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

\[
Q(N, K) = \frac{N(N-1) \cdots (N-K+1)}{N^K} = \frac{N!}{(N-K)!NK}
\]

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

\[
1 - \frac{N!}{(N-K)!NK} = t
\]

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 \( Q(N, K) \geq a \), 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

\[
a \leq Q(N, K) = \left( 1 - \frac{1}{N} \right) \cdots \left( 1 - \frac{K-1}{N} \right)
\]

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

so that \( K(K-1) \leq 2N \log (1/a) \), 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

$$-\log [(1 - t)(1 + \epsilon)] = K + (N - K + \frac{1}{2}) \log (1 - K/N)$$

$$= K - (N - K) \sum_{r=1}^{\infty} \frac{1}{r} \frac{K^r}{N^r} + \frac{1}{2} \log (1 - K/N)$$

$$= K - K - \sum_{r=2}^{\infty} \frac{1}{r} \frac{K^r}{N^{r-1}} + \sum_{r=1}^{\infty} \frac{1}{r} \frac{K^{r+1}}{N^r} + \frac{1}{2} \log (1 - K/N)$$

$$= \frac{1}{2} \log (1 - K/N) + \sum_{n=0}^{\infty} \left( \frac{1}{n + 1} - \frac{1}{n + 2} \right) \frac{K^{n+2}}{N^{n+1}}$$

$$= \frac{1}{2} \log (1 - K/N) + \frac{K^2}{2N} \left[ 1 + \sum_{n=1}^{\infty} \frac{2}{(n + 1)(n + 2)} \frac{K^n}{N^n} \right].$$

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

$$\frac{K^2}{2N} = -\log [(1 - t)(1 + \epsilon)] - \epsilon'' \quad \text{and} \quad 1 + \epsilon'$$
so that
\[
\frac{K^2}{[L(t)]^2N} - \log [(1 - t)(1 + e)] - e'' - (1 + e') \log (1 - t) \to 1.
\]

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 \(N!/[N - K)!NK]\_ \approx 1 - t\) and \(K_1\) is the integer nearest \(L(t)\sqrt{N}\).

<table>
<thead>
<tr>
<th>(t)</th>
<th>(N)</th>
<th>(K)</th>
<th>(K_1)</th>
<th>(K)</th>
<th>(K_1)</th>
<th>(K)</th>
<th>(K_1)</th>
<th>(K)</th>
<th>(K_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(31/32)</td>
<td>(4)</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>(4)</td>
<td>(4)</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>(5)</td>
<td>(8)</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>9</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>(6)</td>
<td>(32)</td>
<td>13</td>
<td>12</td>
<td>20</td>
<td>19</td>
<td>27</td>
<td>27</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>(7)</td>
<td>(9)</td>
<td>5</td>
<td>4</td>
<td>25</td>
<td>24</td>
<td>38</td>
<td>38</td>
<td>54</td>
<td>53</td>
</tr>
<tr>
<td>(8)</td>
<td>(16)</td>
<td>76</td>
<td>75</td>
<td>107</td>
<td>107</td>
<td>168</td>
<td>168</td>
<td>337</td>
<td>337</td>
</tr>
<tr>
<td>(9)</td>
<td>(129)</td>
<td>130</td>
<td>129</td>
<td>604</td>
<td>603</td>
<td>853</td>
<td>853</td>
<td>1348</td>
<td>1348</td>
</tr>
<tr>
<td>(10)</td>
<td>(258)</td>
<td>259</td>
<td>258</td>
<td>1206</td>
<td>1206</td>
<td>1706</td>
<td>1705</td>
<td>2696</td>
<td>2696</td>
</tr>
<tr>
<td>(11)</td>
<td>(516)</td>
<td>517</td>
<td>516</td>
<td>2412</td>
<td>2411</td>
<td>3411</td>
<td>3410</td>
<td>5392</td>
<td>5392</td>
</tr>
<tr>
<td>(12)</td>
<td>(1032)</td>
<td>1033</td>
<td>1032</td>
<td>4823</td>
<td>4823</td>
<td>6821</td>
<td>6820</td>
<td>10784</td>
<td>10784</td>
</tr>
<tr>
<td>(13)</td>
<td>(2064)</td>
<td>2065</td>
<td>2064</td>
<td>9645</td>
<td>9645</td>
<td>13641</td>
<td>13641</td>
<td>21567</td>
<td>21568</td>
</tr>
<tr>
<td>(14)</td>
<td>(4129)</td>
<td>4130</td>
<td>4129</td>
<td>19291</td>
<td>19291</td>
<td>27282</td>
<td>27281</td>
<td>43135</td>
<td>43135</td>
</tr>
<tr>
<td>(15)</td>
<td>(8257)</td>
<td>8258</td>
<td>8257</td>
<td>38581</td>
<td>38581</td>
<td>54563</td>
<td>54562</td>
<td>86270</td>
<td>86271</td>
</tr>
<tr>
<td>(16)</td>
<td>(16514)</td>
<td>16515</td>
<td>16514</td>
<td>77164</td>
<td>77163</td>
<td>109125</td>
<td>109125</td>
<td>172541</td>
<td>172541</td>
</tr>
<tr>
<td>(17)</td>
<td>(33028)</td>
<td>33029</td>
<td>33028</td>
<td>154326</td>
<td>154325</td>
<td>218250</td>
<td>218249</td>
<td>345082</td>
<td>345082</td>
</tr>
</tbody>
</table>

**AN ANALYTICAL EXPRESSION FOR \([X]\)**

ARTHUR PORGES, Los Angeles City College

**THEOREM.** Given the functions
\[
F(X) = \frac{\text{Arcsin} \left( \frac{\sin \pi x}{\pi} \right)}{\pi} \quad \text{and} \quad G(X) = \lim_{N \to \infty} \left\{ 1 + \left| \sin \pi (X - F(X)) \right| \right\}^N,
\]
then for all finite \(X\), \([X]\) (the greatest integer not greater than \(X\)) may be written as
\[
[X] = X - \left| F(X) + 2^{1-G(X)} - 1 \right|.
\]

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

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

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