# Rappels de probabilités/statistiques

**G. Perrin**

guillaume.perrin@univ-eiffel.fr

Année 2022-2023






# Plan de la séance

# Plan de la séance

1. Probabilités
2. Rappels de génération aléatoire de v.a.

# Espaces de probabilités

- **But** : Modéliser une expérience aléatoire.
  Exemple 1 : lancer une pièce.
  Exemple 2 : lancer un dé.
  Exemple 3 : choisir un nombre au hasard entre 0 et 1 (fonction rand).
  $\hookrightarrow$ **Espace de probabilité** $(\Omega, \mathcal{A}, \mathbb{P})$.
- $\Omega$ : ensemble fondamental (ensemble de toutes les **réalisations possibles**, notées $\omega \in \Omega$).
  Exemple 1 : $\Omega = \{\text{Pile, Face}\}$.
  Exemple 2 : $\Omega = \{1, \ldots, 6\}$.
  Exemple 3 : $\Omega = [0, 1]$.
- $\mathcal{A}$ : ensemble des **événements possibles** $A \in \mathcal{A}$.
  Exemple 1 : $A =$ "Les 10 premiers résultats sont identiques".
  Exemple 2 : $A =$ "le résultat est paire".
  Exemple 3 : $A =$ "le résultat est plus grand que 1/3.

# Espaces de probabilités

- $\mathbb{P}$ : **mesure de probabilité**, associant à chaque évènement $A \in \mathcal{A}$ une valeur entre 0 et 1, telle que :
  - $\mathbb{P}(\Omega) = 1$,
  - $A_i$ et $A_j$ sont indépendants ssi $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \times \mathbb{P}(A_j)$,
  - si $\{A_j\}_{j \geq 1}$ est une famille au plus dénombrable d'ensembles disjoints de $\mathcal{A}$, alors $\mathbb{P}\left(\cup \{A_j\}_{j\geq 1}\right) = \sum_{j \geq 1} \mathbb{P}(A_j)$.

  Si l'on répète $N$ fois l'expérience aléatoire, et que l'on note $N(A)$ le nombre de fois où l'évènement $A$ s'est produit, la limite de $N(A)/N$ quand $N \to +\infty$ nous donne la **vision intuitive** de la probabilité $\mathbb{P}(A)$ de l'évènement $A$.
  Exemple 1 : $\mathbb{P}$ ("Les 10 premiers résultats sont identiques") $= 0.5^9$.
  Exemple 2 : $\mathbb{P}$ ("le résultat est paire") $= 1/2$.
  Exemple 3 : $\mathbb{P}$ ("le résultat est plus grand que $1/3$) $= 2/3$.

**Exercice** : quelle est la probabilité que deux élèves d'une même classe de $N$ individus soient nés le même jour ?

# Caractérisation d'une variable aléatoire

- Variable aléatoire $X$ = application mesurable de $\Omega$ dans $\mathbb{R}$.
  Une **réalisation** $X(\omega)$, $\omega \in \Omega$, est donc un **nombre**, pouvant prendre des valeurs **discrètes** ou **continues**.
- La loi d'une variable aléatoire est caractérisée par sa fonction de répartition (CDF) $F_X$ :

$$F_X(x) = \mathbb{P}(X \leq x), \quad \mathbb{P}(a < x \leq b) = F_X(b) - F_X(a), \ a < b.$$

- $F_X$ est une fonction non-décroissante et continue à droite, telle que $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$.
- Toute fonction vérifiant ces quatre propriétés est une potentielle CDF, et définit une v.a.
- Dans le cas où la CDF de $X$ a une composante discrète en $a$, alors

$$\mathbb{P}(X = a) = F_X(a) - \lim_{x \to a^-} F_X(x).$$

# Caractérisation d'une variable aléatoire

- Si la CDF de $X$ est continue, alors $X$ est une v.a. continue. De plus, si l'image par $F_X$ de tout ensemble de mesure nulle (pour la mesure de Lebesgue) est de mesure nulle (propriété N de Luzin), alors $F_X$ est absolument continue, et il existe une fonction $f_X$ (appelée densité de probabilité ou PDF) intégrable (au sens de Lebesgue) telle que :

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy, \quad F_X(b) - F_X(a) = \int_a^b f_X(y)dy, \ a < b.$$

- La fonction $f_X$ est égale à la **dérivée** de $F$ presque partout.
- Si $X$ est une v.a. à valeurs discrètes $x_1, x_2, \ldots$, sa CDF est discontinue en ces points, et constante entre eux :

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} \mathbb{P}(x = x_i).$$

- L'ensemble des points de discontinuité de $F_X$ sur tout intervalle est **au plus dénombrable** (Conséquence directe du théorème de Froda).

