Distributions univariées

Mathématiques Générales B
Université de Genève

Sylvain Sardy

17 avril 2008

1. Variables aléatoires

Définition de v.a. réelle : une fonction définie sur l'ensemble des résultats possibles, appelé ensemble fondamental ou univers Ω, d'une expérience aléatoire

$$X : \Omega \longrightarrow X(\Omega) \subset \mathbb{R}.$$

Exemples :

– On jette deux dés, et on considère $X =$ la somme des deux chiffres.
– On joue à pile ou face 10 fois, et on considère $X =$ le nombre de Pile.
– On joue au loto, et on considère $X =$ le nombre de gagants.
– On jette un dé jusqu'à ce qu'on gagne (le 6 sort), et on considère $X =$ nombre de fois où le dé est jeté.

- On joue au loto, et on considère $X =$ la somme gagnée par un joueur.
- On mesure sur le Pont du Mont Blanc la vitesse X des véhicules.

Deux types de variable aléatoire :

- **Discrète** quand $X(\Omega) = \{x_1, x_2, \dots\}$ est fini ou dénombrable.
- **Continue** quand $X(\Omega)$ est un intervalle de $\mathbb{R}$.

Deux types de distribution :

- **Discrète** définie par une fonction de probabilité/fréquence : $\mathrm{P}(\mathrm{X} = \mathrm{x_i})$.
- **Continue** définie par une fonction de densité $f(x)$ telle que

$$\mathrm{P}(a \leqslant X \leqslant b) = \int_a^b f(x) dx.$$

Un trés grand nombre de distribution a été défini et étudié, et sert à modéliser des phénomènes aléatoires :

– Discrète : **Uniforme discrète** (equiprobabilité), **Bernouilli**, **Binomiale**, **Géométrique**, **Hyper-géométrique**, **Poisson**, Multinomiale, Binomiale-Négative, etc.
– Continue : **Uniforme**, **Gaussienne/Normale**, **Chi-deux**, **Exponentielle**, Student, Cauchy, Log-Normale, Beta, Gamma, Weibull, F, Pareto, Rayleigh, Rice, Fisher-Tippett, etc.

On va en définir quelques-unes, puis on étudiera dans un prochain chapitre quelques propriétés (espérance et variance).

2. Distributions discrètes

Distribution uniforme discrète

Une expérience a n réalisations équiprobables $x_1, x_2, \ldots, x_n$, où n est finie. Alors

$$\mathrm{P}(\mathrm{X} = \mathrm{x_i}) = \frac{1}{\mathrm{n}} \quad \text{pour tout i.}$$

Exemple : un dé.

1) A quoi ressemble le graphique de la fonction de probabilité uniforme discrète ?

2) Vérifie-t-on $\sum_i \mathrm{P}(\mathrm{X} = \mathrm{x_i}) = 1$?

Hypergéométrique(N,D,n)

Une population finie de N objets en contient D défectueux. On en envoie n et on note K le nombre d'objets défectueux dans l'échantillon envoyé. Puisqu'il s'agit d'un échantillon sans remise, on sait pas l'analyse combinatoire que

$$\mathrm{P(K = k) = \frac{C_{D,k} C_{N-D,n-k}}{C_{N,n}}}.$$

Exemple : tableau de contingence

	Envoyé	Non-envoyé	Total
Défectueux	k	D-k	D
Non-défectueux	n-k	N-n-D+k	N-D
Total	n	N-n	N

Bernoulli(p) (James Bernoulli 1654–1705)

Une expérience a $n = 2$ réalisations possibles : soit 0 (échec/défectueux), soit 1 (succès), et

$$\mathrm{P}(\mathrm{X} = 1) = 1 - \mathrm{P}(\mathrm{X} = 0) = \mathrm{p}.$$

Exemple : une pièce de monnaie, toute expérience aléatoire avec un succès (1) et un échec (0).

