Chapitre 3

Lois de probabilité

Lois discrètes

2. Loi de Bernoulli (ou loi alternative simple)

On appelle **variable de Bernoulli** ou variable ***indicatrice***, une variable aléatoire X qui peut s'exprimer sous la forme succès $(X=1)$/échec $(X=0)$

Ex: avoir les cheveux longs, présenter un certain génotype, être malade, faire plus de 60 Kg, …

Loi de probabilité: $\boxed{\mathbf{p(X=0)=q \text{ et } p(X=1)=p} \quad \text{avec } p+q=1}$

Elle est appelée **loi de Bernoulli**, notée **B(1,p)**

$$\boxed{E(X) = p} \qquad \boxed{V(X) = pq}$$

X	0	1
$P(X=x_i)$	q	p

Moyenne:

$$E(X) = p(X=0)\times 0 + p(X=1)\times 1 = q\times 0 + p\times 1 = p$$

Variance

$$E(X^2) = p(X=0)\times 0^2 + p(X=1)\times 1^2 = q\times 0 + p\times 1 = p$$

$$\Rightarrow V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p) = pq$$

d'où la loi de Bernoulli où pour $\left|\begin{array}{l} E(X) = p \\ V(X) = pq \end{array}\right.$

Lois discrètes

3. Loi Binomiale

On appelle **variable Binomiale** une variable aléatoire X correspondant au nombre de succès de la répétition **indépendante** de n épreuves de Bernoulli :

Loi de probabilité: $p(X=k)=C_n^k\,p^k\,q^{n-k}$ avec $C_n^k=\dfrac{n!}{k!(n-k)!}$

Elle est appelée **loi Binomiale**, notée **B(n,p)**

$E(X)=np$

$V(X)=npq$

Exemple: Quelle est la probabilité d'avoir 2 filles dans une fratrie de 4 enfants ?

p: probabilité d'avoir une fille à chaque naissance (= 1\2)

X=« nombre de filles »: loi de probabilité = $\mathcal{B}$ (4 ; 1\2)

$$\boxed{P(X=2)=6q^2p^2=0.375}$$

On retrouve ce résultat par le raisonnement suivant:

On peut par ex. avoir la séquence suivante: FFGG

Indépendance => $P(FFGG) = P(F) \times P(F) \times P(G) \times P(G) = p^2q^2$

D'autres séquences marchent aussi: FGFG, FGGF, GFFG, GFGF, GGFF

Ces séquences ont toutes la même probabilité p^2q^2

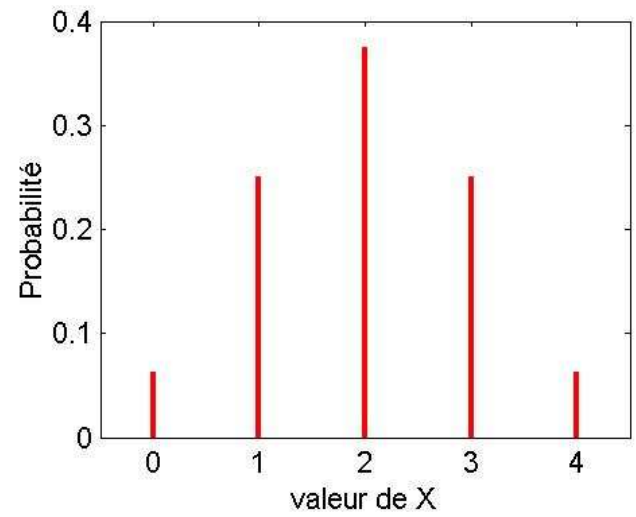
D'où $P(X=2)$ = nombre de séquences $\times\, p^2q^2=6q^2p^2$

Moyenne / variance:

$E(X) = np = 2$ ⇨ 2 filles en moyenne

$V(X) = npq = 2x1\backslash 2 = 1$ ⇨ écart-type = 1

Représentation graphique



Probabilité d'avoir au moins 3 filles

$$P(X >= 3) = p(X=3) + p(X=4)$$
$$= 4p^3q + p^4 = 0.25+0.0625=0.3125$$

Extension:

