List 2 Report

Statistics and Linear Models

Dawid Dieu

5 December 2022

Goal

The goal of this report is to generate data from different distributions and calculate various estimators (e.g. the maximum likelihood estimators). Using those estimators, we calculate probabilities or some variables and compare them against actual true values. We also use a Shapiro–Wilk test to detect if a random variable comes from the normal distribution.

Task 1

In the first task, we are asked to generate $n = 50$ observations from a binomial $b(5,p)$ distribution. Then, we calculate a value of the maximum likelihood estimator of the quantity $P(X \geq 3)$, where $X \sim b(5,p)$. Below we present results for 10 000 experiments. We estimate P using $\hat{p}$. When calculating MSE and bias, we compare it with true P.

p in $b(5,p)$	$\hat{p}$	$P(X \geq 3)$	Variance	MSE	Bias
0.1	0.099915	0.009319	0.000024	0.000025	0.000759
0.3	0.299927	0.165010	0.001420	0.001424	0.001930
0.5	0.500027	0.500050	0.003437	0.003437	0.000050
0.7	0.700181	0.835011	0.001505	0.001509	-0.001909
0.9	0.900129	0.990712	0.000024	0.000024	-0.000728

The first observation would be that $\hat{p}$ is very close to the true p. Therefore when we calculate P it is also close to the true value. Low bias and MSE confirms this statement. The variance is also low because we do quite a lot of experiments (10000), which means that there shouldn't be many outliers in the calculation of P.

Task 2

In the second task, we are asked to generate $n = 50$ observations from a Poisson distribution with the parameter λ. Then, calculate a value of the maximum likelihood estimator of the quantity $P(X = x), x = 0,1,\ldots,10$, where X

~ $\pi(\lambda)$. Below we present results for 10 000 experiments. We estimate P using $\hat{\lambda}$. When calculating MSE and bias, we compare it with true P. We present results only for $\lambda = 5$.

x	$\hat{\lambda}$	$P(X=x)$	Variance	MSE	Bias
0	4.998720	0.007091	0.000005	0.000007	-0.001475
1	5.004460	0.034608	0.000079	0.000081	0.001454
2	5.000988	0.085346	0.000258	0.000486	-0.015111
3	5.002712	0.140484	0.000305	0.000407	0.010110
4	4.997670	0.174521	0.000115	0.000133	0.004161
5	5.004112	0.173769	0.000006	0.000007	0.001243
6	4.997060	0.144688	0.000088	0.000090	0.001535
7	5.001226	0.103875	0.000169	0.000208	0.006263
8	4.996644	0.065272	0.000147	0.000561	-0.020370
9	5.004034	0.036873	0.000083	0.000115	0.005659
10	4.994090	0.018555	0.000033	0.000033	-0.000685

In every experiment, the $\hat{\lambda}$ is quite close to the true λ, which is very good. We do 10 000 experiments for each case of x, so the variance is low. Significantly small MSE and bias mean that our parameter estimation λ is proper.

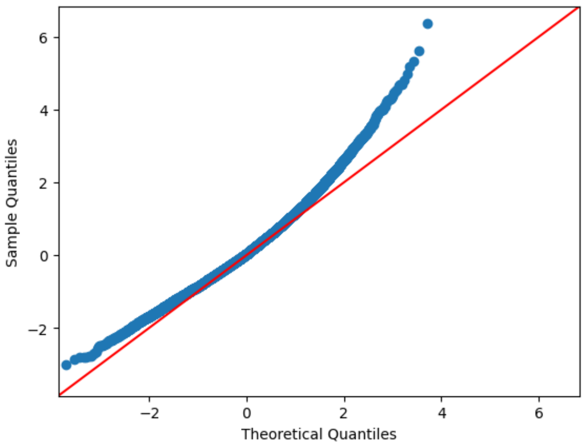
For other lambdas, the estimation is also close to the real value. Quite often, the MSE and bias values are smaller than in the example shown in the table above.

