A Bayesian Version of

The Sign Test

by

Daniel L. Solomon

BU-372-M June, 1971

ABSTRACT

An easily computed Bayesian analogue of the sign test is developed for testing the hypothesis $P(X \geq Y) \geq \theta_o$ against $P(X \geq Y) < \theta_o$, where (X, Y) is a continuous bivariate random variable and θ_o is a specified probability.

A BAYESIAN VERSION OF THE SIGN TEST

By Daniel L. Solomon

Cornell University

SUMMARY

An easily computed Bayesian analogue of the sign test is developed for testing the hypothesis $P(X \geq Y) \geq \theta_0$ against $P(X \geq Y) < \theta_0$, where (X,Y) is a continuous bivariate random variable and θ_0 is a specified probability.

1. INTRODUCTION

There is a large literature on nonparametric and distribution free problems and a large literature on the problems of Bayesian inference but little in their intersection. Some references to the literature can be found in a review paper by Savage (1969). Another important reference is Saxena (1965). The difficulty in attacking these problems will more likely be in obtaining useful formulations than in their solutions.

This paper is designed to illustrate an approach to the area by formulating a Bayesian version of the sign test. A convenient model to have in mind is that for paired comparison experiments. For example, suppose that we have a panel of judges, each of whom is presented with a sample of each of two brands, A and B, of some food product and asked to taste each and state a preference. For our purposes we shall suppose that each judge is forced to state a preference and cannot claim "no preference." It is assumed that the physiological responses to the two treatments have continuous distributions, but that only the statement of preference is observable. An object of the procedure to be derived is to determine "which is the better brand" and to make that question precise.

Formally, suppose that (X_1,Y_1), $(X_2,Y_2),\ldots,(X_n,Y_n)$ is a sample of n independent observations on the continuous bivariate random variable (X,Y), where X_i represents the i^{th} judge's response to brand A and Y_i his response to B. We assume that he will state a preference for A over B if the realized values (x,y) of (X,Y) are such that $x \geq y$, and he will prefer B if $x < y$. Now, let $P(X \geq Y) = \theta$. If $\theta > \frac{1}{2}$ then we will say that A is better than B, if $\theta < \frac{1}{2}$, B is better than A, and if $\theta = \frac{1}{2}$, A and B are equally good. Now θ is unknown and we wish to test hypotheses about θ. Herein lies the first problem. The usual formulation of the sign test (ignoring ties) as found for example in Siegel (1956), is to test the null hypothesis H_1: $\theta = \frac{1}{2}$ against either of the one sided alternatives or the two sided alternative, H_2: $\theta \neq \frac{1}{2}$. Note that the null hypothesis, $P(X - Y \geq 0) = \frac{1}{2}$, is a test of the hypothesis that the <u>median</u> of the difference is zero. Recall the technique: For each of the pairs (x_i,y_i) we determine the sign of the difference $x_i - y_i$ (perhaps this sign is all that is observable). Under the null hypothesis, the number of plus signs is an observation on the binomial distribution with parameters n and $\frac{1}{2}$. Thus by reference to tables of the binomial distribution one can calculate either a significance level or perform a hypothesis test of the appropriate size.

In a Bayesian framework this may not be an acceptable formulation of the problem. The Bayesian wants to put a prior distribution on θ. This prior will likely be a continuous one, and so the null hypothesis, $\theta = \frac{1}{2}$, being a single point in the parameter space will have prior probability zero and therefore posterior probability zero regardless of the data. Therefore, the null hypothesis will, a priori, be rejected. If one insists on testing a simple hypothesis, the prior distribution must put non zero probability on that hypothesis.

Here, instead of attempting to test that simple hypothesis, we shall develop a test of the hypothesis H_1: $\theta \geq \frac{1}{2}$ against the alternative H_2: $\theta < \frac{1}{2}$. In fact, there is no further complication in testing the more general hypothesis H_1: $\theta \geq \theta_o$ against H_2: $\theta < \theta_o$ where $0 < \theta_o < 1$ is some specified number, and so that is what we shall do.

