Apprentissage en ligne et prédiction de séquences individuelles

PESTO 2011

Plan

Présentation

On cherche à prédire séquentiellement

un phénomène (cours de bourse, charge d'électricité, météo)

On n'utilise *aucun modèle* (probabiliste ou autre) sur le phénomène

On s'appuie sur des *experts* plus ou moins fiables

On cherche à faire (au moins) aussi bien que le meilleur expert

Cadre mathématique

Observations $y_1, y_2, \ldots \in \mathcal{Y}$ - on note $\mathbf{y}_t = (y_s)_{s \leq t}$

A l'instant t, l'expert $j \in \{1, \ldots, N\}$ fournit la *prédiction*

$$f_t^j = f_t^j(\mathbf{y}_{t-1}) \in \mathcal{X}$$

où $\mathcal{X}$ est un ensemble pouvant être distinct de $\mathcal{Y}$

On note $\mathbf{f}_t = (f_s^j)_{1 \leq j \leq t, 1 \leq s \leq t}$

La qualité d'une prédiction est quantifiée par la *fonction de perte* $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$

La *perte cumulée* d'une séquence de prédiction $\mathbf{x}_t = (x_1, \ldots, x_n)$ est

$$L_n(\mathbf{x}_t, \mathbf{y}_t) = \sum_{t=1}^{n} \ell(x_t, y_t)$$

Stratégie randomisée

Prédiction séquentielle $\hat{p}_1, \hat{p}_2, \dots \in \mathcal{X}$ telles que :

$$\hat{p}_t = \hat{p}_t(\mathbf{y}_{t-1}, \mathbf{f}_{t-1})$$

Stratégie randomisée :

$$\hat{p}_t = f_t^j \text{ avec probabilité } p_t(j)$$

avec la *pondération* $p_t = (p_t(j))_{1 \leq j \leq N} = p_t(\mathbf{y}_{t-1}, \mathbf{f}_{t-1})$

On veut donc minimiser la perte cumulée du décideur

$$\hat{L}_n(\hat{p}, \mathbf{y}_n) = \sum_{t=1}^{n} \ell(\hat{p}_t, y_t)$$

Regret face au meilleur expert

On cherche à prédire au moins aussi bien que le meilleur expert

On définit le *regret* :

$$R_n(\hat{p}, \mathbf{y}_n) = \max_{1 \leq j \leq N} \hat{L}_n(\hat{p}, \mathbf{y}_n) - L_n(j, \mathbf{y}_n)$$

avec $L_n(j, \mathbf{y}_n) =$ regret cumulé de l'expert j.

Regret dans le pire des cas :

$$R_n(\hat{p}) = \sup_{\mathbf{y}_n \in \mathcal{Y}^n} R_n(\hat{p}, \mathbf{y}_n)$$

But : construire $\hat{p}$ de telle sorte que $\limsup R_n(\hat{p})/n \leq 0$

Théorie des Jeux

La théorie des jeux constitue une approche mathématique de problèmes de stratégie tels qu'on en trouve en recherche opérationnelle et en économie.

Elle étudie les situations où les choix de deux protagonistes — ou davantage — ont des conséquences pour l'un comme pour l'autre.

Elle étudie les comportements — prévus, réels, ou tels que justifiés a posteriori — d'individus face à des situations d'*antagonisme*, et cherche à mettre en évidence des *stratégies optimales*.

Formalisation en théorie des jeux

Cadre : ensemble prédictions $\mathcal{X}$, ensemble d'observations $\mathcal{Y}$, fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ bornée par M
Acteurs : un décideur, un environnement
Déroulement : pour chaque instant $t = 1, 2, \dots$:

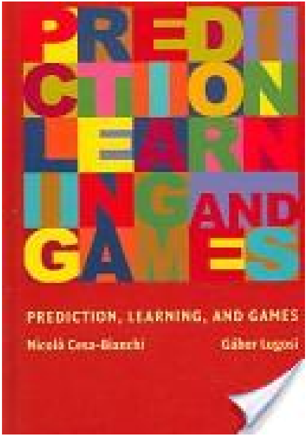
1. les experts publient leurs prédictions $(f_t^j)_{1 \le j \le N}$
2. le décideur choisit une pondération p_t
3. l'environnement choisit une observation y_t
4. le décideur accorde sa confiance à l'expert j avec probabilité $p_t(j)$
5. le décideur découvre l'observation y_t et enregistre les pertes

$$\left(\ell(f_t^j, y_t)\right)_{1 \le j \le N}$$

Références

[Cesa-Bianchi & Lugosi '06]

Prediction, Learning, and Games



[Devroye & Györfi & Lugosi '96] A Probabilistic Theory of Pattern Recognition

[Györfi & Kohler & Krzyzak & Walk '02] A Distribution-Free Theory of Nonparametric Regression

cf. toute la littérature en théorie de l'information : les liens sont plus étroits qu'il n'y paraît

Plan

L'algorithme Exponential Weights (EW)

Stratégie randomisée avec comme choix de pondération :

$$\hat{p}_t(j) = \frac{\exp(-\beta L_{t-1}(j, \mathbf{y}_{t-1}))}{\sum_k \exp(-\beta L_{t-1}(k, \mathbf{y}_{t-1}))}$$

Parfois nommé Hedge, très lié à EXP3 (en feedback bandit).

