HHS Public Access

Author manuscript

Model Assist Stat Appl. Author manuscript; available in PMC 2021 January 22.

Published in final edited form as:

Model Assist Stat Appl. 2020 ; 15(4): 335–349. doi:10.3233/mas-200505.

Approximations of the power functions for Wald, likelihood ratio, and score tests and their applications to linear and logistic regressions

Eugene Demidenko

Dartmouth College, Hanover, NH 03755, USA

Summary.

Traditionally, asymptotic tests are studied and applied under local alternative (Aivazian, et al., 1985). There exists a widespread opinion that the Wald, likelihood ratio, and score tests are asymptotically equivalent. We dispel this myth by showing that These tests have different statistical power in the presence of nuisance parameters. The local properties of the tests are described in terms of the first and second derivative evaluated at the null hypothesis. The comparison of the tests are illustrated with two popular regression models: linear regression with random predictor and logistic regression with binary covariate. We study the aberrant behavior of the tests when the distance between the null and alternative does not vanish with the sample size. We demonstrate that these tests have different asymptotic power. In particular, the score test is generally asymptotically biased but slightly superior for linear regression in a close neighborhood of the null. The power approximations are confirmed through simulations.

Keywords

Effective sample size; GLM; Linear regression; Logistic regression; Local alternative; Sample size determination

1. Introduction

Uniformly most powerful tests exist only in rare statistical models - usually they exist for linear model with fixed/nonrandom predictors and normal distribution (Aivazian et al. 1985; Lehmann and Romano, 2005). On the other hand, asymptotic tests, such as the Wald, likelihood ratio, and score tests can be applied to a much wider variety of statistical distributions and models when the sample size, n, increases to infinity. For a short review of these test we refer the reader to a recent book by the author [9].There exists a widespread opinion that the three tests are asymptotically optimal and equivalent for large n, as stated by Rayner (1997), Young and Smith (2005), among many others. Because of this opinion, not much preference is given to what test to use to determine the desired sample size.

Usually, the treatment of asymptotic tests is reduced to the analysis with local alternatives, or in the terminology of Cox and Hinkley (1974), *contiguous alternatives*, based on the

eugened@dartmouth.edu.

concept of contiguity (Lehmann and Romano, 2005). It is a textbook result that the Wald and likelihood ratio tests are equivalent for alternatives at the distance of $O(n^{-1/2})$ from the null value. The following warning on page 156 from a book by Robert Serfling (1980) is the impetus to the present work: "Therefore, under appropriate regularity conditions, the statistics λ_n, W_n and V_n are asymptotically *equivalent* in distribution, both under the null hypothesis and under local alternatives converging sufficiently fast. However, at fixed alternatives these equivalences are not anticipated to hold." (Here, λ_n, W_n and V_n are likelihood ratio, Wald and score statistics, respectively.) Moreover, we argue that the study with contiguous alternatives is not appealing from the practical point of view. For example, it's not applicable to the sample size determination where the alternative is fixed.

The goal of the present work is to analyze the two-sided tests in the presence of nuisance parameters using the power function with the emphasis on the *global (or fixed)* alternatives. Respectively, the local properties are expressed in terms of the first and second derivative evaluated at the null value of the parameter. The tests are then illustrated with linear regression and normally distributed predictor when the variance is unknown and logistic regression with Bernoulli covariate for which the asymptotic power function is derived in closed form.

It is natural to expect that the rate of rejecting the null hypothesis increases as the alternative gets farther from the null. However, Hauck and Donner (1977) and Væth (1985) showed that the Wald test may be aberrant in this respect - we investigate the aberrant behavior of the likelihood ratio and score tests as well.

Much of the effort, with some controversy, has been made on the comparison of the tests using the method of contiguous alternatives. In particular, several claims have been made that the Rao's score test is superior to Wald and likelihood ratio test (Chandra and Joshi, 1983; Chandra and Mukerjee, 1984). We confirm this claim for a linear model with random predictors in a close proximity to the null. Otherwise, these tests are different and it is impossible to claim an overall champion. Moreover, the score test is usually biased.

The organization of the paper is as follows. After introducing notation and definitions, we derive the power function for the Wald, likelihood ratio, and score tests in each of the following sections and illustrate it with linear and logistic regression models. The three tests are compared in terms of power.

2. Notation and definitions

Throuthout the paper, we deal with iid observations $\{\mathbf{z}_i, i = 1, \ldots, n\}$ having common density $f = f(\mathbf{z}; \boldsymbol{\theta})$ dependent on a vector parameter $\boldsymbol{\theta} = (\beta, \boldsymbol{\gamma})$; formally, $\theta_1 = \beta$. The first component, β, is treated as the parameter of interest and $\boldsymbol{\gamma}$ is treated as a p-dimensional vector of nuisance parameters (Pawitan, 2001). The null hypothesis is composite, $H_0 : \beta = \beta_0$ with the two-sided alternative, $H_A : \beta \neq \beta_0$. For expository purposes, we shall assume that $\beta_0 = 0$, so that the null hypothesis takes the form

$$H_0 : \beta = 0. \tag{1}$$

If $T = T(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ is a test statistic such that the null is rejected when $T > c$, the power function for (1) can be expressed as a function of the alternative,

$$P(\beta; \gamma) = \Pr(T > c \mid \boldsymbol{\theta}). \tag{2}$$

Here, $P(0; \boldsymbol{\gamma}) = \alpha$ is the the significance level or the size of the test (typically, $\alpha = 0.05$) and $\mid \boldsymbol{\theta}$ means that the probability (2) is computed under the assumption that the true parameter is $\boldsymbol{\theta}$. We vary β but fix $\boldsymbol{\gamma}$, so we treat (2) as a function of β. In this paper, the power function (2) is computed, or more precisely, approximated, for large n. This, fortunately, implies that $P(0)$ does not depend on nuisance parameters. We assume that the necessary regularity conditions for asymptotic maximum likelihood estimation and hypothesis testing are fulfilled, such as the support of f does not depend on the parameter, differentiation under the expectation is valid and the power function is twice differentiable (Casella and Berger, 1990; Schervish, 1995; Bickel and Doksum, 2001).

The log-likelihood function is

$$l(\beta, \gamma) = \sum_{i=1}^{n} \ln f(\mathbf{z}_i; \beta, \gamma). \tag{3}$$

The maximum likelihood estimator (MLE), $(\hat{\beta}_{ML}, \hat{\gamma}_{ML})$, is the solution of $1 + p$ score equations

$$\frac{\partial l(\beta, \gamma)}{\partial \beta} = 0, \quad \frac{\partial l(\beta, \gamma)}{\partial \gamma} = \mathbf{0}. \tag{4}$$

The Fisher expected information matrix is defined as

$$\begin{bmatrix} I_{11} & \mathbf{I}'_{12} \\ \mathbf{I}_{12} & \mathbf{I}_{22} \end{bmatrix} = -E \begin{bmatrix} \frac{\partial^2 \ln f}{\partial \beta^2} & \frac{\partial^2 \ln f}{\partial \beta \partial \gamma} \\ \frac{\partial^2 \ln f}{\partial \gamma \partial \beta} & \frac{\partial^2 \ln f}{\partial \gamma^2} \end{bmatrix}. \tag{5}$$

(As a part of regularity conditions we assume that the information matrix is nonsingular.) Asymptotically, the variance of $\hat{\beta}_{ML}$ can be derived from the (1, 1)th element of the inverse matrix,

$$var\left(\hat{\beta}_{ML}\right) = \frac{1}{n} V(\beta), \tag{6}$$

where

$$V(\beta) = \frac{1}{I_{11} - \mathbf{I}'_{12} \mathbf{I}_{22}^{-1} \mathbf{I}_{12}} \tag{7}$$

is the *variance function* evaluated at the alternative, β. In this formula, V is evaluated at the true value of the nuisance parameter, $\boldsymbol{\gamma}$. When studying the likelihood ratio and score tests,

we shall evaluate V at other values which will be indicated as $V(\beta|\boldsymbol{\gamma_0})$. Sometimes, it is more convenient (especially for the Wald test) to deal with the *standard deviation (SD) function*, $\sigma_\beta = \sqrt{V(\beta)}$.

