

Statistique et Informatique (3IN005)

2020-2021

Nicolas Baskiotis

Sorbonne Université
équipe MLIA, Laboratoire d'Informatique de Paris 6 (LIP6)

Cours 5 :

Variable aléatoire réelle continue
Théorème central limite
Applications

- 1 Variable aléatoire réelle et théorie de la mesure
- 2 Densité de probabilité
- 3 Lois usuelles
- 4 Théorème central limite
- 5 Applications
- 6 Exemples d'applications

Motivation

Modéliser des résultats d'expériences aléatoires pouvant être des réels quelconques.

Par exemple : mesures de temps, de distances, d'espaces.

Exemple : Probabilité continue uniforme sur $[0, 1]$

On souhaite créer une mesure telle que chaque valeur est équiprobable.
Plus exactement : si $0 \leq a < b \leq 1$, $P([a, b]) = b - a$.

Plus généralement : on souhaite associer une mesure de probabilité aux *intervalles* de \mathbb{R} .

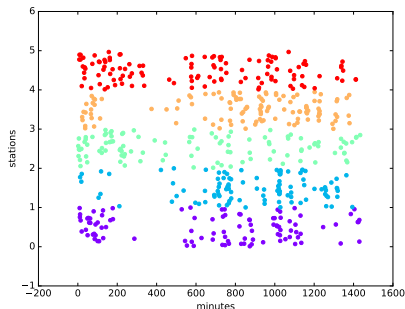
Mesures de probabilités sur \mathbb{R}

Motivation

Modéliser des résultats d'expériences aléatoires pouvant être des réels quelconques.

Par exemple : mesures de temps, de distances, d'espaces.

Exemple : Emprunt de vélo



Quelques détails techniques...

Objectif

- Probabilité d'un point : en général, nulle (sinon, probabilité discrète)
- ⇒ définir la probabilité sur un ensemble de points (infini non dénombrable, par exemple des intervalles) et qui respecte les propriétés des probabilités discrètes.

Problème...

Selon la théorie usuelle (théorie des ensembles + axiome du choix) :
Il *n'existe pas* de fonction P telle que :

- 1 P est une mesure de probabilité,
- 2 P peut être calculée pour n'importe quel sous-ensemble de $[0, 1]$,
- 3 pour tout intervalle $[a, b] \subset [0, 1] : P([a, b]) = b - a$.

Autrement dit :

Il existe des ensembles de \mathbb{R} qui sont *non mesurables* : ensembles de Vitali par exemple. (Non-mesurable ~ contradiction s'ils l'étaient).

Mesures de probabilités : définition générale (1)

Solution : les Tribus où comment se débarrasser des sous-ensembles qui nous intéressent pas

Une tribu \mathcal{T} sur un ensemble Ω contient des sous-ensembles de Ω et vérifie :

- ❶ $\Omega \in \mathcal{T}$,
- ❷ si $E \in \mathcal{T}$, alors $\Omega \setminus E \in \mathcal{T}$,
- ❸ si $(E_i)_{i \geq 1}$ est une suite d'ensembles appartenant à \mathcal{T} , alors $\bigcup_{i \geq 1} E_i \in \mathcal{T}$.

Note : l'ensemble des parties de Ω est une tribu.

Tribu de Borel

- La tribu de Borel sur \mathbb{R} , notée \mathcal{B} , est la plus petite tribu contenant tous les intervalles de \mathbb{R} , par exemple tous les ensembles du type : $\bigcup_{i \in \mathbb{N}} [a_i, b_i[$
- la tribu de Borel sur \mathbb{R}^n , notée \mathcal{B}_n est la plus petite tribu contenant les produits cartésiens de n ensembles de \mathcal{B} :
 $\{I_1 \times I_2 \times \dots \times I_n \mid \forall k, I_k \in \mathcal{B}\} \subset \mathcal{B}_n$.

Les tribus de Borel contiennent les ensembles intéressants du point de vue des probabilités.

Mesures de probabilités : définition générale (2)

Espace mesurable

Un espace mesurable est un couple (Ω, \mathcal{T}) , où Ω est un ensemble et \mathcal{T} est une tribu sur Ω . (Note : $(\mathbb{R}, \mathcal{B})$ est donc un espace mesurable.)

Mesure de probabilité

Soit (Ω, \mathcal{T}) un espace mesurable.

