Mechanics Transport Communications
Academic journal
ISSN 1312-3823
issue 3, 2008
article № 0303
http://www.mtc-aj.com

STATISTICAL HYPOTHESIS TESTING AND THEIR APPLICATION IN RISK TECHNICAL SYSTEMS

Antonio ANDONOV, Angel TANEV
ang_tanev@abv.bg

Higher School of Transport, Sofia 158 Geo Milev Str.
BULGARIA

Abstract: *In this paper we consider how can be defined the presence or absence of useful signal when we have a low ratio signal/noise. An optimal Neyman-Pearson decision rule is applied to distinguish the different hypothesis. Finally, an example showing how many measurements we need to conduct in order to satisfy given requirement regarding the probability of missed detection in case of given probability of false alarm.*

Key words: *probability of missed detection, probability of false alarm, Neyman-Pearson's rule*

1. INTRODUCTION

Let the following sample is given: $x = (x_1, ..., x_n)$ and we have to take decision for the probability distribution based on two hypothesis:

H_0: x distributed with law $p_0(x)$;
H_1: x distributed with law $p_1(x)$ (if x-continuous, then $p_0(x)$, $p_1(x)$- densities; if discrete -probabilities).

Receiving the sequence $x = (x_1, ..., x_n)$ we have to accept one of the two decisions: "true H_0" (it will be noted as 0) or "true H_1" (it will be noted as 1). It is related to definition of decision function $\delta(x)$, corresponding to two levels: 0 and 1, i.e. following definition:

$$(1) \qquad \delta(x) = \begin{cases} 0, \text{ if } x \in \Gamma_0, \\ 1, \text{ if } x \in \Gamma_1 \end{cases}$$

Using the decision function $\delta(x)$ the following errors are possible [1]:

Error 1: α *- to accept* H_1 *when* H_0 *is true;*
Error 2: β *- to accept* H_0 *when* H_1 *is true.*

We desire the errors mentioned above to be approximately zero. But, if one of them is decreasing, for example α, then the other one- β will increase. There are different approaches about the definitions of optimal rules.

2. PRESENTAION OF NEYMAN-PEARSON OPTIMAL DECISION RULE

An optimal rule is such one which possess given probability of error 1: α, and the probability of error 2: β is minimum. The rule $\delta(x)$ is optimal if:

$$(2) \qquad \beta(\Gamma) = \min_{\Gamma'} \beta(\Gamma'),$$

with the condition $\alpha(\Gamma') \le \alpha_0$.

It turns out, for the optimal rule that the area Γ_1 is:

$$(3) \qquad \Gamma_1 = \left\{ x : \frac{p_1(x)}{p_0(x)} \ge h \right\},$$

where: h is obtained from the condition $\alpha(h) = \alpha_0$

3. EXAMPLE OF THE NEYMAN-PEARSON's RULE APPLICATION

Let on the sensor input (fig.1) is applied signal S which can accept two values [2]:

S = 0 (no signal), $S = a \neq 0$ (there is useful signal).

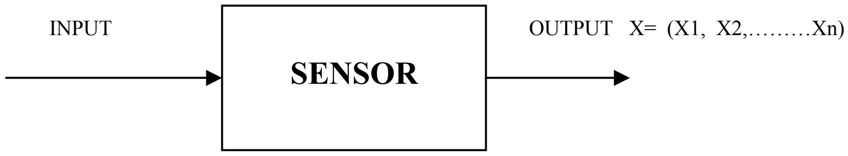

Fig.1. Sensor

In the channel there is additive random error ε, normal distributed with mean value $M\varepsilon = 0$ и variance $D\varepsilon = \sigma^2$, the result is $x' = S + \varepsilon$. The measurements are repeated n times , then on the output we have the sequence $x = (x_1, \ldots, x_n)$ and we have to take decision for the presence of the signal (H_1: $S = a$) or absence (H_0: $S = 0$). It requires to create the decision rule δ which posses given probability of error 1: α (probability of false alarm):

$$\alpha \equiv P\,(\text{accept } H_1 \mid H_0) = \alpha_0$$

and the probability of error 2: β (probability of missed detection) is **minimum**.
Considering the independence of the errors, and taking into account the presence (H_1) or absence (H_0) of the signal, we obtain:

$$(4)\quad p_1(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - a)^2}{2\sigma^2}}, \quad p_0(x) = = \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_i^2}{2\sigma^2}}.$$

