



40
29

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

**INTRODUCCION AL ANALISIS DE DATOS
DIRECCIONALES**

T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I O

P R E S E N T A

ALICIA DEL ROSARIO NAVA CARDONA

**TESIS CON
FALLA DE ORIGEN**

MEXICO, D. F.

1990.





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

INTRODUCCION.	4
-----------------------	---

CAPITULO 1.

ESTADISTICA DESCRIPTIVA

1.1. Introducción.	6
1.2. Los datos direccionales.	6
1.2.1. Clasificación de fenómenos que involucran direcciones	
1.3. El análisis direccional y su relación con la estadística descriptiva.	8
1.3.1. La dirección cero.	
1.3.2. Representación en diagramas de datos no agrupados.	
1.3.3. Representación en diagramas de datos agrupados.	
1.3.4. Relación entre distintas unidades de medidas angulares.	
1.4. Medidas de tendencia central.	16
1.4.1. La dirección media o ángulo medio.	
1.4.2. La mediana direccional.	
1.4.3. La dirección modal.	
1.5. Medidas de dispersión.	28
1.5.1. Una medida de concentración.	
1.5.2. La varianza y la desviación estándar direccional.	
1.5.3. Otras medidas de dispersión.	
1.6. Muestras multimodales.	30

CAPITULO 2.

FUNCIONES DE DISTRIBUCION

2.1. Introducción.	32
2.2. La funcion de distribucion	32
2.2.1.La funcion de densidad circular.	
2.3. La funcion característica.	35
2.3.1.Propiedades de la función característica	
2.4. Momentos de las distribuciones circulares. 39	
2.4.1.Primer momento Trigonométrico.	
2.4.2.Momentos Trigonometricos de orden mayor.	
2.4.3.Momentos Trigonométricos centrales.	
2.5. Funciones de distribucion circulares. . .	43
2.5.1.Distribuciones simetricas.	
2.5.2.Modelos Circulares.	
2.5.3.Distribuciones envueltas.	
2.5.4.Distribuciones Multimodales.	
2.5.5.Representacion gráfica en coordenadas rectangulares.	

CAPITULO 3.

INFERENCIA ESTADISTICA

3.1. Introducción.	57
3.2. Estimación de parametros.	57
3.2.1.Estimación puntual maximoverosimil.	
3.2.2.Intervalos de confianza.	
3.3. Pruebas de hipotesis.	66
3.3.1.Pruebas de hipotesis sobre los parámetros k y μ .	
3.3.2.Pruebas de uniformidad.	
3.3.3.Pruebas de bondad de ajuste.	

CAPITULO 4.

EJEMPLO

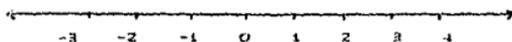
4.1. Introduccion.	77
4.2. Analisis de la precipitacion pluvial.	77
4.2.2. Planteamiento estadistico.	
4.2.3. Analisis estadistico de los datos.	
4.2.4. Pruebas de hipotesis.	
CONCLUSIONES.	92
APENDICE.	93
BIBLIOGRAFIA.	94

INTRODUCCION

El presente trabajo es una revisión bibliográfica de la metodología desarrollada para analizar estadísticamente datos direccionales.

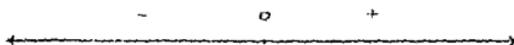
En los cursos de estadística matemática, generalmente se analizan datos que por ser numéricos, pueden ser graficados a lo largo del eje real; ejemplos de tales datos son los provenientes de poblaciones cuya distribución es Binomial, Poisson ó Normal. En el presente trabajo, se hace referencia al análisis estadístico de dichos datos, como: Estadística lineal.

Se le da tal nombre debido a que los datos pueden ser graficados en una línea, el eje real \mathbb{R} .



Existen, sin embargo, fenómenos cuya principal manifestación es un conjunto de direcciones, esto es, un conjunto de líneas o sentidos, sobre los cuales se encamina o dirige el objeto de estudio.

Ahora, si se intenta graficar un conjunto de direcciones en el eje real, solo se puede hacer en dos direcciones ó sentidos, el positivo y el negativo, a partir del cero.



Pero ya no impartan los números que están representados en la línea, sino solo la dirección. Este es un caso muy específica y se puede decir que el más trivial. Dado que, por ejemplo, en un plano a partir del punto $(0,0)$ podemos encontrar un conjunto más amplia de posibles direcciones y aun más en \mathbb{R}^n .

Fig.

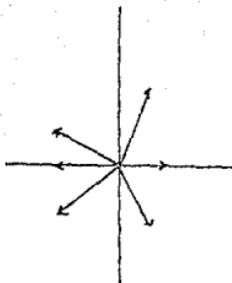
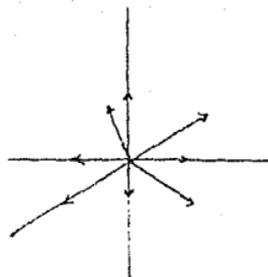


Fig.



Esto puede dar una primera idea de que, al tener un conjunto de direcciones, no se tiene un conjunto de datos usuales de estadística lineal. Es por esto que, se ha desarrollado una metodología especial para analizarlos estadísticamente. Esto es, dado un conjunto de direcciones, determinar cual es la dirección preferida, cuánto se alejan ó desvían de esta las direcciones observadas, etc. siendo el propósito principal de este trabajo el mostrar dicha metodología.

Con este fin, la tesis se divide en cuatro capítulos:

El primer capítulo se dedica a las técnicas de estadística descriptiva.

En el segundo capítulo se hace una revisión de las funciones de distribución de poblaciones direccionales.

El tercer capítulo se destina a las técnicas de inferencia estadística.

Finalmente, el capítulo cuarto muestra una aplicación de la metodología mostrada en los primeros tres capítulos.

CAPITULO 1

ESTADISTICA DESCRIPTIVA

1.1. Introducción.

A manera de motivación en el presente capítulo se define a los datos direccionales.

Se da una clasificación de tres distintos tipos de datos direccionales y se mencionan algunas de las áreas científicas que los estudian.

Siendo éste el capítulo que se dedica a la estadística descriptiva, se examinan los métodos empleados para ilustrar gráficamente las observaciones.

Se mencionan las distintas medidas de unidades angulares y sus relaciones.

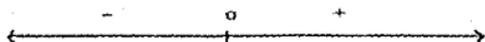
Con el propósito de describirlos numéricamente, se analizan las medidas de tendencia central, así como las medidas de dispersión más importantes.

Para finalizar, se presenta una breve sección dedicada al análisis descriptivo de muestras multimodales.

1.2. Los datos direccionales.

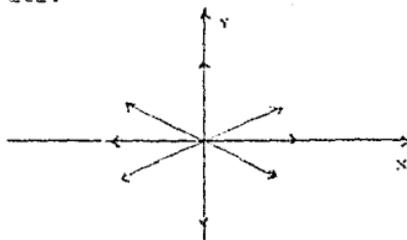
Se define a los datos direccionales como un conjunto de direcciones, líneas o sentidos que toma o sigue un cierto objeto de estudio.

Las direcciones se pueden considerar en el eje real, a partir del cero, como dirección positiva si se considera el sentido hacia los números positivos, negativa si se considera el sentido hacia los números negativos. Entonces, en el eje real, solo se tienen dos distintas direcciones, una hacia 'adelante' y otra hacia 'atras'.



En el plano, se tiene una gama mas amplia de direcciones; hacia 'adelante', hacia 'atras', hacia 'arriba', hacia 'abajo', etc.

Fig



Y se tienen aun mas en \mathbb{R}^3 . De hecho podemos encontrar direcciones en cualquier dimension pero las aplicaciones practicas se dan generalmente en dos o tres dimensiones.

El material de la presente tesis unicamente presenta el analisis para datos en dos dimensiones.

1.2.1. Clasificacion de fenomenos que involucran direcciones.

Un solo dato direccional, es una observacion de la direccion que siguió el objeto estudiado.

En \mathbb{R}^2 se pueden encontrar tres grupos distintos de datos direccionales:

El primer grupo es aquel que incluye a fenomenos que de manera natural presentan observaciones de direcciones; un ejemplo claro de tales fenomenos, se presenta en meteorología al observar las direcciones del viento.

El segundo grupo está formado por las observaciones, llamadas 'observaciones con eje', debido a que solo se registra el eje o línea que siguió, sin considerar el sentido en que lo hizo. Por ejemplo si siguió el eje Norte-sur, sólo importa saber que el objeto de estudio, siguió dicha línea, sin importar si lo hizo en direccion del Norte o del Sur. Un ejemplo de tales observaciones se presentan en Geología, cuando se realizan estudios de magnetismo remanente, para

determinar los distintos desplazamientos del eje magnético de la tierra a través de los años.

El tercer grupo de datos direccionales considera una gran variedad de observaciones de fenómenos que se desarrollan en el tiempo, dichos fenómenos presentan un ciclo o periodicidad fija conocida. El ejemplo analizado en el capítulo cuarto, pertenece a este último grupo.

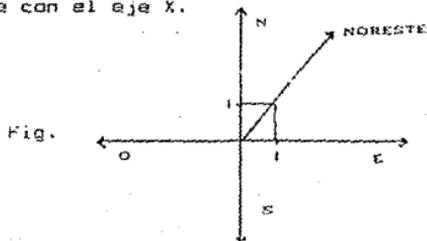
Diversas áreas científicas se interesan en el análisis de datos direccionales, entre las más importantes se encuentran la Biología, la Economía, la Física, la Psicología la Medicina y las ya mencionadas Geología y Meteorología.

1.3. El análisis direccional y su relación con la estadística descriptiva.

En el análisis de datos direccionales se presentan las mismas necesidades de organizar, resumir y mostrar en forma clara la información contenida en una muestra; así, para comenzar se analizan distintas técnicas empleadas para representar datos direccionales.

Dado que solo se analizan observaciones direccionales que se encuentran en \mathbb{P}^2 es posible graficarlas en el plano \overline{XY}

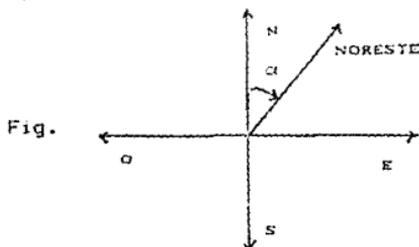
Por ejemplo, si se observa la dirección Noroeste, se puede asociar el eje \overline{Y} , con el eje Norte-Sur, y el eje Oeste-Este con el eje \overline{X} .



Entonces la dirección Noroeste se puede representar por las coordenadas $(1,1)$ o $(2,2)$ o $(3,3)$ etc. ya que cualquiera de ellas son puntos que representan un vector con la misma

dirección Noreste.

Otra forma de representar dicho vector es considerando el ángulo α que forma con respecto a alguna de los ejes. Por ejemplo, se puede medir el ángulo en azimuts, como lo hacen los geógrafos, esto es, se mide el ángulo α a partir de la dirección Norte en el sentido de las manecillas del reloj.



Así $\alpha = 45^\circ$ es la medida en azimuts que determina el vector de la dirección Noreste.

1.3.1. La dirección cero.

En el Análisis de Datos Direccionales para graficar y en general, analizar un conjunto de datos direccionales es necesario seleccionar una dirección a la cual se le llama dirección cero. La dirección cero puede ser alguna de las direcciones obtenidas en la muestra o cualquier otra.

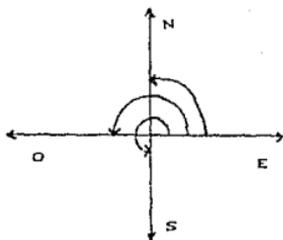
Esta dirección es importante debido a que, a partir de ella, se le asocia un ángulo θ° a cada una de las observaciones de la muestra. La dirección cero recibe este nombre por asociarse el ángulo $\theta^\circ = 0^\circ$ a ésta.

Para asociar un ángulo θ° a cada una de las observaciones en la muestra, se mide el ángulo θ° en grados a partir de la dirección cero en sentido contrario a las manecillas del reloj.

Por ejemplo si se tiene el siguiente conjunto de observaciones direccionales: Sur, Norte, Sur, Este, Oeste

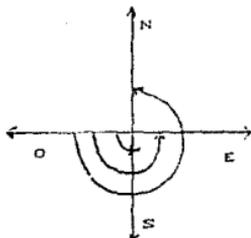
Si se desea tomar el Este como dirección cero, entonces los ángulos asociados al conjunto de observaciones

direccionales son:



Dirección	Angulo θ°
Sur	270°
Norte	90°
Sur	270°
Este	0°
Oeste	180°

Si se desea tomar el Oeste como direccion cero se tiene:

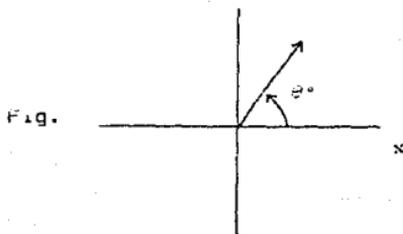


Dirección	Angulo θ°
Sur	90°
Norte	270°
Sur	90°
Este	180°
Oeste	0°

Las direcciones observadas siguen siendo las mismas y solo depende de la dirección cero elegida el valor del ángulo θ° que se asocia a cada una de las observaciones.

1.3.2. Representación en diagramas de datos no agrupados.

Para graficar un conjunto de datos direccionales no agrupados inicialmente se selecciona la dirección cero. Esta dirección se asocia al eje X positivo.

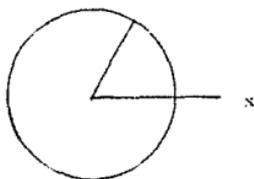


El análisis de datos direccionales, solo se interesa en las direcciones, sin considerar el peso, velocidad o fuerza

con que son tomadas. Una simple observacion se puede representar por un vector unitario y el angulo que hace con respecto a la direccion cero es el correspondiente θ° asociado a la observacion.

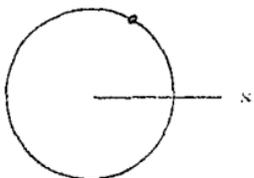
Para representar en forma facil un conjunto de datos direccionales, se dibuja un circulo unitario y por cada observacion se dibuja un radio del circulo unitario.

Fig.



La otra forma alternativa, es representar cada dato por un punto en la circunferencia del circulo unitario, cuya masa asignada es la misma para cada observacion. Tales puntos se determinan por la interseccion del vector unitario, mencionado anteriormente, y la circunferencia.

Fig.



1.3.3. Representacion en diagramas de datos agrupados.

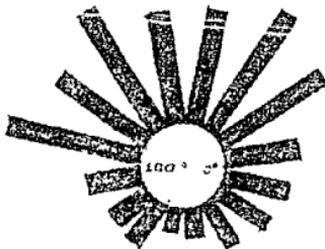
Los datos direccionales se pueden agrupar adoptando los mismos procedimientos que en la linea real. El rango en el caso direccional de $(0^\circ, 360^\circ)$ se puede dividir en un numero dado de intervalos de clase y la frecuencia correspondiente a cada clase se puede contar como ocurre en el caso lineal; es decir se escoge un numero dado de clases y la longitud de los intervalos de clase se calcula como en los datos lineales.

con la diferencia de que en el caso circular puede haber un intervalo tal como $330^{\circ}-30^{\circ}$ que contiene simultáneamente los ángulos 359° y 0° . Los ángulos de cada intervalo siempre serán medidos en dirección contraria a las manecillas del reloj.

1.3.3.a. El histograma circular.

Se pueden representar los datos en un histograma análogo al usado en la línea; se toma un círculo unitario y correspondiendo a cada intervalo, se construye un bloque en la circunferencia, cuya área sea proporcional a la frecuencia del intervalo correspondiente y cuya longitud en la base sea igual a la longitud del intervalo de clase.

Fig.



Si se considera las mismas técnicas que se emplean en la línea, se puede construir también el correspondiente polígono de frecuencia, al marcar los puntos medios de la parte alta de cada bloque y uniéndolos consecutivamente se tiene la representación circular de dicho polígono.

Fig.

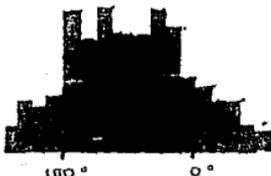


1.3.3.b. El histograma lineal.

Otra representación usual es hacer un corte en la base del histograma circular teniendo así un segmento de amplitud 360° . El punto de corte se puede seleccionar como más convenga. Si los datos tienen una dirección preferida es decir, una moda direccional, es recomendable hacer un corte tal que al extender el histograma circular, el centro del histograma lineal corresponda aproximadamente a esta moda. Un corte cerca de la moda, puede dar una impresión errónea de los datos.

El histograma lineal para datos direccionales, sólo se puede considerar después de imaginarlo dibujado alrededor de la circunferencia de un círculo. En tal representación el final de los ejes existe sólo como una conveniencia y para enfatizar este factor el primer bloque del correspondiente histograma lineal se repite al final. El histograma lineal es preferido al circular, debido a la similitud que tiene con los histogramas de estadística lineal, hecha que facilita su interpretación.

Fig.



1.3.3.c. El diagrama de rosa.

La última representación gráfica de datos direccionales agrupados es el diagrama de rosa circular.

Correspondiendo a cada intervalo se construye un sector en forma de "pay" con centro en el origen cuya ángulo sea:

$\theta^\circ = 360^\circ/k$, donde k es el número de intervalos; la longitud del radio es proporcional a la frecuencia de cada clase.

El área de cada sector en un diagrama de rosa varía de acuerdo al cuadrado de las frecuencias. Si se desean hacer proporcionales las áreas a las frecuencias observadas, es recomendable graficar el diagrama de rosa en base a la raíz cuadrada de las frecuencias en lugar de las frecuencias mismas, a este último diagrama se le llama diagrama de rosa equi-área.



Fig.

Los histogramas lineales conservan áreas en forma natural y son comparativamente simples de construir. El diagrama de rosa, el histograma circular y el polígono de frecuencia son en ocasiones descritos como diagramas polares de abanico.

1.3.4. Relación entre distintas unidades de medidas angulares.

En la sección 1.3.1 se vio que una observación direccional medida en grados se denota por θ° , y representa el ángulo hecho por el vector con respecto a la dirección cero elegida, considerando la dirección contraria a las manecillas del reloj. Sin embargo, para propósitos teóricos y prácticos es necesario convertir, el ángulo θ° en grados, al ángulo θ en radianes; lo cual se logra utilizando la siguiente igualdad:

$$\theta^\circ = 180^\circ \left[\frac{\theta}{\pi} \right]$$

$$\Rightarrow \theta = \pi \left[\frac{\theta^\circ}{180^\circ} \right]$$

Si el rango de θ° es $0 < \theta^\circ < 360^\circ$, el rango de θ será $0 < \theta < 2\pi$. Esto es, si un arco de un círculo con radio unitario sustituye un ángulo de θ° grados en el centro del círculo, entonces la longitud del arco será θ .

En algunos casos las observaciones direccionales se pueden concentrar en el rango $(0^\circ, 180^\circ)$ o $(0^\circ, 90^\circ)$. Si se tienen observaciones en el rango $(0^\circ, 180^\circ)$, éstas se pueden convertir al rango $(0^\circ, 360^\circ)$ doblando ó haciendo doble cada ángulo, es decir, calculando $\theta'^\circ = 2\theta^\circ$. En general el rango $(0^\circ, 360^\circ/l)$ se puede convertir al rango $(0, 2\pi)$ por:

$$\theta' = l \pi \theta^\circ / 180^\circ$$

En el caso de observaciones periódicas se pueden realizar conversiones de tiempo, al reducirlos modulo algún periodo en ángulos. La longitud del periodo de tiempo se asocia con 360° . Por ejemplo, un periodo de 24 horas corresponde a una vuelta completa de 360° , entonces, se puede comparar una hora con un ángulo de $360^\circ/24 = 15^\circ$ y la mitad del día con 180° , etc.

Entre los periodos de tiempo más usuales, se encuentran un día o un año para un periodo de tiempo menor a un año el t -ésimo día se puede transformar a ángulos por la relación:

$$\theta^\circ = 360^\circ \cdot t / 365$$

Para el periodo de 1 año, si se desea clasificar en meses, los intervalos de clase y los puntos medios, se pueden calcular usando la anterior relación, sin embargo, es necesario llevar acabo un ajuste en las frecuencias correspondientes a cada mes debido a que la longitud real de

los meses no es la misma para todos. Así si f_i es la frecuencia observada, correspondiente al i -ésimo mes y n_i el número de días del i -ésimo mes se considera a la nueva frecuencia ajustada, f'_i como:

$$f'_i = 30 \frac{f_i}{n_i}$$

1.4. MEDIDAS DE TENDENCIA CENTRAL.

El análisis de datos direccionales también considera algunas medidas numéricas que permiten describir, brevemente, a las observaciones de una muestra. En esta sección, se analizan las medidas de tendencia central, entre las cuales se encuentran la media, la mediana y la moda. Estas medidas son análogas a las de estadística lineal.

Inicialmente se analiza la forma óptima de calcular la media de una muestra.

Se ha visto en la sección anterior que es posible asociar a las direcciones un ángulo θ° ; entonces, si se tiene la siguiente muestra formada por 3 direcciones

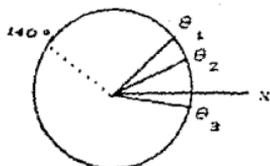
$$\theta_1^\circ = 45^\circ \quad \theta_2^\circ = 25^\circ \quad \text{y} \quad \theta_3^\circ = 350^\circ$$

y se desea determinar una dirección media ó ángulo medio de estas 3 observaciones. Se puede calcular la media aritmética

$$\frac{1}{3} [\theta_1^\circ + \theta_2^\circ + \theta_3^\circ] = \frac{1}{3} [45^\circ + 25^\circ + 350^\circ] = 140^\circ$$

se pueden graficar los 3 ángulos muestrales y el ángulo medio que resulta de aplicar la media aritmética

fig.



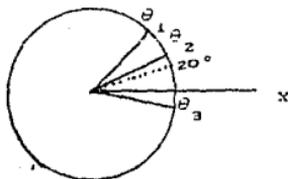
Se puede observar que el ángulo obtenido, no proporciona

un ángulo que describa la dirección media de las 3 observaciones. Por otra parte, si se reemplaza 350° por el ángulo -10° , que es equivalente, y se calcula de nuevo la media aritmética

$$\frac{1}{3} [45^\circ + 25^\circ - 10^\circ] = 20^\circ$$

Graficando el nuevo ángulo medio

Fig.

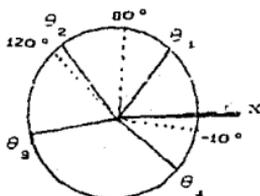


Se puede apreciar que el resultado obtenido es un ángulo medio ó dirección media más adecuada.

Si se intenta de nuevo pero ahora con 4 ángulos, es decir, si se tiene la siguiente muestra:

$$\theta_1^\circ = 50^\circ \quad \theta_2^\circ = 120^\circ \quad \theta_3^\circ = 190^\circ \quad \text{y} \quad \theta_4^\circ = 320^\circ$$

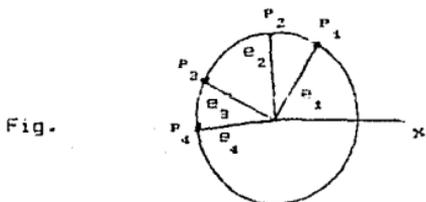
Fig.



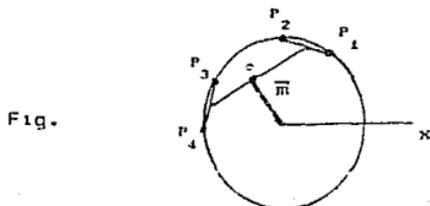
La media aritmética será 125° . Si se reemplaza 320° por -40° , el resultado será 80° y si además 190° es reemplazado por -170° la media aritmética se reduce a -10° . De modo, que dependiera del manejo de los datos el resultado que se obtenga, entonces, se puede dar el caso de obtener más de un resultado aceptable lo que lleva a tener ambigüedades, y por lo tanto, no es conveniente utilizar el cálculo de la media aritmética para buscar un ángulo medio ó dirección media.

