AP 2018–03

Chair of Applied Stochastics and Risk Management

Faculty of Economic and Social Sciences
Department of Mathematics and Statistics

Working Paper

The Likelihood-Ratio Test for ∨-Hypotheses

Gabriel Frahm

March 12, 2018

The Likelihood-Ratio Test for ∨-Hypotheses

Gabriel Frahm

Helmut Schmidt University
Faculty of Economic and Social Sciences
Department of Mathematics and Statistics
Chair of Applied Stochastics and Risk Management
Holstenhofweg 85, D-22043 Hamburg, Germany

URL: www.hsu-hh.de/stochastik
Phone: +49 (0)40 6541-2791
E-mail: frahm@hsu-hh.de

Working Paper
Please use only the latest version of the manuscript. Distribution is unlimited.

Supervised by: Prof. Dr. Gabriel Frahm
Chair of Applied Stochastics and
Risk Management

URL: www.hsu-hh.de/stochastik

The Likelihood-Ratio Test for $\vee$-Hypotheses*

Gabriel Frahm†
Helmut Schmidt University
Department of Mathematics and Statistics
Chair of Applied Stochastics and
Risk Management

March 12, 2018

Abstract

The union of a finite number of single null hypotheses is referred to as a $\vee$-hypothesis, which can be rejected if and only if we are able to reject each single null hypothesis. This simple testing procedure is referred to as a $\vee$-hypothesis test. If the $\vee$-hypothesis is homogeneous, i.e., if all single null hypotheses are either one-sided or two-sided, the $\vee$-hypothesis test represents a likelihood-ratio test. It ignores the asymptotic dependence structure of the asymptotically sufficient statistic and even the number of single null hypotheses is irrelevant for calculating the critical threshold of the log-likelihood ratio. By contrast, if the $\vee$-hypothesis is heterogeneous, the $\vee$-hypothesis test is no longer a likelihood-ratio test and it is less conservative than the latter. Nonetheless, the likelihood-ratio test can be modified after which it becomes less conservative than the $\vee$-hypothesis test.

Keywords: Joint hypothesis, likelihood-ratio test, Mahalanobis distance, t-statistic.

MSC: 62F03, 62G10

*I would like to thank Tobias Wickern for many nice and illuminating discussions about the topic.
†Phone: +49 40 6541-2791, e-mail: frahm@hsu-hh.de.

1. Motivation

The principal goal of every hypothesis test is to reject the null hypothesis, H_0, in favor of the alternative hypothesis H_1. If H_0 cannot be rejected our empirical evidence in favor of H_1 is not strong enough. By contrast, if H_0 can be rejected the result is significant and we may support H_1. Let $\{H_{01}, H_{02}, \ldots, H_{0m}\}$ be some set of null hypotheses. Throughout this work, $A \wedge B$ shall be the intersection of two sets A and B, whereas $a \wedge b$ denotes the minimum of two real numbers a and b. The symbol $\vee$ indicates either the union or the maximum, depending on whether we consider sets or real numbers.

Consider the joint hypotheses

- $H_{0\wedge} := \bigwedge_{i=1}^m H_{0i}$ vs. $H_{1\wedge} := \bigvee_{i=1}^m \neg H_{0i}$ and
- $H_{0\vee} := \bigvee_{i=1}^m H_{0i}$ vs. $H_{1\vee} := \bigwedge_{i=1}^m \neg H_{0i}$,

where $\neg$ means "not." The null hypothesis $H_{0\wedge}$ is referred to as a $\wedge$-hypothesis, whereas $H_{0\vee}$ is said to be a $\vee$-hypothesis. Hence, $H_{0\wedge}$ represents the intersection of a finite number of single null hypotheses, whereas $H_{0\vee}$ is union of all single null hypotheses.

Assume that we want to test for $H_{0\wedge}$, which means that we aim at rejecting the $\wedge$-hypothesis. Further, suppose that we have a single test for each null hypothesis H_{0i} on a significance level α_i. Without any further information, we could reject $H_{0\wedge}$ whenever *at least* one single hypothesis test leads to a rejection. I call this procedure a $\wedge$-hypothesis test. Let A_i be the event in which H_{0i} is rejected and note that

$$\mathbb{P}\left(\bigcup_{i=1}^m A_i\right) \leq \sum_{i=1}^m \mathbb{P}(A_i).$$

If the null hypothesis $H_{0\wedge}$ is true it holds that $\mathbb{P}(A_i) \leq \alpha_i$ and thus, in order to guarantee that the $\wedge$-hypothesis test works on some significance level α, we should have that $\sum_{i=1}^m \alpha_i \leq \alpha$. The most simple choice of significance levels is the Bonferroni correction $\alpha_i = \alpha/m$ for $i = 1, 2, \ldots, m$. However, it is well-known that the Bonferroni correction is very conservative and can often be improved by taking the dependence structure of the single test statistics into account. However, in that case the joint hypothesis test is no longer a $\wedge$-hypothesis test.

Now, consider the null hypothesis $H_{0\vee}$ and assume that this is rejected whenever *each* single hypothesis H_{0i} is rejected. This procedure is referred to as a $\vee$-hypothesis test. It holds that

$$\mathbb{P}\left(\bigcap_{i=1}^m A_i\right) \leq \bigwedge_{i=1}^m \mathbb{P}(A_i)$$

and if the null hypothesis $H_{0\vee}$ is true we have that $\mathbb{P}(A_i) \leq \alpha_i$ for some single hypothesis test. This means that our joint test for the $\vee$-hypothesis works on the significance level α whenever $\bigvee_{i=1}^m \alpha_i \leq \alpha$. The least conservative choice of significance levels is $\alpha_1, \alpha_2, \ldots, \alpha_m = \alpha$.

