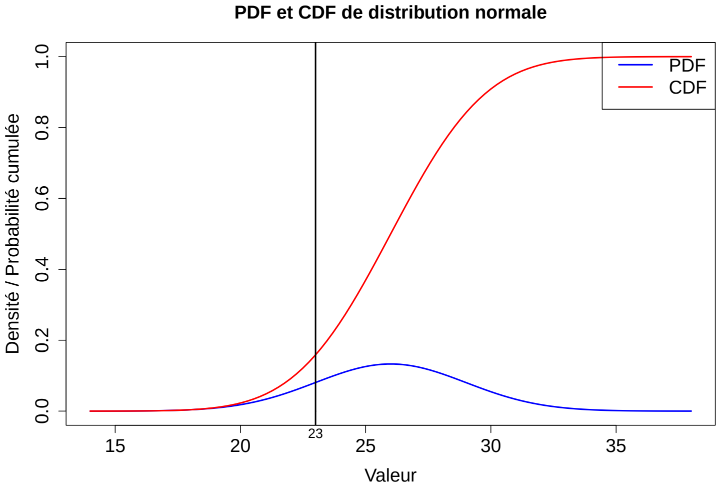
Quelle est la probabilité qu'un élève de la classe ait plus de 23 ans ?

Appliquant la définition de CDF, considérons la probabilité que un étudiant ait 23 ans ou moins, $F(X = 23)$ où X est la variable aléatoire de l'âge des étudiants.

Variables aléatoires continues (suite)



Variables aléatoires continues (suite)



Variables aléatoires continues (suite)

Valeur CDF où $X = 23$, c'est-à-dire, $F(23)$

```
cdf <- pnorm(23, mean = mean_value, sd = sd_value)
round(cdf, 4)
```

```
## [1] 0.1587
```

Variables aléatoires continues (suite)

Valeur CDF où $X = 23$, c'est-à-dire, $F(23)$

```r
cdf <- pnorm(23, mean = mean_value, sd = sd_value)
round(cdf, 4)
```

```
## [1] 0.1587
```

La probabilité qu'un étudiant ait plus de 23 ans, $1 - F(23)$

```r
round(1 - cdf, 4)
```

```
## [1] 0.8413
```

Espérance

Variables aléatoires discrètes

Définition L'espérance d'une variable aléatoire discrète est une mesure centrale, représentant la moyenne pondérée de toutes les valeurs possibles que la variable peut prendre. Chaque valeur est pondérée par sa probabilité. Pour une variable aléatoire discrète X, l'espérance, notée $E(X)$, est calculée comme suit :

$$E(X) = \sum_{x} x \cdot P(X = x)$$

où x représente les valeurs que X peut prendre et $P(X = x)$ est la probabilité que X soit égale à x.

Variables aléatoires discrètes (suite)

Pour un dé à six faces...

$$E(X) = \sum_{x=1}^{6} x \cdot P(X = x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

Ce qui donne :

$$E(X) = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{21}{6} = 3.5$$

Cela signifie que sur de nombreux lancers, la moyenne des résultats tend vers 3.5.

Variables aléatoires continues

Définition L'espérance d'une variable aléatoire continue est la moyenne pondérée de toutes ses valeurs possibles. Pour une variable aléatoire continue X avec une fonction de densité de probabilité $f(x)$, l'espérance, notée $E(X)$, est donnée par l'intégrale sur tout l'espace où X est défini :

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Cette intégrale représente la moyenne des valeurs pondérées par leur probabilité de survenir.

Variables aléatoires continues

Définition L'espérance d'une variable aléatoire continue est la moyenne pondérée de toutes ses valeurs possibles. Pour une variable aléatoire continue X avec une fonction de densité de probabilité $f(x)$, l'espérance, notée $E(X)$, est donnée par l'intégrale sur tout l'espace où X est défini :

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Cette intégrale représente la moyenne des valeurs pondérées par leur probabilité de survenir.

Quelques exemples ?

Variables aléatoires continues (suite)

Distribution Normale (Gaussienne)

Cette distribution est couramment utilisée pour modéliser des phénomènes naturels.

