Exercises for MA 413− **Statistics for Data Science**

This sheet will be handed out lecture 16/11/2019*, overlapping a little with material from the previous week and later material.*

1. Let $Y_1, \ldots Y_n \overset{\text{iid}}{\sim}$ Gamma$(3, \lambda)$, for $n \geq 2$.
   (a) Determine the MLE of λ.
   (b) What is the expectation of $\hat{\lambda}$ the MLE?
   (c) Construct an unbiased estimator of λ, let us denote this by $\tilde{\lambda}_c = c\hat{\lambda}$.
   (d) Find the MSE of $\hat{\lambda}$ and of $\tilde{\lambda}_c$, and determine which estimator is preferred.
   (e) Would another loss function be more appropriate? What does this imply?

2. Assume you observe $Y_1, \ldots Y_n$ from the Bernoulli (θ), independently. Construct the MLE estimator. Construct an alternative estimator of λ, let us denote this by $\tilde{\lambda}_c = c\hat{\lambda}$, for some unspecified c.
   (a) Find the MSE of the MLE $\hat{\theta}$.
   (b) For a fixed but unknown value of c find the MSE of $\tilde{\lambda}_c$
   (c) Determine the optimal value of $c \in \mathbb{R}^+$.

3. Consider using the asymmetric loss function

$$\mathcal{L}(a, b) = a/b - 1 - \log(a/b)$$

   For the problem of Qn 1 write down the expected loss for the MLE as well as the bias corrected estimator.

4. Let $X_1, \ldots, X_n$ have a Poisson distribution with mean μ. Our null hypothesis is $\mu = \mu_0$ and our alternative hypothesis is $\mu = \mu_1$, where $\mu_0 < \mu_1$. Describe how to test the hypothesis.

5. Let $X_1, \ldots, X_n$ have a Gaussian distribution with unknown mean μ, and a known variance σ_0^2. Our null hypothesis is $\mu = \mu_0$ and our alternative hypothesis is $\mu = \mu_1$, where $\mu_0 < \mu_1$. Describe how to test the hypothesis.

6. Consider Stein's unit variance log-likelihood for $Y_i \sim N(\mu_i, 1)$, independently,

$$\ell(\boldsymbol{\mu}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\{Y_i - \mu_i\}^2$$

   (a) Determine $\nabla_{\mu_i}\ell(\boldsymbol{\mu})$. Solve for $\nabla_{\mu_i}\ell(\boldsymbol{\mu}) = 0$
   (b) Show that this solution is a maximum.
   (c) Determine the expected square loss of that estimator.

(d) Define the estimator (for some specified $\theta \in \mathbb{R}^+$) :

$$\widetilde{\mu}_i = \begin{cases} Y_i & \text{if} \quad |Y_i| > \theta \\ 0 & \text{if} \quad |Y_i| \leq \theta \end{cases} \tag{1}$$

Show that the expected loss of $\widetilde{\mu}_i$ is

$$\mathcal{L}\left(\widetilde{\mu}_i, \mu_i\right) = \mathrm{E}\left\{\left(\widetilde{\mu}_i - \mu_i\right)^2\right\} \tag{2}$$

$$= 2\int_\theta^\infty \left(y-\mu_i\right)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\mu_i)^2} dy + 2\mu_i^2 \left\{\Phi\left(\theta-\mu_i\right) - \left(\Phi\left(-\mu_i\right)\right\}\right. . \tag{3}$$

We also define

$$I\left(\mu_i, \theta\right) = \left[2\left(y-\mu_i\right)\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\mu_i)^2}\right]_\theta^\infty - 2\int_\theta^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\mu_i)^2} dy \tag{4}$$

Simplify these expressions (3) and (4) additionally.

(e) Combining these expressions (3) and (4) we get another for the estimation risk of $\widetilde{\mu}_i$. Plot this using matlab as a function of μ_i taking i) $\theta = 0.1$, ii) $\theta = 1$ and iii) $\theta = 2\log(n)$. How do these cases vary? What does this tell us about the estimation problem? It is reasonable to use unscaled numbers (in σ) as the variance is fixed to unity.

7. Let $Y_1, \ldots, Y_n$ be iid uniform random variables on the interval $(0, \phi)$. Let M_n be the maximum of the n random variables. We wish to test $H_0 : \phi = 1/2$ versus $H_1 : \phi > 1/2$. The Wald test is not appropriate as M_n does not converge to a Gaussian. Suppose that we decide to test this hypothesis by rejecting H_0 when $M_n > r$.

   (a) Determine the power function of this test.

   (b) What choice of r will enable us to have a test of size 0.1?

   (c) Please determine the p -value of a sample size $n = 22$ with $M_n = 0.47$. What conclusion would you make about H_0?

   (d) In a sample of $n = 20$ with $M_n = 0.51$, what is the p -value? What conclusion about H_0 would you make?

8. Let $Y_1, \ldots, Y_n$ be independent $N(\mu, 1)$ random variables. Consider testing the hypothesis

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu = 1$$

   (a) Let the rejection region be of the form $\{\underline{Y} : T(\underline{Y}) > c\}$ (for $c > 0$) , where T is the sample mean. Find c so that the test has size α.

   (b) Find $\beta(1)$.

   (c) How does $\beta(1)$ behave as $n \to \infty$?

9. Assume we take a random sample $Y_1, \ldots, Y_n$ from a Poisson (μ) distribution. For some fixed and given μ_0 describe the Wald test for

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

10. Assume you have $Y_1, \ldots, Y_n \sim N\left(\mu, \sigma^2\right)$. For a fixed and given $\sigma_0 > 0$, describe the Wald test for

$$H_0 : \sigma = \sigma_0 \quad \text{vs} \quad H_1 : \sigma \neq \sigma_0$$

You may assume μ known and given.