

Università degli studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



EFFETTO DELLA CORREZIONE DI BARTLETT PER VERIFICHE DI IPOTESI AD ALTA DIMENSIONALITÀ

Relatore: Prof. Nicola Sartori
Dipartimento di Scienze Statistiche

Laureando: Matteo Favero
Matricola N 1061632

Anno Accademico 2016/2017

Indice

| | |
|---|-----------|
| Introduzione | 5 |
| 1 Un problema di dimensionalità | 7 |
| 1.1 Introduzione | 7 |
| 1.2 Il Test Rapporto di Verosimiglianza | 7 |
| 1.2.1 Un problema di dimensionalità | 8 |
| 2 Due alternative al test rapporto di verosimiglianza | 11 |
| 2.1 Introduzione | 11 |
| 2.2 Test Direzionale per famiglie esponenziali lineari | 12 |
| 2.3 Test di Skovgaard per famiglie esponenziali lineari | 14 |
| 3 Una proposta basata sulla correzione di Bartlett | 17 |
| 3.1 Introduzione | 17 |
| 3.2 Una procedura basata sul ricampionamento bootstrap | 17 |
| 3.3 Correzione di Bartlett | 18 |
| 3.4 Una variante della correzione di Bartlett | 20 |
| 3.5 Analisi di due dataset reali e studi di simulazione | 21 |
| 3.5.1 Introduzione | 21 |
| 3.5.2 Analisi dei dati <i>aircondit</i> | 21 |
| 3.5.3 Studio di simulazione per il confronto tra medie esponenziali | 24 |
| 3.5.4 Analisi dei dati <i>gravity</i> | 32 |
| 3.5.5 Studio di simulazione per il confronto tra varianze normali | 34 |

| | |
|------------------------------|-----------|
| Conclusioni | 41 |
| Bibliografia | 44 |
| A Codice R utilizzato | 47 |

Introduzione

Non sempre nel campo dell'inferenza statistica basata sulla verosimiglianza, e in particolare nell'ambito della verifica di ipotesi, gli strumenti più comuni a disposizione riescono a garantire dei risultati efficienti ed affidabili: ci sono infatti alcune situazioni estreme in cui è necessario ricorrere a strumenti più complessi per superare problemi che rendono gli strumenti di base inaffidabili. Il problema principale affrontato in questo studio è quello in cui la dimensione parametrica del modello è dello stesso ordine di grandezza della numerosità campionaria.

Verrà illustrato, attraverso analisi di dati reali e simulazioni, come il test log-rapporto di verosimiglianza si comporta in queste situazioni estreme, e in seguito verrà messo a confronto con alcuni test alternativi, più adatti al tipo di situazione, di cui viene valutata l'affidabilità.

Entrando nel dettaglio, nel Capitolo 1 viene richiamato brevemente il test rapporto di verosimiglianza e le sue proprietà, analizzandone l'affidabilità e presentando il problema dello sbilanciamento della dimensionalità tra parametro e numerosità del campione.

Nel Capitolo 2 vengono presentate due alternative al test log-rapporto di verosimiglianza, rispettivamente il test direzionale e il test aggiustato di Skovgaard.

Nel Capitolo 3 vengono proposte 3 diverse varianti del test log-rapporto di verosimiglianza, basate sulla tecnica di simulazione bootstrap, per risolvere il problema di partenza. Vengono poi confrontati tutti i test discussi, attraverso un'analisi di due dataset reali e due studi di simulazione, per due tipi differenti di verifica di ipotesi.

Capitolo 1

Un problema di dimensionalità

1.1 Introduzione

In questo capitolo verranno presentati brevemente alcuni degli strumenti dell'inferenza statistica basata sulla verosimiglianza che verranno poi utilizzati nei capitoli successivi. Verrà prestata particolare attenzione ad alcune delle loro problematiche, per la risoluzione delle quali verranno valutate delle proposte alternative nei capitoli successivi.

1.2 Il Test Rapporto di Verosimiglianza

Uno degli strumenti classici quando si parla di inferenza statistica parametrica basata sulla verosimiglianza è il test rapporto di verosimiglianza (Azzalini, 2008, paragrafo 4.2.1). Supponiamo di essere in un modello statistico parametrico per una osservazione y univariata o multivariata, con parametro $\theta \in \Theta \subseteq \mathbb{R}^p$, con $p > 1$, e funzione di log-verosimiglianza

$$l(\theta) = \log L(\theta) = c + \log f(y; \theta),$$

dove $f(y, \theta)$ è la densità, o funzione di probabilità, del modello.

Il test rapporto di verosimiglianza viene utilizzato per verificare il sistema di ipotesi

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1. \end{cases}$$

Il test assume normalmente la forma del rapporto tra la verosimiglianza massimizzata sotto l'ipotesi nulla e il massimo globale della verosimiglianza, anche se è di uso più comunque, nonché pratico, utilizzare la sua trasformazione monotona logaritmica

$$w(\theta_0) = 2[l(\hat{\theta}) - l(\theta_0)].$$

Questo test ha distribuzione asintotica chi quadrato, con gradi di libertà pari al numero di parametri vincolati dall'ipotesi nulla. Supponendo di avere p_0 vincoli, il test avrà dunque una distribuzione asintotica $\chi_{p_0}^2$, ed il p -value sarà ottenuto confrontando il valore osservato del test, w^{oss} , con la suddetta distribuzione:

$$\alpha_1^{oss} = Pr(\chi_{p_0}^2 > w^{oss}). \quad (1.1)$$

1.2.1 Un problema di dimensionalità

È di interesse in questo studio valutare il comportamento del test rapporto di verosimiglianza in situazioni estreme, ovvero quando la dimensione del parametro θ è dello stesso ordine di grandezza della numerosità del campione a disposizione. A tal fine consideriamo il seguente esempio:

Supponiamo di avere J campioni casuali semplici indipendenti da variabili casuali $Y_{ij} \sim Exp(1/\theta_j)$, con $i = 1, \dots, n_j$ le numerosità di ogni campione, $n = \sum_{j=1}^J n_j$ la numerosità campionaria totale e $j = 1, \dots, J$ la dimensione del parametro ($p = J$). Supponiamo inoltre di essere interessati a verificare l'ipotesi che tutti i parametri θ_j siano uguali tra di loro e cioè $H_0 : \theta_1 = \theta_2 = \dots = \theta_J$.

In condizioni di regolarità si assume che la dimensione campionaria sia sostanzialmente maggiore di quella parametrica ($n \gg J$), e in tali situazioni l'approssimazione χ_{J-1}^2 per la distribuzione nulla del test rapporto di verosimiglianza è affidabile, ossia produce p -value, utilizzando (1) che sono approssimativamente uniformemente distribuiti nell'intervallo $(0, 1)$, sotto l'ipotesi nulla. Se invece la dimensione parametrica è dello stesso ordine della dimensione campionaria, $J = O(n)$, il test rapporto di verosimiglianza tende a produrre p -value che non sono distribuiti in maniera uniforme in $(0, 1)$ sotto l'ipotesi nulla, in particolar modo tende in generale a sottostimare i suddetti p -value. Per evidenziare quanto appena detto, è stata effettuata una

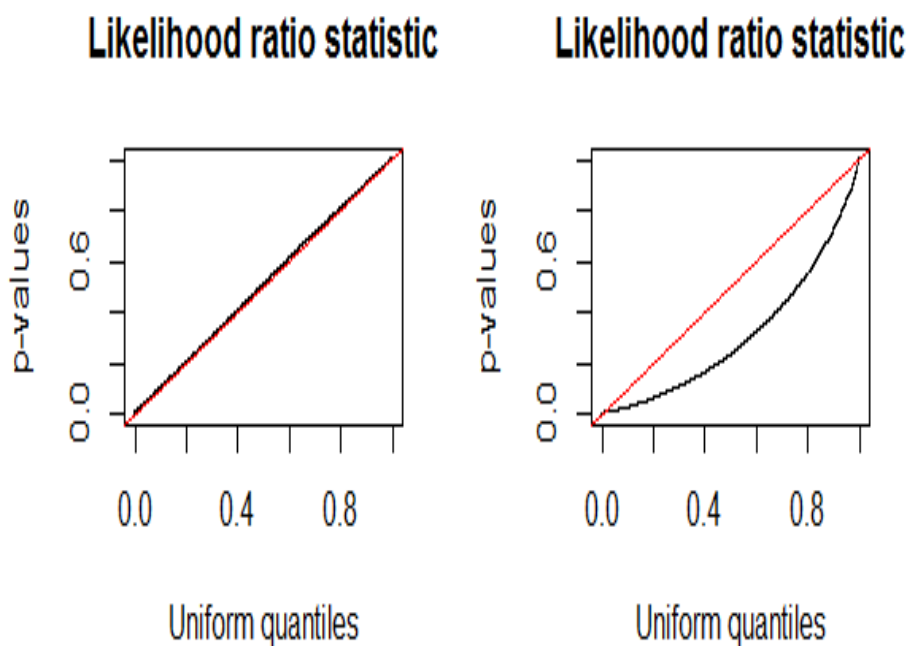


Figura 1.1: Confronto della distribuzione dei p -value al variare di n_j e J ($n_j = 500$, $J = 100$ a sinistra, $n_j = 5$, $J = 1000$ a destra)

simulazione (con $R = 10^4$ replicazioni) per ognuna delle due situazioni descritte: $n_j = 500$ e $J = 10$ nella prima, $n_j = 5$ e $J = 1000$ nella seconda. Come riportato in Figura 1.1, i p -value si distribuiscono in maniera uniforme

quando la numerosità campionaria è sostanzialmente maggiore della dimensione parametrica, mentre se la dimensione parametrica risulta equivalente alla dimensione campionaria, i p -value tendono a venire sottostimati dal test rapporto di verosimiglianza, che risulta dunque inaffidabile. Nei capitoli successivi verranno presentate delle alternative all'usuale approssimazione asintotica del test rapporto di verosimiglianza, per cercare di ottenere dei p -value uniformemente distribuiti anche in situazioni di numerosità estreme, come precedentemente riportato.

Capitolo 2

Due alternative al test rapporto di verosimiglianza

2.1 Introduzione

È stato osservato nel capitolo precedente che il test log rapporto di verosimiglianza perde la sua affidabilità quando la dimensione parametrica è dello stesso ordine di grandezza della numerosità campionaria. Questo richiede l'utilizzo di approcci differenti per garantire l'uniformità dei p -value anche in situazioni estreme. Esistono in letteratura alcune proposte per il calcolo dei p -value in verifiche d'ipotesi su un vettore di parametri chiamate test direzionali, che verranno brevemente presentate in questo capitolo. I test direzionali sono stati inizialmente introdotti da Fraser e Massam (1985) e Skovgaard (1988), e partono da una misura della distanza dall'ipotesi, per calcolare in seguito i p -value sulla base dell'ampiezza di questa distanza, condizionatamente alla direzione di allontanamento dall'ipotesi. I due approcci riportati in seguito fanno riferimento a Davison et al.(2014).

2.2 Test Direzionale per famiglie esponenziali lineari

Si assuma di essere in presenza di una famiglia esponenziale con parametro canonico $\varphi = \varphi(\theta)$ e statistica sufficiente $u = u(y)$,

$$f(y; \theta) = \exp[\varphi(\theta)^\top u(y) - K\{\varphi(\theta)\}]h(y), \quad (2.1)$$

con funzione di log verosimiglianza $l(\theta; y) = \varphi(\theta)^\top u(y) - K\{\varphi(\theta)\}$. È di interesse al fine di ottenere una quantità chiave per il calcolo dei p -value direzionali centrare la statistica sufficiente nel punto osservato $u^0 = u(y^0)$, quindi si definisce $s = u - u^0$ e

$$l(\theta; s) = \varphi(\theta)^\top s + l^0(\theta) = \varphi(\theta)^\top (u - u^0) + l(\theta; u^0). \quad (2.2)$$

Si assuma anche che entrambi i parametri di interesse e disturbo siano lineari in φ , quindi si può ridefinire $\varphi = \theta = (\psi, \lambda)$ e

$$l(\varphi; s) = \psi^\top s_1 + \lambda^\top s_2 + l^0(\psi, \lambda). \quad (2.3)$$

La distribuzione condizionata di s_1 dato s_2 dipende solo da φ , e ha anch'essa la forma di una famiglia esponenziale:

$$f(s_1|s_2; \psi) = \exp\{\psi^\top s_1 - \tilde{K}(\varphi)\}\tilde{h}_2(s_1), \quad (2.4)$$

anche se \tilde{K} e \tilde{h}_2 raramente possono essere calcolate esplicitamente. Il condizionamento al valore osservato di s_2 equivale a fissare la stima di massima verosimiglianza condizionata sotto H_ψ , ossia $\hat{\varphi} = (\psi, \hat{\lambda}_\psi)$, indicata anche con $\hat{\varphi}_\psi^0$ quando calcolata in y^0 . Questa approssimazione alla densità condizionata può essere espressa come

$$f(s; \psi)ds = c \exp[l(\hat{\varphi}_\psi^0; s) - l\{\hat{\varphi}(s); s\} \times |J_{\varphi\varphi}\{\hat{\varphi}(s); s\}|^{-1/2} ds, s \in \mathcal{L}^0, \quad (2.5)$$

dove c normalizza la densità condizionata e \mathcal{L}^0 è il piano d -dimensionale definito fissando $s_2 = 0$, oppure fissando $\hat{\lambda}_\Psi = \hat{\lambda}_\Psi^0$. La stima $\hat{\varphi}(s)$ è la soluzione in φ dell'equazione score da (2.3), e $J_{\varphi\varphi}(\varphi; s) = -\partial^2 l(\varphi; s) / \partial \varphi \partial \varphi^\top$.

La densità (2.5) è alla base dell'inferenza sul parametro ψ , e può essere utilizzata per ricavare un test log rapporto di verosimiglianza per H_ψ , con distribuzione asintotica χ_d^2 sotto l'ipotesi nulla. Per ottenere p -value più accurati ci si condiziona alla direzione che si forma unendo s_0 e s_ψ , che è il valore di s per cui $\hat{\varphi} = \hat{\varphi}_\psi^0$. Viene quindi calcolato un test unidimensionale misurando la grandezza di s , che porterà a rifiutare l'ipotesi nulla tanto più grande sarà la distanza tra s^0 (valore associato al campione) e s_ψ (valore associato all'ipotesi nulla). Dalla (2.3) si ricava

$$s_\psi = -l_\varphi^0(\hat{\varphi}_\psi^0) = \begin{bmatrix} l_\psi^0(\hat{\varphi}_\psi^0) \\ 0 \end{bmatrix}, \quad (2.6)$$

e il valore di s_ψ dipende da y^0 . Riparametrizzando in t si ottiene

$$s(t) = s_\psi + t(s^0 - s_\psi) = (1 - t)s_\psi.$$

Le stime di massima verosimiglianza $\hat{\varphi}(s)$ della (2.5) variano al variare di $s(t)$, tracciando una curva nello spazio parametrico che passa sia attraverso la stima di massima verosimiglianza vincolata ($t = 0$) che attraverso la stima di massima verosimiglianza ($t = 1$). È possibile ora riscrivere la (2.5) riparametrizzata in t :

$$h(t; \psi) = f\{s(t); \psi\} = c \exp[l\{\hat{\varphi}_\psi^0; s(t)\} - l\{\hat{\varphi}\{s(t)\}; s(t)\}] \times |J_{\varphi\varphi}[\hat{\varphi}\{s(t)\}; s(t)]|^{-1/2}. \quad (2.7)$$

Lo stimatore di massima verosimiglianza vincolata è dato da

$\hat{\varphi}_\psi = \operatorname{argsup}_{\psi(\varphi)=\varphi} l^0(\varphi)$. La divergenza dall'ipotesi nulla sarà quindi calcolata utilizzando la distribuzione condizionata di $\|s\|$ dato il vettore $a = s/\|s\|$,

ottenuta da (2.5) effettuando un cambio di variabile da s a $(\|s\|, a)$. Lo Jacobiano di questa trasformazione è proporzionale a t^{d-1} , con t che rappresenta la distanza radiale, mentre a il vettore dei $d - 1$ angoli. Dalla (2.7) si ricava

$$t^{d-1}h(t, \psi) \propto t^{d-1} \exp[l\{\hat{\varphi}_\psi^0; s(t)\} - l\{\hat{\varphi}; s(t)\}] \times |J_{\varphi\varphi}(\hat{\varphi})|^{-1/2}. \quad (2.8)$$

Il test direzionale calcola il p -value attraverso $h(t; \Psi)$, come la probabilità che $s(t)$ almeno tanto distante da s_Ψ quanto il valore osservato 0, condizionatamente al fatto di essere nella parte con $t > 0$ di $s(t)$. Il p -value è dunque

$$p(\Psi) = \frac{\int_1^{t_{max}} t^{d-1}h(t; \Psi)dt}{\int_0^{t_{max}} t^{d-1}h(t; \Psi)dt}, \quad (2.9)$$

con $t = 0$ che corrisponde a $s = s_\Psi$, mentre $t = 1$ corrisponde al valore osservato $s^0 = 0$. Il valore t_{max} è il massimo valore per cui esiste la stima di massima verosimiglianza. L'accuratezza teorica del test deriva da quella della (2.5), quindi c'è un errore relativo al peggio di $O(n^{-1})$, mentre in caso di risposte continue l'errore relativo è di $O(n^{-3/2})$.

