

Uppgifter till 'Matematiska och statistiska metoder för genletning'

Ola Hössjer

March 20, 2006

1 Sannolikhet för nedärvning från samma farförälder

Uppgiften är att härleda formeln

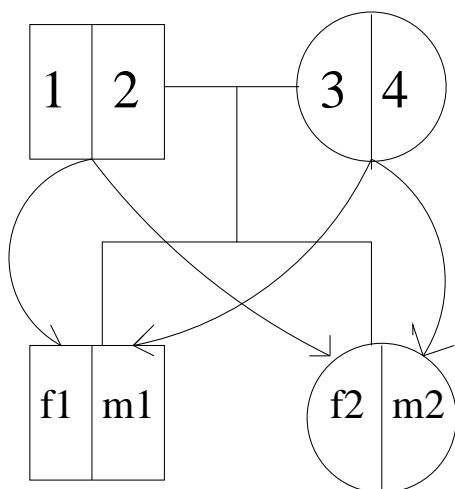
$$p = \frac{1 + q(\psi^2 - 1)}{1 + q(\psi^2 - 1) + (1 + q(\psi - 1))^2}, \quad (1)$$

för sannolikheten att ett affekterat syskonpar har ärvt ner DNA från samma farförälder kring sjukdomsgenen. (Sannolikheten är givetvis densamma att de ärvt ner DNA från samma morförälder.) De ingående parametrarna är $q = P(A)$, sannolikheten för den form A av genen som ökar risken för sjukdomen, och ψ , som anger hur mycket risken att bli sjuk ökar för varje kopia av A . Dvs

$$\begin{aligned} P(\text{sjuk}|aa) &= f, \\ P(\text{sjuk}|Aa) &= f\psi, \\ P(\text{sjuk}|AA) &= f\psi^2. \end{aligned}$$

där f är sjukdomsriskerna för en normalperson (med genuppsättning två normalformer a av genen).

Anta att syskon 1 har genparet f_1m_1 , där f_1 ärvt ner från pappa och m_1 från mamma. Alltså är f_1 a eller A beroende av vad syskon 1 ärvt ner från pappan och m_1 är a eller A beroende av vad syskon 1 ärvt ner från mamman. På samma sätt antar vi att syskon 2 har genparet f_2m_2 , där f_2 ärvt ner från pappa och m_2 från mamma.



1=farfars gen = A eller a
 2=farmore gen = A eller a
 3=morfars gen = A eller a
 4=mormors gen = A eller a

1,2,3 och 4 är A eller a oberoend
 av varandra

För uppgiften ska vi (utan bevis) anta att p fås som sannolikheten att en artificiell affekterad person med genuppsättning $f1f2$ har ärvt ner DNA från samma farförälder. Personen (kallad 3) är artificiell, eftersom han fått båda sina gener från pappan. Vi inför händelserna

- D = '3 är sjuk'
- B = ' $f1$ och $f2$ kommer från samma farförälder'
- C_i = ' $f1f2$ består av i stycken A ', $i = 0, 1, 2$.

Uppgiften är då att bestämma den betingade sannolikheten

$$p = P(B|D) = \frac{P(D \cap B)}{P(D)},$$

vilken är sannolikheten för B givet att D har inträffat. När du läser uppgiften har du nytta av följande identiteter:

$$\begin{aligned} P(D \cap B) &= P(D|B)P(B) \\ P(D) &= P(D|B)P(B) + P(D|B^c)P(B^c) \\ P(D|B) &= P(D|C_0)P(C_0|B) + P(D|C_2)P(C_2|B) \\ P(D|B^c) &= P(D|C_0)P(C_0|B^c) + P(D|C_1)P(C_1|B^c) + P(D|C_2)P(C_2|B^c) \\ P(B^c) &= 1 - P(B), \end{aligned}$$

där B^c är komplementhändelsen till B , dvs 'icke B '.

