PROBABILITES

Lois des probabilités

Faculté de Médecine, UDL, SBA
1ère année Médecine

24 Janvier 2024

Plan de cours

1. Lois de probabilité continues: Loi normale
2. Lois de probabilité discrètes
3. Exercices

Loi normale: Les fondamentaux

- La distribution normale est une distribution théorique, (idéalisation mathématique qui ne se rencontre jamais exactement dans la nature).
- Mais de nombreuses distributions réellement observées s'en rapprochent et ont cette fameuse forme de "cloche" (beaucoup d'individus autour de la moyenne, de moins en moins au fur à mesure qu'on s'en éloigne, et ceci de façon symétrique).
- Très utilisée en statistiques inférentielles : une moyenne calculée sur un échantillon est une v.a. qui tend à suivre une loi normale quand la taille de l'échantillon augmente, même si la population initiale a une tout autre distribution

Le théorème Central-limit 1

- Le TCL: la somme d'un très grand nombre de variables aléatoires de loi quelconque suit approximativement une loi normale (en fait, sans rentrer dans le détail des hypothèses, il nous dit que la variable $X = X_1 + X_2 + \ldots X_n$ tend à suivre une loi normale quand n tend vers l'infini).
- Permet de comprendre pourquoi autant de distributions observées dans la réalité ont approximativement cette forme de cloche : elles décrivent des phénomènes qui résultent de l'**addition** d'un **grand** nombre de causes de fluctuation **indépendantes**.
  - Exemple : La taille ou le poids des hommes est une $v.a.r.$ qui peut être considérée comme la somme très nombreux de $v.a.$ correspondant à des facteurs (génétique, climatique, diététiques, etc) indépendantes entre eux.

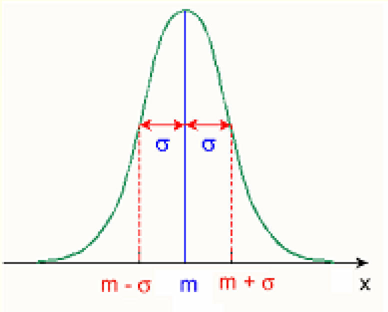
Le théorème Central-limit 3

- Permet d'approcher beaucoup de lois par une loi normale, pour peu que la variable étudiée s'exprime comme une somme d'un grand nombre de variables indépendantes.
- C'est le cas notamment de la variable binomiale (somme de n variables de Bernoulli indépendantes), dont la loi " tend à prendre la forme d'une cloche" quand n augmente.
- Cela reste possible même quand on ne connaît pas loi des variables X_i.

Loi normale centrée réduite: Définition

- Loi continue la plus importante
- Densité: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}$
- $\mathbb{E}(X) = m = 0$ (l'espérance: moyenne)
- $Var(X) = \sigma^2 = 1$ (variance)
- Écart-type $= \sigma$

Allure de la loi $\mathcal{N}(0, 1)$



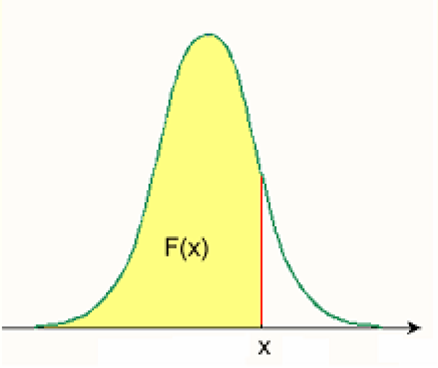
- Courbe de la densité
- Surface sous la courbe = 1
- Loi symétrique
- Axe de symétrie = espérance = m
- Maximum sur l'axe de symétrie
- Ecart-type = distance entre axe de symétrie et point d'inflexion

Fonction de répartition

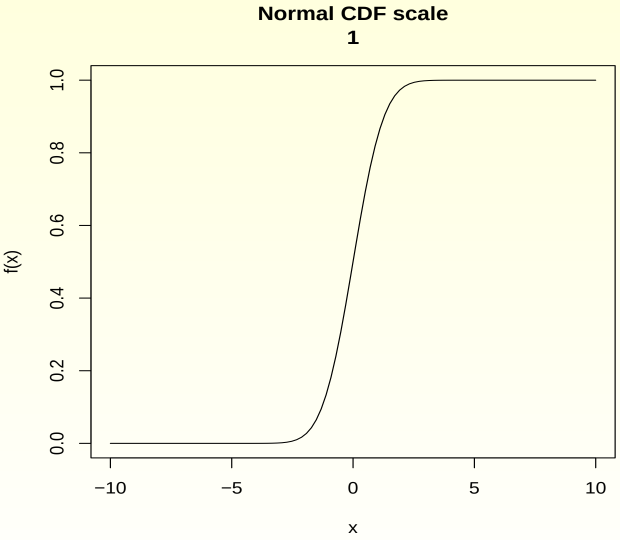
- La fonction de répartition joue un rôle important.
- Rappel: la fonction de répartition F est définie par:

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

- C'est l'aire en jaune sur la figure.



