Procesamiento y Análisis de Datos Astronómicos

13.- Distribuciones en dos Dimensiones II

R. Gil-Hutton

Marzo 2020

Actividades:

Práctica 11:

- Utilizando la descomposición en Componentes Principales obtenida haga un agrupamiento jerárquico de los objetos. Indique cuál sería un valor máximo para definir grupos.
- Simule una muestra limitada por flujo (luminosidad, magnitud absoluta) de objetos distribuídos en un gran volumen de espacio utilizando para el flujo una distribución en ley de potencias del tipo:

$$\rho(> f) \propto f^{-\gamma}$$

  y una distribución volumétrica uniforme para la distancia. Para armar la muestra fije un valor límite para el flujo de manera arbitraria.

Actividades:

Práctica 11:

- Con esta muestra:
  - aplique el método V_{max} y vea si es posible recuperar la distribución de flujo utilizada.
  - produzca errores utilizando bootstrap y compare los resultados con errores basados en $\sqrt{N}$.
  - asigne a cada objeto dos flujos diferentes, detecte y utilice el método de Kaplan-Meier para normalizar la distribución. Aparece alguna correlación?.

Actividades:

- Una distribución con una ley de potencia del tipo $\rho(> f) \propto f^{-\gamma}$ es una distribución acumulativa porque $\rho(> f)$ corresponde a todos los elementos con valores mayores o iguales a f.
- Supongamos una distribución de luminosidad con valores en el rango $[1, 10]$ y $\gamma = 2$. Voy a usar un histograma de 0,25 para cada bin.
- La constante de proporcionalidad se determina asumiendo que para $\rho(> 10) = 1$.

```
In [2]: lum=np.linspace(10,1,37)

In [3]: kk=1./10**(-2)

In [4]: rho=kk*lum**(-2)

In [5]: rho=rho/rho[-1]
```

Actividades:

- Ahora extraigo al azar las luminosidades de una población simulada de 20000 objetos. En el array **val** voy a guardar en cada columna luminosidad, distancia y flujo observado:

```
In [6]: rr=np.random.random(20000)

In [7]: val=np.zeros((len(rr),3))

In [8]: for ii in range(len(rr)):
   ...:     jj=0
   ...:     while(rr[ii] > rho[jj]):
   ...:         jj+=1
   ...:     val[ii,0]=lum[jj]+(np.random.random()-0.5)*0.25
   ...:
```

Actividades:

- Asumo un volumen total de 1000 unidades cúbicas, lo divido en 100 cáscaras de igual volumen y distribuyo los objetos con probabilidad uniforme:

```
In [9]: rad=np.zeros(100)

In [10]: rad[0]=10.**(1/3)

In [11]: for ii in range(1,100):
    ...:     rad[ii]=(10.+rad[ii-1]**3)**(1/3)
    ...:

In [12]: bb=np.random.randint(0,100,size=20000)

In [13]: for ii in range(20000):
    ...:     if(bb[ii] != 0):
    ...:         val[:,1]=(rad[bb]-rad[bb-1])*np.random.random()+rad[bb-1]
    ...:     else:
    ...:         val[:,1]=rad[bb]*np.random.random()
    ...:
```

Actividades:

- Calculo el flujo observable, fijo un límite inferior de detección y extraigo la muestra:

```
In [14]: val[:,2]=val[:,0]/val[:,1]**2

In [15]: np.min(val[:,2]),np.max(val[:,2])
Out[15]: (0.008825323755358806, 1.347497375858959)

In [16]: inx=np.where(val[:,2] > 0.05)

In [17]: np.shape(inx)
Out[17]: (1, 4315)

In [18]: mue=val[inx[0],:]
```

Función de correlación angular:

- Si se desea medir la magnitud de un cierto agrupamiento de objetos se necesita definir un estadístico que lo cuantifique.
- Como cualquier agrupamiento de objetos incrementa el número de pares cercanos, la función de correlación angular punto a punto cuantifica ese valor en función de la separación angular.
- Formalmente, es el incremento fraccional, relativo a una distribución de objetos al azar, en la probabilidad de encontrar objetos en dos ángulos sólidos separados por una cierta distancia angular.

