TENTAMEN I SF1923/SF1924 SANNOLIKHETSTEORI OCH STATISTIK,
TISDAG 31 MAJ 2022 KL 8.00–13.00.

Examinator: Björn-Olof Skytt, 08-790 86 49.

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik (utdelas vid tentamen), miniräknare.

Tentamen består av två delar, benämnda del I och del II. Del I består av uppgifterna 1-12. På denna del skall endast svar anges, antingen i form av ett numeriskt värde med tre värdesiffrors noggrannhet eller i form av val av ett av de möjliga svarsalternativen. Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1-3, utan får tillgodoräkna sig dessa tre uppgifter. Gränsen för godkänt är 9 poäng. Möjlighet att komplettera ges för tentander med 8 poäng.
Del II består av uppgifterna 13-16 och varje korrekt lösning ger 10 poäng. Del II rättas bara för studenter som är godkända på del I och poäng på del II krävs för högre betyg än E. På denna del skall resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar skall förklaras och definieras och numeriska svar skall anges med minst två värdesiffrors noggrannhet. Studenter som är godkända på datorlaborationen får 4 bonuspoäng på del II på ordinarie tentamenstillfället och det första omtentamenstillfället.

Tentamen kommer att vara rättad inom tre arbetsveckor från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Del I

Uppgift 1

Om en person har en viss sjukdom så visar ett test för sjukdomen positivt med sannolikhet 0.98. Om personen inte har sjukdomen visar testet negativt med sannolikhet 0.97. Antag att förekomsten är 5 på 1000 att någon i populationen har sjukdomen. Antag att testet visar positivt. Vad är då sannolikheten att personen inte har sjukdomen?

A: 0.141

B: 0.196

C: 0.804

D: 0.859

Uppgift 2

Den stokastiska variabeln X har täthetsfunktionen

$$f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1\,, \\ 0\,, & \text{annars}\ . \end{cases}$$

En ny stokastisk variabel Y bildas genom $Y = \quad 2X + 1$. Beräkna $P(Y > \frac{1}{2})$.

A: $\frac{1}{16}$

B: $\frac{1}{8}$

C: $\frac{1}{4}$

D: $\frac{1}{2}$

Uppgift 3

Låt X och Y vara två diskreta stokastiska variabler med simultan sannolikhetsfunktion $p_{X,Y}(x, y)$ vars värden ges av tabellen:

$p_{X,Y}(x,y)$	$y = 0$	$y = 1$	$y = 2$
$x = 0$	0.1	0.2	0.1
$x = 1$	0.1	0.2	0
$x = 2$	0.1	0.1	0.1

Beräkna $C(X, Y)$.

A: -0.01

B: 0

C: 0.04

D: 0.09

Uppgift 4

Ett åkeri har 18 lastbilar. Av dessa har 4 stycken körförbud. Vid en poliskontroll tas 4 av åkeriets lastbilar slummpmässigt ut för kontroll. Vad är sannolikheten att exakt 2 av de av polisen kontrollerade lastbilarna har körförbud?

A: 0.121

B: 0.145

C: 0.178

D: 0.212

Uppgift 5

Det är för närvarande väldigt högt tryck på upplysningstjänsten 11313. Antag att man ringer och får beskedet att det är 100 samtal före i kön. Antag vidare att det endast är en person som besvarar samtal och att längden på varje samtal antas vara exponentialfördelad med väntevärde 4 minuter. Längden på ett samtal antas också vara oberoende av längden på andra samtal. Beräkna approximativt sannolikheten att kötiden blir mer än 6 timmar.

A: 0.055

B: 0.159

C: 0.841

D: 0.945

Uppgift 6

Antalet samtal till ett företags callcenter under en tiominutersperiod antas vara Poissonfördelat med väntevärde 1.3. Vad är sannolikheten att det kommer fler än 2 men färre än 5 samtal under perioden 10.30   10.50 ?

A: 0.13

B: 0.36

C: 0.43

D: 0.68

Uppgift 7

Antag att $X \in N(-2.7\theta, 0.1)$ och $Y \in N(7.9\theta, 0.4)$ där X och Y är oberoende stokastiska variabler. Beräkna Minsta-Kvadrat-skattningen av θ givet utfallen x = −3.6 och y = 9.3.

A: -2.84

B: 1.19

C: 1.28

D: 2.85

Uppgift 8

Låt X och Y vara två oberoende s.v. sådana att $X \in U(0, 4\theta)$ och $Y \in U(0, 2\theta)$.

$$\theta^*_{obs} = \frac{x + 2y}{4}$$

och

$$\hat{\theta}_{obs} = \frac{x + y}{3}$$

är två skattningar av θ.

A: Båda skattningarna är lika effektiva.

