

Tema 7: Métodos de aproximación para la estimación Bayesiana

Conchi Ausín

Departamento de Estadística

Universidad Carlos III de Madrid

concepcion.ausin@uc3m.es

CESGA, Noviembre 2012

Introducción

Ya hemos dicho que en la mayoría de los casos no es fácil calcular analíticamente la distribución a posteriori:

$$\pi(\theta | \mathbf{x}) = \frac{f(\mathbf{x} | \theta) \pi(\theta)}{\int f(\mathbf{x} | \theta) \pi(\theta) d\theta},$$

ni tampoco la media y varianza a posteriori, las distribuciones predictivas, intervalos creíbles, etc.

Sin embargo, podemos usar técnicas de aproximación como son:

- Integración numérica
- Aproximaciones asintóticas
- Simulación Monte Carlo:
 - con métodos directos
 - con cadenas de Markov

Contenidos

1. Integración numérica
2. Aproximaciones asintóticas
3. Aproximaciones Monte Carlo
4. Comentarios finales

Integración numérica

Consiste en aproximar una integral usando una suma ponderada de algunos valores del integrando,

$$\int_a^b f(x) dx \approx \sum_{i=1}^n \omega_i f_i$$

donde $f_i = f(x_i)$ y ω_i es el peso asociado a x_i .

Supongamos que dividimos el intervalo $[a, b]$ en n puntos, $x_i = x_1 + ih$, para $i = 1, \dots, n-1$, con $x_1 = a$, $x_n = b$ y $h = (b-a)/n$.

Lo más simple es usar la **regla trapezoidal** que aproxima $f(x)$ en $[x_1, x_2]$ con una línea recta,

$$\int_{x_1}^{x_2} f(x) dx \approx \frac{(f_1 + f_2)h}{2},$$

y repetirlo para cada par de puntos del intervalo.

Integración numérica

En general, es mejor usar la **regla de Simpson**, que usa una polinomio de orden 2 uniendo los puntos (x_1, f_1) , (x_2, f_2) y (x_3, f_3) ,

$$\int_{x_1}^{x_3} f(x) dx \approx \frac{h}{3} (f_1 + 4f_2 + f_3).$$

Se puede usar también esta regla repetidamente para aproximar la integral completa en $[x_1, x_n]$, (n debe ser impar),

$$\int_{x_1}^{x_n} f(x) dx = \frac{h}{3} (f_1 + 4f_2 + 2f_3 + 4f_4 + 2f_5 + \dots + 4f_{n-1} + f_n).$$

Ejemplo 7.1. Aproximar la constante de integración de la distribución a posteriori de θ en el ejemplo de las caras del tema anterior,

$$\pi(\theta | \mathbf{x}) \propto \theta^9 (1 - \theta)^3, \quad 0 < \theta < 1.$$

Integración numérica

Comentarios:

- Existen otros procedimientos de integración numérica más sofisticados que la regla de Simpson simple, como son la regla de Simpson adaptada y otros basados en polinomios ortogonales, como la **cuadratura Gaussiana**.
- El principal problema con la integración numérica aparece cuando el **número de parámetros es grande**. Según aumenta la dimensión de la integral, el número de funciones a evaluar para obtener una buena aproximación aumenta considerablemente.

Aproximaciones asintóticas

Dados los datos, $\mathbf{x} = \{x_1, \dots, x_n\}$, si el tamaño muestral, n , es muy grande, la influencia de la distribución a priori será muy pequeña.

Por ejemplo, si $X|\mu, \sigma^2 \sim N(\mu, \sigma^2)$ y usamos una distribución conjugada a priori, entonces,

$$E[\mu | \mathbf{x}] = \frac{cm + n\bar{x}}{c + n} \rightarrow \bar{x},$$

cuando $n \rightarrow \infty$.

De hecho, las propiedades de las distribuciones a posteriori son serán similares a las de los estimadores máximo verosímiles cuando $n \rightarrow \infty$.

Concretamente, si la distribución a posteriori es unimodal y prácticamente simétrica, se puede aproximar por una distribución normal cuando el tamaño muestral es muy grande.

