LECTURE VI. SUMS OF INDEPENDENT RANDOM VARIABLES WITH DENSITIES

An identity is derived in Lemma 2 for sums of independent random variables having a probability density function. This is similar to the appropriate specialization of Lemma I.3 in the argument leading to the simplest normal approximation theorem in Corollary III.1, but has the advantage that terms involving a difference f(W')-f(W) are replaced by terms involving a derivative f'(W). This should make it possible to derive better approximation theorems in this case. I have not had any real success with this approach but it looks promising. Some auxiliary results such as Lemmas 1 and 3 should be useful in discussing approximation by distributions other than normal. The work is also related to Pearson's family of densities. Finally I should mention that this is a limiting case of an approach to the discrete case that I hope to discuss in a separate paper.

<u>Lemma 1</u>: Let X be a real random variable distributed according to a probability density function p with

$$EX = \int_{-\infty}^{\infty} xp(x)dx = 0, \tag{1}$$

and let $\tau: R \to R$ be defined by

$$\tau(x) = \frac{\int_x^{\infty} yp(y)dy}{p(x)} = -\frac{\int_{-\infty}^{x} yp(y)dy}{p(x)}. \tag{2}$$

Then for any continuous and piecewise continuously differentiable function $f: R \to R$ for which

$$E|f'(X)|\tau(X) < \infty, \tag{3}$$

we have

(4) $$E[\tau(X)f'(X) - Xf(X)] = 0.$$

Proof:

(5) $$Ef'(X)\tau(X)$$

$$= \int_0^\infty f'(x) \frac{\int_x^\infty yp(y)dy}{p(x)} p(x)dx - \int_{-\infty}^0 f'(x) \frac{\int_{-\infty}^x yp(y)dy}{p(x)} p(x)dx$$

$$= \int_0^\infty yp(y)(\int_0^y f'(x)dx)dy - \int_{-\infty}^0 yp(y)(\int_y^0 f'(x)dx)dy$$

$$= \int_\infty^\infty y[f(y)-f(0)]p(y)dy = EXf(X).$$

Lemma 2: Let $X_1,\ldots,X_n$ be independent real random variables having probability density functions $p_1,\ldots,p_n$ respectively with

(6) $$EX_i = \int xp_i(x)dx = 0,$$

and let τ_i: $R \to R$ be defined by

(7) $$\tau_i(x) = \frac{\int_x^\infty yp_i(y)dy}{p_i(x)} = -\frac{\int_{-\infty}^x yp_i(y)dy}{p_i(x)}.$$

Then, for any continuous and piecewise continuously differentiable function f such that, for all $i \in \{1,\ldots,n\}$,

(8) $$E|f'(W)|\tau_i(X_i) < \infty,$$

where

(9) $$W = \sum_{i=1}^n X_i,$$

we have

(10) $$E[(\sum_{i=1}^n E^W \tau_i(X_i))f'(W) - Wf(W)] = 0.$$

Proof: For each $i \in \{1,\ldots,n\}$, let $X^{(i)}$ denote the collection of random variables $\{X_j\}_{j\neq i}$. Because X_i is independent of $X^{(i)}$ we can apply Lemma 1 to X_i conditionally given $X^{(i)}$ to conclude that

$$E^{X^{(i)}}[\tau_i(X_i)f'(W) - X_i f(W)] = 0. \tag{11}$$

In (4) I have replaced f by the function

$$x \mapsto f(x + \sum_{j\neq i} X_j). \tag{12}$$

Taking unconditional expectation of (11), summing over i, and inserting E^W appropriately, we obtain (10).

Next we need to reformulate this lemma in order to emphasize its relevance to the normal approximation problem. This will be similar to the transition from Lemma I.3 to Lemma III.1. I shall need two preliminary lemmas.

