9-2012

The Noisy Secretary Problem and Some Results on Extreme Concomitant Variables

Abba M. Krieger
University of Pennsylvania

Ester Samuel-Cahn
Hebrew University of Jerusalem

Recommended Citation

Krieger, A. M., & Samuel-Cahn, E. (2012). The Noisy Secretary Problem and Some Results on Extreme Concomitant Variables. *Journal of Applied Probability, 49* (3), 821-837. http://dx.doi.org/10.1017/S0021900200009566

The Noisy Secretary Problem and Some Results on Extreme Concomitant Variables

Abstract
The classical secretary problem for selecting the best item is studied when the actual values of the items are observed with noise. One of the main appeals of the secretary problem is that the optimal strategy is able to find the best observation with the nontrivial probability of about 0.37, even when the number of observations is arbitrarily large. The results are strikingly different when the quality of the secretaries are observed with noise. If there is no noise, then the only information that is needed is whether an observation is the best among those already observed. Since observations are assumed to be i.i.d. this is distribution free. In the case of noisy data, the results are no longer distrubtion free. Furthermore, one needs to know the rank of the noisy observation among those already seen. Finally, the probability of finding the best secretary often goes to 0 as the number of obsevations, n, goes to infinity. The results depend heavily on the behavior of pn, the probability that the observation that is best among the noisy observations is also best among the noiseless observations. Results involving optimal strategies if all that is available is noisy data are described and examples are given to elucidate the results.

Keywords
Optimal stopping rule, best choice secretary problem, noisy data

Disciplines
Business | Statistics and Probability

The Noisy Secretary Problem and Some Results on Extreme Concomitant Variables

Abba M. Krieger * and Ester Samuel-Cahn†

February 1, 2012

Abstract

The classical secretary problem for selecting the best item is studied when the actual values of the items are observed with noise. One of the main appeals of the secretary problem is that the optimal strategy is able to find the best observation with the nontrivial probability of about 0.37, even when the number of observations is arbitrarily large. The results are strikingly different when the quality of the secretaries are observed with noise. If there is no noise, then the only information that is needed is whether an observation is the best among those already observed. Since observations are assumed to be i.i.d. this is distribution free. In the case of noisy data, the results are no longer distrubtion free. Furthermore, one needs to know the rank of the noisy observation among those already seen. Finally, the probability of finding the best secretary often goes to 0 as the number of obsevations, n, goes to infinity. The results depend

*Postal address: Department of Statistics of the Wharton School of the University of Pennsylvania, Philadelphia 19104, PA. Email address: krieger@wharton.upenn.edu. Supported by funds from the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem.

†Postal address: Department of Statistics and Center for Rationality, The Hebrew University of Jerusalem, Jerusalem 91905 Israel. E-mail address: scahn@mscc.huji.ac.il. Supported by the Israel Science Foundation Grant 467/04.

heavily on the behavior of p_n, the probability that the observation that is best among the noisy observations is also best among the noiseless observations. Results involving optimal strategies if all that is available is noisy data are described and examples are given to elucidate the results.

Keywords: Optimal stopping rule, best choice secretary problem, noisy data

1 Introduction

The "Best Choice Secretary Problem" is classical because it is surprising that there is a rule which enables finding the best secretary with non-zero probability even if the number of secretaries that are considered is arbitrarily large. But what happens if the qualities of the secretaries at the time of decision are only known subject to noise? This paper considers various aspects of this problem. First, the optimal rule when there is no noise is no longer optimal when measurements are made with noise. Second, in many cases the probability of finding the best secretary now goes to zero, albeit slowly, in the number of secretaries, n, that are considered. Third, the results are sensitive to distributional assumptions, unlike the classical secretary problem, and there are distributions for which the probability of finding the best goes to the same limit as in the noiseless case, and other distributions where the probability goes to 0, as n goes to infinity.

In the classical "Best Choice Secretary Problem" the underlying assumption is that ranks are sequentially obtained from n i.i.d. continuous random variables, X_i, $i = 1, \ldots, n$. This total number, n, is called the horizon and is assumed known. Only the relative ranks,

$$RR\left(X_i\right) := \sum_{j=1}^{i} \mathbb{1}_{[X_i \leq X_j]}$$

are observed. The goal is to maximize the probability of picking the X_i which is maximal, i.e. the i for which the absolute rank

$$AR\left(X_i\right) := \sum_{j=1}^{n} \mathbb{1}_{[X_i \leq X_j]},$$

equals 1. The well-known optimal solution is to let a certain number, $N(n)$, go by, and pick the first item thereafter, $i > N(n)$, for which $RR\left(X_i\right) = 1$. If no such i exists stop at n, anyway. It is well-known that $N(n)/n$ tends to e^{-1} as $n \to \infty$, and the optimal probability, W_n, of picking the best tends to e^{-1} as $n \to \infty$. See e.g. Gilbert and Mosteller (1966) or S. Samuels (1991) with generalizations in Bruss (2000) and Gnedin (2007). When applying this rule it suffices to know whether the present item is relatively best, i.e., if $RB(X_i) = 1$, where $RB(X_i) = 1$ if $RR(X_i) = 1$ and $RB(X_i) = 0$ otherwise. Clearly the solution to this classical problem is distribution free.

In the present paper we consider the case where the relative ranks are not those of the X_i's themselves, but of i.i.d. Y_i's, where $Y_i = X_i + \epsilon_i$, and the ϵ_i are i.i.d. noise (or error) variables, independent of the X's. The goal is the same as before, viz., to maximize the probability of selecting the i for which the X-value is maximal. The optimal rule, and the optimal probability of picking the best X_i are no longer distribution free. Denote the optimal probability when $RR(Y_i)$ are known by $W_n(X, \epsilon)$, where n is the known horizon, and the optimal probability of selecting the best when only the $RB(Y_i)$ are known, by $W_n^*(X, \epsilon)$.

If one uses the classical rule on the noisy data then clearly there is a probability that goes to e^{-1} of finding the best Y_i, as $n \to \infty$. But if one finds the best Y has one found the best X which is what is desired? The difference between the classical secretary problem (i.e., without noise) and the noisy secretary problem depends heavily on the value of p_n, where

$$p_n = p_n(X, \epsilon) = \mathbb{P}\left(\arg\max_{i \le n} X_i = \arg\max_{i \le n} Y_i\right), \tag{1.1}$$

i.e., p_n is the probability that the location of the maximal X is the same as the location of the maximal Y. The behavior of p_n, as mentioned above, is crucial to the values of $W_n(X, \epsilon)$ and $W_n^*(X, \epsilon)$.

The main results in the present paper about finding the best X from noisy data are:

S1. For any X and ϵ, if the observed values are the $RB(Y_i)$ only, $W_n^*(X, \epsilon) \ge W_n p_n$.

S2. If only the RB's are observed, the optimal value $N(n, X, \epsilon)$ after which one should pick the first item for which $RB(Y_i) = 1$, satisfies $N(n, X, \epsilon) \leq N(n)$, i.e., one should stop earlier than in the classical case.

S3. If the $RR(Y_i)$ are sequentially available, it is no longer optimal to base the stopping rule on the $RB(Y_i)$'s only.

Some of the results depend on the probability that the best X in n items is the m^{th} best Y. To this end, for any i let

$$\begin{aligned} p_{nm} &= \mathbb{P}\left(AR(Y_i) = m \mid AR(X_i) = 1\right) = \mathbb{P}\left(AR(X_i) = 1 \mid AR(Y_i) = m\right) \\ &= \sum_{i=1}^{n} \mathbb{P}\left([AR(X_i) = 1] \cap [AR(Y_i) = m]\right). \end{aligned} \tag{1.2}$$

S4. The optimal rule, which can in principle be found by backward induction once p_{nm} is determined is of the form: There exist integer values $1 \leq k_1 \leq \cdots \leq k_n = n$, not necessarily distinct, such that one should stop with the smallest i such that $i < k_j$ and $RR(Y_i) < j$.

