Monte Carlo Integration

Expectation:

1) for discrete random variable X

$$E[g(X)] = \sum_{k=1}^{n} g(x_k)\, P(X = x_k)$$

1st moment: $g(x) = x$: $E(X) = \sum_{k=1}^{n} x_k P(X = x_k)$

2nd moment $g(x) = x^2$ $E(X^2) = \sum_{k=1}^{n} x_k^2 P(X = x_k)$

2) for continuous random variables X

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

Example: exponential random variable

$f(x) = \lambda e^{-\lambda x} \cdot 1_{\{x \geq 0\}}$, where $\lambda > 0$ - rate parameter

$X \sim Exp(\lambda)$ $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{0}^{\infty} x \lambda e^{-\lambda x} = [$ integration by parts $] = \ldots = \frac{1}{\lambda}$.

Expectations are linear operators:

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

This does not hold for variances in general,

$$\left(Var(aX) = E\left[(aX - E(aX))^2\right] \right)$$

but if $X_1, \ldots, X_k$ are independent, then

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i)$$

Strong Law of Large Numbers (SLLN)

Let $X_1, X_2, \ldots$ be independent and identically distributed (iid) random variables with

$$\mu = E(X_1) < \infty \quad \text{Then}$$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu$$

SLLN says that empirical averages of iid random variables converge to the theoretical average/expectation.

Monte Carlo Integration

<u>Objective</u>: $E[h(X)] = \int h(x) f(x)\, dx$, where X is a random variable with probability density function $f(x)$ or

$$E(h(X)) = \sum_{k=1}^{\infty} h(x_k)\, p_k \quad ,$$

where X is a discrete random variable with prob. mass function $p_1, p_2, \ldots$

If $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ and $E(h(X_1)) < \infty$, then

$$\text{SLLN} \Rightarrow \frac{1}{n} \sum_{i=1}^{n} h(X_i) \to E(h(X_1))$$

$$\bar{h}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i) \approx E(h(X_1))$$

Central Limit Theorem (CLT)

$X_1, X_2, \ldots$ be iid random variables with $\mu = E(X_1) < \infty$ and $0 < \sigma^2 = Var(X_1) < \infty$ and let

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then

$$\frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \sim N(0,1) \text{ for large } n$$

approximately

Informally, CLT says that for large n, the empirical average $\bar{X}_n$ behaves as $N(\mu, \frac{\sigma^2}{n})$

Scaling of the variance by $\frac{1}{n}$ implies that averaging reducing variability, which is intuitive.

Recall that during Monte Carlo integration we use $\bar{h}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$ to approximate $E(h(X))$

CLT $\Rightarrow$ $\bar{h}_n \sim N(\mu, \frac{\sigma^2}{n})$

approximately

or

$\bar{h}_n \sim N(\mu, \frac{v_n}{n})$, where

$v_n = \frac{1}{n-1}\sum_{i=1}^{n}[h(X_i) - \bar{h}_n]^2$ – sample variance

We can form 95% confidence interval

$\bar{h}_n \pm 1.96\sqrt{\frac{v_n}{n}}$ – Monte Carlo error

Importance Sampling

Objective: $E_f[h(X)] = \int h(x) f(x)dx$

target density

We cannot or don't want to sample from the target density. Instead, we want to sample from some other, perhaps simpler, distribution with density $g(x)$.

$$E_f[h(x)] = \int h(x) f(x)\, dx = \int \underbrace{\boxed{h(x)\frac{f(x)}{g(x)}}}_{z(x)} g(x)\, dx =$$

$$= \int z(y)\, g(y)\, dy = E_g[z(y)] =$$

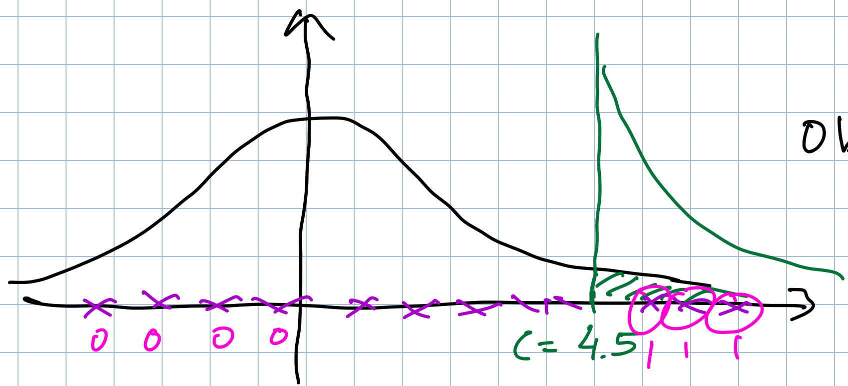
$$= E_g\left[h(y)\frac{f(y)}{g(y)}\right], \text{ where } y \sim g(y)$$

This construction suggests that we can generate iid $y_1, \dots, y_n \overset{iid}{\sim} g(y)$ and use SLLN to arrive at

$$E_f[h(x)] \approx \frac{1}{n}\sum_{i=1}^{n} \boxed{\frac{f(y_i)}{g(y_i)}}\, h(y_i)$$

w_i - importance sampling weights

Example: approximation of the tail prob of $N(0,1)$



$Z \sim N(0,1)$

Objective: $P(Z>c) =$

$= E(1_{\{Z>c\}})$

1_A ↑ event $Pr(A)=p$ $1_A \sim \text{Bernoulli}(p)$

$$E(1_A) = 1\cdot p + 0\cdot(1-p) = p = P(A)$$

Naive Monte Carlo: $Z_1, \dots, Z_n \overset{iid}{\sim} N(0,1)$

Then $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{Z_i > c\}} \approx E(1_{\{Z>c\}}) = P(Z>c)$

Importance sampling:

Simulate $y_1, \dots, y_n \overset{iid}{\sim}$ Shifted-Exp$(c, 1)$

Shifted exponential density: $g(y) = e^{-(y-c)} \cdot 1_{\{y>c\}}$

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{\varphi(y_i)}{g(y_i)} \cancel{\cdot 1_{\{y_i>c\}}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\varphi(y_i)}{g(y_i)}$$

$\varphi(x)$ - density of $N(0,1)$