TOPIC: Importance Sampling / reweighting

DATE:

FILE UNDER:

PAGE:

$Var(Y_i) =$

$Y_i = \frac{1}{3}$

$Y_i = \begin{cases} 1/3 & 1/2 \\ 0 & 1/2 \end{cases}$

$Var(Y_i) = \mathbb{E}[Y^2] - \left(\frac{1}{6}\right)^2$

$= \frac{1}{9}\frac{1}{2}$

$\frac{1}{18} - \frac{1}{36}$

Roll die n times, get n/6 1's

Estimate $\cancel{\mathbb{P}(\theta)}\ \mathbb{P}(X = 1) \underset{\hat{\theta}}{=} \sum \mathbb{1}\{X_i = 1\}/n$

$Var(\hat{\theta}) = \frac{5}{36n}$

$1 \quad \cancel{2} \quad \cancel{3} \quad 4 \quad 5 \quad 6$

$\underbrace{1 \quad 1 \quad 1}$

Let $Y_i = \begin{cases} 1/3 & \text{if } X_i = 1 \\ 0 & \text{if } X_i = 4, 5, 6 \end{cases}$

$\mathbb{E}\, Y_i = 1/6, \quad Var(Y_i) = \frac{1}{36}$

$\hat{\theta} = \frac{1}{n} \sum Y_i$

$Var(\hat{\theta}) = \frac{1}{36n}$

TOPIC: DATE:

FILE UNDER: PAGE:

$$\cancel{\hat{\mu}_n = \frac{\sum \frac{f(x_i)}{g(x_i)} h(x_i)}{\sum \frac{f(x_i)}{g(x_i)}} = \frac{\sum w(x_i)\, h(x_i)}{\sum w(x_i)}}$$

With rejection sampling, we can sample from f given a candidate density g.

What if we want to estimate $\mathbb{E}_f[h(X)]$ for some $h: \mathbb{R}^k \to \mathbb{R}$?

Obvious way: Sample $X_1, \ldots, X_n \sim f$ and use

$$\mathbb{E}_f[h(X)] \approx \frac{1}{n} \sum_{i=1}^{n} h(X_i) = \hat{\mu}_n$$

$$\sqrt{n}(\hat{\mu}_n - \mu_n) \longrightarrow N(0, \sigma^2)$$

But sampling from f is hard, so we use RS with cand. dens. g.

In order to obtain sample of size n, we need on avg, $c \times n$ samples from g, where

$$c = \sup f/g \quad (\text{assumed} < \infty)$$

We reject (on avg.) $(c-1) \times n$ of the samples from g!

Those samples belong in the domain of f, but they may be over/under-represented

e.g. if g has heavier tails, there will be too many extreme values — RS rejects those.

But maybe we can down/up weight values to get ~~an~~ the right answer.

TOPIC:

DATE:

FILE UNDER:

PAGE:

No rejection sampling here

Note that

$$\mathbb{E}_f[h(X)] = \mathbb{E}_g\left[\frac{f(X)}{g(X)} h(X)\right].$$

If $X_1, \dots, X_n \sim g$ then

$$\mathbb{E}_f[h(X)] \approx \frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)} h(X_i)$$

$$= \frac{1}{n}\sum \underbrace{W_i}_{\text{importance weights}} h(X_i) = \tilde{\mu}_h$$

$\Rightarrow$ Assumes we can compute f and g exactly.

Notice that if $f = g$, our estimator is just $\frac{1}{n}\sum_{i=1}^{n} h(X_i)$.

~~Ex~~ If the density of X_i w.r.t g is larger than its density w.r.t f, then $h(X_i)$ is down weighted in the sum (vice versa)

Comparison

Rejection Sampling: Sample directly from f, take averages

Importance Sampling: Sample from g, reweight by f/g

For computing expectations IS is much more efficient because there is no rejection

TOPIC:

DATE:

FILE UNDER:

PAGE:

~~What if we only know~~

RS: Let $c = \sup f/g$. Then

sample $X_1, \ldots, X_n \sim g$ and $U_1, \ldots, U_n \sim \text{Unif}(0,1)$.

$$\hat{\mu}_h = \frac{\sum \mathbb{1}\{U_i \le \frac{f(X_i)}{c g(X_i)}\} h(X_i)}{\sum \mathbb{1}\{U_i \le \frac{f(X_i)}{c g(X_i)}\}}$$

IS: Sample $X_1, \ldots, X_n \sim g$

$$\tilde{\mu}_h = \frac{1}{n} \sum \frac{f(X_i)}{g(X_i)} h(X_i)$$

1. IS estimator ($\tilde{\mu}_h$) is "smoother" than $\hat{\mu}_h$ and should have lower variance
2. RS requires $\sup f/g < \infty$. But IS does not need c.
3. IS computations cannot be done on log scale

What if we only know $f^* = c_1 f$ and $g^* = c_2 g$. Then use

$$\hat{\mu}_h^* = \frac{\frac{1}{n} \sum_{i=1}^n \frac{f^*(X_i)}{g^*(X_i)} h(X_i)}{\frac{1}{n} \sum_{i=1}^n \frac{f^*(X_i)}{g^*(X_i)}} \quad \text{(ratio estimator)}$$

Since f^* and g^* are in numerator and denominator, constants drop out.