S5. $\lim_{n\to\infty} W_n(X, \epsilon) = 0$ if and only if $\lim_{n\to\infty} p_n(X, \epsilon) = 0$.

Results S1 through S5 are proved in Section 4.

In the next section, we consider the important example where X and ϵ are Normally distributed. This example elucidates points S1, S2, S3, and S4 above and point P1 below. As is apparent from the above list of results and will be even more apparent from the example, the $p_n(X, \epsilon)$ values play an important role in the above statements. These values are also of intrinsic interest. Hence we discuss $p_n(X, \epsilon)$ in Section 3. Denote the distribution of X by F with density f and the distribution of ϵ by G with density g. We show

P1. Suppose that $\sup\{x : F(x) < 1\} = \infty$, and that the $\lim_{x\to\infty} \frac{f(x+d)}{f(x)}$ exists for all $d > 0$. Then a necessary and sufficient condition for $\lim_{n\to\infty} p_n(X, \epsilon) = 0$ for *all* distributions G is that $\lim_{x\to\infty} \frac{f(x+d)}{f(x)} = 0$ for every fixed d.

P2. If $\sup\{x : F(x) < 1\} = c < \infty$ then $\lim_{n\to\infty} p_n = 0$ for all G.

P3. For any given F there exists a distribution G such that $\lim_{n\to\infty} p_n(X, \epsilon) = 0$.

P4. There exist distributions F and G such that $\lim_{n\to\infty} p_n(X, \epsilon) = 1$.

P5. The $p_n(X, \epsilon)$ values are *not* necessarily monotone in n.

Additional examples are given in Section 5. Because noise (errors) are often assumed to be normal, special attention is given to the case where G is normal. The examples include cases, such as the exponential, Pareto with parameter 1, and the case where both F and G are normal.

2 Normal-Normal Example

In order to illustrate the results, we consider the case where $X_i \sim \mathcal{N}\left(0,\, \rho^2\right)$ and $Y_i = X_i + \epsilon_i$ where $\epsilon_i \sim \mathcal{N}\left(0,\, 1-\rho^2\right)$ and all X_i and ϵ_i are independent. Hence, the concomitant variable, Y_i (c.f., David and Nagaraja (2003)) is $\mathcal{N}(0,\, 1)$ with correlation of ρ with X_i.

There are two kinds of results mentioned in the introduction. One result considers the behavior of p_n, the probability that the index for which Y_i is maximum agrees with the index for which X_i is maximum. Since a normal distribution satisfies the condition that $\lim_{x\to\infty} \frac{f(x+d)}{f(x)} = 0$ for all d, p_n goes to zero (see P1). Table 1 shows how p_n varies as a function of n and ρ. The values in Table 1 are found by simulation with $10,000$ replications.

It is interesting to note how sensitive p_n is to ρ. In fact, in Ledford and Tawn (1998) it is shown that

$$\lim_{n\to\infty} D_n p_n = \left(\Gamma\left(\frac{1}{1+\rho}\right)\right)^2 \tag{2.1}$$

where $D_n = (1-\rho^2)^{1/2}\{(4\pi \log n)^{\rho} n^{1-\rho}\}^{\frac{1}{1+\rho}}$.

Table 1: Probability that the Observation with Largest X also yields the Largest Y

	n				
ρ	10	50	100	1000	10000
0.5	0.291	0.138	0.107	0.044	0.018
0.6	0.353	0.191	0.155	0.076	0.039
0.7	0.423	0.261	0.222	0.125	0.073
0.8	0.519	0.360	0.324	0.210	0.151
0.9	0.647	0.512	0.481	0.370	0.303

We next illustrate by simulation with $10,000$ replications what occurs with a secretary-like decision rule. This depends on p_{nm} in (1.2), the probability that the m^{th} largest Y corresponds to the observation that has the largest X value. We estimated (see Table 2) this quantity by simulation with $10,000$ replications. If one employed the secretary rule, with $N(n) = n/e$, with $n = 10,000$

Table 2: Probability that index for which the best X value is attained yields a Y value that is m^{th} absolute best: $n = 10,000$; $\rho = 0.9$

m	1	2	3	4	5	6	7	8	9	10
p_{nm}	0.303	0.137	0.083	0.065	0.050	0.036	0.032	0.028	0.022	0.018

and $\rho = 0.9$, one would find the largest X, observing only $RB(Y_i)$, with probability of 0.138, as compared to W_n of approximately $e^{-1} = 0.368$, the value if there was no noise (or equivalently $\rho = 1$). The probability of 0.138 is greater than $W_n p_n = 0.303e^{-1} = 0.112$. The reason is that the secretary rule, when it stops at the relative best Y, might be stopping at an observation that is, say, the second best Y in absolute rank. The second best absolute rank of Y has a probability of

0.137 of being the observation with best X. This would add to the probability that the rule chooses a Y with best X; in fact, the probability of $0.138 - 0.112 = 0.026$ is attributable to stopping at a Y which ultimately is not the best Y, but corresponds to the X which is the best.

As mentioned in S2 of the Introduction, choosing $N(n) = n/e$ in the secretary-like rule might not be the best choice. In fact, it is stated that the value of $N(n, X, \epsilon) \leq N(n)$. In the present example, it is found that $N(10,000, X, \epsilon) = 2,740 < 3,678 = 10,000/e$. When $n = 10,000$ and $\rho = 0.9$, the optimal secretary-like rule has probability equal to 0.141 of finding the best X.

In order to show that the secretary rule is not necessarily optimal when there is noise, as mentioned in S4 above, we consider $n = 5$ items. The optimal classical secretary rule on the Y values has $N(5) = 2$ (i.e., two items are allowed to pass before selection). This results in stopping at the i^{th} best Y with respective probabilities of $\frac{13}{30}, \frac{7}{30}, \frac{4}{30}, \frac{3}{30}, \frac{3}{30}$. On the other hand, if we apply the same rule, but also stop at the next to last item if it is either the relative best or second best then this results in stopping at the i^{th} best Y with respective probabilities of $\frac{12}{30}, \frac{9}{30}, \frac{5}{30}, \frac{2}{30}, \frac{2}{30}$. These probabilities are derived by simple calculation.

The five probabilities for each of the cases need to be weighted by the probability that the i^{th} best Y corresponds to the best X. We find the respective probabilities when $\rho = 0.5$ to be $0.4110, 0.2490, 0.1675, 0.1104$, and 0.0621. These probabilities are found by simulation with ten million replications to ensure accuracy. Finally, the optimal secretary rule finds the best X with probability 0.2758 as compared to 0.2785 if we stop with relative rank of two on the next to the last observation, conditional on not having stopped earlier.

3 Probability That The Best Concomitant Observation is the Best Observation

In the present section we prove P1 through P5 of the Introduction.

Let $X_1, \ldots, X_n$ and $\epsilon_1, \ldots, \epsilon_n$ be independent random variables, where the X_i's are i.i.d. with distribution F, and ϵ_i are i.i.d with distribution G. Let $Y_i = X_i + \epsilon_i$. Our interest is in the behavior of p_n given in (1.1).

