Homework 7 (due 18/Nov/2013)

Try all the exercises. Submit only those marked with an asterisk (*).

1. (*) Let $X_1, \ldots, X_n$ be i.i.d. samples from a parametric family of discrete distributions. In each of the following cases, find the MLE for the unknown parameter(s).

1. X_i are i.i.d. $\text{Bin}(N, p)$ where N is known and p is unknown.
2. X_i are i.i.d. $\text{Pois}(\lambda)$ where λ is unknown.
3. X_i are i.i.d. $\text{Geo}(p)$ where p is unknown.

2. Let $X_1, \ldots, X_n$ be i.i.d. samples from a parametric family of densities. In each of the following cases, find the MLE for the unknown parameter(s).

1. X_i are i.i.d. $\text{Gamma}(\nu, \lambda)$ where ν is known and λ is unknown.
2. X_i are i.i.d. $\text{Unif}[a, b]$ where a, b are unknown.
3. X_i are i.i.d. $N(\mu, \sigma^2)$ where μ, σ^2 are unknown.

3.

1. Let $X_1, \ldots, X_n$ be i.i.d. $\text{Unif}([0,1])$. If M_n is a sample median, show that $\mathbf{P}\{|M_n - \frac{1}{2}| > \delta\} \to 0$ for any $\delta > 0$, as $n \to \infty$.
2. If X_i are i.i.d from some density $f(x)$ (assume that the median is uniquely defined), deduce that the sample median M_n gets close to the population median $\mathbf{m}$ in the same sense, i.e., $\mathbf{P}\{|M_n - \mathbf{m}| > \delta\} \to 0$ for any $\delta > 0$, as $n \to \infty$.
3. More generally, for any $0 < q < 1$, show that the sample q quantile $M_n^{(q)}$ is close to the population q-quantile in the same sense.

[**Hint:** In the first part, observe that $M_n \le t$ if and only if more than half of the X_is are below t. As for the second part, try to deduce it from the first instead of re-doing the proof all over again!]

4. Let $X_1, \ldots, X_n$ be i.i.d. $\text{Exp}(\lambda)$. Let $\theta = \log \lambda$. Let $\gamma = \int_0^\infty \log t \; e^{-t} \, dt$.

1. Show that $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n (\gamma - \log X_i)$ is an unbiased estimate for θ.
2. Compute the m.s.e of $\hat{\theta}$.
3. Explain how you would give an $(1-\alpha)$-confidence interval for λ, based on $\hat{\theta}$. [**Hint:** If $X \sim \text{Exp}(\lambda)$, the distribution of $\log X + \log \lambda$ does not depend on λ.]

5. (*) In `http://math.iisc.ernet.in/~manju/UGstatprob/newcomb_lightspeed.txt` you will see the data from Simon Newcomb's experiment on the time taken (in nanoseconds) by light to travel 7442 meters at sea level.

1. Compute the sample mean and sample standard deviation.
2. Assuming normal distribution for the data, compute a confidence interval for the time taken. What confidence interval does it give for the speed of light (in meters per second)?
3. Repeat the same after dropping the smallest two measurements (declared 'outliers').

6. This is the description of the data given in `http://math.iisc.ernet.in/~manju/UGstatprob/horsekicks.txt`. In each year from 1875 to 1894, the number of cavalrymen who died due to horse-kicks in the Prussian army of the time was counted. The data was collected in 14 different army-units (of equal size), which is what is indicated in the 14 columns following the year column.

Assume that the number of deaths per army-unit per year is a random variable having a Poisson(λ) and that the number of deaths in different units or in different years are independent.

Estimate λ from the given data. Compute the expected frequencies of deaths per units per year from your estimate (and compare with the actual figures).

Quite challenging problems/theorems - optional!

We give two problems addressing two issues that we did not consider in class. One is that in many examples (eg., i.i.d. Exp(λ) data), the sample mean is the UMVU (*uniformly minimum variance unbiased estimate*). Second is that the maximum likelihood estimate is a reasonable choice, at least in the sense that for large sample of data, the MLE is close to the actual value of the parameter with high probability.

