LECTURE VIII. POISSON APPROXIMATIONS

An example of approximation by the Poisson distribution has already been given in the seventh lecture. Here I shall discuss this subject in the context of the abstract formalism of the first lecture, with special emphasis on the classical problem of the total number of occurrences of a large number of independent random events with small probabilities. Most of this work was done by Chen (1975a).

<u>Theorem 1</u>: In order that the random variable W taking values in Z^+, the set of all non-negative integers, have a Poisson distribution with parameter λ it is necessary and sufficient that, for all bounded functions $f: Z^+ \to R$,

(1) $$E[\lambda f(W+1) - Wf(W)] = 0.$$

<u>Proof of necessity</u>: Suppose W has a Poisson distribution with parameter λ, that is, for all $w \in Z^+$

(2) $$P\{W=w\} = e^{-\lambda} \frac{\lambda^w}{w!}.$$

Then, for all bounded $f: Z^+ \to R$

(3) $$EWf(W) = e^{-\lambda} \sum_{w=0}^{\infty} wf(w) \frac{\lambda^w}{w!}$$

$$= e^{-\lambda}\lambda \sum_{w'=0}^{\infty} f(w'+1) \frac{\lambda^{w'}}{w'!} = \lambda Ef(W+1).$$

Observe that the value of f(0) is irrelevant to this result. Of course the identity (3) does not really require f to be bounded. It is valid if the expectation on either side exists.

Proof of sufficiency: Assuming that W is a random variable taking values in Z^+ and that (1) holds for all bounded $f: Z^+ \to R$, let $h: Z^+ \to R$ be a bounded function and define

$$\mathcal{P}_\lambda h = e^{-\lambda} \sum_{w=0}^{\infty} h(w) \frac{\lambda^w}{w!}. \tag{4}$$

We shall see that the equation

$$\lambda f(w+1) - wf(w) = h(w) - \mathcal{P}_\lambda h \tag{5}$$

has a bounded solution f. Then it will follow from (1) that

$$E[h(W) - \mathcal{P}_\lambda h] = E[\lambda f(W+1) - Wf(W)] = 0, \tag{6}$$

which is the desired result.

A bounded solution f of (5) can be constructed by choosing f(0) arbitrarily, say

$$f(0) = 0 \tag{7}$$

and, for $w \geq 1$, defining f(w) by

$$\begin{aligned} f(w) &= - \sum_{\ell=w}^{\infty} \frac{(w-1)!}{\ell!} \lambda^{\ell-w}[h(\ell)-\mathcal{P}_\lambda h] \\ &= \sum_{\ell=0}^{w-1} \frac{(w-1)!}{\ell!} \lambda^{\ell-w}[h(\ell)-\mathcal{P}_\lambda h]. \end{aligned} \tag{8}$$

The equality of the two expressions for f follows from

$$\begin{aligned} &\sum_{\ell=0}^{\infty} \frac{(w-1)!}{\ell!} \lambda^{\ell-w}[h(\ell)-\mathcal{P}_\lambda h] \\ &= (w-1)!\, \lambda^{-w} e^{\lambda} \mathcal{P}_\lambda[h-\mathcal{P}_\lambda h] = 0. \end{aligned} \tag{9}$$

If h is bounded, say

$$|h(w)| \leq C \tag{10}$$

for all $w \in Z^+$, then

$$(11) \qquad |f(w)| \leq 2C \sum_{\ell=w}^{\infty} \frac{(w-1)!}{\ell!} \lambda^{\ell-w} \leq 2Ce^{\lambda},$$

again for all $w \geq 1$, so that f is also bounded. To verify (5) for $w \geq 1$ we substitute (8) into the left-hand-side of (5), obtaining

$$(12) \qquad \begin{aligned} &\lambda f(w+1) - wf(w) \\ &= -\lambda \sum_{\ell=w+1}^{\infty} \frac{w!}{\ell!} \lambda^{\ell-(w+1)}[h(\ell)-\mathcal{P}_{\lambda}h] \\ &\quad + w \sum_{\ell=w}^{\infty} \frac{(w-1)!}{\ell!} \lambda^{\ell-w}[h(\ell)-\mathcal{P}_{\lambda}h] \\ &= h(w) - \mathcal{P}_{\lambda}h. \end{aligned}$$

On the other hand, for $w = 0$ the left-hand side of (5) is

$$(13) \qquad \lambda f(1) = h(0) - \mathcal{P}_{\lambda}h$$

by the second form of (8). This completes the proof of Lemma 1.

