KTH Matematik

Avd. Matematisk statistik

TENTAMEN I SF1917/SF1919 TILLÄMPAD STATISTIK,
TISDAG 11 JANUARI 2022 KL 08.00–13.00

Examinator: Mykola Shykula, 08-790 6644.

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 april 2022. 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 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 april 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

Olivia har pollenallergi och blir täppt i näsan med sannolikhet 75% givet att det finns höga halter av pollen i luften. I vanliga fall (dvs. givet att det inte är höga halter av pollen i luften) har hon bara 12.5% sannolikhet att vara täppt i näsan. En morgon visar pollenprognosen att sannolikheten för höga pollenhalter är 20% i staden där hon bor, men hon märker ändå att hon blir täppt i näsan. Vad är sannolikheten att det är höga pollenhalter i luften givet att Olivia har täppt näsa?

A: 0.15

B: 0.40

C: 0.60

D: 0.75

Uppgift 2

Den stokastiska variabeln (X, Y) har täthetsfunktionen

$$f_{X,Y}(x, y) = \begin{cases} c\,x^{-1}e^{-x^2 y}, & \text{om} \;\; x \geq 2,\; y \geq 0, \\ 0, & \text{annars.} \end{cases}$$

Bestäm konstanten c.

A: $c = 8$

B: $c = 4$

C: $c = 2$

D: $c = 1$

Uppgift 3

De diskreta stokastiska variablerna X och Y är oberoende och har sannolikhetsfunktionerna

$$p_X(k) = \frac{2^k}{k!}\,e^{-2} \quad \text{för} \;\; k = 0, 1, 2, 3, \ldots$$

respektive

$$p_Y(j) = \frac{3^j}{j!}\,e^{-3} \quad \text{för} \;\; j = 0, 1, 2, 3, \ldots.$$

Beräkna väntevärdet $E(X^2 - XY)$.

A: 2

B: 0

C: -2

D: -4

Uppgift 4

Låt $X \in U(0, 2)$ och $Y \in U(0, 3)$ vara två oberoende stokastiska variabler och sätt $Z = \min(X, Y)$. Beräkna sannolikheten $P(Z \leq 1)$.

Uppgift 5

En kommitté med 8 medlemmar ska bildas av en grupp bestående av 8 män och 8 kvinnor. Om valet av kommittémedlemmar görs slumpmässigt, vad är sannolikheten att exakt hälften av dessa ledamöter är kvinnor?

A: 0.14

B: 0.19

C: 0.27

D: 0.38

Uppgift 6

Låt X och Y vara oberoende normalfördelade stokastiska variabler där vi har väntevärdena $E(X) = 1$ och $E(Y) = 2$. Standardavvikelserna är $D(X) = D(Y) = 0.1$.
Beräkna $P(|X + Y - 3| < 0.01)$

A: 0.03

B: 0.06

C: 0.94

D: 0.97

Uppgift 7

En vaccinmottagning har drop-in och håller öppet sex timmar på lördag samt fyra timmar på söndag. Vi kan därför anta att antalet personer som vaccineras på lördagen följer en $\mathrm{Po}(6\theta)$-fördelning och antalet personer på söndagen följer en $\mathrm{Po}(4\theta)$-fördelning. Om 43 personer vaccinerades på lördagen och 37 på söndagen, beräkna minsta-kvadrat-skattningen för θ.

Uppgift 8

Vid en kvalitetskontroll plockar man slumpmässigt ut 1000 enheter från produktionen och antar att varje enhet man undersöker har samma sannolikhet att vara felaktig. Antal felaktiga enheter som man hittar kan ses som en stokastisk variabel X. Anta att 90 av de 1000 undersökta enheterna är felaktiga och vi därmed skattar andelen felaktiga i produktionen till 0.09. Ange medelfelet för denna skattning.

A: 0.009

B: 0.018

C: 0.030

D: 0.045

Uppgift 9

Från ett normalfördelat stickprov erhålls observationerna 2.82, 4.25, 2.96, 3.73, 3.32. Ange ett 95% uppåt begränsat konfidensintervall för väntevärdet.