1.4.1. La dirección media o ángulo medio.

Si se grafican los valores de una muestra aleatoria de tamaño n , de una población de variables aleatorias direccionales, en dicho diagrama, se llama e_i al vector unitario que representa al ángulo θ_i° y P_i es el punto en la circunferencia del círculo unitario correspondiente al ángulo θ_i° .



A cada P_i se le asigna una masa de igual valor y se encuentra, entonces, el centro de masa c de éstos puntos también llamada centro de gravedad.



Si este centro es distinto del origen, la línea \overline{OC} define una dirección a la cual se le llama dirección media de la muestra.

Los vectores unitarios $e_1, e_2, e_3, \dots, e_n$ fijan la ubicación de cada P_i en el círculo unitario.

Por definición $|e_i| = 1$ para todo $i=1, \dots, n$

Se calcula el vector resultante R

$$R = e_1 + e_2 + e_3 + \dots + e_n$$

Sea \bar{m} el vector medio de la muestra $\bar{m} = R / n$ y sea

$LR = |R|$ la longitud del vector resultante; entonces, la longitud del vector medio ó dirección media r es,

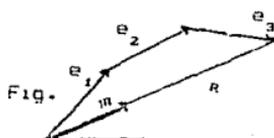
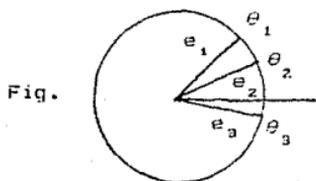
$$r = LR / n$$

La siguiente figura ilustra el procedimiento para la muestra de tamaño 3.

$$\theta_1^\circ = 45^\circ$$

$$\theta_2^\circ = 25^\circ$$

$$\theta_3^\circ = 350^\circ$$



Si se cambia θ_3° por -10° se obtiene el mismo resultado. El centro de masa c puede tocar la circunferencia del círculo unitario, pero sólo en casos excepcionales, cuando todas las masas se encuentran juntas en un solo punto.

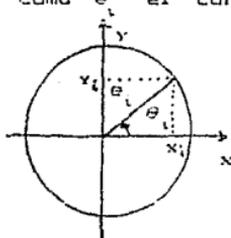
Entonces

$$0 \leq LR \leq n \quad \text{y} \quad 0 \leq r \leq 1$$

Bajo este principio, se encuentra la dirección media utilizando las coordenadas cartesianas de cada P_i :

Primeramente, utilizando un sistema de coordenadas rectangulares con \underline{x} y \underline{y} ejes y origen O , se grafica un ángulo θ_i° observado así como e_i el correspondiente vector unitario.

Fig.



Sean X_i y Y_i las coordenadas rectangulares de e_i .

entonces por definición de seno y coseno.

$$X_i = \cos \theta_i \quad Y_i = \sin \theta_i$$

con θ_i el valor del ángulo θ_i convertido a radianes.

De esta forma el centro de masa de estos puntos es

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n \cos \theta_i \quad \text{y}$$

$$\bar{S} = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n \sin \theta_i$$

Que están determinando las coordenadas cartesianas del vector medio; nuevamente LR es la longitud del vector resultante R cuyas coordenadas son

$$\left(\sum \cos \theta_i, \sum \sin \theta_i \right)$$

entonces

$$LR = \left[\left(\sum_{i=1}^n \cos \theta_i \right)^2 + \left(\sum_{i=1}^n \sin \theta_i \right)^2 \right]^{\frac{1}{2}}$$

Como $LR = nr$ entonces se tiene que

$$r = \frac{1}{n} \left[\left(\sum_{i=1}^n \cos \theta_i \right)^2 + \left(\sum_{i=1}^n \sin \theta_i \right)^2 \right]^{\frac{1}{2}}$$

$$r = (\bar{C}^2 + \bar{S}^2)^{1/2}$$

Se puede dar el caso en que $\bar{C} = 0$ y $\bar{S} = 0$ entonces $r = 0$, y el vector medio \bar{m} es igual al vector cero; pero fuera de este caso, el vector medio tiene un bien definido ángulo medido a partir del eje positivo x ; a éste ángulo se le llama ángulo medio de la muestra y lo denotaremos por \bar{x}_0 , el cual está determinado por la solución de las ecuaciones siguientes:

$$\bar{C} = r \cos \bar{x}_0 \quad \text{y} \quad \bar{S} = r \sin \bar{x}_0$$

Entonces, es conveniente calcular:

$$\bar{x}_0 = \begin{cases} \arctan(\bar{S}/\bar{C}) & \text{si } \bar{C} > 0 \text{ y } \bar{S} > 0 \\ \pi + \arctan(\bar{S}/\bar{C}) & \text{si } \bar{C} < 0 \\ 2\pi + \arctan(\bar{S}/\bar{C}) & \text{si } \bar{S} < 0 \text{ y } \bar{C} > 0 \end{cases}$$

Casos excepcionales son :

$$\bar{x}_0 = \begin{cases} 90^\circ & \text{si } \bar{C} = 0 \text{ y } \bar{S} > 0 \\ 270^\circ & \text{si } \bar{C} = 0 \text{ y } \bar{S} < 0 \\ \text{indeterminado} & \text{si } \bar{C} = 0 \text{ y } \bar{S} = 0 \end{cases}$$

Como una revisión en calculos, se pueden usar las formulas siguientes:

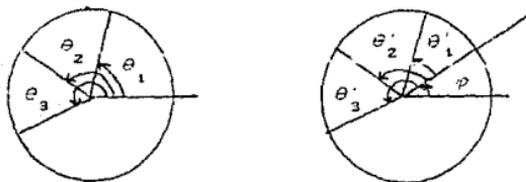
$$\cos \bar{x}_0 = \bar{C}/r \quad \text{sen } \bar{x}_0 = \bar{S}/r$$

1.4.1.a. La dirección media vs. cambio de dirección cero

Como el centro de masa se define independientemente de un sistema de coordenadas entonces, la longitud del vector medio no depende de la dirección cero. Supongase que se rota la dirección cero al ángulo p entonces los ángulos asociados a la muestra $\theta_1, \theta_2, \theta_3, \dots, \theta_n$ se convierten en

$$\theta'_i = (\theta_i - p) \bmod 2\pi \quad i=1,2,\dots,n$$

Fig.



Es posible calcular el correspondiente vector medio dados los nuevos ángulos θ'_i ($i=1,2,\dots,n$).

Por medio de las estadísticas siguientes:

dados los nuevos ángulos θ'_i $i=1,2,\dots,n$.

Por medio de las estadísticas siguientes:

$$\bar{C}' = \frac{1}{n} \sum_{i=1}^n \cos \theta'_i$$

$$\bar{S}' = \frac{1}{n} \sum_{i=1}^n \operatorname{sen} \theta'_i$$

Entonces

$$\bar{C}' = r \cos \bar{x}'_0 \quad \text{y} \quad \bar{S}' = r \operatorname{sen} \bar{x}'_0$$

donde

$$\bar{x}'_0 = (\bar{x}_0 - \rho) \bmod 2\pi$$

y

$$r' = \frac{1}{n} (C'^2 + S'^2)^{1/2}$$

conservándose la longitud r , del vector medio, invariante. Esto es posible comprobarlo algebraicamente analizando r'

$$\begin{aligned} C'^2 + S'^2 &= \left[\sum \cos (\theta_i - \rho) \right]^2 + \left[\sum \operatorname{sen} (\theta_i - \rho) \right]^2 \\ &= \left[\sum (\cos \theta_i \cos \rho + \operatorname{sen} \theta_i \operatorname{sen} \rho) \right]^2 + \\ &\quad + \left[\sum (\operatorname{sen} \theta_i \cos \rho - \cos \theta_i \operatorname{sen} \rho) \right]^2 \\ &= \left[\sum \cos \theta_i \right]^2 \cos^2 \rho + \left[\sum \operatorname{sen} \theta_i \right]^2 \operatorname{sen}^2 \rho + \\ &\quad + \left[\sum \operatorname{sen} \theta_i \right]^2 \cos^2 \rho + \left[\sum \cos \theta_i \right]^2 \operatorname{sen}^2 \rho \\ &= \left[\sum \cos \theta_i \right]^2 + \left[\sum \operatorname{sen} \theta_i \right]^2 = C^2 + S^2 \end{aligned}$$

Por lo tanto $r' = r$.

Es decir, la longitud r del vector medio permanece invariante bajo cambios de la dirección cero.

1.4.1.b. Calculo de la dirección media para datos agrupados.

Al igual que en estadística lineal si se desea calcular la media a partir de datos agrupados, se supone que los datos, pertenecientes a un mismo intervalo de clase, pueden ser identificados por el ángulo medio del límite superior e inferior del intervalo de clase. Ahora, si se llama k al número de intervalos de clase, y λ a la longitud de cada intervalo de clase entonces

$$\lambda = 2\pi / k \quad \text{o} \quad \lambda = 360^\circ / k$$

Si $\bar{\theta}_1, \bar{\theta}_2, \dots, \bar{\theta}_k$ son los puntos medios de los k arcos, medidos en radianes, y f_1, f_2, \dots, f_k las frecuencias correspondientes a cada intervalo, entonces:

$$n = f_1 + f_2 + \dots + f_k$$

es el tamaño de la muestra.

Calculando los componentes del vector medio m de la siguiente manera,

$$\bar{C} = \frac{1}{n} \left[f_1 \cos \bar{\theta}_1 + f_2 \cos \bar{\theta}_2 + \dots + f_k \cos \bar{\theta}_k \right]$$

$$\bar{S} = \frac{1}{n} \left[f_1 \sin \bar{\theta}_1 + f_2 \sin \bar{\theta}_2 + \dots + f_k \sin \bar{\theta}_k \right]$$

Para calcular la longitud del vector medio r y el ángulo medio \bar{x}_c , se utilizan los mismos procedimientos que en datos no agrupados, considerando los nuevos componentes del vector medio m que se han presentado en esta sección.

1.4.1.c. Propiedades de la dirección media.

Algunas propiedades importantes de \bar{x}_c , como medida de tendencia central circular, se analizan a continuación.

Primero, se considera la siguiente identidad trigonomé-

trica:

$$\operatorname{sen}(\theta_i - \bar{x}_o) = \operatorname{sen} \theta_i \cos \bar{x}_o - \cos \theta_i \operatorname{sen} \bar{x}_o$$

sumando desde $i=1$ hasta n , ambos lados de la igualdad

$$\sum_{i=1}^n \operatorname{sen}(\theta_i - \bar{x}_o) = \sum_{i=1}^n \left(\operatorname{sen} \theta_i \cos \bar{x}_o - \cos \theta_i \operatorname{sen} \bar{x}_o \right)$$

$$\sum_{i=1}^n \operatorname{sen}(\theta_i - \bar{x}_o) = \cos \bar{x}_o \sum_{i=1}^n \operatorname{sen} \theta_i - \operatorname{sen} \bar{x}_o \sum_{i=1}^n \cos \theta_i$$

así

$$\sum_{i=1}^n \operatorname{sen}(\theta_i - \bar{x}_o) = \left[\frac{\bar{C}}{r} \right] n\bar{S} - \left[\frac{\bar{S}}{r} \right] n\bar{C} = 0.$$

entonces

$$\sum_{i=1}^n \operatorname{sen}(\theta_i - \bar{x}_o) = 0.$$

lo cual, es análogo a lo que ocurre con la ecuación,

$$\sum_{i=1}^n (X_i - \bar{X}) = 0.$$

en el caso de estadística lineal.

Considerando ahora, la igualdad trigonométrica siguiente:

$$\cos(\theta_i - \bar{x}_o) = \cos \theta_i \cos \bar{x}_o + \operatorname{sen} \theta_i \operatorname{sen} \bar{x}_o$$

y sumando de nuevo desde $i=1$ hasta n , se tiene,

$$\sum_{i=1}^n \cos(\theta_i - \bar{x}_o) = \sum_{i=1}^n \left(\cos \theta_i \cos \bar{x}_o + \operatorname{sen} \theta_i \operatorname{sen} \bar{x}_o \right)$$

$$\sum_{i=1}^n \cos(\theta_i - \bar{x}_o) = \cos \bar{x}_o \sum_{i=1}^n \cos \theta_i + \operatorname{sen} \bar{x}_o \sum_{i=1}^n \operatorname{sen} \theta_i$$

$$\sum_{i=1}^n \cos(\theta_i - \bar{x}_o) = \left[\frac{\bar{C}}{r} \right] n\bar{C} + \left[\frac{\bar{S}}{r} \right] n\bar{S}$$

entonces

$$\sum_{i=1}^n \cos (\theta_i - \bar{x}_0) = \frac{n}{r} (\bar{C}^2 + \bar{S}^2) = \frac{n}{r} r^2$$

por lo tanto
$$\sum_{i=1}^n \cos (\theta_i - \bar{x}_0) = nr$$

Ahora, si se divide entre n , se multiplica por -1 y se le suma 1 a ambos lados de la expresión se tiene:

$$1 - \frac{1}{n} \sum_{i=1}^n \cos (\theta_i - \bar{x}_0) = 1 - r$$

que es igual a
$$\frac{1}{n} \sum_{i=1}^n (1 - \cos (\theta_i - \bar{x}_0)) = 1 - r$$

multiplicando por 2
$$\frac{1}{n} \sum_{i=1}^n 2(1 - \cos (\theta_i - \bar{x}_0)) = 2(1 - r)$$

pero para desviaciones $(\theta_i - \bar{x}_0)$ pequeñas, se cumple que

$$(\theta_i - \bar{x}_0)^2 \approx 2(1 - \cos (\theta_i - \bar{x}_0))$$

en cuyo caso es posible aproximar la sumatoria,

$$\frac{1}{n} \sum_{i=1}^n (\theta_i - \bar{x}_0)^2 \approx 2(1 - r)$$

Que como se analizará en la sec. 1.5.2. es análoga a

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2$$

en el análisis estadístico lineal.

Resumiendo las analogías se tiene:

ESTADÍSTICA LINEAL

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

$$(x_i - \bar{x})^2$$

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = S^2$$

ESTADÍSTICA CIRCULAR

$$\sum_{i=1}^n \text{sen}(\theta_i - \bar{x}_0) = 0$$

$$2(1 - \cos(\theta_i - \bar{x}_0))$$

$$\frac{1}{n} \sum_{i=1}^n 2(1 - \cos(\theta_i - \bar{x}_0)) = 2(1-r)$$

Tales analogías, ayudan a encontrar diversas medidas estadísticas para variables direccionales.

1.4.2. La mediana direccional.

Si se tiene un conjunto de observaciones direccionales $\theta_1, \theta_2, \dots, \theta_n$ se puede calcular una medición análoga a la mediana de la estadística lineal. Para este propósito se divide el círculo por un diámetro tal que la mitad de los puntos muestrales queden situados en una mitad del círculo y en la otra mitad quede el resto de la muestra.

Si denotamos los extremos del diámetro por P y Q asignando P al extremo en que se encuentren más concentrados los puntos muestrales, entonces, al vector \overline{OP} se le llama vector de la dirección mediana de la muestra.

Para calcular el ángulo se procede igual que en la sección 2.2.1. en el cálculo de \bar{x} , considerando los componentes correspondientes al vector \overline{OP} .

Si se tiene una distribución simétrica, la mediana se encuentra en el eje de simetría. La mediana puede no ser única, sin embargo, toda distribución unimodal tiene una mediana única.

1.4.2.a. La mediana direccional para datos agrupados.

Si se tienen datos agrupados, se puede calcular los componentes de la mediana direccional, empleando procedimientos análogos a los de estadística lineal.

Primero, se determina el intervalo de clase en el cual se encuentra la mediana direccional y se calcula :

$$M = L_i + \frac{j}{f_i} (U_i - L_i)$$

Donde L_i es el límite inferior del intervalo que contiene a la mediana, U_i el límite superior del intervalo que contiene a la mediana, j el número de observaciones que aun faltan para alcanzar la mediana una vez que se ha alcanzado el límite inferior del intervalo que contiene a la mediana y f_i la frecuencia del intervalo que contiene a la mediana.¹

1.4.3. La dirección modal.

La moda de un conjunto de observaciones direccionales se define de igual manera que en estadística lineal, como la dirección o punto en el círculo que ocurre con mayor frecuencia. Llámese a este punto M_o , entonces \overline{OM}_o es el vector correspondiente a la dirección modal.

Para el cálculo de la moda de datos agrupados; ésta se determina de acuerdo a las frecuencias de clase, obteniéndose, al igual que en estadística lineal, la clase modal, al observar la clase cuya frecuencia de clase sea mayor.

1. - La técnica empleada es análoga a la de estadística lineal. Esto se debe a que dentro de un intervalo la desviación es tan pequeña que se puede considerar lineal.

1.5. MEDIDAS DE DISPERSION.

Las medidas de tendencia central que se han analizado, para una distribución unimodal nos pueden indicar una dirección preferida, pero no contienen ninguna información de que tan dispersos o concentrados se encuentran los datos alrededor de esta dirección, la presente sección se dedica al análisis de distintas medidas de dispersión o variación.

1.5.1. Una medida de concentración.

Se menciona en la sección 1.4.1. el caso extremo en que todos los puntos muestrales caen en el mismo punto del círculo unitario, es decir todas las observaciones señalan la misma dirección. Este es el caso de máxima concentración y la longitud r del vector medio, es igual a 1. A medida que la concentración disminuye, el valor r se reduce. Si no hay concentración alrededor de una dirección específica, r toma el valor de cero.

Entonces, si se tiene una muestra unimodal, se cuenta con una medida de concentración en la longitud r del vector medio.

1.5.2. La varianza direccional y la desviación estandar direccional.

La longitud del vector medio, r , puede ser utilizado, también, para construir una medida de dispersión. Dado que decrece de 1 a cero cuando la dispersión aumenta, entonces es natural considerar $(1-r)$ como una medida de dispersión.

Ahora, recordando las analogías analizadas en la sección

1.4.1.c.

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2 \quad \text{y} \quad \frac{1}{n} \sum_{i=1}^n 2(1 - \cos(\theta_i - \bar{\theta})) = 2(1-r)$$

$2(1-r)$ es una medida de dispersión análoga a la estadística S^2 en el caso lineal; entonces, se define como varianza direccional a $s^2 = 2(1-r)$,

$$y \quad a \quad s = \left[2(1-r) \right]^{1/2} \quad \text{con } r = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{x}_\theta)$$

como la desviación estandar direccional¹ la cual está medida en radianes; si se desea tener el valor en grados se realiza la correspondiente conversión.

Esta última medida no es la única que se aproxima a la desviación estandar en forma asintótica, existen otras posibilidades, como por ejemplo:

$$s_o = (-2 \ln r)^{1/2}$$

que es propuesta por Mardia (1971); para valores de r cercanos a 1, s y s_o difieren muy poco, pero si r tiende a cero, s_o diverge a infinito mientras que s permanece finito.

1.5.3. Otras medidas de dispersión.

Para el caso de muestras grandes es posible considerar el rango como otra medida de dispersión.

El rango en el análisis de datos direccionales, se define como la longitud del arco más pequeño que contiene a todos los puntos muestrales. Se puede obtener de la siguiente manera.

Si $\theta_{(1)}, \theta_{(2)}, \dots, \theta_{(n)}$ son las estadísticas de orden para $\theta_1, \theta_2, \dots, \theta_n$ con $(0 < \theta_i \leq 2\pi)$ se definen las medidas de longitudes de arcos entre puntos adyacentes:

$$T_i = \theta_{(i)} - \theta_{(i-1)} \quad (i=1, \dots, n-1)$$

$$T_n = 2\pi - \theta_{(n)} + \theta_{(1)}$$

El rango circular está dado por :

1. - Esta medida fue propuesta por Batschelet (1981)

$$W = 2\pi - \max(T_1, T_2, \dots, T_n)$$

Para distribuciones circulares continuas el soporte es usualmente $(0, 2\pi)$, así que, en estos casos el rango es 2π .

Una última medida de dispersión, que se puede utilizar, es el rango intercuartílico, el cual se define como la suma de las longitudes de dos arcos; el primero de ellos se determina, midiendo a partir de la mediana, en sentido contrario a las manecillas del reloj, el arco tal que contiene una cuarta parte de las observaciones; el segundo arco se determina de la misma manera solo que se mide siguiendo el sentido de las manecillas del reloj.

1.5. MUESTRAS MULTIMODALES.

Todas las estadísticas propuestas hasta el momento, ya sean medidas de tendencia central ó medidas de dispersión, describirán adecuadamente el comportamiento de las observaciones si éstas presentan una distribución unimodal. Sin embargo, en diversas aplicaciones, la distribución que tienen las observaciones es multimodal; presentándose aunado a esto, el hecho de tener las modas a 180° de separación en el caso de ser bimodales y a 90° de separación en el caso de ser cuatrimodales.

Un ejemplo de datos que siempre proporcionan muestras bimodales, son los datos con eje mencionados en la sección 1.2.1. en los que solo interesa observar el ángulo de inclinación del eje, sin importar su dirección, razón por la cual, al graficarlos en un diagrama circular, obtenemos dos puntos opuestos por cada observación, dicho diagrama toma entonces, la forma de una figura simétrica centralmente. La posición de una de estas observaciones puede ser fijada si sólo consideramos su posición en el rango $(0^\circ, 180^\circ)$, dado que su posición en el rango $(180^\circ, 360^\circ)$ se puede determinar a partir de la primera. Para poder analizar estadísticamente este tipo de datos, se procede de la siguiente manera: dada

distribución es multimodal, con v modas, se modifica multiplicando cada ángulo por v .

$$v\theta_1, v\theta_2, \dots, v\theta_n$$

Una vez modificada la muestra, se reduce modulo 360° , con lo cual se obtiene una muestra unimodal, y se puede trabajar empleando las mediciones de las secciones anteriores. Para distinguirlas de las mediciones hechas a una muestra no modificada el vector medio, se denotará por m_v el ángulo medio por \bar{x}_v , la longitud del vector medio por r_v y la desviación estándar angular por s_v .

Si se deseara obtener \bar{x}_v en las medidas originales, simplemente se calcula;

$$\bar{x}_v = \bar{x}_0 / v$$

con lo cual obtenemos una de las v direcciones medias. si se desean obtener las $(v-1)$ restantes unicamente es necesario sumarle a \bar{x}_v el factor $2\pi/v$ en forma sucesiva para obtenerlas.

Si se desea tener la desviación estándar, también en las unidades originales, se calcula

$$s_1 = 1 - (1 - s_v^2)^{1/v}$$

CAPITULO 2

FUNCIONES DE DISTRIBUCION

2.1. INTRODUCCION.

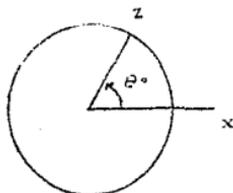
El objetivo principal al analizar una muestra de datos direccionales, es obtener información sobre la población de la cual fue obtenida. Para describir el comportamiento poblacional de variables aleatorias direccionales, se han desarrollado modelos matemáticos analogos a los de estadística lineal. En este capítulo se presentan estos modelos.

Inicialmente se estudia la función de distribución circular y su función de densidad; se presenta la función característica y su importancia en el desarrollo de funciones de distribución circulares; se analizan los momentos trigonométricos y para finalizar se presentan algunos modelos matemáticos que, análogamente como ocurre en estadística lineal, describen distribuciones poblacionales.

2.2. LA FUNCION DE DISTRIBUCION.

Sea Z una variable aleatoria que toma valores en la circunferencia de un círculo unitario. $x^2 + y^2 = 1$, del plano \overline{XY} ; se pueden identificar los posibles valores de Z con ángulos θ , medidos en radianes, a partir del eje X en sentido contrario a las manecillas del reloj, de forma tal que, el ángulo θ representa el punto $(\cos \theta, \sin \theta)$, con $0 \leq \theta \leq 2\pi$.

Fig.



