$P(A\cup B) = P(A)+P(B)-P(A\cap B)$

$P(X\in I) = \int_I p(x)dx$ P = Proba, p = fct° densité

si $I =]-\infty; x[$ ⇒ fct° Répartit° F

$F(x) = P(X<x) = \int_{-\infty}^{x} p(y)dy$

discret : $E(X) = \sum x_k p_k = \mu$

$V(X) = \sum (x_k - E(X))^2 p_k = \sigma^2$

continue : $E(X) = \int x\, p(x)dx$

$V(X) = \int (x-E(X))^2 p(x)dx$

$E(aX+b) = aE(X)+b$ $V(aX+b) = a^2V(X)$

$E(X+Y) = E(X)+E(Y)$

$V(X+Y) = V(X)+V(Y)$ si X et Y indep

sinon $= V(X)+V(Y)+2cov(X,Y)$

$cov(X,Y) = E((X-E(X))(Y-E(Y)))$ discret

$= \iint (x-E(x))(y-E(y))\, p(x,y)\,dx\,dy$ continu

échantillon : $\{(x_1,y_1),\dots,(x_k,y_k),\dots,(x_n,y_n)\}$

$cov(X,Y) = \frac{1}{n}\sum_k (x_k-\bar{x})(y_k-\bar{y})$

corrélation linéaire : $r = cov(X,Y)/\sigma_x\sigma_y$

Loi jointe : $P(A,B) = P(A|B)P(B)$

Loi marginale : $P(A) = \sum_i P(A,B=i) = \sum_i P(A|B=i)P(B_i)$

Bayes : $P(A|B) = \frac{P(A|B)P(A)}{P(B)}$ vraisemblance, ← prior, ← évidence, postérieure

$P(A|B,C) = \frac{P(A,B,C)}{P(B,C)} = \frac{P(B|A,C)P(A|C)}{P(B|C)}$

$P(A,B,C) = P(A,B|C)P(C)$

indep : $P(A,B) = P(A)P(B)$ ou $P(A|B) = P(A)$

mutuellem indep : $P(x_1,\dots,x_n) = \prod P(X_k=x_k) = \prod_{k=1}^{n} P_{X_k}(x_k)$ (vrai ≠ 2 à 2)

indep condi à Z : $P(X,Y|Z) = P(X|Z)P(Y|Z)$

si $P(Y|Z)>0$: $P(X|Y,Z) = P(X|Z)$

si $P(X|Z)>0$: $P(Y|X,Z) = P(Y|Z)$

Loi Bernoulli : $q = 1-p$ (p proba succès)

$P(X=1) = p$ $P(X=0) = 1-p$

$E(X) = p$ $V(X) = p(1-p)$ X : nb succès

Binomiale : $X\sim \mathcal{B}(n,p)$ n Bernoulli iid

$P(X=k) = C_n^k p^k (1-p)^{n-k}$ $E(X) = np$

$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$ $V(X) = np(1-p)$

Normale : $X\sim \mathcal{N}(\mu,\sigma^2)$

$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$

$Y = aX+b \sim \mathcal{N}(a\mu+b, a^2\sigma^2)$

$Z = \frac{X-\mu}{\sigma}\sim\mathcal{N}(0;1)$ centrée réduite

TCL : $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\sim\mathcal{N}(0;1)$

mutuellem indep X_n $(n\in\mathbb{N})$

Normale 2D : $\rho = \frac{cov(X,Y)}{\sigma_x\sigma_y}$

$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)\right)$

Loi faible : indep 2 à 2

$\bar{X}_n = \frac{\sum_k^n X_k}{n}$ converge en proba vers m

(moy empiriq) $E(\bar{X}_n) = m$ $V(\bar{X}_n) = \frac{\sigma^2}{n}$

$P(\lim_{n\to\infty} X_n = a) = 1 \Leftrightarrow P(\lim_{n\to\infty}\sup_{k>n}|X_k-a|\geq\varepsilon) = 0$

Loi forte : mut indep, converge sûrem vers m

$X_n\to X$ si $F_n(x)\to F(x)$ en tt F avec F_n fct° Répartit° de X_n

Expo : $E(X) = 1/\lambda$ $V(X) = 1/\lambda^2$

Répartition : $F(x) = 1-e^{-\lambda x}$

densité : $f(x) = \lambda e^{-\lambda x}$

Géométriq : (discret expo) $E(X) = \frac{1}{p}$

$P(X=k) = p(1-p)^{1-k}$ $V(X) = \frac{1-p}{p^2}$

Poisson : $\lambda>0$ (discret)

$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$ $E(X) = V(X) = \lambda$

Lois conjuguées : $\pi(\theta)$ et $\pi(\theta|x)$ sont conj si $\pi(x|\theta)$ appartient à la m̂ famille de loi que $\pi(\theta)$

Loi Beta : $Beta(\theta,a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ $E(X) = \frac{a}{a+b}$

$V(X) = \frac{ab}{(a+b)^2(a+b+1)}$

$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt$

(Loi binomiale et β conjuguées)

Loi a priori : $\pi(\theta) = Beta(\theta,a,b)$

fct° vraisemblance : $\pi(x|\theta) = \theta^x(1-\theta)^{n-x}$ (avec $x = \#(x_i=1)$)

Loi a posteriori : $\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{\sum_\theta \pi(x|\theta)\pi(\theta)} \propto \pi(x|\theta)\pi(\theta)$

$\propto \theta^{x+a-1}(1-\theta)^{b+n-x-1}$

$\Rightarrow \pi(\theta|x)\sim Beta(\theta; x+a; b+n-x)$

Loi Γ : conjuguée avec Normale (σ^2 connue)

$X\sim\Gamma(x,k,\theta)$

$f(x,k,\theta) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)}$ $\forall x,k,\theta>0$

$\Gamma(k) = \int_0^{+\infty} t^{k-1}e^{-t}dt$ $E(X) = k\theta$ $V(X) = k\theta^2$

k entier : $\Gamma(x,k,\theta)$ = loi k var indep suivant une loi exp d'espérance θ

Uniforme : $f(x) = \frac{1}{a-b}$ $a<b$

$E(X) = a+b/2$ $V(X) = (b-a)^2/12$

vraisemblance : proba obtenir cet échantillon sachant loi P

$L(x) = P(x_1,\dots,x_n) = \prod_{i=1}^{n} P(x_i)$

ex : hyp : θ_1, θ_2 $(n = 80)$

$P(n|\theta_1) = \mathcal{N}(\mu_1,\sigma_1^2) = (100, 20^2)$

$P(n|\theta_2) = \mathcal{N}(\mu_2,\sigma_2^2) = (85, 5^2)$

$L(x,\theta_1) = p(80|\theta_1) \simeq 0{,}012$

$= \frac{1}{20\times\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{80-100}{20}\right)^2\right)$

$L(x,\theta_2)\simeq 0{,}048$ donc θ_2

ex : estimer $\Theta = \{\theta_1,\theta_2\}$

$X\in\{$pile, face$\}$

échantillon : PPFFPPFPP

$L(x,\Theta) = \prod^7 P(P|\Theta)\times\prod^3 P(F|\Theta)$

$\theta_1 = 0{,}75$ $\theta_2 = 0{,}5$ $\theta_3 = 0{,}25$

$L(x,\theta_1) = 0{,}75^7\times 0{,}25^3 \simeq 0{,}002086$

$L(x,\theta_1) = 0{,}5^7\times 0{,}5^3 \simeq 0{,}000975$

$L(x,\theta_3) = 0{,}25^7\times 0{,}75^3 \simeq 0{,}00026$

θ_1 plus vraisemblable que θ_2 et θ_3

estimateur max L :