At first glance, similar to the Bonferroni test, the $\vee$-hypothesis test might seem to suffer from a lack of power. In this work, I investigate the question of whether one can improve the test by taking the dependence structure of the single test statistics into account. In a quite general framework, I show that the $\vee$-hypothesis test represents a likelihood-ratio test. Hence, rejecting $H_{0\vee}$ whenever each single test rejects H_{0i} on the significance level α just *means* to apply a likelihood-ratio test on the same significance level. The given result demonstrates that the (asymptotic) correlations between the single test statistics are irrelevant when applying a likelihood-ratio test for $H_{0\vee}$. In contrast to any hypothesis test for $H_{0\wedge}$, the $\vee$-hypothesis test does not depend on the number of single null hypotheses. The precise meaning of the latter statement will become clear during the subsequent analysis.

Before proceeding further, I would like to mention that neither the $\wedge$- nor the $\vee$-hypothesis test represents a multiple test, where $\{H_{01}, H_{02}, \ldots, H_{0m}\}$ is considered a family of null hypotheses. The principal goal of multiple testing is to reject as many null hypotheses as possible without exceeding some family-wise error rate (Lehmann and Romano, 2005, Chapter 9). By contrast, the hypothesis tests described above aim at rejecting a *joint* null hypothesis, i.e., $H_{0\wedge}$ or $H_{0\vee}$, respectively, and so we need not consider any family-wise error rate.

The following examples shall illustrate why tests for the $\vee$-hypothesis play a fundamental role in many practical applications.

Example 1: Consider a linear regression model $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m + u$, in which the parameter vector $\beta = (\beta_0, \beta_1, \ldots, \beta_m)$ is unknown. A typical question is whether the chosen regressors $X_1, X_2, \ldots, X_m$ are significant. This means that we want to test

$$H_{0\vee}\colon \bigvee_{i=1}^{m} \beta_i = 0 \quad \text{vs.} \quad H_{1\vee}\colon \bigwedge_{i=1}^{m} \beta_i \neq 0.$$

Note that the classical F-test is made for the $\wedge$-hypothesis $H_{0\wedge}\colon \bigwedge_{i=1}^{m} \beta_i = 0$. If that leads to a rejection, we may support the alternative hypothesis $H_{1\wedge}\colon \bigvee_{i=1}^{m} \beta_i \neq 0$. Put another way, we may suspect that *any* regressor is significant. By contrast, if we are able to reject the $\vee$-hypothesis $H_{0\vee}$ we know that *all* regressors are significant.

Example 2: Suppose that the therapeutic effects of m different treatments are investigated in a clinical study. Let $\theta_1, \theta_2, \ldots, \theta_m$ be the true but unknown effects of the given treatments and θ_0 the (placebo) effect of a control group. We could be interested to know whether Treatment m is optimal among all considered treatments. The corresponding hypotheses are given by

$$H_{0\vee}\colon \bigvee_{i=0}^{m-1} \theta_m < \theta_i \quad \text{and} \quad H_{1\vee}\colon \bigwedge_{i=0}^{m-1} \theta_m \geq \theta_i.$$

Hence, Treatment m proves to be optimal if $H_{0\vee}$ can be rejected. This is completely different from testing $H_{0\wedge}\colon \bigwedge_{i=0}^{m-1} \theta_m \geq \theta_i$ vs. $H_{1\wedge}\colon \bigvee_{i=0}^{m-1} \theta_m < \theta_i$, which enables us to prove that Treatment m is *not* optimal.

Example 3: We observe m variables $X_1, X_2, \ldots, X_m$ in some population with n individuals. Now, we want to know whether the mean of X_i exceeds some threshold $\tau_i \in \mathbb{R}$ for all $i \in$

$\{1, 2, \ldots, m\}$. That is, we aim at testing

$$H_{0\vee}\colon \bigvee_{i=1}^{m} \mathbf{E}(X_i) \leq \tau_i \quad \text{vs.} \quad H_{1\vee}\colon \bigwedge_{i=1}^{m} \mathbf{E}(X_i) > \tau_i.$$

The given examples are by far not exhaustive and the reader can find many other situations in which a ∨-hypothesis occurs.

2. General Framework

Let $(\Omega, \mathcal{F})$ be a sample space that is equipped with an indexed probability measure $\mathbb{P}_\theta$, where $\theta \in \Theta \subseteq \mathbb{R}^d$ represents an unknown parameter vector and Θ is an open subset of $\mathbb{R}^d$. Further, let $\mathcal{X}_n = \begin{bmatrix} X_1 \; X_2 \; \cdots \; X_n \end{bmatrix}$ be any sample of random quantities that are measurable on $(\Omega, \mathcal{F})$. For a sufficiently large sample size n, there shall exist a measurable test statistic $\theta_n\colon \mathcal{X}_n \mapsto \mathbb{R}^d$ such that

$$\sqrt{n}\,\big(\theta_n - \theta\big) \rightsquigarrow \mathcal{N}_d\big(\mathbf{0}, \Omega\big).$$

Here "$\rightsquigarrow$" denotes convergence in distribution, $\mathbf{0}$ is a d-dimensional vector of zeros,[1] and the asymptotic covariance matrix $\Omega \in \mathbb{R}^{d \times d}$ is supposed to be positive definite.

