STAT 403

4/16/25

Lecture 7

OH 2-3 PDL
B-321

Lecture Notes III – Importance sampling and rejection sampling

Marina Meila
mmp@stat.washington.edu

Department of Statistics
University of Washington

April 2025

Importance sampling

Rejection sampling

The variance of $\hat{\mu}_{N,q}$

$\text{Var}\,\hat{\mu}_{f,q} \;=\; \frac{1}{N}\left(\int \tilde{f}^2(x)q(x)dx - \mu_f^2\right)$

- Only $M \;=\; \int \tilde{f}^2(x)q(x)dx$ depends on q

$$M \;=\; \int \frac{f^2(x)p^2(x)}{q^2(x)}q(x)dx \;=\; \int \frac{f^2(x)p^2(x)}{q(x)}dx$$

- Want q that makes M as small as possible
- $\boxed{q^*(x) \propto f(x)p(x)}$
- Let's see why. First we need to normalize q

$$q^*(x) \;=\; \frac{f(x)p(x)}{Z} \quad \text{with } Z \;=\; \int f(x)p(x)dx \;=\; \mu_f \tag{1}$$

- Now calculate M for q^*

$$M^* \;=\; \int \frac{f^2(x)p^2(x)}{\frac{f(x)p(x)}{\mu_f}}dx \;=\; \mu_f \underbrace{\int f(x)p(x)dx}_{\mu_f} \;=\; \mu_f^2 \tag{2}$$

- Finally, the variance $\text{Var}\,\hat{\mu}_{f,q} \;=\; \frac{1}{N}(M^* - \mu_f^2) \;=\; 0$!

- Theoretically, $N = 1$ sample is enough!
- But, q^* depends on the true μ_f that we are trying to estimate!

Importance sampling in practice

1. sample from

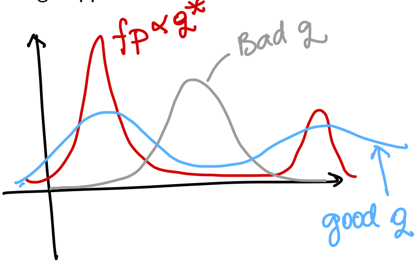
Theory $\boxed{q^*(x) \propto f(x)p(x)}$

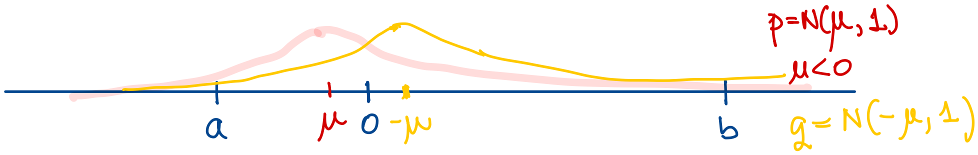
2. Average $\tilde{f} = \frac{f \cdot p}{q^*}$

$E_{q^*}[\tilde{f}] = E_p[f]$

smaller variance

- In practice, we want a distribution q so that
  - $q \approx q^*$
  - q is easy to sample from
  - $q(x)$ is easy to calculate for any x
- Rule of thumb 1: q should have modes where $f(x)p(x)$ is large
- Rule of thumb 2: avoid $q(x) \ll p(x)f(x)$ (tails of q should not decrease too fast!)
  When $f(x)p(x)$ is far from uniform, even a very rough approximation can reduce variance by orders of magnitude.





Wanted $\Pr[S_T \geq b] \ll 1$ $\quad S_k = X_1 + X_2 + \cdots X_k \quad X_{1,2,\ldots} \sim p$

$E[S_k] = k\mu$

$\operatorname{Var} S = k$

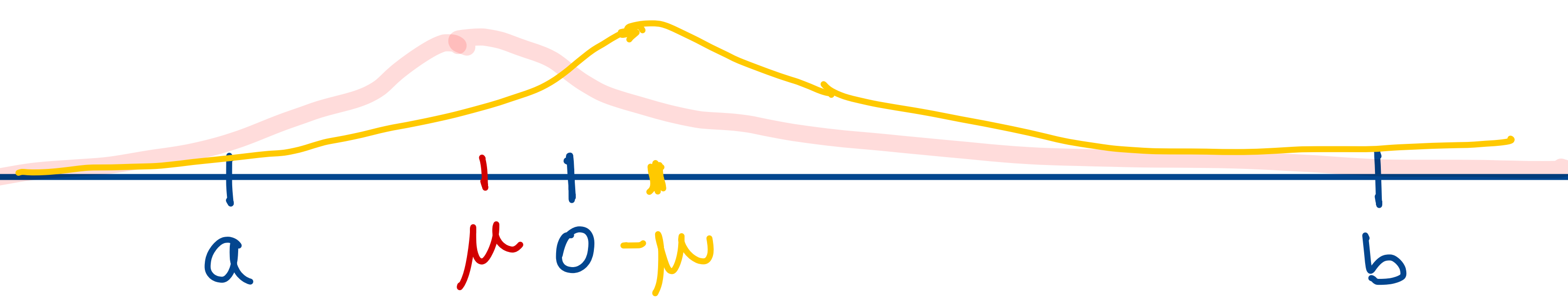
$T = \min_k \{S_k \notin [a,b]\}$

$$\tilde{f} = \frac{pf}{g} = \begin{cases} 0 \\ \frac{p}{g} = \dfrac{e^{-\frac{(S_T - T\mu)^2}{2T}}}{e^{-\frac{(S_T + T\mu)}{2T}}} = e^{-\frac{2\cancel{T}\mu S_T}{\cancel{T}}} \end{cases}$$