Aproximaciones asintóticas

Dado $X|\theta \sim f(\cdot|\theta)$, y una a priori $\pi(\theta)$, cuando $n \rightarrow \infty$, se aproximan:

1. $\theta | \mathbf{x} \approx \mathcal{N}(E[\theta|\mathbf{x}], V[\theta|\mathbf{x}])$.
2. $\theta | \mathbf{x} \approx \mathcal{N}(\hat{\theta}, I(\hat{\theta})^{-1})$, donde $\hat{\theta}$ es la moda a posteriori y

$$I(\theta) = -\frac{d^2}{d\theta^2} \log(\pi(\theta|\mathbf{x})).$$

3. $\theta | \mathbf{x} \approx \mathcal{N}(\hat{\theta}_{MV}, I^*(\hat{\theta}_{MV})^{-1})$, donde $\hat{\theta}_{MV}$ es el estimador MV de θ y,

$$I^*(\theta) = -\frac{d^2}{d\theta^2} \log(f(\mathbf{x}|\theta)).$$

4. $\theta | \mathbf{x} \approx \mathcal{N}(\hat{\theta}_{MLE}, I^{**}(\hat{\theta}_{MLE})^{-1})$, donde,

$$I^{**}(\theta) = -nE_X \left[\frac{d^2}{d\theta^2} \log(f(X|\theta)) \right].$$

Aproximaciones asintóticas

Ejemplo 7.2. Usar algunas aproximaciones asintóticas para la distribución a posteriori de la probabilidad de cara, θ :

- Cuando se han observado 20 caras en 30 lanzamientos.
- Cuando se han observado 2 caras en 3 lanzamientos.

Comentarios:

- Claramente, las aproximaciones asintóticas son sólo válidas cuando el tamaño muestral es grande, lo cual no siempre es factible en la práctica.
- Otras alternativas más útiles para aproximar medidas a posteriori son los métodos de simulación Monte Carlo.

Aproximación Monte Carlo

Supongamos que sabemos cómo simular una muestra, $\{\theta_1, \dots, \theta_M\}$, de valores de la distribución a posteriori, $\pi(\theta | \mathbf{x})$. Entonces, podemos aproximar la media a posteriori de cualquier función de los parámetros, θ , mediante:

$$E[\theta | \mathbf{x}] = \int g(\theta) \pi(\theta | \mathbf{x}) d\theta \approx \frac{1}{M} \sum_{m=1}^M \theta_m$$

Análogamente, podemos aproximar otros momentos a posteriori, probabilidades a posteriori :

$$\Pr(\theta < c | \mathbf{x}) \approx \frac{1}{M} \sum_{m=1}^M I(\theta_m < c),$$

o intervalos creíbles de nivel $(1 - \alpha)\%$ usando los cuantiles de la muestra $\{\theta_1, \dots, \theta_M\}$.

Aproximación Monte Carlo

Ejemplo 7.3. Considerar los datos de longitudes del caparazón de 200 cangrejos del tema anterior. Aproximar mediante simulación Monte Carlo:

- La probabilidad a posteriori de que la media de la longitud sea mayor de 30 milímetros.
- Un intervalo creíble al 95 % para dicha media.
- Comparar los resultados que obtuvimos analíticamente.

Comentarios:

- En la mayoría de los casos sin embargo, no hay forma directa de simular de $\pi(\theta \mid \mathbf{x})$. En estos casos, se puede usar otros métodos MC que veremos a continuación: el muestreo de importancia y el método de rechazo, etc.

Muestreo de importancia

A menudo es difícil simular de la distribución a posteriori pero se puede simular de otra densidad $h(\theta)$ que llamaremos **función de importancia**. Entonces, se puede aproximar:

$$\begin{aligned} E[\theta | \mathbf{x}] &= \int \theta \pi(\theta | \mathbf{x}) d\theta = \int \theta \frac{\pi(\theta) f(\mathbf{x} | \theta)}{\int \pi(u) f(\mathbf{x} | u) du} d\theta \\ &= \int \theta \frac{\frac{\pi(\theta) f(\mathbf{x} | \theta)}{h(\theta)} h(\theta)}{\int \frac{\pi(u) f(\mathbf{x} | u)}{h(u)} h(u) du} d\theta = \frac{\int \theta \omega(\theta) h(\theta) d\theta}{\int \omega(u) h(u) du} d\theta, \end{aligned}$$

donde,

$$\omega(\theta) = \frac{\pi(\theta) f(\mathbf{x} | \theta)}{h(\theta)}.$$

Luego, usando una muestra $\{\theta_1, \dots, \theta_M\}$ simulada de $h(\theta)$, se puede aproximar:

$$E[\theta | \mathbf{x}] \approx \frac{\sum_{m=1}^M \omega(\theta_m) \theta_m}{\sum_{m=1}^M \omega(\theta_m)}$$