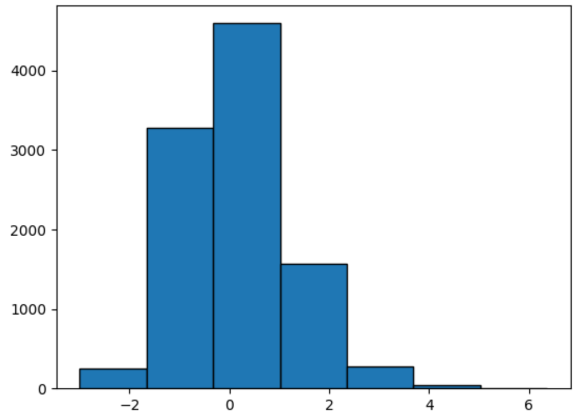
Task 3

In the third task, we are asked to generate $n = 50$ observations from a beta distribution $beta(\theta,1)$. We calculate a value of the maximum likelihood estimator $I(\hat{\theta})$ of the Fisher information for the parameter θ. We calculate $I(\hat{\theta})$ using $\hat{\theta}$. Then we generate $n = 50$ observations again, and we estimate a new $\hat{\theta}_1$. Now using the Fisher information estimated independently at the beginning, we calculate a new variable $Y = \sqrt{nI(\hat{\theta})}(\hat{\theta}_1 - \theta)$. Below we present results for 10 000 experiments with Q-Q plots and histograms. When plotting the histograms, the number of bins is equal to $\sqrt(n)$.

θ in $beta(\theta,1)$	$\hat{\theta}$	$I(\hat{\theta})$	Y
0.5	0.509930	4.079	0.147
1	1.021760	1.019	0.139
2	2.039365	0.255	0.150
5	5.113494	0.041	0.141

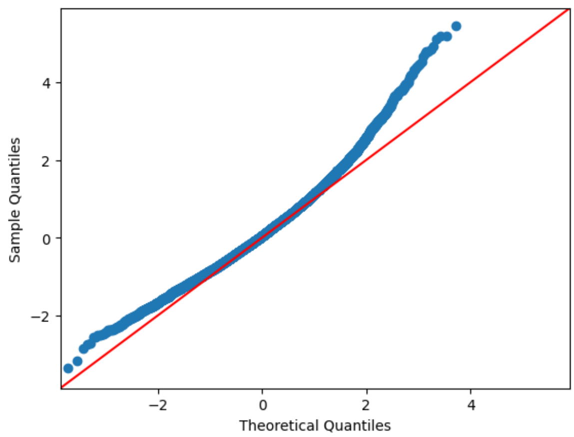
a) $\theta = 0.5$

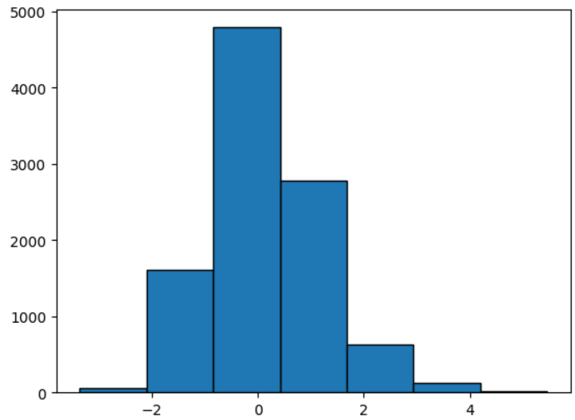




Test shapiro: ShapiroResult(statistic=0.9772, pvalue=6.380e-27)

