TOPIC:

DATE:

FILE UNDER:

PAGE:

$$\hat{l}(\tau) \doteq \sum_{i=1}^{n} \log\left[\frac{1}{m}\sum_{k=1}^{m} p(y_i|u_{ik})\frac{\varphi(u_{ik}|\tau)}{\varphi(u_{ik}|\tau_0)}\right]$$

$$u_{ik} \sim \varphi(u_i|\tau)$$

Could use a general candidate dist. $g(u|\tau_0)$

why?

Ideal (efficient) candidate is proportional to

$$p(y_i|u_i)\,\varphi(u_i|\tau)$$

$$= g(u_i|y_i,\tau)$$

with $\tau=\hat{\tau}$, but that solves our problem!

Geyer 1990:

1. Choose τ_0
2. Simulate $u_{i1},\ldots,u_{im} \sim p(u_i|y_i,\tau_0)$ for $i=1,\ldots,n$
3. Maximize $\hat{l}(\tau)$ to get $\hat{\tau}$
4. Set $\tau_0=\hat{\tau}$
5. Goto (2) until $\|\tau_0-\hat{\tau}\|<\varepsilon$

TOPIC:

DATE:

FILE UNDER:

PAGE:

For IS we do not need $\sup f/g < \infty$ but what is required?

Recall Cramer's Theorem (Delta Method)

Let $Y_1, \ldots, Y_n$ be s.t. $\sqrt{n}(\bar{Y} - \mu) \to N(0, \Sigma)$

Where $\mu = \mathbb{E}[Y_i]$. Let $g: \mathbb{R}^k \to \mathbb{R}$ be a differentiable map, then

$$\sqrt{n}(g(\bar{Y}) - g(\mu)) \longrightarrow N(0, g'(\mu)^T \Sigma g'(\mu))$$

Importance Sampling

Let $Y_i = \begin{pmatrix} h(X_i) w_i \\ w_i \end{pmatrix}$ where $X_i \sim g$, $w_i = \frac{f(x_i)}{g(x_i)}$

So $\mathbb{E}_g Y_i = \begin{pmatrix} \mathbb{E} h(X_i) \\ 1 \end{pmatrix} = \begin{pmatrix} \mu_h \\ 1 \end{pmatrix}$.

Note that we can estimate $Var(Y_i)$ consistently with

$$\hat{\Sigma} = \begin{bmatrix} \text{Sample Var}\{h(x_i)w_i\} & \text{Sample Cov}\{h(x_i)w_i, w_i\} \\ \text{Sample Cov}\{h(x_i)w_i, w_i\} & \text{Sample Var}\{w_i\} \end{bmatrix}$$

Let $V(a,b) = \frac{a}{b}$. Then $\hat{\mu}_h = V(\frac{1}{n}\sum h(x_i)w_i, \frac{1}{n}\sum w_i)$.

Also $V'(a,b) = (1/b, -a/b^2)$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Variance estimates for IS

Cramer's Theorem: Suppose $Y_1, \dots, Y_n$ are s.t. $\sqrt{n}(\bar{Y} - \mu) \longrightarrow N(0, \Sigma)$ where $\mu = \mathbb{E}\bar{Y}$. Let $g: \mathbb{R}^K \to \mathbb{R}$ be a differentiable map. Then

$$\sqrt{n}(g(\bar{Y}) - g(\mu)) \longrightarrow N(0, g'(\mu)^T \Sigma g'(\mu))$$

~~Let~~ f is target density

g is proposal density

$\cancel{\mu_h = \mathbb{E}_g h(x)}$

$$\mu_h = \mathbb{E}_f h(x)$$

Let $Y_i = \begin{pmatrix} h(x_i) w_i \\ w_i \end{pmatrix}$ where $X_i \sim g$

and $w_i = \frac{f(x_i)}{g(x_i)}$

$\cancel{\mathbb{E}_g[m(Y)] = \mathbb{E}_g}$

$\cancel{\mathbb{E}_g Y_i = \begin{pmatrix} \mathbb{E}_g\left[h(x_i)\frac{f(x_i)}{g(x_i)}\right] \\ \mathbb{E}_g\left[\frac{f(x_i)}{g(x_i)}\right] \end{pmatrix} = \begin{pmatrix} \mathbb{E}_f h(x) \\ 1 \end{pmatrix} = \begin{pmatrix} \mu_h \\ 1 \end{pmatrix}}$