En base a esta representación, se define la función de distribución de la variable aleatoria Z por la ecuación:

$$F(\theta) = \Pr(0 < Z \leq \theta), \quad 0 < \theta \leq 2\pi$$

Esta probabilidad se obtiene, al igual que en estadística lineal, calculando

$$\Pr(0 < Z \leq \theta) = \int_0^\theta dF(u)$$

Donde la integral, es una integral de Lebesgue Stieljes.⁽¹⁾

Un contraste que surge con respecto a las funciones de distribución lineales,⁽²⁾ es el hecho de que el rango de Z siempre es finito, lo que implica que el rango en que $F(\theta)$ existe es también finito.

El rango más grande que se puede tener, es aquel que comprende la circunferencia del círculo unitario por completo $(0, 2\pi)$. Es posible tener una función de distribución circular cuya rango sea un subconjunto de $(0, 2\pi)$ en cuyo caso se tiene un rango más pequeña, pero no existen rangos mayores a $(0, 2\pi)$.

La función de distribución circular tiene las siguientes propiedades:

- i) $F(2\pi) = \lim_{\theta \rightarrow 2\pi} F(\theta) = 1$
- ii) $F(0) = \lim_{\theta \rightarrow 0} F(\theta) = 0$
- iii) $F(\theta)$ es una función monótona no decreciente.
- iv) $F(\theta)$ es una función continua por la derecha.

Las propiedades i) e ii) marcan la dependencia de la función de distribución circular con respecto a la dirección

1. - Rudin (1960).

2. - Funciones de distribución usuales de Estadística lineal, como la Normal, Binomial, etc.

cero elegida, razón por la cual, si se desea considerar a θ en el rango $-\pi < \theta \leq \pi$ entonces la dirección cero es $-\pi$, y es necesario añadir la constante:

$$A = \Pr(-\pi < \theta \leq 0) = \int_{-\pi}^0 dF(u)$$

a la función de distribución $F(\theta)$ para que se cumplan las propiedades:

$$i) F(\pi) = \lim_{\theta \rightarrow \pi} F(\theta) = 1$$

$$ii) F(-\pi) = \lim_{\theta \rightarrow -\pi} F(\theta) = 0$$

2.2.1. La función de densidad circular.

Sea Z una variable aleatoria continua que toma valores en la circunferencia de un círculo unitario, si existe una función f tal que:

$$F(\theta) = \int_0^{\theta} dF(u) = \int_0^{\theta} f(u) du \quad 0 < \theta \leq 2\pi$$

Entonces la función de distribución $F(\theta)$ de la variable aleatoria Z se dice que es una función absolutamente continua.

Definición:

Al igual que en estadística lineal, se define a la función de densidad de la variable aleatoria Z como la función f tal que:

$$F(\theta) = \int_0^{\theta} f(u) du \quad 0 < \theta \leq 2\pi$$

Si existe la función de densidad $f(\theta)$, es posible obtenerla a partir de la función de distribución, calculando:

$$f(\theta) = \frac{dF(\theta)}{d\theta} = \lim_{\Delta \rightarrow 0} \frac{F(\theta + \Delta\theta) - F(\theta - \Delta\theta)}{2\Delta\theta}$$

De donde $2\Delta\theta(f(\theta)) \approx F(\theta + \Delta\theta) - F(\theta - \Delta\theta)$ así, la

probabilidad de que Z este en un pequeño intervalo que contiene al valor θ es aproximadamente igual a $f(\theta)$ veces la longitud del intervalo.

Las funciones de densidad que describen el comportamiento de poblaciones direccionales deben cumplir con las siguientes propiedades:

$$\begin{aligned}
 & i) \quad f(\theta) \geq 0 \qquad \qquad \qquad 0 < \theta \leq 2\pi \\
 & ii) \quad \int_0^{2\pi} f(\theta) d\theta = 1
 \end{aligned}$$

Es posible pensar en variables aleatorias circulares discretas, pero en la práctica son difíciles de encontrar, para estas variables la correspondiente función de densidad es

$$f(\theta) = P(Z = \theta), \quad 0 < \theta \leq 2\pi$$

y la función de distribución

$$F(\theta) = \text{Pr}(0 < Z \leq \theta), \quad 0 < \theta \leq 2\pi$$

2.3. LA FUNCION CARACTERISTICA.

Una herramienta importante de la teoría de la probabilidad que es útil en el estudio de convergencia en distribuciones es la función característica. En la presente sección se analiza la función característica de variables aleatorias direccionales.

Definición:

Sea θ una variable aleatoria asociada a una distribución circular F . La esperanza de la variable θ se define en forma análoga que en estadística lineal como:

$$E(\theta) = \int_0^{2\pi} \theta \, dF(\theta)$$

Antes de definir la función característica circular se analiza primero a la variable aleatoria $\exp(it\theta)$, la cual es posible reescribirla como:

$$e^{it\theta} = \cos(t\theta) + i \sin(t\theta)$$

$e^{it\theta}$ es una variable aleatoria compleja de módulo uno:

$$|e^{it\theta}| = (\cos^2(t\theta) + \sin^2(t\theta))^{1/2} = 1$$

Calculando la esperanza de esta variable aleatoria se tiene:

$$E(e^{it\theta}) = \int_0^{2\pi} e^{it\theta} \, dF(\theta)$$

Donde $F(\theta)$ es la función de distribución asociada a la variable aleatoria θ .

Si se calcula $E(e^{it(\theta + 2\pi)})$ se obtiene,

$$E(e^{it(\theta + 2\pi)}) = \int_0^{2\pi} e^{it(\theta + 2\pi)} \, dF(\theta)$$

Cuando t pertenece a los Enteros, ($t \in \mathbb{Z}$), esta última expresión se reduce a

$$\int_0^{2\pi} e^{it\theta} e^{it2\pi} \, dF(\theta) = e^{it2\pi} \int_0^{2\pi} e^{it\theta} \, dF(\theta) = \int_0^{2\pi} e^{it\theta} \, dF(\theta)$$

dado que $e^{it2\pi} = \cos(t2\pi) + i \sin(t2\pi) = 1$ si $t \in \mathbb{Z}$

Entonces, considerando este hecho, se define la función característica para variables aleatorias circulares como:

$$\phi(p) = \int_0^{2\pi} e^{ip\theta} \, dF(\theta) \quad p = 0, \pm 1, \pm 2, \dots$$

Se puede reescribir de la siguiente manera,

$$\begin{aligned}\phi(p) &= \int_0^{2\pi} e^{ip\theta} dF(\theta) = \int_0^{2\pi} (\cos(p\theta) + i \sin(p\theta)) dF(\theta) \\ &= \int_0^{2\pi} \cos(p\theta) dF(\theta) + i \int_0^{2\pi} \sin(p\theta) dF(\theta)\end{aligned}$$

luego,

$$\phi(p) = E(\cos(p\theta)) + i E(\sin(p\theta))$$

finalmente si se nombra

$$\alpha_p = E(\cos(p\theta)) \quad \text{y} \quad \beta_p = E(\sin(p\theta))$$

entonces

$$\phi(p) = \alpha_p + i\beta_p$$

Es sencillo comprobar que $\alpha_p = \alpha_{-p}$ y $\beta_p = -\beta_{-p}$

2.3.1. Propiedades de la función característica.

- i) $\phi(0) = 1$
- ii) $\overline{\phi(p)} = \phi(-p)$
- iii) $|\phi(p)| \leq 1$

$$\text{Dem. } |\phi(p)| = |E(e^{ip\theta})| \leq E|e^{ip\theta}| = 1$$

iv) Unicidad y el leorema de inversión.

Al igual que en la línea, una función de distribución $F(\theta)$ está determinada en forma única por su función característica $\phi(\theta)$ e inversamente, conociendo la función característica se puede, en principio, conocer la función de distribución. Existen una gran colección de fórmulas de inversión, pero en la práctica, estas no son fáciles de aplicar; aun en el caso lineal. Sin embargo, las funciones características en general, son un caso especial de las transformadas de Fourier. En especial, los números $2\pi\phi(p)$ son los coeficientes de Fourier-Stieltjes de $F(\theta)$ y

además,

$$\frac{dF(\theta)}{d\theta} = \left[\frac{1}{2\pi} \right] \sum_{p=-\infty}^{\infty} \phi(p) e^{-ip\theta}$$

Donde $\phi(p)$ es la función característica asociada a la función de distribución $F(\theta)$. Si la serie $\sum_{p=1}^{\infty} |\phi(p)|^2$ es convergente entonces la variable aleatoria θ tiene una densidad la cual está definida casi donde quiera por:

$$f(\theta) = \left[\frac{1}{2\pi} \right] \sum_{p=-\infty}^{\infty} \phi(p) e^{-ip\theta}$$

y es posible reescribirla de la siguiente manera,

$$f(\theta) = \left[\frac{1}{2\pi} \right] \left[1 + 2 \sum_{p=1}^{\infty} \left[\alpha_p \cos p\theta + \beta_p \sin p\theta \right] \right]$$

o, independencia y convolución.

Sean θ_1 y θ_2 dos variables aleatorias independientes, cuya función característica conjunta está definida por:

$$\phi(p, q) = E(e^{ip\theta_1 + iq\theta_2}) \quad p, q \in \mathbb{Z}$$

Sean $\phi(p)$ y $\phi(q)$ las funciones características marginales de (θ_1, θ_2) entonces, al igual que en el caso lineal, $\phi(p, q) = \phi(p)\phi(q)$. Este resultado se generaliza para el caso de n variables aleatorias independientes.

Ahora, sean θ y ξ dos variables aleatorias independientes, entonces, al igual que en la línea, la función de distribución de $\zeta = (\theta + \xi) \bmod 2\pi$ está dada por:

$$F(\zeta) = \int_0^{2\pi} dF(\xi - \theta) dF(\theta)$$

Si una de las variables tiene una densidad, la convolución también tiene una densidad. Cuando θ y ξ tienen ambas una densidad se reduce a

$$F(\xi) = \int_0^{2\pi} f_1(\xi - \theta) f_2(\theta) d\theta$$

2.4. MOMENTOS DE LAS DISTRIBUCIONES CIRCULARES.

Las distribuciones desarrolladas en estadística lineal pueden caracterizarse también por sus momentos. El primer momento o esperanza de la variable aleatoria x proporciona la media de la población μ .

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx = \mu$$

Y se define el segundo momento por :

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \mu$$

Una vez calculados ambos momentos se obtiene la varianza poblacional $\sigma^2 = E(x^2) - (E(x))^2$

Para definir otras medidas se utilizan momentos de orden superior, y se calcula $E(x^k)$ $k=1,2,3,\dots$

En el estudio de distribuciones circulares también es posible calcular momentos para poder describirlas. Sin embargo, los momentos empleados en estadística lineal no pueden ser aplicados a las distribuciones circulares debido a la periodicidad que se presenta en éstas. Se define un nuevo tipo de momentos llamados momentos trigonométricos.

2.4.1. Primer momento trigonométrico.

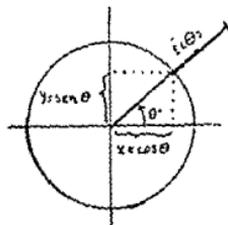
Si se tiene una función de distribución circular $F(\theta)$ con función de densidad $f(\theta)$, $0 < \theta \leq 2\pi$; entonces, para

cada θ es posible asociar un vector unitario con componentes,

$$x = \cos \theta \quad , \quad y = \sin \theta$$

Luego para cada pareja (x,y) , definida así, se le tiene asignado un valor de $f(\theta)$.

Fig.



Es posible calcular el valor medio de x e y por medio de las fórmulas,

$$\alpha_1 = \int_0^{2\pi} x f(\theta) d\theta = \int_0^{2\pi} \cos \theta f(\theta) d\theta$$

$$\beta_1 = \int_0^{2\pi} y f(\theta) d\theta = \int_0^{2\pi} \sin \theta f(\theta) d\theta$$

Los valores medios α_1 y β_1 son los componentes del vector medio que determina la media direccional. Debido a las funciones trigonométricas que se tienen bajo el signo de integral, a este vector se le llama momento trigonométrico.

$$K_1 = \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix}$$

Este momento trigonométrico se define como el primer momento. Es posible convertirla a coordenadas polares,

$$M_1 = (\rho_1, \mu_1)$$

$$\text{con } \rho_1 = (\alpha_1^2 + \beta_1^2)^{1/2}$$

$$\mu_1 = \begin{cases} \arctan(\beta_1/\alpha_1) & \text{si } \alpha_1 > 0. \\ 180^\circ + \arctan(\beta_1/\alpha_1) & \text{si } \alpha_1 < 0. \end{cases}$$

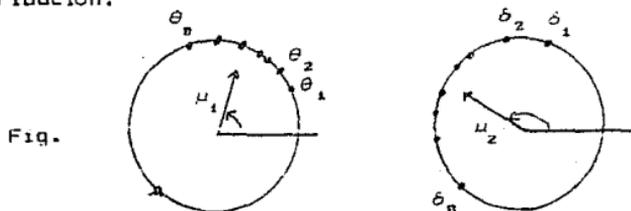
o en casos excepcionales,

$$\mu_1 = \begin{cases} 90^\circ & \text{si } \alpha_1 = 0 \text{ y } \beta_1 > 0 \\ 270^\circ & \text{si } \alpha_1 = 0 \text{ y } \beta_1 < 0 \\ \text{indeterminada} & \text{si } \alpha_1 = 0 \text{ y } \beta_1 = 0 \end{cases}$$

ρ_1 es la longitud del vector medio y μ_1 es el ángulo medio. Entonces M_1 proporciona una dirección media y además una medida de concentración alrededor de la dirección media, lo que muestra el hecho de que los momentos trigonométricos contienen más información que un momento ordinario.

2.4.2. Momentos trigonométricos de orden mayor.

Para obtener el segundo momento trigonométrico M_2 se hace doble cada ángulo, es decir, para cada ángulo θ se calcula $\delta = 2\theta$. Los ángulos δ presentan una nueva distribución.



La nueva distribución tiene su propio vector medio

$$M_2 = \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix}$$

$$\text{donde } \alpha_2 = \int_0^{2\pi} \cos 2\theta f(\theta) d\theta$$

$$\beta_2 = \int_0^{2\pi} \text{sen } 2\theta f(\theta) d\theta$$

Al vector medio M_2 se le llama segundo momento trigonométrico; es posible calcular $M_2 = (\rho_2, \mu_2)$ de la misma forma en que se hace para M_1 .

En general, los momentos trigonométricos de cualquier orden se definen por:

$$M_n = \begin{bmatrix} \alpha_n \\ \beta_n \end{bmatrix} \quad n = 1, 2, 3, \dots$$

$$\text{donde } \alpha_n = \int_0^{2\pi} \cos n\theta f(\theta) d\theta$$

$$\beta_n = \int_0^{2\pi} \text{sen } n\theta f(\theta) d\theta$$

Si se cambia la n por p y se reescribe α_p y β_p de la siguiente manera, $\alpha_p = E(\cos p\theta)$, $\beta_p = E(\text{sen } p\theta)$ se puede recordar de la sección 2.3.1 la definición de la función característica $\phi(p) = \alpha_p + i\beta_p$ $p = 0, \pm 1, \pm 2, \dots$

Entonces la sucesión de momentos trigonométricos quedan determinados por la función característica. Consecuentemente, todos los resultados de la función característica se aplican a los momentos trigonométricos.

2.4.3. Momentos trigonométricos centrales.

El p -ésimo momento trigonométrico central se define como

$$M_p = E(e^{ip(\theta - \mu)}) = \alpha_p + i\beta_p$$

$$\text{donde } \bar{\alpha}_p = E(\cos p(\theta - \mu_1)) = \rho_p \cos(\mu_p^\circ - p\mu_1)$$

$$\bar{\beta}_p = E(\sin p(\theta - \mu_1)) = \rho_p \sin(\mu_p^\circ - p\mu_1)$$

En particular para $p = 1$

$$\bar{\alpha}_1 = E(\cos(\theta - \mu_1)) = \rho_1$$

$$\bar{\beta}_1 = E(\sin(\theta - \mu_1)) = 0$$

Con lo cual se define la varianza circular como

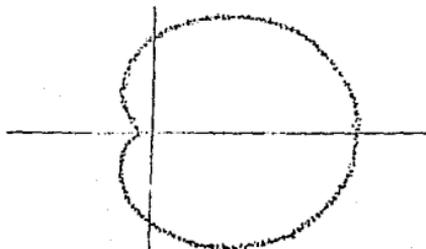
$$V_c = 1 - \bar{\rho}_1^2$$

$V_c = 0$ si y solo si la distribución se concentra en un solo punto $\theta = \mu_1$. Esta propiedad es similar a la de σ^2 en la línea. Si $V_c = 1$, se dice que la distribución tiene una variación máxima y es tan dispersa que no hay concentración hacia ninguna dirección en particular.

2.5. FUNCIONES DE DISTRIBUCION CIRCULARES.

En esta sección se presentan algunas funciones e distribución circulares. Se les llama circulares debido a que el análisis de direcciones se restringe, en el presente trabajo, al caso de poblaciones en dimensión 2. Las distribuciones de este tipo de poblaciones se representan en el plano cartesiano, tomándose como base la circunferencia de un círculo unitario. Se puede tener una idea intuitiva de una distribución circular al observar la siguiente grafica de una función de densidad $f(\theta)$ en coordenadas polares,

Fig.



En contraste con las funciones de distribución en estadística lineal, el rango de las distribuciones circulares es siempre finito, siempre está entre 0° y 360° o en radianes de 0 a 2π . Por lo general, las distribuciones circulares teóricas son continuas, es decir, existe una función continua del ángulo θ . La función de densidad es siempre positiva o cero y la integral sobre el rango completo es igual a 1.

2.5.1. Distribuciones simétricas.

Definición: Se dice que una función de distribución es simétrica alrededor del ángulo ξ , si su función de densidad cumple con la siguiente igualdad:

$$f(\xi - \theta) = f(\xi + \theta)$$

Si esto ocurre, también se dice que es simétrica alrededor del ángulo $\xi + \pi$.

Las distribuciones simétricas unimodales tienen la moda, la media y la mediana direccional determinadas por una misma dirección.

Los modelos de distribuciones circulares que se presentan a continuación en su mayoría se trata de distribuciones simétricas.

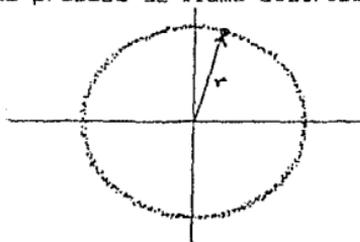
2.5.2. Modelos circulares.

Los modelos circulares desarrollados para poblaciones direccionales, son en cierta forma análogos a los modelos lineales. En esta sección se presentan los modelos discretos de la distribución Puntual, Lattice y Uniforme discreta; así como los modelos de cuatro distribuciones continuas: la distribución Uniforme, la Cardioide, la Triangular y la distribución von Mises.

2.5.2.a. La distribución puntual.

Si solo una dirección, en el plano, es escogida con probabilidad 1, es decir, si $P(\theta = \mu_0) = 1$, entonces la distribución que produce se llama distribución puntual.

Fig.



La función característica de esta distribución es,

$$\phi(p) = e^{ip\mu_0} = \cos p\mu_0 + i \sin p\mu_0.$$

Si se calcula el primer momento central, se tiene el caso en que $\bar{\mu}_1 = 1$, consecuentemente $V_0 = 1 - \bar{\mu}_1^2 = 0$ que indica una concentración máxima.

2.5.2.b. Distribución Lattice.

Si un conjunto discreto de direcciones equidistantes son observadas, no necesariamente con la misma probabilidad, se tiene entonces una distribución llamada Lattice.

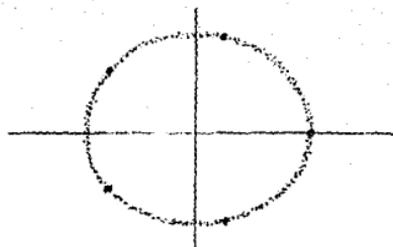
Se define a la distribución Lattice por,

$$P(\theta = v + 2\pi r/m) = p_r \quad r = 1, 2, 3, \dots, m-1.$$

$$\text{con } p_r \geq 0 \quad \text{y} \quad \sum_{r=0}^{m-1} p_r = 1.$$

Esta distribución puede ser imaginada como concentraciones en los vértices de un polígono regular de m lados.

Fig.



Si se tiene el caso en que $\rho_r = 1/m$ se le llama distribución discreta uniforme. Para $v=0$ la función característica está dada por:

$$\phi(p) = \sum_{r=0}^{m-1} \rho_r e^{2\pi r p i / m}$$

2.5.2.c. La distribución uniforme.

Si θ se distribuye uniformemente sobre $(0, 2\pi)$, la función de probabilidad de θ , medida en radianes es:

$$f(\theta) = 1/2\pi \quad 0 < \theta \leq 2\pi$$

Esta función es constante sobre la circunferencia del círculo unitario.

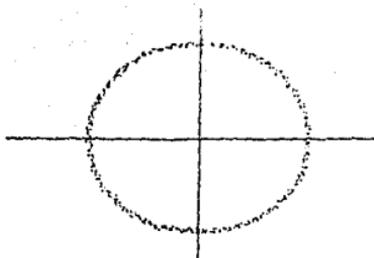
La función característica está dada por:

$$\phi(p) = \left[\frac{1}{2\pi} \right] \int_0^{2\pi} e^{i p \theta} d\theta$$

entonces

$$\phi(p) = \begin{cases} 1 & \text{si } p = 0 \\ \frac{1}{2\pi p} \left[e^{2\pi p i} - 1 \right] & \text{en otra caso.} \end{cases}$$

Fig.



Se puede considerar como centro de simetría el origen. Como no existe una dirección preferida, el valor de ρ_1 es cero y consecuentemente $V_0 = 1 - \rho_1 = 1$, es decir, no hay concentración hacia ninguna dirección en particular.

2.5.2.d. La distribución cardiode.

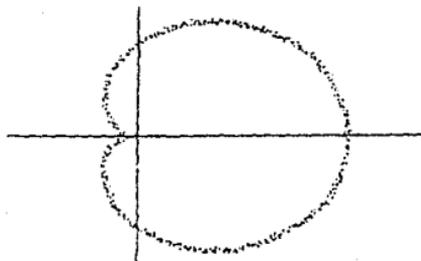
Esta distribución esta determinada por la función

$$f(\theta) = (2\pi)^{-1} [1 + 2\rho \cos(\theta - \theta_0)]$$

con $0 < \theta \leq 2\pi$, $|\rho| < 1/2$

La representación polar de esta densidad es una curva cardiode.

Fig.



La distribución es simétrica y unimodal. Para $\rho=0$ se reduce a la distribución uniforme. En esta distribución $\theta_0 = \mu_0$ y $V_0 = 1 - \rho_1$.

2.5.2.e. La distribución triangular.

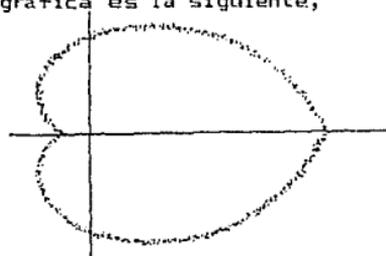
Se dice que una variable aleatoria θ tiene una distribución triangular si su función de densidad es

$$f(\theta) = \frac{1}{2\pi} \left\{ 4 - \pi^2 \rho + 2\pi \rho |\pi - \theta| \right\}$$

con $0 < \theta \leq 2\pi$, $\rho \leq \frac{4}{\pi^2}$

Esta distribución es simétrica alrededor de $\theta = \pi$, su representación gráfica es la siguiente,

Fig.



Si $\rho = 0$ se convierte en la distribución uniforme. Para esta distribución $\alpha_{2p} = 0$, $\alpha_{2p-1} = \rho / (2p-1)^2$ y $\beta_p = 0$.

2.5.2.f. La distribución von Mises.

