PAMS's project

Alfonso Mateos Vicente

`alfonso.mateos-vicente@eleves.enpc.fr`

École des Ponts ParisTech

1 Simulating Random Variables

The aim of this section is to introduce the simulation of random variables from a probability distribution alredy given. We will focus on two methods: Inverse CDF method (also known as Inversion Method) and Rejection Sampling method (also known as Acceptance-Rejection method).

1.1 Inverse CDF method

Inverse CDF method is a basic method to generate pseudo-random numbers from any probability distribution given its cumulative distribution function. First of all, we can make the following supposition: Let X be a random variable with cumulative distribution function F and probability density function f. Then, the cumulative distribution function of $Y = F^{-1}(U)$ behaves like F, so the probability density function of Y is f. Knowing this, the idea is to generate a uniform random variable U and then apply the inverse of the cumulative distribution function F^{-1} to obtain a random variable with the desired distribution. The Algorithm **??** is as follows:

Algorithm 1: Inverse CDF mehtod
1. Generate a set of random numbers $U \sim \mathcal{U}(0,1)$;
2. Find the inverse of the cumulative distribution function F^{-1} ;
3. Apply the inverse to the set of random numbers $X = F^{-1}(U)$;

Let's see this with some examples. Using the exponential distribution, we know its probability density function is $f(x) = \lambda e^{-\lambda x}$. Also, we alredy know its cumulative distribution function which is $F(x) = 1 - e^{-\lambda x}$. Then, the inverse of the cumulative distribution function is:

$$F^{-1}(x) = -\frac{1}{\lambda} \ln(1-x) \tag{1}$$

So, we can generate a set of random numbers $U \sim \mathcal{U}(0,1)$ and apply the inverse to obtain a set of random numbers $X = F^{-1}(U) \sim \mathcal{E}(\lambda)$. The following Figure **??** shows the histogram of the generated random numbers and the probability density function of the exponential distribution.
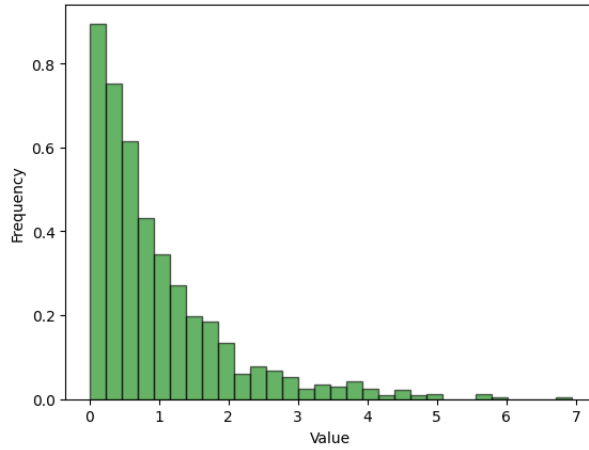
Figure 1: Histogram of Equation (**??**) with $\lambda = 1$ applied to $\mathcal{U}$.

With this picture, we can see that the histogram of the generated random numbers aligns remarkably well with the target distribution, however, we need to make more trials to be sure. We alredy know the theoreticall mean and variance of the exponential distribution, so we can compare them with the mean of the samples while the number of trials increases. The following Figure **??** and Figure **??** shows the error to the theoretical mean as the number of trials increases for the uniform and exponential distribution.
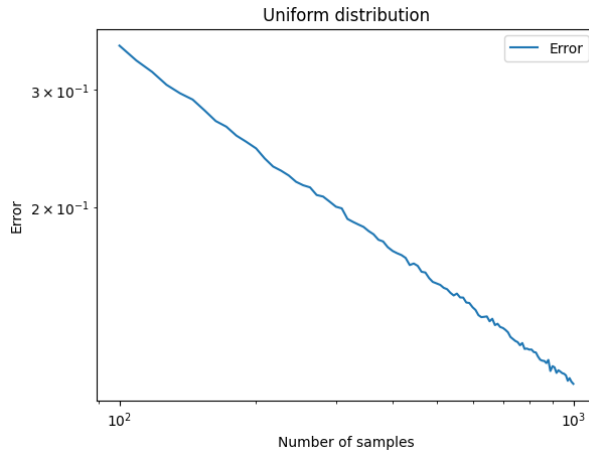

