Introducción

El estimador de Gibbs es una técnica para generar una muestra aleatoria de una distribución marginal indirectamente, es decir, sin la necesidad de calcular dicha distribución. Se trata de un caso especial del algoritmo de Metrópolis-Hastings y, por lo tanto, de las cadenas de markov montecarlo (MCMC).

La mayoría de apliaciones de este estimador han sido en estadística bayesiana aunque también es muy útil en estadística clasíca.

El objetivo de este trabajo es entender el funcionamiendo de este estimador explorando casos simples y su aplicación práctica viendo algun ejemplo así como su implementación en el lenguaje R.

Inicialmente vamos a ver la idea básica de este estimador y la motivación para su aplicación.

El estimador de Gibbs. Motivación

Supongamos que tenemos un función de densidad conjunta de una variable multidimensional $f_{XY_1...Yp}$ y estamos interesados en obtener alguna característica de la función de densidad marginal como la media o la varianza. El enfoque más directo y natural sería calcular dicha función de densidad marginal y usar esta distribución para calcular la característica.

$f_X(x) = \int f_{XY_1...Yp}(x, y_1, ..., y_p)dy_1...dy_p$

Sin embargo, hay muchos casos en los cuales la integral es muy dificil de calcular analítica o numéricamente o no disponemos de la distribución conjunta. En estos casos, el muestreador de Gibbs nos ofrece un método alternativo para obtener f(x). El muestreador de Gibbs nos permite generar una muestra de f(x) sin la necesidad de calcular f(x). Así, si simulamos una muestra suficientemente grande, la media o varianza u otra característica de $f(x)$ puede ser calculada haciendo uso de que la función de distribución empirica converge a la función de distribución(por LDGN, montecarlo):

$\lim \frac{1}{m} \sum x_i = \int xf(x)dx = E(X)$

Para entender como se construye este estimador, por razones de simplicidad, vamos a estudiar el caso en el que tenemos dos variables aleatorias (X,Y).

El estimador de Gibbs. Idea. Caso bivariante

Dada una variable aleatoria bidimensional (X, Y). El estimador de Gibbs genera una muestra de $f(x)$ muestreando las distribuciónes condicionadas $f_{X|Y}(x|y)$ y $f_{Y|X}(y|x)$ que son conocidas.

Esto se hace generando una sequencia o muestra de Gibbs $Y_0^{'}, X_0^{'}, Y_1^{'}, X_1^{'}, Y_2^{'}, X_2^{'}...Y_k^{'}, X_k^{'}$

Se debe especificar el valor inicial $Y_0^{'}$, y el resto se calcula iterativamente generando valores a partir de las distribuciónes condicionadas:

$X_j^{'} \sim f(x|Y_j^{'} = y_j^{'})$

$Y_{j+1}^{'} \sim f(y|X_j^{'} = x_j^{'})$

Se puede comprobar que esta secuencia esta basada en propiedades de cadenas de markov. Bajo condiciones generales, la distribución de $X_k^{'}$ converge a $f_X(x)$ cuando $k \to \infty$. Así, para un k suficientemente grande, el valor final $X_k^{'} = x_k^{'}$ será una observación de la muestra de $f_X(x)$.

Existen diferentes maneras para conseguir la convergencia (en distribución) de la muestra de Gibbs para obtener una aproximación de $f(x)$:

Una alternativa seria generar m muestras de Gibbs independientes, de longitud k, y despues usar el valor final de $x_k^{'}$ de cada secuencia. La eleccion del valor de k la discutiremos mas adelante.

Otra alternativa es usar el estimador de Gibbs para estimar $f(x)$ promediando la densidad condicionada final de cada secuencia de Gibbs, es decir, la media de la densidad condicionada sera una aproximación de $f(x)$ y se puede estimar $f(x)$ como sigue

$\hat{f}(x) = \frac{1}{m} \sum f(x|y_i)$

donde $y_1, ..., y_m$ es la secuecia de las observaciónes finales Y de cada secuencia de Gibbs. La teoría detrás de esto es que el valor esperado de la densidad condicional es:

$E[f(x|Y)] = \int f(x|y)f(y)dy = f(x)$

Notar que si la variable X es discreta analogamente podemos estimar las probabilidades marginales de X

$\hat{P}(X = x) = \frac{1}{m} \sum P(X = x|Y_i = y_i)$.

