

# Probabilistic Modeling of Golf Shots and Shot Risk

September 14, 2025

## 1 Part 1: Dispersion Modeling

For a given set of  $N$  shots we can model the set of shots with a bivariate gaussian distribution. In general, the model parameters  $\vec{\mu}$  and  $\Sigma$  uncertainty will scale with  $1/\sqrt{N}$ . We will test and compare the fractional uncertainty of different samples sizes and then compare how consistent additional real world shots are with the underlying distributions. This will set us up to understand how the distributions change given the sample size, which then will be later reflected in odds evaluation for shot recommendations. This test also will evaluate if the gaussian interpretation is reasonable. The mathematical formulation is as follows:

Suppose we have a set of  $N = 50$  shots. From these shots we will form 4 distributions  $f_{n_i}$  which parameters are estimated from a subset  $n_i$  number of shots.

$$\begin{aligned} f_{n_{10}}(\vec{k}_{10}) \\ f_{n_{20}}(\vec{k}_{20}) \\ f_{n_{35}}(\vec{k}_{35}) \\ f_{n_{50}}(\vec{k}_{50}) \end{aligned} \tag{1}$$

the parameter vector  $\vec{k}_i = (\vec{\mu}, \Sigma)$  with  $i$  being the samples size from  $n_i$ . Next we take  $j$  new shots in the real world and see if they are consistent with the predicted dispersion model. We will evaluate the consistency in two ways: (1) per shot consistency and (2) binomial probabilities. For (1) we can look at the distribution of Mahalanobis distances  $d^2$  such that:

$$d^2(x_j)_{n_i} = (x_j - \vec{\mu}_{n_i})^T \Sigma_{n_i} (x_j - \vec{\mu}_{n_i}) \tag{2}$$

Next for 2 we do a simple and interpretable counting of where things land via binomial probabilities. In the asymptotic limit we know a-priori where shots should land with respect to the gaussian, based on  $\sigma$  bands i.e. 68% of shots will land within  $1\sigma$ , 95% of shots will land within  $2\sigma$  and so on. From  $j$  shots we will count the trials and successes that the shots land within the expected ranges and compare this to the asymptotic limit.

## 1.1 Experimental instructions

We will evaluate the distributions for at least 3 clubs - long distance, medium distance, and short distance

### 1 Build our dispersion models

- For each club take 50 shots in the golf sim, recording where the shots land (carry+total+elevation)

### 2 Test our dispersion models

- For each club take 20 shots on a real course, aiming at the pin. record where the shots land (total+offside)
- We will adjust the models based on elevation in post-processing
- Make sure the weather is ideal: reasonable temps, humidity, and very little wind

## 2 Shot recommendation dependence with dispersion modeling

Using a dispersion model, we can evaluate the probability that a shot will land in a particular course area e.g. water hazard, bunker, fairway, green etc. From this course area probability, we can develop a heuristic that helps maximizing favorable probabilities, like fairway and green, and minimizes the risk of shots falling into hazards. To assess our shot risk we can adjust the aimline or club usage. Here we will evaluate recommendation sensitivity to the underlying sample size. For this test, we will curate our shots by choosing positions over holes with a high density of hazards where we can expect reasonable differences in recommendations. We also want the highest samples size models for each club since we will be switching between clubs per recommendation (if necessary). We define 5 shot positions,  $p_i$ , to test the recommendation dependence to  $n_i$ . To do this we need each club in the bag, and then we form a set of (at least 2)  $f_{n_i}(\vec{k}_{n_i})$  distributions. Then we take  $j$  new shots based on the recommendation of  $f_{n_i}$ . For each new shot we evaluate the Strokes Gained (SG) and ultimately assess the mean SG dependence on  $n_i$

## 2.1 Experimental instructions

### 1 Build the rest of the dispersion models

- For the rest of your clubs take 50 shots in the golf sim, recording where the shots land (carry+total+elevation)
- Use these dispersions for the recommendations in the next step

### 2 Test the recommendations of $n_i$ set of models

- Take 10 shots from the same position on a hole with no club recommendation in the real world, record where the shots land (total distance)(evaluating SG)
  - Take 10 shots from the same position on a hole using the  $f_{10}$  recommendation in the real world, record where the shots land (total distance) (evaluating SG)
  - Take 10 shots from the same position on a hole using the  $f_{50}$  recommendation in the real world, record where the shots land (total distance) (evaluating SG)
  - We will adjust the models based on elevation in post-processing
  - Make sure the weather is ideal: reasonable temps, humidity, and very little wind
- 3 Repeat Step 2 for a few more different shot positions on holes with different hazards

### 3 Shot recommendation impact on golf scores

Here we will put everything together, using our shot recommendations we will evaluate the impact on an overall golf score by seeing if the recommendations reduce the average shots taken to reach the green and measure improvement on a shot by shot basis by seeing if there is an average improvement to SG. We will evaluate this over 9 holes:  $H_1, H_2, \dots, H_9$  using 2 players. For the first set of control holes, each player will play the holes in sequence with no club recommendations. To then test the recommendations, all 9 holes will be played in sequence again where 1 player uses the most ideal  $f_{50}$  shot recommendations for every shot. The second player will use no recommendations and will always shoot first to avoid decision bias from the recommendations. In total, each hole will be played 4 times between the two players.

#### 3.1 Experimental instructions

- 1 Play the control games
  - 2 players play 9 holes in real life twice with no club recommendation
  - record the number of shots to green for each hole for each player (total distance per shot)
  - record the SG for each shot for each player
- 2 Test the recommendations from the best model
  - Play the same 9 holes in real life twice where one player uses shot recommendations for every shot
  - record the number of shots to green for each hole for each player (total distance per shot)

- record the SG for each shot for each player