Principe de grandes déviations pour la suite des estimateurs empiriques de l'entropie d'une loi. Applications.

Philippe Regnault

Laboratoire de Mathématiques Nicolas Oresme
Université de Caen – Basse Normandie

CJPS 2012 - 17/04/2012

Entropie de Shannon d'une probabilité ; estimation

PGD pour l'estimateur de l'entropie d'une loi à partir d'un échantillon

Géométrie de l'information

Approximation numérique de la fonction de taux

L'entropie de Shannon : une mesure de l'incertitude

Notation : $\mathcal{M}_1(E)$ est l'ensemble des probabilités sur l'ensemble fini E.

Shannon (1948) introduit l'entropie d'une probabilité pour **mesurer l'incertitude** d'une source d'information.

Définition

Étant donnée une probabilité $P \in \mathcal{M}_1(E)$, l'entropie $\mathbb{S}(P)$ de P est

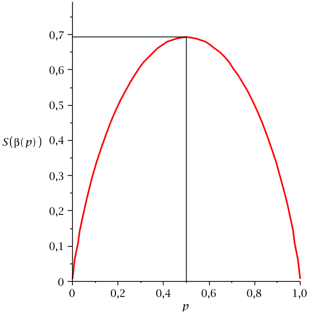
$$\mathbb{S}(P) = -\sum_{i \in E} P(i) \log P(i),$$

avec la convention $0 \log 0 = 0$.

l'entropie de Shannon : une mesure de l'incertitude

L'entropie d'une loi P sur un ensemble fini est

- maximale, et vaut $\log |E|$, lorsque P est la loi uniforme U,
- minimale, et vaut 0, lorsque P est une mesure de Dirac.



Définition

Soient $P, Q \in \mathcal{M}_1(E)$. L'information de Kullback-Leibler de Q par rapport à P est

$$\mathbb{K}(Q|P) = \sum_{i \in E} Q(i) \log \frac{Q(i)}{P(i)},$$

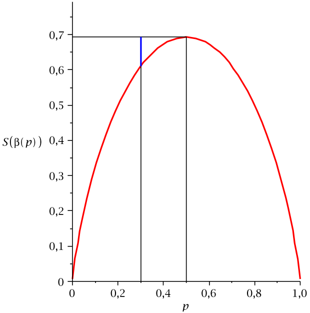
avec les conventions $0 \log 0/0 = 0$, $0 \log(0/x) = 0$ et $x \log(x/0) = +\infty$, $x \in]0, 1[$.

En particulier, $\mathbb{K}(Q|U) = \log |E| - \mathbb{S}(Q)$.

l'entropie de Shannon : une mesure de l'incertitude

L'entropie d'une loi P sur un ensemble fini est

- maximale, et vaut $\log|E|$, lorsque P est la loi uniforme U,
- minimale, et vaut 0, lorsque P est une mesure de Dirac.



Définition

Soient $P, Q \in \mathcal{M}_1(E)$. L'information de Kullback-Leibler de Q par rapport à P est

$$\mathbb{K}(Q|P) = \sum_{i \in E} Q(i) \log \frac{Q(i)}{P(i)},$$

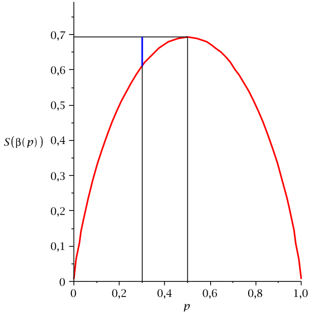
avec les conventions $0 \log 0/0 = 0$, $0 \log(0/x) = 0$ et $x \log(x/0) = +\infty$, $x \in]0,1[$.

En particulier, $\mathbb{K}(Q|U) = \log|E| - \mathbb{S}(Q)$.

l'entropie de Shannon : une mesure de l'incertitude

L'entropie d'une loi P sur un ensemble fini est

- maximale, et vaut $\log|E|$, lorsque P est la loi uniforme U,
- minimale, et vaut 0, lorsque P est une mesure de Dirac.



Définition

Soient $P, Q \in \mathcal{M}_1(E)$. L'information de Kullback-Leibler de Q par rapport à P est

$$\mathbb{K}(Q|P) = \sum_{i \in E} Q(i) \log \frac{Q(i)}{P(i)},$$

avec les conventions $0 \log 0/0 = 0$, $0 \log(0/x) = 0$ et $x \log(x/0) = +\infty$, $x \in]0, 1[$.

En particulier, $\mathbb{K}(Q|U) = \log|E| - \mathbb{S}(Q)$.

Quelques domaines d'application de l'entropie

- **Statistique :** Les estimateurs d'une loi maximisant l'entropie sous contraintes de moments possèdent de bonnes propriétés asymptotiques, en lien avec le maximum de vraisemblance, la méthode de Bayes,...
- **Grandes déviations :** La divergence de Kullback-Leibler est impliquée dans le principe de grandes déviations (PGD) de Sanov.
- **Preuves alternatives de théorèmes limites** : Le PGD de Sanov implique la loi des grands nombres ; le théorème de la limite centrée découle de la convergence en loi sous contrainte de covariance vers le maximum d'entropie (la loi normale).
- **Théorie de l'information :** Shannon et son entropie en posent les bases.

  L'entropie est la borne inférieure pour le taux de compression des codes irréductibles.
- **Économie :** L'entropie de Theil mesure les inégalités de répartition des richesses.

Estimation de l'entropie ; cas indépendant

Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $P \in \mathcal{M}_1(E)$.
Soit $\widehat{P}_n$ la probabilité empirique associée à $(X_1, \ldots, X_n)$ définie par

$$\widehat{P}_n(i) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k = i\}}, \quad i \in E.$$

Soit $\widehat{S}_n = \mathbb{S}(\widehat{P}_n)$ l'estimateur par plug-in de $\mathbb{S}(P)$.