The random quantities $X_1, X_2, \ldots, X_n$ may dependent on each other. Nonetheless, in many applications the weak convergence property of $\sqrt{n}\,\big(\theta_n - \theta\big)$ follows from the Central Limit Theorem, which can be guaranteed under mild regularity conditions such as ergodicity and strong mixing (Bradley, 2005, Hayashi, 2000, Chapter 2 and 6). Alternatively, we could consider $\sqrt{n}\,\big(\theta_n - \theta\big)$ an asymptotically sufficient statistic in the context of local asymptotic normality (Le Cam, 1986, van der Vaart, 2002).

Our ∨-hypothesis reads

$$H_{0\vee}\colon \left(\bigvee_{i=1}^{l} w_i\theta = \theta_{0i}\right) \vee \left(\bigvee_{i=l+1}^{m} w_i\theta \leq \theta_{0i}\right)$$

with $0 \leq l \leq m$ and $m > 1$. Here, $w_i := \begin{bmatrix} w_{i1} \; w_{i2} \; \cdots \; w_{id} \end{bmatrix}$ is any row vector of real numbers and $\theta_{0i} \in \mathbb{R}$ for $i = 1, 2, \ldots, m$. The ∨-hypothesis can be understood, equivalently, as a subset of $\mathbb{R}^d$, viz. $H_{0\vee} = \bigcup_{i=1}^{m} H_{0i}$ with

- $H_{0i} = \big\{\theta \in \mathbb{R}^d\colon w_i\theta = \theta_{0i}\big\}$ for $i = 1, 2, \ldots, l$ and
- $H_{0i} = \big\{\theta \in \mathbb{R}^d\colon w_i\theta \leq \theta_{0i}\big\}$ for $i = l+1, l+2, \ldots, m$.

In this case, each single null hypothesis, H_{0i}, represents either an affine hyperplane or an affine half-space in $\mathbb{R}^d$. I assume that $\Theta \cap H_{0\vee} \neq \emptyset$ in order to avoid any triviality.

Define the quantities $W := \big[w_{ij}\big] \in \mathbb{R}^{m \times d}$ and $\theta_0 := (\theta_{01}, \theta_{02}, \ldots, \theta_{0m}) \in \mathbb{R}^m$, where θ_0 represents a *column* vector in $\mathbb{R}^m$, so that the weak convergence property of $\sqrt{n}\,\big(\theta_n - \theta\big)$ reduces

[1] Throughout this work, the number of dimensions of $\mathbf{0}$ shall always be clear from the context.

to

$$\sqrt{n}\,(\mu_n - \mu) \rightsquigarrow \mathcal{N}_m(\mathbf{0}, \Sigma) \tag{1}$$

with $\mu_n := W\theta_n - \theta_0$, $\mu := W\theta - \theta_0$, and $\Sigma := W\Omega W'$. I suppose that the asymptotic covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is positive definite, which implies that $m \leq d$ and that the row vectors of W are linearly independent.

After the re-parameterization the parameter set turns into $\mathcal{P} := W\Theta - \theta_0$, which is an open subset of $\mathbb{R}^m$. Moreover, the ∨-hypothesis can be reformulated in a more convenient way as

$$H_{0\vee}\colon \left(\bigvee_{i=1}^{l} \mu_i = 0\right) \vee \left(\bigvee_{i=l+1}^{m} \mu_i \leq 0\right),$$

where μ_i denotes the ith component of μ. Alternatively, we can interpret the ∨-hypothesis in the topological sense as $H_{0\vee} = \bigcup_{i=1}^{m} H_{0i}$ with $H_{0i} = \{\mu \in \mathbb{R}^m : \mu_i = 0\}$ for $i = 1, 2, \ldots, l$ and $H_{0i} = \{\mu \in \mathbb{R}^m : \mu_i \leq 0\}$ for $i = l+1, l+2, \ldots, m$.[2] However, the reader should distinguish between μ_n and μ_i. The former represents an m-dimensional estimator for μ, whereas the latter is a real number. In most practical applications n is greater than m and thus no confusion arises.

The main conclusions of this work do not change if we substitute any single null hypothesis $H_{0i}\colon \mu_i \leq 0$ with $H'_{0i}\colon \mu_i < 0$ and so I will ignore strict inequalities without loss of generality. I say that $H_{0\vee}$ is homogeneous if and only if *all* single null hypotheses are either one-sided or two-sided. Put another way, $H_{0\vee}$ is heterogenous if and only if $0 < l < m$. The ∨-hypotheses given by the three examples in the introduction are homogeneous. It seems to me that this is the typical case in most practical applications and, at the end of this work, the reader will see that the heterogeneous case is a little bit more intricate than the homogeneous one.

3. Main Results

Consider the matrix decomposition $\Sigma = \sigma\rho\,\sigma$, where σ is an $m \times m$ diagonal matrix and $\rho > 0$ is an $m \times m$ correlation matrix. Since Σ is positive definite both σ and ρ must be positive definite, too. Let $\|\cdot\|$ be the Euclidean norm and $\|\cdot\|_\Sigma$ the Mahalanobis norm with respect to Σ, i.e., $\|x\|^2 = x'x$ and $\|x\|_\Sigma^2 = x'\Sigma^{-1}x$ for each $x \in \mathbb{R}^m$. The corresponding distance between some point $x \in \mathbb{R}^m$ and any nonempty subset $\mathcal{S}$ of $\mathbb{R}^m$ is defined as

$$\|x - \mathcal{S}\| := \inf_{y \in \mathcal{S}} \|x - y\| \quad \text{and} \quad \|x - \mathcal{S}\|_\Sigma := \inf_{y \in \mathcal{S}} \|x - y\|_\Sigma\,,$$

respectively. It is clear that $\|x - \mathcal{S}\| = \|x - \mathcal{S}\|_\Sigma = 0$ whenever $x \in \mathcal{S}$. A subset $\mathcal{C} \subseteq \mathbb{R}^m$ is said to be a (pointed) cone if and only if $z \in \mathcal{C} \Rightarrow \gamma z \in \mathcal{C}$ for all $\gamma \geq 0$.