Table 5.2. Results of a Simulation Experiment to Find $\theta = P(S_T \geq b)$ for $T = \min\{n \mid S_n \leq a$ or $S_n \geq b\}$ with $S_n = X_1 + \cdots + X_n$, $X_i \sim N(\mu, 1)$ and $a = -4$, $b = 7$

μ	Direct $\hat{\theta}$	s.e. σ	$\hat{\theta}$	s.e. σ	Variance Reduction
0	0.389	0.0049	0.389	0.0055	1
−0.1	0.149	0.0035	0.147	0.0010	12
−0.2	0.041	0.0020	0.0412	1.8×10^{-4}	110
−0.3	0.011	0.0010	0.00996	3.8×10^{-5}	750
−0.5	0.0005	0.0007	0.000505	2.3×10^{-6}	9,600

$5e^{-4}$ $\quad 7e^{-4}$



Rejection sampling

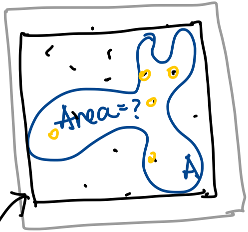
F^{-1}

- Sampling from $F(x)$ when we only know $p(x) = F'(x)$
- (or, when X is multidimensional)

Given $p(x)$ a density

Want An algorithm to get samples from $p(x)$

Ex: sample from uniform(A)



MC for Area

$$\widehat{\text{Area}} = \frac{\# X_i \in A}{N}$$

fewer X_i's rejected

$X_i \sim \text{uni}[\text{Rectangle}]$

Rejection sampling

- Sampling from $F(x)$ when we only know $p(x) = F'(x)$
- (or, when X is multidimensional)

Given $p(x)$ a density

Want An algorithm to get samples from $p(x)$

- Will use $q(x)$ another distribution → $p(x) > 0 \Rightarrow q(x) > 0$
- Let $\boxed{M = \sup_x \frac{p(x)}{q(x)}}$

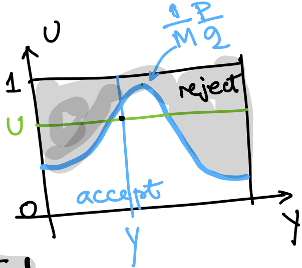
  Note that $p(x) > 0$ implies $q(x) > 0$

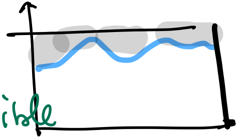
- Rejection Sampling Algorithm
  1. sample $Y \sim q$
  2. sample $U \sim \mathrm{unif}[0,1]$
  3. if $U < \frac{1}{M}\frac{p(Y)}{q(Y)}$ then output $X = Y$ accept y

     else go to step 1.

     ≤ 1

     otherwise no output (reject y)





good q: $q \propto p$

as close to p as possible

Practical Rejection Sampling. What is a good q?

- Intuitively, q should be as close to p as possible
- Acceptance probability $Pr\left[U < \frac{1}{M}\frac{p(Y)}{q(Y)}\right] = \frac{1}{M}$
- Therefore, a good q will have $M = \sup_x \frac{p(x)}{q(x)}$ small
  (Note that $M \geq 1$)

$\Rightarrow$ $q(x)$ not too close to 0 even when
p small

Examples

- $p = N(0,1)$

1) Cauchy: $g = \frac{1}{\pi}\frac{1}{1+x^2}$

$$F = \frac{1}{\pi}\left(\frac{\pi}{2} + \tan^{-1}(x)\right)$$

$$\boxed{M} = \sup_x \frac{p(x)}{g(x)} = \sup_x \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{\frac{1}{\pi}\frac{1}{1+x^2}} = \sup_x \sqrt{\frac{\pi}{2}}(1+x^2)e^{-\frac{x^2}{2}}$$

$$= \sqrt{\frac{\pi}{2}}(1+1)e^{-\frac{1}{2}} = \sqrt{\frac{\pi}{2e}}\,2 \approx \boxed{1.5}$$

$$\left((1+x^2)e^{-\frac{x^2}{2}}\right)' = \underbrace{2x}_{(1+x^2)'} e^{-\frac{x^2}{2}} - \underbrace{x e^{-\frac{x^2}{2}}}_{e^{-\frac{x^2}{2}}}(1+x)^2 = e^{-\frac{x^2}{2}}(x - x^3)$$

$x = 0 \Rightarrow \sqrt{\frac{\pi}{2}}$ $\quad x = \pm 1 \Rightarrow \boxed{\sqrt{\frac{\pi}{2}}\,2e^{-1/2}}$