Fairness

Jeff Edmonds *
York University
jeff@cse.yorku.ca

Karan Singh
York University
singkara@cse.yorku.ca

Ruth Urner
York University
ruth@cse.yorku.ca

March 16, 2021

JEFF IS EDITING THIS
Please don't change
$\widehat{t}_g$

1 New Fun

Theorem: Here we only consider $\widehat{x}_g$ that are non-extreme performance scores, i.e. $r(\widehat{x}_g) = 2$. Let $\langle t_a, e_a, \widehat{x}_a\rangle$ and $\langle t_b, e_b, \widehat{x}_b\rangle$ denote the talent, environment, and measured performance scores for groups A and B. Let $\widehat{t}_a$ and $\widehat{t}_b$ denote desired talent thresholds for acceptance of the person from each group. Let $\widehat{x}_g$ denote the measured performance of both the A and B person that you are deciding between. Below we define the relationship $\widehat{t}_b = F(\widehat{t}_a)$ so that

$$Pr(t_b{=}\widehat{t}_b|\widehat{x}_b{=}\widehat{x}_g) = Pr(t_a{=}\widehat{t}_a|\widehat{x}_a{=}\widehat{x}_g) \text{ (or equivalently)}$$
$$Pr(t_b{\geq}\widehat{t}_b|\widehat{x}_b{=}\widehat{x}_g) = Pr(t_a{\geq}\widehat{t}_a|\widehat{x}_a{=}\widehat{x}_g)$$

- Suppose the talent distribution T is uniform.
  Suppose the measure of performance is the sum $\widehat{x}_g = T_g{+}E_g$ of the talent and environment for $g \in \{A, B\}$.
  Suppose group A is advantaged by having its environment distribution k more than that for group B, i.e. $E_a = E_b{+}k$.
  $\rightarrow$ Then $\widehat{t}_b = \widehat{t}_a{+}k$.
  The effect of this is that if you take the same number of randomly chosen A and B people and sort each group according to their talent, then the talent of the i^{th} B person will be k higher then the i^{th} A person.

- If we temporarily relax the restriction that $\widehat{x}_g$ that is non-extreme performance score, i.e. $r(\widehat{x}_g) = 2$, then this result generalizes to
  $\rightarrow$ Then $\widehat{t}_b = \widehat{t}_a{+}\frac{r(\widehat{x}_g)}{2}k$
  ** Jeff has not check this one ***
  and $Exp(t_b|\widehat{x}_b{=}\widehat{x}_g) = Exp(t_a|\widehat{x}_a{=}\widehat{x}_g) + \frac{r(\widehat{x}_g)}{2}k$.

*Thanks to Frances for her encouragement.

- Generalize, to allowing $E_a = d \cdot E_b + k$, i.e. the same distribution whose standard deviation has been scaled by d and mean raised by k.
  $\rightarrow$ In this case, $\widehat{t_b} = \frac{\widehat{t_a} + k + (d-1)x}{d}$.

- We can prove the same result when both E_a and E_b are arbitrary distributions. We will model this by defining distribution $E_a = E_a(E_b)$, i.e. we randomly choose a value e'_a from the distribution E_b and then set $e_a = E_a(e'_a)$ where the later E_a denotes an arbitrary function.
  $\rightarrow$ In this case, $\widehat{t_b} = x - E_a^{-1}(x - \widehat{t_a})$.

- The only restrictions now are that T is uniform, $\widehat{x}_g = t_g + e_g$, and $r(\widehat{x}_g) = 2$. With this we also get
$$Exp(t_b|\widehat{x}_b = \widehat{x}_g) - Exp(t_a|\widehat{x}_a = \widehat{x}_g) = Exp(e_a) - Exp(e_b)$$

- Suppose now that we are allowed to produce the measured performance score not just by $\widehat{x}_g = t_g + e_g$, but more generally by $\widehat{x}_g = X(t_g, e_g)$ for an arbitrary increasing function X. Let $e_g = E_{\widehat{x}_g}(t_g)$ and $t_g = T_{\widehat{x}_g}(e_g)$ be two inverse functions.
  Finally, let us allow the talent distribution T to be arbitrary by choosing p uniformly from $[0, 1]$ and setting $t_g = T(p)$ for an arbitrary function T. Note that $Pr[t_g \leq T(p)] = p$.
  $\rightarrow$ This most general result becomes $\widehat{t_b} = T_{\widehat{x}_g}(E_a^{-1}(E_{\widehat{x}_g}(\widehat{t_a})))$.

