CHAPTER 11

Some proofs

This chapter contains some of the proofs of the empirical likelihood theorems. The arguments presented here are more difficult than the proofs that appear throughout the text. For the most advanced material the reader is referred to the published literature, as outlined in Chapter 13.

11.1 Lemmas

This section presents some lemmas needed to prove the ELT's. They are used to handle some technical details showing that the Lagrange multiplier λ is asymptotically small, and to show that higher order terms in some Taylor series can be neglected.

A distribution with nondegenerate variance matrix on $\mathbb{R}^p$ cannot put all of its probability on a half space defined as one side of a hyperplane through its mean. A stronger conclusion is that , for any distribution, there is some $\varepsilon > 0$ that provides a uniform lower bound on the probability of all possible half-spaces defined by hyperplanes through the mean.

Lemma 11.1 *Let F_0 be a distribution on $\mathbb{R}^p$ with mean μ_0 and finite variance matrix V_0 of full rank p. Let Θ be the set of unit vectors in $\mathbb{R}^p$. Then for $X \sim F_0$*

$$\inf_{\theta \in \Theta} \Pr\left((X - \mu_0)' \theta > 0\right) > 0.$$

Proof. Without loss of generality take $\mu_0 = 0$. Suppose that the infimum above is 0. Then there exists a sequence θ_n such that $\Pr(X'\theta_n > 0) < 1/n$. By compactness of Θ there is a convergent subsequence $\theta_n^* \to \theta^*$. Let $H = \{X \mid X'\theta^* > 0\}$. Then $1_{X'\theta_n^* > 0} \to 1_{X'\theta^* > 0}$ holds at any $X \in H$. Now by Lebesgue's dominated convergence theorem

$$\begin{aligned}
\Pr\left(X'\theta^* > 0\right) &= \int_H 1_{X'\theta^* > 0} dF_0\left(X\right) \\
&= \lim_{n \to \infty} \int_H 1_{X'\theta_n^* > 0} dF_0\left(X\right) \\
&\le \lim_{n \to \infty} \Pr\left(X'\theta_n^* > 0\right) \\
&= 0.
\end{aligned}$$

Since $X'\theta^*$ has mean zero we must also have $\Pr(X'\theta^* < 0) = 0$ from which

$X'\theta^* = 0$ with probability 1. Then $\mathrm{Var}(X'\theta^*) = 0$ contradicting the assumption on V_0. $\square$

The next lemma shows that for random variables with a finite variance, the largest value in a sample of size n cannot grow to infinity as fast as $n^{1/2}$.

Lemma 11.2 *Let Y_i be independent random variables with a common distribution and $E(Y_i^2) < \infty$. Let $Z_n = \max_{1\le i\le n} |Y_i|$. Then $Z_n = o(n^{1/2})$.*

Proof. Since $E(Y_i^2) < \infty$, we have $\sum_{i=1}^n \Pr(Y_i^2 > n) < \infty$. Therefore, by the Borel-Cantelli lemma there is probability 1 that $|Y_n| > n^{1/2}$ happens only for finitely many n. This implies that there are only finitely many n for which $Z_n > n^{1/2}$. A similar argument shows that for any $A > 0$, there are only finitely many n for which $Z_n > An^{1/2}$, and hence

$$\limsup_{n\to\infty} Z_n n^{-1/2} \le A$$

holds with probability 1. The probability 1 applies simultaneously over any countable set of values for A so $Z_n = o(n^{1/2})$. $\square$

Similarly, a finite variance bounds how fast a sample third moment can diverge to infinity.

Lemma 11.3 *Let Y_i be independent random variables with common distribution and suppose that $E(Y_i^2) < \infty$. Then*

$$\frac{1}{n}\sum_{i=1}^n |Y_i|^3 = o(n^{1/2}).$$

Proof. Write

$$\frac{1}{n}\sum_{i=1}^n |Y_i|^3 \le \frac{Z_n}{n}\sum_{i=1}^n Y_i^2$$

where $Z_n = \max_{1\le i\le n} |Y_i|$. The result follows by Lemma 11.2 applied to Z_n and the strong law of large numbers applied to the average of Y_i^2. $\square$

The next lemma shows how quickly the probability that the maximum observation exceeds $n^{1/2}$ decreases to zero with increasing n.