Si X et Y sont 2 v.a. indépendantes suivant des lois binomiales respectivement de paramètres $X \sim \mathcal{B}(n,p)$ et $Y \sim \mathcal{B}(m,p)$, alors la v.a. $Z = X+Y \sim \mathcal{B}(n+m, p)$

Ex: X =« Nombre de fumeurs hommes dans une entreprise »
Y =« Nombre de fumeurs femmes dans une entreprise (la même) »
Z = X+Y = « Nombre total de fumeurs dans l'entreprise »

En admettant que la proportion p de fumeurs est la même chez les hommes et les femmes, alors

$X \sim B(n,p)$ et $Y \sim B(m,p)$ donc $Z = X+Y \sim B(n+m, p)$

Lois discrètes

4. Loi de Poisson

La loi de **Poisson** est souvent utilisée pour compter l'occurrence d'événements lorsque ceux-ci se produisent les uns à la suite des autres, de façon aléatoire dans le temps.

<u>Exemples</u>: nombre d'accidents de voiture, d'ouragans,
nombre d'individus capturés dans un piège,…

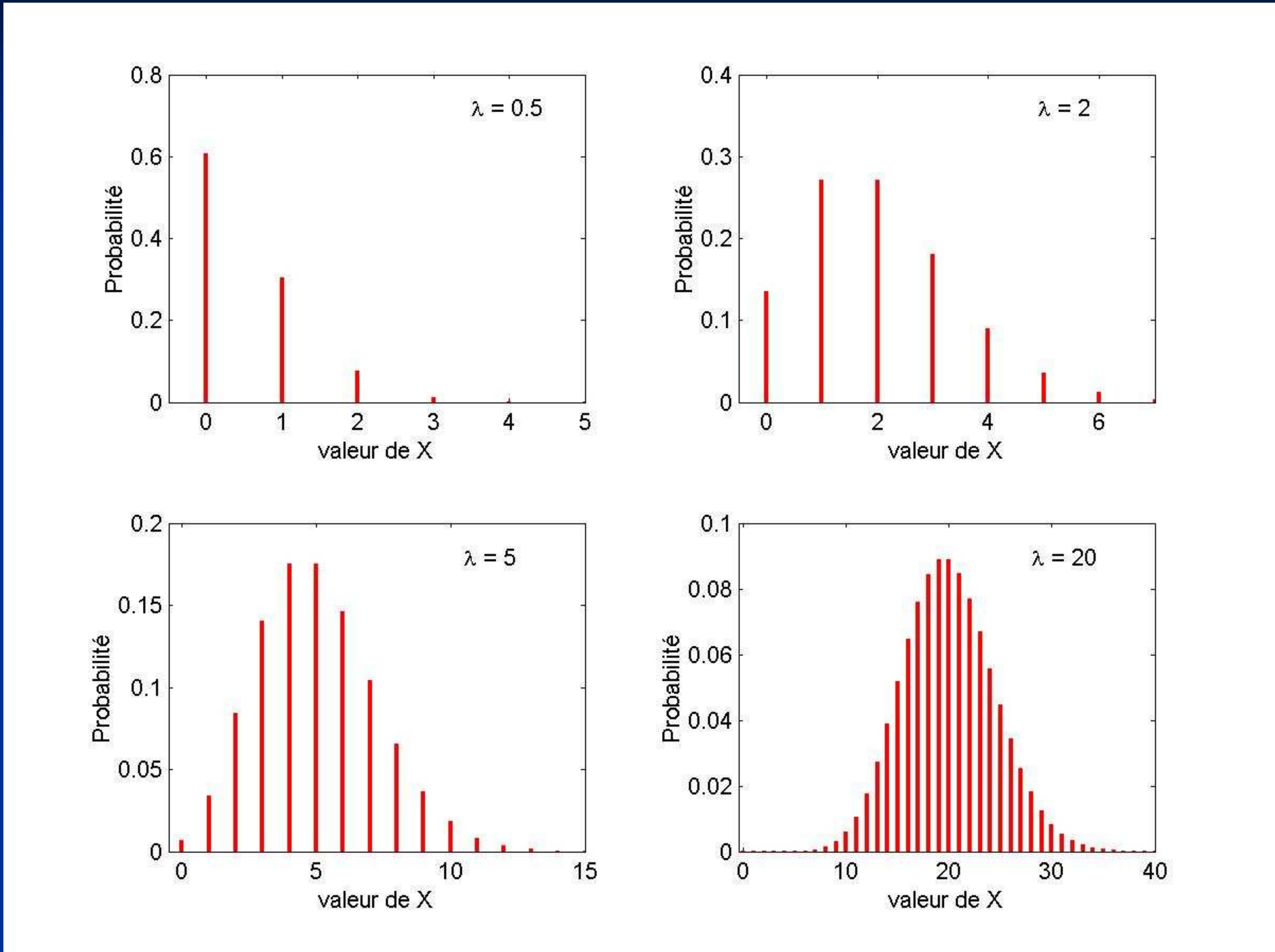
Loi de probabilité: $$p(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{avec } \lambda > 0$$

