

Università degli Studi di Padova

FACOLTÀ DI SCIENZE STATISTICHE
Corso di Laurea Specialistica in Statistica e Informatica

Tesi di Laurea

DISTRIBUZIONI CIRCOLARI ASIMMETRICHE E BIMODALI PER MODELLARE LE DIREZIONI DEL VENTO

**WRAPPED SKEW-SYMMETRIC
DISTRIBUTIONS BY MODELLING THE WIND
DIRECTION**

Candidato:

Estefanía Hernández Sánchez

Matricola: **589755**

Relatore:

Bruno Scarpa

Anno Accademico 2010-2011

Ai miei genitori.

Indice generale

INTRODUZIONE.....	1
1. DESCRIZIONE DEI DATI E CONTESTO AMBIENTALE.....	5
1.1 Il vento e le direzioni del vento.....	5
1.1.1 Caratteristiche tecniche del vento.....	8
1.2 Caratteristiche ambientali della zona in indagine.....	9
1.2.1 Situazione geografica.....	9
1.2.2 Il clima della zona di studio.....	12
1.2.3 Circolazioni del vento nell'interno della provincia di Alicante.....	13
1.3 Campione in indagine e stazione meteorologica analizzata.....	16
1.3.1 Stazione Oregon Scientific WMR928NX.....	16
1.3.2 Descrizione dei dati.....	17
2. STATISTICA DIREZIONALE E DATI CIRCOLARI.....	21
2.1 Definizioni e considerazioni preliminari.....	22
2.2 Variabili casuali direzionali.....	25
2.2.1 Funzione di probabilità e funzione di distribuzione.....	26
2.2.2 Funzione Caratteristica e alcuni momenti di popolazione.....	27
2.2.3 Parametri di forma delle distribuzioni di probabilità.....	29
2.3 Statistiche campionarie.....	31
2.3.1 Momenti trigonometrici campionari.....	32
2.3.2 Media direzionale e lunghezza media risultante.....	34
2.3.3 Varianza circolare campionaria.....	36
2.3.4 Altre misure riassuntive direzionali di posizione, scala e concentrazione....	37
2.4 Modelli statistici circolari notevoli.....	38
2.4.1 Approccio circolare e distribuzioni Wrapped.....	40
2.4.2 Alcune leggi di probabilità asimmetriche.....	42
3. FAMIGLIE DI DISTRIBUZIONI SIMMETRICHE PERTURBATE.....	43
3.1 Distribuzione normale asimmetrica sulla retta reale.....	45
3.1.1 Parametrizzazione diretta e parametrizzazione centrata.....	47
3.2 Distribuzione normale asimmetrica sul cerchio unitario.....	48

3.2.1	Parametrizzazione centrata e circolare.....	50
3.2.2	Metodi di stima e altre riparametrazioni.....	51
3.3	Distribuzione normale asimmetrica circolare modificata.....	54
3.3.1	Forme della distribuzione.....	55
4.	ANALISI STATISTICA.....	61
4.1	Caratteristiche del campione e specificazione del modello statistico.....	62
4.1.1	Asimmetria delle direzioni del vento.....	63
4.1.2	Bimodalità delle direzioni del vento.....	64
4.2	Stima e adattamento dei modelli circolari	67
4.2.1	Vento di giugno.....	68
4.2.2	Brezza di mare.....	70
4.2.3	Brezza di terra.....	71
4.2.4	Test rapporto di verosimiglianza.....	73
4.3	Dipendenza temporale.....	75
	CONCLUSIONI.....	79
A.	CODICE R.....	81
A.1	Distribuzione normale asimmetrica circolare.....	81
A.2	Distribuzione normale asimmetrica circolare modificata.....	85
A.3	Esempi di grafici e calcoli.....	86
	INDICE DELLE TABELLE.....	lxxxix
	INDICE DELLE FIGURE	xc
	BIBLIOGRAFIA.....	92

INTRODUZIONE

«La conoscenza statistica induttiva è inerente alle scienze sperimentali per la sua stessa natura, dato che in queste ultime, quando introduciamo una legge per scoprire un certo tipo di fenomeno, non possiamo fare altro che verificarla tramite nuovi esperimenti e provare a dare una misura della fiducia che si meritano tali esperimenti come verifica della legge. Questo è quello che permette la teoria della inferenza statistica, che fornisce, in termini probabilistici, una misura di tale fiducia».

Sixto Ríos (1913 – 2008).

Le direzioni del vento possono essere considerate come esperimenti aleatori, dato che se misuriamo queste direzioni, alla stessa ora e nelle stesse condizioni, non sempre otteniamo il medesimo risultato. Poiché stiamo parlando di un fenomeno climatologico, le condizioni possono essere molto variabili e dipenderanno soprattutto del periodo dell'anno in cui siamo. Per ciò, fissiamo lo studio sulle direzioni del vento nel mese di giugno, considerando che è un mese più meno stabile per altri fenomeni climatologici e ci permette di dire che le osservazioni, in una specifica zona geografica, sono tutte sotto le stesse condizioni. L'interesse quindi è verificare come si comporta questo fenomeno aleatorio e cioè, in termini statistici, descrivere con che probabilità ogni direzione si verifichi nel mese di giugno.

Partendo allora da un insieme di dati del 2009, vogliamo adattare un modello probabilistico, con idonee situazioni dei parametri di tale modello, che ci permetta di riassumere efficientemente i dati e generalizzare il comportamento delle direzioni del vento, nel mese scelto e nella determinata zona di studio, di qualunque anno.

In questo contesto ambientale, la nostra attenzione sui metodi statistici si concentra sui dati circolari, o anche chiamati dati direzionali, dato che sono le misurazioni, in gradi o radianti, delle direzioni. La particolarità più rilevante di questi dati “speciali” è che non hanno uno *zero naturale* e cioè, qualsiasi direzione può essere arbitrariamente scelta come zero.

Poiché il nostro interesse è soltanto “descrittivo”, attraverso un modello statistico, si lascia magari per un futuro e in collaborazione con specialisti del settore, lo studio sulle

2 Introduzione

cause ed effetti di questa caratteristica del vento, come per esempio l'alterazione della direzione dovuta ad altre variabili o l'esistenza di una possibile relazione con la velocità o la temperatura. Intendiamo allora che per modellare le direzioni del vento, in questo lavoro di tesi stiamo trattando con un'unica variabile casuale, di tipo circolare, e dove la sua struttura probabilistica vuole essere spiegata tramite un modello matematico per poter scoprire l'andamento delle direzioni più probabili di un determinato periodo temporale e in una specifica zona di studio. Per questo scopo, si assume che le probabilità delle osservazioni sperimentali possano essere descritte attraverso una funzione matematica nota che dipende da un certo numero di parametri, ovvero, attraverso una distribuzione di probabilità nota.

Le leggi probabilistiche, che consideriamo possano adattarsi meglio alla nostra tipologia di dati, devono ovviamente essere adeguate ai dati circolari. Tuttavia partiamo dalla distribuzione normale asimmetrica sulla retta reale, presentata da Azzalini (1985), per le sue buone proprietà e la capacità di adattamento alla descrizione di fenomeni empirici, e proseguiamo con lo studio sullo sviluppo teorico di questa distribuzione realizzato da Pewsey (2000), che permette l'applicazione ai dati circolari presentando, in questo modo, la distribuzione normale asimmetrica circolare. Dopo di che, tra l'illustrazione dello stato attuale nel settore delle teorie delle distribuzioni, proponiamo una “nuova” legge probabilistica facendo una modifica alla distribuzione di Pewsey (2000), fondamentalmente analoga a essa ma con variazioni sulla forma della densità. Seguendo il lemma delle famiglie delle distribuzioni simmetriche perturbate (classe alla quale appartengono la normale asimmetrica e la normale asimmetrica circolare), presente in Azzalini (2005), aggiungiamo un parametro secondario a tale distribuzione che ci permette di ottenere una classe di distribuzioni bimodali e sperimentare un nuovo modello sul cerchio, non presente in letteratura.

La scelta di usare queste famiglie di distribuzioni nel cerchio è stata fatta, naturalmente, perché pensiamo si possano adeguare correttamente ai nostri dati in indagine e inoltre è dovuta agli importanti vantaggi verificati negli ultimi anni con l'uso delle distribuzioni asimmetriche, questo comporta che questi modelli siano considerati i più idonei al nostro lavoro di tesi sperimentale.

Dunque, lo scopo di verificare se un determinato campione proveniente di dati

ambientali segue una determinata distribuzione di probabilità diventa complesso, considerando che lavoriamo con un tipo di dato non lineare e che, per la natura dei dati, sono state richieste modifiche delle distribuzioni note per l'adeguamento alla descrizione delle osservazioni. Così la struttura della tesi viene suddivisa in capitoli e di seguito spieghiamo in particolare cosa tratteremo in ognuno di essi:

- **Analisi del contesto ambientale.** Descrizione dei dati con i primi grafici esplorativi, mostrando le caratteristiche tecniche del settore e la zona in studio, e ricercando il modo più appropriato di analizzare le nostre direzioni del vento secondo gli studi già fatti da esperti del ambito climatologico.
- **Studio dei dati circolari.** Cenni teorici più rilevanti della particolare tipologia dei dati osservati, che determineranno poi l'analisi del nostro campione, presentando gli aspetti statistici e le distribuzioni più rilevanti collegati a questi. Discuteremo inoltre l'importanza della asimmetria dei dati che non possiamo trascurare, vista la scelta delle distribuzioni ritenute compatibili con i dati.
- **Distribuzioni di probabilità d'interesse.** Descrizione delle famiglie di distribuzioni simmetriche perturbate e applicazione di queste sul cerchio. Il nostro interesse si focalizzerà sulla sperimentazione di una nuova distribuzione di probabilità che pensiamo possa descrivere meglio il comportamento del nostro fenomeno in indagine.
- **Analisi del campione.** Analisi della capacità di adattamento, dei modelli presentati prima, alle nostre osservazioni sulle direzioni del vento.

Infine, una volta conosciuta la distribuzione di probabilità del nostro fenomeno aleatorio, saremo in grado di descrivere il suo comportamento generale in termini probabilistici.

CAPITOLO 1

DESCRIZIONE DEI DATI E CONTESTO AMBIENTALE

La statistica ambientale, si interessa dei metodi statistici legati alle scienze ambientali. Gli esempi più comuni che di solito troviamo d'interesse di studio per l'ambiente possono essere la qualità dell'aria o dell'acqua nello spazio e nel tempo, gli effetti dell'inquinamento nella natura, alterazioni del clima e ancora altri; che grazie alle nuove tecnologie, forniscono grandi quantità di dati che possiamo interpretare tramite i metodi statistici. L'uso della statistica nel monitoraggio e la protezione dell'ambiente è in attiva crescita d'importanza dato che tramite questa disciplina possiamo essere in grado di capire meglio le situazioni ambientali o i problemi della natura (che purtroppo sono sempre più presenti dovuti all'impatto delle attività umane sugli ecosistemi) che ci circondano ogni giorno.

In quanto allo studio delle direzioni del vento come particolare caratteristica di un elemento meteorologico siamo legati all'ambiente, come fenomeno non deterministico (che mostra una variabilità non trascurabile) vogliamo analizzare il suo comportamento. Per questo obiettivo è necessario capire approssimativamente la natura dei dati specificando a grande idee, che cosa è il vento e come si comporta nella specifica zona di studio e nel periodo prescelto. In questa sezione dunque, analizzeremo le caratteristiche della zona in cui sono stati presi i dati e presenteremo una piccola descrizione dei dati stesi e di come sono stati raccolti.

1.1 Il vento e le direzioni del vento

Un'indagine sul comportamento del vento è importante al fine della conoscenza delle risorse e dei rischi ambientali del territorio. Il vento, influisce sulle strutture

tecniche e naturali, e ha valenza in numerosi campi molti diversi tra di loro, come: l'agricoltura, lo sport e la salute dell'uomo. Una ricerca di questo tipo potrebbe essere utilizzata come studio previo a un'implementazione di energia eolica ed essere d'interesse in ambito meteorologico o aeronautico.

Le direzioni del vento sono definite come le direzioni di provenienza e normalmente sono indicate tramite i punti cardinali; cioè intendendo, per esempio, con un vento da Est l'aria che proviene da Est ed è diretta verso Ovest. Utilizzando la graduazione del cerchio in 360 gradi, possiamo anche esprimere la direzione del vento in gradi, ottenendo una descrizione più dettagliata. Comunemente si usano i gradi in azimut rispetto al nord, ovvero gradi misurati in senso orario dal nord magnetico. In questo modo, un vento di 0° corrisponde ad un vento da Nord, 90° ad un vento da Est, 180° da Sud, 270° da Ovest e 360° viene normalmente escluso dato che indicherebbe di nuovo una direzione proveniente da Nord.

Una rappresentazione delle direzioni del vento sul cerchio fu originariamente considerata per la navigazione ed è la conosciuta *rosa dei venti*, nella Figura 1.1 possiamo osservare come attraverso i punti cardinali vengono definite, con i suoi nomi principali, le direzioni delle masse d'aria provenienti dalle diverse parti del mondo.

Ovviamente, ognuna di queste masse d'aria avrà determinate caratteristiche in funzione da dove arriva e le possibili assunzioni che potremo fare su tali caratteristiche dipenderanno fortemente dalla situazione geografica d'interesse. Per esempio, in Europa centrale, a grandi linee, dai quadranti settentrionali arriverà generalmente aria fredda polare, da Ovest aria mite atlantica, da Sud aria calda umida del Mediterraneo e da Est aria secca continentale; tuttavia la situazione cambierà completamente se ci troviamo nell'emisfero sud del pianeta.

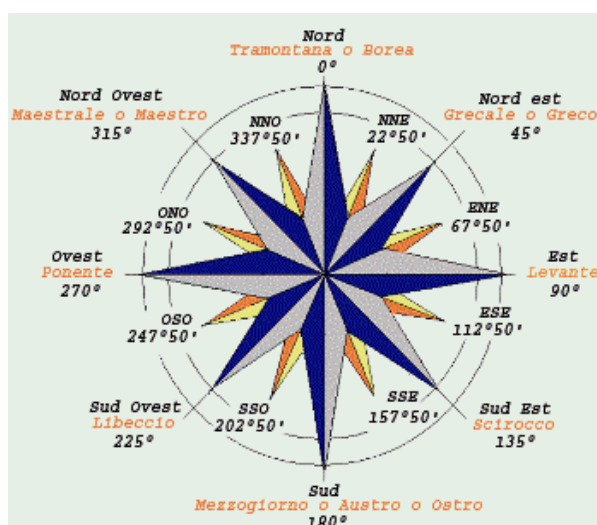


Figura 1.1: Rosa dei venti a 16 punti

Per misurare la direzione del vento ci basta una semplice banderuola che possa girare per indicare qualsiasi direzione di provenienza. In meteorologia lo strumento tecnico usato per questa misurazione è l'anemometro e misura tanto velocità come direzione del vento. Nella Figura 1.2 possiamo vedere le due parti principali di tale strumento, la banderuola e le coppe, dove queste ultime indicano l'intensità del fenomeno in funzione del numero di giri. La forma di questo strumento può variare lievemente secondo il tipo di stazione meteorologica, ma avrà sempre queste due componenti fondamentali.



Figura 1.2: Anemometro WSTX20

1.1.1 Caratteristiche tecniche del vento

Il vento è l'elemento attivo e dinamico dell'aria. La superficie terrestre si riscalda disugualmente per effetto della radiazione solare e l'aria, in contatto con queste superfici non uniformemente riscaldate, si riscalda anche essa in diverse quantità in modo che compaiono settori d'aria calda e settori d'aria fredda. Questo comportamento provoca flussi convettivi¹ d'aria che è quello che conosciamo come vento. L'interazione gravitazionale e la stabilità generale della atmosfera e il suo scarso spessore, fanno che i movimenti verticali d'aria siano un centinaio di volte minori che i movimenti orizzontali (eccetto in casi particolari, come il caso delle tempeste convettive); di fatto, in questo studio si ritengono soltanto i movimenti orizzontali.

Esistono tre controlli sul movimento orizzontale dell'aria vicino alla superficie terrestre:

1. *La forza del gradiente orizzontale (F_{gradp}).*

Il gradiente orizzontale di pressione è la variazione di pressione per unità di distanza, dove tale distanza è misurata perpendicolarmente alle isobare. La forza che soffre il vento dovuto al gradiente di pressione, è determinata dalla seguente equazione: $a = (1/d) \text{grad}p$, dove d = densità e p = pressione.

In quanto all'influenza del gradiente di pressione nella forza del vento, abbiamo la seguente regola: *“Il vento va dalla alta alla bassa pressione e la sua velocità è direttamente proporzionale al gradiente di pressione”*, oppure in altre parole: il vento è forte dove le isobare sono molto vicine e debole dove sono più lontane.

2. *La forza di Coriolis ($F_{Coriolis}$).*

L'effetto di Coriolis è una forza apparente, che si osserva in un sistema di riferimento in rotazione. La sua direzione è sempre perpendicolare alla velocità e fa girare il vento a destra sul emisfero nord e a sinistra sul emisfero sud. Esplicitando in termini matematici, la forza di Coriolis ha modulo: $2vw \text{sen} \varphi$, dove v è il modulo della velocità, w è il modulo della velocità angolare della rotazione della Terra e φ è la latitudine del posto.

3. *La forza di frizione del suolo (F_{roz}).*

L'attrito è un effetto che si apprezza soltanto nella Cappa Limite Planetaria, approssimativamente dalla superficie della Terra fino a 1km di spessore. Da i 1000 metri in su, l'attrito è trascurabile e il vento segue parallelo alle isobare.

¹ La convezione è una delle tre forme di trasferire il calore e si caratterizza perché si realizza tramite un fluido (aria, acqua) che trasporta il calore tra zone con diverse temperature.

Dunque considerando queste tre forze, l'accelerazione del vento, av , risponderà alla forza risultante:

$$av = F_{gradp} + F_{Coriolis} + F_{roz}$$

Questa forza avrà, in generale, una accelerazione tangenziale e una accelerazione normale. Nella seguente Figura 1.3 rappresentato il vento "reale", v , dove possiamo vedere le linee *isobariche*, che sono il luogo geometrico dei punti con stessa pressione. L'angolo α dipende dell'effetto di attrito, della latitudine φ e del sistema di pressioni.

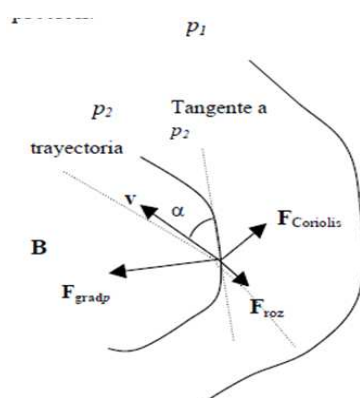


Figura 1.3: Forza risultante del vento. (Fonte: Joaquín Giménez Torres, 2007).

1.2 Caratteristiche ambientali della zona in indagine

1.2.1 Situazione geografica

I dati del vento che useremo per questo studio vengono raccolti nel villaggio di *Villena*, provincia di *Alicante*, con coordinate geografiche (38°38'47"N, 0°53'10"W), nella regione della *Comunidad Valenciana*, Spagna. Concretamente questa località si situa nell'estremo nordoccidentale della provincia *Alicantina* ed è strutturata intorno all'asse fluviale del fiume *Vinalopó*. Questo fiume dà il nome alla *comarca*² dell'*Alto Vinalopó*, che ha una estensione di 630,4 km² distribuiti in sette diversi comuni: *Banyeres*, *Beneixama*, *Biar*, il *Camp de Mirra*, la *Canyada*, *Sax* e *Villena*; questa ultima a una altitudine di 504 metri sul livello del mare con un territorio comunale che copre una estensione di 345,6 km².

² Comarca: Suddivisione territoriale che comprende diversi comuni di una medesima provincia.

Da un punto di vista dalla geomorfologia³, *Villena* si trova in una valle trasversale alle montagne che formano il *Subsistema Prebético* che è un sottosistema delle catene montuose del *Sistema Bético* o equivalentemente chiamato *Cordilleras Béticas*. Il *Sistema Bético* è quello più meridionale dei sistemi montani della Penisola Iberica, trovandosi nella parte sud e sudest della medesima, con una longitudine superiore ai 600 km dallo Stretto di Gibilterra fino al *Cabo de la Nao* (punta più orientale della provincia di Alicante), e continuando ancora sotto il mare per comparire nelle Isole Baleari. Presentiamo di seguito la Figura 1.4 con una mappa della Spagna e la segnalazione della zona di studio.



Figura 1.4: Mappa Fisica della Spagna e localizzazione del comune di Villena.

Questa situazione geografica spiega le caratteristiche della *comarca del Alto Vinalopó* e quelle delle valle e catene montuose che la circondano. Più nello specifico, *Villena* è situata praticamente nel bordo settentrionale del *Subsistema Prebético*. Mostriamo ancora un'altra Figura, la 1.5, per vedere meglio le unità o zone che formano il *Sistema Bético*.

³ Geomorfologia: Branca della geografia fisica che studia le forme che costituiscono il rilievo del territorio.

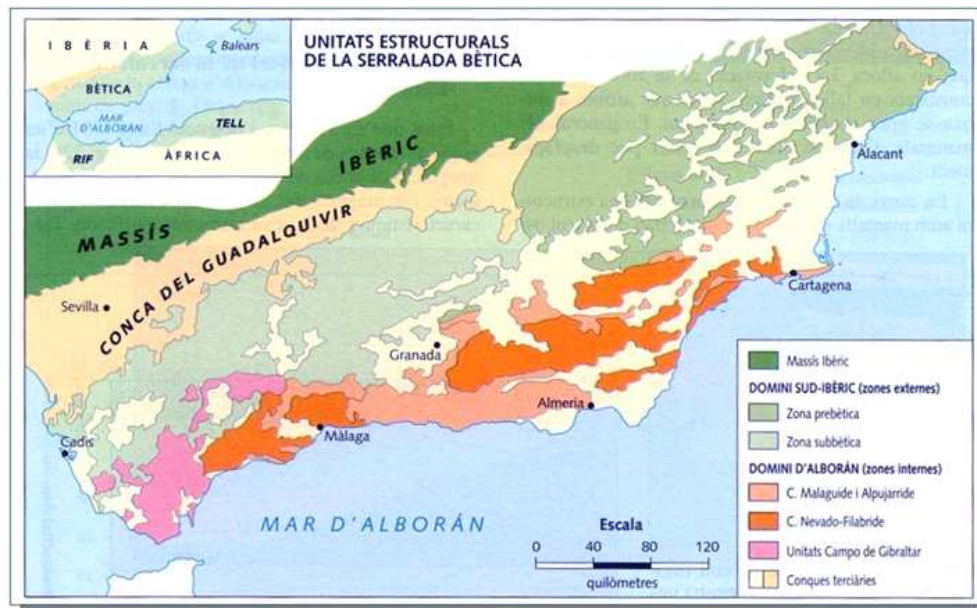


Figura 1.5: Mappa strutturato delle catene montuose delle “Cordilleras Béticas”

Quindi, la nostra località di studio si trova nel limite tra l'unità centrale e quella orientale del *Subsistema Prebético*, concretamente nella valle del *Vinalopó*, che è quello che separa una unità dell'altra. Aumentando ancora di più “lo zoom” della nostra zona d'interesse, presentiamo la Figura 1.6 dove possiamo osservare che le montagne che circondano *Villena* hanno tutte quante una direzione Sudest-Nordest, e inoltre possono essere classificate secondo l'unità geomorfologica e la zona alla quale appartengono all'interno del *Subsistema Prebético*, cioè:

- Catene Montuose del *Prebético Centrale*: Montagne di *Salinas*, Montagne del *Castellar*, Montagne di *Enmedio* e *Picachos de Cabrera*.
- Catene Montuose del *Prebético Orientale*: Montagne della *Solana*, Montagne del *Morrón*, Montagne della *Villa*, Montagne di *Peña Rubia* e Montagne del *Fraile*.
- Il *Valle del Vinalopó*: valle di faglia suddivisa longitudinalmente per una delineazione di formazioni diapiriche⁴, che è quello che nella zona autoctona si conosce con il nome de *Los Cabezos*.

La stazione meteorologica da dove provengono i dati d'indagine per questo studio statistico si situa nella zona del *Valle del Vinalopó*, concretamente sul fianco di una di queste strutture diapirica, chiamata *Cabezo Redondo*.

4 Diapirismo: Fenomeno di risalita di masse rocciose profonde con perforazione delle rocce sovrastanti

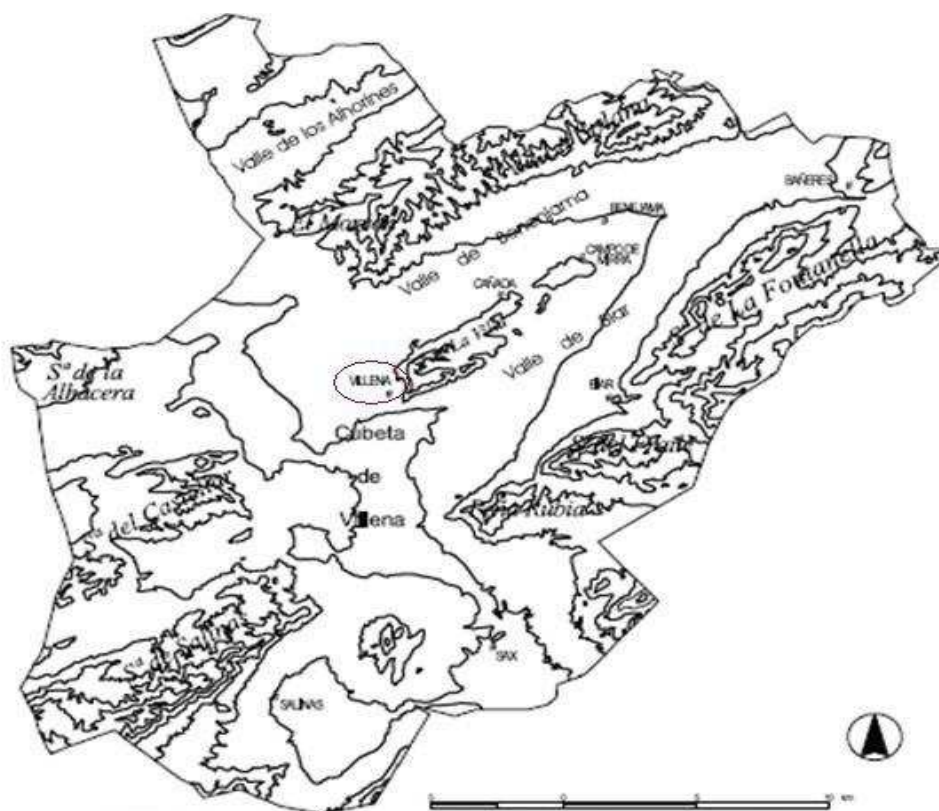


Figura 1.6: Mappa del Alto Vinalopó (Alicante - Spagna). (Fonte: Azorín Molina, 2002:113)

1.2.2 Il clima della zona di studio

Il clima in *Villena* è Temperato Caldo, o equivalentemente, è un clima Mediterraneo. Si caratterizza per avere temperature non molto estreme, in generale, con una stagione secca in state e due massimi pluviometrici, uno in primavera e l'altro in autunno, essendo quest'ultimo il più importante. Questa situazione climatica è molto influenzata dall'Anticiclone delle Azzorre, che in estate si muove verso nord, occupa la latitudine della Penisola Iberica e blocca l'entrata di tempeste provenienti dall'Oceano Atlantico, queste tempeste si muovono in direzione Ovest-Est dovuto alla Forza di Coriolis. Tuttavia non potendo essere in altro modo, il rilievo montano condiziona il clima di ogni regione, e in questo modo in *Villena* si produce un microclima dentro al clima mediterraneo, dovuto alla sua altitudine sul livello del mare (504 m), alle montagne che separano questa località dal litorale e alla situazione in una valle profonda e circondata di catene montuose e valli più elevate, tutte queste caratteristiche

contribuiscono ad avere un clima più estremo in quanto alla ampiezza e oscillazioni di temperatura.

