quelques formules de dénombrement

- nombre de manières de placer Amandine, Bertrand et Cécile sur un banc ?

  permutations : il y a $n! = 1 \times 2 \times \ldots \times n$ manières d'ordonner n éléments

- urne avec 10 numéros, 3 tirages avec remise, nombre de tirages (ordonnés) possibles ?

  ***p*-listes** : le nombre de manières de fabriquer une liste (ordonnée) de n-éléments tous compris entre 1 et p est p^n

quelques formules de dénombrement

- 10 candidats à une épreuve sportive, combien de podiums possibles ?

arrangements : le nombre de manières de classer p personnes choisies parmi n est

$$A_n^p = n \times (n-1) \times \ldots \times (n-p+1) = \frac{n!}{(n-p)!}$$

- 10 candidats pour trois postes identiques, combien de recrutements possibles ?

combinaisons : le nombre de manières de choisir p éléments parmi n est

$$\binom{n}{p} = \frac{n \times (n-1) \times \ldots \times (n-p+1)}{p \times (p-1) \times \ldots \times 1} = \frac{n!}{p!(n-p)!}$$

attention : simplifier les factorielles aux numérateur/dénominateur !

évènements, loi de probabilité

- **expérience aléatoire**, issue, univers Ω (fini, dénombrable, non dénombrable)
- **évènement** A : partie de Ω (évènement élémentaire : partie à un élément)
- évènement contraire $\bar{A}$, union ("ou"), intersection ("et")
- évènements **incompatibles** : si $A \cap B = \emptyset$

- **probabilité** : $p : \mathcal{P}(\Omega) \to [0;1]$ telle que
  $$p(\Omega) = 1 \quad \text{et} \quad p(A \cup B) = p(A) + p(B) \text{ si } A, B \text{ incompatibles}$$
  conséquences :
  - $0 \leq p(A) \leq 1$ pour tout événement A
  - $p(\emptyset) = 0, \quad p(\Omega) = 1$
  - $p(\bar{A}) = 1 - p(A)$ pour tout événement A
  - $p(A \cup B) + p(A \cap B) = p(A) + p(B)$ pour tous événements A, B

- si des évènements sont **indépendants** : $p(A \cap B) = p(A) \times p(B)$ (voir plus loin)

- on parle d'espace probabilisé (Ω, p)

- **univers finis** : $p(A) = p(a_1) + \ldots + p(a_n)$ **équiprobabilité** : $p(A) = \dfrac{\text{card } A}{\text{card } \Omega}$

- **univers continus** : fonction densité, on remplace les sommes par des intégrales, les événements élémentaires par les segments

exemples élémentaires

- **dé équilibré à 6 faces**
  $A =$ « obtenir un nombre impair »,
  $B =$ « obtenir un nombre inférieur ou égal à 4 » :

  $p(A) = 1/2$, $p(B) = 2/3$, $p(A \cap B) = 1/3$, $p(A \cup B) = 5/6$

- **radioactivité**
  probabilité que n atomes se désintègrent en une durée t : $p_n = \frac{\Lambda^n t^n}{n!} e^{-\Lambda t}$

  $p($« moins de 3 atomes se désintègrent » $)$ = $p_0 + p_1 + p_2 + p_3$
  $p($« au moins 10 atomes se désintègrent »$)$ = $1 - (p_0 + ... + p_9)$

- **proba uniforme sur** $[0, 1]$
  fonction densité ?

  $p(\mathbb{R}) = 1$, $p([0, 1]) = 1$, $p([-1/2, 2/3]) = 2/3$

  $p(0) = p(1/2) = p(5) = 0$

- **remarque :** pas de proba uniforme sur un ensemble infini discret

exercice

Une femme est enceinte de jumeaux.

Quel(s) évènement(s) est (sont) le(s) plus fréquent(s) :

1 - elle attend deux garçons ?

2 - elle attend deux filles ?

3 - elle attend un garçon et une fille ?

probabilités conditionnelles

- **exemple :** on jette deux dés équilibrés ;
  probabilité d'obtenir une somme de 6 sachant que l'un des dés a fait un 2 ?

- Si $p(A) \neq 0$, on définit la **probabilité de B sachant A**

$$p(B|A) = \frac{p(A \cap B)}{p(A)}$$

- **exemple :** 24 alternants sur 26 ont leur DUT, 67 non-alternants sur 72

  $p(\text{« DUT »}|\text{« Alt »}) = 24/26$, $p(\text{« DUT »}|\text{« non Alt. »}) = 67/72$
  $p(\text{« Alt. »}|\text{« DUT »}) = 24/91$, $p(\text{« non Alt. »}|\text{« DUT »}) = 67/91$

- Plus généralement, $p(A|B) = \dfrac{p(B|A)p(A)}{p(B)}$

formules

- **formule des probabilités totales :**
  $A_1, \ldots, A_n$ incompatibles et B ne se produit que si l'un des A_i s'est produit :
  $p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + \ldots + p(B|A_n)p(A_n)$

  en particulier, pour les deux évènements A et $\bar{A}$:
  $p(B) = p(B|A)p(A) + p(B|\bar{A})p(\bar{A})$

- **formule de Bayes :**

$$p(A_k|B) = \frac{p(A_k)p(B|A_k)}{\sum_{i=1}^{n} p(A_i)p(B|A_i)}$$

- **exemple :** maladie rare touchant une personne sur 10000
  Un test détecte 99% des personnes infectées, avec $0,5\%$ de « faux positifs »
  Probabilité qu'une personne dont le test est positif soit effectivement malade ?

- **évènements indépendants :** si $p(A|B) = p(A)$ ou $p(B|A) = p(B)$ ou $p(A \cap B) = p(A)p(B)$

  attention, A et B ne sont jamais incompatibles **et** indépendants, sauf si l'un est impossible

variables aléatoires

- **variable aléatoire** : application $X : \Omega \to \mathbb{R}$
  - résultat obtenu en jetant un dé
  - mesure d'une tige métallique
  - nombre de pannes quotidiennes d'une machine
  - note à un examen
- **fonction de répartition** de X : $F_X(x) = p(]-\infty, x[) = p(X < x)$

  croissante, positive, limite 0 en $-\infty$, limite 1 en $+\infty$
- $p(X \geq x) = 1 - p(X < x) = 1 - F_X(x)$