Lemma 11.4 *Let Y_i be independent random variables with common distribution and suppose that $E(|Y_i|^3) < \infty$. Define $Z_n = \max_{1\le i\le n} |Y_i|$ and let $A > 0$. Then*

$$\Pr(Z_n > An^{1/2}) = O(n^{-1/2}).$$

Proof.

$$\begin{aligned} n^{1/2}\Pr\left(Z_n > An^{1/2}\right) &\le n^{3/2}\Pr\left(|Y_1| > An^{1/2}\right) \\ &\le n^{3/2}E\left(|Y_1|^3\right)/(An^{1/2})^3 \\ &= A^{-3}E(|Y_1|^3) \\ &< \infty. \end{aligned}$$

□

11.2 Univariate and Vector ELT

The univariate ELT, Theorem 2.2, is a special case of the vector ELT, Theorem 3.2, corresponding to dimension $d = 1$. In this section we prove Theorem 3.2 of Chapter 3.

The Lagrange multiplier λ plays a key role in the proof. The proof goes in stages. First we show that $\lambda = O_p(n^{-1/2})$. Then, knowing $\lambda = O_p(n^{-1/2})$, we show that $\lambda = S^{-1}(\bar{X} - \mu_0) + o_p(n^{-1/2})$, for a certain sample covariance matrix S. Plugging this expression for λ into the profile empirical log likelihood ratio statistic, applying a central limit theorem, and verifying that some other terms are negligible completes the proof.

Proof. [Proof of Theorem 3.2, Vector ELT] Without loss of generality $q = p$. Otherwise we can replace X_i by a subset of q components having a variance of full rank. Convexity of $C_{r,n}$ follows easily by the same argument used in Chapter 2 to show that confidence regions for the univariate mean are intervals.

Let Θ denote the set of unit vectors in $\mathbb{R}^p$. By Lemma 11.1

$$\inf_{\theta\in\Theta} F_0\left(\{\theta'(X-\mu_0) > 0\}\right) > 0.$$

By a version of the Glivenko-Cantelli theorem for uniform convergence over half spaces

$$\sup_{\theta\in\Theta} |F_0(\{\theta'(X-\mu_0) > 0\}) - F_n(\{\theta'(X-\mu_0) > 0\})| \to 0$$

with probability 1 as $n \to \infty$. It follows that with probability tending to 1 that the mean μ_0 is inside the convex hull of X_i.

When the mean is inside the convex hull of the X_i, then there is a unique set of weights $w_i > 0$ with $\sum_{i=1}^n w_i = 1$ and $\sum_{i=1}^n w_i(X_i - \mu_0) = 0$ for which $\prod_{i=1}^n nw_i$ is maximized. By the arguments in Chapter 3 these maximizing weights may be written

$$w_i = \frac{1}{n}\frac{1}{1 + \lambda'(X_i - \mu_0)},$$

where the vector $\lambda = \lambda(\mu_0) \in \mathbb{R}^p$ satisfies p equations given by

$$g(\lambda) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{X_i - \mu_0}{1 + \lambda'(X_i - \mu_0)} = 0. \tag{11.1}$$

The next step is to bound the magnitude of λ. Let $\lambda = \|\lambda\|\theta$ where $\theta \in \Theta$ is a unit vector. Introduce

$$Y_i = \lambda'(X_i - \mu_0), \quad \text{and} \quad Z_n^\star = \max_{1 \le i \le n} \|X_i - \mu_0\|.$$

