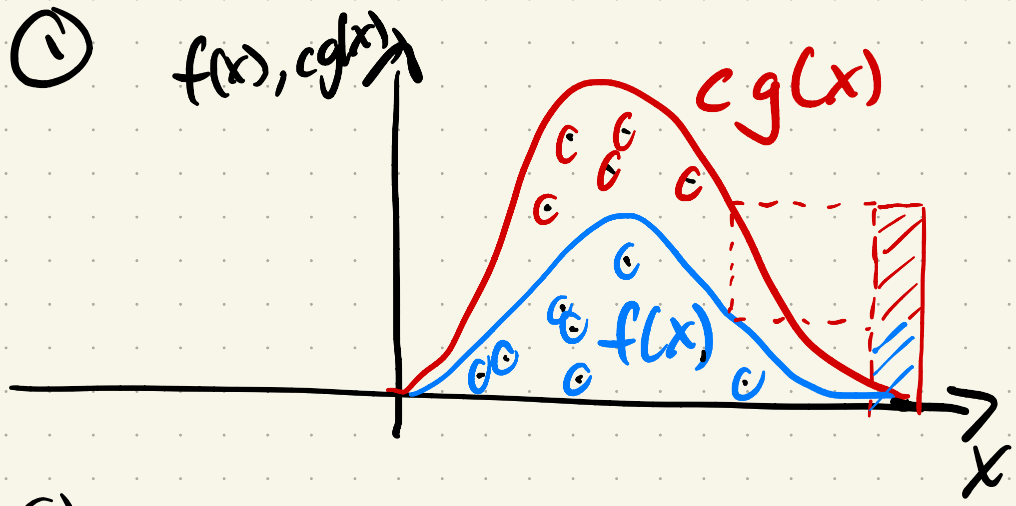
Oral Exam Prep

Random Number Generation

Rejection Sampling

① 

Proposal Distr. $g(x)$

Target Distr. $f(x)$

Form

$$cg(x) \geq f(x)$$

$\boxed{c \geq 1}$ ex: $g \in N(0,1)$

② Draw $Y \sim g$ (interpret prob) c scale param

③ Draw $U \sim \mathbb{1}_{(0,1)}$

④ Reject Y if $Ucg(Y) > f(Y)$ and return ②

⑤ Else, set $X = Y$ and save it as $X \sim f$ and return ①

No bias

- Obtains random draws X_i from target distr $f(x)$
- Sample candidates from an easier distr and correct the sample probability with the help of rejecting candidates.

Pros

- Unbiased i.i.d $X_i \sim f$
- Enables sampling from unknown CDF's
- Flexible, can handle any distributions really

Cons

- Risks being inefficient if diff $|cg(x) - f(x)|$ is large and many rejects, large K.
- Modeling prop. distr. $\underbrace{cg(x)}_{\text{can be hard}}$

Rejection sampling

$P(X \leq Y) = P(Y \leq y \mid \text{accept}$

$$P(X \leq y) \overset{\text{assert}}{=} P(Y \leq y \mid \text{accepted sample})$$

$$= P\left(Y \leq y \mid U \leq \frac{f(Y)}{cg(Y)}\right)$$

$= P(Y \leq y \cap U \leq \frac{f(u)}{cg(u)}$

$= \int_{-\infty}^{y} \int_{0}^{f(z)/cg(z)} du\, dz$

$$\overset{\text{Cond. Prob. Def.}}{=} \frac{P\left(Y \leq y \cap U \leq \frac{f(Y)}{cg(Y)}\right)}{P\left(U \leq \frac{f(Y)}{cg(Y)}\right)}$$

$\left(P(Y < y) = \int_{-\infty}^{y} f_Y(z)\, dz\right)$

$$\overset{\text{CDF def}}{=} \frac{\int_{-\infty}^{y} \int_{0}^{f(z)/cg(z)} du\, g(z)\, dz}{\int_{-\infty}^{\infty} \int_{0}^{f(z)/cg(z)} du\, g(z)\, dz}$$

the same but $\int_{-\infty}^{y}$ on denominator

density def $(= 1)$

$$= \int_{-\infty}^{y} f(z)\, dz \qquad \text{Q.E.D}$$

unbiased i.i.d f

Monte Carlo integration

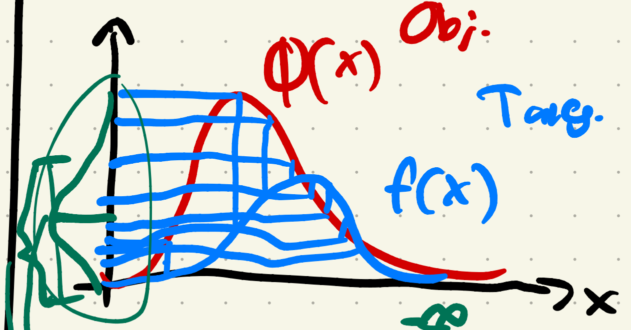
Explain

(A) Basic Monte Carlo Sampler

(B) Importance Sampling

(C) Self-normalized Importance Sampling

Visual!