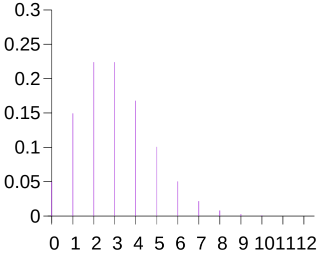
  pour une variable continue, inégalité stricte ou large : $p(X < x) = p(X \leq x)$

exemples de variables aléatoires discrètes

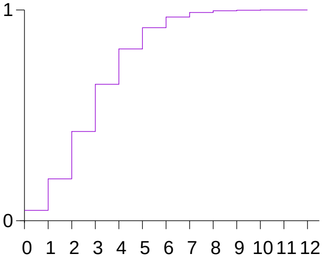
- **exemple 1 :** pile ou face, $X = 0$ si pile et $X = 1$ si face
  $p(X = 1) = 1/2$, $p(X = 0) = 1/2$

- **exemple 2 :** désintégration nucléaire
  X le nombre d'atomes désintégrés sur un intervalle de temps fixé :
  $p(X = k) = \dfrac{\Lambda^k}{k!}e^{-\Lambda}$ (où Λ nombre moyen d'atomes désintégrés sur l'intervalle)

valeurs des $p_k = p(X = k)$



fonction de répartition

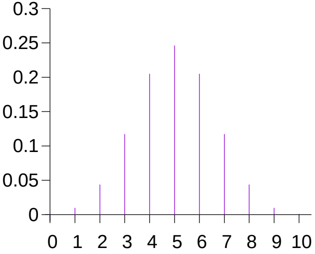


exemple de variables aléatoires discrètes

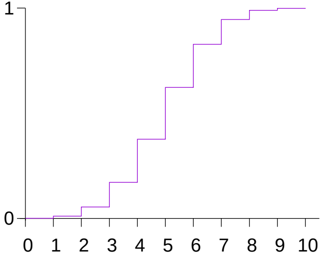
- **exemple 3 :** X nombre de résultats « face » en 10 lancers d'une pièce équilibrée ?

$$p(X = k) = \frac{\binom{10}{k}}{2^{10}}$$

valeurs des $p_k = p(X = k)$



fonction de répartition



espérance et variance

- pour une variable X de valeurs $x_1, \ldots, x_n$ avec probas $p_1, \ldots, p_n$:
  - **espérance** $E(X) = \displaystyle\sum_{i=1}^{n} p_i x_i = p_1 x_1 + p_2 x_2 + \ldots$
    et si f fonction : $E(f(X)) = \sum_{i=1}^{n} p_i f(x_i) = p_1 f(x_1) + p_2 f(x_2) + \ldots$
  - **variance** $\mathrm{Var}(X) = \displaystyle\sum_{i=1}^{n} p_i (x_i - E(X))^2 = p_1 (x_1 - E(X))^2 + p_2 (x_2 - E(X))^2 + \ldots$
  - **écart-type** $\sigma(X) = \sqrt{\mathrm{Var}(X)}$
- **propriétés**
  - l'espérance est linéaire : $E(X + Y) = E(X) + E(Y)$ et $E(aX) = aE(X)$
  - $E(X + b) = E(X) + b$
  - $\mathrm{Var}(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$
  - $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$ et $\sigma(aX) = |a|\sigma(X)$
  - $\mathrm{Var}(X + b) = \mathrm{Var}(X)$ et $\sigma(X + b) = \sigma(X)$
  - X centrée si $E(X) = 0$, réduite si $\mathrm{Var}(X) = 1$

    $\dfrac{X - E(X)}{\sigma(X)}$ est centrée réduite

variables aléatoires continues

- variable X continue :
  - densité f_X : $f_X(x)dx = p(x \leq X < x + dx)$ remplace p_i,
    les intégrales remplacent les sommes
  - $p(X \in A) = \int_A f_X(x)\,dx$
  - $F_X(x) = p(X < x) = \int_{-\infty}^{x} f_X(x)\,dx$
  - $p(a \leq X \leq b) = \int_a^b f_X(x)\,dx = F_X(b) - F_X(a)$
  - $p(X = x) = 0$,
- **espérance** $E(X) = \displaystyle\int_{\mathbb{R}} x\, f_X(x) \;\; dx$
- **variance** $\mathrm{Var}(X) = \displaystyle\int_{\mathbb{R}} f_X(x)\,(x - E(X))^2 \;\; dx$
- **écart-type** $\sigma(X) = \sqrt{\mathrm{Var}(X)}$
- et les mêmes propriétés (somme, produit, ...) que dans le cas discret

couples de variables aléatoires discrètes

- X et Y variables discrètes sur un même univers : on considère le **couple** (X, Y)

  loi de (X, Y) ou **loi conjointe** de X et Y : c'est l'ensemble des $p_{xy} = p(X = x, Y = y)$

$$p((X, Y) \in A) = \sum_{(x,y)\in A} p_{xy}$$

- **loi marginale** en X : $p(X = x) = \sum_y p(X = x, Y = y) = \sum_y p_{xy}$
  loi marginale en Y : $p(Y = y) = \sum_x p(X = x, Y = y) = \sum_x p_{xy}$

- X et Y **indépendantes** si $p(X = x, Y = y) = p(X = x) \times p(Y = y)$ pour tous x, y

  intérêt : la connaissance des lois marginales suffit pour connaître la loi conjointe

couples de variables aléatoires discrètes

- **exemple :** 4 boules indiscernables marquées 1 à 4, on en tire deux.
  X le plus petit des deux numéros, Y le plus grand.
  Loi conjointe et lois marginales ? Indépendance ?

$X \mid Y$	2	3	4	loi marginale en X
1	$1/6$	$1/6$	$1/6$	$1/2$
2	0	$1/6$	$1/6$	$1/3$
3	0	0	$1/6$	$1/6$
loi marginale en Y	$1/6$	$1/3$	$1/2$	1

couples de variables aléatoires continues

- **variables continues :** densité conjointe $f(x, y)$, qui donne les probabilités

$$p((X, Y) \in A) = \int\int_A f(x, y)\, dxdy$$

- **lois marginales**

$$f_X(x) = \int_{\mathbb{R}} f(x, y)\, dy \quad \text{et} \quad f_Y(y) = \int_{\mathbb{R}} f(x, y)\, dx$$

- **indépendance :** si pour tous a, b, c, d,

$$p(a \leq X \leq b, c \leq Y \leq d) = p(a \leq X \leq b) \times p(c \leq Y \leq d)$$

propriétés

- $E(X+Y) = E(X) + E(Y)$
- **dans le cas général**, $E(XY) = \sum_{x,y} p(X=x, Y=y)xy$
  (ou $E(XY) = \int\int f(x,y)xy\, dxdy$ pour des variables continues)
- **si** X, Y **indépendantes**, $E(XY) = E(X)E(Y)$
  et $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
- **covariance** $\mathrm{cov}(X,Y) = E(\,(X - E(X))(Y - E(Y))\,)$
  et **coefficient de corrélation** $\rho(X,Y) = \dfrac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)}$
- $\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y)$
  si X et Y indépendantes, $\mathrm{cov}(X,Y) = \rho(X,Y) = 0$ (**réciproque fausse !!!**)
- $\mathrm{Var}(aX+bY) = a^2\mathrm{Var}(X) + 2ab\mathrm{cov}(X,Y) + b^2\mathrm{Var}(Y)$

retour à l'exemple

$X \mid Y$	2	3	4	loi marginale en X
1	$1/6$	$1/6$	$1/6$	$1/2$
2	0	$1/6$	$1/6$	$1/3$
3	0	0	$1/6$	$1/6$
loi marginale en Y	$1/6$	$1/3$	$1/2$	1

