

On the use of saddlepoint approximations
in highdimensional inference

Jens Ledet Jensen



On the use of saddlepoint approximations in highdimensional inference

Jens Ledet Jensen

Department of Mathematics, Aarhus University, Denmark, jlj@math.au.dk

Abstract

Inference in high dimensional parameter space poses many challenges. One of these is the possible use of saddlepoint approximations. Motivated by a recent use of the saddlepoint approximation to construct a conditional test, we argue that the precision is questionable. We illustrate this by an example giving a 50% relative error in the calculation of the p -value. A power study of the underlying test reveals a low power in many situations. As an alternative it is suggested to use the likelihood ratio test.

1 Introduction

In small sample asymptotics many procedures are based on the saddlepoint approximation to a density either directly or indirectly. This was recently taken a step further in Davison et al. (2014), where inference for a vector-valued parameter is considered. The essence of that paper is to use a saddlepoint approximation, reduce to a conditional test and then use numerical integration to obtain the p -value. The method is successfully applied to a number of cases, including a high dimensional setting as well, leading to the conclusion that the proposed approach is “extremely accurate ... also in high-dimensional settings”. With this note I want to point to the fact that inference based on the saddlepoint approximation in high dimensional settings should *not* be expected to work. I will explain why this is so, explain why the approximation did work in the example of Davison et al. (2014) and make a small change of the example to illustrate that the method can fail. Furthermore, I will discuss different aspects of power and suggest a general approximation to the distribution of the likelihood ratio statistic.

2 The saddlepoint approximation in high dimensional settings

The saddlepoint approximation to the density $f_n(\cdot)$ of $\bar{y} = \frac{1}{n}(y_1 + \dots + y_n) \in \mathbf{R}^m$ takes the form

$$f_n(y) = \varphi(\hat{\theta})^n e^{-n\hat{\theta} \cdot y} (2\pi)^{-m/2} |\Sigma(\hat{\theta})|^{-1/2} (1 + O(\frac{1}{n})),$$

where $\varphi(\theta) = Ee^{\theta \cdot y}$ is the moment generating function, $\Sigma(\theta)$ is the second derivative of $\kappa(\theta) = \log \varphi(\theta)$ and the saddlepoint $\hat{\theta}$ is determined by $(\partial \kappa / \partial \theta)(\hat{\theta}) = y$ (Jensen (1995)). The relative error $O(\frac{1}{n})$ can be expanded as $h(y)/n + O(\frac{1}{n^2})$. Consider now the situation where $y \in \mathbf{R}^m$, where m is large, and where all coordinates are independent. In this situation the saddlepoint approximation becomes the product of the saddlepoint approximations for the individual coordinates:

$$f_n(y) = \left(\prod_{j=1}^m \frac{\varphi_j(\hat{\theta}_j)^n e^{-n\hat{\theta}_j y_j}}{\sqrt{2\pi \Sigma_j(\hat{\theta}_j)}} \right) \prod_{j=1}^m \left(1 + \frac{h_j(y_j)}{n} + O\left(\frac{1}{n^2}\right) \right), \quad (2.1)$$

where φ_j , θ_j and Σ_j refer to the j 'th coordinate of y . It is clear that the last product will generally not be $1 + O(\frac{1}{n})$ when m is large. *When m is of order n or larger, the saddlepoint approximation cannot be trusted.* Figure 1 shows an example where all the coordinates have the same distribution with density given by (3.1) below. On the left is the difference of the logarithm of the saddlepoint approximation and the logarithm of the exact density for one coordinate, where the saddlepoint approximation is normalized to integrate to one when $\theta = 1$. On the right is the difference for the full density with $m = 1000$, that is, the logarithm of the last product of (2.1), for randomly generated points with $\theta = 0.1$ (and with the saddlepoint approximation normalized to integrate to one when $\theta = 0.1$). The relative error is less than five percent for one variable, but for the full density with 1000 variables the saddlepoint approximation can be off by a factor 4. This makes it impossible to compare the density at two randomly generated points via the saddlepoint approximation. The actual setting for the considered example is described in the next section.