# Caractérisation d'une variable aléatoire

Preuve : soit $[a, b]$ un intervalle, $0 < \alpha < \beta$. On définit alors $\mathcal{S}_{\alpha,\beta} := \left\{x_1 < \ldots < x_{N(\alpha,\beta)}\right\}$ comme l'ensemble des $N(\alpha, \beta)$ points (de discontinuité potentiels) de cet intervalle tels que pour tout $x_i \in \mathcal{S}_{\alpha,\beta}$, $\alpha \leq F_X(x_i^+) - F_X(x_i^-) < \beta$. Par propriété de la CDF, on déduit :

$$F_X(b) - F_X(a) \geq \sum_{i=1}^{N(\alpha,\beta)} F_X(x_i^+) - F_X(x_i^-) \geq N(\alpha, \beta)\alpha.$$

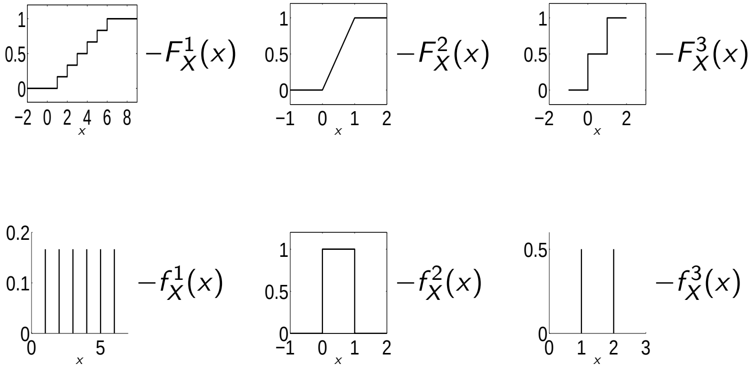
Comme $F_X(b) - F_X(a)$ est finie, on en déduit que la valeur de $N(\alpha, \beta)$ est finie (ou nulle). L'ensemble des points de discontinuité, $\mathcal{S}$, de $F_X$ sur $[a, b]$ peut alors s'écrire comme l'union des ensembles suivants :

$$\mathcal{S} = \mathcal{S}_{1,1+\epsilon} \cup \bigcup_{n=1}^{+\infty} \mathcal{S}_{\frac{1}{n},\frac{1}{n-1}},$$

c'est à dire une union d'ensemble disjoints au plus dénombrables, $\mathcal{S}$ est ainsi au plus dénombrable.

# Caractérisation d'une variable aléatoire

Exercice : reconnaître sur ces figures les CDF et PDF associées au lancé de dé, de pièce, et au tirage uniforme sur [0,1].



# Caractérisation d'une variable aléatoire

Exemples de distributions classiques :

- Distribution **uniforme** $\mathcal{U}(a,b)$ : $f_X(x) = \frac{1}{b-a}\mathbb{I}_{[a,b]}(x)$.

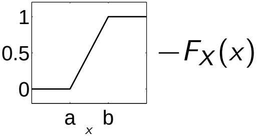



FIGURE: Loi uniforme

- Distribution **exponentielle** $\mathcal{E}(\mu)$ : $f_X(x) = \frac{1}{\mu}\exp(-x/\mu)$.

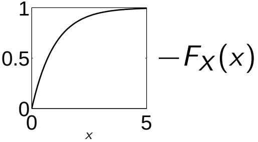



FIGURE: Loi exponentielle

# Caractérisation d'une variable aléatoire

Exemples de distributions classiques :

- Distribution **gaussienne** $\mathcal{N}(\mu, \sigma)$ : $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$.

