

# On Statistical Hypothesis Testing via Simulation Method

*B. Dimitrov, D. Green, Jr., V. Rykov, and P. Stanchev*

Department of Science & Mathematics, Kettering University, Flint, Michigan, USA

1700 West Third Avenue, Flint, Michigan 48504-4898, USA

*e-mail {bdimitro, dgreen, vrykov, pstanche}@Kettering.edu*

A procedure for calculating critical level and power of likelihood ratio test, based on a Monte-Carlo simulation method is proposed. General principles of software building for its realization are given. Some examples of its application are shown.

## 1 Introduction

In this paper we show how the present day fast computer could solve non-standard old statistical problems. In most cases statisticians work with approximations of test statistics distributions, and then use respective statistical tables. When approximations do not work the problem is usually tabled. We propose such simulation approach which we do believe could be helpful in many case.

The problem of statistical hypothesis testing is very important for many applications. In the notable but rare case, it is possible to find some simple test statistic having a standard distribution. However, in the general case the statistics based on the *Likelihood Ratio Test* (LRT) does not usually have one of the known standard distributions. The problem could be overcome with the help of an appropriate simulation method. This method was first used in [3] for a specific case of almost lack of memory (ALM) distributions. In this paper we propose a general approach for using the method, describe its general principles and algorithms, show how to build up an appropriate software, and illustrate with examples its application.

## 2 LRT and the Simulation Approach

It is well known according to Neyman-Pearson theory [8], that the *most powerful test* for testing a null hypothesis  $H_0 : f(x) = f_0(x)$  versus an alternative  $H_1 : f(x) = f_1(x)$  is the LRT. For this test, the *critical region*  $W$  for a sample  $x_1, \dots, x_n$  of size  $n$  has the form

$$W = \left\{ (x_1, \dots, x_n) : w(x_1, \dots, x_n) = \frac{f_1(x_1, \dots, x_n)}{f_0(x_1, \dots, x_n)} > t \right\},$$

where  $f_0(x_1, \dots, x_n)$  and  $f_1(x_1, \dots, x_n)$  are joint probability densities of the distributions of observations (the *likelihood functions*) under hypotheses  $H_0$  and  $H_1$  with probability density functions (p.d.f.)  $f_0(\cdot)$  and  $f_1(\cdot)$  respectively. The notation

$$w(x_1, \dots, x_n) = \frac{f_1(x_1, \dots, x_n)}{f_0(x_1, \dots, x_n)}$$

is used for *test's statistic*. For independent observations this statistic can be represented in the form

$$w(x_1, \dots, x_n) = \frac{f_1(x_1, \dots, x_n)}{f_0(x_1, \dots, x_n)} = \prod_{1 \leq i \leq n} \frac{f_1(x_i)}{f_0(x_i)}.$$

Considering the observations  $x_1, \dots, x_n$  as independent realizations of random variable (i.i.d. r.v.)  $X$  with p.d.f.  $f_0(\cdot)$  the *significance level* of the test is

$$\alpha = P_{H_0} \{W\} = P_{H_0} \{w(X_1, \dots, X_n) > t_\alpha\}. \quad (1)$$

Here an appropriate *critical value*  $t_\alpha$  for any given significance level  $\alpha$  is the smallest solution of equation (1).

Analogously, considering the same observations  $x_1, \dots, x_n$  as independent realizations of random variable  $Y$  with p.d.f.  $f_1(\cdot)$ , *the power of the test* is

$$\pi_\alpha = P_{H_1} \{W\} = P_{H_1} \{w(Y_1, \dots, Y_n) > t_\alpha\}. \quad (2)$$

Thus, to find the critical value for a given significance level  $\alpha$  and the power of the test  $\pi(\alpha)$ , a statistician needs to know the distributions of the test statistic  $w$  under hypotheses  $H_0$  and  $H_1$ .

For parametric hypothesis testing the problem becomes more complicated because in such cases one has to be able to find a free of parameters distribution of this statistic.

