

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**Modelli Matematici per il Gioco dei Darts e Analisi
Statistica dello Score**

Relatore: Marco Ferrante
Dipartimento di Matematica Pura ed Applicata

Laureando: Marco D'Agostini
Matricola: 1058935

Anno Accademico 2015/2016

A mio padre

*" ...senza trattare niente con chi,
ha già fissato il prezzo al mercato,
dei nostri sogni e dentro ai nostri giorni
e per la nostra vita..."*

Liberi Tutti - Subsonica

Indice

Introduzione	1
1 Introduzione al Gioco e Cenni Storici	3
1.1 Le Origini	3
1.2 Cenni sul Regolamento	5
1.3 Modalità di Gioco e Varianti	6
1.3.1 Round the clock(l'orologio)	7
1.3.2 Criquet	8
1.3.3 Killer	8
1.3.4 Split	8
1.4 Tornei e Giocatori	9
1.4.1 PDC e BDO	9
1.4.2 Il PDC World Darts Championship	9
2 Le Freccette da un Punto di vista Dinamico	13
2.1 Il moto parabolico	15
2.2 La Distribuzione di Probabilità di (X,Y)	21
3 Varianza di Tiro con l'Algoritmo EM	25
3.1 Don't try for the triple 20	25
3.2 L'Algoritmo EM	30
3.2.1 Pregi e Difetti	32
3.2.2 Estensioni dell'EM	33
3.3 Stima della varianza di tiro di un giocatore	37
3.4 Un modello Gaussiano più generale	44
3.4.1 Stima della Covarianza	44
3.4.2 Calcolo della Heatmap	45
4 Simulazioni e Analisi Empirica del modello dinamico	49
4.1 La Velocità r	50
4.2 Gli Angoli φ e θ	53
4.3 DM Heatmap	57

4.4	Confronto tra EM Heatmap e DM Heatmap	58
4.4.1	Heatmaps di un giocatore di livello medio	59
4.4.2	Le Heatmaps di Phil Taylor	60
5	Conclusioni	63
A	La Trasformata di Fourier	65
B	Codici Utilizzati	67
B.1	Codici R	67
B.2	Codici Matlab	69
	Bibliografia	79

Elenco delle figure

1.1	Dartboard classica utilizzata nei tornei professionistici	4
1.2	Misure e distanze dalla dartboard utilizzate da regolamento	5
1.3	Valore dei doppi e tripli sulla Dartboard	7
2.1	Posizione di lancio classica nel gioco delle freccette.	13
2.2	Rappresentazione grafica del moto parabolico.	15
2.3	Rappresentazione di una freccetta sul piano	17
2.4	Raffigurazione geometrica del lancio di un dart verso bersaglio.	18
2.5	Punti di massimo che deve raggiungere la velocità dato φ , con $\theta=0$	19
2.6	Distribuzione grafica di $Y X=x$	22
2.7	Distribuzione congiunta (x,y)	23
3.1	Distanze (raggi) in mm dal centro	27
3.2	Simple EM Heatmap per un giocatore di livello medio	40
3.3	General EM Heatmap di un giocatore di livello medio	47
3.4	General EM Heatmap di Phil Taylor	48
4.1	Distribuzioni (X,Y) marginali e congiunte	49
4.2	Distribuzione del lanci sul bersaglio, variando σ_r	51
4.3	Distribuzione dei darts, con σ_r e σ_φ fissati e variando σ_θ	54
4.4	Distribuzione dei darts, fissato σ_r e variando i σ_θ e σ_φ	55
4.5	DM Heatmaps al variare di σ_φ e σ_θ	57
4.6	DM e EM Heatmaps di un giocatore di livello medio	59
4.7	Confronto tra la EM e DM Heatmap di Phil Taylor	61

Introduzione

La leggenda vuole che, durante la Guerra dei Cent'anni, in un giorno di brutto tempo, gli arcieri inglesi, al riparo in un granaio, si divertissero a lanciare delle frecce su un ceppo di legno. Fecero in fretta ad accorciare i loro proiettili fino ad ottenere dei «dardi», simili alle freccette attuali.

Lo sport dei Darts nato agli inizi del '900, ha avuto momenti di alti e bassi, con un boom negli anni '80, stoppato alla fine del decennio a causa dell'uso di alcool dei giocatori durante i match. Nel 2007, la World Dart Federation ha inoltrato la candidatura a Londra 2012, per ottenere le freccette come sport dimostrativo, ma la richiesta è stata rifiutata dagli organizzatori. Purtroppo il risultato è tutt'altro che inatteso. Le freccette, infatti, devono fare i conti con la loro cattiva reputazione, praticate per lo più nell'atmosfera fumosa e alcolica dei pub, hanno avuto molte difficoltà ad affrancarsi dall'immagine di passatempo per uomini ubriachi.

Questa disciplina però non è solo birra e pubs, ma unisce la precisione dell'attività fisica e allo stesso tempo promuove il calcolo mentale. È un modo semplice e piacevole per migliorare le proprie capacità matematiche. Il campione olandese, Raymond Van Barneveld dice di allenarsi diverse ore al giorno, contando e ricontando per trovare il modo migliore di raggiungere l'obiettivo. Anche se per favorire la concentrazione ogni tanto non disdegna una birra.

In questo lavoro di tesi abbiamo trattato il gioco dei Darts dal punto di vista matematico con uno studio focalizzato sulla massimizzazione dello score di tiro. Abbiamo sviluppato dei modelli matematico-fisico-dinamici, sul lancio della freccia verso il bersaglio. L'analisi fisica del contesto è realizzata modellando il moto parabolico, ampliandolo con l'uso delle coordinate sferiche. Abbiamo studiato così la distribuzione dei lanci sul bersaglio dati angoli e velocità di tiro e impostando le variabili con una distribuzione normale. Tramite degli script di Matlab abbiamo implementato i nostri modelli, ricavandoci degli output grafici sulle distribuzioni e sulle Heatmap, utili per osservare le zone ottime di tiro di ogni tiratore.

In aggiunta, tramite un paper dell'Università di Stanford, abbiamo esaminato la massimizzazione del punteggio in merito alla precisione di tiro. In questo caso, l'analisi è stata effettuata con l'uso dell'Expectation Maximization-Algorithm, approfondendo le sue aree di utilizzo e alcune sue proprietà. Con l'Algoritmo EM abbiamo sviluppato due diversi modelli, uno più semplice e uno po' più esteso. Tramite una library di R legata all'algoritmo, ci siamo ricavati diverse Heatmaps, commentandole graficamente e comparandole con quelle del nostro modello fisico. L'obbiettivo dell'elaborato, sarà quello di cercare di modellare la nostra distribuzione (X,Y) , come si spalma sul bersaglio e successivamente provare a trovare dei raffronti con il modello costruito con l'EM Algorithm. Osservando le diverse caratteristiche dei modelli, facendo un'analisi empirica e dei confronti grafici tra di essi.

Lo studio si sviluppa in cinque Capitoli, nel Capitolo 1 verranno citati riferimenti storici, regolamento e varie curiosità del gioco delle freccette. Nel Capitolo 2 ci sarà l'analisi del modello dinamico sui darts. Il Capitolo 3 sarà riferito al Modello EM, partendo dalle sue proprietà e campi in cui è utilizzato e arrivando allo studio di 2 modelli con i relativi output grafici (mappe di calore). Il quarto capitolo sarà un'estensione del secondo, verrà realizzata un'analisi empirica sui grafici del modello fisico, ricavati con Matlab e ci focalizzeremo sulle Heatmaps, poi confrontate con quelle del Capitolo 3. L'ultimo capitolo riguarderà le considerazioni conclusive.

Capitolo 1

Introduzione al Gioco e Cenni Storici

1.1 Le Origini

E' difficile dare un esatta collocazione temporale al gioco delle freccette, pare che la parola darts si trovi già nei dizionari dell'"Oxford English Dictionary" del 1314, anche se il significato veniva attribuito per altre cose. Leggende narrano che i padri pellegrini praticavano una specie di tiro al bersaglio, usando la base di una botte di legno e delle piccole frecce, per passare il tempo durante la traversata del 1620 sul "Mayflowers". Solo alla fine del secolo scorso, nel 1896 appare il bersaglio numerato, opera di un falegname di Bury di nome Brian Gamlin.

L'anno chiave per il gioco dei darts è sicuramente il 1908, per la sentenza chiamata "Foot Anakin". Pare che in quegli anni questo sport, non ancora riconosciuto, fosse ritenuto un gioco d'azzardo e quindi vietato nei Pubs. Il fatto avvenne in un tribunale di Leeds, nel processo verso il Signor Anakin, accusato di gioco d'azzardo nel suo Pub. Foot preparò una memoria difensiva a dir poco inusuale, il giorno del processo si presentò in aula con un set di freccette ed un "Yorkshire Dart Board" (senza tripli e senza anello esterno del bull). Ottenne il permesso di sistemare il bersaglio nell'aula e si accinse a dimostrare che il gioco delle freccette era senz'altro di abilità. Nell'atmosfera austera della sala Foot centrò tre volte lo spicchio del 20, poi chiese al collegio giudicante di provare a fare altrettanto. Il giudice ordinò a un giovane messo di tentare l'impresa, ma egli mancò il bersaglio in tutti e tre tiri. A questo punto mentre il giudice meditava il sig. Foot mise tre freccette sul doppio venti.

"Potete farlo di nuovo, Sig. Anakin?" chiese il giudice

"Sì", rispose Foot

Ritornò alla postazione e centrò di nuovo tre volte lo spicchio del venti. A quel

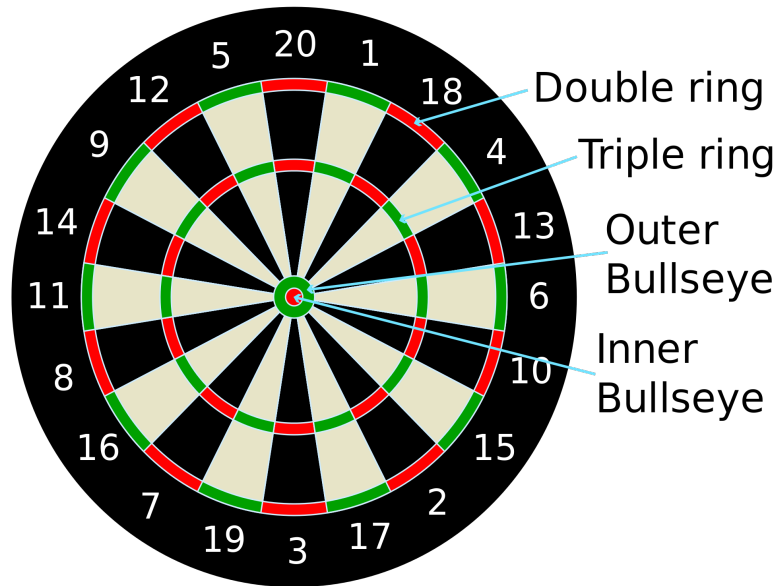


Figura 1.1: Dartboard classica utilizzata nei tornei professionistici

punto tra la meraviglia della gente risuonano le storiche parole del giudice: "Questo non è un gioco d'azzardo" e la causa venne così archiviata.

Altra curiosità riguarda la linea di tiro, chiamata "Oche". Pare che nel 1908 la Birreria di "S. Hokey Sons" molto conosciuta nel Devon, Dorset e Cornovaglia, consegnasse la birra in casse di legno della misura di tre piedi (0,91 m.). In alcuni Pub di Portsmouth venne l'idea di usare tre casse messe in fila dal muro come demarcazione della linea di tiro, ecco perchè la frase "Toe to the Oche". Nel 1911 però, la birreria venne venduta e i nuovi proprietari introdussero un nuovo tipo di cassa per la consegna che misurava 2 piedi (0,61 m.). I pubs si adeguarono usando 4 casse e riducendo così la linea di tiro ad un più consona 8 piedi (2,44 m.) che resistette sino ai nostri giorni, quando nel 1977 al consiglio direttivo della "World Darts Federation" fu definita la distanza tuttora in vigore di 7ft. 9 1/4in, equivalente a 2.37 metri. Per l'altezza invece, pare non ci sia mai stata nessuna contestazione, è sempre stata universalmente 5ft. 8in. (1,73m.). Statistiche storiche segnalano che il primo 180 (T20) cui si ha traccia è del 1902, in un pub del Sussex, da parte di un certo John Reader, che lavorava come assistente in una impresa di pompe funebri (si veda S [4]).

1.2 Cenni sul Regolamento

Le partite ufficiali di Darts si svolgono tra 2 giocatori, uno contro l'altro, su tornei suddivisi quindi in playoff. Ogni partita è suddivisa in leg dispari e può essere strutturata in più set. Ogni set comprende più legs, dove il vincitore di un set è colui che vince più legs, cosicché il vincitore della partita sarà quello che vince più set. Per capire meglio il contesto, si tratta di un match uno contro uno come nel tennis, con la differenza che i games sono chiamati legs. I tornei più importanti vanno dai 3 ai 9 legs, alcuni prevedono che gli incontri siano formati da diversi set, ognuno dei quali composto da legs. Per la vittoria possono servire, ad esempio, 5 set vinti (su 9) al meglio dei 3 leg ciascuno. L'area di gioco è una porzione che deve essere ben individuabile all'interno del luogo di svolgimento di una gara. Comprende una pedana e un bersaglio costruito in fibre di setola con un diametro di 453 millimetri. Il centro del bersaglio viene fissato a un'altezza da terra di 173 cm, con una linea di tiro di 237 cm. Le freccette non devono superare la lunghezza totale di 30.5 cm, nè pesare più di 50 grammi e ogni giocatore lancia un massimo di 3 frecce per turno (F.I.G.F., B [2]). Il bersaglio è diviso da fili metallici (spider) in 20 segmenti di uguale dimensione da 1 al 20, ma non in ordinati consecutivamente e con un centro diviso in 2 settori concentrici. Il settore denominato col numero 20 va posizionato in alto al centro. Ogni segmento è diviso in varie porzioni, che a seconda del punto di arrivo della freccetta, determina il punteggio in maniera diversa:

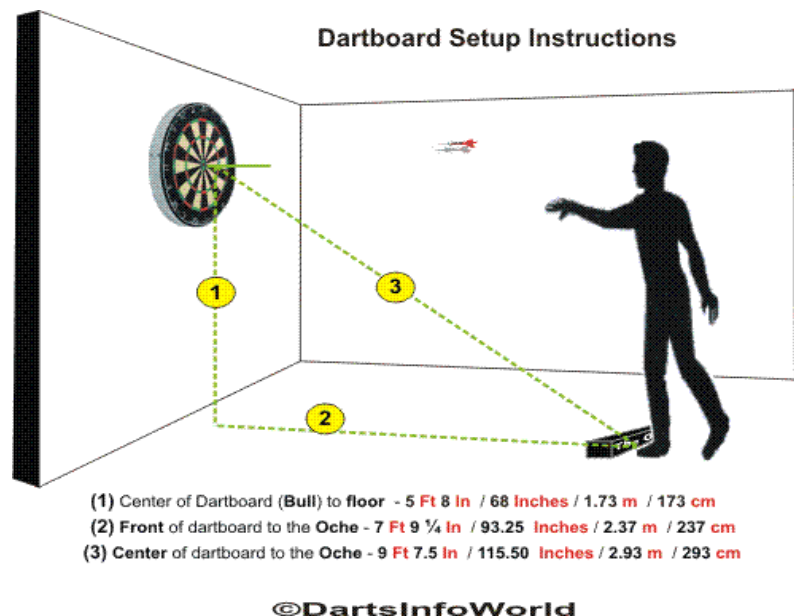


Figura 1.2: Misure e distanze dalla dartboard utilizzate da regolamento

- Colpendo l'anello esterno ("doppio"), colorato di verde o rosso, il punteggio viene raddoppiato (indicheremo con la lettera "D", es. D20).
- Colpendo l'anello più interno("triplo"), colorato di verde rosso, il punteggio viene triplicato. Qui è presente il punteggio più alto che un giocatore possa realizzare, colpendo tutte e tre le frecce a disposizione nel triplo venti (indicheremo con la lettera "T", es. T20).
- Colpendo l'area verde del cerchio esterno del centro, si realizzano 25 punti, mentre il centro del bersaglio rosso, bull's eye, vale 50 punti.
- Colpendo le sezioni interno o esterne, il punteggio equivale ai punti indicati sui bordi esterni del settore (indicheremo con la lettera "S", es. S20).
- Una freccia che va al di fuori del bersaglio varrà zero.

Molti di voi si saranno chiesti il perchè di questa strana distribuzione dei punteggi sulla dartboard. Il sistema di numerazione del bersaglio è allocato in questa maniera con lo scopo di ridurre i lanci con punteggi fortuiti, abbassandone così la probabilità. I numeri sono disposti in questo ordine in modo da premiare la precisione e punire la scarsa mira. Disponendo in questo modo numeri alti vicino a numeri bassi, eludiamo tiri poco accurati producano punteggi alti. Come si può vedere un errore verso lo spicchio del venti può portare al 5 o all'1 come punteggio.(si veda S [4])

1.3 Modalità di Gioco e Varianti

501

Le modalità di gioco più note sono : "301","501","701" e "1001". Il gioco consiste nel sottrarre i punti che realizza il giocatore al suo punteggio iniziale, su ogni turno di tiro, composto complessivamente da 3 tiri(freccette), con lo scopo di arrivare a zero partendo da una modalità di gioco(es. 301). La regola per chiudere il leg sarà quella di realizzare un double come ultimo tiro. Nel caso il giocatore non ci riuscisse, o si avvicinerà a zero per poi chiudere successivamente, o in caso contrario si andrà a sfiorare il punteggio("going bust") e il lancio sarà nullo. Per lancio nullo si intende per esempio, con un punteggio 8, mirando verso il D4 per chiudere, con un'errata mira verso lo spicchio sovrastante(18) si andrà in "bust". Da lì si ritorna sempre a 8 e il turno passerà direttamente al nostro avversario, anche se ci rimangono ancora 2 frecce da tirare. Questa regola viene anche chiamata "Regola del Sovrapunteggio" ed è applicata anche nel caso si

arrivasse a uno, poichè da lì non sarà più possibile finire il match con un doppio (F.I.G.F., B [2]).



Figura 1.3: Valore dei doppi e tripli sulla Dartboard

Per capire meglio il contesto questi sono alcuni esempi di chiusura di un leg:

- triplo 20 (60), triplo 19 (57) e doppio 12 (24)
- triplo 20 (60), triplo 17 (51) e doppio 15 (30)
- triplo 20 (60), triplo 15 (45) e doppio 18 (36)

1.3.1 Round the clock(l'orologio)

Gioco individuale per qualsiasi numero di partecipanti, ideale per migliorare la propria mira. Ogni giocatore lancia a turno tre freccette. Dopo un double di partenza, ogni giocatore deve tirare una freccetta in ogni settore, nella sequenza da 1 a 20; naturalmente le freccette che colpiscono settori diversi da quelli desiderati, non vengono contati e vince il primo che termina la sequenza. Una variante prevede un secondo giro, in cui vanno colpiti tutti gli anelli doppi dei numeri, e un terzo, per gli anelli tripli (F.I.G.F. , S [3]).

1.3.2 Cricket

Il gioco del Cricket viene applicato solo nei settori che vanno dal 15 al 20 più il bull(ordine a discrezione del giocatore). Ogni giocatore, deve colpire per 3 volte lo stesso numero singolo, per "chiudere" il settore e poter passare agli altri. Si possono colpire anche i settori multipli(doppi e tripli), in modo da velocizzare le giocate chiudendo prima. Il gioco in questione propone la variante del Cut-Throat: un giocatore può "caricare" i punti all'avversario, tirando ad un settore già chiuso da lui, ma non ancora completato dall'avversario. In questa variante vince chi ha meno punti. (F.I.G.F. , S [3])

1.3.3 Killer

Gioco ideale se si è tra i 4 e gli 8 partecipanti, può essere spietato e divertente. Il gioco richiede buona mira, buona strategia e un pizzico di fortuna, comprende numerose varianti e lo scopo è quello di eliminare gli avversari rimanendo così l'unico superstite. Il gioco inizia tirando verso il bersaglio con la mano opposta a quella che si usa, la zona in cui planerà la freccia sarà il nostro numero. In questo modo ogni giocatore ha un proprio numero personale, se si colpisce un numero già uscito si deve ritirare la freccia.

Una volta che ognuno ha un numero e si è determinato l'ordine di lancio, ogni giocatore lancia a turno tre freccette cercando di colpire il doppio del suo numero. Da qui diventa un assassino e mirando verso i doppi dei propri avversari può eliminarli. Ogni giocatore ha disposizione 3 vite, al termine delle quali viene eliminato dal gioco e l'ultimo che rimane in gioco vince. Dopo aver colpito il proprio doppio, se si ricentra per sbaglio si perde una vita, rischiando di uccidere se stessi (F.I.G.F., S [3]).

1.3.4 Split

E' un gioco per un numero qualsiasi di persone, dove si lanciano tre frecce per ogni turno. All'inizio si sceglie una sequenza del tipo: 15,16,un qualsiasi doppio,17,18,un qualsiasi triplo,19,20 e il bull. A ogni lancio viene sommato il valore del settore colpito al punteggio totale e nel caso il giocatore non riesca a colpire lo spicchio richiesto con le tre freccette a disposizione, allora il punteggio totale viene dimezzato. In questo gioco vince chi totalizza il numero maggiore di punti (si veda S [4]).