Función de correlación angular:

- Si $\delta\Omega_1$ y $\delta\Omega_2$ son dos ángulos sólidos separados por una distancia angular θ, la probabilidad de encontrar objetos en ambos es:

$$\delta\mathcal{P} = \zeta^2[1 + \omega(\theta)]\delta\Omega_1\delta\Omega_2$$

  donde ζ es la densidad superficial de objetos y $\omega(\theta)$ es la función de correlación angular.
- $\omega(\theta)$ es un estadístico simple que es fácil de medir y de interpretar, y permite revelar efectos sistemáticos en datos observacionales.
- $\omega(\theta)$ es un estadístico conveniente para comparar datos con predicciones teóricas y entre diferentes muestras observacionales.

Función de correlación angular:

Para obtener un estimador de $\omega(\theta)$ para una muestra de n objetos tenemos que:

- medir las separaciones θ de todos los pares posibles.
- agrupar las separaciones en intervalos de clase para obtener el número de pares en cada uno de ellos, $DD(\theta)$.
- calcular el correspondiente número de pares en cada intervalo asumiendo una distribución al azar para el valor medio de la densidad superficial, $RR(\theta)$: para un intervalo centrado en θ y con ancho $\delta\theta$, $RR(\theta) = \frac{1}{2} n \zeta 2\pi\theta\delta\theta$.

Entonces, un estimador simple para $\omega(\theta)$ es:

$$\omega_0(\theta) = \frac{DD(\theta)}{RR(\theta)} - 1$$

Función de correlación angular:

- $\omega(\theta)$ presenta algunos problemas debido a su simpleza, como la dificultad para estimar su error en áreas pequeñas, sufre de errores de correlación entre bines cercanos y presenta también efectos de borde importantes para muestras con agrupamiento débil.
- Para encontrar un estimador mejorado se requiere medir el área media de cada bin ($\approx 2\pi\theta\delta\theta$) usando una integración de Monte Carlo, para generar una distribución de comparación al azar de r puntos sobre la misma área. El resultado es el estimador "natural" de Blake (2002):

$$\omega_1(\theta) = \frac{r(r-1)}{n(n-1)} \frac{DD(\theta)}{RR(\theta)} - 1$$

Función de correlación angular:

Como el estimador natural de Blake también tiene sus problemas se han propuesto otros estimadores. Si $DR(\theta)$ es el número de pares considerando objetos y valores al azar para cada intervalo, tenemos:.

- el estimador de Peebles (Davis and Peebles, Astrophys. J. **267**, 465, 1983):

$$\omega_2(\theta) = \frac{2r}{n(n-1)} \frac{DD(\theta)}{RR(\theta)} - 1$$

- el estimador de Landy y Szalay (Landy and Szalay, Astrophys. J. **412**, 64, 1993):

$$\omega_3(\theta) = \frac{r(r-1)}{n(n-1)} \frac{DD(\theta)}{RR(\theta)} - \frac{(r-1)}{n} \frac{DD(\theta)}{DR(\theta)} + 1$$

Función de correlación angular:

- el estimador de Hamilton (Hamilton, Astrophys. J. **417**, 19, 1993):

$$\omega_4(\theta) = \frac{4nr}{(n-1)(r-1)} \frac{DD(\theta)RR(\theta)}{DR(\theta)^2} - 1$$

Para obtener $RR(\theta)$ y $DR(\theta)$ ae pueden elegir dos formas:

- elegir un conjunto muy grande de datos al azar (tiempo de cálculo).
- elegir muchos conjuntos pequeños de datos al azar y promediar sus resultados (si el número de conjuntos es m con r elementos cada uno, una guía razonable es $r/n \sim 1$ y $m \gg 1$, preferentemente $m \geq 10$).