To avoid calculations of these functions we propose to use the simulation method. This means that instead of searching for exact statistical distributions, we will calculate appropriate empirical distributions as their estimations. This method gives desired results due the fact (based on the *Strong Law of Large Numbers*) that the empirical distribution function of the test statistic converges with probability one to the theoretical distribution.

In the following, due to numerical reasons, instead of statistic  $w$  we will use its natural logarithm, and for simplicity we will denote this statistic with the same letter,  $w$ ,

$$w = \ln \prod_{1 \leq i \leq n} \frac{f_1(x_i)}{f_0(x_i)} = \sum_{1 \leq i \leq n} (\ln f_1(x_i) - \ln f_0(x_i)). \quad (3)$$

Due to additional statistical reasons, instead of the cumulative distribution functions (CDF) of the statistic  $w$  under hypotheses  $H_0$  and  $H_1$ , we will use their tails,

$$\overline{F}_0(t) = P_{H_0} \{w(X_1, \dots, X_n) > t\}, \quad (4)$$

and

$$\overline{F}_1(t) = P_{H_1} \{w(Y_1, \dots, Y_n) > t\}. \quad (5)$$

For large size samples,  $n \gg 1$ , it is possible to use a simpler approach based on the *Central Limit Theorem*. It is well known that this theorem provides a normal approximation of the distribution for sums of i.i.d. r.v.'s under conditions of existence of finite second moments. This would allow one to calculate and use only two moments of the test's statistic  $w$  and then to calculate the appropriate significance level and power of the test making use of the respective normal approximation.

To show how it works, let us denote by  $U$  and  $V$  the r.v.'s

$$U = \ln f_1(X) - \ln f_0(X), \quad V = \ln f_1(Y) - \ln f_0(Y),$$

where  $X$  and  $Y$  are taken from distributions with densities  $f_0(\cdot)$  and  $f_1(\cdot)$  respectively, corresponding to hypotheses  $H_0$  and  $H_1$ . Denote by  $\mu_U$ ,  $\mu_V$  and  $\sigma_U^2$ ,  $\sigma_V^2$  their expectations and variances respectively, when they exist. Then, for large samples,  $n \gg 1$ , under null hypothesis,

the test's statistic  $w$  has approximately normal distribution with parameters  $n\mu_U$ , and  $n\sigma_U^2$ . This means that the significant level  $t_\alpha$  for given value of  $\alpha$  can be found from the equation

$$\alpha = P_{H_0} \{w(X_1, \dots, X_n) > t\} = P_{H_0} \left\{ \frac{w - n\mu_U}{\sigma_U \sqrt{n}} > \frac{t_\alpha - n\mu_U}{\sigma_U \sqrt{n}} \right\} \approx 1 - \Phi \left( \frac{t_\alpha - n\mu_U}{\sigma_U \sqrt{n}} \right),$$

or equivalently

$$\frac{t_\alpha - n\mu_U}{\sigma_U \sqrt{n}} \approx z_{1-\alpha}.$$

Here  $z_{1-\alpha}$  is the  $(1-\alpha)$ -quantile of the standard normal distribution. Thus, the critical value  $t_\alpha$  for the test statistic  $w$  at a given significance level  $\alpha$  is

$$t_\alpha \approx n\mu_U + z_{1-\alpha} \sigma_U \sqrt{n}. \quad (6)$$

The power of the test equals

$$\begin{aligned} \pi_\alpha &= P_{H_1} \{w(Y_1, \dots, Y_n) > t_\alpha\} = P_{H_1} \left\{ \frac{w - n\mu_V}{\sigma_V \sqrt{n}} > \frac{t_\alpha - n\mu_V}{\sigma_V \sqrt{n}} \right\} = \\ &= 1 - \Phi \left( \frac{t_\alpha - n\mu_V}{\sigma_V \sqrt{n}} \right) = 1 - \Phi \left( \frac{\mu_U - \mu_V}{\sigma_V} \sqrt{n} + z_{1-\alpha} \frac{\sigma_U}{\sigma_V} \right). \end{aligned} \quad (7)$$

From this equality it is possible to see that the power of the test mainly depends on the difference in expectations of the r.v.'s  $U$  and  $V$ .

