Màxima versemblança

Aleix Torres i Camps

1 Pròleg

En aquest document s'explica mitjançant un entrada escalonada què és i per a què serveix la funció de màxima versemblança (*maximum likelihood*) i la seva acompanyant més freqüent, la funció de màxima logversemblança (*maximum loglikelihood*). L'objectiu del mateix és descriure el context en que es situa a cada moment, per a que així fer més intuitius els passos que segueixen a la deducció de la funció de màxima versemblança. També s'inclouen, en tot moment, **exemples** amb la finalitat de fer el document més pràctic i allunyar-nos puntualment de l'abstractis-me i la generalització.

2 Què és la funció de màxima versemblança?

La funció de densitat (i de probabilitat):

Sigui Y una variable aleatòria que pot dependre del paràmetre θ, aquest poguent ser un vector $\theta = (\theta_1, \cdots, \theta_k)$, direm que la seva funció de densitat (o de probabilitat si és discreta) és:

$$\boxed{f_Y(y;\theta) \qquad \text{on } y \text{ és la variable de la funció}}$$

En el cas d'una mostra aleatòria de grandària N, suposant rèpliques independents $\{Y_1, Y_2, \cdots, Y_N\}$ la funció de densitat de la mostra és

$$\boxed{f_{\{Y_1,Y_2,\cdots,Y_N\}}(y_1, y_2, \cdots, y_N;\theta) = \prod_{i=1}^{N} f_{Y_i}(y_i;\theta)}$$

és a dir, el producte de les funcions de densitat de les Y_i.

Recordatori

La funció de densitat dona la probabilitat que la variable aleatória caigui en un conjunt de valors. Però també es pot servir per comparar probabilitats de dos punts concrets, ja que un és més probable que l'altre si el valor de la funció de densitat en aquell punt és més gran que el valor per l'altre punt.

Per les variables aleatories discretes, les funcions de probabilitat ja donen la probabilitat de valors concrets i, per tant, també es pot fer servir per comparar les probabilitats de dos punts concrets.

Exemple de funció de densitat (*distribució de Poisson*)

Suposant un succés aleatori que cada variable de la mostra és independent identicament distribuida (i.i.d.) i segueixen un *distribució de Poisson*, és a dir, $Y_i \sim Poisson(\lambda)$, cada una té per funció de probabilitat:

$$\boxed{f_{Y_i}(y_i;\lambda) = e^{-\lambda}\frac{\lambda^{y_i}}{y_i!}}$$

I, per tant, la funció de probabilitat d'una mostra de grandària N és

$$\boxed{f_Y(y;\lambda) = \prod_{i=1}^{N} e^{-\lambda}\frac{\lambda^{y_i}}{y_i} = e^{-N\lambda}\frac{\lambda^{\sum_{i=1}^{N} y_i}}{\prod_{i=1}^{N} y_i!}}$$

La funció de densitat (i de probabilitat) d'un model:

El següent pas és considerar què passaria si les variables $Y = \{Y_1, Y_2, \cdots, Y_N\}$ depenguessin dels valors d'unes variables explicatives X, quan $\{X_1 = x_1, X_2 = x_2, \cdots, X_N = x_N\}$ (on cada X_i pot pendre un vector de valors). Aleshores la funció de densitat de la variable $Y_i|X_i$ és

$$\boxed{f_{Y_i|X_i}(y_i; x_i, \theta)}$$

I la del model és:

$$\boxed{f_{Y|X}(y; x, \theta) = \prod_{i=1}^{N} f_{Y_i|X_i}(y_i; x_i, \theta)}$$

que depèn de la variable $y = (y_1, y_2, \cdots, y_N)$.

