Table des matières

I Variance, covariance

I.1 Rappels sur la variance

I.1.1 Définition

Soit X une variable aléatoire discrète à valeurs dans $\{x_n|\ n \in \mathbb{N}\}$. On dit que X est d'espérance finie (ou admet un moment d'ordre 1) ssi $\sum_{n\in\mathbb{N}} x_n\mathbb{P}(X = x_n)$ converge **absolument**.

Dans ce cas, on appelle **espérance de** X et on note $E(X)$ le réel $\sum_{n=0}^{+\infty} x_n\mathbb{P}(X = x_n)$.

I.1.2 Définition

Soit X une variable aléatoire discrète. On dit que X est de variance finie (ou que X admet un moment d'ordre 2) ssi X^2 est d'espérance finie et alors la variance de X est $V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$.

Dans ce cas, l'écart type de X est $\sigma(X) = \sqrt{V(X)}$.

I.1.3 Proposition

Soit X une variable aléatoire discrète de variance finie. $V(X) = 0$ ssi X est presque sûrement constante, c'est à dire qu'elle prend une seule valeur avec une probabilité 1 (et les éventuelles autres avec un probabilité 0).

Preuve.
Si X est constante presque sûrement, alors $E(X) - X = 0$ avec une probabilité 1 et donc $V(X) = 0$.

Réciproquement, supposons $V(X) = 0$. Alors $E((X - E(X))^2) = 0$ ie $\sum_{n=0}^{+\infty}(x_n - E(X))^2\mathbb{P}(X = x_n) = 0$. Mais cette série est à terme positifs donc tous ses termes sont nuls ie on a soit $x_n = E(X)$ soit $\mathbb{P}(X = x_n) = 0$. ∎

I.1.4 Proposition

Soit X une variable aléatoire discrète possédant une variance finie. Alors pour $a, b \in \mathbb{R}$, $aX + b$ est de variance finie et $V(aX + b) = a^2V(X)$.

I.2 Covariance

I.2.1 Définition-Proposition

Soient X, Y deux variables aléatoires discrètes. Si X et Y admettent un moment d'ordre 2 alors la variable $(X - E(X))(Y - E(Y))$ est d'espérance finie.

Dans ce cas on appelle covariance de X et Y le réel

$$\mathrm{cov}(X, Y) = E((X - E(X))(Y - E(Y)))$$

.

I.2.2 Remarque

On a $\mathrm{cov}(X, X) = V(X)$.

I.2.3 Proposition

Dans les conditions de la définition précédente :

1. $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$.
2. **Si** X et Y sont indépendantes **alors** $\mathrm{cov}(X, Y) = 0$.
3. la covariance est bilinéaire et symétrique.

Preuve.
Simple utilisation de la linéarité de l'espérance, en plus de la propriété $E(a) = a$ quand a est une constante.

Le deuxième point est une conséquence directe d'un théorème du chapitre sur les probabilités.

La symétrie est évidente, la bilinéarité conséquence simple de la linéarité de l'espérance. ∎

I.2.4 Proposition
Soit X, Y deux variables aléatoires admettant une variance finie.

$$V(X+Y) = V(X) + V(Y) + 2\operatorname{cov}(X,Y)$$

Preuve.
En effet, $V(X+Y) = E((X+Y)^2) - (E(X+Y))^2 = E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y^2)$. ∎

I.2.5 Exemple
Ainsi pour des variables indépendantes, $V(X+Y) = V(X) + V(Y)$ et plus généralement la variance d'une somme de variables mutuellement indépendantes est la somme des variances.

Rappelons une application importante, posons $(X_i)_{i\in[\![1,n]\!]}$ des variables aléatoires mutuellement indépendante de même loi de Bernoulli de paramètre $p \in]0,1[$.

Alors $S = \sum\limits_{i=1}^{n} X_i$ suit une loi binomiale de paramètres n et p.

Or $V(X_0) = E(X_0^2) - E(X_0)^2 = p - p^2 = p(1-p)$ et donc $V(S) = np(1-p)$.

Finissons le rappel par $E(S) = nE(X_0) = np$ par linéarité de l'espérance.