In some cases the parameters  $\mu_U$ ,  $\mu_V$  and  $\sigma_U^2$ ,  $\sigma_V^2$  can be calculated in closed (explicit) form. In general it is possible to estimate them also with the help of Monte-Carlo techniques and then use the respective estimated values instead of the exact ones. Appropriate algorithms for calculating the empirical cumulative distribution functions (CDF) of the test's statistic under hypotheses  $H_0$  and  $H_1$  for both cases are described below.

### 3 Algorithms

In this section two algorithms for calculation of the tails of CDF of LRT's statistic  $w$  under both null and alternative hypothesis (the null  $H_0 : f(x) = f_0(x)$  and the alternative  $H_1 : f(x) = f_1(x)$ ), based on Monte-Carlo method are proposed. One algorithm can be applied for any sample size  $n$ . The second algorithm should be used for large samples,  $n \gg 1$ , mainly when the parameters  $\mu_U$ ,  $\mu_V$  and  $\sigma_U^2$ ,  $\sigma_V^2$  are finite.

#### **Algorithm 1. LRT for any sample size**

**Begin.** Select the p.d.f.'s  $f_0(\cdot)$  and  $f_1(\cdot)$ , and the sample size  $n$ .

**Step 1.** Generate a sequence of  $N$  random samples  $(x_1^{(j)}, \dots, x_n^{(j)})$ ,  $j=1, \dots, N$ , from a distribution with p.d.f.  $f_0(\cdot)$ , and calculate  $N$  values of the test's statistics

$$w_j = w(x_1^{(j)}, \dots, x_n^{(j)}) = \sum_{1 \leq i \leq n} (\ln f_1(x_i^{(j)}) - \ln f_0(x_i^{(j)})), \quad j=1, \dots, N. \quad (8)$$

**Step 2.** Calculate the complementary empirical distribution function

$$\bar{F}_{0,N}(t) = \frac{1}{N} \{ \text{number of } w_j \text{'s } > t \}, \quad t > 0.$$

**Step 3.** Calculate the critical value  $t_\alpha$  for the test statistic  $w$  at a given significance level  $\alpha$  as the smallest solution of the equation

$$\bar{F}_{0,N}(t) = \alpha.$$

**Step 4.** Generate a sequence of  $N$  random samples  $(y_1^{(j)}, \dots, y_n^{(j)})$ ,  $j=1, \dots, N$ , from a distribution with p.d.f.  $f_1(\cdot)$ , and calculate the values of the analogous to (8) test statistics  $w_j$  with  $y_i^{(j)}$ 's instead of  $x_i^{(j)}$ 's.

**Step 5.** Calculate the complementary empirical distribution function for the new sample

$$\bar{F}_{1,N}(t) = \frac{1}{N} \{ \text{number of } w_j \text{'s } > t \}, \quad t > 0.$$

**Step 6.** Calculate the power of the test statistic  $w$  at the given significance level  $\alpha$  from the equation

$$\bar{F}_{1,N}(t_\alpha) = \pi_\alpha.$$

**Step 7.** **Enter the application's data:** For a given user's sample  $(x_1, \dots, x_n)$ , calculate the test statistic

$$w = w(x_1, \dots, x_n) = \sum_{1 \leq i \leq n} (\ln f_1(x_i) - \ln f_0(x_i)).$$

Calculate the  $p$ -value for testing the null hypothesis  $H_0: f(x) = f_0(x)$  versus the alternative  $H_1: f(x) = f_1(x)$  by making use of the Likelihood Ratio Test from the equation

$$\bar{F}_{0,N}(w) = p\text{-value}.$$

Make a decision by comparing the calculated  $p$ -value and  $\alpha$ . Alternatively, reject the hypothesis  $H_0$  if the inequality  $w > t_\alpha$  holds.