Fonction de répartition

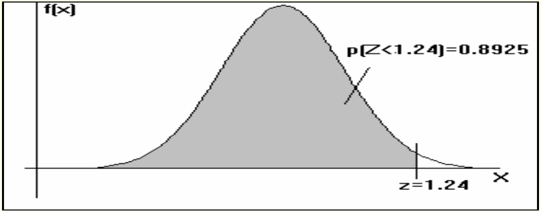


Calculs de probabilités sur une loi normale: Tableau de la loi normale

- **Un gros inconvénient**: on ne connaît pas de primitive de la fonction e^{-x^2} (pas l'expression algébrique de la fonction de répartition $F(x)$).
- Par des techniques de calcul numérique on a pu constituer des **tables** donnant $F(x)$ afin de déterminer les probabilités associées aux événements qui nous intéressent.
- Pour tous les calculs, on se ramène à la fonction de répartition de la loi $\mathcal{N}(0,1)$, dite loi normale centrée réduite.

TABLE DE LA LOI NORMALE CENTREE

Lecture de la table: Pour z=1.24 (intersection de la ligne 1.2 et de la
on a la proportion P(Z < 1,24) = 0.8925



P(Z > 1,9
P(Z > 2,5
P(Z > 3,2

Rappels:

1/ P(Z > z) = 1 - P(Z < z) et 2/ P(Z < -z) = P(Z > z)
Exemple: Sachant P(Z < 1,24) = 0,8925, on en déduit:
1/ (P(Z > 1,24) = 1 - P(Z < 1,24) = 1- 0,8925 = 0,1075
2/ P(Z < -1,24) = P(Z > 1,24) = 0,1075

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608

Calculs de probabilités sur une loi normale: Tableau de la loi normale

- La table donnant, pour différentes valeurs de $x > 0$, les valeurs de $F(x)$, soit $\mathbf{Pr}(X \leq x)$
- On y lit par exemple :
  - $F(0) = 0$
  - $F(1) = 0.8413$
  - $F(0,5) = 0,6915$
  - $F(1,96) = 0,9750$
- dont on déduit par symétrie :
  - $F(-1) = 1 - F(1) = 1 - 0,8413 = 0,1587$
  - $F(-0,5) = 1 - F(0,5) = 1 - 0,6915 = 0,3085$
  - $F(-1,96) = 1 - P(1,96) = 1 - 0,9750 = 0,0250$

Exemple de calculs de quantiles: utilisation du tableau

- Soit $X \rightsquigarrow \mathcal{N}(0,1)$.
- $\mathbf{Pr}(X \leq x) = ?$, x quantile donné, soit le quantile x, la probabilité étant donnée.
- Le tableau de la loi normale centrée réduite donne les valeurs de $F(x) = \mathbf{Pr}(X \leq x)$, $x \geq 0$.

Exemple de calculs de quantiles: utilisation du tableau

- **Que vaut $\Pr(X \leq 1.91)$?**
  - on décompose 1.91 en 1.9 et 0.01 : $1.91 = 1.9 + 0.01$
  - on se place à l'**intersection** de la **ligne** $u = 1.9$ **et** la **colonne** $u = 0.01$ et on **lit la probabilité** associée 0.9719.