I.2.6 Théorème (Cauchy-Schwartz)
On a $|\operatorname{cov}(X,Y)| \leqslant \sqrt{V(X)V(Y)}$

Preuve.
Soit $t \in \mathbb{R}$. Alors $V(X+tY) = \cdots = t^2V(Y) + 2t\operatorname{cov}(X,Y) + V(X)$ qui est de degré 2 (si $V(Y) \neq 0$) et positif. Le discriminant est donc négatif. ∎

I.2.7 Définition
Soient X, Y deux variables aléatoires de variance finie et non nulle. Le coefficient de corrélation de X et Y est

$$\operatorname{cor}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)} \in [-1,1]$$

I.2.8 Interprétation

L'étude du cas d'égalité dans l'inégalité de Cauchy-Schwartz précédente, permet de montrer que $\operatorname{cor}(X,Y) = \pm 1$ ssi $Y = aX + b$ pour a, b des réels. De plus, si X et Y sont indépendantes, $\operatorname{cor}(X,Y) = 0$. On peut "donc" interpréter ce coefficient comme une mesure du lien (autrement appelé corrélation) qui existe entre X et Y.

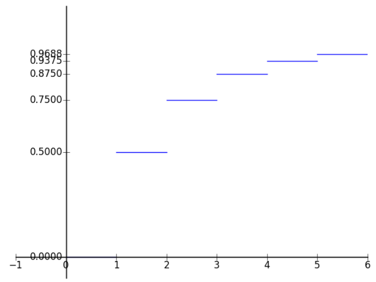
II Fonctions et probabilités

II.1 Fonction de répartition

II.1.1 Définition
Soit X une variable aléatoire discrète. On appelle fonction de répartition de X et on note F_X la fonction

$$F_X : \begin{cases} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \mathbb{P}(X \leqslant x) \end{cases}$$

II.1.2 Exemple
Tracer un partie de la courbe représentative de F_X où X suit une loi géométrique de paramètre $\frac{1}{2}$.



II.1.3 Remarque

Imaginons que l'on connaisse la fonction de répartition F_X d'une VA X mais pas sa loi. Notons $X(\Omega) = \{x_n | \ n \in \mathbb{N}\}$ l'ensemble des valeurs de X où on a ordonnée les x_n, ie la suite (x_n) est croissante.

Alors $\mathbb{P}(X = x_0) = F_X(x_0)$ et pour tout $n \neq 0$, $\mathbb{P}(X = x_n) = F_X(x_n) - F_X(X_{n-1})$

II.1.4 Proposition
Avec les notations de la définition, on a :

1. F_X est croissante sur $\mathbb{R}$.
2. $\lim\limits_{x\to-\infty} F_X(x) = 0$.
3. $\lim\limits_{x\to+\infty} F_X(x) = 1$.

Preuve.
Il s'agit d'utiliser les propriétés de $\mathbb{P}$ suivantes : croissance, limite de la probabilité d'une suite décroissante d'événements et limite de la probabilité d'une suite croissante d'événements. ■

II.1.5 Exemple
En pratique, il est parfois plus pratique de calculer des probabilités de la forme $\mathbb{P}(X \leqslant n)$, ce qui revient à calculer la fonction de répartition sans le dire.

Soient par exemple X, Y deux VA indépendantes de loi géométrique de paramètre $p \in]0,1[$ et $Z = \min(X,Y)$. Calculer $1 - F_Z$

Pour $n \in \mathbb{N}\backslash\{0\}$, $\mathbb{P}(Z > n) = \mathbb{P}((X > n) \cap (Y > n)) = \mathbb{P}(X > n)\mathbb{P}(Y > n) = ((1-p)^2)^n$.

On en déduit facilement que $Z \hookrightarrow \mathcal{G}(2p - p^2)$.

II.2 Fonction génératrice

II.2.1 Une série entière

Soit X une VA à valeurs dans $\mathbb{N}$ (son ensemble de valeurs est un sous-ensemble de $\mathbb{N}$).

Considérons la série entière $\sum\limits_{n\geqslant 0} \mathbb{P}(X = n)t^n$. Comme cette série converge absolument pour $t = \pm 1$, son rayon de convergence vaut au moins 1.

II.2.2 Définition
Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La fonction génératrice (ou série génératrice) de X est la fonction

$$G_X : t \mapsto E(t^X) = \sum_{n=0}^{+\infty} \mathbb{P}(X = n)t^n$$

G_X est définie au moins sur le segment $[-1,1]$, $\mathcal{C}^\infty$ sur $]-1,1[$ et $G_X(1) = 1$.

II.2.3 Remarque

Par unicité des coefficients d'une série entière de rayon de convergence non nul, la loi de X est entièrement déterminée par la fonction G_X.

II.2.4 Valeurs manquantes

On convient de poser $\mathbb{P}(X = n) = 0$ pour tous les n qui ne sont pas des valeurs de X. En particulier, pour une variable aléatoire sur un univers fini G_X est polynomiale !

II.2.5 Exemple
Calculons les fonctions génératrices pour les lois usuelles.

