

# Cours de probabilités

SAPHIRE - 215 Probabilités

Solal Nathan

Alexandre Iooss

## Table des matières

<b>1</b>	<b>Introduction, philosophie, exemples</b>	<b>2</b>
1.1	Logique booléenne . . . . .	2
1.2	Probabilité conditionnelle et axiome fondateur . . . . .	2
1.3	Tirage sans remise, loi hypergéométrique . . . . .	3
1.4	Tirage avec remise, loi binomiale . . . . .	4
1.5	Loi de poisson . . . . .	5
1.6	Applications de la probabilité dans différents domaines de recherches . . . . .	5
<b>2</b>	<b>Probabilités continues</b>	<b>6</b>
2.1	Fonction de répartition et de densité de probabilité . . . . .	6
2.2	Loi Beta . . . . .	7
2.3	Espérance et variance . . . . .	7
2.4	Résolution de problème inverse . . . . .	8
2.5	Fonctions caractéristiques . . . . .	8
2.6	Loi multi-variée . . . . .	9
2.7	Marginalisation . . . . .	9
2.8	Covariance et corrélation . . . . .	9
2.9	Changement de variable . . . . .	10
2.10	Somme de variables indépendantes . . . . .	11
2.11	Loi Gaussienne, Minimiser le bruit . . . . .	11
2.12	Loi de Laplace . . . . .	11
2.13	Test d'hypothèse . . . . .	12
<b>3</b>	<b>Description approchée des lois de probabilité</b>	<b>12</b>
3.1	Loi Gaussienne . . . . .	12
3.2	Filtre de Kalman . . . . .	13

école \_\_\_\_\_  
normale \_\_\_\_\_  
supérieure \_\_\_\_\_  
paris—saclay \_\_\_\_\_

université  
PARIS-SACLAY



# 1 Introduction, philosophie, exemples

## 1.1 Logique booléenne

Voir UE 231.

## 1.2 Probabilité conditionnelle et axiome fondateur

**Règle de la Somme** : Soient A et B deux propositions logiques,

$$P(A|B) + P(\bar{A}|B) = 1$$

**Règle du Produit** : Soient A, B et C trois propositions logiques,

$$P(AB|C) = P(A|BC)P(B|C) = P(B|AC)P(A|C)$$

### 1.2.1 Propriétés des probabilités conditionnelles

**Théorème de Bayes** : Soient A, B et C trois variables logiques,

$$P(A|BC) = \frac{P(B|AC)P(A|C)}{P(B|C)}$$

*Démonstration* avec la règle du produit et  $P(B|C) \neq 0$ .

**Formule de Poincaré (somme)** : Soient A, B et C trois propositions logiques.

$$P(A + B|C) = P(A|C) + P(B|C) - P(AB|C)$$

On généralise,

$$P\left(\sum_{k=1}^N A_k|C\right) = \sum_{k=1}^N (-1)^{k-1} \sum_{1 < i_1 < \dots < i_k < N} P(A_{i_1} A_{i_2} \dots A_{i_k}|C)$$

**Indépendance** : A et B sont indépendantes si et seulement si

$$P(A|BC) = P(A|C) \quad \text{ou} \quad P(B|AC) = P(B|C)$$

**Mutuellement exclusives** : Si A et B sont des propositions mutuellement exclusives alors

$$P(AB|C) = 0 \quad \text{et} \quad P(A + B|C) = P(A|C) + P(B|C)$$

### 1.2.2 Propositions exhaustives et mutuellement exclusives

Soit  $(A_n)$  un ensemble de propositions logiques. Elles sont **exhaustives et mutuellement exclusives** si seulement si

$$\forall i, j \in \mathbb{N}, i \neq j, P(A_i A_j | C) = 0 \quad \text{et} \quad P\left(\sum A_n | C\right) = 1$$

Soit  $(A_n)$  un ensemble de propositions logiques exhaustives et mutuellement exclusives et B, C deux propositions logiques quelconques,

$$P(B|C) = \sum_n P(B|A_n C) P(A_n | C)$$

Alors par le théorème de Bayes,

$$P(A_n | BC) = \frac{P(B|A_n C) P(A_n | C)}{\sum_i P(B|A_i C) P(A_i | C)}$$

### 1.3 Tirage sans remise, loi hypergéométrique

On pose :

- une urne C contenant N boules (M noires, N - M blanches) ;
- l'évènement B « on tire une boule blanche » ;
- l'évènement N « on tire une boule noire ».

