



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1917/SF1919 SANNOLIKHETSTEORI OCH STATISTIK,  
ONSDAG 3 APRIL 2024 KL 8.00–13.00.

*Examinator:* Mykola Shykula, 08-790 66 44

*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 och varje korrekt svar ger 1 poäng. På denna del ska endast svar anges, i form av val av ett av de möjliga svarsalternativen. Svaren ska anges på svarsblanketten (utdelas vid tentamen). Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1–3, utan får tillgodoräkna sig dessa tre uppgifter (i svarsblanketten anges ordet "Bonus"). Studenter som är godkända på den andra datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges ordet "Bonus"). 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 eller får komplettera del I och poäng på del II krävs för högre betyg än E. På denna del ska resonemang och uträkningar vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar ska förklaras och definieras samt numeriska svar ska anges med minst tre värdesiffrors noggrannhet. Studenter som är godkända på den andra datorlaborationen får 3 bonuspoäng på del II.

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

Händelserna  $F$  och  $G$  är oberoende med  $P(F) = 0.4$  och  $P(G) = 0.4$ . Låt  $S = F \cap G$ . Bestäm den betingade sannolikheten  $P(F^* | S^*)$ .

A: 0.714

B: 0.4

C: 1

D: 0.19

**Uppgift 2**

Antag att  $X$  och  $Y$  är stokastiska variabler så att  $V(X) = 1$ ,  $C(X, Y) = 1$ ,  $V(2X + 3Y) = 39$ . Beräkna  $D(Y)$ .

A: 1.599

B: 4.690

C: 8.063

D: 8.368

**Uppgift 3**

Varje dag på förskolan löper Oscar en risk att smittas av förkylning. Sannolikheten att han smittas under en dag är 0.013. Låt  $X$  vara antalet dagar utan att han smittas (dvs. om han smittas på dag ett har vi  $X = 0$ ). Vad är sannolikheten att  $X > 3$ ?

A: 0.974

B: 0.962

C: 0.949

D: 0.987

**Uppgift 4**

Antag att  $X$  är  $U(0, 5)$ -fördelad, att  $Y$  är  $\text{Exp}(1)$ -fördelad och att  $X$  och  $Y$  är oberoende. Beräkna sannolikheten  $P(\min(X, Y) \leq 3)$ .

A: 0.0498

B: 0.43

C: 0.57

D: 0.98

**Uppgift 5**

Låt  $X$  och  $Y$  vara kontinuerliga och oberoende stokastiska variabler. Deras täthetsfunktioner är  $f_X(x)$  respektive  $f_Y(y)$ , där

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty,$$
$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \quad -\infty < y < \infty.$$

Beräkna sannolikheten  $P(|X + Y| \leq 1)$ , där  $|\cdot|$  betecknar absolutbeloppet.

- A: ca 0.76
- B: ca 0.52
- C: ca 0.69
- D: ca 0.38

**Uppgift 6**

Låt  $X_1$  vara Po(5)-fördelad,  $X_2$  vara Po(3)-fördelad och  $X_3$  vara Po(1)-fördelad. Antag att  $X_1$ ,  $X_2$  och  $X_3$  är oberoende. Låt  $Y = X_1 + X_2 + X_3$ . Beräkna sannolikheten  $P(7 < Y \leq 9)$ .

- A: 0.2635
- B: 0.3806
- C: 0.4033
- D: 0.5874

**Uppgift 7**

Låt  $X_i$ ,  $i = 1, 2, 3, 4$  vara oberoende stokastiska variabler med väntevärde  $\mu$  och standardavvikelse  $\sigma$ . Antag att  $\mu$  skattas med  $\mu_{\text{obs}}^* = (x_1 + 2x_2 + 3x_3 + 4x_4)/10$  och  $\sigma$  skattas med stickprovsstandardavvikelsen  $s$ , där  $s^2 = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2$ . Låt utfallen på  $X_i$  vara  $x_1 = 92$ ,  $x_2 = 85$ ,  $x_3 = 101$  och  $x_4 = 98$ . Bestäm medelfelet för skattningen av  $\mu$ .