La variable aleatoria circular θ tiene una distribución von Mises si su función de densidad está dada por:

$$f(\theta) = (2\pi I_0(k))^{-1} e^{k \cos(\theta - \mu_0)}$$

$$0 < \theta \leq 2\pi \quad , \quad k > 0 \quad \text{y} \quad 0 \leq \mu_0 < 2\pi .$$

donde $I_0(k)$ es una función de Bessel modificada de primera clase⁽¹⁾,

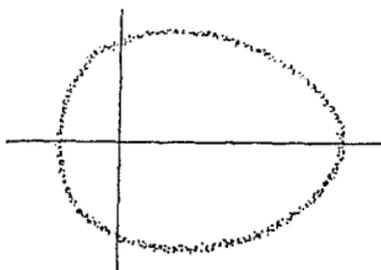
$$I_0(k) = \frac{1}{2\pi} \int_0^{2\pi} e^{k \cos \theta} d\theta = \sum_{r=0}^{\infty} (r!)^{-2} (1/2 k)^{2r}$$

La función de densidad von Mises tiene dos parámetros $k \geq 0$ y μ_0 . La función toma su máximo valor en $\theta = \mu_0$, entonces μ_0 es su moda y por ser simétrica con respecto a su

1. -Mardia(1972) Apéndice 1.

moda, μ_0 , es también su media direccional. El parámetro k es una medida de concentración, si k es grande la distribución está más concentrada alrededor de la media μ_0 , si $k = 0$ la distribución von Mises se convierte en una distribución uniforme.

Fig.



La longitud del vector medio está determinada por

$$\rho = A(k)$$

La función $A(k)$ es un cociente de funciones de Bessel de primera clase de orden uno y cero.⁽⁴⁾

$$A(k) = \frac{I_1(k)}{I_0(k)} \quad k \geq 0$$

La función $A(k)$ es monótonamente creciente a partir del valor $A(0)=0$ en adelante. Alcanza el valor 1 asintóticamente cuando k tiende a infinito.

La distribución von Mises es frecuentemente usada y tiene una importancia similar a la distribución Normal en estadística lineal. Esta distribución fue introducida por Von Mises en 1918.

2.5.3. Distribuciones Envueltas.

Las distribuciones envueltas son transformaciones de algunos modelos de estadística lineal.

Dada una distribución en la línea es posible envolverla alrededor de la circunferencia de un círculo de radio

1. -Mardia(1972) Apéndice 1.

unitario mediante el siguiente procedimiento: a partir de un punto fijo en el círculo por ejemplo $\theta=0$, se hace coincidir la media de la distribución lineal, luego se dan vueltas en sentido positivo y en sentido negativo, y se cargan los distintos sectores del círculo con la densidad de probabilidad que cae en cada vuelta.

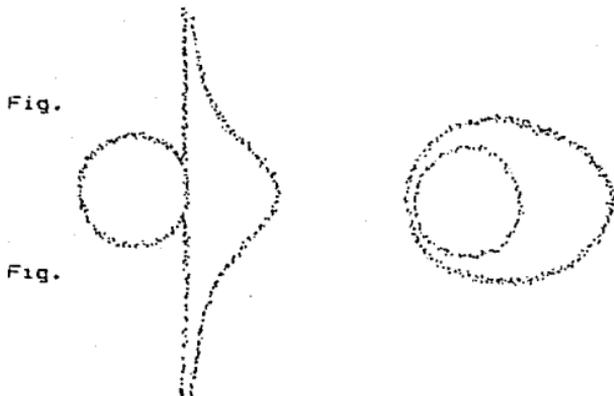


Fig.

Fig.

La distribución resultante es circular y se le llama "Distribución envuelta".

Matemáticamente, si x es una variable aleatoria en la línea con función de distribución $F(x)$, entonces la variable aleatoria x_v de la distribución envuelta está dada por:

$$x_v = x(\text{mod } 2\pi)$$

La función de distribución envuelta está dada por:

$$F_v(\theta) = \sum_{k=-\infty}^{\infty} \{F(\theta+2\pi k) - F(2\pi k)\} \quad 0 < \theta \leq 2\pi$$

Si x tiene una función de densidad, la correspondiente función de densidad de x_v es:

$$f_v(\theta) = \sum_{k=-\infty}^{\infty} f(\theta+2\pi k) \quad 0 < \theta \leq 2\pi$$

Las funciones de distribución envueltas cumplen con las siguientes propiedades:

1) $\phi(t) = \phi(p)$

donde $\phi(t)$ es la función característica de la función de densidad lineal y $\phi(p)$ es la función característica de la correspondiente función envuelta.

Dem.

$$\begin{aligned} \phi(p) &= \int_0^{2\pi} e^{i p \theta} dF_V(\theta) = \sum_{k=-\infty}^{\infty} \int_{2\pi k}^{2\pi(k+1)} e^{i p \theta} dF(\theta) \\ &= \int_{-\infty}^{\infty} e^{i p x} dF(x) = \phi(t). \end{aligned}$$

ii) Si $\phi(t)$ es integrable entonces x tiene una densidad

$$f_V(\theta) = 1/2\pi \sum_{p=-\infty}^{\infty} \phi(p) e^{i p \theta}$$

iii) El proceso de desenvolver una distribución en el círculo no da lugar a una distribución única en la línea. El proceso de envoltura no es reversible.

2.5.3.a. La distribución Poisson envuelta.

La función Poisson está definida por

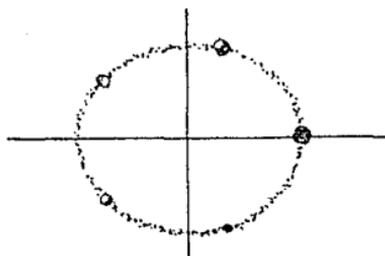
$$P(x; \lambda) = e^{-\lambda} \lambda^x / x! \quad x = 0, 1, 2, \dots \quad \lambda \geq 0$$

La distribución de Poisson envuelta se define por:

$$P(\theta = 2\pi r/m) = \sum_{k=0}^{\infty} P(r+k m; \lambda) \quad r = 0, 1, 2, \dots, m-1.$$

el entero m indica el número de puntos en la circunferencia que se cargarán con la probabilidad. Por ejemplo si $m=5$ su gráfica es:

Fig.



El punto $\theta=0$ tiene la siguiente probabilidad,

$$P(\theta=0) = \sum_{k=0}^{\infty} P(5k; \lambda) = \sum_{k=0}^{\infty} e^{-\lambda} \lambda^{5k} / (5k)!$$

$$P(\theta=0) = e^{-\lambda} + e^{-\lambda} \lambda^5 / 5! + e^{-\lambda} \lambda^{10} / 10! + \dots$$

El punto $\theta = (2\pi/5)$ tiene la probabilidad,

$$P(\theta=2\pi/5) = \sum_{k=0}^{\infty} P(1+5k; \lambda) = \sum_{k=0}^{\infty} e^{-\lambda} \lambda^{(1+5k)} / (1+5k)!$$

$$= e^{-\lambda} \lambda + e^{-\lambda} \lambda^6 / 6! + e^{-\lambda} \lambda^{11} / 11! + \dots$$

Estas funciones son funciones decrecientes de r .

La función característica de θ es

$$\phi(p) = \exp(-\lambda + \lambda e^{2\pi p i / m})$$

2.5.3.b. La distribución Normal envuelta.

Sea x que se distribuye $N(0, \sigma^2)$, la función de densidad de la distribución Normal envuelta está dada por,

$$f_{\nu}(\theta) = (2\pi\sigma^2)^{-1/2} \sum_{k=-\infty}^{\infty} \exp\{-(\theta + 2\pi k)^2 / 2\sigma^2\} \quad 0 < \theta \leq 2\pi$$

$$\phi(t) = \exp\{-\frac{1}{2}t^2\sigma^2\} \quad \text{entonces} \quad \phi(p) = \exp\{-\frac{1}{2}p^2\sigma^2\}$$

$$\text{donde } \alpha_p = \exp\{-\frac{1}{2}p^2\sigma^2\} \quad \text{y} \quad \beta_p = 0.$$

Usando la representación de $f(\theta)$ presentada en la sección 2.3.1. se obtiene la sig. representación de la misma densidad.

$$f_{\nu}(\theta) = \left(1 + 2 \sum_{p=1}^{\infty} \rho^p \cos p\theta \right) / (2\pi)$$

$$0 < \theta \leq 2\pi \quad 0 \leq \rho \leq 1.$$

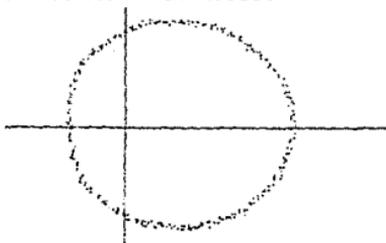
donde

$$\rho = e\left(-\frac{1}{2}\nu^2\right), \quad \nu^2 = -2 \log \rho \quad \dots$$

$$\mu_{\nu} = 0 \quad \vee \quad \nu_{\nu} = 1 - \rho$$

La distribución es unimodal y simétrica alrededor de $\theta = 0$. Cuando $\rho \rightarrow 0$ la función de densidad tiende a la función de densidad Uniforme. Si $\rho = 1$ la distribución se concentra en un solo punto. La gráfica es similar a la gráfica de la distribución von Mises.

Fig.



2.5.3.c. La distribución Cauchy Envuelta.

$$f(x;a) = \frac{a}{\pi(a^2 + x^2)} \quad -\infty < x < \infty$$

es la función de densidad de la distribución Cauchy en la línea. La función característica correspondiente es,

$$f(t) = \exp\{-a|t|\}$$

Entonces utilizando el mismo procedimiento de la sección 2.3.1. se define a la distribución Cauchy envuelta

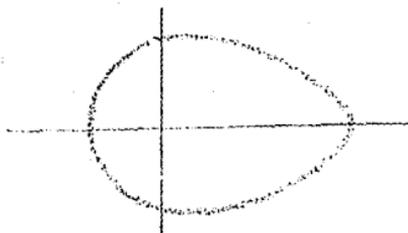
$$f(\theta;a) = (2\pi)^{-1} \left(1 + 2 \sum_{p=1}^{\infty} \rho^p \cos p\theta \right)$$

donde $\rho = e^{-\alpha}$, $\nu = 1-\rho$

$$f(p) = \rho |p|^\alpha , \quad \alpha_p = \rho^p , \quad \beta_p = 0$$

Cuando $\rho \rightarrow 0$ la distribución tiende a la distribución uniforme y cuando $\rho \rightarrow 1$ se concentra en un solo punto.

Fig.



2.5.4. Distribuciones Multimodales.

Algunos experimentos dan lugar a distribuciones multimodales, un ejemplo son las observaciones con eje que se nombran en la sección 1.2., en las cuales, solo se observa la posición de los ejes, creando una distribución bimodal con modas a 180° de separación.

Para describir y modelar poblaciones cuya distribución es multimodal se ha recurrido a mezclar las distribuciones unimodales. A continuación se presentan dos formas distintas de crear una distribución multimodal.

Si se tiene una distribución bimodal circular, cuyas modas están a 180° de separación, se puede describir por la función.

$$g(\theta) = \lambda f(\theta) + (1-\lambda) f(\theta+\pi), \quad 0 \leq \theta \leq 2\pi, \quad 0 \leq \lambda \leq 1$$

Donde $f(\theta)$ es una función de densidad circular unimodal. Otra forma de crear una distribución multimodal es realizando la transformación de los ángulos

$$\theta \rightarrow \theta + \frac{k}{n} 2\pi \text{ (mod } 2\pi)$$

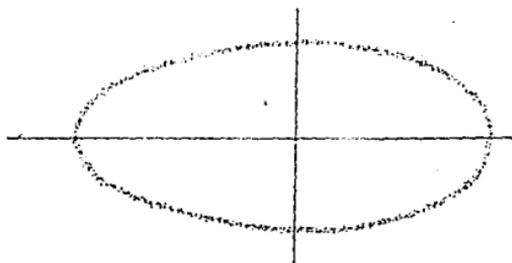
Con esta transformación se obtiene una nueva población cuya distribución que es unimodal y es posible ajustarle cualquiera de los modelos unimodales.

Un ejemplo de una distribución multimodal de la forma von Mises es

$$f(\theta) = (2\pi I_0(k))^{-1} \exp\{k \cos v(\theta - \mu)\}$$

Donde v denota el número de modas presente en la distribución multimodal.

Fig.



2.5.5. Representación gráfica en coordenadas rectangulares.

En general, las gráficas en coordenadas polares, de las funciones de densidad, son muy comunes pero también es posible graficarlas en coordenadas rectangulares.

Para elaborar una gráfica en coordenadas rectangulares es necesario elegir un punto de corte de la gráfica circular, una vez elegido se grafica a partir de ese punto la correspondiente gráfica en coordenadas rectangulares.

A continuación se muestran tres distintos puntos de corte de una misma gráfica circular y las correspondientes gráficas en coordenadas rectangulares.

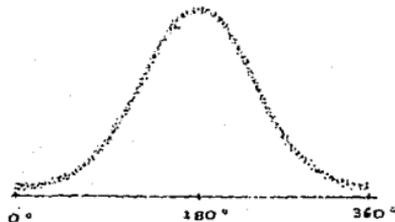


Fig.

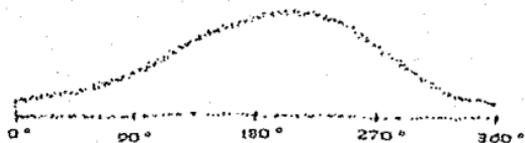


Fig.



Fig.



Como se puede apreciar con la primera grafica lineal, es posible producir graficas parecidas a las de estadística lineal cuya interpretación se puede realizar en forma análoga a las graficas de estadística lineal.

Ahora al observar las otras dos graficas es posible notar que se tiene una desventaja, el punto inicial en la línea es arbitrario y este hecho puede causar confusión al interpretar la gráfica. Sin embargo, conociendo este riesgo se pueden elaborar graficas en coordenadas rectangulares teniendo cuidado de seleccionar un punto de corte tal que al extender la gráfica el centro de ésta coincida aproximadamente con la moda de la distribución. Del mismo modo, para el caso de distribuciones multimodales el punto de corte se puede controlar y seleccionar como mas convenga.

CAPITULO 3

INFERENCIA ESTADISTICA

3.1. INTRODUCCION.

El analizar una muestra de datos direccionales se hace con el propósito de resumir la información que ésta contiene; este resumen se obtiene calculando las estadísticas analizadas en el capítulo 1.

Una vez que se ha resumido la información, se pretende encontrar un modelo que describa el comportamiento de la población de la cual proviene la muestra.

En el capítulo 2 se han presentado diversas funciones de distribución circulares, las cuales son determinadas por algunos parámetros. El presente capítulo se dedica a la estimación de parámetros y ajuste de funciones de distribución a una muestra de observaciones direccionales.

Siendo importante comprobar si la estimación y/o ajuste son aceptables se presentan algunas de las pruebas de hipótesis empleadas en el análisis de datos direccionales.

3.2. ESTIMACION DE PARAMETROS.

Sea Z una variable aleatoria circular, con función de distribución conocida $F(z; \delta)$, de la cual sólo se desconoce el valor del parámetro δ entonces, al igual que en estadística lineal, es posible estimar el valor del parámetro desconocido si se cuenta con una muestra aleatoria de valores z_1, z_2, \dots, z_n obtenidos de la población cuya función de distribución es $F(z; \delta)$.

Esta estimación puede ser de dos formas: puntual o por intervalos. En este capítulo se presenta la aplicación del método de estimación puntual, usado en estadística lineal, llamado método de máxima verosimilitud.

3.2.1. Estimación puntual máximoverosimil.

El principio de máxima verosimilitud empleado en la estimación puntual de parámetros en estadística lineal, también es empleado en el análisis de datos direccionales.

Este método consiste en seleccionar como estimadores aquellos valores de los parámetros que maximizan la función de verosimilitud.

La función de verosimilitud se define como la función de densidad conjunta de n variables aleatorias

$f_{x_1, \dots, x_n}(x_1, x_2, \dots, x_n; \delta)$
considerándola como una función del parámetro desconocido δ , ésta se denota por $L(\delta)$.

Las funciones de verosimilitud en general satisfacen condiciones de regularidad y la solución de la ecuación

$$\frac{dL(\delta)}{d\delta} = 0$$

es el estimador máximoverosimil, si éste determina un máximo de la función de verosimilitud y se denota por $\hat{\delta}$.

Las funciones $L(\delta)$ y $\log L(\delta)$ alcanzan ambas su máximo en el mismo valor de $\hat{\delta}$. Siendo más fácil encontrar el máximo de la función $\log L(\delta)$, suele utilizarse ésta con más frecuencia. A continuación se aplica este método para estimar los parámetros de una distribución von Mises.

La función de densidad von Mises con parámetros k y μ es:

$$f(\theta) = (2\pi I_0(k))^{-1} \exp\{k \cos(\theta - \mu)\}$$

Para estimar los parámetros, con el método de máxima verosimilitud, se considera una muestra aleatoria proveniente de una distribución von Mises y se evalúa la función de densidad conjunta de la muestra.

$$f(\theta_1, \theta_2, \dots, \theta_n; k, \mu) = \prod_{i=1}^n f(\theta_i; k, \mu)$$

Entonces $L(k, \mu) = \prod_{i=1}^n f(\theta_i; k, \mu)$ es la función de verosimilitud.

$$L(k, \mu) = (2\pi I_0(k))^{-n} \exp\{k \cos(\theta_1 - \mu)\} \dots \exp\{k \cos(\theta_n - \mu)\}$$

$$= (2\pi I_0(k))^{-n} \exp\{k[\cos(\theta_1 - \mu) + \dots + \cos(\theta_n - \mu)]\}$$

calculando $\ln L(k, \mu)$

$$\ln L(k, \mu) = -n \ln 2\pi - n \ln I_0(k) + k \sum_{i=1}^n \cos(\theta_i - \mu)$$

Para obtener los estimadores máximo verosímiles se obtiene la derivada de $\ln L$ con respecto a k y μ .

$$\frac{\partial \ln L}{\partial \mu} = k \sum_{i=1}^n \sin(\theta_i - \mu) \dots \dots \dots (1)$$

$$\frac{\partial \ln L}{\partial k} = -n I_0'(k) / I_0(k) + \sum_{i=1}^n \cos(\theta_i - \mu) \dots \dots \dots (2)$$

Iguando a cero la ec. (1)

$$\frac{\partial \ln L}{\partial \mu} = 0 \rightarrow \sum_{i=1}^n \sin(\theta_i - \hat{\mu}) = 0$$

Esta última igualdad se cumple cuando $\hat{\mu} = \bar{x}$ como ya se vió en la sección 1.4.1. del capítulo 1. Por lo tanto el estimador máximo verosímil de el parámetro μ de una distribución von Mises es el ángulo medio muestral \bar{x}_0 .

Para calcular el estimador máximo verosímil del parámetro k , se iguala la ecuación (2) a cero.

$$\frac{\partial \ln L}{\partial k} = 0 \rightarrow I_0'(\hat{k}) / I_0(\hat{k}) = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \hat{\mu})$$

En la sección 1.4.1.c. se vió que

$$\frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \mu) = r$$

Sea la función

$$A(k) = I'_0(k)/I_0(k)$$

Entonces $A(\hat{k}) = \hat{r}$ y por lo tanto $\hat{k} = A^{-1}(\hat{r})$

La solución de esta ecuación es obtenida numericamente.

Algunos valores de la función $A(k)$ y $A^{-1}(k)$ se dan en

la Tabla 1.

Para comprobar si los estimadores encontrados son máximos se aplica el criterio de la segunda derivada. La función $\ln L(\mu, k)$ tiene un máximo en $(\hat{\mu}, \hat{k})$ si cumple las siguientes dos condiciones:

$$I. \left[\frac{\partial^2 \ln L(\hat{\mu}, \hat{k})}{\partial k^2} \right] \left[\frac{\partial^2 \ln L(\hat{\mu}, \hat{k})}{\partial \mu^2} \right] - \left[\frac{\partial^2 \ln L(\hat{\mu}, \hat{k})}{\partial k \partial \mu} \right]^2 > 0$$

$$II. \frac{\partial^2 \ln L(\hat{\mu}, \hat{k})}{\partial \mu^2} < 0$$

Entonces se calculan las segundas derivadas

$$\frac{\partial^2 \ln L}{\partial k^2} = -n \frac{I_0''(k) I_0'(k) - I_0'(k) I_0''(k)}{I_0^2(k)} \Bigg|_{k=\hat{k}} = -n A'(k)$$

$$\frac{\partial^2 \ln L}{\partial k \partial \mu} = \sum_{i=1}^n \sin(\theta_i - \mu) \Bigg|_{\mu=\hat{\mu}} = 0$$

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -k \sum_{i=1}^n \cos(\theta_i - \mu) \Bigg|_{\mu=\hat{\mu}} = -\hat{k} n A(\hat{k}) = -\hat{k} n \hat{r}$$

La segunda condición del criterio de la segunda derivada se cumple puesto que $(-\hat{k} n \hat{r}) < 0$.

Ahora checando la condición I. se tiene que

$$(-n A'(k))(-k n A'(k)) - \theta > \theta$$

Dado que⁽⁴⁾

$$0 < A(k) < 1 \quad \text{y} \quad 0 < A'(k) < 1/2$$

entonces también se cumple la condición 1.; de esta forma se ha comprobado que \hat{k} y $\hat{\mu}$ son los estimadores máximo verosímiles de los parámetros k y μ de una distribución von Mises.

Para el caso de distribuciones multimodales que se basan en la distribución von Mises es posible estimar sus parámetros por el mismo método obteniéndose,

$$\hat{\mu} = \frac{1}{v} \hat{\mu}_v, \quad A(\hat{k}) = r_v$$

Donde v es el número de modas, $\hat{\mu}_v$ el ángulo medio de la muestra de ángulos $\theta^* = v\theta \pmod{2\pi}$ y r_v la longitud del vector medio de la misma muestra θ^* .

3.2.2. Intervalos de confianza.

En la sección 3.2.1. se han calculado estimadores puntuales del parámetro de concentración k y del ángulo medio μ de la distribución von Mises. En la presente sección se ve que al igual que en estadística lineal es posible obtener intervalos de confianza de éstos parámetros.

Estos intervalos se construyen en base a una técnica desarrollada por Stephens (1962), Brown y Mewaldt (1968), la cual fué publicada finalmente por Batschelet (1972).

3.2.2.a. Intervalos de confianza para μ .

Para estimar el intervalo de confianza para μ se cuentan con dos distintos mapas. El primero Fig. A ayuda a determinar los límites del intervalo con un coeficiente de confianza del 95%, el segundo Fig. B está elaborado con un coeficiente de confianza del 99%.

1.- De acuerdo al Apéndice A en J. Watson (1981)

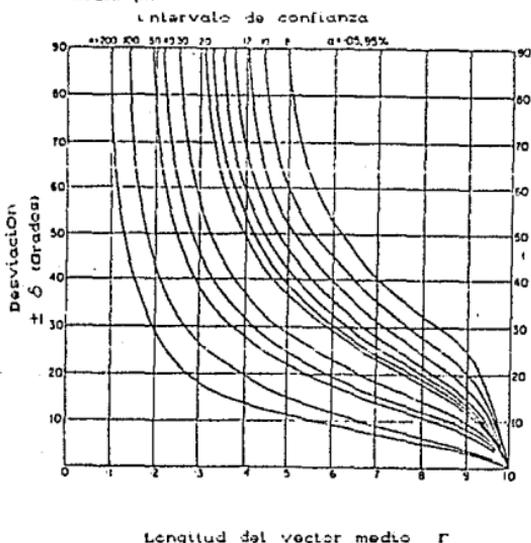
Estos mapas fueron construidos suponiendo que se cuenta con una muestra unimodal con $\mu = 0$. Se grafico para un tamaño de muestra fijo n , considerando un R_α y un S_α con valores fijos, se determina el valor de \bar{C} de la ecuación siguiente:

$$R_\alpha = (S_\alpha^2 + C^2)^{1/2}$$

ó de manera equivalente para un valor fijo $R = R_0$ se obtiene el correspondiente valor de \bar{C} con ayuda de la ecuación vista en la sección 1.4.1. : $\bar{C} = R_0 \cos \delta$, $0 < \delta < \frac{1}{2}\pi$ con la cual, si se mantiene fijo R_0 y se varía C se obtienen distintos valores del ángulo δ , con el cual finalmente se calcula el intervalo de confianza para μ , utilizando el siguiente estimador de los límites del intervalo: $\bar{x}_0 \pm \delta$

La Fig. A y la Fig. B proporcionan los valores correspondientes de δ dados r y n que se calculan en base a la muestra.