Quelle est la probabilité de tirer 2 boules noires ?

Quelle est la probabilité de tirer r boules noires et n - r boules blanches ?

$$P(N_1 \dots N_r) = \frac{M!(N-r)!}{N!(M-r)!}$$
$$P(B_1 \dots B_b) = \frac{(N-M)!(N-M)!}{N!(N-M-B)!}$$

La probabilité de tirer r boules noires puis n - r boules blanches est donc

$$P(N_1 \dots N_r B_{r+1} \dots B_n) = \frac{M!(N-M)!(N-n)!}{(M-r)!(N-M-n-r)!N!}$$

Soit A « on tire exactement r boules noires en n tirages », alors

$$P(A|C) = \frac{\binom{M}{r} \binom{N-M}{n-r}}{\binom{N}{n}}$$

$P(A|C)$  suit la loi hypergéométrique,

$$\mathcal{H}_y(r|N, M, n) = \frac{\binom{M}{r} \binom{N-M}{n-r}}{\binom{N}{n}}$$

avec  $0 \leq r \leq n \leq N$ .

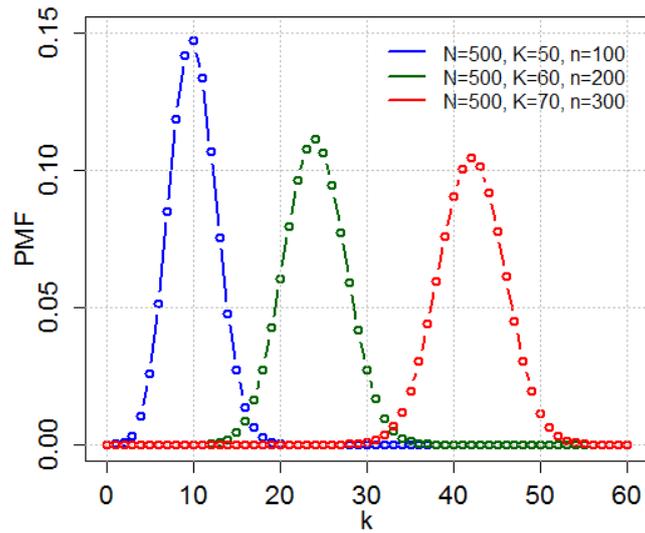


FIG. 1 : Loi hypergéométrique  $\mathcal{H}_y(k|N, K, n)$

## 1.4 Tirage avec remise, loi binomiale

On pose :

- une urne C contenant  $N$  boules ( $M$  noires,  $N - M$  blanches) ;
- l'évènement  $B$  « on tire une boule blanche au  $i$ -ème tirage » ;
- l'évènement  $N$  « on tire une boule noire au  $i$ -ème tirage ».
- l'évènement  $A$  « on tire  $r$  boules noire au cours des  $n$  premiers tirages ».

On pose  $p = M/N$ .  $A$  suit la loi binomiale

$$\mathcal{B}(p, n) = C_r^n p^r (1 - p)^{n-r}$$

La moyenne est  $np$  et la variance est  $n(1 - p)p$ .

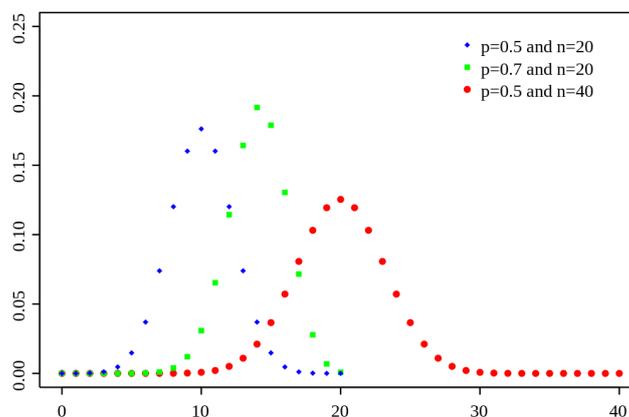


FIG. 2 : Loi binomiale  $\mathcal{B}(p, n)$

Les probabilité modélisent ce que l'on connaît/connait pas (passé/présent/futur). Il n'y a pas forcément de lien de causalité. C'est la connaissance des chose qui est modélisée.