- A: 3.35
- B: 3.87
- C: 6.12
- D: 7.07

**Uppgift 8**

Låt  $\bar{x} = 137.0$ ,  $\bar{y} = 118.0$ ,  $s_x = 9.5$ ,  $s_y = 3.0$ ,  $n_x = 10$  och  $n_y = 10$  vara givet och antag att  $X_i \in N(\mu_x, \sigma_x)$  och  $Y_i \in N(\mu_y, \sigma_y)$  samt att alla dessa stokastiska variabler är oberoende. Antag även att  $\sigma_x$  och  $\sigma_y$  är okända och olika.

Ange undre gränsen för det tvåsidiga konfidensintervallet  $I_{\mu_x - \mu_y}$  som har den approximativa konfidensgraden 99%.

- A: 9.9
- B: 10.9
- C: 11.7
- D: 13.9

**Uppgift 9**

Ur ett stort varuparti tog man ut ett stickprov på 500 enheter. Av dessa fann man att 87 var felaktiga. Ange övre gränsen för det ensidiga uppåt begränsade konfidensintervallet för andelen felaktiga enheter i partiet som har den approximativa konfidensgraden 99%.

- A: 0.174
- B: 0.213
- C: 0.218
- D: 0.226

**Uppgift 10**

$X_1 \in U(0, \theta)$ ,  $X_2 \in U(0, 2\theta)$ , och  $X_3 \in U(0, 3\theta)$  där  $X_1$ ,  $X_2$  och  $X_3$  är oberoende stokastiska variabler.

För att skatta  $\theta$  kan man använda sig av

$$\theta_{\text{obs}}^* = \frac{x_1 + x_2 + x_3}{3} \quad \text{och}$$

$$\hat{\theta}_{\text{obs}} = \frac{x_1}{2} + \frac{x_2}{4} + \frac{x_3}{6}.$$

Vilket av nedanstående påståenden är sant?

- A:  $\theta_{\text{obs}}^*$  är den effektivaste skattningen av  $\theta$  av de två.
- B:  $\hat{\theta}_{\text{obs}}$  är den effektivaste skattningen av  $\theta$  av de två.
- C: Bägge skattningarna är lika effektiva.
- D: Man kan inte avgöra vilken av skattningarna som är effektivast, eftersom minst en av dem inte är väntevärdesriktig.

**Uppgift 11**

Låt  $X$  vara  $N(\mu, \sigma)$ -fördelad, där  $\sigma = 3$ , och anta att vi testar  $H_0 : \mu = 0$  mot  $H_1 : \mu > 0$ . Låt  $x_1, x_2, \dots, x_9$  vara nio oberoende mätningar av  $X$ . Vi beräknar testvariabeln  $z = \frac{\bar{x}}{\sigma/\sqrt{9}}$ , där  $\bar{x}$  är medelvärdet, och vi förkastar  $H_0$  på nivån 10% om  $z > 1.282$ . Bestäm testets styrka för alternativet  $H_1 : \mu = 2.0$ .

A: 0.59

B: 0.76

C: 0.90

D: 0.98

**Uppgift 12**

Ett bibliotek försöker få bukt med sina problem med sent återlämnade böcker genom att höja avgiften som måste betalas vid sen återlämning. Efter att ha provat olika avgiften kom de fram till följande resultat:

Avgift (kr)	0 kr	5 kr	10 kr	20 kr
Snittid för sen återlämning (dagar)	11.2	9.7	10.1	8.1

Sätt upp en enkel linjär regressionsmodell för återlämningstiden som funktion av avgiften och skatta lutningen för denna.

A: 0.59

B: 11.0

C: -0.14

D: -30.6

**Var god vänd!**

## Del II

### Uppgift 13

Ett spelbolag har ställt upp en enarmad bandit som har sannolikheten 2.64% att ge en vinst på 1000 kr givet en insats på 30 kr. Antag att den spelas på 100 gånger per dag. Hur många dagar tar det innan banditen med minst 95% sannolikhet har genererat en nettovinst för spelbolaget på minst 10 000 kr? Motivera eventuella antaganden eller approximationer. (10 p)

### Uppgift 14

Den tvådimensionella stokastiska variabeln  $(X, Y)$  har täthetsfunktionen

$$f_{X,Y}(x, y) = \begin{cases} e^{-y}, & 0 < x < y < \infty, \\ 0, & \text{för övrigt.} \end{cases}$$

Bestäm väntevärdet  $E(XY)$ . (10 p)

### Uppgift 15

Antag att månadshyrorna i kronor för studenter i en mindre svensk stad beskrivs av en normalfördelad population med medelvärde  $\mu$  och standardavvikelse  $\sigma = 620$  kronor.

