

# An Optimization Problem With a Surprisingly Simple Solution

D. Drinen, K. G. Kennedy, and W. M. Priestley

## 1 INTRODUCTION.

Suppose you and  $n$  of your friends play the following game. A random number from the uniform distribution on  $[0, 1]$  will be generated. This number is called the *target*. Each of you will independently guess what the target number will be and the person whose guess is closest will be declared the winner. In order to investigate an optimal strategy for this game, we need to assume something about your friends' guesses.

Consider first the case where you have complete knowledge of all your friends' guesses before you make yours. In that case the optimal strategy is trivial: simply order their guesses and then find the largest gap between successive guesses. If that gap is at least twice as large as that between 0 and the smallest guess and also that between the largest guess and 1, then position yourself halfway between those two guesses. If not, then position yourself as close as the rules allow to the left of the smallest guess or the right of the largest guess, as appropriate. Your friends would likely not find this game to be worth playing and even you would probably not find it interesting.

Let us now assume that you do not have exact knowledge of your friends' guesses. Most of the papers on similar games (see, for example, [1], [7], [2], and [3]) make the assumption that your friends will behave game-theoretically. That is, they will make guesses based on what they think your guess will be, knowing that your guess also depends on what they are likely to guess. With this approach it seems necessary also to assume the number of players is small, for if  $n$  is large such strategical thinking is probably not as feasible and the computations may become much more difficult.

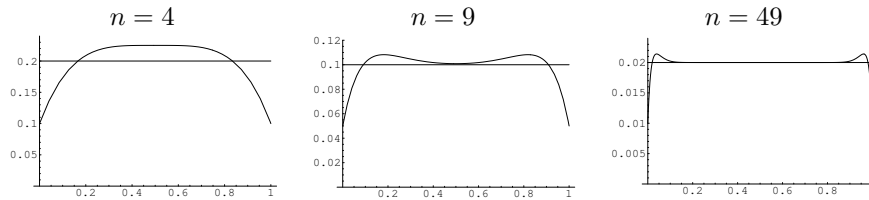
Here we are primarily interested in large  $n$ , so we take a different viewpoint: we assume that you have *probabilistic* knowledge about your friends' guesses. That is, you do not know what the guesses will be, but you know that they will all come from a particular probability distribution. We will consider several distributions in the sections ahead, but for now let us focus on the mathematics of a relatively simple case. If your friends are all computers programmed to guess a uniform random number between 0 and 1, then the optimal strategy can be found.

Before we do it, though, take a moment to guess what the answer is. If you



In Section 4 we will establish in formula (6) an explicit expression for  $W_n$ . For now, we can get a feel for the behavior of  $W_n$  by looking at its graph for various values of  $n$ . Figure 1 shows graphs of  $W_n(g)$  for  $n = 4, 9,$  and  $49$ . Also shown on each plot is a horizontal line at  $1/(n + 1)$ .

Figure 1: Win probability as a function of your guess, with target and guesses distributed uniformly on  $[0, 1]$ .



Two symmetric small peaks materialize in the graph of  $W_n$  when  $n = 5$  and evolve thereafter in an intriguing way. One can see that the peak on the graph of  $W_n$  (for the sake of definiteness, we will always focus our attention on the left peak) moves toward zero as  $n$  gets larger. Also, the range of choices for  $g$  that yield win probabilities nontrivially greater than  $1/(n + 1)$  becomes narrower. Hence, if you know  $n$  exactly you can make an optimal choice for  $g$ , but if you only have the vague knowledge that  $n$  is large, then you might not wish to try for the optimal value. If you think  $n$  is 500 for example, then the guess you make will be disappointing if  $n$  turns out to be only 200.

So if a few thousand of your friends are making uniform random guesses at a uniform random number, then it does not much matter what your guess is. Perhaps this is not surprising. It may be surprising, however, that it does not matter that the distribution is uniform. If the target number is exponentially distributed, for example, and your friends' guesses are also exponentially distributed (with the same parameter), is an analogue of Theorem 1 still true? It is. A main result of this paper is the following generalization of Theorem 1.

**Theorem 2.** *Let  $X_1, X_2, \dots, X_n,$  and  $T$  be independent random variables from the same continuous distribution having density function  $f$ . Let  $g$  be any number for which both  $\lim_{t \rightarrow g^-} f(t)$  and  $\lim_{t \rightarrow g^+} f(t)$  exist and are nonzero. Define*

$$W_n(g) = P(|g - T| < \min_i |X_i - T|).$$

Then

$$\frac{W_n(g)}{1/(n + 1)} \rightarrow 1$$

as  $n$  approaches infinity.

