STA401

Statistique et Calcul des Probabilités

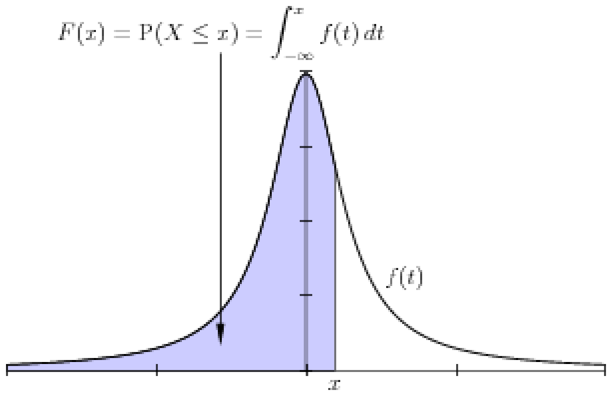
Responsable *: Carole Durand-Desprez*

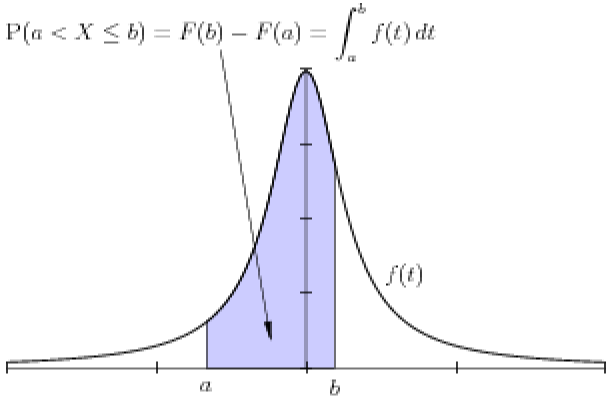
CHAPITRE 2 : Lois continues "autour de la loi Normale"

1. Lois continues - Généralités

X une variable aléatoire, Ω = I (intervalle de toutes les réalisations) $\longrightarrow$ continue

Les probabilités sont définies sur des intervalles par la fonction de répartition
$\Leftrightarrow$ aire sous la courbe de densité f.





$$P(\Omega) = 1$$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

$$f(x) = F'(x)$$

$$P(a \leq X \leq b) = \int_{a}^{b} f(t)dt$$

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$$

$$P(X \geq a) = 1 - P(X \leq a)$$

$$E[X] = \mu = \int x\, f(x) dx$$
$$V[X] = \sigma^2 = E\left[X - E[X]\right]^2 = E[X^2] - E[X]^2$$
$$V[X] = \int (x-\mu)^2 f(x) dx = \int x^2 f(x) dx - \mu^2$$

Propriétés : (identiques au cas discret)

$$E[aX + bY] = aE[X] + bE[Y]$$
$$V[X] = E\left[X - E[X]\right]^2 = E[X^2] - (E[X])^2$$
$$V[aX + b] = a^2 V[X]$$

Inégalité de Markov : X v.a. positive, pour tout a > 0 :

$$\boxed{P(X \geq a) \leq \frac{E(X)}{a}}$$

Dém :
$$E(X) = \int_0^\infty x\, f(x) dx = \int_0^a x\, f(x) dx + \int_a^\infty x\, f(x) dx \geq \int_a^\infty x\, f(x) dx$$
$$E(X) \geq a \int_a^\infty f(x) dx \geq aP(X \geq a)$$

Inégalité de Bienaymé Tchebychev

Soit X v.a. positive, pour tout a > 0 :

$$\boxed{P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}}$$

La loi Uniforme continue sur [a ; b] (vue au lycée)

Densité : $$f(x) = \frac{1}{b-a} \quad si \quad a \le x \le b \,; \quad f(x) = 0 \quad sinon$$

Espérance et variance : $$E[X] = \frac{a+b}{2} \qquad V[X] = \frac{(b-a)^2}{12}$$

La loi Exponentielle de paramètre λ (avec λ > 0) :

Densité : $$f(x) = \lambda e^{-\lambda x} \quad si \quad x \ge 0 \,; \quad f(x) = 0 \quad sinon$$

Espérance : $$E[X] = \frac{1}{\lambda}$$

Variance : $$V[X] = \frac{1}{\lambda^2}$$

(2 IPP)