Función de correlación angular:

Hay dos efectos instrumentales que impactan seriamente en $\omega(\theta)$: errores de calibración en gran escala y alta resolución:

- Los errores de calibración en gran escala se producen por cambios en la densidad superficial. El número de pares de objetos depende de la densidad superficial local ($DD(\theta) \propto \bar{\zeta^2}$) pero el número de pares al azar depende de la densidad superficial global ($RR(\theta) \propto (\bar{\zeta})^2$), al ser $\bar{\zeta^2} > (\bar{\zeta})^2$ se incrementa $\omega(\theta)$ en:

$$\Delta\omega(\theta) = \frac{\bar{\zeta^2}}{(\bar{\zeta})^2} - 1 = \bar{\delta^2}$$

donde $\delta = (\zeta - \bar{\zeta})/\bar{\zeta}$ es la sobre densidad superficial.

Función de correlación angular:

- En el caso de alta resolución, si la resolución del telescopio es suficiente para separar los objetos en más de un elemento hay un peligro real de contaminar las mediciones de $\omega(\theta)$ por una contribución en exceso por pares aparentes debido a partes del mismo objeto.
- Este problema es mucho más serio en radio que en el óptico.
- Para poder evaluarlo se convierten algunos de los n objetos en objetos múltiples con un promedio de $\bar{c}$ componentes, tomando ángulos θ lo suficientemente pequeños para que la función esté dominada por pares próximos.

Función de correlación angular:

- Si e es la fracción de valores originales fragmentados en objetos múltiples, y $f(\theta)\delta\theta$ la fracción de pares con separaciones entre θ y $\theta + \delta\theta$, el número de pares con separaciones en ese intervalo es $nef(\theta)\delta\theta$.
- Como el número de pares esperados al azar es $(\bar{c})^2 n\zeta\pi\theta\delta\theta$, la función de correlación angular para ángulos pequeños estará desplazada por:

$$\Delta\omega(\theta) = \frac{ef(\theta)}{(\bar{c})^2\zeta\pi\theta}$$

Conteo en celdas:

- Una segunda técnica posible para cuantificar el agrupamiento de objetos es el conteo en celdas, que es una manera tradicional de hacerlo.
- Simplemente se divide el cielo en celdas de área y forma fija y se cuentan los objetos en cada celda, lo que provee la probabilidad $p(N)$ de encontrar N objetos en una celda. Si no hay agrupamiento $p(N)$ corresponde a una distribución de Poisson.
- Las propiedades del agrupamiento quedan completamente caracterizadas por esta distribución, la cual provee más información que la función de correlación angular.

Conteo en celdas:

- Si consideramos una distribución de objetos que no muestra agrupamiento, para una densidad superficial ζ el número medio de objetos esperados en una celda de área S es $< N >= \zeta \times S$, y la distribución de probabilidades $p(N)$ será una distribución de Poisson con media y varianza iguales a $< N >$.
- Una distribución de objetos agrupados producirá una varianza mayor a la de una distribución al azar debido a la presencia de celdas sin objetos junto a otras con grandes concentraciones. Para cuantificar esta propiedad se define el estimador:

$$y = \frac{\mu_2 - \bar{N}}{(\bar{N})^2}$$

donde μ_2 es el segundo momento de la distribución (varianza) y $\bar{N}$ es el valor medio de la muestra total.

Conteo en celdas:

- El tercer momento de la distribución μ_3 (sesgo) que es de importancia especial porque mide la asimetría del agrupamiento respecto de una distribución de Poisson indicando la presencia de colas con celdas de gran número de objetos.
- La medida del sesgo es de mucha importancia en el estudio de galaxias ya que cuantifica el grado de agrupamiento gravitatorio no lineal.
- Como en una distribución de Poisson la media, varianza y sesgo son iguales a $< N >$, el siguiente estadístico cuantifica el incremento en sesgo debido al agrupamiento:

$$z = \frac{\mu_3 - 3\mu_2 + 2\bar{N}}{(\bar{N})^3}$$

Conteo en celdas:

Al igual que en el caso de la función de correlación angular, el conteo en celdas sufre de errores de calibración en gran escala y alta resolución:

- Los errores de calibración en gran escala en este caso incrementan la varianza y el sesgo en:

$$\Delta y = \frac{\bar{\zeta^2}}{(\bar{\zeta})^2} - 1 = \bar{\delta^2}$$

$$\Delta z = \frac{\bar{\zeta^3}}{(\bar{\zeta})^3} - 1 = \bar{\delta^3}$$

donde δ es la sobre densidad superficial.