B: Man kan inte avgöra vilken av skattningarna som är effektivast, eftersom minst en av skattningarna inte är väntevärdesriktig.

C: θ^*_{obs} är effektivast.

D: $\hat{\theta}_{obs}$ är effektivast.

Uppgift 9

Låt $X \in N(\mu_X, \sigma)$ och $Y \in N(\mu_Y, \sigma)$. Vi gör 10 observationer på X och får att $\bar{x} = 19.5$ och $s^2_X = 4.06$. Vi gör även 10 observationer på Y och får att $\bar{y} = 13.9$ och $s^2_Y = 3.48$. Vi antar att alla observationer är utfall av stokastiska variabler som alla är oberoende. Ta fram ett tvåsidigt konfidensintervall för skillnaden $\mu_X \quad \mu_Y$ som har konfidensgrad 95%. Ange övre gränsen för detta.

A: 7.03

B: 7.10

C: 7.30

D: 7.42

Uppgift 10

För att undersöka kvalitén hos de kretsar som ett visst företag levererar så tar man ett stickprov på 800 st kretsar och finner att 742 av dessa fungerar. Bilda ett ensidigt nedåt begränsat konfidensintervall för andelen fungerande kretsar. Använd approximativ konfidensgrad 90% och ange nedre gränsen för detta konfidensintervall.

A: 0.841

B: 0.893

C: 0.912

D: 0.916

Uppgift 11

Anta att $X \in Bin(8, p)$ och låt H_0 vara att $p = 0.1$. Vi vill testa H_0 mot alternativet $H_1 : p = 0.25$ och förkastar H_0 till förmån för mothypotesen H_1 om vi får observerationen $x > 2$. Bestäm testets styrka.

A: 0.04

B: 0.19

C: 0.32

D: 0.63

Uppgift 12

Följande datamaterial beskriver hur försäljningen av en vara beror av hur mycket som spenderats på reklam.

Reklam (kkr)	28	39	45	53	59
Försäljning (kkr)	315	335	340	350	352

Utifrån datamaterialet ovan skattas följande regressionsmodell (dvs med både linjär och kvadratisk term)

$$y_i = \quad + \beta x_i + \gamma x_i^2 + \varepsilon_i, \ i = 1, \ldots, 5,$$

där y_i = försäljning (kkr) beror av x_i = reklamkostnad (kkr) och ε_i betecknar slumpmässiga fel. Minsta-kvadrat-skattningarna av regressionskoefficienterna $\ , \beta$ och γ blev $\ ^*_{obs} = 239.3$, $\beta^*_{obs} = 3.42$ respektive $\gamma^*_{obs} = \ \ 0.026$.
95%-iga konfidensintervall $I_\beta(0.95)$ och $I_\gamma(0.95)$ har beräknats, $I_\beta(0.95) = (0.88, 5.97)$ samt $I_\gamma(0.95) = (\ \ 0.055, 0.0035)$.
Man kontrollerar vidare om den effekt som reklamkostnaden har på försäljningen är signifikant, dvs man testar $H_{0,\beta} : \beta = 0$ mot $H_{1,\beta} : \beta \neq 0$ samt $H_{0,\gamma} : \gamma = 0$ mot $H_{1,\gamma} : \gamma \neq 0$. Vidare beräknas två P-värden för de två testen.
Vilken slutsats kan man dra på 5% signifikansnivå?

A: P-värdet för testet $H_{0,\beta} : \beta = 0$ är mindre än 0.05, och P-värdet för testet $H_{0,\gamma} : \gamma = 0$ är större än 0.05 och därmed är den linjära termen signifikant medan den kvadratiska termen inte är signifikant .

B: P-värdet för testet $H_{0,\beta} : \beta = 0$ är större än 0.05, och P-värdet för testet $H_{0,\gamma} : \gamma = 0$ är mindre än 0.05 och därmed är den linjära termen signifikant medan den kvadratiska termen inte är signifikant .

C: P-värdet för testet $H_{0,\beta} : \beta = 0$ är större än 0.05, och P-värdet för testet $H_{0,\gamma} : \gamma = 0$ är mindre än 0.05 och därmed är den linjära termen inte signifikant medan den kvadratiska termen är signifikant .

D: P-värdet för testet $H_{0,\beta} : \beta = 0$ är mindre än 0.05, och P-värdet för testet $H_{0,\gamma} : \gamma = 0$ är större än 0.05 och därmed är den linjära termen inte signifikant medan den kvadratiska termen är signifikant .

