18.650. Statistics for Applications
Fall 2016. Problem Set 4

Due Friday, Oct. 7 at 12 noon

NOTE: there was a typo in the definition of the Log-normal pdf

Problem 1 Maximum likelihood and Fisher information

For each of the following distributions, compute the maximum likelihood estimator for the unknown (one or two dimensional) parameter, based on a sample of n i.i.d. random variables with that distribution. In each case, is the Fisher information well defined ? If yes, compute it.

1. $\text{Ber}(p)$, $0 < p < 1$;
2. Poisson with parameter $\lambda > 0$:

$$\mathbb{P}_\lambda[X = k] = e^{-\lambda}\frac{\lambda^k}{k!}, \quad \forall k \in \mathbb{N};$$

3. Exponential with parameter $\lambda > 0$, with density

$$f_\lambda(x) = \lambda e^{-\lambda x}, \quad \forall x > 0;$$

4. Gaussian with parameters $\mu \in \mathbb{R}, \sigma^2 > 0$ (recall that the second parameter is σ^2, not σ);
5. Shifted exponential distribution with parameters $a \in \mathbb{R}, \lambda > 0$ with density

$$f_{a,\lambda}(x) = \lambda e^{-\lambda(x-a)}\mathbb{1}_{x \geq a}, \quad \forall x \in \mathbb{R};$$

6. Log-normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, with density

$$f_{\mu,\sigma^2}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2}, \quad \forall x > 0.$$

Problem 2 Method of moments

For each distribution of Problem 1, find the moment estimator for the unknown parameter, based on a sample of n i.i.d. random variables.

Problem 3 Censored data

In a given population, n individuals are sampled randomly, with replacement, and each sampled individual is asked whether his/her salary is greater than some fixed threshold z. Assume that the salary of a randomly chosen individual has the exponential distribution with unknown parameter λ. Asking whether the salary overcomes a given threshold rathen than directly asking for the salary increases the number people that are willing to answer and decreases the number of mistakes in the collected answers. Denote by $X_1, \ldots, X_n$ the binary responses ($X_i \in \{0, 1\}, i = 1, \ldots, n$) of the n sampled individuals.

1. What is the distribution of the X_i's ?
2. Let $\bar{X}_n$ be the proportion of sampled individuals whose response was 1 (corresponding to *Yes*). Prove that $\bar{X}_n$ is asymptotically normal and compute the asymptotic variance.
3. Find a function f such that $f(\bar{X}_n)$ is a consistent estimator of λ.
4. Prove that $f(\bar{X}_n)$ is asymptotically normal and compute the asymptotic variance.
5. What equation must z satisfy in order to minimize the asymptotic variance computed in Question 4 ? Write this equation in the form $g_\lambda(z) = z$, where g_λ is a function that depends on the unknown parameter λ.
6. Let $Y_1, \ldots, Y_n$ be the salaries of the n sampled people.
   a) If one could actually observe $Y_1, \ldots, Y_n$, what would be the statistical model ?
   b) In that case, what would be the Fisher information (as a function of the unknown parameter λ ? Denote it by $I_Y(\lambda)$.
   c) In the model where only the X_i's are observed (with fixed threshold z), what is the Fisher information ? Denote it by $I_X(\lambda)$.
   d) Compare $I_Y(\lambda)$ and $I_X(\lambda)$: Which one is the largest ? How do you interpret this fact ?