$t = f(x) = \arg\max_\theta L(x,\theta)$

$= \arg\max_\theta P(x_1,\dots,x_n|\theta)$

$= \arg\max_\theta \prod_{i=1}^{n} P(x_i|\theta)$

$= \arg\max_\theta \ln(L(x,\theta))$

$= \arg\max \sum \ln P(x_i|\theta)$

$\Rightarrow \frac{\partial \ln P(x_i|\theta)}{\partial\theta} = 0$

ex : $X\in\{0;1\}$ (face, pile)

$X\sim\mathcal{B}(1,p)$ n lancers

$P(X=x|p) = p^x(1-p)^{1-x}$

obs : $x = \{x_1,\dots,x_n\}$

$P(x|p) = \prod P(x_i|p)$

$= \prod p^{x_i}(1-p)^{1-x_i} = L(x,p)$

$\ln L = \sum(x_i\ln p + (1-x_i)\ln(1-p))$

$\frac{\partial \ln L}{\partial p} = \frac{1}{p}\sum x_i - \frac{n-\sum x_i}{1-p} = 0$

$\Rightarrow p = \frac{1}{n}\sum x_i$

ex : $X\sim\mathcal{N}(\mu,\sigma^2)$ $\sigma = 1$ θ?

$L(x,\theta) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}(x_i-\theta)^2\right)\right)$

$\ln L = \frac{-n}{2}\ln 2\pi - \frac{1}{2}\sum(x_i-\theta)^2$

$\frac{\partial \ln L}{\partial\theta} = 0 \Rightarrow \theta = \frac{1}{n}\sum x_i = \bar{x}$

θ ? σ² ? $\ln L(x,\theta)$

$= \frac{-n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$

$\frac{\partial \ln L}{\partial\mu} = \frac{1}{\sigma^2}\sum(x_i-\mu) = 0$

$\frac{\partial \ln L}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(x_i-\mu)^2 = 0$

$\mu = \frac{1}{n}\sum x_i = \bar{x}$ $\sigma^2 = \frac{1}{n}\sum(x_i-\bar{x})^2 = s_n^2$

max a posteriori (MAP) :

X : obs x   Θ : param θ

$\pi(x,\theta) = \Pi(X=x,\Theta=\theta)$

$\pi(x) = \Pi(X=x) = \int_{\theta\in\Theta}\Pi(X=x,\Theta=\theta)\,d\theta$

$= \int\sum_{x\in X}\pi(x,\theta)\,d\theta\,dx$

$\pi(x|\theta) = \Pi(X=x|\Theta=\theta) = \frac{\pi(x,\theta)}{\pi(\theta)}$

$\pi(\theta|x) = \Pi(\Theta=\theta|X=x) = \frac{\pi(x,\theta)}{\pi(x)}$

$\pi(\theta)$ proba a priori de θ (avant obs)

$\pi(\theta|x)$ proba a posteriori de θ

$\pi(x|\theta)$ vraisemblance $L(x,\theta)$

$\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{\int\sum\pi(x,\theta)d\theta} = \frac{\pi(x|\theta)\pi(\theta)}{\int\sum\pi(x|\theta)\pi(\theta)d\theta}$

estimateur MAP : $t = \arg\max_\theta \pi(\theta|x)$

ex : $X\in\{0;1\}$ (face, pile)

$X\sim\mathcal{B}(1,\theta)$ $P(X=x|\theta) = \theta^x(1-\theta)^{1-x}$

échantillon : PPP $(n=3)$

max L → θ = 1 $\pi(\theta_4) = 3/32$

$\Theta = \{\theta_1 = 1, \theta_2 = 2/3, \theta_3 = 1/2, \theta_4 = 1/3\}$

$\pi(\theta_1) = 1/32$; $\pi(\theta_2) = 1/4$; $\pi(\theta_3) = 1/2$

$L(x,\theta_1) = \pi(x|\theta_1) = \prod_{i=1}^{3} P(x_i|\theta_1) = 1$

$L(x,\theta_2) = \left(\frac{2}{3}\right)^3\times\left(1-\frac{2}{3}\right)^0 = \left(\frac{2}{3}\right)^3 \simeq 0{,}296$

$L(x,\theta_3) = 0{,}125$ $L(x,\theta_4)\simeq 0{,}037$

$\pi(\theta_1|x) = \frac{L(x,\theta_1)\pi(\theta_1)}{\sum L(x,\theta)\pi(\theta)} \propto L(x,\theta_1)\pi(\theta_1)$

$= 1\times 1/32 \simeq 0{,}03125$

$\pi(\theta_2|x) = (2/3)^3\times 1/4 \simeq 0{,}074$

$\pi(\theta_3|x)\simeq 0{,}0625$ $\pi(\theta_4|x)\simeq 0{,}008$

donc $\Theta = \theta_2 \Rightarrow X\sim\mathcal{B}(1;\frac{2}{3})$

$\Rightarrow P(X=F)\neq 0$

Expectat° - Maximizat° (EM) :

x° : données obs x^h : manquantes

$\ln L(x^\circ,\theta) = \sum\ln P(x_i^\circ|\theta)$

$= \sum\ln\left(\sum_{x^h\in X^h} P(x_i^\circ,x_i^h|\theta)\right)$

$Q_i(x_i^h)$ loi de proba qlconq

$\ln L = \sum\ln\left(\sum Q_i(x_i^h)\frac{P(x_i^\circ,x_i^h|\theta)}{Q_i(x_i^h)}\right)$

inégalité de Jensen :

$\ln\left(\sum\lambda_i y_i\right)\geq\sum\lambda_i\ln(y_i)$

$\ln L \geq \sum_i\sum_{x^h} Q_i(x_i^h)\ln\left(\frac{P(x_i^\circ,x_i^h|\theta)}{Q_i(x_i^h)}\right)$

Jensen ⇒ égalité ssi $\frac{P}{Q}$ = cte

choisir $Q_i(x_i^h)\propto P(x_i^\circ,x_i^h|\theta)$

$\Rightarrow Q_i(x_i^h) = P(x_i^h|x_i^\circ,\theta)$

algo : ① choisir val init $\Theta = \Theta^\circ$

② étape E : $\forall i\in\{1,\dots,n\}$

$Q_i^{t+1}(x_i^h) = P(x_i^h|x_i^\circ,\theta^t)$

$\ln L^{t+1}(x^\circ,\theta) = \sum\sum Q_i^{t+1}(x_i^h)\ln\left(\frac{P}{Q}\right)$

③ étape M : $\Theta^{t+1} = \arg\max_\theta \ln L^{t+1}(x^\circ;\theta)$

④ tant que pas convergence, revenir ②

(convergence : Θ^{t+1} = opti local par max L)

ex : $A\in\{a,b\}$ et $C\in\{c,d\}$

$P(A,C|\Theta) = \begin{bmatrix}\theta_{ac} & \theta_{ad}\\ \theta_{bc} & \theta_{bd}\end{bmatrix} \Rightarrow \Theta$

