LECTURE IV. THE NUMBER OF ONES IN THE BINARY EXPANSION OF A RANDOM INTEGER

Let n be a natural number and X a random variable uniformly distributed over the set $\{0,\ldots,n-1\}$. We shall see that for large n the number of ones in the binary expansion of X has approximately a binomial distribution, the distribution of the number of successes in k independent trials with probability one-half, where k is determined by

$$2^{k-1} < n \leq 2^k. \tag{1}$$

The expected value of the number of ones in this expansion was studied as a function of n by Delange (1975). In Diaconis (1977) the present problem was studied by the method of the third lecture. Here I shall give a slightly different treatment in order to emphasize the notion of approximation by the binomial distribution rather than the asymptotically equivalent normal distribution. At the end of the lecture I shall also sketch a proof of the same result by an induction argument, not related to the main ideas of this series of lectures.

Let n be a natural number and X a random variable uniformly distributed over the set $\{0,\ldots,n-1\}$. For the binary expansions of n-1 and X, I shall write

$$a = n-1 = \sum_{i=1}^{k} a_i 2^{k-i}, \tag{2}$$

and

$$X = \sum_{i=1}^{k} X_i 2^{k-i}. \tag{3}$$

We are interested in the distribution of

$$W = \sum_{i=1}^{k} X_i. \tag{4}$$

When $n = 2^k$, with k a non-negative integer, the distribution of W is the binomial distribution for k trials with probability $\frac{1}{2}$. It is intuitively plausible that this should also hold approximately for all large n with k defined by (1) and we shall see that this is true in a fairly strong sense.

In order to prove this, let I be a random variable uniformly distributed over the set $\{1,\ldots,k\}$ independent of X, and let the random variable X' be defined by

$$X' = \sum_{i=1}^{k} X'_i 2^{k-i}, \tag{5}$$

where

$$X'_i = \begin{cases} X_i & \text{if } i \neq I \\ 1-X_I & \text{if } i=I \text{ and this choice yields } X' < n \\ 0 & \text{if } i=I,\ X_I=0 \text{ and } X+2^{k-I} \geq n. \end{cases} \tag{6}$$

Also let

$$W' = \sum_{i=1}^{k} X'_i. \tag{7}$$

The ordered pair (X,X') of random variables is exchangeable, and thus the pair (W,W') is also exchangeable. Because the function

$$(w,w') \mapsto f(w)\mathcal{I}\{w'=w+1\} - f(w')\mathcal{I}\{w=w'+1\} \tag{8}$$

is antisymmetric in the sense of (I.5) for all $f: \{0,\ldots,k\} \to R$, we have

$$\begin{aligned} 0 &= E[f(W)\mathcal{I}\{W'=W+1\} - f(W')\mathcal{I}\{W=W'+1\}] \\ &= E[f(W)P^X\{W'=W+1\} - f(W-1)P^X\{W'=W-1\}] \\ &= E[f(W)(1-\tfrac{W+Q}{k}) - f(W-1)\tfrac{W}{k}], \end{aligned} \tag{9}$$

where

$$Q = |\{j:X_j=0 \ \&\ X+2^{k-j} \geq n\}|. \tag{10}$$

I have used the fact that

$$P^X\{W'=W-1\} = P^X\{X_I=1\} = \frac{W}{k}, \tag{11}$$

and

$$P^X\{W'=W+1\} = P^X\{X_I=0 \ \& \ X_I'=1\}$$
$$= 1 - \frac{W+Q}{k}. \tag{12}$$

Multiplying (9) by k we obtain

$$E[(k-W-Q)f(W) - Wf(W-1)] = 0. \tag{13}$$

Motivated by the fact that, as we shall see later,

$$EQ \leq 2, \tag{14}$$

we introduce a function $g: \{0,\ldots,k\} \to R$ related to f by

$$(k-w)f(w) - wf(w-1) = g(w). \tag{15}$$

Then (13) can be rewritten in the form

$$E[g(W) - Qf(W)] = 0. \tag{16}$$

We shall need the following.