7. Let $f_\theta(x)$ is a collection of densities parameterized by a real number θ. Suppose $X_1, X_2, \ldots$ be i.i.d. samples from f_{θ_0} where θ_0 is fixed (we pretend it is unknown and give estimates for it). Let $\hat{\theta}_n$ be the MLE based on $X_1, \ldots, X_n$, i.e., $\hat{\theta}_n$ maximizes $\ell_n(\theta) = \sum_{i=1}^n \log f_\theta(X_i)$ (we write ℓ_n for simplicity but note that it depends on $X_1, X_2, \ldots, X_n$).

1. Fix Define $\ell^*(\theta) = \mathbf{E}_{\theta_0}[\log f_\theta(X_1)]$. Show that $\ell^*(\theta) < \ell^*(\theta_0)$ with equality if and only if $\theta = \theta_0$. (This last statement is true only if we assume that the densities f_θ are distinct for distinct values of θ, which is a very reasonable assumption!). [**Hint:** For any two densities f and g, show that $\int_{\mathbb{R}} \log\left(\frac{f(x)}{g(x)}\right) g(x)dx \leq 0$.]
2. Show that $\frac{1}{n}\ell_n(\theta) \xrightarrow{P} \ell^*(\theta)$ for all θ where the convergence is in the sense that $\mathbf{P}\{|\frac{1}{n}\ell_n(\theta) - \ell^*(\theta)| \geq \delta\} \to 0$ as $n \to \infty$, for every $\delta > 0$.
3. $\hat{\theta}$ maximizes $\frac{1}{n}\ell_n(\theta)$ and θ_0 maximizes $\ell^*(\theta)$. Further the two functions $\frac{1}{n}\ell_n(\theta)$ and $\ell^*(\theta)$ are close to each other. Convince yourself that under some conditions (but not always) this implies that $\hat{\theta}$ and θ_0 must be close to each other. [**Note:** Obviously this last part is vague. The point is that one can impose various conditions on the densities f_θ which ensure that it works. It is enough to get the heuristic idea here.]

8. Let $X_1, \ldots, X_n$ be i.i.d. Ber(p). We want to show that among all unbiased estimates of p, the one with least variance (for any value of p) is $\overline{X}_n$. Let $T : \{0,1\}^n \to [0,1]$ be any unbiased estimate (i.e., if we see the data $(X_1, \ldots, X_n)$, the guess for p would be $T(X_1, \ldots, X_n)$).

1. Define $S(x_1, \ldots, x_n) = \frac{1}{n!} \sum_{\pi \in S_n} T(x_{\pi_1}, \ldots, x_{\pi_n})$, i.e., permute the arguments x_is in all possible ways and average out the values of T obtained. Show that
   (a) $S(X_1, \ldots, X_n)$ is an unbiased estimate of p.
   (b) $S(X_1, \ldots, X_n)$ depends on $\overline{X}_n$ only. That is, $S(x_1, \ldots, x_n) = S(y_1, \ldots, y_n)$ if $\bar{x} = \bar{y}$.
   (c) $\mathrm{Var}(S(X_1, \ldots, X_n)) \leq \mathrm{Var}(T(X_1, \ldots, X_n))$.
2. From the second part above, we may write $S(X_1, \ldots, X_n)$ as $g(\overline{X}_n)$ for some function $g : [0,1] \to [0,1]$. By the unbiasedness, $\mathbf{E}_p[g(\overline{X}_n)] = p$ for all $p \in [0,1]$. Show that this implies that $g(\overline{X}_n) = \overline{X}_n)$. [**Hint:** Recall that $X_1 + \ldots + X_n$ has binomial distribution.]
3. Conclude that $\overline{X}_n$ is the UNMVU for this problem.

[**Remark:** The proof that $\overline{X}_n$ (or other specific estimates) are UMVU in other problems is somewhat more difficult, although the same as above once one understands conditional probability well.]