By a *continuous* distribution we mean one with a continuous cumulative distribution function (cdf), though its associated density function (pdf) need

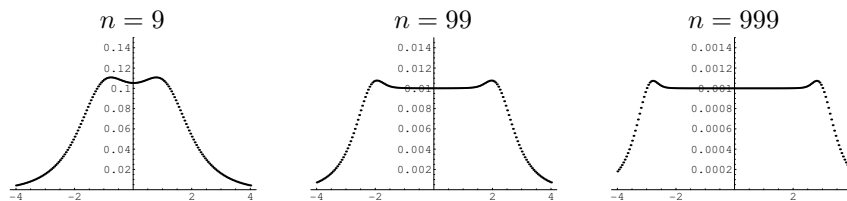
not be continuous. We restrict our attention to continuous random variables (those associated with continuous distributions), where the probability of a tie in the game described at the beginning of the paper will be zero.

To get a feel for the content of this theorem, consider the case where the target random number  $T$  is a standard normal random variable (mean 0, standard deviation 1), and your friends' guesses are also distributed as independent standard normal random variables. The target number is likely to be near zero, so you might be tempted to guess zero or something close to it. But the guesses are likely to be clustered around zero as well, so a guess near zero will win only if the target number ends up in a very narrow range around your guess. At the other extreme, if you make a guess that is far away from zero, you have carved out a wider range of winning possibilities for the target number. But because you are far from zero, the probability of the target's landing in that wide range may (or may not) be less than the probability of landing in some narrow range near zero.

Roughly speaking, your choice is between claiming a tall skinny piece of the distribution (a guess near zero), a short fat piece (a guess far from zero), or something in between. The restriction on  $g$  in the statement of the theorem simply means that your guess is reasonable in some sense. So Theorem 2 says it doesn't matter what you choose, as long as your guess is reasonable. Guess whatever you want. As long as  $n$  is large enough, the crowding out of guesses near the tall part of the distribution will almost exactly counterbalance the unlikelihood of guesses in the short part, making your guess no better or worse than a random guess from the same distribution as the target number.

In Figure 2 we see Mathematica-generated graphs of the function  $W_n(g)$  for  $n = 9, 99,$  and  $999$  in the case where the  $X_i$ 's and  $T$  come from a standard normal distribution. Note the asymptotic convergence to a horizontal line at  $1/(n+1)$ , and also note the small peaks moving outward.

Figure 2: Win probability as a function of your guess, with target and guesses distributed as standard normals.



In the next section we will state and prove a general formula for  $W_n(g)$ , and the following section contains the proof of Theorem 2. In Section 4 we specialize once again to the setting considered in Theorem 1 and explore a few interesting aspects of the uniform case. Section 5 will compare graphs of  $W_n(g)$

for several different probability distributions, whose behavior is clarified by a generalization of Theorem 2. We will close in Section 6 with remarks about related questions.

## 2 A GENERAL FORMULA FOR $W_n(g)$ .

**Theorem 3.** *Suppose  $X_1, X_2, \dots, X_n$  are independent identically distributed continuous random variables with pdf  $f_X$  and cdf  $F_X$ . Suppose that  $T$  is a continuous random variable with pdf  $f_T$ , and that  $X_1, X_2, \dots, X_n, T$  are independent. For any real number  $g$ , let  $W_n(g) = P(|g - T| < \min_i |X_i - T|)$ . Then*

$$W_n(g) = \int_{-\infty}^g (1 - F_X(g) + F_X(2t - g))^n f_T(t) dt + \int_g^{\infty} (1 - F_X(2t - g) + F_X(g))^n f_T(t) dt. \quad (1)$$

Before beginning the proof, we should note that  $W_n(g)$  can also be written as

$$\int_{-\infty}^{\infty} h_g(t)^n f_T(t) dt, \quad (2)$$

where  $h_g(t)$  is the probability that  $X_1$  does not lie between  $g$  and  $2t - g$ . Formula (1) is handy in generating Mathematica plots as in Figure 2, while the conceptually simpler formula (2) will be seen to suggest a proof of Theorem 2.

To prove Theorem 3, we start by noting that

$$W_n(g) = P(\text{you win}) = P(T \leq g, \text{you win}) + P(T > g, \text{you win}).$$

We will show that these two terms correspond to the two integrals in formula (1).

If you guess  $g$ , you win if and only if each of  $X_1, \dots, X_n$  lies outside the interval of length  $2|T - g|$  centered at  $T$ , as seen in Figure 3.

We will consider in detail the case when  $T \leq g$ .

$$P(T \leq g, \text{you win}) = P(T \leq g, \text{and all } X_i \text{ are outside of } [2T - g, g]).$$

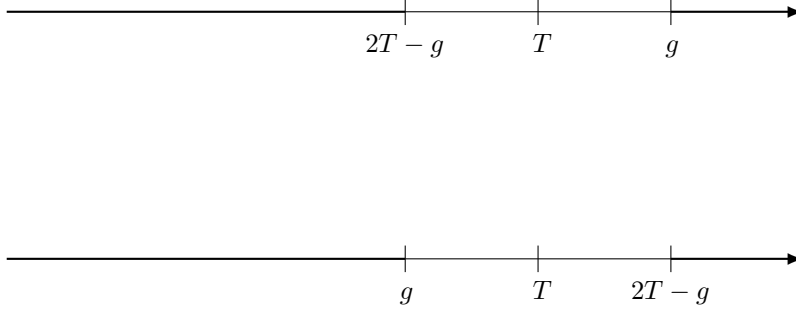
If you are to win, then  $k$  of the  $X_i$ 's must be greater than  $g$  and  $n - k$  must be less than  $2T - g$ , for some  $k$  between 0 and  $n$ . So your probability of winning is

$$\sum_{k=0}^n \binom{n}{k} P(T \leq g, X_1, \dots, X_k > g, \text{ and } X_{k+1}, \dots, X_n < 2T - g).$$

