Fundamentos de Estadística Bayesiana

Juan Carlos Martínez-Ovando

ITAM

[Mini] Taller de Métodos Numéricos y Estadísticos en Cosmología 2017
Cinvestav, CDMX
6 de abril de 2017

1. Incertidumbre y aleatoriedad

1.1. Consideraciones

Incertidumbre $\leftrightarrow$ Desconocimiento

Datos

El proceso que genera los **datos**

$$\{x_1, \ldots, x_n\},$$

es desconocido.

Aleatoriedad intrínseca

Nuestro desconocimiento acerca de los **datos** lo manifestamos suponiendo **aleatoriedad intrínseca** empleando un modelo

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

1.1. Consideraciones

Incertidumbre $\leftrightarrow$ Desconocimiento

Datos

El proceso que genera los **datos**

$$\{x_1, \ldots, x_n\},$$

es desconocido.

Aleatoriedad intrínseca

Nuestro desconocimiento acerca de los **datos** lo manifestamos suponiendo **aleatoriedad intrínseca** empleando un modelo

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

1.1. Consideraciones

Incertidumbre $\leftrightarrow$ Desconocimiento

Datos

El proceso que genera los **datos**

$$\{x_1, \ldots, x_n\},$$

es desconocido.

Aleatoriedad intrínseca

Nuestro desconocimiento acerca de los **datos** lo manifestamos suponiendo **aleatoriedad intrínseca** empleando un modelo

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

1.2. Dependencia

Es un supuesto atribuible al modelo $\mathbb{P}$ y **no** a los **datos**.

Independencia

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{j=1}^{n} \mathbb{P}(X_j \leq x_j).$$

Intercambiabilidad

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_{\sigma(1)} \leq x_1, \ldots, X_{\sigma(n)} \leq x_n),$$

para toda permutación

$$\{\sigma(1), \ldots, \sigma(n)\} \text{ de } \{1, \ldots, n\}.$$

1.2. Dependencia

Es un supuesto atribuible al modelo $\mathbb{P}$ y **no** a los **datos**.

Independencia

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{j=1}^{n} \mathbb{P}(X_j \leq x_j).$$

Intercambiabilidad

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_{\sigma(1)} \leq x_1, \ldots, X_{\sigma(n)} \leq x_n),$$

para toda permutación

$$\{\sigma(1), \ldots, \sigma(n)\} \text{ de } \{1, \ldots, n\}.$$

1.2. Dependencia

Es un supuesto atribuible al modelo $\mathbb{P}$ y **no** a los **datos**.

Independencia

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{j=1}^{n} \mathbb{P}(X_j \leq x_j).$$

Intercambiabilidad

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_{\sigma(1)} \leq x_1, \ldots, X_{\sigma(n)} \leq x_n),$$

para toda permutación

$$\{\sigma(1), \ldots, \sigma(n)\} \text{ de } \{1, \ldots, n\}.$$

2. Subjetividad y modelación

2.1. Subjetividad

Intercambiabilidad *supone* que el orden en el cual los datos son recolectados/observados es indiferente.

Teorema de representación

Atribuible a Bruno de Finetti, bajo intercambiabilidad

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_\Theta \prod_{i=1}^n F_X(x_i|\theta)\Pi(d\theta)$$

- $F_X(x_i|\theta)$ es una distribución de probabilidades
- θ es un objeto estocástico **no observable** común a todos los datos (a.k.a. parámetro)
- $\Pi(\theta)$ es una medida de probabilidad

2.1. Subjetividad

Intercambiabilidad *supone* que el orden en el cual los datos son recolectados/observados es indiferente.