Antag att en undersökning av 12 studenter gav stickprovsmedelvärdet 3509 kronor. Tag fram ett 95%-igt konfidensintervall för den genomsnittliga månadshyran  $\mu$ , och beräkna p-värdet för testet av hypotesen  $H_0: \mu = 3700$  kronor mot hypotesen  $H_1: \mu \neq 3700$  kronor. Kan  $H_0$  förkastas på risknivån 5%? Motivera din slutsats ordentligt. (10 p)

### Uppgift 16

Låt  $X$  vara fördelad enligt  $\text{Bin}(n, p)$  och anta att  $p$  är *känd*. Beräkna ML-skattningen för  $n$  givet en observation  $X = x$ . (10 p)

**Lycka till!**



Avd. Matematisk statistik

**KTH Matematik**

LÖSNINGSFÖRSLAG TENTAMEN I SF1917/SF1918  
SANNOLIKHETSTEORI OCH STATISTIK,  
ONSDAG 3 APRIL 2024 KL 8.00–13.00.

**Del I, Svar.**

1. A

2. A

3. C

4. D

5. B

6. A

7. B

8. B

9. B

10. D

11. B

12. C

**Del I, Lösningsförslag.****Uppgift 1**

Rita Venn-diagram. Vi har:

$$P(F^*|S^*) = \frac{P(F^* \cap S^*)}{P(S^*)} = \frac{1 - P(F)}{1 - P(F)P(G)} = \frac{1 - 0.4}{1 - (0.4)(0.4)} \approx 0.714$$

Svaret är A.

**Uppgift 2**

$$V(2X + 3Y) = 2^2V(X) + 3^2V(Y) + 2 \cdot 2 \cdot 3 \cdot C(X, Y) = 75$$

vilket ger

$$V(Y) = \frac{1}{9}(75 - 4 \cdot 1 - 2 \cdot 2 \cdot 3 \cdot 4) = \frac{47}{9} \approx 2.556$$

så vi har  $D(Y) = \sqrt{V(Y)} = \sqrt{2.556} \approx 1.599$ , alltså alternativ A.

**Uppgift 3**

Sannolikheten att Oscar klarar sig  $x$  dagar utan att smittas men smittas på dag  $x + 1$  är  $(1 - 0.013)^x 0.013$ , dvs.  $X$  har sannolikhetsfunktionen

$$p_X(x) = (1 - 0.013)^x 0.013 = 0.987^x 0.013$$

Vi får alltså

$$P(X > 3) = 1 - P(X \leq 3) = 1 - (p_X(0) + p_X(1) + p_X(2) + p_X(3)) = 1 - 0.013(1 + 0.987 + 0.987^2 + 0.987^3) = 0.949$$

Svaret är därmed C.

**Uppgift 4**

$$\begin{aligned} P(\min(X, Y) \leq 3) &= 1 - P(\min(X, Y) \geq 3) = [\text{ober}] = \\ &= 1 - P(X \geq 3) \cdot P(Y \geq 3) = 1 - \left(\frac{2}{5}\right) \cdot \left(\int_3^\infty e^{-x} dx\right) = 1 - \frac{2}{5} \cdot e^{-3} \approx 0.98 \end{aligned}$$

Svaret är D.

**Uppgift 5**

Täthetsfunktioner för  $X$  respektive  $Y$  ger att:

1)  $X \in N(0, 1)$ , dvs  $\mu_X = 0$ ,  $\sigma_X = 1$ ,

2)  $Y \in N(0, 1)$ , dvs  $\mu_Y = 0$ ,  $\sigma_Y = 1$ .

Det är känt även att  $X$  och  $Y$  är oberoende.

Följaktligen,  $\eta = X + Y \in N(0, \sqrt{2})$ .

Därför,  $P(|\eta| \leq 1) = P(-1 \leq \eta \leq 1) = P(\eta \leq 1) - P(\eta \leq -1) =$

$$= \Phi\left(\frac{1-0}{\sqrt{2}}\right) - \Phi\left(\frac{-1-0}{\sqrt{2}}\right) = 2\Phi\left(\frac{1}{\sqrt{2}}\right) - 1 =$$

$$= 2 \cdot \Phi(0.71) - 1 \simeq |\text{Tabell 1}| \simeq 2 \cdot (0.7611) - 1 \simeq 0.5222$$

Svaret är B.

