

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

**Tid:** Tisdagen den 13 december, 2011 kl 08.30 - 12.30 i V-huset.

**Examinator:** Erik Kristiansson

**Jour:** Erik Kristiansson, tel 070-5259751, Malin Östenson, tel 5316.

**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 en av Göteborgs gymnasieskolor finns det 865 elever varav 50 sympatiserar med Centerpartiet (C). I samband med skolvalet 2010 genomfördes en opinionsundersökning där 80 slumpmässigt valda elever svarade på vilket parti de sympatiserade med. Låt  $X$  vara antalet elever som svarade att de sympatiserade med Centerpartiet.
  - (a) Vad har  $X$  för exakt fördelning? Ange även väntevärdet och variansen för  $X$ .
  - (b) Beräkna den approximativa binomialfördelningen för  $X$ .
  - (c) Beräkna en approximativ sannolikhet för att Centerpartiet får minst 10% av sympatierna.

(5 p)

2. Doktoranderna Fredrik och Anna ska steka köttbullar till jul. Fredrik påstår att köttbullarna som han rullar är mindre än de Anna rullar och för att visa detta genomför han en statistisk undersökning. Fredrik drar på måfå  $m = 10$  köttbullar som han själv har rullat och  $n = 8$  köttbullar som Anna har rullat. Fredrik mäter därefter diametern på köttbullarna  $(x_1, \dots, x_{10}$  samt  $y_1, \dots, y_8)$  och beräknar att  $\bar{x} = 2.22$ ,  $s_X = 0.427$ ,  $\bar{y} = 3.01$ ,  $s_Y = 0.922$  (alla värden i cm). Baserat på erfarenheter från tidigare jular vet Fredrik att diametern på en köttbulle kan antas vara normalfördelad.
  - (a) Formulera lämpliga fördelningsantaganden samt en noll- och alternativhypotes för att testa om Fredriks köttbullar är mindre än Annas.
  - (b) Testa nollhypotesen mot den alternativa hypotesen under antagandet att variansen i de båda stickproven är lika. Använd signifikansnivån  $\alpha = 0.05$ .

- (c) Beräkna p-värdet för för testet ovan. Har Fredrik rätt i sitt påstående.

(4 p)

3. Låt  $(X, Y)$  vara en tvådimensionell stokastisk variabel med täthetsfunktion

$$f_{(X,Y)}(x, y) = cye^{-(xy+2y)}, \quad x, y \geq 0.$$

- (a) Bestäm konstanten  $c$  så att  $f_{X,Y}$  är en giltig täthetsfunktion.  
(b) Bestäm de marginella täthetsfunktionerna för  $X$  och  $Y$ .  
(c) Är  $X$  och  $Y$  oberoende?

(5 p)

4. Låt  $X_1, \dots, X_n$  vara ett oberoende stickprov från en exponentialfördelning med tätheten

$$f(x) = \frac{1}{\mu}e^{-x/\mu}, \quad x \geq 0$$

- (a) Använd maximum likelihood-metoden för att härleda en punktskattning  $\hat{\mu}_{ML}$  för parametern  $\mu$ . Är punktskattaren väntevärdesriktig?  
(b) Visa att den alternativa punktskattaren

$$\mu_1^* = n \times \min(X_1, \dots, X_n)$$

är en väntevärdesriktig punktskattare av  $\mu$ .

- (c) Vilken av punktskattarna  $\mu_{ML}^*$  och  $\mu_1^*$  är effektivast, d.v.s. vilken har lägst varians?

(5 p)

5. Låt  $X_1, \dots, X_n$  vara ett stickprov från en normalfördelning med väntevärde  $\mu$  och varians  $\sigma^2 = 1$ . Nollhypotesen  $H_0 : \mu = \mu_0$  testas mot den alternativa hypotesen  $H_A : \mu > \mu_0$  med hjälp av teststatistikan

$$u = \frac{\bar{x} - \mu_0}{1/\sqrt{n}}.$$

$H_0$  förkastas om  $u > c$ .

- (a) Förklara vad som menas med ett typ-II-fel. När inträffar ett typ-II-fel i testet ovan?

- (b) Härled ett generellt uttryck för styrkefunktionen  $1 - \beta(\mu)$  (d.v.s. för alla  $n$ ,  $\mu$ ,  $\mu_0$ , och  $c$ ).
- (c) Antag att  $\mu_0 = 0$  och det sanna värdet är  $\mu = 1$ . Om testets signifikansnivån är 0.05, hur många observationer behövs för att testets styrka ska bli 0.90? Hur många observationer behövs för att testets styrka ska bli 0.90 om signifikansnivån istället sätts till 0.01? Diskutera resultaten.

(5 p)

6. I en läkemedelsstudie deltar 500 slumpmässiga individer. En individ kan avbryta sitt deltagande i studien och detta kan ske av två orsaker: (i) till följd av en allvarlig biverkning och (ii) till följd av andra orsaker. Antag att 3% av individerna får någon form av biverkning och att 25% av de som får en biverkning avbryter sin medverkan. Av de som inte får några biverkningar kommer 0.26% av deltagarna att avbryta sitt deltagande.

- (a) Beräkna sannolikheten att en individ avbryter studien.
- (b) Beräkna sannolikheten att en person som har avbrutit studien har en biverkning.

(4 p)

7. 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 "DFS" with "EGYT".

In this case, what is the score of an optimal global alignment? Give the alignment(s) with this score.

	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	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4

(4 p)

GOD JUL!!