asympTest: an R package for performing parametric statistical tests and confidence intervals based on the central limit theorem

J.-F. Coeurjolly[1], R. Drouilhet[1], P. Lafaye de Micheaux[1] and J.-F. Robineau[2]
[1] SAGAG Team, Laboratory Jean Kuntzmann/ CNRS, Grenoble (France), [2] CQLS, France

February 26, 2009

Abstract

This paper describes an R package implementing large sample tests and confidence intervals (based on the central limit theorem) for various parameters. The one and two sample mean and variance contexts are considered. The statistics for all the tests are expressed in the same form, which facilitates their presentation. In the variance parameter cases, the asymptotic robustness of the classical tests depends on the departure of the data distribution from normality measured in terms of the kurtosis of the distribution.

Keywords: parametric tests and confidence intervals, central limit theorem, R package

1 Introduction

When you are interested in testing a variance parameter for a large sample in the non Gaussian framework, it is not easy to find a test implemented in the standard statistical software. In fact, we could not find one! The only available tool is the chi-square variance test tailor-made to the Gaussian context. This test is however commonly used by practicians even if the Gaussian assumption fails. We applied it to a data set of size $n = 1000$ with an empirical distribution very different from a normal distribution. The p-value of the chi-square variance test with alternative hypothesis $\mathbf{H_1}$: $\sigma^2 < 1$ is 4.79% which leads us to accept the alternative hypothesis at level $\alpha = 5\%$. Is it reasonable to use this test when we know that it cannot be used in the non Gaussian framework because of its sensitivity to departures from normality, *e.g.* Box (1953)?

From a mathematical point of view, one may wonder why no alternative test has been yet implemented since the asymptotic properties of the sample variance are well-known. Things are much easier in the sample mean study because the one sample t-test is known to be robust to departures from normality for large samples, *e.g.* Ozgur and Strasser (2004). This results from a direct application of the central limit theorem. The same remarks are valid when comparing the two-sample t-test (difference of means test) which is robust for large samples and the Fisher test (ratio of variances test) which is not.

In the statistical framework, one may be a simple user, a tool developer, a theoretician or any combination of the above. There are natural interactions between these different communities and it is expected that their knowledge should be shared, above all for tasks that have now become very basic. It is well-known for a theoretician, see *e.g.* Casella and Berger (1990), that a common method for constructing a large sample test statistic may be based on an estimator that has an asymptotic normal distribution. Suppose we wish to test a hypothesis about a parameter θ, and $\widehat{\theta}_n$ is some estimator of θ based on a sample of size n. If we can prove some form of the central limit theorem to show that as $n \to +\infty$, $(\widehat{\theta}_n - \theta)/\widehat{\sigma}_{\widehat{\theta}} \xrightarrow{d} \mathcal{N}(0,1)$ where $\widehat{\sigma}^2_{\widehat{\theta}}$ is a convergent (in probability) estimate of $Var(\widehat{\theta}_n)$, then one has the basis for an approximate test. This scheme based on the central limit theorem will be called the CLT procedure. We have already specified that we did not find any alternative to the chi-square variance test for testing a variance when the normality assumption fails. On the contrary, the problem of the robustness (to departures from normality) of tests for comparing two (or more) variances has been widely treated in the literature, see *e.g.* Box (1953), Conover et al. (1981), Tiku et al. (1986), Pan (1999) and the references therein. Some alternative procedures to the Fisher test are implemented in `R`: the Bartlett test (`bartlett.test`), the Fligner test (`fligner.test`), the Levene test (`levene.test` available in the `lawstat` package), etc. However to our best knowledge, for large samples, simple alternatives based on the CLT procedure have never been proposed or implemented.

The main objective of this paper is to propose a unified framework, based on the CLT procedure, for large samples to test various parameters such as the mean, the variance, the difference or ratio of means or variances. This approach also allows direct derivation of asymptotic confidence intervals. Tests and confidence intervals are then implemented in our new `R` package, called `asympTest`. This modest contribution also solves the problem of finding a robust (to non-normality) alternative to the chi-square variance test for large samples. It also provides a very simple alternative to the Fisher test. However, note that the purpose of this paper is not to compare our tests to their competitors in terms of power. Finally, a first course of statistical inference usually presents mean tests in both Gaussian and asymptotical frameworks and variance tests restricted to the Gaussian case. The unified approach presented here is is very similar to the classical t-test from a mathematical point of view and gives us the opportunity to propose a more complete teaching framework with no additional difficulty.

The paper is organized as follows. Section 2 deals with the mathematical concepts and describes our main notation. We also propose a mathematical explanation of the reason why the chi-square variance test and the Fisher test are not appropriate even when the sample size is very large. Finally, a general framework is also proposed that allows us to derive some asymptotic statistical tests for the mean, the variance and the difference (and ratio) of means or variances. In Section 3, the `R` package `asympTest` is presented. It notably includes the procedures described in the previous section. Finally, Sections 4, 5 and 6 are devoted to some discussions and the proofs of our results.

2 Mathematical development

2.1 Notation

For one-sample tests, let us denote by $\mathbf{Y} = (Y_1, \ldots, Y_n)$ a sample of n independent and identically distributed random variables with mean μ and variance σ^2. These parameters are classically estimated by

$$\widehat{\mu}(\mathbf{Y}) = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad \widehat{\sigma^2}(\mathbf{Y}) = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \widehat{\mu}(\mathbf{Y}))^2 .$$

In the two-sample context, let $\mathbf{Y^{(1)}} = \left(Y_1^{(1)}, \ldots, Y_{n^{(1)}}^{(1)}\right)$ and $\mathbf{Y^{(2)}} = \left(Y_1^{(2)}, \ldots, Y_{n^{(2)}}^{(2)}\right)$ denote two independent samples of $n^{(1)}$ and $n^{(2)}$ random variables with respective means $\mu^{(1)}$ and $\mu^{(2)}$ and variances $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$. We also define the following parameters and their estimated versions by denoting $\mathbf{Y} = \left(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}\right)$:

- Difference of (weighted) means: $d_\mu = \mu^{(1)} - \rho \times \mu^{(2)}$ ($\rho \in \mathbb{R}$) and $\widehat{d_\mu}\left(\mathbf{Y}\right) = \widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right) - \rho \times \widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)$.
- Difference of (weighted) variances: $d_{\sigma^2} = \sigma^2_{(1)} - \rho \times \sigma^2_{(2)}$ ($\rho \in \mathbb{R}$) and $\widehat{d_{\sigma^2}}\left(\mathbf{Y}\right) = \widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right) - \rho \times \widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)$.
- Ratio of means: $r_\mu = \frac{\mu^{(1)}}{\mu^{(2)}}$ and $\widehat{r_\mu}\left(\mathbf{Y}\right) = \frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)}$.
- Ratio of variances: $r_{\sigma^2} = \frac{\sigma^2_{(1)}}{\sigma^2_{(2)}}$ and $\widehat{r_{\sigma^2}}\left(\mathbf{Y}\right) = \frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}$.

The known parameter ρ is, in our definition of d_μ and d_{σ^2}, intrinsically nonnegative. But note that there is no mathematical problem to deal with negative values of ρ.

In order to compare the Gaussian framework and the general one, we propose to denote by $\mathbf{Y}^G$ (resp. $\mathbf{Y}^{(1),G}$, $\mathbf{Y}^{(2),G}$) a vector (resp. two independent vectors) of n (resp. $n^{(1)}$ and $n^{(2)}$) Gaussian random variables with mean μ (resp. $\mu^{(1)}$ and $\mu^{(2)}$) and variance σ^2 (resp. $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$). In the two-sample context, let us also denote $\mathbf{Y}^G = \left(\mathbf{Y}^{(1),G}, \mathbf{Y}^{(2),G}\right)$.