1) 2)

Géométrique(p)

On répète une expérience de Bernoulli jusqu'à ce qu'un succès arrive. Soit N le nombre d'essais nécessaires, alors

$$\mathrm{P}(\mathrm{N} = \mathrm{n}) = (1 - \mathrm{p})^{\mathrm{n}-1}\mathrm{p}, \quad \mathrm{n} = 1, 2, \ldots$$

1) 2)

Binomiale(n,p)

On réalise une expérience de Bernoulli n fois de façon indépendante. On s'intéresse au nombre S de succès

$$\mathrm{P}(\mathrm{S}=\mathrm{k}) = \left(\begin{array}{c} n \\ k \end{array} \right) \mathrm{p}^{\mathrm{k}}(1-\mathrm{p})^{\mathrm{n}-\mathrm{k}} \quad \text{pour} \quad \mathrm{k} \in \{0,1,2,\ldots,\mathrm{n}\}.$$

(Démonstration dans le cours précédent)

Le support de la (distribution) Binomiale est fini : $\{0,1,2,\ldots,n\}$.

1) Illustration de la Binomiale

2)

Poisson(λ) (Siméon-Denis Poisson 1781 – 1840)

Exemples de phénomènes modélisables avec une distribution de Poisson :

- le nombre de voiture passant à un certain point d'une autoroute pendant une unité de temps.
- le nombre d'erreur d'orthographe par page
- le nombre d'appel téléphonique à une centrale d'appels.
- le nombre de fois qu'un server internet est visité par minute.
- le nombre de mutations d'une certaine séquence d'ADN après une certaine dose de radiation.
- le nombre de particules émises par un objet radioactif pendant une minute.
- le nombre de pins dans une forêt feuillus/épineux.
- le nombre de virus pouvant infecter une cellule.
- le nombre de rayon gamma sur un capteur pendant un interval de temps

Modélise la probabilité qu'un nombre X d'évènements aléatoires arrivent pendant une période de temps (ou d'espace) t d'un processus de Poisson :

- les évènements arrivent avec un taux constant λ,
- les évènements sont indépendants de la date de l'évènement qui vient de se produire.

$$\mathrm{P(X = k) = \exp(-\lambda t)\frac{(\lambda t)^k}{k!}}.$$

Contrairement à la Binomiale, le support de la Poisson n'est pas fini $\{0, 1, 2, \ldots\}$.

1) Illustration du processus de Poisson et de la distribution Poisson

2) (Série de Taylor : $\exp(x) = \sum_{i=0}^{\infty} \frac{x^i}{i!}$)

Approximation de la distribution Binomiale par la distribution de Poisson

La loi des évènements rares : Soit $S_n \sim \text{Binomiale}(n, p_n)$. Si $p_n \to 0$ et $\lambda_n = np_n \to \lambda$ quand $n \to \infty$, alors

$$\mathrm{P}(\mathrm{S_n} = \mathrm{k}) \to \exp(-\lambda)\frac{\lambda^{\mathrm{k}}}{\mathrm{k}!}.$$

En d'autres termes, si un grand nombre d'évènements indépendants a une faible probabilité individuelle de succès, alors le nombre total de succès a approximativement une distribution de Poisson.

Démonstration :

