

Fecha de Entrega: **Mayo 6 de 2015 antes de las 23:59 COT**

Instrucciones de Entrega

Todo el código fuente y los datos se debe encontrar en un repositorio público en github con un commit final hecho antes de la fecha de entrega. El nombre del repositorio debe ser **CM20151_HW7_Apellido1Apellido2**. El link al repositorio lo deben enviar a través de **sicuplplus** antes de la fecha/hora límite. Se hará una entrega parcial de progreso el Jueves 30 de Abril 07:00 COT. Esa entrega entrega parcial es obligatoria para que su taller tenga nota distinta de cero.

En cada parte del ejercicio se entrega 1/3 de los puntos si el código propuesto es razonable, 1/3 si se puede ejecutar y 1/3 si entrega resultados correctos.

1. 40 pt **Cuánto cuesta llenar el álbum del Mundial?**

Es Mayo de 2014 y usted está muy emocionado con la Copa Mundial de Fútbol, y en particular, con llenar el álbum del mundial. Para llenarlo usted debe comprar láminas (una a la vez) seleccionándolas de manera aleatoria. Asuma que el álbum completo tiene un total de 640 láminas. Entre estas hay 40 láminas especiales y 40 láminas holográficas. Cada una tiene un costo de 400 COP. Su objetivo es llenar todo el álbum y estimar el costo de hacerlo!

- (a) 15 pt Asuma que todas las cartas tienen la misma probabilidad de ser seleccionadas. Sea X una variable aleatoria que denota el costo de conseguir todas las fichas del álbum. Simule la variable aleatoria X para aproximar $E(x)$ y $var(x)$. Para un tamaño de muestra m , la media y varianza muestrales son μ_m y σ_m^2 respectivamente. Grafique μ_m y σ_m^2 en el eje vertical y $m=10:10:10^5$ en el eje horizontal. Cuánto cuesta llenar el álbum? Explique su respuesta.
- (b) 15 pt Cuál tamaño de muestra m se necesita para obtener 95% de confianza en que $|\mu_m - \mu| \leq 10^{-6}$?
- (c) 10 pt Asuma lo siguiente: $P_{holograficas} = \frac{1}{3}P_{normales}$ y $P_{especiales} = \frac{1}{6}P_{normales}$ donde P es la probabilidad de obtener una lámina de cada uno de los tipos. Responda las preguntas anteriores con estas condiciones.

2. 80 pt Campo Magnético Solar

El **Solar Dynamic Observatory (SDO)**, tiene a bordo un instrumento que mide el campo magnético del sol. En el archivo `hmi.m_45s.magnetogram.subregion_x1y1.fits` (descargar desde <http://bit.ly/1Gsq0eP>) se encuentran las mediciones del campo magnético en una región del sol. Los ejes x y y representan las coordenadas de la región del sol que se está observando. El eje z representa la variación en el tiempo. Esto quiere decir que en $z[0]$ es el campo magnético de la región en un instante de tiempo. $z[1]$ el campo magnético en el siguiente instante de tiempo. El tiempo de cada coordenada se puede encontrar en el archivo `times.csv`.

Sugerencia: Para leer los tiempos en `times.csv` cree un script de R y use el paquete `lubridate` y la función `diff`. Con esto puede sacar un arreglo con el tiempo transcurrido en segundos (o minutos).

Las variaciones del campo magnético solar se pueden explicar mediante 3 modelos:

Lineal:

$$B(t) = a + bt \quad (1)$$

Lineal gaussiano:

$$B(t) = c + dt + \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] \quad (2)$$

Lineal paso:

$$B(t) = f + gt + h\left(1 + \frac{2}{\pi} \tan^{-1}[n(t - t_0)]\right) \quad (3)$$

Donde $a, b, c, d, e, f, g, h, \mu, \sigma, n$ y t_0 son los parámetros libres de los modelos. t es el tiempo.

Para leer el archivo `.fits` lea el notebook `reading_fits.ipynb`.

- (a) 20 pt Escriba 3 códigos en Python. Cada código debe ajustar el modelo a todos los datos observados. Cada código debe escribir un archivo con el nombre de la observación en la columna 1 y los parámetros que mejor se ajustan incluyendo el "likelihood".^{en} las demás columnas. Debe incluir un encabezado que explique que es cada columna. Cada caminata debe tener 20000 pasos y el código debería correr en ~ 20 min.
- (b) 25 pt Escriba un código en python que a partir de los 3 archivos anteriores. Reconozca cuál es el modelo que mejor se ajusta a cada observación. Grafique dicha observación con el mejor modelo, en la gráfica deben aparecer los mejores parámetros incluido el "likelihood". El código debe escribir un archivo `bestmodels.txt` con el nombre de la observación en la primera columna, en la segunda el nombre del modelo y en las siguientes los mejores parámetros del modelo que ajustan dicha observación.
- (c) 10 pt Todo esto debe funcionar con un Makefile. Al final deben salir 100 gráficas.
- (d) 25 pt **Bono:** Repita el primer punto usando la librería `emcee`

Variaciones de los parámetros:

Para las condiciones iniciales tener en cuenta lo siguiente: $a, c, f \sim [-200, 200]$ y $b, d, g \sim [-1, 1]$, $h \sim [-100, 100]$, $n \sim [0, 100]$, $t_0 \sim [80, 350]$, $\mu \sim 100$ y $\sigma \sim 40$