

Bayes Teoria, tehtävät-2, (28.01.2011)

1. Geenettinen kytkentä ja haplotyyppaus

Oletetaan että diploidi yksilö on heterotsygottinen lokuksissa $l = 1, 2, 3$, eli niissä geenikohdissa hänen perimä koostu pareista

$$\begin{aligned}l = 1 & \quad G_1 = \{A, a\} \\l = 2 & \quad G_2 = \{B, b\} \\l = 3 & \quad G_3 = \{C, c\}\end{aligned}$$

jossa alleli-tyypit $A \neq a$, $B \neq b$ ja $C \neq c$. Oletetaan että lokuksen järjestyks on $1 \leftrightarrow 2 \leftrightarrow 3$.

Lokusten väliset rekombinatio todennäköisyydet ovat $\rho(1, 2) = 0.1$ ja $\rho(2, 3) = 0.2$.

Kysymys 1

Olettamalla rekombinaatioiden riippumattomuus eri geenien välissä, laske todennäköisyys että yhdessä gameetissa ei tapahdu rekombinaatiota lokusten 1,2,3: välissä

Kysymys 2

Laske todennäköisyys että yhdessä gameetissa tapahtuu kahta rekombinaatiota (rekombinaatio välissä $[1, 2]$ ja rekombinaatio välissä $[2, 3]$).

Näistä genotyypeistä voi muodostaa $2^3 = 8$ haplotyyppiä. Kyseisestä yksilöltä on kerätty $N = 100$ gametteja, eli siemeniä jotka mitattiin laboratoriossa. Siemen sisältää yhden haplotyyppi joka on koottu yksilön haplotyyppi-parista rekombinaatio-todennäköisyyksien mukaisesti. Eri rekombinantteja esintyi siemenien datassa kappaleita

$$\#(a, b, c) = 7$$

$$\#(a, b, C) = 3$$

$$\#(a, B, c) = 8$$

$$\#(a, B, C) = 37$$

$$\#(A, b, c) = 34$$

$$\#(A, b, C) = 7$$

$$\#(A, B, c) = 0$$

$$\#(A, B, C) = 4$$

Kysymys 4 Datan perusteella mikä on yksilön todennäköisimmin haplotyyppi?

Kysymys 5 Laske posteriori todennäköisyys datan ehdolla että yksilön haplotyyppit ovat (a, B, C) ja (A, b, c) .

Kysymys 6 Oletetaan nyt että rekombinaatio todennäköisyydet $\rho(1, 2), \rho(2, 3)$ ovat tuntemattomia. Silloin datan todennäköisyys on $\rho(1, 2), \rho(2, 3)$ funktio.

Miten lähtisit laskemaan parametrin $(\rho(1, 2), \rho(2, 3))$ suurimman uskottavuuden estimaattori?

Vihjeet Yksilön haplotyyppit ovat komplementaariset, siksi on 4 eri vaihtoehtoa, jotka ovat a priori yhtä todennäköisiä.

I: (a, B, C) ja (A, b, c) , vai

II: (a, b, c) ja (A, B, C) , vai

III: (a, b, C) ja (A, B, c) , vai

IV: (a, B, c) ja (A, b, C) .

Jokaiselle vaihtoehdolle laske ensin datan ehdollinen todennäköisyys rekombinaatio todennäköisyyden perusteella, ja sitten laske posterioria Bayesin kaavalla.

Taksulaskin tai tietokone tulee tarpeen.