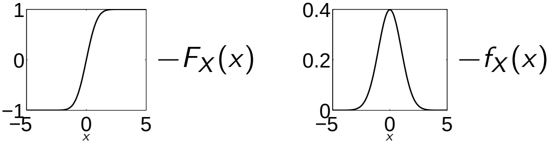



FIGURE: Loi gaussienne

- Distribution **log-normale** $W \sim \mathcal{N}(\mu, \sigma)$, $X = \exp(W)$ :

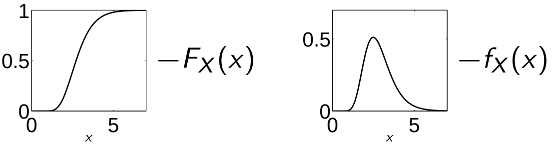



FIGURE: Loi log-normale

# Caractérisation d'une variable aléatoire

Retrouver les distributions associées aux configurations suivantes :

- Je modélise les cours des actions, les taux d'intérêt, le nombre de mots dans une phrase, ou encore la répartition des revenus dans la population. Qui suis-je ?
- Je caractérise la durée de vie des composants électronique et des atomes radioactifs. Qui suis-je ?
- Je sers de guide dans la recherche en sciences physique et sociale, en médecine, en agriculture et en ingénierie, je suis un outil indispensable pour l'analyse et l'interprétation des données de base obtenues par l'observation et l'expérience (ex : mouvement brownien...). Qui suis-je ?
- Je maximise l'incertitude associée à une plage de variation, et suis utilisé pour générer tout type de loi. Qui suis-je ?

# Caractérisation d'une variable aléatoire

**Loi des maxima** : soit $\{X_i,\ 1 \leq i \leq n\}$ $n$ v.a. indépendantes de même fonction de répartition $F_X$. On définit alors la nouvelle v.a. $Y_n = \max_{1 \leq i \leq n} X_i$, de fonction de répartition $P_n$, telle que :

$$P_n(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X_1 \leq y) \cap \cdots \cap \mathbb{P}(X_n \leq y) = \{F_X(y)\}^n .$$

La **loi d'extremum généralisée**, $\Phi_G(y) = \exp\left\{ -\left[ 1 + \xi \left( \frac{y-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}$ permet ainsi de modéliser (après identification de ses paramètres $\xi,\ \mu,\ \sigma$) le comportement asymptotique de nombreuses lois $F_X$ :

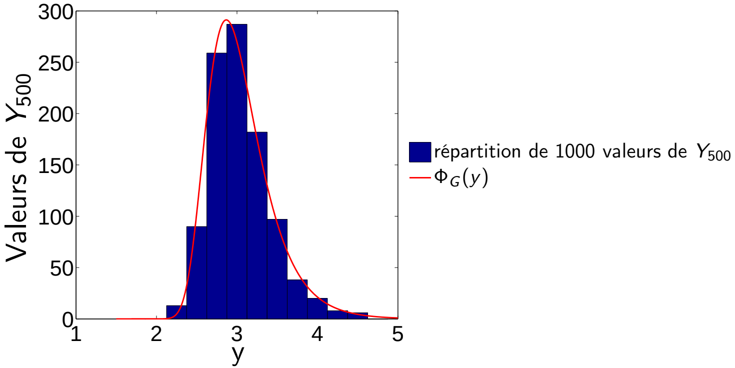
$$\lim_{n \to +\infty} P_n(y) \approx \Phi_G(y).$$

- $\xi \to 0$ : **Gumbel** $\leftrightarrow$ $F_X(x) \approx 1 - a\exp(-bx)$,
- $\xi > 0$ : **Frechet** $\leftrightarrow$ $F_X(x) \approx 1 - a/x^k$,
- $\xi < 0$ : **Weibull** $\leftrightarrow$ $F_X(x) \approx 1 - a(X_0 - x)^k$.

.

# Caractérisation d'une variable aléatoire

**Loi des maxima - exemple** : $X_i \sim \mathcal{N}(0,1), \; Y_{500} = \max_{1 \leq i \leq 500} X_i$.

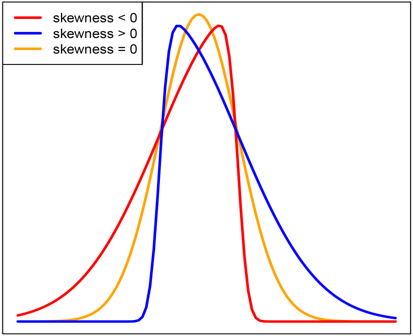


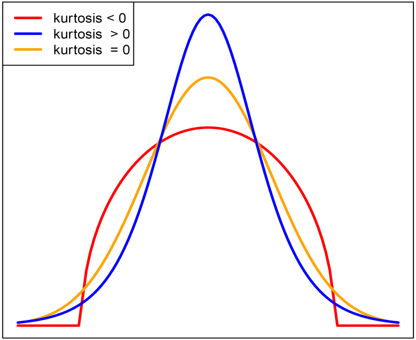
# Caractérisation d'une variable aléatoire

