Assignment 3, due February 11

Corrections: (none yet)

1. Calculate the Fisher information for a univariate normal that determines the accuracy of the maximum likelihood estimate of μ and of $v = \sigma^2$ (two separate calculations). Use this to determine the asymptotic variance of the maximum likelihood estimate of μ and σ^2 from n independent samples. Show that, in the case of μ, this agrees with the direct estimate of the variance of

$$\widehat{\mu}_{\text{ml}} = \overline{X} ,$$

and

$$\widehat{\sigma^2}_{\text{ml}} = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 .$$

For the second, we know that the variance of the right side is related to the variance of a χ^2_{n-1} random variable.

2. This exercise concerns the tails of the t random variable. Let $V = \sum_{k=1}^{n} Z_k^2 \sim \chi^2_n$, and Z be an independent standard normal. Then define

$$T = \frac{Z}{\sqrt{\frac{1}{n}V}} . \qquad (1)$$

(a) Correct the theoretical treatment in class. Let $f_n(t)$ be the probability density for T. Let U be the denominator in (1). Show that $\Pr\left(U < \frac{1}{t}\right) \approx Ct^{-q}$ for large t, and identify q. Use this to argue that

$$1 - F_n(t) = \Pr\left(T > t\right) \approx Ct^{-q} , \qquad (2)$$

for the same q. Here $F_n(t)$ is the cumulative distribution function for the t random variable with n degrees of freedom, i.e. the random variable defined by (1).

(b) Write a program in Matlab that generates a large number of t random variables: T_k for $k = 1, \ldots, L$. Use these samples to compute and plot the *empirical* CDF

$$F_n^e(t) = \frac{1}{L}\#\left\{k \mid T_k < t\right\} . \qquad (3)$$

The superscript is for "empirical". The right side is the number of samples so that $T_k < t$. Clearly $F_n^e(t) \to 0$ as $t \to -\infty$ and $F_n^e(t) \to 1$

as $t \to \infty$. Make a plot of $F_n^e(t)$ for t in the range $-4 \le t \le 4$, $n = 4$ and $n = 20$, and $L = 10^6$ or $L = 10^7$ (depending on how big your computer allows L to be). The large n graph should look like $N(t)$, the cumulative normal. The smaller n graph should have a wider variation.

(c) Estimate q in (2) by making a plot of $\log(1 - F_n^e(t))$ against $\log(t)$ for $t > 3$ and $n = 4$. Show that q is the slope of the plot for large t. Try to estimate this slope from the plot. Note, that the plot gets the most interesting (for estimating q) right where the data ends. Does your estimate of q here agree with your theoretical prediction from part (a)?

3. Consider a different fat tail model with probability density

$$f(x,p,\xi) \;=\; \begin{cases} C\,(\,1\,+\,\xi x\,)^{-p} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases} \qquad (4)$$

(a) Calculate C as a function of the parameters p and ξ. Do this using the requirement that $\int_0^\infty f(x)dx = 1$.

(b) Suppose there is a data set $\vec{X} = X_1, \ldots, X_n$ that consists of n independent samples of f. Write the log likelihood function as a function of the data set and the parameters p and ξ.

(c) Differentiate with respect to ξ and p to find the two equations you would have to solve to find the maximum likelihood estimates $\widehat{p}_{\text{ml}}$ and $\widehat{\xi}_{\text{ml}}$.

(d) Suppose ξ is known. Write the integral that represents the Fisher information for the parameter p. It is possible to calculate this integral in closed form, but I am not asking you to do so.