The Annals of Probability
1974, Vol. 2, No. 2, 256–269

THE POISSON APPROXIMATION FOR DEPENDENT EVENTS[1]

BY DAVID FREEDMAN

University of California, Berkeley

Consider a sequence of dependent events, where each has uniformly small conditional probability given the past, and the sum of the conditional probabilities is approximately constant at a. Then the number of events which occur is approximately Poisson with parameter a. An explicit bound is given on the variation distance.

1. Introduction. Suppose you have a sequence of dependent events, such that each event has small conditional probability given the past, and the sum of the conditional probabilities is approximately constant at a. Then the number of the events which occur is approximately Poisson with parameter a.

First, a review of the well-known case of independence.

(1) DEFINITION. If μ and μ' are two probabilities on $(\mathscr{X}, \Sigma)$, then

$$d(\mu, \mu') = \sup_{A \in \Sigma} |\mu(A) - \mu'(A)| .$$

If Y and Y' have distribution μ and μ', then $d(Y, Y') = d(\mu, \mu')$.

(2) FACTS. (a) The function d metrizes the space of probabilities on $(\mathscr{X}, \Sigma)$.
(b) If Σ is the σ-field of all subsets of the countable set $\mathscr{X}$, then

$$d(\mu, \mu') = \tfrac{1}{2} \sum_{x \in \mathscr{X}} |\mu(x) - \mu'(x)| .$$

Call d the *variation metric*. In fact,

$$2d(\mu, \mu') = ||\mu - \mu'||$$

is the total mass of the finite signed measure $\mu - \mu'$, because $\mu(\mathscr{X}) = \mu'(\mathscr{X}) = 1$.

Next, a calculus estimate. This introduces two constants K_1 and K_2 which appear in the main estimates (4) and (5), the easy proof is omitted.

(3) LEMMA. *Let* $0 < q < 1$. *Suppose* $0 \leqq p \leqq q$.

(a) $p \leqq -\log(1-p) \leqq p + \frac{1}{2}p^2/(1-p)$;
(b) $-\log(1-p) \leqq K_1(q)p$, *where*

$$K_1(q) = q^{-1}[-\log(1-q)] = 1 + \frac{q}{2} + \frac{q^2}{3} + \cdots .$$

Received September 11, 1972; revised May 24, 1973.

[1] Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, USAF, under Grant AFOSR-71-2100A. This manuscript is submitted for publication with the understanding that the United States government is authorized to reproduce and distribute reprints for governmental purposes.

AMS 1970 *subject classifications*. Primary 60F05.

Key words and phrases. Poisson approximation, stopping time, invariance principle.

(c) $-\log(1-p) \leqq p + K_2(q)p^2$, *where*

$$K_2(q) = q^{-2}[-\log(1-q) - q] = \tfrac{1}{2} + \frac{q}{3} + \frac{q^2}{4} + \cdots .$$

(d) $K_1(\tfrac{1}{10}) \leqq \tfrac{19}{18}$ *and* $K_2(\tfrac{1}{10}) \leqq \tfrac{5}{9}$.

Here is an estimate for the independent case. In a sense, exactly the same estimate holds for the dependent case: a precise statement is in (5) below. As usual, 1_A is 1 on A and 0 off A.

(4) THEOREM. *Let $A_1, \cdots, A_n$ be independent events. Let $N = \sum_1^n 1_{A_i}$ be the number which occur. Let*

$$a = \textstyle\sum_1^n P(A_1) \qquad \textit{and} \qquad \varepsilon = \sum_1^n P(A_i)^2 .$$

Suppose $\varepsilon < 1$. Let N^* be Poisson with parameter a. Define K_1, K_2 by (3). *Let*

$$\alpha = \tfrac{1}{2}K_1(\varepsilon^{\frac{1}{2}})^2 + K_2(\varepsilon^{\frac{1}{2}}) = 1 + \tfrac{5}{6}\varepsilon^{\frac{1}{2}} + \cdots .$$

Then

$$d(N, N^*) \leqq \alpha\varepsilon .$$

If $\varepsilon \leqq \frac{1}{100}$, then $\alpha \leqq \frac{9}{8}$.

I learned this inequality from Hodges and LeCam (1960), and will review their ingenious proof later, in Section 2, because most of it works in the dependent case. Of course,

$$\textstyle\sum_1^n P(A_i)^2 \leqq [\max_i P(A_i)] \cdot \sum_1^n P(A_i) .$$

If $\sum_1^n P(A_i) = a$ is moderate, and the $P(A_i)$ are small, then $\sum_1^n P(A_i)^2$ is small. However, $\sum_1^n P(A_i)^2$ can be small even for large $[\max_i P(A_i)] \cdot \sum_1^n P(A_i)$. So a bound in terms of $\sum_1^n P(A_i)^2$ is superior to one in terms of $\max_i P(A_i)$ and $\sum_1^n P(A_i)$.

Now suppose that $A_1, A_2, \cdots$ are dependent events. Let $\mathscr{F}_i$ be the field generated by $A_1, \cdots, A_i$, and let

$$p_i = P(A_i \mid \mathscr{F}_{i-1}) ,$$

so p_i is an $\mathscr{F}_{i-1}$-measurable random variable and $0 \leqq p_i \leqq 1$. To begin with, suppose the p_i are individually small, but have a large sum: say

$$p_i \leqq \eta \quad \text{everywhere} \quad \text{and} \quad \textstyle\sum p_i \geqq T .$$

Let $a \leqq T$. Let $\tau(a)$ be the least n with $p_1 + \cdots + p_n \geqq a$, so $p_1 + \cdots + p_{\tau(a)}$ is nearly a and $p_1^2 + \cdots + p_{\tau(a)}^2$ is small. Explicitly,

$$a \leqq p_1 + \cdots + p_{\tau(a)} \leqq a + \eta$$
$$p_1^2 + \cdots + p_{\tau(a)}^2 \leqq \eta(p_1 + \cdots + p_{\tau(a)}) \leqq \eta(a + \eta) .$$

Let N be the number of A_i which occur with $i \leqq \tau(a)$: formally, $N = \sum_1^{\tau(a)} 1_{A_i}$. Then N is approximately Poisson with parameter a. I will now give an explicit bound in a more general situation.