Caso bivariante Ejemplo

Dada la distribución conjunta de una variable bidimensional (X, Y)

$f_{XY}(x, y) = \binom{n}{x} y^{x+\alpha-1}(1-y)^{n-x+\beta-1}, x = 0, 1, ..., n, 0 \leq y \leq 1$

Estamos interesados en calcular alguna característica de la distribución marginal $f(x)$ de X. El estimador de Gibbs nos permite generar una muestra $f(x)$ siguiendo el esquema explicado anteriormente, sabiendo que las distribuciónes condicionadas son

$f(x|y) \sim Binomial(n, y)$, $f(y|x) \sim Beta(x + \alpha, n - x - \beta)$

El siguiente codigo en R muestra la implementación del estimador de Gibbs para generar dicha muestra en este ejemplo

```
Gibbs = function(m,k,n,ini,alpha,beta){
  # Creamos un vector donde guardaremos los valores de la muestra de X
  muestra_x = numeric()
  # Creamos un bucle para generar m secuecnias de gibbs
  for (j in 1:m){
    # Creamos 2 vectores donde guardaremos los valores de las variables X e Y
    # de la secuencia de gibbs
    x = numeric()
    h = numeric()
    # Introducimos el valor inicial de la variable Y
    h[1] = ini
    # Creamos un bucle para calcular la secuencia de gibbs iterativamente
    for (i in 1:k){
      # Utilizando las distribuciónes condicionadas
      x[i] = rbinom(1,n,h[i])
      h[i+1] = rbeta(1,alpha,beta)
    }
    # Anadimos el valor final de la variable X de la secuencia de gibbs
    muestra_x[j] = x[k]
  }
  # Y devolvemos la muestra de X de tamano m generada a partir de m secuencias de gibbs
  muestra_x
}
muestra_Gibbs = Gibbs(500,100,16,0.45,2,4)
```

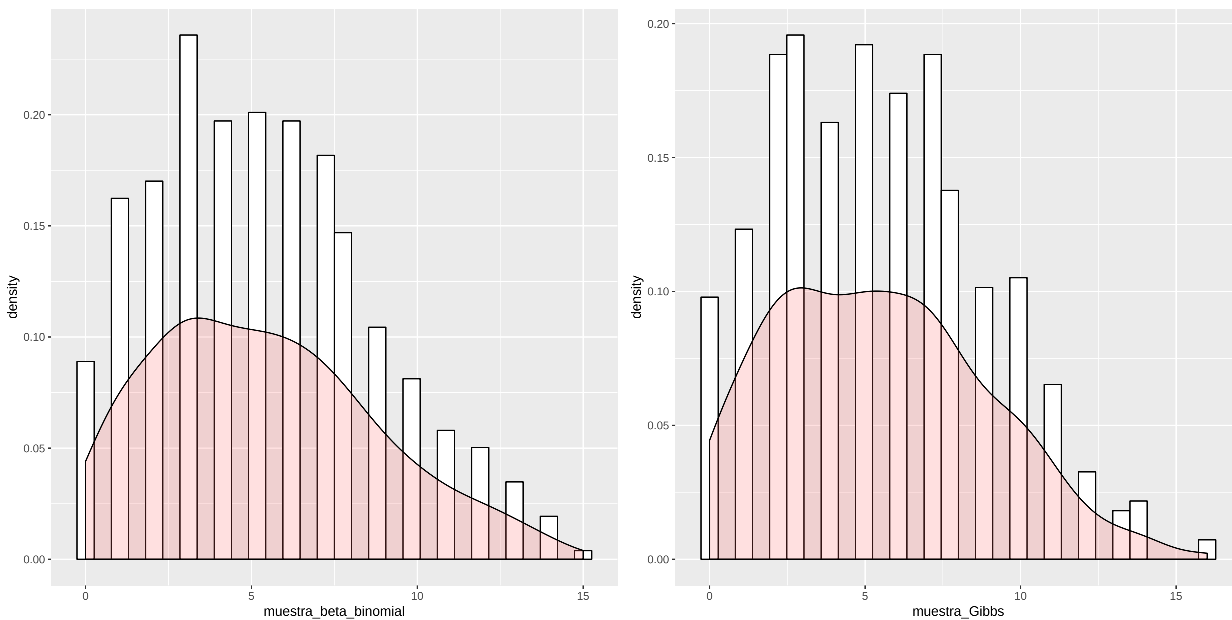
Una observación importante es que si nos damos cuenta, el estimador de Gibbs no es realmente necesario en este ejemplo ya que la distribución marginal $f(x)$ puede ser calculada, conociendo la distribución conjunta, a partir de la definicion de distribución condicionada $f(x) = f(x, y)/f(y|x)$. Procediendo de esta manera obtendriamos que la distribución marginal $f(x)$ es una Beta-Binomial, es decir, $X \sim Beta-Binomial(n, \alpha, \beta)$