La dinamica atmosferica di questa zona e le caratteristiche climatologiche derivano dalla sua appartenenza alla gran zona di circolazione del vento da Ovest e più concretamente alla sua ubicazione nella valle del Mediterraneo Occidentale. Per analizzare questa dinamica atmosferica, il primo fatto da considerare è di origine topografico, come abbiamo fatto nel punto 1.2.1, e cioè la chiusura montagnosa di questa valle assegna una grande importanza meteorologica ai corridoi di comunicazione che formano un canale, ovvero un cammino di penetrazione, che permette di lasciare il passo a correnti d'aria e anche a temporali frontali e depressioni mobili. Un altro fatto fondamentale per analizzare la meteorologia della valle è la presenza di masse d'aria globali presenti nel mediterraneo, ognuna con caratteristiche diverse.

Tuttavia, il centro di azione più importante sul mediterraneo occidentale, e in particolare sulla Penisola Iberica, è il già nominato Anticiclone delle Azzorre con alte pressioni tropicali e presente a tutti i livelli. Questo è un anticiclone che oscilla tutto l'anno in base al bilancio delle stagioni e che assume un ruolo prevalente specialmente intenso durante l'estate. Nel ambito della Penisola Iberica l'arrivo dell'estate viene caratterizzata per il guadagno in latitudine dell'Anticiclone delle Azzorre che si situa in questo periodo all'ovest della penisola, a latitudine 40-45° Nord. La presenza di questo centro di azione blocca e impedisce il passo delle tempeste atlantiche che restano in questo modo situate sul nord. In questo bilancio della dinamica atmosferica estiva non dobbiamo dimenticare l'importanza alle situazioni in superficie e in altura, cioè come abbiamo considerato prima a livello della crosta terrestre abbiamo particolari condizioni climatiche ma diversi, ovviamente, da quelle che si manifestano in altura.

1.2.3 Circolazioni del vento nell'interno della provincia di Alicante

Nella pubblicazione di Azorín-Molina et al. (2011), “*A multi-year study of sea breezes in a Mediterranean coastal site: Alicante (Spain)*”, troviamo informazione di

grande interesse per questo punto; inoltre Azorín-Molina (2002) indaga sulla medesima zona geografica e circa nello stesso periodo dell'anno ed è per questo che il contenuto di questo capitolo è in parte una traduzione in italiano degli aspetti più importanti per il nostro elaborato.

Per analizzare le circolazioni del vento della provincia di Alicante, e in particolare della *comarca del Alto Vinalopó*, dobbiamo di nuovo prestare attenzione alla situazione geografica della zona descritta nel punto 1.2.1; in dettaglio nelle Figure 1.5 e 1.6 è di grande importanza l'individuazione della linea di frattura disposta trasversalmente all'asse delle *Cordilleras Béticas*. Questa fascia si conosce con il nome del *corredor del Vinalopó* ed è l'unica apertura favorevole verso il sudest, ovvero sulla costa *alicantina*. A conseguenza, si giustifica la genesi dei fronti di brezze attivi nell'interno della provincia di Alicante, che a differenza del resto della Penisola Iberica (che viene isolata di qualsiasi influenza mediterranea attraverso le sue piccole catene montuose), i venti locali della situazione geografica cui in indagine hanno probabili brezze marine anche se con qualche sfumatura e insieme a brezze di valle e montagna.

La prima spiegazione teorica del sistema delle brezze di mare e di terra è del celebre teorema di Vilhelm Bjerknes (1862-1951):

«se i gradienti di pressione e temperatura hanno direzioni diverse si crea energia che forza l'aria a circolare nel senso che porta dall'estremo del vettore gradiente di pressione al vettore di gradiente di temperatura per la traiettoria più corta».

Le brezze di mare e di terra rappresentano venti di carattere periodici e orario manifestato in funzione dei fattori termici, che hanno il compito di modificare i campi di pressione in situazioni atmosferiche con gradienti barici piccoli, in questo modo sono venti soffici in superficie e che operano in alternanza in periodi di 24 ore. Molto semplicemente e in forma schematica il meccanismo funzionerebbe nel seguente modo: in estate, in zona costiera, il riscaldamento della superficie di terra è intenso durante il giorno, l'aria si fa più leggera, si dilata e tende infine ad ascendere (processo termodinamico) essendo il vuoto relativo generato nella terra, riempito con l'arrivo d'aria di precedenza marittima a minore temperatura. A questo meccanismo lo chiamiamo **brezza di mare** e soffia circa mezza mattinata per raggiungere maggiore intensità dopo le ore più calde. La notte, invece, le condizioni si invertono, la brezza

soffia da terra a mare, e il meccanismo viene denominato **brezza di terra**. Un processo simile si sviluppa sulle cime riscaldate per il Sole, dando luogo alla apparizione nei settori continentali di quelle che vengono chiamate **brezze di valle** e **brezze di montagna**; analogamente sono un processo termodinamico e favoriscono un'ascesa obliqua dell'aria durante il giorno, seguendo il percorso del versante, che colma la valle durante la notte. Nella seguente Figura 1.7 possiamo osservare i diversi comportamenti della brezza di mare, o equivalentemente di valle se siamo in zona montuosa, e l'inversione della direzione del vento quando si produce la brezza di terra, o a seconda della nostra situazione, della brezza di montagna.

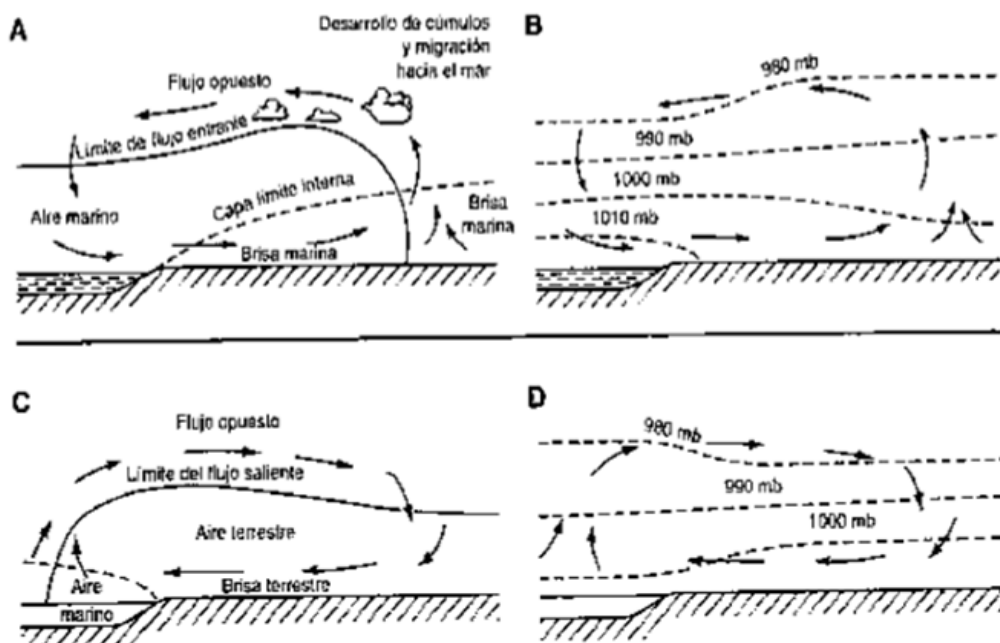


Figura 1.7. A e B: circolazione di brezza marina e distribuzione della pressione all'inizio del pomeriggio con tempo anticiclonico. C e D: stessa situazione di notte, brezza di terra. (Fonte: Azorin-Molina (2002))

Quindi, ci aspettiamo che le frequenze del vento nel mese di giugno, sia una distribuzione fortemente segnalata per le circolazioni dei venti locali (brezze di mare-terra e brezze di valle-montagna). L'apertura della valle di *Villena* verso il SE, attraverso il fiume del *Vinalopó*, permette l'entrata quasi giornaliera della brezza marina con direzione SSE e SE dopo il mezzogiorno. Dovremo anche osservare un'elevata frequenza del vento del S, che si stabilisce normalmente prima della brezza marina e che corrisponde alla brezza di valle. Per la spiegazione di brezza che abbiamo dato precedentemente e il fatto dell'inversione nelle ore notturne, l'aria fredda più pesante

tende a riempire la valle di *Villena* e ci aspettiamo quindi una frequenza elevata anche della circolazione del vento di NW durante la notte che è associato alla brezza di terra.

1.3 Campione in indagine e stazione meteorologica analizzata

I dati a disposizione utilizzati per questo studio, sono scaricabili su richiesta nel sito <http://www.meteoclimatic.com/perfil/ESPVA0300000003400A>. Meteoclimatic è un portale Internet di una grande rete di stazioni meteorologiche non professionisti (di osservatori privati) automatiche in tempo reale e che comprende posti nella Penisola Iberica, nelle Isole Baleari e nelle Isole Canarie, nel sud della Francia e nell'Africa vicina allo Stretto di Gibilterra; è un'importante data base di risorse meteorologiche in Spagna. In particolare la nostra stazione meteo in analisi, "*Villena-Los Cabezos*", è proprietà dell'osservatore privato Jerónimo Hernández Calatayud ed è una stazione di qualità eccezionale secondo questo sito Internet.

1.3.1 Stazione Oregon Scientific WMR928NX

Grazie alle stazioni meteorologiche diffuse sulla superficie terrestre, oggi giorno è possibile avere a disposizione una gran quantità di dati climatici. I nostri dati sulle direzioni del vento sono stati misurati con una stazione meteorologica Oregon Scientific WMR928NX; questo è un modello semi professionista che può essere acquisito in negozi specializzati. Tale strumento viene fornito di un software per la gestione dei dati rilevati su PC e permette di monitorare i seguenti dati climatici:

- Temperatura dell'aria interiore ed esteriore
- Umidità relativa interiore ed esteriore
- Punto di rugiada interiore ed esteriore
- Pressione barometrica
- Velocità e direzione del vento
- Precipitazioni

Inoltre la WMR928NX è anche dotata di altre funzionalità come: ora e data radiocontrollata, previsione del tempo entro un raggio di 32 - 48 Km, allarme selezionabile per determinate condizioni climatiche e memoria per le letture massime e minime.

Tutte queste caratteristiche vengono misurate con tre sensori che vanno installati all'esterno e sono: l'anemometro, il termo-igrometro e il pluviometro; e uno installato all'interno, il barometro-termo-igrometro. Questi quattro sensori hanno una trasmissione di dati alla consol principale della stazione meteo, per radio in una frequenza di 433 Mhz, per cui non occorrono cablaggi tra le unità. Inoltre i componenti esterni hanno ognuno la sua placca solare e il suo trasmettitore indipendenti, affinché possano essere ubicati in posti diversi, in modo tale da ottenere rilevamenti ottimali delle condizioni climatiche per le quali gli strumenti sono stati progettati. La stazione meteorologica ha una gamma di ricezione di 100 metri in un'area aperta e quindi le unità devono essere posizionate entro tale gamma e il percorso di trasmissione deve essere libero da interferenze e ostacoli. La consol meteo interiore che è l'unità principale, riceve i dati per radio dei quattro sensori e mostra tutta l'informazione simultaneamente in uno schermo tattile. Ha una uscita RS232 che ci permette collegare la consol a una porta seriale di un computer per esportare i dati ottenuti, tramite lo specifico cavo e connettore jack, che vengono forniti con l'acquisto della stazione meteorologica, così come il software proporzionato, *Virtual Weather Station Base Edition*, utile poi per la gestione dei dati.

1.3.2 Descrizione dei dati

Per il nostro lavoro abbiamo scelto di studiare il comportamento delle direzioni del vento nel mese di giugno 2009. Il software informatico della stazione meteorologica prende misurazioni ogni 20 minuti e ogni giorno fornisce un file con i risultati di tutte le variabili climatologiche considerate nel punto 1.3.1. Da questo insieme di dati noi abbiamo elaborato un file definitivo con il seguente formato:

Date	Time	WindDir
01/06/09	00:20	34
01/06/09	00:40	1
01/06/09	01:00	NA
...
30/06/09	23:20	NA
30/06/09	23:40	131

Tabella 1.1: Descrizione dei dati

Ovviamente la nostra variabile d'interesse è WindDir, cioè la direzione del vento e viene indicata, come abbiamo descritto nell'introduzione, in gradi azimut rispetto al nord e cioè sotto forma di “bussola”. Occasionalmente la registrazione dei dati della stazione meteorologica può perdere qualche dato dovuta alla sconnessione temporale del PC al quale è collegato, per fortuna questo capita poche volte dato che il computer dell'osservatore è dotato di un gruppo statico di continuità, ovvero un dispositivo *UPS* (dal inglese: *Uninterruptible Power Supply*), che è un'apparecchiatura utilizzata per mantenere costante l'alimentazione elettrica sul PC. Inoltre, il software che memorizza i dati, indica sempre la direzione che marca la banderuola del anemometro, anche quando questa è ferma (ovvero quando la velocità del vento è pari a zero), però tecnicamente non dovrebbe indicare nessuna direzione se vento non c'è. In entrambi queste due situazioni, abbiamo segnalato l'osservazione con “NA” (dal inglese “*Not Available*”, cioè non disponibile). In questo modo delle 2160 osservazioni che dovremo avere, 327 sono valori mancanti (*missing data*) e abbiamo quindi 1833 valori effettivi.

Nella seguente Figura 1.8 presentiamo il campione in indagine suddiviso in giorni, tramite le linee a trattini verticali. In questo modo pensiamo sia possibile osservare se, come abbiamo detto precedentemente nel punto 1.2.3, ci sono venti locali periodici giornalieri e cioè, qualche tipo di brezza di mare e di brezza di valle procedente dal S-SE e l'alternante brezza di terra procedente dal NW.

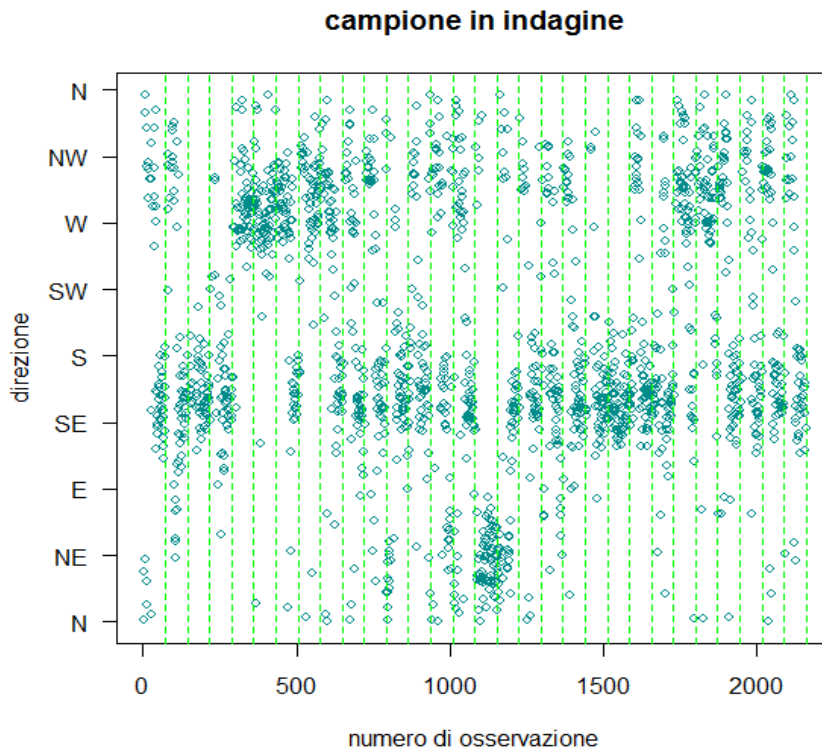


Figura 1.8: Osservazioni del mese di giugno 2009 (separati in giorni) in funzione della loro direzione del vento

Effettivamente in questa Figura, si vede un dominio delle direzioni dei venti intorno al SSE, che indica chiaramente un peso delle circolazioni locali di brezze di mare e di valle. La direzione che ci aspettiamo dalla brezza di terra, cioè il NW, si vede anche abbastanza dominante con molte osservazione che ricadono in questa direzione. Puntualmente, entrano in gioco altri venti come quelli di ponente (W) o quelli grecalesi (NE) che possono anche essere comuni vista la Figura 1.6, dato che possono entrare ed uscire attraverso la valle di *Benejama*. Alcuni punti isolati possono indicare i “outlier”, ovvero i valori anomali. Comunque in generale, osserviamo che ogni giorno di giugno ha l'incirca la stessa distribuzione dei punti, variando forse a seconda se nella giornata a fatto più caldo o meno.

Per completare il grafico precedente, presentiamo di seguito nella Figura 1.9, un altro grafico che conta il numero di osservazioni che ricadono in ogni direzione; cioè l'istogramma campionario delle direzioni del vento in *Villena* nel mese di giugno 2009. Grafico meramente descrittivo, dove l'asse x , cioè l'asse delle direzioni in questo caso, è suddivise in intervalli di 10° e l'asse y indica, come abbiamo detto, il numero di

osservazioni osservate di ogni direzione.

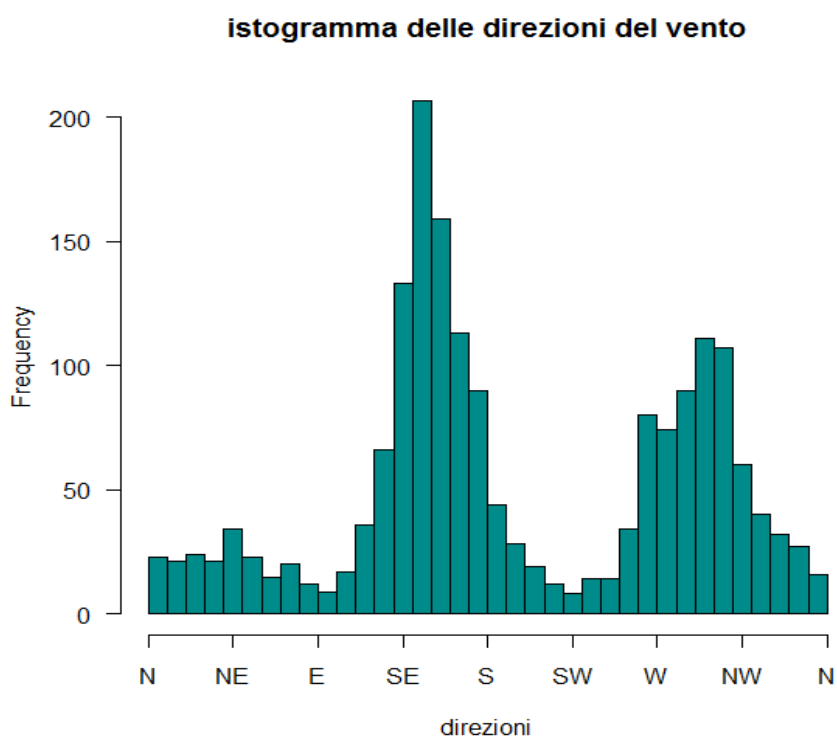


Figura 1.9: Frequenza delle direzioni del vento in Villena nel mese di giugno 2009

La Figura 1.9 rappresenta perfettamente le due direzioni dominanti del vento nella valle di *Villena* che ci aspettavamo; e cioè SSE che indica la brezza di valle e la brezza marina, e NW che indica la brezza di terra notturna. Le giustificazioni di questo comportamenti sono state ben motivate nel punto 1.2.3.

CAPITOLO 2

STATISTICA DIREZIONALE E DATI CIRCOLARI

Diverse scienze si avvalgono dell'uso delle analisi dei *dati circolari* per lo studio di alcuni fenomeni caratterizzati da misure direzionali. Le analisi su questa particolare tipologia di dati si colloca in un settore della statistica chiamato *statistica direzionale*. In questo capitolo si vuole spiegare cosa sono i dati circolari e illustrare un modesto riassunto dei cenni teorici più rilevanti di questo settore della statistica, come ad esempio: le variabili casuali direzionali, i modelli parametrici circolari più importanti, i principali parametri di popolazione con gli analoghi campionari collegate a questa tipologia di dati e l'importanza di misurare la asimmetria.

A livello storico i primi esempi applicativi li troviamo in Astronomia come lo studio delle inclinazioni delle orbite planetarie in funzione del tempo (Funkhouser, 1936) o in Biologia nel fenomeno della migrazione degli animali. Caratteristico è l'esempio attuato da Stephens nel 1969, che presentiamo di seguito nella Figura 2.1, dell'orientamento di 76 tartarughe dopo aver depresso le uova.

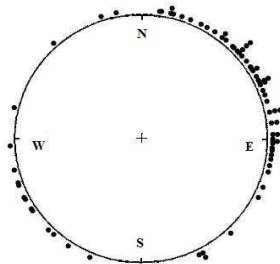


Figura 2.1: Grafico circolare dell'orientamento di 76 tartarughe. (Fonte: *Directional Statistics*, 2000: 9)

Altre ricerche coinvolgono la Geografia (carte isomagnetiche), Geologia (orientamento dei piani di frattura delle placche tettoniche) o la Oceanografia (direzioni delle correnti marine). Continuando, in qualsiasi campo di ricerca che abbia un'insieme di dati dal quale possiamo misurare l'orientamento di ogni direzione.

Il campo di interesse dove concentreremo la nostra analisi sarà la Meteorologia e specificatamente nello studio delle direzioni del vento.

I principali testi che introducono questa metodologia statistica sono Mardia (1972) e Fisher (1993), e più recentemente Mardia and Jupp (2000) e Jammalamadaka and SenGupta (2001), ma ci sono tanti altri autori che hanno dedicato il loro lavoro a questo campo di ricerca.

2.1 Definizioni e considerazioni preliminari

Definiamo i dati circolari (o analogamente, dati direzionali) come le misurazioni delle direzioni, ovvero come un insieme di osservazioni di angoli matematici. Tali osservazioni possono essere viste come punti nel cerchio unitario o come vettori unitari sul piano. Ovviamente, dato che stiamo lavorando con angoli sarà necessario usare funzioni trigonometriche, che vengono anche chiamate funzioni circolari. In generale, un vettore in uno spazio p -dimensionale ($p \geq 2$) viene chiamato *dato direzionale* o *dato sferico*; i dati circolari sono dati sferici in uno spazio bidimensionale. È per questo che il settore della statistica che si occupa di questi dati è noto come statistica direzionale.

Lo spazio campionario di questa tipologia di dati ha una geometria circolare, ed è possibile in questo modo rappresentare tutte le nostre osservazioni in un cerchio di raggio unitario. Da qui il nome di dati circolari che preferiamo usare per questo lavoro di tesi. Il supporto di tali dati, non è quindi la retta reale ma il cerchio unitario compatto. Per cui una osservazione circolare, che denoteremo d'ora in avanti con θ , ha valori soltanto nell'intervallo $[0, 2\pi)$ radianti, o l'equivalente in gradi angolari.

Esistono due principali strumenti di misura circolare, che permettono di distinguere due registrazioni diverse di tale tipologia di dati:

- a *bussola*, nel caso di direzioni nello spazio, come le direzioni del vento;
- a *orologio*, dove con una semplice proporzione possiamo trasformare gli istanti temporali in angoli. Esempio un campione di momenti di arrivo dei pazienti in

un ospedale nell'arco di 24 ore.

Da un punto di vista statistico induttivo, il modo in cui procedere per lo studio dei dati circolari è uguale a quello dei dati lineari; dunque partiremo da un campione casuale rappresentativo della popolazione d'interesse e tramite le misurazioni della caratteristica d'indagine, cercheremo di ricostruire matematicamente il comportamento di tale caratteristica. Tuttavia è importante specificare che le convenzionali statistiche riassuntive, come la media o la varianza campionaria, non possono essere calcolate con il solito approccio lineare dato che nel cerchio non abbiamo un punto di partenza determinato, come lo zero nei dati lineari, e l'origine è arbitraria quindi può essere collocata a nostro piacere. Per questa motivazione è necessario scegliere un **orientamento del cerchio** e specificare una **direzione iniziale**.

Di solito per misurare angoli l'origine è nel punto $(x=1, y=0)$ in coordinate cartesiane e ci muoviamo in senso anti-orario per rappresentarli. A volte però, è utile per una migliore comprensione dei dati, vedere le osservazioni stampate come in una bussola anziché come angoli matematici, cioè dove la direzione iniziale in tale caso si deve mettere nel punto $(x=0, y=1)$ in coordinate cartesiane e l'orientamento del cerchio è in senso orario. Quest'ultima rappresentazione dei angoli è una misura azimut rispetto al Nord ed, come già accennato nel punto 1.1, è la distanza angolare compresa tra la direzione del nord e la direzione in cui cade veramente l'osservazione, per ciò un oggetto posto ad est, ad esempio, ha un azimut di 90° rispetto al nord. Nel caso del vento succede il viceversa ovvero abbiamo le osservazioni misurate in azimut e tal volta ci interesserà stamparle come angoli matematici. Per ottenere una visualizzazione di questo tipo basta fare la trasformazione: $\phi = 90^\circ - \theta$ dove θ è l'osservazione in azimut, ovvero vista come punto cardinale, e la rappresentazione di ϕ viene allora disegnata come angolo matematico.

Un altro aspetto importante da considerare è che, come abbiamo accennato nella definizione dei dati circolari, una direzione nel piano può essere pressa come vettore unitario $\mathbf{x} = (x_1, x_2)$ o equivalentemente come un punto nel cerchio unitario z e cioè come angolo θ o come unità di numero complesso $z = e^{i\theta} = (\cos\theta + i \sin\theta)$.⁵

⁵ La rappresentazione dei numeri complessi in coordinate polari si fa con l'esponenziale. In questo modo tutti i punti $e^{i\theta}$ con $\theta \in [0, 2\pi)$ sappiamo che descrivono una circonferenza di raggio 1.

