



## A statistician plays darts

Ryan J. Tibshirani, Andrew Price and Jonathan Taylor

*Stanford University, USA*

[Received June 2009. Revised April 2010]

**Summary.** 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 level of skill. By modelling a dart throw as a two-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 heat map, where the bright regions correspond to the aiming locations with high (expected) pay-offs. 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 methods; Statistics of games

### 1. Introduction

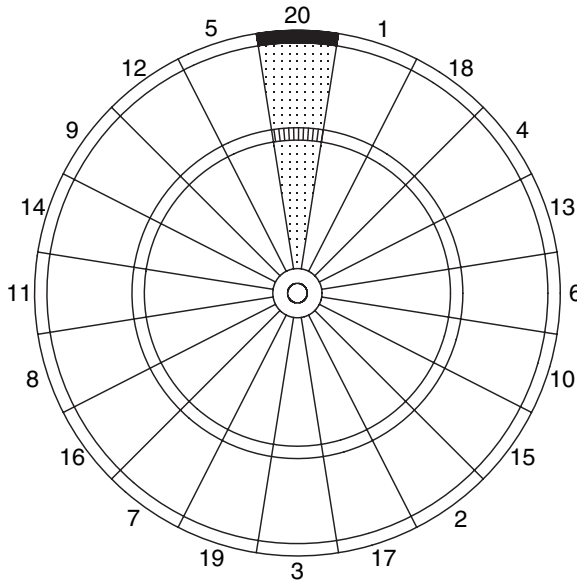
Familiar to most, the game of darts is played by throwing small metal missiles (darts) at a circular target (dartboard). Fig. 1 shows a standard dartboard. A player receives a different score for landing a dart in different sections of the board. In most common dart games, the board's small concentric circle, called the 'double bulls-eye' (DB) or just 'bulls-eye', is worth 50 points. The surrounding ring, called the 'single bulls-eye' (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 'doubles' ring and a 'triples' 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 a Gaussian model), there are some surprising places that 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 DB. The algorithm is a straightforward implementation of the EM algorithm (Dempster *et al.*, 1977), and the simple model that 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 (Liu, 2008) to approximate the expectations in the E-steps. The M-steps, however, 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: Stern (1997) compared aiming at the triple 19 and triple 20 for players with an advanced level of accuracy, and

*Address for correspondence:* Ryan J. Tibshirani, Department of Statistics, Stanford University, 390 Serra Mall, Stanford, CA 94306, USA.  
E-mail: ryantibs@stanford.edu



**Fig. 1.** Standard dartboard: , single 20 region, worth 20 points; , double 20 region, worth 40 points; , triple 20 region, worth 60 points

Percy (1999) considered aiming at the high scoring triples and bulls-eyes for players at an amateur level. In a study on decision theory, Kording (2007) displayed a heat map where the colours reflect the expected score as a function of the aiming point on the dartboard. In this paper we also compute heat maps of the expected score function but, in addition, we propose a method to estimate a player’s level of skill by using the EM algorithm. Therefore any player can obtain a personalized heat map, so long as he or she is willing to aim 50 or so throws at the bulls-eye.

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 Kohler (1982) for an example that takes such matters into consideration. In contrast, 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 and pitching in baseball).

Software for our algorithms is available as an R package (R Development Core Team, 2008) 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 two-dimensional position of a dart throw, and let  $s(Z)$  denote the score. Then the expected score is

$$E[s(Z)] = 50 P(Z \in DB) + 25 P(Z \in SB) + \sum_{i=1}^{20} \{i P(Z \in Si) + 2i P(Z \in Di) + 3i P(Z \in Ti)\}$$

where  $Si$ ,  $Di$  and  $Ti$  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  $B$ , i.e. for any region  $S$

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

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

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

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, a beginner will occasionally miss the board entirely, which corresponds to a score of 0. But, more importantly, most beginners aim at the 20 region; since this is adjacent to the 5 and 1 regions, 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 centre of the board, and consider the model

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

where  $I$  is the  $2 \times 2$  identity matrix. The point  $\mu = (\mu_x, \mu_y)$  represents the location at which the player is aiming, and  $\sigma^2$  controls the size of the error  $\varepsilon$ . (Smaller  $\sigma^2$  means a more accurate player.) Given this set-up, our question becomes: what choice of  $\mu$  produces the largest value of  $E_{\mu, \sigma^2}[s(Z)]$ ?