By definition, τ is a *stopping time* relative to $\{\mathscr{F}_i\}$ provided: τ takes the values $1, 2, \cdots, \infty$; and $\{\tau = i\} \in \mathscr{F}_i$ for all $i = 1, 2, \cdots$. So $\tau(a)$ is a stopping time.

(5) THEOREM. *Let $A_1, A_2, \cdots$ be dependent events. Let $\mathscr{F}_i$ be the field generated by $A_1, \cdots, A_i$, and let*

$$p_i = P(A_i \mid \mathscr{F}_{i-1}) .$$

Let τ be a stopping time relative to $\{\mathscr{F}_i\}$. Let N be the number of A_i which occur with $i \leqq \tau$:

$$N = \sum_{i=1}^{\tau} 1_{A_i} .$$

Let $a \leqq b$ be nonnegative real numbers. Let ε, δ be nonnegative real numbers less than 1. *Suppose*

$$P\{a \leqq p_1 + \cdots + p_\tau \leqq b \text{ and } p_1^2 + \cdots + p_\tau^2 \leqq \varepsilon\} \geqq 1 - \delta .$$

Let N^* be Poisson with parameter a. Define α as in (4).

Then

$$d(N, N^*) \leqq \alpha\varepsilon + (b - a) + 2\delta .$$

I will prove this in Section 3. If $A_1, A_2, \cdots$ are independent, then $p_i = P(A_i)$. If also τ is constant at n, you can choose $a = \sum_1^n p_i$ and $\varepsilon = \sum_1^n p_i^2$ and $\delta = 0$. So (5) includes (4).

NOTATION. Throughout this paper, $(\Omega, \mathscr{F}, P)$ is a probability triple; $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}$ are σ-fields; $A_i \in \mathscr{F}_i$ and $W_i = 1_{A_i}$ is 1 on A_i and 0 off A_i, while $p_i = P(A_i \mid \mathscr{F}_{i-1}) = P\{W_i = 1 \mid \mathscr{F}_{i-1}\}$. Except in Section 5, $\mathscr{F}_i$ is the field generated by $A_1, \cdots, A_i$; so $\mathscr{F}_0 = \{\phi, \Omega\}$.

Sometimes, it is useful to have information about the whole trajectory of partial sums $W_1, W_1 + W_2, \cdots$. If you think of W_1 as having time coordinate p_1, $W_1 + W_2$ as having time coordinate $p_1 + p_2, \cdots$ and in general $W_1 + W_2 + \cdots + W_n$ as having time coordinate $p_1 + p_2 + \cdots + p_n$, the whole process is approximately Poisson, provided the p_i are small and Σp_i is large, as will be shown in Section 4.

Sometimes, it is convenient to have $\mathscr{F}_i$ larger than the field generated by $A_1, \cdots, A_i$; this will be discussed in Section 5.

There is a beautiful result of Aryeh Dvoretzky (unpublished), which extends the general central limit theorem for triangular arrays to dependent variables. It seems that the Normal approximation and Poisson approximation have quite a wide domain, by comparison with the compound Poisson approximation. To illustrate this point, consider dependent variables W_i which take the three values 0 and ± 1. Let $\mathscr{F}_i$ be the σ-field generated by $W_1, \cdots, W_i$. Let

$$p_i = P\{W_1 = 1 \mid \mathscr{F}_{i-1}\} \quad \text{and} \quad q_i = P\{W_i = -1 \mid \mathscr{F}_{i-1}\} .$$

Suppose τ is a stopping time such that

$$\begin{aligned} &p_1 + \cdots + p_\tau \quad \text{is nearly constant at} \quad c\,, \\ &q_1 + \cdots + q_\tau \quad \text{is nearly constant at} \quad d \quad \text{and} \\ &p_1, \cdots, p_\tau\,, \quad q_1, \cdots, q_\tau \quad \text{are small}\,. \end{aligned}$$

Then $W_1 + \cdots + W_\tau$ is nearly compound Poisson with canonical measure c at 1 and d at -1. However, there are usually no such stopping times: if τ makes $p_1 + \cdots + p_\tau$ nearly constant, it will usually make $q_1 + \cdots + q_\tau$ extremely variable, and there is nothing to say.

Here is a sample of the pathology. I will construct two very different processes with $p_1 + \cdots + p_\tau = 1$ and the same $q_1 + \cdots + q_\tau$. One will have $W_1 + \cdots + W_\tau \geqq -1$ everywhere, but equal to -1 with positive probability, even in the limit when $p \to 0$. This prevents $W_1 + \cdots + W_\tau$ from being approximately compound Poisson. The other process will haue $W_1 + \cdots + W_\tau$ unbounded above and below.

The first construction. Initially, let

$$p_i = P\{W_i = 1 \mid \mathscr{F}_{i-1}\} = p \qquad \text{and} \qquad q_i = P\{W_i = -1 \mid \mathscr{F}_{i-1}\} = p\,.$$

So $W_1, W_2, \cdots$ are independent, ± 1 with equal probability p, and 0 with the remaining probability $1 - 2p$. Keep this up until the first n with $W_1 + \cdots + W_n = -1$. From that n on, let

$$p_i = P\{W_i = 1 \mid \mathscr{F}_{i-1}\} = p \qquad \text{and} \qquad q_i = P\{W_i = -1 \mid \mathscr{F}_{i-1}\} = 0\,.$$

Let τ be the least m with $p_1 + \cdots + p_m = 1$.

The second construction is like the first, except you change from the 3-valued to the 2-valued W's at the first n with $W_1 + \cdots + W_n = 1$.

2. The independent case.

(6) LEMMA. *If Y and Y' are measurable mappings from $(\Omega, \mathscr{F}, P)$ to $(\mathscr{X}, \Sigma)$, then*

$$d(Y, Y') \leqq P_*\{Y \neq Y'\}\,,$$

where P_* is inner measure.