In the sequel, we will use the notation $:=$ to define some quantity. For some random variable Z and some distribution $\mathcal{L}$, $Z \rightsquigarrow \mathcal{L}$ (resp. $\overset{approx}{\rightsquigarrow} \mathcal{L}$) means that Z follows (resp. approximately follows) the distribution $\mathcal{L}$.

2.2 About the Chi-square and Fisher tests

In this section, we concentrate on parameters σ^2 and r_{σ^2}. The classical statistics of the chi-square test of variance and the Fisher test of ratio of variances are defined by

$$\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}) := (n-1)\frac{\widehat{\sigma^2}\left(\mathbf{Y}\right)}{\sigma^2} \qquad \text{and} \qquad \Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}) := \frac{\widehat{r_{\sigma^2}}\left(\mathbf{Y^{(1)}}, \mathbf{Y^{(2)}}\right)}{r_{\sigma^2}}.$$

The notation $\Lambda^G_{\widehat{\sigma^2},\sigma^2}$ expresses that the statistic is a measure of the departure of $\widehat{\sigma^2}$ from σ^2 in the Gaussian framework. When the data are Gaussian, it is well-known that

$$\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}^G) \rightsquigarrow \chi^2(n-1) \qquad \text{and} \qquad \Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}^G) \rightsquigarrow \mathcal{F}\left(n^{(1)}-1, n^{(2)}-1\right).$$

However, as shown in Fig. 1, both results become untrue (even approximately) under non-normality assumption. The theoretical reason may be explained as follows. Let $n^{(1)} = n$ and assume that there exists $\alpha > 0$ such that $n^{(2)} = \alpha n^{(1)}$. Assume also that Y (resp. $Y^{(1)}$ and $Y^{(2)}$) has a finite kurtosis $k := E((Y-\mu)^4)/Var(Y)$ (resp. $k^{(1)}$ and $k^{(2)}$). Then, as $n \to +\infty$,

$$\frac{\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}) - (n-1)}{\sqrt{n-1}} \xrightarrow{d} \mathcal{N}(0, k-1) \tag{1}$$

$$\sqrt{n}\left(\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}) - 1\right) \xrightarrow{d} \mathcal{N}\left(0, k^{(1)} - 1 + \frac{k^{(2)}-1}{\alpha}\right). \tag{2}$$

This result is a consequence of a more general one stated in Section 2.3 and proved in Section 6. Equations (1) and (2) lead to the following approximations

$$\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}) \overset{approx}{\rightsquigarrow} \mathcal{N}\left(n-1,(k-1)(n-1)\right) \quad \text{and} \quad \Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}) \overset{approx}{\rightsquigarrow} \mathcal{N}\left(1, \frac{k^{(1)}-1}{n^{(1)}} + \frac{k^{(2)}-1}{n^{(2)}}\right).$$

When data are Gaussian ans when n is large, one obtains the well-known approximations

$$\chi^2(n-1) \simeq \mathcal{N}\left(n-1, 2(n-1)\right) \qquad \text{and} \qquad \mathcal{F}\left(n^{(1)}, n^{(2)}\right) \simeq \mathcal{N}\left(1, \frac{2}{n^{(1)}} + \frac{2}{n^{(2)}}\right)$$

since $k = k^{(1)} = k^{(2)} = 3$ We can underline that $\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y})$ (resp. $\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y})$) and $\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}^G)$ (resp. $\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}^G)$) differ in terms of asymptotical variances. More precisely, the gap between the two frameworks is essentially governed by the kurtosis. Indeed, as $n \to +\infty$

$$\frac{Var(\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}))}{Var(\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}^G))} \to \frac{k-1}{2} \qquad \text{and} \qquad \frac{Var(\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}))}{Var(\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}^G))} \to \frac{k^{(1)}-1+\frac{k^{(2)}-1}{\alpha}}{2\left(1+\frac{1}{\alpha}\right)}.$$

Tab. 1 proposes the computations of these asymptotic ratios for different distributions. This allows the reader to assess the risk of using the classical statistics $\Lambda^G_{\widehat{\sigma^2},\sigma^2}$ and $\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}$ under the non-normality assumption even when the size of the sample is large.

	$Y \overset{d}{=} Y^{(1)} \overset{d}{=} Y^{(2)} \rightsquigarrow \mathcal{L}$		
Test	$\mathcal{L} = \chi^2(\nu)$	$\mathcal{L} = \mathcal{E}(\lambda)$	$\mathcal{L} = \mathcal{U}([a,b])$
Variance test	$1+\frac{6}{\nu}$	4	$\frac{2}{5}$
Ratio of variances test	$1+\frac{6}{\nu}$	4	$\frac{2}{5}$

Table 1: Ratio of asymptotic variances (non gaussian/gaussian) $\frac{k-1}{2}$ and $\frac{k^{(1)}-1+\frac{k^{(2)}-1}{\alpha}}{2\left(1+\frac{1}{\alpha}\right)}$ in the case where $k = k^{(1)} = k^{(2)}$ and $\alpha = 1$.

2.3 Large sample tests based on the central limit theorem

The parameters σ^2, d_μ and d_{σ^2} can be viewed as particular means. Therefore, the idea (widely used in asymptotic theory) is to design asymptotic tests and confidence intervals thanks to the central limit theorem. The variables $\widehat{r_\mu}(\mathbf{Y}) - r_\mu$ and $\widehat{r_{\sigma^2}}(\mathbf{Y}) - r_{\sigma^2}$ can also be expressed in terms of means to which a central limit theorem can be applied. In order to unify asymptotic results, we define θ as one of the parameters μ, σ^2, d_μ, d_{σ^2}, r_μ and r_{σ^2}. By applying a central limit theorem, the law of large numbers and Slutsky's theorem (see Section 6), one obtains, as $n \to +\infty$,

$$\widehat{\Delta}_{\widehat{\theta},\theta}(\mathbf{Y}) := \frac{\widehat{\theta}(\mathbf{Y}) - \theta}{\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})} \xrightarrow{d} \mathcal{N}(0,1), \tag{3}$$

where $\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})$ is the standard error of $\widehat{\theta}(\mathbf{Y})$. The assumptions under which the central limit theorem can be applied and the definition of $\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})$ are stated in Tab. 2. `R` functions have been implemented to evaluate the standard errors of the estimates of θ, see Tab. 3.

