Lecture III. A NORMAL APPROXIMATION THEOREM

The basic Lemmas I.3 and I.4 are elaborated to obtain a normal approximation theorem. This is applied to the special case of a sum of independent random variables, the study of which was begun in the first lecture, and also to the study of the sum of a random diagonal. The substantial shortcomings of these results, which will be described later in this lecture, will be overcome to some extent in later lectures.

Lemma 1: Let (W,W') be an exchangeable pair of real random variables such that

$$E^W W' = (1-\lambda)W \tag{1}$$

with

$$0 < \lambda < 1 \tag{2}$$

and let h: $R \to R$ be a bounded continuous function with bounded piecewise continuous derivative h'. Then, with $U_N h$ defined by (II.4), that is

$$(U_N h)(w) = e^{\frac{1}{2}w^2}\int_{-\infty}^{w} [h(x)-Nh]e^{-\frac{1}{2}x^2}\, dx \tag{3}$$

$$= -e^{\frac{1}{2}w^2}\int_{w}^{\infty}[h(x)-Nh]e^{-\frac{1}{2}x^2}\, dx,$$

where Nh is defined by (II.2), that is

$$Nh = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} h(x)e^{-\frac{1}{2}x^2}\, dx, \tag{4}$$

we have

$$(5) \qquad Eh(W) = Nh + E(U_N h)'(W)[1 - \frac{1}{2\lambda} E^W(W'-W)^2]$$
$$+ \frac{1}{2\lambda} \int E(W'-W)(z - \frac{W+W'}{2})[\mathcal{I}\{z \leq W'\} - \mathcal{I}\{z \leq W\}]d(U_N h)'(z).$$

Proof: We have to express the remainder in (I.45), where f was written instead of $U_N h$, as the sum of the last two terms on the right hand side of (5). In fact

$$(6) \qquad E[(U_N h)'(W) - \frac{1}{2\lambda}(W'-W)((U_N h)(W') - (U_N h)(W))]$$
$$= E(U_N h)'(W)[1 - \frac{1}{2\lambda} E^W(W'-W)^2]$$
$$- \frac{1}{2\lambda} E(W'-W)[(U_N h)(W')-(U_N h)(W)-(W'-W)(U_N h)'(W)].$$

But, for $W < W'$,

$$(7) \qquad (U_N h)(W') - (U_N h)(W)-(W'-W)(U_N h)'(W)$$
$$= \int_W^{W'} [(U_N h)'(z) - (U_N h)'(W)]dz$$
$$= \int_W^{W'} (\int_W^z (U_N h)''(y)dy)dz = \int_W^{W'} (W'-y)(U_N h)''(y)dy,$$

and, for $W' < W$,

$$(8) \qquad (U_N h)(W') - (U_N h)(W) - (W'-w)(U_N h)'(W)$$
$$= - \int_{W'}^{W} [(U_N h)'(z) - (U_N h)'(W)]dz$$
$$= \int_{W'}^{W} (\int_z^W (U_N h)''(y)dy)dz = \int_{W'}^{W} (y-W')(U_N h)''(y)dy.$$

Now (7) and (8) can be combined to obtain, for all W, W',

$$(9) \qquad (U_N h)(W')-(U_N h)(W)-(W'-W)(U_N h)'(W)$$
$$= \int(W'-y)[\mathcal{I}\{y \leq W'\} - \mathcal{I}\{y \leq W\}](U_N h)''(y)dy.$$

Taking expectation, applying the exchangeability of W and W' again, and using (6) and (I.45), we obtain (5).