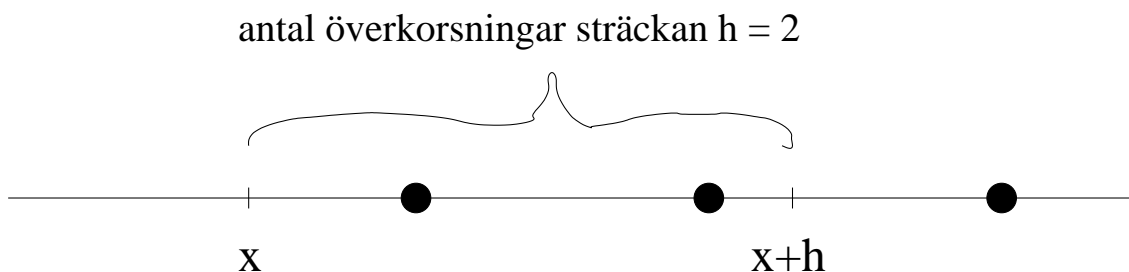
Ledning: Som hjälp på traven, notera att t ex

$$\begin{aligned} P(D|C_1) &= f\psi, \\ P(C_0|B) &= 1 - q, \\ P(C_1|B^c) &= 2q(1 - q). \end{aligned}$$

Ty, givet B , så vet vi att $f_1 = f_2$ och alltså måste C_0 eller C_2 inträffa, med sannolikheter $1 - q$ och q . Given B^c , så väljs f_1 och p_2 oberoende av varandra.

2 Poissonmodell för överkorsningar

Anta att vi har normerat längdskalan för en kromosom så att överkorsningar uppkommer med intensitet 1 (dvs ungefär 10^8 baspar per längdenhet, även om detta varierar längs kromosomerna). I figuren nedan är överkorsningarna markerade som ifyllda cirklar.



Enligt Poissonmodellen har N_h , antalet överkorsningar längs sträckan h , sannolikhetsfördelningen

$$P(N_h = k) = e^{-h} \frac{h^k}{k!}, \quad k = 0, 1, 2, \dots \quad (2)$$

Man kan visa denna formel, utgående från egenskaperna 1-3 hos Poissonmodellen för överkorsningar (se föreläsningen) genom att lösa en serie differentialekvationer.

Det förväntade eller genomsnittliga antalet överkorsningar längs sträckan h skriver vi som $E(N_h)$ (engelskans 'Expectation'). Vi har att

$$E(N_h) = \sum_{k=0}^{\infty} k \cdot P(N_h = k).$$

1. Visa att (2) är en sannolikhetsfördelning, dvs att $\sum_{k=0}^{\infty} P(N_h = k) = 1$.
2. Visa att $E(N_h) = h$.

Ledning: Du har nytta av Taylorutvecklingen

$$e^y = \sum_{k=0}^{\infty} \frac{y^k}{k!}$$

av exponentialfunktionen kring 0.

Nästa uppgift är att använda Poissonmodellen för att härleda att $p(x)$, sannolikheten att ett affekterat syskonpar ärver ner DNA från samma farförälder vi position x , ges av

$$p(x) = 0.5 + (p - 0.5)e^{-2|x-\tau|}, \quad (3)$$

där τ är positionen för sjukdomsgenen och $p = p(\tau)$ sannolikheten (1) att ärva ner DNA från samma farförälder vi sjukdomsgenen.

För att visa (3) så betingar vi på huruvida syskonen har ärvt ner samma DNA vid τ . Låt B och D beteckna samma händelser som i avsnitt 1, inför

$F =$ '3 ärver ner DNA från samma farförälder i position x ',

och låt $N_{|x-\tau|}$ beteckna antalet överkorsningar mellan x och τ . Vi har då

$$\begin{aligned} p(x) &= P(F|D) \\ &= P(F|B)P(B|D) + P(F|B^c)P(B^c|D) \\ &= P(F|B)p + P(F|B^c)(1-p) \\ &= P(N_{|x-\tau|} \text{ jämn})p + P(N_{|x-\tau|} \text{ udda})(1-p). \end{aligned}$$

I andra likheten utnyttjade vi att för att F ska inträffa givet D , så finns två möjligheter: Antingen inträffar B eller B^c (givet D), dvs antingen ärver 3 ner DNA från samma eller olika farföräldrar i sjukdomsgenen τ . $P(F|B)$ är sannolikheten att 3 ärver ner DNA från samma farförälder i position x givet att han även gör det i position τ . Men det är sannolikheten för att ett jämnt antal överkorsningar ägt rum mellan τ och x , eftersom varje överkorsning växlar 'nedärvare' mellan farfar och farmor.

3. Visa att $P(N_h \text{ jämn}) = 0.5(1 + e^{-2h})$ och $P(N_h \text{ udda}) = 0.5(1 - e^{-2h})$.

4. Härled (3).

Ledning: I uppgift 3 har du nytta av att Taylorutveckla e^h och e^{-h} .