It holds that $\mathcal{P}_0 := \mathcal{P} \cap H_{0\vee} \neq \emptyset$ and I make the modest assumption that $\mu_n \in \mathcal{P}$.[3] The

[2] For notational convenience, but also because the re-parameterization does not alter the null hypotheses in the logical sense, I refrain from choosing different symbols for $H_{0\vee}, H_{01}, H_{02}, \ldots, H_{0m}$ before and after the re-parameterization.

[3] Any assertion about some random quantity is meant to be true $\mathbb{P}_\theta$-almost surely unless otherwise stated. Hence, the statement "$\mu_n \in \mathcal{P}$" means that $\mathbb{P}_\theta(\mu_n \in \mathcal{P}) = 1$.

log-likelihood ratio (see, e.g., van der Vaart, 1998, p. 228) of our experiment is given by

$$\begin{aligned}\Lambda_n &:= 2\log\frac{\sup_{y\in\mathcal{P}}\exp\left[-\frac{1}{2}(\mu_n-y)'(\Sigma/n)^{-1}(\mu_n-y)\right]}{\sup_{y\in\mathcal{P}_0}\exp\left[-\frac{1}{2}(\mu_n-y)'(\Sigma/n)^{-1}(\mu_n-y)\right]}\\ &= \inf_{y\in\mathcal{P}_0} n\,(\mu_n-y)'\Sigma^{-1}(\mu_n-y) = n\|\mu_n-\mathcal{P}_0\|_\Sigma^2\,.\end{aligned}$$

Note that Λ_n and $\Lambda_{0n} := n\|\mu_n - H_{0\vee}\|_\Sigma^2$ are asymptotically equivalent, i.e., $|\Lambda_n - \Lambda_{0n}| \xrightarrow{p} 0$, whenever $\mu \in \mathcal{P}_0$. In fact, since $\mathcal{P}$ is an open subset of $\mathbb{R}^m$ and $\mu_n \xrightarrow{p} \mu$, we always can find an open ball in $\mathcal{P}$ around μ with radius $\varepsilon > 0$ such that the event $\|\mu_n - \mu\| < \varepsilon$ occurs with any arbitrarily high probability if the number of observations, n, is sufficiently large. Moreover, if the ∨-hypothesis is true we even can make ε small enough such that $\Lambda_{0n} = \Lambda_n$ whenever $\|\mu_n - \mu\| < \varepsilon$. Put another way, the restriction imposed by $\mathcal{P}$ is asymptotically negligible and so we may focus on Λ_{0n} in order to derive the asymptotic distribution of Λ_n.

A subset $\mathcal{S} \subset \mathbb{R}^m$ that is obtained by setting precisely $k \in \{1, 2, \ldots, m\}$ dimensions of $\mathbb{R}^m$ to zero is said to be an $(m-k)$-dimensional canonical subspace of $\mathbb{R}^m$. In the case of $k = 1$, i.e., if we eliminate only one dimension of $\mathbb{R}^m$, the canonical subspace represents a (linear) hyperplane in $\mathbb{R}^m$. This will be referred to as a canonical hyperplane in $\mathbb{R}^m$. Consider the decompositions $x = (x_1, x_2)$ and

$$\Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix},$$

where $x_1 \in \mathbb{R}^k$ and $\Sigma_{11} \in \mathbb{R}^{k\times k}$ belong to the zero dimensions of $\mathcal{S}$. If $\mathcal{S}$ is a canonical hyperplane we have that $\Sigma_{11} \equiv \sigma_1^2$.

The following proposition will serve as a basic result in the subsequent analysis. It implies that the Mahalanobis distance of any point $x \in \mathbb{R}^m$ with respect to a canonical hyperplane does not depend on the correlation matrix ρ that is implied by Σ.

Proposition 1. *If $\mathcal{S}$ is an $(m-k)$-dimensional canonical subspace of $\mathbb{R}^m$ then*

$$\|x - \mathcal{S}\|_\Sigma = \|x_1\|_{\Sigma_{11}}$$

for all $x \in \mathbb{R}^m$, where $x_1 \in \mathbb{R}^k$ and $\Sigma_{11} \in \mathbb{R}^{k\times k}$ belong to the zero dimensions of $\mathcal{S}$. In particular, if $\mathcal{S}$ is a canonical hyperplane in $\mathbb{R}^m$ it holds that $\|x - \mathcal{S}\|_\Sigma = |x_1|/\sigma_1$.