In accordance with equation (3), the decision for the presence of the signal can accept H_1, if

$$(5)\quad \Gamma_1 = \left\{ x : \ln \frac{p_1(x)}{p_0(x)} \geq \ln h_1 \equiv h_1 \right\} = \left\{ x\!: \frac{1}{2\sigma^2} \left(2a \sum_{i=1}^{n} x_i - na^2 \right) \geq h_1 \right\} = \left\{ x\!: \sum_{i-1}^{n} x_i \geq h_2 \equiv \frac{h_1 2\sigma^2 + na^2}{2a} \right\}$$

Therefore, if:

$$(6)\quad \sum_{i=1}^{n} x_i \geq h_2$$

then we accept H_1; otherwise we accept H_0. The threshold is obtained by the expression:

$$(7)\quad \alpha(h_2) = P\{\text{accept } H_1 / H_0\} = = P\left(\sum_{i=1}^{n} x_i \geq h_2 / H_0 \right) = \alpha_0.$$

If the hypothesis H_0 is true, then $\sum_{i=1}^{n} x_i$ is normal distributed with mean value 0 and variance $n.\sigma^2$, therefore the last condition is:

$$(8)\quad \alpha(h_2) = 1 - \Phi\left(h_2 / \sqrt{n\sigma^2}\right) = \alpha_0,$$

Thus, the threshold is [3]:

$$(9)\quad h_2 = \sigma \sqrt{n}\, Q(1 - \alpha_0),$$

where: Φ(x)- function of normal distribution with $N(0,1)$; $Q(1 - \alpha_0)$- quantile according to $(1 - \alpha_0)$ of this distribution.

Now we define the probability β for the procedure (6) taking into account (9). If the H_1 is true, then the $\sum_{i=1}^{n} x_i$ is normal distributed with mean value $n.a$ and variance $n.\sigma^2$, and then the error 2: β is:

$$\beta = P\ (\text{accept } H_0 / H_1) = P\ \{\sum_{i=1}^{n} x_i < h_2 / H_1\} = \Phi$$

$$(10) \quad \left(\frac{h_2 - na}{\sigma\sqrt{n}} \right) = \Phi(Q - \frac{a}{\sigma}\sqrt{n}),$$

Let we have already defined in test conditions that the mean value of useful signal $a = 0.2$ and the standard deviation is $\sigma = 1$ (the error in this case is 5 times higher than the useful signal), the number of measurements *n=500,* $\alpha = 10^{-2}$ and we obtain:

$$\boldsymbol{h_2 = 1 \cdot \sqrt{500} \cdot 2.33 = 52, \quad \beta = \Phi(2.33 - 0.2 \cdot 22.4) = \Phi(-2.14) = 1.6 \,.\, 10^{-2};}$$

We see that the error 2: β is not high- about $\mathbf{10^{-2}}$

4. MODELING BY USING EXCEL

The example which has been given above is subjected to modeling by using EXCEL. This one is related to the real case of measuring the signals by using sensors. We generated two samples with number of measurements *n=500* in accordance with the hypothesis H_1 and H_0. After that we created the histograms (within the range of -2.5 and 2.5 with 20 intervals).

The simulation of these two samples has been performed by excel: **TOOLS→DATA ANALYSIS**. Then following window open: **ANALYSIS TOOLS** and you have to choose: **RANDOM NUMBER GENERATION**. After that we created histograms of the samples by using: **HISTOGRAM** and showing given intervals.

For the hypothesis (noise in the system) H_0 **(fig.2):**

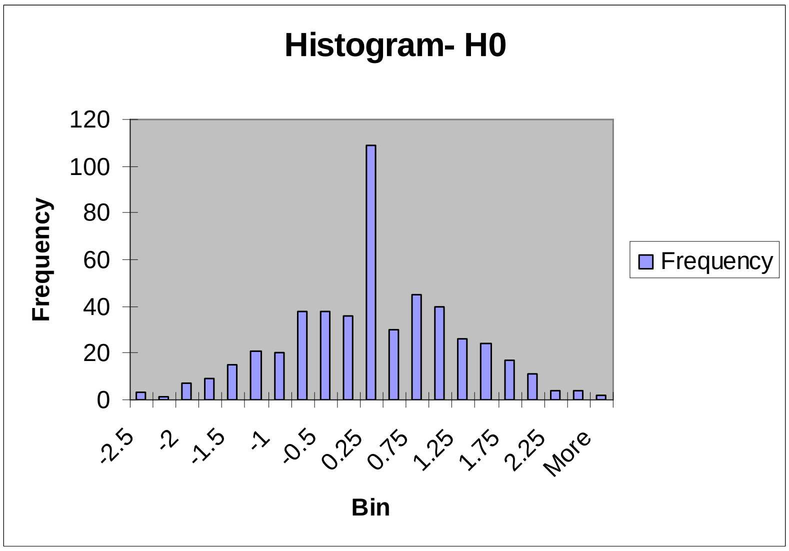

Fig.2. Hypothesis H0 (noise in the system)

