Processus de Markov : formulaire de probabilités

Gabriel Lang

Université Paris-Saclay, AgroParisTech

MAEE, master BEE
Math 1, master EvoGem

Loi de Bernoulli

La variable X vaut 0 avec probabilité $1 - p$ ou 1 avec probabilité p.

Propriétés

- $\mathbb{E}(X) = 0 \times \mathbb{P}(X = 0) + 1 \times \mathbb{P}(X = 1) = p$
- $\mathbb{E}(X^2) = 0 \times \mathbb{P}(X^2 = 0) + 1 \times \mathbb{P}(X^2 = 1) = p$
- $\text{Var}\,(X) = p - p^2 = p(1 - p)$

Loi binomiale

On tire n variables de Bernoulli X_i indépendantes de paramètre p. La variable X est la somme de ces variables.
Ex : roulette, pari sur Noir, nombre de paris gagnants dans une série de n coups.

Propriétés

- $\mathbb{P}(X = k) = \frac{n!}{k!(n-k)!}p^k(1-p)^k$
- $\mathbb{E}(X) = n \times \mathbb{E}(X_i) = np$
- $\text{Var}\,(X) = n\text{Var}\,(X_i) = np(1-p)$

Loi geométrique

On répète le tirage indépendant de variables de Bernoulli $B(p)$ jusqu'au premier succès à la date G.

Définition de la loi géométrique de probabilité de succès p

$\mathbb{P}(G=k)=p(1-p)^{k-1}$, $\mathbb{P}(G\geq k)=(1-p)^k$, $E(G)=1/p$ et $\mathrm{Var}\,(G)=\frac{1-p}{p^2}$.

Propriétés

Soit $G=\min(G_1,G_2)$ où G_1 et G_2 indépendantes géométriques de paramètre p_1 et p_2. Alors G est une loi géométrique de paramètre $p_1+p_2-p_1p_2$. De plus
$\mathbb{P}(G=G_1)=p_1/(p_1+p_2-p_1p_2)$ et $\mathbb{P}(G=G_2)=p_2/(p_1+p_2-p_1p_2)$

Preuve : $\mathbb{P}(G\geq k)=\mathbb{P}(G_1\geq k)\mathbb{P}(G_2\geq k)=((1-p_1)(1-p_2))^k$. D'où
$p=1-(1-p_1)(1-p_2)$.
$\mathbb{P}(G_1=k,G_2>k)=p_1(1-p_1)^{k-1}(1-p_2)^k$, $\mathbb{P}(G_1=k,G_2=k)=p_1p_2(1-p_1)^{k-1}(1-p_2)^{k-1}$.
$\mathbb{P}(G=G_1=k)=p_1((1-p_1)(1-p_2))^{k-1}$ donc $\mathbb{P}(G_1=k|G=k)=p_1/(p_1+p_2-p_1p_2)$.

Loi binomiale négative

La variable X de loi binomiale négative $BN(n,p)$ est le temps nécessaire pour avoir gagné n fois dans un jeu de pile ou face répété. On tire n variables de géométrique G_i indépendantes de paramètre p. X est la somme de ces variables.
Ex : Modèle de naissance pure à partir d'une population de n individus.

Propriétés

- Pour $k \geq n$, $\mathbb{P}(X=k) = \begin{pmatrix} k-1 \\ n-1 \end{pmatrix} p^k (1-p)^{n-1}$
- $\mathbb{E}(X) = n \times \mathbb{E}(X_i) = \frac{n}{p}$
- $\mathrm{Var}\,(X) = n\mathrm{Var}\,(X_i) = \frac{n(1-p)}{p^2}$

Loi binomiale négative

Preuve : X étant la somme de n variables géométriques indépendantes :

$$\mathbb{P}(X=k) = \sum_{k_1+\ldots+k_n=k} p(1-p)^{k_1} \times \cdots \times p(1-p)^{k_n} = \sum_{k_1+\ldots+k_n=k} p^n(1-p)^k.$$

Il faut compter le nombre de façon de décomposer k en n nombres. Cela revient à déposer n objets identiques dans k cases. Chaque configuration peut être lue comme une suite alternée d'objets et de séparateurs entre cases. Il y a k objets et $n-1$ séparateurs de cases, ce qui fait $(n+k-1)!$ ordres possibles. Cependant les $n-1$ séparateurs et les k objets sont identiques et leurs permutations donnent la même décomposition du nombre k. Il faut donc diviser par $k!(n-1)!$. Le nombre de décomposition est donc $(n+k-1)!/k!(n-1)!$.
Ce n'est pas le bon compte car tous les nombres k_i doivent être strictement positifs : aucune case ne doit être vide. On prélève donc n objets qu'on distribue à raison d'un pour chaque case. Les $k-n$ autres sont repartis comme précédemment avec les $n-1$ séparateurs. Il y a donc $(k-1)!/(k-n)!(n-1)!$ décompositions.

