Two-Step Drawing from Urns

Stephan Kolassa and Stefan Schwarz

Institut für Angewandte Mathematik, Friedrich-Schiller-Universität Jena, 07740 Jena, Germany, {kolassa,schwarz}@minet.uni-jena.de

Abstract. Consider the following situation of two-step shortlisting: two experts Alice and Bob are faced with a large number of alternatives which they can only observe imprecisely. They have to choose one of the alternatives, without knowing which one is best. Alice first compiles a shortlist of alternatives by choosing her k best observations. Bob then chooses his best observation among the shortlisted alternatives. Previous research showed that this procedure sometimes yielded worse results than if a single expert made the entire decision himself. Here, we consider an urn containing $n-1$ homogeneous balls and one ball with larger weight. When drawing balls at random from the urn, the probability of drawing any one ball is proportional to its weight. Alice draws k balls and puts them in another urn, from which Bob then draws a single ball. Which value of k maximizes the probability that Bob draws the distinguished ball?

1 Introduction

Situations of two-step shortlisting by imperfect experts can be formally characterized as follows: consider two experts Alice and Bob, faced with a large number of alternatives which they can evaluate only imprecisely. Alice chooses a subset of k alternatives, and Bob then chooses one alternative from this *shortlist* compiled by Alice, with the goal of choosing the unknown best one among the original alternatives. One hopes that the imprecisions of the experts cancel each other out to a certain degree, leading to better results than if only a single expert made the decision.

Previous research on two-step shortlisting by imperfect experts yielded surprising results [1,3]. Even if both experts are equally imprecise, as measured by the distributions of stochastic noise terms, the shortlisting can actually lead to worse results than if a single expert made the entire decision alone: a *shortlisting valley* appeared.

Here, we examine a model of sequential drawing from an urn. The objective of Alice and Bob is to draw the heaviest ball. We restrict ourselves to the special case of $n-1$ homogeneous balls and one distinguished ball which is heavier than the others. Initial results for this model have been given in [4].

The probability of drawing the heaviest ball first increases and then decreases with the shortlist size k (Prop. 1). The optimal value of k is largest for small weights of the distinguished ball. It decreases monotonically from $\sqrt{2n}$ to 2 with increasing distinguished weight (Prop. 2 and 3).

2 The Model

We consider an urn $\mathcal{U}$ containing $n \geq 3$ balls, which we identify with the numbers $1, \ldots, n$. Ball 1 has weight $p > 1$, and the other balls have weight 1. Whenever a ball is drawn from an urn, the probability of any particular ball to be drawn is proportional to its weight.

Alice draws k balls at random from $\mathcal{U}$, one after the other, and puts them in a second urn $\mathcal{S}$ (for "shortlist"). Bob then draws a single ball from $\mathcal{S}$.

We are interested in how the probability that Bob draws the distinguished ball (the *hitting ratio*) varies with $k \in \{1, \ldots, n\}$. In particular, we will investigate the value k^* which yields the highest hitting ratio.

3 Unimodality

Proposition 1. *The hitting ratio is unimodal in k, i.e., it first increases and then decreases.*

It turns out that the analysis is facilitated by assuming different "subjective" weights of ball 1 for Alice and Bob: we let $p > 1$ denote the "subjective" weight of ball 1 for Alice as above and $q > 1$ the "subjective" weight for Bob.

Proof. The probability for Alice to pass on the correct ball in $\mathcal{S}$ is the complement of the probability that she passes on only "bad" balls:

$$1 - \frac{n-1}{n-1+p} \cdot \frac{n-2}{n-2+p} \cdot \ldots \cdot \frac{n-k}{n-k+p}.$$

The hitting ratio h for parameters p, k and q therefore is

$$h(n \xrightarrow{p} k \xrightarrow{q} 1) = \left(1 - \frac{n-1}{n-1+p} \cdot \ldots \cdot \frac{n-k}{n-k+p}\right) \cdot \frac{q}{k-1+q}. \tag{1}$$

First, for fixed n, k and p, we compute the value $q_k(p)$ for which

$$h(n \xrightarrow{p} k \xrightarrow{q_k(p)} 1) = h(n \xrightarrow{p} k+1 \xrightarrow{q_k(p)} 1) \quad \text{for } 1 \leq k < n.$$

Some calculations yield that

$$q_k(p) = \frac{(n-1+p)\ldots(n-k+p)(n-k-1+p)}{(n-1)\ldots(n-k)p} - \frac{n-k-1}{p} - k \tag{2}$$

and that

$$q_1(p) < q_2(p) < \cdots < q_{n-2}(p) < q_{n-1}(p).$$

We have $h(n \xrightarrow{p} k \xrightarrow{q} 1) < h(n \xrightarrow{p} k+1 \xrightarrow{q} 1)$ for $q > q_k(p)$ and $h(n \xrightarrow{p} k \xrightarrow{q} 1) > h(n \xrightarrow{p} k+1 \xrightarrow{q} 1)$ for $q < q_k(p)$.