A: $(-\infty, 3.84]$

B: $(-\infty, 3.97]$

C: $(-\infty, 4.14]$

D: $(-\infty, 4.66]$

Uppgift 10

Låt $\bar{x} = 137.0$, $\bar{y} = 208.0$, $s_x^2 = 92.5$, $s_y^2 = 163.0$, $n_x = 5$ och $n_y = 5$ 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. Ange undre gränsen för det approximativt 99%-iga tvåsidiga konfidensintervallet $I_{\mu_x - \mu_y}$ om vi antar att $\sigma_x \neq \sigma_y$.

A: -85.0

B: -89.4

C: -95.0

D: -98.8

Uppgift 11

Antag att den stokastiska variabeln X är exponentialfördelad med intensiteten λ. Dvs $X \in \text{Exp}(\lambda)$. Vi ställer upp $H_0 : \mu = 3$, där $\mu = E(X)$. Vi vill testa nollhypotesen H_0 mot alternativet $H_1 : \mu = 6$ och förkastar H_0 till förmån för mothypotesen H_1 om observationen $x > 5$. Bestäm testets styrka.

A: 0.19

B: 0.30

C: 0.43

D: 0.55

Uppgift 12

I ett avancerat växthus utförs ett experiment för att avgöra om mängden belysning påverkar hur mycket jordgubbar växer. Belysningen mäts med hjälp av ett belysningsindex och jordgubbarnas vikt mäts i gram. De första fyra erhållna observationerna följer nedan.

Belysningsindex	5	9	13	15
Jordgubbsvikt (g)	17	29	36	45

Det är rimligt att tro att det föreligger ett linjärt samband mellan variablerna. Utifrån datamaterialet ovan skattas en linjär regressionsmodell

$$y_i = \alpha + \beta x_i + \varepsilon_i,\ i = 1, \ldots, 4,$$

där y_i = jordgubbsvikt (g) beror av x_i = belysningsindex och ε_i betecknar slumpmässiga fel. MK-skattningarna av regressionskoefficienterna α och β blev $\alpha^*_{obs} = 4.08$ respektive $\beta^*_{obs} = 2.63$. Vilket av de fyra svarsalternativen nedan motsvarar ett 90% konfidensintervall för den effekt som belysningindex har på jordgubbsvikten, dvs I_β, om man vet att effekten i fråga *är signifikant* på nivån 5% (dvs P-värdet= 0.011).

A: $[1.83, 3.43]$

B: $[0.63, 4.33]$

C: $[-0.09, 5.36]$

D: Inget av alternativen A–C är korrekt.

Var god vänd!

Del II

Uppgift 13

Under pandemin passar en biograf på att renovera sin salong. De har då en möjlighet att utöka antalet platser i salongen för att få plats med mer besökare, men vill inte lägga till fler platser än nödvändigt. För att bilda sig en uppfattning om behovet hos deras potentiella kunder skickas en enkät ut i närområdet där varje hushåll fick svara om de kunde tänka sig att gå på bio (när pandemin är över) samt hur många personer ur hushållet som i så fall skulle gå. Från denna enkät fann de att sannolikheten att ingen från hushållet skulle gå på bio var 0.6, medan sannolikheten att en skulle gå var 0.1, att två skulle gå var 0.2, att tre skulle gå var 0.05, och att fyra skulle gå var 0.05. Inom gångavstånd till biografen finns 325 hushåll och vi antar att antalet personer som kan tänka sig gå på bio är oberoende mellan hushållen.
Givet denna information, vad är det minsta antalet platser som salongen borde ha för att sannolikheten att få en fullsatt salong ska vara mindre än 0.05? Motivera eventuella antaganden.

(10 p)

Uppgift 14

Det avrundningsfel som fås då man avrundar till närmaste heltal kan betraktas som en stokastisk variabel med likformig fördelning på intervallet $(-0.5, 0.5)$. Antag att man avrundar två tal och låter Z vara det totala felet.

a) Bestäm fördelningsfunktionen för Z och beräkna $P(Z > 0.5)$. (7 p)

b) Bestäm täthetsfunktionen för Z och rita upp den. (3 p)

