A Statistician Plays Darts

Ryan J. Tibshirani* Andrew Price† Jonathan Taylor‡

Abstract

Darts is enjoyed both as a pub game and as a professional competitive activity. Yet most players aim for the highest scoring region of the board, regardless of their skill level. By modeling a dart throw as a 2-dimensional Gaussian random variable, we show that this is not always the optimal strategy. We develop a method, using the EM algorithm, for a player to obtain a personalized heatmap, where the bright regions correspond to the aiming locations with high (expected) payoffs. This method does not depend in any way on our Gaussian assumption, and we discuss alternative models as well.
Keywords: *EM algorithm, importance sampling, Monte Carlo, statistics of games*

1 Introduction

Familiar to most, the game of darts is played by throwing small metal missiles (darts) at a circular target (dartboard). Figure 1 shows a standard dartboard. A player receives a different score for landing a dart in different sections of the board. In most common dart games, the board's small concentric circle, called the "double bullseye" (DB) or just "bullseye", is worth 50 points. The surrounding ring, called the "single bullseye" (SB), is worth 25. The rest of the board is divided into 20 pie-sliced sections, each having a different point value from 1 to 20. There is a "double" ring and a "triple" ring spanning these pie-slices, which multiply the score by a factor of 2 or 3, respectively.

Not being expert dart players, but statisticians, we were curious whether there is some way to optimize our score. In Section 2, under a simple Gaussian model for dart throws, we describe an efficient method to try to optimize your score by choosing an optimal location at which to aim. If you can throw relatively accurately (as measured by the variance in the Gaussian model), there are some surprising places you might consider aiming the dart.

The optimal aiming spot changes depending on the variance. Hence we describe an algorithm by which you can estimate your variance based on the scores of as few as 50 throws aimed at the double bullseye. The algorithm is a straightforward implementation of the EM algorithm [DLR77], and the simple model we consider allows a closed-form solution. In Sections 3 and 4 we consider more realistic models, Gaussian with general covariance and skew-Gaussian, and we turn to importance sampling [Liu08] to approximate the expectations in the E-steps. The M-steps, on the other hand, remain analogous to the maximum likelihood calculations; therefore we feel that these provide nice teaching examples to introduce the EM algorithm in conjunction with Monte Carlo methods.

Not surprisingly, we are not the first to consider optimal scoring for darts: [Ste97] compares aiming at the T19 and T20 for players with an advanced level of accuracy, and [Per99] considers aiming at the high-scoring triples and bullseyes for players at an amateur level. In a study on

*Dept. of Statistics, Stanford University, ryantibs@stanford.edu
†Dept. of Electrical Engineering, Stanford University, adprice@stanford.edu
‡Dept. of Statistics, Stanford University, jtaylo@stanford.edu

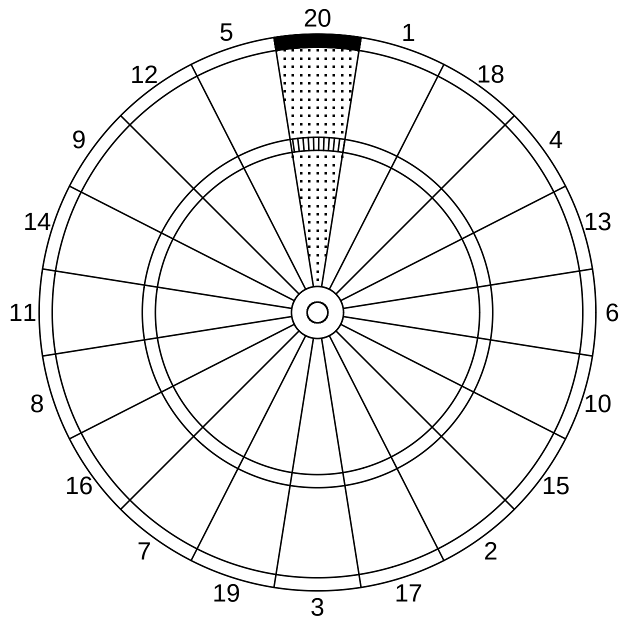

Figure 1: *The standard dartboard. The dotted region is called "single 20"* (S20)*, worth 20 points; the solid region is called "double 20"* (D20)*, worth 40 points; the striped region is called "triple 20"* (T20)*, worth 60 points.*

decision theory, [Kor07] displays a heatmap where the colors reflect the expected score as a function of the aiming point on the dartboard. In this paper we also compute heatmaps of the expected score function, but in addition, we propose a method to estimate a player's skill level using the EM algorithm. Therefore any player can obtain personalized heatmap, so long as he or she is willing to aim 50 or so throws at the bullseye.

It is important to note that we are not proposing an optimal strategy for a specific darts game. In some settings, a player may need to aim at a specific region and it may not make sense for the player to try to maximize his or her score. See [Koh82] for an example of paper that takes such matters into consideration. On the other hand, our analysis is focused on simply maximizing one's expected score. This can be appropriate for situations that arise in many common darts games, and may even be applicable to other problems that involve aiming at targets with interesting geometry (e.g. shooting or military applications, pitching in baseball).

Software for our algorithms is available as an R package [R D08], and also in the form of a simple web application. Both can be found at `http://stat.stanford.edu/~ryantibs/darts/`.

2 A mathematical model of darts

Let Z be a random variable denoting the 2-dimensional position of a dart throw, and let $s(Z)$ denote the score. Then the expected score is

$$\begin{aligned} \mathrm{E}[s(Z)] = {} & 50 \cdot \mathrm{P}(Z \in \mathrm{DB}) + 25 \cdot \mathrm{P}(Z \in \mathrm{SB}) + \\ & \sum_{i=1}^{20} \Big[i \cdot \mathrm{P}(Z \in \mathrm{S}i) + 2i \cdot \mathrm{P}(Z \in \mathrm{D}i) + 3i \cdot \mathrm{P}(Z \in \mathrm{T}i) \Big], \end{aligned}$$

where $\mathrm{S}i, \mathrm{D}i$ and $\mathrm{T}i$ are the single, double and triple regions of pie-slice i.