Lemma 3: Let τ be a continuous and strictly positive valued function on an open interval (a,b) with

$$a < 0 < b, \tag{13}$$

where we may have $a = -\infty$ or $b = +\infty$ or both, and suppose

$$\int_0^b \frac{ydy}{\tau(y)} = \int_0^{-a} \frac{ydy}{\tau(-y)} = \infty. \tag{14}$$

Then there exists a unique probability density function p on (a,b) having mean 0 such that, for all $x \in (a,b)$

$$\tau(x) = \frac{\int_x^b yp(y)dy}{p(x)}. \tag{15}$$

This density p is given by

$$p(x) = \frac{1}{C} \cdot \frac{e^{-\int_0^x \frac{ydy}{\tau(y)}}}{\tau(x)}, \tag{16}$$

where

$$C = \int_a^b e^{-\int_0^z \frac{ydy}{\tau(y)}} \frac{dz}{\tau(z)}. \tag{17}$$

Proof: Multiplying (15) through by $p(x)$ and differentiating, we obtain

(18) $$\tau(x)p'(x) + (x+\tau'(x))p(x) = 0.$$

Dividing by $p(x)\tau(x)$ and integrating, we find

(19) $$\log p(x) - \log p(0) = -\int_0^x \frac{y+\tau'(y)}{\tau(y)}\,dy$$

$$= -\int_0^x \frac{ydy}{\tau(y)} - [\log \tau(x) - \log \tau(0)],$$

which is (16). Thus there is at most one probability density function p satisfying (15). It remains to verify that, subject to (14), this p does actually satisfy (15). We have

(20) $$\frac{\int_x^b zp(z)dz}{p(x)} = \frac{\int_x^b e^{-\int_0^z \frac{ydy}{\tau(y)}} \frac{zdz}{\tau(z)}}{\frac{1}{\tau(x)} e^{-\int_0^x \frac{ydy}{\tau(y)}}} = \tau(x) \frac{\int_x^b d(-e^{-\int_0^z \frac{ydy}{\tau(y)}})}{e^{-\int_0^x \frac{ydy}{\tau(y)}}}$$

$$= \tau(x).$$

One can verify similarly that

(21) $$\int_a^b xp(x)dx = 0.$$

Lemma 4: Suppose τ: $R \to R$ satisfies the conditions of Lemma 3. Then, for given bounded piecewise continuous h: $(a,b) \to R$, the differential equation

(22) $$\tau(w)f'(w) - wf(w) = h(w)$$

has a bounded continuous and piecewise continuously differentiable solution f: $(a,b) \to R$ if and only if

(23) $$E_{(\tau)}h = 0,$$

where $E_{(\tau)}$ is expectation under the density p defined in (16), that is

(24) $$E_{(\tau)}h = \frac{\int_a^b h(s)e^{-\int_0^x \frac{ydy}{\tau(y)}} \frac{dx}{\tau(x)}}{\int_a^b e^{-\int_0^x \frac{ydy}{\tau(y)}} \frac{dx}{\tau(x)}}.$$

When (23) is satisfied, the unique bounded solution f of (22) is given by

$$(25) \qquad f(w) = \int_a^w h(x) e^{\int_x^w \frac{ydy}{\tau(y)}} \frac{dx}{\tau(x)}$$

$$= -\int_w^b h(x) e^{-\int_w^x \frac{ydy}{\tau(y)}} \frac{dx}{\tau(x)} .$$

Proof: It follows from the elementary theory of first order linear differential equations that the differential equation (22) has a one-parameter family of solutions given by

$$(26) \qquad f(w) = e^{\int_0^w \frac{ydy}{\tau(y)}} [C' + \int_a^w h(x) e^{-\int_0^x \frac{ydy}{\tau(y)}} \frac{dx}{\tau(x)}]$$

and no other solution. Because of (14) the first factor in (26) approaches ∞ as w approaches a or b. Letting w approach a we see that in order for f to be bounded we must have

$$(27) \qquad C' = 0.$$

Letting w approach b we see that in order for f to be bounded we must also have

$$(28) \qquad C' = -\int_a^b h(x) e^{-\int_0^x \frac{ydy}{\tau(y)}} \frac{dx}{\tau(x)} .$$

The necessity of (23) follows from (27) and (28). Substituting (27) in (26) we obtain (25). Here and in Lemma 3 I have used the convention that

$$(29) \qquad \int_\alpha^\beta g(x)dx = -\int_\beta^\alpha g(x)dx.$$

It will be useful to fit these results into the abstract framework of the lower line of Diagram (I.28). With the present specializations I shall formulate this as

