For each set of twins s, let $(1, s)$ denote the twin that smokes, and let $(2, s)$ denote the non-smoking twin. Let T_{is} denote random smoking indicators; $T_{is} = 1$ if the ith unit in the sth twin pair smokes, $i = 1, 2$. For this study, for each pair s, the smoking indicators are observed to be $T_{1s} = 1$ and $T_{2s} = 0$. The statistician models the probability that a subject smokes in the following way:

$$\log\left(\frac{P(T_{is} = 1)}{1 - P(T_{is} = 1)}\right) = \alpha + \kappa_1 h_{is} + \kappa_2 w_{is} + \gamma u_{is} \tag{1}$$

where h_{is} and w_{is} are the height and weight of twin (i, s), and u_{is} is the value of an unobserved covariate for that twin. The statistician also assumes that any subject cannot influence any other subject to smoke or not smoke (smoking is independent across all subjects).

Show that, under this model, the probability that subject $(1, s)$ is a smoker is:

$$P(T_{1s} = 1 | T_{1s} + T_{2s} = 1) = \frac{e^{\gamma u_{1s}}}{e^{\gamma u_{1s}} + e^{\gamma u_{2s}}} \tag{2}$$

Hint: Use $P(A|B) = P(A \cap B)/P(B)$, and find an expression for

$$\frac{P(T_{1s} = 1 \cap T_{2s} = 0)}{P(T_{1s} = 0 \cap T_{2s} = 1)} = \left(\frac{P(T_{1s} = 1)}{1 - P(T_{1s} = 1)}\right)\left(\frac{P(T_{2s} = 0)}{1 - P(T_{2s} = 0)}\right)$$

c) Suppose that $0 \leq u_{is} \leq 1$ and that $\gamma > 0$. Find an upper and lower bound (sharper than just 1 and 0) for the probability $P(T_{1s} = 1 | T_{1s} + T_{2s} = 1)$. Denote these bounds by p_s^+ and p_s^- respectively. Do the same for $P(T_{1s} = 0 | T_{1s} + T_{2s} = 1)$. Comment, in one sentence, on how these bounds change if $\gamma < 0$.

d) Let y_{is} denote the 40-yard dash time of subject (i, s) in milliseconds. Let Z_s denote an indicator variable for the smoker having the faster 40-yard dash time: $Z_s = 1$ if and only if the smoking twin has a faster 40-yard dash time than the non-smoking twin. Let d_s denote the rank of $|y_{1s} - y_{2s}|$; higher ranks denote larger absolute values. Assume there are no ties between y_{1s} and y_{2s} within any twin pair s, and that $|y_{1s} - y_{2s}| \neq |y_{1t} - y_{2t}|$ for all distinct twin pairs s, t.

The Wilcoxon signed rank statistic is:

$$W = \sum_{s=1}^{n} d_s Z_s.$$

Let Z_s^+ and Z_s^- be independent and identically distributed bernoulli random variables (or indicator variables) with $P(Z_s^+ = 1) = p_s^+$ and $P(Z_s^- = 1) = p_s^-$. Consider the following random variables:

$$\begin{aligned} W^+ &= \sum_{s=1}^{n} d_s Z_s^+ \\ W^- &= \sum_{s=1}^{n} d_s Z_s^- \end{aligned}$$

Show that, under the null hypothesis that smoking does not effect 40-yard dash times, the following property holds:

$$\mathbb{E}(W^-) \leq \mathbb{E}(W | T_{1s} + T_{2s} = 1) \leq \mathbb{E}(W^+)$$

e) In fact, it can be shown that under this null hypothesis, for any a:

$$P(W^- \geq a) \leq P(W \geq a | T_{1s} + T_{2s} = 1) \leq P(W^+ \geq a) \tag{3}$$

Discuss, in about 3 -5 sentences or so, how property (3) can be exploited to test the exact null of no treatment effect.

f) [BONUS QUESTION] Prove property (3).