Now we describe the local properties of the tests in terms of the first and second derivatives of the power function evaluated at the null, $\beta = 0$; see Rao (1973, p. 454). First, since we require the tests to have size α we assume

$$P(\beta)|_{\beta=0} = \alpha\,. \tag{8}$$

Second, we want the test to be asymptotically (locally) unbiased, which means that the first derivative at the null is zero,

$$P'(0) = \left.\frac{dP}{d\beta}\right|_{\beta=0} = 0. \tag{9}$$

Third, the local power for an unbiased test is determined by the second derivative,

$$P''(0) = \left.\frac{d^2P}{d\beta^2}\right|_{\beta=0}. \tag{10}$$

The higher the value of the second derivative, the more powerful the test is for local alternatives. In fact, we can approximate the power as $P(\beta) = \alpha + \frac{1}{2}P''(0)\beta^2$ in the neighborhood of the null. Finally, the test is consistent if $\lim_{n\to\infty} P(\beta) = 1$ for all $\beta \neq 0$.

Now we derive the power functions for the three tests and illustrate them with popular statistical models: linear and logistic regressions.

3. Wald test

The Wald test is based on the fact that under the null and large n, the ratio of the MLE to its standard error, sometimes called the Z-score, is normally distributed,

$$\frac{\hat{\beta}_{ML}}{SE\left(\hat{\beta}_{ML}\right)} \sim \mathcal{N}(0,1)$$

for large n. The test statistic is the absolute value of the Z-score with $c = Z_{1-\alpha/2}$, the $(1-\alpha/2)$ quantile of the standard normal distribution, namely, $Z_{1-\alpha/2} = \Phi(1-\alpha/2)$. The power function of the Wald test in large sample can be approximated as follows

$$P_W(\beta) \simeq \Phi\left(-Z_{1-\alpha/2} + \frac{\beta}{\sigma_\beta}\sqrt{n}\right) + \Phi\left(-Z_{1-\alpha/2} - \frac{\beta}{\sigma_\beta}\sqrt{n}\right), \tag{11}$$

where sign $\simeq$ means the approximation for large n. The validity of the approximations follows from Slutsky's theorem and uniform convergence follows from continuity of derivatives (see details in Bickel and Doksum, 2001; Demidenko, 2004, p. 644). Apparently, the larger the n the better approximation.

Property (8) holds for the Wald power. It is easy to see that Wald power is symmetric about the null, $\beta = 0$. Consequently, the Wald power is asymptotically unbiased, which can be verified directly by evaluating the derivative at zero,

$$\left.\frac{dP_W}{d\beta}\right|_{\beta=0} = \left[\phi(-Z_{1-\alpha/2})\frac{\sqrt{n}}{\sigma_0} - \phi(-Z_{1-\alpha/2})\frac{\sqrt{n}}{\sigma_0}\right] \times \left.\frac{dV}{d\beta}\right|_{\beta=0} = 0,$$

where $\phi = \Phi'$, the standard normal density. The test is consistent (or better to say asymptotically consistent) because $\lim_{n\to\infty} P_W(\beta) = 1$ if $\beta \neq 0$.

The second derivative of the power function (11) at the null takes the form

$$\left.\frac{d^2P}{d\beta^2}\right|_{\beta=0} = \frac{2n}{V(0)}\phi(Z_{1-\alpha/2})Z_{1-\alpha/2}, \tag{12}$$

so approximately in the neighbourhood of zero $P_W(\beta) \simeq \alpha + n\phi + (Z_{1-\alpha/2})\, Z_{1-\alpha/2}\beta^2/V(0)$.

As follows from (11), the Wald power simplifies under *z-parametrization*,

$$P_W(z) = \Phi\left(-Z_{1-\alpha/2} + z\sqrt{n}\right) + \Phi\left(-Z_{1-\alpha/2} - z\sqrt{n}\right), \tag{13}$$

where

$$z = \frac{\beta}{\sigma_\beta} \tag{14}$$

can be treated as a theoretical counterpart of the Z-score statistic. We refer to (14) as the *z-ratio* at the alternative, β. In applications, the *z-ratio* is called the effect size (Cohen, 1994). It is easy to show that the Wald power is an increasing function for $\beta > 0$ and a decreasing function of z for $\beta < 0$. In other words, the larger the absolute value of the z-function, the greater the power. The z-parametrization reappears in the power functions of the likelihood ratio and score tests for a linear model.

3.1. Linear regression

We illustrate the calculation of the Wald power with linear regression under normal distribution. Note that typically, the power analysis is conducted for the case when predictor x is fixed/nonrandom. We however, develop the power analysis for a more complicated model when x is random, or more precisely, normally distributed. Thus, vector $\mathbf{z}$ has a multivariate normal distribution; the first component is the dependent variable, y; the second component is the variable of interest, x; the rest are covariates combined in vector $\mathbf{u}$. We test the significance of the coefficient, β, at the variable of interest, x. Without loss of generality, we can assume that the means of all regression variables are zero and the marginal variances and covariances are known:

$$\begin{bmatrix} x \\ \mathbf{u} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma'_{xu} \\ \sigma_{xu} & \Sigma_u \end{bmatrix}\right).$$

The linear regression is specified through conditional mean as

$$E(y \mid x, \mathbf{u}) = \beta x + \boldsymbol{\tau}'\mathbf{u}, \tag{15}$$

where $var(y|x, \mathbf{u}) = \sigma^2$. The conditional (error) variance, σ^2, is unknown and subject to estimation along with the vector of regression coefficients. Linear regression is used as an example of hypothesis testing in many texts, but σ^2 is assumed known (Cox and Hinkley, 1974). A more realistic set up with unknown variance brings up several surprises, as we shall see later.

Since the distribution is normal and marginal variance-covariance parameters are known, the log density of $\mathbf{z}$, up to a constant, is given by

$$\ln f(\mathbf{z}; \beta, \gamma) = -\frac{1}{2\sigma^2}(y - \beta x - \boldsymbol{\tau}'\mathbf{u})^2 - \frac{1}{2}\ln\sigma^2, \tag{16}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\tau}, \sigma^2)$. The information matrix takes the form

$$\frac{1}{\sigma^2}\begin{bmatrix} \sigma_x^2 & \boldsymbol{\sigma}_{xu}' & 0 \\ \boldsymbol{\sigma}_{xu} & \Sigma_u & 0 \\ 0 & 0 & \frac{1}{2}\sigma^{-2} \end{bmatrix},$$

so that the Wald power is given by (11) with

$$\sigma_\beta = \frac{\sigma}{\sqrt{\sigma_x^2 - \boldsymbol{\sigma}_{xu}'\Sigma_u^{-1}\boldsymbol{\sigma}_{xu}}}. \tag{17}$$

The Wald power function has a familiar V-shape, and it monotonically increases with $|\beta|$ and approaches 1 - no surprises so far.

3.2. Logistic regression with binary covariate

In general regression problems, the distribution of $\mathbf{z}$ is defined through conditional distribution as $f(\mathbf{z}; \boldsymbol{\theta}) = f(y|\mathbf{x}; \boldsymbol{\theta})\, f(\mathbf{x}; \boldsymbol{\theta})$, where y is treated as the dependent variable and $\mathbf{x}$ as the vector of covariates. As in the previous example, statistical hypothesis is concerned with testing of the slope coefficient, β.

Here, we illustrate the Wald test with logistic regression and binary covariate. This means that both y and x are binary. The probability of the dependent variable is defined through conditional probability as

$$\mathrm{pr}(y = 1 \mid x) = \frac{e^{\gamma + \beta x}}{1 + e^{\gamma + \beta x}}.$$

The marginal probability of x is specified as $\Pr(x = 1) = p_x$, which can be assumed known without loss of generality. Logistic regression is a member of the family of generalized

linear models (GLM, Bickel and Doksum, 2001). After elementary calculations, the information matrix takes the form

$$\begin{bmatrix} \frac{GBp_x}{(1+GB)^2} + \frac{G(1-p_x)}{(1+G)^2} & \frac{GBp_x}{(1+GB)^2} \\ \frac{GBp_x}{(1+GB)^2} & \frac{GBp_x}{(1+GB)^2} \end{bmatrix},$$

where to shorten the notation, we let $G = e^{\gamma}$ and $B = e^{\beta}$. Applying formula (7), we derive the variance function:

$$V(\beta) = \frac{p_x(1+G)^2 B + (1-p_x)(1+GB)^2}{p_x(1-p_x)GB}. \tag{18}$$

For a small alternative, the Wald power can be approximated quadratically as

$$P_W(\beta) \simeq \alpha + \frac{p_x(1-p_x)G}{(1+G)^2} n\phi(Z_{1-\alpha/2})Z_{1-\alpha/2}\beta^2. \tag{19}$$

As follows from this formula, maximum power for small alternatives occurs for a symmetrically distributed covariate, $p_x = 1/2$. Using elementary calculus, one can show that symmetric distribution of the dependent variable leads to maximum power, $G = 1$ and $P(y = 1) = 1/2$. Combining these facts, we obtain an upper bound for the Wald power, $P_W(\beta) \leq \alpha + (n/16)\phi(Z_{1-\alpha/2})\, Z_{1-\alpha/2}\beta^2$. With a popular choice, $\alpha = 0.05$, we have $P_W(\beta) \leq 0.05 + 0.0036n\beta^2$. For example, to detect $\beta = 1$ with power 0.8, one needs at least $n = 208$ observations.