### 2.1. Choosing where to aim

For a given  $\sigma^2$ , consider choosing  $\mu$  to maximize

$$E_{\mu, \sigma^2}[s(Z)] = \iint \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{\|(x, y) - \mu\|^2}{2\sigma^2}\right\} s(x, y) dx dy. \quad (1)$$

Although this is too difficult to approach analytically, we note that this 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  $f_{\sigma^2}$  with the score  $s$ . In fact, by the convolution theorem

$$f_{\sigma^2} * s = \mathcal{F}^{-1}\{\mathcal{F}(f_{\sigma^2})\mathcal{F}(s)\}$$

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

Once we have computed this convolution, we have the expected score (1) evaluated at every  $\mu$  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 (Stern, 1997; Percy, 1999), with the authors using instead naive Monte Carlo methods to approximate the above expectations. This convolution

approach is especially useful for creating a heat map of the expected score, which would be infeasible to compute using Monte Carlo methods.

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

The bright colours (from yellow to white) correspond to the high expected scores. It is important to note that the heat maps change considerably as we vary  $\sigma$ . For  $\sigma = 0$  (perfect accuracy), the optimal  $\mu$  lies in the triple 20 area, which is the highest scoring region of the board. When  $\sigma = 5$ , the best place to aim is still (the centre of) the triple 20 area. But, for  $\sigma = 26.9$ , it turns out that the best place to aim is in the triple 19 region, close to the border that 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 centre, and the optimal spot is to the lower left of the bulls-eye.

### 2.2. Estimating the accuracy of a player

Since the optimal location  $\mu^*(\sigma)$  depends strongly on  $\sigma$ , we consider a method for estimating a player's  $\sigma^2$  so that he or she can implement the optimal strategy. Suppose that a player throws  $n$  independent darts, aiming each time at the centre of the board. If we knew the board positions  $Z_1, \dots, Z_n$ , the standard sample variance calculation would provide an estimate of  $\sigma^2$ . 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 just to record the score of each dart throw.

In what follows, we use just the scores to arrive at an estimate of  $\sigma^2$ . This may seem surprising at first, because there seems relatively little information to estimate  $\sigma^2$  just knowing the score, which for most numbers (e.g. 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 bulls-eyes, 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 (Dempster *et al.*, 1977). This algorithm, which is 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 centre  $\mu = 0$ , and the unobserved data are the positions  $Z = (Z_1, \dots, Z_n)$  where the darts actually landed.

Let  $l(\sigma^2; X, Z)$  denote the complete-data log-likelihood. The EM algorithm (in this case estimating only one parameter,  $\sigma^2$ ) begins with an initial estimate  $\sigma_0^2$ , and then repeats the following two steps until convergence:

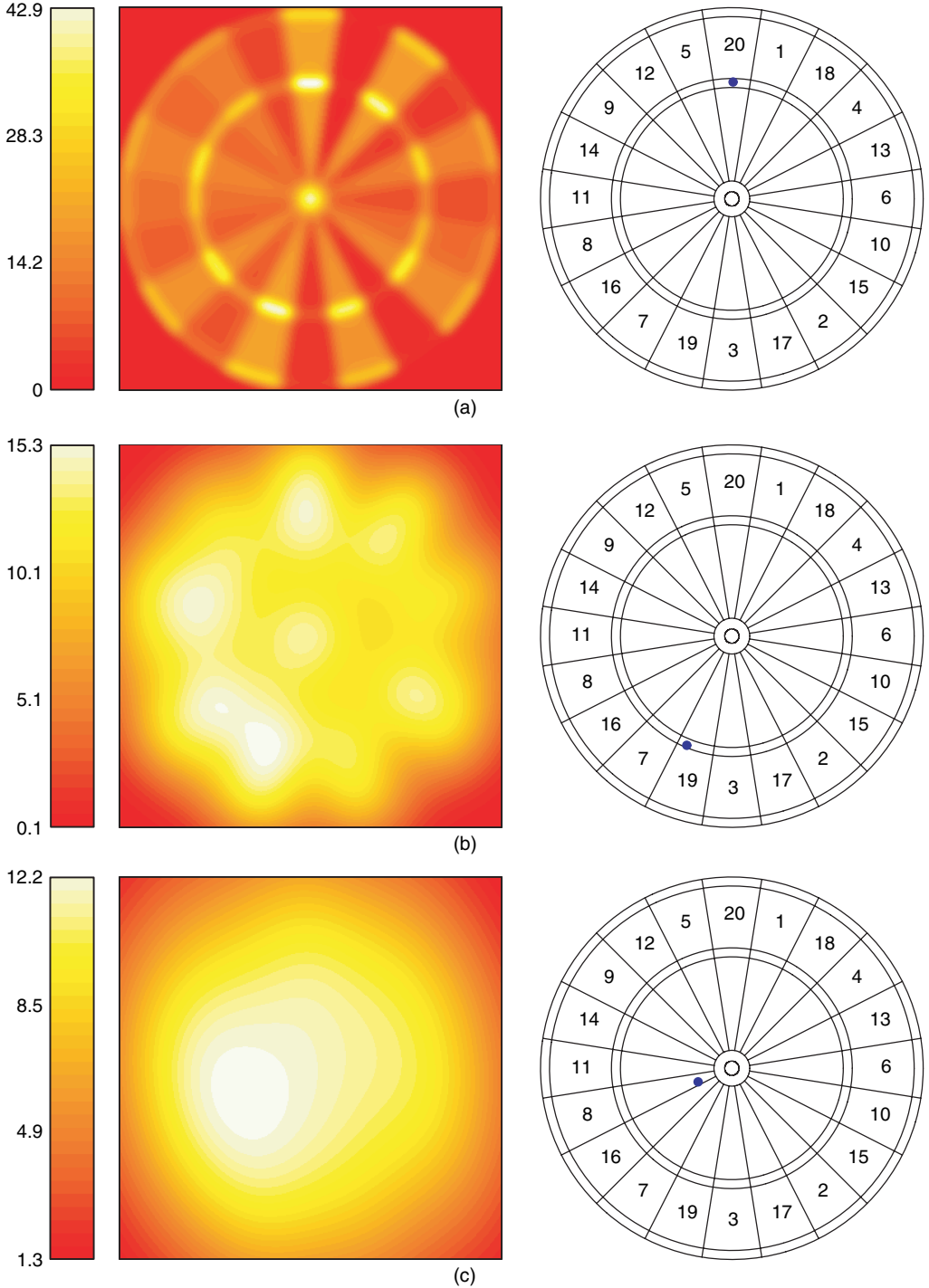
- (a) *E-step*—compute  $Q(\sigma^2) = E_{\sigma_0^2}[l(\sigma^2; X, Z)|X]$ ;
- (b) *M-step*—let  $\sigma_{t+1}^2 = \arg \max_{\sigma^2} \{Q(\sigma^2)\}$ .

