Conditional Monte Carlo

(Another variance reduction technique)

Let α be distributed with pdf $f(\alpha)$.

Want to estimate $\theta = E[\phi(\alpha)]$

$$= \int_{\mathcal{A}} \phi(\alpha) f(\alpha)\, d\alpha$$

But either $\mathcal{A}$ is an awkward space or $f(\alpha)$ is hard to sample. Consider

$$\mathcal{C} = \mathcal{A} \times \mathcal{B} \qquad c = (a, b)$$

$\mathcal{C}$ — new space

$$\gamma = (\alpha, \beta)$$

γ is distributed with pdf $h(c)$

Define transformation Jacobians:

$$J(c) = J(a,b) = \frac{da\,db}{dc}$$

Let $g(c) = g(a,b)$ defined on $\mathcal{C}$

with $$G(a) = \int_{\mathcal{B}} g(a,b)\,db \neq 0 \quad \forall a \in \mathcal{A}$$

Assume $h(c) \neq 0$ (the pdf). Define the weight

$$w(c) = f(a)\,g(c)\,J(c)\,(G(a)\,h(c))^{-1}, \text{ so}$$

$$\int_{\mathcal{C}} \phi(a)\,f(a)\,da = \int_{\mathcal{A}} \frac{\phi(a)\,f(a)}{G(a)}\,da \int_{\mathcal{B}} g(a,b)\,db$$

$$= \int_{\mathcal{A}\times\mathcal{B}} \frac{\phi(\alpha) f(\alpha) g(c)}{G(\alpha) h(c)} h(c)\, d\alpha\, db = \int_{\mathcal{C}} \phi(\alpha) w(c) h(c) \frac{d\alpha db}{J(c)}$$

$$= \int_{\mathcal{C}} \phi(\alpha) w(c) h(c)\, dc$$

Thus $t = \phi(\alpha) w(\gamma)$ sampled with pdf $h(c)$ is an unbiased estimator of θ.

Notes:

Can choose $\mathcal{B}$ and $h(c)$ for convenience

$g(a,b) = g(c)$ is like an importance function.

Now assume $f(\alpha) = f(a, b_0)$ is the conditional pdf of $h(c)$ given $b = b_0$. Let $\psi(b)$ be the "$\mathcal{B}$" part of the pdf so that

$$h(c)\, dc = f(a,b)\, da\, db$$

$$\Rightarrow \quad J(c) = h(c)\,(f(a,b)\,\psi(b))^{-1}$$

Let $b = b_0$

$$J(a, b_0) = h(a, b_0)\left[f(a)\,\psi(b_0)\right]^{-1}$$

Recall: $w(c) = f(a)\,g(c)\,J(c)\left[G(a)\,h(c)\right]^{-1}$ so

$$w(c) = \frac{h(a, b_0)}{h(a, b_0)} \cdot \frac{J(a, b)}{J(a, b_0)} \cdot \frac{g(a, b)}{\psi(b_0)\,G(a)}$$

only "a" dependence

Let $\gamma = (\alpha, \beta)$ be distributed over $\mathcal{C}$ with $h(c)$, then

$t = \phi(\alpha)\,w(\gamma)$ is an unbiased conditional estimator of θ given $b = b_0$.

Now suppose $\mathcal{C} = \mathcal{A} \times \mathcal{B}$ is Euclidean and

$$\text{Dim}(\mathcal{B}) = 1.$$

So $c = (a, b) = c = (a, b(c))$

b: scalar

Suppose $b(c)$ is a positive-homogeneous function of the first degree

$b(\rho c) = \rho^d b(c)$, $d = 1$ (Spherical symmetry)

the equation $b(c) = b_0 \Rightarrow b(\rho c) = \rho b_0$

note $b(0) = 0\, b(c) = 0 \Rightarrow$ Concentric regions of co-dimension 1 in $\mathcal{C}$.

$\forall c$ define $\lambda = b_0 / b(c)$ be a scale factor for c

Let $\mathcal{A}$ be the space of direction is $\mathcal{C}$, i.e. unit vectors. Then $c(a, b)$ ← unit vector from c to the origin

So $\rho c = (a, \rho b) = (a, \rho b_0/\lambda)$ ← so that the straight line from the origin cuts the surface $b(c) = b_0$ at the point

$$(a, b_0) = (a, \lambda b_0/\lambda) = \lambda c$$

When da sweeps out a solid angle (over the sphere) db a straight line, thus

dc is a cone and

$$J(c) = \frac{da\,db}{dc} = \Lambda(a)\, b^{m-1}$$

So $$\frac{J(a,b)}{J(a,b_0)} = \frac{\Lambda(a)\, b^{m-1}}{\Lambda(a)\, b_1^{m-1}} = \lambda^{m-1}$$ as above.