Borne de regret pour EW

Théorème: Le regret de l'algorithme EW face à la meilleure stratégie constante vérifie :

$$\mathbb{E}\left[R_n(\hat{p})\right] \leq \frac{\log(N)}{\beta} + \frac{M^2\beta}{8}n$$

En particulier, pour $\beta = 1/M\sqrt{8\log(N)/n}$, on obtient :

$$\mathbb{E}\left[R_n(\hat{p})\right] \leq M\sqrt{\frac{n}{2}\log N}$$

Remarque : l'espérance $\mathbb{E}$ porte sur la seule randomisation des choix selon les pondérations

Poursuite du meilleur expert

On cherche à faire au moins aussi bien que *la meilleure séquence d'experts*

$$\mathbf{j}_n = (j_0, \ldots, j_0, j_1, \ldots, j_1, j_2, \ldots, j_2, \ldots, j_k, \ldots, j_k)$$

où $h(\mathbf{j}_n) \stackrel{\text{def}}{=} k$ est le nombre de changements d'experts

On définit le *regret face la meilleure séquence de k experts* :

$$R_n^k(\hat{p}, \mathbf{y}_n) = \max_{h(\mathbf{j}_n) \leq k} \hat{L}_n(\hat{p}, \mathbf{y}_n) - L_n(\mathbf{j}_n, \mathbf{y}_n)$$

Regret dans le pire des cas :

$$R_n^k(\hat{p}) = \sup_{\mathbf{y}_n \in \mathcal{Y}^n} R_n^k(\hat{p}, \mathbf{y}_n)$$

L'algorithme EW.S

On définit par récurrence le poids w_t^j de l'expert j à l'instant t de la façon suivante :

- $\forall j \in \{1, \ldots, N\}, w_0^j = 1/N$
- après avoir vu l'observation y_t, on prend $v_t^j = w_{t-1}^j \exp(-\beta \ell(f_t^j, y_t))$ et
$$w_t^j = (1-\varepsilon)\, v_t^j + \frac{\varepsilon}{N} \sum_j v_t^j$$

L'algorithme EW.S est la stratégie randomisée prenant pour pondération

$$p_t(j) = \frac{w_{t-1}^j}{\sum_k w_{t-1}^k}$$

Borne de regret pour EW.S

Soit $H(p) = -p\log(p) - (1-p)\log(1-p)$ la fonction d'entropie binaire

Théorème: Le regret de la stratégie EW.S vérifie :

$$\mathbb{E}\left[R_n^k(\hat{p})\right] \leq \frac{k+1}{\beta}\log N - \frac{1}{\beta}\log\left(\left(\frac{\varepsilon}{N}\right)^k (1-\varepsilon)^{n-k-1}\right) + \frac{M^2\beta}{8}n$$

En particulier, en choisissant $\varepsilon = k/(n-1)$ et
$\beta = \sqrt{8/n\,((k+1)\log(N) + (n-1)h(k/(n-1)))}$, on a :

$$\mathbb{E}\left[R_n^k(\hat{p})\right] \leq M\sqrt{\frac{n}{2}\left((k+1)\log(N) + (n-1)H\left(\frac{k}{n-1}\right)\right)}$$
$$\leq M\sqrt{\frac{n}{2}\left((k+1)\log(N) + k\log\left(\frac{en}{k}\right)\right)}$$

Variante : Agrégation d'experts

Prédiction séquentielle $\hat{p}_1, \hat{p}_2, \dots \in \mathcal{X}$ *convexe* telles que :

$$\hat{p}_t = \hat{p}_t(\mathbf{y}_{t-1}, \mathbf{f}_{t-1})$$

Stratégie d'agrégation :

$$\hat{p}_t = \sum_{j=1}^{N} p_t(j) f_t^j$$

avec la *pondération* $p_t = (p_t(j))_{1 \le j \le N} = p_t(\mathbf{y}_{t-1}, \mathbf{f}_{t-1})$

On veut donc minimiser la perte cumulée du décideur

$$\hat{L}_n = \sum_{t=1}^{n} \ell(\hat{p}_t, y_t)$$

Formalisation en théorie des jeux

Cadre : ensemble *convexe* de prédictions $\mathcal{X}$, ensemble d'observations $\mathcal{Y}$, fonction de perte *convexe* $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ bornée par M