Teorema de representación

Atribuible a Bruno de Finetti, bajo intercambiabilidad

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_\Theta \prod_{i=1}^{n} F_X(x_i|\theta)\Pi(d\theta)$$

- $F_X(x_i|\theta)$ es una distribución de probabilidades
- θ es un objeto estocástico **no observable** común a todos los datos (a.k.a. parámetro)
- $\Pi(\theta)$ es una medida de probabilidad

2.2. Juicio inicial

Teorema de representación

El Teorema de Representación es de **existencia** más **no de unicidad**

Probabilidades subjetivas

- La **aleatoriedad intrínseca** no se interpreta como un **límite de frecuencias**
- Sino, **aleatoriedad intrínseca** es un **juicio individual subjetivo** (en función de nuestro conocimiento/desconocimiento)

Distribución inicial

- La distribución $\Pi(\theta)$ se interpreta como el juicio o creencia individual acerca del **modelo** antes de observar los **datos**.
- Aunque $\Pi(\theta)$ esta definida sobre Θ, en realidad representa una medida de probabilidad sobre la colección de modelos $F(x|\theta)$.

2.2. Juicio inicial

Teorema de representación

El Teorema de Representación es de **existencia** más **no de unicidad**

Probabilidades subjetivas

- La **aleatoriedad intrínseca** no se interpreta como un **límite de frecuencias**
- Sino, **aleatoriedad intrínseca** es un **juicio individual subjetivo** (en función de nuestro conocimiento/desconocimiento)

Distribución inicial

- La distribución $\Pi(\theta)$ se interpreta como el juicio o creencia individual acerca del **modelo** antes de observar los **datos**.
- Aunque $\Pi(\theta)$ esta definida sobre Θ, en realidad representa una medida de probabilidad sobre la colección de modelos $F(x|\theta)$.

2.2. Juicio inicial

Teorema de representación

El Teorema de Representación es de **existencia** más **no de unicidad**

Probabilidades subjetivas

- La **aleatoriedad intrínseca** no se interpreta como un **límite de frecuencias**
- Sino, **aleatoriedad intrínseca** es un **juicio individual subjetivo** (en función de nuestro conocimiento/desconocimiento)

Distribución inicial

- La distribución $\Pi(\theta)$ se interpreta como el juicio o creencia individual acerca del **modelo** antes de observar los **datos**.
- Aunque $\Pi(\theta)$ esta definida sobre Θ, en realidad representa una medida de probabilidad sobre la colección de modelos $F(x|\theta)$.

2.3. Ejemplo: Bernoulli-beta

Variables observables

Supongamos que los **datos** son tales que

$$x_i = \begin{cases} 1 & \text{con base en la ocurrencia de un evento} \\ 0 & \text{no ocurrencia del evento} \end{cases}$$

Desconocimiento/aleatoriedad

Desconocer el **mecanismo generador** de x_is induce **aleatoriedad intrínseca**

$$f(x_i = j) = \begin{cases} \theta & \text{si } j = 1 \\ (1-\theta) & \text{si } j = 0 \end{cases}$$

con $0 \leq \theta \leq 1$.

Distribución Bernoulli

Lo anterior es **equivalente** a suponer que los datos son **generados** con una distribución Bernoulli, i.e.

$$x|\theta \sim F(x|\theta) = Ber(x_i|\theta).$$

2.3. Ejemplo: Bernoulli-beta

Variables observables

Supongamos que los **datos** son tales que

$$x_i = \begin{cases} 1 & \text{con base en la ocurrencia de un evento} \\ 0 & \text{no ocurrencia del evento} \end{cases}$$

Desconocimiento/aleatoriedad

Desconocer el **mecanismo generador** de x_is induce **aleatoriedad intrínseca**

$$f(x_i = j) = \begin{cases} \theta & \text{si } j = 1 \\ (1-\theta) & \text{si } j = 0 \end{cases}$$

con $0 \leq \theta \leq 1$.

Distribución Bernoulli

Lo anterior es **equivalente** a suponer que los datos son **generados** con una distribución Bernoulli, i.e.

$$x|\theta \sim F(x|\theta) = Ber(x_i|\theta).$$

2.3. Ejemplo: Bernoulli-beta

Variables observables

Supongamos que los **datos** son tales que

$$x_i = \begin{cases} 1 & \text{con base en la ocurrencia de un evento} \\ 0 & \text{no ocurrencia del evento} \end{cases}$$

Desconocimiento/aleatoriedad

Desconocer el **mecanismo generador** de x_is induce **aleatoriedad intrínseca**

$$f(x_i = j) = \begin{cases} \theta & \text{si } j = 1 \\ (1-\theta) & \text{si } j = 0 \end{cases}$$

con $0 \leq \theta \leq 1$.