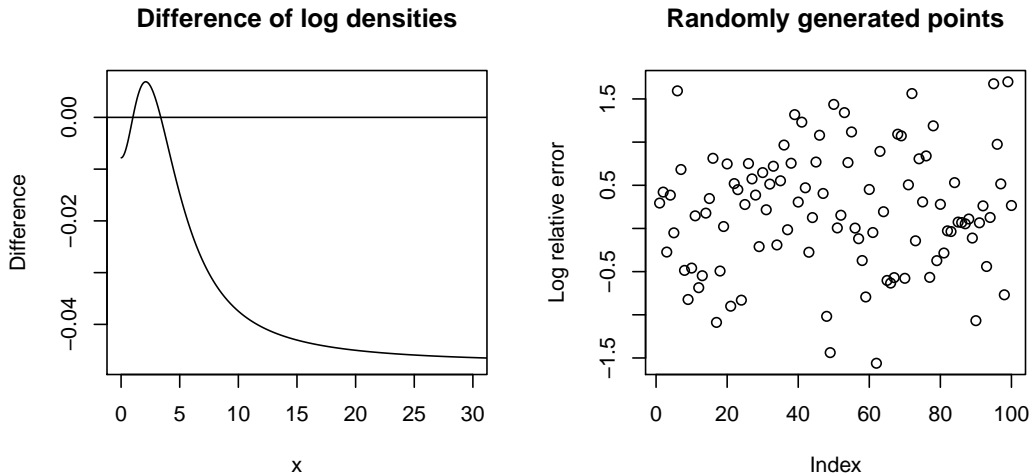


Figure 1: On the left is the error of the saddlepoint approximation to the logarithm of the density (3.1). On the right is the accumulated error for 1000 variables at 100 randomly generated points with $\theta = 0.1$ in the density (3.1).

3 Example

We consider a modification of the example in section 5.2 of Davison et al. (2014). We have $m = 1000$ independent variables (called *groups* in Davison et al. (2014))

and in the j 'th group we have an observation from the density

$$\theta_j(1 + \theta_j)(1 - e^{-y_j})e^{-\theta_j y_j}. \quad (3.1)$$

Davison et al. (2014) used instead the Gamma density $\theta_j^5 y_j^4 e^{-\theta_j y_j} / 24$. The aim is to test the hypothesis $\theta_1 = \theta_2 = \dots = \theta_m$. Under the hypothesis $\bar{y} = \frac{1}{m}(y_1 + \dots + y_m)$ is sufficient and Davison et al. (2014) consider the conditional distribution given \bar{y} . Considering $(y_1 - \bar{y}, \dots, y_{m-1} - \bar{y})$ conditioned on \bar{y} the authors next consider the conditional distribution given the direction. They formalize this by looking at points $(z_1, \dots, z_{m-1}) = (\bar{y}, \dots, \bar{y}) + t(y_1 - \bar{y}, \dots, y_{m-1} - \bar{y})$ as a function of t , which introduces the Jacobian t^{p-2} in the density. In the conditional distribution of t the p -value is the probability of getting a value above $t = 1$. Denoting the saddlepoint density by f_{sp} we end up with

$$p\text{-value} \approx \frac{\int_1^{t_0} g_{\text{sp}}(t) dt}{\int_0^{t_0} g_{\text{sp}}(t) dt} = \int_1^{t_0} \bar{g}_{\text{sp}}(t) dt, \quad (3.2)$$

where

$$g_{\text{sp}}(t) = f_{\text{sp}}(\bar{y} + t(y_1 - \bar{y}), \dots, \bar{y} + t(y_m - \bar{y}))t^{p-2}, \quad (3.3)$$

and t_0 is the maximal allowed value of t along the chosen direction, that is, $t_0 = \bar{y} / (\bar{y} - \min\{y_i\})$, corresponding to the restriction that all variables must be positive. The function $\bar{g}_{\text{sp}}(t)$ is $g_{\text{sp}}(t)$ normalized to integrate to one.

