Online Appendix for Constrained Retrospective Search

Can Urgun* Leeat Yariv†

December 31, 2020

1 Constrained Search with I.I.D. Samples

The existence of an outside option governed by $X_0 = M_0 = 0$ implies that each sample is effectively sampled from a censored normal. For our characterization of the optimal policy, we need to derive the distribution of the first-order statistic of n censored normal distributions. The distribution function for a normal variable with mean 0 and standard deviation σ, censored at 0 is given by:

$$f(x;\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} I_{x>1},$$

where $I_{x>1}$ is the indicator function for $x > 1$. From here we see that a censored normal with standard deviation σ has the same distribution as a censored normal with standard deviation 1 multiplied by σ, much like the uncensored normal.[1] Thus, the first-order statistic of n censored normals with standard deviation σ has the same distribution as $\sigma Y_{(n)}$ where $Y_{(n)}$ is the first-order statistic of n censored normals with standard deviation 1. Thus, the problem of the decision maker can be written as

$$\max_{n,\sigma} \sigma Y_{(n)} - nc(\sigma),$$

which leads to the result in Proposition 2.

*Department of Economics, Princeton University. Email: curgun@princeton.edu
†Department of Economics, Princeton University. Email: lyariv@princeton.edu

[1] This scale-invariance property only holds when censoring is at 0.

2 The Impacts of Constraints

To our knowledge, bounds on the order statistics of censored normal variables are not readily available. We now derive an upper bound for $Y_{(n)}$. Let $t > 0$ be arbitrary and $\{X_i\}_i$ be a sequence of i.i.d. censored normal variables with standard deviation 1. By Jensen's inequality,

$$e^{(tE(Y_{(n)}))} \leq E(e^{tY_{(n)}}) = E(\max_{i \in n} e^{tX_i}).$$

Since $X_i \geq 0$ for all i, their maximum is lower than their sum. Thus,

$$E(\max_{i \in n} e^{tY_{(n)}}) \leq \sum_{i=1}^{n} E(e^{tX_i}) = n\left(\frac{1}{2}e^{\frac{t^2}{2}}(1 + erf(\frac{t}{2})\right),$$

where the last equality follows from taking the expectation and erf denotes the Gaussian error function.

By definition, $erf(\frac{t}{2}) \leq 1$. Combining these inequalities we have:

$$e^{(tE(Y_{(n)}))} \leq n\left(\frac{1}{2}e^{\frac{t^2}{2}}(1+1)\right).$$

Taking log of both sides and dividing by t yields

$$E(Y_{(n)}) \leq \frac{logn}{t} + \frac{t}{2}.$$

Minimizing the right hand side for a sharper upper bound implies $t = \sqrt{2logn}$, which generates our desired bound:

$$E(Y_{(n)}) \leq \sqrt{2logn}.$$

Since the expected payoff from any sample of n needs to account for their cost, this bound also offers an upper bound on the expected payoffs: for any number n of samples, $\bar{V}^{iid} < \sigma\sqrt{2logn}$.[2] Thus, as n gets large, $\bar{V}^{iid}$ cannot grow faster than the $\sqrt{2logn}$, which leads to the asymptotic inefficiency in Corollary 4.

[2]This bound is a frequently-used bound for the first-order statistic of normals, which implies that, for low n, it is not a sharp bound for the statics of variables following a standard normal distribution. Indeed, the censored distribution always has a higher mean and first order stochastically dominates the uncensored distribution.