STATS 406 Fall 2016: Lab 12

1 The basic Monte Carlo method

The goal is to evaluate the integral $\pi(h) = \int h(x)\pi(x)dx = E(h(X))$, where π is a density on $\mathbb{R}^d$ and h is a function on $\mathbb{R}^d$, random vector X has density π. When analytical calculation and numerical methods both don't work well, we can use Monte Carlo methods.

The basic Monte Carlo method is to generate $X_1, \ldots, X_n$ i.i.d. from density π and use

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

to estimate the integral. $\pi_n(h)$ is called the Monte Carlo estimate. By the Law of Large Numbers, this quantity converges to $E(h(x))$.

The standard deviation of $\pi_n(h)$ can be estimated by $\frac{s_n(h)}{\sqrt{n}}$, the Monte Carlo error. Here $s_n(h)^2 = \frac{1}{n-1}\sum_{k=1}^{n}(h(X_k) - \pi_n(h))^2$.

The $\pi_n(h) \pm z_{\alpha/2}\frac{s_n(h)}{\sqrt{n}}$ gives the approximate $(1-\alpha)$ level Monte Carlo confidence interval of $E(h(X))$.

Examples: see lecture notes and Lab 9, Lab 10.

2 Another Monte Carlo method: importance sampling

Importance sampling is a Monte Carlo technique based on transforming the integral into another representation of expectation:

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

where g is another density on $\mathbb{R}^d$ and $\omega(x) = \frac{\pi(x)}{g(x)}$. Here the random vector Y has density g, not π.

When the ratio ω can be fully calculated

Similar as in the basic Monte Carlo case, we generate $Y_1, \ldots, Y_n$ i.i.d. from density g and use

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

to estimate the integral. The Monte Carlo error is given by $\frac{s_{n,IS}(h)}{\sqrt{n}}$ where $s_{n,IS}(h)^2 = \frac{1}{n-1}\sum_{k=1}^{n}(h(Y_k)\omega(Y_k) - \pi_{n,IS}(h))^2$. And the confidence interval is $\pi_{n,IS}(h) \pm z_{\alpha/2}\frac{s_{n,IS}(h)}{\sqrt{n}}$.

When the ratio ω is known up to a constant

In some cases the ratio ω is only known up to a constant, $\tilde{\omega}(x) = C\omega(x)$ where C is not feasible to calculate. Then we can generate $Y_1, \ldots, Y_n$ i.i.d. from density g use the following

$$\tilde{\pi}_{n,IS} = \frac{\sum_{k=1}^{n} h(Y_k)\tilde{\omega}(Y_k)}{\sum_{k=1}^{n} \tilde{\omega}(Y_k)}$$

to estimate the integral.

Rule: we calculate

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

If this is small, say less than 5, then the method is reliable and we can use the estimate. Let $Z_{n,k} = \frac{\tilde{\omega}(Y_k)}{\frac{1}{n}\sum_{i=1}^{n}\tilde{\omega}(Y_i)}h(Y_k)$, then we have $\tilde{\pi}_{n,IS} = \frac{1}{n}\sum_{k=1}^{n} Z_{n,k}$. Then the Monte Carlo error is given by $\sqrt{\hat{\sigma}^2_{n,IS}/n}$, where $\hat{\sigma}^2_{n,IS} = \frac{1}{n-1}\sum_{k=1}^{n}(Z_{n,k} - \tilde{\pi}_{n,IS})^2$. And the confidence interval is $\tilde{\pi}_{n,IS} \pm z_{\alpha/2}\frac{\hat{\sigma}_{n,IS}}{\sqrt{n}}$.

Examples: see lecture notes and Lab 10.

3 Application of Monte Carlo methods: Bayesian inference

Importance sampling has application in Bayesian inference, where the posterior distribution of interest is usually only known up to a constant.

In Bayesian inference, assume the observed data $y_1, ..., y_n$ has density $f(y|\theta)$ where θ is the parameter. θ is viewed as random, and one wants to make inference about the posterior density $\pi(\theta|y_1, ..., y_n)$ where

$$\pi(\theta|y_1, ..., y_n) = \frac{p(\theta, y_1, ..., y_n)}{p(y_1, ..., y_n)} = \frac{f(y_1, ..., y_n|\theta)\pi(\theta)}{p(y_1, ..., y_n)}$$

The normalizing constant in the denominator is often not feasible to calculate. We can apply importance sampling to the unnormalized posterior density

$$f(y_1, ..., y_n|\theta)\pi(\theta)$$

and estimate quantities of interest such as the mean of the posterior distribution.

Examples: see lecture notes and Lab 11.

4 Application of Monte Carlo methods: bootstrap

Suppose $\mathcal{X} = \{X_1, \ldots, X_n\}$ are i.i.d. samples from a cumulative distribution function F_{θ_0}, and we have an estimate $\hat{\theta}$ for the true unknown parameter θ_0. Here we view θ_0 as fixed. We can use bootstrap to estimate quantities such as the MSE, i.e. mean square error:

$$\text{MSE}(\hat{\theta}) = E((\hat{\theta} - \theta_0)^2)$$

The basic steps of bootstrap (taking MSE as an example) are

1. Draw bootstrap samples $\mathcal{X}_1^*, ..., \mathcal{X}_K^*$, each of size n, based on the original sample $\mathcal{X}$.
2. Compute sample statistic $\hat{\theta}_k^*$ for each sample $\mathcal{X}_k^*$.
3. Estimate the MSE by $\frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}_k^* - \hat{\theta})^2$, where $\hat{\theta}$ is the estimate from $\mathcal{X}$.

Parametric bootstrap

Suppose $X_1, \ldots, X_n$ are from a parametric family with density f_{θ_0}, in step 1 in bootstrap, we generate data from the density $f_{\hat{\theta}}$. Here we plug in the estimate $\hat{\theta}$ from the original sample.

Nonparametric bootstrap

In step 1 in bootstrap, we draw samples from the empirical distribution of $\mathcal{X} = \{X_1, \ldots, X_n\}$, which is a discrete probability distribution that put mass $1/n$ on each of the data point.

Examples: see lecture notes and Lab 11.