Black and White Balls Wager

Start with a set of m white balls labeled 1 to m, n black balls labeled 1 to n, and m empty urns. The $n + m$ balls are distributed into the m urns in the following way.

First, the m white balls are randomly distributed into the m urns subject to the restriction that each urn gets exactly one white ball. (Randomly distributed $\equiv$ all permissible allocations are equally probable.)

Second, the n black balls are put into the m urns, one ball at a time, according to the following rules :

(*i*) a black ball is picked at random from those remaining (that is, each remaining ball is equally likely to be picked at each stage.)

(*ii*) the probability that this black ball will land in a given urn is *directly proportional* to the total number of balls (black and white) already in that urn.

After all the balls have been loaded into the urns, a player can make one of two different bets.

Bet 1 : A player can pay $\$d$ to find out the location of a <u>white ball</u> of his choice. The player will receive $\$b_1$ for each <u>black ball</u> found in the urn where his white ball is located.

Bet 2 : A player can pay $\$d$ to find out the location of a <u>black ball</u> of his choice. The player will receive $\$b_2$ for each <u>black ball</u> found in the urn where his black ball is located, except for the black ball actually chosen by the player, for which he receives nothing.

Which of these two bets has the higher expected payoff when $b_1 = 3$, $b_2 = 2$, $n = 100$ and $m = 10$?

What general strategy should our player take for deciding between bets 1 and 2 ?

Solution

Let X_j = the number of black balls in the j^{th} urn, $j = 1, 2, \ldots, m$ after all n black balls have been distributed into the m urns. By the randomization introduced into the loading of the balls into the urns, $X_1, X_2, \ldots, X_m$ are identically distributed random variables. Thus,

$$\mu = E(X_1) = E(X_2) = \ldots = E(X_m).$$

Let

T = the number of black balls in the urn containing the white ball chosen in Bet 1.

We note that choosing a white ball is equivalent to choosing an urn at random from the m urns. That is, $E(T) = \mu$.

However, $X_1 + X_2 + \ldots + X_m = n$, thus

$$E\Big(X_1 + X_2 + \ldots + X_m\Big) = E(X_1) + E(X_2) + \ldots + E(X_m) = E(n) = n$$

and hence, $m\,E(X_1) = n$, and $\mu = E(X_1) = \frac{n}{m}$. So, the expected payoff of Bet 1 is

$$b_1 \cdot E(T) = b_1 \cdot \mu = b_1 \cdot \frac{n}{m}.$$

Let

Y = the number of black balls in the urn containing the black ball chosen in Bet 2,

and

$$P\Big(Y = j\Big) = \sum_{k=1}^{m} P(\text{ball picked comes from box } k \text{ and box } k \text{ has } j \text{ balls in it})$$

$$= m \cdot P\Big(\text{ball picked comes from box 1} \mid \text{box 1 has } j \text{ balls in it}\Big) \times P(\text{box 1 has } j \text{ balls in it})$$

$$= m \cdot \frac{j}{n} \cdot P(\text{box 1 has } j \text{ balls in it})$$

$$E(Y) = \sum_{j=1}^{n} j \cdot P\Big(Y = j\Big) = \sum_{j=1}^{n} j \cdot m \cdot \frac{j}{n} \cdot P(\text{box 1 has } j \text{ balls in it})$$

$$= \frac{m}{n} \sum_{j=1}^{n} j^2 \cdot P(X_1 = j) = \frac{m}{n} E\Big((X_1)^2\Big)$$

$$= \frac{m}{n}\left(\Big(E(X_1)\Big)^2 + \operatorname{Var}(X_1)\right) = \frac{m}{n}\Big((\frac{n}{m})^2 + \operatorname{Var}(X_1)\Big)$$

$$= \mu + \frac{\operatorname{Var}(X_1)}{\mu}.$$

Let $W_k = (W_{1,k}, W_{2,k}, \ldots, W_{m,k})$ where

$W_{j,k} =$ the total number of balls in the j^{th} urn after the k^{th} black ball is distributed.

