

# Asymptotic expansion for the test for sphericity

Max Turgeon

03/02/2020

We assume  $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim N_p(\mu, \Sigma)$  with  $\Sigma$  positive definite. Write  $V = n\hat{\Sigma}$ , where  $(\bar{\mathbf{Y}}, \hat{\Sigma})$  is the (unrestricted) MLE for the multivariate normal distribution.

*Sphericity* means the different components of  $\mathbf{Y}$  are **uncorrelated** and have the **same variance**. In other words, we are looking at the following null hypothesis:

$$H_0 : \Sigma = \sigma^2 I_p, \quad \sigma^2 > 0.$$

## Likelihood Ratio Test

We have

$$\begin{aligned} L(\hat{\mathbf{Y}}, \sigma^2 I_p) &= (2\pi)^{-np/2} |\sigma^2 I_p|^{-n/2} \exp\left(-\frac{1}{2} \text{tr}((\sigma^2 I_p)^{-1} V)\right) \\ &= (2\pi\sigma^2)^{-np/2} \exp\left(-\frac{1}{2\sigma^2} \text{tr}(V)\right). \end{aligned}$$

Taking the derivative of the logarithm and setting it equal to zero, we find that  $L(\hat{\mathbf{Y}}, \sigma^2 I_p)$  is maximised when

$$\widehat{\sigma^2} = \frac{\text{tr} V}{np}.$$

We then get

$$\begin{aligned} L(\hat{\mathbf{Y}}, \widehat{\sigma^2} I_p) &= (2\pi\widehat{\sigma^2})^{-np/2} \exp\left(-\frac{1}{2\widehat{\sigma^2}} \text{tr}(V)\right) \\ &= (2\pi)^{-np/2} \left(\frac{\text{tr} V}{np}\right)^{-np/2} \exp\left(-\frac{np}{2}\right). \end{aligned}$$

Therefore, we have

$$\begin{aligned} \Lambda &= \frac{(2\pi)^{-np/2} \left(\frac{\text{tr} V}{np}\right)^{-np/2} \exp\left(-\frac{np}{2}\right)}{\exp(-np/2) (2\pi)^{-np/2} |\hat{\Sigma}|^{-n/2}} \\ &= \frac{\left(\frac{\text{tr} V}{np}\right)^{-np/2}}{|n^{-1} V|^{-n/2}} \\ &= \left(\frac{|V|}{(\text{tr} V/p)^p}\right)^{n/2}. \end{aligned}$$

We can also rewrite this as follows: let  $l_1 \geq \dots \geq l_p$  be the eigenvalues of  $V$ . We have

$$\begin{aligned}\Lambda^{2/n} &= \frac{|V|}{(\text{tr}V/p)^p} \\ &= \frac{\prod_{j=1}^p l_j}{\left(\frac{1}{p} \sum_{j=1}^p l_j\right)^p} \\ &= \left( \frac{\prod_{j=1}^p l_j^{1/p}}{\frac{1}{p} \sum_{j=1}^p l_j} \right)^p.\end{aligned}$$

In other words, the modified LRT  $\tilde{\Lambda} = \Lambda^{2/n}$  is the ratio of the geometric to the arithmetic mean of the eigenvalues of  $V$  (all raised to the power  $p$ ).

Note that under  $H_0$ , there is only one free parameter, namely  $\sigma^2$ . Therefore the asymptotic theory of likelihood ratio tests implies that

$$-2 \log \Lambda \rightarrow \chi^2 \left( \frac{1}{2} p(p+1) - 1 \right).$$

We will provide a better approximation using asymptotic expansions.

First, we need to compute the moments of  $\Lambda$ . We start with the following lemma:

**Lemma**

Under the null hypothesis, the random variables  $\text{tr}V$  and  $\frac{\det V}{(\text{tr}V)^p}$  are independent.

*Proof*

Recall that  $V \sim W_p(n-1, \sigma^2 I_p)$ , and so its distribution only depends on  $\sigma^2$ . Hence, the distribution of  $\frac{\det V}{(\text{tr}V)^p}$  does *not* depend on  $\sigma^2$ , and therefore it is an ancillary statistic.