Perhaps the simplest model is to suppose that Z is uniformly distributed on the board B, that is, for any region S

$$\mathrm{P}(Z \in S) = \frac{\mathrm{area}(S \cap B)}{\mathrm{area}(B)}.$$

Using the board measurements given in A.1, we can compute the appropriate probabilities (areas) to get

$$\mathrm{E}[s(Z)] = \frac{370619.8075}{28900} \approx 12.82.$$

Surprisingly, this is a higher average than is achieved by many beginning players. (The first author scored an average of 11.65 over 100 throws, and he was trying his best!) How can this be? First of all, a beginner will occasionally miss the board entirely, which corresponds to a score of 0. But more importantly, most beginners aim at the 20; since this is adjacent to the 5 and 1, it may not be advantageous for a sufficiently inaccurate player to aim here.

A follow-up question is: where is the best place to aim? As the uniform model is not a very realistic model for dart throws, we turn to the Gaussian model as a natural extension. Later, in Section 3, we consider a Gaussian model with a general covariance matrix. Here we consider a simpler spherical model. Let the origin $(0,0)$ correspond to the center of the board, and consider the model

$$Z = \mu + \varepsilon, \ \ \varepsilon \sim \mathcal{N}(0, \sigma^2 I),$$

where I is the 2×2 identity matrix. The point $\mu = (\mu_x, \mu_y)$ represents the location at which the player is aiming, and σ^2 controls the size of the error ε. (Smaller σ^2 means a more accurate player.) Given this setup, our question becomes: what choice of μ produces the largest value of $\mathrm{E}_{\mu,\sigma^2}[s(Z)]$?

2.1 Choosing where to aim

For a given σ^2, consider choosing μ to maximize

$$\mathrm{E}_{\mu,\sigma^2}[s(Z)] = \iint \frac{1}{2\pi\sigma^2} e^{-\|(x,y)-\mu\|^2/2\sigma^2} \, s(x,y)\, dx\, dy. \tag{1}$$

While this is too difficult to approach analytically, we note that the above quantity is simply

$$(f_{\sigma^2} * s)(\mu),$$

where $*$ represents a convolution, in this case, the convolution of the bivariate $\mathcal{N}(0, \sigma^2 I)$ density f_{σ^2} with the score s. In fact, by the convolution theorem

$$f_{\sigma^2} * s = \mathcal{F}^{-1}\big[\mathcal{F}(f_{\sigma^2}) \cdot \mathcal{F}(s)\big],$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and inverse Fourier transform, respectively. Thus we can make two 2-dimensional arrays of the Gaussian density and the score function evaluated, say, on a millimeter scale across the dartboard, and rapidly compute their convolution using two FFTs (fast Fourier transform) and one inverse FFT.

Once we have computed this convolution, we have the expected score (1) evaluated at every μ on a fine grid. It is interesting to note that this simple convolution idea was not noted in the previous work on statistical modelling of darts [Ste97, Per99], with the authors using instead naive Monte Carlo to approximate the above expectations. This convolution approach is especially useful for creating a heatmap of the expected score, which would be infeasible using Monte Carlo methods.

Some heatmaps are shown in Figure 2, for $\sigma = 5, 26.9$, and 64.6. The latter two values were chosen because, as we shall see shortly, these are estimates of σ that correspond to author 2 and author 1, respectively. Here σ is given in millimeters; for reference, the board has a radius of 170 mm, and recall the rest of the measurements in A.1.

The bright colors (yellow through white) correspond to the high expected scores. It is important to note that the heatmaps change considerably as we vary σ. For $\sigma = 0$ (perfect accuracy), the optimal μ lies in the T20, the highest scoring region of the board. When $\sigma = 5$, the best place to aim is still (the center of) the T20. But for $\sigma = 26.9$, it turns out that the best place to aim is in the T19, close to the the border it shares with the 7. For $\sigma = 64.6$, one can achieve essentially the same (maximum) expected score by aiming in a large spot around the center, and the optimal spot is to the lower-left of the bullseye.

2.2 Estimating the accuracy of a player

Since the optimal location $\mu^*(\sigma)$ depends strongly on σ, we consider a method for estimating a player's σ^2 so that he or she can implement the optimal strategy. Suppose a player throws n independent darts, aiming each time at the center of the board. If we knew the board positions $Z_1, \dots Z_n$, the standard sample variance calculation would provide an estimate of σ^2. However, having a player record the position of each throw would be too time-consuming and prone to measurement error. Also, few players would want to do this for a large number of throws; it is much easier instead to just record the score of each dart throw.

In what follows, we use just the scores to arrive at an estimate of σ^2. This may seem surprising at first, because there seems relatively little information to estimate σ^2 just knowing the score, which for most numbers (for example, 13), restricts the position to lie in a relatively large region (pie-slice) of the board. This ambiguity is resolved by scores uniquely corresponding to the bullseyes, double rings, and triple rings, and so it is helpful to record many scores. Unlike recording the positions, it seems a reasonable task to record at least $n = 50$ scores.

Since we observe incomplete data, this problem is well-suited to an application of the EM algorithm [DLR77]. This algorithm, used widely in applied statistics, was introduced for problems in which maximization of a likelihood based on complete (but unobserved) data Z is simple, and the distribution of the unobserved Z based on the observations X is somewhat tractable or at least easy to simulate from. In our setting, the observed data are the scores $X = (X_1, \dots X_n)$ for a player aiming n darts at the center $\mu = 0$, and the unobserved data are the positions $Z = (Z_1, \dots Z_n)$ where the darts actually landed.

Let $\ell(\sigma^2; X, Z)$ denote the complete data log-likelihood. The EM algorithm (in this case estimating only one parameter, σ^2) begins with an initial estimate σ_0^2, and then repeats the following two steps until convergence:

E-step: compute $Q(\sigma^2) = \mathrm{E}_{\sigma_t^2}[\ell(\sigma^2; X, Z)|X]$;

M-step: let $\sigma_{t+1}^2 = \mathrm{argmax}_{\sigma^2}\ Q(\sigma^2)$.