Uppgift 15

Man önskar jämföra ljudisoleringsförmågan hos två typer av isoleringar. Man har 70 mätvärden $x_1, \ldots, x_{70}$ från den ena typen och 40 mätvärden $y_1, \ldots, y_{40}$ från den andra typen. Följande resultat erhölls:

$$\sum_{i=1}^{70} x_i = 140, \quad \sum_{i=1}^{70} x_i^2 = 435, \qquad \sum_{i=1}^{40} y_i = 60, \quad \sum_{i=1}^{40} y_i^2 = 158.$$

Vi antar att mätvärdena är utfall av oberoende stokastiska variabler $X_1, \ldots, X_{70}$ resp. $Y_1, \ldots, Y_{40}$ som är $N(m_X, \sigma_X)$- resp. $N(m_Y, \sigma_Y)$-fördelade.

a) Bestäm ett exakt 95% konfidensintervall för $m_X - m_Y$ om man dessutom antar att σ_X och σ_Y är lika. (7 p)

b) Testa hypotesen $H_0 : m_X = m_Y + 2$ mot alternativet $H_1 : m_X \neq m_Y + 2$. (3 p)

Uppgift 16

Anta att vi räknar förekomsten av olika mätvärdespar (x, y) och erhåller följande tabell:

Antal	$y = 0$	$y = 1$	$y = 2$
$x = 0$	91	158	73
$x = 1$	183	365	161
$x = 2$	126	199	97

a) Om x och y är utfall av de stokastiska variablerna X och Y, är det rimligt att anta att X och Y är *oberoende*? Motivera ditt svar. (5 p)

b) Låt oss göra antagandet att X och Y faktiskt är oberoende och att $X \in \text{Bin}(2, p)$ medan $Y \in \text{Bin}(2, 1 - p)$. Beräkna ML-skattningen för p. (5 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1917/SF1919 TILLÄMPAD STATISTIK, TISDAG 11 JANUARI 2022 KL 8.00–13.00.

Del I, Svar:

1. C

2. A

3. B

4. 2/3

5. D

6. B

7. 7.81

8. A

9. B

10. B

11. C

12. A

Del I, Lösningsförslag:

Uppgift 1

Låt A vara händelsen att Olivia är täppt i näsan och B vara händelsen att det är höga pollenhalter ute. Vi har då, enligt problemlydelsen, att $P(A|B) = 0.75$, $P(A|B^\star) = 0.125$ och $P(B) = 0.2$. Komplementsatsen ger då $P(B^\star) = 1 - P(B) = 0.8$ och lagen om total sannolikhet säger att

$$P(A) = P(A|B)P(B) + P(A|B^\star)P(B^\star) = 0.75 \cdot 0.2 + 0.125 \cdot 0.8 = 0.25.$$

Med hjälp av Bayes sats får vi då

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{0.75 \cdot 0.2}{0.25} = 0.60.$$

Uppgift 2

Följande gäller: $\iint_{R^2} f_{X,Y}(x,y)dxdy = 1$. Vi utvecklar det uttrycket, integrerar och sedan löser ut ekvationen för c:

$$c\int_2^\infty \frac{dx}{x}\left(\int_0^\infty e^{-x^2y}dy\right) = 1,$$

$$c\int_2^\infty \frac{dx}{x}\left[-\frac{e^{-x^2y}}{x^2}\right]_{y=0}^{y=\infty} = 1,$$

$$c\int_2^\infty \frac{dx}{x^3} = 1,$$

$$c\left[-\frac{x^{-2}}{2}\right]_{x=2}^{x=\infty} = 1.$$

Alltså, får vi att $c/8 = 1$, dvs $c = 8$.

Uppgift 3

X och Y är oberoende och enligt FS §3 (dvs hur sannolikhetsfunktionerna ser ut) är de båda Poissonfördelade s.v. , dvs $X \in \mathrm{Po}(2)$ och $Y \in \mathrm{Po}(3)$. Vidare, eftersom X och Y är oberoende, gäller att $E(XY) = E(X)E(Y)$. Slutligen, eftersom $X \in \mathrm{Po}(2)$ och $Y \in \mathrm{Po}(3)$, har vi:
$E(X^2 - XY) = E(X^2) - E(XY) = V(X) + \big(E(X)\big)^2 - E(X)E(Y) = 2 + 2^2 - 2 \cdot 3 = 2 + 4 - 6 = 0.$