Let $X^n_{(1)} \geq X^n_{(2)} \geq \ldots \geq X^n_{(n)}$ be the order statistics of $X_1, \ldots, X_n$. Let $Y^n_{[j]} = X^n_{(j)} + \epsilon^*_j$, where the ϵ^*_j are i.i.d. and a random permuation of $\epsilon_1, \ldots, \epsilon_n$. The variables $Y^n_{[j]}$ are called the concomitant random variables, i.e. the random variable Y_i that ″belongs to″ $X^n_{(j)}$.

For any (cumulative) distribution H, let $x_H = sup\{x : H(x) < 1\}$.

Theorem 3.1. *If F is such that for every fixed $c > 0$ and fixed integer k*

$$\lim_{n\to\infty} \mathbb{P}\left(X^n_{(1)} - X^n_{(k)} < c\right) = 1, \tag{3.1}$$

then $\lim_{n\to\infty} p_n = 0$, *for all* G.

Proof. Let $n > k$, and fix c of (3.1).

$$p_n = \mathbb{P}\left(Y^n_{[1]} = \max_{i=1,\ldots,n} Y_i\right) < \mathbb{P}\left(Y^n_{[1]} > \max_{j=2,\ldots,k} Y^n_{[j]}\right) = \mathbb{P}\left(X^n_{(1)} + \epsilon^*_1 > \max_{j=2,\ldots,k}(X^n_{(j)} + \epsilon^*_j)\right). \tag{3.2}$$

We shall show that for any $\delta > 0$ and n sufficiently large, $p_n < \delta$. Let x_0 be such that $G(x_0) = 1 - \delta/4$.

$$\mathbb{P}\left(\max_{j=2,\ldots,k} \epsilon^*_j < \epsilon^*_1 + c\right) \leq \int_{-\infty}^{x_0} \Big[G(x+c)\Big]^{k-1} g(x)dx + \delta/4 < \Big[G(x_0+c)\Big]^{k-1} + \delta/4 < \delta/2, \tag{3.3}$$

provided one chooses k large enough for $\Big[G(x_0+c)\Big]^{k-1} < \delta/4$. If $x_G = \infty$ this is always possible, but if $x_G < \infty$ one may have to repalce the original c by a smaller value, c_0, such that $G(x_0+c_0) < 1$.

Now using (3.1) with c and k satisfying (3.3), pick n sufficiently large for

$$\mathbb{P}\left(X^n_{(1)} - X^n_{(k)} < c\right) > 1 - \delta/2 \text{ for all } n > N(\delta).$$

Let A_n denote the event $\{X^n_{(1)} - X^n_{(k)} < c\}$ and $\overline{A}_n$ its complement. Then, for that c and k we can continue (3.2)

$$\begin{aligned}
\mathbb{P}\left(X^n_{(1)} + \epsilon^*_1 > \max_{j=2,\dots,k}(X^n_{(j)} + \epsilon^*_j)\right) &\leq \mathbb{P}\left(X^n_{(1)} + \epsilon^*_1 > \max_{j=2,\dots,k}(X^n_{(j)} + \epsilon^*_j \mid A_n)\right)\mathbb{P}(A_n) + \mathbb{P}\left(\overline{A}_n\right) \\
&< \mathbb{P}\left(X^n_{(1)} + \epsilon^*_1 > X^n_{(1)} - c + \max_{j=2,\dots,k}\epsilon^*_j\right) + \delta/2 \\
&= \mathbb{P}\left(\epsilon^*_1 + c > \max_{j=2,\dots,k}\epsilon^*_j\right) + \delta/2 < \delta/2 + \delta/2 = \delta, \qquad (3.4)
\end{aligned}$$

where the last inequality in (3.4) uses (3.3). □

Corollary 3.1. *P2 holds.*

Proof. If $x_F < \infty$ then clearly (3.1) holds. □

Theorem 3.2. *A sufficient condition for* $\lim_{n\to\infty} p_n(X, \epsilon) = 0$ *for all G is that* $x_F = \infty$ *and that for any fixed* $d > 0$

$$\lim_{x\to\infty} \mathbb{P}(X \geq x + d \mid X > x) = 0 \qquad (3.5)$$

or, equivalently

$$\lim_{x\to\infty} f(x + d)/f(x) = 0. \qquad (3.6)$$

Proof. We shall show that (3.5) implies (3.1). For a given c and k let $d = c/(k-1)$. Suppose that (3.5) holds, and let x_0 be such that

$$\mathbb{P}(X \geq x + d \mid X > x) < \frac{\delta}{2(k-1)} \quad \text{for all } x > x_0. \qquad (3.7)$$

Let N be sufficiently large, such that for all $n \geq N$

$$\mathbb{P}\left(X^n_{(k)} > x_0\right) > 1 - \delta/2. \qquad (3.8)$$

Let B_n be the event $\{X^n_{(k)} > x_0\}$. Then

$$\mathbb{P}\left(X^n_{(1)} - X^n_{(k)} < c\right) = \mathbb{P}\left(\sum_{i=1}^{k-1}(X^n_{(i)} - X^n_{(i+1)}) < c\right)$$

$$
\begin{aligned}
&> \mathbb{P}\left(\cap_{i=1}^{k-1}\{X_{(i)}^n - X_{(i+1)}^n < d\}\right) \\
&> \mathbb{P}\left(\cap_{i=1}^{k-1}\{X_{(i)}^n - X_{(i+1)}^n < d\} \mid B_n\right)\mathbb{P}(B_n) \\
&> \left[1 - \sum_{i=1}^{k-1}\mathbb{P}\left(\{X_{(i)}^n - X_{(i+1)}^n \geq d\} \mid B_n\right)\right](1-\delta/2) \qquad \text{by (3.8)} \\
&> \left(1-(k-1)\frac{\delta}{2(k-1)}\right)(1-\delta/2) > 1-\delta \qquad \text{by (3.7).}
\end{aligned}
$$

Since $\delta > 0$ was arbitrarily small, (3.1) holds, and the result follows.

Since (3.5) can be written as

$$
\lim_{x\to\infty}(1-F(x+d))/(1-F(x)),
$$

we can, by L'Hopital's rule, take the limit of derivatives, which yields (3.6). □

Theorem 3.3. *Let* $x_F = \infty$ *and* $d > 0$*, and assume that* $\lim_{x\to\infty}\mathbb{P}(X \geq x+d \mid X > x)$ *exists and is equal to* a*, where* $a > 0$*. Then there exists a* G*, and* $\epsilon \sim G$ *such that* $\liminf_{n\to\infty}\ p_n(X,\epsilon) > a-\gamma$ *for any* $\gamma > 0$*.*

This theorem establishes the necessary statement of P1.

Proof. Fix $\gamma > 0$ and let $\delta = \gamma/(1+a)$. There exists an x_0 such that $P(X \geq x+d \mid X > x) \geq a-\delta$ for all $x > x_0$. Since $X_{(2)}^n$ goes to infinity with probability one as $n \to \infty$ (and it is stochastically increasing in n), there exists an N and $x_1 > x_0$ such that $P(X_{(2)}^n > x_1) \geq 1-\delta$ for all $n > N$. Let ϵ have a uniform distribution on $[0,d]$. We make use the following result which is straightforward to verify. Let $X_1,...,X_n$ be i.i.d. continuous random variables with distribution F. Then

$$
\mathbb{P}\left(X_{(1)}^n \geq x+d \mid X_{(2)}^n = x\right) = \mathbb{P}(X \geq x+d \mid X > x),
$$

where $X \sim F$. Hence,

$$
p_n(X,\epsilon) \;=\; \mathbb{P}\left(X_{(1)}^n + \epsilon_1^* > \max_{j=2,\ldots,n}(X_{(j)}^n + \epsilon_j^*)\right) \geq \mathbb{P}\left(X_{(1)}^n \geq X_{(2)}^n + d\right)
$$

$$
\begin{aligned}
&= \int_{w=-\infty}^{\infty} \mathbb{P}\left(X^n_{(1)} \geq w + d \mid X^n_{(2)} = w\right) f_{X^n_{(2)}}(w)dw \\
&\geq \int_{w=x_1}^{\infty} \mathbb{P}\left(X^n_{(1)} \geq w + d \mid X^n_{(2)} = w\right) f_{X^n_{(2)}}(w)dw \\
&\geq (a-\delta)\mathbb{P}\left(X^n_{(2)} > x_1\right) \geq (a-\delta)(1-\delta) > a - \gamma. \qquad (3.9)
\end{aligned}
$$