With $\mu = 0$, the complete data log-likelihood is (up to a constant)

$$\ell(\sigma^2; X, Z) = \begin{cases} -n \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Z_{i,x}^2 + Z_{i,y}^2) & \text{if } X_i = s(Z_i)\ \forall i \\ -\infty & \text{otherwise,} \end{cases}$$

Therefore the expectation in the E-step is

$$\mathrm{E}_{\sigma_0^2}[\ell(\sigma^2; X, Z)|X] = -n \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n \mathrm{E}_{\sigma_0^2}(Z_{i,x}^2 + Z_{i,y}^2 | X_i).$$

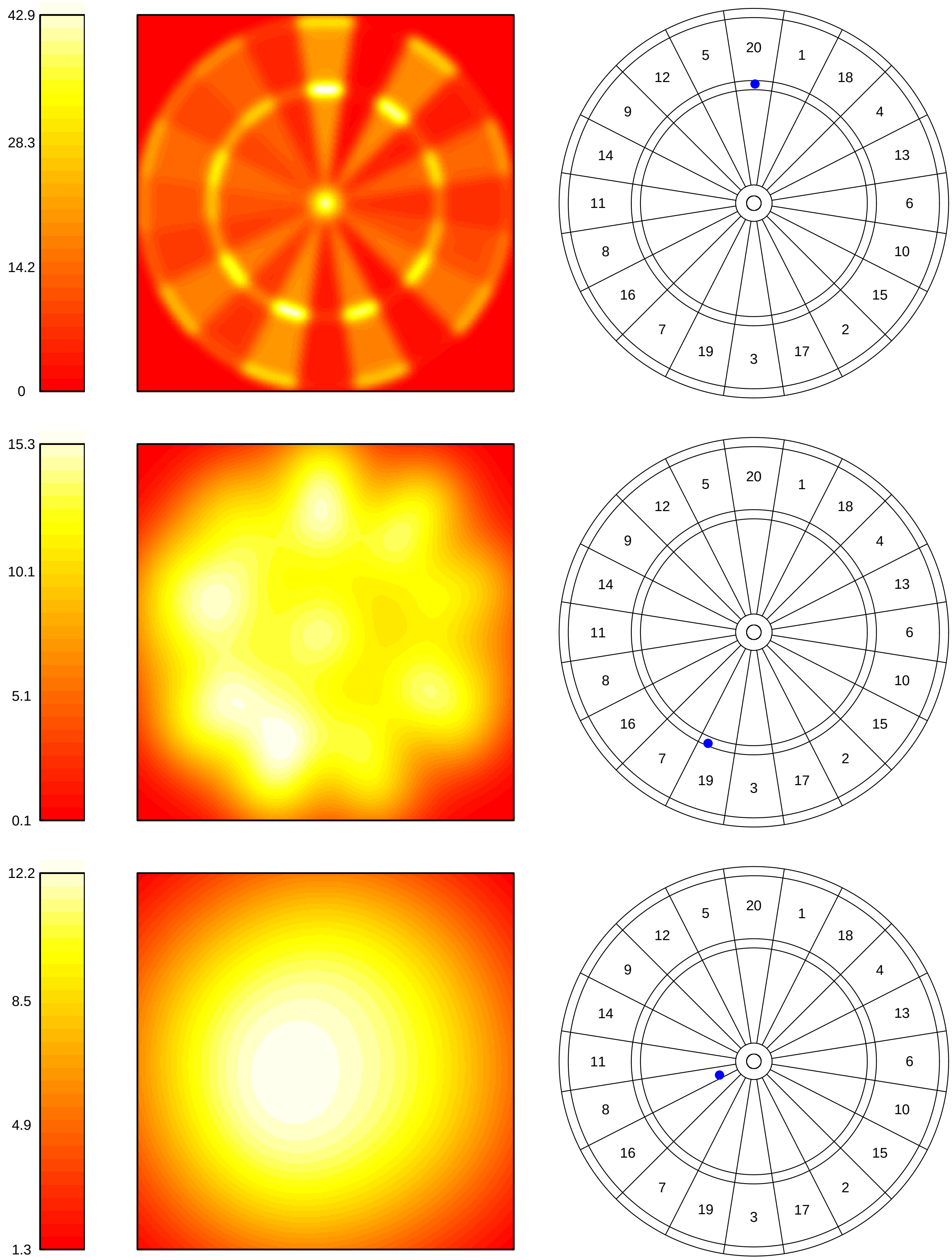

Figure 2: *Heatmaps of* $\mathrm{E}_{\mu,\sigma^2}[s(Z)]$ *for to* $\sigma = 5, 26.9$*, and* 64.6 *(arranged from top to bottom). The color gradient for each plot is scaled to its own range of scores. Adjacent to each heatmap, the optimal aiming location is given by a blue dot on the dartboard.*

We are left with the task of computing the above expectations in the summation. It turns out that these can be computed algebraically, using the symmetry of our Gaussian distribution; for details see A.2.

As for the M-step, note that $C = \sum_{i=1}^{n} \mathrm{E}_{\sigma_0^2}(Z_{i,x}^2 + Z_{i,y}^2 | X_i)$ does not depend on σ^2. Hence we choose σ^2 to maximize $-n \log \sigma^2 - C/2\sigma^2$, which gives $\sigma^2 = C/2n$.

In practice, the EM algorithm gives quite an accurate estimate of σ^2, even when n is only moderately large. Figure 3 considers the case when $n = 50$: for each $\sigma = 1, \ldots 100$, we generated independent $Z_1, \ldots Z_n \sim \mathcal{N}(0, \sigma^2 I)$. We computed the maximum likelihood estimate of σ^2 based on the complete data $(Z_1, \ldots Z_n)$ (shown in blue), which is simply

$$\widehat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{2n} \sum_{i=1}^{n} (Z_{i,x}^2 + Z_{i,y}^2),$$

and compared this with the EM estimate based on the scores $(X_1, \ldots X_n)$ (shown in red). The two estimates are very close for all values of σ.

