Question (1988 STEP III Q16)

Balls are chosen at random without replacement from an urn originally containing m red balls and $M - m$ green balls. Find the probability that exactly k red balls will be chosen in n choices $(0 \leqslant k \leqslant m, 0 \leqslant n \leqslant M)$. The random variables X_i $(i = 1, 2, \ldots, n)$ are defined for $n \leqslant M$ by

$$X_i = \begin{cases} 0 & \text{if the } i\text{th ball chosen is green} \\ 1 & \text{if the } i\text{th ball chosen is red.} \end{cases}$$

Show that

(i) $\mathrm{P}(X_i = 1) = \dfrac{m}{M}$.

(ii) $\mathrm{P}(X_i = 1 \text{ and } X_j = 1) = \dfrac{m(m-1)}{M(M-1)}$, for $i \neq j$.

Find the mean and variance of the random variable X defined by

$$X = \sum_{i=1}^{n} X_i.$$

There are $\binom{m}{k}\binom{M-m}{n-k}$ ways to choose k red and and $n-k$ green balls out of a total $\binom{M}{n}$ ways to choose balls. Therefore the probability is:

$$\mathbb{P}(\text{exactly } k \text{ red balls in } n \text{ choices}) = \frac{\binom{m}{k}\binom{M-m}{n-k}}{\binom{M}{n}}$$

(i) Note that there is nothing special about the ith ball chosen. (We could consider all draws look at the ith ball, or consider all draws apply a permutation to make the ith ball the first ball, and both would look like identical sequences). Therefore $\mathbb{P}(X_i = 1) = \mathbb{P}(X_1 = 1) = \frac{m}{M}$.

(ii) Similarly we could apply a permutation to all sequences which takes the ith ball to the first ball and the jth ball to the second ball, therefore:

$$\begin{aligned} \mathbb{P}(X_i = 1, X_j = 1) &= \mathbb{P}(X_1 = 1, X_2 = 1) \\ &= \mathbb{P}(X_1 = 1) \cdot \mathbb{P}(X_2 = 1 | X_1 = 1) \\ &= \frac{m}{M} \cdot \frac{m-1}{M-1} \\ &= \frac{m(m-1)}{M(M-1)} \end{aligned}$$

So:

$$\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{n} X_i)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathbb{E}(X_i) \\
&= \sum_{i=1}^{n} 1 \cdot \mathbb{P}(X_i = 1) \\
&= \sum_{i=1}^{n} \frac{m}{M} \\
&= \frac{mn}{M}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}(X^2) &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_i X_j\right] \\
&= \sum_{i=1}^{n} \mathbb{E}(X_i^2) + 2\sum_{i<j} \mathbb{E}(X_i X_j) \\
&= \frac{nm}{M} + n(n-1)\frac{m(m-1)}{M(M-1)} \\
\text{Var}(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\
&= \frac{nm}{M} + n(n-1)\frac{m(m-1)}{M(M-1)} - \frac{n^2m^2}{M^2} \\
&= \frac{nm}{M}\left(1 - \frac{nm}{M} + (n-1)\frac{m-1}{M-1}\right) \\
&= \frac{nm}{M}\left(\frac{M(M-1) - (M-1)nm + (n-1)(m-1)M}{M(M-1)}\right) \\
&= \frac{nm}{M}\frac{(M-m)(M-n)}{M(M-1)} \\
&= n\frac{m}{M}\frac{M-m}{M}\frac{M-n}{M-1}
\end{aligned}
$$

Note: This is a very nice way of deriving the mean and variance of the hypergeometric distribution

Question (1994 STEP I Q13)

I have a bag containing M tokens, m of which are red. I remove n tokens from the bag at random without replacement. Let

$$X_i = \begin{cases} 1 & \text{if the ith token I remove is red;} \\ 0 & \text{otherwise.} \end{cases}$$

Let X be the total number of red tokens I remove.

(i) Explain briefly why $X = X_1 + X_2 + \cdots + X_n$.

(ii) Find the expectation $\mathrm{E}(X_i)$.

(iii) Show that $\mathrm{E}(X) = mn/M$.

(iv) Find $\mathrm{P}(X = k)$ for $k = 0, 1, 2, \ldots, n$.

(v) Deduce that

$$\sum_{k=1}^{n} k \binom{m}{k} \binom{M-m}{n-k} = m \binom{M-1}{n-1}.$$

(i) The left hand side counts the number of red tokens we have taken. The right hand side counts the number of red tokens we have taken at each point, across all points. Therefore these must be the same.

(ii) $\mathbb{E}[X_i] = \mathbb{P}(i\text{th token is red}) = \frac{m}{M}$ (since there is nothing special about the ith token.