- On note $\mathbb{E}[X]$ (resp. $\mathbb{E}[f(X)]$) la **moyenne** (espérance) d'une variable aléatoire $X$ (resp. d'une fonction de $X$) de densité $f_X$ :
$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)dx, \quad \mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) f_X(x)dx.$$
- La distribution de $X$ peut ainsi être également caractérisée par ses **moments** $m_k$ d'ordre $k$, qui, s'ils existent, se définissent par :
$$m_k = \int_{\mathbb{R}} x^k f_X(x)dx = \mathbb{E}\left[X^k\right].$$
- Au niveau vocabulaire et notation, on écrit généralement :
  - $\mu = \mathbb{E}[X]$, la **moyenne** de $X$,
  - $\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right] = m_2 - m_1^2$ la **variance** de $X$ ($\sigma \leftrightarrow$ **écart-type**), qui mesure la dispersion de $X$ autour de sa moyenne,
  - $\gamma_1 = \mathbb{E}\left[((X-\mu)/\sigma)^3\right]$ le **skewness** ou coefficient d'**asymétrie**,
  - $\gamma_2 = \mathbb{E}\left[((X-\mu)/\sigma)^4\right] - 3$ le **kurtosis** ou coefficient d'**aplatissement**.

# Caractérisation d'une variable aléatoire

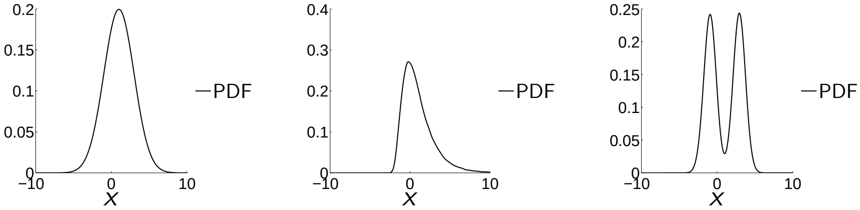
Illustration graphique des coefficients d'asymétrie et d'aplatissement.



# Caractérisation d'une variable aléatoire

Exercice : classer les distributions suivantes en fonction de leurs moyennes, variances, asymétries et aplatissements.




La connaissance de la moyenne et de la variance est nécessaire mais NON suffisante pour la caractérisation d'une variable aléatoire !

# Caractérisation d'une variable aléatoire

Vocabulaire et notation :

- **Coefficient de variation** $\delta$ :

$$\delta = \sigma/\mu.$$

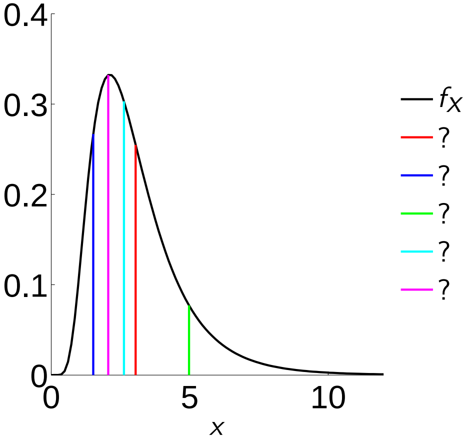
- **Quantile** $q_\alpha$ d'ordre $0 \leq \alpha \leq 1$ :

$$\mathbb{P}(X \leq q_\alpha) = \alpha.$$

- **Médiane** $m = q_{0.5}$.
- **Mode** $M = \arg\max_x f_X(x)$.