2.3 Test di Skovgaard per famiglie esponenziali lineari

Skovgaard (2001) propone un aggiustamento diretto del test w introdotto nel Capitolo 1. Questo aggiustamento assume la forma di

$$w^*(\psi) = w(\psi)\left(1 - \frac{\log \gamma(\psi)}{w(\psi)}\right)^2, \quad (2.10)$$

dove $\gamma(\psi)$ è dato da

$$\gamma(\psi) = \frac{\{(s - s_\Psi)^\top J_{\varphi\varphi}^{-1}(\hat{\varphi}_\Psi)(s - s_\Psi)\}^{d/2}}{w^{d/2-1}(\hat{\varphi} - \hat{\varphi}_\Psi)^\top (s - s_\Psi)} \left\{ \frac{|J_{\varphi\varphi}(\hat{\varphi}_\Psi)|}{|J_{\varphi\varphi}(\hat{\varphi})|} \right\}^{1/2}. \quad (2.11)$$

Il p -value è ottenuto con $s = 0$ in corrispondenza di $y = y^0$, e questa correzione (2.11) permette di aggiustare il test w in qualsiasi punto dello spazio parametrico.

Capitolo 3

Una proposta basata sulla correzione di Bartlett

3.1 Introduzione

In questo capitolo verranno presentate alcune proposte per contrastare il problema della elevata dimensionalità parametrica nel test rapporto di verosimiglianza basate su di una modifica del test rapporto di verosimiglianza stesso. Verranno prese in considerazione tre proposte differenti, collegate tra loro dall'uso della tecnica di ricampionamento bootstrap. Infine, le proposte presentate finora verranno applicate nello studio di alcuni dataset reali, e in alcuni studi di simulazione per verificarne le prestazioni.

3.2 Una procedura basata sul ricampionamento bootstrap

La prima proposta che viene presentata è basata sulla tecnica bootstrap. Il bootstrap è una tecnica per fare inferenza basata su aspetti computazionali, in particolare di simulazione. Il principio su cui si basa la tecnica bootstrap è quello di valutare le proprietà di una determinata tecnica statistica utilizzando una stima dell'ignoto processo generatore dei dati, ottenuta dal campione osservato. Spesso, questo viene fatto attraverso simulazione e cioè generando

diversi campioni dalla distribuzione stimata dei dati e ricalcolando la quantità di interesse, come ad esempio un test, su ognuno di questi in modo da poterne valutare la distribuzione. In questo caso, si utilizzerà il bootstrap, e in particolare il bootstrap parametrico, per calcolare migliori approssimazioni analitiche per la distribuzione di w o per calcolare direttamente un p -value approssimato. Nel dettaglio, la tecnica bootstrap viene utilizzata per simulare B campioni bootstrap sotto l'ipotesi nulla stimata, a partire dai dati osservati. Su B campioni bootstrap viene in seguito calcolato il test rapporto di verosimiglianza per il parametro di interesse, test che produce un valore osservato w^b , con $b = 1, \dots, B$. Essendo a questo punto in possesso del valore osservato del test originale w^{oss} e dei B valori osservati dei test bootstrap w^b , il p -value del test con procedura bootstrap sarà ottenuto calcolando la media del numero di volte che un valore osservato bootstrap w^b è maggiore del valore osservato originale w^{oss} , ovvero

$$\alpha_2^{oss} = \frac{1}{B} \sum_{b=1}^B I(w^b > w^{oss}). \quad (3.1)$$

Attraverso l'utilizzo di questa tecnica di simulazione, siamo in grado di arginare il problema della scarsa numerosità campionaria e riusciamo ad ottenere un risultato affidabile anche quando la dimensione parametrica è molto elevata. In questo studio è stato deciso di utilizzare $B = 2000$ campioni bootstrap, il che significa che per il calcolo del p -value vengono estratti 2000 campioni casuali sotto l'ipotesi nulla stimata su cui viene calcolato il test rapporto di verosimiglianza. È stato inoltre tenuto conto non solo dei p -value ottenuto con 2000 campioni bootstrap, ma anche di quelli ottenuti con rispettivamente 500 e 1000 campioni bootstrap, al fine di valutare l'impatto di B sull'accuratezza delle varie approssimazioni.

3.3 Correzione di Bartlett

La seconda proposta che viene presentata è una modifica diretta al test $w(\theta)$ presentato nel primo capitolo, e viene chiamata correzione di Bartlett

(Pace e Salvan, 2001, paragrafo 6.8). Questa correzione viene applicata in modo da far coincidere il valore atteso del test corretto w_B con quello di un $\chi_{p_0}^2$, che come visto in precedenza è la distribuzione asintotica del test $w(\theta)$ di partenza. Così facendo si va a centrare la media del test con la media della distribuzione asintotica. Il valore atteso di un chi quadrato sono i suoi gradi libertà ($E(\chi_d^2) = d$), da qui il test rapporto di verosimiglianza con correzione di Bartlett assume la forma

$$w_B(\theta) = \frac{w(\theta)}{E_{H_0}\{w(\theta)\}} p_0 .$$

In questo modo si ottiene che $E(w_B) \simeq p_0$, mentre il p -value si calcola allo stesso modo del test rapporto di verosimiglianza originale

$$\alpha_3^{oss} = Pr\{\chi_{p_0}^2 > w_B^{oss}(\theta)\}. \quad (3.2)$$

Il solo aggiustamento del valore atteso del test w ha un importante effetto su tutta la distribuzione del test, che risulta essere centrata con la distribuzione del $\chi_{p_0}^2$, avendo la stessa media.

È di interesse notare che il valore atteso $E_{H_0}(w(\theta))$ non è in genere calcolabile in maniera analitica. Uno dei modi possibili per ottenere questa quantità è tramite sviluppi asintotici, via che però può essere lunga, complessa, e non produce sempre risultati ottimali. Una strada alternativa è invece quella di sostituire il valore atteso con una sua stima, ottenibile ricorrendo alla simulazione via bootstrap parametrico. Simulando B campioni di numerosità n sotto l'ipotesi nulla stimata, possono essere calcolati i test $w^b(\theta)$, con $b = 1, \dots, B$, e da qui ricavare il valore atteso desiderato è relativamente semplice:

$$\hat{E}_{H_0}(w(\theta)) = \frac{1}{B} \sum_{b=1}^B w_b(\theta)$$

e di conseguenza il test diventa:

$$\hat{w}_B(\theta) = \frac{w(\theta)}{\hat{E}_{H_0}\{w(\theta)\}} p_0 .$$

3.4 Una variante della correzione di Bartlett

L'ultima proposta che viene presentata riprende il concetto di base della proposta precedente, con la differenza che non solo la media viene fatta coincidere con quella della distribuzione asintotica $\chi_{p_0}^2$, ma viene fatta coincidere anche la varianza. Conoscendo che la varianza di un Chi-quadrato corrisponde a due volte i suoi gradi di libertà, il sistema di riferimento per questa proposta diventa:

$$\begin{cases} E_{H_0}\left\{\frac{w(\theta)b}{a}\right\} = b, \\ \text{Var}_{H_0}\left\{\frac{w(\theta)b}{a}\right\} = 2b \end{cases}$$

che porta alle soluzioni, per a e b :

$$\begin{cases} a = E_{H_0}\{w(\theta)\} \\ b = \frac{2\{E_{H_0}(w(\theta))\}^2}{\text{Var}_{H_0}\{w(\theta)\}} \end{cases}$$

Come illustrato in precedenza, il valore atteso e la varianza del test $w(\theta)$ possono essere sostituiti con la loro stima, calcolata via bootstrap parametrico. In questo caso, la stima della varianza di $w(\theta)$ è data da

$$\widehat{\text{Var}}_{H_0}(w(\theta)) = \frac{1}{B} \sum_{b=1}^B \{w^b - \hat{E}_{H_0}(w^b)\}^2$$

In questo modo vengono centrate sia la media che la varianza, e la distribuzione asintotica del test

$$w_{BC}^{\hat{}}(\theta) = \frac{w_{oss}(\theta)\hat{b}}{\hat{a}}$$

diventa un $\chi_{\hat{b}}^2$, per costruzione del test. I valori \hat{a} e \hat{b} indicano le stime dei valori a e b presentati sopra, calcolati via bootstrap parametrico.

Il calcolo del p -value per questo test con correzione di Bartlett centrato assume dunque la forma

$$\alpha_4^{oss} = Pr(\chi_{\hat{b}}^2 > w_{BC}^{oss}(\theta)). \quad (3.3)$$

La stima del valore atteso e della varianza di $w(\theta)$ possono essere effettuate sfruttando gli stessi B campioni bootstrap utilizzati per il calcolo del test di tipo bootstrap (3.2), rendendo quindi possibile il calcolo di tre diversi p-value senza dover ripetere tre volte la stessa simulazione bootstrap.

Ci si attende in generale che un numero B relativamente basso ($B = 500$) sia meno influente per (3.2) e (3.3) che non per (3.1).

3.5 Analisi di due dataset reali e studi di simulazione

3.5.1 Introduzione

In questa sezione vengono presentate delle applicazioni pratiche dei test presentati in precedenza, su dei dati reali. Sono stati presi in considerazione due dataset, *aircondit* e *gravity* (presenti entrambi nel pacchetto R *boot*), per ognuno dei quali sono state effettuate un'analisi, rispettivamente confronto tra medie esponenziali e confronto tra varianze normali (Davison et al., 2014), ed in seguito uno studio di simulazione per verificare l'accuratezza dei test proposti.

3.5.2 Analisi dei dati *aircondit*

Il dataset *aircondit* contiene i dati relativi agli intervalli di tempo tra dei guasti negli impianti di condizionamento di 10 Boeing 720. Il dataset è suddiviso in due colonne, una riportante gli intervalli di tempo tra i guasti, e una riportante il numero del Boeing interessato. Siamo quindi in presenza di $J = 10$ gruppi di numerosità diversa tra loro, dei quali vengono riportati media e numerosità campionaria nella Tabella 3.1 e la distribuzione empirica al variare dei gruppi in Figura 3.1.

| | g1 | g2 | g3 | g4 | g5 | g6 | g7 | g8 | g9 | g10 |
|-----------|-------|-------|--------|--------|------|-------|-------|-----|--------|-----|
| \bar{y} | 95.69 | 83.51 | 121.26 | 130.85 | 59.6 | 76.81 | 64.12 | 200 | 108.08 | 82 |
| n | 23 | 29 | 15 | 14 | 30 | 27 | 24 | 9 | 12 | 16 |

Tabella 3.1: Media e numerosità campionaria per ognuno dei 10 gruppi del dataset *aircondit*

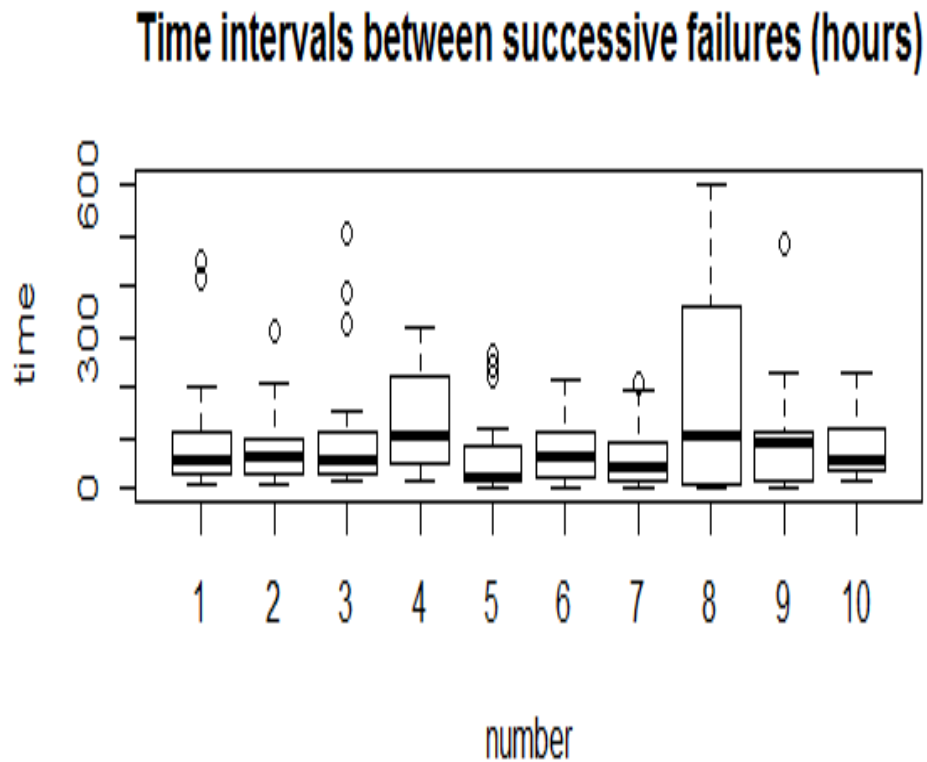


Figura 3.1: Boxplot dei dati *aircondit* al variare dei gruppi

Si ipotizza che i dati siano un campione di osservazioni indipendenti da $Y_{ij} \sim \exp(1/\theta_j)$, $j = 1, \dots, J$ e $i = 1, \dots, n_j$. L'ipotesi che si vuole verificare è quella di omogeneità dei parametri nei g gruppi, $\theta_1 = \dots = \theta_J$, contro l'ipotesi

alternativa che almeno una delle equazioni non sia verificata. La funzione di log-verosimiglianza per il parametro $\theta = (\theta_1, \dots, \theta_J)$ assume la forma

$$l(\theta; y) = \sum_{i=1}^J (-u_i \theta_i + n_i \log \theta_i),$$

con $u_i = n_i \bar{y}_i = \sum_{j=1}^{n_i} y_{ij}$, il parametro canonico è $\varphi = -\theta$, e la statistica sufficiente è $u = (u_1, \dots, u_J)$. L'ipotesi assume quindi la forma

$$H_0 = \theta_1 = \dots = \theta_J$$

Le stime di massima verosimiglianza libere e vincolate sono rispettivamente

$$\hat{\theta} = (\bar{y}_1^{-1}, \dots, \bar{y}_J^{-1}), \hat{\theta}^0 = (\bar{y}^{-1}, \dots, \bar{y}^{-1})$$

dove \bar{y} è la media globale tra tutti i gruppi ($\bar{y} = \sum_{i=1}^J n_i \bar{y}_i / \sum_{i=1}^J n_i$). La statistica log-rapporto di verosimiglianza è

$$w = 2 \sum_{i=1}^J n_i \log(\hat{\theta}_i / \hat{\theta}^0) = 2 \sum_{i=1}^J n_i \log(\bar{y} / \bar{y}_i)$$

che, come accennato in precedenza, ha una distribuzione asintotica χ_{J-1}^2 sotto l'ipotesi nulla.