TOPIC:

DATE:

FILE UNDER:

PAGE:

$\mathbb{E}\bar{Y} = \begin{pmatrix} \mu_h \\ 1 \end{pmatrix}$

$m(\mathbb{E}\bar{Y}) = \mu_h$

Let $m\begin{pmatrix} a \\ b \end{pmatrix} = \frac{a}{b}$. $m'\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1/b \\ -a/b^2 \end{pmatrix}$

$$m(\bar{Y}) = m\begin{pmatrix} \frac{1}{n}\sum h(x_i) w_i \\ \frac{1}{n}\sum w_i \end{pmatrix} = \frac{\frac{1}{n}\sum h(x_i) w_i}{\frac{1}{n}\sum w_i}$$

Note $\mathbb{E}_g Y_i = \begin{pmatrix} \mathbb{E}_g\left[h(x_i)\frac{f(x_i)}{g(x_i)}\right] \\ \mathbb{E}_g\left[\frac{f(x_i)}{g(x_i)}\right] \end{pmatrix} = \begin{pmatrix} \mathbb{E}_f\left[h(\cancel{x_i})\right] \\ 1 \end{pmatrix}$

$$= \begin{pmatrix} \mu_h \\ 1 \end{pmatrix}$$

We can estimate $\mathrm{Var}(Y_i)$ consistently with

$$\hat{\Sigma} = \begin{pmatrix} \hat{\mathrm{Var}}\{h(x_i)w_i\} & \hat{\mathrm{Cov}}\{h(x_i)w_i, w_i\} \\ \hat{\mathrm{Cov}}\{h(x_i)w_i, w_i\} & \hat{\mathrm{Var}}\{\cancel{h(x_i)}\, w_i\} \end{pmatrix}$$

Cramer's Thm says:

$$\sqrt{n}\left(\frac{m(\bar{Y}) - \mu_h}{m'(\bar{Y})^T \hat{\Sigma}\, m'(\bar{Y})}\right) \longrightarrow N(0,1)$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

$\mathbb{E}X^2 - (\mathbb{E}X)^2 \geq 0$

$(\mathbb{E}X)^2 \leq \mathbb{E}X^2$

Cramer's Theorem says:

$$\sqrt{n}\left(\frac{V(\bar{y}) - \mu_h}{V'(\bar{y})^T \hat{\Sigma} V'(\bar{y})}\right) \longrightarrow N(0,1)$$

And

$$V'(\bar{y})^T \hat{\Sigma} V'(\bar{y}) =$$

$$n\left(\frac{\sum h(x_i)w_i}{\sum w_i}\right)^2 \left(\frac{\sum h(x_i)^2 w_i^2}{\left(\sum h(x_i)w_i\right)^2} - 2\frac{\sum h(x_i)w_i^2}{\left(\sum h(x_i)w_i\right)\left(\sum w_i\right)} + \frac{\sum w_i^2}{\left(\sum w_i\right)^2}\right)$$

$\Rightarrow$ We need

$$\mathbb{E}_g\left[h(x)^2 W^2\right] = \mathbb{E}\left[\left(h(x)\frac{f(x)}{g(x)}\right)^2\right] < \infty$$

$$\mathbb{E}\left[W^2\right] = \mathbb{E}\left[\left(\frac{f(x)}{g(x)}\right)^2\right] < \infty$$

$$\mathbb{E}\left[h(x)W^2\right] = \mathbb{E}\left[h(x)\left(\frac{f(x)}{g(x)}\right)^2\right] < \infty$$

All true if $\frac{f(x)}{g(x)}$ is bounded, which is required for RS.