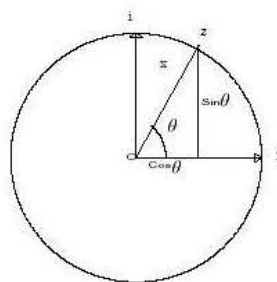


Figura 2.2: Direzione generica rappresentata per il vettore x , l'angolo θ e il numero complesso z . (Fonte: *Directional Statistics*, 2000: 14)

Queste diverse forme di rappresentare la stessa misura di direzione sono naturalmente collegate tra di loro tramite le **coordinate polari o cartesiane**. Se specifichiamo una osservazione come angolo (dalla direzione iniziale al punto sul cerchio corrispondente alla misurazione in questione), stiamo implicitamente usando un vettore unitario in coordinate polari e cioè abbiamo il vettore: (lunghezza del vettore risultante = $\rho = 1$, angolo = θ), dove il dato osservato è la seconda componente di questo vettore e può essere equivalentemente espresso come unità di numero complesso $z = e^{i\theta}$, come si vede nella Figura 2.2. Considerando, però, θ in radianti per semplicità nel verificare le relazioni, notiamo che se la stessa misurazione è definita come vettore (unitario nel piano), solitamente (anche se formalmente possiamo usare sia il sistema di riferimento polare o sia il sistema di riferimento cartesiano) vogliamo fare riferimento al vettore in coordinate cartesiane, quindi dobbiamo indicare le due componenti del vettore e fare la trasformazione: $(x = \rho \cos\theta, y = \rho \sin\theta)$ che anche in questo caso può essere equivalentemente visto come un punto nel cerchio tramite un numero complesso $z = x + iy$.

Dunque, dovremo prestare particolare attenzione alle diverse trasformazioni che dovranno essere fatte, tenendo pure conto che nella conversione a coordinate polari non abbiamo un unico punto nel piano per ogni coppia, dato che θ e $\theta + 2\pi$ sono lo stesso angolo e quindi, i possibili angoli sono tutti quelli che differiscono da esso stesso in multipli di 2π ovvero in aritmetica modulare, gli angoli con modulo 2π che verranno indicati con: $(\text{mod } 2\pi)$. Tutte le operazioni aritmetiche che faremo di seguito nel cerchio, avranno modulo 2π ma per semplicità di lettura, non sempre verrà specificato.

2.2 Variabili casuali direzionali

Siano $\theta_1, \dots, \theta_n$ realizzazioni di una variabile aleatoria Θ che definisce un insieme di direzioni tramite angoli, matematici o in azimut, definiamo allora Θ come una variabile aleatoria direzionale, o equivalentemente, come una variabile casuale circolare.

È da precisare che a seconda se rappresentiamo una generica osservazione circolare θ_i (per $i=1, \dots, n$) come angolo o come numero complesso, possiamo distinguere tra due diversi approcci della statistica direzionale:

- Approccio intrinseco (“*Intrinsic Approach*”): rappresentazione dell'osservazione come **angolo**, quindi le direzioni sono considerate come punti nel cerchio stesso, ed è per questo che possiamo dire che è un approccio intrinseco alle statistiche direzionali.
- Approccio interno (“*Embedding Approach*”): rappresentazione dell'osservazione come **numero complesso**, quindi consideriamo le direzioni come punti speciali nel piano.

Questi due approcci vengono principalmente definiti per calcolare le statistiche direzionali, tuttavia ci servono anche per intendere le variabili casuali direzionali e definire poi i modelli statistici parametrici in base a un approccio o all'altro (le famiglie di distribuzione più importanti per questi due approcci saranno viste nel punto 2.4). Secondo Ferrari (2009), nell'approccio intrinseco il cerchio è utilizzato come spazio campionario e quindi le distribuzioni di probabilità sono definite direttamente sul cerchio; nell'approccio interno invece, lo spazio campionario è considerato come parte di uno spazio più ampio e le distribuzioni in $S^{(n-1)} = \{y \in \mathbb{R}^n : \|y\| = 1\}$ (ovvero, lo spazio campionario circolare) possono essere ottenute come proiezione radiale⁶ delle distribuzioni lineari in \mathbb{R}^n .

⁶ Si dice *proiezione radiale* la funzione continua $g : \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}^n \setminus \{0\}$ che a x associa $\frac{x}{\|x\|}$

2.2.1 Funzione di probabilità e funzione di distribuzione

La funzione di probabilità di una generica variabile casuale associa ad ogni possibile risultato di un esperimento aleatorio (ovvero, ad ogni punto dello spazio campionario di tale variabile aleatoria) la probabilità esistente che detta variabile assuma tale punto. Nel caso in cui la generica variabile casuale sia continua, si parla di **funzione di densità di probabilità**, e infatti, questo è il caso delle variabili aleatorie direzionali, dato che, come indicato nel punto 2.1, le osservazioni circolari cadono soltanto dentro l'intervallo $[0, 2\pi)$, che è un intervallo continuo.

Allora, sia $f(\theta)$ la funzione di densità di probabilità di una variabile aleatoria continua circolare Θ , definiamo $f(\theta)$ come una funzione non-negativa continua tale che:

$$f(\theta + 2\pi) = f(\theta) \quad (2.1)$$

e cioè $f(\theta)$ è periodica con periodo 2π .

Inoltre:

$$\int_0^{2\pi} f(\theta) d\theta = 1 \quad (2.2)$$

Com'è noto, dalla funzione di probabilità di una variabile casuale possiamo ricavare la sua **funzione di distribuzione** -cumulativa- (o anche chiamata funzione di ripartizione). Tale funzione associa ad ogni punto dello spazio campionario (nel caso delle variabili direzionali: θ), la probabilità dell'evento "la variabile casuale Θ assume valori minori o uguali a θ ". La funzione di distribuzione di una variabile aleatoria direzionale viene indicata con $F(\theta)$ e, una volta scelte la direzione iniziale e l'orientazione del cerchio, si può definire (dato che abbiamo specificato Θ variabile casuale continua) attraverso l'integrale di $f(\theta)$, cioè:

$$F(\theta) = \int_0^{\theta} f(\phi) d\phi \quad (2.3)$$

In particolare, $F(2\pi) = 1$.

La funzione di distribuzione determina univocamente la distribuzione di probabilità (cioè, la legge di probabilità) di una generica variabile casuale. Questa legge, assegna ad ogni sottoinsieme dei possibili valori di Θ (nel nostro caso), la probabilità che tale variabile assuma valore in quel sottoinsieme.

2.2.2 Funzione Caratteristica e alcuni momenti di popolazione

Un altro modo per specificare le distribuzioni di probabilità delle variabile casuali è tramite la funzione caratteristica dato che, tra queste ultime e le funzioni di distribuzioni esiste una biiezione, quindi la funzione caratteristica è unica ed esiste per ogni variabile casuale.

Definiamo la **funzione caratteristica** di una variabile casuale Θ , analogamente della sua funzione di distribuzione, come la funzione che associa a $p \in \mathbb{R}$ la $E[e^{ip\Theta}]$ dove E indica il valore atteso di una variabile aleatoria. In questo lavoro denotiamo la funzione caratteristica con $\{\Psi_p : p=0, \pm 1, \dots\}$ dove $\Psi_p : \mathbb{R} \rightarrow \mathbb{C}$ è una sequenza doppiamente infinita di numeri complessi, quindi tramite la definizione appena data possiamo scrivere formalmente:

$$\Psi_p = E[e^{ip\Theta}] = \int_0^{2\pi} e^{ip\Theta} f(\theta) d\theta, \quad p=0, \pm 1, \pm 2, \dots \quad (2.4)$$

In particolare, $|\Psi_p| \leq 1$.

Allora, dato che Θ definisce angoli casuali, usando le proprietà dei numeri complessi sappiamo che $e^{ip\Theta} = \cos p\Theta + i \sin p\Theta$ quindi, di forma analoga, è possibile riscrivere la funzione caratteristica definita in (2.4) in termini di parte reale e parte immaginaria, e cioè:

$$\Psi_p = \int_0^{2\pi} \cos p\Theta f(\theta) d\theta + i \int_0^{2\pi} \sin p\Theta f(\theta) d\theta = \alpha_p + i\beta_p \quad (2.5)$$

dove:

$$\alpha_p = E(\cos p\Theta) \quad \text{e} \quad \beta_p = E(\sin p\Theta) \quad (2.6)$$

sono le coordinate cartesiane della funzione caratteristica definita in (2.4) per la variabile aleatoria direzionale Θ .

Questa ultima specificazione data in (2.6) indica come possiamo ottenere i **momenti trigonometrici di popolazione di ordine p sulla direzione zero**, dove il concetto di momento trigonometrico verrà spiegato meglio nella sezione 2.3.1, si veda in particolare il vettore campionario (a, b) definito in (2.18), analogo al vettore di popolazione (α, β) . Le due componenti di questo vettore di popolazione possono essere chiamate,

rispettivamente, come **momento coseno** di popolazione e **momento seno** di popolazione (ovviamente, di ordine p).

Nel caso in cui la funzione caratteristica ha $p=1$ otteniamo i **parametri di popolazione** più importanti per la statistica direzionale, dove specificando la funzione caratteristica con una trasformazione in coordinate polari del vettore appena definito in coordinate cartesiane, otteniamo, tramite la nuova notazione, la seguente relazione:

$$\Psi_1 = \alpha_1 + i\beta_1 = \rho e^{i\mu} \quad (2.7)$$

dove μ è la **media direzionale di popolazione** e ρ la **lunghezza media risultante**. Nel paragrafo 2.3.2 verranno spiegati gli equivalenti campionari, che sono le quantità riassuntive che descrivono un generico campione; i parametri di popolazione, invece, si differenziano da questi perché sono considerati, non come una semplice quantità numerica ma, ancora come variabili aleatorie.

Come nel caso lineare, possiamo calcolare altri momenti di una variabile casuale, centrati questa volta sulla direzione media della distribuzione della variabile aleatoria invece che sullo zero, questi momenti vengono chiamati nella statistica direzionale con il nome di **momenti trigonometrici di popolazione di ordine p sulla direzione μ** e vengono definiti da:

$$\bar{\alpha}_p = E\{\cos p(\Theta - \mu)\} \quad \text{e} \quad \bar{\beta}_p = E\{\sin p(\Theta - \mu)\} \quad (2.8)$$

L'equivalente campionario verrà definito in (2.18). Come nel caso di (2.6), le due componenti di questo vettore di popolazione saranno ulteriormente chiamate, rispettivamente, come (p -esimo) **momento coseno di popolazione sulla direzione μ** e **momento seno di popolazione sulla direzione μ** .

Altri parametri circolari di popolazione importanti sono: la mediana circolare, la moda circolare, la varianza circolare, la deviazione standard circolare e la dispersione circolare. Tutte questi sono ottenuti con le stesse formule campionarie, che verranno di seguito definite in 2.3, cambiando le quantità riassuntive per i momenti della popolazione. Le principali proprietà sui momenti trigonometrici e i parametri di popolazione che abbiamo appena definito, saranno spiegati di seguito per le quantità campionarie, dato che valgono sia per il campione, sia per l'intera popolazione.

2.2.3 Parametri di forma delle distribuzioni di probabilità

I parametri di forma, come la asimmetria e la curtosi, di una popolazione statistica sono una caratteristica molto importante che non può essere ignorata. Per ciò prestiamo particolare attenzione e dedichiamo questo paragrafo alle quantità di popolazione che indicano questa peculiarità delle distribuzioni e all'importanza della asimmetria dei dati e a come questa può essere misurata.

Per **asimmetria dei dati** intendiamo un campione che non è perfettamente simmetrico alla posizione centrale della sua struttura di frequenze di probabilità. Ovvero, nella **retta reale**, una distribuzione si dice “asimmetrica” quando la sua forma non si presenta speculare rispetto ad un valore fissato x_0 (posizione centrale). Quindi le distribuzioni in \mathbb{R} si distinguono in due tipi di asimmetria:

- **Asimmetria positiva** quando la forma della distribuzione di probabilità è caratterizzata da una coda allungata verso destra.
- **Asimmetria negativa** quando la forma è invece caratterizzata da una coda allungata verso sinistra.

La principale regola che può aiutarci a capire la forma di una variabile aleatoria nella retta reale si basa su due principali proprietà del modello Normale, o in generale, di tutti i modelli statistici simmetrici. Tutti questi verificano che:

- i. Valore atteso, mediana e moda (se è unica) coincidono
- ii. I momenti centrali di ordine dispari sono nulli

Un altro modo possibile per vedere se la distribuzione di probabilità di una variabile aleatoria lineare ha una mancanza di simmetria è tramite il calcolo di alcune indici di asimmetria, come per esempio il coefficiente di asimmetria di Fisher, che è quello più usato, e viene definito come:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \in [-1, 1] \quad (2.9)$$

dove μ_3 è il momento centrale di ordine tre della variabile aleatoria lineare in questione e σ^3 la deviazione standard al cubo di tale variabile casuale. Quando γ_1 è uguale a zero la distribuzione è simmetrica, invece $\gamma_1 < 0$ indica che è una

distribuzione asimmetrica negativa e $\gamma_1 > 0$ indica una asimmetria positiva. Da tenere conto però che $\gamma_1 = 0$ è una condizione necessaria ma non sufficiente per la simmetria, ovvero, possono esistere distribuzioni asimmetriche con coefficiente di Fisher uguale a zero. Un altro indice lineare importante usato nella statistica descrittiva è l'indice di Pearson.

Nel **cerchio** invece, la asimmetria di una distribuzione circolare può essere misurata dal parametro di asimmetria:

$$s = \frac{\bar{\beta}_2}{(1-\rho)^{3/2}} \quad (2.10)$$

dove $\bar{\beta}_2$ è il momento seno di popolazione sulla direzione μ di ordine due (definito in (2.8)) e ρ è la lunghezza media risultante di popolazione (definito da (2.7)). Per una distribuzione simmetrica unimodale il parametro s deve essere pari a zero.

Ulteriormente, diremo che una distribuzione di una variabile aleatoria direzionale è **simmetrica su μ** se tale distribuzione è invariante sotto la trasformazione:

$$\theta \rightarrow \mu - \theta \quad (2.11)$$

Cioè, se la distribuzione è simmetrica sulla direzione μ allora la funzione di densità di tale variabile circolare deve verificare che:

$$f(\theta - \mu) = f(\mu - \theta) \quad (2.12)$$

Ovviamente, per definizione dei dati circolari, se una distribuzione è simmetrica su $\theta = \mu$ allora è anche simmetrica su $\theta = \mu + \pi$

Inoltre, come nel caso lineare, una distribuzione circolare simmetrica verifica che:

- i. Se la distribuzione è unimodale allora la direzione media, la mediana e la moda sono tutte tre la stessa direzione.
- ii. I momenti seno di popolazione (o anche campionari) sulla direzione μ , definiti in (2.8) sono pari a zero.

Partendo da questa seconda proprietà introduciamo un altro parametro di asimmetria circolare, che è quello che usa Pewsey (2000), e cioè il secondo momento seno sulla direzione media:

$$\bar{\beta}_2 = E\{\sin 2(\Theta - \mu)\}, \quad -1 \leq \bar{\beta}_2 \leq 1 \quad (2.13)$$

dove μ è la media direzionale di popolazione definita dalla (2.7).

Il parametro di curtosi riguarda lo studio del maggiore o minore concentrazione dei dati intorno alla posizione centrale della forma della distribuzione e conseguentemente il maggiore o minore peso delle code rispetto a questa posizione. Questa misura viene data dalla (2.14).

$$k = \frac{\bar{\alpha}_2 - \rho^4}{(1 - \rho)^2} \quad (2.14)$$

dove $\bar{\alpha}_2$ è il secondo momento coseno di popolazione sulla direzione media definito dalla (2.8) e ρ è la lunghezza media risultante definiti dalla (2.7). In data set unimodali con un “picco” il quale si adatta bene a una distribuzione Normale circolare, definita in 2.4.1 dalla (2.41), k è vicina a zero; quindi al crescer di k aumenta la concentrazione sulla direzione di riferimento.

Nel paragrafo 2.3.4 vedremo le quantità campionari di questi parametri di popolazione.

2.3 Statistiche campionarie

Il modo appropriato in cui procedere per costruire statistiche riassuntive per dati circolari è considerare i punti del cerchio come **vettori unitari** nel piano in coordinate cartesiane, e cioè i vettori $(\cos\theta, \sin\theta)$ (in modo che realizziamo una specie di “linearizzazione” ed è possibile lavorare normalmente con ascisse e ordinate). Una volta realizzati i calcoli corrispondenti si torna al punto di partenza, per dare una quantità riassuntiva del campione, con una ultima conversione a coordinate polari del vettore risultante. Ovviamente, sempre con θ in radianti per verificare le relazioni.

Allora, supponiamo che ci vengono forniti gli angoli $\theta_1, \dots, \theta_n$ come misura di una direzione (siano azimut o angoli matematici), un modo erroneo di procedere sarebbe fare direttamente una media aritmetica di questi angoli, dato che lo stesso angolo può essere indicato con “diversi numeri” dipendendo, come abbiamo spiegato in 2.1, dell'origine prescelto. Perciò dobbiamo definire quantità riassuntive campionarie con un approccio diverso dal solito, in forma vettoriale e che ci permetta di considerare tutti i vettori osservati.

Dunque, facendo una trasformazione della variabile aleatoria Θ che abbiamo dichiarato nel punto 2.2 e che definisce una direzione tramite un angolo, matematico o in azimut, in:

$$\mathbf{X} = (\cos \Theta, \sin \Theta) \quad (2.15)$$

(cioè, trasformando di coordinate polari a coordinate cartesiane) possiamo poi usare l'approccio intrinseco (trasformando di nuovo a coordinate polari) o interno (specificando il vettore risultante in numero complesso). La trasformazione, ovviamente, avviene per tutto il campione $\theta_1, \dots, \theta_n$, quindi otteniamo un nuovo campione x_1, \dots, x_n della variabile aleatoria direzionale \mathbf{X} , dove x_i , è una generica osservazione con coordinate cartesiane $(\cos \theta_i, \sin \theta_i)$.

2.3.1 Momenti trigonometrici campionari

Ricordiamo prima cosa sono i *momenti teorici nella retta reale*. Un momento di ordine p di una generica variabile casuale lineare univariata Y , è definito come la media della p -esima potenza dei valori osservati, ovvero:

$$\mu_p = E(Y^p) \quad (2.16)$$

Mentre il momento centrale di ordine p è la media della p -esima potenza dello scarto alla media, cioè:

$$\bar{\mu}_p = E((Y - E(Y))^p) \quad (2.17)$$

È noto che il momento di ordine 1 è il valore atteso della variabile casuale in questione e il momento centrale di ordine 2 è la varianza.

La funzione generatrice dei momenti di una variabile casuale lineare genera, come il suo nome indica, i differenti momenti di tale variabile aleatoria. Un problema di questa funzione è che i momenti della variabile aleatoria e la stessa funzione generatrice dei momenti potrebbe non sempre esistere, perciò è preferibile usare al suo posto la funzione caratteristica che esiste sempre e che per i dati circolari è stata definita dalla (2.4) nel paragrafo 2.2.2.

Nel cerchio, per calcolare i momenti di una variabile aleatoria dobbiamo considerare le coordinate cartesiane della variabile casuale direzionale, come spiegato nella

introduzione di 2.2. Quindi consideriamo $X=(\cos \Theta, \sin \Theta)$, per poter definire i momenti che abbiamo dichiarato nella retta reale, con la differenza che, ovviamente, il generico momento sarà un vettore, e non un unico valore, che ci indicherà una direzione e che sarà possibile trasformare in coordinate polari.

L'equivalente campionario quindi di (2.16) per una variabile aleatoria circolare è il vettore (a_p, b_p) (dove a indica la lunghezza dell'asse x e b la lunghezza dell'asse y in coordinate cartesiane) e viene chiamato **vettore dei momenti trigonometrici campionari di ordine p sulla direzione zero**, definito come:

$$\left(a_p = 1/n \sum_{i=1}^n \cos p \theta_i, \quad b_p = 1/n \sum_{i=1}^n \sin p \theta_i \right) \quad (2.18)$$

Naturalmente, possiamo scrivere questo vettore in forma di numero complesso come:

$$m'_p = a_p + i b_p \quad (2.19)$$

Questa formula è l'equivalente campionario della formula della funzione caratteristica di un angolo aleatorio, dato nella (2.5), possiamo quindi intuire che da (2.19) saranno anche ricavati media direzionale campionaria e lunghezza media risultante campionaria (comunque verrà specificato meglio nel seguente paragrafo 2.3.2).

Analogamente ai momenti trigonometrici campionari di ordine p sulla direzione zero, l'equivalente campionario del momento centrale della retta lineare definito in (2.17) è il **vettore dei momenti trigonometrici campionari di ordine p sulla direzione media**, definito da:

$$\left(\bar{a}_p = 1/n \sum_{i=1}^n \cos p(\theta_i - \bar{\theta}), \quad \bar{b}_p = 1/n \sum_{i=1}^n \sin p(\theta_i - \bar{\theta}) \right) \quad (2.20)$$

Ovviamente, anche questo vettore può essere scritto come unità di numero complesso nella seguente forma:

$$m_p = \bar{a}_p + i \bar{b}_p \quad (2.21)$$

dove un ordine $p > 1$ fa riferimento a un campione $p\theta_1, \dots, p\theta_n$.

Inoltre, per dati lineari, sappiamo che una importante proprietà del valore atteso è la *proprietà di baricentro* che afferma che il momento centrale di ordine 1, definito in (2.17), è sempre uguale a zero. Un'affermazione simile può essere fatta per dati circolari, facendo riferimento al momento trigonometrico campionario di ordine 1 sulla direzione media, definito in (2.20).

Mardia and Jupp (2000) nella sezione 2.4 dimostrano che la somma di deviazioni sulla media è sempre zero, anche nei dati circolari, con la differenza però che in questo caso solo la componente seno del vettore si annulla, e cioè:

$$\bar{b}_1 = 1/n \sum_{i=1}^n \sin(\theta_i - \bar{\theta}) = 0 \quad (2.22)$$

Quindi la (2.21) con $p=1$ risulta $m_1 = \bar{a}_1$

2.3.2 Media direzionale e lunghezza media risultante

Per ottenere una direzione media ed essere sicuri che non cambierà a seconda della scelta dell'origine, dobbiamo considerare il vettore somma delle osservazioni del nostro campione in coordinate cartesiane, cioè $\mathbf{x}_1 + \dots + \mathbf{x}_n$, dove \mathbf{x}_i , per $i = 1, \dots, n$, è una realizzazione della variabile aleatoria direzionale $\mathbf{X} = (\cos \Theta, \sin \Theta)$ definita in (2.15).

Allora, la **direzione media campionaria** della variabile aleatoria circolare Θ (che è, come già indicato, la componente angolare della trasformata a coordinate polari della variabile aleatoria \mathbf{X}) la indichiamo con $\bar{\theta}$ di $\theta_1, \dots, \theta_n$ ed è analogamente la componente angolare del vettore somma risultante trasformata in coordinate polari.

Inoltre, la lunghezza di questo vettore risultante in coordinate polari sarà anche in grado di indicarci se i dati sono molto dispersi o meno, ed è quello che chiamiamo **lunghezza media risultante** e indichiamo con R . Da notare che per definizione, se $R=0$ la media direzionale campionaria $\bar{\theta}$ è indefinita.

Per dare una definizione più formale facciamo riferimento al vettore del momento trigonometrico campionario di ordine 1 sulla direzione zero definito in (2.18), dato che come abbiamo accennato prima, il momento di popolazione equivalente era stato definito dal medesimo vettore di popolazione in coordinate polari.

Allora, per ogni $\mathbf{x}_i = (\cos \theta_i, \sin \theta_i)$ per $i = 1, \dots, n$, il vettore risultante:

$$\mathbf{x}_1 + \dots + \mathbf{x}_n = \left(\sum_{i=1}^n \cos \theta_i, \sum_{i=1}^n \sin \theta_i \right) \quad (2.23)$$

è quindi, banale che tale vettore corrisponde a $n(a_1, b_1)$ dove (a_1, b_1) è il 1°

momento trigonometrico campionario sulla direzione zero e n indica la numerosità campionaria dunque è soltanto una costante di proporzionalità che, non cambia l'angolo $\bar{\theta}$ che ci fornisce la direzione media ma, realizza una standardizzazione della lunghezza media risultante R .

Indichiamo allora, questa standardizzazione della lunghezza media risultante con \bar{R} dove la relazione tra entrambe le misure è semplicemente $\bar{R} = nR$. Questa ultima considerazione gioca un ruolo importante a nostro favore, dato che usando (a_i, b_i) (invece di $x_i + \dots + x_n$) $\bar{R} \in [0,1]$ e allora se le direzioni θ_i , $i = 1, \dots, n$, sono ben raggruppate \bar{R} sarà quasi uno, invece se le direzioni sono lontane \bar{R} sarà quasi zero (tuttavia è importante precisare che non vale il viceversa, cioè $\bar{R}=0$ non implica una dispersione uniforme sul cerchio). Dunque con R standardizzato riusciremo a capire meglio la concentrazione dei dati circolari.

Formalizziamo quindi entrambe le definizioni della media direzionale campionaria e la lunghezza media risultante, tramite il vettore del momento trigonometrico campionario di ordine 1 sulla direzione zero, che dopo le verifiche che abbiamo fatto e scritto in numero complesso (prima con coordinate cartesiane e poi con coordinate polari) viene:

$m'_1 = a_1 + i b_1 = |\bar{R}| e^{i\bar{\theta}}$ e che come indicato nel punto 2.2.2, sono l'equivalenti campionari della (2.7). In questo modo, trasformando a coordinate polari il vettore (a_i, b_i) con le solite formule di conversione e tenendo conto che nelle inverse delle funzioni trigonometriche si usa solo una parte della funzione ed è quindi necessario identificare i diversi casi, possiamo definire $\bar{\theta}$ se e solo se $\bar{R} > 0$ come:

$$\bar{\theta} = \begin{cases} \tan^{-1}(b_1/a_1), & a_1 > 0, b_1 > 0 \\ \pi + \tan^{-1}(b_1/a_1), & a_1 < 0 \\ 2\pi + \tan^{-1}(b_1/a_1), & a_1 > 0, b_1 < 0 \end{cases} \quad (2.24)$$

dove $\tan^{-1}(x) \in [-\pi/2, \pi/2]$. Allora, dato che noi lavoriamo in $[0, 2\pi)$ radianti, se $\bar{\theta} < 2\pi$ basta fare $\theta + 2\pi$.