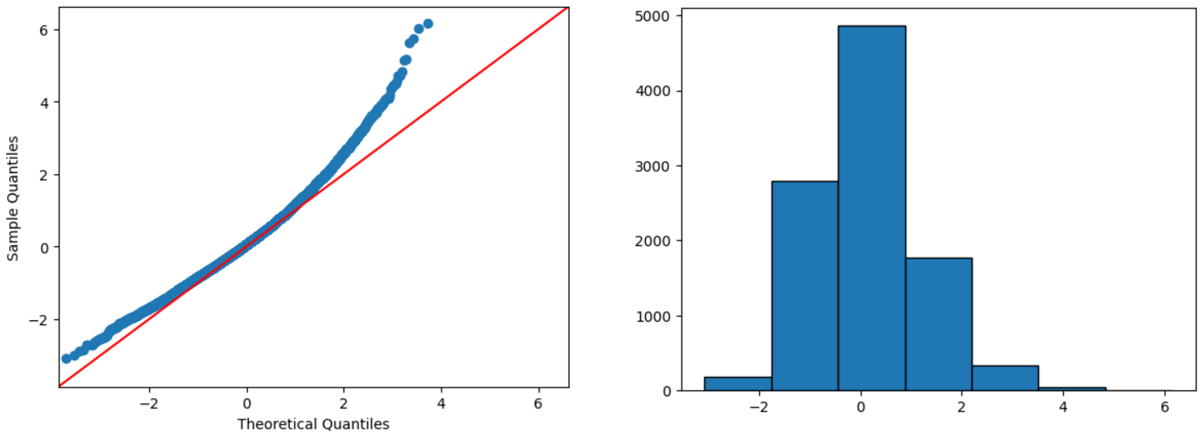
b) $\theta = 1$





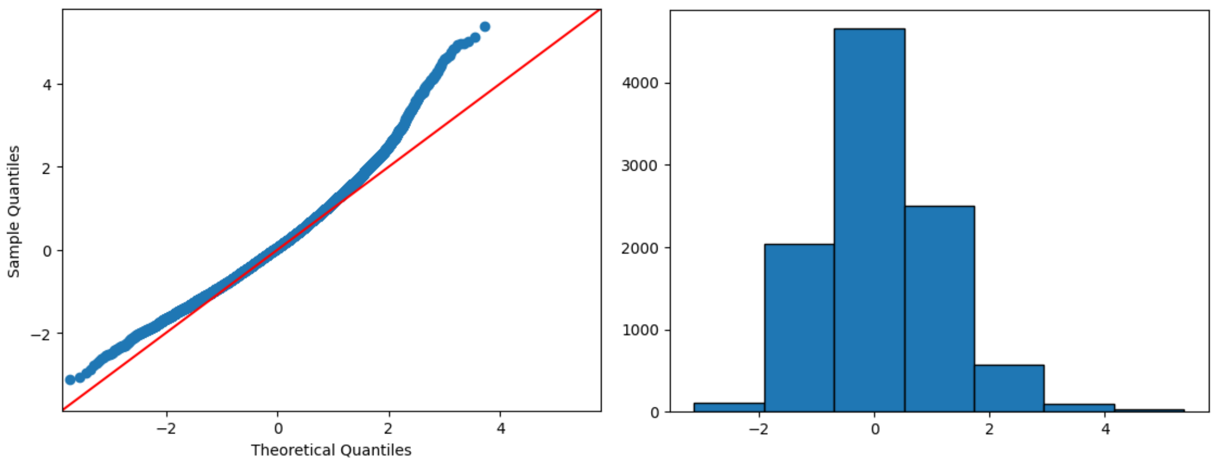
Test shapiro: ShapiroResult(statistic=0.9779, pvalue=3.177e-27)

c) $\theta = 2$



Test shapiro: ShapiroResult(statistic=0.9845, pvalue=8.043e-23)

d) $\theta = 5$



Test shapiro: ShapiroResult(statistic=0.9777, pvalue=5.783e-27)

