

Tentamenskrivning: TMS145 - Grundkurs i matematisk statistik och bioinformatik, 7,5 hp.

Tid: Tisdag den 16 december, 2008 kl 8.30 - 12.30 Eklandagatan 86.

Examinator: Olle Nerman, tel 7723565.

Jour: Frank Eriksson, tel 073-7263051.

Hjälpmedel: Egen handskrivnen formelsamling (fyra A4-sidor) samt med skrivningen utdelade tabellsidor.

Maxpoäng: 32. För godkänt krävs minst 15 poäng totalt och minst 4 poäng på sannolikhetssteori- och statistik-delen vardera samt minst 3 poäng på bioinformatikdelen. För betyget 4 krävs 20 poäng och för betyget 5 25 poäng.

Sannolikhetssteori

1. I en skolklass om 30 elever är 18 flickor och 12 pojkar. En grupp av tre elever väljs ut för att representera klassen i skolans elevråd, så att varje grupp om tre elever har samma sannolikhet att bli vald. Hur stor är sannolikheten att minst en flicka kommer att finnas bland representanterna? (3p)
2. Tjockleken av en på måfå vald pocketbok i ordinarie (enorma) sortimentet i en stor bokhandel har väntevärdet $\mu = 1,5$ cm och standardavvikelsen $\sigma = 0,5$ cm. Om du köper 50 slumpmässigt valda pocketböcker från bokhandeln vad är då den ungefärliga sannolikheten att de får plats (i en rad) i din bokhylla med "innerbredd" 80 cm? (3 p)
3. Kasta två tärningar.
 - a Är händelserna $A = \{\text{summan av tärningarna} = 12\}$ och $B = \{\text{skillnaden på tärningspoängen} = 0\}$ oberoende? (2p)
 - b Är händelserna $C = \{\text{summan av tärningarna} = 7\}$ och $D = \{\text{första tärningen visar en trea}\}$ oberoende? (1p)
 - c Beräkna de betingade sannolikheterna $P(A|B)$ respektive $P(C|D)$. (1p)
4.
 - a Varför kan man säga att en "säker" händelse A , som har sannolikheten $P(A) = 1$ är oberoende av sig själv? (1p)
 - b Finns det några andra x för vilket $P(A) = x$ medför samma konstanta slutsats? I så fall vilka x ? (1p)

Statistik

4. En paretofördelning har täthetsfunktion

$$f(x) = \frac{\theta}{x^{\theta+1}}, \quad x \geq 1$$

där vi vet att $\theta > 1$. Du har tillgång till

- a Vilket är väntevärdet i en sådan paretofördelning? (1p)

Tänk dig nu ett stickprov från en sådan paretofördelning, x_1, \dots, x_n , med oberoende variabler.

- b Hur ser momentskattaren av θ ut? (2p)

- c Hur ser maximum likelihoodskattaren av θ ut? (2p)

5. Medeltorktiden för en typ av färg är 12 minuter. En ny tillsats testas för att se om den förkortar torktiden. 16 ytor målas och medelvärdet av torktiden observeras till $\bar{x} = 11.1$ och stickprovsvariansen till $s^2 = 0.36$.

- a Bilda ett uppåt begränsat konfidensintervall för den nya medeltorktiden med konfidensgrad 0.95 under antagande om normalfördelning. (2p)

- b Tolka om intervallet till ett test av

$$H_0 : \text{medeltorktiden för den nya färgen} \geq 12$$

Vad blir din slutsats och vilken signifikansnivå har testet? (2p)

6. En gen förekommer som allelerna A och a . För att undersöka om populationen befinner sig i Hardy-Weinbergjämvikt, det vill säga om genotyperna AA , Aa och aa förekommer i proportionerna θ^2 , $2\theta(1-\theta)$ respektive $(1-\theta)^2$, har 80 individer på måfå valts för genotypning med följande resultat.

Typ	AA	Aa	aa
Antal	12	36	32

Det är känt att andelen A -gener i populationen är $1/4$. Är den givna modellen rimlig? (3p)

Bioinformatik

7. Sekvensbioinformatik

- (a) Assuming a match score of 2, a mismatch score of -1 and a gap score of -2, derive the score matrix for a global alignment of "ATTA" and "ATTTTA".

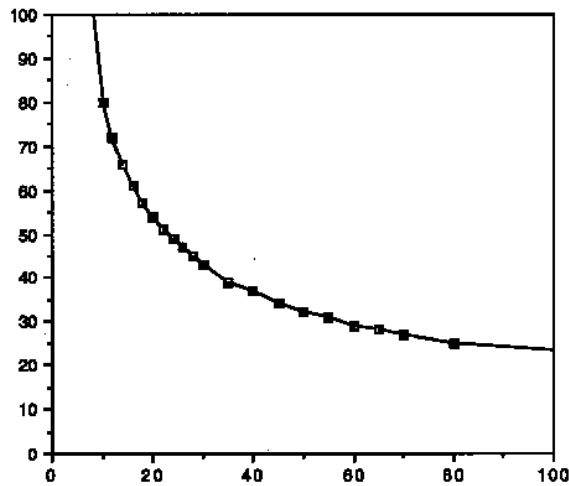
In this case, what is the score of an optimal global alignment? How many alignments have this optimal score (remember: each path represents a different alignment)? What are these alignments? (2p)

- (b) How are z-scores used in assessing the significance of global alignment scores? Explain how a z-score is calculated. (2p)

8. Strukturbioinformatik.

(a) An HSSP-curve is sketched below.

- i Explain what the horizontal and vertical axes represent.
- ii Points are plotted on an HSSP curve using two different symbols (e.g. a cross and a small square). Explain what these points represent.
- iii Explain the significance of the region above the curve. (3p)



(b) Describe a pattern in a multiple sequence alignment that would suggest the presence of an alpha-helix.

Explain why that pattern suggests an alpha-helix.

(2p)