Appendix to An evaluation of common methods for dichotomization of continuous variables to discriminate disease status

Prince Nelson, Sybil L.
Department of Public Health Sciences
Medical Univeristy of South Carolina

Ramakrishnan, Viswanathan
Department of Public Health Sciences
Medical Univeristy of South Carolina

Nietert, Paul J.
Department of Public Health Sciences
Medical Univeristy of South Carolina

Kamen, Diane L.
Department of Medicine
Medical Univeristy of South Carolina

Ramos, Paula S.
Department of Medicine
Medical Univeristy of South Carolina

Wolf, Bethany J.
Department of Public Health Sciences
Medical Univeristy of South Carolina

January 2016

1 Appendix

A.1 *Important terms*

For the theoretical investigation of dichotomization methods, we considered a true threshold of X called T such that $P_{Y=1|X\geq T} > P_{Y=1|X<T}$. For each possible threshold chosen, t_x, there are three possibilities: $t_x < T$, $t_x = T$, and $t_x > T$. Each possible threshold, t_x creates new cell values of a $2X2$ contingency table.

1. $t_x < T$

$$\begin{aligned} a &= P_{X\geq T}P_{Y=1|X\geq T} + (P_{X<T} - P_{X<t_x})P_{Y=1|X<T} \\ b &= P_{X\geq t_x} - (P_{X\geq T}P_{Y=1|X\geq T} - (P_{X<T} - P_{X<t_x})P_{Y=1|X<T}) \\ c &= (P_{X<t_x})P_{Y=1|X<T} \\ d &= (P_{X<t_x}) - (P_{X<t_x})P_{Y=1|X<T} \end{aligned} \tag{1}$$

2. $t_x = T$,

$$\begin{aligned} a &= P_{X\geq T}P_{Y=1|X\geq T} \\ b &= P_{X\geq T} - P_{X\geq T}P_{Y=1|X\geq T} \\ c &= P_{X\leq T}P_{Y=1|X<T} \\ d &= P_{X\leq T} - P_{X<T}P_{Y=1|X<T} \end{aligned} \tag{2}$$

3. $t_x > T$

$$\begin{aligned} a &= P_{X\geq t_x}P_{Y=1|X\geq T} \\ b &= P_{X\geq t_x} - P_{X\geq t_x}P_{Y=1|X\geq T} \\ c &= P_{X<T}P_{Y=1|X<T} + (P_{X<t_x} - P_{X<T})P_{Y=1|X\geq T} \\ d &= (P_{X<t_x} - (P_{X<T}P_{Y=1|X<T} - (P_{X<t_x} - P_{X<T})P_{Y=1|X\geq T}) \end{aligned} \tag{3}$$

A.2 *Proof of Theorem 1 for Youden's Statistic*

Let X be a random variable and Y a dichotomous variable. Also, let T be a threshold such that, $P_{Y=1|X\geq T} > P_{Y=1|X<T}$. There are three possibe cases that can occur when selecting a threshold for X, t_x: (1) $t_x < T$,(2) $t_x = T$, and (3) $t_x > T$. The expression for the Youden's Statistic, $\frac{a}{a+c} + \frac{d}{b+d} - 1$, for each case can be found using the expressions for a,b,c, and defined in equations 1, 2, and 3. We can then show that the Youden's Statistic is maximized when $t_x = T$.

A.2.a Consider the case where $P_{X>t_x} > P_{X>T}$. Start with what is given

$$P_{Y=1|X\geq T} > P_{Y=1|X<T}$$

Multiply both sides by $P_{X\geq T}$

$$P_{X\geq T}P_{Y=1|X\geq T} > P_{X\geq T}P_{Y=1|X<T}$$

On the right hand side, let $P_{X\geq T}$ =$1 - P_{X<T}$

$$P_{X\geq T}P_{Y=1|X\geq T} > (1 - P_{X<T})P_{Y=1|X<T}$$

Add $P_{X<T}P_{Y=1|X<T}$ to both sides

$$P_{X\geq T}P_{Y=1|X\geq T} + P_{X<T}P_{Y=1|X<T} > P_{Y=1|X<T}$$

On the left hand side let $P_{X\geq T}P_{Y=1|X\geq T} + P_{X<T}P_{Y=1|X<T} = P_{Y=1}$

$$P_{Y=1} > P_{Y=1|X<T}$$

Multiply by $P_{t_x<X<T}$

$$P_{t_x<X<T}P_{Y=1} > P_{t_x<X<T}P_{Y=1|X<T}$$

Note $P_{t_x<X<T}$=$P_{X<T} - P_{x<t_x}$

$$P_{X<T}P_{Y=1} - P_{x<t_x}P_{Y=1} > P_{t_x<X<T}P_{Y=1|X<T}$$

Add $P_{x<t_x}P_{Y=1}$ to both sides

$$P_{X<T}P_{Y=1} > P_{t_x<X<T}P_{Y=1|X<T} + P_{x<t_x}P_{Y=1}$$

Subtract $P_{X<T}P_{Y=1}P_{Y=1|X<T}$ from both sides

$$\begin{aligned}&P_{X<T}P_{Y=1} - P_{X<T}P_{Y=1}P_{Y=1|X<T}\\&> P_{t_x<X<T}P_{Y=1|X<T} + P_{x<t_x}P_{Y=1} - P_{X<T}P_{Y=1}P_{Y=1|X<T}\end{aligned}$$

On the right hand side, split $P_{X<T}P_{Y=1}P_{Y=1|X<T}$ into $P_{x<t_x}P_{Y=1}P_{Y=1|X<T}+$ $P_{t_x<X<T}P_{Y=1}P_{Y=1|X<T}$

$$\begin{aligned}&P_{X<T}P_{Y=1} - P_{X<T}P_{Y=1}P_{Y=1|X<T} > P_{t_x<X<T}P_{Y=1|X<T}\\&+ P_{x<t_x}P_{Y=1} - P_{x<t_x}P_{Y=1}P_{Y=1|X<T} - P_{t_x<X<T}P_{Y=1}P_{Y=1|X<T}\end{aligned}$$

Add $(1 - P_{Y=1})P_{X\geq T}P_{Y=1|x\geq T}$ to both sides

$$\begin{aligned}&(1 - P_{Y=1})P_{X\geq T}P_{Y=1|x\geq T}P_{X<T}P_{Y=1} - P_{X<T}P_{Y=1}P_{Y=1|X<T}\\&> (1 - P_{Y=1})P_{X\geq T}P_{Y=1|x\geq T}P_{t_x<X<T}P_{Y=1|X<T}\\&+ P_{x<t_x}P_{Y=1} - P_{x<t_x}P_{Y=1}P_{Y=1|X<T} - P_{t_x<X<T}P_{Y=1}P_{Y=1|X<T}\end{aligned}$$

Factor both sides

$$\begin{aligned}&(1 - P_{Y=1})P_{X\geq T}P_{Y=1|x\geq T}P_{Y=1}(P_{X<T} - P_{X<T}P_{Y=1|X<T})\\&> (1 - P_{Y=1})(P_{X\geq T}P_{Y=1|x\geq T} + P_{t_x<X<T}P_{Y=1|X<T})\\&+ P_{Y=1}(P_{x<t_x}(1 - P_{Y=1|X<T}))\end{aligned}$$

Divide both sides by $(1 - P_{Y+1})$ and $P_{Y=1}$

$$\frac{(1-P_{Y=1})P_{X\geq T}P_{Y=1|x\geq T}P_{Y=1}(P_{X<T} - P_{X<T}P_{Y=1|X<T})}{(1-P_{Y+1})P_{Y=1}}$$
$$> \frac{(1-P_{Y=1})(P_{X\geq T}P_{Y=1|x\geq T} + P_{t_x<X<T}P_{Y=1|X<T}) + P_{Y=1}(P_{x<t_x}(1-P_{Y=1|X<T}))}{(1-P_{Y+1})P_{Y=1}}$$

Simplify

$$\frac{P_{X\geq T}P_{Y=1|x\geq T}}{P_{Y=1}} + \frac{(P_{X<T} - P_{X<T}P_{Y=1|X<T})}{(1-P_{Y=1})}$$
$$> \frac{(P_{X\geq T}P_{Y=1|x\geq T} + P_{t_x<X<T}P_{Y=1|X<T})}{P_{Y=1}}$$
$$+ \frac{P_{x<t_x}(1-P_{Y=1|X<T})}{(1-P_{Y=1})}$$

Subtract 1 from both sides

$$\frac{P_{X\geq T}P_{Y=1|x\geq T}}{P_{Y=1}} + \frac{(P_{X<T} - P_{X<T}P_{Y=1|X<T})}{(1-P_{Y=1})} - 1$$
$$> \frac{(P_{X\geq T}P_{Y=1|x\geq T} + P_{t_x<X<T}P_{Y=1|X<T})}{P_{Y=1}}$$
$$+ \frac{P_{x<t_x}(1-P_{Y=1|X<T})}{(1-P_{Y=1})} - 1$$

We have

$$\frac{a_{t_x=T}}{(a+c)_{t_x=T}} + \frac{d_{t_x=T}}{(b+d)_{t_x=T}} - 1 > \frac{a_{t_x<T}}{(a+c)_{t_x<T}} + \frac{d_{t_x<T}}{(b+d)_{t_x<T}} - 1$$

Thus, $Youd_{t_x=T} > Youd_{t_x<T}$.