Lemma 1: For given $g: \{0,\ldots,k\} \to R$, in order that there exist a function $f: \{0,\ldots,k-1\} \to R$ such that (15) holds for all $w \in \{0,\ldots,k\}$ it is necessary and sufficient that

$$B_{k,.5}g = 0 \tag{17}$$

where

$$B_{k,.5}g = \frac{1}{2^k} \sum_{w=0}^{k} \binom{k}{w} g(w). \tag{18}$$

When this condition holds, the unique solution f of (15) is given by

$$f(w) = \sum_{v=0}^{w} \frac{\binom{k}{v}}{\binom{k}{w}} \frac{g(v)}{k-w}$$
$$= -\sum_{v=w+1}^{k} \frac{\binom{k}{v}}{\binom{k}{w}} \cdot \frac{g(v)}{k-w} \tag{19}$$

for all $w \in \{0,\ldots,k-1\}$.

Proof: First we observe that the values f(-1) and f(k), which are

undefined, are multiplied by 0 when they occur in (15), so that no ambiguity arises. The necessity of (17) follows from the case $n = 2^k$ of (13) since $Q = 0$ identically in this case and thus (16) implies

$$0 = Eg(W) = B_{k,.5}g. \tag{20}$$

When (17) holds, we can verify that the first form of (19) satisfies (15) by direct substitution: For $w < k$

$$(k-w)f(w) - wf(w-1) \tag{21}$$
$$= \sum_{v=0}^{w} \frac{\binom{k}{v}}{\binom{k}{w}} g(v) - \sum_{v=0}^{w-1} \frac{\binom{k}{v}}{\binom{k}{w-1}} \cdot \frac{w}{k-w+1} g(v) = g(w),$$

while, for $w = k$,

$$(k-w)f(w) - wf(w-1) = -kf(k-1) \tag{22}$$
$$= -\sum_{v=0}^{k-1} \frac{\binom{k}{v}}{\binom{k}{k}} g(v) = g(w),$$

by (17) and (18). The equality of the two forms of (19) follows from (17) and (18), and the uniqueness of the solution f is readily verified by induction on w.

In order to approximate $E\, h(W)$ where $h: \{0,\ldots,k\} \to R$ is given, we write

$$g = h - B_{k,.5}h \tag{23}$$

and define f by (19) so that (16) implies that

$$E\, h(W) = B_{k,.5}h + E\, Qf(W). \tag{24}$$

It remains to bound $E\, Qf(W)$.

Let us look at the special case $h = h_{w_0}$, the indicator function defined by

$$h_{w_0}(w) = \begin{cases} 1 & \text{if } w=w_0 \\ 0 & \text{if } w\neq w_0. \end{cases} \tag{25}$$

The function f_{w_0}, related to h_{w_0} as f is related to h by (19) and (23), is given by

$$(26)\qquad f_{w_0}(w) = \begin{cases} \dfrac{-B_{k,.5}{}^{h}w_0}{k-w} \sum\limits_{v=0}^{w} \dfrac{\binom{k}{v}}{\binom{k}{w}} & \text{if } w < w_0 \\ \dfrac{B_{k,.5}{}^{h}w_0}{k-w} \sum\limits_{v=w+1}^{k} \dfrac{\binom{k}{v}}{\binom{k}{w}} & \text{if } w \geq w_0. \end{cases}$$

It is not difficult to see that

$$(27)\qquad |f_{w_0}(w)| \leq \frac{2}{k}.$$

In fact, consider the function ρ defined by

$$(28)\qquad \rho(w) = \sum_{v=0}^{w} \frac{\binom{k}{v}}{\binom{k}{w}}.$$

By writing this in the form

$$(29)\qquad \rho(w) = 1 + \frac{w}{k-w+1} + \frac{w(w-1)}{(k-w+1)(k-w+2)} + \ldots,$$

we see that ρ is an increasing function. Thus, for $w < w_0$

$$(30)\qquad \begin{aligned} 0 > f_{w_0}(w) &\geq f_{w_0}(w_0-1) \\ &= -\frac{\binom{k}{w_0}}{2^k} \cdot \frac{1}{k-w_0+1}\, \rho(w_0-1) \\ &= -\frac{\binom{k}{w_0}}{\binom{k}{w_0-1}} \cdot \frac{1}{k-w_0+1} \quad \frac{1}{2^k} \sum_{v=0}^{w_0-1} \binom{k}{v} \\ &= -\frac{1}{w_0}\frac{1}{2^k} \sum_{v=0}^{w_0-1} \binom{k}{v} > -\frac{2}{k}. \end{aligned}$$