TOPIC: Bnstat 778

DATE: 2007-11-30

FILE UNDER:

PAGE:

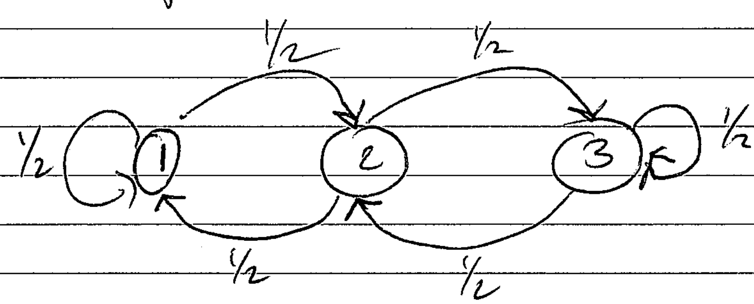
Markov Chains

⟹ A Markov Chain is a stochastic process $\{X_i\}$ which satisfies

$$P(X_i \mid X_{i-1}, \ldots, X_0) = P(X_i \mid X_{i-1})$$

so that the current state of the chain X_i only depends on the previous value X_{i-1}.

⟹ The possible values $\{X_i\}$ can take is called the state space



$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$ Transition matrix (Kernel)

$$P(X_n = j \mid X_{n-1} = i) = P_{ij}$$

Suppose we start the chain with initial dist π_0, what is distribution after n iterations?

$(0, 0, 1)$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Given transition matrix P

$$\mathbb{P}(X_n = j \mid X_0 = i) = (P^n)_{ij}$$

where $P^n = \underbrace{P \times P \times P \times \cdots \times P}_{n \text{ times}}$

$$\pi_n = \pi_0 P^n = \pi_0 \underbrace{P \times P \times P \times \cdots \times P}_{n \text{ times}}$$

Let π^* be a vector (distribution) such that

$$\pi_* P = \pi_*$$

Call π_* a stationary distribution and a MC ~~that~~ ~~[illegible]~~ is said to be stationary if it reaches ~~has~~ this distribution.

Basic limit theorem says that under some conditions $\|\pi_* - \pi_n\| \longrightarrow 0$ as $n \longrightarrow \infty$.

No matter how we start the chain (π_0) π_n approaches π_*.

TOPIC:

DATE:

FILE UNDER:

PAGE:

Assumptions:

1. Irreducible: $P_{ij}^{n} > 0$ for some n and for every i, j

All pairs of i, j are accessible from one another

2. Aperiodic: the chain does not make deterministic visits to a subset of the state space.

ex: $X_0 = 0$ $\quad X_n = X_{n-1} + \varepsilon_n$

$$\varepsilon_n = \begin{cases} 1 & 1/2 \\ -1 & 1/2 \end{cases}$$

For n even, the chain hits subset of even #s
For n odd " " " " " odd #s

3. ~~Positive Recurrent: Every state is visited i.o., The waiting time between visits from state any 2 states is finite~~

3. Stationary dist exists

TOPIC:

DATE:

FILE UNDER:

PAGE:

$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Let $X_0, X_1, \ldots$ be an irreducible, aperiodic, Markov Chain w/ stationary dist. π.

Let $X_0 \sim \pi_0$, some arbitrary distribution. Then

$$\pi_n(i) \longrightarrow \pi(i) \quad \forall i \text{ as } n \longrightarrow \infty$$

~~Irreducible — all states communicate w/ each other~~

Time Reversibility

A Markov Chain is time reversible if

$$(X_0, X_1, \ldots X_n) \stackrel{D}{=} (X_n, X_{n-1}, \ldots X_0)$$

If $\{X_n\}$ is TR, then

$$(X_0, X_1) \stackrel{D}{=} (X_1, X_0)$$

$$\Rightarrow \quad X_0 \stackrel{D}{=} X_1$$

$$\Rightarrow \quad \pi_1 = \pi_0.$$

Since $\pi_1 = \pi_0 P$, the initial dist. is stationary

Let $\pi = \pi_0$. TR chains have

$$\underbrace{\pi_i P_{ij}}_{\text{go from } i \to j} = \underbrace{\pi_j P_{ji}}_{\text{go from } j \to i}$$