Acteurs : un décideur, un environnement

Déroulement : pour chaque instant $t = 1, 2, \dots$:

- on recueille les expertises $(f_t^j)_{1 \leq j \leq N}$
- le décideur choisit une pondération p_t et prédit : $\hat{p}_t = \sum_{j=1}^N p_t(j) f_t^j$
- l'environnement choisit *simultanément* une observation y_t
- le décideur découvre l'observation y_t et enregistre les pertes $\left(\ell(f_t^j, y_t)\right)_{1 \leq j \leq N}$

Objectifs de l'agrégation

Pour toute combinaison convexe q, on définit son regret :

$$L_n(q) = \sum_{t=1}^{n} \ell\left(\sum_{j=1}^{N} q_j f_t^j, y_t\right)$$

Dans ce cadre, le regret du décideur est défini par rapport à la pondération q est :

$$R_n(\hat{p}, q) = \sup_q \hat{L}_n(\hat{p}, \mathbf{y}_n) - L_n(q, \mathbf{y}_n)$$

Objectif du décideur : construire des stratégies de prédiction pour lesquelles, quelle que soit la suite d'observations $y_1, y_2, \ldots$, le décideur fait mieux que la *meilleure combinaison convexe d'experts* :

$$\limsup \frac{1}{n} \max_q R_n(\hat{p}, q) \leq 0$$

L'algorithme EW

Théorème: La pondération définie par

$$\hat{p}_t(j) = \frac{\exp(-\beta L_{t-1}(j, \mathbf{y}_{t-1}))}{\sum_k \exp(-\beta L_{t-1}(k, \mathbf{y}_{t-1}))}$$

est telle que :

$$R_n(\hat{p}) \leq \frac{\log(N)}{\beta} + \frac{M^2\beta}{8}n$$

En particulier, pour $\beta = 1/M\sqrt{8\log(N)/m}$ le regret face à la meilleure pondération convexe vérifie :

$$R_n(\hat{p}) \leq M\sqrt{\frac{n}{2}\log N}$$

Remarques sur ce cadre

Cadre *méta-statistique* : chaque expert peut avoir son modèle...

L'*agrégation* fait souvent mieux que la *sélection*

Ici, on s'intéresse à de l'*agrégation séquentielle*

La formulation choisie requiert des stratégies *robustes*

Plan

Prévision de charge

Cf thèse de Yannig Goude (EDF & Université Paris-Sud)
Une problématique cruciale de la production électrique est la *prévision de charge*

On dispose de *plusieurs modèles* plus ou moins évolués / robustes / expérimentés

Le but est d'agréger leurs résultats en les utilisant comme des emphboîtes noires

Prédiction de la qualité de l'air

Cf Gilles Stoltz (CNRS) et Vivient Mallet (INRIA), Journal of Geophysical Research

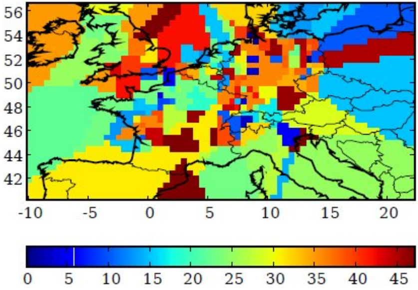
Objectif : prédire, jour après jour, les hauteurs des pics d'ozone du lendemain (ou les concentrations horaires, heure après heure)

Moyens : réseau de stations météorologiques à travers l'Europe (été 2001)

Experts : 48 prédicteurs fondamentaux, chacun construit à partir

- d'un modèle physico-chimique
- d'un schéma numérique de résolution approché des EDP en jeu
- d'un jeu de données

Tous les experts sont utiles



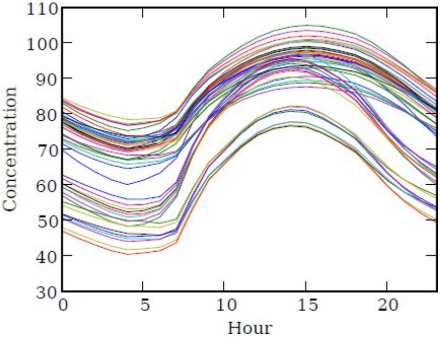


Figure: A gauche : Coloration de l'Europe en fonction de l'indice du meilleur expert local. A droite : Profils moyens de prédiction sur une journée (moyennes spatiales et temporelles, en $\mu g/m^3$)

Résultats

Moyenne	M. fondamental	M. convexe	M. linéaire	EW	
24.41	22.43	21.45	19.24	11.99	21.47

Ci-dessus, les erreurs quadratiques moyennes (en $\mu g/m^3$)