Del II

Uppgift 13

En stad utgörs av fyra öar—vi kan kalla dem A, B, C och D—vilka är sammanbundna med broar enligt följande: en bro mellan A och B, en bro mellan B och C, en bro mellan C och D, samt två broar mellan A och C. Sannolikheten att en bro är öppen är p och broarna är öppna oberoende av varandra.

a) Vad är sannolikheten att precis två broar är öppna? (2 p)

b) Vad är sannolikheten att man kan ta sig från ön A till B givet att det är precis två broar öppna? (3 p)

c) Antag att du befinner dig på ön A. Vad är sannolikheten att broarna är öppna på ett sätt så att du kan ta dig från A till ön D? (5 p)

Uppgift 14

Ett företag vill testa en ny kampanj och har delat in potentiella kunder i tre segment: (i) Gen(eration)-X, (ii) Gen-Y och (iii) slumpmässig (ingen åldersrestriktion). Varje potentiell kund får ett meddelande med en länk till en hemsida för att registrera sig hos företaget. Huruvida en kund klickar på länken eller ej registreras och resultatet ges i tabellen nedan. **??**

Segment	Kontaktade	Klicks	Klick-rate
Slump	6132	464	7.57 %
Gen-X	6080	363	5.97%
Gen-Y	6111	665	10.88%

Undersök på nivån $= 0.05$ om det finns någon skillnad i kundernas beteende mellan de olika segmenten. Var noga med att ange dina hypoteser och motivera dina slutsatser. (10 p)

Uppgift 15

Betrakta polynomet

$$x^2 + 2Yx + Z \tag{1}$$

där koefficienterna är två diskreta stokastiska variabler Y och Z som är oberoende och likformigt fördelade på $\{0, 1, 2, 3, 4\}$.

a) Vad är sannolikheten att nollställena till (**??**) är reella? (6 p)

b) Vad är sannolikheten att (**??**) har endast en rot (en dubbelrot)? (4 p)

Uppgift 16

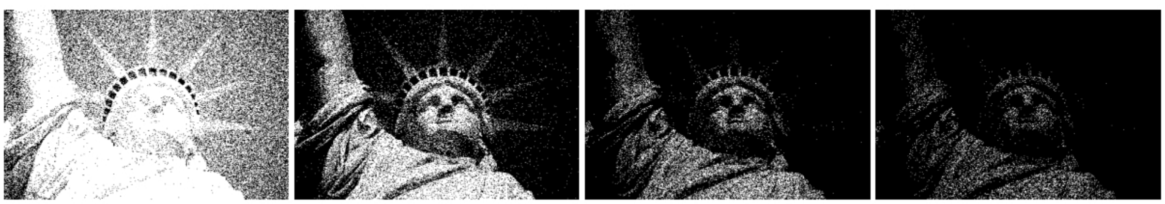
En Quanta Bildsensor (QBS) är en bildsensor som endast kan registrera ett fåtal fotoelektroner per pixel innan den mättas. För ett givet motiv kan antalet inkommande fotoelektroner X för en enstaka pixel antas vara Poisson-fördelat med en okänd parameter λ. I QBS:en registreras dock inte värdet X för en pixel utan istället används så kallat *enkel-bit* eller *multi-bit*. I enkel-bit så registreras

$$Y = \begin{cases} 1, & X \geq q, \\ 0, & X < q, \end{cases} \tag{2}$$

för någon tröskel $q \in$. I multi-bit registreras istället

$$Y = \begin{cases} X, & X < q, \\ q, & X \geq q, \end{cases} \tag{3}$$

åter igen för någon tröskel q. Bägge kan ses som en kvantisering av det inkommande flödet och valet av tröskelvr̈det q påverkar hur bilderna som genereras av QBS:en ser ut; t.ex. visar Figur **??** hur olika q påverkar en bild i fallet med enkel-bit kvantisering.



$q = 1$ $q = 5$ $q = 8$ $q = 10$

Figur 1: Olika versioner av en bild beroende på tröskelvärdet q i enkel-bit kvantisering för en QBS.

a) Härled sannolikhetsfunktionen för både enkel-bit och multi-bit fallen, för en godtycklig tröskel $q \in$. (3 p)

b) Beräkna väntevärdet $E[Y]$ i fallet där Y är en enkel-bit Poisson-variabel, definierad i (**??**). (2 p)

c) Betrakta nu fallet med enkel-bit kvantisering (**??**) med $q = 1$ och observationer $y_1, \ldots, y_n$ av ett oberoende stickprov $Y_1, \ldots, Y_n$. Härled maximum likelihood-skattningen av parametern λ. (5 p)

I (a) och (b) behöver svaren inte anges på sluten form, det går bra att uttrycka dem i termer av den den inkompletta Gamma-funktionen, som de nieras som, för $l \in$,

$$\Psi_l(\theta) = \frac{1}{(l)} \int_\theta^\infty t^{l\ \ 1} e^{\ \ t} dt, \quad \theta > 0.$$

En alternativ representation för $\Psi_l(\theta)$ är

$$\Psi_l(\theta) = \sum_{k=0}^{l\ \ 1} \frac{\theta^k e^{\ \ \theta}}{k!}.$$

Avd Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1923/SF1924 SANNOLIKHETSTEORI OCH STATISTIK, TISDAG 31 MAJ 2022 KL 8.00–13.00.