For the density (3.1) the saddlepoint approximation reduces to

$$1 - e^{-x} \approx \frac{e^{\hat{\theta}x}}{\hat{\theta}(1 + \hat{\theta})\sqrt{2\pi\{1/\hat{\theta}^2 + 1/(1 + \hat{\theta})^2\}}}, \quad \hat{\theta} = \frac{2 - x + \sqrt{4 + x^2}}{2x}.$$

The difference on a log scale of the left and right hand side (the latter renormalized) is shown in the left part of figure 1. The right part of figure 1 shows the relative error for the multivariate saddlepoint approximation at randomly generated points. However, this does not translate directly into the relative error of $g_{\text{sp}}(t)$ for various values of t , since only the point $t = 1$ corresponds to a random point. The left part of figure 2 shows an example of the deviation of $\bar{g}_{\text{sp}}(t)$ from the exact density on a log scale for a random sample generated with $\theta = 0.1$. For $t = 1$ the relative error is almost zero, but for $t = 0$ the relative error is so large that the approximation cannot be used. However, the figure also displays the logarithm of the exact density, showing that the region where the relative error is large has very small probability. The p -value, which is the probability for $t > 1$, is therefore well approximated with the exact value being 0.081 and the approximation being 0.077.

The right part of figure 2 shows an example where data is generated with half of the variables having $\theta = 0.1$ and the other half having $\theta = 0.01$. In this case there is a fairly large variation in the relative error in the range where the exact density is high. For this example the exact p -value is 0.020 and the approximation is 0.010 giving a 50% relative error.

The example considered here shows that typically there is a huge variation in the last product of (2.1) when considered a function of t , $0 < t < t_0$. However,

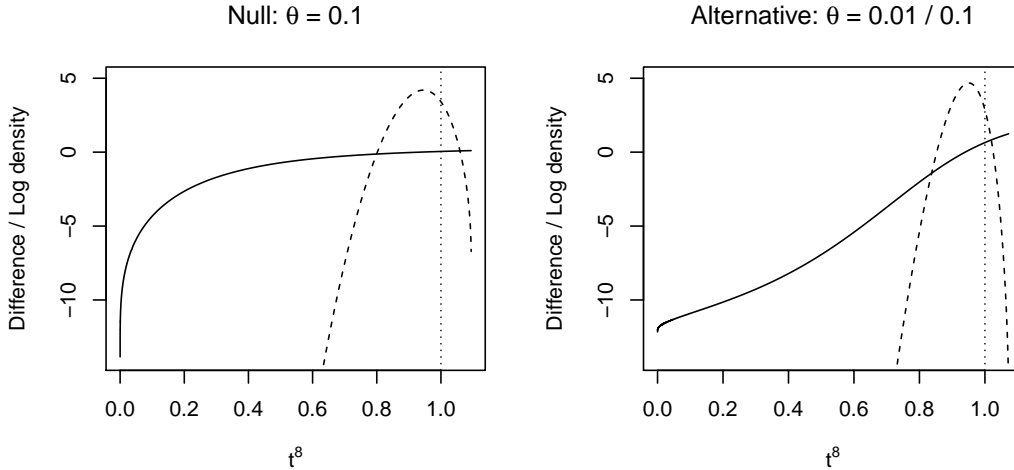


Figure 2: The full drawn curve displays the difference on a log scale between the exact density and the normalized saddlepoint approximation $\bar{g}_{\text{sp}}(t)$. The dashed line is the logarithm of the exact density, and the dotted vertical line highlights $t = 1$. For visual effect t^8 is used instead of t for the abscisse. On the left is a situation with data generated under the null hypothesis with 1000 variables having $\theta = 0.1$ and on the right a situation with 200 variables having $\theta = 0.1$ and 500 variables having $\theta = 0.01$.

when restricted to the region with a high density the variation becomes smaller although for the examples considered the relative error of the p -value can still be up to 50%. It seems likely that other examples can be constructed leading to even larger relative errors. For the example used in this section the relative error of the saddlepoint approximation for one variable is small all over the range of the variable. Using instead an example where the relative error for one variable is unbounded, when approaching the boundary of the support of the variable, larger errors for the multivariate case are to be expected.