Exemple de calculs de quantiles: utilisation du tableau

- **Que vaut $\Pr(X \leq -1.91)$?**
  - ◇ La table ne donne pas les **quantile négatifs**.
  - ◇ Du fait que la densité est **symétrique**, on peut écrire $\Pr(X \leq -1.91) = \Pr(X \geq 1.91)$.
  - ◇ La table donne les probabilités associées à l'événement $X \leq x$ et non $X \geq x$.
  - ◇ On a: $\Pr(X \geq 1.91) = 1 - \Pr(X < 1.91)$ et ainsi,

$$
\begin{aligned}
\Pr(X \leq -1.91) = \Pr(X \geq 1.91) &= 1 - \Pr(X < 1.91) \\
&= 1 - 0.9719 = 0.0281
\end{aligned}
$$

Utilisation du tableau

- **Que vaut $\mathbf{Pr}(0.5 < X \leq 1.01)$?**
  - On voit facilement que l'événement $(0.5 < X \leq 1.01)$ peut se décomposer en $(X \leq 1.01)$ et $(X \leq 0.5)$ et ainsi,

$$\begin{aligned}
\Pr(0.5 < X \leq 1.01) &= \Pr(X \leq 1.01) - \Pr(X \leq 0.5) \\
&= 0.8438 - 0.6915 = 0.1523
\end{aligned}$$

Utilisation du tableau

- **Que vaut x tel que $\Pr(X \leq x) = 0.8315$?**
  - Il faut chercher la valeur 0.8315 à l'**intérieur de la table** et lire le quantile associé en **additionnant** les valeurs de u associées à la **ligne et** à la **colonne** de 0.8315, c'est-à-dire $x = 0.9 + 0.06 = 0.96$.

Utilisation du tableau

- **Que vaut x tel que $\mathbf{Pr}(X \leq x) = 0.2358$?**
  - Cette valeur **ne se trouve pas à l'intérieur de la table**. La probabilité 0.2358 étant **inférieur** = 0.5, cela signifie que le quantile est nécessairement **négatif**.
  - On peut écrire $P(X \leq x) = 0.2358 = 1 - \Pr(X \leq -x)$ et $-x$ tel que $\Pr(X \leq -x) = 1 - 0.2358 = 0.7642$, c'est-à-dire $-x = 0.72$ et $x = -0.72$.

Loi normale quelconque

- Si μ est un nombre réel, σ un nombre positif, et X une variable de loi normale $\mathcal{N}(0, 1)$, la variable aléatoire

$$Z = \mu + \sigma X$$

obtenue à partir de X par changement d'origine et d'unité, a pour densité au point z la fonction f telle que

$$f(z) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{(z-\mu)^2}{2\sigma^2}) \quad z \in \mathbb{R}\,.$$

- On dit alors que Z suit la loi normale $\mathcal{N}(\mu, \sigma^2)$.

Espérance et Variance

- Nous avons déjà vu quel était l'effet d'un changement d'origine et d'unité sur l'espérance et la variance:
  - $\mathbb{E}(Z) = \mathbb{E}(\mu + \sigma X) = \mu$
  - $Var(Z) = \sigma^2$

Fonction de répartition

- La probabilité pour que Z soit inférieur ou égal à une valeur c est égale à

$$F(c) = \int_{-\infty}^{c} f(z)dz\,.$$

- La probabilité pour que Z appartienne à un intervalle $I =]a, b]$ vaut donc

$$\mathbf{Pr}(a < Z \leq b) = \int_{a}^{b} f(z)dz = F(b) - F(a)$$

Variable centrée réduite

Définition

La variable centrée réduite associée à la variable X d'espérance $\mathbb{E}(X)$ et de variance $\sigma^2(X)$ est:

$$X' = \frac{X - \mathbb{E}(X)}{\sigma(X)}\,.$$

- On vérifiera que $\mathbb{E}(X') = 0$ et $Var(X) = 1$.

Loi $\mathcal{N}(\mu, \sigma^2)$ et probabilités

- Soit $Z \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. On cherche : $\mathbf{Pr}(a \leq Z \leq b) =?$
  - ◇ Seule $\mathcal{N}(0, 1)$ est tabulée
  - ◇ Mais $X = \frac{Z-\mu}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$ (on va **centrer et réduire** pour obtenir la probabilité)
    - ▷ En effet: $\mathbf{Pr}(Z \leq z) = \mathbf{Pr}(\mu + \sigma X \leq z) = \mathbf{Pr}(X \leq \frac{z-\mu}{\sigma}) = F(\frac{z-\mu}{\sigma})$.
  - ◇ Il suffit de connaître la fonction de répartition de la loi normale réduite $\mathcal{N}(0, 1)$ pour pouvoir calculer toute probabilité relative à une variable normale quelconque $\mathcal{N}(\mu, \sigma^2)$.

Exemple de calculs de quantiles

- Soit $Z \rightsquigarrow \mathcal{N}(20, 36)$
- **Que vaut $\Pr(Z \leq 22.4)$?**
  - ◇ On peut écrire:

$$\Pr(Z \leq 22.4) = \Pr(X \leq \frac{22.4 - 20}{6})$$

    où X est une variable de loi $\mathcal{N}(0, 1)$.
  - ◇ Ainsi: $\Pr(Z \leq 22.4) = \Pr(X \leq 0.4) = 0.6554$.