Variables de Bernoulli dépendantes

Trio de variables de Bernoulli dépendantes

On définit un triplet de variables de Bernoulli (e_1, e_2, e_3) tel que $\{e_i = 1\} = \{X = i\}$.

- $\mathbb{E}(e_i) = p_i$
- $\mathrm{Var}\,(e_i) = p_i(1 - p_i)$.
- $\mathbb{E}(e_i e_j) = 0$ car ce produit est nul par définition.
- $\mathrm{Cov}\,(e_i, e_j) = -p_i p_j$.

Loi trinomiale

Soit une variable X à trois éventualités $\{1, 2, 3\}$ de probabilités (p_1, p_2, p_3).
Ex : roulette pari sur un carré de 12 nombres.
On tire n variables X_i indépendantes suivant la loi précédente; et on considère les n triplets $(e_{i,1}, e_{i,2}, e_{i,3})$ correspondants. Le triplet (v_1, v_2, v_3) est le nombre de variables valant respectivement 1,2 et 3.

Trio de variables binomiales dépendantes et loi trinomiale

- Les v_i sont des variables binomiales $B(n, p_i)$ dépendantes $(v_1 + v_2 + v_3 = n)$.
- La loi trinomiale est la loi jointe $\mathbb{P}(v_1 = k_1, v_2 = k_2, v_3 = k_3) = \frac{n!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}$ si $k_1 + k_2 + k_3 = n$.
- $\mathbb{E}(v_i) = np_i$.
- $\text{Var}\,(v_i) = np_i(1 - p_i)$.
- $\mathbb{E}(v_i v_j) = n\mathbb{E}(e_i e_j) = n(n-1)p_i p_j$.
- $\text{Cov}\,(v_i, v_j) = -np_i p_j$.

Loi trinomiale

On veut calculer $\mathbb{E}(v_1 v_2) = \sum_{k_1=0}^{n} \sum_{k_2=0}^{n-k_1} k_1 k_2 \frac{n!}{k_1!k_2!(n-k_1-k_2)!} p_1^{k_1} p_2^{k_2} p_3^{n-k_1-k_2}$.
Formule du trinôme :
$f(x,y,z) = (x+y+z)^n = \sum_{k_1=0}^{n} \sum_{k_2=0}^{n-k_1} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} x^{k_1} y^{k_2} z^{n-k_1-k_2}$.
On dérive cette expression par rapport à x puis par rapport à y:

$$n(n-1)(x+y+z)^{n-2} = \sum_{k_1=0}^{n} \sum_{k_2=0}^{n-k_1} k_1 k_2 \frac{n!}{k_1!k_2!(n-k_1-k_2)!} x^{k_1-1} y^{k_2-1} z^{n-k_1-k_2}$$

On remarque alors:

$$n(n-1)(x+y+z)^{n-2} xy = \sum_{k_1=0}^{n} \sum_{k_2=0}^{n-k_1} k_1 k_2 \frac{n!}{k_1!k_2!(n-k_1-k_2)!} x^{k_1} y^{k_2} z^{n-k_1-k_2}$$

Donc $\mathbb{E}(v_1 v_2) = n(n-1)(p_1+p_2+p_3)^{n-2} p_1 p_2 = n(n-1) p_1 p_2$.

Variables à k éventualités

Soit une variable X à k éventualités $\{1, \ldots, k\}$ de probabilités $(p_1, \ldots, p_k)$. On suppose que p_1 est non nul. On tire une suite de variables indépendantes X_i de même loi que X.

Loi géométrique de saut

On définit la variable Y comme le premier temps i où X_i est différent de 1. Donc Y suit une loi géométrique de probabilité de succès $1 - p_1$.

Au temps $i = Y$, la variable X_i devient différente de 1. Elle saute sur une des autres valeurs entre 2 et k.

Choix du saut

La probabilité du choix $X_i = j$ est $\frac{p_j}{1-p_1} = \frac{p_j}{\sum_{l \neq 1} p_l}$. Elle est proportionnelle à p_j.