Distribución Bernoulli

Lo anterior es **equivalente** a suponer que los datos son **generados** con una distribución Bernoulli, i.e.

$$x|\theta \sim F(x|\theta) = Ber(x_i|\theta).$$

2.3. Ejemplo: Bernoulli-beta

Configuración

Cada valor específico de θ en $(0, 1)$ define un modelo de probabilidad particular, por ejemplo:

- M1 $Ber(x|0{,}1)$
- M2 $Ber(x|0{,}33)$
- M3 $Ber(x|0{,}999)$
- M4 $Ber(x|0{,}001)$
- M5 etc.

Juicio inicial

Definido como la **creencia** acerca de la **plausibilidad** de cada configuración para describir los **datos** que estan por observarse.

Posibles alternativas:

- $\Pi_1(\theta) = \sum_{k=1}^{K} \alpha_k \delta_{\theta_k}(\theta)$ para un conjunto de θ_ks en $(0, 1)$
- $\Pi_2(\theta) = U(\theta|0, 1)$
- $\Pi_3(\theta) = Be(\theta|a, b)$
- entre muchas otras,

2.3. Ejemplo: Bernoulli-beta

Configuración

Cada valor específico de θ en $(0, 1)$ define un modelo de probabilidad particular, por ejemplo:

- M1 $Ber(x|0{,}1)$
- M2 $Ber(x|0{,}33)$
- M3 $Ber(x|0{,}999)$
- M4 $Ber(x|0{,}001)$
- M5 etc.

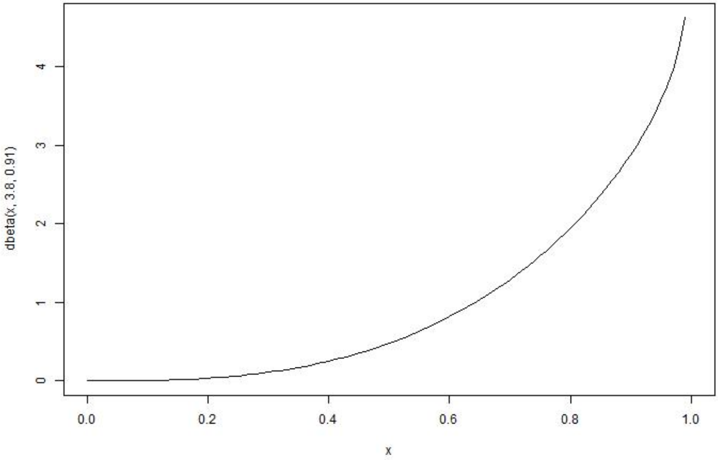
Juicio inicial

Definido como la **creencia** acerca de la **plausibilidad** de cada configuración para describir los **datos** que estan por observarse.
Posibles alternativas:

- $\Pi_1(\theta) = \sum_{k=1}^{K} \alpha_k \delta_{\theta_k}(\theta)$ para un conjunto de θ_ks en $(0, 1)$
- $\Pi_2(\theta) = U(\theta|0, 1)$
- $\Pi_3(\theta) = Be(\theta|a, b)$
- entre muchas otras,

2.3. Ejemplo: Bernoulli-beta

Figura: Una posible distribución $\Pi(\theta)$



2.3. Ejemplo: Bernoulli-beta

Reflexión

Recordemos:

- Los **datos** x_is son **observables**
- La variable (**parámetro**) θ **no es observable**.
- $\Pi(\theta)$ **no es única**, sino **subjetiva**

