M1AM – Applied Probability and Statistics

Practical Session 2

Foreworld : The objective of this work is the numerical study of different parametric estimators in the exponential model. In the first exercise, we illustrate the asymptotic theorems of the Law of Large Numbers and the central limit theorem for the Bernoulli, Gaussian, Uniform and Exponential models. The second exercise involves the study of two estimators of the parameter of the exponential model at fixed sample size and in the asymptotic case. The third exercise looks at confidence interval estimation in the exponential model for any sample size n and then in the asymptotic case.

Exercice 1

1. **Bernoulli Model**
   Note that the random generator, cdf, density and quantile function of the Binomial variable are provided by R with `rbinom`, `pbinom`, `dbinom` and `qbinom`. First set $n = 100$, $N = 10\ 000$ and $p = 0.2$.

   (a) Simulate the realization of a sample of size n of a Bernoulli random variable with parameter p.

   (b) Calculate the empirical means of the first k values in the sample for k ranging from 1 to n.

   (c) Plot $(k, \bar{x}_k)_{k=1,\ldots,n}$ and add the horizontal line passing through p (use functions `plot(..., type="l")` and `abline()`).

   (d) Run several times your script and observe the results. Comment.

   (e) Simulate N realizations of a n-sample and arrange them in a matrix with N rows and n columns.

   (f) Calculate the N realisations of $\bar{X}_n$

   (g) Represent the distribution obtained (with function `hist()`). Add on this graphic the density function of the normal distribution which is known to approximate the data (think about the central limit theorem). Comment on the accuracy of the approximation. What do you observe if you vary the value of n? of N? of p?

   (h) Draw a graph to compare the empirical cumulative distribution and the cumulative distribution function of the normal distribution $\mathcal{N}(0.2, 0.2 * 0.8/n)$. Conclude.

   (i) Another way of diagnosing the normality of a sample is the quantile-quantile plot (`qqplot()`). Try it on your data and conclude.

2. **Gaussian Model**

   Make the same numerical experiments as for the Bernoulli model with $\mu = p$ and $\sigma^2 = p(1-p)$.

3. **Uniform Model**

   Make the same numerical experiments as for the Bernoulli model with a uniform model over $[0, \theta]$ with θ such that $E(X) = p$.

4. **Exponential Model**

   Make the same numerical experiments as for the Bernoulli model with an exponential model of parameter θ such that $E(X) = p$.

5. **Rate of convergence**

   (a) Compare the a.s. convergences of $\bar{X}_n$ for the four different parametric models by plotting all the convergence curves, with different colors, on the same graphic. Comment.

   (b) Compute for the uniforme model $(0.5 \cdot max_{i=1,\ldots,k}(x_i))_k$ and add the convergence curve for this estimator to the previous graphic. Is the convergence of this sequence faster? Comment.

Exercice 2

We consider here the exponential model with parameter θ and with sample of size n. We propose to study numerically and compare the two following estimators of θ:

$$\hat{\theta}_{1,n} = \frac{n}{\sum_{i=1}^{n} X_i} \quad \text{and} \quad \hat{\theta}_{2,n} = \sqrt{\frac{2n}{\sum_{i=1}^{n} X_i^2}}$$

1. Set $N = 1000$, $n = 10$ and $\theta = 0.25$. Simulate N realizations of a n-sample with probability distribution $\mathcal{E}(\theta)$ and arrange them in a matrix with N rows and n columns. Compute the N realisations of both estimators.

2. Evaluate the bias and variance of each estimator. Deduce the empirical quadratic risk of each estimator. Which of the two has the smallest risk? What about the difference of these risks when n is much larger than 10 (for example $n \geq 100$)?

3. What are the theoritical limit distributions (for large n) of these two estimators (see Exercice 4.6)? Comment.

4. Show graphically that the observed distribution of each estimator is correctly approximated by the limit law established above.

Exercice 3

We want to study now the following confidence interval for θ in the exponential model for a sample of any size n possibly small (for ex. $n = 10$). For $0 < p < 1$, $0 < \alpha < 1$ and $z_q^{(\nu)}$ the quantile of order q $(0 < q < 1)$ of the $\mathcal{X}_\nu^2$ set :

$$I_n(\alpha, p) = \left[\frac{z_{p\alpha}^{(2n)}}{2n\bar{X}_n}, \frac{z_{1-(1-p)\alpha}^{(2n)}}{2n\bar{X}_n} \right].$$

1. Prove that $I_n(p, \alpha)$ is a confidence interval of level $1 - \alpha$ for the unknown parameter θ and any n.

2. Verify it numerically with simulated data.

3. For $n = 10$ find numerically the best choice of p to obtain the smaller width of interval for $\alpha = 10\%$.

4. Would the choice of p be the same for $n > 100$?

5. Calculate an asymptotic confidence interval of level $1 - \alpha$ with symmetric risks for the parameter θ using the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{1,n} - \theta)$ stated in Exercice 2. Compare it with the previous one.