2. THE MODEL

For $i = 1,2,\ldots,n$, let Z_i be the random variable defined by $Z_i = 1$ if $X_i - Y_i \geq 0$ and $Z_i = 0$ otherwise. Thus $P(Z_i = 1) = 1 - P(Z_i = 0) = \theta$, $S = \Sigma Z_i$ is the number of times X equals or exceeds Y, and for given θ, S has the binomial distribution with parameters n and θ. Furthermore, S is a sufficient statistic for the family of distributions of the Z's indexed by θ and so can serve as a basis for the test.

If the problem is viewed as a two action decision problem in which action a_1 corresponds to the acceptance of H_1 and action a_2 corresponds to acting as if H_2 were true, then the loss function for the problem is assumed to be reasonably represented by the values in the following table:

<u>$L(a,\theta)$</u>

	$\theta \epsilon H_1$	$\theta \epsilon H_2$
a_1	0	ℓ_2
a_2	ℓ_1	0

That is ℓ_1 is the loss associated with making a type one error, and ℓ_2 with making a type two error.

From the Bayesian viewpoint, θ is a given value of the random variable $\tilde{\theta}$, for which we must specify a prior distribution, i.e., a probability distribution on the interval (0,1). We now suggest that the experimenter choose his prior from the natural conjugate class which in this case, since the sampling distribution is binomial, is the class of Beta distributions. The motivation for the suggestion is of course that it makes the problem tractable. However, for the suggestion to be reasonable, the class must be rich enough to contain the experimenter's actual prior, or at least a good approximation. This two parameter family contains symmetric distributions, both positively and negatively skewed distributions, (all unimodal, however), U-shaped and J-shaped distributions, and as a degenerate case, the uniform distribution on (0,1).

The sampling distribution is

$$f_{S|\tilde{\theta}}(s|\theta) = P(S = s|\theta) = \binom{n}{s}\theta^{s}(1 - \theta)^{n-s} \;;\; s = 0,1,2,\ldots,n, \quad 0 < \theta < 1,$$

and so the natural conjugate family is the class of distributions with densities of the form:

$$f_{\tilde{\theta}}(\theta) = \frac{1}{B(s',n'-s')}\,\theta^{s'-1}(1-\theta)^{n'-s'-1}; \quad 0 < \theta < 1, \quad n' > s' > 0.$$

Here $B(\cdot,\cdot)$ is the Beta function, and s' and n' are parameters of the prior distribution which has been reparameterized from the usual form of the Beta density for reasons which will become apparent. Now the posterior density of $\tilde{\theta}$ corresponding to this prior is

$$f_{\underline{\tilde\theta}|S}(\theta|s) = K(s)f_{S|\underline{\tilde\theta}}(s|\theta)f_{\underline{\tilde\theta}}(\theta)$$

$$= K(s)\theta^{s+s'-1}(1-\theta)^{(n+n')-(s+s')-1}$$

$$= \frac{1}{B(s'',n''-s'')}\theta^{s''-1}(1-\theta)^{n''-s''-1} \qquad 0 < \theta < 1,$$

which is the Beta density with parameters s'' and $n''-s''$, where $s'' = s + s'$ and $n'' = n + n'$. $K(s)$ is of course the reciprocal of the marginal density of S.

Now to calculate a Bayes rule, i.e., a test, let δ be the decision rule defined by

$$\delta(s) = \begin{cases} a_1 & \text{if } s \in A_1 \\ a_2 & \text{if } s \in A_2 = \bar{A}_1, \end{cases}$$

where A_1 is the set of values of S leading to acceptance of H_1 under δ. Thus, specifying a test is equivalent to specifying A_1. The Bayes risk is

$$BR(\delta) = E_{\underline{\tilde\theta}}E_{S|\underline{\tilde\theta}}L(\delta(S),\underline{\tilde\theta}) = E_S E_{\underline{\tilde\theta}|S}L(\delta(S),\underline{\tilde\theta})$$

$$= E_S\begin{cases} \ell_2 P(\underline{\tilde\theta} < \theta_o|S = s) & \text{if } s \in A_1 \\ \ell_1 P(\underline{\tilde\theta} \geq \theta_o|S = s) & \text{if } s \in A_2 \end{cases}$$

$$= \sum_{s\in A_1} \ell_2 P(\underline{\tilde\theta} < \theta_o|S = s)f_S(s) + \sum_{s\in A_2} \ell_1 P(\underline{\tilde\theta} \geq \theta_o|S = s)f_S(s).$$