□

In Section 5 we consider two examples, Example 5.1 and 5.2, where F has an Exponential distribution. Depending on G, $\lim_{n\to\infty} p_n = 0$ or $\liminf_{n\to\infty}\ p_n > 0$. The Exponential distribution is of special interest, since for it

$$
\mathbb{P}\left(X > x + d \mid X > x\right) = e^{-d},
$$

i.e., independent of x, and thus can be considered as a borderline case. For many well known distributions, if (3.5) fails, the limit in the left hand side of (3.5) will be one.

To show P3 we need the following lemma.

Lemma 3.1. *For any continuous distribution F and X_i i.i.d. distributed F, there exists a distribution H and i.i.d. Z_i, independent of the X_i's, distributed H, such that*

$$
\lim_{n\to\infty} \mathbb{P}\left(X^n_{(1)} > Z^n_{(1)}\right) = 0. \qquad (3.10)
$$

Proof. We shall first prove the statement where F is the uniform distribition on $[0,1]$. Then for any continuous H

$$
\mathbb{P}\left(X^n_{(1)} > Z^n_{(1)}\right) = n\int_0^1 [H(x)]^n x^{n-1} dx.
$$

Now let $H(x) = 1 - (1-x)^{1/2}$ for $0 \leq x \leq 1$. We shall show that (3.10) holds.

$$
\begin{aligned}
\mathbb{P}\left(X^n_{(1)} > Z^n_{(1)}\right) &= n\int_0^1 [1-(1-x)^{1/2}]^n x^{n-1} dx \\
&= 2n\int_0^1 (1-y)^n (1-y^2)^{n-1} y dy \qquad (3.11)
\end{aligned}
$$

where we have made the change of variable, $y = (1-x)^{1/2}$. Now the last expression in (3.11) equals

$$\begin{aligned} &2n \int_0^{n^{-2/3}} (1-y^n)(1-y^2)^{n-1} y dy + 2n \int_{n^{-2/3}}^1 (1-y)^n (1-y^2)^{n-1} y dy \\ &< 2n \int_0^{n^{-2/3}} y dy + 2n \int_{n^{-2/3}}^1 (1-y)^n dy = n^{-1/3} + \frac{2n}{n+1} \left(1-n^{-2/3}\right)^{n+1}. \end{aligned} \tag{3.12}$$

The first term in the right hand side of (3.12) clearly tends to 0. The last term is less than $2(1-n^{-2/3})^n = 2\left[(1-1/m)^m\right]^{\sqrt{m}}$, where $m = n^{2/3}$, which also goes to 0, since the value in the bracket is arbitrarily close to e^{-1}. This proves (3.10) for F uniformily distributed on $[0,1]$.

To generalize, add $*$ to all of the previous variables, i.e., $X_1^*, X_2^*, \ldots, Z_1^*, Z_2^*, \ldots$ and H^*, and note that H^* has all of its mass on $[0,1]$. Now consider any continuous F, and its inverse F^{-1}. Let $X_i = F^{-1}(X_i^*)$ and $Z_i = F^{-1}(Z_i^*)$. Then the X_i's are i.i.d. with distribution F and Z_i are i.i.d. with distribution $H(x) = H^*(F(x))$. But since $F^{-1}(x)$ is monotone increasing, clearly

$$\mathbb{P}\left(X_{(1)}^n > Z_{(1)}^n\right) = \mathbb{P}\left(X_{(1)}^{*n} > Z_{(1)}^{*n}\right)$$

and (3.10) follows. □

Theorem 3.4. *For any F there exists G such that* $\lim_{n\to\infty} p_n(X,\epsilon) = 0$.

Proof. If $x_F < \infty$ the result follows from Corollary 3.1. Thus assume $x_F = \infty$.

We consider first $P(X_i \geq 0) = 1$. Note that ϵ_i^* is the ϵ that is associated with the i^{th} largest X_i. Then $Y_{[1]}^n = X_{(1)}^n + \epsilon_1^*$. We choose G so that $\epsilon^* > 0$. We want to show $p_n := P(Y_{[1]}^n = max_{j=1,\ldots,n} Y_j) \to 0$. Now

$$\begin{aligned} \mathbb{P}\left(Y_{[1]}^n > max_{j=2,\ldots,n} Y_{[j]}\right) &= \mathbb{P}\left(X_{(1)}^n + \epsilon_1^* > max_{j=2,\ldots,n,}(X_{(j)} + \epsilon_j^*)\right) \\ &\leq \mathbb{P}\left(X_{(1)}^n + \epsilon_1^* > max_{j=2,\ldots,n} \epsilon_j^*\right) = \mathbb{P}\left(X_{(1)}^n > \epsilon_{(1)}^{n-1} - \epsilon_1^*\right) \\ &\leq \mathbb{P}\left(X_{(1)}^n > \epsilon_{(1)}^n - 2\epsilon_1^*\right), \end{aligned} \tag{3.13}$$

where the first inequality uses $X_j \geq 0$, and the second inequality uses $\epsilon_1^* \geq 0$ and $\epsilon_{(1)}^n \leq \epsilon_{(1)}^{n-1} + \epsilon_1^*$. Consider Z_i of Lemma 3.1. Since $X_{(1)}^n \to \infty$ a.s. as $n \to \infty$ it follows that $Z_{(1)}^n \to \infty$, and we may take $Z_i \geq 0$. Let $\delta > 0$. For proper choice of G we shall show that $p_n < \delta$ for all n sufficiently large. Let $\epsilon_i = 2Z_i$, where Z_i satsfies (3.10), and let z_0 be a constant such that $P(4Z_i > z_0) < \delta/4$, and let n be so large that $P(Z^n_{(1)} < z_0) < \delta/4$. With this choice we can continue the inequality in (3.13), obtaining

$$\begin{aligned}
\mathbb{P}\left(X^n_{(1)} > 2Z^n_{(1)} - 4Z_{i_n}\right) &< \delta/4 + \mathbb{P}\left(X^n_{(1)} > 2Z^n_{(1)} - z_0\right) \\
&< \delta/4 + \mathbb{P}\left(Z^n_{(1)} \leq z_0\right) + \mathbb{P}\left(X^n_{(1)} > Z^n_{(1)}\right) \\
&< \frac{2\delta}{4} + \mathbb{P}\left(X^n_{(1)} > Z^n_{(1)}\right).
\end{aligned}$$

Now, by (3.10) one can take n sufficiently large for the last term to be less than $\delta/4$, thus $p_n < \frac{3}{4}\delta$ for all n sufficiently large.

