ESTIMATION IN A CERTAIN PROBABILITY PROBLEM

GERALD A. HEUER, Concordia College and Remington Rand Univac

An example sometimes quoted as a "nonintuitive" probability is the following: Given a collection of K people, find the probability that some two of them were born on the same day of the year (assuming that no one was born on February 29, and that people are born with equal probability on other days). If K is 23, the probability slightly exceeds 1/2, and if K is 50, the odds are about 33 to 1 in favor of such a pair.

This is of course a special case of the following problem. Let an experiment with N equally likely outcomes be performed K times. What is the probability, $P(N, K)$, that at least one of the outcomes occurs twice? $P(N, K)=1-Q(N, K)$ where $Q(N, K)=N(N-1)\cdots(N-K+1)/N^K=N!/[(N-K)!N^K]$ is the probability that all K outcomes are distinct $(K<N)$. For given N and K this may be evaluated directly, or approximated with the aid of Stirling's formula.

The problem becomes more difficult, however, if one assigns values to $P(N, K)$ and N, and attempts to solve for K. We are then confronted with an equation

$$1-\frac{N!}{(N-K)!N^K}=t$$

or, if Stirling's formula is used,

$$\left(\frac{N}{N-K}\right)^{N-K+1/2}e^{-K}=1-t,$$

to be solved for K, where $t=P(N, K)$. Such transcendental equations can be solved approximately, of course, when N and t are specified, but to exhibit $K=f(N, t)$ explicitly seems rather difficult. It is the purpose of this note to show that for $0<t<1$, K is given asymptotically by $L(t)\sqrt{N}$, where $L(t)=\sqrt{-2\log(1-t)}$. Furthermore, except for extreme values of t, this approximation is very good even for small N.

We first prove the

LEMMA. *If $Q(N, K)\geqq a$, where a is constant, $0<a<1$, then $K/N\to 0$ as $N\to\infty$.*

*Proof.** Since $1-x<e^{-x}$ for all x, we have

$$a\leqq Q(N,K)=\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{K-1}{N}\right)$$

$$\leqq\exp\left[-\left(\frac{1}{N}+\cdots+\frac{K-1}{N}\right)\right]=\exp\left[-\frac{K(K-1)}{2N}\right],$$

so that $K(K-1)\leqq 2N\log(1/a)$, which implies the conclusion of the lemma.

* The author is indebted to the referee for suggesting this simpler proof of the lemma.

THEOREM. *Let K, N, and t be related by the equation*

$$1-\frac{N!}{(N-K)!N^K}=t.$$

(We may assume all three are continuous real variables, $0<K<N$.*) Then*

$$\frac{K^2}{-2N\log(1-t)}\to 1 \text{ as } N\to\infty.$$

Thus K is given asymptotically by $K=L(t)\sqrt{N}$, *where* $L(t)=\sqrt{-2\log(1-t)}$.

Proof. By Stirling's inequality we have

$$\left(\frac{N}{N-K}\right)^{N-K+1/2}e^{-K}\left(\frac{12(N-K)-1}{12(N-K)}\right)$$

$$<\frac{N!}{(N-K)!N^K}<\left(\frac{N}{N-K}\right)^{N-K+1/2}e^{-K}\frac{12N}{(12N-1)}.$$

Since the middle member above is $1-t$, we have, by inverting,

$$(1-K/N)^{N-K+1/2}e^K=1/[(1-t)(1+\epsilon)],$$

where $\epsilon\to 0$ as $N\to\infty$ (since $N-K\to\infty$ when $N\to\infty$).

Taking logarithms gives

$$\begin{aligned}-\log[(1-t)(1+\epsilon)] &= K+(N-K+\tfrac{1}{2})\log(1-K/N)\\ &= K-(N-K)\sum_{r=1}^{\infty}\frac{1}{r}\frac{K^r}{N^r}+\frac{1}{2}\log(1-K/N)\\ &= K-K-\sum_{r=2}^{\infty}\frac{1}{r}\frac{K^r}{N^{r-1}}+\sum_{r=1}^{\infty}\frac{1}{r}\frac{K^{r+1}}{N^r}+\frac{1}{2}\log(1-K/N)\\ &= \frac{1}{2}\log(1-K/N)+\sum_{n=0}^{\infty}\left(\frac{1}{n+1}-\frac{1}{n+2}\right)\frac{K^{n+2}}{N^{n+1}}\\ &= \frac{1}{2}\log(1-K/N)+\frac{K^2}{2N}\left[1+\sum_{n=1}^{\infty}\frac{2}{(n+1)(n+2)}\frac{K^n}{N^n}\right].\end{aligned}$$