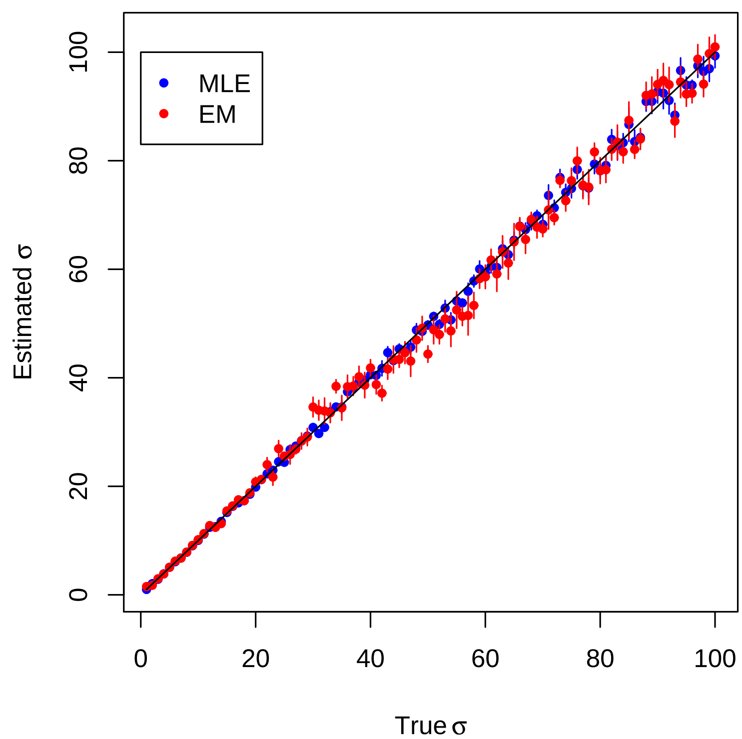

Figure 3: *The MLE and EM estimate, from* $n = 50$ *points drawn independently from* $\mathcal{N}(0, \sigma^2 I)$*, and* σ *ranging over* $1, 2, \ldots 100$. *For each* σ *we actually repeated this 10 times; shown are the mean plus and minus one standard error over these trials.*

Author 1 and author 2 each threw 100 darts at the bullseye and recorded their scores, from which we estimate their standard deviations to be $\sigma_1 = 64.6$ and $\sigma_2 = 26.9$, respectively. Thus Figure 2 shows their personalized heatmaps. To maximize his expected score, author 1 should be aiming at the S8, close to the bullseye. Meanwhile, author 2 (who is a fairly skilled darts player) should be aiming at the T19.

3 A more general Gaussian model

In this section, we consider a more general Gaussian model for throwing errors

$$\varepsilon \sim \mathcal{N}(0, \Sigma),$$

which allows for an arbitrary covariance matrix Σ. This flexibility is important, as a player's distribution of throwing errors may not be circularly symmetric. For example, it is common for most players to have a smaller variance in the horizontal direction than in the vertical one, since the throwing motion is up-and-down with no (intentional) lateral component. Also, a right-handed player may possess a different "tilt" to his or her error distribution (defined by the sign of the correlation) than a left-handed player. In this new setting, we follow the same approach as before: first we estimate model parameters using the EM algorithm, then we compute a heatmap of the expected score function.

3.1 Estimating the covariance

We can estimate Σ using a similar EM strategy as before, having observed the scores $X_1, \ldots X_n$ of throws aimed at the board's center, but not the positions $Z_1, \ldots Z_n$. As $\mu = 0$, the complete data log-likelihood is

$$\ell(\Sigma; X, Z) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^{n} Z_i^T \Sigma^{-1} Z_i,$$

with $X_i = s(Z_i)$ for all i. It is convenient to simplify

$$\sum_{i=1}^{n} Z_i^T \Sigma^{-1} Z_i = \operatorname{tr}\left(\Sigma^{-1} \sum_{i=1}^{n} Z_i Z_i^T\right),$$

using the fact that trace is linear and invariant under commutation. Thus we must compute

$$\mathrm{E}_{\Sigma_0}[\ell(\Sigma; X, Z)|X] = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \operatorname{tr}\left(\Sigma^{-1} \sum_{i=1}^{n} \mathrm{E}_{\Sigma_0}(Z_i Z_i^T | X_i)\right).$$

Maximization over Σ is a problem identical to that of maximum likelihood for a multivariate Gaussian with unknown covariance. Hence the usual maximum likelihood calculations (see [MKB79]) give

$$\Sigma = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{\Sigma_0}(Z_i Z_i^T | X_i).$$

The expectations above can no longer be done in closed form as in the simple Gaussian case. Hence we use importance sampling [Liu08] which is a popular and useful Monte Carlo technique to approximate expectations that may be otherwise difficult to compute. For example, consider the term

$$\mathrm{E}_{\Sigma_0}(Z_{i,x}^2 | X_i) = \iint x^2 \, p(x, y) \, dx \, dy,$$

where p is the density of $Z_i | X_i$ (Gaussian conditional on being in the region of the board defined by the score X_i). In practice, it is hard to draw samples from this distribution, and hence it is hard to estimate the expectation by simple Monte Carlo simulation. The idea of importance sampling is to replace samples from p with samples from some q that is "close" to p but easier to draw from. As long as $p = 0$ whenever $q = 0$, we can write

$$\iint x^2 p(x,y)\,dx\,dy = \iint x^2 w(x,y) q(x,y)\,dx\,dy,$$

where $w = p/q$. Drawing samples $z_1, \ldots z_m$ from q, we estimate the above by

$$\frac{1}{m}\sum_{j=1}^{m} z_{i,x}^2 w(z_{i,x}, z_{i,y}),$$

or, if the density is known only up to some constant

$$\frac{\frac{1}{m}\sum_{j=1}^{m} z_{i,x}^2 w(z_{i,x}, z_{i,y})}{\frac{1}{m}\sum_{j=1}^{m} w(z_{i,x}, z_{i,y})}.$$

There are many choices for q, and the optimal q, measured in terms of the variance of the estimate, is proportional to $x^2 \cdot p(x,y)$ [Liu08]. In our case, we choose q to be the uniform distribution over the region of the board defined by the score X_i, because these distributions are easy to draw from. The weights in this case are easily seen to be just $w(x,y) = f_{\Sigma_0}(x,y)$, the bivariate Gaussian density with covariance Σ_0.