The trouble with Wald power (11) begins when it is computed for large alternatives, because as follows from (18), the z-ratio (14) vanishes at infinity, $\lim_{\beta\to\pm\infty} z = 0$. The fact that the Wald power for logistic regression may fall for large alternative was noticed by Hauck and Donner (1977); they termed it *aberrant* behavior. This negative property of the Wald test was later studied in a broader context of a family of exponential distributions by Væth (1985), and for a linear model by Fears et al. (1996). For large alternatives the Wald power decreases to the size of the test—see Figure 1 for a geometrical illustration. The power function is computed for $\alpha = 5\%$ and $\Pr(y = 1|\beta = 0) = G/(1 + G) = 0.25$, and $p_x = \Pr(x = 1) = 0.2$ with $n = 100$. The quadratic approximation (19) is valid in a close neighborhood of the origin. The Wald power falls at $\beta = -2.4$ and beyond this point decreases back to α when $\beta \to -\infty$. We show four simulation results for alternatives $\beta = -3, -2, 1/2$, and 1 (the number of experiments equals 5,000). For large negative values of β, the power approximation (11) is not valid because $n = 100$ is too small. To account for change in the variance function at the alternative versus the variance at the null, we introduce the *effective sample size*,

$$n_e = \frac{V(0)}{V(\beta)} n. \tag{20}$$

The effective sample size can be interpreted as the sample size modification for detection of a large alternative versus a local alternative due to the change of variance. For example, the effective sample size to detect alternative $\beta = -2$ with $n = 100$, as follows from (18), is

$$n_e = \frac{B(1+G)^2}{p_X(1+G)^2 B + (1-p_X)(1+GB)^2} n = 32.$$

The effective sample size, one third the size of the original, explains why simulations at $\beta = -2$ in Figure 1 do not quite match the power: $n_e = 32$ is simply not large enough for the Central Limit Theorem to give a good approximation.

As follows from z-parametrization (13), the value of the alternative where the monotonicity of the Wald power falls satisfies the equation

$$\beta \frac{d \ln \sigma_\beta}{d\beta} = 1, \tag{21}$$

which will be referred to as the Wald *break-down* equation. In Figure 2, we show the region within which the Wald power behaves well, meaning that it increases for $\beta > 0$ and decreases for $\beta < 0$. The boundary of the shaded region is where its derivative turns zero as defined by equation (21). Roughly, one could say that the Wald power for logistic regression is well-behaved for alternatives in the interval (−2, 2).

4. Likelihood ratio test

The test statistic of the likelihood ratio (LR) test is

$$T = 2\left[\max_{\beta,\gamma} l(\beta,\gamma) - \max_{\gamma} l(0,\gamma)\right],$$

which under the null has χ^2-distribution with one degree of freedom, $\chi^2(1)$. Hence, according to the likelihood ratio test, we reject the null if $T > q_{1-\alpha}$, where $q_{1-\alpha}$ is the $(1-\alpha)$th quantile. Under the alternative, the maximizer of $l(0, \boldsymbol{\gamma})$, which is referred to as *a profile nuisance parameter*, $\hat{\gamma}_0$, yields a biased estimator of $\boldsymbol{\gamma}$. To derive the power function, we represent the rejection probability in the following way (Self et al., 1992),

$$\begin{aligned}&\mathrm{pr}(T > q_{1-\alpha})\\ &= \mathrm{pr}\left\{2\left[\left(\max_{\beta,\gamma} l(\beta,\gamma) - l(\beta,\gamma)\right) + \left(l(\beta,\gamma) - \max_{\gamma} l(0,\gamma)\right)\right] > q_{1-\alpha}\right\}.\end{aligned}$$

Using the biased estimation equation theory, one can prove that the maximizer of $l(0, \boldsymbol{\gamma})$ converges to $\boldsymbol{\gamma_0} = \boldsymbol{\gamma_0}(\beta, \boldsymbol{\gamma})$, which is the solution to the equation

$$E_{(\beta,\gamma)}\left(\frac{\partial \ln f(\mathbf{z}; 0, \gamma_0)}{\partial \gamma_0}\right) = \mathbf{0}, \tag{22}$$

which will be referred to the *limit profile nuisance parameter*. Since

$$\lim_{n \to \infty} \frac{1}{n}\left(l(\beta, \gamma) - \max_{\gamma} l(0, \gamma) = E_{(\beta, \gamma)}(\ln f(\mathbf{z}; \beta, \gamma) - \ln f(\mathbf{z}; 0, \gamma_0))\right.$$

with probability 1, the distribution of T under the alternative is $\chi^2(1)$ with the noncentrality parameter $2n\eta(\beta)$, where

$$\eta(\beta) = E_{(\beta, \gamma)}(\ln f(\mathbf{z}; \beta, \gamma) - \ln f(\mathbf{z}; 0, \gamma_0)) \tag{23}$$

will be referred to as the η-function. Note that the η-function is nonnegative, specifically, $\eta(\beta) > 0$ for all $\boldsymbol{\gamma} \neq \boldsymbol{\gamma_0}$. Notice that for a local alternative ($\beta \to 0$), we have $\boldsymbol{\gamma_0} \to \boldsymbol{\gamma}$ and $\eta(\beta) \to 0$. The noncentrality parameter depends on β and $\boldsymbol{\gamma}$, but we fix the latter and consider η only as a function of β.

Finally, the power function of the LR test can be expressed via Φ as

$$P_{LR}(\beta) = \Phi\left(-Z_{1-\alpha/2} + \sqrt{2n\eta(\beta)}\right) + \Phi\left(-Z_{1-\alpha/2} - \sqrt{2n\eta(\beta)}\right). \tag{24}$$

An advantage of expressing the power in terms of Φ, is that one can apply it for a signed LR test, which is useful for one-sided hypotheses (Severeni, 2000). Note that the power is completely specified by function η.

The local equivalence of the Wald and LR tests has been established using the method of local alternatives, $\beta = O(0)$, Serfling (1980). We prove this by showing that the second derivatives of the power functions evaluated at $\beta = 0$ are the same.

Theorem 1.—The Wald and LR tests are asymptotically unbiased (the first derivative of the power function vanishes at the null) and locally equivalent,

$$\left.\frac{d^2 P_W}{d\beta^2}\right|_{\beta=0} = \left.\frac{d^2 P_{LR}}{d\beta^2}\right|_{\beta=0}.$$

The proof is found in the Appendix.

Note that if the η-function (23) is an increasing function of β then the likelihood ratio power increases with distance from the null. Now we derive the power function for the two regressions, as we did for the Wald power.

4.1. Linear regression

We derive the likelihood ratio power for linear regression from Section 3.1. The regression is specified by equation (15) with the log density (16). First, we find the limit profile nuisance parameter, $\boldsymbol{\gamma_0} = \boldsymbol{\gamma_0}(\beta, \boldsymbol{\gamma})$ as the solution to (22), which for linear regression is equivalent to the pair of equations, $E[(y - \boldsymbol{\tau}_*'\mathbf{u})\mathbf{u}] = \mathbf{0}$ and $\sigma_*^2 = E(y - \boldsymbol{\tau}_*'\mathbf{u})^2$. In this section, we omit the subscript $(\beta, \boldsymbol{\gamma})$ on the expectation for brevity. Using some algebra, we obtain

$$E[(y-\boldsymbol{\tau}_*'\mathbf{u})\mathbf{u}] = E[(y-\beta x-\boldsymbol{\tau}'\mathbf{u}+\beta x-(\boldsymbol{\tau}_*-\boldsymbol{\tau})'\mathbf{u})\mathbf{u}]$$
$$= E[(\varepsilon+\beta x-(\boldsymbol{\tau}_*-\boldsymbol{\tau})'\mathbf{u})\mathbf{u}] = \beta\boldsymbol{\sigma}_{xu}-\boldsymbol{\Sigma}_u(\boldsymbol{\tau}_*-\boldsymbol{\tau}),$$

which yields the solution

$$\boldsymbol{\tau}_* = \boldsymbol{\tau}+\beta\,\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\sigma}_{xu}, \tag{25}$$

$$\sigma_*^2 = \sigma^2+\beta^2\left(\sigma_x^2-\boldsymbol{\sigma}_{xu}'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\sigma}_{xu}\right). \tag{26}$$

Now we find the noncentrality parameter. Specifically, we want to express function (23) through unknown parameters β, $\boldsymbol{\tau}$, and σ^2. From (16) we obtain

$$\eta(\beta) = -\frac{1}{2}+\frac{1}{2\sigma_*^2}E(y-\boldsymbol{\tau}_*'\mathbf{u})^2+\frac{1}{2}\ln\frac{\sigma_*^2}{\sigma^2} = \frac{1}{2}\ln\left(1+\frac{\beta^2}{\sigma_\beta^2}\right),$$

where σ_β is given by (17). Finally, the power function of the LR test on the z-scale (14) can be approximated as

$$P_{LR}(z) = \Phi\left(-Z_{1-\alpha/2}+\sqrt{n}\sqrt{\ln\left(1+z^2\right)}\right)+\Phi\left(-Z_{1-\alpha/2}-\sqrt{n}\sqrt{\ln\left(1+z^2\right)}\right). \tag{27}$$

Clearly, the Wald and LR power are different for linear regression contrary to the case when σ^2 is known. Further test comparisons are found in Section 6.