Assume $g(a,b)$ is a function of b and b_0 only i.e.

$$g(a,b) = \varepsilon(\lambda, b_0)$$

Also $$G(a) = \int_{\partial B} g(a,b)\, db = \text{constant} = 1$$

by scaling.

So $$w(c) = \frac{h(\lambda c)}{h(c)}\, \lambda^{m-1}\, \frac{\varepsilon(\lambda, b_0)}{\psi(b_0)}$$

This is a special weight Now an application.

Consider $z_1, z_2, \ldots, z_n$ $(n \geq 4)$ a set of n i.i.d. $N(0,1)$ variables. Sort them so that $z_1 \to z_n$ are ranked in increasing magnitude

$\beta = z_n - z_1$ the maximum difference

$\sigma = \max(z_n - z_2, z_{n-1} - z_1)$

$\tau = \max(z_n - z_3, z_{n-1} - z_2, z_{n-2} - z_1)$

Given $\varepsilon > 0$ and small, $S, T \Rightarrow$

$$P[\sigma \geq S \text{ or } \tau \geq T] < \varepsilon$$

Find B so that

$$P[\sigma \geq S \text{ or } \tau \geq T \text{ or } \beta \geq B] = \varepsilon \quad (*)$$

Approach 1: Direct simulation. Say

$\varepsilon = 0.05$, $S = 3.31$, $T = 3.17$, $n = 4$

only $N\varepsilon$ events with B chosen correctly will satisfy $*$ as an inequality.

Thus this is a low efficiency, high variance approach

$$P[\sigma \geq S \text{ or } \tau \geq T \text{ or } \beta \geq B] \triangleq Q(B) = \varepsilon$$

$$Q(b) = \int_T^b P[\sigma \geq S \text{ or } \tau \geq T \mid \beta = b_0]\, \psi(b_0)\, db_0 + \int_b^\infty \psi(b_0)\, db_0$$

$\psi(b) =$ pdf for β and can do this integral

Consider $P[\sigma \geq S \text{ or } \tau \geq T \mid \beta = b_0]\, \psi(b_0)$. If this can be estimated then

a) Evaluate for various values of b

b) " second integral

c) Tabulate $Q(b)$ & interpolate

d) invert $Q(b) = \varepsilon$

Define $\eta = \gamma + \bar{\xi}(1,1,1,\ldots,1)$

(can use this and previous results to solve this problem. (See H&H pp. 81-84)

Note: This process is complicated but has similarities to other approaches where what is known is segregated from what must still be sampled.

Markov Chains

Assume a system has discrete states $\{S_i\}$ and we sample the system at discrete times $\{t_j\}$

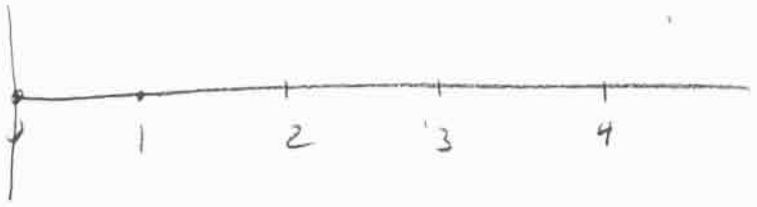
$$X_t = \text{state at time } t$$

$$P\left[X_{t_{n+1}} = S_j \mid X_{t_1} = S_{i_1}, \ldots, X_{t_n} = S_{t_n}\right] \circledast$$

If $\circledast = P[X_t = S_j \mid X_{t-1} = S_{i_{t-1}}]$, then the system is memory-less or has the Markov property or is Markoffian.

$$p_{ij} = P[X_t = S_j \mid X_{t-1} = S_i] = P[S_i \to S_j]$$

An example



$$p_{ij} = \begin{cases} \frac{1}{2} & |i-j| = 1 \ , \ i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$p_{0j} = \begin{cases} 1 & \text{if } j = 1 \\ 0 & \text{otherwise} \end{cases}$$

Reflecting symmetric random walk.

Another example (to work out yourselves!)

