Sesión 6: Muestreador de Gibbs

Aplicaciones en Computación Estadística

Natalie Julian - www.nataliejulian.com

Estadística UC y Data Scientist en Zippedi Inc.

Muestreador de Gibbs

Inferencia Bayesiana

1. El procedimiento de Gibbs Sampler es uno de los procedimientos MCMC (Markov chain Monte Carlo) más importante.
2. Nuestro objetivo es producir valores de la distribución a posteriori y resumir la distribución a posteriori usando estos valores.
3. MCMC es una clase de algoritmos que produce una cadena de valores simulados de una distribución donde cada valor es dependiente del anterior.
4. Si la cadena satisface algunas condiciones, entonces la cadena converge a la distribución estacionaria (en nuestro caso, la distribucion a posteriori).

Algoritmo

Nos interesa obtener valores de $\theta = (\theta_1, \theta_2, ..., \theta_p)$ mediante sus respectivas distribuciones a posteriori.

- P1 Especificar valores iniciales $\theta^{(0)}$ y definir el número de iteraciones T
- P2 Para $t = 1, ..., T$
  - 0.1 Simular $\theta_1^{(t)} \sim p(\theta_1 | resto^{(t)})$
  - 0.2 Simular $\theta_2^{(t)} \sim p(\theta_2 | resto^{(t)})$
  - 0.3 $\vdots$
  - 0.4 Simular $\theta_p^{(t)} \sim p(\theta_p | resto^{(t)})$
- P3 Quemar las primeras b observaciones (*¿Por qué quemar? Leer aquí!*)
- P4 Calcular estadísticas de resumen usando los resultados de las últimas $T - b$ iteraciones.

Muestreador de Gibbs - Modelo jerárquico

Número de matrimonios durante segunda guerra mundial

Considere el número de matrimonios por cada 1000 personas en Italia, desde el año 1936 hasta 1951. La pregunta de interés es si es conveniente modelar las tasas de matrimonio durante la segunda guerra mundial del mismo modo que antes y después de ella. Sea y_i el número de matrimonios en el año i desde el año 1936 hasta 1951.

Los valores y_i son:

7 9 8 7 7 6 6 5 5 7 9 10 8 8 8 7

Considere que $y_i \sim Pois(\lambda_i)$. No tiene sentido asumir que λ_i es constante para todos los años, más aún, dado un contexto particular como la segunda guerra mundial.

Número de matrimonios durante segunda guerra mundial

Dado lo anterior, tiene sentido asumir una distribución a priori para λ_i, parámetro del cual depende y_i.
Considere el siguiente modelo jerárquico considerando una priori Gamma para λ_i:

$$y_i|\lambda_i \sim Pois(\lambda_i)$$

$$\lambda_i|\beta \sim Gamma(\alpha, \beta)$$

$$\beta \sim Gamma(\gamma, \delta)$$

Con α, γ y δ fijos.

Pasos para Muestrear utilizando Gibbs Sampler

Determinar una expresión para la posteriori de θ

Pasos para Muestrear utilizando Gibbs Sampler

Determinar una expresión para la posteriori de θ
Notar que $\theta = (\lambda_1, \lambda_2, ..., \lambda_n, \beta)$.

Pasos para Muestrear utilizando Gibbs Sampler

Determinar una expresión para la posteriori de θ
Notar que $\theta = (\lambda_1, \lambda_2, ..., \lambda_n, \beta)$.
Es decir, queremos $p(\lambda_1, \lambda_2, ..., \lambda_n, \beta | resto)$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar una expresión para la posteriori de θ
Notar que $\theta = (\lambda_1, \lambda_2, ..., \lambda_n, \beta)$.
Es decir, queremos $p(\lambda_1, \lambda_2, ..., \lambda_n, \beta | resto)$

Y ya sabemos que:

$$
\begin{aligned}
y_i|\lambda_i &\sim Pois(\lambda_i) \\
\lambda_i|\beta &\sim Gamma(\alpha, \beta) \\
\beta &\sim Gamma(\gamma, \delta)
\end{aligned}
$$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar una expresión para la posteriori de θ
Notar que $\theta = (\lambda_1, \lambda_2, ..., \lambda_n, \beta)$.
Es decir, queremos $p(\lambda_1, \lambda_2, ..., \lambda_n, \beta | resto)$

Y ya sabemos que:

$$y_i | \lambda_i \sim Pois(\lambda_i)$$

$$\lambda_i | \beta \sim Gamma(\alpha, \beta)$$

$$\beta \sim Gamma(\gamma, \delta)$$

Y también sabemos que $posteriori \propto verosimilitud \cdot priori$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar una expresión para la posteriori de θ
Notar que $\theta = (\lambda_1, \lambda_2, ..., \lambda_n, \beta)$.
Es decir, queremos $p(\lambda_1, \lambda_2, ..., \lambda_n, \beta | resto)$