Expression Mathématique: Si X suit une distribution normale avec une moyenne μ et un écart-type σ, alors:

$$E(X) = \mu$$

Exemple : Les tailles des adultes dans une population peuvent souvent être modélisées par une distribution normale. Par exemple, si la taille moyenne des hommes adultes dans une certaine population est de 175 cm avec un écart-type de 10 cm, alors $E(X) = 175$ cm.

Variables aléatoires continues (suite)

Distribution Uniforme

Cette distribution est utilisée lorsque chaque intervalle de la plage de valeurs a une probabilité égale.

Expression Mathématique: Pour une distribution uniforme continue entre a et b:

$$E(X) = \frac{a+b}{2}$$

Exemple : Supposons qu'un bus arrive à un arrêt toutes les 20 minutes, et vous arrivez à l'arrêt de bus à un moment aléatoire. Le temps d'attente peut être modélisé comme une distribution uniforme entre 0 et 20 minutes, donc $E(X) = \frac{0+20}{2} = 10$ minutes.

Variables aléatoires continues (suite)

Distribution Exponentielle

Cette distribution peut représenter le temps d'attente ou le temps entre les événements dans un processus de Poisson.

Expression Mathématique: Si X suit une distribution exponentielle avec un taux λ:

$$E(X) = \frac{1}{\lambda}$$

Exemple : Imaginons un service client où l'on observe en moyenne l'arrivée d'un client toutes les 15 minutes. Cela signifie que le taux d'arrivée λ est de $\frac{1}{15}$ clients par minute. Le temps moyen d'attente jusqu'à l'arrivée du prochain client est l'inverse du taux d'arrivée. Par conséquent, l'espérance $E(X)$ ou le temps d'attente moyen est de $\frac{1}{\frac{1}{15}} = 15$ minutes.

Propriétés de l'espérance

Espérance d'une Constante :

- L'espérance d'une constante est égale à la constante elle-même :
$$E(c) = c$$
- Cette propriété est vraie car la distribution de probabilité d'une constante est un point unique à la valeur de la constante.

Propriétés de l'espérance (suite)

Linéarité :

- Pour deux variables aléatoires X et Y, et deux constantes a et b, l'espérance est linéaire :

$$E(aX + bY) = aE(X) + bE(Y)$$

- Cela signifie que l'espérance d'une combinaison linéaire de variables aléatoires est égale à la même combinaison linéaire de leurs espérances.

Propriétés de l'espérance (suite)

Multiplication par une Constante :

- Pour une variable aléatoire X et une constante c :

$$E(cX) = cE(X)$$

- Multiplier une variable aléatoire par une constante multiplie son espérance par cette constante.

Propriétés de l'espérance (suite)

Additivité :

- L'espérance de la somme de deux ou plusieurs variables aléatoires est égale à la somme de leurs espérances :

$$E(X+Y) = E(X) + E(Y)$$

- Ceci est vrai que X et Y soient indépendantes ou non.

Propriétés de l'espérance (suite)

Produit de Variables Indépendantes :

- Si deux variables aléatoires X et Y sont indépendantes, alors :

$$E(XY) = E(X)E(Y)$$

- L'espérance du produit de variables indépendantes est le produit de leurs espérances.

Variance

Variables aléatoires discrètes

Définition La variance d'une variable aléatoire discrète mesure la quantité moyenne de déviation des valeurs de la variable par rapport à la moyenne. Elle est calculée comme la moyenne des carrés des différences entre chaque valeur et la moyenne. Pour une variable aléatoire discrète X avec des valeurs possibles $x_1, x_2, ..., x_n$ et des probabilités $P(X = x_i)$, la variance, notée $\text{Var}(X)$, est donnée par :

$$\text{Var}(X) = \sum (x_i - \mu)^2 P(X = x_i)$$

où μ est la moyenne (valeur attendue) de X.

Variables aléatoires discrètes (suite)

Pour d'un dé à six faces...