Conteo en celdas:

- En el caso de los errores por alta resolución, los desplazamientos para la varianza y sesgo son (Blake 2002):

$$\Delta y = \frac{1}{\zeta S}\left[\frac{2e+6f}{1+e+2f}\right]$$

$$\Delta z = \frac{1}{(\zeta S)^2}\left[\frac{6f}{1+e+2f}\right]$$

  donde S es el área de la celda, ζ es la densidad superficial, e la fracción de objetos dobles y f la fracción de objetos triples.

- Es bastante crítica la presencia de objetos múltiples para la estimación del sesgo. Si bien el sesgo es casi insensible a los objetos dobles, es al mismo tiempo muy sensible a los triples.

Espectro de potencia angular:

El espectro de potencia angular, c_l, asume que la densidad superficial de objetos se expresa como una suma de fluctuaciones de densidad angular con diferentes frecuencias.

- Se parte de un campo de densidad contínuo:

$$\zeta(\phi, \theta) = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} a_{l,m} \mathcal{Y}_{l,m}(\phi, \theta)$$

  donde $a_{l,m}$ son coeficientes, $\mathcal{Y}_{l,m}$ son armónicos esféricos, $0 \leq \theta \leq 2\pi$ y $-\pi/2 \leq \phi \leq \pi/2$.

- La probabilidad de encontrar un objeto en un elemento de ángulo sólido $\delta\Omega$ en la posición (θ, ϕ) es:

$$\delta p = \zeta(\phi, \theta)\delta\Omega$$

Espectro de potencia angular:

- En esta formulación, el espectro de potencia angular se define como:

$$|a_{l,m}|^2 = c_l$$

donde la asunción de isotropía asegura que $|a_{l,m}|^2$ es una función sólo de l y no de m.

- Tanto $a_{l,m}$ como $\mathcal{Y}_{l,m}$ son cantidades complejas y, debido a que $\zeta(\phi, \theta)$ es real, tenemos que $a_{l,-m} = a^*_{l,m}$.
- En este modelo para el campo de densidad las partes reales e imaginarias de $a_{l,m}$ se obtienen independientemente de distribuciones de tal modo que la normalización satisfaga $|a_{l,m}|^2 = c_l$.

Espectro de potencia angular:

- Si consideramos un survey de todo el cielo, como los $\mathcal{Y}_{l,m}$ son ortonormales es posible invertir las ecuaciones para obtener $a_{l,m}$:

$$a_{l,m} = \int \zeta(\phi, \theta) \mathcal{Y}^{*}_{l,m}(\phi, \theta) d\Omega$$

lo que sugiere que un estimador de $a_{l,m}$ es la suma de los armónicos esféricos en las n posiciones (θ_i, ϕ_i):

$$E(a_{l,m}) = \sum_{i=1}^{n} \mathcal{Y}^{*}_{l,m}(\phi_i, \theta_i)$$

y un estimador para c_l es:

$$E(c_{l,m}) = |E(a_{l,m})|^2 - \zeta_0 \qquad c_l = < E(c_{l,m}) >$$

donde ζ_0 es la densidad superficial media que se incluye para compensar por la distribución discreta.

- Si consideramos un survey de cielo parcial hay que hacer algunas modificaciones pero se termina obteniendo:

$$E(c_{l,m}) = \frac{|E(a_{l,m}) - \zeta_0 \mathcal{I}_{l,m}|^2}{\mathcal{J}_{l,m}} - \zeta_0$$

donde $\zeta_0 = n/\Delta\Omega$ siendo $\Delta\Omega$ el área del survey y:

$$\mathcal{I}_{l,m} = \int_{\Delta\Omega} \mathcal{Y}^*_{l,m} d\Omega \qquad \mathcal{J}_{l,m} = \int_{\Delta\Omega} |\mathcal{Y}_{l,m}|^2 d\Omega$$

Espectro de potencia angular:

- En teoría el espectro de potencia angular, c_l, es equivalente a la función de correlación angular, $\omega(\theta)$. Las dos cantidades están relacionadas por:

$$c_l = 2\pi\zeta_0^2 \int_{-1}^{+1} \omega(\theta)\mathcal{P}_l(\cos\theta)d(\cos\theta)$$

$$\omega(\theta) = \frac{1}{4\pi\zeta_0^2} \sum_{l=1}^{\infty} (2l+1)c_l\mathcal{P}_l(\cos\theta)$$

donde $\mathcal{P}_l$ son polinomios de Legendre.