Modelación

Así, el modelo subjetivo es una medida de probabilidad conjunta de **observables** y **no observables**,

$$
\begin{aligned}
\mathbb{P}(x_1, \ldots, x_n, \theta) &= F(x_1, \ldots, x_n|\theta) \times \Pi(\theta) \\
&= \prod_{i=1}^{n} F(x_i|\theta) \times \Pi(\theta)
\end{aligned}
$$

2.3. Ejemplo: Bernoulli-beta

Reflexión

Recordemos:

- Los **datos** x_is son **observables**
- La variable (**parámetro**) θ **no es observable**.
- $\Pi(\theta)$ **no es única**, sino **subjetiva**

Modelación

Así, el modelo subjetivo es una medida de probabilidad conjunta de **observables** y **no observables**,

$$\begin{aligned}\mathbb{P}(x_1, \ldots, x_n, \theta) &= F(x_1, \ldots, x_n|\theta) \times \Pi(\theta) \\ &= \prod_{i=1}^{n} F(x_i|\theta) \times \Pi(\theta)\end{aligned}$$

3. Aprendizaje y prediccón

3.1. Aprendizaje

Aprendizaje bayesiano/inferencia

El **aprendizaje bayesiano**, *con base en un conjunto de* **datos**, x_i*s consiste en la actualización de* $\Pi(\theta)$ *condicional en los* **datos**, *i.e.*

$$\Pi(\theta|x_1,\ldots,x_n) = \frac{F(x_1,\ldots,x_n|\theta)\Pi(\theta)}{\mathbb{P}(x_1,\ldots,x_n)}$$

donde

$$\mathbb{P}(x_1,\ldots,x_n) = \int_{\Theta} F(x_1,\ldots,x_n|\theta)\Pi(d\theta)$$

Las ecuaciones anteriores se refieren al **Teorema de Bayes** de probabilidades inversas.

Comentario

- El denominador en $\Pi(\theta|x_1,\ldots,x_n)$, i.e. $\mathbb{P}(x_1,\ldots,x_n)$ se conoce como la constante de normalización de la distribución actualizada para θ
- Generalmente, la constante de normalización no puede obtenerse de manera analiítica cerrada, pero puede aproximarse empleando **métodos numéricos**.

3.1. Aprendizaje

Aprendizaje bayesiano/inferencia

El **aprendizaje bayesiano**, *con base en un conjunto de* **datos**, x_i*s consiste en la actualización de* $\Pi(\theta)$ *condicional en los* **datos**, *i.e.*

$$\Pi(\theta|x_1,\ldots,x_n) = \frac{F(x_1,\ldots,x_n|\theta)\Pi(\theta)}{\mathbb{P}(x_1,\ldots,x_n)}$$

donde

$$\mathbb{P}(x_1,\ldots,x_n) = \int_\Theta F(x_1,\ldots,x_n|\theta)\Pi(d\theta)$$

Las ecuaciones anteriores se refieren al **Teorema de Bayes** de probabilidades inversas.

Comentario

- El denominador en $\Pi(\theta|x_1,\ldots,x_n)$, i.e. $\mathbb{P}(x_1,\ldots,x_n)$ se conoce como la constante de normalización de la distribución actualizada para θ
- Generalmente, la constante de normalización no puede obtenerse de manera analíitica cerrada, pero puede aproximarse empleando **métodos numéricos**.

3.2. Ejemplo: Bernoulli-beta

Especificación

Continuemos con el modelo Bernoulli-beta del ejemplo 2.3.

$$
\begin{aligned}
f(x_1, \ldots, x_n|\theta) &= \prod_{i=1}^{n} Ber(x_i|\theta) \\
&\propto \theta^{\#\{x_i : x_i = 1\}} (1-\theta)^{\#\{x_i : x_i = 0\}}
\end{aligned}
$$

y

$$
\begin{aligned}
\pi(\theta) &= Be(\theta|a, b) \\
&\propto \theta^{a-1}(1-\theta)^{b-1}\mathbb{I}_{(0,1)}(\theta).
\end{aligned}
$$