Exemple de funció de densitat d'un model (*recta de regressió*)

Siguin $Y_i|X_i$ unes variables aleatòries que segueixen una $N(\beta_0+\beta_1 x_i, \sigma^2)$. Donats els valors $\{X_1 = x_1, \cdots, X_N = x_N\}$, la funció de densitat de cada Y_i és

$$\boxed{f_{Y_i|X_i}(y_i; x_i, \beta, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i-(\beta_0+\beta_1 x_i)}{\sigma}\right)^2}}$$

per tant, la funció de densitat del model és

$$\boxed{f_{Y|X}(y; x, \beta, \sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i-(\beta_0+\beta_1 x_i)}{\sigma}\right)^2}}$$

Continuació de l'exemple, ara amb valors numérics (*recta de regressió*)

Suposem que, per estudiants de sexe masculí i amb edat entre els 18 i els 26 anys, la funció de densitat del seu pes (en Kg) coneixent la seva alçada (en cm) segueix una $N(-33 + 0.62 \cdot \mathit{Alçada}, 8^2)$. Per un estudiant que medeixi 170 cm tindrem:

$$\boxed{f_{y_1}(y_1) = \frac{1}{\sqrt{2\pi 8^2}} e^{-\frac{1}{2}\left(\frac{y_1-(-33+0.62\cdot 170)}{8}\right)^2} = 0.0499 e^{-\frac{1}{128}(y_1-72.4)^2}}$$

I, per un conjunt d'estudiants amb alçades $X = \{170, 188, 178, 192, 184\}$ tindrem:

$$\boxed{f_Y(y) = (0.0499)^5 e^{-\frac{1}{128}\sum_{i=1}^{5}(y_i-(-33+0.62x_i))^2}}$$

Canvi de paper entre les Y_i i dels paràmetres θ.

Fins ara, suposavem que de les variables aleatòries Y_i en coneixíem els paràmetres i podiem utilitzar-la per calcular probabilitats. Ara, podem pensar-ho al revés, seguim suposant que les variables Y_i segueix una distribució que depen d'uns paràmetres, però aquests paràmetres els desconeixem. I el que tenim és una sèrie de valors experimentals, és a dir, per cada Y_i coneixem y_i. Aleshores, per una mostra tindrem:

$$\boxed{\mathcal{L}(\theta; y) = \prod_{i=1}^{N} f_{Y_i}(y_i; \theta) \qquad \text{On } \theta \text{ és la variable de la funció.}}$$

A la funció $\mathcal{L}(\theta; y)$ l'anomenarem. I, al seu logaritme l'anomenarem funció de logversemblança:

$$\boxed{\mathscr{L}(\theta; y) = \log \mathcal{L}(\theta; y) = \sum_{i=1}^{N} \log f_{Y_i}(y_i; \theta)}$$

funció de versemblança

Canvi de paper entre les Y_i i dels paràmetres θ en un model.

Ara fem el mateix canvi, però per les variables aleatòries $Y_i|X_i$. A més, ara coneixem el valor experimental de y_i i x_i, però també desconeixem els parametres. Tambe anomenarem funció de versemblança a:

$$\boxed{\mathcal{L}(\theta; y, x) = \prod_{i=1}^{N} f_{Y_i|X_i}(y_i; x_i, \theta) \qquad \text{On } \theta \text{ és la variable de la funció.}}$$

I funció de logversemblança a:

$$\boxed{\mathscr{L}(\theta; y, x) = \log \mathcal{L}(\theta; y, x) = \sum_{i=1}^{N} \log f_{Y_i|X_i}(y_i; x_i, \theta)}$$

Exemple de funció de funció de versemblança I (*distribució de Poisson*)

Suposant un succés aleatori que cada variable de la mostra és independent identicament distribuida (i.i.d.) i segueixen un *distribució de Poisson*, és a dir, $Y_i \sim Poisson(\lambda)$, cada una té per funció de probabilitat:

$$\boxed{f_{Y_i}(y_i; \lambda) = e^{-\lambda} \frac{\lambda^{y_i}}{y_i!}}$$

I, per tant, la funció de probabilitat d'una mostra de grandària N és

$$\boxed{f_Y(y; \lambda) = \prod_{i=1}^{N} e^{-\lambda} \frac{\lambda^{y_i}}{y_i} = e^{-N\lambda} \frac{\lambda^{\sum_{i=1}^{N} y_i}}{\prod_{i=1}^{N} y_i!}}$$

---

3 Exemples detallats