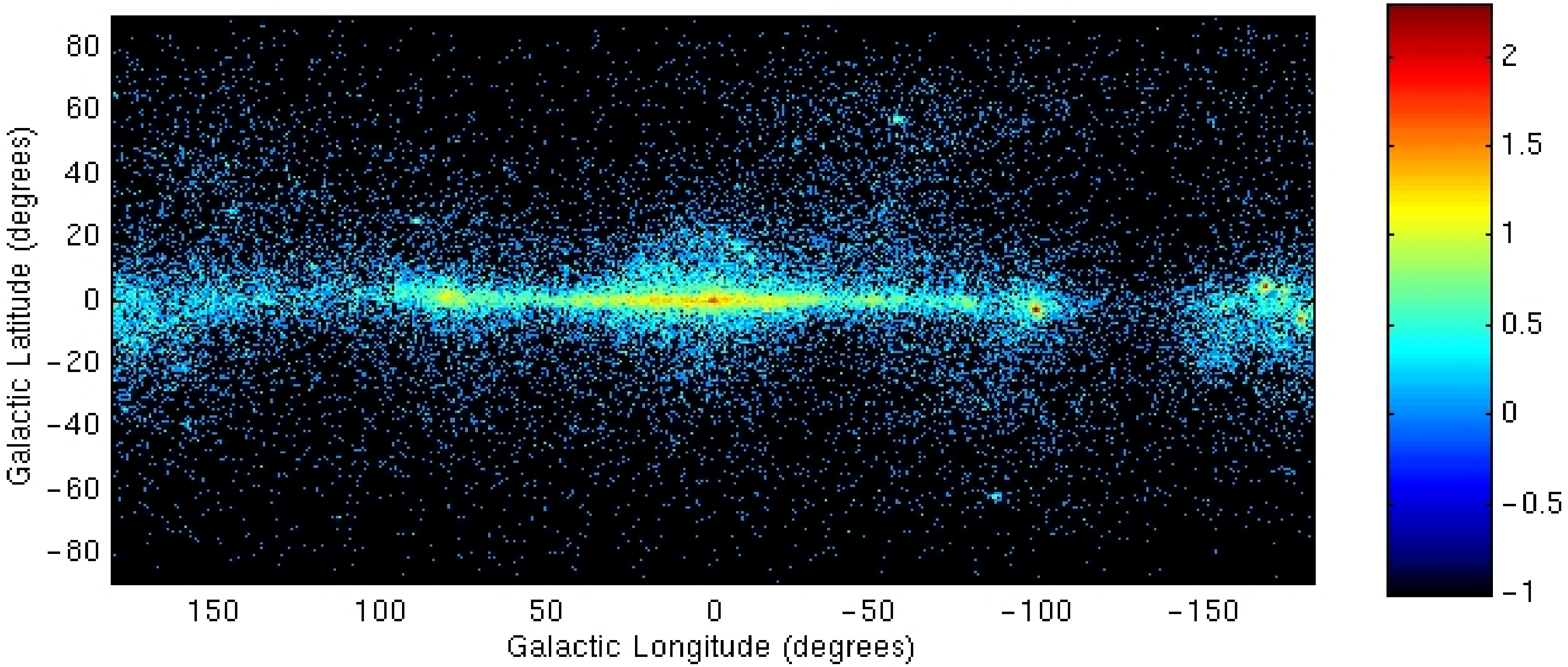
$$\mathrm{P}(\mathrm{S_n} = 0) = (1 - \mathrm{p_n})^{\mathrm{n}} = (1 - \frac{\lambda_{\mathrm{n}}}{\mathrm{n}})^{\mathrm{n}} \to \exp(-\lambda)$$

car $\lambda_n \to \lambda$ quand $n \to \infty$.

$$\begin{aligned}
\mathrm{P}(\mathrm{S_n} = \mathrm{k}) &= C_{n,k}(\frac{\lambda_n}{n})^k(1 - \frac{\lambda_n}{n})^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{\lambda_n^k}{k!}\left(1 - \frac{\lambda_n}{n}\right)^n\left(1 - \frac{\lambda_n}{n}\right)^{-k} \\
&\to \qquad\qquad \square
\end{aligned}$$

Illustration de la convergence

Application : Chaque pixel de cette image reflète un comptage de rayon gamma pendant une unité de temps pour une région de l'univers.



Application : On jette deux dés 12 fois et s'intéresse à $D =$ le nombre de fois où un double Six apparait.

On connait la distribution exacte de D n'est-ce pas ?

• Que vaut $\mathrm{P(D = k)}$ quand $k = 0$?

L'approximation par la distribution de Poisson donne :

$$\mathrm{P(D = 0)} \approx \exp(-1/3) = 0.7165.$$

• Que vaut $\mathrm{P(D = k)}$ quand $k = 2$? L'approximation par la distribution de Poisson donne :

$$\mathrm{P(D = 2)} \approx \exp(-1/3)(\frac{1}{3})^2 \frac{1}{2!} = 0.0398$$

Application : Les compagnies aériennes pratiquent l'overbooking. Supposons qu'elles vendent 200 tickets pour 198 places et que la probabilité qu'un passager ne se présente pas est de 1%. En utilisant l'approximation de Poisson, quelle est la probabilité qu'il y a une place pour toutes les personnes qui se sont présentées.

