# Sample Size Determination for Some Common Nonparametric Tests

GOTTFRIED E. NOETHER*

The article discusses the problem of determining the number of observations required by some common nonparametric tests, so that the tests have power at least $1 - \beta$ against alternatives that differ sufficiently from the hypothesis being tested. It is shown that the number of observations depends on certain simple probabilities. A method is suggested for fixing the value of the appropriate probability when determining sample size.

KEY WORDS: Sign test; Wilcoxon one-sample test; Wilcoxon two-sample test; Kendall test of independence; Odds ratio.

## 1. INTRODUCTION

The problem of determining an adequate sample size for a test of a hypothesis can be formulated as follows. We want an $\alpha$-level test to reject the hypothesis being tested with probability at least $1 - \beta$, whenever the alternative differs sufficiently from the null hypothesis. For parametric tests, it is natural to express the requirement "differs sufficiently from the null hypothesis" in terms of the parameter being tested. But how should the requirement be interpreted for nonparametric tests? We shall see that associated with some common nonparametric tests, there exist certain probabilities that can be used to measure distance from the null hypothesis for purposes of sample size determination.

Throughout the article, we assume that the distribution of a test statistic $T$ is approximately normal with mean $\mu(T)$ and standard deviation $\sigma(T)$. In particular, the mean and standard deviation of $T$ under the null hypothesis will be denoted by $\mu_0(T)$ and $\sigma_0(T)$. For the nonparametric tests to be discussed, the assumption of the approximate normality of the test statistics is sufficiently accurate for most practical purposes, unless the sample sizes involved are quite small.

For simplicity, we discuss sample size determination for an upper-tailed test. The result remains valid for a lower-tailed test and requires only an obvious modification for a two-tailed test. If then $Z$ denotes a standard normal variable and $z_\alpha$ denotes its upper $\alpha$-level significance point, the critical region for the test is given by $T > \mu_0(T) + z_\alpha\sigma_0(T)$. The power of the test against the alternative $H_a$ is given by

$$\begin{aligned}\text{Power} &= P(T > \mu_0(T) + z_\alpha\sigma_0(T) \mid H_a) \\ &= P\left(\frac{T - \mu(T)}{\sigma(T)} > \frac{\mu_0(T) - \mu(T) - z_\alpha\sigma_0(T)}{\sigma(T)}\right) \\ &= P\left(Z > \frac{\mu_0(T) - \mu(T)}{\rho\sigma_0(T)} + \frac{z_\alpha}{\rho}\right),\end{aligned}$$

where $\rho = \sigma(T)/\sigma_0(T)$. The power of this test equals $1 - \beta$ when the expression on the right of the inequality sign equals $-z_\beta$ or if

$$Q(T) = \left[\frac{\mu(T) - \mu_0(T)}{\sigma_0(T)}\right]^2$$

equals $(z_\alpha + \rho z_\beta)^2$. In general, the value of $\rho$ is unknown. But for alternatives that do not differ too much from the null hypothesis, it will often be appropriate to assume that $\sigma(T)$ is close to $\sigma_0(T)$ or, equivalently, that we may set $\rho = 1$. If we refer to $Q(T)$ as the noncentrality factor for the test $T$, under the stated conditions an approximation to the sample size is obtained by setting the noncentrality factor equal to $(z_\alpha + z_\beta)^2$ and then solving the resulting equation for the number of observations. It is easily seen that a lower-tailed test leads to exactly the same solution.

We illustrate our result with the classical one-sample problem. Let $X_1, \ldots, X_N$ constitute a random sample from a population with mean $\mu$ and standard deviation $\sigma$. We want to test the hypothesis that $\mu$ equals $\mu_0$ by using the test statistic $T = \overline{X}$. Then $\mu(\overline{X}) = \mu$, $\sigma^2(\overline{X}) = \sigma^2/N$, and $Q(\overline{X}) = N(\mu - \mu_0)^2/\sigma^2$. We then have the well-known result that the required sample size is given by

$$N = \frac{(z_\alpha + z_\beta)^2}{[(\mu - \mu_0)/\sigma]^2}.$$

## 2. NONPARAMETRIC ONE-SAMPLE LOCATION TESTS

Let $X_1, \ldots, X_N$ constitute a random sample from a population with median $\eta$. We want to test the hypothesis that $\eta = \eta_0$. Without loss of generality, we assume that $\eta_0 = 0$.