Uppgift 4

Först har vi

$$P(Z \le 1) = P(X \le 1 \text{ eller } Y \le 1) = 1 - P(X > 1 \text{ och } Y > 1).$$

Eftersom X och Y är oberoende stokastiska variabler har vi nu

$$P(X > 1)P(Y > 1) = (1 - P(X \le 1))(1 - P(Y \le 1)) = (1 - 1/2)(1 - 1/3) = 1/2 \cdot 2/3 = 1/3.$$

Vi stoppar nu in detta i det första uttrycket och får

$$P(Z \le 1) == 1 - P(X > 1 \text{ och } Y > 1) = 1 - 1/3 = 2/3.$$

Uppgift 5

Här har vi dragning utan återläggning vilket medför att vi har hypergeometrisk fördelning. Sätt X = antal kvinnor i kommittén. Då gäller att $X \in Hyp(16, 8, 0.5)$. Då blir

$$p_X(8) = \frac{\binom{8}{4}\binom{8}{4}}{\binom{16}{8}} = \frac{70 \cdot 70}{12870} = 0.38$$

Uppgift 6

Sätt $Z = X + Y - 3 \Rightarrow Z \in N(1 + 2 - 3, \sqrt{0.1^2 + 0.1^2}) = N(0, 0.1\sqrt{2})$
$P(|Z| < 0.01) = P(-0.01 < Z < 0.01) =$[Gör om till $N(0, 1)$]=

$$= P(\frac{-0.01}{0.1\sqrt{2}} < \frac{Z}{0.1\sqrt{2}} < \frac{0.01}{0.1\sqrt{2}}) = \Phi(\frac{0.01}{0.1\sqrt{2}}) - \Phi(\frac{-0.01}{0.1\sqrt{2}}) = 2\Phi(\frac{0.1}{\sqrt{2}}) - 1 \approx 2\Phi(0.07) - 1 \approx 2 \cdot 0.53 - 1 = 0.06$$

Uppgift 7

F.S §9.2 $\Rightarrow Q = \sum_{i=1}^{n}(x_i - E(X_i))^2 = (43 - 6\theta)^2 + (37 - 4\theta)^2$

$$\frac{dQ}{d\theta} = 2(43 - 6\theta)(-6) + 2(37 - 4\theta)(-4) = 104\theta - 812$$

$$\frac{dQ}{d\theta} = 0 \Rightarrow \theta = 7.81$$

$$\frac{d^2Q}{d\theta^2} = 104 \Rightarrow \theta = 7.81 > 0$$

Det betyder att $\theta = 7.81$ minimerar Q och därmed är MK-skattningen av $\theta = 7.81$.

Uppgift 8

$X \in Bin(1000, 0.09)$. $p^*_{obs} = \frac{x}{n} = \frac{90}{1000} = 0.09$.

$$V(p^*) = V(\frac{X}{n}) = \frac{V(X)}{n^2}$$

Eftersom X är Binomialfördelad så är $V(X) = np(1 - p)$ (Se §3 i F.S.)
Detta leder till att

$$V(p^*) = \frac{p(1 - p)}{n}$$

vilket i sin tur leder till att medelfelet blir

$$D^*_{obs}(p^*) = \sqrt{\frac{p^*_{obs}(1 - p^*_{obs})}{n}} = \sqrt{\frac{p^*_{obs}(1 - p^*_{obs})}{n}} = \sqrt{\frac{0.09(1 - 0.09)}{1000}} = 0.00854 \approx 0.09$$

Uppgift 9

Från datat får vi att

$$\bar{x} = 3.416 \quad \text{och} \quad s \approx 0.5845.$$

Det sökta konfidensintervallet är då

$$I_\mu = \left(-\infty, \bar{x} + \mathrm{t}_{0.05}(4)\frac{s}{\sqrt{5}}\right] \approx \left(-\infty, 3.416 + 2.13\frac{0.5845}{\sqrt{5}}\right] \approx (-\infty, 3.97].$$

Uppgift 10

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

$$I_{\mu_x-\mu_y} = \bar{x} - \bar{y} \pm \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \cdot \lambda_{\frac{\alpha}{2}} = 137 - 208 \pm \sqrt{\frac{92.5}{5} + \frac{163.0}{5}} \cdot \lambda_{0.005} =$$

$$= 137 - 208 \pm \sqrt{\frac{92.5}{5} + \frac{163.0}{5}} \cdot 2.5758 = -71 \pm 18.4$$

Undre gränsen blir alltså -89.4.

Uppgift 11

Styrkan hos testet är P(förkasta H_0) om H_1 är sann. Dvs vi ska räkna ut $PX > 5)$ om $\mu = 6$ $X \in Exp(\lambda)$ så vi ska räkna ut

$$\int_5^\infty f_X(x)dx = \int_5^\infty \lambda e^{-\lambda x} dx$$

om $\mu = 6$ dvs om $\lambda = 1/6$.

$$\int_5^\infty \frac{1}{6} e^{-\frac{1}{6}x} dx = [-e^{-\frac{1}{6}x}]_5^\infty = e^{-\frac{5}{6}} = 0.43$$

Uppgift 12

1) Konfidensintervallet I_β måste ha $\beta^*_{obs} = 2.63$ som mittpunkt, eftersom intervallet är symmetriskt runt β^*_{obs}.
2) Konfidensintervallet I_β ska *inte* innehålla noll, eftersom effekten som belysningsindex har på jordgubbsvikten är signifikant.
Det finns bara ett intervall (nämligen, i svarsalternativ A) som uppfyller de två kraven. Enkla beräkningar enligt FS §13.2 (där $s = 2.1113$ från FS §13.1d, och $t_{p/2}(n-2) = t_{0.05}(2) = 2.92$) bekräftar att approximativt $I_\beta = [1.83, 3.43]$ är det 90% k.i. för den effekt som belysningindex har på jordgubbsvikten.

Del II

Uppgift 13

Låt X_i vara antalet personer som kommer gå på bio från det i:te hushållet. I så fall har vi

$$P(X_i = x) = \begin{cases} 0.6, & x = 0, \\ 0.1, & x = 1 \\ 0.2, & x = 2, \\ 0.05, & x = 3, \\ 0.05, & x = 4. \end{cases}$$

Väntevärdet och variansen för X_i fås till

$$E(X_i) = 0.85 \quad \text{och} \quad V(X_i) = 1.4275.$$

Om vi tar $Y = \sum_{i=1}^{325} X_i$ har vi

$$E(Y) = 325E(X_i) = 276.25 \quad \text{och} \quad V(Y) = 325V(X_i) = 463.9375,$$

eftersom antalet personer från varje hushåll är oberoende av varandra. Antalet personer från varje hushåll är likafördelade och oberoende samt antalet hushåll är stort, så vi kan tillämpa centrala gränsvärdessatsen, vilket ger att Y är approximativt normalfördelat. Med detta får vi att det y som uppfyller

$$P(Y > y) = 0.05$$

är approximativt $E(Y) + \lambda_{0.05}D(Y) = 276.25 + 1.6449 \cdot \sqrt{463.9375} \approx 311.68$. Vi söker det minsta heltalet k som uppfyller $P(Y > k) < 0.05$. Sannolikheten $P(Y > k)$ avtar med ökande y, vilket ger att $P(Y > 311) > 0.05$ medan $P(Y > 312) < 0.05$. Svaret är alltså 312.

Uppgift 14

Svar: 14a) Fördelningsfunktionen $F_Z(z)$:

$$F_Z(z) = \begin{cases} 0, & z \leq -1, \\ \frac{(1+z)^2}{2}, & -1 < z \leq 0, \\ 1 - \frac{(1-z)^2}{2}, & 0 < z \leq 1, \\ 1, & z > 1. \end{cases}$$

Svar: 14b) Täthetsfunktionen $f_Z(z) = F'_Z(z)$:

$$f_Z(z) = \begin{cases} 1 + z, & -1 \leq z \leq 0, \\ 1 - z, & 0 < z \leq 1, \\ 0, & |z| > 1. \end{cases}$$

Lösningsförslag: (kommer)
a) Låt $X \in U(-0.5, 0.5)$ och $Y \in U(-0.5, 0.5)$ vara två avrundningsfel när man avrundar två tal,

X och Y är oberoende. Det totala felet $Z = X + Y$.