Proof. Note that

$$\|x - \mathcal{S}\|_\Sigma^2 = \inf_{y\in\mathcal{S}} (x-y)'\Sigma^{-1}(x-y).$$

Finding the infimum represents a convex minimization problem and, since Σ is positive definite, the solution is unique. The Lagrangian reads

$$L(x,\lambda) = (x-y)'\Sigma^{-1}(x-y) + \sum_{i=1}^{k}\lambda_i \mathbf{e}_i' y$$

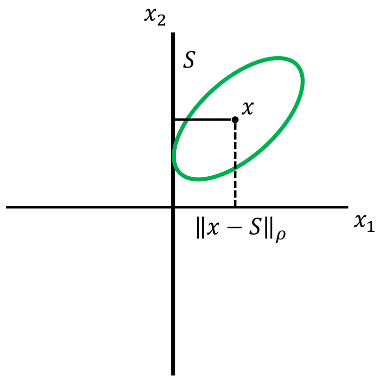

Figure 1: Mahalanobis distance of a point $x \in \mathbb{R}^2$ to the vertical line $\mathcal{S}$.

with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ and its partial derivative with respect to x is

$$-2\,\Sigma^{-1}(x-y) + \sum_{i=1}^{k} \lambda_i \mathbf{e}_i = -2\,\Sigma^{-1}(x-y) + \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \lambda,$$

where $\mathbf{I}$ denotes the $k \times k$ identity matrix and $\mathbf{0}$ is an $(m-k) \times k$ matrix of zeros. It follows that

$$y = x - \frac{1}{2}\Sigma \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \lambda = x - \frac{1}{2} \begin{bmatrix} \Sigma_{11} \\ \Sigma_{21} \end{bmatrix} \lambda$$

and from $x_1 - \frac{1}{2}\Sigma_{11}\lambda = 0$ we conclude that $\lambda = 2\Sigma_{11}^{-1}x_1$. Hence, we obtain $y_2 = x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1$ and thus

$$x - y = \begin{bmatrix} \mathbf{I} \\ \Sigma_{21}\Sigma_{11}^{-1} \end{bmatrix} x_1\,.$$

Moreover, it is well-known that

$$\Sigma^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

with $B_{11} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}B_{22}\Sigma_{21}\Sigma_{11}^{-1}$, $B_{12} = -\Sigma_{11}^{-1}\Sigma_{12}B_{22}$, and $B_{21} = -B_{22}\Sigma_{21}\Sigma_{11}^{-1}$. This means that

$$\Sigma^{-1}(x-y) = \begin{bmatrix} \Sigma_{11}^{-1} \\ \mathbf{0} \end{bmatrix} x_1$$

and thus $(x-y)'\Sigma^{-1}(x-y) = x_1'\Sigma_{11}^{-1}x_1$. We conclude that

$$\|x - \mathcal{S}\|_\Sigma = \sqrt{(x-y)'\Sigma^{-1}(x-y)} = \sqrt{x_1'\Sigma_{11}^{-1}x_1} = \|x_1\|_{\Sigma_{11}}.$$

The rest of the proof is trivial. □

Proposition 1 is illustrated in Figure 1, where we can see that the Mahalanobis distance of $x \in \mathbb{R}^2$ to the vertical line equals $|x_1|$. Here, without loss of generality, the Mahalanobis norm refers to the correlation matrix ρ rather than the covariance matrix Σ.

The next proposition represents the key observation of this work. It will be used later on in

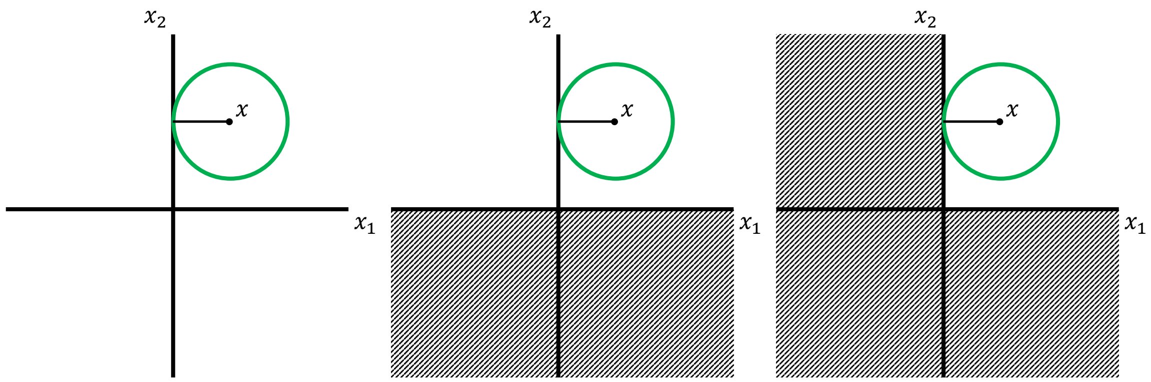

Figure 2: Cones in $\mathbb{R}^2$ satisfying the condition of Proposition 2.

order to derive the asymptotic distribution of Λ_n.

Proposition 2. *If $\mathcal{C} \subset \mathbb{R}^m$ is a cone such that $\|z - \mathcal{C}\| = \bigwedge_{i=1}^m |z_i|$ for all $z = (z_1, z_2, \ldots, z_m) \in \mathbb{R}^m \setminus \mathcal{C}$ then we have that*

$$\|x - \mathcal{C}\|_\Sigma = \bigwedge_{i=1}^m \frac{|x_i|}{\sigma_i}$$

for all $x = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m \setminus \mathcal{C}$ but $\|x - \mathcal{C}\|_\Sigma = 0$ for all $x \in \mathcal{C}$.