Calculate the *probability of committing an error of type II* (when testing the null hypothesis  $H_0: f(x) = f_0(x)$  versus the alternative  $H_1: f(x) = f_1(x)$  by making use of the Likelihood Ratio Test by the simulation method) from the equation

$$1 - \bar{F}_{1,N}(w) = \beta - \text{the probability of type II error}.$$

**Step 8.** Print results:

- The chosen null hypothesis  $H_0: f(x) = f_0(x)$  and alternative hypothesis  $H_1: f(x) = f_1(x)$ , the selected significance level  $\alpha$ , and the sample size  $n$ .
- The  $p$ -value of the test
- The power of the test,  $\pi_\alpha = 1 - \beta$ .
- The calculated value of the test statistic  $w$ , and the calculated by simulation critical value  $t_\alpha$
- The graphs of the tails of the empirical CDFs  $\bar{F}_{0,N}(t)$  and  $\bar{F}_{1,N}(t)$ .

**End.**

For large size samples when the second moments of the r.v.'s  $U$ , and  $V$  exist, it is possible to modify and simplify the simulation algorithm as shown below.

## Algorithm 2. LRT for large samples.

**Begin.** Select the p.d.f.'s  $f_0(\cdot)$  and  $f_1(\cdot)$ , and the sample size  $n$ .

**Step 1.** Generate a sequence of  $N$  random variables  $(x_1, \dots, x_N)$ , from a distribution with p.d.f.  $f_0(\cdot)$ , and calculate  $N$  values of the statistics

$$u_j = u(x_j) = \ln f_1(x_j) - \ln f_0(x_j), \quad j=1, \dots, N. \quad (9)$$

and its sample mean  $\bar{u}$ , and sample variance  $s_u^2$  according to

$$\bar{u} = \frac{1}{N} \sum_{1 \leq j \leq N} u_j, \quad s_u^2 = \frac{1}{N} \sum_{1 \leq j \leq N} (u_j - \bar{u})^2 = \frac{1}{N} \sum_{1 \leq j \leq N} u_j^2 - (\bar{u})^2. \quad (10)$$

**Step 2.** Calculate the critical value  $t_\alpha$  for the test statistic  $w$  at a given significance level  $\alpha$  from the equation

$$t_\alpha \approx n\bar{u} + z_{1-\alpha} \cdot s_u \cdot \sqrt{n}, \quad (11)$$

where  $z_{1-\alpha}$  is the  $(1-\alpha)$ -quantile of the standard normal distribution.

**Step 3.** Generate a sequence of  $N$  random variables  $(y_1, \dots, y_N)$ , from a distribution with p.d.f.  $f_1(\cdot)$ , and calculate  $N$  values of the statistics

$$v_j = v(y_j) = \ln f_1(y_j) - \ln f_0(y_j), \quad j=1, \dots, N. \quad (12)$$

and its sample mean  $\bar{v}$ , and sample variance  $s_v^2$  according to rules (10) for the

data (12).

**Step 4.** Calculate the power of the test at the given significance level  $\alpha$  from the equation

$$\pi_\alpha = 1 - \Phi \left( \frac{\bar{u} - \bar{v}}{s_v} \sqrt{n} + z_{1-\alpha} \frac{s_u}{s_v} \right). \quad (13)$$

where  $\Phi(x)$  is the c.d.f. of the standard normal distribution.

**Step 5.** **Enter the application's data:** For a given user's sample  $(x_1, \dots, x_n)$ , calculate the test

statistic

$$w = w(x_1, \dots, x_n) = \sum_{1 \leq i \leq n} (\ln f_1(x_i) - \ln f_0(x_i)).$$

Calculate the  $p$ -value for testing the null hypothesis  $H_0: f(x) = f_0(x)$  versus the alternative  $H_1: f(x) = f_1(x)$  by making us of the Likelihood Ratio Test from the equation

$$p\text{-value} = P_{H_0} \{w(X_1, \dots, X_n) > w\} \approx 1 - \Phi \left( \frac{w - n \cdot \bar{u}}{s_u \sqrt{n}} \right),$$

where  $w$  is the calculated statistic from the sample. Make a decision by comparing the calculated  $p$ -value and  $\alpha$ . Alternatively, reject the hypothesis  $H_0$  if the inequality  $w > t_\alpha$  holds, where  $t_\alpha$  is calculated by (11).

