Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1920/SF1921 SANNOLIKHETSLÄRA OCH STATISTIK,
MÅNDAG 14 MARS 2022 KL 08.00–13.00

Examinator: Mykola Shykula.

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. Svaren anges på svarsblanketten.
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). Detta gäller endast på den här tentan och vid omtentamen i juni 2022. 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 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 den andra datorlaborationen får 3 bonuspoäng på del II på det här ordinarie tentamenstillfället och omtentamenstillfället i juni 2022.

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

Låt A, B och C vara oberoende händelser på ett utfallsrum. Antag att $P(A) = P(B) = 0.5$ och $P(C) = 0.3$. Beräkna sannolikheten att högst två av de tre händelserna inträffar.

Uppgift 2

Den stokastiska variabeln X har täthetsfunktionen

$$f_X(x) = \begin{cases} \frac{c}{x^4}, & \text{om} \quad x \geq 1, \\ 0, & \text{annars,} \end{cases}$$

där c är en konstant. Beräkna sannolikheten $P(X \geq 5)$.

A: 0.8

B: 0.08

C: 0.008

D: 0.0008

Uppgift 3

De stokastiska variablerna X och Y har den simultana sannolikhetsfunktionen

$p_{X,Y}(j,k)$	0	1	2
0	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{12}$

Således antar X alltså värdena 0 eller 1, medan Y antar värdena 0, 1, eller 2. Bestäm $V(X^2)$.

A: 0.06

B: 0.24

C: 0.49

D: 0.58

Uppgift 4

En restaurang serverar 8 förrätter med fisk, 12 med nötkött och 10 med fjäderfä. Om kunderna beställer från dessa slumpmässigt (dvs var och en av de trettio förrätterna har samma chans att väljas vid varje beställning), vad är då sannolikheten för att exakt två av de kommande fyra kunderna beställer fiskrätter?

A: 0.07

B: 0.12

C: 0.18

D: 0.23

Uppgift 5

Låt X och Y vara två oberoende Poissonfördelade stokastiska variabler. $E(X) = E(Y) = 2$. Bestäm sannolikheten att $X + Y$ antar ett värde större än 5, givet att $X + Y$ är större än 2.

A: 0.28

B: 0.41

C: 0.57

D: 0.76

Uppgift 6

Antag att X och Y är två oberoende Normalfördelade stokastiska variabler med väntevärdena $E(X) = E(Y) = 0$ och standardavvikelserna $D(X) = D(Y) = 3$. Beräkna $P(|X - Y| > 2)$.

A: 0.08

B: 0.36

C: 0.64

D: 0.92

Uppgift 7

Låt datamängden 82, 91, 122 vara given. Data anses vara utfall av stokastiska variabler X_i, där X_i:na antas vara oberoende och $X_i \in Po(\mu)$. Bestäm övre gränsen till det ensidigt uppåt begränsade konfidensintervallet för μ som har den approximativa konfidensgraden 95%.

A: 107.75

B: 109.55

C: 114.64

D: 117.77

Uppgift 8

Låt x_i, $i = 1, \ldots, n$ vara oberoende observationer av $X_i \in Exp(\lambda)$, $i = 1, \ldots, n$ där λ är en okänd parameter. ML-skattningen av väntevärdet $\mu = 1/\lambda$ ges då av

$$\mu^*_{obs} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Vad av följande är rimligt att använda som medelfel för ML-skattningen?

A: $\dfrac{1}{n\lambda^2}$

B: $\dfrac{\bar{x}^2}{n}$

C: $\dfrac{1}{\sqrt{n}\lambda}$

D: $\dfrac{\bar{x}}{\sqrt{n}}$

Uppgift 9

Låt datamängden 28, 19, 22 vara given. Data anses vara utfall av stokastiska variabler X_i, där X_i:na antas vara oberoende och likafördelade mellan θ och 7θ. Dvs $X_i \in U(\theta, 7\theta)$. Bestäm Minsta-Kvadrat-skattningen av θ m.h.a. dessa tre observationer.