3.2 Computing the heatmap

Having estimated a player's covariance Σ, a personalized heatmap can be constructed just as before. The expected score if the player tries to aim at a location μ is

$$(f_\Sigma * s)(\mu).$$

Again we approximate this by evaluating f_Σ and s over a grid and taking the convolution of these two 2-d arrays, which can be quickly computed using two FFTs and one inverse FFT.

From their same set of $n = 100$ scores (as before), we estimate the covariances for author 1 and author 2 to be

$$\Sigma_1 = \begin{bmatrix} 1820.6 & -471.1 \\ -471.1 & 4702.2 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 320.5 & -154.2 \\ -154.2 & 1530.9 \end{bmatrix},$$

respectively. See Figure 4 for their personalized heatmaps.

The flexibility in this new model leads to some interesting results. For example, consider the case of author 2: from the scores of his 100 throws aimed at the bullseye, recall that we estimate his marginal standard deviation to be $\sigma = 26.9$ according to the simple Gaussian model. The corresponding heatmap instructs him to aim at the T19. However, under the more general Gaussian model, we estimate his x and y standard deviations to be $\sigma_x = 17.9$ and $\sigma_y = 39.1$, and the new heatmap tells him to aim slightly above the T20. This change occurs because the general model can adapt to the fact that author 2 has substantially better accuracy in the x direction. Intuitively, he should be aiming at the 20 since his darts will often remain in this (vertical) pie-slice, and he won't hit the 5 or 1 (horizontal errors) often enough for it to be worthwhile aiming elsewhere.

4 Model extensions and considerations

The Gaussian distribution is a natural model in the EM context because of its simplicity and its ubiquity in statistics. Additionally, there are many studies from cognitive science indicating that in

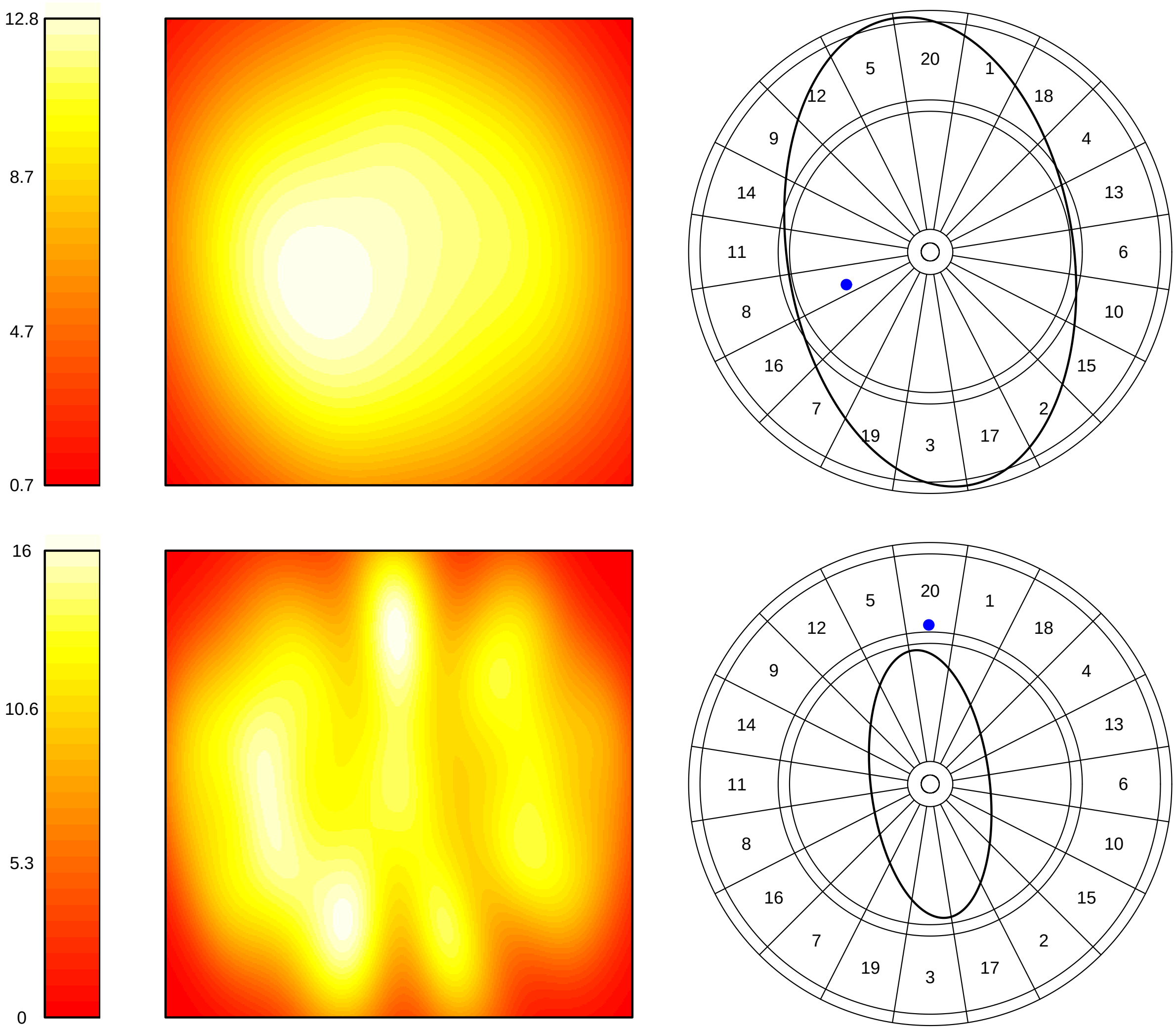

Figure 4: *Author 1's and author 2's covariances Σ_1, Σ_2 were estimated, and shown above are their personalized heatmaps (from top to bottom). Drawn on each dartboard is an ellipse denoting the 70% level set of $\mathcal{N}(0, \Sigma_i)$, and the optimal location is marked with a blue dot.*

motor control, movement errors are indeed Gaussian (see [TGM$^+$05], for example). In the context of dart throwing, however, it may be that the errors in the y direction are skewed downwards. An argument for this comes from an analysis of a player's dart-throwing motion: in the vertical direction, the throwing motion is mostly flat with a sharp drop at the end, and hence more darts could veer towards the floor than head for the ceiling. Below we investigate a distribution that allows for this possibility.