A.2.b Now consider the case where $P_{X>t_x} < P_{X>T}$. Start with what is given

$$P_{Y=1|X\geq T} > P_{Y=1|X<T}$$

Multiply both sides by $P_{X<T}$

$$P_{X<T}P_{Y=1|X\geq T} > P_{X<T}P_{Y=1|X<T}$$

On the left, let $P_{X<T} = (1 - P_{Y=1|X\geq T})$

$$P_{Y=1|X\geq T} - P_{X>T}P_{Y=1|X\geq T} > P_{X<T}P_{Y=1|X<T}$$

Add $P_{X>T}P_{Y=1|X\geq T}$ to both sides

$$P_{Y=1|X\geq T} > P_{X<T}P_{Y=1|X<T} + P_{X>T}P_{Y=1|X\geq T}$$

Note $P_{Y=1} = P_{X<T}P_{Y=1|X<T} + P_{X>T}P_{Y=1|X\geq T}$

$$P_{Y=1|X\geq T} > P_{Y=1}$$

Multiply both sides by $P_{T<X<t_x}$

$$P_{T<X<t_x}P_{Y=1|X\geq T} > P_{T<X<t_x}P_{Y=1}$$

Subtract $P_{T<X<t_x}P_{Y=1}P_{Y=1|X\geq T}$ from both sides

$$P_{T<X<t_x}P_{Y=1|X\geq T} - P_{T<X<t_x}P_{Y=1}P_{Y=1|X\geq T} > P_{T<X<t_x}P_{Y=1} - P_{T<X<t_x}P_{Y=1}P_{Y=1|X\geq T}$$

Let $P_{T<X<t_x} = P_{X\geq T} - P_{X\geq t_x}$

$$(P_{X\geq T} - P_{X\geq t_x})P_{Y=1|X\geq T} - (P_{X\geq T} - P_{X\geq t_x})P_{Y=1}P_{Y=1|X\geq T} > P_{T<X<t_x}P_{Y=1} - P_{T<X<t_x}P_{Y=1}P_{Y=1|X\geq T}$$

Distribute

$$\begin{aligned}&P_{X\geq T}P_{Y=1|X\geq T} - P_{X\geq t_x}P_{Y=1|X\geq T}\\ &- P_{X\geq T}P_{Y=1}P_{Y=1|X\geq T} + P_{X\geq t_x}P_{Y=1}P_{Y=1|X\geq T}\\ &> P_{T<X<t_x}P_{Y=1} - P_{T<X<t_x}P_{Y=1}P_{Y=1|X\geq T}\end{aligned}$$

Add $P_{X\geq t_x}P_{Y=1|X\geq T}$ to both sides and subtract $P_{X\geq t_x}P_{Y=1}P_{Y=1|X\geq T}$ from both sides

$$\begin{aligned}&P_{X\geq T}P_{Y=1|X\geq T} - P_{X\geq T}P_{Y=1}P_{Y=1|X\geq T}\\ &> P_{X\geq t_x}P_{Y=1|X\geq T} - P_{X\geq t_x}P_{Y=1}P_{Y=1|X\geq T} + P_{T<X<t_x}P_{Y=1} - P_{T<X<t_x}P_{Y=1}P_{Y=1|X\geq T}\end{aligned}$$

Note $P_{T<X<t_x}P_{Y=1} = (P_{X<t_x} - P_{X<T})P_{Y=1}$ and add $P_{X<T}P_{Y=1}$ to both sides

$$\begin{aligned}&P_{X\geq T}P_{Y=1|X\geq T} - P_{X\geq T}P_{Y=1}P_{Y=1|X\geq T} + P_{X<T}P_{Y=1}\\ &> P_{X\geq t_x}P_{Y=1|X\geq T} - P_{X\geq t_x}P_{Y=1}P_{Y=1|X\geq T} + P_{X<t_x}P_{Y=1} - P_{T<X<t_x}P_{Y=1}P_{Y=1|X\geq T}\end{aligned}$$

Subtract $P_{X<T}P_{Y=1}P_{Y=1|X<T}$ from both sides

$$\begin{aligned}&P_{X\geq T}P_{Y=1|X\geq T} - P_{X\geq T}P_{Y=1}P_{Y=1|X\geq T} + P_{X<T}P_{Y=1} - P_{X<T}P_{Y=1}P_{Y=1|X<T}\\ &> P_{X\geq t_x}P_{Y=1|X\geq T} - P_{X\geq t_x}P_{Y=1}P_{Y=1|X\geq T}\\ &+ P_{X<t_x}P_{Y=1} - P_{X<T}P_{Y=1}P_{Y=1|X<T} - P_{T<X<t_x}P_{Y=1}P_{Y=1|X\geq T}\end{aligned}$$

Divide both sides by $P_{Y=1}(1 - P_{Y=1})$ and factor

$$\frac{P_{X\geq T}P_{Y=1|X\geq T}}{P_{Y=1}}+\frac{P_{X<T}(1-P_{Y=1|X<T})}{(1-P_{Y=1})}>\frac{P_{X\geq t_x}P_{Y=1|X\geq T}}{P_{Y=1}}$$
$$+\frac{P_{X<t_x}-P_{X<T}P_{Y=1|X<T}-P_{T<X<t_x}P_{Y=1|X\geq T}}{(1-P_{Y=1})}$$

We have

$$\frac{a_{t_x=T}}{(a+c)_{t_x=T}}+\frac{d_{t_x=T}}{(b+d)_{t_x=T}}-1>\frac{a_{t_x>T}}{(a+c)_{t_x>T}}+\frac{d_{t_x>T}}{(b+d)_{t_x>T}}-1$$

Thus, $Youd_{t_x=t} > Youd_{t_x>t}$. If the expression for $Youd_{t_x=T}$ is greater than the expression for $Youd_{t_x<T}$ and the expression for $Youd_{t_x=T}$ is greater than the expression for $Youd_{t_x>T}$ then it shows that the Youden's Statistic is the highest when $t_x = T$.

A.3 *Proof of Theorem 1 for Gini Index*

Let X be a random variable and Y a dichotomous variable. Also, let T be a threshold such that, $P_{Y=1|X\geq T} > P_{Y=1|X<T}$. There are three possibe cases that can occur when selecting a threshold for X, t_x: (1) $t_x < T$,(2) $t_x = T$, and (3) $t_x > T$. The expression for the Gini Index, $(P_y(1-P_y)) - (\frac{ab}{a+b}+\frac{cd}{c+d})$, for each case can be found using the expressions for a,b,c, and defined in equations 1, 2, and 3. We can then show that the Gini Index is maximized when $t_x = T$.