The lemma shows that $K/N\to 0$ as $N\to\infty$, so we have $-\log[(1-t)(1+\epsilon)]=\epsilon'+[K^2/(2N)](1+\epsilon'')$, where $\epsilon'=\frac{1}{2}\log(1-K/N)\to 0$ as $N\to\infty$ and

$$0<\epsilon''=\sum_{n=1}^{\infty}\frac{2}{(n+1)(n+2)}\frac{K^n}{N^n}<\sum_{n=1}^{\infty}\frac{K^n}{N^n}=\frac{K/N}{1-K/N}\to 0$$

as $N\to\infty$. Thus

$$\frac{K^2}{2N}=\frac{-\log[(1-t)(1+\epsilon)]-\epsilon''}{1+\epsilon'},$$

so that

$$\frac{K^2}{[L(t)]^2 N} \qquad \frac{-\log[(1-t)(1+\epsilon)] - \epsilon''}{-(1+\epsilon')\log(1-t)} \to 1.$$

It is of some interest to see how good the approximation is for several values of N. The author has done this with the aid of the Univac 1103A and some of the results are tabulated below. In this table, K is the least integer such that $N!/[(N-K)!N^K] \leqq 1-t$ and K_1 is the integer nearest $L(t)\sqrt{N}$.

	$t=\frac{1}{32}$		$t=\frac{1}{4}$		$t=\frac{1}{2}$		$t=\frac{3}{4}$		$t=31/32$	
N	K	K_1	K	K_1	K	K_1	K	K_1	K	K_1
4	2	1	2	2	3	2	4	3	5	5
4^2	2	1	4	3	5	5	7	7	10	11
4^3	3	2	7	6	10	9	14	13	21	21
4^4	5	4	13	12	20	19	27	27	42	42
4^5	9	8	25	24	38	38	54	53	84	84
4^6	17	16	49	49	76	75	107	107	168	168
4^7	33	32	98	97	151	151	214	213	337	337
4^8	65	65	195	194	302	301	427	426	674	674
4^9	130	129	389	388	604	603	853	853	1348	1348
4^{10}	259	258	778	777	1206	1206	1706	1705	2696	2696
4^{11}	517	516	1554	1553	2412	2411	3411	3410	5392	5392
4^{12}	1033	1032	3108	3107	4823	4823	6821	6820	10784	10784
4^{13}	2065	2064	6215	6214	9646	9645	13641	13641	21567	21568
4^{14}	4130	4129	12429	12428	19291	19291	27282	27281	43135	43135
4^{15}	8258	8257	24856	24855	38582	38581	54563	54562	86270	86271
4^{16}	16515	16514	49712	49711	77164	77163	109125	109125	172541	172541
4^{17}	33029	33028	99423	99422	154326	154325	218250	218249	345082	345082

AN ANALYTICAL EXPRESSION FOR $[X]$

ARTHUR PORGES, Los Angeles City College

THEOREM. *Given the functions*

$$F(X) = \frac{\operatorname{Arcsin}|\sin \pi x|}{\pi} \quad \textit{and} \quad G(X) = \lim_{N\to\infty} \{1 + |\sin \pi(X - F(X))|\}^N,$$

then for all finite X, $[X]$ (the greatest integer not greater than X) may be written as

$$[X] = X - |F(X) + 2^{1-G(X)} - 1|.$$

To prove the theorem it is helpful to establish two lemmas.

LEMMA 1. *For K an integer and $0 \leqq b \leqq \frac{1}{2}$, $F(K+b) = b$; for $\frac{1}{2} < b < 1$, $F(K+b) = 1 - b$.*

LEMMA 2. *For K an integer and $0 \leqq b \leqq \frac{1}{2}$, $G(K+b) = 1$; for $\frac{1}{2} < b < 1$, $G(K+b) = \infty$.*