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

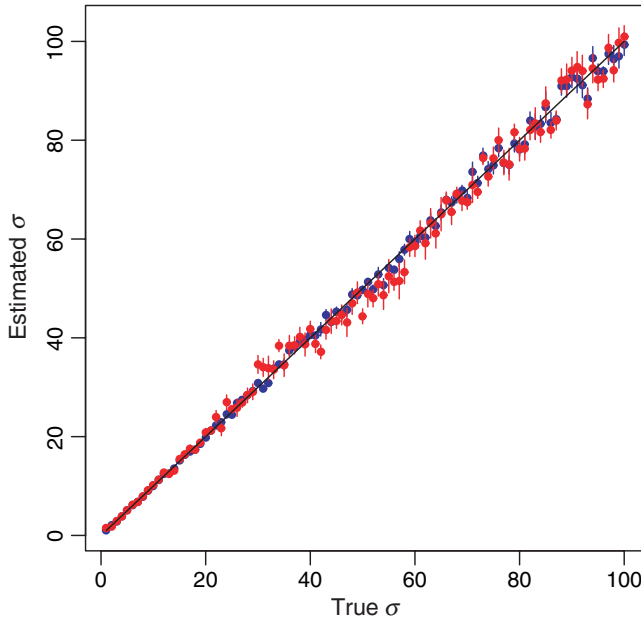
$$l(\sigma^2; X, Z) = \begin{cases} -n \log(\sigma^2) - (1/2\sigma^2) \sum_{i=1}^n (Z_{i,x}^2 + Z_{i,y}^2) & \text{if } X_i = s(Z_i) \forall i, \\ -\infty & \text{otherwise.} \end{cases}$$

Therefore the expectation in the E-step is

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



**Fig. 2.** Heat maps of  $E_{\mu, \sigma^2}[s(Z)]$  for (a)  $\sigma = 5$ , (b)  $\sigma = 26.9$  and (c)  $\sigma = 64.6$ : the colour gradient for each plot is scaled to its own range of scores; ●, optimal aiming location



**Fig. 3.** Maximum likelihood (●) and EM estimate (●), from  $n = 50$  points drawn independently from  $\mathcal{N}(0, \sigma^2 I)$ , and  $\sigma$  ranging over  $1, 2, \dots, 100$ : for each  $\sigma$  we repeated this 10 times; shown are the mean  $\pm 1$  standard error over these trials

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 Appendix A.2.

As for the M-step, note that  $C = \sum_{i=1}^n E_{\sigma^2} [Z_{i,x}^2 + Z_{i,y}^2 | X_i]$  does not depend on  $\sigma^2$ . Hence we choose  $\sigma^2$  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  $\sigma^2$ , even when  $n$  is only moderately large. Fig. 3 considers the case when  $n = 50$ : for each  $\sigma = 1, \dots, 100$ , we generated independent  $Z_1, \dots, Z_n \sim \mathcal{N}(0, \sigma^2 I)$ . We computed the maximum likelihood estimate of  $\sigma^2$  based on the complete data  $(Z_1, \dots, Z_n)$  (shown in blue), which is simply

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

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

Author 1 and author 2 each threw 100 darts at the bulls-eye and recorded their scores, from which we estimate their standard deviations to be  $\sigma_1 = 64.6$  and  $\sigma_2 = 26.9$  respectively. Thus Fig. 2 shows their personalized heat maps. To maximize his expected score, author 1 should be aiming at the single 8 region, close to the bulls-eye. Meanwhile, author 2 (who is a fairly skilled darts player) should be aiming at the triple 19 region.

### 3. 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  $\Sigma$ . 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 direction, since the throwing motion is up and down with no (intentional) lateral component. Also, a right-handed player may have 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 by using the EM algorithm; then we compute a heat map of the expected score function.

### 3.1. Estimating the covariance

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

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

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

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

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

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

Maximization over  $\Sigma$  is a problem which is identical to that of maximum likelihood for a multivariate Gaussian distribution with unknown covariance. Hence the usual maximum likelihood calculations (see Mardia *et al.* (1979)) give

$$\Sigma = \frac{1}{n} \sum_{i=1}^n E_{\Sigma_0}[Z_i Z_i^T | X_i].$$

The expectations above can no longer be done in closed form as in the simple Gaussian case. Hence we use importance sampling (Liu, 2008), which is a popular and useful Monte Carlo technique to approximate expectations that may be otherwise difficult to compute. For example, consider the term

$$E_{\Sigma_0}[Z_{i,x}^2 | X_i] = \iint x^2 p(x, y) dx dy,$$

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 difficult to draw samples from this distribution, and hence it is difficult 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  $z_1, \dots, z_m$  from  $q$ , we estimate this quantity 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  $x^2 p(x, y)$  (Liu, 2008). In our case, we choose  $q$  to be the uniform distribution over the region of the board that is 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  $w(x, y) = f_{\Sigma_0}(x, y)$ , which is the bivariate Gaussian density with covariance  $\Sigma_0$ .