4.2. Logistic regression with binary covariate

Now we apply the LR test to the logistic regression with a binary covariate from Section 3.2. For this statistical model, we have

$$\ln f(y,x;\beta,\gamma) = y\gamma+\beta(xy)-\ln\left(1+e^{\gamma+\beta x}\right). \tag{28}$$

To calculate the noncentrality parameter, we need the expectation:

$$E[\ln f(y,x;\beta,\gamma)]$$
$$= \gamma\,\mathrm{pr}(y=1)+\beta\,\mathrm{pr}(y=1,x=1)-p_x\ln(1+GB)-(1-p_x)\ln(1+G)$$
$$= \gamma\left[\frac{GB}{1+GB}p_x+\frac{G}{1+G}(1-p_x)\right]+\beta\frac{GB}{1+GB}p_x-p_x\ln(1+GB)-(1-p_x)\ln(1+G).$$

To find γ_0, we take the derivative of (28) at $\gamma=\gamma_0$ and $\beta=0$,

$$\frac{d\ln f(y,x;0,\gamma_0)}{d\gamma} = y-\frac{e^{\gamma 0}}{1+e^{\gamma 0}}.$$

We find the limit profile nuisance parameter, γ_0, as the solution to the equation

$$E_{(\gamma,\beta)}\left(\frac{d\ln f(y,x;0,\gamma_0)}{d\gamma}\right) = E_{(\gamma,\beta)}(y) - \frac{e^{\gamma 0}}{1+e^{\gamma 0}} = 0$$

or equivalently

$$\frac{GB}{1+GB}p_x + \frac{G}{1+G}(1-p_x) = \frac{G_0}{1+G_0},$$

where following our convention $G_0 = e^{\gamma 0}$. After some simplification, we express G_0 as a function of true β and γ as

$$G_0 = G_0(\gamma,\beta) = G\frac{1+GB+(B-1)p_x}{1+GB-(B-1)Gp_x}. \tag{29}$$

Finally, the η-function takes the form

$$\begin{aligned}\eta(\beta) &= (\gamma-\gamma_0)\left[\frac{GB}{1+GB}p_x + \frac{G}{1+G}(1-p_x)\right]\\ &+ \beta\frac{GB}{1+GB}p_x - p_x\ln(1+GB) - (1-p_x)\ln(1+G) + \ln(1+G_0).\end{aligned}$$

From elementary calculus, it is easy to show that the η-function has finite limits when $\beta \to \infty$ or $\beta \to -\infty$, meaning that the power does not approach 1 when the alternative goes to infinity. However, for mild values of γ and p_x, those limits are very close to 1 when n is fairly large. More details are found in Section 6.

5. The score test

The score test was originally developed by Rao (1948). This test is especially simple when the null hypothesis is simple. However, in the presence of nuisance parameters we need to derive the MLE for $\boldsymbol{\gamma}$ under the restrictive model $\beta = 0$ (see Pawitan (2001) for more detail). The idea of the score test is easy: if the null hypothesis is true then the derivative of the log-likelihood function evaluated at null should be close to zero. Specifically, let $\hat{\gamma}_0$ return the maximum of the log-likelihood function under $\beta = 0$. In other words, $\hat{\gamma}_0$ is the MLE of the profile likelihood. It is proven that the distribution of the normalized derivative,

$$S = \sqrt{\frac{V(0\mid\hat{\gamma}_0)}{n}}\frac{\partial l(0,\hat{\gamma}_0)}{\partial\beta} \tag{30}$$

is asymptotically normal with zero mean and unit variance, where $V(0\mid\hat{\gamma}_0)$ is the asymptotic variance (7) evaluated at $\beta = 0$ and $\gamma = \hat{\gamma}_0$.

The derivation of the power function of the score test is similar to the derivation of the likelihood ratio test, mainly because both use the concept of the profile likelihood. Let $\boldsymbol{\gamma_0} = \boldsymbol{\gamma_0}(\beta, \boldsymbol{\gamma})$ be the solution to (22). Then the power function of the score test is derived from the following approximations somewhat similar to what we used for the Wald test:

$$
\begin{aligned}
P_S(\beta) &= \mathrm{pr}(|S| > Z_{1-\alpha/2}) \\
&= \mathrm{pr}\left\{\sqrt{\frac{V(0\mid\hat{\gamma}_0)}{n}}\left|\frac{\partial l(\beta,\gamma)}{\partial\beta} + \left(\frac{\partial l(0,\hat{\gamma}_0)}{\partial\beta} - \frac{\partial l(\beta,\gamma)}{\partial\beta}\right)\right| > Z_{1-\alpha/2}\right\} \\
&\simeq \mathrm{pr}\left\{\left|\sqrt{\frac{V(0\mid\hat{\gamma}_0)}{V(\beta)}}Z + \sqrt{n}\sqrt{V(0\mid\hat{\gamma}_0)}\frac{1}{n}\left(\frac{\partial l(0,\hat{\gamma}_0)}{\partial\beta} - \frac{\partial l(\beta,\gamma)}{\partial\beta}\right)\right| > Z_{1-\alpha/2}\right\},
\end{aligned}
$$

Where $Z \sim \mathcal{N}(0,1)$. In terms of Φ, the power function can be approximated as

$$
\begin{aligned}
P_S(\beta) = \Phi\left(-r(\beta,\gamma)Z_{1-\alpha/2} - \sqrt{n}\sqrt{V(\beta)}\delta(\beta)\right) \\
+ \Phi\left(-r(\beta,\gamma)Z_{1-\alpha/2} + \sqrt{n}\sqrt{V(\beta)}\delta(\beta)\right),
\end{aligned} \tag{31}
$$

Where $r(\beta,\gamma) = \sqrt{V(\beta)/V(0\mid\gamma_0)}$, and

$$
\delta(\beta) = E_{(\beta,\gamma)}\left(\frac{d\ln f(\mathbf{z};0,\gamma_0)}{d\beta}\right) \tag{32}
$$

will be referred to as the delta-function. When $\beta \to 0$, we have $V(0|\boldsymbol{\gamma_0}) \to V(0)$ and $\delta(\beta) \to 0$, implying that $P_S(\beta) \to 2\Phi(-Z_{1-\alpha/2}) = \alpha$, the power of the score test at the null equals the size of the test (8). Regarding the unbiasedeness of the score test, it is easy to see that

$$
\left.\frac{dP_S}{d\beta}\right|_{\beta=0} = -\frac{\phi(-Z_{1-\alpha/2})Z_{1-\alpha/2}}{\sqrt{V(\beta)V(0\mid\gamma_0)}} \times \left.\frac{dV}{d\beta}\right|_{\beta=0}, \tag{33}
$$

so that it is unbiased if and only if the derivative of the variance is zero. For linear regression V does not depend on β and therefore the score test is unbiased. However in general the score test is asymptotically biased. We shall find out in Section 6 that this test is not equivalent to the Wald or likelihood ratio test even in the neighborhood of the null.

5.1. Linear regression

We illustrate the score power with linear regression as we did earlier for the Wald and likelihood ratio tests. The needed $\boldsymbol{\tau}_*$ and σ_*^2 are defined by (25) and (26). Now we express function (32) through parameters β, $\boldsymbol{\tau}$ and σ^2. Assuming that the expectation is taken under the true parameters (β, $\boldsymbol{\tau}$, σ^2), we obtain $\delta(\beta) = -\beta/\left(\sigma_\beta^2 + \beta^2\right)$ implying that on the z-scale the power takes the form

$$
P_S(z) = \Phi\left(-\frac{Z_{1-\alpha/2}}{\sqrt{1+z^2}} - \frac{z}{1+z^2}\sqrt{n}\right) + \Phi\left(-\frac{Z_{1-\alpha/2}}{\sqrt{1+z^2}} + \frac{z}{1+z^2}\sqrt{n}\right). \tag{34}
$$

It is easy to see that $P_S(0) = \alpha$, and condition (8) holds. After some algebra, one may check that the score test is asymptotically unbiased—condition (9) holds. We shall compare this power approximation to that of the other two tests in Section 6.