Proof. Note that

$$\|x - \mathcal{C}\|_\Sigma^2 = \inf_{y \in \mathcal{C}} (x - y)'\Sigma^{-1}(x - y)$$

and so we obtain $\|x - \mathcal{C}\|_\Sigma = 0$ in the case of $x \in \mathcal{C}$. Recall that Σ is positive definite. Hence, if x does not belong to $\mathcal{C}$ the infimum must be attained on the boundary of $\mathcal{C}$. Due to the property of $\mathcal{C}$ that is expressed by the proposition, this can be found on a canonical hyperplane of $\mathbb{R}^m$. Thus, it can be determined by setting, successively, each component of y to zero and minimizing the quadratic form with respect to the other components of y. The desired infimum corresponds to the smallest value of the quadratic forms that have been obtained for every canonical hyperplane. Consider, without loss of generality, the partitions

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where x_1 denotes the first component of x, y_1 is the first component of y, and $\Sigma_{11} \equiv \sigma_1^2$ symbolizes the upper left element of Σ. From Proposition 1 we know that $\|x - \mathcal{S}\|_\Sigma = |x_1|/\sigma_1$ and so we conclude that

$$\|x - \mathcal{C}\|_\Sigma = \bigwedge_{i=1}^m \frac{|x_i|}{\sigma_i}$$

for all $x \in \mathbb{R}^m \setminus \mathcal{C}$. □

Figure 2 illustrates some cones in $\mathbb{R}^2$ that satisfy the condition of Proposition 2.[4] In particular,

[4] A typical counterexample is any canonical hyperplane or half-space in $\mathbb{R}^m$ with $m > 1$.

the set $H_{0\vee}$ represents a cone in $\mathbb{R}^m$ that satisfies the given condition. Thus, we have that

$$\Lambda_{0n} = n\|\mu_n - H_{0\vee}\|^2_\Sigma = \mathbf{1}_{\mu_n \notin H_{0\vee}} \left(\bigwedge_{i=1}^m \frac{|\mu_{in}|}{\sigma_i/\sqrt{n}} \right)^2 .^5$$

The quintessence is that we may ignore the asymptotic correlation matrix of $\sqrt{n}\,(\mu_n - \mu)$, i.e., ρ, when calculating Λ_{0n}. The likelihood-ratio test rejects $H_{0\vee}$ if Λ_n exceeds a critical threshold $\tau^2 > 0$. We already know that Λ_n and Λ_{0n} are asymptotically equivalent. Hence, in our context, this (essentially) means that $\Lambda_{0n} > \tau^2$, i.e., that *each* $\sqrt{n}\,|\mu_{in}|/\sigma_i$ exceeds τ provided $\mu_n \notin H_{0\vee}$.

Although we need not take ρ into account when calculating Λ_{0n}, its asymptotic distribution may well depend on the asymptotic correlation matrix of $\sqrt{n}\,(\mu_n - \mu)$. The set $H_{0\vee}$ is Chernoff regular and so we have that $\sqrt{n}\,(H_{0\vee} - \mu) \to \mathcal{T}_0(\mu)$, where the convergence takes place in the Painlevé-Kuratowski sense (Geyer, 1994). The limit of $\sqrt{n}\,(H_{0\vee} - \mu)$, i.e., $\mathcal{T}_0(\mu)$, represents the tangent cone of $H_{0\vee}$ at μ and, since $\mathcal{P}$ is open, it holds that $\sqrt{n}\,(\mathcal{P}_0 - \mu) \to \mathcal{T}_0(\mu)$. I make the following implicit assumption, which is hopefully satisfied in most practical applications:

$$\left\| \sqrt{n}\,(\mu_n - \mu) - \sqrt{n}\,(H_{0\vee} - \mu) \right\|^2_\Sigma = \left\| \sqrt{n}\,(\mu_n - \mu) - \mathcal{T}_0(\mu) \right\|^2_\Sigma + o_{\mathrm{P}}(1).$$

The following theorem provides the asymptotic distribution of Λ_{0n} and thus of Λ_n.

Theorem 1. $\Lambda_{0n}, \Lambda_n \rightsquigarrow \left\| \mathcal{N}_m(\mathbf{0}, \rho) - \mathcal{T}_0(\mu) \right\|^2_\rho$

Proof. Note that

$$\begin{aligned}
\Lambda_{0n} &= \inf_{y \in H_{0\vee}} n\,(\mu_n - y)'\Sigma^{-1}(\mu_n - y) \\
&= \inf_{y \in H_{0\vee}} \left[\sqrt{n}\,(\mu_n - \mu) - \sqrt{n}\,(y - \mu) \right]' \Sigma^{-1} \left[\sqrt{n}\,(\mu_n - \mu) - \sqrt{n}\,(y - \mu) \right] \\
&= \|\sqrt{n}\,(\mu_n - \mu) - \sqrt{n}\,(H_{0\vee} - \mu)\|^2_\Sigma = \|\sqrt{n}\,(\mu_n - \mu) - \mathcal{T}_0(\mu)\|^2_\Sigma + o_{\mathrm{P}}(1) \\
&= \|\sqrt{n}\,\sigma^{-1}(\mu_n - \mu) - \sigma^{-1}\mathcal{T}_0(\mu)\|^2_\rho + o_{\mathrm{P}}(1)
\end{aligned}$$

with $\sqrt{n}\,\sigma^{-1}(\mu_n - \mu) \rightsquigarrow \mathcal{N}_m(\mathbf{0}, \rho)$ and $\sigma^{-1}\mathcal{T}_0(\mu) = \mathcal{T}_0(\mu)$. Moreover, the distance $\|x - \mathcal{T}_0(\mu)\|_\rho$ is continuous in x and so, from the Continuous Mapping Theorem and Slutsky's Theorem, we conclude that

$$\Lambda_{0n} = \|\sqrt{n}\,\sigma^{-1}(\mu_n - \mu) - \mathcal{T}_0(\mu)\|^2_\rho + o_{\mathrm{P}}(1) \rightsquigarrow \|\mathcal{N}_m(\mathbf{0}, \rho) - \mathcal{T}_0(\mu)\|^2_\rho .$$