- The only remaining cases to consider are when the range of talent is smaller than the range of environment so that the performance score $\widehat{x}_g$ is extreme in both directions, i.e. $r(\widehat{x}_g) = 0$.
  For here, we will assume that the talent distributions are the same T and uniform, the environment distributions $E_a = E_b + k$ are the same but shifted, and the measured performance score is $\widehat{x}_g = t_g + e_g$. In this case, not only may there not be a gap between the expected talent of person B over person A, but the direction may be reversed.
  $\rightarrow$ If E_b is *sub-exponential*, then $Exp(t_b|\widehat{x}_b = \widehat{x}_g) \geq Exp(t_a|\widehat{x}_a = \widehat{x}_g)$.
  $\rightarrow$ Otherwise it could be that for all $\widehat{x}_g$ (with $r(\widehat{x}_g) = 0$) $Exp(t_b|\widehat{x}_b = \widehat{x}_g) < Exp(t_a|\widehat{x}_a = \widehat{x}_g)$.

==========

Suppose our talent distribution is defined only within the range $[t^{min}, t^{max}]$ and the environment distribution within $[e_b^{min}, e_b^{max}]$. Suppose we condition on the fact that the performance score is given by $\widehat{x}_b = t_b + e_b$ is fixed to some value $\widehat{x}_g$. Rearranging and considering the environment range gives that $t_b = x - e_b \in [x - e_b^{max}, x - e_b^{min}]$. If $\widehat{x}_g$ is an extreme low value, then low range $x - e_b^{max}$ is smaller than the talent low range t^{min} and hence the bound t^{min} kicks in. Similarly, if $\widehat{x}_g$ is an extreme height value, then the high range $x - e_b^{min}$ is trumped by t^{max}. We define $r(\widehat{x}_g)$ to be the number of endpoint for which this does not happen. For example, if the range $[t^{min}, t^{max}]$ on talent is much wider than that on environment and $\widehat{x}_g$ is not an extreme value, then the range $[x - e_b^{max}, x - e_b^{min}]$ is a subset of the range $[t^{min}, t^{max}]$. In this case we say $r(x) = 2$.

==========

Proof: The proof will first assume that T is uniform.

We denote the talent of a random group $g \in \{A, B\}$ person by t_g which is drawn from the distribution T. If T is a continuous random variable, then $Pr(t_g = t) = 0$ for any exact value of t. The standard way of dealing with this is to define the density function $P_T(t)$ so that $Pr(t_g \in [t, t+\delta t]) = \delta t \cdot P_T(t)$. Similarly, we denote his environment value by e_g which is drawn from the distribution E_g with density function $P_{E_g}(e)$.

Because these two random variables are independent, we can define the cross density function $P_g(t, e) = P_T(t) \times P_{E_g}(e)$ so that $Pr(t_g \in [t, t+\delta t] \ \& \ e_g \in [e, e+\delta e]) = \delta t \cdot P_T(t) \times \delta e \cdot P_{E_g}(t) = \delta t \delta e \cdot P_g(t, e)$. Imagine raising a third dimension coming out of the page on the $\langle T, E \rangle$ rectangle in the figure, so that its height at location $\langle t, e \rangle$ is $P_g(t, e)$.