Su questo dataset sono stati applicati i vari test proposti e discussi finora, rispettivamente il test log rapporto di verosimiglianza w , i test direzionali e w^* di Skovgaard, e le tre varianti bootstrap, con correzione di Bartlett e con correzione di Bartlett centrata. Di questi test sono stati dunque raccolti e confrontati tra di loro i vari p -value. Per i tre test coinvolgenti la procedura bootstrap, sono stati riportati i p -value relativi a tre diversi valori di B (numero di campioni bootstrap utilizzati), al fine di valutare l'impatto di B nell'accuratezza dei test.

| | lrt | dir | sko | bo1 | bo2 | bo3 |
|---------------|-----|-----|-----|-----|-----|-----|
| p -value(%) | 2.2 | 1.9 | 2.2 | 2.7 | 3.0 | 2.4 |

Tabella 3.2: p -value(%) osservati nel dataset *aircondit*. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

| | ba1 | ba2 | ba3 | bace1 | bace2 | bace3 |
|---------------|-----|-----|-----|-------|-------|-------|
| p -value(%) | 1.8 | 1.7 | 1.8 | 2.3 | 1.9 | 1.9 |

Tabella 3.3: p -value(%) osservati nel dataset *aircondit*. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

Notiamo che tutti i test producono p -value decisamente bassi e portano al rifiuto dell'ipotesi nulla. I p -value prodotti dai test direzionale e di Skovgaard sono molto simili a quello prodotto dal test w , mentre per quanto riguarda i test proposti nel Capitolo 3, quello basato sul ricampionamento bootstrap produce p -value più elevati, mentre le due varianti della correzione di Bartlett producono p -value inferiori a quello del test w originale, anche se tutti portano comunque al rifiuto dell'ipotesi nulla. Questo risultato era atteso, visto che le numerosità n_j sono abbastanza grandi e $J = 10$. Nello specifico, notiamo che B non sembra avere un effetto sistematico sui p -value, anche se tutti i p -value con $B = 2000$ risultano inferiori ai p -value con $B = 500$.

3.5.3 Studio di simulazione per il confronto tra medie esponenziali

Per verificare l'affidabilità dei test proposti, ossia che in condizioni di regolarità producano p -value uniformemente distribuiti nell'intervallo $(0, 1)$, sotto l'ipotesi nulla, è stato condotto uno studio di simulazione per il confronto tra medie esponenziali. A questo scopo si sono simulati campioni casuali da una distribuzione esponenziale di parametro θ fissato ($\theta = 0.01$), divisi in g gruppi di numerosità n . Sono state esaminate diverse combinazioni di n e g , per verificare l'adattamento dei test proposti a situazioni più o meno estreme. Nello specifico, le combinazioni utilizzate sono $(n = 5, g = 100)$, $(n = 10, g = 100)$, $(n = 5, g = 1000)$, e per ogni combinazione sono state effettuate $R = 10000$ replicazioni. Per ognuna delle combinazioni verrà presentata la distribuzione empirica simulata.

| Nominal (%) | 1.0 | 2.5 | 5.0 | 10.0 | 25.0 | 50.0 | 75.0 | 90.0 | 95.0 | 97.5 | 99.0 |
|----------------|-----|-----|-----|------|------|------|------|------|------|------|------|
| LRT | 2.1 | 4.7 | 8.2 | 14.9 | 32.4 | 58.4 | 82.0 | 93.4 | 96.6 | 98.4 | 99.4 |
| Bootstrap1 | 1.0 | 2.9 | 5.3 | 10.0 | 24.6 | 49.4 | 74.8 | 90.2 | 94.8 | 97.4 | 98.7 |
| Bootstrap2 | 1.1 | 2.7 | 5.3 | 10.0 | 24.6 | 49.4 | 74.7 | 90.3 | 94.8 | 97.3 | 98.8 |
| Bootstrap3 | 1.1 | 2.7 | 5.3 | 10.0 | 24.4 | 49.4 | 74.9 | 90.4 | 94.9 | 97.4 | 98.8 |
| Bartlett1 | 1.1 | 2.7 | 5.2 | 10.0 | 24.6 | 49.4 | 74.9 | 90.4 | 94.9 | 97.4 | 98.9 |
| Bartlett2 | 1.1 | 2.7 | 5.2 | 10.1 | 24.5 | 49.4 | 74.8 | 90.4 | 94.9 | 97.4 | 98.9 |
| Bartlett3 | 1.1 | 2.7 | 5.2 | 10.0 | 24.5 | 49.4 | 74.9 | 90.4 | 94.9 | 97.4 | 98.9 |
| Bartlett alt.1 | 1.1 | 2.8 | 5.3 | 10.1 | 24.6 | 49.4 | 74.9 | 90.4 | 95.0 | 97.4 | 98.8 |
| Bartlett alt.2 | 1.1 | 2.8 | 5.3 | 10.1 | 24.6 | 49.4 | 74.9 | 90.4 | 94.9 | 97.3 | 98.9 |
| Bartlett alt.3 | 1.2 | 2.7 | 5.2 | 10.0 | 24.6 | 49.4 | 74.9 | 90.4 | 94.9 | 97.4 | 98.9 |
| directional | 1.1 | 2.8 | 5.0 | 10.0 | 24.8 | 49.2 | 74.8 | 90.2 | 94.7 | 97.4 | 99.0 |
| Skovgaard | 0.2 | 0.4 | 0.9 | 2.0 | 6.6 | 19.1 | 40.1 | 62.6 | 75.2 | 83.1 | 90.4 |
| sim-error | 0.1 | 0.2 | 0.2 | 0.3 | 0.4 | 0.5 | 0.4 | 0.3 | 0.2 | 0.2 | 0.1 |

Tabella 3.4: Distribuzione empirica simulata (%) dei p -value per verificare l'omogeneità dei parametri esponenziali, con $n = 5$, $g = 100$ e $R = 10000$ replicazioni. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

| Nominal (%) | 1.0 | 2.5 | 5.0 | 10.0 | 25.0 | 50.0 | 75.0 | 90.0 | 95.0 | 97.5 | 99.0 |
|----------------|-----|-----|-----|------|------|------|------|------|------|------|------|
| LRT | 1.4 | 3.4 | 6.4 | 12.3 | 28.8 | 54.2 | 78.8 | 92.0 | 96.2 | 98.1 | 99.3 |
| Bootstrap1 | 1.0 | 2.7 | 5.0 | 10.0 | 24.9 | 49.9 | 75.2 | 90.1 | 95.0 | 97.5 | 98.9 |
| Bootstrap2 | 1.0 | 2.6 | 4.9 | 10.1 | 24.8 | 50.0 | 75.2 | 89.9 | 95.1 | 97.5 | 98.9 |
| Bootstrap3 | 0.9 | 2.6 | 5.0 | 10.0 | 24.8 | 49.9 | 75.1 | 90.0 | 95.1 | 97.6 | 99.0 |
| Bartlett1 | 0.9 | 2.5 | 5.0 | 10.0 | 25.0 | 49.8 | 75.1 | 90.1 | 95.1 | 97.5 | 99.1 |
| Bartlett2 | 0.9 | 2.6 | 5.0 | 10.0 | 24.9 | 50.0 | 75.3 | 90.1 | 95.1 | 97.5 | 99.1 |
| Bartlett3 | 0.9 | 2.6 | 5.0 | 10.1 | 24.9 | 50.0 | 75.2 | 90.1 | 95.2 | 97.5 | 99.1 |
| Bartlett alt.1 | 1.0 | 2.6 | 5.1 | 10.0 | 25.0 | 49.9 | 75.2 | 90.1 | 95.2 | 97.6 | 99.0 |
| Bartlett alt.2 | 0.9 | 2.6 | 5.1 | 10.1 | 24.9 | 50.0 | 75.2 | 90.0 | 95.1 | 97.5 | 99.1 |
| Bartlett alt.3 | 0.9 | 2.6 | 5.0 | 10.1 | 24.9 | 50.0 | 75.2 | 90.1 | 95.1 | 97.5 | 99.1 |
| directional | 1.0 | 2.5 | 4.9 | 9.9 | 25.0 | 49.7 | 75.1 | 90.3 | 95.2 | 97.6 | 99.1 |
| Skovgaard | 0.4 | 0.9 | 2.1 | 4.5 | 14.0 | 32.9 | 58.6 | 78.8 | 87.5 | 92.7 | 96.6 |
| sim-error | 0.1 | 0.2 | 0.2 | 0.3 | 0.4 | 0.5 | 0.4 | 0.3 | 0.2 | 0.2 | 0.1 |

Tabella 3.5: Distribuzione empirica simulata (%) dei p -value per verificare l'omogeneità dei parametri esponenziali, con $n = 10$, $g = 100$ e $R = 10000$ replicazioni. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

| Nominal (%) | 1.0 | 2.5 | 5.0 | 10.0 | 25.0 | 50.0 | 75.0 | 90.0 | 95.0 | 97.5 | 99.0 |
|----------------|-----|------|------|------|------|------|------|------|------|------|------|
| LRT | 5.5 | 11.1 | 18.4 | 29.1 | 51.5 | 76.3 | 91.6 | 97.6 | 99.0 | 99.7 | 99.9 |
| Bootstrap1 | 1.0 | 2.4 | 4.8 | 9.9 | 24.8 | 48.9 | 74.4 | 89.6 | 94.5 | 97.1 | 98.8 |
| Bootstrap2 | 0.9 | 2.4 | 4.7 | 9.9 | 24.9 | 49.0 | 74.6 | 89.8 | 94.6 | 97.2 | 98.8 |
| Bootstrap3 | 0.9 | 2.4 | 4.8 | 9.8 | 24.9 | 49.1 | 74.6 | 89.8 | 94.6 | 97.2 | 98.9 |
| Bartlett1 | 0.9 | 2.3 | 4.7 | 9.9 | 24.9 | 49.0 | 74.6 | 89.9 | 94.7 | 97.3 | 99.0 |
| Bartlett2 | 0.9 | 2.4 | 4.7 | 9.9 | 24.9 | 49.0 | 74.7 | 89.9 | 94.7 | 97.3 | 98.9 |
| Bartlett3 | 0.9 | 2.4 | 4.7 | 9.8 | 25.0 | 49.2 | 74.7 | 89.8 | 94.6 | 97.3 | 98.9 |
| Bartlett alt.1 | 1.0 | 2.4 | 4.8 | 10.0 | 24.9 | 49.0 | 74.6 | 89.6 | 94.7 | 97.3 | 98.9 |
| Bartlett alt.2 | 0.9 | 2.4 | 4.7 | 10.0 | 24.9 | 49.0 | 74.7 | 89.7 | 94.6 | 97.3 | 98.9 |
| Bartlett alt.3 | 0.9 | 2.4 | 4.8 | 9.8 | 24.9 | 49.2 | 74.7 | 89.8 | 94.6 | 97.3 | 98.9 |
| directional | 1.0 | 2.4 | 4.7 | 10.0 | 25.1 | 49.5 | 74.9 | 89.8 | 94.8 | 97.5 | 98.9 |
| Skovgaard | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 1.2 | 4.6 | 9.2 | 14.7 | 23.7 |
| sim-error | 0.1 | 0.2 | 0.2 | 0.3 | 0.4 | 0.5 | 0.4 | 0.3 | 0.2 | 0.2 | 0.1 |

Tabella 3.6: Distribuzione empirica simulata (%) dei p -value per verificare l'omogeneità dei parametri esponenziali, con $n = 5$, $g = 1000$ e $R = 10000$ repliche. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

Dai grafici riportati nelle Figure 3.1-3.13 e dai risultati nelle Tabelle 3.4-3.6 è possibile osservare la tendenza del test log-rapporto di verosimiglianza a sottostimare i p -value mentre il test aggiustato di Skovgaard ha la tendenza a sovrastimare i p -value in condizioni di estremità. Il test direzionale e le varianti proposte basate sul ricampionamento bootstrap sembrano essere affidabili. È di interesse notare come la sovrastima/sottostima dei p -value da parte dei test log-rapporto di verosimiglianza e Skovgaard risulta più evidente man mano che il divario tra n e g aumenta, mentre nel caso meno estremo ($n = 10$ e $g = 100$) è meno pronunciata.

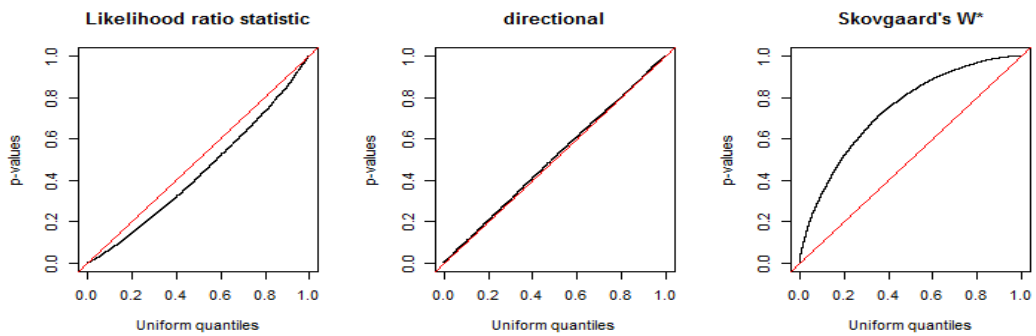


Figura 3.2: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 100$, con $R = 10000$ replicazioni

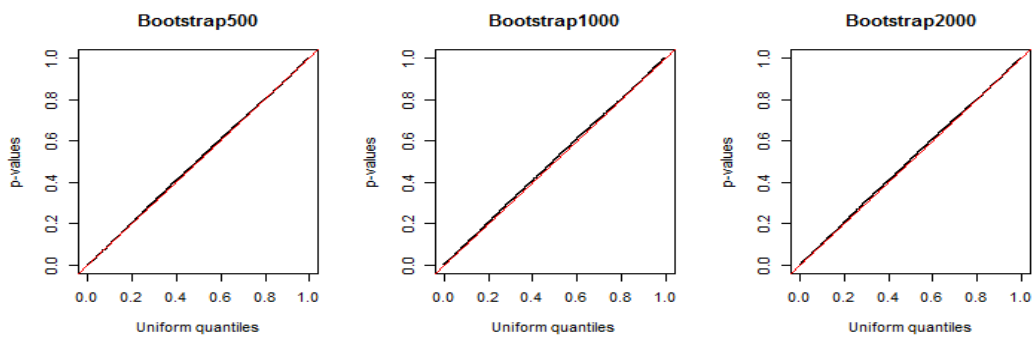


Figura 3.3: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 100$, con $R = 10000$ replicazioni

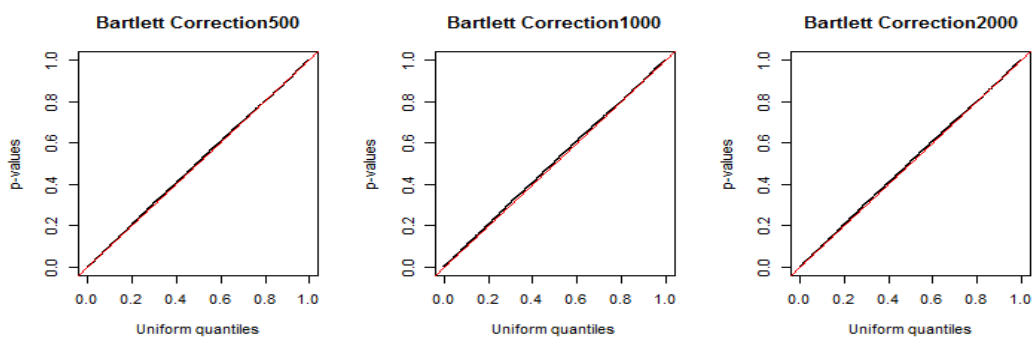


Figura 3.4: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 100$, con $R = 10000$ replicazioni

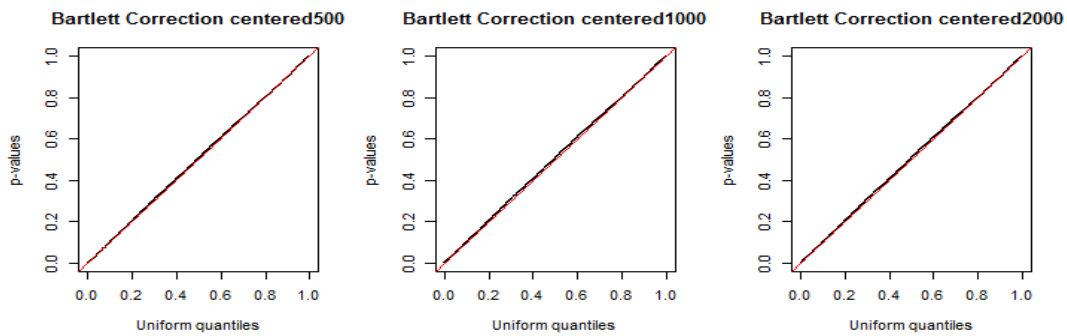


Figura 3.5: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 100$, con $R = 10000$ replicazioni