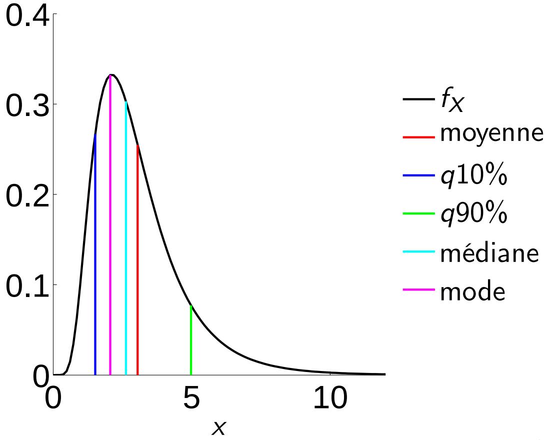
# Caractérisation d'une variable aléatoire

Pour la variable $X$ de PDF $f_X$ représentée sur cette figure, retrouver les couleurs associées aux quantiles à 10% et 90%, le mode, la médiane et la moyenne.



# Caractérisation d'une variable aléatoire

Pour la variable $X$ de PDF $f_X$ représentée sur cette figure, retrouver les couleurs associées aux quantiles à 10% et 90%, le mode, la médiane et la moyenne.



# Plan de la séance

1. Probabilités

2. Rappels de génération aléatoire de v.a.

# Génération de réalisations i.i.d. d'une variable aléatoire

On se donne une v.a. $X$ caractérisée par sa CDF $F_X$ (connue), et on nomme (si elle existe) $f_X$ sa PDF.

**Objectif**

Générer $N$ rélisations indépendantes de $X$

- Méthode de la CDF inverse.
- Méthode de réjection.
- Méthode de chaînes de Markov.

# Génération de réalisations i.i.d. d'une variable aléatoire

La méthode de la CDF inverse est la méthode la plus utilisée pour simuler une réalisation d'une v.a. à partir de sa CDF. On procède en deux étapes :

1. Considérer U uniformément distribué sur $[0, 1]$,
2. Former $X = F_X^{-1}(U)$.

En effet :

$$\begin{aligned} \mathbb{P}(F_X^{-1}(U) \leq x) &= \mathbb{P}(U \leq F_X(x)), \\ &= F_X(x). \end{aligned}$$

# Génération de réalisations i.i.d. d'une variable aléatoire

Exercice : expliquer comment générer à partir de la loi uniforme une variable aléatoire

- de loi triangulaire $\mathcal{T}(a, b, c)$,
- de loi Gaussienne $\mathcal{N}(\mu, \sigma^2)$.

# Génération de réalisations i.i.d. d'une variable aléatoire

Alternativement, la méthode de réjection est également très utilisée.

## Cadre d'application

- $f_X$ à support borné $[x_m, x_M]$
- $f_X$ borné : $f_X(x) \leq s$.
  On introduit alors la notion d'aire $\mathcal{A} = s(x_M - x_m) \geq 1$.

## Algorithme

- Tirer U uniformément distribué sur $[x_m, x_M]$.
- Tirer $\eta$ uniformément distribué sur $[0, s]$.
- Si $\eta \leq f_X(U)$, alors X=U.

# Génération de réalisations i.i.d. d'une variable aléatoire

## Remarques sur la méthode de réjection

- Probabilité d'acceptation : $1/\mathcal{A}$.
- Temps d'attente $T$ : $\mathbb{P}(T=k) = \frac{1}{\mathcal{A}}\left(1-\frac{1}{\mathcal{A}}\right)^{k-1}$.
- $\mathbb{E}\left[T\right] = \mathcal{A}$.

$\Rightarrow$ Problème d'application pour des lois très piquées. $\Rightarrow$ Principalement utilisée pour la génération de lois contraintes, si bien que la CDF est dure (ou impossible) à calculer.

# Génération de réalisations i.i.d. d'une variable aléatoire

Exercice : expliquer comment générer

- à partir de la loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ une variable aléatoire gaussienne tronquée sur $[a, b]$,
- à partir de la loi uniforme une variable aléatoire uniforme sur [a,b] telle que son carré soit plus petit que sa valeur.

# Génération de réalisations i.i.d. d'une variable aléatoire

- Il peut arriver que la PDF $f_X$ soit trop compliquée ou pas assez bien connue pour qu'on puisse appliquer les précédentes méthodes.
- Ceci arrive typiquement lorsque $f_X$ n'est connue qu'à une constante multiplicative près (cadre bayésien).
- Les méthodes de **Monte Carlo par chaîne de Markov** (MCMC) sont spécialement conçues pour générer des réalisations de $X$ dans ce cadre.
  - $\rightarrow$ construire une chaîne de Markov ergodique, dont l'unique probabilité invariante est $f_X$.
  - $\rightarrow$ plusieurs déclinaisons possibles d'algorithmes MCMC.

