Fair Sample Selection

Markus G. Kuhn

Computer Laboratory
University of Cambridge

2000-05-26

http://www.cl.cam.ac.uk/~mgk25/fair-samples.dvi

Fair Sample Selection I

Problem: Alice receives a sequence of objects $O_1, \ldots, O_n$. She does not know n in advance. She can hold in her hand only one single object at a time. She must decide for each new object O_i whether to keep the new one or whether to hold on to the one already in her hand. The other object must be discarded and can't be retrieved later.

Alice wants that all O_i ($1 \leq i \leq n$) have an equal probability $S_i = 1/n$ of becoming the sample that she holds in her hand at the end.

Is there a strategy (i.e., a set of probabilities K_i for keeping O_i on arrival and discarding the previous sample) that achieves exactly that?

Fair Sample Selection II

Solution:

Keep O_i on arrival with probability $K_i = \frac{1}{i}$.

Reason:

We obviously need $K_n = \frac{1}{n}$, because $K_n = S_n$. Since Alice does not know n, each O_i could be the last one. So if there is a solution at all, it must be $K_i = \frac{1}{i}$. This actually works, because

$$\begin{aligned} S_i &= K_i \cdot \prod_{j=i+1}^{n} (1 - K_j) \\ &= \frac{1}{i} \cdot \frac{i}{i+1} \cdot \frac{i+1}{i+2} \cdot \ldots \cdot \frac{n-1}{n} = \frac{1}{n} \end{aligned}$$

Fair Sample Selection III

What if Alice can keep m objects in her hand?

Solution:

First keep all of $O_1, \ldots, O_m$ and then for $i > m$ with probability $K_i = \frac{m}{i}$ discard a randomly selected previous sample and keep O_i on arrival.

Reason:

Obviously $S_1 = \cdots = S_m$ and for $i \geq m$:

$$
\begin{aligned}
S_i &= K_i \cdot \prod_{j=i+1}^{n} (1 - K_j \cdot \frac{1}{m}) \\
&= \frac{m}{i} \cdot \prod_{j=i+1}^{n} (1 - \frac{m}{i} \cdot \frac{1}{m}) \\
&= \frac{m}{i} \cdot \frac{i}{i+1} \cdot \frac{i+1}{i+2} \cdot \cdots \cdot \frac{n-1}{n} = \frac{m}{n}
\end{aligned}
$$