Chapter 2 Appendices

Appendix 2A. Distributional Results in the Central Case

2A.1. Distributions

Assume all distributions are central. We will take advantage of the well-known property of hierarchical (or nested) models that for $k \geq k_*$ and L > 0,

$$SSE_k - SSE_{k+L} \sim \sigma_*^2 \chi_L^2, \tag{2A.1}$$

$$SSE_k \sim \sigma_*^2 \chi_{n-k}^2 \tag{2A.2}$$

and

$$SSE_k - SSE_{k+L}$$
 is independent of SSE_{k+L} . (2A.3)

We will also need the distribution of linear combinations of SSE_k and SSE_{k+L} . Consider the linear combination $aSSE_k - bSSE_{k+L}$ where a and b are scalars. It follows from Eq. (2A.2) that

$$E[aSSE_k - bSSE_{k+L}] = a(n-k)\sigma_*^2 - b(n-k-L)\sigma_*^2$$
 (2A.4)

and

$$var[aSSE_k - bSSE_{k+L}] = var[aSSE_k - aSSE_{k+L} + aSSE_{k+L} - bSSE_{k+L}]$$
$$= var[a(SSE_k - SSE_{k+L}) + (a - b)SSE_{k+L}].$$

Applying Eqs. (2A.1)-(2A.3), we have

$$var[aSSE_k - bSSE_{k+L}] = 2a^2L\sigma_*^4 + 2(a-b)^2(n-k-L)\sigma_*^4.$$
 (2A.5)

Since many model selection criteria (such as AIC) use some function of $\log(\mathrm{SSE}_k)$, we will also introduce some useful distributional results involving $\log(\mathrm{SSE}_k)$. It can be shown (Gradshteyn, 1965, p. 576) that

$$\int_0^\infty z^{\nu-1} e^{-\mu z} \log(z) dz = \frac{1}{\mu^{\nu}} \Gamma(\nu) \big[\psi(\nu) - \log(\mu) \big], \ \mu > 0, \nu > 0,$$

where

$$\psi(\nu) = -C - \sum_{i=0}^{\infty} \left(\frac{1}{j+\nu} - \frac{1}{j+1} \right),$$

 $C = 0.577 \ 215 \ 664 \ 901$ is Eu $Z \sim \chi_m^2$ with m degrees of fr

$$E[\log(Z)] = \int_{\mathbb{R}^n} dz$$

which has no closed-form sol

$$E[\log(\mathrm{SSE}_k)]$$

Although ψ has no closed-fuseful for computing exact ϵ pp. 943–945):

 $\psi(v$

where

$$\psi(\frac{1}{2}) = -$$

This recursion will be used derived in Section 2.3 as we tion 2.6.

The distribution of differ involved. Since we have diffe

$$\log(SSE_{k+L})$$

Let $Q = SSE_{k+L}/SSE_k$. Ass (2A.3), we have

In nested models these two χ^2

 $Q \sim$

Since $Q = SSE_{k+L}/SSE_k$, the

$$\log \left(\frac{\text{SSE}_{k+1}}{\text{SSE}_k} \right)$$

dvantage of the wellor $k \ge k_*$ and L > 0,

(2A.1)

$$+L$$
 (2A.3)

nations of SSE_k and $_{+L}$ where a and b are

 $\mathbf{E}_{k+L} - b \mathbf{SSE}_{k+L}$] - $b) \mathbf{SSE}_{k+L}$].

$$-L)\sigma_*^4$$
. (2A.5)

use some function of nal results involving

$$u > 0, \nu > 0$$

Appendix 2A. Distributional Results in the Central Case

67

C = 0.577 215 664 901 is Euler's constant, and ψ is Euler's psi function. For $Z \sim \chi_m^2$ with m degrees of freedom,

$$\begin{split} E\big[\log(Z)\big] &= \int_0^\infty \log(z) \frac{1}{2^{m/2} \Gamma(m/2)} z^{m/2-1} e^{-z/2} dz \\ &= \log(2) + \psi\left(\frac{m}{2}\right), \end{split}$$

which has no closed-form solution. For $\mathrm{SSE}_k \sim \sigma_*^2 \chi_{n-k}^2$

$$E\left[\log(\mathrm{SSE}_k)\right] = \log(\sigma_*^2) + \log(2) + \psi\left(\frac{n-k}{2}\right). \tag{2A.6}$$

Although ψ has no closed-form solution, a simple recursion exists which is useful for computing exact expectations in small samples (Gradshteyn, 1965 pp. 943–945):

$$\psi(v+1) = \psi(v) + \frac{1}{v}, \ v > 0, \tag{2A.7}$$

where

$$\psi(\frac{1}{2}) = -C - 2\log(2) \text{ and } \psi(1) = -C.$$
 (2A.8)

This recursion will be used to check the accuracy of the Taylor expansion derived in Section 2.3 as well as in studying small-sample properties in Section 2.6.