Let $A_{j,k}$ be the event that the k^{th} black ball distributed lands in the j^{th} urn. By assumption, the probability that the k^{th} black ball will land in the j^{th} urn is directly proportional to $\omega_{j,k-1}$, the total number of balls (black and white) already in that urn. That is,

$$P\left(A_{j,k} \,\middle|\, W_{k-1} \right) = \alpha(k, m) \cdot \omega_{j,k-1}$$

where $\alpha(k,m)$ is the proportionality factor. By the law of total probability,

$$1 = \sum_{j=1}^{m} P\left(A_{j,k} \,\middle|\, W_{k-1} \right) = \sum_{j=1}^{m} \alpha(k, m) \cdot \omega_{j,k-1}$$

$$= \alpha(k, m) \cdot \sum_{j=1}^{m} \omega_{j,k-1} = \alpha(k, m) \cdot (m + k - 1)$$

Thus,

$$\alpha(k, m) = \frac{1}{(m+k-1)}.$$

It follows that for $k = 0, 1, 2, \ldots, n-1$

$$P\left(W_{1,n} = j \mid W_{1,n-1} = k+1 \right) = \begin{cases} 1 - \left(\frac{k+1}{m+n-1} \right) & j = k+1 \\ \frac{k+1}{m+n-1} & j = k+2 \\ 0 & \text{else} \end{cases}$$

Now, $X_1 = W_{1,n} - 1$, thus

$$\mathrm{Var}(X_1) = \mathrm{Var}(W_{1,n} - 1) = \mathrm{Var}(W_{1,n}).$$

It is a well known result that for general random variables X and Y, that

$$\mathrm{Var}(X) = \mathrm{Var}\Big(E(X \mid Y)\Big) + E\Big(\mathrm{Var}(X \mid Y)\Big).$$

Using this result, we have

$$\mathrm{Var}(X_1) = \mathrm{Var}(W_{1,n}) = \mathrm{Var}\Big(E(W_{1,n} \mid W_{1,n-1})\Big) + E\Big(\mathrm{Var}(W_{1,n} \mid W_{1,n-1})\Big).$$

From the conditional distribution of $W_{1,n}$ given $W_{1,n-1}$, we have

$$E\Big(W_{1,n} \,\Big|\, W_{1,n-1} = \omega\Big) = \omega\left(1 - \left(\frac{\omega}{m+n-1}\right)\right) + (1+\omega)\left(\frac{\omega}{m+n-1}\right)$$

$$= \omega\left(\frac{m+n}{m+n-1}\right)$$

and

$$\mathrm{Var}\Big(E(W_{1,n} \mid W_{1,n-1})\Big) = \mathrm{Var}\left(W_{1,n-1} \cdot \left(\frac{m+n}{m+n-1}\right)\right)$$

$$= \left(\frac{m+n}{m+n-1}\right)^2 \mathrm{Var}(W_{1,n-1}).$$

$$\mathrm{Var}\Big(W_{1,n} \,\Big|\, W_{1,n-1} = \omega\Big) = E\left(\Big(W_{1,n} - E(W_{1,n} \mid W_{1,n-1} = \omega)\Big)^2 \,\Bigg|\, W_{1,n-1} = \omega\right)$$

$$= \Big(\omega - E(W_{1,n} \mid W_{1,n-1} = \omega)\Big)^2 \cdot P\Big(W_{1,n} = \omega \,\Big|\, W_{1,n-1} = \omega\Big)$$

$$+ \Big(\omega + 1 - E(W_{1,n} \mid W_{1,n-1} = \omega)\Big)^2 \cdot P\Big(W_{1,n} = \omega + 1 \,\Big|\, W_{1,n-1} = \omega\Big)$$

$$= \left(\omega - \omega\left(\frac{m+n}{m+n-1}\right)\right)^2 \left(1 - \left(\frac{\omega}{m+n-1}\right)\right)$$

