

Tentamenskrivning: TMS145 - Grundkurs i matematisk statistik och bioinformatik, 7.5p.

Tid: Tisdagen den 18 december, 2012 kl 08.30 - 12.30 i V-huset.

Examinator: Erik Kristiansson

Jour: Erik Kristiansson, tel 070-5259751.

Hjälpmedel: kalkylator, egen handskrivna formelsamling (fyra A4-sidor) samt med skrivningen utdelade tabellsidor.

Max är 32 poäng. För godkänt krävs minst 15 poäng, för betyget 4 krävs 21 poäng och för 5 krävs 26 poäng. Uppgifterna kommer inte i svårighetsordning.

1. På ett kasino kan följande spel spelas. Ett rättvist mynt kastas 4 gånger och antalet klave X räknas. Därefter får spelaren en vinst på Y kronor enligt formeln $Y = X^2$.

- (a) Bestäm sannolikhetsfunktionen p_X för X . Vad har X för fördelning?
- (b) Beräkna den förväntade vinsten, d.v.s. $E[Y]$.
- (c) För kasinobesökare kostar spelet 6 kronor att spela. Antag att 100 besökare spelar spelet under en och samma kväll. Beräkna approximativt sannolikheten att de utbetalade vinsterna överstiger de totala intäkterna, d.v.s. sannolikheten att kasinot går med förlust.

(5 p)

2. Staffan S driver en chark som säljer förkokta julskinkor. Den rekommenderar salthalten i en förkokt julskinka är 0.75 gram per 100 gram julskinka. Staffan S misstänker att ett parti med julskinkor har för hög salthalt. För att undersöka detta tas prover från 20 slumpmässigt utvalda julskinkor och mängden salt (g/100 g julskinka) uppmäts (x_1, \dots, x_{20}) . Stickprovsmedelvärdet beräknas till $\bar{x} = 0.93$ (g/100 g julskinka). Av tidigare erfarenheter vet Staffan S att salthalten hos julskinkor är normalfördelad med känd varians $\sigma^2 = 0.103$.

- (a) Formulera fördelningsantaganden och hypoteser. Använd ett ensidigt hypotestest för att avgöra om salthalten för julskinkorna i partiet är över rekommendationen på 0.75. Signifikansnivån ska vara 0.01.

- (b) Beräkna p-värdet för testet i uppgift (a).

(4 p)

3. Låt X och Y vara två stokastiska variabler.
- (a) Förklara begreppen kovarians och korrelation. Hur hänger de ihop och vad är skillnaden?
 - (b) Antag att väntevärdet för X är μ och att variansen är σ^2 . Låt $Y = aX + b$ där $a > 0$ och b är konstanter. Visa att korrelationen mellan X och Y är 1.

(4 p)

4. Låt X_1, \dots, X_n vara ett stickprov från en Poissonfördelning med sannolikhetsfunktion

$$p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \geq 0.$$

- (a) Använd maximum likelihood-metoden för att härleda en punktskattning $\hat{\lambda}_{ML}$ av λ .
- (b) Visa att λ_{ML}^* är väntevärdesriktig.
- (c) Antag att vi observerar x_1, \dots, x_5 till 8, 8, 6, 4, 12. Beräkna värdet på $\hat{\lambda}_{ML}$ och dess standardfel.

(5 p)

5. Vid tillverkning av två kugghjul kan de båda felen beskrivas som en två-dimensionell stokastiska variabeln (X, Y) med täthetsfunktion

$$f_{X,Y}(x, y) = \frac{4}{5}(xy + x + y), \quad 0 < x < 1, 0 < y < 1.$$

- (a) Beräkna väntevärdet för det totala felet $X + Y$.
- (b) Beräkna sannolikheten att det totala felet $X + Y$ är mindre än 1.

(5 p)

6. I en vallokalsundersökning tillfrågas 500 slumpmässigt utvalda individer om deras partitillhörighet. Av de tillfrågade personerna uppger 263 att dom tänker rösta på Matematikpartiet.

- (a) Beräkna ett approximativt konfidensintervall för proportionen väljare som stödjer Matematikpartiet. Konfidensgraden ska vara 0.99. Kan vi vara säkra på att matematiskpartiet har egen majoritet?
- (b) Hur många personer måste tillfrågas för att längden på konfidensintervallet inte ska överstiga 2 procentenheter, d.v.s. 0.02. Gör beräkningen under antagandet att $\hat{p} = 0.5$.

(5 p)

7. Bioinformatics

- (a) When performing global pairwise sequence alignment with a dynamic programming algorithm (the Needleman-Wunsch algorithm), each path through the matrix corresponds to an alignment of the sequences being compared. For each of the matrices below, give the alignment that corresponds to the shown path.

i)

		E	S	V	A
A					
D					
T					
L					

ii)

		E	S	V	A
A					
D					
T					
L					

- (b) Using a gap score of -2 and match/mismatch scores taken from the PAM250 substitution matrix (given below), derive the score matrix for a global alignment of "ADTL" with "ESVA". In this case, what is the score of an optimal global alignment? Give the alignment(s) with this score.

PAM250 substitution matrix:

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	2																			
R	-2	6																		
N	0	0	2																	
D	0	-1	2	4																
C	-2	-4	-4	-5	4															
Q	0	1	1	2	-5	4														
E	0	-1	1	3	-5	2	4													
G	1	-3	0	1	-3	-1	0	5												
H	-1	2	2	1	-3	3	1	-2	6											
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5										
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6									
K	-1	3	1	0	-5	1	0	-2	0	-2	-3	5								
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6							
F	-4	-4	-4	-6	-4	-5	-5	-5	-2	1	2	-5	0	9						
P	1	0	-1	-1	-3	0	-1	-1	0	-2	-3	-1	-2	-5	6					
S	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	3				
T	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-2	0	1	3			
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17		
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10	
V	0	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4	

(4 p)

GOD JUL!!