### 3.2. Computing the heat map

Having estimated a player’s covariance  $\Sigma$ , a personalized heat map can be constructed just as before. The expected score if the player tries to aim at a location  $\mu$  is

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

Again we approximate this by evaluating  $f_{\Sigma}$  and  $s$  over a grid and taking the convolution of these two two-dimensional arrays, which can be quickly computed by using two fast Fourier transforms and one inverse fast Fourier transform.

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

$$\Sigma_1 = \begin{pmatrix} 1820.6 & -471.1 \\ -471.1 & 4702.2 \end{pmatrix},$$

$$\Sigma_2 = \begin{pmatrix} 320.5 & -154.2 \\ -154.2 & 1530.9 \end{pmatrix}$$

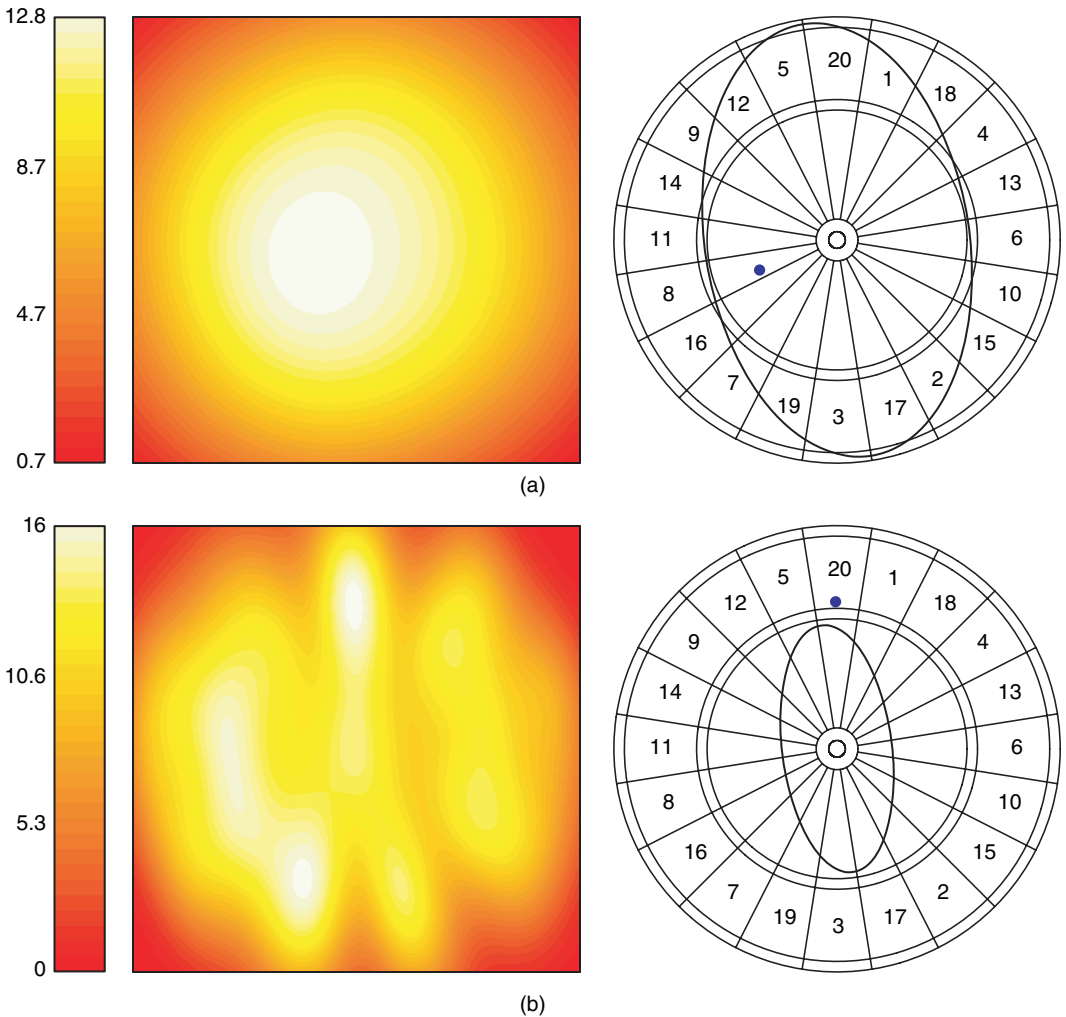
respectively. See Fig. 4 for their personalized heat maps.

