Importance Sampling

Jessi Cisewski Carnegie Mellon University

June 2014

Outline

- Recall: Monte Carlo integration
- Importance Sampling
- Examples of Importance Sampling





(a) Monte Carlo, Monaco

(b) Monte Carlo Casino

★ Some content and examples from Wasserman (2004)

Jessi Cisewski (CMU)

Importance Sampling

Simple illustration: what is π ?

$$\frac{Area_{\circ}}{Area_{\Box}} = \frac{\pi r^2}{(2r)(2r)} = \frac{\pi}{4}$$



Monte Carlo Integration: motivation

$$I=\int_a^b h(y)dy$$

- Goal: evaluate this integral
- Sometimes we can find I (e.g. if $h(\cdot)$ is a function from Calc I)
- But sometimes we can't and need a way to approximate *I*. Monte Carlo methods are one (of many) approaches to do this.

The Law of Large Numbers

While nothing is more uncertain than the duration of a single life, nothing is more certain than the average duration of a thousand lives.

 \sim Elizur Wright



Figure: Elizur Wright (1804 - 1885), American mathematician, the "father of life insurance", "father of insurance regulation" (http://en.wikipedia.org)

Law of Large Numbers

- The Law of Large Numbers describes what happens when performing the same experiment many times.
- After many trials, the average of the results should be close to the expected value and will be more accurate with more trials.
- For Monte Carlo simulation, this means that we can learn properties of a random variable (mean, variance, etc.) simply by simulating it over many trials.

Suppose we want to estimate the probability, p, of a coin landing "heads up". How many times should we flip the coin?



Law of Large Numbers (LLN)

Given an independent and identically distributed sequence of random variables Y_1, Y_2, \ldots, Y_n with $\overline{Y}_n = n^{-1} \sum_{i=1}^n Y_i$ and $E(Y_i) = \mu$, then for every $\epsilon > 0$

$$P(|\bar{Y}_n-\mu|>\epsilon)\longrightarrow 0,$$

as $n \longrightarrow \infty$.

Monte Carlo Integration

General idea

Monte Carlo methods are a form of stochastic integration used to approximate expectations by invoking the law of large numbers.

$$I = \int_a^b h(y) dy = \int_a^b w(y) f(y) dy = E_f(w(Y))$$

where $f(y) = \frac{1}{b-a}$ and $w(y) = h(y) \cdot (b-a)$

- $f(y) = \frac{1}{b-a}$ is the pdf of a U(a,b) random variable
- By the LLN, if we take an iid sample of size N from U(a, b), we can estimate I as

$$\hat{I} = N^{-1} \sum_{i=1}^{N} w(Y_i) \longrightarrow E(w(Y)) = I$$

Monte Carlo Integration: standard error

$$I = \int_a^b h(y) dy = \int_a^b w(y) f(y) dy = E_f(w(Y))$$

- Monte Carlo estimator: $\hat{I} = N^{-1} \sum_{i=1}^{N} w(Y_i)$
- Standard error of estimator: $\hat{SE} = \frac{s}{\sqrt{N}}$ where

$$s^{2} = (N-1)^{-1} \sum_{i=1}^{N} \left(w(Y_{i}) - \hat{I} \right)^{2}$$

Monte Carlo Integration: Gaussian CDF example*

• Goal: estimate $F_Y(y) = P(Y \le y) = E\left[I_{(-\infty,y)}(Y)\right]$ where $Y \sim N(0,1)$:

$$F(Y \le y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \int_{-\infty}^{\infty} h(t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

where h(t) = 1 if t < y and h(t) = 0 if $t \ge y$

Draw an iid sample Y₁,..., Y_N from a N(0,1), then the estimator is

$$\hat{I} = N^{-1} \sum_{i=1}^{N} h(Y_i) = \frac{\# \text{ draws } < x}{N}$$

★ Example 24.2 of Wasserman (2004)

Importance Sampling: motivation

- Standard Monte Carlo integration is great if you can sample from the *target* distribution (i.e. the desired distribution)
 - \longrightarrow But what if you can't sample from the target?
- Idea of importance sampling: draw the sample from a *proposal* distribution and re-weight the integral using *importance weights* so that the correct distribution is targeted



Monte Carlo Integration \longrightarrow Importance Sampling

$$I=\int h(y)f(y)dy$$

- h is some function and f is the probability density function of Y
- When the density *f* is difficult to sample from, importance sampling can be used
- Rather than sampling from *f*, you specify a different probability density function, *g*, as the proposal distribution.

$$I = \int h(y)f(y)dy = \int h(y)\frac{f(y)}{g(y)}g(y)dy = \int \frac{h(y)f(y)}{g(y)}g(y)dy$$