Per la formula matematica di \bar{R} dato che per definizione è la lunghezza del vettore risultante somma $x_i + \dots + x_n$ e abbiamo verificato che questo vettore è proporzionale a (a_i, b_i) ci basta calcolare, tramite il modulo vettoriale, la distanza dal centro al vertice del vettore dei momenti trigonometrici campionari di ordine 1 sulla direzione zero e

cioè:

$$\bar{R} = \sqrt{a_1^2 + b_1^2}, \quad \bar{R} \in [0,1] \quad (2.25)$$

Inoltre, vista la (2.22), un altro modo per definire \bar{R} viene dato tramite il momento trigonometrico campionario di ordine 1 sulla direzione media e cioè:

$$m_1 = \bar{a}_1 = 1/n \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = \bar{R} \quad (2.26)$$

per la dimostrazione di questo risultato si rimanda a *Mardia & Jupp (2000)* : 15 e 21.

Mostriamo ora due importanti **proprietà** di queste due quantità campionarie che abbiamo definito.

i. La media direzionale $\bar{\theta}$ è **EQUIVARIANTE** sotto rotazioni.

Cioè, se supponiamo che una nuova direzione iniziale sia stata scelta, facendo angolo α con l'originale direzione iniziale, allora i punti del nuovo campione corrispondono ai angoli $\theta_j' = \theta_j - \alpha$ con $j = 1, \dots, n$ e la media direzionale campionaria di questi nuovi dati è: $\bar{\theta}' = \bar{\theta} - \alpha$

ii. La lunghezza media risultante \bar{R} è invece **INVARIANTE** sotto rotazioni e sotto riflessioni⁷, e cioè $\bar{R}' = \bar{R}$.

Ovvero la lunghezza del vettore non cambia sotto le diverse trasformazioni.

2.3.3 Varianza circolare campionaria

Anche se nelle statistiche descrittive lineari la varianza viene normalmente definita da (2.17) con $p=2$, nel caso dei dati circolari viene invece usata una diversa definizione. In ogni modo il secondo ordine del momento trigonometrico campionario sulla direzione media, continua comunque ad essere di particolare interesse, come abbiamo visto nella (2.13).

Nel punto precedente, 2.3.2, abbiamo indicato che \bar{R} fornisce una misura della concentrazione dei dati, tale misura però è un indice di concentrazione inverso rispetto a come solitamente siamo abituati a valutare. Dunque, col fine di paragonare la

⁷ Se intende con riflessione che si conserva oltre alle relazioni angolari reciproche (rotazione) anche l'orientazione.

dispersione dei dati circolari con dati nella retta reale è necessario introdurre un'altra misura che però ovviamente, dipenderà fortemente da essa.

Definiamo quindi la **varianza circolare** come:

$$V = 1 - \bar{R}, \quad V \in [0,1] \quad (2.27)$$

dove è ovvio che l'interpretazione in riferimento a \bar{R} è invertita; cioè, più piccolo è il valore della varianza circolare, più concentrata sarà la distribuzione; e viceversa, più grande è la varianza circolare campionaria, più dispersi saranno i nostri dati.

2.3.4 Altre misure riassuntive direzionali di posizione, scala e concentrazione

Un'altra misura di posizione, diversa dalla media direzionale per dati circolari, è la così chiamata **mediana circolare** che viene definita tramite le sue proprietà. Allora la mediana campionaria direzionale $\tilde{\theta}$ dei angoli $\theta_1, \dots, \theta_n$ è un qualsiasi angolo ϕ tale che:

- i. la metà dei punti dati si trovano nell'arco $[\phi, \phi + \pi)$ e l'altra metà dei punti si trovano nel arco $(\phi - \pi, \phi]$.
- ii. la maggior parte dei punti dati sono più vicini a ϕ che a $\phi + \pi$

La **moda circolare** campionaria, come nel caso lineare, è il valore che più volte si ripete nel campione e cioè la direzione “preferita”.

Le misure di scala, sono anche quantità molto importanti per descrivere un campione di dati. Altri indici di dispersione, diverse della lunghezza media risultante e la varianza circolare, definiti rispettivamente in 2.3.2 e 2.3.3, sono i seguenti:

1. **Deviazione standard circolare:**

$$\sigma = \{-2 \log(1 - V)\}^{(1/2)} = \{-2 \log \bar{R}\}^{(1/2)} \quad (2.28)$$

2. **Dispersione circolare**, definita da:

$$\delta = \frac{1 - \bar{R}_2}{2 \bar{R}^2} \quad (2.29)$$

dove, in quest'ultima, come indicato precedentemente, \bar{R}_2 fa riferimento alla lunghezza media risultante del campione dei angoli raddoppiato $2\theta_1, \dots, 2\theta_n$ e viene calcolato tramite il vettore dei momenti trigonometrici sulla direzione zero, definito in (2.19), e che data la lunghezza media risultante vista in 2.3.2, possiamo ora generalizzare tale definizione con la scrittura in coordinate polari, cioè:

$$m'_p = a_p + i b_p = |\bar{R}_p| e^{i\bar{\theta}_p} \quad (2.30)$$

da dove possiamo ottenere le diverse quantità \bar{R}_p e $\bar{\theta}_p$.

L'analogia quantità riassuntiva del parametro di popolazione s definito da (2.10), è la **misura di asimmetria** campionaria:

$$\hat{s} = \frac{\bar{R}_2 \sin(\bar{\theta}_2 - 2\bar{\theta})}{(1 - \bar{R})^{3/2}} \quad (2.31)$$

dove $\bar{\theta}_2$ e \bar{R}_2 si ottengono in questo caso da m'_2 . Per un campione unimodale, \hat{s} vicino a zero indica simmetria.

Infine, l'analogia quantità campionaria di (2.14) che ci serve per valutare la concentrazione di un campione di dati circolari è la misura di **curtosi** campionaria definita da:

$$\hat{k} = \frac{\bar{R}_2 \cos(\bar{\theta}_2 - 2\bar{\theta}) - \bar{R}^4}{(1 - \bar{R})^2} \quad (2.32)$$

dove come nella formula precedente, $\bar{\theta}_2$ e \bar{R}_2 si ottengono da m'_2 .

È importante precisare che ne \hat{s} ne \hat{k} sono esattamente uguali alle sue versioni di popolazione, in particolare al posto di $\bar{\beta}_2$ ed $\bar{\alpha}_2$ non abbiamo i analoghi campionari \bar{b}_2 e \bar{a}_2 . Questo si deve alla interpretazione che possono prendere i momenti trigonometrici di popolazione nelle distribuzioni circolari molto concentrate, per una maggiore chiarezza si veda *Mardia & Jupp (2000) : 21 e 22*.

2.4 Modelli statistici circolari notevoli

Come abbiamo spiegato nel punto 2.2, ci sono due diversi approcci per il trattamento di dati circolari, l'approccio intrinseco e l'approccio interno. D'altronde un

altro approccio più generale per definire distribuzioni di probabilità nel cerchio è la così chiamata procedura “*wrapping*” e che chiameremo Approccio circolare. Questo si basa nel fatto che una distribuzione di probabilità sul cerchio può essere ottenuta “avvolgendo” una qualsiasi distribuzione di probabilità nota nella retta reale intorno a una circonferenza di un cerchio di raggio unitario. Tale approccio verrà spiegato meglio nel paragrafo 2.4.1. Di seguito invece presentiamo le leggi di distribuzioni ottenute tramite l'approccio intrinseco e l'approccio interno. Infine, nel punto 2.4.2, nomineremo alcune leggi di probabilità asimmetriche sul cerchio, senza entrare in dettaglio, dato che presteremo più attenzione a quelle che interessano per questo elaborato nel capitolo 3.

Le distribuzioni di probabilità più importanti seguendo l'approccio intrinseco sono:

DISTRIBUZIONE UNIFORME CIRCOLARE U_c

È la più semplice distribuzione nel cerchio, dove tutte le direzioni tra 0° e 360° sono equiprobabili. La funzione di densità di probabilità è data da:

$$f(\theta) = \frac{1}{2\pi}; \quad 0 \leq \theta < 2\pi \quad (2.33)$$

DISTRIBUZIONE VON MISES $VM(\mu, k)$

È il modello più usato comunemente per dati circolari. Introdotta da von Mises (1918). Ha due tipi di parametri, uno per definire la posizione e l'altro per descrivere la concentrazione sulla direzione di referenza della distribuzione. È una distribuzione unimodale simmetrica su $\theta = \mu$ e ha la seguente funzione di densità di probabilità:

$$f(\theta; \mu, k) = \frac{1}{2\pi I_0(k)} \exp\{k \cos(\theta - \mu)\}; \quad 0 \leq \theta < 2\pi \quad (2.34)$$

dove $0 \leq \mu < 2\pi$ e $k \geq 0$. La lunghezza media risultante ρ può essere ottenuta come $\rho = A(k) = I_1(k)/I_0(k) \in [0,1]$, dove $I_p(\cdot)$ è la *Funzione di Bessel Modificata* di prima specie e ordine p .

DISTRIBUZIONE CARDIOID $C(\mu, \rho)$

È stata introdotta da Jeffreys (1948) e a volte viene indicata come *distribuzione coseno*. Come la precedente è una distribuzione unimodale simmetrica con due parametri e la sua funzione di densità è:

$$f(\theta; \mu, \rho) = \frac{1}{2\pi} (1 + 2\rho \cos(\theta - \mu)); \quad 0 \leq \theta < 2\pi \quad (2.35)$$

con $0 \leq \mu < 2\pi$ e $0 \leq \rho \leq 1/2$.

Un' esempio invece di modello che segue l'approccio interno e cioè, una distribuzione di probabilità che ha massa concentrata nel cerchio unitario (quindi può essere considerata come una distribuzione particolare del piano) è la distribuzione:

PROJECTED NORMAL DISTRIBUTION $PN_p(\mu, \Sigma)$

Ottenuta proiettando una distribuzione normale p-dimensionale lineare. Dove se \mathbf{Y} è un vettore aleatorio in \mathbb{R} tale che $P(\mathbf{Y}=\mathbf{0})=0$, allora $\|\mathbf{Y}\|^{-1}\mathbf{Y}$ (dove $\|\mathbf{Y}\|$ è la norma di \mathbf{Y}) è un punto aleatorio nel cerchio unitario. Per la funzione di densità si rimanda alla sezione 3.5.6 di *Mardia and Jupp (2000)*.

2.4.1 Approccio circolare e distribuzioni Wrapped

Sia Y una variabile aleatoria lineare con funzione di probabilità $f_Y(y)$ allora la corrispondente variabile aleatoria circolare Θ_w è ottenuta con la seguente procedura “wrapping”:

$$\Theta_w = Y \pmod{2\pi} \quad (2.36)$$

Se il cerchio è definito con numeri complessi unitari allora l'applicazione “wrapping” $y \rightarrow \theta_w$ può essere scritta come:

$$y \rightarrow e^{2\pi i \theta} \quad (2.37)$$

Dunque, per ottenere la **funzione di densità di probabilità** $f_w(\theta)$ della variabile aleatoria Θ_w possiamo avvolgere $f_Y(y)$ definita in \mathbb{R} intorno alla circonferenza di un cerchio di raggio unitario, nel seguente modo:

$$f_w(\theta) = \sum_{k=-\infty}^{\infty} f_Y(\theta + 2k\pi); \quad 0 \leq \theta < 2\pi \quad (2.38)$$

Le proprietà date in (2.1) e (2.2) valgono anche per $f_w(\theta)$.

Di conseguenza, data la variabile aleatoria circolare Θ_w definita in $[0, 2\pi)$ attraverso la trasformazione $(\theta + 2k\pi)$ con la variabile $k \in \mathbb{Z}$ inosservabile, stiamo estendendo il supporto di Θ_w a \mathbb{R} in modo che possiamo applicare la funzione di densità $f_Y(y)$ nella retta reale, in argomento $(\theta + 2k\pi)$.

Inoltre, nel punto 3.2.1, in particolare dalla (2.3), abbiamo visto come ottenere una funzione di distribuzione di una variabile aleatoria circolare. Con l'approccio circolare la **funzione di distribuzione** di Θ viene definita nel seguente modo:

$$F_w(\theta) = \sum_{k=-\infty}^{\infty} F_Y(\theta + 2k\pi) - F_Y(2k\pi); \quad 0 \leq \theta < 2\pi \quad (2.39)$$

dove F_Y è la funzione di distribuzione di Y e come in 3.2.1 $F_w(2\pi) = 1$.

Le principali **proprietà** dell'approccio “wrapping” sono:

- i. Se Y è infinitamente divisibile, allora Θ_w è infinitamente divisibile.
- ii. $(y_1 + y_2)_w = y_{1w} + y_{2w}$
- iii. Se la funzione caratteristica di Y è $\Psi_Y(t)$, allora la funzione caratteristica $\{\Psi_p: p=0, \pm 1, \dots\}$ di Θ_w è data da:

$$\Psi_p = \Psi_Y(p) \quad (2.40)$$

Altre proprietà sono ripotate in *Mardia and Jupp (2000)* sezione 3.5.7.

Tra queste distribuzioni che vengono chiamate “*Wrapped Distributions*” le più importanti sono:

DISTRIBUZIONE WRAPPED NORMAL $WN(\mu, \rho)$

Ottenuta tramite la funzione di densità della variabile casuale normale nella retta reale e applicando la (2.38). La distribuzione di densità di questo modello viene definita da:

$$f(\theta; \mu, \rho) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left\{-\frac{(\theta - \mu + 2\pi k)^2}{2\sigma^2}\right\}; \quad 0 \leq \theta < 2\pi \quad (2.41)$$

DISTRIBUZIONE WRAPPED CAUCHY $wC(\mu, \rho)$

Con distribuzione di densità:

$$f(\theta; \mu, \rho) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}; \quad 0 \leq \theta < 2\pi \quad (2.42)$$

DISTRIBUZIONE WRAPPED JONES & PEWSEY $JP(\mu, k, \psi)$

Famiglia di distribuzioni simmetriche unimodali da Jones & Pewsey (2005), include al suo interno diverse distribuzioni circolari, e ha funzione di densità:

$$f(\theta; \mu, k, \psi) = \frac{(\cosh(k\psi) + \sinh(k\psi)\cos(\theta - \mu))^{1/\psi}}{2\pi P_{1/\psi}(\cosh(k\psi))}; \quad 0 \leq \theta < 2\pi \quad (2.43)$$

Un'ampia classe di distribuzioni di probabilità (di fatto, qualsiasi distribuzione lineare nota) può essere adattata ai dati direzionali tramite l'approccio circolare.

2.4.2 Alcune leggi di probabilità asimmetriche

- Papakonstantinou (1979): Modello la cui forma dipende da due parametri, dove uno di essi determina la asimmetria.
- Batschelet (1981): Estensione della distribuzione di von Mises. Due parametri.
- Pewsey (2000): *Wrapped Skew-normal on the circle* con parametri di posizione, scala e asimmetria.

Quest'ultimo sarà di particolare interesse per questo studio e lo vedremo più dettagliatamente nel capitolo 3. Altri ancora, tali come *Generalized von mises distribution* (famiglia esponenziale di distribuzioni asimmetriche bimodali), o alcuni casi speciali di *Kato & Jones distributions* sono ben riassunti da Rotolo, che ha studiato anche lui alcune famiglie di distribuzioni nel cerchio.

CAPITOLO 3

FAMIGLIE DI DISTRIBUZIONI SIMMETRICHE PERTURBATE

In questo capitolo presentiamo i modelli probabilistici che pensiamo possano adattarsi meglio al nostro campione dei dati e analizziamo le distribuzioni di probabilità d'interesse: la Normale Asimmetrica sulla retta reale (Azzalini, 1985), la Normale Asimmetrica Circolare (Pewsey, 2000) e una modifica della Normale Asimmetrica Circolare. Questi tre modelli di probabilità appartengono alla famiglia di distribuzioni simmetriche perturbate, la quale può essere abbreviata come *classe SS* (dall'inglese Skew-Symmetric); queste particolari distribuzioni, combinando densità simmetriche con funzioni di asimmetria, permettono di adattare modelli statistici più adeguati a descrivere la realtà, dando l'opportunità di catturare asimmetria, code pesanti e multimodalità. Quest'ultima caratteristica, la multimodalità, è quella che vorremo ottenere modificando la distribuzione normale asimmetrica circolare, dato che la nostra distribuzione di frequenze del vento, vista in Figura 1.9, presenta una chiara bimodalità.

Secondo Elal-Olivero, *et al.* (2009) il principale obiettivo delle distribuzioni *Skew* è indagare su modelli statistici parametrici che permettano di modellare dati provenienti da distribuzioni asimmetriche unimodali e bimodali.

Naturalmente, nel momento in cui dobbiamo specificare un modello statistico per il campione dei dati e cioè individuare una famiglia di distribuzioni ritenuta compatibile con tale campione, l'importanza dell'asimmetria dei dati non è da trascurare. Tuttavia, normalmente le metodologie consolidate per l'analisi dei dati stabiliscono di assumere distribuzioni simmetriche (come il modello Normale per dati lineari o von Mises per dati circolari), quando invece, forse sarebbe più coerente cercare di adattare i procedimenti statistici alla vera natura dei dati. Un approccio che di solito si usa per risolvere il problema dell'asimmetria si basa sulle trasformazioni dei dati, come ad esempio, la trasformazione Box-Cox; tuttavia, questa procedura può presentare alcuni

problemi come l'interpretazione dei risultati ottenuti dai dati trasformati. Perciò consideriamo di notevole importanza l'alternativa di disporre di modelli probabilistici con parametri al suo interno che riescano a modellare l'asimmetria dei dati.

Inoltre, A.Pewsey (2000) indica che modellare i dati circolari nella scala originale è una necessità di anche maggiore importanza, dato che il cerchio è compatto (chiuso e limitato) e nemmeno le solite procedure di standardizzazione, come le trasformazioni della distribuzione originaria in una distribuzione espressa in una unità di misura standard, hanno l'equivalente nel cerchio dato che qualsiasi forma di trasformazione diversa della rotazione o riflessione cambia le relative posizioni delle osservazioni.

Il termine *Skew-Symmetric* viene di solito usato per fare riferimento alle densità continue che hanno una forma del seguente tipo:

$$f(x) = 2f_0(x)G\{w(x)\}, \quad x \in \mathbb{R}^d \quad (3.1)$$

dove $f_0(x)$ è una funzione di densità tale che $f_0(x) = f_0(-x)$ e G è una funzione di distribuzione nella retta reale tale che G' esiste ed è una funzione simmetrica attorno a zero; $w(\cdot)$ è una qualsiasi funzione dispari nel senso che $w(-x) = -w(x)$.

Per il nostro lavoro considereremo soltanto il caso $d=1$ dato che siamo in ambito univariato.

Questo **Lemma** presente in Azzalini (2005), ci garantisce che $f(x)$ è una funzione di densità che avrà un effetto di perturbazione della densità simmetrica $f_0(x)$, che può essere non solo di asimmetria, ed è avvenuto grazie al fattore $G\{w(x)\}$.

Pensiamo allora che le nostre famiglie di distribuzioni d'interesse, possano essere riassunte nella Tabella 3.1, dove presentiamo gli “ingredienti” per creare la forma standard, rispettivamente, della normale asimmetrica sulla retta reale, la normale asimmetrica sul cerchio e la nuova distribuzione che chiameremo *normale asimmetrica circolare modificata*.

f_0	G	$w(\cdot)$
densità $N(0,1)$ $\phi(x), x \in \mathbb{R}$	ripartizione $N(0,1)$ $\Phi\{\cdot\}$	λx (con $\lambda \in \mathbb{R}$ parametro di asimmetria)
densità $WN(0,1)$ $\phi_w(\theta), \theta \in [0, 2\pi)$	ripartizione $WN(0,1)$ $\Phi_w\{\cdot\}$	$\lambda \theta$ (con $\lambda \in \mathbb{R}$ parametro di asimmetria)
densità $WN(0,1)$ $\phi_w(\theta), \theta \in [0, 2\pi)$	ripartizione $WN(0,1)$ $\Phi_w\{\cdot\}$	$\alpha_1 \theta + \alpha_2 \theta^3$

Tabella 3.1: Famiglie di distribuzioni simmetriche perturbate d'interesse per i nostri dati

Da questa tabella si ottengono le forme standard rispettivamente:

- della famiglia di distribuzioni di densità $f(x; \lambda)$, che corrisponde al modello normale asimmetrico sulla retta reale;
- della famiglia $f(\theta; \lambda)$, corrispondente alla normale asimmetrica sul cerchio;
- della famiglia $f(\theta; \alpha_1, \alpha_2)$, che indica la normale asimmetrica circolare modificata.

Per $\lambda \in \mathbb{R}$ fissato, le densità $f(x; \lambda)$ e $f(\theta; \lambda)$ sono strettamente unimodali; invece per $-\infty < \alpha_j < \infty$, per $j=1,2$, fissati, la nuova distribuzione $f(\theta; \alpha_1, \alpha_2)$ può essere bimodale.

Queste famiglie flessibili di distribuzioni, lineari (nel primo caso) o circolari (negli altri due casi), verranno presentate rispettivamente nei paragrafi 3.1, 3.2 e 3.3, prestando particolar interesse alla forma della distribuzione normale asimmetrica circolare modificata. Tuttavia, per le distribuzioni circolari presentate nella Tabella 3.1, che ha uno scopo meramente descrittivo, non presenteremo quella che abbiamo chiamato forma standard, ma il suo costrutto più generale attraverso l'approccio circolare visto in 2.4.1.

3.1 Distribuzione normale asimmetrica sulla retta reale

Per presentare il funzionamento di questa distribuzione facendo riferimento alla classe SS e in particolare alla forma (3.1), assumiamo di essere nel caso univariato e

fissiamo $w(x)=\lambda x$, dove x è una realizzazione di una variabile casuale X e $\lambda \in \mathbb{R}$ è un parametro di forma. Allora, la **distribuzione normale asimmetrica**, proposta da Azzalini (1985), è una distribuzione di probabilità continua denotata, nella forma standard, con $SN(\lambda)$ e con una funzione di densità di probabilità:

$$f(x; \lambda) = 2\phi(x)\Phi(\lambda x), \quad -\infty < x < \infty, \quad -\infty < \lambda < \infty \quad (3.2)$$

dove $\phi(\cdot)$ e $\Phi(\cdot)$ indicano rispettivamente la funzione di densità e la funzione di distribuzione di una normale standard. Quando X assume la forma di una normale asimmetrica standard viene indicato con $X \sim SN(\lambda)$.

Mostriamo di seguito, alcune delle proprietà più importanti di quest'ultima funzione di densità.

- i. Se $\lambda=0$, otteniamo la densità $N(0,1)$
- ii. Se $X \sim SN(\lambda)$, allora $-X \sim SN(-\lambda)$
- iii. Quando $\lambda \rightarrow \infty$, la (3.2) converge puntualmente alla densità *half-normal*, cioè $2\phi(x)$ per $x \geq 0$
- iv. Se $X \sim SN(\lambda)$, allora $X^2 \sim X_1^2$.

Per altre proprietà e dimostrazioni si veda Azzalini (2005).

Una delle principali motivazioni, per cui abbiamo scelto di lavorare con la famiglia di distribuzioni *skew-normal*, è la gran quantità di vantaggi che ci fornisce questa famiglia parametrica di densità di probabilità, come ad esempio:

1. La più importante distribuzione della statistica, cioè la normale, è un punto interno alla distribuzione SN , e non un caso limite, come usualmente succede per altre famiglie parametriche; per questo motivo, la classe SN conserva alcune delle proprietà formali della classe normale.
2. Il semplice meccanismo con cui si riesce a descrivere e a modellare le variabili aleatorie è di grande interesse nella applicabilità di questa distribuzione, e ancor di più nella pratica per la modellazione dei dati, soprattutto se si preferisce considerare l'asimmetria non come un problema dei dati ma si vuole modellare anche questa tramite un modello statistico parametrico che la consideri.
3. La libera scelta di una qualsiasi funzione dispari $w(\cdot)$ di (3.1), fornisce un'ampia flessibilità, estendendo i risultati oltre alla sola asimmetria.

3.1.1 Parametrizzazione diretta e parametrizzazione centrata

La classe SN può essere generalizzata includendo i parametri di posizione e scala, identificati rispettivamente da ξ ed η . In questo modo se $X \sim SN(\lambda)$ definita da (3.2) con $\lambda \in \mathbb{R}$ e $Y = \xi + \eta X$ dove $\xi \in \mathbb{R}$ e $\eta > 0$, allora anche Y segue una distribuzione normale asimmetrica, dipendente in questo caso dai parametri (ξ, η, λ) e con funzione di densità:

$$f(y; \xi, \eta, \lambda) = \frac{2}{\eta} \phi\left(\frac{y-\xi}{\eta}\right) \Phi\left\{\lambda\left(\frac{y-\xi}{\eta}\right)\right\}; \quad -\infty < y < \infty \quad (3.3)$$

La variabile casuale Y viene quindi indicata tramite $Y \sim SN_D(\xi, \eta, \lambda)$, dove “ D piccolo” fa riferimento alla **parametrizzazione diretta**. Inoltre è noto che:

$$E\{Y\} = \xi + \eta b \delta, \quad \text{Var}\{Y\} = \eta^2 \{1 - (b \delta)^2\},$$

$$y_1 = \frac{4 - \pi}{2} \frac{(b \delta)^3}{(1 - (b \delta)^2)^{3/2}}, \quad y_2 = 2(\pi - 3) \frac{(b \delta)^4}{(1 - (b \delta)^2)^2},$$

dove $\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}} \in (-1, 1)$, $b = \sqrt{2/\pi}$ e y_1 e y_2 sono rispettivamente il coefficiente di asimmetria e di curtosi.

Per risolvere alcuni problemi inferenziali della parametrizzazione diretta (con $\lambda = 0$) si fa un cambio di parametri da (ξ, η, λ) a (μ, σ, y_1) , attraverso una standardizzazione della variabile X e creando una nuova variabile casuale Y_C , che viene indicata con la **parametrizzazione centrata** della distribuzione normale asimmetrica, dove i parametri diretti vengono espressi tramite i parametri centrati nel seguente modo:

$$\xi = \mu - c y_1^{1/3} \sigma, \quad \eta = \sigma \sqrt{1 + c^2 y_1^{2/3}}, \quad \lambda = \frac{c y_1^{1/3}}{\sqrt{b^2 + c^2 (b^2 - 1) y_1^{2/3}}}, \quad (3.4)$$

per $b = \sqrt{2/\pi}$ e $c = \{2/(4 - \pi)\}^{1/3}$ costanti.

Quindi, Y_C è una variabile casuale normale asimmetrica con parametrizzazione centrata che ha media $\mu \in \mathbb{R}$, deviazione standard $\sigma > 0$ e coefficiente di asimmetria $y_1 \in (-1, 1)$ e dove denotiamo tale relazione con $Y_C \sim SN_C(\mu, \sigma, y_1)$. La densità della variabile casuale Y_C è nella forma (3.3) mettendo al posto dei parametri diretti i corrispondenti delle relazioni definiti in (3.4).

