B. Online Appendix: Proof of Theorem 1 for a Single Lottery

B.1. **Markov Chain Description.**

We begin with a Markov Chain description of the Deferred Acceptance algorithm with a single lottery. Our description uses the implementation of Deferred Acceptance described by McVitie and Wilson (1971). This implementation starts by placing students in a linear order, and proceeds in ρn "rounds." At the beginning of round t, student s_t proposes to her first choice school. If this school has a vacancy, then the student is accepted and the round ends. Otherwise, the school rejects one student, who goes on to propose to the next school on her list. This process continues until either a vacancy is filled or a student is rejected from the last school on her list. At that point, round t ends with an assignment that is stable in the submarket consisting of only students $\{s_1, s_2, \ldots, s_t\}$.

We follow in the footsteps of Pittel (1992) and Immorlica and Mahdian (2005) by deploying what they call the principle of deferred decisions. This means that we generate student preferences on demand, waiting to reveal the next school on a student's list until that student is about to propose. In the case where priorities are determined by a single lottery, we simplify our analysis by placing students in decreasing order of priority. This implies that in each round t, only student s_t can be rejected: she will never trigger the rejection of students processed before her.

We now give a formal description of our Markov Chain for Deferred Acceptance with a single lottery. We let $\mathcal{S}$ be the set of students, and $\mathcal{H}$ be the set of high schools. The state of our Markov chain is $X = (L, \mathcal{I}, \mathcal{M})$, where

- $L \in (\mathbb{N} \cup \{\emptyset\})^{\mathcal{S}}$ tracks the length of student lists:
  $L_s \in \mathbb{N}$ is the number of schools listed by s, and $L_s = \emptyset$ if this has not been determined.
- $\mathcal{I} \subseteq \{(s,h) : s \in \mathcal{S}, h \in \mathcal{H}\}$ tracks the schools that each student has proposed to:
  $(s,h) \in \mathcal{I}$ if s has proposed to h. We define
  $$\mathcal{I}_s = \{h : (s,h) \in \mathcal{I}\}, \qquad \mathcal{I}_h = \{s : (s,h) \in \mathcal{I}\}.$$
- $\mathcal{M} \subseteq \mathcal{I}$ tracks the set of tentative assignments:
  $(s,h) \in \mathcal{M}$ if s is tentatively matched to h. We define
  $$\mathcal{M}_s = \{h : (s,h) \in \mathcal{M}\}, \qquad \mathcal{M}_h = \{s : (s,h) \in \mathcal{M}\}.$$

The evolution of this chain is as follows.

Algorithm 1 (DA with a Single Lottery)**.**
$\mathcal{I} = \mathcal{M} = \emptyset$, $L_s = \emptyset$ for all s.
for $t \in \{1, 2, \ldots, |\mathcal{S}|\}$ ***do***
  $s = s_t$, $L_s \sim \mathcal{L}$.
  while $|\mathcal{I}_s| \leq L_s$ && $\mathcal{M}_s = \emptyset$ ***do***
    Sample h uniformly from $\mathcal{H} \backslash \mathcal{I}_s$ and set $\mathcal{I} \leftarrow \mathcal{I} \cup (s,h)$.
    if $|\mathcal{I}_h| \leq C_h$ ***then*** *s fills a vacancy: $\mathcal{M} \leftarrow \mathcal{M} \cup (s,h)$.*
  end while
end for

For the case where schools differ in their popularity, only a single modification to the above algorithm is needed. Let p_h be the popularity of h. For any non-empty subset $H \subseteq \mathcal{H}$, define the probability distribution $P(H)$ over $\mathcal{H}$ by

$$P_h(H) = \frac{p_h \mathbf{1}(h \in H)}{\sum_{h' \in H} p_{h'}}.$$

Then the only change required is that each time that a new proposal is made, the identity of school h is sampled from $P(\mathcal{H} \backslash \mathcal{I}_s)$, rather than the uniform distribution on $\mathcal{H} \backslash \mathcal{I}_s$.

B.2. **The Differential Equation Method.** The state space for the Markov chain above is very large, as it includes the application history for each student. To address this, we use the "differential equation" method developed by Wormald (1999).

At a high level, the idea underlying this technique is that a lower-dimensional state space "nearly suffices" to describe the evolution of the Markov chain. More specifically, we will show that there exists a reduced state $Y(t) = Y(X(t)) \in \mathbb{R}^d$ (where the dimension d does not depend on n) and a function $f : \mathbb{R}^d \to \mathbb{R}^d$ such that for any state $X(t)$,

$$\mathbb{E}[Y(t+1) - Y(t) | X(t)] \approx f(Y(t)/n),$$

with error that vanishes as n grows.

We use the following special case of Theorem 5.1 from Wormald (1999).

Theorem 4. *For $n \in \mathbb{N}$, let $X^n = \{X^n(t)\}_{t \geq 1}$ be a discrete time Markov chain in the space $\mathcal{X}^n$. Suppose that there exist $d \in \mathbb{N}$, continuous functions $Y^n : \mathcal{X}^n \to \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}^d$, and a bounded set $\mathcal{Y} \subset \mathbb{R}^d$ such that:*

(i) (Initial Condition) There exists $y^0 \in \mathcal{Y}$ such that $Y^n(X^n(0))/n = y^0$ for all n.

(ii) (Boundedness) There exists $\beta > 0$ such that

$$\max_{n,t \in \mathbb{N}} ||Y^n(X^n(t+1)) - Y^n(X^n(t))||_\infty \leq \beta. \tag{60}$$

(iii) (Lipschitz hypothesis) There exists $L > 0$ such that for all $y, \tilde{y} \in \mathcal{Y}$,

$$||f(y) - f(\tilde{y})||_\infty \leq L \, ||y - \tilde{y}||_\infty \, . \tag{61}$$

(iv) (Trend hypothesis) There exists $\delta > 0$ such that if $Y^n(X^n(t))/n \in \mathcal{Y}$, then

$$||\mathbb{E}[Y^n(X^n(t+1)) - Y^n(X^n(t)) | X^n(t)] - f(Y^n(X^n(t))/n)||_\infty \leq \delta/n. \tag{62}$$

Then the following are true:

(a) (Unique solution) There is a unique function $\hat{y} : \mathbb{R}_+ \to \mathbb{R}^d$ satisfying

$$\hat{y}(0) = y_0, \qquad \hat{y}_i'(s) = f_i(\hat{y}(s)). \tag{63}$$

(b) (Convergence) Define $\sigma = \inf\{s \geq 0 : \hat{y}(s) \notin \mathcal{Y}\}$. For any $s < \sigma$ and $\epsilon > 0$, as $n \to \infty$,

$$\mathbb{P}(||Y^n(X^n(\lfloor ns \rfloor))/n - \hat{y}(s)||_\infty > \epsilon) \to 0. \tag{64}$$

The proof of this result is given by Wormald (1999). We briefly outline how each of the conditions above is used in the proof. In what follows, we drop the superscript n to reduce clutter, and use the shorthand $Y(t)$ in place of $Y^n(X^n(t))$.

The most fundamental condition is (iv), which states that the expected change in Y is well-approximated by the function f. Due to random fluctuation, the realized value of $Y(t+1)-Y(t)$ could be far from its expectation. However, over a large number of steps, this randomness should cancel out; we might hope that

$$Y(t+w)-Y(t) \approx \sum_{k=0}^{w-1} f(Y(t+k)/n) \approx wf(Y(t)/n),$$

provided that two conditions hold:

1. $Y(t+1)-Y(t)$ is unlikely to be large. This implies that
   - $Y(t+w)-Y(t) = \sum_{k=0}^{w-1} Y(t+k+1)-Y(t+k)$ is not dominated by any large jumps
   - $Y(t+k)/n \approx Y(t)/n$, so long as k is small relative to n.
2. f is sufficiently smooth, so that $f(Y(t+k)/n) \approx f(Y(t)/n)$.

These are exactly the conditions (ii) and (iii). The proof of Theorem 4 follows this reasoning, with $w = n^\alpha$ for some $\alpha \in (0,1)$, so that the conditions above hold. These allow for the application of a well-known Chernoff-style large deviation bounds, which establish that the sum of these many small terms is unlikely to be far from its expected value. In other words, $Y(t)/n \approx \hat{y}(t/n)$, where $\hat{y}(\cdot)$ is the solution to the differential equation (63).

The notation in Theorem 4 is somewhat different from that of Theorem 5.1 in Wormald (1999). This is partly because our setting does not require the full power of that theorem: our initial state y_0 is identical for all n, our bounded condition (ii) holds deterministically (the γ that appears in his paper is equal to zero), and our β does not depend on n. Furthermore, we strengthen his condition (iv) by explicitly assuming that his λ_1 term is $\mathcal{O}(1/n)$. Finally, whereas his conclusion (b) is made messy by specifying a rate of convergence, our weaker but simpler-to-parse conclusion (b) suffices for our purposes.

B.3. Defining Summary Statistics and Verifying Bounded, Lipschitz, Trend Conditions.

Throughout, we drop the dependence of Y and X on n. We begin by defining our summary statistics Y. Define

$$\mathcal{J} = \{j \in \mathbb{N} : 0 \leq j < C\} \tag{65}$$

$$\mathcal{Z} = \{(k,l) \in \mathbb{N}^2 : 1 \leq k \leq l \leq \ell\}, \tag{66}$$

and recall that ℓ represents the maximum length of student lists.

Given state $X = \{L, \mathcal{I}, \mathcal{M}\}$, for $j \in \mathcal{J}$ and $(k,l) \in \mathcal{Z}$ define

$$Y_j(X) = |\{h : |\mathcal{I}_h| = j\}| \tag{67}$$

$$Y_{kl}(X) = |\{s : L_s = l, |\mathcal{I}_s| = k, \mathcal{M}_s \neq \emptyset\}|\,. \tag{68}$$

In English, Y_j is the number of schools who have received exactly j proposals, while Y_{kl} is the number of students who list l schools and are matched to their k^{th} choice. The variables Y_{kl} are the measures of student welfare addressed in Theorem 1. Meanwhile, the variables Y_j are useful for tracking the number of schools with availability, which determines how likely it is that future students will be able to match to their top choices.

Because there are only n schools, (67) implies

$$\frac{1}{n}\sum_{j\in\mathcal{J}} Y_j(t) \le \frac{1}{n}|\mathcal{H}| = 1.$$

Meanwhile, each school can accept at most C students, so (68) implies

$$\frac{1}{n}\sum_{(k,l)\in\mathcal{Z}} Y_{kl}(t) \le \frac{1}{n}|\{s : \mathcal{M}_s \neq \emptyset\}| \le C.$$

Therefore, we define

$$\mathcal{Y} = \{y \in \mathbb{R}_+^{\mathcal{J}} \times \mathbb{R}_+^{\mathcal{Z}} : \sum_{j\in\mathcal{J}} y_j \le 1, \sum_{(k,l)\in\mathcal{Z}} y_{kl} \le C.\}$$

For $y \in \mathcal{Y}$, $j \in \mathcal{J}$, and $(k,l) \in \mathcal{Z}$, define

$$V(y) = \sum_{j<C} y_j, \tag{69}$$

$$f_j(y) = (y_{j-1} - y_j)\mu(V(y)), \tag{70}$$

$$f_{kl}(y) = \mathcal{L}(l)V(y)(1 - V(y))^{k-1}. \tag{71}$$

When $j = 0$ in (70), we define the value $y_{-1} = 0$ for convenience.

In what follows, we use $Y_j(t)$ and $Y_{kl}(t)$ as shorthand for $Y_j(X(t))$ and $Y_{kl}(X(t))$.

It is clear that the initial condition *(i)* is met, with y_0 equal to the zero vector. The following Lemmas establish that conditions *(ii)*, *(iii)*, *(iv)* (stated in equations (60), (61), (62)) also hold.

Lemma 13 (Bounded)**.** *Under Algorithm 1, if Y_j and Y_{kl} are defined by (67) and (68), then (60) holds with $\beta = 1$.*

Proof. Examining Algorithm 1, note that the values $\{Y_j\}_{j\in\mathcal{J}}$ and $\{Y_{kl}\}_{(k,l)\in\mathcal{Z}}$ change only when a proposal is sent to a school that had previously received $j < C$ proposals and is accepted. In this case, Y_{j+1} increases by one, Y_j decreases by one, exactly one Y_{kl} increases by one, and round t ends. It follows that $||Y(t+1) - Y(t)||_\infty \le 1$, so (60) holds with $\beta = 1$. □

Lemma 14 (Lipschitz)**.** *If the functions f_j and f_{kl} are defined by (70) and (71) then (61) holds with $L = \max(C, C\ell^2/2 + 2\ell)$.*

Proof. We first note that

$$|V(y) - V(\tilde{y})| \le \sum_{j=0}^{C-1} |y_j - \tilde{y}_j| \le C\,||y - \tilde{y}||_\infty. \tag{72}$$

Because the function $h(p) = p(1-p)^{k-1}$ is Lipschitz with constant 1, it follows from (71) that

$$|f_{kl}(y) - f_{kl}(\tilde{y})| \le \mathcal{L}(l)\,|V(y) - V(\tilde{y})| \le C\,||y - \tilde{y}||_\infty.$$

Meanwhile,

$$f_j(y) - f_j(\tilde{y}) = (y_{j-1} - y_j)(\mu(V(y)) - \mu(V(\tilde{y}))) + (y_{j-1} - \tilde{y}_{j-1} + \tilde{y}_j - y_j)\mu(V(\tilde{y})),$$

and therefore

$$\begin{aligned}|f_j(y) - f_j(\tilde{y})| &\le |y_{j-1} - y_j|\,|\mu(V(y)) - \mu(V(\tilde{y}))| + (|y_{j-1} - \tilde{y}_{j-1}| + |\tilde{y}_j - y_j|)\mu(V(\tilde{y})) \\ &\le |y_{j-1} - y_j|\,|V(y) - V(\tilde{y})|\,\ell^2/2 + (|y_{j-1} - \tilde{y}_{j-1}| + |\tilde{y}_j - y_j|)\ell \\ &\le C\ell^2/2\,||y - \tilde{y}||_\infty + 2\ell\,||y - \tilde{y}||_\infty\,,\end{aligned}$$

where the second line follows from Lemma 4, and the third from (72) and the fact $|y_{j-1} - y_j| \le 1$. □

Define

$$g(v, k) = \binom{n-v}{k} / \binom{n}{k}.$$

This gives the chance that none of the first k schools on a student's list have vacancies, given that a total of v out of n schools have vacancies.

Lemma 15 (Trend). *Under Algorithm 1 with a single lottery, if $\{Y_j\}_{0\le j<C}$ and $\{Y_{kl}\}_{1\le k\le l\le \ell}$ are defined as in (67) and (68), then*

$$\mathbb{E}[Y_{kl}(t+1) - Y_{kl}(t)|X(t)] = \mathcal{L}(l)\Big(g\big(V(Y(t)), k-1\big) - g\big(V(Y(t)), k\big)\Big). \tag{73}$$

$$\mathbb{E}[Y_j(t+1) - Y_j(t)|X(t)] = (Y_{j-1}(t) - Y_j(t))\sum_{k=0}^{\ell-1}\frac{\mathcal{L}(>k)}{n-k}g(V(Y(t)), k). \tag{74}$$

Furthermore, if the functions f_j and f_{kl} are defined by (70) and (71), then (62) holds with $\delta = \ell^3$.

Proof. Note that Y_{kl} increases by one if and only if student s_t lists exactly l schools (which occurs with probability $\mathcal{L}(l)$) and matches to her k^{th} choice (which occurs with probability $g(V(Y(t)), k-1) - g(V(Y(t)), k)$). This directly implies (73).

Meanwhile, Y_j increases by one if student s_t matches to a school with $j-1$ other proposals, and decreases by one if student s_t matches to a school with j other proposals. The probability that student s_t is rejected from her first k schools and lists a school with $j < C$ other proposals as her $(k+1)^{st}$ choice is $\mathcal{L}(>k)g(V(Y(t)), k)\frac{Y_j}{n-k}$. Summing across k yields (74).

To prove that (62) holds, we note that $g(v, k) \approx \big(1 - \frac{v}{n}\big)^k$. More precisely, for $1 \le k \le \ell$,

$$\left(\frac{(n-v-\ell+1)_+}{n}\right)^k \le g(v, k) \le \left(\frac{n-v}{n}\right)^k.$$

It follows that

$$\begin{aligned}0 \le \Big(1 - \frac{v}{n}\Big)^k - g(v, k) &\le \Big(1 - \frac{v}{n}\Big)^k - \left(\frac{(n-v-\ell+1)_+}{n}\right)^k \\ &\le k\left(\frac{n-v}{n} - \frac{(n-v-\ell+1)_+}{n}\right) \\ &\le \frac{\ell^2}{n},\end{aligned} \tag{75}$$

where the second line uses the fact that for $a, b \in [0,1]$ with $a \geq b$, $a^k - b^k \leq k(a-b)$. In the remainder of the proof, we use Y in place of $Y(t)$ to reduce notational clutter. From (75) and the definition of f_{kl} in (71), we have

$$\begin{aligned} &|\,\mathbb{E}[Y_{kl}(t+1) - Y_{kl}(t)|X(t)] - f_{kl}(Y(t)/n)| \\ &\qquad = \mathcal{L}(l)\left|g(V(Y), k-1) - (1 - V(Y/n))^{k-1} + (1 - V(Y/n))^k - g(V(Y), k)\right| \\ &\qquad \leq \mathcal{L}(l)\ell^2/n \\ &\qquad \leq \ell^2/n. \end{aligned}$$

Furthermore, (74) implies that

$$\begin{aligned} |\mathbb{E}[Y_j(t+1) - Y_j(t)|X(t)] - f_j(Y(t)/n)| &= \left|\left(\frac{Y_{j-1}}{n} - \frac{Y_j}{n}\right)\sum_{k=0}^{\ell-1}\mathcal{L}(>k)\left(\frac{n}{n-k}g(V(Y), k) - (1 - V(Y/n))^k\right)\right| \\ &\leq \left|\frac{Y_{j-1}}{n} - \frac{Y_j}{n}\right|\sum_{k=0}^{\ell-1}\mathcal{L}(>k)\left|\frac{n}{n-k}g(V(Y), k) - (1 - V(Y/n))^k\right|. \\ &\leq \sum_{k=0}^{\ell-1}\mathcal{L}(>k)\ell^2/n \\ &\leq \ell^3/n. \end{aligned}$$

This establishes (62). Note that the third line follow from (75): if the rightmost term in absolute values is negative, then it is at most ℓ^2/n by (75), while if it is positive, then (75) implies that it is at most $\frac{k}{n-k}g(V(y), k)$, which is at most $2\ell/n \leq \ell^2/n$ (for all $n \geq 2\ell$). $\square$

B.4. **Solving The Differential Equation.**

Lemmas 13, 14 and 15 establish that conditions (60), (61) and (62) of Theorem 4 are met. Therefore, (64) implies that for any $\epsilon > 0$ and $1 \leq k \leq l \leq \ell$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k'=1}^{k} Y_{k'l}(\rho n) - \sum_{k'=1}^{k}\hat{y}_{k'l}(\rho)\right| > \epsilon\right) \to 0.$$

Note that

$$G^n(k, l) = \sum_{k'=1}^{k} Y_{k'l}(\rho n).$$

Therefore, to complete the proof of Theorem 1 in the case of a single lottery, all that remains is to show that for any $1 \leq k \leq l \leq \ell$,

$$\sum_{k'=1}^{k}\hat{y}_{k'l}(\rho) = \rho\mathcal{L}(l)F_S(k). \tag{76}$$

This is established by the following Lemma.

Lemma 16. *When f_j and f_{kl} are defined by (70) and (71), the solution $\hat{y}$ to (63) is given by*

$$\hat{y}_j(s) = p_j(\Lambda_S(1 - s/\rho)), \tag{77}$$

$$\hat{y}_{kl}(s) = \rho\mathcal{L}(l) \int_{1-s/\rho}^{1} \mathcal{V}(\Lambda_S(t))(1 - \mathcal{V}(\Lambda_S(t)))^{k-1} dt, \tag{78}$$

where p_j is defined in (15), Λ_S in (6), and $\mathcal{V}$ in (8). Furthermore, (76) holds for any $1 \le k \le l \le \ell$.

Proof. These clearly satisfy the initial condition: $\hat{y}_0(0) = 1$, and $\hat{y}_{kl}(0) = 0$ for $k, l \in \mathcal{Z}$. Furthermore,

$$\begin{aligned}
\hat{y}_j'(s) &= p_j'(\Lambda_S(1 - s/\rho))\Lambda_S'(1 - s/\rho)(-1/\rho) \\
&= p_j'(\Lambda_S(1 - s/\rho))\mu(\mathcal{V}(\Lambda_S(1 - s/\rho))) \\
&= (\hat{y}_{j-1}(s) - \hat{y}_j(s))\mu(\mathcal{V}(\Lambda_S(1 - s/\rho))) \\
&= (\hat{y}_{j-1}(s) - \hat{y}_j(s))\mu(V(\hat{y}(s))). \\
&= f_j(\hat{y}(s)).
\end{aligned}$$

The first equality follows from differentiating (77), the second from applying (6), the third from the fact that $p_j'(\lambda) = p_{j-1}(\lambda) - p_j(\lambda)$ and (77), the fourth from the fact that

$$\mathcal{V}(\Lambda_S(1 - s/\rho))) = \sum_{j<C} p_j(\Lambda_S(1 - s/\rho)) = V(\hat{y}(s)), \tag{79}$$

and the last from (70). Meanwhile, differentiating (78) and applying (79) and (71) yields

$$\begin{aligned}
\hat{y}_{kl}'(s) &= \mathcal{L}(l)\mathcal{V}(\Lambda_S(1 - s/\rho))(1 - \mathcal{V}(\Lambda_S(1 - s/\rho)))^{k-1} \\
&= \mathcal{L}(l)V(\hat{y}(s))(1 - V(\hat{y}(s)))^{k-1} \\
&= f_{kl}(\hat{y}(s)).
\end{aligned}$$

This establishes that $\hat{y}$ solves (63).

Meanwhile (76) follows from (78) and (7):

$$\begin{aligned}
\sum_{k'=1}^{k} \hat{y}_{k'l}(\rho) &= \rho\mathcal{L}(l) \int_0^1 \sum_{k'=1}^{k} \mathcal{V}(\Lambda_S(t))(1 - \mathcal{V}(\Lambda_S(t)))^{k'-1} dt \\
&= \rho\mathcal{L}(l) \int_0^1 1 - (1 - \mathcal{V}(\Lambda_S(t)))^k dt \\
&= \rho\mathcal{L}(l)F_S(k).
\end{aligned}$$

□

C. Proof of Theorem 1 for Independent Lotteries

C.1. **Markov Chain Description and Summary Statistics.** As was the case for a single lottery, our Markov Chain for independent lotteries uses the sequential proposal order of McVitie and Wilson (1971), along with the principle of deferred decisions. The state of our chain is still given by $X = (L, \mathcal{I}, \mathcal{M})$, where L tracks student list lengths, $\mathcal{I}$ tracks the set of proposals, and $\mathcal{M}$ tracks the current tentative assignment. For a single lottery, we invited students to propose in decreasing order of priority. For independent lotteries, the ordering of students is arbitrary, and we wait to reveal how a school ranks students that have proposed to it until that school is forced to reject some student.

Algorithm 2 (DA with Independent Loteries)**.**
$\mathcal{I} = \mathcal{M} = \emptyset$, $L_s = \emptyset$ for all s.
for $t \in \{1, 2, \ldots, |\mathcal{S}|\}$ ***do***
  $s = s_t$, $L_s \sim \mathcal{L}$.
  while $|\mathcal{I}_s| \leq L_s$ && $\mathcal{M}_s = \emptyset$ ***do***
    Sample h uniformly from $\mathcal{H} \backslash \mathcal{I}_s$ and set $\mathcal{I} \leftarrow \mathcal{I} \cup (s, h)$.
    if $|\mathcal{I}_h| \leq C_h$ ***then*** *s fills a vacancy: $\mathcal{M} \leftarrow \mathcal{M} \cup (s, h)$.*
    else *with probability $C_h/|\mathcal{I}_h|$, s triggers a rejection:*
      Select s' uniformly at random from $\mathcal{M}_h$
      $\mathcal{M} \leftarrow \mathcal{M} \cup (s, h) \backslash (s', h)$
      $s \leftarrow s'$
    end if
  end while
end for

The only difference between Algorithm 2 and Algorithm 1 for a single lottery is the "else" clause above, which accounts for the fact that the proposing student can trigger the rejection of another student. This introduces significant additional complexity to the proof. Under single lotteries, when a student proposed to a school that had already received at least C proposals, that student was rejected. As a result, it sufficed to track the number of schools that had received j proposals for each $j < C$. By contrast, with independent lotteries, we need to also track this number for $j \geq C$. Furthermore, the probability that a new proposal triggers the rejection of a student who was tentatively matched to that school depends on the number of students who have previously proposed to that school: the higher this number, the less likely that a new proposal will be accepted. For this reason, our summary statistics must track not only the number of students who list l schools and are tentatively assigned to their k^{th} choice, but also the number of proposals received by the schools to which these students are matched.

More precisely, given state $X = \{L, \mathcal{I}, \mathcal{M})$, we define the following summary statistics:

$$\mathcal{S}_{jkl}(X) = \{s : \mathcal{M}_s \neq \emptyset, |\mathcal{I}_{\mathcal{M}_s}| = j, |\mathcal{I}_s| = k, L_s = l\}. \tag{80}$$

$$Y_{jkl}(X) = |\mathcal{S}_{jkl}(X)|. \tag{81}$$

In English, $\mathcal{S}_{jkl}$ is the set of students who are tentatively matched to a school which has received interest from a total of j students, and rank that school k^{th} among l listed schools. Meanwhile, Y_{jkl} is the number of such students.

C.2. **The Differential Equation Method.** The complexities discussed above necessitate modifications to Theorem 4. Because we now need to track the number of schools with j proposals for $j > C$, the dimension d of our summary statistics must grow with n (rather than remaining bounded, as was the case for a single lottery). Wormald (1999) notes that the proof can accommodate this modification. A second challenge is that the difference $Y(t+1) - Y(t)$ is no longer bounded by an absolute constant. With a single lottery, the number of proposals in round t is at most the length of the list of student s_t, which is at most ℓ by assumption. By contrast, with independent lotteries, a student can trigger a rejection chain that is longer than the length of any single student's list. Fortunately, Theorem 5.1 of Wormald (1999) allows for the bound on $Y(t+1)-Y(t)$ to hold with high probability, rather than with probability one. This is sufficient for our purposes: because the probability that the next proposal is sent to a school with a vacancy is lower-bounded by a constant, the total number of proposals in a round is stochastically dominated by a geometric random variable, whose tail probabilities decay exponentially. One final change is that it is most convenient to use the L_1-norm rather than the L_∞-norm. The L_1-norm is also used by Wormald (1995), and Wormald (1999) notes that it gives an equivalent result. Combining these changes, we use an adaptation of Theorem 5.1 from Wormald (1999) in which the dimension of the summary statistics varies with n, and which replaces conditions *(ii)*, *(iii)*, *(iv)* from Theorem 4 with the following:

(ii') (Boundedness) There exists β_n and γ_n such that if $Y^n(X^n(t))/n \in \mathcal{Y}$, then

$$\mathbb{P}(||Y^n(X^n(t+1)) - Y^n(X^n(t))||_1 > \beta_n) < \gamma_n. \tag{82}$$

Furthermore, there exists $d_n \in \mathbb{N}$ such that $Y^n_{jkl}(X^n(t)) = 0$ if $j > d_n$, and as $n \to \infty$, $\beta_n \cdot n^{-1/4} \to 0$ and $n \cdot d_n \cdot \gamma_n \to 0$.

(iii') (Lipschitz hypothesis) There exists $L > 0$ such that for all $y, \tilde{y} \in \mathcal{Y}$,

$$||f(y) - f(\tilde{y})||_1 \leq L\, ||y - \tilde{y}||_1 \,. \tag{83}$$

(iv') (Trend hypothesis) There exists $\delta > 0$ such that if $Y^n(X^n(t))/n \in \mathcal{Y}$,

$$||\mathbb{E}[Y^n(X^n(t+1)) - Y^n(X^n(t))|X^n(t)] - f(Y^n(X^n(t))/n)||_1 \leq \delta/n. \tag{84}$$

We briefly comment on the latter part of the boundedness condition *(ii')*. Wormald's result establishes that the probability of a large deviation is at most $\mathcal{O}(nd_n\gamma_n + n\beta_n/\lambda_n e^{-n\lambda_n^3/\beta_n^3})$, where λ_n is $o(1)$ but can go to zero arbitrarily slowly. For this reason, we require that $nd_n\gamma_n \to 0$. Meanwhile, if $\beta_n n^{-1/4} \to 0$, then the latter error term is at most $n^{5/4}/\lambda_n \cdot e^{-n^{1/4}\lambda_n^3}$, which tends to zero so long as λ_n decays sufficiently slowly.

C.3. **Analysis of a single proposal.** Although our eventual goal is to determine the expected change in our summary statistics Y due to the addition of a single student (or in other words, the expected change during a single pass through the 'for' loop of Algorithm 2), we start by studying the expected change in Y from a single proposal (in other words, a single pass through the 'while' loop of Algorithm 2). In this subsection, we fix the round t, and use X^i to denote the state after the i^{th} proposal of the round – that is, after the i^{th} time through the while loop of Algorithm 2.

By construction of Algorithm 2, at each state $X = (L, \mathcal{I}, \mathcal{M})$, there is at most one student who is not tentatively assigned and has not proposed to all schools on her list. That is, there is at most one $s = s(X)$ for whom $\mathcal{M}_s = \emptyset$ and $|\mathcal{I}_s| < L_s$. This student will propose to the next school on her list. We now give exact expressions for the expected change in Y from this proposal. Given a state $X = (L, \mathcal{I}, \mathcal{M})$ such that $s(X) \neq \emptyset$, we define

$$\tilde{S}_{jkl}(X) = \mathcal{S}_{jkl} \backslash \{s' : \mathcal{M}_{s'} \in \mathcal{I}_{s(X)}\} \tag{85}$$

$$\tilde{Y}_{jkl}(X) = |\tilde{\mathcal{S}}_{jkl}(X)|, \tag{86}$$

$$\tilde{n}(X) = |\mathcal{H} \backslash \mathcal{I}_{s(X)}| = n - |\mathcal{I}_{s(X)}|, \tag{87}$$

$$\mathcal{Y} = \{y \in \mathbb{R}_+^{\mathbb{N} \times \mathcal{Z}} : ||y||_1 \leq C, v(y) > \mathcal{V}(\rho\ell)\}. \tag{88}$$

The values $\tilde{Y}$ and $\tilde{n}$ are analogs of Y and n which reflect the fact that the next school to receive a proposal from s is sampled from $\mathcal{H} \backslash \mathcal{I}_s$, rather than the full set of schools $\mathcal{H}$. Note that $\sum_{j \in \mathbb{N}, (k,l) \in \mathcal{Z}} Y_{jkl}$ is equal to the total number of students who are tentatively assigned, which can never exceed Cn, so $||Y(X)/n||_1$ is always at most C. Similarly, $\sum_{j \in \mathbb{N}, (k,l) \in \mathcal{Z}} \tilde{Y}_{jkl}$ is equal to the total number of students who are tentatively assigned to a school that s has yet to propose to, which cannot exceed $C\tilde{n}$, so $||\tilde{Y}(X)/\tilde{n}||_1 \leq C$. The condition $v(y) > \mathcal{V}(\rho\ell)$ is needed to establish the Lipschitz condition *(iii')*, and will be discussed in more detail later.

In order to track the evolution of Y, it is important to keep track of the rank that the proposing student $s(X)$ assigns to the next school on his or her list. We define

$$\tilde{Z}(X) = (|\mathcal{I}_{s(X)}| + 1, L_{s(X)}) \in \mathcal{Z}, \tag{89}$$

which tracks both the rank of the school that s is about to propose to, and the total number of schools listed by s. That is, $\tilde{Z}(X) = (k, l)$ if $s(X)$ is about to propose to the k^{th} of l listed schools. We also define $\tilde{Z}(X) = \emptyset$ if $s(X) = \emptyset$ (that is, if the round has just ended).

Our next lemma gives the exact expected change in Y from a single proposal, expressed as a function of $\tilde{Y}$, $\tilde{n}$, and $\tilde{Z}$. To state it, we first define some notation. For $(k, l) \in \mathcal{Z}$, and $Z \in \mathcal{Z} \cup \{\emptyset\}$, let $\mathbf{1}_{kl}(Z)$ be the indicator that $Z = (k, l)$, and let $\mathbf{1}_{\mathcal{Z}}(Z)$ be the indicator that $Z \in \mathcal{Z}$ (equivalently, that $Z \neq \emptyset$). In what follows, we use Y^i, $\tilde{Y}^i$, $\tilde{n}^i$ and $\tilde{Z}^i$ as shorthand for $Y(X^i)$, $\tilde{Y}(X^i)$, $\tilde{n}(X^i)$ and $\tilde{Z}(X^i)$, respectively.

Lemma 17 (Expected change from a single proposal)**.** *Given any $y \in \mathcal{Y}$, for any $j \in \mathbb{N}$ and $(k, l) \in \mathcal{Z}$, define*

$$\Delta_{jkl}(y) = \left(1 - \frac{\mathbf{1}(j > C)}{j}\right) y_{(j-1)kl} - y_{jkl}. \tag{90}$$

$$P_j(y) = \frac{1}{\min(j, C)} \sum_{1 \leq k \leq l \leq \ell} y_{jkl}, \qquad P_0(y) = 1 - \sum_{j \geq 1} P_j(y). \tag{91}$$

$$v(y) = \sum_{j=0}^{C-1} P_j(y) \tag{92}$$

$$A_j(y) = P_{j-1}(y) \min\left(1, C/j\right). \tag{93}$$

$$a(y) = \sum_{j \geq 1} A_j(y). \tag{94}$$

$$D_{kl}(y) = \sum_{j \geq C} \frac{y_{jkl}}{j+1}. \tag{95}$$

Then

$$\mathbb{E}[Y^{i+1}_{jkl} - Y^{i}_{jkl} \mid X^i] = \quad \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i)\Delta_{jkl}(\tilde{Y}^i/\tilde{n}^i) + \mathbf{1}_{kl}(\tilde{Z}^i)A_j(\tilde{Y}^i/\tilde{n}^i), \tag{96}$$

$$\mathbb{E}[\mathbf{1}_{kl}(\tilde{Z}^{i+1}) \mid X^i] = \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i)D_{(k-1)l}(\tilde{Y}^i/\tilde{n}^i) + \mathbf{1}_{(k-1)l}(\tilde{Z}^i)(1 - a(\tilde{Y}^i/\tilde{n}^i)), \tag{97}$$

$$\mathbb{E}[\mathbf{1}_{\emptyset}(\tilde{Z}^{i+1}) \mid X^i] = v(\tilde{Y}^i/\tilde{n}) + D_{kl}(\tilde{Y}^i/\tilde{n}) + \sum_{l=1}^{\ell} \mathbf{1}_{ll}(\tilde{Z}^i)(1 - a(\tilde{Y}^i/\tilde{n})). \tag{98}$$

Proof. In this proof, we drop all superscript i's to reduce clutter (so for example we write Y_{jkl} instead of Y^i_{jkl}). As we will see below, y, P, A, D have the following interpretations:

- y_{jkl} represents the expected number of students in $\mathcal{S}_{jkl}(X)$ who are tentatively matched to the next school that student s proposes to.
- $P_j(y)$ represents the probability that the next proposal goes to a school that has received j *other* proposals,
- $A_j(y)$ represents the probability that the next proposal is the j^{th} sent to that school, and is accepted (and therefore $a(y)$ represents the probability that the next proposal is accepted).
- $D_{kl}(y)$ represents the probability that the next proposal causes the displacement of another student from her k^{th} choice of l listed schools.

In particular, if $s = s(X)$ is the next student to propose, then

1. Each school $h \in \mathcal{H}\backslash\mathcal{I}_s$ has a $1/|\mathcal{H}\backslash\mathcal{I}_s| = 1/\tilde{n}$ chance of receiving the next proposal.
2. The chance that the next proposal is the j^{th} proposal received by the school is $\mathbb{P}(h \in \mathcal{H}_{j-1}) = P_{j-1}(\tilde{Y}/\tilde{n})$.
3. The j^{th} proposal at a school has a $\min(C, j)/j$ chance of being accepted.
4. Each student previously matched to a school receiving a j^{th} proposal is rejected with probability $\mathbf{1}(j > C)/j$

We start by proving (96). Recall that Y_{jkl} is a sum of indicators across students. We analyze the expected change in Y_{jkl} by partitioning students into three groups:

(I) Students matched to schools in $\mathcal{I}_s$ (that is, schools where the next student to propose has already been rejected).
(II) Students matched to schools in $\mathcal{H}\backslash\mathcal{I}_s$ (that is, schools where the next student to propose has not yet proposed).
(III) Students that are currently unmatched.

For the first group, the next proposal cannot be sent to their school, so their contribution to Y_{jkl} cannot change.

Fix a student s' in the second group. This student's contribution to Y_{jkl} changes only if the school to which she is assigned receives the next proposal. Because the next proposal is sent to a uniformly random school in $\mathcal{H}\backslash\mathcal{I}_s$, the probability of this event is $1/\tilde{n}$. On this event,

- If $s' \in \tilde{S}_{jkl}$, then s' previously contributed to Y_{jkl}, and no longer does (because the school has now received $j + 1$ proposals).
- If $s' \in \tilde{S}_{(j-1)kl}$, then s' previously did not contribute to Y_{jkl}, but now does *unless* the proposal causes s' to be rejected. If $j \leq C$, new proposals will fill a vacancy, and not cause any rejections. If $j > C$, then the proposing student is accepted *if* their priority is among

the top C of the j proposals that the school has received. Because priorities are iid uniform across student-school pairs, the probability of this is C/j, and in this case, each student that was previously matched to the school has an identical $1/C$ chance of being rejected. Thus, the chance that s' is not rejected is $(1 - \mathbf{1}(j > C)/j)$.

- If $s' \in \tilde{S}_{j'kl}$ for $j' \notin \{j-1, j\}$, then s' does not contribute to Y_{jkl} before or after the proposal.

Putting these facts together, the overall expected change in Y_{jkl} due to the second group of students is $-\tilde{Y}_{jkl}/\tilde{n} + (1 - \mathbf{1}(j > C)/j)\tilde{Y}_{(j-1)kl}/\tilde{n} = \Delta_{jkl}(\tilde{Y}/\tilde{n})$.

Finally, no unmatched student contributes to Y_{jkl} before the proposal, and only the current proposer s can contribute to Y_{jkl} after the proposal. Furthermore, s can contribute to Y_{jkl} only if s is proposing to her k^{th} choice of l listed schools – that is, if $\tilde{Z} = (k, l)$. In addition, s must propose to a school that has received $j - 1$ other proposals, and be accepted. The probability of the first event is $P_{j-1}(\tilde{Y}/\tilde{n})$, as $\frac{1}{\min(C,j-1)} \sum_{(k,l)\in\mathcal{Z}} \tilde{Y}_{(j-1)kl}$ gives the number of schools with $j - 1$ proposals that s has not proposed to, and each such school has a $1/\tilde{n}$ chance of receiving the next proposal from s. Furthermore, if s becomes the j^{th} student to propose to a school, s is accepted if among the top $\min(C, j)$ proposing students, which occurs with probability $\min(C, j)/j$. Putting it all together, the probability that the proposing student s contributes to Y_{jkl} after the proposal is $\mathbf{1}_{kl}(\tilde{Z}^i)A_j(\tilde{Y}^i/\tilde{n}^i)$.

We now turn to (97). In order for there to be an $(i + 1)^{st}$ proposal, the i^{th} proposal must either be rejected or trigger the rejection of another student. In either case, $\tilde{Z}^{i+1} = (k, l)$ if and only if the student making the $(i + 1)^{st}$ proposal was just rejected from her $(k - 1)^{st}$ choice of l listed schools. As argued above, the probability that the i^{th} proposal is rejected is $1-a(\tilde{Y}/\tilde{n})$. Meanwhile, the probability that this proposal triggers the rejection of a student from her $(k - 1)^{st}$ choice of l schools is $D_{(k-1)l}(\tilde{Y}/\tilde{n})$, because each student contributing to $\tilde{Y}_{j(k-1)l}$ is rejected with probability $1/\tilde{n} \times 1/(j + 1)$, as explained above. □

Lemma 18 (Lipschitz properties for the expected change from a proposal)**.** *For any $y, \tilde{y} \in \mathcal{Y}$,*

$$||\Delta(y) - \Delta(\tilde{y})||_1 \leq 2\, ||y - \tilde{y}||_1 \,. \tag{99}$$

$$|a(y) - a(\tilde{y})| \leq ||A(y) - A(\tilde{y})||_1 \leq ||P(y) - P(\tilde{y})||_1 \,. \tag{100}$$

$$\sum_{(k,l)\in\mathcal{Z}} |D_{kl}(y) - D_{kl}(\tilde{y})| \leq \frac{1}{C}\, ||y - \tilde{y}||_1 \,. \tag{101}$$

Proof. Note that

$$\begin{aligned} |\Delta_{jkl}(y) - \Delta_{jkl}(\tilde{y})| &= \left|(y_{(j-1)kl} - \tilde{y}_{(j-1)kl})(1 - \mathbf{1}(j > C)/j) + y_{jkl} - \tilde{y}_{jkl}\right| \\ &\leq \left|y_{(j-1)kl} - \tilde{y}_{(j-1)kl}\right| + |y_{jkl} - \tilde{y}_{jkl}| \,. \end{aligned}$$

Summing across j, k, l yields (99).

Meanwhile, (100) holds because

$$\begin{aligned}|a(y)-a(\tilde{y})| &= \left|\sum_{j\geq 1} A_j(y)-A_j(\tilde{y})\right| \\ &\leq \sum_{j\geq 1} |A_j(y)-A_j(\tilde{y})| \\ &= \sum_{j\geq 1} |P_{j-1}(y)-P_{j-1}(\tilde{y})| \min(C,j)/j \\ &\leq \sum_{j\geq 1} |P_{j-1}(y)-P_{j-1}(\tilde{y})|\,.\end{aligned}$$

Finally, (101) follows from summing the following across k,l:

$$\begin{aligned}|D_{kl}(y)-D_{kl}(\tilde{y})| &= \left|\sum_{j\geq C} \frac{y_{jkl}-\tilde{y}_{jkl}}{j+1}\right| \\ &\leq \sum_{j\geq C} \frac{|y_{jkl}-\tilde{y}_{jkl}|}{j+1} \\ &\leq \sum_{j\geq C} \frac{|y_{jkl}-\tilde{y}_{jkl}|}{C}.\end{aligned}$$

□

Lemma 19 (Bounding the difference between sampling with and without replacement)**.** *For any X that can arise in the execution of Algorithm 2, the following hold:*

$$\left|\left|Y/n-\tilde{Y}/\tilde{n}\right|\right|_1 \leq \frac{2C(\ell-1)}{n}. \tag{102}$$

$$\left|\left|P(Y/n)-P(\tilde{Y}/\tilde{n})\right|\right|_1 \leq \frac{2(\ell-1)}{n}. \tag{103}$$

$$\left|\left|\Delta(Y/n)-\Delta(\tilde{Y}/\tilde{n})\right|\right|_1 \leq \frac{4C(\ell-1)}{n}. \tag{104}$$

$$\left|\left|A(Y/n)-A(\tilde{Y}/\tilde{n})\right|\right|_1 \leq \frac{2(\ell-1)}{n}. \tag{105}$$

$$\left|\left|D(Y/n)-D(\tilde{Y}/\tilde{n})\right|\right|_1 \leq \frac{2(\ell-1)}{n}. \tag{106}$$

Proof. Let $s=s(X)$ be the student that is about to propose, and note that

$$\begin{aligned}\frac{Y_{jkl}}{n}-\frac{\tilde{Y}_{jkl}}{\tilde{n}} &= \frac{Y_{jkl}-\tilde{Y}_{jkl}}{n}-\frac{(n-\tilde{n})}{n}\frac{\tilde{Y}_{jkl}}{\tilde{n}} \\ &= \frac{1}{n}|\mathcal{S}_{jkl}\cap\{s':\mathcal{M}_{s'}\in\mathcal{I}_s\}|-\frac{|\mathcal{I}_s|}{n}\frac{|\mathcal{S}_{jkl}\backslash\{s':\mathcal{M}_{s'}\in\mathcal{I}_s\}|}{|\mathcal{H}\backslash\mathcal{I}_s|}.\end{aligned} \tag{107}$$

Furthermore,

$$\sum_{k,l} |\mathcal{S}_{jkl} \cap \{s' : \mathcal{M}_{s'} \in \mathcal{I}_s\}| = |\{s' : \mathcal{M}_{s'} \in \mathcal{H}_j \cap \mathcal{I}_s\}| = |\mathcal{H}_j \cap \mathcal{I}_s| \min(C, j). \tag{108}$$

$$\sum_{k,l} |\mathcal{S}_{jkl} \backslash \{s' : \mathcal{M}_{s'} \in \mathcal{I}_s\}| = |\{s' : \mathcal{M}_{s'} \in \mathcal{H}_j \backslash \mathcal{I}_s\}| = |\mathcal{H}_j \backslash \mathcal{I}_s| \min(C, j). \tag{109}$$

Combining (107), (108), and (109) yields (102):

$$\begin{aligned}\sum_{j,k,l} \left| \frac{Y_{jkl}}{n} - \frac{\tilde{Y}_{jkl}}{\tilde{n}} \right| &\le \frac{1}{n} \sum_j \left(|\mathcal{H}_j \cap \mathcal{I}_s| \min(C, j) + |\mathcal{I}_s| \frac{|\mathcal{H}_j \backslash \mathcal{I}_s| \min(C, j)}{|\mathcal{H} \backslash \mathcal{I}_s|} \right). \\ &\le \frac{2C|\mathcal{I}_s|}{n} \le \frac{2C(\ell - 1)}{n}.\end{aligned}$$

Furthermore, (103) follows from (91), (107), (108), and (109):

$$\begin{aligned}P_j(Y/n) - P_j(\tilde{Y}/\tilde{n}) &= \frac{1}{\min(C, j)} \sum_{(k,l) \in \mathcal{Z}} \left(\frac{Y_{jkl}}{n} - \frac{\tilde{Y}_{jkl}}{\tilde{n}} \right) \\ &\le \frac{1}{n} \sum_j \left(|\mathcal{H}_j \cap \mathcal{I}_s| + |\mathcal{I}_s| \frac{|\mathcal{H}_j \backslash \mathcal{I}_s|}{|\mathcal{H} \backslash \mathcal{I}_s|} \right). \\ &= \frac{2|\mathcal{I}_s|}{n} \le \frac{2(\ell - 1)}{n}.\end{aligned}$$

Finally, (104) (105), (106) follow from (102), (103), and Lemma 18. $\square$

Lemma 20 (Bounding the change from a single step)**.** *The following hold with probability one:*

$$\left|\left| Y_{jkl}^{i+1}/n - Y_{jkl}^i/n \right|\right|_1 \le 2C/n. \tag{110}$$

$$\left|\left| P(Y^{i+1}/n) - P(Y^i/n) \right|\right|_1 \le 2/n. \tag{111}$$

$$\left|\left| \Delta(Y^{i+1}/n) - \Delta(Y^i/n) \right|\right|_1 \le 4C/n. \tag{112}$$

$$\left|\left| A(Y^{i+1}/n) - A(Y^i/n) \right|\right|_1 \le 2/n. \tag{113}$$

$$\left|\left| D(Y^{i+1}/n) - D(Y^i/n) \right|\right|_1 \le 2/n. \tag{114}$$

Proof. Note that if the next proposal is sent to a school $h \in \mathcal{H}_{j-1}$, then the values $\{Y_{(j-1)kl}\}_{k,l}$ all weakly decrease, the values $\{Y_{jkl}\}_{k,l}$ all weakly increase, and $Y_{j'kl}$ is constant for $j' \notin \{j-1, j\}$. Only students that were previously matched to h stop contributing to $\sum_{k,l} Y_{(j-1)kl}$, and only those matched to h are new contributors to $\sum_{k,l} Y_{jkl}$, so

$$\begin{aligned}\sum_{(k,l) \in \mathcal{Z}} \left| Y_{(j-1)kl}^{i+1} - Y_{(j-1)kl}^i \right| &= \min(C, j-1) \le C, \\ \sum_{(k,l) \in \mathcal{Z}} \left| Y_{jkl}^{i+1} - Y_{jkl}^i \right| &= \min(C, j) \qquad \le C.\end{aligned}$$

These jointly imply (110). Furthermore, combining these inequalities with (91) implies (111). Finally, combining (110) and (111) with Lemma 18 immediately yields (112), (113), and (114). $\square$

C.4. **Analysis of A Full Round.** We now study the expected change in the variables Y_{jkl} over a full round (pass through the 'for' loop) of Algorithm 2. Although our preceding analysis of a single proposal will be helpful, several challenges remain. First, the number of proposals in the round is unknown. Second, (97) reveals that the expected change in Y_{jkl} from a proposal depends on the value $\mathbf{1}_{(k-1)l}(\tilde{Z})$. In other words, we must understand not only the total number of proposals, but also how often a student proposes to her k^{th} of l listed schools (for each k and l). A third challenge is that the probability of each proposal being accepted evolves over the course of the round.

We address these challenges by introducing a Markov chain Z on $\mathcal{Z} \cup \{\emptyset\}$. This chain is intended to approximate the evolution of the true variable $\tilde{Z}$. Compared to $\tilde{Z}$, transition dynamics for Z are simplified in two ways. First, transition probabilities for Z are based only the summary statistics Y: while the evolution of $\tilde{Z}$ accounts for the fact that students sample schools without replacement, Z implicitly assumes that students sample schools with replacement. Second, the transition probabilities for Z are fixed throughout a round: although the evolution of $\tilde{Z}$ accounts for the fact that competition rises as more proposals are made, Z implicitly assumes that the changes during a single round of proposals are small enough to ignore.

We use the simplified chain Z to define a function f which approximates the expected change in the variables Y over the course of a round.

C.4.1. *Defining the chain Z and the function f.* Given any $y \in \mathcal{Y}$, we define a Markov Chain Z_y on $\mathcal{Z} \cup \{\emptyset\}$. It starts in state $Z_y^0 = (1, l)$, where $l \sim \mathcal{L}$. The state $\emptyset$ is absorbing: it represents a student filling a vacant position or being rejected from the last school on her list, thereby ending the round. We let Z_y^i denote the state of this chain after i transitions. From state $Z_y^i \neq \emptyset$, move to $\emptyset$ with probability

$$\mathbb{E}[\mathbf{1}_\emptyset(Z_y^{i+1}) \mid Z_y^i] = v(y) + \sum_{l=1}^{\ell} D_{ll}(y) + \sum_{l=1}^{\ell} \mathbf{1}_{ll}(Z_y^i)(1 - a(y)), \tag{115}$$

and move to state (k, l) with probability

$$\mathbb{E}[\mathbf{1}_{kl}(Z_y^{i+1}) \mid Z_y^i] = \mathbf{1}_{\mathcal{Z}}(Z_y^i) D_{(k-1)l}(y) + \mathbf{1}_{(k-1)l}(Z_y^i)(1 - a(y)). \tag{116}$$

We now define the functions f which will approximate the change in Y. For $y \in \mathcal{Y}$, $j \in \mathbb{N}$ and $(k, l) \in \mathcal{Z}$, define

$$f_{jkl}(y) = Q(y)\Delta_{jkl}(y) + q_{kl}(y)A_j(y), \tag{117}$$

where

$$Q(y) = \frac{\mu(a(y))}{1 - \sum_{1 \le k \le l \le \ell} \sum_{k'=1}^{k-1} D_{k'l}(y)(1 - a(y))^{k-k'-1}}. \tag{118}$$

$$q_{kl}(y) = \begin{cases} \mathcal{L}(l) & : k = 1. \\ \mathcal{L}(l)(1 - a(y))^{k-1} + Q(y) \sum_{k'=1}^{k-1} D_{k'\ell}(y)(1 - a(y))^{k-k'-1} & : k \ge 2. \end{cases} \tag{119}$$

These functions are opaque, but the following lemma establishes a close relationship between f_{jkl} and the Markov chain Z_y. It uses the fact that for any $q \in [0,1]$,

$$\sum_{1\le k\le l\le \ell} \mathcal{L}(l)(1-q)^{k-1} = \sum_{1\le l\le \ell} \mathcal{L}(l)\frac{1-(1-q)^l}{q} = \mu(q). \tag{120}$$

To preview our next step, note the close relationship between the expression in Lemma 21 and the expected change in Y given by (96) in Lemma 17.

Lemma 21 (Connection to Markov Chain)**.** *For any $y \in \mathcal{Y}$ and any $(k,l) \in \mathcal{Z}$,*

$$f_{jkl}(y) = \mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(y) + \mathbf{1}_{kl}(Z_y^i)A_j(y)\right].$$

Proof of Lemma 21. Define

$$\hat{q}_{kl}(y) = \mathbb{E}[\sum_{i=0}^{\infty} \mathbf{1}_{kl}(Z_y^i)]. \tag{121}$$

$$\hat{Q}(y) = \mathbb{E}[\sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(Z_y^i)]. \tag{122}$$

To complete the proof, it suffices to show that $q_{kl}(y) = \hat{q}_{kl}(y)$ and $Q(y) = \hat{Q}(y)$. Note that $\mathbb{E}[\mathbf{1}_{1l}(Z^0)] = \mathcal{L}(l)$ and $\sum_{i=1}^{\infty} \mathbf{1}_{1l}(Z^i) = 0$, so $\hat{q}_{1l}(y) = \mathcal{L}(l) = q_{1l}(y)$. Furthermore, for $k \ge 2$,

$$\begin{aligned}
\hat{q}_{kl}(y) &= \mathbb{E}\left[\sum_{i=1}^{\infty} \mathbf{1}_{kl}(Z^i)\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{\infty} \mathbb{E}[\mathbf{1}_{kl}(Z^{i+1}) \mid Z^i]\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(Z^i)D_{(k-1)l}(y) + \mathbf{1}_{(k-1)l}(Z^i)(1-a(y))\right] \\
&= \hat{Q}(y)D_{(k-1)l}(y) + \hat{q}_{(k-1)l}(y)(1-a(y)).
\end{aligned}$$

The first equality follows because $\mathbf{1}_{kl}(Z^0) = 0$ for $k \ge 2$, the second from changing indices and the law of iterated expectation, the third from (116), and the final equality by definition of $\hat{q}_{kl}$ and $\hat{Q}$. Recursively expanding this expression yields

$$\hat{q}_{kl}(y) = \mathcal{L}(l)(1-a(y))^{k-1} + \hat{Q}(y)\sum_{k'=1}^{k-1} D_{k'\ell}(y)(1-a(y))^{k-k'-1}. \tag{123}$$

This matches (119), so long as we show that $\hat{Q}(y) = Q(y)$. Summing (123) across k and l and applying (120) yields

$$\hat{Q}(y) = \sum_{1\le k\le l\le \ell} \hat{q}_{kl}(y) = \mu(a(y)) + \hat{Q}(y)\sum_{1\le k\le l\le \ell}\sum_{k'=1}^{k-1} D_{k'l}(y)(1-a(y))^{k-k'-1}.$$

Solving for $\hat{Q}(y)$ establishes that $\hat{Q}(y) = Q(y)$. $\square$

Before establishing that the modified bounded, lipschitz, and trend conditions *(ii')*, *(iii')* and *(iv')* hold for this choice of f, we provide one helpful lemma, which establishes several bounds on the duration of each round.

Lemma 22. *Fix $t \leq |S|$, and let $\tilde{T}$ denote the number of proposals (passes through the 'while' loop) in the t^{th} round of Algorithm 2. For $i \in \mathbb{N}$, let X^i be state of the Markov chain after $\min(i, \tilde{T})$ proposals, and define $Y^i = Y(X^i)$ as in (81). Then*

$$\mathbb{P}(\tilde{T} > i \mid X^0) = \mathbb{E}[\mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i) \mid X^0] \leq (1 - v(Y^0/n))^i. \tag{124}$$

$$\mathbb{E}[\max(\tilde{T} - i, 0) \mid X^i] = \mathbb{E}[\sum_{i'=i}^{\infty} \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^{i'}) \mid X^i] \leq 1/v(Y^0/n). \tag{125}$$

$$\mathbb{E}[\max(\tilde{T} - i, 0)^2 \mid X^i] = \mathbb{E}\left[\left(\sum_{i'=i}^{\infty} \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^{i'})\right)^2 \middle| X^i\right] \leq 2/v(Y^0/n)^2. \tag{126}$$

Furthermore, if $y \in \mathcal{Y}$ and Z_y is defined as in (115) and (116), then for any $i \in \mathbb{N}$,

$$\mathbb{E}[\mathbf{1}_{\mathcal{Z}}(Z_y^i)] \leq (1 - v(y))^i. \tag{127}$$

$$\mathbb{E}[\sum_{i'=i}^{\infty} \mathbf{1}_{\mathcal{Z}}(Z_y^{i'}) \mid Z_y^i] \leq 1/v(y). \tag{128}$$

Proof. We claim that for all i,

$$\mathbb{P}(\tilde{Z}^{i+1} = \emptyset | X^i) \geq v(\tilde{Z}^i/\tilde{n}) \geq v(Y^0/n). \tag{129}$$

This follows from (98), along with the observation that for $j < C$ and all i such that $\tilde{Z}^i \neq \emptyset$,

$$\tilde{Y}^i_{jkl} = Y^i_{jkl} = Y^0_{jkl}. \tag{130}$$

The first equality follows because $\mathcal{I}_{s^i}$ does not include any schools that have received fewer than C applications (otherwise, s^i would currently be tentatively accepted, and would not be proposing). The second equality in (130) holds because none of the proposals in the current round have gone to a school with a vacancy (otherwise, the round would be over and we would have $\tilde{Z}^i = \emptyset$). Because $\tilde{n}^i \leq n$, it follows from (130) that for $j < C$, $\tilde{Y}^i_{jkl}/\tilde{n}^i \geq Y^0_{jkl}/n$, and therefore by (91) $P_j(\tilde{Y}^i/\tilde{n}^i) \geq P_j(Y^0/n)$. This implies the second inequality in (129).

It follows immediately from (129) that the probability that $\tilde{Z}^i \neq \emptyset$ is at most $(1 - v(Y^0/n))^i$. In other words, (124) holds. Similarly, it follows that from any point, the expected number of remaining steps is at most $1/v(Y^0/n)$, so (125) holds. Finally, (126) follows from upper-bounding the left side by the square of a geometric random variable with success probability $v(Y^0/n)$.

Meanwhile, note that (115) implies that at each step, Z_y transitions to $\emptyset$ with probability at least $v(y)$. From this, (127) and (128) follow by arguments analogous to those outlined above. □

C.4.2. *Verifying the modified conditions (ii'), (iii'), (iv').* We note that the initial condition *(i)* of Theorem 4 holds with y^0 a vector of all zeros. We next establish that the modified conditions *(ii')*, *(iii')*, *(iv')* hold.

Lemma 23 (Bounded Condition)**.** *Let X be the Markov chain arising from Algorithm 2, with $X(t)$ denoting the state after t times through the 'for' loop. Let $\mathcal{Y}$ be defined as in (88), and let Y and f be defined as in (81) and (117), respectively. If $d_n = \rho n$, $\beta_n = 2C\log^2(n)$, and $\gamma_n = (1-\mathcal{V}(\rho\ell))^{\log^2(n)}$, then condition (ii') holds.*

Proof. By Lemma 20, in order for $Y^\infty - Y^0$ to be larger than β_n, the round must go more than $\log^2(n)$ steps. By (124) in Lemma 22, the probability of this is at most $(1 - v(Y^0/n))^{\log^2(n)}$. By the definition of $\mathcal{Y}$ in (88), if $Y^0/n \in \mathcal{Y}$, then this probability is at most $(1 - \mathcal{V}(\rho\ell))^{\log^2(n)} = \gamma_n$.

Because no school can receive more than ρn proposals, it is clear that $Y^n_{jkl}(t) = 0$ for $j > d_n$. Furthermore, it is clear that as $n \to \infty$, $\beta_n n^{-1/4} \to 0$ and $n \cdot d_n \cdot \gamma_n = \rho n^{2+\log(1-\mathcal{V}(\rho\ell))\log(n)} \to 0$. $\square$

Lemma 24 (Lipschitz Condition)**.** *Let X be the Markov chain arising from Algorithm 2, with $X(t)$ denoting the state after t times through the 'for' loop. Let $\mathcal{Y}$ be defined as in (88), and let Y and f be defined as in (81) and (117), respectively. Condition (iii') holds with $L = 3/\mathcal{V}(\rho\ell) + 4(2C+1)/\mathcal{V}(\rho\ell)^2$.*

Proof. Let $y, \tilde{y} \in \mathcal{Y}$. The high-level idea of the proof is to use Lemma 21, which shows that $f(y)$ and $f(\tilde{y})$ can be expressed as a sum over the steps before the Markov Chains Z_y and $Z_{\tilde{y}}$ reach the state $\emptyset$. Lemma 22 bounds the expected number of these steps, and the contribution from each step is bounded below.

By Lemma 21, we have

$$\begin{aligned}
||f(y) - f(\tilde{y})||_1 &= \sum_{j,k,l}\left|\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(y) + \mathbf{1}_{kl}(Z_y^i)A_j(y) - \mathbf{1}_{\mathcal{Z}}(Z_{\tilde{y}}^i)\Delta_{jkl}(\tilde{y}) - \mathbf{1}_{kl}(Z_{\tilde{y}}^i)A_j(\tilde{y})\right]\right| \\
&\le \sum_{i=0}^{\infty}\sum_{j,k,l}\left|\mathbb{E}\left[\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(y) + \mathbf{1}_{kl}(Z_y^i)A_j(y) - \mathbf{1}_{\mathcal{Z}}(Z_{\tilde{y}}^i)\Delta_{jkl}(\tilde{y}) - \mathbf{1}_{kl}(Z_{\tilde{y}}^i)A_j(\tilde{y})\right]\right| \\
&\le \sum_{i=0}^{\infty}\sum_{j,k,l}\left|\mathbb{E}\left[\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(y) + \mathbf{1}_{kl}(Z_y^i)A_j(y) - \mathbf{1}_{\mathcal{Z}}(Z_{y}^i)\Delta_{jkl}(\tilde{y}) - \mathbf{1}_{kl}(Z_{y}^i)A_j(\tilde{y})\right]\right| &\quad (131)\\
&\qquad + \left|\mathbb{E}\left[\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(\tilde{y}) + \mathbf{1}_{kl}(Z_y^i)A_j(\tilde{y}) - \mathbf{1}_{\mathcal{Z}}(Z_{\tilde{y}}^i)\Delta_{jkl}(\tilde{y}) - \mathbf{1}_{kl}(Z_{\tilde{y}}^i)A_j(\tilde{y})\right]\right|, &\quad (132)
\end{aligned}$$

where the final line follows from adding and subtracting $\mathbb{E}[\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(\tilde{y}) + \mathbf{1}_{kl}(Z_y^i)A_j(\tilde{y})]$ and applying the triangle inequality. We will bound the terms in (131) and (132). We claim that

$$\begin{aligned}
\sum_{j,k,l}\left|\mathbb{E}\left[\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(y) + \mathbf{1}_{kl}(Z_y^i)A_j(y)\right.\right. &\left.\left. -\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(\tilde{y}) - \mathbf{1}_{kl}(Z_y^i)A_j(\tilde{y})\right]\right| \\
&\le \mathbb{E}[\mathbf{1}_{\mathcal{Z}}(Z_y^i)]\left(||\Delta(y) - \Delta(\tilde{y})||_1 + ||A(y) - A(\tilde{y})||_1\right) \\
&\le \mathbb{E}[\mathbf{1}_{\mathcal{Z}}(Z_y^i)] \times 3\,||y - \tilde{y}||_1 .
\end{aligned}$$

The first inequality uses the triangle inequality, and the second uses Lemma 18, which states that $||\Delta(y) - \Delta(\tilde{y})||_1 \le 2\,||y - \tilde{y}||_1$ and $||A(y) - A(\tilde{y})||_1 \le ||P(y) - P(\tilde{y})||_1 \le ||y - \tilde{y}||_1$. Therefore, the term in (131) can be upper-bounded as follows:

$$\begin{aligned}\sum_{i=0}^{\infty}\sum_{j,k,l}\left|\mathbb{E}\left[\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(y)+\mathbf{1}_{kl}(Z_y^i)A_j(y)-\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(\tilde{y})\right.\right.&\left.\left.-\mathbf{1}_{kl}(Z_y^i)A_j(\tilde{y})\right]\right|\\ &\leq 3\left|\left|y-\tilde{y}\right|\right|_1\sum_{i=0}^{\infty}\mathbb{E}[\mathbf{1}_{\mathcal{Z}}(Z_y^i)]\\ &\leq 3\left|\left|y-\tilde{y}\right|\right|_1/\mathcal{V}(\rho\ell), \end{aligned}\tag{133}$$

where we have used (128) in Lemma 22 and the fact that $v(y)>\mathcal{V}(\rho\ell)$ by definition of $\mathcal{Y}$ in (88).

Turning to (132), we claim that

$$\sum_{j,k,l}\left|\mathbb{E}\left[\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(\tilde{y})+\mathbf{1}_{kl}(Z_y^i)A_j(\tilde{y})-\mathbf{1}_{\mathcal{Z}}(Z_{\tilde{y}}^i)\Delta_{jkl}(\tilde{y})-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)A_j(\tilde{y})\right]\right|$$

$$\leq\left|\mathbb{E}[\mathbf{1}_{\mathcal{Z}}(Z_y^i)-\mathbf{1}_{\mathcal{Z}}(Z_{\tilde{y}}^i)]\right|\left|\left|\Delta(\tilde{y})\right|\right|_1+\sum_{k,l}\left|\mathbb{E}[\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)]\right|\left|\left|A(\tilde{y})\right|\right|_1.\tag{134}$$

$$\leq\left(\left|\left|\Delta(\tilde{y})\right|\right|_1+\left|\left|A(\tilde{y})\right|\right|_1\right)\sum_{k,l}\left|\mathbb{E}[\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)]\right|\tag{135}$$

$$\leq(2C+1)\sum_{k,l}\left|\mathbb{E}[\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)]\right|.\tag{136}$$

Note that (134) follows from the triangle inequality, (135) from the fact that

$$\left|\mathbb{E}[\mathbf{1}_{\mathcal{Z}}(Z_y^i)-\mathbf{1}_{\mathcal{Z}}(Z_{\tilde{y}}^i)]\right|=\left|\mathbb{E}\left[\sum_{k,l}\left(\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)\right)\right]\right|\leq\sum_{k,l}\left|\mathbb{E}[\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)]\right|,\tag{137}$$

and (136) because $\left|\left|A(\tilde{y})\right|\right|_1\leq 1$ from the definition of A in (93), and $\left|\left|\Delta(\tilde{y})\right|\right|_1\leq 2\left|\left|\tilde{y}\right|\right|_1\leq 2C$ from the definition of Δ in (90) and the definition of $\mathcal{Y}$ in (88).

To upper bound the quantity in (137), we couple Z_y and $Z_{\tilde{y}}$ by maximizing, for each $(k,l)\in\mathcal{Z}$, the probability that both chains move simultaneously to state (k,l):

$$\mathbb{E}[\mathbf{1}_{kl}(Z_y^{i+1})\mathbf{1}_{kl}(Z_{\tilde{y}}^{i+1})\mid Z_y^i,Z_{\tilde{y}}^i]=\min\left(\mathbb{E}[\mathbf{1}_{kl}(Z_y^{i+1})\mid Z_y^i],\mathbb{E}[\mathbf{1}_{kl}(Z_{\tilde{y}}^{i+1})\mid Z_{\tilde{y}}^i]\right).\tag{138}$$

Our argument proceeds as follows. At each step, the fact that $y,\tilde{y}\in\mathcal{Y}$ and the definition of $\mathcal{Y}$ in (88) and the transition probabilities in (115) and (116) ensure that the chance that both chains end in the next step can be lower-bounded by $\mathcal{V}(\rho\ell)$:

$$\mathbb{P}(Z_y^{i+1}=Z_{\tilde{y}}^{i+1}=\emptyset|Z_y^i,Z_{\tilde{y}}^i)\geq\mathcal{V}(\rho\ell).\tag{139}$$

Meanwhile, whenever $Z_y^i=Z_{\tilde{y}}^i$, the probability that the two chains diverge in the next step is upper-bounded as follows:

$$\begin{aligned}\mathbb{P}(Z_y^{i+1}\neq Z_{\tilde{y}}^{i+1}|Z_y^i,Z_{\tilde{y}}^i)&\leq\sum_{(k,l)\in\mathcal{Z}}\left|\mathbb{E}[\mathbf{1}_{kl}(Z_y^{i+1})\mid Z_y^i]-\mathbb{E}[\mathbf{1}_{kl}(Z_{\tilde{y}}^{i+1})\mid Z_{\tilde{y}}^i]\right|\\ &\leq\left|\left|D(y)-D(\tilde{y})\right|\right|_1+\left|a(y)-a(\tilde{y})\right|\\ &\leq 2\left|\left|y-\tilde{y}\right|\right|_1.\end{aligned}\tag{140}$$

The second inequality above follows from the transition probabilities of the chains given in (116) and the assumption $Z_y^i = Z_{\tilde{y}}^i$, while the third follows from the Lipschitz properties of D and a established by Lemma 18.

Combining (139) and (140) implies that the probability that the chains Z_y and $Z_{\tilde{y}}$ ever diverge is at most $2\,||y-\tilde{y}||_1/\mathcal{V}(\rho\ell)$:

$$\mathbb{P}(\sum_{i=0}^{\infty}\sum_{k,l\in\mathcal{Z}}\left|\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)\right|>0)\leq 2\,||y-\tilde{y}||_1/\mathcal{V}(\rho\ell). \tag{141}$$

When the chains do diverge, we can upper-bound the expected total time that they spend in different states by the time until both have reached state $\emptyset$:

$$\mathbb{E}\left[\sum_{i=i'}^{\infty}\sum_{k,l\in\mathcal{Z}}\left|\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)\right|\middle|Z_y^{i'},Z_{\tilde{y}}^{i'}\right]\leq\mathbb{E}\left[\sum_{i=i'}^{\infty}\mathbf{1}_{\mathcal{Z}}(Z_y^i)+\mathbf{1}_{\mathcal{Z}}(Z_{\tilde{y}}^i)\middle|Z_y^{i'},Z_{\tilde{y}}^{i'}\right]\leq 2/\mathcal{V}(\rho\ell), \tag{142}$$

where the final inequality again uses (128) in Lemma 22 and the fact that $y,\tilde{y}\in\mathcal{Y}$. Combining (141) and (142) yields that

$$\mathbb{E}\left[\sum_{i=0}^{\infty}\sum_{k,l\in\mathcal{Z}}\left|\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)\right|\right]\leq 4\,||y-\tilde{y}||_1/\mathcal{V}(\rho\ell)^2. \tag{143}$$

Noting that

$$\sum_{i=0}^{\infty}\sum_{k,l\in\mathcal{Z}}\left|\mathbb{E}[\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)]\right|\leq\mathbb{E}\left[\sum_{i=0}^{\infty}\sum_{k,l\in\mathcal{Z}}\left|\mathbf{1}_{kl}(Z_y^i)-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)\right|\right],$$

equations (136) and (143) imply that the expression in (132) is upper-bounded as follows:

$$\begin{aligned}\sum_{i=0}^{\infty}\sum_{j,k,l}\left|\mathbb{E}\left[\mathbf{1}_{\mathcal{Z}}(Z_y^i)\Delta_{jkl}(\tilde{y})+\mathbf{1}_{kl}(Z_y^i)A_j(\tilde{y})-\mathbf{1}_{\mathcal{Z}}(Z_{\tilde{y}}^i)\Delta_{jkl}(\tilde{y})\right.\right.&\left.\left.-\mathbf{1}_{kl}(Z_{\tilde{y}}^i)A_j(\tilde{y})\right]\right|\\ &\leq(2C+1)\times 4\,||y-\tilde{y}||_1/\mathcal{V}(\rho\ell)^2.\end{aligned} \tag{144}$$

Substituting (133) and (144) into (131) and (132) completes the proof. □

Lemma 25 (Trend Condition)**.** *Let X be the Markov chain arising from Algorithm 2, with $X(t)$ denoting the state after t times through the 'for' loop. Let $\mathcal{Y}$ be defined as in (88), and let Y and f be defined as in (81) and (117), respectively. Then condition (iv') holds with $\delta=5\ell(4C+2)/\mathcal{V}(\rho\ell)^3$.*

Proof. Throughout, we fix t and let Y^i denote the value of Y after i passes through the 'while' loop in Algorithm 2, with $Y^0=Y(t)$ and $Y^\infty=Y(t+1)$. By definition of $\mathcal{Y}$ in (88), if $Y^0/n\in\mathcal{Y}$, then $v(Y^0/n)>\mathcal{V}(\rho\ell)$. Therefore, it suffices to show that

$$\left|\left|\mathbb{E}[Y^\infty-Y^0]-f(Y^0/n)\right|\right|_1\leq\frac{1}{n}\frac{5\ell(4C+2)}{v(Y^0/n)^3}. \tag{145}$$

Note that

$$\mathbb{E}[Y^{\infty}_{jkl} - Y^0_{jkl}] = \mathbb{E}\left[\sum_{i=0}^{\infty} Y^{i+1}_{jkl} - Y^i_{jkl}\right]$$
$$= \mathbb{E}\left[\sum_{i=0}^{\infty} \mathbb{E}[Y^{i+1}_{jkl} - Y^i_{jkl} \mid X^i]\right]$$
$$= \mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i)\Delta_{jkl}(\tilde{Y}^i/\tilde{n}^i) + \mathbf{1}_{kl}(\tilde{Z}^i)A_j(\tilde{Y}^i/\tilde{n}^i)\right]. \tag{146}$$

In what follows we use $\tilde{\Delta}^i, \Delta^i, \tilde{A}^i$ and A^i as shorthand for $\Delta(\tilde{Y}^i/\tilde{n}^i), \Delta(Y^i/n), A(\tilde{Y}^i/\tilde{n}^i)$ and $A(Y^i/n)$, respectively. Applying (146) and Lemma 21, we have

$$\mathbb{E}[Y^{\infty}_{jkl} - Y^0_{jkl}] - f_{jkl}(Y^0/n) = \mathbb{E}\left[\sum_{i=0}^{\infty}\left(\mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i)\tilde{\Delta}^i_{jkl} + \mathbf{1}_{kl}(\tilde{Z}^i)\tilde{A}^i_j\right) - \sum_{i=0}^{\infty}\left(\mathbf{1}_{\mathcal{Z}}(Z^i)\Delta^0_{jkl} + \mathbf{1}_{kl}(Z^i)A^0_j\right)\right]$$
$$= \mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i)(\tilde{\Delta}^i_{jkl} - \Delta^i_{jkl}) + \mathbf{1}_{kl}(\tilde{Z}^i)(\tilde{A}^i_j - A^i_j)\right] \tag{147}$$
$$+ \mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i)(\Delta^i_{jkl} - \Delta^0_{jkl}) + \mathbf{1}_{kl}(\tilde{Z}^i)(A^i_j - A^0_j)\right] \tag{148}$$
$$+ \mathbb{E}\left[\sum_{i=0}^{\infty}(\mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i) - \mathbf{1}_{\mathcal{Z}}(Z^i))\Delta^0_{jkl} + (\mathbf{1}_{kl}(\tilde{Z}^i) - \mathbf{1}_{kl}(Z^i))A^0_j\right]. \tag{149}$$

In what follows, let $\tilde{T} = \sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i)$.

By Lemma 19 equations (104) and (105) we have $\left|\left|\tilde{\Delta}^i - \Delta^i\right|\right|_1 \leq 4C(\ell-1)/n$ and $\left|\left|\tilde{A}^i - A^i\right|\right|_1 \leq 2(\ell-1)/n$, so

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i)\left|\left|\tilde{\Delta}^i - \Delta^i\right|\right|_1 + \mathbf{1}_{kl}(\tilde{Z}_i)\left|\left|\tilde{A}^i - A^i\right|\right|_1\right] \leq \frac{1}{n}(4C+2)(\ell-1)\,\mathbb{E}[\tilde{T}]$$
$$\leq \frac{1}{n}\frac{(4C+2)(\ell-1)}{v(Y^0/n)}$$
$$\leq \frac{1}{n}\frac{(4C+2)(\ell-1)}{v(Y^0/n)^3}, \tag{150}$$

where the second inequality follows from (125) in Lemma 22.

From Lemma 20 equations (112) and (113) we have $||\Delta^i - \Delta^0||_1 \leq 4Ci/n$ and $||A^i - A^0||_1 \leq 2i/n$, and therefore

$$\begin{aligned}\mathbb{E}\left[\sum_{i=0}^{\infty} \mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i)\,||\Delta^i - \Delta^0||_1 + \mathbf{1}_{kl}(\tilde{Z}^i)\,||A^i - A^0||_1\right] &\leq \mathbb{E}\left[\sum_{i=0}^{\tilde{T}-1}(2C+1)(2i/n)\right] \\ &\leq \frac{1}{n}(2C+1)\,\mathbb{E}[\tilde{T}^2] \\ &\leq \frac{1}{n}\frac{4C+2}{v(Y^0/n)^2} \\ &\leq \frac{1}{n}\frac{4C+2}{v(Y^0/n)^3}, \qquad (151)\end{aligned}$$

where the penultimate inequality follows from (126) in Lemma 22.

Turning to (149),

$$\begin{aligned}&\left|\mathbb{E}\left[\sum_{i=0}^{\infty}(\mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i) - \mathbf{1}_{\mathcal{Z}}(Z^i))\,||\Delta^0||_1 + (\mathbf{1}_{kl}(\tilde{Z}^i) - \mathbf{1}_{kl}(Z^i))\,||A^0||_1\right]\right| \\ &\qquad\leq (2C+1)\max\left(\left|\mathbb{E}\left[\sum_{i=0}^{\infty}\mathbf{1}_{\mathcal{Z}}(\tilde{Z}^i) - \mathbf{1}_{\mathcal{Z}}(Z^i)\right]\right|, \left|\mathbb{E}\left[\sum_{i=0}^{\infty}\mathbf{1}_{kl}(\tilde{Z}^i) - \mathbf{1}_{kl}(Z^i)\right]\right|\right) \\ &\qquad\leq \frac{1}{n}\frac{4\ell(4C+2)}{v(Y^0/n)^3}, \qquad (152)\end{aligned}$$

where the first inequality follows from the facts that $||\Delta^0||_1 = ||\Delta(Y^0/n)||_1 \leq 2\,||Y^0/n||_1 \leq 2C$ (the first inequality follows from (99)) and $||A^0||_1 \leq 1$, and the second from Lemma 26 below. Substituting the bounds from (150), (151), and (152) into (147), (148), and (149) establishes (145), completing the proof. □

Lemma 26 (Coupling Z and $\tilde{Z}$.)**.** *For any $S \subseteq \mathcal{Z}$,*

$$\left|\mathbb{E}\left[\sum_{i=0}^{\infty}\mathbf{1}_S(\tilde{Z}^i) - \mathbf{1}_S(Z^i_{Y^0/n})\right]\right| \leq \frac{1}{n}\cdot\frac{8\ell}{v(Y^0/n)^3}.$$

Proof of Lemma 26. Throughout, we write Z in place of $Z_{Y^0/n}$. The idea of this proof is the same as that of Lemma 24: we will couple $\tilde{Z}$ and Z to maximize the chance that they end up in the same state in step: for each $(k,l) \in \mathcal{Z}$ we have

$$\mathbb{E}[\mathbf{1}_{kl}(\tilde{Z}^{i+1})\mathbf{1}_{kl}(Z^{i+1}) \mid X^i, Z^i] = \min\left(\mathbb{E}[\mathbf{1}_{kl}(\tilde{Z}^{i+1}) \mid X^i], \mathbb{E}[\mathbf{1}_{kl}(Z^{i+1}) \mid Z^i]\right). \qquad (153)$$

Let T^* be the first time that the chains diverge. That is,

$$T^* = \min\{i : \tilde{Z}^i \neq Z^i\}, \qquad (154)$$

with $T^* = \emptyset$ if $\tilde{Z}^i = Z^i$ for all i. We will show that $T^* = \emptyset$ with probability $1 - \mathcal{O}(1/n)$. Intuitively, this is because the expected number of proposals is at most $1/v(Y^0/n)$ by Lemma 22, and at each proposal the probability of the two chains diverging is $\mathcal{O}(1/n)$. Furthermore, when the chains do diverge, this divergence lasts for only $\mathcal{O}(1)$ steps.

Being more precise, note that

$$\sum_{i=0}^{\infty} \mathbf{1}_S(\tilde{Z}^i) - \mathbf{1}_S(\tilde{Z}^i) = \sum_{i=1}^{\infty} \mathbf{1}(T^* = i) \left(\sum_{i'=i}^{\infty} \mathbf{1}_S(\tilde{Z}^{i'}) - \sum_{i'=i}^{\infty} \mathbf{1}_S(Z^{i'}) \right),$$

and therefore

$$\mathbb{E}\left[\left|\sum_{i=1}^{\infty} \mathbf{1}_S(\tilde{Z}^i) - \mathbf{1}_S(Z^i)\right|\right] \le \sum_{i=1}^{\infty} \mathbb{P}(T^* = i)\, \mathbb{E}\left[\left|\sum_{i'=i}^{\infty} \mathbf{1}_S(\tilde{Z}^{i'}) - \sum_{i'=i}^{\infty} \mathbf{1}_S(Z^{i'})\right| \mid T^* = i\right]. \tag{155}$$

$$\le \sum_{i=1}^{\infty} \mathbb{P}(T^* = i)\, \mathbb{E}\left[\sum_{i'=i}^{\infty} \mathbf{1}_S(\tilde{Z}^{i'}) + \sum_{i'=i}^{\infty} \mathbf{1}_S(Z^{i'}) \mid T^* = i\right] \tag{156}$$

$$\le \frac{2}{v(Y^0/n)} \sum_{i=1}^{\infty} \mathbb{P}(T^* = i), \tag{157}$$

with the final inequality following from (125) and (128) of Lemma 22.

All that remains is to bound the probability that the chains diverge, given by $\sum_{i=1}^{\infty} \mathbb{P}(T^* = i)$. We start by noting that if $\tilde{Z}^i = Z^i = Z \in \mathcal{Z}$, then

$$\begin{aligned}
\mathbb{P}(\tilde{Z}^{i+1} \neq Z^{i+1} \mid X^i, Z^i) &\le \sum_{(k,l)\in\mathcal{Z}} \left| \mathbb{E}[\mathbf{1}_{kl}(\tilde{Z}^{i+1}) \mid X^i] - \mathbb{E}[\mathbf{1}_{kl}(Z^{i+1}) \mid Z^i] \right| \\
&\le \sum_{(k,l)\in\mathcal{Z}} \left| D(\tilde{Y}^i/\tilde{n}^i) - D_{(k-1)l}(Y^0/n) + \mathbf{1}_{(k-1)l}(Z)(a(Y^0/n) - a(\tilde{Y}^i/\tilde{n}^i)) \right| \\
&\le \left|\left| D(\tilde{Y}^i/\tilde{n}^i) - D(Y^0/n) \right|\right|_1 + \left|\left| P(Y^0/n) - P(\tilde{Y}^i/\tilde{n}^i) \right|\right|_1 \\
&\le \frac{4(i+\ell-1)}{n}.
\end{aligned} \tag{158}$$

The second line follows from (97) from Lemma 17 and (116), the third uses (100) from Lemma 18, and the final line follows from the triangle inequality and Lemmas 19 and 20.

We use this to conclude that

$$\begin{aligned}
\mathbb{P}(T^* = i+1) &\le \mathbb{P}(\tilde{Z}^i = Z^i \neq \emptyset)\, \mathbb{P}(\tilde{Z}^{i+1} \neq \hat{Z}^{i+1} | \tilde{Z}^i = Z^i \neq \emptyset) \\
&\le (1 - v(Y^0/n))^i \cdot \frac{4(i+\ell-1)}{n},
\end{aligned} \tag{159}$$

where the second line follows from (124) in Lemma 22 and (158). By (159), we have

$$\begin{aligned}
\sum_{i\ge1} \mathbb{P}(T^* = i) &\le \sum_{i\ge1} (1 - v(Y^0/n))^{i-1} \frac{4(i-1+\ell-1)}{n} \\
&= \frac{4}{n} \left(\frac{\ell-1}{v(Y^0/n)} + \frac{1 - v(Y^0/n)}{v(Y^0/n)^2} \right) \\
&\le \frac{4}{n} \left(\frac{\ell-1}{v(Y^0/n)} + \frac{1}{v(Y^0/n)^2} \right). \\
&\le \frac{4}{n} \frac{\ell}{v(Y^0/n)^2}.
\end{aligned} \tag{160}$$

Combining (157) and (160) completes the proof. $\square$

C.5. **Solving The Differential Equation.**

Section C.4 shows that when X is the Markov chain arising from Algorithm 2 and Y and f are defined as in (81) and (117), the bounded, lipchitz, and trend conditions *(ii')*, *(iii')*, and *(iv')* are satisfied. It follows that if $\hat{y}$ is the solution to

$$\hat{y}(0) = y_0, \qquad \hat{y}'_{jkl}(s) = f_{jkl}(\hat{y}(s)). \tag{161}$$

then for any $\epsilon > 0$, as $n \to \infty$,

$$\mathbb{P}\left(\left|\left|Y^n_{jkl}(\rho n)/n - \hat{y}(\rho)\right|\right|_1 > \epsilon\right) \to 0.$$

By definition of G^n and the definition of Y^n_{jkl} in (81), we have

$$G^n(k,l) = \sum_{k'=1}^{k} \sum_{j\geq 1} Y^n_{jk'l}(\rho n).$$

Therefore, to prove Theorem 1 in the case of independent lotteries, we must show that

$$\sum_{k'=1}^{k} \sum_{j\geq 1} \hat{y}_{jk'l}(\rho) = \rho \mathcal{L}(l) F_I(k). \tag{162}$$

This is established by the following Lemma.

Lemma 27 (Solving the Differential Equation)**.** *Fix $C, \ell \in \mathbb{N}$ and $\mathcal{L}$ such that $\mathcal{L}(>\ell) = 0$. For any $\rho \geq 0$, define $\lambda(\rho)$ to be the unique solution λ to*

$$\lambda = \rho \cdot \mu(\mathcal{A}(\lambda)). \tag{163}$$

If f is given by (117) and $\hat{y}$ is defined as the solution to (161) with initial condition $y^0_{jkl} = 0$ for $j \geq 1$ and $1 \leq k \leq l \leq \ell$, then

$$\hat{y}_{jkl}(\rho) = \rho \mathcal{L}(l)(1 - \mathcal{A}(\lambda(\rho)))^{k-1} p_{j-1}(\lambda(\rho)) \min(1, C/j), \tag{164}$$

Furthermore, for any $1 \leq k \leq l \leq \ell$, (162) holds.

Proof. To keep notation uncluttered, throughout, we omit the dependence of $\hat{y}$ and λ on ρ.

We first prove that if $\hat{y}$ is defined by (164), then (162) holds. By definition of p_j in (15) and $\mathcal{A}$ in (2),

$$\sum_{j\geq 1} p_{j-1}(\lambda) \min(C/j, 1) = \mathcal{A}(\lambda),$$

and therefore

$$\sum_{j\geq 1} \hat{y}_{jk'l} = \rho \mathcal{L}(l)(1 - \mathcal{A}(\lambda))^{k'-1} \mathcal{A}(\lambda).$$

Summing this for k' from 1 to k immediately establishes (162).

We now show that if $\hat{y}$ is defined by (164), then it does indeed satisfy (161). Clearly, when $\rho = 0$, (164) implies that $\hat{y}_{jkl}(0) = 0$, so the initial condition of (161) holds.

We now show the differential component of (161). Differentiating (15) reveals that for $j \geq 1$,

$$p'_{j-1}(\lambda) = p_{j-2}(\lambda) - p_{j-1}(\lambda).$$

(When $j = 1$, we take $p_{-1}(\lambda) = 0$ by convention.) Therefore, differentiating (164) yields

$$\hat{y}'_{jkl} = \rho\mathcal{L}(l)\left(p_{j-2}(\lambda) - p_{j-1}(\lambda)\right)\min(C/j,1)(1-\mathcal{A}(\lambda))^{k-1}\lambda' \tag{165}$$

$$- \rho\mathcal{L}(l)p_{j-1}(\lambda)\min(C/j,1)(k-1)(1-\mathcal{A}(\lambda))^{k-2}\mathcal{A}'(\lambda)\lambda' \tag{166}$$

$$+ \mathcal{L}(l)p_{j-1}(\lambda)\min(C/j,1)(1-\mathcal{A}(\lambda))^{k-1} \tag{167}$$

We first consider the term in (165), and note that

$$\begin{aligned}\rho\mathcal{L}(l)\left(p_{j-2}(\lambda) - p_{j-1}(\lambda)\right)\min(C/j,1)(1-\mathcal{A}(\lambda))^{k-1} &= \frac{\min(C/j,1)}{\min(C/(j-1),1)}\hat{y}_{(j-1)kl} - \hat{y}_{jkl} \\ &= \left(1 - \frac{\mathbf{1}(j>C)}{j}\right)\hat{y}_{(j-1)kl} - \hat{y}_{jkl} \\ &= \Delta_{jkl}(\hat{y}),\end{aligned} \tag{168}$$

where the second line follows from considering each of the cases $j \le C$ and $j > C$.

Next, we claim that

$$p_{j-1}(\lambda)\min(C/j,1) = A_j(\hat{y}) \tag{169}$$

$$\mathcal{A}(\lambda) = a(\hat{y}). \tag{170}$$

$$-\rho\mathcal{L}(l)(k-1)(1-\mathcal{A}(\lambda))^{k-2}\mathcal{A}'(\lambda) = \sum_{k'=1}^{k-1} D_{k'\ell}(\hat{y})(1-a(\hat{y}))^{k-k'-1}. \tag{171}$$

$$\lambda'(\rho) = Q(\hat{y}). \tag{172}$$

Substituting these expressions into (166) and (167) yields

$$-\rho\mathcal{L}(l)p_{j-1}(\lambda)\min(C/j,1)(k-1)(1-\mathcal{A}(\lambda))^{k-2}\mathcal{A}'(\lambda)\lambda'(\rho) = A_j(\hat{y})Q(\hat{y})\sum_{k'=1}^{k-1} D_{k'\ell}(\hat{y})(1-a(\hat{y}))^{k-k'-1} \tag{173}$$

$$\mathcal{L}(l)p_{j-1}(\lambda)\min(C/j,1)(1-\mathcal{A}(\lambda))^{k-1} = \mathcal{L}(l)A_j(\hat{y})(1-a(\hat{y}))^{k-1} \tag{174}$$

Substituting (168), (173), (174) into the expression for $\hat{y}'_{jkl}$ in (165), (166) and (167) reveals that

$$\hat{y}'_{jkl} = Q(\hat{y})\Delta_{jkl}(\hat{y}) + q_{kl}(\hat{y})A_j(\hat{y}) = f_{jkl}(\hat{y}).$$

All that remains is to prove (169), (170), (171) and (172). For $j \ge 1$, we claim that

$$\begin{aligned}P_j(\hat{y}) &= \frac{1}{\min(C,j)}\sum_{1\le k\le l\le \ell}\hat{y}_{jkl} \\ &= \rho\mu(\mathcal{A}(\lambda))p_{j-1}(\lambda)/j \\ &= \rho\mu(\mathcal{A}(\lambda))p_j(\lambda)/\lambda \\ &= p_j(\lambda).\end{aligned} \tag{175}$$

The first equality restates (91), the second follows from (164) and (120), the third from (15) (which implies $p_{j-1}(\lambda)/j = p_j(\lambda)/\lambda$), and the fourth from (163). It follows that (175) also holds for $j = 0$. Note that (93) and (175) jointly imply (169).

Turning to (170) yields

$$\begin{aligned} a(\hat{y}) &= \sum_{0\le j<C} P_j(\hat{y}) + \sum_{j\ge C} P_j(\hat{y})\frac{C}{j+1}. \\ &= \sum_{j\ge 0} P_j(\hat{y})\frac{\min(C,j+1)}{j+1} \\ &= \sum_{j\ge 0} p_j(\lambda)\frac{\min(C,j+1)}{j+1} \\ &= \mathcal{A}(\lambda). \end{aligned} \tag{176}$$

The first line follows from combining (91), (95) and (94), the second is elementary, the third follows from (175), the fourth from (15), and the final line from (2).

Turning to (171), we have

$$\begin{aligned} \sum_{k'=1}^{k-1} D_{k'l}(\hat{y})(1-a(\hat{y}))^{k'-1} &= \rho\mathcal{L}(l)(k-1)(1-\mathcal{A}(\lambda))^{k-2}\sum_{j\ge C} p_{j-1}(\lambda)\frac{C}{j(j+1)} \\ &= \rho\mathcal{L}(l)(k-1)(1-\mathcal{A}(\lambda))^{k-2}\sum_{j\ge C}\frac{p_j(\lambda)}{\lambda}\frac{C}{j+1} \\ &= \rho\mathcal{L}(l)(k-1)(1-\mathcal{A}(\lambda))^{k-2}\frac{\mathcal{A}(\lambda)-\mathcal{V}(\lambda)}{\lambda} \\ &= -\rho\mathcal{L}(l)(k-1)(1-\mathcal{A}(\lambda))^{k-2}\mathcal{A}'(\lambda). \end{aligned}$$

The first line uses the definition of $\hat{y}$ in (164) as well as (176), the second uses (15), the third uses (1) and (2), and the final uses Lemma 3.

Finally, we turn to (172). By (119), (176) and (171) we have

$$\begin{aligned} Q(\hat{y}) &= \frac{\mu(\mathcal{A}(\lambda))}{1+\rho\mathcal{A}'(\lambda)\sum_{1\le k\le l\le \ell}\mathcal{L}(l)(k-1)(1-\mathcal{A}(\lambda))^{k-1}} \\ &= \frac{\mu(\mathcal{A}(\lambda))}{1-\rho\mathcal{A}'(\lambda)\mu'(\mathcal{A}(\lambda))} \\ &= \lambda'(\rho). \end{aligned} \tag{177}$$

The second line uses the fact that for $q \in [0,1]$,

$$\begin{aligned} \mu'(q) &= \frac{d}{dq}\sum_{k=1}^{\ell}\mathcal{L}(\ge k)(1-q)^{k-1} \\ &= -\sum_{k=1}^{\ell}\mathcal{L}(\ge k)(k-1)(1-q)^{k-2} \\ &= -\sum_{1\le k\le l\le \ell}\mathcal{L}(l)(k-1)(1-q)^{k-2}, \end{aligned}$$

while (177) follows from differentiating (163) and solving the resulting equation for λ':

$$\lambda' = \mu(\mathcal{A}(\lambda)) + \rho\mu'(\mathcal{A}(\lambda))\mathcal{A}'(\lambda)\lambda'.$$

□