PROOF. Let $A \in \Sigma$. Then

$$|P(Y \in A) - P(Y' \in A)| \leqq P\{(Y \in A) \triangle (Y' \in A)\}\,,$$

where $\triangle$ is symmetric difference. But $(Y \in A) \triangle (Y' \in A)$ is an $\mathscr{F}$-measurable subset of $\{Y \neq Y'\}$. □

(7) LEMMA. *If Y is Poisson with parameter λ, then*

(a) $P\{Y \geqq 1\} \leqq \lambda$
(b) $P\{Y \geqq 2\} \leqq \frac{1}{2}\lambda^2$.

PROOF. *Claim* (a). Observe $P\{Y = 0\} = e^{-\lambda} \geqq 1 - \lambda$.
Claim (b). Check that

$$\begin{aligned} P\{Y \geqq 2\} &= e^{-\lambda} \textstyle\sum_{j=2}^\infty \dfrac{\lambda^j}{j!} \\ &= \tfrac{1}{2}\lambda^2 e^{-\lambda} \textstyle\sum_{j=2}^\infty \dfrac{2\lambda^{j-2}}{j!} \leqq \tfrac{1}{2}\lambda^2\,, \end{aligned}$$

because $2/j! \leqq 1/(j-2)!$ for $j \geqq 2$. □

(8) LEMMA. *Suppose Y is Poisson with parameter a, and Y' is Poisson with parameter $a' > a$. Then*

$$d(Y, Y') \leqq a' - a .$$

PROOF. Let Z be independent of Y and Poisson with parameter $a' - a$. Then $Y + Z$ is distributed like Y'. With the help of (6) and (7a),

$$d(Y, Y') = d(Y, Y + Z) \leqq P(Z \neq 0) \leqq (a' - a) . \qquad \square$$

PROOF OF (4). Construct independent Poisson variables $Y_1, \cdots, Y_n$ on some triple $(\Omega', \mathscr{F}', P')$, where Y_i has parameter

$$\lambda_i = -\log(1 - p_i) \qquad \text{and} \qquad p_i = P(A_i) .$$

Then let $X_i = 0$ when $Y_i = 0$, and $X_i = 1$ when $Y_i \geqq 1$. Now $X_1, \cdots, X_n$ are independent 0–1 variables, and

$$P'(X_i = 0) = P'(Y_i = 0) = e^{-\lambda_i} = 1 - p_i ,$$

so $P'(X_i = 1) = P(A_i)$. That is, $X_1 + \cdots + X_n$ is distributed like $N = 1_{A_1} + \cdots + 1_{A_n}$. So

(9) $$d(N, N^*) = d(X_1 + \cdots + X_n, N^*) .$$

But $X_1 + \cdots + X_n = Y_1 + \cdots + Y_n$ except on $\bigcup_{i=1}^n \{Y_i \geqq 2\}$. Let $k_1 = K_1(\varepsilon^{\frac{1}{2}})$. Check

(10) $$\begin{aligned} d(X_1 + \cdots + X_n, Y_1 + \cdots + Y_n) &\leqq \textstyle\sum_{i=1}^n P\{Y_i \geqq 2\} && \text{by (6)} \\ &\leqq \tfrac{1}{2}\textstyle\sum_{i=1}^n \lambda_i^2 && \text{by (7b)} \\ &\leqq \tfrac{1}{2}k_1^2\varepsilon && \text{by (3b)} , \end{aligned}$$

where $\varepsilon = \sum_{i=1}^n p_i^2$. But $Y_1 + \cdots + Y_n$ is Poisson with parameter $\lambda_1 + \cdots + \lambda_n$. Let $k_2 = K_2(\varepsilon^{\frac{1}{2}})$. As (3a/c) imply, $p_i \leqq \lambda_i \leqq p_i + k_2 p_i^2$, so

$$a = \textstyle\sum_{i=1}^n p_i \leqq \sum_{i=1}^n \lambda_i \leqq a + k_2\varepsilon .$$

Now N^* is Poisson with parameter a, so (8) shows

(11) $$d(N^*, Y_1 + \cdots + Y_n) \leqq k_2\varepsilon.$$

Combine (9), (10) and (11) by (2a), to see

$$d(N, N^*) \leqq \tfrac{1}{2}k_1^2\varepsilon + k_2\varepsilon = \alpha\varepsilon .$$

The bound for α can be established this way. If $\varepsilon \leqq \frac{1}{100}$, then $\varepsilon^{\frac{1}{2}} \leqq \frac{1}{10}$. So (3d) shows

$$\alpha \leqq \tfrac{1}{2}(\tfrac{19}{18})^2 + \tfrac{5}{9} \leqq \tfrac{9}{8} .$$

3. The dependent case. Here is the leading special case of (5).

(12) PROPOSITION. *Inequality* (5) *holds provided τ is constant, say at n, and $\delta = 0$, so*

(13) $$a \leqq p_1 + \cdots + p_n \leqq b \qquad \text{and} \qquad p_1^2 + \cdots + p_n^2 \leqq \varepsilon \quad \text{everywhere.}$$

PROOF. On some triple $(\Omega', \mathscr{F}', P')$, construct a unit Poisson process Y. That is, for all $\omega \in \Omega'$,

(14a) $$Y(0, \omega) = 0\,.$$

(14b) $Y(\cdot, \omega)$ is a right-continuous, non-decreasing, integer-valued step function, with jumps of height 1, and only finitely many jumps in finite time intervals.

Let $\Sigma(t)$ be the σ-field spanned by $Y(s)$ for $s \leqq t$. Then Y has the main defining property

(15) $P'\{Y(t+u) - Y(t) = j \mid G\} = e^{-u}u^j/j!$ for all nonnegative t and u, nonnegative integer j, and all sets $G \in \Sigma(t)$.