Since Λ_n and Λ_{0n} are asymptotically equivalent, we obtain the same result for Λ_n. □

It is worth emphasizing that the asymptotic results presented in this work do not change if we substitute σ with some estimator σ_n such that $\sqrt{n}\,\sigma_n^{-1}(\mu_n - \mu) \rightsquigarrow \mathcal{N}_m(\mathbf{0}, \rho)$ and

$$\left\| \sqrt{n}\,\sigma_n^{-1}(\mu_n - \mu) - \sqrt{n}\,\sigma_n^{-1}(H_{0\vee} - \mu) \right\|^2_\rho = \left\| \sqrt{n}\,\sigma_n^{-1}(\mu_n - \mu) - \mathcal{T}_0(\mu) \right\|^2_\rho + o_{\mathrm{P}}(1).$$

[5] The symbol $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function, i.e., $\mathbf{1}_A = 1$ if the assertion A is true and $\mathbf{1}_A = 0$ otherwise.

The correlation matrix ρ need not be known at all.

Theorem 1 can be used in order to derive the critical threshold τ^2 for the likelihood-ratio test. For this purpose, I distinguish between the homogeneous and the heterogeneous case.

3.1. The Homogeneous Case

Now, let us come back to the $\vee$-hypothesis test, which is described in the introduction, and suppose that $H_{0\vee}$ is homogeneous. Then the critical thresholds are identical, i.e., in the one-sided case ($l = 0$) we reject H_{0i} if and only if

$$t_{in} := \frac{\mu_{in}}{\sigma_i / \sqrt{n}} > z_\alpha,$$

where z_α denotes the $(1-\alpha)$-quantile of the standard normal distribution, whereas in the two-sided case ($l = m$) we reject H_{0i} if and only if $|t_{in}| > z_{\alpha/2}$. In any case, the $\vee$-hypothesis test rejects $H_{0\vee}$ if and only if (the absolute value of) each single t-statistic exceeds the *same* critical threshold. Put another way, $H_{0\vee}$ is rejected whenever

$$\sqrt{\Lambda_n} \approx \sqrt{\Lambda_{0n}} = \mathbf{1}_{\mu_n \notin H_{0\vee}} \bigwedge_{i=1}^{m} |t_{in}| > \begin{cases} z_\alpha, & l = 0 \\ z_{\alpha/2}, & l = m \end{cases}.$$

We conclude that the $\vee$-hypothesis test represents a likelihood-ratio test with critical threshold $\tau^2 = z_\alpha^2$ or $\tau^2 = z_{\alpha/2}^2$, respectively.

We already know that the hypothesis test obeys the significance level α. Nonetheless, the reader might ask whether it is possible to choose a smaller threshold in order to increase the power of the likelihood-ratio test without destroying the given significance level. This question can be answered by Theorem 1, which reveals the asymptotic distribution of Λ_{0n}. We can see that the worst case under the null hypothesis $H_{0\vee}$ is any situation in which one and *only* one component of μ equals zero, i.e., if $\mathcal{T}_0(\mu)$ represents either a canonical hyperplane ($l = m$) or a canonical half-space ($l = 0$) in $\mathbb{R}^m$. One can imagine that this is precisely the case in which the $(1-\alpha)$-quantile of $\left\|\mathcal{N}_m(\mathbf{0},\rho) - \mathcal{T}_0(\mu)\right\|_\rho^2$ becomes maximal.

First of all assume that $l = m$ and suppose, without loss of generality, that

$$\mathcal{T}_0(\mu) = \Big\{x = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m : x_1 = 0\Big\}.$$

This means that only the first component of μ is zero. Hence, the first single null hypothesis H_{01} is satisfied, but each other single null hypothesis is violated. From Proposition 1 we know that $\|x - \mathcal{T}_0(\mu)\|_\rho = |x_1|$. Hence, since H_{01} is two-sided, we obtain

$$\|X - \mathcal{T}_0(\mu)\|_\rho^2 \sim \chi_1^2$$

with $X \sim \mathcal{N}_m(\mathbf{0},\rho)$. Since we have that $l = m$ the same conclusion can be made after setting any other component of μ to zero and assuming that all other components are distinct from zero.

By contrast, if H_{01} is one-sided it follows that $\|X - \mathcal{T}_0(\mu)\|_\rho^2$ is distributed like $\max\{\zeta, 0\}^2$ with

$\zeta \sim \mathcal{N}(0,1)$. This is a standard result of likelihood theory. To the best of my knowledge it goes back to Chernoff (1954).[6] Note that $F_{\chi^2_1}(z^2_{\alpha/2}) = 1-\alpha$ and thus

$$F_{\max\{\zeta,0\}^2}(z^2_\alpha) = 0.5 + 0.5\underbrace{F_{\chi^2_1}(z^2_\alpha)}_{=1-2\alpha} = 1-\alpha.$$

Thus, in the case of $l = 0$, the best choice for τ^2 is z^2_α, whereas for $l = m$ we should choose $\tau^2 = z^2_{\alpha/2}$. This means that the ∨-hypothesis test, in fact, represents a likelihood-ratio test that has a Type-I error probability of α in the worst case. That is, given the significance level α, we cannot increase the power of the likelihood-ratio test by choosing a smaller threshold. Moreover, the ∨-hypothesis test inherits the asymptotic optimality properties of likelihood-ratio tests that are known from likelihood theory (see, e.g., van der Vaart, 1998, Chapter 15 and 16). In particular, if $\mathcal{T}_0(\mu)$ represents a canonical hyperplane or half-space in $\mathbb{R}^m$ then the likelihood-ratio test proves to be uniformly most powerful (van der Vaart, 1998, Proposition 15.2 and p. 236).