Une mesure sur (Ω, \mathcal{T}) est une fonction $P : \mathcal{T} \rightarrow [0, +\infty[$ telle que :

- ❶ $P(\emptyset) = 0$,
- ❷ $\forall E \in \mathcal{T}, P(E) \geq 0$,
- ❸ si $(E_i)_{i \geq 1}$ sont deux à deux disjoints, et $\forall i, E_i \in \mathcal{T}$, alors

$$P\left(\bigcup_{i \geq 1} E_i\right) = \sum_{i \geq 1} P(E_i).$$

- ❹ Si de plus $P(\Omega) = 1$, alors P est une *mesure de probabilité*.

Cette définition généralise la définition du cours 1 en prenant $\mathcal{P}(\Omega)$ comme tribu sur Ω lorsque Ω est discret.

Mesure de Borel

Fait marquant 2

Il existe une mesure λ sur $(\mathbb{R}, \mathcal{B})$, appelée mesure de Borel, telle que :

pour tout intervalle $[a, b]$ de \mathbb{R} ($a \leq b$), $\lambda([a, b]) = b - a$.

La mesure de probabilité uniforme sur un intervalle $[A, B]$, $A < B$ est alors :

$$\forall I \in \mathcal{B}, P(I) = \frac{\lambda(I \cap [A, B])}{B - A}$$

Mesure de Borel sur \mathbb{R}^n

La mesure de Borel sur $(\mathbb{R}^n, \mathcal{B}_n)$, notée λ_n est définie par :

$$\forall I_1 \in \mathcal{B}, \dots, I_n \in \mathcal{B}, \lambda_n(I_1 \times \dots \times I_n) = \prod_{k=1}^n \lambda(I_k)$$

Par exemple : $\lambda_2([0, 1/2] \times [0, 1/2]) = 1/4$.

λ_2 correspond à l'aire d'une figure dans le plan, λ_3 au volume d'un objet dans l'espace à 3 dimensions.

Plan

- 1 Variable aléatoire réelle et théorie de la mesure
- 2 Densité de probabilité**
- 3 Lois usuelles
- 4 Théorème central limite
- 5 Applications
- 6 Exemples d'applications

Densité de probabilité

Définition

Une mesure de probabilité P sur $(\mathbb{R}^n, \mathcal{B}_n)$ admet une *fonction de densité* $p : \mathbb{R}^n \rightarrow \mathbb{R}$ si : pour tout $I \subset \mathcal{B}_n$, $P(I) = \int_{x \in I} p(x) d\lambda_n(x)$

Exemples

- Si P est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$ qui admet comme fonction de densité p , alors pour tout $a < b$, on a :

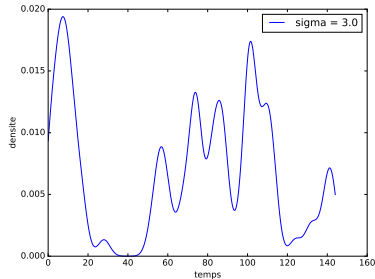
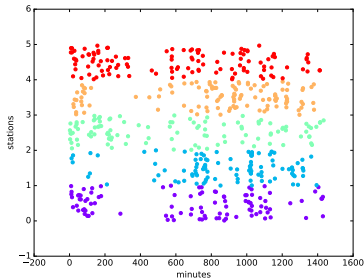
$$P([a, b]) = \int_a^b p(x) dx \quad (\text{avec les notations usuelles de l'intégrale sur } \mathbb{R}).$$

- La loi uniforme sur $[a, b]$ admet comme fonction de densité :

$$p(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases},$$

- une v.a. réelle à valeurs dans un ensemble discret *n'a pas* de fonction de densité.

Exemples : emprunt de velib



Densité de la probabilité d'emprunt d'un vélib à une borne donnée durant une journée.

Variables aléatoires à valeurs réelles

Variable aléatoire réelle

Soit P est une mesure de probabilité sur (Ω, \mathcal{T}) .

Une variable aléatoire réelle est une fonction $X : \Omega \mapsto \mathbb{R}$ telle que :

$$\forall I \in \mathcal{B}, X^{-1}(I) \in \mathcal{T}$$

X induit une mesure de probabilité P_X sur $(\mathbb{R}, \mathcal{B})$ par :

$$P_X(I) = P(X \in I) = P(X^{-1}(I)).$$

Note : cette définition généralise notre définition de variable aléatoires à valeurs réelles sur des ensembles discrets.

Fonctions de répartition et de densité d'une v.a.r.