5.2. Logistic regression

As in Section 4.2, the limit profile nuisance parameter G_0 is found as the solution to equation (22) provided by expression (29). Thus, to find the power, it suffices to find the delta-function (32). After some algebra and using (28), we obtain

$$\delta(\beta) = E_{(\gamma,\beta)}\frac{\ln f(y,x;0,\gamma_0)}{\partial\beta} = \frac{GB}{1+GB} - \frac{G_0}{1+G_0}p_x\,. \tag{35}$$

The variance function, V was derived earlier and is given by (18); the variance at null is

$$V(0\mid\gamma_0) = \frac{(1+G_0)^2}{p_x(1-p_x)G_0}\,.$$

Unlike for linear regression, the score test is asymptotically biased for logistic regression. Namely, the first derivative at $\beta = 0$ given by (33) is not zero. We shall compare this test to the other two tests in Section 6.

6. Tests comparison

In this section, we compare the Wald, LR, and score tests using two regression models as examples: linear regression with an unknown σ^2 under normal distribution and logistic regression with binary covariate.

We start our analysis with linear regression. It is convenient to compare the tests on the z-scale using representations (13), (27) and (34). From (14), it is easy to see that for each test the second derivatives on the original β-scale and the z-scale are related through a coefficient, which is the reciprocal of the variance at the null,

$$\left.\frac{d^2P}{d\beta^2}\right|_{\beta=0} = \frac{1}{V(0)}\left.\frac{d^2P}{dz^2}\right|_{z=0}\,.$$

Theorem 2.

All three tests are asymptotically unbiased for linear regression under a normal distribution with an unknown σ^2 (the first derivative of the power function evaluated at the null vanishes). The power function of the Wald test is uniformly greater than that of the LR test ($\beta \neq 0$). The second derivatives of the three tests at the null are as follows:

$$\left.\frac{d^2P_W}{dz^2}\right|_{z=0} = \left.\frac{d^2P_{LR}}{dz^2}\right|_{z=0} = 2\phi(Z_{1-\alpha/2})Z_{1-\alpha/2}n, \tag{36}$$

$$\left.\frac{d^2P_S}{dz^2}\right|_{z=0} = 2\phi(Z_{1-\alpha/2})Z_{1-\alpha/2}(n+1)\,. \tag{37}$$

Thus, the score test is slightly superior in a close neighborhood of the null.

Proof.

The asymptotic unbiasedeness of the tests (the first derivatives vanish at zero) have been noted previously. The superiority of the Wald over the LR test follows from an elementary inequality $\sqrt{\ln(1+z^2)} < |z|$ for all $z \neq 0$. The second derivative of the Wald test at $\beta = 0$ has been derived earlier (12). The second derivative of the LR test follows from (34). The evaluation of the second derivative of the score test (37) is somewhat tedious but straightforward. A slight superiority of the score test in the neighborhood of the null follows from the fact that the difference of the derivatives is $2\phi(Z_{1-\alpha/2})Z_{1-\alpha/2}$ although this difference rapidly diminishes on the relative scale as $n \to \infty$.■

We illustrate the power approximation of the three tests in Figure 3 with linear regression $y = \beta x + \tau u + \varepsilon$ where all random variables have normal distributions with zero mean, $var(x) = \sigma_x^2$, $var(u) = \sigma_u^2$, $var(\varepsilon) = \sigma^2$ The power functions of the Wald, LR, and score tests are plotted as functions of the z-ratio (14) with $n = 100$, $\sigma_x = \sigma = 1$, $\rho_{xu} = 0.9$ and $\tau = 1$, assuming that the significance level $|z| = 5\%$. Notice that the power functions are very close in the interval $|z| < 0.2$ with the maximum difference around $z = 0.5$.

Now we compare the three tests for logistic regression with a binary covariate. See Figure 4 as an example with $\alpha = 0.05$, $p_x = \Pr(x = 1) = 0.2$, $\Pr(y = 1|x = 0) = 0.1$ ($G = 1/9$), and $n = 500$. We intentionally use a rare parameter set up ($p_x = 0.2$) because at symmetry ($p_x = 0.5$ and $G = 1$) the power functions are practically identical. Recall the Wald and LR power functions have zero derivatives and the same second derivative at $\beta = 0$. However, the first derivative of the score test is positive at $\beta = 0$ meaning that the power of the score test is slightly less in the neighborhood to the left of the null. Recall that the Wald power falls to α when $\beta \to -\infty$ or $\beta \to \infty$ which is not true for the LR and score tests. Although the LR power does not approach 1 when $\beta \to -\infty$ the limit is very close to 1. In order to test the power approximations, we conducted simulations for $\beta = -1$ and $\beta = 0.5$ (the number of experiments is 5,000)–simulations and power approximations are in good agreement.

7. Discussion and summary points

Asymptotic tests are at the heart of statistical inference. Early authors studied which test is superior. For example, Madansky (1989) raised this question, and in early edition of Rao's book it was conjectured that the score test is locally more powerful than the other two. (There is no such conjecture in the latest, 1973 edition of the book!) Yes, as follows from Theorem 2, there exists a slight superiority of the score test for linear regression in a close proximity to the null. However, for other statistical models this superiority disappears-moreover, the score test is generally biased (the first derivative is not zero) which means that for a double-sided test the score power will be less on one side of the null.

The main points of the paper are as follows:

- The Wald test may exhibit aberrant behavior for some statistical models, namely, the power increases and then drops to the size of the test as the alternative approaches negative infinity. For logistic regression with a binary covariate, the

Wald test behaves well for alternatives in the interval (−2, 2). The chance that the likelihood ratio or score test have aberrant behavior is smaller, but still exists.

- The Wald and LR tests are asymptotically unbiased (the first derivative of their power function vanishes at the null), but the score test is generally biased. The Wald and LR tests are locally equivalent (the second derivatives are the same at the null), but the score test is not.
- It is well known that all three tests are the same for a linear model under the normal distribution when the error variance, σ^2 is known. However, when σ^2 is unknown (nuisance parameter) the tests are different. The power of the Wald test is higher than that of the LR test for all alternatives, but locally the tests are equivalent. The score test for this model is also asymptotically unbiased and has slightly higher power in close proximity to the null.
- The behavior of the tests changes from distribution to distribution and from one statistical model to another. Even members of the same family of generalized linear model (linear and logistic) have quite different test properties. Especially drastic are the differences for low probability and/or larger true values of nuisance parameters.
- Simulations show a good agreement with power approximations especially for moderate parameter values.
- No test is superior over another in the whole range of statistical models and distributions. For example, tests are close for moderate parameter values, particularly when p_x is around 1/2. However, tests may be very different with extreme parameter values, such as for rare events in logistic regression ($p_x \simeq 0$). Thus, different tests may be optimal for different statistical models, moreover for different ranges of the alternative.

Besides theoretical interest, the power analysis has an important application for study design, particularly in epidemiology and clinical trials. Usually, one wants to determine the sample size required to achieve a specified power or minimum detectable difference (the alternative beta value). Several commercial and noncommercial software packages are available on the market. However, there is no consensus on what test to use as the basis for the sample size determination–some authors use the Wald test (Whittemore, 1981), some the likelihood ratio test (Self et al., 1992 and Shieh, 2000) and some the score test (Self and Mautitzen, 1988) mainly because of the popular opinion that asymptotic tests are all the same. True, when alternatives are close to null and nuisance parameters do not take extreme values, the powers look alike. However, in some cases, we plan studies for large alternatives such as in epidemiologic studies with gene—gene or gene—environment interaction or observational studies of rare diseases (cancer), see Duchateau (1998) and Gauderman (2002) for examples. In such studies, the choice of the test becomes crucial. The test used for the sample size determination should be the same as the test in future significance testing (Demidenko 2007, 2008).

Much additional work is to be done in the following areas: (1) Extend our power analysis to one-sided tests and the multivariate null hypothesis; (2) identify what asymptotic tests are

optimal for specific statistical models, such as members of generalized linear model family, and (3) improve power approximations for extreme values of alternative and nuisance parameters, particularly when the effective sample size is small.