(30) $$\mathfrak{F}_0 \underset{U_{(\tau)}}{\overset{T_{(\tau)}}{\rightleftarrows}} \mathcal{X}_0 \underset{\iota_0}{\overset{E_{(\tau)}}{\rightleftarrows}} R.$$

Here $\mathcal{X}_0$ is the linear space of all bounded piecewise continuous $h: (a,b) \to R$ and $E_{(\tau)}: \mathcal{X}_0 \to R$ is defined by (24). Also $\mathfrak{F}_0$ is the linear space of all continuous and piecewise continuously differentiable $f: (a,b) \to R$ for which the function

(31) $$w \mapsto |wf(w)| + |\tau(w)f'(w)|$$

is bounded. The linear mapping $T_{(\tau)}: \mathfrak{F}_0 \to \mathcal{X}_0$ is defined by

(32) $$(T_{(\tau)}f)(w) = \tau(w)f'(w) - wf(w)$$

and the linear mapping $U_{(\tau)}: \mathcal{X}_0 \to \mathfrak{F}_0$ by

(33) $$(U_{(\tau)}h)(w) = \int_a^w [h(x) - E_{(\tau)}h]e^{\int_x^w \frac{ydy}{\tau(y)}} \frac{dx}{\tau(x)}$$

$$= -\int_w^b [h(x) - E_{(\tau)}h]e^{-\int_w^x \frac{ydy}{\tau(y)}} \frac{dx}{\tau(x)}.$$

We must verify that these formulas define linear mappings between the appropriate spaces. The linearity is obvious as is the fact that $T_{(\tau)}$ is a mapping on $\mathfrak{F}_0$ to $\mathcal{X}_0$ because of the rather artificial definition of $\mathfrak{F}_0$. It follows from Lemma 4 that $U_{(\tau)}h$ satisfies the differential equation

(34) $$\tau(w)(U_{(\tau)}h)'(w) - w(U_{(\tau)}h)(w) = h(w) - E_{(\tau)}h.$$

Thus, to prove that, for $h \in \mathcal{X}_0$, $U_{(\tau)}h \in \mathfrak{F}_0$ we need only show that the function $w \mapsto w(U_{(\tau)}h)(w)$ is bounded. From the second form of (33) we have, for $w \geq 0$

(35) $$w|(U_{(\tau)}h)(w)| \leq 2(\sup|h|)w\int_w^b e^{\int_0^w \frac{ydy}{\tau(y)} - \int_0^x \frac{ydy}{\tau(y)}} \frac{dx}{\tau(x)}$$

$$\leq 2 \sup |h|$$

since, with

(36) $$u(x) = \int_0^x \frac{ydy}{\tau(y)}$$

we have

(37) $$\int_w^b e^{\int_0^w \frac{ydy}{t(y)} - \int_0^x \frac{ydy}{t(y)}} \frac{dx}{\tau(x)} = \int_w^b e^{u(w)-u(x)} \frac{du(x)}{x} < \frac{1}{w}.$$

A similar result for $w \leq 0$ follows from the first form of (33). Thus $U_{(\tau)}h \in \mathfrak{F}_0$. The identity (I.30), in this case

(38) $$T_{(\tau)} \circ U_{(\tau)} = I_{\mathcal{X}_0} - \iota_0 \circ E_{(\tau)},$$

was proved in Lemma 4. Thus the diagram (30) satisfies all the conditions imposed on the lower row of Diagram (I.28).

The function τ, related to the density p by (2), takes a very simple form when p is one of Pearson's family of probability density functions (with mean 0).