### Uppgift 6

Eftersom  $Y = X_1 + X_2 + X_3$  är fördelad enligt  $Po(9)$  har vi från Tabell 5:

$$\begin{aligned} P(7 < Y \leq 9) &= P(Y \leq 9) - P(Y \leq 7) = \\ &= 0.58741 - 0.32390 \approx 0.2635 \end{aligned}$$

Svaret är A.

### Uppgift 7

$$\begin{aligned} V(\mu_{\text{obs}}^*) &= V\left(\frac{X_1 + 2X_2 + 3X_3 + 4X_4}{10}\right) \\ &= \frac{1}{100} (V(X_1) + 4V(X_2) + 9V(X_3) + 16V(X_4)) \\ &= \frac{1}{100} (30\sigma^2) \\ &= \frac{3}{10}\sigma^2 \end{aligned}$$

Alltså blir

$$D(\mu_{\text{obs}}^*) = \sqrt{\frac{3}{10}} \sigma$$

Vi uppskattar  $\sigma$  med  $s$ , så att  $d(\mu_{\text{obs}}^*) = \sqrt{\frac{3}{10}} s$ .

Då  $x_1 = 92$ ,  $x_2 = 85$ ,  $x_3 = 101$  och  $x_4 = 98$  blir  $\bar{x} = 94$  och  $\sum_{i=1}^4 (x_i - \bar{x})^2 = 2^2 + 9^2 + 7^2 + 4^2 = 150$ . Eftersom  $n = 4$  blir  $s^2 = \sum_{i=1}^4 (x_i - \bar{x})^2 / 3 = 50$  och således  $s = \sqrt{50}$ . Alltså blir

$$d(\mu_{\text{obs}}^*) = \sqrt{\frac{3}{10}} s = \sqrt{\frac{3}{10}} \cdot \sqrt{50} = 3.87$$

### Uppgift 8

Eftersom standardavvikelserna är okända och olika fås enligt §12.3 följande approximativa konfidensintervall:

$$\begin{aligned} I_{\mu_x - \mu_y} &= \bar{x} - \bar{y} \pm \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \cdot \lambda_{\frac{\alpha}{2}} = 137 - 118 \pm \sqrt{\frac{9.5^2}{10} + \frac{3.0^2}{10}} \cdot \lambda_{0.005} = \\ &= 19 \pm \sqrt{\frac{9.5^2}{10} + \frac{3.0^2}{10}} \cdot 2.5758 = 19 \pm 8.115 \end{aligned}$$

Undre gränsen blir alltså  $10.885 \approx 10.9$ .

**Uppgift 9**

Låt  $X$  vara antalet felaktiga enheter vi finner.  $X \in \text{Bin}(n, p) = \text{Bin}(500, p)$ .

Eftersom  $np_{obs}^*(1 - p_{obs}^*) = 500 \cdot \frac{87}{500} \cdot (1 - \frac{87}{500}) \geq 10$  så är  $X$  approximativt Normalfördelad och p.g.a. att  $p^* = \frac{X}{n}$  är en linjärkombination av  $X$ , så kan vi använda oss av §12.3 i F.S. och får då

$$I_p = p_{obs}^* \pm D_{obs}^* \cdot \lambda_{\frac{\alpha}{2}}$$

Vi får då det tvåsidiga konfidensintervallet till

$$I_p = p_{obs}^* \pm \sqrt{\frac{p_{obs}^* \cdot (1 - p_{obs}^*)}{n}} \cdot \lambda_{\frac{\alpha}{2}}$$

Det ensidigt uppåt begränsade konfidensintervallet blir då

$$\begin{aligned} I_p &= (-\infty, p_{obs}^* + \sqrt{\frac{p_{obs}^* \cdot (1 - p_{obs}^*)}{n}} \cdot \lambda_{\alpha}) = (-\infty, \frac{87}{500} + \sqrt{\left(\frac{87}{500} \cdot (1 - \frac{87}{500})\right)} \cdot \lambda_{0.01}) = \\ &= (-\infty, \frac{87}{500} + \sqrt{\frac{87 \cdot (1 - \frac{87}{500})}{500}} \cdot 2.3263) = (-\infty, 0.213) \end{aligned}$$

**Uppgift 10**

Vi börjar med att kontrollera om skattningarna är väntevärdesriktiga. Dvs. om  $E(\theta^*) = \theta$  och om  $E(\hat{\theta}) = \theta$ .

$$E(\theta^*) = E\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{\theta + \theta + \frac{3\theta}{2}}{3} = \theta$$

$$E(\hat{\theta}) = E\left(\frac{X_1}{2} + \frac{X_2}{4} + \frac{X_3}{6}\right) = \frac{\theta}{2} + \frac{\theta}{4} + \frac{3\theta}{6} = \frac{3\theta}{4} \neq \theta$$

Alltså är D rätt svarsalternativ.