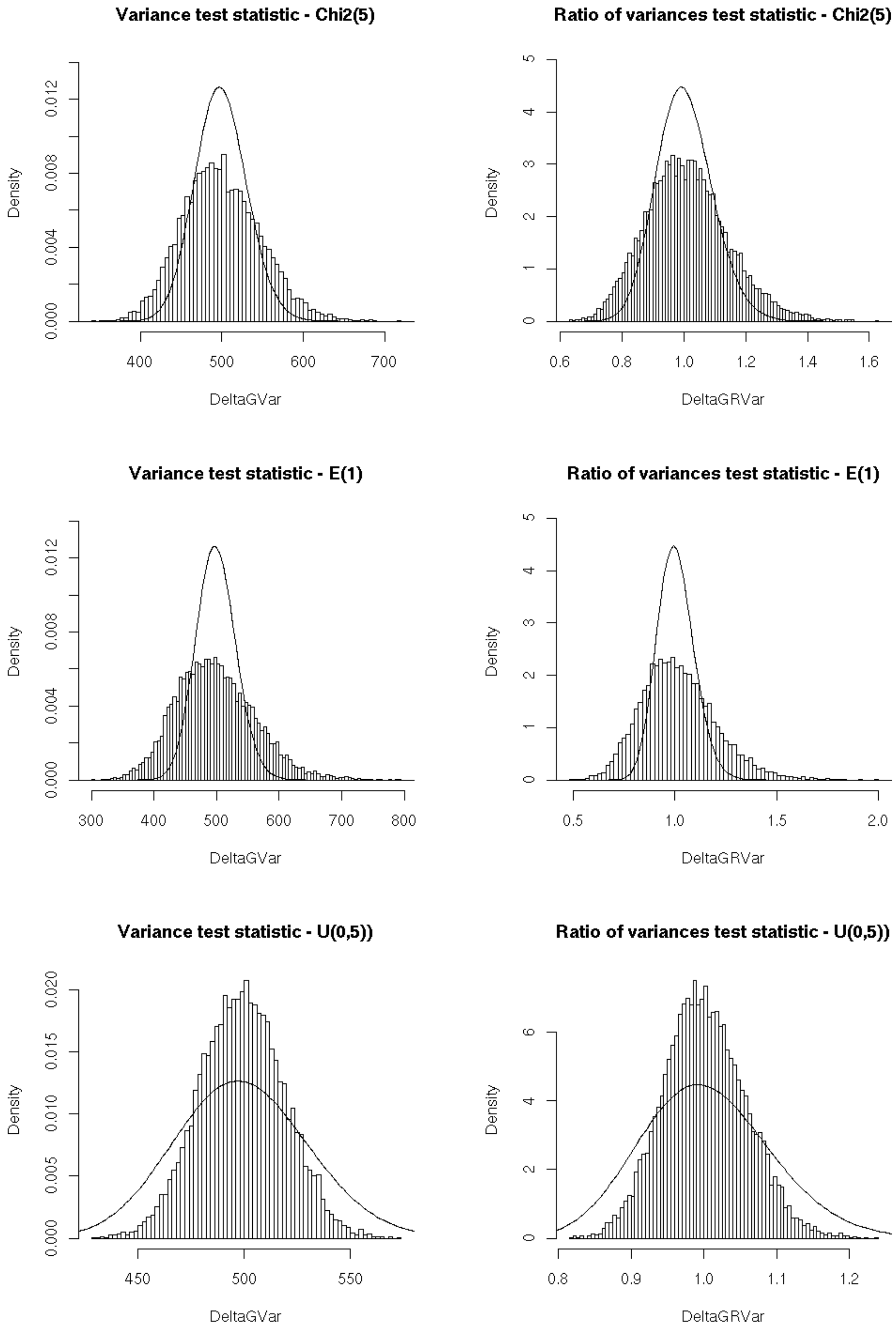


Figure 1: Histograms of $m = 10000$ replications of test statistics of variance test (left) and ratio of variance tests (right) in the Gaussian context. The simulation has been done as follows: $n = n^{(1)} = n^{(2)} = 500$, $Y \stackrel{d}{=} Y^{(1)} \stackrel{d}{=} Y^{(2)} \rightsquigarrow \chi^2(5)$ (top), $\mathcal{E}(1)$ (middle) and $\mathcal{U}([0,5])$ (bottom).

θ	Assumptions	$\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})$
μ	$E\left(Y_i^2\right) < +\infty$	$\sqrt{\frac{\widehat{\sigma^2}(\mathbf{Y})}{n}}$
σ^2	$E\left(Y_i^4\right) < +\infty$	$\sqrt{\frac{\widehat{\sigma^2_{\ddot{Y}}}\left(\ddot{\mathbf{Y}}\right)}{n}}$
d_μ	$E\left(\left(Y_i^{(j)}\right)^2\right) < +\infty,\ j = 1, 2$	$\sqrt{\frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{n^{(1)}} + \rho^2 \times \frac{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n^{(2)}}}$
d_{σ^2}	$E\left(\left(Y_i^{(j)}\right)^4\right) < +\infty,\ j = 1, 2$	$\sqrt{\frac{\widehat{\sigma^2_{\ddot{Y}^{(1)}}}\left(\ddot{\mathbf{Y}}^{(\mathbf{1})}\right)}{n^{(1)}} + \rho^2 \times \frac{\widehat{\sigma^2_{\ddot{Y}^{(2)}}}\left(\ddot{\mathbf{Y}}^{(\mathbf{2})}\right)}{n^{(2)}}}$
r_μ	$\mu^{(2)} \neq 0,\ E\left(\left(Y_i^{(j)}\right)^2\right) < +\infty,\ j = 1, 2$	$\frac{1}{\left\lvert\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)\right\rvert}\sqrt{\frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{n^{(1)}} + \widehat{r_\mu}\left(\mathbf{Y}\right)^2 \times \frac{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n^{(2)}}}$
r_{σ^2}	$\sigma^2_{(2)} \neq 0,\ E\left(\left(Y_i^{(j)}\right)^4\right) < +\infty,\ j = 1, 2$	$\frac{1}{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}\sqrt{\frac{\widehat{\sigma^2_{\ddot{Y}^{(1)}}}\left(\ddot{\mathbf{Y}}^{(\mathbf{1})}\right)}{n^{(1)}} + \widehat{r_{\sigma^2}}\left(\mathbf{Y}\right)^2 \frac{\widehat{\sigma^2_{\ddot{Y}^{(2)}}}\left(\ddot{\mathbf{Y}}^{(\mathbf{2})}\right)}{n^{(2)}}}$

Table 2: Standard errors of estimates of θ. For the sake of simplicity we denote by $\ddot{Y} := (Y-\mu)^2$, $\ddot{\mathbf{Y}} := \left(\mathbf{Y} - \widehat{\mu}\left(\mathbf{Y}\right)\right)^2$, $\sigma^2_{\ddot{Y}} := Var\left(\ddot{Y}\right)$ and $\widehat{\sigma^2_{\ddot{Y}}}\left(\ddot{\mathbf{Y}}\right) := \frac{1}{n-1}\sum_{i=1}^n \left(\left(Y_i - \widehat{\mu}\left(\mathbf{Y}\right)\right)^2 - \widehat{\sigma^2}\left(\mathbf{Y}\right)\right)^2$.

θ	Dataset(s)	$\widehat{\sigma_{\widehat{\theta}}}(\mathbf{y})$ in `R`
μ	`y`	`seMean(y)`
σ^2	`y`	`seVar(y)`
d_μ	`y1, y2`	`seDMean(y1,y2,rho=1)`
d_{σ^2}	`y1, y2`	`seDVar(y1,y2,rho=1)`
r_μ	`y1, y2`	`seRMean(y1,y2)`
r_{σ^2}	`y1, y2`	`seRVar(y1,y2)`

Table 3: Standard errors of estimates of θ in `R`.

Remark 1 *The figures Fig. 2, Fig. 3 and Fig. 4 allow the reader to illustrate the mathematical result (3).*

Remark 2 *The asymptotic result (3) allows us to easily construct statistical hypothesis tests and confidence intervals, see e.g. Casella and Berger (1990) p. 385 for development.*

Remark 3 *The alternative hypothesis* $\mathbf{H_1}$ *comparing the ratio of means or variances to some reference value* r_0 *may be expressed in terms of the comparison of the weighted differences* $d_\mu := \mu^{(1)} - r_0\mu^{(2)}$ *or* $d_{\sigma^2} := \sigma^2_{(1)} - r_0\sigma^2_{(2)}$ *to 0, the value of* ρ *being fixed to* r_0*.*

3 Using asympTest

The `R` package `asympTest` consists of a main function `asymp.test` and six auxiliary ones designed to compute standard errors of estimates of different parameters, see Tab. 3. The auxiliary functions should not be very useful for the user, except if he/she wants to compute himself/herself the confidence interval. The function `asymp.test` has been written in the same spirit as the `R` functions `t.test` or `var.test`. The arguments of `asymp.test`, its value and the resulting outputs are inspired from the ones of `t.test` or `var.test`. In particular, the function `asympt.test` returns an object of class "htest" (which is the general class of test objects in `R`, see R Development Core Team (2004)).