We will now rewrite the above as a sum of integrals of the joint pdf  $f_{X_1, \dots, X_n, T}$ :

$$\sum_{k=0}^n \binom{n}{k} \int_{-\infty}^g \underbrace{\int_g^{\infty} \cdots \int_g^{\infty}}_{k \text{ times}} \underbrace{\int_{-\infty}^{2t-g} \cdots \int_{-\infty}^{2t-g}}_{n-k \text{ times}} f_{X_1, \dots, X_n, T}(x_1, \dots, x_n, t) dx_n \cdots dx_1 dt.$$

Figure 3:



By our independence assumption this becomes

$$\sum_{k=0}^n \binom{n}{k} \int_{-\infty}^g \int_g^{\infty} f_{X_1}(x_1) dx_1 \cdots \int_g^{\infty} f_{X_k}(x_k) dx_k \int_{-\infty}^{2t-g} f_{X_{k+1}}(x_{k+1}) dx_{k+1} \cdots \int_{-\infty}^{2t-g} f_{X_n}(x_n) dx_n f_T(t) dt.$$

Each of the latter  $n - k$  integrals is equal to  $F_X(2t - g)$  and each of the former  $k$  integrals is equal to  $1 - F_X(g)$ . So we have

$$\sum_{k=0}^n \binom{n}{k} \int_{-\infty}^g (1 - F_X(g))^k F_X(2t - g)^{n-k} f_T(t) dt.$$

Passing the sum inside the integral and applying the binomial theorem yields

$$P(T \leq g, \text{you win}) = \int_{-\infty}^g (1 - F_X(g) + F_X(2t - g))^n f_T(t) dt.$$

A similar computation produces the analogous result in the  $T > g$  case:

$$P(T > g, \text{you win}) = \int_g^{\infty} (1 - F_X(2t - g) + F_X(g))^n f_T(t) dt.$$

These two equalities imply formula (1), which completes the proof of Theorem 3. The rewriting of formula (1) as (2) is straightforward.

### 3 PROOF OF THEOREM 2.

The representation of  $W_n(g)$  in terms of  $h_g(t)$  as given in equation (2) motivates our proof. Fix  $g$  as in the hypotheses of Theorem 2 and fix  $\epsilon > 0$ . Focussing

for now on the positivity of the right-hand limit at  $g$ , we first bracket this limit close enough to obtain positive numbers  $p_1$ ,  $p_2$ , and  $\delta$  so that  $1 - \epsilon < p_1/p_2 < p_2/p_1 < 1 + \epsilon$  and

$$p_1 < f(x) < p_2 \quad (3)$$

for  $x$  in  $(g, g + \delta)$ . If necessary, we shrink  $\delta$  so that  $p_2\delta < 1$ . Note that if  $t \geq g$ ,

$$h_g(t) = P(X_1 \notin [g, 2t - g]) = 1 - \int_g^{2t-g} f(x) dx.$$

To use inequality (3) in estimating the integrand we must have  $x$  in  $(g, g + \delta)$ , for which it is sufficient to restrict  $t$  to the interval  $(g, g + \delta/2)$ . Thus, the right inequality in (3) implies

$$h_g(t) \geq 1 - 2p_2(t - g) \text{ for all } t \text{ in } (g, g + \delta/2). \quad (4)$$

Note also that  $p_2\delta < 1$  implies that the right side of inequality (4) is strictly positive. Now, the left inequality in (3), together with (4) and a simple antiderivative, can be used to conclude that

$$\begin{aligned} \int_g^{g+\delta/2} (n+1)h_g(t)^n f(t) dt &\geq \frac{p_1}{p_2} \left( \frac{1}{2} - \frac{1}{2}(1 - p_2\delta)^{n+1} \right) \\ &> \frac{p_1}{p_2} \left( \frac{1}{2} - \frac{\epsilon}{2} \right) \end{aligned}$$

if  $n$  is sufficiently large. Because  $p_1/p_2 > 1 - \epsilon$  it follows that for such  $n$ ,

$$\frac{1}{2} - \epsilon < \int_g^{g+\delta/2} (n+1)h_g(t)^n f(t) dt.$$

Similar estimates using inequality (3) in the other direction show that

$$\int_g^{g+\delta/2} (n+1)h_g(t)^n f(t) dt < \frac{1}{2} + \frac{\epsilon}{2},$$

and hence for sufficiently large  $n$  we have

$$\frac{1}{2} - \epsilon < \int_g^{g+\delta/2} (n+1)h_g(t)^n f(t) dt < \frac{1}{2} + \frac{\epsilon}{2}. \quad (5)$$