On peut facilement approximer une loi hypergéométrique par une loi binomiale, qui est bien plus facile à calculer. C'est pour ça qu'elle est bien plus utilisée en pratique.

## 1.5 Loi de poisson

La loi de poisson de paramètre  $\lambda$  s'exprime

$$P(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Se retrouve avec le développement limité d'exponentielle car  $\sum_k P(k|\lambda) = 1$ .

La **moyenne** est  $\lambda$  et la **variance** est  $\lambda$ .

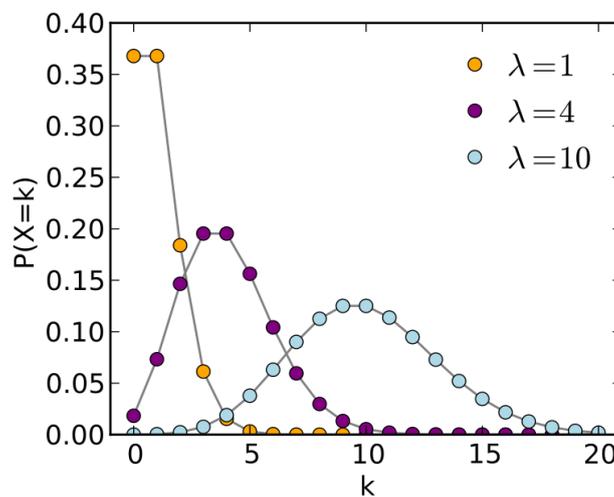


FIG. 3 : Loi de poisson

**Exemple d'utilisation** : Lors de la mesure d'antineutrinos au fond des océans, pendant le temps  $\Delta T$ . Soit l'évènement  $E$  « on détecte un neutrino », alors  $p(E|\lambda, \Delta T, C) = \lambda \Delta T$

On découpe le temps de mesure en  $n$  intervalles  $\Delta T/n$  dans lequel occurrent 0 ou 1 détections. On modélise ceci avec une loi binomiale qui tend vers la loi de poisson :

$$P(r|\lambda, n, t) = \binom{n}{r} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r} \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^r}{r!}$$

## 1.6 Applications de la probabilité dans différents domaines de recherches

### 1.6.1 Correction de défauts sans des images astrophysique

Mesures sol des rayonnement infrarouge impossible. On envoie des télescopes dans l'espace (Hubble, Spitzer). Données caractéristiques des composants chimiques dans le spectre des nuages de gaz.

On diminue la confiance en la donnée si il y a eu un événements juste avant (exemple : parasite à haute énergie qui vient pendant quelques temps sur le semi-conducteur)

### 1.6.2 Mission Herschel/Plank

Toujours des satellite pour mesurer dans l'infrarouge

Reconstruction sur-résolue grâce à plusieurs mesure de la même chose. On peut alors corriger les stries (dérive du détecteur) et augmenter la résolution

### 1.6.3 Imagerie médicale

Tomographie émission de petite onde. Permet de détecter les tumeurs/cancers beaucoup plus tôt qu'avant. Permet d'éviter la prolifération des métastases. Statistique poissonienne.

### 1.6.4 Tomographie dentaire

Problème des artefacts métalliques dans l'imagerie dentaire. Reconstitution d'un signal bruité. On associe une confiance aux ondes en fonction de la manière dont elles sont atténuées.

## 2 Probabilités continues

Le GPS par exemple est une grandeur d'intérêt qui possède des valeurs réelles.

