LECTURE IX. SUMS OF INDEPENDENT IDENTICALLY DISTRIBUTED RANDOM VARIABLES

Here I shall give an essentially self-contained derivation of the Berry-Esseen Theorem for sums of independently identically distributed random variables. Some of the analytic results of the second lecture will be used.

In order to establish the framework for a basic lemma, let $X_1, \ldots, X_{n+1}$ be independent random variables with common c.d.f. μ such that

$$EX_i = \int_{-\infty}^{\infty} t d\mu(t) = 0 \tag{1}$$

and

$$EX_i^2 = \int_{-\infty}^{\infty} t^2 d\mu(t) = \frac{1}{n}. \tag{2}$$

Also let

$$K(t) = n\int_t^{\infty} u d\mu(u) = -n\int_{-\infty}^{t} u d\mu(u). \tag{3}$$

Then K is a probability density function. The positivity of K follows from the first form of (3) when t is positive and from the second form when t is negative. The integral of K is easily seen to be 1:

$$\int_{-\infty}^{\infty} K(t)dt = \int_{-\infty}^{0} K(t)dt + \int_0^{\infty} K(t)dt \tag{4}$$

$$= n\left[\int_0^\infty dt\int_t^\infty u\,d\mu(u) - \int_{-\infty}^0 dt\int_{-\infty}^t u\,d\mu(u)\right]$$

$$= n\int_{-\infty}^\infty u^2 d\mu(u) = 1.$$

Let Z_n be a random variable with probability density function K, independent of $X_1,\ldots,X_n$. Also, for given bounded piecewise continuous h: $R \to R$. Let

$$\text{(5)} \qquad Nh = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty h(t)e^{-\frac{t^2}{2}}\,dt,$$

and let f: $R \to R$ be the unique bounded solution of the differential equation

$$\text{(6)} \qquad f'(w) - wf(w) = h(w) - Nh.$$

Finally let

$$\text{(7)} \qquad W = \sum_1^n X_i.$$

<u>Lemma 1</u>: Under the conditions of the preceding paragraph

$$\text{(8)} \qquad Eh(W+Z_n) = Nh + EW[f(W+X_{n+1}) - f(W+Z_n)] - EZ_nf(W+Z_n).$$

Proof: We have

$$\text{(9)} \qquad 0 = E\left(\sum_{i=1}^n X_i - nX_{n+1}\right)f\left(\sum_{j=1}^{n+1} X_j\right)$$

$$= E[Wf(W+X_{n+1}) - n\int tf(W+t)d\mu(t)]$$

$$= E[Wf(W+X_{n+1}) - f'(W+Z_n)]$$

$$= E[Wf(W+X_{n+1}) - (W+Z_n)f(W+Z_n) - h(W+Z_n) + Nh].$$

At the first equality I have used the exchangeability of the independent identically distributed random variables $X_1,\ldots,X_{n+1}$ to conclude that

(10) $$E(X_i - X_{n+1})f(\sum_1^{n+1} X_j) = 0.$$

At the second equality the fact that X_{n+1} is distributed according to μ, independent of W, implies that

(11) $$\begin{aligned} EX_{n+1}f(\sum_{j=1}^{n+1} X_j) &= EE^W X_{n+1}f(W+X_{n+1}) \\ &= E\int tf(W+t)d\mu(t). \end{aligned}$$

For the third equality we use (1) and (3) to conclude that

(12) $$\begin{aligned} &nE\int tf(W+t)d\mu(t) \\ &= nE\int t[f(W+t) - f(W)]d\mu(t) \\ &= nE[\int_0^\infty t\int_0^t f'(W+u)dud\mu(t) - \int_{-\infty}^0 t\int_t^0 f'(W+u)dud\mu(t)] \\ &= E\int f'(W+u)K(u)du = Ef'(W+Z_n). \end{aligned}$$

The final equality in (9) is obtained by substituting for $f'(W+Z_n)$ the expression given by (6). Of course (8) is an immediate consequence of (9).

<u>Theorem 1</u>: If $Y_1, Y_2, \ldots$ are independent identically distributed random variables with

(13) $$EY_i = 0,$$

(14) $$EY_i^2 = 1,$$

and

(15) $$\beta_3 = E|Y_i|^3 < \infty,$$

then, for all natural numbers n, and all $t \in R$

(16) $$|P\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i \le t\right\} - \Phi(t)| \le \frac{6\beta_3}{\sqrt{n}}.$$

Before going through the details required for a careful treatment of this result by the present method, I shall briefly indicate the simple basic idea.

we define the X_i of the lemma in terms of the Y_i of the theorem by

$$X_i = \frac{Y_i}{\sqrt{n}} \tag{17}$$

and of course take μ to be the cumulative distribution function of the X_i. Then, by computation similar to (4),

$$E|Z_n| = \int |t| K(t)dt = \frac{\beta_3}{2\sqrt{n}} . \tag{18}$$

We specialize the function h of Lemma 1 to h_t defined by

$$h_t(x) = \begin{cases} 1 & \text{if} \quad x \leq t \\ 0 & \text{if} \quad x > t . \end{cases} \tag{19}$$

