

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

Tid: Lördag den 18 april 2009, kl 14:00-18:00 Väg och vatten

Examinator: Olle Nerman, tel 7723565.

Jour: Frank Eriksson, tel 073-7263051.

Hjälpmedel: Egen handskriven 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. Vad är sannolikheten för två oberoende händelser A och B med sannolikheterna $P(A) = 0,2$ resp $P(B) = 0,9$ att

a de inträffar båda två?

(1p)

b inte någon inträffar?

(1p)

c inte någon av dem inträffar någon enda gång vid 3 oberoende upprepningar av försöket?

(2p)

2. Avrundningsfelen vid en viss typ av vägningar kan antas vara likformigt fördelade i intervallet $-0,005$ och $0,005$. Om verkliga vikterna varierar mycket mellan 100 olika enheter som vägs, så kan de 100 mätfelen dessutom betraktas som oberoende.

a Vilket väntevärde har mätfelet för en enskild enhet?

(1p)

b Vilken varians har mätfelet för en enskild enhet?

(1p)

c Bestäm approximativt sannolikheten att absolutbeloppet av summan av mätfelen för de hundra enheterna är $< 0,02$

(2p)

Vänd!

3. Antag att du har en kontinuerligt fördelad stokastisk variabel X med frekvensfunktionen $f(x) = (2x + 2)C$ på intervallet $[-1, 1]$

a Bestäm konstanten C .

(1p)

b Bestäm frekvensfunktionen för $10X$

(1p)

c Bestäm fördelningsfunktionen för $10X$

(2p)

Statistik

4. Låt x_1, \dots, x_n vara ett stickprov från oberoende $Gamma(k, \theta)$ -fördelade variabler. $X \sim Gamma(k, \theta)$ innebär att X har frekvensfunktion $f_X(s) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}$, $x > 0$, $k > 0$, $\theta > 0$ samt $E[X] = k\theta$.

a Härled maximum likelihoodskattaren för θ .

(2p)

b Är detta en väntevärdesriktig (unbiased) skattare?

(1p)

5. En forskare hävdar att åtminstone 10% av alla hockeyhjälmor har ett visst tillverkningsfel som skulle kunna ge upphov till skador hos bäraren. Då man kontrollerar 400 hjälmor upptäcker man att 32 av dessa har det påtalade felet.

a Ta reda på, med hjälp av ett lämpligt hypotestest på signifikansnivå 0,01, om detta stödjer forskarens påstående.

(2p)

b Finn testets p-värde.

(1p)

c Vilken bredd skulle ett 99%-igt konfidensintervall baserat på informationen ovan ha?

(1p)

Vänd!

6. För att undersöka hur temperaturen förändras ju längre norrut i Sverige som man kommer, så mättes årsmedeltemperaturen år 2004 i 11 svenska orter. De återges i följande tabell tillsammans med orternas latituder.

| Ort | Latitud | Medeltemperatur — |
|-----------|---------|-------------------|
| Jokkmokk | 66,6 | -0,6 |
| Umeå | 63,5 | 4,0 |
| Östersund | 63,1 | 4,2 |
| Gävle | 60,4 | 5,8 |
| Karlstad | 59,2 | 7,0 |
| Stockholm | 59,3 | 7,6 |
| Göteborg | 57,8 | 7,7 |
| Jönköping | 57,4 | 6,0 |
| Visby | 57,6 | 7,6 |
| Kalmar | 56,7 | 7,5 |
| Lund | 55,7 | 8,5 |