A.3.a Consider the case where $P_{X>t_x} > P_{X>T}$. Start with what is given

$$P_{Y=1|X\geq T} > P_{Y=1|X<T}$$

Subtract $P_{Y=1|X<T}$ from both sides

$$0 < P_{Y=1|X\geq T} - P_{Y=1|X<T}$$

Square both sides

$$0 < (P_{Y=1|X\geq T} - P_{Y=1|X<T})^2$$

Multiply

$$0 < P^2_{Y=1|X\geq T} - 2P_{Y=1|X\geq T}P_{Y=1|X<T} + P^2_{Y=1|X<T}$$

Subtract $P^2_{Y=1|X\geq T}$ and $P^2_{Y=1|X<T}$ from both sides

$$-P^2_{Y=1|X\geq T} - P^2_{Y=1|X<T} < -2P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Multiply by $P_{X\geq T}P_{t_x<X<T}$

$$(-P_{t_x<X<T})P_{X\geq T}P^2_{Y=1|X\geq T}+(-P_{X\geq T})P_{t_x<X<T}P^2_{Y=1|X<T} < -2P_{X\geq T}P_{Y=1|X\geq T}P_{t_x<X<T}P_{Y=1|X<T}$$

Note $-P_{t_x<X<T} = P_{X\geq T} - P_{X>t_x}$ and $-P_{X\geq T} = P_{t_x<X<T} - P_{X>t_x}$

$$\begin{aligned}&(P_{X\geq T} - P_{X>t_x})P_{X\geq T}P^2_{Y=1|X\geq T} + (P_{t_x<X<T} - P_{X>t_x})P_{t_x<X<T}P^2_{Y=1|X<T}\\ <& -2P_{X\geq T}P_{Y=1|X\geq T}P_{t_x<X<T}P_{Y=1|X<T}\end{aligned}$$

Expand left hand side

$$\begin{aligned}&(P_{X>t}P_{Y=1|X>t})^2 - P_{X>t_x}P_{X>t}(P_{Y=1|X>t})^2 + (P_{t_x<X<t}P_{Y=1|X<t}))^2 - P_{X>t_x}(P_{t_x<X<t})(P_{Y=1|X<t})^2\\ <& -2P_{X\geq T}P_{Y=1|X\geq T}P_{t_x<X<T}P_{Y=1|X<T}\end{aligned}$$

Subtract $(P_{X\geq T}P_{Y=1|X\geq T})^2$ and $(P_{t_x<X<T}P_{Y=1|X<T}))^2$ from both sides

$$\begin{aligned}&-P_{X\geq t_x}P_{X\geq T}(P_{Y=1|X\geq T})^2 - P_{X\geq t_x}(P_{t_x<X<T})(P_{Y=1|X<T})^2\\ <& -(P_{X\geq T}P_{Y=1|X\geq T})^2 - (P_{t_x<X<T}P_{Y=1|X<T}))^2 - 2P_{X\geq T}P_{Y=1|X\geq T}P_{t_x<X<T}P_{Y=1|X<T}\end{aligned}$$

Factor right hand side

$$\begin{aligned}&-P_{X\geq t_x}P_{X\geq T}(P_{Y=1|X\geq T})^2 - P_{X\geq t_x}(P_{t_x<X<T})(P_{Y=1|X<T})^2\\ <& -((P_{X\geq T}P_{Y=1|X\geq T} + (P_{t_x<X<T})P_{Y=1|X<t}))^2\end{aligned}$$

Add $P_{X\geq t_x}(P_{t_x<X<T})P_{Y=1|X<T}$ and $P_{X\geq t_x}P_{X\geq T}P_{Y=1|X\geq T}$

$$\begin{aligned}&P_{X\geq t_x}P_{X\geq T}P_{Y=1|X} - P_{X\geq t_x}P_{X\geq T}(P_{Y=1|X\geq T})^2 + P_{X\geq t_x}(P_{t_x<X<T})P_{Y=1|x<T} - P_{X\geq t_x}(P_{t_x<X<T})(P_{Y=1|X<T}\\ <& P_{X\geq t_x}P_{X\geq T}P_{Y=1|X\geq T} + P_{X\geq t_x}(P_{t_x<X<T})P_{Y=1|X<T} - ((P_{X\geq t}P_{Y=1|X\geq T} + (P_{t_x<X<T})P_{Y=1|X<T}))^2\end{aligned}$$

Factor

$$\begin{aligned}&P_{X\geq t_x}P_{X\geq T}P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T}) + P_{X\geq t_x}(P_{t_x<X<T})P_{Y=1|x<T}(1 - P_{Y=1|X<T})\\ <&(P_{X\geq T}P_{Y=1|X\geq T} + (P_{t_x<X<T})P_{Y=1|X<T})(P_{X\geq t_x} - (P_{X\geq T}P_{Y=1|X\geq T} + (P_{t_x<X<T})P_{Y=1|X<T}))\end{aligned}$$

Divide by $P_{X\geq t_x}$

$$\begin{aligned}&P_{X\geq T}P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T}) + (P_{t_x<X<T})P_{Y=1|x<T}(1 - P_{Y=1|X<T})\\ <&\frac{(P_{X\geq T}P_{Y=1|X\geq T} + (P_{t_x<x<T})P_{Y=1|X<T})(P_{X\geq t_x} - (P_{X\geq T}P_{Y=1|X\geq T} + (P_{t_x<x<T})P_{Y=1|X<T}))}{P_{X\geq t_x}}\end{aligned}$$

Note $P_{t_x<X<T} = P_{X<T} - P_{X<t_x}$.

$$\begin{aligned}&P_{X\geq T}P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T}) + (P_{X<T} - P_{X<t_x})P_{Y=1|x<T}(1 - P_{Y=1|X<T})\\ <&\frac{(P_{X\geq T}P_{Y=1|X\geq T} + (P_{t_x<x<T})P_{Y=1|X<T})(P_{X\geq t_x} - (P_{X\geq T}P_{Y=1|X\geq T} + (P_{t_x<x<T})P_{Y=1|X<T}))}{P_{X\geq t_x}}\end{aligned}$$

Distribute

$$
\begin{aligned}
&P_{X\geq T}P_{Y=1|X\geq T}(1-P_{Y=1|X\geq T})+P_{X<T}P_{Y=1|X<T}(1-P_{Y=1|X<T})-(P_{X<t_x}P_{Y=1|X<T})(1-P_{Y=1|X<T}))\\
<&\frac{(P_{X\geq T}P_{Y=1|X\geq T}+(P_{t_x<X<T})P_{Y=1|X<T})(P_{X\geq t_x}-(P_{X\geq T}P_{Y=1|X\geq T}+(P_{t_x<X<T})P_{Y=1|X<T}))}{P_{X\geq t_x}}
\end{aligned}
$$

Add $(P_{X<t_x}P_{Y=1|X<T})(1-P_{Y=1|X<T}))$ to both sides

$$
\begin{aligned}
&P_{X\geq T}P_{Y=1|X\geq T}(1-P_{Y=1|X\geq T})+P_{X<T}P_{Y=1|X<T}(1-P_{Y=1|X<T})\\
<&(\frac{(P_{X\geq T}P_{Y=1|X\geq T}+(P_{t_x<X<T})P_{Y=1|X<T})(P_{X\geq t_x}-(P_{X\geq T}P_{Y=1|X\geq T}+(P_{t_x<X<T})P_{Y=1|X<T}))}{P_{X\geq t_x}}\\
&+(P_{X<t_x}P_{Y=1|X<T})(1-P_{Y=1|X<T}))
\end{aligned}
$$

$$
\begin{aligned}
&(\frac{(P_{X\geq T}P_{Y=1|X\geq T})(P_{X\geq T}-P_{X\geq T}P_{Y=1|X\geq T})}{P_{X\geq T}}+\frac{(P_{X<T}P_{Y=1|X<T})(P_{X<T}-P_{X<T}P_{Y=1|X<T})}{P_{X<T}})\\
<&(\frac{(P_{X\geq T}P_{Y=1|X\geq T}+(P_{t_x<x<T})P_{Y=1|X<T})(P_{X\geq t_x}-(P_{X\geq T}P_{Y=1|X\geq T}+(P_{t_x<x<T})P_{Y=1|X<T}))}{P_{X\geq t_x}}\\
&+\frac{((P_{X<t_x})P_{Y=1|X<T})((P_{X<t_x})-(P_{X<t_x})P_{Y=1|X<T})}{(P_{X<t_x})})
\end{aligned}
$$