- $E(X) = 1/2 \times 1 + 1/3 \times 2 + 1/6 \times 3 = 5/3$
  $E(Y) = 1/6 \times 2 + 1/3 \times 3 + 1/2 \times 4 = 10/3$
- $E(XY) = 35/6$
- covariance $5/18 \neq 0$, donc les variables ne sont pas indépendantes

loi de Bernoulli

- **épreuve de Bernoulli** : « succés/échec » de probas p et $q = 1 - p$
- X vaut 1 en cas de succés, 0 en cas d'échec : $p(X = 0) = 1 - p$, $p(X = 1) = p$
- la loi de X est appelée **loi de Bernoulli** de paramètre p
- $E(X) = p(X = 0) \times 0 + p(X = 1) \times 1 = p$
- $\mathrm{Var}(X) = E(X^2) - E(X)^2 = p(1 - p)$

loi binomiale

- **schéma de Bernoulli** : on répète n épreuves de Bernoulli identiques indépendantes de même paramètre p

- alors la variable $X =$ « nombre de succés » suit une **loi binomiale** de paramètres n et p

- pour tout k entre 0 et n, $p(X = k) = \binom{n}{k} p^k q^{n-k}$

- $E(X) = np, \qquad \mathrm{Var}(X) = np(1 - p)$

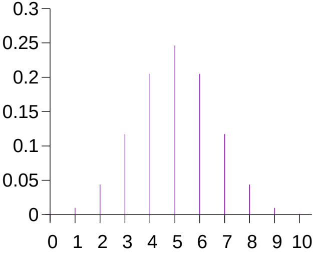
- **exemple 1 :** 5% de chances d'être admis à un concours, sur 40 candidats, espérance et variance du nombre d'admis ?
  loi binomiale de paramètres $n = 40$, $p = 0.05$, espérance 2, variance 1.90, écart-type 1.38

loi binomiale : exemple

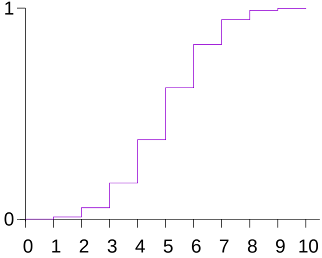
- **exemple 2 :** nombre de « face » en 10 lancers d'une pièce équilibrée ?

$$p(X = k) = \frac{\binom{10}{k}}{2^{10}}$$

valeurs des $p_k = p(X = k)$



fonction de répartition



désintégration radioactive

trois hypothèses :

- nombre moyen de désintégrations durant l'intervalle Δt : $\lambda \Delta t$
  (λ constante de radioactivité)
- deux désintégrations jamais simultanées
- désintégrations indépendantes

- alors : probabilité qu'un atome se désintègre durant $[t, t + dt]$: λdt
- calcul de $p_k(t)$, proba que k atomes se désintègrent sur $[0, t]$?

désintégration radioactive

- $p_0(t+dt) = p_0(t)(1-\lambda dt)$
- donc équation $p_0'(t) = -\lambda p_0(t)$,
- donc $p_0(t) = e^{-\lambda t} \qquad (p_0(0) = 1)$

- $p_k(t+dt) = p_k(t)(1-\lambda dt) + p_{k-1}(t)\lambda dt$
- donc $p_k'(t) = -\lambda p_k(t) + \lambda p_{k-1}(t)$
- si p_{k-1} est connu, on peut donc calculer p_k...

- $p_1'(t) = -\lambda p_1(t) + \lambda e^{-\lambda t} \qquad \Rightarrow \qquad p_1(t) = \lambda t e^{-\lambda t}$
- de même, par récurrence, pour tout k on a $p_k(t) = \dfrac{\lambda^k t^k}{k!} e^{-\lambda t}$

c'est un cas typique d'apparition de la loi de Poisson en physique

loi de Poisson

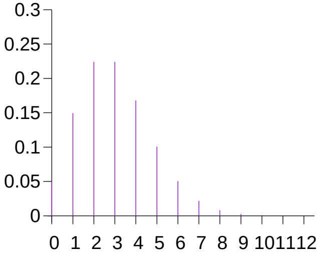
- **c'est la « lois des événements rares »**
  événements indépendants, non simultanés, en nombre moyen Λ sur une unité de temps

  alors $X =$ « nombre d'événements durant l'unité de temps » suit une **loi de Poisson de paramètre** Λ

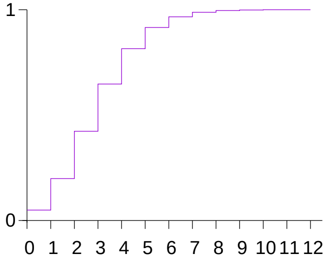
- $$p(X=k) = \frac{\Lambda^k}{k!}e^{-\Lambda} \qquad E(X) = \Lambda \qquad \mathrm{Var}(X) = \Lambda$$

- exemple $\Lambda = 3$:

valeurs des p_k



fonction de répartition



loi de Poisson

exemple type : un standard téléphonique reçoit en moyenne 10 appels par heure, donc $10 \times 5/60$ par période de 5 minutes, $10/4$ par période de 15 minutes, ...

- probabilité pour qu'en une heure il reçoive exactement 2 appels ?

$$\frac{e^{-10}10^2}{2!} \simeq 0,23\%$$

- probabilité pour qu'en 5 minutes il reçoive exactement 3 appels ?

$$\frac{e^{-5/6}(\frac{5}{6})^3}{3!} \simeq 4,2\%$$

- probabilité pour qu'en 15 minutes, il reçoive au moins 2 appels ?

$$p(X \geq 2) = 1 - p(X < 2) =$$
$$1 - p(X = 0) - p(X = 1) = 1 - e^{-5/2} - e^{-5/2}(\tfrac{5}{2}) \simeq 71,3\%$$

- probabilité pour qu'en une heure, il reçoive au plus 8 appels ?

on trouve $33,28\%$, sans calculs !