A	C
a	?
b	?
a	d
b	d
a	c

par max L : (A tjrs obs)

$\theta_{ac}+\theta_{ad} = 3/5$ $\theta_{bc}+\theta_{bd} = 2/5$

init possible : $\Theta^\circ = \{\theta^\circ_{ab} = 0.3, \theta^\circ_{ad} = 0.3, \theta^\circ_{bc} = 0.2, \theta^\circ_{bd} = 0.2\}$

étape E : $Q_i'(x_i^h) = P(x_i^h|x_i^\circ,\Theta^\circ)$

$Q_1'(c) = P(C|A=a,\Theta^\circ) = [0.5; 0.5]$

$= \frac{P(A=a,C|\Theta^\circ)}{\sum_C P(A=a,C|\Theta^\circ)} = \left[\frac{0{,}3}{0{,}6};\frac{0{,}3}{0{,}6}\right]$ $\theta_{ab}+\theta_{bd} = 0{,}6$

$Q_2'(c) = P(C|A=b,\Theta^\circ) = [0.5; 0.5]$

$P(x_i^h,x_i^\circ|\Theta^\circ) = \begin{bmatrix}0.3 & 0.3\\ 0.2 & 0.2\end{bmatrix}$

$\ln L^{t+1}(x^\circ,\Theta)$

$= \sum\sum Q_i^{t+1}(x_i^h)\ln\left(\frac{P(x_i^\circ,x_i^h|\Theta^\circ)}{Q_i^{t+1}(x_i^h)}\right)$

$= \sum\sum Q_i^{t+1}\left(\ln P(x_i^\circ,x_i^h|\Theta^\circ) - \ln Q_i^{t+1}\right)$

étape M : $\Theta^{t+1} = \arg\max_\theta \ln L^{t+1}(x^\circ;\theta)$

$= \arg\max_\theta \sum\sum Q^{t+1}\ln(P(x_i^\circ,x_i^h|\theta))$

A	C	Q_i^{t+1}
a	c	0.5
a	d	0.5
b	c	0.5
b	d	0.5
a	d	1
b	d	1
a	c	1

$= \arg\max_\theta\big((0.5+1)\ln\theta_{ac} + (0.5+1)\ln\theta_{ad} + 0.5\ln\theta_{bc} + (0.5+1)\ln\theta_{bd}\big)$

contrainte : $\sum\theta_{ij} = 1$

$\theta_{bc} = 1-\theta_{ac}-\theta_{ad}-\theta_{bd}$

dérivées partielles puis annuler

$\Theta' = \{\theta'_{ac} = 3/10;\ \theta'_{ad} = 3/10;\ \theta'_{bc} = 1/10;\ \theta'_{bd} = 3/10\}$

Mixture de gaussienne :

$\Theta = \{\mu_1,\mu_2,\sigma_1,\sigma_2,\pi_1,\pi_2\}$

$p(x|\Theta) = \pi_1\mathcal{N}(\mu_1,\sigma_1^2)+\pi_2\mathcal{N}(\mu_2,\sigma_2^2)$

$L(x,\theta) = \prod p(x_i|\theta)$

$= \prod_i\sum_k\pi_k\frac{1}{\sqrt{2\pi}\sigma_k}\exp\left(-\frac{1}{2}\left(\frac{x_i-\mu_k}{\sigma_k}\right)^2\right)$

$\ln L = \sum\ln\sum_k\pi_k\dots$ compliqué à max analytiqm

x_i appartient à une classe $y_{k(i)}$ non obs $\sim\mathcal{N}(\mu_{k(i)},\sigma_{k(i)})$

$P(X,Y_{k(i)}|\theta) = P(X|Y_{k(i)},\theta)P(Y|\theta)$

$= \begin{bmatrix}\mathcal{N}(\mu_1,\sigma_1^2)\pi_1\\ \mathcal{N}(\mu_2,\sigma_2^2)\pi_2\end{bmatrix}$ ①

$P(Y|X,\theta) = \frac{P(x_i,Y_{k(i)}|\theta)}{p(x_i|\theta)}\propto$ ①

$\arg\max \ln L^{t+1}(x^\circ,\Theta)$

$= \arg\max_\theta \sum Q_i^{t+1}(y_1)\ln\left(\pi_1\frac{1}{\sqrt{\dots}}\exp(-1\dots)\right) + Q_i^{t+1}(y_2)\ln(\pi_2\dots)$

MC : PNSPNSP   n = 7

état	S	N	P
nb apparit°	2	2	3
freq	2/7	2/7	3/7

nb apparit° bigrammes :

	S	N	P
S	0	0	2
N	2	0	0
P	0	2	0

$A = \begin{bmatrix}0 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}$

$\pi = [0\ 0\ 1]$ on a 1 seq qui commence par P

Student : $\nu\geq 1$

$f(x) = \frac{1}{\sqrt{k\pi}}\frac{\Gamma(\frac{k+1}{2})}{\Gamma(k/2)}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}$

$E(X) = 0$ ($\nu>1$) $V(X) = \frac{1}{\nu-2}$ ($\nu>2$)

Test d'indépendance :

$\alpha = P(\text{Rejet } H_0 | H_0 \text{ vraie})$

$\beta = P(\text{Rejet } H_1 | H_1 \text{ vraie})$

puissance $= 1-\beta$

ex : $n = 25$ $X\sim\mathcal{N}(\mu;100)$

H_0 : $\mu = 10$ H_1 : $\mu>10$

sous H_0 : $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-10}{2}\sim\mathcal{N}(0;1)$

Région critiq : "rejet H_0 si $\bar{x}>14$"

$\alpha = P(\bar{X}>14|\mu=10)$

$= P\left(\frac{\bar{X}-10}{2}>\frac{14-10}{2}\middle|\mu=10\right)$

$\alpha = P\left(\frac{\bar{X}-10}{2}>2\right)\simeq 0{,}0228$

sous H_1 : supposons $\mu = 11$

$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-11}{2}\sim\mathcal{N}(0;1)$

$1-\beta(11) = P(\text{Rejet } H_0|H_1\text{ vraie})$

$= P(\bar{X}>14|\mu=11)$

$= P\left(\frac{\bar{X}-11}{2}>\frac{14-11}{2}\right) = P\left(\frac{\bar{X}-11}{2}>1{,}5\right)$

$\simeq 0{,}0668$ avec la table

ex : $X\sim\mathcal{N}(\mu,6^2)$ $n = 9$

$\{10, 8, 13, 20, 12, 14, 9, 17, 15\}$

Les notes ont baissées ? $\alpha = 5\%$

H_0 : $\mu = 14$ H_1 : $\mu<14$

sous H_0 : $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-14}{2}\sim\mathcal{N}(0;1)$

$P\left(\frac{\bar{X}-14}{2}<\frac{c-14}{2}\right) = 0{,}05$ utiliser la table

$\frac{c-14}{2}\simeq 1{,}645 \Rightarrow c = 10{,}71$

⇒ Rejet H_0 si $\bar{x}<10{,}71$

or tableau → $\bar{x} = 12$ donc pas rejet H_0

puissance test : "H_1 : $\mu = 12$"

