UGA
IUT II Département informatique

TD 1 : Introduction aux lois continues

Préambule

Vous trouverez une version du cours en ligne avec des rappels sur les probabilités discrètes à l'adresse suivante :
`https://fcorset.github.io/cours/cours.html`

Notions de cours

On distingue en mathématiques la branche des probabilités de celle de la statistique. Pour faire simple les probabilités traitent de modèles mathématiques et de leurs propriétés alors que la statistique utilise les probabilités pour les appliquer à des données réelles.

Nous commencerons par une brève introduction aux probabilités afin de connaitre le langage et les objets mathématiques utiles à la statistique proprement dite. Chaque notion théorique sera dans la mesure du possible illustrée par des exemples numériques ou des simulations.
Nous utiliserons pour les illustrations le langage R.

Prérequis :
- savoir calculer des primitives simples
- savoir dériver

Objectifs :

- connaître les définitions de variable aléatoire, loi, fonction de densité, fonction de répartition.
- savoir calculer une espérance écrite sous forme d'une intégrale "simple".
- savoir calculer une variance écrite sous forme d'une intégrale "simple".

Côté mathématique

Définition 0.1 *Une "variable aléatoire" X est une fonction mathématique qui associe, à chaque élément d'un espace Ω :*
- soit une valeur entière et l'on parle alors de variable aléatoire discrète
- soit un nombre réel et l'on parle alors de variable aléatoire continue.

L'adjectif "aléatoire" est là pour exprimer le fait que l'espace de départ et celui d'arrivée ne sont pas quelconques mais qu'ils possèdent des propriétés particulières notamment celle d'être dotés de probabilités.
Cet espace de départ mystérieux (et totalement abstrait) ne nous intéressera jamais et nous ne chercherons donc pas à en savoir plus sur lui.....par contre toute la suite du cours va porter sur l'espace d'arrivée de cette fonction.
Le terme "variable aléatoire" désigne bien une fonction contrairement à l'habitude qui est de noter f une fonction et $f(x)$ l'image d'une valeur de départ x. En probabilité et en statistique, la lettre X désignera bien la fonction et non pas une valeur possible de la fonction ; c'est d'ailleurs pour cela que le plus souvent les variables aléatoires sont notées à l'aide de lettres majuscules pour ne pas confondre avec le petit x d'une fonction "classique".

Loi d'une variable aléatoire réelle continue
Dans le cas classique d'une fonction, pour la décrire, pour pouvoir la calculer et pour l'étudier, bref pour s'en servir, on donne sa formule ; par exemple $f(x) = x^2 + 31 + \frac{1}{\cos(x)}$.
Avec les variables aléatoires c'est plus compliqué.... Pour décrire une variable aléatoire il faut donner sa loi. Pour cela, il existe plusieurs façons (qui heureusement sont équivalentes entre elles, si on a l'une d'entre elles

on a les autres) :

- **première possibilité :** donner la fonction de densité associée à la variable aléatoire
- **deuxième possibilité :** donner la fonction de répartition associée à la variable aléatoire
- **troisième possibilité :** donner son nom (parmi une liste de noms de lois connues voir par exemple wikipedia).

Côté réalité
Il est possible de regarder en vrai une variable aléatoire, ou presque.... vu que cela n'existe pas !
Pour cela il suffit d'utiliser un générateur de nombres aléatoires. On pourrait presque considérer qu'une variable aléatoire et un générateur de nombres aléatoires c'est la même chose ! L'idée étant qu'un générateur aléatoire produit des nombres mais que ce qui est important ici n'est pas tant quels nombres il produit mais comment.

Densité et fonction de répartition d'une variable quantitative continue

La variable aléatoire X associée à une fonction f donnée et définie sur $\mathbb{R}$ représente le fait de tirer un nombre au hasard avec la probabilité suivante :

$$\text{Proba}(X \leq t) = \int_{-\infty}^{t} f(x)dx = F(t)$$

où t est un réel fixé.

1. Naturellement, cette écriture n'a de sens que si :
   (a) f est une fonction positive sur $\mathbb{R}$
   (b) $\displaystyle\int_{-\infty}^{+\infty} f(x)dx = 1.$
   f est appelée "**densité**" de X.