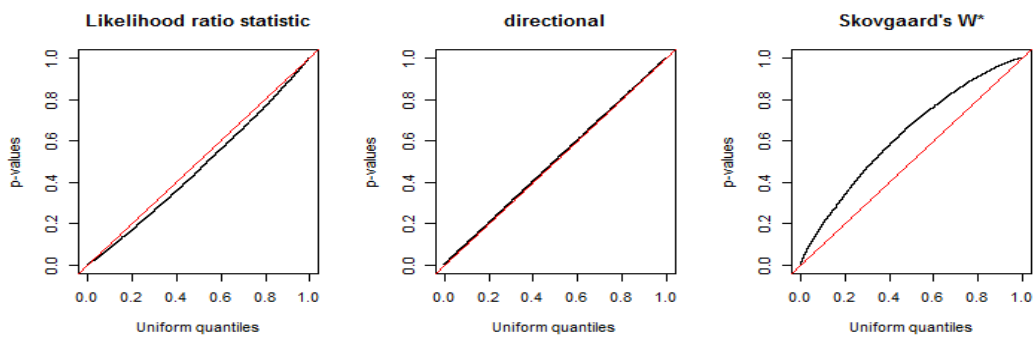


Figura 3.6: Grafici delle distribuzioni dei p -value con $n = 10$ e $g = 100$, con $R = 10000$ replicazioni

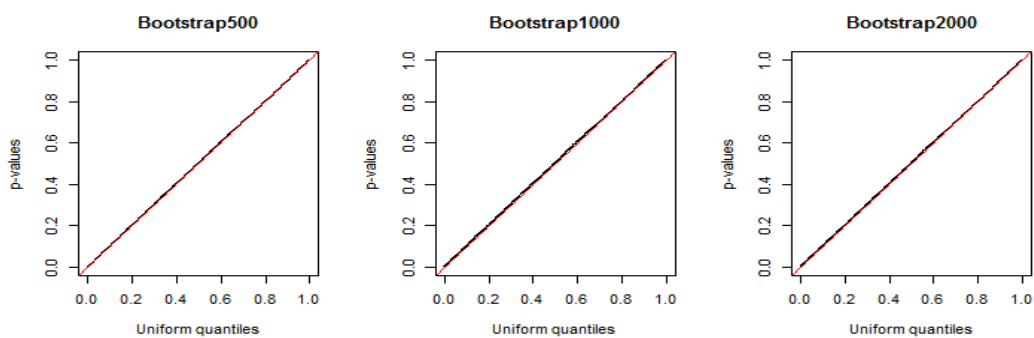


Figura 3.7: Grafici delle distribuzioni dei p -value con $n = 10$ e $g = 100$, con $R = 10000$ replicazioni

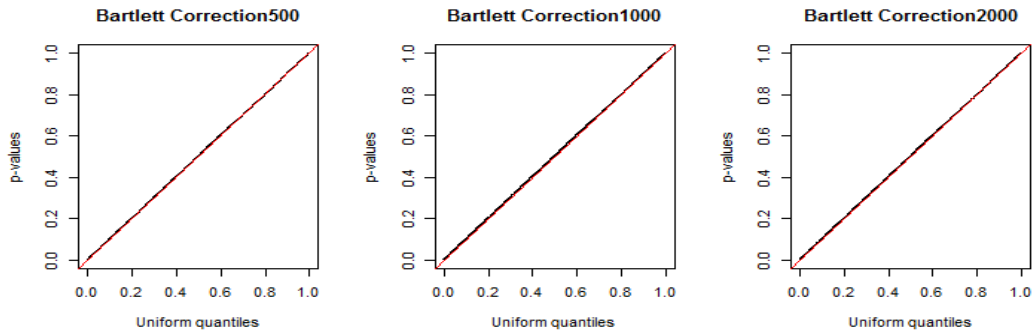


Figura 3.8: Grafici delle distribuzioni dei p -value con $n = 10$ e $g = 100$, con $R = 10000$ replicazioni

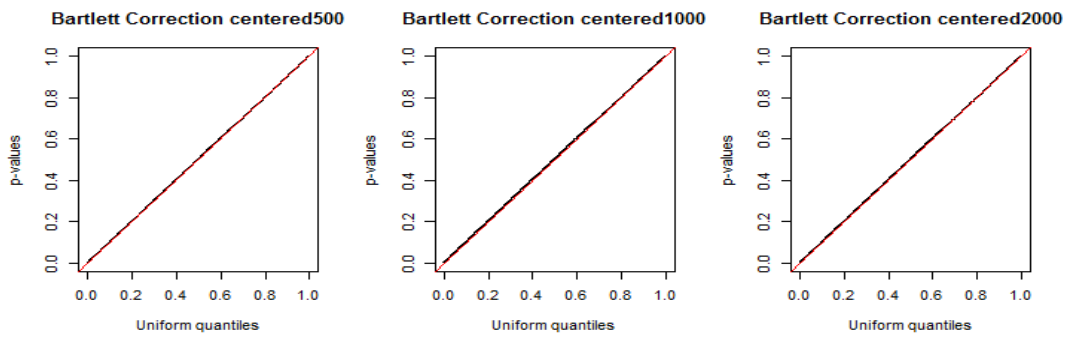


Figura 3.9: Grafici delle distribuzioni dei p -value con $n = 10$ e $g = 100$, con $R = 10000$ replicazioni

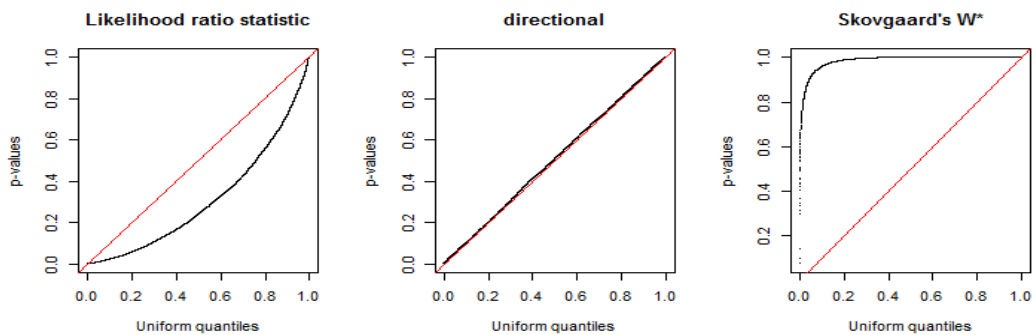


Figura 3.10: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 1000$, con $R = 10000$ replicazioni

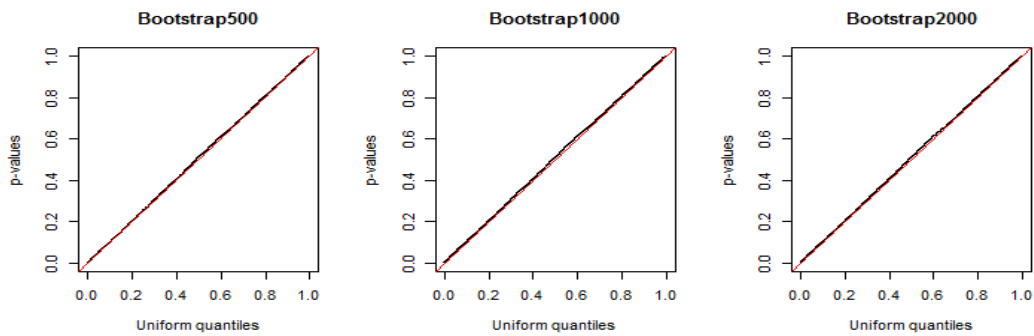


Figura 3.11: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 1000$, con $R = 10000$ repliche

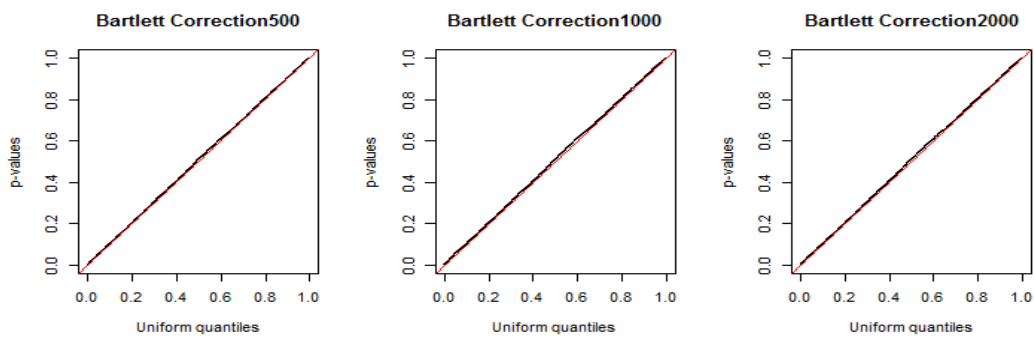


Figura 3.12: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 1000$, con $R = 10000$ repliche

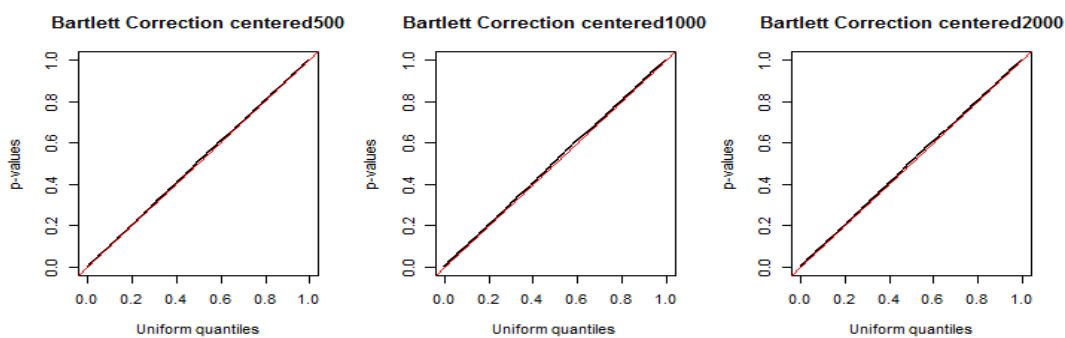


Figura 3.13: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 1000$, con $R = 10000$ repliche

3.5.4 Analisi dei dati *gravity*

Il dataset *gravity* contiene i dati relativi a 8 esperimenti effettuati per stimare l'accelerazione di gravità g effettuati a Washington tra il 1934 ed il 1935. Il dataset presenta due colonne, una contenente le deviazioni delle stime da 980000 centimetri per secondo quadrato, l'altra contenente a quale esperimento appartiene la stima corrispondente. Siamo quindi in presenza di $J = 8$ gruppi aventi numerosità diversa tra loro, dei quali vengono riportati media, varianza e numerosità campionaria nella Tabella 3.7 e la distribuzione empirica al variare dei gruppi in Figura 3.14.

| | g1 | g2 | g3 | g4 | g5 | g6 | g7 | g8 |
|------------|--------|--------|--------|-------|-------|-------|-------|-------|
| \bar{y} | 66.37 | 89.90 | 77.33 | 81.37 | 75.25 | 78.90 | 77.53 | 80.38 |
| σ^2 | 370.55 | 233.89 | 248.25 | 68.83 | 13.35 | 34.09 | 22.43 | 11.25 |
| n | 8 | 11 | 9 | 8 | 8 | 11 | 13 | 13 |

Tabella 3.7: Media, varianza e numerosità campionaria per ognuno degli 8 gruppi del dataset *gravity*

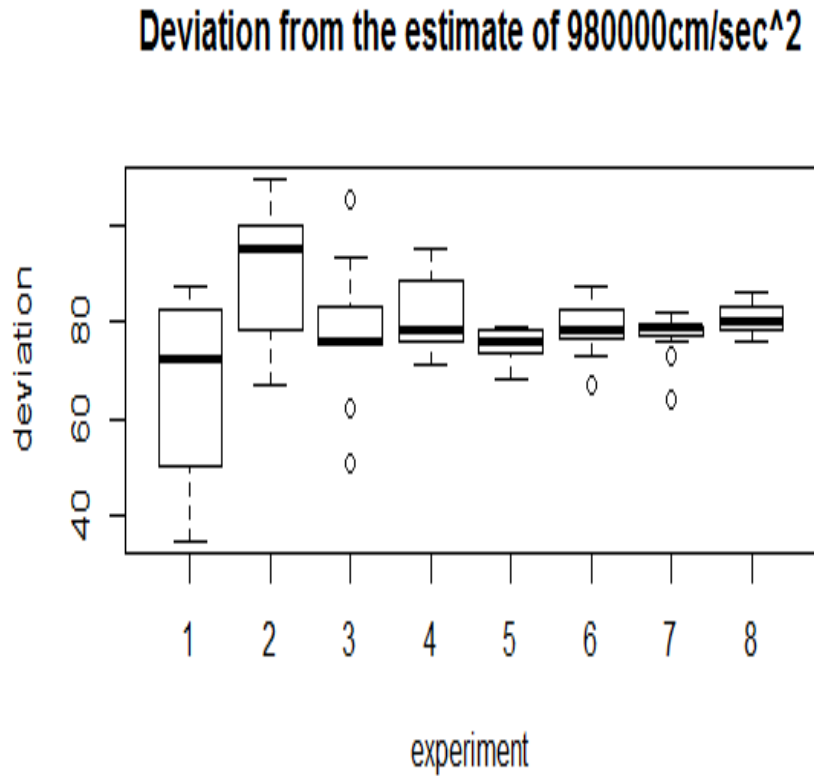


Figura 3.14: Boxplot dei dati *gravity* al variare dei gruppi

Si assume che i dati siano un campione di osservazioni indipendenti da una variabile casuale $Y_{ij} \sim N(\mu_j, \sigma_j^2)$, con $i = 1, \dots, n_j$ e $j = 1, \dots, J$. L'ipotesi che si vuole verificare in questo secondo caso è quella di omogeneità delle varianze tra i J gruppi, ossia

$$H_0 : \sigma_1^2 = \dots = \sigma_J^2,$$

contro l'ipotesi alternativa che almeno una delle uguaglianze non sia verificata. La funzione di log-verosimiglianza per il parametro $\theta = (\mu_1, \dots, \mu_J, \sigma_1^2, \dots, \sigma_J^2)$ è

$$l(\theta; y) = -\frac{1}{2} \sum_{i=1}^J \left\{ n_i \log \sigma_i^2 + \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 \right\}.$$

Le stime di massima verosimiglianza libere e vincolate sono rispettivamente

$$\hat{\theta} = (\bar{y}_1, \dots, \bar{y}_J, v_1^2, \dots, v_J^2), \hat{\theta}_0 = (\bar{y}_1, \dots, \bar{y}_J, \bar{v}^2, \dots, \bar{v}^2),$$

dove $\bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} (y_{ij}, v_i^2 = n_i^{-1} \sum_{j=1}^{n_i} / y_{ij} - \bar{y}_i)^2$, e $\bar{v}^2 = \sum_{i=1}^J n_i v_i^2 / \sum_{i=1}^J n_i$. La statistica del rapporto di log-verosimiglianza è dunque

$$w = \sum_{i=1}^J n_i \log(\bar{v}^2 / v_i^2),$$

che, sotto l'ipotesi nulla, ha una distribuzione asintotica χ_{J-1}^2 .

