Probabilités, Statistiques, Combinatoire - CM 11

Adrian Tanasă

Université de Bordeaux - Licence Informatique

Notions vu le cours dernier

- loi jointe de deux v. a.
- indépendance de v. a.
- description d'une v. a. en fonction d'autres v. a.
- lois de Bernoulli, géométrique, binomiale, Poisson

Plan du cours d'aujourd'hui

- Retour sur la notion d'indépendance de v. a.
- Espérance et variance de variables aléatoires

Indépendance : définition

Définition

- Deux variables aléatoires X et Y sont **indépendantes** si, les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants :

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \ \forall x, y$$

- **Généralisation.** Des variables aléatoires $X_1, \dots, X_n$ sont indépendantes si, les événement $\{X_1 = a_1\}$, $\{X_2 = a_2\}$,... $\{X_n = a_n\}$ sont indépendants :

$$\mathbb{P}(X_1 = a_1, \dots, X_n = a_n) = \mathbb{P}(X_1 = a_1) \dots \mathbb{P}(X_n = a_n), \ \forall a_1, \dots, a_n$$

Indépendance : quelques conséquences

- Si X et Y sont indépendantes, alors pour n'importe quels ensembles A et B, les événements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants (pas seulement les cas où A et B ne contiennent qu'une valeur chacun).

- Si X et Y sont indépendantes, et si f et g sont des fonctions quelconques, alors $f(X)$ et $g(Y)$ sont indépendantes.

- Généralisation : si $X_1, \ldots, X_n$ sont des variables indépendantes, et si $Y = f(X_1, \ldots, X_k)$, et $Z = g(X_{k+1}, \ldots, X_n)$, Y et Z sont indépendantes.

Espérance et variance

- **Espérance** : décrit l'ordre de grandeur des valeurs (c'est une **moyenne**)
- **Variance** : décrit la *"dispersion"* par rapport à la moyenne

Espérance d'une variable aléatoire

- **Définition :** X variable aléatoire à valeurs réelles, V son ensemble de valeurs :

$$\mathbf{E}(X) = \sum_{v \in V} v \mathbb{P}(X = v)$$

Propriétés de l'espérance

- Si X est une variable aléatoire *positive* (*i.e.* $\mathbb{P}(X \geq 0) = 1$), alors $\mathbf{E}(X) \geq 0$
- **Linéarité :** $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$
- **Linéarité (bis) :** $\mathbf{E}(aX) = a\mathbf{E}(X)$ (pour a constante)
- **Si** X et Y sont des variables aléatoires indépendantes, alors

$$\mathbf{E}(X \cdot Y) = \mathbf{E}(X) \cdot \mathbf{E}(Y).$$

Rappel loi "usuelles" (revoir dernier cours)

- **Bernoulli :** variable aléatoire X valant 0 ou 1 ; $p = \mathbb{P}(X = 1)$.
- **Géométrique :** nombre d'essais jusqu'au premier succès dans une séquence d'essais indépendants et de même probabilité de succès ; $\mathbb{P}(G = k) = (1 - p)^{k-1}p$ pour $k > 0$.
- **Binomiale :** nombre d'essais réussis parmi n essais indépendants et de même probabilité de succès ; $\mathbb{P}(B = k) = \binom{n}{k}p^k(1 - p)^{n-k}$ pour $0 \leq k \leq n$.
- **Poisson :** $\mathbb{P}(P = k) = e^{-x}x^k/k!$

Espérances des lois usuelles

Loi	param.	conditions	espérance
Bernoulli	p	$0 < p < 1$	p
Binomiale	(n, p)	$n \in \mathbb{N}^*$, $0 < p < 1$	$n \cdot p$
Géométrique	p	$0 < p < 1$	$1/p$
Poisson	x	$x > 0$	x

- calculs en TD

Espérance de $f(X)$

- Soit X une v. a. dont on connaît la loi ; et soit $f : \mathbb{R} \to \mathbb{R}$ une fonction.
- On veut calculer $\mathbf{E}(f(X))$.
- Cas particulier : si f est de la forme $f(x) = ax + b$ (a et b constantes), $\mathbf{E}(aX + b) = a\mathbf{E}(X) + b$ (par linéarité)
- Sinon, il faut calculer :

$$\mathbf{E}(f(X)) = \sum_{v \in V} f(v)\mathbb{P}(X = v)$$

Exemple calcul espérance

- X : tirage d'un dé ; valeurs $\{1, 2, 3, 4, 5, 6\}$, loi uniforme
- $Z = (X - 3)^2$: valeurs $\{0, 1, 4, 9\}$
- On peut calculer la loi de Z :
  - $\mathbb{P}(Z = 0) = \mathbb{P}(Z = 9) = 1/6$
  - $\mathbb{P}(Z = 1) = \mathbb{P}(Z = 4) = 2/6$
- On en déduit directement $\mathbf{E}(Z)$:

$$\mathbf{E}(Z) = \frac{0}{6} + \frac{9}{6} + \frac{1 \times 2}{6} + \frac{4 \times 2}{6} = \frac{19}{6}$$

- On a aussi l'autre formule :

$$\mathbf{E}(Z) = \frac{4 + 1 + 0 + 1 + 4 + 9}{6} = \frac{19}{6}$$

Variance

- La **variance** d'une variable aléatoire est une "mesure" de sa tendance à dévier de son espérance.
- **Définition :** si X est une variable aléatoire, et $m = \mathbf{E}(X)$ est son espérance, alors on définit la **variance** de X, notée $\mathbf{Var}(X)$, comme étant

$$\mathbf{Var}(X) = \mathbf{E}((X-m)^2)$$

- **Remarque :** d'après la définition, on a forcément $\mathbf{Var}(X) \geq 0$.
- **Proposition :** la formule suivante est également valable :