Figure 2: Error to the theoretical mean as the number of trials increases for the uniform distribution
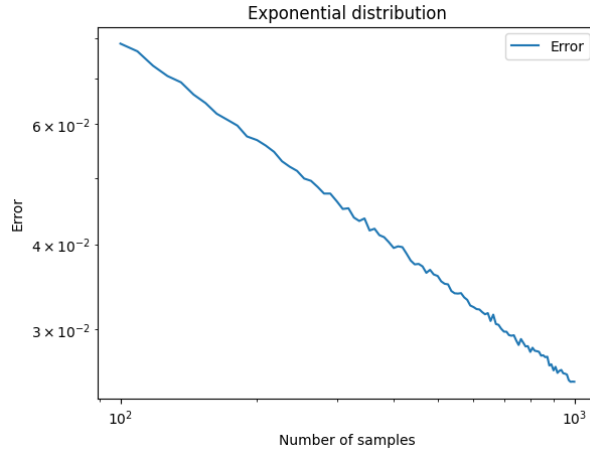
Figure 3: Error to the theoretical mean as the number of trials increases for the exponential distribution

In drawing conclusions from the observed alignment of the histograms of generated random numbers with the target distribution, there appears to be a noteworthy correlation, as evidenced by Figures **??** and **??**. These figures tentatively suggest a coherence between the theoretical and empirical means, hinting at a plausible reliability of the inverse CDF methods used. The error plots seem to indicate a convergence, providing a preliminary yet cautious optimism about the accuracy of the generated numbers in adhering to the desired distributions.

1.2 Rejection Sampling method

The Rejection Sampling method is a method to generate pseudo-random numbers for any probability distribution given its probability density function. The idea is to generate a set of random numbers from a probability distribution that is easy to sample from and then reject the numbers that are not in the desired distribution. The Algorithm **??** is as follows:

Algorithm 2: Rejection Sampling mehtod
---
 1. Generate a set of random numbers $X \sim g(x)$;
 2. Generate a set of random numbers $U \sim \mathcal{U}(0, 1)$;
 3. If $U \leq \frac{f(X)}{Mg(X)}$ then accept X, otherwise reject X ;

Being $g(x)$ the probability density function with which we alredy know how to generate random numbers; $f(x)$ the target probability density function; and M a factor we can choose manually and can be optimized.

Let's see this with one example. We want to generate a set of random numbers from the following probability density function:

$$f(x) = 0.3e^{-0.2x^2} + 0.7e^{-0.2(x-5)^2} \tag{2}$$

In the other side, we will use the normal distribution as the easy distribution to sample from. So applying the algorithm, we can generate a set of random numbers $X \sim \mathcal{N}(\mu, \sigma)$ and a set of random numbers $U \sim \mathcal{U}(0, 1)$. Then, we accept the numbers that are in the desired distribution.

First of all, it is important to note that we have to choose μ, σ and M in order to envelope the desired distribution with our normal distribution maximizing the acceptance rate. Since the bimodal distribution is almost between -5 and 10, we can choose $\mu = 2.5$ and $\sigma = 3$. Then, to maximize the acceptance rate, we have to envelope the desired distribution but with the minimun space between them as possible, so we should take M with the following formula:

$$M = \max_{x \in \mathbb{R}} \frac{f(x)}{g(x)} \tag{3}$$

Being $g(x)$ the probability density function of the normal distribution and $f(x)$ the probability density function of the desired distribution. In this case, $M \approx 2.15$.

The following Figure **??** shows the histogram of the generated random numbers and the probability density function of the desired and the easy distribution.
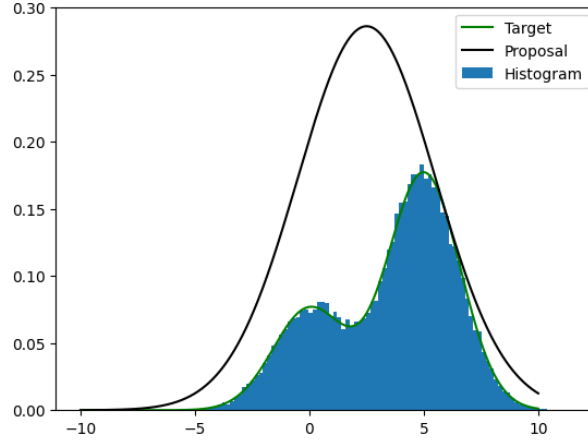

Figure 4: Histogram of the generated random numbers, Equation (**??**) and PDF of $\mathcal{N}(2.5, 3)$ scaled by $M \approx 2.15$.