It was proved in the second lecture that the corresponding bounded solution f_t of the differential equation (6) and its derivative f_t' are bounded by one in absolute value. Thus (8) implies that

$$\begin{aligned} |P\{W+Z_n \leq t\} - \Phi(t)| &= |Eh_t(W+Z_n) - Nh| \\ &\leq E|W|(|X_{n+1}| + |Z_n|) + E|Z_n| \leq \frac{\beta_3+1}{\sqrt{n}} . \end{aligned} \tag{20}$$

The final inequality follows from (18) and the fact that W_n is independent of X_{n+1} and Z_n and

$$E|X_{n+1}| \leq \sqrt{EX_{n+1}^2} = \frac{1}{\sqrt{n}} \tag{21}$$

and

$$E|W| \leq \sqrt{EW^2} = 1. \tag{22}$$

It remains only to apply an unsmoothing lemma to derive (16) from (20) and (18). Such a lemma was used by Loève (1977), p. 295 to prove the Berry-Esseen Theorem using the method of characteristic functions. I shall give a self-contained treatment of a similar lemma, partly because of the difficulty of setting up a correspondence between Loève's notation and mine and partly because it may be possible to apply the basic idea of this lemma to other problems. The present version of this lemma could easily be improved, but it seems desirable to avoid complications that are not essential for the present application.

Lemma 2: If F, G, and H are three cumulative distribution functions on the real line and A, μ real numbers such that

$$H' \leq A \tag{23}$$

$$|(F-H)*G| \leq \mu \tag{24}$$

and

$$\gamma = \int |t| dG(t) < \infty, \tag{25}$$

then

$$|F-H| \leq \mu+6\gamma A+4\sqrt{\gamma A(\mu+2\gamma A)} \leq 3\mu+12\gamma A. \tag{26}$$

(Of course H is understood to be the indefinite integral of its derivative H', and (23), (24), and (26) are to be interpreted as pointwise inequalities).

Before proving the lemma, I shall use it to complete the proof of Theorem 1. The distributions H, F, and G of the lemma are identified with Φ, the distribution of W, and the distribution of Z_n, respectively. Then (23) holds with

$$A = \frac{1}{\sqrt{2\pi}}, \tag{27}$$

and (25) holds with

$$\gamma = \frac{\beta_3}{2\sqrt{n}} \tag{28}$$

because of (18). Finally (24) holds with

$$\mu = \frac{\beta_3+1}{\sqrt{n}} \tag{29}$$

because of (20). Then Lemma 2 with

(30) $$A = \frac{1}{\sqrt{2\pi}}$$

and

(31) $$\gamma = \frac{\beta_3}{2\sqrt{n}}$$

yields

(32) $$\begin{aligned} &|P\{W \le t\} - \Phi(t)| \\ &\le \mu + 6\gamma A + 4\sqrt{\gamma A(\mu + 2\gamma A)} \\ &= \frac{\beta_3+1}{\sqrt{n}} + \frac{3\beta_3}{\sqrt{2\pi n}} + 4\sqrt{\frac{\beta_3}{2\sqrt{2\pi n}}\left(\frac{\beta_3+1}{\sqrt{n}} + \frac{\beta_3}{\sqrt{2\pi n}}\right)} \\ &< \frac{6\beta_3}{\sqrt{n}} . \end{aligned}$$

Proof of Lemma 2: Let

(33) $$\rho_0 = -\inf(F-H)$$

and

(34) $$\rho_1 = \sup(F-H).$$

The basic idea of the proof can be seen from Fig. 1 below although the details are moderately complicated. If, at some point x_0, $F(x_0) - (H(x_0)-\rho_0)$ is a large multiple of $\mu+\gamma A$, a similarly large lower bound for $(F-H)*G$ at some x in the horizontal portion of the graph can be obtained from the lower bound for F given by the solid portion of the graph and can then be compared with (24). An analogous argument in the other direction enables us to complete the proof that ρ_0 and ρ_1 cannot be too large.