A: 3.83

B: 4.00

C: 4.66

D: 5.75

Uppgift 10

Antag att $X \in \text{ffg}(p)$ och låt H_0 vara att $p = 0.1$. Vi vill testa H_0 mot alternativet $H_1 : p > 0.1$ och förkastar H_0 om vi observerar ett litet värde x på X.
Antag att vi observerat $x = 3$. Bestäm testets P-värde.

A: 0.18

B: 0.21

C: 0.24

D: 0.27

Uppgift 11

I ett land med 8 stora partier har två väljarundersökningar gjorts med tre månaders mellanrum. För att avgöra om den politiska opinionen förblivit oförändrad mellan undersökningarna gör man ett homogenitetstest och får teststorheten $Q = 19.7$. Den politiska opinionen representeras här av fördelningen mellan de 8 partierna.
Vilket av följande påståenden är sant?

A: H_0 kan varken förkastas på risknivån 1% eller risknivån 5%

B: H_0 kan både förkastas på risknivån 1% och risknivån 5%

C: H_0 kan förkastas på risknivån 1%, men inte på risknivån 5%

D: H_0 kan förkastas på risknivån 5%, men inte på risknivån 1%

Uppgift 12

Vid enkel linjär regression gäller följande samband

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

där man observerat paren (x_i, y_i), $i = 1, \dots, n$ och ε_i betecknar de slumpmässiga felen. Man önskar testa nollhypotesen $H_0 : \beta = 0$ mot $H_1 : \beta \neq 0$. Ett 99%-igt konfidensintervall $I_\beta(0.99)$ och ett 95%-igt konfidensintervall $I_\beta(0.95)$ har beräknats: $I_\beta(0.99) = (-0.414, 4.194)$, $I_\beta(0.95) = (0.210, 3.570)$. Vilket av följande påståenden är sant?

A: Vi kan förkasta H_0 och kan därmed förkasta att y_i:na beror av x_i:na på risknivån 1%

B: Vi kan förkasta H_0 och kan därmed förkasta att y_i:na beror av x_i:na på risknivån 5%

C: Vi kan förkasta H_0 och kan därmed inte förkasta att y_i:na beror av x_i:na på risknivån 1%

D: Vi kan förkasta H_0 och kan därmed inte förkasta att y_i:na beror av x_i:na på risknivån 5%

Var god vänd!

Del II

Uppgift 13

En komponent består av *två delar* med livslängder som kan uppfattas som exponentialfördelade med väntevärden 6 månader respektive 3 månader. Komponenten fungerar så länge som båda delarna fungerar. För att höja livslängden parallellkopplas två sådana komponenter till ett system som fungerar så länge som minst en av komponenterna fungerar. Alla de 4 delarnas livslängder är oberoende.

Beräkna sannolikheten att systemet fungerar åtminstone två månader. (10 p)

Uppgift 14

I ett system sitter en komponent som har en livslängd X timmar som är exponentialfördelad med väntevärde 100 timmar. Så snart komponenten brister byts den ut mot en ny likadan som i sin tur byts ut då den brister o.s.v. Utbytestiden är exakt 10 timmar, och under denna tid är systemet nere. Systemet är nere enbart då komponentbyte görs. Livslängderna av komponenterna är oberoende.

Beräkna sannolikheten att systemet fungerar minst 8000 timmar under ett år (8760 timmar). Välmotiverade approximationer är tillåtna. (10 p)

Uppgift 15

För att jämföra två metoder att bestämma halten koppar i legeringar analyserades fem olika legeringar med båda metoderna. Resultat (%koppar):

Legering	1	2	3	4	5
Metod B	13.4	17.9	9.7	15.0	12.9
Metod A	13.9	17.6	10.3	15.4	13.3

Antag normalfördelade observationer och undersök om det föreligger någon skillnad mellan metoderna A och B på signifikansnivån 5%. Ange tydligt alla dina hypoteser och motivera din slutsats noga. (10 p)

Uppgift 16

Ett sätt att skatta storleken av en okänd population, t.ex. ett fiskbestånd i en sjö, är att använda “capture-recapture-tekniken”.

