Probabilités et Statistiques

Dans cette partie, on présente quelques résultats en probabilités et statistiques dans le cadre de ce cours. Pour plus d'information, vous pouvez vous référer au cours STT-1000, à Wasserman (2010) (en anglais) et à Delmas (2013) (en français).

1 Modéliser le hasard

Beaucoup de phénomènes réels ne sont pas prévisibles et généralement, leurs résultats contiennent une certaine variabilité. Cette variabilité est prise en compte grâce à une mesure de l'incertitude que l'on appelle **mesure de probabilités**.

Définition

L'**espace d'évènements** S est l'ensemble de tous les résultats possibles d'un phénomène. Un **évènement** est un sous-ensemble de l'espace d'évènements S.

Exemples

1. Si l'expérience consiste à lancer un pièce, $S = \{0, 1\}$. Le résultat de cette expérience ne peut pas être connu à l'avance. Par exemple, $E = \{1\}$ est un évènement de S.
2. Si on s'intéresse à la durée de vie d'un téléphone, $S = \mathbb{R}_+$. On peut aussi choisir $S = [0, M]$, car cette durée de vie n'est probablement pas infini ! L'évènement $E = [10, \infty)$ représente l'évènement "une durée de vie de plus de 10 unités de temps".
3. Pour le nombre de jours sans neige à Québec dans l'année, on peut choisir $S = \mathbb{N}$. L'évènement $E = (0, 5]$ représente l'évènement "moins de 5 jours sans neige à Québec dans l'année".

Définition

Une **mesure de probabilités** $\mathbb{P}$ sur S est une application (fonction) définie sur l'espace d'évènements et satisfaisant les propriétés suivantes :

1. Pour chaque évènement E, $\mathbb{P}(E) \in [0, 1]$.
2. $\mathbb{P}(S) = 1$.
3. Soient $E_1, E_2, ...$, une séquence d'évènements (finie ou infinie) mutuellement exclusive, i.e. $\forall i \neq j, E_i \cap E_j = \emptyset$. On a

$$\mathbb{P}(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mathbb{P}(E_n).$$

On appelle $\mathbb{P}(E)$, la probabilité de l'évènement E.

La définition de mesures de probabilités peut être subjective et lié à l'expérience du statisticien. En reprenant l'exemple 3 sur le nombre de jours sans neige à Québec dans l'année. Une personne venant d'arriver au Canada peut vouloir donner la même probabilité à chacun des jours, alors qu'un Québécois aura plus d'information et pourras faire varier les probabilités en fonction de cette connaissance.

Définition

Deux évènements E et F sont dits **indépendants** si $\mathbb{P}(E \cap F) = \mathbb{P}(E) \times \mathbb{P}(F)$.

Définition

Soient E et F, deux évènements, la **probabilité conditionelle** que E se réalise sachant que F s'est réalisé est définie par :

$$\mathbb{P}(E \mid F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}.$$

De façon intuitive, deux évènements sont indépendants si la connaissance de l'un ne donne aucune information sur la réalisation de l'autre. On a aussi $\mathbb{P}(E \mid F) = \mathbb{P}(E)$.

2 Variables aléatoires

En probabilité, la convention est d'exprimer le résultat d'expériences comme la valeur d'une fonction appelé **variable aléatoire**. Cette caractérisation est toujours possible.

Définition

Soit une variable aléatoire X. La **distribution** de cette variable aléatoire est définie par l'application $A \mapsto \mathbb{P}(X \in A)$.

Définition

Soit une variable aléatoire X. Cette variable aléatoire est **discrète** si elle prend, au plus, un nombre dénombrable de valeurs. Dans ce cas, la distribution de X est donnée par les probabilités $\mathbb{P}(X = x)$ pour tout résultat x.

Définition

Soit une variable aléatoire X. Cette variable aléatoire est **continue** si les probabilités $\mathbb{P}(X \in A)$ sont données par des intégrales de la forme $\int_A f(x)dx$ où $f : \mathbb{R}^d \to \mathbb{R}_+$ est une fonction intégrable tel que $\int_{\mathbb{R}^d} f(x)dx = 1$. Notons que, pour un résultat x fixé, $\mathbb{P}(X = x) = 0$.

Définition

Soit une variable aléatoire X. L'**espérance mathématique** $\mathbb{E}(X)$ de X est la valeur moyenne du résultat de X par rapport à sa distribution de probabilité. L'espérance est généralement noté μ.

Soit F un ensemble dénombrable. Une variable aléatoire discrète X a pour espérance $\mathbb{E}(X) = \sum_{x \in F} x\mathbb{P}(X = x)$. Soit une variable aléatoire continue X ayant pour densité f, son espérance est donnée par $\mathbb{E}(X) = \int_{\mathbb{R}^d} xf(x)dx$.

Théorème de transfert

Soit une variable aléatoire X. Soit $g : \mathbb{R}^d \mapsto \mathbb{R}$ une fonction telle que $\mathbb{E}[g(X)]$ existe. On a :

1. Si X est une variable aléatoire discrète, $\mathbb{E}[g(X)] = \sum_{x \in F} g(x)\mathbb{P}(X = x)$;
2. Si X est une variable aléatoire continue de densité f, $\mathbb{E}[g(X)] = \int_{\mathbb{R}^d} g(x)f(x)dx$.

Propriétés : Linéarité de l'espérance

Soient X et Y, deux variables aléatoires, dont les espérances sont définies et soit $\lambda \in R$. On a :

1. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
2. $\mathbb{E}(\lambda X) = \lambda\mathbb{E}(X)$.

