Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1924 SANNOLIKHETSTEORI OCH STATISTIK,
MÅNDAG 12 AUGUSTI 2019 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 preliminärt 9 poäng. Möjlighet att komplettera ges för tentander med, preliminärt, 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

För händelserna A, B och C gäller att $P(A \cap B \cap C) = 0.1$, $P(A) = 0.5$ och $P(B \mid A) = 0.4$. Beräkna $P(C \mid A \cap B)$.

A: 0.10

B: 0.25

C: 0.40

D: 0.50

Uppgift 2

En kontinuerlig stokastisk variabel X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ \frac{1}{4} \cdot x^2 & \text{om } 0 \leq x < 2, \\ 1 & \text{om } x \geq 2. \end{cases}$$

Beräkna $D(X)$.

A: 0.22

B: 0.47

C: 0.69

D: 0.89

Uppgift 3

Låt $X_1, X_2, ...X_{20}$ vara stokastiska variabler sådana att varje $X_i \in U[0, 6]$, och $Y_1, Y_2, ...Y_{30}$ vara stokastiska variabler sådana att varje $Y_i \in U[0, 2]$. Vidare är alla stokastiska variabler oberoende. Bestäm $V(\bar{X} - \bar{Y})$.

Uppgift 4

$X \in N(2, 2)$. Bestäm a så att $P(|X| < a) = 0.95$.

Uppgift 5

I fotbolls-VM 2019 fanns 6 grupper med 4 lag i varje. Vidare från gruppspelet gick förutom de två bästa lagen från varje grupp även de 4 bästa grupptreorna. Låt oss anta att var och en av de 6 grupperna har samma sannolikhet att ”producera” en grupptrea som går vidare. Bestäm då sannolikheten att de 4 grupptreorna som går vidare kommer från grupperna A,B,C, och E.

A: 0.002

B: 0.033

C: 0.042

D: 0.067

Uppgift 6

X och Y är oberoende stokastiska variabler sådana att $X \in Po(0.4)$ och $Y \in Po(1.2)$. Bestäm $P(X + Y = 2)$.

Uppgift 7

Antag $X \in N(\mu, \sigma)$ och $Y \in N(2\mu, 2\sigma)$. Vidare antar vi att de stokastiska variablerna X och Y är oberoende. Vi skattar μ med

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

och skattar σ med

$$\sigma^*_{obs} = \sqrt{(x - \mu^*_{obs})^2 + (\frac{y}{2} - \mu^*_{obs})^2}$$

Vi har vidare fått de observerade värdena $x = 7$ och $y = 15$. Beräkna medelfelet av vår skattning av μ utgående från denna information.

A: 0.25

B: 0.50

C: 0.75

D: 1.00

Uppgift 8

En stokastisk variabel X har täthetsfunktionen

$$f_X(x) = \begin{cases} \theta \cdot x^{\theta - 1} & , \quad 0 < x < 1 \\ 0 & , \quad \text{f.ö.} \end{cases}$$

Vi har två oberoende observationer: $x_1 = \frac{1}{2}$ och $x_2 = \frac{1}{4}$.

Bestäm Minsta-kvadrat-skattningen av θ utifrån detta.

A: 0.2

B: 0.4

C: 0.6

D: 0.8

Uppgift 9

Vid ett reningsverk mäts dagligen syrekoncentrationen i vattnet (mg/l). Mätningarnas resultat anses vara utfall av oberoende stokastiska variabler som följer $N(\mu, \sigma)$-fördelning där σ betecknar standardavvikelsen. För en månads mätningar blev summan av de 31 mätvärdena 69.70 och stickprovsvariansen 1.2772. Ange den övre gränsen till det ensidigt uppåt begränsade 95%-iga konfidensintervallet för μ.

Uppgift 10

Antag att $X \in Bin(200, p)$. Vi gör ett försök och får $x = 23$. Ange den övre gränsen för det tvåsidiga konfidensintervallet I_p för p. Använd approximativ konfidensgrad 95%.