$1-\beta(12) = P(\bar{X}<10{,}71)$

$= P\left(\frac{\bar{X}-12}{2}<-0{,}645\right)\simeq 25{,}95\%$

Loi du χ² : pop répartie en k classes

hyp : répartit° des classes connues → p_r = proba individu classe c_r

N_r = v.a. "nb ind tirés de c_r"

pour chq ind → p_r chances d'être dans c_r

X_i^r = v.a. succès si i dans c_r

$X_i^r\sim\mathcal{B}(1,p_r)$ $N_r\sim\mathcal{B}(n,p_r)$

n ind $N_r\sim\mathcal{N}$ si n grand

$D_n^2 = \sum_{i=1}^{k}\frac{(N_r-np_r)^2}{np_r}$

$D_n\sim\chi^2_{k-1}$ $E(X) = r$ $V(X) = 2r$

ddl = (col − 1)(lig − 1)

ex : n = 10 cabines $\alpha = 5\%$

n° cabine	1	2	3	4	5	6	7	8	9	10
nb clients	24	14	18	20	23	13	23	24	23	18

H_0 : "répartit° des clients est uniforme"

H_1 : "pas uniforme"

H_0 : 20 clients / cabine (moy tab)

X_i var "effectif" pour cabine i

$D^2 = \sum_{i=1}^{10}\frac{(X_i-20)^2}{20}$ $D^2\sim\chi^2_9$

$\alpha = 0{,}05 = P(\text{Rejet } H_0|H_0\text{ vraie})$

$= P(D^2>d_\alpha|D^2\sim\chi^2_9)$

$\Rightarrow d_\alpha = 16{,}9$ (table χ²)

calcul val d obs sur échanti :

$d^2 = \frac{1}{20}\left((14-20)^2+(24-20)^2+\dots+(23-20)^2\right)$

$\simeq 7{,}6$

donc $d^2<D^2\Rightarrow H_0$ accepté

ex : C_1 : notes <8 C_2 : [8;12[C_3 : ≥12

X : v.a. "note 1ère session"

Y : v.a. "2ème session"

X et Y sont indep ? n = 100

X\Y	C_1	C_2	C_3	total
C_1	2	13	6	21
C_2	11	27	13	51
C_3	3	17	8	28
total	16	57	27	100

si X et Y indep : $\alpha = 10\%$

X\Y	C_1	C_2	C_3
C_1	3.36	11.97	5.67
C_2	8.16	29.07	13.77
C_3	4.48	15.96	7.56

$\frac{16\times 21}{100} = 3.36$ $d^2 = 2{,}42$

$D^2\sim\chi^2_4\Rightarrow d_\alpha^2 = 7{,}78\Rightarrow d^2<d_\alpha^2$

⇒ X et Y indep

chaîne de Markov (CM) :

ex : $\pi = [0.24; 0.36; 0.40]$ init

transi : $A = \begin{bmatrix}0.2 & 0.4 & 0.4\\ 0.3 & 0.4 & 0.3\\ 0.2 & 0.3 & 0.5\end{bmatrix}$ Paris $X_0\sim\pi$ $P(X_{t+1}|X_t)$ $\Omega = \{S,N,P\}$

$= \begin{bmatrix}P(S|S) & P(N|S) & P(P|S)\\ P(S|N) & P(N|N) & P(P|N)\\ P(S|P) & P(N|P) & P(P|P)\end{bmatrix}$ $X = (X_0,X_1,\dots,X_n)$ degré 1

$P(X) = P(X_0)P(X_1|X_0)P(X_2|X_1,X_0)$

$= P(X_0)P(X_1|X_0)P(X_2|X_1)$

degré d : $P(X_{t+d}|X_t,\dots,X_{t+d-1})$

séq : Q = NNSNNPPNPSSP

$P(Q) = P(\pi_N)\times P(N|N)\times P(S|N)\times\dots\times P(S|S)\times P(P|S)$

$= 0.36\times 0.4\times 0.3\times\dots\times 0.2\times 0.4$

$= 0.000001493$

$P(S = (S_1,S_2,\dots,S_n)) = \pi_{S_1}\prod_{t=2}^{N} a_{S_{t-1}S_t}$

générer séq aléatoire : U = [0.31, 0.63, 0.92, 0.87, 0.01, 0.35, 0.01, 0.43, 0.55]

$P(X_{t+1}|S) = (S = 0.2; N = 0.4; P = 0.4)$

$P(X_{t+1} = S|S) = P(U<0{,}2) = 0{,}2$

$P(X_{t+1} = N|S) = P(U\in[0.2;0.6]) = 0.4$

$P(X_{t+1} = P|S) = P(U>0{,}6) = 0.4$

t = 1 : $\pi = [0.24; 0.39; 0.4]$

$u_1 = 0{,}31\in[0.24; 0.24+0.39] \in [0.24; 0.6]\Rightarrow N$

t = 2 : $P(A|N) = [0.3; 0.4; 0.3]$

$u_2 = 0{,}63\in[0.3; 0.7]\Rightarrow N$

etc. ⇒ NNPPSNSNN

Longueur moy seq consécutif N ?

Y = nb jours nuageux consécutifs

$P(Y=k) = P(X_1 = N, X_2 = N,\dots, X_k = N, X_{k+1}\neq N)$

$= P(X_1 = N)P(X_2 = N|X_1 = N)\times\dots\times P(X_k = N|X_{k-1} = N)P(X_{k+1}\neq N|X_k = N)$

$= 1\times P(X_{t+1} = N|X_t = N)^{k-1}\times(1-P(X_{t+1} = N|X_t = N))$

$= 0.4^{k-1}(1-0.4) = 0.4^{k-1}\times 0.6$

$= (1-p)^k p$ ⇒ Géométriq

$E(Y) = \frac{1}{p} = \frac{1}{0.6} = \frac{5}{3} = 1{,}667$

ergodiq = irréductible + apériodiq

tous les états sont atteignables à partir de n'importe quel état

(période = pgcd longueur chemin)

A^2 = matrice transit° après 2 jours

$A^2 = \begin{bmatrix}0.24 & 0.36 & 0.4\\ 0.24 & 0.37 & 0.39\\ 0.23 & 0.35 & 0.42\end{bmatrix}$ =1 =1 =1

proba stationnaire :

$\mu A = \mu$ avec $\sum\mu_i = 1$

$\begin{cases}2\mu_1+3\mu_2+2\mu_3 = 10\mu_1\\ 4\mu_1+4\mu_2+3\mu_3 = 10\mu_2\\ 4\mu_1+3\mu_2+5\mu_3 = 10\mu_3\\ \mu_1+\mu_2+\mu_3 = 1\end{cases}$

proba tps diff au bout de 2 jours :

$p = 1-\mu(S)A^2(S,S)-\mu(N)A^2(N,N)-\mu(P)A^2(P,P)$

$p = 1-0.24\times 0.24-0.36\times 0.37-0.4\times 0.42 = 0.64$

$A_2 = \begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.4 & 0.4 & 0.2\\ 0.2 & 0.5 & 0.3\end{bmatrix}$ Marseille

$A^k = \begin{bmatrix}0.4 & 0.38 & 0.22\\ 0.4 & 0.38 & 0.22\\ 0.4 & 0.38 & 0.22\end{bmatrix}$ k>5