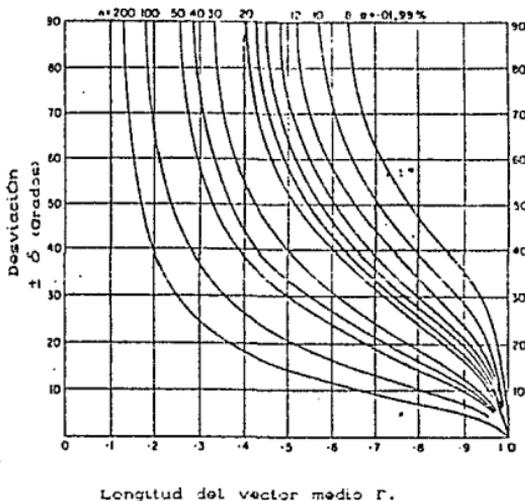
FIG. A. Mapa I para obtener un intervalo de confianza del 95% para la dirección media μ .



I. - Reproducido de Batschelet(1981).

Fig. B. Mapa para obtener un intervalo de confianza del 99% para la dirección media μ .

Intervalo de confianza.



I. - Reproducido de Batschelet(1981).

3.2.2.b. Intervalo de confianza para k.

El intervalo de confianza para k también es calculado en forma numérica y su resultado se presenta en forma de un mapa de gráficas. La Fig. C y la Fig. D proporcionan los límites de confianza (k_1, k_2) tales que:

$$P(k_1 < k < k_2) = 1 - \alpha$$

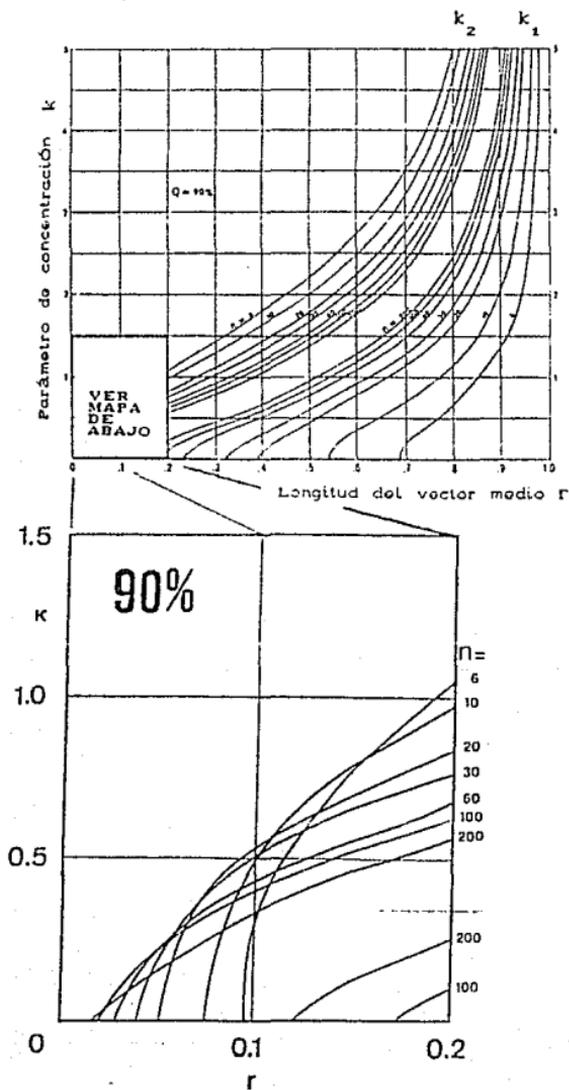
donde $(1-\alpha)$ es el coeficiente de confianza.

Es necesario, calcular la longitud del vector medio r de una muestra aleatoria de tamaño n para consultar los límites del intervalo directamente en la Fig. C ó la Fig. D dependiendo del coeficiente de confianza que se desee considerar.

FIG. C. Mapa I para determinar un intervalo de confianza del 90% del parámetro de concentración k . El tamaño de la muestra es n .

Las curvas k_2 determinan el límite inferior.

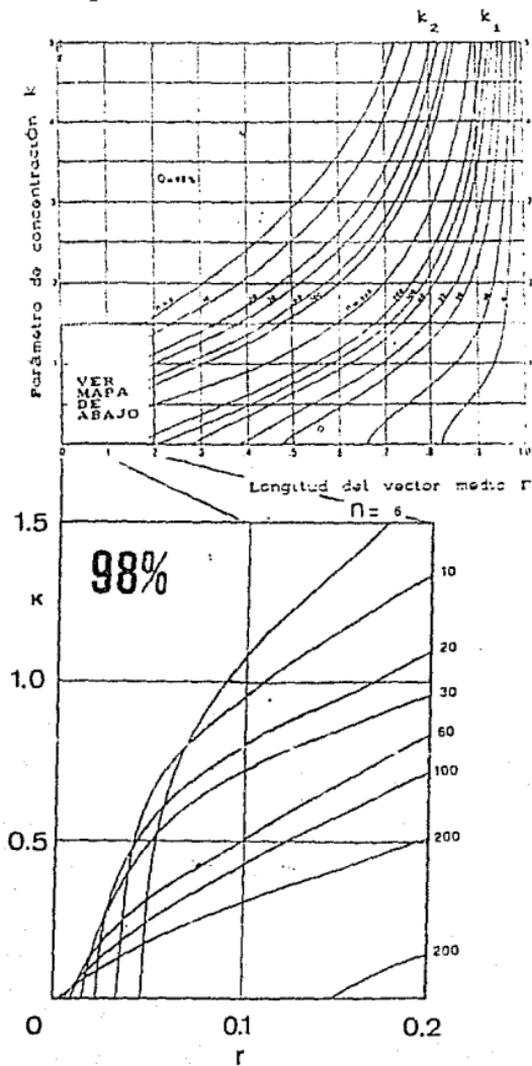
Las curvas k_1 determinan el límite superior.



I. Reproducido de Batschelet(1981).

FIG. D. Mapa I para determinar un intervalo de confianza del 98% del parámetro de concentración k . El tamaño de la muestra es n .

Las curvas k_1 determinan el límite inferior.
 Las curvas k_2 determinan el límite superior.



I. Reproducido de Batschelet(1981)

3.3. PRUEBAS DE HIPOTESIS.

Al igual que en estadística lineal, la experiencia de distintas investigaciones o el desarrollo de la teoría del fenómeno de interés dan lugar a conjeturas o hipótesis. Estas hipótesis es posible, en algunos casos, verificar si se cumplen o no. La teoría clásica de Neyman-Pearson se aplica igualmente a la hipótesis circulares, ya que también se consideran dos hipótesis; la hipótesis de interés o nula y la hipótesis alternativa. Se tienen presentes los dos tipos de errores que se pueden cometer, el error tipo 1, que consiste en rechazar la hipótesis nula cuando esta es cierta y el error tipo 2 que consiste en no rechazar la hipótesis nula siendo que esta es falsa.

Se cuenta, al igual que en estadística lineal, el nivel de significancia α que indica la probabilidad de cometer el error tipo 1.

Dependiendo del valor α y de la hipótesis nula se construye la región crítica o de rechazo.

3.3.1. Pruebas de hipótesis sobre los parámetros μ y k .

Es posible usar los intervalos de confianza para probar hipótesis respecto al parámetro μ , por ejemplo, si se desea probar si el ángulo μ difiere significativamente de una dirección fija dada por el ángulo μ_0 , para realizar la hipótesis se determinan los límites de confianza $\bar{x} \pm z \sigma$, si μ_0 cae fuera del intervalo se concluye que el ángulo μ difiere del ángulo μ_0 , es decir que la prueba fue significativa, en caso contrario se concluye que la prueba no fue significativa.

Del mismo modo, el intervalo de confianza para el parámetro k también puede ser usado para realizar pruebas de hipótesis. Si por ejemplo, una hipótesis nula establece que $k > 2$ se chequea el límite superior de confianza de k ,

si este limite es menor que 2 se rechaza la hipótesis nula; en caso de ocurrir lo contrario se concluye que no hay evidencia suficiente para probar que la hipótesis nula es falsa.

3.3.2. Pruebas de uniformidad.

En el análisis de datos direccionales las pruebas de hipótesis más frecuentemente usadas son las pruebas de uniformidad, es decir, aquellas en que la hipótesis nula establece que la población se distribuye uniformemente alrededor del círculo unitario.

Se presentan dos pruebas de uniformidad: la prueba Rayleigh y la prueba Hodges-Ajne.

3.3.2. La prueba Rayleigh.

Para aplicar esta prueba es necesario tener una muestra aleatoria $\theta_1, \theta_2, \dots, \theta_n$ de datos direccionales. Se pretende probar si la población de la cual fue obtenida la muestra presenta tendencia a seguir una dirección preferida es decir si difiere significativamente de tener una distribución uniforme.

Entonces se puede escribir, (considerando fija $\mu = 0$):

$$H_0 : \nu = 0$$

$$H_1 : k \neq 0$$

Se utiliza el Lema fundamental de Neyman-Fearson para encontrar la estadística de prueba con la correspondiente región de rechazo.

Bajo H_0 la función de verosimilitud es:

$$L_0 = 1/(2\pi)^n$$

Bajo H_1 se niega el hecho de que se tiene una distribución uniforme y se propone como función alternativa a la distribución von Mises. Entonces la función de verosimilitud sobre Ω es:

$$L_1 = \{2\pi I_0(\hat{k})\}^{-n} \exp\{n \hat{k} \hat{r}\}$$

Donde $\hat{r} = A(\hat{k})$, $r^2 = \bar{C}^2 + \bar{S}^2$

La región crítica es $\lambda < K$,

Con

$$\lambda = \frac{(2\pi)^{-n}}{(2\pi I_0(\hat{k}))^{-n} \exp(n \hat{k} \hat{r})}$$

Reduciendo esta expresión

$$\lambda = (I_0(\hat{k}))^n \exp(-n \hat{k} \hat{r})$$

Se analiza el comportamiento de éste cociente para obtener una región de rechazo más sencilla. Se calcula:

$$\ln \lambda = n \ln I_0(\hat{k}) - (n \hat{k} A(\hat{k}))$$

Se deriva con respecto a \hat{k}

$$\frac{d(\ln \lambda)}{d\hat{k}} = n \left[\frac{I_0'(\hat{k})}{I_0(\hat{k})} - \hat{k} A'(\hat{k}) - A(\hat{k}) \right]$$

Usando el hecho de que $I_0'(\hat{k}) = I_1(\hat{k})$

$$\frac{d(\ln \lambda)}{d\hat{k}} = -n \hat{k} A'(\hat{k})$$

Entonces $\frac{d(\ln \lambda)}{d\hat{k}}$ es negativo dado que $k \geq 0$ y como

ya se mencionó en la sección 3.2.1 $A'(\hat{k}) \geq 0$.

Así que, se puede afirmar que λ es una función decreciente de \hat{k} , mientras que $A'(\hat{k})$ es una función monótona creciente de \hat{r} y la región crítica $\lambda < K$ se reduce a $\hat{r} > K$.

Donde r es la longitud del vector medio calculado en base a la muestra. Los valores de K se pueden consultar en la Tabla 2.

Para n grande se usa el estadístico de prueba $3nr^2$ que tiene una distribución Ji-cuadrada con tres grados de libertad.

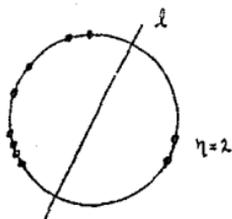
Intuitivamente esta prueba está basada en la longitud del vector medio \bar{r} , basándose en el hecho de que las muestras provenientes de poblaciones cuya distribución es uniforme presentan valores de \bar{r} cercanos a cero.

3.3.2. b. Prueba Hodges-Ajne.

Se tiene una muestra aleatoria de observaciones direccionales $\theta_1, \theta_2, \dots, \theta_n$ y se desea probar si la población de la cual proviene tiene una distribución uniforme.

Para aplicar esta prueba se grafican las n observaciones de la muestra alrededor del círculo unitario. Se dibuja uno de los diámetros del círculo, al cual se la llama l . El diámetro l divide al círculo en dos arcos iguales; entonces se rota este diámetro hasta tener el mínimo de observaciones en uno de los dos arcos del círculo. Se denota por η al número mínimo de observaciones.

Fig.



Si el valor de η es muy pequeño con respecto al número total de observaciones, entonces se dice que la muestra presenta una dirección preferida sobre las demás.

Sea θ_i una observación de la muestra, se realiza la transformación,

$$\theta_i^* = \theta_i \pmod{\pi} \quad 0 < \theta \leq 2\pi, \quad i = 1, 2, \dots, n$$

Se define $\theta_{(i)}^*$ $i = 1, 2, \dots, n$ como los correspondientes valores ordenados de θ_i^* .

Sea z_i definido de la siguiente manera,

$$z_i = \begin{cases} 1 & \text{si } \theta_{(i)}^* = \theta_j & \text{para alguna } j \\ -1 & \text{si } \theta_{(i)}^* = \theta_j - \pi & \text{para alguna } j \end{cases}$$

Obteniendo entonces el vector $Z = (z_1, z_2, \dots, z_n)$ con ayuda del cual es posible saber cuales observaciones caen en cada arco, dependiendo de si se trata de un 1 o un -1.

Rotando el diámetro l se producen $2n$ valores de Z . Es posible obtener los distintos valores del vector Z producidos con cada rotación si se aplica el operador $\xi(Z)$ $2n$ veces.

$$\xi(Z) = (z_2, z_3, \dots, z_n, -z_1)$$

Sea K_r , $r = 1, 2, \dots, 2n$ el número de veces que apareció el número -1 dentro del vector Z al aplicar por r -ésima vez el operador $\xi(Z)$, entonces

$$\eta = \min \{K_1, K_2, \dots, K_{2n}\}, \quad 0 \leq \eta \leq \frac{1}{2}n$$

K_r cumple las siguientes propiedades:

$$\begin{aligned} K_{i+1} &= K_i + z_i \\ K_i &= n - K_{n-i} \quad , \quad i = 1, 2, \dots, n \end{aligned}$$

Si se calcula $S_i = z_1 + z_2 + \dots + z_i$

$$\begin{aligned} K_i &= \frac{1}{2} (n - S_i) \\ K_{i+1} &= S_i + K_i \quad , \quad i = 1, 2, \dots, n-1 \\ K_{n+1} &= n - K_1 = K_1 + S_n \end{aligned}$$

Luego entonces es posible expresar a η de la siguiente manera:

$$\eta = \frac{1}{2}n - \max \left\{ -\frac{1}{2}S_n + \max_{0 \leq i \leq n} S_i, \frac{1}{2}S_n - \min_{0 \leq i \leq n} S_i \right\}$$

Que tiene la siguiente función de distribución⁽¹⁾

$$P(\eta \leq t) = 2^{-n+1} (n - 2t) \sum_{k=0}^r V(t, -k)$$

1. - De acuerdo a Mardia(1972)

donde

$$V(t, -k) = \binom{n}{t-k(n-2t)}$$

con $0 \leq \eta \leq \frac{n}{2}$, $r = t/(n-2t)$

Para $t < \frac{n}{2}$ se tiene que $(n-2t) > \frac{n}{2}$ y $r < 1$. Si este es el caso la función de distribución de η se reduce⁽¹⁾ a.

$$P(\eta \leq t) = 2^{-n+1} (n-2t) \binom{n}{t} , \quad t < \frac{n}{2}$$

La hipótesis de uniformidad se rechaza para valores pequeños de η . Se puede escribir la región de rechazo \mathcal{C} como

$$\mathcal{C} = \left\{ \theta \mid \eta \leq \eta_0 \right\}$$

donde el valor de η_0 se puede consultar en la Tabla 3.

3.3.2.c. Prueba Hodges-Ajne para datos agrupados.

En el caso de datos agrupados se pueden considerar las frecuencias como si estuvieran concentradas en los puntos medios de cada intervalo. Si $N(\theta)$ es el número total de frecuencias en el arco $(\theta, \pi+\theta)$ se tiene que

$$\eta = \min_{0 \leq \theta \leq \pi} N(\theta)$$

Si se tienen r intervalos de clase de igual longitud con frecuencia f_i ($i = 1, 2, \dots, r$) donde $f_{i+r} = f_i$ entonces si se calcula

$$A_j = \sum_{i=1}^k f_{i+j} \quad j = 0, 1, \dots, r$$

se tiene que $\eta = \min_j (A_j)$ y $A_j + A_{j+k} = n$

Se define $B = \max_j |A_j - A_{j+k}|$, entonces,

1. - De acuerdo a Mardia(1972)

$$B = \max_j |A_j - (n - A_j)| = \max_j \{n - 2A_j, 2A_j - n\}$$

Como $\eta = \min_j A_j$ y $(n - \eta) = \max_j A_j$ entonces es

posible reescribir B de la manera siguiente:

$$B = \max \{n - 2(n - \eta), 2(n - \eta) - n\} = \max \{-n + 2\eta, n - 2\eta\}$$

entonces

$$B = n - 2\eta$$

El valor de B es importante debido a que forma parte de la estadística de prueba en el caso de datos agrupados, dada por:

$$\eta = B/\sqrt{n}$$

Bajo la hipótesis de uniformidad B probablemente tomara un valor pequeño en cuyo caso la región de rechazo es

$$\mathcal{R} = \{ \theta \mid \eta \geq \eta_0 \}$$

Solo se cuenta con dos valores de η_0 , los cuales fueron calculados por David y Newell (1965).

Para	$\alpha = 0.05$	$\eta_0 = 2.53$
	$\alpha = 0.01$	$\eta_0 = 3.09$

3.3.3. Pruebas de bondad de ajuste.

Los modelos teóricos desarrollados para describir a las poblaciones direccionales se intenta que sean la más simple posible, mientras que los fenómenos aleatorios que se estudian son en general muy complejos debido a la gran variedad de factores que influyen en su acurrencia. Al ajustar un modelo a una muestra no se puede esperar que la distribución teórica coincida exactamente con la "verdadera" distribución del fenómeno; sin embargo debe ajustarse una función que en general, no contradiga el comportamiento de la distribución presentada por la muestra.

En esta sección se analizan algunas pruebas de bondad de

ajuste empleadas en el análisis de datos direccionales.

3.3.3.a. Prueba Ji-cuadrada.

Esta prueba se aplica en estadística lineal y se ha aplicado al análisis de datos direccionales con iguales resultados.

Dada una muestra aleatoria de observaciones direccionales $\theta_1, \theta_2, \dots, \theta_n$ se desea probar si la distribución de la muestra puede ser descrita por una distribución teórica dada.

Para este propósito se agrupan los datos en intervalos o grupos no necesariamente de igual longitud, con la única restricción de conservar $f_i \geq 4$, donde f_i es la frecuencia del intervalo i .

La hipótesis nula y la alternativa son:

H_0 : La distribución poblacional coincide con la distribución teórica propuesta.

H_1 : La distribución poblacional no coincide con la distribución teórica propuesta.

La estadística de prueba es la misma que se utiliza en estadística lineal:

$$T = \sum_{i=1}^k (f_i - e_i)^2 / e_i$$

donde

k es el número de grupos

f_i es la frecuencia de la muestra en el intervalo i

e_i es la frecuencia esperada dentro de el intervalo i (la cual se calcula en base a la distribución teórica dada).

La distribución de T es Ji-cuadrada con $(k-1)$ grados de

libertad. Si algun parámetro de la distribución debe ser estimado se resta un grado de libertad a la distribución, al igual que ocurre en estadística lineal.

La región crítica es :

$$\mathcal{C} = \left\{ \theta \mid T > \chi^2_{(k-1)}^{1-\alpha} \right\}$$

Valores de la distribución Ji-cuadrada se dan en la Tabla 4.

Esta prueba también puede ser usada para probar uniformidad; si se considera que la distribución de la muestra es uniforme.

3.3.3.b. Prueba de Kuiper.

Esta prueba fue concebida por Kolmogorov y es frecuentemente usada en estadística lineal. Kuiper en 1960 adaptó la prueba a las condiciones especiales de las distribuciones circulares.

Al igual que en estadística lineal se define a la función de distribución empírica :

$$S_n(\theta) = \frac{i}{n} \quad \theta_{(i)} \leq \theta \leq \theta_{(i+1)}, \quad i=0,1,\dots,n$$

$$\text{con } \theta_{(0)} = 0 \quad \text{y} \quad \theta_{(n+1)} = 2\pi$$

Los resultados de la función de distribución empírica en la línea son igualmente válidos para la función de distribución empírica circular.

Las hipótesis nula y alternativa son :

$$H_0: F(\underline{\theta}) = F_0(\underline{\theta}) \quad \text{vs.} \quad H_1: F(\underline{\theta}) \neq F_0(\underline{\theta})$$

Para una función de distribución $F_0(\underline{\theta})$ dada.

Al igual que en el caso lineal se definen los valores D_n^+ y D_n^- por:

$$D_n^+ = \sup_{\theta} \{ S_n(\theta) - F_0(\theta) \}$$

$$D_n^- = \sup_{\theta} \{ F_0(\theta) - S_n(\theta) \}$$

Tanto D_n^+ como D_n^- dependen de la dirección cero que se escoja. Kuiper en 1960 definió, en base a éstas una nueva estadística de prueba:

$$V_n = D_n^+ + D_n^-$$

la cual es independiente de la dirección cero que se escoja.

Mardia en 1972 considera una representación alternativa de V_n , en la cual se tiene que

$$D_n^+ = \max_{0 \leq i \leq n} \sup_{\substack{\theta_{(i)} \leq \theta < \theta_{(i+1)}}} \left\{ \frac{i}{n} - F_0(\theta) \right\}$$

$$D_n^- = \max_{0 \leq i \leq n} \left\{ \frac{i}{n} - \inf_{\substack{\theta_{(i)} \leq \theta < \theta_{(i+1)}}} F_0(\theta) \right\}$$

$$\text{entonces } D_n^+ = \max_{0 \leq i \leq n} \left\{ \frac{i}{n} - U_i \right\}$$

donde $U_0 = 0$, $U_i = F_0(\theta_{(i)})$ $i = 1, 2, \dots, n$.

Con un razonamiento igual es posible obtener D_n^-

$$D_n^- = \max_{0 \leq i \leq n} \left\{ U_i - \frac{(i+1)}{n} \right\}$$

De esta forma se tiene que :

$$V_n = \max_{0 \leq i \leq n} \left\{ \frac{i}{n} - U_i \right\} + \max_{0 \leq i \leq n} \left\{ U_i - \frac{(i+1)}{n} \right\}$$

Kuiper en 1960 calculó la función de distribución de la siguiente estadística de prueba

$$T = n^{1/2} V_n = n^{1/2} [D_n^+ + D_n^-]$$

Obteniendo la siguiente distribución asintótica:

$$\Pr [T \leq j] = 1 - 2 \sum_{m=1}^{\infty} [4m^2 j^2 - 1] + \left[\frac{\theta z}{3\gamma_1} \right] \sum_{m=1}^{\infty} m^2 [4m^2 j^2 - 3] e^{-2m \frac{z}{j} z} + O(n^{-1})$$

con ayuda de la cual calculó los valores críticos $K(\alpha)$ que se utilizan en la región de rechazo

$$\mathcal{R} = \left\{ \theta \mid T > K(\alpha) \right\}$$

los valores de $K(\alpha)$ se pueden consultar en la Tabla 5.