$$
\begin{aligned}
&-(\frac{(P_{X\geq T}P_{Y=1|X\geq T})(P_{X\geq T}-P_{X\geq T}P_{Y=1|X\geq T})}{P_{X\geq T}P_{Y=1|X\geq T}+P_{X\geq T}-P_{X\geq T}P_{Y=1|X\geq T}}+\frac{(P_{X<T}P_{Y=1|X<T})(P_{X<T}-P_{X<T}P_{Y=1|X<T})}{P_{X<T}P_{Y=1|X<T}+P_{X<T}-P_{X<T}P_{Y=1|X<T}})\\
>&-(\frac{(P_{X\geq T}P_{Y=1|X\geq T}+(P_{t_x<x<T})P_{Y=1|X<T})(P_{X\geq t_x}-(P_{X\geq T}P_{Y=1|X\geq T}-(P_{t_x<x<T})P_{Y=1|X<T}))}{P_{X\geq T}P_{Y=1|X\geq T}+(P_{t_x<x<T})P_{Y=1|X<T}+P_{X\geq t_x}-(P_{X\geq T}P_{Y=1|X\geq T}-(P_{t_x<x<T})P_{Y=1|X<T})}\\
&+\frac{((P_{X<t_x})P_{Y=1|X<T})((P_{X<t_x})-(P_{X<t_x})P_{Y=1|X<T})}{(P_{X<t_x})P_{Y=1|X<T}+(P_{X<t_x})-(P_{X<t_x})P_{Y=1|X<T}})
\end{aligned}
$$

$$(P_y(1-P_y))-(\frac{ab}{a+b}+\frac{cd}{c+d})>(P_y(1-P_y))-(\frac{ab}{a+b}+\frac{cd}{c+d})$$

Thus,

$$Gini_{t_x=t}>Gini_{t_x<t}$$

A.3.b Now consider the case where $P_{X>t_x}<P_{X>T}$. Start with what is given

$$P_{Y=1|X\geq T}>P_{Y=1|X<T}$$

Subtract $P_{Y=1|X<T}$ from both sides

$$0 < P_{Y=1|X\geq T} - P_{Y=1|X<T}$$

Square both sides

$$0 < (P_{Y=1|X\geq T} - P_{Y=1|X<T})^2$$

Expand

$$0 < (P_{Y=1|X\geq T})^2 - 2P_{Y=1|X<T}P_{Y=1|X\geq T} + (P_{Y=1|X<T})^2$$

Add $2P_{Y=1|X<T}P_{Y=1|X\geq T}$ to both sides

$$2P_{Y=1|X<T}P_{Y=1|X\geq T} < (P_{Y=1|X\geq T})^2 + (P_{Y=1|X<T})^2$$

Multiply both sides by $P_{X<T}P_{t<X<t_x}($

$$2P_{X<T}P_{Y=1|X<T}P_{t<X<t_x}P_{Y=1|X\geq T} < P_{X<T}P_{t<X<t_x}(P_{Y=1|X\geq T})^2 + P_{t<X<t_x}P_{X<T}(P_{Y=1|X<T})^2$$

Multiply by -1

$$-2P_{X<T}P_{Y=1|X<T}P_{t<X<t_x}P_{Y=1|X\geq T} < (-P_{X<T})P_{t<X<t_x}(P_{Y=1|X\geq T})^2 + (-P_{t<X<t_x})P_{X<T}(P_{Y=1|X<T})^2$$

Note $-P_{t<X<t_x}$ is $(P_{t<X<t_x} - P_{X<t_x})$ to $-P_{X<T}$ and $(P_{X<T} - P_{X<t_x})$

$$\begin{aligned}&-2P_{X<T}P_{Y=1|X<T}P_{t<X<t_x}P_{Y=1|X\geq T}\\ &< (P_{t<X<t_x} - P_{X<t_x})P_{t<X<t_x}(P_{Y=1|X\geq T})^2 + (P_{X<T} - P_{X<t_x})P_{X<T}(P_{Y=1|X<T})^2\end{aligned}$$

Distribute

$$\begin{aligned}&-2P_{X<Tt}P_{Y=1|X<T}P_{t<X<t_x}P_{Y=1|X\geq T}\\ &< -P_{X<t_x}P_{t<X<t_x}(P_{Y=1|X\geq T})^2 + P^2_{t<X<t_x}P^2_{Y=1|X\geq T} - P_{X<t_x}P_{X<T}(P_{Y=1|X<T})^2 + P^2_{X<T}P^2_{Y=1|X<T}\end{aligned}$$

Subtract $P^2_{X<T}P^2_{Y=1|X<T}$ and $P^2_{t<X<t_x}P^2_{Y=1|X\geq T}$ from both sides

$$\begin{aligned}&-P^2_{X<T}P^2_{Y=1|X<T} - P^2_{t<X<t_x}P^2_{Y=1|X\geq T} - 2P_{X<T}P_{Y=1|X<T}P_{t<X<t_x}P_{Y=1|X\geq T}\\ &< -P_{X<t_x}P_{t<X<t_x}(P_{Y=1|X\geq T})^2 - P_{X<t_x}P_{X<T}(P_{Y=1|X<T})^2\end{aligned}$$

Factor left hand side

$$\begin{aligned}&-(P_{X<T}P_{Y=1|X<T} + (P_{t<X<t_x})P_{Y=1|X\geq T})^2\\ &< -P_{X<t_x}P_{t<X<t_x}(P_{Y=1|X\geq T})^2 - P_{X<t_x}P_{X<t}(P_{Y=1|X<T})^2\end{aligned}$$

Add $P_{X<t_x}P_{t<X<t_x}P_{Y=1|X\geq T}$ and $P_{X<t_x}P_{X<T}P_{Y=1|X<T}$ to both sides

$$P_{X<t_x} P_{X<T} P_{Y=1|X<T} + P_{X<t_x} P_{t<X<t_x} P_{Y=1|X\geq T} - (P_{X<T} P_{Y=1|X<T} + (P_{t<X<t_x}) P_{Y=1|X\geq T})^2$$
$$< P_{X<t_x} P_{t<X<t_x} P_{Y=1|X\geq T} - P_{X<t_x} P_{t<X<t_x} (P_{Y=1|X\geq T})^2 + P_{X<t_x} P_{X<T} P_{Y=1|X<T} - P_{X<t_x} P_{X<T} (P_{Y=1|X<}$$

Factor out $P_{X<t_x}$

$$P_{X<t_x}(P_{X<T} P_{Y=1|X<T} + (P_{t<X<t_x}) P_{Y=1|X\geq T}) - (P_{X<T} P_{Y=1|X<T} + (P_{t<X<t_x}) P_{Y=1|X})^2$$
$$< P_{X<t_x} P_{t<X<t_x} P_{Y=1|X\geq T} - P_{X<t_x} P_{t<X<t_x} (P_{Y=1|X\geq T})^2 + P_{X<t_x} P_{X<T} P_{Y=1|X<T} - P_{X<t_x} P_{X<T} (P_{Y=1|X<}$$

Factor

$$(P_{X<T} P_{Y=1|X<T} + (P_{t<X<t_x}) P_{Y=1|X\geq T})(P_{X<t_x} - (P_{X<T} P_{Y=1|X<T} - P_{t<X<t_x} P_{Y=1|X\geq T}))$$
$$< P_{X<t_x} P_{t<X<t_x} P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T}) + P_{X<t_x} P_{X<T} P_{Y=1|X<T}(1 - P_{Y=1|X<T})$$

Divide by $P_{X<t_x}$

$$\frac{(P_{X<T} P_{Y=1|X<T} + (P_{t<X<t_x}) P_{Y=1|X\geq T})(P_{X<t_x} - (P_{X<T} P_{Y=1|X<T} - P_{t<X<t_x} P_{Y=1|X\geq T}))}{P_{X<t_x}}$$
$$< P_{t<X<t_x} P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T}) + P_{X<T} P_{Y=1|X<T}(1 - P_{Y=1|X<T})$$

Separate $P_{t<X<t_x}$ term

$$\frac{(P_{X<t} P_{Y=1|X<t} + (P_{t<X<t_x}) P_{Y=1|X>t})(P_{X<t_x} - (P_{X<t} P_{Y=1|X<t} - P_{t<X<t_x} P_{Y=1|X>t}))}{P_{X<t_x}}$$
$$< P_{X\geq T} P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T}) + P_{X<T} P_{Y=1|X<T}(1 - P_{Y=1|X<T}) - P_{X>t_x} P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T})$$