Elle est appelée **loi de Poisson**, notée **P(λ)**

$$E(X) = \lambda$$

$$V(X) = \lambda$$

Représentation graphique: diagramme en bâtons



Approximation de la loi Binomiale par une loi de Poisson

Lorsque le nombre d'épreuves dans une loi Binomiale tend vers l'infini et que la probabilité de succès tend vers 0, une v.a. binomiale tend vers la loi de Poisson

$X \sim \mathcal{B}(n, p)$ et $n>=50$ et $np<=5$ alors $X \to \mathcal{P}(\lambda)$ avec $\lambda=np=E(X)$

Exemple:

Dans une population, la proba. d'apparition d'une maladie génétique est de 0.001.

On tire un échantillon de 100 individus.

Quelle est la proba. que 2 individus soient malades ?

X: nb d'individus parmi 100

$X \sim \mathcal{B}(100, 0.001)$

$$P(X=2) = \frac{100!}{2!\,98!}(0.001)^2\,(0.999)^{98} = 0.00449$$

Approximation par la loi de Poisson:

$\lambda = np = 0.1 \qquad X \sim \mathcal{P}(0.1)$

$$P(X=2) = e^{-0.1}\frac{(0.1)^2}{2!} = 0.00452 \sim 0.00449$$

Lois discrètes

6. Loi géométrique (loi de Pascal)

Une variable géométrique correspond au nombre d'épreuves de Bernoulli nécessaires pour obtenir 1 succès

<u>Exemple</u>: En période d'épidémie de grippe, X = « nombre de patients se présentant chez le médecin avec une autre maladie respiratoire entre deux cas de grippe »

Loi de probabilité: $$p(X = n) = q^{n-1}\,p$$

$$E(X) = \frac{1}{p}$$

$$V(X) = \frac{q}{p^2}$$

Lois discrètes

6. Loi Binomiale négative

Une variable binomiale négative correspond au nombre d'épreuves nécessaires pour obtenir k succès:

Exemple: En période d'épidémie de grippe, X = « nombre de patients se présentant chez le médecin avec une autre maladie respiratoire avant le k-ième cas de grippe »

Loi de probabilité: $$p(X = n) = C_{n-1}^{k-1}\, q^{n-k}\, p^{k}$$

$$E(X) = \frac{k}{p}$$

$$V(X) = k\,\frac{q}{p^{2}}$$

Lois discrètes

7. Loi hypergéométrique

N boules dans une urne, n boules blanches (les autres noires)
X = « nombre de boules blanches parmi les m premières boules tirées »