1.4 Tornei e Giocatori

1.4.1 PDC e BDO

Potrebbe sembrare banale, ma spiegare federazioni, tornei e campionati non è del tutto facile nei darts.

Cerchiamo di capire come funziona il sistema darts nel mondo, abbiamo innanzitutto due federazioni: la Professional Darts Corporation (PDC) e la British Darts Organisation (BDO). La BDO è (o meglio era) l'organizzazione principe di questo sport, fondata nel 1973, con lo scopo di far crescere il gioco delle freccette fino a farle diventare uno sport popolarissimo, con diversi tornei trasmessi nel Regno Unito dalla BBC e da ITV e molti sponsor pronti ad investire. Le freccette ebbero un periodo di popolarità, con giocatori che divennero veri e propri professionisti, fino alla fine degli anni '80, quando ci fu un drastico calo di questo gioco. La crisi dovuta all'immagine dei giocatori legata all'uso di alcool in pedana, non convinse molti sponsor a investire. Qui, molti giocatori decisero di prendere in mano la situazione e i migliori sedici giocatori di allora, fondarono nel 1994 la Professional Dart Corporation, con l'obbiettivo di aumentare l'esposizione mediatica e migliorare l'immagine delle freccette. Questo portò a una spaccatura tra la BDO e la PDC, i giocatori di quest'ultima vennero sospesi da qualsiasi organizzazione della BDO, formando dei nuovi tornei con l'aiuto di sponsor e finanziamenti. La battaglia si inasprì nel 1997, ma un accordo tra le due organizzazioni stipulò la libertà dei giocatori di partecipare a tornei di entrambe le organizzazioni.

La PDC dopo un difficile inizio, con un forte ostacolo della controparte, riuscì a creare una struttura di tornei professionistici, anche grazie all'arrivo della piattaforma satellitare Sky. Mentre la BDO, grazie alla visibilità garantita dalla BBC e dalla sua forte struttura storica, continuò a sfornare nuovi giocatori, fino a quando nel 2003 il governo britannico vietò le sponsorizzazioni di tabacco negli sport. Quando l'Embassy, sponsor principale della BDO, si dovette ritirare, la PDC divenne organizzazione leader nei tornei di darts.

Oggi la differenza tra le due federazioni è incolmabile, il livello della PDC è nettamente superiore a quello della BDO, a livello di sponsor, prize money e copertura televisiva. Se la PDC è un'organizzazione ottima per i migliori giocatori al mondo e per lo sport a livello professionistico, la BDO rimane fondamentale per le freccette essendo l'organizzazione che si occupa dello sport a livello amatoriale, giovanile e femminile (si veda S [11]).

1.4.2 Il PDC World Darts Championship

Spettacolo e divertimento, condito da alti montepremi, tornei in giro per il mondo e format per la qualificazioni ai tornei televisivi; in questo si può definire la programmazione offerta dalla Professional Darts Corporation. Il complesso

panorama dei tornei di freccette a livello mondiale possiamo dire che segue il modello Tennis ATP. Come per la stagione del tennis abbiamo diversi tornei di livello mondiale e nazionale, dai principali come Roland Garros, Wimbledon, US Open, AUS Open ai minori come Roma o Madrid; per i Darts il meccanismo è pressochè esattamente lo stesso.

Si inizia la stagione con il The Master, in cui i migliori 16 players del mondo si sfidano in una due giorni molto intensa e di grande spettacolo. Si inizia a Febbraio con la Premier League of Darts, torneo a giornate, comprende i migliori dieci al mondo, con un match a settimana in diverse città inglesi e irlandesi.

L'UK Open invece dà la possibilità anche ai giocatori amatoriali di sfidare i giocatori più forti e di utilizzare la formula del random draw alla fine di ogni round, senza alcun tipo di teste di serie. A Francoforte in giugno ci sarà World Cup of Darts, torneo composto da 32 Nazionali, formate da 2 giocatori cadauna, che si affrontano nel singolo e doppio. Uno dei tornei principali, un classico della PDC, è il World Matchplay di Blackpool, che si svolge a Luglio. L'estate offre numerosi tornei di esibizione, che la PDC organizza per far espandere le freccette in tutto il globo, toccherà nel 2016 Asia ed Oceania. A ottobre a Dublino, si svolge il World Grand Prix, unico torneo dell'anno che viene disputato con la regola del Double in/Double Out, dove bisogna aprire la partita con un doppio. L'unico vero torneo dell'anno in cui si affrontano sia giocatori PDC che BDO è il Grand Slam of Darts dal 12 al 20 Novembre. La stagione termina con le Players Championship Finals e con il torneo più famoso, svolto a dicembre, a Londra, il World Darts Championship (Campionato del Mondo di Darts) legato alla partnership di William Hill porta montepremi finali elevatissimi (si veda S [11]).

Tra questi tornei solo alcuni di essi influiscono sul ranking dei giocatori, per esempio alcuni come la World Cup of Darts non viene conteggiato. Le Partite sono disputate esattamente come nel tennis a playoff. Per esempio, si ha un campione pari di giocatori e da lì si va dai sedicesimi di finale fino alla fase conclusiva della finale, in scontri uno contro uno. In ogni torneo, addirittura in ogni fase, può varare il numero di set o di leg con il quale si gioca. Spesso più si va avanti nel torneo, più per vincere una partita servono più set.

Tra i giocatori più famosi di questo sport possiamo citare: Phil Taylor (The Power), britannico e leggenda vivente di questo sport, considerato tra i migliori con 207 tornei vinti, tra cui 81 major e 16 campionati del mondo. Unico giocatore nella storia ad essere riuscito nell'impresa di realizzare due "nine-dart finish" (chiudere il leg in solo 9 tiri) in una sola partita. E' uno dei fondatori della PDC (si veda S [7]).

Altri giocatori noti da segnalare sono: Gary Anderson, giocatore scozzese soprannominato The Flying Scotsman, vincitore delle ultime due edizioni (2014,2015) della World Darts Championship, numerosi tornei come la Premier League of

Darts, Players Championship Finals. Adrian Lewis(Jackpot), inglese e molto più giovane degli ultimi due, vincitore di 2 PDC World Championship e altri tornei minori come gli UK Open e l'European Championship (si veda S [5]).

Capitolo 2

Le Freccette da un Punto di vista Dinamico

In questo capitolo ci soffermeremo sulla componente dinamica del gioco delle freccette, prima di fare un'analisi statistica sul gioco in sé. Il moto di una freccetta lanciata verso il Bullseye è molto simile, per esempio, a quello di un lancio di una pietra o di una palla da baseball di un lanciatore verso il battitore, quindi si tratta di un moto parabolico. La curva può essere più alta o più bassa, in conseguenza a quanta forza utilizzi per tirare le freccette. Una buona tecnica di tiro, segue la freccetta lungo tutta la parabola di volo durante l'accelerazione e deve garantire, che tale parabola venga seguita dalla freccetta, anche dopo il lancio dalla mano.

Affinchè manteniamo la freccia nella giusta posizione dobbiamo prima guardare i meccanismi del braccio di tiro. Esso può essere descritto come "una macchina di 3 leve" fissate l'una a l'altra da 2 giunti mobili e da 1 giunto fisso: i giunti mobili saranno il gomito e il polso, il giunto fisso sarà la spalla mentre le tre leve saranno il braccio, l'avambraccio e la mano.

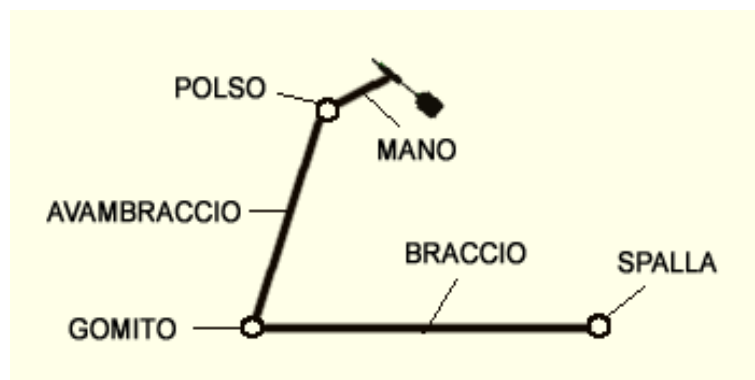


Figura 2.1: Posizione di lancio classica nel gioco delle freccette.

La spalla è l'unico punto dell'intero processo che non cambia mai la posizione e non va mossa durante il tiro. La posizione del gomito invece rimane inalterata durante il caricamento e va spostata in avanti durante la fase di accelerazione. In questo caso abbiamo teorie contrastanti con chi sostiene che la posizione del gomito non vada modificata durante il tiro. Un lancio a gomito fisso anticiperebbe il rilascio della freccia, mentre se si vuole mantenere la parabola il braccio andrebbe disteso seguendo essa più a lungo possibile.

Anche il polso spesso è al centro di molte discussioni, in cui tiene la possibilità di eseguire "la mossa della frusta". Molti giocatori di freccette professionisti, tendono a lanciare le freccette dando un colpo secco col polso, come se esso fosse la parte finale di una frusta. Questo approccio aiuta nell'accelerazione e schioccando il polso, le freccette andranno più veloci, di conseguenza le altre parti del braccio potranno essere mosse più lentamente, migliorando la precisione. Questa particolarità aggiunge però un altro elemento da controllare nel movimento di tiro, portando a una maggiore attenzione e aumentando le possibilità di errore (si veda S [4]).

Nel nostro caso, l'analisi verrà eseguita studiando la posizione direzionale della mano sulla freccia, la quale, dà un angolo φ per mirare il dart dall'basso verso l'alto e un angolo θ nella direzione sinistra-destra.

2.1 Il moto parabolico

Soffermiamoci ora sullo studio della componente fisica delle freccette, come detto in precedenza, dettate da un moto parabolico. Il problema nasce dal tentativo di modellare inanzitutto il lancio di una singola freccetta e la distribuzione di probabilità che descrive questo evento. Il moto parabolico è un tipo di moto bidimensionale esprimibile tramite la combinazione del moto rettilineo uniforme e il moto uniformemente accelerato (si veda S [6]). Esso è descritto mediante la relazione della cinematica che lega i vettori posizione, velocità e accelerazione. Le componenti principali di tale moto sono:

- Tutta la massa e la geometria del corpo sono concentrate in un unico punto
- L'accelerazione del moto è verticale e il suo modulo è $g=9.81m/s^2$ che corrisponde all'accelerazione di gravità della crosta terrestre
- Forme di attriti legati alla resistenza dell'aria vengono trascurati

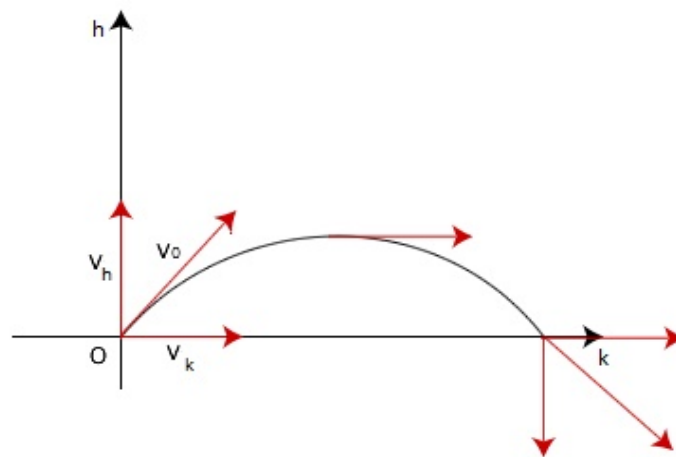


Figura 2.2: Rappresentazione grafica del moto parabolico.

Un corpo lanciato all'istante $t=0$ nell'origine O , in un sistema di coordinate cartesiane OKH , con velocità iniziale di modulo V_0 formi un angolo φ con l'asse orizzontale. Dalle leggi del moto uniformemente accelerato si ha:

$$V(t) = V_0 + \int_0^t a(t) dt$$

Dove:

V_0 = Velocità iniziale

$a(t)$ = Funzione costante con valore pari a $-g$ diretta lungo la perpendicolare al terreno

La velocità giace sempre nel piano formato dai vettori costanti V_0 e g , ovvero quello su cui si svolge il moto. Il vettore velocità può essere scomposto lungo le due componenti h e k :

$$V_0 = v_0 \cos\varphi \cdot \widehat{v}_k + v_0 \sin\varphi \cdot \widehat{v}_h$$

Dalla relazione precedente andiamo ad ottenere:

$$\begin{aligned} V(t) &= V_0 + \int_0^t a(t) dt \\ &= V_0 - g \cdot t \cdot \widehat{v}_h \\ &= v_0 \cos\varphi \cdot \widehat{v}_k + (v_0 \sin\varphi - g \cdot t) \widehat{v}_h \end{aligned} \quad (2.1)$$

Proiettando le velocità sugli assi si ottengono le componenti:

$$v_k = v_0 \cos\varphi$$

$$v_h = v_0 \sin\varphi - g \cdot t$$

Dove il parametro t della legge oraria $k(t)$ sarà:

$$t = \frac{k}{v_0 \cos\varphi}$$

In questo caso stiamo studiando le componenti del vettore velocità in una dimensione bi-dimensionale, usando le coordinate polari, nel nostro caso l'analisi che prenderemo in esame sarà ancora più ampliata. Analizzeremo il moto parabolico di una freccia usando un ulteriore componente Y e un angolo θ che rappresenterà la virata della freccia del tiratore (da sinistra verso destra o da destra verso sinistra), mentre l'angolo φ rappresenterà l'angolo di ampiezza (verticale) presente anche nel moto parabolico a due dimensioni.

Immaginiamo che il giocatore voglia colpire il punto (d, x, y) del bersaglio, spesso sarà il T20. La sua posizione di lancio della sua mano sia $(0, 0, 0) = 0$ qualunque sia la velocità impressa. Analizzeremo tutto questo, con un'ipotesi forte, cioè che il movimento del gomito non verrà considerato, ma esso sarà fisso creando così un angolo φ . Data la nostra posizione di partenza (d, x, y) con $d=2.37$ m, possiamo risalire alla velocità v_0 da imprimere per colpirlo.

$$\vec{v}_0 = r(\cos\theta \cos\varphi; \cos\varphi \sin\theta, \sin\varphi)$$

è interamente determinato da (r, φ, θ) .

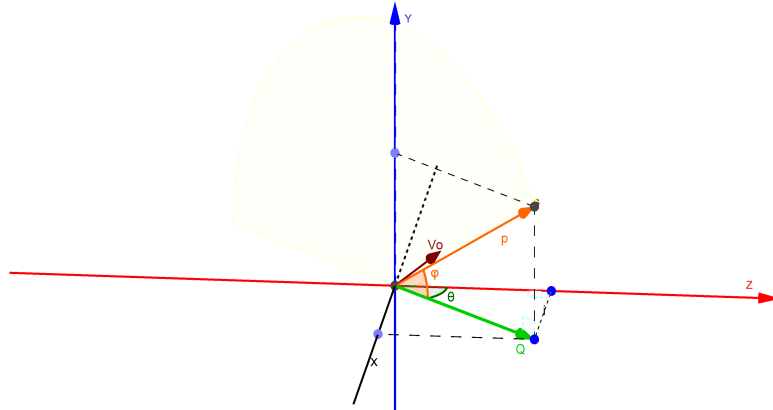


Figura 2.3: Rappresentazione di una freccetta sul piano, in funzione di V_0, θ e φ .

Da qui possiamo applicare una semplice sostituzione con il caso visto precedentemente relativo al moto parabolico bidimensionale. Nel nostro caso la componente k sarà divisa in 2 parti formate da: gli assi della x e della z , mentre la componente delle ordinate h sarà rappresentata da y . Procedendo quindi con la sostituzione e tramite l'uso delle coordinate sferiche, le nostre componenti saranno:

$$v_x = r \cdot \cos\varphi \cos\theta$$

$$v_y = r \cdot \cos\varphi \sin\theta$$

$$v_z = r \cdot \sin\varphi - gt$$

$$r \in (0, \infty), \theta \in (0, 2\pi), \varphi \in (0, \pi)$$

integrando su t ricaveremo le leggi orarie dei moti lungo gli assi x, y , e z :

$$z(t) = r \cdot \cos\varphi \cos\theta t$$

$$x(t) = r \cdot \cos\varphi \sin\theta t$$

$$y(t) = r \cdot \sin\varphi - g \frac{t^2}{2}$$

La traiettoria la ricaveremo eliminando la variabile temporale, realizzando un rapporto tra le componenti. L'intersezione della parabola con il piano dunque sarà:

$$\frac{x}{z} = \frac{r \cdot \cos\varphi \cos\theta t}{r \cdot \cos\varphi \sin\theta t} = X$$

$$\frac{y}{z} = \frac{r \cdot \sin\varphi d}{r \cdot \cos\varphi \cos\theta} - \frac{g d^2}{2 r^2 \cos^2\varphi \cos^2\theta} = Y$$

dove:

$$g = 9.81 \text{ m/s}^2$$

$$t = \frac{d}{r \cos\varphi \cos\theta}$$

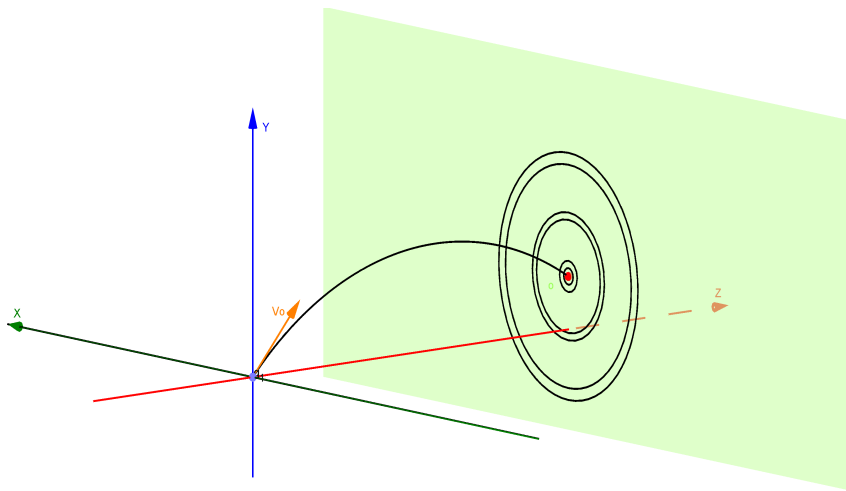


Figura 2.4: Rappresentazione geometrica del lancio di un dardo verso bersaglio.

L'obiettivo del giocatore è quello di mirare la freccia in un determinato punto, in base all'andamento della partita e a quanto sia più o meno vicino a chiudere il suo 301 o 501. Con lo scopo di colpire un punto (d, \bar{x}, \bar{y}) , le terne utili sono date da:

$$\theta = \arctan\left(\frac{\bar{x}}{d}\right)$$

$\varphi \in \left[\arctan\frac{\bar{y}}{\sqrt{z^2 + x^2}}, \frac{\pi}{2} \right]$ dove $\varphi \in \arctan\frac{\bar{y}}{\sqrt{x^2 + z^2}} = \varphi_{min}$ corrisponde all'angolo limite,

r: dato un angolo φ , la velocità in modulo deve essere tale che:

$$Y = \frac{d \sin\varphi}{\cos\varphi \cos\theta} - \frac{g d^2}{2 r^2 \cos^2\varphi \cos^2\theta}$$

da qui invertendo la formula, troviamo la r^2

$$r^2 = \frac{g d^2}{2 \left(\frac{d \sin \varphi \cos \varphi}{\cos \theta} - \bar{z} \cos^2 \varphi \right) \cos^2 \theta} = \frac{g d^2}{2 (d \sin \varphi \cos \varphi \cos \theta - \bar{z} \cos^2 \varphi \cos^2 \theta)}$$

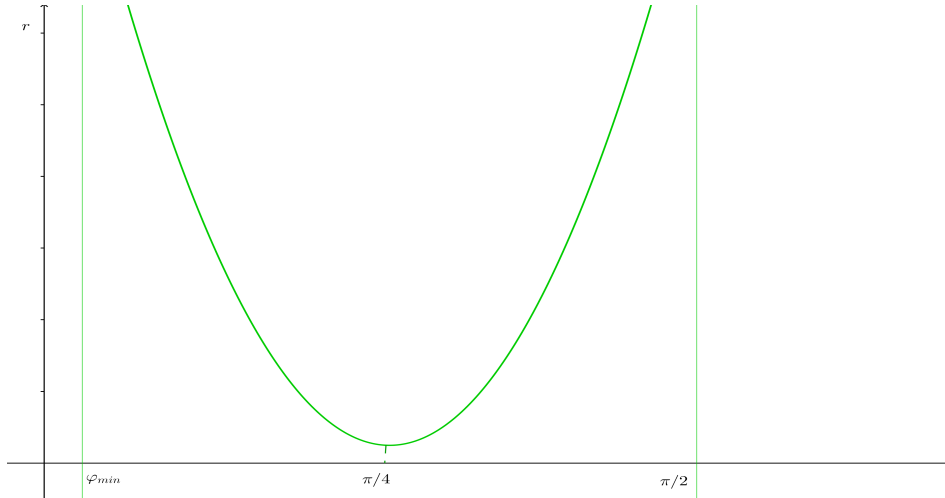


Figura 2.5: Punti di massimo che deve raggiungere la velocità dato φ , con $\theta=0$.

Ovvero una funzione con questo profilo (si veda Figura 2.5).