Now, given that  $\widehat{\sigma^2} = \frac{\text{tr}V}{np}$  and given that the multivariate normal forms an exponential family, we know that  $(\tilde{\mathbf{Y}}, \text{tr}V)$  is a minimal sufficient and complete statistic. Therefore, we can conclude by using Basu's theorem.  $\square$

Now, going back to  $\tilde{\Lambda}$ , note that we have

$$\tilde{\Lambda} \left( \frac{1}{p} \text{tr}V \right)^p = |V|.$$

Using our lemma above, for any  $h$ , we can write

$$\begin{aligned}E|V|^h &= E \left( \tilde{\Lambda} \left( \frac{1}{p} \text{tr}V \right)^p \right)^h \\ &= E \tilde{\Lambda}^h E \left( \frac{1}{p} \text{tr}V \right)^{ph}.\end{aligned}$$

In other words, we have

$$E(\tilde{\Lambda}^h) = \frac{E(|V|^h)}{E \left( \left( \frac{1}{p} \text{tr}V \right)^{ph} \right)}.$$

Recall the following two results:

1. If  $W \sim W_p(m, I_p)$ , then  $\text{tr}W \sim \chi^2(mp)$ .
2. If  $W \sim W_p(m, \Sigma)$ , then  $|W| \sim |\Sigma| \prod_{j=1}^p \chi^2(m-p+j)$ .

Therefore, we can get all the moments of  $\tilde{\Lambda}$  from the moments of chi-squared distributions.

### Proposition

Let  $X \sim \chi^2(d)$ . Then for  $h > -\frac{1}{2}d$ , we have

$$E(X^h) = 2^h \frac{\Gamma(\frac{1}{2}d + h)}{\Gamma(\frac{1}{2}d)}.$$

Putting all this together, we get the following theorem:

### Theorem

The moments of the modified LRT statistic are given by

$$E(\tilde{\Lambda}^h) = p^{ph} \frac{\Gamma(\frac{1}{2}(n-1)p)}{\Gamma(\frac{1}{2}(n-1)p + ph)} \frac{\Gamma_p(\frac{1}{2}(n-1) + h)}{\Gamma_p(\frac{1}{2}(n-1))}.$$

*Proof*

This follows from our discussion above, the moments of the chi-squared distribution, and the fact that  $V \sim \sigma^2 W_p(n-1)$ .  $\square$

### Asymptotic expansion

For this section, we are following Chapter 12 of Bilodeau & Brenner. In a 1949 *Biometrika* paper, George Box studied the distribution theory of a very general class of likelihood ratio tests. It can be applied any time the moments of the likelihood ratio  $\Lambda$  (or some power  $W = \Lambda^d$  thereof) have the following expression:

$$E(W^h) = K \left( \frac{\prod_{j=1}^b y_j^{y_j}}{\prod_{k=1}^a x_k^{x_k}} \right)^h \frac{\prod_{k=1}^a \Gamma(x_k(1+h) + \zeta_k)}{\prod_{j=1}^b \Gamma(y_j(1+h) + \eta_j)}, \quad (1)$$

such that

$$\sum_{j=1}^b y_j = \sum_{k=1}^a x_k$$

and  $K$  is a constant such that  $E(\tilde{\Lambda}^0) = 1$ .

In the context of the test for sphericity, we can take  $W = \Lambda^{m/n}$  with  $m = n - 1$ , and we get

$$\begin{aligned} E(W^h) &= E(\Lambda^{mh/n}) \\ &= E(\tilde{\Lambda}^{mh/2}) \\ &= p^{pmh/2} \frac{\Gamma(\frac{1}{2}mp)}{\Gamma(\frac{1}{2}mp + \frac{1}{2}pmh)} \frac{\Gamma_p(\frac{1}{2}m + \frac{1}{2}mh)}{\Gamma_p(\frac{1}{2}m)} \\ &= \left( \frac{\Gamma(\frac{1}{2}mp)}{\Gamma_p(\frac{1}{2}m)} \right) p^{pmh/2} \frac{\Gamma_p(\frac{1}{2}m(1+h))}{\Gamma(\frac{1}{2}mp(1+h))} \\ &= \left( \pi^{p(p-1)/4} \frac{\Gamma(\frac{1}{2}mp)}{\Gamma_p(\frac{1}{2}m)} \right) \left( p^{pm/2} \right)^h \frac{\prod_{k=1}^p \Gamma_p(\frac{1}{2}m(1+h) - \frac{1}{2}(k-1))}{\Gamma(\frac{1}{2}mp(1+h))}. \end{aligned}$$