4.1 Skew-Gaussian

In this setting we model the x and y coordinates of ε as independent Gaussian and skew-Gaussian, respectively. We have

$$\varepsilon_x \sim \mathcal{N}(0, \sigma^2), \quad \varepsilon_y \sim \mathcal{SN}(0, \omega^2, \alpha)$$

and so we have three parameters to estimate. With $\mu = 0$, the complete data log-likelihood is

$$\ell(\sigma^2, \omega^2, \alpha; X, Z) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} Z_{i,x}^2 - n \log \omega - \frac{1}{2\omega^2} \sum_{i=1}^{n} Z_{i,y}^2 + \sum_{i=1}^{n} \log \Phi \left(\frac{\alpha Z_{i,y}}{\omega} \right),$$

with $X_i = s(Z_i)$ for all i. Examining the above, we can decouple this into two separate problems: one in estimating σ^2, and the other in estimating ω^2, α. In the first problem we compute

$$C_1 = \sum_{i=1}^{n} \mathrm{E}_{\sigma_0^2}(Z_{i,x}^2 | X_i),$$

which can be done in closed form (see A.3), and then we take the maximizing value $\sigma^2 = C_1/n$. In the second we must consider

$$C_2 = \sum_{i=1}^{n} \mathrm{E}_{\omega_0^2, \alpha_0}(Z_{i,y}^2 | X_i), \quad C_3 = \sum_{i=1}^{n} \mathrm{E}_{\omega_0^2, \alpha_0} \left[\log \Phi \left(\frac{\alpha Z_{i,y}}{\omega} \right) \right].$$

We can compute C_2 by importance sampling, again choosing the proposal density q to be the uniform distribution over the appropriate board region. At first glance, the term C_3 causes a bit of trouble because the parameters over which we need to maximize, ω^2 and α, are entangled in the expectation. However, we can use the highly accurate piecewise-quadratic approximation

$$\log \Phi(x) \approx a + bx + cx^2, \quad (a, b, c) = \begin{cases} (-0.693, 0.727, -0.412) & \text{if } x \le 0 \\ (-0.693, 0.758, -0.232) & \text{if } 0 < x \le 1.5 \\ (-0.306, 0.221, -0.040) & \text{if } 1.5 < x. \end{cases}$$

(See A.4 for derivation details.) Then with

$$K_1 = \sum_{i=1}^{n} \mathrm{E}_{\omega_0^2, \alpha_0}[b(Z_{i,y}) \cdot Z_{i,y} | X_i], \quad K_2 = \sum_{i=1}^{n} \mathrm{E}_{\omega_0^2, \alpha_0}[c(Z_{i,y}) \cdot Z_{i,y}^2 | X_i],$$

computed via importance sampling, maximization over ω^2 and α yields the simple updates

$$\omega^2 = C_2/n, \quad \alpha = -K_1/K_2 \cdot \sqrt{C_2/n}.$$

Notice that these updates would be analogous to the ML solutions, had we again used the piecewise-quadratic approximation for $\log \Phi$.

Once we have the estimates $\sigma^2, \omega^2, \alpha$, the heatmap is again given by the convolution

$$f_{\sigma^2, \omega^2, \alpha} * s,$$

where $f_{\sigma^2, \omega^2, \alpha}$ is the product of the $\mathcal{N}(0, \sigma^2)$ and $\mathcal{SN}(0, \omega^2, \alpha)$ densities. We estimated these parameters for author 1 and author 2, using the scores of their $n = 100$ throws aimed at the board's center. As expected, the skewness parameter α is negative in both cases, meaning that there is a downwards vertical skew. However, the size of the skew is not large enough to produce heatmaps that differ significantly from Figure 4, and hence we omit them here.

4.2 Non-constant variance and non-independence of throws

A player's variance may decrease as the game progresses, since he or she may improve with practice. With this in mind, it is important that a player is sufficiently "warmed up" before he or she throws darts at the bullseye to get an estimate of their model parameters, and hence their personalized heatmap. Moreover, we can offer an argument for the optimal strategy being robust against small changes in accuracy. Consider the simple Gaussian model of Section 2, and recall that a player's accuracy was parametrized by the marginal variance parameter σ^2. Shown in Figure 5 is the optimal aiming location $\mu^*(\sigma) = \text{argmax}_\mu \ \text{E}_{\mu,\sigma^2}[s(Z)]$ as σ varies from 0 to 100, calculated at increments of 0.1. The path appears to be continuous except for a single jump at $\sigma = 16.4$. Aside from being interesting, this is important because it indicates that small changes in σ amount to small changes in the optimal strategy (again, except for σ in an interval around 16.4).