- Basharin (1959) prouve que $\widehat{S}_n$ converge presque sûrement vers $\mathbb{S}(P)$, lorsque n tend vers l'infini.
- Zubkov (1973) et Harris (1977) prouvent que
  - si $P \neq U$, alors $\sqrt{n}(\widehat{S}_n - \mathbb{S}(P))$ est asymptotiquement normal de variance asymptotique explicite ;
  - si $P = U$, alors $2n(\mathbb{S}(U) - \widehat{S}_n)$ converge en loi vers une combinaison linéaire de variables indépendantes de loi $\chi^2(1)$.
- Girardin et Regnault (2011) établissent un principe de grandes déviations pour la suite $(\widehat{S}_n)_{n \in \mathbb{N}^*}$.

Estimation de l'entropie ; cas indépendant

Soit $(X_1, \dots, X_n)$ un n-échantillon de loi $P \in \mathcal{M}_1(E)$.
Soit $\widehat{P}_n$ la probabilité empirique associée à $(X_1, \dots, X_n)$ définie par

$$\widehat{P}_n(i) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k = i\}}, \quad i \in E.$$

Soit $\widehat{S}_n = \mathbb{S}(\widehat{P}_n)$ l'estimateur par plug-in de $\mathbb{S}(P)$.

- Basharin (1959) prouve que $\widehat{S}_n$ converge presque sûrement vers $\mathbb{S}(P)$, lorsque n tend vers l'infini.
- Zubkov (1973) et Harris (1977) prouvent que
  - si $P \neq U$, alors $\sqrt{n}(\widehat{S}_n - \mathbb{S}(P))$ est asymptotiquement normal de variance asymptotique explicite ;
  - si $P = U$, alors $2n(\mathbb{S}(U) - \widehat{S}_n)$ converge en loi vers une combinaison linéaire de variables indépendantes de loi $\chi^2(1)$.
- Girardin et Regnault (2011) établissent un principe de grandes déviations pour la suite $(\widehat{S}_n)_{n \in \mathbb{N}^*}$.

Estimation de l'entropie ; cas indépendant

Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $P \in \mathcal{M}_1(E)$.
Soit $\widehat{P}_n$ la probabilité empirique associée à $(X_1, \ldots, X_n)$ définie par

$$\widehat{P}_n(i) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k = i\}}, \quad i \in E.$$

Soit $\widehat{S}_n = \mathbb{S}(\widehat{P}_n)$ l'estimateur par plug-in de $\mathbb{S}(P)$.

- Basharin (1959) prouve que $\widehat{S}_n$ converge presque sûrement vers $\mathbb{S}(P)$, lorsque n tend vers l'infini.
- Zubkov (1973) et Harris (1977) prouvent que
  - si $P \neq U$, alors $\sqrt{n}(\widehat{S}_n - \mathbb{S}(P))$ est asymptotiquement normal de variance asymptotique explicite ;
  - si $P = U$, alors $2n(\mathbb{S}(U) - \widehat{S}_n)$ converge en loi vers une combinaison linéaire de variables indépendantes de loi $\chi^2(1)$.
- Girardin et Regnault (2011) établissent un principe de grandes déviations pour la suite $(\widehat{S}_n)_{n \in \mathbb{N}^*}$.

Estimation de l'entropie ; cas indépendant

Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $P \in \mathcal{M}_1(E)$.
Soit $\widehat{P}_n$ la probabilité empirique associée à $(X_1, \ldots, X_n)$ définie par

$$\widehat{P}_n(i) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k = i\}}, \quad i \in E.$$

Soit $\widehat{S}_n = \mathbb{S}(\widehat{P}_n)$ l'estimateur par plug-in de $\mathbb{S}(P)$.

- Basharin (1959) prouve que $\widehat{S}_n$ converge presque sûrement vers $\mathbb{S}(P)$, lorsque n tend vers l'infini.
- Zubkov (1973) et Harris (1977) prouvent que
  - si $P \neq U$, alors $\sqrt{n}(\widehat{S}_n - \mathbb{S}(P))$ est asymptotiquement normal de variance asymptotique explicite ;
  - si $P = U$, alors $2n(\mathbb{S}(U) - \widehat{S}_n)$ converge en loi vers une combinaison linéaire de variables indépendantes de loi $\chi^2(1)$.
- Girardin et Regnault (2011) établissent un principe de grandes déviations pour la suite $(\widehat{S}_n)_{n\in\mathbb{N}^*}$.

Qu'est-ce qu'un principe de grandes déviations ?

Un exemple : Soient $X_1, \ldots, X_n$ indépendantes, de même loi P, d'espérance m, de variance σ^2.

- **LGN :** $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ converge presque sûrement vers m.
- **TLC :** Pour tout $\delta > 0$, $\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}|S_n - m| \geq \delta\right) \simeq \frac{2}{\sqrt{2\pi}}\int_\delta^{+\infty} \exp(-x^2/2)dx$.

Si P est une **loi gaussienne**, on a

$$\mathbb{P}(|S_n - m| \geq \delta) = \frac{2}{\sqrt{2\pi}}\int_{\delta\sqrt{n}}^{+\infty} \exp(-x^2/2\sigma^2)dx,$$

de sorte que (**PGD**)

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(|S_n - m| \geq \delta) = -\frac{\delta^2}{2\sigma^2},$$

soit, de manière heuristique, $\mathbb{P}(|S_n - m| \geq \delta) \overset{n\to\infty}{\simeq} \exp(-n\delta^2/2\sigma^2)$.
Si P n'est **pas gaussienne**, un PGD subsiste (Cramér, 1936), de la forme

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(|S_n - m| \geq \delta) = -\inf\{I_P(x), x : |x - m| \geq \delta\},$$

où I_P est une fonction réelle dépendant de P, appelée fonction de taux.