### 2.1 Sign Test

As our test statistic, we use the quantity $S = \#$(positive observations). Then $\mu(S) = Np$, where $p = P(X > 0)$. In particular, $\mu_0 = N/2$. Further, $\sigma^2(S) = Np(1 - p)$ and $\sigma_0^2(S) = N/4$. Then $Q(S) = 4N(p - \frac{1}{2})^2$ and the required sample size $N = N(S)$ is given by

$$N(S) = \frac{(z_\alpha + z_\beta)^2}{4(p - \frac{1}{2})^2},$$

which depends on how much the true probability $p$ that an observation exceeds the hypothetical median deviates from the hypothetical probability $\frac{1}{2}$.

This result raises a practical problem for the experimenter. What value $p$ should he choose for the alternative?

* Gottfried E. Noether is Professor Emeritus, Department of Statistics, University of Connecticut, Storrs, CT 06268.


Journal of the American Statistical Association
June 1987, Vol. 82, No. 398, Theory and Methods

He may want to rely on past information or use a pilot sample to obtain an estimate of $p$. Alternatively, the experimenter may want to fix in his mind some value for the odds ratio $r = p/(1 - p) = P(\text{positive observation})/P(\text{negative observation})$. The associated value of $p$ is $r/(1 + r)$. An experimenter who is willing to tolerate accepting a hypothetical median with probability at most $\beta$, if in fact the probability that an observation is greater than the hypothetical median is at least twice as large as the probability that it is smaller than the hypothetical median, would use the value $p = \frac{2}{3}$ when computing the required sample size. Settling on an odds ratio to determine the required sample size is certainly no more arbitrary than settling on a value for $(\mu - \mu_0)/\sigma$ in the classical one-sample problem, when the standard deviation is unknown.

For the sign test, $\rho = \sigma(S)/\sigma_0(S) = 2\sqrt{p(1 - p)}$. Since, under the alternative, $\rho$ is smaller than 1, the suggested sample size $N(S)$ is conservative. The degree of conservativeness of $N(S)$ is indicated by $N(S)/N(\rho)$, where $N(\rho) = (z_\alpha + \rho z_\beta)^2/4(p - \frac{1}{2})^2$ is the normal approximation to the sample size based on the actual value $\sigma(S)$. When $\beta = \alpha$, $N(S)/N(\rho) = 4/[1 + 2\sqrt{p(1 - p)}]^2$. For some typical value of the odds ratio $r$, we find

| $r$ | $p$ | $N(S)/N(\rho)$ |
|---|---|---|
| 1.5 | .60 | 1.02 |
| 2 | $\frac{2}{3}$ | 1.06 |
| 3 | .75 | 1.15. |

When $\beta > \alpha$, the ratio $N(S)/N(\rho)$ is smaller.

As an illustration, we tabulate $N(S)$ and $N(\rho)$ for $\alpha = \beta = .10$. The table also gives values of the approximation $N(a) = [(z_\alpha + z_\beta)/\arcsin(2p - 1)]^2$, suggested by a referee, and the sample size $N(b)$ of the exact binomial test randomized to have $\alpha = .10$.

| $p$ | $N(S)$ | $N(\rho)$ | $N(a)$ | $N(b)$ |
|---|---|---|---|---|
| .60 | 164.4 | 161.0 | 162.1 | 162 |
| $\frac{2}{3}$ | 59.2 | 55.8 | 56.9 | 57 |
| .75 | 26.3 | 22.9 | 24.0 | 24 |

All three approximations would seem to give practically useful results, the arcsin approximation being closest to the randomized exact test sample size.

