Problem 1.) I will break this into two parts:

(1) Proving $\tilde{w}^{(m)} = p(\tilde{x}^{(m)}|X_{i\cdot} = x_{i\cdot}, X_{\cdot j} = x_{\cdot j}, p_{ij} = p_{i\cdot}p_{\cdot j})$. In other words, the probability of a specific table in T_x given the row and column counts are fixed to the observed values and the null hypothesis is true (column and row probabilities are independent).

(2) Given 1, prove if you create a rejection region $\{\tilde{x}^{(j)} : \tilde{w}^{(j)} = \tilde{w}^{(i_t)} \text{ for } t \geq (m+1)\}$ as described in the question will give a test of size $\leq \alpha$.

Part 1.) We know that the probability of any specific $\tilde{x}$ only given $p_{ij} = p_{i\cdot}p_{\cdot j}$ is just multinomial(n, p) where $p \in \Delta_{k_1 k_2}$ and $p_{ij} = p_{i\cdot}p_{\cdot j}$. To add the condition on the row/column counts, we can normalize by summing these probabilities over all $\tilde{x} \in T_x$, that is, all possible tables for which the row and column count condition holds. This means

$$p(\tilde{x}^{(m)}|X_{i\cdot} = x_{i\cdot}, X_{\cdot j} = x_{\cdot j}, p_{ij} = p_{i\cdot}p_{\cdot j}) = \frac{p(\tilde{x}^{(m)}|p_{ij} = p_{i\cdot}p_{\cdot j})}{\sum_{i=1}^{M} p(\tilde{x}^{(i)}|p_{ij} = p_{i\cdot}p_{\cdot j})}. \tag{1}$$

Now we will examine $p(\tilde{x}^{(m)}|p_{ij} = p_{i\cdot}p_{\cdot j})$. Before we do, observe that $w^{(m)}$ defined in the HW is the multinomial normalizing constant. Also, for notational convenience, let $x_{ij} = \tilde{x}_{ij}^{(m)}$. Then we have,

$$\begin{aligned}
p(\tilde{x}^{(m)}|p_{ij} = p_{i\cdot}p_{\cdot j}) &= w^{(m)}(p_{11})^{x_{11}}(p_{12})^{x_{12}}...(p_{k_1k_2})^{x_{k_1k_2}} \\
&= w^{(m)}(p_{1\cdot}p_{\cdot 1})^{x_{11}}(p_{1\cdot}p_{\cdot 2})^{x_{12}}...(p_{k_1\cdot}p_{\cdot k_2})^{x_{k_1k_2}} && \text{by independence assumption} \\
&= w^{(m)}(p_{1\cdot})^{x_{11}}(p_{\cdot 1})^{x_{11}}(p_{1\cdot})^{x_{12}}(p_{\cdot 2})^{x_{12}}...(p_{k_1\cdot})^{x_{k_1k_2}}(p_{\cdot k_2})^{x_{k_1k_2}}
\end{aligned}$$

Now to simplify, we need to combine terms and add up the exponents. But the question is: For a specific p, say $p_{1\cdot}$, what will the exponent be? $p_{1\cdot}$ will appear once for each cell in the first row, and no others. So when we combine all the $p_{1\cdot}$'s and add the exponents, in the exponent we will get $\sum_{j=1}^{k_2} x_{1j} = x_{1\cdot}$, the row 1 total. Applying this logic, we will get

$$= w^{(m)}(p_{1\cdot})^{x_{1\cdot}}(p_{2\cdot})^{x_{2\cdot}}...(p_{k_1\cdot})^{x_{k_1\cdot}}(p_{\cdot 1})^{x_{\cdot 1}}(p_{\cdot 2})^{x_{\cdot 2}}...(p_{\cdot k_2})^{x_{\cdot k_2}}.$$

So everything besides $w^{(m)}$ depends only on the row and column totals. But since every $\tilde{x} \in T_x$ has the same row and column totals, the piece besides $w^{(m)}$ will be exactly the same for all $\tilde{x} \in T_x$. Plugging this back into (1), we get

$$p(\tilde{x}^{(m)}|X_{i\cdot} = x_{i\cdot}, X_{\cdot j} = x_{\cdot j}, p_{ij} = p_{i\cdot}p_{\cdot j}) = \frac{w^{(m)}(p_{1\cdot})^{x_{1\cdot}}(p_{2\cdot})^{x_{2\cdot}}...(p_{k_1\cdot})^{x_{k_1\cdot}}(p_{\cdot 1})^{x_{\cdot 1}}(p_{\cdot 2})^{x_{\cdot 2}}...(p_{\cdot k_2})^{x_{\cdot k_2}}}{\sum_{i=1}^{M} w^{(i)}(p_{1\cdot})^{x_{1\cdot}}(p_{2\cdot})^{x_{2\cdot}}...(p_{k_1\cdot})^{x_{k_1\cdot}}(p_{\cdot 1})^{x_{\cdot 1}}(p_{\cdot 2})^{x_{\cdot 2}}...(p_{\cdot k_2})^{x_{\cdot k_2}}}$$
$$= \frac{w^{(m)}}{\sum_{i=1}^{M} w^{(i)}}$$
$$= \tilde{w}^{(m)}$$

as desired.

