Simulación ITAM
11 de mayo de 2018 Jorge de la Vega
2018-I

Tarea 5.

La fecha de entrega es el **28 de mayo**. Tarea optativa, el que la entregue (correcta) tiene un punto más en el promedio de los parciales.

Lecturas

- Robert & Casella Capítulos 6 y 7
- Dagpunar Capítulos 5 y 8
- Efron y Gong: A leisurely look at the Boostrap, the Jackknife and Cross-Validation
- Chib y Greenberg: Understanding the Metropolis-Hastings Algorithm
- Casella & George. Explaining the Gibbs sampler.

Problemas

1. En una prueba real, 12 lotes de mantequilla de cacahuate tienen residuos de aflatoxin en partes por mil millones de 4.94, 5.06, 4.53, 5.07, 4.99, 5.16, 4.38, 4.43, 4.93, 4.72, 4.92, y 4.96.
   - ¿Cuántas posibles muestras bootstrap hay en estos datos?
   - Usando `R` y la función `sample`, o una tabla de números aleatorios, generar 100 remuestras de los datos de la muestra. Para cada una de estas remuestras, obtener la media. Comparar la media de las medias obtenidas en las remuestras con la media de la muestra original.
   - Encontrar de las 100 remuestras, un intervalo de confianza del 95 % para la media.

2. El número de accidentes aéreos de 1983 a 2006 fueron 23, 16, 21, 24, 34, 30, 28, 24, 26, 18, 23, 23, 36, 37, 49, 50, 51, 56, 46, 41, 54, 30, 40, 31.
   - Para la muestra de datos, calcular la media y su error estándar (a partir de la desviación estándar), así como la mediana.
   - Usando `R`, calcular estimados bootstraps de la media y la mediana con estimados de sus errores estándar, usando $B = 1000$ remuestras. También calcular la mediana de las medianas muestrales.

- ¿Cómo se comparan los dos incisos anteriores?

3. Usar los datos `law` del paquete bootstrap para calcular, usando el método jackknife-after-bootstrap el error estándar del estimado bootstrap de $se(\hat{\rho})$.

4. Las siguientes 12 observaciones son los tiempos, en horas, entre fallas de equipo de aire acondicionado:

$$3, 5, 7, 18, 43, 85, 91, 98, 100, 130, 230, 487$$

Suponiendo que los tiempos siguen una distribución exponencial con parámetro λ, obtener el estimador de máxima verosimilitud de λ y utilizar bootstrap para estimar el sesgo y el error estándar del estimado.

5. Los datos `scor` del paquete `bootstrap` corresponden a califcaciones de 88 estudiantes que realizaron exámenes en 5 materias. Las primeras dos materias (mecánica, algebra lineal) los exámenes fueron a libro cerrado, y las últimas tres (álgebra, análisis y estadística) fueron a libro abierto. Cada renglón de los datos pertenece a un estudiante y sus cinco calificaciones:

```
library(bootstrap)
data(scor)
head(scor)
```

```
  mec vec alg ana sta
1  77  82  67  67  81
2  63  78  80  70  81
3  75  73  71  66  81
4  55  72  63  70  68
5  63  63  65  70  63
6  53  61  72  64  73
```

- Hacer la gráfica matriz de dispersión de puntos de los datos para las cinco calificaciones.
- Obtener estimados boostrap para los errores estándar de $\hat{\rho}_{mec,vec}$, $\hat{\rho}_{alg,ana}$, $\hat{\rho}_{alg,sta}$, $\hat{\rho}_{ana,sta}$.
- Los scores de los 5 exámenes tienen una matriz de varianza-covarianza de 5×5 Σ, con eigenvalores positivos $\lambda_1 > \cdots > \lambda_5$. En el análisis de componentes principales,

$$\theta = \frac{\lambda_1}{\sum_{j=1}^{5} \lambda_j}$$

mide la proporción de varianza explicada por la primera componente. Supongamos que $\hat{\lambda}_1 > \cdots > \hat{\lambda}_5$ son los eigenvalores de $\hat{\Sigma}$, el estimador máximo-verosímil de Σ. Calcular $\hat{\theta}$. Utilizar bootstrap para estimar el sesgo y el error estándar de $\hat{\theta}$.

- Con relación al ejercicio anterior, obtener los estimadores jackknife de sesgo y error estándar.
- Calcular intervalos del 95 % percentíl y BC_a para $\hat{\theta}$.

6. IMplementar la prueba de Separman de correlación de rangos bivariada para independencia como una prueba de permutación. Comparar el ASL de la prueba de permutación con el p-value reportado por la función `cor.test` en las mismas muestras.