PROBABILITÉS ET STATISTIQUE
5. Variables aléatoires discrètes

Carmelo Vaccaro

Université de Paris-Saclay

M1 - Informatique Science des Données (ISD) - Apprentissage
2022/23: premier semestre

Contenu du chapitre

1. Variables aléatoires et distributions de probabilité

Variables aléatoires

Une *variable aléatoire (réelle)* est une variable qui prend une valeur numérique pour chaque événement simple.

La définition formelle est la suivante. Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilités. Une fonction $X : \Omega \to \mathbb{R}$ est une *variable aléatoire (réelle)* si pour tout $a, b \in \mathbb{R}$ tels que $a \leq b$ on a que $X^{-1}(]a, b]) \in \mathcal{A}$.

Cela veut dire qu'on peut calculer la probabilité de l'image inverse par rapport à X de tout intervalle ouvert à gauche et fermé à droite.

On peut prouver que si X est une variable aléatoire, alors l'image inverse de tout intervalle (ouvert/fermé à gauche, ouvert/fermé à droite, fini ou infini, singleton) appartient à $\mathcal{A}$.

Distributions de probabilité

La *distribution de probabilité* d'une variable aléatoire donne la probabilité de chaque valeur de la variable aléatoire.

La définition formelle est la suivante. Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilités et $X : \Omega \to \mathbb{R}$ est une variable aléatoire. Soient $a, b \in \mathbb{R}$ tels que $a \leq b$. Alors par définition l'image inverse des intervalles suivants appartient à $\mathcal{A}$ et donc on peut calculer la probabilité :

$$]a, b[,\ ]a, b],\ [a, b[,\ [a, b],\ \{a\}, [a, +\infty[,\ ]a, +\infty[,\ ]-\infty, a[,\ ]-\infty, a].$$

La *distribution de probabilité* de X est la fonction qui donne la probabilité de ces intervalles.

Notation pour les distributions de probabilités

Normalement on dénote les distributions de probabilités de la façon suivante :

$P\Big(X^{-1}(]a,b[)\Big)$ est dénoté $P(a < X < b)$;

$P\Big(X^{-1}(]a,b])\Big)$ est dénoté $P(a < X \leq b)$;

... ;

$P\Big(X^{-1}(a)\Big)$ est dénoté $P(X = a)$;

$P\Big(X^{-1}([a,+\infty[)\Big)$ est dénoté $P(X \geq a)$;

...

Propriétés des distributions de probabilité

1. Les valeurs des probabilités sont comprises entre 0 et 1, c'est-à-dire que $0 \leq P(X = x) \leq 1$ pour chaque valeur x que la variable aléatoire peut assumer.

2. La somme de toutes les probabilités doit être égale à 1, c'est-à-dire $\sum P(X = x) = 1$ où la somme porte sur toutes les valeurs x que la variable aléatoire peut assumer.

Variables aléatoires discrètes et continues (1/2)

Soit X une variable aléatoire, soient a, b deux nombre réels tels que $a \leq b$. Nous voulons calculer la probabilité que X soit comprise entre a et b, $P(a \leq X \leq b)$.

Nous avons que $P(a \leq X \leq b)$ est égale à la somme de tous les termes $P(X = c)$ pour tous les nombres réels c tel que $a \leq c \leq b$.

Il y a deux cas spéciales qui méritent attention :

1. X assume un nombre fini ou infini discret de valeurs : on dit que X est une variable aléatoire *discrète*.
2. Pour tout nombre réel d on a que $P(X = d)$ est égale à 0 : on dit que X est une variable aléatoire *continue*.

Variables aléatoires discrètes et continues (2/2)

Si X est une variable aléatoire discrète, alors

$$P(a \leq X \leq b) = \sum_{x: a \leq x \leq b} P(X = x),$$

où la somme est une série si le nombre de valeurs assumé par X est infini.

Si X est une variable aléatoire continue, alors on peut montrer qu'elle admet une *densité de probabilité*, c'est-à-dire il existe une fonction f, appelée la *densité de probabilité de* X, telle que

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Variables aléatoires discrètes et continues : remarques

Remarque. On peut montrer qu'une variable aléatoire non-constante ne peut pas être discrète et continue en même temps.

Remarque. Il y a des variables aléatoires qui ne sont ni discrètes ni continues.