Substituting $1/(1 + Y_i) = 1 - Y_i/(1 + Y_i)$ into $\theta' g(\lambda) = 0$ and simplifying, we find that

$$\|\lambda\|\, \theta' \widetilde{S} \theta = \theta'(\bar{X} - \mu_0) \tag{11.2}$$

where

$$\widetilde{S} = \frac{1}{n}\sum_{i=1}^{n} \frac{(X_i - \mu_0)(X_i - \mu_0)'}{1 + Y_i}. \tag{11.3}$$

Let

$$S = \frac{1}{n}\sum_{i=1}^{n} (X_i - \mu_0)(X_i - \mu_0)'.$$

Every $w_i > 0$, so $1 + Y_i > 0$ and therefore

$$\begin{aligned}
\|\lambda\|\, \theta' S \theta &\le \|\lambda\|\, \theta' \widetilde{S} \theta (1 + \max_i Y_i) \\
&\le \|\lambda\|\, \theta' \widetilde{S} \theta (1 + \|\lambda\| Z_n^\star) \\
&= \theta'(\bar{X} - \mu_0)(1 + \|\lambda\| Z_n^\star),
\end{aligned}$$

by (11.2) and so

$$\|\lambda\| \left(\theta' S \theta - Z_n^\star \theta'(\bar{X} - \mu_0) \right) \le \theta'(\bar{X} - \mu_0).$$

Now $\sigma_1 + o_p(1) \ge \theta' S \theta \ge \sigma_p + o_p(1)$, where $\sigma_1 \ge \sigma_p > 0$ are the largest and smallest eigenvalues of $\mathrm{Var}(X_i)$. Also by Lemma 11.2, $Z_n^\star = o(n^{1/2})$. The central limit theorem applied to the vector $\bar{X} - \mu_0$ implies that $\theta'(\bar{X} - \mu_0) = O_p(n^{-1/2})$. It follows that

$$\|\lambda\| \left(\theta' S \theta + o_p(1) \right) = O_p(n^{-1/2}),$$

and hence

$$\|\lambda\| = O_p(n^{-1/2}).$$

Having established an order bound for $\|\lambda\|$, we have from Lemma 11.2 that

$$\max_{1 \le i \le n} |Y_i| = O_p(n^{-1/2}) o(n^{1/2}) = o_p(1). \tag{11.4}$$

Now

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)\left(1-Y_i+Y_i^2/(1-Y_i)\right) \\
&= \bar{X}-\mu_0-S\lambda+\frac{1}{n}\sum_{i=1}^{n}\frac{(X_i-\mu_0)Y_i^2}{1-Y_i}. \qquad (11.5)
\end{aligned}
$$

The final term in (11.5) above has a norm bounded by

$$
\frac{1}{n}\sum_{i=1}^{n}\|X_i-\mu_0\|^3\,\|\lambda\|^2\,|1-Y_i|^{-1}=o(n^{1/2})O_p(n^{-1})O_p(1)=o_p(n^{-1/2}),
$$

using Lemma 11.1 from which we find

$$
\lambda = S^{-1}(\bar{X}-\mu_0)+\beta,
$$

where

$$
\beta = o_p(n^{-1/2}).
$$

By (11.4), we may write

$$
\log(1+Y_i)=Y_i-\frac{1}{2}Y_i^2+\eta_i,
$$

where for some finite $B>0$

$$
\Pr(|\eta_i|\le B|Y_i|^3,\quad 1\le i\le n)\to 1,
$$

as $n\to\infty$.

Now we may write

$$
\begin{aligned}
-2\log\mathcal{R}\,(\mu_0) &= -2\sum_{i=1}^{n}\log(nw_i) \\
&= 2\sum_{i=1}^{n}\log(1+Y_i) \\
&= 2\sum_{i=1}^{n}Y_i-\sum_{i=1}^{n}Y_i^2+2\sum_{i=1}^{n}\eta_i \\
&= 2n\lambda'(\bar{X}-\mu_0)-n\lambda'S\lambda+2\sum_{i=1}^{n}\eta_i \\
&= n(\bar{X}-\mu_0)S^{-1}(\bar{X}-\mu_0)-n\beta'S^{-1}\beta+2\sum_{i=1}^{n}\eta_i.
\end{aligned}
$$

