Monte Carlo Methods: Monte Carlo Integration

Monte Carlo integration

- Let π be a density on $\mathbb{R}^d$. For example if d $= 1$, π is a density on $\mathbb{R}$.
- Let $h : \mathbb{R}^d \to \mathbb{R}$ a function. Suppose we want to evaluate the integral $\pi(h) = \int h(x)\pi(x)dx = \mathbb{E}(h(X))$.
- For instance h could be the function $h(x_1 \ldots, x_d) = x_1$. In which case we want to compute

$$\pi(h) = \int x_1 \pi(x) dx = \mathbb{E}(X_1),$$

the expectation of the first component of X.

Monte Carlo integration

- Very often, such integral can be hard to evaluate analytically.
- If d is not too large, we can use numerical methods (simpson, trapezoid rules, quadrature methods etc...).
- But for d large ($d \geq 4$), numerical methods become inefficient and we have to turn to Monte Carlo methods.

Monte Carlo integration

- The basic idea of Monte Carlo integration is very simple: generate $X_1, \ldots, X_n$ i.i.d. from π.
- By the LLN $n^{-1} \sum_{k=1}^{n} h(X_k)$ converge to $E(h(X))$ (we will use the notation $\pi(h)$ to denote $E(h(X))$).
- In other words, we approximate the integral $\pi(h) = E(h(X))$ by

$$\pi_n(h) = \frac{1}{n} \sum_{k=1}^{n} h(X_k).$$

- $\pi_n(h)$ is called the Monte Carlo estimate of $\pi(h)$ or the Basic Monte Carlo estimate of $\pi(h)$.

Monte Carlo integration

$\pi_n(h)$ has the following properties

Theorem

1. *By the Law of large numbers,* $\pi_n(h)$ *converges to* $\pi(h)$ *as* $n \to \infty$.
2. $\mathbb{E}(\pi_n(h)) = \pi(h)$ *and* $Var(\pi_n(h)) = \frac{\sigma^2(h)}{n}$ *where* $\sigma^2(h) = Var(h(X)) = \mathbb{E}(h^2(X)) - (\mathbb{E}(h(X)))^2$.
3. *By the central limit theorem*

$$\pi_n(h) \approx \mathcal{N}\left(\pi(h), \frac{\sigma^2(h)}{n}\right).$$

Monte Carlo integration

- According to the CLT, the fluctuations or precision of $\pi_n(h)$ is given by $\frac{\sigma^2(h)}{n}$ or its square root $\frac{\sigma(h)}{\sqrt{n}}$.
- We can estimate $\sigma^2(h)$ by

$$s_n^2(h) = \frac{1}{n-1}\sum_{k=1}^{n}\left(h(X_k) - \pi_n(h)\right)^2 .$$

- The quantity $\frac{s_n(h)}{\sqrt{n}}$ is known as the Monte Carlo error.
- In any Monte Carlo experiment, always report the Monte Carlo error or the Monte Carlo confidence interval

$$\pi_n(h) \pm z_{\alpha/2}\frac{s_n(h)}{\sqrt{n}}.$$

Monte Carlo integration

Example

- Suppose we want to calculate the integral

$$I = \int_0^1 (\cos(50x) + \sin(20x))^2 \, dx.$$

- It is possible to do this analytically. The exact value is 0.965.
- We can also approximate I very well using numerical integration methods (function `integrate` in R)
- To use Monte Carlo, we could see I as $\mathbb{E}(h(U))$ where $h(x) = (\cos(50x) + \sin(20x))^2$ and $U \sim \mathcal{U}(0,1)$. Then we generate $U_1, \ldots, U_n$ i.i.d. from $\mathcal{U}(0,1)$ and approximate I by $\hat{I}_n = n^{-1} \sum_{k=1}^{n} h(U_k)$.