Y ya sabemos que:

$$y_i|\lambda_i \sim Pois(\lambda_i)$$

$$\lambda_i|\beta \sim Gamma(\alpha, \beta)$$

$$\beta \sim Gamma(\gamma, \delta)$$

Y también sabemos que $posteriori \propto verosimilitud \cdot prioris$, entonces, la posteriori:

$$p(\lambda_1, ..., \lambda_n, \beta | resto) \propto \prod_{i=1}^{n} p(y_i|\lambda_i) \cdot p(\lambda_i|\beta) \cdot p(\beta)$$

Determinar una expresión para la posteriori de θ

De lo anterior, llegamos a:

$p(\lambda_1, ..., \lambda_n, \beta | resto) \propto$

$$\left(\prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right)\left(\prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}\right)\frac{\delta^{\gamma}}{\Gamma(\gamma)}\beta^{\gamma-1}e^{-\delta\beta}$$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar una expresión para la posteriori de θ

De lo anterior, llegamos a:

$p(\lambda_1, ..., \lambda_n, \beta | resto) \propto$

$$\left(\prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right)\left(\prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}\right)\frac{\delta^{\gamma}}{\Gamma(\gamma)}\beta^{\gamma-1}e^{-\delta\beta}$$

Juntando las productorias:

$$p(\lambda_1, ..., \lambda_n, \beta | resto) \propto \left(\prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}\right)\frac{\delta^{\gamma}}{\Gamma(\gamma)}\beta^{\gamma-1}e^{-\delta\beta}$$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar las distribuciones condicionales completas

Es decir, queremos $p(\lambda_i|resto)$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar las distribuciones condicionales completas
Es decir, queremos $p(\lambda_i|resto)$ (*Multiplicar todo lo que dependa de* λ_i).

Determinar las distribuciones condicionales completas
Es decir, queremos $p(\lambda_i|resto)$ (*Multiplicar todo lo que dependa de* λ_i).

$$p(\lambda_i|resto) \propto \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}$$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar las distribuciones condicionales completas
Es decir, queremos $p(\lambda_i|resto)$ (*Multiplicar todo lo que dependa de* λ_i).

$$p(\lambda_i|resto) \propto \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}$$

Podemos obviar algunas constantes:

$$p(\lambda_i|resto) \propto e^{-\lambda_i}\lambda_i^{y_i}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}$$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar las distribuciones condicionales completas
Es decir, queremos $p(\lambda_i|resto)$ (*Multiplicar todo lo que dependa de* λ_i).

$$p(\lambda_i|resto) \propto \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}$$

Podemos obviar algunas constantes:

$$p(\lambda_i|resto) \propto e^{-\lambda_i}\lambda_i^{y_i}\lambda_i^{\alpha-1}e^{-\beta\lambda_i} = \lambda_i^{(\alpha+y_i)-1}e^{-\lambda(\beta+1)}$$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar las distribuciones condicionales completas
Es decir, queremos $p(\lambda_i|resto)$ (*Multiplicar todo lo que dependa de* λ_i).

$$p(\lambda_i|resto) \propto \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}$$

Podemos obviar algunas constantes:

$$p(\lambda_i|resto) \propto e^{-\lambda_i}\lambda_i^{y_i}\lambda_i^{\alpha-1}e^{-\beta\lambda_i} = \lambda_i^{(\alpha+y_i)-1}e^{-\lambda_i(\beta+1)}$$

Notar que $\lambda_i|resto \sim Gamma(\alpha + y_i, \beta + 1)$

Determinar las distribuciones condicionales completas

$$p(\beta|resto) \propto \left(\prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta\lambda_i} \right) \frac{\delta^{\gamma}}{\Gamma(\gamma)} \beta^{\gamma-1} e^{-\delta\beta}$$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar las distribuciones condicionales completas

$$p(\beta|resto) \propto \left(\prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta\lambda_i}\right) \frac{\delta^{\gamma}}{\Gamma(\gamma)} \beta^{\gamma-1} e^{-\delta\beta}$$

Obviamos algunas constantes:

$$p(\beta|resto) \propto \beta^{n\alpha} e^{-\beta\sum_{i=1}^{n}\lambda_i} \beta^{\gamma-1} e^{-\delta\beta}$$

