Datos direccionales.

Ánalisis descriptivo.

Leyenda Rodríguez, María

17 de febrero de 2011

Introducción

Distribuciones notables

Test de uniformidad

Series de tiempo con datos direccionales

- En diversos campos surgen problemas estadísticos dónde los datos son dados mediante medidas angulares dando la orientación o angulos en el plano (datos circulares) o en el espacio (datos esféricos).
- Datos circulares es el caso más simple de esta categoría de datos llamada **datos direccionales**, dónde la única respuesta es no escalar, pero es angular o direccional.

Notación

- Las direcciones en el plano pueden ser observadas como vectores en el plano o como puntos en el círculo unidad. Tambien se pueden observar, como ángulos o como números complejos.
- Escogemos direccción y orientación inicial.

Luego cada punto **x** en el círculo unidad puede ser representado por un ángulo θ o por un número complejo unitario z.

$$x = (cos\theta, sen\theta)$$
$$z = e^{i\theta} = cos\theta + isen\theta$$

Medidas de localización

- Sean $x_1, \ldots, x_n$ cuyos ángulos correspondientes θ_i, $i = 1, \ldots, n$.
- La dirección media $\bar{\theta}$ de $\theta_1, \ldots, \theta_n$ es la dirección de $x_1 + \ldots + x_n$. Es tambien el centro de masa $\bar{x}$ de $x_1, \ldots, xn$.
- Las coordenadas catesianas de x_j son $(cos\theta_j, sen\theta_j)$
- Las coordenadas cartesianas del centro de masa son $(\bar{C}, \bar{S})$.

$$\bar{C} = \frac{1}{n} \sum_{j=1}^{n} cos\theta_j$$

$$\bar{S} = \frac{1}{n} \sum_{j=1}^{n} sen\theta_j$$

Medidas de localización

- Dónde $\bar{\theta}$ es la solución de las ecuaciones

$$\bar{C} = \bar{R}cos\bar{\theta}$$
$$\bar{S} = \bar{R}sen\bar{\theta}$$

- La media de la longitud resultante $\bar{R}$ es dada por

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$$

Medidas de localización

- Notemos que $\bar{\theta}$ no está definida cuando $\bar{R} = 0$
- Cuando $\bar{R} > 0$, $\bar{\theta}$ es dada por

$$\bar{\theta} = \begin{cases} tan^{-1}(\bar{S}/\bar{C}) & \text{si} \quad \bar{C} \geq 0 \\ tan^{-1}(\bar{S}/\bar{C}) + \pi & \text{si} \quad \bar{C} < 0 \end{cases}$$

Notemos que en el contexto de estadística circular $\bar{\theta}$ no es la media $(\theta_1 + \ldots + \theta_n)/n$

Medidas de localización

- Notemos que la media muestral de la dirección es equivariante bajo rotación.
- Diferentes estadísticos utilizando diferentes sistemas de coordenadas estarán de acuerdo en donde está la media la muestral, a pesar de que puede usar números diferentes para describir su posición

Medidas de localización

- La mediana muestral de la dirección $\bar{\theta}$ de los ángulos $(\theta_1, \ldots, \theta_n)/n$ es algún ángulo ϕ tal que
  1. La mitad de los puntos se encuentran en el arco $[\phi, \phi + \pi)$
  2. la mayoría de los puntos están más cerca de ϕ que de $\phi + \pi$
- Cuando n es par, la mediana coincide con uno de ellos.
- Cuando n es impar, es conveniente tomar la mediana como el punto medio de dos puntos adyacentes adecuados.

Medidas de concentración y dispersión

- $\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$, entonces si $x_1, \ldots, x_n$ son vectores unitarios, $0 \leq \bar{R} \leq 1$.
- Si las direcciones $\theta_1, \ldots, \theta_n$ están estrechamente agrupadas luego $\bar{R} = 1$.
- Si $\theta_1, \ldots, \theta_n$ están muy dispersas luego $\bar{R} = 1$ será practicamente 0.

Por tanto, $\bar{R}$ **es una medida de concentración del conjunto de datos.**.

- $\bar{R}$ es invariante bajo rotaciones.
- La longitud resultante $\bar{R}$ es la longitud del vector resultante $x_1 + \ldots + x_n$.
- $R = n\bar{R}$

Medidas de concentración y dispersión

- La media de longitud resultante $\bar{R}$ es la medida de dispersión más importante.
- A veces emplearemos otras medidas de dispersión de datos circulares, para comparar con datos en linea: la medida más simple es la **varianza circular muestral**

$$V = 1 - \bar{R}$$

Medidas de concentración y dispersión

Otras medidas de dispersión

- Distancia entre dos ángulos θ y ξ

$$min(\theta - \xi, 2\pi - (\theta - \xi)) = \pi - |\pi - |\theta - \xi||$$

- Medida de dispersión de los ángulos $\theta_1, \ldots, \theta_n$ sobre un ángulo dado α es

$$d_0(\alpha) = \frac{1}{n} \sum_{i=1}^{n} (\pi - |\pi - |\theta_i - \alpha||)$$

- la función d_0 toma el mínimo en la mediana muestral $\tilde{\theta}$. La desviación circular media es $d_0(\tilde{\theta})$

Distribución circular

- Una distribución circular es una distribución de probabilidad la cual está concentrada en la circunferencia de círculo unidad.
- Las distribuciones circulres son de dos tipos:
  1. Discretas- asignan masas de probabilidad solo a un número de direcciones
  2. Absolutamente contínuas
- Existe la funciím de densidad $f(\theta)$
  1. $f(\theta) \geq 0$
  2. $\int_0^{2\pi} f(\theta) d\theta = 1$
  3. $f(\theta) = f(\theta + k2\pi)$ para algún entero k. (f es periódica)

Distribución lattice

$$Pr(\theta = v + \frac{2\pi r}{m}) = p_r \ r = 0, 1, \cdots, m-1$$

$$p_r \geq 0, \ \sum_{r=0}^{m-1} p_r = 1$$

- Los puntos $\frac{2\pi r}{m}$ son los vértices de un polígono de m lados inscrito en el círculo unidad.
- Si los pesos son $p_r = \frac{1}{m}$, entonces se le denomina distribución uniforme discretizada en m puntos.