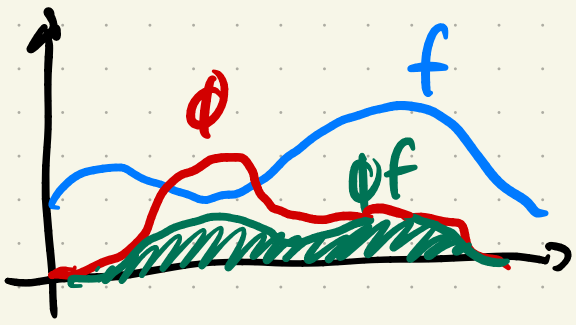
$E_f(\phi(x)) = \int_{-\infty}^{\infty} \phi(x) f(x) dx$

$\sigma^2(\phi)$ inference

Intro: Random sampling to estimate integrand $\phi(x)$ useful for problems that have difficult analytical solutions (or has no analytical solution).

Curse of dimensionality makes integrals vry diff as $d \to$.

We seek



(A) The foundation of basic MC sampling is the law of large numbers with $\phi(x_i)$ over n i.i.d $X_i \sim f(x)$ target dens converges to $\longrightarrow E_f(\phi(x))$ as $n \to \infty$

Analytical Problem

$$\tau \equiv E_f(\phi(X)) = \int_A \phi(x) f(x) dx \ , \ A \subseteq \mathbb{R}^d$$

Numerical Solution MC sampler

$$\tau_N \equiv \frac{1}{N} \sum_{i=1}^{N} \phi(X_i) \approx \tau \ , \ \text{as } N \text{ is large}$$

Asserting $N \to \infty$, then $\tau_N = \tau$

Monte Carlo integration

Problem

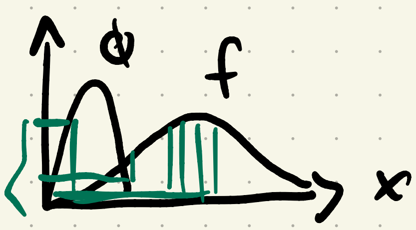
$$\tau \equiv E(\phi(X)) = \int_A \phi(x) f(x)\,dx$$

X: rand var within $A \subseteq \mathbb{R}^d$ (where $d \in \mathbb{N}$ may be large)

$f: A \to \mathbb{R}_+$ is density of X (target)

$\phi: A \to \mathbb{R}$ objective function

(B)



MC effectiveness (i.e $\sigma(\phi)$) diminishes when ϕ obj and f target are dissimilar and many draws at small $\phi(x)$ not contr. to estimate.

Importance sampling is introducing instrumental density

g - instrumental density, $g(x) = 0 \Rightarrow f(x) = 0$

To make them more similar.

$$\Rightarrow \quad \tau = E(\phi(x)) = \int_{f(x)>0} \phi(x) f(x)\,dx = \int_{g(x)>0} \phi(x) \frac{f(x)}{g(x)} g(x)\,dx =$$

$$= \int_{g(x)>0} \phi(x) w(x) g(x)\,dx = E_g(\phi(x) w(x))$$

w - Imp weight function

Monte Carlo integration

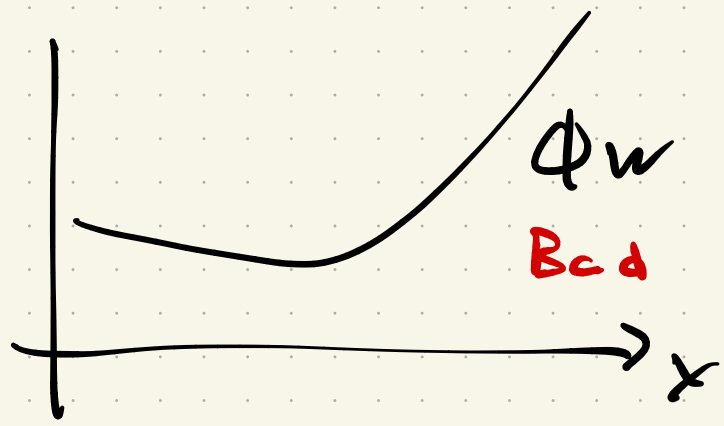
(B)

