

Oppgave, EM-algoritmen

Anta at vi trekker et utvalg på n individ fra en populasjon i Hardy-Weinberg likevekt og at homozygoter AA ikke kan skilles fra heterozygoter av type Aa . Hvilken fordeling har det observerte antallet homozygoter av type aa , n_{aa} ? Hva blir sannsynlighetsmaksimeringsestimaten av P_{aa} ? Og av allelfrekvensen p_a ?

Anta at vi i stedet ønsker å estimere p_a ved hjelp av EM-algoritmen. Hva blir forventningsverdien til de manglende dataene n_{AA}^* og n_{Aa}^* betinget på observerte data og gitt forrige estimat av p_a (E-steget i algoritmen)? Hva blir estimaten av p_a basert på de "fullstendige" dataene $(n_{AA}^*, n_{Aa}^*, n_{aa})$ (M-steget)?

Vis at estimaten av p_a konvergerer mot sannsynlighetsmaksimeringsestimaten av p_a , altså at algoritmen fungerer i denne situasjonen.