PROBABILITÉS

Table des matières

Contextualisation

Un outil informatique est souvent confronté à des situations incertaines. Il est par exemple conçu pour répondre à des requêtes utilisateurs, mais ni la fréquence de ces requêtes ni leurs complexités ne sont connues au moment de la conception. Au mieux sont-elles estimées, mais alors on est déjà dans un contexte incertain. Dans de telles situations, il faut faire intervenir un modèle probabiliste.

Il est alors fréquent que les variables aléatoires impliquées puissent prendre un nombre infini de valeurs. Si tous les entiers, par exemple, sont des valeurs possibles, il faut considérer les probabilités de chaque valeur élémentaire $n \in \mathbb{N}$. On définit ainsi la distribution, qu'on utilise de la même manière que dans le cas fini, à la différence près que les sommes qui interviennent sont maintenant des sommes infinies. Ainsi, les espérances et variances sont des limites de séries numériques.

Dans la pratique, il faudrait stocker en mémoire toute la distribution. Ce serait un tableau de taille infinie contenant la probabilité d'occurrence de chaque valeur, ce qui n'est pas réalisable. Il faut donc condenser l'information contenue dans la distribution. Les notions d'espérance et de variance peuvent servir à cela. Celle de fonction génératrice peut aussi être efficace.

De nombreuses situations font appel à des variables aléatoires infinies entières :

- nombre de requêtes sur un serveur dans un intervalle de temps donné ;
- nombre de fois où il faut tenter une expérience avant de la réussir ;
- complexité d'un algorithme, nombre d'opérations élémentaires exigé par son exécution ;
- etc.

Attendus

Voilà les compétences que vous devez acquérir, sur lesquelles vous serez évalués :

- comprendre la notion de fonction génératrice et sa relation à la distribution d'une variable aléatoire ;
- savoir traiter des variables aléatoires prenant un nombre infini de valeurs entières ;
- savoir associer des situations concrètes à des distributions connues.

Les sections qui suivent sont consacrées à des points particuliers de ces compétences :

Fonction génératrice d'une variable aléatoire entière finie :

- comprendre la relation entre fonction génératrice et distribution ;
- savoir basculer d'un raisonnement « distribution » à un raisonnement « fonction génératrice » pour résoudre une question de probabilité.

Séries entières :

- relever les erreurs de raisonnement qui peuvent découler d'une utilisation abusives des propriétés spécifiques aux séries entières ;
- savoir calculer le rayon de convergence d'une série entière ;
- connaître et manipuler les séries de référence.

Variable aléatoire entière infinie :

- étendre des raisonnements, familiers pour les variables aléatoires finies, au cas infini ;
- savoir utiliser les séries génératrices pour répondre à des questions de probabilités ;
- associer des situations concrètes à des distributions infinies connues.

1 Fonction génératrice d'une variable aléatoire entière finie

1.1 Résumé

Quand on combine plusieurs variables aléatoires, il peut être difficile de déterminer la distribution de la variable résultante par un raisonnement purement probabiliste. La notion de fonction génératrice permet parfois de résoudre efficacement cette question, par des opérations élémentaires sur les polynômes.

La fonction génératrice d'une variable aléatoire contient toute l'information donnée par sa distribution. Cette information est simplement donnée autrement, donc utilisable autrement.

1.2 Exercices

Exercice 1.1

Soit X le résultat du lancé d'un dé équilibré.

1. Quelle est la fonction génératrice G_X de la variable X ?
2. En déduire $\mathrm{E}(X)$ et $\mathrm{Var}(X)$.

Exercice 1.2

Soit X une variable aléatoire à valeurs entières dont la fonction génératrice est

$$G_X(t) = a(2t+1)^2$$

1. Quelle est la valeur de a ?
2. Déterminer la distribution de X.
3. Déterminer son espérance et sa variance.

Exercice 1.3

Soient $(p_x, p_y) \in [0,1]^2$, X et Y deux variables de Bernoulli de paramètres p_x et p_y, c'est-à-dire qu'elles sont à valeurs dans $\{0,1\}$ et que

$$P(X{=}1) = p_x \qquad \text{et} \qquad P(Y{=}1) = p_y$$

Ces deux variables sont supposées indépendantes.

