Homework Assignment 1

Due Jan 31st midnight via Canvas

SDS 384-11 Theoretical Statistics

1. We will examine asymptotic equivalence in this question.

   (a) Show that two sequences of normalized R.V.'s (mean 0 and variance 1) are asymptotically equivalent if their correlation converges to one. Conclude that if $(X_n - E[X_n])/\sqrt{\text{var}(X_n)} \xrightarrow{d} X$ and if $corr(X_n, Y_n) \to 1$, then $(Y_n - EY_n)/\sqrt{\text{var}(Y_n)} \xrightarrow{d} X$.

   (b) Suppose X_n, Y_n have zero mean and equal variance. If $X_n \xrightarrow{d} X$ and $corr(X_n, Y_n) \to 1$, is it true that $Y_n \xrightarrow{d} X$?

2. The following inequality bounds the worst case error that may be made using a Poisson Approximation. It is also known as Le Cam's inequality. Let $X_1, \ldots, X_n$ be i.i.d Bernoulli R.V.'s with $P(X_i = 1) = p_i$. Let $S_n = \sum_i X_i$ and let $\lambda = \sum_i p_i$, and let Z be an R.V. with the Poisson(λ) distribution, i.e. $\mathcal{P}(\lambda)$. Show that for all sets A,

$$|P(S_n \in A) - P(Z \in A)| \leq \sum_i p_i^2.$$

   Hint: We will prove this using a coupling argument, i.e. we will use a construction which defines S_n and Z to be on the same probability space, so that they are close. Let $U \sim Uniform(0,1)$ be i.i.d uniform R.V.'s. Now let $X_i = 1(U_i \geq 1 - p_i)$. Now let $Y_i = 0$ if $U_i < e^{-p_i}$. Construct the rest of Y_i's PMF using U_i such that $Y_i \sim \mathcal{P}(p_i)$. Now show $|P(S_n \in A) - P(Z \in A)| \leq \sum_i P(X_i \neq Y_i)$. Finish the rest of the proof.

3. Suppose $X_1, \ldots, X_n$ are i.i.d random variables with mean μ and variance σ^2. Let $T_n = \sum_i z_{ni} X_i$, $\mu_n = E[T_n]$ and $\sigma_n^2 = \text{var}(T_n)$. Using the Lindeberg-Feller theorem show that

$$\frac{T_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1),$$

   provided $\max_{j \leq n} z_{nj}^2 / \sum_j z_{nj}^2 \to 0$.

4. If $X_n \xrightarrow{d} X \sim Poisson(\lambda)$, is it necessarily true that $E[g(X_n)] \to E[g(X)]$?

   (a) $g(x) = 1(x \in (0, 10))$

   (b) $g(x) = e^{-x^2}$

   (c) $g(x) = sgn(cos(x))$ [$sgn(x) = 1$ if $x > 0$, -1 if $x < 0$ and 0 if $x = 0$.]

   (d) $g(x) = x$

5. Let $X_1, X_2 \dots$ be i.i.d. Uniform on $\{1, \dots, n\}$. In this question, we will deal with the famous Coupon collector's problem. Think of X_i as the i^{th} card you have picked from a set of possibilities (i.e. n cards). Your i^{th} pick is independent of all previous picks. Define the first time to get k different items as

$$\tau_k^n = \inf\{m : |\{X_1, \dots, X_m\}| = k\}$$

Clearly $\tau_1^n = 1$. Assume that $\tau_0^n = 0$. Now set $X_{n,k} := \tau_k^n - \tau_{k-1}^n$ as the time to get a different card from the first $k-1$.

(a) $X_{n,k}$ is distributed as geometric($p_{n,k}$). What is $p_{n,k}$?

(b) Let $T_n = \tau_n^n$. Calculate ET_n and bound var(T_n). You can assume that the variance of a geometric(p) random variable is upper bounded by $1/p^2$.

(c) Show that

$$\frac{T_n}{n \log n} \xrightarrow{P} 1$$

Hint 1: You can use the following method of bounding the different series you come across in your calculations.

$$\sum_{m=1}^{n} \frac{1}{m} \geq \int_1^n \frac{dx}{x} \geq \sum_{m=2}^{n} \frac{1}{m}.$$

Hint 2: You may find it easy to first establish $\frac{T_n - ET_n}{n \log n} \xrightarrow{P} 0$.