

Inversio-ongelmien laskennallinen peruskurssi

Luento 7

Kevät 2012

1 Tilastolliset inversio-ongelmat

Tilastollinen inversio perustuu seuraaviin periaatteisiin:

1. Kaikki mallissa olevat muuttujat mallinnetaan satunnaismuuttujilla.
2. Niiden realisaatioiden satunnaisuus kuvaa tiedon astetta (degree of information).
3. Tämä tiedon määrä on ”koodattu” todennäköisyysjakaumiin.
4. Tilastollisen inversio-ongelman ratkaisu on tuntemattoman muuttujan a posteriori -todennäköisyysjakauma.

Varsinkin viimeinen kohta erottaa tilastollisen lähestymistavan perinteisistä tavoista. Regularisaatiomenetelmät tuottavat yksittäisen estimaatin tuntemattomasta, kun taas tilastolliset menetelmät tuottavat todennäköisyysjakauman, jota voidaan käyttää varsinaisten estimaattien tekoon. Näin ollen tärkein kysymys ei ole, ”mikä on muuttujan arvo?”, vaan ”kuinka paljon ja millaista informaatiota meillä on tuntemattomasta muuttujasta?”.

2 Todennäköisyysavaruuksista ja todennäköisyyksistä

Määritelmä 2.1. Olkoon Ω avaruus ja Σ kokoelma sen osajoukkoja. Σ on σ -algebra, jos

1. $\Omega \in \Sigma$;

2. Jos $A \in \Sigma$, niin $A^c = \Omega \setminus A \in \Sigma$;
3. Jos $A_i \in \Sigma$, $i \in \mathbb{N}$, niin $\cup_{i=1}^{\infty} A_i \in \Sigma$.

Määritelmä 2.2. Olkoon Σ σ -algebra avaruudessa Ω . Kuvaus $\mu : \Sigma \rightarrow \mathbb{R}$ on mitta, jos:

1. $\mu(A) \geq 0$ kaikilla $A \in \Sigma$;
2. $\mu(\emptyset) = 0$;
3. Jos joukot $A_i \in \Sigma$ ovat pistevieraita, niin

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad (\sigma\text{-additiivisuus})$$

Määritelmä 2.3. Mittaa μ kutsutaan äärelliseksi, jos $\mu(\Omega) < \infty$. Mittaa kutsutaan σ -äärelliseksi, jos on olemassa jono osajoukkoja $A_i \in \Sigma$, $i \in \mathbb{N}$, siten, että

$$\Omega = \bigcup_{i=1}^{\infty} A_i, \quad \mu(A_i) < \infty \text{ kaikilla } i.$$

Erityisesti, jos $\mu(\Omega) = 1$, kutsutaan mittaa μ todennäköisyysmitaksi.

Määritelmä 2.4. Olkoon Σ σ -algebra avaruudessa Ω ja olkoon P todennäköisyysmitta. Kolmikkoa (Ω, Σ, P) kutsutaan todennäköisyysavaruudeksi. Avaruutta Ω kutsutaan otosavaruudeksi (sample space), Σ on tapahtumajoukko (set of events) ja $P(A)$ on tapahtuman $A \in \Sigma$ todennäköisyys.

Määritelmä 2.5. Olkoon (Ω, Σ, P) todennäköisyysavaruus. Tapahtumia $A, B \in \Sigma$ kutsutaan riippumattomiksi (independent), jos

$$P(A \cap B) = P(A)P(B).$$

Yleisemmin perhettä $\{A_t | t \in T\} \subset \Sigma$, missä T on mielivaltainen indeksijoukko, kutsutaan riippumattomaksi, jos

$$P\left(\bigcap_{i=1}^m A_{t_i}\right) = \prod_{i=1}^m P(A_{t_i})$$

kaikilla äärellisillä indeksien $t_i \in T$ valinnoilla.

Määritelmä 2.6. Olkoon (Ω, Σ, P) todennäköisyysavaruus. $\mathbb{R}^n$ -arvoinen satunnaismuuttuja X on mitallinen kuvaus

$$X : \Omega \rightarrow \mathbb{R}^n,$$

toisin sanoen $X^{-1}(B) \in \Sigma$ kaikilla avoimilla joukoilla $B \subset \mathbb{R}^n$.

Huomautus 2.7. Jatkossa satunnaismuuttujia merkitään isoilla kirjaimilla ja niiden realisaatioita pienillä kirjaimilla, eli $X(\omega) = x$, kun $\omega \in \Omega$.