### 2.1 Fonction de répartition et de densité de probabilité

**Fonction de répartition**  $F$  (intégration) :

$$F(t_0) = P(t < t_0|C)$$

d'où la probabilité d'un intervalle  $[a, b]$  :

$$P(t \in [a, b]|C) = F(b) - F(a)$$

**Propriétés** :  $F(t) : \mathbb{R} \rightarrow [0, 1]$ ,  $F(t)$  est croissante, continue à gauche.  $\lim_{t \rightarrow -\infty} F(t) = 0$  et  $\lim_{t \rightarrow +\infty} F(t) = 1$ .

**Fonction de densité de probabilité** : Si  $F$  est dérivable,

$$f(t) = F'(t)$$

**Propriétés** : Comme  $F$  est croissante,  $f$  est positive. et  $\int_{\mathbb{R}} f(t) dt = 1$ .

*Attention !*  $f$  n'est pas une probabilité.

$$P(t \in [x, x + dx]|C) = f(x) dx$$

### 2.1.1 Tirage dans une urne

**Exemple :** Tirage dans une urne

On connaît  $P(r|p, n, C) = \mathcal{B}(n, p)$ , on cherche  $P(p|r, n, C)$  grâce à la règle de Bayes.

**Principe d'indifférence :** On ne privilégie pas de valeur a priori (loi uniforme).

**Application :**

$$P(A_p|r, N, C) = \frac{P(r|A_p, N, C)P(A_p|C)}{P(r|N, C)}$$

$$f(p|r, N, C) = \frac{P(r|p, N, C)f(p|C) dp}{P(r|N, C)}$$

Calcul du dénominateur par marginalisation :

$$f(p|r, N, C) = \frac{P(r|p, N, C)f(p|C)}{\int_0^1 P(r|p, N, C)f(p|C) dp}$$

et la formule d'Euler  $\int_0^1 p^r (1-p)^{(N-r)} dp = \frac{r!(N-r)!}{(N+1)!}$

## 2.2 Loi Beta

$$\mathcal{B}_e(p|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

**La moyenne** est  $\frac{\alpha}{\alpha+\beta}$  si  $\alpha, \beta > 0$ . **La variance** est  $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ .

L'estimateur de la moyenne a posteriori est plus robuste (il minimise le risque quadratique).

## 2.3 Espérance et variance

**Espérance dans le cas continu :**

$$E\{h(x)\} = \int_{\mathbb{R}} h(x)f(x|C) dx$$

**Espérance dans le cas discret :**

$$E\{h(x)\} = \sum_i h(x_i)P(x = x_i|C)$$

**Linéarité :**  $E\{a + \lambda b\} = E\{a\} + \lambda E\{b\}$

**Moments** :  $M_n = \int_{\mathbb{R}} x^n f(x|C) dx$

**Variance dans le cas continu** :

$$var\{x\} = E\{(x - E\{x\})^2\}$$

## 2.4 Résolution de problème inverse

On fait de l'estimation de paramètre à partir des données d'entrée. Résolution d'un problème inverse.

En pratique on va chercher un estimateur de  $\theta$ , noté  $\hat{\theta}$ .

On définit alors l'erreur

$$e = \theta - \hat{\theta}$$

On cherche alors le meilleur estimateur possible : un estimateur optimum, qui dépend du critère de qualité utilisé. Le plus utilisé est celui de l'Erreur Quadratique Moyenne.

On définit également le biais :

$$biais(\hat{\theta}) = E\{e\}$$

$$var(\hat{\theta}) = E\{(\hat{\theta} - E\{\hat{\theta}\})^2\}$$

Si on essaie de minimiser à la fois le biais et la variance alors on minimise l'erreur quadratique moyenne :

$$EQM = E\{(\theta - \hat{\theta})^2\} = var\{\hat{\theta}\} + biais(\hat{\theta})^2$$

Pour minimiser l'EQM on prend l'espérance a posteriori (EAP) :

$$\hat{\theta}^{EAP} = E\{\theta\} = \int \theta f(\theta|D, C) d\theta$$

## 2.5 Fonctions caractéristiques

Les fonctions caractéristiques sont l'analogie des transformées de Fourier dans le cas des probabilités continues. Trois façon de représenter la même chose : fonction de répartition (utile dans certains cas non linéaire), fonction de densité de probabilité (utile presque tout le temps) et enfin fonction caractéristique (utile quand on somme deux variables indépendantes).