(iii) Therefore $\mathbb{E}[X] = \mathbb{E}[X_1 + \cdots + X_n] = n\mathbb{E}[X_i] = \frac{nm}{M}$

(iv) $\mathbb{P}(X = k) = \binom{m}{k}\binom{M-m}{n-k}/\binom{M}{n}$ since this is the number of ways we can choose k of the m red objects, $n-k$ of the $M-m$ non-red objects divided by the total number of ways we can choose our n tokens.

(v)

$$\begin{aligned} \frac{mn}{M} &= \mathbb{E}[X] \\ &= \sum_{k=1}^{n} k\mathbb{P}(X = k) \\ &= \sum_{k=1}^{n} k \binom{m}{k} \binom{M-m}{n-k} / \binom{M}{n} \end{aligned}$$

$$\Rightarrow \quad \sum_{k=1}^{n} k \binom{m}{k} \binom{M-m}{n-k} = m \frac{n}{M} \binom{M}{n} = m \binom{M-1}{n-1}$$

This question is a nice example of how to find the mean of the hypergeometric distribution

Question (1996 STEP I Q12)

An examiner has to assign a mark between 1 and m inclusive to each of n examination scripts ($n \leqslant m$). He does this randomly, but never assigns the same mark twice. If K is the highest mark that he assigns, explain why

$$\mathrm{P}(K = k) = \binom{k-1}{n-1} \Big/ \binom{m}{n}$$

for $n \leqslant k \leqslant m$, and deduce that

$$\sum_{k=n}^{m} \binom{k-1}{n-1} = \binom{m}{n}.$$

Find the expected value of K.

If the highest mark is k, then there are $n-1$ remaining marks to give, and they have to be chosen from the numbers $1, 2, \ldots, k-1$, ie in $\binom{k-1}{n-1}$ ways. There are n numbers to be chosen from $1, 2, \ldots, m$ in total, therefore $\mathbb{P}(K = k) = \binom{k-1}{n-1} \Big/ \binom{m}{n}$

Since K can take any of the values $n, \cdots, m$, we must have

$$\begin{aligned}
1 &= \sum_{k=n}^{m} \mathbb{P}(K = k) \\
&= \sum_{k=n}^{m} \binom{k-1}{n-1} \Big/ \binom{m}{n} \\
\Rightarrow \qquad \binom{m}{n} &= \sum_{k=n}^{m} \binom{k-1}{n-1}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(K) &= \sum_{k=n}^{m} k \cdot \mathbb{P}(K = k) \\
&= \sum_{k=n}^{m} k \cdot \binom{k-1}{n-1} \Big/ \binom{m}{n} \\
&= n \binom{m}{n}^{-1} \sum_{k=n}^{m} \frac{k}{n} \cdot \binom{k-1}{n-1} \\
&= n \binom{m}{n}^{-1} \sum_{k=n}^{m} \binom{k}{n} \\
&= n \binom{m}{n}^{-1} \sum_{k=n+1}^{m+1} \binom{k-1}{n+1-1} \\
&= n \binom{m}{n}^{-1} \binom{m+1}{n+1}
\end{aligned}$$

$$= n \cdot \frac{m+1}{n+1}$$

Question (2000 STEP II Q13)

A group of biologists attempts to estimate the magnitude, N, of an island population of voles (*Microtus agrestis*). Accordingly, the biologists capture a random sample of 200 voles, mark them and release them. A second random sample of 200 voles is then taken of which 11 are found to be marked. Show that the probability, p_N, of this occurrence is given by

$$p_N = k\frac{\big((N-200)!\big)^2}{N!(N-389)!},$$

where k is independent of N. The biologists then estimate N by calculating the value of N for which p_N is a maximum. Find this estimate. All unmarked voles in the second sample are marked and then the entire sample is released. Subsequently a third random sample of 200 voles is taken. Write down the probability that this sample contains exactly j marked voles, leaving your answer in terms of binomial coefficients. Deduce that

$$\sum_{j=0}^{200}\binom{389}{j}\binom{3247}{200-j} = \binom{3636}{200}.$$

There will be 200 marked vols out of N, and we are finding 11 of them. There are $\binom{200}{11}$ ways to chose the 11 marked voles and $\binom{N-200}{200-11}$ ways to choose the unmarked voles. The total number of ways to choose 200 voles is $\binom{N}{200}$. Therefore the probability is

$$\begin{aligned} p_N &= \frac{\binom{200}{11}\cdot\binom{N-200}{200-11}}{\binom{N}{200}} \\ &= \binom{200}{11}\cdot\frac{\frac{(N-200)!}{(189)!(N-389)!}}{\frac{N!}{(N-200)!(200)!}} \\ &= \binom{200}{11}\frac{200!}{189!}\frac{\big((N-200)!\big)^2}{N!(N-389)!} \end{aligned}$$

As required and $k = \binom{200}{11}\frac{200!}{189!}$.

