Estimación y Teorema del limite Central

Cristian Guarnizo-Lemus
cristianguarnizo@itm.edu.co

Maestria en Automatización y Control Industrial

Estimación o Inferencia

Proceso por el cual se estima o calcula el valor de variables o parámetros desconocidos. Dentro de los modelos probabilísticos existen

- Estimación puntual: calculo puntual de los valores de los parámetros. Por ejemplo los valores de los parámetros que maximicen la verosimilitud ($p(\mathcal{D}|\theta)$), o el máximo a posteriori ($p(\mathcal{D}|\theta)p(\theta)$).
- Inferencia Bayesiana: calculo del posterior de los parámetros dados las observaciones, $p(\theta|\mathcal{D})$. Si la verosimilitud y el prior son conjugados, entonces el posterior esta distribuido por el mismo tipo de distribución del prior. Si el posterior no es tratable, se recurre a inferencia aproximada usando técnicas de muestreo o variacionales.

Contenido

Estimación Puntual

Consiste en estimar los valores de los parámetros que maximicen una función de probabilidad o un criterio de error.

1. Función de costo: Error medio cuadrático, Verosimilitud, Posterior.
2. Técnicas de optimización: Derivada (pendientes iguales a cero), Gradiente, búsquedas aleatorias.

Contenido

Inferencia Bayesiana

Consiste en determinar el posterior de los parámetros, la cual es la distribucion de los parametros dados los datos

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

1. Prior Conjugado: para ciertas distribuciones de verosimilitud y prior, se puede encontrar que el posterior tiene las misma forma que el prior.
2. Posterior no tratable: Se puede aproximar por medio de técnicas de muestreo: MCMC, MH, HMC, Gibbs sampling. O por métodos variacionales. Para modelos de variable latente se puede emplear el algoritmo EM.

Inferencia Bayesiana - Software

Actualmente existen paquetes de software orientados a estimar distribuciones de variables empleando muestreo, por ejemplo:

- Edward - TensorFlow.
- Pyro - PyTorch.
- MXFusion - MXNet.
- Stan - C++ (R, Python).

Contenido

Propiedades de MLE

- Si la fdp de $p(\mathcal{D}|\theta)$ del conjunto de datos $\mathcal{D}$ satisface condiciones de regularidad, entonces el MLE del parámetro desconocido θ esta distribuido asintóticamente (para N grande) como

$$\widehat{\theta} \overset{a}{\sim} \mathcal{N}\left(\theta, I^{-1}(\theta)\right)$$

  donde $I(\theta)$ es el coeficiente de información de Fisher.
- **Condiciones de regularidad**: las derivadas de primer y segundo orden de la verosimilitud logarítmica existen, y el coeficiente de información de Fisher es diferente de cero.
- De la distribución asintótica se observa que el MLE es asintóticamente insesgado.

Propiedades de MLE

- Una característica importante del MLE es que siempre es posible encontrarlo si no analíticamente, sí numéricamente.
- Es posible emplear métodos de optimización no lineal como Newton-Raphson.
- El punto que encuentre el algoritmo de optimización puede no ser el mínimo global, si no un mínimo local.

Máxima verosimilitud

Calcular el estimador de máxima verosimilitud para un conjunto de datos $\mathcal{D} = \{x_1, \ldots, x_N\}$, donde los datos iid y están modelados de la siguiente manera

$$x_n \sim \mathcal{N}(\mu, \sigma^2).$$

La verosimilitud se define como

$$p(\mathcal{D}|\mu, \sigma) = \prod_{n=1}^{N} \mathcal{N}(\mu, \sigma^2) = \prod_{n=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$

Contenido

Notación

Permite generar muestras de una distribución en particular, por ejemplo si se quiere decir que X es una variable aleatoria definida por una distribución normal:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

donde el símbolo $\sim$ significa “esta distribuida” y la letra $\mathcal{N}$ corresponde a “normal”.

Transformación lineal

Una transformación lineal de X es $X' = aX + b$, donde a y b son números reales. Una familia de distribuciones es cerrada bajo la transformación lineal si X' esta en la misma familia de X. La distribución normal tiene esta propiedad; si $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$X' \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

Transformación lineal

Las distribuciones Normales también son cerrada bajo la adición. Si $Z = X + Y$ y $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y $Y \sim \mu_{\dagger}, \sigma_{\mathcal{Y}}^{\in}$ entonces

$$Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

En el caso particular de $Z = X + X$, se tiene

$$Z \sim \mathcal{N}(2\mu_x, 2\sigma_X^2).$$

Y en general si extraemos n valores de X y los sumamos, se tiene

$$Z \sim \mathcal{N}(n\mu_x, n\sigma_X^2).$$

Contenido

Distribuciones muestrales

Podemos expresar la distribución muestral de $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, donde cada x_i es una muestra de $X \sim \mathcal{N}(\mu, \sigma^2)$.
Asumiendo que la suma de las X's es Y, entonces

$$Y \sim \mathcal{N}(n\mu, n\sigma^2)$$

Y usando el resultado anterior

$$Z \sim \mathcal{N}(\mu, \sigma^2/n)$$

La distribución de Z es la distribución muestral de $\bar{x}$.
La media de Z es μ lo que indica que $\bar{x}$ es un estimador insesgado (unbiased) de μ. La varianza de la distribución es σ^2/n.

Contenido

Teorema del limite central

El teorema nos indica que la suma de muestras de cualquier distribución converge a una distribución normal. Mas específicamente, si la distribución tiene media μ y desviación σ, la distribución de la suma es aproximadamente $\mathcal{N}(n\mu, n\sigma^2)$.

Referencias

- S. Shamugan. “ Random Signals: Detection, Estimation and Data Analysis”, 1988.