CMS-COR-SAP. Exercise 4

Oleh Shkalikov

November 17, 2022

1 Proofs of the Weak Law of Large Numbers

a) The characteristic function is an alternative way to describe a random variable and is defined as $\varphi_x(t) = \mathbb{E}[e^{itx}]$ for a random variable X. Prove the weak law of large numbers by showing that the characteristic function of $\overline{X}_N = \frac{1}{N}\sum\limits_{i=1}^{N} X_i$ approaches $e^{it\mu}$ as $N \to \infty$.

Solution Let's write down characteristic function for mean and use properties of independent random variable to derive the result

$$\varphi_{\overline{X}_N}(t) = \varphi_{\sum\limits_{i=1}^{N} X_i}\left(\frac{t}{N}\right) = \prod_{i=1}^{N} \varphi_{X_i}\left(\frac{t}{N}\right) = \left(\varphi_{X_1}\left(\frac{t}{N}\right)\right)^N$$

Then expand this expression into series and use the fact, that $\varphi^k(0) = i^k\,\mathbb{E}\,X^k$:

$$\lim_{N\to\infty}\left(\varphi_{X_1}\left(\frac{t}{N}\right)\right)^N = \lim_{N\to\infty}\left(\mathbb{E}\,e^{\left(i\frac{t}{N}X_1\right)}\right)^N = \lim_{N\to\infty}\left(1 + \frac{i\mu t}{N} + o(t)\right)^N = e^{i\mu t}$$

b) Chebyshev's inequality states that, for a random variable X with finite mean μ and variance σ^2, no more than $\frac{1}{k^2}$ of a distribution's values can be more than k standard deviations away from the mean μ.

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Another form of this inequality is:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \tag{1}$$

Prove the weak law of large numbers using Chebyshev's inequality by showing:

$$\lim_{N\to\infty} P(|\overline{X}_N - \mu| \geq \epsilon) = 0$$

Solution Let's calculate the variance of the empirical mean of i.i.d. random values:

$$\operatorname{var}\frac{1}{N}X_i = \frac{1}{N^2}\sum_{i=1}^{N}\operatorname{var}X_1 = \frac{\operatorname{var}X_i}{N} = \frac{\sigma^2}{N}$$

Therefore putting this variance into the (1) we get:

$$P(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2} \overset{N\to\infty}{\to} 0$$

2 Monte Carlo Integration

Find $\int\limits_0^1 e^x dx$ by Monte Carlo integration using a sample size of $N = 20$ random numbers from a uniform distribution $U(0, 1)$. Compute the mean and variance of the estimator. How does it compare with the analytical mean and variance?

Solution according to law of Large Numbers the mean of estimator has to be equal(go to) to the integral value. So, let's calculate this value analytically:

$$\int\limits_0^1 e^x dx = e - 1 \approx 1.718282$$

. The analytical variance of this estimator is the following

$$\operatorname{var} \frac{1}{N} \sum_{i=1}^{N} e^{X_i} = \frac{1}{N^2} \sum_{i=1}^{N} \operatorname{var} e^{X_i} = \frac{\operatorname{var} e^{X_1}}{N}$$

In this task we are working with sampling from uniform distribution, then the variance is equals to:

$$\operatorname{var} e^{X_1} = \int\limits_0^1 x^2 e^x dx - \left(\int\limits_0^1 x e^x dx \right)^2$$

Let's calculate the first integral by calculation by parts:

$$\int\limits_0^1 x^2 e^x dx = x^2 e^x \Big|_0^1 - 2 \int\limits_0^1 x e^x = x^2 e^x \Big|_0^1 - 2x e^x \Big|_0^1 + \int\limits_0^1 e^x = 1$$

From this derivation we have that

$$\int\limits_0^1 x e^x = \frac{e-1}{2}$$

Thus, the expected variance of Monte Carlo estimator is

$$\frac{\operatorname{var} e^{X_1}}{N} = \frac{1 - \left(\frac{e-1}{2}\right)^2}{N} = \frac{1 - \left(\frac{e-1}{2}\right)^2}{20} \approx 0.013094$$

During our experiment (we have ran estimator 100 times) we got as an empirical mean 1.7328645 and 0.014310 as an empirical variance, that perfectly match theoretical results.

