A probability model for golf putting*

Andrew Gelman† and Deborah Nolan‡

October 17, 2001

Abstract

We derive a model, using trigonometry and the normal distribution, for the probability that a golf putt is successful. We describe a class activity in which we lead the students through the steps of examining the data, considering possible models, constructing a probability model, and checking the fit. The model is, of necessity, oversimplified, a point which the class discusses at the end of the demonstration.

1 Looking at data on golf putts

Golf is a harder game than it looks. A study of professional golf players found that they made less than 60% of their five-foot putts. Figure 1 shows the success rate of golf putts as a function of distance from the hole. (We found these data in the textbook by Berry, 1995, and this example is discussed further in Gelman and Nolan, 2001. Further quantitative information on golf putting appears in Pelz, 1989.) What do these data tell us about the accuracy of pro golf putts?

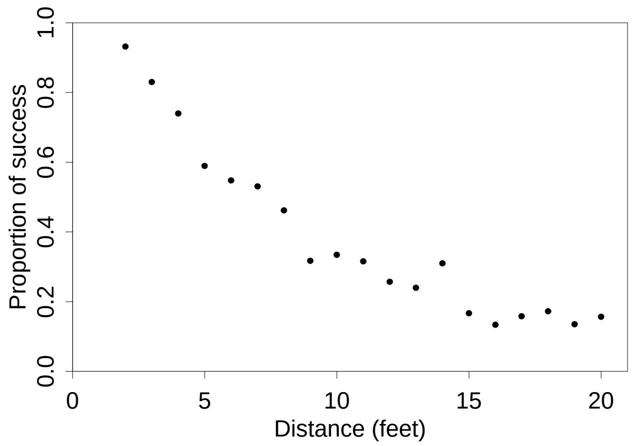

Figure 1: Success rate of golf putts, as a function of distance from the hole.

*To appear in *Teaching Statistics*. We thank the editor for helpful comments.
†Department of Statistics, Columbia University, New York, gelman@stat.columbia.edu
‡Department of Statistics, University of California, Berkeley, nolan@stat.berkeley.edu

We use these data to motivate a derivation of a probability model for our mathematical statistics class. It is a nice example because it involves some trigonometry, and it gives the students a sense of the interplay between mathematics, probability, and statistics. It is fun to see that, with a little mathematical reasoning, we can learn something new.

We start by estimating the standard error of the estimated probability of success at each distance, so that we have a sense of how closely our model should be expected to fit the data. Each point in Figure 1 is an estimate of the form y/n (for example, $208/353 = 59\%$ of five-foot putts were made) with an estimated standard error of $\sqrt{(y/n)(1-y/n)/n}$. Figure 2 repeats the graph with ± 2 standard errors, which correspond to approximate 95% confidence intervals.

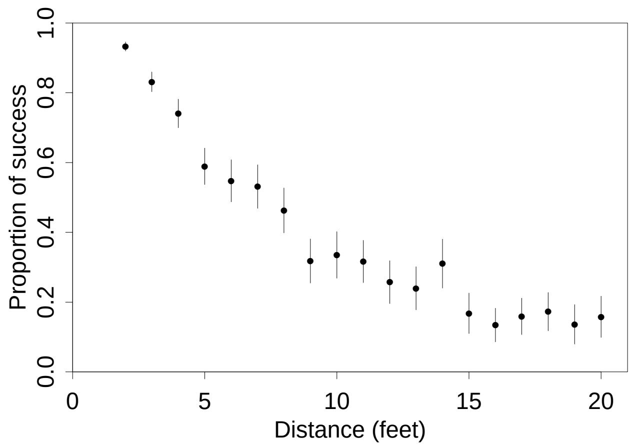

Figure 2: Success rate of golf putts, as a function of distance from the hole, with vertical lines showing 95% confidence intervals based on the normal approximation to the binomial distribution.

2 Constructing a probability model

We now ask what sort of model could fit the data in Figure 2. With the class, we discuss the possibilities. Clearly a linear regression is inappropriate, given the evident curve in the data. What about a quadratic? This runs into problems because the probabilities are bounded between 0 and 1. What should happen at the extremes of the distance x from the hole? The probability of success must have an asymptote and approach 0 as $x \to \infty$. Also, shots from a zero distance must go in, and so the success probability at distance 0 must be 1.

We then sketch a sketch of the golf shot (see Figure 3) on the blackboard. Simple trigonometry shows that the shot goes in the hole if its angle is less, in absolute value, than the threshold angle $\theta_0 = \sin^{-1}((R-r)/x)$. The students can work in pairs or small groups here to discover this relation.

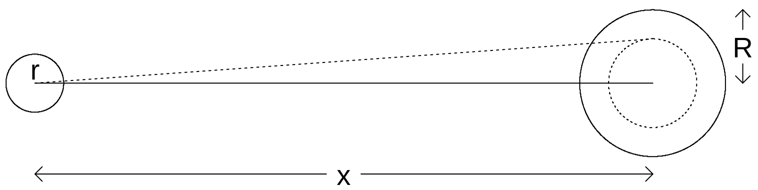

Figure 3: Diagram showing the probability model for the accuracy of golf putts. The golf ball has diameter $2r = 1.68''$ and the hole has diameter $2R = 4.25''$. The shot goes in the hole if the error in its angle is less than $\theta_0 = \sin^{-1}((R - r)/x)$; this is the angle between the solid and dotted lines in the picture.