A: 0.152

B: 0.159

C: 0.162

D: 0.776

Uppgift 11

Vi antar att $X \in Exp(\lambda)$. Nollhypotesen H_0 är att $\mu = 4$, där $\mu = E(X)$. Mothypotesen H_1 är att $\mu > 4$. Vi har en observation : $x = 8$. Bestäm testets p-värde.

A: 13.5%

B: 39.3%

C: 60.7%

D: 86.5%

Uppgift 12

Den trendiga badbutikskedjan Poolen och Plurret säljer bl.a. badbyxor. Badbyxorna finns i fyra olika färger. Tabellen nedan visar antalet sålda badbyxor av respektive färg i respektive års undersökning.

Färg	Blå	Svart	Grön	Röd
Föregående år	943	357	498	202
Innevarande år	860	347	476	317

Nollhypotesen H_0 är att fördelningen av färgerna är oförändrad mellan de båda undersökningstillfällena.
Mothypotesen H_1 är då att det skett en förändring.
Man gör en viss typ av χ^2-test och får Q= 29.94
Vilken slutsats kan man dra då man fått dessa data?

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%.

Del II

Uppgift 13

Företagen A och B levererar en viss typ av maskinkomponenter. Livslängden för en maskinkomponent från företag A kan betraktas som en Exp(1)-fördelad stokastisk variabel och livslängden för en maskinkomponent från företag B kan betraktas som en Exp(2)-fördelad stokastisk variabel. Vid en industrianläggning där dessa komponenter används i stor skala förvaras 100 komponenter från företag A och 200 komponenter från företag B. Dessutom har komponenterna blandats och inga yttre markeringar avslöjar tillverkaren.

a) Vad är sannolikheten för att en komponent plockad på måfå ur industrianläggningens lager kommer att ha en livslängd som överstiger 2? (5 p)

b) Givet att ett en komponent plockad på måfå ur industrianläggningens lager har en livslängd som överstiger 2, vad är sannolikheten för att komponenten kommer från företag B? (5 p)

Uppgift 14

En person erbjuds att mot insatsen a kr få delta i följande spel: En vanlig tärning kastas upprepade gånger, tills ettan kommer upp för första gången varvid spelet slutar. För varje kast som ej givit en etta får personen 1 kr. Bestäm insatsen a så att spelet blir rättvist, dvs så att a är lika med väntevärdet av antalet kronor som personen erhåller. (10 p)

Uppgift 15

För att undersöka om en förpackningsmaskin är rätt inställd görs följande försök. Fem enheter tas ut. Dessa väges först med innehåll och sedan utan innehåll. Följande resultat erhölls:

Enhet nr	1	2	3	4	5
(i)	114	124	115	117	123
(ii)	18	23	19	20	24

där (i) står för uppmätt vikt för förpackning *med* innehåll och (ii) står för uppmätt vikt för förpackning *utan* innehåll.
Vi antar följande statistiska modell. Vägningarna av de tomma förpackningarna betraktas som utfall av oberoende stokastiska variabler $X_1, \ldots, X_5$, där $X_k \in N(\mu_k, \sigma_1)$. Konstanterna $\mu_1, \ldots, \mu_5$ är således förpackningarnas ”sanna” vikter. Vägningarna av de fyllda förpackningarna betraktas som utfall av oberoende stokastiska variabler $Y_1, \ldots, Y_5$ där $Y_k \in N(\mu_k + \Delta, \sigma_2)$. Konstanten Δ svarar således mot den vikt som förpackningsmaskinen är inställd på att fylla i. Både σ_1, som svarar mot spridningen hos vågens mätfel och σ_2, som svarar mot spridningen hos summan av förpackningsmaskinens och vågens mätfel är okända.

a) Bestäm ett 95% konfidensintervall för Δ. (7 p)

b) Testa hypotesen $\Delta = 100$ mot alternativet $\Delta \neq 100$ på signifikansnivån 5%. Det ska klart framgå om hypotesen förkastas eller ej. (3 p)