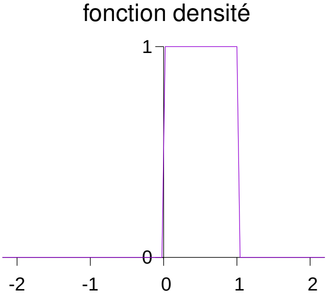
table de la loi de Poisson

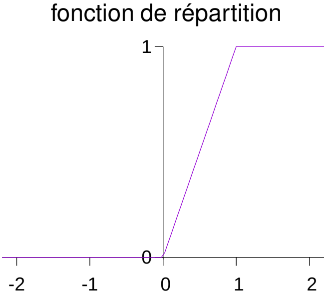
$p(X \leq 8)$ si $X \sim P(10)$?

k	$\Lambda = 10$ $p(X = k)$	$p(X \leq k)$
0	0.0	0.0
1	0.0005	0.0005
2	0.0023	0.0028
3	0.0076	0.0103
4	0.0189	0.0293
5	0.0378	0.0671
6	0.0631	0.1301
7	0.0901	0.2202
8	0.1126	0.3328
9	0.1251	0.4579
10	0.1251	0.583
11	0.1137	0.6968
12	0.0948	0.7916
13	0.0729	0.8645
14	0.0521	0.9165

loi uniforme

exemple : **loi uniforme sur** $[0, 1]$





loi uniforme

X suit une loi uniforme sur $[a;b]$ si

$$f_X(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{1}{b-a} & \text{si } x \in [a;b] \\ 0 & \text{si } x > b \end{cases},$$

$$F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } x \in [a;b] \\ 1 & \text{si } x > b \end{cases},$$

alors

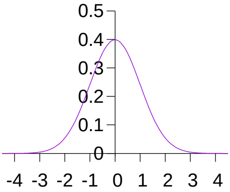
$$E(X) = \frac{a+b}{2},$$

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}, \qquad \sigma(X) = \frac{b-a}{2\sqrt{3}}$$

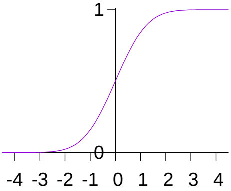
loi normale

- **loi normale** : loi empirique pour répartition des mesures, loi limite pour d'autres lois
- si $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, loi normale $N(\mu, \sigma^2)$ de paramètres μ et σ^2
- $F_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} \, dt$
- $E(X) = \mu, \qquad \mathrm{Var}(X) = \sigma^2$
- **loi normale** $N(0,1)$:

fonction densité



fonction de répartition



loi normale

$(\mu = -2, \sigma = 2)$ $(\mu = 0, \sigma = 1)$ $(\mu = \frac{5}{2}, \sigma = \frac{1}{2})$

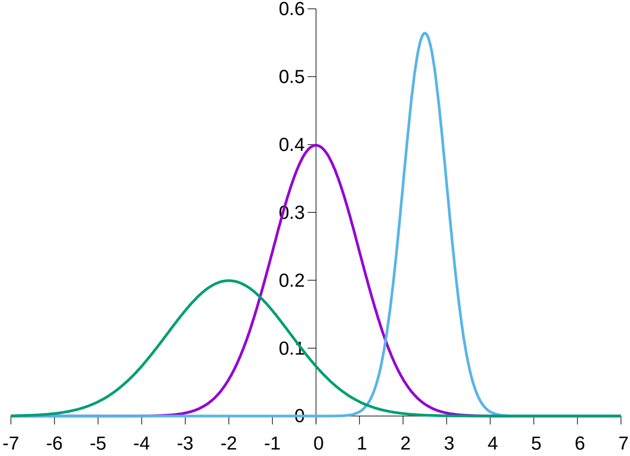


table de la loi normale centrée réduite

elle donne les $p(Z < z)$ pour les valeurs de z entre 0 et 4 :

	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

loi normale : utilisation de la table

- calcul exact de $F_X(x)$ impossible !
  on se ramène a la variable $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$:
  on applique la relation $p(X \leq b) = p(\ \ Z \leq \frac{b-\mu}{\sigma}\ \ )$,
  puis les $p(Z < z)$ sont directement lus dans la table

- **exemple :** si $X \sim N(1,4)$
  - $p(X \leq 3)$? c'est $p(\dfrac{X-1}{2} \leq 1)$, soit 0.8413
  - $p(X > 2)$? c'est $p(\dfrac{X-1}{2} > 1/2) = 1 - p(\dfrac{X-1}{2} < 1/2) = 0.3085$
  - $p(-2.5 < X \leq 3)$? c'est $p(-1.75 < \dfrac{X-1}{2} < 1) = p(-1.75 < Z < 1) =$
    $p(Z < 1) - p(Z < -1.75) = 0.8413 - (1 - p(Z < +1.75)) =$
    $0.8413 - 1 + p(Z < 1.75) = 0.8413 - 1 + 0.9599 = 0.8012$
  - α tel que $p(X \leq \alpha) = 50\%$? (α médiane !)
    $p(\dfrac{X-1}{2} \leq \dfrac{\alpha-1}{2}) = 50\%$, donc $\alpha - 1 = 2 \times 0$, $\alpha = 1$ (médiane = moyenne)

loi normale

interprétation de l'écart-type

- si $X \sim N(\mu, \sigma^2)$, on a $E(X) = \mu$, $\sigma(X) = \sigma$

$$\begin{aligned}
p(E(X) - t\sigma(X) < X < E(X) + t\sigma(X)) &= p(\mu - t\sigma < X < \mu + t\sigma) \\
&= p(-t < \frac{X-\mu}{\sigma} < t) \\
&= p(-t < Z < t) \\
&= p(Z < t) - p(Z < -t) \\
&= p(Z < t) - 1 + p(Z < t) \\
&= 2p(Z < t) - 1
\end{aligned}$$

- $t = 1, 2, 3$?

  Les $p(Z < t)$ valent 0.8413, 0.9772 et 0.9987,

  donc les $p(\mu - t\sigma < X < \mu + t\sigma)$ valent 68.26%, 95.44% et 99.74%

loi normale

- **produit d'une loi normale par une constante**
  si $X \sim N(\mu, \sigma^2)$ et $a \in \mathbb{R}$, alors
  $$aX \sim N(a\mu, a^2\sigma^2)$$
- **somme de deux lois normales**
  si $X \sim N(\mu_1, \sigma_1^2)$ et $Y \sim N(\mu_2, \sigma_2^2)$, et si X, Y **indépendantes**, alors
  $$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$
- **somme de** ***n*** **lois normales**
  si $X_1, X_2, \ldots, X_n$ sont des variables **indépendantes** de même loi $N(\mu, \sigma^2)$, alors
  $$X_1 + \ldots + X_n \sim N(n\mu, n\sigma^2)$$
- **moyenne de lois normales**
  si $X_1, X_2, \ldots, X_n$ sont des variables **indépendantes** de même loi $N(\mu, \sigma^2)$, alors
  $$\bar{X} = \frac{X_1 + \ldots + X_n}{n} \sim N(\mu, \sigma^2/n)$$

approximation d'une loi par une autre

- $p(X = k)$, $p(X < x)$ difficiles/longs à calculer $\longrightarrow$ tables

- limiter le nombre de tables :
  - $B(n, p)$: autant de tables que de valeurs de n et p
  - $P(\Lambda)$: autant de tables que de valeurs de Λ
  - $N(\mu, \sigma^2)$: une seule table !

