A Decision-Theoretic Approach To the Problem of Testing a Null Hypothesis*

by

Herman Rubin

Department of Statistics
Division of Mathematical Sciences

Mimeograph Series No. 259

June 1971

*Research was supported in part by the Office of Naval Research Contract N00014-67-A-0226-0008,Project Number NR042-216 at Purdue University.

A DECISION-THEORETIC APPROACH TO THE PROBLEM OF TESTING A NULL HYPOTHESIS*

By Herman Rubin

Purdue University

1. *Summary*. We consider the testing of the "*null hypothesis*" $\theta = 0$ against the one-dimensional alternative $\theta \neq 0$. In most problems, the investigator knows that $\theta = 0$ is unreasonable, and would prefer to "*accept*" $\theta = 0$ if $|\theta|$ is sufficiently small. We make the assumption that the problem is sufficiently regular, that is, that the likelihood function is sufficiently close to that of a sample from a normal distribution with mean θ and variance 1, after normalization if necessary. We give a mathematical formulation of this problem and investigate the solution. It is shown that a crude procedure based on a "*small sample*" treatment and a "*very large sample*" treatment can be very bad in the transition region; also, there is not enough information in those treatments to get robust results. Further work is contemplated to see if a small amount of additional information will suffice to obtain robust procedures using only information which the user can reasonably supply.

---

*Research was supported in part by the Office of Naval Research Contract N00014-67-A-0226-0008. Project Number NR042-216 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

2. *Mathematical treatment.* Let X be the mean of a sample of size n from $N(\theta,1)$. Let the weight density for accepting $\theta = 0$ be $c\theta^2$, and let the weight measure for rejecting $\theta = 0$ be μ. Then the risk of accepting if $|X| < \xi$ is

$$(1)\qquad \int_{-\infty}^{\infty} \int_{-\xi}^{\xi} c\theta^2 \sqrt{\frac{n}{2\pi}}\; e^{-n(x-\theta)^2/2}\, dx\, d\theta + \int \int_{|x|>\xi} \sqrt{\frac{n}{2\pi}}\; e^{-(x-\theta)^2/2}\, d\mu(\theta) \quad .$$

Now the first integral in (1) can readily be evaluated as

$$(2)\qquad 2c\left(\frac{\xi}{n} + \frac{\xi^3}{3}\right) \quad .$$

Suppose $\theta = \alpha\phi$, $x = \alpha y$, $\xi = \alpha\eta$, $n = \alpha^{-2}m$. Then (1) becomes

$$(3)\qquad 2c\alpha^3\left(\frac{\eta}{m} + \frac{\eta^3}{3}\right) + \int\int_{|y|>\eta} \sqrt{\frac{m}{2\pi}}\; e^{-m(y-\phi)^2/2}\, d\mu(\alpha\phi) \quad .$$

Suppose $|\mu| < \infty$. Then if α is chosen so that

$$(4)\qquad c\alpha^3 = \frac{1}{\sqrt{2\pi}}\,|\mu| \quad ,$$

and $d\nu(\phi) = d\mu(\alpha\phi)/|\mu|$, the risk is

$$(5)\qquad |\mu|\left[\frac{2}{\sqrt{2\pi}}\left(\frac{\eta}{m} + \frac{\eta^3}{3}\right) + \nu * N\left(0,\frac{1}{m}\right)(\{y: |y| > \eta\})\right] \quad .$$

We will take this as our standard form. Suppose ν is symmetric. Then it is easily seen by differentiating that

$$(6)\qquad \frac{1}{\sqrt{2\pi}}\left(\frac{1}{m} + \eta^2\right) = \frac{d\; \nu * N\left(0,\frac{1}{m}\right)(\eta)}{d\eta} \quad .$$

Unfortunately, the solution of this equation for the optimal $\hat{\eta}$ depends heavily on ν. Let us first see what happens in two special cases.

The case in which ν is concentrated at 0 corresponds to the situation in which there is positive prior probability that $\theta = 0$, and for any deviation from $\theta = 0$ rejection would be preferred. In this case, (6) becomes

$$\frac{1}{\sqrt{2\pi}}\left(\frac{1}{m} + \eta^2\right) = \frac{\sqrt{m}}{\sqrt{2\pi}} e^{-\frac{1}{2} m \eta^2} , \tag{7}$$

so that if $m \geq 1$ acceptance is possible. (This is the reason for choosing the particular normalization.) Let us call the solution for this case η_N.

Another case is that in which the sample size is so large that (6) is approximately

$$\frac{1}{\sqrt{2\pi}} \eta^2 = \frac{d\nu(\eta)}{d\eta} . \tag{8}$$

This is the case in which sampling error is unimportant and the question is merely whether θ is small enough that $\theta = 0$ should be accepted. Let the optimal η for this be η_D.

The first simple procedure which comes to mind is to consider $\eta^* = \max(\eta_N, \eta_D)$. That this can be very bad is easily seen computationally in the case ν normal.

From a theoretical analysis of the problem with 0-th power loss for ν normal, a procedure $\tilde{\eta}$ suggested itself. Let P_N be the probability of type I error under the null hypothesis of rejection beyond η_N and let P_D be the probability of rejection beyond η_D. Then the probability of rejection beyond $\tilde{\eta}$ is $P_N P_D$. This does not give as good results far away from the critical values of n as η^*, but rarely is much worse if ν is normal.