Now we consider the integral  $\int_{g+\delta/2}^{\infty} (n+1)h_g(t)^n f(t) dt$ . The function  $h_g$  is decreasing on  $(g, \infty)$ , so  $h_g(t) \leq h_g(g + \delta/2) \leq 1 - p_1\delta$  if  $t > g + \delta/2$ . Thus, for sufficiently large  $n$  we have

$$\begin{aligned} 0 \leq \int_{g+\delta/2}^{\infty} (n+1)h_g(t)^n f(t) dt &\leq (n+1)(1 - p_1\delta)^n \int_{g+\delta/2}^{\infty} f(t) dt \\ &< (n+1)(1 - p_1\delta)^n \\ &< \frac{\epsilon}{2}. \end{aligned}$$

Combining the above with inequality (5), we see that for large  $n$ ,

$$\frac{1}{2} - \epsilon < \int_g^\infty (n+1)h_g(t)^n f(t) dt < \frac{1}{2} + \epsilon,$$

so that

$$\lim_{n \rightarrow \infty} \int_g^\infty (n+1)h_g(t)^n f(t) dt = \frac{1}{2}.$$

Similar reasoning applies to the integral on  $(-\infty, g)$ , so

$$\lim_{n \rightarrow \infty} \int_{-\infty}^g (n+1)h_g(t)^n f(t) dt = 1,$$

and the result follows.

## 4 THE UNIFORM CASE.

We now specialize to the game described at the beginning of Section 1, where the other players' guesses and the target are uniformly distributed with cdf  $F$ . In this case,

$$f(x) = f_{X_1}(x) = f_{X_2}(x) = \dots = f_{X_n}(x) = f_T(x) = \begin{cases} 0 & \text{if } x \notin [0, 1], \\ 1 & \text{if } x \in [0, 1]. \end{cases}$$

Since  $F(2t-g)$  is 1 for  $2t-g \geq 1$ , 0 for  $2t-g \leq 0$ , and equal to  $2t-g$  when  $2t-g \in (0, 1)$ , we have:

$$F(2t-g) = \begin{cases} 0 & \text{if } t \leq \frac{g}{2}, \\ 2t-g & \text{if } \frac{g}{2} < t < \frac{1+g}{2}, \\ 1 & \text{if } t \geq \frac{1+g}{2}. \end{cases}$$

Assume  $0 \leq g \leq 1$ . To calculate  $W_n(g)$  here, we will break up the first integral in equation (1) as

$$\int_0^{\frac{g}{2}} (1 - F(g) + F(2t-g))^n f(t) dt + \int_{\frac{g}{2}}^g (1 - F(g) + F(2t-g))^n f(t) dt,$$

and the second integral as

$$\int_g^{\frac{1+g}{2}} (1 + F(g) - F(2t-g))^n f(t) dt + \int_{\frac{1+g}{2}}^1 (1 + F(g) - F(2t-g))^n f(t) dt.$$

Applying what we know about the functions  $F$  and  $f$ , we deduce that  $W_n(g)$  is given by

$$\int_0^{\frac{g}{2}} (1-g)^n dt + \int_{\frac{g}{2}}^g (1-2g+2t)^n dt + \int_g^{\frac{1+g}{2}} (1+2g-2t)^n dt + \int_{\frac{1+g}{2}}^1 g^n dt.$$

Routine integration yields

$$W_n(g) = \frac{1}{n+1} + \frac{1}{2} \left( g(1-g)^n + g^n(1-g) - \frac{1}{n+1}(g^{n+1} + (1-g)^{n+1}) \right). \quad (6)$$

Rearranging terms, we have

$$2(n+1)W_n(g) - 2 = -g^{n+1} + (n+1)g^n(1-g) + (n+1)(1-g)^n g - (1-g)^{n+1}, \quad (7)$$

showing that the graph of  $W_n$  has the same shape as the polynomial in  $g$  on the right-hand side but shifted and scaled.

For large  $n$ , this polynomial is easily seen to take very small positive values on a large portion of the center of the interval  $[0, 1]$ , but its behavior near the endpoints requires more scrutiny. Despite taking its minimum value of  $-1$  at the endpoints  $0$  and  $1$ , the polynomial takes a maximum value approaching  $e^{-2}$  close to the points  $g_1 = 2/(n+2)$  and  $g_2 = n/(n+2)$ , thereby producing through formula (6) the small peaks in the graph of  $W_n$  when  $n \geq 5$ .