- de la moyenne des prédictions des 48 modèles
- du meilleur modèle fondamental parmi les 48
- de la meilleure combinaison convexe
- de la meilleure combinaison linéaire
- de l'algorithme EW

$\Longrightarrow$ la meilleure combinaison convexe constante est battue

on peut faire mieux (voir Stoltz & Mallet)

Gestion de portefeuilles

[Cover '91] Universal Portfolios
[Györfi & Urban & Vajda '07] Kernel-based semi-log-optimal portfolio selection strategies

N actifs $\{1, \dots, N\}$, prix Z_t^j

Market Vector x_t = vecteur d'évolution des prix $x_t^j = Z_t^j / Z_{t-1}^j$

Position Q_t = vecteur des fractions d'investissement $Q_t^j =$ part du patrimoine dans l'actif j à l'instant t.

Wealth Factor = rendement du placement

$$S_n(Qn, \mathbf{x}_n) = \prod_{t=1}^{n} \left(\sum_{j=1}^{N} x_t^j Q_t^j \right)$$

Plan

Prédiction de séquences individuelles

- Objectifs
- Résultats
- Exemples d'applications

Filtrage pour HMM discrets

- Chaînes de Markov Cachées
- Filtrage

Application à la prédiction de séquences individuelles

- Construction d'un modèle instrumental
- Analyse de la performance
- Optimalité, Extensions, Discussion
- Ouverture sur les problèmes de bandits

Chaînes de Markov

Système dynamique à temps discret $(X_t)_{t\geq 0}$, où $X_t \in \mathcal{X}$ espace d'états (fini)

Propriété de Markov:

$$P(X_{t+1} = x | X_0, \dots, X_t) = P(X_{t+1} = x | X_t)$$

Toute l'information pertinente pour la prédiction du futur est contenue dans l'état présent

Une chaîne de Markov *homogène* sur $\mathcal{X}$ est définie par une distribution initiale χ_0 et par un *noyau de transition*

$$p(y|x) = P(X_{t+1} = y | X_t = x)$$

$$P(X_0 = x_0, \dots, X_n = x_n) = \chi(x_0) \prod_{t=1}^{n} m(x_{t-1}, x_t)$$

Chaînes de Markov Cachées

Soit $(X_t)_{t\geq 0}$ une chaîne de Markov

Soit $(V_t)_{t\geq 1}$ une suite de v.a. i.i.d. uniformes sur $[0,1]$

Soit $h_t : \mathcal{X} \times [0,1] \to \mathcal{Y}$ une application (mesurable), et soit

$$Y_t = h_t(X_t, V_t)$$

Le processus $(X,Y)_t$ est appelé *chaîne de Markov cachée*

Omniprésent en traitement du signal (traitement de la parole, déconvolution, robotique, génomique, etc...)

Propriétés

Notations: Pour $n > 0$, on note $\mathbf{X}_n = (X_0, \dots, X_n)$, $\mathbf{x}_n = (x_0, \dots, x_n)$, $\mathbf{Y}_n = (Y_1, \dots, Y_n)$, $\mathbf{y}_n = (y_1, \dots, y_n)$

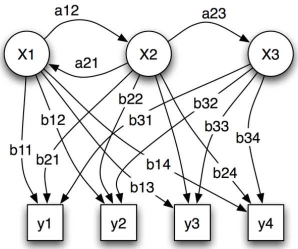
On note $g_t(x, dy) = P(Y_t \in dy | X_t = x)$ la *densité d'émission*

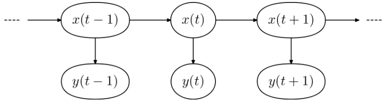
Alors $((X_t, Y_t))_{t \geq 1}$ est une CdM et

$$P(\mathbf{X}_n = \mathbf{x}_n, \mathbf{Y}_n = \mathbf{y}_n) = \chi(x_0) \prod_{t=1}^{n} m(x_{t-1}, x_t) g_t(x_t, y_t)$$

Problème classique : seule $(Y_t)_{t \geq 1}$ est observé, $(X_t)_{t \geq 0}$ est caché et doit être retrouvé.

Chaînes de Markov Cachées : modèle graphique





$$\begin{cases} X_t &= f_n(X_{t-1}, U_t) \\ Y_t &= h_n(X_t, V_t) \end{cases}$$

Vraisemblance

La densité de l'observation $\mathbf{y}_n$ s'écrit :

$$\Lambda_n(\mathbf{y}_n) = \sum_{\mathbf{x}_n \in \mathcal{X}^{n+1}} P(\mathbf{X} = \mathbf{x}) \prod_{t=1}^{n} g(x_t, y_t)$$

$$= \sum_{x_{0:n} \in \mathcal{X}^{n+1}} \chi(x_0) \prod_{t=1}^{n} m(x_{t-1}, x_t) g(x_t, y_t)$$

$$= \sum_{x \in X} P\left(X_n = x \middle| \mathbf{Y}_{n-1} = \mathbf{y}_{n-1}\right) g(x, y_n)$$

