

## Statistics 406 Problem Set 9

Due in lab, Tuesday December 5th

1. Here we will see how  $\hat{\sigma}^2$  can be biased downward when the data are positively correlated. We will work with data sets of size  $p = 20$ , and consider two different ways to produce positively correlated data.
  - (a) Start with  $p$  iid standard normal values, then replace a subset of  $k$  values with an exact copy of a different subset of  $k$  values. Do this for  $k = 0, \dots, 8$ , and use simulation to estimate the bias of the sample variance for each value of  $k$ .
  - (b) Start with  $p$  iid standard normal values  $Z_1, \dots, Z_p$ , then take one additional independent standard normal value  $X$ , and add it onto each of the  $p$  values. The data  $Z_1 + X, \dots, Z_p + X$  are positively correlated. You can calculate the bias of  $\hat{\sigma}^2$  in this case by mathematical analysis rather than simulation. To do this, first calculate the population variance of these data. Then consider how  $\hat{\sigma}^2$  for the  $Z_i + X$  data differs from the sample variance that you would get from just the  $Z_i$ . From these two considerations you can determine the bias of  $\hat{\sigma}^2$ . Once you have done this, use simulation to confirm your findings.
2. Use simulation to show that the MSE of  $\hat{\sigma}^2$  is greater when using either of the two types of positively correlated data in problem 1, relative to a set of  $p$  iid values having the same mean and variance. For both of the two cases of dependence, comment on whether the increased MSE is coming mainly from bias, variance, or both.