A partir de aquí podriamos obtener directamente cualquier característica de $f(x)$ analíticamente o estimandola a través de una muestra generada por $f(x)$.

Por tanto, queda claro que el estimador de Gibbs no es necesario cuando estamos en un contexto bivariante donde la distribución conjunta puede ser calculada.

A pesar de ello este ejemplo es muy ilustrativo para entender el método y nos permite comprobar su eficacia.

El siguiente gráfico muestra 2 histogramas para una muestra de la distribución marginal $f(x)$. La primera es la generada mediante el estimador de Gibbs (tomando el valor final de la variable X de cada secuencia de Gibbs) y la segunda es una muestra obtenida directamente de la distribución Beta-Binomial.



Como se puede obserbar, los dos histogramas son muy similares y es un claro ejemplo de que el esquema de generacion de variables aleatorias del estimador de Gibbs es una alternativa bastante buena a la generacion de variables alteatorias a partir de la distribución marginal.

Como notamos anteriormente, el estimador de Gibbs no es necesario cuando podemos obtener la distribución conjunta. Sin embrago, este estimador puede ser indispensable en situaciones donde la distribución conjunta y por ende la distribución marginal no pueden ser calculadas.

Caso bivariante Ejemplo

Si queremos usar el estimador de Gibbs para estimar $f(x)$ promediando la densidad condicionada final de cada secuencia de Gibbs, es decir, tomando la media de la densidad condicionada será una aproximación de $f(x)$ y se puede estimar $f(x)$ como sigue

$\hat{f}(x) = \frac{1}{m} \sum f(x|y_i)$

La implementaciòn en R es muy similar a la anterior pero ahora guardamos los valores finales de la variable Y de cada secuencia de Gibbs para despues promediar la densidad condicionada.

```
Gibbs = function(m,k,n,ini,alpha,beta){
  # Creamos un vector donde guardaremos los valores finales de la variable Y
  # de cada secuencia de gibbs
  muestra_y=numeric()
  # Creamos un bucle para generar m secuencias de gibbs
  for (j in 1:m){
    # Creamos 2 vectores donde guardaremos los valores de las variables X e Y
```

```r
    # de la secuencia de gibbs
    x = numeric()
    h = numeric()
    # Introducimos el valor inicial de la variable Y
    h[1] = ini
    # Creamos un bucle para calcular la secuencia de gibbs iterativamente
    for (i in 1:k){
      # Utilizando las distribucións condicionadas
      x[i] = rbinom(1,n,h[i])
      h[i+1] = rbeta(1,alpha,beta)
    }
    # Anadimos el valor final de la variable Y de la secuencia de gibbs
    muestra_y[j]=h[k]
  }
  # Y devolvemos la muestra de Y de tamano m generada a partir de m secuencias de gibbs
  muestra_y
}
muestra_y_sec_Gibbs = Gibbs(500,100,16,0.45,2,4)

# Creamos un vector donde guardaremos los valores de la muestra de X a partir
# de la función de densidad estimada
f_est = numeric()
# Creamos un bucle para estimar la función de densidad promediando la densidad condicionada
for (x in seq(1,16,by=1)){
  # Calculamos la media de la densidad condionada usando los valores finales de Y de cada secuencia
  media_condicionada = mean(dbinom(x,16,muestra_y_sec_Gibbs))
  f_est[x]=media_condicionada
}
```

Caso bivariante Detalles

El estimador de Gibbs puede pensarse como la implementación próctica del hecho de que conocer las distribuciónes condicionadas es suficiente para determinar la distribución conjunta(si existe). Supongamos que tenemos dos variables aleatorias (X, Y) y conocemos las densidades condicionadas $f_{X|Y}(x|y)$ y $f_{Y|X}(y|x)$.