1. Soit $X \hookrightarrow \mathcal{B}(p)$ (Bernoulli, $p \in]0,1[$).
   Alors $G_X : t \mapsto (1-p)t^0 + pt^1 = 1 - p + pt$.
2. Soit $X \hookrightarrow \mathcal{B}(n,p)$ (binomiale, $n \in \mathbb{N}\backslash\{0\}$, $p \in]0,1[$).
   Alors $G_X : t \mapsto \sum\limits_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}t^k = (1-p+pt)^n$.
3. Soit $X \hookrightarrow \mathcal{G}(p)$ (où $p \in]0,1[$).
   Pour $t \in]-1,1[$ (au moins), $G_X(t) = \sum\limits_{n=1}^{+\infty} p(1-p)^{n-1}t^n = pt\sum\limits_{n=0}^{+\infty}((1-p)t)^n$.
   Comme $|1-p| < 1$, on a bien $|(1-p)t| < 1$ et donc $G_X(t) = \frac{pt}{1-(1-p)t}$.
   Remarquons que le rayon de convergence de la série est $\frac{1}{1-p} > 1$ et l'égalité précédente reste vraie.
4. Soit $X \hookrightarrow \mathcal{P}(\lambda)$ ($\lambda > 0$).
   Pour $t \in]-1,1[$, $G_X(t) = \sum\limits_{n=0}^{+\infty} e^{-\lambda}\frac{\lambda^n}{n!}t^n = e^{-\lambda}e^{\lambda t} = e^{\lambda(t-1)}$.

II.2.6 Exercice
Déterminer la fonction génératrice d'une loi uniforme sur $[\![1,n]\!]$.

II.2.7 Théorème
Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ et G_X sa fonction génératrice.

1. X possède une espérance finie ssi G_X est dérivable en 1 et alors $E(X) = G'_X(1)$.
2. X possède une variance finie ssi G_X est deux fois dérivable en 1 et alors

$$V(X) = G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2.$$

II.2.8 Retrouver les formules

Tout d'abord, on a $E(X)=\sum\limits_{n=1}^{+\infty} n\mathbb{P}(X=n)$ et $E(X^2)=\sum\limits_{n=1}^{+\infty} n^2\mathbb{P}(X=n)$ (théorème de transfert).

De plus, on supposant la dérivabilité terme à terme, $G'_X(t)=\sum\limits_{n=1}^{+\infty} n\mathbb{P}(X=n)t^{n-1}$ donc on a bien $G'_X(1)=E(X)$.

De plus, $G''_X(t)=\sum\limits_{n=2}^{+\infty} n(n-1)\mathbb{P}(X=n)t^{n-1}$ donc $G''_X(1)=E(X^2)-E(X)$.

II.2.9 Exemple
Retrouvons l'espérance et la variance des lois géométriques et de Poisson.

1. Soit $X\hookrightarrow\mathcal{G}(p)$ (où $p\in]0,1[$).
   G_X des DSE sur $]-\frac{1}{1-p},\frac{1}{1-p}[$ donc est deux fois dérivable en 1.
   De plus, $G_X(t)=\frac{pt}{1-(1-p)t}$ donc $G'_X(t)=\frac{p(1-(1-p)t)+pt\times(1-p)}{(1-(1-p)t)^2}=\frac{p}{(1-(1-p)t)^2}$ donc $E(X)=\frac{p}{p^2}=\frac{1}{p}$.
   De même $G''(t)=p\times(-2)\times(-(1-p))\times(1-(1-p)t)^{-3}$ donc $G''(1)=\frac{2p(1-p)}{p^3}=\frac{2(1-p)}{p^2}=E(X^2)-E(X)$.
   Ainsi $V(X)=E(X^2)-(E(X))^2=\frac{2(1-p)}{p^2}+\frac{1}{p}-\frac{1}{p^2}=\frac{1-p}{p^2}$.
2. Soit $X\hookrightarrow\mathcal{P}(\lambda)$ ($\lambda>0$).
   Cette fois G_X est DSE sur $\mathbb{R}$ donc dérivable deux fois en 1.
   De plus $G'_X(t)=\lambda e^{\lambda(t-1)}$ et $G''_X(1)=\lambda^2 e^{\lambda(t-1)}$.
   Ainsi $E(X)=G'_X(1)=\lambda$ et $V(X)=\lambda^2+\lambda-\lambda^2=\lambda$.