$\Omega = [0,1]^2$ (unit square) discretized with

points $(x_i, y_j) = (i\Delta, j\Delta)$, $\Delta = \frac{1}{N}$, $0 \le i,j \le N$.

1. The $\partial\Omega$ is absorbing
2. $P(ij \to ke) = \begin{cases} 0 & \text{if same} \\ \frac{1}{4} & \text{if neighbors} \\ 0 & \text{if not neighbors} \end{cases}$

Problem Produce an ordering on the ij and then
$\downarrow$
k

define the transition probabilities.

Consider f a function on the states $\{S_i\}$

$$\overline{I}_t = \frac{1}{t}\sum_{s=1}^{t} f(X_s)$$

Consider all the "frequency" measure

$$\Delta_t(i) = \frac{1}{t} \sum_{s=1}^{t} \chi_s(i)$$

$$\chi_t(i) = \begin{cases} 1 & \text{if } X_t = S_i \\ 0 & \text{otherwise} \end{cases}$$

"average time in state i"

Consider $Y_t = \sum_i \Delta_t(i) f(S_i)$ ⊛

Note: The presumption that ⊛ $= \frac{1}{t}\sum_{s=1}^{t} f(X_s)$ as $t \to \infty$ is called the "Ergodic Hypothesis"

Define $p_{ij}^{(n)} = P[X_t = S_j \mid X_{t-n} = S_i]$ so we know by the Markoffian assumption

$$p_{ij}^{(1)} = p_{ij} \quad ; \quad p_{ij}^{(n+1)} = \sum_k p_{ik}^{(n)} p_{kj}$$

$$[\bar{\bar{P}}]_{ij} = p_{ij} \quad ; \quad [\bar{\bar{P}}^{(n)}]_{ij} = p_{ij}^{(n)}$$

So $\bar{\bar{P}}^{(n)} = \bar{\bar{P}}^n$. Also $\bar{\bar{P}}$ is a stochastic matrix

$$\forall_{i,j} \quad p_{ij} \geq 0$$

$$\sum_j p_{ij} = 1, \forall_i$$ (note: add extra state if < 1, an absorbing state)

Can prove that $\bar{\bar{P}}^n$ is stochastic as well by induction.

First Passage Probabilities

$$f_{ij}^{(n)} = P\left[X_t = S_j, X_{t-1} \cdots, X_{t-n+1} \neq S_j \mid X_{t-n} = S_i\right]$$

probability of starting at i and reaching j for the first time in n steps (transitions)

$$f_{ij}^{(1)} = p_{ij}, \quad f_{ij}^{(n+1)} = p_{ij}^{(n+1)} - \sum_{r=1}^{n} f_{ij}^{(r)} p_{jj}^{(n-r+1)}$$

Mean First Passage Time

$$m_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)}$$

assuming you can get from i to j: $\sum_{n=1}^{\infty} f_{ij}^{(n)} = 1$

Definitions:

S_i is recurrent if $\sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$

S_i is null if $m_{ii} = \infty$

S_i is positive if $m_{ii} < \infty$

S_i is periodic with period d if $p_{ii}^{(n)} \neq 0$ only when $n | d$. $d = 1 \rightarrow S_i$ is aperiodic

S_i, S_j are "mutually accessible" if $\exists\, m, n \rightarrow$

$$p_{ij}^{(n)} \neq 0, \quad p_{ji}^{(m)} \neq 0$$

then S_i and S_j belong to the same class, and classes are either non-recurrent, positive, or null.

An irreducible Markov chain have all same class states

I. If all states are positive & aperiodic

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j = m_{jj}^{-1} \quad (*)$$

an π_j satisfies $\pi_j > 0$ $\sum_j \pi_j = 1$ &

$$\pi_j = \sum_i \pi_i p_{ij}.$$

$$\bar{\pi} = \bar{\pi}\bar{\bar{P}}$$ (stationary distribution)

II. If $\bar{\pi}$ is such that and the states are aperiodic then $\bar{\pi}$ is unique and $(*)$ holds

III. If the states are positive then

$$\lim_{t\to\infty} \Delta_t(j) \longrightarrow m_{jj}^{-1}$$ with probability one

IV. If the above conditions hold then if

$$m_{ii}^{(2)} = \sum_n n^2 f_{ii}^{(n)} < \infty \quad \text{and}$$

$$E\left[\left(\Delta_t c_{ij} - m_{ij}^{-1}\right)^2\right] = \mathcal{O}\left(\frac{1}{t}\right)$$

V. If conditions above hold then if f is bounded then $\left\{ \frac{1}{t} \sum_{s=1}^{t} f(X_s) - \sum_j m_{jj}^{-1} f(S_j) \right\}$ is $\mathcal{O}\left(\frac{1}{t}\right)$. [Chung]

Solution of Linear Equations

$$\bar{x} = \bar{a} + \bar{\bar{H}}\bar{x}$$

Define: $\|\bar{\bar{H}}\| = \max_i \left(\sum_j |h_{ij}| \right) = \gamma$

Curtiss results: (Paper included)

When is it cheaper to solve x_i via MC?