Qu'est-ce qu'un principe de grandes déviations ?

Un exemple : Soient $X_1, \ldots, X_n$ indépendantes, de même loi P, d'espérance m, de variance σ^2.

- **LGN :** $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ converge presque sûrement vers m.
- **TLC :** Pour tout $\delta > 0$, $\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}|S_n - m| \geq \delta\right) \simeq \frac{2}{\sqrt{2\pi}}\int_\delta^{+\infty} \exp(-x^2/2)dx$.

Si P est une **loi gaussienne**, on a

$$\mathbb{P}\left(|S_n - m| \geq \delta\right) = \frac{2}{\sqrt{2\pi}}\int_{\delta\sqrt{n}}^{+\infty} \exp(-x^2/2\sigma^2)dx,$$

de sorte que (**PGD**)

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left(|S_n - m| \geq \delta\right) = -\frac{\delta^2}{2\sigma^2},$$

soit, de manière heuristique, $\mathbb{P}\left(|S_n - m| \geq \delta\right) \overset{n\to\infty}{\simeq} \exp(-n\delta^2/2\sigma^2)$.

Si P n'est **pas gaussienne**, un PGD subsiste (Cramér, 1936), de la forme

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left(|S_n - m| \geq \delta\right) = -\inf\{I_P(x), x : |x - m| \geq \delta\},$$

où I_P est une fonction réelle dépendant de P, appelée fonction de taux.

Qu'est-ce qu'un principe de grandes déviations ?

Un exemple : Soient $X_1, \ldots, X_n$ indépendantes, de même loi P, d'espérance m, de variance σ^2.

- **LGN :** $S_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ converge presque sûrement vers m.
- **TLC :** Pour tout $\delta > 0$, $\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}|S_n - m| \geq \delta\right) \simeq \frac{2}{\sqrt{2\pi}} \int_{\delta}^{+\infty} \exp(-x^2/2)dx.$

Si P est une **loi gaussienne**, on a

$$\mathbb{P}\left(|S_n - m| \geq \delta\right) = \frac{2}{\sqrt{2\pi}} \int_{\delta\sqrt{n}}^{+\infty} \exp(-x^2/2\sigma^2)dx,$$

de sorte que (**PGD**)

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(|S_n - m| \geq \delta\right) = -\frac{\delta^2}{2\sigma^2},$$

soit, de manière heuristique, $\mathbb{P}\left(|S_n - m| \geq \delta\right) \overset{n\to\infty}{\simeq} \exp(-n\delta^2/2\sigma^2)$.
Si P n'est **pas gaussienne**, un PGD subsiste (Cramér, 1936), de la forme

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(|S_n - m| \geq \delta\right) = -\inf\{I_P(x), x : |x - m| \geq \delta\},$$

où I_P est une fonction réelle dépendant de P, appelée fonction de taux.

Le PGD de Sanov et le principe de contraction

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires à valeurs dans E fini, indépendantes, de loi P.
Soit $\widehat{P}_n$ la **loi empirique** associée à $X_1, \dots, X_n$.

Théorème (Sanov, 1961)

Pour toute partie B de $\mathcal{M}_1(E)$ d'intérieur non vide, on a

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\widehat{P}_n \in B) = - \inf_{Q\in B} \mathbb{K}(Q|P).$$

Principe de contraction : Si f est une fonction continue de $\mathcal{M}_1(E)$ dans $\mathbb{R}$, alors pour tout borélien A de $\mathbb{R}$, d'intérieur non vide, on a

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(f(\widehat{P}_n) \in I) = - \inf_{x\in I} \inf\{\mathbb{K}(Q|P) : f(Q) = x\}.$$

Le principe de contraction, appliqué pour $f = \mathbb{S}$, fournit un principe de grandes déviations pour la suite des estimateurs par plug-in $\widehat{S}_n = \mathbb{S}(\widehat{P}_n)$ de $\mathbb{S}(P)$.

Le PGD de Sanov et le principe de contraction

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires à valeurs dans E fini, indépendantes, de loi P.
Soit $\widehat{P}_n$ la **loi empirique** associée à $X_1, \dots, X_n$.

Théorème (Sanov, 1961)

Pour toute partie B de $\mathcal{M}_1(E)$ d'intérieur non vide, on a

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\widehat{P}_n \in B) = - \inf_{Q\in B} \mathbb{K}(Q|P).$$

Principe de contraction : Si f est une fonction continue de $\mathcal{M}_1(E)$ dans $\mathbb{R}$, alors pour tout borélien A de $\mathbb{R}$, d'intérieur non vide, on a

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(f(\widehat{P}_n) \in I) = - \inf_{x\in I} \inf\{\mathbb{K}(Q|P) : f(Q) = x\}.$$

Le principe de contraction, appliqué pour $f = \mathbb{S}$, fournit un principe de grandes déviations pour la suite des estimateurs par plug-in $\widehat{S}_n = \mathbb{S}(\widehat{P}_n)$ de $\mathbb{S}(P)$.

Le PGD de Sanov et le principe de contraction

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires à valeurs dans E fini, indépendantes, de loi P.
Soit $\widehat{P}_n$ la **loi empirique** associée à $X_1, \dots, X_n$.

