Control Variates

④ β^* in $V(Z)$

$$V(\phi)\left(1+\left(-\frac{C(\phi(x),Y)}{V(Y)}\right)^2\frac{1}{V(\phi(x))}V(Y)+2\left(-\frac{C(\phi(x),Y)^2}{V(Y)V(\phi)}\right)\right)$$

$$= V(\phi)\left(1+\frac{C(\phi,Y)^2}{V(Y)V(\phi)}-2\frac{C(\phi,Y)^2}{V(Y)V(\phi)}\right)$$

$$= V(\phi)\left(1-\frac{C(\phi,Y)^2}{V(Y)V(\phi)}\right) = V(\phi)(1-\rho_{\phi,Y}^2)$$

⑤ If $\rho^2>0$, then $V(\phi)$ is reduced.

- $V(\phi)$ reduced

- Known $E(Y)$ req.
- Finding Y such that $\rho_{\phi,Y}>0$
- Time Complexity to generate Z must be worth it.

Var. Reduction for MC methods

Seek $\tau = E(\phi(x))$

• **Antithetic Sampling**: Variance reduction by sampling antithetic variables $\tilde{V}$ in pairs with standard sampling V creating

1

$W = \frac{V+\tilde{V}}{2}$, unbiased $E(W) = \tau$

Variance $V(W) = V\left(\frac{V+\tilde{V}}{2}\right) =$

$$= \frac{1}{4} V(V) + V(\tilde{V}) + 2C(V,\tilde{V})$$

$$= \frac{1}{2}\left(V(V) + \underbrace{C(V,\tilde{V})}\right)$$

If $C(V,\tilde{V}) < 0$ then comp. work gained

2

$$\varepsilon\sqrt{N} > \lambda_{\alpha/2} D(V)$$

$$\Leftrightarrow \varepsilon^2 N > \lambda_{\alpha/2}^2 V(V)$$

$$\Leftrightarrow N_V > \lambda_{\alpha/2}^2 \frac{V(V)}{\varepsilon^2}$$

Find

$$2\lambda_{\alpha/2}^2 \frac{V(W)}{\varepsilon^2} < \lambda_{\alpha/2}^2 \frac{V(V)}{\varepsilon^2}$$

$$\Leftrightarrow \cancel{V(V)} + C(V,\tilde{V}) < \cancel{V(V)}$$

$$\underline{C(V,\tilde{V}) < 0}$$, then win!

Var. Reduction for MC methods

Antithetic

③ Apply theorem:

Theorem

Let $V = \varphi(U)$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is a monotone function. Moreover, assume that there exists a non-increasing transform $T : \mathbb{R} \to \mathbb{R}$ such that $U \stackrel{d.}{=} T(U)$. Then $V = \varphi(U)$ and $\tilde{V} = \varphi(T(U))$ are identically distributed and

$$\mathbb{C}(V, \tilde{V}) = \mathbb{C}(\varphi(U), \varphi(T(U))) \leq 0.$$

An important application of this theorem is the following: Let F be a distribution function and ϕ a monotone function. Then, letting $U \sim \mathcal{U}(0,1)$, $T(u) = 1 - u$, and $\varphi(u) = \phi(F^{-1}(u))$ yields, for $V = \phi(F^{-1}(U))$ and $\tilde{V} = \phi(F^{-1}(1-U))$,

Same random varible

$$V \stackrel{d.}{=} \tilde{V} \quad \text{and} \quad \mathbb{C}(V, \tilde{V}) \leq 0.$$