Now we want to express this in the form of a diagram

$$(14) \qquad \mathcal{F}_0 \underset{U_\lambda}{\overset{T_\lambda}{\rightleftarrows}} \mathcal{X}_0 \underset{\iota_0}{\overset{\mathcal{P}_\lambda}{\rightleftarrows}} R,$$

a special case of the lower line of diagram (I.28). The choice of the linear spaces $\mathcal{X}_0$ and $\mathcal{F}_0$ is largely arbitrary, but for definiteness I define $\mathcal{X}_0$ to be the space of functions of at most exponential growth, that is, functions $h: Z^+ \to R$ for which there exist positive constants A and B such that, for all $w \in Z^+$,

$$(15) \qquad |h(w)| \leq Ae^{Bw},$$

and I take

$$(16) \qquad \mathcal{F}_0 = \mathcal{X}_0 \cap \{f: f(0) = 0\}.$$

The linear mappings occurring in the diagram (14) are defined by

(17) $$(T_\lambda f)(w) = \lambda f(w+1) - wf(w)$$

(18) $$(U_\lambda h)(w) = -\sum_{\ell=w}^{\infty} \frac{(w-1)!}{\ell!} \lambda^{\ell-w}[h(\ell)-\mathcal{P}_\lambda h]$$

$$= \sum_{\ell=0}^{w-1} \frac{(w-1)!}{\ell!} \lambda^{\ell-w}[h(\ell)-\mathcal{P}_\lambda h]$$

$$= \sum_{\ell=0}^{\infty} \frac{(w-1)!}{\ell!} \lambda^{\ell-w}(\mathcal{J}\{\ell\leq w-1\} - e^{-\lambda} \sum_{\ell'=0}^{w-1} \frac{\lambda^{\ell'}}{\ell'!})h(\ell),$$

with $\mathcal{P}_\lambda$ given by (4) and ι_0 an inclusion mapping. The final form of (18) follows easily from (4). Here $T_\lambda f$ occurs on the left-hand side of (5) and $U_\lambda h$ was denoted by f in (8). We should verify that, for $h \in \mathcal{X}_0$, $U_\lambda h \in \mathcal{F}_0$. Assuming (15), we have, by the first form of (18),

(19) $$|(U_\lambda h)(w)|$$

$$\leq (A+|\mathcal{P}_\lambda h|) \sum_{\ell=w}^{\infty} \frac{(w-1)!}{\ell!} \lambda^{\ell-w} e^{B\ell}$$

$$\leq (A+|\mathcal{P}_\lambda h|)e^{Bw} \sum_{\ell'=0}^{\infty} \frac{1}{\ell'!} (\lambda e^B)^{\ell'},$$

which is of at most exponential growth.

Now let us look at the case where W is the number of occurrences of a large number of independent events. More precisely, let $X_1,\ldots,X_n$ be independent random variables taking on only the values 0 and 1 and let

(20) $$p_i = P\{X_i=1\},$$

(21) $$\lambda = \sum_{i=1}^{n} p_i$$

and

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

It has long been known that if all the p_i are small then W has approximately a Poisson distribution with parameter λ, but I believe a result essentially equivalent to (43) below (but with a constant factor greater than 1 on the r.h.s.) was first obtained by Le Cam (1960). It was later proved, using essentially the present method, by Chen (1975a).

Let $X_1^*,\ldots,X_n^*$ be independent random variables independent of the X_i and suppose that, for each i, X_i^* has the same distribution as X_i. With I uniformly distributed in $\{1,\ldots,n\}$ independent of the X_i and X_i^* let

$$W' = W - X_I + X_I^*. \tag{23}$$

Then (W,W') is an exchangeable pair of random variables since we can think of first determining I, then the unordered pair $\{X_I, X_I^*\}$, then the X_j for $j \neq I$ and finally choosing one of the two elements of the unordered pair $\{X_I, X_I^*\}$, with probability one-half independent of all the other choices, to be X_I. Then, for arbitrary $f: Z^+ \to R$,

$$\begin{aligned}(24)\qquad 0 &= EE^X[f(W')\mathscr{I}\{W'=W+1\} - f(W)\mathscr{I}\{W=W'+1\}] \\ &= E[f(W+1)P^X\{W'=W+1\} - f(W)P^X\{W'=W-1\}] \\ &= \frac{1}{n} E[f(W+1)\sum_{j=1}^{n} p_j(1-\mathscr{I}\{X_j=1\}) \\ &\qquad - f(W)\sum_{j=1}^{n}(1-p_j)\mathscr{I}\{X_j=1\}].\end{aligned}$$