It is well known that the relative error is constant for the saddlepoint approximation to the density of a Gamma distribution (the approximation reduces to Stirling's approximation to the Gamma function). When all variables have a Gamma distribution the last product in (2.1) becomes a constant and therefore disappears after a renormalization. The approximation $\bar{g}_{\text{sp}}(t)$ used to calculate the p -value therefore becomes exact. So the p -value in section 5.2 of Davison et al. (2014) is not only a good approximation, but actually exact.

4 Power

In the previous section I illustrated that a priori one cannot trust the approximative conditional p -values introduced by Davison et al. (2014) when used in very high dimensional settings. Setting aside the approximation issue I want next to consider the approach via a conditional distribution in terms of the power of the test. Continuing with the example of section 3, the power under various alternatives is found from simulations based on the exact conditional p -value given the direction of $(y_1 - \bar{y}, \dots, y_{m-1} - \bar{y})$, and given \bar{y} . For a shorthand notation we denote this the conditional test, and compare below with a test we denote the "marginal" test

although we consider the conditional distribution given \bar{y} .

To evaluate the performance of the conditional test for $\theta_1 = \dots = \theta_m$ in terms of power we compare with a simple natural alternative test, namely the test that rejects large values of $\sum_{i=1}^m (y_i - \bar{y})^2/m$. We can formulate this as a comparison of the variance of a mixture distribution with the variance of a non-mixture. Let $\kappa_j(\theta)$ be the j 'th cumulant and consider the case where $\theta_1, \dots, \theta_m$ consist of k different values $\theta^1, \dots, \theta^k$ appearing in the mixture proportions $\alpha_1, \dots, \alpha_k$. The variance of the mixture is

$$\sum_{j=1}^k \alpha_j \kappa_2(\theta^j) + \sum_{j=1}^k \alpha_j (\kappa_1(\theta^j) - \mu)^2, \quad \mu = \sum_{j=1}^k \alpha_j \kappa_1(\theta^j), \quad (4.1)$$

which is to be compared with the variance for the case $\theta_1 = \dots = \theta_m = \theta_\mu$ with $\kappa_1(\theta_\mu) = \mu$, the latter variance being $\kappa_2(\theta_\mu)$. The last term of the first expression in (4.1) is clearly positive. If therefore $\sum_j \alpha_j \kappa''(\theta^j) > \kappa_2(\theta_\mu)$ we have that the mixture variance (4.1) is bigger than the non-mixture variance $\kappa_2(\theta_\mu)$, making the use of $\sum_i (y_i - \bar{y})^2/m$ as a test statistic reasonable. Letting $\mu^j = \kappa_1(\theta^j)$ we ask whether

$$\sum_{j=1}^k \alpha_j \kappa_2((\kappa_1)^{-1}(\mu^j)) > \kappa_2\left((\kappa_1)^{-1}\left(\sum_{j=1}^k \alpha_j \mu^j\right)\right) \quad (4.2)$$

holds, that is, we ask whether the function $\kappa_2((\kappa_1)^{-1}(\mu))$ is convex. The second derivative of this function is

$$\kappa_4(\theta)/\kappa_2(\theta)^2 - \kappa_3(\theta)^2/\kappa_2(\theta)^3 \quad \text{with } \kappa_1(\theta) = \mu.$$

A simple calculation with the cumulant transform $\kappa(s) = \log(\theta(1+\theta)/\{(\theta-s) \cdot (1+\theta-s)\})$ shows that the above second derivative is positive and (4.2) holds.

