Statistics 523, Problem Set 4

Wellner; 4/22/2020

Reading: Chen, Goldstein, & Shao, chapter 2.
(2009) AMM paper by Goldstein
start reading Shorack, PfS (2017), Chapter 11, Section 1, pages 271-277.

Reminder: Project outlines due Monday, 4 May
Midterm Exam, Friday, 8 May

Due: Wednesday, April 29, 2020.

1. Exercise 10.2.1, PfS (2017), page 235: the following are equivalent:
   (i) The random variables are *uan*; that is,
   $\max_{k\le n} P(|X_{nk} - \mu_{nk}| > \epsilon) \to 0$ for all $\epsilon > 0$.
   (ii) $\max_{k\le n} |\phi_{nk}(t) - 1| \to 0$ uniformly on every finite interval of t's.
   (iii) $\max_{k\le n} E(X_{nk}^2 \wedge 1) \to 0$.

2. Suppose that $\{b_i\}_{i=1}^N$ and $\{c_i\}_{i=1}^N$ are two sequences of real numbers, and write $c(i) \equiv c_i$. Suppose that $\underline{R} = (R_1, \ldots, R_N)$ is distributed uniformly over Π_N, the collection of all permutations of $\{1, \ldots, N\}$; i.e. $P(\underline{R} = \underline{r}) = 1/N!$ for all $\underline{r} \in \Pi_N$. Let $S \equiv S_N \equiv \sum_{j=1}^N b_j c(R_j)$. (i) Show that $Var(S) = (N-1)^{-1} B_N^2 \cdot C_N^2$ where $B_N^2 = \sum_{j=1}^N (b_j - \overline{b}_N)^2$ and $C_N^2 = \sum_{j=1}^N (c_j - \overline{c}_N)^2$.
   (ii) What is said in Chen, Goldstein, and Shao (2011) about the asymptotic normality of $(S_N - E(S_N))/\sqrt{Var(S_N)}$?
   (iii) Do they say anything about the rate of convergence to normality of $S_N - E(S_N))/\sqrt{Var(S_N)}$?

3. Suppose that $X_1, \ldots, X_n$ are the numbers resulting from sampling without replacement from an urn consisting of balls with the numbers $a_1, \ldots, a_N$ on the N balls. Let $\overline{a}_N \equiv \overline{a} \equiv N^{-1} \sum_1^N a_i$ and $\sigma_a^2 \equiv N^{-1} \sum_{i=1}^N (a_i - \overline{a})^2$. Let $T_n \equiv X_1 + \cdots + X_n$.
   (i) Verify that for $j \ne k$, $j, k \in \{1, \ldots, N\}$,
   $$Cov[X_j, X_k] = Cov[X_1, X_2] = -\frac{\sigma_a^2}{N-1}$$
   and that
   $$Var(T_n/n) = \frac{\sigma_a^2}{n}\left(1 - \frac{n-1}{N-1}\right).$$
   The factor $(1-(n-1)/(N-1))$ is sometimes called the *finite-sampling correction factor*; note that the variance of the mean is *smaller* than the variance of the mean under sampling with replacement (namely $n^{-1}\sigma_a^2$).
   (ii) Is there any connection with the previous problem, # 2?

4. Suppose that $Y_1, Y_2, \ldots$ are i.i.d. with distribution function G and characteristic function $\varphi(t) = E\exp(itY_1)$. Let N_λ be a random variable with Poisson(λ) distribution and assume that N_λ is independent of the $\{Y_i\}$'s. Let $S \equiv S_\lambda \equiv \sum_{j=1}^{N_\lambda} Y_j$. Find the characteristic function ϕ_S of S.

5. **Bonus problem 1: Beyond zero-bias Stein identities:** There are several other ways of using the Stein identity (or characterization of the normal distribution) to prove central limit theorems and establish rates of convergence. Some of the most important of these are explained in Chen, Goldstein, and Shao (2011), chapter 2:
   (a) For sums of independent random variables: the K function approach:
   (b) Exchangeable pairs.
   (c) Size biased transformation.
   For your choice of either of (a) or (c), explain the basic identity analogous to the identity $E[Wf(W)] = \sigma^2 E[f'(W^*)]$ used in the zero-biased distribution approach.