To see this more clearly, let us first restrict our attention to  $g$  in the interval  $[0, 1/2]$ . Here the first two terms of the right-hand side of equation (7) approach  $0$  uniformly, so the two sides of the approximation

$$2(n+1)W_n(g) - 2 \approx (n+1)(1-g)^n g - (1-g)^{n+1}$$

become uniformly close on  $[0, 1/2]$  as  $n$  increases. It is a routine calculus exercise to verify that the right-hand side attains at  $g = g_1 = 2/(n+2)$  its maximum value of  $(1 - 2/(n+2))^n$ , which approaches  $e^{-2}$  as  $n \rightarrow \infty$ . From this it follows that

$$(n+1)W_n(g_1) \rightarrow 1 + \frac{e^{-2}}{2}, \quad (8)$$

a result that, with a little more effort, can be derived directly from formula (6) via l'Hôpital's Rule. By symmetry of  $W_n$  the same limit is approached near the right endpoint when  $g = g_2 = n/(n+2)$ . Note that (8) holds even though Theorem 2 guarantees that for any fixed  $g$  in the open interval  $(0, 1)$  we have  $(n+1)W_n(g) \rightarrow 1$ .

The overshoot revealed by (8) may remind us of the Gibbs phenomenon encountered in approximations by Fourier series near a jump discontinuity of the limit function [4, pp. 62–66]. Here, this behavior may be related to constraints encountered in approximations by Bernstein polynomials, as we shall note below. Let us first summarize, however, the results of the foregoing discussion of the uniform case.

**Remark 4.** If  $n \geq 5$ , the maximum of  $W_n(g)$  given by formula (6) occurs at approximately  $g_1 = 2/(n+2)$  and  $g_2 = n/(n+2)$ , and  $\max W_n(g)/(1/(n+1))$  approaches  $1 + 1/(2e^2)$  as  $n$  approaches infinity.

The implications for anyone who is playing the guessing game described at the beginning of the paper are as follows:

- If you know precisely how many players  $n$  are in the game, and if  $n$  is large, you should guess  $2/(n+2)$  or, equivalently, guess  $n/(n+2)$ .
- This strategy will provide you with a win probability approximately  $1/(2e^2) \approx 6.767\%$  better than a uniform random guess.

Readers familiar with approximation properties expected of Bernstein polynomials [5] may be unsurprised by (8), once it is recognized that the polynomial on the right-hand side of formula (7) is the  $(n+1)$ th Bernstein approximation of a function  $f$  specified to be identically zero on  $[0, 1]$  except near the endpoints, where  $f(0) = -1 = f(1)$  and  $f(1/(n+1)) = 1 = f(n/(n+1))$ . The polynomial must then agree with  $f$  at the endpoints and must “try” to approximate  $f$  uniformly on  $[0, 1]$ , while of course encountering difficulty because of the quick fluctuations of  $f$  near the endpoints. The resulting trade-off produces the shape of the graph of this Bernstein polynomial and, through formula (6), produces overshoots in  $W_n(g)$  near the endpoints.

The uniform case discussed in this section is one of the few situations where we have been able to simplify equation (1) to obtain an explicit formula for  $W_n(g)$ . As we shall observe in the next section, though, the overshoots in the uniform case appear to manifest themselves in more general settings at points naturally corresponding to  $g_1$  and  $g_2$ . In fact, the existence, for large  $n$ , of overshoots near 0 and 1 in the uniform case may be seen to be due not so much to the fact that these points are endpoints, but that they are “unreasonable guesses” that lie in the support of the pdf under consideration.

## 5 NON-UNIFORM DISTRIBUTIONS.

For non-uniform distributions, the integrals in equation (1) are very difficult or impossible to evaluate symbolically, so we asked Mathematica to estimate them numerically. Shown here are the graphs of  $W_n$  when  $n = 9, 99,$  and  $999$  for several different distributions. In accordance with Theorem 2, the graphs of  $W_n$  approach  $1/(n+1)$  asymptotically at all reasonable values of  $g$ . We will comment shortly on some of the unreasonable guesses.

Figure 4: Win probability as a function of your guess, with target and guesses distributed according to the pdf  $f(x) = 2x$  on  $[0, 1]$ .

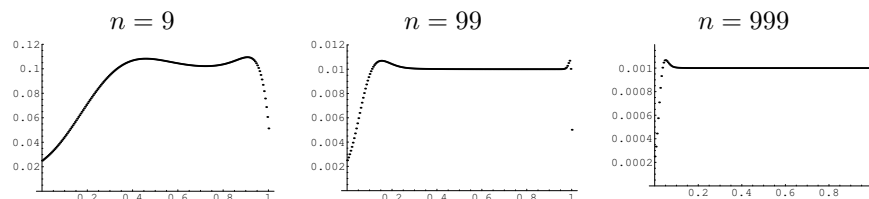
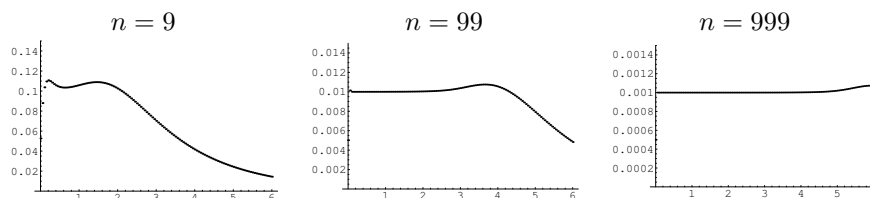


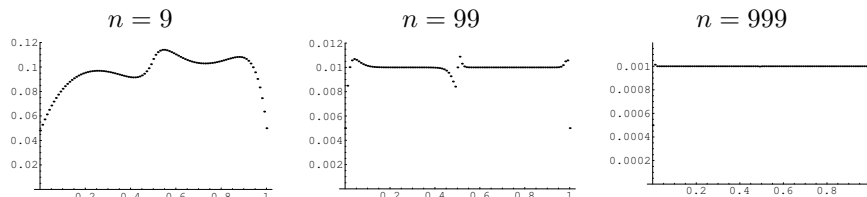
Figure 5: Win probability as a function of your guess, with target and guesses distributed according to the pdf  $f(x) = e^{-x}$  on  $[0, \infty)$ .