Theorem 1: Under the assumptions of Lemma 1,

(10) $$|Eh(W)-Nh|$$

$$\leq 2\sup|h-Nh|\sqrt{E[1-E^W\frac{1}{2\lambda}(W'-W)^2]^2}+\frac{1}{4\lambda}\sup|h'|E|W'-W|^3$$

and, for all real w_0,

(11) $$|P\{W \leq w_0\} - \Phi(w_0)|$$

$$\leq 2\sqrt{E[1-\frac{1}{2\lambda}E^W(W'-W)^2]^2}+(2\pi)^{-\frac{1}{4}}\sqrt{\frac{1}{\lambda}E|W'-W|^3}.$$

Proof: We recall that by Lemma II.3,

(12) $$\sup|(U_Nh)'| \leq 2\sup|h-Nh|$$

and

(13) $$\sup|(U_Nh)''| \leq 2\sup|h'|.$$

Thus from Lemma 1 we obtain

(14) $$|Eh(W)-Nh| = |E(U_Nh)'(W)[1-\frac{1}{2\lambda}E^W(W'-W)^2]$$

$$+\frac{1}{2\lambda}\int E(W'-W)(z-\frac{W+W'}{2})[\mathcal{I}\{z \leq W'\}-\mathcal{I}\{z \leq W\}](U_Nh)''(z)dz|$$

$$\leq \sup|(U_Nh)'|E|1-\frac{1}{2\lambda}E^W(W'-W)^2|$$

$$+\sup|(U_Nh)''|\frac{1}{2\lambda}E\int_{W\wedge W'}^{W\vee W'}|W'-W||z-\frac{W+W'}{2}|dz$$

$$\leq 2\sup|h-Nh|\sqrt{E[1-\frac{1}{2\lambda}E^W(W'-W)^2]^2}$$

$$+2\sup|h'|\cdot\frac{1}{2\lambda}E\frac{|W'-W|^3}{4},$$

which is (10). Here I have written

(15) $$a \wedge b = \min(a,b)$$

and

(16) $$a \vee b = \max(a,b).$$

In order the prove (11), we shall apply (10) with $h = h_{w_0,\Delta}$ defined for $\Delta > 0$ by

(17)
$$h_{w_0,\Delta}(w) = \begin{cases} 1 & \text{if } w \le w_0 \\ 1 - \frac{w-w_0}{\Delta} & \text{if } w_0 \le w \le w_0 + \Delta \\ 0 & \text{if } w \ge w_0 + \Delta. \end{cases}$$

Then

(18)
$$Eh_{w_0-\Delta,\Delta}(W) \le P\{W \le w_0\} \le Eh_{w_0,\Delta}(W),$$

and

(19)
$$\sup|h_{w_0,\Delta} - Nh_{w_0,\Delta}| \le 1$$

and

(20)
$$\sup|h'_{w_0,\Delta}| \le \frac{1}{\Delta}.$$

Thus, using (10) we obtain

(21)
$$\begin{aligned} P\{W \le w_0\} &\le Eh_{w_0,\Delta}(W) \\ &\le Nh_{w_0,\Delta} + 2\sqrt{E[1 - \tfrac{1}{2\lambda} E^W(W'-W)^2]^2} + \tfrac{1}{4\lambda\Delta} E|W'-W|^3 \\ &\le \Phi(w_0) + \tfrac{\Delta}{\sqrt{2\pi}} + 2\sqrt{E[1 - \tfrac{1}{2\lambda} E^W(W'-W)^2)]^2} + \tfrac{1}{4\lambda\Delta} E|W'-W|^3. \end{aligned}$$

The upper bound on the right hand side is minimized by

(22)
$$\Delta = \frac{(2\pi)^{\frac{1}{4}}}{2} \sqrt{\frac{1}{\lambda} E|W'-W|^3}$$

yielding

(23)
$$\begin{aligned} P\{W \le w_0\} \le \Phi(w_0) &+ 2\sqrt{E[1 - \tfrac{1}{2\lambda} E^W(W'-W)^2]^2} \\ &+ (2\pi)^{-\frac{1}{4}} \sqrt{\tfrac{1}{\lambda} E|W'-W|^3}. \end{aligned}$$

The bound for the error in (11) follows from (23) and the corresponding lower bound.