Antag att det i en sjö finns N fiskar. Man fångar f stycken av dessa, märker dem och släpper dem tillbaka i sjön. Sedan fångar man n fiskar och noterar hur många av dem som är märkta. Beteckna detta antal med X.

Vi antar att varje fisk har samma sannolikhet att bli fångad.

a) Vilken fördelning har X? (2 p)

b) Ge en lämplig väntevärdesriktig skattning av sannolikheten att fånga en fisk som är märkt, och beräkna därur en skattning av populationsstorleken N. Ange det numeriska värdet på skattad populationsstorlek om $f = n = 1000$ och man återfångade $x = 20$ märkta fiskar. (3 p)

Var god vänd!

c) Beräkna ett approximativt 95 % konfidensintervall för populationsstorleken N då n, f och x antar värden som i uppgift b).

Ledning: Beräkna först ett konfidensintervall för sannolikheten att fånga en fisk som är märkt. (5 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG
TENTAMEN I SF1920/SF1921 SANNOLIKHETSLÄRA OCH STATISTIK,
MÅNDAG 14 MARS 2022 KL 8.00–13.00.

Del I, Svar

1. 0.925

2. C

3. B

4. D

5. A

6. C

7. A

8. D

9. D

10. D

11. B

12. D

Del I, Lösningsförslag

Uppgift 1

P(högst två händelser inträffar) = 1-P(alla tre händelserna inträffar) $= 1 - 0.5 * 0.5 * 0.3 = 0.925$

Uppgift 2

$F_X(x) = \int_1^{\infty} \frac{c}{x^4} dx = 1 - \frac{3c}{x^3}$
$F_X(1) = 0 \Rightarrow c = \frac{1}{3}$
$P(X \geq 5) = 1 - F_X(5) = 1 - (1 - \frac{1}{5^3}) = 0.008$

Uppgift 3

Sätt $Z = X^2 \Rightarrow p_Z(0) = \frac{5}{12}$ och $p_Z(1) = \frac{7}{12}$

$E(Z) = 0 \cdot \frac{5}{12} + 1 \cdot \frac{7}{12} = \cdot \frac{7}{12}$
$E(Z^2) = 0^2 \cdot \frac{5}{12} + 1^2 \cdot \frac{7}{12} = \cdot \frac{7}{12}$
$V(Z) = E(Z^2) - E^2(Z) = \frac{7}{12} - (\frac{7}{12})^2 = 0.243$

Uppgift 4

Antal fiskrätter som beställs kan sägas vara en stokastisk variabel X
där $X \in Bin(4, \frac{8}{30}) = Bin(4, \frac{4}{15})$
$p_X(2) = \binom{4}{2}(\frac{4}{15})^2 \cdot (1 - \frac{4}{15})^{4-2} = 0.23$

Uppgift 5

Eftersom X och Y är två oberoende Poissonfördelade stokastiska variabler med väntevärdena $E(X) = E(Y) = 2$, så gäller att $Z = X + Y \in Po(2 + 2) = Po(4)$.

$$P(Z > 5|Z > 2) = \frac{P(Z > 5 \cap Z > 2)}{P(Z > 2)} = \frac{P(Z > 5)}{P(Z > 2)} = \frac{1 - P(Z \leq 5)}{1 - P(Z \leq 2)} = [\text{tab } 5] = \frac{1 - 0.78513}{1 - 023810} = 0.28$$

Uppgift 6

$X \in N(0, 3)$och$Y \in N(0, 3)$.Sätt$Z = X - Y \Rightarrow Z \in N(0 - 0, \sqrt{3^2 + 3^2}) = N(0, \sqrt{18})$
$P(|Z| > 2) = [\text{p.g.a.symmetri kring 0}] = 2P(Z > 2) =$ gör om till $N(0, 1) =$

$$= 2(1 - \Phi(\frac{2 - 0}{\sqrt{18}})) \approx 2(1 - \Phi(0.47)) = 2(1 - 0.6808) = 0.64$$