$\mu = [0.4\ 0.38\ 0.22]$

Paris ou Marseille ? Q = SSPPNS

$P(Q|\theta_P) = 0.24\times 0.2\times 0.4\times 0.5\times 0.3\times 0.3$

$P(Q|\theta_M) = 0.4\times 0.5\times 0.2\times 0.5\times 0.5\times 0.4$

$\frac{P(Q|\theta_P)}{P(Q|\theta_M)} = \frac{0.4\times 0.5}{0.24\times 0.3} > 1$

⇒ Marseille plus vraisemblable

χ² : k > 0 ddl

$f(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)}x^{k/2-1}e^{-x/2}$

$E(X) = k$ $V(X) = 2k$

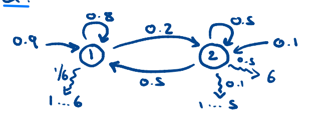
MC cachée (MMC/HMM):

séq états: $S=(s_1,\dots,s_T)$ cachés
séq obs: $X=(x_1,\dots,x_T)$
$\pi_i=p(s_1=i|\lambda)$ proba init
$a_{ij}=p(s_t=j|s_{t-1}=i,\lambda)$ état actu selon préd
$b_i(x_t)=p(x_t|s_t=i,\lambda)$ obs selon état
$p(x_1^T|\lambda)=\Sigma_{s_1^T}(p(x_1^T,s_1^T|\lambda))$
$=\sum_{s_1^T}\prod_t p(s_t|s_{t-1})p(x_t|s_t)$
MMC ordre 1

Récurrence:
init: $\alpha_1(i)=p(x_1^1,s_1=i|\lambda)=\pi_i b_i(x_1)$
itérat°: $\alpha_t(j)=(\sum_i^N\alpha_{t-1}(i)a_{ij})b_j(x_t)$
arrêt: $p(x_1^T|\lambda)=\Sigma\alpha_T(i)$
$s_1^{T*}=\arg\max_{s_1^T}p(x_1^T|\lambda)$ MMC ordre 1
$=\arg\max_{s_1^T}\prod_t p(s_t|s_{t-1})p(x_t|s_t)$

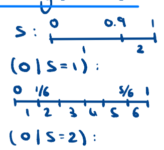
Viterbi: (backforward)
$\delta_t(i)=\max_{s_1^{t-1}}p(s_1^{t-1},s_t=i,x_1^t|\lambda)$
δ = stockage proba
ψ = stockage indices des états
init: $\delta_1(i)=\pi_i b_i(x_1)$ $\psi_1(i)=0$
ité: $\delta_t(j)=(\max_j\delta_{t-1}(i)a_{ij})b_j(x_t)$
$\psi_t(j)=\arg\max_{i\in[1,N]}\delta_{t-1}(i)a_{ij}$
arrêt: $s^*=\max_i\delta_T(i)$
chemin: (à l'envers)
$q_T^*=\arg\max\delta_T(i)$
$q_t^*=\psi_{t+1}(q_{t+1}^*)$

calcul des β:
$\beta_t(i)=p(x_{t+1}^T|s_t=i,\lambda)$ (sym α)
init (arbitraire): $\forall i,\ \beta_T(i)=1$
Récursion: ① transi i vers j
② obs x_{t+1} (à partir de j)
③ continuer sur $\beta_{t+1}(j)$
$\beta_t(i)=\sum_j^N a_{ij}b_j(x_{t+1})\beta_{t+1}(j)$

ex:


$\pi=[0.9\ \ 0.1]$ $B=\begin{bmatrix}\frac16&\cdots&\frac16&\frac16\\0,1&\cdots&0,1&0,5\end{bmatrix}$
$\lambda=\begin{bmatrix}0.8&0.2\\0.5&0.5\end{bmatrix}$

calculer les proba:
. 1 fois dé pipé, 2 fois pipé
. séq 1,2,6 avec états précédents
$P(S=(1,2,2))=0,9\times0,2\times0,5$
commencer avec dé 1 — 1 à 2 — reste avec 2
$P(x=(1,2,6)|S=(1,2,2))$
$=P(x=1|S=1)P(x=2|S=2)P(x=6|2)$
$=1/6\times0.1\times0.5\simeq0.00833$

séq états et obs générées:
0.10: $s_1=1$
0.55: $o_1=4$
0.45: $s_2=1$
0.30: $o_2=2$
0.95: $s_3=2$
0.23: $o_3=3$


calculer proba séq:
$S=\{2,6,6,6,3\}$
$\alpha_t(j)=p(x_1^t,s_t=j|\lambda)$
$=\sum_i p(x_1^t,s_t=j,s_{t-1}=i|\lambda)$
transi $i\to j$ + émission x_t
$=\sum p(s_t=j|\cancel{x_1^t},s_{t-1}=i,\lambda)\times p(x_t|\cancel{x_1^t},s_t=j,\cancel{s_{t-1}=i})\times p(x_1^t,s_{t-1}=i|\lambda)$
$=\sum p(x_1^{t-1},s_{t-1}=i|\lambda)a_{ij}b_j(x_t)$
$=(\sum\alpha_{t-1}(i)a_{ij})b_j(x_t)$
$p(x_1^T|\lambda)=\sum\alpha_T(i)$

t	x_t	$\alpha_t(1)$	$\alpha_t(2)$
0	2	$0.9\times1/6=0.15$	$0.1\times0.1=0.01$
1	6	$\frac16(0.15\times0.8+0.01\times0.5)=0.02083$	$0.5(0.15\times0.2+0.01\times0.5)=0.01750$
2	6	$\frac16(0.02083\times0.8+0.01750\times0.5)=0.00424$	$0.5(0.02083\times0.2+0.01750\times0.5)=0.00646$
3	6	0.00110	0.00204
4	3	0.00032	0.00012

$p(x_1^T)=0.00032+0.00012=0.00044$

évolut° de α_t en fct° t:
$\sum_j\alpha_t(j)=\sum_j(\sum_i\alpha_{t-1}(i)a_{ij})b_j(x_t)$
$=\sum\sum\alpha_{t-1}(i)a_{ij}b_j(x_t)$
$=\sum_i\alpha_{t-1}(i)\underbrace{\sum_j a_{ij}b_j(x_t)}_{<1}$
car $\sum_j a_{ij}=1$ et $b_j(x_i)<1$
donc $\sum_j\alpha_t(j)<\sum_j\alpha_{t-1}(j)$
=> $\alpha_t(j)\searrow$ qd $t\nearrow$ tend vers 0
$\log(\alpha_t(j))$ impossible car 0

problème approx num:
$\alpha_t^+(j)=p(x_t,s_t=j|x_1^{t-1},\lambda)$
$\Omega_t=p(x_t|x_1^{t-1},\lambda)$
$=p(x_t|x_1,\dots,x_{t-1},\lambda)$
$=\sum p(x_t,s_t=j|x_1,\dots,x_{t-1},\lambda)$
$=\sum\alpha_t^+(j)$
$\alpha_{t+1}^+(j)=p(x_{t+1},s_t=j|x_1^t)$
$=p(x_{t+1}|s_{t+1}=j,\cancel{x_1^t})p(s_{t+1}=j|x_1^t)$
$=p(x_{t+1}|s_{t+1}=j)p(s_{t+1}=j|x_1^t)$
$=b_j(x_{t+1})\sum p(s_{t+1}=j,s_t=i|x_1^t)$
$=b_j(x_{t+1})\sum p(s_{t+1}=j|s_t=i,\cancel{x_1^t})\times p(s_t=i|x_1^t)$
$=b_j(x_{t+1})$
$\sum a_{ij}p(s_t=i|x_1^{t-1},x_t)\times\frac{p(x_t|x_1^{t-1})}{p(x_t|x_1^{t-1})}$
$=b_j(x_{t+1})\sum(a_{ij}\frac{p(s_t=i,x_t|x_1^{t-1})}{p(x_t|x_1^{t-1})})$
$=b_j(x_{t+1})\sum_i\frac{\alpha_t^+(i)}{\Omega_t}$
or $p(x_1^T|\lambda)=\prod_{t=1}^T p(x_t|x_1^{t-1},\lambda)$
donc $\ln p(x_1^T|\lambda)=\sum\ln\Omega_t$

t	x_t	$\alpha_t^+(1)$	$\alpha_t^+(2)$	$\ln\Omega_t$
0	2	0.15	0.01	-1.8326
1	6	0.1302⊗	0.1094	-1.4288
2	6	0.1105	0.1684	-1.2766
3	6	0.1031	0.1906	-1.2251
4	3	0.1009	0.0395	-1.9638