- on approche une loi par une autre "plus simple"

approximation d'une loi binomiale par une loi de Poisson

- p petit : événement rare, et espérance np et variance $np(1-p)$ proches
  $\longrightarrow$ on remplace $B(n,p)$ par $P(np)$ si $n \geq 30$, $p \leq 0.1$, $np \leq 5$

- **exemple :** 40 candidats, 5% de chances d'être admis.
  $n = 40 > 30$, $p = 0,05 < 0.1$, $np = 2 \leq 5$ $\longrightarrow$ on remplace $B(40, 0.05)$ par $P(2)$.

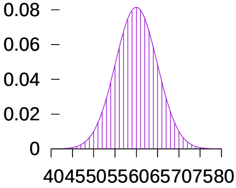
k	$p(X = k)$ pour $B(40, 0.05)$	$p(X = k)$ pour $P(2)$
0	0.1285	0.1353
1	0.2706	0.2707
2	0.2777	0.2707
3	0.1851	0.1804
4	0.0901	0.0902
5	0.0342	0.0361
6	0.0105	0.0120
7	0.0027	0.0034
8	0.0006	0.0009
9	0.0001	0.0002
10+	0.0000	0.0000

et, par exemple, $p(X \leq 5) \simeq 0.9834$ (valeur exacte 0.986..)

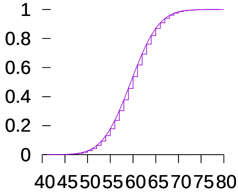
approximation d'une loi binomiale par une loi normale

- si p pas petit, $B(n,p)$ approchée par $N(np, np(1-p))$
  ($n \geq 50$ et $np(1-p) \geq 18$)
- mais $p(X = k) = 0$ pour une loi continue $\longrightarrow$ $p(k - 0.5 < X < k + 0.5)$
- **exemple :** $B(100, 0.6)$ et $N(60, 24)$, graphique :

densité et valeur des p_k



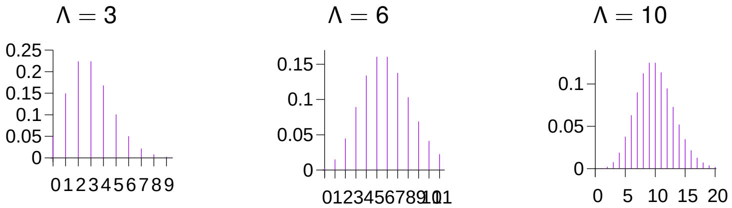
fonctions de répartition



	$B(n=100, p=0.6)$	$N(\mu = np = 60, \sigma^2 = 24)$	$N(\mu = np = 60, \sigma^2 = 24)$
k	$p(X=k)$	$F(k+\frac{1}{2}) - F(k-\frac{1}{2})$	$F(k+1) - F(k)$
40-	0.0	0.0	0.0
50	0.0103	0.0102	0.0125
55	0.0478	0.0484	0.0534
60	0.0812	0.0813	0.0809
65	0.0491	0.0484	0.0434
70	0.01	0.0102	0.0082
79+	0.0	0.0	0.0

approximation d'une loi de Poisson par une loi normale

- $\Lambda > 10$: la loi de Poisson devient symétrique



- on remplace $P(\Lambda)$ par $N(\Lambda, \Lambda)$

- $p(X = k) \longrightarrow p(k - 0.5 < X < k + 0.5)$
  $p(X \leq k) \longrightarrow p(X \leq k + 0.5)$

loi faible des grands nombres

lien entre probabilités et statistiques :

- jeu de pile ou face : "pile" avec proba $1/2$
- on joue une fois : soit "pile", soit "face"
- on joue deux fois : pas forcément un "pile", un "face"...
- mais si f est la fréquence de "pile", si le nombre de tirages augmente, f tend vers $1/2$

énoncé de la **loi faible des grands nombres** :

dans une expérience aléatoire, on considère un événement de probabilité p.
Si on répète l'expérience n fois en notant f_n la fréquence d'apparition du résultat,
$f_n \to p$ quand $n \to +\infty$.

mais **le hasard n'a pas de mémoire**

échantillonnage

- X variable de loi connue, espérance μ, variance σ^2
- propriétés d'un échantillon de taille n : moyenne, variance ?
- on étudie la variable $\bar{X} = \dfrac{X_1 + \ldots + X_n}{n}$ où les X_i sont des variables indépendantes de même loi que X
- $\bar{X}$: même espérance que X, mais variance σ^2/n
- probabilité que les valeurs effectivement mesurées soient proches de μ ?

échantillonnage : loi normale

- $X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n)$

- **exemple :** on fabrique des disques de diamètre $D \sim N(15, 0.25^2)$.

  Diamètre d'un disque au hasard ? Dans 95.4% des cas dans
  $[15 - 2 \times 0.25, 15 + 2 \times 0.25] = [14.5, 15.5]$

  On en prend 10 au hasard, diamètre moyen du lot ?
  La variance est $0.25^2/10$ soit 0.00625, l'écart-type 0.079

  $\bar{D} \sim N(15, 0.00625)$ donc dans 95.4% des cas dans
  $[15 - 2 \times 0.079, 15 + 2 \times 0.079] = [14.842, 15.158]$

  Dans 98% des cas dans $[15 - t \times 0.079, 15 + t \times 0.079]$ avec t tel que
  $2F(t) - 1 = 0.98$ soit $F(t) = 0.99$.

  On utilise la table de la loi normale : $t = 2.33$, un intervalle de confiance 98% pour le diamètre moyen d'un lot de 10 disques est
  $[15 - 2.33 \times 0.079, 15 + 2.33 \times 0.079] = [14.817, 15.183]$

échantillonnage : loi quelconque

théorème central limite
si n est assez grand et si $X_1, X_2, \dots, X_n$ sont des variables aléatoires identiques et indépendantes, d'espérance μ et variance σ^2,

$\bar{X} \sim N(\mu, \sigma^2/n)$

exemple : fabrique de piles, tension espérance 1.6V, écart-type 0.1V.

Tension moyenne sur un échantillon de 50 ? On considère qu'elle suit une loi $N(1.6, 0.1^2/50)$ (écart-type 0.0141) sans faire l'hypothèse que c'est le cas pour chaque tension d'une pile.