Uppgift 7

Eftersom väntevärdet $E(X) = \mu$ så skattar vi μ med $\bar{x} = \frac{82+91+122}{3} = 98.33$ Dvs $\mu^*_{obs} = 98.33 \geq 15$ och då kan vi approximera $Po(\mu)$ till $N(\mu, \sqrt{\mu})$ enl §6 i Formelsamlingen och därmed bilda ett konfidensintervall med approximativ konfidensgrad enl. §12.3.
Enligt §12.3 blir då konfidensintervallet $I_\mu = \mu^*_{obs} \pm D^*_{obs}(\mu^*) \cdot \lambda_{\frac{\alpha}{2}}$. Men nu har vi ett uppåt begränsat

konfidensintervall och då fås $I_\mu = \mu^*_{obs} + D^*_{obs}(\mu^*) \cdot \lambda_\alpha$
$\mu^*_{obs} = \bar{x} \Rightarrow D(\mu^*) = D(\bar{X}) = \frac{D(X)}{\sqrt{3}} = \frac{\sqrt{\mu}}{\sqrt{3}} \Rightarrow D^*_{obs}(\mu^*) = \frac{\sqrt{\bar{x}}}{\sqrt{3}}$
Dvs.$I_\mu = \bar{x} + \frac{\sqrt{\bar{x}}}{\sqrt{3}} \cdot \lambda_{0.05} = 98.33 + \frac{\sqrt{98.33}}{\sqrt{3}} \cdot 1.6449 = 107.75$

Uppgift 8

Medelfelet är ju enligt §9.3 i F.S $D^*_{obs}(\theta^*)$. I detta fall $D^*_{obs}(\mu^*)$.
$D(\mu^*) = D(\bar{X}) = \frac{D(X_i)}{\sqrt{n}}$
[Enl §4 är$D(X_i) = \mu$] $\Rightarrow D(\mu^*) = \frac{\mu}{\sqrt{n}} \Rightarrow D^*_{obs}(\mu^*) = \frac{\bar{x}}{\sqrt{n}}$

Uppgift 9

Enligt § 9.2 har vi

$$Q = \sum_{i=1}^{3}(x_i - \mu_i(\mu))^2 = \sum_{i=1}^{3}(x_i - E(X_i))^2 = (28-4\theta)^2 + (19-4\theta)^2 + (22-4\theta)^2$$

$$\frac{dQ}{d\theta} = 0 \Rightarrow 2(28-4\theta)4 + 2(19-4\theta)4 + 2(22-4\theta)4 = 0 \Rightarrow 69 = 12\theta \Rightarrow \theta^*_{obs_{MK}} = \frac{69}{12} = 5.75$$

Uppgift 10

P-värdet är sannolikheten att förkasta nollhypotesen om nollhypotesen är sann utgående från den observation vi fått. I detta fall blr P-värdet
$P(X \leq 3)$om$p = 0.1$ Då fås

$$P(X \leq 3) = \sum_{i=1}^{3} P(X = i) = p + p(1-p) + p(1-p)^2 = 0.1 + 0.1 \cdot 0.9 + 0.1 \cdot 0.9^2 = 0.27$$

Uppgift 11

$\chi^2_{0.01}((8-1)(2-1)) = \chi^2_{0.01}(7)$. Nollhypotesen H_0 är att skillnad ej föreligger. Vi går in i tab 4 i F.S och ser att $\chi^2_{0.01}(7) = 18.5$ och att $\chi^2_{0.05}(7) = 14.1$. $Q = 19$ är större än både 14.1 och 18.5, så på båda risknivåerna kan vi förkasta att det inte föreligger skillnad.

Uppgift 12

$\beta = 0 \notin I_\beta(0.95)$, så vi kan förkasta att $\beta = 0$ på 5%-nivån. Dämed kan vi inte förkasta att y_i:na beror av x_i:na på risknivån 5%.