Part 2.) For now, let's ignore how the $\tilde{w}$'s are reordered, because in terms of proving it is a size $\leq \alpha$ test, it does not matter (that is not to say it isn't important how we reorder them, more on that later). Given any ordering, the test boils down to the following (Note, I changed some of the wording from the question because I think this is clearer. Some thought will show this is equivalent) :

- Find the *smallest* m such that $\tilde{w}^{(i_1)} + \tilde{w}^{(i_2)} + ... + \tilde{w}^{(i_m)} > 1 - \alpha$. Note that this implies that $w^{(i_{m+1})} + ... + w^{(i_M)} < \alpha$.
- Notice that any observed x must be $\tilde{x}^{(i)} \in T_x$ for some $i = 1, ..., M$. For a specific observed x, fail to reject if x is in $\{\tilde{x}^{(i_1)}, \tilde{x}^{(i_2)}, ..., \tilde{x}^{(i_m)}\}$ and reject if x is in $\{\tilde{x}^{(i_{m+1})}, ..., \tilde{x}^{(i_M)}\}$.

Since we proved that $\tilde{w}^{(m)}$ gives the probability $\tilde{x}^{(m)}$ under the null hypothesis, the probability of getting an x in the rejection region under the null hypothesis is $p(\tilde{x}^{(i_{m+1})}, ..., \tilde{x}^{(i_M)}) = \tilde{w}^{(i_{m+1})} + ... + \tilde{w}^{(i_M)} < \alpha$. This defines a test of size $\leq \alpha$.

The ordering is important because it determines the power of the test. The idea is in order to maximize power, put the most likely $\tilde{x}$'s into the fail to reject region and the least likely $\tilde{x}$'s into the rejection region (under the null hypothesis). The way it is described in the HW, they are ordered from smallest to largest based on their pearson chi-square statistic, which in a sense orders them from most likely to least likely (because being larger on chi-square means more extreme). So by the way we construct our regions, the fail to reject region is comprised of the m most likely $\tilde{x}$'s under the null, and the rejection region is comprised of the $M - m$ least likely $\tilde{x}$'s under the null.

It is somewhat confusing to me why we use $S(x)$. In the case where you have to compute all the $\tilde{w}$'s anyway, why not order them in terms of $\tilde{w}$ from biggest to smallest and do the same thing? That would be an EXACT ordering from most to least likely. While ordering by the chi-square would at the very least be close to ordering by $\tilde{w}$, it is not clear to me how to would show they are exactly the same. Either way, why even introduce the chi-square when we already have the $\tilde{w}$'s? It seems like it must come from cases where all of this is approximated and so the $\tilde{w}$'s are never explicitely calculated. I'm not sure...

Problem 2a.)

$$\begin{aligned} S(X) &= \sum \frac{(observed - expected)^2}{expected} \\ &= \frac{(140 - \frac{1}{3}500)^2}{\frac{1}{3}500} + \frac{(165 - \frac{1}{3}500)^2}{\frac{1}{3}500} + \frac{(195 - \frac{1}{3}500)^2}{\frac{1}{3}500} \\ &= 9.1 \end{aligned}$$

The degrees of freedom are the dimension of the unrestricted, which is 2 since p must sum to 1, minus the dimension under the null, which is 0. This gives us $df = 2$. So

$$\begin{aligned} p &= 1 - F_2(9.1) \\ &= 0.0106 \end{aligned}$$

where F_ν is cdf of χ^2_ν. Notice it is only a one-sided tail, because values close to 0 indicate the observed counts are really close to the expected under the null, so only the upper tail represents more extreme values.

Problem 2b.) $H_0 : p = (a, a, 1 - 2a)$ where $a \in [0, 1/2]$.