Remarque. On peut montrer qu'une variable aléatoire continue non-constante assume une infinité continue de valeurs.

Remarque. Il y a des variables aléatoires qui assument une infinité continue de valeurs mais qui ne sont pas continues d'après la définition.

Exemples de variables aléatoires discrètes

1. Le nombre d'œufs qu'une poule pond en une journée est une variable aléatoire discrète car ses seules valeurs possibles sont 0, ou 1, ou 2, et ainsi de suite. Aucune poule ne peut pondre 2,343115 œufs, ce qui aurait été possible si les données provenaient d'une échelle continue.
2. Le nombre d'étudiants présents en classe un jour donné est un nombre entier et est donc une variable aléatoire discrète.

Exemples de variables aléatoires continues

1. La quantité de lait produite par une vache en une journée est une variable aléatoire continue car elle peut avoir n'importe quelle valeur sur un intervalle continu. Au cours d'une seule journée, une vache peut produire une quantité de lait dont la valeur peut être comprise entre 0 et 20 litres. Il serait possible d'obtenir 4,123456 litres, car la vache n'est pas limitée aux quantités discrètes de 0, 1, 2, 3, 4 ou 5 litres.
2. La mesure de la tension pour une pile AA peut être n'importe quelle valeur comprise entre 1,1 et 1,5 volts. Il s'agit donc d'une variable aléatoire continue.

2. Variables aléatoires discrètes

Moyenne, variance et écart type d'une distribution de probabilité discrète

On peut montrer qu'on peut calculer la moyenne et la variance d'une variable aléatoire à partir de sa distribution de probabilité. En particulier

- La moyenne μ est égale à $\sum(x \cdot P(X = x))$, où la somme porte sur toutes les valeurs x que la variable aléatoire peut assumer.
- La variance est égale à $\sum(x - \mu)^2 \cdot P(X = x) = \sum \left(x^2 \cdot P(X = x)\right) - \mu^2$.
- L'écart-type est égal à la racine carrée de la variance.

Ces sommes sont des séries si la variable aléatoire assume un nombre infini discret de valeurs ou des intégrales si la la variable aléatoire est continue.

Exemple

Trouvez la moyenne, la variance et l'écart-type de la distribution de probabilité donnée dans le tableau suivant

x	$P(x)$
0	0.001
1	0.015
2	0.088
3	0.264
4	0.396
5	0.237

Solution

La moyenne est

$$0 \cdot 0{,}001 + 1 \cdot 0{,}015 + 2 \cdot 0{,}088 + 3 \cdot 0{,}264 + 4 \cdot 0{,}396 + 5 \cdot 0{,}237 = 3{,}752.$$

La variance est

$$(0 - 3{,}752)^2 \cdot 0{,}001 + (1 - 3{,}752)^2 \cdot 0{,}015 + (2 - 3{,}752)^2 \cdot 0{,}088 +$$

$$(3-3{,}752)^2 \cdot 0{,}264 + (4-3{,}752)^2 \cdot 0{,}396 + (5-3{,}752)^2 \cdot 0{,}237 \approx 0{,}941 \approx 0{,}94.$$

L'écart type est

$$\sqrt{0{,}941} \approx 0{,}97.$$

Exercice

Calculez la moyenne μ, la variance σ^2 et l'écart-type σ de la distribution de probabilité suivante :
$S = \{0{,}5,\ 2{,}5,\ 3\}$ et $P(0{,}5) = \frac{1}{3}$, $P(2{,}5) = \frac{1}{3}$, $P(3) = \frac{1}{3}$.

3. Variables aléatoires binomiales

Variables aléatoires binomiales : définition

Soit $X : \Omega \to \mathbb{N}$ une variable aléatoire (qui est discrète car l'ensemble d'arrivée est constitué par les nombres naturels) et soient $n \in \mathbb{N}$, $p \in [0, 1]$

On dit que X est une **variable aléatoire binomiale de paramètres** n **et** p si pour tout $k \in \mathcal{N}$ on a que

$$P(X = k) = \frac{n!}{(n-k)!k!}\, p^k (1-p)^{n-k}.$$

Variables aléatoires binomiales : remarque

Une variable aléatoire binomiale est définie par sa distribution.