2. Cette probabilité est notée $F(t)$: F, vue comme une fonction de t définie sur $\mathbb{R}$, est appelée **fonction de répartition** de X. La valeur $F(t)$ peut être vue comme l'aire de la surface délimitée par la demi-droite $]-\infty, t]$, la droite $x = t$ et la courbe représentative de f.
3. L'espérance de X correspond à la valeur suivante :
$$E(X) = \int_{-\infty}^{+\infty} xf(x)\,dx.$$
4. La variance de X est :
$$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - \left(\int_{-\infty}^{+\infty} xf(x)\,dx\right)^2.$$
5. L'écart-type de X notée σ_X est : $\sigma_X = \sqrt{\text{Var}(X)}$.

Ce que vous devez être capable de faire, est de répondre aux questions suivantes :

Exercice 1 *Exercice de cours*

1. Sous quelles conditions une fonction F peut-elle être considérée comme une fonction de répartition d'une variable aléatoire continue X ?

2. Sous quelles conditions une fonction f peut-elle être considérée comme une densité associée à une variable aléatoire X ?
3. Soit X une variable aléatoire continue de densité f.
   (a) En supposant que les quantités existent, donner les définitions de l'espérance de X notée $\mathbb{E}(X)$ et de la variance de X notée $\text{Var}(X)$.
   (b) Comment obtient-on la fonction de répartition F de X ?
   (c) Écrire en fonction de F la probabilité que X soit comprise entre -1 et 5.
   (d) Définir la médiane de X.

Définitions et quelques résultats

Définition 0.2 *Soient X et Y deux variables aléatoires. Soient $(a, b) \in \mathbb{R}^2$.*
On appelle Covariance de X et Y la valeur suivante :

$$\text{Cov}(X, Y) = \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]).$$

On appelle Coefficient de corrélation entre X et Y la valeur suivante :

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \, \sigma_Y}.$$

Ce coefficient est compris entre -1 et 1.

Proposition 0.3 *L'espérance est linéaire : à savoir, $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.*
En particulier, $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.
La variance vérifie : $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\,\text{Cov}(X, Y)$.

Théorème 0.4 *Si les variables X et Y sont indépendantes, alors : $\text{Cov}(X, Y) = 0$ ou encore $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*

Exercice 2 *Propriétés de l'espérance et de la variance*
Soient X et Y deux variables aléatoires. Soient $(a, b) \in \mathbb{R}^2$.

1. Montrer que $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
2. Montrer, en utilisant la linéarité de l'espérance, que $\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$.
3. Montrer, en utilisant la linéarité de l'espérance, que $\text{Var}(aX+bY) = a^2\text{Var}(X)+b^2\text{Var}(Y)+2ab\,\text{Cov}(X, Y)$.
4. Que devient la propriété précédente si les variables X et Y sont indépendantes ?

Exercices

Exercice 3 *Une loi quelconque*
Soit k un réel. On considère la fonction f suivante définie sur $\mathbb{R}$ par :

$$f(x) = \begin{cases} kx^2 & \text{si } 0 \leq x \leq 2. \\ x + 1 & \text{si } -1 < x < 0 \\ 0 & \text{sinon} \end{cases}$$

1) Donner les conditions sur f pour que f soit la densité d'une variable aléatoire X continue.
2) Déterminer k afin que f soit une densité de probabilité.
3) Soit X la variable aléatoire continue ayant pour densité f. Calculer $E(X)$.
4) Déterminer la fonction de répartition de X.
5) En déduire les probabilités suivantes :

$$p(X < 1/2), p(X < 5), p(1 < X < 4), p(X = 1).$$

Exercice 4 *Une loi quelconque*
Soit $k \geq 0$ et X une variable aléatoire ayant pour densité la fonction suivante :

$$f(x) = \begin{cases} x+1 & \text{si } |x| \leq k \\ 0 & \text{sinon} \end{cases}$$

1) Tracer la courbe représentant f et déterminer k pour que f soit une densité.
2) Calculer la fonction de répartition F de X et la représenter graphiquement.
3) Calculer $\mathbb{E}(X)$, $\mathrm{Var}(X)$ et $\mathbb{E}(1+X^2)$.
4) Soit $T = 1 + X^2$. Déterminer $P(T \leq x)$. En déduire la densité de T et retrouver le dernier résultat de la question précédente.