In the limit as $n\to\infty$,

$$
n(\bar{X}-\mu_0)S^{-1}(\bar{X}-\mu_0)\to\chi^2_{(p)}
$$

in distribution,

$$n\beta' S^{-1}\beta = n o_p(n^{-1/2}) O_p(1) o_p(n^{-1/2}) = o_p(1),$$

and

$$\left|\sum_{i=1}^{n} \eta_i\right| \le B\,\|\lambda\|^3 \sum_{i=1}^{n} \|X_i - \mu_0\|^2 = O_p(n^{-3/2}) o_p(n^{3/2}) = o_p(1).$$

Therefore $-2\log \mathcal{R}(\mu_0) \to \chi^2_{(p)}$ in distribution. □

11.3 Triangular array ELT

In this section we prove Theorem 4.1 of Chapter 4.3. The proof follows the same lines as the proof of the Vector ELT in Chapter 11.2.

Proof. [Proof of Theorem 4.1, Triangular Array ELT] Without loss of generality we may take $\mu_n = 0$ and $\sigma_{1n} = 1$, and then seek the asymptotic distribution of $-2\log \mathcal{R}(0)$. This is equivalent to reformulating the problem with $\sigma_{1n}^{-1/2}(Z_{in} - \mu_n)$ in place of Z_{in}. To simplify the notation, we drop the second subscript n, using Z_i for Z_{in}. Let

$$\hat{V}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'.$$

By assumption (4.4) we may assume that the convex hull of Z_i contains the origin. It then follows by Lagrange multiplier arguments that

$$\mathcal{R}(0) = \prod_{i=1}^{n} \frac{1}{1+\lambda' Z_i}$$

where $\lambda = \lambda(0)$ is uniquely determined by

$$\sum_{i=1}^{n} \frac{Z_i}{1+\lambda' Z_i} = 0.$$

Write $\lambda = \|\lambda\|\theta$ where $\theta'\theta = 1$. Let

$$Y_i = \lambda' Z_i, \quad \text{and} \quad Z^\star = \max_{1\le i\le n} \|Z_i\|.$$

By an argument used in the proof of the Vector ELT, we obtain

$$\|\lambda\| \left(\theta' \hat{V}\theta - Z_n^\star \theta' \bar{Z}\right) \le \theta' \bar{Z}.$$

From assumption (4.5) the variance of each entry in $\hat{V}_n$ tends to zero and so by Chebychev's inequality $\hat{V}_n - V_n = o_p(1)$. Then by assumption (4.6) on V_n, we obtain

$$c + o_p(1) \le \theta' \hat{V}\theta \le 1 + o_p(1),$$

where $c > 0$ is the constant in that assumption.

Assumption (4.5) also implies that $Z_n^\star = o(n^{1/2})$, which in turn implies Lindeberg's condition for Z_i. The central limit theorem applied to the vector $\bar{Z}$ implies that $\theta'\bar{Z} = O_p(n^{-1/2})$. It follows that

$$\|\lambda\| \left(\theta'\hat{V}\theta - o_p(1)\right) = O_p(n^{-1/2}),$$

and hence $|\lambda| = O_p(n^{-1/2})$.

The rest of the proof follows the same argument as used for the Vector ELT, and so $-2\log\mathcal{R}(0) \to \chi^2_{(p)}$ in distribution. $\square$

11.4 Multi-sample ELT

This section presents a nonrigorous argument that empirical likelihood inferences have a χ^2 calibration in multi-sample settings. For simplicity we consider the case of two samples and a single estimating equation for a scalar parameter. A rigorous argument would have to impose explicit moment conditions in order to bound the Lagrange multiplier, along the lines of the proof in Chapter 11.2 of the vector ELT.