Täthetsfunktionerna för X och Y är $f_X(x) = 1_{\{-0.5 \le x \le 0.5\}}(x)$ resp $f_Y(y) = 1_{\{-0.5 \le y \le 0.5\}}(y)$.

För $-1 \le z \le 0$, har vi enligt geometri (rita område $\{(x,y) \in [-0.5, 0.5]^2 : x + y \le z\}$, dvs för någon z sådan att $-1 \le z \le 0$):

$$F_Z(z) = P(Z \le z) = P(X + Y \le z) = \iint_{\{(x,y)\in[-0.5,0.5]^2 : x+y \le z\}} f_{X,Y}(x,y)dxdy =$$

$$= \iint_{\{(x,y)\in[-0.5,0.5]^2 : x+y \le z\}} f_X(x) f_Y(y) dxdy = \int_{-0.5}^{z+0.5} dx \left(\int_{-0.5}^{z-x} dy \right) = \int_{-0.5}^{z+0.5} (z - x + 0.5) dx =$$

$$= \left[-\frac{(z-x+0.5)^2}{2} \right]_{x=-0.5}^{x=z+0.5} = \frac{(1+z)^2}{2}.$$

På samma sätt, för $0 \le z \le 1$, har vi enligt geometri:

$$F_Z(z) = \iint_{\{(x,y)\in[-0.5,0.5]^2 : x+y \le z\}} f_X(x) f_Y(y) dxdy = 1 - \int_{z-0.5}^{0.5} dx \left(\int_{z-x}^{0.5} dy \right) =$$

$$= 1 - \frac{(1-z)^2}{2}.$$

Sammanfattningsvis, fördelningsfunktionen $F_Z(z)$ för Z:

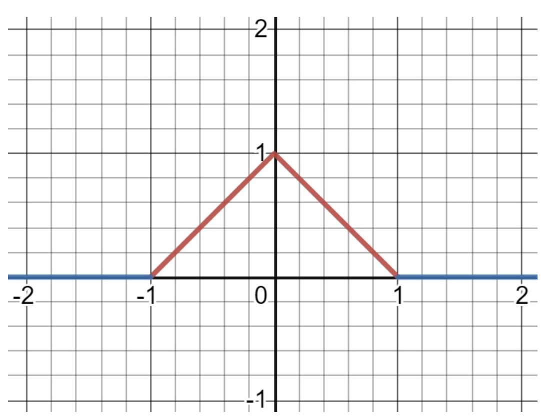
$$F_Z(z) = \begin{cases} 0, & z \le -1, \\ \frac{(1+z)^2}{2}, & -1 < z \le 0, \\ 1 - \frac{(1-z)^2}{2}, & 0 < z \le 1, \\ 1, & z > 1. \end{cases}$$

Och, vi har därför: $P(Z > 0.5) = 1 - P(Z \le 0.5) = 1 - F_Z(0.5) = (1 - 0.5)^2/2 = 0.125$.

b) Genom att derivera $F_Z(z)$ i a) får man täthetsfunktionen $f_Z(z)$ för Z, dvs $f_Z(z) = F_Z'(z)$:

$$f_Z(z) = \begin{cases} 1 + z, & -1 \le z \le 0, \\ 1 - z, & 0 < z \le 1, \\ 0, & |z| > 1. \end{cases}$$

Täthetsfunktionen $f_Z(z)$ finns uppritad nedanför.



Uppgift 15

a) Vi utnyttjar att

$$\frac{\bar{X} - \bar{Y} - (m_X - m_Y)}{s\sqrt{\frac{1}{70} + \frac{1}{40}}} \quad \text{är } t(70 + 40 - 2)\text{-fördelad}$$

Vi har

$$\bar{x} - \bar{y} = \frac{140}{70} - \frac{60}{40} = 0.5,$$

$$s^2 = \frac{69 \cdot s_X^2 + 39 \cdot s_Y^2}{69 + 39} = \frac{1}{108}\left(\sum_{i=1}^{70} x_i^2 - \frac{1}{70}\left(\sum_{i=1}^{70} x_i\right)^2 + \sum_{i=1}^{40} y_i^2 - \frac{1}{40}\left(\sum_{i=1}^{40} y_i\right)^2\right) = \frac{223}{108}$$

och $t_{0.025}(108) \approx 1.98$ Detta ger

$$I_{m_X - m_Y} = \bar{x} - \bar{y} \pm t_{0.025}(108) s\sqrt{\frac{1}{70} + \frac{1}{40}} = \underline{0.5 \pm 0.56}.$$

b) Eftersom $I_{m_X - m_Y}$ ej omsluter 2, så <u>förkastas H_0</u>.

