Atšķirības Beijesa un klasiskās statistikas metodēs

Aleksis Jurševskis

LU

08.10.2012

Prezentācijas struktūra

1. ieskats Beijesa metodēs;
2. diskusija starp Dž. Robinsu, L. Vasermanu un K. Simsu

Klasisko (frekventistu) statistisko metožu pamatprincipi

1. Varbūtība ir frekvence ar kādu notiek notikums. Varbūtība ir notikuma objektīva īpašība.
2. Parametri ir nezināmas konstantes.
3. Statistikas metodes jākonstruē tā, lai tām piemistu labas, iepriekš definētas īpašības.
4. Hipotēzes var tikai noraidīt vai pieņemt ar noteiktu ticamību. Hipotēzes nevar apstiprināt.

Beijesa statistisko metožu pamatprincipi

1. Varbūtība ir personisks uzskats (degree of belief) par notikuma iespējamību.
2. Lai arī parametri ir konstantes, to novērtēšanai tiek pieņemts, ka tie ir gadījuma lielumi.
3. Informāciju par parametru varam iegūt, izpētot tā sadalījumu.
4. Hipotēzes varam tikai salīdzināt savā starpā. Hipotēzes nevar noraidīt, tās var tikai aizstāt ar labākām.

Piemērs

$$Y = \alpha + \beta X$$

Beijesietis α un β uztver kā gadījuma lielumus, piemēram,
$\alpha \sim N(0, 1)$;
$\beta \sim U(0, \inf)$,
tādejādi iegūstot šaurākus ticamības intervālus α un β pēc ievades modelī.

Beijesa metode

1. Izvēlas aprioro (prior) blīvuma funkciju $f(\theta)$ katram no nezināmajiem parametriem.
2. Izvēlas modeli $f(x|\theta)$.
3. Novēro datus $X_1, ..., X_n$. Izmantojot Beijesa teorēmu, ar simulācijām (Markova ķēdes Montekarlo algoritmi) vai skaitlisko integrēšanu iegūst aposterioro (posterior) sadalījumu $f(\theta|X_1, ..., X_n)$.

Viens g.l. X, diskrētas blīvuma funkcijas

$$\mathbb{P}(\Theta = \theta | X = x) = \frac{\mathbb{P}(X = x, \Theta = \theta)}{\mathbb{P}(X = x)} =$$

$$= \frac{\mathbb{P}(X = x | \Theta = \theta)\mathbb{P}(\Theta = \theta)}{\sum_\theta \mathbb{P}(X = x | \Theta = \theta)\mathbb{P}(\Theta = \theta)}$$

Viens g.l. X, nepārtrauktas blīvuma funkcijas

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

n iid g.l. $X_1, ..., X_n$

$$f(x_1, ..., x_n|\theta) = \prod_{i=1}^{n} f(x_i, \theta) = \mathcal{L}_n(\theta)$$

Apzīmējam $(X_1, ..., X_n)$ ar X^n un $(x_1, ..., x_n)$ ar x^n.

Posteriorais sadalījums

$$f(\theta|x^n) = \frac{f(x^n|\theta)f(\theta)}{\int f(x^n|\theta)f(\theta)d\theta} = \frac{\mathcal{L}_n(\theta)f(\theta)}{c_n} \propto \mathcal{L}_n(\theta)f(\theta)$$

Parametru novērtējumi

Modelis ar vienu parametru

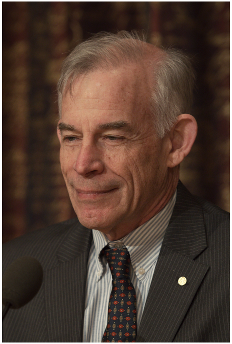
$$\hat{\theta} = \int \theta \mathrm{f}(\theta|\mathrm{x_n}) \mathrm{d}\theta$$

Modelis ar p parametriem $\theta = (\theta_1, ..., \theta_\mathrm{p})$

$$\mathrm{f}(\theta_1|\mathrm{x^n}) = \int ... \int \mathrm{f}(\theta_1, ..., \theta_\mathrm{p}|\mathrm{x^n}) \mathrm{d}\theta_2 ... \mathrm{d}\theta_\mathrm{p}$$

$$\hat{\theta}_1 = \int \theta_1 \mathrm{f}(\theta_1|\mathrm{x^n}) \mathrm{d}\theta_1$$

James Robins & Larry Wasserman 

vs Christopher Sims 

Veselības ministrija (VM), lai plānotu slimnīcu budžetus, vēlas noskaidrot, cik lielai Latvija iedzīvotāju daļai nākošgad būs insults. Tā pirms gada izvēlējās 5000 iedzīvotājus $(X_1, ..., X_{5000})$ un katram no tiem noskaidroja 300 faktorus: augumu, vecumu, iepriekšējas saslimšanas, utt. $(X_i \in [0,1]^{300})$.

Daļa no 5000 tika atlasīti novērošanai ar Bernulli gadījuma lielumu R_i ($R_i = 1$: novērots; $R_i = 0$: nenovērots), turklāt varbūtība tikt novērota nav visiem vienāda, to nosaka zināma funkcija $\pi(X) \equiv \mathbb{E}[R|X]$. No šiem novērotajiem daļai bija insults $(Y = 1)$, pārējiem nē $(Y = 0)$.