Exemple de calculs de quantiles

- Soit une $v.a.r\ Z \rightsquigarrow \mathcal{N}(400, (15.4)^2)$.
- Trouver la valeur de z pour que $\mathbf{Pr}(Z > z) = 0.67$.
  - ◇ On a:

$$\begin{aligned} \mathbf{Pr}(Z > z) &= \mathbf{Pr}(\tfrac{Z-100}{15.4} > \tfrac{z-400}{15.4}) \\ &= \mathbf{Pr}(X > t) = 0.67 = F(0.44) \end{aligned}$$

    , $\Rightarrow |\frac{z-400}{15.4}| = |t| = 0.44$.
  - ◇ Mais $\mathbf{Pr}(X > t) = 0.67 > 0.5 \Rightarrow t < 0$, $\Rightarrow t = -0.44$,
    $\Rightarrow z = t * 15.4 + 400 = 393$.

Somme de variables normale indépendantes

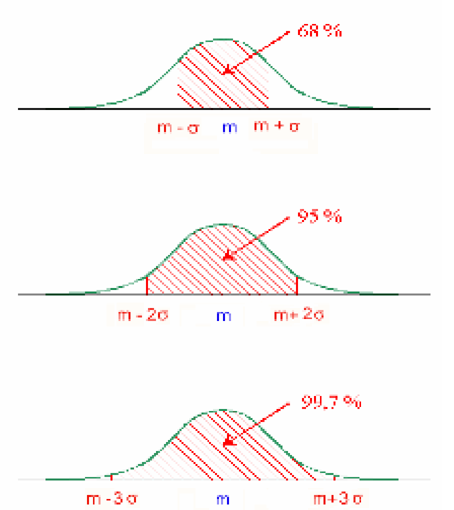
- Si X et Y sont deux variables aléatoires normales indépendantes avec $X \leadsto \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \leadsto \mathcal{N}(\mu_2, \sigma_2^2)$, alors $X + Y \leadsto \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
- On voit que ce résultat se généralise à la somme de n'importe quel nombre de variables normales indépendantes.
- si X et Y sont deux variables aléatoires normales indépendantes avec $X \leadsto \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \leadsto \mathcal{N}(\mu_2, \sigma_2^2)$, la variable $X - Y$ suit la loi $\leadsto \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

Variables normale indépendantes: exemple

La taille X des hommes en Algérie est modélisée par une loi normale $\mathcal{N}(172, 196)$ (unité : le cm).

1. Prop de l'algérien a une taille inférieure à 160cm?
2. Prop de l'algérien mesure plus de deux mètres?
3. Prop des algériens mesure entre 165 et 185cm?
4. Si on classait 10000 algériens choisis au hasard par ordre de taille croissante, quelle serait la taille du 9000-ième?
5. La taille Y des algériennes est modélisée par une loi normale $\mathcal{N}(162, 144)$. Quelle est la prob pour qu'un homme soit plus grand qu'une femme (choisie au hasard) ?

Quelques ordres de grandeur utiles à retenir



- Une variable normale a "95 chances sur 100 d'être située entre : moyenne moins 2 écarts-types et moyenne plus 2 écarts types " (la vraie valeur n'est pas 2 mais 1,96)
- Une variable normale est presque certainement située entre : moyenne moins 3 écarts-types et moyenne plus 3 écarts-types.

Les lois discrètes ne concernant que des variables dont les valeurs sont des nombres entiers.

Loi de Bernoulli

- On considère une expérience n'ayant que deux résultats possibles
  - exemple: succès et échec (ou présence et absence d'une certaine caractéristiques), mâle ou femelle, mort ou vivant, ect.
- On introduit la variable aléatoire X qui associe la valeur 0 à l'échec (intervention chirurgicale) et la valeurs 1 au succès. Cette variable aléatoire est appelée variable de Bernoulli.