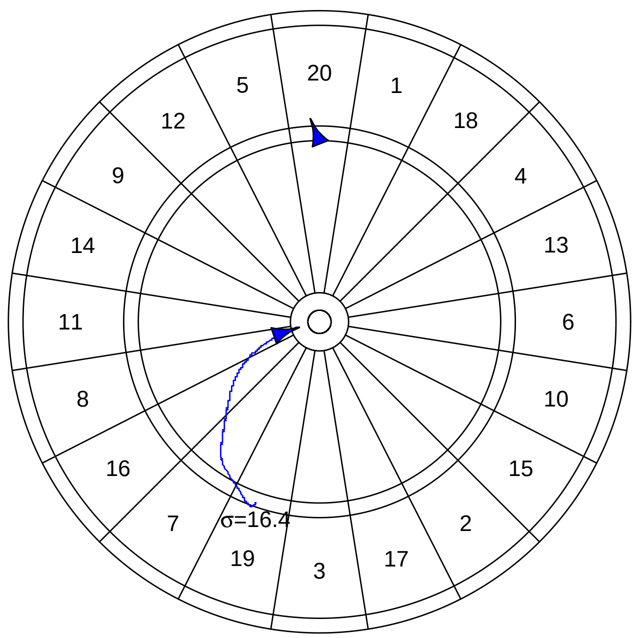

Figure 5: *Path of the optimal location μ^* parametrized by σ. Starting at $\sigma = 0$, the optimal μ is in the center of the* T20*, and moves slightly up and to the left. Then it jumps to the* T19 *at $\sigma = 16.4$. From here it curls into the center of the board, stopping a bit lower than and the left of the bullseye at $\sigma = 100$.*

Furthermore, the assumption that dart throws are independent seems unlikely to be true in reality. Muscle memory plays a large role in any game that requires considerable control of fine motor skills. It can be both frustrating to repeat a misthrow, and joyous to rehit the T20, with a high amount of precision and seemingly little effort on a successive dart throw. Though accounting for this dependence can become very complicated, a simplified model may be worth considering. For instance, we might view the current throw as a mixture of two Gaussians, one centered at the spot where a playing is aiming and the other centered at the spot that this player hit previously. Another example from the time series literature would be an autoregressive model, in which the current throw is Gaussian conditional on the previous throws.

5 Discussion

We have developed a method for obtaining a personalized strategy, under various models for dart throws. This strategy is based on the scores of a player's throws aimed at the bullseye (as opposed to, for example, the positions of these throws) and therefore it is practically feasible for a player to gather the needed data. Finally, the strategy is represented by a heatmap of the expected score as a function of the aiming point.

Recall the simple Gaussian model presented in Section 2: here we were mainly concerned with the optimal aiming location. Consider the optimal (expected) score itself: not surprisingly, the optimal score decreases as the variance σ^2 increases. In fact, this optimal score curve is very steep, and it nearly achieves exponential decline. One might ask whether there was much was much thought put into the design of the current dartboard's arrangement of the numbers $1, \ldots 20$. In researching this question, we found that the person credited with devising this arrangement is Brian Gamlin, a carpenter from Bury, Lancashire, in 1896 [Cha09]. Gamlin boasted that his arrangement penalized drunkards for their inaccuracy, but still it remained unclear how he chose the particular sequence of numbers.

Therefore we decided to develop a quantitative measure for the difficulty of an arbitrary arrangement. Since every arrangement yields a different optimal score curve, we simply chose the integral under this curve (over some finite limits) as our measure of difficulty. Hence a lower value corresponds to a more challenging arrangement, and we sought the arrangement that minimized this criterion. Using the Metropolis-Hastings algorithm [Liu08], we managed to find an arrangement that achieves lower value of this integral than the current board; in fact, its optimal score curve lies below that of the current arrangement for every σ^2.

Interestingly enough, the arrangement we found is simply a mirror image of an arrangement given by [Cur04], which was proposed because it maximizes the sum of absolute differences between adjacent numbers. Though this seems to be inspired by mathematical elegance more than reality, it turned out to be unbeatable by our Metropolis-Hastings search! Supplementary materials (including a longer discussion of our search for challenging arrangements) are available at `http://stat.stanford.edu/~ryantibs/darts/`.

Acknowledgements

We would like to thank Rob Tibshirani for stimulating discussions and his great input. We would also like to thank Patrick Chaplin for his eager help concerning the history of the dartboard's arrangement. Finally we would like to thank the Joint Editor and referees whose comments led to significant improvements in this article.

A Appendix

A.1 Dartboard measurements

Here are the relevant dartboard measurements, taken from the British Darts Organization playing rules [Ald06]. All measurements are in millimeters.

Center to DB wire	6.35
Center to SB wire	15.9
Center to inner triple wire	99
Center to outer triple wire	107
Center to inner double wire	162
Center to outer double wire	170

A.2 Computing conditional expectations for the simple Gaussian EM

Recall that we are in the setting $Z_i \sim \mathcal{N}(0, \sigma_0^2 I)$ and we are to compute the conditional expectation $\mathrm{E}(Z_{i,x}^2 + Z_{i,y}^2 | X_i)$, where X_i denotes the score $X_i = s(Z_i)$. In general, we can describe a score X_i as being achieved by landing in $\cup_j A_j$, where each region A_j can be expressed as $[r_{j,1}, r_{j,2}] \times [\theta_{j,1}, \theta_{j,2}]$ in polar coordinates. For example, the score $X_i = 20$ can be achieved by landing in 3 such regions: the two S20 chunks and the D10. So

$$\begin{aligned}
\mathrm{E}(Z_{i,x}^2 + Z_{i,y}^2 | X_i) &= \mathrm{E}(Z_{i,x}^2 + Z_{i,y}^2 | Z_i \in \cup_j A_j) \\
&= \frac{\sum_j \iint_{A_j} (x^2 + y^2) e^{-(x^2+y^2)/2\sigma_0^2} \, dx \, dy}{\sum_j \iint_{A_j} e^{-(x^2+y^2)/2\sigma_0^2} \, dx \, dy} \\
&= \frac{\sum_j \int_{r_{j,1}}^{r_{j,2}} \int_{\theta_{j,1}}^{\theta_{j,2}} r^3 e^{-r^2/2\sigma_0^2} \, d\theta \, dr}{\sum_j \int_{r_{j,1}}^{r_{j,2}} \int_{\theta_{j,1}}^{\theta_{j,2}} r e^{-r^2/2\sigma_0^2} \, d\theta \, dr},
\end{aligned}$$

where we used a change of variables to polar coordinates in the last step. The integrals over θ will contribute a common factor of

$$\theta_{j,2} - \theta_{j,1} = \begin{cases} 2\pi & \text{if } X_i = 25 \text{ or } 50 \\ \pi/10 & \text{otherwise,} \end{cases}$$

to both the numerator and denominator, and hence this will cancel. The integrals over r can be computed exactly (using integration by parts in the numerator), and therefore we are left with

$$\mathrm{E}(Z_{i,x}^2 + Z_{i,y}^2 | X_i) = \frac{\sum_j [(r_{j,1}^2 + 2\sigma_0^2) e^{-r_{j,1}/2\sigma_0^2} - (r_{j,2}^2 + 2\sigma_0^2) e^{-r_{j,2}/2\sigma_0^2}]}{\sum_j (e^{-r_{j,1}/2\sigma_0^2} - e^{-r_{j,2}/2\sigma_0^2})}.$$