In Section 4 we found that the uniform distribution of guesses produces the function  $W_n$  given by equation (6), whose graph exhibits small peaks near  $g_1 = 2/(n+2)$  and  $g_2 = n/(n+2)$  if  $n \geq 5$ . For other distributions we might be led on intuitive grounds (see [6, Proposition 9.3.1]) to expect the resulting  $W_n$  to approximate the composition of the function of equation (6) with the cdf  $F$  of the distribution in question, provided we are at a point where  $F$  has a positive derivative. We should then generally expect Mathematica to plot new peaks near  $F^{-1}(g_1)$  and  $F^{-1}(g_2)$  when  $n \geq 5$ , and this seems to be the case in Figures 2, 4, and 5. In Figure 4, for example, the peaks seem to be located, as we should expect, near  $\sqrt{g_1}$  and  $\sqrt{g_2}$ , for  $n = 9, 99$ , and  $999$ , respectively.

Surprises may occur, however, in more complicated settings where the cdf fails to be differentiable or where its derivative vanishes, as seen in Figures 6, 7, and 8. Also, at the end of this section we note what can happen if the graph of the cdf has a vertical tangent.

Figure 6: Win probability as a function of your guess, with target and guesses distributed according to the “piecewise uniform” distribution with pdf  $f(x) = 1/2$  on  $[0, 1/2]$  and  $f(x) = 3/2$  on  $(1/2, 1]$ .



Recall that the pdf of the distribution need not be continuous in order for Theorem 2 to hold, but the example in Figure 6 shows that some intriguing behavior may occur near a point of discontinuity. In particular, we did not expect to see a local minimum just to the left of  $1/2$ .

The two distributions used in Figures 7 and 8 are very similar and in each

Figure 7: Win probability as a function of your guess, with target and guesses distributed according to the pdf  $f(x) = 4|x - 1/2|$  on the unit interval.

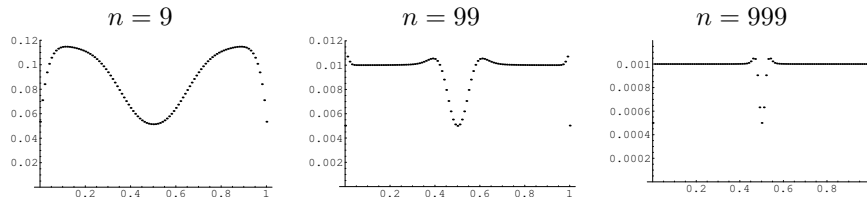
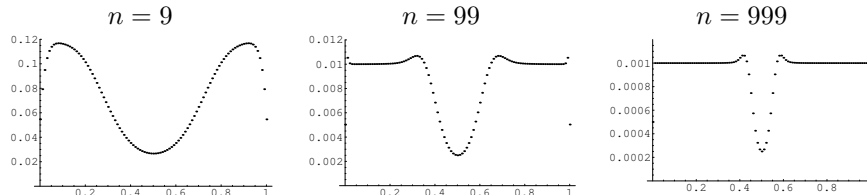


Figure 8: Win probability as a function of your guess, with target and guesses distributed according to the pdf  $f(x) = 12(x - 1/2)^2$  on the unit interval.



case the hypotheses of Theorem 2 fail to be satisfied at  $g = 1/2$ . We should therefore not be too surprised that  $W_n(1/2)$  does not approach  $1/(n+1)$  asymptotically. That the limiting value of  $(n+1)W_n(1/2)$  differs between the two cases is attributable to the following generalization of Theorem 2.

**Theorem 5.** *Let  $X_1, X_2, \dots, X_n$ , and  $T$  be independent random variables from the same continuous distribution and let  $f$  be their common pdf. For a given  $g$ , suppose there exist real numbers  $k$  and  $k'$ , both greater than  $-1$ , such that  $\lim_{t \rightarrow g^+} \frac{f(t)}{|t - g|^k}$  and  $\lim_{t \rightarrow g^-} \frac{f(t)}{|t - g|^{k'}}$  both exist and are nonzero. Then*

$$\lim_{n \rightarrow \infty} (n+1)W_n(g) = \frac{1}{2^{k+1}} + \frac{1}{2^{k'+1}}. \quad (9)$$

Theorem 2 is Theorem 5 with  $k = k' = 0$ . Fortunately, the proof of Theorem 2 generalizes to this larger setting, although we encounter slightly more complicated expressions. The requirement  $p_2\delta < 1$  made at the outset of the proof of Theorem 2 must be modified to ensure that  $p_2\delta^{k+1} < k+1$  in general. Other modifications suggest themselves. We ask the reader to fill in the details of the proof of Theorem 5.