Aprendizaje bayesiano

Así, la distribución actualizada para θ es,

$$
\pi(\theta|x_1, \ldots, x_n) \propto \theta^{\#\{x_i : x_i = 1\}+a-1}(1-\theta)^{\#\{x_i : x_i = 0\}+b-1}\mathbb{I}_{(0,1)}(\theta).
$$

3.2. Ejemplo: Bernoulli-beta

Especificación

Continuemos con el modelo Bernoulli-beta del ejemplo 2.3.

$$\begin{aligned} f(x_1,\ldots,x_n|\theta) &= \prod_{i=1}^{n} Ber(x_i|\theta) \\ &\propto \theta^{\#\{x_i:x_i=1\}}(1-\theta)^{\#\{x_i:x_i=0\}} \end{aligned}$$

y

$$\begin{aligned} \pi(\theta) &= Be(\theta|a,b) \\ &\propto \theta^{a-1}(1-\theta)^{b-1}\mathbb{I}_{(0,1)}(\theta). \end{aligned}$$

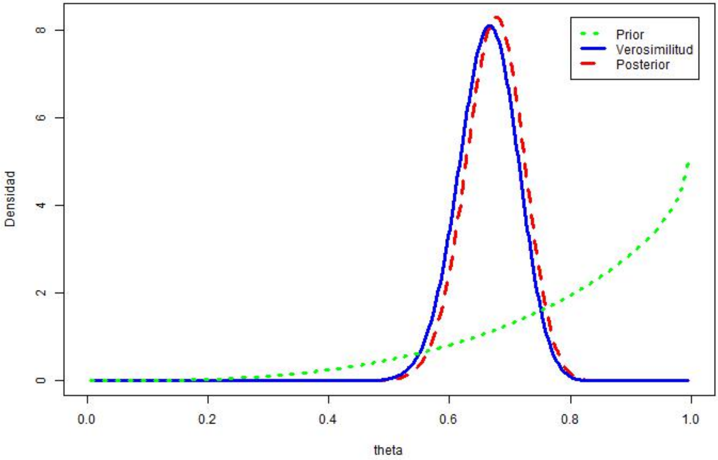
Aprendizaje bayesiano

Así, la distribución actualizada para θ es,

$$\pi(\theta|x_1,\ldots,x_n) \propto \theta^{\#\{x_i:x_i=1\}+a-1}(1-\theta)^{\#\{x_i:x_i=0\}+b-1}\mathbb{I}_{(0,1)}(\theta).$$

3.2. Ejemplo: Bernoulli-beta

Figura: Distribución $\Pi(\theta)$ actualizada con base en los **datos**



3.3. Predicción

Reflexión

Predicción es el objetivo último cuando se define un modelo estocástico.

Se **anticipan** *posibles resultados de eventos observables con base en* **información pasada.***.*

Distribución predictiva

La **previsión** de posibles esenarios no observados aun se produce con

$$\mathbb{P}(x^f|x_1,\ldots,x_n) \quad = \quad \int_\Theta \underbrace{F(x^f|\theta)}_{\text{Intrínseca}} \underbrace{\Pi(d\theta|x_1,\ldots,x_n)}_{\text{Epistémica}}.$$

Cuando la constante de normalización de $\Pi(\theta|x_1,\ldots,x_n)$ no se puede calcular de manera analítica cerrada, la distribución predictiva puede aproximarse mediante el método de Monte Carlos, donde

$$\widehat{\mathbb{P}}(x^f|x_1,\ldots,x_n) \quad = \quad \frac{1}{M}\sum_{m=1}^{M} F(x^f|\theta^{(m)}).$$

3.3. Predicción

Reflexión

Predicción es el objetivo último cuando se define un modelo estocástico.

Se **anticipan** *posibles resultados de eventos observables con base en* **información pasada.**.