Soit S le nombre total de passagers qui ne se présentent pas. On s'intéresse à $\mathrm{P}(\mathrm{S} \geqslant 2) = 1 - (\mathrm{P}(\mathrm{S} = 0) + \mathrm{P}(\mathrm{S} = 1))$. Or $S \sim \mathrm{Bin}(n, p)$, où $n = 200$ est grand et $p = 0.01$ est petit. On peut donc utiliser l'approximation Poisson(λ) avec $\lambda = np = 2$ et

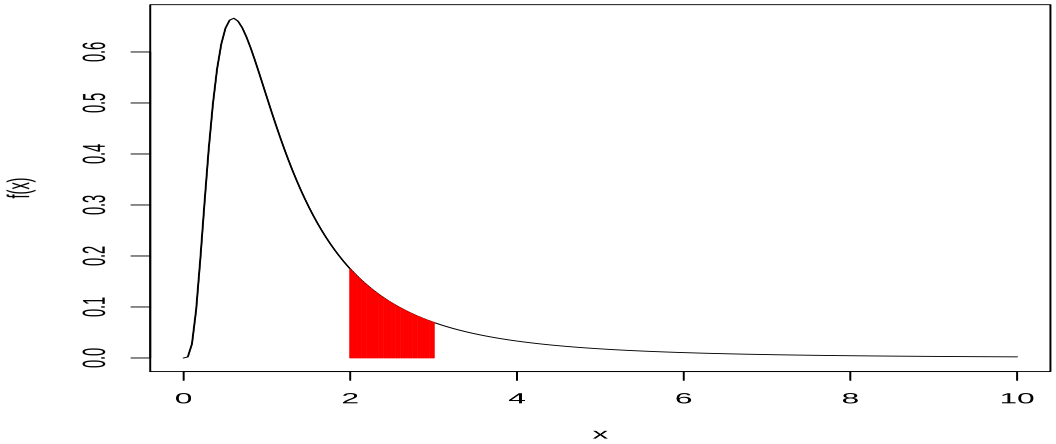
$$\begin{aligned} \mathrm{P}(\mathrm{S} \geqslant 2) &\approx 1 - \exp(-2) - \exp(-2)2^1/1! \\ &= 1 - 3\exp(-2). \end{aligned}$$

3. Distributions continues

Une v.a. X a une distribution continue avec **fonction de densité** $f(x)$ si pour tout a et b :

$$\mathrm{P}(\mathrm{a} \leqslant \mathrm{X} \leqslant \mathrm{b}) = \int_{\mathrm{a}}^{\mathrm{b}} \mathrm{f}(\mathrm{x})\mathrm{dx}.$$

Géométriquement, $\mathrm{P}(\mathrm{a} \leqslant \mathrm{X} \leqslant \mathrm{b})$ est l'aire sous la courbe f entre a et b.



La fonction de densité $f(x)$ en x joue le rôle de $\mathrm{P(X = x)}$ même si $\mathrm{P(X = x) = 0}$ pour une distribution continue. En effet :

$$\mathrm{P(x \leqslant X \leqslant X + \Delta x) = \int_{x}^{x+\Delta x} f(y)dy \approx f(x)\Delta x}$$

quand Δx est petit. Donc quand $\Delta x \to 0$, $\mathrm{P(x \leqslant X \leqslant X{+}\Delta x) \to P(X = x) = 0}$.

Par contre $f(x)$ renseigne sur la vraisemblance de trouver X proche de x.

Définition Toute fonction f telle que

1. $f(x) \geqslant 0$ pour tout x
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

est une fonction de densité.

Définition Toute variable aléatoire (discrète, continue ou mixte) a une fonction de répartition

$$F(x) = \mathrm{P(X \leqslant x)}.$$

Si X a une fonction de densité f, alors

$$F(x) = \mathrm{P(-\infty < X \leqslant x)} = \int_{-\infty}^{\mathrm{x}} \mathrm{f(y)dy}.$$

Donc F est la primive de F telle que $F(+\infty) = 1$.

Pac conséquent $F'(x) = f(x)$. Donc F est une fonction non-décroissante.

Utilité de la fonction de répartition F

De nombreuses tables et softwares donnent $F(x)$ pour des valeurs de x :

Table

Caclulateur

Dans `R` : ppois, punif, pnorm, pchisq, ...