Quindi se chiamiamo F la mappa che manda (r, φ, θ) nel punto corrispondente sul bersaglio:

$$F(r, \varphi, \theta) = \begin{cases} z = d \\ x = d \tan \theta \\ y = \frac{d \sin \varphi}{\cos \varphi \cos \theta} - \frac{g d^2}{2 r^2 \cos^2 \varphi \cos^2 \theta} = \frac{d \tan \varphi}{\cos \theta} - \frac{g d^2}{2 r^2 \cos^2 \varphi \cos^2 \theta} \end{cases}$$

Allora le terne:

(r, φ, θ) per cui $F(r, \varphi, \theta) = (d, \vec{x}, \vec{y})$ sono da immaginarsi come curve sul piano

$\theta = \arctan\left(\frac{\vec{x}}{d}\right)$ che seziona lo spazio.

Al variare di $(d, \vec{x}, \vec{y}) \in \text{Piano}\langle x = z \rangle$ le fibre $\overleftarrow{F}(d, \vec{y}, \vec{z})$ esauriscono i punti dello spazio

$$D :=]0, \infty[, x] - \frac{\pi}{2}, \frac{\pi}{2} [, x] - \frac{\pi}{2}, \frac{\pi}{2} [\subseteq \mathfrak{R}^3$$

Per cui se $(r, \varphi, \theta) \in D$, esiste una ed una sola curva del tipo sopra rappresentato che lo contenga.

2.2 La Distribuzione di Probabilità di (X,Y)

Possiamo immaginare che attorno il valore scelto di \bar{v}_0 per il lancio (scelta che il giocatore fa mirando) si sviluppino 3 traiettorie:

$$\begin{cases} W_r = r - \bar{r} \sim N(0, \sigma_r) \\ W_\varphi = \tan(\varphi - \bar{\varphi}) \sim N(0, \sigma_\varphi) \\ W_\theta = \tan(\theta - \bar{\theta}) \sim N(0, \sigma_\theta) \end{cases}$$

Di modo che se il giocatore cerca di lanciare con velocità iniziale data da $(\bar{r}, \bar{\varphi}, \bar{\theta})$, lancerà invece con velocità (r, φ, θ) con probabilità:

$$f_{W_r, W_\varphi, W_\theta}(r - \bar{r}, \varphi - \bar{\varphi}, \theta - \bar{\theta}) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} r - \bar{r} \\ \tan(\varphi - \bar{\varphi}) \\ \tan(\theta - \bar{\theta}) \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} r - \bar{r} \\ \tan(\varphi - \bar{\varphi}) \\ \tan(\theta - \bar{\theta}) \end{pmatrix} \right\}$$

Dove:

$$\Sigma = \begin{pmatrix} \sigma_r^2 & 0 & 0 \\ 0 & \sigma_\varphi^2 & 0 \\ 0 & 0 & \sigma_\theta^2 \end{pmatrix} = \begin{pmatrix} \sigma_r^2 & 0 & 0 \\ 0 & \sigma_\varphi^2 & 0 \\ 0 & 0 & \sigma_\theta^2 \end{pmatrix}$$

Possiamo assumere che gli errori di mira siano pressochè indipendenti, la distribuzione diventa:

$$f_{W_r, W_\varphi, W_\theta}(r - \bar{r}, \varphi - \bar{\varphi}, \theta - \bar{\theta}) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp \left\{ -\frac{1}{2} \left(\frac{r - \bar{r}}{\sigma_r^2} + \tan \frac{\varphi - \bar{\varphi}}{\sigma_\varphi^2} + \tan \frac{\theta - \bar{\theta}}{\sigma_\theta^2} \right) \right\}$$

Il nostro obiettivo è quello di studiare come questa distribuzione si trasformi sul bersaglio. Sappiamo che:

$$\tan \theta = \frac{\sin \theta}{\cos \theta} = \frac{\sqrt{1 - \cos^2 \theta}}{\cos \theta}$$

$$\cos \theta = \frac{1}{\sqrt{\tan^2 \theta + 1}} = \frac{1}{\sqrt{W_\theta^2 + 1}} \quad \sin \theta = \frac{\tan \theta}{\sqrt{\tan^2 \theta + 1}} = \frac{W_\theta}{\sqrt{W_\theta^2 + 1}}$$

Consideriamo la legge congiunta del vettore aleatorio (X, Y) , dove:

$$\begin{cases} X = d \cdot W_\theta \\ Y = d \cdot W_\varphi \sqrt{1 + W_\theta^2} - \frac{g d^2 (1 + W_\theta^2) (1 + W_\varphi^2)}{2 W_r^2} \end{cases}$$

Vediamo subito che $X \sim N(0, d^2 \sigma_\theta)$ e studiamo la distribuzione di $Y|X=x$. Abbiamo che:

$$P[Y \leq y | X = x] = P \left[d \cdot W_\varphi \sqrt{1 + \frac{x^2}{d^2}} - \frac{g d^2 \left(1 + \frac{x^2}{d^2}\right) (1 + W_\varphi^2)}{2 W_r^2} \leq y \right]$$

Non siamo in grado di calcolare in forma chiusa questa probabilità, anche se per esempio è possibile, grazie all'ausilio di Geogebra, disegnare la regione per x fissato:

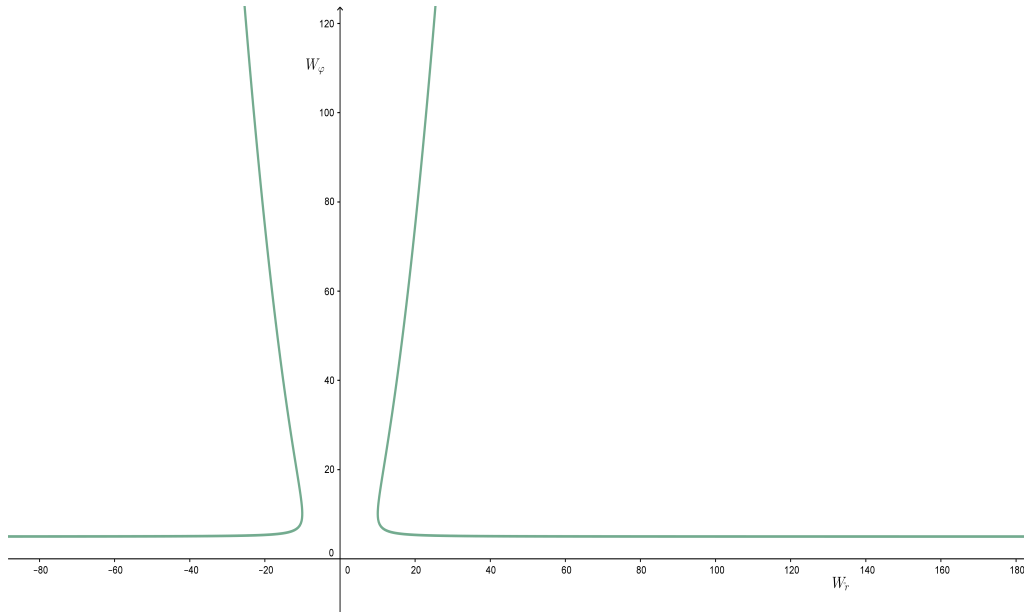


Figura 2.6: Distribuzione grafica di $Y|X=x$, ponendo W_φ in ordinata e W_θ in ascissa.

$$\left\{ (W_\varphi, W_r) : C_1 \cdot W_\varphi - C_2 \cdot \frac{1 + W_\varphi^2}{W_r^2} \leq y \right\}$$

al variare di $Y \in \mathfrak{R}$, dove $C_1 = C_1(x) = d \cdot \sqrt{1 + \frac{x^2}{d^2}}$ e $C_2 = C_2(x) = \frac{g d^2 \left(1 + \frac{x^2}{d^2}\right)}{2}$.

Non riuscendo a trovare una distribuzione (X,Y) matematicamente, tramite la Figura 2.7 vediamo come i lanci si distribuiscono sul bersaglio, simulando in questo caso 1000 tiri. Vediamo dal grafico in 3D, come entrambi le componenti mostrino un andamento normale, il modello si adatta a bene ai dati, anche pur avendo in y una concavità più estesa. La causa di questa anomalia, la vedremo più avanti, con lo studio delle componenti singolarmente, nel capitolo 4.

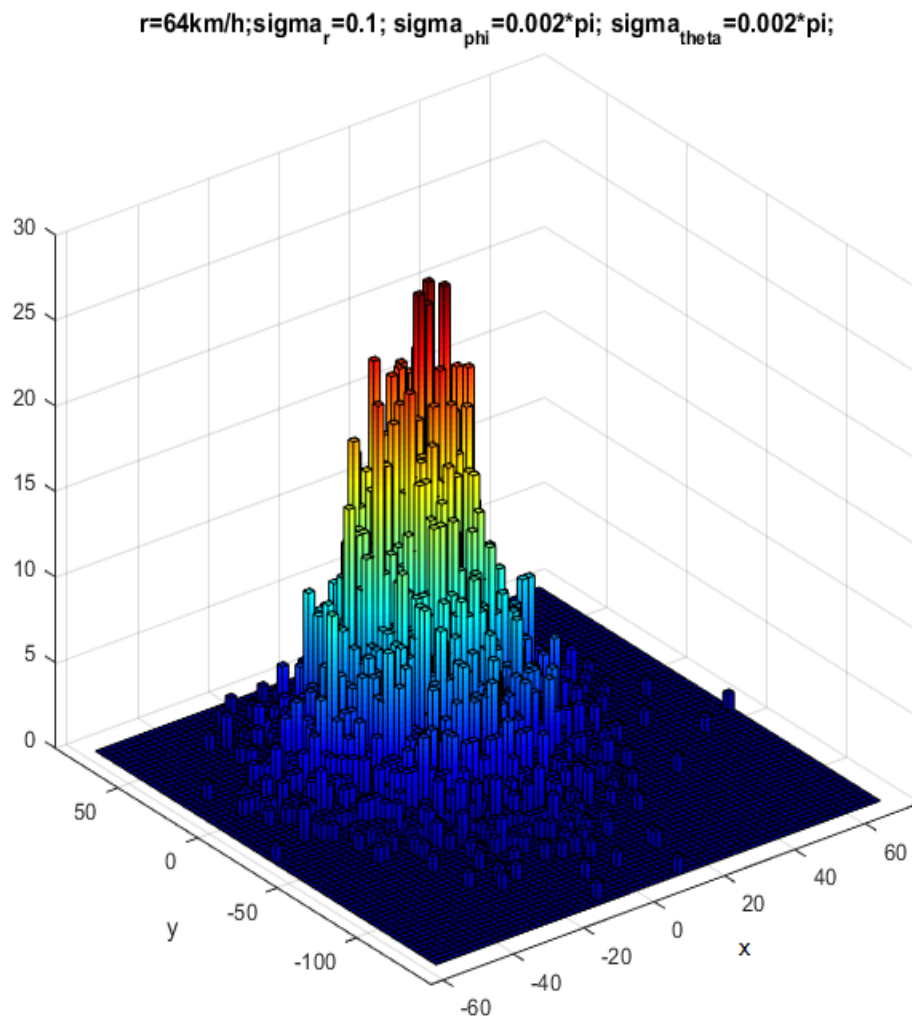


Figura 2.7: Distribuzione grafica di (X,Y) , simulando 1000 lanci, con matlab

Capitolo 3

Varianza di Tiro con l'Algortimo EM

3.1 Il triplo 20 non è per tutti

Il gioco dei Darts come tutti sappiamo si basa sulle capacità di ogni singolo giocatore. Sulla sua precisione, freddezza e calma. Se nel precedente capitolo abbiamo trattato il lancio delle freccette da un punto di vista fisico, stavolta ci soffermeremo di più sul lato prettamente statistico.

Le abilità che hanno i giocatori professionisti non sono le stesse che hanno gli amici che si trovano al Pub tra una pinta di birra e l'altra. Per un principiante lo score medio su 100 tiri potrebbe risultare addirittura più basso dello score medio usando una distribuzione uniforme. Questo può succedere principalmente per 2 motivi.

Il primo si basa sul fatto che in un Modello Uniforme non viene preso in considerazione lo zero, in poche parole non viene presa in considerazione l'ipotesi che la freccia manchi il bersaglio. Il secondo fattore riguarda la composizione della dartboard. Come ben si sa dal regolamento, vince il giocatore che colleziona più leg e di conseguenza set. Ogni giocatore mira verso lo spicchio superiore che cade perpendicolare sul centro, lo spicchio del venti, precisamente il Triplo 20, che porta 60 punti al tiratore e velocizza il suo countdown verso lo zero. Ma per un principiante vale la pena puntare lì?

Sicuramente no, poichè lo spicchio centrale è adiacente a quello del 5 e dell' 1 e ciò rischierebbe di far perdere molti punti verso la chiusura della partita.

Allora qual'è la migliore zona della tavola dove puntare? Questo dipende dal livello di abilità di ognuno di noi (si veda B [7](2011)). Proponiamo con un modello in cui poniamo lo 0 come centro del segnapunti e Z una variabile bi-dimensionale:

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

μ corrisponde alla posizione alla quale il giocatore vuole mirare, mentre ε è l'errore gaussiano a media Zero e varianza $\sigma^2 I$. In poche parole, ogni lancio che viene effettuato verso la dartboard, giungerà verso il punto di target scelto e avrà un certo margine di errore. Questo dipenderà dalla precisione del tiratore, misurata con la deviazione standardizzata. Più la deviazione standard sarà piccola, più il giocatore avrà un'ottima precisione e ciò implica una maggiore probabilità di colpire la zona che si è prefissato. Idem viceversa, un'alta varianza del giocatore, favorisce una mira spesso errata. Ora il problema che ci poniamo è: in base alla nostra accuratezza, dove ci conviene tirare per massimizzare il nostro score?

L'idea è quella di costruire una heatmap, in base alla varianza di tiro di ognuno di noi, mostri in quale zona del bersaglio dobbiamo concentrarci. Per esempio, chi ha una devianza di 5 mm, dovrebbe mirare tranquillamente al T20, invece conviene puntare al T19 chi ha una varianza di tiro più elevata, per esempio di $\sigma = 25$ mm.

L'uso di queste heatmaps personalizzate può aiutarci quindi a capire come massimizzare il nostro score, ma come potremmo stimare la nostra σ ?

Uno dei metodi più semplici e diretti, è quello di registrare tot tiri verso il bull's eye e calcolare la distanza tra il SB e le freccette affisse sul bersaglio. Chiaramente più tiri eseguiamo, più la stima dell'accuratezza sarà precisa (minimo 100 lanci). Nel paragrafo successivo quindi ci soffermeremo sullo studio di un algoritmo costruito per per la stima della deviazione standard di un un giocatore, in base alle sue abilità, con l'obiettivo che porti a massimizzare il proprio score (si veda B [7]).

Non bisogna essere per forza esperti giocatori di freccette, ma avere un buon uso della statistica, se si vuole trovare la via per ottimizzare il proprio punteggio in questo gioco.

Con un modello gaussiano possiamo spiegare un efficiente metodo per massimizzare il nostro score, scegliendo una zona del bersaglio ben definita nella quale puntare. L'area di punto ottimale cambia in base alla varianza di ognuno di noi, quindi spiegheremo un algoritmo da cui stimeremo la varianza di tiro, basata sul punteggio di 100 tiri effettuati verso la tavola puntando al double bullseye. Il Modello in uso sarà una chiara implementazione dell' Algoritmo EM o comunque una forma molto simile a esso.

Poniamo Z come una variabile Bi-dimensionale e notiamo $s(Z)$ come lo score. Cosichè lo score previsto sarà:

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

dove con Si,Di e Ti sono rispettivamente i singoli, doppi e tripli.

Precedentemente, abbiamo sostenuto che un principiante potrebbe tranquillamente far più punti, usando una distribuzione uniforme, invece di utilizzare la dartboard classica. Supponendo allora che Z sia distribuita uniformemente sulla tavola B(Figura 3.1), per ogni regione di S :

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

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

Figura 3.1: Distanze (raggi) in mm dal centro

Usando le misure standard del bersaglio circolare, possiamo calcolare il valore atteso del punteggio ottenuto usando la distribuzione uniforme:

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

Spesso questo valore risulta più alto rispetto a giocatori beginners che giocano mirando verso il T20.

Riprendiamo ora ad analizzare il modello $Z = \mu + \varepsilon$, quale scelta di μ produce il più grande valore di $E_{\mu, \sigma^2}[s(Z)]$?

Per un dato σ , considero di scegliere μ da massimizzare :

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

Un approccio analitico risulta troppo complicato, possiamo semplificare l'espressione in esame in questo modo:

$$(f_\sigma^2 * s)(\mu)$$

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

Faremo in questo caso un'operazione di convoluzione bivariata $N(0, \sigma^2 I)$ dove s rappresenta lo score e f_σ^2 la nostra densità.

Dalla teoria sulla convoluzione, \mathcal{F} e \mathcal{F}^{-1} sono rispettivamente la trasformata di Fourier e la sua inversa.

Costruiamo così 2 matrici bi-dimensionali sulla densità gaussiana e sulla stima degli score, usando una scala in millimetri per la dartboard. Calcolando le convoluzioni usando 2 FFTs (Fast Fourier Transform) e una inversa FFT otteniamo così il punteggio stimato per ogni μ fissato presente in una griglia accurata.

Questa procedura sarà utile a creare la nostra heatmap giallo-rossa dello score stimato dove colori luminosi mostrano la realizzazione di un alto score se il giocatore mira in quel punto. Al variare delle varianze, questi grafici cambiano notevolmente. A varianza pari a zero il nostro μ sarà distribuito facilmente attorno al T20 e da qui, all'aumentare del σ , cambieranno le zone della tavola dove tirare la freccia (TIBSHIRANI, PRICE, TAYLOR; B [5] (2011)).

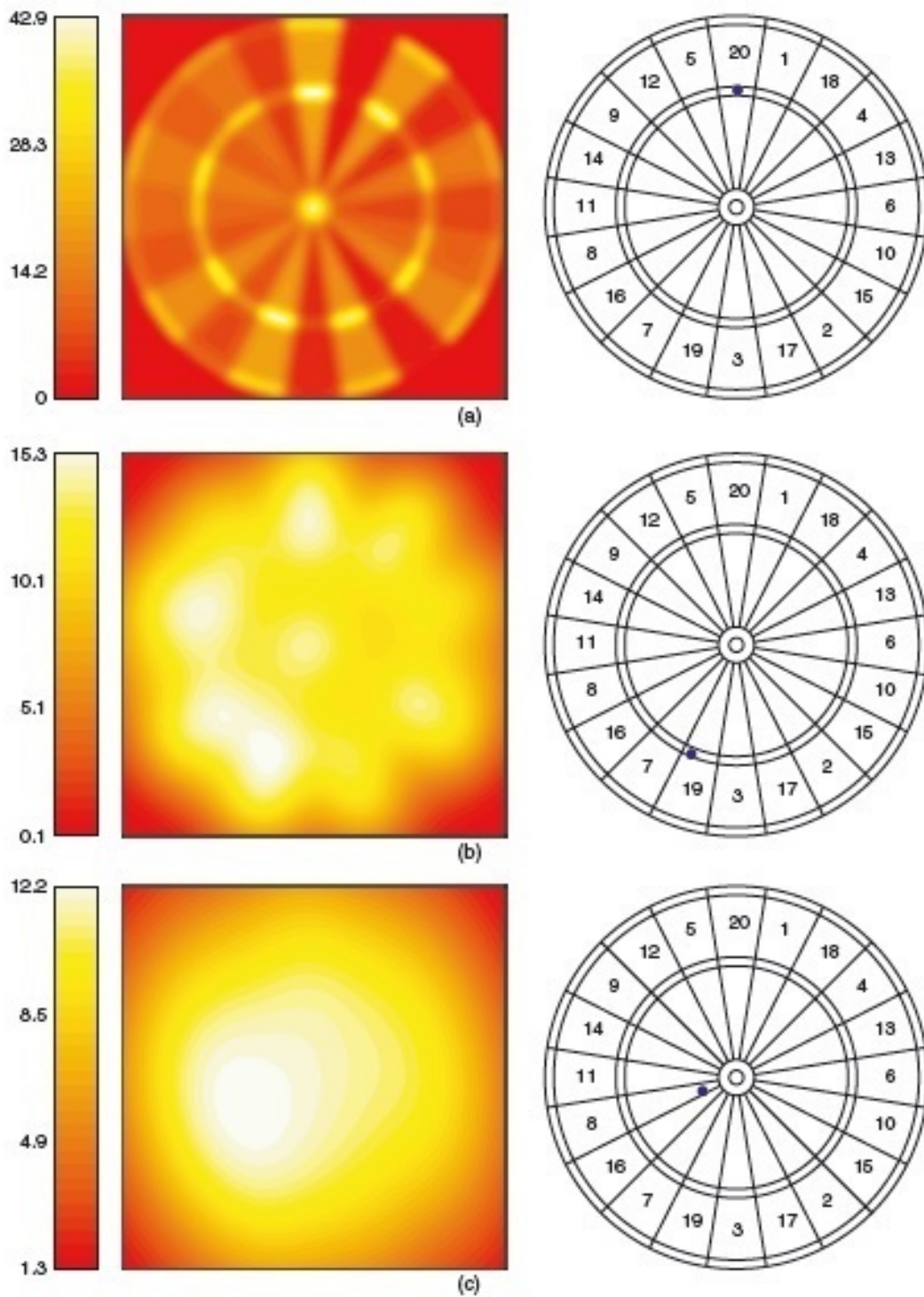


Figura 3.1a Heatmaps di $E_{\mu, \sigma^2}[s(Z)]$ per $(a) = 5$, $(b) = 26.9$, $(c) = 64.6$. La scala colori è in base al range di score e il punto blu rappresenta la zona di tiro ottimale.