La moyenne (valeur attendue) du lancer de dé, μ, est :

$$\mu = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

La variance est alors calculée comme suit :

$$\text{Var}(X) = \sum_{i=1}^{6}(x_i - 3.5)^2 \times \frac{1}{6}$$

Cela implique de sommer les carrés des différences de chaque valeur faciale à partir de la moyenne, chacun multiplié par la probabilité $\frac{1}{6}$. Le résultat donne la variance du lancer de dé, indiquant à quel point les résultats s'écartent de la moyenne de 3.5.

Variables aléatoires continues

Définition La variance d'une variable aléatoire continue mesure la dispersion de ses valeurs autour de la moyenne, indiquant à quel point les valeurs de la variable s'écartent en moyenne de la moyenne. Pour une variable aléatoire continue X ayant une fonction de densité de probabilité $f(x)$ et un valeur attendue μ, la variance $\mathrm{Var}(X)$ est définie :

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

Cet intégral représente la moyenne des différences au carré entre chaque valeur possible de X et la moyenne μ, pondérée par la probabilité de chaque valeur.

Variables aléatoires continues

Définition La variance d'une variable aléatoire continue mesure la dispersion de ses valeurs autour de la moyenne, indiquant à quel point les valeurs de la variable s'écartent en moyenne de la moyenne. Pour une variable aléatoire continue X ayant une fonction de densité de probabilité $f(x)$ et un valeur attendue μ, la variance $\mathrm{Var}(X)$ est définie :

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

Cet intégral représente la moyenne des différences au carré entre chaque valeur possible de X et la moyenne μ, pondérée par la probabilité de chaque valeur.

À retenir... Une variance élevée signifie que les valeurs de X sont plus dispersées autour de μ, tandis qu'une faible variance indique qu'elles sont plus regroupées autour de μ.

Variables aléatoires continues (suite)

Distribution Normale

Expression Mathématique : Pour une distribution normale avec une moyenne μ et un écart-type σ, la variance est donnée par :

$$\text{Var}(X) = \sigma^2$$

Exemple : Dans une population des adultes dont les tailles sont normalement distribuées avec une moyenne de 175 cm et un écart-type de 10 cm, la variance des tailles est $10^2 = 100\,\text{cm}^2$.

Variables aléatoires continues (suite)

Distribution Uniforme

Expression Mathématique : Pour une distribution uniforme allant de a à b, la variance est :

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

Exemple : Si un bus arrive de façon aléatoire entre 14h00 et 15h00 (120 à 180 minutes), la variance du temps d'arrivée est $\frac{(180-120)^2}{12} = 300\,\text{minutes}^2$.

Variables aléatoires continues (suite)

Distribution Exponentielle

Expression Mathématique : Pour une distribution exponentielle avec un taux λ, la variance est :

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Exemple : Si les clients arrivent à un taux de 3 par heure (donc $\lambda = 3$), la variance du temps entre les arrivées est $\frac{1}{3^2} = \frac{1}{9}$ heures^2.

Propriétés de la variance

Non-Négativité :

- La variance est toujours non-négative : $\mathrm{Var}(X) \geq 0$.
- Cela est dû au fait que la variance est la moyenne des différences au carré par rapport à la moyenne, et les carrés sont toujours non-négatifs.

Propriétés de la variance (suite)

Variance d'une Constante et Variance Zéro :

- La variance d'une constante est zéro : $\text{Var}(c) = 0$.
- Si la variance d'une variable aléatoire est zéro, alors cette variable est presque sûrement une constante.
- Une constante ne varie pas, donc son écart par rapport à sa moyenne (elle-même) est toujours nul.

Propriétés de la variance (suite)

Linéarité :

- La variance d'une variable aléatoire multipliée par une constante est égale au carré de cette constante multiplié par la variance de la variable : $\text{Var}(aX) = a^2\text{Var}(X)$.

Propriétés de la variance (suite)

Somme des Variances :

- Pour deux variables aléatoires indépendantes X et Y, la variance de leur somme est la somme de leurs variances : $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.
- Cette propriété ne s'applique pas nécessairement si X et Y ne sont pas indépendantes.