A.3 Computing conditional expectations for the skew-Gaussian EM

Here we have $Z_{i,x} \sim \mathcal{N}(0, \sigma_0^2)$ (recall that it is the y component $Z_{i,y}$ that is skewed), and we need to compute the conditional expectation $\mathrm{E}(Z_{i,x} | X_i)$. Following the same arguments as A.2, we have

$$\mathrm{E}(Z_{i,x}^2 | X_i) = \frac{\sum_j \int_{r_{j,1}}^{r_{j,2}} \int_{\theta_{j,1}}^{\theta_{j,2}} r^3 \cos^2\theta e^{-r^2/2\sigma_0^2} \, d\theta \, dr}{\sum_j \int_{r_{j,1}}^{r_{j,2}} \int_{\theta_{j,1}}^{\theta_{j,2}} r e^{-r^2/2\sigma_0^2} \, d\theta \, dr}.$$

This is only slightly more complicated, since the integrals over θ no longer cancel. We compute

$$\int_{\theta_{j,1}}^{\theta_{j,2}} \cos^2\theta \, d\theta = \triangle\theta_j/2 + [\sin(2\theta_{j,2}) - \sin(2\theta_{j,1})]/4,$$

where $\triangle\theta_j = \theta_{j,2} - \theta_{j,1}$, and the integrals over r are the same as before, giving

$$\mathrm{E}(Z_{i,x}^2 | X_i) = \frac{\sum_j [(r_{j,1}^2 + 2\sigma_0^2) e^{-r_{j,1}/2\sigma_0^2} - (r_{j,2}^2 + 2\sigma_0^2) e^{-r_{j,2}/2\sigma_0^2}] \cdot [2\triangle\theta_j + \sin(2\theta_{j,2}) - \sin(2\theta_{j,1})]}{\sum_j (e^{-r_{j,1}/2\sigma_0^2} - e^{-r_{j,2}/2\sigma_0^2}) \cdot 4\triangle\theta_j}.$$

A.4 Approximation of the logarithm of the standard normal CDF

We take a very simple-minded approach to approximating $\log \Phi(x)$ with a piecewise-quadratic function $a + bx + cx^2$: on each of the intervals $[-3, 0]$, $[0, 1.5]$, $[1.5, 3]$, we obtain the coefficients (a, b, c) using ordinary least squares and a fine grid of points. This gives the coefficient values

$$(a, b, c) = \begin{cases} (-0.693, 0.727, -0.412) & \text{if } x \le 0 \\ (-0.693, 0.758, -0.232) & \text{if } 0 < x \le 1.5 \\ (-0.306, 0.221, -0.040) & \text{if } 1.5 < x. \end{cases}$$

In Figure 6 we plotted $\log \Phi(x)$ for $x \in [-3, 3]$, and on top we plotted the approximation, with the colors coding the regions. The approximation is very accurate over $[-3, 3]$, and a standard normal random variable lies in this interval with probability > 0.999.

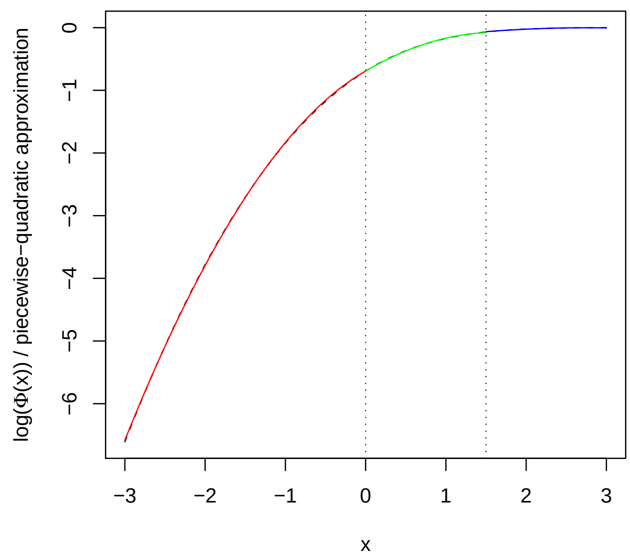

Figure 6: *The function* $\log \Phi(x)$ *is plotted in black, and its piecewise-quadratic approximation is plotted in color.*

References

[Ald06] D. Alderman. BDO playing rules for the sport of darts. `http://www.bdodarts.com/play_rules.htm`, 2006.

[Cha09] P. Chaplin. *Darts in England 1900-39: A social history.* Manchester University Press, Manchester, 2009.

[Cur04] S. A. Curtis. Darts and hoopla board design. *Information Processing Letters*, 92(1):53–56, 2004.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–38, 1977.

[Koh82] D. Kohler. Optimal strategies for the game of darts. *Journal of the Operational Research Society*, 33(10):871–884, 1982.

[Kor07] K. Kording. Decision theory: what "should" the nervous system do? *Science*, 318(5850):606–610, 2007.

[Liu08] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York, 2008.

[MKB79] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.

[Per99] D. Percy. Winning darts. *Mathematics Today*, 35(2):54–57, 1999.

[R D08] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[Ste97] H.S. Stern. Shooting darts. *In the column: A statistician reads the sports pages, Chance*, 10(3):16–19, 1997.

[TGM+05] J. Trommershauser, S. Gepshtein, L. T. Maloney, M. S. Landy, and M. S. Banks. Optimal compensation for changes in task-relevant movement variability. *The Journal of Neuroscience*, 25(31):7169–7178, 2005.