Algorithm

① $X_i \sim g$ $\quad \left(\frac{f(x)}{g(x)} \right)$

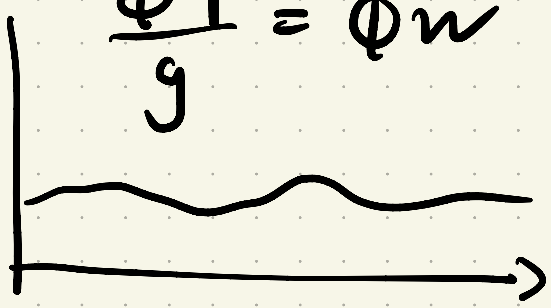
② $\tau_N = \frac{1}{N} \sum_{i}^{N} \phi(x_i) w(x_i)$

$$V(\tau_N) = \frac{1}{N} V_g\big(\phi(x) w(x)\big)$$

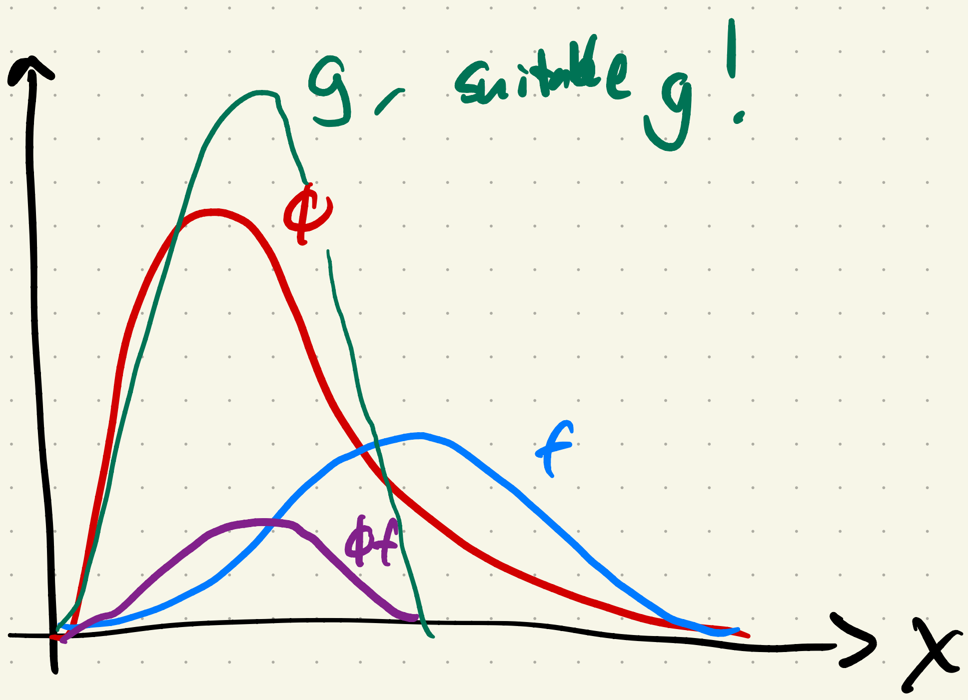


Good

$\frac{\phi f}{g} = \phi w$



- $\phi(x) w$ should be close to constant in support of g.
- g should be chosen by scale-location-transformation.



Monte Carlo integration

ⓒ Self Normalized IS

$\underbrace{z(x)}_{\text{we can sample}} = c f(x)$, we do not know normalizing const c!

$$\tau = E_f(\phi(x)) \overset{\text{def}}{=} \int_X \phi(x) f(x)\,dx = \frac{c\int_{f>0} \phi(x) f(x)\,dx}{c\int_{f>0} f(x)\,dx} \quad (= 1)$$

$$= \frac{\int_{g>0} \phi(x)\, c\frac{f(x)}{g(x)}\, g(x)\,dx}{\int_{g>0} c\frac{f(x)}{g(x)}\, g(x)\,dx}$$

$$= \frac{\int_{g>0} \phi(x) \frac{z(x)}{g(x)} g(x)\,dx}{\int_{g>0} \frac{z(x)}{g(x)} g(x)\,dx}$$

$w(x) = \frac{z(x)}{g(x)}$

$$= \frac{\int_{g>0} \phi(x) w(x) g(x)}{\int_{g>0} w(x) g(x)\,dx} = \frac{E_g(\phi(X) w(X))}{E_g(w(X))}$$

Monte Carlo integration

(c) with

$X_i \sim g \in \mathcal{N}(\mu, \sigma)$

$E_g(\phi(X) w(X))$, $E_g(w(X))$ may be estimated

$$\tau = \frac{E_g(\phi(X) w(X))}{E_g(w(X))} \approx \frac{\frac{1}{N}\sum_i^N \phi(X_i) w(X_i)}{\frac{1}{N}\sum_j^N w(X_j)} =$$

$$= \sum_i^N \underbrace{\frac{w(X_i)}{\underbrace{\sum_j^N w(X_j)}_{c = E_g(w(X))}}}_{\tilde{w}(X_i)} \phi(X_i) \quad \text{SIS}$$

Normalized constant

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

Monte Carlo integration

(A)
- Prove the MC sampler satisfies CLT and provide asymptotic variance expression. (B) Relation to Imp. Sampl.? (C) How is this useful for CI? (D) How can this be used for designing the instrumental distribution for a Imp. Sampl. problem?