Our goal is to compute $Pr(t_g \in [\widehat{t}_g, \widehat{t}_g + \delta t] | x_g = \widehat{x}_g)$. Using the standard formula $Pr(t_g \in [\widehat{t}_g, \widehat{t}_g + \delta t] \& x_g = \widehat{x}_g) / Pr(x_g = \widehat{x}_g)$ is awkard because the later is zero unless we allow x_g to fall into some infintesemally small range. This performance is given by $x = t+e$ or more generally by $x = X(t, e)$. These equations narrow our $\langle T, E \rangle$ rectangle of possibilities to the 1-dimensional (possibly curved) line defined as $L = \{\langle t, e \rangle \, | \widehat{x}_g = X(t, e)\}$. Let $|L|$ denote its length and $|P_g(L)|$ the area over this line and under the surface $P_g(t, e)$. In order to translate from 2 to 1-dims, define L' to be 2-dim curved road with width ϵ along L and total area $|L'| = \epsilon |L|$. For all of these points $\langle t, e \rangle$ in L', we have that $X(t_g, e_g)$ is infintesimally close to $\widehat{x}_g$. Let use denote this with $x_g \approx \widehat{x}_g$. Because $P_g(t, e)$ the our probability density function, $Pr(x_g \approx \widehat{x}_g)$ is equal to the volume $|P_g(L')|$ over L' and under the surface $P_g(t, e)$. As we do in calculus, ϵ is small enough that $P_g(t, e)$ does vary much across the width of L'. Hence, its volume $|P_g(L')|$ is given by $\epsilon |P_g(L)|$. The next step is define S to be the portion of the line L for which $t \in [\widehat{t}_g, \widehat{t}_g + \delta t]$. and to define S' to be the corresponding portion within L', i.e. this line sigment when given ϵ width. Let $|S|$ and $|S'| = \epsilon |S|$ denote their respective length and area. Let $|P_g(S)|$ and $|P_g(S')|$ denote the respective area and volume over them and under the surface $P_g(t, e)$. As we do in calculus, $P_g(t, e)$ does vary much within S', hence $|P_g(S)| = |S| P_g(\widehat{t}_g, \widehat{e}_g)$ and $|P_g(S')| = |S'| P_g(\widehat{t}_g, \widehat{e}_g) = \epsilon |P_g(S)|$. We are now ready to compute $Pr(t_g \in [\widehat{t}_g, \widehat{t}_g + \delta t] | x_g = \widehat{x}_g) = Pr(S|L) = Pr(S'|L') = Pr(S')/Pr(L') = |P_g(S')|/|P_g(L')| = |P_g(S)|/|P_g(L)| = |S| P_g(\widehat{t}_g, \widehat{e}_g / |P_g(L)|$, where $\widehat{x}_g = X(\widehat{t}_g, \widehat{e}_g)$. It follows that density function of this probability is $P_{\langle x_g = \widehat{x}_g \rangle}(t) = \frac{|S|}{\delta t} \cdot P_g(t, X_E^{-1}(\widehat{x}_g, \widehat{t}_g)) / |P_g(L)|$. Here $|P_g(L)|$ depends one on the fixed value $\widehat{x}_g$. We will ignore it and other "contant" factors that scale a density function so that the area under it is one. However, $\frac{|S|}{\delta t}$ is more problematic, because it relates to the slope of the curved line L. If L is defined by $x = t+e$, then this slope is constant. But with the more general defintion $x = X(t, e)$, this slope $\frac{|S(\widehat{t}_g)|}{\delta t}$ depends on $\widehat{t}_g$. Luckly, we won't need to compute it because it will be the same for both groups $g \in \{A, B\}$.

We will now (temporarily) use the restriction that the talent distribution T is uniform. This means that the density function $P_T(t)$ is constant everywhere in its range and zero elsewhere. Because we decided not to care about the area under our density functions, we might as well assume that $P_T(t) = 1$ within the range. We also have the restriction that $\widehat{x}_g$ is such that $r(\widehat{x}_g) = 2$. This means that the talent range $[x - e_g^{max}, x - e_g^{min}]$ imposed by the environment is a subset of the range $[t^{min}, t^{max}]$ imposed by the talent. This means that for all values of t that we care about $P_T(t) = 1$. This gives us that the density function for $Pr(t_g = t | x_g = \widehat{x}_g)$ is the awkward term $\frac{|S(\widehat{t}_g)|}{\delta t}$ times $P_{\langle x_g = \widehat{x}_g \rangle}(t) = P_g(t, X_E^{-1}(x, t)) =$

$P_T(t) \times P_{E_g}(X_E^{-1}(x,t)) = 1 \times P_{E_g}(X_E^{-1}(x,t)) = P_{E_g}(X_E^{-1}(x,t))$.