Remember $W_i = 1_{A_i}$, so $N = W_1 + \cdots + W_n$. The partial sums $W_1, W_1 + W_2, \cdots, W_1 + \cdots + W_n$ can be replicated by looking at Y at the random times $\Lambda_1, \Lambda_2, \cdots, \Lambda_n$, as follows. Let $\phi(j)$ be 0 when $j = 0$, and 1 when $j \geqq 1$. For convenience, let $\Lambda_0 = 0$. Let $p_1 = P(A_1)$ and $\lambda_1 = -\log(1 - p_1)$ and $\Lambda_1 = \lambda_1$ and $X_1 = \phi[Y(\Lambda_1)]$. So

$$P'\{X_1 = 0\} = e^{-\lambda_1} = 1 - p_1 = P\{W_1 = 0\}\,,$$

and X_1 is distributed like W_1. Furthermore, $\{X_1 = x_1\} \in \Sigma(\lambda_1)$, where $x_1, \cdots, x_n$ are 0 or 1. Proceeding inductively, let $1 \leqq m < n$. Let

$$\begin{aligned} p_{m+1}(x_1, \cdots, x_m) &= P\{W_{m+1} = 1 \mid W_1 = x_1, \cdots, W_m = x_m\} \\ \lambda_{m+1}(x_1, \cdots, x_m) &= -\log[1 - p_{m+1}(x_1, \cdots, x_m)] \\ \Lambda_{m+1}(x_1, \cdots, x_m) &= \lambda_1 + \lambda_2(x_1) + \cdots + \lambda_{m+1}(x_1, \cdots, x_m) \\ \Lambda_{m+1} &= \Lambda_{m+1}(X_1, \cdots, X_m) \\ X_{m+1} &= \phi[Y(\Lambda_{m+1}) - Y(\Lambda_m)]\,. \end{aligned}$$

Make the inductive assumption that $\{X_1 = x_1, \cdots, X_m = x_m\} \in \Sigma(\Lambda_m(x_1, \cdots, x_{m-1}))$. So (15) shows:

(16) Given $\{X_1 = x_1, \cdots, X_m = x_m\}$, the difference $Y(\Lambda_{m+1}) - Y(\Lambda_m)$ is conditionally Poisson with parameter $\lambda_{m+1}(x_1, \cdots, x_m)$.

In particular,

$$\begin{aligned} P'\{X_{m+1} = 0 \mid X_1 = x_1, \cdots, X_m = x_m\} &= \exp[-\lambda_{m+1}(x_1, \cdots, x_m)] \\ &= 1 - p_{m+1}(x_1, \cdots, x_m) \\ &= P\{W_{m+1} = 0 \mid W_1 = x_1, \cdots, x_m\}\,. \end{aligned}$$

Make the inductive assumption that $(X_1, \cdots, X_m)$ is distributed like $(W_1, \cdots, W_m)$; so $(X_1, \cdots, X_{m+1})$ is distributed like $(W_1, \cdots, W_m, W_{m+1})$. For the next move, confirm that $\{X_1 = x_1, \cdots, X_{m+1} = x_{m+1}\} \in \Sigma(\Lambda_{m+1}(x_1, \cdots, x_m))$. When the induction is over, $(X_1, \cdots, X_n)$ is distributed like $(W_1, \cdots, W_n)$, so

(17) $N = W_1 + \cdots + W_n$ is distributed like $(X_1, \cdots, X_n)$, and

$$d(N, N^*) = d(X_1 + \cdots + X_n, N^*)\,.$$

This completes the embedding. So $\Lambda_0 = 0$ and $\Lambda_1 = \lambda_1$ are constant, but $\Lambda_2 = \lambda_1 + \lambda_2(X_1)$ is random; as are $\Lambda_3, \cdots, \Lambda_n$. Also, $X_1 = \phi[Y(\Lambda_1) - Y(\Lambda_0)]$, by (14a).

For the next step, let

$$(18)\qquad C_m = \{Y(\Lambda_m) - Y(\Lambda_{m-1}) \geqq 2\} \quad \text{and} \quad C = \bigcup_{m=1}^n C_m\,.$$

Check that

$$X_m = Y(\Lambda_m) - Y(\Lambda_{m-1}) \quad \text{off} \quad C_m\,,$$

so

$$(19)\qquad X_1 + \cdots + X_m = Y(\Lambda_m) \quad \text{for} \quad m = 1, \cdots, n \quad \text{off} \quad C\,.$$

I must now estimate $P'(C)$. Relation (16) implies that $P'\{C_m \mid X_1 = x_1, \cdots, X_{m-1} = x_{m-1}\}$ is the probability that a Poisson variable with parameter $\lambda_m(x_1, \cdots, x_{m-1})$ is 2 or more. So (7b) shows

$$P'\{C_m \mid X_1 = x_1, \cdots, X_{m-1} = x_{m-1}\} \leqq \tfrac{1}{2}\lambda_m(x_1, \cdots, x_{m-1})^2\,.$$

Use (3b) with $k_1 = K_1(\varepsilon^{\frac{1}{2}})$, and assumption (13):

$$P'\{C_m \mid X_1 = x_1, \cdots, X_{m-1} = x_{m-1}\} \leqq \tfrac{1}{2}k_1^2 p_m(x_1, \cdots, x_{m-1})^2\,,$$

and

$$P'\{C_m\} \leqq \tfrac{1}{2}k_1^2 E\{p_m(x_1, \cdots, x_{m-1})^2\}\,.$$

But $P'(C) \leqq \sum_{m=1}^n P'(C_m)$, so

$$(20)\qquad P'(C) \leqq \tfrac{1}{2}k_1^2 E\{\textstyle\sum_{m=1}^n p_m(x_1, \cdots, x_{m-1})^2\}\,.$$

Use assumption (13) again:

$$(21)\qquad P'(C) \leqq \tfrac{1}{2}k_1^2\varepsilon\,.$$

Use (19) with $m = n$:

$$P'\{X_1 + \cdots + X_n \neq Y(\Lambda_n)\} \leqq P'(C) \leqq \tfrac{1}{2}k_1^2\varepsilon\,.$$