Distribución uniforme

- A menudo usada como modelo nulo.
- Es la única distribución en el círculo la cual es invariante bajo rotación y reflexión.
- Su función de masa de probabilidad

$$f(\theta) = \frac{1}{2\phi}$$

Distribución von Mises

Definición

$$g(\theta;\mu,k) = \frac{1}{2\pi I_0(k)} e^{k\cos(\theta-\mu)}$$

- I_0 denota la función de Bessel modificada de primer tipo y orden 0,

$$I_0(k) = \frac{1}{2\pi}\int_0^{2\pi} e^{k\cos\theta}\, d\theta$$

$$I_0(k) = \sum_{r=0}^{\text{ínf}} \frac{1}{(r!)^2}\left(\frac{k}{2}\right)^{2r}$$

- El parámetro μ es la media de las direcciones
- k es el parámetro de concentración.

Distribución von Mises

Relación v,R,S_n^2 y parámetro de concentración

1. Generamos una muestra de tamaño 3000 para cada parámetro k (75 valores)
2. Calculamos para cada una de las muestras
   - la varianza definida de forma análoga que en el caso lineal con respecto a la distancia euclídea y a la distancia definida por la longitud de arco,
   - el parámetro de concentración $R = \sqrt{\bar{C}^2 + \bar{S}^2}$ y
   - la desviación circular estandard $\sqrt{2(1-\bar{R})}$.
3. Representamos la estimador local-lineal de cada una de las estimaciones anteriores frente el parámetro de von-Mises y la función 1/k.

Distribución von Mises

Relación v,R,S_n^2 y parámetro de concentración

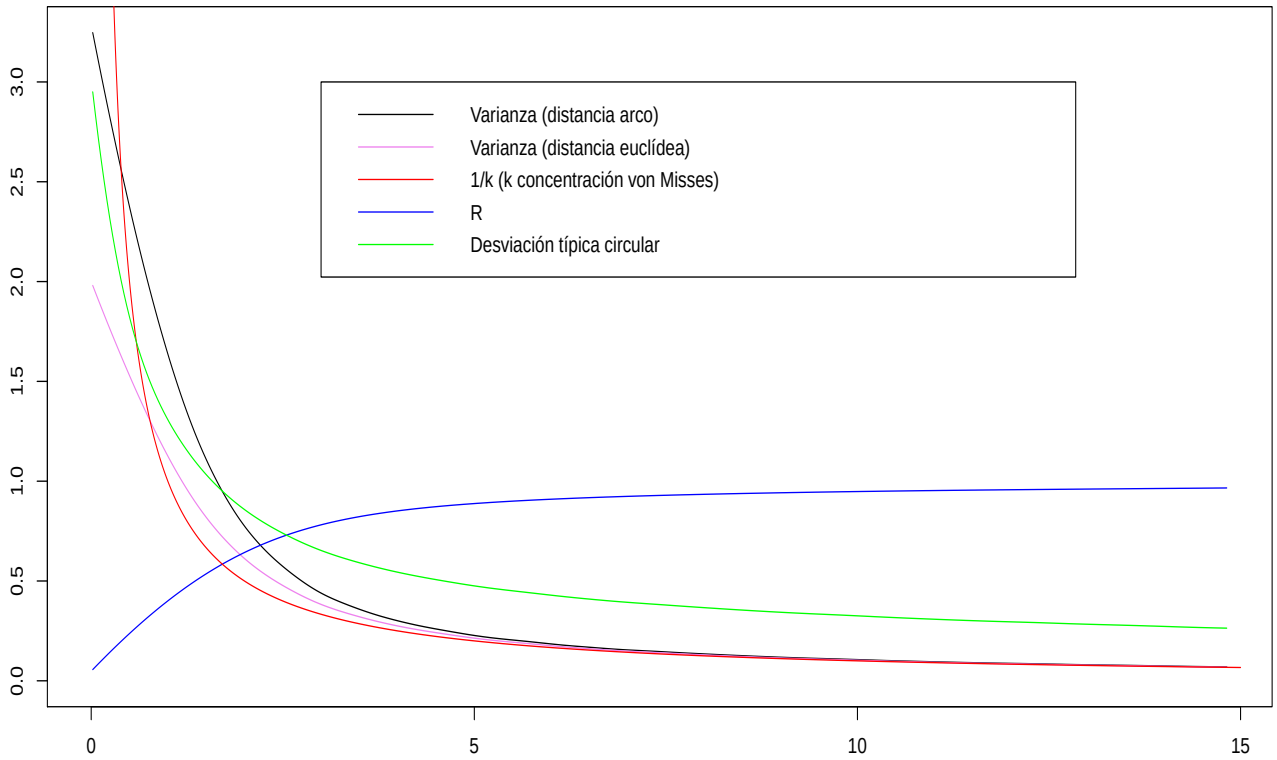
$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} d^2(\theta_i, \theta)$$