3.2 L'Algoritmo EM

L'Algoritmo Expectation-Maximization, è un algoritmo che viene applicato per il calcolo delle stime di massima verosomiglianza, nel caso vi siano dati incompleti, è una valida alternativa all'utilizzo di metodi numerici come Newton-Raphson o Fisher-Scoring.

I punti di forza di questo algoritmo sono la generalità dei contesti in cui si può applicare, può essere utilizzato anche in problemi non riconducibili a dati mancanti e una semplicità di programmazione. L'algoritmo si basa sul rimpiazzare i dati mancanti con valori stimati opportunamente, da qui stimare i parametri a partire dall'insieme dei dati completi e dai parametri ottenuti è possibile ristimare i dati mancanti fino alla convergenza. Questo problema di massimizzazione della log-verosomiglianza è posto come un problema statistico e non semplicemente come un problema numerico.

L'Algoritmo riesce a emergere in moltissimi ambiti e viene applicato in molti problemi statistici in cui vi sono dati mancanti, come ad esempio: l'ANOVA, l'analisi delle tabelle di contingenza o le serie storiche. Prende anche in considerazione situazioni, in cui spesso non vengono considerati come problemi di dati incompleti come la stima dei modelli a componenti di varianza, i minimi quadrati pesati iterati, l'analisi fattoriale o l'analisi di dati censurati.

L'assenza di valori mancanti può portare a numerose difficoltà nel contesto statistico. C'è una perdita di efficienza delle stime poichè la numerosità campionaria viene ridotta e la matrice dei dati non è più rettangolare, contenendo delle celle vuote i metodi di analisi multivariata non sono direttamente applicabili. Il problema più grave che si verifica però è quello legato alla presenza di distorsione nelle stime di parametri. Questo si verifica quando le ragioni della mancanza di un dato, dipende dal valore del dato stesso, si parla allora di "Dati Mancanti Non a caso". Quindi sarà importante capire se i dati sono mancanti a caso o no.

Da qui ci sono diverse classificazioni di dati mancanti: i dati MAR sono dati mancanti quando la probabilità che un dato sia mancante, non dipende dal valore del dato stesso. Mentre i dati OAR si riferiscono ai dati osservati a caso, quando la probabilità che un dato sia osservato non dipende dal valore dei dati osservati. Nel caso in cui i dati, sia mancanti che osservati, non dipendono dai valori dei dati stessi, vengono chiamati Dati mancanti completamente a caso (MCAR), nel caso opposto il meccanismo è detto non ignorabile.

In un problema in cui abbiamo una funzione di densità $f(Y|\theta)$ dove Y è una Variabile Casuale contenente valori osservati e mancanti(MAR).

$$Y = (Y_{obs}, Y_{mis})$$

con Y_{obs} = valori osservati e Y_{mis} = valori mancanti. Il nostro scopo è quello di massimizzare rispetto a θ la verosomiglianza dei dati osservati. Nel caso però non vi siano soluzioni esatte nella risoluzione delle equazioni di verosomiglianza, possono essere applicati metodi iterativi, usando Newton-Rapson o Fisher-Scoring.

$$\theta^{(t+1)} = \theta^{(t)} + J^{-1}(\theta^{(t)}|Y_{obs}) \cdot u(\theta^{(t)}|Y_{obs})$$

Gli algoritmi iterativi presentano però dei problemi:

- Possono non convergere quando si è in presenza di strane funzioni di verosomiglianza, poichè non assicurano di aumentare la log-verosomiglianza ad ogni iterazione
- Richiedono il calcolo e l'inversione delle matrici, che diventa laborioso per matrici di grandi dimensioni, con un vettore θ grande.

Cosicchè un'alternativa ai metodo iterativi, è l'uso dell'Algoritmo EM. Esso esprime una relazione tra le SMV ricavate da $l(\theta|Y_{obs})$ e da $l(\theta|Y)$, quindi tra la Log-Verosomiglianza dei dati osservati e la Log-Verosomiglianza dei dati completi. In questo modo massimizzeremo la $l(\theta|Y)$, che ha una forma molto più trattabile rispetto $l(\theta|Y_{obs})$.

L'Algoritmo EM si divide in 4 fasi:

1. Sostituire i valori mancanti con dei valori stimati
2. Stimare i Parametri
3. Ristimare i dati mancanti, assumendo che le nuove stime dei parametri siano corrette
4. Ristimare i parametri, ripetendo la procedura fino alla convergenza

Passo E: Va a trovare i valori attesi condizionati dei dati mancanti, dati i valori osservati e le correnti stime dei parametri. In seguito i valori mancanti (o meglio le funzioni di Y_{mis} che compaiono nella verosomiglianza dei dati completi) vanno sostituiti con questi valori attesi condizionati.

Passo M: Calcolo la SMV di θ come se i dati mancanti non esistessero, fa un calcolo sui dati completati. Quindi il passo M usa gli stessi metodi computazionali che si usano per trovare la SMV per i dati completi.

Come detto in precedenza, l'Algoritmo è applicabile ad una vasta classe di modelli, in particolare è utile quando i dati completi provengono da una famiglia

esponenziale. In questo caso il passo E, si basa sul calcolo del valore atteso condizionato delle statistiche sufficienti per i dati completi e il passo M che segue è molto semplice (Giraldo, B [3](2004)).

3.2.1 Pregi e Difetti

Soffermiamoci ora su una lista di pregi e difetti dell'algoritmo EM.

I **Pregi** sono principalmente i seguenti:

- Facile da implementare. Il Passo E e M sono basati su calcoli fatti su dati completi.
- Non prevede calcolo e inversione di matrici di informazione.
- Numericamente stabile, ogni iterazione aumenta la Log-Verosomiglianza.
- Esplicita la natura del problema dei dati mancanti. Il passo E è un passo di imputazione e completa i dati mancanti, mentre il passo M calcola SMV sui dati completi.
- Se $\theta^{(0)}$ giace all'interno dello spazio dei parametri allora anche tutti i valori iterati $\theta^{(1)} \dots \theta^{(t)}$ vi appartengono.

Tra i **Difetti** dell'algoritmo EM:

- Ricorrere all'Algoritmo SEM, al Metodo di Louis o al bootstrap per arrivare agli Errori Standard.
- Convien solo quando il passo E può essere calcolato direttamente (Es. Famiglie Esponenziali regolari), altrimenti bisogna iterare 2 volte.
- Convergenza ad una SMV (a un massimo globale) non è garantita.
- Tasso di convergenza lento soprattutto se ci sono molti dati mancanti.

3.2.2 Estensioni dell'EM

In questa sezione vedremo delle varianti dell'Expectation Maximization Algorithm che utilizzeremo nel nostro modello per i Darts. Illustreremo il caso dell'EM per le famiglie esponenziali, il Metodo Montecarlo e l'applicazione sulla Normale Multivariata.

EM per Famiglie Esponenziali

L'algoritmo diventa facile da interpretare quando i dati completi Y appartengono a una Famiglia Esponenziale regolare, vale dire nella forma:

$$f(Y|\theta) = b(Y) \cdot \exp(s(Y) \cdot \theta) \cdot \frac{1}{a(\theta)}$$

dove:

θ è un vettore di parametri ($d \times 1$)

$s(Y)$ è un vettore di statistiche sufficienti per dati completi ($d \times 1$)

a e b sono funzioni rispettivamente di θ e Y

Passo E: Stima la statistica sufficiente per i dati completi $s(Y)$ con

$$s^{(t+1)} = E[s(Y)|Y_{obs}, \theta^{(t)}]$$

Passo M: determina la nuova stima $\theta^{(t+1)}$ di θ risolvendo le equazioni di verosomiglianza:

$$E[s(Y)|\theta] = s^{(t)}$$

che equivalgono alle equazioni di verosomiglianza per di dati completi con $s(Y)$ sostituito da $s^{(t)}$. Queste equazioni possono spesso essere risolte risolte per θ esplicitamente. Il problema computazionale è effettivamente confinato al passo E, alla stima della statistica $s(Y)$.

Nelle famiglie esponenziali, la decomposizione dell'informazione osservata è la differenza tra la varianza non condizionata e varianza condizionata alla statistica sufficiente(dati osservati). Il rapporto tra le due quantità determina il tasso di convergenza(Giraldo, B [3](2004)).

$$I(\theta|Y_{obs}) = \text{var}(s(Y)|\theta) - \text{var}(s(Y)|Y_{obs}; \theta)$$

Applicazione dell'EM nella Normale Multivariata

In questa sezione, mostreremo l'uso dell'Algoritmo EM applicato alla distribuzione Normale. Abbiamo un campione di variabili casuali indipendenti e identicamente distribuite di numerosità n distribuite $Y_i \sim N(\mu, \sigma^2)$. La Verosomiglianza dei dati completi sarà:

$$L(\mu, \sigma^2 | y) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{y_i - \mu}{\sigma}\right)^2\right)$$

dove la log-verosomiglianza sarà:

$$l(\mu, \sigma^2 | y) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2$$

Massimizzando rispetto a μ e σ^2 e ponendo le equazioni di verosomiglianza a zero, otteniamo le SMV dei parametri:

$$\hat{\mu}^2 = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2$$

Da qui supponiamo di lavorare con $n - m$ valori mancanti, il campione si può scrivere come $Y = (Y_{obs}, Y_{mis})$ dove Y_{obs} sono gli m valori osservati, mentre Y_{mis} sono gli $n - m$ valori mancanti. Supponiamo che siamo in presenza di dati MAR.

$$f(Y|\theta) = \prod_{i=1}^n f(Y_i|\theta) = \prod_{i=1}^m f(Y_i|\theta) \cdot \prod_{i=m+1}^n f(Y_i|\theta) = f(Y_{obs}|\theta) \cdot f(Y_{mis}|\theta)$$

Massimizzando $L(\mu, \sigma^2 | Y_{obs})$ rispetto a θ porta alle SMV:

$$\hat{\mu} = \frac{1}{m} \sum_{i=1}^m y_i$$

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{\mu})^2$$

Iniziamo a studiare i passi dell'algoritmo EM in questo esempio. La log-verosomiglianza dei dati completi è lineare nelle statistiche sufficienti, che in una funzione Normale sono: $(\sum_{i=1}^n y_i ; \sum_{i=1}^n y_i^2)$.

Quindi il **Passo E** dell'Algoritmo:

$$E \left[\sum_{i=1}^n y_i | \theta^{(t)}, Y_{obs} \right] = \sum_{i=1}^m y_i + (n - m) \cdot \mu^{(t)}$$

$$E \left[\sum_{i=1}^n y_i^2 | \theta^{(t)}, Y_{obs} \right] = \sum_{i=1}^n y_i^2 + (n - m) \cdot (\mu^{(t)2} + \sigma^{(t)2})$$

Il **Passo M** usa le stesse espressioni delle SMV per i dati completi con i valori correnti delle statistiche sufficienti calcolati al passo E, al posto delle statistiche sufficienti non osservate:

$$\mu^{(t+1)} = \frac{E[\sum_{i=1}^n y_i | \theta^{(t)}, Y_{obs}]}{n}$$

$$\sigma^{2(t+1)} = \frac{E[\sum_{i=1}^n y_i^2 | \theta^{(t)}, Y_{obs}]}{n} - (\mu^{(t+1)})^2$$

Poniamo ora $\mu^{(t)} = \mu^{(t+1)} = \hat{\mu}$ e $\sigma^{(t)2} = \sigma^{(t+1)2} = \hat{\sigma}^2$ nelle equazioni sopra si mostra che queste iterazioni convergono a:

$$\hat{\mu} = \frac{1}{m} \sum_{i=1}^m y_i$$

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{\mu})^2$$

Queste sono le SMV dei parametri calcolate direttamente massimizzando la log-verosomiglianza dei dati osservati (Giraldo, B [3](2004)).

MCEM

L'Algoritmo Montecarlo EM è uno dei metodi che possono essere intrapresi quando il passo E è difficile da calcolare. Data la stima corrente $\theta^{(t)}$ il passo E richiede il calcolo di:

$$Q(\theta | \theta^{(t)}) = \int l(\theta | Y) f(Y_{mis} | Y_{obs}; \theta = \theta^{(t)}) dY_{mis}$$

Come visto prima, in caso di famiglie esponenziali, il calcolo di questo integrale si riduce al calcolo del valore medio condizionato alle statistiche sufficienti, mentre per gli altri casi tale calcolo può risultare laborioso. Con l'applicazione del Metodo Montecarlo, facilitiamo il **Passo E**:

1. estrai $Y_{1,mis}, \dots, Y_{m,mis} (i.i.d.) \sim f(Y_{mis} | Y_{obs}; \theta = \theta^{(t)})$

2. poni $\widehat{Q}(\theta | \theta^{(t)}) = \frac{1}{m} \sum_{j=1}^m l(\theta^{(t)} | Y_{obs}, Y_{j,mis})$

Questo approccio è un'imputazione multipla di Y_{mis} ad ogni passo E. Queste imputazioni, nel **Passo M**, vengono sintetizzate in una stima di $\widehat{Q}(\theta | \theta^{(t)})$, massimizzata successivamente ottenendo θ^{t+1} .

L'Algoritmo MCEM mostra purtroppo anche dei limiti:

- L'approssimazione $\widehat{Q}(\theta|\theta^{(t)})$ non garantisce più la convergenza monotona dell'EM.
- Estrarre i valori dalla distribuzione $f(Y_{mis}|Y_{obs}; \theta = \theta^{(t)})$ può talvolta essere un problema di difficile soluzione (Giraldo, B [3](2004)).

3.3 Stima della varianza di tiro di un giocatore

Il punto ottimale di tiro $\mu^*(\sigma)$ dipende fortemente da σ , da cui costruiremo la nostra dartboard. Per ricavare il valore della varianza, il metodo di tirare N frecce verso il centro del bersaglio e misurare la distanza, dal centro al dart può richiedere troppo tempo ed errori di misurazione. Quindi la tecnica che useremo per ricavarci la nostra varianza di tiro sarà basata sui punteggi e non sulla posizione delle frecce sul bersaglio. La cosa potrebbe sorprendere, poichè vediamo solo una piccola informazione nella stima di σ^2 avendo solo lo score. Cosicché sfrutteremo le aree dei bulleyes, dei tripli e dei doppi, utilizzando dei valori unici per ogni area, per aiutarci a una più facile raccolta dei dati.

Denotiamo con Z la variabile relativa alla posizione della freccia sul bersaglio e con X lo score che otteniamo. Svilupperemo il contesto utilizzando l'EM algorithm, utilizzato per problemi di massimizzazione di verosomiglianze semplici (con variabile di riferimento non osservata) e studi di variabili non osservate (Z) implementate su osservazioni di altre (X). Nel nostro caso il dato osservato saranno gli scores $X = (X_1, \dots, X_n)$ per un giocatore che lancia n freccette verso il centro ($\mu = 0$). Mentre il dato non osservato sarà la posizione $Z = (Z_1, \dots, Z_n)$ dove la freccia plana (TIBSHIRANI, PRICE, TAYLOR; B [5](2011)).

Siamo in presenza di un caso OAR, la probabilità che un dato sia osservato non dipende dal valore dei dati osservati. I nostri dati presi in esame sono gli score, essi non dipendono da loro stessi, ma da in che punto del segnapunti si trova la freccia. Per esempio per fare 20 punti con un tiro possiamo colpire i 2 spicchi del 20 e lo spicchio esterno del doppio 10, quindi gli score X sono condizionati alla posizione Z .

La nostra logverosomiglianza sarà $l(\sigma^2; X, Z)$. L'EM algorithm è focalizzato sulla stima di σ^2 , che inizia stimando σ_0^2 e ripetendo le due fasi fino alla convergenza.

$$\begin{cases} X \text{ Score} \\ Z \text{ Posizione} \end{cases} \begin{cases} x\text{-componente orizzontale sul bersaglio} \\ y\text{-componente verticale sul bersaglio} \end{cases}$$

Partendo dalla Verosomiglianza:

$$L(\mu, \sigma^2; X, Z) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(Z_i - \mu)^2\right\} \quad (3.1)$$

E-Step: calcolo $Q(\sigma^2) = E_{\sigma_t^2}[l(\sigma^2; X, Z)|X]$

la stima della fase E è data da:

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

Andiamo a trovare il valore atteso $E_{\sigma_0^2}(Z_{i,x}^2 + Z_{i,y}^2|X_i)$, dove $Z_i \sim N(0, \sigma_0^2)$ e $X_i = S(Z_i)$ denota gli score. Possiamo definire gli score come la realizzazione della freccia in $U_j A_j$, dove ogni regione A_j può essere espressa come $[r_{j,1}, r_{j,2}] \times [\theta_{j,1}, \theta_{j,2}]$ in coordinate polari. Come detto prima, la realizzazione di uno score può essere fatta in diversi modi e in diverse regioni, come l'esempio fatto in precedenza, su come ottenere 20 punti con un tiro.

$$\begin{aligned} E(Z_{i,x}^2 + Z_{i,y}^2|X_i) &= E(Z_{i,x}^2 + Z_{i,y}^2|Z_i \in U_j A_j) \\ &= \frac{\sum_j \int \int_{A_j} (x^2 + y^2) e^{-(x^2+y^2)/2\sigma_0^2} dx dy}{\sum_j \int \int_{A_j} e^{-(x^2+y^2)/2\sigma_0^2} dx dy} \\ &= \frac{\sum_j \int_{r_{j,1}}^{r_{j,2}} \int_{\theta_{j,1}}^{\theta_{j,2}} r^3 e^{-r^2/2\sigma_0^2} d\theta dr}{\sum_j \int_{r_{j,1}}^{r_{j,2}} \int_{\theta_{j,1}}^{\theta_{j,2}} r e^{-r^2/2\sigma_0^2} d\theta dr} \end{aligned} \quad (3.2)$$

Nell'ultimo step abbiamo usato un cambio di variabili, utilizzando le coordinate polari. Integrando su θ abbiamo 2 possibili soluzioni:

$$\theta_{j,2} - \theta_{j,1} = \begin{cases} 2\pi & \text{se } X_i = 25 \text{ o } 50 \\ \frac{\pi}{10} & \text{altro} \end{cases}$$

Usando l'integrazione per parti il risultato dell'integrale sarà:

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

M-Step: $\sigma_{t+1}^2 = \operatorname{argmax}_{\sigma^2} Q(\sigma^2)$

Con $\mu = 0$ la verosomiglianza è data da:

$$l(\sigma^2, X, Z) = \begin{cases} -n \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Z_{i,x}^2 + Z_{i,y}^2), & \text{se } X_i = S(Z_i) \forall i \\ -\infty & \text{altro} \end{cases}$$

Per semplicità chiameremo $C = E_{\sigma_0^2}[l(\sigma^2; X, Z)|X]$ che non dipende da σ^2 . Il nostro obiettivo è quello di scegliere la nostra varianza da massimizzare, quindi:

$$l^*(\sigma^2; X, Z)|X) = -n \log \sigma^2 - \frac{C}{2\sigma^2}$$

da qui ricaveremo: $\hat{\sigma}^2 = \frac{C}{2n}$

L'algoritmo EM ci da una buona stima di σ^2 per n molto grande. Possiamo anche calcolare la varianza tramite il procedimento di massima verosomiglianza e ci da:

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

Dall'immagine vediamo come le due stime sono molto simili tra di loro.

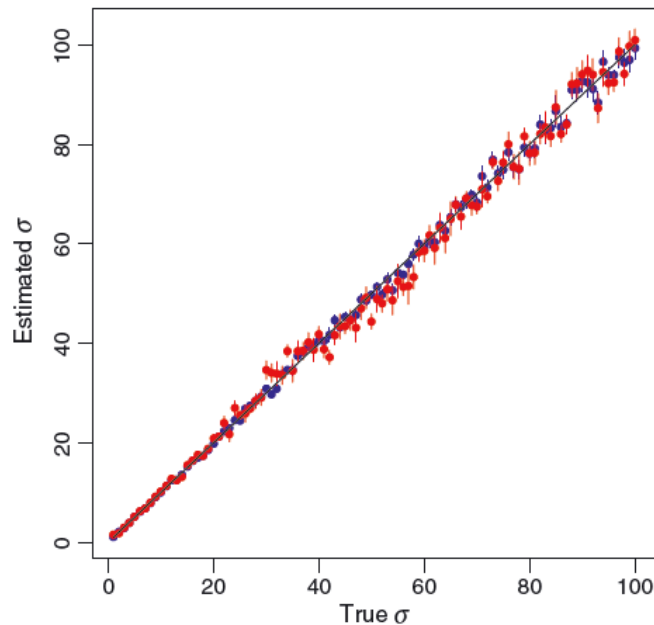


Fig.3.1b Maximum likelihood (●) and EM estimate (●), from $n = 50$ points drawn independently from $\mathcal{N}(0, \sigma^2 I)$, and σ ranging over $1, 2, \dots, 100$: for each σ we repeated this 10 times; shown are the mean ± 1 standard error over these trials

Simple EM e Heatmaps

Dopo aver visto la parte teorica riguardo l'uso di questo algoritmo e le varie proprietà in se, andiamo ad applicare degli esempi pratici per capire meglio il modello. Tramite un Package di R di Ryan Tibshirani (si veda S [10]), vediamo lo studio delle heatmap nella dartboard e dove un giocatore mira per massimizzare il proprio score.