If X_i can take on negative values, but is bounded from below by some $c < 0$, shift X_i by c to obtain $\hat{X}_i = X_i - c \geq 0$. Let $\hat{Y}_j = \hat{X}_j + \epsilon_j$. Then

$$\begin{aligned}
\hat{p}_n &:= \mathbb{P}\left(\hat{Y}^n_{[1]} > max_{j=2,\dots,n}\hat{Y}_{[j]}\right) = \mathbb{P}\left(\hat{X}^n_{(1)} + \epsilon^*_1 > max_{j=2,\dots n}(\hat{X}_{(j)} + \epsilon^*_j)\right) \\
&= \mathbb{P}\left(X^n_{(1)} - c + \epsilon^*_1 > max_{j=2,\dots,n}(X_{(j)} + \epsilon^*_j) - c\right) \\
&= \mathbb{P}\left((X^n_{(1)} + \epsilon^*_1 > max_{j=2,\dots,n}(X_{(j)} + \epsilon^*_j)\right) = p_n.
\end{aligned}$$

But $\hat{X}_i \geq 0$, so if ϵ_i are chosen so that $\hat{p}_n \to 0$, the same ϵ_i will do for the original X_i, and $p_n < \frac{3}{4}\delta$ for all n sufficiently large.

Now consider X_i which are not bounded below. Take n sufficiently large so that

$P(X^n_{(1)} < c) < \delta/4$. On $\{X^n_{(1)} \geq c\}$ we may replace X_i by $\tilde{X}_i = max(X_i, c)$ which are bounded, and obtain $p_n < \delta/4 + \frac{3}{4}\delta = \delta$. □

Note that Theorem 3.4 establishes P3. P4 is established through Example 5.4 where

$F(x) = (1 - x^{-1})\mathbb{1}_{[x>1]}$ and G is $N(0, 1)$. P5 also follows from that example.

4 Results for the Noisy Secretary Problem

In this section, we prove or illustrate the five results in the Introduction, labeled S1 through S5. There are two versions of the secretary problem in the presence of noise. In one problem, we only observe whether a noisy observation is the relative best Y. In another problem, we observe $RR(Y_i)$, that is the relative rank of the noisy observation amongst observations we observed so far. It is important to note that if we observed X_i, or equivalently there was no noise, then there would not be a distinction between these two problems. If there is no noise, then it is obvious that one should not stop if an observation is not the relative best.

The first result, S1, relates the probability of finding the best X in two versions, that is knowing $RB(X_i)$ as compared to knowing $RB(Y_i)$. S2 considers the rule that only uses $RB(Y_i)$. Specifically, let a certain number of observations go by and then stop at the first i such that $RB(Y_i) = 1$. The main finding is that it is optimal to let fewer observations go by when the data are noisy than in the classical secretary problem (where it is optimal to let approximately n/e observations go by). S3 indicates that the two versions of the problem do not necessarily have the same solution. This is shown by example.

In the last two results we consider the problem where the relative ranks of the noisy data are available. An algortihm that produces the optimal solution for this problem, based on dynamic programming, is described in the discussion to S4 in a similar treatment as in Ferguson (2008). Finally, it is shown that the probability of finding the best X goes to zero, when and only when $\lim_{n\to\infty} p_n(X, \epsilon) = 0$. This is in contrast to the classical secretary problem where the probability of choosing the best X goes to e^{-1}. The results that follow depend on

$$q_{nijk} = \mathbb{P}\left(AR(Y_i) = k \mid RR(Y_i) = j\right), \tag{4.1}$$

for $1 \le j \le i$ and $j \le k \le n + j - i$. This probability, which is negative hypergeometric requires that the first i items include $j - 1$ of the observations that have absolute rank of at most $k - 1$ and the remaining $i-j$ items from among those with absolute rank exceeding k. Hence,

$$q_{nijk} = \frac{\binom{k-1}{j-1}\binom{n-k}{i-j}}{\binom{n}{i}}. \tag{4.2}$$

This is given in Ferguson (2008, Chapter 2 page 2.4).

In order to show that S1 holds, suppose one uses the simple classical rule which maximizes the probability of finding the maximal Y. The probability of finding the maximal Y is W_n. The probability that this is also the maximal X is p_n. In addition, this rule may pick a Y-value which turns out not to be the maximal Y, but still could be the Y-value that corresponds to the maximal X, i.e., $Y^n_{[1]}$. Thus, with this rule, one achieves a value which is at least $W_n p_n$. As this rule may not be optimal, (see S2), an optimal rule may achieve an even higher value.

To show S2, we consider optimal rules assuming we only know whether an observation is the relative best. These rules can be characterized by an integer $S(n)$ which implies that the stopping time is the first time that $RB(Y_i) = 1$ for $i > S(n)$. This is akin to the elegant result in Bruss (2000) where it is shown how to obtain the secretary rule by summing odds. The difference is that now we do not observe the variables I_i that indicate whether we have a relative record among the X values at the i^{th} observation. Rather we observe the noisy data, which indicates whether we have a relative record among the Y values. We prove the following theorem that relates the best secretary rule in the classical problem to that in the noisy problem:

Theorem 4.1. *Let $N(n)$ be the number of observations in the classical secretary rule such that we stop the first time, i, for which $i > N(n)$ and $RB(X_i) = 1$. The optimal value $N(n, X, \epsilon)$ after which one should pick the first item for which $RB(Y_i) = 1$ in the noisy case satisfies $N(n, X, \epsilon) \leq N(n)$.*

Let $S^*(n) = N(n, X, \epsilon)$, denote the optimal stopping rule in the noisy case, which depends on the horizon, n, and the distributions of X and ϵ. Let $N(n) \approx n/e$, be the analog to $S^*(n)$ in the classical secretary rule, which is based on $RB(X_i)$. Then the above theorem shows that $S^*(n) \leq N(n)$ and hence one should stop no later when there is noise, than when there is no noise.

Proof. Consider the rule in the noisy case where S (for $1 \le S \le n-1$) items are allowed to go by and we stop at the first $i > S$ for which $RB(Y_i) = 1$ (otherwise stop at n, anyway). Let

$$r_{in} = \mathbb{P}\left(AR(X_i) = 1 \mid RB(Y_i) = 1\right) = \sum_{k=1}^{n-i+1} q_{ni1k} p_{nk}, \tag{4.3}$$

where q_{ni1k} is given in (4.1) and p_{nk} is given in (1.2). Hence, the probability that this rule chooses the best X is

$$\mathbb{P}\left(S\right) := \sum_{i=S+1}^{n} \frac{1}{i} \frac{S}{i-1} r_{in},$$

where the $1/i$ term is the probability that Y_i has relative rank of 1 and $S/(i-1)$ is the probability that the best in the first $i-1$ observations is among the first S items (so that one does not stop before the i^{th} observation).

Let $r^*_{in} = \frac{r_{in}}{i/n}$. This implies that

$$\mathbb{P}\left(S\right) = \frac{S}{n} \sum_{i=S+1}^{n} \frac{r^*_{in}}{i-1}.$$

Consider

$$\begin{aligned}
\mathbb{P}\left(S+1\right) - \mathbb{P}\left(S\right) &= \frac{S+1}{n} \sum_{i=S+2}^{n} \frac{r^*_{in}}{i-1} - \frac{S}{n} \sum_{i=S+1}^{n} \frac{r^*_{in}}{i-1} \\
&= \frac{S+1}{n} \sum_{i=S+2}^{n} \frac{r^*_{in}}{i-1} - \frac{S}{n} \Big[\sum_{i=S+2}^{n} \frac{r^*_{in}}{i-1} + \frac{r^*_{S+1,n}}{S} \Big] \\
&= \frac{1}{n} \Big\{ \sum_{i=S+2}^{n} \frac{r^*_{in}}{i-1} - r^*_{S+1,n} \Big\}.
\end{aligned} \tag{4.4}$$

The above expression is nonnegative if and only if

$$\sum_{i=S+2}^{n} \frac{1}{i-1} \frac{r^*_{in}}{r^*_{S+1,n}} \ge 1. \tag{4.5}$$

Note that for the classical secretary problem $r^*_{in} = 1$.