**Fonction caractéristique** :

$$\phi_x(u) = E\{e^{jux}\}$$

$\phi_x(u)$  est continue, symétrique hermitienne, majorée par  $\phi_x(0) = 1$ .

## 2.6 Loi multi-variée

Généralisation aux fonctions vectoriels pour les lois multivariés :

$$F(\underline{r}) = F(r_1, r_2, \dots, r_n) = P(q_1 < r_1, \dots, q_n < r_n) = P(\underline{q} < \underline{r})$$

On définit la fonction de densité de probabilité de la même façon.

## 2.7 Marginalisation

Si on a une loi jointe  $f(x, y)$  alors en marginalisant on retrouve :

$$f(y) = \int_{\mathbb{R}} f(x, y) dx$$

On intègre en projection sur les marges pour obtenir la loi.

On peut déterminer la fonction de répartition composante par composante : (sous réserve d'indépendance)

$$F(\underline{r}) = \prod_i F_i(r_i)$$

De façon analogue :

$$f(\underline{r}) = \prod_i f_i(r_i)$$

Si les variables sont indépendantes  $\Rightarrow$  fonction séparable

## 2.8 Covariance et corrélation

**La covariance :**

$$Cov\{x_k, x_j\} = E\{(x_k - E\{x_k\})(x_j - E\{x_j\})\}$$

Caractériser la dépendance partielle entre deux variables.

**Coefficient de corrélation :**

$$\rho_{kj} = \frac{Cov\{x_k, x_j\}}{\sqrt{var(x_k)var(x_j)}}$$

**Matrice de covariance :**

$$\Sigma = \begin{pmatrix} \vdots & & \\ \dots Cov\{x_k, x_j\} \dots & & \\ \vdots & & \end{pmatrix} = E\{(x - m_x)(x - m_x)^t\}$$

**Fonction d'auto-corrélation :**

$$\gamma_x(k) = \text{Cov} \{x_j, x_{j+k}\} \quad \forall j$$

Matrice à structure Toeplitz (si les variables sont stationnaires i.e. ne dépendent que de  $(k - j)$ ), diagonalisable par TF.

La densité spectrale de puissance  $\Gamma$  ou spectre du signal est la transformée de Fourier de la fonction de corrélation.

## 2.9 Changement de variable

Il faut faire le changement de variable dans la probabilité, par exemple dans la fonction de répartition, mais pas directement dans la fonction de densité de probabilité par exemple.

### 2.9.1 Cas uni-varié

$y = \phi(x)$ ,  $\phi$  une fonction continue différentiable,  $\phi^{-1}$  est définie et  $\mathcal{U}$  un intervalle de  $\mathbb{R}$

$$F_x(t) = P(x < t) = P(\phi^{-1}(y) < t) = P(y < \phi(t)) = F_y(\phi(t))$$

Donc par dérivation

$$f_x(x) = \phi'(x)f_y(\phi(x))$$

Donc par changement de variable

$$f_y(y) = \frac{f_x(\phi^{-1}(y))}{\phi'(\phi^{-1}(y))}$$

### 2.9.2 Cas multi-varié

$y = \phi(x)$ ,  $\phi$  une fonction continue différentiable,  $\phi^{-1}$  est définie et  $\mathcal{U}$  un intervalle de  $\mathbb{R}$  et  $J_\phi$  la matrice jacobienne de  $\phi$ .

$$J_\phi = \begin{pmatrix} \vdots \\ \dots \frac{\partial \phi_i(x)}{\partial x_j} \dots \\ \vdots \end{pmatrix}$$

Alors on a :

$$f_y(y) = f_x(\phi^{-1}(y)) |J_{\phi^{-1}}|$$

## 2.10 Somme de variables indépendantes

On pose  $z = x + y$ , par changement de variable  $s = x + y$

$$P(s \leq z < s + ds) = \iint f(x, y) dx dy = \int_{\mathbb{R}} f(x, s - x) dx ds = f_z(s) ds$$

Donc si  $x$  et  $y$  sont indépendantes,  $f(x, y) = f_x(x)f_y(y)$  donc

$$f_z = f_x * f_y$$

## 2.11 Loi Gaussienne, Minimiser le bruit

Loi gaussienne :

$$\mathcal{N}(x|m, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

Problématique, minimiser le bruit dans un ensemble de  $N$  mesures. La moyenne fonctionne toujours mais n'est pas forcément la solution la plus optimale.

