

Report HW2

Simone De Carli Damiano Salvaterra
simone.decarli@studenti.unitn.it damiano.salvaterra@studenti.unitn.it

Exercise 1

The simulations were performed by drawing samples from a Poisson process in a fixed time $[0, T]$ with $T = 100$ and fixed number of arrivals N , then $\lambda = \frac{N}{T}$ followed.

Part 1

The first part consisted in drawing N uniformly distributed arrivals in the interval $[0, T]$, computing the inter-arrival verifying their exponential distribution by:

- computing the empirical mean and variance of the inter-arrival times and comparing them with the theoretical values
- aggregating all the inter-arrival from each process realization and plotting the histogram of the inter-arrival times with the theoretical PDF of the exponential distribution
- plotting the Q-Q plot of the inter-arrival times against the theoretical exponential distribution

computing and comparing empirical and theoretical metrics like mean and variance, fitting the data to a distribution and plotting the PDF and by plotting the Q-Q plot.

TODO: Add plots, results and comments

Part 2

In the second method we generated N exponentially distributed inter-arrival times. This required to check that the sum of N exponentially

Exercise 2

Rejection Sampling

To apply the rejection sampling, we need to find a distribution such that we can bound the probability $f(x)$ such that

$$\frac{f(x)}{g(x)} \leq c$$

A valid proposal distribution can be

$$g(x) = \begin{cases} k x^2 & \text{if } -3 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

where k is such that $\int_{-3}^3 g(x) dx = 1$. Such k is given by

$$\int_{-3}^3 g(x) dx = \int_{-3}^3 kx^2 dx = 18k \stackrel{!}{=} 1 \Rightarrow k = \frac{1}{18}$$

The proposal distribution is then $g(x) = \frac{x^2}{18}$ and the bound is

$$\frac{\frac{1}{A}x^2 \sin^2(\pi x)}{\frac{x^2}{18}} = \frac{18}{A} \sin^2(\pi x) \leq c \Rightarrow c = \frac{18}{A} = 2.03435386. \quad (1)$$

To draw a sample from the distribution $g(x)$ we can apply the CDF-Inversion method, so we derive the inverse of the CDF $G(x)$:

$$G(x) = \int_{-3}^x g(t) dt = \int_{-3}^x \frac{t^2}{18} dt = \frac{1}{2} + \frac{x^3}{54} \Rightarrow G^{-1}(u) = 3\sqrt[3]{2(u - \frac{1}{2})} \quad \text{for } u \in [0, 1].$$

We can now choose 2 ways of applying the rejection sampling, the first with the knowledge of A , which allow us to have an higher acceptance rate, or without knowledge of A , which allow us to know the target distribution up to a constant but it does not let us to bound it so tightly, having a drop in the acceptance rate. We report both methods in the following.

Algorithm with the full knowledge of $f(x)$

In this case we assume to know the value of A , so that we know also the value of c from 1, so we can bound $f(x)$ with the function $cg(x) = \frac{18}{A} \frac{x^2}{18} = \frac{x^2}{A}$ which is basically the envelope of $f(x)$. Thus, the algorithm is the one reported in Algorithm 1.

Algorithm 1 Rejection sampling with full knowledge of $f(x)$

- 1: Draw $u_1 \sim \mathcal{U}[0, 1]$
 - 2: Compute $x = G^{-1}(u_1)$
 - 3: Draw $u_2 \sim \mathcal{U}[0, c \cdot g(x)]$
 - 4: **if** $u_2 < f(x)$ **then**
 - 5: Accept x
 - 6: **else**
 - 7: Go back to Step 1
 - 8: **end if**
-

The distribution resulting from Algorithm 1 is shown in Figure 1a. The acceptance rate of this algorithm in 10^8 iterations is $\frac{10^8}{49155544} \approx 0.49$. The distribution of the drawn samples is shown against the theoretical one in Figure 1a, while the accepted samples and the rejected samples (from the proposal) are shown in 2a.

Algorithm with knowledge of $f(x)$ up to a constant

Given that $f(x) = \frac{1}{A}x^2 \sin^2(\pi x) = \frac{1}{A}f^n(x)$, if we do not know the value of A , we can still bound the non-normalized distribution $f^n(x)$:

$$f^n(x) = x^2 \sin^2(\pi x) \leq x^2 \leq 9 = M \quad \text{in } -3 \leq x \leq 3 \quad (2)$$

We can use this upper bound to apply rejection sampling as in Algorithm 2.