Loi jointe

Loi jointe

La **distribution de probabilité jointe** décrit la probabilité que les deux variables aléatoires X et Y prennent certaines valeurs simultanément.

Figure 2: Probabilité jointe

Variables aléatoires discrètes

Ornwipa Thamsuwan

Pour des **variables aléatoires discrètes**, la fonction de masse de probabilité jointe (PMF) de X et Y est notée $P(X = x, Y = y)$ et représente la probabilité que X prenne la valeur x et Y la valeur y simultanément. Elle s'exprime comme :

$$P(X = x, Y = y)$$

Variables aléatoires discrètes (exemple)

Examinons la relation entre le moment de la journée et l'intérêt d'un chat à poursuivre un pointeur laser.

Définissons deux variables aléatoires discrètes :

- X : Moment de la journée, où $X = 1$ représente le soir et $X = 0$ le matin.
- Y : Activité du chat, avec $Y = 1$ indiquant que le chat poursuit activement un pointeur laser, et $Y = 0$ signifiant qu'il n'est pas intéressé.

Variables aléatoires discrètes (exemple)

$P(X = 1, Y = 1)$ représente la probabilité qu'il soit le soir (X=1) et que le chat poursuive activement le pointeur laser (Y=1). Par exemple, supposons que le chat soit plus énergique et joueur le soir, $P(X = 1, Y = 1) = 0.7$.

Variables aléatoires discrètes (exemple)

$P(X = 1, Y = 1)$ représente la probabilité qu'il soit le soir (X=1) et que le chat poursuive activement le pointeur laser (Y=1). Par exemple, supposons que le chat soit plus énergique et joueur le soir, $P(X = 1, Y = 1) = 0.7$.

$P(X = 0, Y = 1)$ indique la probabilité qu'il soit le matin (X=0) et que le chat poursuive activement le pointeur laser (Y=1). Car les chats sont souvent moins actifs le matin, préférant se reposer ou dormir après une nuit d'activité, $P(X = 0, Y = 1) = 0.3$.

Variables aléatoires continues

Pour des **variables aléatoires continues**, la probabilité jointe est représentée par une fonction de densité de probabilité jointe (PDF), notée $f_{X,Y}(x,y)$. Elle décrit la densité de probabilité à tout point dans la plage de X et Y. La probabilité que X et Y tombent dans des intervalles spécifiques est donnée par l'intégrale de la PDF jointe sur ces intervalles :

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y)\, dx\, dy$$

Variables aléatoires continues (exemple)

Étudions la fascination d'un chat nommé Julio pour la poursuite d'un pointeur laser au cours de la journée.

Définissons deux variables aléatoires continues :

- X : Moment de la journée, mesuré en heures (allant de 0 à 24, où 0 représente minuit, 12 représente midi, etc.).
- Y : Vitesse de Julio (en mètres par seconde) en poursuivant le pointeur laser.

Variables aléatoires continues (exemple)

La distribution de probabilité jointe $f_{X,Y}(x, y)$ explore la relation entre le moment de la journée et la vitesse de poursuite de Julio.

Variables aléatoires continues (exemple)

La distribution de probabilité jointe $f_{X,Y}(x,y)$ explore la relation entre le moment de la journée et la vitesse de poursuite de Julio.

Disons qu'il est plus active et donc plus rapide à certains moments de la journée, comme tôt le matin et tard le soir.

Par exemple, la densité de probabilité jointe pourrait être plus élevée pour que Julio sprinte plus vite (valeurs élevées de Y) d'environ à 6-7h (disons $6 < X \leq 7$) et à nouveau à 19-22h (disons $19 < X \leq 22$).

Variables aléatoires continues (exemple)

La distribution de probabilité jointe $f_{X,Y}(x,y)$ explore la relation entre le moment de la journée et la vitesse de poursuite de Julio.

Disons qu'il est plus active et donc plus rapide à certains moments de la journée, comme tôt le matin et tard le soir.

Par exemple, la densité de probabilité jointe pourrait être plus élevée pour que Julio sprinte plus vite (valeurs élevées de Y) d'environ à 6-7h (disons $6 < X \leq 7$) et à nouveau à 19-22h (disons $19 < X \leq 22$).