If we can show that r^*_{in} decreases as i increases, then the optimal S must necessarily be smaller than the corresponding value for the classical secretary problem, as desired. Note that p_{nk} does not depend on i and as i increases, the number of terms in (4.3) decreases. Hence it is sufficient to show that $\frac{q_{ni1k}}{i/n}$ decreases in i for any n and k. But $q_{ni11} = \frac{i}{n}$ since we need the best Y to be among the first i items. In general, for $k > 1$,

$$q_{ni1k} = \frac{\binom{n-k}{i-1}}{\binom{n}{i}}$$

Hence,

$$\frac{q_{ni1k}}{i/n} = \frac{(n-k)!}{(n-1)!} \frac{(n-i)!}{(n+1-i-k)!}$$

The above expression clearly decreases in i for any $k > 1$ and n. □

We provide an example that shows that S3 holds, that is, it is better to stop in some cases with $RR(Y_i) > 1$. The smallest such example for which this can occur is $n = 4$, where it might be better to take the third observation if it is second best among the first three Y-values. What follows is such an example involving exponential random variables.

Example 4.1: The solution based on RR is better than the solution based on RB.

Let X_i be i.i.d. Exponential with $\mu = 1$ and ϵ_i be i.i.d. Exponential with $\mu = c$, where μ is the mean. Let $Y^3_{[i]} = X^3_{(i)} + \epsilon^*_i$. We want to show that it is better to stop at $n = 3$ if $RR(Y_3) = 2$. To this end, we need to consider

$$\gamma_{ci} = \mathbb{P}\left(Y^3_{[1]} = Y^3_{(i)}\right) \text{ for } i = 1, 2, 3.$$

Specifically, since there is a $\frac{3}{4}$ chance that the best X in four observations is among the first three observations, we need to show that $3\gamma_{c2}/4 > \frac{1}{4}$ or $\gamma_{c2} > \frac{1}{3}$.

Lemma 4.1. *Let c be the mean in the Exponential distribution of the epsilons. Then,*

$$\gamma_{c1} = 1 + \frac{2c^2}{3(2c+1)(c+2)} - \frac{c(4c+1)}{2(2c+1)(c+1)},$$

$$\gamma_{c2} = \frac{c(4c+1)}{2(2c+1)(c+1)} - \frac{4c^2}{3(2c+1)(c+2)}.$$

$$\gamma_{c3} = \frac{2c^2}{3(2c+1)(c+2)}.$$

Before we prove the lemma we observe that if $c = 3$, then $\gamma_{c1} = \frac{133}{280}$, $\gamma_{c2} = \frac{99}{280}$ and $\gamma_{c3} = \frac{48}{280}$. Since γ_{c2} is approximately $0.35357 > \frac{1}{3}$ this is an example where it is better to stop at $n = 3$ if we observe the second largest Y-value, from among the first three Y-values. The largest value that γ_{c2} can achieve is 0.36275. This occurs when $c = 5.535$.

Proof. We begin with γ_{c3} as it is the easiest and highlights the argument. The probability of interest is

$$\mathbb{P}\left([X^3_{(1)} + \epsilon^*_1 < X^3_{(2)} + \epsilon^*_2] \cap [X^3_{(1)} + \epsilon^*_1 < X^3_{(3)} + \epsilon^*_3]\right).$$

First note that since the X_i are i.i.d. standard exponential random variables, we can express the resulting order statistics as: $X^3_{(3)} = E_3$, $X^3_{(2)} = E_3 + E_2$ and $X^3_{(1)} = E_3 + E_2 + E_1$, where E_i are independent exponential random variables with mean of $1/i$. In order to evaluate the above probability, consider two events: $A(x,y) = \{X^3_{(2)} + \epsilon^*_2 > X^3_{(1)} + \epsilon^*_1 \mid ([\epsilon^*_1 = x] \cap [E_1 = y])\}$ and similarly $B(x,y) = \{X^3_{(3)} + \epsilon^*_3 > X^3_{(1)} + \epsilon^*_1 \mid ([\epsilon^*_1 = x] \cap [E_1 = y])\}$. First, $A(x,y)$ and $B(x,y)$ are independent conditional on E_1. This follows since $X^3_{(1)} - X^3_{(2)} = E_1$, hence $A(x,y)$ only depends on ϵ^*_2, x, and y and $X^3_{(1)} - X^3_{(3)} = E_1 + E_2$, hence $B(x,y)$ only depends on E_2, ϵ^*_3, x and y. Second, $P[A(x,y)] = e^{-(y+x)/c}$ and

$$P[B(x,y)] = \int_{v=0}^{\infty} e^{-(y+x+v)/c} 2e^{-2v} dv = \frac{2ce^{-(y+x)/c}}{2c+1}.$$

Therefore,

$$\gamma_{c3} = \frac{2c}{2c+1} \int_{x=0}^{\infty} \int_{y=0}^{\infty} e^{-2(y+x)/c} e^{-x} \frac{1}{c} e^{-y/c} dy dx = \frac{2c^2}{3(2c+1)(c+2)}.$$

To obtain γ_{c1} we use a similar argument. Hence,

$$\gamma_{c1} = \int_{x=0}^{\infty} \int_{y=0}^{\infty} P[\overline{A(x,y)}] P[\overline{B(x,y)}] e^{-x} \frac{1}{c} e^{-y/c} dy dx.$$

Finally, $\gamma_{c2} = 1 - (\gamma_{c1} + \gamma_{c3})$. □

In S4, the method for finding the optimal rule when relative ranks are observed is described. The optimal rule, which can in principle be found by backward induction, is of the form: There exist integer values $0 \leq k_1 \leq \cdots \leq k_n = n$ not necessarily distinct, such that one should stop with the smallest i such that $RR(Y_i) \leq k_i$. The obvious way to proceed, which we programmed, is by backward induction. Once we determine p_{nm} by simulation, the backward induction, which we outline below, is distribution free.

At observation i, we need to decide whether we should stop or not if $RR(Y_i) = j$, for $1 \leq j \leq i$. If we were to stop when $RR(Y_i) = j$, then the probability that $AR(X_i) = 1$ is

$$f_{nij} = \sum_{v=j}^{n-i+j} q_{nijv} p_{nv}.$$

i.e. f_{nij} is the probability that X_i is the best among all of the n X-observations, conditional on Y_i being the j^{th} best from among the first i Y-observations.

We need to keep track of R_{ni}, which is the probability of getting the best X if the optimal rule is followed from observation i and thereafter. To complete the discussion we need to show how R_{ni} is determined recursively, beginning with the last observation, n, and going backwards.

Note that $R_{nn} = \frac{1}{n}$. For any i, let k_i be the largest j such that $f_{nij} > R_{n,i+1}$. Then,

$$R_{ni} = \frac{1}{i} \sum_{j=1}^{k_i} f_{nij} + R_{n,i+1} \frac{(i - k_i)}{i}.$$

The form of the solution, as claimed above, that the maximum $RR(Y_i)$ at which we would stop at observation i is non-decreasing in i, is intuitive. It also follows from the solution described above since R_{ni} is non-increasing in i, p_{nj} is clearly non-increasing in j, and there exists a k_0 which depends on i such that $q_{nijk} \leq q_{n,i+1,j,k}$ only when $k \leq k_0$. The last two statements imply that f_{nij} increases as i increases.

S5 is straightforward. No rule can be better than the rule that finds the observation that is best amongst the Y values with certainty. But if $\lim_{n \to \infty} p_n(X, \epsilon) = 0$, even this rule satisfies

$\lim_{n\to\infty} W_n(X,\epsilon) = 0$. The converse follows from S1.

