Jeff is DONE editing this.
Though we may still consider more general T and E_A.

1 Non-Extreme X Values

Our primary goal is to estimate the talent $t \in T$ of a person from her performance score $x \in X$. The noise making this estimating hard is the person's environment $e \in E$. In an *extreme* case, the range within which these environment values lie is wider than that for the talent. In this case, this noise overwhelms our signal and all the information about the talent is lost. In this section, we give quite a comprehensive version of the theorem under the condition that the performance measure x is non-extreme, i.e. $r(x) = 2$. More formally, this means that the talent range $[X - e^{max}, X - e^{min}]$ imposed by the environment is a subset of the range $[t^{min}, t^{max}]$ imposed by the talent.

We say that group A is privileged over group B if their environment distributions are such that when ever $Pr(E_B \geq e_B) = Pr(E_A \geq e_A)$ we have that $e_B \ll e_A$. This says that if the A person received environment value e_A and the B received e_B, then they are at the same percentiles within their respective groups. However, because group A is privileged over B, the A person would have a significantly better environment value, giving $e_B \ll e_A$.

Consider the following story. Your job is to choose who to accept for some job/university. Being a mediumly desired job, everyone who applies happens to have performance level exactly x. Your goal of course is to accept people whose talent is as high as possible. This paper explains why you should favor people from the disadvantaged B group over those from the privileged A group. The first step is to prove

$$Exp(T_B|x{=}x) - Exp(T_A|X{=}x) = Exp(E_A) - Exp(E_B).$$

But we can say more as follows. Choose N people from group A and N from B randomly conditioned on their performances being x. Sort each group by talent into two parallel lines. For each percentile $p \in [0, 1]$, get the pN^{th} person in each line to shake hands. Let t_A and t_B denote their respective talent. This can be expressed as

$$Pr(T_B {\geq} t_B|X_B{=}x) = Pr(T_A {\geq} t_A|X_A{=}x)$$

This might be useful if you suspect that those people whose talent is higher than percentile p within the privileged group A and higher than the same percentile p within the disadvantaged group B will likely accept a better offer somewhere else. Or maybe p is the risk level you are willing to take. Either way our goal is to compare these two talent levels by defining the function $t_B = F(t_A)$ mapping between them and by proving that $t_B \gg t_A$.

Theorem: Here we only consider x_g that are non-extreme performance scores, i.e. $r(x_g) = 2$. Suppose the talent distribution T is uniform. The environment distributions E_A and E_B can be anything. We are assuming the measure of performance is the sum $X_g = T_g + E_g$ of the talent and environment for $g \in \{A, B\}$. It follows that

$$Exp(T_B|x{=}x) - Exp(T_A|X{=}x) = Exp(E_A) - Exp(E_B).$$

If their environment distributions E_A and E_B are such that group A is privileged over group B, then

$$Pr(T_B \geq t_B | X_B = x_g) = Pr(T_A \geq t_A | X_A = x_g) \ \Rightarrow \ t_B \gg t_A.$$

Suppose further that $E_A = E_B + K$, then $t_B = t_A + K$.
Slightly more generally, if $E_A = d \cdot E_B + k$, then $t_B = t_A + \frac{k}{d} + \frac{d-1}{d} e_A$.
Suppose much more generally, both groups have the same but general talent distribution T and performance is computed by a more general function $X_g = X(T_g, E_g)$, however, these are restricted so that within the range $t \in [t_A, t_B]$, $Pr(T \in [t, t+\delta t]) \in [s^{-1}, s] \times Pr(T \in [t', t'+\delta t])$ and $X_g = X(T_g, E_g) \approx u \cdot T_g + v \cdot E_g + x_0$. Then the function $t_B = F(t_A)$ changes to $t_B \in [s^{-1}, s] \times F(t_A)$.