Calculate the *probability of committing an error of type II* (when testing the null hypothesis  $H_0 : f(x) = f_0(x)$  versus the alternative  $H_1 : f(x) = f_1(x)$  by making use of the LRT by the simulation method) from the equation  $\beta = 1 - \pi_\alpha$  with the  $\pi_\alpha$  calculated in Step 4.

**Step 6.** Print results:

- The chosen null hypothesis  $H_0 : f(x) = f_0(x)$  and alternative hypothesis  $H_1 : f(x) = f_1(x)$ , the selected significance level  $\alpha$ , and the sample size  $n$  ;
- The  $p$ -value of the test;
- The power of the test,  $\pi_\alpha = 1 - \beta$  ;
- The calculated test statistic  $w$ , and the calculated by simulation critical value  $t_\alpha$  ;
- The graphs of the tails of the CDFs

$$\bar{F}_{0,N}(t) = 1 - \Phi\left(\frac{t - n \cdot \bar{u}}{s_u \sqrt{n}}\right), \text{ and } \bar{F}_{1,N}(t) = 1 - \Phi\left(\frac{t - n \cdot \bar{v}}{s_v \sqrt{n}}\right).$$

**End.**

## 4 The Software

For practical application of the above algorithms an appropriate software should be elaborated. The software should have a friendly interface, which allows work in two different regimes: *individual* (customized), and *automatic*.

In the individual regime only particular observations are tested for any pair of given null and alternative hypotheses. Automatic regime allows one to calculate and show the significance level and power functions as functions of the test's statistic, and also as functions of some parameters of the model. In this way it would allow one to investigate some parametric models.

The interface includes the main menu, which allows the users to choose:

- the regime for investigation;
- the p.d.f. for hull and alternative hypotheses from a given list of distributions, which include almost all standard discrete and continuous distributions, or
- propose an option to the user for selecting probability distribution's formulae or tables of his/her own choice.

The submenu allows:

- one to choose the parameter values for hypothesis testing for individual regime; or
- one to choose the intervals and steps of increment for parameters varying for the problem investigated in an automatic regime.

The software allows also different type of presentation of the results: numerical, graphical, comparison with respect to various variables, or with respect to family of functions. These and other appropriate possibilities make the content of the design menu.

The design will be based on the new technologies presented in [8].

## 5 An Example

Below we consider one example on which the work of algorithms in the previous section will be illustrated.

**Example.** An ALM distribution versus other ALM distribution with uniform distribution.

It is known that when in the ALM distribution

$$f(x) = (1 - a) \cdot a^{\left[\frac{x}{c}\right]} f_Y\left(x - \left[\frac{x}{c}\right]c\right), \quad (14)$$

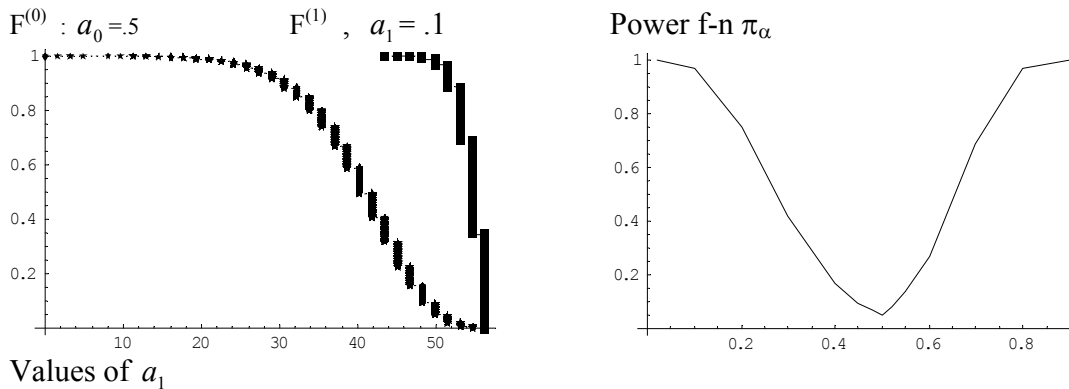
where  $a$  is a parameter of distribution,  $c$  is the length of a period, and  $f_Y(x)$  is an arbitrary distribution on the interval  $[0,c)$ . More details about ALM distributions can be found in [5].