Démo : $$E[X] = \int_0^{\infty} \lambda x e^{-\lambda x} dx = \lim_{t\to\infty} \left[-x e^{-\lambda x}\right]_0^t + \lim_{t\to\infty} \int_0^t e^{-\lambda x}\, dx \qquad (IPP)$$

$$E[X] = \lim_{t\to\infty} \left[-x e^{-\lambda x}\right]_0^t - \frac{1}{\lambda} \lim_{t\to\infty} \left[e^{-\lambda x}\right]_0^t = \frac{1}{\lambda}$$

Propriété : $$P(X > s + t \mid X > t) = P(X > s)$$ (loi sans mémoire, sans vieillissement)

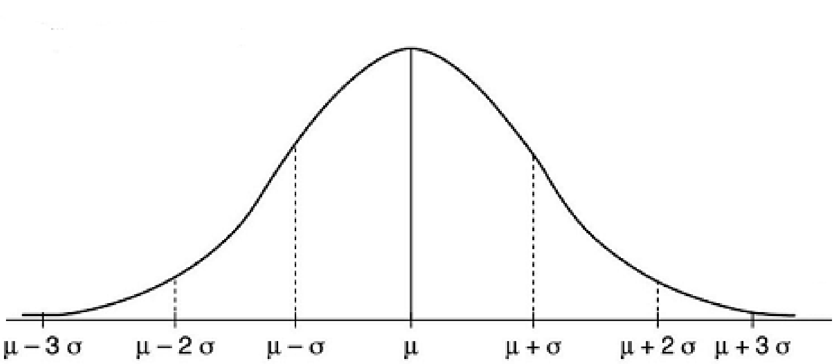
Démo : $$P(X > s + t \mid X > t) = \frac{P(X > s + t)}{P(X > t)} \;; \; de\ plus \quad P(X > s + t) = e^{-\lambda(s+t)}$$

$$et \quad P(X > t) = e^{-\lambda t}. \;\; On\ en\ déduit\ l'égalité\ voulue.$$

2. La loi Normale (Gauss) :

La loi Normale $N(\mu;\sigma^2)$:

Densité : $$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

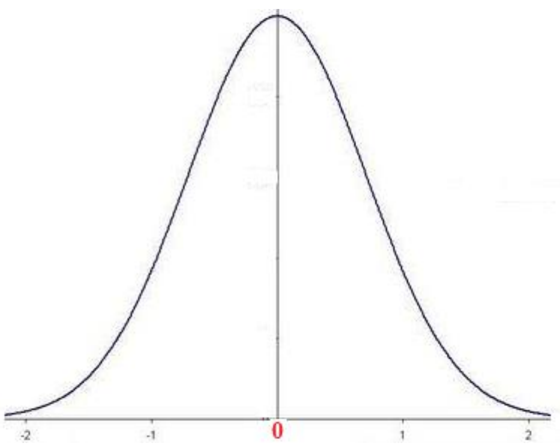


Espérance : $$E[X]=\mu$$

Variance : $$V[X]=\sigma^2$$

La loi Normale centrée réduite $N(0;1)$:

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



Espérance : $$E[X]=0$$

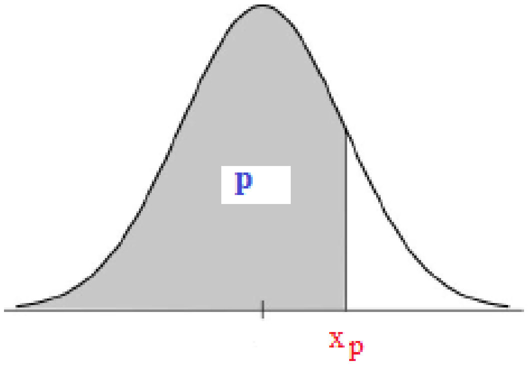
Variance : $$V[X]=1$$

Lecture des tables statistiques (pas de calcul d'intégrales !) :

Loi de Gauss (Normale), Student, Khi-deux, Fisher ...

Notation importante :

$$x_p \quad quantile \;\; d'ordre \;\; p: \quad F(x_p) = P(X \leq x_p) = p$$



Convention – Notation :

Pour la loi Normale centrée-réduite, on notera u_p le quantile d'ordre p

Pour la loi de Student, on notera t_p le quantile d'ordre p

Pour la loi du Khi-deux, on notera z_p le quantile d'ordre p

Pour la loi de Fisher, on notera f_p le quantile d'ordre p

Propriétés :

Si X suit $N(\mu;\sigma^2)$ et a un réel $\longrightarrow$ $X+a$ suit $N(\mu+a\ ;\ \sigma^2)$

