

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

Tid: Lördag den 10 april, 2010 14:00-18:00, Väg och vatten.

Examinator: Olle Nerman, tel 7723565.

Jour: Frank Eriksson, tel 0737263051, Alexandra Jauhiainen tel 073-7168778

Hjälpmedel: valfri miniräknare, egen handskrivna 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 statistikdelen vardera samt minst 3 poäng på bioinformatikdelen.

Sannolikhetssteori

1. Anna och Niklas befinner sig i en fruktträdgård. De har plockat sju frukter av synnerligen aptitligt utseende. De är lyckligt ovetande om att tre av frukterna är giftiga. Om Niklas sätter i sig fyra frukter valda på måfå, och Anna de övriga tre, hur stor är sannolikheten att båda blir förgiftade? (2p)
2. Helena är en ivrig trädgårdsentusiast. Hon har en fröpåse med texten "grobarhet 80%". Vi antar att fröna utvecklas oberoende av varandra. Om Helena sätter n frön, låt den stokastiska variabeln X beteckna antal frön som utvecklas.
 - a) Förklara och motivera vilken fördelning X har och ange dess väntevärde och varians. (2p)
 - b) Beräkna sannolikheten att av 20 sådda frön åtminstone 17 utvecklas. (2p)
 - c) Kommentera oberoendeantagandet. Är det rimligt? (1p)

Vänd!

3. a) Ett bussbolag som har 121 bussar önskar dimensionera en reparationsverkstad. Man önskar få en uppfattning om det totala antalet större reparationer som behöver göras under en månad. Man vet att antalet gånger som en buss behöver genomgå en större reparation under en månad beskrivs av en stokastisk variabel X med sannolikhetsfunktionen

$$P_X(x) = \begin{cases} 0.3 & \text{om } x = 0 \\ 0.5 & \text{om } x = 1 \\ 0.2 & \text{om } x = 2 \\ 0 & \text{annars} \end{cases}$$

Beräkna med lämplig approximation sannolikheten att minst 100 och högst 120 större reparationer totalt behöver göras under en månad. (3p)

- b) Låt den stokastiska variabeln X vara likformigt fördelad på intervallet $(0,1)$. Bestäm täthetsfunktionen $f_Y(y)$ där $Y = (X - 1)^2$. (2p)

Observera att det inte finns något samband mellan a) och b).

Vänd!

Statistik

4. Punktskattning

- a) Definiera följande begrepp (matematiskt): Stickprov (*sample*), skattare (*point estimator*), skattning (*point estimate*) samt väntevärdesriktig skattare (*unbiased estimator*).

2p

Låt X_1, \dots, X_n vara ett stickprov från $Exp(\lambda)$ -fördelningen, dvs $f_{X_i}(x) = \lambda e^{-\lambda x}$, $x \geq 0$, och $E[X_i] = 1/\lambda$.

- b) Härled maximum-likelihood-skattaren för λ .

2p

- c) Antag att vi har gjort de två observationerna $x_1 = 2$ och $x_2 = 6$. Skatta λ med hjälp av skattaren i b).

1p

5. En slumpalsgenerator jag har i min dator påstår sig ge observationer från en likformig fördelning på intervallet $[0, 1]$. Vi tror inte riktigt på detta och bestämmer oss för att undersöka påståendet. Vi genererar därför 200 slumpstal och delar dem i de 10 delintervallen

$[0, 0.1)$, $[0.1, 0.2)$, $[0.2, 0.3)$, $[0.3, 0.4)$, \dots , $[0.8, 0.9)$, $[0.9, 1]$.

Undersök om vi bör skriva och klaga på tillverkaren av slumpalsgeneratorn då vi observerar följande frekvenser i vardera delintervall

19, 29, 20, 21, 19, 18, 18, 16, 21, 19.

3p

Vänd!

6. Frank har krockat med sin Vespa igen och den behöver ny lack. På grund av nya miljökrav får han inte använda originallacken från femtiotalet. Femtioårslacken torkade på i snitt tre minuter. Tillverkaren påstår att torktiden för den nya lacken kortare, vilket vi vill undersöka.

a) Ställ upp den noll- och alternativ(mot)hypotes som krävs för att kunna påvisa att den nya lacken torkar snabbare.

1p

b) Redogör för de praktiska konsekvenserna av ett fel av typ I.

1p

c) Vid sexton försök där små plåtbitar lackades med den nya lacken uppmättes följande torktider

1.4	2.1	2.8	0.9
2.4	1.7	3.7	2.7
2.6	1.9	2.8	2.8
2.2	2.2	3.4	1.9

där $\bar{x} = 2.3438$, $s^2 = 0.5106$. Testa hypotesen från fråga a) på nivån $\alpha = 0.05$. Vilken slutsats drar du?

2p

Vänd!

Bioinformatik

7. Sequence Alignment

- a) (i) What is the Hamming distance between strings *GTATG* and *GTTCG*? Explain your answer.

(1p)

- (ii) What is the Levenshtein distance between strings *GTATG* and *GTTCG*? Explain your answer.

(1p)

- b) 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 *GTATG* with *GTTCG*.

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)

8. Structural Bioinformatics

- a) In describing protein conformation, what are the phi and psi angles? Which atoms define each of these? Draw a sketch to illustrate your answer.

(2p)

- b) What are the main steps in the comparative modelling process?

(2p)