Sambandet mellan årsmedeltemperatur och latitud kan sammanfattas med en regressionsmodell

```
lm(formula = y ~ x)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|--------|--------|--------|
| -1.6397 | -0.5078 | 0.3306 | 0.6731 | 1.3348 |

Coefficients:

| | Estimate | Std. Error |
|-------------|----------|------------|
| (Intercept) | 49.16360 | 5.72184 |
| x | -0.72341 | 0.09562 |

Residual standard error: 1.015 on 9 degrees of freedom

Analysis of Variance Table

Response: y

| | Df | Sum Sq | Mean Sq |
|-----------|----|--------|---------|
| x | 1 | 59.025 | 59.025 |
| Residuals | 9 | 9.281 | 1.031 |

Vänd!

a Ange uttrycket (som en ekvation) för den linjära regressionsmodellen i exemplet ovan. Ange skattningarna av samtliga parametrar.

(2p)

b Testa på signifikansnivå $\alpha = 0.01$ hurivida årsmedeltemperaturen 2004 sjönk ju längre norrut man kom.

(1p)

c I en regressionsanalys arbetar man ofta med kvadratsummor. Beskriv de intressanta kvadratsummor som man tittar på och vilken information man kan få från dessa.

(2p)

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 "GTTA" and "GTCCA".

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 The BLOSUM62 is shown below. Comment on the score between W and Y, and the score between L and D.

(1p)

c Calculate the score of the following multiple alignment using the BLOSUM62 matrix and the sum of pairs method:

Sequence 1: GAHV

Sequence 2: GATA

Sequence 3: GSSV

(1p)

Vänd!

BLOSUM62 Matrix:

| | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y | V |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|
| A | 4 | | | | | | | | | | | | | | | | | | | |
| R | -1 | 5 | | | | | | | | | | | | | | | | | | |
| N | -2 | 0 | 6 | | | | | | | | | | | | | | | | | |
| D | -2 | -2 | 1 | 6 | | | | | | | | | | | | | | | | |
| C | 0 | -3 | -3 | -3 | 9 | | | | | | | | | | | | | | | |
| Q | -1 | 1 | 0 | 0 | -3 | 5 | | | | | | | | | | | | | | |
| E | -1 | 0 | 0 | 2 | -4 | 2 | 5 | | | | | | | | | | | | | |
| G | 0 | -2 | 0 | -1 | -3 | -2 | -2 | 6 | | | | | | | | | | | | |
| H | -2 | 0 | 1 | -1 | -3 | 0 | 0 | -2 | 8 | | | | | | | | | | | |
| I | -1 | -3 | -3 | -3 | -1 | -3 | -3 | -4 | -3 | 4 | | | | | | | | | | |
| L | -1 | -2 | -3 | -4 | -1 | -2 | -3 | -4 | -3 | 2 | 4 | | | | | | | | | |
| K | -1 | 2 | 0 | -1 | -3 | 1 | 1 | -2 | -1 | -3 | -2 | 5 | | | | | | | | |
| M | -1 | -1 | -2 | -3 | -1 | 0 | -2 | -3 | -2 | 1 | 2 | -1 | 5 | | | | | | | |
| F | -2 | -3 | -3 | -3 | -2 | -3 | -3 | -3 | -1 | 0 | 0 | -3 | 0 | 6 | | | | | | |
| P | -1 | -2 | -2 | -1 | -3 | -1 | -1 | -2 | -2 | -3 | -3 | -1 | -2 | -4 | 7 | | | | | |
| S | 1 | -1 | 1 | 0 | -1 | 0 | 0 | 0 | -1 | -2 | -2 | 0 | -1 | -2 | -1 | 4 | | | | |
| T | 0 | -1 | 0 | -1 | -1 | -1 | -1 | -2 | -2 | -1 | -1 | -1 | -1 | -2 | -1 | 1 | 5 | | | |
| W | -3 | -3 | -4 | -4 | -2 | -2 | -3 | -2 | -2 | -3 | -2 | -3 | -1 | 1 | -4 | -3 | -2 | 11 | | |
| Y | -2 | -2 | -2 | -3 | -2 | -1 | -2 | -3 | 2 | -1 | -1 | -2 | -1 | 3 | -3 | -2 | -2 | 2 | 7 | |
| V | 0 | -3 | -3 | -3 | -1 | -2 | -2 | -3 | -3 | 3 | 1 | -2 | 1 | -1 | -2 | -2 | 0 | -3 | -1 | 4 |

8. Strukturbioinformatik.

- a
 - i Describe how an anti-parallel beta-sheet can be identified by the DSSP program.
 - ii Explain whether a proline residue can be present in an anti-parallel beta-sheet.

(3p)

- b Describe one of the planarity checks performed by PROCHECK. State which atoms are involved in this check.

(1p)