Theorem 1: Let p be a probability density function on an open interval (a,b) and let τ be related to p by (2). Then in order that the function τ have the form

(39) $$\tau(x) = \alpha x^2 + \beta x + \gamma$$

with α, β, and γ constant, it is necessary and sufficient that p satisfy the differential equation

(40) $$p'(x) = - \frac{(2\alpha+1)x+\beta}{\alpha x^2+\beta x+\gamma}\, p(x).$$

Proof: Because of (2), (39) is equivalent to

(41) $$\int_x^b yp(y)dy - (\alpha x^2+\beta x+\gamma)p(x) = 0.$$

By differentiation this implies that

(42) $$-xp(x) - (2\alpha x+\beta)p(x) - (\alpha x^2+\beta x+\gamma)p'(x) = 0,$$

which is (40). Of course we must have

(43) $$\tau(a) = 0$$

if a is finite and

(44) $$\tau(b) = 0$$

if b is finite. Then the argument from (41) to (40) by way of (42) can be reversed. If b is finite we use (44) and if b is infinite we use the existence of the expectation to go from (42) to (41) by integration. For a discussion of Pearson's curves see for example Kendall and Stuart (1963), Vol. I, pp. 148-154.

We can apply these considerations to the approximation of the distribution of a sum of independent random variables with densities by a nearly arbitrary distribution having a density.

Theorem 2: Let $p_1,\ldots,p_n$ be probability density functions with

(45) $$\int xp_i(x)dx = 0$$

for all $i \in \{1,\ldots,n\}$ and let $X_1,\ldots,X_n$ be independent random variables with densities $p_1,\ldots,p_n$. Also let p be a probability density function with mean 0 on an open interval (a,b), continuous and nonvanishing, and let the $\{\tau_i\}$ and τ be related to the $\{p_i\}$ and p as in (2). Then, for any bounded piecewise continuous $h: (a,b) \to R$,

(46) $$Eh(W) = E_{(\tau)}h + E[\tau(W) - \sum_{i=1}^{n}\tau_i(X_i)](U_{(\tau)}h)'(W),$$

where $E_{(\tau)}$ is defined by (24) and $U_{(\tau)}$ by (33) and

(47) $$W = \sum_{i=1}^{n}X_i.$$

Proof: By Lemma 2 and the definition of $E_{(\tau)}$ and $U_{(\tau)}$ we have

(48) $$0 = E[(\sum_{i=1}^{n}\tau_i(X_i))(U_{(\tau)}h)'(W) - W(U_{(\tau)}h)(W)]$$

$$= E[\sum_{i=1}^{n}\tau_i(X_i) - \tau(W)](U_{(\tau)}h)'(W)$$

$$+ E[\tau(W)(U_{(\tau)}h)'(W) - W(U_{(\tau)}h)(W)]$$

$$= E[\sum_{i=1}^{n}\tau_i(X_i) - \tau(W)](U_{(\tau)}h)'(W) + Eh(W) - E_{(\tau)}h.$$

I had originally intended this lecture as preparation for the detailed study of the normal approximation problem in the continuous case. When it became clear that such a study would not be made, I decided to include this lecture anyway because it is reasonably simple and seems likely to be useful eventually. Lemma 1 gives a characterization of an arbitrary random variable having a density function (and mean 0) that is analogous to the characterization of a standard normal variable by $E[f'(W)-Wf(W)] = 0$. Lemma 2 suggests a possible way of applying this to the study of sums of independent random variables. Lemmas 3 and 4 are preliminary to the specialization, below (30), of the lower row of the fundamental diagram (I.28) to the present situation. Theorem 2 continues the development of Lemma 2 in the light of the basic formalism.

A mildly surprising relation of these ideas to Pearson's curves is expressed by Theorem 1. This was discovered by accident. After defining the function τ associated with a density function p by (2), I decided to compute it in a number of special cases: normal, uniform, χ^2, and Student's t, and was surprised to find that, in every case, τ was a polynomial of degree at most two. I eventually realized that these densities were all special cases of Pearson's curves.

The ideas of this lecture originated as a limiting case of an approach to discrete problems that fits more naturally into the basic abstract formalism. Unfortunately I cannot include a detailed treatment in these lectures because the results are fragmentary although promising. For a brief description let us look at the case of a sum of independent random variables. In the development starting with (I.51), instead of replacing X_I by an independent X_I^*, we replace X_I by an adjacent value or itself (which I shall still call X_I^*) in such a way that (W,W') with W' defined by (I.52) is an exchangeable pair and, for some λ, $E^W W' = (1-\lambda)W$. It turns out that there is a unique way of doing this. In this way $|W'-W|$ is made smaller, sometimes at the cost of substantial increase in complication.