At the last equality sign the following argument has been used. With probability $\frac{1}{n}$, I takes any particular value j. Given this, the conditional probability that $W'=W+1$ is $\mathscr{I}\{X_j=0\}p_j$ and the conditional probability that $W'=W-1$ is $\mathscr{I}\{X_j=1\}(1-p_j)$. We can rewrite the final form of (24) as

$$\begin{aligned}(25)\qquad & E[\lambda f(W+1) - Wf(W)] \\ &= E(\sum p_j\mathscr{I}\{X_j=1\})(f(W+1) - f(W)).\end{aligned}$$

Substituting

$$f = U_\lambda h \tag{26}$$

in (24) with $h\colon Z^+ \to R$ an arbitrary bounded function and U_λ defined by (18) we obtain

$$\begin{aligned} Eh(W) &= \mathcal{P}_\lambda h + E\Big(\sum_{j=1}^{n} p_j \mathcal{I}\{X_j=1\}\Big)(U_\lambda h(W+1) - U_\lambda h(W)) \\ &= \mathcal{P}_\lambda h + \sum_{j=1}^{n} p_j^2 \, EV_\lambda h(W_j) \end{aligned} \tag{27}$$

where

$$W_j = \sum_{j' \neq j} X_{j'} \tag{28}$$

and

$$(V_\lambda h)(w) = U_\lambda h(w+2) - U_\lambda h(w+1). \tag{29}$$

We can apply the same identity (27) to evaluate the right-hand side of (27), obtaining

$$\begin{aligned} Eh(W) &= \mathcal{P}_\lambda h + \sum_{j=1}^{n} p_j^2 \, \mathcal{P}_{\lambda - p_j} V_\lambda h \\ &\quad + \sum_{j_1=1}^{n} \sum_{j_2 \neq j_1} p_{j_1}^2 p_{j_2}^2 \, EV_{\lambda - p_j} V_\lambda h(W_{j_1,j_2}), \end{aligned} \tag{30}$$

where

$$W_{j_1,j_2} = \sum_{j \neq j_1, j_2} X_j. \tag{31}$$

It is clear how to iterate further, but we shall have enough to do in studying the remainders in (27) and (30).

For this purpose we must bound $V_\lambda h$ for appropriate choice of h, in particular for $h=h_A$, defined for $A \subset Z^+$ by

(32) $$h_A(w) = \begin{cases} 1 & \text{if } w \in A \\ 0 & \text{if } w \notin A \end{cases}$$

We can rewrite the final expression (18) for $U_\lambda h$ in the form

(33) $$(U_\lambda h)(w) = e^\lambda (w-1)!\lambda^{-w}(\mathcal{P}_\lambda(hh_{C_{w-1}}) - \mathcal{P}_\lambda h\, \mathcal{P}_\lambda h_{C_{w-1}})$$

where

(34) $$C_w = \{0,\dots,w\}.$$

Essentially this form was used by Barbour (1982) to obtain (42) in applying this method to the study of random graphs. I shall discuss this work in a later lecture. This expression is also instructive in suggesting that similar results hold for approximation by other distributions.

Let us start by observing that $U_\lambda h_{w_0}$ can be expressed in the form

(35) $$(U_\lambda h_{w_0})(w) = \begin{cases} \dfrac{(w-1)!}{w_0!}\lambda^{w_0-w}\mathcal{P}_\lambda h_{C^c_{w-1}} & \text{if } w_0 < w \\ -\dfrac{(w-1)!}{w_0!}\lambda^{w_0-w}\mathcal{P}_\lambda h_{C_{w-1}} & \text{if } w_0 \geq w \end{cases}$$

From this and the definition (29) of V_λ it follows that

(36) $$(V_\lambda h_{w_0})(w) = \begin{cases} -\lambda^{w_0-w-2}\dfrac{w!}{w_0!}[(w+1)\mathcal{P}_\lambda h_{C_{w+1}} - \lambda\mathcal{P}_\lambda h_{C_w}] & \text{if } w \leq w_0-2 \\ \lambda^{w_0-w-2}\dfrac{w!}{w_0!}[(w+1)\mathcal{P}_\lambda h_{C^c_{w+1}} + \lambda\mathcal{P}_\lambda h_{C_w}] & \text{if } w = w_0-1 \\ \lambda^{w_0-w-2}\dfrac{w!}{w_0!}[(w+1)\mathcal{P}_\lambda h_{C^c_{w+1}} - \lambda\mathcal{P}_\lambda h_{C^c_w}] & \text{if } w \geq w_0. \end{cases}$$