Pour une loi gaussienne pour le bruit on obtient un loi gaussienne pour  $y$  grâce à un changement de variable :

$$\underline{y} = \underline{1}x_0 + \underline{b}$$

On pose  $\underline{b} = \underline{y} - \underline{1}x_0$

→ Vraisemblance :  $f(y|x_0)$

Pour une gaussienne : max = moyenne =  $\frac{1}{N}\sum y$   
et la variance

$$\sigma_x = \frac{\sigma_b}{\sqrt{N}}$$

, donc le bruit décroît bien en "racine de  $N$ "

Le modèle gaussien pour le bruit n'est pas le seul, et pas forcément le meilleur.

## 2.12 Loi de Laplace

Loi de Laplace :

$$\mathcal{L}_{ap}(p|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

La **moyenne** est la **médiane** qui vaut  $\mu$ . La **variance** est  $2b^2$ .

## 2.13 Test d'hypothèse

Tester la véracité de différentes hypothèses  $H_1, H_2, \dots, H_n$ .

On utilise Bayes. Fonctionne bien dans le cas binaire,  $H_1 = H$  et  $H_2 = \bar{H}$ .

Cote :

$$C(H|DI) = \frac{Pr(H|DI)}{Pr(\bar{H}|DI)}$$

Évidence :

$$\mathcal{E}(H|DI) = 10 \log_{10} C(H|DI)$$

On peut développer en somme si les données sont indépendantes. En général  $D_i$  sachant  $H_i$  indépendant mais  $D_i$  sachant  $\bar{H}_i$  par forcément.

Attention :  $P(A|B) = 1 - P(\bar{A}|B) \neq 1 - P(A|\bar{B})$

## 3 Description approchée des lois de probabilité

### 3.1 Loi Gaussienne

Il faut une fonction séparable et isotrope de  $\mathbb{R}^2$ . La seule fonction qui vérifie ces deux propriétés et la gaussienne.

$$\mathcal{N}(x|m, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

#### 3.1.1 Exemple : Voiture autonome

Problème :

- vitesse élevée
- mesures fréquentes
- mesures bruitées
- données manquantes (et difficile à estimer)
- limité en terme de calcul (temps réel sur le  $\mu\text{C}$ )

Traitement récursif, par exemple faire une moyenne récursive -> problème si la valeur change en fonction du temps on faudrait oublier les anciennes données

$$\text{moy}(x_n) = \frac{N-1}{N} \text{moy}(x_{n-1}) + \frac{y_n}{N}$$

On introduit ce qu'on appelle un facteur d'oubli on ne peut pas faire de moyenne glissante de manière récursive mais on peut mettre une pondération plus forte sur les termes les plus récents.

### 3.2 Filtre de Kalman

deux étapes : prédictions, mises à jours

On connaît  $P(x_{N-1}|y_1, \dots, y_{N-1}) = P(x_N|x_{N-1}, \dots, x_1) = P(x_N|x_{N-1})$  C'est une loi markovienne

\* protip : quand on a une loi jointe et que l'on souhaite faire disparaître l'un des termes on marginalise i.e. on intègre.

Quand on travail avec des loi gaussienne il suffit de déterminer la moyenne et la matrice de covariance, sachant qu si l'a priori et le posteriori sont des gaussiennes alors le résultats sera une gaussienne. Dans le cas contraire on est dans le domaine de la recherche.

Pour une Loi Gaussienne on peut trouver ces valeurs, et donc déterminer entièrement la loi résultante pour l'étape de prédiction et de mise à jour.

$$E\{x_N\} = \dots = F\hat{x}_{N_1|N-1} = \dots = F\Sigma_{N+1|N-1}F^t + \Sigma_b$$