Finally, the study of the tests in this paper, as well as in many others, relies on standard conditions on the density function, such as independence of the density support on the parameter. In my private conversation with S.A. Aivazian, he told me that one of the problems he considered in his doctoral dissertation, was studying superefficient estimation and the respective statistical inference when the support of the distribution dependens on the unknown parameter, like in the uniform distribution with unknown upper limit.

8.: Appendix.

Proof of Theorem 1

The fact that the first derivative of the Wald power vanishes at $\beta = 0$ was proven in Section 3. For the LR power, we have

$$\frac{dP_{LR}}{d\beta} = \sqrt{\frac{n}{2}}\left[\phi\left(-Z_{1-\alpha/2} + \sqrt{2n\eta(\beta)}\right) - \phi\left(Z_{1-\alpha/2} + \sqrt{2n\eta(\beta)}\right)\right]\frac{d\eta/d\beta}{\sqrt{\eta(\beta)}}.$$

This implies that the LR test is asymptotically unbiased (9) because $\eta(0) = 0$ (we show below that $d\eta/d\beta/\sqrt{\eta(\beta)}$ is a finite number at zero).

Now we evaluate the second derivative of the power function at zero. Using the property that $d\phi/dx = -x\phi(x)$, we obtain

$$\left.\frac{d^2P_{LR}}{d\beta^2}\right|_{\beta=0} = n\phi\left(Z_{1-\alpha/2}\right)Z_{1-\alpha/2}\left.\frac{(d\eta/d\beta)^2}{\eta(\beta)}\right|_{\beta=0}. \tag{38}$$

Since the numerator and denominator with the η-function are both zero we need to consider the limit. From L'Hopital's rule

$$\lim_{\beta\to 0}\frac{(d\eta/d\beta)^2}{\eta(\beta)} = 2\lim_{\beta\to 0}\frac{(d\eta/d\beta)\left(d^2\eta/d\beta^2\right)}{d\eta/d\beta} = 2\left.\frac{d^2\eta}{d\beta^2}\right|_{\beta=0},$$

so that (38) is rewritten as

$$\left.\frac{d^2P_{LR}}{d\beta^2}\right|_{\beta=0} = 2n\phi(Z_{1-\alpha/2})Z_{1-\alpha/2}\left.\frac{d^2\eta}{d\beta^2}\right|_{\beta=0}.$$

Thus, the evaluation of the second derivative of the power reduces to the evaluation of the second derivative of the η-function. We aim to find the second derivative of η at zero. Interchanging expectation and differentiation from (23), we obtain $\frac{d}{d\beta}E_{(\beta,\gamma)}$. This implies

$$\frac{d\eta}{d\beta} = -\frac{d}{d\beta}E_{(\beta,\gamma)}(\ln f(\mathbf{z};0,\gamma_0)) = -\left(\frac{\partial}{\partial\gamma_0}E_{(\beta,\gamma)}(\ln f(\mathbf{z};0,\gamma_0))\right)'\left(\frac{d\gamma_0}{d\beta}\right).$$

From differentiation of an implicit function, we have

$$\frac{d\gamma_0}{d\beta} = -\left[\frac{\partial}{\partial\gamma}E_{(\beta,\gamma)}\left(\frac{\partial\ln f(\mathbf{z};0,\gamma)}{\partial\gamma}\right)\right]^{-1}\left[\frac{d}{d\beta}E_{(\beta,\gamma)}\left(\frac{\partial\ln f(\mathbf{z};0,\gamma)}{\partial\gamma}\right)\right],$$

but

$$\frac{d}{d\beta}E_{(\beta,\gamma)}\left(\frac{\partial\ln f(\mathbf{z};0,\gamma)}{\partial\gamma}\right) = \mathbf{I}_{12}, \quad \frac{\partial}{\partial\gamma}E_{(\beta,\gamma)}\left(\frac{\partial\ln f(\mathbf{z};0,\gamma)}{\partial\gamma}\right) = \mathbf{I}_{22},$$

which results in $d\gamma_0/d\beta = -\mathbf{I}_{22}^{-1}\mathbf{I}_{12}$. Finally, the second derivative of the η-function at zero is

$$\begin{aligned}&\frac{d^2}{d\beta^2}E_{(\beta,\gamma)}(\ln f(\mathbf{z};\beta,\gamma)) - \left(\frac{d\gamma_0}{d\beta}\right)'\left(\frac{\partial^2}{\partial\gamma_0}E_{(\beta,\gamma)}(\ln f(\mathbf{z};0,\gamma_0))\right)\left(\frac{d\gamma_0}{d\beta}\bigg|_{\beta=0}\right)\\ &= I_{11} - \mathbf{I}_{12}'\mathbf{I}_{22}^{-1}\mathbf{I}_{22}\mathbf{I}_{22}^{-1}\mathbf{I}_{12} = I_{11} - \mathbf{I}_{12}'\mathbf{I}_{22}^{-1}\mathbf{I}_{12}.\end{aligned}$$

Combining the results, we arrive at the second derivative at the null

$$\frac{d^2P_{LR}}{d\beta^2}\bigg|_{\beta=0} = 2n\phi(Z_{1-\alpha/2})Z_{1-\alpha/2}\left(I_{11} - \mathbf{I}_{12}'\mathbf{I}_{22}^{-1}\mathbf{I}_{12}\right),$$

which coincides with the second derivative of the Wald test (12). This means that the Wald and LR tests are locally equivalent.

References

Aivazian SA, Yenyukov IS, Meshalkin LD (1985). Applied Statistics, vol. 2 Investigation of Dependencies (in Russian). Moscow: Finasi i Statistika.

Bickel PJ and Doksum KA (2001). Mathematical Statistics Basic Ideas and Selected Topics. Upper Saddle River, NJ: Prentice Hall.

Casella G and Berger RL (1990). Statistical Inference. Belmont, CA: Duxbury Press.

Chandra TK and Joshi SN (1983). Comparison of the likelihood ratio, Rao's and Wald's tests and a conjecture of C.R. Rao. Sankhyā Ser. A 45:226–246.

Chandra TK and Mukerjee R (1984). On the optimality of the Rao's statistic. Communications in Statistics. Theory and Methods 13:1507–1515.

Cohen J (1994). The earth is round ($p < .05$). American Psychologist 49:997–1003.

Cox DR and Hinkley DV (1974). Theoretical Statistics. Boca Raton, FL: Chapman & Hall.

Demidenko E (2013). Mixed Models: Theory and Applications. 2nd. ed. Hoboken, NJ: Wiley.

Demidenko E (2020). Advanced Statistics with Applications in R. Hoboken, NJ: Wiley www.eugened.org

Demidenko E (2007). Sample size determination for logistic regression revisited. Statistics in Medicine 26:3385–3394. [PubMed: 17149799]

Demidenko E (2008). Sample size and optimal design for logistic regression with binary interaction. Statistics in Medicine 27:36–46. [PubMed: 17634969]

Duchateau L, McDermott B, and Rowlands GR (1998). Power evaluation of small drug and vaccine experiments with binary outcomes. Statistics in Medicine 17:111–120. [PubMed: 9463854]

Fears TR, Benicou J, and Gail MH (1996). A reminder of the fallibility of the Wald statistics, The American Statistician, 50:226–227.

Gauderman WJ (2002). Sample size requirements for association studies of gene-gene interaction. American Journal of Epidemiology 155:478–484. [PubMed: 11867360]

Hauck WW and Donner A (1977). Wald's test as applied to hypotheses in logit analysis. Journal of American Statistical Association 72:851–853.

Lehmann EL and Romano JP (2005). Testing Statistical Hypotheses, 3rd edition New York: Springer.

Madansky A (1989). A comparison of the likelihood ratio, Wald, and Rao tests In: Contributions to Probability and Statistics, (Eds: Gleser LJ et al.). New York: Springer-Verlag.

McCullagh P (1987). Tensor Methods in Statistics. London: Chapman & Hall.

Pawitan Y (2001). In All Likelihood. Statistical Modeling and Inference Using Likelihood. Oxford: Clarendon Press.

Rao CR (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. Proc. Camb. Phil. Soc 44:50–57.

Rao CR (1973). Linear Statistical Inference and Its Applications. 2nd ed New York: Wiley.

Rayner JC (1997). The asymptotically optimal tests, The Statistician 46, 337–346.

Schervish MJ (1995). Theory of Statistics. New York: Springer–Verlag.

Self G and Mautitzen RH (1988). Power/sample size calculations for generalized linear models. Biometrics 44:79–86.

Self G, Mautitzen RH, and Ohara J (1992). Power calculations for likelihood ratio tests in generalized linear models. Biometrics 48:31–39.