Algorithm 2 Rejection sampling with knowledge of $f(x)$ up to a constant

```
1: Draw  $x \sim \mathcal{U}[-3, 3]$ 
2: Draw  $u \sim \mathcal{U}[0, M]$ 
3: if  $u \leq f^n(x)$  then
4:   Accept  $x$ 
5: else
6:   Go back to Step 1
7: end if
```

We see that in this algorithm we do not employ A at all, and the proposal distribution is a uniform distribution. We build a “bounding box” around the scaled distribution $f^n(x)$ and accept only those points that fall under the plot of $f^n(x)$. This approach works, but, clearly, being less precise in the bounding, the acceptance rate drops: the acceptance rate on 10^8 iterations is $\frac{10^8}{16381697} \approx 0.16$. The distribution of the drawn samples is shown against the theoretical one in Figure 1b., while the accepted samples and the rejected samples (from the proposal) are shown in Figure 2b.

Confidence intervals

For the computation of the confidence intervals, since the data are i.i.d., we use the order statistics and the binomial distribution (approximated with a normal, since $n = 200$ is large enough) for the CIs of the median and the 0.9-quantile. For a quantile q , the formula for CI $[X_{(j)}, X_{(k)}]$ is (using the normal approximation of the binomial):

$$j \approx \lfloor nq - 1.96\sqrt{nq(q - q)} \rfloor, \quad k \approx \lfloor nq + 1.96\sqrt{nq(q - q)} \rfloor + 1.$$

Instead, for the mean, we exploit the central limit Theorem (since we have enough data points and the distribution is symmetric) and obtain the confidence interval:

$$\hat{\mu}_n \pm \eta \frac{s_n}{\sqrt{n}}$$

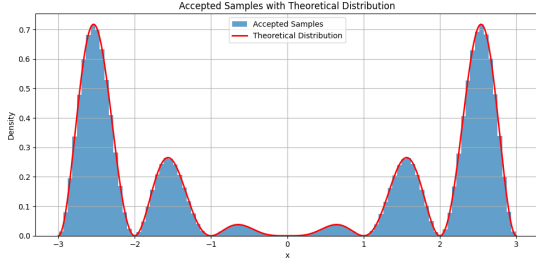
For the bootstrap procedure, the only assumption is to have i.i.d. data, so we can also apply it.

The plots of the median and the 0.9-quantile computed with the two different approaches are shown in Figure 3.

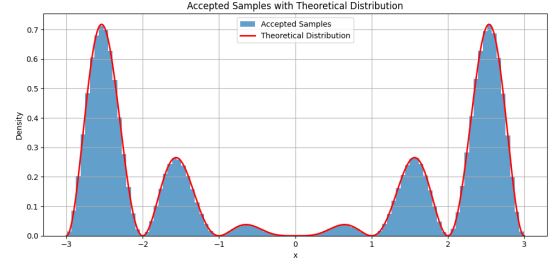
The plots for the mean are shown in Figure 4.

Statistical significance of the mean’s confidence intervals

We subdivided 20000 variates into 100 disjoint sets and for each of them we computed the 95% confidence intervals, both with the Gaussian approximation approach and the Bootstrap approach. As expected, the number of confidence intervals containing the true mean is 95 in the case of the Gaussian approximation and 94 in the case of the bootstrap CIs (we can see that the bootstrap method slightly underestimates the CI, so we find less accurate CIs). Figure 5 shows the distribution of the sample mean computed in the 100 sets, showing that the Central Limit Theorem holds over this statistic.

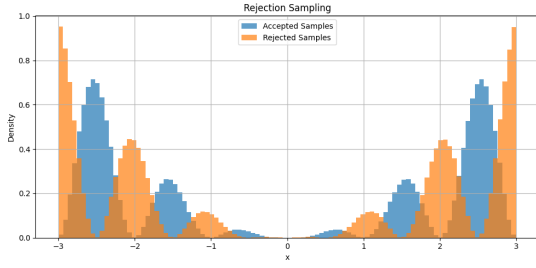


(a) Rejection sampling with Algorithm 1.

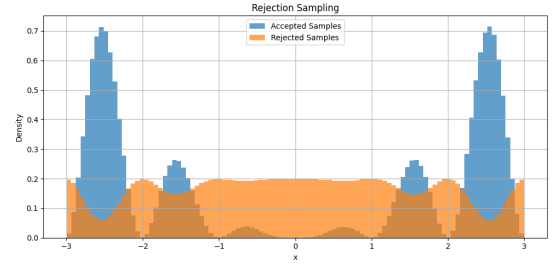


(b) Rejection sampling with Algorithm 2.