Pas d'expression close exploitable, mais un algorithme récursif pour la calculer

Plan

Filtrage

Fixons une observation $\mathbf{Y}_n = \mathbf{y}_n$

Le **filtre** $\phi_t \in \mathcal{M}_1(\mathcal{X})$ est défini pour $0 \leq t \leq n$ par :

$$\phi_t(x) = P(X_t = x_t | Y_{1:t} = y_{1:t})$$

Propriété de récurrence : $\phi_0(x) = \chi$ et pour $t \geq 1$:

$$\phi_t(x') = \frac{1}{\gamma_t} \sum_{x \in \mathcal{X}} \phi_{t-1}(x) m(x, x') g_t(x', y_t)$$

où

$$\gamma_t = P(Y_t = y_t | Y_{1:t-1} = y_{1:t-1}) = \sum_{x, x' \in \mathcal{X}} \phi_{t-1}(x) m(x, x') g_t(x', y_t)$$

Prédiction, Lissage, etc...

Filtre de prédiction :

$$\phi_{t+1|t}(x) \stackrel{\text{def}}{=} P(X_{t+1} = x | Y_{1:t} = y_{1:t}) = \sum_{y \in \mathcal{X}} \phi_t(y) m(y, x)$$

Lissage : $\phi_{t|n}(x_t) \stackrel{\text{def}}{=} P(X_t = x_t | Y_{1:n} = y_{1:n})$

Chemin de Viterbi :

$$\hat{\mathbf{X}}_n = \underset{x_{0:n} \in \mathcal{X}^{n+1}}{\text{argmax}}\, P(X_{0:n} = x_{0:n} | Y_{1:t} = y_{1:t})$$

$\Longrightarrow$ tous ces problèmes ont une solution en $O(n|\mathcal{X}|^2)$ par la programmation dynamique

Exemple : suivi d'un robot

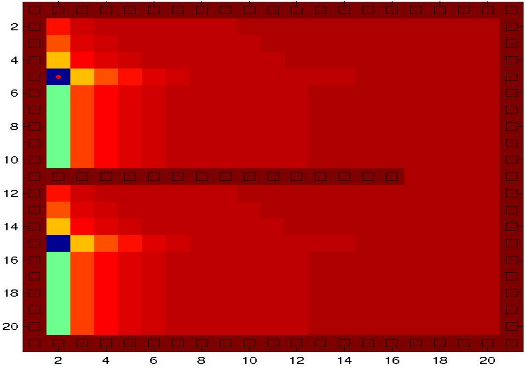
Le robot connaît la configuration de la pièce.

L'espace est discrétisé en petites cases rectangulaires

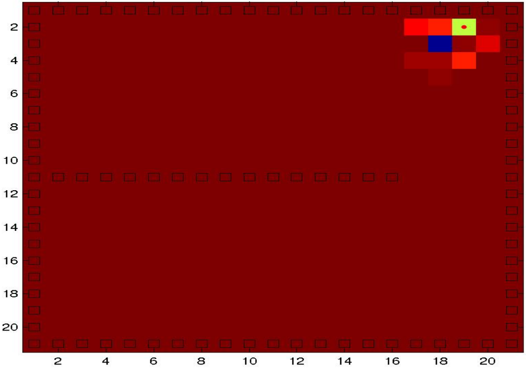
Il ne voit que ses distances aux murs autour de lui, bruitées (probabilité ε de voir une donnée incorrecte)

En se déplaçant, il finit par se localiser

Démo : suivi d'un robot



Démo : suivi d'un robot



Plan

Prédiction de séquences individuelles

- Objectifs
- Résultats
- Exemples d'applications

Filtrage pour HMM discrets

- Chaînes de Markov Cachées
- Filtrage

Application à la prédiction de séquences individuelles

- Construction d'un modèle instrumental
- Analyse de la performance
- Optimalité, Extensions, Discussion
- Ouverture sur les problèmes de bandits

Modèle instrumental pour la poursuite du meilleur expert

Idée : si $J_t^* = \mathrm{argmin}_j \, \ell_t(f_t^j, y_t)$ est le meilleur expert à l'instant t, on va faire *comme si la suite des couples $(J_t^*, y_t)_{t\geq 1}$ était une chaîne de Markov cachée* !

