LECTURE I. THE BASIC APPROACH

In this first lecture I shall describe the abstract approach and one way to specialize it to the normal approximation problem. The latter will be illustrated by a very superficial treatment of the familiar problem of sums of independent random variables. Some of the technical details will be postponed until the next two lectures.

Let $(\Omega,\mathcal{B},P)$ be a probability space and H a real-valued random variable on this space with $E|H| < \infty$, where E is the operation of expectation under P. My aim is to discuss the problem of approximating EH, which often arises in the following way. We are given a real-valued random variable W on $(\Omega,\mathcal{B},P)$ and want to approximate the cumulative distribution function of W. If we choose $H = h_{w_0}(W)$ where

(1) $$h_{w_0}(w) = \begin{cases} 1 & \text{if } w \leq w_0 \\ 0 & \text{if } w > w_0, \end{cases}$$

then

(2) $$P\{W \leq w_0\} = EH.$$

The approach will be based on the following easy lemma together with a bit of linear algebra.

<u>Lemma 1</u>: Let $(\Omega_1,\mathcal{B}_1,P_1)$ be a probability space with associated expectation operation E_1, and let (X,X') be an exchangeable pair of mappings of

$(\Omega_1,\mathcal{B}_1,P_1)$ into another probability space $(\Omega,\mathcal{B},P)$ in the sense that

$$(3) \qquad P_1\{X \varepsilon B\} = P(B)$$

for all $B \varepsilon \mathcal{B}$ and

$$(4) \qquad P_1\{X \varepsilon B \;\&\; X' \varepsilon B'\} = P_1\{X \varepsilon B' \;\&\; X' \varepsilon B\}$$

for all $B, B' \varepsilon \mathcal{B}$. Then, for all measurable functions $F\colon\ \Omega^2 \to R$ that are antisymmetric in the sense that, for all $x, x' \varepsilon \Omega$

$$(5) \qquad F(x,x') = -F(x',x)$$

we have

$$(6) \qquad E_1E_1^X F(X,X') = 0,$$

provided, of course,that

$$(7) \qquad E_1|F(X,X')| < \infty.$$

In (6), E_1^X denotes conditional expectation given X.

Proof. We have

$$(8) \qquad E_1F(X,X') = E_1F(X',X) = E_1(-F(X,X')) = -E_1F(X,X').$$

The first equality uses the exchangeability of X and X' and the second uses the antisymmetry of F. It follows that

$$(9) \qquad 0 = E_1F(X,X') = E_1E_1^X F(X,X'),$$

which is (6).

This result is useful in that it gives us, ordinarily, a large class of random variables, the $E_1^X F(X,X')$, whose expectations are 0. In fact, under appropriate conditions these are all of the random variables on $(\Omega,\mathcal{B},P)$ whose expectation is 0. This is stated in the following lemma, the proof of which will be postponed until the second lecture.

<u>Lemma 2</u>: With Ω_1 finite, let $(\Omega_1,\mathcal{B}_1,P_1)$, $(\Omega,\mathcal{B},P)$, and (X,X') be as in Lemma 1. Suppose also that, for every $x,x^* \varepsilon \Omega$ there exists a finite sequence $x_1,\ldots,x_k$ of elements of Ω such that

(10) $$x = x_1 \quad \text{and} \quad x^* = x_k$$

and, for all $i \in \{1,\ldots,k-1\}$

(11) $$P_1\{X = x_i \;\&\; X' = x_{i+1}\} > 0.$$

Then for every function $g: \Omega \to R$ such that

(12) $$Eg(X) = 0$$

there exists an antisymmetric function $F: \Omega^2 \to R$ such that

(13) $$g(X) = E_1^X F(X,X').$$

It may be useful to say a few words about notation. First there is very little conceptual loss in considering only the case of finite Ω_1 although many applications will require the more general results. Second, I shall often ignore the distinction between P and P_1 and between E and E_1 and use P(or E) for P_1(or E_1). I shall often suppress the reference to $(\Omega_1,\mathcal{B}_1,P_1)$ except when verifying the correctness of one or two examples of a type of construction that will be used repeatedly. In more formal notation (3) and (4) could be written

(14) $$P_1(\overset{-1}{X}(B)) = P(B)$$

and

(15) $$P_1(\overset{-1}{X}(B) \cap \overset{-1}{X'}(B')) = P_1(\overset{-1}{X}(B') \cap \overset{-1}{X'}(B)).$$