1. If $\gamma > 1$, MC method fails

(An interesting question:

Given: $\bar{\bar{A}}\,\bar{y} = \bar{b}$ when can you rewrite

this as $\bar{x} = \bar{q} + \bar{\bar{H}}\,\bar{x}$ $\quad \|\bar{\bar{H}}\| < 1$?)

2. If $\gamma = 0.9$ MC is less efficient

a) $n \leq 554$ to 1%

b) $n \leq 84$ to 10%

3. If $\gamma = 0.5$

a) $n \leq 151$ to 1%

b) $n \leq 20$ to 10%

Accuracy is the % of max fail.

Method of von Neumann + Ulam

Define $\bar{\bar{P}}$: $p_{ij} \geq 0$, $\sum_j p_{ij} \leq 1$

$p_{ij} > 0$ when $h_{ij} \neq 0$

define $p_i = 1 - \sum_j p_{ij}$ (could add $n+1$'s state as the absorbing state!)

$$v_{ij} = \begin{cases} h_{ij}/p_{ij} & \text{when } p_{ij} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

$\bar{\bar{P}}$ is the transition probability matrix for a random walk on n states with absorbtion.

Suppose we have walked $\gamma = (i_0, i_1, \cdots, i_k)$, and terminated

$$P[i_{m+1} = j \mid i_m = i, k > m] = p_{ij} \quad \text{continuation}$$

$$P[k = m \mid i_m = i, k > m-1] = p_i \quad \text{termination}$$

$$V_m(\gamma) = \prod_{j=1}^{m} V_{i_{j-1} i_j} \qquad m \le k$$

$$X(\gamma) = V_k(\gamma) \frac{q_{i_k}}{p_{i_k}}$$ (walk terminates in k steps at i_k)

$$E\left[X(\gamma) \mid i_0 = i\right] =$$

$$\sum_{k=0}^{\infty} \left(\sum_{i_1} \cdots \sum_{i_k} p_{i i_1} \cdots p_{i_{k-1} i_k} p_{i_k} V_{i i_1} \cdots V_{i_{k-1} i_k} \frac{q_{i_k}}{p_{i_k}} \right)$$

$$= \sum_{k=0}^{\infty} \left(\sum_{i_1} \cdots \sum_{i_k} h_{i i_1} \cdots h_{i_{k-1} i_k} q_{i_k} \right)$$

$$= \left[\bar{\bar{I}}\right]_i + \left[\bar{\bar{H}}\,\bar{q}\right]_i + \left[\bar{\bar{H}}^2 q\right]_i + \cdots$$

$$= \left[\left\{ \bar{\bar{I}} + \bar{\bar{H}} + \bar{\bar{H}}^2 + \cdots \right\} \bar{q} \right]_i = \left[\left\{ \sum_{i=0}^{\infty} \bar{\bar{H}}^i \right\} \bar{q} \right]_i$$

$$= \left[\left(\bar{\bar{I}} - \bar{\bar{H}} \right)^{-1} \bar{q} \right]_i = \left[\bar{X}\right]_i$$

Wasow considered the modified statistic:

$$X^*(\gamma) = \sum_{m=0}^{k} V_m(\gamma)\, q_{im}$$

$$\mathrm{Var}[X^*(\gamma)] < \mathrm{Var}[X(\gamma)] \quad \text{when}$$

$p_r \leq \frac{r_j}{2 - r_j}$ ← prob. of never revisiting j.