When our only restrictions are that T is uniform and $\widehat{x}_g = t_g + e_g$, the probability density function of $Pr(t_g = t | x_g = \widehat{x}_g)$ is $P_{E_g}(X_E^{-1}(x,t)) = P_{E_g}(x-t)$. Due to the linearity of expectation, $Exp(t_g | x_g = \widehat{x}_g) = x - Exp(e_g)$. The result follows that $Exp(t_b | \widehat{x}_b = \widehat{x}_g) - Exp(t_a | \widehat{x}_a = \widehat{x}_g) = Exp(e_a) - Exp(e_b)$.

Lets now look further into the relationship between the distributions E_a and E_b. Initially suppose that $E_a = d \cdot E_b + k$, i.e. we randomly choose a value e'_a from the distribution E_b with probability density $P_{E_b}(e)$ and then set $e_a = d \cdot e'_a + k$. Solving gives $e'_a = \frac{e_a - k}{d}$. It follows that $Pr(e_a \in [e, e+\delta e]) = Pr(e'_a \in [\frac{e-k}{d}, \frac{e-k}{d} + \frac{\delta e}{d}]) = \frac{\delta e}{d} P_{E_b}(\frac{e-k}{d})$. We can ignore the $\frac{1}{d}$ because presumably P_{E_b} already has area one. More generally, $e_a = E_a(e'_a)$, $e'_a = E_a^{-1}(e_a)$, and $Pr(e_a = e)$ has density function $P_{E_a}(e_a) = P_{E_b}(E_a^{-1}(e_a))$.

Combining these two ideas gives that the density function for $Pr(t_a = t | \widehat{x}_a = \widehat{x}_g)$ is $P_{E_a}(X_E^{-1}(x,t)) = P_{E_b}(E_a^{-1}(X_E^{-1}(x,t)))$. Just to check the accuracy of our figure, if $\widehat{x}_a = t_a + e_a$ and $e_a = d \cdot e'_a + k$, then the line to which we restrict the $\langle T, E \rangle$ rectangle is $\widehat{x}_a = t_a + d \cdot e'_a + k$ or $t_a = \widehat{x}_a - d \cdot e'_a - k$. Note this lowers the group B line by k and makes its slope $-d$ instead of -1.

We can obtain the result $Pr(t_b = c_b | \widehat{x}_b = \widehat{x}_g) = Pr(t_a = c_a | \widehat{x}_a = \widehat{x}_g)$ by forcing their probability density functions to be the same, namely $P_{E_b}(X_E^{-1}(x, c_b)) = P_{E_b}(E_a^{-1}(X_E^{-1}(x, c_a)))$, which is obtained by forcing $X_E^{-1}(x, c_b) = E_a^{-1}(X_E^{-1}(x, c_a))$, which is obtained by forcing $c_b = X_T^{-1}(x, X_E^{-1}(x, c_b)) = X_T^{-1}(x, E_a^{-1}(X_E^{-1}(x, c_a)))$.

That is ugly. Lets look at our specific examples. Suppose $x = X(t,e) = t + e$. Then $X_T^{-1}(x,e) = x - e$ and $X_E^{-1}(x,t) = x - t$. Suppose further that $e_a = E_a(e'_a) = d \cdot e'_a + k$. Then $E_a^{-1}(e_a) = \frac{e_a - k}{d}$. This gives $c_b = X_T^{-1}(\widehat{x}_g, E_a^{-1}(X_E^{-1}(\widehat{x}_g, c_a))) = x - E_a^{-1}(x - c_a) = x - \frac{x - c_a - k}{d} = \frac{c_a + k + (d-1)x}{d}$. If further, we set $d = 1$, then we get $c_b = c_a + k$.