Exemple: échantillonnage

Loi de probabilité:

$$p(X = k) = \frac{C_n^k \, C_{N-n}^{m-k}}{C_N^m}$$

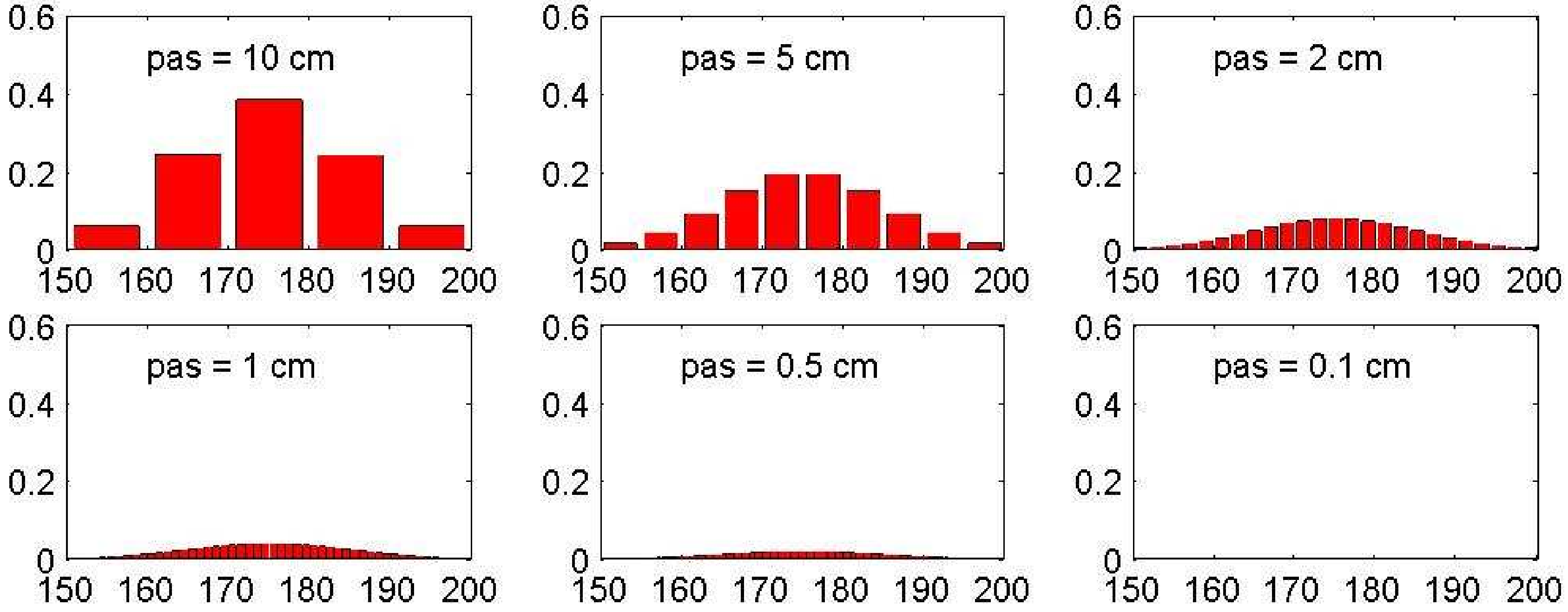
$$E(X) = \frac{n\,m}{N}$$

$$V(X) = \frac{m\ n\ (N - n)\ (N - m)}{(N - 1)N^2}$$

Variable aléatoire continue

On mesure la taille chez tous les hommes de France
Représentation = discrétisation avec un pas donné