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2.$$

- (C'est souvent cette formule qui est utilisée pour un calcul pratique ; mais par exemple, il n'est pas aisé sur cette formule de voir que la variance ne peut pas être négative)

Propriétés de la variance

- La variance est toujours positive ou nulle ; et elle ne peut être nulle que si la variable aléatoire est une variable aléatoire **constante** ($\mathbb{P}(X = a) = 1$, pour une certaine valeur a).
- On note parfois σ^2 pour la variance ; σ (la **racine carrée** de la variance) est appelé **écart-type** de la variable.
- Si a est une constante, $\mathbf{Var}(aX) = a^2\mathbf{Var}(X)$;
- Si X s'exprime dans une "unité" (mètres, kilogrammes, kilomètres par heure...), sa variance s'exprime dans le **carré** de cette unité ; l'écart-type est de nouveau dans la même unité que X.

Variances classiques

- Variable de Bernoulli : $\mathbf{Var}(B) = p - p^2$.
- Variable de Poisson : $\mathbf{E}(P^2) = x^2 + x$, $\mathbf{E}(P) = x$, donc $\mathbf{Var}(P) = x$.
- Variable géométrique : $\mathbf{E}(G^2) = \frac{2-p}{p^2}$, ce qui entraîne

$$\mathbf{Var}(G) = \frac{1-p}{p^2}$$

Variance d'une somme

- Soient X et Y deux v. a. et soient $m = \mathbf{E}(X)$ et $m' = \mathbf{E}(Y)$ *but :* calculer $\mathbf{Var}(X+Y)$.
- On va calculer $\mathbf{Var}(X+Y) = \mathbf{E}((X+Y)^2) - (\mathbf{E}(X+Y))^2$.
- Partie facile : $\mathbf{E}(X+Y) = m + m'$ donc $(\mathbf{E}(X+Y))^2 = m^2 + 2m.m' + m'^2$.
- Pour $\mathbf{E}((X+Y)^2)$, on développe : $(X+Y)^2 = X^2 + 2XY + Y^2$, donc par linéarité de l'espérance,
$$\mathbf{E}((X+Y)^2) = \mathbf{E}(X^2) + 2\mathbf{E}(XY) + \mathbf{E}(Y^2).$$
- Par soustraction :
$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) + 2(\mathbf{E}(XY) - m.m')$$
- En règle générale, ça s'arrête là (il faut calculer $\mathbf{E}(XY)$)
- **Si X et Y sont indépendantes**, on peut simplifier : $\mathbf{E}(XY) = m.m'$ et donc
$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$$

Généralisation ; variance d'une binomiale

- On généralise la formule de la variance à la somme de n variables indépendantes : **Si** n **variables** $X_1, \ldots, X_n$ **sont indépendantes**, alors

$$\mathbf{Var}(X_1 + \cdots + X_n) = \mathbf{Var}(X_1) + \cdots + \mathbf{Var}(X_n)$$

- En particulier, en tenant compte de la description d'une binomiale comme somme de Bernoulli indépendantes, on obtient sa variance : $n.p.(1-p)$.

Variances des autres lois "classiques"

- Variable de Bernoulli : $\mathbf{Var}(B) = p - p^2$.
- Variable de Poisson : $\mathbf{E}(P^2) = x^2 + x$, $\mathbf{E}(P) = x$, donc $\mathbf{Var}(P) = x$.
- Variable géométrique : $\mathbf{E}(G^2) = \frac{2-p}{p^2}$, ce qui entraîne

$$\mathbf{Var}(G) = \frac{1-p}{p^2}$$

Récapitulatif : espérances et variances

Loi	param.	espérance	variance
Bernoulli	p	p	$p(1-p)$
Binomiale	(n, p)	$n.p$	$n.p.(1-p)$
Géométrique	p	$1/p$	$(1-p)/p^2$
Poisson	x	x	x

Inégalités basées sur espérance et variance

- **Idée générale :** trouver des propriétés de la forme, "la probabilité que telle variable aléatoire prenne une valeur [plus grande,plus petite] que [valeur] est petite"
- On va en voir deux, assez générales : l'**inégalité de Markov**, et l'**inégalité de Tchebycheff**.

Inégalité de Markov

- **Inégalité de Markov :** si X est une variable aléatoire **à valeurs positives**, d'espérance m, alors pour tout nombre a,

$$\mathbb{P}(X \geq a.m) \leq \frac{1}{a}$$

- Autre formulation : $\mathbb{P}(X > b) \leq \frac{m}{b}$
- Attention : il faut que la variable soit positive ; ne marche pas si on n'a pas de telle garantie !
- (une v.a. positive a probabilité au plus $1/10$ d'être 10 fois plus grande que son espérance)
- **Remarque :** l'inégalité ne dit rien d'intéressant pour $a \leq 1$.

Inégalité de Tchebycheff

- **Théorème :** si X est une variable aléatoire d'espérance m et de variance $v = \sigma^2$, alors on a, pour tout réel $a > 0$,

$$\mathbb{P}(|X - m| \geq a\sigma) \leq \frac{1}{a^2}.$$

- **Remarque :** la condition "$|X - m| \geq a.\sigma$" correspond à "$X \geq m + a.\sigma$ ou $X \leq m - a.\sigma$" : on regarde la probabilité d'être "loin de l'espérance"
- En particulier ($a = 3$), $\mathbb{P}(|X - m| > 3\sigma) \leq 1/9$, donc

$$\mathbb{P}(X \in [m - 3\sigma, m + 3\sigma]) \geq 8/9$$

- **Remarque :** pas de condition du type "X positive", mais en revance il faut connaître la variance.
- Cette inégalité va nous permettre de montrer la "loi des grands nombres"