Add $P_{X>t_x} P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T})$ to both sides

$$P_{X>t_x} P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T})$$
$$+ \frac{(P_{X<T} P_{Y=1|X<T} + (P_{t<X<t_x}) P_{Y=1|X\geq T})(P_{X<t_x} - (P_{X<T} P_{Y=1|X<T} - P_{t<X<t_x} P_{Y=1|X\geq T}))}{P_{X<t_x}}$$
$$< P_{X\geq T} P_{Y=1|X\geq T}(1 - P_{Y=1|X\geq T}) + P_{X<T} P_{Y=1|X<T}(1 - P_{Y=1|X<T})$$

Multiply by $P_{X>t_x}$ and $P_{X\geq T}$. Divide by $P_{X>t_x} P_{Y=1|X\geq T} + (P_{X>t_x} - P_{X>t_x} P_{Y=1|X\geq T}$ and $P_{X<t} P_{Y=1|X<T} + (P_{t<X<t_x}) P_{Y=1|X\geq T} + (P_{X<t_x} - (P_{X<T} P_{Y=1|X<T} - P_{t<X<t_x} P_{Y=1|X>t}))$

$$\frac{P_{X>t_x} P_{Y=1|X\geq T}(P_{X>t_x} - P_{X>t_x} P_{Y=1|X>t})}{P_{X>t_x} P_{Y=1|X\geq T} + (P_{X>t_x} - P_{X>t_x} P_{Y=1|X\geq T})}$$
$$+ \frac{(P_{X<t} P_{Y=1|X<T} + (P_{t<X<t_x}) P_{Y=1|X\geq T})(P_{X<t_x} - (P_{X<T} P_{Y=1|X<T} - P_{t<X<t_x} P_{Y=1|X\geq T}))}{P_{X<T} P_{Y=1|X<T} + (P_{t<X<t_x}) P_{Y=1|X\geq T} + (P_{X<t_x} - (P_{X<T} P_{Y=1|X<T} - P_{t<X<t_x} P_{Y=1|X\geq T}))}$$
$$< \frac{(P_{X\geq T} P_{Y=1|X\geq T})(P_{X\geq T} - P_{X\geq T} P_{Y=1|X\geq T})}{P_{X\geq T} P_{Y=1|X\geq T} + P_{X\geq T} - P_{X\geq T} P_{Y=1|X\geq T}} + \frac{(P_{X<T} P_{Y=1|X<T})(P_{X<T} - P_{X<T} P_{Y=1|X<T})}{P_{X<T} P_{Y=1|X<T} + P_{X<T} - P_{X<T} P_{Y=1|X<T}}$$

Thus $Gini_{t_x>T} < Gini_{t_x=T}$ If the expression for $Gini_{t_x=T}$ is greater than the expression for $Gini_{t_x<T}$ and the expression for $Gini_{t_x=T}$ is greater than the expression for $Gini_{t_x>T}$ then it shows that the Gini Index is the highest when $t_x = T$.

A.4 *Proof of Theorem 1 for the chi-square Statistic*

Let X be a random variable and Y a dichotomous variable. Also, let T be a threshold such that, $P_{Y=1|X\geq T} > P_{Y=1|X<T}$. There are three possibe cases that can occur when selecting a threshold for X, t_x: (1) $t_x < T$,(2) $t_x = T$, and (3) $t_x > T$. The expression for the chi-square, $\frac{(ad-bc)^2}{(a+b)(c+d)(b+d)(a+c)}$, for each case can be found using the expressions for a,b,c, and defined in equations 1, 2, and 3. We can then show that the chi-square is maximized when $t_x = T$.

A.4.a Consider the case where $P_{X>t_x} > P_{X>T}$. Start with what is given

$$P_{Y=1|X\geq T} > P_{Y=1|X<T}$$

Subtract $P_{Y=1|X<T}$ from both sides

$$P_{Y=1|X\geq T} - P_{Y=1|X<T} > 0$$

Square both sides

$$(P_{Y=1|X\geq T} - P_{Y=1|X<T})^2 > 0$$

Square both sides

$$(P_{Y=1|X\geq T})^2 - 2P_{Y=1|X\geq T}P_{Y=1|X<T} + (P_{Y=1|X<T})^2 > 0$$

Add $2P_{Y=1|X\geq T}P_{Y=1|X<T}$ to both sides

$$(P_{Y=1|X\geq T})^2 + (P_{Y=1|X<T})^2 > 2P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Multiply both sides by $(P_{x<T})^2 - (P_{x<t_x})^2$

$$((P_{x<T})^2 - (P_{x<t_x})^2)(P_{Y=1|X\geq T})^2 + ((P_{x<T})^2 - (P_{x<t_x})^2)(P_{Y=1|X<T})^2$$
$$> 2((P_{x<T})^2 - (P_{x<t_x})^2)P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Distribute

$$(P_{x<T})^2(P_{Y=1|X\geq T})^2 - (P_{x<t_x})^2(P_{Y=1|X\geq T})^2 + (P_{x<T})^2(P_{Y=1|X<T})^2 - (P_{x<t_x})^2(P_{Y=1|X<T})^2$$
$$> 2(P_{x<T})^2P_{Y=1|X\geq T}P_{Y=1|X<T} - 2(P_{x<t_x})^2P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Subtract $2(P_{x<T})^2P_{Y=1|X\geq T}P_{Y=1|X<T}$, add $(P_{x<t_x})^2(P_{Y=1|X\geq T})^2$ and $(P_{x<t_x})^2(P_{Y=1|X<T})^2$.

$$(P_{x<T})^2(P_{Y=1|X\geq T})^2 + (P_{x<T})^2(P_{Y=1|X<T})^2 - 2(P_{x<T})^2P_{Y=1|X\geq T}P_{Y=1|X<T}$$
$$> (P_{x<t_x})^2(P_{Y=1|X\geq T})^2 + (P_{x<t_x})^2(P_{Y=1|X<T})^2 - 2(P_{x<t_x})^2P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Multiply both sides by $(P_{X \geq T})^2$.

$$(P_{X\geq T})^2(P_{x<T})^2(P_{Y=1|X\geq T})^2 + (P_{X\geq T})^2(P_{x<T})^2(P_{Y=1|X<T})^2 - 2(P_{X\geq T})^2(P_{x<T})^2 P_{Y=1|X\geq T}P_{Y=1|X<T}$$
$$> (P_{X\geq T})^2(P_{x<t_x})^2(P_{Y=1|X\geq T})^2 + (P_{X\geq T})^2(P_{x<t_x})^2(P_{Y=1|X<T})^2 - 2(P_{X\geq T})^2(P_{x<t_x})^2 P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Factor by difference of squares

$$(P_{X<T}P_{X\geq T}P_{Y=1|X\geq T} - P_{X<T}P_{X\geq T}P_{Y=1|X<T})^2 > (P_{X<t_x}P_{X\geq T}P_{Y=1|X\geq T} - P_{X<t_x}P_{X\geq T}P_{Y=1|X<T})^2$$

On the left side, add and subtract $P_{X\geq T}P_{Y=1|X\geq T}P_{X<T}P_{Y=1|X<T}$

$$(P_{X<T}P_{X\geq T}P_{Y=1|X\geq T} - P_{X\geq T}P_{Y=1|X\geq T}P_{X<T}P_{Y=1|X<T}$$
$$- P_{X<T}P_{X\geq T}P_{Y=1|X<T} + P_{X\geq T}P_{Y=1|X\geq T}P_{X<T}P_{Y=1|X<T})^2$$
$$> (P_{X<t_x}P_{X\geq T}P_{Y=1|X\geq T} - P_{X<t_x}P_{X\geq T}P_{Y=1|X<T})^2$$

Thus, the left side factors by difference of squares

$$(P_{X\geq T}P_{Y=1|X\geq T}(P_{X<T} - P_{X<T}P_{Y=1|X<T}) - (P_{X\geq T} - P_{X\geq T}P_{Y=1|X\geq T})(P_{X<T}P_{Y=1|X<T}))^2$$
$$> (P_{X<t_x}P_{X\geq T}P_{Y=1|X\geq T} - P_{X<t_x}P_{X\geq T}P_{Y=1|X<T})^2$$