The main arguments of the function `asymp.test` are:

- `x`: vector of data values.
- `y`: optional vector of data values.
- `parameter`: parameter under testing, must be one of "mean", "var", "dMean", "dVar", "rMean", "rVar".
- `alternative`: alternative hypothesis, must be one of "two.sided" (default), "greater" or "less".
- `reference`: reference value of the parameter under the null hypothesis.
- `conf.level`: confidence level of the interval (default is 0.95). The type-one error is then fixed to 1-`conf.level`.
- `rho`: optional parameter (only used for parameters "dMean" and "dVar") for penalization (or enhancement) of the contribution of the second parameter.

The user may only specify the first letters of the parameter or alternative.

In order to illustrate this function, let us consider the `iris` data available in `R`. This famous (Fisher's or Anderson's Fisher (1935); Anderson (1935)) data set gives the measurements (in centimeters) of the four variables sepal length and width, and petal length and width, for 50 flowers from each species of iris: setosa, versicolor, and virginica.

```
> data(iris)
> attach(iris)
> names(iris)
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
> levels(iris$Species)
[1] "setosa"     "versicolor" "virginica"
```

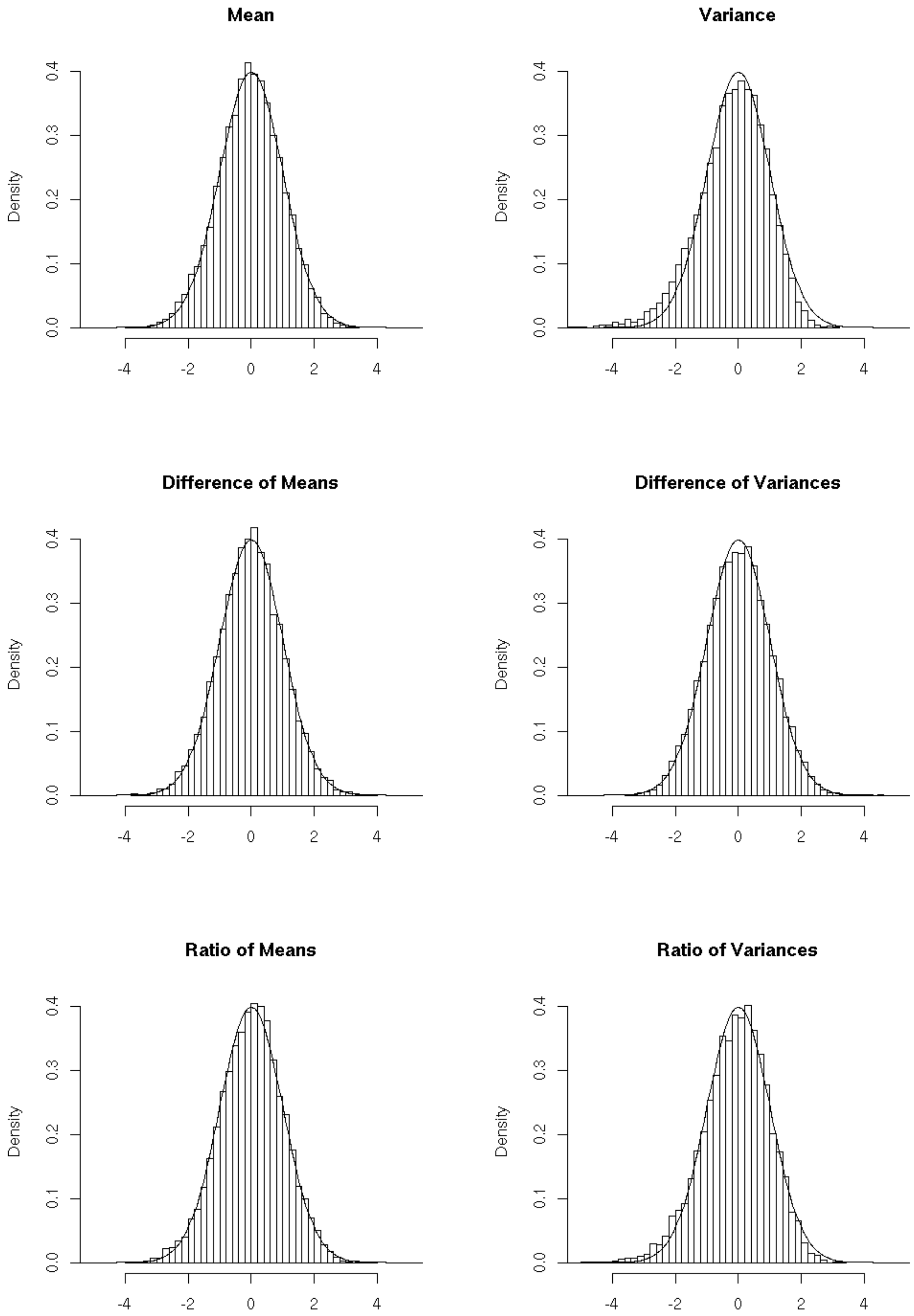


Figure 2: Histograms of $m = 10000$ replications of $\frac{\widehat{\theta}(\mathbf{Y})-\theta}{\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})}$ for $\theta = \mu$, σ^2, d_μ, d_{σ^2}, r_μ and r_{σ^2}. The simulation has been done as follows: $n = n^{(1)} = n^{(2)} = 500$, $Y^{(1)} \rightsquigarrow \chi^2(5)$, $Y^{(2)} \rightsquigarrow \chi^2(5)$.