Por definición

$f_X(x) = \int f_{XY}(x,y)dy$ donde $f_{XY}(x,y)$ es la distribución conjunta(desconocida). Si usamos ahora la definición de distribución condicionada $f(x) = f(x,y)/f(y|x)$ tenemos que:

$f_X(x) = \int f_{X|Y}(x|y) f_Y(y)dy$

y si procedemos analogamente podemos sustituir $f_Y(y)$ en la ecuacion obteniendo que

$f_X(x) = \int f_{X|Y}(x|y)(\int f_{Y|X}(y|t)f_X(t)dt)dy = \int h(x,t)f_X(t)dt$

donde $h(x,t) = \int f_{X|Y}(x|y)f_{Y|X}(y|t)dy$

Esta ecuación representa el esquema del estimador de Gibbs. Así, cuando $k \to \infty$:

$f_{x_k|x_0}(x|x_0) \to f_X(x)$ y

$f_{x_k|x_{k-1}}(x|t) \to h(x,t)$

Aunque la distribución conjunta de X e Y determina todas las distribuciónes condicionadas y marginales, no siempre se da el caso en el que un conjunto de distribuciónes condicionadas determinan una distribución marginal adeacuada(y por ende una distribución conjunta adecuada).

Caso multivariante

Cuando el numero de variables es mayor de 2, las relaciones entre las distrubuciones condicionadas, las marginales y la conjunta se vuelven mas complejas.

Hay diferentes maneras para conseguir una distribución marginal ya que se pueden usar diferentes conjuntos de distribuciónes condicionadas para calcular la distribución marginal de interés. Estas metodologías son parte de técnicas generales de muestreo por substitución.

En el caso de dos variables todos los algoritmos de muestreo por sustitución son los mismos. El caso de tres variables es suficiente para mostrar las diferencias entre estos algoritmos.

Supongamos que queremos calcular la distribución marginal $f_X(x)$ en un problema con variables aleatorias X, Y y Z. Podemos considerar el par (Y, Z) como una variable aleatoria particular, y analogamente al caso bivariante tendriamos:

$f_X(x) = \int[\int\int f_{X|YZ}(x|y,z)f_{YZ|X}(y,z|t)dydz]f_X(t)dt$

Ahora muestreando iterativamente las distribuciónes condicionadas $f_{X|YZ}$ y $f_{YZ|X}$ siguiendo el siguiente esquema

$X_j^{'} \sim f(x|Y_j^{'} = y_j^{'}, Z_j^{'} = z_j^{'})$

$Y_{j+1}^{'} \sim f(y|X_j^{'} = x_j^{'}, Z_j^{'} = z_j^{'})$

$Z_{j+1}^{'} \sim f(z|X_j^{'} = x_j^{'}, Y_j^{'} = y_j^{'})$

Se genera la siguiente secuencia de Gibbs

$Y_0^{'}, Z_0^{'}, X_0^{'}, Y_1^{'}, Z_1^{'}, X_1^{'}, Y_2^{'}, Z_2^{'}, X_2^{'}...Y_k^{'}, Z_k^{'}, X_k^{'}$

con la propiedad tal que tomando un k suficientemente grande, el punto $X_k^{'} = x_k^{'}$ es una observación de la muestra de $f_X(x)$.

Una característica que define al estimador de Gibbs es que siempre usa el conjunto completo de distribuciones condicionadas univariantes para definir la iteración ya que este conjunto es suficiente para determinar la distribucion conjunta(y alguna marginal).

Aplicación

El estimador de Gibbs se ha aplicado en multitud de problemas en estadística clásica y bayesiana. Para la estadística bayesiana el estimador de Gibbs es principalmente usado para generar distribuciones a posteriori mientras que en estadística clásica el mayor uso es para el cálculo de la función de verosimilitud y características de estimadores de verosimilitud.

El gran poder del estimador de Gibbs es que es mas efectivo cuanto mayor se incremente la dimension del problema. Esto es debido a que el estimador de Gibbs nos permite 'esquivar' la tarea de calcular una integral, que puede ser una tarea complicada en grandes dimensiones.

El ultimo valor de este estimador es su potencial práctico. Muchos trabajos diferentes han aplicado este estimador desde aplicaciones para los modelos lineales generalizados hasta para modelos mixtos, etc.

Referencias:

- The American Statistician, Vol. 46, No. 3. (Aug., 1992), pp. 167-174. George Casella; Edward I. George