- Modèle de **Markov** sur la suite des meilleurs experts :

$$m(j,k) = (1-\varepsilon)\delta_{j,k} + \frac{\varepsilon}{N}$$

- Loi de **Gibbs** sur les pertes :

$$g_t(j,y) = \exp\left(-\beta\ell(f_t^j, y)\right)$$

où β désigne une *température inverse* (cf. méca stat)

Filtrage dans le modèle instrumental

On prend pour χ_0 la loi uniforme sur l'ensemble $\{1, \ldots, N\}$ des experts

EW.S utilise le filtre de prédiction comme pondération pour le choix du prochain expert :

$$\begin{aligned} P(J^*_{t+1} = j | y_{1:t}) = \phi_{t+1|t}(j) &= \sum_k \phi_t(k) m(k,j) \\ &= (1-\varepsilon)\, \phi_t(j) + \frac{\varepsilon}{N} \sum_k \phi_t(k) \end{aligned}$$

Expression récursive pour le filtre :

$$P(J^*_t = j | y_{1:t}) = \phi_t(j) = \frac{\phi_{t|t-1}(j) g_t(x, y_t)}{\gamma_t}$$

Plan

Inégalité de Hoeffding

Lemme de Hoeffding : si X est une variable aléatoire à valeur dans l'intervalle $[a, b]$, alors

$$\forall \lambda > 0,\ \log \mathbb{E}\left[\exp(\lambda X)\right] \leq \lambda \mathbb{E}[X] + \frac{(b-a)^2}{8}\lambda^2$$

Corollaire : si $X_1, \ldots, X_t$ sont des variables aléatoires i.i.d. à valeur dans l'intervalle $[a, b]$, alors

$$\forall \varepsilon > 0,\ P\left(|\bar{X}_t - \mathbb{E}[X]| > \varepsilon\right) \leq 2 \exp\left(-\frac{2t\varepsilon^2}{(b-a)^2}\right)$$

Rq : le lemme d'Hoeffding est une “réciproque” à l'inégalité de Jensen

Minoration de la vraisemblance

On minore brutalement la vraisemblance des observations dans le modèle instrumental :

$$\begin{aligned}
\log \Lambda_n(y_1, \dots, y_n) &= \log \left(\sum_{j_{0:n}} \chi_0(j_0) \prod_{t=1}^{n} m(j_{t-1}, j_t) g_t(j_t, y_t) \right) \\
&\geq \log \left(\chi_0(j_0^*) \prod_{t=1}^{n} m(j_{t-1}^*, j_t^*) g_t(j_t^*, y_t) \right) \\
&= \log \left[\frac{1}{N} \left(\frac{\varepsilon}{N} \right)^{h(j^*)} \log\left(1 - \varepsilon\right)^{n - h(j^*)} \right] - \beta L_n(\mathbf{j}_n^*)
\end{aligned}$$

pour toute suite de choix d'experts $\mathbf{j}_n$.

Majoration de la vraisemblance

D'autre part, grâce au lemme de Hoeffding :

$$
\begin{aligned}
\log \Lambda(y_t|y_{t-1}) &= \log \sum_k \phi_{t-1}(k) \exp\left(-\beta \ell(f_t^k, y_t)\right) \\
&\leq -\beta \mathbb{E}\left[\ell(f_t^{J_t}, y_t)\right] + \frac{M^2\beta^2}{8}
\end{aligned}
$$

et donc

$$
\log \Lambda_n(y_1, \dots, y_n) \leq -\beta \mathbb{E}\left[\hat{L}(\hat{p}, \mathbf{y}_n)\right] + \frac{nM^2\beta^2}{8}
$$

En recollant les morceaux

On obtient donc :

$$\mathbb{E}\left[\hat{L}(\hat{p}, \mathbf{y}_n)\right] \leq L_n(\mathbf{j}_n^*) - \frac{1}{\beta}\log\left[\frac{1}{N}\left(\frac{\varepsilon}{N}\right)^{h(j^*)}\log(1-\varepsilon)^{n-h(j^*)}\right] + \frac{M^2\beta}{8}n$$

$$= L_n(\mathbf{j}_n^*) + \frac{1}{\beta}(1+h(j^*))\log(N)$$

$$- \frac{1}{\beta}\log\left[\varepsilon^{h(j^*)}\log(1-\varepsilon)^{n-h(j^*)}\right] + \frac{M^2\beta}{8}n$$

En prenant $\varepsilon = k/(n-1)$, on voit que pour $j^* \leq k$:

$$-\log\left[\varepsilon^{h(j^*)}\log(1-\varepsilon)^{n-h(j^*)}\right] \leq (n-1)H(\varepsilon)$$

Optimisation de la borne

$$\mathbb{E}\left[\hat{L}(\hat{p}, \mathbf{y}_n)\right] \leq L_n(\mathbf{j}_n^*) + \frac{(1+k)\log(N) + (n-1)H\left(\frac{k}{n-1}\right)}{\beta} + \frac{M^2\beta}{8}n$$

Enfin, on optimise en β : on choisit

$$\beta = \frac{1}{M}\sqrt{\frac{8}{n}\left((1+k)\log(N) + (n-1)H\left(\frac{k}{n-1}\right)\right)}$$

ce qui donne :

$$\mathbb{E}\left[\hat{L}(\hat{p}, \mathbf{y}_n)\right] \leq M\sqrt{\frac{n}{2}\left((1+k)\log(N) + (n-1)H\left(\frac{k}{n-1}\right)\right)}$$

Regret face au meilleur expert

Apparaît comme un cas particulier quand $\varepsilon = 0$

Interprétation bayésienne : le meilleur expert suit la loi χ_0 (uniforme sur $\mathcal{X}$), et conditionnellement à lui les observations sont i.i.d.