$$\theta = min_y \frac{1}{n}\sum_{i=1}^{n} d^2(\theta_i, y)$$

$$d^2(\theta_i, \theta j) = \pi - |\pi - |\theta_i - \theta_j||$$

$$d^2(\theta_i, \theta j) = 2(1 - (cos(\theta_i - \theta_j)))$$

Distribución von Mises



Distribución von Mises

Definición

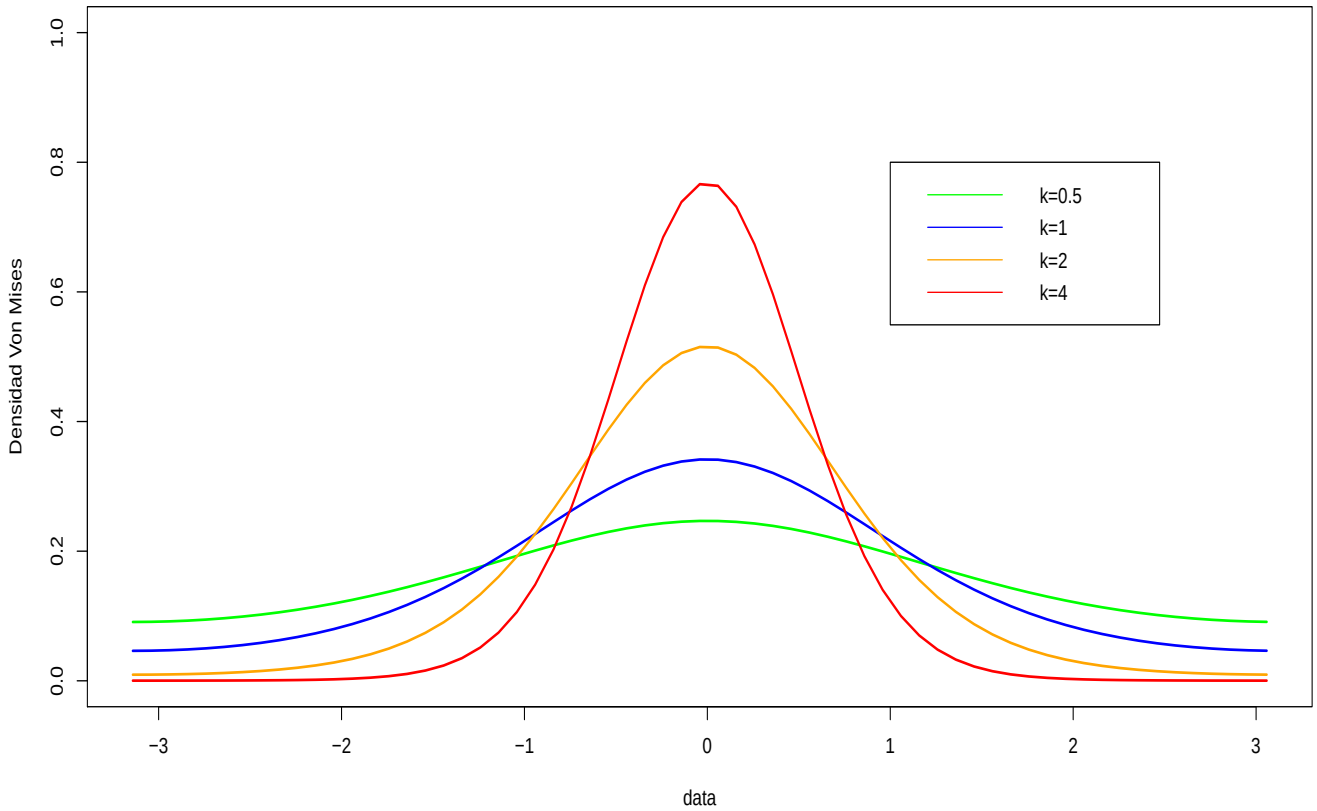
- Notemos que $M(\mu + \pi, k)$ y $M(\mu, -k)$ son la misma distribución.
- Para eliminar la indeterminación de los parámetros μ,k es usual tomar $k \geq 0$.

Distribución von Mises

Forma

- La distribución es unimodal y es simétrica sobre $\theta = \mu$
- La moda se encuentra en $\theta = \mu$ y la antimoda en $\theta = \mu + \pi$
- la relación de la moda de la densidad y la antimoda viene dada por e^{2k}, así que cuanto mayor sea el valor de k, mayor es el agrupamiento acerca de la moda.

Distribución von Mises



Distribución von Mises

Relación con otras distribuciones

- Si k=0, entonces $M(\mu, k)$ es la distribución uniforme.
- La aproximación $exp(x) \cong 1 + x$, muestra que para k pequeños $M(\mu, k) \cong C(\mu, k/2)$, dónde $C(\mu, k/2)$ denota una distribución Cardioid.
- Cualquier von Mises puede ser aproximada por una distribución normal wrapped.

$$M(\mu, k) \cong WN(\mu, A(k)), \ k \to \infty$$
$$A_k = I_1(k)/I_0(k)$$

Distribución von Mises

Función de distribución

$$F(\theta; 0, k) = \frac{1}{I_0(k)} \int_0^{\theta} e^{kcosu} du$$

Distribución von Mises

Convolución

- La convolución de dos von Mises no es una distribución von Mises.
- La convolución de dos distribuciones normal wrapped, $WN(\mu_1, A(k_1))$ y $WN(\mu_2, A(k_2))$ es la distribución wrapped normal $WN(\mu_1 + \mu_2, A(k_1)A(k_2))$. La cual puede ser aproximada por$M(\mu_1 + \mu_2, A^{-1}(A(k_1)A(k_2)))$

$$\theta_1 + \theta_2 \sim (M(\mu_1 + \mu_2, A^{-1}(A(k_1)A(k_2)))$$

Distribución Cardioid

- Perturbación de la densidad uniforme por la función coseno da lugar a una distribución Cardioid $C(\mu, \rho)$
- Función de masa de probabilidad

$$f(\theta) = \frac{1}{2\pi}(1 + 2\rho cos(\theta - \mu)) \ |\rho| < \frac{1}{2}$$

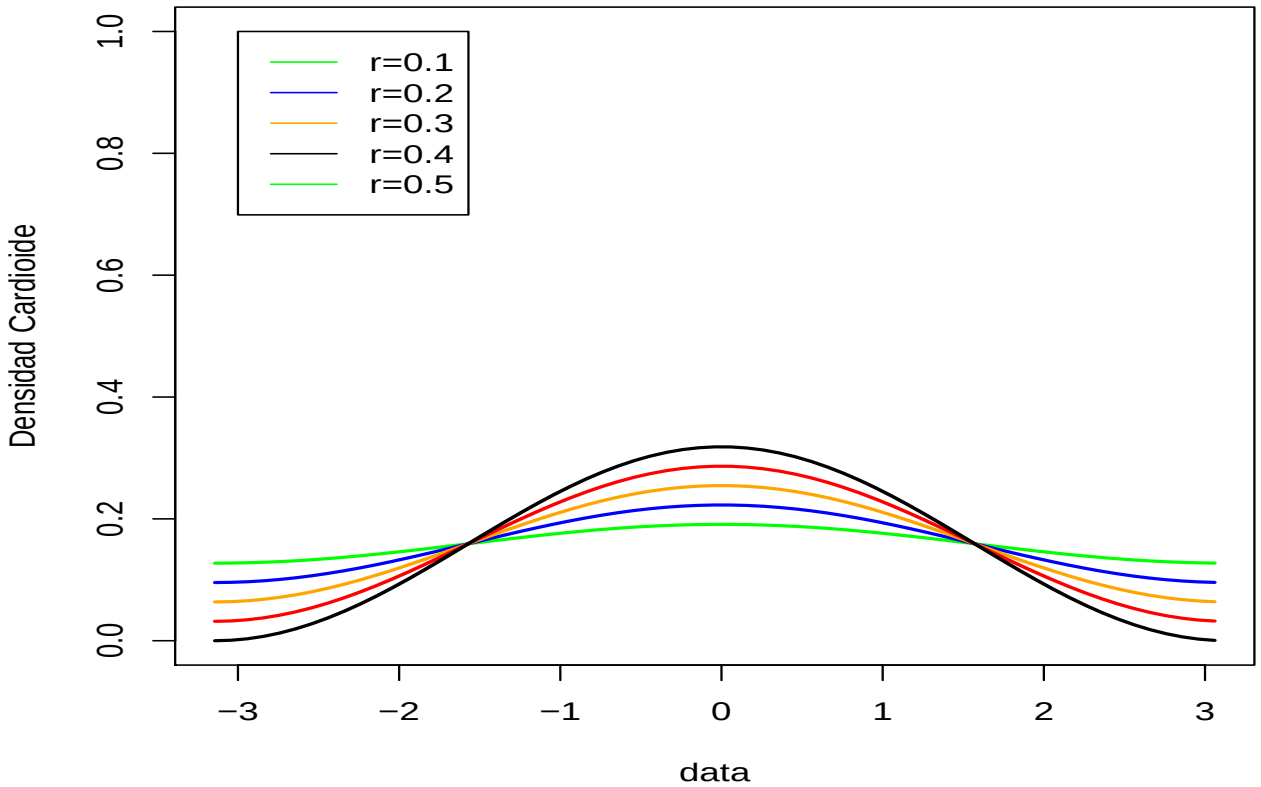
- La media de longitud resultante es ρ
- La media de la dirección es μ
- La distribución es simétrica y unimodal en μ (si $\rho > 0$).
- Si $\rho = 0$ la distribución se reduce a la distribución uniforme.
- El principal uso de estas distribuciones es como aproximaciones de poca concentración a las distribuciones von Mises