The final inequality follows from the fact that the immediately preceding expression decreases as w_0 increases from 1 to $[\frac{k+1}{2}]$ and for $w_0 \geq [\frac{k+1}{2}]$ is clearly greater than $-\frac{2}{k}$. The corresponding inequality for $w \geq w_0$,

$$(31)\qquad 0 < f_{w_0}(w) \leq f_{w_0}(w_0) < \frac{2}{k}$$

follows from the symmetry property

$$f_{k-w_0}(k-w-1) = f_{w_0}(w). \tag{32}$$

Now let us prove (14) in order to complete the proof of the binomial approximation for the distribution of W. The behavior of EQ as a function of n was studied by Delange (1975) but for our purpose the following simple argument of Diaconis (1977) will suffice. By the definition (10) of Q,

$$\begin{aligned} E\,Q &\le \sum_{j=1}^{k} P\{X \ge n-2^{k-j}\} \\ &= \sum_{j=1}^{k} \frac{2^{k-j}}{n} \le \frac{2^k}{2^{k-1}} = 2. \end{aligned} \tag{33}$$

Of course Q is non-negative. It follows from (24), (27), and (33) that

$$\begin{aligned} |P\{W=w_0\} - \frac{1}{2^k}\binom{k}{w_0}| &= |Eh_{w_0}(W) - B_{k,.5}\,h_{w_0}| \\ &= |EQf_{w_0}(W)| \le \frac{2}{k} E|Q| \le \frac{4}{k}. \end{aligned} \tag{34}$$

Thus the probability that W has a given value differs from the corresponding binomial probability by $O(\frac{1}{k})$. Since this binomial probability is of the exact order of $k^{-\frac{1}{2}}$ in the main part of the distribution, the bound is reasonably satisfactory in some respects.

Now let us look more carefully at the way this fits into the abstract framework of the first lecture. The underlying sample space Ω is $\{0,\ldots,n-1\}$, the probability measure P is the uniform distribution in Ω and the random variable W is the number of ones in the binary expansion of the random number X, uniformly distributed over $\{0,\ldots,n-1\}$. The exchangeable pair (X,X') is constructed from X by choosing a random number I uniformly distributed in $\{1,\ldots,k\}$ independent of X where k is related to n by $2^{k-1} < n \le 2^k$ and changing the coefficient of 2^{k-I} in the binary expansion of X to obtain X' provided this is less than n. Otherwise X'=X.

The space $\mathfrak{F}_0$ of diagram (I.28) consists of the functions $f: \{0,\ldots,k-1\} \to R$ and $\alpha: \mathfrak{F}_0 \to \mathfrak{F}$ is defined by

(35) $$(\alpha f)(X,X') = f(W)\mathcal{J}\{W'=W+1\} - f(W')\mathcal{J}\{W=W'+1\}.$$

The linear mapping T_0: $\mathfrak{F}_0 \to \mathcal{X}_0$, where $\mathcal{X}_0$ is the space of all h: $\{0,\ldots,k\} \to R$ is defined by

(36) $$(T_0 f)(W) = [(k-W)f(W) - Wf(W-1)]/k$$

with the convention that the result of multiplying the undefined f(-1) or f(k) by zero is zero. Of course $\imath$: $\mathcal{X}_0 \to \mathcal{X}$ is the appropriate inclusion mapping, that is

(37) $$(\imath h)(X) = h(W)$$

in the present notation. The linear functional E_0: $\mathcal{X}_0 \to R$ is $B_{k,.5}$, expectation under the appropriate binomial distribution, defined by (18), and the linear mapping U_0: $\mathcal{X}_0 \to \mathfrak{F}_0$ is defined by

(38) $$U_0 h = kf$$

given by (23) and (19). The identity (I.30) states that the f defined by (23) and (19) satisfies (15). Finally, (I.33) is specialized to (24) since $T\circ\alpha - \imath\circ T_0$: $\mathfrak{F}_0 \to \mathcal{X}_0$ is given by

(39) $$(T\circ\alpha - \imath\circ T_0)f(X)$$

$$= E^X[f(W)\mathcal{J}\{W'=W+1\} - f(W')\mathcal{J}\{W=W'+1\}] - [f(W)(1-\frac{W}{k}) - f(W-1)\frac{W}{k}]$$

$$= f(W)P^X\{W'=W+1\} - f(W-1)P^X\{W'=W-1\} - [f(W)(1-\frac{W}{k}) - f(W-1)\frac{W}{k}]$$

$$= -\frac{Q}{k} f(W).$$

Substituting in (I.33) and using (38) we obtain (24). Of course T: $\mathfrak{F} \to \mathcal{X}$ is defined as usual by (I.20).