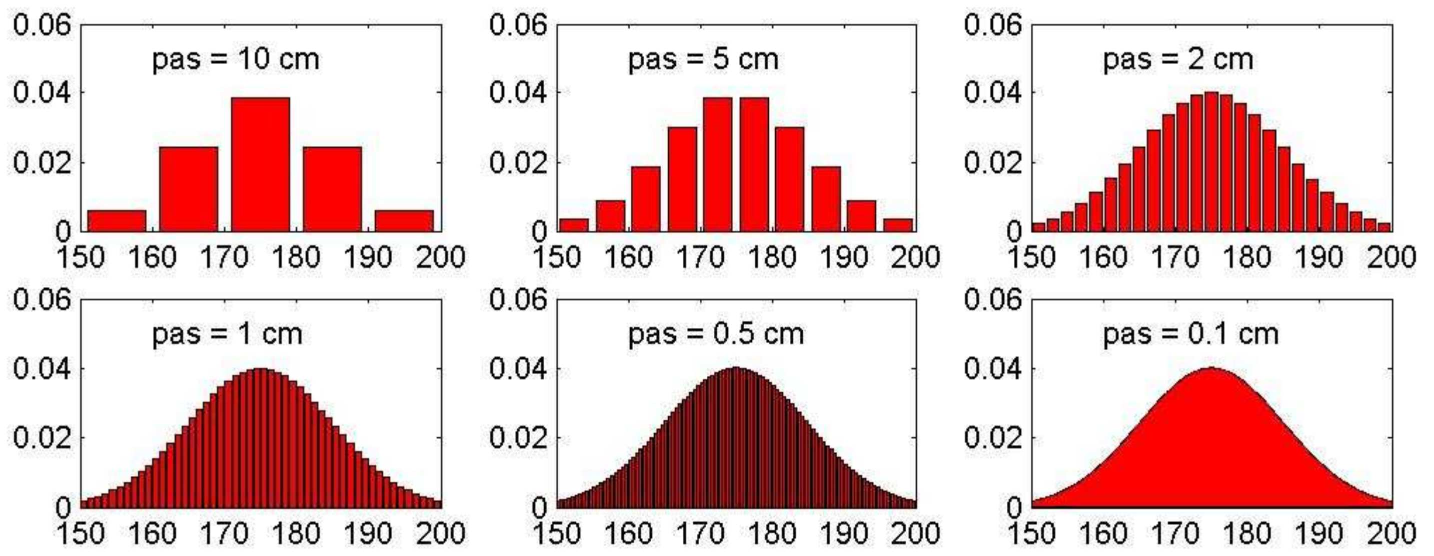
Ex: pas = 10 cm : P(150<X<160), …, P(190<X<200)



Variable aléatoire continue

Plus le pas est petit moins il y a de gens dans l'intervalle
$\Rightarrow$ On ne voit plus rien

Pour avoir une mesure indépendante du pas, on remet à l'échelle en divisant par le pas: on parle de densité de probabilité: P(150<X<150+pas) / pas, …, P(200-pas<X<200) / pas



Avec un pas suffisamment petit on retrouve toujours la même courbe bien lisse qui ne dépend plus du pas:
c'est la fonction densité de probabilité f de la variable aléatoire X

Variable aléatoire continue

Remarque:

On ne peut mesurer que $\mathbf{p(a \leq X \leq b)}$ car $\mathbf{p(X=a)=0}$

On a alors:

$$p(a < X < b) = \int_a^b f(x)dx$$

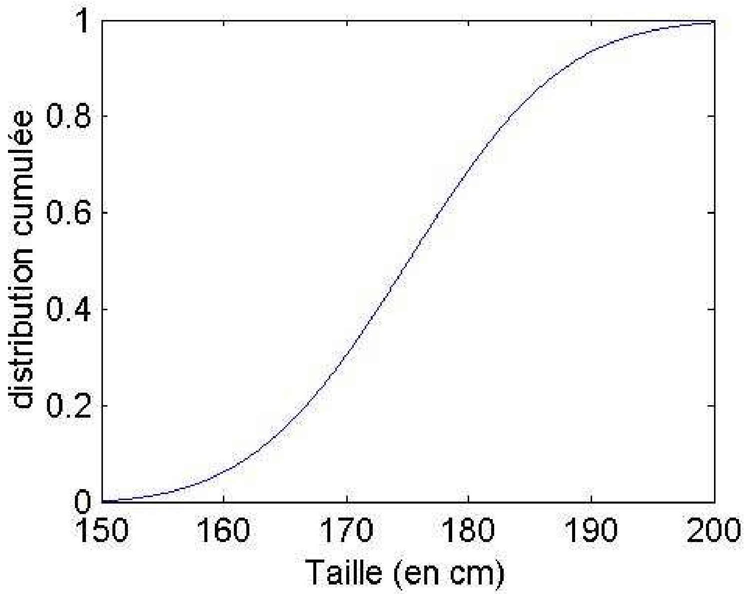
Malheureusement (ou heureusement), en pratique le plus souvent on ne peut pas calculer ces intégrales

$\Rightarrow$Table (cf TD pour leur utilisation)

Variable aléatoire continue

Fonction de répartition

$$F(u) = p(X < u) = \int_{-\infty}^{u} f(x)dx$$



La fonction de répartition correspond aux probabilités cumulées associées à la v.a. X continue sur l'intervalle d'étude.