Adjoint Method: (Backwards walking)

Define $\bar{\bar{Q}}$ as follows: $q_{ij} \geq 0$, $\sum_j q_{ij} \leq 1$,

$q_{ij} > 0$ when $h_{ji} \neq 0$,

$$q_i = 1 - \sum_j q_{ij}.$$

$$w_{ij} = \begin{cases} h_{ji}/q_{ij} & \text{when } q_{ij} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Define $\pi_i \geq 0$, $\sum_i \pi_i = 1$ $\alpha_i = \frac{q_i}{\pi_i}$

(starting probability)

$$W_m(\gamma) = \alpha_{i_0} \prod_{j=1}^{m} w_{i_{j-1} i_j}$$

Choose $i_0 = i$ with probability π_i and then

$$[\bar{x}]_j = E\left[W_k(\gamma) \frac{\delta_{i_k j}}{q_{i i_k}} \right] =$$

$$= E\left[\sum_{m=0}^{k} W_m(\gamma) \delta_{i_m j} \right]$$

Note, the adjoint allows for estimation of all of the elements of $[\bar{x}]$ by using different j's above.

Monte Carlo Methods for Linear Equations (supplemental)

Variance of the Ulam/von Neumann method.

To solve $\bar{x} = \bar{\bar{H}}\bar{x} + \bar{a}$

we get:

$\gamma = (i_0, i_1, \ldots, i_k)$, p_{ij} +

$v_{ij} = \frac{h_{ij}}{p_{ij}}$, $p_i = 1 - \sum_j p_{ij}$

$X(\gamma) = v_{i_0 i_1} v_{i_1 i_2} \cdots v_{i_{k-2} i_{k-1}} \frac{a_{i_k}}{p_{i_k}}$

$E[X(\gamma) \mid i_0 = i] = \left[(\bar{\bar{I}} - \bar{\bar{H}})^{-1} \bar{a}\right]_i = [\bar{x}]_i$

What is $\mathrm{var}[X(\gamma) \mid i_0 = i]$?

Let $\bar{\bar{R}}$ be s.t.

$r_{ij} = h_{ij} v_{ij} = p_{ij} v_{ij}^2$

Assume $\|R\| < 1$, then the Neumann series

$$\sum_{i=0}^{\infty} R^i \rightarrow (I - R)^{-1} = T$$

Then $Var[X(\gamma) \mid i_0 = i, i_n = j] =$

$$t_{ij} p_j^{-1} - (\bar{\bar{B}}^{-1})_{ij}^2 = \sigma^2_{ij}$$

where $\bar{\bar{B}}^{-1} = (\bar{\bar{I}} - \bar{\bar{H}})^{-1}$

$$\sigma^2_{ij} = E\left[\left(X(\gamma)_{ij} - (\bar{\bar{B}}^{-1})_{ij}\right)^2\right]$$

$$= E(X(\gamma)^2_{ij}) - (\bar{\bar{B}}^{-1})^2_{ij}$$

$v_{ij}^2 p_{ij}^2$

$\|$

$r_{ij} p_{ij}$

$$\cdots \sum_{i=0}^{\infty} R^i = T$$

kt.

$$= t_{ij} p_j^{-1} - (\bar{\bar{B}}^{-1})_{ij}^2$$

Note that as $\rho_j \to 0$ $\sigma^2_{ij} \to \infty$.

Two solutions

1) $\rho_j = 1 - \sum_k \rho_{jk}$, can scale $\sum \rho_{jk}$ to anything to choose a good value for ρ_j, say $\frac{1}{2}$

2) Can use a biased variant of the Ulam/von Neumann method

Construct p_{ij} as before, but with $\sum_k p_{jk} = 1$, again can be done by scaling. Choose m and construct $\gamma = (i_0, i_1, i_2, \ldots, i_m)$, and

$$X(\gamma) = V_{i_0 i_1} \cdots V_{i_{m-1} i_m} a_{i_m}$$

$$E[X(\gamma) \mid i_0 = i] \approx [\bar{x}]_i$$

$$= \left[\left(\sum_{\lambda=0}^{m} \bar{\bar{H}}^{\lambda}\right)\bar{x}\right]_i$$ so there is a truncation error, but no

problem w/ variance exploding.

Implementation & Complexity

consider $i \to j$ transition

compute $\sum_{k=1}^{k} p_{ik}$ to get PDF

for transition, construct vector of percentiles for transition for each state. In compressed form needs $\mathcal{O}(NZ)$, NZ = # of nonzeros in $\bar{\bar{H}}$.

Costs 1 RV/step, $\mathcal{O}(|\ln(1-\rho^*)|^{-1})$ steps

where $\rho^* = \min_{\delta} p_{ij}$

Note this is independent of n. So solving the system completely is $\mathcal{O}(n)$ disregarding variance and accuracy.