Théorème (Sanov, 1961)

Pour toute partie B de $\mathcal{M}_1(E)$ d'intérieur non vide, on a

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\widehat{P}_n \in B) = - \inf_{Q\in B} \mathbb{K}(Q|P).$$

Principe de contraction : Si f est une fonction continue de $\mathcal{M}_1(E)$ dans $\mathbb{R}$, alors pour tout borélien A de $\mathbb{R}$, d'intérieur non vide, on a

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(f(\widehat{P}_n) \in I) = - \inf_{x\in I} \inf\{\mathbb{K}(Q|P) : f(Q) = x\}.$$

Le principe de contraction, appliqué pour $f = \mathbb{S}$, fournit un principe de grandes déviations pour la suite des estimateurs par plug-in $\widehat{S}_n = \mathbb{S}(\widehat{P}_n)$ de $\mathbb{S}(P)$.

Le PGD pour l'estimateur par plug-in de l'entropie

Théorème (Girardin et Regnault, 2011)

Pour tout borélien A de $\mathbb{R}$,

$$\mathbb{P}\left(\widehat{S}_n \in A\right) \leq (n+1)^{|E|+1} \exp\left(-n \inf_{s\in\overline{A}} I_{\mathbb{S}}(P,s)\right);$$

de plus, si A est d'intérieur non vide,

$$\lim_{n\to+\infty} \frac{1}{n} \log \mathbb{P}\left(\widehat{S}_n \in A\right) = -\inf_{s\in\overline{A}} I_{\mathbb{S}}(P,s),$$

avec

$$I_{\mathbb{S}}(P,s) = \begin{cases} \inf\{\mathbb{K}(Q|P) : \mathbb{S}(Q) = s\} & \text{si } s \in [0, \log|E|], \\ +\infty & \text{sinon.} \end{cases}$$

En particulier, pour $s \in [0 \log|E|]$,
$I_{\mathbb{S}}(P,s) = \inf\{\mathbb{K}(Q|P) : \mathbb{K}(Q|U) = \log|E| - s\}$ est la distance, au sens de la divergence de Kullback-Leibler, de P à la sphère centrée en U, de rayon $\log|E| - s$.

La transformation escorte

Définition

La transformation escorte est l'application

$$\begin{aligned} \mathcal{E} \quad : \quad & \mathbb{R} \times \mathcal{M}_1(E) \to \mathcal{M}_1(E) \\ & (k, P) \mapsto E_P^k : i \mapsto P(i)^k / \sum_{j \in E} P(j)^k, \end{aligned}$$

Étant donnée $P \in \mathcal{M}_1(E)$, les k-transformations E_P^k de P sont appelées lois escortes de P.

Les lois escortes sont introduites par Beck et Schlögl (1993), indépendamment de toute considération géometrique, en théorie du chaos, afin d'étudier les propriétés thermodynamiques de systèmes microscopiques.

Nouveau point de vue : Les lois escortes de P sont les homothétiques de P dans la géométrie de l'information.
Elles donnent la direction de la plus grande variation d'information pour P.

Projection au sens de Kullback-Leibler

Soit $P \in \mathcal{M}_1^*(E)$.
Soient $p^* = \max_{i \in E} P(i)$ et
$m = |\{i \in E : P(i) = p^*\}|$ le nombre de modes de P.
Soient $\mathbb{S}(P)$ l'entropie de P et $s \in \mathbb{R}$.

Théorème (Girardin et Regnault, 2011)

$$I_{\mathbb{S}}(P, s) = \begin{cases} \mathbb{K}(E_P^k | P) & \text{si } s \in]\log m, \log |E|], \\ -\log p^* - s & \text{si } s \in [0, \log m], \\ +\infty & \text{sinon,} \end{cases}$$

avec $k > 0$ tel que $\mathbb{S}(E_P^k) = s$.

Projection au sens de Kullback-Leibler

Soit $P \in \mathcal{M}_1^*(E)$.
Soient $p^* = \max_{i \in E} P(i)$ et
$m = |\{i \in E : P(i) = p^*\}|$ le nombre de modes de P.
Soient $\mathbb{S}(P)$ l'entropie de P et $s \in \mathbb{R}$.

Théorème (Girardin et Regnault, 2011)

$$I_{\mathbb{S}}(P, s) = \begin{cases} \mathbb{K}(E_P^k | P) & \text{si } s \in]\log m, \log |E|], \\ -\log p^* - s & \text{si } s \in [0, \log m], \\ +\infty & \text{sinon}, \end{cases}$$

avec $k > 0$ tel que $\mathbb{S}(E_P^k) = s$.

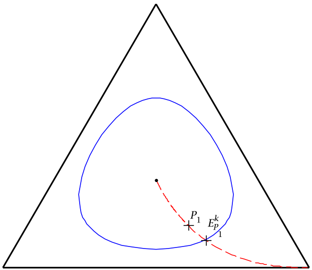


Fig.: $\mathbb{K}$-projection sur la shpère $\mathcal{S}_{\mathbb{K}}(\log(3) - s)$, pour $s > \log(2)$.

Projection au sens de Kullback-Leibler

Soit $P \in \mathcal{M}_1^*(E)$.
Soient $p^* = \max_{i \in E} P(i)$ et $m = |\{i \in E : P(i) = p^*\}|$ le nombre de modes de P.
Soient $\mathbb{S}(P)$ l'entropie de P et $s \in \mathbb{R}$.

Théorème (Girardin et Regnault, 2011)

$$I_{\mathbb{S}}(P,s) = \begin{cases} \mathbb{K}(E_P^k|P) & \text{si } s \in]\log m, \log |E||, \\ -\log p^* - s & \text{si } s \in [0, \log m], \\ +\infty & \text{sinon,} \end{cases}$$

avec $k > 0$ tel que $\mathbb{S}(E_P^k) = s$.