Using the conditional distribution of $\sum_i (y_i - \bar{y})^2$ given \bar{y} is equivalent to considering the conditional distribution of $\sum_i y_i^2$ given \bar{y} . It is possible to establish a saddlepoint approximation for this conditional distribution, but since m is large a simple normal approximation will suffice. The normal approximation is found by calculating the moments of (Y, Y^2) and replacing θ by $\hat{\theta} = (2 - \bar{y} + \sqrt{4 + \bar{y}^2})/(2\bar{y})$. This gives

$$\frac{1}{p} \sum_{i=1}^p y_i^2 | \bar{y} \sim N(\hat{\kappa}_2 - \hat{\kappa}_1^2, \hat{\kappa}_4 + 2\hat{\kappa}_2^2 - \hat{\kappa}_3^2/\hat{\kappa}_2) \quad (4.3)$$

with $\hat{\kappa}_j = a_j(1/\hat{\theta}^j + 1/(1+\hat{\theta})^j)$, $a_1 = a_2 = 1$, $a_3 = 2$ and $a_4 = 6$.

In a first test we consider a situation with 100 variables having the value $\theta_1 = 1$ and 900 variables having the value $\theta_2 = 2$. The result of the simulations can be seen in Table 1. The conditional test of Davison et al. (2014) has power 0.11 whereas the marginal test based on (4.3) has a much larger power of 0.74. This is a remarkable large difference that calls for an explanation. Looking at the left part of Figure 3 we see that there is not much relation between the conditional test and the marginal test. The right part of Figure 3 shows a strong correlation between the conditional test and $\min_i \{y_i\}$, in that a small value of the minimum correlates with a small value of the p -value. Actually, a plot based on simulated values under the null hypothesis $\theta_1 =$

$\dots = \theta_m$ shows the same strong correlation between the p -value of the conditional test and the minimum of the observations. A small value of the minimum implies a small value of the upper point t_0 of the range of t in the conditional test.

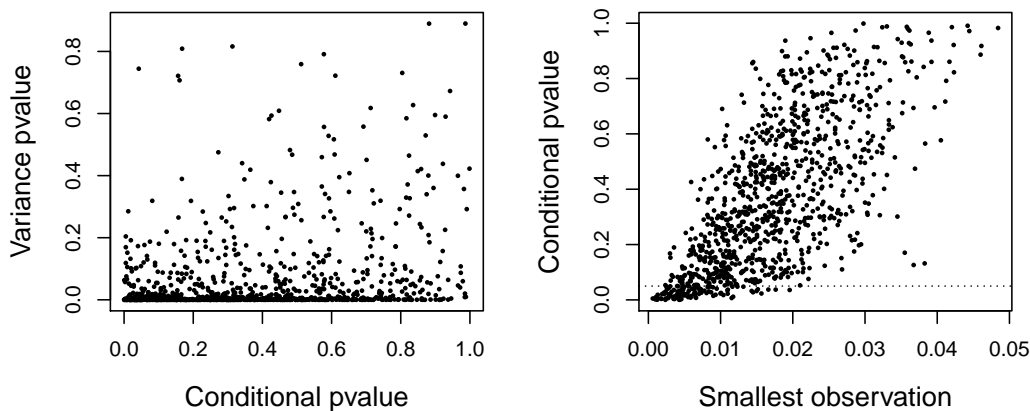


Figure 3: Result of 1000 simulated p -values based on either the conditional test or the marginal test (variance p -value). The right plot shows the relation between the conditional p -value and the minimum of the m variables.

If, generally, the conditional test rejects mainly when $\min_i\{y_i\}$ is small, the test should have very low power for an alternative that does not produce small values of the minimum. We therefore consider a situation with 999 variables having the value $\theta_1 = 1$ and one variable having the value $\theta_2 = 0.1$. The probability that the observation from $\theta = 0.1$ is larger than the 95% quantile of the maximum of 999 observations from $\theta = 1$ is 0.38. This sets an upper limit for the power we can expect to find. Table 1 shows that the power of the conditional test is 0.06, that is, not much bigger than the level of the test. Contrary to this, the marginal test has power 0.33. Indeed, a plot reveals (not shown) that there is no relation between the largest observation and the conditional p -value.