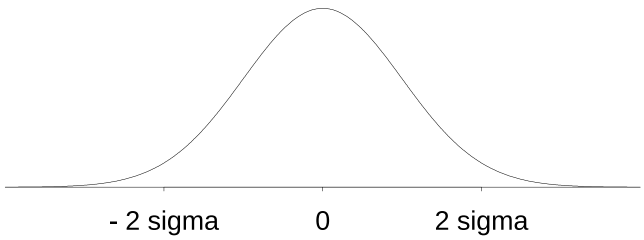

Figure 4: Assumed normal distribution for the angle of error of the golf shot: $\theta = 0$ is a perfect shot, and the shot goes in the hole if $|\theta| < \theta_0$, where the threshold error, θ_0, depends on the distance x from the hole.

How does this translate into the probability of a successful shot? The only random variable here is θ, and so we need to assign a distribution to it. A normal distribution seems reasonable (why?), presumably centered at $\theta = 0$ (assuming that shots do not systematically list to the left or the right), in which case the only parameter is the standard deviation, σ (see Figure 4).

Using this distribution, the probability the ball goes into the hole is,

$$\Pr(\text{success of a shot from distance } x) = 2\Phi\left[\frac{1}{\sigma}\sin^{-1}\left(\frac{R-r}{x}\right)\right] - 1,$$

where Φ is the normal cumulative distribution function. (If $x < R - r$, then $\sin^{-1}((R-r)/x)$ is not defined, but in this case the model is not needed since the ball is already in the hole!)

The unknown parameter σ can be estimated by fitting to the data in Figure 2. (We fit this using nonlinear least squares, but we do not go into this in class—we just say that we fit the curve to the data.) The resulting estimate is $\hat{\sigma} = 0.026$ (which, when multiplied by $180/\pi$, comes out to $1.5°$), and the fitted curve is shown overlain on the data in Figure 5. The model fits pretty well.

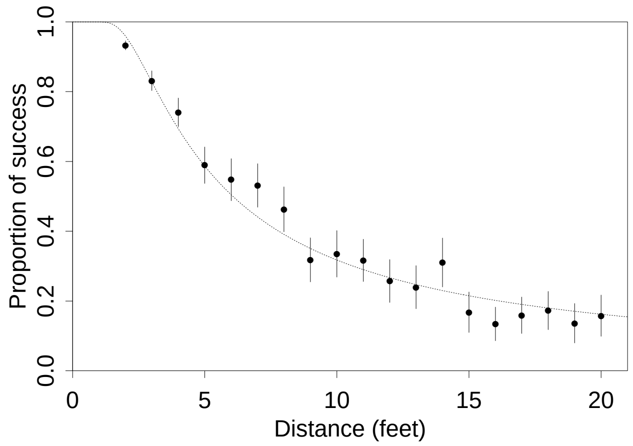

Figure 5: Success rate of golf putts, as a function of distance from the hole, along with a fitted curve, Pr (success of a shot from distance x)= $2\Phi(\sin^{-1}[(R-r)/x]/\sigma)-1$, with $\sigma = 0.026$ (that is, $1.5°$) estimated from the data.

3 Checking the fit of the model to the data

The fit of the curve is not perfect, however. As the vertical bars on the graph indicate, several of the 95% confidence intervals do not intersect the curve, and we can formally check this with a χ^2 test. We introduce the notation $i = 1, \ldots, 19$ for the data points in Figure 1, and x_i, y_i, n_i for the distance to the hole, the number of shots made at this distance, the number of shots attempted, respectively. Goodness of fit can then be tested using the Pearson statistic,

$$\sum_i \frac{(y_i - \mathrm{E}(y_i))^2}{\mathrm{var}(y_i)},$$

where $\mathrm{E}(y_i) = n_i \mathrm{Pr}(\text{success from distance } x_i)$ and $\mathrm{var}(y_i) = n_i \mathrm{E}(y_i)(1-\mathrm{E}(y_i))$. The expected values $\mathrm{E}(y_i)$ must be calculated given the estimated parameter σ. The value of the test statistic is 58 for the data and fitted curve in Figure 5. The nonlinear least squares method of estimating σ that we have used ensures that the approximate distribution of this statistic if the model is correct is χ^2_{18} (19 data points minus one degree of freedom for the parameter σ estimated from the data). Our result is clearly statistically significant (the 95th percentile of the χ^2_{18} distribution is 29).

So the model does not fit perfectly. Nonetheless, it seems pretty reasonable. At this point, we ask the students if they can see any reasons why the model might not be correct. One serious flaw is that the model does not allow for shots that miss because they are too short. In addition, the binomial error model assumes that the probability of success depends only on distance, which ignores variation among golf greens, playing conditions, and abilities of pro golfers. Including these complexities into the model would be difficult, but we explain to the students that the current model, for all its flaws, yields some insight into putting and into why the curve of success probabilities looks the way it does.

This demonstration can be elaborated upon by actually bringing a putter and a golf ball into class, having the students take shots, and marking off the distribution of their error angles.

References

Berry, D. (1995). *Statistics: A Bayesian Perspective.* Belmont, Calif.: Duxbury Press.

Gelman, A., and Nolan, D. (2001). *Bag of Tricks for Teaching Statistics.* Oxford University Press, to appear.

Pelz, D. (1989). *Putt Like the Pros.* Harper Collins.