Analogamente a quanto effettuato sul dataset *aircondit* nell'esempio precedente, anche in questo caso sono stati applicati tutti i test discussi precedentemente, e ne vengono riportati in seguito i p -value ottenuti.

| | lrt | dir | sko | bo1 | bo2 | bo3 |
|------------|-------|----------|----------|-----|-----|-----|
| p -value | 1e-10 | 2.13e-08 | 2.26e-08 | 0 | 0 | 0 |

Tabella 3.8: p -value osservati nel dataset *gravity*. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

| | ba1 | ba2 | ba3 | bace1 | bace2 | bace3 |
|------------|---------|---------|---------|-------|---------|---------|
| p -value | 3.9e-09 | 6.9e-09 | 6.8e-09 | 2e-09 | 2.7e-09 | 2.8e-09 |

Tabella 3.9: p -value osservati nel dataset *gravity*. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

Tutti i p -value sono praticamente prossimi allo 0, fortissimo segno contro l'ipotesi nulla. L'ordine di grandezza dei vari p -value è talmente piccolo che è difficile fare paragoni, ma si può notare come tutte i test proposti come alternativa al log rapporto di verosimiglianza producano p -value più elevati.

3.5.5 Studio di simulazione per il confronto tra varianze normali

Per verificare l'affidabilità dei vari test proposti anche con questo secondo tipo di verifica di ipotesi, è stato condotto un secondo studio di simulazione.

In questo caso sono stati simulati y_{ij} campioni casuali indipendenti con distribuzione $N(\mu_i, \sigma_i^2)$, per $i = 1, \dots, g$, $j = 1, \dots, n_i$. Il parametro σ^2 è fissato (si suppone di simulare sotto l'ipotesi nulla), e anche in questo caso sono state esaminate le combinazioni di n e g utilizzate nello studio di simulazione precedente. Per ogni combinazione viene presentata la distribuzione empirica simulata.

| Nominal (%) | 1.0 | 2.5 | 5.0 | 10.0 | 25.0 | 50.0 | 75.0 | 90.0 | 95.0 | 97.5 | 99.0 |
|----------------|------|------|------|------|------|------|------|-------|-------|-------|-------|
| LRT | 45.8 | 58.6 | 69.8 | 80.1 | 91.7 | 97.8 | 99.5 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bootstrap1 | 1.2 | 2.6 | 4.9 | 10.0 | 24.9 | 49.1 | 74.6 | 89.3 | 94.2 | 97.2 | 98.6 |
| Bootstrap2 | 1.1 | 2.5 | 4.8 | 9.9 | 25.0 | 49.2 | 74.6 | 89.3 | 94.4 | 97.2 | 98.8 |
| Bootstrap3 | 1.2 | 2.6 | 4.8 | 9.8 | 25.1 | 49.2 | 74.6 | 89.3 | 94.4 | 97.3 | 98.9 |
| Bartlett1 | 1.1 | 2.6 | 4.7 | 9.9 | 25.0 | 49.2 | 74.8 | 89.4 | 94.6 | 97.4 | 99.0 |
| Bartlett2 | 1.1 | 2.6 | 4.8 | 9.9 | 25.1 | 49.2 | 74.9 | 89.4 | 94.6 | 97.5 | 98.9 |
| Bartlett3 | 1.1 | 2.6 | 4.8 | 9.9 | 25.0 | 49.1 | 74.8 | 89.4 | 94.5 | 97.5 | 98.9 |
| Bartlett alt.1 | 1.2 | 2.7 | 4.9 | 10.1 | 25.0 | 49.2 | 74.6 | 89.3 | 94.4 | 97.3 | 98.9 |
| Bartlett alt.2 | 1.2 | 2.6 | 4.9 | 10.0 | 25.2 | 49.2 | 74.7 | 89.3 | 94.5 | 97.3 | 99.0 |
| Bartlett alt.3 | 1.2 | 2.7 | 4.8 | 10.0 | 25.1 | 49.1 | 74.6 | 89.3 | 94.4 | 97.4 | 98.9 |
| directional | 1.2 | 2.8 | 5.4 | 9.9 | 24.8 | 49.7 | 74.7 | 89.8 | 94.8 | 97.3 | 98.9 |
| Skovgaard | 0.0 | 0.0 | 0.0 | 0.2 | 0.7 | 3.2 | 10.5 | 24.5 | 35.9 | 46.8 | 59.7 |
| sim-error | 0.1 | 0.2 | 0.2 | 0.3 | 0.4 | 0.5 | 0.4 | 0.3 | 0.2 | 0.2 | 0.1 |

Tabella 3.10: Distribuzione empirica simulata (%) dei p -value per verificare l'omogeneità delle varianze normali, con $n = 5$, $g = 100$ e $R = 10000$ replicazioni. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

| Nominal (%) | 1.0 | 2.5 | 5.0 | 10.0 | 25.0 | 50.0 | 75.0 | 90.0 | 95.0 | 97.5 | 99.0 |
|----------------|------|------|------|------|------|------|------|------|------|------|-------|
| LRT | 10.4 | 18.4 | 27.1 | 40.0 | 63.4 | 84.1 | 95.3 | 98.8 | 99.5 | 99.8 | 100.0 |
| Bootstrap1 | 0.9 | 2.7 | 5.0 | 9.8 | 24.6 | 50.2 | 75.2 | 90.2 | 95.4 | 97.8 | 99.0 |
| Bootstrap2 | 0.9 | 2.5 | 5.1 | 9.8 | 24.5 | 50.3 | 75.4 | 90.2 | 95.4 | 97.7 | 99.0 |
| Bootstrap3 | 0.9 | 2.6 | 5.1 | 9.8 | 24.5 | 50.1 | 75.4 | 90.3 | 95.4 | 97.9 | 99.1 |
| Bartlett1 | 0.8 | 2.6 | 5.1 | 9.8 | 24.6 | 50.3 | 75.5 | 90.3 | 95.4 | 97.9 | 99.1 |
| Bartlett2 | 0.8 | 2.6 | 5.1 | 9.8 | 24.7 | 50.3 | 75.6 | 90.3 | 95.5 | 97.9 | 99.1 |
| Bartlett3 | 0.9 | 2.6 | 5.0 | 9.7 | 24.7 | 50.1 | 75.5 | 90.3 | 95.5 | 97.9 | 99.1 |
| Bartlett alt.1 | 0.9 | 2.5 | 5.1 | 9.9 | 24.6 | 50.3 | 75.4 | 90.3 | 95.4 | 97.8 | 99.1 |
| Bartlett alt.2 | 0.9 | 2.6 | 5.1 | 9.8 | 24.6 | 50.3 | 75.5 | 90.4 | 95.4 | 97.8 | 99.1 |
| Bartlett alt.3 | 0.9 | 2.6 | 5.1 | 9.8 | 24.7 | 50.2 | 75.5 | 90.3 | 95.4 | 97.9 | 99.1 |
| directional | 1.1 | 2.6 | 5.1 | 9.7 | 25.3 | 49.8 | 75.7 | 90.8 | 95.3 | 97.7 | 99.1 |
| Skovgaard | 0.1 | 0.3 | 0.8 | 1.9 | 6.6 | 19.0 | 40.0 | 62.7 | 74.9 | 83.3 | 90.8 |
| sim-error | 0.1 | 0.2 | 0.2 | 0.3 | 0.4 | 0.5 | 0.4 | 0.3 | 0.2 | 0.2 | 0.1 |

Tabella 3.11: Distribuzione empirica simulata (%) dei p -value per verificare l'omogeneità delle varianze normali, con $n = 10$, $g = 100$ e $R = 10000$ replicazioni. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

| Nominal (%) | 1.0 | 2.5 | 5.0 | 10.0 | 25.0 | 50.0 | 75.0 | 90.0 | 95.0 | 97.5 | 99.0 |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| LRT | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bootstrap1 | 1.3 | 3.0 | 5.1 | 10.7 | 25.0 | 49.5 | 75.0 | 89.6 | 94.8 | 97.5 | 98.9 |
| Bootstrap2 | 1.2 | 2.8 | 5.2 | 10.6 | 25.1 | 49.6 | 74.8 | 89.9 | 94.8 | 97.5 | 99.1 |
| Bootstrap3 | 1.3 | 2.8 | 5.2 | 10.7 | 25.0 | 49.3 | 75.0 | 89.7 | 95.0 | 97.5 | 99.1 |
| Bartlett1 | 1.2 | 2.8 | 4.9 | 10.4 | 25.0 | 49.6 | 75.1 | 90.0 | 95.2 | 97.5 | 99.2 |
| Bartlett2 | 1.2 | 2.8 | 5.0 | 10.5 | 25.1 | 49.7 | 75.0 | 90.0 | 95.2 | 97.5 | 99.1 |
| Bartlett3 | 1.2 | 2.8 | 5.1 | 10.6 | 25.1 | 49.4 | 75.0 | 90.0 | 95.2 | 97.6 | 99.2 |
| Bartlett alt.1 | 1.2 | 2.8 | 5.1 | 10.7 | 25.2 | 49.5 | 75.0 | 89.8 | 95.1 | 97.5 | 99.2 |
| Bartlett alt.2 | 1.2 | 2.9 | 5.2 | 10.7 | 25.1 | 49.7 | 74.9 | 89.9 | 95.0 | 97.6 | 99.2 |
| Bartlett alt.3 | 1.3 | 2.8 | 5.2 | 10.7 | 25.1 | 49.4 | 75.0 | 89.7 | 95.1 | 97.5 | 99.2 |
| directional | 0.9 | 2.4 | 5.2 | 10.2 | 25.5 | 50.2 | 75.5 | 90.5 | 95.1 | 97.4 | 98.8 |
| Skovgaard | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| sim-error | 0.1 | 0.2 | 0.3 | 0.4 | 0.6 | 0.7 | 0.6 | 0.4 | 0.3 | 0.2 | 0.1 |

Tabella 3.12: Distribuzione empirica simulata (%) dei p -value per verificare l'omogeneità delle varianze normali, con $n = 5$, $g = 1000$ e $R = 10000$ replicazioni. Per i test bootstrap, Bartlett e Bartlett centrato sono riportati i valori dei p -value per tre valori di B (500,1000,2000)

Dai grafici riportati nelle Figure 3.14-3.26 e dai risultati nelle tabelle 3.10-3.12 è possibile osservare nuovamente la tendenza del test log-rapporto di verosimiglianza e del test di Skovgaard a rispettivamente sottostimare e sovrastimare i p -value, questa volta in maniera molto più marcata. I rimanenti test restano, come previsto, più affidabili anche in questo secondo studio. Di nuovo, è apprezzabile la differenza per quanto riguarda la sovrastima e la sottostima tra la situazione meno estrema ($n = 10$, $g = 100$) e quella più estrema ($n = 5$, $g = 1000$).

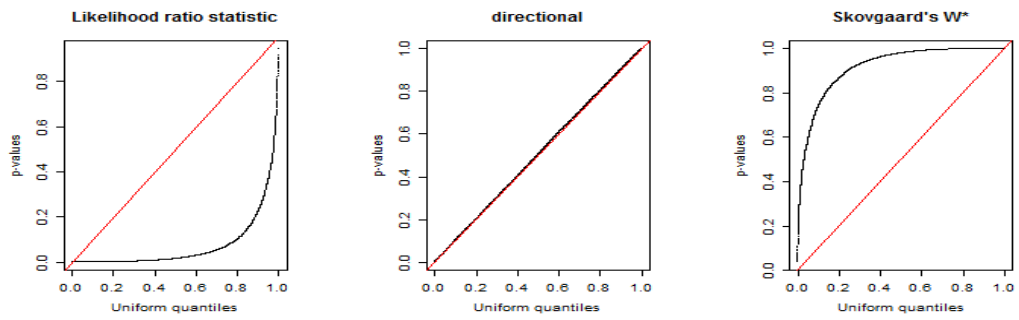


Figura 3.15: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 100$, con $R = 10000$ replicazioni

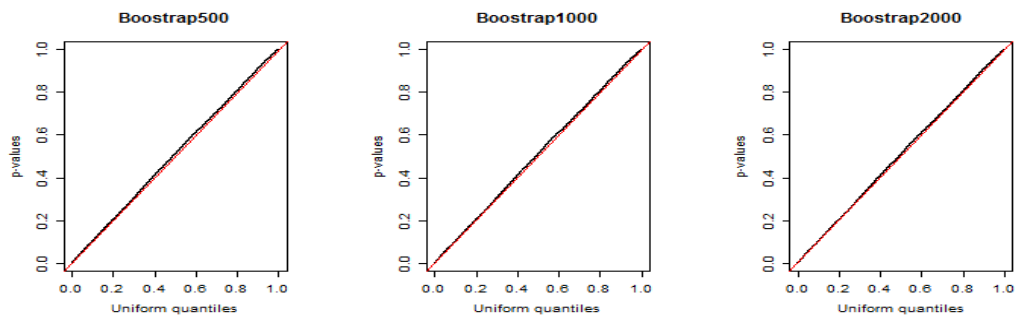


Figura 3.16: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 100$, con $R = 10000$ replicazioni

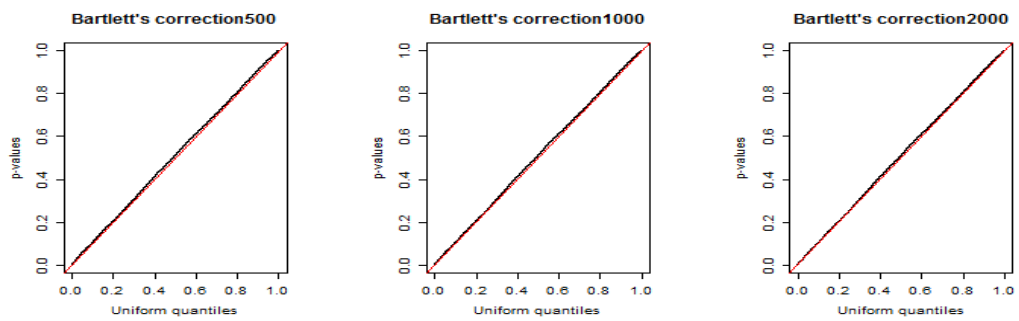


Figura 3.17: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 100$, con $R = 10000$ replicazioni

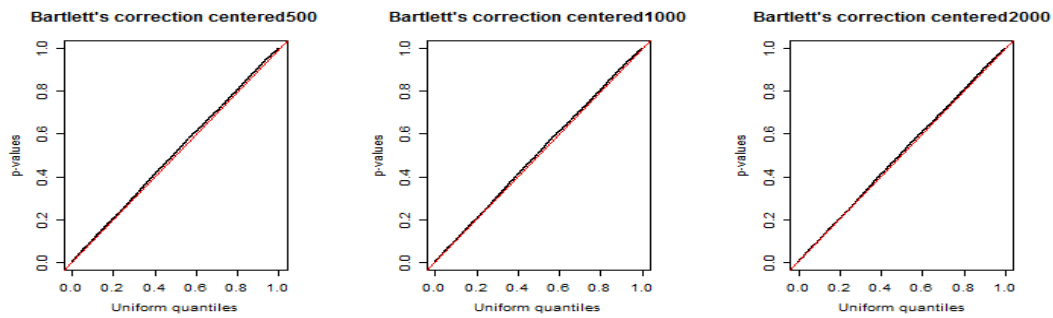


Figura 3.18: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 100$, con $R = 10000$ replicazioni

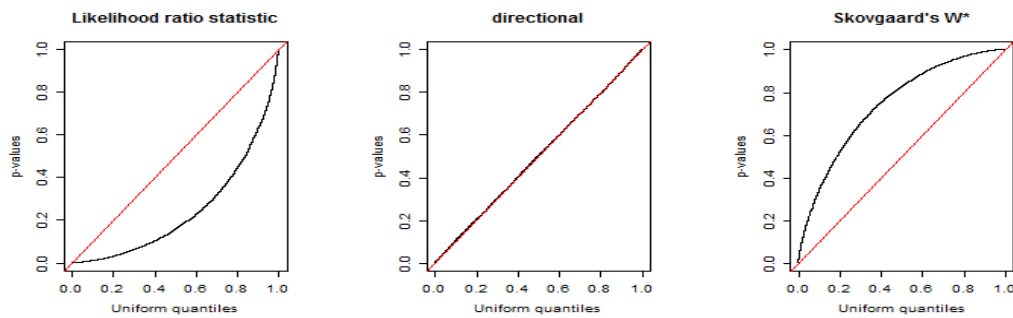


Figura 3.19: Grafici delle distribuzioni dei p -value con $n = 10$ e $g = 100$, con $R = 10000$ replicazioni