With this picture, we can see that the histogram of the generated random numbers aligns remarkably well with the target distribution, however, we need to make more trials to be sure. As we did in the previous section, we can compare the mean of the samples with the theoretical mean as the number of trials increases. The following Figure **??** shows the error to the theoretical mean. As disclaimer, we know the mean of the function using the following formula: $\int_{-\infty}^{\infty} x f(x) dx$. However, we can't compute this integral analytically, so we will use the following approximation: $\frac{1}{n} \sum_{i=1}^{n} x_i$.
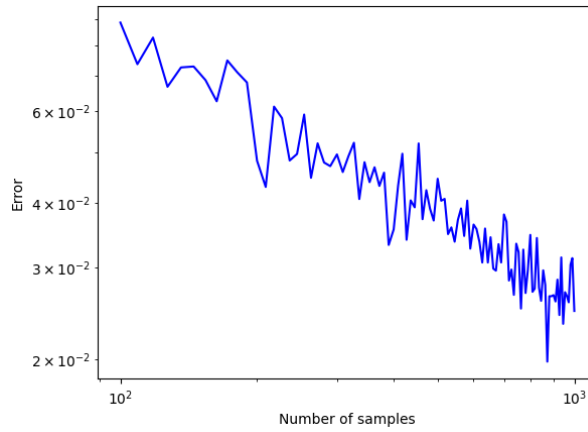

Figure 5: Error to the theoretical mean as the number of trials increases

In drawing conclusions from the observed alignment of the histograms of generated random numbers with the target distribution, there appears to be a noteworthy correlation, as evidenced by Figure **??**. This figure tentatively suggests a coherence between the theoretical and empirical means, hinting at a plausible reliability of the rejection sampling methods used. The error plot seems to indicate a convergence, providing a preliminary yet cautious optimism about the accuracy of the generated numbers in adhering to the desired distributions.

1.3 Empirical validation of the Law of Large Numbers

The Law of Large Numbers posits that, as the number of trials in an experiment increases, the average of the results obtained should converge towards the expected value of the probability distribution in question. Formally, the law can be expressed as follows:

Theorem 1.1 (Law of Large Numbers). Let $\{X_1, X_2, ...\}$ be a sequence of independent and identically distributed random variables drawn from a distribution of mean μ. And let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the average of the first n elements in the sequence. Then, as n approaches infinity, the random variables $\bar{X}_n$ converge in probability to μ:

$$\bar{X}_n \rightarrow \mu \ \ as \ n \rightarrow \infty \tag{4}$$

Given our established ability to generate random numbers reflecting specific probability distributions, we can select a distribution, generate a set of numbers corresponding to it, and then contrast the calculated averages and means of the target distribution while escalating the number of trials. To illustrate, employing the normal distribution with a mean of 0 provides the results depicted in Figure **??**.
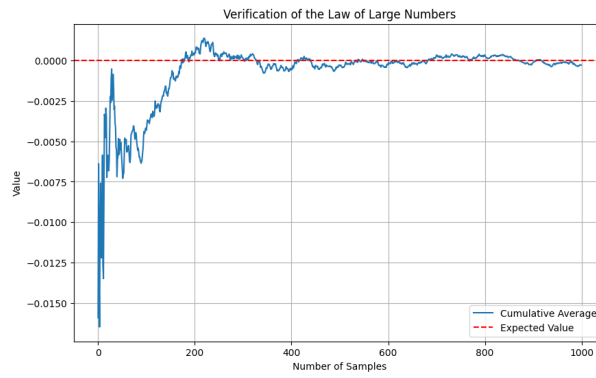

Figure 6: Comparison of the method's average with the mean of the normal distribution (0)