Pasos para Muestrear utilizando Gibbs Sampler

Determinar las distribuciones condicionales completas

$$p(\beta|resto) \propto \left(\prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta\lambda_i}\right) \frac{\delta^{\gamma}}{\Gamma(\gamma)} \beta^{\gamma-1} e^{-\delta\beta}$$

Obviamos algunas constantes:

$$p(\beta|resto) \propto \beta^{n\alpha} e^{-\beta \sum_{i=1}^{n} \lambda_i} \beta^{\gamma-1} e^{-\delta\beta} = \beta^{(n\alpha+\gamma)-1} e^{-\beta(\delta+\sum_{i=1}^{n} \lambda_i)}$$

Determinar las distribuciones condicionales completas

$$p(\beta|resto) \propto \left(\prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}\right)\frac{\delta^{\gamma}}{\Gamma(\gamma)}\beta^{\gamma-1}e^{-\delta\beta}$$

Obviamos algunas constantes:

$$p(\beta|resto) \propto \beta^{n\alpha}e^{-\beta\sum_{i=1}^{n}\lambda_i}\beta^{\gamma-1}e^{-\delta\beta} = \beta^{(n\alpha+\gamma)-1}e^{-\beta(\delta+\sum_{i=1}^{n}\lambda_i)}$$

Por lo tanto,

$$\beta|resto \sim Gamma(n\alpha+\gamma, \delta+\sum_{i=1}^{n}\lambda_i)$$

Pasos para Muestrear utilizando Gibbs Sampler

¿Qué ganamos?

¿Qué ganamos?

$$\lambda_i | resto \sim Gamma(\alpha + y_i, \beta + 1)$$

$$\beta | resto \sim Gamma(n\alpha + \gamma, \delta + \sum_{i=1}^{n} \lambda_i)$$

Implementación

Obtendremos una cadena de largo $T = 20000$ fijando $\alpha = 1$, $\delta = 1$ y $\gamma = 1$. Y tomando como valores iniciales $\lambda_i^{(0)} = 1$ y $\beta^{(0)} = 1$.

En R

```
#Caso Poisson con prioris gamma - Modelo jerárquico
años<-1936:1951
y<-c(7,9,8,7,7,6,6,5,5,7,9,10,8,8,8,7) #Datos y_i
(n<-length(y))

T <- 20000
beta <- rep(NA, T)  #Cadena para beta
lambda <- matrix(0, nrow=T, ncol=n) #n cadenas, una por cada lambda_i

#### Fijamos alpha, delta y gamma (son fijos)
alpha <- 1
delta <- 1
gamma <- 1

###Valores iniciales:
beta[1] <- 1
lambda[1,] <- rep(1, n)

#Muestreamos a partir de las distribuciones condicionales completas!

for(i in 2:T) #Partimos desde 2, pues la posición 1 ya se ha llenado
  {
    for(j in 1:n) #Se debe usar doble índice porque tenemos n lambdas
    {
      #p(lambda_i|resto) ~ Gamma(alpha+y_i, beta+1)
      lambda[i,j] <- rgamma(1, alpha + y[j], beta[i - 1] + 1)

      #p(beta|resto) ~ Gamma(n*alpha+gamma, delta+sum(lambda_i))
      beta[i] <- rgamma(1, n*alpha + gamma, delta + sum(lambda[i,]))
    }
}
```

En R

```
cbind(y, round(apply(lambda[4000:T,], 2, mean), 2), años)
       y      años
 [1,]  7 6.96 1936
 [2,]  9 8.71 1937
 [3,]  8 7.86 1938
 [4,]  7 6.99 1939
 [5,]  7 6.99 1940
 [6,]  6 6.13 1941
 [7,]  6 6.11 1942
 [8,]  5 5.24 1943
 [9,]  5 5.22 1944
[10,]  7 6.98 1945
[11,]  9 8.72 1946
[12,] 10 9.61 1947
[13,]  8 7.84 1948
[14,]  8 7.86 1949
[15,]  8 7.87 1950
[16,]  7 7.04 1951
```

Muestreador de Gibbs - Datos faltantes

Animales en categorías

Ejemplo de Laird, Dempster and Rubin (1977)

197 animales distribuidos multinomialmente en cuatro categorías

$$x = (x_1, x_2, x_3, x_4) = (125, 18, 20, 34)$$

Con probabilidades:

$$(p_1, p_2, p_3, p_4) = (\frac{1}{2} + \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4})$$

Supongamos que la primera celda, x_1 se divide en dos, y quedan z animales en una y en la otra $x_1 - z$ con probabilidades $\frac{1}{2}$ y $\frac{\theta}{4}$ respectivamente.