Le Mappe di calore sono rappresentazioni grafiche di dati, i quali vengono suddivisi in una matrice rappresentata da colori. Questi spesso vanno dal freddo al caldo, più i valori sono elevati più la heatmap avrà un colore caldo, o viceversa.

Prendiamo in esame un giocatore di medio livello, il quale tira 100 frecette cercando di fare centro sul bulleye, studiamo poi il suo sigma e successivamente come massimizzare il suo punteggio.

```
x = c(12,16,19,3,17,1,25,19,17,50,18,1,3,17,2,2,13,18,16,
      2,25,5,5,1,5,4,17,25,25,50,3,7,17,17,3,3,3,7,11,10,25,
      1,19,15,4,1,5,12,17,16,50,20,20,20,25,50,2,17,3,20,
      20,20,5,1,18,15,2,3,25,12,9,3,3,19,16,20,5,5,1,4,15,
      16,5,20,16,2,25,6,12,25,11,25,7,2,5,19,17,17,2,12)
```

In 100 tiri va a colpire per 4 volte il bulleye rosso da 50 punti e per 10 volte l'anello verde del 25. Da qui applichiamo l'algoritmo EM andando a stimare la nostra varianza di tiro:

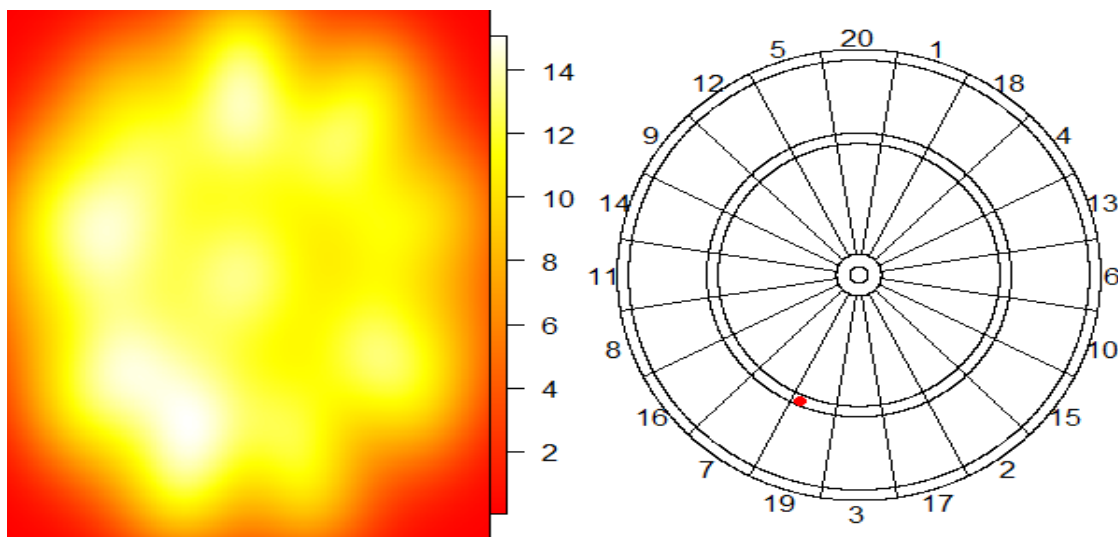


Figura 3.2: Simple EM Heatmap per un giocatore di livello medio

```
> a = simpleEM(x, niter=100)
> s = a$s.final
> s
[1] 799.028
```

σ^2 è pari a 799.028 corrisponde a una deviazione standard $\sigma=28.27$ mm. Vediamo come si distribuiscono gli score sul bersaglio ricavandoci l'heatmap:

```
> par(mfrow=c(1,2))
> drawHeatmap(e)
> drawBoard(new=TRUE)
> drawAimSpot(e, col = "red")
```

Possiamo vedere dalla heatmap come questo giocatore di livello medio riesca a raggiungere score abbastanza elevati nella parte sud-ovest del bersaglio, dove il colore risulta di un giallo molto acceso. La sua deviazione standard di 28.27 mm risulta troppo alta e andar a cercare il T20 per migliorare il proprio score porterebbe più danni che benefici. La figura di destra infatti consiglia il T19 come zona su cui il giocatore deve concentrarsi per massimizzare il proprio punteggio.

Andamento Heatmaps per σ che tende all'infinito

Come anticipato all'inizio capitolo, per un principante è conveniente puntare al Triplo 20?

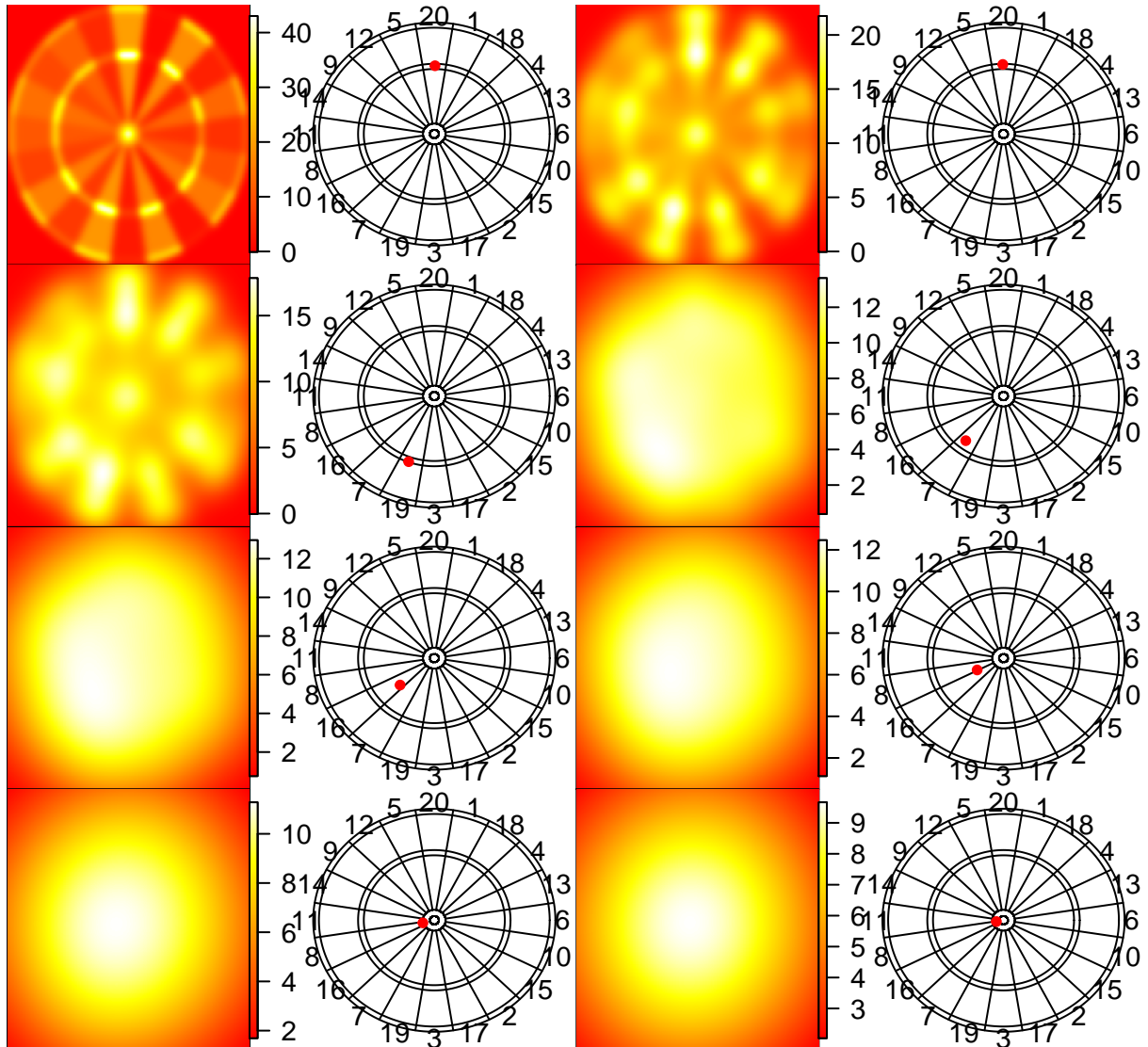


Figura 3.2a Punto ottimo di tiro e Heatmaps, relative al $\sigma=s$, partendo da $s=5$ mm(alto a sinistra), aumentando fino a $s=100$ mm.

La risposta come si può vedere dall'immagine seguente è no. Il suo scopo è quello di far più punti possibili, ma ciò dipende dalla sua abilità cioè in che zona della tavola puntare. Vediamo che un giocatore con un $\sigma=5=s$ mm non ha nessun problema a tirare verso il T20 e lì dovrà mirare per usare meno turni possibili

per chiudere il suo leg. Man mano che la varianza di tiro aumenta, i fuochi gialli delle heatmap s'ingrandiscono e cambiano le zone dove tirare il dart. Un giocatore con una varianza di tiro più ampia ($\sigma=20$ mm) dovrà mirare verso il T19, se non vuole perdere troppi punti fra l'S1 e l'S5. Da qui per $s \rightarrow \infty$ la zona di ottimizzazione risale dall'S19 verso destra, passando per lo spicchio inferiore dell'S11 e terminando nel bullseye.

Dunque se siete alle prime armi con il mondo delle freccette e vi imbattete nello sfidare un vostro amico, il punto dove si deve mirare non sarà il T20 ma esattamente il centro rosso del bersaglio.

3.4 Un modello Gaussiano più generale

In questa sezione svilupperemo un più generale modello gaussiano con errori di tiro distribuiti come:

$$\varepsilon \sim N(0, \Sigma)$$

avendo una matrice di covarianza Σ . Questa aggiunta è molto importante, un giocatore di freccette ha una distribuzione di tiro che potrebbe non essere simmetricamente circolare. E' molto comune per molti giocatori avere una varianza orizzontalmente piccola, ampia verticalmente. In poche parole per un professionista che mira verso il T20 è molto comune un errore di tiro che vada negli S20, invece che negli spicchi a destra e sinistra dell'1 e del 5. Altra cosa, un destro ha diversi andamenti degli errori di distribuzione rispetto a un mancino.

In questa sezione, faremo un'analisi simile a prima: una stima dei parametri usando l'EM algoritmo e successivamente una heatmap sulla stima degli score (TIBSHIRANI, PRICE, TAYLOR; B [5] (2011)).

3.4.1 Stima della Covarianza

Possiamo stimare Σ usando un modello simile al EM Algorithm visto in precedenza. Osserviamo gli score (X_1, \dots, X_n) dei tiri verso il centro del bersaglio, senza guardare le posizioni delle frecce (Z_1, \dots, Z_n) . Come $\mu = 0$, la Log-Verosomiglianza sarà:

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

con $X_i = s(Z_i)$ per tutti gli i . Semplificando:

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

Possiamo ora calcolare:

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

Massimizzando su Σ è un problema identico alla massimizzazione della verosomiglianza per la gaussiana multivariata con parametro di covarianza sconosciuto. Da qui la massimizzazione della verosomiglianza porta a:

$$\Sigma = \frac{1}{n} \sum_{i=1}^n E_{\Sigma_0}(Z_i^T Z_i | X_i)$$

La stima non può essere fatta in forma chiusa come nel caso della normale multivariata. Da qui useremo un utile campionamento: la tecnica Monte Carlo (si veda MCEM, pag. 35), per trovare un'approssimazione della stima difficile da calcolare:

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

Dove p è la nostra densità $Z_i|X_i$. Non è facile rappresentare questa distribuzione e fare una stima con il Metodo Montecarlo. L'idea è quella di replicare il campione con valori q vicini a p e più facili da stimare. Purchè $p = 0$ quando $q = 0$, possiamo scrivere:

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

dove $w = \frac{p}{q}$. Definiamo le variabili Z_1, \dots, Z_m da q e stimiamo le quantità sopra indicate:

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

o se la densità contiene delle costanti:

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

Il punto ottimale di q misurato in termini stima della varianza è proporzionale a $x^2 \cdot p(x, y)$. Nel nostro caso sceglieremo q distribuito come un'Uniforme sulla regione della dartboard definita dallo score X_i , poichè queste distribuzioni risultano più facili da campionare. I pesi, come si può vedere, sono $w(x, y) = f_{\Sigma_0}(x, y)$, dove $f_{\Sigma_0}(x, y)$ è la densità di una Normale bivariata con matrice di covarianza Σ_0 (TIBSHIRANI, PRICE, TAYLOR; B [5](2011)).

3.4.2 Calcolo della Heatmap

Dopo aver stimato la Σ del giocatore, possiamo procedere con la costruzione della Heatmap. La stima dello score del giocatore che mira verso la posizione μ è:

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

Approssimando la stima di f_Σ e s sulla griglia e prendendo la convoluzione di queste 2 matrici, possiamo velocemente calcolare le covarianze usando due trasformate di Fourier (FFTs) e un'inversa di Fourier (FFT). Facciamo un esempio su 100 tiri effettuati, le matrici di covarianze si riferiscono ai punteggi dei due autori del Paper "A Statistician plays Darts".

$$\Sigma_1 = \begin{bmatrix} 1820.6 & -471.1 \\ -471.1 & 470.2 \end{bmatrix}$$

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

Il nuovo modello risulta più flessibile del precedente e porta a interessanti risultati. Consideriamo l'autore 2, dove nel modello gaussiano gli veniva indicato il T19 come area ottimale, con una deviazione marginale standard $\sigma = 26$.

Sotto il modello gaussiano generale, dopo aver stimato la deviazione in x e y corrispondenti a $\sigma_x = 17.9$ e $\sigma_y = 39.1$, la general heatmap gli mostra invece di puntare il dart attorno al T20. Questo implica che il giocatore 2, ha una migliore precisione in direzione orizzontale e in questo modo i suoi darts puntando verso il T20, raramente andranno nelle sezione dell'1 e del 5.

Esempio di Heatmap di un giocatore di medio livello

Analizziamo ora l'esempio visto nella sezione precedente, di un tiratore che esegue 100 tiri verso il centro del bersaglio, studiando però ora il modello più esteso.

```
> x = c(12,16,19,3,17,1,25,19,17,50,18,1,3,17,2,2,13,18,16,
        2,25,5,5,1,5,4,17,25,25,50,3,7,17,17,3,3,3,7,11,10,25,
        1,19,15,4,1,5,12,17,16,50,20,20,20,25,50,2,17,3,20,
        20,20,5,1,18,15,2,3,25,12,9,3,3,19,16,20,5,5,1,4,15,
        16,5,20,16,2,25,6,12,25,11,25,7,2,5,19,17,17,2,12)
> a = generaleEM(x, niter=100, seed=0)
> Sig = a$Sig.final
> Sig
[1] 333.6949 1674.0911 -151.5163
```

Dalla matrice di varianza e covarianza, possiamo vedere come il giocatore in questione abbia una varianza di tiro non di certo circolare, ma come previsto, con una forte componente verticale e leggermente orizzontale. La deviazione standard di x è $\sigma_x=18.2$ mm mentre quella di y raggiunge i 4 cm, $\sigma_y=40.9$ mm. Le variabili mostrano una covarianza negativa, la precisione di tiro verticale influenza negativamente la precisione orizzontale. Possiamo quindi prevedere che se un giocatore cercasse di migliorare la sua varianza di tiro lungo l'asse y , ne risentirebbe l'accuratezza sul piano x .

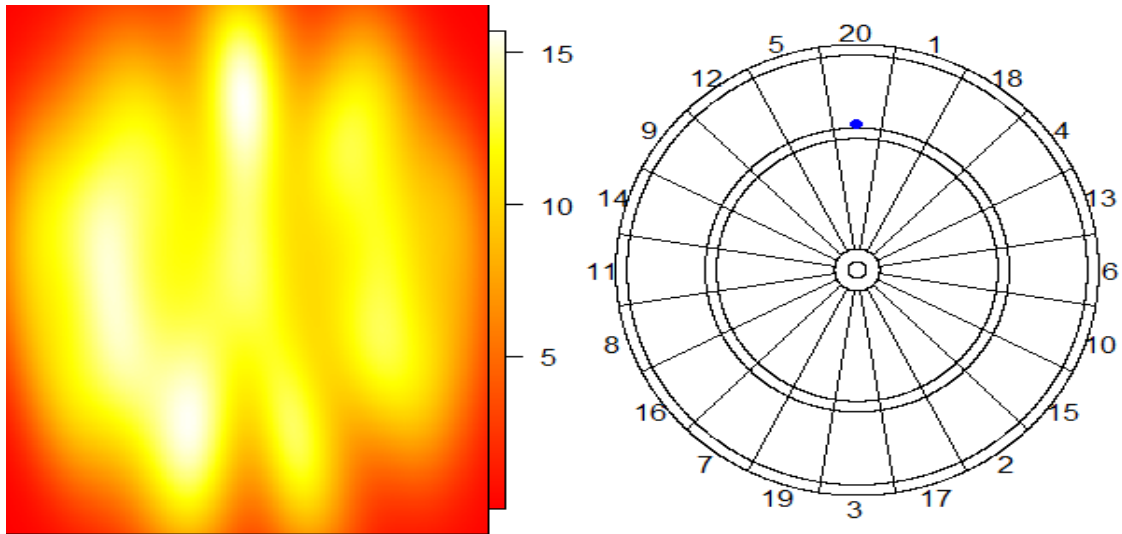


Figura 3.3: General EM Heatmap di un giocatore di livello medio

$$\Sigma = \begin{bmatrix} 333.69 & -151.52 \\ -151.52 & 1674.09 \end{bmatrix}$$

```
> e = generalExpScores(Sig=(c(333.69,1674.09,-151.52)))
> par(mfrow=c(1,2))
> drawHeatmap(e)
> drawBoard(new=TRUE)
> drawAimSpot(e, col="blue")
```

Dalla heatmap possiamo vedere come le zone in cui riesce a realizzare alti score sono attorno allo spicchio del venti e nelle zone in basso a sinistra tra l' S19 e l'S16. Data la sua buona precisione di tiro nella direzione verticale, gli viene indicato lo spicchio del 20, per massimizzare il proprio score.

L'Heatmap di Phil Taylor

Nei precedenti paragrafi, tramite l'algoritmo EM abbiamo capito parecchie cose sull'andamento di tiro nel bersaglio. Per esempio un giocatore che non abbia mai praticato questo sport se voglia minimizzare i propri leg, non deve puntare al T20, ma al centro; oppure che la varianza di tiro oscilla molto di più verticalmente piuttosto che orizzontalmente. Ma quanto vale l'accuratezza di un giocatore professionista?

Dato il modello generale gaussiano, che si basa su lanci verso il centro del bersaglio, non è stato facile raccogliere dei dati per studiare questo caso. I giocatori durante i match, lanciano la frecce mirando sempre al T20 o T19, a meno che

siano a -140 punti per chiudere il leg. Quindi tramite due eventi organizzati da sponsor, il "Betway Bullseye Challenge" e il "Bullseyes For Beer | Channelbee", presi da due video su YouTube(si veda S [1] [2]), abbiamo raccolto 18 tiri realizzati da Phil Taylor, verso il Bulleye e applicato su di essi l'algoritmo EM, per vedere la sua Heatmap:

```
> x = c(50,25,20,8,25,25,25,50,50,50,50,50,50,25,50,25,25)
> a = generalEM(x, niter=18, seed=0)
> Sig = a$Sig.final
> Sig
[1] 41.31716 49.48925 8.13580
```

$$\Sigma_{PT} = \begin{bmatrix} 41.32 & 8.14 \\ 8.14 & 49.49 \end{bmatrix}$$

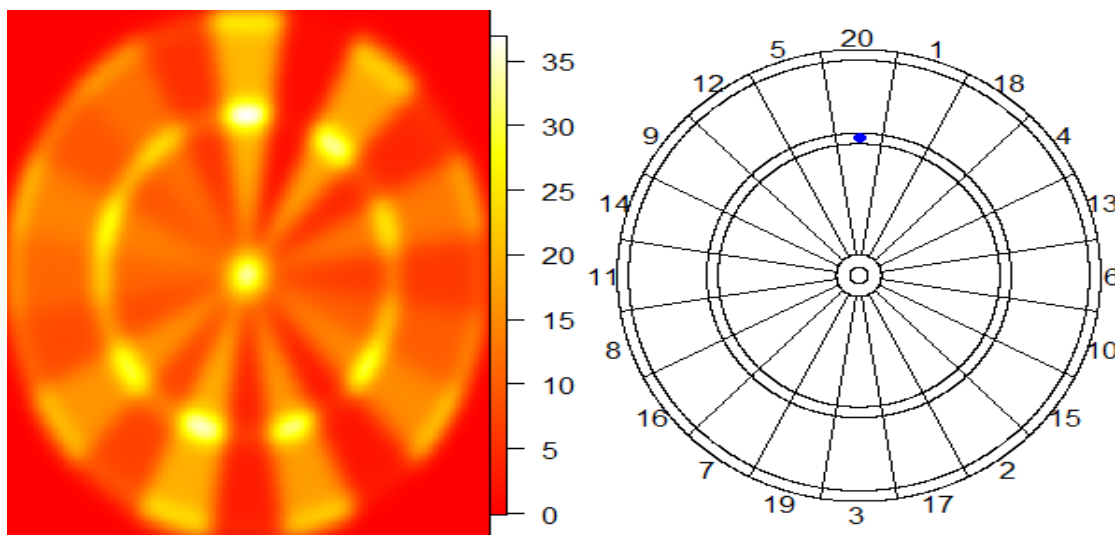


Figura 3.4: General EM Heatmap di Phil Taylor

La varianza di tiro di Taylor si dimostra esser stupefacente. Eseguendo 6 lanci verso il bersaglio, composti da 3 freccette per lancio, colpisce 9 volte l'anello centrale e 7 quello dello small bull. Vediamo come giocatori professionisti del suo calibro abbiano una precisione di tiro quasi circolare: $\sigma_x = 6.43$ mm e $\sigma_y = 7.03$ mm. Ecco perchè, sapendo che l'altezza di ogni triplo e doppio è di 8 mm, per un giocatore con un'accuratezza così elevata, risulta naturale mirare verso il T20 con lo scopo di massimizzare il proprio punteggio. In questo caso la covariata è addirittura negativa, x e y subiscono oscillazione concordi, una migliore precisione di Taylor dall'alto verso il basso, influirà positivamente sulla precisione di sinistra-destra.