Problem 2c.) By invariance of MLE, we can find the MLE for a and that will determine the MLE for p. We have

$$\begin{aligned} l_x(a) &= constant + (x_1 + x_2)log(a) + (x_3)log(1 - 2a) \\ \frac{d}{da} l_x(a) &= \frac{x_1 + x_2}{a} + \frac{x_3}{1 - 2a}(-2) \\ \frac{2x_3}{1 - 2a} &= \frac{x_1 + x_2}{a} && \text{set to 0} \\ 2ax_3 &= (x_1 + x2) - 2a(x_1 + x_2) \\ \hat{a} &= \frac{x_1 + x_2}{2(x_1 + x_2 + x_3)} \\ \hat{a} &= \frac{x_1 + x_2}{2n} \end{aligned}$$

which means $\hat{p} = (\frac{x_1+x_2}{2n}, \frac{x_1+x_2}{2n}, \frac{x_3}{n})$.

Problem 2d.) The MLE under the null in this case is $\hat{p} = (.305, .305, .39)$. Then we get

$$\begin{aligned} S(x) &= \frac{(140 - (.305)500)^2}{(.305)500} + \frac{(165 - (.305)500)^2}{(.305)500} + \frac{(195 - (.39)500)^2}{(.39)500} \\ &= 2.05 \end{aligned}$$

The degrees of freedom are the dimension of the unrestricted, which is 2 since p must sum to 1, minus the dimension under the null, which is 1. This gives us $df = 1$. So

$$\begin{aligned} p &= 1 - F_1(2.05) \\ &= 0.1523 \end{aligned}$$

where F_ν is cdf of χ^2_ν.

Problem 3a.) In order for cigerettes to not have an effect for anyone, we must have the effect for non-blacks, β_2, equal zero and simultaneously the effect for blacks, $\beta_2 + \beta_4$, equal 0. This is equivalent to $\beta_2 = 0$ and $\beta_4 = 0$. For this test we have,

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad a_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where the wald statistic is

$$W(x) = (A^T\hat{\beta} - a_0)^T (A^T I_x^{-1} A)^{-1} (A^T\hat{\beta} - a_0)$$

Notice this is different then the notes because I'm using the information of n data points instead of just one, and we have $I_x^{-1} = n * I_1^{-1}$. Plugging in what we are given, this becomes $W(x) = 7.685$. The degrees of freedom are the dimension of the unrestricted, which is 4, minus the dimension under the null, which is 2. This gives us $df = 2$. So

$$\begin{aligned} p &= 1 - F_2(7.685) \\ &= 0.0214 \end{aligned}$$

where F_ν is cdf of χ^2_ν. This indicates there is evidence, at least at the .05 level, that smoking has an effect on the probability of low birthweight.

Problem 3b.) In order for the effect of cigerettes to be the same effect on blacks and non-blacks, we must have $\beta_2 = \beta_2 + \beta_4$. This is equivalent to $\beta_4 = 0$. For this test we have,

$$A = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad a_0 = \begin{bmatrix} 0 \end{bmatrix}$$

where the wald statistic is

$$W(x) = (A^T\hat{\beta} - a_0)^T (A^T I_x^{-1} A)^{-1} (A^T\hat{\beta} - a_0).$$

Plugging in what we are given, this becomes $W(x) = 0.0018$. The degrees of freedom are the dimension of the unrestricted, which is 4, minus the dimension under the null, which is 3. This gives us $df = 1$. So

$$\begin{aligned} p &= 1 - F_1(0.0018) \\ &= 0.966 \end{aligned}$$

where F_ν is cdf of χ^2_ν. This indicates there is no evidence that blacks and non-blacks are effected differently by smoking in terms of their probability of low birthweight.

Problem 4) The relative efficiency between 2 estimators is defined as the ratio of their asymptotic variances. So to find the efficiency of X_{med} relative to $\bar{x}$, we get $\frac{1/(4f_0(0)^2}{\sigma^2(f_0)} = (4f_0(0)^2\sigma^2(f_0))^{-1}$. Also note, for all distributions in the following, they are mean 0 which implies $\sigma^2(f_0) = var(f_0)$.

a.)

- $var(f_0) = 1$
- $f_0(0) = 1/\sqrt{2\pi}$
- $ARE = (var(f_0) * 4 * f_0(0)^2)^{-1} = \frac{\pi}{2}$
- This is more than one, so the mean is more efficient than the median.

b.)

- $var(f_0) = 2$
- $f_0(0) = 1/2$
- $ARE = (var(f_0) * 4 * f_0(0)^2)^{-1} = \frac{1}{2}$
- This is less than one, so the median is more efficient than the mean.

c.)

- $var(f_0) = \pi/3$
- $f_0(0) = \frac{1}{(1+1)^2} = 1/4$
- $ARE = (var(f_0) * 4 * f_0(0)^2)^{-1} = \frac{12}{\pi}$
- This is greater than one, so the mean is more efficient than the median.