Monte Carlo integration

```
hfun=function(x){
    return((cos(50*x)+sin(20*x))^2)
}

nSple=10000
uniforms=runif(nSple)
hvalues=hfun(uniforms)
Ihat=mean(hvalues) #Monte Carlo estimate
se=sqrt(var(hvalues)/nSple) # Monte Carlo error
z=qnorm(0.975)
CI=c(Ihat-z*se,Ihat+z*se) # 95% confid. interv.
##We report either Ihat and se, or CI.
```

Monte Carlo integration

- An important step to apply the Monte Carlo technique to an integral $I = \int v(x)dx$ is to rewrite I as an expectation.
- This step requires some care because it can be done in many different ways. For instance, suppose we want to approximate $I = \int_0^1 6x^2(1-x)dx$.
- We can write $I = \mathbb{E}(h_1(U))$, where $h_1(x) = 6x^2(1-x)$ and $U \sim \mathbf{U}(0,1)$.
- But we could also write $I = \mathbb{E}(h_2(V))$, were $h_2(x) = x$, and $V \sim \text{Beta}(2,2)$, the Beta distribution with parameter $(2,2)$.

Monte Carlo integration

- Another example. Consider $I = \int_0^1 (\cos(50x) + \sin(20x))^2 \, dx$ seen above. We could have also used the Beta distribution. How?
- Let $f_{\alpha,\beta}(x) = \frac{1}{\text{Beta}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0 < x < 1$, be the density of the Beta distribution. We can write.

$$I = \int_0^1 \frac{(\cos(50x) + \sin(20x))^2}{f_{\alpha,\beta}(x)} f_{\alpha,\beta}(x) dx.$$

- We have re-written the same integral as a different expectation $\mathbb{E}(h(X))$, where here $X \sim \text{Beta}(\alpha,\beta)$, and

$$h(x) = \frac{(\cos(50x) + \sin(20x))^2}{f_{\alpha,\beta}(x)}.$$

Monte Carlo integration

- The idea of moving from one expectation representation to another, for the same integral, is known as importance sampling.
- We develop this in the next section.

Importance sampling

- We have discussed the Basic Monte idea to approximate an integral $\int h(x)\pi(x)dx$ where π is a density.
- We generate $X_1, \ldots, X_n$ i.i.d. from π and compute $n^{-1}\sum_{k=1}^{n} h(X_k)$.
- But we can write an integral in many different ways. Let g be another density such that $g(x) > 0$ whenever $h(x)\pi(x) > 0$. Then write

$$\int h(x)\pi(x)dx = \int h(x)\frac{\pi(x)}{g(x)}g(x)dx = \int h(x)\omega(x)g(x)dx,$$

  where $\omega(x) = \frac{\pi(x)}{g(x)}$. and g another density.
- Now we can apply the Basic Monte Carlo and sample from g.

Importance sampling

How does it work?

To evaluate $\pi(h) = \int h(x)\pi(x)dx$ we introduce another density g and define the importance ratio $\omega(x) = \frac{\pi(x)}{g(x)}$:

$$\pi(h) = \int h(x)\pi(x)dx = \int h(x)\omega(x)g(x)dx.$$

- Then we generate $Y_1, \ldots, Y_n$ i.i.d. from g and approximate $\pi(h)$ by

$$\pi_{n,IS}(h) = \frac{1}{n}\sum_{k=1}^{n} h(Y_k)\omega(Y_k).$$

Importance sampling

- As in Basic Monte Carlo

$$\frac{1}{n}\sum_{k=1}^{n} h(Y_k)\omega(Y_k) \approx \mathcal{N}\left(\pi(h), \frac{\sigma_{IS}^2(h)}{n}\right),$$

where $\sigma_{IS}^2(h) = \mathrm{Var}\left(h(Y)\omega(Y)\right)$, and we can estimate $\sigma_{IS}^2(h)$ by

$$s_{n,IS}^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(h(Y_k)\omega(Y_k) - \pi_{n,IS}(h)\right)^2.$$

- So again we either report $\pi_{n,IS}(h)$ and $\sqrt{s_{n,IS}^2(h)/n}$ or the confidence interval $\pi_{n,IS}(h) \pm z_{\alpha/2}\sqrt{s_{n,IS}^2(h)/n}$.

Importance sampling

How do we choose g?

- Coverage issue: if $\pi(x) > 0$ for some x, make sure $g(x) > 0$. For example if $\pi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ $x \in \mathbb{R}$ don't use $g(x) = e^{-x}$ for $x > 0$ and $g(x) = 0$ otherwise.
- Resemblance: make sure g is close to $|h(x)|\pi(x)$.
- Stability: make sure the function $\omega(x)$ is bounded: $\pi(x) \leq Mg(x)$ for some finite $M > 0$.