Cette fonction est croissante et varie entre 0 et 1

Variable aléatoire continue

Espérance

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

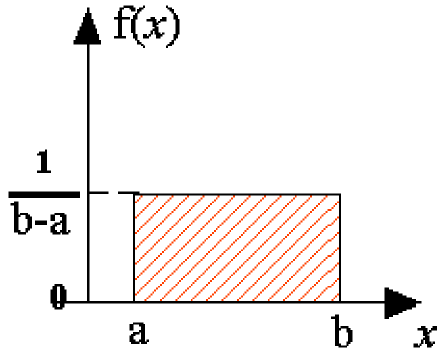
Variance

$$V(X) = \int_{-\infty}^{+\infty} f(x)\left[x - E(X)\right]^2 dx = E(X^2) - \left[E(X)\right]^2$$

Lois continues

1. Loi uniforme continue

Une variable continue est **uniforme** sur [a,b] si la valeur de sa fonction densité de probabilité est constante entre a et b et nulle en dehors



Fonction de densité de probabilité

Densité de probabilité:

$$f(x) = \frac{1}{b-a} \text{ si } x \in [a,b] \text{ et } f(x) = 0 \text{ si } x \notin [a,b]$$

$$p(\alpha < X < \beta) = \int_\alpha^\beta \frac{1}{b-a} dx = \frac{1}{b-a}[x]_\alpha^\beta = \frac{\beta-\alpha}{b-a}$$

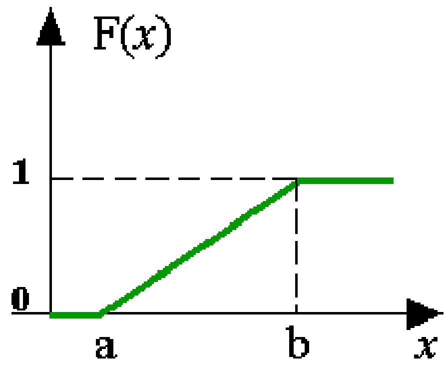
Lois continues

1. Loi uniforme continue

Fonction de répartition:

$$F(x) = \frac{x-a}{b-a} \text{ si } x \in [a,b]$$

$$F(x) = 0 \text{ si } x < a, \; F(x) = 1 \text{ si } x > b$$



Fonction de répartition

$$E(X) = \frac{b+a}{2}$$

$$V(X) = \frac{(b-a)^2}{12}$$

Lois continues

2. Loi de Gauss ou loi normale

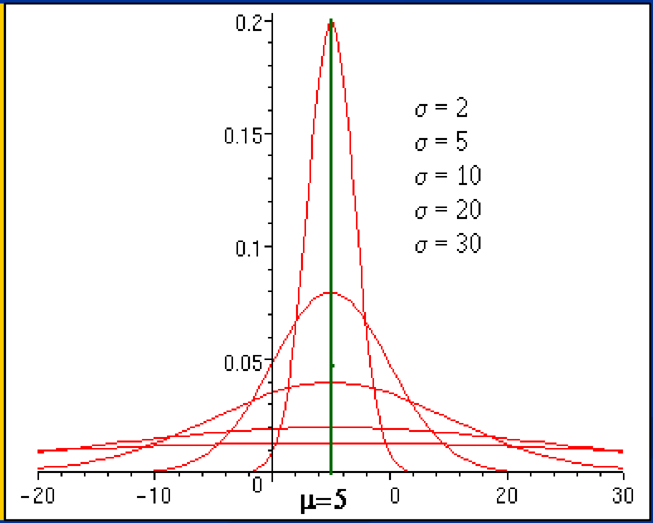
La loi la plus souvent utilisée en biologie (cf théorème central limite)