Sequential Monte Carlo:

(iterative refinement of the above)

Suppose we have an estimate $\hat{\bar{x}}$ of $\bar{x}$, define

$$\bar{y} = \bar{x} - \hat{\bar{x}}$$

$$\bar{d} = \bar{a} + \bar{\bar{H}}\hat{\bar{x}} - \hat{\bar{x}}, \quad \text{recall} \quad \bar{x} = \bar{a} + \bar{\bar{H}}\bar{x}$$

$$-\hat{\bar{x}} = -\bar{a} + \bar{\bar{H}}\hat{\bar{x}} + \bar{d}$$

$$\bar{x} - \hat{\bar{x}} = \bar{a} + \bar{\bar{H}}\bar{x} - \bar{a} + \bar{\bar{H}}\hat{\bar{x}} + \bar{d}$$

$$\bar{y} = \bar{a} + \bar{\bar{H}}\bar{y} - \bar{a} + \bar{d} = \bar{d} + \bar{\bar{H}}\bar{y}$$

Can redo with $\bar{\bar{H}}$ & $\bar{d}$ for $\bar{y}$.

If one improves at regular intervals, after the s^{th} improvement we get a reduction of

$$\left(\frac{\sigma}{1-\sigma}\right)^{s-1}, \quad \sigma = \max_i \left\{ \sum_j \frac{h_{ij}^2}{p_{ij}} \right\}$$

$\sigma < \frac{1}{2}$ is good and can be found only when $\| \bar{\bar{H}} \| < \frac{1}{2}$.

Fredholm Integral Equations of the First Kind.

Consider $f(\bar{x}) = g(\bar{x}) + \int K(\bar{x}, \bar{y}) f(\bar{y}) \, d\bar{y}$

g: given

K: Kernel function

Methods of solution:

1) Replace $\int$ with summation (quadrature/cubature rule) $\rightarrow$ Matrix equation for unknown node values.

(Modern equivalent is to use FEM to discretize. Also BEM are applicable as well as BEM + FMM.)

2) Assume $\{\phi_i\}$ is a complete set of functions.

$$g(\bar{x}) = \sum_{i=0}^{\infty} \left[\int g(\bar{x}) \phi_i(\bar{x}) d\bar{x} \right] \phi_i(\bar{x})$$

↑ Known (g_i)

$$f(\bar{x}) = \sum_{i=0}^{\infty} x_i \phi_i(\bar{x})$$

↑ unknown

We know $\int K(\bar{x}, \bar{y}) \phi_j(\bar{y}) d\bar{y} = \sum_{i=0}^{\infty} h_{ij} \phi_i(\bar{x})$

↑ Known

Note: in many situations $K(\bar{x}, \bar{y})$ is easily represented via the $\{\phi_i\}$'s

$$K(\bar{x}, \bar{y}) = \sum_{i=0}^{\infty} \phi_i(\bar{x}) \phi_i(\bar{y}) \lambda_i^{-1}$$

for symmetric kernels ($K(\bar{x}, \bar{y}) = K(\bar{y}, \bar{x})$)

λ_i are eigenvalues of the equation

$$\phi_i(\bar{x}) = \lambda_i \int K(\bar{x}, \bar{y}) \phi_i(\bar{y}) \, d\bar{y}$$

Thus method 2) is a "spectral" method.

$$\Rightarrow \quad \bar{x} = \bar{a} + \bar{\bar{H}} \bar{x}$$

$$[\bar{x}] = x_i \,, \quad [\bar{a}]_i = a_i \,, \quad [\bar{\bar{H}}]_{ij} = h_{ij}$$

Linear system guaranteed to converge.

Note $\bar{\bar{I}} + \bar{\bar{H}} + \bar{\bar{H}}^2 + \cdots \approx (\bar{\bar{I}} - \bar{\bar{H}})^{-1}$

↑ called the Neumann series

converges $\| \bar{\bar{H}} \| < 1$.

3) Functional iteration (recall th. proof of existence and uniqueness for ODEs, Picard iteration)

$f_0(\bar{x})$ is a guess for $f(\bar{x})$, define

$$f_{n+1}(\bar{x}) = g(\bar{y}) + \int K(\bar{x}, \bar{y}) f_n(\bar{y})\, d\bar{y}$$

If $\lim_{n\to\infty} f_n(\bar{x}) \longrightarrow f(\bar{x})$ then $f(\bar{y})$ satisfies the integral equation.