Distribución Cardioid

$$\theta_i \sim C(\mu_i, \rho_i)(i = 1, 2) \Rightarrow \theta_1 + \theta_2 \sim C(\mu_1 + \mu_2, \rho_2\rho_2)$$

El conjunto de distribuciones Cardioids es cerrado bajo convolución

Distribución Cardioid



Distribución Normal proyectada

- Distribuciones en el ciclo pueden seer obtenidas por proyección radial de la distribución radial en el plano.
- Sea x un vector aleatorio bidimensional, Pr(x=0)=0. Luego $||x||^{-1}x$ es un punto aleatorio sobre el círculo unidad.
- Un ejemplo importante es aquel en el que x tiene una distribución normal bivariante $N_2(\mu, \sum)$, en cuyo caso $||x||^{-1}x$ se dice que sigue una distribución normal proyectada $PN_2(\mu, \sum)$.

Distribución Normal proyectada

Aplicación

Meteorología: cuando la velocidad del viento es modelada por una distribución normal bivariante, la distribución marginal resultante para la dirección del viento es una distribución normal proyectada.

Distribución Normal proyectada

$$p(\theta;\mu,\sum) = \frac{\phi(\theta;0,\sum) + |\sum|^{-1/2} D(\theta)\phi(D(\theta))\phi(|\sum|^{-1/2}(x^T\sum^{-1}x)^{-1/2}\mu \wedge x)}{x^T\sum^{-1}x}$$

- $\phi(\cdot;0,\sum)$ denota la función de masa de probabilidad, ϕ y Φ denotan la función de masa de probabilidad y la función de densidad acumulada de N(0,1), $x = (cos\theta, sen\theta)^T$

$$D(\theta) = \frac{\mu^T\sum^{-1}x}{(x^T\sum^{-1}x)^{1/2}}$$

$$\mu \wedge x = \mu_1 sen\theta - \mu_2 cos\theta \ \ \mu = (\mu_1,\mu_2)^T$$

Distribución Wrapped

Definición

- Dada una distribución en la línea, se puede envolver alrededor de la circunferencia del círculo de radio uno.
- Si x es una variable aleatoria en la línea, la variable aleatoria correspondiente x_w de la distribución wrapped viene dada por

$$x_w = x(mod 2\pi)$$

- Si el círculo es identificado con el conjunto de números complejos de módulo la unidad luego el mapa wraping $x \to x_w$ pueden ser escrito como

$$x \to e^{2\pi i x}$$

Distribución Wrapped

Definición

- Si x tiene como función de distribución F $\Rightarrow$ la función de distribución F_w de x_w viene dada por

$$F_w(\theta) = \sum_{k=-\infty}^{\infty} \{F(\theta + 2\pi k) - F(2\pi k)\}$$

- Si x tiene como función de masa de probilidad f $\Rightarrow$ la función de masa de probabilidad f_w de x_w es

$$f_w(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2\pi k)$$

Distribución Wrapped

Propiedades

1. $(x+y)_w = x_w + y_w$
2. Si la función característica de x es ϕ entonces la función característica $\{\phi_p : p = 0, \pm 1, \ldots\}$ de x_w es dada por $\phi_p = \phi(p)$
3. Si ϕ es integrable entonces x tiene función una densidad y

$$f_w(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2\pi k) = \frac{1}{2\pi}[1 + 2\sum_{p=1}^{\infty}(\alpha_p cosp\theta + \beta_p senp\theta)]$$

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

Distribución Wrapped

Propiedades

- Hay muchas distribuciones en las líneas las cuales pueden ser envueltas en cualquier distribución dada en el círculo.
- Sea g la función de masa de probabilidad de una distribución en el círculo y define una función de masa de probabilidad en la línea por

$$f(x) = p_r g(x)$$
$$2\pi r < x \leq 2\pi(r+1) \quad r = 0, \pm 1, \pm 2, \ldots,$$

- p_r son números no negativos tales que, $\sum_{r=-\infty}^{\infty} p_r = 1$.

Distribución Wrapped

Distribución Poisson Wraped

- Así como módulo de reducción de 2π ajusta la línea en el círculo.
- Reducción módulo $2\pi m$ (m entero positivo) ajusta los enteros en el grupo de las m-raíces de uno, considerado como un subgrupo del ciclo.
- x variable aleatoria en los enteros, luego x_w

$$x_w = 2\pi x (mod 2\pi m)$$

- es una variable lattice $\{2\pi r/m : 0, 1, \ldots, m-1\}$ en el círculo.

Distribución Wrapped

Distribución Poisson Wraped

- La función de probabilidad de x_w es dada por

$$Pr(x_w = \frac{2\pi r}{m}) = \sum_{k=0}^{\infty} p(r + km) \; r = 0, 1, \ldots, m-1,$$

dónde p es la función de probabilidad de x.