Capitolo 4

Simulazioni e Analisi Empirica del modello dinamico

In questa sezione faremo uno studio empirico del modello dinamico, utilizzando degli script con matlab. Studieremo come si distribuisce il modello sul bersaglio data la velocità, l'ampiezza dell'angolo φ e dell'angolo θ . Lo studio sarà fatto simulando il lancio di 1000 frecce verso il bersaglio e variando i nostri tre input iniziali, vediamo come variano i nostri output grafici e i qqplot.

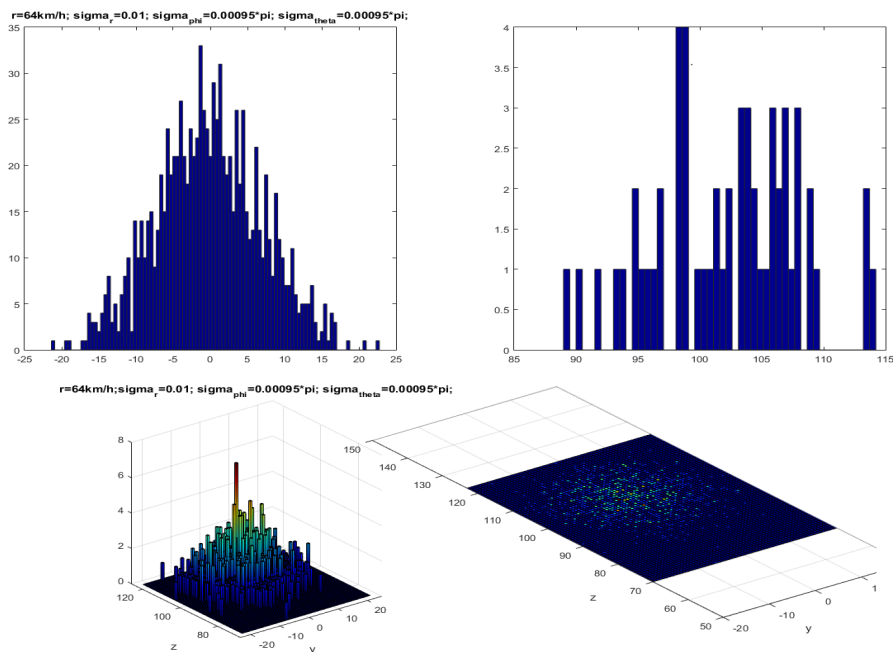


Figura 4.1: Distribuzioni di Y e X marginali e congiunta in 3D

La velocità media dei darts è di 64 km/h, il σ_r rappresenta una variazione percentuale della velocità media. Per esempio un $\sigma_r = 10\%$ significa che la nostra deviazione standard della velocità sarà di 6.4 km/h. Vedremo più avanti che questa componente influirà una delle due variabili (X,Y).

σ_φ invece rappresenta la deviazione standard relativa all'inclinazione verticale della freccia sulla mano. Il movimento effettuato per lanciare la freccia, ha un determinato angolo φ misurato in radianti. σ_φ è lo scarto quadratico medio, riferito all'angolo φ , cioè l'errore dell'angolo di lancio in direzione verticale.

σ_θ invece rappresenta la deviazione standard angolare in orizzontale. E' l'errore di tiro sulla mano in direzione "sinistra-destra", corrispondente all'angolo θ . Anche in questo caso viene rappresentata in radianti. Analizzeremo quindi quanto influisce e come si distribuisce la velocità sulla dartboard, le componenti φ e θ , vedremo l'heatmap costruita in base alla matrice degli score e faremo un confronto infine con le heatmaps dell'Algoritmo EM.

4.1 La Velocità r

La Velocità iniziale, con cui il giocatore lancia la freccia verso la dartboard, è segnata con V_0 , nel moto parabolico che abbiamo studiato nel capitolo 2, tramite le coordinate sferiche, l'abbiamo chiamata r e rappresenta la distanza dal polo, agendo su gli angoli φ e θ . E' una grandezza vettoriale, specificata da modulo, direzione e verso, ed è il rapporto tra lo spostamento percorso in una certa direzione e l'intervallo di tempo impiegato.

Quanto e come influisce la componente r sulla distribuzione dei tiri?

Come ben sappiamo, una pallina lanciata contro un muro a una certa velocità v colpirà un determinato punto. Se lanciamo la pallina con una forza maggiore, colpirà un punto più alto, mentre con una forza minore, andrà in un punto più basso. Il meccanismo della freccia è esattamente lo stesso: più ampio sarà il nostro sigma σ_r e più la distribuzione dei tiri sarà in verticale, viceversa se meno ampio. La componente r quindi possiamo definirla una componente aggiuntiva all'angolo φ , o meglio va ad influire su Y . Dalla nostra ricerca non siamo riusciti a risalire al range di velocità che hanno i giocatori professionisti, pur sapendo che la loro media di tiro è di 64 km/h, non siamo riusciti però a stabilire di quanto varia il loro σ_r . Negli studi successivi, relativi ad angoli e heatmaps, abbiamo fissato un σ_r pari a 1 %. Dalla formula:

$$Y = \frac{d \tan \varphi}{\cos \theta} - \frac{g d^2}{2 r^2 \cos^2 \varphi \cos^2 \theta}$$

vediamo come la componente r è in relazione con Y .

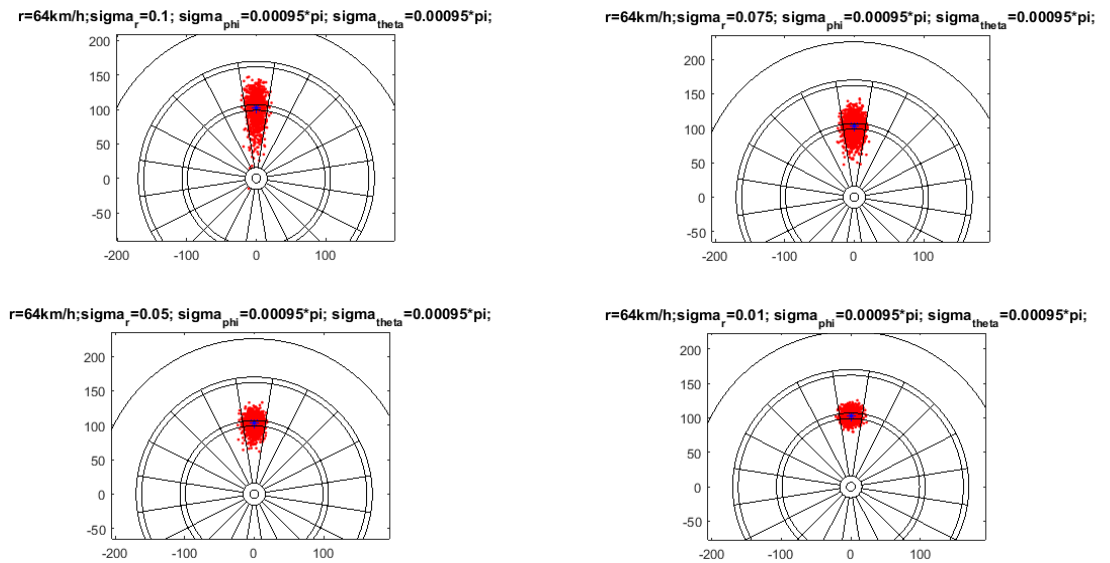


Figura 4.2: La distribuzione di 1000 darts verso il bersaglio, dati $\sigma_\theta = \sigma_\varphi = 0.17^\circ$ al variare di σ_r .

Analizziamo ora graficamente le componente σ_r , tenendo fissati i σ_θ e σ_φ . Un giocatore professionista, ha uno score medio sulle tre frecce di 110 punti a turno, quindi ha circa una media di 36.66 punti/tiro mirando verso il T20. Per ottenere una media così alta, le sue deviazioni standard angolari devono essere molto basse. Noi con lo scopo di studiare r le teniamo fisse e uguali, $\sigma_\varphi = \sigma_\theta \simeq \frac{\pi}{1053} \simeq 0.00095\pi \simeq 0.17^\circ$. Dalla Figura 4.2 vediamo graficamente quello che abbiamo detto poco sopra. Partendo da un $\sigma_r = 10\%$ vediamo come la componente r si sviluppi perpendicolare all'asse y , tendendo ad essere un'aggiunta di Y . Il sigma del primo quadrante è di 6.4 km/h man mano che diminuiamo il suo valore, la componente velocità va ad annullarsi ed a essere quasi ininfluenza.

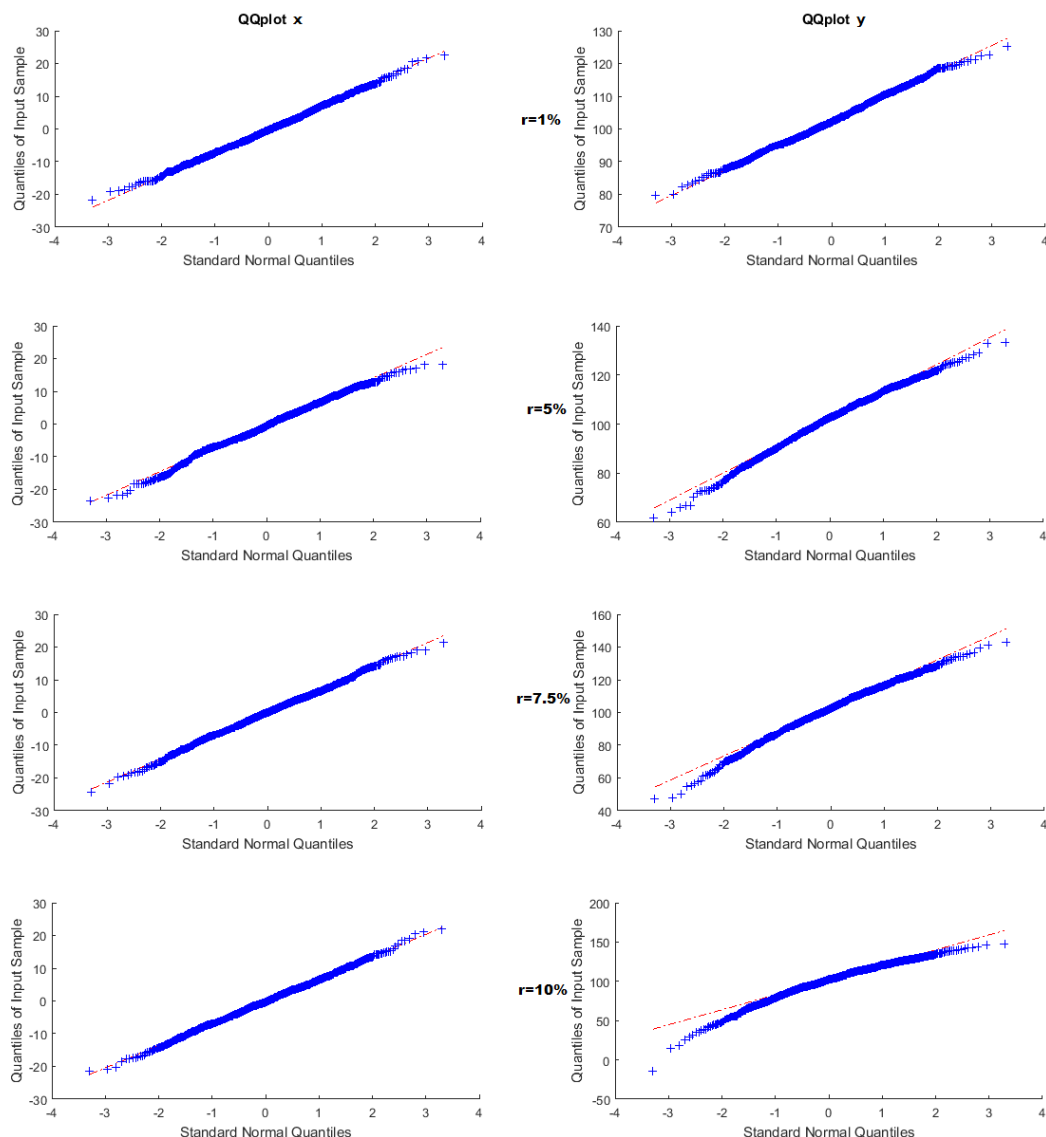


Figura 4.2a QQplots di X e Y, messi a confronto, in relazione alla velocità.

La figura invece mostra la rappresentazione grafica dei quantili della distribuzione, qqplot che confrontano la distribuzione cumulata della variabile osservata con la distribuzione cumulata della normale. Notiamo come aumentando σ_r %, i qqplot relativi all'asse Y perdino in normalità, subendo degli scostamenti sulle code e prendendo una pendenza negativa della curva. L'asse X mantiene invece sempre una buona normalità, non dipendendo dalla velocità.

4.2 Gli Angoli φ e θ

Dopo aver studiato la componente della velocità r , ora soffermiamoci su come varia la distribuzione dei tiri sul bersaglio, in base alle varianze dei due angoli θ e φ . Ogni giocatore tira con un determinato angolo orizzontale θ e verticale φ , verso il bersaglio, mirando il T20. L'angolo di lancio per realizzare i 60 punti con una sola freccia è di $\varphi \simeq 5^\circ$, mentre l'angolo θ sarà a 0° . Vediamo il tutto nello specifico.

$$\theta = \arctan\left(\frac{x}{d}\right) = \arctan\left(\frac{0}{d}\right) = 0^\circ$$

Sapendo che:

$$\begin{cases} \theta = 0 \rightarrow \cos\theta = 1 \\ Y = \frac{d \sin\varphi}{\cos\varphi \cos\theta} - \frac{g d^2}{2 r^2 \cos^2\varphi \cos^2\theta} = d \tan\varphi - \frac{g d^2}{2 r^2 \cos^2\varphi} \end{cases}$$

Da qui possiamo ricavarci l'angolo di tiro φ :

$$\min \left\{ \varphi = \frac{r^2 \pm \sqrt{r^4 - 2 \cdot g y r^2 - g^2 d^2}}{g \cdot d} \right\}$$

Dei quali verrà preso il minimo tra i 2 e i nostri valori saranno:

$$r = 64 \text{ km/h} = 17777 \text{ mm/s}$$

$$g = 9.8 \text{ m/s}^2 = 9810 \text{ mm/s}^2$$

$$y = 103 \text{ mm (distanza dal centro al T20)}$$

$$d = 2.3 \text{ m} = 2370 \text{ mm}$$

Sostituendo otterremo un angolo di tiro φ pari a 4.72° . Variando pian piano i nostri σ e tenendo costante r , quanto e come cambierà la distribuzione sul bersaglio? Un giocatore professionista ha un σ di tiro molto basso, intorno ai 0.17° , vediamo ampliando questi gradi come si amplieranno i tiri sulla dartboard. Facciamo questo test, sempre su 1000 tiri e aumentando gli angoli con delle soglie di 0.25° . Quindi analizzeremo le soglie di $0.17^\circ, 0.25^\circ, 0.5^\circ, 0.75^\circ, 1^\circ$:

$$\begin{cases} 0.17^\circ = 0.00095\pi = \frac{\pi}{1052} \\ 0.25^\circ = 0.0013\pi = \frac{\pi}{769} \\ 0.5^\circ = 0.0027\pi = \frac{\pi}{370} \\ 0.75^\circ = 0.0041\pi = \frac{\pi}{244} \\ 1^\circ = 0.0055\pi = \frac{\pi}{180} \end{cases}$$

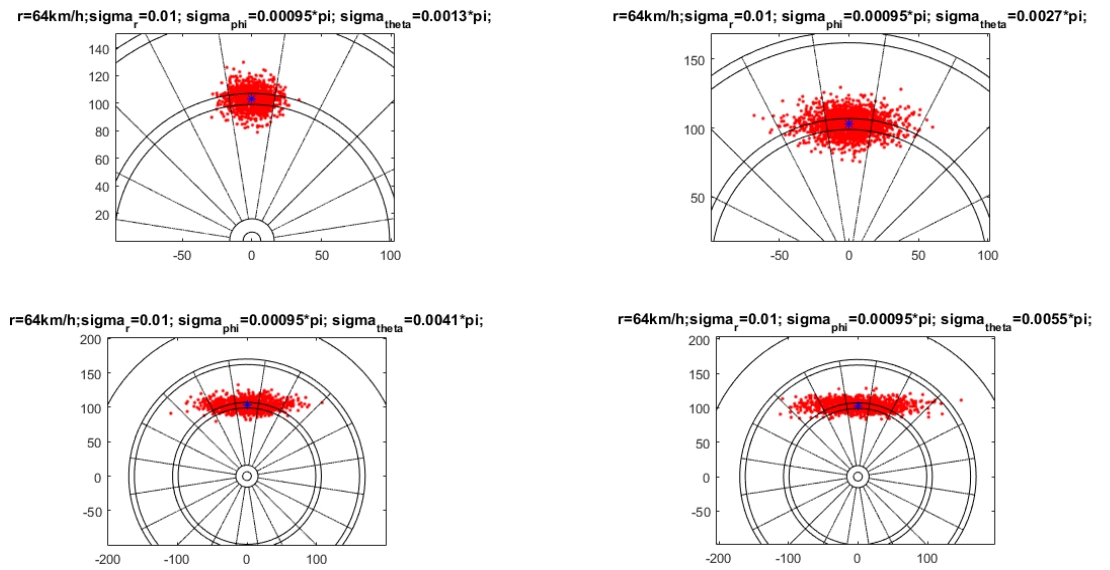


Figura 4.3: Distribuzione di 1000 darts verso il bersaglio, dati $\sigma_r = 1\%$ e $\sigma_\varphi = 0.17^\circ$, al variare di σ_θ .

Nel primo caso teniamo l'angolo φ costante ad un livello pro di 17° e andiamo ad aumentare man man di grado θ . Vediamo dalla Figura 4.3, come i punti sul bersaglio si estendono in orizzontale all'aumentare della devianza standardizzata di tiro. Tramite una funzione che stima il nostro score, si ha una buona media di tiro nel primo quadrante che è di ≈ 34 punti/tiro, mentre nel quadrante in basso a destra, con un $\sigma_\theta=1^\circ$, si ha uno score medio mediocre di 17 punti/tiro.

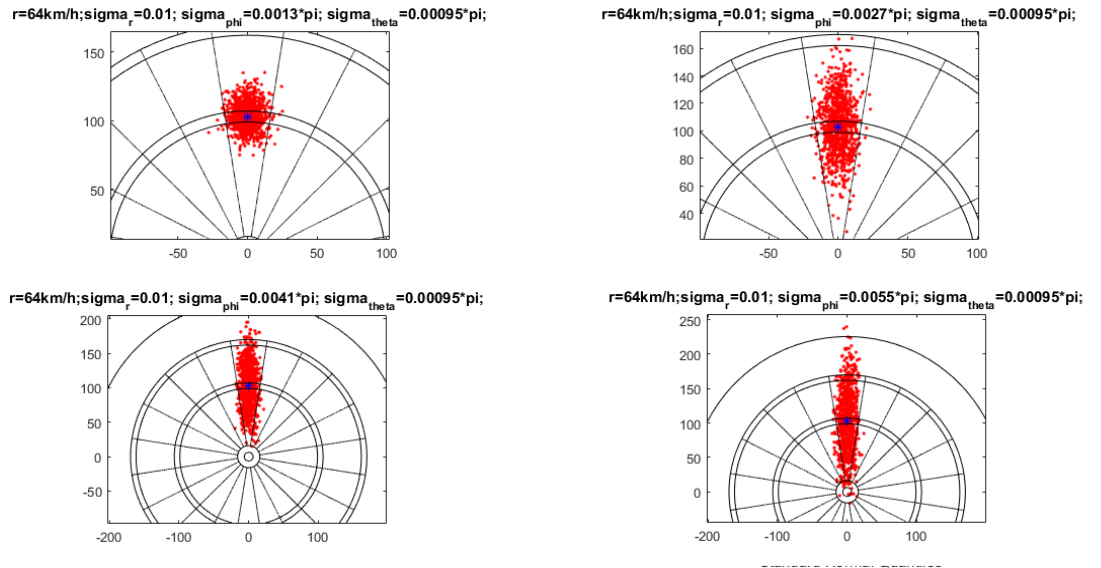
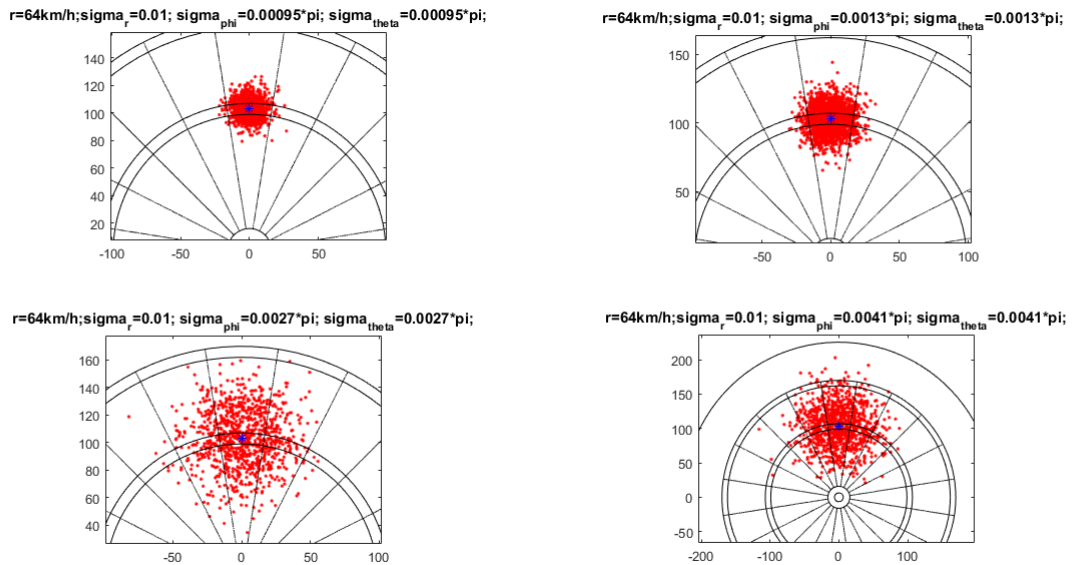


Figura 4.4: Distribuzione di 1000 darts verso il bersaglio, dato $\sigma_r = 1\%$, sopra con $\sigma_\theta = 0.17^\circ$ al variare di σ_φ , sotto al variare di $\sigma_\theta = \sigma_\varphi$



Sappiamo che molti giocatori hanno una deviazione standard alta rispetto a Y e bassa rispetto a X . In Figura 4.4 vediamo infatti come varia Y , modificando il σ_φ e tenendo un $\sigma_\theta = 17^\circ$. I tiri sul bersaglio si dispongono lungo lo spicchio del 20, ci si aspetta quindi che un giocatore con un basso σ_θ e un elevato σ_φ abbia una media punti maggiore al caso inverso visto prima. Tirando la freccia verso il

T20, seppur con una devianza verticale alta di 1° , si riescono ad ottenere punteggi medi attorno ai 22.5 punti. Quindi il dart rientrerà spesso nello spicchio dell' S20.