It will be convenient to summarize the results of Lemmas 1 and 2 by introducing the diagram

(16) $$\mathfrak{F} \xrightarrow{T} \mathcal{X} \xrightarrow{E} R$$

Here R is the linear space of real numbers, $\mathcal{X}$ the space of all measurable $h: \Omega \to R$ such that

(17) $$E|h(X)| < \infty,$$

and $\mathfrak{F}$ the space of all antisymmetric $F: \Omega^2 \to R$ such that

(18) $$E|F(X,X')| < \infty.$$

The linear mapping $E: \mathcal{X} \to R$ is the expectation operator, with a slight confusion of notation in that

(19) $$Eh = Eh(X).$$

The linear mapping $T\colon\ \mathfrak{F} \to \mathcal{X}$ is defined by the condition that

(20) $$(TF)(X) = E^X F(X,X').$$

Then Lemma 1 asserts that, if (X,X') is an exchangeable pair, then

(21) $$E \circ T = 0,$$

where $E \circ T$ is the composition, defined by

(22) $$(E \circ T)F = E(TF).$$

The identity (21) can be expressed in terms of kernels and images. If $\mathcal{X}$ and $\mathcal{Y}$ are linear spaces and $T\colon\ \mathcal{X} \to \mathcal{Y}$ a linear mapping, the linear spaces $\ker T \subset \mathcal{X}$ and $\operatorname{im} T \subset \mathcal{Y}$ are defined by

(23) $$\ker T = \{x\colon\ Tx = 0\}$$

and

(24) $$\operatorname{im} T = \{Tx\colon\ x \in \mathcal{X}\}.$$

Then (21) can be restated as

(25) $$\ker E \supset \operatorname{im} T.$$

Also, Lemma 2 asserts that, under the additional connectedness condition stated there,

(26) $$\ker E = \operatorname{im} T.$$

Next I shall introduce an approximation

(27) $$\mathfrak{F}_0 \xrightarrow{T_0} \mathcal{X}_0 \xrightarrow{E_0} R$$

to the diagram (16), leading to the larger diagram

(28) $$\begin{tikzcd}
\mathfrak{F} \arrow[r, "T"] & \mathcal{X} \arrow[dr, "E"] & \\
 & & R \arrow[dl, shift left, "\iota_0"] \\
\mathfrak{F}_0 \arrow[uu, "\alpha"] \arrow[r, shift left, "T_0"] & \mathcal{X}_0 \arrow[uu, "\iota"'] \arrow[l, shift left, "U_0"] \arrow[ur, shift left, "E_0"] &
\end{tikzcd}$$

Here $\mathcal{X}_0$ is a linear subspace of $\mathcal{X}$, $\mathfrak{F}_0$ a linear space, and $\iota\colon\ \mathcal{X}_0 \to \mathcal{X}$ and $\iota_0\colon\ R \to \mathcal{X}_0$ the appropriate inclusion mappings. Thus, for $c \in R$, $\iota_0 c$ is the random variable taking the constant value c, and $\mathcal{X}_0$ is required to contain this random variable. Also

(29) $$\iota h = h$$

for $h \in \mathcal{X}_0$. The linear mapping E_0: $\mathcal{X}_0 \to R$ is intended as an approximation to $E \circ \iota$, and it is also intended that $\iota \circ T_0$ be an approximation to $T \circ \alpha$ but of course these are not formal conditions. In addition to (21) it is assumed that

(30) $$T_0 \circ U_0 = I_{\mathcal{X}_0} - \iota_0 \circ E_0,$$

where $I_{\mathcal{X}_0}$ is the identity mapping of $\mathcal{X}_0$. Then for $h \in \mathcal{X}_0$

(31) $$\begin{aligned} 0 &= E((T \circ \alpha \circ U_0)h) \\ &= E((\iota \circ T_0 \circ U_0)h) + E((T \circ \alpha - \iota \circ T_0) \circ U_0)h \\ &= E(\iota \circ (I_{\mathcal{X}_0} - \iota_0 \circ E_0)h) + E((T \circ \alpha - \iota \circ T_0) \circ U_0)h \\ &= Eh - E_0 h + E((T \circ \alpha - \iota \circ T_0) \circ U_0)h. \end{aligned}$$

The first equality uses (21) and the third uses (30). The final equality uses the fact that for $c \in R$

(32) $$Ec = c.$$

The identity (31) can be rewritten as

(33) $$Eh = E_0 h + E((T \circ \alpha - \iota \circ T_0) \circ U_0)h.$$

This can be thought of as asserting that Eh is approximated by $E_0 h$ with an error given by the second term on the right hand side. If the choices have been made appropriately it may not be too difficult to bound the remainder term.