Importance sampling

Exercise: Recall that $\sigma^2(h) = \mathrm{Var}(h(X))$ (where $X \sim \pi$). Suppose that $\sigma^2(h) < \infty$ and we choose g such that $\pi(x) \leq Mg(x)$ for some $M < \infty$. Show that $\mathrm{Var}\left(h(Y)\omega(Y)\right)$ (where $Y \sim g$) is also finite.

Importance sampling

Example: Compute $p = \Pr(Z \geq a)$ for $a = 2.3$, where $Z \sim N(0, 1)$.

- We can write

$$p = \int_{-\infty}^{+\infty} h(x)\pi(x)dx,$$

where π is the density of $N(0, 1)$, and

$$h(x) = \begin{cases} 1 & \text{if } x \in [a, \infty), \\ 0 & \text{otherwise.} \end{cases}$$

- Basic Monte Carlo: Generate $Z_1, \ldots Z_n \sim N(0, 1)$ and form

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} h(X_i),$$

Importance sampling

```
hfun =function(x,a) (x>=a)
##
a=2.3
n=20000
z=qnorm(0.975)
answer=1-pnorm(a)
#Estimate p by Basic Monte
Z=rnorm(n);
hX=hfun(Z,a)
p_hat=mean(hX);MC_err=sqrt(var(hX)/n)
CI=c(p_hat-z*MC_err,p_hat+z*MC_err)
```

Importance sampling

We could also use importance sampling by generating from $g_\alpha(x) = \alpha e^{-\alpha(x-a)}$, for $x \geq a$.

$$p = \int_a^{+\infty} \pi(x)dx = \int_a^{+\infty} \frac{\pi(x)}{g_\alpha(x)} g_\alpha(x)dx = \int_a^{+\infty} \omega(x)g_\alpha(x)dx.$$

where

$$\omega(x) = \frac{1}{\alpha\sqrt{2\pi}} e^{\alpha(x-a)-\frac{1}{2}x^2}.$$

```
omfun =function(x,a,alpha){
y=(1/sqrt(2*pi))*(1/alpha)
            *exp(-0.5*x^2+alpha*(x-a))
return(y)
}
```

Importance sampling

```
#Estimate by Importance Smpling
alpha=a;
Y=a + rexp(n,alpha)
hY=hfun(Y,a)*omfun(Y,a,alpha)
p_hat2=mean(hY);
MC_err2=sqrt(var(hY)/n)
CI=c(p_hat2-z*MC_err2,p_hat2+z*MC_err2)
```

Importance sampling

- Consider again importance sampling for calculating $\pi(h) = \int h(x)\pi(x)dx$:
$$\pi(h) = \int h(x)\pi(x)dx = \int h(x)\omega(x)g(x)dx,$$
where $\omega(x) = \pi(x)/g(x)$.
- In many applications of IS the importance ratio ω cannot be computed. The prime example where this happens is in Bayesian inference.
- Often all that can be computed is $\tilde{\omega}(x) = C\omega(x)$ for some constant C.
- In these cases we do IS as follows: generate $Y_1, \ldots, Y_n$ from g and form
$$\tilde{\pi}_{n,IS} = \frac{\sum_{k=1}^{n} h(Y_k)\tilde{\omega}(Y_k)}{\sum_{k=1}^{n} \tilde{\omega}(Y_k)}.$$

Importance sampling

- Notice that

$$
\begin{aligned}
\tilde{\pi}_{n,IS} &= \frac{\sum_{k=1}^{n} h(Y_k)\tilde{\omega}(Y_k)}{\sum_{k=1}^{n} \tilde{\omega}(Y_k)} \\
&= \frac{\frac{1}{n}\sum_{k=1}^{n} h(Y_k)\omega(Y_k)}{\frac{1}{n}\sum_{k=1}^{n} \omega(Y_k)} \\
&\to \frac{\int h(x)\pi(x)dx}{1}.
\end{aligned}
$$

- Hence it is also correct to use this estimate.