Fonction de répartition

Soit X une variable aléatoire réelle sur (Ω, \mathcal{T}, P) .

La fonction de répartition de X , notée F_X , est définie par :

$$F_X : \left(\begin{array}{ll} \mathbb{R} & \rightarrow [0, 1] \\ t & \mapsto P(X \leq t) \end{array} \right)$$

On a alors $P(a < X \leq b) = F_X(b) - F_X(a)$.

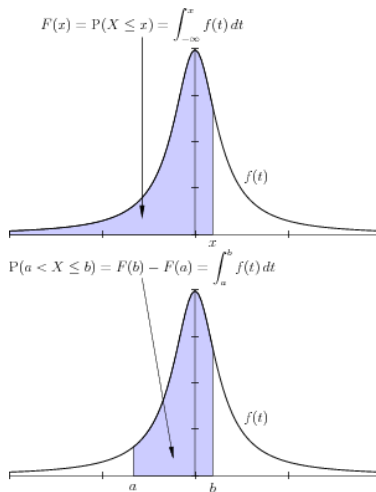
Fonction de densité

Une v.a. réelle X admet une *fonction de densité* p_X , si, pour tout $x \in \mathbb{R}$:

$$F_X(x) = \int_{-\infty}^x p_X(u) du.$$

- On a alors, pour $a < b$, $P(a < X \leq b) = \int_a^b p_X(u) du$.
- Si X est une v.a.r. telle que sa fonction de répartition est dérivable, alors $p_X = F'_X$ est une densité de X .

Fonctions de répartition et de densité d'une v.a.r.



Espérance et variance d'une v.a.r. continue

Définition et propriétés

- Soit X une v.a.r. sur espace probabilisé $(\mathbb{R}^n, \mathcal{B}_n, P)$ où P a une fonction de densité p . L'espérance de X est alors définie par :

$$\mathbb{E}(X) = \int_{\omega \in \Omega} X(\omega) p(\omega) d\lambda_n(\omega).$$

- Soit X , une v.a.r. de densité p_X . L'espérance de X est définie par :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} u p_X(u) du.$$

- La variance de X est définie par : $V(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$.

- Soit $f : \mathbb{R} \mapsto \mathbb{R}$. Alors : $\mathbb{E}(f(X)) = \int_{-\infty}^{+\infty} f(u) p_X(u) du$

Les résultats montrés pour les v.a.r. à valeurs discrètes restent vrais :

- Les propriétés de l'espérance et de la variance,
- les inégalités de Markov et Tchebychev, et la loi des grands nombres.

Application : Méthodes de Monte-Carlo

Types d'algorithmes probabilistes

- Algorithmes Las Vegas : renvoient toujours la réponse correcte, mais dans un temps variable,
- algorithmes de Monte-Carlo : les ressources sont limitées a priori, mais le résultat peut ne pas être exact.

Exemple : approximation de nombres réels

Principe de la méthode (exemples : approximation de $\ln 2$ et de π) :

- 1 écrire le nombre comme l'espérance d'une fonction d'une v.a.r. à densité,

$$\ln 2 = \int_0^1 \frac{1}{1+u} du, \quad \frac{\pi}{4} = \int_{u=0}^1 \int_{v=0}^1 I_{\{u^2+v^2 \leq 1\}} dudv$$

- 2 générer n valeurs selon la densité, calculer la fonction pour chaque valeur échantillonnée,
- 3 loi des grands nombres : la moyenne des n valeurs calculées tend vers le nombre qu'on souhaite calculer.

Plan

- 1 Variable aléatoire réelle et théorie de la mesure
- 2 Densité de probabilité
- 3 Lois usuelles**
- 4 Théorème central limite
- 5 Applications
- 6 Exemples d'applications

Loi exponentielle

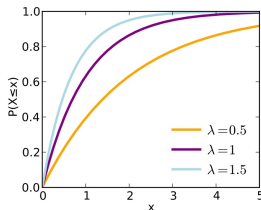
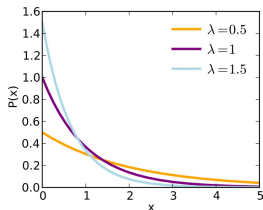
Définition et propriétés

Une v.a.r. X suit une loi exponentielle de paramètre $\lambda > 0$ si elle admet comme densité de probabilité :

$$p_X(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{sinon} \end{cases}.$$

On a alors :

- $\mathbb{E}(X) = \frac{1}{\lambda}$, $V(X) = \frac{1}{\lambda^2}$, $F_X(t) = P(X \leq t) = 1 - e^{-\lambda t}$.
- La loi exponentielle est l'analogue continu de la loi géométrique,
- elle représente le temps d'attente avant la réalisation d'un événement.