What remains is to prove the result when the talent distribution T is not uniform. We will do this by choosing p uniformly from $[0,1]$ and setting $t_g = T(p)$ for an arbitrary function T. Note that $Pr[t_g \leq T(p)] = p$. We have us this talent in the equation $\widehat{x}_g = X(t_g, e_g) = X(T(p_g), e_g)$. Rename p_g to be our "talent" measure and $X'(p_g, e_g) = X(T(p_g), e_g)$. Because this talent measure is uniform, we get the result that $p_b = X_T'^{-1}(\widehat{x}_g, E_a^{-1}(X_E'^{-1}(\widehat{x}_g, p_a))) = T^{-1}(X_T^{-1}(\widehat{x}_g, E_a^{-1}(X_E^{-1}(\widehat{x}_g, T(p_a)))))$. However, we do not want our results in terms of these pseudo talents but in terms the our actual talent measure $c_g = T(p_g)$. This gives our original result $c_b = X_T'^{-1}(\widehat{x}_g, E_a^{-1}(X_E'^{-1}(\widehat{x}_g, c_a)))$. Interesting that this relationship does not depend on the distribution T.

End Proof

==========

Hey! $\widehat{t}_b \gg \widehat{t}_a$, i.e. by some real amount.

Proof:

Rearranging $\widehat{t}_b = X_T^{-1}(\widehat{x}_g, E_a^{-1}(X_E^{-1}(\widehat{x}_g, \widehat{t}_a)))$.

$X_E^{-1}(\widehat{x}_g, \widehat{t}_b) = E_a^{-1}(X_E^{-1}(\widehat{x}_g, \widehat{t}_a))$.

$E_a(X_E^{-1}(\widehat{x}_g, \widehat{t}_b)) = X_E^{-1}(\widehat{x}_g, \widehat{t}_a)$.

Because group A is privileged, $E_a(e_b) = e_a \Rightarrow e_b \ll e_a$.

Hence, $X_E^{-1}(\widehat{x}_g, \widehat{t}_b) \ll X_E^{-1}(\widehat{x}_g, \widehat{t}_a)$.

Performance $X(t,e)$ must increases with talent and with environment.

Hence if $x = X(t,e)$ is fixed, then as e increases from LHS to RHS, t decrease.

Hence, $\widehat{t}_b \gg \widehat{t}_a$.

Theorem: Here we only consider $\widehat{x}_g$ that are non-extreme performance scores, i.e. $r(\widehat{x}_g) = 2$. Let $\langle t_a, e_a, \widehat{x}_a \rangle$ and $\langle t_b, e_b, \widehat{x}_b \rangle$ denote the talent, environment, and measured performance scores for groups A and B. Let c_a and c_b denote desired talent thresholds for acceptance of the person from each group. Let $\widehat{x}_g$ denote the measured performance of both the A and B person that you are deciding between. Below we define the relationship $c_b = F(c_a)$ $\widehat{t}_b = F_{\widehat{x}}(\widehat{t}_a)$ so that

$$Pr(t_b \in [\widehat{t}_b, \widehat{t}_b + \delta\widehat{t}]) = Pr(t_a \in [\widehat{t}_a, \widehat{t}_a + \delta\widehat{t}])$$

$$Pr(t_b = c_b | \widehat{x}_b = \widehat{x}_g) = Pr(t_a = c_a | \widehat{x}_a = \widehat{x}_g) \text{ (or equivalently)}$$
$$Pr(t_b \geq c_b | \widehat{x}_b = \widehat{x}_g) = Pr(t_a \geq c_a | \widehat{x}_a = \widehat{x}_g)$$

- Suppose the talent distribution T is uniform.
  Suppose the measure of performance is the sum $\widehat{x}_g = T_g + E_g$ of the talent and environment for $g \in \{A, B\}$.
  Suppose group A is advantaged by having its environment distribution k more than that for group B, i.e. $E_a = E_b + k$.
  $\rightarrow$ Then $c_b = c_a + k$.
  The effect of this is that if you take the same number of randomly chosen A and B people and sort each group according to their talent, then the talent of the i^{th} B person will be k higher then the i^{th} A person.

- If we temporarily relax the restriction that $\widehat{x}_g$ that is non-extreme performance score, i.e. $r(\widehat{x}_g) = 2$, then this result generalizes to
  $\rightarrow$ Then $c_b = c_a + \frac{r(\widehat{x}_g)}{2} k$
  ** Jeff has not check this one ***
  and $Exp(t_b | \widehat{x}_b = \widehat{x}_g) = Exp(t_a | \widehat{x}_a = \widehat{x}_g) + \frac{r(\widehat{x}_g)}{2} k$.