Here for the ALM distribution in the null hypothesis  $H_0$  we choose  $f_0(x)$ , presented by (14) with parameters chosen in the following way

$$c=1, \quad a_0=.5, \quad f_{Y,0}(x)=1 \text{ for } 0 \leq x \leq 1. \quad (15)$$

This means that the r.v.  $X$  with distribution (14) is based on the uniform distribution of  $Y_0$  on  $[0,1]$  (cycle of length 1), and probability for jump over a cycle without success is  $a_0=.5$ . Any other choice of the parameter  $a_0 \neq .5$  will produce an ALM distribution  $f_1(x)$ , different from the chosen  $f_0(x)$ . And this p.d.f.  $f_1(x)$  will appear in our considerations as an alternative hypothesis  $H_1$ .

Thus, we study the likelihood ratio test according to Algorithms 1 and 2 above with the choice for the p.d.f.  $f_0(x)$ , with  $a_0=.5$ , and choosing various other values for parameter  $a_1 \neq .5$ . In studying the power function dependence on significance level  $\alpha$  we select  $a_1 = .05, .1, .15, \dots, .9, .95$ ;  $N = 10000$ ,  $n = 10$ ,  $c = 1$ .



**Fig. 1. Cumulative distribution functions for the test statistic and the power function of the test**

The results for the power function in this case of significance level  $\alpha = .05$  are shown on Fig. 1.

## 6 Conclusions

The problem of hypotheses testing arises in many statistical applications. In analytical form its solution can be done for a very limited number of cases. The method proposed in this paper gives the solution for practically all cases. Nevertheless, for its practical realization special computer tools with friendly interface are needed. This work is now in the progress, and we show here

some examples of the approach used for some special case of distributions, - so called almost lack of memory distributions.

## References

[1] **S.Chukova, and B.Dimitrov** (1992) On distributions having the almost lack of memory property. *Journal of Applied Probability*, **29** (3), 691-698.

[2] **S.Chukova, B.Dimitrov, and Z. Khalil** (1993) A characterization of probability distributions similar to the exponential. *The Canadian Journal of Statistics*, **21** (3), 269-276.

[3] **B.Dimitrov, Z.Khalil, M.Ghitany, and V.Rykov** (2001) Likelihood Ratio Test for Almost Lack of Memory Distributions. *Technical Report No. 1/01*, November 2001, Concordia University, 2001.

[4] **B.Dimitrov, and Z.Khalil** (1992) A class of new probability distributions for modeling environmental evolution with periodic behavior. *Environmetrics*, **3** (4), 447-464.

[5] **B.Dimitrov, S.Chukova, and D.Green Jr.**(1997) Distributions in Periodic Random Environment and Their Applications. *SIAM J.on Appl. Math.*, **57** (2), 501-517.

[6] **B.Dimitrov, Z.Khalil, and L.Le Normand** (1993) A statistical study of periodic environmental processes, *Proceedings of the International Congress on Modelling and Simulation*, Dec. 6-10, 1993, The Univ. of Western Australia, Perth, **v. 41**, pp. 169 - 174.

[7] **R.A.Johnson** (1994) *Miller & Freund's Probability and Statistics for Engineers*, 5th edition, Prentice Hall, New Jersey 07632.

[8] **J.Carroll** (2002) *Human - Computer Interaction in the New Millennium*, ACM Press, NY.

[9] **A.M.Law and W.D.Kelton** (2000) *Simulation Modeling and Analysis*, 3d edition, McGraw-Hill, N.Y.