On the right side, note $-P_{X\geq T} = (P_{X<T} - 1)$

$$(P_{X\geq T}P_{Y=1|X\geq T}(P_{X<T} - P_{X<T}P_{Y=1|X<T}) - (P_{X\geq T} - P_{X\geq T}P_{Y=1|X\geq T})(P_{X<T}P_{Y=1|X<T}))^2$$
$$> (P_{X<t_x}P_{X\geq T}P_{Y=1|X\geq T} + P_{X<t_x}P_{Y=1|X<T}(P_{X<T} - 1))^2$$

Distribute on the right

$$(P_{X\geq T}P_{Y=1|X\geq T}(P_{X<T} - P_{X<T}P_{Y=1|X<T}) - (P_{X\geq T} - P_{X\geq T}P_{Y=1|X\geq T})(P_{X<T}P_{Y=1|X<T}))^2$$
$$> (P_{X<t_x}P_{X\geq T}P_{Y=1|X\geq T} + P_{X<t_x}P_{Y=1|X<T}P_{X<T} - P_{X<t_x}P_{Y=1|X<T})^2$$

Also note $P_{X<t_x} + P_{X>t_x} = 1$. So, multiply by 1 on the right

$$(P_{X\geq T}P_{Y=1|X\geq T}(P_{X<T} - P_{X<T}P_{Y=1|X<T}) - (P_{X\geq T} - P_{X\geq T}P_{Y=1|X\geq T})(P_{X<T}P_{Y=1|X<T}))^2$$
$$> (P_{X<t_x}P_{X\geq T}P_{Y=1|X\geq T} + P_{X<t_x}P_{Y=1|X<T}P_{X<T} - P_{X<t_x}P_{Y=1|X<T}(P_{X<t_x} + P_{X>t_x}))^2$$

Distribute on the right

$$(P_{X\geq T}P_{Y=1|X\geq T}(P_{X<T} - P_{X<T}P_{Y=1|X<T}) - (P_{X\geq T} - P_{X\geq T}P_{Y=1|X\geq T})(P_{X<T}P_{Y=1|X<T}))^2$$
$$> (P_{X<t_x}P_{X\geq T}P_{Y=1|X\geq T} + P_{X<t_x}P_{Y=1|X<T}P_{X<T}$$
$$- P_{X<t_x}P_{Y=1|X<T}P_{X<t_x} - P_{X<t_x}P_{Y=1|X<T}P_{X>t_x})^2$$

On the right side, add and subtract $(P_{X\geq T}P_{Y=1|X\geq T} + (P_{X<T} - P_{X<t_x})P_{Y=1|X<T})P_{X<t_x}P_{Y=1|X<T}$.

$$(P_{X\geq T}P_{Y=1|X\geq T}(P_{X<T}-P_{X<T}P_{Y=1|X<T})-(P_{X\geq T}-P_{X\geq T}P_{Y=1|X\geq T})(P_{X<T}P_{Y=1|X<T}))^2$$
$$>(P_{X<t_x}P_{X\geq T}P_{Y=1|X\geq T}+P_{X<t_x}P_{Y=1|X<T}P_{X<T}$$
$$-(P_{X\geq T}P_{Y=1|X\geq T}+(P_{X<T}-P_{X<t_x})P_{Y=1|X<T})P_{X<t_x}P_{Y=1|X<T}$$
$$-P_{X<t_x}P_{Y=1|X<T}P_{X<t_x}-P_{X<t_x}P_{Y=1|X<T}P_{X\geq t_x}$$
$$+(P_{X\geq T}P_{Y=1|X\geq T}+(P_{X<T}-P_{X<t_x})P_{Y=1|X<T})P_{X<t_x}P_{Y=1|X<T})^2$$

Factor the right side,

$$(P_{X\geq T}P_{Y=1|X\geq T}(P_{X<T}-P_{X<T}P_{Y=1|X<T})-(P_{X\geq T}-P_{X\geq T}P_{Y=1|X\geq T})(P_{X<T}P_{Y=1|X<T}))^2$$
$$>((P_{X\geq T}P_{Y=1|X\geq T}+(P_{X<T}-P_{X<T})P_{Y=1|X<T})((P_{X<T}-P_{X<T})P_{Y=1|X<T})$$
$$-(P_{X\geq t_x}-(P_{X\geq T}P_{Y=1|X\geq T}+(P_{X<T}-P_{X<T})P_{Y=1|X<T}))(P_{X<t_x}P_{Y=1|X<T}))^2$$

Divide both sides by $P_{X\geq T}(1-P_{Y=1})P_{X<T}P_{Y=1}$ and we have

$$\chi^2_{t_x=T}>\chi^2_{t_x<T}$$

A.4.b Now consider the case where $P_{X>t_x}<P_{X>T}$. Start with what is given

$$P_{Y=1|X\geq T}>P_{Y=1|X<T}$$

Subtract $P_{Y=1|X<T}$ from both sides

$$P_{Y=1|X\geq T}-P_{Y=1|X<T}>0$$

Square both sides

$$(P_{Y=1|X\geq T}-P_{Y=1|X<T})^2>0$$

Square both sides

$$(P_{Y=1|X\geq T})^2-2P_{Y=1|X\geq T}P_{Y=1|X<T}+(P_{Y=1|X<T})^2>0$$

Add $2P_{Y=1|X\geq T}P_{Y=1|X<T}$ to both sides

$$(P_{Y=1|X\geq T})^2+(P_{Y=1|X<T})^2>2P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Multiply both sides by $(P_{x>T})^2-(P_{x>t_x})^2$

$$((P_{x>T})^2-(P_{x>t_x})^2)(P_{Y=1|X\geq T})^2+((P_{x>T})^2-(P_{x>t_x})^2)(P_{Y=1|X<T})^2$$
$$>2((P_{x>T})^2-(P_{x>t_x})^2)P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Distribute

$$(P_{x>T})^2(P_{Y=1|X\geq T})^2-(P_{x>t_x})^2(P_{Y=1|X\geq T})^2+(P_{x>T})^2(P_{Y=1|X<T})^2-(P_{x>t_x})^2(P_{Y=1|X<T})^2$$
$$>2(P_{x>T})^2P_{Y=1|X\geq T}P_{Y=1|X<T}-(P_{x>t_x})^2P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Rearrange terms

$$(P_{x>T})^2(P_{Y=1|X\geq T})^2 - 2(P_{x>T})^2P_{Y=1|X\geq T}P_{Y=1|X<T} + (P_{x>T})^2(P_{Y=1|X<T})^2$$
$$> (P_{x>t_x})^2(P_{Y=1|X\geq T})^2 - 2(P_{x>t_x})^2P_{Y=1|X\geq T}P_{Y=1|X<T} + (P_{x>t_x})^2(P_{Y=1|X<T})^2$$

Multiply by $(P_{X<T})^2$

$$(P_{x>T})^2(P_{X<T})^2(P_{Y=1|X\geq T})^2 - 2(P_{X<T})^2(P_{x>T})^2P_{Y=1|X\geq T}P_{Y=1|X<T} + (P_{x>T})^2(P_{X<T})^2(P_{Y=1|X<T})^2$$
$$> (P_{x>t_x})^2(P_{X<T})^2(P_{Y=1|X\geq T})^2 - 2(P_{X<T})^2(P_{x>t_x})^2P_{Y=1|X\geq T}P_{Y=1|X<T} + (P_{X<T})^2(P_{x>t_x})^2(P_{Y=1|X<T})^2$$

Factor

$$((P_{x>T})(P_{X<T})(P_{Y=1|X\geq T}) - (P_{x>T})^2(P_{X<T})(P_{Y=1|X<T}))^2$$
$$> ((P_{x>t_x})(P_{X<T})(P_{Y=1|X\geq T}) - (P_{X<T})(P_{x>t_x})(P_{Y=1|X<T}))^2$$

Therefore,

$$(a_{t_x=T}d_{t_x=T} - b_{t_x=T}c_{t_x=T})^2 > (a_{t_x>T}d_{t_x>T} - b_{t_x>T}c_{t_x>T})^2$$

and we have, $\chi^2_{t_x=T} > \chi^2_{t_x>T}$ If the expression for $chi^2_{t_x=T}$ is greater than the expression for $chi^2_{t_x<T}$ and the expression for $chi^2_{t_x=T}$ is greater than the expression for $chi^2_{t_x>T}$ then it shows that the chi-square is the highest when $t_x = T$.