The distribution of differences between $\log(\text{SSE}_k)$ and $\log(\text{SSE}_{k+L})$ is more involved. Since we have differences of logs,

$$\log(SSE_{k+L}) - \log(SSE_k) = \log\left(\frac{SSE_{k+L}}{SSE_k}\right)$$

Let $Q = SSE_{k+L}/SSE_k$. Assuming nested models and applying Eqs. (2A.1)–(2A.3), we have

$$Q \sim \frac{\chi_{n-k-L}^2}{\chi_{n-k-L}^2 + \chi_L^2}.$$

In nested models these two χ^2 are independent, and Q has the Beta distribution

$$Q \sim \operatorname{Beta}\left(\frac{n-k-L}{2}, \frac{L}{2}\right)$$
 .

Since $Q = SSE_{k+L}/SSE_k$, the log-distribution is

$$\log\left(\frac{\mathrm{SSE}_{k+L}}{\mathrm{SSE}_k}\right) \sim \log\text{-Beta}\left(\frac{n-k-L}{2}, \frac{L}{2}\right).$$
 (2A.9)

Appendix 2A

where

2A.2. Appr

AIC, Al computing r of $log(SSE_k)$ approximati: Beginning w Taylor expar Suppose

and

Numerically paring the re good for E le mation with tion is used t in Section 2.3 moments of l

2A.3. Appro Since Q =

However, ther $(= \log(Q))$ in (n-k-L)/(

log(

It can be shown (Gradshteyn, 1965 p. 538 and p. 541) that

$$\int_0^1 t^{\mu-1} (1-t^r)^{\nu-1} \log(t) dt = \frac{1}{r^2} \dot{B} \left(\frac{\mu}{r} + \nu, \nu \right) \left[\psi \left(\frac{\mu}{r} \right) - \psi \left(\frac{\mu}{r} + \nu \right) \right]$$

and

$$\begin{split} \int_0^1 t^{\mu-1} (1-t^r)^{\nu-1} \log^2(t) dt = & B\left(\frac{\mu}{r} + \nu, \nu\right) \left[\left(\psi\left(\frac{\mu}{r}\right) - \psi\left(\frac{\mu}{r} + \nu\right)\right)^2 \right. \\ & + \psi'(\mu) - \psi'(\mu + \nu) \right], \end{split}$$

where

$$\psi'(v) = \sum_{j=0}^{\infty} \frac{1}{(j+v)^2}$$

and $\mu > 0, \nu > 0$. For $Q \sim Beta(\frac{m}{2}, \frac{L}{2})$.

$$E\left[\log(Q)\right] = \int_0^1 \log(t)B^{-1}\left(\frac{m}{2}, \frac{L}{2}\right)t^{\frac{m}{2}-1}(1-t)^{\frac{L}{2}-1}dt$$
$$= \psi\left(\frac{m}{2}\right) - \psi\left(\frac{m}{2} + \frac{L}{2}\right)$$

and

$$\begin{split} E\left[\log^2(Q)\right] &= \int_0^1 \log^2(t) B^{-1}\left(\frac{m}{2}, \frac{L}{2}\right) t^{\frac{m}{2}-1} (1-t)^{\frac{L}{2}-1} dt \\ &= \left(\psi\left(\frac{m}{2}\right) - \psi\left(\frac{m}{2} + \frac{L}{2}\right)\right)^2 + \psi'\left(\frac{m}{2}\right) - \psi'\left(\frac{m}{2} + \frac{L}{2}\right). \end{split}$$

Hence,

$$var\left[\log(Q)\right] = \psi'\left(\frac{m}{2}\right) - \psi'\left(\frac{m}{2} + \frac{L}{2}\right)$$

and

$$var\left[\log\left(\frac{\mathrm{SSE}_{k+L}}{\mathrm{SSE}_{k}}\right)\right] = \psi'\left(\frac{n-k-L}{2}\right) - \psi'\left(\frac{n-k}{2}\right). \tag{2A.10}$$

Eq. (2A.10) also has no closed-form, but again, convenient recursions for ψ' exist which are useful in small samples (Gradshteyn, 1965 pp. 945-946):

$$\psi'\left(\frac{m}{2}\right) = \psi'\left(\frac{m}{2} - 1\right) - \frac{4}{(m-2)^2},$$
 (2A.11)