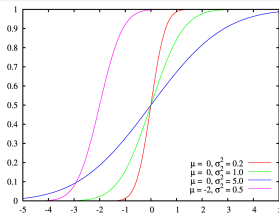
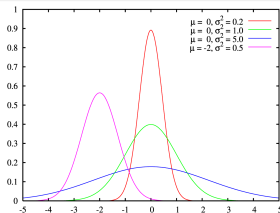


Définition et propriétés

Une v.a.r. X suit une loi normale (ou gaussienne) de paramètres μ et σ^2 si elle admet comme densité de probabilité :

$$p_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

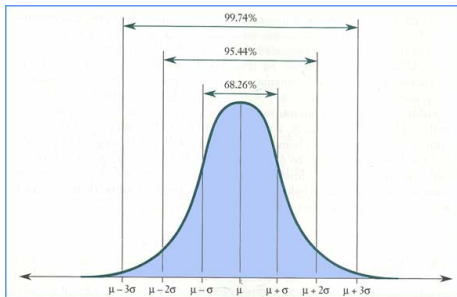
- $\mathbb{E}(X) = \mu$, $V(X) = \sigma^2$ (σ est donc l'écart-type de X).
- Si $\mu = 0$, on parle de loi *centrée*.
- Si $\mu = 0$ et $\sigma = 1$ alors X suit une loi normale *centrée réduite*.



Loi normale (2)

Propriétés additionnelles

- Symétrie : $F_X(\mu + t) = 1 - F_X(\mu - t)$ ($\Leftrightarrow P(X \leq \mu - t) = P(X \geq \mu + t)$),
- Si X suit une loi normale de paramètres (μ, σ^2) , alors $Y = \alpha X + \beta$ suit une loi normale de paramètres $\alpha\mu + \beta$ et $\alpha^2\sigma^2$.
En particulier, $\frac{X - \mu}{\sigma}$ suit une loi normale centrée réduite.
- Si X et X' sont indépendantes et suivent respectivement une loi normale de paramètres (μ, σ^2) et (μ', σ'^2) , alors $X + X'$ suit une loi normale de paramètres $\mu + \mu'$ et $\sigma^2 + \sigma'^2$.



Plan

- 1 Variable aléatoire réelle et théorie de la mesure
- 2 Densité de probabilité
- 3 Lois usuelles
- 4 Théorème central limite**
- 5 Applications
- 6 Exemples d'applications

Théorème central limite

Énoncé du théorème

Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes et identiquement distribuées, d'espérances et de variances finies.

Les X_n peuvent suivre n'importe quelle loi dès que ces deux conditions sont respectées.

On note $\mu = \mathbb{E}(X_n)$ et $\sigma = \sqrt{V(X_n)}$ l'espérance et l'écart-type de X_n .

- En notant $S_n = \sum_{k \leq n} X_k$, on a :

$\frac{S_n - n\mu}{\sigma\sqrt{n}}$ a une espérance de 0 et un écart-type de 1

- De plus, pour tout t :

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) = \Phi(t)$$

$$\text{où } \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}x^2} dx.$$

Φ est la fonction de répartition de la loi normale centrée réduite.

Théorème central limite

Énoncé du théorème

Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes et identiquement distribuées, d'espérances et de variances finies.

Les X_n peuvent suivre n'importe quelle loi dès que ces deux conditions sont respectées.

On note $\mu = \mathbb{E}(X_n)$ et $\sigma = \sqrt{V(X_n)}$ l'espérance et l'écart-type de X_n .

- En notant $S_n = \sum_{k \leq n} X_k$, on a :

$\frac{S_n - n\mu}{\sigma\sqrt{n}}$ a une espérance de 0 et un écart-type de 1

- De plus, pour tout t :

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) = \Phi(t)$$

où $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}x^2} dx.$

Φ est la fonction de répartition de la loi normale centrée réduite.

Théorème central limite

interprétation

$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) = P(Y \leq t)$ où Y suit une loi normale centrée réduite.

Formulation alternative informelle :

lorsque n est grand : $P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) \approx P(Y \leq t)$

ou encore (toujours lorsque n est grand) :

$$P(S_n \leq z) \approx P\left(Y \leq \frac{z - n\mu}{\sigma\sqrt{n}}\right)$$

Exemple : X_n sont des v.a. de Bernoulli de paramètre p

lorsque n est grand :

$$P(S_n \leq k) = \sum_{i=0}^k C_n^i p^i (1-p)^{n-i} \approx P\left(Y \leq \frac{k - np}{\sqrt{np(1-p)}}\right)$$

Théorème central limite

interprétation

$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) = P(Y \leq t)$ où Y suit une loi normale centrée réduite.

Formulation alternative informelle :

lorsque n est grand : $P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) \approx P(Y \leq t)$

ou encore (toujours lorsque n est grand) :

$$P(S_n \leq z) \approx P\left(Y \leq \frac{z - n\mu}{\sigma\sqrt{n}}\right)$$

Exemple : X_n sont des v.a. de Bernoulli de paramètre p

lorsque n est grand :

$$P(S_n \leq k) = \sum_{i=0}^k C_n^i p^i (1-p)^{n-i} \approx P\left(Y \leq \frac{k - np}{\sqrt{np(1-p)}}\right)$$

Théorème central limite

interprétation

$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) = P(Y \leq t)$ où Y suit une loi normale centrée réduite.

Formulation alternative informelle :

lorsque n est grand : $P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) \approx P(Y \leq t)$

ou encore (toujours lorsque n est grand) :

$$P(S_n \leq z) \approx P\left(Y \leq \frac{z - n\mu}{\sigma\sqrt{n}}\right)$$

Exemple : X_n sont des v.a. de Bernoulli de paramètre p

lorsque n est grand :

$$P(S_n \leq k) = \sum_{i=0}^k C_n^i p^i (1-p)^{n-i} \approx P\left(Y \leq \frac{k - np}{\sqrt{np(1-p)}}\right)$$

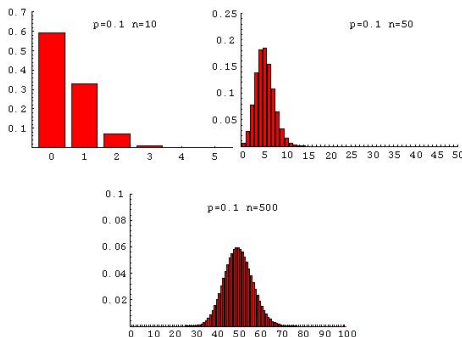
Théorème central limite

Exemple : les X_n sont des v.a. de Bernoulli de paramètre p

lorsque n est grand :

$$P(S_n \leq k) = \sum_{i=0}^k C_n^i p^i (1-p)^{n-i} \approx P\left(Y \leq \frac{k - np}{\sqrt{np(1-p)}}\right)$$

En pratique : une loi binomiale peut être approximée par une loi normale si $np \geq 10$ et $n(1-p) \geq 10$



Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes et identiquement distribuées. On note μ leur espérance et σ leur écart-type.

On note $S_n = \sum_{k \leq n} X_k$.

- Loi des grands nombres :

$$\forall t > 0, P\left(-t \leq \frac{S_n}{n} - \mu \leq t\right) \xrightarrow{n \rightarrow \infty} 1.$$

- Théorème central limite :

$$\forall t > 0, P\left(-\frac{\sigma t}{\sqrt{n}} \leq \frac{S_n}{n} - \mu \leq \frac{\sigma t}{\sqrt{n}}\right) \xrightarrow{n \rightarrow \infty} 2\Phi(t) - 1$$

Plan

- 1 Variable aléatoire réelle et théorie de la mesure
- 2 Densité de probabilité
- 3 Lois usuelles
- 4 Théorème central limite
- 5 Applications**
- 6 Exemples d'applications

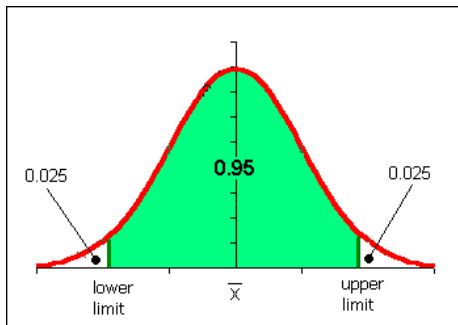
Intervalle de confiance

Définition

Soit X , une variable aléatoire réelle.