For all x and x_0,

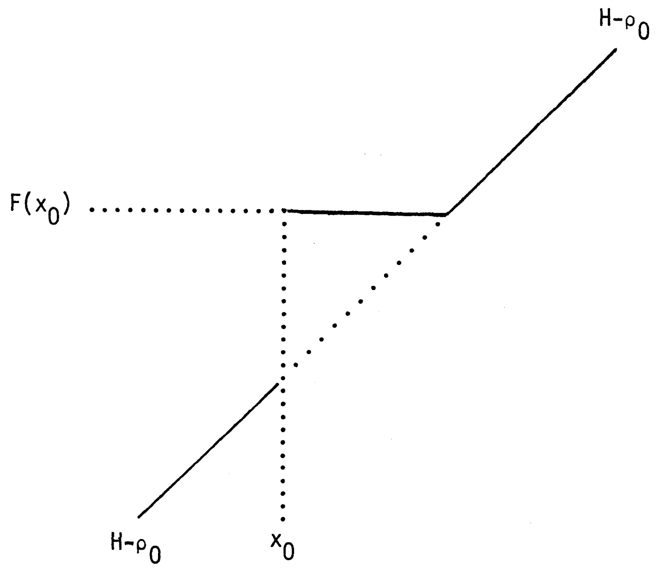

Fig. 1

$$\mu \geq \int [F(x-y) - H(x-y)]dG(y) \tag{35}$$

$$\geq -\rho_0 + \int_{-\infty}^{x-x_0} [F(x_0) - (H(x-y)-\rho_0)]_+ dG(y)$$

$$\geq \int_B^{B+D} [F(x_0) - H(x_0) + \rho_0 - A(x-x_0-y)]dG(y)-\rho_0$$

$$= A \int_B^{B+D} (y-B)dG(y) - \rho_0,$$

$$= A \int_B^{B+D} (y-B)dG(y) - \rho_0,$$

where

$$B = x - x_0 - \frac{F(x_0)-H(x_0)+\rho_0}{A} \tag{36}$$

and

$$D = \frac{F(x_0) - H(x_0) + \rho_0}{A} . \tag{37}$$

The first inequality in (35) used (24), the second used the lower bound for F in Fig. 1 and the third used (23). Now, for some $\alpha \in (0,1)$ choose

(38) $$x = x_0 + \frac{\alpha}{2} D.$$

Then (35) yields

(39) $$\mu + \rho_0 \geq A\alpha D \int_{B+\alpha D}^{B+D} dG(y) \geq A\alpha D \left(1 - \frac{2\gamma}{(1-\alpha)D}\right)$$

$$= \alpha[F(x_0) - H(x_0) + \rho] - 2\frac{\alpha}{1-\alpha}\gamma A.$$

The second inequality uses Markov's inequality and (25). Consequently

(40) $$F(x_0) - H(x_0) \leq \frac{1}{\alpha}\left[\mu + \rho_0 + 2\frac{\alpha}{1-\alpha}\gamma A\right] - \rho_0$$

$$= \frac{\mu+\rho_0}{\alpha} + \frac{2\gamma A}{1-\alpha} - \rho_0.$$

The upper bound in (40) is minimized by

(41) $$\alpha = \frac{\sqrt{\mu+\rho_0}}{\sqrt{\mu+\rho_0} + \sqrt{2\gamma A}} .$$

Since x_0 was arbitrary

(42) $$\rho_1 \leq (\sqrt{\mu+\rho_0} + \sqrt{2\gamma A})^2 - \rho_0$$

$$= \mu + 2\gamma A + 2\sqrt{2\gamma A}\,\sqrt{\mu+\rho_0} .$$

Applying the same argument with F and H replaced by 1-F and 1-H we obtain, by symmetry,

(43) $$\rho_0 \leq \mu + 2\gamma A + 2\sqrt{2\gamma A}\,\sqrt{\mu+\rho_1} .$$

Starting with the trivial bound $\rho_0 \leq 1$ it is easy to prove by induction, using (42) and (43) and the fact that the r.h.s. of (42) is an increasing function of ρ_0, that

(44) $$\rho_0, \rho_1 \leq \rho^*$$

where ρ^* is the unique solution of the equation

(45) $$\rho^* = \mu + 2\gamma A + 2\sqrt{2\gamma A}\,\sqrt{\mu+\rho^*} = Q_1 + Q_2\sqrt{\mu+\rho^*} ,$$

say, that is the larger solution of

(46)
$$(\rho^*-Q_1)^2 = Q_2^2(\mu+\rho^*),$$

which is

(47)
$$\rho^* = Q_1 + \frac{Q_2^2}{2} + \sqrt{Q_2^2(Q_1+\mu)+Q_2^4/4}$$
$$= \mu + 6\gamma A + 4\sqrt{\gamma A(\mu+2\gamma A)} \leq 3\mu + 12\gamma A.$$

This completes the proof of Lemma 2.

It would be tedious but not really difficult to improve the constant multiple 6 on the right hand side of (16) in Theorem 1. The bound in Lemma 2 could be substantially improved by imposing the condition that G has a unimodal density, which is satisfied in our case, and by using a more careful argument than (39). Also, the crude bound (20) can be improved.

I have given an essentially self-contained proof of the Berry-Esseen theorem in the special case of a sum of independent identically distributed random variables. Lemma 1 plays much the same role as Lemma III.1 did in a more general context. The completion of the proof starting from Lemma 1 is essentially a routine matter using known results. The remainder on the right hand side is of the order of $\frac{1}{\sqrt{n}}$ since, with h the indicator function of the event $\{W \leq w_0\}$, f is bounded and satisfies a Lipschitz condition with constant 1 (by Lemma II.2) and the random variables X_{n+1} and Z_n have first absolute moments of the order of $\frac{1}{\sqrt{n}}$. This leaves us with an unsmoothing problem, that of going from $Eh(W+Z_n)$ to $Eh(W)$. This task is accomplished by Lemma 2, which is essentially the same as that used by Loève in a proof by the method of characteristic functions. It may not be difficult to modify this proof to include some modern improvements on the theorem in this special case. Some such results results, proved by traditional methods, are given by Petrov (1975).