ANALOGS OF RECORDS: RELATIVE SEQUENTIAL SELECTIONS — RELAXED OR CONSTRAINED

PEICHAO PENG AND J. MICHAEL STEELE

ABSTRACT. We consider selection processes that can provide either a multiplicative relaxation or a multiplicative constriction of the classical process of selected records in a sequence of independent observations. In the relaxed case, we find that the number of selections satisfies a CLT with a different normalization than Rényi's classical CLT for records, and in the constricted case we find convergence to an unbounded random variable with all moments. We also find refinements of some classical facts for the number of records in an independent sample, including an exact formula for the expected number of records given a specified level for the zero'th record. Further we note that a Markov chain central to our analysis has a stationary distribution that satisfies a non-autonomous pantographic equation.

KEY WORDS. records, central limit theorem, Rényi's CLT for records, sequential selection, Dobrushin coefficient, pantograph equation, non-homogenous Markov chains.

MATHEMATICS SUBJECT CLASSIFICATION (2010). Primary: 60C05, 60G40, 90C40; Secondary: 60F99, 90C27, 90C39

1. RELAXED OR CONSTRAINED SEQUENTIAL SELECTION PROCESSES

Let $X_i,\ i = 1, 2, \ldots$ be a sequence of independent random variables with a common continuous distribution F with support in $[0, \infty)$, and let ρ denote a non-negative constant. If we then set $\tau_1 = 1$ and define a sequence of stopping times by taking

$$\tau_k = \min\{j : X_j \geq \rho X_{\tau_{k-1}}\} \quad \text{for } k > 1, \tag{1}$$

then the random variables of main interest here are then given by

$$R_n(\rho) = \max\{k : \tau_k \leq n\}. \tag{2}$$

When $\rho = 1$ the times τ_k, $k = 1, 2, \ldots$ are precisely the times at which new record values are observed, and $R_n(1)$ is the total number of records that are observed in the time interval $[1 : n] = \{1, 2, \ldots, n\}$.

The random variable $R_n(1)$ has been well understood for a long time. In particular, Rényi (1962) found among other things that $\mathbb{E}[R_n(1)] \sim \log n$ and $\text{Var}[R_n(1)] \sim \log n$; moreover, he found that one has

$$\frac{R_n(1) - \log n}{\sqrt{\log n}} \Rightarrow N(0, 1), \tag{3}$$

Peichao Peng: Department of Statistics, The Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA, 19104. Email address: ppeichao@wharton.upenn.edu.
J. M. Steele: Department of Statistics, The Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA, 19104. Email address: steele@wharton.upenn.edu.

where, as usual, the symbol $\Rightarrow$ denotes convergence in distribution and $N(0,1)$ denotes the standard Gaussian distribution.

The cases $\rho \in (0,1)$ and $\rho \in (1,\infty)$ seem not to have been considered previously, and they lead to some novel phenomena. First, the distribution of $R_n(\rho)$ depends on F while in Rényi's case it does not. Moreover, one finds that in both of the new cases the asymptotic behavior is unlike the behavior that one finds for Rényi's classical record process $\{R_n(1) : n \geq 1\}$.

The most interesting case is when $\rho \in (0,1)$ where, in comparison to the classical record process, one has *relaxed* the condition that is imposed on sequential selections. In this case, one again has a central limit theorem, but the underlying process differs notably from Rényi's. In particular, the mean and variance grow linearly when $\rho \in (0,1)$, and the summands are no longer independent.

Here, for the sake of brevity, we say a distribution function is in the *selection class* $\mathcal{S}_L$ if there is an $L \in (0,\infty)$ such that $F(0) = 0$, $F(L) = 1$, and F is continuous and strictly monotone on $(0,L)$. For example, the uniform distribution on $[0,1]$ is in $\mathcal{S}_1$, and for any $L > 0$ the truncated exponential distribution $F(x) = (1-e^{-x})/(1-e^{-L})$ is in $\mathcal{S}_L$. For these examples, one has a density, but, in general, a distribution in $\mathcal{S}_L$ need not have a density.

Theorem 1 (Mean, Variance, and CLT when $0 < \rho < 1$). *If X_i, $i = 1,2,\ldots$ are independent and if $F \in \mathcal{S}_L$, then there are constants $\mu_\rho(F) > 0$ and $\sigma_\rho(F) > 0$ such that*

$$\mathbb{E}[R_n(\rho)] \sim n\mu_\rho(F) \quad \textit{and} \quad \mathrm{Var}[R_n(\rho)] \sim n\sigma_\rho^2(F), \tag{4}$$

and one has a central limit theorem

$$\frac{R_n(\rho) - n\mu_\rho(F)}{\sigma_\rho(F)\sqrt{n}} \Rightarrow N(0,1). \tag{5}$$

After we develop some useful connections with the theory of Markov chains in Sections 2 and 3, the proof of Theorem 1 is given in Section 4. The main issues are the proofs of the relations (4) and the proof of $\sigma_\rho^2(F) > 0$. Once these facts are in hand, the convergence (5) then follows from general theory; for example, one can obtain (5) directly from Arlotto and Steele (2016) Theorem 1, Corollary 2. Alternatively, with a page or two of extra work, one can obtain (5) by first generalizing other known central limit theorems for additive functionals of Markov processes. In either case, the proof that $\sigma_\rho^2(F) > 0$ presents itself as the make-or-break step.

For a general $F \in \mathcal{S}_L$, the task of determining the constants $\mu_\rho(F)$ and $\sigma_\rho(F)$ seems intractable. Nevertheless, in leading case when F is the uniform distribution U on $[0,1]$, there is an explicit series formula for the mean.

Theorem 2 (Moments for Uniform Distribution). *If $\rho \in (0,1)$ and if the random variables X_i, $i = 1,2,\ldots$ have the uniform distribution U on $[0,1]$, then we have*

$$\begin{aligned}\mu_\rho(U) = 1 - \frac{\rho}{2} - \frac{\rho}{3}\left(\rho - \frac{\rho^2}{2}\right) - \frac{\rho}{4}\left(\rho - \frac{\rho^2}{2}\right)\left(\rho - \frac{\rho^3}{3}\right) \\ - \frac{\rho}{5}\left(\rho - \frac{\rho^2}{2}\right)\left(\rho - \frac{\rho^3}{3}\right)\left(\rho - \frac{\rho^4}{4}\right)\cdots\end{aligned} \tag{6}$$

The proof of Theorem 2 is given in Section 5 and 6 where we also develop an equation of the pantograph type for the stationary distribution of the Markov chain given by the selected values. We do not solve this equation, but we use it to derive the Mellin transform for the stationary distribution. This is used in turn to get our formula (6) for $\mu_\rho(U)$. There is little hope of finding a correspondingly explicit representation for $\sigma^2_\rho(F)$ even for $F = U$. Nevertheless, we do find in Section 4 that there is a more abstract series representation (26) for $\sigma^2_\rho(F)$.

When $\rho > 1$, one no longer has a central limit theorem. Instead one has almost sure convergence to an unbounded non-negative random variable that has a well-behaved moment generating function.

Theorem 3 (Distributional Limit when $\rho > 1$)**.** *If $F \in \mathcal{S}_L$ and if $F(x) = O(x)$ in the neighbourhood of* 0*, then for each $\rho > 1$ there exists an unbounded random variable N_ρ with moment generating function*

$$\mathbb{E}[\exp(sN_\rho)] < \infty \quad \textit{for } |s| < \log \rho, \tag{7}$$

such that with probability one $R_n(\rho) \nearrow N_\rho$ as $n \to \infty$.

Sequential selection with $\rho > 1$ is notably less interesting than the cases with $\rho = 1$ or $\rho \in (0, 1)$ where one finds central limit theorems of two different kinds. Nevertheless, completeness calls for the consideration of $\rho > 1$, and we give a brief analysis of this case in Section 7.

Section 8 then develops several refinements of Rényi's classic formula for the expected number of records. For example, consider the number $R^x_n(1)$ of records that are larger than x. When F is the uniform distribution on $[0, 1]$, we find

$$\mathbb{E}[R^x_n(1)] = \sum_{k=1}^{n} \frac{1 - x^k}{k} = H_n - \sum_{k=1}^{n} \frac{x^k}{k}. \tag{8}$$

This formula recaptures Rényi's classic harmonic sum when we set $x = 0$, yet the proof of (8) has nothing in common with the classic arguments of Rényi (1962). Moreover, the methods that lead one to (8) yield further generalizations for the quantities $\mathbb{E}[R^x_n(\rho)]$ and $\lim_n\{\mathbb{E}[R^x_n(\rho)] - \mathbb{E}[R^y_n(\rho)]\}$.

In Section 9 we examine more fully the senses in which the values chosen by the selection process (1) can be viewed as relaxed records, constrained records, or creatures of another breed. In particular, we note that in the relaxed case $\rho \in (0, 1)$, the selected values can differ greatly from any notion of approximate record, even though our sequential selection processes and various approximate record processes both contain the record process as a limiting cases when the time horizon is finite. Finally, in Section 10, we make brief note of three further connections between the theory of relative sequential selections and Rényi's theory of records.

2. Representation as a Markov Additive Functional

For $k = 1, 2, \ldots$ we take Y_k to be the last value that has been accepted by the selection process during the time interval $[1 : k]$; that is, we set

$$Y_k = X_{\tau_j} \quad \text{where } j = \max\{m : \tau_m \le k\}.$$

The values Y_k, $k = 1, 2, \ldots$ determine a Markov chain where if one is in state x then one stays in state x with probability $F(\rho x)$ and with probability $1 - F(\rho x)$ one moves to a point y in the set $[\rho x, 1] \setminus \{x\}$ that is chosen according to the probability measure $dF(y)/(1-F(\rho x))$. In other words, if we also set $Y_0 = 0$ then the process $\{Y_k : k \in [0:\infty)\}$, has the transition kernel

$$K_{\rho,F}(x,A) = F(\rho x)\mathbb{1}(x \in A) + \int_{\rho x}^{L} \mathbb{1}(y \in A)\, dF(y). \tag{9}$$

Now, in terms of the Markov chain $\{Y_k : 0 = 1, 2, \ldots\}$ we have the representation

$$R_n(\rho) = \sum_{k=1}^{n} \mathbb{1}[Y_k \neq Y_{k-1}], \tag{10}$$

since we accept a new value precisely at the times when the state of the chain $\{Y_n\}$ changes. Most of Theorem 1 follows from this representation after we establish a few analytic properties of the Markov chain $\{Y_n\}$.

Remark 4. Here one should note that by definition $R_n(\rho)$ is a function of the independent random variables $\{X_1, X_2, \ldots X_n\}$, and we simply write $\mathbb{E}[R_n(\rho)]$ and $\mathrm{Var}[R_n(\rho)]$ when $R_n(\rho)$ is viewed in this way. On the other hand, by the representation (10), we can also view $R_n(\rho)$ as a function of $\{Y_1, Y_2, \ldots, Y_n\}$ and the distribution of this sequence depends on the initial distribution of the Markov chain. When we take the second point of view it is natural (and necessary) to write $\mathbb{E}_\mu[R_n(\rho)]$ and $\mathrm{Var}_\mu[R_n(\rho)]$ whenever Y_0 has the distribution μ. By construction, we always have $\mathrm{Var}[R_n(\rho)] = \mathrm{Var}_0[R_n(\rho)]$ and $\mathbb{E}[R_n(\rho)] = \mathbb{E}_0[R_n(\rho)]$.

3. The Dobrushin Coefficient and Its Consequences

There are several ways one can investigate the Markov chain defined by (9), but here it is especially efficient to first estimate its Dobrushin coefficient.

Definition 5 (Dobrushin Coefficient)**.** If K is a Markov transition function on a Borel state space $\mathcal{X}$ and if $\mathcal{B}(\mathcal{X})$ denotes the collection of Borel subsets of $\mathcal{X}$, then the *Dobrushin coefficient* $\delta(K)$ of the kernel K is defined by

$$\delta(K) = \sup_{x_1, x_2 \in \mathcal{X},\, A \in \mathcal{B}(\mathcal{X})} |K(x_1, A) - K(x_2, A)|.$$

Lemma 6 (Dobrushin Coefficient for $K_{\rho,F}$)**.** *If $F \in \mathcal{S}_L$ and if $K_{\rho,F}$ is the transition kernel given by* (9)*, then one has*

$$\delta(K_{\rho,F}) \leq F(\rho L) < 1. \tag{11}$$

Proof. If we assume $x_1 < x_2$, then for any Borel set $A \subset [0, L]$ we have from (9) that

$$\begin{aligned}
\Delta &\stackrel{\text{def}}{=} K_{\rho,F}(x_1, A) - K_{\rho,F}(x_2, A) \\
&= F(\rho x_1)\mathbb{1}(x_1 \in A) + \int_{\rho x_1}^{1} \mathbb{1}(y \in A)\, dF(y) - F(\rho x_2)\mathbb{1}(x_2 \in A) - \int_{\rho x_2}^{1} \mathbb{1}(y \in A)\, dF(y) \\
&= F(\rho x_1)\mathbb{1}(x_1 \in A) - F(\rho x_2)\mathbb{1}(x_2 \in A) + \int_{\rho x_1}^{\rho x_2} \mathbb{1}(y \in A)\, dF(y).
\end{aligned}$$

After majorizing the positive terms, we see from monotonicity of F that

$$\Delta \leq F(\rho x_1) + \{F(\rho x_2) - F(\rho x_1)\} = F(\rho x_2) \leq F(\rho L).$$

On the other hand, if we keep just the one negative term in the sum for Δ, then we have

$$\Delta \geq -F(\rho x_2) \geq -F(\rho L),$$

and these two bounds on Δ complete the proof of (11). $\square$

Nagaev (2015) proved that for any Markov chain with kernel K and Dobrushin coefficient $\delta(K) < 1$, there is a probability measure ν on the state space $\mathcal{X}$ that is stationary under K, and, most notably, if $K^{(n)}$ denotes the n step transition kernel, then one has the total variation bound

$$|K^{(n)}(x, A) - \nu(A)| \leq 2[\delta(K)]^n \quad \text{for all } x \in \mathcal{X} \text{ and } A \in \mathcal{B}(\mathcal{X}). \tag{12}$$

Now, we let $\{Z_n : n = 0, 1, 2, \ldots\}$ be the Markov chain associated with the kernel K, and we write $\mathbb{E}_x$ and $\mathbb{E}_\nu$ for the corresponding expectation operators where accordingly as $Z_0 = x \in \mathcal{X}$ or $Z_0 \sim \nu$. By the total variation bound (12) and approximation by step functions, one can then check that for any bounded measurable $g : D = \mathcal{X}^4 \to \mathbb{R}$, one has for any fixed $0 \leq i \leq j \leq k$ and $n \to \infty$ that

$$\mathbb{E}_x[g(Z_n, Z_{n+i}, Z_{n+j}, Z_{n+k})] - \mathbb{E}_\nu[g(Z_0, Z_i, Z_j, Z_k)] = O(||g||_\infty[\delta(K)]^n), \tag{13}$$

where here we set $||g||_\infty = \sup_{v \in D} |g(v)|$.

The implied constant in (13) is absolute; in fact, it can be taken to be 4. Naturally, we also have analogous relations for functions of fewer than four variables or more than four variables. Here we only need (13) and its analog for functions of two variables.

4. When $\rho < 1$: Proof of Theorem 1

We now restrict attention to the Markov chain with transition kernel $K_{\rho,F}(\cdot, \cdot)$ given by (9). By the bound (11) we have $\delta \equiv \delta(K_{\rho,F}) < 1$, so the stationary distribution exists. We consider two initial distributions: in the first case we take $Y_0 \equiv 0$; and in the second case we assume that Y_0 has the stationary distribution ν. By the two variable analog of (13) for $g(Y_n, Y_{n+1}) = \mathbb{1}(Y_n \neq Y_{n+1})$ we have

$$\mathbb{E}_0[\mathbb{1}(Y_{k-1} \neq Y_k)] = \mathbb{E}_\nu[\mathbb{1}(Y_0 \neq Y_1)] + O(\delta^k). \tag{14}$$

From (14) and the representation (10), we see by geometric summation that

$$\begin{aligned} \mathbb{E}_0[R_n(\rho)] &= \sum_{k=1}^{n} \mathbb{E}_0[\mathbb{1}(Y_{k-1} \neq Y_k)] = n\mathbb{E}_\nu[\mathbb{1}(Y_0 \neq Y_1)] + O(1) \\ &= n \int_{\mathcal{X}} \int_{\mathcal{X}} \mathbb{1}[s \neq t] K_{\rho,F}(s, dt)\, d\nu(s) + O(1), \end{aligned} \tag{15}$$

where the double integral is just $\mathbb{E}_\nu[\mathbb{1}(Y_0 \neq Y_1)]$ written in longhand. This gives us the first assertion (4) of Theorem 1 in a form that is a bit more explicit; in particular, (15) tells us that in (4) we have

$$\mu_\rho(F) = \int_{\mathcal{X}} \int_{\mathcal{X}} \mathbb{1}[s \neq t] K_{\rho,F}(s, dt)\, d\nu(s). \tag{16}$$

To find the asymptotic variance of $R_n(\rho)$, we introduce two sequences of random variables:

$$U_k = \mathbb{1}[Y_{k-1} \neq Y_k] - \mathbb{E}_0(\mathbb{1}[Y_{k-1} \neq Y_k]) \text{ and } V_k = \mathbb{1}[Y_{k-1} \neq Y_k] - \mathbb{E}_\nu(\mathbb{1}[Y_{k-1} \neq Y_k]).$$

Both U_k and V_k are functions of the Markov process $\{Y_k : k = 0, 1, \ldots\}$, so in particular, both $\{U_n\}$ and $\{V_n\}$ depend on the initial value Y_0. For clarity one should note that U_k has mean zero when $Y_0 \equiv 0$ and V_k has mean zero when Y_0 follows the stationary distribution ν.

The random variables U_k and V_k differ by a *constant* that depends on k, and by (14) the constant is not larger than $O(\delta^k)$. Thus, by the representation (10), we have

$$\text{Var}[R_n(\rho)] = \text{Var}_0[R_n(\rho)] = \mathbb{E}_0\Big[\big(\sum_{k=1}^{n} U_k\big)^2\Big] = \mathbb{E}_0\Big[\big(\sum_{k=1}^{n} V_k\big)^2\Big] + O(1). \tag{17}$$

Now, we expand the sum in (17) and write

$$\mathbb{E}_0\Big[\big(\sum_{k=1}^{n} V_k\big)^2\Big] = \sum_{k=1}^{n} \mathbb{E}_0[V_k^2] + 2\sum_{i=1}^{n-1}\sum_{j=1}^{n-i} \mathbb{E}_0[V_i V_{i+j}] \stackrel{\text{def}}{=} A_n + B_n. \tag{18}$$

To estimate the first sum A_n of (18), we apply (13) just as we did in the derivation of (14), and this time we find

$$\mathbb{E}_0[V_k^2] = \mathbb{E}_\nu[V_k^2] + O(\delta^k) = \mathbb{E}_\nu[V_1^2] + O(\delta^k).$$

Summation then gives us

$$A_n = n\mathbb{E}_\nu[V_1^2] + O(1). \tag{19}$$

To deal with the double sum B_n, we first need a lemma to help us estimate some of the summands of B_n.

Lemma 7. *For any initial distribution μ one has*

$$\mathbb{E}_\mu[V_i V_{i+j}] = O(\delta^j) \quad \textit{for all } i, j \geq 0. \tag{20}$$

Proof. To exploit the Markov property for the chain $\{Y_n : n = 0, 1, \ldots\}$, we first condition on Y_{i-1} and Y_i and note that

$$\mathbb{E}_\mu[V_i V_{i+j}] = \mathbb{E}_\mu[V_i \mathbb{E}_\mu[V_{i+j} \,|\, Y_{i-1}, Y_i]] = \mathbb{E}_\mu[V_i \mathbb{E}_\mu[V_{i+j} \,|\, Y_i]] = \mathbb{E}_\mu[V_i \mathbb{E}_{Y_i}[V_j]]. \tag{21}$$

If we use (13) as before, then we see that for all $x \in \mathcal{X}$ we have

$$\mathbb{E}_x[V_n] = \mathbb{E}_\nu[V_n] + O(\delta^n),$$

and the implied constant does not depend on x. When we insert this into (21) and recall that the definition of V_n gives us $\mathbb{E}_\nu[V_n] = 0$, the proof of the lemma is complete. $\square$

Lemma 7 helps us deal with cross terms with large j, but we also need a relation that deals with arbitrary j. Here, we again use (13) to get for all $j \geq 0$ that

$$\mathbb{E}_0[V_i V_{i+j}] = \mathbb{E}_\nu[V_i V_{i+j}] + O(\delta^i) = \mathbb{E}_\nu[V_1 V_{1+j}] + O(\delta^i). \tag{22}$$

Now, to calculate B_n, we first apply Lemma 7 to the cross terms $\mathbb{E}_0[V_i V_{i+j}]$ where $i \leq j$ and then apply (22) to the rest to obtain

$$B_n = 2\sum_{j=1}^{n-1}\sum_{i=1}^{j} O(\delta^j) + 2\sum_{j=1}^{\lfloor (n-1)/2 \rfloor}\sum_{i=j+1}^{n-1} \Big\{\mathbb{E}_\nu[V_1 V_{1+j}] + O(\delta^i)\Big\}. \tag{23}$$

We have the sums

$$\sum_{j=1}^{n-1}\sum_{i=1}^{j} O(\delta^j) = \sum_{j=1}^{n-1} O(j\delta^j) = O(1), \qquad \sum_{j=1}^{\lfloor (n-1)/2 \rfloor}\sum_{i=j+1}^{n-1} O(\delta^i) = \sum_{j=2}^{n-1} O(j\delta^j) = O(1)$$

and we have the sum

$$\sum_{j=1}^{\lfloor (n-1)/2 \rfloor} \sum_{i=j+1}^{n-1} \mathbb{E}_\nu[V_1 V_{1+j}] = \sum_{j=1}^{\lfloor (n-1)/2 \rfloor} (n-j-1)\mathbb{E}_\nu[V_1 V_{1+j}],$$

so (23) becomes

$$B_n = 2 \sum_{j=1}^{\lfloor (n-1)/2 \rfloor} (n-j-1)\mathbb{E}_\nu[V_1 V_{1+j}] + O(1). \tag{24}$$

In summary, (17), (18), (19) and (24) give us the key relation

$$\frac{1}{n}\mathrm{Var}[R_n(\rho)] = \mathbb{E}_\nu[V_1^2] + 2 \sum_{j=1}^{\lfloor (n-1)/2 \rfloor} \left(1 - \frac{j+1}{n}\right) \mathbb{E}_\nu[V_1 V_{1+j}] + O(1/n), \tag{25}$$

and by Lemma 7 the summands are absolutely convergent, so we can take the limit in (25) to get

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{Var}[R_n(\rho)] = \mathbb{E}_\nu[V_1^2] + 2\sum_{j=1}^{\infty} \mathbb{E}_\nu[V_1 V_{1+j}] \stackrel{\text{def}}{=} \sigma_\rho^2(F). \tag{26}$$

This completes the proof of the asymptotic relations for the mean and variance of $R_n(\rho)$.

When these relations are coupled with the bound (11) on the Dobrushin coefficient, the rest of the proof of the central limit theorem of Theorem 1 is almost on automatic pilot. The key remaining step is the proof that the constant $\sigma_\rho^2(F)$ defined by (26) is strictly positive. Once this is done, the CLT (5) follows immediately from Theorem 1 of Arlotto and Steele (2016).

To work toward a lower bound for $\sigma_\rho^2(F)$, we let $\mathcal{F}_e$ be the σ-field generated by the evenly indexed terms $Y_0, Y_2, Y_4, \ldots$, and, in order to facilitate calculations that are conditional on the "even σ-field" $\mathcal{F}_e$, we write

$$R_{2n}(\rho) = \sum_{j=0}^{n-1} W_j \quad \text{where } W_j = \mathbb{1}(Y_{2j+1} \neq Y_{2j}) + \mathbb{1}(Y_{2j+2} \neq Y_{2j+1}).$$

We already know by (26) that $\mathrm{Var}[R_n(\rho)] = \mathrm{Var}_0[R_n(\rho)] \sim n\sigma_\rho^2(F)$, and we have also shown that $\mathrm{Var}_0[R_n(\rho)] \sim \mathrm{Var}_\nu[R_n(\rho)]$. Thus, to show $\sigma_\rho^2(F) > 0$, it suffices to show that there is a constant $\alpha > 0$ such that $\mathrm{Var}_\nu[R_{2n}(\rho)] \geq n\alpha$ for all $n \geq 1$. We begin by studying the conditional variances of the individual summands of $R_{2n}(\rho)$.

For specificity, we should also note that for each j the distribution of W_j given $\mathcal{F}_e$ does not depend on the initial distribution of the Markov chain; accordingly we simply write $\mathrm{Var}[W_j|\mathcal{F}_e]$ for the corresponding conditional variance. On the other hand, the distribution of the random variable $\mathrm{Var}[W_j|\mathcal{F}_e]$ depends on the distribution of Y_0, so, for its expectation when $Y_0 \sim \nu$, we need to write $\mathbb{E}_\nu[\mathrm{Var}[W_j|\mathcal{F}_e]]$.

Lemma 8. *For all $\rho \in (0,1)$ and $F \in \mathcal{S}_L$, there exists a constant $\alpha_F(\rho) > 0$ for which one has*

$$\mathbb{E}_\nu[\mathrm{Var}[W_j|\mathcal{F}_e]] = \mathbb{E}_\nu[\mathrm{Var}[W_j|Y_{2j}, Y_{2j+2}]] \geq \alpha_F(\rho) \quad \textit{for all } j \geq 0.$$

Proof. When we condition on $\mathcal{F}_e = \sigma\{Y_0, Y_2, \ldots\}$, the distribution of W_j requires the consideration of two cases. First, if we have $Y_{2j} = Y_{2j+2}$, then with probability one we have $Y_{2j} = Y_{2j+1} = Y_{2j+2}$ and hence $W_j = 0$. Second, given $\mathcal{F}_e$ with $Y_{2j} \neq Y_{2j+2}$, then we have

$$W_j = \begin{cases} 0, & \text{with probability } 0, \\ 1, & \text{with probability } F(\rho Y_{2j}), \\ 2, & \text{with probability } 1 - F(\rho Y_{2j}). \end{cases} \tag{27}$$

From the representation (27) and the strict monotonicity of $F \in \mathcal{S}_L$, we see there is a constant $C_F(\rho) > 0$ such that for all $j \geq 0$,

$$\mathrm{Var}[W_j \,|\, Y_{2j}, Y_{2j+2}] \geq C_F(\rho) \mathbb{1}[Y_{2j} \neq Y_{2j+2},\ \rho L \leq Y_{2j}, Y_{2j+2} \leq L]. \tag{28}$$

If we set $Z = \mathbb{1}[Y_{2j} \neq Y_{2j+2},\ \rho L \leq Y_{2j}, Y_{2j+2} \leq L]$ and

$$A = \{Y_{2j} \in [\rho L, L]\}, B = \{Y_{2j+1} \in [\rho L, L], Y_{2j+1} \neq Y_{2j}\}, C = \{Y_{2j+2} \in [\rho L, L]\}.$$

Then $Z \geq \mathbb{1}(A \cap B \cap C)$ and

$$\mathbb{E}_\nu[Z] \geq P_\nu(A \cap B \cap C) = P_\nu(A) P_\nu(B|A) P_\nu(C|A, B).$$

Each term on the right hand side is at least $1 - F(\rho L)$ because any upcoming observation that falls within $[\rho L, L]$ will be accepted. This gives us

$$\mathbb{E}_\nu[Z] \geq (1 - F(\rho L))^3,$$

so by (28) one can take $\alpha_F(\rho) \equiv C_F(\rho)(1 - F(\rho L))^3 > 0$ to complete the proof of the lemma. □

This is last of the tools we need to get a non-trivial lower bound for $\sigma_\rho^2(F)$. By the law of total variance and by Lemma 8, we have

$$\begin{aligned} \mathrm{Var}[R_{2n}(\rho)] &= \mathbb{E}[\mathrm{Var}[R_{2n}(\rho)|\mathcal{F}_e]] + \mathrm{Var}[\mathbb{E}[R_{2n}(\rho)]|\mathcal{F}_e]] \\ &\geq \mathbb{E}[\mathrm{Var}[R_{2n}(\rho)|\mathcal{F}_e]] = \mathbb{E}\big[\sum_{j=1}^{n} \mathrm{Var}[W_j|\mathcal{F}_e]\big] \geq n\alpha_F(\rho) \end{aligned} \tag{29}$$

where the last equality is due to the independence between W_i and W_j given $\mathcal{F}_e$ when $i \neq j$.

Finally, given Lemma 8 and our earlier observations, the proof of Theorem 1 is complete.

5. Proof of Theorem 2

Before we take up the proof of Theorem 2 in earnest, it will be useful to know that when F is the uniform distribution we can work with the density of the stationary distribution of $K_{\rho,F}$. To get the required absolute continuity we begin with a general inequality.

Proposition 9. *If $F \in \mathcal{S}_L$ and if ν is the stationary measure for the transition kernel $K_{\rho,F}$ given by* (9)*, then for all Borel $A \subset \mathcal{X}$ one has*

$$\nu(A) \leq \frac{1}{1 - F(\rho L)} \int_0^L \mathbb{1}(y \in A) F(dy). \tag{30}$$

Proof. Stationarity of ν and the definition of $K_{\rho,F}$ give us

$$\begin{aligned}\nu(A) &= \int_{\mathcal{X}} K_{\rho,F}(x,A)\nu(dx) \\ &= \int_{\mathcal{X}} \mathbb{1}(x\in A)F(\rho x)\nu(dx) + \int_0^L \int_{\mathcal{X}} \mathbb{1}(y\in A)\mathbb{1}(\rho x\le y\le L)\nu(dx)F(dy) \\ &\le \nu(A)F(\rho L) + \int_0^L \mathbb{1}(y\in A)F(dy),\end{aligned}$$

from which we get (30). □

From (30) we see that ν is always absolutely continuous with respect to F. Consequently, if F is absolutely continuous with respect to Lebesgue measure dx, then both ν and F have densities with respect to dx.

Now we take F to be the uniform distribution on $[0,1]$, and we simply write K_ρ, M_ρ and m_ρ for the corresponding transition kernel, stationary distribution function and density function. The definition of K_ρ and equation of stationarity now tell us

$$m_\rho(y) = \int_0^1 m_\rho(x)K_\rho(x,y)\,dx = \rho y m_\rho(y) + \int_0^1 m_\rho(x)\mathbb{1}(\rho x\le y)\,dx,$$

or, in other words,

$$m_\rho(y) - \rho y m_\rho(y) = M_\rho(y/\rho) \quad \text{for all } y\in[0,1]. \tag{31}$$

Perhaps the quickest way to extract what we need from this key identity is to first introduce the Mellin transform of $m(\cdot)$:

$$\phi(s,\rho) \overset{\text{def}}{=} \int_0^1 x^s m_\rho(x)\,dx.$$

From (31) and the fact that $M_\rho(x)=1$ for $x\ge 1$ we then find

$$\phi(s,\rho) - \rho\phi(s+1,\rho) = \int_0^1 x^s M_\rho(x/\rho)\,dx = \int_0^\rho x^s M_\rho(x/\rho)\,dx + \frac{1-\rho^{s+1}}{s+1}. \tag{32}$$

A change of variables and integration by parts give us

$$\int_0^\rho x^s M_\rho(x/\rho)\,dx = \rho^{s+1}\int_0^1 u^s M_\rho(u)\,du = \rho^{s+1}\frac{1-\phi(s+1,\rho)}{s+1},$$

so (32) becomes

$$\phi(s,\rho) - \rho\phi(s+1,\rho) = \frac{1-\rho^{s+1}\phi(s+1,\rho)}{s+1}, \tag{33}$$

which we can rewrite as a recursion,

$$\phi(s,\rho) = \frac{1}{1+s} + \left(\rho - \frac{\rho^{s+1}}{s+1}\right)\phi(s+1,\rho). \tag{34}$$

Proposition 10 (Mellin Transform of the Stationary Density)**.** *We have*

$$\phi(s,\rho) = \sum_{k=0}^\infty a_k(s) \quad \textit{where} \quad a_0(s) = \frac{1}{1+s} \quad \textit{and} \tag{35}$$

$$a_k(s) = \frac{1}{s+k+1}\prod_{i=1}^k \left(\rho - \frac{\rho^{s+i}}{s+i}\right) \quad \textit{for } k\ge 1.$$

Proof. We just need to check that (35) satisfies the recursion (34). In fact we have

$$\begin{aligned}\left(\rho - \frac{\rho^{s+1}}{s+1}\right) a_k(s+1) &= \frac{1}{s+k+2}\left(\rho - \frac{\rho^{s+1}}{s+1}\right)\prod_{i=1}^{k}\left(\rho - \frac{\rho^{s+1+i}}{s+1+i}\right) \\ &= \frac{1}{s+k+2}\prod_{i=1}^{k+1}\left(\rho - \frac{\rho^{s+i}}{s+i}\right) = a_{k+1}(s),\end{aligned}$$

so summing from $k = 0$ to ∞ gives us

$$\left(\rho - \frac{\rho^{s+1}}{s+1}\right)\phi(s+1,\rho) = \sum_{k=0}^{\infty} a_{k+1}(s) = \sum_{k=1}^{\infty} a_k(s).$$

Since $a_0(s) = 1/(1+s)$, we have proved

$$\left(\rho - \frac{\rho^{s+1}}{s+1}\right)\phi(s+1,\rho) = \sum_{k=1}^{\infty} a_k(s) = \phi(s,\rho) - \frac{1}{1+s},$$

giving us the required recursion (34). □

For the first moment of $m_\rho(\cdot)$ we therefore find

$$\int_0^1 x m_\rho(x)\, dx = \phi(1,\rho) = \frac{1}{2} + \frac{1}{3}\left(\rho - \frac{\rho^2}{2}\right) + \frac{1}{4}\left(\rho - \frac{\rho^2}{2}\right)\left(\rho - \frac{\rho^3}{3}\right) + \cdots,$$

and this is just what we need to complete the calculation of $\mu_\rho(U)$. Specifically, if we specialize the general formula (16) for $\mu_\rho(F)$ to the uniform distribution function U, we get some substantial simplification. Specifically, we have

$$\begin{aligned}\mu_\rho(U) = \mathbb{E}_\nu[\mathbb{1}(Y_0 \neq Y_1)] &= \int_0^1\int_0^1 K_\rho(x,y)\mathbb{1}(x \neq y) m_\rho(x)\, dx\, dy \\ &= \int_0^1\int_0^1 \mathbb{1}(\rho x \leq y) m_\rho(x) dx dy = \int_0^1 (1-\rho x) m_\rho(x)\, dx = 1 - \rho\phi(1,\rho),\end{aligned}$$

and, together with the expansion for $\phi(1,\rho)$, this completes the proof of the first assertion (6) of Theorem 2.

We will see in Section 6 that (31) is from an interesting class of equations with a long history and a rich theory. Nevertheless, there are situations where one can make use of (31) without knowing its solution and without appealing to the wider theory. Specifically, the next proposition illustrates the qualitative use of (31) and gives some properties that will be used in Section 10.

Proposition 11 (Features of the Stationary Density)**.** *The probability density $m_\rho(\cdot)$ on $[0,1]$ that satisfies the equation of stationarity (31) is a continuous, strictly increasing function on $(0,1)$. Moreover, it is strictly convex on $(0,\rho)$ and strictly convex on $(\rho,1]$, but it is not convex in any neighborhood of ρ. In particular, $m'_\rho(\cdot)$ has a jump discontinuity at ρ.*

Proof. If we write (31) as $m_\rho(y) = (1-\rho y)^{-1} M_\rho(y/\rho)$, then from the fact that $y \mapsto M_\rho(y/\rho)$ is non-decreasing and $(1-\rho y)^{-1}$ is strictly increasing, we see that $m_\rho(\cdot)$ is strictly increasing. Also, from the continuity of $M_\rho(\cdot)$ we see that $m_\rho(\cdot)$ is continuous, and, since $M_\rho(\cdot)$ is the integral of $m_\rho(\cdot)$, we see that $M_\rho(\cdot)$ is continuously differentiable. From $M_\rho(0) = 0$ and (31) we have $m_\rho(0) = 0$ from which we also find $M'_\rho(0) = m_\rho(0) = 0$.

For all $x \in [1, \infty)$ we have $M_\rho(x) = 1$, so (31) gives us an explicit formula

$$m_\rho(y) = \frac{1}{1-\rho y} \quad \text{for } y \in [\rho, 1]. \tag{36}$$

On the other hand, for $y \in [0, \rho)$ the relation (31) is an integral equation which by differentiation gives us

$$m'_\rho(y) = \frac{m_\rho(y/\rho)}{\rho(1-\rho y)} + \frac{\rho M_\rho(y/\rho)}{(1-\rho y)^2} \quad \text{for } y \in [0, \rho). \tag{37}$$

From (37) and the fact that $m_\rho(\cdot)$ and $M_\rho(\cdot)$ are both monotone non-decreasing, we find that $m_\rho(\cdot)$ is strictly convex on $(0, \rho)$, while from (36) we see immediately that $m_\rho(\cdot)$ is strictly convex on $[\rho, 1]$.

To check that $m_\rho(\cdot)$ fails to be convex on $[0, 1]$, we first recall that we have $m_\rho(1) = 1/(1-\rho)$ and $M_\rho(1) = 1$, so by (31) and (37) the one-sided derivatives at ρ are given by

$$m'_\rho(\rho^+) = \frac{\rho}{(1-\rho^2)^2} < \frac{1}{\rho(1-\rho)(1-\rho^2)} + \frac{\rho}{(1-\rho^2)^2} = m'_\rho(\rho^-).$$

This tells us that $m_\rho(\cdot)$ is not differentiable at ρ, and, moreover, since $m'_\rho(\cdot)$ has a negative jump discontinuity at ρ, we see that $m_\rho(\cdot)$ is not convex on $[0, 1]$. $\square$

As an application of the proposition, one should note that the jump discontinuity of $m'_\rho(\cdot)$ tells us that some plausible solution methods for (37) cannot work. For example, there can be no solution of (37) on $[0, 1]$ that is given by a power series (or a Frobenius series), even though we have the nice power series representation (35) for the Mellin transform of $m_\rho(\cdot)$.

6. The Stationary Measure and the Pantograph Equation

The first-order non-autonomous *pantograph equation* for $\lambda \in (0, \infty)$ is the functional differential equation

$$H'(t) = a(t)H(t) + b(t)H(\lambda t) \quad t \geq 0. \tag{38}$$

The connection to the problems considered here is that for $0 < \rho < 1$ the equation (31) for the distribution function M_ρ of the stationary measure of the transition kernel $K_{\rho,U}(\cdot, \cdot) \equiv K_\rho(\cdot, \cdot)$ can be written as

$$M'_\rho(t) = \frac{1}{1-\rho t} M_\rho(t/\rho) \quad \text{for } 0 \leq t < 1. \tag{39}$$

Thus, on the interval $[0, 1]$, the distribution function M_ρ satisfies the pantograph equation (38) with $a(t) = 0$, $b(t) = 1/(1-\rho t)$, and $\lambda = 1/\rho > 1$.

The pantograph equation occurs in many contexts, perhaps the earliest of which was a number of theoretic investigations of Mahler (1940) that exploited the equation $H'(t) = bH(\lambda t)$, $H(0) = 1$ and its solution

$$H(t) = \sum_{j=0}^{\infty} \frac{1}{j!} \lambda^{j(j-1)/2} b^j t^j, \tag{40}$$

which is an elegant — and useful — generalization of the exponential function.

The two-term pantograph equation (38) has mostly commonly occurred in the autonomous case where $a(t)$ and $b(t)$ are constant, and the equation got its name from Fox et al. (1971) where the autonomous equation was used to model the collection of current by the pantograph (or flat pan connection head) of a tram. The subsequent investigation of Kato and McLeod (1971) showed the full richness of the equation, and, ever since, the pantograph equation has been regularly studied and applied, see e.g. Iserles (1993), Derfel and Iserles (1997), Guglielmi and Zennaro (2003), Saadatmandi and Dehghan (2009), Yusufoğlu (2010), and Hsiao (2015), all of which contain many references.

In the non-autonomous case, essentially all work on (38) has been asymptotic or numerical. Moreover, all of the recent work focuses on the case when $\lambda \in (0, 1)$, and there is a sound scientific reason for this. Specifically, for (38) to be useful in an engineering or scientific context, it seems natural to assume that it is a *causal* equation; that is, the current rate of change $H'(t)$ is required to be determined by information that is available at time t.

A noteworthy feature of the stationarity equation (39) is that it is *not* a causal equation; one has $\lambda = 1/\rho > 1$. The other interesting feature of (39) is that it was essentially solved in Section 5, at least in the sense that Proposition 10 gives explicit series expansion of its Mellin transform.

Mellin transforms have rarely been used in the theory of the pantograph equation; we know of only one other case. Specifically, van Brunt and Wake (2011) used Mellin transforms to study a second order non-autonomous pantograph equation. Intriguingly, their equation was also acausal, and it also had a probabilistic origin. Specifically, it arose as the Fokker-Plank equation in a diffusion model for a population of cells, and the acausal parameter came from a splitting constant for cell division.

We do not make further use the pantograph equation here, but, given the richness of its theory, the connection seems worth noting. Benefits may even flow both ways. For example, calculations like those of Section 5 provide explicit Mellin transforms for the solutions of some other pantograph equations in addition to (39).

7. When $\rho > 1$: The Proof of Theorem 3

We now consider an infinite sequence $X_1, X_2, \ldots$ of independent random variables with distribution $F \in \mathcal{S}_L$. We then fix $\rho > 1$, and we again use the recursive definition (1) to specify the set of selection times $\{\tau_k : k = 1, 2, \ldots\}$. If we then set

$$N_\rho = \min\{k : X_{\tau_k} \in (L/\rho, L]\} \text{ and } M_\rho = \min\{\tau_k : X_{\tau_k} \in (L/\rho, L]\}$$

then the number of selections one makes from $\{X_1, X_2, \ldots, X_n\}$ is simply given by $R_n(\rho) = R_{\min(n, M_\rho)}(\rho)$, since, after a selection larger than L/ρ has been made, no further selections are possible. Also, for each $\omega \in \{\omega : M_\rho(\omega) < \infty\}$, we have

$$R_n(\rho) = R_{\min(n, M_\rho)}(\rho) \nearrow R_{M_\rho}(\rho) = N_\rho \quad \text{as } n \to \infty,$$

so the main task is to prove the moment generating function bound (7).

Since each value accepted by the selection process with $\rho > 1$ must be at least a factor of ρ greater than the preceding selection we have the bounds

$$N_\rho \le \max\{k : \rho^{k-1} X_1 \le L\} \le 1 + \log L / \log \rho - \log X_1 / \log \rho,$$

so for the moment generating function we find

$$\mathbb{E}[\exp(s N_\rho)] \le \exp(s) L^{s/\log \rho} \mathbb{E}[X_1^{-s/\log \rho}] = \exp(s) L^{s/\log \rho} \int_0^L x^{-s/\log \rho} dF.$$

We know the integral is finite when $F(x) = O(x)$ near 0 and $|s| < \log \rho$, and this gives us (7).

To show N_ρ is unbounded, we first fix an integer $M > 1$, and we consider the disjoint subintervals $\{I_1, I_2, \dots, I_M\}$ of $[0, L]$ that are defined by setting

$$I_k = [a_k, b_k] = \left[\frac{(\rho-1)L}{\rho^M - 1} \sum_{i=1}^{k-1} \rho^i, \frac{(\rho-1)L}{\rho^M - 1} \sum_{i=0}^{k-1} \rho^i \right], \quad 1 \le k \le M.$$

The main feature here is that one has $a_{k+1}/b_k = \rho > 1$ for all $1 \le k < M$. If we have $X_i \in I_i$ for $i = 1, 2, \dots M$, then all of the observations X_1, X_2, ..., X_M are selected, so we always have the inequality

$$\prod_{k=1}^{M} \mathbb{1}(X_k \in I_k) \le \mathbb{1}(N_\rho \ge M).$$

Finally, by the independence of the variables X_k, $1 \le k \le M$ and the strict monotonicity of F, we see that the expectation of the product is strictly positive. This gives us $P(N_\rho \ge M) > 0$ for all $M \ge 1$. Since M is arbitrary, we see that N_ρ is unbounded, and the proof of the theorem is complete.

8. Complements to Classical Record Theory

Here we consider the calculation of the expected number of selections where we assume that there was a selection made at "time zero" that had value $x \in [0, 1]$. Formally, we modify the definition (2) by first setting $\tau_1 = \min\{j : X_j \ge \rho x\}$. Next, for $k \ge 2$ we define τ_k as before by setting $\tau_k = \min\{j : X_j \ge \rho X_{\tau_{k-1}}\}$, and finally we set

$$R_n^x(\rho) = \max\{k : \tau_k \le n\}. \tag{41}$$

In this notation, Rényi's classical formula for the expected number of records is

$$\mathbb{E}[R_n^0(1)] = \sum_{k=1}^{n} \frac{1}{k} \stackrel{\text{def}}{=} H_n, \tag{42}$$

and the main goal of this section is to generalize this result in two ways. The immediate goal is to show that

$$\mathbb{E}[R_n^x(1)] = H_n - \sum_{k=1}^{n} \frac{x^k}{k}, \tag{43}$$

and then in Theorem 12 we will get a closely related formula for $\mathbb{E}[R_n^x(\rho)]$.

We begin by using first step analysis to get a useful recursion for the quantities

$$g_{n,\rho}(x) \stackrel{\text{def}}{=} \mathbb{E}[R_n^x(\rho)] \quad \text{and} \quad g_n(x) \stackrel{\text{def}}{=} g_{n,1}(x).$$

Specifically, if we consider the first observation $y = X_1$, then X_1 is not accepted if $y \le \rho x$, and this happens with probability ρx. On the other hand if $y = X_1 \in [\rho x, 1]$ we do accept X_1, and accordingly we find the basic recurrence relation

$$g_{n+1,\rho}(x) = \rho x g_{n,\rho}(x) + \int_{\rho x}^{1} [1 + g_{n,\rho}(y)]\, dy. \tag{44}$$

For general $\rho \in (0, 1)$, this equation offers considerable resistance; in essence, it is a linearized non-autonomous pantograph equation in integrated form. Nevertheless, one can use (44) to extract some interesting information, including refinements of some classical facts.

For example, if we take $\rho = 1$ in (44), then we can make some quick progress. Specifically, if we write $g_n(x)$ for $g_{n,1}(x)$ then differentiation and a nice cancellation give us

$$g'_{n+1}(x) = xg'_n(x) - 1. \tag{45}$$

We have $g_1(x) = 1 - x$, so $g'_1(x) = -1$ and repeated applications of (45) give us

$$g'_2(x) = -x - 1,\ g'_3(x) = -x^2 - x - 1, \text{ and } g'_4(x) = -x^3 - x^2 - x - 1.$$

In general, one has

$$g'_n(x) = -x^{n-1} - x^{n-2} - \cdots - 1 = -\frac{1 - x^n}{1 - x}, \tag{46}$$

so integration over $[0, x]$ gives us

$$g_n(x) = g_n(0) - x - \frac{x^2}{2} - \cdots - \frac{x^n}{n}. \tag{47}$$

Now if we use the basic recursion (44) with $x = 0$ and $\rho = 1$ we have from (47) that

$$\begin{aligned} g_{n+1}(0) &= 1 + \int_0^1 g_n(y)\, dy = g_n(0) + 1 - \frac{1}{1 \cdot 2} - \frac{1}{2 \cdot 3} - \cdots - \frac{1}{n \cdot (n+1)} \\ &= g_n(0) + \frac{1}{n+1}. \end{aligned}$$

By telescoping we then recover Rényi's formula $g_n(0) = H_n$, but from (47) we now also find our refinement of Rényi's formula (and its approximation):

$$\mathbb{E}[R_n^x(1)] = H_n - \sum_{k=1}^n \frac{x^k}{k} = \log n - \sum_{k=1}^n \frac{x^k}{k} + \gamma + \frac{1}{2n} + O(1/n^2), \tag{48}$$

where $\gamma = 0.577\cdots$ is Euler's constant.

For any $0 < \rho < 1$, one can derive a representation of $\mathbb{E}[R_n^x(\rho)]$ that is only a little less explicit than (43). The correcting term is again a truncated power series, but, in the general case, the principal term $g_{n,\rho}(0)$ is no longer a well-known quantity.

Theorem 12. *For all $0 < \rho \le 1$ and $0 \le x \le 1$ we have*

$$g_{n,\rho}(x) = g_{n,\rho}(0) - \sum_{i=1}^n a_i x^i, \tag{49}$$

where $a_1 = \rho$, $a_2 = \rho(\rho - \rho^2/2)$, and $a_i = (\rho - \rho^i/i)a_{i-1}$ for all $i \ge 2$.

Proof. To argue by induction, we first recall that $g_{1,\rho}(x) = \rho x$ for all $0 \le x \le 1$, and this gives us by direct evaluation that (49) holds for $n = 1$. Next, from the basic recursion (44) we have

$$g_{n+1,\rho}(0) = \int_0^1 [1 + g_{n,\rho}(y)]dy \quad \text{and} \quad g_{n+1,\rho}(x) = \rho x g_{n,\rho}(x) + \int_{\rho x}^1 [1 + g_{n,\rho}(y)]dy,$$

so taking the difference gives us

$$g_{k+1,\rho}(0) - g_{k+1,\rho}(x) = \rho x + \int_0^{\rho x} [g_{n,\rho}(y) - g_{n,\rho}(x)]dy. \tag{50}$$

By the induction hypothesis we can expand the last integrand as

$$g_{n,\rho}(y) - g_{n,\rho}(x) = \sum_{i=1}^{n} a_i(x^i - y^i), \tag{51}$$

so from (50) and the defining relation $a_{i+1} = (\rho - \rho^{i+1}/(i+1))a_i$ we have

$$\int_0^{\rho x} \sum_{i=1}^{n} a_i(x^i - y^i)dy = \sum_{i=1}^{n} a_i \left(\rho - \frac{\rho^{i+1}}{i+1}\right) x^{i+1} = \sum_{i=1}^{n} a_{i+1}x^{i+1}. \tag{52}$$

Finally, from (50) and (52) we then get

$$g_{n+1,\rho}(0) - g_{n+1,\rho}(x) = \sum_{i=1}^{n+1} a_i x^i,$$

which completes the induction step. □

Since $0 < a_i \le \rho^i$, the identity (49) has an immediate corollary that underscores an informative difference between the case when $\rho \in (0,1)$ and the case when $\rho = 1$. Specifically, for $\rho = 1$ we see from (43) that the influence of x is unbounded, while the next corollary tells us that for $0 < \rho < 1$ the influence of the initial value x has only a *bounded influence.*

Corollary 13 (Insensitivity of the Initial Constraint)**.** *For all* $\rho \in (0,1)$*,* $n \ge 0$*, and all* $0 \le x \le y \le 1$*, one has*

$$0 \le g_{n,\rho}(x) - g_{n,\rho}(y) \le \frac{\rho}{1-\rho}. \tag{53}$$

The bounds (53) suggest that we should take limits, and from the geometric convergence in (51), we can define a continuous anti-symmetric function $B : [0,1]^2 \to \mathbb{R}$ by setting

$$\sum_{i=1}^{\infty} a_i(y^i - x^i) = \lim_{n\to\infty} \{g_{n,\rho}(x) - g_{n,\rho}(y)\} \stackrel{\text{def}}{=} B(x,y). \tag{54}$$

A useful feature of this function is that it leads to an alternative characterization of $\mu_\rho(U)$, and it gives second proof of the series representation (6).

To derive the characterization, we subtract $g_{n,\rho}(x)$ from both sides of the basic recursion (44), and we simplify to get the identity

$$g_{n+1,\rho}(x) - g_{n,\rho}(x) = (1-\rho x) + \int_{\rho x}^{1} \{g_{n,\rho}(y) - g_{n,\rho}(x)\}\, dy.$$

Now, if we set $x = 0$ in the defining relation (54) and apply antisymmetry of $B(\cdot,\cdot)$, then we see that as $n \to \infty$ one has

$$g_{n+1,\rho}(0) - g_{n,\rho}(0) = 1 + \int_0^1 B_\rho(y,0)\, dy + o(1) = 1 - \int_0^1 B_\rho(0,y)\, dy + o(1).$$

We now sum over $n \in [0:N]$. By telescoping, division by $N+1$, and taking limits we get a new formula for the mean $\mu_\rho(U)$ given by Theorem 2:

$$\mu_\rho(U) = 1 - \int_0^1 B_\rho(0,y)\, dy. \tag{55}$$

Finally, if we substitute the series expansion (54) for $B_\rho(0,y)$ into (55), we see that term-by-term integration of (55) gives us a second derivation of the original formula (6) for $\mu_\rho(U)$. In a sense, this integration also explains the presence of the harmonic factors 1/2, 1/3, 1/4, ... in (6).

9. More Records: Relaxed or Constrained

For any $\rho \in (0, \infty)$ and any $F \in \mathcal{S}_L$, we can consider the set of selected values $\mathcal{A}(\rho) = \{X_{\tau_1}, X_{\tau_2}, \ldots\}$; these are formally defined by the stopping time recursion (1). The set $\mathcal{A}(1)$ is exactly the set of record values, and more generally we have the relations

$$\rho \in (0,1) \Rightarrow \mathcal{A}(1) \subset \mathcal{A}(\rho) \quad \text{and} \quad \rho \in (1,\infty) \Rightarrow \mathcal{A}(\rho) \subset \mathcal{A}(1), \tag{56}$$

which give us a more explicit sense in which $\rho \in (0,1)$ *relaxes* the record condition and $\rho \in (1, \infty)$ further *constrains* the record condition.

The first relation of (56) is obvious since whenever X_k is a record, then we have $X_k \geq X_{\tau_i} \geq \rho X_{\tau_i}$ for all $\tau_i < k$. It is rather less obvious that for $\rho > 1$ one has the complementary relation $\mathcal{A}(\rho) \subset \mathcal{A}(1)$. To prove this by induction, we first note that $X_{\tau_1} = X_1 \in \mathcal{A}(\rho)$, and by definition X_1 is a record. Now we suppose by induction that the first n elements $\{X_{\tau_1}, X_{\tau_2}, \ldots, X_{\tau_n}\}$ of $\mathcal{A}(\rho)$ are also all records.

There are two cases to consider. First, if $\tau_{n+1} < \infty$ and $X_{\tau_n} = x$, then we have $\tau_{n+1} = \min\{k : X_k \geq \rho x\}$. This tells us that τ_{n+1} is the first entrance time of the process $X_1, X_2, \ldots$ into the interval $[\rho x, L]$. Since all such first entrance times are also record times, we see that $X_{\tau_{n+1}}$ is a record, and induction gives us $\mathcal{A}(\rho) \subset \mathcal{A}(1)$. In the second case we have $\tau_{n+1} = \infty$, and $\mathcal{A}(\rho) = \{X_{\tau_1}, X_{\tau_2}, \ldots, X_{\tau_n}\}$. We already have from our induction hypothesis that $\{X_{\tau_1}, X_{\tau_2}, \ldots, X_{\tau_n}\} \subset \mathcal{A}(1)$, so again we get $\mathcal{A}(\rho) \subset \mathcal{A}(1)$.

Despite the first relation of (56), it is generally inappropriate to think of the values $\mathcal{A}(\rho) = \{X_{\tau_j} : j = 1, 2, \ldots\}$ are anything like "approximate records" when $0 < \rho < 1$. To make this distinction explicit, fix $0 < \epsilon < 1$ and consider the events

$$A_k = \big\{X_k \text{ is selected and } X_k \leq \epsilon \max\{X_i : i \leq k\}\big\}, \tag{57}$$

where the random variables X_i, $i = 1, 2\ldots$ are independent and uniformly distributed on $[0, 1]$.

When A_k occurs, the selected value X_k is only a small fraction of the current maximum, so it is not an approximate record (or a near-record) in any reasonable sense. Nevertheless, with probability one, infinitely many of the events $A_1, A_2, \ldots$ will occur, so infinitely often the selected values are quite unlike records.

To see this, we first note by Proposition 11 that for any $\epsilon > 0$ both of the sets $[0, \epsilon/2]$ and $[1/2, 1]$ have positive probability under the stationary measure ν for the associated Markov chain $\{Y_n : n = 1, 2, \ldots\}$ of Section 2. Thus, they are also both recurrent sets for the chain. Now, if at time k the chain enters $[0, \epsilon/2]$ after having entered $[1/2, 1]$ at some time previous to k, then the event A_k also occurs. The positive recurrence of the respective sets then tells us that infinitely many of the events $\{A_k : k = 1, 2, \ldots\}$ will occur with probability one.

This construction shows that there is a disconnection between the theory of the selection process with $0 < \rho < 1$ and the theory of the near records such as studied in Balakrishnan et al. (2005), Gouet et al. (2007) or Gouet et al. (2012), but this construction does not tell the whole story. In Section 8 we saw several instances where the technology of selection processes could inform us about the classical record process. Still, it is reasonable to expect that one has at least some analogous carry forward to the theory of near-records, but here we cannot pursue that point except to acknowledge the possibility.

10. Connections and Directions

The theory of relative sequential selections lets one embed Rényi's record process into a parametric family of processes with a parameter $\rho \in (0, \infty)$ where the sequential selections are made easier when $\rho \in (0, 1)$ and where they are made harder when $\rho \in (1, \infty)$. If the horizon is finite, then the parametric processes have immediate contiguity relations; specifically, the finite dimensional distributions depend continuously on the parameter ρ. Still, as the table below reminds us, the processes show singular differences when ρ is fixed and n tends to infinity.

	Expectation	Variance	CLT
$\rho < 1$	$\sim n\mu_\rho(F)$	$\sim n\sigma_\rho^2(F)$	Yes
$\rho = 1$	$\sim \log n$	$\sim \log n$	Yes
$\rho > 1$	Convergence to random variable a.s.		

There are several natural directions that have not been explored here. First, there are the issues of rates of convergence. In Rényi's case this is relatively easy, since one has access to the full theory of sums of independent random variables. On the other hand, it would be quite difficult to obtain a rate result when $\rho \in (0, 1)$ since even the basic CLT depends on the theory of functions of non-homogenous Markov chains.

Second, there is the possibility of related Poisson laws. Thus, for example, in Rényi's framework one can easily show that the number of records between time n and time $2n$ is approximately Poisson with mean $\lambda = \ln 2$. On the other hand, when $\rho \in (0, 1)$ the linear growth rate of the mean and variance tells us that as far as this example goes there is no directly analogous Poisson law. Naturally, there can be — and probably are — Poisson laws that are more distantly related.

Finally, one can consider the theory of selection processes where one does not require F to have compact support. Such extensions are feasible and interesting. Nevertheless, they are also intrinsically more complicated. For example, to guarantee good asymptotic behavior of the mean and variance of the number of selections when $\rho \in (0, 1)$, one probably needs to assume that the tail map $x \mapsto 1 - F(x)$ has regular variation. Still, even with such an assumption, proper analogs of Theorems 1 and 3 are not easily formulated — or proved.

References

Arlotto, A. and Steele, J. M. (2016), 'A central limit theorem for temporally nonhomogenous markov chains with applications to dynamic programming', to appear in *Mathematics of Operations Research (eprint arXiv:1505.00749)* .

Balakrishnan, N., Pakes, A. G. and Stepanov, A. (2005), 'On the number and sum of near-record observations', *Adv. in Appl. Probab.* **37**(3), 765–780.

Derfel, G. and Iserles, A. (1997), 'The pantograph equation in the complex plane', *J. Math. Anal. Appl.* **213**(1), 117–132.

Fox, L., Mayers, D. F., Ockendon, J. R. and Tayler, A. B. (1971), 'On a functional differential equation', *J. Inst. Math. Appl.* **8**, 271–307.

Gouet, R., López, F. J. and Sanz, G. (2007), 'Asymptotic normality for the counting process of weak records and δ-records in discrete models', *Bernoulli* **13**(3), 754–781.

Gouet, R., López, F. J. and Sanz, G. (2012), 'Central limit theorem for the number of near-records', *Comm. Statist. Theory Methods* **41**(2), 309–324.

Guglielmi, N. and Zennaro, M. (2003), 'Stability of one-leg Θ-methods for the variable coefficient pantograph equation on the quasi-geometric mesh', *IMA J. Numer. Anal.* **23**(3), 421–438.

Hsiao, C.-H. (2015), 'A Haar wavelets method of solving differential equations characterizing the dynamics of a current collection system for an electric locomotive', *Appl. Math. Comput.* **265**, 928–935.

Iserles, A. (1993), 'On the generalized pantograph functional-differential equation', *European J. Appl. Math.* **4**(1), 1–38.

Kato, T. and McLeod, J. B. (1971), 'The functional-differential equation $y'(x) = ay(\lambda x) + by(x)$', *Bull. Amer. Math. Soc.* **77**, 891–937.

Mahler, K. (1940), 'On a special functional equation', *J. London Math. Soc.* **15**, 115–123.

Nagaev, S. V. (2015), 'The spectral method and ergodic theorems for general Markov chains', *Izv. Ross. Akad. Nauk Ser. Mat.* **79**(2), 101–136.

Rényi, A. (1962), 'Théorie des éléments saillants d'une suite d'observations', *Ann. Fac. Sci. Univ. Clermont-Ferrand No.* **8**, 7–13.

Saadatmandi, A. and Dehghan, M. (2009), 'Variational iteration method for solving a generalized pantograph equation', *Comput. Math. Appl.* **58**(11-12), 2190–2196.

van Brunt, B. and Wake, G. C. (2011), 'A Mellin transform solution to a second-order pantograph equation with linear dispersion arising in a cell growth model', *European J. Appl. Math.* **22**(2), 151–168.

Yusufoğlu, E. (2010), 'An efficient algorithm for solving generalized pantograph equations with linear functional argument', *Appl. Math. Comput.* **217**(7), 3591–3595.