- Generalize, to allowing $E_a = d \cdot E_b + k$, i.e. the same distribution whose standard deviation has been scaled by d and mean raised by k.
  $\rightarrow$ In this case, $c_b = \frac{c_a + k + (d-1)x}{d}$.

- We can prove the same result when both E_a and E_b are arbitrary distributions. We will model this by defining distribution $E_a = E_a(E_b)$, i.e. we randomly choose a value e'_a from the distribution E_b and then set $e_a = E_a(e'_a)$ where the later E_a denotes an arbitrary function.
  $\rightarrow$ In this case, $c_b = x - E_a^{-1}(x - c_a)$.

- The only restrictions now are that T is uniform, $\widehat{x}_g = t_g + e_g$, and $r(\widehat{x}_g) = 2$. With this we also get
  $$Exp(t_b | \widehat{x}_b = \widehat{x}_g) - Exp(t_a | \widehat{x}_a = \widehat{x}_g) = Exp(e_a) - Exp(e_b)$$

- Suppose now that we are allowed to produce the measured performance score not just by $\widehat{x}_g = t_g + e_g$, but more generally by $\widehat{x}_g = X(t_g, e_g)$ for an arbitrary increasing function X. Let $e_g = X_E^{-1}(\widehat{x}_g, t_g)$ and $t_g = X_T^{-1}(\widehat{x}_g, e_g)$ be two inverse functions.
  Finally, let us allow the talent distribution T to be arbitrary by choosing p uniformly from $[0, 1]$ and setting $t_g = T(p)$ for an arbitrary function T. Note that $Pr[t_g \leq T(p)] = p$.
  $\rightarrow$ This most general result becomes $c_b = X_T^{-1}(\widehat{x}_g, E_a^{-1}(X_E^{-1}(\widehat{x}_g, c_a)))$.

- The only remaining cases to consider are when the range of talent is smaller than the range of environment so that the performance score $\widehat{x}_g$ is extreme in both directions, i.e. $r(\widehat{x}_g) = 0$.
  For here, we will assume that the talent distributions are the same T and uniform, the environment distributions $E_a = E_b + k$ are the same but shifted, and the measured performance score is $\widehat{x}_g = t_g + e_g$. In this case, not only may there not be a gap between the expected talent of person B over person A, but the direction may be reversed.
  $\rightarrow$ If E_b is *sub-exponential*, then $Exp(t_b|\widehat{x}_b = \widehat{x}_g) \geq Exp(t_a|\widehat{x}_a = \widehat{x}_g)$.
  $\rightarrow$ Otherwise it could be that for all $\widehat{x}_g$ (with $r(\widehat{x}_g) = 0$) $Exp(t_b|\widehat{x}_b = \widehat{x}_g) < Exp(t_a|\widehat{x}_a = \widehat{x}_g)$.

==========

Suppose our talent distribution is defined only within the range $[t^{min}, t^{max}]$ and the environment distribution within $[e_b^{min}, e_b^{max}]$. Suppose we condition on the fact that the performance score is given by $\widehat{x}_b = t_b + e_b$ is fixed to some value $\widehat{x}_g$. Rearranging and considering the environment range gives that $t_b = x - e_b \in [x - e_b^{max}, x - e_b^{min}]$. If $\widehat{x}_g$ is an extreme low value, then low range $x - e_b^{max}$ is smaller than the talent low range t^{min} and hence the bound t^{min} kicks in. Similarly, if $\widehat{x}_g$ is an extreme height value, then the high range $x - e_b^{min}$ is trumped by t^{max}. We define $r(\widehat{x}_g)$ to be the number of endpoint for which this does not happen. For example, if the range $[t^{min}, t^{max}]$ on talent is much wider than that on environment and $\widehat{x}_g$ is not an extreme value, then the range $[x - e_b^{max}, x - e_b^{min}]$ is a subset of the range $[t^{min}, t^{max}]$. In this case we say $r(x) = 2$.

==========