We observe that

(37) $$V_\lambda h_{w_0}(w) \begin{cases} > 0 & \text{if} \quad w = w_0-1 \\ < 0 & \text{if} \quad w \neq w_0-1. \end{cases}$$

The case $w = w_0-1$ is obvious. For the second part of (27), the case $w \leq w_0-2$ follows from

(38) $$(w+1)\mathcal{P}_\lambda h_{C_{w+1}} - \lambda \mathcal{P}_\lambda h_{C_w}$$

$$= e^{-\lambda} \sum_{k=0}^{w+1} \frac{\lambda^k}{k!} (w+1-k) > 0$$

and the case $w \geq w_0$ follows from

(39) $$(w+1)\mathcal{P}_\lambda h_{C^c_{w+1}} - \lambda \mathcal{P}_\lambda h_{C^c_w}$$

$$= e^{-\lambda} \sum_{k=w+2}^{\infty} \frac{\lambda^k}{k!} (w+1-k) < 0.$$

Thus for any $w \in Z^+$ and $A \subset Z^+$

(40) $$(V_\lambda h_{w+1})(w) \geq (V_\lambda h_A)(w) \geq (V_\lambda h_{\{w+1\}^c})(w).$$

But, by the middle case of (36)

(41) $$(V_\lambda h_{w+1})(w) = \lambda^{-1}[\mathcal{P}_\lambda h_{C^c_{w+1}} + \frac{\lambda}{w+1} \mathcal{P}_\lambda h_{C_w}]$$

$$\leq \lambda^{-1} \wedge 1$$

by (38) and (39). Since $V_\lambda 1 = 0$, (40) and (41) yield

(42) $$|V_\lambda h_A(w)| \leq \lambda^{-1} \wedge 1$$

for all $w \in Z^+$ and $A \subset Z^+$. The bound (42) is sharp in the limit as λ approaches 0 or ∞, at least if one takes a supremum over w.

Finally, from (27) and (42) we obtain, for any $A \subset Z^+$ and $\lambda \in R^+$

(43) $$|P\{W \in A\} - \mathcal{P}_\lambda h_A| \leq (\sum_{i=1}^{n} p_i^2)(\lambda^{-1} \wedge 1).$$

Recall that W is the number of successes in n independent trials with probabilities $p_1,\ldots,p_n$ and λ is given by (21). We can also use (30) to obtain a more accurate, but of course more complicated, approximation. From (40) it follows easily that if we define

(44) $$||h||_\infty = \sup_{w \in Z^+} |h(w)|,$$

then, for all $w \in Z^+$,

(45) $$|V_\lambda h(w)| \leq 2||h||_\infty(\lambda^{-1} \wedge 1).$$

Using (45) with h replaced by $V_\lambda h_A$ and λ replaced by λ-p we obtain, also using (42),

(46) $$|V_{\lambda-p}V_\lambda h_A(w)| \leq 2(\lambda^{-1} \wedge 1)((\lambda-p)^{-1} \wedge 1)$$

and thus, by (30)

(47) $$|P\{W \in A\} - [\mathcal{P}_\lambda h_A + \sum_{j=1}^{n} p_j^2 \, \mathcal{P}_{\lambda-p_j} V_\lambda h_A]$$
$$\leq 2(\sum p_j^2)^2(\lambda^{-1} \wedge 1)((\lambda-1)^{-1} \wedge 1).$$

In order to apply this we would have to study $\mathcal{P}_{\lambda-p}V_\lambda h_A$, or at least to be able to compute it. It may also be desirable to improve the bound (46).

I shall close this lecture with a brief treatment of the general problem of Poisson approximation for the distribution of a sum of random variables, not necessarily independent, taking on only the values zero and one. Let $X_1,\ldots,X_n$ be such random variables, let

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

and let I be uniformly distributed over $1,\ldots,n$ independent of $X_1,\ldots,X_n$. We shall see that, roughly speaking, for fixed λ = EW, the random variable W has approximately a Poisson distribution with parameter λ if the conditional distribution of W given $X_I=1$ is nearly the same as the unconditional distribution of W+1. The converse is almost true, but it requires an additional condition, for example that the variance of W is not too large.

