ANSWER TO QUESTION ON WEIGHTED DATA.

This question contained two important errors: you were asked to show that the mean of y^* was m, when in fact it was $6m$, and you were asked to show that the sample mean of y_i/w_i was a consistent estimator of the mean of y_i/w_i (which it is) and that is is asymptotically normal (which it is not, because it has infinite variance). The answer below notes these errors and proceeds, but it was not realistic to expect the exam-taker to see around these errors under time pressure.

(5) **515a** We have observations on $\{y_i, i = 1, \ldots, N\}$. The sample has been subject to selection, meaning that there were i.i.d. draws y_j^*, w_j^* made from a population, then with probability w_j^* the observation appeared as an element (y_i, w_i) in our observed sample, while with probability $1 - w_j^*$ the observation was discarded. We are interested in the population mean of y_j^* before selection, i.e. $E[y/w]/E[1/w]$. Suppose we know the distribution of y conditional on w, and its pdf is given by

$$y_i \mid w_i \sim \frac{w_i}{m} e^{-\frac{w_i y_i}{m}} .$$

The marginal distrbution of w_j^* is Beta(2,2), i.e. w has pdf $6w \cdot (1 - w)$. This means the marginal distribution of the observed w_i in our sample is Beta(3,2) and thus that in our sample $E[1/w_i] = 2$.

(a) The population mean we are interested in is then $\frac{1}{2}E[y_i/w_i]$. Show that this is just m. Show also that the sample mean of $\frac{1}{2}y_i/w_i$ is a consistent and asymptotically normal estimator for m.

Actually, it's not just m. I lost track of a normalizing constant.

$$E[y^*] = E[y/w]/E[1/w] = 6m ..$$

Here's the derivation: The distribution given for $y \mid w$ is a $\mathrm{Gamma}(1, \mathrm{w_i/m})$ distribution, and thus has mean m/w_i. The Beta(3,2) pdf is $12w^2(1 - w)$, where 12 is the inverse of

$$B(3, 2) = \frac{\Gamma(3)\Gamma(2)}{\Gamma(5)} = \frac{2!1!}{4!} = \frac{1}{12} .$$

The unconditional expectation of y_i/w_i in the observed sample is therefore

$$\int_0^1 E[y|w] \cdot \frac{1}{w} 12w^2(1 - w)\, dw = \int_0^1 m \cdot 12(1 - w)\, dw = 6m.$$

Since y/w is i.i.d. and we have verified it has finite expectation, we can be sure its sample mean converges a.s. to its expectation, and therefore its sample mean times $\frac{1}{2}$ converges to the target.

However, the $\mathrm{Gamma}(1, w/m)$ has variance m^2/w^2. Thus $\mathrm{Var}[y/w \mid w] = m^2/w^4$. Since the Beta(3,2) density is $O(w^2)$ as $w \to 0$, it is clear that m^2/w^4 times the Beta(3,2) density is not integrable, and thus that the variance of y/w is infinite.

Date: May 17, 2018.

(b) Find the posterior mean of m under a flat prior, and show that it also is consistent and asymptotically normal.

The likelihood is

$$\prod_1^N w_i m^{-N} e^{-\frac{\sum y_i w_i}{m}} \prod_1^N \big(w_i(1-w_i)\big).$$

The part of this that depends on m is proportional to an inverse gamma distribution with scale parameter $\sum_i y_i w_i$ and shape parameter $N-1$. The posterior mean for m is therefore

$$\frac{\sum y_i w_i}{N-2}.$$

$y_i w_i$ can easily be seen to be of finite variance and finite mean. Since $E[y \mid w] = m/w$, $E[yw \mid w] = m$. Thus the sample mean of $y_i w_i$, multiplied by 6, is a consistent and asymptotically normal estimate of our target parameter, and asymptotically equivalent to the posterior mean of m..

(c) Which estimator is better? Explain why.

Obviously, since it has finite variance, the estimator based on the posterior mean of m is better than the estimator based on the sample mean of y/w. Another way to see this is that y has bigger variance when w is smaller, so a good estimator should put more weight on observations with large w.