En revanche, la densité de probabilité jointe serait plus faible autour de midi ou tard dans la nuit, des moments où Julio préfère faire la sieste au soleil ou sur un coussin confortable.

Indépendance

Indépendance

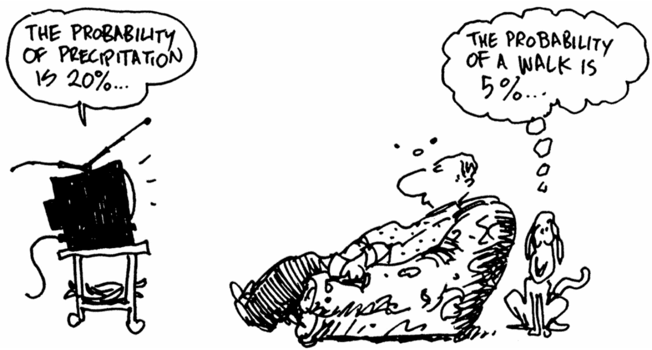


Figure 3: Indépendance ou non ?

Indépendance

Deux variables aléatoires X et Y sont dites indépendantes si la réalisation de l'une ne modifie pas la probabilité de réalisation de l'autre. En d'autres termes, la connaissance du résultat de X n'apporte aucune information sur le résultat de Y, et vice versa.

Indépendance

Deux variables aléatoires X et Y sont dites indépendantes si la réalisation de l'une ne modifie pas la probabilité de réalisation de l'autre. En d'autres termes, la connaissance du résultat de X n'apporte aucune information sur le résultat de Y, et vice versa.

Définition

Mathématiquement, deux variables aléatoires X et Y sont indépendantes **si et seulement si** pour tous les ensembles A et B :

$$P(X \in A, Y \in B) = P(X \in A) \times P(Y \in B)$$

Indépendance (suite)

Pour des variables discrètes, cela signifie que :

$$P(X = x, Y = y) = P(X = x) \times P(Y = y)$$

Indépendance (suite)

Pour des variables discrètes, cela signifie que :

$$P(X = x, Y = y) = P(X = x) \times P(Y = y)$$

Pour des variables continues, l'indépendance implique que la fonction de densité de probabilité conjointe s'est exprimée comme le produit des fonctions de densité marginales :

$$f_{X,Y}(x, y) = f_X(x) \times f_Y(y)$$

Indépendance (suite)

Pour des variables discrètes, cela signifie que :

$$P(X = x, Y = y) = P(X = x) \times P(Y = y)$$

Pour des variables continues, l'indépendance implique que la fonction de densité de probabilité conjointe s'est exprimée comme le produit des fonctions de densité marginales :

$$f_{X,Y}(x, y) = f_X(x) \times f_Y(y)$$

L'indépendance est cruciale pour de nombreuses méthodes d'analyse statistique.

- À étudier et utiliser dans les prochains séances . . .

Indépendance (suite)

Récapitulatif

Lorsque deux variables aléatoires X et Y sont indépendantes ...

- **Espérance du Produit :**

Indépendance (suite)

Récapitulatif

Lorsque deux variables aléatoires X et Y sont indépendantes . . .

- **Espérance du Produit :**
- Si X et Y sont indépendantes, alors l'espérance de leur produit est le produit de leurs espérances :

$$E(XY) = E(X)E(Y)$$

Indépendance (suite)

Récapitulatif

Lorsque deux variables aléatoires X et Y sont indépendantes ...

- **Espérance du Produit :**
- Si X et Y sont indépendantes, alors l'espérance de leur produit est le produit de leurs espérances :

$$E(XY) = E(X)E(Y)$$

- **Variance de la Somme :**

Indépendance (suite)

Récapitulatif

Lorsque deux variables aléatoires X et Y sont indépendantes
...