Del I

Del I Svar:

1. D

2. A

3. A

4. C

5. C

6. B

7. B

8. C

9. D

10. D

11. C

12. A

Del I Lösningsförslag:

Uppgift 1

Låt A beteckna att testet visar positivt medan S betecknar att personen har sjukdomen. Då söker vi $P(S^*|A)$ vilket kan beräknas genom Bayes sats

$$P(S^*|A) = \frac{P(S^* \cap A)}{P(A)} = \frac{P(A|S^*)P(S^*)}{P(A|S^*)P(S^*) + P(A|S)P(S)} = \frac{0.03 \cdot 0.995}{0.03 \cdot 0.995 + 0.98 \cdot 0.005} = 0.859.$$

Uppgift 2

$$P(Y > \frac{1}{2}) = P(-2X+1 > \frac{1}{2}) = P(-2X > -\frac{1}{2}) = P(X < \frac{1}{4}) = \int_0^{1/4} f_X(x)dx = \int_0^{1/4} 2xdx = [x^2]_0^{1/4} = \frac{1}{16}$$

Uppgift 3

$C(X,Y) = E(XY) - E(X)E(Y).$

$E(XY) = \sum xyp_{X,Y}(x,y) = 1 \cdot 1 \cdot 0.2 + 2 \cdot 1 \cdot 0.1 + 2 \cdot 2 \cdot 0.1 = 0.8.$

$E(X) = \sum xp_X(x) = 1 \cdot 0.3 + 2 \cdot 0.3 = 0.9.$

$E(Y) = \sum yp_Y(y) = 1 \cdot 0.5 + 2 \cdot 0.2 = 0.9.$

Dvs. $C(X,Y) = 0.8 - 0.9 \cdot 0.9 = -0.01.$

Uppgift 4

Låt X vara antalet lastbilar med körförbud som polisen kontrollerat.
Då gäller att $X \in Hyp(N,n,p) = Hyp(18,4,\frac{4}{18})$ och då blir

$$p_X(2) = \frac{\binom{Np}{2}\binom{N(1-p)}{n-2}}{\binom{N}{n}} = \frac{\binom{4}{2}\binom{14}{2}}{\binom{18}{4}} = \frac{6 \cdot 91}{3060} = 0.178$$

Uppgift 5

Låt X_i vara längden av samtal i. $X_i \in Exp(\frac{1}{4})$. Då gäller att $\mu = E(X_i) = 4$ och att $\sigma = D(X_i) = 4$. $Y = X_1 + X_2 + ...X_{100}$ är då enligt Centrala Gränsvärdessatsen approximativt Normalfördelad så att $Y \sim N(n\mu, \sigma\sqrt{n}) = N(100 \cdot 4, 4 \cdot \sqrt{100}) = N(400, 40)$

$$P(Y > 6 \cdot 60) = [\text{gör om till} N(0,1)] = P(\frac{Y-400}{40} > \frac{360-400}{40}) = [Z = \frac{Y-400}{40}] = P(Z > -1) =$$

$$= 1 - P(Z < -1) = 1 - \Phi(-1) = 1 - [1 - \Phi(1)] = 0.841$$

Uppgift 6

Låt X vara antalet samtal den första 10-minutersperioden och Y vara antalet samtal den andra 10-minutersperioden.
X och Y är oberoende och $X \in Po(1.3)$ och $Y \in Po(1.3)$. Då gäller att $Z = X+Y \in Po(1.3+1.3) = Po(2.6)$. Då gäller att $P(2 < Z < 5) =$ [tab 5]$= P(Z \leq 4) \quad P(Z \leq 2) = 0.87742 \quad 0.51843 = 0.35899 \approx 0.36$

Uppgift 7

MK-skattning: Se §9.2 i F.S.

$$Q = \sum_{i=1}^{2} (x_i \quad \mu_i(\theta))^2 = (\quad 3.6 + 2.7\theta)^2 + (9.3 \quad 7.9\theta)^2$$

$$\frac{dQ}{d\theta} = 2(\quad 3.6 + 2.7\theta)2.7 + 2(9.3 \quad 7.9)(\quad 7.9)$$

$$\frac{dQ}{d\theta} = 0 \Rightarrow (2.7^2 + 7.9^2)\theta \quad (3.6 \cdot 2.7 + 7.9 \cdot 9.3) = 0$$

$$\Rightarrow \theta^*_{obs_{MK}} = 1.19$$

Uppgift 8

Om man har två väntevärdesriktiga skattningar av θ så är det den skattningen av de två som har lägst varians som är den effektivaste.Vi kontrollerar först om skattningarna är väntevärdesriktiga. Då ska väntevärdena för skattningarna av θ vara lika med θ.