Distribución predictiva

La **previsión** de posibles esenarios no observados aun se produce con

$$\mathbb{P}(x^f|x_1,\ldots,x_n) \quad = \quad \int_\Theta \underbrace{F(x^f|\theta)}_{\text{Intrínseca}} \underbrace{\Pi(d\theta|x_1,\ldots,x_n)}_{\text{Epistémica}}.$$

Cuando la constante de normalización de $\Pi(\theta|x_1,\ldots,x_n)$ no se puede calcular de manera analítica cerrada, la distribución predictiva puede aproximarse mediante el método de Monte Carlos, donde

$$\widehat{\mathbb{P}}(x^f|x_1,\ldots,x_n) \quad = \quad \frac{1}{M}\sum_{m=1}^{M} F(x^f|\theta^{(m)}).$$

3.4. Resultados importantes

Aspectos importantes

Conforme el conjunto de **datos** *es más grande, la incertidumbre epistémica se reduce.*

En el caso límite, la **incertidumbre epistémica** *se desvance, i.e.*

$$\lim_{n\to\infty} \pi(\theta|x_1,\ldots,x_n) = \delta_{\{\theta^*\}}(\theta), \tag{1}$$

donde

θ^* *es el "verdadero" valor de* θ.

Conciliación de opiniones

Se ha demostrado que aun en el caso donde dos individuos asignan distribuciones iniciales distintas, digamos Π_1 *y* Π_2*, el proceso de aprendizaje de ambos converge con el tamaño de los* **datos***, i.e.*

$$\lim_{n\to\infty} \pi_1(\theta|x_1,\ldots,x_n) = \lim_{n\to\infty} \pi_2(\theta|x_1,\ldots,x_n) \tag{2}$$

y ambos convergen al "verdadero" valor de θ^*.

3.4. Resultados importantes

Aspectos importantes

Conforme el conjunto de **datos** *es más grande, la incertidumbre epistémica se reduce.*

En el caso límite, la **incertidumbre epistémica** *se desvance, i.e.*

$$\lim_{n\to\infty} \pi(\theta|x_1,\ldots,x_n) = \delta_{\{\theta^*\}}(\theta), \tag{1}$$

donde

θ^* *es el "verdadero" valor de* θ.

Conciliación de opiniones

Se ha demostrado que aun en el caso donde dos individuos asignan distribuciones iniciales distintas, digamos Π_1 *y* Π_2, *el proceso de aprendizaje de ambos converge con el tamaño de los* **datos**, *i.e.*

$$\lim_{n\to\infty} \pi_1(\theta|x_1,\ldots,x_n) = \lim_{n\to\infty} \pi_2(\theta|x_1,\ldots,x_n) \tag{2}$$

y ambos convergen al "verdadero" valor de θ^*.

3.5. Inferencia

Inferencia bayesiana

Siendo que toda la información contenida en los datos se resume en

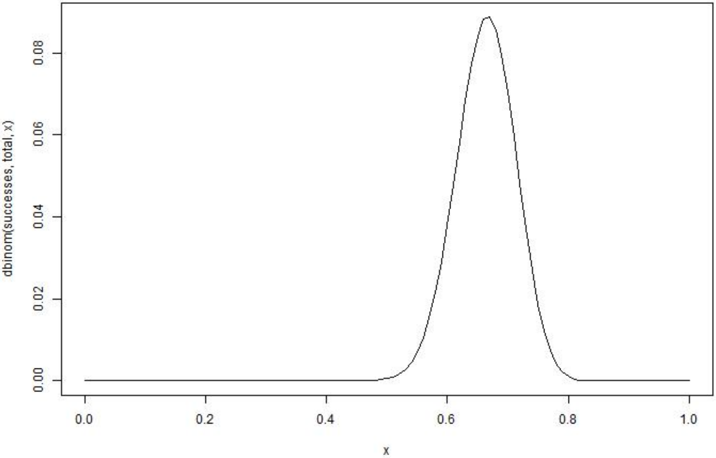
$$\Pi(\theta|x_1, \ldots, x_n),$$

cualquier problema inferencial puede derivarse de esta distribución, e.g.

