Data Analytics – Lab Instruction 10

Importance Sampling

There are several applications where we want to estimate

$$\theta = E[f(x)] = \int f(x)p(x)dx$$

we can re-write the above equation as

$$\theta = E[f(x)] = \int f(x)\frac{p(x)}{q(x)}q(x)dx = E[f(x)w(x)]$$

where $w(x) = \frac{p(x)}{q(x)}$.

Therefore θ can be numerically estimated as

$$\theta \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)\, w(x_i)$$

Example

Assume that we want to estimate the probability $P(X > 5)$ where X has a Cauchy distribution:

$$f(x) = \frac{1}{\pi\,(1 + x^2)} \qquad -\infty < x < +\infty$$

We therefore want to find

$$\theta = \int_5^\infty f(x)dx$$

The easiest method would be to simulate values from the Cauchy distribution directly and approximate $P(X > 5)$ by the proportion of the simulated values which are bigger than 5.

The problem is that the variance of this estimator is very large as in Cauchy distribution samples are rarely exceed 5 (about 6%).

For a second, we assume that the question is not about to calculate a probability, but is to think of θ not as the area under a curve.

for large xs, $f(x)$ is approximately equal to $\frac{1}{\pi x^2}$. We therefore like to generate a probability fuction close to this simplified $f(x)$, which would be

$$q(x) = \frac{5}{x^2} \quad x > 5$$

We then can rewrite θ as

$$\theta = \int_{5}^{\infty} \frac{f(x)}{q(x)} q(x) dx$$

We can easily generate sample from $q(x)$ using inversion method from a Uniform distribution (How? Do it!).

Therefore θ can be estimated using

$$\theta \approx \frac{1}{n} \sum_{i=1}^{n} \frac{x_i^2}{5\pi \left(1 + x_i^2\right)}$$