Serfling RJ (1980) Approximation Theorems of Mathematical Statistics. New York: Wiley.

Severeni TA (2000). Likelihood Methods in Statistics. Oxford: Oxford University Press.

Shieh G (2000). On power and sample size calculations for likelihood ratio tests in generalized linear models. Biometrics 56:1192–1196. [PubMed: 11129478]

Væth M (1985). On the use of Wald's test in exponential families. International Statistical Review 53:199–214.

Whittemore AS (1981). Sample size for logistic regression with small response probability. Journal of the American Statistical Association 76:27–32.

Young GA and Smith RL (2005). Essentials of Statistical Inference. Cambridge: Cambridge University Press.

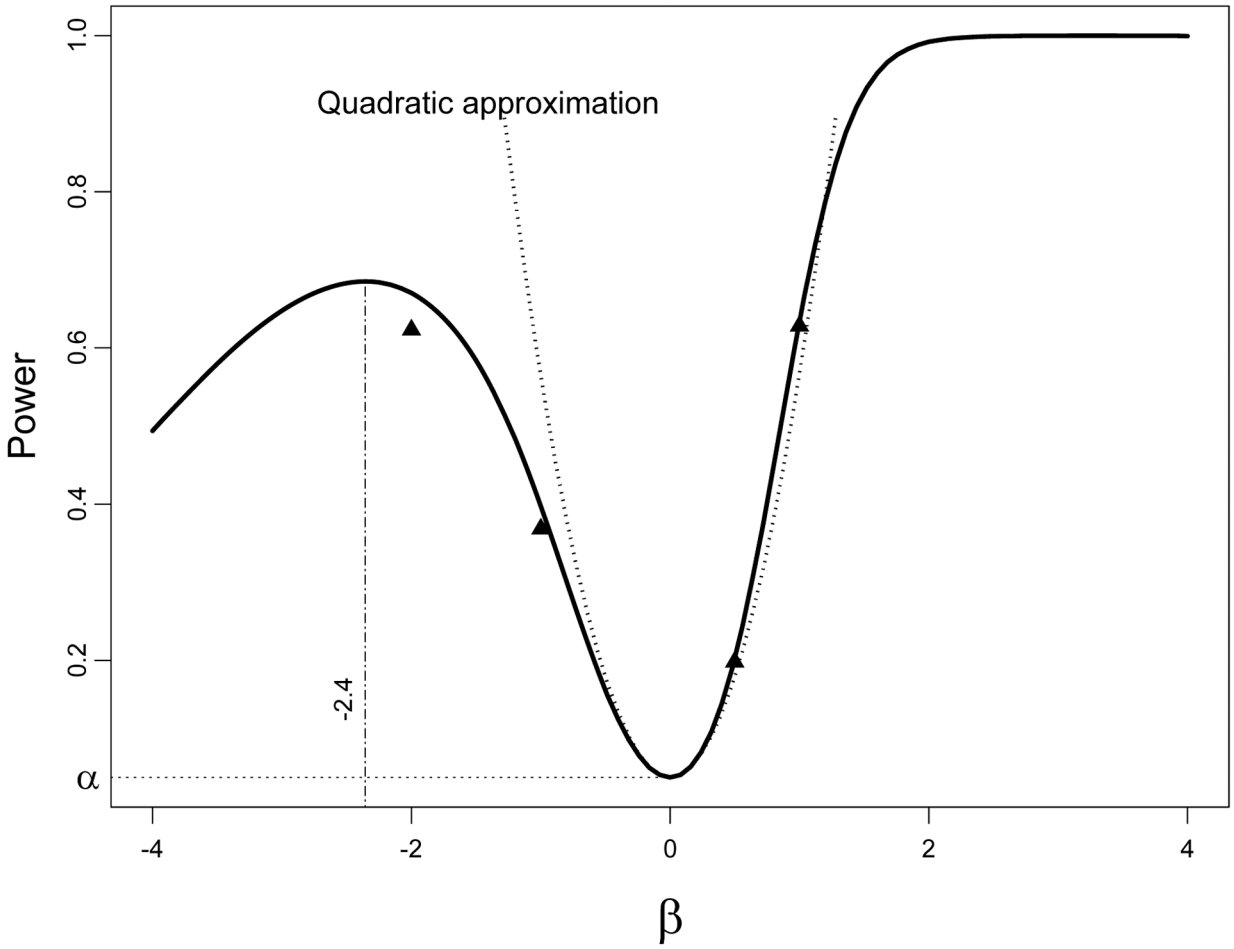

Figure 1.

The Wald power and its quadratic approximation for logistic regression with a binary covariate, $\alpha = 0.05$, $\Pr(y = 1|\beta = 0) = 0.25$, $\Pr(x = 1) = 0.2$, and $n = 100$. The Wald power falls at $\beta = -2.4$. Simulations (▲) are good match of the power approximation especially for positive alternatives (N_{exp}=5,000).

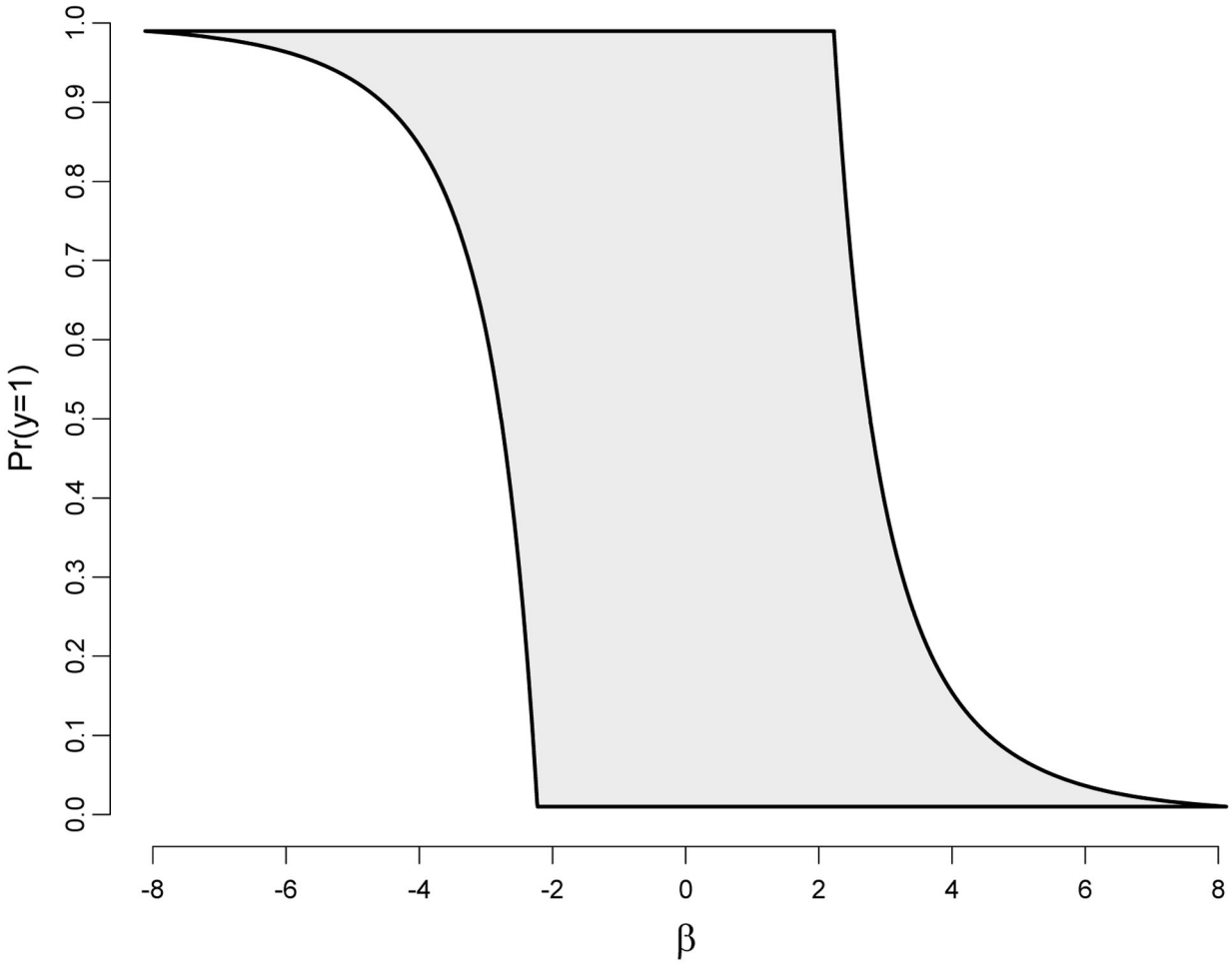

Figure 2.