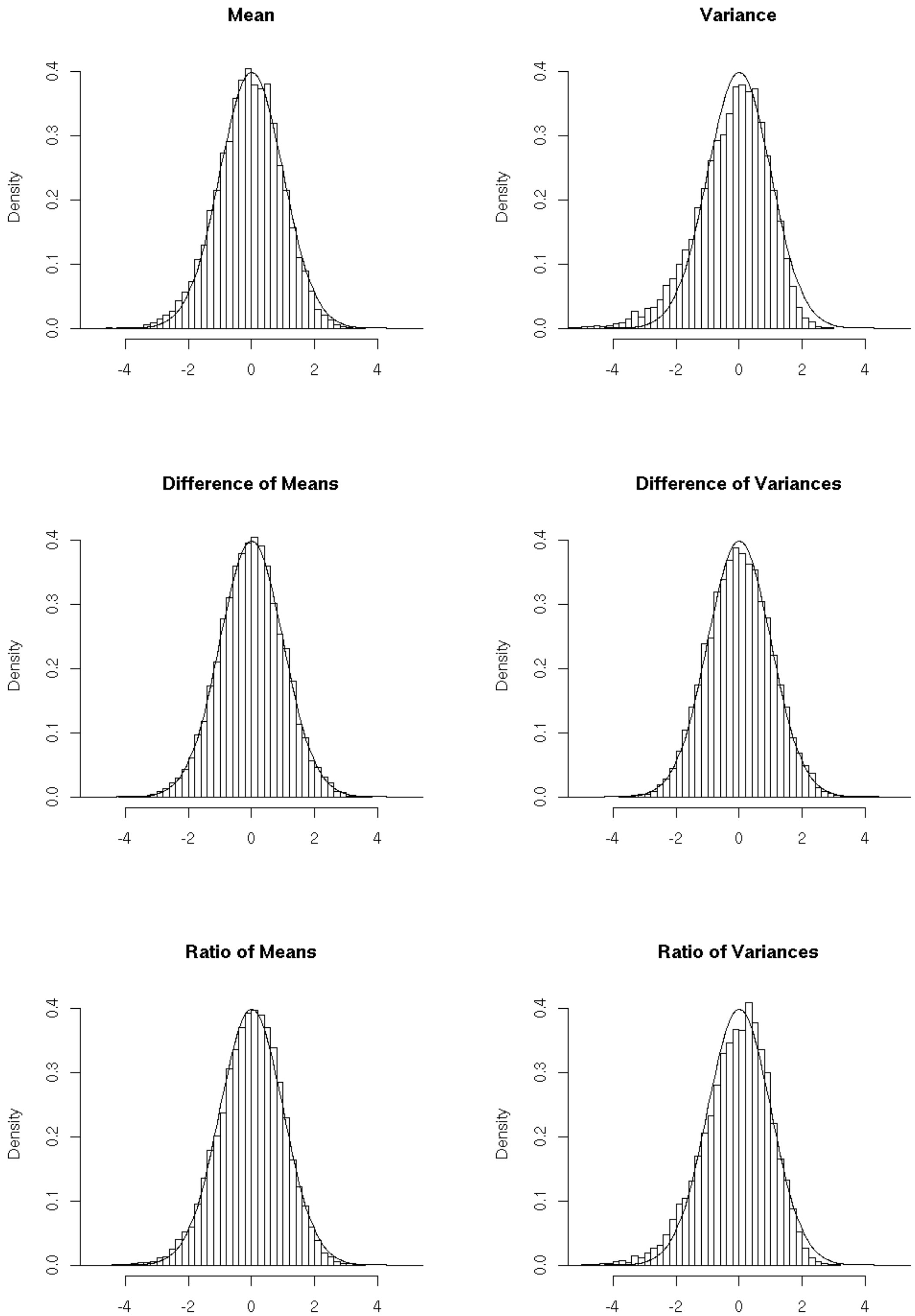


Figure 3: Histograms of $m = 10000$ replications of $\frac{\widehat{\theta}(\mathbf{Y})-\theta}{\widehat{\sigma_{\widehat{\theta}}(\mathbf{Y})}}$ for $\theta = \mu$, σ^2, d_μ, d_{σ^2}, r_μ and r_{σ^2}. The simulation has been done as follows: $n = n^{(1)} = n^{(2)} = 500$, $Y^{(1)} \rightsquigarrow \mathcal{E}(1)$, $Y^{(2)} \rightsquigarrow \mathcal{E}(1)$.

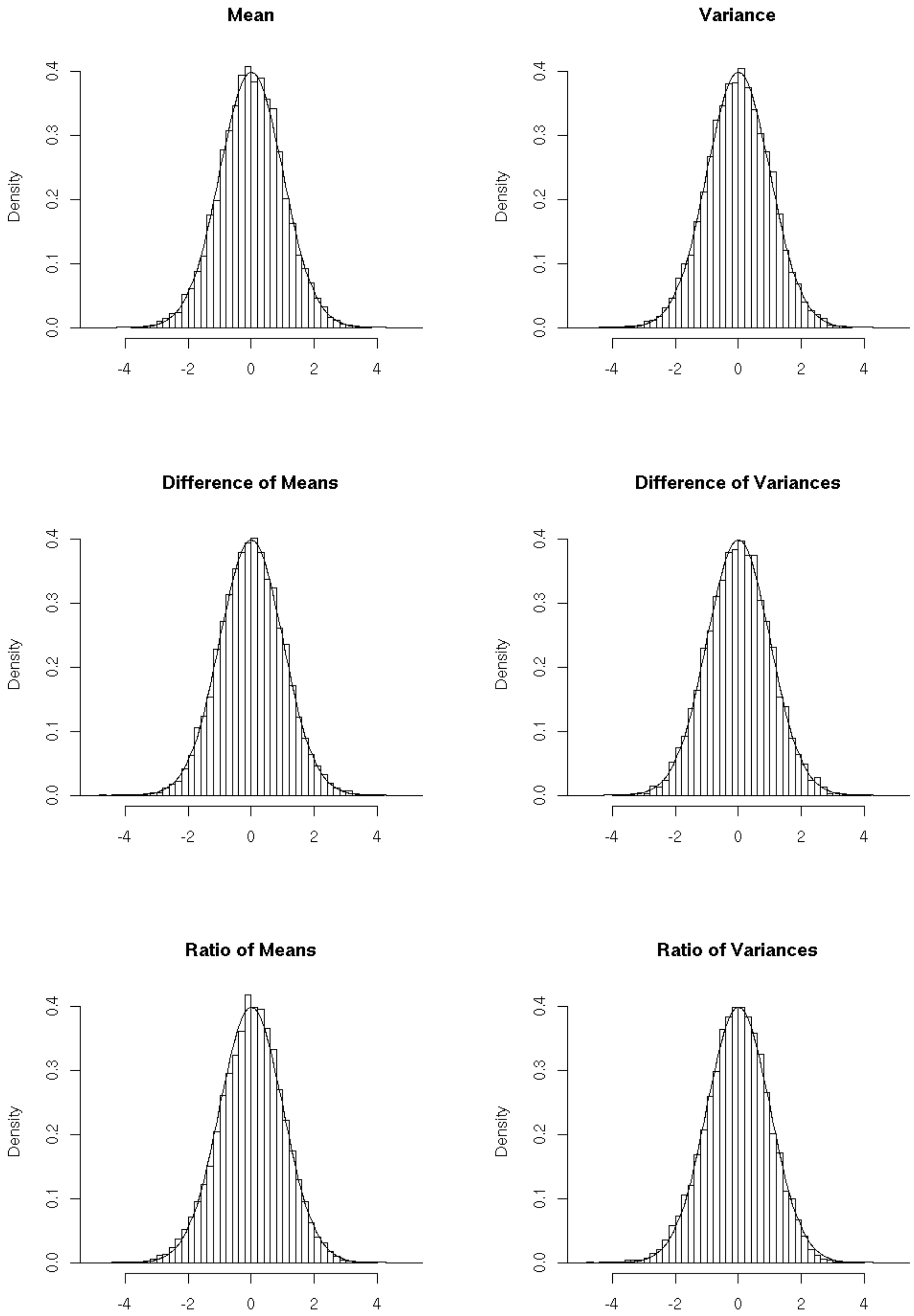


Figure 4: Histograms of $m = 10000$ replications of $\frac{\widehat{\theta}(\mathbf{Y})-\theta}{\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})}$ for $\theta = \mu$, σ^2, d_μ, d_{σ^2}, r_μ and r_{σ^2}. The simulation has been done as follows: $n = n^{(1)} = n^{(2)} = 500$, $Y^{(1)} \rightsquigarrow \mathcal{U}([0,5])$, $Y^{(2)} \rightsquigarrow \mathcal{E}(0.5)$.

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
setosa	0.4595	0.2715	0.0548	0.0000
versicolor	0.4647	0.3380	0.1585	0.0273
virginica	0.2583	0.1809	0.1098	0.0870

The following table presents the p-values of all the Shapiro-Wilk normality tests for the different variables and the three species.