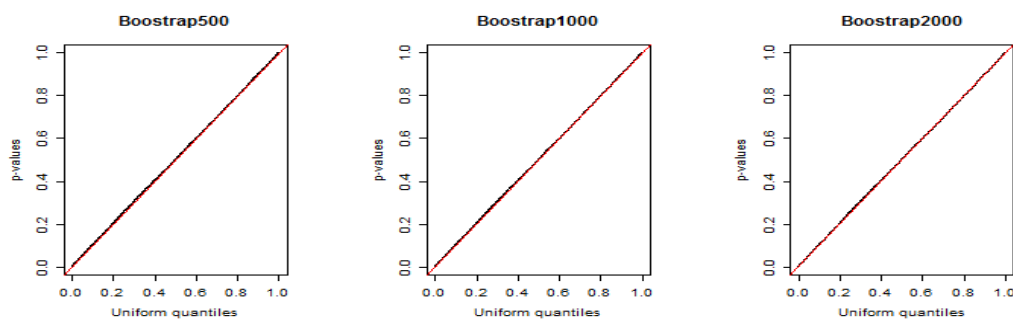


Figura 3.20: Grafici delle distribuzioni dei p -value con $n = 10$ e $g = 100$, con $R = 10000$ replicazioni

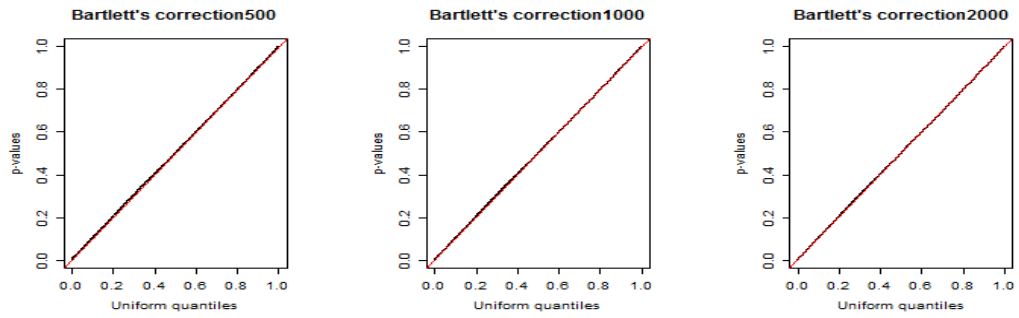


Figura 3.21: Grafici delle distribuzioni dei p -value con $n = 10$ e $g = 100$, con $R = 10000$ replicazioni

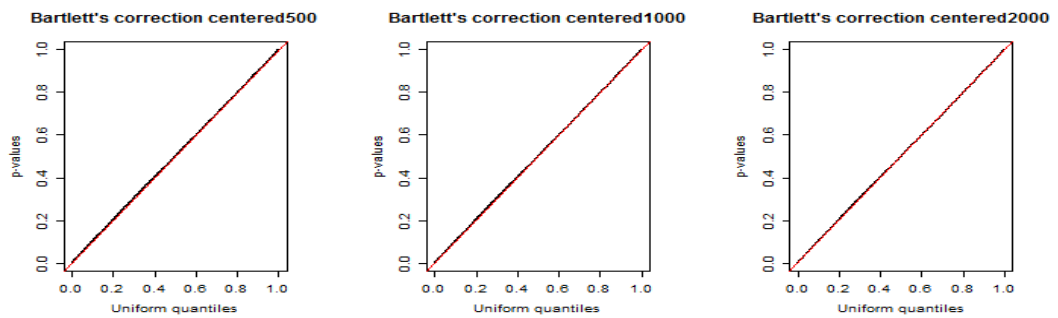


Figura 3.22: Grafici delle distribuzioni dei p -value con $n = 10$ e $g = 100$, con $R = 10000$ replicazioni

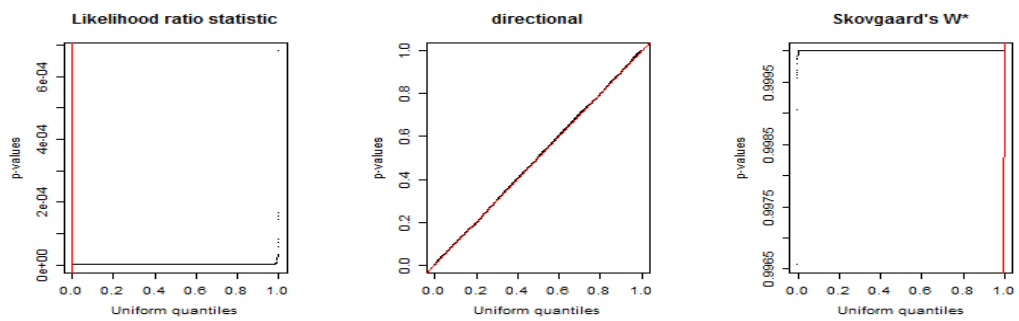


Figura 3.23: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 1000$, con $R = 10000$ replicazioni

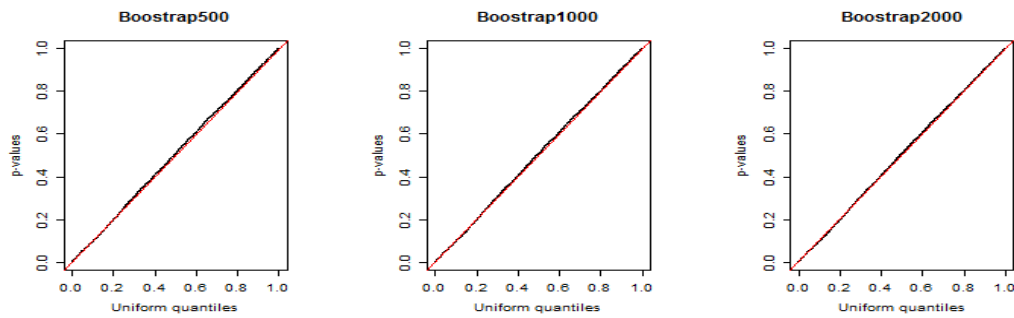


Figura 3.24: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 1000$, con $R = 10000$ replicazioni

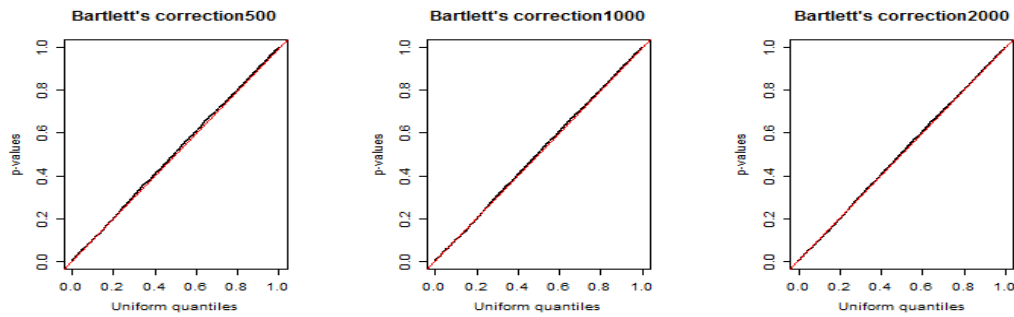


Figura 3.25: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 1000$, con $R = 10000$ replicazioni

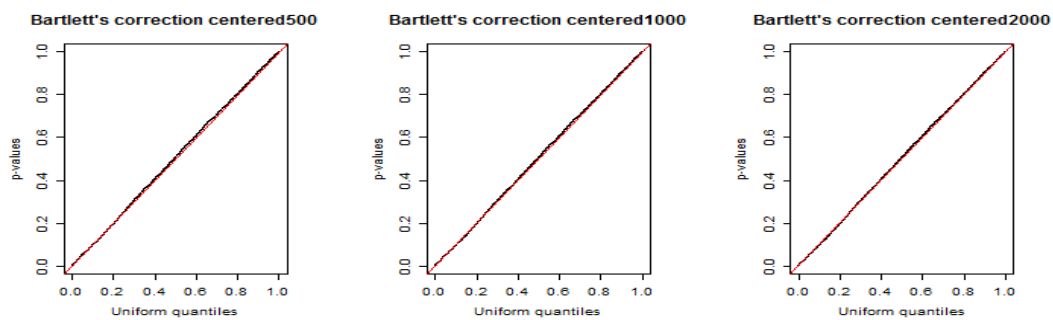


Figura 3.26: Grafici delle distribuzioni dei p -value con $n = 5$ e $g = 1000$, con $R = 10000$ replicazioni

Conclusioni

In questo studio sono stati messi a confronto diversi test di verifica di ipotesi basati sulla verosimiglianza, al fine di contrastare il problema di una bassa accuratezza in una situazione di elevata dimensionalità parametrica contro una scarsa numerosità campionaria.

È stato osservato come l'usuale approssimazione asintotica per la distribuzione del test *log*-rapporto di verosimiglianza perda di affidabilità in questo tipo di situazioni, e sono state discusse alcune alternative attraverso due analisi di dati e due studi di simulazione in diversi scenari, piuttosto estremi.

Sono stati presentati nel Capitolo 2 il test direzionale e il test aggiustato di Skovgaard, alternative al test *log*-rapporto di verosimiglianza presenti in letteratura. Dagli studi di simulazione, il test direzionale si è dimostrato affidabile per quanto riguarda il problema in esame, mentre il test aggiustato di Skovgaard ha avuto la tendenza a sovrastimare i vari *p*-value all'aumentare dell'estremità della situazione.

Sono state infine proposti nel Capitolo 3 altre tre approssimazioni alternative per la distribuzione del test *log*-rapporto di verosimiglianza, basati sull'uso della tecnica di simulazione bootstrap e su delle correzioni del test *log*-rapporto di verosimiglianza stesso. Questi test sono stati chiamati test bootstrap, test *w* con correzione di Bartlett e test *w* con correzione di Bartlett centrato (o alternativo). È stato inoltre osservato l'andamento dei *p*-value al variare del numero di campioni bootstrap utilizzati per le varie simulazioni. Gli studi di simulazione hanno dimostrato che queste tre proposte sono affidabili per quanto riguarda gli scenari, piuttosto estremi, presi in esame. Non è stata rilevata una particolare influenza del numero di campioni bootstrap utilizzati. Questi studi preliminari andranno approfonditi in modelli

più complessi, anche non appartenenti a famiglie esponenziali.

Bibliografia

- A., Azzalini. *Inferenza Statistica una presentazione basata sul concetto di verosimiglianza*. Milano: Springer Verlag, 2008.
- Davison A. C. Fraser D. A. S., Reid N. Sartori N. «Accurate Directional Inference for Vector Parameters in Linear Exponential Families». In: *Journal of the American Statistical Association*. B (2014), pp. 302–314.
- Fraser, D.A.S. e H. Massam. «Conical Tests: Observed Levels of Significance and Confidence Regions». In: *Statistics Hefte* (1985), pp. 26, 1–17.
- Pace L., Salvani A. *Introduzione alla statistica II*. Milano: Cedam, 2001.
- Skovgaard, L. M. «Saddlepoint Expansions for Directional Test Probabilities». In: *Journal of the Royal Statistical Society*. B (1988), pp. 50, 3–32.

Appendice A

Codice R utilizzato

Codice A.1: Funzioni utilizzate per le analisi e le simulazioni

```
# exponential samples
#####

W.fun <- function(data)
{
  n <- tapply(data$y,data$group,length)
  # MLE
  ybar <- tapply(data$y,data$group,mean)
  thetahat <- 1/ybar
  # common MLE under H0
  thetatilde <- sum(n)/sum(n*ybar)
  # LRT for homogeneity of variances
  w <- 2*sum(n*log(thetahat/thetatilde))
  list(w=w, th=thetahat, tt=thetatilde)
}

w.funbo <- function(data) # for boot function only
{
  n <- tapply(data$y,data$group,length)
  # MLE
  ybar <- tapply(data$y,data$group,mean)
```

```

thetahat <- 1/ybar
# common MLE under H0
thetatilde <- sum(n)/sum(n*ybar)
# LRT for homogeneity of variances
w <- 2*sum(n*log(thetahat/thetatilde))
w
}

sim.fun <- function(data,par)
{
  data$y <- rexp(length(data$y),par)
  data
}

directional.p <- function(y,group,plot=FALSE)
{
  I <- nlevels(group)
  # group sample sizes
  n <- tapply(y,group,FUN=length)
  # sample means
  ybar <- tapply(y,group,mean)
  # common mean
  ytilde <- sum(n*ybar)/sum(n) #mean(y)
  tsup <- ytilde/(ytilde-min(ybar))
  # ff <- function(x) x^(I-2)*prod((1-x*(ytilde-ybar)/ytilde)^(n-1))/
  # ytilde^I
  # ff <- function(x) x^(I-2)*prod((1-x*(ytilde-ybar)/ytilde)^(n-1))/
  # ytilde^I
  # logarithmic scale to improve stability
  ff <- function(x) exp((I-2)*log(x)+sum((n-1)*log(1-x*(ytilde-ybar)/
  ytilde)))-I*log(ytilde))
  ff.v <- Vectorize(ff)
  if (plot) plot(ff.v,0,tsup)
  up <- integrate(ff.v,lower=1,upper=tsup)
  down <- integrate(ff.v,lower=0,upper=1)
  up$value/(up$value+down$value)
}

```



```

}

directional.p2 <- function(y,group,plot=FALSE)
{ # directional p-value with nuisance correction
  I <- nlevels(group)
  # group sample sizes
  n <- tapply(y,group,FUN=length)
  # sample means
  ybar <- tapply(y,group,mean)
  # common mean
  ytilde <- sum(n*ybar)/sum(n) #mean(y)
  tsup <- ytilde/(ytilde-min(ybar))
  #   ff <- function(x) x^(I-2)*prod((1-x*(ytilde-ybar)/ytilde)^(n-1))/
    ytilde^I * sqrt(sum(1/(ytilde-x*(ytilde-ybar))^2))
  ff <- function(x) x^(I-2)*prod((1-x*(ytilde-ybar)/ytilde)^(n-1))/ytilde^
    I /sqrt(sum(1/(ytilde-x*(ytilde-ybar))^2))*sqrt(sum(n*ybar/(ytilde-
    x*(ytilde-ybar))))
  ff.v <- Vectorize(ff)
  if (plot) plot(ff.v,0,tsup)
  up <- integrate(ff.v,lower=1,upper=tsup)
  down <- integrate(ff.v,lower=0,upper=1)
  up$value/(up$value+down$value)
}

W.skovgaard <- function(y,group)
{ # saddlepoint function (without nuisance adjustment)

  I <- nlevels(group)
  d <- I-1
  n <- tapply(y,group,length)
  ybar <- tapply(y,group,mean)
  ytilde <- sum(n*ybar)/sum(n)
  #LRT
  w <- 2*sum(n*log(ytilde/ybar))
  # #gamma
  # num <- (sum(n*((ybar-ytilde)/ytilde)^2))^(d/2)

```

```

# den <- sum(n*(ybar-ytilde)^2/ybar/ytilde)
# fact <- (prod(ytilde/ybar))
# gamma <- num/den/w^((d/2)-1)*fact
# wstar1 <- w*(1-log(gamma)/w)^2
# wstar2 <- w-2*log(gamma)
#log(gamma)
lnum <- (d/2)*log(sum(n*((ybar-ytilde)/ytilde)^2))
lden <- log(sum(n*(ybar-ytilde)^2/ybar/ytilde))
lfact <- sum(log(ytilde/ybar))
lgamma <- lnum-lden-((d/2)-1)*log(w)+lfact
wstar1 <- w*(1-lgamma/w)^2
wstar2 <- w-2*lgamma
p1 <- pchisq(wstar1,d,lower=FALSE)
p2 <- pchisq(wstar2,d,lower=FALSE)
list(wstar1=wstar1,p.value1=p1,wstar2=wstar2,p.value2=p2)
}

bt.simula <- function(theta,n,Nsim=10000,trace=TRUE)
{
  group <- factor(rep(1:length(n),n))
  I <- nlevels(group)
  fo.p <- bo500.p <- bo1000.p <- bo.p <- ba500.p <- ba1000.p <- ba.p <-
    bace500.p <- bace1000.p <- bace.p <- d1.p <- d2.p <- sk.p1 <- sk.p2
    <- numeric(Nsim)
  for (i in 1:Nsim)
  {
    if (trace) if (i%1000==0) cat(i,'\n')
    y <- rexp(sum(n),rate=theta)
    dati.i <- data.frame(y=y,group=group)
    # first order
    #####
    w <- W.fun(dati.i)
    fo.p[i] <- pchisq(w$I,I-1,lower=FALSE)
    # bootstrap
    #####
  }
}