5 Examples

In this section, we consider four examples to illustrate interesting findings concerning the behavior of p_n. It follows from (2.1) that when X is Normal and ϵ is Normal the probability that the best concomitant observation is the best observation goes to 0, if $\rho < 1$. In Example 5.1, we show that the probability does not go to zero if the X distribution is exponential.

Example 5.1: F =Exponential, G =Normal, and $\liminf_{n\to\infty}\ p_n > 0$.

We want to know how likely it is that the largest among Y_i has the same index as the largest among X_i.

Hence we need to consider

$$\mathbb{P}\left(X^n_{(1)} + \epsilon^*_1 > X^n_{(j)} + \epsilon^*_j, j = 2, \ldots, n\right)$$

where ϵ^*_j are i.i.d. standard normal. Let $A_j = \{X^n_{(1)} - X^n_{(j)} > \epsilon^*_j - \epsilon^*_1\}$. Note that A_j depends on n, but for ease of notation, we do not include n as a superscript, because the value of n remains fixed in this argument. We want to show that $P(A_2 \cap A_3 \cap ... \cap A_n)$ goes to a constant greater than zero.

It suffices to show that $P(A_2 \cap A_3 \cap ... \cap A_n \mid \epsilon^*_1 = z)$ goes to a constant greater than zero as $n \to \infty$ for any z. This is the case since $P(A_2 \cap A_3 \cap ... \cap A_n) = \int_z P(A_2 \cap A_3 \cap ... \cap A_n \mid \epsilon^*_1 = z)\phi(z)dz$ where ϕ is the density of the standard normal. Since the conditional probability in the integral increases in z, $P(A_2 \cap A_3 \cap ... \cap A_n) \geq P(A_2 \cap A_3 \cap ... \cap A_n \mid \epsilon^*_1 = z)P(Z > z)$ where Z is a standard normal random variable independent of the E_is.

But $P(A_2 \cap A_3 \cap ... \cap A_n \mid \epsilon^*_1 = z) = 1 - P(B_1 \cup \ldots \cup B_{n-1}) \geq 1 - \sum_{j=1}^{n-1} P(B_j)$ where $B_j = \{X_{(1)} - X_{(j+1)} \leq \epsilon^*_{j+1} - z\}$.

To evaluate $P(B_j)$ note that for i.i.d. exponential, $E_i = X^n_{(i)} - X^n_{(i+1)}$ are independent exponential with mean of $1/i$. Hence $P(B_j) = P(E_1 + \ldots + E_j + Z \leq -z)$.

It suffices to show that $\sum_{i=1}^{n-1} P(B_i) < 1$ as $n \to \infty$. We use the following Chernoff bound: $P(H \leq a) \leq e^{-ta} M(t)$ for all $t < 0$ where M is the moment generating function of H. The moment generating function of the random variable $E_1 + \ldots + E_j + Z$ is

$$M_j(t) = e^{t^2/2} \prod_{i=1}^{j} \frac{i}{i-t}.$$

We now choose $z = -a$ to be 2 and t to be -2. This implies that

$$P(B_j) \leq e^{-4} e^2 \prod_{i=1}^{j} \frac{i}{i+2} = \frac{2e^{-2}}{(j+1)(j+2)}.$$

Hence

$$\sum_{j=1}^{n-1} P(B_j) \leq 2e^{-2} \sum_{j=1}^{n-1} \frac{1}{(j+1)(j+2)} = e^{-2}[1 - 2/(n+1)] < 1.$$

We showed that when X and ϵ have normal distributions p_n goes to zero. This is intuitive because the difference between the largest relatively few X values are arbitrarily close to each other in probability as n gets large. In the exponential case, however, the expected difference between the largest and second largest observation is one and hence the largest X values do not become indistiguishable as in the normal case. Nevertheless, Example 5.2 shows that when the distributions of X and ϵ are both exponential, and hence the error term is sufficiently large, $\lim_{n\to\infty} p_n(X, \epsilon) = 0$.

Example 5.2: F =Exponential, G =Exponential and $\lim_{n\to\infty} p_n = 0$.

Assume we observe $X_1, \ldots,$ and $\epsilon_1, \ldots,$ as i.i.d. exponential with equal mean, without loss of generality taken to be one. Let $Y_i = X_i + \epsilon_i$. Let $x_N = X^N_{(1)}$ and $y_N = Y^N_{(1)}$. Let $n > N$ be an observation after N. The probability that we have a record at n in at least one of the X sequence or Y sequence is

$$\mathbb{P}\left(\left[X_n > x_N\right] \cup \left[Y_n > y_N\right]\right).$$

We want to determine

$$\mathbb{P}\left(\left[X_n > x_N\right] \cap \left[Y_n > y_N\right] \mid \left[X_n > x_N\right] \cup \left[Y_n > y_N\right]\right)$$

$$= \frac{P(\left[X_n > x_N\right] \cap \left[Y_n > y_N\right])}{P(\left[X_n > x_N\right] \cup \left[Y_n > y_N\right])} \leq \frac{P(\left[X_n > x_N\right] \cap \left[Y_n > y_N\right])}{\max\left[P(X_n > x_N), P(Y_n > y_N)\right]}. \quad (5.1)$$

Since X_n is exponential with mean one and Y_n is Gamma (2,1) it follows that

$$\mathbb{P}\left(X_n > x_N\right) = e^{-x_N}$$

and

$$\mathbb{P}\left(Y_n > y_N\right) = (1 + y_N)e^{-y_N}.$$

Furthermore,

$$\begin{aligned} \mathbb{P}\left(\left[X_n > x_N\right] \cap \left[Y_n > y_N\right]\right) &= \int_{u=x_N}^{u=y_N} (e^{-u})(e^{u-y_N})du + \int_{u=y_N}^{\infty} e^{-u}du \\ &= e^{-y_N}(y_N - x_N) + e^{-y_N} \\ &= e^{-y_N}(y_N - x_N + 1). \end{aligned} \quad (5.2)$$

For any $\delta > 0$, let

$$y(\delta) = \min\left\{y \middle| \frac{\log(1+y)+1}{y+1} < \delta/2\right\}.$$

Let N be sufficiently large so that $P(Y^m_{(1)} > y(\delta)) > 1 - \delta/2$ for all $m > N$. We need to consider two cases:

1. If $e^{-x_N} > (1 + y_N)e^{-y_N}$, or equivalently $e^{y_N - x_N} > (1 + y_N)$ then the right hand side of (5.1) is $e^{-y_N + x_N}(y_N - x_N + 1) < \frac{\log(1+y_N)+1}{1+y_N} < \delta/2$. The next to last inequality follows becuase $e^{-u}(u+1)$ is a decreasing function.

2. If $e^{-x_N} \leq (1 + y_N)e^{-y_N}$ or equivalently $e^{y_N - x_N} \leq (1 + y_N)$ then the right hand side of (5.1) is $\frac{y_N - x_N + 1}{y_N + 1} \leq \frac{\log(1+y_N)+1}{1+y_N} < \delta/2$.

Finally, if N is sufficiently large then

$$
\begin{aligned}
& \mathbb{P}\left(\left[X_n > x_N\right] \cap \left[Y_n > y_N\right] \mid \left(\left[X_n > x_N\right] \cup \left[Y_n > y_N\right]\right)\right) \\
\leq \; & \mathbb{P}\left(\left[X_n > x_N\right] \cap \left[[Y_n > y_N\right] \mid \left\{\left[X_n > x_N\right] \cup \left[Y_n > y_N\right]\right\} \cap \left[Y_{(1)}^n > y(\delta)\right]\right)(1-\delta/2)+\delta/2 \\
< \; & (\delta/2)(1-\delta/2)+\delta/2 < \delta.
\end{aligned}
$$

One might conjecture that if X and ϵ have the same distributions then p_n goes to zero as in the normal and exponential cases. But if the tail of X is sufficently fat then p_n need not go to zero. The intuition is that the largest X is likely to be a lot larger than the second largest X. This is in essence, what is shown in the following example.

Example 5.3: F =Pareto, G = Pareto and $\liminf_{n\to\infty}\ p_n > 0$.

Let $F(x) = G(x) = 1 - 1/x^\alpha$ for $x \geq 1$. (Note that for $\alpha = 1$, the tail behavior of this distribution is the same as that of a Cauchy distribution.)

We want to show that $\liminf_{n\to\infty}\ p_n > 0$.

Claim: If we can show that

$$\mathbb{P}\left(X_{(1)}^n > X_{(2)}^n + \epsilon_{(2)}^n\right) > \delta \tag{5.3}$$

for some $\delta > 0$, as $n \to \infty$, we are done. The reason for this is that if the second largest ϵ is with the second largest X then the only other possible observation that has higher Y than the Y with index corresponding to the largest X is the one with the highest ϵ. But it is just as likely that the index with the highest X and the index with the highest ϵ has the highest Y value.

Note that $X_i^\alpha = 1/U_i$ where the U_i are i.i.d. uniform (0,1) and similarly let $\epsilon_i^\alpha = 1/V_i$ where V_i are i.i.d. uniform (0,1).

We will show that $\lim P(X_{(1)}^n - X_{(2)}^n > (n+1)/2) > 0$ and that $\lim P(\epsilon_{(2)}^n < (n+1)/2) > 0$

from which (5.3) follows. But

$$\begin{aligned}
&\mathbb{P}\left([X^n_{(1)} > (n+1)^{1/\alpha}] \cap [X^n_{(2)} < ((n+1)/2)^{1/\alpha}]\right) \\
&= \mathbb{P}\left([U^n_{(n)} < 1/(n+1)] \cap [U^n_{(n-1)} > 2/(n+1)]\right) \\
&= n\Big[\frac{1}{n+1}\Big(1-\frac{2}{n+1}\Big)^{n-1}\Big] \to e^{-2}
\end{aligned}$$

where $U^n_{(i)}$ is the i^{th} largest uniform that generates $X_{(n+1-i)}$. Note that $(n+1)^{1/\alpha} - ((n+1)/2)^{1/\alpha} = ((n+1))^{1/\alpha}/c$ where $c = \frac{2^{1/\alpha}}{2^{1/\alpha}-1}$. Finally, since $P(\epsilon^n_{(2)} < ((n+1)/c)^{1/\alpha})$ is $P(V^n_{(n-1)} > c/(n+1))$ goes to $(c+1)e^{-c}$, where $V^n_{(i)}$ is the i^{th} largest uniform that generates $\epsilon^n_{(n+1-i)}$, we are done.

It is somewhat intuitive that as n increases, the probability that the observation that is the largest X is also the largest Y, decreases. But that is not necessarily the case. In fact, the next example says more. The above probability in this example goes to one as n goes to infinity.

Example 5.4: F =Pareto, $G =$ Normal, and $\lim_{n\to\infty} p_n = 1$.

Let $X_1, \ldots, X_n$ as i.i.d. with $f(x)$ as in Example 5.3 with $\alpha = 1$, that is, $F(x) = 1 - 1/x$ for $x \geq 1$. Assume we observe $Y_i = X_i + \epsilon_i$ where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. and normally distributed.

Claim: $P(X^n_{(1)} - X^n_{(2)} \geq z_n) \to 1$ as $n \to \infty$ where $z_n = n^{1-\delta}$ for any $\delta > 0$.

Since the maximum of n Normally distributed random variables is of order $(\log n)^{1/2}$ for large n the above claim shows that the largest from among the X distribution must also be the largest from among the Y distribution with probability tending to 1.

Proof of claim: Let $X^n_{(i)} = \frac{1}{U^n_{(n+1-i)}}$.

$$\begin{aligned}
P\Big(X^n_{(1)} - X^n_{(2)} \geq z_n\Big) &= P\Big(\frac{1}{U^n_{(n)}} - \frac{1}{U^n_{(n-1)}} \geq z_n\Big) \\
&= P\Big(\frac{1}{U^n_{(n)}} \geq z_n + \frac{1}{U^n_{(n-1)}}\Big) = P\Big(U^n_{(n)} \leq \frac{U^n_{(n-1)}}{1+z_n U^n_{(n-1)}}\Big) \\
&= \int_{t=0}^{1} P\Big(U_{(n)} \leq \frac{U^n_{(n-1)}}{1+z_n U^n_{(n-1)}} \mid U^n_{(n-1)} = t\Big) f_{n-1}(t)dt \\
&= \int_{t=0}^{1} \frac{1}{1+z_n t} f_{n-1}(t)dt
\end{aligned}$$

where f_i is the density for the i^{th} largest order statistics from a uniform (0,1) distribution which is Beta($\alpha = n+1-i$,$\beta = i$). But $U_{(n-1)} \sim$ Beta($\alpha = 2$, $\beta = n-1$). Hence $E(U^n_{(n-1)}) = \frac{2}{n+1}$ and $Var(U^n_{(n-1)}) = \frac{2(n-1)}{(n+1)^2(n+2)} = O\big(\frac{1}{n^2}\big)$. So $P(U^n_{(n-1)} < \frac{1}{n^{1-\delta/2}}) \to 1$ as $n \to \infty$. This implies that

$$\begin{aligned} P(X^n_{(1)} - X^n_{(2)} \geq n^{1-\delta}) &\geq \frac{1}{1+n^{1-\delta}n^{-(1-\delta/2)}} P\Big(U^n_{(n-1)} < \frac{1}{n^{1-\delta/2}}\Big) \\ &= \frac{1}{1+n^{-\delta/2}} P\Big(U^n_{(n-1)} < \frac{1}{n^{1-\delta/2}}\Big) \to 1 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Remark 5.1: The assumption that G has a normal distribution in Example 5.4 is easily relaxed. All that is needed in the proof is that maximum of n random variables be $o(n^{1-\delta})$ for any $\delta > 0$.

Remark 5.2: Note that Example 5.4 establishes P5. Clearly, for small n the value of p_n here is not equal to 1. But if $\lim_{n\to\infty} p_n = 1$, the p_n sequence cannot be monotone.

Acknowledgment: The authors would like to thank Moshe Pollak for invaluable insights he provided during the early stages of this research. The authors would also like to thank an anonymous referee for extremely helpful comments.

References

[1] Bruss, F.T.(2000). Sum the odds to one and stop. *The Annals of Probability* ,28, 1384-1391.

[2] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics* . Third Edition. Johm Wiley & Sons, Inc., Jersey.

[3] Ferguson, T.S.(2008).*Optimal Stopping and Applications*. Electronic text http://www.math.ucla.edu/˜ tom?Stopping?contents.html

[4] Gnedin, A.V. (2007). Optimal stopping with rank-dependent loss *Journal of Applied Probability*, 44, 996-1011.

[5] Gilbert, J.P. and Mosteller F.(1966). Recognizing the maximum of a sequence. *Journal of the American Statistical Association*, 61, 35-73.

[6] Ledford, A.W. and Tawn, J.A. (1998). Concomitant tail behavior for extremes. *Advances of Applied Probability*, 30, 197-215.

[7] Samuels, S.M. (1991). *Handbook of sequential analysis, Editors: B. K. Ghosh and P. K. Sen* Statist. Textbooks Monogr., Dekker, New York, 118, 381-405.