And for the second one (useful signal) H_1 **(fig.3):**

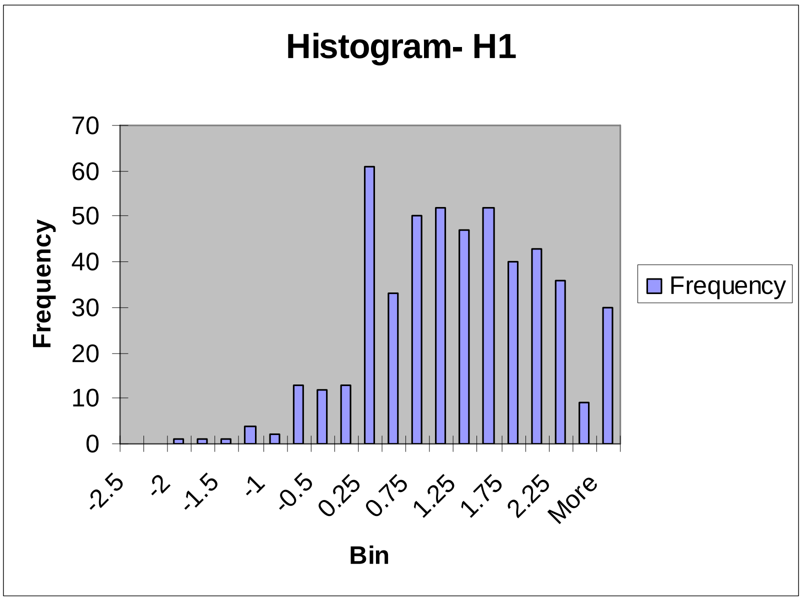

Fig.3. Hypothesis H1 (presence of useful signal)

We defined the summations of the observations for each sample and apply the decision rule (6) with the threshold (9). We can conclude that in each case the decision rule gives the correct value.
The results are shown in excel file:

5. CONCLUSIONS

5.1. a model (approach) for statistical testing hypothesis has been considered.

5.2. a simulation of two samples are shown by using EXCEL.

5.3. histograms are created in order to show the different hypothesis H_0 (noise) and H_1 (useful signal).

5.4. the optimal decision rule (Neyman-Pearson rule) has been applied to define the presence or absence of the signal in noise environment (to verify the hypothesis).

5.5. the probability of missed detection (error 2: β) is obtained based on the number of measurements, useful signal a and standard deviation σ.

REFERENCES:

[1]. Gindev E., *An introduction of engineering reliability*, Part 1, "Marin Drinov" Academic publishing house, Sofia, 2000.

[2]. Akashi, H., and H. Kumamoto, "Random Sampling Approach to State Estimation in Switching Environments", *Automatica*, Vol. 13, 1987.

[3]. Anderson, B. D. O., and J.B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1999.

СТАТИСТИЧЕСКО ТЕСТВАНЕ НА ХИПОТЕЗИ И ТЯХНОТО ПРИЛОЖЕНИЕ В РИСКОВИ ТЕХНИЧЕСКИ СИСТЕМИ

Антонио АНДОНОВ, Ангел ТАНЕВ

ВТУ „Т. Каблешков", София, К-ра „СОТС"
БЪЛГАРИЯ

Резюме: *В този доклад се разглежда дефинирането на наличието или отсъствието на полезен сигнал, когато отношението сигнал/шум е ниско. Приложен е критерия на Нейман-Пирсън за да различи различните хипотези. Показан е пример колко измервания трябва да бъдат направени за да се удовлетвори изискването за вероятността за пропуснато откриване при дадена вероятност за лъжлива тревога.*

Ключови думи: *вероятност за пропуснато откриване, вероятност за лъжлива тревога, Нейман-Пирсън критерий*