3.2 Distribuzione normale asimmetrica sul cerchio unitario

Nella Tabella 3.1, abbiamo definito $f(\theta; \lambda)$ come una forma standard della distribuzione normale asimmetrica circolare tramite la funzione di densità e di ripartizione di una normale circolare, WN , con parametri $\mu=0$ e $\rho=1$. È da precisare che questo approccio non si trova in letteratura, anche se dalla costruzione del Lemma dato in (3.1) può essere ragionevole dato che si verificano le condizioni. Lo scopo della Tabella 3.1 è soltanto descrittivo e la definizione della normale asimmetrica sul cerchio unitario tramite la classe SS è da verificare magari in un lavoro futuro. Il modo in cui viene definita questa distribuzione in letteratura è tramite l'applicazione dell'approccio circolare, visto in 2.4.1, col quale possiamo “avvolgere” su un cerchio di raggio unitario la distribuzione SN , ottenendo così la distribuzione normale asimmetrica circolare (abbreviata con l'acronimo $WSNC$, dall'inglese: *Wrapped Skew-Normal on the Circle*). Questo è quello che viene fatto da Pewsey (2000) per la SN_D e da Pewsey (2006) per la SN_C e che noi seguiamo come riferimento per presentare questa distribuzione.

Sia Y una variabile aleatoria continua SN_D con funzione di densità definita da (3.3), possiamo allora ottenere, per la (2.36) e la (2.38) definite nel capitolo precedente, $\Theta = Y \pmod{2\pi}$ dove Θ ha una distribuzione **normale asimmetrica circolare** con funzione di densità di probabilità uguale a:

$$f(\theta; \xi, \eta, \lambda) = \frac{2}{\eta} \sum_{k=-\infty}^{\infty} \phi\left(\frac{\theta + 2\pi k - \xi}{\eta}\right) \Phi\left\{\lambda\left(\frac{\theta + 2\pi k - \xi}{\eta}\right)\right\}, \quad 0 \leq \theta < 2\pi \quad (3.5)$$

dove in questo caso $0 \leq \xi < 2\pi$, $\eta > 0$ e $-\infty < \lambda < \infty$. Rispettivamente ognuno di questi parametri controlla la posizione, la dispersione e la forma. Dipendendo dei valori di η e λ otteniamo alcuni casi speciali della distribuzione. Vediamo di seguito la Tabella 3.2 che riassume queste situazioni.

η	λ	Distribuzione
$\rightarrow 0$	qualsiasi	Un punto della distribuzione
$\rightarrow \infty$	qualsiasi	Uniforme Circolare
> 0	0	Wrapped Normal
> 0	$\rightarrow \pm \infty$	Wrapped half normal / negativa

Tabella 3.2: Casi limiti della WSNC con parametrizzazione diretta

Rappresentando la densità (3.5) possiamo capire meglio qual è la forma della distribuzione. Nella seguente Figura 3.1, fissiamo $\xi=0$ e $\eta=1$ e disegniamo la densità (3.5) per alcuni valori del parametro di asimmetria λ . È importante specificare che la densità di ogni punto $\theta \in [0, 2\pi)$ di una distribuzione circolare è l'area che insiste tra il cerchio unitario e la curva disegnata rappresentante di ogni forma del modello probabilistico con i parametri specificati.

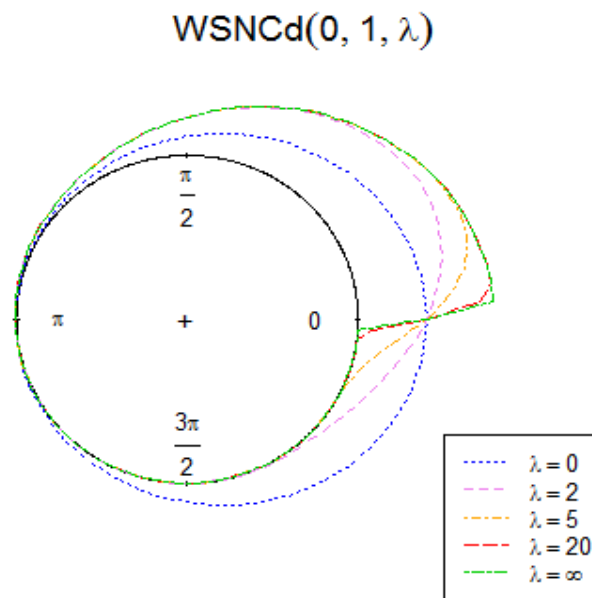


Figura 3.1: Densità nel cerchio unitario per $WSNC_D(0,1,\lambda)$

Precisiamo che per la stampa delle densità teoriche e dei modelli si userà la definizione formale di angolo matematico e cioè rappresentiamo l'origine nel punto (1,0) in coordinate cartesiane e seguiamo una rotazione anti-oraria. Nella stampa invece del campione, che vedremo nel Capitolo 4, si preferirà usare una rappresentazione in forma di bussola.

La $WSNC$ è allora un modello parametrico che riesce a captare le variazioni di asimmetria dei dati circolari e che verrà indicato, quando scritto nella forma (3.5), con

$\Theta \sim WSNC_D(\xi, \eta, \lambda)$ dove “ D piccolo”, ovviamente, continua ad indicare la parametrizzazione diretta.

3.2.1 Parametrizzazione centrata e circolare

Come abbiamo accennato prima, nel 3.1.1, la parametrizzazione diretta della distribuzione normale asimmetrica sulla retta reale è una parametrizzazione ridondante quando $\lambda=0$. Secondo Pewsey (2000), lo stesso problema risulta nel caso della $WSNC_D(\xi, \eta, \lambda)$ e le stesse implicazioni per l'uso della stima di massima verosimiglianza sono analoghe al caso lineare, quindi sarebbe ancora consigliabile usare la parametrizzazione centrata.

Pewsey (2006) propone di applicare ancora l'approccio circolare, spiegato in 2.4.1, in questo caso alla variabile casuale $Y_C \sim SN_C(\mu, \sigma, \gamma_1)$ definita tramite le relazioni di (3.4) applicate in (3.3), in modo da ottenere Θ_C variabile casuale circolare distribuita come una $WSNC$ con parametri centrati (μ, σ, γ_1) e con una densità:

$$f(\theta; \mu, \sigma, \gamma_1) = \frac{2}{\sigma \sqrt{1+c^2 \gamma_1^{2/3}}} \sum_{k=-\infty}^{\infty} \phi \left(\frac{1}{\sqrt{1+c^2 \gamma_1^{2/3}}} \left(\frac{\theta + 2\pi k - \mu}{\sigma} + c \gamma_1^{1/3} \right) \right) \cdot \Phi \left\{ \frac{c \gamma_1^{1/3}}{\sqrt{[b^2 + c^2(b^2 - 1) \gamma_1^{2/3}](1 + c^2 \gamma_1^{2/3})}} \left(\frac{\theta + 2\pi k - \mu}{\sigma} + c \gamma_1^{1/3} \right) \right\}, \quad 0 \leq \theta < 2\pi \quad (3.6)$$

Purtroppo nel cerchio, come già nominato nell'introduzione di questo capitolo, non esistono le solite trasformazioni di standardizzazione e quindi i parametri (μ, σ, γ_1) di (3.6) non sono gli equivalenti circolari di media, deviazione standard e coefficiente di asimmetria. Tuttavia, perseguendo la logica della parametrizzazione centrata Pewsey (2000) introduce gli analoghi circolari dei parametri centrati, ovvero i parametri circolari $(\mu, \rho, \bar{\beta}_2)$ che come per qualsiasi altra distribuzione circolare vengono definiti da (2.7) e (2.8) con $p=2$. Allora, scriveremo $WSNC(\mu, \rho, \bar{\beta}_2)$ per fare riferimento alla **parametrizzazione circolare** della distribuzione normale asimmetrica sul cerchio unitario.

3.2.2 Metodi di stima e altre riparametrizzazioni

Al fine di stimare i parametri della distribuzione normale asimmetrica circolare, è utile fare prima un chiarimento sulle diverse parametrizzazioni di tale modello.

Finora, per la distribuzione $WSNC$ sono state presentate tre diverse parametrizzazioni:

- quella diretta, tramite: $\xi \in [0, 2\pi)$, $\eta > 0$ e $\lambda \in \mathbb{R}$;
- quella centrata con: $\mu \in [0, 2\pi)$, $\sigma > 0$ e $\gamma_1 \in (-1, 1)$;
- l'equivalente a quest'ultima cioè quella circolare: $\mu \in [0, 2\pi)$, $\rho \in [0, 1]$ e $-1 \leq \bar{\beta}_2 \leq 1$.

Diverse volte in questo capitolo abbiamo accennato ad alcuni problemi inferenziali riguardo a $\lambda=0$ nel caso della SN_D e la $WSNC_D$, questo fatto ci ha portato a presentare una riparametrizzazione dei parametri centrati ai parametri diretti, vista in (3.4), per introdurre la parametrizzazione centrata della distribuzione SN sulla retta reale.

Come abbiamo indicato nel paragrafo precedente 3.2.1, Pewsey (2006) suggerisce che anche sul cerchio la parametrizzazione centrata è quella più idonea per fare inferenza basata sul metodo della massima verosimiglianza, le stime di verosimiglianza dei parametri centrati hanno un comportamento migliore rispetto alla parametrizzazione diretta e i valori stimati sono sempre ammissibili. Tuttavia ci avverte anche della possibilità di ottenere stime limite di $\gamma_1 \in (-1, 1)$ (soprattutto per piccoli campioni con alta dispersione o provenienti da una distribuzione fortemente asimmetrica). Inoltre, dato che la parametrizzazione centrata non equivale a nessuna quantità d'interesse nel cerchio, quando stimiamo i parametri della variabile casuale Θ_C , definita da (3.6), è ancora necessario fare una riparametrizzazione ai parametri diretti della distribuzione, che avviene tramite le stesse relazioni definite dalla (3.4). Viceversa, se può essere utile, è anche possibile calcolare i parametri centrati dai parametri diretti, usando le relazioni:

$$\begin{aligned} \mu \pmod{2\pi} &= \xi \pmod{2\pi} + b \eta \delta, & \sigma &= \eta \sqrt{1 - b^2 \delta^2}, \\ \gamma_1 &= \frac{b \delta^3 (2b^2 - 1)}{(1 - b^2 \delta^2)^{3/2}}, \end{aligned} \quad (3.7)$$

dove $b = \sqrt{2/\pi}$ e $\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}} \in (-1, 1)$.

Riguardo alla parametrizzazione circolare, $(\mu, \rho, \bar{\beta}_2)$, presentata nel paragrafo 3.2.1,

diciamo che tali parametri possono essere stimati banalmente tramite il metodo dei momenti, uguagliando semplicemente i momenti teorici ai momenti campionari. Questa parametrizzazione può essere ottenuta anche in termini dei parametri diretti; per questo scopo si usano alcune identità, dichiarate in Pewsey (2000) a partire dalla funzione caratteristica ottenuta come in (2.4) e (2.40), alle quali segue:

$$\mu = \tan^{-1} \left[\frac{\sin \xi + \zeta(\delta \eta) \cos \xi}{\cos \xi - \zeta(\delta \eta) \sin \xi} \right],$$

$$\rho = \omega \sqrt{1 + \zeta^2(\delta \eta)}, \quad (3.8)$$

$$\bar{\beta}_2 = \omega^4 \left(\frac{\zeta(2\delta \eta) \{1 - \zeta^2(\delta \eta)\} - 2\zeta(\delta \eta)}{1 + \zeta^2(\delta \eta)} \right),$$

dove $\omega = e^{-\frac{1}{2}\eta^2}$ e $\zeta(x) = \int_0^x b e^{u^2/2} du$ t.c. $\zeta(-x) = -\zeta(x)$ e $b = \sqrt{2/\pi}$.

Dunque i valori dei parametri diretti possono essere ottenuti da quelli circolari, risolvendo prima numericamente per $\delta \eta$ in:

$$\frac{\bar{\beta}_2}{\rho^4} = \frac{\zeta(2\delta \eta) \{1 - \zeta^2(\delta \eta)\} - 2\zeta(\delta \eta)}{(1 + \zeta^2(\delta \eta))^3}. \quad (3.9)$$

Il valore di η e anche quello di δ possono essere ottenuti usando:

$$\eta = \sqrt{-2 \log \rho + \log \{1 + \zeta^2(\delta \eta)\}}. \quad (3.10)$$

Infine il valore di $\xi \pmod{\pi}$ è la soluzione a:

$$\tan(\xi) = \left[\frac{\tan \mu - \zeta(\delta \eta)}{1 + \zeta(\delta \eta) \tan \mu} \right]. \quad (3.11)$$

Quest'ultima soddisfa la formula riguardo μ di (3.8).

In questo modo è possibile riparametrizzare anche i parametri circolari a parametri diretti. Tuttavia Pewsey (2006) ci avverte della possibilità che la stima del parametro diretto δ (usato per specificare la asimmetria λ) ottenuto attraverso i parametri circolari, possa essere non ammissibile ovvero che non rientri nei ranghi associati al parametro; comunque ci indica anche che il metodo dei momenti fornisce in generale una stima più precisa per i casi della *WSNC* relativamente dispersi e vicini alla simmetria.

Allora, per le stime dei parametri della normale asimmetrica circolare usiamo due diversi approcci:

1. Metodo dell'analogia, per stimare i parametri circolari della variabile aleatoria

direzionale $\Theta \sim WSNC(\mu, \rho, \bar{\beta}_2)$.

2. Metodo della massima verosimiglianza considerando che le osservazioni sono realizzazioni indipendenti⁸ ed identicamente distribuite secondo una variabile casuale Θ_c , con funzione di densità (3.6) e parametri centrati (μ, σ, γ_1) . La funzione di log-verosimiglianza viene calcolata tramite la somma dei logaritmi della (3.6), cioè:

$$\begin{aligned}
 l(\mu, \sigma, \gamma_1; \theta) = & n \log 2 - n \log \sigma - \frac{n}{2} \log(1 + c^2 \gamma_1^{2/3}) \\
 & + \sum_{i=1}^n \log \sum_{k=-\infty}^{\infty} \phi \left(\frac{1}{\sqrt{1 + c^2 \gamma_1^{2/3}}} \left(\frac{\theta + 2\pi k - \mu}{\sigma} + c \gamma_1^{1/3} \right) \right) \\
 & \cdot \Phi \left\{ \frac{c \gamma_1^{1/3}}{\sqrt{[b^2 + c^2(b^2 - 1)\gamma_1^{2/3}](1 + c^2 \gamma_1^{2/3})}} \left(\frac{\theta + 2\pi k - \mu}{\sigma} + c \gamma_1^{1/3} \right) \right\}
 \end{aligned} \tag{3.12}$$

Per la sommatoria da $-\infty$ a ∞ Pewsey (2006) indica che bastano i primi termini centrali, noi useremo da -5 a 5 . Per risolvere l'equazione di verosimiglianza usiamo il metodo Nelder-Mead per massimizzare, ovvero passando la log-verosimiglianza negativa, con valori iniziali dati a caso.

Dopo di che, col fine di trovare la stessa parametrizzazione sia per la verosimiglianza, sia per il metodo dei momenti, si riparametrizzano entrambe le stime ottenute dai parametri circolari e quelle dai parametri centrati, a parametri diretti. Per chiarire le formule presentate fin ora riguardo le diverse riparametrizzazioni, presentiamo di seguito la Tabella 3.3, che mostra come le tre possibili parametrizzazioni della *WSNC* possono essere ottenute attraverso le altre.

Riparametrizzazione di: (Parametri stimati)

A:	Diretti	Centrati	Circolari
Diretti	-	Riparametrizzazione di (μ, σ, γ_1) a (ξ, η, λ) . Tramite (3.4).	Riparametrizzazione di $(\mu, \rho, \bar{\beta}_2)$ a (ξ, η, λ) . Tramite (3.9)-(3.11).
Centrati	Riparametrizzazione di (ξ, η, λ) a (μ, σ, γ_1) . Tramite (3.7).	-	#
Circolari	Riparametrizzazione di (ξ, η, λ) a $(\mu, \rho, \bar{\beta}_2)$. (3.8)	#	-

Tabella 3.3: Riparametrizzazioni per le stime di *WSNC*.

⁸ Discuteremo questa forte assunzione nell'ultima sezione del prossimo capitolo.

Il simbolo # indica che la stima dei parametri centrati non ha un equivalente nel cerchio, come abbiamo visto in 4.3.1, e di conseguenza i parametri circolari non possono essere calcolati direttamente da quelli centrati e viceversa.

Pertanto sarà possibile riferirsi, generalmente, ai parametri diretti in entrambe i metodi di stima e saremo quindi in grado di poter disegnare e paragonare i diversi modelli ottenuti a seconda del metodo inferenziale utilizzato.

3.3 Distribuzione normale asimmetrica circolare modificata

Come già accennato nell'introduzione di questo capitolo e nel paragrafo precedente 3.2, la forma standard di questa distribuzione $f(\theta; \alpha_1, \alpha_2)$, descritta nella Tabella 3.1, non è quella che viene mostrata in questo punto. Tuttavia in questo caso facciamo comunque uso del Lemma, presentato in (3.1), di Azzalini (1985) e consideriamo che dalla forma standard $X \sim SN(\lambda)$ definita in (3.2) possiamo modificare $w(x) = \lambda x$ in $w(x) = \alpha_1 x + \alpha_2 x^3$. Ottenendo in questo caso:

$$f(x; \alpha_1, \alpha_2) = 2 \phi(x) \Phi(\alpha_1 x + \alpha_2 x^3), \quad -\infty < x < \infty, \quad -\infty < \alpha_1 < \infty, \quad -\infty < \alpha_2 < \infty \quad (3.13)$$

Attraverso la parametrizzazione diretta, possiamo ottenere l'analogo di (3.3) con le corrispondenti modifiche fatte in (3.12), in modo da avere $Y \sim SN_D(\xi, \eta, \alpha_1, \alpha_2)$. Allora, partendo da questa distribuzione e ripetendo i passi di Pewsey (2000) che abbiamo spiegato in 3.2 (e cioè usando l'approccio circolare tramite la (2.36) e la (2.38) del Capitolo precedente) possiamo ottenere la nuova distribuzione circolare:

$$f(\theta; \xi, \eta, \alpha_1, \alpha_2) = \frac{2}{\eta} \sum_{k=-\infty}^{\infty} \phi\left(\frac{(\theta + 2\pi k - \xi)}{\eta}\right) \Phi\left\{\alpha_1 \left(\frac{\theta + 2\pi k - \xi}{\eta}\right) + \alpha_2 \left(\frac{\theta + 2\pi k - \xi}{\eta}\right)^3\right\} \quad (3.14)$$

per $\theta \in [0, 2\pi)$, $\xi \in [0, 2\pi)$, $\eta > 0$ e $\alpha_1, \alpha_2 \in \mathbb{R}$. Che è analoga alla $WSNC_D(\xi, \eta, \lambda)$ di Pewsey (2000) definita da (3.5), con la sola modifica della aggiunta di un nuovo parametro che ci permette di osservare diverse forme della distribuzione. Dalla costruzione del Lemma (3.1) e dal approccio teorico circolare del Capitolo 2, possiamo verificare che la distribuzione *normale asimmetrica circolare*

modificata, che pensiamo possa essere indicata d'ora in avanti in questo elaborato con *WSSNC* per fare riferimento alla *WSNC* di Pewsey (2000) e alla classe *SS* presentata in Azzalini (2005), è ancora una densità di una variabile casuale continua circolare Θ .

Quindi, scriveremo $\Theta \sim WSSNC(\xi, \eta, \alpha_1, \alpha_2)$ per indicare che una generica variabile casuale circolare segue una distribuzione normale asimmetrica circolare modificata. Dove ognuno di questi parametri indica posizione, dispersione (nel caso dei primi due, come per la *WSNC*) e gli ultimi due controllano la forma della distribuzione (come vedremo di seguito in 3.3.1). La stima di questi parametri diretti, avverrà tramite il metodo della massima verosimiglianza, dove in questo caso la funzione di log-verosimiglianza viene calcolata tramite la somma dei logaritmi della (3.10) e come per la *WSNC*, usiamo i primi 11 termini centrati per la sommatoria infinita. Risolviamo l'equazione di verosimiglianza in R tramite il metodo Nelder-Mead per massimizzazione dato che passiamo la log-verosimiglianza negativa, con valori iniziali casuali.

Il valore medio, la varianza e altri momenti della distribuzione, dovrebbero essere calcolati a partire dalla funzione caratteristica definita in (2.4), tuttavia in questo elaborato non vengono forniti esplicitamente dato che non sono immediati e forse nemmeno calcolabili.

3.3.1 Forme della distribuzione

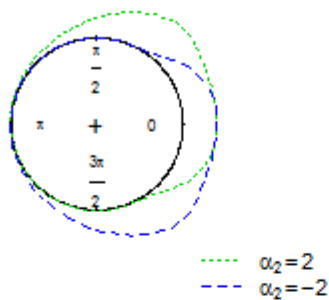
Col fine di vedere quali sono le modifiche rispetto alla *WSNC* e verificare se è possibile ottenere anche alcuni casi speciali di questa nuova distribuzione, *WSSNC*, rappresentiamo la forma della densità (3.10). Fissando i valori $\xi=0$ e $\eta=1$, come abbiamo fatto per la *WSNC*, e variando i valori di $\alpha_1, \alpha_2 \in \mathbb{R}$.

- Casi banali:
 - $\alpha_2=0$ equivale alla $WSNC_D(0, 1, \lambda)$.
 - $\alpha_1=\alpha_2=0$ equivale alla $WN(0, 1)$.

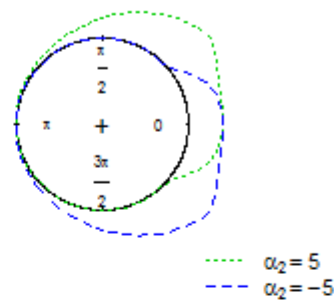
- 1^a prova: Figura 3.2.
 - Fissiamo $\alpha_1=0$ e diamo valori positivi e negativi ad α_2 .

Rappresentiamo le forme della distribuzione per $\alpha_2 > 0$ con le linee a puntini di colore verde e per $\alpha_2 < 0$ con le linee a trattini di colore blue. Si può osservare che entrambe le linee hanno forme analoghe però con l'orientamento della “gobba” da una parte del cerchio o dall'altra. La forma della distribuzione mantiene una piccola “gobba” indicata da α_1 , mentre le variazioni del secondo parametro di forma fanno sì che la distribuzione sia mano a mano più concentrata in un punto all'aumento di α_2 . Pertanto per valori di α_2 vicini allo zero abbiamo una forma più regolare dove si osservano comunque due zone di alta probabilità, per valori che tendono a infinito queste due zone sono maggiormente accentuate.

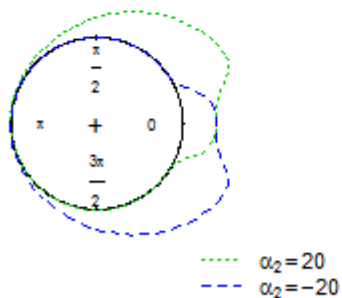
WSSNC(0, 1, 0, α_2)



WSSNC(0, 1, 0, α_2)



WSSNC(0, 1, 0, α_2)



WSSNC(0, 1, 0, α_2)

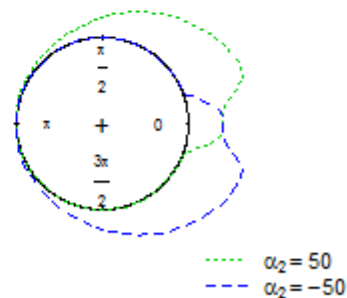


Figura 3.2: Forme della WSSNC per diversi valori di α_2 .

- 2^a prova: Figura 3.3.
 - Valori positivi e negativi per α_1 e α_2 .

Fissiamo nella prima riga di grafici rispettivamente, $\alpha_2=1, 5$, e facciamo variare il valore di α_1 tra -5 e 5, dove come nella Figura precedente, le curve di colore verde sono per i valori positivi e le curve di colore blue per i valori negativi. In questo caso però le forme non sono analoghe tra la parte superiore e la parte inferiore del cerchio. I puntini indicano $\alpha_1=-5, 5$ e i trattini $\alpha_1=-2, 2$. Il caso analogo viene realizzato nella seconda riga di grafici, fissando ora α_1 agli stessi valori e facendo variare α_2 nello stesso campo di variazione. La linea rossa a puntini-trattini indica in ogni caso il parametro che facciamo variare uguale a zero. In generale, si osserva che, se entrambi i valori dei due parametri di forma sono positivi, la distribuzione individua soltanto una direzione più probabile, invece se uno dei parametri di forma è positivo e l'altro è negativo la distribuzione riesce a catturare due direzioni più probabili, potendo essere possibile in questo modo adattare il modello a popolazioni che verificano bimodalità.

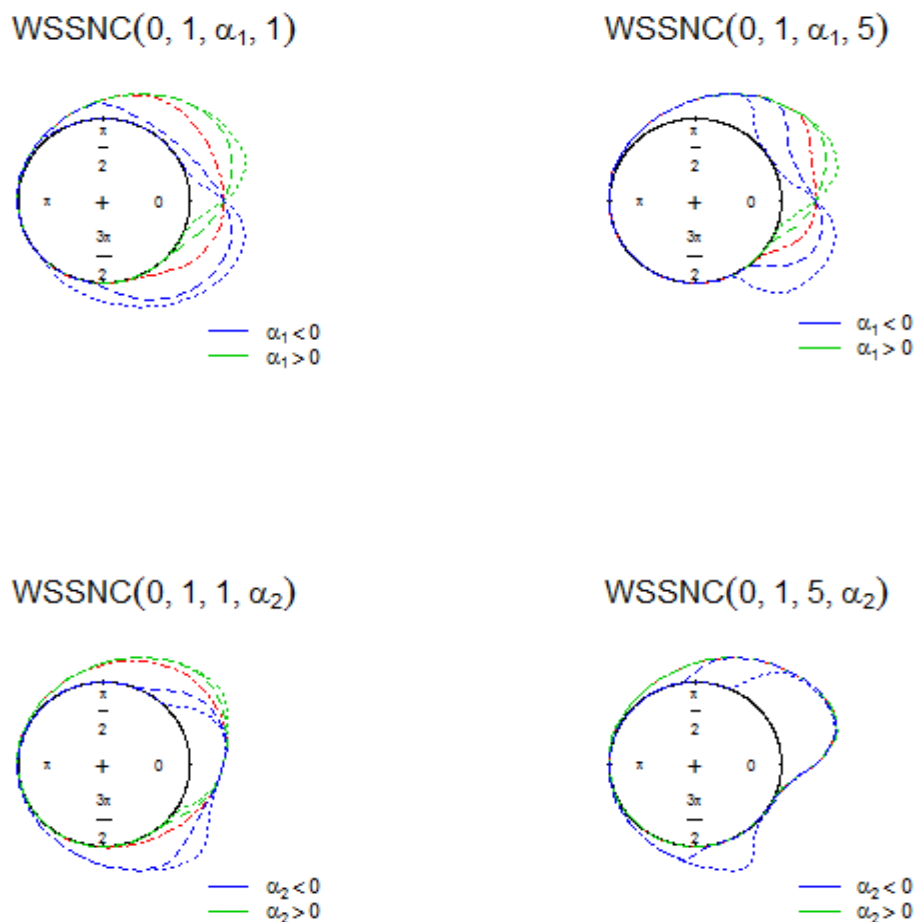


Figura 3.3: Forme della WSSNC per valori di $\alpha_j=(-5,-2,0,2,5)$ per $j=1,2$.