Instead of following the approach used to derive (25), I shall formulate the argument in the manner of Chen (1975a). Let

$$\lambda = EW = E\sum_{i=1}^{n} X_i = nEX_I. \tag{49}$$

Then, for bounded $f:Z^+ \to R$,

$$\begin{aligned} EW\ f(W) &= \sum_{i=1}^{n} EX_i f(W) = n\ E\ X_I f(W) \\ &= n\ E\ X_I E[f(W)|X_I=1] = \lambda\ E\ [f(W)|X_I=1] \\ &= \lambda\ E\ f(W+1) + \lambda\{E[f(W)|X_I=1] - Ef(W+1)\}. \end{aligned} \tag{50}$$

For arbitrary bounded $h:Z^+ \to R$, substitute $f = U_\lambda h$ in (50) to obtain

$$Eh(W) = \mathcal{P}_\lambda h + \lambda\{E(U_\lambda h)(W+1) - E[(U_\lambda h)(W)|X_I=1]\}. \tag{51}$$

Theorem 2: There exist two functions $\alpha,\beta:(0,\infty)^2 \to (0,\infty)$ such that, for any W, $\mathcal{X}$, λ, and I as above, if we define

$$\varepsilon = \sup_{A\subset Z^+}[P\{W\in A\} - \mathcal{P}_\lambda h_A] \tag{52}$$

and

$$\delta = \sup_{A\subset Z^+}[P\{W+1\in A\} - P\{W\in A|X_I=1\}]. \tag{53}$$

where h_A is defined by (32), then,

(i) for any $\varepsilon' > 0$, if $\delta < \alpha(\varepsilon',\lambda)$ then $\varepsilon < \varepsilon'$, and

(ii) for any $\delta' > 0$, if $\varepsilon < \beta(\delta',\lambda)$ and

$$EW^2 \le 2(1+\lambda^2) = \rho, \tag{54}$$

say, then $\delta < \delta'$.

Proof of (i): This follows easily from (51) and the boundedness of the operator U_λ (in the space of bounded functions on Z^+ to R with the bound as norm), which was proved in (11). More precisely, from (51) with $h=h_A$ we obtain

(55) $$\varepsilon = \sup_{A\subset Z^+}[P\{W\in A\}-\mathcal{P}_\lambda h_A]$$

$$\leq (\sup_{A,w}(U_\lambda h_A((w))\sup_{A\subset Z^+}[P\{W+1\in A\}-P\{W\in A|X_I=1\}]$$

$$\leq 2\lambda e^\lambda \delta,$$

so that we can take

(56) $$\alpha(\varepsilon',\lambda) = \frac{1}{2}\lambda^{-1}e^{-\lambda}\varepsilon'.$$

<u>Proof of (ii)</u>: Our aim is to show that, subject to (54), if δ, defined by (53), is not small, then ε, defined by (52), is not small either. For this purpose substitute $f=h_{\{w_0+1,w_0+2,\ldots\}}$ in the next to last form of (50) to obtain

(57) $$P\{W \geq w_0|X_I=1\} = \frac{1}{\lambda} EW \mathcal{A}\{W \geq w_0\} = \tau,$$

say. Later I shall use (54) to choose w_0 appropriately. Then for the maximizing A in (53)

(58) $$|P\{W+1\in A \cap \{0,\ldots,w_0-1\}\}- P\{W\in A \cap \{0,\ldots,w_0-1\}|X_I=1\}|$$

$$\geq |P\{W+1\in A\} - P\{W\in A|X_I=1\}| - P\{W+1 \geq w_0\} - P\{W \geq w_0|X_I=1\}$$

$$\geq \delta - \frac{\lambda}{w_0-1} - \tau.$$

Thus there exists $w_1 \leq w_0-1$ such that

(59) $$|P\{W+1=w_1\} - P\{W=w_1|X_I=1\}| \geq \frac{1}{w_0}(\delta - \frac{\lambda}{w_0-1} - \tau).$$