Määritelmä 2.8. Satunnaismuuttuja $X \in \mathbb{R}^n$ generoi todennäköisyysmitan μ_X avaruudessa $\mathbb{R}^n$ kaavalla

$$\mu_X(B) = P(X^{-1}(B)).$$

Tätä mittaa kutsutaan X :n todennäköisyysjakaumaksi. Satunnaismuuttujan $X = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ todennäköisyysfunktio $F : \mathbb{R}^n \rightarrow [0, 1]$ on

$$F(X) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \quad X = (x_1, x_2, \dots, x_n).$$

Määritelmä 2.9. Satunnaismuuttujan $X \in \mathbb{R}^n$ odotusarvo on

$$\mathbb{E}\{X\} = \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbb{R}^n} x d\mu_X(x),$$

sikäli, kun integraali suppnee. Samoin muuttujan X korrelaatio- ja kovarianssimatriisit määritellään kaavoilla

$$\begin{aligned} \text{corr}(X) &= \mathbb{E}\{XX^T\} \\ &= \int_{\Omega} X(\omega)X(\omega)^T dP(\omega) \\ &= \int_{\mathbb{R}^n} xx^T d\mu_X(x) \in \mathbb{R}^{n \times n}, \end{aligned}$$

ja

$$\begin{aligned} \text{cov}(X) &= \text{corr}(X - \mathbb{E}\{X\}) \\ &= \mathbb{E}\{(X - \mathbb{E}\{X\})(X - \mathbb{E}\{X\})^T\} \\ &= \int_{\mathbb{R}^n} (x - \mathbb{E}\{X\})(x - \mathbb{E}\{X\})^T d\mu_X(x) \\ &= \text{corr}(X) - \mathbb{E}\{X\}\mathbb{E}\{X\}^T, \end{aligned}$$

jos yo. integraalit ovat olemassa.

Määritelmä 2.10. Satunnaismuuttujaa X kutsutaan absoluuttisesti jatkuvaksi (Lebesguen mitan suhteen), jos

$$m(B) = 0 \Rightarrow \mu_X(B) = 0,$$

missä B on Borelin joukko. Tällöin satunnaismuuttuja X määrittelee todennäköisyystiheyden π_X kaavalla

$$P(X^{-1}(B)) = \mu_X(B) = \int_B \pi_X(x) dx.$$

Määritelmä 2.11. Olkoon $X_1 \in \mathbb{R}^n$ ja $X_2 \in \mathbb{R}^m$ kaksi samassa todennäköisyysavaruudessa Ω määriteltyä satunnaismuuttujaa. Niiden yhteistodennäköisyysjakauma (joint probability distribution) määritellään kaavalla

$$\begin{aligned} \mu_{X_1 X_2} : \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^m) &\rightarrow \mathbb{R}_+, \\ (B_1, B_2) &\mapsto P(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)). \end{aligned}$$

Toisin sanoen yhteistodennäköisyysjakauma on satunnaismuuttujien tulon todennäköisyysjakauma:

$$X_1 \times X_2 : \Omega \rightarrow \mathbb{R}^n \times \mathbb{R}^m, \quad \omega \mapsto (X_1(\omega), X_2(\omega)).$$

Määritelmä 2.12. Satunnaismuuttujia x_1 ja x_2 sanotaan riippumattomiksi, jos

$$\mu_{X_1 X_2}(B_1, B_2) = \mu_{X_1}(B_1) \mu_{X_2}(B_2).$$

Jos X_1 ja X_2 ovat absoluuttisesti jatkuvia, voidaan riippumattomuus ilmaista tiheysfunktioiden avulla:

$$\pi_{X_1 X_2}(x_1, x_2) = \pi_{X_1}(x_1) \pi_{X_2}(x_2).$$

3 Ehdolliset todennäköisyydet

Määritelmä 3.1. Olkoon $A, B \in \Sigma$ ja $P(B) > 0$. Tapahtuman A ehdollinen todennäköisyys ehdolla B määritellään kaavalla

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Nyt pätee

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c), \quad B^c = \Omega \setminus B.$$

Yleisemmin, jos $\Omega = \bigcup_{i \in I} B_i$, missä I on numeroituva indeksijoukko, niin

$$P(A) = \sum_{i \in I} P(A|B_i)P(B_i).$$

Tarkastellaan kahta satunnaismuuttujaa $X_i \in \mathbb{R}^{n_i}$, $i = 1, 2$. Käytetään jatkossa seuraavaa merkintätapaa: jos $B_i \in \mathbb{R}^{n_i}$ ovat kaksi Borelin joukkoa, niin merkitään joukon $B_1 \times B_2$ yhteistodennäköisyyttä

$$\mu_{X_1 X_2}(B_1, B_2) = \mu(B_1, B_2) = P(X_1 \in B_1, X_2 \in B_2).$$

Samoin, jos X_1 ja X_2 ovat absoluuttisesti jatkuvia, niin

$$\mu(B_1, B_2) = \int_{B_1 \times B_2} \pi(x_1, x_2) dx_1 dx_2.$$

Käyttäen näitä merkintöjä, määritellään X_1 :n marginaalitodennäköisyys

$$\mu(B_1) = \mu(B_1, \mathbb{R}^{n_2}) = P(X_1 \in B_1),$$

eli muuttujan $X_1 \in B_1$ todennäköisyys huolimatta muuttujan X_2 arvosta.