- Si x sigue una Poisson con media λ, luego x_w tiene distribución Poisson wrapped con función de probabilidad

$$Pr(x_w = \frac{2\pi r}{m}) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{r+km}}{(r + km)!} \; r = 0, 1, \ldots, m-1,$$

Distribución Wrapped

Distribución Poisson Wraped

- Ball y Blackwel (1992) han escrito estas probabilidades de forma finita

$$Pr(x_w = \frac{2\pi r}{m}) = \frac{1}{m} \sum_{k=0}^{\infty} e^{w^j \lambda} w^{-rj}$$

- w es una raiz m-ésima raiz de uno
- Función caraterística de θ es $\phi_p = exp\left\{\lambda(1 - e^{2\pi ip/m})\right\}$
- La convolución de distribuciones Poisson wrapped de parámetro λ_1 y λ_2 es una distribución Poisson wrapped de parámetro $\lambda_1 + \lambda_2$

Distribución Wrapped

Distribución Normal Wrapped

- $WN(\mu, \rho)$ es obtenido mediante "wrapping"la distribución $WN(\mu, \sigma^2)$ en el círculo

$$\sigma^2 = -2log\rho$$
$$\rho = e^{-\sigma^2/2}$$

- Función de masa de probabilidad

$$\phi_w(\theta; \mu, \rho) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} exp\frac{-(\theta - \mu + 2\pi k)^2}{2\sigma^2}$$

$$\phi_w(\theta; \mu, \rho) = \frac{1}{2\pi}\left\{1 + 2\sum_{p=1}^{\infty} \rho^{p^2} cosp(\theta - \mu)\right\} \quad 0 \leq \rho \leq 1$$

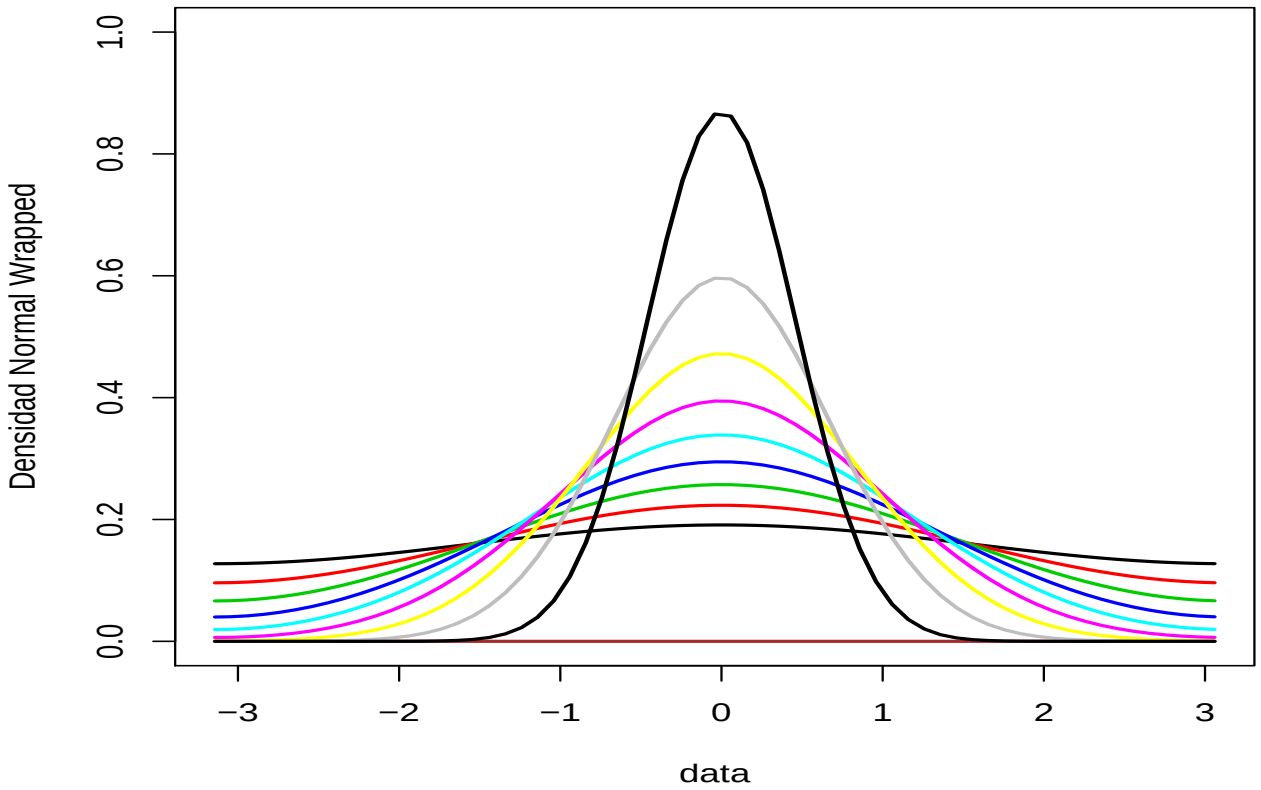
Distribución Wrapped

Distribución Normal Wrapped

- $WN(\mu, \rho)$ es unimodal y asimétrica sobre su moda μ.
- $\rho \to 0$, $WN(\mu, \rho)$ tiende a la distribución uniforme, mientras $\rho \to 1$, tiende a una distribución puntual en μ.

Distribución Wrapped

Distribución Normal Wrapped



Distribución Wrapped

Distribución Cauchy Wrapped

- Distribución en la línea real con densidad

$$f(x;\mu,a)=\frac{1}{\pi}\frac{a}{a^2+(x-\mu)^2},\ -\infty<\mu<\infty,\ a>0$$

- Dendidad de la distribución Cauchy Wrapped

$$c(\theta;\mu,\rho)=\sum_{k=-\infty}^{\infty}f(\theta+2\pi k;\mu,a)=\frac{1}{2\pi}\left\{1+2\sum_{p=1}^{\infty}\rho^p cosp(\theta-\mu)\right\}$$

- $\rho=e^{-a}$

Distribución Wrapped

Distribución Cauchy Wrapped

- De considerar la parte real de la serie geométrica se sigue que $\sum_{p=1}^{\infty} \rho^p e^{-ip(\theta-\mu)}$ se sigue que

$$c(\theta;\mu,\rho) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho cos(\theta-\mu)}$$

- La media de la dirección es $\mu(mod 2\pi)$ y la media de longitud resultante es ρ.
- Denotamos esta distribución por $WC(\mu,\rho)$