Animales en categorías

Ejemplo de Laird, Dempster and Rubin (1977)

197 animales distribuidos multinomialmente en cuatro categorías

$$x = (x_1, x_2, x_3, x_4) = (125, 18, 20, 34)$$

Con probabilidades:

$$(p_1, p_2, p_3, p_4) = (\frac{1}{2} + \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4})$$

Supongamos que la primera celda, x_1 se divide en dos, y quedan z animales en una y en la otra $x_1 - z$ con probabilidades $\frac{1}{2}$ y $\frac{\theta}{4}$ respectivamente.

Los datos completos son $(z, x_1 - z, x_2, x_3, x_4)$ con probabilidades $(\frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4})$.

La verosimilitud de los datos completos es:

$$L(\theta; x, z) \propto (\frac{1}{2})^{z}(\frac{\theta}{4})^{x_1 - z}(\frac{1-\theta}{4})^{x_2+x_3}(\frac{\theta}{4})^{x_4}$$

Notar que los parámetros son θ y z (lo variante).

Animales en categorías

La verosimilitud de los datos completos es:

$$
\begin{aligned}
L(\theta; x, z) &\propto (\frac{1}{2})^{z}(\frac{\theta}{4})^{x_1 - z}(\frac{1-\theta}{4})^{x_2+x_3}(\frac{\theta}{4})^{x_4} \\
&\propto \theta^{x_1+x_4-z}(1-\theta)^{x_2+x_3} \\
&= \theta^{159-z}(1-\theta)^{38}
\end{aligned}
$$

Consideremos como priori que $\theta \sim U(0,1)$ entonces,

$$p(\theta) = 1 \quad \forall \theta \in (0,1)$$

.

Muestreo de Gibbs

Consideremos como priori que $\theta \sim U(0,1)$ entonces,

$$p(\theta) = 1 \quad \forall\theta \in (0,1)$$

Tenemos que la condicional completa es:

$$p(\theta|z,x) \propto \theta^{159-z}(1-\theta)^{38} \cdot 1$$

Muestreo de Gibbs

Consideremos como priori que $\theta \sim U(0,1)$ entonces,

$$p(\theta) = 1 \quad \forall \theta \in (0,1)$$

Tenemos que la condicional completa es:

$$p(\theta|z,x) \propto \theta^{159-z}(1-\theta)^{38} \cdot 1$$

Y por lo tanto

$$\theta|z,x \sim beta(160-z,39)$$

Muestreo de Gibbs

Consideremos como priori que $\theta \sim U(0,1)$ entonces, $p(\theta) = 1 \quad \forall\theta \in (0,1)$.

Tenemos que la condicional completa es:

$$p(\theta|z,x) \propto \theta^{159-z}(1-\theta)^{38} \cdot 1$$

Y por lo tanto

$$\theta|z,x \sim beta(160-z, 39)$$

Muestreo de Gibbs

Nos faltaría conocer $p(z|\theta, x)$. Sabemos que z es una variable de conteo que a lo más podría ser x_1 y además recordemos que la probabilidad asociada a z era $\frac{1}{2}$ versus la probabilidad asociada a $x_1 - z$ era $\frac{\theta}{4}$. Por lo tanto,

$$z|\theta, x \sim Bin(x_1, \frac{\frac{1}{2}}{\frac{1}{2} + \frac{\theta}{4}})$$

Muestreo de Gibbs

Así, el muestreador de Gibbs simula:

$$\theta|z,x \sim beta(160 - z, 39)$$

$$z|\theta,x \sim Bin(x_1, \frac{\frac{1}{2}}{\frac{1}{2} + \frac{\theta}{4}})$$

Implementación

Obtendremos una cadena de largo 1000, partiremos con un valor inicial de $\theta^{(0)} = 0.1$ y para $z^{(0)} = 1$.

En R

```
T <- 1000

theta <- rep(NA, T)

theta[1] <- 0.1 #Valor inicial theta

z <- rep(NA, T)

z[1] <- 1  #Valor inicial z

for(i in 2:T)
  {
    #p(z|resto) ~ bin(x1, (1/2)/(1/2+theta/4))
    z[i] <- rbinom(1, 125, ((1/2)/((1/2) + (theta[i-1]/4))))

    #p(theta|resto) ~ beta(160-z, 39)
    theta[i] <- rbeta(1, 160 - z[i], 39)
}
```

En R

```
mean(z) #zeta: n° de animales que quedó en la primera posición
[1] 95.564

mean(z[200:T]) #Quema
[1] 95.64045

mean(theta)  #theta: parámetro de la multinomial
[1] 0.6205774

mean(theta[200:T]) #Quema
[1] 0.6215774
```