échantillonnage : application aux pourcentages

- si une proportion p de la population possède une caractéristique donnée, quelle sera la proportion de personnes à la posséder dans un échantillon de n personnes ?
- X nombre de personnes concernées dans l'échantillon : $X \sim B(n, p)$.
- Pour n « grand » ($n \geq 50$ et $np(1-p) \geq 18$) : $X \sim N(np, np(1-p))$.
- On s'intéresse à la fréquence $f = X/n$, soit X multiplié par $1/n$.
- Alors l'espérance est multipliée par $1/n$, la variance par $1/n^2$:
- ainsi, $f \sim N(p, p(1-p)/n)$.

échantillonnage : application aux pourcentages

exemple : 10% de gauchers dans la population.
Probabilité d'en avoir plus de 15% sur les 250 étudiants du département MP ?

$250 > 50$ et $250 \times 0.9 \times 0.1 > 18$ donc on peut utiliser l'approximation par une loi normale.
$f \sim N(0.1, 0.09/250)$ donc
$p(f > 0.15) = p(Z > 2.635) = 1 - p(Z < 2.64) = 1 - 0.9959 = 0.004.$

échantillonnage/estimation

- **échantillonnage** :
  on connaît la population, on veut en déduire les propriétés des échantillons.
  facile

- **estimation** :
  on connaît un échantillon, on veut estimer les propriétés de la population.
  plus dur !

  - **estimateur robuste** : peu sensible à de rares valeurs extrèmes
  - **estimation sans biais** : si on fait la moyenne sur un très grand nombre d'échantillons,

    on tend vers la valeur exacte
  - plus l'échantillon est important, plus l'estimation est bonne

estimation : deux exemples

exemple 1 : diamètre de 300 tiges

diamètre d en mm	42	43	44	45	46	47	48
effectif	13	26	63	149	30	15	4

exemple 2 : 5 mesures d'une tension u : en volts, 11.83, 11.91, 12.01, 12.03, 12.1

estimation : estimation ponctuelle

X variable aléatoire d'espérance μ et de variance σ^2 inconnues,
mais on connait un échantillon de valeurs $x_1, x_2, \ldots, x_n$

estimation de la moyenne : $E(X) = \mu \simeq \bar{x}$

ici, diamètre moyen 44.73, tension 11.97

estimation : estimation ponctuelle

estimation de la variance : prendre

$$\sigma^2 = \mathrm{Var}(X) \qquad \simeq \qquad \sigma'^2 = \frac{(x_1 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n} \text{ ?}$$

non : c'est un **estimateur biaisé** !

grand nombre de mesures, sur plusieurs échantillons de taille n :

moyenne des variances estimées $\longrightarrow \dfrac{n-1}{n}\sigma^2$, on **sous-estime** la variance

$s^2 = \dfrac{(x_1 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1}$ est l'estimateur non biaisé de σ^2

ici, $\sigma' = 1.136$, $s = 1.138$ pour le premier exemple,
$\sigma' = 0.095$, $s = 0.106$ pour le second.

estimation : estimation par intervalle de confiance

problème de l'estimation ponctuelle : deux échantillons, deux valeurs de μ !

estimation par intervalles de confiance :
on veut déterminer un intervalle qui contient μ avec une probabilité maîtrisée

pour obtenir une affirmation de la forme :

« le diamètre moyen a 95% de chances d'être dans $[44.60, 44.86]$ »

ou plutôt :

« un intervalle de confiance 95% pour le diamètre est $[44.60, 44.86]$ »

estimation : intervalle de confiance de la moyenne, cas idéal

cas idéal : X suit une loi normale d'espérance μ et variance σ^2, avec σ connu (peu réaliste mais simple, pour commencer). On veut estimer μ à partir d'un échantillon.

Plus précisément on cherche un intervalle qui ait une probabilité fixée de contenir μ.

Alors $\bar{X} \sim N(\mu, \sigma^2/n)$, donc $\sqrt{n}\dfrac{\bar{X}-\mu}{\sigma} \sim N(0,1)$ (loi centrée réduite).

Si $t > 0$, $p(\mu - t\sigma/\sqrt{n} \leq \bar{X} \leq \mu + t\sigma/\sqrt{n}) = p(-t \leq \sqrt{n}\dfrac{\bar{X}-\mu}{\sigma} \leq t) = 2F(t) - 1$,

en soustrayant $\bar{X}$ et μ puis en multipliant par -1 :
$p(\bar{X} - t\sigma/\sqrt{n} \leq \mu \leq \bar{X} + t\sigma/\sqrt{n}) = 2F(t) - 1$.

Ainsi, l'intervalle $[\bar{X} - t\sigma/\sqrt{n}, \bar{X} + t\sigma/\sqrt{n}]$ a une probabilité $2F(t) - 1$ de contenir μ.

Si on veut un intervalle de confiance c fixée, il reste à trouver à l'aide de la table de la loi normale c tel que $2F(t) - 1 = c$.

Ainsi dans l'exemple des tiges on peut dire que μ est dans $[44.60, 44.86]$ avec une confiance 95%, ou dans $[44.53, 44.92]$ avec confiance 99.7%.

estimation : intervalle de confiance de la moyenne, cas général

cas général : si X suit une loi normale d'espérance μ et variance σ^2, tous deux inconnus ;

si on note $\bar{X}$ la moyenne de n tirages, et s l'estimation ponctuelle de σ,

alors $\sqrt{n}\dfrac{\bar{X}-\mu}{s}$ suit une loi de Student à $n-1$ degrés de liberté.

Donc comme dans l'exemple idéal précédent, une fois fixé le **risque** α (5%, 1%, etc...), la table de la loi de Student à $\nu = n-1$ degré de liberté la valeur $t(\alpha)$ telle que

$$p(\bar{X}-t(\alpha)s/\sqrt{n} \leq \mu \leq \bar{X}+t(\alpha)s/\sqrt{n}) \quad = \quad p(-t(\alpha) \leq \sqrt{n}\frac{\bar{X}-\mu}{s} \leq t(\alpha)) \quad = \quad 1-\alpha,$$

exemples : $\alpha = 5\%$, $n = 5, 10, 100, +\infty$? On trouve dans la table :