- **Estimación puntual.-** Media, moda o mediana de la distribución.
- **Estimación por intervalos.-** Regiones inter-cuantílicas de la distribución.
- **Pruebas de hipótesis.-** Con probabilidades cuantificadas respecto a la distribución.

3.5. Inferencia

Figura: Distribución final de θ en el modelo Bernoulli-beta



4. Herramientas computacionales

4.1. Variables latentes

Definición

Para un modelo de probabilididad donde

$$x_i \sim F(x_i|\theta),$$

tal que la distribución puede expresarse como

$$F(x_i|\theta) = \int F(x_i|\gamma)G(d\gamma|\theta),$$

de define γ como una variable latente para el dato x_i.

Interpretación

- Las **variables latentes** con variables no observables que descomponen el modelo en partes analíticamente manejables.
- En particular, se emplean para simplificar la distribución conjunta (verosimilitud) de un conjunto de datos.
- *Se emplean tanto en el paradigma bayesiano como frecuentista de aprendizaje.*

4.1. Variables latentes

Definición

Para un modelo de probabilididad donde

$$x_i \sim F(x_i|\theta),$$

tal que la distribución puede expresarse como

$$F(x_i|\theta) = \int F(x_i|\gamma)G(d\gamma|\theta),$$

de define γ como una variable latente para el dato x_i.

Interpretación

- Las **variables latentes** con variables no observables que descomponen el modelo en partes analíticamente manejables.
- En particular, se emplean para simplificar la distribución conjunta (verosimilitud) de un conjunto de datos.
- *Se emplean tanto en el paradigma bayesiano como frecuentista de aprendizaje.*

4.1. Variables latentes

Ejemplo: Modelo tipo mezcla

Para un modelo de probabilidad donde

$$x_i \sim F(x_i|\theta) = \sum_{j=1}^{J} w_j N(x_i|\mu_k, 1),$$

donde $w_j \geq 1$ y $\sum_{j=1}^{J} w_j = 1$.

Verosimilitud

El aprendizaje con base en un conjunto de datos, $x_1, \ldots, x_n$, descansa en la funcion de verosimilitud,

$$lik\left(\{w_j, \mu_j\}_{j=1}^{J} | \{x_i\}_{i=1}^{n}\right) = \prod_{i=1}^{n} \sum_{j=1}^{J} w_j N(x_i|\mu_j, 1),$$

la cual resulta intratable.

4.1. Variables latentes

Ejemplo: Modelo tipo mezcla

Para un modelo de probabilidad donde

$$x_i \sim F(x_i|\theta) = \sum_{j=1}^{J} w_j N(x_i|\mu_k, 1),$$

donde $w_j \geq 1$ y $\sum_{j=1}^{J} w_j = 1$.

Verosimilitud

El aprendizaje con base en un conjunto de datos, $x_1, \ldots, x_n$, descansa en la funcion de verosimilitud,

$$lik\left(\{w_j, \mu_j\}_{j=1}^{J} | \{x_i\}_{i=1}^{n}\right) = \prod_{i=1}^{n} \sum_{j=1}^{J} w_j N(x_i|\mu_j, 1),$$

la cual resulta intratable.

4.1. Variables latentes

Ejemplo: Modelo tipo mezcla

Introduciendo una variable latente

$$z_i = j \quad \text{con} \quad \mathbb{P}(z_i = j) = w_j,$$

tenemos que

$$x_i | z_i \sim N(x_i | \mu_j, 1).$$

integrando z_i recuperamos el modelo original de mezclas.

Verosimilitud extendida

Introduciendo las variables latentes en el proceso de aprendizaje, definimos la siguiente extensión de verosimilitud,

$$lik\left(\{w_j, \mu_j\}_{j=1}^{J}, \{z_i\}_{i=1}^{n} | \{x_i\}_{i=1}^{n}\right) = \prod_{i=1}^{n} \mathbb{I}(z_i = j) w_j N(x_i | \mu_j, 1),$$

y ahora **aprendemos** de los **parámetros** y **variables latentes** simultáneamente.