Uppgift 16

I en undersökning av Sveriges älgstam var man intresserad av förekomsten av en viss gen, som antingen saknas, finns i enkel uppsättning eller finns i dubbel uppsättning hos de olika älgarna. Man kontrollerade 1000 på måfå valda älgar och konstaterade att 11 st hade genen i dubbel uppsättning, 187 st hade den i enkel uppsättning och att de övriga 802 älgarna saknade genen. Som teoretisk modell antog man att antalet gener av den aktuella typen hos en älg är $\text{Bin}(2, p)$-fördelad och att de olika undersökta älgarnas gener var oberoende av varandra.

a) Beräkna Maximum Likelihood-skattningen av p baserad på ovanstående data. (5 p)

b) Teknologen Osquarulda känner inte till ML-metoden, men kom på intuitiva grunder fram till att p borde skattas med $p^* = \dfrac{x_1 + 2x_2}{2000}$ där x_1 = antalet älgar (av de 1000 undersökta) som hade genen i enkel uppsättning och x_2 = antalet älgar som hade genen i dubbel uppsättning. Undersök om p^* är en väntevärdesriktig skattning av p. (2 p)

c) Osquarulda vill dessutom räkna ut ett approximativt 95%-igt konfidensintervall för p baserad på sin skattning. Din uppgift är att utföra detta! (3 p)

<u>Ledning och varning:</u> x_1 och x_2 är <u>ej</u> utfall av oberoende stokastiska variabler. Fundera i stället över tolkningen av $x_1 + 2x_2$ när Du löser c-delen.

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1924 SANNOLIKHETSTEORI OCH STATISTIK, MÅNDAG 12 AUGUSTI 2019 KL 8.00–13.00.

Del I

Uppgift 1

Vi har

$$P(C \mid A \cap B) = \{\text{def. av betingning}\} = \frac{P(C \cap A \cap B)}{P(A \cap B)} = \frac{P(A \cap B \cap C)}{P(A \cap B)}$$

$$= \frac{0.1}{P(B \mid A)P(A)} = \frac{0.1}{0.4 \cdot 0.5} = \frac{0.1}{0.2} = \underline{0.5}.$$

Uppgift 2

Vi har

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ \frac{1}{4}x^2 & \text{om } 0 \leq x < 2, \\ 1 & \text{om } x \geq 2, \end{cases}$$

vilket ger

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

Detta ger

$$E(X) = \int_0^2 x \cdot \frac{1}{2}x\,dx = \left[\frac{1}{6}x^3\right]_0^2 = \underline{\frac{4}{3}} \quad \text{och} \quad E(X^2) = \int_0^2 x^2 \cdot \frac{1}{2}x\,dx = \left[\frac{1}{8}x^4\right]_0^2 = 2,$$

vilket ger

$$V(X) = E(X^2) - (E(X))^2 = 2 - \frac{16}{9} = \frac{2}{9} \text{ och således } D(X) = \sqrt{\frac{2}{9}} = \underline{0.471}.$$

Uppgift 3

$$V(\bar{X} - \bar{Y}) = \text{ober} = V(\bar{X}) + V(\bar{Y}) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_x^2}{n_x} = \frac{3}{20} + \frac{\frac{1}{3}}{30} = 0.161$$

Uppgift 4

$P(|X| < a) = 0.95 \Leftrightarrow P(-a < X < a) = 0.95.$

Gör om till N(0,1) då $X \in N(2,2)$.

$$P\left(\frac{-a-2}{2} < \frac{X-2}{2} < \frac{a-2}{2}\right) = 0.95$$

$$Y = \frac{X-2}{2} \in N(0,1)$$

$$P\left(\frac{-a-2}{2} < Y < \frac{a-2}{2}\right) = 0.95$$

$$\Leftrightarrow$$

$$P(Y > \frac{a-2}{2}) = 0.025 \Leftrightarrow \frac{a-2}{2} = \lambda_{0.025} = 1.96$$

$$\Rightarrow a = 2 \cdot 1.96 + 2 = 5.92$$

Uppgift 5

$P(ABCE)$= Antal gynnsamma utfall delat med antal möjliga utfall. =

$$= \frac{1}{\binom{6}{4}} = \frac{1}{15} = 0.0667$$

Uppgift 6

Eftersom X och Y är oberoende och $X \in Po(0.4)$ och $Y \in Po(1.2)$ gäller att $Z = X + Y \in Po(0.4 + 1.2) = Po(1.6)$. Då gäller att $P(Z = 2) =$ [tab 5]$= P(Z \leq 2) - P(Z \leq 1) = 0.78336 - 0.52493 = 0.25843$