## 2.2 Wilcoxon One-Sample Test

The Wilcoxon statistic for testing the hypothesis that a symmetric population is centered at zero is usually computed as $W = \sum \text{rank}|X|$, where the summation extends over the positive observations. For our purposes, an equivalent computational form is preferable:

$$W = \#[\text{positive}(X_i + X_j)], \qquad 1 \le i \le j \le N.$$

We find $\mu(W) = Np + \frac{1}{2}N(N - 1)p'$, where again $p = P(X > 0)$ and $p' = P(X + X' > 0)$, $X$ and $X'$ being two independent observations. Under the null hypothesis, $p = p' = \frac{1}{2}$ and $\sigma_0^2(W) = N(N + 1)(2N + 1)/24$. Then

$$Q(W) = \frac{[N(p - \frac{1}{2}) + \frac{1}{2}N(N - 1)(p' - \frac{1}{2})]^2}{N(N + 1)(2N + 1)/24} \doteq 3N(p' - \tfrac{1}{2})^2$$

for sufficiently large $N$. It follows that the required sample size $N(W)$ for the Wilcoxon one-sample test is given by

$$N(W) = \frac{(z_\alpha + z_\beta)^2}{3(p' - \frac{1}{2})^2}.$$

The comments concerning an appropriate choice of $p$ for the sign test apply equally to the choice of $p'$. In particular, for a given value of the odds ratio $r' = P(X + X' > 0)/P(X + X' < 0)$, we get $p' = r'/(1 + r')$.

## 2.3 Comparison of $N(W)$ and $N(S)$

From the formulas for $N(W)$ and $N(S)$, we see that the Wilcoxon test requires fewer (more) observations than the sign test depending on whether $|p' - \frac{1}{2}|/|p - \frac{1}{2}|$ is greater (smaller) than $\sqrt{\frac{4}{3}} = 1.15$. It is then of some interest to investigate the relationship between $p'$ and $p$ in greater detail.

Let $X = U + \eta$ $(\eta > 0)$, where the random variable $U$ is symmetric about 0 with density $f(u)$ and cdf $F(u)$. Then

$$p = P(X > 0) = P(U > -\eta) = F(\eta).$$

Correspondingly we find

$$p' = P(X + X' > 0) = P(U + U' > -2\eta) = F^*(2\eta),$$

where $F^*(u)$ is the cdf of the random variable $U + U'$ [the convolution of $F(u)$].

For a number of distributions, $p$ and $p'$ can be evaluated in closed form:

1. $U \sim$ uniform on $(-\frac{1}{2}, \frac{1}{2})$:

   $$p = \tfrac{1}{2} + \eta, \qquad p' = \tfrac{1}{2} + 2\eta(1 - \eta), \qquad \eta < \tfrac{1}{2}.$$

2. $U \sim N(0, 1)$:

   $$p = \Phi(\eta), \qquad p' = \Phi(\eta\sqrt{2}),$$

   where $\Phi(u)$ = standard normal cdf.

3. $U \sim$ Laplace (0, 1):

   $$p = 1 - \tfrac{1}{2}e^{-\eta}, \qquad p' = 1 - \tfrac{1}{2}(1 + \eta)e^{-2\eta}.$$

   For $\eta$ near zero, we have the first order approximations $p = p' = \frac{1}{2}(1 + \eta)$.

4. $U \sim$ Cauchy (0, 1):

   $$p = \tfrac{1}{2} + 1/\pi \tan^{-1}\eta = p'.$$

These results suggest that for long-tailed distributions like the Cauchy and the Laplace distributions, the values of $p'$ and $p$ are very nearly equal. For these distributions, the value of $p$ used to determine the sample size for the sign test can also be used to determine the sample size for the Wilcoxon one-sample test. For relatively short-tailed distributions like the normal, $p'$ values are somewhat larger than the corresponding $p$ values, as the following table shows:

| $p$ | $r$ | $p'$ | $r'$ |
|---|---|---|---|
| .55 | 1.22 | .57 | 1.33 |
| .60 | 1.50 | .64 | 1.78 |
| .65 | 1.86 | .71 | 2.45 |
| .70 | 2.33 | .77 | 3.35. |

