Statistique et Informatique (LI323)

Cours 6

Nicolas Baskiotis
nicolas.baskiotis@lip6.fr

Université Pierre et Marie Curie (UPMC)
Laboratoire d'Informatique de Paris 6 (LIP6)

Plan

- Lois de probabilités continues,
- théorème de la limite centrale,
- applications.

Loi exponentielle

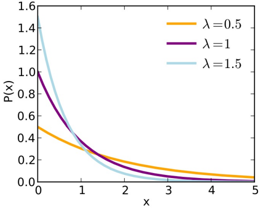
Définition et propriétés

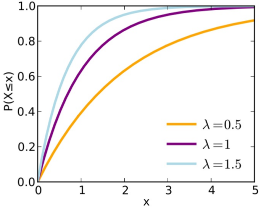
Une v.a.r. X suit une loi exponentielle de paramètree $\lambda > 0$ si elle admet comme densité de probabilité :

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

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.





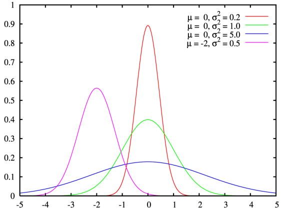
Loi normale

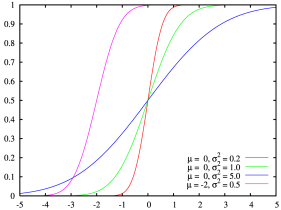
Définition et propriétés

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*.

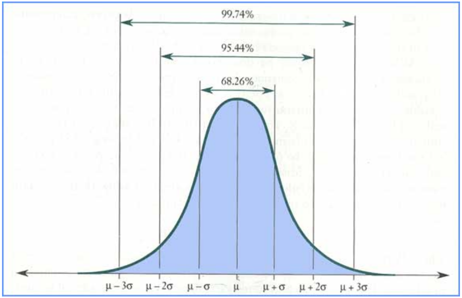




Loi normale (2)

Propriétés additionelles

- 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 $\sigma^2 + \sigma'^2$.



Théorème de la limite centrale

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 :

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

- De plus, pour tout t :

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

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

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

Théorème de la limite centrale

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 :

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

- De plus, pour tout t :

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

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 de la limite centrale

interprétation

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

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

Un exemple : les X_n sont des variables 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 de la limite centrale

interprétation

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

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

Un exemple : les X_n sont des variables 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 de la limite centrale

interprétation

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

Formulation alternative informelle :

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

ou encore (toujours lorsque n est grand) :

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

Un exemple : les X_n sont des variables de Bernoulli de paramètre p

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

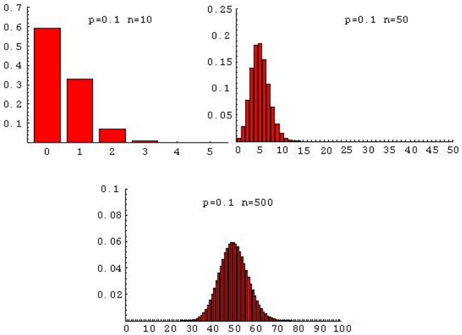
Théorème de la limite centrale

Un exemple : les X_n sont des variables 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 $n(1-p) \geq 10$



TLC et loi des grands nombres

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 :

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

- Théorème de la limite centrale :

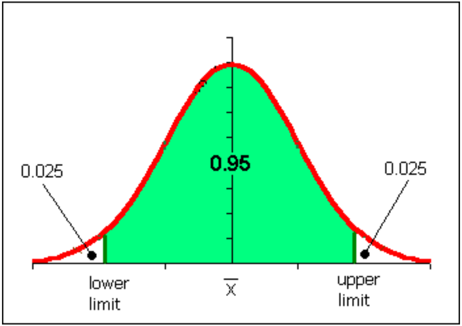
$$\forall t > 0, P\left(-\frac{\sigma t}{\sqrt{n}} \leq \frac{S_n}{n} - \mu \leq \frac{\sigma t}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{} 2\Phi(t) - 1$$

Intervalles 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:

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



Intervalles de confiance

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

Soient $X_1, ..., X_n$, 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 $X = \sum_{i=1}^{n} X_n$. $\dfrac{X - n\mu}{\sigma\sqrt{n}}$ suit une loi normale centrée réduite.

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

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

on a $P\big(u(X) < \mu < v(X)\big) \geq c/100$.
c'est un intervalle de confiance à $c\%$ pour le paramètre μ.

Applications : dilemne de l'exploration/exploitation

Question : probleme 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 l_t la machine jouée au coup t, x_{l_t} la v.a. du gain associé,
- regret après n coups : $R_n = n * \nu_{i^*} - \sum_{t=1}^{n} x_{l_t}$
- 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 ou la machine i a été joué 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 :

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

Utilisations :

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

Applications : 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 $(x_1, \cdots, x_t)$ 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