This is consistent with the general form above, if we take

$$\begin{aligned} a &= p, & x_k &= \frac{1}{2}m, & \zeta_k &= -\frac{1}{2}(k-1), \\ b &= 1, & y_1 &= \frac{1}{2}mp, & \eta_1 &= 0, \end{aligned}$$

and

$$K = \pi^{p(p-1)/4} \frac{\Gamma(\frac{1}{2}mp)}{\Gamma_p(\frac{1}{2}m)}.$$

Going back to Equation 1, let  $M = -2 \log W$ . For a real number  $0 < \rho \leq 1$ , we can look at the *characteristic function* of  $\rho M$ :

$$\begin{aligned} \varphi_{\rho M}(t) &= E(\exp(it\rho M)) \\ &= E(W^{-2it\rho}) \\ &= K \left( \frac{\prod_{j=1}^b y_j^{y_j}}{\prod_{k=1}^a x_k^{x_k}} \right)^{-2it\rho} \frac{\prod_{k=1}^a \Gamma(x_k(1-2it\rho) + \zeta_k)}{\prod_{j=1}^b \Gamma(y_j(1-2it\rho) + \eta_j)}. \end{aligned}$$

Taking the logarithm of the characteristic function gives us the *cumulant function*:

$$\begin{aligned} K_{\rho M}(t) &= \log \varphi_{\rho M}(t) \\ &= \log K - 2it\rho \left( \sum_{j=1}^b y_j \log y_j - \sum_{k=1}^a x_k \log x_k \right) \\ &\quad + \sum_{k=1}^a \log \Gamma(x_k(1-2it\rho) + \zeta_k) - \sum_{j=1}^b \log \Gamma(y_j(1-2it\rho) + \eta_j). \end{aligned}$$

If we set  $\beta_k = (1-\rho)x_k$  and  $\epsilon_j = (1-\rho)y_j$ , we can write

$$K_{\rho M}(t) = g(t) - g(0),$$

where

$$\begin{aligned} g(t) &= 2it\rho \left( \sum_{k=1}^a x_k \log x_k - \sum_{j=1}^b y_j \log y_j \right) \\ &\quad + \sum_{k=1}^a \log \Gamma(\rho x_k(1-2it) + \beta_k + \zeta_k) - \sum_{j=1}^b \log \Gamma(\rho y_j(1-2it) + \epsilon_j + \eta_j). \end{aligned}$$

Next, we can approximate the cumulant function by using the following asymptotic expansion of the log-gamma function: for  $h$  bounded and  $|z| \rightarrow \infty$ , we have

$$\begin{aligned} \log \Gamma(z+h) &= \log \sqrt{2\pi} + \left( z+h - \frac{1}{2} \right) \log z - z \\ &\quad - \sum_{\alpha=1}^l \frac{B_{\alpha+1}(h)}{\alpha(\alpha+1)} z^{-\alpha} + O(z^{-(l+1)}), \end{aligned}$$

where  $B_r(h)$  are Bernoulli polynomials. We can control the accuracy of our approximation by choosing  $l$  appropriately.

Using this expansion with the function  $g$ , we get

$$K_{\rho M}(t) = -\frac{1}{2}f \log(1-2it) + \sum_{\alpha=1}^l \omega_{\alpha}((1-2it)^{-\alpha} - 1) + O(n^{-(l+1)}),$$

where

$$f = -2 \left( \sum_{k=1}^a \zeta_k - \sum_{j=1}^b \eta_j - \frac{1}{2}(a-b) \right),$$

and

$$\omega_\alpha = \frac{(-1)^{\alpha+1}}{\alpha(\alpha+1)} \left( \sum_{k=1}^a \frac{B_{\alpha+1}(\beta_k + \zeta_k)}{(\rho x_k)^\alpha} - \sum_{j=1}^b \frac{B_{\alpha+1}(\epsilon_j + \eta_j)}{(\rho y_j)^\alpha} \right).$$

Two important observations:

1.  $\beta_k$  and  $\epsilon_j$  are  $O(1)$ ;
2. If  $x_k$  and  $y_j$  are  $O(n)$ , then  $\omega_\alpha$  is  $O(n^{-\alpha})$ .

Now, we have an approximation of the cumulant function, which can be converted to an approximation of the characteristic function through exponentiation:

$$\begin{aligned} \varphi_{\rho M}(t) &= \exp(K_{\rho M}(t)) \\ &= \exp \left( -\frac{1}{2}f \log(1-2it) + \sum_{\alpha=1}^l \omega_\alpha ((1-2it)^{-\alpha} - 1) + O(n^{-(l+1)}) \right) \\ &= (1-2it)^{-f/2} \prod_{\alpha=1}^l \exp(\omega_\alpha ((1-2it)^{-\alpha} - 1)) \cdot O(1+n^{-(l+1)}) \\ &= (1-2it)^{-f/2} \prod_{\alpha=1}^l \exp(\omega_\alpha (1-2it)^{-\alpha}) \prod_{\alpha=1}^l \exp(-\omega_\alpha) + O(n^{-(l+1)}) \\ &= (1-2it)^{-f/2} \prod_{\alpha=1}^l \sum_{k=0}^{\infty} \frac{\omega_\alpha^k}{k!} (1-2it)^{-\alpha k} \prod_{\alpha=1}^l \sum_{k=0}^{\infty} (-1)^k \frac{\omega_\alpha^k}{k!} + O(n^{-(l+1)}). \end{aligned}$$