To minimize the Bayes risk, it will suffice to select a rule δ which assigns a_1 to those observations s for which the contribution to the first sum would be less than the contribution to the second. That is, $\delta(s) = a_1$ if $\ell_2 P(\underline{\tilde\theta} < \theta_o|S = s) < \ell_1 P(\underline{\tilde\theta} \geq \theta_o|S = s)$ or equivalently if $\ell_2 P(H_2|S = s) < \ell_1 P(H_1|S = s)$. Thus a rule which selects H_1, if its posterior probability is larger than $\ell (\equiv \ell_2/\ell_1)$ times the posterior probability of H_2, is a Bayes rule. Recalling that the posterior distribution is the Beta with parameters s'' and n''-s'', and that $P(H_1|S = s) = 1 - P(H_2|S = s)$, the inequality becomes

$$I_{\theta_o}(s'', n''-s'') < \frac{1}{1+\ell} \tag{1}$$

where I is the incomplete Beta function:

$$I_x(a,b) \equiv \frac{1}{B(a,b)} \int_o^x t^{a-1}(1-t)^{b-1}dt.$$

Thus to test H_1: $\theta \geq \theta_o$ against H_2: $\theta < \theta_o$ with prior parameters s', n' - s', loss ratio ℓ and n observations, observe S = s as the number of times $X \geq Y$, calculate s'' = s + s', n'' = n + n' and accept H_1 if and only if (1) is satisfied. The test thus is simple to compute and uses readily available tables.

As a special case, if $\ell = 1$ and $\theta_o = \frac{1}{2}$, then (1) becomes

$$I_{\frac{1}{2}}(s'', n''-s'') < \frac{1}{2}$$

which, it can be shown, is equivalent to $s''/n'' > \frac{1}{2}$. But $s''/n'' = \mathbf{E}\left[\tilde{\theta}|S = s\right]$, is the posterior mean of $\tilde{\theta}$.

For a numerical example, we return to the tasting problem. Suppose that the brand corresponding to response Y is an existing brand and that to X is from a new "improved" process and is being considered to replace the standard. Suppose further that cost and sales considerations are such that in order for the change of process to be profitable, at least $\frac{2}{3}$ of the product buying population would have to prefer the new product. Thus we want to test H_1: $\theta \geq \frac{2}{3}$ against H_2: $\theta < \frac{2}{3}$. Suppose also, that the cost structure is such that it is 1.5 times as expensive to make a type II error as a type I, so that $\ell = 1.5$. A type II error leads to set-up costs for change in production and advertising costs both of which fail to bring increased sales. A type I error leads to a loss of the potential increase in sales with the new product.

Finally, we need a prior distribution, i.e., s' and n'. The company set out to produce an improved product, and so they are reasonably certain that they have. The prior distribution should therefore put much probability in the region of large values of $\tilde{\theta}$, that is, large probabilities of preference for the new product. The experimenter might specify for example, that the mean of his prior is $s'/n' = \frac{3}{4}$. Now, considering Beta distributions with mean $\frac{3}{4}$ and perhaps specifying some other characteristic of the prior distribution, say variance or coefficient of skewness, he determines s' and n'. For purposes of this example, suppose the prior has $s' = 27$ and $n' = 36$. Now we present each of $n = 32$ judges with a sample of each product and insist that he state a preference. Suppose that the "improved" product (X) is preferred to the old (Y) $s = 21$ times; then we will accept H_1 and produce the new version only if

$$I_{\frac{2}{3}}(21 + 27,\ 32 + 36 - 21 - 27) < \frac{1}{1+1.5} = .4 \quad .$$

From tables of the incomplete Beta function, we find that $I_{\frac{2}{3}}(48,20) \doteq .24 < .4$ and thus we make the new product.

REFERENCES

Savage, I. R. (1969). Nonparametric Statistics: a personal review. Sankhyā, Ser. A, 31, 107-144.

Saxena, K. M. Lal (1965). Some nonparametric Bayesian estimation problems. Florida State University Statistics Report M100.

Siegel, S. (1956). Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill.