Selection Problems in the Presence of Implicit Bias

Jon Kleinberg
Cornell University

Manish Raghavan
Cornell University

Abstract

Over the past two decades, the notion of implicit bias has come to serve as an important component in our understanding of discrimination in activities such as hiring, promotion, and school admissions. Research on implicit bias posits that when people evaluate others – for example, in a hiring context – their unconscious biases about membership in particular groups can have an effect on their decision-making, even when they have no deliberate intention to discriminate against members of these groups. A growing body of experimental work has pointed to the effect that implicit bias can have in producing adverse outcomes.

Here we propose a theoretical model for studying the effects of implicit bias on selection decisions, and a way of analyzing possible procedural remedies for implicit bias within this model. A canonical situation represented by our model is a hiring setting: a recruiting committee is trying to choose a set of finalists to interview among the applicants for a job, evaluating these applicants based on their future potential, but their estimates of potential are skewed by implicit bias against members of one group. In this model, we show that measures such as the Rooney Rule, a requirement that at least one of the finalists be chosen from the affected group, can not only improve the representation of this affected group, but also lead to higher payoffs in absolute terms for the organization performing the recruiting. However, identifying the conditions under which such measures can lead to improved payoffs involves subtle trade-offs between the extent of the bias and the underlying distribution of applicant characteristics, leading to novel theoretical questions about order statistics in the presence of probabilistic side information.

1 Introduction

Over the past two decades, the notion of *implicit bias* [13] has come to provide on important perspective on the nature of discrimination. Research on implicit bias argues that unconscious attitudes toward members of different demographic groups — for example, defined by gender, race, ethnicity, national origin, sexual orientation, and other characteristics — can have a non-trivial impact on the way in which we evaluate members of these groups; and this in turn may affect outcomes in employment [2, 3, 21], education [22], law [14, 15], medicine [12], and other societal institutions.

In the context of a process like hiring, implicit bias thus shifts the question of bias and discrimination to be not just about identifying bad actors who are intentionally discriminating, but also about the tendency of all of us to reach discriminatory conclusions based on the unconscious application of stereotypes. An understanding of these issues also helps inform the design of interventions to mitigate implicit bias — when essentially all of us have a latent tendency toward low-level discrimination, a set of broader practices may be needed to guide the process toward the desired outcome.

A basic mechanism: The Rooney Rule. One of the most basic and widely adopted mechanisms in practice for addressing implicit bias in hiring and selection is the *Rooney Rule* [7], which, roughly speaking, requires that in recruiting for a job opening, one of the candidates interviewed must come from an underrepresented group. The Rooney Rule is named for a protocol adopted by the National

Football League (NFL) in 2002 in response to widespread concern over the low representation of African-Americans in head coaching positions; it required that when a team is searching for a new head coach, at least one minority candidate must be interviewed for the position. Subsequently the Rooney Rule has become a guideline adopted in many areas of business [5]; for example, in 2015 then-President Obama exhorted leading tech firms to use the Rooney Rule for hiring executives, and in recent years companies including Amazon, Facebook, Microsoft, and Pinterest have adopted a version of the Rooney Rule requiring that at least one candidate interviewed must be a woman or a member of an underrepresented minority group [18]. In 2017, a much-awaited set of recommendations made by Eric Holder and colleagues to address workplace bias at Uber advocated for the use of the Rooney Rule as one of its key points [8, 19].

The Rooney Rule is the subject of ongoing debate, and one crucial aspect of this debate is the following tension. On one side is the argument that implicit (or explicit) bias is preventing deserving candidates from underrepresented groups from being fairly considered, and the Rooney Rule is providing a force that counter-balances and partially offsets the consequences of this underlying bias. On the other side is the concern that if a job search process produces a short-list of top candidates all from the majority group, it may be because these are genuinely the strongest candidates despite the underlying bias — particularly if there is a shortage of available candidates from other groups. In this case, wholesale use of the Rooney Rule may lead firms to consider weaker candidates from underrepresented groups, which works against the elimination of unconscious stereotypes. Of course, there are other reasons to seek diversity in recruiting that may involve broader considerations or longer time horizons than just the specific applicants being evaluated; but even these lines of argument generally incorporate the more local question of the effect on the set of applicants.

Given the widespread consideration of the Rooney Rule from both legal and empirical perspectives [7], it is striking that prior work has not attempted to formalize the inherently mathematical question that forms a crucial ingredient in these debates: given some estimates of the extent of bias and the prevalence of available minority candidates, does the expected quality of the candidates being interviewed by a hiring committee go up or down when the Rooney Rule is implemented? When the bias is large and there are many minority candidates, it is quite possible that a hiring committee's bias has caused it to choose a weaker candidate over a stronger minority one, and the Rooney Rule may be strengthening the pool of interviewees by reversing this decision and swapping the stronger minority candidate in. But when the bias is small or there are few minority candidates, the Rule might be reversing a decision that in fact chose the stronger applicant.

In this paper, we propose a formalization of this family of questions, via a simplified model of selection with implicit bias, and we give a tight analysis of the consequences of using the Rooney Rule in this setting. In particular, when selecting for a fixed number of slots, we identify a sharp threshold on the effectiveness of the Rooney Rule in our model that depends on three parameters: not just the extent of bias and the the prevalence of available minority candidates, but a third quantity as well — essentially, a parameter governing the distribution of candidates' expected future job performance. We emphasize that our model is deliberately stylized, to abstract the trade-offs as cleanly as possible. Moreover, in interpreting these results, we emphasize a point noted above, that there are other reasons to consider using the Rooney Rule beyond the issues that motivate this particular formulation; but an understanding of the trade-offs in our model seems informative in any broader debate about such hiring and selection measures.

We now describe the basic ingredients of our model, followed by a summary of the main results.

1.1 A Model of Selection with Implicit Bias

Our model is based on the following scenario. Suppose that a hiring committee is trying to fill an open job position, and it would like to choose the $k \geq 2$ best candidates as *finalists* to interview from among a large set of applicants. We will think of k as a small constant, and indeed most of the subtlety of the question already arises for the case $k = 2$, when just two finalists must be selected.

X-candidates and Y-candidates. The set of all applicants is partitioned into two groups X and Y, where we think of Y as the majority group, and X as a minority group within the domain that may be subject to bias. For some positive real number $\alpha \leq 1$ and a natural number n, there are n applicants from group Y and αn applicants from group X. If a candidate i belongs to X, we will refer to them as an *X-candidate*, and if i belongs to Y, we will refer to them as a *Y-candidate*. (The reader is welcome, for example, to think of the setting of academic hiring, with X as candidates from a group that is underrepresented in the field, but the formulation is general.)

Each candidate i has a (hidden) numerical value that we call their *potential*, representing their future performance over the course of their career. For example, in faculty hiring, we might think of the potential of each applicant in terms of some numerical proxy like their future lifetime citation count (with the caveat that any numerical measure will of course be an imperfect representation). Or in hiring executives, the potential of each applicant could be some measure of the revenue they will bring to the firm.

We assume that there is a common distribution Z that these numerical potentials come from: each potential is an independent draw from Z. (Thus, the applicants can have widely differing values for their numerical potentials; they just arise as draws from a common distribution.) For notational purposes, when i is an X-candidate, we write their potential as X_i, and when j is a Y-candidate, we write their potential as Y_j. We note an important modeling decision in this formulation: we are assuming that all X_i and all Y_j values come from this same distribution Z. While it is also of interest to consider the case in which the numerical potentials of the two groups X and Y are drawn from different group-specific distributions, we focus on the case of identical distributions for two reasons. First, there are many settings where differences between the underlying distributions for different groups appear to be small compared to the bias-related effects we are seeking to measure; and second, in any formal analysis of bias between groups, the setting in which the groups begin with identical distributions is arguably the first fundamental special case that needs to be understood.

In the domains that we are considering — hiring executives, faculty members, athletes, performers — there is a natural functional form for the distribution Z of potentials, and this is the family of *power laws* (also known as *Pareto distributions*), with $\Pr[Z \geq t] = t^{-(1+\delta)}$ and support $[1, \infty)$ for a fixed $\delta > 0$. Extensive empirical work has argued that the distribution of individual output in a wide range of creative professions can be approximated by power law distributions with small positive values of δ [6]. For example, the distribution of lifetime citation counts is well-approximated by a power law, as are the lifetime downloads, views, or sales by performers, authors, and other artists. In the last part of the paper, we also consider the case in which the potentials are drawn from a distribution with bounded support, but for most of the paper we will focus on power laws.

Selection with Bias. Given the set of applicants, the hiring committee would like to choose k *finalists* to interview. The *utility* achieved by the committee is the sum of the potentials of the k finalists it chooses; the committee's goal is to maximize its utility.[1]

[1] Since our goal is to model processes like the Rooney Rule, which apply to the selection of finalists for interviewing, rather than to the hiring decision itself, we treat the choice of k finalists as the endpoint rather than modeling the interviews that subsequently ensue.

If the committee could exactly evaluate the potential of each applicant, then it would have a straightforward way to maximize the utility of the set of finalists: simply sort all applicants by potential, and choose the top k as finalists. The key feature of the situation we would like to capture, however, is that the committee is biased in its evaluations; we look for a model that incorporates this bias as cleanly as possible.

Empirical work in some of our core motivating settings — such as the evaluation of scientists and faculty candidates — indicates that evaluation committees often systematically downweight female and minority candidates of a given level of achievement, both in head-to-head comparisons and in ranking using numerical scores [23]. It is thus natural to model the hiring committee's evaluations as follows: they correctly estimate the potential of a Y-applicant j at the true value Y_j, but they estimate the potential of an X-applicant i at a reduced value $\tilde{X}_i < X_i$. They then rank candidates by these values $\{Y_j\}$ and $\{\tilde{X}_i\}$, and they choose the top k according to this biased ranking.

For most of the paper, we focus on the case of *multiplicative bias*, in which $\tilde{X}_i = X_i/\beta$ for a bias parameter[2] $\beta > 1$. This is a reasonable approximation to empirical data from human-subject studies [23]; and moreover, for power law distributions this multiplicative form is in a strong sense the "right" parametrization of the bias, since biases that grow either faster or slower than multiplicatively have a very simple asymptotic behavior in the power law case.

In this aspect of the model, as in others, we seek the cleanest formulation that exposes the key underlying issues; for example, it would be an interesting extension to consider versions in which the estimates for each individual are perturbed by random noise. A line of previous work [4, 10, 11] has analyzed models of ranking under noisy perturbations; while our scenario is quite different in that the entities being ranked are partitioned into a fixed set of groups with potentially different levels of bias and noise, it would be natural to see if these techniques could potentially be extended to handle noise in the context of implicit bias.

1.2 Main Questions and Results

This then is the basic model in which we analyze interventions with the structure of the Rooney Rule: (i) a set of n Y-applicants and αn X-applicants each have an independent future potential drawn from a power law distribution; (ii) a hiring committee ranks the applicants according to a sorted order in which each X-applicant's potential is divided down by $\beta > 1$, and chooses the top k in this ordering as *finalists*; and (iii) the hiring committee's *utility* is the sum of the potentials of the k finalists.

Qualitatively, the motivation for the Rooney Rule in such settings is that hiring committees are either unwilling or unable to reasonably correct for their bias in performing such rankings, and therefore cannot be relied on to interview X-candidates on their own. The difficulty in removing this skew from such evaluations is a signature aspect of phenomena around implicit bias.

The decision to impose the Rooney Rule is made at the outset, before the actual values of the potentials $\{Y_j\}$ and $\{\tilde{X}_i\}$ are materialized. All that is known at the point of this initial decision to use the Rule or not are the parameters of the domain: the bias β, the relative abundance of X-candidates α, the power law exponent $1 + \delta$, and the number of finalists to be chosen k. The question is: as a function of these parameters, will the use of the Rooney Rule produce a positive or negative expected change in utility, where the expectation is taken over the random draws of applicant values? We note that one could instead ask about the probability that the Rooney Rule produces a positive change in utility as opposed to the expected change; in fact, our techniques naturally extend to characterize not only the expected change, but the probability that this change is positive, as we will show in Section 2.

Our model lets us make precise the trade-off in utility that underpins the use of the Rooney Rule. If the committee selects an X-candidate on its own — even using its biased ranking — then their choice

[2] When $\beta = 1$, the ranking has no bias.

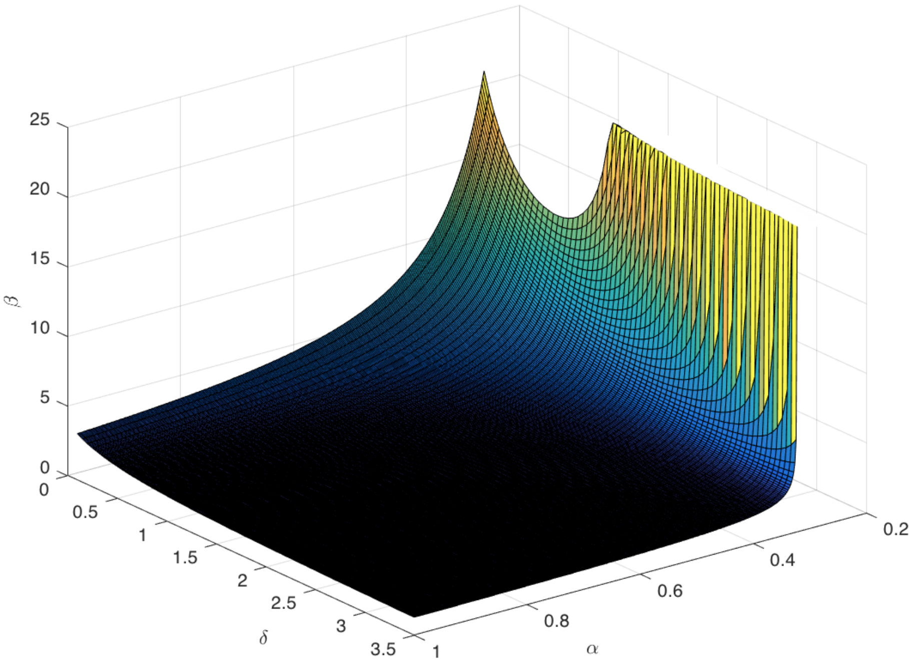

Figure 1: Fixing $k = 2$, the (α, β, δ) values for which the Rooney Rule produces a positive expected change for sufficiently large n lie above a surface (depicted in the figure) defined by the function $\phi_2(\alpha, \beta, \delta) = 1$.

already satisfies the conditions of the Rule. But if all k finalists are Y-candidates, then the Rooney Rule requires that the committee replace the lowest-ranked of these finalists j with the highest-ranked X-candidate i. Because i was not already a finalist, we know that $\tilde{X}_i = X_i/\beta < Y_j$. But to see whether this yields a positive change in utility, we need to understand which of X_i or Y_j has a larger expected value, conditional on the information contained in the committee's decision, that $X_i/\beta < Y_j$.

Our main result is an exact characterization of when the Rooney Rule produces a positive expected change in terms of the four underlying parameters, showing that it non-trivially depends on all four. For the following theorem, and for the remainder of the paper, we assume $0 < \alpha \leq 1$, $\beta > 1$, and $\delta > 0$. We begin with the case where $k = 2$.

Theorem 1.1. *For $k = 2$ and sufficiently large n, the Rooney Rule produces a positive expected change if and only if $\phi_2(\alpha, \beta, \delta) > 1$ where*

$$\phi_2(\alpha, \beta, \delta) = \frac{\alpha^{1/(1+\delta)} \left[1 - (1 + c^{-1})^{-\delta/(1+\delta)} \left[1 + \frac{\delta}{1+\delta}(1 + c)^{-1}\right]\right]}{\frac{\delta}{1+\delta}(1 + c)^{-1-\delta/(1+\delta)}} \tag{1}$$

and $c = \alpha\beta^{-(1+\delta)}$. Moreover, $\phi_2(\alpha, \beta, \delta)$ is increasing in β, so for fixed α and δ there exists β^* such that $\phi_2(\alpha, \beta, \delta) > 1$ if and only if $\beta > \beta^*$.

Thus, we have an explicit characterization for when the Rooney Rule produces positive expected change. The following theorem extends this to larger values of k.

Theorem 1.2. *There is an explicit function $\phi_k(\alpha, \beta, \delta)$ such that the Rooney Rule produces a positive expected change, for n sufficiently large and $k = O(\ln n)$, if and only if $\phi_k(\alpha, \beta, \delta) > 1$.*

Interestingly, even for larger values of k, there are parts of the parameter space for which the Rooney Rule produces a positive expected change and parts for which the Rooney Rule produces a negative expected change, independent of the number of applicants n.

Figure 1 depicts a view of the function ϕ_2, by showing the points in three-dimensional (α, β, δ) space for which ϕ takes the value 1. The values for which the Rooney Rule produces a positive expected change for sufficiently large n lie above this surface.

The surface in Figure 1 is fairly complex, and it displays unexpected non-monotonic behavior. For example, on certain regions of fixed (α, β), it is non-monotonic in δ, a fact which is not a priori obvious: there are choices of α and β for which the Rooney Rule produces a positive expected change at certain "intermediate" values of δ, but not at values of δ that are sufficiently smaller or sufficiently larger. Moreover, there exist (α, δ) pairs above which the surface does not exist. (One example in Figure 1 occurs at $\alpha \approx 0.3$ and $\delta \approx 3$). Characterizing the function ϕ and its level set $\phi = 1$ is challenging, and it is noteworthy that the complexity of this function is arising from our relatively bare-bones formulation of the trade-off in the Rooney Rule; this suggests the function and its properties are capturing something inherent in the process of biased selection.

One monotonocity result we are able to establish for the function ϕ is the following, showing that for fixed (α, β, δ), increasing the number of positions can't make the Rooney Rule go from beneficial to harmful.

Theorem 1.3. *For sufficiently large n and $k = O(\ln n)$, if the Rooney Rule produces a positive expected change at a given number of finalists k, it also produces a positive expected change when there are $k+1$ finalists (at the same (α, β, δ)).*

We prove these theorems through an analysis of the *order statistics* of the underlying power law distribution. Specifically, if we draw m samples from the power law Z and sort them in ascending order from lowest to highest, then the ℓ^{th} item in the sorted list is a random variable denoted $Z_{(\ell:m)}$. To analyze the effect of the Rooney Rule, we are comparing $Y_{(n-k+1:n)}$ with $X_{(\alpha n:\alpha n)}$. Crucially, we are concerned with their expected values conditional on the fact that the committee chose the k^{th}-ranked Y-candidate over the top-ranked X-candidate, implying as noted above that $X_{(\alpha n:\alpha n)}/\beta < Y_{(n-k+1:n)}$. The crucial comparison is therefore between $\mathbb{E}\left[Y_{(n-k+1:n)} | X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right]$ and $\mathbb{E}\left[X_{(\alpha n:\alpha n)} | X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right]$. Order statistics conditional on this type of side information turn out to behave in complex ways, and hence the core of the analysis is in dealing with these types of conditional order statistics for power law distributions.

More generally, given the ubiquity of power law distributions [6], we find it surprising how little is known about how their order statistics behave qualitatively. In this respect, the techniques we provide may prove to be independently useful in other applications. For example, we develop a tight asymptotic characterization of the expectations of order statistics from a power law distribution that to our knowledge is novel.

We also note that although our results are expressed for sufficiently large n, the convergence to the asymptotic behavior happens very quickly as n grows; to handle fixed values of n, we need only modify the bounds by correction terms that grow like $\left(1 \pm O\left(\frac{(\ln n)^2}{n}\right)\right)$. In particular, the errors in the asymptotic analysis are small once n reaches 50, which is reasonable for settings in which a job opening receives many applications.

Estimating the level of bias β. The analysis techniques we develop for proving Theorem 1.2 can also be used for related problems in this model. A specific question we are able to address is the problem of estimating the amount of bias from a history of hiring decisions.

In particular, suppose that over m years the hiring committee makes one offer per year; in N of the m years this offer goes to an X-candidate, and in $m-N$ of the m years this offer goes to a Y-candidate. Which value of the bias parameter β maximizes the probability of this sequence of observations?

We provide a tight characterization of the solution to this question, finding again that it depends not only on α (in this case, the sequence of α values for each year), but also on the power law exponent $1+\delta$. The solution has a qualitatively natural structure, and produces $\beta = 1$ (corresponding to no bias) as the estimate when the fraction of X-candidates hired over the m years is equal to the expected number that would be hired under random selection.

Generalizations to other distributions. Finally, at the end of the paper we consider how to adapt our approach for classes of distributions other than power laws. A different category of distributions that can be motivated by the considerations discussed here is the set of bounded distributions, which take values only over a finite interval. Just as power laws are characteristic of the performance of employees in certain professions, bounded distributions are appropriate when there are absolute constraints on the maximum effect a single employee can have.

Moreover, bounded distributions are also of interest because they contain the uniform distribution on $[0, 1]$ as a special case. We can think of this special case as describing an instance in which each candidate is associated with their *quantile* (between 0 and 1) in a ranking of the entire population, and the bias then operates on this quantile value, reducing it in the case of X-candidates.

For bounded distributions, we can handle much more general forms for the bias — essentially, any function that reduces the values X_i strictly below the maximum of the distribution (for instance, a bias that always prefers a Y-candidate to an X-candidate when they are within some ε of each other). When $k = 2$ and there are equal numbers of X-candidates and Y-candidates, we show that for any bounded distribution and any such bias, the Rooney Rule produces a positive expected change in utility for all sufficiently large n.

1.3 An Illustrative Special Case: Infinite Bias

To illustrate some of the basic considerations that go into our analysis and its interpretation, we begin with a simple special case that we can think of as "infinite bias" — the committee deterministically ranks every Y-candidate above every X-candidate. This case already exhibits structurally rich behavior, although the complexity is enormously less than the case of general β. We also focus here on $k = 2$. In terms of Figure 1, we can visualize the infinite bias case as if we are looking down at the plot from infinitely high up; thus, reasoning about infinite bias amounts to determining which parts of the (α, δ) plane are covered by the surface $\phi_2(\alpha, \beta, \delta) = 1$.

With infinite bias, the committee is guaranteed to choose the two highest-ranked Y-candidates in the absence of an intervention; with the Rooney Rule, the committee will choose the highest-ranked Y-candidate and the highest-ranked X-candidate. As we discuss in the next section, for power law distributions with exponent $1+\delta$, if z^* is the expected maximum of n draws from the distribution, then (i) the expected value of the second-largest of the n draws is $\frac{\delta}{(1+\delta)} z^*$; and (ii) the expected maximum of αn draws from the distribution is asymptotically $\alpha^{1/(1+\delta)} z^*$.

This lets us directly evaluate the utility consequences of the intervention. If there is no intervention, the utility of the committee's decision will be $\left(1 + \frac{\delta}{1+\delta}\right) z^*$, and if the Rooney Rule is used, the utility of the committee's decision will be $(1+\alpha^{1/(1+\delta)})z^*$. Thus, the Rooney Rule produces positive expected change in utility if and only if $\alpha^{1/(1+\delta)} > \frac{\delta}{(1+\delta)}$; that is, if and only if $\alpha > \left(\frac{\delta}{1+\delta}\right)^{1+\delta}$.

In addition to providing a simple closed-form expression for when to use the Rooney Rule in this setting, the condition itself leads to some counter-intuitive consequences. In particular, the closed-form expression for the condition makes it clear that *for every* $\alpha > 0$, there exists a sufficiently small $\delta > 0$ so that when the distribution of applicant potentials is a power law with exponent $1 + \delta$, using the Rooney Rule produces the higher expected utility. In other words, with a power law exponent close to 1, it's a better strategy to commit one of the two offers to the X-candidates, even though they form an extremely small fraction of the population.

This appears to come perilously close to contradicting the following argument. We can arbitrarily divide the Y-candidates into two sets A and B of $n/2$ each; and if $\alpha < 1/2$, each of A and B is larger than the set of all X-candidates. Let a^* be the top candidate in A and b^* be the top candidate in B. Each of a^* and b^* has at least the expected value of the top X-candidate, and moreover, one of them is the top Y-candidate overall. So how can it be that choosing a^* and b^* fails to improve on the result of using the Rooney Rule?

The resolution is to notice that using the Rooney Rule still involves hiring the *top* Y-candidate. So it's not that the Rooney Rule chooses one of a^* or b^* at random, together with the top X-candidate. Rather, it chooses the *better* of a^* and b^*, along with the top X-candidate. The real point is that power law distributions have so much probability in the tail of the distribution that the best person among a set of αn can easily have a higher expected value than the second-best person among a set of n, even when α is quite small. This is a key property of power law distributions that helps explain what's happening both in this example and in our analysis.

1.4 A Non-Monotonicity Effect

As noted above, much of the complexity in the analysis arises from working with expected values of random variables conditioned on the outcomes of certain biased comparisons. One might hope that expected values conditional on these types of comparisons had tractable properties that facilitated the analysis, but this is not the case; in fact, these conditional expectations exhibit some complicated and fairly counter-intuitive behavior. To familiarize the reader with some of these phenomena — both as preparation for the subsequent sections, but also as an interesting end in itself — we offer the following example.

Much of our analysis involves quantities like $\mathbb{E}[X|X > \beta Y]$ — the conditional expectation of X, given that it exceeds some other random variable Y multiplied by a bias parameter. (We will also be analyzing the version in which the inequality goes in the other direction, but we'll focus on the current expression for now.) If we choose X and Y as independent random variables both drawn from a distribution Z, and then view the conditional expectation as a function just of the bias parameter β, what can we say about the properties of this function $f(\beta) = \mathbb{E}[X|X > \beta Y]$?

Intuitively we'd expect $f(\beta)$ to be monotonically increasing in β — indeed, as β increases, we're putting a stricter lower bound on X, and so this ought to raise the conditional expectation of X.

The surprise is that this is not true in general; we can construct independent random variables X and Y for which $f(\beta)$ is not monotonically increasing. In fact, the random variables are very simple: we can have each of X and Y take values independently and uniformly from the finite set $\{1, 5, 9, 13\}$. Now, the event $X > 2Y$ consists of four possible pairs of (X, Y) values: (5,1), (9,1), (13,1), and (13,5). Thus, $f(2) = \mathbb{E}[X|X > 2Y] = 10$. In contrast, the event $X > 3Y$ consists of three possible pairs of (X, Y) values: (5,1), (9,1), and (13,1). Thus, $f(3) = 9$, which is a smaller value, despite the fact that X is required to be a larger multiple of Y.

The surprising content of this example has a fairly sharp formulation in terms of a story about recruiting. Suppose that two academic departments, Department A and Department B, both engage in hiring each year. In our stylized setting, each interviews one X-candidate and one Y-candidate each year, and hires one of them. Each candidate comes from the uniform distribution on $\{1, 5, 9, 13\}$. Departments A and B are both biased in their hiring: A only hires the X-candidate in a given year if they're more than twice as good as the Y-candidate, while B only hires the X-candidate in a given year if they're more than three times as good as the Y-candidate.

Clearly this bias hurts the average quality of both departments, B more so than A. But you might intuitively expect that at least if you looked at the X-candidates that B has actually hired, they'd be of higher average quality than the X-candidates that A has hired — simply because they had to pass through a stronger filter to get hired. In fact, however, this isn't the case: despite the fact that B imposes a stronger filter, the calculations performed above for this example show that the average quality of the X-candidates B hires is 9, while the average quality of the X-candidates A hires is 10.

This non-monotonicity property shows that the conditional expectations we work with in the analysis can be pathologically behaved for arbitrary (even relatively simple) distributions. However, we will see that with power law distributions we are able — with some work — to avoid these difficulties; and part of our analysis will include a set of explicit monotonicity results.

2 Biased Selection with Power Law Distributions

Recall that for a random variable Z, we use $Z_{(\ell:m)}$ to denote the ℓ^{th} order statistic in m draws from Z: the value in position ℓ when we sort m independent draws from Z from lowest to highest. Recall also that when selecting k finalists, the Rooney Rule improves expected utility exactly when

$$\mathbb{E}\left[X_{(\alpha n:\alpha n)} - Y_{(n-k+1:n)} | X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right] > 0.$$

Using linearity of expectation and the fact that $\Pr[A|B]\Pr[B] = \Pr[A \cdot \mathbb{1}_B]$, this is equivalent to

$$\frac{\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}}\right]}{\mathbb{E}\left[Y_{(n-k+1:n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}}\right]} > 1. \tag{2}$$

We will show an asymptotically tight characterization of the tuples of parameters $(k, \alpha, \beta, \delta)$ for which this condition holds, up to an error term on the order of $O\left(\frac{(\ln n)^2}{n}\right)$. In order to better understand the terms in (2), we begin with some necessary background.

2.1 Preliminaries

Fact 1. Let $f_{(p:m)}$ and $F_{(p:m)}$ be, respectively, the probability density function and cumulative distribution function of the p^{th} order statistic out of m draws from the power law distribution with parameter δ. Using definitions from [9],

$$f_{(p:m)}(x) = (1+\delta)(m-p+1)\binom{m}{p-1}\left(1 - x^{-(1+\delta)}\right)^{p-1}\left(x^{-(1+\delta)}\right)^{m-p+1} x^{-1}$$

and

$$F_{(p:m)}(x) = \sum_{j=p}^{m}\binom{m}{j}\left(1 - x^{-(1+\delta)}\right)^{j}\left(x^{-(1+\delta)}\right)^{m-j}.$$

Definition 2. We define

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}\,dt.$$

$\Gamma(\cdot)$ is considered the continuous relaxation of the factorial, and it satisfies

$$\Gamma(a+1) = a\Gamma(a).$$

If a is a positive integer, $\Gamma(a+1) = a!$. Furthermore, $\Gamma(a) > 1$ for $0 < a < 1$ and $\Gamma(a) < 1$ for $1 < a < 2$.

2.2 The Case where $k = 2$

For simplicity, we begin with the case where we're selecting $k = 2$ finalists. In this section, we will make several approximations, growing tight with large n, that we treat formally in Appendices A and B. This section is intended to demonstrate the techniques needed to understand the condition (2). In the case where $k = 2$, always selecting an X-candidate increases expected utility if and only if

$$\frac{\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-1:n)}}\right]}{\mathbb{E}\left[Y_{(n-1:n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-1:n)}}\right]} > 1. \tag{3}$$

Theorems B.1 and B.2 in Appendix B give tight approximations to these quantities; here, we provide an outline for how to find them. For the sake of exposition, we'll only show this for the denominator in this section, which is slightly simpler to approximate. We begin with

$$\mathbb{E}\left[Y_{(n-1:n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-1:n)}}\right] = \int_1^\infty y f_{(n-1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy.$$

Letting $c = \alpha\beta^{-(1+\delta)}$, we can use Lemma D.2 and some manipulation to approximate this by

$$(1+\delta)n(n-1)\int_1^\infty \left(1 - y^{-(1+\delta)}\right)^{n(1+c)-2} \left(y^{-(1+\delta)}\right)^2 dy.$$

Conveniently, the function being integrated is (up to a constant factor) $y \cdot f_{(n(1+c)-1:n(1+c))}(y)$, i.e. y times the probability density function of the second-highest order statistic from $n(1 + c)$ samples. Since

$$\begin{aligned}\mathbb{E}\left[Z_{(n(1+c)-1:n(1+c))}\right] &= \int_1^\infty z f_{(n(1+c)-1:n(1+c))}(z)\, dz \\ &= (1+\delta)n(1+c)(n(1+c)-1)\int_1^\infty \left(1 - z^{-(1+\delta)}\right)^{n(1+c)-2} \left(z^{-(1+\delta)}\right)^2 dz,\end{aligned}$$

we have

$$\mathbb{E}\left[Y_{(n-1:n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-1:n)}}\right] \approx \frac{1}{(1+c)^2}\mathbb{E}\left[Z_{(n(1+c)-1:n(1+c))}\right].$$

Then, we can use Lemmas D.10 and D.11 to get $\mathbb{E}\left[Z_{(n(1+c)-1:n(1+c))}\right] \approx (1+c)^{1/(1+\delta)}\mathbb{E}\left[Y_{(n-1:n)}\right]$, meaning that

$$\mathbb{E}\left[Y_{(n-1:n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-1:n)}}\right] \approx (1+c)^{-(1+\delta/(1+\delta))}\mathbb{E}\left[Y_{(n-1:n)}\right]. \tag{4}$$

For the numerator of (3), a slightly more involved calculation yields

$$\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-1:n)}}\right] \approx \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right]\left[1 - (1+c^{-1})^{-\delta/(1+\delta)}\left[1 + \frac{\delta}{1+\delta}(1+c)^{-1}\right]\right]. \tag{5}$$

By Lemmas D.10 and D.11, $\mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] \approx \Gamma\left(\frac{\delta}{1+\delta}\right)(\alpha n)^{1/(1+\delta)}$ and $\mathbb{E}\left[Y_{(n-1:n)}\right] \approx \Gamma\left(1 + \frac{\delta}{1+\delta}\right)n^{1/(1+\delta)}$. Recall that, up to the approximations we made, the Rooney Rule improves utility in expectation if and only if the ratio between (5) and (4) is larger than 1. Therefore, the following theorem holds:

Theorem 2.1. *For sufficiently large n, the Rooney Rule with $k = 2$ improves utility in expectation if and only if*

$$\frac{\alpha^{1/(1+\delta)}\left[1 - (1+c^{-1})^{-\delta/(1+\delta)}\left[1 + \frac{\delta}{1+\delta}(1+c)^{-1}\right]\right]}{\frac{\delta}{1+\delta}(1+c)^{-1-\delta/(1+\delta)}} > 1. \tag{6}$$

where $c = \alpha\beta^{-(1+\delta)}$.

Note that in the limit as $\beta \to \infty$, $c \to 0$, and the entire expression goes to $\alpha^{1/(1+\delta)}(1+\delta)/\delta$, as noted in Section 1.3. Although the full expression in the statement of Theorem 2.1 is complex, it can be directly evaluated, giving a tight characterization of when the Rule yields increased utility in expectation.

With this result, we could ask for a fixed α and δ how to characterize the set of β such that the condition in (6) holds. In fact, we can show that this expression is monotonically increasing in β.

Theorem 2.2. *The left hand side of* (6) *is decreasing in c and therefore increasing in β. Hence for fixed α and δ there exists β^* such that* (6) *holds if and only if $\beta > \beta^*$.*

Non-monotonicity in δ. From Theorem 2.1, we can gain some intuition for the non-monotonicity in δ shown in Figure 1. For $\alpha < e^{-1}$, we can show that even with infinite bias, the Rooney Rule has a negative effect on utility for sufficiently large δ. Intuitively, this is because the condition for positive change with infinite bias is $\alpha > \left(\frac{\delta}{1+\delta}\right)^{1+\delta}$, which can be written as $\alpha > \left(1-\frac{1}{d}\right)^d$ for $d = 1+\delta$. Since this converges to e^{-1} from below, for sufficiently large δ and $\alpha < e^{-1}$, we have $\alpha < \left(\frac{\delta}{1+\delta}\right)^{1+\delta}$. On the other hand, as $\delta \to 0$, the Rooney Rule has a more negative effect on utility. For instance, $\phi_2(.3, 10, 1) > 1$ but $\phi_2(.3, 10, .5) < 1$. Intuitively, this non-monotonicity arises from the fact that for large δ and small α, the Rooney Rule always has a negative impact on utility, while for very small δ, samples are very far from each other, meaning that the bias has less effect on the ranking.

2.3 The General Case

We can extend these techniques to handle larger values of k. For $k \in [n]$, we define

$$r_k(\alpha,\beta,\delta) = \frac{\mathbb{E}\left[X_{(\alpha n:\alpha n)} | X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right]}{\mathbb{E}\left[Y_{(n-k+1:n)} | X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right]} = \frac{\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}}\right]}{\mathbb{E}\left[Y_{(n-k+1:n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}}\right]}.$$

We can see that the Rooney Rule improves expected utility when selecting k candidates if and only if $r_k > 1$. While r_k depends on n, we will show that it is a very weak dependence: for small k, as n increases, r_k converges to a function of $(\alpha, \beta, \delta, k)$ up to a $1 + O((\ln n)^2/n)$ multiplicative factor. To make this precise, we define the following notion of asymptotic equivalence:

Definition 3. For nonnegative functions $f(n)$ and $g(n)$, define

$$f(n) \approxeq g(n)$$

if and only if there exist $a > 0$ and $n_0 > 0$ such that

$$\frac{f(n)}{g(n)} \leq 1 + \frac{a(\ln n)^2}{n} \quad \text{and} \quad \frac{g(n)}{f(n)} \leq 1 + \frac{a(\ln n)^2}{n}$$

for all $n \geq n_0$. In other words, $f(n) = g(n)\left(1 \pm O\left(\frac{(\ln n)^2}{n}\right)\right)$. When being explicit about a and n_0, we'll write $f(n) \approxeq_{a;n_0} g(n)$.

Appendix C contains a series of lemmas establishing how to rigorously manipulate equivalences of this form. Now, we formally define a tight approximation to r_k, which serves as an expanded restatement of Theorem 1.2 from the introduction.

Theorem 2.3. *For $k \in [n]$, define*

$$\phi_k(\alpha,\beta,\delta) = \frac{\alpha^{1/(1+\delta)} c^{\delta/(1+\delta)} (1+c)^{k-1}}{\binom{k-1-\frac{1}{1+\delta}}{k-1}} \left[(1+c^{-1})^{\delta/(1+\delta)} - \sum_{j=0}^{k-1} \binom{j-\frac{1}{1+\delta}}{j} (1+c)^{-j} \right] \tag{7}$$

where $c = \alpha\beta^{-(1+\delta)}$. Note that ϕ_k does not depend on n. When (α,β,δ) are fixed, we will simply write this as ϕ_k. For $k \le ((1-c^2)\ln n)/2$, we have

$$r_k \approxeq \phi_k,$$

and therefore the Rooney Rule improves expected utility for sufficiently large n if and only if $\phi_k > 1$.

This condition tightly characterizes when the Rooney Rule improves expected utility, and its asymptotic nature in n becomes accurate even for moderately small n: for example, when $n = 50$, the error between r_k and ϕ_k is around 1% for reasonable choices of (α,β,δ).

Increasing k. Consider the scenario in which we're selecting k candidates, and for the given parameter values, the Rooney Rule improves our expected utility. If we were to instead select $k+1$ candidates, should we still be reserving a spot for an X-candidate? Intuitively, as k increases, the Rule is less likely to change our selections, since we're more likely to have already chosen an X-candidate; however, it is not a priori obvious whether increasing k should make it better for us to use the Rooney Rule (because we have more slots, so we're losing less by reserving one) or worse (because as we take more candidates, we stop needing a reserved slot).

In fact, we can apply Theorem 2.3 to understand how r_k changes with k. The following theorem, proven in Appendix B, is an expanded restatement of Theorem 1.3, showing that if the Rooney Rule yields an improvement in expected quality when selecting k candidates, it will do so when selecting $k+1$ candidates as well.

Theorem 2.4. *For $k \le ((1-c^2)\ln n)/2$, we have $\phi_{k+1} > \phi_k$, and therefore for sufficiently large n, we have $r_{k+1} > r_k$.*

Finally, using these techniques, we can provide a tight characterization of the probability that the Rooney Rule produces a positive change. Specifically, we find the probability that the Rooney Rule has a positive effect conditioned on the event that it changes the outcome.

Theorem 2.5.

$$\Pr\left[X_{(\alpha n:\alpha n)} > Y_{(n-k+1:n)} | X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right] \approxeq 1 - \left(\frac{1+\alpha\beta^{-(1+\delta)}}{1+\alpha}\right)^k.$$

To determine whether the Rooney Rule is more likely than not to produce a positive effect (conditioned on changing the outcome), we can compare the right-hand side to $1/2$.

Note that in the case of infinite bias, the right-hand side becomes $1-(1+\alpha)^{-k}$, and thus, the Rooney Rule produces positive change with probability at least $1/2$ if and only if $\alpha \ge \sqrt[k]{2} - 1$. It is interesting to observe that this means with infinite bias, the condition is independent of δ; in contrast, when considering the effect on the expected value with infinite bias, as we did in Section 1.3, the expected change in utility due to the Rooney Rule did depend on δ.

2.4 Maximum Likelihood Estimation of β

The techniques established thus far make it possible to answer other related questions, including the following type of question that we consider in this section: "Given some historical data on past selections, can we estimate the bias present in the data?" For example, suppose that for the last m years, a firm has selected one candidate for each year i out of a pool of $\alpha_i n_i$ X-candidates and n_i Y-candidates. If all applicants are assumed to come from the same underlying distribution, then it is easy to see that the expected number of X-selections (in the absence of bias) should be

$$\sum_{i=1}^{m} \frac{\alpha_i}{1+\alpha_i},$$

regardless of what distribution the applicants come from. However, if there is bias in the selection procedure, then this quantity now depends on the bias model and parameters of the distribution. In particular, in our model, we can use Theorem B.3 to get

$$\Pr\left[X_{(\alpha n:\alpha n)} < \beta Y_{(n:n)}\right] \approx \frac{1}{1+\alpha\beta^{-(1+\delta)}}.$$

This gives us the following approximation for the likelihood of the data $D = (M_1, \ldots, M_m)$ given β, where M_i is 1 if an X-candidate was selected in year i and 0 otherwise:

$$\prod_{i=1}^{m} (1-M_i) \cdot \frac{1}{1+\alpha_i\beta^{-(1+\delta)}} + M_i \cdot \frac{\alpha_i\beta^{-(1+\delta)}}{1+\alpha_i\beta^{-(1+\delta)}}.$$

Taking logarithms, this is

$$\sum_{i:M_i=1} \log(\alpha_i\beta^{-(1+\delta)}) - \sum_{i=1}^{m} \log(1+\alpha_i\beta^{-(1+\delta)}),$$

and maximizing this is equivalent to maximizing

$$\sum_{i:M_i=1} \log(\beta^{-(1+\delta)}) - \sum_{i=1}^{m} \log(1+\alpha_i\beta^{-(1+\delta)}) = N\log(\beta^{-(1+\delta)}) - \sum_{i=1}^{m} \log(1+\alpha_i\beta^{-(1+\delta)})$$

where N is the number of X-candidates selected. Taking the derivative with respect to β, we get

$$-(1+\delta)N\beta^{-1} + (1+\delta)\sum_{i=1}^{m} \frac{\alpha_i\beta^{-(2+\delta)}}{1+\alpha_i\beta^{-(1+\delta)}}.$$

Setting this equal to 0 and canceling common terms, we have

$$\sum_{i=1}^{m} \frac{1}{1+\alpha_i^{-1}\beta^{1+\delta}} = N$$

Since each $1/(1+\alpha_i^{-1}\beta^{1+\delta})$ is strictly monotonically decreasing in β, there is a unique $\hat{\beta}$ for which equality holds, meaning that the likelihood is uniquely maximized by $\hat{\beta}$, up to the $1 \pm O((\ln n)^2/n)$ approximation we made for $\Pr\left[X_{(\alpha n:\alpha n)} < \beta Y_{(n:n)}\right]$. In the special case where $\alpha_i = \alpha$ for $i = 1, \ldots, m$, then the solution is given by

$$\hat{\beta} = \left(\left(\frac{m}{N} - 1\right)\alpha\right)^{1/(1+\delta)}.$$

3 Biased Selection with Bounded Distributions

In this section, we consider a model in which applicants come from a distribution with bounded support. Qualitatively, one would expect different results here from those with power law distributions because in a model with bounded distributions, we expect that for large n, the top order statistics of any distribution will concentrate around the maximum of that distribution. As a result, when there is even a small amount of bias against one population, for large n the probability that *any* of the samples with the highest perceived quality come from that population goes to 0. This means that the Rooney Rule has an effect with high probability, and the effect is positive if the unconditional expectation of the top X-candidate is larger than the unconditional expectation of the Y-candidate that it replaces.

We focus on the case when $\alpha = 1$, meaning we have equal numbers of applicants from both populations. We use the same order statistic notation as before. While all of our previous results have modeled the bias as a multiplicative factor β, we can in fact show that in the bounded distribution setting, for any model of bias $\tilde{X}_{(k:n)} = b(X_{(k:n)})$ such that $b(x) < T$ for $x \geq 0$, where T is strictly less than the maximum of the distribution, the Rooney Rule increases expected utility. Unlike in the previous section the following theorem and analysis are by no means a tight characterization; instead, this is an existence proof that for bounded distributions, there is always a large enough n such that the Rooney Rule improves utility in expectation. We prove our results for continuous distributions with support $[0, 1]$, but a simple scaling argument shows that this extends to any continuous distribution with bounded nonnegative support – specifically, we scale a distribution such that $\inf_{x:f(x)>0} = 0$ and $\sup_{x:f(x)>0} = 1$.

Theorem 3.1. *If f is a continuous probability density function on $[0, 1]$ such that $\sup_{x:f(x)>0} = 1$ and $\tilde{X}_{(n:n)} = b(X_{(n:n)})$ is never more than $T < 1$, then for large enough n,*

$$\mathbb{E}\left[X_{(n:n)} - Y_{(n-1:n)} | b(X_{(n:n)}) < Y_{(n-1:n)}\right] > 0.$$

While we the defer the full proof to Appendix E, the strategy for the proof is as follows:

1. With high probability, $X_{(n:n)}$ and $Y_{(n-1:n)}$ are both large.
2. Whenever $X_{(n:n)}$ and $Y_{(n-1:n)}$ are large, $X_{(n:n)}$ is significantly larger than $Y_{(n-1:n)}$.
3. The gain from switching from $Y_{(n-1:n)}$ to $X_{(n:n)}$ when $X_{(n:n)}$ and $Y_{(n-1:n)}$ are both large outweighs the loss when at least one of them is not large.

4 Conclusion

In this work we have presented a model for implicit bias in a selection problem motivated by settings including hiring and admissions, and we analyzed the Rooney Rule, which can improve the quality of the resulting choices. For one of the most natural settings of the problem, when candidates are drawn from a power-law distribution, we found a tight characterization of the conditions under which the Rooney Rule improves the quality of the outcome. In the process, we identified a number of counter-intuitive effects at work, which we believe may also help provide insight into how we can reason about implicit bias. Our techniques also provided a natural solution to an inference problem in which we estimate parameters of a biased decision-making process. Finally, we performed a similar type of analysis on general bounded distributions.

There are a number of further directions in which these issues could be investigated. One intriguing direction is to consider the possible connections to the theory of optimal delegation (see e.g. [1]).[3]

[3] We thank Ilya Segal for suggesting this connection to us.

In the study of delegation, a *principal* wants a task carried out, but this task can only be performed by an *agent* who may have a utility function that is different from the principal's. In an important family of these models, the principal's only recourse is to impose a restriction on the set of possible actions taken by the agent, creating a more constrained task for the agent to perform, in a way that can potentially improve the quality of the eventual outcome from the principal's perspective. Our analysis of the Rooney Rule can be viewed as taking place from the point of view of a principal who is trying to recruit k candidates, but where the process must be delegated to an agent whose utilities for X-candidates and Y-candidates are different from the principal's, and who is the only party able to evaluate these candidates' potentials. The Rooney Rule, requiring that the agent select at least one X-candidate, is an example of a mechanism that the principal could impose to restrict the agent's set of possible actions, potentially improving the quality of the selected candidates as measured by the principal. More generally, it is interesting to ask whether there are other contexts where such a link between delegation and this type of biased selection provides insight.

Our framework also makes it possible to naturally explore extensions of the basic model. First, the model can be generalized to include noisy observations, potentially with a different level of noise for each group. It would also be interesting to analyze generalizations of the Rooney Rule; for example, if we were to define the ℓ^{th}-*order Rooney Rule* to be the requirement that at least ℓ of k finalists must be from an underrepresented group, we could ask which ℓ produces the greatest increase in utility for a given set of parameters. Finally, we could benefit from a deeper undestanding of the function ϕ that appears in our main theorems. For example, while we showed in Theorem 1.3 that ϕ is monotone in β for $k = 2$, Figure 1 shows that ϕ is clearly not monotone in δ. A better understaning of the function ϕ may lead to new insights into our model and into the phenomena it seeks to capture.

Acknowledgements We thank Eric Parsonnet for his invaluable technical insights. This work was supported in part by a Simons Investigator Grant and an NSF Graduate Fellowship.

References

[1] Ricardo Alonso and Niko Matouschek. Optimal delegation. *Review of Economic Studies*, 75(1):259–293, January 2008.

[2] Marianne Bertrand and Sendhil Mullainathan. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013, September 2004.

[3] Iris Bohnet, Alexandra van Geen, and Max Bazerman. When performance trumps gender bias: Joint vs. separate evaluation. *Management Science*, 62(5):1225–1234, 2016.

[4] Mark Braverman and Elchanan Mossel. Sorting from noisy information. *arXiv preprint arXiv:0910.1191*, 2009.

[5] Marilyn Cavicchia. How to fight implicit bias? With conscious thought, diversity expert tells NABE. *American Bar Association: Bar Leader*, 40(1), 2015.

[6] Aaron Clauset, Cosma R. Shalizi, and Mark E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[7] Brian W. Collins. Tackling unconscious bias in hiring practices: The plight of the Rooney Rule. *NYU Law Review*, 82(3), 2007.

[8] Covington and Burling. Recommendations to Uber, 13 June 2017.

[9] H. A. David and H. N. Nagaraja. *Basic Distribution Theory*, pages 9–32. John Wiley & Sons, Inc., 2005.

[10] Uriel Feige, Prabhakar Raghavan, David Peleg, and Eli Upfal. Computing with noisy information. *SIAM Journal on Computing*, 23(5):1001–1018, 1994.

[11] Qiang Fu and Jingfeng Lu. Micro foundations of multi-prize lottery contests: a perspective of noisy performance ranking. *Social Choice and Welfare*, 38(3):497–517, 2012.

[12] Alexander R. Green, Dana R. Carney, Daniel J. Pallin, Long H. Ngo, Kristal L. Raymond, Lisa I. Iezzoni, and Mahzarin R. Banaji. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22(9):1231–1238, 2007.

[13] Anthony G. Greenwald and Mahzarin R. Banaji. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1):4–27, 1995.

[14] Anthony G. Greenwald and Linda Hamilton Krieger. Implicit bias: Scientific foundations. *California Law Review*, 94:945–967, 2006.

[15] Christine Jolls and Cass R. Sunstein. The law of implicit bias. *California Law Review*, 94:969–996, 2006.

[16] Manuel Lopez and James Marengo. An upper bound for the expected difference between order statistics. *Mathematics Magazine*, 84(5):365–369, 2011.

[17] Henrick John Malik. Exact moments of order statistics from the pareto distribution. *Scandinavian Actuarial Journal*, 1966(3-4):144–157, 1966.

[18] Christina Passariello. Tech firms borrow football play to increase hiring of women. *Wall Street Journal*, 27 September 2016.

[19] Hamza Shaban. What is the “Rooney Rule” that Uber just adopted? *Washington Post*, 13 June 2017.

[20] Francesco Giacomo Tricomi and Arthur Erdélyi. The asymptotic expansion of a ratio of gamma functions. *Pacific J. Math*, 1(1):133–142, 1951.

[21] Eric Luis Uhlmann and Geoffrey L. Cohen. Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, 16(6):474–480, 2005.

[22] Linda van den Bergh, Eddie Denessen, Lisette Hornstra, Marinus Voeten, and Rob W. Holland. The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Education Research Journal*, 47(2):497–527, 2010.

[23] Christine Wenneras and Agnes Wold. Nepotism and sexism in peer-review. *Nature*, 387:341–343, 1997.

A Missing Proofs for Section 2

Proof of Theorems 2.1 and 2.3. We can expand the statement in Theorem B.1 to

$$
\begin{aligned}
&\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}}\right] \\
&\approx \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right]\left[1-(1+c^{-1})^{-\delta/(1+\delta)}\sum_{j=0}^{k-1}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j}\right] \\
&\approx (\alpha n)^{1/(1+\delta)}\Gamma\left(\frac{\delta}{1+\delta}\right)\left[1-(1+c^{-1})^{-\delta/(1+\delta)}\sum_{j=0}^{k-1}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j}\right] \qquad \text{(By Lemma D.10)}
\end{aligned}
$$

This gives us a ratio

$$
\begin{aligned}
r_k(\alpha,\beta,\delta) &= \frac{\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}}\right]}{\mathbb{E}\left[Y_{(n-k+1:n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}}\right]} \\
&\approx \frac{(\alpha n)^{1/(1+\delta)}\Gamma\left(\frac{\delta}{1+\delta}\right)\left[1-(1+c^{-1})^{-\delta/(1+\delta)}\sum_{j=0}^{k-1}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j}\right]}{(1+c)^{-(k-1/(1+\delta))}\frac{\Gamma\left(k-\frac{1}{1+\delta}\right)}{\Gamma(k)}n^{1/(1+\delta)}} \\
&\qquad \text{(Using Theorem B.2)} \\
&= \frac{\alpha^{1/(1+\delta)}\Gamma(k)\Gamma\left(\frac{\delta}{1+\delta}\right)(1+c)^{k-1/(1+\delta)}}{\Gamma\left(k-\frac{1}{1+\delta}\right)}\left[1-(c^{-1}(1+c))^{-\delta/(1+\delta)}\sum_{j=0}^{k-1}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j}\right] \\
&= \frac{\alpha^{1/(1+\delta)}c^{\delta/(1+\delta)}\Gamma(k)\Gamma\left(\frac{\delta}{1+\delta}\right)(1+c)^{k-1}}{\Gamma\left(k-\frac{1}{1+\delta}\right)}\left[(1+c^{-1})^{\delta/(1+\delta)}-\sum_{j=0}^{k-1}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j}\right] \\
&= \frac{\alpha^{1/(1+\delta)}c^{\delta/(1+\delta)}(1+c)^{k-1}}{\binom{k-1-\frac{1}{1+\delta}}{k-1}}\left[(1+c^{-1})^{\delta/(1+\delta)}-\sum_{j=0}^{k-1}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j}\right]
\end{aligned}
$$

□

Proof of Theorem 2.2. Since the only influence of β is through c and c is decreasing in β, it is sufficient to show that

$$
\frac{\alpha^{1/(1+\delta)}\left[1-(1+c^{-1})^{-\delta/(1+\delta)}\left[1+\frac{\delta}{1+\delta}(1+c)^{-1}\right]\right]}{\frac{\delta}{1+\delta}(1+c)^{-1-\delta/(1+\delta)}}
$$

is decreasing in c. Ignoring constants, this is

$$
\begin{aligned}
&\propto c^{\delta/(1+\delta)}(1+c)\left[(1+c^{-1})^{\delta/(1+\delta)}-1-\frac{\delta}{1+\delta}(1+c)^{-1}\right] \\
&= (1+c)^{1+\delta/(1+\delta)}-c^{\delta/(1+\delta)}(1+c)-\frac{\delta}{1+\delta}c^{\delta/(1+\delta)} \\
&= (1+c)^{1+\delta/(1+\delta)}-c^{1+\delta/(1+\delta)}-\left(1+\frac{\delta}{1+\delta}\right)c^{\delta/(1+\delta)}
\end{aligned}
$$

This has derivative

$$\begin{aligned}
&\frac{\delta}{dc}(1+c)^{1+\delta/(1+\delta)} - c^{1+\delta/(1+\delta)} - \left(1+\frac{\delta}{1+\delta}\right)c^{\delta/(1+\delta)} \\
&\quad = \left(1+\frac{\delta}{1+\delta}\right)(1+c)^{\delta/(1+\delta)} - \left(1+\frac{\delta}{1+\delta}\right)c^{\delta/(1+\delta)} + \left(\frac{\delta}{1+\delta}\right)\left(1+\frac{\delta}{1+\delta}\right)c^{-1/(1+\delta)},
\end{aligned}$$

which is negative if and only if

$$\begin{aligned}
&(1+c)^{\delta/(1+\delta)} < c^{\delta/(1+\delta)} + \frac{\delta}{1+\delta}c^{-1/(1+\delta)} \\
\iff &(1+c)^{\delta/(1+\delta)}c^{-\delta/(1+\delta)} < 1 + \frac{\delta}{1+\delta}c^{-1} \\
\iff &(1+c^{-1})^{\delta/(1+\delta)} < 1 + \frac{\delta}{1+\delta}c^{-1}.
\end{aligned}$$

This is true by Lemma D.9, which proves the theorem. □

Proof of Theorem 2.4. By Theorem 2.3,

$$\phi_k(\alpha,\beta,\delta) = \frac{\alpha^{1/(1+\delta)}c^{\delta/(1+\delta)}\Gamma(k)\Gamma\left(\frac{\delta}{1+\delta}\right)(1+c)^{k-1}}{\Gamma\left(k-\frac{1}{1+\delta}\right)}\left[(1+c^{-1})^{\delta/(1+\delta)} - \sum_{j=0}^{k-1}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j}\right]$$

We use the fact that for $a, b \in \mathbb{Z}$ and $s \in \mathbb{R}$

$$\frac{\Gamma(s-a+1)}{\Gamma(s-b+1)} = (-1)^{b-a}\frac{\Gamma(b-s)}{\Gamma(a-s)}.$$

If the summation went to ∞, it would be

$$\begin{aligned}
\sum_{j=0}^{\infty}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j} &= \sum_{j=0}^{\infty}(1+c)^{-j}\frac{\Gamma\left(j+\frac{\delta}{1+\delta}\right)}{\Gamma\left(\frac{\delta}{1+\delta}\right)\Gamma(j+1)} \\
&= \sum_{j=0}^{\infty}(1+c)^{-j}(-1)^j\frac{\Gamma\left(1-\frac{\delta}{1+\delta}\right)}{\Gamma\left(-j+1+\frac{\delta}{1+\delta}\right)\Gamma(j+1)} \\
&= \sum_{j=0}^{\infty}\binom{-\frac{\delta}{1+\delta}}{j}(-(1+c)^{-1})^j \\
&= (1-(1+c)^{-1})^{-\delta/(1+\delta)} \\
&= (1+c^{-1})^{\delta/(1+\delta)}
\end{aligned}$$

Therefore,

$$\sum_{j=0}^{k-1}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j} = (1+c^{-1})^{\delta/(1+\delta)} - \sum_{j=k}^{\infty}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j}.$$

Plugging this in,

$$\begin{aligned}
\phi_k(\alpha,\beta,\delta) &= \frac{\alpha^{1/(1+\delta)}c^{\delta/(1+\delta)}\Gamma(k)\Gamma\left(\frac{\delta}{1+\delta}\right)(1+c)^{k-1}}{\Gamma\left(k-\frac{1}{1+\delta}\right)}\sum_{j=k}^{\infty}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j} \\
&= \frac{\alpha^{1/(1+\delta)}c^{\delta/(1+\delta)}\Gamma(k)\Gamma\left(\frac{\delta}{1+\delta}\right)}{\Gamma\left(k-\frac{1}{1+\delta}\right)(1+c)}\sum_{j=0}^{\infty}\binom{j+k-\frac{1}{1+\delta}}{j+k}(1+c)^{-j}
\end{aligned}$$

With this, we can take

$$\begin{aligned}
&\phi_{k+1}(\alpha,\beta,\delta)-\phi_k(\alpha,\beta,\delta)\\
&\quad=\frac{\alpha^{1/(1+\delta)}c^{\delta/(1+\delta)}\Gamma\left(\frac{\delta}{1+\delta}\right)}{(1+c)}\left[\frac{\Gamma(k+1)}{\Gamma\left(k+\frac{\delta}{1+\delta}\right)}\sum_{j=0}^{\infty}\binom{j+k+1-\frac{1}{1+\delta}}{j+k+1}(1+c)^{-j}\right.\\
&\quad\left.-\frac{\Gamma(k)}{\Gamma\left(k-\frac{1}{1+\delta}\right)}\sum_{j=0}^{\infty}\binom{j+k-\frac{1}{1+\delta}}{j+k}(1+c)^{-j}\right]\\
&\quad=\frac{\alpha^{1/(1+\delta)}c^{\delta/(1+\delta)}\Gamma(k)\Gamma\left(\frac{\delta}{1+\delta}\right)}{\Gamma\left(k-\frac{1}{1+\delta}\right)(1+c)}\left[\frac{k}{k-\frac{1}{1+\delta}}\sum_{j=0}^{\infty}\binom{j+k+1-\frac{1}{1+\delta}}{j+k+1}(1+c)^{-j}\right.\\
&\quad\left.-\sum_{j=0}^{\infty}\binom{j+k-\frac{1}{1+\delta}}{j+k}(1+c)^{-j}\right]\\
&\quad=\frac{\alpha^{1/(1+\delta)}c^{\delta/(1+\delta)}\Gamma(k)\Gamma\left(\frac{\delta}{1+\delta}\right)}{\Gamma\left(k-\frac{1}{1+\delta}\right)(1+c)}\sum_{j=0}^{\infty}(1+c)^{-j}\left[\frac{k}{k-\frac{1}{1+\delta}}\binom{j+k+1-\frac{1}{1+\delta}}{j+k+1}-\binom{j+k-\frac{1}{1+\delta}}{j+k}\right]
\end{aligned}$$

Thus, to show that $\phi_{k+1}>\phi_k$, it is sufficient to show that for $j\geq 0$,

$$\begin{aligned}
\frac{k}{k-\frac{1}{1+\delta}}\binom{j+k+1-\frac{1}{1+\delta}}{j+k+1}-\binom{j+k-\frac{1}{1+\delta}}{j+k}&>0\\
\frac{k}{k-\frac{1}{1+\delta}}\frac{\Gamma\left(j+k+1+\frac{\delta}{1+\delta}\right)}{\Gamma(j+k+2)\Gamma\left(\frac{\delta}{1+\delta}\right)}-\frac{\Gamma\left(j+k+\frac{\delta}{1+\delta}\right)}{\Gamma(j+k+1)\Gamma\left(\frac{\delta}{1+\delta}\right)}&>0\\
\frac{k}{k-\frac{1}{1+\delta}}\frac{j+k+\frac{\delta}{1+\delta}}{j+k+1}-1&>0 \qquad (\Gamma(x+1)=x\Gamma(x))\\
\frac{k-\frac{1}{1+\delta}+(j+1)}{k+(j+1)}&>\frac{k-\frac{1}{1+\delta}}{k}
\end{aligned}$$

The last inequality holds by Lemma D.4. As a result a result, $\phi_{k+1}>\phi_k$, proving the theorem. □

Proof Theorem 2.5. We want to find

$$\Pr\left[X_{(\alpha n:\alpha n)}>Y_{(n-k+1:n)}|X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}\right],$$

or equivalently,

$$\Pr\left[X_{(\alpha n:\alpha n)}<Y_{(n-k+1:n)}|X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}\right].$$

This can be written as

$$\frac{\Pr\left[X_{(\alpha n:\alpha n)}<Y_{(n-k+1:n)}\cap X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}\right]}{\Pr\left[X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}\right]}=\frac{\Pr\left[X_{(\alpha n:\alpha n)}<Y_{(n-k+1:n)}\right]}{\Pr\left[X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}\right]}. \tag{8}$$

By Theorem B.3, the numerator can be approximated by $(1+\alpha)^{-k}$ while the denominaotr is approximately $(1+\alpha\beta^{-(1+\delta)})^{-k}$. Thus, we have

$$\Pr\left[X_{(\alpha n:\alpha n)}<Y_{(n-k+1:n)}|X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}\right]\approx\frac{(1+\alpha\beta^{-(1+\delta)})^k}{(1+\alpha)^k}=\left(\frac{1+\alpha\beta^{-(1+\delta)}}{1+\alpha}\right)^k,$$

and therefore

$$\Pr\left[X_{(\alpha n:\alpha n)} > Y_{(n-k+1:n)} | X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right] \approxeq 1 - \left(\frac{1+\alpha\beta^{-(1+\delta)}}{1+\alpha}\right)^k.$$

□

B Additional Theorems for Power Laws

Theorem B.1.

$$\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}}\right] \approxeq \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right]\left[1-(1+c^{-1})^{-\delta/(1+\delta)}\sum_{j=0}^{k-1}\binom{j-\frac{1}{1+\delta}}{j}(1+c)^{-j}\right]$$

where $c = \alpha\beta^{-(1+\delta)}$.

Proof. First, observe that

$$\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}<\beta Y_{(n-k+1:n)}}\right] = \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] - \mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}\geq\beta Y_{(n-k+1:n)}}\right].$$

Next, we use the fact that

$$\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}\geq\beta Y_{(n-k+1:n)}}\right] = \int_\beta^\infty x f_{(\alpha n:\alpha n)}(x) F_{(n-k+1:n)}\left(\frac{x}{\beta}\right)\,dx.$$

We know that

$$\begin{aligned}\int_\beta^\infty x f_{(\alpha n:\alpha n)}(x) F_{(n-k+1:n)}\left(\frac{x}{\beta}\right)\,dx &= \int_\beta^{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}} x f_{(\alpha n:\alpha n)}(x) F_{(n-k+1:n)}\left(\frac{x}{\beta}\right)\,dx \\ &\quad + \int_{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}}^\infty x f_{(\alpha n:\alpha n)}(x) F_{(n-k+1:n)}\left(\frac{x}{\beta}\right)\,dx, \qquad (9)\end{aligned}$$

and

$$\begin{aligned}\int_\beta^{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}} x f_{(\alpha n:\alpha n)}(x) F_{(n-k+1:n)}\left(\frac{x}{\beta}\right)\,dx &\leq \left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)} F_{(\alpha n:\alpha n)}\left(\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}\right) \\ &\leq \left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)} \cdot \frac{1}{n}\end{aligned}$$

by Lemma D.6. The second term of (9) is

$$\begin{aligned}&(1+\delta)\alpha n \int_{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}}^\infty (1-x^{-(1+\delta)})^{\alpha n-1} x^{-(1+\delta)} \sum_{j=0}^{k-1}\binom{n}{j}\left(1-\left(\frac{x}{\beta}\right)^{-(1+\delta)}\right)^{n-j}\left(\frac{x}{\beta}\right)^{-j(1+\delta)}\,dx \\ &= (1+\delta)\alpha n \sum_{j=0}^{k-1}\binom{n}{j}\beta^{j(1+\delta)}\int_{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}}^\infty (1-x^{-(1+\delta)})^{\alpha n-1}\left(x^{-(1+\delta)}\right)^{j+1}\left(1-\left(\frac{x}{\beta}\right)^{-(1+\delta)}\right)^{n-j}\,dx\end{aligned}$$

Next, we show that for $x \geq \left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}$,

$$\left(1-\left(\frac{x}{\beta}\right)^{-(1+\delta)}\right)^{n-j} \approxeq (1-x^{-(1+\delta)})^{\beta^{1+\delta}n-j}.$$

We begin with

$$\left(1-\left(\frac{x}{\beta}\right)^{-(1+\delta)}\right)^{n-j} \approxeq (1-x^{-(1+\delta)})^{\beta^{1+\delta}(n-j)} = (1-x^{-(1+\delta)})^{\beta^{1+\delta}n-j}(1-x^{-(1+\delta)})^{-j(\beta^{1+\delta}-1)}.$$

Note that $(1-x^{-(1+\delta)})^{-j(\beta^{1+\delta}-1)} \geq 1$, and by Lemma D.5,

$$(1-x^{-(1+\delta)})^{-j(\beta^{1+\delta}-1)} = 1+j(\beta^{1+\delta}-1)x^{-(1+\delta)} + O\left(\frac{1}{n}\right) \approxeq 1.$$

because $j \leq \ln n$. Thus, $(1-x^{-(1+\delta)})^{\beta^{1+\delta}n-j}(1-x^{-(1+\delta)})^{-j(\beta^{1+\delta}-1)} \approxeq (1-x^{-(1+\delta)})^{\beta^{1+\delta}n-j}$. Therefore, this becomes

$$(1+\delta)\alpha n\sum_{j=0}^{k-1}\binom{n}{j}\beta^{j(1+\delta)}\int_{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}}^{\infty}\left(1-x^{-(1+\delta)}\right)^{\beta^{1+\delta}n(1+c)-j-1}\left(x^{-(1+\delta)}\right)^{j+1}\,dx.$$

We'll now try to relate the jth term in this summation to the order statistic $Z_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}$. We know that

$$\begin{aligned}
&\mathbb{E}\left[Z_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}\right]\\
&= \int_1^{\infty} z f_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}(z)\,dz\\
&= (1+\delta)(j+1)\binom{\beta^{1+\delta}n(1+c)}{j+1}\int_1^{\infty}\left(1-z^{-(1+\delta)}\right)^{\beta^{1+\delta}n(1+c)-j-1}\left(z^{-(1+\delta)}\right)^{j+1}\,dz.
\end{aligned}$$

Using this, we have

$$\begin{aligned}
\int_{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}}^{\infty} x f_{(\alpha n:\alpha n)}(x)F_{(n-k+1:n)}\left(\frac{x}{\beta}\right)\,dx \approxeq \sum_{j=0}^{k-1}\frac{\alpha n\beta^{j(1+\delta)}\binom{n}{j}}{(j+1)\binom{\beta^{1+\delta}n(1+c)}{j+1}}&\Bigg[\mathbb{E}\left[Z_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}\right]\\
&-\int_1^{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}} z f_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}(z)\,dz\Bigg]
\end{aligned}$$

We'll show that this last multiplicative term is approximately $\mathbb{E}\left[Z_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}\right]$. Observe that

$$\begin{aligned}
&\int_1^{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}} z f_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}(z)\,dz\\
&\qquad\leq \left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}\int_1^{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}} f_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}(z)\,dz\\
&\qquad= \left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)} F_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}\left(\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}\right)\\
&\qquad\leq \left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}\frac{\sqrt{k}}{n}
\end{aligned}$$

by Lemma D.3. This means

$$\begin{aligned}
&\sum_{j=0}^{k-1} \frac{\alpha n \beta^{j(1+\delta)} \binom{n}{j}}{(j+1)\binom{\beta^{1+\delta} n(1+c)}{j+1}} \mathbb{E}\left[Z_{(\beta^{1+\delta} n(1+c)-j:\beta^{1+\delta} n(1+c))}\right] \\
&\quad \geq \sum_{j=0}^{k-1} \frac{\alpha n \beta^{j(1+\delta)} \binom{n}{j}}{(j+1)\binom{\beta^{1+\delta} n(1+c)}{j+1}} \left[\mathbb{E}\left[Z_{(\beta^{1+\delta} n(1+c)-j:\beta^{1+\delta} n(1+c))}\right] - \left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)} \frac{\sqrt{k}}{n}\right] \\
&\quad \approx \sum_{j=0}^{k-1} \frac{\alpha n \beta^{j(1+\delta)} \binom{n}{j}}{(j+1)\binom{\beta^{1+\delta} n(1+c)}{j+1}} \mathbb{E}\left[Z_{(\beta^{1+\delta} n(1+c)-j:\beta^{1+\delta} n(1+c))}\right] \qquad \text{(by Lemma D.7)}
\end{aligned}$$

Next, we deal with the $n\beta^{j(1+\delta)}\binom{n}{j}/((j+1)\binom{\beta^{1+\delta}n(1+c)}{j+1})$ terms. These are

$$\frac{n \beta^{j(1+\delta)} \binom{n}{j}}{(j+1)\binom{\beta^{1+\delta} n(1+c)}{j+1}} = \frac{n(n-1)\cdots(n-j+1)}{\beta^{1+\delta} n(1+c)(\beta^{1+\delta} n(1+c)-1)\cdots(\beta^{1+\delta} n(1+c)-j+1)} \cdot \frac{n\beta^{j(1+\delta)}}{\beta^{1+\delta} n(1+c)-j}. \tag{10}$$

Each term $(n-\ell)/(\beta^{1+\delta}n(1+c)-\ell)$ is between $1/(\beta^{1+\delta}(1+c))$ and $1/(\beta^{1+\delta}(1+c))\cdot(1-\ell/n)$. This means

$$\frac{1}{(\beta^{1+\delta}(1+c))^j} \geq \prod_{\ell=0}^{j} \frac{n-\ell}{\beta^{1+\delta} n(1+c)-\ell} \geq \prod_{\ell=0}^{j} \frac{1}{\beta^{1+\delta}(1+c)}\left(1-\frac{\ell}{n}\right) \geq \left(1-\frac{j^2}{n}\right) \approx \frac{1}{(\beta^{1+\delta}(1+c))^j}$$

since $j \leq k \leq ((1-c^2)/2)\ln n$. Multiplying by the second term in (10), which is

$$\frac{n\beta^{j(1+\delta)}}{\beta^{1+\delta} n(1+c)-j} \approx \frac{\beta^{(j-1)(1+\delta)}}{1+c},$$

we have

$$\frac{n \beta^{j(1+\delta)} \binom{n}{j}}{(j+1)\binom{\beta^{1+\delta} n(1+c)}{j+1}} \approx \frac{1}{\beta^{1+\delta}(1+c)^{j+1}}.$$

As a result,

$$\begin{aligned}
\int_{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}}^{\infty} x f_{(\alpha n:\alpha n)}(x) F_{(n-k+1:n)}\left(\frac{x}{\beta}\right) dx &\approx \sum_{j=0}^{k-1} \frac{\alpha}{\beta^{1+\delta}(1+c)^{j+1}} \mathbb{E}\left[Z_{(\beta^{1+\delta} n(1+c)-j:\beta^{1+\delta} n(1+c))}\right] \\
&= \sum_{j=0}^{k-1} \frac{c}{(1+c)^{j+1}} \mathbb{E}\left[Z_{(\beta^{1+\delta} n(1+c)-j:\beta^{1+\delta} n(1+c))}\right] \qquad (11)
\end{aligned}$$

Finally, note that

$$\begin{aligned}
\mathbb{E}\left[Z_{(\beta^{1+\delta} n(1+c)-j:\beta^{1+\delta} n(1+c))}\right] &= \mathbb{E}\left[Z_{(n\beta^{1+\delta}(1+c):\beta^{1+\delta} n(1+c))}\right] \frac{\Gamma(j+\delta/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(j+1)} \\
&\approx (\beta^{1+\delta} n(1+c))^{1/(1+\delta)} \frac{\Gamma(j+\delta/(1+\delta))}{\Gamma(j+1)} \\
&= \beta(1+c)^{1/(1+\delta)} n^{1/(1+\delta)} \frac{\Gamma(j+\delta/(1+\delta))}{\Gamma(j+1)} \\
&= \frac{\beta}{\alpha^{1/(1+\delta)}}(1+c)^{1/(1+\delta)}(\alpha n)^{1/(1+\delta)} \frac{\Gamma(j+\delta/(1+\delta))}{\Gamma(j+1)} \\
&\approx c^{-1/(1+\delta)}(1+c)^{1/(1+\delta)} \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] \frac{\Gamma(j+\delta/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(j+1)}
\end{aligned}$$

Substituting back to (11),

$$\int_{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}}^{\infty} x f_{(\alpha n:\alpha n)}(x) \approx \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] c^{\delta/(1+\delta)} \sum_{j=0}^{k-1} (1+c)^{-(j+\delta/(1+\delta))} \frac{\Gamma(j+\delta/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(j+1)}$$

Going back to (9),

$$\begin{aligned}
\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} > \beta Y_{(n-k+1:n)}}\right] &\approx \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] c^{\delta/(1+\delta)} \sum_{j=0}^{k-1} (1+c)^{-(j+\delta/(1+\delta))} \frac{\Gamma(j+\delta/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(j+1)} \\
&+ \int_{\beta}^{\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}} x f_{(\alpha n:\alpha n)}(x) F_{(n-k+1:n)}\left(\frac{x}{\beta}\right) dx \\
&\leq \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] c^{\delta/(1+\delta)} \sum_{j=0}^{k-1} (1+c)^{-(j+\delta/(1+\delta))} \frac{\Gamma(j+\delta/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(j+1)} \\
&+ \left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)} \frac{(\ln n)^2}{n} \\
&\approx \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] c^{\delta/(1+\delta)} \sum_{j=0}^{k-1} (1+c)^{-(j+\delta/(1+\delta))} \frac{\Gamma(j+\delta/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(j+1)}
\end{aligned}$$

Therefore,

$$\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}}\right] \approx \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] \left[1 - c^{\delta/(1+\delta)} \sum_{j=0}^{k-1} (1+c)^{-(j+\delta/(1+\delta))} \frac{\Gamma\left(j+\frac{\delta}{1+\delta}\right)}{\Gamma\left(\frac{\delta}{1+\delta}\right)\Gamma(j+1)}\right].$$

We can simplify this to

$$\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}}\right] \approx \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] \left[1 - (1+c^{-1})^{-\delta/(1+\delta)} \sum_{j=0}^{k-1} (1+c)^{-j} \frac{\Gamma\left(j+\frac{\delta}{1+\delta}\right)}{\Gamma\left(\frac{\delta}{1+\delta}\right)\Gamma(j+1)}\right].$$

Using the definition

$$\binom{a}{b} = \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)},$$

this is

$$\mathbb{E}\left[X_{(\alpha n:\alpha n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}}\right] \approx \mathbb{E}\left[X_{(\alpha n:\alpha n)}\right] \left[1 - (1+c^{-1})^{-\delta/(1+\delta)} \sum_{j=0}^{k-1} \binom{j-\frac{1}{1+\delta}}{j} (1+c)^{-j}\right].$$

□

Theorem B.2.

$$\mathbb{E}\left[Y_{(n-k+1:n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}}\right] \approx (1+\alpha\beta^{-(1+\delta)})^{-(k-1/(1+\delta))} \mathbb{E}\left[Y_{(n-k+1:n)}\right]$$

Proof. We begin with

$$\mathbb{E}\left[Y_{(n-k+1:n)} \cdot \mathbb{1}_{X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}}\right] = \int_1^\infty y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy.$$

Let $c = \alpha\beta^{-(1+\delta)}$. Break this up into

$$\begin{aligned}\int_1^\infty y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy = \int_1^{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}} & y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy \\ + \int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^\infty & y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy. \end{aligned} \tag{12}$$

The first term is

$$\begin{aligned}\int_1^{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}} & y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy \\ & \le F_{(\alpha n:\alpha n)}\left(\beta \left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}\right) \int_1^{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}} y f_{(n-k+1:n)}(y)\, dy \\ & \le F_{(\alpha n:\alpha n)}\left(\beta \left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}\right) \mathbb{E}\left[Y_{(n-k+1:n)}\right] \\ & \le \frac{\mathbb{E}\left[Y_{(n-k+1:n)}\right]}{n}\end{aligned}$$

by Lemma D.6.

For the second term in (12), we have

$$\begin{aligned}\int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^\infty & y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy \\ & = (1+\delta)k\binom{n}{k} \int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^\infty \left(1 - y^{-(1+\delta)}\right)^{n-k} \left(y^{-(1+\delta)}\right)^k \left(1 - (\beta y)^{-(1+\delta)}\right)^{\alpha n} dy\end{aligned}$$

By Lemma D.2, for all $y \ge (cn/\ln n)^{1/(1+\delta)}$,

$$\left(1 - (\beta y)^{-(1+\delta)}\right)^{\alpha n} \approxeq \left(1 - y^{-(1+\delta)}\right)^{cn}.$$

Therefore,

$$\begin{aligned}\int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^\infty & y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy \\ & \approxeq (1+\delta)k\binom{n}{k} \int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^\infty \left(1 - y^{-(1+\delta)}\right)^{n-k+cn} \left(y^{-(1+\delta)}\right)^k dy \\ & = (1+\delta)k\binom{n}{k} \int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^\infty \left(1 - y^{-(1+\delta)}\right)^{n(1+c)-k} \left(y^{-(1+\delta)}\right)^k dy.\end{aligned}$$

We'll now try to relate this to the order statistic $Z_{(n(1+c)-k+1:n(1+c))}$. We know that

$$\begin{aligned}\mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right] &= \int_1^\infty z f_{(n(1+c)-k+1:n(1+c))}(z)\, dz \\ &= (1+\delta)k\binom{n(1+c)}{k} \int_1^\infty \left(1 - z^{-(1+\delta)}\right)^{n(1+c)-k} \left(z^{-(1+\delta)}\right)^k dz.\end{aligned}$$

Using this, we have

$$\int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^{\infty} y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy$$
$$\approx \frac{\binom{n}{k}}{\binom{n(1+c)}{k}} \left[\mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))} \right] - \int_{1}^{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}} y f_{(n(1+c)-k+1:n(1+c))}(y)\, dy \right]. \quad (13)$$

From here, we'll show that the term being subtracted is only a $\frac{\sqrt{\ln n}}{n}$ fraction of $\mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right]$. To do so, note that

$$\begin{aligned}
\int_{1}^{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}} y f_{(n(1+c)-k+1:n(1+c))}(y)\, dy &\leq \left(\frac{cn}{\ln n}\right)^{1/(1+\delta)} \int_{1}^{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}} f_{(n(1+c)-k+1:n(1+c))}(y)\, dy \\
&= \left(\frac{cn}{\ln n}\right)^{1/(1+\delta)} F_{(n(1+c)-k+1:n(1+c))}\left(\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)} \right) \\
&\leq \left(\frac{cn}{\ln n}\right)^{1/(1+\delta)} \left(\frac{\sqrt{k}}{n}\right)
\end{aligned}$$

By Lemma D.1. Lemma D.7 gives us

$$\begin{aligned}
\mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right] &\geq \mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right] - \int_{1}^{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}} y f_{(n(1+c)-k+1:n(1+c))}(y)\, dy \\
&\geq \mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right] - \left(\frac{cn}{\ln n}\right)^{1/(1+\delta)} \left(\frac{\sqrt{k}}{n}\right) \\
&\geq \mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right] \left(1 - \frac{\sqrt{k}}{n}\right) \\
&\approx \mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right]
\end{aligned}$$

Combining with (13), Lemma C.6 yields

$$\int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^{\infty} y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy \approx \frac{\binom{n}{k}}{\binom{n(1+c)}{k}} \mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right] \quad (14)$$

By Lemma D.8,

$$\frac{\binom{n}{k}}{\binom{n(1+c)}{k}} \approx \frac{1}{(1+c)^k}.$$

Putting this into (14),

$$\int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^{\infty} y f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy \approx \frac{1}{(1+c)^k} \mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right].$$

Finally, note that

$$\begin{aligned}
\mathbb{E}\left[Z_{(n(1+c)-k+1:n(1+c))}\right] &= \mathbb{E}\left[Z_{(n(1+c):n(1+c))}\right]\frac{\Gamma(k-1/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(k)} \\
&\approxeq (n(1+c))^{1/(1+\delta)}\frac{\Gamma(k-1/(1+\delta))}{\Gamma(k)} \\
&= (1+c)^{1/(1+\delta)}n^{1/(1+\delta)}\frac{\Gamma(k-1/(1+\delta))}{\Gamma(k)} \\
&\approxeq (1+c)^{1/(1+\delta)}\mathbb{E}\left[Y_{(n:n)}\right]\frac{\Gamma(k-1/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(k)} \\
&= (1+c)^{1/(1+\delta)}\mathbb{E}\left[Y_{(n-k+1:n)}\right]
\end{aligned}$$

Substituting into (12),

$$\begin{aligned}
\mathbb{E}\left[Y_{(n-k+1:n)}\cdot \mathbb{1}_{X_{(\alpha n:\alpha n)}\leq \beta Y_{(n-k+1:n)}}\right] &\approxeq \mathbb{E}\left[Y_{(n-k+1:n)}\right]\left((1+c)^{-(k-1/(1+\delta))}+\frac{1}{n}\right) \\
&\approxeq \mathbb{E}\left[Y_{(n-k+1:n)}\right](1+\alpha\beta^{-(1+\delta)})^{-(k-1/(1+\delta))}
\end{aligned}$$

since $c = a\beta^{-1(1+\delta)}$, proving the theorem. □

Theorem B.3.

$$\Pr\left[X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right] \approxeq (1+c)^{-k}.$$

Proof. Begin with

$$\begin{aligned}
\Pr\left[X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right] &= \int_1^\infty f_{(n-k+1:n)}(y)F_{(\alpha n:\alpha n)}(\beta y)\,dy \\
&= \int_1^{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}} f_{(n-k+1:n)}(y)F_{(\alpha n:\alpha n)}(\beta y)\,dy \\
&\quad + \int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^\infty f_{(n-k+1:n)}(y)F_{(\alpha n:\alpha n)}(\beta y)\,dy \qquad (15)
\end{aligned}$$

Observe that

$$\begin{aligned}
\int_1^{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}} f_{(n-k+1:n)}(y)F_{(\alpha n:\alpha n)}(\beta y)\,dy &\leq F_{(\alpha n:\alpha n)}\left(\beta\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}\right)F_{(n-k+1:n)}\left(\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}\right) \\
&\leq F_{(\alpha n:\alpha n)}\left(\beta\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}\right) \\
&\leq \left(1-\beta^{-(1+\delta)}\left(\frac{\ln n}{cn}\right)\right)^{\alpha n} \\
&\leq \exp\left(-\alpha\beta^{-(1+\delta)}\frac{\ln n}{cn}\right) \\
&= \frac{1}{n}
\end{aligned}$$

Next, we have

$$\int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^\infty f_{(n-k+1:n)}(y)F_{(\alpha n:\alpha n)}(\beta y)\,dy = \int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^\infty (1-(\beta y)^{-(1+\delta)})^{\alpha n} f_{(n-k+1:n)}(y)\,dy$$

By Lemma D.2, for $y \geq (cn/\ln n)^{1/(1+\delta)}$,

$$\left(1-(\beta y)^{-(1+\delta)}\right)^{\alpha n} \approxeq \left(1-y^{-(1+\delta)}\right)^{cn},$$

so

$$\begin{aligned}
&\int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^{\infty} f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy \\
&\qquad\approxeq \int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^{\infty} (1-y^{-(1+\delta)})^{cn} f_{(n-k+1:n)}(y)\, dy \\
&\qquad= (1+\delta)k\binom{n}{k} \int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^{\infty} (1-y^{-(1+\delta)})^{n(1+c)-k} (y^{-(1+\delta)})^{k} y^{-1}\, dy \\
&\qquad= \frac{\binom{n}{k}}{\binom{n(1+c)}{k}} \int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^{\infty} f_{(n(1+c)-k+1:n(1+c))}\, dy
\end{aligned}$$

From Lemma D.1, we have

$$F_{(n(1+c)-k+1:n(1+c))}\left(\left(\frac{cn}{\ln n}^{1/(1+\delta)}\right)\right) \leq \frac{\sqrt{k}}{n},$$

so

$$\begin{aligned}
\int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^{\infty} f_{(n(1+c)-k+1:n(1+c))}\, dy &= 1 - F_{(n(1+c)-k+1:n(1+c))}\left(\left(\frac{cn}{\ln n}^{1/(1+\delta)}\right)\right) \\
&\geq 1 - \frac{\sqrt{k}}{n} \\
&\approxeq 1.
\end{aligned}$$

Therefore,

$$\int_{\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}}^{\infty} f_{(n-k+1:n)}(y) F_{(\alpha n:\alpha n)}(\beta y)\, dy \approxeq \frac{\binom{n}{k}}{\binom{n(1+c)}{k}} \approxeq \frac{1}{(1+c)^k}$$

by Lemma D.8. By (15), this means

$$\Pr\left[X_{(\alpha n:\alpha n)} < \beta Y_{(n-k+1:n)}\right] \approxeq (1+c)^{-k}.$$

□

C Lemmas for the Equivalence Definition

Lemma C.1 (Transitivity)**.** *If $f(n) \approxeq_{a_1;n_1} g(n)$ and $g(n) \approxeq_{a_2;n_2} h(n)$, then $f(n) \approxeq h(n)$.*

Proof.

$$\begin{aligned}
\frac{f(n)}{h(n)} &= \frac{f(n)}{g(n)} \cdot \frac{g(n)}{h(n)} \\
&\leq \left(1 + \frac{a_1(\ln n)^2}{n}\right)\left(1 + \frac{a_2(\ln n)^2}{n}\right) \\
&\leq 1 + \frac{(a_1+a_2)(\ln n)^2}{n} + \frac{a_1 a_2 (\ln n)^4}{n^2} \\
&\leq 1 + \frac{(a_1+a_2+a_1a_2)(\ln n)^2}{n}
\end{aligned}$$

for all $n \geq \max(n_1, n_2)$, since $n \geq (\ln n)^2$. A symmetric argument holds for $h(n)/f(n)$. Thus, $f(n) \approxeq_{a_1+a_2+a_1a_2;\max(n_1,n_2)} h(n)$. $\square$

Lemma C.2 (Linearity)**.** *If $f_1(n) \approxeq_{a_1;n_1} g_1(n)$ and $f_1(n) \approxeq_{a_2;n_2} g_2(n)$, then $bf_1(n) + cf_2(n) \approxeq bg_1(n) + cg_2(n)$.*

Proof. By Lemma C.7,

$$\frac{bf_1(n) + cf_2(n)}{bg_1(n) + cg_2(n)} \leq \max\left(\frac{f_1(n)}{g_1(n)}, \frac{f_2(n)}{g_2(n)}\right) \leq \frac{\max(a_1, a_2)(\ln n)^2}{n}$$

for $n \geq \max(n_1, n_2)$. A symmetric argument holds for the reciprocal. Therefore,

$$bf_1(n) + cf_2(n) \approxeq_{\max(a_1,a_2);\max(n_1,n_2)} bg_1(n) + cg_2(n).$$

$\square$

Lemma C.3 (Integrals)**.** *If $f(x, n) \approxeq_{a;n_0} g(x, n)$, then*

$$\int f(x, n)\, dx \approxeq \int g(x, n)\, dx$$

Proof.

$$\frac{\int f(x, n)\, dx}{\int g(x, n)\, dx} = \frac{\int g(x, n)\frac{f(x,n)}{g(x,n)}\, dx}{\int g(x, n)\, dx} \leq \frac{\int g(x, n)\left(1 + \frac{a(\ln n)^2}{n}\right)\, dx}{\int g(x, n)\, dx} \leq 1 + \frac{a(\ln n)^2}{n}$$

for $n \geq n_0$. A symmetric argument holds for the repciprocal, proving the lemma. $\square$

Lemma C.4. *If $f_1(n) \approxeq_{a_1;n_1} g_1(n)$ and $f_2(n) \approxeq_{a_2;n_2} g_2(n)$, then*

$$f_1(n)f_2(n) \approxeq g_1(n)g_2(n).$$

Proof.

$$\begin{aligned}
\frac{f_1(n)f_2(n)}{g_1(n)g_2(n)} &= \frac{f_1(n)}{g_1(n)} \cdot \frac{f_2(n)}{g_2(n)} \\
&\leq \left(1 + \frac{a_1(\ln n)^2}{n}\right)\left(1 + \frac{a_2(\ln n)^2}{n}\right) \\
&\leq 1 + \frac{(a_1 + a_2)(\ln n)^2}{n} + \frac{a_1a_2(\ln n)^4}{n^2} \\
&\leq 1 + \frac{(a_1 + a_2 + a_1a_2)(\ln n)^2}{n}
\end{aligned}$$

for all $n \geq \max(n_1, n_2)$, since $n \geq (\ln n)^2$. A symmetric argument holds for the reciprocal. Thus, $f_1(n)f_2(n) \approxeq_{a_1+a_2+a_1a_2;\max(n_1,n_2)} g_1(n)g_2(n)$. $\square$

Lemma C.5. *If $f(n) \approxeq_{a;n_0} g(n)$, then $\frac{1}{f(n)} \approxeq \frac{1}{g(n)}$.*

Proof.

$$\frac{1/f(n)}{1/g(n)} = \frac{g(n)}{f(n)} \le 1 + \frac{a(\ln n)^2}{n}$$

for $n \ge n_0$. A symmetric argument holds for the reciprocal. ☐

Lemma C.6. *If $g_1(n) \le f(n) \le g_2(n)$, $g_1(n) \approxeq h(n)$, and $g_2(n) \approxeq h(n)$, then $f(n) \approxeq h(n)$.*

Proof.

$$\frac{f(n)}{h(n)} \le \frac{g_2(n)}{h(n)}$$

and

$$\frac{h(n)}{f(n)} \le \frac{h(n)}{g_1(n)},$$

proving the lemma by definition. ☐

Fact 4. For all $x \ge 1$, $\ln x \le x$ and $(\ln x)^2 \le x$.

Lemma C.7. *For $a, b, c, d > 0$, if $\frac{a}{b} \le \frac{c}{d}$, then*

$$\frac{a}{b} \le \frac{a+c}{b+d} \le \frac{c}{d}.$$

Proof. Since $\frac{a}{b} \le \frac{c}{d}$, $\frac{d}{b} \le \frac{c}{a}$. Therefore,

$$\frac{a+c}{b+d} = \frac{a}{b} \cdot \frac{1+c/a}{1+d/b} \ge \frac{a}{b} \cdot \frac{1+d/b}{1+d/b} = \frac{a}{b}.$$

Similarly,

$$\frac{a+c}{b+d} = \frac{c}{d} \cdot \frac{1+a/c}{1+b/d} \le \frac{c}{d} \cdot \frac{1+b/d}{1+b/d} = \frac{c}{d}.$$

☐

Lemma C.8.

$$\frac{a-(\ln n)^2/n}{b} \approxeq \frac{a}{b}$$

Proof.

$$\frac{\frac{a-(\ln n)^2/n}{b}}{\frac{a}{b}} = 1 - \frac{(\ln n)^2}{n} \le 1$$

$$\frac{\frac{a}{b}}{\frac{a-(\ln n)^2/n}{b}} = \frac{1}{1-\frac{(\ln n)^2}{an}} = 1 + \frac{\frac{(\ln n)^2}{an}}{1-\frac{(\ln n)^2}{an}} \le 1 + \frac{2(\ln n)^2}{an}$$

for $n \ge 16/a^4$. ☐

D Lemmas for Appendix B

Lemma D.1. *For $k \leq (1-c)\ln n$,*

$$F_{(n(1+c)-k+1:n(1+c))}\left(\left(\frac{cn}{\ln n}^{1/(1+\delta)}\right)\right) \leq \frac{\sqrt{k}}{n}.$$

Proof. We can write

$$\begin{aligned}
F_{(n(1+c)-k+1:n(1+c))}\left(\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}\right) &= \sum_{j=0}^{k-1}\binom{n(1+c)}{j}\left(1-\frac{\ln n}{cn}\right)^{n(1+c)-j}\left(\frac{\ln n}{cn}\right)^{j} \\
&\leq \sum_{j=0}^{k-1}\frac{(n(1+c))^j}{j!}\exp\left(-\left(\frac{\ln n}{cn}\right)(n(1+c)-j)\right)\left(\frac{\ln n}{cn}\right)^{j} \\
&= \sum_{j=0}^{k-1}\frac{1}{j!}\left(\frac{(1+c)\ln n}{c}\right)^{j}\exp\left(-\ln n\left(1+c^{-1}\left(1-\frac{j}{n}\right)\right)\right) \\
&= \frac{1}{n}\sum_{j=0}^{k-1}\frac{1}{j!}\left((1+c^{-1})\ln n\right)^{j}\left(\frac{1}{n}\right)^{c^{-1}\left(1-\frac{j}{n}\right)} \\
&\leq \frac{1}{n^2}+\frac{1}{n}\sum_{j=1}^{k-1}\frac{1}{\sqrt{2\pi j}}\left(\frac{e(1+c^{-1})\ln n}{j}\right)^{j}\left(\frac{1}{n}\right)^{c^{-1}\left(1-\frac{j}{n}\right)} \qquad (16)
\end{aligned}$$

by Stirling's approximation. The term

$$\left(\frac{e(1+c^{-1})\ln n}{j}\right)^{j}$$

is increasing whenever it's natural log,

$$j\left(1+\ln(1+c^{-1})+\ln\ln n-\ln j\right),$$

is increasing. This has derivative

$$1+\ln(1+c^{-1})+\ln\ln n-\ln j-1 = \ln(1+c^{-1})+\ln\ln n-\ln j \geq \ln\ln n-\ln j.$$

Thus, it is increasing for $j \leq \ln n$. For $j \leq (1-c)\ln n$, we have

$$\begin{aligned}
\left(\frac{e(1+c^{-1})\ln n}{j}\right)^{j} &\leq \left(\frac{e(1+c^{-1})\ln n}{(1-c)\ln n}\right)^{(1-c)\ln n} \\
&= \left(\frac{e(1+c^{-1})}{1-c}\right)^{(1-c)\ln n} \\
&= \exp\left(1+\ln(1+c^{-1})-\ln(1-c)\right)^{(1-c)\ln n} \\
&= \exp\left(\ln n\right)^{(1+\ln(1+c^{-1})-\ln(1-c))(1-c)} \\
&= n^{(1+\ln(1+c^{-1})-\ln(1-c))(1-c)} \\
&\leq n^{(1+c^{-1}+c+c^2)(1-c)} \\
&= n^{1+c^{-1}+c+c^2-c-1-c^2-c^3} \\
&= n^{c^{-1}-c^3} \\
&\leq n^{c^{-1}(1-j/n)}
\end{aligned}$$

for sufficiently large n, since $j \leq (1-c)\ln n$. Combining this with (16), we have

$$\begin{aligned} F_{(n(1+c)-k+1:n(1+c))}\left(\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}\right) &\leq \frac{1}{n^2} + \frac{1}{n\sqrt{2\pi}} \sum_{j=1}^{k-1} \frac{1}{\sqrt{j}} \\ &\leq \frac{1}{n^2} + \frac{1}{n\sqrt{2\pi}} \left(1 + \int_1^k \frac{1}{\sqrt{j}}\, dj\right) \\ &\leq \frac{1}{n^2} + \frac{\sqrt{k}}{n\sqrt{2\pi}} \\ &\leq \frac{\sqrt{k}}{n} \end{aligned}$$

□

Lemma D.2. *For $y \geq (cn/\ln n)^{1/(1+\delta)}$,*

$$\left(1-(\beta y)^{-(1+\delta)}\right)^{\alpha n} \approxeq \left(1-y^{-(1+\delta)}\right)^{cn},$$

Proof. We know that $1-(\beta y)^{-(1+\delta)} \geq (1-y^{-(1+\delta)})^{\beta^{-(1+\delta)}}$ from the Taylor expansion, giving us

$$\left(1-(\beta y)^{-(1+\delta)}\right)^{\alpha n} \geq \left((1-y^{-(1+\delta)})^{\beta^{-(1+\delta)}}\right)^{\alpha n} = (1-y^{-(1+\delta)})^{cn}.$$

On the other hand, for $y \geq (cn/\ln n)^{1/(1+\delta)}$,

$$\left(1-(\beta y)^{-(1+\delta)}\right)^{\alpha n} \leq \exp\left(-cy^{-(1+\delta)}n\right) \leq \frac{\left(1-y^{-(1+\delta)}\right)^{cn}}{1-cny^{-2(1+\delta)}} \leq \frac{\left(1-y^{-(1+\delta)}\right)^{cn}}{1-\frac{(\ln n)^2}{cn}} \approxeq \left(1-y^{-(1+\delta)}\right)^{cn}.$$

□

Lemma D.3. *For $k \leq ((1-c^2)\ln n)/2$,*

$$F_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}\left(\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}\right) \leq \frac{\sqrt{k}}{n},$$

Proof. We begin with

$$\begin{aligned} &F_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}\left(\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}\right) \\ &= \sum_{\ell=0}^{j} \binom{\beta^{1+\delta}n(1+c)}{\ell} \left(1-\frac{\ln n}{\alpha n}\right)^{\beta^{1+\delta}n(1+c)-\ell} \left(\frac{\ln n}{\alpha n}\right)^{\ell} \\ &\leq \sum_{\ell=0}^{j} \frac{(\beta^{1+\delta}n(1+c))^{\ell}}{\ell!} \exp\left(-\left(\frac{\ln n}{\alpha n}\right)(\beta^{1+\delta}n(1+c)-\ell)\right) \left(\frac{\ln n}{\alpha n}\right)^{\ell} \\ &= \sum_{\ell=0}^{j} \frac{1}{\ell!} \left(\frac{\beta^{1+\delta}(1+c)\ln n}{c}\right)^{\ell} \exp\left(-\ln n\left(1+c^{-1}\left(1-\frac{\ell}{n}\right)\right)\right) \\ &= \frac{1}{n} \sum_{\ell=0}^{j} \frac{1}{\ell!} \left(\beta^{1+\delta}(1+c^{-1})\ln n\right)^{\ell} \left(\frac{1}{n}\right)^{c^{-1}\left(1-\frac{\ell}{n}\right)} \\ &\leq \frac{1}{n^2} + \frac{1}{n} \sum_{\ell=1}^{j} \frac{1}{\sqrt{2\pi\ell}} \left(\frac{e\beta^{1+\delta}(1+c^{-1})\ln n}{\ell}\right)^{\ell} \left(\frac{1}{n}\right)^{c^{-1}\left(1-\frac{\ell}{n}\right)} \qquad (17) \end{aligned}$$

We apply a similar argument as in Lemma D.1, showing that for $\ell \leq ((1-c^2)/2)\ln n$,

$$\begin{aligned}
\left(\frac{e\beta^{1+\delta}(1+c^{-1})\ln n}{\ell}\right)^{\ell} &\leq \left(\frac{ec^{-1}(1+c^{-1})\ln n}{(1-c)\ln n}\right)^{((1-c^2)/2)\ln n} \\
&= \left(\frac{ec^{-1}(1+c^{-1})}{1-c}\right)^{(1-c)\ln n} \\
&= \exp\left(1+\ln c^{-1}+\ln(1+c^{-1})-\ln(1-c)\right)^{((1-c^2)/2)\ln n} \\
&= \exp\left(\ln n\right)^{(1+\ln c^{-1}+\ln(1+c^{-1})-\ln(1-c))((1-c^2)/2)} \\
&= n^{(1+\ln c^{-1}+\ln(1+c^{-1})-\ln(1-c))((1-c^2)/2)} \\
&\leq n^{(1+(c^{-1}-1)+c^{-1}+c+c^2)((1-c^2)/2)} \\
&= n^{(2c^{-1}+c+c^2)((1-c^2)/2)} \\
&\leq n^{(2c^{-1}+2c)((1-c^2)/2)} && (c \leq 1) \\
&= n^{c^{-1}(1+c^2)(1-c^2)} \\
&= n^{c^{-1}(1-c^4)} \\
&\leq n^{c^{-1}(1-\ell/n)}
\end{aligned}$$

for sufficiently large n. This gives us

$$F_{(\beta^{1+\delta}n(1+c)-j:\beta^{1+\delta}n(1+c))}\left(\left(\frac{\alpha n}{\ln n}\right)^{1/(1+\delta)}\right) \leq \frac{1}{n^2}+\frac{1}{n}\sum_{\ell=1}^{j}\frac{1}{\sqrt{2\pi\ell}} \leq \frac{\sqrt{k}}{n},$$

□

Lemma D.4. *For $0 < a < b$ and $c > 0$,*

$$\frac{a+c}{b+c} > \frac{a}{b}$$

Proof.

$$\frac{a+c}{b+c} = \frac{a(1+c/a)}{b(1+c/b)} > \frac{a(1+c/b)}{b(1+c/b)} = \frac{a}{b}$$

□

Lemma D.5. *For $0 \leq y \leq a_1 \cdot \frac{\ln n}{n}$ and $|z| \leq a_2 \ln n$,*

$$|(1-y)^z - (1-yz)| = O\left(\frac{1}{n}\right)$$

Proof. By Taylor's theorem,

$$f(y) = (1-y)^z = 1 - yz \pm \frac{f''(\varepsilon)}{2}y^2$$

for some $0 \leq \varepsilon \leq y$. Note that

$$f''(\varepsilon) = z(z-1)(1-\varepsilon)^{z-2} \leq |z(z-1)|\exp(-\varepsilon(z-2)) \leq |z(z-1)|\exp(\varepsilon|z-2|).$$

Since $\varepsilon \leq y \leq a_1 \cdot \frac{\ln n}{n}$ and $|z| \leq a_2 \ln n$,

$$|z(z-1)| \exp(\varepsilon|z-2|) \leq a_2^2 (\ln n)^2 n^{-2} n^{a_1|z-2|/n}.$$

This gives us

$$\frac{f''(\varepsilon)}{2} y^2 \leq \frac{a_2^2 (\ln n)^2 n^{-2} n^{a_1|z-2|/n}}{2} a_1^2 (\ln n)^2 n^{-2} \leq a_1^2 a_2^2 (\ln n)^4 n^{-(2-a_1(a_2 \ln n+2)/n)}.$$

Using $\ln n = n^{\ln\ln n/\ln n}$, this is

$$\frac{a_1^2 a_2^2}{n} n^{-(1-a_1(a_2 \ln n+2)/n - 4\ln\ln n/\ln n)}.$$

For sufficiently large n, $a_1(a_2 \ln n + 2)/n + 4\ln\ln n/\ln n \leq 1$, so

$$\frac{a_1^2 a_2^2}{n} n^{-(1-a_1(\ln n+2)/n - 4\ln\ln n/\ln n)} \leq \frac{a_1^2 a_2^2}{n} = O(1/n),$$

which proves the lemma. □

Lemma D.6.

$$F_{(an:an)}\left(b\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}\right) \leq n^{-ab^{-(1+\delta)}/c}$$

Proof.

$$\begin{aligned} F_{(an:an)}\left(b\left(\frac{cn}{\ln n}\right)^{1/(1+\delta)}\right) &= \left(1 - b^{-(1+\delta)}\frac{\ln n}{cn}\right)^{an} \\ &\leq \exp\left(-\frac{ab^{-(1+\delta)}}{c}\ln n\right) \\ &= n^{-ab^{-(1+\delta)}/c} \end{aligned}$$

□

Lemma D.7.

$$\mathbb{E}\left[Z_{(Cn-\ln n+1:Cn)}\right] \geq \left(\frac{Cn}{\ln n}\right)^{1/(1+\delta)}$$

for $C \geq 1$ and sufficiently large n.

Proof.

$$
\begin{aligned}
\mathbb{E}\left[Z_{(Cn-\ln n+1:Cn)}\right] &= \mathbb{E}\left[Z_{(Cn:Cn)}\right]\prod_{j=1}^{\ln n-1}\left(1-\frac{1}{(1+\delta)j}\right) \\
&= \mathbb{E}\left[Z_{(Cn:Cn)}\right]\prod_{j=1}^{\ln n-1}\left(\frac{(1+\delta)j-1}{(1+\delta)j}\right) \\
&= \mathbb{E}\left[Z_{(Cn:Cn)}\right]\prod_{j=1}^{\ln n-1}\left(\frac{j-1/(1+\delta)}{j}\right) \\
&= \mathbb{E}\left[Z_{(Cn:Cn)}\right]\frac{\Gamma(\ln n-1/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(\ln n)} \\
&\geq \Gamma\left(\frac{\delta}{1+\delta}\right)(Cn)^{1/(1+\delta)}\frac{\Gamma(\ln n-1/(1+\delta))}{\Gamma(\delta/(1+\delta))\Gamma(\ln n)} && \text{(by Lemma D.10)} \\
&= (Cn)^{1/(1+\delta)}\frac{\Gamma(\ln n-1/(1+\delta))}{\Gamma(\ln n)} \\
&\geq \left(\frac{Cn}{\ln n}\right)^{1/(1+\delta)}\left(1+\frac{\frac{1}{1+\delta}\cdot\frac{1+2\delta}{1+\delta}}{\ln n}-O\left(\frac{1}{(\ln n)^2}\right)\right) && \text{(by [20])} \\
&\geq \left(\frac{Cn}{\ln n}\right)^{1/(1+\delta)} && \text{(for sufficiently large } n\text{)}
\end{aligned}
$$

□

Lemma D.8. *For $k = O(\ln n)$,*

$$\frac{\binom{n}{k}}{\binom{n(1+c)}{k}} \approx \frac{1}{(1+c)^k}.$$

Proof.

$$\frac{\binom{n}{k}}{\binom{n(1+c)}{k}} = \frac{n(n-1)\cdots(n-k+1)}{n(1+c)(n(1+c)-1)\cdots(n(1+c)-k+1)}.$$

Each term $(n-j)/(n(1+c)-j)$ is between $1/(1+c)$ and $(1-j/n)/(1+c)$. Therefore, the entire product is at least

$$\prod_{j=0}^{k-1}\frac{1}{1+c}\left(1-\frac{j}{n}\right) = \frac{1}{(1+c)^k}\prod_{j=0}^{k-1}\left(1-\frac{j}{n}\right) \geq \frac{1}{(1+c)^k}\left(1-\frac{k^2}{n}\right)$$

and at most $1/(1+c)^k$. This means that

$$\frac{1}{(1+c)^k} \geq \frac{\binom{n}{k}}{\binom{n(1+c)}{k}} \geq \frac{1}{(1+c)^k}\left(1-\frac{(\ln n)^2}{n}\right) \approx \frac{1}{(1+c)^k}$$

□

Lemma D.9. *For $0 < z < 1$, and $y \geq 0$,*

$$(1+y)^z < 1+yz.$$

Proof. Let $w = z^{-1}$. Then, the lemma is true if and only if for $w > 1$,

$$1 + y < \left(1 + \frac{y}{w}\right)^w.$$

Note that for $w = 1$, we have equality. We will show that the function

$$f(w) = \left(1 + \frac{y}{w}\right)^w$$

has nonnegative derivative for $w \geq 1$. This is equivalent to showing the same for its log, which is

$$\begin{aligned}
\frac{d}{dw} \log f(w) &= \frac{d}{dw} w \log\left(1 + \frac{y}{w}\right) \\
&= \log\left(1 + \frac{y}{w}\right) + \frac{w}{1 + \frac{y}{w}} \cdot \left(-\frac{y}{w^2}\right) \\
&= \log\left(1 + \frac{y}{w}\right) - \frac{\frac{y}{w}}{1 + \frac{y}{w}}
\end{aligned}$$

Let $x = 1 + \frac{y}{w}$. Then, the lemma is true if for $x > 1$,

$$\begin{aligned}
\log(x) - \frac{x-1}{x} &> 0 \\
x \log(x) &> x - 1
\end{aligned}$$

Both are 0 at $x = 1$, but the left hand side has derivative $1 + \log(x)$ while the right hand side has derivative 1, so left hand side will be strictly larger than the right hand side for $x > 1$. $\square$

Lemma D.10.

$$\mathbb{E}\left[Z_{(m:m)}\right] \approx \Gamma\left(\frac{\delta}{1+\delta}\right) m^{1/(1+\delta)}.$$

Also,

$$\mathbb{E}\left[Z_{(m:m)}\right] \geq \Gamma\left(\frac{\delta}{1+\delta}\right) m^{1/(1+\delta)}.$$

Proof. From [17], we have

$$\mathbb{E}\left[Z_{(m:m)}\right] = \frac{\Gamma(m+1)\Gamma\left(1 - \frac{1}{1+\delta}\right)}{\Gamma\left(m + \frac{\delta}{1+\delta}\right)}.$$

By [20],

$$\frac{\Gamma(m+1)}{\Gamma\left(m + \frac{\delta}{1+\delta}\right)} = m^{1/(1+\delta)} \left(1 + \frac{\left(\frac{1}{1+\delta}\right)\left(\frac{\delta}{1+\delta}\right)}{2m} + O\left(\frac{1}{m^2}\right)\right) \geq m^{1/(1+\delta)}$$

This means

$$\Gamma\left(\frac{\delta}{1+\delta}\right) m^{1/(1+\delta)} \leq \mathbb{E}\left[Z_{(m:m)}\right] \leq \Gamma\left(\frac{\delta}{1+\delta}\right) m^{1/(1+\delta)} \left(1 + O\left(\frac{1}{n}\right)\right),$$

so

$$\Gamma\left(\frac{\delta}{1+\delta}\right) m^{1/(1+\delta)} \approx \mathbb{E}\left[Z_{(m:m)}\right].$$

$\square$

Lemma D.11 ([17], Formula 1)**.**

$$\mathbb{E}\left[Z_{(m-k:m)}\right] = \left(1 - \frac{1}{k(1+\delta)}\right) \mathbb{E}\left[Z_{(m-k+1:m)}\right]$$

E Lemmas and Proofs for Section 3

Proof of Theorem 3.1. To proceed, we need some notation. Let L be the event that $X_{(n-1:n)} \geq T \cap Y_{(n-1:n)} \geq T$ (the samples are "large"). Let G be the event that $b(X_{(n:n)}) < Y_{(n-1:n)}$, meaning G is the event that the policy has an effect. Let D be the random variable $X_{(n:n)} - Y_{(n-1:n)}$. We want to show that $\mathbb{E}[D|G] > 0$. To do so, we observe that by Lemma E.1, is sufficient to show that $\mathbb{E}[D|L] > \frac{\Pr[\overline{L}]}{\Pr[L]}$. By Lemma E.2, we know that $\Pr[\overline{L}] \leq 2nF(T)^{n-1}$. To complete the proof, we need to show that $\mathbb{E}[D|L]$ is large, which we do via Lemma E.3.

Since $\Pr[L] \geq 1 - 2nF(T)^{n-1}$, there exists N_1 such that for all $n \geq N_1$, $\Pr[L] \geq 1/2$. Using Lemma E.3, if $n \geq N_1$, it is sufficient to have

$$\begin{aligned}
\mathbb{E}[D|L] &> \frac{\Pr[\overline{L}]}{\frac{1}{2}} \\
K(F(T)+\eta)^{n-1} &> 4nF(T)^{n-1} \\
\left(1 + \frac{\eta}{F(T)}\right)^n &> \frac{4n}{K} \\
n \log\left(1 + \frac{\eta}{F(T)}\right) &> \log n + \log\left(\frac{4}{K}\right) \\
\sqrt{n} \log\left(1 + \frac{\eta}{F(T)}\right) &> 2 && (n \geq 4/K, \text{ using } \sqrt{n} > \log n) \\
n &> 4\left(\log\left(1 + \frac{\eta}{F(T)}\right)\right)^{-2} = N_2
\end{aligned}$$

Thus, for $n > \max\{N_1, N_2, 4/K\}$, $\mathbb{E}[D|L] > \frac{\Pr[\overline{L}]}{\Pr[L]}$, which by Lemma E.3 implies that $\mathbb{E}[D|G] > 0$. This completes the proof of Theorem 3.1. □

Lemma E.1. *If $L \Rightarrow G$ and $D \geq -1$, then $\mathbb{E}[D|L] > \frac{\Pr[\overline{L}]}{\Pr[L]}$ implies $\mathbb{E}[D|G] > 0$.*

Proof.

$$\begin{aligned}
\mathbb{E}[D|G] &= \mathbb{E}[D \cdot \mathbb{1}_L|G] + \mathbb{E}\left[D \cdot \mathbb{1}_{\overline{L}}|G\right] \\
&= \frac{\mathbb{E}[D \cdot \mathbb{1}_L \cdot \mathbb{1}_G] + \mathbb{E}\left[D \cdot \mathbb{1}_{\overline{L}} \cdot \mathbb{1}_G\right]}{\Pr[G]} \\
&= \frac{\mathbb{E}[D \cdot \mathbb{1}_L] + \mathbb{E}\left[D \cdot \mathbb{1}_{\overline{L}} \cdot \mathbb{1}_G\right]}{\Pr[G]} && (L \Rightarrow G) \\
&\geq \frac{\mathbb{E}[D \cdot \mathbb{1}_L] - \mathbb{E}\left[\mathbb{1}_{\overline{L}} \cdot \mathbb{1}_G\right]}{\Pr[G]} && (D \geq -1) \\
&\geq \frac{\mathbb{E}[D \cdot \mathbb{1}_L] - \mathbb{E}\left[\mathbb{1}_{\overline{L}}\right]}{\Pr[G]} && (\mathbb{1}_G \leq 1) \\
&= \frac{\mathbb{E}[D|L]\Pr[L] - \Pr\left[\overline{L}\right]}{\Pr[G]}
\end{aligned}$$

$$
\begin{aligned}
&\frac{\mathbb{E}\left[D|L\right]\Pr\left[L\right]-\Pr\left[\overline{L}\right]}{\Pr\left[G\right]}>0\\
\iff &\mathbb{E}\left[D|L\right]\Pr\left[L\right]-\Pr\left[\overline{L}\right]>0\\
\iff &\mathbb{E}\left[D|L\right]>\frac{\Pr\left[\overline{L}\right]}{\Pr\left[L\right]}
\end{aligned}
$$

□

Lemma E.2. *For $X_{(n-1:n)}$, $Y_{(n-1:n)}$ order statistics from a distribution with support on $[0,1]$,*

$$\Pr\left[X_{(n-1:n)}\geq T\cap Y_{(n-1:n)}\geq T\right]\leq 2nF(T)^{n-1}.$$

Proof.

$$
\begin{aligned}
\Pr\left[X_{(n-1:n)}\geq T\cap Y_{(n-1:n)}\geq T\right]&=\Pr\left[X_{(n-1:n)}\geq T\right]\Pr\left[Y_{(n-1:n)}\geq T\right]\\
&=(1-F_{(n-1)}(T))^2\\
&=(1-nF(T)^{n-1}(1-F(T))-F(T)^n)^2\\
&=(1-nF(T)^{n-1}+(n-1)F(T)^n)^2\\
&\geq(1-nF(T)^{n-1})^2\\
&\geq 1-2nF(T)^{n-1}
\end{aligned}
$$

$$\Pr\left[\overline{L}\right]=1-\Pr\left[L\right]\leq 2nF(T)^{n-1}$$

□

Lemma E.3. *There exist constants $\eta>0$ and $K>0$ such that* $\mathbb{E}\left[D|L\right]\geq K(F(T)+\eta)^{n-1}$

Proof. First, let f_Z and F_Z be the pdf and cdf respectively of $Y|Y\geq T$, i.e. $F_Z(x)=\frac{F(x)-F(T)}{1-F(T)}$ and $f_Z=F_Z'$. Note that

$$
\begin{aligned}
\mathbb{E}\left[D|L\right]&=\mathbb{E}\left[X_{(n:n)}-Y_{(n-1:n)}|L\right]\\
&=\mathbb{E}\left[X_{(n:n)}|X_{(n-1:n)}\geq T\right]-\mathbb{E}\left[Y_{(n-1:n)}|Y_{(n-1:n)}\geq T\right]\\
&=\mathbb{E}\left[Y_{(n:n)}|Y_{(n-1:n)}\geq T\right]-\mathbb{E}\left[Y_{(n-1:n)}|Y_{(n-1:n)}\geq T\right]\\
&=\mathbb{E}\left[Y_{(n:n)}-Y_{(n-1:n)}|Y_{(n-1:n)}\geq T\right]
\end{aligned}
$$

Let M be a random variable corresponding to the number of samples from $Y_1,\ldots,Y_n$ that are larger than T. We can rewrite this as

$$
\begin{aligned}
\mathbb{E}\left[D|L\right]&=\sum_{m=2}^{M}\mathbb{E}\left[Y_{(n:n)}-Y_{(n-1:n)}|Y_{(n-1:n)}\geq T,M=m\right]\Pr\left[M=m|Y_{(n-1:n)}\geq T\right]\\
&=\sum_{m=2}^{M}\mathbb{E}\left[Y_{(n:n)}-Y_{(n-1:n)}|M=m\right]\Pr\left[M=m|Y_{(n-1:n)}\geq T\right]\qquad(M\geq 2\implies Y_{(n-1:n)}\geq T)
\end{aligned}
$$

Conditioning on $M=m$, $Y_{(n:n)}$ and $Y_{(n-1:n)}$ have the same distributions as $Z_{(m:m)}$ and $Z_{(m-1:m)}$ respectively, where $Z_{(k:m)}$ is the kth order statistic of random variables $Z_1,Z_2,\ldots,Z_m$ drawn from the distribution with cdf F_Z. We will use $F_{Z,(k:m)}$ to denote the cdf of $Z_{(k:m)}$. Thus, $\mathbb{E}\left[Y_{(n:n)} - Y_{(n-1:n)}|M = m\right] = \mathbb{E}\left[Z_{(m:m)} - Z_{(m-1:m)}\right]$. Using an analysis similar to that of [16],

$$\begin{aligned}
\mathbb{E}\left[Z_{(m:m)} - Z_{(m-1:m)}\right] &= \int_T^1 (1 - F_{Z,(m:m)}(x)) - (1 - F_{Z,(m-1:m)}(x))\, dx \\
&= \int_T^1 F_{Z,(m-1:m)} - F_{Z,(m:m)}(x)\, dx \\
&= \int_T^1 \binom{m}{m-1} F_Z(x)^{m-1}(1 - F_Z(x))\, dx \\
&\geq \int_T^1 F_Z(x)^{m-1}(1 - F_Z(x))\, dx
\end{aligned}$$

Choose $\eta \in (0, 1 - F(T))$ and $\eta' \in (\eta, 1 - F(T))$. Let $r = F_Z^{-1}(F(T) + \eta)$ and $r' = F_Z^{-1}(F(T) + \eta')$. Note that $T < r < r' < 1$ because otherwise F_Z would have infinite slope at r or r', which is impossible because f_Z is continuous over a compact set and therefore has a finite maxmium. Moreover, it must be the case that $F(T) < 1$ because by assumption, $\sup_{x:f(x)>0} = 1$. If $F(T)$ were 1, this would imply that $\sup_{x:f(x)>0} = T < 1$, which is a contradiction.

$$\begin{aligned}
\int_T^1 F_Z(x)^{m-1}(1 - F_Z(x))\, dx &\geq \int_r^1 F_Z(x)^{m-1}(1 - F_Z(x)) \\
&\geq \int_r^1 F_Z(r)^{m-1}(1 - F_Z(x)) \\
&= (F(T) + \eta)^{m-1} \int_r^1 1 - F_Z(x)\, dx \\
&\geq (F(T) + \eta)^{n-1} \int_r^1 1 - F_Z(x)\, dx \\
&\geq (F(T) + \eta)^{n-1} \int_r^{r'} 1 - F_Z(x)\, dx \\
&\geq (F(T) + \eta)^{n-1} \int_r^{r'} 1 - F_Z(r')\, dx \qquad (F_Z(x) \leq F_Z(r') \text{ for } x \leq r') \\
&= (F(T) + \eta)^{n-1}(r' - r)(1 - (F(T) + \eta')) \\
&= (F(T) + \eta)^{n-1}[F_Z^{-1}(F(T) + \eta') - F_Z^{-1}(F(T) + \eta)](1 - F(T) - \eta') \\
&= K(F(T) + \eta)^{n-1}
\end{aligned}$$

where $K = [F_Z^{-1}(F(T) + \eta') - F_Z^{-1}(F(T) + \eta)](1 - F(T) - \eta')$. Since this is independent of m, we have

$$\begin{aligned}
\mathbb{E}\left[D|L\right] &= \sum_{m=2}^{n} \mathbb{E}\left[Y_{(n:n)} - Y_{(n-1:n)}|M = m\right] \Pr\left[M = m|Y_{(n-1:n)} \geq T\right] \\
&\geq \sum_{m=2}^{n} K(F(T) + \eta)^{n-1} \Pr\left[M = m|Y_{(n-1:n)} \geq T\right] \\
&= K(F(T) + \eta)^{n-1}
\end{aligned}$$

$\square$