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 bulls-eye, recall that we estimate his marginal standard deviation to be  $\sigma = 26.9$  according to the simple Gaussian model. The corresponding heat map instructs him to aim at the triple 19 region. 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 heat map tells him to aim slightly above the triple 20 region. 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 area since his darts will often remain in this (vertical) pie slice, and he will not hit the 5 or 1 areas (horizontal errors) sufficiently often 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 motor control, movement errors are indeed Gaussian (see Trommershauser *et al.* (2005), 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





**Fig. 4.** (a) Author 1's and (b) author 2's covariances  $\Sigma_i$  and their personalized heat maps: drawn on each dartboard is an ellipse denoting the 70% level set of  $\mathcal{N}(0, \Sigma_i)$  and the optimal location (●)

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 distribution

In this setting we model the  $x$ - and  $y$ -co-ordinates of  $\varepsilon$  as independent Gaussian and skew Gaussian respectively. We have

$$\begin{aligned} \varepsilon_x &\sim \mathcal{N}(0, \sigma^2), \\ \varepsilon_y &\sim \mathcal{SN}(0, \omega^2, \alpha) \end{aligned}$$

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

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

with  $X_i = s(Z_i)$  for all  $i$ . Examining this equation, we can decouple it into two separate problems: one in estimating  $\sigma^2$ , and the other in estimating  $\omega^2$  and  $\alpha$ . In the first problem we compute

$$C_1 = \sum_{i=1}^n E_{\sigma_0^2} [Z_{i,x}^2 | X_i]$$

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

$$C_2 = \sum_{i=1}^n E_{\omega_0^2, \alpha_0} [Z_{i,y}^2 | X_i],$$

$$C_3 = \sum_{i=1}^n E_{\omega_0^2, \alpha_0} \left[ \log \left\{ \Phi \left( \frac{\alpha Z_{i,y}}{\omega} \right) \right\} \right].$$

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 little trouble because the parameters over which we need to maximize,  $\omega^2$  and  $\alpha$ , are entangled in the expectation. However, we can use the highly accurate piecewise quadratic approximation

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

(see Appendix A.4 for details of the derivation). Then with

$$K_1 = \sum_{i=1}^n E_{\omega_0^2, \alpha_0} [b(Z_{i,y}) Z_{i,y} | X_i],$$

$$K_2 = \sum_{i=1}^n E_{\omega_0^2, \alpha_0} [c(Z_{i,y}) Z_{i,y}^2 | X_i]$$

computed via importance sampling, maximization over  $\omega^2$  and  $\alpha$  yields the simple updates

$$\omega^2 = C_2/n,$$

$$\alpha = -(K_1/K_2) \sqrt{C_2/n}.$$

Note that these updates would be analogous to the maximum likelihood solutions, if we had again used the piecewise quadratic approximation for  $\log(\Phi)$ .