Symbolically: $\int K(\bar{x}, \bar{y})\, d\bar{y} \Longrightarrow \bar{\bar{K}}$

$$f_0$$
$$f_1 = g + \bar{\bar{K}} f_0$$
$$f_2 = g + \bar{\bar{K}} f_1 = g + \bar{\bar{K}} g + \bar{\bar{K}}^2 f_0$$
$$f_3 = g + \bar{\bar{K}} f_2 = g + \bar{\bar{K}} g + \bar{\bar{K}}^2 g + \bar{\bar{K}}^3 f_0$$

In general $f_n = \left(\sum_{i=0}^{n-1} \bar{\bar{K}}^i \right) g + \bar{\bar{K}}^n f_0$

Note: Since $\|\bar{\bar{K}}\| < 1$ $\lim_{n \to \infty} \bar{\bar{K}}^n f_0 = 0$

Or let $f_0 = g$.

$$f_n = \left(\sum_{i=0}^{n} \bar{\bar{K}}^i \right) g \approx \left(\bar{\bar{I}} - \bar{\bar{K}} \right)^{-1} g$$

$$f \left(\bar{\bar{I}} - \bar{\bar{K}} \right) = f - \bar{\bar{K}} f = g$$

$$\Rightarrow f = g + \bar{\bar{K}} f$$

Depends on convergence of a Neumann series.

How do we apply $\bar{\bar{K}}$?

1) Deterministically
2) Via Monte Carlo

Monte Carlo: Define a function $p(\bar{x}, \bar{y}) \geq 0$

↑ transition probability from $\bar{x}$ to $\bar{y}$

1) $\int p(\bar{x}, \bar{y})\, d\bar{y} \leq 1$

2) $p(\bar{x}) = 1 - \int p(\bar{x}, \bar{y})\, dy$

3) $v(\bar{x}, \bar{y}) = K(\bar{x}, \bar{y}) / p(\bar{x}, \bar{y})$

4) $\gamma = \{\bar{x}_0, \bar{x}_1, \bar{x}_2, \cdots, \bar{x}_m\}$ random walk with $p(\bar{x}, \bar{y})$ transition probabilities

5) $V_m(\gamma) = \prod_{i=1}^{m} v(\bar{x}_{i-1}, \bar{x}_i)$

5) $X(\gamma) = V_m(\gamma)\, g(\bar{x}_m) / p(\bar{x}_m)$

$$E[X(\gamma) \mid \bar{x}_0 = x] = f(\bar{x})$$

all provided the Neumann series converges, which occurs if

$$\|\bar{\bar{K}}\| = \sup_{\bar{x}} \int |K(\bar{x}, \bar{y})|\, dy < 1$$

The Dirichlet Problem

$$\Delta u = 0 \quad \text{on} \quad \Omega \subset \mathbb{R}^2$$

$$u = f \quad \text{on} \quad \partial\Omega$$

Monte Carlo method first proposed in CFL paper !!!

Cover Ω with a mesh with spacing h

$$\Delta_h u(x,y) \approx \left\{ \begin{matrix} u(x+h,y) + u(x,y+h) \\ + u(x-h,y) + u(x,y-h) - 4u(x,y) \end{matrix} \right\} h^{-2}$$

$$+ O(h^2)$$

rewrite this as

$$u(x,y) = \frac{1}{4}\left[\begin{matrix} u(x+h,y) + u(x,y+h) + \\ u(x-h,y) + u(x,y-h) \end{matrix} \right]$$

Now start walk at $(x,y) = P$ in the interior Walk via $\frac{1}{4}$ probabilities to neighbors until you hit the boundary, at point Q

$$u(x,y) = u(P) = E[f(Q)]$$

first hitting location of walks starting at P and first-passage to boundary at Q.

Defer Proof

Adjoint Method:

1. Choose Q on $\partial\Omega$ with $p(Q)$ probability
2. Walk via $\frac{1}{4}$ walk

$\forall\ P \in \Omega / \partial\Omega$: $n(P)$ number of visits to P

3. Exit region at again.

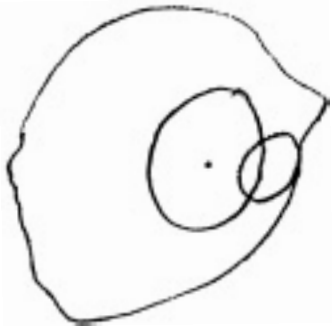
$$E\left[\frac{1}{q}\, n(P) f(Q)/p(Q)\right] = u(P)$$

Note: Via the Curtiss analysis this is very bad as $\| \bar{\bar{H}} \| = 1$.

