The Prisoner's Dilemma

ROMAN WONG
Washington & Jefferson College
Washington, PA 15301

The following is an interesting problem in probability: *The prisoner's dilemma*: A prisoner was to be executed and he begged the king for mercy. The king decided to give the prisoner a chance. He gave the prisoner 50 white balls and 50 black balls, all identical in shape. The prisoner was supposed to distribute these balls into two identical bags in any way he liked and then pick one bag at random and draw one ball at random from that bag. His life would be spared if the ball drawn was a white ball. (There are several other problems under the same name. For instance, see [1].)

The question now of course is to decide how the balls should be distributed so that the prisoner has the best (and the worst) chance to live. For a small number of balls, one can easily find the answer by computing the probabilities of all possible distributions. A solution to the above problem can be found in [2]. For a large number of balls, a calculus approach is more desirable.

Assume now that there are N white balls and N black balls, $N \geq 1$, to be distributed into two identical bags. Let w and b be the number of white and black balls in one of the bags, $0 \leq w,\ b \leq N$. Consider the domains

$$\mathfrak{D} = \{(w,b) \in \mathbb{Z} \times \mathbb{Z} \mid 0 \leq w,\ b \leq N,\ w+b \neq 0,\ w+b \neq 2N\}$$

and

$$\mathfrak{D}' = \{(w,b) \in \mathbb{R} \times \mathbb{R} \mid 0 \leq w,\ b \leq N,\ w+b \neq 0,\ w+b \neq 2N\}.$$

The probability that a ball drawn at random from one of the bags is a white ball is given by

$$P(w,b) = \frac{1}{2}\left(\frac{w}{w+b} + \frac{N-w}{2N-w-b}\right), \qquad (w,b) \in \mathfrak{D}. \tag{1}$$

We want to find the maximum and minimum of the function $P(w,b)$ in $\mathfrak{D}'$. The partial derivatives of P are rather messy. Nevertheless, it can be shown that

$$\frac{\partial P}{\partial w} = \frac{1}{2}\left[\frac{(w+b)(2N-w-b)(N-2w) - (3Nw - 2w^2 - 2wb + Nb)(2N-2w-2b)}{(w+b)^2(2N-w-b)^2}\right]$$

and

$$\frac{\partial P}{\partial b} = \frac{1}{2}\left[\frac{(w+b)(2N-w-b)(3N-4w-2b) - (3Nw - 2w^2 - 2wb + Nb)(2N-2w-2b)}{(w+b)^2(2N-w-b)^2}\right].$$

To find critical points of P, we set both partial derivatives to zero. This leads to

$$(w+b)(2N-w-b)(N-2w) = (w+b)(2N-w-b)(3N-4w-2b).$$

Thus,

$$(w+b)(2N-w-b)(2w+2b-2N)=0.$$

Since $w+b \neq 0$ and $w+b \neq 2N$, we must have $w+b=N$. Putting this back into $\partial P/\partial w = 0$, we find that $(w,b)=(N/2, N/2)$ is the only critical point of P in $\mathfrak{D}'$. From (1), we have $P(N/2, N/2) = 1/2$. In fact, $P(w,b)|_{w+b=N} = 1/2$. We shall see that this value is neither the maximum nor the minimum for P in $\mathfrak{D}'$.

We now consider the value of $P(w,b)$ on the boundary of $\mathfrak{D}'$. First we set $w=0$. Then

$$P(0,b) = \frac{N}{2}(2N-b)^{-1} \quad \text{so that} \quad \frac{\partial}{\partial b}P(0,b) = \left(\frac{N}{2}\right)(2N-b)^{-2},$$

which is always positive for any $(0,b) \in \mathfrak{D}$. Thus $P(w,b)$ is increasing along $w=0$ and so the maximum of $P(0,b)$ along the border $w=0$ is attained at $b=N$ with maximum value

$$P(0,N) = \frac{N}{2(2N-N)} = \frac{1}{2}.$$

By symmetry, $P(N,0) = 1/2$ is the maximum of $P(N,b)$ along the border $w=N$.

We now consider the border $b=0$.

$$P(w,0) = \frac{1}{2}\left(1+\frac{N-w}{2N-w}\right) = \frac{1}{2}\left(\frac{3N-2w}{2N-w}\right)$$

and so

$$\frac{\partial P(w,0)}{\partial w} = \frac{1}{2}\left[\frac{(2N-w)(-2)+(3N-2w)}{(2N-w)^2}\right] = \frac{-N}{2(2N-w)^2} < 0 \text{ in } \mathfrak{D}'.$$

Thus $P(w,0)$ is decreasing along the border $b=0$ and so the maximum of $P(w,0)$ in $\mathfrak{D}$ is attained at $w=1$ with maximum value

$$P(1,0) = \frac{1}{2}\left(\frac{3N-2}{2N-1}\right). \tag{2}$$

By symmetry, the maximum of P along the border $b=N$ is attained at $w=N-1$ with the same maximum value given by (2). Since $N \geq 1$, $3N-2 \geq 2N-1 > 0$ so that

$$P(N-1,N) = P(1,0) = \frac{1}{2}\left(\frac{3N-2}{2N-1}\right) \geq \frac{1}{2}(1) = \frac{1}{2},$$

with equality only when $N=1$. Thus we have shown that the maximum of $P(w,b)$ is attained at $(1,0)$ and $(N-1,N)$ with maximum given by (2). Also, as a function of N,

$$\frac{dP(1,0)}{dN} = \frac{1}{2(2N-1)^2} > 0,$$

so that the prisoner's chance of survival increases when more balls are used. The best chance for his survival is thus equal to

$$\lim_{N\mapsto\infty}\left(\frac{3N-2}{2(2N-1)}\right) = \frac{3}{4}.$$

The above discussions also indicate that the minimum value of $P(w, b)$ is attained on the border at $(0, 1)$ and $(N, N-1)$ with minimum value equal to

$$\frac{N}{2(2N-1)} \leq \frac{N}{2(2N-N)} = \frac{1}{2}.$$

Since $N/2(2N-1)$ is a decreasing function in N, the worst chance for the prisoner's survival is $1/4$.