Proof: We will drop the subscript $g \in \{A, B\}$, when the statements apply to either group. If T is a continuous random variable, then $Pr(T = t) = 0$ for any specific value of t. The standard way of dealing with this is to define the density function $P_T(t)$ so that $Pr(T \in [t, t+\delta t]) = \delta t \cdot P_T(t)$. Similarly, we denote his environment value by e which is drawn from the distribution E with density function $P_E(e)$. Because these two random variables are independent, we can define the cross density function $P(t, e) = P_T(t) \times P_E(e)$ so that $Pr(T \in [t, t+\delta t] \ \& \ E \in [e, e+\delta e]) = \delta t \cdot P_T(t) \times \delta e \cdot P_E(t) = \delta t \delta e \cdot P(t, e)$. 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 $P(t, e)$.

We will now 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. In order to be able to ignore ugly constants, Though it is standard to define density functions so that the area under them is one, it is also natural to relax this condition and then to divide by the area when one wants to compute a probability. This allows us to define $P_T(t) = 1$ within the range. We also have the restriction that x is such that $r(x) = 2$. This means that the talent range $[X - e^{max}, X - e^{min}]$ 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 $P_T(t) = 1$, giving $P(t, x-t) = P_T(t) \times P_E(e) = P_E(e)$.

Fix a performance value x of which we will require of all group g people that we are considering for acceptance. The performance of a person with talent T and environment E is given by $X = T + E$. Just to check the accuracy of our figure, if $E_A = d \cdot E_B + K$, then the line to which we restrict the $\langle T, E \rangle$ rectangle is $x = T_A + d \cdot E_B + K$ or $T_A = x - d \cdot E_B - K$. Note this lowers the group B line by k and makes its slope $-d$ instead of -1.

Our goal is prove that the probability density function $P_x(t) = Pr(T \in [t, t+\delta t] | X = x)/\delta t$ of the distribution on talents t that arise under this condition is simply $P(t, x-t)$. If X were computed by some more complex function, this would not be the case. Using the standard formula $Pr(T \in [t, t+\delta t] \ \& \ X = x)/Pr(X = x)$ is awkward because the later is zero. Conditioning on our probability space amounts to narrowing our $\langle T, E \rangle$ rectangle of possibilities to the 1-dimensional line defined by $\{\langle T, E \rangle \, | x = T + E\}$. Lets us define the infinitesimal rectangle of possibilities $S_t = \{T \in [t, t+\delta t]\} \times \{E \in [x-t-\delta t, x-t]\}$. Within this, X is sufficiently close to x, the density function $P(t, e)$ is sufficiently constant. Hence, we will approximate $Pr(T \in [t, t+\delta t] \ \& \ X = x)$ with $Pr(S_t)$, which is $P(t, x-t) \cdot (\delta t)^2$. Lets return to the awkward fact that $Pr(X = x)$ is zero. Let's define $S_x = \bigcup_t S_t$ to be the union of all of our rectangles within which X is sufficiently close to x. Then we will replacing $Pr(X=x)$ with $Pr(S_x)$. Lets denote this probability with $p_x\delta t$. We are now able to determine the probability $Pr(T \in [t, t+\delta t]|X = x) = Pr(S_t|S_x) = Pr(S_t)/Pr(S_x) = [P(t, x-t)\cdot(\delta t)^2] / (p_x \cdot \delta t)$. Our density function $P_x(t)$ is this divided by δt. Because we decided not to care about the area under our density functions, we get the density function $P_x(t) = P(t, x-t) = P_E(x-t)$.