Distribución Wrapped

Distribución Cauchy Wrapped

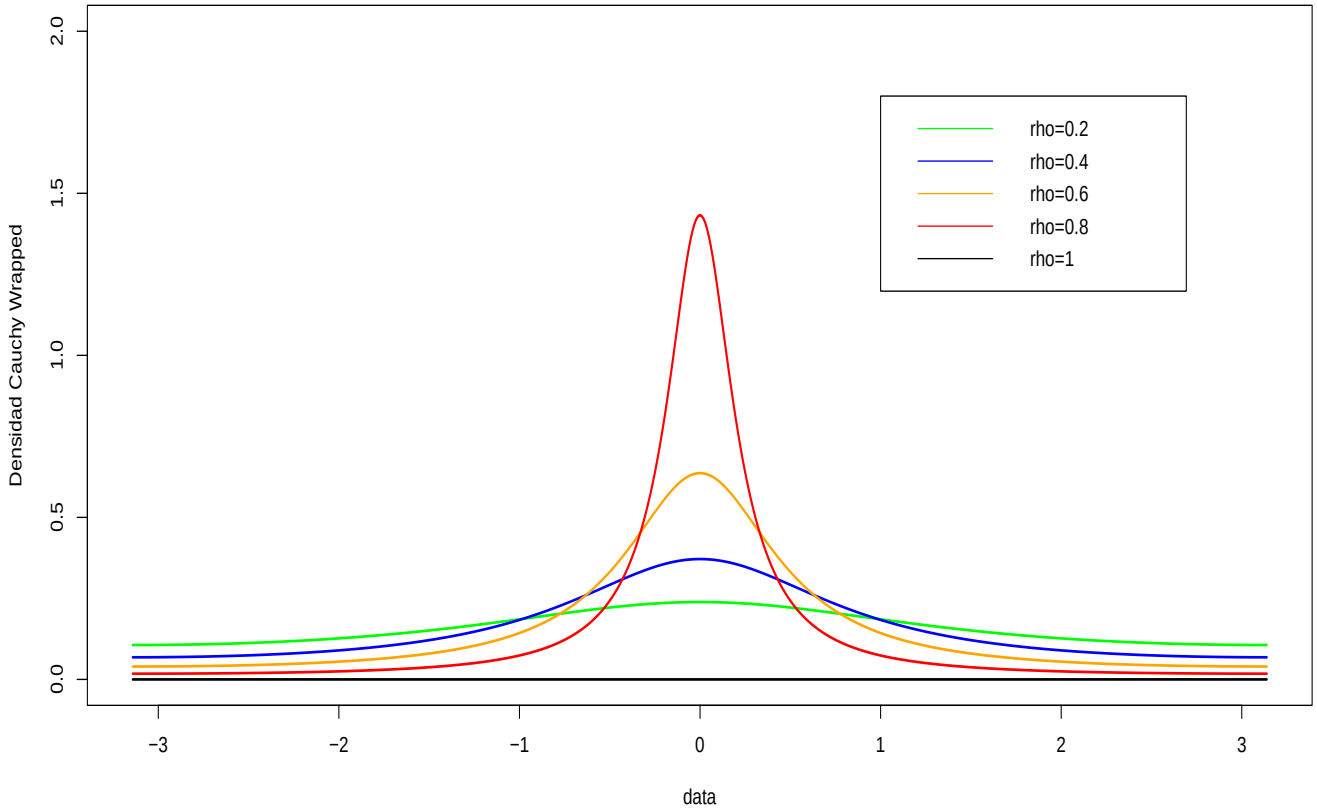
- Cuando $\rho \longrightarrow 0$ tiende a la distribución uniforme.
- Cuando $\rho \longrightarrow 1$, está concentrada en μ.
- Función de distribución

$$F(\theta) - F(\mu) = \frac{1}{2\pi} cos^{-1} \left\{ \frac{(1+\rho^2)cos(\theta-\mu) - 2\rho}{1+\rho^2 - 2\rho cos(\theta-\mu)} \right\}$$
$$\mu \leq \theta \leq \mu + \pi$$

- La connvolución de las distribuciones Cauchy Wrapped $WC(\mu_1, \rho_1)$ y $WC(\mu_2, \rho_2)$ es la distribución Cauchy Wrapped $WC(\mu_1 + \mu_2, \rho_1\rho_2)$

Distribución Wrapped

Distribución Cauchy Wrapped



Distribución Wrapped

Distribución Cauchy Wrapped

$$\theta \sim PN(0, \sigma) \Rightarrow 2\theta \sim WC(\mu, \rho)$$

$$\rho^2 = \frac{tr(\sigma) - 2|\sigma|^{1/2}}{tr(\sigma) + 2|\sigma|^{1/2}}$$

$$\sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Test de Rayleigh

- $E[(cos\theta), sen(\theta)^T] = 0$ cuando θ sigue una distribución uniforme, es razonable rechazar la uniformidad cuando el vector muestral de medias $(\bar{C}, \bar{S})$ está lejos de cero, es decir cuando $\bar{R}$ es grande.

Estadístico

Es útil tomar para el test el siguiente estadístico $2n\bar{R}^2$. El cual bajo uniformidad

$$2n\bar{R}^2 \sim \chi^2_2$$

Test de Rayleigh

Dirección media dada

- Se rechaza la uniformidad para valores grandes de

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

- Bajo uniformidad

$$2n\bar{C}^2 \sim \chi^2_1$$

Test de Kuiper

Test para la distribución linear están basados en la desviación entre la distribución empírica y la hipotética función de distribución.

Distribución

- Las observaciones ordenadas $\theta_1, \ldots, \theta_n$ aumentan con $\theta_0 = 0$ y $\theta_{n+1} = 2\pi$
- S_n es definida por

$$S_n(\theta) = i/n \texttt{ Si } \theta_i \leq \theta < \theta_n \; i = 0, 1, \ldots, n$$

Test de Kuiper

- $S_n(\theta)$ y $F(\theta)$. Es útil

$$U_i = \frac{\theta_i}{2\pi}, \ i = 1, \ldots, n+1$$

- Considerando el estadístico de Kolmogorov-Smirnov, sugiere que la distribución círcular sea rechazada para grandes valores de $max(D_n^+, D_n^-)$, dónde

$$D_n^+ = sup_\theta(S_n(\theta) - F(\theta))$$
$$D_n^- = sup_\theta(F(\theta) - S_n(\theta))$$

- $F(\theta) = \theta/2\pi$
- Kuiper(1990) $V_n = D_n^+ + D_n^-$

Test de Kuiper

- Kuiper(1990)

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

$$V_n^* = n^{1/2} V_n (1 + \frac{0.155}{\sqrt{n}} + \frac{0.24}{n})$$

- La distribución bajo uniformidad de V_n^* varía muy poco con n $n \geq 8$.