Reversing the above example by having 999 variables with the value $\theta_1 = 1$ and one variable having the value $\theta_2 = 100$ the conditional test finds the cases with a small value of the variable with $\theta = 100$. The probability that the observation from $\theta = 100$ is smaller than the 95% quantile of the minimum of 999 observations from $\theta = 1$ is 0.16. This fits well with the power for the conditional test seen in Table 1. In this case the marginal test based on $\sum_i(y_i - \bar{y})^2$ has almost no power, since one observation close to zero gives only a very small increase of the variance.

Table 1: Simulated power for two tests based on 10000 simulated cases. In all three settings the variables fall in two groups according to the value of θ , with the number of variables in the group given in parenthesis after the value of θ .

Setting	Conditional Test	Marginal Test
$\theta = 1$ (100); $\theta = 2$ (900)	0.11	0.74
$\theta = 1$ (999); $\theta = 0.1$ (1)	0.06	0.33
$\theta = 1$ (999); $\theta = 100$ (1)	0.19	0.05

5 Likelihood ratio statistic

As argued above we cannot generally trust the saddlepoint approximation in high dimensional space. To use the saddlepoint approximation we need to consider a low dimensional summary statistic. In the example of this paper I used $\sum_i (Y_i, Y_i^2)$. A natural alternative is to consider the likelihood ratio statistic conditional on the value of the minimal sufficient statistic under the null hypothesis. Generalizing the example with exponential rates in Davison et al. (2014) and the example of this paper let us consider m independent variables, each variable y_j being an observation from an exponential family with density $\exp\{\theta_j y_j - \kappa_j(\theta_j)\}$. Under the hypothesis $\theta_1 = \dots = \theta_m$ the statistic $\bar{y} = \sum_j y_j/m$ is minimal sufficient and the likelihood ratio statistic is

$$w = 2 \left\{ \sum_{j=1}^m (\hat{\theta}_j y_j - \kappa_j(\hat{\theta}_j)) - (m\hat{\theta}\bar{y} - \sum_j \kappa_j(\hat{\theta})) \right\}, \quad \kappa'_j(\hat{\theta}_j) = y_j, \quad \sum_{j=1}^m \kappa'_j(\hat{\theta}) = m\bar{y}.$$

The conditional distribution of w given \bar{y} is equivalent to the conditional distribution of $h = \sum_j (\hat{\theta}_j y_j - \kappa_j(\hat{\theta}_j))$ given \bar{y} .

Consider the two dimensional exponential families with densities

$$\exp\{\lambda(\hat{\theta}_j y_j - \kappa_j(\hat{\theta}_j)) + \theta y_j - \tilde{\kappa}_j(\lambda, \theta)\},$$

for $j = 1, \dots, m$. Testing $\lambda = 0$ leads to the same conditional distribution as above, and the saddlepoint approximation (Jensen (1995)) can be expressed as a normal approximation to the modified likelihood ratio statistic $r_L = r + (1/r) \log(u/r)$, where

$$\begin{aligned} r &= \text{sgn}(\tilde{\lambda}) \sqrt{2} \left\{ \tilde{\lambda} h + m(\tilde{\theta} - \hat{\theta})\bar{y} - \sum_j \tilde{\kappa}_j(\tilde{\lambda}, \tilde{\theta}) + \sum_j \kappa_j(\hat{\theta}) \right\}^{1/2}, \\ \sum_j \frac{\partial \tilde{\kappa}_j}{\partial(\lambda, \theta)}(\tilde{\lambda}, \tilde{\theta}) &= (h, m\bar{y}), \quad \sum_j \kappa'_j(\hat{\theta}) = m\bar{y}, \\ u &= \tilde{\lambda} \sqrt{|\tilde{\Sigma}|/\hat{\Sigma}_{22}}, \quad \hat{\Sigma}_{22} = \sum_j \kappa''_j(\hat{\theta}), \quad \tilde{\Sigma} = \sum_j \frac{\partial^2 \tilde{\kappa}_j}{\partial(\lambda, \theta) \partial(\lambda, \theta)^T}(\tilde{\lambda}, \tilde{\theta}). \end{aligned}$$