Now let us specialize these considerations to a method of treating the normal approximation problem. This will then be specialized to the case of a sum of independent random variables.

Lemma 3: Let (W,W') be an exchangeable pair of real random variables such that

(34) $$E^W W' = (1-\lambda)W$$

and

(35) $$E(W'-W)^2 = 2\lambda$$

with

(36) $$0 < \lambda < 1.$$

Then for every piecewise continuous function $f\colon R \to R$ for which there exists $C > 0$ such that, for all $w \varepsilon R$,

(37) $$|f(w)| \leq C(1+|w|),$$

we have

(38) $$E[Wf(W) - \frac{1}{2\lambda}(W'-W)(f(W')-f(W))] = 0.$$

Proof: Since the function $F\colon R^2 \to R$ defined by

(39) $$F(w,w') = (w'-w)(f(w)+f(w'))$$

is antisymmetric in the sense of (5), we have

(40) $$\begin{aligned} 0 &= E(W'-W)[f(W)+f(W')] \\ &= E(W'-W)\{2f(W)+[f(W')-f(W)]\} \\ &= 2E(E^W W'-W)f(W)+E(W'-W)[f(W')-f(W)] \\ &= -2\lambda EWf(W)+E(W'-W)[f(W')-f(W)]. \end{aligned}$$

which is essentially the same as (38).

Lemma 4: With W, W' and λ as in Lemma 3, let $h\colon R \to R$ be a piecewise continuous function for which there exists $C' > 0$ such that, for all $w \varepsilon R$

(41) $$|h(w)| \leq C'(1+w^2).$$

Let

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

and let f be the unique solution of the differential equation

(43) $$f'(w)-wf(w) = h(w)-Nh$$

satisfying (37), that is

$$(44) \qquad f(w) = e^{\frac{1}{2}w^2}\int_{-\infty}^{w}[h(x)-Nh]e^{-\frac{1}{2}x^2}dx$$

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

Then

$$(45) \qquad Eh(W) = Nh+E[f'(W) - \frac{1}{2\lambda}E^W(W'-W)(f(W')-f(W))].$$

Proof: The equality of the two alternative expressions in (44) follows from

$$(46) \qquad \int_{-\infty}^{\infty}[h(x)-Nh]e^{-\frac{1}{2}^2}dx = 0,$$

which is a consequence of (42). Differentiating f defined by (44) we see that it satisfies (43). From the fact that, for $w > 1$

$$(47) \quad |f(w)| \leq e^{\frac{1}{2}w^2}\int_{w}^{\infty}\frac{x}{w}|h(x)-Nh|e^{-\frac{1}{2}x^2}dx \leq \frac{e^{\frac{1}{2}w^2}}{w}\int_{w}^{\infty}C''(1+x^2)xe^{-\frac{1}{2}x^2}dx \leq \frac{C}{w}(1+w^2),$$

for appropriate C'' and C, together with a similar result for $w < -1$ and a trivial bound for $-1 \leq w \leq 1$, it follows that f satisfies (37). The uniqueness of the solution of (43) satisfying (37) follows from the fact that the solutions of the homogeneous equation corresponding to (43) are $C'''e^{w^2/2}$.

Substituting this f in (38) we obtain

$$(48) \qquad 0 = E[Wf(W) - \frac{1}{2\lambda}(W'-W)(f(W')-f(W))]$$

$$= E[Wf(W)-f'(W)] + E[f'(W) - \frac{1}{2\lambda}(W'-W)(f(W')-f(W)]$$

$$= E[Nh-h(W)] + E[f'(W) - \frac{1}{2\lambda}E^W(W'-W)(f(W')-f(W))]$$

which is (45).