Del II, Lösningsförslag

Uppgift 13

Låt X beteckna en komponents livslängd.
Exponentialfördelning ger $P(X > x) = \frac{1}{\mu}\int_x^\infty e^{-x/\mu}dx = e^{-x/\mu}$.
Oberoende ger

$$P(X > 2) = P(\text{komponentens båda delar håller 2 månader}) = e^{-2/6} \cdot e^{-2/3} = e^{-1}.$$

Oberoende och parallellkopplingen ger för systemet

$$P(\text{systemet fungerar minst 2 månader}) =$$

$$= 1 - P(\text{ingen av systemets komponenter håller 2 månader}) = 1 - (1 - e^{-1})^2 \approx \underline{0.60}.$$

Uppgift 14

Att systemet fungerar minst 8000 timmar under ett år är ekvivalent med att systemet ligger nere högst 760 timmar, vilket är ekvivalent med att högst 76 komponentbyten påbörjats under ett år. Låt X_i vara livslängden av komponent nummer i, $i = 1, 2, \ldots, 77$ som alltså är Exp(100). Därför är $E(X_i) = 100$ och variansen $V(X_i) = 100^2$. Vi skall beräkna sannolikheten för händelsen $X_1 + 10 + X_2 + 10 + \cdots + X_{76} + 10 + X_{77} > 8760$ (tiden fram till att det 77:e bytet påbörjas är mer än ett år), dvs händelsen $X_1 + X_2 + \cdots + X_{76} + X_{77} > 8000$.
Av centrala gränsvärdessatsen fås att

$$Y = X_1 + X_2 + \cdots + X_{77} \approx \mathrm{N}(77 \cdot 100, 100\sqrt{77})$$

Härav erhåller vi att

$$P(Y > 8000) = P\left(\frac{Y - 7700}{100\sqrt{77}} > \frac{8000 - 7700}{100\sqrt{77}}\right) \approx 1 - \Phi\left(\frac{8000 - 7700}{100\sqrt{77}}\right) =$$
$$1 - \Phi(0.3419) = \underline{0.37}$$

Uppgift 15

Parade observationer, dvs att för legering i är mätvärdena utfall av $N(\mu_i, \sigma_1)$ respektive $N(\mu_i + \Delta, \sigma_2)$. Alla variabler antas oberoende. Vi bildar parvisa skillnader, z_i: 0.5 -0.3 0.6 0.4 0.4. Man erhåller att $\bar{z}$=0,32 och s_z=0.3564. Det ger oss konfidensintervallet $I_{\mu_z} = \bar{z} \pm t_{0{,}025}(4)s_z/\sqrt{5}$ dvs $\underline{0{,}32 \pm 0{,}44}$ eftersom $t_{0{,}025}(4) = 2{,}78$.
Nollhypotesen H_0 är att vi inte har någon skillnad, dvs att$\Delta = 0$. Mothypotesen är att vi har skillnad, dvs att$\Delta \neq 0$.
Eftersom $0 \in I_{\mu_z}$ kan vi inte förkasta H_0 på signifikansnivån 5%

Uppgift 16

a) X har uppkommit genom att ur en population om N element (fiskar) slumpmässigt utan återläggning dra n stycken. Därför är $X \in \mathrm{Hyp}(N, n, p)$ där p är proportionen märkta fiskar, $p = f/N$.

b) p skattas lämpligen med relativa antalet märkta fiskar som fångats i andra omgången, d.v.s. med $p^* = \frac{x}{n}$. Denna skattning är väntevärdesriktig ty

$$E(p^*) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p$$

Eftersom $p = f/N$, skattas N lämpligen med $N^* = f/p^* = fn/x$. Med $f = n = 1000$ och $x = 20$ erhålles $p^* = 20/1000 = 0.02$ och $N^* = 1000^2/20 = \underline{50000}$.

c) Vi ser att n är litet i förhållande till populationsstorleken N. Vi har därför att $X \approx$Bin(n, p). Ett approximativt 95 % konfidensintervall för p ges därför av

$$p^* \pm \lambda_{0.025}\sqrt{p^*(1-p^*)/n}$$

varur intervallet $0.02 \pm 0.00868 = (0.01132, 0.02868)$ erhålls. Eftersom $N = 1000/p^*$ erhåller vi intervallet

$$(\frac{1000}{0.02868}, \frac{1000}{0.01132}) \approx \underline{(34900,\ 88300)}$$

som ett approximativt 95 % konfidensintervall för N.