```

```

out <- boot(dati.i,w.funbo,sim="parametric",ran.gen=sim.fun,mle=w$tt,R
  =2000,parallel="multicore",ncpus=3)
bo500.p[i] <- mean(out$t[1:500]>out$t0)
bo1000.p[i] <- mean(out$t[1:1000]>out$t0)
bo.p[i] <- mean(out$t>out$t0)
# bartlett correction
#####
w.b500 <- w$(I-1)/mean(out$t[1:500])
ba500.p[i] <- pchisq(w.b500,I-1,lower=F)
w.b1000 <- w$(I-1)/mean(out$t[1:1000])
ba1000.p[i] <- pchisq(w.b1000,I-1,lower=F)
w.b <- w$(I-1)/mean(out$t)
ba.p[i] <- pchisq(w.b,I-1,lower=F)
# bartlett correction centered
#####
a500 <- mean(out$t[1:500])
b500 <- (2*a500^2)/var(out$t[1:500])
w.bc500 <- w$b500/a500
bace500.p[i] <- pchisq(w.bc500,b500,lower=F)
a1000 <- mean(out$t[1:1000])
b1000 <- (2*a1000^2)/var(out$t[1:1000])
w.bc1000 <- w$b1000/a1000
bace1000.p[i] <- pchisq(w.bc1000,b1000,lower=F)
a <- mean(out$t)
b <- (2*a^2)/var(out$t)
w.bc <- w$b/a
bace.p[i] <- pchisq(w.bc,b,lower=F)
# directional p-values
#####
d1.p[i] <- directional.p(y,group)
d2.p[i] <- directional.p2(y,group)
# Skovgaard p-values
#####
app <- W.skovgaard(y,group)
sk.p1[i] <- app$p.value1
sk.p2[i] <- app$p.value2

```

```

}
list(first.order=fo.p,bootstrap500=bo500.p,bootstrap1000=bo1000.p,
      bootstrap=bo.p,bartlett500=ba500.p,bartlett1000=ba1000.p,bartlett=ba
      .p,bartlett.centered500=bace500.p,bartlett.centered1000=bace1000.p,
      bartlett.centered=bace.p,directional1=d1.p,directional2=d2.p,
      skovgaard1=sk.p1,skovgaard2=sk.p2,theta=theta,n=n)
}

summary.sim <- function(out,levels=c
  (0.01,0.025,0.05,0.10,0.25,0.5,0.75,0.9,0.95,0.975,0.99))
{
  fo <- sapply(levels,function(x) mean(out$first.order <x))
  bo500 <- sapply(levels,function(x) mean(out$bootstrap500 <x))
  bo1000 <- sapply(levels,function(x) mean(out$bootstrap1000 <x))
  bo <- sapply(levels,function(x) mean(out$bootstrap <x))
  ba500 <- sapply(levels,function(x) mean(out$bartlett500 <x))
  ba1000 <- sapply(levels,function(x) mean(out$bartlett1000 <x))
  ba <- sapply(levels,function(x) mean(out$bartlett <x))
  bace500 <- sapply(levels,function(x) mean(out$bartlett.centered500 <x))
  bace1000 <- sapply(levels,function(x) mean(out$bartlett.centered1000 <x)
    )
  bace <- sapply(levels,function(x) mean(out$bartlett.centered <x))
  dir1 <- sapply(levels,function(x) mean(out$directional1 <x))
  sk1 <- sapply(levels,function(x) mean(out$skovgaard1 <x))
  sim.se <- sqrt(levels*(1-levels)/length(out$first.order))
  res <- rbind(levels, fo,bo500,bo1000,bo,ba500,ba1000,ba,bace500,bace1000,
    bace,dir1,sk1,sim.se)
  rownames(res) <- c("nominal","lrt","bootstrap500","bootstrap1000","
    bootsrap","Bartlett500","Bartlett1000","Bartlett","Bartlett
    centered500","Bartlett centered1000","Bartlett centered","
    directional","Skovgaard","sim-error")
  res
}

plot.sim <- function(out,step=1)
{

```

```
par(mfrow=c(4,3))
index=seq(1,length(out$first.order),by=step)
x=sort(out$first.order)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Likelihood ratio statistic")
x=sort(out$bootstrap500)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bootstrap500")
x=sort(out$bootstrap1000)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bootstrap1000")
x=sort(out$bootstrap)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bootstrap")
x=sort(out$bartlett500)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett Correction500")
x=sort(out$bartlett1000)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett Correction1000")
x=sort(out$bartlett)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
```

```

title("Bartlett Correction")
x=sort(out$bartlett.centered500)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett Correction centered500")
x=sort(out$bartlett.centered1000)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett Correction centered1000")
x=sort(out$bartlett.centered)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett Correction centered")
x=sort(out$directional1)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("directional")
x=sort(out$skovgaard1)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab="p-values",pch=".")
abline(0,1,col=2)
# title(expression(paste("Skovgaard's ",W^"*")))
title("Skovgaard's W*")
###
# Skovgaard's W** is always worse than W*
# hence we skip it
}

# normal samples
#####

W.funCVN <- function(data)

```

```

{
  n <- tapply(data$y, data$group, length)
  muhat <- tapply(data$y, data$group, mean)
  mubar <- sum(n*muhat)/sum(n)
  sigma2hat <- tapply(data$y, data$group, FUN=var)*(n-1)/n
  # common MLE under H0
  sigma2hatbar <- sum(n*sigma2hat)/sum(n)
  # LRT for homogeneity of variances
  w <- sum(n*log(sigma2hatbar/sigma2hat))
  list(w=w, muhat=muhat, mubar=mubar, s2h=sigma2hat, s2hb=sigma2hatbar)
}

W.funboCVN <- function(data) # use this function only in the boot command
  (boot doesn't allow lists)
{
  n <- tapply(data$y, data$group, length)
  sigma2hat <- tapply(data$y, data$group, FUN=var)*(n-1)/n
  # common MLE under H0
  sigma2hatbar <- sum(n*sigma2hat)/sum(n)
  # LRT for homogeneity of variances
  w <- sum(n*log(sigma2hatbar/sigma2hat))
}

sim.funCVN <- function(data, par)
{
  data$y <- rnorm(length(data$y), par)
  data
}

directional.p3 <- function(y, group, plot=FALSE)
{ # this is the "analytical" version of directional.p2 (they give exactly
  the same result)
  I <- nlevels(group)
  # group sample sizes
  n <- tapply(y, group, FUN=length)
  # MLE of variances

```

```

sigma2hat <- tapply(y,group,var)*(n-1)/n
# common MLE under H0
sigma2hatbar <- sum(n*sigma2hat)/sum(n)
tsup <- sigma2hatbar/(sigma2hatbar-min(sigma2hat))
# ff <- function(x) x^(I-2)*prod((x*sigma2hat+(1-x)*sigma2hatbar)^((n-3)
  /2))
# use logarithm for numerical stability
ff <- function(x) exp((I-2)*log(x)+sum(((n-3)/2)*log(x*sigma2hat+(1-x)*
  sigma2hatbar))-sum(((n-3)/2)*log(sigma2hatbar))
#(NOTE: the last term is a constant used for rescaling and numerical
  stability)
ff.v <- Vectorize(ff)
if (plot) plot(ff.v,0,tsup)
up <- integrate(ff.v,lower=1,upper=tsup)
down <- integrate(ff.v,lower=0,upper=1)
# this version is more stable in certain instances and avoid results >1
up$value/(down$value+up$value)
}

W.skovgaard2 <- function(y,group)
{ # saddlepoint function (without nuisance adjustment)
  I <- nlevels(group)
  d <- I-1
  n <- tapply(y,group,length)
  v2 <- tapply(y,group,var)*(n-1)/n
  v2bar <- sum(n*v2)/sum(n)
  #LRT
  w <- sum(n*log(v2bar/v2))
  # #gamma
  # num <- (0.5*sum(n*((v2-v2bar)/v2bar)^2))^(d/2)
  # den <- 0.5*sum(n*(v2-v2bar)^2/v2/v2bar)
  # fact <- sqrt(prod((v2bar/v2)^3))
  # gamma <- num/den/w^((d/2)-1)*fact
  # wstar1 <- w*(1-log(gamma)/w)^2
  # wstar2 <- w-2*log(gamma)
  #log(gamma): numerically more stable when d is large

```



```

lnum <- (d/2)*log(0.5*sum(n*((v2-v2bar)/v2bar)^2))
lden <- log(0.5*sum(n*(v2-v2bar)^2/v2/v2bar))
lfact <- (3/2)*sum(log(v2bar/v2))
lgamma <- lnum-lden-((d/2)-1)*log(w)+lfact
wstar1 <- w*(1-lgamma/w)^2
wstar2 <- w-2*lgamma
p1 <- pchisq(wstar1,d,lower=FALSE)
p2 <- pchisq(wstar2,d,lower=FALSE)
list(wstar1=wstar1,p.value1=p1,wstar2=wstar2,p.value2=p2)
}

bt.simula2 <- function(mu,sigma2,n,Nsim=10000,trace=TRUE)
{
  group <- factor(rep(1:length(n),n))
  I <- nlevels(group)
  fo.p <- bo500.p <- bo1000.p <- bo.p <- bt500.p <- bt1000.p <- bt.p <-
    btc500.p <- btc1000.p <- btc.p <- d2.p <- sk.p1 <- sk.p2 <- numeric(
    Nsim)
  for (i in 1:Nsim)
  {
    if (trace) if (i%%1==0) cat(i,'\n')
    y <- rnorm(n[group],mean=mu[group],sd=sqrt(sigma2))
    dati.i <- data.frame(y=y,group=group)
    # first order
    #####
    w <- W.funCVN(dati.i)
    fo.p[i] <- pchisq(w$w,I-1,lower=FALSE)
    # bootstrap
    #####
    out <- boot(dati.i, W.funboCVN, sim="parametric", ran.gen=sim.funCVN,
      mle=w$s2hb,R=2000,parallel="multicore",ncpus=3)
    bo500.p[i] <- mean(out$t[1:500]>out$t0)
    bo1000.p[i] <- mean(out$t[1:1000]>out$t0)
    bo.p[i] <- mean(out$t>out$t0)
    # Bartlett correction
    #####

```

```

w.b500 <- w$w*(I-1)/mean(out$t[1:500])
bt500.p[i] <- pchisq(w.b500,I-1,lower=F)
w.b1000 <- w$w*(I-1)/mean(out$t[1:1000])
bt1000.p[i] <- pchisq(w.b1000,I-1,lower=F)
w.b <- w$w*(I-1)/mean(out$t)
bt.p[i] <- pchisq(w.b,I-1,lower=F)
# Bartlett correction centered
#####
a500 <- mean(out$t[1:500])
b500 <- (2*a500^2)/var(out$t[1:500])
w.bc500 <- w$w*b500/a500
btc500.p[i] <- pchisq(w.bc500,b500,lower=F)
a1000 <- mean(out$t[1:1000])
b1000 <- (2*a1000^2)/var(out$t[1:1000])
w.bc1000 <- w$w*b1000/a1000
btc1000.p[i] <- pchisq(w.bc1000,b1000,lower=F)
a <- mean(out$t)
b <- (2*a^2)/var(out$t)
w.bc <- w$w*b/a
btc.p[i] <- pchisq(w.bc,b,lower=F)
# directional p-values
#####
d2.p[i] <- directional.p3(y,group)
# Skovgaard p-values
#####
app <- W.skovgaard2(y,group)
sk.p1[i] <- app$p.value1
sk.p2[i] <- app$p.value2
}
list(first.order=fo.p,bootstrap500=bo500.p,bootstrap1000=bo1000.p,
      bootstrap=bo.p,bartlett500=bt500.p,bartlett1000=bt1000.p,bartlett=bt
      .p,bartlett.centered500=btc500.p,bartlett.centered1000=btc1000.p,
      bartlett.centered=btc.p,directional=d2.p,skovgaard1=sk.p1,skovgaard2
      =sk.p2,mu=mu,sigma2=sigma2,n=n)
}

```

```

summary.sim2 <- function(out, levels=c
  (0.01,0.025,0.05,0.10,0.25,0.5,0.75,0.9,0.95,0.975,0.99))
{
  fo <- sapply(levels,function(x) mean(out$first.order <x))
  dir2 <- sapply(levels,function(x) mean(out$directional <x))
  boot500 <- sapply(levels,function(x) mean(out$bootstrap500 <x))
  boot1000 <- sapply(levels,function(x) mean(out$bootstrap1000 <x))
  boot <- sapply(levels,function(x) mean(out$bootstrap <x))
  bart500 <- sapply(levels,function(x) mean(out$bartlett500 <x))
  bart1000 <- sapply(levels,function(x) mean(out$bartlett1000 <x))
  bart <- sapply(levels,function(x) mean(out$bartlett <x))
  bart.c500 <- sapply(levels,function(x) mean(out$bartlett.centered500 <x)
    )
  bart.c1000 <- sapply(levels,function(x) mean(out$bartlett.centered1000 <
    x))
  bart.c <- sapply(levels,function(x) mean(out$bartlett.centered <x))
  sk1 <- sapply(levels,function(x) mean(out$skovgaard1 <x))
  sk2 <- sapply(levels,function(x) mean(out$skovgaard2 <x))
  sim.se <- sqrt(levels*(1-levels)/length(out$first.order))
  res <- rbind(levels, fo, dir2, boot500, boot1000, boot, bart500, bart1000, bart,
    bart.c500, bart.c1000, bart.c, sk1, sk2, sim.se)
  rownames(res) <- c("nominal", "first-order", "directional", "bootstrap500",
    "bootstrap1000", "bootstrap", "bartlett500", "bartlett1000", "bartlett",
    "bartlett.centered500", "bartlett.centered1000", "bartlett.centered", "
    Skovgaard1", "Skovgaard2", "sim-error")
  res
}

plot.sim2 <- function(out, step=1)
{
  index=seq(1, length(out$first.order), by=step)
  par(mfrow=c(4, 3), cex=0.6, mar=c(5, 5, 5, 5), mex=0.8)
  x=sort(out$first.order)
  x=x[index]
  plot(ppoints(x), x, xlab='Uniform quantiles', ylab='p-values', pch=".")
  abline(0, 1, col=2)
}

```

```
title("Likelihood ratio statistic")
x=sort(out$bootstrap500)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bootstrap500")
x=sort(out$bootstrap1000)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bootstrap1000")
x=sort(out$bootstrap)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bootstrap")
x=sort(out$bartlett500)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett's correction500")
x=sort(out$bartlett1000)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett's correction1000")
x=sort(out$bartlett)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett's correction")
x=sort(out$bartlett.centered500)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett's correction centered500")
```

```

x=sort(out$bartlett.centered1000)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett's correction centered1000")
x=sort(out$bartlett.centered)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("Bartlett's correction centered")
x=sort(out$directional)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab='p-values',pch=".")
abline(0,1,col=2)
title("directional")
x=sort(out$skovgaard1)
x=x[index]
plot(ppoints(x),x,xlab='Uniform quantiles',ylab="p-values",pch=".")
abline(0,1,col=2)
# title(expression(paste("Skovgaard's ",W^"*")))
title("Skovgaard's W*")
###
# Skovgaard's W** is always worse than W*
# hence we skip it
}

```

Codice A.2: Codice utilizzato per le analisi e le simulazioni

```

# exponential samples
#####
# first we try the different methods on the aircondit dataset, then we do
# a simulation to explore the validity of them
## data loading ##
aircondit<-read.table("aircondit.dat",header=T)
aircondit$group <- factor(aircondit$group)