So (6) shows

$$(22)\qquad d(X_1 + \cdots + X_n, Y(\Lambda_n)) \leqq P'(C) \leqq \tfrac{1}{2}k_1^2\varepsilon\,.$$

Remember that N^* is Poisson with parameter a. The next job is to estimate $d(Y(\Lambda_n), N^*) = d(Y(\Lambda_n), Y(a))$. Relations (3a/c) with $k_2 = K_2(\varepsilon^{\frac{1}{2}})$, and assumption (12), shows

$$p_m(x_1, \cdots, x_{m-1}) \leqq \lambda_m(x_1, \cdots, x_{m-1}) \leqq p_m(x_1, \cdots, x_{m-1}) + k_2 p_m(x_1, \cdots, x_{m-1})^2\,;$$

so assumption (13) makes

$$a \leqq \Lambda_n \leqq b + k_2\varepsilon \quad \text{everywhere}\,.$$

Now (14b) shows $\{Y(\Lambda_n) \neq Y(a)\} \subset \{Y(b + k_2\varepsilon) - Y(a) \geqq 1\}$. Use (15) and (7a):

$$\begin{aligned} P'\{Y(\Lambda_n\} \neq Y(a)\} &\leqq P'\{Y(b + k_2\varepsilon) - Y(a) \geqq 1\} \\ &\leqq k_2\varepsilon + b - a\,. \end{aligned}$$

Now (6) shows

(23) $$d[Y(\Lambda_n), N^*] \leqq k_2\varepsilon + b - a\,.$$

Combine (17), (22) and (23):

$$d(N, N^*) \leqq \tfrac{1}{2}k_1{}^2\varepsilon + k_2\varepsilon + b - a\,. \qquad \square$$

The general case of (5) can be reduced to the special case (12), as follows. First, a can be slightly decreased because the bound is continuous. Then, τ can be replaced by $\tau' = \max(\tau, n)$, so as to get a uniformly bounded stopping time. Then, τ' can be replaced by τ'', the largest $m \leqq \tau'$ with $p_1 + \cdots + p_m \leqq b$ and $p_1^2 + \cdots + p_m^2 \leqq \varepsilon$. Then, b and ε can be increased slightly and the game continued with (say) tossing an η coin after time τ'', so as to build Σp_i up to a. When all is said and done, the new events will satisfy (13) and will not differ appreciably from the original events.

Here is a preliminary.

(24) LEMMA. *Let $A_1, A_2, \cdots, A_n$ be events. Let $\mathscr{F}_i$ be the field generated by $A_1, \cdots, A_i$. Let $\tau \leqq n$ be a stopping time relative to $\{\mathscr{F}_i\}$. For $i = 1, \cdots, n$, let $A_i{}^* = A_i \cap \{\tau \geqq i\}$ and let $\mathscr{F}_i{}^*$ be the field generated by $A_1{}^*, \cdots, A_i{}^*$. Then τ is a stopping time relative to $\{\mathscr{F}_i{}^*\}$, and*

$$\begin{aligned} P\{A_i{}^* \mid \mathscr{F}_{i-1}^*\} &= P\{A_i \mid \mathscr{F}_{i-1}\} && \textit{on} \quad \{\tau \geqq i\} \\ &= 0 && \textit{on} \quad \{\tau < i\}\,. \end{aligned}$$

Finally, $\sum_1^\tau 1_{A_i} = \sum_1^n 1_{A_i{}^*}$.

PROOF. To make it formal, I will argue by induction that

(25) if $A \in \mathscr{F}_i$, then $A \cap \{\tau \geqq i\} \in \mathscr{F}_i{}^*$; especially $\{\tau = i\} \in \mathscr{F}_i{}^*$.

This is clear for $i = 1$. Suppose it for $i \leqq j$. Then

$$\{\tau \geqq j + 1\} = \Omega\backslash\{\tau \leqq j\} \in \mathscr{F}_j$$

and

$$\{\tau \geqq j + 1\} \in \mathscr{F}_j{}^*\,.$$

Now the collection of $A \in \mathscr{F}_{j+1}$ with $A \cap \{\tau \geqq j + 1\} \in \mathscr{F}_j{}^*$ is plainly a field, which includes the basic sets $A_1, \cdots, A_{j+1}$, so is all of $\mathscr{F}_{j+1}$. This proves (25). With its help, you can verify that $\{\tau \geqq i\} \in \mathscr{F}_{i-1}^*$, and then (24) is both obvious and easy to verify. $\square$

The next proposition is the main step in reducing (5) to (12).

(26) PROPOSITION. *Inequality (5) holds provided τ is uniformly bounded, with 2δ reduced to δ.*

PROOF. Suppose $\tau \leqq n$ everywhere. Let σ be the largest $m \leqq \tau$ with

$$p_1 + \cdots + p_m \leqq b \qquad \text{and} \qquad p_1^2 + \cdots + p_m^2 \leqq \varepsilon\,.$$

Then $\sigma \leqq \tau$; and σ is a stopping time, because p_{m+1} is $\mathscr{F}_m$-measurable.

For $i = 1, \cdots, n$, let

$$W_i = 1_{A_i};$$
$$A_i^* = A_i \cap \{\sigma \geqq i\} \quad \text{and} \quad W_i^* = 1_{A_i^*};$$
$$\mathscr{F}_i^* \quad \text{be the } \sigma\text{-field generated by} \quad A_1^*, \cdots, A_i^*;$$
$$p_i^* = P\{A_i^* \mid \mathscr{F}_{i-1}^*\}.$$

With the help of (24), you can check

$$p_1^* + \cdots + p_n^* = p_1 + \cdots + p_\sigma \leqq b \tag{27}$$

$$p_1^{*2} + \cdots + p_n^{*2} = p_1^2 + \cdots + p_\sigma^2 \leqq \varepsilon. \tag{28}$$

Let

$$G = \{a \leqq p_1 + \cdots + p_\tau \leqq b \text{ and } p_1^2 + \cdots + p_\tau^2 \leqq \varepsilon\}. \tag{29}$$