Refining Figure **??** to display the absolute error in relation to the mean, and plotting it on a logarithmic scale, yields Figure **??**.
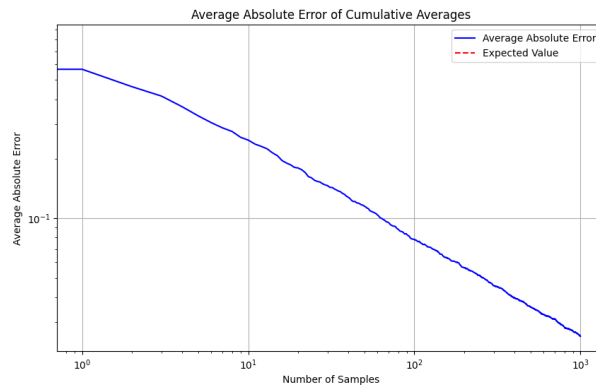

Figure 7: Logarithmic depiction of the absolute error relative to the mean

The visual representation in Figure **??** substantiates that the error indeed diminishes as the number of trials augments, validating the assertion of the Law of Large Numbers that the experimental mean approaches the theoretical mean with an increasing number of observations.

1.4 Empirical Validation of the Central Limit Theorem

The Central Limit Theorem (CLT) posits a pivotal foundational theorem in probability theory, signifying that, irrespective of the shape of the original distribution, the distribution of sample means will approx-

imate a normal distribution as the sample size burgeons. To empirically validate this theorem, we can simulate a series of random numbers from any given probability distribution and systematically calculate their mean. By perpetuating this process, we can construct a histogram of the means and scrutinize whether it converges to a normal distribution, aligning with the theorem's prediction. Formally, the theorem can be expressed as follows:

Theorem 1.2 (Lindeberg–Lévy CLT)**.** Suppose $\{X_1, X_2, ..., X_n\}$ is a sequence of independent and identically distributed random variables drawn from a distribution of mean $\mathbb{E}(X_i) = \mu$ and finite variance $\mathbb{V}(X_i) = \sigma^2$. Then, as n approaches infinity, the random variables $\sqrt{n}(\bar{X}_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \tag{5}$$

For illustrative purposes, consider the uniform distribution, which yields a histogram as delineated in Figure **??**.
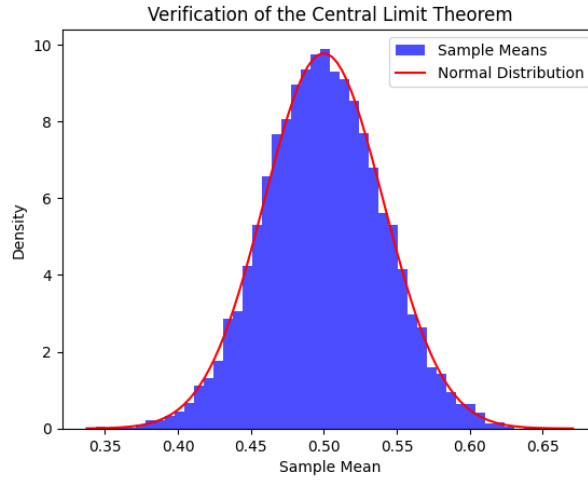

Figure 8: Histogram illustrating the convergence of the means of the uniform distribution

In congruence with the methodology espoused in previous sections, we have juxtaposed the empirical mean of the samples with the theoretical mean, progressively augmenting the number of trials. Figure **??** delineates the deviation from the theoretical mean, and Figure **??** represents the deviation from the theoretical variance, both as functions of the number of trials.
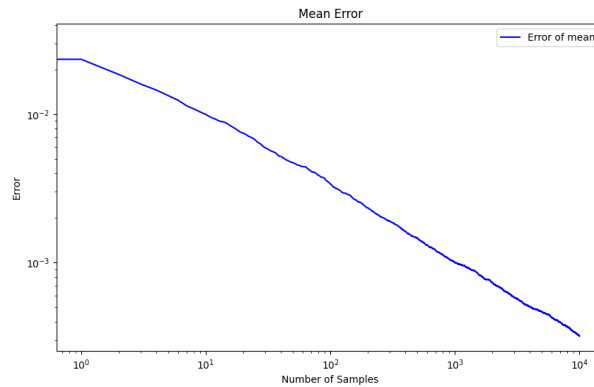