$U \sim \mathcal{U}(0,1)$,

non-incr. Transform

$T(u) = 1 - u$,

Monotone func

$\varphi(u) = \phi(F^{-1}(u))$

$V = \phi(F^{-1}(U))$, $\tilde{V} = \phi(F^{-1}(1-U))$

$V = \tilde{V}$ and $C(V, \tilde{V}) \leq 0$

Var. Reduction for MC methods

④ Ex application

$$\tau = \int_{-\pi/2}^{\pi/2} \exp(\cos^2(x))\,dx =$$

$$= \int_0^{\frac{\pi}{2}} \underbrace{2\frac{\pi}{2}\exp(\cos^2(x))}_{\phi(x)} \underbrace{\frac{2}{\pi}}_{f(x)}\,dx$$

$$= E_f(\phi(X))$$

$$\begin{cases} U = \cos^2(X), & X \sim \mathcal{U}(0, \pi/2) \\ T(u) = 1 - u & \text{Non-increasing} \\ \varphi(u) = 2\frac{\pi}{2}\exp(u) & \text{Monotone func} \end{cases}$$

$$T(U) = 1 - \cos^2(x) = \sin^2(x) = \cos^2\left(\tfrac{\pi}{2} - X\right) \overset{d}{=} \cos^2(X) = U$$

Thus,

$$C\left(\pi\exp(\cos^2(X)), \pi\exp(1 - \cos^2(X))\right) \leq 0$$

Seq. MC (particle filter) methods

τ_n sequences of statistics!

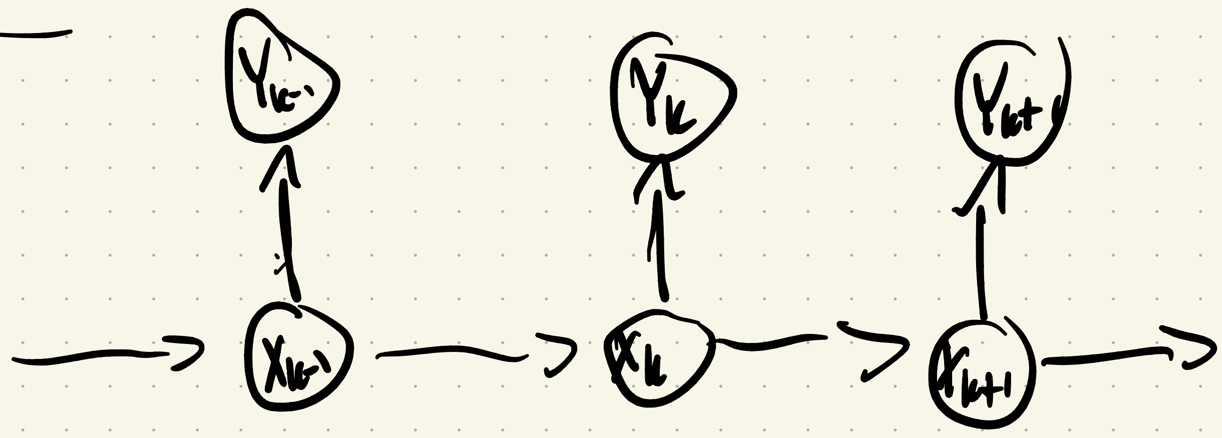
SMC

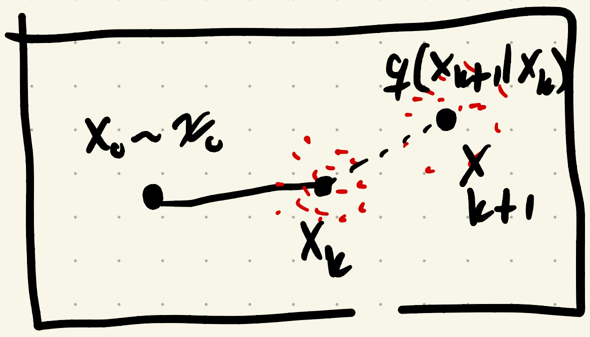
$$E_{f_n}(\phi(x_{0:n})) \approx \sum_i^N \frac{w_n^i}{\sum_j^N w_n^j} \phi(x_i^{0:n}) \quad \text{SIS}$$

$$\tau_n = E_f(\phi(\underline{X}_{0:n})) = \int_X \phi(x_{0:n}) f_n(x_{0:n})\, dx_{0:n}$$

Ex application:

HHM





HHM

$$Y_k | X_k = x_k \sim \overbrace{p(y_k | x_k)}^{\text{observation dens}}$$

$$X_{k+1} | X_k = x_{k+1} \sim \overbrace{q(x_{k+1} | x_k)}^{\text{transition density}}$$

$$X_0 \sim \mathcal{X}(x_0) \quad \text{Initial Distr}$$

Gives density fn

$\longrightarrow$ Smoothing

Seq

$$f_n(x_{0:n} | y_{0:n}) = \frac{\mathcal{X}(x_0)\, p(y_0|x_0) \prod_{k=1}^n p(y_k|x_k)\, q(x_k|x_{k-1})}{L_n(y_{0:n})} \quad \text{ML}$$

ML density observation

Seq. MC (particle filter) methods

SMC again

MC

$$\tau_n = E(\phi(X_{0:n}) \mid Y_{0:n} = y_{0:n}) =$$

$$= \int \phi(x_{0:n}) f_n(x_{0:n} \mid y_{0:n})\, dx_0 \cdots dx_n$$

Put in $f_n(x_{0:n} \mid y_{0:n})$

$$= \int \phi(x_{0:n}) \frac{z(x_{0:n})}{c_n}\, dx_0 \cdots dx_n$$

SIS

Self Normalized Importance Sampling.

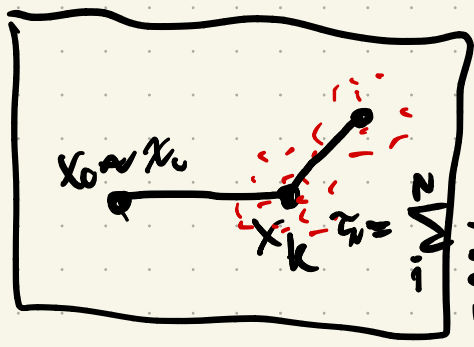
$\boxed{\text{SIS algorithm}}$

Particles w/ weights

$t=0$

Init Particles

For $i = 1 \to N$
  draw $X_i^0 \sim g_0$
  set $w_0^i = \frac{z_0(X_i^0)}{g_0(X_i^0)}$
  return (X_i^0, w_0^i)



$\tau_n = \sum_i^N \frac{w(x_i)}{\sum_j w(x_j)} \phi(x_i)$

for $k = 0, 1, 2 \ldots$

Online

For all particles

  for $i = 1 \to N$

Instrumental given now

  draw $X^{k+1} \sim g_{k+1}(x_{k+1} \mid X_i^{0:k})$

$$W_{k+1}^i = \frac{z_{n+1}(X_i^{0:k+1})}{g_{n+1}(X_i^{0:k+1})}$$

Append states

  set $X_i^{0:k+1} \leftarrow (X_i^{0:k}, X_i^{k+1})$

Update Weights

  set $W_{k+1}^i \leftarrow \frac{z_{k+1}(X_i^{0:k+1})}{z_k(X_i^{0:k}) g_{k+1}(X_i^{k+1} \mid X_i^{0:k})} W_k^i$

  Return $(X_i^{0:k+1}, W_{k+1}^i)$

State predictions with weights

Seq. MC (particle filter) methods

SIS

Pros

- Online updating of γ_n
- Flexible acc through time
- Works w. non-linear
- Easy implementation!
- Low variance

Cons

- Weight degeneration as $n \to$
- Dimensionality harder

- Weight degeneration causes SIS to be uneffective. SISR solves this problem.

Improve with selection: SISR SIS-Resampling

- Delete samples with small weights.

- Draw with replacement among $X_i^{0:k}$ with probabilities $\tilde{X}_i^{0:n} \sim \frac{W_n^i}{\sum_j^N W_n^j}$, total N particles

Resampled $\tilde{X}_i^{0:k}$ are set $\tilde{w}_n^i = 1$

No bias

$$E\left(\frac{1}{N}\sum_{i=1}^{N}\phi(\tilde{X}_i^{0:n})\right) = E\left(\sum_{i=1}^{N}\frac{W_n^i}{\sum_{j=1}^{N}W_n^j}\phi(X_i^{0:n})\right)$$