A.5 *Proof of Theorem 1 for Relative Risk*

Let X be a random variable and Y a dichotomous variable. Also, let T be a threshold such that, $P_{Y=1|X\geq T} > P_{Y=1|X<T}$. There are three possibe cases that can occur when selecting a threshold for X, t_x: (1) $t_x < T$,(2) $t_x = T$, and (3) $t_x > T$. The expression for the Relative Risk, $\frac{a/(a+b)}{c/(c+d)}$, for each case can be found using the expressions for a,b,c, and defined in equations 1, 2, and 3. We can then show that the Relative Risk is maximized when $t_x = T$.

A.5.a Consider the case where $P_{X>t_x} > P_{X>T}$. Start with what is given

$$P_{Y=1|X\geq T} > P_{Y=1|X<T}$$

Multiply both sides by $P_{Y=1|X<T}$

$$P_{Y=1|X\geq T}P_{Y=1|X<T} > (P_{Y=1|X<T})^2$$

Set equal to 0

$$0 > -P_{Y=1|X\geq T}P_{Y=1|X<T} + (P_{Y=1|X<T})^2$$

Multiply both sides by $(P_{X<t_x} - P_{X<T})$

$$0 > -P_{Y=1|X\geq T}P_{Y=1|X<T}(P_{X<t_x} - P_{X<T}) + (P_{X<t_x} - P_{X<T})(P_{Y=1|X<T})^2$$

Replace $P_{X<t_x}$ with $(1 - P_{X>t_x})$ and $P_{X<T}$ with $(1 - P_{X\geq T})$

$$0 > -P_{Y=1|X\geq T}P_{Y=1|X<T}((1-P_{X>t_x})-(1-P_{X\geq T}))+(P_{X<t_x}-P_{X<T})(P_{Y=1|X<T})^2$$

Distribute

$$0 > -P_{Y=1|X\geq T}P_{Y=1|X<T}P_{X>t_x}+P_{Y=1|X\geq T}P_{Y=1|X<T}P_{X\geq T})+(P_{X<t_x}-P_{X<T})(P_{Y=1|X<T})^2$$

Add $P_{Y=1|X\geq T}P_{Y=1|X<T}P_{X>t_x}$ to both sides

$$P_{Y=1|X\geq T}P_{Y=1|X<T}P_{X>t_x} > P_{Y=1|X\geq T}P_{Y=1|X<T}P_{X\geq T})+(P_{X<t_x}-P_{X<T})(P_{Y=1|X<T})^2$$

Factor out $P_{Y=1|X<T}$ from the left side

$$P_{Y=1|X\geq T}P_{Y=1|X<T}P_{X>t_x} > P_{Y=1|X<T}(P_{Y=1|X\geq T}P_{X\geq T}+(P_{X<t_x}-P_{X<T})P_{Y=1|X<T})$$

Divide both sides by $(P_{Y=1|X<T})^2P_{X>t_x}$

$$\frac{P_{Y=1|X\geq T}}{P_{Y=1|X<T}} > \frac{P_{Y=1|X\geq T}P_{X\geq T} + (P_{X<t_x} - P_{X<T})P_{Y=1|X<T}}{P_{Y=1|X<T}P_{X>t_x}}$$

Multiply the left side by $\frac{P_{X\geq T}P_{X<T}}{P_{X\geq T}P_{X<T}}$ and the right side by $\frac{P_{X<t_x}}{P_{X<t_x}}$

$$\frac{P_{X\geq T}P_{Y=1|X\geq T}P_{X<T}}{P_{X<T}P_{Y=1|X<T}P_{X\geq T}} > \frac{P_{Y=1|X\geq T}P_{X\geq T} + (P_{X<t_x} - P_{X<T})P_{Y=1|X<T}P_{X<t_x}}{P_{X<t_x}P_{Y=1|X<T}P_{X>t_x}}$$

Thus $RR_{t_x=T} > RR_{t_x<T}$

A.5.b Now consider the case where $P_{X>t_x} < P_{X>T}$. Start with what is given

$$P_{Y=1|X\geq T} > P_{Y=1|X<T}$$

Multiply both sides by $P_{Y=1|X\geq T}$

$$(P_{Y=1|X\geq T})^2 > P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Set equal to 0

$$(P_{Y=1|X\geq T})^2 - P_{Y=1|X\geq T}P_{Y=1|X<T} > 0$$

Multiply both sides by $(P_{X<t_x} - P_{X<T})$

$$(P_{X<t_x} - P_{X<T})(P_{Y=1|X\geq T})^2 - (P_{X<t_x} - P_{X<T})P_{Y=1|X\geq T}P_{Y=1|X<T} > 0$$

Factor out a negative 1

$$(P_{X<t_x} - P_{X<T})(P_{Y=1|X\geq T})^2 + (P_{X<T} - P_{X<t_x})P_{Y=1|X\geq T}P_{Y=1|X<T} > 0$$

Distribute

$$P_{X<t_x}(P_{Y=1|X\geq T})^2 - P_{X<T}(P_{Y=1|X\geq T})^2 + P_{X<T}P_{Y=1|X\geq T}P_{Y=1|X<T} - P_{X<t_x}P_{Y=1|X\geq T}P_{Y=1|X<T} > 0$$

Add $P_{X<t_x}P_{Y=1|X\geq T}P_{Y=1|X<T}$ to both sides

$$P_{X<t_x}(P_{Y=1|X\geq T})^2 - P_{X<T}(P_{Y=1|X\geq T})^2 + P_{X<T}P_{Y=1|X\geq T}P_{Y=1|X<T} > P_{X<t_x}P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Factor out $P_{Y=1|X\geq T}$ from the left

$$P_{Y=1|X\geq T}(P_{X<t_x}(P_{Y=1|X\geq T}) - P_{X<T}(P_{Y=1|X\geq T}) + P_{X<T}P_{Y=1|X<T}) > P_{X<t_x}P_{Y=1|X\geq T}P_{Y=1|X<T}$$

Divide both sides by $P_{Y=1|X<T}$ and $(P_{X<t_x}(P_{Y=1|X\geq T}) - P_{X<T}(P_{Y=1|X\geq T}) + P_{X<T}P_{Y=1|X<T})$

$$\frac{P_{Y=1|X\geq T}}{P_{Y=1|X<T}} > \frac{P_{X<t_x}P_{Y=1|X\geq T}}{(P_{X<t_x}(P_{Y=1|X\geq T}) - P_{X<T}(P_{Y=1|X\geq T}) + P_{X<T}P_{Y=1|X<T})}$$

Multiply by $\frac{P_{X>T}}{P_{X>T}}$, $\frac{P_{X<T}}{P_{X<T}}$ and $\frac{P_{X>t_x}}{P_{X>t_x}}$

$$\frac{P_{X>T}P_{Y=1|X\geq T}P_{X<T}}{P_{X<T}P_{Y=1|X<T}P_{X>T}} > \frac{P_{X>t_x}P_{X<t_x}P_{Y=1|X\geq T}}{(P_{X<t_x}(P_{Y=1|X\geq T}) - P_{X<T}(P_{Y=1|X\geq T}) + P_{X<T}P_{Y=1|X<T})P_{X>t_x}}$$

Thus $RR_{t_x=T} > RR_{t_x>T}$ If the expression for $RR_{t_x=T}$ is greater than the expression for $RR_{t_x<T}$ and the expression for $RR_{t_x=T}$ is greater than the expression for $RR_{t_x>T}$ then it shows that the Relative Risk is the highest when $t_x = T$.