Finalmente, la prueba de Kuiper al igual que la prueba Ji-cuadrada puede ser usada para probar uniformidad considerando

$$F_0(\theta) = 1/2\pi$$

Las pruebas presentadas en esta sección son las más comunmente usadas en el análisis de datos direccionales, sin embargo se han desarrollado diversas pruebas estadísticas para probar bondad de ajuste ó simplemente uniformidad; cada una de las cuales ofrece, ya sea, mayor rapidez, facilidad de aplicación ó mayor precisión. Siendo el objetivo del presente trabajo dar una visión general del análisis de datos direccionales solo se han incluida algunas de las más importantes.

CAPITULO 4

EJEMPLO

4.1. Introducción.

En este capítulo se presentan un estudio en el que se aplica la metodología presentada en los tres primeros capítulos. Como se menciona en la sección 1.2.1. del capítulo 1, el análisis de datos direccionales puede ser aplicado cuando se analizan fenómenos que se desarrollan al transcurrir el tiempo y que presentan una cierta periodicidad conocida. Este es el caso del estudio que se presenta a continuación.

4.2. Análisis de la precipitación pluvial.

El estudio que se presenta a continuación es un análisis del comportamiento de la precipitación pluvial a través del año. Este se realiza en base a mediciones en ml. de la precipitación pluvial a través de los días de 44 años, de 1931 a 1974. Estas mediciones se encuentran concentradas en grupos de 5 días formando un total de 73 grupos por año. En el caso de años bisiestos el último grupo es de seis días.¹⁾

El objetivo de este estudio es explicar el comportamiento del fenómeno aleatorio con ayuda del análisis de datos direccionales. Para tal efecto se considera cada ml. como una ocurrencia del fenómeno.

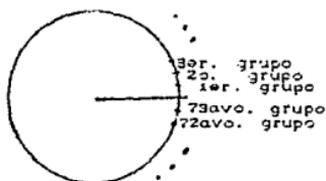
1.- De esta forma se encuentran organizados los datos en la tesis de Pérez (1978) de la cual fueron tomados.

4.2.1. Planteamiento estadístico.

La precipitación pluvial es considerada un fenómeno aleatorio que se desarrolla en el tiempo cuya periodicidad es anual

Con el propósito de estudiar este fenómeno con ayuda del análisis de datos direccionales se ha asociado el periodo, de un año, a una vuelta completa del círculo. Dado que se tienen divididos los datos de cada año en 73 grupos (cada uno de los cuales representan las observaciones de 5 días) entonces se ha dividido el círculo en 73 arcos, a cada uno de los cuales se asocia la información de los 73 grupos de un año.

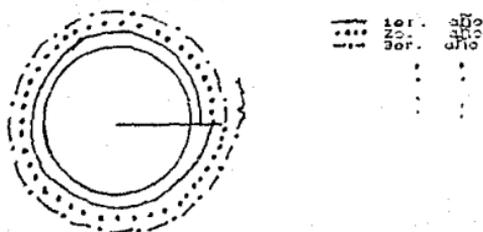
Fig.



De esta forma se cuenta con una muestra de 73 observaciones $\theta_1, \theta_2, \dots, \theta_{73}$ determinada por los ángulos medios de cada arco.

La frecuencia asociada a cada arco o intervalo de clase se calcula de la siguiente manera: Dado que se cuenta con las mediciones de 44 años, se procede a envolver las mediciones de cada año en el círculo.

Fig.



El proceso de envoltura se elabora de forma tal que la

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

frecuencia del primer arco corresponde a la suma de las frecuencias observadas en los primeros cinco días de cada uno de los 44 años. El segundo arco tiene asociado como frecuencia a la suma de las frecuencias observadas en los siguientes cinco días de cada uno de los 44 años. Y de igual forma se obtiene la frecuencia de cada uno de los arcos restantes. Cada ml. medido de precipitación pluvial se considera una ocurrencia por esta razón se puede llegar a tener en algún grupo una frecuencia por ejemplo de 0.8 ml., es decir, no necesariamente se trata de un número entero.

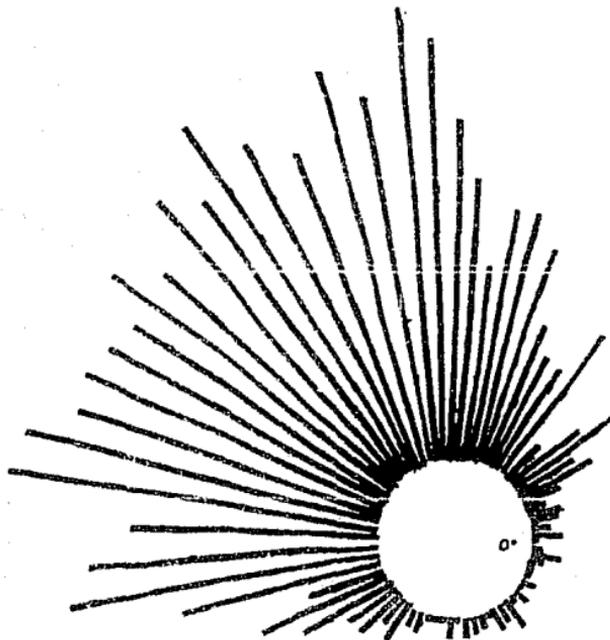
La tabla 4.1.a. presenta el ángulo medio de los 73 intervalos con la correspondiente frecuencia asociada a cada intervalo, una vez que se ha realizado el proceso de envoltura.

TABLA 4.1.a.

Ángulo Medio	F	F x COS(θ)	F x SEN(θ)
2.47	15.00	14.1001	2.4852
7.40	69.10	67.2522	9.7676
12.34	42.10	41.1291	8.4072
17.28	89.20	85.2183	15.4251
22.17	75.20	69.7072	28.2170
27.12	67.10	61.6242	24.7425
32.05	67.00	62.7200	27.9793
36.99	109.40	105.4606	48.6476
41.77	79.20	77.1125	34.8677
46.65	79.20	76.3240	34.0255
51.70	139.50	132.4494	68.9771
56.71	150.100	143.6264	85.2760
61.68	230.00	195.0293	139.2468
66.56	75.00	70.7272	32.9886
71.51	449.50	412.067	215.8622
76.44	746.10	687.6625	401.3625
81.27	510.00	477.2287	212.2272
86.10	580.00	537.0000	247.9672
91.15	50.20	47.2116	20.1258
96.16	451.70	425.5041	241.9802
101.19	85.10	81.2547	41.3009
106.02	777.00	718.2271	384.2376
110.90	625.20	582.4761	304.1691
115.64	1941.20	1824.8044	1026.2824
120.60	841.20	781.0960	392.2522
125.75	624.20	588.1027	296.4670
130.60	779.20	736.9892	368.6720
135.62	637.10	596.7821	295.2160
140.55	440.20	412.2345	204.7075
145.48	781.20	734.6677	368.2596
150.41	840.20	770.0025	411.9817
155.34	721.70	653.6257	341.0886
160.10	823.10	774.7903	374.2121
165.11	714.20	680.6194	307.9099
170.16	736.00	707.7441	326.1245
175.17	755.00	712.1918	331.8907
180.10	742.10	702.1036	327.1166
185.06	840.00	781.0000	371.2166
190.00	822.10	768.4013	364.1512
195.00	770.20	736.0000	341.3001
200.00	670.20	634.0021	296.2060
205.00	420.00	382.0037	194.5007
210.00	177.30	142.7059	74.1074
215.00	105.00	82.7971	44.1061
220.00	237.00	165.0612	142.1284
225.00	12.20	11.771	1.06730
230.00	10.20	9.81241	1.07425
235.10	8.20	7.74443	0.84066
240.11	181.10	168.1094	84.11820
245.04	67.00	62.9255	28.1052
250.10	46.10	42.9264	18.1172
255.00	16.00	14.1001	2.4852
260.04	16.00	14.1001	2.4852
265.00	11.30	10.7940	1.1769
270.00	9.70	9.2061	0.9720
275.00	11.30	10.7940	1.1769
280.00	60.20	57.0067	27.9872
285.00	67.20	63.7027	30.1093
290.00	20.10	19.2664	8.9106
295.00	15.00	14.1001	2.4852
300.00	20.00	19.2664	8.9106
305.00	10.00	9.6332	4.4553
310.00	5.00	4.8166	2.2277
315.00	2.00	1.9284	0.9640
320.00	1.00	0.9642	0.4820
325.00	1.00	0.9642	0.4820
330.00	1.00	0.9642	0.4820
335.00	1.00	0.9642	0.4820
340.00	1.00	0.9642	0.4820
345.00	1.00	0.9642	0.4820
350.00	1.00	0.9642	0.4820
355.00	1.00	0.9642	0.4820
360.00	1.00	0.9642	0.4820

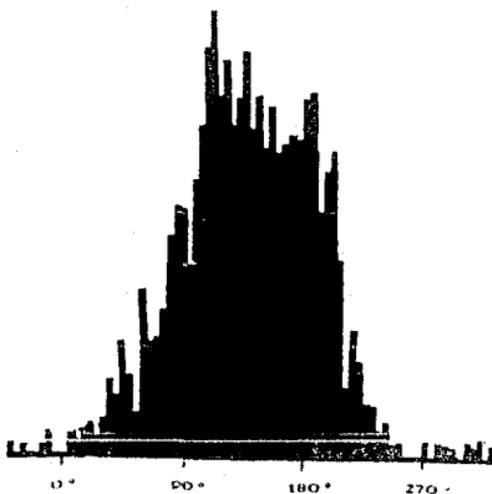
4.1.3. Análisis estadístico de los datos.

Una vez que se cuenta con una muestra de datos direccionales se puede comenzar el análisis estadístico graficando las observaciones. La gráfica 1 presenta el Histograma Circular de los datos.



Este se grafica considerando el ángulo 0° como el punto donde el año empieza, se avanza en sentido contrario de las manecillas del reloj hasta llegar al punto determinado por el ángulo 360° , en el cual se considera que el año termina.

La gráfica 2 presenta el Histograma Lineal correspondiente al anterior Histograma Circular.



El punto de corte del Histograma circular se eligió en el ángulo 115.50° el cual corresponde al límite inferior del intervalo de clase No. 65, la frecuencia de clase de este intervalo se repite al final del Histograma Lineal.

En ambas gráficas es posible apreciar que las observaciones presentan una distribución unimodal aproximadamente simétrica. La clase modal es la correspondiente al intervalo de clase No. 24 cuyos límites son

$$(113.25 - 30^\circ, 118.21 - 18^\circ)$$

que en término de días se trata del periodo entre el día 115 del año y el día No. 120 del año (estos corresponden a los últimos cinco días del mes de abril). La frecuencia del intervalo es de 1041.30 ml.

Para encontrar la mediana direccional de los datos se elaboró la tabla 4.1.b. En la primera columna se muestra el límite superior del intervalo de clase, la segunda columna muestra la frecuencia observada en el intervalo, la tercera

columna se muestran las frecuencias acumuladas. La primera mitad de esta columna se repite en la columna 4. Para obtener la columna 5, se restan las entradas de la columna 4 de las de la columna 3. Así la columna 5 muestra la frecuencia de los intervalos $(\theta_{i-180^\circ}, \theta_i)$ cuyos límites se muestran en la columna 6.

TABLA 4.1.b

LÍMITE SUPERIOR θ_i	FRECUENCIA	FRECUENCIA ACUMULADA	FRECUENCIA $(\theta_{i-180^\circ}, \theta_i)$	FRECUENCIA INTERVALO $(100-\theta_{i-1}, \theta_i)$	INTERVALO $(100-\theta_{i-1}, \theta_i)$
1.47	7.50	7.50	7.50		
4.03	7.50	15.00	7.50		
7.40	15.00	30.00	15.00		
9.86	22.50	52.50	22.50		
12.33	30.00	82.50	30.00		
14.79	37.50	120.00	37.50		
17.26	45.00	165.00	45.00		
19.72	52.50	217.50	52.50		
22.19	60.00	277.50	60.00		
24.66	67.50	345.00	67.50		
27.12	75.00	420.00	75.00		
29.59	82.50	502.50	82.50		
32.06	90.00	592.50	90.00		
34.52	97.50	690.00	97.50		
36.99	105.00	795.00	105.00		
39.46	112.50	907.50	112.50		
41.92	120.00	1027.50	120.00		
44.39	127.50	1155.00	127.50		
46.85	135.00	1290.00	135.00		
49.32	142.50	1432.50	142.50		
51.79	150.00	1582.50	150.00		
54.25	157.50	1740.00	157.50		
56.71	165.00	1905.00	165.00		
59.18	172.50	2077.50	172.50		
61.64	180.00	2257.50	180.00		
64.11	187.50	2445.00	187.50		
66.58	195.00	2640.00	195.00		
69.04	202.50	2842.50	202.50		
71.51	210.00	3052.50	210.00		
73.98	217.50	3270.00	217.50		
76.44	225.00	3495.00	225.00		
78.91	232.50	3727.50	232.50		
81.37	240.00	3967.50	240.00		
83.84	247.50	4215.00	247.50		
86.30	255.00	4470.00	255.00		
88.77	262.50	4732.50	262.50		
91.23	270.00	5002.50	270.00		
93.70	277.50	5280.00	277.50		
96.16	285.00	5565.00	285.00		
98.63	292.50	5857.50	292.50		
101.10	300.00	6157.50	300.00		
103.56	307.50	6465.00	307.50		
106.03	315.00	6780.00	315.00		
108.49	322.50	7102.50	322.50		
110.96	330.00	7432.50	330.00		
113.42	337.50	7770.00	337.50		
115.89	345.00	8115.00	345.00		
118.35	352.50	8467.50	352.50		
120.82	360.00	8827.50	360.00		
123.28	367.50	9195.00	367.50		
125.75	375.00	9570.00	375.00		
128.21	382.50	9952.50	382.50		
130.68	390.00	10342.50	390.00		
133.14	397.50	10740.00	397.50		
135.61	405.00	11145.00	405.00		
138.07	412.50	11557.50	412.50		
140.54	420.00	11977.50	420.00		
143.00	427.50	12405.00	427.50		
145.47	435.00	12840.00	435.00		
147.93	442.50	13282.50	442.50		
150.41	450.00	13732.50	450.00		
152.87	457.50	14190.00	457.50		
155.34	465.00	14655.00	465.00		
157.80	472.50	15127.50	472.50		
160.27	480.00	15607.50	480.00		
162.74	487.50	16095.00	487.50		
165.20	495.00	16590.00	495.00		
167.67	502.50	17092.50	502.50		
170.14	510.00	17602.50	510.00		
172.60	517.50	18120.00	517.50		
175.07	525.00	18645.00	525.00		
177.53	532.50	19177.50	532.50		
180.00	540.00	19717.50	540.00		
182.46	547.50	20265.00	547.50		
184.93	555.00	20819.50	555.00		
187.39	562.50	21381.50	562.50		
189.86	570.00	21951.50	570.00		
192.32	577.50	22529.00	577.50		
194.79	585.00	23113.50	585.00		
197.25	592.50	23706.00	592.50		
199.72	600.00	24306.00	600.00		
202.18	607.50	24913.50	607.50		
204.65	615.00	25528.50	615.00		
207.11	622.50	26151.00	622.50		
209.58	630.00	26781.00	630.00		
212.04	637.50	27418.50	637.50		
214.51	645.00	28063.50	645.00		
216.97	652.50	28716.00	652.50		
219.44	660.00	29376.00	660.00		
221.90	667.50	30043.50	667.50		
224.37	675.00	30718.50	675.00		
226.83	682.50	31401.00	682.50		
229.30	690.00	32091.00	690.00		
231.76	697.50	32788.50	697.50		
234.23	705.00	33493.50	705.00		
236.69	712.50	34206.00	712.50		
239.16	720.00	34926.00	720.00		
241.62	727.50	35653.50	727.50		
244.09	735.00	36388.50	735.00		
246.55	742.50	37131.00	742.50		
249.02	750.00	37881.00	750.00		
251.48	757.50	38638.50	757.50		
253.95	765.00	39403.50	765.00		
256.41	772.50	40176.00	772.50		
258.88	780.00	40956.00	780.00		
261.34	787.50	41743.50	787.50		
263.81	795.00	42538.50	795.00		
266.27	802.50	43341.00	802.50		
268.74	810.00	44151.00	810.00		
271.20	817.50	44968.50	817.50		
273.67	825.00	45793.50	825.00		
276.13	832.50	46626.00	832.50		
278.60	840.00	47466.00	840.00		
281.06	847.50	48313.50	847.50		
283.53	855.00	49168.50	855.00		
285.99	862.50	50031.00	862.50		
288.46	870.00	50901.00	870.00		
290.92	877.50	51778.50	877.50		
293.39	885.00	52663.50	885.00		
295.85	892.50	53556.00	892.50		
298.32	900.00	54456.00	900.00		
300.78	907.50	55363.50	907.50		
303.25	915.00	56278.50	915.00		
305.71	922.50	57201.00	922.50		
308.18	930.00	58131.00	930.00		
310.64	937.50	59068.50	937.50		
313.11	945.00	60013.50	945.00		
315.57	952.50	60966.00	952.50		
318.04	960.00	61926.00	960.00		
320.50	967.50	62893.50	967.50		
322.97	975.00	63868.50	975.00		
325.43	982.50	64851.00	982.50		
327.90	990.00	65841.00	990.00		
330.36	997.50	66838.50	997.50		
332.83	1005.00	67843.50	1005.00		
335.29	1012.50	68856.00	1012.50		
337.76	1020.00	69876.00	1020.00		
340.22	1027.50	70903.50	1027.50		
342.69	1035.00	71938.50	1035.00		
345.15	1042.50	72981.00	1042.50		
347.62	1050.00	74031.00	1050.00		
350.08	1057.50	75088.50	1057.50		
352.55	1065.00	76153.50	1065.00		
355.01	1072.50	77226.00	1072.50		
357.48	1080.00	78306.00	1080.00		
360.00	1087.50	79393.50	1087.50		
362.56	1095.00	80488.50	1095.00		
365.12	1102.50	81591.00	1102.50		
367.68	1110.00	82701.00	1110.00		
370.24	1117.50	83818.50	1117.50		
372.80	1125.00	84943.50	1125.00		
375.36	1132.50	86076.00	1132.50		
377.92	1140.00	87216.00	1140.00		
380.48	1147.50	88363.50	1147.50		
383.04	1155.00	89518.50	1155.00		
385.60	1162.50	90681.00	1162.50		
388.16	1170.00	91851.00	1170.00		
390.72	1177.50	93028.50	1177.50		
393.28	1185.00	94213.50	1185.00		
395.84	1192.50	95406.00	1192.50		
398.40	1200.00	96606.00	1200.00		
400.96	1207.50	97813.50	1207.50		
403.52	1215.00	99028.50	1215.00		
406.08	1222.50	100251.00	1222.50		
408.64	1230.00	101481.00	1230.00		
411.20	1237.50	102718.50	1237.50		
413.76	1245.00	103963.50	1245.00		
416.32	1252.50	105216.00	1252.50		
418.88	1260.00	106476.00	1260.00		
421.44	1267.50	107743.50	1267.50		
424.00	1275.00	109018.50	1275.00		
426.56	1282.50	110301.00	1282.50		
429.12	1290.00	111591.00	1290.00		
431.68	1297.50	112888.50	1297.50		
434.24	1305.00	114193.50	1305.00		
436.80	1312.50	115506.00	1312.50		
439.36	1320.00	116826.00	1320.00		
441.92	1327.50	118153.50	1327.50		
444.48	1335.00	119488.50	1335.00		
447.04	1342.50	120831.00	1342.50		
449.60	1350.00	122181.00	1350.00		
452.16	1357.50	123538.50	1357.50		
454.72	1365.00	124903.50	1365.00		
457.28	1372.50	126276.00	1372.50		
459.84	1380.00	127656.00	1380.00		
462.40	1387.50	129043.50	1387.50		
464.96	1395.00	130438.50	1395.00		
467.52	1402.50	131841.00	1402.50		
470.08	1410.00	133251.00	1410.00		
472.64	1417.50	134668.50	1417.50		
475.20	1425.00	136093.50	1425.00		
477.76	1432.50	137526.00	1432.50		
480.32	1440.00	138966.00	1440.00		
482.88	1447.50	140413.50	1447.50		
485.44	1455.00	141868.50	1455.00		
488.00	1462.50	143331.00	1462.50		
490.56	1470.00	144801.00	1470.00		
493.12	1477.50	146278.50	1477.50		
495.68	1485.00	147763.50	1485.00		
498.24	1492.50	149256.00	1492.50		
500.80	1500.00	150756.00	1500.00		
503.36	1507.50	152263.50	1507.50		
505.92	1515.00	153778.50	1515.00		
508.48	1522.50	155			

TABLA 4.1.b. (continuación)