Then

$$\sigma = \tau \quad \text{on} \quad G, \tag{30}$$

so (24) shows

$$W_1^* + \cdots + W_n^* = W_1 + \cdots + W_\sigma = W_1 + \cdots + W_\tau \quad \text{on} \quad G. \tag{31}$$

The next job is to get the sum of the conditional probabilities up to a. This is a problem if $b = a$ or if $\sum_1^n p_i^{*2} = \varepsilon$. So, introduce $b' > b$ and $\varepsilon' > \varepsilon$, but close. Choose a positive integer n' with

$$n' > \max\left\{a, \frac{a}{b' - b}, \frac{a^2}{\varepsilon' - \varepsilon}\right\}.$$

Let $\eta = a/n'$. So without having the upper bounds (27) and (28),

$$\eta < 1 \quad \text{and} \quad \eta < b' - b \quad \text{and} \quad n'\eta^2 < \varepsilon' - \varepsilon.$$

Without real loss, suppose there are independent events $A_{n+1}^*, \cdots, A_{n+n'}^*$, with common probability η, independent of $\mathscr{F}_n^*$. Let $W_1^* = 1_{A_i^*}$ and let $\mathscr{F}_i^*$ be the field generated by $A_1^*, \cdots, A_i^*$, for $i = n + 1, \cdots, n + n'$. Then

$$p_i^* = P\{A_i^* \mid \mathscr{F}_{i-1}^*\} = \eta \quad \text{for} \quad i = n + 1, \cdots, n + n'.$$

Let σ' be the least $m \geqq n$ with $p_1^* + \cdots + p_m^* \geqq a$. Then σ' is a stopping time and $\sigma' \leqq n + n'$, because $n'\eta = a$. From (27),

$$a \leqq p_1^* + \cdots + p_{\sigma'}^* \leqq a + \eta \leqq b', \tag{32}$$

because $\eta < b' - b$. From (28).

$$p_1^{*2} + \cdots + p_{\sigma'}^{*2} \leqq \varepsilon + n'\eta^2 \leqq \varepsilon', \tag{33}$$

because $n'\eta^2 < \varepsilon' - \varepsilon$. Remember G from (29). Now $\sigma = \tau$ on G by (30), so $p_1^* + \cdots + p_n^* = p_1 + \cdots + p_\sigma = p_1 + \cdots + p_\tau \geqq a$ on G by (27), and $\sigma' = n$ on G. Then (31) shows

$$\begin{aligned} W_1^* + \cdots + W_{\sigma'}^* &= W_1^* + \cdots + W_n^* \\ &= W_1 + \cdots + W_\sigma = W_1 + \cdots + W_\tau \quad \text{on} \quad G. \end{aligned} \tag{34}$$

Now let

$$A_i' = A_i^* \cap \{\sigma' \geqq i\} \qquad \text{and} \qquad W_i' = 1_{A_i'} \,;$$
$$\mathscr{F}_i' \text{ be the field generated by } A_1', \cdots, A_i' \,;$$
$$p_i' = P\{A_i' \mid \mathscr{F}_{i-1}'\}$$

for $i = 1, \cdots, n + n'$. Relations (24) and (32) and (33) show

$$a \leqq p_1' + \cdots + p_{n+n'}' \leqq b' \qquad \text{and} \qquad p_1'^2 + \cdots + p_{n+n'}'^2 \leqq \varepsilon' \,. \tag{35}$$

Similarly, (24) and (34) show

$$W_1 + \cdots + W_\tau = W_1' + \cdots + W_{n+n'}' \qquad \text{on} \quad G \,. \tag{36}$$

Proposition (12) on the W_i' shows

$$d(W_1' + \cdots + W_{n+n'}', N^*) \leqq \alpha\varepsilon' + b' - a \,,$$

where N^* is Poisson with parameter a. Condition (13) holds by (35). Remember $P(G) \geqq 1 - \delta$ by assumption. Relations (6) and (36) show

$$d(W_1 + \cdots + W_\tau, W_1' + \cdots + W_{n+n'}') \leqq \delta \,.$$

So (2a) makes

$$d(W_1 + \cdots + W_\tau, W_1' + \cdots + W_{n+n'}') \leqq \alpha\varepsilon' + b' - a + \delta \,.$$

Since $b' > b$ and $\varepsilon' > \varepsilon$ were arbitrary, inequality (5) is true for uniformly bounded τ, with 2δ reduced to δ. □

PROOF OF (5). As $n \uparrow \infty$, the stopping time $\tau_n = \min\{\tau, n\}$ increases to τ. And $p_1 + \cdots + p_\tau \leqq b$ makes $W_1 + \cdots + W_\tau$ converge a.e. So

$$p_1 + \cdots + p_{\tau_n} \uparrow p_1 + \cdots + p_\tau$$
$$p_1^2 + \cdots + p_{\tau_n} \uparrow p_1^2 + \cdots + p_\tau^2$$
$$W_1 + \cdots + W_{\tau_n} = W_1 + \cdots + W_\tau$$

for large enough random n, a.e. on the set $\{a \leqq p_1 + \cdots + p_\tau \leqq b\}$. If $\tau = \infty$, however, $p_1 + \cdots + p_{\tau_n}$ may never reach a. So let $a' < a$ and $\delta' > \delta$. Let $\eta > 0$ be small. For large enough n,

$$P\{a' \leqq p_1 + \cdots + p_{\tau_n} \leqq b \text{ and } p_1^2 + \cdots + p_{\tau_n}^2 \leqq \varepsilon\} \geqq 1 - \delta' \tag{37}$$

$$P\{W_1 + \cdots + W_{\tau_n} = W_1 + \cdots + W_\tau\} \geqq 1 - \delta - \eta \,. \tag{38}$$

Let N' be Poisson with parameter a', while N^* is Poisson with parameter a. Then (37), and (26) on τ_n, show

$$d(W_1 + \cdots + W_{\tau_n}, N') \leqq \alpha\varepsilon + b - a' + \delta' \,.$$