1. Déterminer leurs fonctions caractéristiques G_X et G_Y.
2. Existe-t-il des valeurs p_x et p_y telles que $X+Y$ soit uniformément distribuée sur $\{0,1,2\}$?

Exercice 1.4

Soient $n \in \mathbb{N}^*$, $p \in [0,1]$, puis n variables $X_1, X_2, \cdots, X_n$ de Bernoulli indépendantes de paramètre p. Soit $Y = X_1 + \cdots + X_n$. On rappelle que Y suit alors une loi binomiale de paramètres n et p :

$$Y \rightsquigarrow \mathrm{B}(n,p)$$

1. Déterminer les fonctions génératrices G_{X_i} pour $i \in [\![1, n]\!]$.
2. Déterminer la fonction génératrice G_Y de la variable Y.
3. En déduire la distribution de Y, puis $\mathrm{E}(Y)$ et $\mathrm{Var}(Y)$.

2 Séries entières

2.1 Résumé

Quand une variable aléatoire peut prendre une infinité de valeurs entières, sa fonction génératrice n'est plus un polynôme mais une série entière.

Avant d'utiliser ces objets dans le cadre probabiliste, il faut savoir les manier et les comprendre.

2.2 Exercices

Exercice 2.5

Soient les séries $\sum f_n$ et $\sum g_n$ dont les termes généraux sont définis par

$$\forall x \in [0,2], f_n(x) = \frac{x^{n+1}}{1+x^{n+1}} - \frac{x^n}{1+x^n} \qquad \text{et} \qquad \forall x \in \mathbb{R}, g_n(x) = \sqrt{x^2 + \frac{1}{n+1}} - \sqrt{x^2 + \frac{1}{n}}$$

1. Vérifier que les fonctions f_n et g_n sont continues de dérivables.
2. Montrer que pour tout $x \in [0,2]$, la série numérique $\sum f_n(x)$ converge.
   Expliciter la fonction $F : x \longmapsto \sum_{n=0}^{+\infty} f_n(x)$.
3. La fonction F est-elle continue sur $[0,2]$? Dérivable ?
4. Montrer que pour tout $x \in \mathbb{R}$, la série numérique $\sum g_n(x)$ converge.
   Expliciter la fonction $G : x \longmapsto \sum_{n=1}^{+\infty} g_n(x)$.
5. La fonction G est-elle continue sur $\mathbb{R}$? Dérivable ?

Exercice 2.6

1. Rappeler le développement en série entière $\sum a_n x^n$ de la fonction $f : x \longmapsto \dfrac{1}{1-x}$ et son rayon de convergence r.
2. Pour tout $x \in]-r, r[$, on définit la suite numérique $\big(R_n(x)\big)$ des restes de la série numérique $\sum a_n x^n$:
   $$R_n(x) = f(x) - \sum_{k=0}^{n} a_k x^k = \sum_{k=n+1}^{+\infty} a_k x^k$$
   a. Soit $x \in]-r, r[$. A-t-on $R_n(x) = o(x^n)$ quand n tend vers $+\infty$?
   b. Soit $n \in \mathbb{N}$. A-t-on $R_n(x) = o(x^n)$ quand x tend vers 0 ?

Exercice 2.7

Déterminer le rayon de convergence des séries entières suivantes :

1. $\sum \dfrac{2^n}{n!} x^n$
2. $\sum \dfrac{(n!)^2}{(2n)!} x^n$
3. $\sum (-1)^n n x^n$

Exercice 2.8

Déterminer les développements en série entière en 0, avec leurs rayons de convergence, des fonctions suivantes :

1. $\dfrac{1}{(1-x)^r}$ où $r \in \mathbb{N}$
2. $-\dfrac{\ln(1-x)}{1-x}$
3. $\big(\ln(1-x)\big)^2$

Exercice 2.9

Déterminer sous forme condensée les sommes suivantes et leurs rayons de convergence :

1. $\displaystyle\sum_{n=1}^{+\infty} \frac{x^n}{n}$
2. $\displaystyle\sum_{n=1}^{+\infty} nx^n$
3. $\displaystyle\sum_{n=0}^{+\infty} n^2x^n$

3 Variable aléatoire entière infinie

3.1 Résumé

Les variables aléatoires rencontrées dans la vie prennent souvent une infinité de valeurs. Un cas fréquent est celui où les valeurs possibles sont toutes des nombres entiers.
Ainsi, par exemple :

– le nombre de requêtes parvenant à un serveur dans un intervalle de temps donné prend des valeurs entières qui ne peuvent pas être bornées à priori ;
– il en est de même du nombre de fois une expérience aléatoire doit être répétée avant de se conclure par un succès.