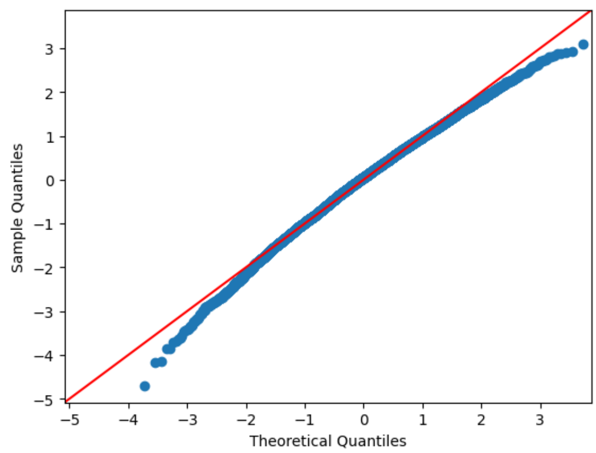
The distribution does not look normal when we look at the Q-Q plot. The Shapiro test result confirms it. In all cases, the p-value is small, and we reject the null hypothesis that the data is drawn from the normal distribution.

But when we increase the number of observations we generate, we can see that the p-value can increase significantly. In the table below, one can see that for

N equal to 1000, 5000, and 10000, we would not reject the null hypothesis as the p-value is more significant than 0.05.

N	p-value
50	1.93e-27
500	0.000433
1000	0.07801
5000	0.2752
10000	0.06818

Note: If for calculating, Y we use the same $\hat{\theta}$ that we use for calculating the Fisher information $I(\hat{\theta})$, i.e. $Y = \sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta)$, the Q-Q plot looks different. It goes downwards at the ends.



Task 4

In the fourth task, we are asked to generate $n = 50$ observations from a Laplace distribution with the parameters θ and σ. We calculate the values of four estimators.

(i) $\hat{\theta}_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$

(ii) $\hat{\theta}_2 = Me\{X_1, \ldots, X_n\}$,

(iii) $\hat{\theta}_3 = \sum_{i=1}^{n} w_i X_i, \sum_{i=1}^{n} w_i = 1, 0 \leq w_i \leq 1, i = 1,\ldots,n$

with an arbitrary weights selection,

(iv) $\hat{\theta}_4 = \sum_{i=1}^{n} w_i X_{i:n}$, where $X_{1:n} \leq \ldots \leq X_{n:n}$ are the order statistics from the sample $X_1, \ldots, X_n$, $w_i = \phi(\Phi^{-1}(\frac{i-1}{n})) - \phi(\Phi^{-1}(\frac{i}{n}))$, while ϕ is the density and Φ is the cumulative distribution function of the standard normal $N(0,1)$ distribution.

Below we present results for 10 000 experiments.

a) $n = 50, \theta = 1, \sigma = 1$

	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
variance	0.040131	0.024236	0.041071	0.040192
MSE	0.040132	0.024236	0.041114	0.152828
Bias	-0.000927	-0.000155	-0.006553	0.335613

b) $n = 50, \theta = 4, \sigma = 1$

	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
variance	0.039816	0.023968	0.040827	0.039383
MSE	0.039820	0.023969	0.040882	7.140216
Bias	-0.001968	-0.001238	-0.007372	-2.664739

c) $n = 50, \theta = 1, \sigma = 2$

	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
variance	0.160300	0.095915	0.164238	0.158630
MSE	0.160303	0.095916	0.164316	2.951103
Bias	0.001696	-0.000959	-0.008833	1.671069

The optimal estimator of the Laplace distribution is the sample median θ_2. As we can see in our results, this estimator has the lowest MSE in all cases. In c) the standard deviation is 2, and we draw only 50 samples; that's why we can observe the most significant variance in the estimators. The first estimator θ_1 - the sample mean, does not perform that poorly, as the distribution is symmetrical.

Task 5

In the last task, we will discuss the experiment's outcomes if we use different values of n, i.e. $n = 20$ and $n = 100$. The results are not surprising. If we decrease the number of samples, let's say to $n = 20$ the variance grows. Therefore we are more likely to see a more considerable bias as our estimator is the mean of the "unfortunate" sample. When we test for the normality of the population that consists of only 20 samples, the chances of being normal are even lower. If we increase the sample size, the opposite holds. The estimators are more accurate. The variance drops. The statistic tests for normality have a higher chance of not rejecting the null hypothesis that the sample of observations comes from the normal distribution.