Muestreo de importancia

Ejemplo 7.4. Usando una función de importancia uniforme, aproximar la constante de integración y la media de la distribución a posteriori de θ en el ejemplo de las caras del tema anterior,

$$\pi(\theta | \mathbf{x}) \propto \theta^9(1 - \theta)^3, \quad 0 < \theta < 1.$$

Comentarios:

- La eficiencia del algoritmo depende mucho de la función de importancia. En particular, $h(\theta)$ tiene que tener colas más pesadas que $\pi(\theta | \mathbf{x})$ para que la varianza del estimador de $E[\theta | \mathbf{x}]$ sea finita.
- Además, la función de importancia debe parecerse a la distribución a posteriori. Por ejemplo, si el centro de $\pi(\theta | \mathbf{x})$ está en la cola de $h(\theta)$, la mayoría de los pesos serán muy pequeños y la estimación de la integral sólo dependerá unos pocos pesos grandes.

Muestreo de importancia

El muestreo de importancia no proporciona en principio una muestra de la distribución a posteriori, sino estimaciones de los momentos a posteriori.

Sin embargo, se puede obtener una muestra aproximada de $\pi(\theta | \mathbf{x})$ por submuestreo mediante el [algoritmo SIR](#). Para ello, se normalizan primero los pesos para que sumen uno:

$$\omega_m = \frac{\omega(\theta_m)}{\sum_{m=1}^M \omega(\theta_m)},$$

luego, se simula una muestra aproximada, $\tilde{\theta}$, de tamaño $J < M$ tomando $\tilde{\theta}_j = \theta_m$ con probabilidad ω_m para $m = 1, \dots, M$ y $j = 1, \dots, J$.

Ejemplo 7.5. Obtener una remuestra de tamaño $J = 1000$ de la distribución a posteriori de θ a partir de la muestra de importancia obtenida en el ejemplo anterior.

Método del rechazo

Otro método que permite simular valores de $\pi(\theta \mid \mathbf{x})$ a partir de valores simulados de otra densidad $h(\theta)$, denominada **densidad candidata**, que verifique que $\pi(\theta \mid \mathbf{x}) < Mh(\theta)$ para alguna constante $M > 0$:

Para $m = 1, \dots, M$:

1. Simular $\tilde{\theta}_m \sim h(\theta)$.
2. Simular $u_m \sim \mathcal{U}(0, 1)$.
3. Si $Mu_m h(\tilde{\theta}_m) < \pi(\tilde{\theta}_m \mid \mathbf{x})$, tomar $\theta_m = \tilde{\theta}_m$.
4. Si no, ir a 1.

Método del rechazo

Ejemplo 7.6. Considerar los datos de longitudes del caparazón de 200 cangrejos del tema anterior. Suponer que se impone a priori que la media de la longitud del caparazón debe ser mayor que 31 milímetros. Simular valores de la distribución a posteriori de la media mediante el método de rechazo.

Comentarios:

- El gran problema del método del rechazo es encontrar una buena densidad candidata que sólo rechace algunos de los valores simulados.
- Se puede demostrar que la **probabilidad de aceptar** un valor simulado es $1/M$.
- Existen varios refinamientos del algoritmo de rechazo tales como los métodos de envoltura (*envelope methods*) o el algoritmo ARS (*adaptive rejection sampling*), etc.

Comentarios finales

- En este tema hemos visto varios métodos para aproximar la distribución a posteriori cuando no es analíticamente tratable.
- Sin embargo, hemos visto que todos estos métodos ofrecen distintos problemas:
 - Los métodos de integración numérica son muy costosos.
 - Los de aproximación asintótica no son útiles en muestras pequeñas.
 - Los métodos de simulación Monte Carlo directa requieren poder simular de la distribución a posteriori o de una densidad candidata parecida lo cual casi nunca es fácil en la práctica.
- En el tema siguiente, veremos los métodos MCMC, que son también métodos de simulación Monte Carlo para generar valores de la distribución a posteriori de los parámetros.
- Sin embargo, los métodos MCMC proporcionan procedimientos para los casos en los que la forma de la posteriori es casi desconocida y/o el número de parámetros a estimar es muy elevado.