4.1. Variables latentes

Ejemplo: Modelo tipo mezcla

Introduciendo una variable latente

$$z_i = j \quad \text{con} \quad \mathbb{P}(z_i = j) = w_j,$$

tenemos que

$$x_i | z_i \sim N(x_i | \mu_j, 1).$$

integrando z_i recuperamos el modelo original de mezclas.

Verosimilitud extendida

Introduciendo las variables latentes en el proceso de aprendizaje, definimos la siguiente extensión de verosimilitud,

$$lik\left(\{w_j, \mu_j\}_{j=1}^{J}, \{z_i\}_{i=1}^{n} | \{x_i\}_{i=1}^{n}\right) = \prod_{i=1}^{n} \mathbb{I}(z_i = j) w_j N(x_i | \mu_j, 1),$$

y ahora **aprendemos** de los **parámetros** y **variables latentes** simultáneamente.

4.2. Aproximaciones numéricas

Racionalidad

Constantes de normalización de la distribución final de **parámetros** *y* **variables latentes** *usualmete no se pueden calcular analíticamente.*

Idea

Usualmente, se generan **muestras aleatorias** *de la distribución final, y características de interés se aproximan empleando el método de Monte Carlo.*

Ejemplo: Mezcla de Modelos

Se generan

$$\left\{ \left(w_j^{(m)}, \mu_j^{(m)} \right)_{j=1}^{J}, \left(z_i^{(m)} \right)_{i=1}^{n} \right\}_{m=1}^{M}$$

para un M relativamente "grande".

Estas muestras serán presumiblemente obtenidas de la distribución final de interés.

4.2. Aproximaciones numéricas

Racionalidad

Constantes de normalización de la distribución final de **parámetros** *y* **variables latentes** *usualmete no se pueden calcular analíticamente.*

Idea

Usualmente, se generan **muestras aleatorias** *de la distribución final, y características de interés se aproximan empleando el método de Monte Carlo.*

Ejemplo: Mezcla de Modelos

Se generan

$$\left\{ \left(w_j^{(m)}, \mu_j^{(m)} \right)_{j=1}^{J}, \left(z_i^{(m)} \right)_{i=1}^{n} \right\}_{m=1}^{M}$$

para un M relativamente "grande".

Estas muestras serán presumiblemente obtenidas de la distribución final de interés.

4.2. Aproximaciones numéricas

Racionalidad

Constantes de normalización de la distribución final de **parámetros** *y* **variables latentes** *usualmete no se pueden calcular analíticamente.*

Idea

Usualmente, se generan **muestras aleatorias** *de la distribución final, y características de interés se aproximan empleando el método de Monte Carlo.*

Ejemplo: Mezcla de Modelos

Se generan

$$\left\{ \left(w_j^{(m)}, \mu_j^{(m)} \right)_{j=1}^{J}, \left(z_i^{(m)} \right)_{i=1}^{n} \right\}_{m=1}^{M}$$

para un M relativamente "grande".

Estas muestras serán presumiblemente obtenidas de la distribución final de interés.

4.2. Aproximaciones numéricas

Algorimos

- Métodos de aceptació y rechazo (muestreo por importancia)
- Métodos de Monte Carlo vía Cadenas de Markov
  - Metropolis-Hastings
  - Gibbs sampler
- Métodos de Monte Carlo secuenciales
- Métodos variacionales
- Slice-sampler
- Métodos bayesianos computacionales aproximados (ABC)

Ejercicios

Veamos algunos ejemplos en R...

4.2. Aproximaciones numéricas

Algorimos

- Métodos de aceptació y rechazo (muestreo por importancia)
- Métodos de Monte Carlo vía Cadenas de Markov
  - Metropolis-Hastings
  - Gibbs sampler
- Métodos de Monte Carlo secuenciales
- Métodos variacionales
- Slice-sampler
- Métodos bayesianos computacionales aproximados (ABC)

Ejercicios

Veamos algunos ejemplos en `R`...