October 11, 2010

Splitting Between Two Jars

For some positive integer N there are N red balls and N blue balls, and two jars. All the balls are distributed in some way between the two jars, and then a ball is randomly selected by first choosing a jar at random, and then choosing a ball at random from the jar. How should one distribute the balls to get the highest probability of drawing a red ball? (Hint: the optimal probability is more than 1/2).

Solution Two solutions are presented. Both write a function giving the probability of choosing a red ball and seek to maximize it. One does this using calculus, the other using combinatorial reasoning. **Solution 1 (Calculus):** (Courtesy of Tom Edgar) Let z be the number of total balls in jar 1 and let x be the total number of red balls in jar 1.

Then, since one has an equal chance of picking either jar, the probability of picking jar 1 is 0.5 and likewise for jar 2.

So the probability of getting a red ball is given by the multivariate function

$$f(x,z) = 0.5\frac{x}{z} + 0.5\frac{N-x}{2N-z}$$

We want to maximize f, so take the partial derivatives, set them equal to zero and solve

$$f_x = 0.5\frac{1}{z} - 0.5\frac{1}{2N-z}$$

and

$$f_z = -0.5\frac{x}{z^2} + 0.5\frac{N-x}{(1-z)^2}$$

and we want to maximize on $1 \leq z \leq 100$ and $1 \leq x \leq 50$ (ruling out $(0,0)$ for now) and $x \leq z$.

It turns out there is a critical point, but it ends up being a saddle point.

So, that means that the maximum is on the boundary. The boundary looks like a square with a trapezoid cut out of the bottom (corresponding to the case where $z < 50$ and $x > z$).

Turning to the boundary cases, the interesting case is the $x = z$ line. This line contains no critical points and the derivative is negative so the function is decreasing along it. This means f is minimized at the bottom of the line: the case $x = z = 1$.

So, this gives the probability of picking a red ball as: $0.5(1/1)+0.5(49/99) = 0.7474\ldots = 74/99$.

Solution 2 (Algebra): Pick a jar and let (r, b) represent the number of red and blue balls respectively in that jar. Let N_a be the total number of red balls and N_b the total number of blue balls. We fix $N = N_a = N_b$. The function

$$f(r,b) = \frac{1}{2}\left(\frac{r}{r+b} + \frac{N-r}{2N-r-b}\right)$$

gives the probability of drawing a red ball in terms of (r, b). Our goal is to maximize this function.

The jar we picked contains (r, b) so the other jar contains $(N-r, N-b)$. If we picked the other jar in the beginning, we would get the same overall probability, giving the identity $f(r, b) = f(N-r, N-b)$. Thus, if $r \neq b$ we may assume, by possibly switching jars, that $r > b$.

Claim 1. *Let a and b be integers.*
(1) If $a > b > 0$ then $\frac{a}{a+b} < \frac{a-1}{a+b-2}$.
(2) If $b > a > 0$ then $\frac{a}{a+b} > \frac{a-1}{a+b-2}$

Proof. (1) Observe $0 < a+b-2$. Suppose for contradiction $\frac{a}{a+b} \geq \frac{a-1}{a+b-2}$. Cross multiplying, expanding both sides, and canceling like terms gives $-2a \geq -a-b$, which is equivalent to $b \geq a$. This is a contradiction, proving the claim.

(2) Similar. ☐

Claim 2. $f(1, 0) > 1/2$

Proof. Since $\frac{N-1}{2N-1} > 0$, $f(1, 0) = \frac{1}{2}(1 + \frac{N-1}{2N-1}) > 1/2$. ☐

The next claim is the crucial idea: moving a red and a blue ball to the other jar always improves our odds of picking a red ball.

Claim 3. *If $r > b > 0$ then $f(r-1, b-1) > f(r, b)$.*

Proof. We will show $f(r-1, b-1) - f(r, b) > 0$, for which it is enough to show $2f(r-1, b-1) - 2f(r, b) > 0$. Substituting in the definition of f gives

$$\frac{r-1}{r+b-2} + \frac{N-r+1}{2N-r-b+2} - \frac{r}{r+b} - \frac{N-r}{2N-r-b}$$

Regroup as

$$\left(\frac{r-1}{r+b-2} - \frac{r}{r+b}\right) + \left(\frac{N-r+1}{2N-r-b+2} - \frac{N-r}{2N-r-b}\right)$$

Since $r > b > 0$, the left group is > 0, using claim 1.1. In the same way the right group is > 0 since $N-b+1 > N-r+1 > 0$. Thus the above expression is greater than 0, which is what we needed to show. ☐

Claim 4. *The case $(1, 0)$ is optimal.*

Proof. It suffices to show for every pair $(r, b) \neq (1, 0)$ with $r \geq b$ that there is some other (r', b'), $r' \geq b'$ with $f(r', b') > f(r, b)$.

First consider the case $(r, 0)$ with $r > 1$. Then by moving a single red ball to the other jar we see that the probability of drawing a red ball in the first jar stays the same $(= 1)$ and the probability of drawing a red ball in the other jar improves. Thus $f(r-1, 0) > f(r, 0)$ for $r > 1$.

Next consider the case (r, b) with $r = b$. Then $f(r, b) = 1/2$, and so $f(1, 0) > f(r, b)$ by claim 2.

Finally, suppose we have (r, b) with $r > b > 0$. Then by moving one red ball and one blue ball to the other jar we reach the state $(r-1, b-1)$. By claim 3 $f(r-1, b-1) > f(r, b)$, and since $r > b$, we have $r - 1 > b - 1$.

Induction then shows $(1, 0)$ is optimal. □

This same argument should work in the case $N_a \neq N_b$, but I haven't tried it yet.