Let us concentrate on the variable Petal.Width for which the Gaussian assumption seems to be wrong for each one of the three species. The empirical means and variances are given below:

```
> by(Petal.Width,Species,function(e) c(mean=mean(e),var=var(e)))
Species: setosa
      mean        var
0.24600000 0.01110612
------------------------------------------------------------
Species: versicolor
      mean        var
1.32600000 0.03910612
------------------------------------------------------------
Species: virginica
      mean        var
2.02600000 0.07543265
```

Is the mean petal width of setosa species less than 0.5 ?

```
> require(asympTest)
> asymp.test(Petal.Width[Species=="setosa"],par="mean",alt="l",ref=0.5)

        One-sample asymptotic mean test

data:  Petal.Width[Species == "setosa"]
statistic = -17.0427, p-value < 2.2e-16
alternative hypothesis: true mean is less than 0.5
95 percent confidence interval:
      -Inf 0.2705145
sample estimates:
 mean
0.246
```

Is the mean petal width of virginica species larger than the versicolor one ?

```
> asymp.test(Petal.Width[Species=="virginica"],
+ Petal.Width[Species=="versicolor"],"dMean","g",0)

        Two-sample asymptotic difference of means test

data:  Petal.Width[Species == "virginica"] and Petal.Width[Species == "versicolor"]
statistic = 14.6254, p-value < 2.2e-16
alternative hypothesis: true difference of means is greater than 0
```

```
95 percent confidence interval:
 0.621274      Inf
sample estimates:
difference of means
                0.7
```

Is the mean petal width of virginica species 4 times larger than the setosa one ?

```
> asymp.test(Petal.Width[Species=="virginica"],
+ Petal.Width[Species=="setosa"],"rMean","g",4)

        Two-sample asymptotic ratio of means test

data:  Petal.Width[Species == "virginica"] and Petal.Width[Species == "setosa"]
statistic = 8.0936, p-value = 3.331e-16
alternative hypothesis: true ratio of means is greater than 4
95 percent confidence interval:
 7.374946      Inf
sample estimates:
ratio of means
      8.235772
```

This may also be done via a difference of weighted means test.

```
> asymp.test(Petal.Width[Species=="virginica"],
+ Petal.Width[Species=="setosa"],"dMean","g",0,rho=4)

        Two-sample asymptotic difference of (weighted) means test

data:  Petal.Width[Species == "virginica"] and Petal.Width[Species == "setosa"]
statistic = 14.6447, p-value < 2.2e-16
alternative hypothesis: true difference of (weighted) means is greater than 0
95 percent confidence interval:
 0.9249653       Inf
sample estimates:
difference of (weighted) means
                         1.042
```

4 Type I error risks

4.1 Comparison between classical and asymptotic variance tests

In the context of large samples, two simulation studies are proposed in order to show the lack of reliability of the classical tests for variance parameters compared with the asymptotic tests studied in this paper. For each of the following examples, 10000 simulations of samples of size $n = 1000$ have been performed.

1. Let us consider testing $\mathbf{H_0} : \sigma^2 = 1$ versus $\mathbf{H_1} : \sigma^2 < 1$ with data sampled from distribution $\mathcal{E}(1)$ (i.e., under $\mathbf{H_0}$).

   In the case where $\alpha = 5\%$, the probability of accepting the alternative hypothesis is 8.91% for the asymptotic test and 20.99% for the chi-square test.

	false	true
false	0.7901	0.0000
true	0.1208	0.0891

Table 4: Acceptance of $\mathbf{H_1}$ for the chi-square test (rows) versus the asymptotic test (columns)

2. Let us consider test $\mathbf{H_0} : \sigma^2_{(1)} = \sigma^2_{(2)}$ versus $\mathbf{H_1} : \sigma^2_{(1)} \neq \sigma^2_{(2)}$ with data sampled from distribution $\mathcal{U}([0, 5])$ for both samples (i.e., under $\mathbf{H_0}$).

	false	true
false	0.9498	0.0481
true	0.0000	0.0021

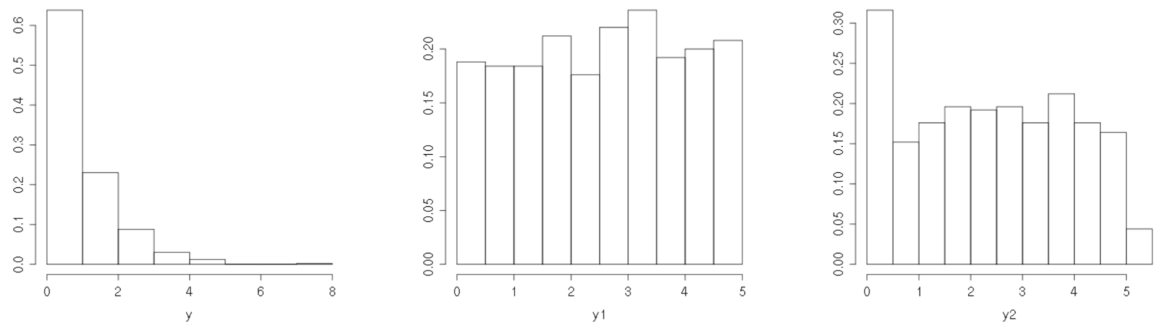
Table 5: Acceptance of $\mathbf{H_1}$ for the Fisher test (rows) versus the asymptotic test (columns)

In the case where $\alpha = 5\%$, the probability of accepting the alternative hypothesis is 5.02% for the asymptotic test and 0.21% for the Fisher test.

In both cases, the probabilities for Type I errors are worse for the classical tests than for the corresponding well-suited asymptotic tests. This is a direct consequence of the previous results summarized by figure Fig. 1.

4.2 Back to the example of the introduction

Now, one may wonder what the consequences of these previous results are for pratical purposes. Let us consider again the example presented in the introduction. In order to illustrate the two-samples case, we propose a second example. In the following `R` outputs, the data of the first (resp. second) example are denoted by `y` (resp. `y1` and `y2`). These samples have a size n=1000 and their empirical distributions, proposed below, do not seem to fit normal distributions.



The following output provides the p-value of the chi-square test for the first example.

```
> pchisq((length(y)-1)*var(y)/1,length(y)-1)
[1] 0.04785152
```

Due to the apparent non normality of the data, one may prefer to apply the corresponding asymptotic test:

```
> asymp.test(y,par="var",alt="l",ref=1)

        One-sample asymptotic variance test

data:  y
statistic = -0.9771, p-value = 0.1643
alternative hypothesis: true variance is less than 1
95 percent confidence interval:
     -Inf 1.070430
sample estimates:
 variance
0.8969455
```

The two decisions do not match, but since the empirical variance is 0.8969455, one may think that σ^2 is slightly inferior to 1. In which case, we have to be cautious because our sample might be of the same kind as the 12.08% of Table 4.

The following output provides the p-value of the Fisher test for the second example.

```
> var.test(y1,y2)

        F test to compare two variances

data:  y1 and y2
F = 0.8874, num df = 499, denom df = 499, p-value = 0.1825
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7444324 1.0578390
sample estimates:
ratio of variances
         0.8874061
```

Due to the apparent non normality of the data, one may prefer to apply the corresponding asymptotic test:

```
> asymp.test(y1,y2,"dVar")

        Two-sample asymptotic difference of variances test

data:  y1 and y2
statistic = -2.0925, p-value = 0.03639
alternative hypothesis: true difference of variances is not equal to 0
95 percent confidence interval:
 -0.49988046 -0.01635305
sample estimates:
difference of variances
             -0.2581168
```

The two decisions do not match, but since the empirical variances are 2.034341 and 2.292458, one may think that $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$ are slightly different. In which case, we have to be cautious because our sample might be of the same kind as the 4.81% of Table 5.