Loi de Bernoulli, Espérance et Variance

- Appelons p la probabilité de l'événement succès:

$$\begin{aligned} \mathbf{Pr}(\{\text{succés}\}) &= \mathbf{Pr}(X = 1) = p \\ \mathbf{Pr}(\{\text{échec}\}) &= \mathbf{Pr}(X = 0) = 1 - p = q \end{aligned}$$

  avec $p + q = 1$
- La loi de X est:

X	0	1
$\mathbf{Pr}(X = x_i)$	q = 1-p	p

- $\mathbb{E}(X) = \mu_X = p$
- $V(X) = \sigma_X^2 = p \times q.$

Loi Binômiale: Définition

- Soient les épreuves répétés et indépendantes d'une même expérience de Bernoulli.
- Chaque expérience n'a que deux résultats possibles: succès ou échec.
- Comme précédemment, appelons p la probabilité de l'événement aléatoire succès.
- À cette expérience multiple on associe une **variable aléatoire** X qui mesure **le nombre de succès obtenus**.
- Nombreuses applications au calcul des fréquences d'allèles et de phénotypes en génétique.

Loi Binômiale: modélisation

- n = nombre d'essais (ou épreuves).
- p = probabilité d'un des deux événements. Probabilité de l'autre: $q = 1 - p$.
- x = **nombre de succès**; nombre de fois où l'événement apparaît au cours de n **épreuves indépendantes et identiques**. C'est la variable aléatoire.

Exemple

- urne contenant 2 catégories de boules: des boules blanches au proportion p et des boules non blanches en proportion $q = 1 - p$.
- Soit l'expérience aléatoire qui, consiste à tirer n boules **avec remise** de cette urne et soit la v.a. X: le nombre de boules blanches obtenus en cours de n tirages.
- $X(\Omega)$= l'ensemble des résultats possibles
  $= \{0, 1, \ldots, n\}$.

Exemple

- X_i: loi de Bernoulli (l'expérience $i^{\text{ième}}$ tirage), donc

$$X_i = \begin{cases} 1 & \text{avec probabilité } p \\ 0 & \text{avec probabilité } q = 1 - p \end{cases}$$

- $X = \sum_{i=1}^{n} X_i = X_1 + \ldots + X_n$, avec $X_i \rightsquigarrow$ Bernoulli(p).

Procédure

- **Question:** On peut se demander avec quelle probabilité on a k succès, c'est-à-dire: $\mathbf{Pr}(X = k) = ?$, $k \in \{0, 1, \ldots, n\}$.
- Rappels: $C_n^k = \frac{n!}{k!(n-k)!}$, est le nombre de manières d'obtenir k succès parmi n
- $p^k q^{n-k}$: est la probabilité d'en obtenir une.
- D'où la probabilité d'avoir exactement k succès est

$$p_k = \mathbf{Pr}(X = k) = C_n^k p^k q^{n-k} , k = 0, 1, \ldots, n.$$

- On vérifie que:

$$\sum_{i=1}^{n} \mathbf{Pr}(X = k) = \sum_{i=1}^{n} C_n^k p^k q^{n-k} = (p + q)^n = 1.$$

Définition

On dit qu'une $v.a.r.$ suit une loi Binômiale de paramètres n et p (construite sur n expériences de Bernoulli **indépendantes**) si l'on a:

- $X(\Omega) = \{0, 1, \ldots, n\}$, et
- $\forall k \in \{0, 1, \ldots, n\}$, on a

$$\mathbf{Pr}(X = k) = C_n^k p^k q^{n-k}, \quad q = 1 - p.$$

- On note $X \rightsquigarrow \mathcal{B}(n, p)$.

Espérance et Variance

- Si $X \rightsquigarrow \mathcal{B}(n,p)$, alors:
  - Espérance:

    $$\boxed{\mathbb{E}(X) = \mu_X = n \times p}$$

  - Variance:

    $$\boxed{V(X) = \sigma_X^2 = n \times p \times q}$$

Exemple

- On jette 6 fois une pièce de monnaie bien équilibrée, on suppose que face est un succès. On a donc: $p = \frac{1}{2}$, $n = 6$.
  a) $\Pr(\text{on ait exactement 2 faces}) = \Pr(X = 2) = C_6^2(\frac{1}{2})^2(\frac{1}{2})^4 = \frac{15}{64}$.
  b)
$$\begin{aligned}
&\Pr(\text{d'avoir au moins 4 faces}) = \\
&\quad = \Pr(X = 4) + \Pr(X = 5) + \Pr(X = 6) \\
&\quad = C_6^4(\tfrac{1}{2})^4(\tfrac{1}{2})^2 + C_6^5(\tfrac{1}{2})^5(\tfrac{1}{2})^1 + C_6^6(\tfrac{1}{2})^6 \quad . \\
&\quad = \tfrac{15}{64} + \tfrac{6}{64} + \tfrac{1}{64} = \tfrac{11}{32}
\end{aligned}$$

Exemple

- Un examen consiste à répondre à 4 questions à coté des quelles on trouve cinq réponses possibles: parmi les réponses, une seule est correcte. Pour réussir l'examen, il faut répondre correctement à trois questions au moins. Quelle est la probabilité pour qu'un étudiant n'ayant pas préparé son examen réussisse?