Let $X_1, \dots, X_n \in \mathbb{R}^p \sim F_0$ and $Y_1, \dots, Y_m \in \mathbb{R}^q \sim G_0$, with all observations independent. Let $\theta \in \mathbb{R}^t$ be defined by $E(h(X,Y,\theta)) = 0$, where $h(x,y,\theta) \in \mathbb{R}^t$. For example when $p = q = t$, $h(X,Y,\theta) = X - Y - \theta$ defines θ as the difference in means $E(X) - E(Y)$. When $p = q = t = 1$, the function $h(X,Y,\theta) = 1_{X>Y} - \theta$ defines θ as the probability that X is larger than Y. We might be interested in testing $\theta = 0$ in the first case and $\theta = 1/2$ in the second case.

The ANOVA setting of Chapter 4.4 applies to group means and functions of group means. The setting here is more general in that the expectation in the estimating equation is with respect to both distributions jointly.

The argument below provides a $\chi^2_{(t)}$ limit, under some mild assumptions. To simplify the presentation $t = 1$ is used, but the argument goes through for general t with natural modifications. We assume that $\min(n,m) \to \infty$, and that $0 < E(h(X,Y,\theta_0)^2) < \infty$. We also rule out cases such as those where $h(X,Y,\theta_0)$ can be written in the form $\phi(X)\eta(Y)$ with $E(\phi(X)) = E(\eta(Y)) = 0$. In such a case, independence of X and Y implies that $E(h(X,Y,\theta_0)) = 0$, and there is no need to infer it from data. The extra assumption we will make is that either $E(E(h(X,Y,\theta_0)|X)^2) > 0$ or $E(E(h(X,Y,\theta_0)|Y)^2) > 0$.

The empirical likelihood ratio may be written

$$R(F,G) = \prod_{i=1}^{n} nu_i \prod_{j=1}^{m} mv_j$$

where F puts weight u_i on X_i and G puts weight v_j on Y_j. We will assume that $u_i \geq 0$, $\sum_{i=1}^n u_i = 1$, $v_j \geq 0$, $\sum_{j=1}^m v_j = 1$, so that the empirical likelihood is a product of two multinomials, one on each sample. The profile empirical likelihood ratio function is

$$\mathcal{R}(\theta) = \max\left\{ R\left(F, G\right) \mid \sum_{i=1}^{n}\sum_{j=1}^{m} u_i v_j H_{ij}\left(\theta\right) = 0 \right\}$$

where $H_{ij}(\theta) = H_{ij} = h(X_i, Y_j, \theta)$, and the simplex constraints on u_i and v_j have been left out to shorten the expression.

Using Lagrange multipliers we find that

$$u_i = \frac{1}{n + \lambda \sum_{r=1}^{m} v_r H_{ir}}$$
$$v_j = \frac{1}{m + \lambda \sum_{s=1}^{n} u_s H_{sj}},$$

where λ is defined by $\sum_{i=1}^{n}\sum_{j=1}^{m} u_i v_j H_{ij} = 0$. Introduce the terms

$$\bar{H}_{i\bullet} = \frac{1}{m}\sum_{j=1}^{m} H_{ij}, \quad \widetilde{H}_{i\bullet} = \sum_{r=1}^{m} v_r H_{rj}$$
$$\bar{H}_{\bullet j} = \frac{1}{n}\sum_{i=1}^{n} H_{ij}, \quad \widetilde{H}_{\bullet j} = \sum_{s=1}^{n} u_s H_{sj},$$

and $\bar{H}_{\bullet\bullet} = (nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}$. Then

$$u_i = \frac{1}{n}\left[1 - \left(\frac{\lambda}{n}\widetilde{H}_{i\bullet}\right) + \left(\frac{\lambda}{n}\widetilde{H}_{i\bullet}\right)^2 - \left(\frac{\lambda}{n}\widetilde{H}_{i\bullet}\right)^3 + \cdots\right], \text{ and}$$
$$v_j = \frac{1}{m}\left[1 - \left(\frac{\lambda}{m}\widetilde{H}_{\bullet j}\right) + \left(\frac{\lambda}{m}\widetilde{H}_{\bullet j}\right)^2 - \left(\frac{\lambda}{m}\widetilde{H}_{\bullet j}\right)^3 + \cdots\right].$$