$n = 5$: $\nu = n-1 = 4$, donc $t(\alpha) = 2.7763$,

$n = 100$, $\nu = n-1 = 99 \simeq 100$, donc $t(\alpha) = 1.984$,

$n = +\infty$, $\nu = n-1 = +\infty$, donc $t(\alpha) = 1.96$: on retrouve la loi normale

estimation : intervalle de confiance de la moyenne, cas général

En résumé :

détermination d'un intervalle de confiance avec la loi de Student

si X suit une **loi normale** de paramètres μ et σ^2 inconnus,
si l'on dispose, à l'aide d'un échantillon d'effectif n, d'une estimation ponctuelle $\bar{x}$ pour l'espérance $\mu = E(X)$ et d'une estimation ponctuelle s pour l'écart-type σ,

alors $\sqrt{n}\dfrac{\bar{X}-\mu}{s} \simeq N(0,1)$, et donc :

il y a une probabilité $1-\alpha$ que l'espérance $E(X) = \mu$ soit dans l'intervalle

$$[\bar{x} - t(\alpha)\tfrac{s}{\sqrt{n}} \ , \ \bar{x} + t(\alpha)\tfrac{s}{\sqrt{n}}],$$

$t(\alpha)$ étant lu dans la table de la **loi de Student** à $n-1$ degrés de liberté.

estimation : intervalle de confiance de la moyenne, exemple

exemple :

mesures de tension, 11.83, 11.91, 12.01, 12.03, 12.1
estimation ponctuelle de la moyenne $\bar{x} = 11.97$
estimation poncutelle de l'écart-type $s = 0.106$

donc intervalle de confiance 95% ?

risque 5%, $t(0.05) = 2.7764$, et donc intervalle
$[11.97 - 2.776 \times 0.106/\sqrt{5}, 11.97 + 2.776 \times 0.106/\sqrt{5}] = [11.83, 12.10]$

(avec loi normale :
$[11.97 - 1.96 \times 0.106/\sqrt{5}, 11.97 + 1.96 \times 0.106/\sqrt{5}] = [11.88, 12.06]$, intervalle légèrement plus petit mais faussement précis)

de même au risque 1%, $t(0.01) = 4.60$, intervalle $[11.75, 12.19]$.

estimation : intervalle de confiance de la moyenne, grand effectif

Si la loi de X, la variance σ^2 et l'espérance μ sont tous inconnus,
que l'on dispose d'estimations ponctuelles pour la moyenne $\bar{x}$ et l'écart-type s,
à l'aide d'un échantillon d'effectif n, **avec** n **grand,** (supérieur à 30),
alors :

- l'espérance $E(X) = \mu$ est dans $[\bar{x} - t\frac{s}{\sqrt{n}}, \bar{x} + t\frac{s}{\sqrt{n}}]$ avec une probabilité de $2F(t) - 1$.
- l'espérance $E(X) = \mu$ est dans $[\bar{x} - \frac{s}{\sqrt{n}}, \bar{x} + \frac{s}{\sqrt{n}}]$ avec une probabilité de 68%,
- l'espérance $E(X) = \mu$ est dans $[\bar{x} - 1.96\frac{s}{\sqrt{n}}, \bar{x} + 1.96\frac{s}{\sqrt{n}}]$ avec une probabilité de 95%,
- l'espérance $E(X) = \mu$ est dans $[\bar{x} - 2\frac{s}{\sqrt{n}}, \bar{x} + 2\frac{s}{\sqrt{n}}]$ avec une probabilité de 95.4%,
- l'espérance $E(X) = \mu$ est dans $[\bar{x} - 3\frac{s}{\sqrt{n}}, \bar{x} + 3\frac{s}{\sqrt{n}}]$ avec une probabilité de 99.7%.

estimation d'une proportion

p proportion de présence d'une caractéristique au sein de la population.

X le nombre d'éléments d'un éhantillon de taille n qui possède cette caractéristique. $F = X/n$ fréquence. Alors un

estimateur ponctuel de p est la fréquence f mesurée sur l'échantillon.

Si n grand ($n \geq 100$) et $0.2 \leq f \leq 0.8$), $F \simeq N(p, p(1-p)/n)$; on remplace $p(1-p)/n$ par l'estimation ponctuelle $f(1-f)/n$:

un intervalle de confiance $2F(t) - 1$ pour la valeur de p est

$$[f - t\sqrt{\frac{f(1-f)}{n}} \;\; ; \;\; f + t\sqrt{\frac{f(1-f)}{n}}].$$

estimation d'une proportion

exemple : un sondage sur 1000 personnes prises au hasard annonce un candidat A à 52%

Proportion p réelle d'intentions de vote dans la population ?

ici $f = 0.52$ et $\sqrt{\dfrac{0.52(1-0.52)}{1000}} = 0.0158$,

donc l'intervalle de confiance 95% est $[0.52 - 1.96 \times 0.0158; 0.52 + 1.96 \times 0.0158]$ soit $[0.489, 0.551]$.

de même au risque 1% : $2F(t) - 1 = 0.99$, $F(t) = 0.995$, $t = 2.57$, intervalle $[0.479, 0.561]$

au risque 10% : $2F(t) - 1 = 0.9$, $F(t) = 0.95$, $t = 1.64$, intervalle $[0.494, 0.546]$

intervalle au risque r commençant à 0.5 ?
$0.52 - 0.0158t = 0.5$ donne $t = 1.27$, donc $F(t) = 0.898$, $2F(t) - 1 = 0.796$...
r dépasse 20%

estimation d'une proportion - étude unilatérale

attention : si la question est d'être élu ou pas élu, on cherche un intervalle limité à 0.5 à gauche, mais pouvant aller jusqu'à 1 (ou en pratique $+\infty$ ce qui change peu la probabilité associée) à droite.

on regarde $p(F - 0.0158t < p)$, avec $t = 1.27$,

et non pas $p(F - 0.0158t < p < F + 0.0158t)$

alors $p(F - 0.0158t < p) = F(1.27) = 0.898$.

on prédit que le candidat sera élu, au risque 10.2% de se tromper

Tests d'hypothèses : objectif

Mesures sur un échantillon compatibles avec les affirmations :

- « la durée de vie en fonctionnement d'un composant est de 1000h »
- « les pièces produites par les usines A et B ont la même longueur »
- « la résistance du composant suit une loi normale d'espérance 10Ω et d'écart-type 0.5Ω ».

?

Tests d'hypothèse : méthode

- H_0 **hypothèse nulle,** à tester
- **risque de première espèce** $\alpha = p(\text{"rejeter } H_0\text{" / "} H_0 \text{ est vraie"})$
- intervalle de confiance de risque α pour l'échantillon, en supposant H_0
- si la valeur pour l'échantillon est hors de l'intervalle, on refuse H_0

Tests d'hypothèse : remarques

remarque 1 : $\beta = p(''\text{accepter } H_0'' / '' H_0 \text{ est fausse}'')$**risque de seconde espèce**

remarque 2 : pas de symétrie entre H_0 et H_1 : H_0 n'est pas prouvée par un test

Test bilatéral de la moyenne

Echantillon 300 disques, diamètre moyen 44.73 mm, $s = 1.296$.

H_0 : « les disques font 45 mm »
$\alpha = 5\%$.