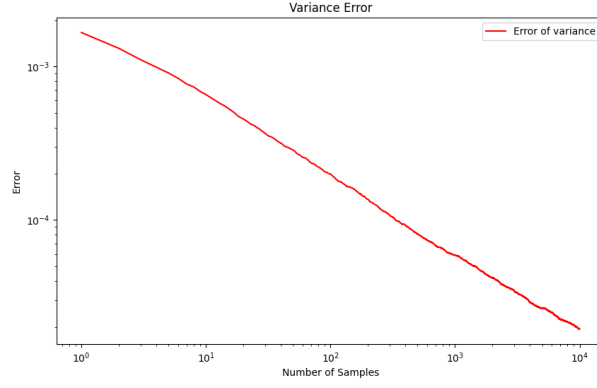Figure 9: Deviation from the theoretical mean as a function of the number of trials

Figure 10: Deviation from the theoretical variance as a function of the number of trials

A discernible insight gleaned from Figure **??** and Figure **??** is the palpable decrement in deviations from the theoretical values as the trials proliferate. This phenomenological observation substantiates the assertions of the Central Limit Theorem, illuminating the convergence of the distribution of sample means to a normal distribution as the sample size escalates, validating the theoretical underpinnings of the theorem through empirical exploration.

1.5 Variance reduction techniques

In the theory of Monte Carlo methods, variance reduction techniques are a pivotal tool to increase the precision of the estimates of the expected value of a random variable. In this section, we will focus on three techniques: Control variates, Importance sampling and Antithetic variates. Also, we will introduce an example problem in which we will apply these techniques to compare them.

First of all, let introduce the variance reduction formally. Let $\{X_1, X_2, ..., X_n\}$ be a sequence of independent and identically distributed random variables drawn from a distribution of mean $\mathbb{E}(X_i) = \mu$ and finite variance $\mathbb{V}(X_i) = \sigma^2$. Then, the expected value of a random variable is defined as follows:

$$\hat{\mathbb{E}}(X) = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{6}$$

So, we want to make zero the variance of our estimation. Since the variance is defined as follows:

$$Var(\hat{\mathbb{E}}(X)) = \frac{Var(X)}{n} \tag{7}$$

Consequently, we are presented with two avenues for optimization: increasing the value of n, or diminishing the variance of $\mathbb{E}$. Assuming that n is predetermined and unalterable, our focus would then shift to minimizing the variance of .

Since we are going to introduce and compare the three techniques, first of all we need to introduce the problem we are going to solve. The problem is that we want to estimate the following integral:

$$I = \int_0^1 x^2 \, dx \tag{8}$$

It should be noted that the selection of the integral for this demonstration was intentional; a readily solvable integral was chosen for its ease of analytical computation, allowing for a straightforward comparison with theoretical values. Nonetheless, the methods illustrated herein are equally applicable and potent for evaluating integrals that pose substantial challenges to analytical computation.

1.5.1 Preliminaries

First of all we have to compute which is the estimation we alredy can have without applying any variance reduction technique. We can compute the integral analytically. We know that the mean of the variable in a probability space is defined as follows:

$$\mathbb{E}(g(X)) = \int_b^a g(x)f(x)\,dx \tag{9}$$

Since $p(x) = \frac{1}{b-a}$ for the uniform distribution, we can apply Equation (**??**) to Equation (**??**) and we obtain:

$$I = \int_0^1 x^2\,dx = \mathbb{E}(g(X)) = \int_0^1 x^2 \frac{1}{1-0}\,dx = \frac{1}{3} \tag{10}$$