Now let us return to the special case of a sum of independent random variables. We need only evaluate the expectations occurring in the bounds on the right hand side of (10) and (11) in the situation discussed at the end of Lecture I starting before (I.51). I shall state the result as

Corollary 1: Let $X_1,\ldots,X_n$ be independent real random variables with

$$EX_i = 0,\ EX_i^2 = \sigma_i^2 \tag{24}$$

and

$$\Sigma\sigma_i^2 = 1. \tag{25}$$

Then, for all real w_0,

$$|P\{W \le w_0\} - \Phi(w_0)| \le \sqrt{\Sigma(EX_i^4-\sigma_i^4)} + (2\pi)^{-\frac{1}{4}}\sqrt{2\Sigma(E|X_i|^3+3\sigma_i^3)} \tag{26}$$

and for any bounded continuous functions h: R → R with bounded piecewise continuous derivative h',

$$|Eh(W)-Nh| \le \sup|h-Nh|\ \sqrt{\Sigma(EX_i^4-\sigma_i^4)} + \frac{1}{2}\sup|h'|\Sigma(E|X_i|^3 + 3\sigma_i^3). \tag{27}$$

Proof: We shall need the fact that, if X_i^* is independent of X_i and has the same distribution as X_i then

$$E^X(X_i^*-X_i)^2 = \sigma_i^2 + X_i^2 \tag{28}$$

and

$$E|X_i^*-X_i|^3 \le 2(E|X_i|^3 + 3(EX_i^2)E|X_i|) \le 2(E|X_i|^3 + 3\sigma_i^3). \tag{29}$$

Of course (29) could be improved substantially. Then the expectations on the right hand side of (10) and (11) can be bounded by using

$$\begin{aligned} E[1 - \tfrac{n}{2}\,E^W(W'-W)^2]^2 &= E[1 - \tfrac{n}{2}\,E^W(X_I^*-X_I)^2]^2 \\ &\le E[1 - \tfrac{n}{2}\,E^X(X_I^*-X_I)^2]^2 = E[1 - \tfrac{1}{2}\,\Sigma\,E^X(X_i^*-X_i)^2]^2 \\ &= E[\tfrac{1}{2}\,\Sigma(X_i^2-\sigma_i^2)]^2 = \tfrac{1}{4}\,\Sigma(EX_i^4-\sigma_i^4) \end{aligned} \tag{30}$$

and

$$E|W'-W|^3 = E|X_I^*-X_I|^3 = \frac{1}{n}\,\Sigma\,E|X_i^*-X_i|^3 \le \frac{2}{n}\,\Sigma(E|X_i|^3 + 3\sigma_i^3). \tag{31}$$

Substituting in (10) and (11) we obtain (26) and (27).

Let us look at this corollary in the special case of a sum of independent identically distributed random variables. Let $Y_1,Y_2,\ldots$ be independent identically distributed random variables with mean 0 and variance 1 and let

$$W = \frac{1}{\sqrt{n}} \Sigma\, Y_i. \tag{32}$$

Then, by applying Corollary 1 to the

$$X_i = \frac{1}{\sqrt{n}} Y_i \tag{33}$$

we see that, with

$$\beta_3 = E|Y_i|^3 \tag{34}$$

and

$$\beta_4 = EY_i^4, \tag{35}$$

we have

$$|P\{W \le w_0\} - \Phi(w_0)| \le \sqrt{\frac{\beta_4 - 1}{n}} + (2\pi n)^{-\frac{1}{4}} \sqrt{2(\beta_3+3)} \tag{36}$$

and

$$|Eh(W)-Nh| \le \sup|h-Nh|\sqrt{\frac{\beta_4-1}{n}} + \frac{1}{2\sqrt{n}} \sup|h'|(\beta_3+3). \tag{37}$$

We see that the bound in (37) is of the right order of magnitude $n^{-\frac{1}{2}}$ provided $\beta_4 < \infty$, but the bound in (36) is too large, being of the order of $n^{-\frac{1}{4}}$. These faults will be repaired in later lectures.