# Génération de réalisations i.i.d. d'une variable aléatoire

## Algorithme de Métropolis-Hastings

- Soit $\boldsymbol{x}_0$ un point de départ, et $(q(\boldsymbol{x}', \boldsymbol{x}))_{\boldsymbol{x} \in \mathbb{R}^d}$ une loi d'exploration ($\leftrightarrow$ une famille de PDFs paramétrées par $\boldsymbol{x}'$ facilement simulables).
- On suppose qu'on a simulé le n$^e$ terme de la chaîne, noté $\boldsymbol{x}_n$.

(1) on fait une proposition $\boldsymbol{x}'$ tirée selon la PDF $q(\boldsymbol{x}_n; \cdot)$.

(2) on calcule le taux d'acceptation $a(\boldsymbol{x}_n; \boldsymbol{x}_0) = min(1; \rho(\boldsymbol{x}_n; \boldsymbol{x}_0))$ avec :

$$\rho(\boldsymbol{x}_n; \boldsymbol{x}') = \frac{p(\boldsymbol{x}')q(\boldsymbol{x}'; \boldsymbol{x}_n)}{p(\boldsymbol{x}_n)q(\boldsymbol{x}_n; \boldsymbol{x}')}.$$

(3) on tire $u \sim U(0; 1)$.

(4) on pose :

$$\boldsymbol{x}_{n+1} = \begin{cases} \boldsymbol{x}' \text{ si } u \leq a(\boldsymbol{x}_n; \boldsymbol{x}') \\ \boldsymbol{x}_n \text{ si } u > a(\boldsymbol{x}_n; \boldsymbol{x}'); \end{cases},$$

$\Rightarrow$ on accepte avec proba. $a(\boldsymbol{x}_n; \boldsymbol{x}')$ la proposition $\boldsymbol{x}'$, on la rejette sinon.

# Génération de réalisations i.i.d. d'une variable aléatoire

## Deux exemples classiques de loi d'exploration

1. $q(\boldsymbol{x}', \boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{x}', \sigma^2 \boldsymbol{I})$ (**réversible**),
2. $q(\boldsymbol{x}', \boldsymbol{x}) \sim \frac{1_{\boldsymbol{x} \in [0,1]^d} 1_{\boldsymbol{x}'-\epsilon \leq \boldsymbol{x} \leq \boldsymbol{x}'+\epsilon}}{\int_{\boldsymbol{x} \in [0,1]^d} 1_{\boldsymbol{x}'-\epsilon \leq \boldsymbol{x} \leq \boldsymbol{x}'+\epsilon} d\boldsymbol{x}}$ (**non réversible**).

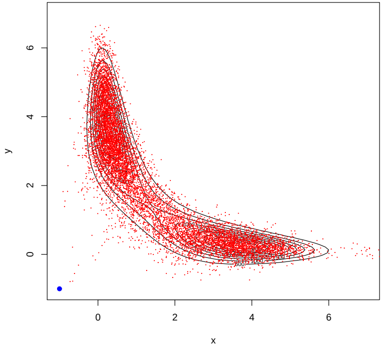
- Ces lois dépendent souvent de paramètres (ici $\sigma$ ou $\epsilon$) qu'il faut calibrer pour avoir un taux d'acceptation ni trop fort (pas assez d'exploration), ni trop faible (on reste coincé).
- Une règle empirique classique est de calibrer ces grandeurs au cours de la simulation de manière à observer un taux d'acceptation constant de l'ordre de 25%.

$\Rightarrow$ On peut alors montrer, sous des conditions assez générales sur $f_X$ et $q$ (par ex., le support de $q$ doit inclure le support de $f_X$), que l'algorithme précédent définit une chaîne de Markov ergodique dont l'unique mesure invariante est $f_X$.

# Génération de réalisations i.i.d. d'une variable aléatoire

## Exemple d=2

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \;\propto\; \exp\left(-0.5(x_1^2 x_2^2 + x_1^2 + x_2^2 - 8x_1 - 8x_2)\right).$$



# Plan de la séance

1. Probabilités

2. Rappels de génération aléatoire de v.a.