



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

TP

26 de mayo de 2016

Inferencia Bayesiana

Integrante	LU	Correo electrónico
Freund, Teodoro	526/15	tfreund95@gmail.com
Radunsky, Andrés	395/15	andres.radunsky@gmail.com



**Facultad de Ciencias Exactas y
Naturales**

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta
Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep.
Argentina

Tel/Fax: (++54 +11) 4576-3300

<http://www.exactas.uba.ar>

0. Enunciado

Observamos 10 tiradas de 3 monedas distintas, de las que sabemos que hay dos monedas comunes y una cargada, aunque ignoramos cargada como. Los números de caras obtenidos para las distintas monedas en las 10 tiradas son 3, 4 y 10.

1. Problema 1: Modelo y Representación Gráfica

1.1. Enunciado

Escriba un modelo que capture el problema enunciado. Realice una representación gráfica del modelo propuesto, utilizando la convención para identificar nodos latentes, observados y determinísticos.

1.2. Resolución

Para la resolución del problema consideramos dos posibles modelos, en ambos tenemos que x_i representa la cantidad de caras que salen en la i -ésima moneda, m representa el número de tiradas, en este caso 10, θ_i es la probabilidad de que la i -ésima moneda salga cara y c es una variable aleatoria con distribución categórica que hace de selector de la moneda cargada.

En el primer modelo se tienen tres variables aleatorias que complementan lo anterior, λ_1 y λ_2 son dos probabilidades independientes, con distribución a priori $Beta(1000, 1000)$ que simbolizan la confianza que tenemos en que hay dos monedas no cargadas. τ es la distribución de la probabilidad de que la moneda cargada salga cara, y a priori sigue una distribución $Beta(1, 1)$, que representa nuestro desconocimiento sobre su estado.

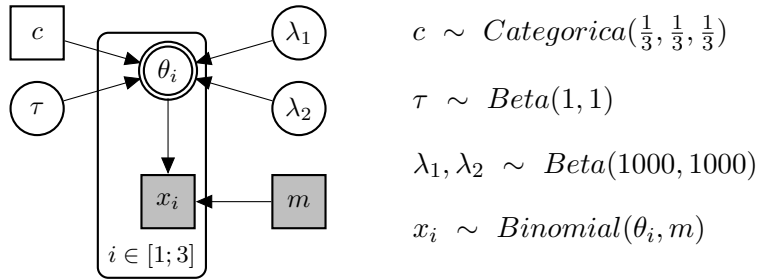
A partir de c , τ , λ_1 y λ_2 se determina θ_i siguiendo las siguientes reglas:

$$c = 1 \implies (\theta_1 = \tau \wedge \theta_2 = \lambda_2 \wedge \theta_3 = \lambda_1)$$

$$c = 2 \implies (\theta_1 = \lambda_1 \wedge \theta_2 = \tau \wedge \theta_3 = \lambda_2)$$

$$c = 3 \implies (\theta_1 = \lambda_2 \wedge \theta_2 = \lambda_1 \wedge \theta_3 = \tau)$$

Se ve de la regla de asignación que el valor de c indica que moneda es la cargada. La asignación de los λ_i se eligió de manera tal de que ninguno se asigne siempre a la misma moneda para los dos casos en que esta no esté cargada.



En nuestro segundo modelo, usamos otro acercamiento, mientras que la idea general es la misma utilizamos un λ_i y un τ_i por moneda, esto nos permite independizar mucho más las muestras de cada una, siendo la única variable que las correlaciona c .

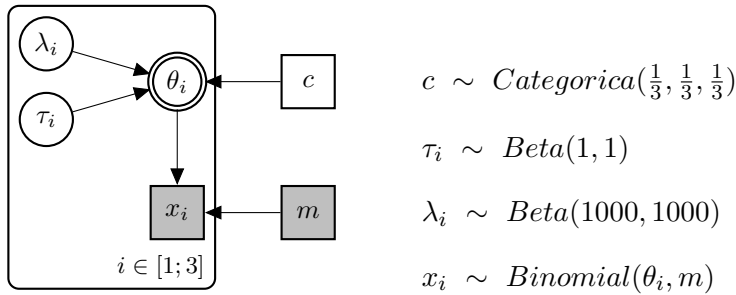


Figura 1: Representación del Segundo Modelo

2. Problema 2: Implementación e Inferencia

2.1. Enunciado

Implemente el modelo en su sistema de inferencia predilecto, y obtenga muestras de la posterior para la variables relevantes. Explícite cuales fueron los parámetros elegidos para el algoritmo de muestreo.

- Realice histogramas de las distintas variables, utilizando un mismo gráfico cuando sea posible/razonable.
- Reporte la media y el desvío estándar para todas las variables inferidas.
- Compute la probabilidad a posteriori de que cada una de las monedas sea la moneda cargada.

2.2. Resolución

Para ambos modelos se utilizaron los mismos parámetros de sampleo (mostrados en el Cuadro 1) para poder compararlos en condiciones similares. Del cuadro se ve que se guardó un numero elevado de samples, esto fue

principalmente para obtener una cantidad considerable de muestras para las variables que raramente se aplican al modelo debido a que se encuentran en zonas de baja probabilidad y obtener así gráficos menos ruidosos (esta consideración se tuvo en cuenta principalmente por el segundo modelo).

```
nchains = 3  nburnin = 3e3  nsamples = 9e5
nthin = 3    doparallel = 0
```

Cuadro 1: Valores de muestreo utilizados en el algoritmo para ambos modelos.

```
model{
  # Variable que decide que moneda esta cargada
  for(i in 1:m){
    p[i] <- 1/m
  }
  c ~ dcat(p[])

  #Priors de las variables.
  #La distribucion depende del valor de k
  Tau ~ dbeta(1, 1)
  Lambda[1] ~ dbeta(1000,1000)
  Lambda[2] ~ dbeta(1000,1000)

  Theta[1] <- equals(c, 1) * Tau + equals(c, 2) * Lambda[1] +
    equals(c,3) * Lambda[2]
  Theta[2] <- equals(c, 1) * Lambda[2] + equals(c,2) * Tau +
    equals(c, 3) * Lambda[1]
  Theta[3] <- equals(c, 1) * Lambda[1] + equals(c,2) * Lambda[2] +
    equals(c, 3) * Tau

  for (i in 1:m){
    k[i] ~ dbin(Theta[i], n)
  }
}
```

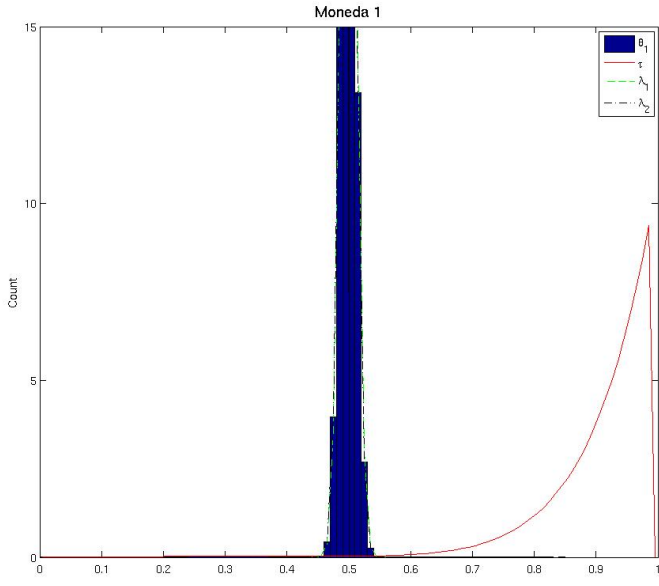
Figura 2: Implementación del primer modelo en JAGS

a) En la Figura 2 se muestra la implementación del primer modelo descrito en la sección 1.2. Se tienen entonces los parámetros c , τ , λ_1 , λ_2 con sus respectivas distribuciones, θ_i determinado con la regla presentada anteriormente y por último k_i con su distribución. Utilizando MATJAGS se

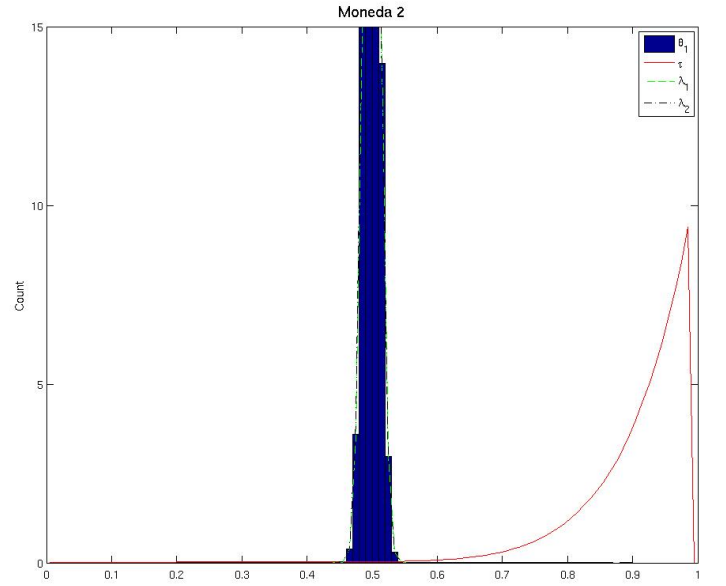
obtuvieron muestras de todos estos parámetros con los que se armaron los gráficos presentados en la Figura 3.

De estos se ve claramente que la distribución de θ_3 es casi idéntica a la de τ , mientras que la de las otras dos monedas es ajustada muy bien por las distribuciones de λ_1 y λ_2 . Esto parecería indicar a primera vista que lo más probable es que la tercera moneda sea la cargada, ya que, recordando el significado de τ , este era la distribución de probabilidad de la moneda cargada. Esta conclusión se ve corroborada por el Gráfico 3(d), ya que se ve como la variable categórica se muestreo muchas más veces indicando que la cargada era la tercer moneda.

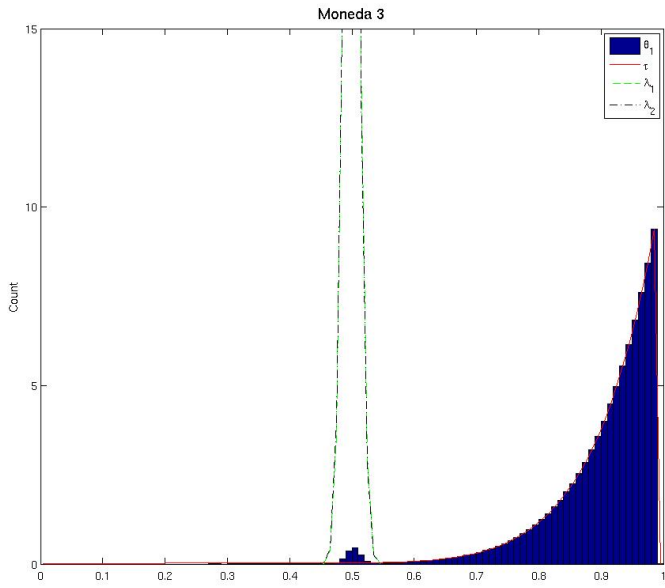
Para el segundo modelo hubo que realizar un procesamiento extra de los datos obtenidos. Dado que cada moneda posee su propio τ y λ , para una moneda en particular estos valores solo tienen sentido en el modelo si en ese paso del sampleo resultó ser o no la cargada respectivamente. Sin embargo, el programa toma muestras de todos los parámetros en cada paso, como consecuencia se tuvieron que filtrar de los resultados obtenidos los valores "basura" de los parámetros, esto se hizo en MATLAB.



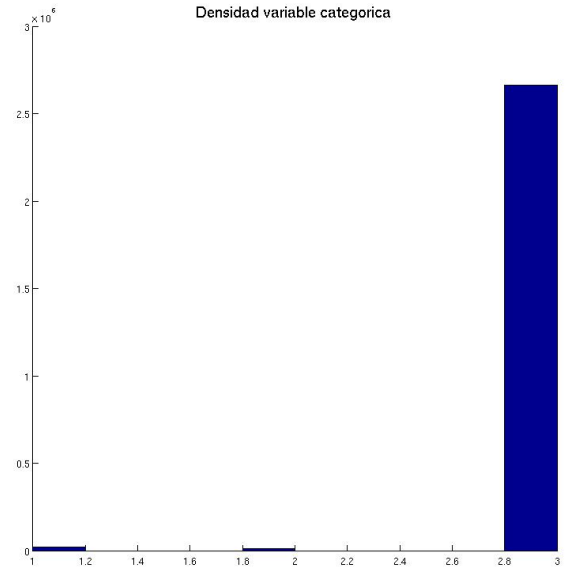
(a) Distribuciones para la moneda 1.



(b) Distribuciones para la moneda 2.



(c) Distribuciones para la moneda 3.



(d) Distribución de la variable categórica.

Figura 3: Se grafican los resultados obtenidos para el primer modelo presentado.

```

model{
  # Variable que decide que moneda esta cargada
  for(i in 1:m){
    p[i] <- 1/m
  }
  c ~ dcat(p[])

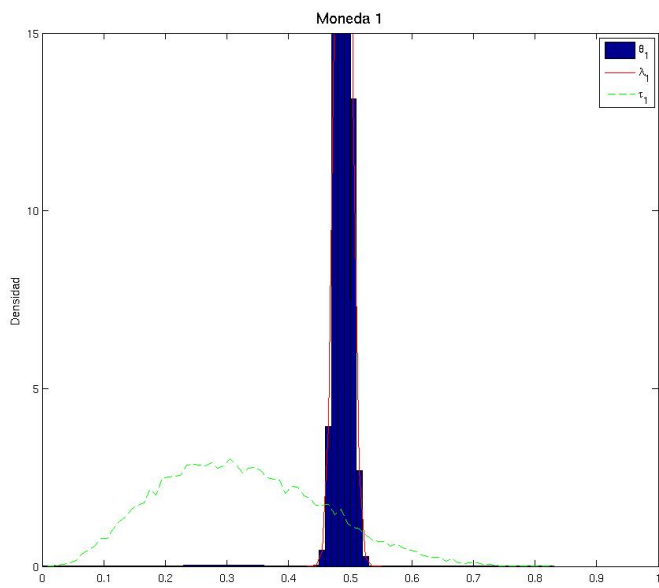
  #Priors de las variables.
  # La distribucion depende del valor de k
  for (i in 1:m){
    Lambda[i] ~ dbeta(1000, 1000)
    Tau[i] ~ dbeta(1, 1)
    Theta[i] <- ifelse(c == i, Tau[i], Lambda[i])
  }

  for (i in 1:m){
    k[i] ~ dbin(Theta[i], n)
  }
}

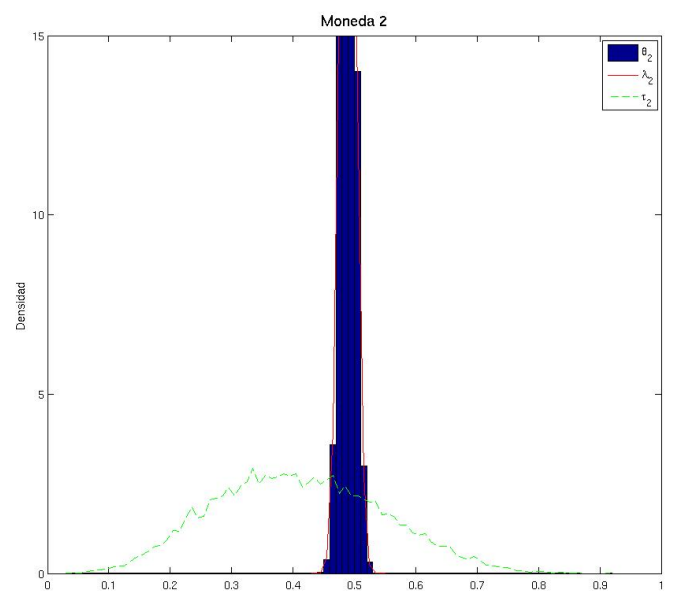
```

Figura 4: Implementación del segundo modelo en JAGS

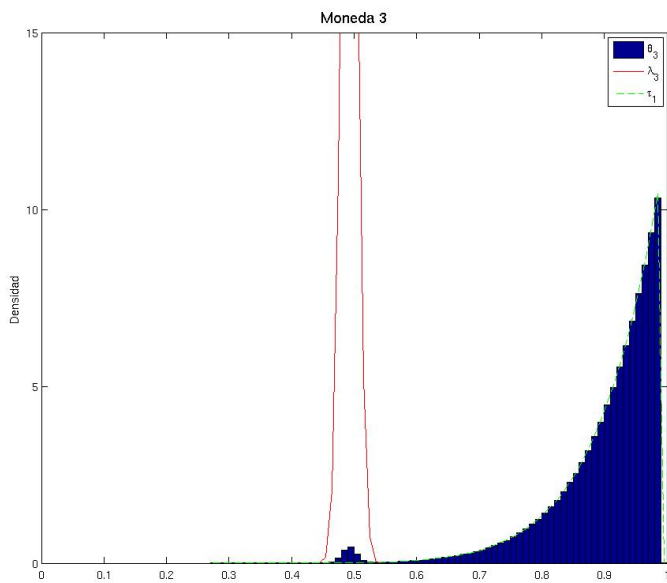
Por otro lado, de este esfuerzo extra se obtuvo un modelo mucho más expresivo, con parámetros con significado más concreto que para el primer modelo. Si se observan los gráficos presentados en la Figura 5 se ve que para cada moneda, λ representa la distribución de probabilidad de que salga cara para esa moneda si resulta no estar cargada mientras que τ lo es en el caso en que sí lo esté. Volviendo a los gráficos del primer modelo, λ_1 y λ_2 son prácticamente idénticas y no parecen estar dando mucha información acerca de cada moneda en particular si no más bien de algo más general de las monedas como conjunto.



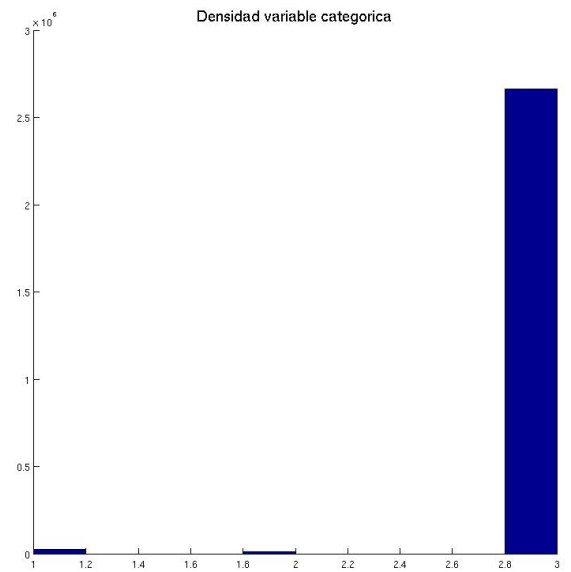
(a) Distribuciones para la moneda 1.



(b) Distribuciones para la moneda 2.



(c) Distribuciones para la moneda 3.



(d) Distribución de la variable categórica.

Figura 5: Se grafican los resultados obtenidos para el segundo modelo presentado.

Se pueden sacar conclusiones similares a las que provee el primer modelo, ya que de los gráficos se observa que las monedas 1 y 2 se distribuyen de

manera muy similar a sus respectivos λ mientras que la 3 es muy similar a su τ . La distribución de la categórica corrobora significativamente el hecho de que la tercera moneda parece ser la cargada.

b) Observando los Cuadros 2 y 3 se puede observar como de ambos modelos se llegan a los mismos valores. Los valores medios de c para ambos modelos con su desviación estándar no presentan diferencias significativas, y dado que indica que $c = 3$ entonces se comparan los valores de τ y τ_3 observando nuevamente que no existen diferencias significativas. Por otro lado, dado que los λ tenían una distribución a priori $Beta \sim (1000, 1000)$ la evidencia observada no fue suficientemente fuerte como para creer que las otras monedas puedan estar cargadas (todos están cerca de 0.5).

	Media	Desvío estándar
τ	0.90954	0.099887
λ_1	0.49951	0.011153
λ_2	0.49903	0.011153
θ_1	0.49764	0.022179
θ_2	0.49910	0.015671
θ_3	0.91135	0.089429
c	2.9787	0.19342

Cuadro 2: Medias y desvíos estándar para el primero modelo

	Media	Desvío estándar
τ_1	0.33264	0.130992
τ_2	0.41831	0.135655
τ_3	0.91662	0.076764
λ_1	0.49901	0.011153
λ_2	0.49901	0.011153
λ_3	0.50254	0.011190
θ_1	0.49760	0.022379
θ_2	0.49912	0.015597
θ_3	0.91113	0.089768
c	2.9783	0.19537

Cuadro 3: Medias y desvíos estándar para el segundo modelo

c) Para computar esta probabilidad estimamos sobre los datos de c para ambos modelos, haciendo para cada posible valor de i (1,2,3) la división entre el número de muestras donde $c = i$ y el número de muestras total (2700000).

$$\begin{array}{l|l} P(c = 1|D) & 0.0082870 \\ P(c = 2|D) & 0.0047163 \\ P(c = 3|D) & 0.98700 \end{array}$$

Cuadro 4: Probabilidad de que cada moneda este cargada, dados los datos, para el primer modelo

$$\begin{array}{l|l} P(c = 1|D) & 0.0084630 \\ P(c = 2|D) & 0.0047889 \\ P(c = 3|D) & 0.98675 \end{array}$$

Cuadro 5: Probabilidad de que cada moneda este cargada, dados los datos, para el segundo modelo

De los cuadros 4 y 5 se ve que las probabilidades difieren en valores muy chicos entre los modelos planteados.

2.3. Conclusión de los modelos.

Como conclusión final del análisis de los modelos se ve que de ambos se llegaron a las mismas conclusiones y que existe un tradeoff entre descriptividad del modelo vs procesamiento de datos entre el primer y segundo modelo. El primer modelo resulta más rápido de implementar si lo que se busca es reconocer la moneda más probable a ser la cargada y su distribución, mientras que el segundo además de reconocer esto, da la distribución para de las monedas para todas las otras posibilidades (dados los mismos datos). Sin embargo obtener un número considerable de muestras para observar todos estos casos (más aún cuando resultan poco probables) requieren realizar un número de muestras muy elevado, por lo que el análisis se vuelve más costoso.

3. Problema 3: Modificaciones al Modelo

3.1. Enunciado

Discuta como modificaría el modelo si en lugar de saber que hay una moneda cargada nos dicen que cada moneda puede estar cargada o no con probabilidad $1/2$ independientemente de las otras monedas. Provea el modelo y su representación gráfica para este caso.

3.2. Resolución

Para este modelo alternativo la principal diferencia es que ya no existe correlación entre las monedas, siendo estas completamente independientes

una de otra. Es por esto que en vez de tener la variable c que indica cual de las tres monedas es la cargada, ahora se tiene una variable, d_i por cada moneda, que indica si la i -esima moneda se encuentra cargada o no. Esta variable sigue una distribución $Bernoulli(\frac{1}{2})$ y se determina θ_i de la siguiente forma:

$$d_i = 0 \implies \theta_i = \lambda_i$$

$$d_i = 1 \implies \theta_i = \tau_i$$

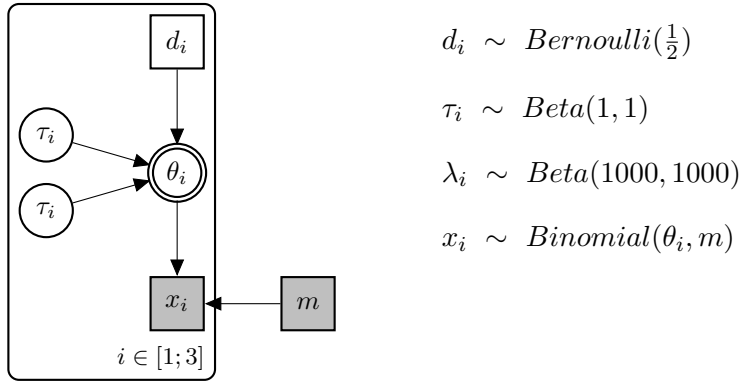


Figura 6: Representación del modelo modificado

4. Problema 4: Predicciones

4.1. Enunciado

Escriba la expresión para la probabilidad de obtener cara en la próxima tirada para cada una de las monedas. ¿Que distintas fuentes de incertidumbre puede identificar en ella?

4.2. Resolución

Para esta resolución utilizaremos el segundo modelo presentado, el que tiene 3 τ_i y 3 λ_i .

Sea k_i una $\text{Binomial}(\theta_i, 1)$, por definición $P(k_i = 1|\theta_i) = \theta_i$ y sea $D = [X_1, X_2, X_3]$ los datos, que siguen una distribución $x_i \sim \text{Binomial}(\theta_i, 10)$ en nuestro modelo. Entonces, lo que estamos buscando es $P(k_i = 1|D)$, para cada i , y tenemos que:

$$P(k_i = 1|D) = \int_0^1 P(k_i = 1|\theta_i) * P(\theta_i|D) d\theta_i = \int_0^1 \theta_i * P(\theta_i|D) d\theta_i \quad (1)$$

Siguiendo nuestro modelo:

$$\begin{aligned}
P(\theta_i|D) &= P(\theta_i|c = i, D) * P(c = i|D) + P(\theta_i|c \neq i, D) * P(c \neq i|D) \\
&= P(\tau_i|c = i, D) * P(c = i|D) + P(\lambda_i|c \neq i, D) * P(c \neq i|D)
\end{aligned}$$

Por el Apéndice:

$$\tau_{i|c=i,D} \sim \text{Beta}(1 + X_i, 1 + 10 - X_i)$$

$$\lambda_{i|c \neq i,D} \sim \text{Beta}(1000 + X_i, 1000 + 10 - X_i)$$

Y, por Bayes, obtenemos:

$$P(c = i|D) = \frac{P(D|c = i)P(c = i)}{P(D)}$$

$$P(D|c = i) = \int_0^1 \int_0^1 \int_0^1 P(D|\theta, c = i)P(\theta|c = i)d\theta$$

$$P(D|\theta, c = i) = P(k_1 == K_1|\theta_1)P(k_2 == K_2|\theta_2)P(k_3 == K_3|\theta_3)$$

$$P(\theta|c = i) = P(\theta_1|c = i)P(\theta_2|c = i)P(\theta_3|c = i)$$

$$\begin{aligned}
P(D|c = i) &= \int_0^1 \int_0^1 \int_0^1 P(k_1 == K_1|\theta_1)P(k_2 == K_2|\theta_2)P(k_3 == K_3|\theta_3) \\
&\quad P(\theta_1|c = i)P(\theta_2|c = i)P(\theta_3|c = i)d\theta_1d\theta_2d\theta_3
\end{aligned}$$

Notar que son independientes, y que θ_i ya esta determinada, porque la variable c ya está dada.

$$P(\theta_i|c = i) = P(\tau_i)$$

$$P(\theta_i|c \neq i) = P(\lambda_i)$$

Y completamos con las siguientes distribuciones conocidas:

$$P(c \neq i|D) = 1 - P(c = i|D)$$

$$P(D) = \sum_{j=1}^3 P(D|c = j)P(c = j)$$

$$P(c = i) = \frac{1}{3}, \forall i$$

$$\tau_i \sim \text{Beta}(1, 1)$$

$$\lambda_i \sim \text{Beta}(1000, 1000)$$

Entonces tenemos por la ecuación 1, y por haber determinado cada una de sus componentes, la probabilidad de que la i -ésima moneda sea cara.

Las principales fuentes de incertidumbre que encontramos son las tres variables aleatorias para cada i , es decir c , τ_i y λ_i , estas distintas incertidumbres se deben a su naturaleza de aleatoriedad y se ven reflejadas en el Cuadro 2 donde el desvío estandar para cada una de estas es distinto de 0. Por otro lado, consideramos que θ_i no es una fuente de incertidumbre ya que se calcula determinísticamente a partir de las nombradas anteriormente, y su desvío es simplemente heredado de estas.

A. Prueba necesaria

Buscamos probar que

$$\tau_i|_{D,c=i} \sim \text{Beta}(1 + X_i, 1 + 10 - X_i)$$

$$P(\tau_i|c = i, D) = \frac{P(D|c = i, \tau_i)P(\tau_i|c = i)}{P(D|c = i)}$$

$$P(D|c = i, \tau_i) = P(x_i = X_i|c = i, \tau_i) \prod_{j=1, j \neq i}^3 P(x_j = X_j|c \neq j)$$

Y ademias:

$$P(D|c = i) = \int_0^1 P(D|c = i, \tau_i)P(\tau_i|c = i)d\tau_i$$

$$P(D|c = i) = \int_0^1 P(x_i = X_i|c = i, \tau_i) \prod_{j=1, j \neq i}^3 P(x_j = X_j|c \neq j)P(\tau_i|c = i)d\tau_i$$

$$P(D|c = i) = \prod_{j=1, j \neq i}^3 P(x_j = X_j|c \neq j) \int_0^1 P(x_i = X_i|c = i, \tau_i)P(\tau_i|c = i)d\tau_i$$

entonces:

$$\begin{aligned} P(\tau_i|c = i, D) &= \frac{P(D|c = i, \tau_i)P(\tau_i|c = i)}{P(D|c = i)} \\ &= \frac{P(x_i = X_i|c = i, \tau_i) \prod_{j=1, j \neq i}^3 P(x_j = X_j|c \neq j)P(\tau_i|c = i)}{\prod_{j=1, j \neq i}^3 P(x_j = X_j|c \neq j) \int_0^1 P(x_i = X_i|c = i, \tau_i)P(\tau_i|c = i)d\tau_i} \\ &= \frac{P(x_i = X_i|c = i, \tau_i)P(\tau_i|c = i)}{\int_0^1 P(x_i = X_i|c = i, \tau_i)P(\tau_i|c = i)d\tau_i} = P(\tau_i|c = i, x_i = X_i) \end{aligned}$$

Y, por lo visto en clase y por tener que $\tau_i|_{c=i} \sim \text{Beta}(1, 1)$ y $x_i|_{c=i, \tau_i} \sim \text{Binomial}(10, \tau_i)$ se tiene que $\tau_i|_{c=i, x_i=X_i} \sim \text{Beta}(1 + X_i, 1 + 10 - X_i)$, entonces $\tau_i|_{D,c=i} \sim \text{Beta}(1 + X_i, 1 + 10 - X_i)$ por la igualdad anterior.

De forma similar se puede probar que

$$\lambda_i|_{D,c=i} \sim \text{Beta}(1000 + X_i, 1000 + 10 - X_i)$$