We want to maximise $\frac{(N-200)!^2}{N!(N-389)!}$, we will do this by comparing consecutive p_N.

$$\begin{aligned} \frac{p_{N+1}}{p_N} &= \frac{\frac{(N+1-200)!^2}{(N+1)!(N+1-389)!}}{\frac{(N-200)!^2}{N!(N-389)!}} \\ &= \frac{(N-199)!^2\cdot N!\cdot(N-389)!}{(N+1)!(N-388)!(N-200)!^2} \\ &= \frac{(N-199)^2\cdot 1\cdot 1}{(N+1)\cdot(N-388)\cdot 1} \end{aligned}$$

$$\begin{aligned} & \frac{p_{N+1}}{p_N} > 1 \\ \Leftrightarrow \quad & \frac{(N-199)^2 \cdot 1 \cdot 1}{(N+1)\cdot(N-388)\cdot 1} > 1 \\ \Leftrightarrow \quad & (N-199)^2 > (N+1)\cdot(N-388) \\ \Leftrightarrow \quad & N^2 - 2\cdot 199N + 199^2 > N^2 - 387N - 388 \\ \Leftrightarrow \quad & -398N + 199^2 > -387N - 388 \\ \Leftrightarrow \quad & 199^2 + 388 > 11N \\ \Leftrightarrow \quad & \frac{199^2+388}{11} > N \\ \Leftrightarrow \quad & 3635\frac{4}{11} > N \end{aligned}$$

Therefore p_N is increasing if $N \leq 3635$, so we should take $N = 3636$.

$$\mathbb{P}(\text{exactly } j \text{ marked voles}) = \frac{\binom{389}{j}\cdot\binom{3636-389}{200-j}}{\binom{3636}{200}}$$

Since

$$\begin{aligned} 1 &= \sum_{j=0}^{200} \mathbb{P}(\text{exactly } j \text{ marked voles}) \\ &= \sum_{j=0}^{200} \frac{\binom{389}{j}\cdot\binom{3247}{200-j}}{\binom{3636}{200}} \\ \Leftrightarrow \quad \binom{3636}{200} &= \sum_{j=0}^{200} \binom{389}{j}\cdot\binom{3247}{200-j} \end{aligned}$$

Question (2001 STEP III Q12)

A bag contains b black balls and w white balls. Balls are drawn at random from the bag and when a white ball is drawn it is put aside.

(i) If the black balls drawn are also put aside, find an expression for the expected number of black balls that have been drawn when the last white ball is removed.

(ii) If instead the black balls drawn are put back into the bag, prove that the expected number of times a black ball has been drawn when the first white ball is removed is b/w. Hence write down, in the form of a sum, an expression for the expected number of times a black ball has been drawn when the last white ball is removed.

Question (2005 STEP III Q13)

A pack of cards consists of $n+1$ cards, which are printed with the integers from 0 to n. A game consists of drawing cards repeatedly at random from the pack until the card printed with 0 is drawn, at which point the game ends. After each draw, the player receives £1 if the card drawn shows any of the integers from 1 to w inclusive but receives nothing if the card drawn shows any of the integers from $w+1$ to n inclusive.

(i) In one version of the game, each card drawn is replaced immediately and randomly in the pack. Explain clearly why the probability that the player wins a total of exactly £3 is equal to the probability of the following event occurring: out of the first four cards drawn which show numbers in the range 0 to w, the numbers on the first three are non-zero and the number on the fourth is zero. Hence show that the probability that the player wins a total of exactly £3 is equal to $\dfrac{w^3}{(w+1)^4}$.

Write down the probability that the player wins a total of exactly £r and hence find the expected total win.

(ii) In another version of the game, each card drawn is removed from the pack. Show that the expected total win in this version is half of the expected total win in the other version.

None

Question (2006 STEP I Q14) **(i)** A bag of sweets contains one red sweet and n blue sweets. I take a sweet from the bag, note its colour, return it to the bag, then shake the bag. I repeat this until the sweet I take is the red one. Find an expression for the probability that I take the red sweet on the rth attempt. What value of n maximises this probability?

(ii) Instead, I take sweets from the bag, without replacing them in the bag, until I take the red sweet. Find an expression for the probability that I take the red sweet on the rth attempt. What value of n maximises this probability?

(i) This is the probability of having the sequence $\underbrace{BB\cdots B}_{r-1 \text{ times}}R$ which has probability $\left(\dfrac{n}{n+1}\right)^{r-1}\dfrac{1}{n+1}$.

Maximising this, is equivalent to maximising log of it, ie

$$\begin{aligned} y &= (r-1)\ln n - r\ln(n+1) \\ \Rightarrow \quad \frac{\mathrm{d}y}{\mathrm{d}n} &= \frac{r-1}{n} - \frac{r}{n+1} \\ &= \frac{(r-1)(n+1)-rn}{n(n+1)} \end{aligned}$$

$$= \frac{r - n - 1}{n(n+1)}$$

Therefore this is maximised when $n = r - 1$

(ii)