Uppgift 7

$$\mu^*_{obs} = \frac{x}{2} + \frac{y}{4} = \frac{7}{2} + \frac{15}{4} = \frac{29}{4}$$

$$V(\mu^*) = \frac{1}{4}V(X) + \frac{1}{16}V(Y) = \frac{1}{4}\sigma^2 + \frac{1}{16} \cdot (2 \cdot \sigma)^2 = \frac{\sigma^2}{2}$$

D.v.s.

$$D(\mu^*) = \frac{\sigma}{\sqrt{2}}$$

medelfelet =

$$D^*(\mu^*)_{obs} = \frac{\sigma^*_{obs}}{\sqrt{2}}$$

$$\sigma^*_{obs} = \sqrt{(x - \mu^*_{obs})^2 + (\frac{y}{2} - \mu^*_{obs})^2} = \sqrt{(7 - \frac{29}{4})^2 + (\frac{15}{2} - \frac{29}{4})^2} = \frac{\sqrt{2}}{4}$$

$\Rightarrow$ medelfelet$= \frac{1}{4} = 0.25$

Uppgift 8

MK-metoden (se §9.2 i F.S.) $Q = \sum_{i=1}^{n}(x_i - \mu_i(\theta_1, \theta_2, ...\theta_k))^2$

Vi behöver

$$\mu(\theta) = E(X) = \int_0^1 \theta \cdot x^\theta dx =$$

$$= \theta \left[\frac{x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$$

$$\Rightarrow Q = \left(\frac{1}{2} - \frac{\theta}{\theta+1}\right)^2 + \left(\frac{1}{4} - \frac{\theta}{\theta+1}\right)^2$$

$$\frac{dQ}{d\theta} = 0 \Rightarrow 2\left[\frac{1}{2} - \frac{\theta}{\theta+1}\right]\left[\frac{-1}{(\theta+1)^2}\right] + 2\left[\frac{1}{4} - \frac{\theta}{\theta+1}\right]\left[\frac{-1}{(\theta+1)^2}\right] = 0$$

$$\Rightarrow 2 \cdot \frac{1}{2} + \frac{2}{4} = \frac{4\theta}{\theta+1}$$

$$\Rightarrow \frac{3}{8} = \frac{\theta}{\theta+1}$$

$\Rightarrow$ MK-skattningen av $\theta = \frac{3}{5} = 0.6$

Uppgift 9

Ett ensidigt uppåt begränsat 95%-igt konfidensintervall för μ då ett stickprov betraktas som observationer på $N(\mu, \sigma)$ ges som

$$I_\mu = (-\infty, \frac{1}{n}\sum_{i=1}^{31} x_i + t_{0.05}(n-1)\frac{s}{\sqrt{n}})$$

Då $n = 31, s^2 = 1.2772$ och $\sum_{i=1}^{31} x_i = 69.70$, samt $t_{0.05}(30) = 1.70$ blir den övre gränsen för $I_\mu = \frac{69.70}{31} + 1.70\frac{\sqrt{1.2772}}{\sqrt{31}} = 2.59$

Uppgift 10

Övre gränsen är

$$p_{obs}^* + \sqrt{\frac{p_{obs}^*(1-p_{obs}^*)}{n}} \cdot \lambda_{\frac{\alpha}{2}} = \frac{23}{200} + \sqrt{\frac{0.115(1-0.115)}{200}} \cdot \lambda_{0.025} =$$

$$= 0.115 + \sqrt{\frac{0.115(1-0.115)}{200}} \cdot 1.96 = 0.159$$

Uppgift 11

p-värdet = P(förkasta H_0) då H_0 är sann = [i detta fall] = $P(X \geq 8)$ om $\mu = 4$.

p-värdet är således

$$= \int_8^\infty \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_8^\infty = e^{-\frac{1}{4}\cdot 8} = 0.135$$

Uppgift 12

Här har vi homogenitetstest. Då är antalet frihetsgrader i detta fall = (r-1)(s-1) = (4-1)(2-1) = 3.