- 3^a prova: Figura 3.4.
 - $\alpha_1 \rightarrow \pm\infty$ e $\alpha_2 \rightarrow \pm\infty$.

In questa ultima prova fissiamo nella prima riga dei grafici rispettivamente $\alpha_2 \rightarrow \pm\infty$, si osserva che di nuovo le forme per $\alpha_2 < 0$ (grafico in alto a destra) sono simmetriche rispetto allo zero alle forme per i valori di $\alpha_2 > 0$ (grafico in alto a sinistra) e analogamente succede nella seconda riga dei grafici per i valori di α_1 fissato. Un'altra rilevante osservazione da fare è che se entrambi i parametri di forma tendono a $+\infty$ la distribuzione è una *Wrapped Half Normal* positiva ed è invece una *Wrapped Half Normal* negativa se entrambi tendono a $-\infty$. Inoltre quando $\alpha_1 \rightarrow \pm\infty$ per qualsiasi valore di α_2 , non molto grande, si verifica ancora la distribuzione *WHN* pos./neg. Nel caso contrario, quando $\alpha_2 \rightarrow \pm\infty$ per qualsiasi valore di α_1 , non molto grande, abbiamo quasi un *WHN* dove si vede anche una piccola “gobba”. Il caso particolare lo otteniamo quando uno dei parametri di forma tende a $+\infty$ e l'altro a $-\infty$, in questo caso si osserva ancora che il modello assegna alta probabilità a due direzioni ben definite, potendo essere corretto per campioni bimodali.

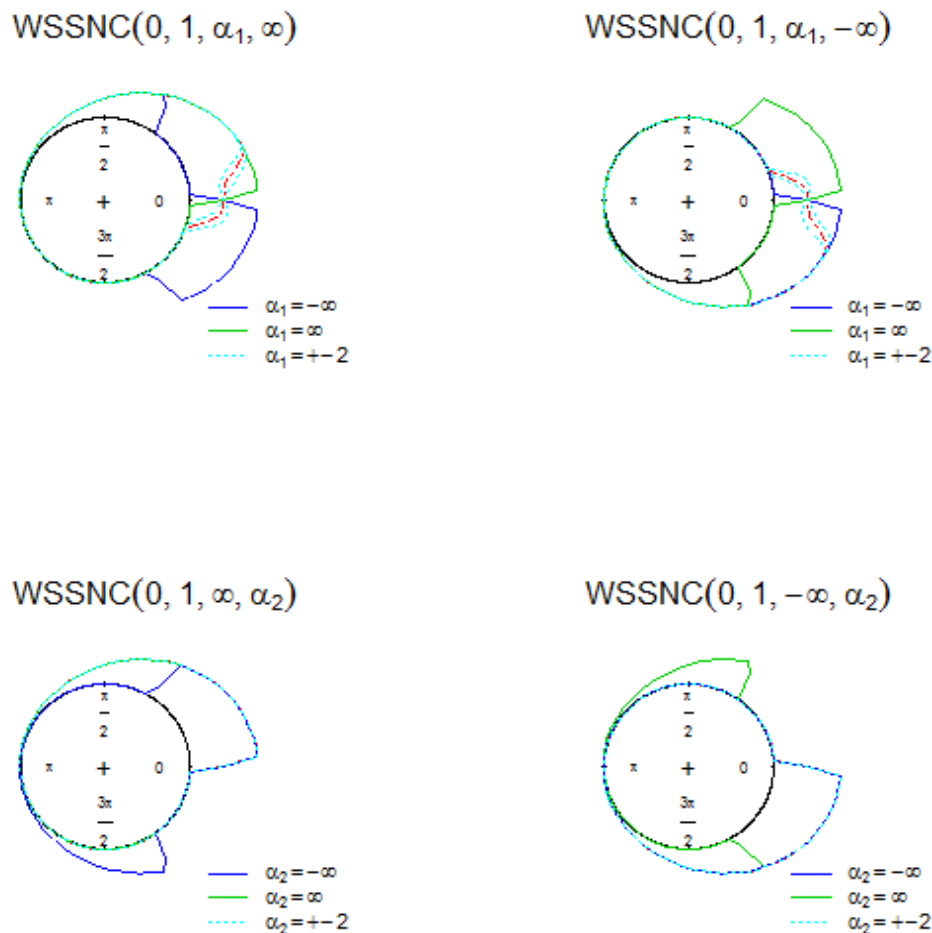


Figura 3.4: Forme della WSSNC per α_1 e α_2 che tendono a infinito

Dunque, viste le Figure 3.2-3.4, possiamo riassumere nella Tabella 3.4 i casi speciali della distribuzione $WSSNC$, in funzione dei valori che prendono α_1 e α_2 .

α_1	α_2	Distribuzione
0	0	Wrapped Normal
qualsiasi	0	Wrapped Skew Normal
0	qualsiasi	Figura 3.2.
$\rightarrow \pm \infty$	$\rightarrow \pm \infty$	Se sono dello stesso segno: WHN pos./neg.
qualsiasi	$\rightarrow \pm \infty$	Figura 3.4, grafico (1,1)
$\rightarrow \pm \infty$	qualsiasi	Wrapped Half Normal pos./neg.

Tabella 3.4: Casi limiti della $WSSNC$.

CAPITOLO 4

ANALISI STATISTICA

Dopo aver presentato nel Capitolo 2 i cenni teorici più importanti per analizzare la tipologia dei dati riguardanti le misurazioni delle direzioni, cioè i dati circolari, e nel Capitolo 3 i modelli statistici più affini alle “nostre” direzioni del vento, vogliamo procedere a modellare l'insieme di queste osservazioni usando una forma particolare dei modelli probabilistici specificati e cioè con i parametri stimati, in modo da ottenere graficamente forme concrete per le distribuzioni e vedere tramite (e a partire da) il campione dei dati di giugno 2009 se queste stime possono adattarsi bene ai nostri dati. Nel Capitolo 1 abbiamo già fatto una breve descrizione di questi dati, indicando la provenienza e le caratteristiche ambientali che ci aspettiamo in questo arco temporale scelto da noi. Nel presente capitolo invece, vogliamo adattare un modello statistico che riesca a captare la variabilità delle direzioni in tale mese nella determinata zona di studio.

Per realizzare questa analisi abbiamo utilizzato il software *R*, dove in particolare per dati circolari sono disponibili alcune librerie, come *CircStat* o *circular*, che permettono di visualizzare in modo appropriato tali dati e calcolare le statistiche riassuntive e le stime per i modelli più comuni. Per i modelli da noi scelti, c'è stato bisogno di creare le funzioni specifiche dato che non erano in queste librerie, tali funzioni e il resto del codice per l'analisi sono disponibili nell'appendice A.

Nel paragrafo 4.1, vedremo le particolarità di questo insieme di dati e in base a queste verificheremo l'individuazione dei modelli statistici più adatti presi in considerazione nel Capitolo 3. Nel paragrafo 4.2, faremo una stima puntuale dei parametri dei modelli scelti e discuteremo sui diversi metodi di stima a disposizione. Una volta stimati i parametri mostreremo graficamente l'adattamento dei modelli e decideremo se sono ragionevoli o meno in base all'analisi grafica risultante, poiché le due famiglie di distribuzioni scelte sono annidate, si può fare anche il test del rapporto di verosimiglianza che non varia per dati circolari. Infine, nel paragrafo 4.3,

questioneremo l'indipendenza assunta dell'insieme dei dati e mostreremo un approccio che potrebbe diminuire la correlazione tra le osservazioni.

4.1 Caratteristiche del campione e specificazione del modello statistico

Nella Figura 1.9 abbiamo già potuto vedere una descrizione del nostro campione, composto da 1833 direzioni del vento misurate in gradi azimut, che rappresentiamo di nuovo nella seguente Figura 4.1, mostrando in questo caso anche una visuale circolare dell'istogramma.

istogramma circolare con i punti osservati delle direzioni del vento

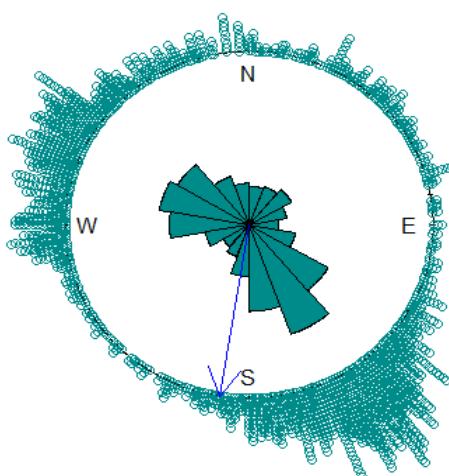


Figura 4.1: Osservazioni delle direzioni del vento nel mese di giugno 2009

In questo grafico circolare vediamo i dati raggruppati in classi di intervalli di 20° dove il raggio di ogni classe è una proporzione della radice quadrata delle relative frequenze cadenti dentro a tale intervallo. La freccia blu, rappresenta la direzione media del campione.

Come spiegato in 2.1, per ottenere le statistiche campionarie presentate in 2.3,

dobbiamo lavorare con il campione in radianti quindi moltiplichiamo i dati per $\pi/180$. Di seguito nella Tabella 4.1 riassumiamo i momenti campionari più importanti.

(a_1, b_1)	(\bar{a}_1, \bar{b}_1)	$(\bar{R}, \bar{\theta})$	V
(-0.209, -0.034)	(0.211, 0.000)	(0.211, 189°)	0.789

Tabella 4.1: Statistiche campionarie delle direzioni del vento di giugno 2009 in Villena.

Come ci aspettavamo la componente seno del momento centrale di ordine 1 (\bar{b}_1) è molto vicina a zero e invece la componente coseno di tale momento (\bar{a}_1) equivale alla lunghezza media risultante (\bar{R}). Tale lunghezza, insieme alla varianza (V), ci indica che le osservazioni sono abbastanza disperse tra di loro. La media direzionale campionaria ($\bar{\theta}$) ottenuta tramite (a_1, b_1) descrive una media sui 189° che equivale a una direzione molto vicina al sud anche se già nel quadrante del SSW (si veda la freccia della Figura 4.1).

La distribuzione empirica è chiaramente bimodale non simmetrica, quindi per modellare le direzioni del vento non sarebbe corretto usare i convenzionali modelli statistici per dati circolari, come la distribuzione di von Mises, la Normale circolare o altri che abbiamo presentato in 2.4, dato che sono distribuzioni unimodali simmetriche.

4.1.1 Asimmetria delle direzioni del vento

Per verificare la asimmetria dei nostri dati possiamo calcolare la componente seno del secondo momento trigonometrico campionario sulla direzione media definito da (2.20) e la misura di asimmetria campionaria definita da (2.31). (Tabella 4.2).

\hat{s}	\bar{b}_2
-0.588	-0.412

Tabella 4.2: Coefficienti di asimmetria campionari

In entrambi i casi non siamo vicino allo zero, quindi la distribuzione empirica non sembra essere simmetrica. Graficamente possiamo verificare questa caratteristica

vedendo se la distribuzione varia sotto la trasformazione $\bar{\theta} - \theta_i$, $i=1, \dots, n$. (Figura 4.2). La distribuzione delle osservazioni è cambiata rispetto alla Figura 1.9, rivelando che non si tratta di una distribuzione simmetrica rispetto alla media del campione. L'ipotesi di non usare un modello simmetrico per i nostri dati sembra essere giusta.

In questo modo, per le motivazioni date nel Capitolo 3 e le verifiche appena fatte, riteniamo sia ragionevole specificare come modello statistico una famiglia di distribuzioni simmetriche perturbate circolari, così come la normale asimmetrica circolare, Pewsey (2000), (anche se possiamo essere quasi certi che non si adatterebbe bene, visto che si riferisce a distribuzioni unimodali) presentata in 3.2, o quella che abbiamo chiamato normale asimmetrica modificata, presentata in 3.3. Nel punto 4.2 vedremo allora quale dei due modelli si adatta meglio ai nostri dati.

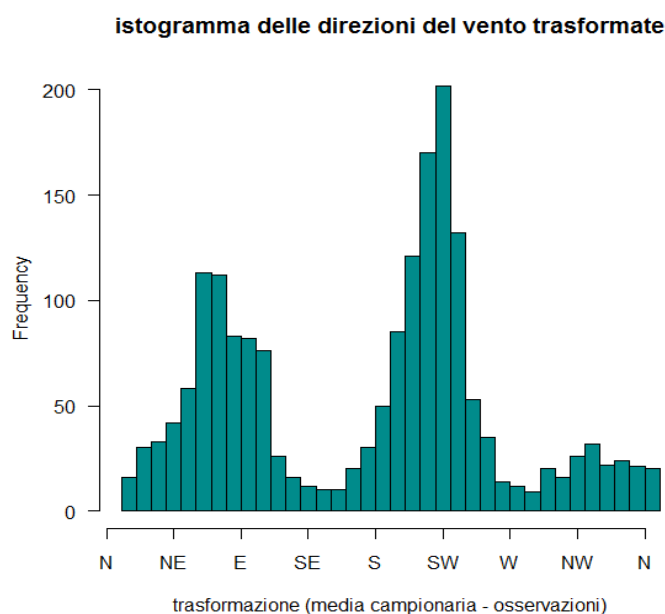


Figura 4.2: Trasformazione $\mu - \theta$ dei dati osservati

4.1.2 Bimodalità delle direzioni del vento

Come già visto precedentemente nella Figura 1.9 e nella Figura 4.1, il nostro campione è chiaramente bimodale. Per cercare di modellare questa bimodalità abbiamo proposto nel 3.3 un nuovo modello statistico, partendo da quello di Pewsey (2000) e cambiando la funzione di perturbazione con l'aggiunta di un nuovo parametro. In questo

paragrafo invece, mostriamo di seguito un altro possibile approccio più semplice per modellare i dati, concentrandoci sulla natura di questi. Vedremo dopo qualche analisi in che modo riusciamo a migliorare l'adattamento.

Un'idea più naturale infatti per considerare la bimodalità è pensare al fatto che ci possano essere due diversi gruppi dentro allo stesso campione, ipotesi molto ragionevole viste le considerazioni ambientali che abbiamo spiegato in 1.2.3, dove abbiamo indicato che ci aspettiamo una alta frequenza di brezza marina, insieme a brezza di valle, proveniente dal SE-S nelle ore più calde e una brezza di terra proveniente invece dal NW nella notte quando scendono le temperature.

Presentiamo dunque un grafico delle 1833 osservazioni delle direzioni del vento del mese di giugno 2009 in funzione dell'ora di misura, per vedere se veramente possiamo individuare i due diversi tipi di brezza rilevati nel Capitolo 1. Si veda la Figura 4.3, i dati forniti dalla stazione climatologica sono in orario solare e quindi non dobbiamo fare nessun tipo di trasformazione.

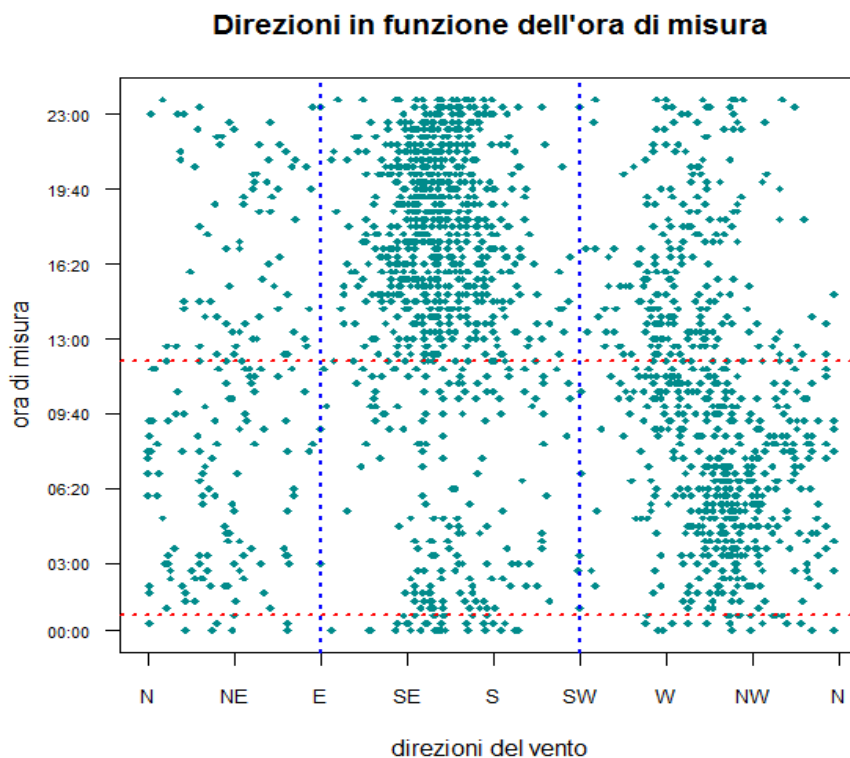


Figura 4.3: Separazione tra brezza di mare-valle e brezza di terra

Nella Figura 4.3, possiamo osservare ancora quali sono le direzioni più frequenti, che come avevamo già indicato si situano all'incirca del SSE e del NWN (si osservino le linee verticali blu che fanno una grossolana separazione), vedendo inoltre l'ora in cui si verifica ogni direzione con una rappresentazione globale di un'intera giornata. Dunque, le due grandi concentrazioni delle nostre direzioni “preferite” cadono approssimativamente, una all'incirca il primo pomeriggio fino ad poco dopo la mezzanotte, quella del SSE; l'altra invece nelle ore del mattino fino approssimativamente a mezzogiorno, quella del NWN. Si vedano le linee orizzontali rosse per una idea approssimata.

Separando il campione tra brezza di mare (che include anche quella di valle ma per semplicità d'ora in avanti la chiamiamo soltanto di mare) equivalente alle ore del pomeriggio-sera (che indichiamo con *ore calde*) e brezza di terra che è quella equivalente alle ore notturne-mattutine (indicate con *ore fredde*), possiamo distinguere il campione in due diversi gruppi rappresentati nella seguente Figura 4.4. Specificamente abbiamo scelto l'insieme dei dati appartenenti alla brezza marina tra le 12:20 e le 00:20 con un numero di osservazioni pari a 1081; invece il gruppo di brezza di terra, dalle 00:40 alle 12:00, ci resta soltanto con 752 osservazioni.

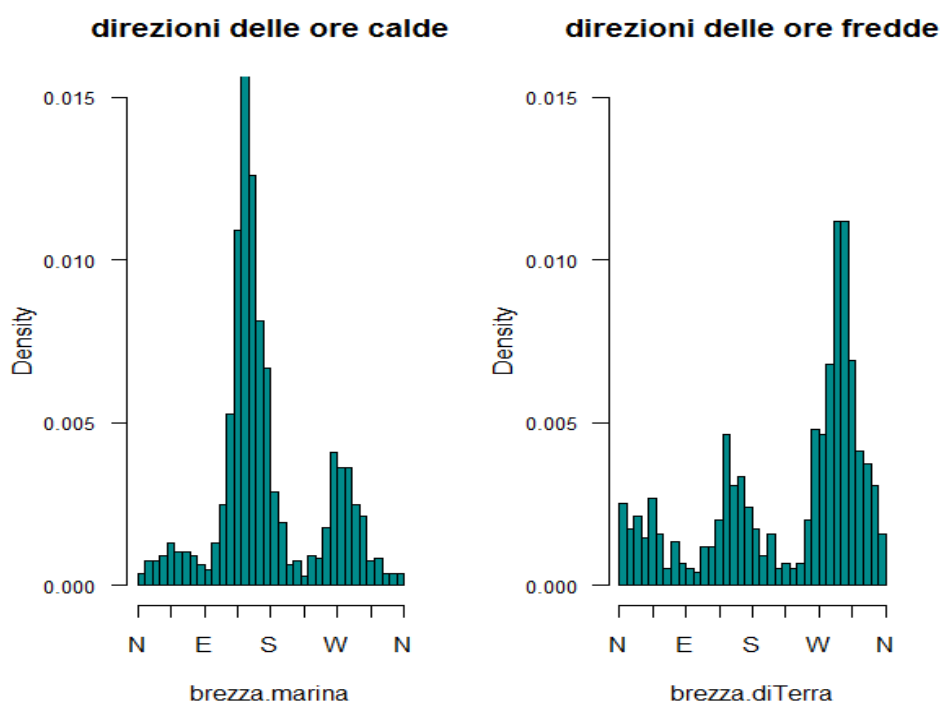


Figura 4.4: Istogrammi della brezza di mare e la brezza di terra

Nella precedente Figura 4.4 osserviamo che, anche se siamo riusciti ad individuare separatamente le due direzioni dominanti dell'intero campione, in entrambi i gruppi abbiamo ancora più di una direzione prevalente e la caratteristica della bimodalità sembra ancora essere presente. Ovviamente ogni giorno alla stessa ora non abbiamo sempre la medesima direzione, quindi i diversi venti di brezza non possono essere individuati soltanto dalle ore della giornata, ma dipenderanno anche da altri fattori, come sicuramente la temperatura; ci sono inoltre i venti di passaggio da un tipo di brezza all'altro che si potrebbero considerare.

In ogni caso visto che, tanto nelle ore fredde come nelle ore calde, la distribuzione potrebbe essere ancora bimodale, cerchiamo di adattare anche a questi due sottocampioni uno dei modelli statistici specificati, cioè *WSNC* o *WSSNC*, col fine di stabilire quale dei due modelli potrebbe adattarsi meglio a questi tipi di dati.

4.2 Stima e adattamento dei modelli circolari

Per ognuno dei campioni descritti in 4.1, siamo interessati ad ottenere una stima puntuale dei parametri dei modelli statistici specificati e cioè della famiglia di distribuzioni normale asimmetrica circolare (classe *WSNC*) e della famiglia di distribuzioni normale asimmetrica circolare modificata (classe *WSSNC*), presentate entrambe nel Capitolo 3. In questo modo, supponendo che il nostro sia un campione casuale semplice con osservazioni indipendenti (assunzione che verrà discussa meglio nel paragrafo 4.3) e che non si verifichino cambiamenti strutturali nell'area osservata, potremo fornire conclusioni su una intera popolazione ovvero, per il primo campione, sulle direzioni del vento nella zona di studio nel mese di giugno di un qualunque anno, e analogamente per le altre due sottopopolazioni di brezza di mare e brezza di terra (sulla base in ogni caso del mese di giugno 2009).

Per l'adattamento dei modelli a questi campioni, cioè per la stima dei parametri, si utilizza per il caso della *WSNC* il metodo dell'analogia per i parametri circolari e il metodo della massima verosimiglianza per i parametri centrati, dopo di che si farà una

riparametrizzazione di entrambe le stime a parametri diretti⁹ e se è possibile anche alle altre parametrizzazioni della distribuzione, come specificato in 3.2.2; per il caso della *WSSNC* si utilizza il metodo della massima verosimiglianza per i parametri diretti (come indicato in 3.3).

Precisiamo che, al contrario della Figura 4.1, per la stampa della stima dei modelli si usa una orientazione del cerchio anti-oraria e con l'origine nel punto (0,1) e cioè una stampa degli angoli come misurazioni matematiche e non come misurazioni azimut (come abbiamo già fatto nel Capitolo 3 per le forme della distribuzione normale asimmetrica circolare modificata). Per l'interpretazione del campione bisogna quindi considerare che lo zero equivale al Nord e di conseguenza π equivale al Sud, $\pi/2$ ad Est e $3\pi/2$ ad Ovest.

4.2.1 Vento di giugno

Di seguito presentiamo due tabelle riassuntive con le stime ottenute tramite i metodi nominati nell'introduzione di questa sezione. Nella prima tabella, corrispondente alle stime di *WSNC*, vengono evidenziate le celle delle stime originarie, ovvero massima verosimiglianza per i parametri centrati (indicando anche il valore della funzione di log-verosimiglianza) e metodo dei momenti sui parametri circolari, le altre celle non evidenziate sono una riparametrizzazione di quelle originarie. Avverrà analogamente per gli altri due campioni.

Le quantità giuste da confrontare per i due diversi metodi di stima sono alla fine le stime dei parametri diretti, dato che le riparametrizzazioni delle stime di massima verosimiglianza avvengono in senso discendente, cioè prima si fa la riparametrizzazione dei P.Centrati a P.Diretti e poi di quest'ultimi a P.Circolari. Nel caso del metodo dei momenti, avviene invece in senso ascendente partendo dalla cella evidenziata.

⁹ Per trasformare i parametri circolari a parametri diretti, si richiedono metodi di integrazione numeriche e tecniche per risolvere equazioni non lineari. Si vedano dalla (3.9)-(3.11). La funzione in R creata per tale scopo è stata impostata come adattamento del codice FORTRAN (tramite la creazione di librerie .dll) di Pewsey (2006).

$WSNC$	Massima Verosimiglianza	Metodo dei Momenti
P. Centrati	(4.535, 1.850, 0.988) con -3106.696	-
P. Diretti	(2.092, 3.065, 23.386)	(1.730, 6.161, -)
P. Circolari	(3.633, 0.303, -0.002e ⁻⁰⁴)	(3.301, 0.211, -0.412)

Tabella 4.3: Stima dei parametri della $WSNC$ delle direzioni del vento 2009

Dunque, dalla Tabella 4.3 possiamo osservare che i problemi di stima discussi in 3.2.2 si verificano in questo campione. La stima di massima verosimiglianza dei parametri centrati fornisce una stima limite del parametro di asimmetria ($\gamma_1 \rightarrow 1$) dovuta magari alla alta dispersione del campione o alla forte asimmetria del medesimo; la stima del metodo dei momenti per i parametri circolari produce invece valori fuori dei limiti nella riparametrizzazione a parametri diretti ($\delta > 1$). Nella seguente Figura 4.5 rappresenteremo graficamente la forma della $WSNC$ con i parametri diretti ottenuti dalle stime di M.V.

La seconda tabella, che in questo campione è la Tabella 4.4, mostra soltanto la stima dei parametri diretti utilizzando il metodo della massima verosimiglianza.

