FIRST-HALF FINAL EXAM

Your exam answers should reach me by email (`sims@princeton.edu`) by 9PM Friday October 9. The exam is meant to take no more than two hours, and exam answers that look like they took longer than that will be downgraded. However there is no strict limit. Answer all four questions. Total available points are 105.

(1) **15 points** Stein proposed an estimator for β in the standard normal linear regression model $y = X\beta + \varepsilon$ with $\varepsilon \mid \{X, \beta\} \sim N(0, \sigma^2 I)$ that has lower expected mean squared error $E[\|\beta - \hat{\beta}\|^2$ than the OLS estimator whenever the length of the β vector exceeds 2.

Does the complete class theorem imply that Stein's estimator must be a Bayes estimator for some prior? Why or why not?

The complete class theorem states that, subject to some regularity conditions, every Bayes estimator is admissible and every admissible estimator is Bayes. By showing that OLS under standard assumptions can be dominated (have lower mean squared error for every possible true β) by another estimator, Stein showed that OLS is neither Bayes nor admissible. However, Stein did not show that his proposed estimator could not be dominated by any other estimator, so it is not necessarily true that his estimator is Bayes or admissible (and in fact it is neither).

We had not discussed Stein's result in class and some read the question as referring to a fictional "Stein" who must be mistaken. OLS is "best linear unbiased" which means it has lower mean squared error than any other linear, unbiased estimator, and under normality assumptions it also is better than any unbiased estimator by this criterion. But Stein's estimator is biased, and better than OLS with the $\left\|\beta - \hat{\beta}\right\|^2$ criterion.

Note that Stein's estimator dominates OLS for $(\beta - \hat{\beta})' Q (\beta - \hat{\beta}$ with $Q = I$ as the error criterion. It does not dominate for other choices of Q. The "best linear unbiased" result holds for any quadratic criterion with positive definite Q.

(2) **45 points** Suppose you have a sample of 20 i.i.d. observations of a random variable Y that is equal to 1 with probability p and is otherwise zero. In your sample, there are 7 ones and 13 zeros.

(a) Using a flat prior on p, find a 95% posterior HPD credibility set for p.
The likelihood is $p^7(1-p)^{13}$, which is proportional to a $\text{Beta}(8, 14)$ pdf. This is a single-peaked function on $(0, 1)$, so the HPD interval is found by looking for a pair (a, b) such that under that distribution the probability of the interval is .95 and the pdf has the same value at both a and b. So if $F(x)$ is the cdf of $\text{Beta}(8, 14)$, $p()$ is its density function, and $Q(p)$ is its quantile function (i.e. the inverse of the cdf), we are

Date: November 30, 2020.

looking for a such that

$$p(a) = p(Q(.95 + F(a)))\,.$$

If we find an a that solves that equation, the interval is $(a, Q(.95 + F(a)))$. You could give this function to a univariate equation solver, or just find it by hand trying values of a between 0 and $Q(.05)$. I did it the latter way, arriving at $(.1734, .5606)$

(b) Though 20 is a small enough sample size to make use of the asymptotic normality of the sample mean of Y possibly a doubtful approximation, go ahead and construct a 95% confidence interval for p based on the asymptotic normal distribution for the sample mean of the Y's.
The sample mean is .35, and the sample standard deviation is .4894, so the estimated standard deviation of the mean is $.4894/\sqrt{20} = .1094$, making the approximate 95% interval $.35 \pm 1.96 * .1094 = (.14, .56)$.

(c) Suggest a 5% significance level frequentist test, without using any asymptotic approximations, of the null hypothesis that p takes on a particular value p^*. Note that number n of 1's in an i.i.d. sample of size N when the probability of a 1 is p is the binomial distribution. In R, **`pbinom`**`()` and **`dbinom`**`()` could be helpful here.
If the true value of p is p^*, we could reject H0: $p = p^*$ when (in R notation) **`pbinom`**`(n, size=20, prob = pstar)` is outside the interval (.025, .975). (Any subinterval of (0,1) of length .95 would work.)

(d) Explain how your suggested test could be used to construct a confidence interval for p that does not rely on asymptotic approximations (It is probably even possible to calculate such an interval under the exam time constraints but this is not required.)
Define the confidence set as the collection of p^* values that are not rejected for the value of n (the number of ones) in the sample. For our hypothetical sample of size 20 with 7 ones, the interval is (.20, .60). R code that solves this:

(3) **20 points** The sum of squared deviations from its mean of the `educ` variable in the AK data set is 3547667.7. The sum of squared residuals from a regression of `educ` on a complete set of the 10 dummies for the `yob` (year of birth) values is 3537568.4. The total number of observations is $N = 329509$. Use these values to construct a 0.1% level F test of the hypothesis that the population means of years of schooling are the same across birth cohorts. Compare the results of this test to what emerges from the BIC criterion.
The F statistic is

$$\frac{(RSSR - RSSU) * (N - 10)}{(RSSU * 9)} = 104.5\,,$$

where $RSSR$ is the sum of squared residuals from the smaller model, RSSU is the sum of squared residuals for the larger model, N is sample size, and 9 is the difference in number of parameters between the two models. We usually reject when F is large, so in this case when the F statistic exceeds the .999 quantile, which for an $F(9, 329499)$ distribution is

3.098. So we reject the null hypothesis overwhelmingly. BIC would instead compare the F statistic to $\log(N) = 12.7$ In this case BIC also strongly favors the larger model.

(4) **25 points** Suppose you have i.i.d. observations on a variable Y that is distributed as $N(\mu, 1)$, but all observations in which $Y_i < 0$ are recorded instead as $Y_i = 0$. Suggest how, by treating the unobserved $Y_i < 0$ values as unknown parameters, you could use Gibbs sampling to sample from the joint posterior on μ and the unobserved negative Y_i's.

The joint pdf of all the Y_i values, conditional on μ, has log pdf (ignoring constants)

$$-\tfrac{1}{2}\sum_{i=1}^{N}(Y_i - \mu)^2 .$$

Conditional on all the Y's, including those not observed, under a flat prior $\mu \sim N(\bar{Y}/N, 1/N)$. So start the algorithm with guessed values of the unobserved Y's, draw from this conditional distribution for μ. Conditional on μ and the observed Y's, the unobserved Y's are independent of each other and truncated normal. That is, one can draw each unobserved Y_i by drawing from a $N(\mu, 1)$ repeatedly until a draw that satisfies $Y_i < 0$ is found. Then go back to drawing μ conditional on all the Y's, etc. Since in each case, whether drawing the unobserved Y_i's or drawing μ, we are drawing from the exactly correct conditional posterior, this procedure has the posterior on μ and the unobserved Y_i's as a fixed point and will converge to that target distribution.

If the sample is small enough, e.g. if there is only one observed Y_i, we might worry about the flat prior. The prior will matter in a small sample, and a flat prior on the whole real line is improper (not integrable), which can cause problems for MCMC algorithms. We could implement a $N(m, s^2)$ conjugate prior on μ in this model by including a dummy observation $Y_{N+1} = m$ and weighting the observation with weight $w_{N+1} = 1/s$. This conjugate prior part of the answer was not expected to be worked out in any detail under exam time pressure.