Statistics 523, Problem Set 4 Solutions

Wellner; 4/26/2017

1. Exercise 7.5.4, PfS, page 146: Suppose that $X_1, \ldots, X_n$ are i.i.d. with continuous distribution function F. Show that with probability 1 all the observations are distinct. [Hint: use corollary 2 to Theorem 5.1.3.]

   Solution: Now if $i \ne j$, it follows that

   $$P(X_i = X_j) = EE(1_{[X_i = X_j]} | X_j) = E\{\Delta F(X_j)\} = E\{0\} = 0$$

   since $\Delta F(x) \equiv F(x) - F(x-) = 0$ for all x by continuity of F. Then

   $$P(\cup_{1 \le i < j \le n} [X_i = X_j]) \le \sum_{1 \le i < j \le n} P(X_i = X_j) = \sum_{1 \le i < j \le n} 0 = 0.$$

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 $S \equiv S_N \equiv \sum_{j=1}^N b_j c(R_j)$. 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 $C_N^2 = \sum_{j=1}^N (c_j - \overline{c}_N)^2$.

   Solution: As I showed in class,

   $$Var(c(R_i)) = N^{-1} \sum_{j=1}^N (c_j - \overline{c})^2$$

   and

   $$Cov(c(R_i), c(R_j)) = -\frac{1}{N(N-1)} \sum_{j=1}^N (c_j - \overline{c})^2.$$

Thus

$$\begin{aligned}
Var(S_N) &= \sum_{i=1}^N b_i^2 Var(c(R_i)) + \sum_{i\neq j} b_i b_j Cov(c(R_i), c(R_j)) \\
&= \frac{1}{N(N-1)} \sum_{j=1}^N (c_j - \overline{c})^2 \left\{ (N-1)\sum_{i=1}^N b_i^2 - \sum_{i\neq j} b_i b_j \right\} \\
&= \frac{1}{N(N-1)} \sum_{j=1}^N (c_j - \overline{c})^2 N \sum_{i=1}^N (b_i - \overline{b})^2 \\
&= \frac{1}{N-1} \sum_{j=1}^N (c_j - \overline{c})^2 \sum_{i=1}^N (b_i - \overline{b})^2 \\
&= \frac{1}{N-1} B_N^2 C_N^2,
\end{aligned}$$

where we used the identity

$$(N-1)\sum_{i=1}^N b_i^2 - \sum_{i\neq j} b_i b_j = N\sum_{i=1}^N b_i^2 - N^2 \overline{b}_N^2 = N\sum_{i=1}^N (b_i - \overline{b})^2 .$$

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 $T_n \equiv X_1 + \cdots + X_n$ Verify that for $j \neq k$, $j, k \in \{1, \ldots, N\}$,

$$Cov[X_j, X_k] = Cov[X_1, X_2] = -\frac{\sigma_a^2}{N-1}$$

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 $n^{-1}\sigma_a^2$).

Solution: As discussed in class $T_n \overset{d}{=} S$ as in Problem 2 with $c_i = a_i$ for $1 \leq i \leq N$ and $b_i = 1_{\{1,\ldots,n\}}(i)$ for $1 \leq i \leq N$. Now $\overline{c}_N = \overline{a}_N$, $\overline{b}_N = n/N$. Furthermore $C_N^2 = \sum_{i=1}^N (a_i - \overline{a})^2$ while

$$\begin{aligned}
B_N^2 &= \sum_{i=1}^N (b_i - \overline{b})^2 = \sum_{i=1}^N b_i^2 - N\overline{b}^2 \\
&= n - N(n/N)^2 = \frac{n}{N}(N-n).
\end{aligned}$$

Thus it follows from the calculation in Problem 2 that

$$\begin{aligned} Cov(X_i, X_j) &= Cov(a(R_i), a(R_j)) = -\frac{1}{N(N-1)} \sum_{i=1}^{N} (a_i - \overline{a})^2 \\ &= -\frac{\sigma_a^2}{N-1} \end{aligned}$$

for $i \neq j$ since $\sigma_a^2 = N^{-1} \sum_1^N (a_i - \overline{a})^2$, and

$$\begin{aligned} Var(T_n/n) &= \frac{1}{n^2} \cdot \frac{1}{N-1} \frac{n}{N} (N-n) \sum_{i=1}^{N} (a_i - \overline{a})^2 \\ &= \frac{\sigma_a^2}{n} \cdot \frac{N-n}{N-1} = \frac{\sigma_a^2}{n} \cdot \left(1 - \frac{n-1}{N-1}\right). \end{aligned}$$

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 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.

   Solution: By conditioning on N_λ we find that

$$\begin{aligned} \phi_S(t) &= E \exp(itS) = E[E(e^{it \sum_{j=1}^{N_\lambda} Y_j} | N_\lambda)] \\ &= E(\varphi(t)^{N_\lambda}) = \sum_{k=0}^{\infty} \varphi(t)^k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda \varphi(t))^k}{k!} = e^{\lambda(\varphi(t)-1)}. \end{aligned}$$