3.2. The Heterogeneous Case

The ∨-hypothesis test represents a proper likelihood-ratio test if $H_{0\vee}$ is homogeneous, but this no longer holds true if $H_{0\vee}$ is heterogenous. The problem is that the likelihood-ratio test does not distinguish between the one-sided and the two-sided single null hypotheses. If we want to conduct the (ordinary) likelihood-ratio test in the heterogenous case we must choose the larger threshold $\tau^2 = z^2_{\alpha/2}$. This threshold applies to each single t-statistic—irrespective of whether it refers to a one-sided or two-sided single null hypothesis.

By contrast, the ∨-hypothesis test provides a lower threshold to the one-sided single null hypotheses and so the likelihood-ratio test is more conservative. However, it can be improved by choosing a critical threshold that depends on $\arg\min_{i\in\{1,2,\ldots,m\}} |t_{in}|$ whenever $\mu_n \notin H_{0\vee}$.[7] The resulting test is said to be a modified likelihood-ratio test. The principal idea goes like this:

- If $\mathcal{T}_0(\mu)$ is a canonical hyperplane we have that $t_{in} \to \infty$ for all $i > l$ and so it cannot happen that $\arg\min |t_{in}| > l$ provided the sample size, n, is large enough.
- By contrast, if $\mathcal{T}_0(\mu)$ is a canonical half-space we have that $|t_{in}| \to \infty$ for all $i \le l$ and thus it cannot happen that $\arg\min |t_{in}| \le l$, given that we have enough observations.

It is clear that the likelihood-ratio test makes sense only if the sample size is sufficiently large but in this case $\arg\min |t_{in}|$ provides us with important information: If $\arg\min |t_{in}| > l$ we know that $\mathcal{T}_0(\mu)$ cannot be a canonical hyperplane and if $\arg\min |t_{in}| \le l$ it cannot be a canonical half-space. Hence, the modified likelihood-ratio test rejects $H_{0\vee}$ whenever

$$\sqrt{\Lambda_{0n}} = \mathbf{1}_{\mu_n \notin H_{0\vee}} \bigwedge_{i=1}^{m} |t_{in}| > \begin{cases} z_\alpha, & \arg\min |t_{in}| > l \\ z_{\alpha/2}, & \arg\min |t_{in}| \le l \end{cases}.$$

[6]For similar results concerning the asymptotic distributions of log-likelihood ratios see, e.g., Self and Liang (1987).

[7]If $\mu_n \in H_{0\vee}$ we have that $\Lambda_{0n} = 0$ and then it is clear that $H_{0\vee}$ cannot be rejected at all. Moreover, in the following I write "$\arg\min |t_{in}|$" instead of $\arg\min_{i\in\{1,2,\ldots,m\}} |t_{in}|$ for notational convenience.

The modified likelihood-ratio test is less conservative than the $\vee$-hypothesis test. This is because in case $\arg\min |t_{in}| > l$ the critical threshold for each two-sided single null hypothesis is lower than that of the $\vee$-hypothesis test. The modified likelihood-ratio test reduces the ordinary one in the case of $l = 0$ or $l = m$, i.e., if the $\vee$-hypothesis is homogeneous.

4. Conclusion

The seemingly naive approach of rejecting a homogeneous $\vee$-hypothesis if we are able to reject each single null hypothesis in $H_{0\vee}$ represents a likelihood-ratio test. Hence, this simple procedure can be justified either by likelihood theory or by local asymptotic normality theory. The likelihood-ratio test ignores the asymptotic dependence structure of $\sqrt{n}\,(\theta_n - \theta)$, i.e., of the asymptotically sufficient statistic of the given experiment. For this reason why we may focus on rejecting each single null hypothesis in order to reject the joint null hypothesis. We even need not take the number of single null hypotheses into account because this has no impact on the worst-case asymptotic distribution of the log-likelihood ratio. In particular, we need not apply a Bonferroni correction, or any similar technique from multiple testing, to control the Type-I error probability. If $H_{0\vee}$ is heterogeneous the $\vee$-hypothesis test is no longer a likelihood-ratio test and the former is less conservative than the latter. Nonetheless, the likelihood-ratio test can simply be modified after which it becomes less conservative than the $\vee$-hypothesis test. The modified likelihood-ratio test corresponds to the ordinary likelihood-ratio test if $H_{0\vee}$ is homogeneous.

References

Bradley, R. (2005): "Basic properties of strong mixing conditions. A survey and some open questions," *Probability Surveys* **2**, pp. 107–144.

Chernoff, H. (1954): "On the distribution of the likelihood ratio," *Annals of Mathematical Statistics* **25**, pp. 573–578.

Geyer, C. (1994): "On the asymptotics of constraint M-estimation," *Annals of Statistics* **22**, pp. 1993–2010.

Hayashi, F. (2000): *Econometrics*, Princeton University Press.

Le Cam, L. (1986): *Asymptotic Methods in Statistical Decision Theory*, Springer.

Lehmann, E., Romano, J. (2005): *Testing Statistical Hypotheses*, Springer.

Self, S., Liang, K.L. (1987): "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions," *Journal of the American Statistical Association* **82**, pp. 605–610.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press.

van der Vaart, A. (2002): "The statistical work of Lucien Le Cam," *Annals of Statistics* **30**, pp. 631–682.