An urn contains \( n \) balls such that each of the consecutive \( n \) integers 1, 2, 3, \ldots, \( n \) is carried by one ball. If \( k \) balls are removed at random, find the mean and variance of the total of their numbers in two cases:

(a) They are not replaced.
(b) They are replaced.

Find also the distribution of the largest number removed in each case.

Following the hint, let \( S_i \) be the number drawn on the \( i \)th ball, and \( M \) the maximum of the numbers drawn. Thus, the sum of the numbers is given by

\[
T = \sum_{i=1}^{k} S_i.
\]

(a.) To compute the expectations we’ll need \( E(S_i) \) and \( E(S_i S_j) \). Without knowing anything else about the value of the other numbers drawn, \( S_i \) is equally likely to be any of the \( n \) numbers, so that

\[
E(S_i) = \frac{\sum_{j=1}^{n} j}{n} = \frac{n(n+1)}{2} \cdot \frac{1}{n} = \frac{n+1}{2}.
\]

Hence

\[
E(T) = E\left(\sum_{i=1}^{k} S_i\right) = \sum_{i=1}^{k} E(S_i) = \frac{k(n+1)}{2}.
\]

Similarly,

\[
E(S_i^2) = \frac{\sum_{j=1}^{n} j^2}{n} = \frac{n(n+1)(2n+1)}{6} \cdot \frac{1}{n} = \frac{(n+1)(2n+1)}{6}.
\]

If \( i \neq j \) then the balls have different numbers and so each of the \( \binom{n}{2} \) pairs is equally likely. Hence,

\[
E(S_i S_j) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} ij = \frac{1}{n(n-1)} \cdot \frac{1}{2} \left( \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} ij - \sum_{i=1}^{n} i^2}{n} \right)
\]

\[
= \frac{2}{n(n-1)} \cdot \frac{1}{2} \left( \frac{\sum_{i=1}^{n} i^2}{n} \right) - \frac{n(n+1)(2n+1)}{6}
\]

\[
= \frac{2}{n(n-1)} \cdot \frac{(n-1)n(n+1)(3n+2)}{24} = \frac{(n+1)(3n+2)}{12}.
\]

Thus we may compute the variance using the computational formula

\[
\text{Var}(T) = E(T^2) - E(T)^2 = E\left(\left[\sum_{i=1}^{n} S_i\right]^2\right) - \frac{k^2(n+1)^2}{4} = \sum_{i,j=1}^{k} E(S_i S_j) - \frac{k^2(n+1)^2}{4}
\]

\[
= k(n+1)(2n+1) \quad + \frac{(k-1)k(n+1)(3n+2)}{12} \quad - \frac{k^2(n+1)^2}{4} = \frac{k(n+1)(n-k)}{12}.
\]
To get the pmf for \( M = \max\{S_1, \ldots, S_k\} \), we observe that the cumulative distribution function
\[
F_M(m) = P(M \leq m) = P(S_i \leq m \text{ for all } 1 \leq i \leq k).
\]
Since all the subsets of \( k \) are equally likely, the probability is just gotten by counting the number of subsets in \( \{1, 2, 3, \ldots, m\} \). Thus
\[
F_M(m) = \frac{{m \choose k}}{{n \choose k}}.
\]
It follows from Pascal’s triangle that for \( k \leq m \leq n \),
\[
f_M(m) = F_M(m) - F_M(m-1) = \frac{{m \choose k}}{{n \choose k}} - \frac{{m-1 \choose k}}{{n \choose k}}.
\]

(b.) Now assume that the draws are made with replacement so now the draws are independent uniform variables. \( S_i \) is equally likely to be any of the \( n \) numbers, so that
\[
E(S_i) = \sum_{j=1}^{n} \frac{j}{n} = \frac{n(n+1)}{2} \cdot \frac{1}{n} = \frac{n+1}{2}.
\]
Hence, again,
\[
E(T) = E\left( \sum_{i=1}^{k} S_i \right) = \sum_{i=1}^{k} E(S_i) = \frac{k(n+1)}{2}.
\]
Similarly,
\[
E(S_i^2) = \sum_{j=1}^{n} \frac{j^2}{n} = \frac{n(n+1)(2n+1)}{6} \cdot \frac{1}{n} = \frac{(n+1)(2n+1)}{6}.
\]
If \( i \neq j \) then by independence,
\[
E(S_iS_j) = E(S_i)E(S_j) = \frac{(n+1)^2}{4}.
\]
Thus we may compute the variance using the computational formula
\[
\text{Var}(T) = E(T^2) - E(T)^2 = \frac{\left( \sum_{i=1}^{n} S_i \right)^2}{4} - \frac{k^2(n+1)^2}{4} = E\left( \sum_{i=1}^{k} S_i \right)^2 - \frac{k^2(n+1)^2}{4}.
\]
\[
= \left( \sum_{i=1}^{k} S_i^2 + \sum_{i \neq j} S_iS_j \right) - \frac{k^2(n+1)^2}{4} = \sum_{i=1}^{n} E(S_i^2) + \sum_{i \neq j} E(S_iS_j) - \frac{k^2(n+1)^2}{4}
\]
\[
= \frac{k(n+1)(2n+1)}{6} + \frac{k(k-1)(n+1)^2}{4} - \frac{k^2(n+1)^2}{4} = \frac{k(n-1)(n+1)}{12}.
\]
Of course this is just the formula for the variance of a sum for independent variables
\[
\text{Var}(T) = \text{Var}\left( \sum_{i=1}^{k} S_i \right) = \sum_{i=1}^{k} \text{Var}(S_i) = \frac{k(n^2-1)}{12}.
\]
To get the pmf for \( M = \max\{S_1, \ldots, S_k\} \), we observe that the cumulative distribution function
\[
F_M(m) = P(M \leq m) = P(S_i \leq m \text{ for all } 1 \leq i \leq k).
\]
By independence, the probability is just gotten by multiplying the $P(S_i \leq m)$. Thus