Exemple

- Si l’étudiant n’a pas préparé son examen, il devra choisir les réponses "au hasard", c’est-à-dire que toutes les réponses ont la même probabilité d’être choisir, en particulier la réponse correcte, il s’en suit que $p = \frac{1}{5}$ pour cet étudiant là.
- Pour chaque étudiant l’examen peut être assimilé à loi de Bernoulli où le succès est la réponse juste.
  $X \rightsquigarrow \mathcal{B}(n, p) = \mathcal{B}(4, \frac{1}{5})$. $(X$: nombre des réponses juste$)$.
- L’événement "réussir l’examen" peut être décrit par: "$(X = 3)$ ou $(X = 4)$"="$(X \leq 3)$", et $\mathbf{Pr}(X \geq 3) = 0.0272$.

Exemple: Application biologique

- Quelle est la probabilité de trouver $x = 0, 1, 2$ ou 3 filles dans des familles de $n = 3$ enfants

Loi de Poisson

- Loi concernant la réalisation d'événements
  - **faiblement probables** (Lois des événements rares)
  - **Indépendants**.
- La variable X est "**nombre d'événement observé (de réalisations) pendant une période de temps donnée**" de ces événements.

Loi de Poisson: Application

- Elle peut s'appliquer au: nombre d'accidents, à l'apparition d'anomalies divers (maladies exceptionnelles), à la gestion des files d'attentes, ou nombre de colonies bactériennes dans une boîte de Pétri, le contrôle de fabrication à très faibles proportion de robuste dans l'industrie pharmaceutique, etc ...

Définition

Définition

- Soit X la variable aléatoire représentant le nombre d'apparitions indépendantes d'un événement faiblement probable dans une population infinie (**varie entre 0 et** ∞).
- La probabilité d'avoir k apparitions de l'événement rare est: $\mathbf{Pr}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}\,, \quad \lambda \in \mathbb{R}^+\,,\, k \in \mathbb{N}$
- notée: $X \rightsquigarrow \mathcal{P}(\lambda)$, dépend d'un paramètre λ, nombre réel strictement positif.

Remarque

1. $X(\Omega) = \mathbb{N}\,,$ et $\displaystyle\sum_{k=0}^{\infty} \mathbf{Pr}(X = k) = 1\,.$
2. Si deux variables aléatoires indépendants X_1 et X_2 sont distribuées selon des lois de Poisson de paramètres λ_1 et λ_2, alors la variable $X_1 + X_2$ est distribuée selon une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.

Espérance et Variance

Si $X \rightsquigarrow \mathcal{P}(\lambda)$, alors

1. L'espérance de X est donnée par:

   $\boxed{\mu_X = \mathbb{E}(X) = \lambda \text{ (moyenne)}}$

2. La variance de X est donnée par:

   $\boxed{\sigma_X^2 = V(X) = \lambda.}$

Remarque

- Si on connaît la probabilité de n'observé aucune événement, c'est-à-dire $\mathbf{Pr}(X=0)=p$, alors, d'après la formule du loi on a: $p=\mathbf{Pr}(X=0)=e^{-\lambda}\frac{\lambda^0}{0!}=e^{-\lambda}$.
- On en déduit: $\lambda=-\ln(p)$ et:

$$\begin{array}{rcccl} \mathbf{Pr}(X=1) & = & e^{-\lambda}\frac{\lambda^1}{1!} & = & p\lambda \\ \mathbf{Pr}(X=2) & = & e^{-\lambda}\frac{\lambda^2}{2!} & = & \mathbf{Pr}(X=1)\frac{\lambda}{2} \\ \vdots & & \vdots & & \vdots \\ \mathbf{Pr}(X=k) & = & e^{-\lambda}\frac{\lambda^k}{k!} & = & \mathbf{Pr}(X=k-1)\frac{\lambda}{k} \end{array}$$

- On peut ainsi calculer facilement de proche en proche les probabilité des diverses valeurs de k.