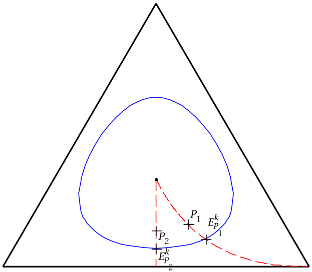


Fig.: $\mathbb{K}$-projection sur la shpère $\mathcal{S}_{\mathbb{K}}(\log(3) - s)$, pour $s > \log(2)$.

Projection au sens de Kullback-Leibler

Soit $P \in \mathcal{M}_1^*(E)$.
Soient $p^* = \max_{i \in E} P(i)$ et
$m = |\{i \in E : P(i) = p^*\}|$ le nombre de modes de P.
Soient $\mathbb{S}(P)$ l'entropie de P et $s \in \mathbb{R}$.

Théorème (Girardin et Regnault, 2011)

$$I_{\mathbb{S}}(P, s) = \begin{cases} \mathbb{K}(E_P^k | P) & \text{si } s \in]\log m, \log |E|], \\ -\log p^* - s & \text{si } s \in [0, \log m], \\ +\infty & \text{sinon,} \end{cases}$$

avec $k > 0$ tel que $\mathbb{S}(E_P^k) = s$.

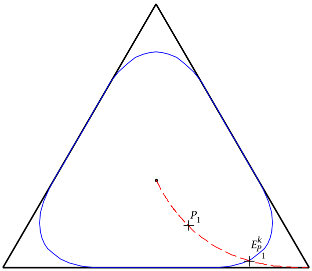


Fig.: $\mathbb{K}$-projection sur la shpère $\mathcal{S}_{\mathbb{K}}(\log(3) - s)$, pour $s < \log(2)$.

Projection au sens de Kullback-Leibler

Soit $P \in \mathcal{M}_1^*(E)$.
Soient $p^* = \max_{i \in E} P(i)$ et
$m = |\{i \in E : P(i) = p^*\}|$ le nombre de modes de P.
Soient $\mathbb{S}(P)$ l'entropie de P et $s \in \mathbb{R}$.

Théorème (Girardin et Regnault, 2011)

$$I_{\mathbb{S}}(P, s) = \begin{cases} \mathbb{K}(E_P^k | P) & \text{si } s \in]\log m, \log |E|], \\ -\log p^* - s & \text{si } s \in [0, \log m], \\ +\infty & \text{sinon,} \end{cases}$$

avec $k > 0$ tel que $\mathbb{S}(E_P^k) = s$.

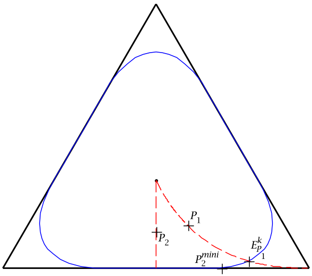


Fig.: $\mathbb{K}$-projection sur la shpère $\mathcal{S}_{\mathbb{K}}(\log(3) - s)$, pour $s < \log(2)$.

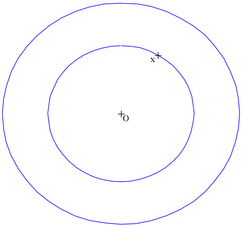
Comparaison entre géométries euclidienne et de l'information

Géométrie euclidienne

Soit $x \in (\mathcal{X}, \langle ., . \rangle)$ non nul.
Soit $r > 0$. On a

$$argmin\{\|y - x\| : \|y\| = r\} = k.x,$$

avec $k = \frac{r}{\|x\|}$.

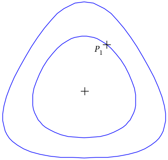


Géométrie de l'information

Soit $P \in \mathcal{M}_1(E)$, $P \neq U$.
Soit $s \in]0, \log(|E|)[$. On a

$$argmin\{\mathbb{K}(Q|P) : \mathbb{S}(Q) = s\} = E_P^k,$$

avec $k > 0$ tel que $\mathbb{S}(E_P^k) = s$.



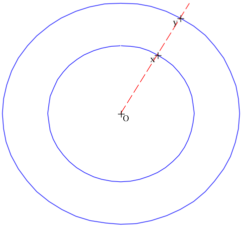
Comparaison entre géométries euclidienne et de l'information

Géométrie euclidienne

Soit $x \in (\mathcal{X}, \langle ., . \rangle)$ non nul.
Soit $r > 0$. On a

$$argmin\{\|y - x\| : \|y\| = r\} = k.x,$$

avec $k = \frac{r}{\|x\|}$.

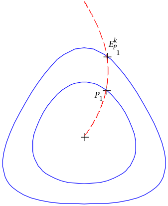


Géométrie de l'information

Soit $P \in \mathcal{M}_1(E)$, $P \neq U$.
Soit $s \in]0, \log(|E|)[$. On a

$$argmin\{\mathbb{K}(Q|P) : \mathbb{S}(Q) = s\} = E_P^k,$$

avec $k > 0$ tel que $\mathbb{S}(E_P^k) = s$.



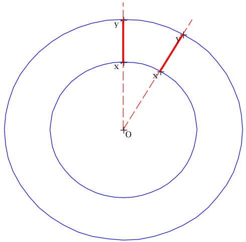
Comparaison entre géométries euclidienne et de l'information

Géométrie euclidienne

Soit $x \in (\mathcal{X}, \langle ., . \rangle)$ non nul.
Soit $r > 0$. On a

$$argmin\{\|y - x\| : \|y\| = r\} = k.x,$$

avec $k = \frac{r}{\|x\|}$.



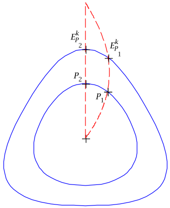
En particulier, k ne dépend de x qu'à travers $\|x\|$.