LÍMITE SUPERIOR	FRECUENCIA ACUMULADA	FRECUENCIA (F _i)	FRECUENCIA (F _i)/N	LÍMITE INFERIOR
241.84	16.15	27055.64	155.74	2258.70
244.11	50.64	33504.66	175.53	2259.50
246.38	100.60	27337.66	167.50	2260.30
248.66	211.04	22706.86	131.70	2261.10
250.91	342.08	22140.66	125.55	2261.85
253.17	484.55	21605.15	120.00	2262.65
255.44	640.55	21102.15	115.00	2263.40
257.69	802.10	20625.60	110.00	2264.15
259.95	971.15	20175.15	105.00	2264.90
262.21	1148.70	19750.15	100.00	2265.65
264.47	1335.15	19350.15	95.00	2266.40
266.73	1530.15	18975.15	90.00	2267.15
268.99	1734.15	18625.15	85.00	2267.90
271.25	1947.15	18300.15	80.00	2268.65
273.51	2169.15	18000.15	75.00	2269.40
275.77	2400.15	17725.15	70.00	2270.15
278.03	2640.15	17475.15	65.00	2270.90
280.29	2890.15	17250.15	60.00	2271.65
282.55	3150.15	17050.15	55.00	2272.40
284.81	3420.15	16875.15	50.00	2273.15
287.07	3700.15	16725.15	45.00	2273.90
289.33	3990.15	16600.15	40.00	2274.65
291.59	4290.15	16500.15	35.00	2275.40
293.85	4600.15	16425.15	30.00	2276.15
296.11	4920.15	16375.15	25.00	2276.90
298.37	5260.15	16350.15	20.00	2277.65
300.63	5610.15	16350.15	15.00	2278.40
302.89	5980.15	16375.15	10.00	2279.15
305.15	6360.15	16425.15	5.00	2279.90
307.41	6760.15	16500.15	0.00	2280.65
309.67	7180.15	16600.15	0.00	2281.40
311.93	7620.15	16725.15	0.00	2282.15
314.19	8080.15	16875.15	0.00	2282.90
316.45	8560.15	17050.15	0.00	2283.65
318.71	9060.15	17250.15	0.00	2284.40
320.97	9580.15	17475.15	0.00	2285.15
323.23	10120.15	17725.15	0.00	2285.90
325.49	10680.15	18000.15	0.00	2286.65
327.75	11260.15	18300.15	0.00	2287.40
330.01	11860.15	18625.15	0.00	2288.15
332.27	12480.15	19000.15	0.00	2288.90
334.53	13120.15	19425.15	0.00	2289.65
336.79	13780.15	19900.15	0.00	2290.40
339.05	14460.15	20425.15	0.00	2291.15
341.31	15160.15	21000.15	0.00	2291.90
343.57	15880.15	21625.15	0.00	2292.65
345.83	16620.15	22300.15	0.00	2293.40
348.09	17380.15	23025.15	0.00	2294.15
350.35	18160.15	23800.15	0.00	2294.90
352.61	18960.15	24625.15	0.00	2295.65
354.87	19780.15	25500.15	0.00	2296.40
357.13	20620.15	26425.15	0.00	2297.15
359.39	21480.15	27400.15	0.00	2297.90
361.65	22360.15	28425.15	0.00	2298.65
363.91	23260.15	29500.15	0.00	2299.40
366.17	24180.15	30625.15	0.00	2300.15
368.43	25120.15	31800.15	0.00	2300.90
370.69	26080.15	33025.15	0.00	2301.65
372.95	27060.15	34300.15	0.00	2302.40
375.21	28060.15	35625.15	0.00	2303.15
377.47	29080.15	37000.15	0.00	2303.90
379.73	30120.15	38425.15	0.00	2304.65
381.99	31180.15	39900.15	0.00	2305.40
384.25	32260.15	41425.15	0.00	2306.15
386.51	33360.15	43000.15	0.00	2306.90
388.77	34480.15	44625.15	0.00	2307.65
391.03	35620.15	46300.15	0.00	2308.40
393.29	36780.15	48025.15	0.00	2309.15
395.55	37960.15	49800.15	0.00	2309.90
397.81	39160.15	51625.15	0.00	2310.65
400.07	40380.15	53500.15	0.00	2311.40
402.33	41620.15	55425.15	0.00	2312.15
404.59	42880.15	57400.15	0.00	2312.90
406.85	44160.15	59425.15	0.00	2313.65
409.11	45460.15	61500.15	0.00	2314.40
411.37	46780.15	63625.15	0.00	2315.15
413.63	48120.15	65800.15	0.00	2315.90
415.89	49480.15	68025.15	0.00	2316.65
418.15	50860.15	70300.15	0.00	2317.40
420.41	52260.15	72625.15	0.00	2318.15
422.67	53680.15	75000.15	0.00	2318.90
424.93	55120.15	77425.15	0.00	2319.65
427.19	56580.15	79900.15	0.00	2320.40
429.45	58060.15	82425.15	0.00	2321.15
431.71	59560.15	85000.15	0.00	2321.90
433.97	61080.15	87625.15	0.00	2322.65
436.23	62620.15	90300.15	0.00	2323.40
438.49	64180.15	93025.15	0.00	2324.15
440.75	65760.15	95800.15	0.00	2324.90
443.01	67360.15	98625.15	0.00	2325.65
445.27	68980.15	101500.15	0.00	2326.40
447.53	70620.15	104425.15	0.00	2327.15
449.79	72280.15	107400.15	0.00	2327.90
452.05	73960.15	110425.15	0.00	2328.65
454.31	75660.15	113500.15	0.00	2329.40
456.57	77380.15	116625.15	0.00	2330.15
458.83	79120.15	119800.15	0.00	2330.90
461.09	80880.15	123025.15	0.00	2331.65
463.35	82660.15	126300.15	0.00	2332.40
465.61	84460.15	129625.15	0.00	2333.15
467.87	86280.15	133000.15	0.00	2333.90
470.13	88120.15	136425.15	0.00	2334.65
472.39	90000.15	140000.15	0.00	2335.40
474.65	91900.15	143625.15	0.00	2336.15
476.91	93820.15	147300.15	0.00	2336.90
479.17	95760.15	151025.15	0.00	2337.65
481.43	97720.15	154800.15	0.00	2338.40
483.69	99700.15	158625.15	0.00	2339.15
485.95	101700.15	162500.15	0.00	2339.90
488.21	103720.15	166425.15	0.00	2340.65
490.47	105760.15	170400.15	0.00	2341.40
492.73	107820.15	174425.15	0.00	2342.15
494.99	109900.15	178500.15	0.00	2342.90
497.25	112000.15	182625.15	0.00	2343.65
499.51	114120.15	186800.15	0.00	2344.40
501.77	116260.15	191025.15	0.00	2345.15
504.03	118420.15	195300.15	0.00	2345.90
506.29	120600.15	199625.15	0.00	2346.65
508.55	122800.15	204000.15	0.00	2347.40
510.81	125020.15	208425.15	0.00	2348.15
513.07	127260.15	212900.15	0.00	2348.90
515.33	129520.15	217425.15	0.00	2349.65
517.59	131800.15	222000.15	0.00	2350.40
519.85	134100.15	226625.15	0.00	2351.15
522.11	136420.15	231300.15	0.00	2351.90
524.37	138760.15	236025.15	0.00	2352.65
526.63	141120.15	240800.15	0.00	2353.40
528.89	143500.15	245625.15	0.00	2354.15
531.15	145900.15	250500.15	0.00	2354.90
533.41	148320.15	255425.15	0.00	2355.65
535.67	150760.15	260400.15	0.00	2356.40
537.93	153220.15	265425.15	0.00	2357.15
540.19	155700.15	270500.15	0.00	2357.90
542.45	158200.15	275625.15	0.00	2358.65
544.71	160720.15	280800.15	0.00	2359.40
546.97	163260.15	286025.15	0.00	2360.15
549.23	165820.15	291300.15	0.00	2360.90
551.49	168400.15	296625.15	0.00	2361.65
553.75	171000.15	302000.15	0.00	2362.40
556.01	173620.15	307425.15	0.00	2363.15
558.27	176260.15	312900.15	0.00	2363.90
560.53	178920.15	318425.15	0.00	2364.65
562.79	181600.15	324000.15	0.00	2365.40
565.05	184300.15	329625.15	0.00	2366.15
567.31	187020.15	335300.15	0.00	2366.90
569.57	189760.15	341025.15	0.00	2367.65
571.83	192520.15	346800.15	0.00	2368.40
574.09	195300.15	352625.15	0.00	2369.15
576.35	198100.15	358500.15	0.00	2369.90
578.61	200920.15	364425.15	0.00	2370.65
580.87	203760.15	370400.15	0.00	2371.40
583.13	206620.15	376425.15	0.00	2372.15
585.39	209500.15	382500.15	0.00	2372.90
587.65	212400.15	388625.15	0.00	2373.65
589.91	215320.15	394800.15	0.00	2374.40
592.17	218260.15	401025.15	0.00	2375.15
594.43	221220.15	407300.15	0.00	2375.90
596.69	224200.15	413625.15	0.00	2376.65
598.95	227200.15	420000.15	0.00	2377.40
601.21	230220.15	426425.15	0.00	2378.15
603.47	233260.15	432900.15	0.00	2378.90
605.73	236320.15	439425.15	0.00	2379.65
607.99	239400.15	446000.15	0.00	2380.40
610.25	242500.15	452625.15	0.00	2381.15
612.51	245620.15	459300.15	0.00	2381.90
614.77	248760.15	466025.15	0.00	2382.65
617.03	251920.15	472800.15	0.00	2383.40
619.29	255100.15	479625.15	0.00	2384.15
621.55	258300.15	486500.15	0.00	2384.90
623.81	261520.15	493425.15	0.00	2385.65
626.07	264760.15	500400.15	0.00	2386.40
628.33	268020.15	507425.15	0.00	2387.15
630.59	271300.15	514500.15	0.00	2387.90
632.85	274600.15	521625.15	0.00	2388.65
635.11	277920.15	528800.15	0.00	2389.40
637.37	281260.15	536025.15	0.00	2390.15
639.63	284620.15	543300.15	0.00	2390.90
641.89	288000.15	550625.15	0.00	2391.65
644.15	291400.15	558000.15	0.00	2392.40
646.41	294820.15	565425.15	0.00	2393.15
648.67	298260.15	572900.15	0.00	2393.90
650.93	301720.15	580425.15	0.00	2394.65
653.19	305200.15	588000.15	0.00	2395.40
655.45	308700.15	595625.15	0.00	2396.15
657.71	312220.15	603300.15	0.00	2396.90
659.97	315760.15	611025.15	0.00	2397.65
662.23	319320.15	618800.15	0.00	2398.40
664.49	322900.15	626625.15	0.00	2399.15
666.75	326500.15	634500.15	0.00	2399.90
669.01	330120.15	642425.15	0.00	2400.65
671.27	333760.15	650400.15	0.00	2401.40
6				

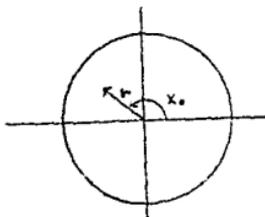
$$M = 140.55 + \frac{12362.9 - 12089.85}{474.25} (143.01 - 140.55) = 141.97$$

Así la mediana direccional es $141^{\circ}57'58''$, en término de días se trata del día No. 144 del año (el cual corresponde al 24 de mayo).

Las coordenadas del vector medio son:

$$(-0.5041, 0.4014)$$

Fig.



De acuerdo a los resultados de la sección 1.4.1. el valor del ángulo medio es $\bar{x}_0 = 141^{\circ}28'11''$ que en término de días se trata del día No. 143 del año (el día correspondiente es el 23 de mayo).

La longitud del vector medio es $r = 0.6444$, $0 \leq r \leq 1$ la cual es utilizada para calcular la variancia presentada en la sección 1.5.2. $s^2 = 40^{\circ}44'58''$, de acuerdo con la variancia propuesta por Batschelet (1979); en término de días es de 41.31 días.

También se calcula la variancia propuesta por Mardia (1972) obteniéndose $s^2 = 53^{\circ}42'51''$ que en término de días es de 54.5 días.

Dado que los datos presentan una distribución unimodal y aproximadamente simétrica se ajustó la distribución von Mises, presentada en el capítulo 2.

$$f(\theta) = (2\pi I_0(k))^{-1} e^{k \cos(\theta - \mu)}$$

$$0 \leq \theta \leq 2\pi, k > 0, 0 \leq \mu < 2\pi$$

De acuerdo al capítulo 3 los estimadores máximo verosímiles de μ y k son:

$$\hat{\mu} = x_0 = 2.469 \quad (\text{en radianes})$$

$$\hat{k} = A^{-1}(r) = A^{-1}(0.6444) = 1.69$$

El valor de k se obtiene con ayuda de la tabla No. 1 .

Así tenemos

$$f(\theta) = (2\pi I_0(1.69))^{-1} e^{1.69 \cos \theta} - 2.469$$

De acuerdo a la sección 3.2.3.a es posible calcular los intervalos de confianza de los parámetros k y μ de una distribución von Mises. Utilizando la Fig. A de la sección 3.2.2.a. para obtener el intervalo de confianza de μ con un coeficiente de confianza del 95% . Para $r = 0.64$ y $n = 24726$ se tiene que $\delta = 9$,

$$\text{así } \mu \in (\bar{x}_0 - \delta, \bar{x}_0 + \delta) = (132.46^\circ, 150.46^\circ).$$

Considerando que el máximo valor de n proporcionado por la tabla es de 200, en base a éste se ha calculado el intervalo de confianza; para un n mayor el intervalo de confianza se reduce, por lo que se puede asegurar que el intervalo encontrado incluye al intervalo de confianza del 95% de confianza.

Para el cálculo del intervalo de confianza del 90% para el parámetro k se utiliza la Fig.C de la sección 3.2.2.b. de nuevo $r = 0.6444$; dado que el máximo n tabulado es 200 es el que se considera.

Así $k \in (1.4, 1.9)$, nuevamente se puede asegurar que este intervalo incluye al intervalo de confianza del 90% para el parámetro k .

4.1.4. Pruebas de hipótesis.

Prueba 1 :

De acuerdo al capítulo 3 es posible aplicar la prueba de hipótesis Rayleigh para determinar si la distribución de las observaciones en el círculo, es uniforme o no.

Una forma alternativa de escribir las hipótesis H_0 y H_1 a la presentada en la sección 3.3.1.a es:

H_0 : La distribución de las observaciones en el círculo

es uniforme.

H_1 : No H_0 .

La estadística de prueba, dado que n es grande, es:

$$3nr^2 = 3(24726)(0.6444)^2 = 30802.5$$

La región de rechazo es:

$$\mathcal{R} = \{ \theta \mid 3nr^2 > \chi_{(3)}^{2,1-\alpha} \}$$

Entonces $\mathcal{R} = \{ \theta \mid 30802.5 > 7.815 \}$ por lo tanto la estadística de prueba cae en la región de rechazo, es decir que hay evidencia suficiente para asegurar que la distribución de las observaciones en el círculo no es uniforme.

Prueba 2 :

Una vez rechazada la hipótesis de uniformidad se aplica la prueba Ji-cuadrada presentada en la sección 3.3.2.a del capítulo 3.

La hipótesis nula a probar es :

$H_0 : \theta \sim$ von Mises con parámetros $k=1.6, \mu = 141.27.47''$

La hipótesis alternativa es :

$H_1 :$ La variable aleatoria θ no se distribuye von Mises con parámetros $k=1.6, \mu = 141.27.47''$

Para aplicar esta prueba se formaron 72 clases, debido a que se cuenta con una tabla que proporciona las función de distribución von Mises calculada en los puntos $\alpha = 5n, n=1,2,\dots,72$.

Dado que los datos originales se encuentran agrupados en 73 grupos se calcularon las frecuencias correspondientes a los nuevos intervalos, suponiendo que los datos se distribuyen uniformemente dentro de cada uno de los 73

grupos.

La tabla 4.1.d. proporciona el limite superior de los 72 intervalos y las correspondientes frecuencias observadas, la columna 3 muestra las frecuencias esperadas en cada intervalo suponiendo una distribución von Mises con $k = 1.6$, $\mu = 141 \cdot 27 \cdot 47''$.

TABLA 4.1.d.

LIMITE SUPERIOR DEL INTERVALO	FRECUENCIA OBSERVADA	FRECUENCIA ESPERADA	FRECUENCIA
11	6	2.25	2.25
10.00	15.00	16.25	16.25
15.00	43.45	44.51	44.51
20.00	47.74	74.14	74.14
25.00	74.57	93.09	93.09
30.00	52.00	75.29	75.29
35.00	107.50	104.76	104.76
40.00	107.94	121.18	121.18
45.00	169.57	129.25	129.25
50.00	271.32	159.43	159.43
55.00	189.24	183.22	183.22
60.00	154.63	197.14	197.14
65.00	301.02	242.07	242.07
70.00	262.14	279.67	279.67
75.00	344.82	317.75	317.75
80.00	334.48	281.25	281.25
85.00	540.02	410.04	410.04
90.00	535.25	462.01	462.01
95.00	530.80	474.89	474.89
100.00	512.72	570.74	570.74
105.00	857.14	706.82	706.82
110.00	642.53	828.23	828.23
115.00	455.75	956.56	956.56
120.00	449.02	1111.01	1111.01
125.00	674.78	1210.21	1210.21
130.00	1014.00	1371.74	1371.74
135.00	999.46	1575.10	1575.10
140.00	521.80	1711.25	1711.25
145.00	870.00	1881.15	1881.15
150.00	811.87	2071.12	2071.12
155.00	908.73	2284.73	2284.73
160.00	775.22	2522.74	2522.74
165.00	705.45	2785.24	2785.24
170.00	734.69	3072.43	3072.43
175.00	756.10	3384.57	3384.57
180.00	751.87	3721.87	3721.87
185.00	802.32	4084.60	4084.60
190.00	658.84	4473.14	4473.14
195.00	714.52	4887.83	4887.83
200.00	637.55	5329.01	5329.01
205.00	787.55	5807.12	5807.12
210.00	585.72	6322.74	6322.74
215.00	581.10	6876.33	6876.33
220.00	225.10	7467.32	7467.32
225.00	444.42	8096.47	8096.47
230.00	177.74	8864.47	8864.47
235.00	133.04	9671.82	9671.82
240.00	133.14	10619.04	10619.04
245.00	62.41	11706.80	11706.80
250.00	46.05	12934.66	12934.66
255.00	27.22	14312.27	14312.27
260.00	16.63	15849.36	15849.36
265.00	14.15	17556.58	17556.58
270.00	40.91	19444.60	19444.60
275.00	21.28	21523.04	21523.04
280.00	48.48	23802.52	23802.52
285.00	40.17	26293.64	26293.64
290.00	23.42	29007.07	29007.07
295.00	49.83	31953.42	31953.42
300.00	20.00	35147.32	35147.32
305.00	22.57	38599.44	38599.44
310.00	24.80	42322.32	42322.32
315.00	41.94	46337.54	46337.54
320.00	27.22	50654.66	50654.66
325.00	39.67	55283.01	55283.01
330.00	20.28	60234.14	60234.14
335.00	25.17	65518.66	65518.66
340.00	11.00	71147.07	71147.07
345.00	17.77	77130.04	77130.04
350.00	25.44	83478.25	83478.25
355.00	17.22	90202.32	90202.32
360.00	10.87	97422.74	97422.74
TOTAL			14100.27



La estadística de prueba que se obtiene es : $T=1600.57$

La region critica es:

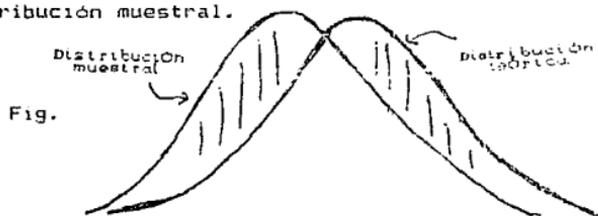
$$S = \{ \theta \mid T > \chi^2_{(60)}^{2, 1-\alpha} \}$$

Sustituyendo los correspondientes valores se obtiene:

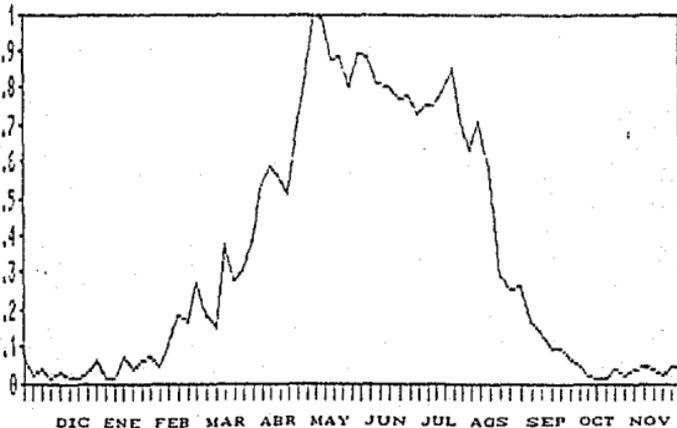
$$\hat{\sigma} = \{ \theta \mid 1600.57 > 90.5 \} \quad \alpha = .05$$

Por lo tanto se rechaza H_0 , es decir, la distribución difiere de una distribución von Mises con parámetros $k = 1.6$ y $\mu = 141^{\circ}27'47''$.

Analizando la tabla 4.1.d se ha marcado una 5a. columna con '-' los renglones en que los valores de la cuarta columna son mayores de 40 y en una 6a. columna se han marcado los renglones en que los valores de la cuarta columna son mayores de 30. Observando estas dos nuevas columnas se observa que se forman dos grupos, esto podría indicar un defasamiento entre la distribución teórica y la distribución muestral.



Para llevar a cabo una comparación gráfica entre la distribución teórica propuesta y la distribución muestral se podría utilizar el polígono de frecuencias del histograma lineal.



Pero es mejor si se cuenta con una gráfica más suave

de la misma distribución muestral. Para obtenerla se utiliza una técnica de estimación de densidades por medio de distribuciones kernel⁽¹⁾.

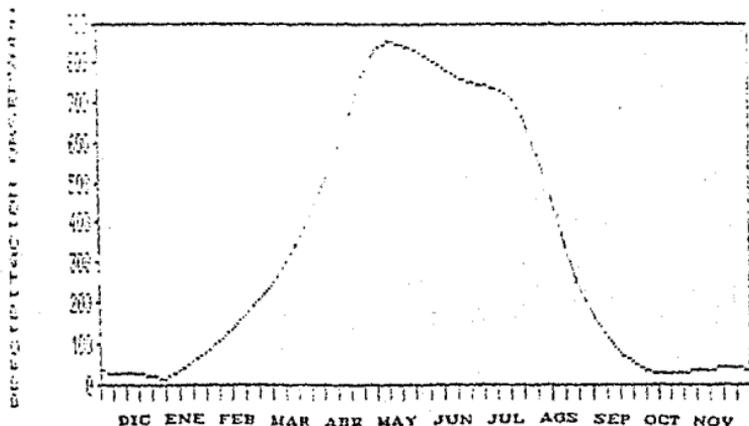
Para lo cual se busca una función kernel o núcleo que permita describir a la distribución muestral en una forma más suave.

En este estudio se utilizó el siguiente núcleo de una distribución Normal.

$$F(t) = \sum_{i=1}^{72} f_i e^{-\lambda(t-\theta_i)^2}$$

Donde θ_i es el punto medio del intervalo i -ésimo y f_i es la frecuencia observada en el i -ésimo intervalo.

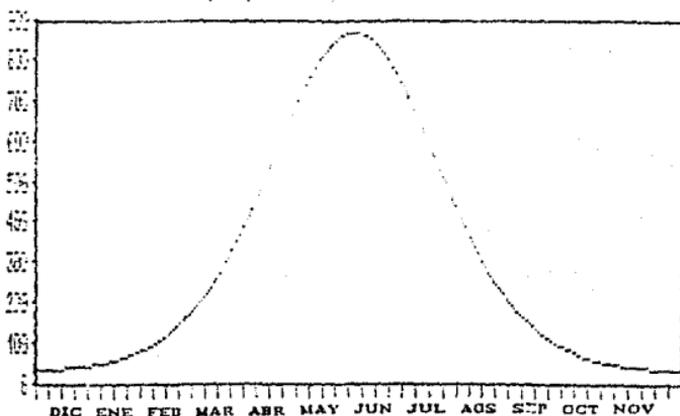
Como se puede notar este núcleo está formado por la suma de 72 núcleos de distribuciones Normales, cada uno de los cuales representa la información contenida en los 72 intervalos en que está dividida la información de la muestra. Graficando el valor de $F(t)$ para 50 distintos puntos del intervalo $(0, 2\pi)$ se obtiene la siguiente gráfica



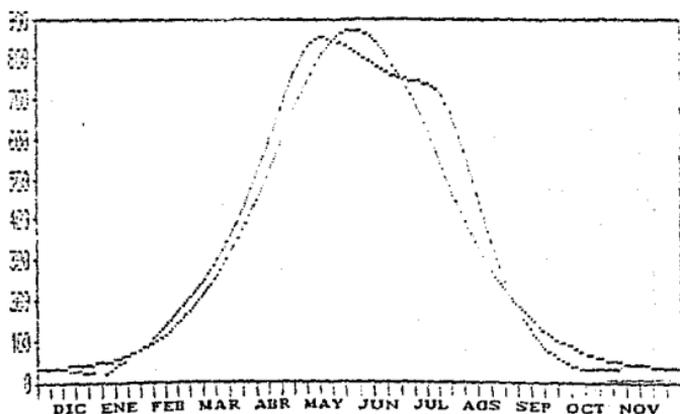
La cual, como se puede apreciar es una representación más suave, que la que proporciona el polígono de frecuencia.

1. - Good y Gaskins (1980).

Comparando esta gráfica con la gráfica de la distribución teórica propuesta, von Mises.



Es posible detectar 3 principales regiones que producen las diferencias más grandes en la cuarta columna de la tabla 4.1.d.



Estas tres regiones se marcan en esta gráfica y en la tabla 4.1.d como I, II y III. Donde la región II se produce debido a que el valor de la densidad muestral es menor que el de la distribución teórica, y las regiones I y III se

producen debido a que el valor de la densidad muestral es mayor al de la distribución teórica.

Aun cuando se tiene una distribución unimodal, ya se ha comprobado que la distribución de la muestra difiere significativamente de una distribución von Mises, también se han localizado las regiones en las que difieren ambas distribuciones; ahora cabría preguntarse si es posible ajustar alguna otra distribución teórica, y la respuesta es si. Quizá podría ajustarse una combinación de distribuciones, pero dado que se tiene un número de observaciones tan grande $n = 24726$, dividido en 72 grupos, cualquier prueba de bondad de ajuste se torna muy sensible y ante la más mínima diferencia entre la distribución teórica que se proponga y la distribución muestral, se rechazara el ajuste. Por esta razón la única distribución que podrá ser aceptada es la distribución que describa fielmente a la distribución muestral, o en otras palabras, la única posible distribución que describirá el comportamiento de la distribución poblacional del fenómeno observado es la distribución muestral misma. Lo cual es aceptable, ya que el tamaño de la muestra es tan grande, se puede considerar que la distribución de la muestra es una buena aproximación de la distribución poblacional.