$\hat{\mu}_h^* \longrightarrow \mu_h$ by Slutsky theorem

TOPIC: | DATE:

FILE UNDER: | PAGE:

Ex. Importance sampling for Bayesian sensitivity analysis.

We have data y with a likelihood $L(y|\theta)$ and a prior for ~~θ~~ $\pi(\theta|\psi_0)$ where ψ_0 is a known hyperparameter. The posterior for θ is

$$p(\theta|y,\psi_0) \propto L(y|\theta)\pi(\theta|\psi_0).$$

Suppose ~~we~~ expend much energy obtaining $\theta_1 \to \theta_n$, a sample from $p(\theta|y,\psi_0)$, and we can compute posterior mean $\mathbb{E}[\theta] = \frac{1}{n}\sum \theta_i$. What if we want to calculate different values of $\mathbb{E}(\theta|y,\psi)$? We do not need to resample $\theta_1, \to \theta_n$, just reweight them. ~~Let $f = p(\theta|y,\psi$~~ Let ψ be the new hyperparameter, $f = p(\theta|y,\psi)$, $\underbrace{g = p(\theta|y,\psi_0)}_{\text{have a sample from this}}$.

$$\mathbb{E}[\theta|y,\psi] \approx \frac{\sum \theta_i \frac{f(\theta_i)}{g(\theta_i)}}{\sum \frac{f(\theta_i)}{g(\theta_i)}} = \frac{\sum \theta_i \frac{p(\theta_i|y,\psi)}{p(\theta_i|y,\psi_0)}}{\sum \frac{p(\theta_i|y,\psi)}{p(\theta_i|y,\psi_0)}}$$

$$= \frac{\sum \theta_i \frac{L(y|\theta)\pi(\theta|\psi)}{L(y|\theta)\pi(\theta|\psi_0)}}{\sum \frac{L(y|\theta)\pi(\theta|\psi)}{L(y|\theta)\pi(\theta|\psi_0)}} = \frac{\sum \theta_i \frac{\pi(\theta|\psi)}{\pi(\theta|\psi_0)}}{\sum \frac{\pi(\theta|\psi)}{\pi(\theta|\psi_0)}}$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Ex. Calculating Marginal likelihoods.

Suppose we have $f(y|u)$, the dist of y given some random effect u and $h(u|\theta)$ the dist. of random effects for parameter θ.

If

$$Y_{ij} \sim N(\mu + u_i, \sigma^2)$$

$$u_i \sim N(0, \theta)$$

We want to maximize

$$L(\theta) = \int f(y|u)\, h(u|\theta)\, du = E_h[f(y|u)]$$

↳ integrate out random effects.

Suppose we simulate $u_1 \ldots u_n$ from a candidate dist $\underbrace{h(u|\theta_0)}_{\text{wrong dist}}$. Then

$$\hat{L}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{h(u_i|\theta)}{h(u_i|\theta_0)} f(y|u_i)$$

Could use a more general candidate dist $g(u|\theta_0)$.

$$\hat{L}(\theta) = \frac{1}{n}\sum \frac{h(u_i|\theta)}{g(u_i|\theta_0)} f(y|u_i)$$

TOPIC:

DATE:

FILE UNDER: Mezzanol Litchhand

PAGE:

MCEM

Ideal candidate is proportional to

$\sum f(y|u) h(u|\theta)$ which is

$p(u|y,\theta)$. Since we want to maximize $\hat{L}$,
the best candidate is $p(u|y,\hat{\theta})$ where
$\hat{\theta}$ maximizes $\hat{L}$. But that solves our
problem. So try (Geyer, 1990)

(1) Simulate $\{u_i\}$ from $p(u|y,\theta_0)$

(2) Max $\hat{L}$ ~~to~~ get θ_1

(3) Set $\theta_0 = \theta_1$. Goto (1)

TOPIC:

DATE:

FILE UNDER:

PAGE:

$Y_1 \to Y_n \sim N(\mu, \sigma^2)$

$\pi(\mu) = N(0, \beta)$

$\pi(\sigma^2) = IG(\alpha, \beta)$

$c = \frac{1}{\sqrt{2\pi}}$ $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \frac{1}{n}\sum(y_i - \bar{y})^2$

$\frac{1}{a^2}$ vs $\frac{5}{3\,6}$

Marginal Likelihoods

$Y_{ij} \sim p(Y \mid u_i)$

$u_i \sim N(0, \tau^2)$

$$L(\tau) = \left\{ \prod_{i=1}^{n} \int \underbrace{\left[\prod_{j=1}^{J} \varphi(Y_{ij} \mid u_i) \right]}_{p(Y_i \mid u_i)} \varphi(u_i \mid \tau) \right\} du_i$$

$$= \prod_{i=1}^{n} \int p(y_i \mid u_i)\, \varphi(u_i \mid \tau)\, du_i$$

Let $u_{i1}, \to u_{iM} \sim \varphi(u_i \mid \tau_0)$

$$\widehat{\quad} \int p(Y_i \mid u_i)\, \varphi(u_i \mid \tau)\, du_i \approx$$

$$\frac{1}{n} \sum_{k=1}^{M} p(Y_i \mid u_{ik}) \frac{\varphi(u_{ik} \mid \tau)}{\varphi(u_{ik} \mid \tau_0)}$$