⊗ $\frac16\times\frac{(0.15\times0.8+0.01\times0.5)}{0.16}$ ← Ω_{t-1}

séq d'états la plus probable:
$\Delta=\begin{bmatrix}-1.9&-3.9&-5.9&-7.9&-9.5\\-4.6&-4.2&-5.6&-7&-10\end{bmatrix}$
$\log A=\begin{bmatrix}-0.2&-1.6\\-0.7&-0.7\end{bmatrix}$ Ψ?
$\Psi=\begin{bmatrix}0&1&1&1&2\\0&1&2&2&2\end{bmatrix}$
$-1.9-0.2>-4.6-0.7\to1$
$-1.9-1.6>-4.6-0.7\to1$
$-4.1>-4.9\to1$ $\quad -5.5<-4.9\to2$
etc.
$s^*=1$
$s_1^{T*}=(1,2,2,\psi_5(1)=2,1)$
$\psi_4(2)$
ligne 1 (car 1 au préd)
colonne 5
→ état avant le + probable
car $-9.5>-10$ (dans Δ)
regarder état final d'abord

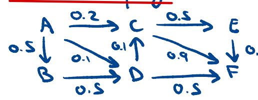
échantillonage:

$E_p(h(x))=\int p(x)h(x)dx$
$\simeq\frac1N\sum_{x_i\sim p(x_i)}h(x_i)$ tirage ∝ selon p
loi des grands nombres
$p(x)=\frac{e^{h(x)}}{\int e^{h(x')}dx'}\propto e^{h(x)}$
$E_p(h(x))=\int\frac{q(x)}{q(x)}p(x)h(x)dx$ $q(x)>0$
$=E_q\left(\frac{p(x)h(x)}{q(x)}\right)$ $p(x)>0$
$\simeq\frac1N\sum_{x_i\sim q(x_i)}\frac{p(x_i)}{q(x_i)}h(x_i)=h_N$
$E(h_N)=E_q(\frac1N\sum\frac{p(x_i)}{q(x_i)}h(x_i))$
$=\frac1N\sum E_q(\dots)=\frac NN(\dots)$
car $(x_i)_i$ m̂ distribut°
$E(h_N)=E_q(h(x))$
$V_q(h_N)=E(h_N^2)-E(h_N)^2$
$=\frac1{N^2}E_q((\Sigma\dots)^2)-\frac1{N^2}E_q(\Sigma\dots)^2$
$=\frac1{N^2}V_q(\Sigma\dots)=\frac1{N^2}\times N\,V_q(\dots)$
$=\frac1N(E_q((\dots)^2)-E_q(\dots)^2)$
$=\frac1N(\int q(x)(\frac{p(x)}{q(x)}h(x))^2dx-\mu^2)$
=> $\arg\min_q V_q(h_N)$ cherche q tq $\min(V_q)$
$=\arg\min_q\int\frac{p(x)^2}{q(x)}h(x)^2dx$
car μ et N sont cte par rapport à q
pour $q=\mathcal N(0;1)=p$: $q(x)=p(x)$
$\int\frac{p(x)^2h(x)^2}{q(x)}dx=\int p(x)h(x)^2dx$
$=\frac1{\sqrt{2\pi}}\int e^{-\frac12x^2}e^{2x}dx=\dots$
avec a = 1/2 et b = 2:
$\frac1{\sqrt{2\pi}}\int e^{\frac{-x^2}{2}+2x}dx=\frac{\sqrt{2\pi}}{\sqrt{2\pi}}e^2=e^2$
pour $q=\mathcal N(1;1)$:
$\int\dots=\frac1{\sqrt{2\pi}}\int e^{2(-\frac{x^2}{2}+x)-\left(\frac{x-1}{2}\right)^2}dx$
a = 1/2 et b = 1: $\int\dots=e$
optimum quand q ∝ ph

Rejection sampling:
$X\sim\mathcal N$ tronquée ∝ $\mathcal N(\frac34;1)$
$f(x)=\begin{cases}C\cdot\frac1{\sigma\sqrt{2\pi}}e^{-\frac12\left(\frac{x-\mu}{\sigma}\right)^2}&a\le x\le b\\0&x<a\text{ et }x>b\end{cases}$
$Z=\frac{X-3}{4}\sim\mathcal N(0;1)$
$P(X\in[0;1])=P(Z\in[-0.75;0.25])$
$=1-(P(Z<-0.75)+P(Z>0.25))$
$=1-(P(Z>0.75)+P(Z>0.25))$
$=1-(0.2266+0.4013)=0.3721$
$c=1/0.3721=2.6874$
pour que $\int_0^1f(x)dx=1$
Monte Carlo: choisir distribut° g(·), facile à échanti, tq ∃ facteur k $\forall x$: $kg(x)\ge f(x)$
① tirer un nb z selon g(·)
② calculer $m_g=kg(z)$
③ tirer un nb u selon dist uniforme sur $[0;m_g]$
④ accepter z comme échantillon si $u\le f(z)$
taux acceptat° lorsque g(·) est loi uni sur [0;1], $kg(\cdot)=\max f(x)$
$P(\text{acceptat°})=E_g(P(\text{accepter}(x)))$
$=\int P(x)P(\text{acc}(x)|x)dx$
$=\int_0^1g(x)\frac{f(x)}{kg(x)}dx$ $\quad P(A)=\int B(A|B)$
$=\frac1k\int_0^1f(x)dx=\frac1k$
or $k=\max f(x)=\frac{c}{\sqrt{2\pi}}$
$\simeq\frac{2.6874}{2.5663}\simeq1.07211$
donc $P(\text{acc})\simeq0.93$
compression: r(x) droite limite inf de f(x), pré-filtrage elem acceptés quand $u\le r(z)$

$r(x)=\min_x f(x)$
$P(\text{pré-}f)=\int_0^1g(x)\frac{r(x)}{kg(x)}dx$
$=\int_0^1r(x)/k\,dx$
$r(x)=f(0)=\frac{c}{\sqrt{2\pi}}e^{-\frac12\left(\frac{-3}{4}\right)^2}$
$=ke^{-9/32}$
donc $P(\text{pré-}f)=\frac{f(0)}{k}=e^{-9/34}\simeq0.75$
75% du tps, on accepte x