$$E(\theta^*) = E(\frac{X + 2Y}{4}) = \frac{E(X) + 2E(Y)}{4} = \frac{2\theta + 2\theta}{4} = \theta$$

$$E(\hat{\theta}) = E(\frac{X + Y}{3}) = \frac{E(X) + E(Y)}{3} = \frac{2\theta + \theta}{3} = \theta$$

$E(\theta^*) = E(\hat{\theta}) = \theta$. Således är båda skattingarna väntevärdesriktiga.

$$V(\theta^*) = V(\frac{X + 2Y}{4}) = \frac{V(X) + 4V(Y)}{16} = \frac{\frac{(4\theta \quad 0)^2}{12} + 4\frac{(2\theta \quad 0)^2}{12}}{16} = \frac{\frac{(16+16)\theta^2}{12}}{16} = \frac{\theta^2}{6}$$

$$V(\hat{\theta}) = V(\frac{X + Y}{3}) = \frac{V(X) + V(Y)}{9} = \frac{\frac{(4\theta \quad 0)^2}{12} + \frac{(2\theta \quad 0)^2}{12}}{9} = \frac{\frac{(16+4)\theta^2}{12}}{12} = \frac{20\theta^2}{9 \cdot 12} = \frac{5\theta^2}{27}$$

$$V(\theta^*) = \frac{\theta^2}{6} = \frac{5\theta^2}{30} < V(\hat{\theta}) = \frac{5\theta^2}{27}$$

Alltså är θ^*_{obs} den effektivaste skattningen av de två

Uppgift 9

Ett tvåsidigt konfidensintervall för skillnaden mellan två stickprovs väntevärden där vi antar att stickproven har samma okända varians fås m.h.a. §12.2 och § 11.2 till

$$I_{\mu_x \;\; \mu_y} = \bar{x} \quad \bar{y} \pm s \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \cdot t_{\frac{}{2}}(n_x + n_y \quad 2)$$

där

$$s^2 = \frac{(n_x \quad 1) \cdot s_x^2 + (n_y \quad 1) \cdot s_y^2}{n_x + n_y \quad 2} = \frac{(10 \quad 1) \cdot 4.06 + (10 \quad 1) \cdot 3.48}{10 + 10 \quad 2} = 3.77$$

Intervallet $I_{\mu_x \;\; \mu_y})$ nu skrivas

$$I_{\mu_x \;\; \mu_y} = 19.5 \quad 13.9 \pm \sqrt{3.77} \cdot \sqrt{\frac{1}{10} + \frac{1}{10}} \cdot t_{0.025}(18) = 5.6 \pm \sqrt{\frac{3.77}{5}} \cdot 2.10 = 5.6 \pm 1.82$$

Viket ger att övre gränsen blir 7.42.

Uppgift 10

Låt X vara antalet fungerande kretsar vi finner. $X \in Bin(n,p) = Bin(800,p)$. Vi får då det tvåsidiga konfidensintervallet till

$$I_p = p^*_{obs} \pm \sqrt{\frac{p^*_{obs} \cdot (1 \quad p^*_{obs})}{n}} \cdot \lambda_{\frac{}{2}}$$

Då blir det ensidigt nedåt begränsade konfidensintervallet

$$I_p = (p^*_{obs} \quad \sqrt{\frac{p^*_{obs} \cdot (1 \quad p^*_{obs})}{n}} \cdot \lambda \;\; , \infty) = (\frac{742}{800} \quad \sqrt{\left(\frac{\frac{742}{800} \cdot (1 \quad \frac{742}{800})}{800}\right)} \cdot \lambda_{0.10}, \infty) =$$

$$= (\frac{742}{800} \quad \sqrt{\left(\frac{\frac{742}{800} \cdot (1 \quad \frac{742}{800})}{800}\right)} \cdot 1.2816, \infty) = (0.916, \infty)$$

Uppgift 11

Styrkan hos testet är $P(\text{förkasta} H_0)$ om H_1 är sann.
Dvs. vi ska ta fram $P(X > 2)$ om $X \in Bin(8, 0.25)$
$P(X > 2) = 1 \quad P(X \leq 2) = [\text{ se tab 6 }] = 1 \quad 0.68 = 0.32$.