II.2.10 Théorème
Si X et Y sont deux variables aléatoires à valeurs dans $\mathbb{N}$ et **indépendantes**, notons R_X et R_Y les rayons de convergence de G_X et G_Y respectivement. Posons également $r=\min(R_X,R_Y)$

Alors G_{X+Y} est de rayon $R\geqslant r$ et

$$\forall t\in]-r,r[\ G_{X+Y}(t)=G_X(t)G_Y(t)$$

Preuve.
Pour $t\in]0,r[$, on pose $f_t:x\mapsto t^x$. Alors $f(X)$ et $f(Y)$ sont indépendantes (cf chapitre de probabilités) et donc $E(t^X)E(t^Y)=E(t^Xt^Y)=E(t^{X+Y})$. Ainsi $t\leqslant R$ et on a bien $G_{X+Y}(t)=G_X(t)G_Y(t)$.

Deuxième méthode. La série produit (de Cauchy) G_XG_Y est de rayon $R\geqslant\min(R_X,R_Y)$ et pour $t\in]-r,r[$

$$G_X(t)G_Y(t)=\sum_{n=0}^{+\infty}c_nt^n$$

où $c_n=\sum\limits_{k=0}^{n}\mathbb{P}(X=k)\mathbb{P}(Y=n-k)=\sum\limits_{k=0}^{n}\mathbb{P}((X=k)\cap(Y=n-k))=\mathbb{P}(X+Y=n)$. ■

II.2.11 Exemple
On peut utiliser ce théorème pour calculer la loi d'une somme de variables indépendantes.

1. Soient $\lambda,\mu>0$ et $X\hookrightarrow\mathcal{P}(\lambda),YX\hookrightarrow\mathcal{P}(\mu)$.
   Alors, pour tout $t\in\mathbb{R}$, $G_X(t)G_X(Y)=e^{\lambda(t-1)}e^{\mu(t-1)}=e^{(\lambda+\mu)(t-1)}=G_{X+Y}(t)$.
   Ainsi $X+Y\hookrightarrow\mathcal{P}(\lambda+\mu)$ (car la fonction génératrice détermine la loi).
2. Lançons deux dés équilibrés à 6 faces et notons X,Y les résultats obtenus pour le premier et le second dé respectivement.
   Donner la loi de $X+Y$ (la somme des deux dés).
   Ici les lois prennent un nombre fini de valeurs et donc les fonctions génératrices sont polynomiales.
   Pour $t\in\mathbb{R}$, $G_X(t)=G_Y(t)=\frac{1}{6}t\sum\limits_{k=0}^{5}t^k$. De plus X et Y sont indépendantes.
   Ainsi $G_{X+Y}(t)=\frac{t^2}{36}\left(\sum\limits_{k=0}^{5}t^k\right)^2=\frac{t^2}{36}(1+2t+3t^2+4t^3+5t^4+6t^5+5t^6+4t^7+3t^8+2t^9+t^{10})$. On obtient

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(X=k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

III Etude asymptotique

III.1 Interprétation de la loi de Poisson

III.1.1 Proposition
Soit $\lambda>0$.

On considère une suite $(X_n)_{n\geqslant1}$ de variables aléatoires telles que $X_n\hookrightarrow\mathcal{B}(n,p_n)$ où $p_n\underset{+\infty}{\sim}\frac{\lambda}{n}$.

Pour $k\in\mathbb{N}$ fixé, on a $\lim\limits_{n\to+\infty}\mathbb{P}(X_n=k)=e^{-\lambda}\frac{\lambda^k}{k!}$.

Preuve.
On cherche à estimer la limite de $\binom{n}{k} p_n^k (1-p_n)^{n-k}$.

Or $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{1}{k!} \prod_{i=0}^{k-1} (n-i) \underset{n\to+\infty}{\sim} \frac{1}{k!} n^k$. (par produit d'un nombre fixé d'équivalents)

De plus, $p_n^k \underset{+\infty}{\sim} \frac{\lambda^k}{n^k}$ (encore une fois, k est fixé).

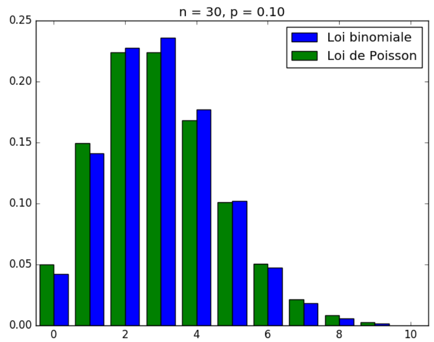
De plus, $(1-p_n)^{n-k} \underset{+\infty}{\sim} (1-p_n)^n$ car $(1-p_n)^{-k} \underset{n\to+\infty}{\to} 1$.