5 Discussion

We have presented an R package implementing large sample tests for various parameters. The interesting point is that each test statistic can be written in the same form, as follows:

$$\widehat{\Delta}_{\widehat{\theta},\theta}(\mathbf{Y}) := \frac{\widehat{\theta}\left(\mathbf{Y}\right)-\theta}{\widehat{\sigma_{\widehat{\theta}}}\left(\mathbf{Y}\right)}.$$

This form clearly expresses the departure of the estimate from the true parameter, normalized by some quantity measuring the precision of the estimate. This approach is then attractive and easy to present. In the Gaussian framework, the one and two-sample t-tests follow this idea whereas the chi-square variance test and the Fisher test do not. One may wonder if it is possible to embed the classical Gaussian framework within this formalism. More precisely, for an estimate $\widehat{\theta}\left(\mathbf{Y}\right)$ of some parameter θ with standard deviation $\sigma_{\widehat{\theta}} = \sqrt{Var(\widehat{\theta}\left(\mathbf{Y}\right))}$, let us propose the test statistic

$$\Delta_{\widehat{\theta},\theta}(\mathbf{Y}) := \frac{\widehat{\theta}\left(\mathbf{Y}\right)-\theta}{\sigma_{\widehat{\theta}}} \text{ or } \frac{\widehat{\theta}\left(\mathbf{Y}\right)-\theta}{\widetilde{\sigma}_{\widehat{\theta}}} \tag{4}$$

in the case where $\sigma_{\widehat{\theta}}$ or possibly some known asymptotic equivalent $\widetilde{\sigma}_{\widehat{\theta}}$ of $\sigma_{\widehat{\theta}}$ only depends on θ, or

$$\widehat{\Delta}_{\widehat{\theta},\theta}(\mathbf{Y}) := \frac{\widehat{\theta}\left(\mathbf{Y}\right)-\theta}{\widehat{\sigma_{\widehat{\theta}}}\left(\mathbf{Y}\right)} \tag{5}$$

otherwise, where $\widehat{\sigma_{\widehat{\theta}}}\left(\mathbf{Y}\right)$ is a consistent estimate of $\sigma_{\widehat{\theta}}$.

In the large sample framework, the parameters μ, σ^2, d_μ, r_μ, d_{σ^2} and r_{σ^2} fall into the case of equation (5) and the corresponding statistic approximately follows a $\mathcal{N}(0,1)$ distribution. In this same context, the proportion parameter falls into the case of equation (4) with $\sigma_{\widehat{p}} = \sqrt{p(1-p)/n}$.

In the Gaussian framework, the parameters μ and d_μ leading to the one-sample and two-sample t-tests also fall into the case of equation (5). Let us now concentrate on parameters σ^2 and r_{σ^2} (corresponding to the chi-square variance test and the Fisher test). It is known that the variance of $\widehat{\sigma^2}\left(\mathbf{Y^G}\right)$ and an asymptotic equivalent of the variance of $\widehat{r_{\sigma^2}}\left(\mathbf{Y^G}\right)$ are of the form

$$\sigma^2_{\widehat{\sigma^2}} = \frac{1}{n}\left(\dot{\mu}_4 - \frac{n-3}{n-1}\dot{\mu}_2^2\right) \quad \text{and} \quad \widetilde{\sigma}^2_{\widehat{r_{\sigma^2}}} = \frac{1}{n^{(1)}}\frac{\dot{\mu}_4^{(1)} - (\dot{\mu}_2^{(1)})^2}{(\dot{\mu}_2^{(2)})^2} + \frac{1}{n^{(2)}} r^2_{\sigma^2}\frac{\dot{\mu}_4^{(2)} - (\dot{\mu}_2^{(2)})^2}{(\dot{\mu}_2^{(2)})^2}$$

where $\dot{\mu}_k = E\Big((Y-E(Y))^k\Big)$ is the $k-$th centered moment of Y (with our notation $\dot{\mu}_2 = \sigma^2$). For Gaussian variables, $\dot{\mu}_4 = 3(\dot{\mu}_2)^2$ which leads to

$$\sigma^2_{\widehat{\sigma^2}} = \frac{2}{n-1}\sigma^4 \quad \text{and} \quad \widetilde{\sigma}^2_{\widehat{r_{\sigma^2}}} = 2r^2_{\sigma^2}\left(\frac{1}{n^{(1)}}+\frac{1}{n^{(2)}}\right).$$

Now, in order to build a test, let us give the distributions of $\Delta^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G})$ and $\Delta^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G})$ expressed in terms of $\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G})$ and $\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G})$:

$$\Delta^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G}) := \frac{\widehat{\sigma^2}\left(\mathbf{Y^G}\right)-\sigma^2}{\sigma_{\widehat{\sigma^2}}} = \frac{\frac{\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G})}{n-1}\sigma^2 - \sigma^2}{\sigma^2\sqrt{2/(n-1)}} \rightsquigarrow \frac{\chi^2(n-1)-(n-1)}{\sqrt{2(n-1)}}$$

and

$$\Delta^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G}) := \frac{\widehat{r_{\sigma^2}}\left(\mathbf{Y^G}\right)-r_{\sigma^2}}{\widetilde{\sigma}_{\widehat{r_{\sigma^2}}}} = \frac{r_{\sigma^2}\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G})-r_{\sigma^2}}{r_{\sigma^2}\sqrt{2/n^{(1)}+2/n^{(2)}}} \rightsquigarrow \frac{\mathcal{F}(n^{(1)}-1,n^{(2)}-1)-1}{\sqrt{2/n^{(1)}+2/n^{(2)}}}.$$

One may propose a name for the previous two free distributions: centered reduced chi-square distribution and centered reduced Fisher distribution respectively denoted by $\chi^2_{cr}(\cdot)$ and $\mathcal{F}_{cr}(\cdot,\cdot)$. If these distributions were implemented such that one may evaluate quantiles and p-values, one could build two new tests directly based on $\Delta^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G})$ and $\Delta^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G})$. Of course, these two new tests would be strictly equivalent to the classical chi-square variance test and Fisher test and would then fall into the same formalism.

Let us now comment on the concept of robustness in the Gaussian framework. In the particular case of mean hypothesis testing, this robustness is expressed by the fact that $\Delta^G_{\widehat{\mu},\mu}(\mathbf{Y^G})$ is equal to $\Delta_{\widehat{\mu},\mu}(\mathbf{Y^G})$ with the same asymptotic distribution $\mathcal{N}(0,1)$. When considering the two new cases $\theta = \sigma^2$ or $\theta = r_{\sigma^2}$, $\Delta^G_{\widehat{\theta},\theta}(\mathbf{Y^G})$ is no longer equal to $\Delta_{\widehat{\theta},\theta}(\mathbf{Y^G})$ but have at least the same asymptotic distribution $\mathcal{N}(0,1)$. However, one can prove that $\Delta^G_{\widehat{\theta},\theta}(\mathbf{Y^G})$ and $\Delta_{\widehat{\theta},\theta}(\mathbf{Y^G})$ are asymptotically equivalent in probability, which may be viewed as some kind of robustness.

6 Proofs

In this section, we only prove (3). The results (1) and (2) are direct consequences. Recall that for each parameter, some assumptions are needed essentially in order to apply the central limit theorem. They are summarized in Tab. 2.

Parameter μ:

This is a direct application of the central limit theorem.

Parameter σ^2:

By definition

$$\widehat{\sigma^2}(\mathbf{Y}) - \sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\left(Y_i - \overline{Y}\right)^2 - \sigma^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left((Y_i-\mu)^2 - \sigma^2\right) - \frac{n}{n-1}\left(\overline{Y}-\mu\right)^2.$$

From the CLT, the law of large numbers and Slutsky's Theorem (see *e.g.* Ferguson (1996)), it comes that as $n \to +\infty$

$$\sqrt{n}\left(\overline{Y}-\mu\right)^2 \xrightarrow{\mathbb{P}} 0.$$

Therefore, as $n \to +\infty$,

$$\sqrt{n}\left(\widehat{\sigma^2}(\mathbf{Y}) - \sigma^2\right) \xrightarrow{d} \mathcal{N}(0, Var((Y-\mu)^2)).$$

Since $\sigma_{\widehat{\sigma^2}} := \sqrt{\frac{Var((Y-\mu)^2)}{n}}$ can be consistently estimated by $\widehat{\sigma_{\widehat{\sigma^2}}}(\mathbf{Y}) := \sqrt{\frac{\widehat{\sigma^2_{\ddot{Y}}}(\ddot{\mathbf{Y}})}{n}}$, see Tab. 2, we obtain as $n \to +\infty$

$$\widehat{\Delta}_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}) := \frac{\widehat{\sigma^2}(\mathbf{Y}) - \sigma^2}{\widehat{\sigma_{\widehat{\sigma^2}}}(\mathbf{Y})} \xrightarrow{d} \mathcal{N}(0,1).$$