Preuve : on définit les événements : $A = (X_i \neq 1)$, $B = (X_i = j)$.
On calcule $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(A)} = \frac{p_j}{1-p_1}$

Loi de Poisson

La loi de Poisson est une loi sur $\mathbb{N}$

Loi de Poisson de paramètre c

- $\mathbb{P}(X = k) = \frac{c^k}{k!} e^{-c}$
- $\mathbb{E}(X) = c$
- $\text{Var}\,(X) = c.$

Par définition de l'exponentielle : $\sum_{k=0} \infty \frac{c^k}{k!} = e^c$, donc la somme des probabilités vaut 1.
Pour calculer l'espérance

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \frac{k}{c^k} k! e^{-c} = e^{-c} \sum_{k=1}^{\infty} \frac{c^k}{(k-1)!}$$

$$= e^{-c} c \sum_{k=1}^{\infty} \frac{c^{k-1}}{(k-1)!} = e^{-c} c \sum_{k=0}^{\infty} \frac{c^k}{k!} = c$$

Loi de Poisson

De même

$$\begin{aligned}\mathbb{E}(X(X-1)) = &= \sum_{k=2}^{\infty} \frac{k}{(} k-1)c^k k! e^{-c} = e^{-c} \sum_{k=2}^{\infty} \frac{c^k}{(k-2)!} \\ &= e^{-c} c^2 \sum_{k=2}^{\infty} \frac{c^{k-2}}{(k-2)!} = e^{-c} c^2 \sum_{k=0}^{\infty} \frac{c^k}{k!} = c^2\end{aligned}$$

On conclut par $\operatorname{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2 = c^2 + c - c^2$.

Formule de Stirling

Développement de factorielle

$$n! = n^n e^{-n}\sqrt{2\pi n}(1 + 1/12n + O(n^{-2}))$$

La preuve consiste à écrire $\ln(n!) = \sum_{i=2}^{n} \ln(i)$ et à comparer à l'intégrale de $\ln(x)$.

Convergence de la loi binomiale vers la loi de Poisson

La preuve utilise la formule de Stirling pour la factorielle et la convergence suivante:

$$\left(1-\frac{b}{n}\right)^n = \exp\left(n\log\left(1-\frac{b}{n}\right)\right) = \exp\left(n\left(-\frac{b}{n}+O(n^{-2})\right)\right)$$
$$= \exp\left(-b+O(n^{-1})\right) = \exp\left(-b\right)+o(1)$$

On considère une variable X_n de loi binomiale $B(n, c/n)$. Alors quand n tend vers l'infini, la loi de X_n tend vers une loi de Poisson.

$$\mathbb{P}(X_n = k) = \frac{n!}{(n-k)!k!}\frac{c^k}{n^k}\left(1-\frac{c}{n}\right)^{n-k} = \frac{c^k}{k!}\frac{n^n e^{-n}}{(n-k)^{n-k}e^{-n+k}}(1+o(1))\frac{1}{n^k}\left(1-\frac{c}{n}\right)^{n-k}$$
$$= \frac{c^k}{k!}\frac{n^n e^{-n}}{\left(1-\frac{k}{n}\right)^{n-k} n^{n-k}e^{-n+k}}\frac{1}{n^k}\left(e^{-c}+o(1)\right)$$
$$= \frac{c^k}{k!}\frac{1}{\left(e^{-k}+o(1)\right)e^k}\left(e^{-c}+o(1)\right) = \frac{c^k e^{-c}}{k!}+o(1)$$

Loi de la date d'extinction de la descendance

Supposons qu'on connait la probabilité d'extinction en moins de t générations $\mathbb{P}(d=t)$. Soit v la variable égale au nombre d'enfants à la première génération. Si $v=j$, pour que l'extinction du lignage ait eu lie à l'instant t+1, il faut que le lignage de chaque enfant se soit éteint en un temps inférieur ou égal à t. Les extinctions sont indépendantes donc la proba d'extinction des j lignages est $(\mathbb{P}(d=t))^j$. Donc $\mathbb{P}(d=t+1|v=j)=(\mathbb{P}(d=t))^j$. Donc

$$\mathbb{P}(d=t+1)=\sum_{j=0}^{\infty}\mathbb{P}(d=t+1|v=j)\mathbb{P}(v=j)=\sum_{j=0}^{\infty}e^{-1}\frac{(\mathbb{P}(d=t))^j}{j!}=e^{\mathbb{P}(d=t)-1}$$