$\chi^2_{0.05}(3) = 7.81$. $\chi^2_{0.01}(3) = 11.3$. $Q = 29.94$. Alltså är Q större än både $\chi^2_{0.05}(3)$ och $\chi^2_{0.01}(3)$.

Detta medför att vi förkastar H_0 både på 5%-nivån och 1%-nivån.

Del II

Uppgift 13

a) Vi inför de båda stokastiska variablerna

$$X_A = \text{livslängd för en komponent från företag } A,$$

$$X_B = \text{livslängd för en komponent från företag } B.$$

Enligt uppgift har vi $X_A \in \text{Exp}(1)$ och $X_B \in \text{Exp}(2)$. Dessutom betraktar vi händelserna

$$L = \{\text{livslängd för en komponent plockad ur förrådet } > 2\},$$

$$A = \{\text{komponenten är från företag } A\} \quad \text{och} \quad B = \{\text{komponenten är från företag } B\}.$$

Enligt uppgift har vi att $P(A) = \frac{1}{3}$ och $P(B) = \frac{2}{3}$. Vi vill bestämma $P(L)$ och enligt lagen om total sannolikhet så gäller

$$P(L) = P(L|A)\,P(A) + P(L|B)\,P(B)\,.$$

Eftersom $X_A \in \text{Exp}(1)$, så har vi att

$$P(L|A) = P(X_A > 2) = \int_2^\infty e^{-x}dx = [-e^{-x}]_2^\infty = e^{-2},$$

och på samma sätt har vi att

$$P(L|B) = P(X_B > 2) = \int_2^\infty 2e^{-2x}dx = [-e^{-2x}]_2^\infty = e^{-4},$$

eftersom $X_B \in \text{Exp}(2)$. Därmed fås

$$P(L) = e^{-2} \cdot \frac{1}{3} + e^{-4} \cdot \frac{2}{3} \approx 0.0573.$$

Svar: Sannolikheten att en på måfå vald komponent har en livslängd som överstiger 2 är 5.7%.

b) Med beteckningarna från uppgift a), så söker vi nu $P(B|L)$. Enligt Bayes sats, så gäller det att

$$P(B|L) = \frac{P(L|B) \cdot P(B)}{P(L)} = \frac{e^{-4} \cdot \frac{2}{3}}{e^{-2} \cdot \frac{1}{3} + e^{-4} \cdot \frac{2}{3}} \approx 0.2130.$$

Svar: Sannolikheten att en på måfå vald komponent som har en livslängd som överstiger 2 kommer från företag B är 21.3%.

Uppgift 14

Vi inför den stokastiska variabeln $X = \{\text{antalet kast till och med första ettan}\}$ och inser att X är ffg(1/6)-fördelad, dvs

$$P(X = n) = \frac{1}{6}\left(\frac{5}{6}\right)^{n-1}, \quad \text{för } n = 1, 2, \ldots$$

Enligt formelsamlingen så gäller $E(X) = \frac{1}{1/6} = 6$. Låt nu den stokastiska variabeln Y beteckna antalet kast före första 1:an, dvs $Y = X - 1$, vilket ger $E(Y) = E(X - 1) = 6 - 1 = 5$. Enligt uppgiften, så ska vi välja $a = E(Y)$, vilket ger $a = 5$.

Svar: Insatsen ska väljas till $a = 5$.

