PB HLTH C240C / STAT C245C

Assignment #2

Due: Thursday, October 16th, 11:59pm

Question 1. MLE of the parameter of a Poisson distribution. Consider the following random sample from a Poisson distribution $X \sim \text{Poi}(\lambda)$.

```
##  [1] 1 2 6 2 3 3 2 3 2 1 5 2 2 4 2 0 2 1 2 1
```

a) Write the log likelihood and plot it for a sensible choice of values. Keep in mind that the Poisson distribution is defined only for $\lambda > 0$.

b) Find the MLE of λ both analytically and numerically.

c) Plot the likelihood function, considering the parametrization $\theta = g(\lambda) = \log \lambda$. What is the parameter space in the new parametrization?

d) Find numerically the MLE of θ, defined in the previous point. Verify numerically that $\theta_n = g(\lambda_n)$.

Question 2. MLE of the shape and rate parameters of a Gamma distribution. Consider the following random sample from a Gamma distribution $Y \sim \text{Gamma}(\alpha, \lambda)$, with density function $f(y; \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha y^{\alpha-1} e^{-\lambda y}$.

```
##  [1] 0.15812 0.30070 0.48016 0.49813 0.20042 0.26716 0.80124
##  [8] 0.10914 0.57169 0.83686 1.57027 0.10458 0.58490 1.14454
## [15] 0.61595 0.28155 0.13236 0.36252 0.08614 0.27907 0.46010
## [22] 0.03824 0.76581 0.30369 0.42404 0.57530 0.26987 0.22416
## [29] 0.07673 1.09659
```

a) Write the log likelihood and plot it for a sensible choice of values. *Hint:* since the parameter is two-dimensional use `contour`, `image`, `persp` or similar functions to plot the likelihood.

b) Find the MLE of $\psi = (\alpha, \lambda)$.

Question 3. Regression. Consider a data structure $X = (W, Y) \sim P$, where $W \in \mathbb{R}$ is a scalar covariate and $Y \in \mathbb{R}$ a scalar outcome. The parameter of interest is the regression function $\psi(W) = E[Y|W]$. A learning set of $n = 200$ IID $(W, Y) \sim P$ pairs and a test set of $N - n = 2,000 - 200$ IID $(W, Y) \sim P$ pairs were generated independently from a mystery data generating distribution P. The data are available at http://daviderisso.wordpress.com/teaching/ph240c14/pb-hlth-240c-assignments/.

a) Using the learning set, estimate the regression function $\psi(W) = E[Y|W]$, using *linear regression* with polynomial functions of varying degree (`lm` function in R) and *local robust weighted regression* with varying span (`loess` function in R). Provide graphical displays of the various fits.

b) Display the *learning set empirical risk* and the *test set empirical risk* of the candidate estimates from part a) for L_2 loss functions. Specify which estimate is optimal.