A.6 *Proof of theorem 1 for Kappa statistic*

Let X be a random variable and Y a dichotomous variable. Also, let T be a threshold such that, $P_{Y=1|X\geq T} > P_{Y=1|X<T}$. There are three possibe cases that can occur when selecting a threshold for X, t_x: (1) $t_x < T$,(2) $t_x = T$, and (3) $t_x > T$. The expression for Kappa, $\frac{(a+d)-((a+b)(a+c)+(c+d)(b+d))}{1-((a+b)(a+c)+(c+d)(b+d))}$, for each case can be found using the expressions for a,b,c, and defined in equations 1, 2, and 3. We can then show that Kappa is maximized when $t_x = T$.

A.6.a First we want to show that $Kappa_{t_x=T} > Kappa_{t_x<T}$
We begin with the true statement

$$P_{X<T} > P_{X<t_x}$$

Note $(P_{X\geq t_x} + P_{X<t_x}) = 1$

$$P_{X<T}(P_{X\geq t_x} + P_{X<t_x}) > P_{X<t_x}$$

Distribute and set equal to 0

$$P_{X<T}P_{X\geq t_x} - P_{X<t_x} + P_{X<T}P_{X<t_x} > 0$$

Factor out $P_{X<t_x}$

$$P_{X<T}P_{X\geq t_x} - P_{X<t_x}(1 - P_{X<T}) > 0$$

Note $1 - P_{X<T} = P_{X\geq T}$

$$P_{X<T}P_{X\geq t_x} - P_{X<t_x}(P_{X\geq T}) > 0$$

Add $P_{X<t_x}(P_{X\geq T})$ to both sides

$$P_{X<T}P_{X\geq t_x} > P_{X<t_x}(P_{X\geq T})$$

Multiply both sides by $(1 - P_{Y=1})$

$$P_{X<T}P_{X\geq t_x}(1 - P_{Y=1}) > P_{X<t_x}(P_{X\geq T})(1 - P_{Y=1})$$

Distribute

$$P_{X<T}P_{X\geq t_x} - P_{X<T}P_{X\geq t_x}P_{Y=1} > P_{X<t_x}P_{X\geq T} - P_{X<t_x}P_{X\geq T}P_{Y=1}$$

Add $P_{X<T}P_{Y=1}P_{X<t_x}$ to both sides

$$P_{X<T}P_{X\geq t_x}+P_{X<T}P_{Y=1}P_{X<t_x}-P_{X<T}P_{X\geq t_x}P_{Y=1} > P_{X<t_x}P_{X\geq T}+P_{X<t_x}P_{Y=1}P_{X<T}-P_{X<t_x}P_{X\geq T}P_{Y=1}$$

Factor both sides

$$P_{X<T}P_{X\geq t_x}+P_{X<T}P_{Y=1}(P_{X<t_x}-P_{X\geq t_x}) > P_{X<t_x}P_{X\geq T}+P_{X<t_x}P_{Y=1}(P_{X<T}-P_{X\geq T})$$

Factor $P_{X<T}$ and P_{t_x}

$$P_{X<T}(P_{X\geq t_x}+P_{Y=1}(P_{X<t_x}-P_{X\geq t_x})) > P_{X<t_x}(P_{X\geq T}+P_{Y=1}(P_{X<T}-P_{X\geq T}))$$

Divide by $P_{X\geq t_x}+P_{Y=1}(P_{X<t_x}-P_{X\geq t_x})$ and $P_{X\geq T}+P_{Y=1}(P_{X<T}-P_{X\geq T})$

$$\frac{P_{X<T}}{P_{X\geq T}+P_{Y=1}(P_{X<T}-P_{X\geq T})} > \frac{P_{X<t_x}}{P_{X\geq t_x}+P_{Y=1}(P_{X<t_x}-P_{X\geq t_x})}$$

Multiply by $2P_{X\geq T}(P_{Y=1|X\geq T}-P_{Y=1|X<T})$

$$\frac{2P_{X\geq T}P_{X<T}(P_{Y=1|X\geq T}-P_{Y=1|X<T})}{P_{X\geq T}+P_{Y=1}(P_{X<T}-P_{X\geq T})} > \frac{2P_{X<t_x}P_{X\geq T}(P_{Y=1|X\geq T}-P_{Y=1|X<T})}{P_{X\geq t_x}+P_{Y=1}(P_{X<t_x}-P_{X\geq t_x})}$$

Thus $Kappa_{t_x=T} > Kappa_{t_x<T}$. If the expression for $Kappa_{t_x=T}$ is greater than the expression for $Kappa_{t_x<T}$ and the expression for $Kappa_{t_x=T}$ is greater than the expression for $Kappa_{t_x>T}$ then it shows that the Kappa is the highest when $t_x = T$.

A.6.b Next we show $Kappa_{t_x=T} < Kappa_{t_x<T}$

$$P_{X\geq T} > P_{X\geq t_x}$$

Note that $P_{X<t_x}+P_{X\geq t_x}=1$

$$P_{X\geq T}(P_{X<t_x}+P_{X\geq t_x}) > P_{X\geq t_x}$$

Distribute and set equal to zero

$$P_{X\geq T}P_{X<t_x}-P_{X\geq t_x}+P_{X\geq T}P_{X\geq t_x} > 0$$

Factor out $P_{X\geq t_x}$

$$P_{X\geq T}P_{X<t_x}-(1-P_{X\geq T})P_{X\geq t_x} > 0$$

Note $1-P_{X\geq T}=P_{X<T}$,

$$P_{X\geq T}P_{X<t_x}-P_{X<T}P_{X\geq t_x} > 0$$

Add $P_{X<T}P_{X\geq t_x}$ to both sides

$$P_{X\geq T}P_{X<t_x} > P_{X<T}P_{X\geq t_x}$$

Subtract $P_{X\geq T}P_{X\geq t_x}$ from both sides

$$P_{X\geq T}P_{X<t_x}-P_{X\geq T}P_{X\geq t_x} > P_{X<T}P_{X\geq t_x}-P_{X\geq T}P_{X\geq t_x}$$

Factor each side and multiply by 2

$$2P_{X\geq T}(P_{X<t_x}-P_{X\geq t_x}) > 2P_{X\geq t_x}(P_{X<t_x}-P_{X\geq t_x})$$

Multiply both sides by $(P = Y = 1|X \geq T - P_{Y=1})$

$$2P_{X\geq T}(P=Y=1|X\geq T-P_{Y=1})(P_{X<t_x}-P_{X\geq t_x}) > 2P_{X\geq t_x}(P=Y=1|X\geq T-P_{Y=1})(P_{X<t_x}-P_{X\geq t_x})$$

Add $2P_{X \geq T}(P = Y = 1|X \geq T - P_{Y=1})P_{X \geq t_x}$ to both sides

$$2P_{X \geq T}(P = Y = 1|X \geq T - P_{Y=1})P_{X \geq t_x} + 2P_{X \geq T}(P = Y = 1|X \geq T - P_{Y=1})(P_{X<t_x} - P_{X \geq t_x}) >$$
$$2P_{X \geq T}(P = Y = 1|X \geq T - P_{Y=1})P_{X \geq t_x} + 2P_{X \geq t_x}(P = Y = 1|X \geq T - P_{Y=1})(P_{X<t_x} - P_{X \geq t_x})$$

Factor both sides

$$(2P_{X \geq T}(P_{Y=1|X \geq T} - P_{Y=1}))(P_{X \geq t_x} + P_{Y=1})(P_{X<t_x} - P_{X \geq t_x})) >$$
$$(2P_{X \geq t_x}(P_{Y=1|X \geq T} - P_{Y=1}))(P_{X \geq T} + P_{Y=1})(P_{X<T} - P_{X \geq T}))$$

Divide both sides by $P_{X \geq t_x} + P_{Y=1})(P_{X<t_x} - P_{X \geq t_x})$ and $P_{X \geq T} + P_{Y=1})(P_{X<T} - P_{X \geq T})$

$$\frac{2P_{X \geq T}(P_{Y=1|X \geq T} - P_{Y=1})}{P_{X \geq T} + P_{Y=1})(P_{X<T} - P_{X \geq T})} > \frac{2P_{X \geq t_x}(P_{Y=1|X \geq T} - P_{Y=1})}{P_{X \geq t_x} + P_{Y=1})(P_{X<t_x} - P_{X \geq t_x})}$$

Thus, $Kappa_{t_x=T} > Kappa_{t_x>T}$.