What does Theorem 5 tell us? At  $g = 0$  in Figure 4, we have essentially a “one-sided” case where  $k = 1$ , so  $(n+1)W_n(0) \rightarrow 1/4$ . In Figure 7, we have

$k = k' = 1$  at  $g = 1/2$ , so  $(n + 1)W_n(1/2) \rightarrow 1/2$ ; while in Figure 8 we have  $k = k' = 2$ , so  $(n + 1)W_n(1/2) \rightarrow 1/4$ .

We rarely deal with unbounded pdf's, but they can produce interesting effects, such as making the right-hand side of equation (9) exceed 1 and forcing the graph of the associated cdf to have a vertical tangent. Consider, for example, the distribution on  $[-1, 1]$  given by the pdf  $f(x) = (1/4)|x|^{-1/2}$ . Here,  $k = k' = -1/2$  at  $g = 0$ , so  $(n + 1)W_n(0) \rightarrow \sqrt{2}$ . Moreover, despite what our foregoing discussions might have led us to expect,  $W_n$  seems to attain its maximum here for all  $n$  at  $g = 0$ .

## 6 RELATED QUESTIONS.

Here we mention some issues that, for lack of space and/or insight, seem to deserve more attention than we can provide.

### 6.1 UNREASONABLE GUESSES AND INDUCED OVERSHOTS?

More precise definitions may help us better study the intriguing relationship suggested in the Mathematica plots we have seen. To that end, we define an *unreasonable guess* to be a point  $g$  lying in the support of the common pdf  $f$  of the  $X_i$ 's and  $T$  such that  $\limsup(n + 1)W_n(g) = L < 1$ . Theorem 5 often enables us to calculate  $L$  easily, and shows that unreasonable guesses include not only endpoints but also points like  $1/2$  in Figures 7 and 8 where the pdf tends to zero.

We will say that we have an *overshoot* at  $g$  if for each neighborhood  $N$  of  $g$ ,  $\liminf M_n = M > 1$ , where  $M_n$  is the supremum of  $(n + 1)W_n$  on  $N$ . Remark 4 confirms that we have overshoots ( $M = 1 + e^{-2}/2$ ) at 0 and at 1 in the uniform case. The heuristic argument in Section 5 suggests that we should expect the same value of  $M$  associated with all unreasonable guesses in Figures 1, 2, 4, and 5 — where in Figure 2 we should consider  $\infty$  and  $-\infty$  to be unreasonable guesses.

We do not know whether there must necessarily be an overshoot at each unreasonable guess. Note, however, that if you decide to guess according to the same pdf  $f$  as your  $n$  friends then you would be indistinguishable from them, so all of you would have the same win probability of  $1/(n + 1)$ . Therefore,

$$\int_{-\infty}^{\infty} W_n(g)f(g) dg = \frac{1}{n + 1}. \quad (10)$$

Hence, for each  $n$ , the continuous function  $(n + 1)W_n$  integrates to 1 against the pdf  $f$ . This remark, together with Theorem 2, seems to make it at least plausible that if  $g$  is a point in the support of  $f$  where the  $\limsup L$  is less than 1, then the  $\liminf M$  should exceed 1, but we don't see a proof.

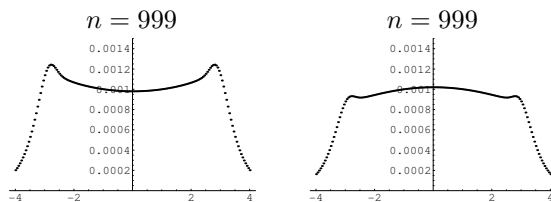
## 6.2 WHEN THE GUESSES DIFFER FROM THE TARGET.

While most of this paper has focussed on the case where the target and your friends' guesses come from the same distribution, equation (1) is general enough to handle situations where the target random variable  $T$  differs from the guesses  $X_1, \dots, X_n$ . Suppose for example that the target is standard normal. Theorem 2 says that, if your friends' guesses are also standard normal, then for large enough  $n$ , any guess you make is essentially as good as any other.

Suppose instead, however, that your friends' guesses cluster around zero more tightly than the target. Then guesses near zero will be crowded out, hence becoming less profitable for you, and guesses far from zero (though not too far) will become more profitable. If, on the other hand, your friends' guesses are slightly more spread out than the target, then guessing near zero becomes a better strategy for you.

Shown in Figure 9 are pictures of  $W_n$  for  $n = 999$  if the target  $T$  is standard normal while your friends' guesses are normally distributed with mean zero and standard deviations .98 and 1.02, respectively. Compare these pictures with Figure 2.

Figure 9: Win probability as a function of your guess, where the distribution of guesses differs from distribution of the target.