Oletetaan nyt, että X_1 ja X_2 ovat absoluuttisesti jatkuvia. Tällöin

$$\mu(B_1) = \int_{B_1 \times \mathbb{R}^{n_2}} \pi(x_1, x_2) dx_1 dx_2 = \int_{B_1} \pi(x_1) dx_1,$$

missä marginaalitodennäköisyystiheys on

$$\pi(x_1) = \int_{\mathbb{R}^{n_2}} \pi(x_1, x_2) dx_2.$$

Määritelmä 3.2. Oletetaan, että $B_i \subset \mathbb{R}^{n_i}$, $i = 1, 2$, ovat Borelin joukkoja, ja että joukolla B_2 on positiivinen marginaalimitta, eli $\mu(B_2) = P(X_2^{-1}(B_2)) > 0$. Määritellään joukon B_1 ehdollinen mitta ehdolla B_2 kaavalla

$$\mu(B_1|B_2) = \frac{\mu(B_1, B_2)}{\mu(B_2)}.$$

Erityisesti, jos satunnaismuuttujat ovat absoluuttisesti jatkuvia, niin

$$\mu(B_1|B_2) = \frac{1}{\mu(B_2)} \int_{B_1 \times B_2} \pi(x_1, x_2) dx_1 dx_2,$$

missä

$$\mu(B_2) = \int_{B_2} \pi(x_2) dx_2 = \int_{\mathbb{R}^{n_1} \times B_2} \pi(x_1, x_2) dx_1 dx_2 > 0.$$

Huomaa, että kun B_2 on kiinnitetty, niin kuvaus $B_1 \mapsto \mu(B_1|B_2)$ on todennäköisyysmitta.

Jos satunnaismuuttujat X_1 ja X_2 ovat riippumattomia, niin

$$\mu(B_1|B_2) = \mu(B_1), \quad \mu(B_2|B_1) = \mu(B_2).$$

Tarkoituksena on määritellä ehdollinen mitta tapauksessa, missä ehtona on kiinnitetty arvo, eli $X_2 = x_2$. Ongelmana on, että yleisessä tapauksessa voi olla, että $P(X_2 = x_2) = 0$, jolloin edellä oleva ehdollisen mitan kaava ei toimi. Kuitenkin voidaan osoittaa

Lemma 3.3. *Oletetaan, että X_1 ja X_2 ovat absoluuttisesti jatkuvia ja että niiden tiheydet ovat jatkuvia. Olkoon $x_2 \in \mathbb{R}^{n_2}$ sellainen piste, että*

$$\pi(x_2) = \int_{\mathbb{R}^{n_1}} \pi(x_1, x_2) dx_1 > 0.$$

Edelleen olkoon $(B_2^{(j)})_{1 \leq j \leq \infty}$ pienenevä jono välejä avaruudessa $\mathbb{R}^k$ siten, että $B_2^{(j)} \searrow \{x_2\}$, toisin sanoen $B_2^{(j+1)} \subset B_2^{(j)}$ ja $\bigcap B_2^{(j)} = \{x_2\}$.

Tällöin raja-arvo

$$\lim_{j \rightarrow \infty} \mu(B_1|B_2^{(j)}) = \mu(B_1|\{x_2\})$$

on olemassa, ja se voidaan laskea integraalina

$$\mu(B_1|\{x_2\}) = \frac{1}{\pi(x_2)} \int_{B_1} \pi(x_1, x_2) dx_1.$$

Nyt ehdollisen mitan määritelmä voidaan laajentaa tapaukseen $B_2 = \{x_2\}$. Jos $\pi(x_2) > 0$, niin ehdollinen todennäköisyys joukolle $B_1 \in \mathbb{R}^{n_1}$ ehdolla $B_2 = \{x_2\}$ on (raja-arvo)

$$\mu(B_1|x_2) = \int_{B_1} \frac{\pi(x_1, x_2)}{\pi(x_2)} dx_1.$$

Yllä olevasta lemmasta huomataan myös, että ehdollinen todennäköisyysmitta $\mu(B_1|x_2)$ on määritelty todennäköisyystiheyden avulla

$$\pi(x_1|x_2) = \frac{\pi(x_1, x_2)}{\pi(x_2)}, \quad \pi(x_2) > 0.$$

Tästä seuraa Bayesin kaava todennäköisyystiheyksille:

$$\pi(x_1, x_2) = \pi(x_1|x_2)\pi(x_2) = \pi(x_2|x_1)\pi(x_1),$$

joka kirjoitetaan usein muotoon

$$\pi(x_1|x_2) = \frac{\pi(x_1)\pi(x_2|x_1)}{\pi(x_2)}.$$