Uppgift 16

a) För att bestämma om X och Y är obereonde använder vi ett χ^2-test. Vi har då att $n = 1453$ och

$$p_{0\cdot}^\star \approx 0.2216 \quad p_{1\cdot}^\star \approx 0.4880 \quad p_{2\cdot}^\star \approx 0.2904$$

samt

$$p_{\cdot 0}^\star \approx 0.2753 \quad p_{\cdot 1}^\star \approx 0.4969 \quad p_{\cdot 2}^\star \approx 0.2278.$$

Detta ger då testvariabeln

$$\begin{aligned} Q &= \sum_{i=0}^{2}\sum_{j=0}^{2} \frac{(x_{ij} - np_{i\cdot}^\star p_{\cdot j}^\star)^2}{np_{i\cdot}^\star p_{\cdot j}^\star} \\ &\approx \frac{(91 - 88.64)^2}{88.64} + \frac{(158 - 160.0)^2}{160.0} + \frac{(73 - 73.35)^2}{73.35} \\ &\quad + \frac{(183 - 195.2)^2}{195.2} + \frac{(365 - 352.3)^2}{352.3} + \frac{(161 - 161.5)^2}{161.5} \\ &\quad + \frac{(126 - 116.2)^2}{116.2} + \frac{(199 - 209.7)^2}{209.7} + \frac{(97 - 96.12)^2}{96.12} \\ &\approx 2.692. \end{aligned}$$

Antalet frihetsgrader är $(3-1)(3-1) = 2 \cdot 2 = 4$, vilket ger kvantilen $\chi^2_{0.05}(4) \approx 9.49$ för $\alpha = 0.05$. Vi förkastar alltså inte nollhypotesen $H_0 : X$ oberoende av Y med signifikansgrad 5% då $Q < \chi^2_{0.05}(4)$.

b) Om X och Y är oberoende har de den simultana sannolikhetsfunktionen

$$p_{X,Y}(x, y) = \binom{2}{x} p^x (1-p)^{2-x} \cdot \binom{2}{y} (1-p)^y p^{2-y} = \binom{2}{x}\binom{2}{y} p^{2+x-y} (1-p)^{2-x+y}.$$

Om vi har oberoende likafördelade stokastiska variabler (X_i, Y_i) ger detta likelihoodfunktionen

$$L(p) = \prod_{i=1}^{n} \binom{2}{x_i}\binom{2}{y_i} p^{2+x_i-y_i}(1-p)^{2-x_i+y_i}.$$

Om vi logaritmerar får vi log-likelihoodfunktionen

$$\log L(p) = (2n + \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i)\log p + (2n - \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i)\log(1-p) + C,$$

där C är en konstant som inte beror på p. Vi tar nu derivatan och sätter denna till noll. Detta ger

$$\frac{2n + \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}{p} - \frac{2n - \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i}{1-p} = 0,$$

vilket kan skrivas om som

$$p = \frac{2n + \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}{4n}.$$

Andraderivatan av log-likelihoodfunktionen (för $p \neq 0, 1$) är

$$-\frac{2n + \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}{p^2} - \frac{2n - \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i}{(1-p)^2}.$$

Denna är negativ för alla värden på $0 < p < 1$, så vi har ett globalt maximum givet att $p \neq 0, 1$. Med våra värden får vi

$$n = 1453, \quad \sum_{i=1}^{n} x_i = 1553, \quad \sum_{i=1}^{n} y_i = 1384,$$

vilket ger $p \approx 0.529$. Eftersom detta värde skiljer sig från noll och ett har vi ett global maximum. Med andra ord är $p^\star_{\text{obs}} \approx 0.529$.