Variables aléatoires binomiales : modélisation

Une variable aléatoire binomiale résulte d'une procédure qui répond à toutes les exigences suivantes :

1. La procédure a un nombre fixe d'essais, égal à n.
2. Les essais doivent être indépendants. (Le résultat d'un essai individuel n'affecte pas les probabilités des autres essais).
3. Tous les résultats de chaque essai doivent être classés en deux catégories (communément appelées *succès* et *échec*).
4. La probabilité d'un succès reste la même dans tous les essais.

Notation pour la distribution de probabilité binomiale

- S et E (succès et échec) désignent les deux catégories possibles de tous les résultats.
- $P(S) = p$ ($p =$ probabilité de réussite)
- $P(E) = 1 - p$ ($1 - p =$ probabilité d'un échec)
- n désigne le nombre fixe d'essais.
- k indique un nombre spécifique de succès dans n essais, donc k peut être n'importe quel nombre entier entre 0 et n, inclus.
- $P(k)$ désigne la probabilité d'obtenir exactement k succès parmi les n essais.

Exemple

Considérons une expérience dans laquelle 5 descendants de petits pois sont générés à partir de 2 parents ayant chacun la combinaison de gènes vert/jaune pour la couleur. La probabilité qu'un petit pois de la descendance ait une couleur verte est de 0,75, c'est-à-dire que $P(\text{verte}) = 0{,}75$.

Supposons que nous voulions trouver la probabilité qu'exactement 3 des 5 petit pois de la descendance aient une couleur verte.

1. Cette procédure donne-t-elle lieu à une distribution binomiale ?
2. Si cette procédure aboutit à une distribution binomiale, identifiez les valeurs de n, k et p.

Solution de 1.

Cette procédure satisfait aux exigences d'une distribution binomiale, comme indiqué ci-dessous.

1. Le nombre d'essais, 5, est fixé.
2. Les 5 essais sont indépendants, car la probabilité qu'un petit pois de la descendance ait une couleur verte n'est pas affectée par le résultat d'un autre pois de la descendance.
3. Chacun des cinq essais comporte deux catégories de résultats : le petit pois a une couleur verte ou pas.
4. Pour chaque petit pois de la descendance, la probabilité qu'il ait une couleur verte est de 3/4 ou 0,75, et cette probabilité reste la même pour chacun des 5 pois.

Solution de 2.

1. Avec 5 petits pois, nous avons $n = 5$.
2. Nous voulons connaître la probabilité d'avoir exactement 3 petits pois avec verts, donc $k = 3$.
3. La probabilité de succès (obtenir un pois avec une cosse verte) pour une sélection est de 0,75, donc $p = 0{,}75$.

Exemple

En supposant que la probabilité qu'un petit pois soit de couleur verte est de 0,75, utilisez la formule de probabilité binomiale pour trouver la probabilité d'obtenir exactement 3 petits pois verts lorsque 5 petits pois de la descendance sont générés.

Autrement dit, trouvez $P(\mathrm{k})$ avec $n = 5$, $k = 3$ et $p = 0{,}75$.

Solution

La formule,

$$P(k) = \frac{n!}{(n-k)!k!} \, p^k (1-p)^{n-k}$$

avec les valeurs $n = 5$, $k = 3$ et $p = 0{,}75$ donne

$$P(3) = \frac{5!}{(5-3)!3!} \, 0{,}75^3 \cdot 0{,}25^{5-3}$$

$$\frac{5!}{2!3!} \, 0{,}75^3 \cdot 0{,}25^2 = 10 \cdot 0{,}421875 \cdot 0{,}0625 \approx 0{,}264.$$

Exercice

La chaîne de restauration rapide McDonald's a un taux de reconnaissance de la marque de 95% dans le monde. En supposant que nous choisissions 4 personnes au hasard, trouvez la probabilité qu'exactement 2 des 4 personnes reconnaissent McDonald's.

Moyenne, variance et écart type de la distribution binomiale

- **Moyenne :** np.
- **Variance :** $np(1-p)$.
- **Écart-type :** $\sqrt{np(1-p)}$.

Exemple

Trouvez la moyenne, la variance et l'écart-type du nombre de petits pois de couleur verte lorsque des groupes de 5 pois sont générés. Supposez qu'il existe une probabilité de 0,75 qu'un petit pois de la descendance soit de couleur verte.

Autrement dit, trouvez $P(\mathrm{k})$ avec $n = 5$, $k = 3$ et $p = 0{,}75$.