But (8) shows $d(N', N^*) \leqq a - a'$. Relations (38) and (7) show

$$d(W_1 + \cdots + W_\tau, W_1 + \cdots + W_{\tau_n}) \leqq \delta + \eta \,.$$

So (2a) makes

$$d(W_1 + \cdots + W_\tau, N^*) \leqq \alpha\varepsilon + b - a' + \delta' + a - a' + \delta + \eta \,.$$

Since $a' < a$ and $\delta' > \delta$ and $\eta > 0$ are free, inequality (5) is proved. □

4. The joint distribution. As usual, let $A_1, A_2, \cdots, A_n$ be events. Let $W_i = 1_{A_i}$. Let $\mathscr{F}_i$ be the field generated by $A_1, \cdots, A_i$, and let $p_i = P\{A_i \mid \mathscr{F}_{i-1}\}$. Let $\tau(t)$ be the least m if any with $p_1 + \cdots + p_m \geqq t$, and $\tau(t) = \infty$ if none. Let

$$S(t) = W_1 + \cdots + W_{\tau(t)} .$$

So S is the trajectory of the process: W_1 at time p_1, $W_1 + W_2$ at time $p_1 + p_2, \cdots, W_1 + \cdots + W_n$ at time $p_1 + \cdots + p_n, \cdots$. Let Y be a unit Poisson process, as defined by (14) and (15). Let a be a positive real number. Let ε and δ be positive real numbers less than 1. Define K_1 and K_2 by (3). Let

$$\beta(\varepsilon, a) = [K_1(\varepsilon) - 1]a + K_1(\varepsilon)\varepsilon = \tfrac{1}{2}\varepsilon a + \varepsilon + O(\varepsilon^2)$$
$$\gamma(\varepsilon, a) = \tfrac{1}{2}K_1(\varepsilon)^2\varepsilon(a + \varepsilon) = \tfrac{1}{2}\varepsilon a + O(\varepsilon^2) .$$

Let k be a positive integer. Let $0 \leqq t_1 < \cdots < t_k \leqq a$. Let

$$\mathbf{S} = (S(t_1), \cdots, S(t_k)) \qquad \text{and} \qquad \mathbf{Y} = (Y(t_1), \cdots, Y(t_k)) .$$

(39) THEOREM. *Suppose there is a stopping time τ with*

$$P\{p_1 + \cdots + p_\tau \geqq a \ \text{ and } \ \max_{i \leqq \tau} p_i \leqq \varepsilon\} \geqq 1 - \delta .$$

Then

$$d(\mathbf{S}, \mathbf{Y}) \leqq k\beta(\varepsilon, a) + \gamma(\varepsilon, a) + 2\delta .$$

If $a \geqq 1$ and $\varepsilon \leqq \frac{1}{10}$, then

$$k\beta(\varepsilon, a) + \gamma(\varepsilon, a) \leqq \tfrac{7}{3}k\varepsilon a .$$

PROOF. You can argue this pretty much like (5). There is no loss in making τ uniformly bounded and $\delta = 0$. Then, you can replace τ by $\tau(a)$, so

$$p_1 + \cdots + p_\tau \leqq a + \varepsilon .$$

Then, you can make τ constant, say at n, by (24). After these reductions, you can use the same construction as in (12). For $0 \leqq t \leqq a$, let $\sigma(t)$ be the smallest m with

$$p_1 + p_2(X_1) + \cdots + p_m(X_1, \cdots, X_{m-1}) \geqq t .$$

Let

$$T(t) = X_1 + \cdots + X_{\sigma(t)}$$
$$\mathbf{T} = (T(t_1), \cdots, T(t_k)) .$$

Since $(X_1, \cdots, X_n)$ is distributed like $(W_1, \cdots, W_n)$, you can conclude $\mathbf{T}$ is distributed like $\mathbf{S}$, and

$$d(\mathbf{S}, \mathbf{Y}) = d(\mathbf{T}, \mathbf{Y}) .$$

Let $L_t = \Lambda_{\sigma(t)}$. So

$$t \leqq p_1 + p_2(X_1) + \cdots + p_{\sigma(t)}(X_1, \cdots, X_{\sigma(t)-1}) \leqq t + \varepsilon .$$

And (3) shows

$$t \leqq L_t \leqq K_1(\varepsilon)(t + \varepsilon) \leqq t + \beta(\varepsilon, a) .$$

So

$$P\{Y(L_t) \neq Y(t)\} \leqq \beta(\varepsilon, a)$$

by (7a). Let

$$\mathbf{Y}' = (Y(L_{t_1}), \cdots, Y(L_{t_k})) .$$

So

$$P\{\mathbf{Y}' \neq \mathbf{Y}\} \leqq k\beta(\varepsilon, a) ,$$

and (6) makes

$$d(\mathbf{Y}, \mathbf{Y}') \leqq k\beta(\varepsilon, a) .$$

Remember the set C defined by (18). Use (19) to check that $\mathbf{T} = \mathbf{Y}'$ off C. So (6) shows

$$d(\mathbf{T}, \mathbf{Y}') \leqq P'(C) .$$

Use (20) with $k_1 = K_1(\varepsilon)$:

$$\begin{aligned} P'(C) &\leqq \tfrac{1}{2}K_1(\varepsilon)^2 \sup \textstyle\sum_{m=1}^{n} p_m{}^2 \\ &\leqq \tfrac{1}{2}K_1(\varepsilon)^2(\sup p_m) \cdot \sup \textstyle\sum_{m=1}^{n} p_m \\ &\leqq \tfrac{1}{2}K_1(\varepsilon)^2\varepsilon(a + \varepsilon) . \end{aligned}$$

So

$$d(\mathbf{T}, \mathbf{Y}') \leqq \gamma(\varepsilon, a) .$$

Overall,

$$\begin{aligned} d(\mathbf{S}, \mathbf{Y}) &= d(\mathbf{T}, \mathbf{Y}) \\ &\leqq d(\mathbf{T}, \mathbf{Y}') + d(\mathbf{Y}', \mathbf{Y}) \\ &\leqq \gamma(\varepsilon, a) + k\beta(\varepsilon, a) . \end{aligned}$$