Uppgift 12

Eftersom $0 \notin I_\beta(0.95)$ så är också P-värdet för testet $H_{0,\beta} : \beta = 0$ mindre än 0.05. Då kan vi förkasta $H_{0,\beta} : \beta = 0$ mot $H_{1,\beta} : \beta \neq 0$ och därmed är den linjära termen signifikant. Eftersom $0 \in I_\gamma(0.95)$ så är också P-värdet för testet $H_{0,\gamma} : \gamma = 0$ större än 0.05. Då kan vi inte förkasta $H_{0,\gamma} : \gamma = 0$ mot $H_{1,\gamma} : \gamma \neq 0$ och därmed är den kvadratiska termen inte signifikant.

Del II

Del II Lösningsförslag:

Uppgift 13

Vi ger här lösningar för bägge tolkningarna av att en bro är öppen: att det betyder att den är öppen för trafik samt att den är uppfälld och alltså ej färdbar. Noterat att för uppgift (a) spelar tolkningen ingen roll.

Version 1: öppen=färdbar
(a) Antal broar som är öppna är $Bin(5, p)$-fördelat, vilket ger att sannolikheten att två broar är öppna är

$$\binom{5}{2} p^2(1 \quad p)^3 = 10p^2(1 \quad p)^3$$

<u>Svar:</u> Sannolikheten att exakt två broar är öppna är $10p^2(1 \quad p)^3$.

(b) Låt E vara händelsen att exakt två broar är öppna och F händelsen att det går att ta sig från A till B. Vi söker $P(F|E)$ som per definition ges av

$$\frac{P(F \cap E)}{P(E)}.$$

Det finns $\binom{5}{2} = 10$ olika sätt för två exakt två broar att vara öppna och dessa utgör händelsen E. Händelsen $F \cap E$ består då av fallen, bland de tio, där det också går att ta sig från A till B. Det finns totalt sex uppsättningar med två broar som ger en sådan väg: fyra möjliga kombinationer som innehåler bron mellan A och B samt två som inte innehåller den bron. Varje specifik kombination av två öppna broar har samma sannolikhet, $p^2(1 \quad p)^3$. Tillsammans ger det att

$$P(F|E) = \frac{6}{10} = 0.6$$

<u>Svar:</u> Sannolikheten att det går att ta sig från A till B givet att exakt två broar är öppna är 0.6.

(c) Om det ska gå att nå ön D måste bron CD vara öppen samt minst en av följande gälla: (i) minst en av de två broarna mellan A och C är öppna, eller (ii) broarna mellan A och B samt mellan B och C är bägge öppna. Då får vi $pP((i) \cup (ii)) = p[P(i) + P(ii) \quad P(i)P(ii)]$. Sammantaget har dessa alternativ då sannolikheten

$$p[[1 \quad (1 \quad p)^2] + p^2 \quad p^2[1 \quad (1 \quad p)^2]] = ... = 2p^2 \quad 2p^4 + p^5.$$

<u>Svar:</u> Sannolikheten att det går att nå ön D är $2p^2 \quad 2p^4 + p^5$.

Version 1: öppen=ej färdbar
(a): som ovan då tolkningen av öppen inte spelar någon roll här.
(b): samma typ av resonemang som ovan leder till sannolikheten $9/10 = 0.9$.

(c) Svaret ovan med p bytt till 1 $\quad p$ kommer nu ge sannolikheten för att kunna nå D:

$$1 \quad p \quad 2p^3 + 3p^4 \quad p^5.$$

Uppgift 14

Vi identifierar det här som en situation där homogenitetstest är passande för att undersöka huruvida det finns en skillnad i beteende mellan de olika segmenten. Som nollhypotes tar vi alltså att det inte är någon skillnad i kundbeteende mellan de tre grupperna (dvs. grupperna är homogena i detta avseende).

Som ett första steg utökar vi därför tabellen med annan nyttig information:

Segment	Kontaktade	Klicks	Ej klick
Slump	6132	464	5668
Gen-X	6080	363	5717
Gen-Y	6111	665	5446
Summa	18323	1492	16831

$$\text{Eftersom } \frac{6080 \times 1492}{18323} \geq 5 \text{ är alla } \frac{n_i \times mj}{N} \geq 5 \text{ och då är det O.K. med homogenitetstest.}$$