Seq. MC (particle filter) methods

No bias

$$E\Big(\underbrace{\frac{1}{N}\sum_{j=1}^{N}\phi(\tilde{X}_j^{0:n})}_{\tilde{\tau}}\Big) = E\Big(\underbrace{\sum_{j=1}^{N}\frac{w_n^i}{\sum_{j=1}^{N} w_n^j}\phi(X_i^{0:n})}_{\tau}\Big)$$

Proof

Rewrite

$$\tilde{\tau}_n = \sum_{k=1}^{N}\frac{N_n^k}{N}\phi(X_k^{0:n}) =$$

$$E\Big(E\Big(\tilde{\tau}_n \,\Big|\, \{X_k^{0:n}\}_{k=1}^{N}\Big)\Big) =$$

number of resamples of sample k.

$$= \frac{1}{N}\sum_{k=1}^{N} E\Big(E\Big(N_n^k \phi(X_k^{0:n}) \,\Big|\, \{X_k^{0:n}\}_{k=1}^{n}\Big)\Big)$$

IDK why

$$= \frac{1}{N}\sum_{k=1}^{N} E\Big(\phi_k^{0:N} E\Big(N_n^k \,\Big|\, \{X_k^{0:n}\}_{k=1}^{n}\Big)\Big)$$

Bernoulli

$$= \frac{1}{N}\sum_{k=1}^{N} E\Big(\phi_k^{0:N} E\Big(N_n^k \,\Big|\, \{\}\Big) \in \mathrm{Bin}\Big(N, \frac{w_i^{0:n}}{\sum_j w_j^{0:n}}\Big)\Big)$$

$$= \cancel{\frac{1}{N}}\sum_{k=1}^{N} E\Big(\phi_k^{0:n} \cancel{N} \frac{w_i^{0:n}}{\sum_j^{N} w_i^{0:n}}\Big) = E(\tau)$$

MCMC methods $\mathcal{X}$ - state,

Monte Carlo Markov Chain

MCMC has unique properties enabling unique modeling, such as Bayesian statistics w. posteriors.

Markov Chain

- Markov chains are a stochastic process of transitions between states where future states are uncorr. with past states or seq of them $\mathcal{X} = \pi \Rightarrow f(X_n) = \pi(X_n) \; \forall n$
- <u>Stationary Distribution</u>. The density function that Markov Chain forms in mean after $t \to \infty$ (expected distr). Chain X_n

$$\int q(x|z)\pi(z)\,dz = \pi(x)$$

- <u>Ergodicity</u>. Property of a Markov Chain that when $t < \infty$ all states will be visited. Guarantee of convergence to stationary distribution.

$$\sup_{A \subseteq X} |P(X_n \in A) - \pi(A)| \to 0 \quad \text{as } n \to \infty$$

- <u>Irreducibility</u>. Given steps $< \infty$ all states can be reached from any start point
- <u>Aperiodicity</u>. No cycle loops ensuring that the walk cannot get stuck.

MCMC methods

- **Geometric Ergodicity**

$$q(z \mid x) \geq \varepsilon \mu(z) \quad \forall x, z \in X$$

$$\sup_{A \subseteq X} \left| P(X_n \in A) - \pi(A) \right| \leq \rho \text{ , there exist } \rho < 1$$

$$n \to \infty$$

Forgets init distr
geometrically fast

Stronger conditions and subset of Ergodicity.
MC is geometrically ergodic if it converges to stationary distr. at exponential (geometric) rate.

$$\left| C(\phi(X_m), \phi(X_n)) \right| \leq C \tilde{\rho}^{|n-m|}$$

$$\tilde{\rho} < 1 , \quad C(\phi) > 0$$

$$\tau_n = \frac{1}{n} \sum_{k=1}^{n} \phi(X_n)$$

MCMC methods

- Proof of large numbers for geometrically ergodic Markov Chains and state CLT.