### 2.4 Pitman Efficiency of the Sign Test Relative to the Wilcoxon Test

The results of Sections 2.1 and 2.2 immediately convert into statements of relative efficiency. Thus the efficiency of the sign test relative to the Wilcoxon one-sample test is given by the ratio

$$e_{S,W} = \frac{N(W)}{N(S)} = \frac{4(p - \frac{1}{2})^2}{3(p' - \frac{1}{2})^2}.$$

According to the results in Section 2.3, relative efficiency depends not only on the population distribution $F(u)$, but also on the specific alternative $\eta$. The dependence on $\eta$ disappears as we let $\eta \to 0$ (Pitman local efficiency). For $\eta$ values near 0, we find

$$p = F(\eta) = \tfrac{1}{2} + \eta f(0)$$

and

$$p' = F^*(2\eta) = \tfrac{1}{2} + 2\eta f^*(0) = \tfrac{1}{2} + 2\eta \int f^2(u)\, du.$$

Then

$$e_{S,W} = \frac{f^2(0)}{3\left[\int f^2(u)\, du\right]^2}.$$

In particular, we have the following results: for the uniform distribution, the relative efficiency is $\frac{1}{3}$; for the normal distribution, $\frac{2}{3}$; for the logistic distribution, $\frac{3}{4}$; and for the Laplace and Cauchy distributions, $\frac{4}{3}$.

## 3. TWO-SAMPLE PROBLEM

Given two independent random samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, we want to test the hypothesis that the two samples have come from the same population against the alternative that $Y$-observations tend to be larger than $X$-observations.

As test statistic, we use the Mann–Whitney statistic $U$ for the Wilcoxon two-sample test,

$$U = \#(Y_j > X_i), \qquad i = 1, \ldots, m; \; j = 1, \ldots, n.$$

We find $\mu(U) = mnp''$, where $p'' = P(Y > X)$. In addition, $\mu_0(U) = \frac{1}{2}mn$ and $\sigma_0^2(U) = mn(N + 1)/12$, where $N = m + n$. In terms of the probability $p''$, the alternative can be stated more precisely as $p'' > \frac{1}{2}$.

Setting $m = cN$, we find $Q(U) = 12c(1 - c)N^2(p'' - \frac{1}{2})/(N + 1)$ and, approximately,

$$N = \frac{(z_\alpha + z_\beta)^2}{12c(1 - c)(p'' - \frac{1}{2})^2}.$$

For a given odds ratio $r'' = P(Y > X)/P(Y < X)$, we find $p'' = r''/(1 + r'')$.

## 4. KENDALL'S TEST OF INDEPENDENCE

Given the bivariate random sample $(X_1, Y_1), \ldots, (X_N, Y_N)$, we want to test the hypothesis that the two random variables $X$ and $Y$ are independent. As test statistic, we use $C = \#$(concordant pairs), where two sample observations $(X, Y)$ and $(X', Y')$ are said to be concordant or discordant, depending on whether $(X - X')(Y - Y')$ is positive or negative. We find $\mu(C) = \frac{1}{2}N(N - 1)p_c$, where $p_c$ is the probability of concordance. Under the hypothesis of independence, $p_c = \frac{1}{2}$ and $\sigma_0^2(C) = N(N - 1)(2N + 5)/72$. Approximately, then,

$$N = \frac{(z_\alpha + z_\beta)^2}{9(p_c - \frac{1}{2})^2}.$$

The practical determination of $p_c$ can be based on the odds ratio $p_c/p_d$, where $p_d = 1 - p_c$ is the probability of discordance. Alternatively, we have $p_c = \frac{1}{2}(1 + \tau)$, where $\tau = p_c - p_d$ is the Kendall rank correlation coefficient. Thus a value of $\tau$ for which an experimenter prefers rejection of the hypothesis of independence determines a corresponding value of $p_c$.

[*Received February 1986. Revised October 1986.*]