MAPSI — cours 2 : Rappels de probabilités et statistiques

Pierre-Henri Wuillemin (& Christophe Gonzales)

LIP6 / ISIR – Sorbonne Université, France

Plan du cours n°2

1. Indépendance mutuelle
2. Indépendance conditionnelle
3. Loi de Bernoulli / binomiale
4. Loi normale
5. Théorème central-limite

Rappel : Indépendance de deux variables discrètes

X et Y sont *indépendantes* si $\forall x \in X, \forall y \in Y$:

les événements $X = x$ et $Y = y$ sont indépendants

1. $\forall x, \forall y,\ P(X = x \cap Y = y) = P(X = x) \times P(Y = y)$
   $$P(X, Y) = P(X) \times P(Y)$$
2. $\forall x, \forall y$ $t.q.$ $P(Y = y) > 0,\ P(X = x | Y = y) = P(X = x)$
3. $\forall y, \forall x$ $t.q.$ $P(X = x) > 0,\ P(Y = y | X = x) = P(Y = y)$

⚠ 2 et 3 : conditionnement = apport d'information

Rappel : Indépendance de deux variables continues

X et Y sont *indépendantes* si $\forall I, \forall J$, intervalles,

les événements $X \in I$ et $Y \in J$ sont indépendants

Il suffit que les fonctions de répartition, F_X, F_Y de X et Y et F_{XY} du couple vérifient :

$$\forall x, y,\ F_{XY}(x, y) = F_X(x) \times F_Y(y)$$

ou encore que les densités de probabilité p_X, p_Y de X et Y et p_{XY} du couple vérifient :

$$\forall x, y,\ p_{XY}(x, y) = p_X(x) \times p_Y(y)$$

Généralisation : indépendance mutuelle de n variables

Définition

Soient n variables aléatoires $(X_1, X_2, \ldots, X_k, \ldots, X_n)$
Elles sont *mutuellement indépendantes* si tout événement lié à une partie d'entre elles est indépendant de tout événement lié à toute autre partie disjointe de la précédente

$\Longrightarrow$ c'est la généralisation naturelle de l'indépendance de deux variables :

Les variables discrètes $(X_1, \ldots, X_k, \ldots, X_n)$ sont mutuellement indépendantes lorsque :

$$\forall x_k, P\left(x_1, \cdots, x_n\right) = \prod_{k=1}^{n} P(X_k = x_k) = \prod_{k=1}^{n} P(x_k)$$

Indépendance mutuelle de n variables

Définition

Soient n variables aléatoires $(X_1, X_2, \ldots, X_k, \ldots, X_n)$
Elles sont *mutuellement indépendantes* si tout événement lié à une partie d'entre elles est indépendant de tout événement lié à toute autre partie disjointe de la précédente

Pour des variables continues $(X_1, \ldots, X_k, \ldots, X_n)$ sont mutuellement indépendantes lorsque :

$$p_{X_1 \ldots X_k \ldots X_n}(x_1, \ldots, x_k, \ldots, x_n) = \prod_{k=1}^{n} p_{X_k}(x_k)$$

L'indépendance mutuelle de n variables entraîne leur indépendance deux à deux.
la réciproque n'est pas vraie

- Soit n dés à 6 faces
- X_k : variable aléatoire indiquant sur quelle face tombe le $k$ème dé
- tous les dés sont différents $\Longrightarrow X_1, \ldots, X_n$ mutuellement indépendantes

$$P(X_1, \ldots, X_k, \ldots, X_n) = \prod_{k=1}^{n} P(X_k)$$

$\Longrightarrow$ stockage mémoire = $6n$ au lieu de 6^n

$n = 10$	60	60 millions
$n = 20$	120	3,6 millions de milliards

Indépendance conditionnelle (1/2)

Indépendance conditionnelle de deux variables discrètes

X et Y sont *indépendantes conditionnellement à* Z si $\forall x, \forall y, \forall z$, les événements $X = x$ et $Y = y$ sont indépendants conditionnellement à $Z = z$

- $P(X{=}x \cap Y{=}y|Z{=}z) = P(X{=}x|Z{=}z) \times P(Y{=}y|Z{=}z)$
- si $P(Y{=}y|Z{=}z) > 0$ alors :

  $P(X{=}x|Y{=}y, Z{=}z) = P(X{=}x|Z{=}z)$
- si $P(X{=}x|Z{=}z) > 0$ alors :

  $P(Y = y|X{=}x, Z{=}z) = P(Y{=}y|Z{=}z)$

Indépendance conditionnelle (2/2)

Indépendance conditionnelle de deux variables discrètes

X et Y sont *indépendantes* conditionnellement à Z si :

- $P(X \cap Y|Z) = P(X|Z) \times P(Y|Z)$
- si $P(Y|Z) > 0$ alors $P(X|Y,Z) = P(X|Z)$
- si $P(X|Z) > 0$ alors $P(Y|X,Z) = P(Y|Z)$

Interprétation

- Conditionnement = apport de connaissances
- Si l'on connaît la valeur de la variable Z, alors connaître celle de Y n'apporte rien sur la connaissance de X

Ces formules s'étendent si X, Y et/ou Z sont remplacés par des ensembles de variables aléatoires disjoints 2 à 2

$$P(A,B|C) = \begin{array}{cccc|l} \multicolumn{2}{c}{a_1} & \multicolumn{2}{c}{a_2} & \\ c_1 & c_2 & c_1 & c_2 & \\ \end{array}$$

$$P(A,B|C) = \begin{pmatrix} 0{,}15 & 0{,}18 & 0{,}07 & 0{,}56 \\ 0{,}15 & 0{,}12 & 0{,}63 & 0{,}14 \end{pmatrix} \begin{array}{l} b_1 \\ b_2 \end{array} = \underbrace{\begin{pmatrix} 0{,}5 & 0{,}6 & 0{,}1 & 0{,}8 \\ 0{,}5 & 0{,}4 & 0{,}9 & 0{,}2 \end{pmatrix}}_{P(B|A,C)} \begin{array}{l} b_1 \\ b_2 \end{array} \times \underbrace{\begin{pmatrix} 0{,}3 & 0{,}7 \end{pmatrix}}_{P(A)}$$

(columns of the matrices: a_1: c_1, c_2; a_2: c_1, c_2; columns of $P(A)$: a_1, a_2)

$$P(I,C|B) = \begin{pmatrix} 0{,}48 & 0{,}08 & 0{,}48 & 0{,}08 \\ 0{,}12 & 0{,}32 & 0{,}12 & 0{,}32 \end{pmatrix} \begin{array}{l} i_1 \\ i_2 \end{array} = \overbrace{\begin{pmatrix} 0{,}8 & 0{,}2 \\ 0{,}2 & 0{,}8 \end{pmatrix}}^{P(I|C)} \begin{array}{l} i_1 \\ i_2 \end{array} \times \overbrace{\begin{pmatrix} 0{,}6 & 0{,}4 \end{pmatrix}}^{P(C)}$$

(columns of the first matrix: b_1: c_1, c_2; b_2: c_1, c_2; columns of $P(I|C)$ and $P(C)$: c_1, c_2)

⚠ probabilités $\Longrightarrow$ produits terme à terme !

- Classe de 40 étudiants
- assertion : «il y a au moins 2 étudiants qui sont nés le même jour»

Avez-vous intérêt à parier 10 € que cette assertion est vraie ?

- $X_i \in \{1, \ldots, 365\}$ le jour de naissance du *i*ème étudiant
- que vaut P(tous les X_i sont différents) ?

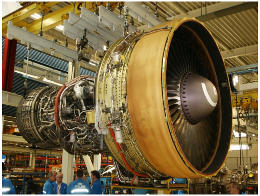
Application des probas conditionnelles (2/2)

$$
\begin{aligned}
\alpha &= P(\text{tous les } X_i \text{ sont différents}) \\
&= P(X_2 \neq X_1) \times P(X_3 \notin \{X_1, X_2\} | X_2 \neq X_1) \times \\
&\quad P(X_4 \notin \{X_1, X_2, X_3\} | X_2 \neq X_1, X_3 \notin \{X_1, X_2\}) \times \ldots \\
&= \prod_{i=2}^{40} P\left(X_i \notin \{X_j : j < i\} \middle| \bigwedge_{j<i} X_j \notin \{X_k : k < j\} \right) \\
&= \prod_{i=2}^{40} \frac{365 - (i-1)}{365} = \prod_{i=1}^{39} \frac{365 - i}{365} \approx 10,87\%
\end{aligned}
$$

$\Longrightarrow$ *en choisissant au hasard une classe de 40 étudiants,*

on a 10,87% de chances que l'assertion soit fausse

Les probas conditionnelles en pratique

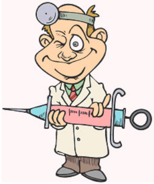


General Electric :

maintenance des moteurs CF6 :

350 variables aléatoires

$> 10^{105}$ événements élémentaires !



Monitoring de patients :

37 variables aléatoires

$> 10^{16}$ événements élémentaires !

Liens entre gènes :

syndrome LQT – marqueur génétique

724 variables aléatoires

$> 10^{277}$ événements élémentaires !

Loi de Bernoulli

Définition

Épreuve de Bernoulli = expérience aléatoire qui ne peut prendre que deux résultats (*succès* et *échec*)

p = proba de succès, et $q = 1 - p$ = proba d'échec.

Loi de Bernoulli

Variable X à support $\mathcal{X} = \{0, 1\}$ telle que :

$P(X = 1) = p$ et $P(X = 0) = 1 - p$

$E(X) = p \quad V(X) = p(1 - p)$

$\Longrightarrow X =$ le nombre de succès de l'épreuve de Bernoulli

Loi binomiale

Définition

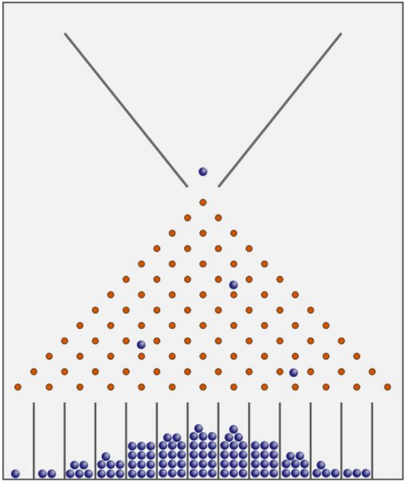
Épreuve binomiale = expérience aléatoire telle que :

1. on répète n fois la même épreuve de Bernoulli,
2. les probas p et q restent inchangées pour chaque épreuve de Bernoulli,
3. les épreuves de Bernoulli sont toutes réalisées indépendamment les unes des autres.

Loi binomiale de paramètres n et p

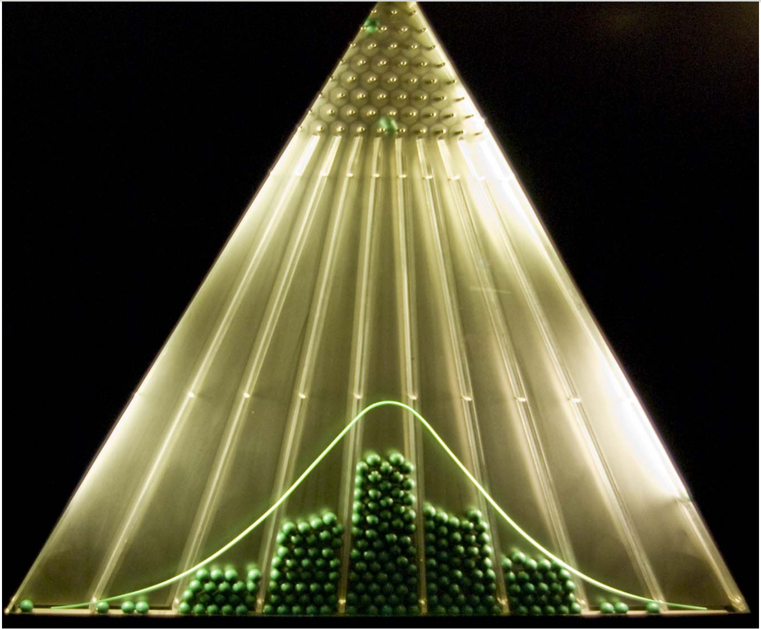
- X = nombre de succès de l'épreuve binomiale
- $X \sim \mathcal{B}(n, p)$
- $P(X = k) = C_n^k \, p^k \, (1 - p)^{n-k}, \forall k = 0, \ldots, n$
- $E(X) = np \quad V(X) = np(1 - p)$

La planche de Galton



- chaque niveau $\Longrightarrow$ expérience de Bernoulli
- $\Longrightarrow X \sim$ loi binomiale

La planche de Galton



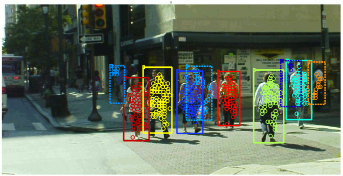
Importance des sommes et moyennes

Master $\Longrightarrow$ moyenne par semestre


La note du Labo Fnac

Le Galaxy S5 possède un écran de 12,9 cm de diagonale et d'une résolution de 1920 x 1080 pixels. La densité de l'affichage est très élevée (432 ppp) et l'écran obtient logiquement la note maximale aux tests de définition. L'angle de vision est large, le S5 obtient donc de bons résultats aux tests de directivité. La fidélité colorimétrique est en aussi hausse par rapport au Galaxy S4 et le contraste est correct, sans plus.

recommandations $\Longrightarrow$ moyenne critères



tracking d'objets par filtre particulaire

Loi normale

⚠ Loi extrêmement importante : souvent une très bonne approximation de la loi réelle

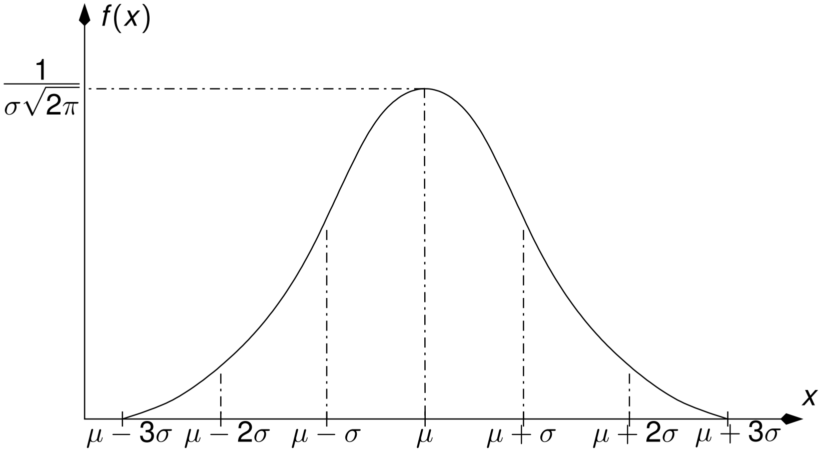
Définition : loi normale de paramètres μ et σ^2

- notée $\mathcal{N}(\mu, \sigma^2)$
- s'applique pour des variables aléatoires continues
- densité positive sur tout $\mathbb{R}$:

$$f(x) = \frac{1}{\sqrt{2\pi}.\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

- $E(X) = \mu \quad V(X) = \sigma^2$

Fonction de densité de la loi normale



Théorème

$X \sim \mathcal{N}(\mu; \sigma^2)$

Alors la variable $Y = aX + b$ obéit à la loi $\mathcal{N}(a\mu + b; a^2\sigma^2)$.

$\Longrightarrow$ toute transformée affine d'une variable aléatoire suivant une loi normale suit aussi une loi normale

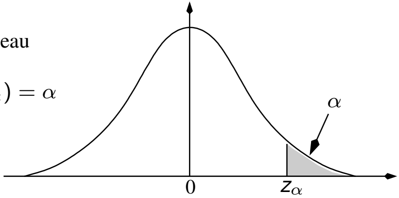
Corollaire

- X une variable aléatoire obéissant à une loi $\mathcal{N}(\mu; \sigma^2)$

  $\Longrightarrow Z = \dfrac{X - \mu}{\sigma}$ suit la loi $\mathcal{N}(0; 1)$

- Z suit une loi normale centrée (à cause de la moyenne en 0) réduite (à cause du σ^2 égal à 1)

valeurs dans le tableau
ci-dessous : les α
tels que $P(Z > z_\alpha) = \alpha$



z_α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0859	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0722	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0466	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