The optimal procedure $\hat{\eta}$ and the risks of $\hat{\eta}$, η^*, and $\tilde{\eta}$ were computed for ν $N(0,\sigma^2)$, $\sigma^2 = 10^{-k}$, $k = 1(1)20$, and $m = 10^j$, $j = .1(.1)20$. The most striking results were at the extreme for $\sigma^2 = 10^{-20}$ (see table). Note that a sample of "size" 10^{20} is 25,000 times as bad as one of half the size for the crude procedure η^*. One might argue that 10^{20} is too large a sample size; however, there is a scale factor involved, and 10^{20} might correspond to a much smaller sample. However, the bad behavior of the crude procedure holds for $\sigma^2 < 10^{-3}$, and a table is included for $\sigma^2 = 10^{-8}$.

While the central limit theorem gives us reason to make a normal approximation for the statistic, there does not seem to be a compelling reason for the weight measure ν to be normal. Computations with ν double-exponential turned out to be feasible, and this was done for scale factors 10^{-k}, $k = .5(.5)10$ and m as before. As is seen in the enclosed tables for scale factors 10^{-10} and 10^{-} the crude procedure η^* shows the same type of behavior as before, but not as extreme; the procedure $\tilde{\eta}$ is not too good, giving risks 45% too high and regrets (excesses of risk over that of knowledge of the parameter) of 80% too high. It is possible to develop an analog of $\tilde{\eta}$ for the double-exponential, but then the double-exponential was chosen only for computational convenience, and no clear brief can be made for it.

In the case of ν the Cauchy distribution, the only case we have done is for $m = 10^{20}$, $\eta_N = \eta_D$. Here η^* gives a risk of 1.56×10^{-27}, $\hat{\eta}$, 1.55×10^{-27}, and $\tilde{\eta}$, 1.95×10^{-27}. This indicates that the tail nature of ν near η_D is very important and further investigation is being made of this problem.

Acknowledgment. The author wishes to acknowledge the invaluable help of Arthur Rubin in the construction and especially the debugging of the programs for the numerical results.

References

1. Rubin, H. "Occam's Razor needs new blades", Purdue Mimeo Series No. 216, to appear in the Proceedings of the Symposium on the Foundations of Statistical Inference held at Waterloo, 1970.

2. Rubin, H. and Sethuraman, J. (1965). "Bayes risk efficiency". *Sankyā*, A, Vol. 27 pp. 347-356.

ν double exponential, scale factor 10^{-10}			
m	10^{26} risk $(\hat{\eta})$	10^{26} risk (η^{*})	10^{26} risk $(\tilde{\eta})$
$10^{17.3}$	365.80	366.10	402.71
$10^{17.7}$	96.52	97.87	119.92
10^{18}	37.01	45.64	52.80
$10^{18.3}$	17.53	4,424.52	26.21
$10^{18.5}$	12.54	2,049,000	17.89
$10^{18.7}$	9.94	6,409.12	13.15
10^{19}	8.02	50.29	9.45
$10^{19.3}$	7.16	9.73	7.75
$10^{19.7}$	6.70	6.88	6.82
∞	6.36	6.36	6.36

ν double exponential, scale factor $10^{-4.5}$			
m	10^{10} risk $(\hat{\eta})$	10^{10} risk (η^{*})	10^{10} risk $(\tilde{\eta})$
10^{7}	26.05	26.27	31.03
$10^{7.3}$	10.39	11.09	13.88
$10^{7.7}$	3.84	19.97	5.56
$10^{7.9}$	2.78	85.98	3.85
10^{8}	2.46	24.54	3.29
$10^{8.3}$	1.92	3.24	2.28
$10^{8.7}$	1.63	1.73	1.73
∞	1.45	1.45	1.45

ν normal $(0,10^{-20})$			
m	10^{27} risk $(\hat{\eta})$	10^{27} risk (η^{*})	10^{27} risk $(\tilde{\eta})$
10^{18}	349.35	351.01	351.48
$10^{18.7}$	34.93	45.60	35.53
10^{19}	13.58	85.00	13.91
$10^{19.3}$	5.60	3,194.46	5.77
$10^{19.7}$	2.01	44,893,000	2.08
10^{20}	1.11	1.112×10^{12}	1.14
∞	.40	.40	.40

ν normal $(0,10^{-8})$			
m	10^{10} risk $(\hat{\eta})$	10^{10} risk (η^{*})	10^{10} risk $(\tilde{\eta})$
10^{6}	636.50	637.17	642.63
10^{7}	28.66	34.44	29.97
$10^{7.3}$	12.21	31.57	12.96
$10^{7.7}$	4.53	369.49	4.86
10^{8}	2.54	10,860	2.71
$10^{8.3}$	1.68	183.85	1.76
$10^{8.7}$	1.21	3.95	1.24
10^{9}	1.06	1.37	1.08
∞	.93	.93	.93

Table 1

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Purdue University	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

A Decision-Theoretic Approach to the Problem of Testing a Null Hypothesis

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report June, 1971

5. AUTHOR(S) *(Last name, first name, initial)*

Rubin, Herman

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June, 1971	6	2

8a. CONTRACT OR GRANT NO. N00014-67-A-0226-0008 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Mimeograph Series #259 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research Washington, D. C.

13. ABSTRACT

We consider the testing of the "null hypothesis" $\theta = 0$ against the one-dimensional alternative $\theta \neq 0$. In most problems, the investigator knows that $\theta = 0$ is unreasonable, and would prefer to "accept" $\theta = 0$ if $|\theta|$ is sufficiently small. We make the assumption that the problem is sufficiently regular, that is, that the likelihood function is sufficiently close to that of a sample from a normal distribution with mean θ and variance 1, after normalization if necessary. We give a mathematical formulation of this problem and investigate the solution. It is shown that a crude procedure based on a "small sample" treatment and a "very large sample" treatment can be very bad in the transition region; also, there is not enough information in those treatments to get robust results.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification