Simulación ITAM
18 de noviembre de 2018 2018-II

Tarea 5.

La fecha de entrega es el **6 de diciembre de 2018**.

Lectura

- Chib & Greenberg. Understanding the Metropolis-Hastings Algorithm.
- Casella & George. Explaining the Gibbs sampler.
- Efron. A leisurely look at the boostrapt, the Jackknife, and Cross-validation.
- Salmon, F. (2009) Recipe for Disaster: The Formula That Killed Wall Street, Wired
- Arturo Ederly: Cópulas y variables aleatorias, una introducción

Problemas

1. Supongan que $Y|\theta \sim \mathcal{G}\left(1, \theta\right)$ y que $\theta \sim IG(\alpha, \beta)$.
   - Encuentren la distribución posterior de θ.
   - Encuentren la media y varianza posterior de θ.
   - Encuentren la moda posterior de θ.
   - Escriban dos ecuaciones integrales que se pueden resolver para encontrar el intervalo de 95 % de colas simétricas para θ.

2. Los siguientes datos corresponden a las horas adicionales de sueño de 10 pacientes tratados con un somnífero B comparado con un somnífero A:

   $$1.2, 2.4, 1.3, 1.3, 0, 1, 1.8, 0.8, 4.6, 1.4$$

   Lleven a cabo un análisis bayesiano de estos datos y extraigan conclusiones, asumiendo que cada componente de la verosimilitud sea:

   - normal
   - $t_{(3)}$
   - $t_{(1)}$

- Bernoulli (de alguna manera que se les ocurra)

En este ejercicio, escriban un código para manejar cualquier integración necesaria y cálculo de probabilidades marginales posteriores.

3. Spiegelhalter et al. (1995) analiza la mortalidad del escarabajo del trigo en la siguiente tabla, usando BUGS.

Dosis	# muertos	# expuestos
w_i	y_i	n_i
1.6907	6	59
1.7242	13	60
1.7552	18	62
1.7842	28	56
1.8113	52	63
1.8369	53	59
1.8610	61	62
1.8839	60	60

Estos autores usaron una parametrización usual en dos parámetros de la forma $p_i \equiv P(\text{muerte}|w_i)$, pero comparan tres funciones ligas diferentes:

$$\begin{aligned} \text{logit:} \quad p_i &= \frac{\exp(\alpha + \beta z_i)}{1 + \exp(\alpha + \beta z_i)} \\ \text{probit:} \quad p_i &= \Phi(\alpha + \beta z_i) \\ \text{complementario log-log:} \quad p_i &= 1 - \exp[-\exp(\alpha + \beta z_i)] \end{aligned}$$

en donde se usa la covariada centrada $z_i = w_i - \bar{w}$ para reducir la correlación entre la ordenada α y la pendiente β. En OpenBUGS el código para implementar este modelo es el que sigue:

```
model{
for (i in 1:k){
y[i] ~ dbin(p[i],n[i])
logit(p[i]) <- alpha + beta*(w[i]-mean(w[])
#       probit(p[i]) <- alpha + beta*(w[i]-mean(w[])
#       cloglog(p[i]) <- alpha + beta*(w[i]-mean(w[])
} #fin del loop i

alpha ~ dnorm(0.0,0.001)
dbeta ~ dnorm(0.0,0.001)
} #fin del código
```

Lo que sigue al símbolo # es un comentario, así que esta versión corresponde al modelo logit. También `dbin` denota la distribución binomial y `dnorm` denota la distribución normal, donde el segundo argumento denota la precisión, no la varianza (entonces las iniciales normales para α y β tienen precisión 0.001, que son aproximadamente iniciales planas (no informativas)). Hacer el análisis en `OpenBUGS` o donde crean conveniente.

4. Consideren las siguientes dos distribuciones condicionales completas, analizadas en el artículo de Casella y George (1992):

$$f(x|y) \propto ye^{-yx}, \qquad 0 < x < B < \infty$$
$$f(y|x) \propto xe^{-xy}, \qquad 0 < y < B < \infty$$

   - Obtener un estimado de la distribución marginal de X cuando $B = 10$ usando el Gibbs sampler.
   - Ahora supongan que $B = \infty$ así que las distribuciones condicionales completas son ahora las ordinarias distribuciones exponenciales no truncadas. Mostrar analíticamente que $f_x(t) = 1/t$ es una solución a la ecuación integral en este caso:
$$f_x(x) = \int \left[\int f_{x|y}(x|y) f_{y|t}(y|t) dy \right] f_x(t) dt$$
¿EL Gibbs sampler convergerá a esta solución?

5. En una prueba real, 12 lotes de mantequilla de cacahuate tienen residuos de aflatoxin en partes por mil millones de:

$$4.94, 5.06, 4.53, 5.07, 4.99, 5.16, 4.38, 4.43, 4.93, 4.72, 4.92, 4.96$$

   - ¿Cuántas posibles muestras boototrap 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.

6. 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?

7. La τ de Kendall entre X y Y es 0.55. Tanto X como Y son positivas. ¿Cuál es la τ entre X y $1/Y$? ¿Cuál es la τ de $1/X$ y $1/Y$?