In order to derive a normal approximation theorem from Lemma 4 we need to bound the second term on the right hand side of (45). Here I shall give only a rough indication of the argument, postponing the details until the third lecture. If h is continuous then f, defined by (44) is continuously differentiable and we can approximate the expectation of the second term in brackets on the right hand side of (45) by

$$\frac{1}{2\lambda} EE^W(W'-W)(f(W')-f(W)) \approx \frac{1}{2\lambda} EE^W(W'-W)^2 f'(W) \approx Ef'(W). \tag{49}$$

The first approximate equality uses the first term of a Taylor series approximation to $f(W')-f(W)$ and the second approximation assumes that

$$\frac{1}{2\lambda} E^W(W'-W)^2 \approx 1. \tag{50}$$

If the left hand side of (50) is approximately constant it must be approximately 1 by (35). Actually (35) was not required in either Lemma 3 or Lemma 4 but it will ordinarily hold in the applications.

Now let us specialize the considerations of Lemma 3 and Lemma 4 to the case of a sum of independent random variables. Let $X_1,\dots,X_n$ be independent real random variables and

$$W = \sum_{i=1}^{n} X_i. \tag{51}$$

In order to apply Lemmas 3 and 4 we introduce additional random variables I, $X_1^*,\dots,X_n^*$ and W' defined in the following way. The random variables I, $X_1,\dots,X_n$, $X_1^*,\dots,X_n^*$ are independent, I is uniformly distributed over the index set $\{1,\dots,n\}$, each X_i^* has the same distribution as the corresponding X_i, and

$$W' = W + (X_I^*-X_I), \tag{52}$$

that is, W' is obtained by replacing X_I in the sum (51) by X_I^*. It is easy to see that (W,W') is an exchangeable pair since we can first choose I, then the X_j for $j \neq I$ and the unordered pair $\{X_I, X_I^*\}$. Finally we choose one element of this pair to make up W and the other to make up W', each of the two possible choices being assigned probability one-half.

Now suppose that, for each $i \in \{1,\dots,n\}$,

$$EX_i = 0 \tag{53}$$

and also

$$\sum_{i=1}^{n} EX_i^2 = 1. \tag{54}$$

Then

$$E^W W' = W+E^W X_I^* - E^W X_I = W+0 - \frac{1}{n}\sum X_i = (1-\frac{1}{n})W \tag{55}$$

and

$$E(W'-W)^2 = E(X_I^*-X_I)^2 = \frac{1}{n}\sum_{i=1}^{n} E(X_i^*-X_i)^2 = \frac{2}{n}\sum EX_i^2 = \frac{2}{n}. \tag{56}$$

Thus (34) and (35) are satisfied with

$$\lambda = \frac{1}{n}. \tag{57}$$

For the approximate equality (50), we have

$$E^X(W'-W)^2 = E^X(X_I^*-X_I)^2 = \frac{1}{n}\sum E^X(X_i^*-X_i)^2 = \frac{1}{n}\sum [EX_i^{*2}+X_i^2] = \frac{1}{n}(1+\sum_1^n X_i^2) \tag{58}$$

and under appropriate conditions, by the law of large numbers, ΣX_i^2 is, with high probability, close to its expected value 1.

The details of this special case will also be postponed to the third lecture. There I shall also apply Lemmas 3 and 4 to the distribution of the sum of a random diagonal, which includes the case of the mean of a sample from a finite population and some other random variables that arise in statistical problems.

This lecture begins with the theoretical framework that will be applied in all the lectures in this series. Our aim is to evaluate expectations of certain functions of a random point X. In Lemma 1 we introduce an exchangeable pair (X,X') and observe that the conditional expectation given X of an anti-symmetric function of X and X' has expectation 0. Following the technical Lemma 2, we examine a way of using this simple fact in the diagram (28). In this way the problem of computing expectations, usually approximately, is divided into a number of steps:

(i) Choose an exchangeable pair, thereby constructing the top row of (28), satisfying (21).

(ii) Construct the bottom row of (28), satisfying (30), in such a way that, roughly speaking, $T\circ\alpha-\iota\circ T_0$ is small.

(iii) Study the boundedness properties of the linear mapping U_0.

(iv) Use these to bound the remainder in (31).

A more abstract and more elegant version of diagram (28) is given in the latter part of Lecture XIV.

The second part of this lecture describes a first specialization of this formalism to the study of normal approximation problems. Lemma 3 corresponds to the top row and left column of diagram (28), and Lemma 4 to the rest of the diagram. A rough indication of the way this can be applied to study the normal approximation for a sum of independent random variables starts with (51). However, no real proof or even precise statement is given for this result until the third lecture, after the study of the boundedness properties of the appropriate specialization of U_0 in Lecture II.