Loi de Poiusson: Exemple

- Si on sait que dans une région, il y a en moyenne 10 cas d'une certain maladie rare pendant une période donnée, alors le nombre de cas observées suit une loi de Poisson de paramètre 10.

Table des lois de Poisson

- Á l'intersection de la colonne λ et de ligne k, figure la probabilité pour qu'une variable de Poisson X de paramètre λ soit égal à la valeur entière k

$$\mathbf{Pr}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Approximation de la loi Binômiale par une loi de Poisson

- Si $X \rightsquigarrow \mathcal{B}(n,p)$,
- Si p est petit (**en pratique** $p < 0.1$**)**, n **assez grand** (supérieur à 50), et $0 \leq n \times p \leq 10$
- Alors, la loi Binômiale peut être approximée par Y, une loi de Poisson de paramètre $\lambda = n \times p$
  - $\Pr([X = k]) = \Pr([Y = k])$
  - La probabilité $\Pr([Y > n])$ est faible, mais non nulle.
- Les calcule sont plus simple avec la loi de Poisson qu'avec la loi Binômiale.

Lien avec la loi Binômiale

- Par exemple, sur un échantillon de grand taille, on étudie le nombre de fois que se produit un événement "rare", c'est-à-dire approximation de la loi Binômiale $\mathcal{B}(n, p)$ lorsque n est grand et p faible (ou $1 - p$ faible) de façons que $\lambda = n * p$ soit l'ordre de l'unité.
- Cela se produit par exemple lorsqu'on considère le nombre d'accident provoquée par vaccin, le nombre de suicide dans une grande ville, pour une période donnée.

Lien avec la loi Binômiale: remarque

1. Notons que puisque X est distribuée selon une loi Binômiale, ses valeurs ne peuvent dépasser n, alors que l'approximation par la loi de Poisson autorise des valeurs supérieurs. Cependant le calcule fournit des probabilité très faibles pour ces valeurs aberrantes.
2. Lorsque $5 \leq n \times p \leq 10$, on doit aux deux approximation (par la loi de Poisson et la loi normale), mais bien sûr, celle de Poisson est d'autant meilleur, et donc préférable, que q est plus proche de 0.

Approximation: Exemple

- La probabilité d'observer une mutation sur un individu étant de 10^{-4}, combien d'individus faut-il s'attendre à examiner pour être pratiquement "sûr" d'observer au moins un mutant ?
- Soit n ce nombre et soit X la $v.a.$ qui associe à ces n individus le nombre de mutant que l'on peut observer parmi eux. Donc, $X \rightsquigarrow \mathcal{B}(n, 10^{-4})$. Comme 10^{-4} est une valeur très faible et que n a certainement une valeur élevée, on peut dire que X suit pratiquement une loi de Poisson de $\mathcal{P}(\lambda = n \times 10^{-4})$.

Exemple

- Si on considère comme pratiquement sûr un événement de probabilité au moins égale à 0.95, on écrit:

$$\mathbf{Pr}(x \geq 1) \geq 0.95 \Leftrightarrow \mathbf{Pr}(x < 1) < 0.05 \,.$$

- Or

$$\begin{aligned}
\mathbf{Pr}(x < 1) = \mathbf{Pr}(x = 0) = e^{-\lambda} < 0.05 \quad &\Leftrightarrow \quad e^{-n*10^{-4}} < 0.05 \\
&\Leftrightarrow \quad n \geq \frac{-\ln 0.05}{10^{-4}} \\
&\Leftrightarrow \quad n \geq 29958
\end{aligned}$$

Approximation normale des lois Binômiales

- Nous utilisons la règle suivante:
  - si $X \rightsquigarrow \mathcal{B}(n,p)$, sa moyenne est $m = n \times p$ et sa variance est $\sigma^2 = n \times p \times q$,
  - si $n > 30$, $n \times p \geq 5$ et $n \times q \geq 5$,
  - on pourra considérer la $v.a.$ $\frac{X-m}{\sigma} \rightsquigarrow \mathcal{N}(0,1)$.

$$\mathrm{Pr}\left(\frac{X-m}{\sigma} < x\right) \approx F(x)\,.$$

  - On a $\mathrm{Pr}([X=k]) \approx \mathrm{Pr}([k-0,5 \leq Y \leq k+0,5])$
  - Les probabilités $\mathrm{Pr}([Y<0])$ et $\mathrm{Pr}([Y>n])$ sont faibles, mais non nulles

Approximation normale des lois de Poisson

- Si $X \rightsquigarrow \mathcal{P}(\lambda)$
- λ grand $(\lambda \geq 15)$,
- on peut utiliser l'approximation normale suivante:

$$X \rightsquigarrow \mathcal{N}(\lambda, \lambda)\,,$$

$\lambda =$ moyenne , $\sigma = \sqrt{\lambda}$.