Test de U^2 de Watson

- Estadístico de Watson

$$U^2 = \sum_{i=1}^{n} \left[U_i - \bar{U} - \frac{i - 1/2}{n} + \frac{1}{2} \right]^2 + \frac{1}{12n}$$

$$\bar{U} = (U_1 + \ldots, U_n)/n$$

$$U^2 = \sum_{i=1}^{n} U_i^2 - n\bar{U} - \frac{2}{n} \sum_{i=1}^{n} iU_i + (n+1)\bar{U} + \frac{n}{12}$$

- El estadístico Cramer-von Mises

$$W^2 = \sum_{i=1}^{n} \left[U_i - \frac{i - 1/2}{n} \right]^2 + \frac{1}{12n}$$

Test de U^2 de Watson

- El estadístico Cramer-von Mises

$$W^2 = \sum_{i=1}^{n} \left[U_i - \frac{i - 1/2}{n} \right]^2 + \frac{1}{12n}$$

- Estadístico modificado

$$U^{*2} = (U^2 - \frac{0.1}{n} - \frac{0.1}{n^2})(1 + \frac{0.8}{n})$$

- La distribución de este estadístico bajo uniformidad varía muy poco con n

Test de U^2 de Watson

- Distribución asintótica de U^2

$$lim_{n\longrightarrow\infty} Pr(U^2 > u) = \sum_{m=1}^{\infty} (-1)^{m-1} e^{-2m^2\pi^2 u}$$

- Esta serie converge rápidamente y puede ser calculada rapidamente.
- Watson(1961) muestra que una muestra de gran tamaño la distribución asintótica de U^2 es la misma que la de V_n^2/π^2

$$W_n(t) = n^{1/2} S_n(t) - F(t)$$

- Dónde $F(t) = t/2\pi$

Test de Hodges-Ajne

- Una estrategia para testar la uniformidad de una distribución en el círculo es que se rechace la uniformidad si el número de puntos observados de una muestra los cuales caen en un conjunto adecuado difieren de un número esperado.
- Dos medidas adecuadas de la discrepancia entre lo esperado y lo observado
  - La diferencia máxima
  - La media de diferencias al cuadrado la diferencia como el conjunto varía en una determinada clase. (Estadístico A_n de Ajne)

Test de Hodges-Ajne

- Sea $N(\theta)$ el número de observaciones en el semiciclo centrado en θ, es decir, en el arco $(\theta - \pi/2, \theta + \pi/2)$
- El test rechaza la uniformidad para valores grandes de $max_\theta N(\theta)$ o para valores pequeños de $m = min_\theta N(\theta)$
- Test introducido por Ajne(1968), usando m-n, el máximo númerode observaciones en un semicírculo.

$$Pr(m \leq t) = 2^{-n+1}(n - 2t)\binom{n}{t}, \ t < \frac{n}{3}$$

Test rango

- w es la la longitud del menor arco que contiene a todas las observaciones

$$w = 2\pi - T \;\; T = max_{1 \leq i \leq n} T_i$$

$$T_i = \theta_i - \theta_{i-1}, \;\; i = 1, \ldots, n-1, \;\; T_n = 2\pi - (\theta_{(n)} - \theta_{(1)})$$

- La hipótesis de uniformidad es rechazada para valores pequeños de w.
- Función de distribución de el rango circular

$$Pr(w \leq r) = \sum_{\infty}^{k=1} \binom{n}{k} \left[1 - k(1 - \frac{r}{2\pi})\right]^{n-1}$$

- $1 - k(1 - \frac{r}{2\pi}) > 0$

Test rango

- Bajo uniformidad $E[T_i] = 2\pi/n$. Por tanto esrazonable rechazar la uniformidad para valores grandes de $L = \frac{1}{2}\sum_{i=1}^{n} |T_i - \frac{2\pi}{n}|$
- Shermam (1950)

$$\frac{n^{1/2}(L - 2\pi/e)}{2\pi(2e^{-1} - 5e^{-2})^{1/2}} \sim N(0,1)$$

Test Ajne An

- Sea $N(\theta)$ el número de observaciones en el semicírculo centrado en θ.
- El valor esperado bajo uniformidad es n/2. Se rechaza la uniformidad para grandes valores de

$$A_n = \frac{1}{2\pi n}\int_0^{2\pi}\left\{N(\theta) - \frac{n}{2}\right\}^2 d\theta \; A_n = n\left(\frac{1}{4} - \frac{\bar{D}_0}{2\pi}\right)$$

$$\bar{D}_0 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{\pi - |(\pi - |\theta_i - \theta_j|)|\}$$

- La distribución de A_n bajo uniformidad es

$$lim_{n\longrightarrow} Pr(A_n > \alpha) = \frac{4}{\pi}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{2k-1} exp\left(-\frac{(2k-1)^2\pi^2\alpha}{2}\right)$$

Test Hermans-Rasson

- Se rechaza la hipótesis de uniformidad para grandes valores de

$$H_n = 2\pi A_n - 2.895 \frac{1}{n} \sum_{j=2}^{n} \sum_{i=1}^{j-1} \left(|sen(\theta_i - \theta_j)| - \frac{2}{\pi} \right)$$

Definición

El vector unitario aleatorio $x = (cos\theta, sin\theta)^T$ tiene distribición von Mises con media μ y concentración $k \geq 0$. Si su densidad es

$$f(x) = [2\pi I_0(k)]^{-1} exp(x^T \alpha)$$

dónde el vector de concentración $\alpha = k(cos\mu, sin\mu)^T$ y $I_v(\cdot)$ función de Bessel modificada de primer tipo y orden v. En este caso podemos escribir

$$x \sim VM(\alpha)$$

El proceso autorregresivo de von Mises

- Breckling(1989), el proceso autorregresivo (AR) de von Mises.
- $\{\theta_t\}$, es un proceso direccional. Si la distribución condicionada de θ_t, dados $(\theta_{t-1}, \ldots, \theta_{t-p})$ es von Mises con vector de concentración $v_t = \sum_{j=1}^{p} k_j (cos\theta_{t-j}, sin\theta_{t-j})^T + (k_0, 0)^T$
- Luego θ_t es un proceso de von Mises

$$f(\theta_t|\theta_{t-1}, \ldots, \theta_{t-p}) = [2\pi I_0(u_t)]^{-1}$$

$$exp\left\{\sum_{j=1}^{p} k_j cos(\theta_t - \theta_{t-j}) + k_0 cos\theta_t\right\}$$

$$u_t = ||v_t||$$

El proceso Linked

- Fisher y Lee (1994), usan la idea de funciones *link* entre los modelos de series de tiempo lineales a un contexto direccional.
- Proceso estacionario direccional $\{\theta_t\}$ con media direccional con media μ es un proceso "link.ªuntoregresivo de medias mobiles (LARMA); si y solo si el proceso $\{x_t\} = \{g^{-1}(\theta_t - \mu)\}$ es un proceso auntoregresivo de medias mobiles (ARMA)
- Dónde g es par , es una función monótona$(-\pi, \pi]$ y g(0)=0
- k-lag función circular de autocorrelación de $\{\theta_t\}$ es dado como

$$\rho_T(k) = \rho_T\{g(X_t), g(X_{t+k})\}$$

$$\rho_T = \frac{E[sen(\Theta_1 - \Theta_2)sen(\Phi_1 - \Phi_2)]}{\{E[sen^2(\Theta_1 - \Theta_2)]E[sen^2(\Phi_1 - \Phi_2)]\}^{1/2}}$$

Proceso autorregresivo wrapped (WAR)

- Dado el proceso lineal AR(p)

$$X_t = \sum_{j=1}^{p} \alpha_j X_{t-j} + \epsilon_t$$

$$\epsilon_t \sim N(0, \sigma^2) iid$$

- El proceso WAR(p) ν_t es definido por $\nu_t = X_t (mod 2\pi)$, $t = 1, \ldots, n$
- Si ν_t es la serie observada, el proceso lineal viene dado por $X_t = \nu_t + 2\pi k_t$. k_t coeficientes wrapping.
- La complejidad de inferencia surge del hecho de que necesitamos estimar los parámetros $\alpha_1, \ldots, \alpha_p, \sigma^2$ de el proceso lineal, dado solo el proceso wrappped.