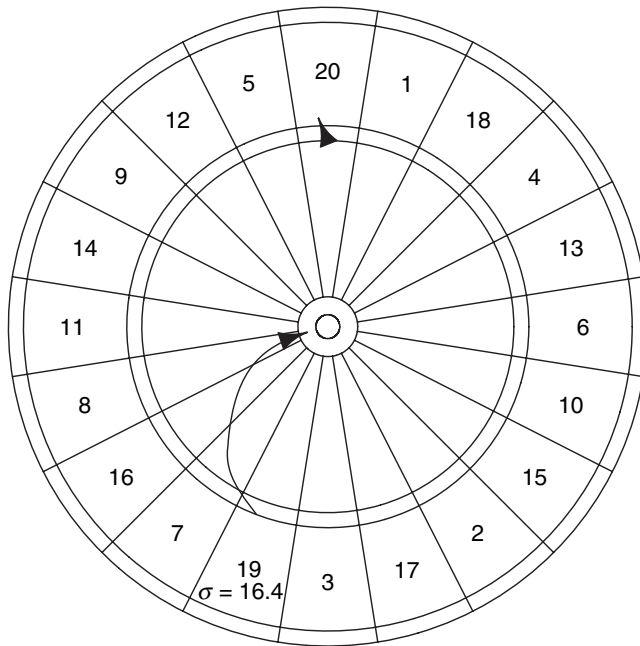
Once we have the estimates  $\sigma^2$ ,  $\omega^2$  and  $\alpha$ , the heat map is again given by the convolution

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

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 centre. As expected, the skewness parameter  $\alpha$  is negative in both cases, meaning that there is a downward vertical skew. However, the size of the skew is not sufficiently large to produce heat maps that differ significantly from Fig. 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



**Fig. 5.** Path of the optimal location  $\mu^*$  parameterized by  $\sigma$ : starting at  $\sigma = 0$ , the optimal  $\mu$  is in the centre of the triple 20 region and moves slightly up and to the left; then it jumps to the triple 19 region at  $\sigma = 16.4$ ; from here it curls into the centre of the board, stopping a little lower than and to the left of the bulls-eye at  $\sigma = 100$

or she throws darts at the bulls-eye to obtain an estimate of their model parameters, and hence their personalized heat map. 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 parameterized by the marginal variance parameter  $\sigma^2$ . Shown in Fig. 5 is the optimal aiming location  $\mu^*(\sigma) = \arg \max_{\mu} \{E_{\mu, \sigma^2}[s(Z)]\}$  as  $\sigma$  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  $\sigma$  amount to small changes in the optimal strategy (again, except for  $\sigma$  in an interval around 16.4).

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 triple 20 region, 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 Gaussian distributions: one centred at the spot where a player is aiming and the other centred 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 bulls-eye (as opposed to, for example, the positions of these throws) and therefore it is practically feasible

for a player to gather the data needed. Finally, the strategy is represented by a heat map of the expected score as a function of the aiming point.

Recall the simple Gaussian model that was 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  $\sigma^2$  increases. In fact, this optimal score curve is very steep, and it nearly achieves exponential decline. We might ask whether much thought was put into the design of the current dartboard's arrangement of the numbers  $1, \dots, 20$ . In researching this question, we found that the person who is credited with devising this arrangement is Brian Gamlin, a carpenter from Bury, Lancashire, in 1896 (Chaplin, 2009). 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 (Liu, 2008) 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  $\sigma^2$ .

Interestingly, the arrangement that we found is simply a mirror image of an arrangement that was given by Curtis (2004), 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 thank Rob Tibshirani for stimulating discussions and his great input. We also thank Patrick Chaplin for his eager help concerning the history of the dartboard's arrangement. Finally we thank the Joint Editor and referees, whose comments led to significant improvements in this paper.

## Appendix A

### A.1. Dartboard measurements

Table 1 gives the relevant dartboard measurements, taken from the British Darts Organization playing rules (Alderman, 2006).