Substituting these values into $\sum_{i=1}^{n}\sum_{j=1}^{m} u_i v_j H_{ij} = 0$, we get

$$0 = \bar{H}_{\bullet\bullet} - \lambda\left[\frac{1}{n^2 m}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\widetilde{H}_{i\bullet} + \frac{1}{nm^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\widetilde{H}_{\bullet j}\right]$$
$$+ \lambda^2\left[\frac{1}{n^3 m}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\widetilde{H}_{i\bullet}^2 + \frac{1}{nm^3}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\widetilde{H}_{\bullet j}^2 + \right.$$
$$\left. \frac{1}{n^2 m^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\widetilde{H}_{i\bullet}\widetilde{H}_{\bullet j}\right] + \cdots.$$

Ignoring higher terms in λ, we find $\lambda \doteq D^{-1}\bar{H}_{\bullet\bullet}$ where

$$D = \frac{1}{n^2m^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\sum_{r=1}^{m} H_{ir} + \frac{1}{n^2m^2}\sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}\sum_{s=1}^{n} H_{sj}.$$

In finding this D, the term

$$\widetilde{H}_{i\bullet} = \bar{H}_{i\bullet} - \frac{\lambda}{m^2}\sum_{r=1}^{m} H_{ir}\widetilde{H}_{\bullet r}$$

has been replaced by $\bar{H}_{i\bullet}$, and $\widetilde{H}_{j\bullet}$ has been replaced by $\bar{H}_{j\bullet}$, with the differences being absorbed into the coefficient of λ^2.

Now keeping terms up to order λ^2 in the profile empirical log likelihood function, we find

$$\begin{aligned} -2\log \mathcal{R}\left(\theta_0\right) &= 2\sum_{i=1}^{n} \log\left(1 + \frac{\lambda}{n}\widetilde{H}_{i\bullet}\right) + 2\sum_{j=1}^{m} \log\left(1 + \frac{\lambda}{m}\widetilde{H}_{\bullet j}\right) \\ &\doteq 2\sum_{i=1}^{n}\left(\frac{\lambda}{n}\widetilde{H}_{i\bullet} - \frac{1}{2}\left(\frac{\lambda}{n}\bar{H}_{i\bullet}\right)^2\right) \\ &\quad + 2\sum_{j=1}^{m}\left(\frac{\lambda}{m}\widetilde{H}_{\bullet j} - \frac{1}{2}\left(\frac{\lambda}{m}\bar{H}_{\bullet j}\right)^2\right). \end{aligned}$$

Replacing $\widetilde{H}$'s by corresponding $\bar{H}$'s and keeping terms to order λ^2, we get

$$\begin{aligned} -2\log \mathcal{R}\left(\theta_0\right) &\doteq 2\sum_{i=1}^{n}\frac{\lambda}{n}\bar{H}_{i\bullet} - \frac{2\lambda^2}{m^2}\sum_{r=1}^{m}\bar{H}_{\bullet r}^2 - \sum_{i=1}^{n}\left(\frac{\lambda}{n}\bar{H}_{i\bullet}\right)^2 \\ &\quad + 2\sum_{j=1}^{m}\frac{\lambda}{m}\bar{H}_{\bullet j} - \frac{2\lambda^2}{n^2}\sum_{s=1}^{n}\bar{H}_{s\bullet}^2 - \sum_{j=1}^{m}\left(\frac{\lambda}{m}\bar{H}_{\bullet j}\right)^2 \\ &= 4\lambda\bar{H}_{\bullet\bullet} - 3\lambda^2\left(\frac{1}{n^2}\sum_{i=1}^{n}\bar{H}_{i\bullet}^2 + \frac{1}{m^2}\sum_{j=1}^{m}\bar{H}_{\bullet j}^2\right) \\ &\doteq \bar{H}_{\bullet\bullet}^2\left(4D^{-1} - 3KD^{-2}\right), \end{aligned}$$

where

$$K = \frac{1}{n^2}\sum_{i=1}^{n}\bar{H}_{i\bullet}^2 + \frac{1}{m^2}\sum_{j=1}^{m}\bar{H}_{\bullet j}^2.$$