EW choisit l'expert selon la *loi a posteriori* dans le modèle de Gibbs

C'est une stratégie de type *SOFT-MAX*

Plan

Prédiction de séquences individuelles

Objectifs
Résultats
Exemples d'applications

Filtrage pour HMM discrets

Chaînes de Markov Cachées
Filtrage

Application à la prédiction de séquences individuelles

Construction d'un modèle instrumental
Analyse de la performance
Optimalité, Extensions, Discussion
Ouverture sur les problèmes de bandits

Optimalité : borne inférieure

Peut-on faire mieux ?

Réponse : non, même si l'environnement n'est pas malicieux mais aléatoire, même si on se compare à la meilleure combinaison convexe

Théorème : pour $\mathcal{X} = [0, 1]$, $\mathcal{Y} = [0, 1]$, et $\ell(x, y) = |x - y|$, toute stratégie encourt un regret face au meilleur expert tel que, pour $N \geq 2$ et $n \geq c \log(N)$,

$$R_n \geq C\sqrt{n \log(N)}$$

Bornes en probabilité

Plaçons-nous dans le cadre du regret face au meilleur expert
En plus des bornes en espérance :

$$\mathbb{E}\left[R_n(\hat{p})\right] \leq M\sqrt{\frac{n}{2}\log N}$$

On montre des bornes du type : *avec probabilité* $1-\delta$*,*

$$R_n(\hat{p}) \leq M\sqrt{\log\left(n\frac{N}{\delta}\right)}$$

Méthode : concentration autour de la moyenne (Hoeffding-Azuma)

Adaptation à l'horizon

On a trouvé qu'il faut choisir

$$\beta = 1/M\sqrt{\frac{8\log(N)}{m}}$$

Que faire si on ne connaît pas M et n ?

Adaptation à l'horizon : en prenant $\beta_t = 1/M\sqrt{4\log(N)/t}$ on obtient un regret :

$$\mathbb{E}\left[R_n(\hat{p})\right] \leq M\sqrt{(n+1)\log N}$$

$\implies$ on perd un facteur $\sqrt{2}$.

cf aussi le "doubling trick"

Adaptation à la borne sur les regret

Adaptation à M : en prenant $\beta_t = \sqrt{\log(N)/\sum_{s<t} \mathbb{V}ar[\ell(f_t^{J_t}, y_t)]}$ on obtient un regret :

$$\mathbb{E}\left[R_n(\hat{p})\right] \leq CM\sqrt{n \log N} + O(1)$$

Mieux : en prenant

$$\beta_t = \min\left\{\left(\max_{j,s<t} \ell(f_s^j, y_t)\right)^{-1}, \gamma\sqrt{\frac{\log(N)}{\sum_{s<t} \mathbb{V}ar[\ell(f_t^{J_t}, y_t)]}}\right\}$$

on obtient un regret :

$$\mathbb{E}\left[R_n(\hat{p})\right] \leq M\sqrt{n \log N} + CM \log(N)$$

Amélioration pour les faibles pertes cumulées

Si l'un des experts est très bon, il se peut que $\min_j L_n(j, \mathbf{y}_n) \ll n$ et que les bornes présentées jusqu'ici soient très pessimistes !

En prenant $\beta = 1$,

$$\mathbb{E}\left[R_n(\hat{p})\right] \leq \frac{e}{e-1}\left(\min_j L_n(j, \mathbf{y}_n) + \log(N)\right)$$

En prenant

$$\beta_t = \min\left\{\left(\max_{j,s<t} \ell(f_s^j, y_t)\right)^{-1}, \sqrt{\frac{\log(N)}{\sum_{s<t} \mathbb{V}ar[\ell(f_t^{J_t}, y_t)]}}\right\}$$

on montre que

$$\mathbb{E}\left[R_n(\hat{p})\right] \leq +C\sqrt{M \min_j L_n(j, \mathbf{y}_n) \log(N)} + C' M \log(N)$$

Remarques sur les modèles instrumentaux

D'autres exemples célèbres où on introduit un modèle probabiliste instrumental :

- *Théorie de l'information* : modèle de Shannon
  modèle instrumental sur les données pour fabriquer des compresseurs
- *Finance mathématique* : modèle de Cox-Ross-Rubinstein
  probabilité risque-neutre = instrument pour calculer le prix de la couverture (déterministe) des options

Codage source

ATCAGAATC → 0011011000110011010

compression sans perte

Winzip, compress, etc.