Densité de probabilité:

notée

$N(\mu, \sigma)$

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



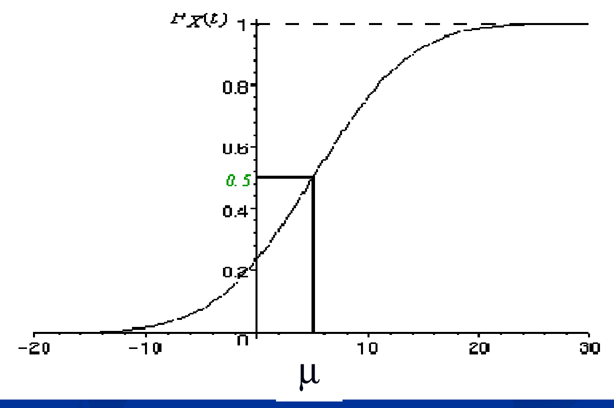
Lois continues

2. Loi de Gauss ou loi normale

Fonction de répartition

$$F(u) = \int_{-\infty}^{u} \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

NON CALCULABLE → TABLES (CF TD)



$$E(X)=\mu$$

$$V(X) = \sigma^2$$

Un petit problème de convention:

La notation $X \sim N(\mu,\sigma)$ signifie: « X suit une loi normale de moyenne μ et d'écart type σ » (convention du cours)

Exemple: N(1,4)= « loi normale de moyenne 1 et d'écart type 4 »

Autre convention: (cf par exemple page wikipedia sur la loi normale)

Le deuxième nombre entre parenthèses indique la variance

Sur l'exemple, N(1,4) signifierait « loi normale de moyenne 1 et de variance 4 (donc d'écart type 2) » (pas la convention du cours)

Lois continues

3. Loi normale réduite

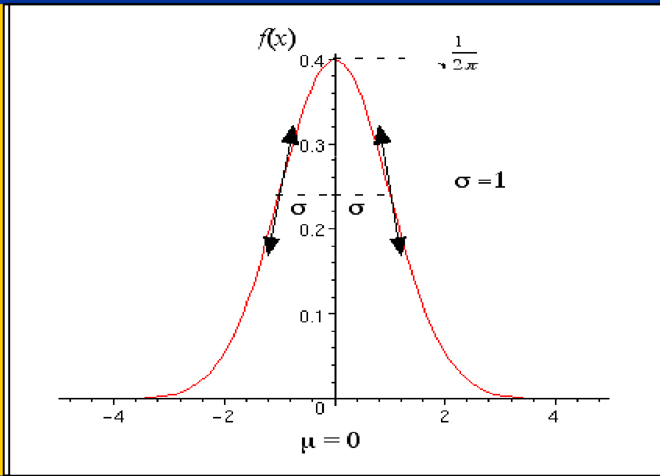
Une variable normale réduite a pour densité de probabilité :

Densité de probabilité:

notée

$N(0,1)$

$$f(x) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{1}{2}x^2}$$



$E(X)=0$

$V(X) = 1$

Lois continues

Centrer et réduire une loi normale

Considérons une varible X suivant une loi normale de moyenne μ et d'écart type σ: X~N(μ, σ)

Alors: $$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

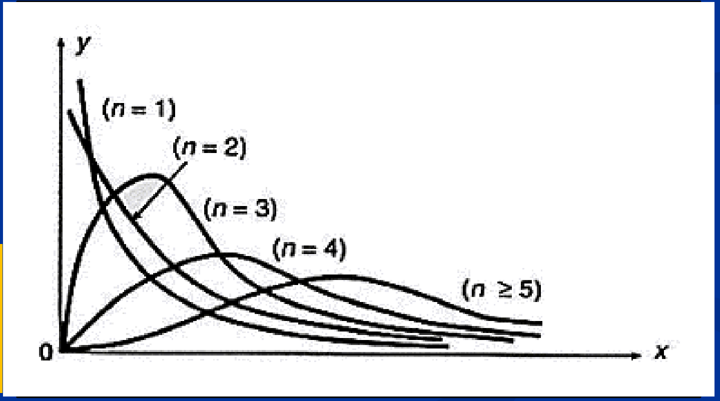
Lois continues

4. Loi du χ^2 de Pearson

On appelle χ^2 à n degrés de liberté la variable aléatoire définie par :

$$\chi^2 = X_1^2 + X_2^2 ... + X_i^2 .... + X_n^2 \quad \text{avec} \quad X_i \sim N(0,1)$$