To complete the argument, we need to show that $\bar{H}_{\bullet\bullet}$, suitably scaled, is asymptotically normally distributed and that the coefficient $4D^{-1} - 3KD^{-2}$of $\bar{H}_{\bullet\bullet}^2$ is a consistent estimator of the suitably scaled variance. To make this argument, we introduce an ANOVA decomposition $h(X, Y, \theta_0) = A(X) + B(Y) + C(X, Y)$. Here $A(X) = E(h(X, Y, \theta_0) \mid X)$, $B(Y)$ is defined similarly and C is found by subtraction. We have $E(C(X,Y) \mid X) = E(C(X,Y) \mid Y) = 0$, and $E(A(X)) = E(B(X)) = 0$. Let $\sigma_A^2 = E(A(X)^2)$, $\sigma_B^2 = E(B(X)^2)$, and $\sigma_C^2 = E(C(X,Y)^2)$. These components are uncorrelated: $E(AB) = E(AC) = E(BC) = 0$.

Now

$$\bar{H}_{\bullet\bullet} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} A_i + B_j + C_{ij}$$

has mean 0 and variance

$$\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m} + \frac{\sigma_C^2}{nm}.$$

The expected value of D is

$$E\left(D\right) = \frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m},$$

and the expected value of K is

$$E\left(K\right) = \frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m} + \frac{\sigma_A^2 + \sigma_B^2 + 2\sigma_C^2}{mn}.$$

Under mild moment conditions D and K approach their expectations with small relative errors, as $m, n \to \infty$, and so

$$\begin{aligned}
&\left(4D^{-1} - 3KD^{-2}\right) V\left(\bar{H}_{\bullet\bullet}\right) \\
&\doteq \left(\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m} + \frac{\sigma_C^2}{nm}\right)\left(\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m} - 3\frac{\sigma_A^2 + \sigma_B^2 + 2\sigma_C^2}{mn}\right)\left(\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}\right)^{-2} \\
&\to 1,
\end{aligned}$$

as $\min(n,m) \to \infty$. This final limit also relies on the condition that at least one of σ_A^2 and σ_B^2 is positive. If $\sigma_A^2 = \sigma_B^2 = 0$, then $E(h(X,Y,\theta_0)) = 0$ follows, as remarked above, from the independence of X and Y.

From this it follows that empirical likelihood on two independent samples has an asymptotic $\chi^2_{(1)}$ distribution, for inferences on θ defined by $E(h(X,Y,\theta)) = 0$. We needed to have $\min(n,m) \to \infty$, but there was no need to have them grow at the same rate. Also, the second moment $E(h(X,Y,\theta)^2)$ must be finite to allow the ANOVA decomposition, and positive to get a central limit theorem for $\bar{H}_{\bullet\bullet}$.

11.5 Bibliographic notes

The proof of the vector ELT is taken from Owen (1990*b*). The triangular array ELT is from Owen (1991). A univariate ELT was proved in Owen (1988*b*) for the mean, as well as for some M-estimates and statistics defined as (Fréchet) differentiable functions of F.

The basic strategy of forcing λ to be small and then making Taylor expansions is also used by Jing Qin in his dissertation (Qin 1992) and in subsequent papers.

For smooth estimating equations Qin (1992) assumes that $\|m(x,\theta)\|^3 \le M(x)$ uniformly in a neighborhood of θ_0 with $E(M(X)) < \infty$. Along with a few other conditions, having $-\log \mathcal{R}(\theta)$ smaller than a certain multiple of $n^{1/3}$ forces $\|\theta - \theta_0\| \le n^{-1/3}$. Then Taylor expansions of the log likelihood can be made in (λ, θ) around $(0, \theta_0)$ jointly.