Si H_0, sur un échantillon de 300 disques, le diamètre moyen sera dans $[45 - 1.96 \times 1.296/\sqrt{300}, 45 + 1.96 \times 1.296/\sqrt{300}] = [44.87, 45.13]$.

44.73 n'est pas dans cet intervalle : on refuse H_0
(en prenant un risque 5% de la refuser à tort)

Test unilatéral de la moyenne

argument de vente : « la durée de vie des composants dépasse 1000h »

échantillon de 50 : durée de vie moyenne 998h, écart-type estimé $s = 15.15h$

$H_0 : \mu \geq 1000$,
$\alpha = 1\%$

si H_0 vraie, $p(\mu - ts/\sqrt{50} \leq \bar{X}) = 99\%$ pour $t = 2.33$, donc dans 99% des cas $\bar{X}$ est dans $[\mu - 2.33 \times 15.15/\sqrt{50}, +\infty[$, a fortiori dans $[1000 - 2.33 \times 15.15/\sqrt{50}, +\infty[$.

L'intervalle est $[995, +\infty[$: on ne refuse pas H_0.

Test de comparaison de moyennes

deux populations,
variables X_1 et X_2 : espérances μ_1 et μ_2, variances σ_1^2 et σ_2^2.
$\mu_1 = \mu_2$?

échantillons effectifs n_1 et n_2, moyennes $\overline{x_1}$ et $\overline{x_2}$, écarts-types estimés s_1 et s_2

$H_0 : \mu_1 = \mu_2$.
Alors $\dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \sim N(0,1)$ (d'après les règles sur les sommes de lois normales indépendantes ; si les effectifs sont "grands")

exemple : $n_1 = n_2 = 40$, $\overline{x_1} = 1025\text{h}$, $\overline{x_1} = 1070\text{h}$, $s_1 = 120\text{h}$, $s_2 = 140\text{h}$

Donc intervalle de risque 5% pour $\overline{X_1} - \overline{X_2}$:
$[-1.96\sqrt{s_1^2/n_1 + s_2^2/n_2}\ ,\ 1.96\sqrt{s_1^2/n_1 + s_2^2/n_2}] = [-57.1, 57.1]$

L'intervalle contient $1070 - 1025 = 45$, la différence n'est pas significative

Test du χ^2

hypothèse portant sur la distribution de l'ensemble des valeurs

exemple : On jette 100 fois de suite un dé à 6 faces. On obtient les résultats :

valeur	1	2	3	4	5	6
effectif	10	17	11	15	19	28

H_0 : le dé est équilibré ?

Données :

- un échantillon $(x_1 = 1, n_1 = 10), (x_2 = 2, n_2 = 17), \ldots$
  (fréquences : $f_1 = 0.1$, $f_2 = 0.17$, $f_3 = 0.11$, $f_4 = 0.15$, $f_5 = 0.19$, $f_6 = 0.28 \ldots$)
- une loi théorique $p_1 = p_2 = \ldots = p_6 \simeq 0.1667$,
  effectifs "théoriques" $np_i = 16.67$ pour chaque valeur

Test du χ^2

- H_0 : « $p_1 = \dfrac{n_1}{n}, \ldots, p_k = \dfrac{n_k}{n}$ »
- risque α
- $\nu = k - 1$ en général ($\nu = k - 2$ si loi binomiale, loi de Poisson ou loi normale)
- $\chi_o^2 = \sum_{i=1}^k \dfrac{(n_i - np_i)^2}{np_i} = n \sum_{i=1}^k \dfrac{(f_i - p_i)^2}{p_i}$
- dans la table du χ^2 à ν degrés de liberté : c^2 telle que $p(\chi^2 \geq c^2) = \alpha$

Alors : on accepte H_0 si $\chi_o^2 \leq c^2$, on refuse H_0 sinon

Test du χ^2

ici, $\chi_o^2 = \frac{6}{100}\sum_{i=1}^{6}(n_i - \frac{100}{6})^2 = \frac{6}{100}\frac{640}{3} = 12.8.$

pour $\alpha = 5\%$, $c^2 = 11.07$ dans la table du χ^2 à 5 degrés de liberté,
donc on refuse H_0, en prenant 5% de risque de le refuser à tort

pour $\alpha = 1\%$, $c^2 = 15.09$ dans la table du χ^2 à 5 degrés de liberté,
donc on accepte H_0, sans maîtriser le de risque de l'accepter à tort

Test du χ^2

remarque 1 : si effectifs théoriques np_i pas tous supérieurs à 5, regrouper plusieurs valeurs

remarque 2 : H_0 acceptée ne signifie pas que l'on a trouvé la loi de X

Test d'indépendance

deux quantités X et Y, valeurs x_i $(1 \leq i \leq r)$ et y_j $(1 \leq i \leq s)$.

Sur un échantillon : $n_{i,j}$, nombre de mesures telles que $X = x_i$, $Y = y_j$.

H_0 : X et Y sont indépendantes

$l_i = \sum_{j=1}^{s} n_{i,j}$ le nombre de mesures telles que $X = x_i$
$c_j = \sum_{i=1}^{r} n_{i,j}$ le nombre de mesures telles que $Y = y_j$

si indépendantes, $n_{i,j} \simeq N_{i,j} = \dfrac{l_i \times c_j}{n}$

Test d'indépendance

- $\chi_o^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{i,j} - N_{i,j})^2}{N_{i,j}}$

- table du χ^2 à $(r-1)(s-1)$: c^2 telle que $p(\chi^2 \geq c^2) = \alpha$

- si $c^2 \geq \chi_o^2$, on accepte H_0, sinon on la refuse.

Test d'indépendance

exemple : 3 modèles pour 200 photocopieurs qui ont 0, 1 ou plus de 2 pannes par an
Nombre de pannes est indépendant du modèle ?

	0	1	2+	
modèle 1	13	19	25	57
modèle 2	28	29	28	85
modèle 3	24	18	16	58
	65	66	69	200

Ici $l_1 = 57$, $l_2 = 85$, $l_3 = 58$, $c_1 = 65$, $c_2 = 66$, $c_3 = 69$, et $n = 200$.

Tableau si indépendance : cases $l_i \times c_j/200$:

	0	1	2+	
modèle 1	18.5	18.9	19.7	57
modèle 2	27.6	28.1	29.3	85
modèle 3	18.9	19.1	20	58
	65	66	69	200

Test d'indépendance

- $\chi_o^2 = \dfrac{(13-18.5)^2}{18.5} + \ldots + \dfrac{(16-20)^2}{20} = 5.473.$

- $\alpha = 5\%$, table du χ^2 à $(3-1)(3-1) = 4$ degrés de liberté : $c^2 = 9.4877$

- on accepte l'hypothèse.