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

Tid: Torsdag den 23 augusti, 2007 kl 14.00 - 18.00 i M-huset.
Examinator: Olle Nerman, tel 7723565.
Jour: Alexandra Jauhiainen, tel 073-7168778, Erik Kristiansson, tel 7725342.
Hjälpmedel: kalkylator, egen handskriven formelsamling (fyra A4 sidor) samt med skrivningen utdelade formel- och tabellsidor.

Maxpoäng: 32. För godkänt krävs minst 15 poäng totalt och minst 4 poäng på sannolikhetsteori- och statistik-delen vardera samt minst 3 poäng på bioinformatikdelen.

Sannolikhetsteori

1 Snickaren Arne bygger ut sin sommarstuga. Spiken som han använder är av dålig kvalité och 20% är defekta.

(a) Om Arne väljer ut 8 spikar slumpmässigt, beräkna sannolikheten att max 2 av dem är defekta. (2p)

(b) Om Arne väljer ut 800 spikar slumpmässigt, beräkna den approximativa sannolikheten att max 200 av dem är defekta. (2p)

2 Resultatet av två mätningar (X och Y) kan beskrivas enligt

$$f_{X,Y}(x,y) = K(x+y) \quad 0 \leq x \leq 1,\ 0 \leq y \leq 1.$$

(a) Bestäm K så att $f_{X,Y}(x,y)$ är en täthetsfunktion. (1p)

(b) Beräkna korrelationen mellan X och Y. Tolka resultatet. Är X och Y oberoende? (3p)

3 En doktorand på matematisk statistik är på väg till jobbet och ska åka spårvagn mellan Redbergsplatsen och Chalmers. På denna sträcka går två spårvagnar, 6:an och 8:an. När doktoranden anländer till hållplatsen, låt X vara tiden tills en spårvagn från linje 6 kommer och Y vara tiden tills en spårvagn från linje 8 kommer. Antag att fördelningen för både X och Y är $\text{Uni}(0,10)$. Antag också att X och Y är oberoende. Låt Z vara tiden tills en spårvagn från någon av linjerna kommer.

(a) Beräkna fördelningen för Z. *Tips:* $Z = \min(X,Y)$. (2p)

(b) Beräkna den förväntade tiden som doktoranden får vänta vid hållplatsen. (1p)

(c) Hur mycket längre blir den förväntade tiden som doktoranden får vänta vid hållplatsen om linje 8 är indragen så att endast spårvagnar från linje 6 går? (1p)

Statistik

4 Miljöteknik-konsulten Henrik arbetar med ett uppdrag på en potentiellt nedsmutsad industritomt. Han har tagit jordprover från två områden, som båda kan vara förorenade. Jordproverna analyseras med avseende på mängden av en viss tungmetall (mg/kg):

```
Område A:
5.53 2.93 2.67 1.69 4.93 6.40 5.66 6.04 4.87
Område B:
12.66 5.95 4.82 6.94 10.14 8.60 9.44 14.75 10.56 11.20 6.04
```

Följande information kan vara användbar: $\bar{x}_A = 4.52$, $\bar{x}_B = 9.19$, $s_A = 1.674$, $s_B = 3.078$

(a) Undersök om de teoretiska varianserna i stickproven kan anses vara lika. Nivå 0.05. (2p)

(b) Konsulten misstänker att område B är mer förorenat än område A. Undersök detta på lämpligt sätt (om möjligt med information från uppgift (a)) på nivå 0.01. (2p)

Lämpligt normalfördelningsantagande får göras.

5 Livslängderna för en viss typ av kondensatorer modelleras med exponentialfördelning med väntevärde $1/\lambda$. Vi har följande observationer från 10 kondensatorer:

```
127.6 340.0 380.5 967.6 22.6 56.0 111.8 118.8 69.3 122.8
```

(a) Härled maximumlikelihood-skattningen för λ och beräkna den för observationerna ovan. (2p)

(b) Ofta undersöker man om en punktskattning har vissa önskvärda egenskaper. Beskriv två sådana egenskaper och varför de är önskvärda. (2p)

6 (a) Till en affär anländer kunder med en intensitet av λ kunder per timme. Antalet kunder, X, som anländer under ett tidsintervall t modelleras som $Poi(\lambda t)$. Under två timmar en vardag anlände 40 kunder.

Gör ett konfidensintervall för λ den aktuella dagen med konfidensgrad approximativt 0.95.

Tips: Här är $X \sim Poi(2\lambda)$. Använd normalapproximation. (2p)

(b) Ulla singlar en slant 10 gånger. Låt X beteckna antalet krona som kommer upp och p beteckna sannolikheten för krona. Ulla använder sig av nollhypotesen $H_0 : p = 0.5$ som förkastas om $x \geq 8$ där x är en observation av X.

Vilken nivå har testet? Vilken mothypotes prövas? (2p)

Bioinformatik

7 Sekvensbioinformatik

Beskriv principiellt hur Needleman-Wunsch algoritm löser parvisa sekvensuppställningsproblem. (4p)

8 Strukturbioinformatik.

(a) Draw a sketch that shows the omega torsion angle in a protein's main chain. Which atoms define this angle? What values are typically observed for this angle? (2p)

(b) Describe the planarity checks performed by PROCHECK. (2p)

Lycka till!