Todennäköisyyslaskentaa

Lassi Miinalainen
lassimii@paju.oulu.fi

14.11.2010

1 Johdanto

Kuten edellisellä kerralla sanoin, todennäköisyyslaskenta on ensisijaisesti *vain* matemaattinen malli, sillä suurin osa todennäköisyyslaskennan avulla mallinnetuista ilmiöistä ovat todellisuudessa deterministisiä (esimerkiksi nopanheitto, tietyssä ihmisjoukossa olevien ihmisten syntymäpäivät jne). Keskeisessä asemassa on kaaoksen käsite ja se, että mallinnettavan ilmiön taustalla olevia muuttujia ei tiedetä riittävän tarkasti, eikä niiden mittaaminen ole ehkä mahdollistakaan.

Tarkoitukseni ei ole pelkästään opettaa kaavoja ja lauseita, vaan myös koettaa näyttää mihin periaatteisiin ne perustuvat. Pelkkä MAOL:in taulukkokirjan mekaaninen soveltaminen ei ole kovinkaan tehokasta: paitsi että unohdat kaavat sillä hetkellä kun suljet kirjan, niin hyvin helposti myös sovellat niitä väärissä tilanteissa. Lisäksi kaikissa kirjoissa voi olla vain äärellinen määrä kaavoja, mutta jos tiedät ne yleiset periaatteet, joista ne on johdettu, pystyt myös itse kehittämään uusia. Tämä ei ole vaikeaa, vaikka voikin kuulostaa siltä. Matematiikka on vain formalisoitua ihmisjärkeä, lauseita todellisuudesta, vain järjestelmällisyyden määrä on eri kuin siinä kuuluisassa maalaisjärjessä. Myös sinä sovellat todennäköisyyslaskentaa tiedostamattasi tehdessäsi päätöksiä epävarmoissa tilanteissa. Todennäköisyyslaskenta matematiikan haarana on vain kussakin tilanteessa olevien mahdollisuuksien ja niistä tehtyjen oletusten kirjoittamista esiin paperille.

2 Symmetrinen todennäköisyys

Kaiken takana on siis jokin satunnainen tapahtuma. Kaikkien mahdollisten lopputulosten joukkoa kutsutaan *alkeistapahtumien joukoksi.* Tässä vaiheessa helpotamme toimintaamme siten, että oletamme erilaisia alkeistapahtumia olevan äärellinen määrä ja että niistä jokainen tapahtuu yhtä suurella todennäköisyydellä $p(a) = 1/n$, jossa n on eri alkeistapahtumien lukumäärä. Tästä voimme johtaa klassisen todennäköisyyden lausekkeen

$$\begin{aligned} p(A) &= p(\{a_1, a_2, a_3 \dots a_k\}) = p(\bigcup_{i=1}^{k} \{a_i\}) \\ &= \sum_{i=1}^{k} p(\{a_i\}) = \underbrace{\frac{1}{n} + \frac{1}{n} + \frac{1}{n} + \dots \frac{1}{n}}_{\text{k kpl}} \\ &= \frac{\text{Joukkoon } A \text{ kuuluvien alkeistapahtumien lukumäärä}}{\text{Kaikkien alkeistapahtumien lukumäärä}} \end{aligned}$$

Pohjimmiltaan lainaamassasi oppikirjassa olleet geometrinen todennäköisyys ja symmetrinen todennäköisyys ovat sama asia: molemmissa halutaan vain tietää ennustettavan tapahtuman suhteellinen osuus kaikista mahdollisista tapahtumista.

Kuten jo viime kerralla totesin on hyvä muistaa seuraavat viisi kaavaa:

1. $$p(A \text{ tai } B) = p(A) + p(B) - p(A \text{ ja } B)$$

2. $$\frac{\text{Joukkoon } A \text{ kuuluvien alkeistapahtumien lukumäärä}}{\text{Kaikkien alkeistapahtumien lukumäärä}}$$

3. $$p(\bar{A}) = 1 - p(A)$$

   Kannattaa muistaa myös, että

   $$p(\text{Jotakin tapahtuu ainakin kerran}) = 1 - p(\text{Jotakin ei tapahdu kertaakaan})$$

4. $$p(A \text{ kun tiedetään että B tapahtuu myös}) := p(A|B) = \frac{p(A \text{ ja } B)}{p(B)}$$

5. $$A \text{ ja } B \text{ keskenään riippumattomia} \iff p(A \text{ ja } B) = p(A) \cdot p(B)$$

Ratkaisemme eräitä tehtäviä oppikirjastasi harjoitukseksi ja laskurutiinin kehittämiseksi.

3 Toistokoe

Edellisen osion menetelmät riittyvät yleensä sellaisiin ilmiöihin, jotka ovat tietyllä tapaa "jakamattomia", esimerkiksi kolikko on joko kokonaan kruuna tai klaava, ihminen on joko värisokea tai ei. Uusia alkeistapauksia voidaan myös muodostaa tällaisten jakamattomien alkeistapausten erilaisista yhdistelmistä. Jos yhden tällaisen hieman monimutkaisemman alkeistapauksen sisäiset oliot ovat toisistaan riippumattomia, voidaan muodostaa toistokoe ja siihen liittyvät kaavat. Sitä ennen hieman kombinatoriikkaa.