Aumentando invece i sigma in modo proporzionale, vediamo come la macchia di punti vada a coprire gran parte del bersaglio. In questo caso abbiamo mantenuto i sigma angolari uguali tra loro e la distribuzione sul bersaglio ha una forma circolare. Se un professionista(primo quadrante) ha una media di tiro 36.5 punti/tiro, aumentando il valore dei sigma congiuntamente le medie si abbassano passando a 30 punti/tiro- 0.25° , 16.5 punti- 0.5° , 13.5 punti- 0.75° .

4.3 DM Heatmap

Oltre allo studio dei grafici relativi a normalità, qqplot e disposizione dei lanci sul bersaglio, è stato realizzato anche uno studio sulle mappe di calore. Le Heatmaps sono delle rappresentazioni di matrici quadrate, dove ogni valore numerico appartenente alla matrice viene convertito a certo livello di colore. Nel nostro caso dati i $\sigma_\varphi, \sigma_\theta, \sigma_r$ il giocatore effettuerà n lanci verso le coordinate (X, Y) .

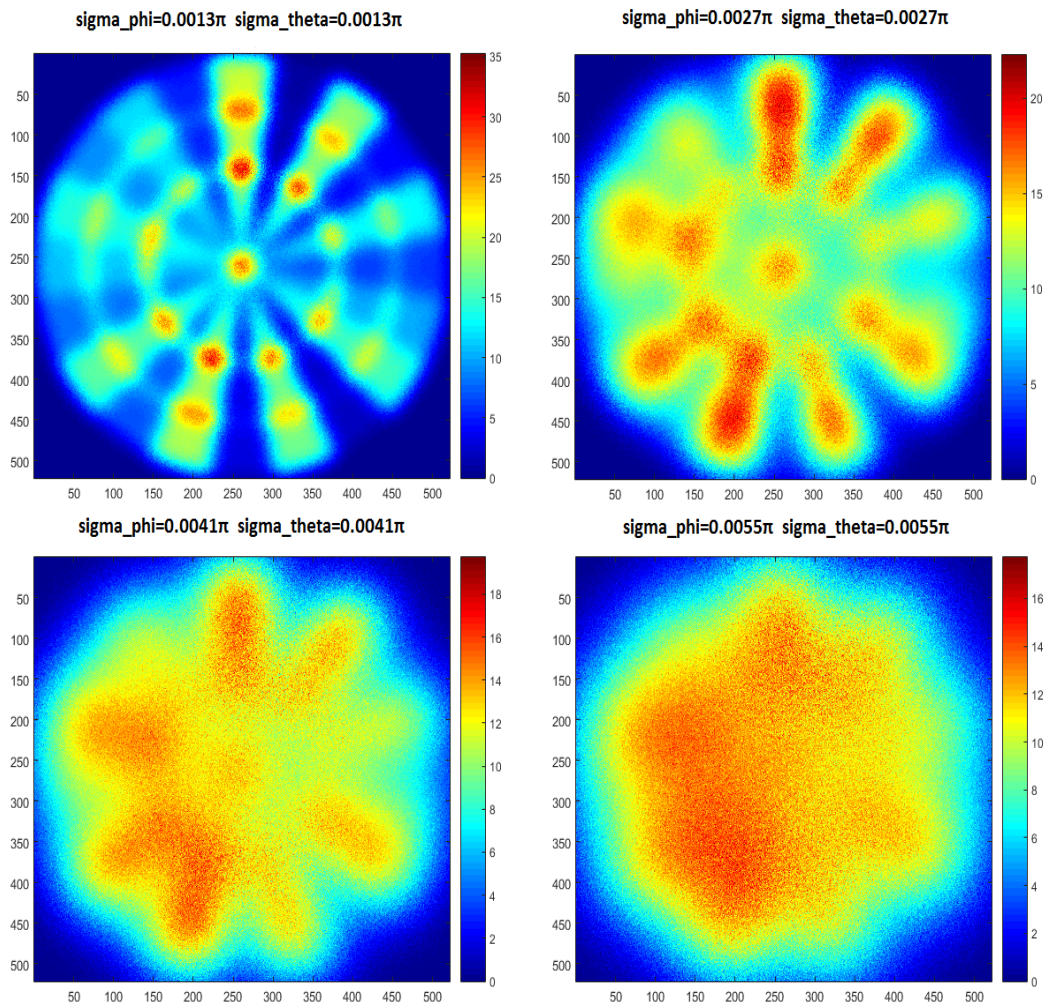


Figura 4.5: DM Heatmaps con $\sigma_\varphi = \sigma_\theta$ e $\sigma_r = 1\%$, in ordine orario $\sigma = 0.25^\circ, 0.50^\circ, 0.75^\circ, 1^\circ$

Andiamo a costruire la mappa relativa al modello fisico-dinamico, che da ora chiameremo "DM Heatmap", usando matlab e impostando per prima cosa il grado di raffinatezza della griglia. Più sarà alta, meglio vedremo le diverse

sfumature delle heatmaps. Avremo quindi una griglia(matrice) $m \times m$, dove ogni coordinata rappresenta un punto del bersaglio, per esempio la coordinata $\frac{m}{2} \times \frac{m}{2}$ sarà il centro del bersaglio oppure la coordinata (1,1) è un tiro mirato(fuori cerchio) in alto a sinistra.

Dopo aver scelto quindi il grado m di raffinatezza della griglia, avremo come output una matrice S di dimensioni $m \times m$ che rappresenterà gli score medi di tiro data la coordinata da mirare. Ogni casella della matrice S contiene un valore score, che verrà convertito nella mappa di calore, formata da colori che vanno dal blu-verde-rosso. Nel nostro caso abbiamo deciso di utilizzare una heatmap con $m=521$, creando una matrice 521×521 , cercando di avere così una buona lettura dei grafici. Il centro del bersaglio sarà la casella 261×261 .

Dalla Figura 4.5 notiamo come all'aumentare delle devianze angolari, i fuochi aumentano. La mappa in alto a sinistra con $\sigma=0.25^\circ$ mostra come il tiratore(vedi legenda) abbia alti scores. Dai colori vediamo come un giocatore con questa accuratezza di tiro, possa benissimo massimizzare i propri punti mirando a posizioni con alti coefficienti di difficoltà come il T20, BullEye o T19. L'aumento del sigma, ci porta a vedere heatmaps che perdono la forma della dartboard e le zone rosse di massimizzazione del punteggio, prima concentrate sui tripli, passano ad occupare quasi interi spicchi. Per esempio con un σ pari a 1° c'è una netta diminuzione degli score medi e la mappa risulta un mix di puntini giallo-rossi.

4.4 Confronto tra EM Heatmap e DM Heatmap

Dopo aver visto la costruzione dell'Heatmap del modello fisicodinamico, cerchiamo di andare a trovare un confronto con quella studiata nel Capitolo 3, relativa all'EM Algoritmo. Partiamo prima di tutto dal fatto che le due Heatmap vengono da modelli sui darts, ma con scopi leggermente diversi.

Con l'EM algorithm, registriamo gli score che realizziamo, tirando verso il bull eye ($\mu = 0$) e successivamente andiamo a stimare la nostra varianza di tiro. Data la nostra deviazione standardizzata costruiamo con R un'heatmap che ci dice come variano i nostri score medi, in relazione a che zona della tavola miriamo. Con una scala di colori disposta in maniera crescente dal rosso al giallo-bianco.

L'Heatmap del nostro modello dinamico è stata fatta con uno studio diverso. Ogni giocatore tira la freccia con una certa velocità e angoli di tiro φ e θ , i quali più saranno precisi nel controllare queste 3 variabili, maggiore sarà la probabilità che la freccia plani nella loro zona desiderata. Come detto in precedenza, il dart colpirà la coordinata (X,Y) sul bersaglio, che corrisponderà a un determinato punteggio, in base alla zona della dartboard in cui si trova.

Quindi le due heatmaps mostrano entrambe gli score medi, in base alla zona che si vuole colpire, ma una ha come input la precisione di tiro e l'altra la sensibilità misurata sulla mano e la forza che si imprime(velocità) . Per realizzare un con-

fronto, abbiamo pensato di utilizzare una funzione in matlab che dati $\sigma_r, \sigma_\varphi, \sigma_\theta$ e il punto (X, Y) da colpire, ci da lo score medio su n tiri effettuati. Poichè il modello dell'EM si basa sul registrare i punteggi mirando verso il centro, con matlab abbiamo applicato un procedimento simile.

4.4.1 Heatmaps di un giocatore di livello medio

Prendendo il caso del capitolo 3, dove un giocatore medio effettua 100 tiri cercando di prendere il 50, abbiamo calcolato la sua media di tiro, pari a $\bar{x} = 13.21$. Successivamente con matlab, per la DM Heatmap, abbiamo impostato 1000 tiri verso il centro $(0,0)$, regolato un $\sigma_r = 1\%$ e modificato le deviazioni standard degli angoli con lo scopo di avere un output simil-uguale allo score medio del modello in R.

```
x = c(12,16,19,3,17,1,25,19,17,50,18,1,3,17,2,2,13,18,16,
      2,25,5,5,1,5,4,17,25,25,50,3,7,17,17,3,3,3,7,11,10,25,
      1,19,15,4,1,5,12,17,16,50,20,20,20,25,50,2,17,3,20,
      20,20,5,1,18,15,2,3,25,12,9,3,3,19,16,20,5,5,1,4,15,
      16,5,20,16,2,25,6,12,25,11,25,7,2,5,19,17,17,2,12)
> mean(x)
[1] 13.21
```

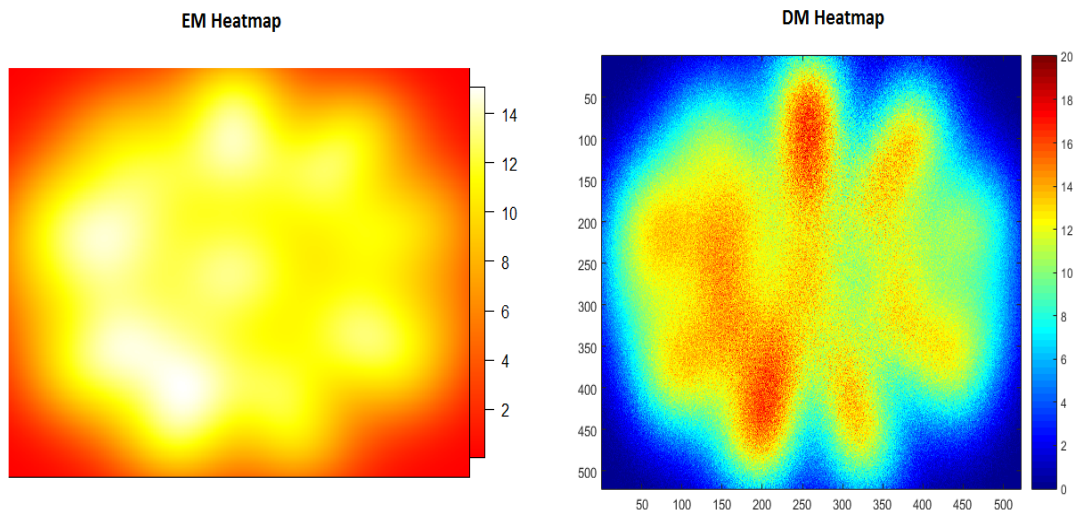


Figura 4.6: Confronto tra EM e DM Heatmap, riferita a un giocatore di livello medio

Il giocatore in questione, nello studio dell'Algoritmo EM, lanciando la freccia verso il centro, ha una media di tiro di 13.21. Mentre la sua matrice di covarianza è pari a:

$$\Sigma = \begin{bmatrix} 333.69 & -151.52 \\ -151.52 & 1674.09 \end{bmatrix}$$

Quindi nelle prove per raggiungere un output uguale abbiamo impostato una devianza di tiro di φ più ampia rispetto a θ . Cosichè un giocatore con $\sigma_r=1\%$, $\sigma_\varphi = \frac{\pi}{180} \simeq 1^\circ$, $\sigma_\theta = \frac{\pi}{322} \simeq 0.56^\circ$ ha uno score medio di 13,20 punti/tiro. Con questi input, abbiamo creato la nostra heatmap e confrontata con l'heatmap del capitolo in R.

Abbiamo deciso di utilizzare colori diversi per il confronto tra le due, con l'aiuto delle scale dei colori poste alla destra dei plot. Osserviamo come in molte zone le densità siano simili tra i due grafici, attorno allo spicchio del 19, i due giocatori mostrano alte densità di score. Anche nell'S20 dal grafico le due heatmap hanno andamenti simili, con alte percentuali di punti, vicino ai 20. Mirando il T20 si può vedere come il tiratore abbia una media bassa-media di tiro, dove la sua alta varianza verticale manda spesso la freccia sopra o sotto il triplo. Dalle mappe di calore si può vedere che entrambe indicano la zone dei T19 e T20 per minimizzare i nostri leg.

4.4.2 Le Heatmaps di Phil Taylor

Approfondiamo il confronto tra Heatmaps, utilizzando anche i dati del più forte giocatore al mondo, Phil Taylor. Dal capitolo precedente sappiamo che la sua deviazione standard di tiro è molto piccola $\sigma_x = 6.42$ mm e $\sigma_y = 7.03$ mm. Questo dato è stato calcolato su solo 18 tiri verso il Bull Eye, presi da due video su YouTube (si veda S [1] [2]).

```
> x = c(50,25,20,8,25,25,25,50,50,50,50,50,50,25,50,25,25)
> mean(x)
[1] 36.27778
```

$$\Sigma_{PT} = \begin{bmatrix} 41.32 & 8.14 \\ 8.14 & 49.49 \end{bmatrix}$$

Da qui abbiamo fatto lo stesso procedimento fatto prima. Ci siamo calcolati la media di tiro di Phil corrispondente a 36.28 punti/tiro verso il centro e siamo passati alla creazione della DM Heatmap. Anche in questo caso, abbiamo tenuto un $\sigma_r=1\%$ e realizzato delle simulazioni modificando man mano i σ_φ e σ_θ cercando di ottenere uno score medio su 1000 lanci attorno al 36.28. Abbiamo deciso di usare uno scarto quadratico medio di φ leggerme superiore vista la matrice di covarianza Σ_{PT} e ottenendo $\sigma_\varphi = \frac{\pi}{1282} = 0.14^\circ$ e $\sigma_\theta = \frac{\pi}{1369} = 0.13^\circ$. Andiamo a confrontare le 2 Heatmaps:

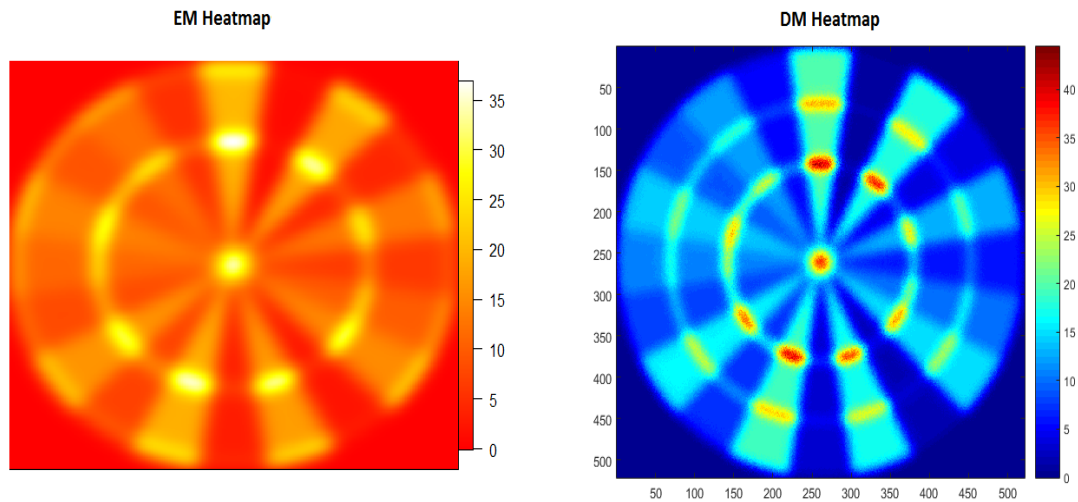


Figura 4.7: Confronto tra la EM e DM Heatmap di Phil Taylor

Vediamo che in questo caso la somiglianza è molto più marcata, questo poichè un'accuratezza di tiro molto più alta, crea una legenda, formata da una scala con molti più valori. Infatti vediamo che il range della scala colori per entrambi va da zero a quaranta. Le mappe di calore, anche se realizzate con colori differenti, mostrano forme e sfocature molto simili tra di loro. Vediamo come Taylor, mirando nel Triplo 20 riesca ad avere una media altissima e se andiamo ad osservare ogni triplo o sezione di ogni heatmap e controlliamo la legenda degli score medi, essi coincidono o sono molto simili tra di loro quasi sempre. Per esempio, mirando verso il T19=57 punti, entrambe le mappe mostrano una media alta, superiore ai 35 punti. Quindi soprattutto da questo confronto, possiamo dire che le Heatmap sono quasi uguali tra loro.

Capitolo 5

Conclusioni

L'obiettivo del nostro studio era quello di analizzare la distribuzione del lancio di una freccia sul bersaglio, dati angoli di tiro e la velocità, facendo un'analisi empirica e un confronto con il modello proposto dai ricercatori inglesi della Stanford University. Dai risultati ottenuti, possiamo notare come gli angoli che danno direzioni a una freccia siano estremamente sensibili. Solo i giocatori professionisti di darts, con la loro leggerezza e precisione, riescono a dare traiettorie perfette e volute. Abbiamo riscontrato che la variazione del raggio di lancio della freccia, verso il bersaglio, dal punto di vista dell'ampiezza, è dovuta strettamente all'angolo di virata θ , o meglio dalla nostra sensibilità nel dare direzione sinistra-destra al dart.

Per quanto riguarda la distribuzione della freccia, verticalmente alla dartboard, abbiamo visto che non solo l'angolo verticale φ va a modellare la nostra Y . Inoltre, la forza impressa porta ad aggiungere il fattore della velocità che insieme alla forza gravitazionale g , crea una distribuzione ellittica di molti giocatori sul bersaglio. Questo porta a spiegare il perchè la maggior parte di giocatori di darts mostrano spesso errori di mira più verticali che orizzontali. Un errato controllo della forza e un calcolo sbagliato della distanza dal bersaglio, possono influire molto sull'errore di tiro di un giocatore.

Con l'Algoritmo EM, abbiamo scoperto un'ottima alternativa ai metodi numerici come Fisher-Scoring o Newton-Raphson, per ovviare il problema di SMV per dei dati mancanti. Osservando i 2 modelli (Simple EM e General EM), costruiti tramite l'algoritmo, vediamo che il secondo garantisce più flessibilità e precisione di risultati. La matrice di varianze-covarianze, fornisce un'aggiunta al modello semplice, che può portare al giocatore zone differenti della dartboard, rispetto a prima, su cui focalizzarsi. Nell'elaborato, abbiamo trattato soprattutto questo fattore, la massimizzazione del proprio punteggio. Le Heatmap mostrano come chi detiene un'ottima precisione, può cercare di ottenere più punti, mirando fa-

cilmente su tripli e doppi, mentre diminuendo l'accuratezza, giocatori alle prime armi hanno una distribuzione di tiro vicino all' U niforme.