$$+ \left(\omega + 1 - \omega\left(\frac{m+n}{m+n-1}\right)\right)^2 \cdot \left(\frac{\omega}{m+n-1}\right)$$

$$= \frac{\omega \cdot (m+n-\omega-1)}{(m+n-1)^2}.$$

Thus,

$$\mathrm{E}\Big(\mathrm{Var}(\mathrm{W}_{1,n} \mid \mathrm{W}_{1,n-1})\Big) = \mathrm{E}\left(\frac{\mathrm{W}_{1,n-1} \cdot (m+n-\mathrm{W}_{1,n-1}-1)}{(m+n-1)^2}\right)$$

$$= \frac{1}{(m+n-1)^2}\left((m+n-1)\,\mathrm{E}\Big(\mathrm{W}_{1,n-1}\Big) - \mathrm{E}\Big((\mathrm{W}_{1,n-1})^2\Big)\right)$$

$$= \frac{1}{(m+n-1)}\left(1 + \frac{n-1}{m}\right)$$

$$- \frac{1}{(m+n-1)^2} \cdot \left(\mathrm{Var}(\mathrm{W}_{1,n-1}) + \left(1 + \frac{n-1}{m}\right)^2\right)$$

$$= \frac{1}{m} \cdot \left(1 - \frac{1}{m}\right) - \frac{1}{(m+n-1)^2}\,\mathrm{Var}(\mathrm{W}_{1,n-1}).$$

Thus,

$$\mathrm{Var}(\mathrm{W}_{1,n}) = \mathrm{Var}\Big(\mathrm{E}(\mathrm{W}_{1,n} \mid \mathrm{W}_{1,n-1})\Big) + \mathrm{E}\Big(\mathrm{Var}(\mathrm{W}_{1,n} \mid \mathrm{W}_{1,n-1})\Big)$$

$$= \left(\frac{m+n}{m+n-1}\right)^2 \mathrm{Var}(\mathrm{W}_{1,n-1}) + \frac{1}{m} \cdot \left(1 - \frac{1}{m}\right) - \frac{1}{(m+n-1)^2}\,\mathrm{Var}(\mathrm{W}_{1,n-1})$$

$$= \left(\frac{m+n+1}{m+n-1}\right) \mathrm{Var}(\mathrm{W}_{1,n-1}) + \frac{1}{m} \cdot \left(1 - \frac{1}{m}\right)$$

where

$$\mathrm{Var}(\mathrm{W}_{1,1}) = \mathrm{E}\Big((\mathrm{W}_{1,1})^2\Big) - \Big(\mathrm{E}(\mathrm{W}_{1,1})\Big)^2 = \mathrm{P}(\mathrm{W}_{1,1}) - \Big(\mathrm{P}(\mathrm{W}_{1,1})\Big)^2$$

$$= \frac{1}{m} - \left(\frac{1}{m}\right)^2.$$

Thus, we have a linear recurrence relation (with nonconstant coefficients) from which we can solve for Var($W_{1,n}$). By successive substitutions, we have, for $n \geq 2$,

$$\mathrm{Var}(W_{1,n}) \;=\; u(n)\,\mathrm{Var}(W_{1,1}) \;+\; u(n)\,\frac{(m-1)}{m^2}\sum_{j=2}^{n}\frac{1}{u(j)}$$

where,

$$u(k) \;=\; \prod_{j=2}^{k}\left(\frac{m+j+1}{m+j-1}\right) \;=\; \frac{(m+k+1)(m+k)}{(m+2)(m+1)}.$$