But : minimiser la longueur de code

Codage source : modèle de Shannon

Source P

= processus stationnaire sur l' **alphabet A**

ici, $A=\{A,C,T,G\}$

code $\phi_n : A \to \{0,1\}^*$

ATCAGAATC → 001101100011001010

message X_1^n (n=9)

compression sans perte

Winzip, compress, etc.

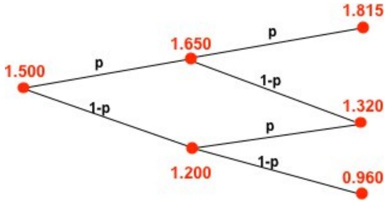
mot de code

$\phi_n(X_1^n)$

But : minimiser la longueur de code moyenne

$$E_P[|\phi_n(X_1^n)|]$$

Modèle de Cox-Ross-Rubinstein



Méthode numérique à temps discret pour l'évaluation des options

On introduit la *probabilité risque-neutre* sous laquelle le prix de l'actif est une *martingale*
Le prix de l'option apparaît alors comme une *espérance conditionnelle*

Limite quand le pas de temps tend vers 0 = modèle de Black-Scholes

Plan

Prédiction de séquences individuelles

Filtrage pour HMM discrets

Application à la prédiction de séquences individuelles

Problèmes de bandits classiques

Dans de nombreuses applications, on n'observe que la perte de l'expert qu'on a choisi (et pas celle des autres)

On parle alors de *problème de bandit* (classique)

Exemples : problèmes de plus court chemin

Ajustement dynamique des prix

Un site commercial cherche à maximiser ses profits pour un bien dont on néglige le coût de production :

- on fixe à l'instant t le prix $\hat{p}_t \in \mathcal{X} = \{14.99, 19.99, 24.99, \dots, 99.99\}$
- chaque client est prêt à payer un prix maximal $y_t \in]0, 100]$
- la perte est

$$\ell(f_t, y_t) = y_t \mathbb{1}_{y_t \leq f_t} + (y_t - f_t)_+ = y_t - f_t \mathbb{1}_{f_t \leq y_t}$$

Le regret est équivalent si on prend la perte $\tilde{\ell}(f_t, y_t) = -f_t \mathbb{1}_{f_t \leq y_t}$, qui elle est observable pour le prix $\hat{p}_t$.

Formalisation en théorie des jeux

Cadre : ensemble prédictions $\mathcal{X}$, ensemble d'observations $\mathcal{Y}$, fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ bornée par M
Acteurs : un décideur, un environnement
Déroulement : pour chaque instant $t = 1, 2, \dots$:

1. les experts publient leurs prédictions $(f_t^j)_{1 \leq j \leq N}$
2. le décideur choisit une pondération p_t
3. l'environnement choisit une observation y_t
4. le décideur accorde sa confiance à l'expert j avec probabilité $p_t(j)$
5. le décideur découvre l'observation y_t et enregistre uniquement sa perte $\ell(f_t^{J_t}, y_t)$

Algorithme EXP3

EXP3 = *EXPonential-weight algorithm for EXPloration and EXPloitation*, variante de l'algorithme *Hedge* (couverture) de [Freund&Schapire, Journal of Computer and System Sciences 55, 119–139 (1997)]

Idée: estimateur *sans biais* de $\ell(f_t^j, y_t)$:

$$\hat{\ell}(f_t^j, y_t) = \frac{\ell(f_t^{J_t}, y_t)}{p_t(j)} \mathbb{1}_{\{J_t = j\}}$$

On estime ainsi les pertes cumulées $\hat{L}_t(j, \mathbf{y}_t)) = \sum_{s=1}^t \hat{\ell}(f_t^j, y_t)$.

EXP3 = stratégie randomisée avec comme choix de pondération :

$$\hat{p}_t(j) = \frac{\exp(-\beta \hat{L}_{t-1}(j, \mathbf{y}_{t-1}))}{\sum_k \exp(-\beta \hat{L}_{t-1}(k, \mathbf{y}_{t-1}))}$$

Borne de regret pour EXP3

Théorème: En choisissant

$$\beta = \frac{1}{M}\sqrt{\frac{2\log(N)}{nN}}$$

le regret de l'algorithme EXP3 face à la meilleure stratégie constante vérifie :

$$\mathbb{E}\left[R_n(\hat{p})\right] \leq M\sqrt{2nN\log(N)}$$

On perd donc un facteur multiplicatif $\sqrt{N}$

Preuve : encore et toujours des arguments de concentration…