Vi har här $s = 3$ serier (segmenten) med $r = 2$ olika utfall ('klicks' och 'ej klick'). Vi bildar kvadratsumman enligt formelsamlingen:

$$\begin{aligned}
Q_{obs} &= \frac{464 \quad \frac{6132\times1492}{18323})^2}{\frac{6132\times1492}{18323}} + \frac{5668 \quad \frac{6132\times16832}{18323})^2}{\frac{6132\times16832}{18323}} + \frac{363 \quad \frac{6080\times1492}{18323})^2}{\frac{6080\times1492}{18323}} + \frac{5717 \quad \frac{6080\times16831}{18323})^2}{\frac{6080\times16832}{18323}} \\
&\quad \frac{665 \quad \frac{6111\times1492}{18323})^2}{\frac{6111\times1492}{18323}} + \frac{5446 \quad \frac{6111\times16831}{18323})^2}{\frac{6111\times16831}{18323}} \\
&= \frac{(464 \quad 499.31)^2}{499.31} + \frac{(5668 \quad 5633.02)^2}{5633.02} + \frac{(363 \quad 495.08)^2}{495.08} + \frac{(5717 \quad 5585.20)^2}{5585.20} \\
&\quad + \frac{(665 \quad 497.60)^2}{497.60} + \frac{(5446 \quad 5613.73)^2}{5613.73} \\
&= 2.497 + 0.217 + 35.237 + 3.110 + 56.316 + 5.012 \\
&= 102.389
\end{aligned}$$

Under noll-hypotesen att det inte är någon skillnad i kundbeteendet mellan grupperna är Q_{obs} ett utfall av den stokastiska variabeln Q som approximativt har en χ^2-fördelning med $(s \quad 1)(r \quad 1) = 2$ frihetsgrader. Med $\quad = 0.05$ är -kantilen i $\chi^2(2)$-fördelningen

$$\chi^2_{0.05}(2) = 5.99.$$

Vi ser därmed att vi på nivå 5% kan förkasta hypotesen om ett homogent beteende mellan de tre segmenten. Som en extra koll kan vi också beräkna p-värdet: med $Q \sim \chi^2(2)$,

$$p = P(Q > Q_{obs}) = P(Q > 102.389) < 2.2 \times 10^{\quad 16},$$

vilket också leder till slutsatsen att förkasta hypotesen om lika beteenden mellan kundsegmenten.

<u>Svar:</u> På nivån 5% kan vi förkasta hypotesen om lika beteende mellan kundsegmenten.

Uppgift 15

Lösningarna till ekvationen

$$x^2 + 2Yx + Z = 0$$

ges av

$$x = \quad Y \pm \sqrt{Y^2 \quad Z}.$$

(a) Vi har reella rötter om $Y^2 \geq Z$. Sannolikheten för reella rötter ges alltså av

$$P(Y^2 \geq Z) = I \quad P(Y^2 < Z).$$

Vi ser att utfallen som ger $\{Y^2 < Z\}$ är paren $(0,1), (0,2), (0,3), (0,4), (1,2), (1,3), (1,4)$. Det är totalt sju utfall av 25 möjliga. Eftersom vi har likformig fördelning för Y och Z samt oberoende ger nu den klassiska sannolikhetsdefinitionen att $P(Y^2 < Z) = 7/25$, vilket ger

$$P(Y^2 \geq Z) = \frac{18}{25}.$$

<u>Svar:</u> Sannolikheten att polynomet har endast reella rötter är 18/25.

(b) Frå formen på lösningar x ser vi att ekvationen har en dubbelrot om $Y^2 = Z$. Gynsamma utfall för den händelsen är paren $(0,0), (1,1)$ och $(2,4)$. Sannolikheten för att ekvationen ska ha en dubbelrot är därmed 3/25.
<u>Svar:</u> Sannolikheten att polynomet har en dubbelrot är 3/25.

Uppgift 16

(a) Vi börjar med fallet med enkel-bit,

$$Y = \begin{cases} 1, & X \geq q, \\ 0, & X < q, \end{cases}$$

med $X \in Po(\lambda)$. Vi ser att Y i själva verket är en Bernoulli-variabel med sannolikhet för 1 resp. 0 som bestäms av valet av q och Poisson-fördelningen. Sannolikhetsfunktionen ges därför av

$$p_Y(1) = P(Y = 1) = \sum_{k=q}^{\infty} \frac{\lambda^k e^{\quad \lambda}}{k!},$$

$$p_Y(0) = P(Y = 0) = \sum_{k=0}^{q \quad 1} \frac{\lambda^k e^{\quad \lambda}}{k!}.$$

Med definitionen av den inkompletta Gamma-funktionen kan vi också uttrycka ovan som

$$p_Y(1) = 1 \quad \Psi_q(\lambda), \quad p_Y(0) = \Psi_q(\lambda).$$

Vi betraktar nu fallet med multi-bit,

$$Y = \begin{cases} X, & X < q, \\ q, & X \geq q. \end{cases}$$

Eftersom $Y = X$ för $X < q$ har vi att, för $y \in \quad ,$

$$p_Y(y) = \frac{\lambda^y e^{\ \lambda}}{y!}, \quad y < q.$$

Utöver alla $y < q$ kan Y endast anta värdet q, vilket sker med sannolikhet

$$p_Y(q) = P(X \geq q) = \sum_{k=q}^{\infty} \frac{\lambda^k e^{\ \lambda}}{k!} = 1 \quad \Psi_q(\lambda).$$

<u>Svar:</u> För enkel-bit fallet ges sannolikhetsfunktionen av $p_Y(0) = \Psi_q(\lambda)$, $p_Y(1) = 1 \quad \Psi_q(\lambda)$. För multi-bit fallet har vi istället $p_Y(y) = \lambda^y e^{\ \lambda/y}$ för $y < q$ och $p_Y(q) = 1 \quad \Psi_q(\lambda)$. [Notera att samtliga svar kan anges med Ψ_q-funktionen ersatt av motsvarande summor].