Exercice 5 Soit k un réel. On considère la fonction f suivante définie sur $\mathbb{R}$ par :

$$f(x) = \begin{cases} k/x & \text{si} & 1 \leq x < e. \\ x & \text{si} & 0 < x < 1 \\ 0 & \text{sinon} & \end{cases}$$

1. Donner les conditions sur f pour que f soit la densité d'une variable aléatoire continue, notée X. Déterminer k afin que f soit une densité de probabilité.
2. Calculer $\mathbb{E}(X)$.
3. Calculer $\mathbb{V}[X]$.
4. Déterminer la fonction de répartition de X.
5. En déduire les probabilités suivantes :

$$p(X < 1),\ p(X > 1/2),\ p(X = 1).$$

Exercice 6 *La loi exponentielle*
Soit λ un nombre réel strictement positif. On considère la fonction f définie sur $\mathbb{R}$ par

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

Partie I.

1. Vérifier que f est la densité de probabilité d'une variable aléatoire X. On dit alors que X suit une loi exponentielle de paramètre λ.
2. Quelle est la fonction de répartition notée F de X ?
3. Calculer l'espérance et la variance de X.

Partie II.
Un corps radioactif est constitué à l'instant $t = 0$ de N atomes. Un atome de ce corps se désintègre au bout d'un temps aléatoire T (en années) suivant une loi exponentielle de paramètre k.
1) Quelle est l'espérance de vie d'un atome ?
2) Calculer la probabilité pour qu'un atome vive au moins une année.
3) Calculer la probabilité $p(t)$ pour qu'un atome se désintègre dans l'intervalle de temps $[0, t]$.
4) On suppose que les N atomes du corps se désintègrent indépendamment les uns des autres et on note X_t la variable aléatoire : "nombre d'atomes s'étant désintégrés dans l'intervalle de temps $[0, t]$". Donner la loi de probabilité de X_t.

Exercice 7 *La loi uniforme*
Soient a et b deux réels, $a < b$. On considère la fonction f définie sur $\mathbb{R}$ par

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

1. Faire la représentation graphique de f.
2. Vérifier que f est la densité de probabilité d'une variable aléatoire X. On dit alors que X suit une loi uniforme sur l'intervalle $[a, b]$.
3. Quelle est la fonction de répartition notée F de X ?
4. Calculer l'espérance et la variance de X.

Exercice 8 *La loi normale*
Soit X une variable aléatoire suivant une loi normale d'espérance μ et de variance σ^2, notée $\mathcal{N}(\mu, \sigma^2)$.

1. Donner la densité de X.
2. Donner la fonction de répartition sous forme intégrale de cette variable aléatoire.
3. Soit $Y = \frac{X-\mu}{\sigma}$, calculer la fonction de répartition de Y. Que peut-on en conclure ?

Exercice 9 Soit X une variable aléatoire de loi normale de moyenne 1 et d'écart type 2.

1. Calculer $P(X < 1,6)$.
2. Trouver u tel que $P(X < u) = 0,95$.

Exercice 10 *Dans le cadre où on dispose de n variables aléatoires $X_1, ..., X_n$ indépendantes et identiquement distribuées suivant une loi normale, alors, la somme de ces variables aléatoires suit aussi une loi normale.*

Un avion long courrier pèse 120 t à vide. Les consignes de sécurité lui interdisent de décoller s'il pèse plus de 130,5 t. Il emporte 100 passagers. On suppose que le poids d'un passager et de ses bagages suit une loi normale de moyenne 100 kg et d'écart type 20 kg. Quelle est la probabilité de décollage ?

Exercice 11 Une usine utilise une machine automatique pour remplir des flacons contenant un certain produit en poudre. Par suite de variations aléatoires dans le mécanisme, le poids de poudre par flacon est une v.a. de loi normale de moyenne m et d'écart-type 1,1 mg. Les flacons sont vendus comme contenant 100 mg de produit.

1. La machine est réglée sur $m = 101,2$ mg. Quelle est la probabilité que le poids de produit dans un flacon soit inférieur au poids annoncé de 100 mg ?
2. Sur quelle valeur de m faut-il régler la machine pour qu'au plus 4% des flacons aient un poids inférieur au poids annoncé de 100 mg ?

Exercice 12 *Loi de la moyenne*
On considère n variables aléatoires X_i indépendantes deux à deux et de même loi. On ne donne pas explicitement cette loi mais on sait que $E(X_i) = \mu$ et que $Var(X_i) = \sigma^2$. On considère la variable Z moyenne des X_i.

1. Déterminer l'espérance de Z.
2. Déterminer la variance de Z.
3. Construire à partir de Z une autre variable qui soit centrée et réduite.