Otras formas de análisis:

- Análisis topológico (Gott et al., Astrophys. J. **340**, 625, 1989).
- Árboles de extensión mínima (Barrow, Bhavsar and Sonada, MNRAS **216**, 17, 1985).
- Teoría de percolación (Dekel and West, Astrophys. J. **288**, 11, 1985).
- Funciones de correlación de alto orden (Peebles and Groth, Astrophys. J. **196**, 1, 1975).
- Análisis fractal (Martinez et al., Astrophys. J. **357**, 50, 1990).

Actividades:

Práctica 12

Para un conjunto de objetos con coordenadas (θ, ϕ):

```
( 54.43  39.01), ( 22.82  23.42), (162.06  -3.00), ( 52.76  41.10), (159.23  62.97),
(194.82  32.25), (195.20  44.26), (103.93  70.33), (103.95  16.42), (138.32 -19.43),
( 38.74  46.30), (113.11  12.45), ( 57.53 -36.07), ( 84.96 -39.78), ( 80.30  25.66),
( 57.29   2.97), ( 88.42   0.80), (129.03   9.28), (132.35   8.32), (-20.55  29.36),
(126.02   9.27), (169.62  64.72), ( 72.02 -12.41), (124.88  10.39), (148.71  16.71),
(112.04  89.31), (193.71  32.61), ( 60.11  13.35), (101.87   8.22), ( 51.07  23.63),
(127.61  -8.69), (128.83 -62.72), ( 69.74  15.08), ( 87.93  31.96), ( 80.57  44.37),
(204.89 -42.29), ( 57.63  55.36), ( 93.32 -18.69), (106.57  23.87), (164.41  38.46),
(162.02  24.01), ( 91.96  17.42), ( 41.53  36.09), (184.32 -27.34), ( 78.42  20.07),
(150.25 -11.29), (  5.11  52.38), ( 60.40  31.39), (170.58   9.92), ( 50.23  89.83),
(104.32  12.71), (105.64  84.09), (160.47  19.51), ( 99.75  -8.25), ( 79.83 -21.24),
(144.63  12.63), ( 60.65 -23.28), (127.11  42.92), ( 91.27  21.76), (123.97  21.09),
(105.32  -7.44), (176.37  52.37), ( 71.76 -14.17), ( 41.53  48.17), (  5.46  16.93),
( 74.95 -40.73), (134.38 -24.06), ( 96.58 -26.26), ( 61.35  19.02), (120.19  -4.08),
( 46.19  -9.34), (121.22 -13.26), ( 93.16 -12.93), (129.26  79.56), (  5.17   1.40),
(140.20  55.48), ( 93.90  37.84), ( 74.99 -46.07), ( 29.86 -27.06), (128.54   7.51),
( 12.61 -12.46), ( 81.03   5.04), ( 40.82   6.88), (116.76  24.85), (176.26  17.35),
(  1.61  38.21), (129.16 -15.80), (127.19  35.96), ( 14.13  20.20), ( 36.87  34.13),
( 85.16  27.67), (136.10  44.62), (101.88  52.10), (211.70  21.57), ( 48.06 -18.70),
( 46.25   8.72), (107.61  29.39), ( 25.37  30.88), (137.28  40.91), (189.28  -6.07)
```

Actividades:

Práctica 12

- Verifique si corresponden o no a una distribución uniforme.
- Calcule su mediana esférica, vector medio, estime sus errores y compruebe si hay simetría rotacional.
- Repita los ejercicios anteriores adicionando a las coordenadas errores al azar distribuídos según von Mises-Fisher con $\kappa = 2$. Compare los resultados.

Entrega

Para la próxima clase

Por consultas: ricardo.gil-hutton@conicet.gov.ar
Grupo de Ciencias Planetarias - CUIM 2