3A Keskihajonta ja korrelaatio

Tuntitehtävät

3A1 (Korrelaatio ja riippuvuus) Diskreettien satunnaismuuttujien X ja Y yhteisjakauma on esitetty allaolevana taulukkona:

X \ Y	-1	0	1
-1	0	$\frac{1}{6}$	$\frac{1}{6}$
0	$\frac{1}{3}$	0	0
1	0	$\frac{1}{6}$	$\frac{1}{6}$

(a) Määritä X:n jakauma, odotusarvo ja keskihajonta.

(b) Määritä Y:n jakauma, odotusarvo ja keskihajonta.

(c) Laske X:n ja Y:n korrelaatio.

(d) Selvitä, ovatko X ja Y riippuvat vai riippumattomat.

3A2 (Nopanheittojen keskiarvo) Tavallista noppaa heitetään monta kertaa peräkkäin. Heittojen tuloksia merkitään $X_1, X_2, \ldots$ ja ne ovat keskenään riippumattomat. Ensimmäisten n:n tuloksen keskiarvoa merkitään $A_n = \frac{1}{n}(X_1 + \ldots + X_n)$.

(a) Laske satunnaismuuttujan X_1 odotusarvo ja keskihajonta.

(b) Määritä satunnaismuuttujan $A_2 = \frac{1}{2}(X_1 + X_2)$ jakauma.
Vihje: Selvitä ensin A_2:n arvojoukko. Tutki sitten, aluksi pienillä arvoilla a, milloin eli millä parin (X_1, X_2) arvoilla toteutuu $A_2 = a$. Koeta yleistää päättelysi.

(c) Laske satunnaismuuttujan A_2 odotusarvo ja keskihajonta. Vertaa muuttujien A_2 ja X_1 keskihajontoja laskemalla niiden suhde.

(d) Laske odotusarvo ja keskihajonta satunnaismuuttujalle

$$A_{100} = \frac{1}{100}(X_1 + X_2 + \cdots + X_{100}).$$

Vertaa muuttujien A_{100} ja X_1 keskihajontoja laskemalla niiden suhde.

Opastus. C-kohdan voit laskea joko b-kohdassa lasketusta jakaumasta, tai voit hyödyntää odotusarvon ja varianssin yhteenlaskuominaisuuksia. D-kohdassa tarkan jakauman selvittäminen olisi hyvin työlästä, joten on parempi käyttää yhteenlaskuominaisuuksia.

Lauseesta 4.9 seuraa, että $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2 \cdot \mathrm{Cov}(X,Y)$. **Riippumattomilla** satunnaismuuttujilla em. kaavasta jää kovarianssitermi nollaksi, ja usealle **riippumattomalle** muuttujalle kaava yleistyy muotoon $\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$. Summien variansseihin tutustutaan tarkemmin luentomonisteen luvussa 5.

Huomaa, että yhteenlaskukaava pätee nimenomaan variansseille, ei keskihajonnoille.

Kotitehtävät

3A3 (Lämpötilamalli) Meteorologi mallintaa tämän ja huomisen päivän lämpötilojen T_0 ja T_1 välistä yhteyttä kaavalla

$$T_1 \;=\; T_0 + \Delta T$$

jossa ΔT kuvaa lämpötilojen muutosta. Satunnaismuuttujat T_0 ja ΔT oletetaan toisistaan riippumattomiksi. Lisäksi tiedetään, että $\mathrm{E}(T_0) = \mu$ ja $\mathrm{Var}(T_0) = \sigma^2$ sekä $\mathrm{E}(\Delta T) = 0$ ja $\mathrm{Var}(\Delta T) = \theta^2$. Mallin parametrit μ, σ ja θ oletetaan ennalta tunnetuiksi, ja lisäksi $\sigma > 0$ ja $\theta \geq 0$.

(a) Määritä $\mathrm{E}(T_1)$.

(b) Määritä $\mathrm{SD}(T_1)$. Vrt. tehtävään 3A2. Miten lasketaan summan varianssi?

(c) Määritä $\mathrm{Cov}(T_1, T_0)$. Käytä kovarianssin bilineaarisuutta.