(b) Givet sannolikhetsfunktionen i (a) har vi vänteärdet

$$\begin{aligned} E[Y] &= 1p_Y(1) + 0p_Y(0) \\ &= \sum_{k=q}^{\infty} \frac{\lambda^k e^{\ \lambda}}{k!} \\ &= 1 \quad \Psi_q(\lambda). \end{aligned}$$

<u>Svar:</u> För enkel-bit ges väntevärdet av $E[Y] = \sum_{k=q}^{\infty} \frac{\lambda^k e^{\ \lambda}}{k} = 1 \quad \Psi_q(\lambda)$.

(c) För enkel-bit med $q = 1$ har Y sannolikhetsfunktion

$$\begin{aligned} p_Y(1) &= P(X \geq 1) = 1 \quad e^{\ \lambda}, \\ p_Y(0) &= e^{\ \lambda}. \end{aligned}$$

Dvs.$Y \in Bin(1, 1 \quad e^{\ \lambda})$. Likelihood-funktionen givet observationerna $y_1, \ldots, y_n$ ges därför av

$$\begin{aligned} L(\lambda; y_1, \ldots, y_n) &= \prod_{i=1}^{n} (1 \quad e^{\ \lambda})^{y_i} (e^{\ \lambda})^{1 \ y_i} \\ &= (1 \quad e^{\ \lambda})^{\ \sum_{i=}^{n} y_i} e^{\ \lambda \left(n \quad \sum_{i=}^{n} y_i\right)} \end{aligned}$$

Motsvarande log-likelihood ges därför av

$$\begin{aligned} l(\lambda; y_1, \ldots, y_n) &= \ln L(\lambda; y_1, \ldots, y_n) \\ &= \ln \ 1 \quad e^{\ \lambda}) \sum_{i=1}^{n} y_i \quad \lambda \left(n \quad \sum_{i=1}^{n} y_i \right). \end{aligned}$$

För att hitta ML-skattningen av λ deriverar vi l:

$$\frac{dl(\lambda; y_1, \ldots, y_n)}{d\lambda} = \frac{e^{\ \lambda}}{1 \quad e^{\ \lambda}} \sum_{i=1}^{n} y_i \quad n + \sum_{i=1}^{n} y_i.$$

Låt $\bar{y}_n = (1/n)\sum_{i=1}^{n} y_i$. Vi sätter ovan till 0 och hittar motsvarande extrempunkt:

$$\begin{aligned}
\frac{dl(\lambda; y_1, \ldots, y_n)}{d\lambda} = 0 &\Leftrightarrow \frac{e^{\;\;\lambda}}{1 \quad e^{\;\;\lambda}} \sum_{i=1}^{n} y_i = n \quad\;\; \sum_{i=1}^{n} y_i \\
&\Leftrightarrow e^{\;\;\lambda}\bar{y}_n = 1 \quad e^{\;\;\lambda} \quad \bar{y}_n + e^{\;\;\lambda}\bar{y}_n \\
&\Leftrightarrow e^{\;\;\lambda} = 1 \quad \bar{y}_n \\
&\Leftrightarrow \quad \lambda = \ln(1 \quad \bar{y}_n).
\end{aligned}$$

Vi har alltså att ML-skattningen ges av

$$\hat{\lambda}_{ML} = \quad \ln(1 \quad \bar{y}_n) = \quad \ln\left(1 \quad \frac{1}{n}\sum_{i=1}^{n} y_i\right).$$

<u>Svar:</u> ML-skattningen av λ är $\hat{\lambda}_{ML} = \quad \ln\ 1 \quad \frac{1}{n}\sum_{i=1}^{n} y_i\big)$.
Notera att del (c) också kan lösas på ett smidigt sätt genom att se att $\sum_i y_i$ är antalet 1:or som observerats, och $n \quad \sum_i y_i$ antalet 0:or, samt att använda sig av en binomialfördelningen med parametrar n och sannolikhet $1 \quad e^{\;\;\lambda}$. Då vet vi vad ML-skattningen av sannolikheten är och akn därifrån hitta ML-skattningen av λ.