(A) As N increases, the distr of the scaled deviation of the MC estimate τ_N from the true τ, $\tau_N - \tau \xrightarrow{d} \mathcal{N}(0, \sigma^2(\phi))$

CLT def gives

$$\sqrt{N}(\tau_N - \tau) \xrightarrow[]{\text{convergence}\ d} \mathcal{N}(0, V(\phi(x)))$$

$\sqrt{N}(\tau_N - \tau)$ is asymptotically normal.

CLT requires:
- rand var "τ_i" are i.i.d. OK
- $V(\phi(x)) < \infty$ OK

Monte Carlo integration

• Asymptotic variance of MC estimator τ_N

Rate of convergence gives for large N

asympt normal

$$V(\sqrt{N}(\tau_N - \tau)) = N\,V(\tau_N - \tau) = V(\phi(x))$$

Implying

$$D(\tau_N - \tau) = \sqrt{V(\tau_N - \tau)} = \sqrt{\frac{V(\phi(x))}{N}}$$

$$= \frac{D(\phi(x))}{\sqrt{N}} = \frac{\sigma(\phi)}{\sqrt{N}}$$

Convergence Rate

$O\left(\frac{1}{\sqrt{N}}\right)$, invariant of d!

$\varphi(\tau)$ has CLT if $\varphi \in C_1$

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

$\varphi(\tau_N)$ is O.K but not unbiased estimator

(B) This relates to IS by

$$\tau_N = E_g(\phi(x) w(x))$$

$$V(\sqrt{N}(\tau_N - \tau)) = N\,V(\tau_N - \tau) = V(\phi(x) w(x))$$

$$\iff D(\tau_N - \tau) = \frac{D(\phi(x) w(x))}{\sqrt{N}}$$

$$c = E(c) = \int \overbrace{c f(x)}^{z(x)} dx = E_g(w(x)) = \int \frac{z(x)}{g(x)} g(x)\,dx = \int w(x) g(x)\,dx$$

$$\approx \frac{1}{N} \sum_{i}^{N} w(x_i)$$

Monte Carlo integration

Ⓒ With inference of $(\tau_N - \tau) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\phi))$, you can construct CI of $\phi(x)$ by

$$I_\alpha = \left(\tau_N - \lambda_{\alpha/2} \frac{\sigma(\phi)}{\sqrt{N}}, \tau_N + \lambda_{\alpha/2} \frac{\sigma(\phi)}{\sqrt{N}} \right)$$

- Where $\sigma^2(\phi)$ is estimated

$$\sigma^2(\phi) \overset{\text{Var. def.}}{=} E(\phi^2(x)) - E(\phi(x))^2$$

$$\approx E(\phi^2(x)) - \tau_N^2$$

$$\overset{\text{Var. Def.}}{=} \frac{1}{N} \sum_i^N \left(\phi(x_i) - \frac{1}{N} \sum_{j=1}^N \phi(x_j) \right)^2$$

Bias corr. estimator

$$\Rightarrow \frac{1}{N-1} \sum_i^N \left(\phi(x_i) - \frac{1}{N} \sum_{j=1}^N \phi(x_j) \right)^2$$

Ⓓ For IS, this can be used for designing the instrumental distribution.

$$I_{\alpha/2} = \left(\tau_N - \lambda_{\alpha/2} \frac{\sigma(\phi w)}{\sqrt{N}}, \tau_N + \lambda_{\alpha/2} \frac{\sigma(\phi w)}{\sqrt{N}} \right)$$

Choose w so that CI is minimized.

Var. Reduction for MC methods

$$E_f(\phi(x)) = E_{\tilde{f}}(\tilde{\phi}(x)), \quad \sigma^2(\tilde{\phi}) < \sigma(\phi)$$

- Explain how $\sigma^2(\phi(x))$ can be reduced using control variates.

E(Y) is known

① Construct $Z = \phi(X) + \beta[Y - E(Y)]$

$= \phi(X) + \beta(Y - m)$

② Verify $E(Z) = E[\phi(X)] + \cancel{E(Y) - E(Y)}$

$= E(\phi(X)) = \tau$

$V(\phi(X)) = \sigma^2(\phi)$

TNV

$V(Z) = V(\phi(X) + \beta Y)$

$= V(\phi(X)) + \beta^2 V(Y)$

$+ 2\beta C(\phi(X), Y)$

③ Find β^*

$$\frac{\partial V(Z)}{\partial \beta} = 2\beta V(Y) + 2C(\phi(X), Y) = 0$$

$$\iff \beta^* = -\frac{C(\phi(X), Y)}{V(Y)}$$