- **Espérance du Produit :**
- Si X et Y sont indépendantes, alors l'espérance de leur produit est le produit de leurs espérances :

$$E(XY) = E(X)E(Y)$$

- **Variance de la Somme :**
- Pour des variables indépendantes, la variance de leur somme est la somme de leurs variances :

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Covariance

Covariance

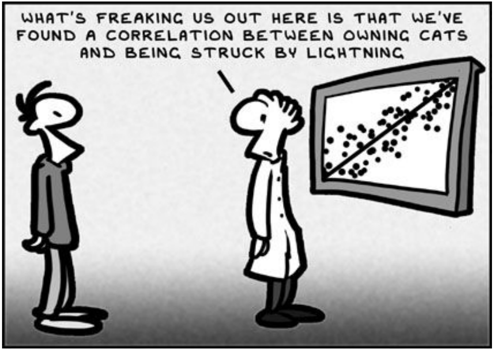


Figure 4: Covariance

Covariance

La covariance est une mesure utilisée en statistique pour déterminer le degré de variation conjointe de deux variables aléatoires. Elle aide à comprendre la relation entre deux variables et si elles ont tendance à évoluer dans la même direction ou dans des directions opposées.

Covariance

La covariance est une mesure utilisée en statistique pour déterminer le degré de variation conjointe de deux variables aléatoires. Elle aide à comprendre la relation entre deux variables et si elles ont tendance à évoluer dans la même direction ou dans des directions opposées.

Définition

La covariance entre deux variables aléatoires X et Y est définie comme la valeur attendue du produit de leurs écarts par rapport à leurs valeurs moyennes respectives.

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

où μ_X et μ_Y sont les moyennes de X et Y, respectivement.

Calcul

Pour des variables discrètes, la covariance peut être calculée en utilisant la formule :

$$\text{Cov}(X, Y) = \sum (x_i - \mu_X)(y_i - \mu_Y)P(x_i, y_i)$$

où x_i et y_i sont les valeurs de X et Y, et $P(x_i, y_i)$ est la probabilité jointe de X et Y.

Calcul

Pour des variables discrètes, la covariance peut être calculée en utilisant la formule :

$$\text{Cov}(X, Y) = \sum (x_i - \mu_X)(y_i - \mu_Y) P(x_i, y_i)$$

où x_i et y_i sont les valeurs de X et Y, et $P(x_i, y_i)$ est la probabilité jointe de X et Y.

Pour des variables continues, elle est calculée en utilisant l'intégrale :

$$\text{Cov}(X, Y) = \int \int (x - \mu_X)(y - \mu_Y) f(x, y)\, dx\, dy$$

où $f(x, y)$ est la fonction de densité de probabilité conjointe de X et Y.

Interprétation

Covariance Positive : Se produit lorsque deux variables ont tendance à évoluer dans la même direction.

- Prix des actions des entreprises technologiques et demande des consommateurs pour la technologie.

Interprétation

Covariance Positive : Se produit lorsque deux variables ont tendance à évoluer dans la même direction.

- Prix des actions des entreprises technologiques et demande des consommateurs pour la technologie.

Covariance Négative : Se produit lorsqu'une variable a tendance à augmenter alors que l'autre diminue.

- Utilisation du chauffage et température extérieure en hiver.

Interprétation

Covariance Positive : Se produit lorsque deux variables ont tendance à évoluer dans la même direction.

- Prix des actions des entreprises technologiques et demande des consommateurs pour la technologie.

Covariance Négative : Se produit lorsqu'une variable a tendance à augmenter alors que l'autre diminue.

- Utilisation du chauffage et température extérieure en hiver.

Covariance Nulle (ou Proche de Zéro) : Indique qu'il n'y a pas de relation entre les deux variables.

- Nombre d'heures d'étude et couleur de cheveux d'un étudiant.

Travaux pratiques

Travaux pratiques

Pour chacun des neuf paramètres dans la base de données sur les diabètes ...

1. Déterminer si la variable est discrète ou continue
2. (Bonus) Créer un histogramme pour présenter la distribution de probabilité
3. Calculer l'espérance et la variance

Pour chaque pair des deux paramètres ...

4. En appliquant des connaissances apprises dans ce cours, déterminer si les deux variables sont indépendantes
5. Calculer les covariances