So, we only have to generate a set of random numbers $X \sim \mathcal{U}(0,1)$, apply $f(x) = x^2$ and get the mean of the sample which is the estimate of the integral. The following Figure **??** shows the error to the theoretical value as the number of trials increases.
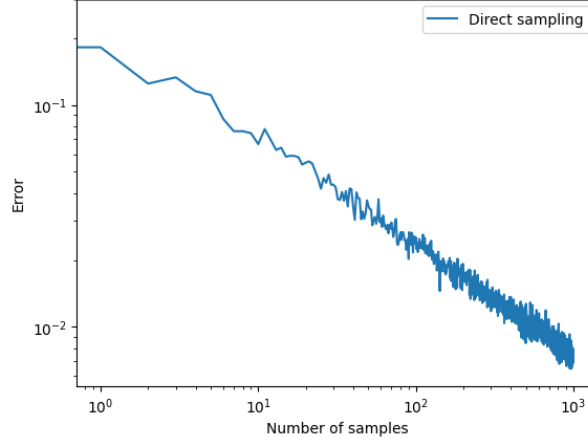

Figure 11: Error to the theoretical value as the number of trials increases using the direct method

1.5.2 Control variates

Control variates is a variance reduction technique with the following idea: Let μ the parameter we want to estimate, and assume we have a statistic Y such that $\mathbb{E}(Y) = \tau$. Then, we can estimate μ by estimating $\mathbb{E}(Y)$ as $\hat{\mathbb{E}}(Y)$ and correcting the bias with the following formula:

$$\hat{\mu} = \mu + c(\tau - \hat{\mathbb{E}}(Y)) \tag{11}$$

Being c a constant which minimize the variance of the estimation. It is computed as follows:

$$c = -\frac{Cov(X,Y)}{Var(Y)} \tag{12}$$

In our case, we can use the knowledge of the mean of the uniform distribution to correct the bias, so in each iteration we can compute the mean of the sample of random numbers of $\mathcal{U}(0,1)$, and knowing that the mean of the uniform distirbution is 0.5 we can apply Equation (**??**) to obtain the estimate of the integral. The following Figure **??** shows the error to the theoretical value as the number of trials increases.
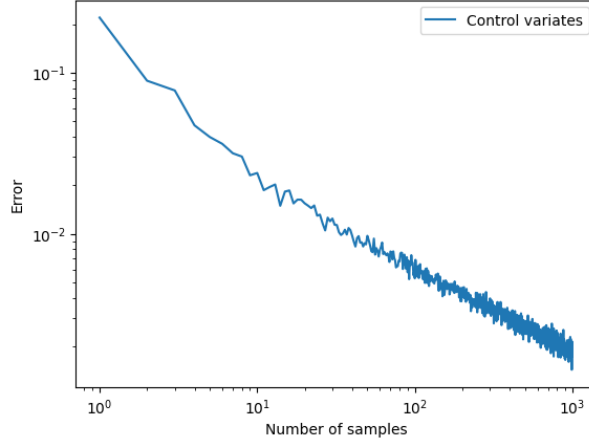
8

Figure 12: Error to the theoretical value as the number of trials increases using the control variates method

1.5.3 Importance sampling

Importance sampling is another variance reduction technique. Instead of using the knowledge of another estimator to reduce the bias of our sample, as done in the control variates method, in this case the idea is using the knowledge of the actual function we want to integrate, so instead of using a uniform distribution, we can use another distribution that is more similar to the function we want to integrate, in order to try more samples in the areas where the function is more important.

In our example, we know that the function is a parabola, so instead of using a sample which follows an uniform distribution, maybe we can use a distribution that fits better with the shape of the function. In this case, we have chosen $Beta(2.9, 1)$. In order to illustrate this, in the following Figure **??** we can see the function we want to integrate, the PDF of $Beta(2.9, 1)$ and the PDF of $\mathcal{U}(0, 1)$.
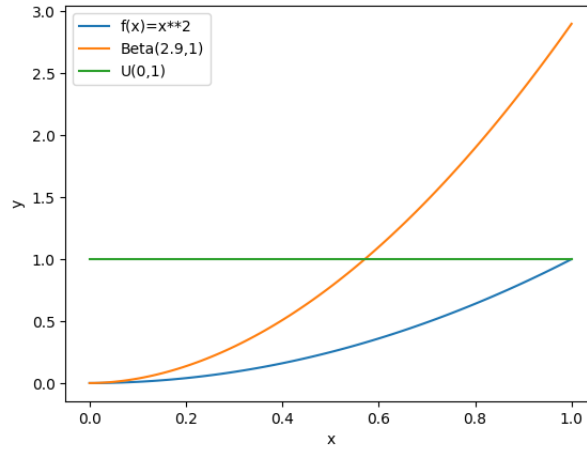

Figure 13: Function we want to integrate, PDF of $Beta(2.9, 1)$ and PDF of $\mathcal{U}(0, 1)$

As we can see, the PDF of $Beta(2.9, 1)$ fits better with the shape of the function we want to integrate, so we can expect that the error will decrease faster than the direct method and the control variates method.