Definējam arī $\theta(X) \equiv \mathbb{E}[Y|X]$, kas nosaka insulta varbūtību atkarībā no veselības stāvokļa. Šī funkcija nav zināma. Turklāt $\pi(X)$ izvēlēta tā, ka $\mathrm{Cov}\{\theta(X), \pi(X)\} \neq 0$.

Modeļa pieņēmumi

n (5000) iid novērojumi

$$(X_1, R_1, RY_1), ..., (X_n, R_n, RY_n).$$

$$X_i \in [0,1]^d, \qquad Y_i \in \{0,1\}, \qquad R_i \in \{0,1\} \qquad d = 300$$

$$\theta(X) = \mathbb{E}[Y|X] \qquad \pi(X) = \mathbb{E}[R|X] \qquad \mathrm{Cov}\{\theta(X), \pi(X)\} \neq 0$$

Mērķis novērtēt

$$\psi \equiv \mathbb{E}[Y] = \mathbb{E}\{\mathbb{E}[Y|X]\} = \mathbb{E}\{\mathbb{E}[Y|X, R=1]\} = \int_{[0,1]^d} \theta(x)dx$$

$$\theta(x) = \mathbb{E}[Y|X = x] =$$
$$= 1 \cdot \mathbb{P}(Y = 1|X = x) + 0 \cdot \mathbb{P}(Y = 0|X = x) = \mathbb{P}(Y = 1|X = x)$$

$$\mathbb{P}(Y = 0|X = x) = 1 - \theta(x)$$

Līdzīgi

$$\pi(x) = \mathbb{P}(R = 1|X = x) \qquad 1 - \pi(x) = \mathbb{P}(R = 0|X = x)$$

Frekventistu pieeja

Horvica-Tompsona novērtējums

$$\hat{\psi} = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i Y_i}{\pi(X_i)}$$

$\hat{\psi}$ ir nenovirzīts un būtisks (consistent), bet neefektīvs (inefficient).
Taču eksistē uzlabotas H-T versijas.

Beijesiešu pieeja

Izvēlas aprioro sadalījumu $W(\theta)$.

Ticamības funkcija vienam novērojumam (x, r, ry)

$$f_X(x) f_{R|X}(r|x) f_{Y|X}(y|x)^r$$

n novērojumu $(X_i, R_i, R_i Y_i)$ ticamības funkcija

$$\prod_{i=1}^{n} f(X_i) f(R_i|X_i) f(Y_i|X_i)^{R_i} =$$

$$= \prod_i \pi(X_i)^{R_i} (1 - \pi(X_i))^{1-R_i} \theta(X_i)^{Y_i R_i} (1 - \theta(X_i))^{(1-Y_i)R_i}$$

Ticamības funkcija

$$\mathcal{L}(\theta) \propto \prod_{i=1}^{n} \theta(X_i)^{Y_i R_i}(1-\theta(X_i))^{(1-Y_i)R_i}$$

Kombinējot $W(\theta)$ un $\mathcal{L}(\theta)$ iegūst posterioro sadalījumu W_n. Tā kā ψ ir funkcija no θ $\left(\psi = \int_{[0,1]^d} \theta(x)dx\right)$, θ posteriorais sadalījums noteiks ψ vērtību.

Teorēma, Robins un Ritovs 1997

Jebkurš novērtējums, kas nav funkcija no $\pi(\cdot)$, nevar būt vienmērīgi būtisks (uniformly consistent).

Tas nozīmē, ka jebkurš novērtējums $\hat{\psi}$, kas nav funkcija no π, nebūs ψ apkārtnē neatkarīgi no novērojumu skaita.

Viena no beijesiešu piedāvātajām atbildēm ir aprioro sadalījumu $\mathrm{W}(\theta)$ padarīt atkarīgu no π.

VM piemērs. Kardiologam ir izveidojies priekšstats par θ, un viņš vēlas to izmantot, analizējot VM datus ar Beijesa metodi. VM ierēdnis paskaidro, kādi apsvērumi par θ tika ņemti vērā izvēloties π, taču kardiologs ierēdņa informāciju neņem vērā, jo par insultiem ir daudz labāk informēts. Šajā gadījumā θ apriorais sadalījums būs neatkarīgs no π.

Turklāt $\mathrm{W}(\theta)$ atkarība no π ir nepieciešams, bet ne pietiekams nosacījums.

Secinājumi

Beijesa metodes, lai arī ar plašu pielietojumu, noteiktās situācijās var nestrādāt. Beijesieši ir ticamības funkcijas vergi (slaves to the likelihood).

Beijesa metodes ir ērtas, ja nepieciešams modelī ievietot personiskos vai ekspertu uzskatus par pētāmo objektu.

Paldies par uzmanību! Vai ir kādi jautājumi?

Avoti

1. Vasermaņa blogs normaldeviate.wordpress.com
2. L. Wasserman "All of Statistics", 2004