Il confronto finale tra le due Heatmaps, ci ha dato come responso, una buona adattabilità ai dati del nostro modello fisico, mostrando una buona somiglianza con gli output dell'EM Algorithm. I due modelli, con scopi simili tra loro, ma diversità dal punto di vista contestuale e dai dati immessi, mostrano dai grafici, aree e punti di tiro tra loro corrispondenti. Il modello da noi sviluppato quindi, spiega in modo efficiente, come componenti fisiche e angolari, influiscono sul punteggio ottenuto, in una partita a freccette.

Appendice A

La Trasformata di Fourier

Data una funzione $f(x) \in \mathcal{F}$, la sua **Trasformata di Fourier** è:

$$\widehat{f}(k) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x) e^{-ikx} dx;$$

qui la variabile k appartiene alla retta reale, $k \in \mathbb{R}$; dunque $\widehat{f}: \mathbb{R} \rightarrow \mathbb{C}$.

L'antitrasformata di Fourier della funzione $\widehat{f}(k)$ è data da:

$$\tilde{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \widehat{f}(k) e^{ikx} dk;$$

Per f continua ed appartenente allo spazio $L[\mathbb{R}]$ vale la formula di ricostruzione :

$$\tilde{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(y) e^{-iky} dy \right) e^{ikx} dk;$$

Nelle applicazioni risulta utile da definire il **Prodotto di Convoluzione** tra 2 funzioni, questo è dato da:

$$(f * g)(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x-y) g(y) dy$$

Il Prodotto di convoluzione è evidentemente commutativo e associativo:

$$f * g = g * f; \quad (f * g) * h = f * (g * h).$$

Dal **Teorema di Convoluzione** sappiamo che La Trasformata di Fourier del prodotto di convoluzione di due funzioni $f, g \in L[\mathbb{R}]$ è pari al prodotto delle Trasformate di Fourier delle due funzioni:

$$T[f * g] = T[f] \cdot T[g]$$

Per la Funzione **Gaussiana**

$$f(x) = \exp \frac{x^2}{2\alpha^2}$$

la trasformata di Fourier risulta essere:

$$\widehat{f}(k) = \alpha \exp \frac{-\alpha^2 k^2}{2}$$

Appendice B

Codici Utilizzati

B.1 Codici R

```
library(darts)
library(fields)

#SimpleEM

# Scores of 100 of my dart throws, aimed at the center of the board

x = c(12,16,19,3,17,1,25,19,17,50,18,1,3,17,2,2,13,18,16,
      2,25,5,5,1,5,4,17,25,25,50,3,7,17,17,3,3,3,7,11,10,25,
      1,19,15,4,1,5,12,17,16,50,20,20,20,25,50,2,17,3,20,
      20,20,5,1,18,15,2,3,25,12,9,3,3,19,16,20,5,5,1,4,15,
      16,5,20,16,2,25,6,12,25,11,25,7,2,5,19,17,17,2,12)

# Get my variance in the simple Gaussian model
a = simpleEM(x, niter=100)

# Check the log likelihood
plot(1:a$niter, a$loglik, type="l", xlab="Iteration", ylab="Log likelihood")

# The EM estimate of my variance
s = a$s.final

# Generate the matrix of expected scores if my variance is s = 799.028,
# in the simple model. This will take a couple of seconds.
e = simpleExpScores(s=799.028)
```

```

# Draw a heatmap
par(mfrow=c(1,2))
image.plot(e, col=heat.colors(256), xaxt='n', yaxt='n')
drawBoard(new=TRUE)
drawAimSpot(e, col = "red")

#General EM

# Scores of 100 of my dart throws, aimed at the center of the board
x = c(12,16,19,3,17,1,25,19,17,50,18,1,3,17,2,2,13,18,16,
      2,25,5,5,1,5,4,17,25,25,50,3,7,17,17,3,3,3,7,11,10,25,
      1,19,15,4,1,5,12,17,16,50,20,20,20,25,50,2,17,3,20,
      20,20,5,1,18,15,2,3,25,12,9,3,3,19,16,20,5,5,1,4,15,
      16,5,20,16,2,25,6,12,25,11,25,7,2,5,19,17,17,2,12)

# Get my variance in the general Gaussian model
a = generalEM(x, niter=100, seed=0)

# The EM estimate of my covariance matrix
Sig = a$Sig.final

# Generate the matrix of expected scores if my covariance is
# Sig=c(15^2,30^2,0), in the general model. This will take a
# couple of seconds.
e = generalExpScores(Sig=(c(15^2,30^2,0)))

# Draw a heatmap
par(mfrow=c(1,2))
image.plot(e, col=heat.colors(256), xaxt='n', yaxt='n')
drawBoard(new=TRUE)
drawAimSpot(e, col = "red")

```


B.2 Codici Matlab

```
#Funzione per realizzare il Bersaglio

function []=draw_dartboard()

t = linspace(0, 2*pi, 100);

% Bullseye (diametro: 12.7 mm / 0.5 inch)
x1 = (6.35)*cos(t);
y1 = -(6.35)*sin(t);

% 25 punti (diametro: 32 mm / 1.25 inch)
x2 = (16)*cos(t);
y2 = -(16)*sin(t);

% Triple interno (spessore anello triple: 8 mm / 0.32-0.38 inch)
x3 = (99)*cos(t);
y3 = -(99)*sin(t);
% Triple esterno (distanza dal centro: 107 mm = 4.2 inch)
x4 = (107)*cos(t);
y4 = -(107)*sin(t);

% Double interno (spessore anello double: 8 mm / 0.32-0.38 inch)
x5 = (162)*cos(t);
y5 = -(162)*sin(t);
% Double esterno (distanza dal centro: 170 mm = 6.7 inch)
x6 = (170)*cos(t);
y6 = -(170)*sin(t);

% Bordo esterno (diametro: 451 mm = 17.75 inch)
x7 = (225.5)*cos(t);
y7 = -(225.5)*sin(t);

for k=0:19
    figure(1)
    hold on
    draw_line([16*cos(pi/20+k*2*pi/20), -16*sin(pi/20+k*2*pi/20)],
              [170*cos(pi/20+k*2*pi/20), -170*sin(pi/20+k*2*pi/20)])
end

hold on
```

```
plot(x1, y1, 'k', x2, y2, 'k', x3, y3, 'k', x4, y4, 'k', x5, y5, 'k',
     x6, y6, 'k', x7, y7, 'k');
daspect([1,1,1])
```

```
end
```

```
function [] = draw_line(p1, p2)
```

```
theta = atan2( p2(2) - p1(2), p2(1) - p1(1));
r = sqrt( (p2(1) - p1(1))^2 + (p2(2) - p1(2))^2);
```

```
line = 0:0.01:r;
x = p1(1) + line*cos(theta);
y = p1(2) + line*sin(theta);
```

```
plot(x, y, 'k')
```

```
end
```

```
#Distribuzione di Y
```

```
function [f_distry, y_distr] = distr_y (theta, sigma_theta, y_min, y_max)
```

```
% Dati di ingresso:
```

```
% theta: angolo (destra-sinistra) esatto per colpire il punto mirato
```

```
% sigma_theta: deviazione standard dell' errore su theta;
```

```
% y_min/y_max: intervallo in cui vogliamo calcolare la distribuzione
```

```
%
```

```
% Dati di uscita:
```

```
% f_distry: distribuzione di y nell'intervallo [y_min, y_max]
```

```
% y_distr: mi serve per disegnare in seguito la distribuzione
```

```
d=2370;
```

```
y_distr=linspace(y_min,y_max,1000);
```

```
f_distry=(1/(sqrt(2*pi)*sigma_theta))*exp(-(atan(y_distr./d)-theta)
.^2/(2*sigma_theta^2)).*(d./(d^2+y_distr.^2));
```

```
end
```

```
#Grafici Vari
```

```
%%% DISTANZA DARTBOARD DAL TIRATORE
```

```
    d=2370;      %millimetri
```

```

%% NUMERO SIMULAZIONI
n=1000;

%% OBIETTIVO (decommentare l'obiettivo interessato)
y=0; z=103; %coordinate triplo 20
r_kmh=64;
r=(r_kmh*100000/3600)*ones(n,1); %r=4930.16*ones(n,1);
%(millimetri/secondi)[phi1, phi2] = rz2phi(r(1),z);
phi=min(phi1, phi2)*ones(n,1); %phi=pi/4*ones(n,1);
theta=atan(y/d)*ones(n,1);

%% ERRORI DI PRECISIONE DEL LANCIO
%Errore modulo (velocità [U+FFFD] del lancio)

sigma_r=r(1)*0.01 %;
e_r=sigma_r*randn(n,1);

%Errore angolo phi (alto-basso)
sigma_phi=0.00095*pi; %sigma_phi=pi/180*0.2 = (PI/180)*10
=Spicchio; %pi/180 = 1 grado
e_phi=sigma_phi*randn(n,1);

%Errore angolo theta (destra-sinistra)
sigma_theta=0.00095*pi;
e_theta=sigma_theta*randn(n,1);

%% PUNTO COLPITO CON PRECISIONE SENZA ERRORI
[X,Y,Z]=throw(d,r,phi,theta);

%% PUNTO REALMENTE COLPITO
[x,y,z]=throw(d,r+e_r,phi+e_phi,theta+e_theta);

s=zeros(n,1);
mean_s=0;
for i=1:n
s(i)=score(y(i),z(i));
mean_s=mean_s+s(i);
end
mean_s=mean_s/n;

%% PLOT DARTBOARD

```

```

figure(1)
hold on
subplot(2,2,1)
plot(y,z,'r.',Y,Z,'b*')
draw_dartboard()
daspect([1,1,1])
titolo=strcat('r=',num2str(r_kmh),'km/h;', 'sigma_r=',num2str(
(sigma_r/r(1)),'; sigma_{phi}=', num2str(sigma_phi/pi),'* pi;', '
sigma_{theta}=', num2str(sigma_theta/pi),'* pi;');
title(titolo)

%%%QQPLOT
subplot(2,2,3)
qqplot(y)
title('QQplot y')

subplot(2,2,4)
qqplot(z)
title('QQplot z')

%%% ISTOGRAMMA 3D
figure(2)
subplot(1,2,1)
hist3([y,z],[70,70])
xlabel('y')
ylabel('z')
set(get(gca,'child'),'FaceColor','interp','CDataMode','auto');
titolo=strcat('r=',num2str(r_kmh),'km/h;', 'sigma_r=',
num2str(sigma_r/r(1)),'; sigma_{phi}=', num2str(sigma_phi/pi),
'* pi;', 'sigma_{theta}=', num2str(sigma_theta/pi),'* pi;');
title(titolo)

subplot(1,2,2)
hist3([y,z],[70,70])
daspect([1,1,100])
xlim([-20,20])
ylim([50,150])
xlabel('y')
ylabel('z')
set(get(gca,'child'),'FaceColor','interp','CDataMode','auto');

%CONFRONTO ISTOGRAMMA DI Y CON LA DISTRIBUZIONE
CALCOLATA

```

```

y_min=floor(min(y));
y_max=ceil(max(y));
[f_distry, y_distr] = distr_y (theta(1),sigma_theta, y_min, y_max);

figure(3)
subplot(1,2,1)
hist(y,100)
titolo=strcat('r=',num2str(r_kmh),'km/h;', ' sigma_r=',num2str(sigma_r
/r(1)),'; sigma_{phi}=',num2str(sigma_phi/pi),'*pi;', ' sigma_{theta}
=', num2str(sigma_theta/pi),'*pi;');
title(titolo)
subplot(1,2,2)
plot(y_distr, f_distry)

%INTERVALLO DI Y
a=-0.5;
b=0.5;

%SELEZIONIAMO I VALORI DI Y CHE CADONO NELL'INTERVALLO [a,b]
ind=find(y>=a & y<=b);

%VALORI Z CORRISPONDENTI AGLI Y SELEZIONATI
z_ind=z(ind);

%ISTOGRAMMA z_ind
subplot(1,2,2)
hist(z_ind,50)

#Heatmaps

m=521; %raffinatezza griglia      (inserire un numero dispari in modo che il
punto mirato sia
al centro della "casella")
y=[];%coordinate y dei punti mirati

n=100; %numero di lanci effettuati su cui viene calcolata la media
S=zeros(m); %matrice S contiene i punteggi medi ottenuti.
           %(immaginando la griglia sopra il bersaglio ,
           % la componente S(1,1) contiene[U+FFFD] il punteggio medio
           % ottenuto da n lanci mirando in alto a sinistra)

```

```

for k=1:floor(m/2)
    y=[-k*451/m;y];
end

y=[y;0;sort(-y)];
z=-y;

for i=1:m
    for j=1:m
        S(i,j) = shot2meanscore (y(j),z(i),n);
    end
end

%HeatMap(S)

%"imagesc" coincide con image solo che riscalda i valori della matrice S in
%modo da usare tutta la colorbar
figure(2)
imagesc(S)
colorbar
title('imagesc(S)')

#Coordinate Punto Colpito

function [x,y,z] = throw(d,r,phi,theta)

g=9810;

%coordinate punto colpito
x=d;
y=d.*tan(theta);
z=(d.*sin(phi))./(cos(phi).*cos(theta))-(g.*d.^2)./(2*r.^2.
*cos(phi).^2.*cos(theta).^2);

end

#Punteggio medio su n lanci effettuati mirando il punto (Y,Z)

function [mean_s] = shot2meanscore (y,z,n)

% Dati di ingresso:
% y: ascissa del punto mirato
% z: ordinata del punto mirato ---> NB: IL CENTRO DEL BERSAGLIO

```

```

HA COORDINATE
(y,z)=(0,0)
% n: numero di lanci effettuati
%
% Dati di uscita:
% mean_p: valore medio dei punteggi ottenuti = (somma punteggi)/(n)

%% DISTANZA TIRATORE
d=2370;      %millimetri

%% CALCOLO (phi,theta) FISSANDO LA VELOCITA' r
r_kmh=64;    %velocita'[U+FFFD] media campionati mondiali (in km/h)
r=(r_kmh*1000000/3600)*ones(n,1);  %velocita'[U+FFFD] in millimetri/secondo

[phi1, phi2] = rz2phi(r(1),z);
phi=min(phi1, phi2)*ones(n,1);

theta=atan(y/d)*ones(n,1);

%% ERRORI DI PRECISIONE DEL LANCIO
%Errore modulo (velocita'[U+FFFD] del lancio)

sigma_r=r(1)*10/1000;
e_r=sigma_r*randn(n,1);

%Errore angolo phi (alto-basso)

sigma_phi=0.00095*pi;
%sigma_phi=pi/180*0.2;  %pi/180 = 1 grado
e_phi=sigma_phi*randn(n,1);

%Errore angolo theta (destra-sinistra)
sigma_theta=0.00095*pi;
e_theta=sigma_theta*randn(n,1);

%% PUNTO COLPITO CON PRECISIONE SENZA ERRORI
[X,Y,Z]=throw(d,r,phi,theta);

%% PUNTO REALMENTE COLPITO
[x,y,z]=throw(d,r+e_r,phi+e_phi,theta+e_theta);

s=zeros(n,1);
mean_s=0;
for i=1:n

```

```

s(i)=score(y(i),z(i));
mean_s=mean_s+s(i);
end
mean_s=mean_s/n;

end

#Funzione che restituisce i valori di phi dati r e z

function [phi1, phi2] = rz2phi(r,z)

g=9810; % (millimetri/secondi^2)
d=2370; % (millimetri)

phi1=atan( (r^2+sqrt(r^4-2*g*z*r^2-g^2*d^2)) / (g*d) );
phi2=atan( (r^2-sqrt(r^4-2*g*z*r^2-g^2*d^2)) / (g*d) );

% phi1=atan( (2*r^2*d+sqrt(4*r^4*d^2-4*g*d^2*(2*r^2*z+g*d^2)))
% / (2*g*d^2) );
% phi2=atan( (2*r^2*d-sqrt(4*r^4*d^2-4*g*d^2*(2*r^2*z+g*d^2)))
% / (2*g*d^2) );

end

%Abbiamo esplicitato phi dall'equazione:
%
% 
$$z = (\sin(\phi)*d) / (\cos(\phi)*\cos(\theta)) + ((-g*d^2) / 2*r^2 * \cos(\phi)^2*\cos(\theta)^2)$$

%
%Ponendo theta=0 ----> cos(theta)=1, abbiamo
%
% 
$$z = (\sin(\phi)*d) / (\cos(\phi)) + ((-g*d^2) / 2*r^2*\cos(\phi)^2)$$

%
% 
$$2*r^2*\cos(\phi)^2*z = 2*r^2*d*\sin(\phi)*\cos(\phi) + g*d^2 = 0$$

%
% 
$$2*r^2*\cos(\phi)^2*z = 2*r^2*d*\sin(\phi)*\cos(\phi) + g*d^2$$

% 
$$(\cos(\phi)^2+\sin(\phi)^2) = 0$$

%
% 
$$(2*r^2*z + g*d^2)*\cos(\phi)^2 - 2*r^2*d*\sin(\phi)*\cos(\phi) + g*d^2*\sin(\phi)^2 = 0$$

%
%Divedendo per cos(phi)^2, abbiamo
%
```



```
%  $g*d^2*\tan(\phi)^2 - 2*r^2*d*\tan(\phi) + (2*r^2*z + g*d^2) = 0$ 
#Coordinate punto colpito dati d,r,phi,theta
function [x,y,z] = throw(d,r,phi,theta)
g=9810;
%coordinate punto colpito
x=d;
y=d.*tan(theta);
z=(d.*sin(phi))./(cos(phi).*cos(theta))-(g.*d.^2)./(2*r.^2.*cos(phi)
.^2.*cos(theta).^2);
end
```


Bibliografia

- [1] J. ALBERT, J. BENNETT, J. COCHRAN(2005), «Anthology of Statistics in Sports», Society for Industrial and Applied Mathematic.

- [2] F.I.G.F. FEDERAZIONE ITALIANA GIOCO FRECCETTE(2014), «REGOLE DI GIOCO E REGOLE DI GARA F.I.G.F.», 13.a edizione 14.03.2014

- [3] A. GIRALDO (1999), «L’algoritmo EM. Aspetti teorici ed applicativi», Technical report series, n.3/1999,

- [4] D. KOHLER(1982), «Optimal Strategies for the Game of Darts», The Journal of the Operational Research Society, Vol. 33, No. 10 (Oct., 1982), pp. 871-884.

- [5] R. J. TIBSHIRANI, A. PRICE, J. TAYLOR(2011), «A statistician Plays Darts», JRSS Series A, Vol. 174, No. 1, 213-226.

- [6] R. J. TIBSHIRANI, A. PRICE, J. TAYLOR(2011), «Supplement to: A Statistician Plays Darts

- [7] «Don’t try for the triple 20», Significance magazine, Vol. 8, No. 1, 46-48, (2011)

- [8] M. TOWNEND(1984), «Mathematics in Sports», Ellis Horwood Limited, New York

Sitografia

- [1] Bullseyes For Beer | Channelbee, <https://www.youtube.com/watch?v=WQgrxzHdqpo>
- [2] Bullseye Challenge | Phil Taylor, <https://www.youtube.com/watch?v=ZTXqlqouaDs>
- [3] F.I.G.F. - Federazione Italiana Gioco Freccette, <http://www.figf-italia.it/>
- [4] Freccette.info, <http://www.freccette.info/>
- [5] Gary Anderson, [https://en.wikipedia.org/wiki/Gary_Anderson_\(darts_player\)](https://en.wikipedia.org/wiki/Gary_Anderson_(darts_player))
- [6] Il Moto Parabolico, https://it.wikipedia.org/wiki/Moto_parabolico
- [7] Phil Taylor career statistics, https://en.wikipedia.org/wiki/Phil_Taylor_career_statistics
- [8] Phil Taylor, [https://en.wikipedia.org/wiki/Phil_Taylor_\(darts_player\)](https://en.wikipedia.org/wiki/Phil_Taylor_(darts_player))
- [9] Professional Darts Corporation, <http://www.pdc.tv/>
- [10] R. J. Tibshirani, A. Price, J. Taylor, A Statistician plays darts, <http://www.stat.cmu.edu/~ryantibs/darts/>
- [11] TUTTODARTS, <https://tuttodarts.wordpress.com/>

Ringraziamenti

Inanzitutto, volevo ringraziare il Prof. Marco Ferrante, per la sua disponibilità e per avermi aiutato a concludere questo lavoro di tesi, oltre che per la sua attitudine nel trasmettere la passione del Calcolo Probabilistico.

La Dott.ssa Stefania Pan, per la gran parte degli scripts realizzati con matlab e per la gentilezza mostrata sempre in questi mesi, aiutandomi a risolvere i miei molti dubbi, via E-mail.

La mia famiglia, che in questi anni, grazie alla loro presenza e totale fiducia nei miei confronti, mi è sempre stata vicino, nonostante a volte, i momenti non facili.

I miei amici più cari, con cui ho trascorso gran parte del tempo libero, durante il periodo universitario (non li cito tutti perchè sono tanti), i miei compagni di Università per i consigli e il tempo passato assieme, e infine, quelli del calcetto, che mi aiutano a staccare la spina.

Ringrazio infine, il mio Campus di Santa Caterina, che forse sarà una delle cose che mi mancherà di più dell'Università.