Loi normale bi-dimensionnelle

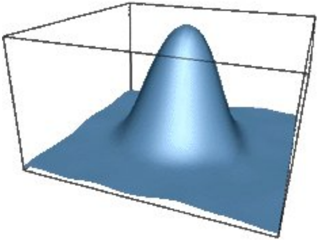
Définition : loi normale bi-dimensionnelle

- couple de variables (X, Y)
- densité dans $\mathbb{R}^2$:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

où $\rho = \frac{cov(X,Y)}{\sigma_x\sigma_y}$ = *coefficient de corrélation linéaire*



Digression : convergenceS des distributions

Convergences pour les distributions

- Convergence en loi : $X_n \stackrel{loi}{\rightarrow} X$
- Convergence en probabilité : $X_n \stackrel{P}{\rightarrow} X$
- Convergence presque sûre : $X_n \stackrel{p.s.}{\rightarrow} X$

Hiérarchie des convergences

$$X_n \stackrel{p.s.}{\rightarrow} X \quad \Rightarrow \quad X_n \stackrel{P}{\rightarrow} X \quad \Rightarrow \quad X_n \stackrel{loi}{\rightarrow} X$$

Convergence en loi

Définition

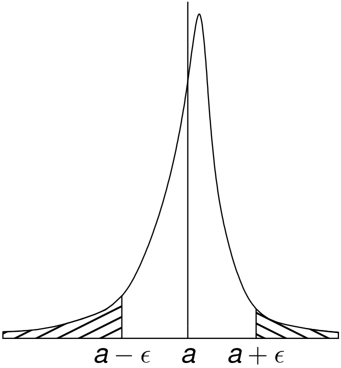
- $(X_n)_{n\in\mathbb{N}}$: suite de variables
- F_n : fonction de répartition de X_n
- X : variable de fonction de répartition F
- La suite X_n *converge en loi* vers X lorsque $F_n(x)$ tend vers $F(x)$ en tout point de continuité x de F

$$\forall x, \lim_{n\to\infty} F_n(x) = F(x)$$

Définition

- $(X_n)_{n\in\mathbb{N}}$: suite de variables
- X : variable aléatoire
- (X_n) *converge en probabilité* vers X si, pour tout $\epsilon > 0$ la probabilité que l'écart absolu entre X_n et X dépasse ϵ tend vers 0 quand $n \to \infty$:

$$\lim_{n\to\infty} P(\mid X_n - X \mid \geq \epsilon) = 0$$



Aire hachurée tend vers 0 quand $n \to \infty$

Convergence presque sûre

Définition

- $(X_n)_{n\in\mathbb{N}}$: suite de variables
- X : variable
- (X_n) *converge presque sûrement* vers X s'il y a une proba 1 que la suite des réalisations des X_n tende vers X :

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$
$$\Longleftrightarrow P\left(\lim_{n\to\infty}\sup_{k\geq n} |X_k - X| \geq \epsilon\right) = 0$$

⚠ Définition la plus exigeante !

Loi faible des grands nombres

Loi faible

- $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires :
  - de même loi
  - d'espérance m
  - possédant une variance σ^2
  - deux à deux indépendantes
- alors la suite des variables $\overline{X}_n = \frac{\sum_{k=1}^n X_k}{n}$ converge en probabilité vers m

$\overline{X}_n$ est appelée *moyenne empirique*

$$E(\overline{X}_n) = m$$

$$V(\overline{X}_n) = \frac{\sigma^2}{n}$$

conséquence : échantillons de grandes tailles $\Longrightarrow$ bonne chance d'estimer m

Loi forte des grands nombres

Loi forte

- $(X_n)_{n \in \mathbb{N}}$: suite de variables aléatoires
  - de même loi
  - d'espérance m
  - possédant une variance σ^2
  - mutuellement indépendantes
- alors la suite des variables $\overline{X}_n = \frac{\sum_{k=1}^{n} X_k}{n}$ converge presque sûrement vers m