**Uppgift 11**

Testets styrka är  $P(\text{förförkasta } H_0) \text{ om } H_1 \text{ är sann.}$

Vi ska alltså räkna ut  $P(Z > 1.282)$  om  $\mu = 2.0$

$$P\left(\frac{\bar{X}}{\sigma/\sqrt{9}} > 1.282\right) = P\left(\frac{\bar{X}}{3/\sqrt{9}} > 1.282\right) = P(\bar{X} > 1.282)$$

Nu gör vi om till  $N(0, 1)$

$$\begin{aligned} P(\bar{X} > 1.282) &= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{9}} > \frac{1.282 - \mu}{\sigma/\sqrt{9}}\right) = [\sigma = 3, \text{ och } \mu = 2 \text{ enligt } H_1] = \\ &= P(W > 1.282 - 2.0) = P(W > -0.718) \approx P(W > -0.72) \text{ där } W \in N(0, 1) \end{aligned}$$

$$P(W > -0.72) = P(W < 0.72) = \Phi(0.72) = 0.76$$

### Uppgift 12

Från datan har vi

$$\sum_i x_i = 35 \quad \sum_i x_i^2 = 525 \quad \sum_i y_i = 39.1 \quad \sum_i y_i^2 = 387.15 \quad \sum_i x_i y_i = 311.5,$$

vilket ger  $\bar{x} = 8.75$ ,  $\bar{y} = 9.775$  och

$$s_{xy} = 311.5 - 4 \cdot 8.75 \cdot 9.775 = -30.625$$

samt

$$s_{xx} = 525 - 4 \cdot 8.75^2 = 218.75.$$

Vi får då

$$\beta_{\text{obs}}^* = \frac{s_{xy}}{s_{xx}} = -0.14,$$

dvs. svaret är C.

**Del II, Lösningsförslag:****Uppgift 13**

Vinst efter  $n$  spel är  $Y = 30n - 1000X$  där  $X$  är  $\text{Bin}(n, 0.0264)$ -fördelad. Vi vet att  $n$  kommer vara minst 100 (en dags spelande), så vi kan använda centrala gränsvärdeessatsen för att approximera  $X$ . Den har då den approximativa fördelningen  $N(0.0264n, 0.1603\sqrt{n})$ . Vinsten  $Y$  efter  $n$  spel har då den approximativa fördelningen

$$N(30n - 1000 \cdot 0.0264n, 1000 \cdot 0.1603\sqrt{n}) = N(3.600n, 160.3\sqrt{n})$$

Vi söker nu  $n$  så att

$$P(Y > 10000) \approx P(3.600n + 160.3\sqrt{n}Z > 10000) = P\left(Z > \frac{10000 - 3.600n}{160.3\sqrt{n}}\right) = 0.95$$

där  $Z$  är standardnormalfördelad. Vi ser då att

$$\frac{10000 - 3.600n}{160.3\sqrt{n}} = \lambda_{0.95} = -\lambda_{0.05} = -1.6449,$$

vilket ger ekvationen (efter en del omskrivning):

$$3.600n - 263.7\sqrt{n} - 10000 = 0.$$

Detta är en andragradsekvation i  $\sqrt{n}$  med lösningarna

$$\sqrt{n} = \frac{263.7 \pm \sqrt{263.7^2 + 4 \cdot 3.600 \cdot 10000}}{2 \cdot 3.600} = 36.625 \pm 64.181.$$

Eftersom  $\sqrt{n} \geq 0$  är den negativa lösningen ogiltig och vi får  $\sqrt{n} = 100.81$ , vilket i sin tur ger  $n = 10162$ .

För att uppnå en sannolikhet på minst 95% att vinsten överstiger 10 000 kr tar då  $\lceil 10162/100 \rceil = 102$  dagar.