We can then get an approximation of order  $l+1$  by computing terms of order up to  $l$  in the Taylor expansions. For example, if we want an approximation of order 2, we compute all linear terms:

$$\begin{aligned} \varphi_{\rho M}(t) &= (1-2it)^{-f/2} (1 + \omega_1(1-2it)^{-1}) (1 - \omega_1) + O(n^{-2}) \\ &= (1-2it)^{-f/2} + \omega_1 \left( (1-2it)^{-(f+2)/2} - (1-2it)^{-f/2} \right) + O(n^{-2}) \\ &= \varphi_f(t) + \omega_1(\varphi_{f+2}(t) - \varphi_f(t)) + O(n^{-2}), \end{aligned}$$

where  $\varphi_f(t)$  is the characteristic function of a chi-square on  $f$  degrees of freedom. By the inversion formula, we have that the density of  $\rho M$  is

$$\chi^2(f) + \omega_1(\chi^2(f+2) - \chi^2(f)) + O(n^{-2}).$$

Remember that we let  $\rho$  be arbitrary; it turns out that we can choose  $\rho$  so that  $\omega_1 = 0$ :

$$\rho = 1 - f^{-1} \left( \sum_{k=1}^a x_k^{-1} (\zeta_k^2 - \zeta_k + \frac{1}{6}) - \sum_{j=1}^b y_j^{-1} (\eta_j^2 - \eta_j + \frac{1}{6}) \right). \quad (2)$$

We have essentially proven (minus some technical details) the following approximation result:

### Theorem (Order 2 approximation)

If  $W$  has moments as in Equation 1, where  $x_k$  and  $y_j$  are  $O(1)$ , then with the choice of  $\rho$  as in Equation 2, we have

$$P(\rho M \leq x) = F_{\chi^2}(x; f) + O(n^{-2}),$$

where  $F_{\chi^2}(x; f) = P(\chi^2(f) \leq x)$ .

Note that this is still a chi-square approximation, just as in the general asymptotic theory of likelihood ratio tests. However, this approximation is generally much better. Moreover, we can make the approximation even more precise by computing more terms in the Taylor expansion. Finally, it is possible to use bootstrap to increase the accuracy in the Theorem above to  $O(n^{-3})$  (see Chapter 14 of Bilodeau & Brenner).

Now, going back to the test of sphericity, we need to compute  $f$  and  $\rho$ :

$$f = \frac{1}{2}(p+2)(p-1),$$
$$\rho = 1 - \frac{2p^2 + p + 2}{6p(n-1)}.$$

In other words, if we let  $\Lambda$  be the likelihood ratio for the test of sphericity, we have

$$-2 \left( \frac{6p(n-1) - (2p^2 + p + 2)}{6pn} \right) \log \Lambda \approx \chi^2 \left( \frac{1}{2}p(p+1) - 1 \right).$$

Compared to the general theory, we are multiplying  $-2 \log \Lambda$  by a constant and correcting the degrees of freedom. This is known as a *Bartlett correction*.

### Simulation

We will compare the two approximations using a short simulation study. We will let  $n = 10$ ,  $p = 2$  and  $\sigma^2 = 1$ .

```
set.seed(7200)

# Simulation parameters
n <- 10
p <- 2
B <- 1000

# Generate data
lrt_dist <- replicate(B, {
  Y <- matrix(rnorm(n*p), ncol = p)
  V <- crossprod(Y)

  # log Lambda
  0.5*n*(log(det(V)) - p*log(mean(diag(V))))
})
```

First, we will look at the general asymptotic theory:

```
df <- choose(p + 1, 2)
general_chisq <- rchisq(B, df = df)
```

Next, we will look at Bartlett's correction:

```
df <- choose(p + 1, 2) - 1
const <- (6*p*(n-1) - (2*p^2 + p + 2))/(6*p*n)
bartlett_chisq <- rchisq(B, df = df)/const
```

We can compare them using the *empirical CDF*:

```
plot(ecdf(-2*lrt_dist), main = "-2 log Lambda")
lines(ecdf(general_chisq), col = 'blue')
lines(ecdf(bartlett_chisq), col = 'red')
legend('bottomright', legend = c("-2log Lambda", "General approx.", "Bartlett"),
      lty = 1, col = c('black', 'blue', 'red'))
```