**Table 1.** Dartboard measurements

<i>Distance</i>	<i>Measurement (mm)</i>
Centre to DB wire	6.35
Centre to SB wire	15.9
Centre to inner triple wire	99
Centre to outer triple wire	107
Centre to inner double wire	162
Centre to outer double wire	170

**A.2. Computing conditional expectations for the simple Gaussian EM algorithm**

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  $E[Z_{i,x}^2 + Z_{i,y}^2 | X_i]$ , 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 co-ordinates. For example, the score  $X_i = 20$  can be achieved by landing in three such regions: the two single 20 chunks and the double 10 region. So

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

where we used a change of variables to polar co-ordinates in the last step. The integrals over  $\theta$  will contribute a common factor of

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

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

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

**A.3. Computing conditional expectations for the skew Gaussian EM algorithm**

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  $E[Z_{i,x} | X_i]$ . Following the same arguments as in Appendix A.2, we have

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

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

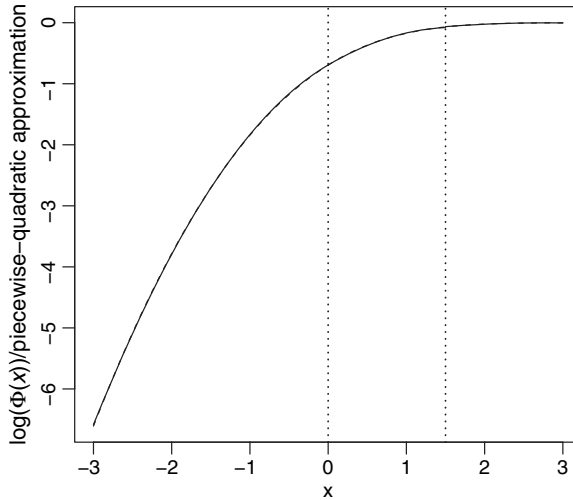
$$\int_{\theta_{j,1}}^{\theta_{j,2}} \cos^2(\theta) d\theta = \Delta\theta_j/2 + \{\sin(2\theta_{j,2}) - \sin(2\theta_{j,1})\}/4$$

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

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

**A.4. Approximation of the logarithm of the standard normal cumulative distribution function**

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  $[-3, 0]$ ,  $[0, 1.5]$  and  $[1.5, 3]$ , we obtain the coefficients  $(a, b, c)$  by using ordinary least squares and a fine grid of points. This gives the coefficient values



**Fig. 6.** Function  $\log\{\Phi(x)\}$  is plotted as a broken curve, and its piecewise quadratic approximation is plotted as a full curve (they are indistinguishable)

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

In Fig. 6 we plotted  $\log\{\Phi(x)\}$  for  $x \in [-3, 3]$ , as a broken curve, and on top we plotted the approximation, as a full curve. The approximation is very accurate over  $[-3, 3]$ , and a standard normal random variable lies in this interval with probability greater than 0.999.

**References**

Alderman, D. (2006) BDO playing rules for the sport of darts. British Darts Organisation, London. (Available from [http://www.bdodarts.com/play\\_rules.htm](http://www.bdodarts.com/play_rules.htm).)  
 Chaplin, P. (2009) *Darts in England 1900-39: a Social History*. Manchester: Manchester University Press.  
 Curtis, S. A. (2004) Darts and hoopla board design. *Inf. Process. Lett.*, **92**, 53–56.  
 Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc. B*, **39**, 1–38.  
 Kohler, D. (1982) Optimal strategies for the game of darts. *J. Oper. Res. Soc.*, **33**, 871–884.  
 Kording, K. (2007) Decision theory: what “should” the nervous system do? *Science*, **318**, 606–610.  
 Liu, J. S. (2008) *Monte Carlo Strategies in Scientific Computing*. New York: Springer.  
 Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) *Multivariate Analysis*. London: Academic Press.  
 Percy, D. (1999) Winning darts. *Math. Today*, **35**, 54–57.  
 R Development Core Team (2008) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.  
 Stern, H. S. (1997) Shooting darts. *Chance*, **10**, 16–19.  
 Trommershauser, J., Gepshtein, S., Maloney, L. T., Landy, M. S. and Banks, M. S. (2005) Optimal compensation for changes in task-relevant movement variability. *J. Neurosci.*, **25**, 7169–7178.