Math 5010 § 1. **Solutions to tenth Homework**
Treibergs March 27, 2009

226[31] *An urn contaains 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 ith 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 $\mathbf{E}(S_i)$ and $\mathbf{E}(S_i S_j)$. 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

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

Hence

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

Similarly,

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

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

$$\begin{aligned}
\mathbf{E}(S_i S_j) &= \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} ij = \frac{2}{n(n-1)} \cdot \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} ij - \sum_{i=1}^{n} i^2\right) \\
&= \frac{2}{n(n-1)} \cdot \frac{1}{2}\left(\left[\frac{n(n+1)}{2}\right]^2 - \frac{n(n+1)(2n+1)}{6}\right) \\
&= \frac{2}{n(n-1)} \cdot \frac{(n-1)n(n+1)(3n+2)}{24} = \frac{(n+1)(3n+2)}{12}.
\end{aligned}$$

Thus we may compute the variance using the computational formula

$$\begin{aligned}
\mathbf{Var}(T) = \mathbf{E}(T^2) - \mathbf{E}(T)^2 &= \mathbf{E}\left(\left[\sum_{i=1}^{n} S_i\right]^2\right) - \frac{k^2(n+1)^2}{4} = \mathbf{E}\left(\sum_{i,j=1}^{k} S_i S_j\right) - \frac{k^2(n+1)^2}{4} \\
&= \mathbf{E}\left(\sum_{i=1}^{k} S_i^2 + \sum_{i \neq j} S_i S_j\right) - \frac{k^2(n+1)^2}{4} = \sum_{i=1}^{k} \mathbf{E}\left(S_i^2\right) + \sum_{i \neq j} \mathbf{E}\left(S_i S_j\right) - \frac{k^2(n+1)^2}{4} \\
&= \frac{k(n+1)(2n+1)}{6} + \frac{(k-1)k(n+1)(3n+2)}{12} - \frac{k^2(n+1)^2}{4} = \frac{k(n+1)(n-k)}{12}.
\end{aligned}$$

To get the pmf for $M = \max\{S_1, \dots, S_k\}$, we observe that the cumulative distribution function

$$F_M(m) = \mathbf{P}(M \le m) = \mathbf{P}(S_i \le m \text{ for all } 1 \le i \le k).$$

Since all the subsets of k are equally likely, the probability is just gotten by counting the number of subsets in $\{1, 2, 3, \dots, m\}$. Thus

$$F_M(m) = \frac{\binom{m}{k}}{\binom{n}{k}}.$$

It follows from Pascal's triangle that for $k \le m \le n$,

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

(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

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

Hence, again,

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

Similarly,

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

If $i \ne j$ then by independence,

$$\mathbf{E}(S_i S_j) = \mathbf{E}(S_i)\,\mathbf{E}(S_j) = \frac{(n+1)^2}{4}.$$

Thus we may compute the variance using the computational formula

$$\begin{aligned}
\mathbf{Var}(T) = \mathbf{E}(T^2) - \mathbf{E}(T)^2 &= \mathbf{E}\left(\left[\sum_{i=1}^{n} S_i\right]^2\right) - \frac{k^2(n+1)^2}{4} = \mathbf{E}\left(\sum_{i,j=1}^{k} S_i S_j\right) - \frac{k^2(n+1)^2}{4} \\
&= \mathbf{E}\left(\sum_{i=1}^{k} S_i^2 + \sum_{i\ne j} S_i S_j\right) - \frac{k^2(n+1)^2}{4} = \sum_{i=1}^{n} \mathbf{E}\left(S_i^2\right) + \sum_{i \ne j} \mathbf{E}\left(S_i S_j\right) - \frac{k^2(n+1)^2}{4} \\
&= \frac{k(n+1)(2n+1)}{6} + \frac{k(k-1)(n+1)^2}{4} - \frac{k^2(n+1)^2}{4} = \frac{k(n-1)(n+1)}{12}.
\end{aligned}$$

Of course this is just the formula for the variance of a sum for independent variables

$$\mathbf{Var}(T) = \mathbf{Var}\left(\sum_{i=1}^{k} S_i\right) = \sum_{i=1}^{k} \mathbf{Var}(S_i) = \frac{k(n^2-1)}{12}.$$

To get the pmf for $M = \max\{S_1, \dots, S_k\}$, we observe that the cumulative distribution function

$$F_M(m) = \mathbf{P}(M \le m) = \mathbf{P}(S_i \le m \text{ for all } 1 \le i \le k).$$

By independence, the probability is just gotten by multiplying the $\mathbf{P}(S_i \le m)$. Thus

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

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

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

226[36] *An urn contains m white balls and $M - m$ black balls. $n \le M$ 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, $0 \le k \le m$ is given by*

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

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

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

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

$I_i = 1$ exactly when the ith 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 $\mathbf{E}(I_i) = \mathbf{P}(i\text{th ball is white}) = \mathbf{P}(\text{first ball is white}) = \mathbf{E}(I_1)$. Since each choice of the first ballis equally likely,

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

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

$$\begin{aligned}\mathbf{E}(I_i I_j) &= \mathbf{P}(I_i = 1 \text{ and } I_j = 1) = \mathbf{P}(I_1 = 1 \text{ and } I_2 = 1)\\ &= \mathbf{P}(I_i = 1)\,\mathbf{P}(I_2 = 1 \mid I_1 = 1) = \frac{m(m-1)}{M(M-1)}.\end{aligned}$$

Now, by linearity,

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

Similarly,

$$\begin{aligned}\mathbf{E}(X^2) &= \mathbf{E}\left(\left[\sum_{i=1}^n I_i\right]^2\right) = \mathbf{E}\left(\sum_{i,j=1}^n I_i I_j\right) = \mathbf{E}\left(\sum_{i=1}^n I_i^2 + \sum_{i\ne j} I_i I_j\right)\\ &= \sum_{i=1}^n \mathbf{E}(I_i^2) + \sum_{i\ne j}\mathbf{E}(I_i I_j) = \frac{mn}{M} + \frac{n(n-1)m(m-1)}{M(M-1)}.\end{aligned}$$

Finally,

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \frac{mn}{M} + \frac{n(n-1)m(m-1)}{M(M-1)} - \frac{m^2n^2}{M^2} = \frac{m(M-m)n(M-n)}{M^2(M-1)}.$$

[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 ith arrival sits at a previously unoccupied table. Then the number of occupied tables is $X = I_1 + \cdots + I_n$. Since noone else has arrived when the first person arrives, always $I_1 = 1$ so $\mathbf{E}(I_i) = 1$. 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

$$\mathbf{E}(I_i) = \mathbf{P}(I_i = 1) = q^{i-1}.$$

This holds for $i = 1$ also. Thus the expectation

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

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