Strategic Costly Sample Selection

Stephan Lauermann Asher Wolinsky

December 2, 2025

Comment: This is work in progress.

Extended abstract

Various interactive scenarios involve the disclosure of a first order statistic by a strategic agent who is interested in affecting the resulting posterior beliefs of an uninformed agent. An example of such a scenario is the case of a researcher who looks for a convincing finding to obtain funding. This paper examines the informativeness of the first order statistic under such circumstances when commitment is not possible.

Much of the large literature on first order statistics is not concerned with strategic interactions. Previous papers that considered these issues in some form include Lauermann and Wolinsky (2017, 2022) and Di Tillio, Ottaviani and Sorenson (2021).

There is a binary state ("low" and "high") and a signal whose probability distribution depends on the state – higher signal realizations are more likely in the high state. The **sender** takes a sample of (conditionally) independent realizations of that signal at a constant cost per realization and then presents the highest signal realization to the **receiver**. Based on sender's report, the receiver who does not observe the state or any other information about the sample, updates her belief of the state. The sender's objective is to maximize the receiver's posterior probability of the high state while economizing on sampling costs. We consider both the case in which the sender is uninformed about the state (and shares the same prior with the receiver) and the case in which he is informed. In equilibrium, the sender selects the sample size optimally (as a function of the state, if informed) and, given the sender's strategy and the actual report, the receiver forms her belief concerning the state.

In the uninformed sender case, there always exists a partially informative equilibrium in which the distribution of beliefs varies across the states and the expected belief is higher in the higher state. In the informed sender case, we prove existence of such an equilibrium under additional conditions (small marginal sampling cost and lower bound on the inverse hazard ratio). The existence of such an equilibrium is not a-priori obvious in the informed sender case, since

the small sampling cost incentivizes the sender to sample more aggressively in the low state to mimic the outcome of the high state.

The main observations include the following. First, simple characterization of the equilibrium belief distribution for the case of small marginal sampling cost. Second, the total equilibrium sampling cost captured the cumulative entropy – a dispersion measure – of the equilibrium belief distribution. One corollary of these two points is that even when the marginal sampling cost is small, the overall cost that the sender incurs is significant. In contrast, in a scenario in which the receiver observes the sample size, that cost would be small when the marginal sampling cost is small. Third, while the receiver gains valuable information relative to the ex-ante position, the sender's sampling effort is pure waste in terms of his ex-ante payoffs, since the overall expected belief that the sampling induces in equilibrium must average to the prior. This point is reminiscent of the familiar "signal jamming" and related insights, but our analysis goes further to characterize the equilibrium magnitudes of this wasted effort and of the information obtained by the receiver in terms of the signal distributions and the prior. Fourth, in the case of an informed sender, the strategic selection may enhance or diminish the informativeness of the first order statistic. This depends on the ratio of the equilibrium sample sizes selected in the two states, which in turn depends on the prior and on the likelihood ratio (of the two states) for high signal realizations. The intuition is that a relatively low top likelihood ratio makes it easier for the sender to mimic the high state result and hence induces relatively more aggressive sampling in the low state, which diminishes the informativeness of the first order statistic. Fifth, the degree to which the strategic sampling enhances (or dampens) the equilibrium informativeness of the first order statistic is not monotone in the likelihood ratio (of the two states) for high signal realizations—at relatively low values of the latter the degree of informativeness decreases and for higher values it increases.

In a variation on the model, the sender can censor the report, i.e., to report any value lower than the highest signal realization. This might be feasible in certain scenarios in which the signal is a cumulative measure that can be partially hidden. In the equilibrium described above, the sender has no incentive to censor. But when the sender is informed, it is conceivable that the possibility of censoring may give rise to pooling equilibria in which a certain threshold value is being reported for all first order statistics above that threshold. Uninteresting equilibria of this type can be sustained by pessimistic off path beliefs. However, even when off-path beliefs are refined to preclude such equilibria, the may be equilibria in which censoring takes place because the informed sender's sample sizes are sufficiently larger in the low state than in the high state leading to decreasing posterior belief over a range of high realizations of the first order statistic. We show that, for some specifications of the underlying state-dependent signal distributions, there is no censoring equilibrium. Thus, under these circumstances, when censoring is allowed, the partially informative equilibrium described above is unique in the informed sender case as well (in the limit as the marginal sampling cost is small). However, it is fairly clear that there are signal distributions that would also support an uninformative equilib-

rium with censoring. But the work on this is in progress and we have not yet fully established the conditions that would give rise to such equilibria.

We also consider briefly a sequential version of the model. It is a sort of a signaling model that has both separating and pooling equilibria.