Gibbs Sampling:



activat° X dépend de ses parents
$P(X=1|Par(X))=1-\prod_{Y\in par(X)}(1-I(Y=1)\theta_{Y,X})$
I fct° indicatrice retourne 1 si son arg vrai (0 sinon)
$\theta_{Y,X}$ proba activat° Y par une parente activée X
supposons F activé
$x=(x_1,\dots,x_n)$ multivariée
$p(x_j|x_1,\dots,x_{j-1},x_{j+1},\dots,x_n)$
$=\frac{p(x_1,\dots,x_n)}{p(x_1,\dots,x_{j-1},x_{j+1},\dots,x_n)}$
$\propto p(x_1,\dots,x_n)$
① select aléa indice $j\in\{1,\dots,n\}$
② calcul des proba $p(x_1,\dots,x_n)$ avec diff val x_j
③ échanti de x_j ∝ val proba jointe associée
$P(A,B,C,D,E)$
$=P(A)P(B|A)P(C|A,D)P(D,AB)$
$P(E|C)P(F|C,D,E)$ graphe
$P(X=x|Z\setminus X)=\frac{P(X=x,Z\setminus X)}{P(Z\setminus X)}$
$=\frac{P(X=x,Z\setminus X)}{\Sigma_{x'}P(X=x',Z\setminus X)}$
$=(P(X=x|par(X))$
$\times\prod_{Y\in enf(X)}P(Y|par(Y)\setminus X,X=x))\div(\Sigma_{x'}\dots)$
$P(A=1|B,C,D,E,F)\propto P(A=1)$
$P(B|A=1)P(C|A=1,D)P(D|A=1,B)$
$P(B=1|A,C,D,E,F)$
$\propto P(B=1|A)P(D|A,B=1)$
$=\frac{P(B=1|A)P(D|A,B=1)}{(P(B=1|A)P(D|A,B=1)+P(B=0|A)P(D|A,B=0))}$
instation init: (1,1,1,1,1,1)
0.8, 0.3, 0.8, 0.7, 0.9
A B C D E
$P(A=1|B=1,C=1,D=1,E=1,F=1)=1$
car F activé, $1>0.8$ => A=1
$P(B=1|\dots)\propto P(B=1|A=1)$
$\times P(D=1|A=1,B=1)$
$=(1-(1-0.5))\times(1-(1-0.5)(1-.05))$
$\simeq0.275$
$P(B=0|\dots)\propto P(B=0|A=1)$
$\times P(D=1|A=1,B=1)$
$=0.5\times0.1=0.05$
$P(B=1|\dots)=\frac{0.275}{0.275+0.05}$
$=0.982>0.3$ => B reste 1
$P(C=1|\dots)=1$ sinon E ne peut pas être activé
$1>0.8$ donc C reste 1
$P(D=1|\dots)=0.632<0.8$
=> D passe à 0
$P(E=1|A,B,C,D=0,F=1)$
$=0.524<0.7$ => E = 0
si Réversibilité d'une chaîne: $\pi(x)P_{x,y}=\pi(y)P_{y,x}$
alors π(x) dist stationnaire
mq $P(X\setminus F|F=1)P_{x,y}$
$=P(Y\setminus F|F=1)P_{y,x}$
$P_{x,y}=\frac{P(Y)}{P(X)+P(Y)}$ $\quad P_{y,x}=\frac{P(X)}{P(X)+P(Y)}$
soit $v=P(Y\setminus F|F=1)\frac{P(X)}{P(Y)}$
$=\frac{P(Y)}{P(F=1)}\times\frac{P(X)}{P(Y)}=\frac{P(X)}{P(F=1)}$
$=P(X\setminus F|F=1)$
donc chaîne réversible si $\pi(x)=P(X\setminus F|F=1)$

Metropolis-Hasting: MCMC

MCMC: dans le but de simuler des var de dist p, produit chaîne Markov ergo dont dist stationnaire est p
alphabet Λ (lettres, ponctuat°, chiffres, ...)
μ dist proba sur Λ
M matrice stochastiq donnant pour chq lettre de Λ la proba de la suiv
fct d'encodage/décodage τ
txt encodé $T=\tau(T')$ T' un txt en remplaçant chq lettre c de T' par τ(c) (substitut°)
obj: décoder $T=(c_1,\dots,c_{|T|})$
$L(T|\mu,M)=\mu(c_1)\prod_{i=2}^{|T|}M(c_{i-1},c_i)$
$\ln L=\ln(\mu(c_1))+\sum\ln M(c_{i-1},c_i)$
$L(\tau)=L(\tau(T)|\mu,M)$
$P(\tau)=\frac{L(\tau)}{\Sigma L(\tau')}$ → impossible à échantillonner

chaîne de Markov:
mot proposé: $q(\tau'|\tau)$
choisit au hasard w_i et $w_j\in\Lambda$
intervertir encodage:
$\tau'(w_i)=\tau(w_j)$ $\quad\tau'(w_j)=\tau(w_i)$
$q(\tau'|\tau)=0$ si (τ,τ') ne respecte pas la règle de déplac
$=\frac1{|\Lambda|^2}$ sinon
$A_{\tau,\tau'}=q(\tau'|\tau)\alpha(\tau,\tau')$ si $\tau\neq\tau'$
$=q(\tau'|\tau)+\sum_{\tau''\neq\tau}q(\tau''|\tau)(1-\alpha(\tau,\tau''))$ sinon
N fois étapes avec τ_0 init:
① calculer τ à partir de τ_t en changeant 2 lettres c_1 et c_2
$\tau(c)=\tau_t(c)$ $c\neq c_1$ $c\neq c_2$
$\tau(c_1)=\tau_t(c_2)$ $\tau(c_2)=\tau_t(c_1)$
② accepter transi τ_t vers τ
proba $\alpha(\tau_t,\tau)=\min\left(1,\frac{L(\tau(T),\mu,M)\,A(\tau_t,\tau)}{L(\tau_t(T),\mu,M)\,A(\tau,\tau_t)}\right)$
$\tau_{t+1}=\tau$, sinon $\tau_{t+1}=\tau_t$
où $A(\tau,\tau')$ proba transi $\tau\to\tau'$
τ_N tirage aléa selon P
τ_{N+kh} $k\in\mathbb N^*$ fixé, $h\in\mathbb N^*$
$\ln\alpha(\tau,\tau')=\ln\min(1,\frac{P(\tau')q(\tau|\tau')}{P(\tau)q(\tau'|\tau)})$
$=\ln(\min(1,\frac{P(\tau')}{P(\tau)}))$
car $q(\tau|\tau')=q(\tau'|\tau)$ fct° mut sym
$=\ln(\min(1,\frac{L(\tau')}{L(\tau)}))$
$=\ln(\min(0,\ln L(\tau')-\ln L(\tau))$

algo: générer x_{t+1} à partir de x_t
① tirer z selon $P(\cdot|x_t)$
② calculer $\alpha(x_t,z)=\min\left(1,\frac{\pi(z)P(x_t|z)}{\pi(x_t)P(z|x_t)}\right)$
③ tirer nb u selon uni[0;1]
④ $x_{t+1}=\begin{cases}z&\text{si }u\le\alpha(x_t,z)\\x_t&\text{sinon}\end{cases}$