(d) Määritä $\mathrm{Cor}(T_1, T_0)$. Ennen kuin lasket korrelaation tarkan arvon, yritä intuitiivisesti päätellä korrelaation tulkinnan kautta miten se käyttäytyy tapauksissa $\theta = 0$ ja $\theta \gg \sigma$ (eli θ paljon suurempi kuin σ).

Tulokset tulee ilmoittaa mahdollisimman yksinkertaisina lausekkeina mallin parametreista μ, σ ja θ.

3A4 (Kustannusfunktion minimointi) Eräs pieni kaupunki sijoittuu itä-länsi-suuntaisen tien varteen niin, että mallinnamme kaupungin 2 kilometrin pituisena reaalilukuvälinä $[0, 2]$. Kaupungin itäosassa asukastiheys on suurempi; jos valitsemme satunnaisen asukkaan, niin hänen asuinpaikkansa X on jatkuva satunnaismuuttuja, jonka tiheysfunktio on $f(x) = x/2$, kun $0 \leq x \leq 2$. (Todellisuudessa populaatio olisi äärellinen ja asuinpaikka diskreetti, mutta oletamme asukkaita olevan niin paljon, että voimme käsitellä heitä jatkuvana massana.)

(a) Laske asuinpaikan odotusarvo $\mu = \mathrm{E}(X)$. Laske asuinpaikan mediaani, eli sellainen piste m että $\mathrm{P}(X \leq m) = \frac{1}{2}$, ts. puolet asukkaista asuu pisteen länsipuolella.

(b) Laske $\mathrm{E}(X^2)$ ja $\mathrm{SD}(X)$.

(c) Abel on kaupunkisuunnittelija, jonka tehtävänä on valita kaupungintalolle paikka a johonkin kohtaa kaupunkia ($0 \leq a \leq 2$). Hän pyrkii minimoimaan *neliöllisen kustannusfunktion* $q(a) = \mathrm{E}\left((X - a)^2\right)$. Toisin sanoen, hän haluaa minimoida kaupungin asukkaiden *keskimääräisen neliöidyn etäisyyden* asuinpaikastaan kaupungintalolle.

Määritä q yksinkertaisena a:stä riippuvana (polynomi)lausekkeena, jossa ei ole satunnaismuuttujia eikä integraaleja. Mikä on tämän funktion $q(a)$ muoto? Muokkaa annettua $q(a)$:n lauseketta sellaiseen muotoon, että pääset käyttämään odotusarvon lineaarisuutta. (Vaihtoehtoisesti voit käyttää odotusarvon muunnoskaavaa ja laskea näin saamasi integraalin.) Funktion muotoa voi tarkastella lausekkeesta tai piirtämällä kuvan.

Etsi se arvo a, jolla $q(a)$ on mahdollisimman pieni. (Minimikohdan pitäisi löytyä helposti derivaatan avulla.) Onko se jompikumpi arvoista μ ja m?

(d) Bertta on toinen suunnittelija. Hänen tehtävänään on valita kirjastolle paikka b johonkin kohtaan kaupunkia ($0 \leq b \leq 2$). Hän pyrkii minimoimaan *lineaarisen kustannusfunktion* $\ell(b) = \mathrm{E}(\,|X - b|\,)$. Toisin sanoen, hän haluaa minimoida kaupungin asukkaiden *keskimääräisen etäisyyden* asuinpaikastaan kirjastoon.

Määritä $\ell(b)$ yksinkertaisena b:stä riippuvana (polynomi)lausekkeena. Aloita odotusarvon muunnoskaavasta. Integraali kannattaa sitten jakaa kahteen osaan, eli väleihin $[0, b]$ ja $(b, 2]$, jotta pääset eroon itseisarvomerkeistä.

Etsi se arvo b, jolla $\ell(b)$ on mahdollisimman pieni. Onko se jompikumpi arvoista μ ja m?

(e) Vapaaehtoinen ylimääräinen lisätehtävä (ei pisteitä): Ovatko Abelin ja Bertan valitsemat paikat samat, vai onko jompikumpi idempänä? Pohdi miksi. Voiko ilmiön ymmärtää intuitiivisesti tarkastelemalla heidän kustannusfunktioitaan? Tapahtuisiko sama ilmiö, vaikka asukkaiden jakauma olisi jokin muu kuin tässä tehtävässä annettu? Entä jos kaupunki ei olekaan jana, vaan alue tasossa?