Finally let us look briefly at the case of a sum of a random diagonal. Let $\{a_{ij}\}_{i,j\varepsilon\{1\ldots n\}}$ be a square array of numbers such that, for every $i \varepsilon \{1,\ldots,n\}$,

$$\sum_j a_{ij} = 0 \tag{38}$$

and, for every $j \varepsilon \{1,\ldots,n\}$,

$$\sum_i a_{ij} = 0, \tag{39}$$

and also

$$\sum_i \sum_j a_{ij}^2 = n-1. \tag{40}$$

Let Π be a random permutation of $\{1,\ldots,n\}$ such that, for every one to one

$\pi\colon\ \{1,\ldots,n\} \to \{1,\ldots,n\}$,

$$P\{\Pi=\pi\} = \frac{1}{n!}\,. \tag{41}$$

We shall be interested in the distribution of

$$W = \Sigma\, a_{i\Pi(i)}. \tag{42}$$

In order to apply Theorem 1, we introduce an ordered pair (I,J) of random variables independent of Π with

$$P\{I=i\ \&\ J=j\} = \frac{1}{n(n-1)} \tag{43}$$

for every $i,j \in \{1,\ldots,n\}$ with $i \neq j$, and define the random permutation Π' by

$$\Pi'(i) = \begin{cases} \Pi(i) & \text{if } i \notin \{I,J\} \\ \Pi(J) & \text{if } i = I \\ \Pi(I) & \text{if } i = J. \end{cases} \tag{44}$$

Finally we define

$$W' = \Sigma\, a_{i\Pi'(i)} = W+\alpha_{I,J,\Pi(I),\Pi(J)} \tag{45}$$

where

$$\alpha_{i,j,k,\ell} = a_{i\ell} + a_{jk} - a_{ik} - a_{j\ell}. \tag{46}$$

Clearly (Π,Π') is an exchangeable pair since we can first choose the pair (I,J), then the $\Pi(i)$ for $i \notin \{I,J\}$ and then, of the two remaining possibilities, choose one to be Π and the other to be Π' with equal probability. It follows that (W,W') is also an exchangeable pair.

We must verify the condition (1) which is part of the hypotheses of Theorem 1 and then bound the quantities occurring on the right hand side of (10) and (11). By (45) and (46),

$$\begin{aligned} (47)\quad E^{\Pi}W' &= W+E^{\Pi}[a_{I\Pi(J)}+a_{J\Pi(I)}-a_{I\Pi(I)}-a_{J\Pi(J)}] \\ &= W + \frac{1}{n(n-1)}\sum_i \sum_{j\neq i}[a_{i\Pi(j)}+a_{j\Pi(i)}-a_{i\Pi(i)}-a_{j\Pi(j)}] \\ &= W + \frac{1}{n(n-1)}\,[-\sum_i a_{i\Pi(i)}-\sum_j a_{j\Pi(j)}-(n-1)\sum_i a_{i\Pi(i)}-(n-1)\sum_j a_{j\Pi(j)}] \\ &= (1 - \frac{2}{n-1})W. \end{aligned}$$

Thus (1) holds with

$$\lambda = \frac{2}{n-1} \, . \tag{48}$$

We shall bound the first term on the right-hand side of (10) or (11) by using the inequality

$$E[1 - \frac{1}{2\lambda} E^{W}(W'-W)^2]^2 \leq E[1 - \frac{1}{2\lambda} E^{\Pi}(W'-W)^2]^2 \tag{49}$$

which is true because Π determines W.

In order to bound the right-hand side of (49) we first compute

$$\begin{aligned} E^{\Pi}(W'-W)^2 &= \frac{1}{n(n-1)} \sum_{i,j} (a_{i\Pi(j)}+a_{j\Pi(i)}-a_{i\Pi(i)}-a_{j\Pi(j)})^2 \\ &= \frac{2}{n(n-1)} [\sum_{i,j} a^2_{i\Pi(j)} + \sum_{i,j} a^2_{i\Pi(i)} + \sum_{i,j} (a_{i\Pi(j)}a_{j\Pi(i)}+a_{i\Pi(i)}a_{j\Pi(j)})] \\ &\quad - \frac{4}{n(n-1)} [\sum_{i,j} a_{i\Pi(j)}a_{i\Pi(i)} + \sum_{i,j} a_{i\Pi(j)}a_{j\Pi(j)}] \\ &= \frac{2}{n(n-1)} [(n-1)+(n+2)\sum_{i} a^2_{i\Pi(i)} + \sum_{i,j} (a_{i\Pi(j)}a_{j\Pi(i)}+a_{i\Pi(i)}a_{j\Pi(j)})] \end{aligned} \tag{50}$$