- On a $\mathbf{Pr}([X = k]) \approx \mathbf{Pr}([k - 0,5 \leq Y \leq k + 0,5])$

Exercice 1

Dans une famille la probabilité d'avoir un enfant gaucher est de 0.25.

a) Quelle est la loi de prob suivie par la variable X le nombres d'enfants gauchers dans une famille de n enfants? Nature de la loi? Formule donnant la probabilité de X.

b) Quelle est la prob d'avoir exactement 2 enfants gauchers dans une famille de 9 enfants?

c) Quelle est la probabilité d'avoir exactement 6 enfants droitiers dans cette famille?

d) Quelle est la probabilité d'avoir au plus 6 enfants droitiers dans cette famille?

Exercice 2

Dans une population humaine, la proportion des gauchers est de 1%. Quelle est la probabilité d'observer au moins 4 gauchers dans un échantillons de 230 personnes?

Exercice 3

Dans une entreprise, il se produit en moyenne 4 accidents de travail par moins. Soit X la variable aléatoire qui représente le nombre d'accidents de travail dans cette entreprise.

1. Donner les valeurs de X.
2. Donner la loi de X.
3. Calculer la probabilité de $x \geq 1$.

Exercice 4

Dans une ville mexicaine, on estime que le nombre moyen de personnes qui attrape quotidiennement le virus H1N1 dû à la grippe porcine est de l'ordre de 20 personne. On note X la variable aléatoire comptabilisant le nombre de personnes ayant été affecté quotidiennement par le virus. On suppose qu'une journée correspond l'unité du temps.

1. Quelle est la loi de probabilité de X ?
2. Peut-on faire une approximation de la loi de X par la loi normale?
3. En utilisant la table statistique de la loi normale, calculer la probabilité pour que la nombre de touristes ayant attrapé le virus le jour de leur arrivée, soit compris entre 5 et 20.

Exercice 5

On suppose que l'âge auquel apparaissent les premiers mots de vocabulaire chez l'enfant suit la loi normale de moyenne 12 mois et d'écart-type 2.5 mois.

1. Quelle est la proportion d'enfants pour lesquels les premiers mots apparaissent avant 9 mois ?
   $(A)\ < 0.01 \qquad (B)\ = 0.11507 \qquad (C)\ = 0.13786 \qquad (D)\ = 0.86214 \qquad (E)\ > 0.88.$
2. Déterminer l'âge au-dessus duquel 2% des enfants prononcent leurs premiers mots.
   $(A)\ < 6 \qquad (B)\ = 6.875 \qquad (C)\ = 17.125 \qquad (D)\ = 17.15 \qquad (E)\ > 18.$
3. Le nouveau né mohamed à 9.5 mois et il n'a pas encore dit son premier mot. Quelle est la probabilité qu'il prononce son premier mot après 9.5 mois ?
   $(A)\ < 0.1 \qquad (B)\ = 0.15866 \qquad (C)\ = 0.84134 \qquad (D)\ = 0.94295 \qquad (E)\ = 1.$

Exercice 5, la suite

4. Le nouveau né Fatima à 9.5 mois et elle n'a pas encore prononcer son premier mot. Quelle est la probabilité qu'elle prononce son premier mot entre 9.5 et 10 mois ? $(A) = 0.06323 \quad (B) = 0.44039 \quad (C) = 0.0532 \quad (D) = 0.33531 \quad (E) > 0.6.$
5. La médiane de la variable "Age" est:
$(A) = 0 \quad (B) = 12 \quad (C) = 13.25 \quad (D) = 14 \quad (E) = 15.$

Bibliographie

- Cours Probabilité, Benchikh Tawfik, http://www.univ-sba.dz/lsps/images/pdf/cours/
- Probabilités et Biostatistique. www.chups.jussieu.fr/polys/biostats/CoursProba1.pdf
- FMPMC Pitié-Salpêtrière-probabilités, cours 1 et 2. www.chups.jussieu.fr/polys/biostats/coursProba-1-2/
- Notions de Probabilités, Roch Giorgi, LERTIM, Faculté de Médecine, Université de la Méditerranée, Marseille, France http://lertim.fr