Géométrie de l'information

Soit $P \in \mathcal{M}_1(E)$, $P \neq U$.
Soit $s \in]0, \log(|E|)[$. On a

$$argmin\{\mathbb{K}(Q|P) : \mathbb{S}(Q) = s\} = E_P^k,$$

avec $k > 0$ tel que $\mathbb{S}(E_P^k) = s$.



$k = k(P, s)$ dépend de la direction de la projection ($k \neq s/\mathbb{S}(P)$).

Quelques propriétés de la transformation escorte

Déterminer une expression analytique pour $I_{\mathbb{S}}(P,.)$ ne semble pas faisable.

Construire une approximation numérique de $I_{\mathbb{S}}(P,.)$ nécessite quelques propriétés de la transformation escorte.

Soit $P \in \mathcal{M}_1^*(E)$, $P \neq U$. Soient $\alpha_1, \dots, \alpha_m$ les modes de P.

Propriétés

- E_P^k converge vers $\sum_{l=1}^m \delta_{\alpha_l}/m$ lorsque k tend vers l'infini.
- E_P^k converge vers la probabilité uniforme sur E lorsque k tend vers 0 ;
- La fonction $\Phi_P : k \in \mathbb{R}_+ \mapsto \mathbb{S}(E_P^k)$ est deux fois continûment dérivable et $\Phi_P'(k) < 0$, pour tout $k > 0$.

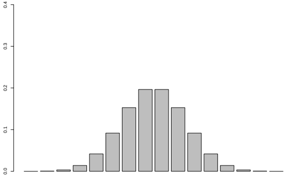


Fig.: Loi Binomiale $\mathbb{B}(15, 1/2)$

$I_{\mathbb{S}}(P,s) = \mathbb{K}(\mathcal{E}(P,\Phi_P^{-1}(s)|P)$, $\quad s \in]\log(m), \log|E|[$ est dérivable par rapport à s.

Quelques propriétés de la transformation escorte

Déterminer une expression analytique pour $I_{\mathbb{S}}(P,.)$ ne semble pas faisable.

Construire une approximation numérique de $I_{\mathbb{S}}(P,.)$ nécessite quelques propriétés de la transformation escorte.

Soit $P \in \mathcal{M}_1^*(E)$, $P \neq U$. Soient $\alpha_1, \dots, \alpha_m$ les modes de P.

Propriétés

- E_P^k converge vers $\sum_{l=1}^m \delta_{\alpha_l}/m$ lorsque k tend vers l'infini.
- E_P^k converge vers la probabilité uniforme sur E lorsque k tend vers 0 ;
- La fonction $\Phi_P : k \in \mathbb{R}_+ \mapsto \mathbb{S}(E_P^k)$ est deux fois continûment dérivable et $\Phi_P'(k) < 0$, pour tout $k > 0$.

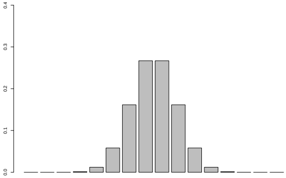


Fig.: 2-escorte de $\mathbb{B}(15, 1/2)$

$I_{\mathbb{S}}(P,s) = \mathbb{K}(\mathcal{E}(P,\Phi_P^{-1}(s)|P)$, $\quad s \in]\log(m), \log|E|[$ est dérivable par rapport à s.

Quelques propriétés de la transformation escorte

Déterminer une expression analytique pour $I_{\mathbb{S}}(P,.)$ ne semble pas faisable.

Construire une approximation numérique de $I_{\mathbb{S}}(P,.)$ nécessite quelques propriétés de la transformation escorte.

Soit $P \in \mathcal{M}_1^*(E)$, $P \neq U$. Soient $\alpha_1, \dots, \alpha_m$ les modes de P.

Propriétés

- E_P^k converge vers $\sum_{l=1}^m \delta_{\alpha_l}/m$ lorsque k tend vers l'infini.
- E_P^k converge vers la probabilité uniforme sur E lorsque k tend vers 0 ;
- La fonction $\Phi_P : k \in \mathbb{R}_+ \mapsto \mathbb{S}(E_P^k)$ est deux fois continûment dérivable et $\Phi_P'(k) < 0$, pour tout $k > 0$.

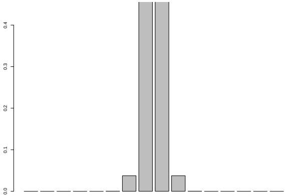


Fig.: 10-escorte de $\mathbb{B}(15, 1/2)$

$I_{\mathbb{S}}(P,s) = \mathbb{K}(\mathcal{E}(P,\Phi_P^{-1}(s)|P), \quad s \in]\log(m), \log|E|[$ est dérivable par rapport à s.

Quelques propriétés de la transformation escorte

Déterminer une expression analytique pour $I_{\mathbb{S}}(P,.)$ ne semble pas faisable.

Construire une approximation numérique de $I_{\mathbb{S}}(P,.)$ nécessite quelques propriétés de la transformation escorte.

Soit $P \in \mathcal{M}_1^*(E)$, $P \neq U$. Soient $\alpha_1, \dots, \alpha_m$ les modes de P.

Propriétés

- E_P^k converge vers $\sum_{l=1}^m \delta_{\alpha_l}/m$ lorsque k tend vers l'infini.
- E_P^k converge vers la probabilité uniforme sur E lorsque k tend vers 0 ;
- La fonction $\Phi_P : k \in \mathbb{R}_+ \mapsto \mathbb{S}(E_P^k)$ est deux fois continûment dérivable et $\Phi_P'(k) < 0$, pour tout $k > 0$.

