Solution

Week 48 (8/11/03)

The hotel problem

In figuring out the probability for success (choosing the cheapest hotel), it is convenient to organize the different cases according to what the highest ranking hotel (in order of cheapness) in the first fraction x is. Let H_1 denote the cheapest hotel, H_2 the second cheapest, etc.

Assume that H_1 is among the first x, which happens with probability x. In this case, there is failure.

Assume that H_2 is the cheapest among the first x, which happens with probability $x(1-x)$. This is the probability that H_2 is in the first x, times the probability that H_1 is not.[1] In this case, we have success.

Assume that H_3 is the cheapest among the first x, which happens with probability $x(1-x)^2$ (again, see the remark below). This is the probability that H_3 is in the first x, times the probability that H_2 is not, times the probability that H_1 also is not. In this case, we have success 1/2 of the time, because only 1/2 of the time H_1 will come before H_2.

Continuing in this fashion, we see that the probability for success, P, is

$$
\begin{aligned}
P(x) &= x(1-x) + \frac{1}{2}x(1-x)^2 + \frac{1}{3}x(1-x)^3 + \cdots \\
&= \sum_{k=1} \frac{1}{k}\, x(1-x)^k. \qquad (1)
\end{aligned}
$$

The $1/k$ factor comes from the probability that H_1 is first among the top k hotels which lie in the final $(1-x)$ fraction. Using the expansion $\ln(1-y) = -(y + y^2/2 + y^3/3 + \cdots)$, with $y = 1-x$, we obtain

$$P(x) = -x \ln x. \qquad (2)$$

Taking the derivative, we see that $P(x)$ is maximized when $x = 1/e$, in which case the value is $1/e$. Therefore, you want to pass up $1/e \approx 37\%$ of the hotels, and then pick the next one that is better than all the ones you've seen. Your chance of getting the best one is $1/e \approx 37\%$.

REMARK: For sufficiently large N, the probabilities in eq. (1) are arbitrarily close to $x(1-x)^k/k$, for small values of k. But each successive term in eq. (1) is suppressed by a factor of at least $(1-1/e)$. The terms therefore become negligibly small at a k value that is independent of N. The only significant contribution to the sum therefore comes from small values of k, for which the given probabilities are essentially correct.

[1] The $(1-x)$ factor is technically not correct, because there are only $N-1$ spots available for H_1, given that H_2 has been placed. But the error is negligible for large N. See the remark below.