The bound for $k\beta(\varepsilon, a) + \gamma(\varepsilon, a)$ can be argued this way. Relation (3a) implies

$$K_1(\varepsilon) - 1 \leqq \frac{1}{2}\frac{\varepsilon}{1-\varepsilon} \leqq \tfrac{5}{9}\varepsilon ;$$

and (3d) shows

$$K_1(\varepsilon)\varepsilon \leqq K_1(\varepsilon)\varepsilon a \leqq \tfrac{19}{18}\varepsilon a .$$

So

$$k\beta(\varepsilon, a) \leqq (\tfrac{5}{9} + \tfrac{19}{18})k\varepsilon a .$$

Furthermore

$$\begin{aligned} \gamma(\varepsilon, a) &\leqq \tfrac{1}{2}K_1(\varepsilon)^2 k\varepsilon(a + \tfrac{1}{10}a) \\ &\leqq \tfrac{1}{2} \cdot (\tfrac{19}{18})^2 \cdot \tfrac{11}{10} \cdot k\varepsilon a . \end{aligned}$$

So

$$\begin{aligned} k\beta(\varepsilon, a) + \gamma(\varepsilon, a) &\leqq (\tfrac{5}{9} + \tfrac{19}{18} + \tfrac{1}{2} \cdot (\tfrac{19}{18})^2 \cdot \tfrac{11}{10})k\varepsilon a \\ &\leqq \tfrac{7}{3}k\varepsilon a . \end{aligned}$$

□

Theorem (39) only deals with finite joint distributions. However, the Skorokhod topology on non-decreasing step-functions is weaker than the pointwise convergence topology: if you know a path on a large enough finite set of times, you know it up to small perturbations of the time scale. So (39) is stronger than the conventional invariance principle.

The only interesting functions on paths are the waiting times. Let $\sigma(m)$ be the least n if any with $W_1 + \cdots + W_n = m$, and $\sigma(m) = \infty$ if none. Let

$$\theta(m) = \textstyle\sum_1^{\sigma(m)} p_i ,$$

the (intrinsic) time to the mth occurrence. Let Y be a unit Poisson process, and let $\theta_m{}^*$ be the least t with $Y(t) = m$.

(40) COROLLARY. *Let* $0 \leqq t_1 < t_1' < t_2 < t_2' < \cdots < t_m < t_m' \leqq a$. *Under the conditions of* (37),

$$|P\{t_\nu \leqq \theta(\nu) \leqq t_\nu' \text{ for } \nu = 1, \cdots, m\} - P\{t_\nu \leqq \theta_\nu{}^* \leqq t_\nu' \text{ for } \nu = 1, \cdots, m\}|$$
$$\leqq 2m\beta(\varepsilon, a) + \gamma(\varepsilon, a) + 2\delta .$$

So the inter-occurrence times are approximately independent and exponential with parameter 1.

5. General σ-fields. Let $(\Omega, \mathscr{F}, P)$ be a probability triple. Let $A_1, A_2, \cdots$ be events, that is, elements of $\mathscr{F}$. Let $\mathscr{F}_i$ be a sub σ-field of $\mathscr{F}$, not necessarily the same generated by $A_1, \cdots, A_i$. Assume:

(41 a) $\qquad A_i \in \mathscr{F}_i \quad$ for $\quad i = 1, 2, \cdots, \qquad$ and

(41 b) $\qquad \mathscr{F}_0 \subset \mathscr{F}_1 \subset \cdots .$

Let

(42) $\qquad p_i = P\{A_i \mid \mathscr{F}_{i-1}\} .$

A *stopping time* τ relative to $\{\mathscr{F}_i\}$ takes the values $0, 1, \cdots, \infty$, and $\{\tau = i\} \in \mathscr{F}_i$ for $i = 0, 1, \cdots$.

(43) PROPOSITION. *Theorems* (5) *and* (6) *remain true when* $\{\mathscr{F}_i\}$ *satisfies* (41), *and* p_i *is defined by* (42), *and* τ *is a stopping time relative to the* $\mathscr{F}_i$.

The proof is more delicate, since the p_i may have a continuous distribution. What you should do is construct on a triple $(\Omega', \mathscr{F}', P')$ a unit Poisson process Y, and independent of Y a sequence $U_1, U_2, \cdots$ of independent uniform variables. Let $\Lambda_0 = 0$. Construct $r_1 = f_1(U_1)$ to be distributed like p_1. Let $\lambda_1 = -\log(1 - r_1)$ and $\Lambda_1 = \lambda_1$. So even Λ_1 is random. Let $X_1 = \phi[Y(\Lambda_1)]$, where $\phi(0) = 0$ and $\phi(n) = 1$ for $n \geqq 1$. By conditioning on r_1, check that (r_1, X_1) is distributed like (p_1, W_1). Construct $r_2 = f_2(r_1, X_1, U_2)$ so (r_1, X_1, r_2) is distributed like (p_1, W_1, p_2). Let $\lambda_2 = -\log(1 - r_2)$ and $\Lambda_2 = \lambda_1 + \lambda_2$. Let $X_2 = \phi[Y(\Lambda_2) - Y(\Lambda_1)]$. By conditioning on (r_1, X_1, r_2), check that (r_1, X_1, r_2, X_2) is distributed like (p_1, W_1, p_2, W_2). In fact, given $r_1 = s$ and $X_1 = x$ and $r_2 = t$, a regular conditional distribution for $Y(\Lambda_2) - Y(\Lambda_1)$ is: the conditional distribution of

$$Y[-\log(1 - s) - \log(1 - t)] - Y[-\log(1 - s)]$$

given $\phi[Y(-\log(1 - s))] = x$. The construction proceeds inductively, and the rest of the argument is the same.

REFERENCES

HODGES, J. L., JR. and LECAM, L. (1960). The Poisson approximation to the binomial distribution. *Ann. Math. Statist.* **31** 737-740.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720