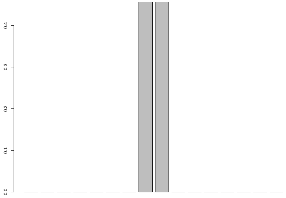


Fig.: 100-escorte de $\mathbb{B}(15, 1/2)$

$I_{\mathbb{S}}(P,s) = \mathbb{K}(\mathcal{E}(P,\Phi_P^{-1}(s)|P)$, $\quad s \in]\log(m), \log|E|[$ est dérivable par rapport à s.

Quelques propriétés de la transformation escorte

Déterminer une expression analytique pour $I_{\mathbb{S}}(P,.)$ ne semble pas faisable.

Construire une approximation numérique de $I_{\mathbb{S}}(P,.)$ nécessite quelques propriétés de la transformation escorte.

Soit $P \in \mathcal{M}_1^*(E)$, $P \neq U$. Soient $\alpha_1, \ldots, \alpha_m$ les modes de P.

Propriétés

- E_P^k converge vers $\sum_{l=1}^m \delta_{\alpha_l}/m$ lorsque k tend vers l'infini.
- E_P^k converge vers la probabilité uniforme sur E lorsque k tend vers 0 ;
- La fonction $\Phi_P : k \in \mathbb{R}_+ \mapsto \mathbb{S}(E_P^k)$ est deux fois continûment dérivable et $\Phi_P'(k) < 0$, pour tout $k > 0$.

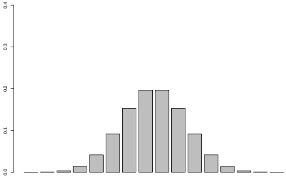


Fig.: Loi binomiale $\mathbb{B}(15, 1/2)$

$I_{\mathbb{S}}(P,s) = \mathbb{K}(\mathcal{E}(P,\Phi_P^{-1}(s)|P)$, $\quad s \in]\log(m), \log|E|[$ est dérivable par rapport à s.

Quelques propriétés de la transformation escorte

Déterminer une expression analytique pour $I_{\mathbb{S}}(P,.)$ ne semble pas faisable.

Construire une approximation numérique de $I_{\mathbb{S}}(P,.)$ nécessite quelques propriétés de la transformation escorte.

Soit $P \in \mathcal{M}_1^*(E)$, $P \neq U$. Soient $\alpha_1, \dots, \alpha_m$ les modes de P.

Propriétés

- E_P^k converge vers $\sum_{l=1}^m \delta_{\alpha_l}/m$ lorsque k tend vers l'infini.
- E_P^k converge vers la probabilité uniforme sur E lorsque k tend vers 0 ;
- La fonction $\Phi_P : k \in \mathbb{R}_+ \mapsto \mathbb{S}(E_P^k)$ est deux fois continûment dérivable et $\Phi_P'(k) < 0$, pour tout $k > 0$.

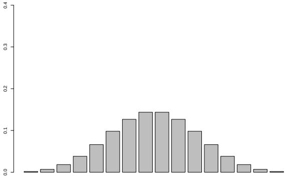


Fig.: $0,5$-escorte de $\mathbb{B}(15, 1/2)$

$I_{\mathbb{S}}(P,s) = \mathbb{K}(\mathcal{E}(P, \Phi_P^{-1}(s)|P)$, $\quad s \in]\log(m), \log|E|[$ est dérivable par rapport à s.

Quelques propriétés de la transformation escorte

Déterminer une expression analytique pour $I_{\mathbb{S}}(P,.)$ ne semble pas faisable.

Construire une approximation numérique de $I_{\mathbb{S}}(P,.)$ nécessite quelques propriétés de la transformation escorte.

Soit $P \in \mathcal{M}_1^*(E)$, $P \neq U$. Soient $\alpha_1, \dots, \alpha_m$ les modes de P.

Propriétés

- E_P^k converge vers $\sum_{l=1}^m \delta_{\alpha_l}/m$ lorsque k tend vers l'infini.
- E_P^k converge vers la probabilité uniforme sur E lorsque k tend vers 0 ;
- La fonction $\Phi_P : k \in \mathbb{R}_+ \mapsto \mathbb{S}(E_P^k)$ est deux fois continûment dérivable et $\Phi_P'(k) < 0$, pour tout $k > 0$.

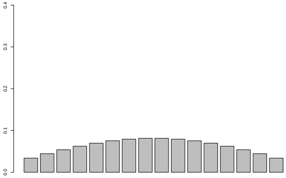


Fig.: $0,1$-escorte de $\mathbb{B}(15,1/2)$

$I_{\mathbb{S}}(P,s) = \mathbb{K}(\mathcal{E}(P,\Phi_P^{-1}(s)|P)$, $\quad s \in]\log(m), \log|E|[$ est dérivable par rapport à s.

Quelques propriétés de la transformation escorte

Déterminer une expression analytique pour $I_{\mathbb{S}}(P,.)$ ne semble pas faisable.

Construire une approximation numérique de $I_{\mathbb{S}}(P,.)$ nécessite quelques propriétés de la transformation escorte.

Soit $P \in \mathcal{M}_1^*(E)$, $P \neq U$. Soient $\alpha_1, \dots, \alpha_m$ les modes de P.

Propriétés

- E_P^k converge vers $\sum_{l=1}^m \delta_{\alpha_l}/m$ lorsque k tend vers l'infini.
- E_P^k converge vers la probabilité uniforme sur E lorsque k tend vers 0 ;
- La fonction $\Phi_P : k \in \mathbb{R}_+ \mapsto \mathbb{S}(E_P^k)$ est deux fois continûment dérivable et $\Phi_P'(k) < 0$, pour tout $k > 0$.

Fig.: $0,01$-escorte de $\mathbb{B}(15, 1/2)$

$I_{\mathbb{S}}(P,s) = \mathbb{K}(\mathcal{E}(P,\Phi_P^{-1}(s)|P)$, $\quad s \in]\log(m), \log|E|[$ est dérivable par rapport à s.