where the sums over the pair of indices i,j are subject to the restriction $i \neq j$. The final equality uses (38), (39), and (40). From (50) we can compute

$$E(W'-W)^2 = EE^{\Pi}(W'-W)^2 = \frac{4}{n-1} = 2\lambda \tag{51}$$

since

$$E \sum_{i} a^2_{i\Pi(i)} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{n} a^2_{ik} = \frac{n-1}{n} \tag{52}$$

and

$$\begin{aligned} E \sum_{i,j} a_{i\Pi(j)}a_{j\Pi(i)} &= E \sum_{i,j} a_{i\Pi(i)}a_{j\Pi(j)} \\ &= \frac{1}{n(n-1)} \sum_{i,j} \sum_{k,\ell} a_{ik}a_{j\ell} = -\frac{1}{n(n-1)} \sum_{i,j} \sum_{k} a_{ik}a_{jk} \\ &= \frac{1}{n(n-1)} \sum_{i,k} a^2_{ik} = \frac{1}{n}, \end{aligned} \tag{53}$$

where the sum over the pair of indices k,ℓ is also subject to the restriction $k \neq \ell$. Substituting $f(w) = w$ in (I.38) and using (51) we obtain

$$EW^2 = 1, \tag{54}$$

which indicates that the scale has been chosen in a natural way.

From (50) it is not hard to see that for some absolute constants C and C'

$$(55)\quad \operatorname{Var} E^{\pi}(W'-W)^2$$
$$\le C\left[\frac{1}{n^2}\sum_i \operatorname{Var} a^2_{i\pi(i)} + \frac{1}{n^4}\sum_{i,j}\operatorname{Var}(a_{i\pi(j)}a_{j\pi(i)} + a_{i\pi(i)}a_{j\pi(j)})\right]$$
$$\le \frac{C'}{n^3}\sum_{i,j} a^4_{ij}$$

and consequently, by (48) and (49),

$$(56)\qquad E[1 - \frac{1}{2\lambda} E^W(W'-W)^2]^2 \le \frac{C'}{n}\sum_{i,j} a^4_{ij}.$$

Similarly, for use in the final remainder term of (10) or (11) we observe that

$$(57)\qquad E|W'-W|^3$$
$$= E|a_{I\pi(J)} + a_{J\pi(I)} - a_{I\pi(I)} - a_{J\pi(J)}|^3$$
$$\le \frac{C}{n^2}\sum_{i,j}|a_{ij}|^3.$$

Thus in the present case (10) and (11) are specialized to

$$(58)\qquad |Eh(W) - Nh|$$
$$\le \frac{C'}{\sqrt{n}}\sup|h|\sqrt{\sum_{i,j} a^4_{ij}} + \frac{C'}{n}\sup|h'|\sum_{i,j}|a_{ij}|^3$$

and

$$(59)\qquad |P\{W \le w_0\} - \Phi(w_0)|$$
$$\le \frac{C'}{\sqrt{n}}\left[\sqrt{\sum_{i,j} a^4_{ij}} + \sqrt{\sum_{i,j}|a_{ij}|^3}\right].$$

In Lemma 1 an identity was derived that makes the conditions for approximate normality of the random variable W fairly clear intuitively, at least in a rough way. Provided that the scale has been chosen in such a way that

$$(60)\qquad 2\lambda = E(W'-W)^2,$$

the first remainder term is small if $\frac{1}{2\lambda} E^W(W'-W)^2$ is nearly constant, and the second remainder term is small if $|W'-W|$ is small, or more precisely if $E|W'-W|^3/E[W'-W]^2$ is small. These rough statements were made precise in Theorem 1. This was specialized in Corollary 1 to sums of independent random variables and then specialized further in (36) and (37) to the identically distributed case. Finally the case of a sum of a random diagonal was considered briefly. A special case of this is the permutation distribution of a sample covariance and, more particularly, the distribution of the mean of a sample from a finite population.