We can use this to test variance homogeneity $\sigma_1^2 = \dots = \sigma_m^2$ in a normal model based on $y_j = \sum_{v=1}^{n_j} (x_{jv} - \bar{x}_j)^2$. In this case we have, with $\gamma_j = (n_j - 1)/2$,

$$\begin{aligned} \kappa_j(\theta) &= -\gamma_j \log(-\theta) + \log(\Gamma(\gamma_j)), \\ \hat{\theta}_j &= -\gamma_j/y_j, \quad \hat{\theta} = -\gamma./y., \quad \hat{\Sigma}_{22} = y.^2/\gamma., \\ \tilde{\kappa}_j(\lambda, \theta) &= -\gamma_j(1 - \lambda) \log(-\theta) + \lambda \gamma_j (\log(\gamma_j) - 1) + \log(\Gamma(\gamma_j(1 - \lambda))), \\ \sum_j \gamma_j \{ \log(1 - \tilde{\lambda}) - \psi(\gamma_j(1 - \tilde{\lambda})) \} &= \gamma. \{ \log(y.) - \log(\gamma.) \} - \sum_j \gamma_j \log(y_j), \\ \tilde{\theta} &= -\gamma.(1 - \tilde{\lambda})/y., \\ |\tilde{\Sigma}| &= \sum_j \gamma_j^2 \psi'(\gamma_j(1 - \tilde{\lambda})) y.^2 / \{ \gamma.(1 - \tilde{\lambda}) \} - y.^2 / (1 - \tilde{\lambda})^2. \end{aligned}$$

Davison et al. (2014) consider a data example with $m = 10$ groups and testing homogeneity of the variances. Their conditional test gives a p -value of 0.0389, whereas the above likelihood ratio based test gives a p -value of 0.0137, the latter agreeing with Bartlett's test (p -value of 0.0136). It seems plausible that the higher p -value of Davison et al. (2014) is caused by a low power of the test similar to the situation discussed in section 4.

The saddlepoint approximation to the marginal distribution of the likelihood ratio statistic for testing homogeneity of variances has been considered in Wu and Wong (2003). The marginal test is formally the same as the conditional test above since the likelihood ratio statistic is independent of the estimate of the common variance, although the two saddlepoint approximations give slightly different results. A slightly different version of the approximation to the marginal distribution of the likelihood ratio statistic was given in Jensen (1991).

The latter reference, Jensen (1991), shows that in a number of cases inference for vector parameters can still be done using the likelihood ratio statistic. However, one should not use the asymptotic chisquared distribution, but rather use an advanced "Bartlett correction". By this is meant that the distribution of the likelihood ratio statistic is approximated by a gamma distribution with the same mean and variance, that is, a scaled chisquared distribution where both the scaling and the degrees of freedom are chosen to match the first two moments. Typically, the first two moments must be found by simulations. These are performed under the null hypothesis with the nuisance parameter equal to its estimate. A refinement of this, but more difficult to implement, would be to estimate the first two moments conditionally on the value of the sufficient statistic under the null hypothesis. Of course one can simulate the p -value directly instead of simulating the moments and using a chisquare approximation, but for small p -values this may require a large number of simulations. To illustrate the approach in practice we can use the example on binary regression considered in Davison et al. (2014) based on the bacteria data from (Venables and Ripley, 2002, section 10.4) (the data are available in R). A Rasch model is used for the presence of bacteria including a person level and a level for each of five time points. To test for no time effect the conditional distribution given the sufficient statistic is very simple to simulate. Figure 4 shows the simulated conditional tail probabilities for the likelihood ratio statistic together with the probabilities from a scaled chisquared distribution obtained from the simulated mean and variance (based on 632000 simulated cases). In the present example the approximation is 1.35 times a chisquared distribution with 3.94 degrees of freedom. This is actually quite close to the traditional Bartlett correction that corresponds to 1.33 times a chisquared distribution with 4 degrees of freedom. The simulated conditional p -value is 0.00493 (95% confidence interval: [0.00476, 0.00511]), and the p -value from the approximating scaled chisquared distribution is 0.00492.