Des variables de ce type se traitent d'une façon similaire aux variables finies, mais les sommes deviennent des sommes infinies, c'est-à-dire des limites de séries.

3.2 Exercices

Exercice 3.10 *loi géométrique*

Un hameçonneur envoie des emails à des individus tirés au hasard, les incitant à donner en ligne leurs coordonnées bancaires. Pour un email envoyé, la probabilité que l'individu visé se fasse piéger est p. On suppose que les réponses des individus sont indépendantes.
Soit X le nombre d'emails envoyés avant que le hameçonneur obtienne sa première réponse.

1. Quelle est la distribution de X ? La représenter graphiquement.
2. Calculer $P(X \in [\![3, 6]\!])$ puis $P(X \geqslant 5)$.
3. Déterminer la fonction génératrice G_X de X.
4. En déduire son espérance et sa variance.

Remarque : on dit alors que X suit une loi géométrique de paramètre p, ce qu'on note

$$X \rightsquigarrow \text{Géom}(p)$$

Exercice 3.11 *loi de Pascal*

Soit $r \in \mathbb{N}^*$. On considère le hameçonneur de l'exercice précédent et on définit Y comme le nombre d'emails qu'il doit envoyer pour obtenir r réponses.

1. Écrire Y comme une somme de variables géométriques.
2. Déterminer la fonction génératrice G_Y de Y.
3. En déduire la distribution de Y, son espérance et sa variance.

Exercice 3.12 *loi de Poisson*

La sécurité routière a constaté sur une année deux accidents corporels en moyenne par quart d'heure. On note X le nombre d'accidents corporels dans le prochain intervalle de temps d'un quart d'heure.

On suppose que, à chaque instant t, la date du prochain accident est indépendante du nombre d'accidents précédents et de leurs dates, et on admet que, moyennant cette hypothèse, il existe $\lambda \in \mathbb{R}_+^*$ tel que

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

1. Représenter graphiquement la distribution de X.
2. Déterminer sa fonction génératrice.
3. En déduire son espérance et sa variance.
4. En tenant compte des observations de la sécurité routière, déterminer λ et en déduire la probabilité des évènements :
   a. aucun accident corporel dans le prochain quart d'heure ;
   b. au moins 3 accidents corporels dans le prochain quart d'heure ;
   c. aucun accident corporel dans l'heure qui vient.

Remarque : on dit alors que X suit une loi de Poisson de paramètre λ, ce qu'on note

$$X \rightsquigarrow \text{Poisson}(\lambda)$$

Exercice 3.13

Soient $(\lambda, \mu) \in \left(\mathbb{R}_+^*\right)^2$, X et Y deux variables aléatoires indépendantes telles que

$$X \rightsquigarrow \text{Poisson}(\lambda) \qquad \text{et} \qquad Y \rightsquigarrow \text{Poisson}(\mu)$$

1. En utilisant les fonctions génératrices de X et Y, déterminer la distribution de $X + Y$.
2. Dans la situation de l'exercice précédent, déterminer la probabilité des évènements :
   a. deux accidents en une demi-heure ;
   b. deux accidents en une heure.

Exercice 3.14

Soient $\lambda \in \mathbb{R}_+^*$ et Y une variable aléatoire telle que $Y \rightsquigarrow \text{Poisson}(\lambda)$.

Pour tout $n \in \mathbb{N}$, on considère une variable binomiale $X_n \rightsquigarrow \mathrm{B}\left(n, \dfrac{\lambda}{n}\right)$.

1. Rappeler les fonctions génératrices G_Y et G_{X_n} des variables Y et X_n.
2. Soit $t \in \mathbb{R}$. Montrer que $\lim\limits_{n\to+\infty} G_{X_n}(t) = G_Y(t)$.

Remarque : on admet que cette propriété implique que «la suite (X_n) converge en loi vers la variable Y», c'est-à-dire que

$$\forall A \subset \mathbb{N}, \quad P(X_n{\in}A) \xrightarrow[n\to+\infty]{} P(Y{\in}A)$$