Si X suit $N(\mu;\sigma^2)$ et a un réel $\longrightarrow$ aX suit $N(a\mu\ ;\ a^2\sigma^2)$

Si X suit $N(\mu_1;\sigma_1^2)$ et Y suit $N(\mu_2;\sigma_2^2)$ indépendantes $\longrightarrow$ $X+Y$ suit $N(\mu_1+\mu_2\ ;\ \sigma_1^2+\sigma_2^2)$
$\longrightarrow$ $X-Y$ suit $N(\mu_1-\mu_2\ ;\ \sigma_1^2+\sigma_2^2)$

Si X_i suit $N(\mu;\sigma^2)$, $i=1,\dots,n$ indépendantes $\longrightarrow$ $X_1+\dots+X_n$ suit $N(n\mu\ ;\ n\sigma^2)$

$\longrightarrow$ $\dfrac{X_1+\dots+X_n}{n}$ suit $N(\mu\ ;\ \sigma^2/n)$

Si X suit $N(\mu;\sigma^2)$ $\longrightarrow$ $Y=(X-\mu)/\sigma$ suit $N(0;1)$ ***[centrage-réduction]***

Exemples :

Si X suit $N(3;4)$ $\longrightarrow$ $2X-1$ suit $N(5\ ;\ 16)$

Si X suit $N(3;4)$ et Y suit $N(1;9)$ indépendantes $\longrightarrow$ $(2X-Y)$ suit $N(5\ ;\ 25)$
$\longrightarrow$ $(2X-Y-5)/5$ suit $N(0\ ;\ 1)$

Lecture de table de la loi Normale N(0;1) - Exemples :

Si X suit la loi Normale *N(0;1)* [loi tabulée] ……. Lecture directe :

$P(X < 0{,}51) = p \longrightarrow p = 0{,}695$
$P(X < u) = 0{,}95 \longrightarrow u = 1{,}6449$
$P(X > u) = 0{,}2$ alors $P(X < u) = 0{,}8 \longrightarrow u = 0{,}8416$

Lecture de table et centrage-réduction

Si X suit la loi $N(\mu;\sigma^2)$ alors $Y = (X-\mu)/\sigma$ suit la loi $N(0;1)$

- $P(X < a) = p \Rightarrow P((X-\mu)/\sigma < (a-\mu)/\sigma) = P(Y < (a-\mu)/\sigma) = p$

  $\longrightarrow$ lecture de u_p sur la table avec : $u_p = (a-\mu)/\sigma \Rightarrow a = \mu + \sigma\, u_p$

- Si X suit la loi $N(3;4)$ alors $Y = (X-3)/2$ suit la loi $N(0;1)$:

  - Si $P(X < 4) = p \Rightarrow P(Y < (4-3)/2) = p \Rightarrow P(Y < 1/2) = p \longrightarrow p = 0{,}6915$
  - Si $P(X < a) = 0{,}9 \Rightarrow P(Y < (a-3)/2) = 0{,}9 \Rightarrow P(Y < u) = 0{,}9$

    $\longrightarrow$ lecture de $u_{0,9} = 1{,}2816 = (a-3)/2 \Rightarrow a = \mu + \sigma\, u = 3+2\text{x}1{,}2816 = 5{,}5632$

3. Théorème Central Limite : (Tous les théorèmes de convergence seront vus en L3)

Théorème Central Limite :

$Soient\ \ X_1,\ldots,X_n\ \ des\ variables\ aléatoires\ indépendantes\ de\ même\ loi, avec$

$$E(X_i)=\mu\quad et\quad V(X_i)=\sigma^2.\quad Soit\quad Y=\sum X_i\quad alors, si\ n\ est\ grand$$

$$\frac{Y-n\mu}{\sigma\sqrt{n}}\quad suit\ \ la\ \ loi\ \ N(0;1)\quad ou\quad Y\ suit\ \ la\ loi\ \ N(n\mu;n\sigma^2)$$

En pratique, on prendra : n>30

Cas particulier :
Moivre-Laplace

Approximation d'une loi Binomiale par une loi Gaussienne

Soient X_i de loi *B(p)* indépendantes, alors $\mu = p$ et $\sigma^2 = p(1\text{-}p)$