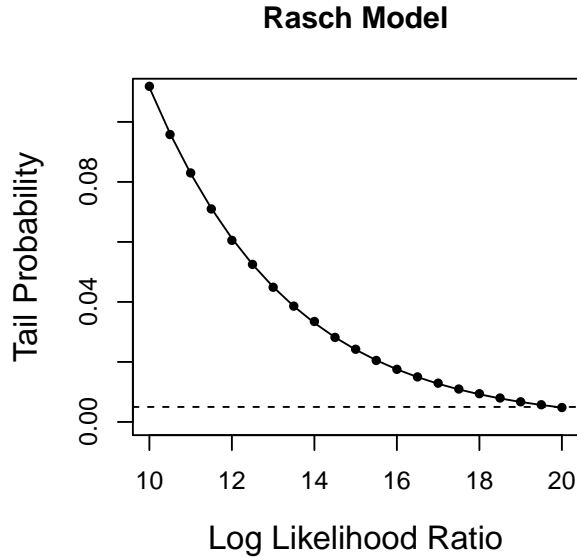


Figure 4: Empirical and approximate tail probabilities for the conditional likelihood ratio in a Rasch model based on 632000 simulated cases. The horizontal dashed line is at 0.005.

6 Summary

We have pointed out that for simple theoretical reasons the saddlepoint approximation to the conditional test of Davison et al. (2014) is not to be trusted in the very high dimensional setting. By way of an example it is shown that indeed the relative error can be large. One could have feared an even worse situation than that seen in the example, but it seems that the conditional distribution is concentrated on a small region where the relative error to the density does not change catastrophically.

The conditional test of Davison et al. (2014) seems to have small power in many relevant situations. It appears to be best suited to alternatives that relate to the geometry of the construction of the test, namely those alternatives that restrict the upper range of the conditional distribution used in the test.

Because generally we cannot trust the saddlepoint approximation in (very) high dimensional space a natural alternative is to use a low dimensional summary statistic. In the example of this paper I used $\sum_i (Y_i, Y_i^2)$ as the summary statistic. This was based on the understanding that many alternatives to the null hypothesis produce a higher variance in the data. In a sense I therefore construct a test aimed at a particular class of alternatives and, indeed, the test has no power for an alternative that tends to produce one or a few number of small values. This is the curse of dimensionality: any test will have high power in certain directions, but not all directions. Although not being based on a low dimensional summary statistic, the test of Davison et al. (2014) does not sidestep this problem, and seems to have high power in a very specific direction only. I used $\sum_i (Y_i, Y_i^2)$ as summary statistic for convenience and for its interpretability. Another logical summary statistic is based on the likelihood ratio statistic as discussed in section 5. For the case of testing homogeneity of variances the relevant moment generating function can be given analytically, but in general cases numerical integration is needed for establishing this function. Using a “Bartlett adjusted” approach to the likelihood ratio statistic seems

to work well in many situations.

I have in this paper focussed on an example that relates directly to the example on comparing exponential rates in Davison et al. (2014). That paper contains other high dimensional examples related to covariance matrices and it seems of interest to consider these situations in terms of the power of the suggested test, and to better delimit when the saddlepoint approximation can be used.

References

- Davison, A.C., Fraser, D.A.S., Reid, N. and Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *Journal of the American Statistical Association*, 109, 302–314.
- Jensen, J.L. (1991). A large deviation type approximation for the “Box class” of likelihood ratio criteria. *Journal of the American Statistical Association*, 86, 437–440.
- Jensen, J.L. (1995). *Saddlepoint Approximations*. Clarendon Press, Oxford.
- Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics With S*. Springer-Verlag New York.
- Wu, J. and Wong, A.C.M. (2003). A note on determining the p -value of Bartlett’s test of homogeneity of variances. *Communications in Statistics - Theory and Methods*, 32, 91–101.