Stat 5102 Notes: More on Asymptotic Distributions for Moments of the Empirical Distribution

Charles J. Geyer

February 9, 2015

The purpose of this handout is to prove the following.

Theorem 1. *Suppose we have independent and identically distributed data from a distribution having moments of order $2k$, and suppose $\overline{X}_n$ denotes the mean of the empirical distribution, $M_{j,n}$ denotes the j-th central moment of the empirical distribution, μ denotes the mean of the distribution of the data, and μ_j denotes the j-th central moment of the distribution of the data. Then*

$$\sqrt{n} \begin{pmatrix} \overline{X}_n - \mu \\ M_{n,2} - \mu_2 \\ \vdots \\ M_{n,k} - \mu_k \end{pmatrix} \xrightarrow{\mathcal{D}} \mathbf{Y} \tag{1}$$

where $\mathbf{Y}$ is multivariate normal having mean vector zero and variance matrix given by

$$\operatorname{var}(Y_1) = \mu_2 \tag{2a}$$

$$\operatorname{cov}(Y_1, Y_j) = \mu_{j+1} - j\mu_{j-1}\mu_2 \tag{2b}$$

and for $j > 1$ and $k > 1$

$$\operatorname{cov}(Y_j, Y_l) = \mu_{j+l} - \mu_j\mu_l - l\mu_{l-1}\mu_{j+1} - j\mu_{j-1}\mu_{l+1} + jl\mu_{j-1}\mu_{l-1}\mu_2 \tag{2c}$$

Corollary 2. *Under the assumptions of the theorem with $k = 2$*

$$\sqrt{n} \begin{pmatrix} \overline{X}_n - \mu \\ V_n - \sigma^2 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathbf{Y}$$

where

$$\operatorname{var} \mathbf{Y} = \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \mu_2^2 \end{pmatrix}$$

The course slides (deck 1, slide 101) only prove the corollary, not the theorem, although they do prove the univariate convergence in distribution result for one component of the vector in the theorem. The asymptotic variance obtained (slide 99) agrees with the $j = l$ case of (2c)

$$\operatorname{var}(Y_j) = \mu_{2j} - \mu_j^2 - 2j\mu_{j-1}\mu_{j+1} + j^2\mu_{j-1}^2\mu_2.$$

The rest of the assertions of the theorem do not appear on the slides, although most of the work involved in proving the theorem is done on the slides.

First we remark that there is something funny about the first components of the vectors $\overline{\mathbf{Z}}_n^*$ (defined on slide 87) and $\overline{\mathbf{Z}}_n$ (defined on slide 93). The first component of $\overline{\mathbf{Z}}_n^*$ is

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) = \overline{X}_n - \mu = A_{1,n} - \alpha_1$$

whereas the first component of $\overline{\mathbf{Z}}_n$ is zero because every first central moment is zero.

Thus $\overline{\mathbf{Z}}_n$ as defined on the slides does not tell us anything about first moments. So we modify our definition of the first component of $\overline{\mathbf{Z}}_n$ (on slide 93) changing the first component like so

$$\overline{\mathbf{Z}}_n = \begin{pmatrix} \overline{X}_n \\ M_{2,n} \\ \vdots \\ M_{k,n} \end{pmatrix} \tag{3}$$

thus preserving information about first moments. So the problem we set out to solve in this document is: What is the asymptotic distribution of the random vector (3)? This is the problem the theorem solves.

We already know from slide 92 that

$$\overline{\mathbf{Z}}_n \xrightarrow{P} \begin{pmatrix} \mu \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix} \tag{4}$$

So we want to look at the asymptotic distribution of the left-hand side of (1). The displayed equation at the top of slide 98, which is

$$\sqrt{n}(M_{j,n} - \mu_j) = \sqrt{n}(M_{j,n}^* - \mu_j) - j\sqrt{n}(\overline{X}_n - \mu)M_{j-1,n}^* + o_p(1), \tag{5}$$

hold jointly for $j = 1, \ldots, k$ as well as for a single j, because there is no difference between joint convergence in probability to a constant and convergence in probability of the separate components to the corresponding constants (Stat 5101, deck 7, slide 78). Thus we can derive the asymptotic distribution of $\overline{\mathbf{Z}}_n$ as redefined in our equation (3) from the asymptotic distribution of $\overline{\mathbf{Z}}_n^*$, which is given on slides 88–89.

