**PROBABILITY AND STATISTICS - HOMEWORK 7**

DUE DATE: 9/NOV/2016 (IN TUTORIALS)

**Homework problems are marked blue (but trying all problems is strongly recommended).**

**1.** Let $X_1, X_2, \dots$ be i.i.d. Uniform$[1,2]$ distribution. Let $S = X_1 + \dots + X_{100}$. Give approximate quantiles at levels $0.01, 0.25, 0.5, 0.75, 0.99$ for $S$. Use CLT and normal distribution tables.

**2.** Let $X_1, \dots, X_n$ 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) and find the bias.

(1) $X_i$ are i.i.d. Ber$(p)$ where $p$ is unknown.
(2) $X_i$ are i.i.d. $N(\mu, \sigma^2)$ where $\mu, \sigma^2$ are unknown.

**3.** Let $X_1, \dots, X_n$ 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) and calculate the bias.

(1) $X_i$ are i.i.d. Geo$(p)$ where $p$ is unknown.
(2) $X_i$ are i.i.d. Unif$[a,b]$ where $a, b$ are unknown.

**4.** Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. samples from a bivariate distribution. Let $\tau = \text{Cov}(X_1, Y_1)$. Let $r_n = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X}_n)(Y_k - \overline{X}_n)$ be the sample covariance.

(1) Show that $r_n$ is a biased estimate for $\tau$ and find the bias.
(2) Modify the estimate $r_n$ to get an unbiased estimate of $\tau$.

[**Remark:** It is often convenient, here and elsewhere, to realise that $\tau = \mathbf{E}[X_1Y_1] - \mathbf{E}[X_1]\mathbf{E}[Y_1]$ and $r_n = (\frac{1}{n}\sum_{k=1}^{n} X_kY_k) - \overline{X}_n\overline{Y}_n$.]

**5.** Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables from a distribution $F$. Let $M_n$ be a median of $X_1, \dots, X_n$. Assume that the distribution $F$ has a unique median, that is there is a unique number $m$ such that $F(m) = \frac{1}{2}$. For any $\delta > 0$ show that $\mathbf{P}\{|M_n - m| \geq \delta\} \to 0$ as $n \to \infty$.

[**Remark:** The above statement justifies using the sample median to estimate the population median, in the sense that at least for large sample sizes, the two are close. Similar justification for using sample mean to estimate expected value came from the law of large numbers]

---

*Date*: November 1, 2016.

The following problem is only for those mathematically minded.

**6.** Let $X_1, X_2, \ldots$ be i.i.d. Pois$(\lambda)$ random variables. Work out the exact distribution of $X_1 + \ldots + X_n$ and use it show the central limit theorem in this case. That is, show that for any $a < b$,

$$\mathbf{P}\left\{a \le \frac{\sqrt{n}(\overline{X}_n - \lambda)}{\sqrt{\lambda}} \le b\right\} \longrightarrow \mathbf{P}\{a \le Z \le b\}$$

where $Z \sim N(0,1)$.

[**Remark:** This is analogous to the two cases of CLT that we showed in class, for exponential and for Bernoulli random variables].

The following problem shows that in certain situations, sums of random variables are approximately Poisson distributed. This gives a hint as to why Poisson distribution arises in many contexts. The question may be ignored safely from the exam point of view.

**7.** Let $X_{n,1}, X_{n,2}, \ldots X_{n,n}$ be i.i.d. Ber$(p_n)$ random variables. Let $S_n = X_{n,1} + \ldots + X_{n,n}$. If $np_n \to \lambda$ (a finite positive number), show that $S_n$ has approximately Pois$(\lambda)$ distribution in the sense that for any $k \in \mathbb{N}$,

$$\mathbf{P}\{S_n = k\} \to e^{-\lambda}\frac{\lambda^k}{k!}.$$

[**Remark:** In contrast, if $np_n \to \infty$, deduce from CLT that $S_n$ has approximately a normal distribution, i.e.,

$$\mathbf{P}\left\{a \le \frac{S_n - \mathbf{E}[S_n]}{\sqrt{\text{Var}(S_n)}} \le b\right\} \to \mathbf{P}\{a \le Z \le b\}$$

for any $a < b$.]