On sait que $Y=\sum_i X_i$ suit $B(n;p)$

Si n est grand, alors Y suit approx. $N\big(np;np(1-p)\big)$

$$\boxed{B(n;p)\approx N\big(np;np(1-p)\big)}$$

En pratique, on prendra : n>30, np>5, n(1-p)>5

Approximation d'une loi de Poisson par une loi Gaussienne

Soit Y une v.a. de loi P(λ). Si λ est grand, alors Y suit approx. $N(\lambda;\lambda)$

$$\boxed{P(\lambda) \approx N(\lambda;\lambda)}$$

En pratique, on prendra : λ>25

Démo : Soit (X_i) un échantillon de taille n de loi P(λ/n), alors Y= Σ X_i suit une loi P(λ) – cf chapitre pécédent -. De plus, E(X_i) = V(X_i) = λ/n. D'après le TCL, Y suit une loi $N(\lambda;\lambda)$

Exemple d'approximation :

Une certaine boîte mail affirme que la probabilité qu'un courriel soit classé en tant que 'spam' alors qu'il ne l'est pas réellement est p=1%.

On prend un échantillon de 1000 mails. X suit la loi Binomiale (1000; 1%)

$$P(X \leq 15) = \binom{1000}{0} 0{,}01^{0}(0{,}99)^{1000} + \dots\dots\dots\dots\dots\dots + \binom{1000}{15} 0{,}01^{15}(0{,}99)^{985} = ???$$

Approximation : *n grand* $\longrightarrow$ *n > 30 ; np=10 > 5 ; n(1-p)=990 > 5*

$$B(n;p) \approx N\big(np; np(1-p)\big) \qquad ici \qquad B(1000;0{,}01) \approx N(10;9{,}9)$$

$$P(X \leq 15) = P\big((X-10)/\sqrt{9{,}9} \leq (15-10)/\sqrt{9{,}9}\big) = P(Y \leq 1{,}59) = 0{,}9441$$

$\longrightarrow$ *Lecture N(0;1)*

4. Intervalle de fluctuation de la proportion :

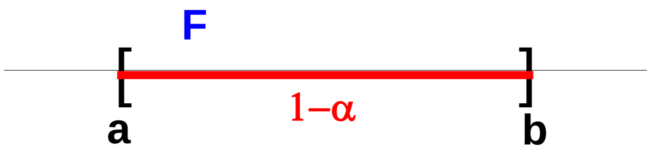
Soit p la probabilité de succès d'un évènement A. Soit un n-échantillon indépendant.

X : “ le nombre d'individus pour lequel l'évènement A a réussi “ , X suit la loi B(n,p)

La fréquence *F=X/n* représente la proportion (dans l'échantillon) de réussites de A.

But : Trouver un intervalle [a ; b] qui contient F avec une probabilité 1 - α :

$$P(a \leq F \leq b) = 1 - \alpha$$



Si n est grand, alors F suit approximativement une loi *N(p, p(1-p)/n)* – TCL (conditions).

$$P\left[\underbrace{(a-p)/\sqrt{p\,(1-p)/n}}_{-u_{1-\alpha/2}} \leq (F-p)/\sqrt{p\,(1-p)/n} \leq \underbrace{(b-p)/\sqrt{p\,(1-p)/n}}_{u_{1-\alpha/2}}\right] = 1-\alpha$$

$$[a;b] = \left[p \, - u_{1-\alpha/2}\sqrt{p\,(1-p)/n} \;\; ; \;\; p \, + u_{1-\alpha/2}\sqrt{p\,(1-p)/n}\right]$$

Bon échantillonnage : si $F\ est\ suffisamment\ proche\ de\ p$

si $F \in [a; b]$

si $X \in [na; nb]$

Exemple :

Dans une boîte mail un courriel a une probabilité de1% d'être classé comme un spam à tort. On prend un échantillon de taille 1000 mails. Soit X le nombre de mails classés „spam“ à tort. On repère 14 mails mal classés dans cet échantillon.
L'échantillon est-il ' bon', 'conforme' ? L'échantillon est-il “représentatif de la population totale“ ?

On cherche à savoir si le nombre de mails mal classés X est cohérent, donc si la fréquence F est bien dans l'intervalle de fluctuation à 95% :

$$I = [0{,}01 - 1{,}96 * 0{,}0031464; 0{,}01 + 1{,}96 * 0{,}0031464] = [0{,}00383; 0{,}01617]$$

Pour 14 mails mal classés dans l'échantillon, on a F= 0,014. On remarque que F est bien dans l'intervalle I. Ou bien : $14 \in [3{,}83; 16{,}17]$

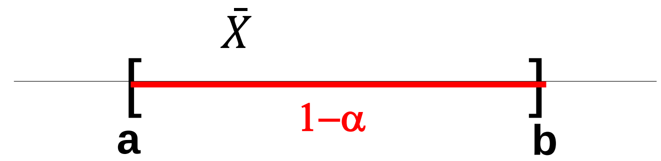
L'échantillon est donc représentatif (ou conforme).