Régression: $Y=f(X,\varepsilon)$ résidus
$Y=\alpha+\beta X+\varepsilon$ $\quad\varepsilon\sim\mathcal N(0;\sigma)$
$P(Y|X,\alpha,\beta,\sigma)$
$=\frac1{\sigma\sqrt{2\pi}}\exp\left(\frac{-1}{2\sigma^2}(Y-(\alpha X^2+\beta))^2\right)$
$E(Y_i)=\alpha+\beta X_i$ $\quad V(Y_i)=\sigma^2$
$Y_i\sim\mathcal N(\alpha+\beta X_i;\sigma)$
$L=\prod_i p(y_i|x_i,\alpha,\beta,\sigma)$
$=\prod\frac1{\sigma\sqrt{2\pi}}\exp\left(\frac{-1}{2\sigma^2}(y_i-(\alpha x_i^2+\beta))^2\right)$
$\ln L=\sum\ln\frac1{\sigma\sqrt{2\pi}}-\sum\frac1{2\sigma^2}(\dots)$
$\ln L=C+\frac1{2\sigma^2}\sum(y_i-(\alpha x_i^2+\beta))^2$
$\frac{\partial\ln L}{\partial\alpha}=\frac{-1}{2\sigma^2}\sum2(x_i^2)(y_i-(\alpha x_i^2+\beta))=0$
=> $\alpha(\Sigma x_i^4)+\beta(\Sigma x_i^2)=\Sigma x_i^2y_i$
$\frac{\partial\ln L}{\partial\beta}=\frac1{2\sigma^2}\sum-2(y_i-(\alpha x_i^2+\beta))=0$
=> $\alpha(\Sigma x_i^2)+\beta N=\Sigma y_i$
$\begin{pmatrix}\Sigma x_i^4&\Sigma x_i^2\\\Sigma x_i^2&N\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix}=\begin{pmatrix}\Sigma x_i^2y_i\\\Sigma y_i\end{pmatrix}$
$\frac{\partial\ln L}{\partial\sigma}=\sum\left(\frac{-1}{\sigma}\right)-\sum\frac12(y_i-\dots)^2(-2/\sigma^3)=0$
=> $\frac{-N}{\sigma}+\frac1\sigma\left(\sum\frac1{\sigma^2}(y_i-(\alpha x_i^2+\beta))^2\right)=0$
$\sigma^2=\frac1N\sum(y_i-(\alpha x_i^2+\beta))^2$

$Y = \alpha_1 X + \beta_1 + \varepsilon$

$Y' = \alpha_2 X^3 + \beta_2 + \varepsilon$

algo EM :

1. init affectation des pts aux 2 modèles
2. param qui max L
3. calcule proba appartenance de chq point à chq modèle

θ_2

$Q_i^0(\theta_1) = p(\theta_1 | y_i \, \theta^0)$

$= p(y_i | \theta_1, \theta^0) \dfrac{p(\theta_1 | \theta^0)}{p(y_i | \theta^0)}$

θ_2

$p(y_i | \theta_1, \theta^0) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(\dfrac{-1}{2\sigma^2} \left(y_i - (\alpha_1^0 x_i + \beta_1^0)\right)^2 \right)$

$\alpha_2^0 x_i^3 + \beta_2^0$

$p(\theta_1 | \theta^0) = p(\theta_1) = 0.5$

$p(y_i | \theta^0) = \sum_j p(y_i, \theta_j | \theta^0)$

$= \sum p(y_i | \theta_j, \theta^0) \, p(\theta_j | \theta^0)$

$\ln L = \sum_i \sum_j Q_i^0(\theta_j) \ln \dfrac{P(y_i, \theta_j)}{Q_i^0(\theta_j)}$

$= \sum \sum Q \ln \dfrac{P(y_i | \theta_j) P(\theta_j)}{Q_i^0(\theta_j)}$

$\dfrac{\partial \ln L}{\partial \alpha_1} = \sum Q_i^0(\theta_1) \times \left(\dfrac{-1}{2\sigma^2}\right) \times (-x_i)(y_i - \alpha_1 x_i - \beta_1) = 0$

$\alpha_1 \sum Q_i^0(\theta_1) x_i^2 + \beta_1 \sum Q_i^0(\theta_1) x_i = \sum Q_i^0(\theta_1) x_i y_i$

Régression linéaire :

$\hat{y}_i = x_i a + b = f(x) \qquad D = \begin{bmatrix} a \\ b \end{bmatrix}$

$X \begin{bmatrix} x_i & 1 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} x_i a + b \\ \vdots \end{bmatrix} Y \qquad \hat{Y} = XD$

$E = \sum_{i=1}^{N} e_i^2 = \sum (f(x_i) - y_i)^2$

$= \sum (a x_i + b - y_i)^2$

$\nabla E = 0$

$\nabla_D E = \begin{bmatrix} \partial E / \partial a \\ \partial E / \partial b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$E = (XD - Y)^T - (XD - Y)$

$\nabla_D E = X^T(XD - Y) + X^T(XD - Y)$

$= 2X^T(XD - Y) = 0$

$2X^T XD - 2X^T Y = 0$

$D = (X^T X)^{-1} X^T Y$

$E(a) = a^2(\sum x_i^2) + a(-2\sum x_i y_i) + y_i^2$

→ parabole convexe

⇒ 1 seul min

descente de gradient :

converge vers un minimum

→ avancem̄ par petit pas ε dans sens du gradient

→ possible si la fct° convexe

sinon plus minima

⇒ min local

ε trop grand ⇒ divergence

$C(w)$ fct° continue dérivable

init w, $t = 1$

repeat | $w_{t+1} = w_t - \varepsilon \dfrac{\partial C}{\partial w}$

| $t = t + 1$

until $C(w)$ n'évolue plus

$E = \sum (a x_i + b - y_i)^2$

$\dfrac{\partial E}{\partial a} = 2 \sum (a x_i + b - y_i) x_i$

$= 2a \sum x_i^2 + 2bN - 2\sum x_i y_i$

$a_{t+1} = a_t + \varepsilon \times 2 \sum_i (a_t x_0 + b_t - y_i) x$

$a^0 > a^*$ (opt)

$\left.\dfrac{\partial E}{\partial a}\right|_{a^*} = 2a^* \sum x_i^2 + C = 0$

$\left.\dfrac{\partial E}{\partial a}\right|_{a_0} = 2a_0 \sum x_i^2 + C > \left.\dfrac{\partial E}{\partial a}\right|_{a^*}$

> 0

$-\varepsilon \dfrac{\partial E}{\partial a} \Rightarrow w \searrow$

$\dfrac{\partial E}{\partial b} = \sum 2(a x_i + b - y_i)$

$\min E \Rightarrow \left.\dfrac{\partial E}{\partial b}\right|_{b^0} = 0$

$b^{t+1} = b^t = -\varepsilon \left.\dfrac{\partial E}{\partial b}\right|_{b^t}$

$\left.\dfrac{\partial E}{\partial b}\right|_{b} - \left.\dfrac{\partial E}{\partial b}\right|_{b^0} = 2bN - 2b^0 N$

$= 2(b - b^0) N$

si $b^t > b^0$: $\left.\dfrac{\partial E}{\partial b}\right|_{b^t} > 0$

$\Rightarrow b^{t+1} < b^t$ (et inversem̄)

non-linéaire :

$\hat{y} = ax + a^2 x + b \qquad a x^3 + b$

$(ax + b)^2 \qquad a \ln x + b$