Un intervalle de confiance à $c\%$ pour un paramètre p est défini par deux fonctions u et v telles que :

$$P(u(X) < p < v(X)) \geq c/100$$



Intervalles de confiance

Exemple : I.C. pour l'espérance d'une loi normale

Soient X_1, \dots, X_n , n v.a.r. indépendantes et suivant une loi normale d'espérance μ et d'écart-type σ .

L'espérance μ est inconnue. On souhaite déterminer un intervalle de valeurs possibles pour μ , en supposant que σ est connue.

On note $X = \sum_{i=1}^n X_i$. $\frac{X - n\mu}{\sigma\sqrt{n}}$ suit une loi normale centrée réduite.

donc $\forall t, P(-t < \frac{X - n\mu}{\sigma\sqrt{n}} \leq t) = 2\Phi(t) - 1$. Donc, en fixant :

- t tel que $2\Phi(t) - 1 = c/100$ (par exemple : $t = 2$ pour 95%),
- $u(X) = \frac{X}{n} - \frac{\sigma t}{\sqrt{n}},$
- $v(X) = \frac{X}{n} + \frac{\sigma t}{\sqrt{n}},$

on a $P(u(X) < \mu < v(X)) \geq c/100$.

c'est un intervalle de confiance à $c\%$ pour le paramètre μ .

Plan

- 1 Variable aléatoire réelle et théorie de la mesure
- 2 Densité de probabilité
- 3 Lois usuelles
- 4 Théorème central limite
- 5 Applications
- 6 Exemples d'applications**

En théorie de l'information

Entropie : quantité d'information

- Je cherche à deviner un nombre entre 0 et 100 en posant des questions.
- quelle question m'apporte le plus d'information ?
 - le nombre est-il pair ?
 - le nombre finit-il par 12 ?
 - le nombre est-il supérieur à 50 ?
- Notion d'entropie : nombre minimum de question à poser pour trouver le nombre.

Définition

- Soit p une v.a. X à n valeurs distinctes i , chacune de probabilité p_i ,
- la probabilité de (x_1, \dots, x_t) tend vers $\prod_{k=1}^n p_k^{tp_k} = (\prod_{k=1}^n p_k^{p_k})^t$,
- l'entropie de p est $H(p) = - \sum_{i=1}^n p_i \log(p_i)$

Utilisations (entre autre)

- Codage de Huffman
- Arbres de décision

Maximum de vraisemblance

Soit

- une famille de modèle $\mathcal{M} = \{\mathcal{M}_\theta\}$, paramétrée par un vecteur réel θ
- Vraisemblance d'un modèle pour des données X :
$$L(X; M_\theta) = p(X|M_\theta) = \prod_{x \in X} p(x|M_\theta)$$
- Maximum de vraisemblance : choix du modèle qui maximise la vraisemblance

Applications :

- classification d'images par histogramme
- classification de séquences : reconnaissance de la parole
- infection et diffusion dans les graphes
- ...

Applications : dilemme de l'exploration/exploitation

Question : problème du bandit-manchot

- soit K machines à sous disponible, toutes différentes,
- connaissant le résultat de mes n dernières tentatives,
- quelle machine jouée pour maximiser mes gains ?

Modélisation :

- chaque machine suit une loi inconnue ν_k d'espérance μ_k ,
- soit μ^* la machine i^* d'espérance maximale,
- soit I_t la machine jouée au coup t , x_{I_t} la v.a. du gain associé,
- regret après n coups : $R_n = n * \nu_{i^*} - \sum_{t=1}^n x_{I_t}$
- minimiser : $\mathbb{E}(R_n) = n\mu^* - \mathbb{E} \sum_{i=1}^K T_i(n) \mu_i$
avec $T_i(n)$ le nombre de fois où la machine i a été jouée durant n coups.

Applications : dilemne de l'exploration/exploitation

Politique de sélection

- ϵ -greedy : jouer avec une certaine probabilité la meilleure machine, au hasard sinon.
- Upper-Confidence Bound : politique optimiste.
 - Inégalité d'Hoeffding : avec une probabilité au moins $1 - \epsilon$,

$$\mathbb{E}X \leq \frac{1}{m} \sum_{s=1}^m X_s + \sqrt{\frac{\log(\epsilon^{-1})}{2m}}$$

- UCB :

$$\operatorname{argmax} \mu_i + \sqrt{\frac{2\log(t+1)}{T_i(t)}}$$

Utilisations :

- marketing ciblé/google ads
- approximation de monte-carlo : I.A pour go