El modelo seno de series de tiempo

- La densidad conjunta de dos variables circulares aleatorias θ_1, θ_2 del modelo del seno es dada por

$$f(\theta_1, \theta_2) = C exp\{k_1 cos(\theta_1 - \mu_1) + k_2 cos(\theta_2 - \mu_2) + \lambda sen(\theta_1 - \mu_1) sen(\theta_2 - \mu_2)$$

$$-\pi \leq \theta_1, \theta_2 \leq \pi$$

$$k_1, k_2 \geq 0$$

$$-\infty < \lambda < \infty$$

$$-\pi \leq \mu_1, \mu_2 \leq \pi$$

$$C^{-1} = 4\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} \left(\frac{\lambda^2}{4k_1 k_2}\right)^m I_m(k_1) I_m(k_2)$$

El modelo seno de series de tiempo

- Dada la distribución condicional de θ_2 dado θ_1 como von Mises con densidad

$$f(\theta_1|\theta_2) = [2\pi I_0(a_1)]^{-1} e^{k_2 cos(\theta_2-\mu_2)+\lambda sin(\theta_1-\mu_1)sen(\theta_2-\mu_2)}$$

$$a_1 = k_2^2 + \lambda^2 sen^2(\theta_1 - \mu_1)^{1/2}$$

- Para formular un modelo de series de tiempo reemplazamos θ_2 por θ_t y θ_1 por θ_{t-1} y usamos la densidad condicional de θ_t dado θ_{t-1}.

El modelo seno de series de tiempo

- Dada la distribución condicional de θ_t dado θ_{t-1} como von Mises con densidad

$$f(\theta_t|\theta_{t-1}) = [2\pi I_0(k_t)]^{-1} e^{k\cos(\theta_t-\mu)+\lambda \sin(\theta_t-\mu)\sin(\theta_{t-1}-\mu)}$$

$$k_t = \left\{k^2 + \lambda^2 \operatorname{sen}(\theta_{t-1}-\mu)\right\}^{1/2}$$

$$\mu_t = \mu + \tan^{-1}\left\{\frac{\lambda}{k}\operatorname{sen}(\theta_{t-1}-\mu)\right\} \quad t = 2,\ldots,n$$

- Como un proceso circular AR(1), el modelo puede ser equivalentemente expresado como

$$\theta_t = \mu + \tan^{-1}\left\{\frac{\lambda}{k}\operatorname{sen}(\theta_{t-1}-\mu)\right\} + \epsilon_t$$

$$\epsilon_t \sim M(0, k_t) \; t = 2,\ldots,n$$

El modelo coseno de series de tiempo

- Modelamos el vector unitario $x_t = (cos\theta_t, sen\theta_t)^T$ condicionado a x_{t-1} como von Mises

$$x_t|x_{t-1} \sim VM(ax_{t-1} + b\mathbf{e})$$

- $a, b > 0$ escalares, $\mathbf{e} = (cos\mu, sen\mu)^T$

$$f(x_t|x_{t-1}) = [2\pi I_0(||ax_{t-1} + b\mathbf{e}||)]^{-1} e^{x_t^T(ax_{t-1}+b\mathbf{e})}$$

El modelo coseno de series de tiempo

$$\theta_t \sim M(\mu_t, k_t)$$

$$\mu_t = tan{-}1(asen\theta_{t-1} + bsen\mu, acos\theta_{t-1} + bcos\mu)$$

$$k_t = \sqrt{a^2 + b^2 + 2abcos(\theta_{t-1} - \mu)}$$

$$tan^{-1}(p, q) \in [-\pi, \pi)$$

- Es el ángulo entre el eje positivo x y el vector (p,q)

$$f(\theta_t|\theta_{t-1}) = \left[2\pi I_0(\sqrt{a^2 + b^2 + 2abcos(\theta_{t-1} - \mu)})\right]^{-1} e^{(acos(\theta_t - \theta_{t-1}) + bcos(\theta_{t-1} - \mu))}$$

El modelo Möbius de series de tiempo

- La componente determinística del modelo de regresión estudiado por Downs y Mardia(2002) une a la variable angular dependiente v con la variable angular independiente u mediante

$$tan\frac{1}{2}(v-\beta) = \omega tan\frac{1}{2}(u-\alpha)$$

$$v = \beta + 2tan^{-1}\left\{\omega tan\frac{1}{2}(u-\alpha)\right\}$$

- $\omega \in [-1,1]$ es el parámetro pendiente.
- $-\pi \leq \alpha, \beta < \pi$ son los parámetros angulares de localización.

El modelo Möbius de series de tiempo

- Reemplazamos el ángulo v por θ_t y el ángulo u por θ_{t-1}, $t = 2, \ldots, n$. $\alpha = \beta$

$$tan\frac{1}{2}(\theta_t - \alpha) = \omega tan\frac{1}{2}(\theta_{t-1} - \alpha)$$

$$\theta_t = \alpha + 2tan^{-1}\left\{\omega tan\frac{1}{2}(\theta_{t-1} - \alpha)\right\}$$

$$\theta_t|\theta_{t-1} \sim M(\alpha + 2tan^{-1}\left\{\omega tan\frac{1}{2}(\theta_{t-1} - \alpha)\right\}, k)$$

El modelo Möbius de series de tiempo

- Por tanto el modelo de serie de tiempo se convierte en

$$\theta_t = \alpha + 2tan^{-1}\left\{\omega tan\frac{1}{2}(\theta_{t-1} - \alpha)\right\} + \epsilon_t$$

$$\epsilon_t \sim M(0, k)$$

$$\mu_t = \alpha + 2tan^{-1}\left\{\omega tan\frac{1}{2}(\theta_{t-1} - \alpha)\right\}$$

Datos direccionales.

Ánalisis descriptivo.

Leyenda Rodríguez, María

17 de febrero de 2011