Problem Set 3.

Problem 1. You have been sent by email the data Y_i of 500 persons, which are the lengths of stay described in part (ii) of the previous problem set.

(i) By equating the expressions for $E(Y_i|I_i = 1, \theta)$ and var$(Y_i|I_i = 1, \theta)$ (in terms of θ_1 and θ_2) to the sample mean and variance of your data Y_i, find estimates of θ_1 and θ_2. This is called a " moment estimation" method.

(ii) Consider the likelihood function of your data as in part (ii) of the previous problem set. For the MLEs of that likelihood, there is no known closed form, but the MLEs can be found numerically by maximization algorithms. By using the algorithm "optim" (see help(optim)) in the statistical environment "R", or any other appropriate algorithm and/or programming of your choice, find a stationary point for the mean θ_1 and variance θ_2 of the length of stay in the target population.

Note: Good starting values are important for stability of the algorithm; use the moment estimates from part (i) above as starting values for the algorithm for finding the MLE.

(iii) You may get some warning messages while using the algorithm, which indicate possible numerical instability of the algorithm. To make sure the converged values of the algorithm are a maximum, check that the second derivative matrix of the log-likelihood, also called the Hessian matrix, is negative definite (as defined in class) when evaluated at the converged values of the algorithm.

Note: you can obtain the Hessian matrix by setting the option "hessian"=T in the algorithm "optim". You can use the result that the Hessian matrix is negative definite if and only if it satisfies the conditions a.-c. of example 7.2.12 of the text (p. 322).

(iv) Find the MLE of the 95th percentile of the distribution of lengths of stay in the target population. (Hint: you can use the invariance property of MLEs, and a numerical method to do the actual computation. For example, check the function "qgamma()" in "R").

(v) Each of your colleagues has been given a different independent set of 500 people from the survey. Using each other's ML estimates, report an estimate of the variance of the MLEs $\hat{\theta}_1$ and $\hat{\theta}_2$, and an estimate of the covariance between $\hat{\theta}_1$ and $\hat{\theta}_2$.

Problem 2. We wish to study the level of a specific radioactive particle in an environment, using a counter. The number X of particles counted by the counter in a time interval of 1 min. is assumed to follow a Poisson distribution. You have been sent 50 measurements, $X_1, ..., X_{50}$, of counts at different 1 min; assume the 50 measurements are i.i.d. from Poisson(μ).

(i) Find the MLE of μ. Show that the MLE is a minimal sufficient statistic for μ.

(ii) Suppose we are interested in $g(\mu) = \text{pr}(X = 0 \mid \mu)$. Find the MLE of $g(\mu)$. Do you think this MLE is biased or unbiased for $g(\mu)$ and why ?

(iii) By considering $X_i^* = 1(X_i = 0)$, $i = 1, ..., 50$, where $1()$ is the indicator function, find an unbiased estimator (and estimate) of $g(\mu)$.

(iv) Find the distribution of $pr(X_1 \mid \bar{X}, \mu)$. (Here, X_1 indicates the first measurement as given to you in random order, and is not necessarily the smallest measurement).

(v) Use your estimator in (iii), your result in (iv) and "Blackwellization", to obtain an unbiased estimator (and estimate) for $g(\mu)$ that has smaller variance than the one in (iii). Is this the minimum variance unbiased estimator for $g(\mu)$, and why or why not ?