$WSSNC$	Massima Verosimiglianza
$(\xi, \eta, \alpha_1, \alpha_2)$	(1.533, 2.404, 3.758, -6.744) con -3057.609

Tabella 4.4: Stima dei parametri della $WSSNC$ delle direzioni del vento 2009

Rappresentiamo dunque, sulla medesima Figura 4.5, la forma anche della $WSSNC$ con i parametri stimati (che secondo le prove sperimentate in 3.3.1, individuerà sicuramente la bimodalità dato che abbiamo $\alpha_1 > 0$ e $\alpha_2 < 0$).

Nella seguente Figura 4.5, vediamo allora come il modello $WSNC$ riesce a evidenziare adeguatamente la asimmetria del campione, tuttavia stima una alta probabilità per osservazioni sul 3° quadrante che sembrano invece non verificarsi con tanta frequenza secondo l'insieme di dati osservati. Il modello $WSSNC$ riesce invece ad individuare meglio le due direzioni più frequenti del nostro campione, anche se stima con alte probabilità direzioni su $\pi/2$ che non sono osservabili.

campione giugno 2009 con stima dei modelli

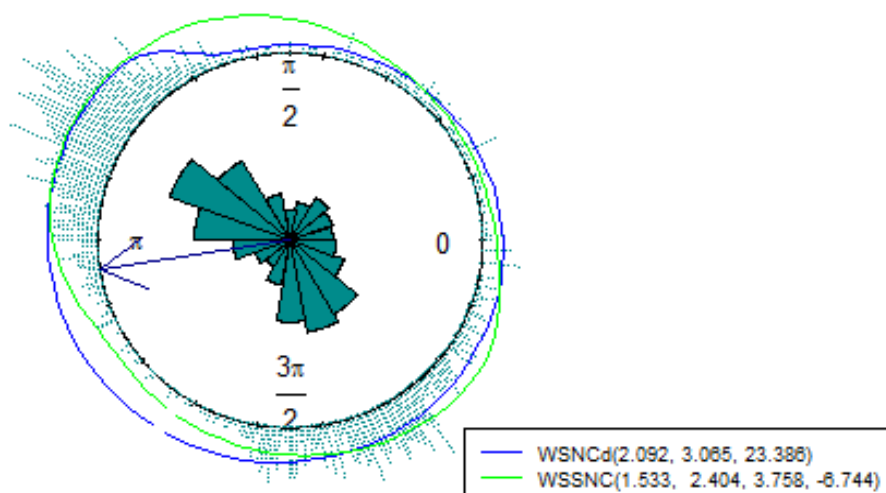


Figura 4.5: Adattamento della WSNC e la WSSNC per il campione delle direzioni del vento di giugno 2009

4.2.2 Brezza di mare

WSNC	M.V.	M.M.
P. Centri	(3.719, 1.314, 0.966) con -1640.013	-
P. Diretti	(1.997, 2.166, 11.705)	(1.429, 2.158, -)
P. Circolari	(3.365, 0.477, -0.0004)	(2.810, 0.515, -0.273)

Tabella 4.5: Stima dei parametri della WSNC delle direzioni della brezza di mare 2009

Come nel caso precedente, la stima di M.V. dei parametri centrati fornisce $\gamma_1 \rightarrow 1$ e la stima del metodo dei momenti per i parametri circolari produce valori fuori dei limiti ($\delta > 1$) nella riparametrizzazione a parametri diretti.

WSSNC	Massima Verosimiglianza
$(\xi, \eta, \alpha_1, \alpha_2)$	(3.202, 1.490, -5.023, 6.565) con -1494.236

Tabella 4.6: Stima dei parametri della WSSNC delle direzioni della brezza di mare 2009

Per confrontare i due modelli rappresentiamo la forma della WSNC con i parametri

diretti ottenuti dalla riparametrizzazione dalle stime dei parametri centrati attraverso il metodo della M.V. e la forma della WSSNC con i parametri diretti stimati. Vediamo di seguito la Figura 4.6.

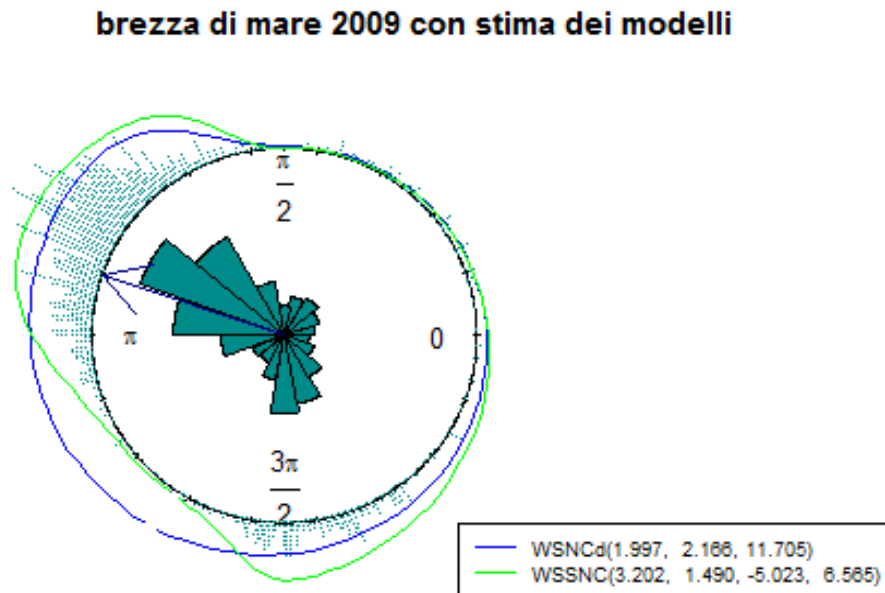


Figura 4.6: Adattamento della WSNC e la WSSNC per il campione di brezza di mare 2009

In questo caso dato che la direzione dominante è molto più frequente di quella verso Ovest (osservabile anche in Figura 4.4), il modello normale asimmetrico circolare modificato riesce a individuare molto bene le due direzioni più frequenti. Il modello normale asimmetrico circolare invece continua a dare una elevata probabilità alle osservazioni sul terzo quadrante, che tuttavia in questo campione non sono quasi presenti.

4.2.3 Brezza di terra

WSNC	M.V.	M.M.
P. Centriati	(0.679, 1.867, 0.991) con -1267.391	(4.771, 1.964, -0.449)
P. Diretti	(4.495, 3.094, 32.898)	(0.485, 2.814, -1.936)
P. Circolari	(6.038, 0.299, $-0.001e^{-04}$)	(5.248, 0.375, 0.120)

Tabella 4.7: Stima dei parametri della WSNC delle direzioni della brezza di terra 2009

Per il campione di brezza di terra, la riparametrizzazione a parametri diretti delle stime del M.M. dei parametri circolari sono ammissibili, ovvero $\delta = -0.943$ è interno ai limiti e allora possiamo calcolare λ . Questo ci permette di paragonare entrambe le stime di M.V. e M.M., anche se sappiamo che lo stimatore di massima verosimiglianza, a parte le sue buone proprietà, è più potente e migliore rispetto a quello del M.M.

<i>WSSNC</i>	Massima Verosimiglianza
$(\xi, \eta, \alpha_1, \alpha_2)$	(0.147, 1.835, -2.942, 2.412) con -1201.635

Tabella 4.8: Stima dei parametri della *WSSNC* delle direzioni della brezza di terra 2009

In Figura 4.7 è disegnata la forma della *WSSNC* con i parametri stimati riportati nella Tabella 4.8. Presentiamo allora in questo caso, tre diverse stime per il campione in indagine e per una maggiore chiarezza mostreremo quattro diversi grafici.

Nei primi due vediamo la forma del modello normale asimmetrico circolare con le stime fornite dal metodo della M.V. (primo grafico a sinistra) e dal metodo dei momenti (secondo grafico a destra). Confrontando questi due metodi di stima, vediamo che la massima verosimiglianza fornisce una approssimazione migliore per il nostro campione, com'era da aspettare, accentuando bene la direzione prevalente; mentre le stime del M.M., sembrano dare quasi una uniformità di equiprobabilità a tutte le osservazioni quindi non appare una buona stima (forse dovuto al fatto che, anche se ammissibile, $\delta \rightarrow -1$). Il modello *WSSNC* (secondo grafico a sinistra) riesce a catturare le due direzioni prevalenti dando anche una maggiore probabilità alle direzioni del Nord che, come abbiamo visto in Figura 4.4, per questa frangia oraria possono essere anche abbastanza frequenti. Nell'ultimo grafico, quello più a destra nella seconda riga, vediamo le tre stime sovrapposte, con la freccia blu che indica la direzione media stimata tramite i metodi dei momenti, in questo modo osserviamo che comunque tutti e tre i modelli danno più probabilità ad una specifica direzione che coincide anche con quella più frequente nel nostro campione.

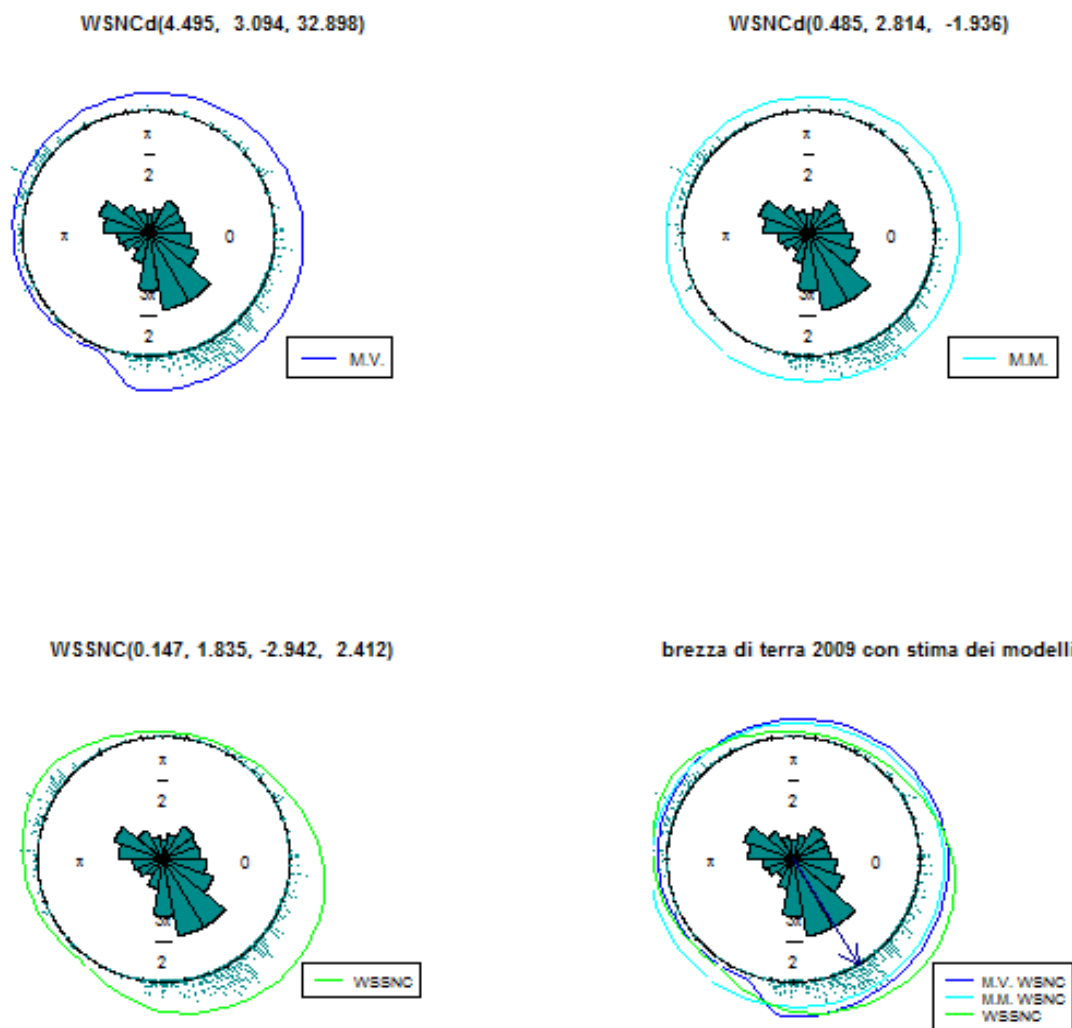


Figura 4.7: Adattamento della WSNC e la WSSNC per il campione di brezza di terra 2009

4.2.4 Test rapporto di verosimiglianza

Col fine d'individuare quali dei due modelli probabilistici specificati è più adatto ai nostri dati, vorremo risolvere il seguente contrasto d'ipotesi:

$$\begin{cases} H_0: \Theta \sim WSNC \\ H_1: \Theta \sim WSSNC \end{cases} \equiv \begin{cases} H_0: \alpha_2 = 0 \\ H_1: \alpha_2 \neq 0 \end{cases} \quad (4.1)$$

ovviamente per ognuno dei campioni, con i corrispondenti parametri stimati usando la M.V.; in modo che, per le teorie ottimali della verifica d'ipotesi, sappiamo che se il rapporto delle verosimiglianze è monotono rispetto ad una certa funzione dei dati (cioè,

una statistica), potremo valutare in base a questa il grado di significatività contro H_0 . Cioè, risolviamo il problema del contrasto definito in (4.1) tramite la statistica test del rapporto di verosimiglianza (abbreviato con TRV).

Allora, supponendo che sia $\hat{\lambda}$ il vettore parametrico stimato, il TRV sarebbe:

$$W = -2 \log \frac{L_0(\hat{\lambda})}{L_1(\hat{\lambda})} = -2(l_0(\hat{\lambda}) - l_1(\hat{\lambda})) \quad (4.2)$$

dove $l_j, j=0,1$, è il logaritmo del valore della funzione di verosimiglianza. W è una funzione dei dati monotona crescente che sotto H_0 , per la teoria asintotica della verosimiglianza e in condizioni di regolarità, sappiamo che converge in distribuzione, quando $n \rightarrow \infty$, a una variabile casuale Chi-quadrato con $p-q$ gradi di libertà, dove p è il numero di parametri della distribuzione sotto H_0 e q è il numero di parametri della distribuzione sotto H_1 , ovvero sia Ω lo spazio parametrico di una data classe di distribuzioni, nel nostro caso abbiamo $\Omega_0 \in \mathbb{R}^3$ e $\Omega_1 \in \mathbb{R}^4$. Allora per la verifica d'ipotesi data in (4.1), $W \sim \chi^2$, e in base a questa distribuzione possiamo calcolarne il livello di significatività osservato, cioè:

$$p\text{-value} = \Pr(W > w_{oss}) \quad (4.3)$$

Procediamo quindi a contrastare per ognuno dei nostri campioni, l'ipotesi nulla presentata in (4.1), dove per il calcolo del w_{oss} di ogni insieme di dati usiamo i valori riportati delle funzioni di log-verosimiglianza nelle corrispondente tabelle precedenti. In questo modo, il valore p dato in (4.2) di ogni campione è:

campione	p-value
Vento di giugno	p.v $\simeq 0$
Brezza di mare	p.v $\simeq 0$
Brezza di terra	p.v $\simeq 0$

Tabella 4.9: p-value di ogni campione per WSSNC contro WSNC

Quindi, in base ai dati ottenuti della Tabella 4.9 e fissando un probabilità di errore del 1%, possiamo dire che H_0 viene in ogni caso rifiutata; pertanto sembrerebbe che la vera legge probabilistica delle direzioni del vento possa appartenere all'insieme delle famiglie di distribuzioni della nuova classe WSSNC.

4.3 Dipendenza temporale

Per fare inferenza su un campione di dati è necessaria l'indipendenza tra le osservazioni, dato che, se questa condizione non viene verificata, i risultati delle analisi potrebbero non essere validi. L'assunzione dell'indipendenza è quindi una premessa molto importante che deve essere presa in considerazione. Nel nostro caso, difficilmente questa condizione può essere assicurata dato che lavoriamo con una serie temporale di misurazioni fatte ogni 20 minuti e nelle serie storiche temporali è naturale che quello che succede “ora” sia relazionato con quello che succederà “dopo”; le osservazioni dunque raccolte sequenzialmente nel tempo sono di solito dipendenti ed è quindi necessario studiare le relazioni dinamiche che ci sono nel tempo. Sfortunatamente per dati circolari non sono tanti gli studi fatti sulla correlazione temporale, soltanto Fisher (1993) parla di come analizzare dati direzionali con questo tipo di struttura, purtroppo però non sono disponibili implementazioni a computer delle procedure di analisi di serie temporali per dati circolari.

Dunque, dato che non possiamo capire nemmeno quanto grandi possano essere le ampiezze delle oscillazioni nel tempo (dato che possono variare in funzione dell'origine scelta) e non abbiamo a disposizione gli strumenti che ci permettono di vedere se esistono relazioni anche tra osservazioni più lontane nel tempo, come la funzione di autocorrelazione; ci aspettiamo che se la dipendenza temporale si verifica solo per periodi relativamente corti, la selezione di punti più distanti nel tempo possa contribuire ad eliminare tale dipendenza. Quindi può essere ragionevole aumentare la numerosità campionaria, prendendo altre misurazioni sotto le stesse condizioni e cioè nei mesi di giugno di altri anni, e selezionare, da questo grande insieme di dati, osservazioni ogni 3 ore piuttosto che ogni 20 minuti, in modo che magari si riesca ad avere una minore presunta dipendenza temporale.

Pertanto prendiamo anche le direzioni del vento in *Villena* nel mese di giugno dell'anno 2008 e 2010 in modo da ottenere un campione totale di 6480 osservazioni (cioè le 2160 osservazioni di un mese per tre mesi). Di questo insieme, prendiamo una direzione ogni nove (la proporzione non altera la struttura probabilistica dell'intero campione) e rimangono quindi 720 osservazioni, più lontane nel tempo tra di loro, che possono

dunque essere indipendenti. Togliendo i dati mancanti e le misurazioni dove abbiamo vento a velocità zero, rimaniamo con solo 588 dati, dei quali rappresentiamo di seguito, nella Figura 4.8, l'istogramma di probabilità messo assieme al campione del solo 2009 per mostrare che i dati sono simili.

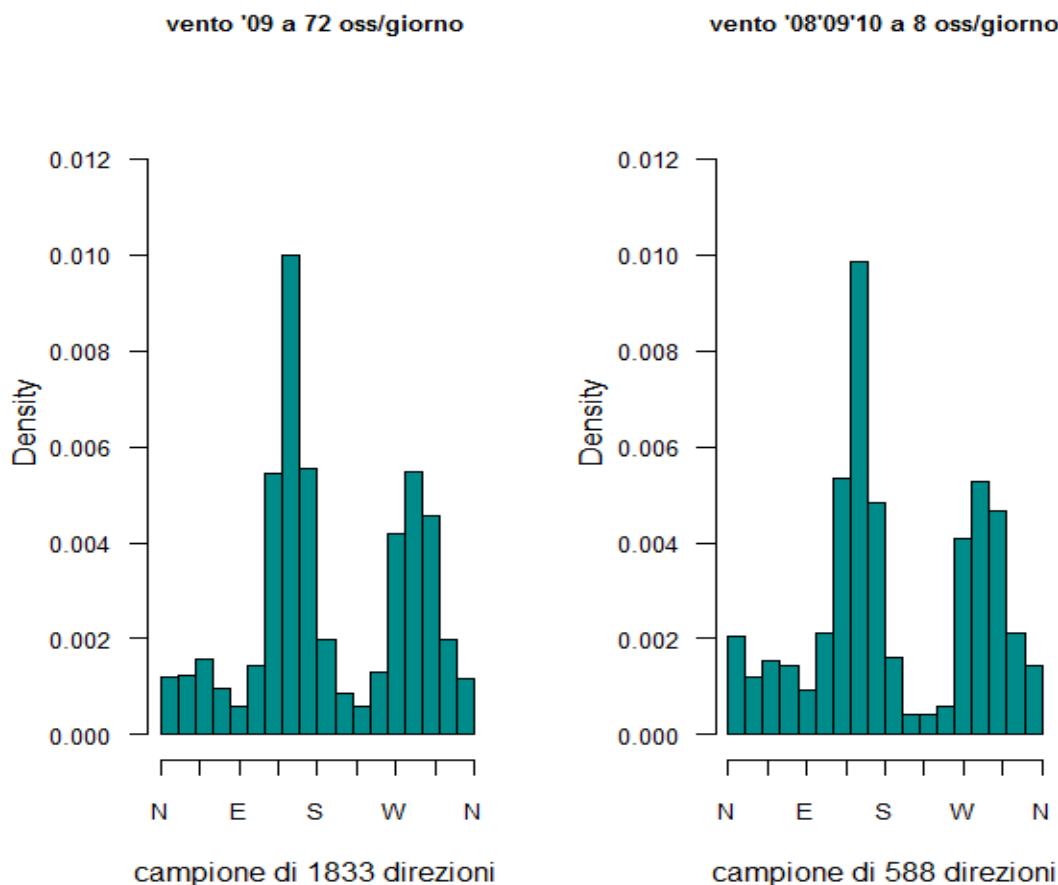


Figura 4.8: Campioni delle direzioni del vento in Villena del 2009 e del 2008, 2009 e 2010

Nella Figura 4.8 abbiamo preso intervalli di direzioni di 20° e come si può osservare, le differenze tra i due campioni sono sottilmente apprezzabili, soltanto per la direzione Nord abbiamo un po' più di frequenza, dovuta magari ai venti puntuali entranti dalla valle di *Benejama*, come abbiamo commentato nella Figura 1.8. Come prima approssimazione, possiamo quindi usare i risultati presentati nel paragrafo 4.2.1 per descrivere le direzioni del vento; comunque ripetiamo le analisi per questo nuovo campione (che si ipotizza, più correttamente, con osservazioni indipendenti) per verificarlo.

<i>WSNC</i>	M.V.	M.M.
P. Centriati	(5.058, 2.279, 0.989) con -1029.478	-
P. Diretti	(2.047, 3.776, 27.718)	(-1.681, 9.964, -)
P. Circolari	(3.614, 0.229, -0.001e ⁻⁷)	(3.031, 0.145, -0.411)

Tabella 4.10: Stima dei parametri della *WSNC* di una proporzione delle direzioni del vento di 2008, 2009 e 2010

I parametri stimati sono un po' cambiati rispetto a quelli riportati nella Tabella 4.3, le stime di M.V. dei parametri centriati sono sottilmente più elevate, vedremo di seguito nella Figura 4.9 se con questi valori stimati se apprezzano cambiamenti di forma nella distribuzione. I parametri circolari sono praticamente analoghi.

<i>WSSNC</i>	Massima Verosimiglianza
$(\xi, \eta, \alpha_1, \alpha_2)$	(4.544, 2.469, 9.455, -38.123) con -957.0502

Tabella 4.11: Stima dei parametri della *WSSNC* di una proporzione delle direzioni del vento di 2008, 2009 e 2010

La stima del parametro di scala $\eta > 0$ rimane invariata, invece le stime degli altri parametri cambiano significativamente rispetto ai parametri stimati del campione delle direzioni del vento di giugno 2009, con valori riportati in Tabella 4.4. Nella seguente Figura 4.9, possiamo osservare cosa implica questo cambiamento.

campione ridotto giugno '08'09'10 con stima dei modelli

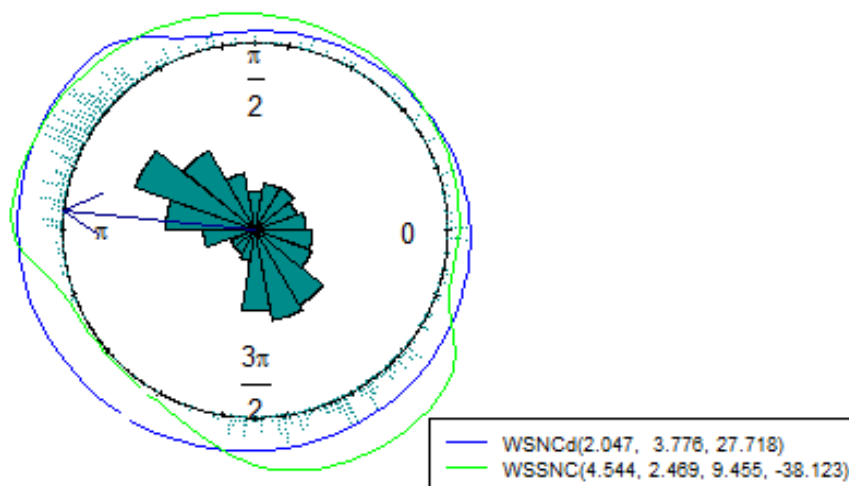


Figura 4.9: Adattamento della *WSNC* e la *WSSNC* per il campione ridotto delle direzioni del vento di giugno 2008, 2009 e 2010

In quanto visto nella Figura 4.5, possiamo dire che le sottili differenze di stima dei parametri del modello *WSNC* riportano risultati analoghi. Invece, le nuove stime di *WSSNC* sembrano migliorare per la descrizione del nostro fenomeno empirico, dato che si assegna probabilità maggiore alle due direzioni dominanti, SE e NW, e poca probabilità alle direzioni poco frequenti, come quelle del SW.

CONCLUSIONI

In questo lavoro di tesi abbiamo presentato una possibile soluzione per modellare dati circolari bimodali.

In particolare, abbiamo studiato le direzioni del vento di una determinata zona della Spagna, e abbiamo provato a modellare la bimodalità che presenta questo insieme di dati, in due diversi modi.

- Uno, centrato sulla natura del campione; differenziando tra la brezza di mare e la brezza di terra. Tuttavia, questo approccio non è stato sufficiente per dividere le due direzioni prevalenti, e in ogni nuovo campione di brezza di mare e brezza di terra si osservano ancora due diverse direzioni dominanti.
- L'altro invece, cercando un modello statistico capace di modellare dati nella scala originaria e di adattare e descrivere correttamente, in senso probabilistico, il comportamento del nostro fenomeno empirico circolare.

Abbiamo quindi sviluppato il secondo approccio e tale ricerca ci ha portato, tramite lo studio della statistica direzionale, a identificare la distribuzione nota normale asimmetrica circolare, che abbiamo potuto modificare grazie alla flessibilità della classe SS dalla quale proviene. Cioè, tramite il Lemma che definisce le distribuzioni simmetriche perturbate è stato possibile aggiungere un nuovo parametro di forma al modello probabilistico $WSNC$, in questo modo abbiamo sperimentato come può comportarsi la nuova distribuzione circolare di probabilità secondo i diversi valori di questo nuovo parametro di forma. Alla fine, si è verificato che il modello introdotto, $WSSNC$, riesce ad individuare anche la bimodalità.

Sul nostro insieme di dati, abbiamo paragonato il modello normale asimmetrico circolare e la variante sviluppata e la stima dei parametri di questi due modelli, in base al nostro campione, ci ha fornito adattamenti probabilistici ragionevoli per entrambi i modelli ma con una preferenza sulla modifica della distribuzione normale asimmetrica circolare (scelta corroborata anche del test rapporto di verosimiglianza).