$$F_M(m) = \left( \frac{m}{n} \right)^k.$$  

It follows that for $1 \leq m \leq n$,

$$f_M(m) = F_M(m) - F_M(m-1) = \left( \frac{m}{n} \right)^k - \left( \frac{m-1}{n} \right)^k.$$

226[36] An urn contains $m$ white balls and $M - m$ black balls. $n \leq M$ balls are chosen at random without replacement. Let $X$ denote the number of white balls among these. Show that the probability that there are exactly $k$ white balls, $0 \leq k \leq m$ is given by

$$f_X(k) = P(X = k) = \frac{\binom{M-m}{n-k}}{\binom{M}{n}}.$$

Show that $X = I_1 + \cdots + I_n$ where $I_i = 0$ or 1 according to whether or not the $i$th ball is black or white. Show that for $i \neq j$,

$$P(I_i = 1) = \frac{m}{M}, \quad P(I_i = 1 \text{ and } I_j = 1) = \frac{m(m-1)}{M(M-1)}.$$

By computing $E(X)$ and $E(X^2)$ or otherwise, find the mean and variance of $X$. [In other words, given a hypergeometric variable $X \sim \text{hyp}(M, m, n)$, use the method of indicators to derive the mean and variance of $X$.]

$I_i = 1$ exactly when the $i$th ball is white, so that $\sum_{i=1}^n I_i$ is the number of white balls chosen. Knowing nothing else about the other $I_j$, by symmetry $E(I_i) = P(\text{ith ball is white}) = P(\text{first ball is white}) = E(I_1)$. Since each choice of the first balls equally likely,

$$E(I_i^2) = E(I_i) = E(I_1) = \frac{m}{M},$$

since $I_i^2 = I_i$. Similarly, by symmetry, the chances of both the $i$th and $j$th balls being white is the same as the first two drawn being white. Thus, for $i \neq j$,

$$E(I_i I_j) = P(I_i = 1 \text{ and } I_j = 1) = P(I_i = 1 \text{ and } I_2 = 1) = P(I_i = 1) P(I_2 = 1 | I_1 = 1) = \frac{m(m-1)}{M(M-1)}.$$

Now, by linearity,

$$E(X) = E\left( \sum_{i=1}^n I_i \right) = \sum_{i=1}^n E(I_i) = \frac{mn}{M}.$$  

Similarly,

$$E(X^2) = E\left( \left( \sum_{i=1}^n I_i \right)^2 \right) = E\left( \sum_{i,j=1}^n I_i I_j \right) = E\left( \sum_{i=1}^n I_i^2 + \sum_{i \neq j} I_i I_j \right) = \sum_{i=1}^n E(I_i^2) + \sum_{i \neq j} E(I_i I_j) = \frac{mn}{M} + \frac{n(n-1)m(m-1)}{M(M-1)}.$$  

Finally,

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{mn}{M} + \frac{n(n-1)m(m-1)}{M(M-1)} - \frac{m^2n^2}{M^2} = \frac{m(M-m)n(M-n)}{M^2(M-1)}.$$
[A.] \( N \) people arrive separately to a professional dinner. Upon arrival, each person looks to see if he or she has any friends among those present. That person then sits at a table of a friend, or at an unoccupied table if none of those present is a friend. Assuming that each of the \( \binom{n}{2} \) pairs of people are, independently, friends with probability \( p \), find the expected number of occupied tables. [S. Ross, “A First Course in Probability,” Ch. 7 Prob. 8.]

According to the hint, using the method of indicators, we let \( I_i = 1 \) or 0 according to whether or not the \( i \)th arrival sits at a previously unoccupied table. Then the number of occupied tables is \( X = I_1 + \cdots + I_n \). Since no one else has arrived when the first person arrives, always \( I_1 = 1 \) so \( \mathbb{E}(I_1) = 1 \). For \( i \geq 2 \), there are \( i - 1 \) people present, so that the probability that none of them is a friend of \( i \) is, by independence, \( q^{i-1} \). Thus

\[
\mathbb{E}(I_i) = \mathbb{P}(I_i = 1) = q^{i-1}.
\]

This holds for \( i = 1 \) also. Thus the expectation

\[
\mathbb{E}(X) = \mathbb{E} \left( \sum_{i=1}^{n} I_i \right) = \sum_{i=1}^{n} \mathbb{E}(I_i) = \sum_{i=1}^{n} q^{i-1} = \frac{1 - q^n}{1 - q} = \frac{1 - q^n}{p}.
\]

Note that \( 0 < \mathbb{E}(I_i) < 1 \) for \( i \geq 2 \) so \( 1 < \mathbb{E}(X) < n \), as we would expect with \( n \geq 2 \) arrivals.