```

```
plot(aircondit$y~aircondit$group,xlab='number',ylab='time',main="Time
  intervals between successive failures (hours)")
dev.copy2eps(file='aircondit-data.eps')
n <- tapply(aircondit$y,aircondit$group,length)
## (roughly) check graphically exponentiality in each sample ##
par(mfrow=c(2,5))
for (i in 1:10)
{
  x=aircondit$y[aircondit$group==i]
  plot(qexp(ppoints(x),rate=1/mean(x)),sort(x),ylab="Sample Quantiles",
    xlab='Theoretical Quantiles')
  abline(0,1)
}

## first order ##
w<-W.fun(aircondit)
w$w
I <- nlevels(aircondit$group)
alpha.oss <- pchisq(w$w,I-1,lower=FALSE)
alpha.oss

## directional p-value ##
dir1 <- directional.p(aircondit$y,aircondit$group)
dir1
dir2 <- directional.p2(aircondit$y,aircondit$group)
dir2

## Skovgaard p-value ##
sko <- W.skovgaard(aircondit$y,aircondit$group)
sko1 <- sko$p.value1
sko2 <- sko$p.value2
sko1
sko2

## bootstrap ##
dati <- data.frame(y=aircondit$y,group=aircondit$group)
```

```
out <- boot(dati,w.funbo,sim="parametric",ran.gen=sim.fun,mle=w$tt,R=2000,
  parallel="multicore",ncpus=3)
bo500.p <- mean(out$t[1:500]>out$t0) # testing variability of p-values
  with different sample size
bo1000.p <- mean(out$t[1:1000]>out$t0)
bo.p <- mean(out$t>out$t0)
bo500.p
bo1000.p
bo.p

## Bartlett correction ##
w.b500 <- w$w*(I-1)/mean(out$t[1:500]) # testing variability of p-values
  with different sample size
ba500.p <- pchisq(w.b500,I-1,lower=F)
w.b1000 <- w$w*(I-1)/mean(out$t[1:1000])
ba1000.p <- pchisq(w.b1000,I-1,lower=F)
w.b <- w$w*(I-1)/mean(out$t)
ba.p <- pchisq(w.b,I-1,lower=F)
ba500.p
ba1000.p
ba.p

## Bartlett correction centered ##
a500 <- mean(out$t[1:500]) # testing variability of p-values with
  different sample size
b500 <- (2*a500^2)/var(out$t[1:500])
w.bc500 <- w$w*b500/a500
bace500.p <- pchisq(w.bc500,b500,lower=F)
a1000 <- mean(out$t[1:1000])
b1000 <- (2*a1000^2)/var(out$t[1:1000])
w.bc1000 <- w$w*b1000/a1000
bace1000.p <- pchisq(w.bc1000,b1000,lower=F)
a <- mean(out$t)
b <- (2*a^2)/var(out$t)
w.bc <- w$w*b/a
bace.p <- pchisq(w.bc,b,lower=F)
```

```

bace500.p
bace1000.p
bace.p
es.acc1 <- data.frame(alpha.oss, dir1, dir2, sko1, sko2, bo500.p, bo1000.p
  , bo.p, ba500.p, ba1000.p, ba.p, bace500.p, bace1000.p, bace.p, row.
  names=c("p.value"))
kable(es.acc1, format="latex")

## simulation with 100 groups of sample size 5 ##

es1g100n5.sim <- bt.simula(theta=w$tt,n=rep(5,100),Nsim=10^4)
save(es1g100n5.sim,file="es1g100n5sim.RData")
plot.sim(es1g100n5.sim) # graphical check
round(summary.sim(es1g100n5.sim),3) # table check

# exploring p-values accuracy as B changes
b500 <- c(mean(es1g100n5.sim$bootstrap500),var(es1g100n5.sim$bootstrap500)
  )
b1000 <- c(mean(es1g100n5.sim$bootstrap1000),var(es1g100n5.sim$
  bootstrap1000))
b2000 <- c(mean(es1g100n5.sim$bootstrap),var(es1g100n5.sim$bootstrap))
ba500 <- c(mean(es1g100n5.sim$bartlett500),var(es1g100n5.sim$bartlett500))
ba1000 <- c(mean(es1g100n5.sim$bartlett1000),var(es1g100n5.sim$
  bartlett1000))
ba2000 <- c(mean(es1g100n5.sim$bartlett),var(es1g100n5.sim$bartlett))
bac500 <- c(mean(es1g100n5.sim$bartlett.centered500),var(es1g100n5.sim$
  bartlett.centered500))
bac1000 <- c(mean(es1g100n5.sim$bartlett.centered1000),var(es1g100n5.sim$
  bartlett.centered1000))
bac2000 <- c(mean(es1g100n5.sim$bartlett.centered),var(es1g100n5.sim$
  bartlett.centered))
p.acc <- data.frame(b500,b1000,b2000,ba500,ba1000,ba2000,bac500,bac1000,
  bac2000,row.names = c("mean","var"))
kable(p.acc,format="latex")

## simulation with 100 groups of sample size 10 ##

```



```
es1g100n10.sim <- bt.simula(theta=w$tt,n=rep(10,100),Nsim=10^4)
save(es1g100n10.sim,file="es1g100n10sim.RData")
plot.sim(es1g100n10.sim) # graphical check
round(summary.sim(es1g100n10.sim),3) # table check

# exploring p-values accuracy as B changes
b500 <- c(mean(es1g100n10.sim$bootstrap500),var(es1g100n10.sim$
  bootstrap500))
b1000 <- c(mean(es1g100n10.sim$bootstrap1000),var(es1g100n10.sim$
  bootstrap1000))
b2000 <- c(mean(es1g100n10.sim$bootstrap),var(es1g100n10.sim$bootstrap))
ba500 <- c(mean(es1g100n10.sim$bartlett500),var(es1g100n10.sim$bartlett500
  ))
ba1000 <- c(mean(es1g100n10.sim$bartlett1000),var(es1g100n10.sim$
  bartlett1000))
ba2000 <- c(mean(es1g100n10.sim$bartlett),var(es1g100n10.sim$bartlett))
bac500 <- c(mean(es1g100n10.sim$bartlett.centered500),var(es1g100n10.sim$
  bartlett.centered500))
bac1000 <- c(mean(es1g100n10.sim$bartlett.centered1000),var(es1g100n10.sim
  $bartlett.centered1000))
bac2000 <- c(mean(es1g100n10.sim$bartlett.centered),var(es1g100n10.sim$
  bartlett.centered))
p.acc1 <- data.frame(b500,b1000,b2000,ba500,ba1000,ba2000,bac500,bac1000,
  bac2000,row.names = c("mean","var"))
kable(p.acc1, format="latex")

## simulation with 1000 groups of sample size 5 ##

es1g1000n5.sim <- bt.simula(theta=w$tt,n=rep(5,1000),Nsim=10^4)
save(es1g1000n5.sim,file="es1g1000n5sim.RData")
plot.sim(es1g1000n5.sim) # graphical check
round(summary.sim(es1g1000n5.sim),3) # table check

# exploring p-values accuracy as B changes
b500 <- c(mean(es1g1000n5.sim$bootstrap500),var(es1g1000n5.sim$
  bootstrap500))
```

```

b1000 <- c(mean(es1g1000n5.sim$bootstrap1000),var(es1g1000n5.sim$
  bootstrap1000))
b2000 <- c(mean(es1g1000n5.sim$bootstrap),var(es1g1000n5.sim$bootstrap))
ba500 <- c(mean(es1g1000n5.sim$bartlett500),var(es1g1000n5.sim$bartlett500
  ))
ba1000 <- c(mean(es1g1000n5.sim$bartlett1000),var(es1g1000n5.sim$
  bartlett1000))
ba2000 <- c(mean(es1g1000n5.sim$bartlett),var(es1g1000n5.sim$bartlett))
bac500 <- c(mean(es1g1000n5.sim$bartlett.centered500),var(es1g1000n5.sim$
  bartlett.centered500))
bac1000 <- c(mean(es1g1000n5.sim$bartlett.centered1000),var(es1g1000n5.sim
  $bartlett.centered1000))
bac2000 <- c(mean(es1g1000n5.sim$bartlett.centered),var(es1g1000n5.sim$
  bartlett.centered))
p.acc2 <- data.frame(b500,b1000,b2000,ba500,ba1000,ba2000,bac500,bac1000,
  bac2000,row.names = c("mean","var"))
kable(p.acc2, format="latex")

# normal samples
#####

# first we try the different methods on the gravity dataset, then we do a
  simulation to explore the validity of them

## data loading ##
gravity<-read.table("gravity.txt",header=T)
gravity$group <- factor(gravity$group)
plot(gravity$y~gravity$group,xlab='experiment',ylab='deviation',main="
  Deviation from the estimate of 980000cm/s^2")
dev.copy2eps(file='gravity-data.eps')

## first order ##
w2<-W.funCVN(gravity)
w2$w
I <- nlevels(gravity$group)

```

```
alpha.oss <- pchisq(w2$w,I-1,lower=FALSE)
alpha.oss

## directional p-value ##
dir3 <- directional.p3(gravity$y,gravity$group)
dir3

## Skovgaard p-value ##
sko <- W.skovgaard2(gravity$y,gravity$group)
sko3 <- sko$p.value1
sko4 <- sko$p.value2
sko3
sko4

## bootstrap ##
dati2 <- data.frame(y=gravity$y,group=gravity$group)
out <- boot(dati2,W.funboCVN,sim="parametric",ran.gen=sim.funCVN,mle=w2$
  s2hb,R=2000,parallel="multicore",ncpus=3)
bo500.p <- mean(out$t[1:500]>out$t0) # testing variability of p-values
  with different sample size
bo1000.p <- mean(out$t[1:1000]>out$t0)
bo.p <- mean(out$t>out$t0)
bo500.p
bo1000.p
bo.p

## Bartlett correction ##
w.b500 <- w2$w*(I-1)/mean(out$t[1:500]) # testing variability of p-values
  with different sample size
ba500.p <- pchisq(w.b500,I-1,lower=F)
w.b1000 <- w2$w*(I-1)/mean(out$t[1:1000])
ba1000.p <- pchisq(w.b1000,I-1,lower=F)
w.b <- w2$w*(I-1)/mean(out$t)
ba.p <- pchisq(w.b,I-1,lower=F)
ba500.p
ba1000.p
```

```

ba.p

## Bartlett correction centered ##
a500 <- mean(out$t[1:500]) # testing variability of p-values with
  different sample size
b500 <- (2*a500^2)/var(out$t[1:500])
w.bc500 <- w2*w*b500/a500
bace500.p <- pchisq(w.bc500,b500,lower=F)
a1000 <- mean(out$t[1:1000])
b1000 <- (2*a1000^2)/var(out$t[1:1000])
w.bc1000 <- w2*w*b1000/a1000
bace1000.p <- pchisq(w.bc1000,b1000,lower=F)
a <- mean(out$t)
b <- (2*a^2)/var(out$t)
w.bc <- w2*w*b/a
bace.p <- pchisq(w.bc,b,lower=F)
bace500.p
bace1000.p
bace.p
es.acc2 <- data.frame(alpha.oss, dir3, sko3, sko4, bo500.p, bo1000.p, bo.p
  , ba500.p, ba1000.p, ba.p, bace500.p, bace1000.p, bace.p, row.names=c(
  "p.value"))
kable(es.acc2, format="latex",digits=15)

## simulation with 100 groups of sample size 5 ##
es2g100n5.sim <- bt.simula2(mu=2*(100-1:100),sigma2=1,n=rep(5,100),Nsim
  =10^4)
save(es2g100n5.sim,file="es2g100n5sim.RData")
plot.sim2(es2g100n5.sim) # graphical check
round(summary.sim2(es2g100n5.sim),3) # table check

# exploring p-values accuracy as B changes
b500 <- c(mean(es2g100n5.sim$bootstrap500),var(es2g100n5.sim$bootstrap500)
  )
b1000 <- c(mean(es2g100n5.sim$bootstrap1000),var(es2g100n5.sim$
  bootstrap1000))

```

```
b2000 <- c(mean(es2g100n5.sim$bootstrap), var(es2g100n5.sim$bootstrap))
ba500 <- c(mean(es2g100n5.sim$bartlett500), var(es2g100n5.sim$bartlett500))
ba1000 <- c(mean(es2g100n5.sim$bartlett1000), var(es2g100n5.sim$
  bartlett1000))
ba2000 <- c(mean(es2g100n5.sim$bartlett), var(es2g100n5.sim$bartlett))
bac500 <- c(mean(es2g100n5.sim$bartlett.centered500), var(es2g100n5.sim$
  bartlett.centered500))
bac1000 <- c(mean(es2g100n5.sim$bartlett.centered1000), var(es2g100n5.sim$
  bartlett.centered1000))
bac2000 <- c(mean(es2g100n5.sim$bartlett.centered), var(es2g100n5.sim$
  bartlett.centered))
p.acc3 <- data.frame(b500, b1000, b2000, ba500, ba1000, ba2000, bac500, bac1000,
  bac2000, row.names = c("mean", "var"))
kable(p.acc3, format="latex")

## simulation with 100 groups of sample size 10 ##
es2g100n10.sim <- bt.simula2(mu=2*(100-1:100), sigma2=1, n=rep(10, 100), Nsim
  =10^4)
save(es2g100n10.sim, file="es2g100n10sim.RData")
plot.sim2(es2g100n10.sim) # graphical check
round(summary.sim2(es2g100n10.sim), 3) # table check

# exploring p-values accuracy as B changes
b500 <- c(mean(es1g100n10.sim$bootstrap500), var(es1g100n10.sim$
  bootstrap500))
b1000 <- c(mean(es1g100n10.sim$bootstrap1000), var(es1g100n10.sim$
  bootstrap1000))
b2000 <- c(mean(es1g100n10.sim$bootstrap), var(es1g100n10.sim$bootstrap))
ba500 <- c(mean(es1g100n10.sim$bartlett500), var(es1g100n10.sim$bartlett500
  ))
ba1000 <- c(mean(es1g100n10.sim$bartlett1000), var(es1g100n10.sim$
  bartlett1000))
ba2000 <- c(mean(es1g100n10.sim$bartlett), var(es1g100n10.sim$bartlett))
bac500 <- c(mean(es1g100n10.sim$bartlett.centered500), var(es1g100n10.sim$
  bartlett.centered500))
bac1000 <- c(mean(es1g100n10.sim$bartlett.centered1000), var(es1g100n10.sim
```

```

    $bartlett.centered1000))
bac2000 <- c(mean(es1g100n10.sim$bartlett.centered),var(es1g100n10.sim$
  bartlett.centered))
p.acc4 <- data.frame(b500,b1000,b2000,ba500,ba1000,ba2000,bac500,bac1000,
  bac2000,row.names = c("mean","var"))
kable(p.acc4, format="latex")

## simulation with 1000 groups of sample size 5 ##
es2g1000n5.sim <- bt.simula2(mu=2*(1000-1:1000),sigma2=1,n=rep(5,1000),
  Nsim=10^4)
save(es2g1000n5.sim,file="es2g1000n5sim.RData")
plot.sim2(es2g1000n5.sim) # graphical check
round(summary.sim2(es2g1000n5.sim),3) # table check

# exploring p-values accuracy as B changes
b500 <- c(mean(es1g1000n5.sim$bootstrap500),var(es1g1000n5.sim$
  bootstrap500))
b1000 <- c(mean(es1g1000n5.sim$bootstrap1000),var(es1g1000n5.sim$
  bootstrap1000))
b2000 <- c(mean(es1g1000n5.sim$bootstrap),var(es1g1000n5.sim$bootstrap))
ba500 <- c(mean(es1g1000n5.sim$bartlett500),var(es1g1000n5.sim$bartlett500
  ))
ba1000 <- c(mean(es1g1000n5.sim$bartlett1000),var(es1g1000n5.sim$
  bartlett1000))
ba2000 <- c(mean(es1g1000n5.sim$bartlett),var(es1g1000n5.sim$bartlett))
bac500 <- c(mean(es1g1000n5.sim$bartlett.centered500),var(es1g1000n5.sim$
  bartlett.centered500))
bac1000 <- c(mean(es1g1000n5.sim$bartlett.centered1000),var(es1g1000n5.sim
  $bartlett.centered1000))
bac2000 <- c(mean(es1g1000n5.sim$bartlett.centered),var(es1g1000n5.sim$
  bartlett.centered))
p.acc5 <- data.frame(b500,b1000,b2000,ba500,ba1000,ba2000,bac500,bac1000,
  bac2000,row.names = c("mean","var"))
kable(p.acc5, format="latex")

```