Figure 1: Rejection sampling of $f(x)$

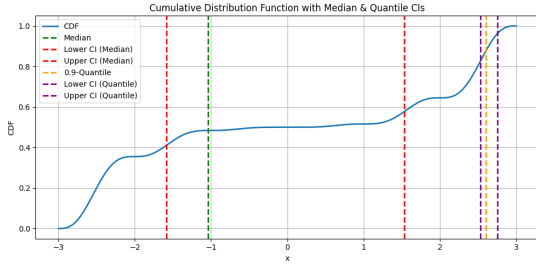


(a) Rejected/accepted samples with Algorithm 1.

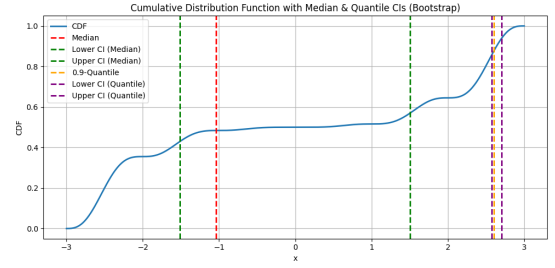


(b) Rejected/accepted samples with Algorithm 2.

Figure 2: Accepted/rejected samples distributions.

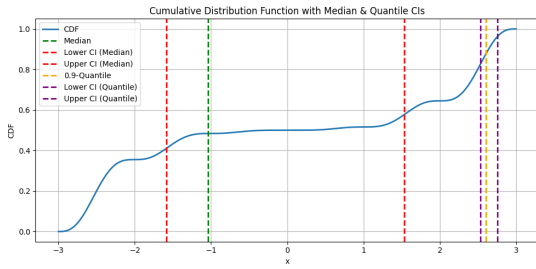


(a) Median and 0.9-Quantile CIs computed with the Binomial formula.

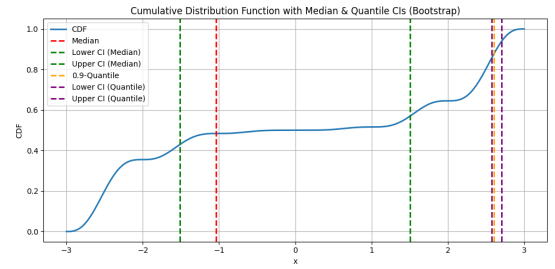


(b) Median and 0.9-Quantile CIs computed with the Bootstrap heuristic.

Figure 3: CDF with median and 0.9-quantile CIs.



(a) Mean CI computed with the Gaussian approximation.



(b) Mean CI computed with the Bootstrap heuristic.

Figure 4: PDF and mean CI.

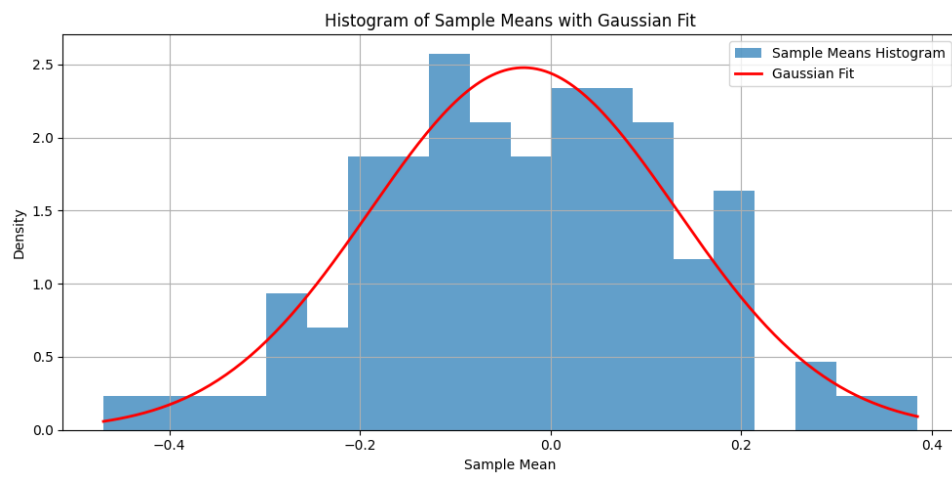


Figure 5: Empirical PDF of the sample mean.