So the idea is to generate a sample of random numbers $X \sim Beta(2.9, 1)$, and take $Y = \frac{f(X)}{g(X)}$ as the estimator of the integral, being $f(x) = x^2$ and $g(x)$ the PDF of $Beta(2.9, 1)$. The following Figure **??** shows the error to the theoretical value as the number of trials increases.
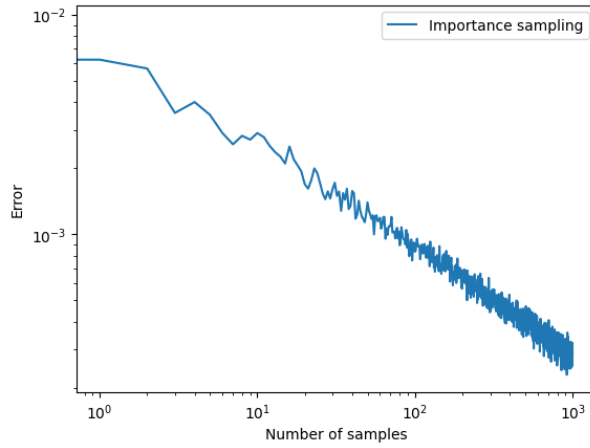
Figure 14: Error to the theoretical value as the number of trials increases using the importance sampling method

1.5.4 Stratified sampling

The idea is to divide the interval of the stratified sampling is pretty simple. We divide the interval in n subintervals, and we generate a sample of random numbers for each subinterval. Doing this, we are trying to reduce the variance of the mean of the distribution we are considering, so the sample is more representative of the distribution. The following Figure **??** shows the error to the theoretical value as the number of trials increases.
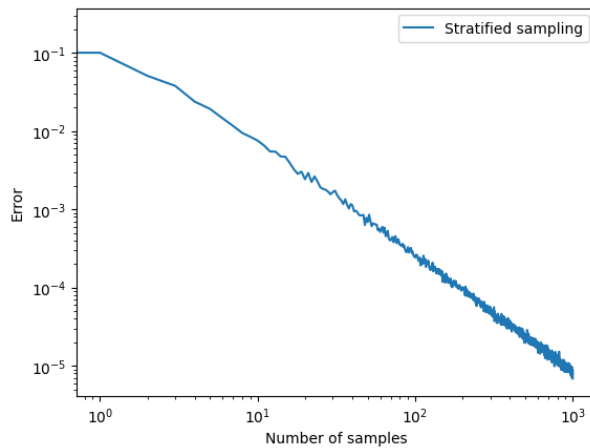

Figure 15: Error to the theoretical value as the number of trials increases using the stratified sampling method

1.5.5 Antithetic variates

The Antithetic variates method consists on taking for each random sample, its antithetic, i.e. the symmetric with respect to the mean of the distribution. The idea is that the variance of the mean of the distribution is reduced, since the mean of the antithetic is the same as the mean of the distribution. In our case, we have generated a sample $\mathcal{U} \sim \mathcal{U}(0, 0.5)$ and its antithetic $mathcalU' = \{1 - x \; : \; x \in mathcalU\}$. Taking $X = \mathcal{U} \bigcup \mathcal{U}'$, we can apply Equation (**??**) to $f(X)$ to obtain the estimate of the integral. The following Figure **??** shows the error to the theoretical value as the number of trials increases.
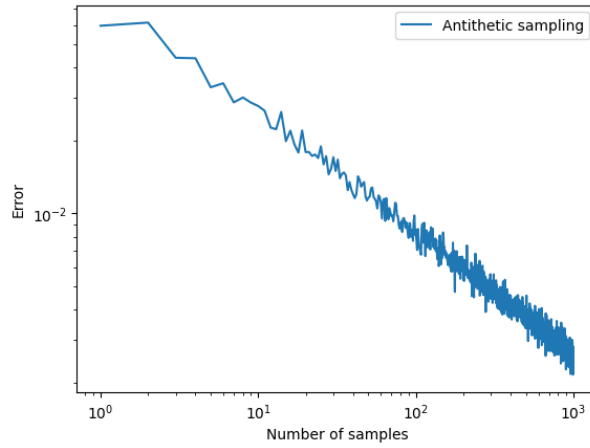
10