Interprétation : échantillon de grande taille $\Longrightarrow$ bonne estimation de m

Théorème central-limite

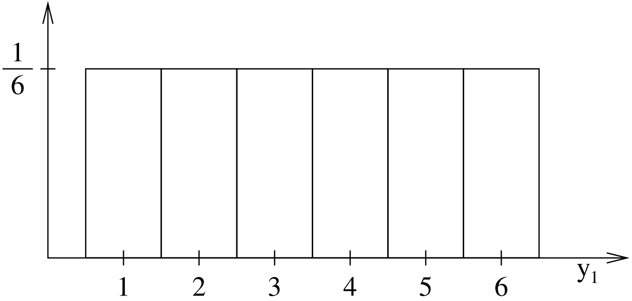
- $(X_n)_{n\in\mathbb{N}}$: suite de variables
  - de même loi
  - d'espérance μ
  - de variance σ^2
  - mutuellement indépendantes
- alors la suite des moyennes empiriques centrées réduites $\frac{\overline{X}_n-\mu}{\sigma\,/\,\sqrt{n}}$ tend en loi vers la loi normale centrée réduite :

$$\frac{\overline{X}_n-\mu}{\sigma\,/\,\sqrt{n}} \xrightarrow{loi} \mathcal{N}(0,1)$$

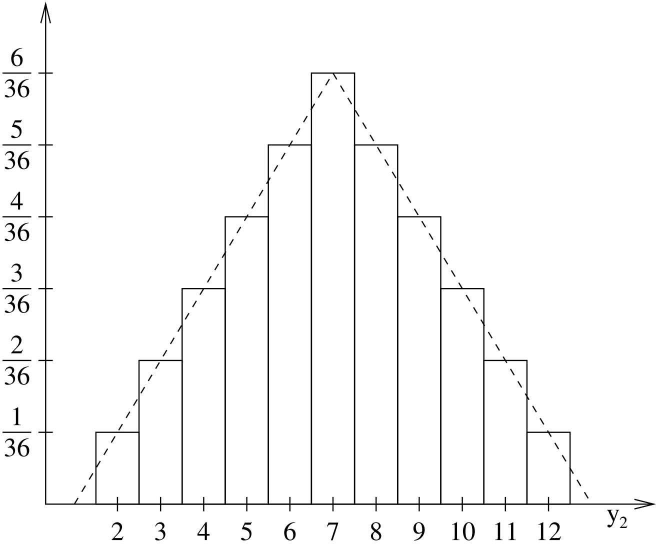
Lancés de dés à 6 faces

$\Longrightarrow$ on compte la somme des résultats des dés

Somme pour 1 jet de dé

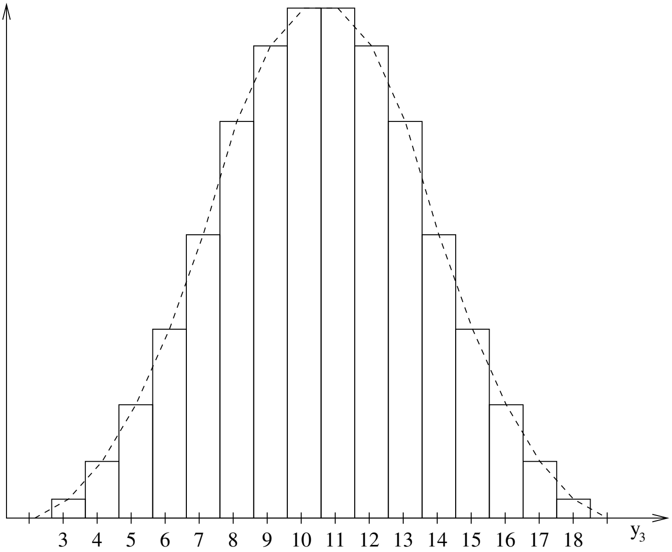


Somme pour 2 jets de dés



Loi normale = limite d'autres lois (3/4)

Somme pour 3 jets de dés



Loi normale = limite d'autres lois (4/4)

Somme pour 4 jets de dés