KCD 03/02/00

Note sur l'algorithme de CHEN DEMPSTER.

On part d'un vecteur $\tilde{\pi}_k$ de proba d'inclusion sur U et $\tilde{p}$ est le plan Poissonien associé.

$$\lambda_k = \log \frac{\tilde{\pi}_k}{1-\tilde{\pi}_k} \quad \left(\tilde{\pi}_k = \frac{\exp \lambda_k}{1+\exp\lambda_k}, \text{ logit !}\right).$$

Pour D partie de C_N (le cube de dimension N, ou plutot l'ensemble de ses sommets), on note allégrement

$$S(D) = \exp \kappa(D) = \sum_{s\in D} \exp \lambda s \quad \rightarrow \quad \tilde{p}(s) = \exp(\lambda s - \kappa(C_N)).$$

Evidemment, le plan conditionnel à D a pour probas $\tilde{p}(s) = \exp(\lambda s - \kappa(D))$ si $s \in D$, 0 sinon.

On s'intéresse à des domaines liés aux trois (pour l'instant) type de contrainte (linéaires).

- taille $|s| = 1.s = n$ $\quad (S(n)\,;\ \kappa(n))$
- Support : $s \subset V$ ou $1'_{\bar{V}}.s = 0$ $\quad (\bar{V} = U - V.)$
- Contenir k $\quad s \ni k$ ou $1'_k.s = 1 \ (= s_k)$

On écrira (n), (n,V), (n,V,k) les domaines liés à ces contraintes.

On veut tirer un échantillon selon $\tilde{p}_n$:

$$\tilde{p}_n(s) = \frac{\exp \lambda s}{S(n)} = \frac{\exp(\lambda s)}{S(n,U)}.$$

Les prob d'inclusion sont :

$$\pi_k^{(n)} = \pi_k^{(n,U)} = \frac{\exp\lambda_k \; S(n-1, U-k)}{S(n,U)} = \frac{S(n,U,k)}{S(n,U)}.$$

façon plus générale, les proba d'inclusion

(m, V) sont :

$$\pi_k^{(m,V)} = \frac{S(m,V,k)}{S(m,V)} = \exp\lambda_k \frac{S(m-1, V-k)}{S(m,V)}.$$

(Rem. $S(m,V,k) = S(m-1, V-k)\exp\lambda_k$ si $m \geq 1$, $k \in V$

[Convention $S(0,U) = 1$?].

<u>Méthode</u> :

1 Calcul de $\pi_k^{(m,V)}$ [création d'une fonction des λ_k].

$$S(m-1, V-k) = S(m-1, V)\left(1 - \pi_k^{(m-1,V)}\right) \quad \text{par définition.}$$

<u>Donc</u> : $\pi_k^{(m,V)} = \exp\lambda_k \left(1 - \pi_k^{m-1,V}\right) * \dfrac{S(m-1,V)}{S(m,V)}$ (si $k \in V$, 0 sinon)

On a evidemment $\pi_k^{0,V} = 0$,

d'où $\pi_k^{1,V} = \exp\lambda_k \Big/ \left(\sum_V \exp\lambda_k\right)$

Comme $\sum_k \pi_k^{m,V} = m$, la récurrence s'organise

$$: \quad S(m,V) = \frac{1}{m} S(m-1,V) . \sum_k \exp\lambda_k \left(1 - \pi_k^{m-1,V}\right)$$

<u>Remarque</u> : Le calcul explicite des $S(m,V)$ n'est pas obligatoire !

Ceci permet de fabriquer la fonction

$$\pi^{(m,V)} = f^{m,V}(\lambda) \quad \text{(ou cce } \tilde{\pi} \text{ au choix).}$$

gorithme :

se base sur l'idée (bête !) suivante.

Soit s un échantillon tiré de $\tilde{P}^{\nu}_{n,V}$.

Prendre un élément k de s au hasard, c'est le tirer avec la proba $p_k = \frac{\pi_k^{nV}}{n}$

(NB : $\sum_k p_k = 1$; $p_k = 0$ si $k \notin V$). Ce qui reste est un échantillon de $\tilde{P}_{n-1, V-k}$. Du moins on pourrait le supposer si on regarde la construction que ça suggère

b) Construction :

① tirer k_1 dans $\frac{\pi(n, U)}{n}$: proba : $\frac{1}{n} \exp \lambda_{k_1} \frac{S(n-1, U_1)}{S(n, U)}$

avec $U_1 = U - k_1$ ... $(U_m = U - \{k_1 \ldots, k_m\} = U_{m-1} - k_m)$.

② tirer k_2 dans $\frac{\pi(n-1, U_1)}{n-1}$ proba $\frac{1}{n-1} \exp \lambda_{k_2} \frac{S(n-2, U_2)}{S(n-1, U_1)}$

... k_{m+1} —— $\frac{\pi(n-m, U_m)}{(n-m)}$ .. $\frac{1}{n-m} \exp \lambda_{k_m} \frac{S(n-(m+1), U_{m+1})}{S(n-m, U_m)}$

Au bout de n étapes, on a tiré un échantillon séquentiel $(k_1, \ldots, k_n)$ avec la proba :

$$\frac{1}{n!} \exp\left(\sum_{i=1}^{n} \lambda_{k_i}\right) \cdot \frac{S(0, U_{m+1})}{S(n, U)} = \frac{1}{n!} \exp \lambda s / S(n, U).$$

Il y a $n!$ façons d'arriver à l'échantillon non ordonné s, chacun ayant la même proba et donc $pr(s) = \exp(\lambda s - \mu(n)) = \tilde{P}_n(s)$, c'est gagné.