Proof of LLN:

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\phi(X_k) - \underbrace{\int_X \phi(x)\pi(x)\,dx}_{\tau}\right| > \varepsilon\right) \to 0 \quad \text{as } n\to\infty$$

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\phi(X_k) - \tau\right| > \varepsilon\right) \underset{\substack{\text{Chebyshev}\\\text{inequality}}}{\leq} \frac{\mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^{n}(\phi(X_k)-\tau)\right|^2\right]}{\varepsilon^2}$$

$$= \frac{1}{n^2\varepsilon^2}\mathbb{E}\left[\left(\sum_{k=1}^{n}(\phi(X_k)-\tau)\right)^2\right]$$

$$= \frac{1}{n^2\varepsilon^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{E}\left[(\phi(X_k)-\tau)(\phi(X_l)-\tau)\right]$$

$$= \frac{1}{n^2\varepsilon^2}\sum_{k=1}^{n}\sum_{l=1}^{n} C(\phi(X_k),\phi(X_l))$$

$$\overset{\text{sym}}{=} \frac{1}{n^2\varepsilon^2}\left(\sum_{k=1}^{n} V(\phi(X_k)) + 2\sum_{k=2}^{n}\sum_{l=1}^{k-1} C(\phi(X_k),\phi(X_l))\right) =$$

$$\left\{\text{due stationarity} \quad C(\phi(X_l),\phi(X_{l+\Delta})) = C(\phi(X_1),\phi(X_{1+\Delta}))\right\}$$

$$\overset{\text{stat}}{=} \frac{1}{\varepsilon^2 n^2}\left(nV(\phi(X_1)) + 2\sum_{k=1}^{n-1}(n-k)\,C(\phi(X_1),\phi(X_{1+k}))\right)$$

$$\leq \frac{1}{\varepsilon^2 n^2}\left(nV(\phi(X_1)) + 2\sum_{k=1}^{n-1}(n-k)\left|C(\phi(X_1),\phi(X_{1+k}))\right|\right)$$

$$\leq \frac{2}{\varepsilon^2 n}\left(V(\phi(X_1)) + \sum_{k=1}^{n-1}\left|C(\phi(X_1),\phi(X_{1+k}))\right|\right)$$

$$= \frac{2}{\varepsilon^2 n}\sum_{k=0}^{n-1}\left|C(\phi(X_1),\phi(X_{1+k}))\right| \leq \frac{2}{\varepsilon^2 n}\underbrace{\sum_{k=0}^{\infty}\left|C(\phi(X_1),\phi(X_{1+k}))\right|}_{<\infty \text{ due to geom. erg.}}$$

$\to 0$ as $n\to\infty$ $\square$

MCMC methods

$X_{k+1}^{0:n}$

- Proof of large numbers for geometrically ergodic Markov Chains and state CLT.

Geometrically ergodic with LLN gives asymptotically stationary distribution.

$X_n, \pi, \forall \ \varepsilon > 0$ $\quad P(|\tau_N - \tau| \geq \varepsilon) \to 0$

$$P\Bigg(\Bigg|\underbrace{\frac{1}{n}\sum_{k=1}^{n} \phi(X_n)}_{\tau_N} - \underbrace{\int \phi(x)\pi(x)\,dx}_{\tau}\Bigg| \geq \varepsilon\Bigg) \longrightarrow 0$$

as $n \to \infty$

$\longrightarrow \frac{1}{n}\sum_{k=1}^{n} \phi(X_k)$ converges to mean $\phi(X_k)$ under π

Proof:

Chebyshev's inequality

$$P(|\tau_N - \tau| \geq \varepsilon) \leq E\Bigg(\underbrace{\Bigg|\frac{1}{n}\sum_{k=1}^{n}\big(\phi(X_n) - \tau\big)\Bigg|^2}_{\varepsilon^2}\Bigg)$$

$= \ldots$

$=$ Stationarity $C(\phi(X_1), \phi(X_{1+\Delta}))$

Show that $\varepsilon^2 \to 0$ as $n \to \infty$

For X_n where $E_\pi(\phi(x)) = \tau_N$ and $V(\phi(x)) = \sigma^2(\phi)$

$$\Longrightarrow \sqrt{N}(\tau_N - \tau) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\phi))$$