**Uppgift 14**

Vi har:

$$E(XY) = \int_0^\infty dx \int_x^\infty xye^{-y} dy = \int_0^\infty x dx \left( \int_x^\infty ye^{-y} dy \right).$$

Mha den partiella integrationen får vi integralen i parantesen till:

$$\int_x^\infty ye^{-y} dy = |y = u, e^{-y} dy = dv| = xe^{-x} + \int_x^\infty e^{-y} dy = xe^{-x} + e^{-x} = (x+1)e^{-x}.$$

Vidare, vi har:

$$E(XY) = \int_0^\infty x(x+1)e^{-x} dx = \int_0^\infty x^2e^{-x} dx + \int_0^\infty xe^{-x} dx = I_1 + I_2,$$

där  $I_1 = E(\xi^2)$  och  $I_2 = E(\xi)$ , för  $\xi \in \text{Exp}(\lambda)$ ,  $\lambda = 1$ . Nu, eftersom  $E(\xi) = \frac{1}{\lambda} = 1$  och  $E(\xi^2) = V(\xi) + (E(\xi))^2 = \frac{1}{\lambda^2} + 1^2 = 2$ , så är det sökta väntevärdet:

$$E(XY) = I_1 + I_2 = E(\xi^2) + E(\xi) = 2 + 1 = 3$$

**Uppgift 15**

Ett 95% konfidensintervall för  $\mu$  ges av

$$\bar{x} \pm \lambda_{0.025} \frac{\sigma}{\sqrt{n}} = 3509 \pm 1.96 \frac{620}{\sqrt{12}} = 3509 \pm 351 = \underline{(3158, 3860)} \text{ kronor.}$$

Hypotesen  $\mu = 3700$  förkastas till förmån för  $\mu \neq 3700$  om teststorheten

$$t_{obs} = \left| \frac{\bar{x} - 3700}{\sigma/\sqrt{n}} \right| = \left| \frac{3509 - 3700}{620/\sqrt{12}} \right| = 1.07$$

är stor och förkastelsegränserna bestäms ur en standardnormalfördelning. Går vi in i tab 1 och använder den baklänges ser vi att arean till höger om 1.07 är  $1 - 0.86 = 0.14$ . Nu är p-värdet p.g.a. att vi ska göra ett tvåsidigt test summan av arean till vänster om  $-1.07$  och arean till höger om  $1.07$  dvs 0.28

Eftersom p-värdet är större än 0.05, så kan inte  $H_0$  förkastas på risknivån 5%. En annan motivering är att  $3700 \in I_\mu$ .

**Uppgift 16**

Log-likelihoodfunktionen ges av

$$\log L(n) = \log \binom{n}{x} + x \log p + (n-x) \log(1-p) = \log n! - \log x! - \log(n-x)! + x \log p + (n-x) \log(1-p).$$

Om vi betraktar skillnaden mellan  $\log L(n)$  och  $\log L(n-1)$  får vi

$$\log L(n) - \log L(n-1) = \log n - \log(n-x) + \log(1-p).$$

För att  $n_{obs}^*$  ska utgöra ett maximum vill vi att denna skillnad ska vara positiv för alla  $n \leq n_{obs}^*$  och negativ för alla  $n > n_{obs}^*$ . Vi har då

$$\log L(n) - \log L(n-1) = \log n - \log(n-x) + \log(1-p) > 0 \Leftrightarrow \frac{n}{n-x} > \frac{1}{1-p} \Leftrightarrow n < \frac{x}{p}.$$

Låt oss anta att  $x/p$  inte är ett heltal. I så fall har vi att  $L(n)$  växer till och med  $\lfloor x/p \rfloor$  ( $x/p$  avrundat nedåt) och sedan avtar. Vi har alltså ML-skattningen  $\lfloor x/p \rfloor$ .

Om  $x/p$  är ett heltal ser vi ovan att  $\log L(n) - \log L(n-1) = 0$  för  $n = x/p$ . Alltså växer  $L(n)$  fram till  $x/p - 1$  och avtar först efter  $x/p$  har passerats. Med andra ord har vi två maxima, och därmed två ML-skattningar:  $x/p - 1$  och  $x/p$ .

Sammanfattningsvis har vi alltså att  $\lfloor x/p \rfloor$  alltid är en ML-skattning för  $n$  men om  $x/p$  är ett heltal så är också  $x/p - 1$  en ML-skattning.