3.1 Kertoma

Kysymys: Miten monella eri tavalla n eri alkiota voidaan laittaa järjestykseen?
Vastaus: $n \cdot (n-1) \cdot (n-2) \cdot \ldots 3 \cdot 2 \cdot 1 := n!$
Seuraavaksi tutor perustelee miksi näin on.

3.2 Variaatio

Kysymys: Miten monta erilaista k alkion pituista jonoa voidaan muodostaa n alkion suuruisesta joukosta?
Vastaus: $(n)_k = n!/k! = n \cdot (n-1) \cdot (n-2) \cdot \ldots (n-k+1)$. Tutor perustelee jälleen miksi näin on.

3.3 Kombinaatio

Kysymys: Miten monta erilaista k alkion suuruista aliryhmää on n alkion suuruisella joukolla?
Vastaus: Toteamme että kukin näistä aliryhmistä voidaan laittaa $k!$ eri tavalla järjestykseen, siten oikea tulos on

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}$$

Esimerkki 1 (Sivun 110 tehtävä 265)**.** Yleensä toistokokeisiin liittyvissä tehtävissä pyydetään lasketaan todennäköisyys sille, että jotakin tapahtuu (tai jokin ominaisuus vaikuttaa, esim värisokeus) tarkasteltavassa koepopulaatiossa n kertaa tai vähemmän. Tällöin on kiusaus määritellä alkeistapauksiksi tapahtumat: jotakin ei tapahdu kertaakaan, jotakin tapahtuu kerran, jotakin tapahtuu kaksi kertaa, jotakin tapahtuu kolme kertaa, neljä kertaa, viisi kertaa ja niin edelleen. Oikea toiminta on kuitenkin valita alkeistapahtumiksi eri kombinaatiot, eri ne eri tavat joilla X voi tapahtua toistokokeen aikana kerran, kaksi kertaa jne. Toisin sanoen koepopulaation alkiot nimetään toisistaan eroavasti. Tämän jälkeen etenemme seuraavasti ...

Yleisessä n-kertaisessa toistokokeessa, jossa jokaisella yksittäisellä tilastoyksiköllä on kaksi vaihtoehtoa ja tilastoyksiköt ovat toisistaan riippumattomia, pätee seuraavaa

$$p(\text{toistokokeessa tapahtuu X yhteensä k kertaa}) = \binom{n}{k} p^k (1-p)^{n-k}$$

jossa $p = p(X$ tapahtuu kun tarkastellaan yksittäistä tilastoalkiota) Tutor perustelee miksi näin on. Huomioitavaa on että tätä kaavaa ei voida enää suoraan johtaa symmetrisestä todennäköisyydestä, koska yksittäisten tilastoalkioiden todennäköisyysjakauma ei ole enää symmetrinen, eli niin sanotusti heitämme painotettua noppaa. Tästä pääsemmekin seuraavaan aiheeseen.

4 Todennäköisyysjakauma

Symmetrisessä todennäköisyydessä jokainen satunnaismuuttujan arvo oli yhtä todennäköinen (tasainen jakauma). Toistokokeen yhteydessä esittelimme sellaisen tapauksen, jossa satunnaismuuttujalla oli vain kaksi mahdollista arvoa, joiden todennäköisyydet myös erosivat toisistaan (bernoullin jakauma). Yleinen todennäköisyyden kaava kaikille erilaisille jakaumille on:

$$p(A) = \sum_{a_i \in A} p(a_i)$$

Nyt siis satunnainen tapahtuma voi käyttäytyä jollakin tapaa suunnatusti, eli se ei saa joka kertaa samaa arvoa, mutta kuitenkin tapahtumat saattavat kasautua jonkin/joidenkin arvojen ympärille. Tätä ominaisuutta ilmaistaan ns. tunnuslukujen avulla:

1. moodi
2. keskiarvo
3. varianssi ja keskihajonta

Moodi on kaikkein yleistetyin tunnusluku, se voidaan määritellä mille tahansa todennäköisyysjakaumalle, myös ns. luokkamuuttujille (silmien väri, hiusten väri). Moodi on se satunnaismuuttujan arvo, jonka todennäköisyys on suurin.

Keskiarvon ja -hajonnan määrittelyyn tarvitaan, että mitattavaa muuttujaa voidaan kuvata numeroilla (pituus, ikä, tulot). Näiden kahden laskemiseen tarvittavat kaavat on määritelty kirjassasi erikseen frekvenssitaululle, jatkuvalle jakaumalle sekä diskreetille jakaumalle. Osoitan kuitenkin että nämä kolme ovat pohjimmiltaan aina sama kaava, vain hieman eri muodossa.