Nótese que todos los valores estadísticos calculados en el presente capítulo tales como moda, media, varianza, etc. continúan proporcionando una descripción adecuada de la distribución, debido a que se tiene una distribución unimodal, aun cuando ésta no haya sido von Mises.

Finalmente, es importante hacer notar que se ha considerado cada ml. medido de precipitación pluvial como una ocurrencia del fenómeno, si en lugar de ml. se hubieran considerado litros algunos resultados cambiarían, en especial los de las pruebas de hipótesis empleadas. Esto no es un problema que resuelva el análisis de datos direccionales y se presenta de igual forma en Estadística lineal.

CONCLUSIONES

La estadística lineal, cuyo estudio y aplicación es muy común en el análisis de fenómenos aleatorios, presenta limitantes al tratar con fenómenos aleatorios direccionales.

Los fenómenos aleatorios direccionales poseen características que los definen y los diferencian de los fenómenos comúnmente estudiados por la estadística lineal.

El análisis de datos direccionales proporciona una teoría estadística y un conjunto de técnicas para analizar éstos fenómenos aleatorios, conservando y considerando las características especiales que los definen.

Su aplicación es importante principalmente en Biología, Meteorología, Astronomía y áreas interesadas en estudios sobre la Tierra y el cosmos que la rodea, en las cuales es posible encontrar análisis de fenómenos que involucran direcciones, cuyo estudio desde un punto de vista de la estadística lineal resultaría difícil, debido principalmente a las diferencias existentes entre las características de los datos que pueden ser representados en el eje real, los cuales son los usuales de estadística lineal, y los datos direccionales, cuya representación en general requiere al menos de dos dimensiones.

La aplicación de estas técnicas en el estudio de fenómenos que se desarrollan en el tiempo y que presentan una periodicidad conocida es de utilidad para poder estudiar su comportamiento dentro del periodo conocido, pudiendo así conocer los momentos más importantes de dicho periodo, teniendo el problema de no contar con pruebas de bondad de ajuste aplicables a distribuciones cuyas frecuencias no sean conteos, como fue el caso del problema analizado en el capítulo 4., siendo este un problema que no es exclusivo del análisis de datos direccionales sino incluso de la estadística lineal.

APENDICE

TABLA 1.

El estimador máximo verosímil k dado r en el caso de la distribución von Mises.

$k=A^{-1}(r)$					
R	k	R	k	R	k
0.00	0.00000	0.33	0.74783	0.70	2.01363
-01	-0.00000	-36	-77241	-71	-2.07685
-02	-0.04001	-37	-79730	-72	-2.14359
-03	-0.06001	-38	-82152	-73	-2.21425
-04	-0.07000	-39	-84512	-74	-2.28920
-05	-1.00113	-40	-87403	-75	-2.36930
-06	-1.10022	-41	-90043	-76	-2.45490
-07	-1.40334	-42	-92720	-77	-2.54646
-08	-1.60551	-43	-95440	-78	-2.64413
-09	-1.80771	-44	-98207	-79	-2.75332
-10	-2.01011	-45	-1.01022	-80	-2.87139
-11	-2.21244	-46	-1.03839	-81	-3.00020
-12	-2.41475	-47	-1.06610	-82	-3.14262
-13	-2.62221	-48	-1.09788	-83	-3.30114
-14	-2.82779	-49	-1.12825	-84	-3.47991
-15	-3.03144	-50	-1.15932	-85	-3.68041
-16	-3.24199	-51	-1.19103	-86	-3.91072
-17	-3.45053	-52	-1.22350	-87	-4.17703
-18	-3.65959	-53	-1.25672	-88	-4.48376
-19	-3.87007	-54	-1.29077	-89	-4.83371
-20	-4.08234	-55	-1.32570	-90	-5.2047
-21	-4.29662	-56	-1.36156	-91	-5.6122
-22	-4.51100	-57	-1.39842	-92	-6.0594
-23	-4.72731	-58	-1.43635	-93	-7.4257
-24	-4.94533	-59	-1.47543	-94	-8.6104
-25	-5.16499	-60	-1.51574	-95	-10.2716
-26	-5.3862	-61	-1.55738	-96	-12.7661
-27	-5.6077	-62	-1.60044	-97	-15.2255
-28	-5.83150	-63	-1.64506	-98	-25.2522
-29	-6.06225	-64	-1.69134	-99	-50.2421
-30	-6.2922	-65	-1.73945	1.00	∞
-31	-6.5242	-66	-1.78933		
-32	-6.7587	-67	-1.84177		
-33	-6.9952	-68	-1.89637		
-34	-7.2336	-69	-1.95357		

Longitud del resultante $r=A(k)$ para la distribución von Mises, dado k .

k	r	k	r	k	r
0.0	0.00000	3.5	0.84110	7.0	0.92553
0.1	-0.0994	3.6	-0.61616	7.1	-9.2661
0.2	-0.9950	3.7	-8.5091	7.2	-9.2770
0.3	-1.4834	3.8	-8.5537	7.3	-9.2874
0.4	-1.9610	3.9	-8.5926	7.4	-9.2975
0.5	-2.4250	4.0	-8.6312	7.5	-9.3072
0.6	-2.8726	4.1	-8.6726	7.6	-9.3168
0.7	-3.3018	4.2	-8.7079	7.7	-9.3260
0.8	-3.7108	4.3	-8.7414	7.8	-9.3350
0.9	-4.0984	4.4	-8.7732	7.9	-9.3438
1.0	-4.4639	4.5	-8.8033	8.0	-9.3524
1.1	-4.8070	4.6	-8.8320	8.1	-9.3607
1.2	-5.1278	4.7	-8.8593	8.2	-9.3683
1.3	-5.4267	4.8	-8.8853	8.3	-9.3767
1.4	-5.7042	4.9	-8.9101	8.4	-9.3844
1.5	-5.9613	5.0	-8.9338	8.5	-9.3919
1.6	-6.1990	5.1	-8.9565	8.6	-9.3991
1.7	-6.4183	5.2	-8.9782	8.7	-9.4064
1.8	-6.6204	5.3	-8.9990	8.8	-9.4134
1.9	-6.8065	5.4	-9.0190	8.9	-9.4202
2.0	-6.9777	5.5	-9.0382	9.0	-9.4269
2.1	-7.1353	5.6	-9.0566	9.2	-9.4398
2.2	-7.2803	5.7	-9.0743	9.4	-9.4521
2.3	-7.4138	5.8	-9.0913	9.6	-9.4639
2.4	-7.5367	5.9	-9.1078	9.8	-9.4752
2.5	-7.6500	6.0	-9.1236	10	-9.4860
2.6	-7.7545	6.1	-9.1389	12	-9.5730
2.7	-7.8511	6.2	-9.1536	15	-9.6407
2.8	-7.9404	6.3	-9.1678	20	-9.7467
2.9	-8.0231	6.4	-9.1816	24	-9.7937
3.0	-8.0999	6.5	-9.1949	30	-9.8319
3.1	-8.1711	6.6	-9.2078	40	-9.8719
3.2	-8.2355	6.7	-9.2202	60	-9.9163
3.3	-8.2993	6.8	-9.2323	120	-9.9582
3.4	-8.3570	6.9	-9.2440	∞	-1.0000

TABLA 2.

valores críticos de la prueba de uniformidad
Rayleigh con la estadística de prueba R^2 .

$$P(R^2 \leq r) = \alpha$$

n	$\alpha \rightarrow$	0.10	0.05	0.025	0.01	0.001
5		0.677	0.734	0.816	0.879	0.991
6		.618	.690	.753	.825	.940
7		.572	.642	.702	.771	.891
8		.535	.602	.660	.725	.847
9		.504	.569	.624	.687	.808
10		.478	.540	.591	.655	.773
11		.456	.516	.567	.627	.745
12		.437	.494	.544	.602	.716
13		.420	.475	.524	.580	.692
14		.405	.458	.505	.560	.667
15		.391	.443	.489	.542	.649
16		.379	.429	.474	.525	.630
17		.367	.417	.460	.510	.613
18		.357	.405	.447	.495	.597
19		.348	.394	.436	.484	.583
20		.339	.385	.425	.472	.567
21		.331	.375	.415	.461	.556
22		.323	.367	.405	.451	.544
23		.316	.359	.397	.441	.533
24		.309	.351	.389	.432	.522
25		.303	.344	.381	.423	.512
30		.277	.315	.348	.387	.470
35		.256	.292	.323	.359	.436
40		.240	.273	.302	.336	.409
45		.226	.257	.285	.318	.386
50		.214	.244	.270	.301	.367
100		.15	.17	.19	.21	.26
$2nR^2 \approx \chi^2$		4.605	5.991	7.378	9.210	13.816

1. Reproducida de Hatschelat(1981).

TABLA 3.

valores críticos para la prueba de Hodges-Ajne
con la estadística de prueba η .

$$P(\eta \leq \eta_\alpha) = \alpha$$

n	$\alpha \rightarrow$	0.10	0.05	0.025	0.01
9	0	0	0	0	0
10	1	0	0	0	0
11	1	0	0	0	0
12	1	0	0	0	0
13	1	0	0	0	0
14	2	1	1	1	0
15	2	2	2	2	1
16	2	2	2	2	1
17	3	2	2	2	1
18	3	3	3	3	2
19	3	3	3	3	2
20	4	3	3	3	2
21	4	4	4	4	2
22	5	4	4	4	3
23	5	4	4	4	3
24	5	5	5	5	3
25	6	5	5	5	4
30	7	7	7	7	5
35	9	9	9	9	7
40	11	10	10	10	9
50	15	14	13	13	12

1. Reproducida de Hatschelat(1981).

TABLA 4.

k	Cuantiles de $F_{k,n}^{(1)}$						
	p = .750	.900	.950	.975	.990	.995	.999
1	1.321	2.706	3.841	5.024	6.635	7.879	10.53
2	2.771	4.605	5.991	7.376	9.210	10.60	13.82
3	4.108	6.251	7.815	9.348	11.34	12.84	16.27
4	5.385	7.779	9.485	11.14	13.23	14.86	18.47
5	6.626	9.216	11.07	12.83	15.09	16.75	20.51
6	7.841	10.64	12.59	14.45	16.81	18.55	22.45
7	9.031	12.02	14.07	16.01	18.48	20.28	24.32
8	10.22	13.36	15.51	17.51	20.09	21.96	26.13
9	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	13.70	17.28	19.68	21.92	24.71	26.76	31.26
12	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	15.98	19.81	22.36	24.74	27.69	29.82	34.55
14	17.12	21.06	23.68	26.12	29.14	31.42	36.12
15	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	19.37	23.54	26.30	28.85	32.03	34.27	39.25
17	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	24.93	29.62	32.67	35.48	38.93	41.40	46.80
22	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	28.24	33.20	36.42	39.37	42.98	45.54	51.18
25	29.34	34.38	37.65	40.65	44.31	46.89	52.62
26	30.43	35.56	38.87	41.92	45.64	48.29	54.05
27	31.53	36.74	40.11	43.19	46.96	49.64	55.48
28	32.62	37.92	41.34	44.46	48.26	50.99	56.89
29	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	45.62	51.91	55.76	59.34	63.69	66.77	75.40
50	56.33	63.17	67.50	71.42	76.15	79.40	86.66
60	66.98	74.10	79.08	83.10	88.18	91.95	99.61
70	77.58	85.33	90.53	95.02	100.4	104.2	112.9
80	88.12	96.54	101.9	106.6	112.3	116.3	124.8
90	98.65	107.6	113.1	118.1	124.1	128.5	137.2
100	109.1	118.3	124.3	129.6	135.4	140.2	149.4
x_p	.675	1.281	1.645	1.920	2.326	2.576	3.090

Si $k > 100$ usar la aproximación:

$$x_p = \frac{1}{2} \left[1 + \sqrt{2k} \right] F$$

TABLA 5.

Valores críticos para la prueba de Kuiper con la estadística de prueba T_n^2 V_n

k	$P(T_n^2 \leq \alpha)$			
	$\alpha = 0.10$	0.05	0.01	0.005
5	1.46	1.37	1.76	1.84
6	1.47	1.38	1.79	1.87
7	1.48	1.40	1.81	1.89
8	1.49	1.41	1.83	1.91
9	1.50	1.42	1.84	1.93
10	1.51	1.43	1.85	1.94
11	1.51	1.43	1.85	1.94
12	1.52	1.44	1.87	1.95
13	1.52	1.44	1.88	1.95
14	1.53	1.45	1.88	1.97
15	1.53	1.45	1.89	1.98
16	1.53	1.46	1.89	1.98
17	1.54	1.46	1.90	1.99
18	1.54	1.46	1.90	1.99
19	1.54	1.46	1.90	2.00
20	1.55	1.47	1.91	2.00
21	1.55	1.47	1.92	2.01
22	1.56	1.48	1.92	2.02
23	1.57	1.48	1.94	2.03
24	1.58	1.49	1.95	2.04
100	1.59	1.72	1.97	2.06
∞	1.62	1.75	2.00	2.10

Reproducida de Watschelet(1983).

TABLE 4.

Función de distribución con Mises parámetro de concentración k , ángulo medio μ 180°

θ	F(θ)										
	$k=0$	$k=2$	$k=4$	$k=6$	$k=8$	$k=10$	$k=12$	$k=14$	$k=16$	$k=18$	$k=20$
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.1369	0.1126	0.0895	0.0697	0.0536	0.0404	0.0291	0.0202	0.0144	0.0103	0.0074
10	0.2778	0.2254	0.1793	0.1397	0.1074	0.0811	0.0594	0.0422	0.0304	0.0219	0.0153
15	0.4177	0.3267	0.2408	0.1708	0.1200	0.0859	0.0622	0.0450	0.0322	0.0228	0.0162
20	0.5556	0.4322	0.3300	0.2426	0.1775	0.1304	0.0974	0.0719	0.0519	0.0374	0.0271
25	0.6944	0.5663	0.4331	0.3357	0.2544	0.1923	0.1459	0.1093	0.0794	0.0574	0.0419
30	0.8333	0.6816	0.5467	0.4304	0.3329	0.2535	0.1903	0.1411	0.1035	0.0765	0.0553
35	0.9722	0.7978	0.6420	0.5071	0.4016	0.3107	0.2366	0.1696	0.1241	0.0906	0.0654
40	1.1111	0.9152	0.7392	0.5861	0.4647	0.3584	0.2730	0.2004	0.1459	0.1079	0.0799
45	1.2500	1.0138	0.8156	0.6429	0.5228	0.4089	0.3162	0.2359	0.1699	0.1259	0.0939
50	1.3889	0.9905	0.7927	0.6327	0.5221	0.4089	0.3162	0.2359	0.1699	0.1259	0.0939
55	1.5278	1.2757	1.0452	0.8409	0.6653	0.5184	0.3983	0.2983	0.2122	0.1544	0.1129
60	1.6667	1.3992	1.1529	0.9311	0.7428	0.5825	0.4509	0.3450	0.2514	0.1865	0.1393
65	1.8056	1.5246	1.2639	1.0295	0.8231	0.6517	0.5062	0.3919	0.2994	0.2204	0.1670
70	1.9444	1.6520	1.3784	1.1300	0.9128	0.7265	0.5513	0.4254	0.3245	0.2425	0.1820
75	2.0833	1.7813	1.4968	1.2346	1.0064	0.8078	0.6003	0.4603	0.3630	0.2745	0.2037
80	2.2222	1.9132	1.6192	1.3445	1.1064	0.8962	0.6716	0.5090	0.4047	0.3101	0.2301
85	2.3611	2.0471	1.7460	1.4602	1.2139	0.9923	0.8028	0.6436	0.4512	0.3422	0.2533
90	2.5000	2.1844	1.8772	1.5801	1.3289	1.0975	0.8974	0.7277	0.5363	0.3843	0.2809
95	2.6389	2.3222	2.0110	1.7004	1.4522	1.2122	1.0025	0.8228	0.6014	0.4515	0.3343
100	2.7778	2.4613	2.1517	1.8242	1.5844	1.3372	1.1191	0.9302	0.6769	0.4930	0.3630
105	2.9167	2.6019	2.2992	2.0000	1.7260	1.4734	1.2483	1.0514	0.8016	0.6150	0.4507
110	3.0556	2.7529	2.4498	2.1354	1.8754	1.6217	1.3913	1.1876	0.9101	0.7011	0.5107
115	3.1944	2.9014	2.6034	2.2814	2.0192	1.9785	1.5491	1.3405	1.0152	0.7962	0.5962
120	3.3333	3.0522	2.7659	2.4402	2.1805	2.1444	1.7226	1.4112	1.1224	0.8565	0.6565
125	3.4722	3.2053	2.9314	2.6107	2.3440	2.3440	1.9125	1.5099	1.2103	0.9302	0.7302
130	3.6111	3.3606	3.1017	2.8420	2.5282	2.5440	2.1194	1.6106	1.3206	1.0394	0.8194
135	3.7500	3.5180	3.2766	3.0327	2.7026	2.8116	2.3454	1.7400	1.4545	1.1645	0.9345
140	3.8889	3.6774	3.4559	3.2366	2.8973	3.0334	2.5977	1.8877	1.6127	1.3127	1.0127
145	4.0278	3.8385	3.6392	3.4533	3.1139	3.3134	2.8433	2.0544	1.7966	1.4866	1.1166
150	4.1667	4.0013	3.8263	3.6843	3.3456	3.5583	3.1172	2.2544	1.9944	1.8044	1.2344
155	4.3056	4.1655	4.0166	3.9828	3.5927	3.8146	3.4040	2.4816	2.2199	2.1599	1.3702
160	4.4444	4.3309	4.2098	4.2941	3.8490	4.0770	3.7079	2.7399	2.5899	2.4769	1.5201
165	4.5833	4.4973	4.4053	4.6095	4.1222	4.3555	4.0208	3.0294	2.9744	2.8744	1.6744
170	4.7222	4.6644	4.6025	4.9379	4.4222	4.6466	4.3423	3.4200	3.4200	3.4200	1.8400
175	4.8611	4.8321	4.8007	5.2684	4.7353	4.9722	4.6722	3.8481	3.8481	3.8481	2.0181
180	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	2.0000
185	5.1389	5.1679	5.1991	5.2316	5.2647	5.2978	5.3304	5.3619	5.3923	5.4214	2.1214
190	5.2778	5.3356	5.3975	5.4621	5.5278	5.5944	5.6577	5.7200	5.7799	5.8378	2.2378
195	5.4167	5.5027	5.5847	5.6693	5.7558	5.8444	5.9322	6.0176	6.1014	6.1825	2.3525
200	5.5556	5.6601	5.7602	5.8559	5.9480	6.0366	6.1221	6.2044	6.2833	6.3599	2.4699
205	5.6944	5.8245	5.9834	6.1372	6.2923	6.4434	6.5900	6.7362	6.8710	6.9952	2.5852
210	5.8333	5.9987	6.1737	6.3537	6.5344	6.7117	6.8828	7.0436	7.1980	7.3422	2.6922
215	5.9722	6.1615	6.3508	6.5457	6.7481	6.9616	7.1660	7.3644	7.5569	7.7427	2.7927
220	6.1111	6.3225	6.5441	6.7674	6.9927	7.2091	7.4181	7.6082	7.7817	7.9391	2.8891
225	6.2500	6.4820	6.7234	6.9673	7.2074	7.4384	7.6465	7.8392	8.0145	8.1731	2.9831
230	6.3889	6.6494	6.9143	7.1580	7.4118	7.6540	7.8806	8.0944	8.2794	8.4409	3.0749
235	6.5278	6.8247	7.0986	7.3413	7.6036	7.8536	8.0875	8.2991	8.4897	8.6598	3.1648
240	6.6667	6.9478	7.2341	7.5168	7.7886	8.0414	8.2774	8.4888	8.6772	8.8435	3.2535
245	6.8056	7.0966	7.3946	7.8446	7.9908	8.2175	8.4509	8.6595	8.8414	9.0038	3.3408
250	6.9444	7.2471	7.5592	7.9426	8.1228	8.3783	8.6083	8.8124	8.9944	9.1429	3.4269
255	7.0833	7.3991	7.7096	7.9920	8.2740	8.4766	8.7517	8.9446	9.1184	9.2830	3.5110
260	7.2222	7.5567	7.8641	8.1418	8.4456	8.6628	8.8909	9.0698	9.2307	9.3761	3.5931
265	7.3611	7.7208	7.9810	8.2794	8.5478	8.7975	9.0375	9.1772	9.3235	9.4545	3.6745
270	7.5000	7.8166	8.1225	8.4099	8.6711	8.9025	9.1026	9.2223	9.3417	9.4501	3.7541
275	7.6389	7.9129	8.2540	8.5338	8.7861	9.0075	9.1922	9.3166	9.4478	9.5587	3.8327
280	7.7778	8.0168	8.3868	8.6535	8.8934	9.1018	9.2824	9.4110	9.5323	9.6649	3.9104
285	7.9167	8.1185	8.5312	8.7692	8.9936	9.1922	9.3591	9.4870	9.6019	9.7315	3.9872
290	8.0556	8.2180	8.6718	8.8814	9.0852	9.2792	9.4320	9.5538	9.6855	9.8150	4.0631
295	8.1944	8.3152	8.8100	8.9902	9.1782	9.3592	9.5012	9.6301	9.7604	9.8904	4.1381
300	8.3333	8.4098	8.9471	9.0969	9.2722	9.4414	9.5741	9.6981	9.8166	9.9356	4.2126
305	8.4722	8.5023	9.0914	9.2001	9.3577	9.5166	9.6415	9.7697	9.8872	9.9951	4.2856
310	8.6111	8.5920	9.2319	9.3001	9.4577	9.5913	9.6949	9.8113	9.9252	9.9954	4.3581
315	8.7500	8.6784	9.3644	9.3921	9.5422	9.6723	9.7638	9.8701	9.9821	9.9944	4.4294
320	8.8889	8.7624	9.5004	9.4791	9.6233	9.7466	9.8314	9.9114	9.9955	9.9994	4.5004
325	9.0278	8.8422	9.6380	9.5629	9.7004	9.8211	9.8999	9.9500	9.9999	9.9999	4.5709
330	9.1667	8.9184	9.7783	9.6436	9.7671	9.8765	9.9309	9.9800	9.9999	9.9999	4.6409
335	9.3056	9.0013	9.9200	9.7216	9.8274	9.9117	9.9411	9.9844	9.9999	9.9999	4.7104
340	9.4444	9.0813	9.9792	9.7974	9.8829	9.9355	9.9670	9.9914	9.9999	9.9999	4.7794
345	9.5833	9.1583	9.9701	9.8702	9.9310	9.9644	9.9899	9.9999	9.9999	9.9999	4.8479
350	9.7222	9.2326	9.9820	9.9400	9.9744	9.9899	9.9999	9.9999	9.9999	9.9999	4.9159
355	9.8611	9.3044	9.9915	9.9901	9.9944	9.9999	9.9999	9.9999	9.9999	9.9999	4.9834
360	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	5.0500

BIBLIOGRAFIA

BATSCHLET, E. (1981). Circular Statistics in Biology, Academic Press Inc. (London) Ltd. 1-211.

DOLCIANI, M. P.; BERMAN, S.L.; WOOTON, W. (1978). Algebra Moderna y Trigonometría, Publicaciones Culturales S. A. México. 371-493.

GOOD, I.J. y GASKINS, R.A. (1980). Density Estimation and Bump-Hunting by the penalized likelihood method exemplified by scattering and meteorite data. Journal of the American Statistical Association, Vol.75 No.369, 42-55.

MARDIA, K.V. (1972). Statistics of directional data, Academic Press Inc. (London) Ltd. 3-128, 224-308.

PEREZ SALVADOR, B.R. (1979). Estimación de la periodicidad en la precipitación pluvial y un estudio de la acumulación de lluvias a través del año. TESIS Actuario. Fac. de ciencias UNAM. México.

RUDIN, W. (1980). Principios de análisis matemático. McGraw-Hill. México. 129-143.

WATSON, G.S. (1983). Statistics on Spheres, John Willey & Sons, Inc. 187-208.