Parameter d_μ:

Recall that $n = n^{(1)}$ and $n^{(2)} = \alpha n^{(1)}$. As $n \to +\infty$,

$$\begin{aligned}\sqrt{n}\left(\widehat{d_\mu}(\mathbf{Y}) - d_\mu\right) &= \sqrt{n}\left(\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right) - \mu^{(1)}\right) - \rho \times \frac{\sqrt{n\alpha}}{\sqrt{\alpha}}\left(\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right) - \mu^{(2)}\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{(1)} + \rho^2 \times \frac{\sigma^2_{(2)}}{\alpha}\right).\end{aligned}$$

Since $\sigma_{\widehat{d_\mu}} := \sqrt{\frac{\sigma^2_{(1)} + \rho^2 \times \frac{\sigma^2_{(2)}}{\alpha}}{n}}$ can be consistently estimated by

$$\sigma_{\widehat{d_\mu}}(\mathbf{Y}) := \sqrt{\frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{n^{(1)}} + \rho^2 \times \frac{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n^{(2)}}},$$

we obtain as $n \to +\infty$

$$\widehat{\Delta}_{\widehat{d_\mu}, d_\mu}(\mathbf{Y}) := \frac{\widehat{d_\mu}(\mathbf{Y}) - d_\mu}{\sigma_{\widehat{d_\mu}}(\mathbf{Y})} \xrightarrow{d} \mathcal{N}(0,1).$$

Parameter d_{σ^2}:

As $n \to +\infty$,

$$\begin{aligned}
\sqrt{n}\left(\widehat{d_{\sigma^2}}(\mathbf{Y}) - d_{\sigma^2}\right) &= \sqrt{n}\left(\widehat{\sigma^2_{(1)}(1)}\left(\mathbf{Y^{(1)}}\right) - \sigma^2_{(1)}\right) - \rho \times \frac{\sqrt{n\alpha}}{\sqrt{\alpha}}\left(\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right) - \sigma^2_{(2)}\right) \\
&\xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{\ddot{Y}^{(1)}} + \rho^2 \times \frac{\sigma^2_{\ddot{Y}^{(2)}}}{\alpha}\right).
\end{aligned}$$

Since $\sigma_{\widehat{d_{\sigma^2}}} := \sqrt{\frac{\sigma^2_{\ddot{Y}^{(1)}} + \rho^2 \frac{\sigma^2_{\ddot{Y}^{(2)}}}{\alpha}}{n}}$ can be consistently estimated by

$$\sigma_{\widehat{d_{\sigma^2}}}(\mathbf{Y}) := \sqrt{\frac{\widehat{\sigma^2_{\ddot{Y}^{(1)}}}\left(\mathbf{\ddot{Y}^{(1)}}\right)}{n} + \rho^2 \times \frac{\widehat{\sigma^2_{\ddot{Y}^{(2)}}}\left(\mathbf{\ddot{Y}^{(2)}}\right)}{n\alpha}} = \sqrt{\frac{\widehat{\sigma^2_{\ddot{Y}^{(1)}}}\left(\mathbf{\ddot{Y}^{(1)}}\right)}{n^{(1)}} + \frac{\widehat{\sigma^2_{\ddot{Y}^{(2)}}}\left(\mathbf{\ddot{Y}^{(2)}}\right)}{n^{(2)}}},$$

we obtain as $n \to +\infty$

$$\frac{\widehat{d_{\sigma^2}}(\mathbf{Y}) - d_{\sigma^2}}{\sigma_{\widehat{d_{\sigma^2}}}(\mathbf{Y})} \xrightarrow{d} \mathcal{N}(0,1).$$

Parameter r_μ:

Using Slutsky's Theorem, one may assert that as $n \to +\infty$

$$\begin{aligned}
\sqrt{n}\left(\widehat{r_\mu}(\mathbf{Y}) - r_\mu\right) &= \sqrt{n}\left(\frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)} - \frac{\mu^{(1)}}{\mu^{(2)}}\right) \\
&= \sqrt{n}\left(\frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right) - \mu^{(1)}}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)} + \mu^{(1)}\left(\frac{1}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)} - \frac{1}{\mu^{(2)}}\right)\right) \\
&\overset{n\to+\infty}{\sim} \sqrt{n}\left(\frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right) - \mu^{(1)}}{\mu^{(2)}} + \frac{\sqrt{n\alpha}}{\sqrt{\alpha}}\frac{\mu^{(1)}}{(\mu^{(2)})^2}\left(\mu^{(2)} - \widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)\right)\right) \\
&\xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2_{(1)}}{(\mu^{(2)})^2} + \left(\frac{\mu^{(1)}}{(\mu^{(2)})^2}\right)^2 \frac{\sigma^2_{(2)}}{\alpha}\right).
\end{aligned}$$

Since, $\sigma_{\widehat{r_\mu}} := \sqrt{\frac{\frac{\sigma^2_{(1)}}{(\mu^{(2)})^2} + \left(\frac{\mu^{(1)}}{(\mu^{(2)})^2}\right)^2 \frac{\sigma^2_{(2)}}{\alpha}}{n}}$ can be consistently estimated by

$$\begin{aligned}
\widehat{\sigma_{\widehat{r_\mu}}}(\mathbf{Y}) &:= \sqrt{\frac{\frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)}}{n} + \frac{\left(\frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)^2}\right)^2 \widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n\alpha}} \\
&= \frac{1}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)} \sqrt{\frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{n^{(1)}} + \widehat{r_\mu}\left(\mathbf{Y}\right) \frac{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n^{(2)}}}
\end{aligned}$$

we obtain as $n \to +\infty$,

$$\frac{\widehat{r_\mu}\left(\mathbf{Y}\right) - r_\mu}{\widehat{\sigma_{\widehat{r_\mu}}}\left(\mathbf{Y}\right)} \xrightarrow{d} \mathcal{N}(0,1).$$

Parameter r_{σ^2}:

The proof follows the ideas developed for the parameters σ^2, d_{σ^2} and r_μ, and thus is left to the reader.

Acknowledgments

We would like to thank Laurence Pierret for having kindly reread our English. She could in no way be held responsible for any remaining mistakes.

References

E. Anderson. The irises of the gaspe peninsula. *Bulletin of the American Iris Society*, 59:2–5, 1935.

G.E.P. Box. Non-normality and tests on variances. *Biometrika*, 40(3/4):318–335, 1953.

G. Casella and R.L. Berger. *Statistical Inference*. Duxbury press, Belmont, California, 1990.

W.J. Conover, M.E. Johnson, and M.M. Johnson. A comparative study of tests for homogeneity of variances with applications to the outer continental shelf bidding data. *Technometrics*, 23 (4):351–361, 1981.

T.S. Ferguson. *A course in large sample theory*. Chapman and Hall, London, UK, 1996.

R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7 (Part II):179–188, 1935.

C. Ozgur and S.E. Strasser. A study of the statistical inference criteria: can we agree on when to use z versus t ? *Decision Sciences Journal of Innovative Education*, 2(2):177–192, 2004.

G. Pan. On a levene type test for equality of two variances. *Journal of Statistical Computation and Simulation*, 63:59–71, 1999.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2004. ISBN ISBN 3-900051-00-3. URL `http://www.R-project.org/`.

M.L. Tiku, W.Y. Tan, and N. Balakrishnan. *Robust inference.* Statistics: Textbooks and Monographs, New York, USA, 1986.