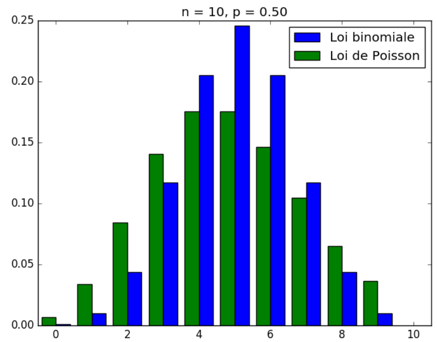
Comme $(1-p_n)^n = e^{n\ln(1-p_n)} = e^{n(-p_n + o_{+\infty}(p_n))} = e^{-\lambda + o_{+\infty}(1)}$ (avec 2 $o(1)$ obtenus en remplaçant p_n par son équivalent dans le o). Ainsi $(1-p_n)^n \underset{+\infty}{\to} e^{-\lambda} \neq 0$ (et donc on peut transformer cette limite en équivalent).

Il n'y a plus qu'à effectuer le produit de nos équivalents.. ∎

III.1.2 En pratique

On peut utiliser une loi de Poisson pour approximer une loi binomiale de paramètre (n,p) dans le cas où $\lambda = np$ n'est "pas trop grand".





III.2 Loi des grands nombres

III.2.1 Théorème (Inégalité de Markov)
Soit X une variable aléatoire réelle à valeurs positives, d'espérance finie.

$$\forall a > 0 \ \mathbb{P}(X \geqslant a) \leqslant \frac{E(X)}{a}.$$

Explication L'idée "grossière" derrière ce théorème est que si l'espérance (la valeur moyenne) de X vaut m, alors X ne prend pas des valeurs trop grande par rapport à m, ou alors avec une probabilité très faible.

III.2.2 Théorème (Inégalité de Bienaymé-Tchebychev)
Soit X une variable aléatoire de variance finie.

$$\forall \varepsilon > 0 \ \mathbb{P}(|X - E(X)| > \varepsilon) \leqslant \frac{V(X)}{\varepsilon^2}.$$

Explication On quantifie cette fois l'écart entre X et sa "moyenne". La variance apparaît naturellement.

III.2.3 Exemple
On pose S_n la moyenne arithmétique de n variables de loi de Bernoulli indépendantes de paramètre p. $S = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Exemple pratique : on dépouille une urne contenant n bulletins dans une élection à deux candidats. $(X_i = 1)$ est l'événement : le i-ème bulletin est pour le candidat A. Ici l'indépendance des variables n'est sûrement pas respecté dans la pratique. Tant pis, poursuivons.

Le but est d'estimer p, la proportion de votant ayant choisi le candidat A. Cette probabilité (théorique) est inconnue au moment de l'expérience.

Alors $E(S) = p$ et $V(S) = \frac{p(1-p)}{n^2}$.

S représente la proportion votes après n dépouillements indépendants. Alors $\mathbb{P}(|S-p| > \varepsilon) \leqslant \frac{p(p-1)}{n\varepsilon^2}$.

On veut $\mathbb{P}(|S-p| > \varepsilon) \leqslant 5\%$. Comment choisir ε ? Il faut $\frac{p(p-1)}{n\varepsilon^2} \leqslant \frac{5}{100}$ soit encore $\varepsilon^2 \leqslant 20\frac{p(1-p)}{n^2}$.

Or $p(1-p) \leqslant \frac{1}{4}$ (étape obligatoire, on ne connaît pas encore p). On a donc $\varepsilon^2 \leqslant \frac{5}{n^2}$.

Ainsi, si on veut une approximation de p à 1% près, on prend $\frac{1}{100} \leqslant \sqrt{5}\frac{1}{n}$ soit encore $n \geqslant 100\sqrt{5}$.

Attention, on à juste le résultat : la probabilité pour que la fréquence théorique s'écarte de plus de 1% de la fréquence observée est $\leqslant \frac{5}{100}$

III.2.4 Théorème (Loi faible des grands nombres)
Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires deux à deux indépendantes et de même loi, admettant un moment d'ordre 2.

On pose, pour $n \geqslant 1$, $S_n = \sum\limits_{k=1}^{n} X_k$ et on note $m = E(X_1)$ l'espérance commune aux X_k.

$$\forall\varepsilon > 0 \ \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right) \underset{n\to+\infty}{\to} 0$$

. Pour un $\varepsilon > 0$ fixé, la limite est nulle.

Explication Ce théorème est la formalisation mathématique d'une idée naturelle.

Je répète n fois la même expérience aléatoire de Bernoulli (paramètre p) sans connaître a priori le paramètre p (on cherche à estimer une fréquence de manière empirique, par exemple pour réaliser un sondage...)

Alors la fréquence moyenne de succès converge vers le paramètre théorique p.