3 Multivariate Monte Carlo Integration

Evaluate the double integral

$$\int\limits_1^3 \int\limits_{-1}^2 (3y - 2x^2) dx dy$$

analytically and using Monte Carlo integration. Use sample sizes of $n = \{10, 50, 100, 500, 1000\}$ and plot the mean and variance vs. sample size. Compare the Monte Carlo integration with the analytical result.

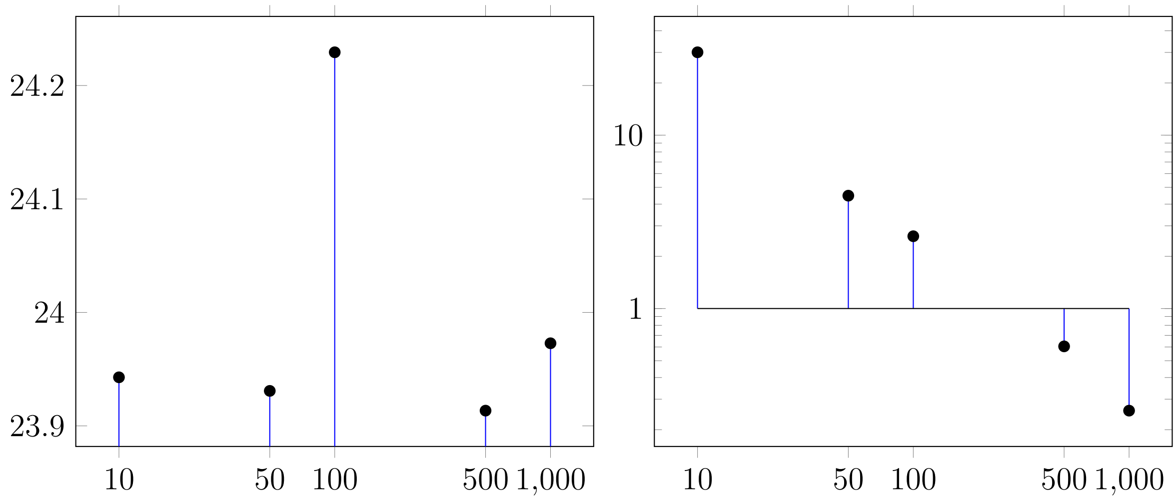
Solution First of all, let's evaluate this integral analytically

$$\int_1^3 \int_{-1}^2 (3y - 2x^2) dx dy =$$

$$\int_1^3 \left(9y - \frac{2x^3}{3} \Bigg|_{-1}^2 \right) dy = \int_1^3 \left(9y - \frac{18}{3} \right) dy =$$

$$\frac{9y^2}{2} \Bigg|_1^3 - \frac{36}{3} = \frac{72}{2} - \frac{36}{3} = 36 - 12 = 24$$

Then, we have evaluated given integral 100 times for each sample size and got the following result:



(a) Means for different N (b) Variations for different N

As we can see, the empirical mean of our estimator is close to the real value of integral and the variance rapidly decrease with growing of the sample size.

4 Importance Sampling

Solve for the expectation of the function $f(x) = 10e^{-2|x-5|}$, where $x \sim U(0, 10)$, using simple Monte Carlo integration and importance sampling from a gaussian distribution $N(5, 1)$. Use a sample size of $N = 200$ and calculate the mean and variance of the estimator. What can you observe about the variance between the simple Monte Carlo integration and importance sampling strategies?

Solution In this task we are supposed to evaluate the following integral:

$$\int_0^{10} e^{-2|x-5|} dx$$

with use of Monte Carlo method with different sampling distribution.

We have ran MC estimators 100 times for each sample distribution and got the following results:

	Uniform	Normal
Mean	1.00069	1.00305
Variance	0.01782	0.00167

Thus, normal sampling for this integral is by far better than sampling from uniform distribution.