↑ walk matrix

Variant #1.

Walk on Spheres



P is an interior point of Ω

Walk: $P' = P$

1. Draw largest sphere centered at P'
2. Pick P'' uniformly on the sphere
3. If $|\partial\Omega - P''| < \varepsilon$ let Q = closest point on $\partial\Omega$ to P''
4. Set $P' = P''$ and go to 1

$u(P) = E[f(Q)]$ with number of walks of $O(|\ln \varepsilon|)$.

Why do these methods really work?

Solutions, u, to

$$\Delta u = 0$$

are called Harmonic Functions.

Facts about harmonic functions:

1. They have the maximum property:

they cannot achieve a maximum/minimum on the interior of Ω. Why?

If so at $\bar{x}$: $\nabla \cdot u(\bar{x}) = 0$ (critical point)

but $\Delta u(\bar{x}) = 0 \Rightarrow$ neither a minimum nor maximum but instead an inflection point or a hyperbolic point!!

2 They have the mean value property:

If Ω_0 = sphere with center at $\bar{x}_0$, then

$$u(\bar{x}_0) = \frac{1}{|\partial\Omega_0|} \int_{\partial\Omega_0} u(\bar{x})\, d\bar{x}$$

$|\partial\Omega_0|$ ↑ area of the sphere

Therefore the "walk on spheres" method of generating random walks obeys the mean value property, recursively.

Note: On the grid, a solution to the discrete Laplacian obeys the discrete version of the maximum principal and mean value property:

$$u(\bar{x}) = \frac{1}{2^s} \sum_{i=1}^{2^s} u(\bar{x} + h\bar{e}_i) .$$

Thus this proves that the discrete Laplace equation can be solved with random walks on grids!!

$$\Delta u = 0 \quad \text{in } \Omega$$
$$u = g \quad \text{in } \partial\Omega$$

Green's Function: $u(\bar{x}) = \int_{\partial\Omega} G(\bar{x}, \bar{y})\, g(\bar{y})\, d\bar{y}$

$$\Delta_{\bar{x}} G(\bar{x}, \bar{y}) = 0 \;, \quad \bar{x} \in \Omega / \partial\Omega \;, \; y \in \partial\Omega$$
$$G(\bar{x}, \bar{y}) = \delta(\bar{x} - \bar{y}) \;, \quad \bar{x}, \bar{y} \in \partial\Omega$$

Can prove that

$p(\bar{x}, \bar{y})$ = "first passage probability from $\bar{x} \in \Omega / \partial\Omega$ to $\bar{y} \in \partial\Omega$"

$= G(\bar{x}, \bar{y})$

Boundary mass

Solving Eigenvalue Problems:

Solve a) $\bar{\bar{H}}\bar{x} = \lambda \bar{x}$

b) $\int K(\bar{x},\bar{y}) f(\bar{y}) d\bar{y} = \lambda f(\bar{x})$

c) $-\Delta u(\bar{x}) + V(\bar{x}) u(\bar{x}) = \lambda u(\bar{x})$

For a) consider the recursion

$$\bar{x}_n = \bar{\bar{H}}\,\bar{x}_{n-1} / \lambda_n \quad \text{where}$$

$$\lambda_n = \| \bar{\bar{H}}\,\bar{x}_{n-1} \| \Rightarrow \| \bar{x}_n \| = 1.$$

This is the power method and from that we know that $\lambda_n \to \lambda$ the largest eigenvalue of $\bar{\bar{H}}$ and so $\bar{x}_n \to \bar{x}$ the corresponding eigenvector.

For b) $f_n(\bar{x}) = \int K(\bar{x},\bar{y}) f_{n-1}(\bar{y})\, d\bar{y}\ \lambda_n^{-1}$

where $\lambda_n = \| \int K(\bar{x},\bar{y}) f_{n-1}(\bar{y})\, d\bar{y} \|$.

as above this is the power method.

How do we do this via Monte Carlo?

a) Use a random walk to apply $\bar{\bar{H}}$ to $\bar{x}$, recursively.

b) Same here. Use random walk to apply $\bar{\bar{K}}$ to $f(\bar{y})$, recursively.

Example (Matrix case): Do yourself.

Aside: Can look at Kac / Donsker method but first we need more probability.