The power function is increasing for positive alternative and decreasing for negative alternative for (β, p_x) within the gray region. Approximately, the Wald power is an increasing function of the distance from the null if $|\beta| \leq 2$.

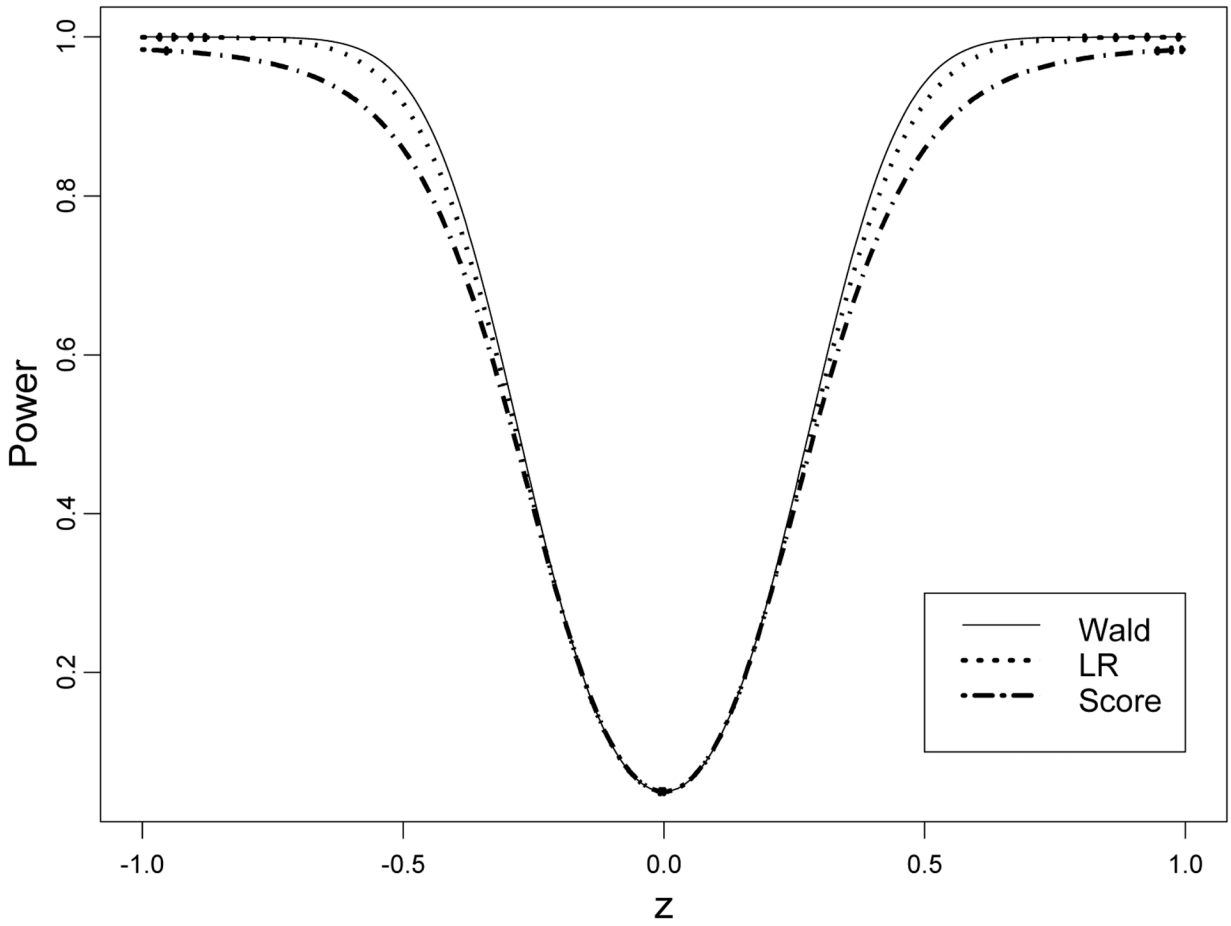

Figure 3.

Power of three tests for linear regression with unknown σ^2 on the z-scale ($n = 100$, $\sigma_x = \sigma = 1$, $\rho_{xu} = 0.9$, $\tau = 1$ with $\alpha = 0.05$). All three tests have the same size (α) and the first derivative vanishes at $\beta = 0$. The Wald and likelihood ratio tests have the same second derivatives at the null, but the second derivative of the score test is higher (actually invisible). However, globally, the Wald test is superior.

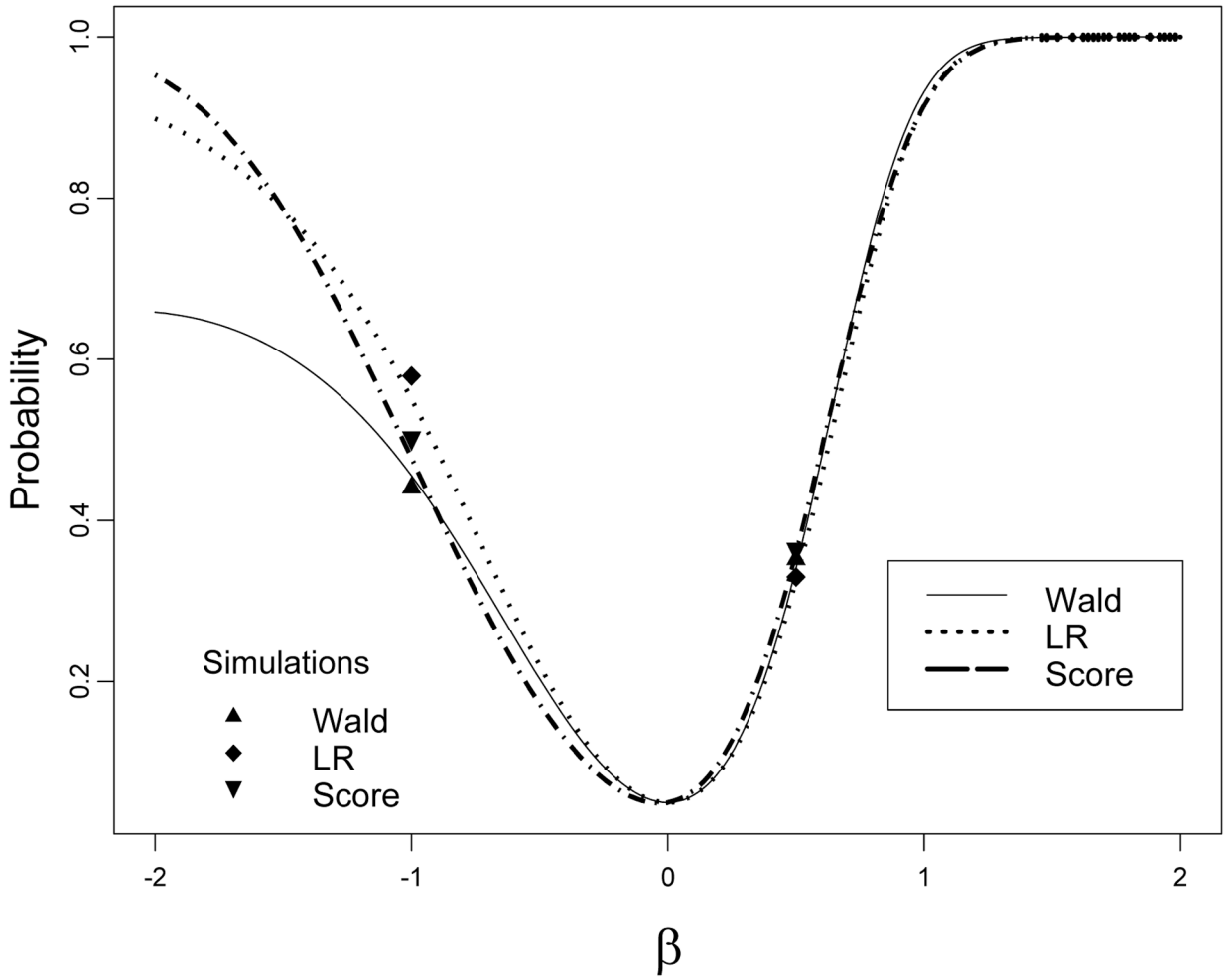

Figure 4.

Three power functions for logistic regression with a binary covariate, $p_X = 0.2$, $G = 1/9$, $n = 500$ with $\alpha = 5\%$. The first derivative of the Wald and likelihood ratio tests vanishes, but the derivative of the score test power is positive at $\beta = 0$. Simulations for $\beta = -1$ and 0.5 are in good agreement with the power approximations.