De plus $\{X \leqslant b\} = \{X \leqslant a\} \bigcup \{a < X \leqslant b\}$, et les deux ensembles sont disjoints donc

$$\mathrm{P}(\mathrm{X} \leqslant \mathrm{b}) = \mathrm{P}(\mathrm{X} \leqslant \mathrm{a}) + \mathrm{P}(\mathrm{a} < \mathrm{X} \leqslant \mathrm{b}).$$

Connaitre F suffit pour calculer

$$\mathrm{P}(\mathrm{a} < \mathrm{X} \leqslant \mathrm{b}) = \mathrm{F}(\mathrm{b}) - \mathrm{F}(\mathrm{a}).$$

Uniform $\mathrm{U}(a,b)$ Une distribution dont la densité est constante sur un intervalle borné (a,b) :

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{sinon} \end{cases} .$$

1) Graphe de la densité ?

2) La constante $c = \frac{1}{b-a}$ satisfait la condition d'une densité que $\int_a^b c \, dx = 1$.

3) Fonction de répartition F et son graphe.

Exponential $\mathrm{Exp}(\lambda)$

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & x \geqslant 0 \\ 0 & \text{sinon} \end{cases}.$$

Cette distribution est souvent utilisée pour modéliser le temps d'attente entre des évènements, par exemple le temps entre l'arrivée de deux clients successifs dans une banque.

1)

2)

3)

La disribution Gaussienne/Normale standard : $\mathrm{N}(0,1)$

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

1) 2) Soit $I = \int \exp(-\frac{1}{2}x^2)dx$. Donc

$$I^2 = \int \exp(-\frac{1}{2}x^2)dx \int \exp(-\frac{1}{2}y^2)dy = \int\int \exp(-\frac{1}{2}(x^2+y^2))\ dx\ dy.$$

En changeant de variables et en passant en coordonnées polaires

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases},$$

qui est une bijection de $(x,y) \in (-\infty,\infty)\times(-\infty,\infty)$ dans $(r,\theta) \in (0,\infty)\times[0,2\pi]$

alors

$$\left| \begin{array}{cc} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{array} \right| = r,$$

donc

$$I^2 = \int_0^{\infty} \int_0^{2\pi} \exp(-\frac{1}{2}r^2) r \; d\theta \; dr. \qquad \square$$

3)

4. Fonctions d'une variable aléatoire

Motivation : Supposons que $X \sim \mathrm{N}(0,1)$.

- Quelle est la distribution de $Y = \mu + \sigma X$, où $\sigma > 0$?

$$\mathrm{P}(\mathrm{Y} \leqslant \mathrm{y}) = \mathrm{P}(\mu + \sigma \mathrm{X} \leqslant \mathrm{y})) = \mathrm{P}(\mathrm{X} \leqslant (\mathrm{y} - \mu)/\sigma) = ?$$

On appelle cette distribution la $\mathrm{N}(\mu, \sigma^2)$.

- Quelle est la distribution de $Y = X^2$?

$$\mathrm{P}(\mathrm{Y} \leqslant \mathrm{y}) = \mathrm{P}(\mathrm{X}^2 \leqslant \mathrm{y}) = \mathrm{P}(|\mathrm{X}| \leqslant \sqrt{\mathrm{y}}) = \mathrm{P}(-\sqrt{\mathrm{y}} \leqslant \mathrm{X} \leqslant \sqrt{\mathrm{y}}) = ?,$$

d'où on dérive la fonction de densité. On appelle cette distribution la χ_1^2 (la distribution chi-deux à 1 degré de liberté).

Théorème : Supposons que X a une fonction de densité f définie sur (a, b), que $r : (a, b) \longrightarrow (\alpha, \beta)$ est continue et strictement croisssante, et que $s : (\alpha, \beta) \longrightarrow (a, b)$ est la fonction inverse de r ($s = r^{-1}$). Alors $Y = r(X)$ a pour densité

$$g(y) = f(s(y))\ s'(y) \quad y \in (\alpha, \beta).$$

Ce théorème s'applique-t-il aux deux exemples précédents ?

Suppose que X a une fonction de répartition F strictement croissante. Quelle est la distribution de $Y = F(X)$?

Inversement si $U \sim \mathrm{U}(0, 1)$, quelle est la distribution de $Y = F^{-1}(U)$?