Solution

On a $n = 5$ et $p = 0{,}75$, donc

$$\mu = np = 5 \cdot 0{,}75 = 30{,}75,$$

$$\sigma^2 = np(1-p) = 5 \cdot 0{,}75 \cdot 0{,}25 = 0{,}9375$$

et

$$\sigma = \sqrt{0{,}9375} \approx 0{,}968.$$

Solution

On a $n = 4$ et $p = 0{,}95$, donc

$$\mu = np = 4 \cdot 0{,}95 = 3{,}8,$$

$$\sigma^2 = np(1 - p) = 4 \cdot 0{,}95 \cdot 0{,}05 = 0{,}19,$$

et

$$\sigma = \sqrt{0{,}19} \approx 0{,}44.$$

4. Variables aléatoires de Poisson

Variables aléatoires de Poisson (1/2)

Une *variable aléatoire de Poisson* est une autre variable aléatoire discrète.

Les variables aléatoires de Poisson sont souvent utilisées pour décrire le comportement d'événements rares (avec de petites probabilités), comme par exemple :

- la désintégration radioactive,
- l'arrivée de personnes dans une file d'attente,
- la nidification d'aigles dans une région,
- l'arrivée de patients dans une salle d'urgence,
- les accidents sur une autoroute,
- des internautes se connectant à un site web.

Variables aléatoires de Poisson : définition

Soit $X : \Omega \to \mathbb{N}$ une variable aléatoire (qui est discrète car l'ensemble d'arrivée est constitué par les nombres naturels) et soit λ un réel positif.

On dit que X est une **variable aléatoire de Poisson de coefficient** λ si pour tout $k \in \mathcal{N}$ on a que

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Variables aléatoires de Poisson : remarques

Une variable aléatoire de Poisson est définie par sa distribution.

La valeur k représente en général le nombre d'occurrences d'un événement dans un intervalle, qui être le temps, la distance, la surface, le volume ou une autre unité.

Variables aléatoires de Poisson : modélisation

Une variable aléatoire de Poisson résulte d'une procédure telle que :

1. k représente le nombre d'occurrences d'un événement *sur un certain intervalle*.
2. Les occurrences doivent être *aléatoires*.
3. Les occurrences doivent être *indépendantes* les unes des autres.
4. Les occurrences doivent être *uniformément distribuées* sur l'intervalle utilisé, c'est-à-dire que la probabilité d'occurrence doit être proportionnelle à la taille de l'intervalle.

Moyenne, variance et écart-type pour la distribution de Poisson

La moyenne et la variance sont égales à λ, l'écart type est égal à $\sqrt{\lambda}$.

Différence entre une distribution de Poisson et une distribution binomiale

La distribution binomiale est affectée par la taille de l'échantillon n et la probabilité p, tandis que la distribution de Poisson est affectée uniquement par le paramètre λ.

Dans une distribution binomiale, les valeurs possibles de la variable aléatoire X sont $0, 1, \cdots, n$, mais pour une distribution de Poisson, les valeurs possibles de X sont $0, 1, 2, \cdots$, sans limite supérieure.

Exemple

Supposons que le nombre de tremblements de terre majeurs dans le monde par an soit modélisé par une distribution de Poisson avec $\lambda = 0{,}93$.

1. Si *P(k)* est la probabilité d'avoir k tremblements de terre majeurs au cours d'une année choisie au hasard, trouvez *P(0)*, *P(1)* et *P(2)*.
2. Trouvez le nombre moyen de tremblements de terre majeurs par an.

Solution de 1

La formule

$$P(k) = \frac{\lambda^k\, e^{-\lambda}}{k!},$$

avec $\lambda = 0{,}93$ et $k = 0, 1, 2$ donne

$$P(0) = \frac{(0{,}93)^0\, e^{-0{,}93}}{0!} = 0{,}395,$$

$$P(1) = \frac{(0{,}93)^1\, e^{-0{,}93}}{1!} = 0{,}367,$$

et

$$P(2) = \frac{(0{,}93)^2\, e^{-0{,}93}}{2!} = 0{,}171.$$

Solution de 2

Le nombre moyen de tremblements de terre majeurs par an est de $\lambda = 0{,}93$.

Exercice

Trouvez la probabilité $P(3)$ pour une distribution de Poisson de $\lambda = 3$.

Fin.