Quelques propriétés de la transformation escorte

Déterminer une expression analytique pour $I_{\mathbb{S}}(P,.)$ ne semble pas faisable.

Construire une approximation numérique de $I_{\mathbb{S}}(P,.)$ nécessite quelques propriétés de la transformation escorte.

Soit $P \in \mathcal{M}_1^*(E)$, $P \neq U$. Soient $\alpha_1, \dots, \alpha_m$ les modes de P.

Propriétés

- E_P^k converge vers $\sum_{l=1}^m \delta_{\alpha_l}/m$ lorsque k tend vers l'infini.
- E_P^k converge vers la probabilité uniforme sur E lorsque k tend vers 0 ;
- La fonction $\Phi_P : k \in \mathbb{R}_+ \mapsto \mathbb{S}(E_P^k)$ est deux fois continûment dérivable et $\Phi'_P(k) < 0$, pour tout $k > 0$.

$I_{\mathbb{S}}(P,s) = \mathbb{K}(\mathcal{E}(P,\Phi_P^{-1}(s)|P), \quad s \in]\log(m), \log|E|[$ est dérivable par rapport à s.

Approximation numérique de la fonction de taux

$I_{\mathbb{S}}(P,.)$ peut être approximée par la fonction continue, affine par morceaux définie par $I_M(s) =$

$$\begin{cases} \alpha \widetilde{I_{\mathbb{S}}}(s_1) + (1-\alpha)(-\log \max_{i\in E}\{P(i)\} - \log m), & s = \alpha s_1 + (1-\alpha)\log m, \\ \alpha \widetilde{I_{\mathbb{S}}}(s_l) + (1-\alpha)\widetilde{I_{\mathbb{S}}}(s_{l-1}), & s = \alpha s_l + (1-\alpha s_{l-1}). \\ \alpha\big(-\log(|E|) - \frac{1}{|E|}\sum_{i\in E}\log P(i)\big) + (1-\alpha)\widetilde{I_{\mathbb{S}}}(s_{M-1}), & s = \alpha \log|E| + (1-\alpha)s_{M-1}, \end{cases}$$

pour

- un ensemble de points équidistribués $\{s_1, \dots, s_{M-1}\} \in]\log m, \log|E|[$,
- des solutions approchées $\widetilde{k_l}$ des vraies solutions k_l des équations $\mathbb{K}(E_P^{k_l} \mid P) = s_l$,
- des approximations de $I_{\mathbb{S}}(P, s_l)$ obtenues en posant $\widetilde{I_{\mathbb{S}}}(s_l) = \mathbb{K}(E_P^{\widetilde{k_l}}|P)$.

Proposition (Girardin et Regnault, 2011)

Pour tout compact $K \subseteq]\log m, \log |E|[$ et tout $s \in K$, on a

$$|I_M(s) - I_{\mathbb{S}}(P, s)| \leq \frac{2}{M},$$

dès que $M \geq (\log |E| - \log m) \max\{\sup_{s \in K} |I_{\mathbb{S}}^{''}(s)|/8, 1\}$.

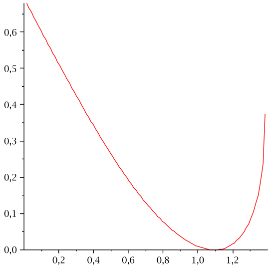


Fig.: Approximation de $I_{\mathbb{S}}(P, .)$ pour $P = [0.5, 0.1, 0.05, 0.35]$.

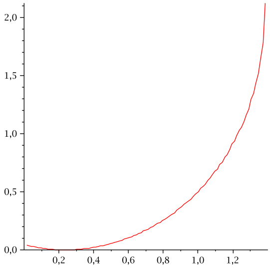


Fig.: Approximation de $I_{\mathbb{S}}(P, .)$ pour $P = [0.95, 0.03, 0.01, 0.01]$.

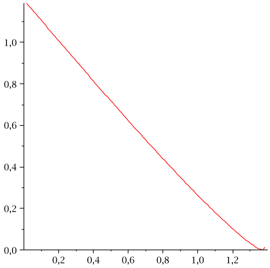


Fig.: Approximation de $I_{\mathbb{S}}(P, .)$ pour $P = [0.2, 0.3, 0.22, 0.28]$.

Applications du PGD

La fonction de taux $I_{\mathbb{S}}(P,.)$ peut être remplacée par son approximation I_M dans les applications impliquant le principe de grandes déviations, sans perte significative de précision.

- Majorations exponentielles explicites de la probabilité des événements rares, comme $(\widehat{S}_n \geq \mathbb{S}(P) + \delta)$ avec $\delta > 0$, asymptotiquement aussi précises que celles fournies par le PGD original.

  Précisément,

$$\mathbb{P}\left(\mathbb{S}(\widehat{P}_n) > \mathbb{S}(P) + \delta\right) \leq (n+1)^{|E|} \exp\left(-n\left[I_M(\mathbb{S}(P)+\delta) - \frac{2}{M}\right]\right).$$

  De même pour $(\mathbb{S}(\widehat{P}_n) \leq \mathbb{S}(P) - \delta)$.

- Tests du niveau d'entropie d'une loi, par seuillage de l'approximation I_M par des valeurs appropriées.

Perspectives

- Étendre le principe de grandes déviations, dans le cas indépendant, aux espaces d'état dénombrables et continus, puis aux chaînes de Markov, aux processus de Markov.
- Poursuivre l'exploration de la géométrie de l'information : généralisation de la projection relativement à d'autres divergences, sur des ensembles contraints autrement.
- Poursuivre l'étude des tests du niveau d'entropie d'une loi.