Preuve

La preuve se déduit du théorème de transfert et de la linéarité de l'addition et de l'intégration.

Définition

Soit X une variable aléatoire telle que l'espérance de son carré existe. La **variance** de X est définie par

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2.$$

La variance mesure la dispersion d'une variable aléatoire autour de sa moyenne. On peut aussi s'intéresser à l'**écart-type**, défini comme la racine carrée de la variance : $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

Définition

Soient X et Y, deux variables aléatoires et A et B, deux évenements. Si les évenements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants, alors on dit que les variables aléatoires X et Y sont **indépendantes**.

De cette définition, on en déduit que :

1. pour des fonctions f et g, les variables aléatoires $f(X)$ et $g(Y)$ sont indépendantes ;
2. si les variables aléatoires X et Y sont à valeurs réelles et que leur espérance existe, alors l'espérance du produit XY existe et $\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y)$.

Définition

Soit X une variable aléatoire. La **fonction de répartition** $F : \mathbb{R} \mapsto [0, 1]$ de X est définie par

$$F(t) = \mathbb{P}(X \leq t), \quad t \in \mathbb{R}.$$

3 Vecteurs aléatoires

Supposons que $X = (X_1, X_2)$ est une variable aléatoire de dimension 2 de densité f_X. On appelle généralement les variables aléatoires de dimension supérieure à 1, des **vecteurs aléatoires**. Les densités de X_1 et X_2 sont appelées les **densités marginales**. Lorsque X_1 et X_2 sont indépendantes, on a :

$$f_X(x, y) = f_{X_1}(x) \cdot f_{X_2}(y), \quad (x, y) \in \mathbb{R}^2.$$

Exemple de la loi normale multidimensionnelle

On dit qu'un vecteur aléatoire X de dimension p suit une loi normale multidimensionnelle de moyenne μ et de variance Σ, si sa densité est donnée par

$$f_X(x) = \frac{1}{(2\pi)^{p/2}} \cdot \frac{1}{(\det\Sigma)^{1/2}} \cdot \exp\left\{-\frac{1}{2}\left(x-\mu\right)^\top \Sigma^{-1}\left(x-\mu\right)\right\}, \quad x \in \mathbb{R}^p.$$

On note $X \sim \mathcal{N}_p(\mu, \Sigma)$.

En statistiques, une quantité importante à mesurer est la dépendance linéaire entre X_1 et X_2. Pour cela, on peut utiliser la covariance ou la correlation.

Définition

Soit $X = (X_1, X_2)$ un vecteur aléatoire tel que l'espérance du carré de X_1 et de X_2 existe. La **covariance** entre X_1 et X_2 est donnée par

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))\right].$$

La **corrélation** entre X_1 et X_2 est une version de la covariance normalisée par l'écart-type des variables aléatoires. Elle est donnée par

$$\mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma(X_1)\sigma(X_2)}.$$

On peut interpréter le signe de la covariance et de la corrélation. Si elles sont strictement positives, X_1 et X_2 ont tendance à aller dans la même direction. Si X_1 augmente, alors X_2 aussi, et inversement. Si elles sont strictement négatives, X_1 et X_2 ont tendance à aller dans des directions opposées. Si X_1 augmente, alors X_2 diminue, et inversement. Si la covariance est égales à 0, il n'y a pas de règles et X_1 et X_2 sont dites orthogonales.

Propriétés

Soit $X = (X_1, X_2)$ un vecteur aléatoire. On a

1. $\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$;
2. $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1)$;
3. $\mathrm{Cov}(X_1 + \lambda Y_1, X_2) = \mathrm{Cov}(X_1, X_2) + \lambda \mathrm{Cov}(Y_1, X_2)$.

Preuve

1. On trouve le résultat en développement le produit dans la définition de la covariance.
2. En utilisant le point 1. et la commutativité de la multiplication.
3. En utilisant le point 1. et la linéarité de l'espérance.

4 Estimation

En practique, nous n'avons pas une connaissance parfaite de nos vecteurs aléatoires, mais seulement des réalisations de ceux-ci (que l'on appelle échantillon). Notons $x_1, \dots, x_n$, n réalisations indépendantes d'un vecteur aléatoire X de moyenne μ et de variance Σ.

L'estimateur de la moyenne μ est donné par

$$\hat{\mu} = \overline{X} := \frac{1}{n} \sum_{i=1}^{n} x_i.$$

L'estimateur de la variance Σ est donné par

$$\widehat{\Sigma} := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu})(x_i - \hat{\mu})^\top.$$

Pourquoi divise t-on cette somme par $n-1$ et non par n pour estimer la variance ? Si l'on divise par n, $\widehat{\Sigma}$ est un estimateur biaisé de la variance. En effet, il faut prendre en compte que l'on utilise un estimateur biaisé de la moyenne dans l'estimateur de la variance et donc corriger pour cette estimation.

Notons $D = \{\mathrm{diag}(\widehat{\Sigma})\}^{1/2}$, la matrice des écarts-types calculés sur l'échantillon. On peut estimer la matrice des corrélations sur l'échantillon par

$$\widehat{R} = D^{-1} \widehat{\Sigma} D^{-1}.$$

Références

Delmas, Jean-François. 2013. *Introduction au calcul des probabilités et à la statistique : exercices, problèmes et corrections (2e édition)*. Les Presses de l'ENSTA.

Wasserman, Larry. 2010. *All of Statistics : A Concise Course in Statistical Inference*. Springer Publishing Company, Incorporated.