The above argument can be extended to any number of bags used. Let k be any integer with $N \geq k \geq 2$. We want to find the maximum and minimum probabilities of drawing a white ball when N white balls and N black balls are distributed among the k bags. Let w_i and b_i be the number of white and black balls respectively to be distributed in bag i, $1 \leq i \leq k$. The probability that a white ball is drawn at random from the k bags is a function of $2k$ variables:

$$P = P(w_1, w_2, \ldots, w_k; b_1, b_2, \ldots, b_k) = \frac{1}{k} \sum_{i=1}^{k} \frac{w_i}{w_i + b_i},$$

where $0 \leq w_i$, $b_i \leq N$, $w_i + b_i \neq 0$ for each i, and

$$\sum_{i=1}^{k} w_i = N = \sum_{i=1}^{k} b_i.$$

With the result for $k = 2$ above and a simple induction on k, one can see that the maximum value of P is attained at points $(w_1, \ldots, w_j, \ldots w_k; b_1, \ldots, b_j, \ldots b_k)$ for each $1 \leq j \leq k$, where

$$w_j = N - k + 1, \; b_j = N \quad \text{and} \quad w_i = 1 \quad \text{and} \quad b_i = 0 \text{ for each } i \neq j.$$

It follows that the maximum probability is equal to

$$\begin{aligned} P_{\max} &= P(1, 1, \ldots, 1, N-k+1; 0, 0, \ldots, 0, N) = \frac{1}{k}\left[k - 1 + \frac{N-k+1}{2N-k+1}\right] \\ &= \frac{2Nk - k^2 + k - N}{2Nk - k^2 + k}. \end{aligned}$$

Thus for a fixed k,

$$\frac{dP_{\max}}{dN} = \frac{k^2 - k}{(2Nk - k^2 + k)^2} > 0$$

so that $P_{\max}$ is an increasing function of N. Taking the limit of $P_{\max}$ as N approaches ∞, we see that the limit probability with a large number of balls is $(2k-1)/2k$. As more bags are used, this probability approaches 1 as a limit.

Similarly, the minimum P value is attained at $(0, \ldots, 0, N; 1, \ldots, 1, N-k+1)$ with minimum probability equal to

$$P_{\min} = P(0, \ldots, 0, N; 1, \ldots, 1, N-k+1) = \frac{N}{k(2N-k+1)} \leq \frac{N}{k(2N-N)} = \frac{1}{2k}.$$

For a fixed k, this minimum probability approaches $1/2k$ when N is large, and approaches 0 when many bags are used. Thus, for example if $k = 5$ and $N = 10$, we see that we should fill each of the four bags with exactly one white ball and no black balls. The remaining bag should contain six white and 10 black balls. With $k = 5$, increasing N implies that $P_{\max} \to 0.90$.

Remark. The above discussion shows a good application of calculus in probability. When $k = 2$, the above result can also be established without using calculus. Denote $w + b$ by t, the number of balls in the first bag. Now (1) can be written as

$$P(w,b) = \frac{1}{2}\left(\frac{w}{t} + \frac{N-w}{2N-t}\right) = \frac{1}{2}\left(\frac{N}{2N-t}\right) + \frac{1}{2}w\left(\frac{1}{t} - \frac{1}{2N-t}\right).$$

The expression in the last set of parentheses is positive when $0 < t < N$ and is 0 if $t = N$. The maximum of $P(w, b)$ thus occurs when $w = t$ with value $P_t = P(t, b) = (1/2)(2 - N/(2N - t))$. The largest value for P_t is attained at $t = 1$ with value

$$\frac{1}{2}\left(2 - \frac{N}{2N-1}\right) = \frac{1}{2}\left(\frac{3N-2}{2N-1}\right).$$

The author is grateful to the referee for pointing out this method.

REFERENCES

1. Solomon Garfunkel, *For All Practical Purposes*, W. H. Freeman and Company, New York, 1988, p. 220–224.
2. Fred Schuh, *The Master Book of Mathematical Recreations*, Dover Publications, Mineola, NY, 1968, p. 180.

An Unconventional Orthonormal Basis Provides an Unexpected Counterexample

GIUSEPPE MARINO
PAOLAMARIA PIETRAMALA
Università della Calabria
Cosenza, Italy

Any good textbook about linear operators on real Hilbert spaces begins with the basic geometry of Hilbert spaces. In this context one introduces orthonormal bases and gives the classical examples of such bases in the spaces $L^2[a,b]$, with $[a,b]$ a compact interval of R: In this way the reader becomes familiar with the Haar basis and, via the Gram-Schmidt orthogonalization procedure, the Legendre, Hermite, and Laguerre polynomials.

We note that the usual examples of orthonormal bases in $L^2[a,b]$ are such that either *all* the functions in the basis are *continuous* functions or *all* the functions in the basis are *discontinuous* functions. So, the following problem arises:

QUESTION 1. *Does there exist an orthonormal basis in $L^2[a,b]$ containing only one discontinuous function?*

Here discontinuous means that such a function is not equal almost everywhere to another continuous function. In Proposition 1 we show that the answer to this question is "yes".