Substitute $f=h_{w_1}$ in (50) to obtain

(60) $$\frac{1}{w_0}(\delta - \frac{\lambda}{w_0-1} - \tau) \leq |E\frac{W}{\lambda}h_{w_1}(W) - Eh_{w_1}(W+1)|$$

$$= |\frac{w_1}{\lambda}P\{W=w_1\} - P\{W=w_1-1\}|.$$

Now if W' is a random variable having a Poisson distribution with parameter λ,

(61) $$\frac{w_1}{\lambda}P\{W'=w_1\} - P\{W'=w_1-1\} = 0.$$

It follows from (60) and (61) that either

(62) $$\frac{w_1}{\lambda}\,|P\{W=w_1\} - P\{W'=w_1\}| \geq \frac{1}{2w_0}\left(\delta - \frac{\lambda}{w_0-1} - \tau\right)$$

or

(63) $$|P\{W=w_1-1\} - P\{W'=w_1-1\}| \geq \frac{1}{2w_0}\left(\delta - \frac{\lambda}{w_0-1} - \tau\right).$$

In either case

(64) $$\varepsilon = \sup_{A\subset Z^+}|P\{W\in A\} - \mathcal{P}_\lambda h_A| \geq \frac{1}{2w_0(1+\frac{w_0}{\lambda})}\left(\delta - \frac{\lambda}{w_0-1} - \tau\right).$$

In order to prove (ii) it remains to use (54) to make an appropriate choice of w_0 for use in (64). By Markov's inequality

(65) $$\tau = P\{W \geq w_0|X_I=1\} \leq \frac{E[W|X_I=1]}{w_0} = \frac{EW^2}{\lambda w_0} \leq \frac{\rho}{w_0} = \frac{2(1+\lambda^2)}{\lambda w_0}.$$

In order to have $\tau \leq \frac{\delta}{2}$, choose w_0 to be an integer not less than $\frac{2\rho}{\lambda\delta}$. Then

(66) $$\varepsilon \geq \frac{\delta}{2\cdot\frac{2\rho}{\lambda\delta}\left(1+\frac{2\rho}{\lambda^2\delta}\right)} = \frac{\lambda^3\delta^2}{8(1+\lambda^2)(\lambda^2\delta+4+4\lambda^2)}.$$

It is now clear that, for fixed λ, if ε is small, δ is also small.

Of course the theorem and its proof leave much to be desired. The bound used for $U_\lambda h_A$ in the proof of (i) is absurdly large in view of (42). The quantitative result (66) in (ii) is even further from being sharp. It would be desirable to bring the upper and lower bounds for ε in terms of δ closer together and to clarify the dependence on λ.

The way in which (51) is commonly applied, as in Chen (1975a), Barbour and Eagleson (1983) and Lectures XII and XIII, is essentially an instance of the method of coupling, although I am not sure this is clearly stated in any of these references. On the same probability space we construct random variables W and W^* in such a way that

(i) as indicated by the notation, the distribution of this W is the same as that of W in (48).

(ii) the distribution of W^* is the same as the conditional distribution of W given $X_I=1$,

(iii) roughly speaking, W and W^* should differ by as little as possible. Then (51) with $h=h_A$ implies

(67)
$$\begin{aligned}&|P\{W\in A\} - \mathcal{P}_\lambda h_A| \\ &= \lambda|E(U_\lambda h_A)(W+1) - E[(U_\lambda h_A)(W)|X_I=1]| \\ &= \lambda|E(U_\lambda h_A)(W+1) - E(U_\lambda h_A)(W^*)| \\ &\leq \lambda(\sup_A V_\lambda h_A)E|W+1-W^*|.\end{aligned}$$

By a more careful analysis of the transition from W to W^* it should be possible, in many cases, to obtain improvements on the Poisson approximation analogous to (47) (or better, some of the results in Chen (1975a)) in the independent case.

In this lecture I have set up the formalism for Poisson approximation and applied it to the distribution of the number of occurrences of independent rare events. In Lemma 1, a Poisson random variable W with parameter λ was characterized by the property that, for all bounded f, $E[\lambda f(W+1)-Wf(W)] = 0$. This led to the specialization of the lower line of diagram (I.28) starting with (14).

Then the special case where W is the number of occurrences of independent rare events was studied, leading to the identity (27). This led to the bound (43) for the error in the Poisson approximation for the distribution of W. I believe this was first obtained by Le Cam (1960). The lecture continued with a rough treatment of an improved approximation to the distribution of W. Finally, in Theorem 2 and the remarks below it, I discussed the general subject of Poisson approximation for dependent trials. Other applications of the Poisson approximation to special problems concerning dependent events will be given in Lectures XII and XIII.