Now let us look at an alternative argument that is in some ways more successful than the above. I shall have to indicate the dependence of X, W and k on a in (2), (3), and (4) by writing X(a), W(a) and k(a). Then, for any h: $\{0,\ldots,k(a)\} \to R$

$$
\begin{aligned}
(40)\qquad & E\,h(W(a)) \\
&= P\{X(a) \le 2^{k(a)-1} - 1\}E[h(W(a))|X(a) \le 2^{k(a)-1} - 1] \\
&\quad + P\{X(a) \ge 2^{k(a)-1}\}E[h(W(a))|X(a) \ge 2^{k(a)-1}] \\
&= \frac{2^{k(a)-1}}{a+1}\,\mathcal{B}_{k(a)-1,.5}\,h + \frac{a+1-2^{k(a)-1}}{a+1}\,E\,h(1+W(a-2^{k(a)-1})).
\end{aligned}
$$

In the second equality I have inserted the probabilities of the two cases and recognized the conditional distribution of W(a) given each of the two conditions. Given that $X(a) \le 2^{k(a)-1}-1$, the conditional distribution of W(a) is that of $W(2^{k(a)-1})$, a binomial distribution with k(a)-1 trials and probability $\frac{1}{2}$. Given that $X(a) \ge 2^{k(a)-1}$ the conditional distribution of W(a) is the unconditional distribution of $1+W(a-2^{k(a)-1})$ since the number of ones in the binary expansion of X(a) is one more than the number of ones in the binary expansion of $X(a)-2^{k(a)-1}$ which, given that $X_1(a) = 1$, is distributed as $X(a-2^{k(a)-1})$.

In order to apply (40) inductively, let

$$
(41)\qquad k_1(a) = k(a)
$$

and, for $j \ge 1$,

$$
(42)\qquad k_{j+1}(a) = k_1\Big(a - \sum_{i=1}^{j} 2^{k_i(a)-1}\Big).
$$

We can express (41) explicitly as

$$
(43)\qquad k_1(a) = 1 + [\log_2 a].
$$

Then, by induction on j, (40) yields

$$
\begin{aligned}
(44)\qquad & E\,h(W(a)) \\
&= \sum_{i=1}^{j} \frac{2^{k_i(a)-1}}{a+1}\,\mathcal{B}_{k_i(a)-1,.5}(w \mapsto h(i-1+w)) \\
&\quad + \frac{a - \sum_{i=1}^{j} 2^{k_i(a)-1} + 1}{a+1}\,E\,h\Big(j+W\Big(a - \sum_{i=1}^{j} 2^{k_i(a)-1}\Big)\Big).
\end{aligned}
$$

In particular, for j equal to j(a), the number of ones in the binary

expansion of a,

$$(45)\quad E\,h(W(a)) = \sum_{i=1}^{j(a)} \frac{2^{k_i(a)-1}}{a+1} B_{k_i(a)-1,.5}(w \mapsto h(i-1+w)) + \frac{h(j(a))}{a+1}$$

$$= \mathcal{B}_{k(a),.5}h + \sum_{i=1}^{j(a)} \frac{2^{k_i(a)-1}}{a+1}[B_{k_i(a)-1,.5}(w \mapsto h(i-1+w)) - \mathcal{B}_{k(a),.5}h]$$

$$+ \frac{1}{a+1}[h(j) - B_{k(a),.5}h].$$

If k is not enormous this can be used to compute E h(W(a)) exactly. In the second form it also yields

$$(46)\qquad |P\{W=w_0\} - \frac{1}{2^k}\frac{k}{w_0}| = O(\frac{1}{k}),$$

and the approximation $B_{k(a),.5}h$ for Eh(W(a)) can readily be improved by using a few of the terms of the summation in the final form of (45) and bounding the rest.

For large n, the number of ones in the binary expansion of a random integer uniformly distributed over {0,...,n-1} has approximately a binomial distribution, that of the number of successes in $[\log_2 n]$ independent trials with probability one-half. The error in the probability of any particular values was shown to be of the order of $(\log_2 n)^{-1}$. The proof by the method of the present series of lectures seems to be quite simple in its basic outline. However, starting just above (40), an alternative proof by induction on n that is more powerful and perhaps simpler was described briefly.