Uppgift 15

Typsituation för stickprov i par. Bilda skillnaderna (med innehåll − utan innehåll). Nya data blir 96, 101, 96, 97, 99, som är utfall av $N(\mu_k + \Delta - \mu_k, \sqrt{\sigma_1^2 + \sigma_2^2}) = N(\Delta, \sigma)$, där $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.
a) Ett 95% konfidensintervall för Δ blir (se formelsamlingen avsnitt 10.1 och 11.2):

$$\bar{x} \pm t_{0.025}(5-1) \cdot \frac{s}{\sqrt{5}}$$

där $\bar{x} = 97.8$ och

$$s^2 = \frac{1}{5-1}(96^2 + 101^2 + 96^2 + 97^2 + 99^2 - 5 \cdot 97.8^2) = 4.7.$$

Konfidensintervallet blir

$$97.8 \pm 2.78 \cdot \frac{\sqrt{4.7}}{\sqrt{5}} = 97.8 \pm 2.7 = (95.1; 100.5).$$

b) Eftersom 100 tillhör intervallet i a-uppgiften kan vi *inte* förkasta hypotesen $H_0 : \Delta = 100$ på signifikansnivån 5%.

Uppgift 16

Låt U_i = antalet av genen som älg nr i har, $i = 1, 2, \cdots, 1000$. $U_1, U_2, \cdots, U_{1000}$ är oberoende och är alla Bin$(2, p)$-fördelade, dvs
$P(U_i = 0) = (1-p)^2$, $P(U_i = 1) = 2p(1-p)$, $P(U_i = 2) = p^2$.
Vi observerar att x_2 st (=11) U_i:n är 2, x_1 st (=187) U_i:n är 1 och att $1000 - x_1 - x_2$ st (=802) st U_i:n är 0. Likelihoodfunktionen blir då

$$L(p) = \prod_{i=1}^{1000} P(U_i = u_i) = (p^2)^{x_2} (2p(1-p))^{x_1} \left((1-p)^2\right)^{1000-x_1-x_2} =$$

$$= p^{2x_2+x_1}(1-p)^{x_1+2000-2x_2-2x_1}2^{x_1} = p^{x_1+2x_2}(1-p)^{2000-2x_2-x_1}2^{x_1}$$

som ger

$$\ln L(p) = (x_1 + 2x_2)\ln p + (2000 - 2x_2 - x_1)\ln(1-p) + x_1 \ln 2.$$

Vi får

$$\frac{d \ln L(p)}{dp} = \frac{x_1 + 2x_2}{p} - \frac{2000 - 2x_2 - x_1}{1-p}$$

$\frac{d \ln L(p)}{dp} = 0$ ger

$$p^* = \frac{x_1 + 2x_2}{2000} = \frac{187 + 2 \cdot 11}{2000} = \underline{0.1045}.$$

b) x_1 är ett utfall av X_1 som är $\text{Bin}(1000, 2p(1-p))$ och vidare är x_2 ett utfall av X_2 som är $\text{Bin}(1000, p^2)$. Vi får

$$E(p^*) = E\left(\frac{X_1 + 2X_2}{2000}\right) = \frac{1}{2000}\left(E(X_1) + 2E(X_2)\right) = \frac{1}{2000}\left(1000 \cdot 2p(1-p) + 2 \cdot 1000p^2\right) = p$$

dvs p^* <u>är</u> en väntevärdesriktig skattning av p.

c) Vi noterar att $x_1 + 2x_2$ är antalet ”genpositioner” som är besatta av genen. Eftersom U_i var $\text{Bin}(2, p)$ och de är oberoende, är det ”oberoende lottningar” vid varje genposition. Alltså gäller att $X_1 + 2X_2$ är $\text{Bin}(2000, p)$.
Eftersom vidare $np(1-p) \approx 2000p^*(1-p^*)$=$2000 \cdot 0.1045(1 - 0.1045)$=$187.1595 \geq 10$ är normalapproximation tillåten, dvs $X_1 + 2X_2$ är approximativt $N\left(2000p, \sqrt{2000p(1-p)}\right)$ och p^* är alltså approximativt $N\left(p, \sqrt{\frac{p(1-p)}{2000}}\right)$. Ett approximativt 95%-igt konfidensintervall för p blir alltså med approximativa metoden (FS 12.3)

$$p^* \pm \lambda_{0.025}\sqrt{\frac{p^*(1-p^*)}{2000}} = 0.1045 \pm 1.9600\sqrt{\frac{0.1045(1-0.1045)}{2000}} = \underline{0.1045 \pm 0.0134}.$$