Importance Sampling

$$I = E_f[h(Y)] = \int \frac{h(y)f(y)}{g(y)}g(y)dy = E_g\left[\frac{h(Y)f(Y)}{g(Y)}\right]$$

Hence, given an iid sample Y_1, \ldots, Y_N from g, our estimator of I becomes

$$\hat{I} = N^{-1} \sum_{i=1}^{N} \frac{h(Y_i)f(Y_i)}{g(Y_i)} \longrightarrow E_g\left[\frac{h(Y)f(Y)}{g(Y)}\right] = I$$

Importance Sampling: selecting the proposal distribution

• The standard error of \hat{l} could be infinite if $g(\cdot)$ is not selected appropriately $\longrightarrow g$ should have thicker tails than f (don't want ratio f/g to get large)

$$E_{g}\left[\left(\frac{h(Y)f(Y)}{g(Y)}\right)^{2}\right] = \int \left(\frac{h(y)f(y)}{g(y)}\right)^{2}g(y)dy$$

- Select a g that has a similar shape to f, but with thicker tails
- Variance of \hat{I} is minimized when $g(y) \propto |f(y)|$
- Want to be able to sample from g(y) with ease

Importance sampling: Illustration

- Goal: estimate P(Y < 0.3) where $Y \sim f$
- Try two proposal distributions: U(0,1) and U(0,4)



Importance sampling: Illustration, continued.

If take 1000 samples of size 100, and find the IS estimates, we get the following *estimated* expected values and variances.

	Expected Value	Variance
Truth	0.206	0
$g_1: U(0,1)$	0.206	0.0014
g ₂ : U(0,4)	0.211	0.0075



Monte Carlo Integration: Gaussian tail probability example*

• Goal: estimate $P(Y \ge 3)$ where $Y \sim N(0,1)$ (Truth is \approx 0.001349)

$$P(Y > 3) = \int_{3}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^{2}/2} dt = \int_{-\infty}^{\infty} h(t) \frac{1}{\sqrt{2\pi}} e^{-t^{2}/2} dt$$

where h(t) = 1 if t > 3 and h(t) = 0 if $t \le 3$

• Draw an iid sample Y_1, \ldots, Y_{100} from a N(0,1), then the estimator is

$$\hat{I} = \frac{1}{100} \sum_{i=1}^{100} h(Y_i) = \frac{\# \text{ draws } > 3}{100}$$

★ Example 24.6 of Wasserman (2004)

Gaussian tail probability example*, continued.

• Draw an iid sample Y_1, \ldots, Y_{100} from a N(0, 1), then the estimator is

$$\hat{U} = \frac{1}{100} \sum_{i=1}^{100} h(Y_i)$$

• Draw an iid sample Y_1, \ldots, Y_{100} from a N(4, 1), then the estimator is

$$\hat{l} = rac{1}{100} \sum_{i=1}^{100} rac{h(Y_i)f(Y_i)}{g(Y_i)}$$

where f is the density of a N(0,1) and g is the density of N(4,1)

★ Example 24.6 of Wasserman (2004)

Gaussian tail probability example*, continued.

If take N samples of size 100, and find the MC and IS estimates, we get the following *estimated* expected values and variances.

$N = 10^{5}$

	Expected Value	Variance
Truth	0.00135	0
Monte Carlo	0.00136	$1.3 imes10^{-5}$
Importance Sampling	0.00135	$9.5 imes10^{-8}$

Extensions of Importance Sampling

- Sequential Importance Sampling
- Sequential Monte Carlo (Particle Filtering)
 → See Doucet et al. (2001)
- Approximate Bayesian Computation → See Turner and Zandt (2012) for a tutorial, and Cameron and Pettitt (2012); Weyant et al. (2013) for applications to astronomy

Bibliography

- Cameron, E. and Pettitt, A. N. (2012), "Approximate Bayesian Computation for Astronomical Model Analysis: A Case Study in Galaxy Demographics and Morphological Transformation at High Redshift," *Monthly Notices of* the Royal Astronomical Society, 425, 44–65.
- Doucet, A., De Freitas, N., and Gordon, N. (2001), Sequential Monte Carlo Methods in Practice, Statistics for Engineering and Information Science, New York: Springer-Verlag.
- Turner, B. M. and Zandt, T. V. (2012), "A tutorial on approximate Bayesian computation," Journal of Mathematical Psychology, 56, 69 – 85.
- Wasserman, L. (2004), All of statistics: a concise course in statistical inference, Springer.
- Weyant, A., Schafer, C., and Wood-Vasey, W. M. (2013), "Likeihood-free cosmological inference with type la supernovae: approximate Bayesian computation for a complete treatment of uncertainty," *The Astrophysical Journal*, 764, 116.