Therefore,

$$\mathrm{Var}(W_{1,n}) \;=\; \frac{(m+n+1)(m+n)}{(m+2)(m+1)}\cdot\frac{(m-1)}{m^2}$$

$$+\;\frac{(m-1)(m+n+1)(m+n)}{m^2}\sum_{j=2}^{n}\frac{1}{(m+j+1)(m+j)}$$

$$=\;\frac{(m+n+1)(m+n)(m-1)}{m^2(m+2)(m+1)}\left(1\;+\;\sum_{j=2}^{n}\frac{(m+2)(m+1)}{(m+j+1)(m+j)}\right)$$

$$=\;\frac{(m+n+1)(m+n)(m-1)}{m^2}\left(\sum_{j=1}^{n}\frac{1}{(m+j+1)(m+j)}\right).$$

However, using partial fractions, we have

$$\sum_{j=2}^{n}\frac{1}{(m+j+1)(m+j)} \;=\; \sum_{j=2}^{n}\frac{1}{(m+j)} \;-\; \sum_{j=2}^{n}\frac{1}{(m+j+1)}$$

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

and hence,

$$\mathrm{Var}(W_{1,n}) \;=\; \frac{(m+n+1)(m+n)(m-1)}{m^2}\left(\sum_{j=1}^{n}\frac{1}{(m+j+1)(m+j)}\right)$$

$$= \frac{(m+n+1)(m+n)(m-1)}{m^2} \left(\frac{n}{(m+1)(m+n+1)} \right)$$

$$= \frac{n(m+n)(m-1)}{m^2(m+1)}.$$

So, finally we have $\text{Var}(X_1) = \text{Var}(W_{1,n}) = \frac{n(m+n)(m-1)}{m^2(m+1)}$, $n \geq 2$. We have already established that

$$\text{Var}(W_{1,1}) = \frac{m-1}{m^2}$$

which agrees with the above general formula, hence we can conclude that

$$\text{Var}(X_1) = \text{Var}(W_{1,n}) = \frac{n(m+n)(m-1)}{m^2(m+1)}, \quad n \geq 1.$$

From here, we have

$$\begin{aligned} E(Y) &= E(\text{number of black balls in the urn containing the black ball chosen in Bet 2}) \\ &= \mu + \frac{\text{Var}(X_1)}{\mu} \\ &= \frac{n}{m} + (\frac{m}{n})\left(\frac{n(m+n)(m-1)}{m^2(m+1)} \right) \\ &= \frac{n}{m} + \left(\frac{(m+n)(m-1)}{m(m+1)} \right) \\ &= 2\mu \left(\frac{m}{m+1} \right) + \left(\frac{m-1}{m+1} \right) \end{aligned}$$

where $\mu = E(\text{number of black balls in the urn containing the white ball chosen in Bet 1})$.

The expected payoff on Bet 1 is $b_1\mu$ and the expected payoff on Bet 2 is

$$b_2 \cdot \left(2\mu \left(\frac{m}{m+1} \right) + \left(\frac{m-1}{m+1} \right) - 1 \right)$$

where the final $1 is subtracted off because by the rules the player receives nothing for the black ball actually chosen. It follows that the expected payoff on Bet 2 equals the expected payoff on Bet 1 whenever

$$b_2 \left(2\mu \left(\frac{m}{m+1} \right) + \left(\frac{m-1}{m+1} \right) - 1 \right) = b_1 \mu$$

This equality simplifies to

$$\frac{b_1}{b_2} = 2 \cdot \left(\frac{m}{m+1} \right) \left(\frac{n-1}{n} \right).$$

Therefore, in order to maximize expected payoff, our player should follow the strategy of

Make Bet 1 whenever

$$\frac{b_1}{b_2} > 2 \cdot \left(\frac{m}{m+1} \right) \left(\frac{n-1}{n} \right)$$

Make Bet 2 whenever

$$\frac{b_1}{b_2} < 2 \cdot \left(\frac{m}{m+1} \right) \left(\frac{n-1}{n} \right)$$

In the particular case when $b_1 = 3$, $b_2 = 2$, $n = 100$ and $m = 10$,

$$\frac{3}{2} < 2 \cdot \frac{10}{11} \cdot \frac{99}{100}$$

and hence our player should take Bet 2 in this case. □