## 6.3 THE PRICE IS RIGHT.

One variation on the game is to play it “Price is Right” style, where the winner is the player whose guess is closest to the target *among all guesses less than the target*. Here, a player is disqualified if he or she “overbids.” Following the strategy outlined in the proof of Theorem 3, the reader may verify that in this case the probability of winning with a guess of  $g$  is given by

$$W_n(g) = \int_g^\infty (1 - F_X(t) + F_X(g))^n f_T(t) dt,$$

where  $F_X$  represents the cdf of the guesses and  $f_T$  represents the pdf of the target.

If the guesses are assumed to come from the same distribution as the target, and if the pdf of the target is a continuous function, then this integral is surprisingly simple to evaluate:

$$W_n(g) = \frac{1}{n+1}(1 - F_X(g)^{n+1}).$$

Note that  $W_n$  is decreasing (no overshoots here), so it follows that the optimal strategy is to make the minimum allowable guess. If the “least guess” is an allowable option, it gives you the optimal win probability of  $1/(n+1)$ . Also note that, in effect, there are  $n+2$  players here, since the house wins if you and your  $n$  friends all overbid.

## 6.4 AN APPLICATION?

We close with a few remarks about the motivation for this problem. This paper arose out of an attempt by the first author to build an abstract model for the optimal selection of an entry in an NCAA basketball tournament prediction contest with a large number of entrants, such as those run each year by major sports websites like ESPN.com and yahoo.com.

There are some vague similarities between a basketball pool and the game described in this paper. In both cases, something will happen, lots of people will make a guess at what that something will be, and the one who is closest will be declared the winner. In the abstract game of this paper, the participant must choose between making a guess in the tall part of the target distribution and potentially being crowded out by many other participants making similar guesses, and making a guess in the short part of the target distribution, which has less likelihood of being close to the target but a better chance of winning if it does happen to be close. In the basketball pool, the former choice corresponds to picking the teams who are favored to win while the latter corresponds to choosing an entry with many long shot teams.

If you believe that the entries in a large pool satisfy something approximating the conditions of Theorem 2, then our results might be loosely interpreted to mean that it doesn’t really matter what kind of entry you fill out. Serious basketball fans might be disappointed to learn that all their basketball knowledge is for naught in these large contests. Casual basketball fans, on the other hand, might find this result liberating. They can fill in their brackets based on the geography of the teams, their uniform color, or the relative fierceness of their mascots, and their chances will be as good as anyone else’s.

While we were pleased with the mathematics that arose from the investigation, even the first author must concede that the connection between our abstract game and its real-life inspiration is probably too loose to offer much practical advice for those who wish to enter such contests.

**Dedication.** This paper is dedicated to Sherwood F. Ebey, whose enthusiasm for the study of probability has proven to be infectious.

## References

- [1] S. Even, The price is right game, this MONTHLY **73** (1966) 180–182.
- [2] T. Feder, Toetjes, this MONTHLY **97** (1990) 785–794.
- [3] T. Ferguson and C. Genest, Toetjes na, *Mathematical Statistics and Applications: Festschrift for Constance van Eeden, IMS Lecture Notes/Monograph Series* **42** (2003) 169–184.
- [4] T. Körner, *Fourier Analysis*, Cambridge University Press, Cambridge, 1988.
- [5] G. Lorentz, *Bernstein Polynomials*, Chelsea Publishing, 1986.
- [6] S. Ross, *Probability Models for Computer Science*, Harcourt/Academic Press, San Diego, 2002.
- [7] J. Steele and J. Zidek, Optimal strategies for second guessers, *Journal of the American Statistical Association* **75** (1980) 596–601.

**Doug Drinen** is an associate professor at the University of the South. He received a B.A. from Trinity University (TX) in 1993 and a Ph.D. from Arizona State in 1999. He only very occasionally thinks about anything other than sports or math, so he is always excited when he can save some time by thinking about both simultaneously.

*Department of Mathematics and Computer Science, Sewanee: The University of the South, Sewanee, TN 37383*  
*ddrinen@sewanee.edu*

**K. Grace Kennedy** graduated from the University of the South in 2006 majoring in mathematics and French literature. She is currently pursuing a Masters 2 in group theory at the Université de Picardie in Amiens, France. In the spirit of the liberal arts tradition, Grace will try to see every work of art in the Louvre that she studied in Art History 103 at Sewanee. Soon, she hopes to return to the US for graduate study on group theory topics related to generalizations of the braid group.

*kgracekennedy@alumni.sewanee.edu*

**Mac Priestley** has taught at the University of the South since 1967, having received a B.A. there in 1962. His Ph.D. degree is from Princeton. He wrote an elementary text, *Calculus: A Liberal Art* (Springer-Verlag, 1998), to support a one-semester general education course for liberal arts majors who may lack a strong background in mathematics. In 2006 he received the Distinguished Teaching Award from the Southeastern Section of the MAA.

*Department of Mathematics and Computer Science, Sewanee: The University of the South, Sewanee, TN 37383*  
*wpriestl@sewanee.edu*