Avremo voluto anche verificare l'indipendenza dei dati osservati, tuttavia lo sviluppo di questo settore della statistica, riguardo ai dati circolari, non è attualmente molto diffuso e molti degli approcci non sono ancora disponibili computazionalmente. Comunque abbiamo presentato un ulteriore approccio empirico per il presunto problema di osservazioni dipendenti e abbiamo osservato che, alla fine i risultati sono approssimativamente analoghi, al meno riguardo la *WSNC*. Si considera allora che come prima approssimazione, possiamo usare i modelli stimati sulla base dei campioni che ipotizzano dipendenza per descrivere le direzioni del vento.

Considerando che la variabilità delle osservazioni, viene in generale descritta meglio dalla nuova distribuzione *WSSNC*, verifichiamo che l'applicazione di questo modello alle direzioni del vento in *Villena* nel mese di giugno riesce a spiegare bene le due direzioni prevalenti del vento, che come visto nel Capitolo 1 per le caratteristiche geomorfologiche della zona in indagine, sono quelle provenienti del S-SE e del NW che equivalgono alla brezza di valle-mare e alla brezza di terra.

Concludendo, possiamo indicare che lo studio di questo insieme di dati ambientali ci ha portato a due considerazioni importanti:

- In termini climatologici; alla verifica delle circolazioni estive di brezza di mare e brezza di terra, non sempre percepite dal grande pubblico, dato che solitamente vengono considerate per il loro effetto di mitigare soltanto le temperature della costa, mentre come abbiamo visto nel nostro caso, si possono anche verificare in luoghi relativamente lontani dal mare.
- In termini statistici, abbiamo individuato una nuova e flessibile distribuzione adatta a dati circolari e inoltre, in base al nostro campione possiamo dire, anche come una buona stima per insiemi di dati bimodali.

Ci auguriamo che tale distribuzione possa essere magari sviluppata meglio in future ricerche, studiandone le proprietà formali e vedendo come si comporta con altri insiemi di dati.

APPENDICE A

CODICE R

A.1 Distribuzione normale asimmetrica circolare

Si riporta il codice R per il calcolo delle diverse riparametrazioni della *WSNC*. Inoltre si presentano anche i calcoli della funzione di densità e funzione di log-verosimiglianza per la parametrizzazione diretta e la parametrizzazione centrata.

```
#FUNZIONE PER CALCOLARE PARAMETRI CENTRATI: (mu, sigma, gamma1)
#DAI PARAMETRI DIRETTI: (xi, eta, lambda)
#ARGOMENTO DELLA FUNZIONE: stima dei parametri diretti
#####
RiparCentrale_Dir <- function(parDir=c(xi, eta, lambda)){

  xi <- parDir[1]
  eta <- parDir[2]
  lambda <- parDir[3]

  b<-sqrt(2/pi)
  c<-(2/(4-pi))^(1/3)
  lambda2 <- (lambda)^2

  delta <- lambda / sqrt( 1 + lambda2)

  mu <- xi + b * eta * delta
  mu <- mu %% (2*pi)
  if(mu < 0) mu <- mu + 2*pi

  sigma <- sqrt( eta^2 * (1 - b^2 * delta^2) )

  gamma1 <- b * delta^3 * (2*b^2 - 1) / (1 - b^2*delta^2)^(3/2)

  list(mu=mu,sigma=sigma,gamma1=gamma1)
}
#####

#FUNZIONE PER CALCOLARE PARAMETRI CIRCOLARI: (mu, rho, b2bar)
#DAI PARAMETRI DIRETTI: (xi, eta, lambda)
#ARGOMENTO DELLA FUNZIONE: stima dei parametri diretti
#####
RiparCircolare_Dir <- function(parDir=c(xi, eta, lambda)){

  xi <- parDir[1]
  eta <- parDir[2]
  lambda <- parDir[3]

  b<-sqrt(2/pi)
  c<-(2/(4-pi))^(1/3)
  lambda2 <- (lambda)^2
```

```

delta <- lambda / sqrt( 1 + lambda2)

arg1<- delta * eta
arg2<- 2 * arg1

integrand <- function(x) {b*exp(((x^2)/2))}

if(arg1 < 0){
  arg1 <- -arg1
  arg2 <- -arg2
  int1 <- -(integrate(integrand, 0, arg1)$value)
  int2 <- -(integrate(integrand, 0, arg2)$value)
}

else if(arg1 >= 0){
  int1 <- integrate(integrand, 0, arg1)$value
  int2 <- integrate(integrand, 0, arg1)$value
}

top <- sin(xi) + int1 * cos(xi)

bottom <- cos(xi) - int1 * sin(xi)

if(bottom > 0 & top >= 0) mu <- atan(top/bottom)

else if(bottom > 0 & top < 0) mu <- 2*pi + atan(top/bottom)

else if(bottom < 0) mu <- pi + atan(top/bottom)

mu <- mu %% (2*pi)
if(mu < 0) mu <- mu + 2*pi

omega<- exp(-eta*eta/2)

rho<- omega * sqrt( 1 + int1^2 )

b2bar<- omega^4 * (int2 * (1 - int1^2) - 2 * int1) / (1 + int1^2)

list(mu=mu, rho=rho, b2bar=b2bar)
}
#####

#FUNZIONE PER CALCOLARE PARAMETRI DIRETTI: (xi, eta, lambda)
#DAI PARAMETRI CIRCOLARI: (mu, rho, beta2_bar)
#ARGOMENTO DELLA FUNZIONE: stima dei parametri circolari e osservazioni campionarie
#####
AP.lettura <- function(){

  setwd("C:/Users/Estefanía/Documents/ESTADISTICA/Universidad_PADOVA/TESI/R definitivo")

  dyn.load("C:/Users/Estefanía/Documents/R_MS DOS/INTCAL.dll")

  dyn.load("C:/Users/Estefanía/Documents/R_MS DOS/XSOLV.dll")

  TABLE<-read.table("modINTGRL.DAT", header=T); attach(TABLE)

  XTAB<-read.table("modXFUN.DAT", header=T); attach(XTAB)
}
#####
RiparDiretta_Cir <- function(parCir=c(theta_bar, R_bar, b2_bar), dati){

  theta_bar <- parCir[1]
  R_bar <- parCir[2]

```



```

b2_bar <- parCir[3]

XFVAL <- b2_bar/(R_bar^4)
AP.lettura()

if (XFVAL <= 0 ){
  XEST <- as.numeric(.Fortran("XSOLV",XFVAL,XVAL,XFUN,X,XINT,XEST=double(1))[6])
  XESINT <- as.numeric(.Fortran("INTCAL",XEST,X,XINT,XESINT=double(1))[4])
}

else{
  XFVAL <- -XFVAL
  XEST<--(as.numeric(.Fortran("XSOLV",XFVAL,XVAL,XFUN,X,XINT,XEST=double(1))[6]))
  XESINT <- -(as.numeric(.Fortran("INTCAL",XEST,X,XINT,XESINT=double(1))[4]))
}

eta <- sqrt(-2*log(R_bar) + log(1+XESINT*XESINT))

delta <- XEST / eta

top <- tan(theta_bar) - XESINT

bottom <- 1 + XESINT * tan(theta_bar)

sol1 <- atan(top/bottom)

if(sol1 < 0){
  sol1 <- sol1 + 2*pi
  sol2 <- sol1 - pi
}

else if(sol1 >= 0) sol2 <- sol1 + pi

omega <- exp(-eta*eta/2)

arg1 <- omega*(cos(sol1)-XEST*sin(sol1))
arg2 <- omega*(cos(sol2)-XEST*sin(sol2))

n <- length(dati)
a1 <- sum(cos(dati))/n; a2 <- sum(cos(2*dati))/n
A <- c(a1,a2)

sol <- A[1]

diff1 <- abs(arg1 - sol)
diff2 <- abs(arg2 - sol)

if(diff1 < diff2) xiEST <- sol1
else xiEST <- sol2

if(xiEST > pi) xiEST <- xiEST -2*pi

list(xi=xiEST, eta=eta, delta=delta)
}
#####

#FUNZIONE PER CALCOLARE PARAMETRI DIRETTI: (xi, eta, lambda)
#DAI PARAMETRI CENTRATI: (mu, sigma, gamma1)
#ARGOMENTO DELLA FUNZIONE: stima dei parametri centrati
#####
RiparDiretta_Cent <- function(par = c(mu,sigma,gamma1)){

  mu <- par[1]
  sigma <- par[2]

```

```

gamma1 <- par[3]

b<-sqrt(2/pi)
c<-((2/(4-pi))^(1/3))

if (gamma1<0) gamma <- -((-gamma1)^(1/3))
else gamma <- (gamma1)^(1/3)

xi <- mu - c * gamma * sigma
xi <- xi %% (2*pi)

if(xi > 2*pi) xi <- xi %% 2*pi
else if(xi < 0) xi <- 2*pi + xi

eta <- sigma * sqrt(1 + c^2 * gamma^2 ) #parametro nu

lambda <- (c * gamma) / sqrt( b^2 + c^2 * (b^2-1) * gamma^2 )

list(xi=xi,eta=eta,lambda=lambda)
}
#####

#FUNZIONE DI DENSITÀ di una WSNC con P.CENTRATA per una osservazione circolare
#ARGOMENTI DELLA FUNZIONE: osservazione_circolare e vettore di parametri centrati di una WSNC
#####
WSNCc.density <- function(oss, par = c(mu,sigma,gamma1), log=FALSE){

  mu <- par[1]
  sigma <- par[2]
  gamma1 <- par[3]

  if(sigma <= 0.00009 | abs(gamma1) >= 0.995){
    dens = 999
  }

  xi <- as.numeric(RiparDiretta_Cent(par)[1])
  eta <- as.numeric(RiparDiretta_Cent(par)[2])
  lambda <- as.numeric(RiparDiretta_Cent(par)[3])

  r<- -5:5
  c.theta <- (oss - xi+2*pi*r)/eta

  dens<- (2/eta) * sum(dnorm(c.theta) *pnorm(lambda * c.theta))

  dens
}
#####
#funzione di densità di una WSNCc per un VETTORE DI OSS.CIRCOLARI
#ARGOMENTO DELLA FUNZIONE: dati circolari in radianti e parametri centrati di una WSNC
#####
WSNCc.density.vec <- function(dati, par=c(mu,sigma,gamma1) ){

  sapply(dati, WSNCc.density, par=par)
}
#####

#funzione di LOG-VEROSIMIGLIANZA NEGATIVA di una WSNC con P.CENTRATA
#ARGOMENTO DELLA FUNZIONE: parametri CENTRATI.
#####
WSNCc.logver<-function(par,THETAoss){

  par[1] %% (2*pi)

```

```

if(abs(par[3]) >= 0.995 | par[2] <= 0 ) logver = -999

logver = - sum(log(WSNCc.density.vec( THETAoss,par )))

logver
}
#####

#Le funzioni di densità e log-verosimiglianza di una WSNC con P.DIRETTA sono analoghe alle funzioni della
P.CENTRATA, senza i controlli su gamma1 e senza le riparametizzazioni.
#####

```

A.2 Distribuzione normale asimmetrica circolare modificata

Si illustra il codice R per il calcolo della funzione di densità e della funzione di log-verosimiglianza della distribuzione *WSSNC*.

```

#FUNZIONE DI DENSITÀ di una WSSNC per una osservazione circolare
#ARGOMENTI DELLA FUNZIONE: osservazione_circolare e vettore di parametri diretti
#####
WSNCdBI.density <- function(oss, par=c(xi, eta, alpha, beta), log=FALSE){

  xi <- par[1]
  eta <- par[2]
  alpha <- par[3]
  beta <- par[4]

  r <- -5:5
  p.theta <- ( oss + 2*pi*r - xi ) / eta

  dens <- ( 2/eta ) * sum( dnorm(p.theta) * pnorm(alpha*p.theta + beta*p.theta^3) )
  dens
}
#####

#funzione di densità di una WSSNC per un VETTORE DI OSS.CIRCOLARI
#ARGOMENTO DELLA FUNZIONE: dati circolari in radianti e parametri diretti
#####
WSNCdBI.density.vec <- function(dati, par=c(xi, eta, alpha, beta) ){

  sapply(dati, WSNCdBI.density, par=par)
}
#####

#FUNZIONE DI LOG-VEROSIMIGLIANZA NEGATIVA di una WSSNC con P.Diretta
#ARGOMENTO DELLA FUNZIONE: parametri diretti
#####
#####
WSNCdBI.logver<-function( par,THETAoss){

  logver = - sum(log(WSNCdBI.density.vec( THETAoss,par )))

  logver
}
#####

```

A.3 Esempi di grafici e calcoli

Esempi sul codice R utile per la rappresentazione di alcuni grafici mostrati nell'elaborato, stima dei momenti campionari e separazione del campione tra brezza di mare e brezza di terra (le stime di massima verosimiglianza sono realizzate tramite la funzione `optim`).

```
#ANALISI DESCRITTIVO GIUGNO 09
#####

direzioni <- WindDir[!is.na(WindDir)] #togliamo i valori persi
#####
plot(WindDir, yaxt="n", col="dark cyan", pch=21,cex=0.8, xlab="numero di osservazione",
ylab="direzione",main="campione in indagine")
axis( 2,c(0,45,90,135,180,225,270,315,360),
      c("N","NE","E","SE","S","SW","W","NW","N"), las=1)
abline(v=c(72,72*2,72*5,72*4,72*3,72*6,72*7,72*8,72*9,
72*10,72*12,72*15,72*14,72*13,72*16,72*17,72*18,72*19,72*11,
72*20,72*21,72*25,72*24,72*23,72*26,72*27,72*28,72*29,72*22,72*30),lty=2,col="green")

hist(direzioni, 36, freq=T, yaxt="n", col="dark cyan", las=1, main="istogramma delle direzioni del vento")
axis( 1,c(0,45,90,135,180,225,270,315,360),
      c("N","NE","E","SE","S","SW","W","NW","N"))
#####

stimaMM.Cir <- function(dati){ #dati in radianti: direzioni * pi/180
  n <- length(dati)

  a1 <- sum(cos(dati))/n; a2 <- sum(cos(2*dati))/n
  b1 <- sum(sin(dati))/n; b2 <- sum(sin(2*dati))/n

  A <- c(a1,a2); B <- c(b1,b2)

  if (A[1]>0 & B[1]>=0)
    theta_bar <- atan(B[1]/A[1])
  else if (A[1]>0 & B[1]<0)
    theta_bar <- (pi*2) + atan(B[1]/A[1])
  else if (A[1]<0)
    theta_bar <- pi + atan(B[1]/A[1])

  R_bar <- sqrt( A[1]^2 + B[1]^2 )
  b2_bar <- (1/length(dati))*sum(sin(2*(dati - theta_bar)))

  if(theta_bar < 0) theta_bar <- (theta_bar + 2*pi)
  theta_bar%%(2*pi)

  list(theta_bar = theta_bar, R_bar = R_bar, b2_bar = b2_bar)
}
#####

##WindDir È UNA VARIABILE CIRCOLARE, allora:
#####
library(circular)
vento<-circular(direzioni,"angles","degree","geographics","2pi","clock")

##momento trigonometrico campionario di ordine 2:
a2=sum(cos(2*rad(direzioni)))/n; b2=sum(sin(2*rad(direzioni)))/n; a2; b2
```

```

## momento trigonometrico campionario centrato di ordine 2:
bar.a2=sum( cos( 2*( rad(direzioni)-media.camp) ) )/n
bar.b2=sum( sin( 2*( rad(direzioni)-media.camp) ) )/n

## coefficiente di asimmetria:
theta2.bar <- media.dir(a2,b2)
R2.bar<-sqrt(a2^2+b2^2)

s.capello <- (R2.bar*sin(theta2.bar-2*media.camp))/(varianza.camp^(3/2))
#####

## BREZZA DI MARE E BREZZA DI TERRA
#####
ore <- Time[!is.na(Time)] #non ci sono valori persi - OK.
n.ore<-as.numeric(ore)
summary(n.ore)

## BREZZA di MARE
brezza.marina<-c()
for(i in 1:length(WindDir)){
  if(n.ore[i] < 3 | n.ore[i] > 37) brezza.marina[i] <- WindDir[i]
  else brezza.marina[i] <- NA
}
brezza.marina
dir.orecalde <- brezza.marina[!is.na(brezza.marina)]
length(brezza.marina) - length(dir.orecalde)

## BREZZA di TERRA
brezza.diTerra<-c()
for(i in 1:length(WindDir)){
  if(n.ore[i] > 2 & n.ore[i] < 38) brezza.diTerra[i] <- WindDir[i]
  else brezza.diTerra[i] <- NA
}
brezza.diTerra
dir.orefredde <- brezza.diTerra[!is.na(brezza.diTerra)]
length(brezza.diTerra) - length(dir.orefredde)
length(dir.orefredde)
#####

## Esempio: GRAFICO FINALE
#####
m<-length(WindDir)
oss.9<-c()
for(i in seq(1,m,9)){
  oss.9[i] <- WindDir[i]
}
length(oss.9)
direzioni.9 <- oss.9[!is.na(oss.9)]

source("logV + densita WSNCd.r")

vento<-circular(rad(direzioni.9),"angles","rad","none","2pi",0,"counter")

y<-rose.diag(vento,bins=18,shrink=1.7,col="dark cyan",prop=1.5, tcl.text=0.2,
  xlim=c(-1,1.8), cex=0.8, ylim=c(-0.8,-0.5),
  title(main="campione ridotto giugno '08'09'10 con stima dei modelli", cex.main=0.8))

points(vento, plot.info=y, stack=T,col="dark cyan", pch=46)
arrows.circular(mean.circular(vento),col="dark blue",length = 0.2)

curve.circular(WSNCd.density.vec(x, stima.mv), add=T, lty=1, col="blue")
curve.circular(WSNCdBI.density.vec(x,stimaBI.mv), add=T, lty=1, col="green")
legend(0.9,-1,legend=c("WSNCd( 2.047, 3.776, 27.718)","WSSNC(4.544, 2.469, 9.455, -38.123)"),
  lty=1,col=c("blue","green"),cex=0.6)

```


Indice delle tabelle

Tabella 1.1: Descrizione dei dati.....	18
Tabella 3.1: Famiglie di distribuzioni simmetriche perturbate d'interesse per i nostri dati	45
Tabella 3.2: Casi limiti della WSNC con parametrizzazione diretta.....	49
Tabella 3.3: Riparametrazioni per le stime di WSNC.....	53
Tabella 3.4: Casi limiti della WSSNC.....	59
Tabella 4.1: Statistiche campionarie delle direzioni del vento di giugno 2009 in Villena.	63
Tabella 4.2: Coefficienti di asimmetria campionari.....	63
Tabella 4.3: Stima dei parametri della WSNC delle direzioni del vento 2009.....	69
Tabella 4.4: Stima dei parametri della WSSNC delle direzioni del vento 2009.....	69
Tabella 4.5: Stima dei parametri della WSNC delle direzioni della brezza di mare 2009	70
Tabella 4.6: Stima dei parametri della WSSNC delle direzioni della brezza di mare 2009	70
Tabella 4.7: Stima dei parametri della WSNC delle direzioni della brezza di terra 2009	71
Tabella 4.8: Stima dei parametri della WSSNC delle direzioni della brezza di terra 2009	72
Tabella 4.9: p-value di ogni campione per WSSNC contro WSNC	74
Tabella 4.10: Stima dei parametri della WSNC di una proporzione delle direzioni del vento di 2008, 2009 e 2010.....	77
Tabella 4.11: Stima dei parametri della WSSNC di una proporzione delle direzioni del vento di 2008, 2009 e 2010.....	77

Indice delle Figure

Figura 1.1: Rosa dei venti a 16 punti.....	7
Figura 1.2: Anemometro WSTX20.....	7
Figura 1.3: Forza risultante del vento. (Fonte: Joaquín Giménez Torres, 2007).....	9
Figura 1.4: Mappa Fisica della Spagna e localizzazione del comune di Villena.....	10
Figura 1.5: Mappa strutturato delle catene montuose delle “Cordilleras Béticas”.....	11
Figura 1.6: Mappa del Alto Vinalopó (Alicante - Spagna). (Fonte: Azorín Molina, 2002:113).....	12
Figura 1.7. A e B: circolazione di brezza marina e distribuzione della pressione all'inizio del pomeriggio con tempo anticiclonico. C e D: stessa situazione di notte, brezza di terra. (Fonte: Azorin-Molina (2002)).....	15
Figura 1.8: Osservazioni del mese di giugno 2009 (separati in giorni) in funzione della loro direzione del vento.....	19
Figura 1.9: Frequenza delle direzioni del vento in Villena nel mese di giugno 2009.....	20
Figura 2.1: Grafico circolare del orientamento di 76 tartarughe. (Fonte: Directional Statistics, 2000: 9).....	21
Figura 2.2: Direzione generica rappresentata per il vettore \mathbf{x} , l'angolo θ e il numero complesso z . (Fonte: Directional Statistics, 2000: 14).....	24
Figura 3.1: Densità nel cerchio unitario per.....	49
Figura 3.2: Forme della WSSNC per diversi valori di.....	56
Figura 3.3: Forme della WSSNC per valori di per $j=1,2$	57
Figura 3.4: Forme della WSSNC perche tendono a infinito.....	58
Figura 4.1: Osservazioni delle direzioni del vento nel mese di giugno 2009.....	62
Figura 4.2: Trasformazione dei dati osservati.....	64
Figura 4.3: Separazione tra brezza di mare-valle e brezza di terra.....	65
Figura 4.4: Istogrammi della brezza di mare e la brezza di terra.....	66
Figura 4.5: Adattamento della WSNC e la WSSNC per il campione delle direzioni del vento di giugno 2009.....	70
Figura 4.6: Adattamento della WSNC e la WSSNC per il campione di brezza di mare 2009.....	71

Figura 4.7: Adattamento della WSNC e la WSSNC per il campione di brezza di terra 2009.....	73
Figura 4.8: Campioni delle direzioni del vento in Villena del 2009 e del 2008, 2009 e 2010.....	76
Figura 4.9: Adattamento della WSNC e la WSSNC per il campione ridotto delle direzioni del vento di giugno 2008, 2009 e 2010.....	77

BIBLIOGRAFIA

Agostinelli, C., (2007), “Robust estimation for circular data.”, *Computational Statistics & Data Analysis*, **51**, 5867-5875.

Azorin-Molina, C., (2002), “La formación de frentes de brisa activos en la comarca alicantina del alto Vinalopó. El episodio atmosférico del 27 de abril de 2001.”, *Investigaciones geográficas*. Universidad de Alicante.

Azorin-Molina, C., Chen, D., Tijmd, S. & Baldie, M., (2011), “A multi-year study of sea breezes in a Mediterranean coastal site: Alicante (Spain).”, *International Journal of Climatology*, **31**, 468-486.

Breiman, L., (2001), “Statistical Modeling: The Two Cultures.”, *Statistical Science*, **16**, 199-231.

Azzalini, A., (1985), “A class of distributions which includes the normal ones.”, *Scandinavian Journal of Statistics*, **12**, 171-178.

Azzalini, A., (2001), *Inferenza statistica. Una presentazione basata sul concetto di verosimiglianza*. Milano: Springer-Verlag Italia.

Azzalini, A., (2005), “The Skew-normal Distribution and Related Multivariate Families.”, *Scandinavian Journal of Statistics*, **32**, 159-188.

Coles, S., (1998), “Inference for circular distributions and processes.”, *Statistics and Computing*, **8**, 105-113.

Elal-Olivero, D., Gómez, H. W. & Quintana, F. A., (2009), “Bayesian Modeling using a Class of Bimodal Skew-Elliptical Distributions.”, *Journal of Statistical Planning and Inference*, **139**, 1484–1492.

Ferrari, C., (2009), “The Wrapping Approach for Circular Data Bayesian Modeling.”, Bologna: Scuola di dottorato in Scienze Economiche e Statistiche. XXI ciclo.

Fisher, N.I., (1993), *Statistical Analysis of Circular Data*. Cambridge: University Press.

Fратиани, S., Cagnazzi, B. & Cremonini R., (2007), *Il vento in Piemonte*. Torino: Alzani arti grafiche.

Gimenez-Torres, J., (2007), *Meteorología y Climatología*. Elche: UMH. Facultad de Ciencias Ambientales.

Jammalamadaka, S.R. & Sengupta, A., (2001), *Topics in Circular Statistics*. Singapore: World Scientific.

Mardia, K.V., (1972), *Statistics of Directional Data*. London: Academic Press Inc.

Mardia, K.V. & Jupp, P.E., (2000), *Directional Statistics*. London: Academic Press Inc.

Matarredona-Coll, E., (1983), *Estudio geografico del alto Vinalopó*. Alicante: Instituto de estudios alicantinos.

McCracken, D., (1974), *Guida alla Programmazione del Fortran IV*. Roma: Edizioni Bizzarri.

Olivares-Pacheco, J. F., Elal-Olivero, D., Gómez, H. W. & Bolfarine H., (2009), “Una reparametrización de la distribución triangular basada en las distribuciones skew-simétricas.”, *Revista Colombiana de Estadística*, **32**, 145-156.

Pace, L. & Salvan, A., (2001), *Introduzione alla statistica - II. Inferenza, Verosimiglianza, Modelli*. Padova: CEDAM.

Pewsey, A., (2000), “The wrapped skew-normal distribution on the circle.”, *Communications in Statistics – Theory and Methods*, **29**, 2459-2472.

Pewsey, A., (2000b), “Problems of inference for Azzalini's skew-normal distribution.”, *J Appl Stat*, **27**, 859-870.

Pewsey, A., (2002), “Testing circular symmetry.”, *The Canadian Journal of Statistics*, **30**, 591-600.

Pewsey, A., (2004), “The large-sample joint distribution of key circular statistics.”, *Metrika*, **60**, 25-32.

Ravindran, P. & Gosh, S.K., (n.d.), “Bayesian Analysis of Circular Data Using Wrapped Distributions.”, North Carolina State University.

Rotolo, F., (2010), “Families of distributions on the circle.”, Padova: Scuola di dottorato in Scienze Statistiche. XXV ciclo.

Umbach, D., (2007), “The effect of the skewing distribution on skew-symmetric families.”, *Soochow Journal of Mathematics*, **33**, 657-668.

Wang, J., Boyer, J & Genton M.G., (2004), “A skew-symmetric representation of multivariate distributions.”, *Statistica Sinica*, **14**, 1259-1270.

Reed, W.J. & Pewsey, A., (2008), “Two nested families of skew-symmetric circular distributions.”,

Yanyuan, M.A. & G.Genton M.G., (2004), “Flexible Class of Skew-Symmetric Distributions.”, *Scandinavian Journal of Statistics*, **31**, 459-468.