Intervalle de fluctuation de la moyenne :

Soit X une variable de loi *N(μ;σ^2)*. Soit un n-échantillon indépendant.
La moyenne empirique $\bar{X}$ suit la loi *N(μ;σ^2/n)*. (μ et σ connus)

But : Trouver un intervalle [a ; b] qui contient $\bar{X}$ avec une probabilité 1 - α :

$$P(a \leq \bar{X} \leq b) = 1 - \alpha$$



Raisonnement analogue à l'intervalle de fluctuation de la proportion : $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ suit une loi *N(0;1)*. Donc,

$$P\left[\underbrace{\frac{a-\mu}{\sigma/\sqrt{n}}}_{-u_{1-\alpha/2}} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \underbrace{\frac{b-\mu}{\sigma/\sqrt{n}}}_{u_{1-\alpha/2}}\right] = 1 - \alpha$$

$$\boxed{[a;b] = \left[\mu - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \; ; \; \mu + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

***Idem** pour l'intervalle de fluctuation asymptotique*
- n grand - dans le cas où X ne serait pas Gaussienne
(à faire en exo)

Exemple :

Le temps de compilation d'une page de codes avec la version 1 d'un logiciel était en moyenne de 0,2 seconde avec un écart type de 0,09 seconde. On suppose que ce temps suit une loi Normale.
On s'intéresse à la version 2 du même logiciel. On prend un échantillon de 150 pages, la moyenne de compilation est de 0,18. Calculer l'intervalle de fluctuation de la moyenne à un niveau de confiance de 0,98, puis conclure si le temps de compilation moyen a varié entre les deux versions.

Dans la version 1, les paramètres sont connus. On cherche à savoir si cette nouvelle version est 'conforme' ou pas à l'ancienne.
Intervalle de fluctuation à 98% de la moyenne :

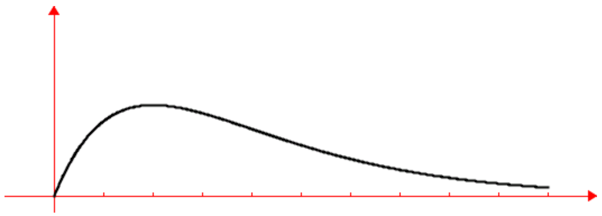
$$I = \left[0{,}2 - 2{,}3263 * 0{,}09/\sqrt{(150)}\,; 0{,}2 + 2{,}3263 * 0{,}09/\sqrt{(150)}\right] = [0{,}183; 0{,}217]$$

On remarque que la moyenne $\bar{x} = 0{,}18$ n'est pas dans l'intervalle I, par valeur inférieure.
On conclut que le temps moyen de compilation a varié depuis l'ancienne version, il a diminué avec un niveau de confiance de 98%.

5. Quelques autres lois usuelles (liens avec la loi normale) :

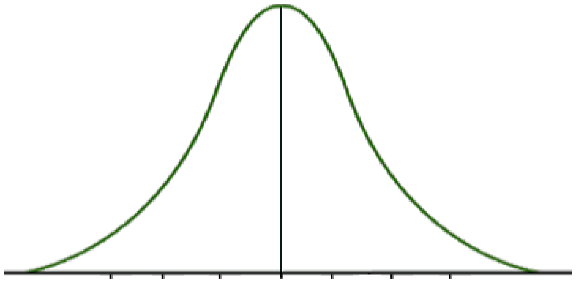
- Soient $X_1, \dots, X_n$ n variables indépendantes de loi $N(0;1)$:

$$Z = \sum X_i^2 \quad suit\ la\ loi\ du\ Khi-deux\ \ \aleph_n^2$$



- Soient X et Y deux variables indépendantes, avec X suit une loi $N(0;1)$ et Y suit une loi $\aleph_n^2$:

$$T = \frac{X}{\sqrt{Y/n}} \quad suit\ la\ loi\ de\ Student\ \ T_{(n)}$$



- Soient X et Y deux variables indépendantes de deux lois du Khi-deux à n et p degrés de liberté :

$$F = \frac{X/n}{Y/p} \quad suit\ la\ loi\ de\ Fisher\ \ F_{(n;p)}$$