Pour avoir définitivement gagné, il faut trouver les λ_k (ou les $\tilde{\pi}_k$) si on connait les $\pi_k = \tilde{\pi}_k^m$.
Nos auteurs proposent une méthode. On doit pouvoir y arriver aussi en utilisant la résolution de $\pi = f_{m,V}(\lambda)$ par la méthode de Newton. En effet :

a) On peut récursivement calculer les $\tilde{\pi}_{kl}^m$ avec les mêmes méthodes :

$$\pi_{k,l}^{m,V} = \frac{\sum S(m, V, (kl))}{S(m,V)} = \exp\lambda_k \exp\lambda_l \frac{S(m-2, V-(k,l))}{S(m,V)}$$

$$= \exp\lambda_k \exp\lambda_l \left(1 - \pi_k^{m-2,V} - \pi_l^{m-2,V} + \pi_{kl}^{m-2,V}\right) * \frac{S(m-2,V)}{S(m,V)}$$

(si $k l \in V$
0 sinon)

On a évidemment $\pi_{kl}^{0,V} = 0$ et $\pi_{kl}^{1,V} = 0$
ce qui permet la recurrence et donc le calcul de $\pi_{(k,l)}^{m,V} = f_2^{m,V}(\lambda)$.

Maintenant comme $\frac{\partial\pi}{d\lambda} = \left[\pi_k \, \pi_{kl}\right] = M(\lambda)$ on peut appliquer la méthode de Newton avec les itérations $\lambda_k^{(0)} = \log\frac{\pi_k}{1-\pi_k}$ et $\lambda^{(n+1)} = \lambda^{(n)} - M(\lambda)^{-1}\pi^{(n)}$

(Les notations ont encore un peu flotté, c'est habituel).

L'alternative, qui revient à reprendre l'idée de CHEW DEMPSTER consiste à repartir de l'équation du haut de la p.2 :

$$\pi_k^{(m,V)} = \exp\lambda_k \cdot \frac{S(m-1, V-k)}{S(m,V)}$$

Partant (par exemple, mais quoi d'autre ?) de $\exp\lambda_k = \frac{\pi_k}{1-\pi_k}$ on calcule les $\pi_k^m(\lambda) = \pi^{((0))}$ (toujours ce flottement dans les notations). On récupère dans le calcul le ratio $r(\lambda) = \frac{S(m-1, V-k)}{S(m,V)}$.

Du coup, on "corrige" λ par $\exp\lambda_k^{(1)} = \frac{\pi_k}{\pi_k^{(0)}} \cdot \frac{1}{r(\lambda)}$

ce qui permet de calculer $\pi^{((1))}$ et d'itérer.

Si il y a des problèmes numériques d'inversion de la matrice M, ça peut être meilleurs. Cependant les erreurs de troncature doivent cumuler assez vite. C'est à essayer

$$\hat{P}^{n}(s\,|\,d) = c^{n}\ \underline{\exp(\lambda(s+d))}$$

$$P(s\,|\,s \supset d) = \frac{\exp(\lambda s' \cancel{\ldots})\ \exp(\lambda d - \mu)}{\sum_{s' \subset V-d} \exp \lambda s'}$$ $\boxed{P_V(d)\,.\ P_{(V-d)}(s')}$

$$P_V(s) = \frac{\exp \lambda s)}{\sum_{s \in V} \exp \lambda \cancel{s}} \quad s \cup s \subset V$$

$\sum_{s \in V} \exp \lambda s \to \exp \mu_V$ $\boxed{\exp(\lambda s - \mu_V)}$

$$P(s\,|\,s \supset d) = \frac{\exp(\lambda(s-d) - \mu_V)\ \exp \lambda d}{\sum_{s' \in V-d} \exp \lambda s'} = \boxed{P_V(d)\, P_{V-d}(s')}$$

P_n

$$P(s\,|\,s \ni k) = P_V(\cancel{(k)})\, P_{V-k}(s'-k$$

$\downarrow$

$\frac{\pi_k}{1-\pi_k}$

$$P(\{k\}) = \pi_k \prod_{V-k} (1-\pi_k)$$
$$= \frac{\pi_k}{1-\pi_k} \prod_V (1-\pi)$$

$$\exp \lambda s \quad \prod_s \left(\frac{\pi}{1-\pi}\right)$$

$$\sum \exp \lambda s$$

$$\sum_{(s)} \left(\frac{\pi}{1-\pi}\right)^s$$

$$\prod \left(\frac{1}{1-\pi}\right) \quad \sum_s \frac{\cancel{\ldots}\ \pi^s (1-\pi)^{v-s}}{1}$$

$\boxed{\mu_V = -\sum_V \log(1-\pi_k)} > 0$

$$\exp(-\mu_V) = \prod_{k \in V} (1-\pi_k)$$
$$= \prod_k \exp \log(1-\pi_k)$$
$$= \exp \sum_V \log(1-\pi_k)$$
$$= \underline{\sum_k} \pi^s (1-\pi)^{v-s} \Big/ \prod_V (1-\pi)$$

$$\frac{1}{\prod_V (1-\pi_k)} = \cancel{\exp}\ \exp \mu$$

$$\exp(-\mu) = \prod_V (1-\pi_j)$$
$$= \exp \sum \log(1-\pi)$$
$$\exp \mu = \sum \exp \lambda s$$
$$= \sum \left(\frac{\pi}{1-\pi}\right)^s$$