Importance sampling

- But the Monte Carlo error for the new estimate is more difficult to evaluate.
- Notice that the quantity $n^{-1}\sum_{k=1}^{n}\tilde{\omega}(Y_k)$ converges to $\int C\omega(x)g(x)dx = C$.
- Thus we can assess the estimate $\tilde{\pi}_{n,IS}$ by looking how well the estimation of C is doing. To do so we compute the sample coefficient of variation:

$$CV = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(\frac{\tilde{\omega}(Y_k)}{\bar{\omega}} - 1\right)^2},$$

where $\bar{\omega} = \frac{1}{n}\sum_{k=1}^{n}\tilde{\omega}(Y_k)$.

Importance sampling

Take home method:

- Calculate the fluctuation of the weights:

$$CV = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(\frac{\tilde{\omega}(Y_k)}{\bar{\omega}} - 1\right)^2},$$

  where $\bar{\omega} = \frac{1}{n}\sum_{k=1}^{n}\tilde{\omega}(Y_k)$.

- If CV is small (say smaller than 5) then compute $\tilde{\pi}_{n,IS}$ and its Monte Carlo error $\sqrt{\hat{\sigma}^2_{n,IS}/n}$.

$$\hat{\sigma}^2_{n,IS} = \frac{1}{n-1}\sum_{k=1}^{n}\left(Z_{n,k} - \tilde{\pi}_{n,IS}\right)^2, \quad Z_{n,k} = \frac{\tilde{\omega}(Y_k)}{\frac{1}{n}\sum_{k=1}^{n}\tilde{\omega}(Y_k)} h(Y_k).$$

- Otherwise $\tilde{\pi}_{n,IS}$ is too unstable and should not be trusted.

Importance sampling

Example:

- We want to sample from the distribution

$$\pi(x) = \frac{1}{C}\frac{1}{1+x^2}e^{-(x-2)^2/2} \tag{1}$$

  and calculate $\delta = \int x\pi(x)dx$.
- C is called the normalizing constant. $C = \int \frac{1}{1+x^2}e^{-(x-2)^2/2}dx$ and is unknown.
- It is thus important to have Monte Carlo methods where the knowledge of C is not required like the weighted average importance sampling.

Importance sampling

- We use importance sampling with the Cauchy distribution $\mathcal{C}(1,1)$.
$$g(x) = \frac{1}{c_0(1+(x-1)^2)},$$
where $c_0 = \pi$ (the irrational number).
- We can plot the functions $\tilde{\pi}(x) = \frac{1}{1+x^2} e^{-(x-2)^2/2}$ and $g(x) = \frac{1}{c_0(1+(x-1)^2)}$ to see how close they are.
- We need the importance function
$$\omega(x) = \frac{\pi(x)}{g(x)} = \frac{c_0}{C} \frac{1+(x-1)^2}{1+x^2} e^{-\frac{1}{2}(x-2)^2}.$$

Importance sampling

- To do importance sampling, we generate $Y_1, \ldots, Y_n$ iid from $\mathcal{C}(1,1)$ and if we knew C we can estimate δ by:

$$\frac{1}{n}\sum_{i=1}^{n} Y_k \omega(Y_k).$$

- Clearly, the problem is that C is not known!
- This is where the weighted average version becomes useful.

Importance sampling

- We drop C (and c_0). The new importance function is
$$\tilde{\omega}(x) = \frac{1+(x-1)^2}{1+x^2} e^{-\frac{1}{2}(x-2)^2}.$$
- The new importance sampling: We sample $Y_1, \ldots, Y_n$ iid from $\mathcal{C}(1,1)$ and estimate δ by:
$$\frac{\sum_{k=1}^n Y_k \tilde{\omega}(Y_k)}{\sum_{i=1}^n \tilde{\omega}(Y_i)}.$$

Importance sampling

To implement this method in R we need to

- write a function *omegafun* that calculates the function $\tilde{\omega}$
- know how to generate from $\mathcal{C}(1,1)$ (rcauchy(n,mu,sigma).
- That's it.

Importance sampling

```
omfun=function(x){
   v=exp(-(2-x)^2/2)*(1+(x-1)^2)/(1+x^2)
}
```

Importance sampling

```
n=10000
Sple=1+rcauchy(n)
Om=omfun(Sple)
est=sum(Sple*Om)/sum(Om)# estimate
CVsq=var(Om/mean(Om))
v1=var((Om/mean(Om))*Sple)
se=sqrt(v1/n)
z=qnorm(0.975)
CI=(est-z*se,est+z*se)
```

Question: Use the function `integrate` to confirm the answer we obtained.