Figure 16: Error to the theoretical value as the number of trials increases using the antithetic variates method

1.5.6 Comparative Review

After unraveled the intricacies of each method and confirming the convergence of each approach described in this study, we now turn to a more focused comparison of these techniques. The goal is simple: find out which method works best for our specific problem. See Figure **??** for a visual representation of the error in relation to the theoretical value as we increase the number of trials for each method.
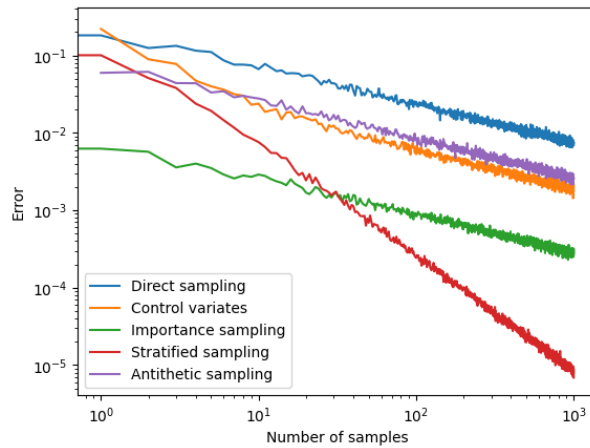

Figure 17: Error variation with increased trials for each method.

Looking at the data, it's clear that stratified sampling takes the lead with its quicker convergence compared to the other methods. However, claiming it as the undisputed champion would be premature, especially when we've explored just one problem. To firm up this initial finding, we need to dig deeper and explore a variety of problems.

Next, we'll broaden our investigation to a more complex problem, applying the methods we've discussed to estimate the following integrals:

$$Parabola = \int_0^1 x^2 \, dx \tag{13}$$

$$Gaussian = \int_0^1 e^{-x^2} \, dx \tag{14}$$

$$Sine = \int_0^1 \sin(x) \, dx \tag{15}$$

11

$$Polynomial = \int_0^1 x^3 - 2x^2 + x \, dx \qquad (16)$$

$$Exponential = \int_0^1 e^x \, dx \qquad (17)$$

Using these examples, we can now draw comparisons between the methods in a broader context. Figure **??** illustrates the error of each method applied to each function.
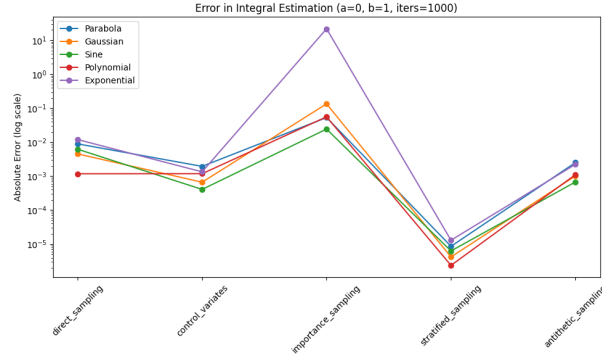

Figure 18: Method-wise error in each function.

Notably, for the importance sampling method, we opted for a normal distribution with the mean and variance of each function to maintain fairness in comparison, avoiding manual distribution selection for every function.

Reviewing the data, stratified sampling again stands out for its rapid convergence. However, the importance sampling lags, a likely outcome of applying a general normal distribution for all functions, each requiring a unique distribution. But to outrightly conclude that stratified sampling is the go-to method would be a rush to judgment, given our exploration is based on limited scenarios. Thorough exploration involving varied problems is essential to cement our initial conclusions while maintaining the professional and scientific rigor of our exploration.