Maximum Likelihood Estimation*

Worksheet 18

1. Let X_1 and X_2 be two independent measurements of some unknown value μ. X_1 has much higher variance than X_2.
   (a) Should your estimate for μ be closer to X_1, to X_2 or be the simple average $(X_1 + X_2)/2$? Explain your choice.
   (b) To test this, let X_1 and X_2 be normal random variables with mean μ and respective variances $\sigma_1^2 = 1/2$ and $\sigma_2^2 = 1/20$. Thus, the densities are
   $$f_{X_1}(x_1) = \frac{1}{\sqrt{\pi}} \exp -(x_1 - \mu)^2, \quad f_{X_2}(x_2) = \frac{\sqrt{10}}{\sqrt{\pi}} \exp -10(x_2 - \mu)^2.$$
   Give the likelihood for the pair X_1, X_2.
   (c) Find the log of the likelihood and use this to find the maximum likelihood estimator $\hat{\mu}$ for μ.
   (d) Let $x_1 = 3.11$ and $x_2 = 3.22$. Find the estimate $\hat{\mu}$.
   (e) Does this answer support your claim in part (a)?
2. Loss of property for insurance purposes is sometimes modeled as a Pareto distribution. An insurance company offiers two insurance policy. If we take the claim amounts (in thousands of dollars), this yields a density of
   $$f_X(x|\beta) = \frac{\beta 5^\beta}{x^{\beta+1}}, \quad x \geq 5. \quad \text{with mean } \mu_X = \frac{5\beta}{\beta - 1} \quad \text{and standard deviation } \sigma_X = \frac{5}{\beta - 1}\sqrt{\frac{\beta}{\beta - 2}}$$
   for a minimum claim of 5 thousand dollars.
   (a) Give the likelihood function for n independent claims for the first insurance policy.
   (b) Find the maximum likelihood estimate $\hat{\beta}$ for β in terms of claims $x_1, x_2, \ldots, x_n$.
   (c) The claims for the insurance policies can be downloaded with the commands
   ```
   > claims<-read.csv("http://math.arizona.edu/~jwatkins/claims5.csv")[,1]
   ```
   for claim data under the insurance policy. Give the maximum likelihood estimate for these data.
   (d) Use numerical summaries, a histogram, and an empirical cumulative distribution function to assist you in providing a description of the data.
   (e) Use the estimated parameter value to estimate the mean and standard deviations of the claims.
   (f) How well does this match the values from the data?