Interviendra par la suite p.e.
pour établir une liaison entre VA



Densité de probabilité compliquée
⇒TABLES (CF TD)

$$E(\chi^2) = n$$

$$V(\chi^2) = 2n$$

Lois continues

Autre exemple de variable suivant un χ^2
X = « taille des étudiantes à Lyon 1»

On choisit deux étudiantes au hasard:
X1 = « taille de la première étudiante »
X2 = « taille de la deuxième étudiante »

La distance de la taille de chaque étudiante à la moyenne est une VA suivant une N(0,1)

On mesure de combien ces deux étudiantes s'écartent de la moyenne:

$$C = \left(\frac{X_1 - \mu}{\sigma}\right)^2 + \left(\frac{X_2 - \mu}{\sigma}\right)^2$$

Alors **$C \sim \chi^2$ (2 ddl)**

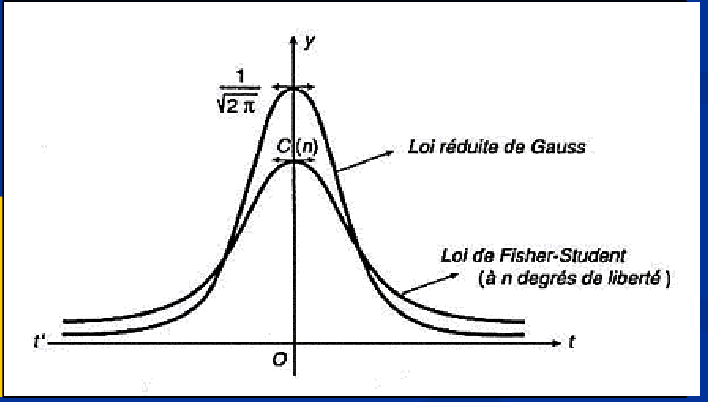
Lois continues

6. Loi de Student

On appelle T à n degrés de liberté la variable aléatoire définie par :

$$T = \frac{U}{\sqrt{V/n}} \quad \text{avec} \quad U \sim \mathcal{N}(\mathbf{0,1}) \text{ et } V \sim \chi^{\mathbf{2}} \text{ à n ddl}$$

Interviendra par la suite p.e. dans l'estimation de moyenne



Densité de probabilité compliquée
⇒TABLES (CF TD)

$E(T)=0$

$$V(T) = \frac{n}{n-2}$$

Théorème central limite

Considérons la ***n* variables aléatoires indépendantes et de même loi** **$X_1 \ldots X_n$** de moyenne μ et d'écart type σ.

Si n est suffisamment grand (souvent n>30), alors on a à peu près:

$$X_1 + X_2 + \ldots + X_n \sim N(n\mu, \sqrt{n}\sigma)$$

Ou encore

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Théorème central limite

Exemple d'application: approximation de la loi binomiale par une loi normale

Rappel: Loi binomiale = « Nombre de succès de la répétition de n épreuves de Bernouilli indépendantes »

Xi = « succès de la i-ème répétition » (*ex*: faire pile au i-ème lancé)
Xi = 1 si succès, 0 si échec

X = « nombre total de succès » = X1+…+Xn

D'après le TCL:

$$X = X_1 + X_2 + \ldots + X_n \sim N(n\mu, \sqrt{n}\sigma)$$

Pour une loi de Bernouilli: $\mu = p, \sigma = \sqrt{pq}$

Donc $X \sim N(np, \sqrt{npq})$

L'approximation est valable dès que n≥30, np ≥ 5 et nq ≥ 5