Introduce the notation

$$\sqrt{n}(\overline{\mathbf{Z}}_n^* - \boldsymbol{\mu}_{\text{central}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{M}_{\text{central}}) \tag{6}$$

where

$$\boldsymbol{\mu}_{\text{central}} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}$$

$$\mathbf{M}_{\text{central}} = \begin{pmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 & \cdots & \mu_{k+1} - \mu_1\mu_k \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 & \cdots & \mu_{k+2} - \mu_2\mu_k \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k+1} - \mu_1\mu_k & \mu_{k+2} - \mu_2\mu_k & \cdots & \mu_{2k} - \mu_k^2 \end{pmatrix}$$

(these come from slides 88-89).

Now, again changing notation from what was on the slides, let $\mathbf{U}$ denote a random vector having the limiting distribution in (6), so we can rewrite that as

$$\sqrt{n}(\overline{\mathbf{Z}}_n^* - \boldsymbol{\mu}_{\text{central}}) \xrightarrow{\mathcal{D}} \mathbf{U} \tag{7}$$

So far this is just a recap of the slides, except that we have decided that we want to define $\overline{\mathbf{Z}}_n$ differently from what it is on the slides, our new definition being (3), and except for our introduction of $\mathbf{U}$ for the random vector that is the limit in (7). Now for something not on the slides.

Putting together (4) and (7) using the multivariable Slutsky theorem (Stat 5101, deck 7, slides 87–90) we obtain

$$\begin{pmatrix} \sqrt{n}(\overline{X}_n - \mu) \\ \sqrt{n}(M_{n,2}^* - \mu_2) \\ \vdots \\ \sqrt{n}(M_{n,k}^* - \mu_k) \\ M_{n,1}^* \\ M_{n,2}^* \\ \vdots \\ M_{n,k}^* \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_k \\ 0 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}$$

then this plus the multivariable continuous mapping theorem (Stat 5101, deck 7, slide 86) plus the multivariable Slutsky theorem plus (5) give

$$\begin{pmatrix} \sqrt{n}(\overline{X}_n - \mu) \\ \sqrt{n}(M_{n,2} - \mu_2) \\ \vdots \\ \sqrt{n}(M_{n,k} - \mu_k) \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} U_1 \\ U_2 - 2U_1\mu_1 \\ \vdots \\ U_k - kU_1\mu_{k-1} \end{pmatrix} \tag{8}$$

This is the same conclusion we reached on slide 98 (of 5102 deck 7) except that now we have been careful and obtained the asymptotic joint distribution not just the marginal for one component.

So now we have to figure out the variance matrix of the limiting random vector in (8). Unlike on the slides, we don't use matrices because the matrices would be too big to fit on the page. Since the random vector in question, the right-hand side of (8) has two kinds of components U_1 and $U_j - jU_1\mu_{j-1}$ for j between 2 and k, we need to figure out three things. The first is obvious. We have $\operatorname{var}(U_1) = \mu_2$ because U_1 is the asymptotic distribution of $\sqrt{n}(\overline{X}_n - \mu)$. This verifies (2a). Then we have

$$\begin{aligned} \operatorname{cov}(U_1, U_j - jU_1\mu_{j-1}) &= \operatorname{cov}(U_1, U_j) - j\mu_{j-1}\operatorname{var}(U_1) \\ &= \mu_{j+1} - \mu_1\mu_j - j\mu_{j-1}\mu_2 \\ &= \mu_{j+1} - j\mu_{j-1}\mu_2 \end{aligned}$$

This verifies (2b). Finally we have

$$\begin{aligned} \operatorname{cov}(U_j - jU_1\mu_{j-1}, U_k - kU_1\mu_{k-1}) &= \operatorname{cov}(U_j, U_k) - k\mu_{k-1}\operatorname{cov}(U_j, U_1) \\ &\qquad - j\mu_{j-1}\operatorname{cov}(U_1, U_k) \\ &\qquad + jk\mu_{j-1}\mu_{k-1}\operatorname{var}(U_1) \\ &= \mu_{j+k} - \mu_j\mu_k - k\mu_{k-1}\mu_{j+1} \\ &\qquad - j\mu_{j-1}\mu_{k+1} + jk\mu_{j-1}\mu_{k-1}\mu_2 \end{aligned}$$

This verifies (2c) and finishes the proof of the theorem.