We are now ready to compare the two groups $g \in \{A, B\}$ using this result. Before we can compare this density function $P_{E_g}(x-t_g)$ for the two groups, we need that the area under them is the same. No matter, what the distribution E_A and E_B are, the areas under their density functions $P_{E_g}(e_g)$ are both one. The area under $P_{E_g}(x-t_g)$ will also be one, as long as when one varies over all values of t_g considered, one gets all possible values of e_g. This is the case, because of the restriction that x is such that the talent range $[X-e^{max}, X-e^{min}]$ imposed by the environment is a subset of the range $[t^{min}, t^{max}]$ imposed by the talent. In conclusion,

$$\text{The density function of } Pr(T_g \in [t_g, t_g+\delta t]|X_g = x) \text{ is } P_{E_g}(x-t_g)$$

Due to the linearity of expectation, $Exp(T_g|X_g=x) = x - Exp(E_g)$. The result follows that

$$Exp(T_B|x=x) - Exp(T_A|X=x) = Exp(E_A) - Exp(E_B).$$

Recall that our second goal is to define a function

$$F_x(t_A) = t_B \text{ so that } Pr(T_B \geq t_B|X_B=x) = Pr(T_A \geq t_A|X_A=x)$$

$$\text{Or equivalently } Pr(T_B \in [t_B, t_B+\delta t]|X_B=x) = Pr(T_A \in [t_A, t_A+\delta t]|X_A=x)$$

To do this this, it is sufficient to equate their density functions and solve for t_B given a each fixed a value for t_A, namely

$$P_{E_B}(x-t_B) = P_{E_A}(x-t_A)$$

Given that these are also the density functions of this other probability, this says

$$Pr(E_B \in [x-t_B, x-t_B+\delta e]) = Pr(E_A \in [x-t_A, x-t_A+\delta e]).$$

Locally, this does not tell us much. However, because we do this simultaneously for every pair $F_x(t_A) = t_B$, we can integrate and get the global statement

$$Pr(E_B \geq x-t_B) = Pr(E_A \geq x-t_A).$$

This says that if the A person received environment value $e_A = x-t_A$ and the B received $e_B = x-t_A$, then they are at the same percentiles within their respective groups. However, because group A is privileged over B, the A person would have a significantly better environment value, giving $e_B \ll e_A$ and hence $t_B \gg t_A$.

In order to be more specific, lets suppose that $E_A = d\cdot E_B + K$, i.e we randomly choose a value E'_A from the distribution E_B and then set $E_A = d\cdot E'_A + K$. Plugging this in gives

$$Pr(E_B \geq x-t_B) = Pr(E_A \geq x-t_A) = Pr(d\cdot E'_A + K \geq x-t_A) = Pr(E'_A \geq d^{-1}(x-t_A-K)).$$

Because E_B and E'_A are drawn from the same distribution, it follows that

$$x - t_B = d^{-1}(x - t_A - K).$$

Solving this gives that

$$F_x(t_A) = t_B = x - \frac{x - t_A - k}{d} = \frac{t_A + K + (d-1)x}{d} = \frac{t_A + K + (d-1)(t_A + e_A)}{d} = t_A + \frac{K}{d} + \frac{d-1}{d} e_A.$$

If further, we set $d = 1$, then we get

$$t_B = t_A + K.$$

Now suppose more generally, performance is computed by $X_g = X(T_g, E_g) = uT_g + vE_g + x_0$. One could achieve the same effect, by first scaling the uniform talent distribution and both environment distributions linearly. This would have no effect on the result.

What remains is to prove the result when the talent distribution T is not uniform. Lets choose p uniformly from $[0,1]$ to denote their pseudo talent and set $T = T(p)$ to be their actual talent, for some arbitrary increasing function T. Because this pseudo talent is uniform, we get the result that $p_B = F(p_A)$ as before, namely if you sort each group, conditioned on $X_A = X_B = x$, then the paired people will have pseudo talents p_A and p_B. If you ask these same people what their actual talent is, they will answer $t_A = T(p_A)$ and $t_B = T(p_B)$. This gives $t_B = T(F(T^{-1}(t_A)))$.

*** Say something about

$Pr(T \in [t, t+\delta t]) \in [s^{-1}, s] \times Pr(T \in [t', t'+\delta t])$, then $t_B = F(t_A)$ changes to $t_B \in [s^{-1}, s] \times F(t_A)$.

These are restricted so that within the range $t \in [t_A, t_B]$, ...

End Proof