TMA4245 Statistikk
Eksamen august 2015

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

Vi tenker oss at total endring i global gjennomsnittstemperatur de neste 50 år kan skrives som en sum $W = X + Y$ hvor $X \sim N(0, \sigma_X^2)$ er endring som følge av naturlige klimasvingninger og $Y \sim N(\mu, \sigma_Y^2)$ er endring som følge av menneskeskapt CO_2-utslipp. Endringen som følge av naturlige svingninger har dermed varians σ_X^2 men er i forventning lik 0. Parameterene μ og σ_Y^2 representerer forventning og usikkerhet omkring effekt av menneskeskapt påvirkning basert på klimamodeller. Vi antar at X og Y er uavhengige stokastiske variabler. Forutsatt at CO_2-utslippene holdes på dagens nivå er det gitt at $\mu = 2°\text{C}$ og at $\sigma_Y = 1.5°\text{C}$. Vi antar videre at $\sigma_X = 0.5°\text{C}$

a) Finn sannsynligheten for at $X > 1°\text{C}$ og sannsynligheten for at total endring i gjennomsnittstemperatur er større enn $5°\text{C}$.

Vi tenker oss videre at samfunnskostnadene knyttet til ødeleggelser som følge av klimaendringer vil være minst dersom global gjennomsnittstemperatur forblir uendret ($W = 0$) og at både en økning og en reduksjon i global gjennomsnittstemperatur vil være kostbart. For å modellere dette tilnærmer vi de totale samfunnskostnadene som påløper årlig som følge av en total temperaturendring W som et 2. ordens Taylor-polynom $g(W) = aW^2$. Anta at $a = 1.1 \cdot 10^{12}$ USD/$°C^2$.

b) Finn forventningsverdien til de årlige samfunnskostnadene $g(W)$ som følge av total endring i gjennomsnittstemperatur W uttrykt ved a, μ, σ_X og σ_Y.

Anta at vi ved å halvere CO_2-utslippene fra dagens nivå greier å halvere μ og σ_Y. Hvor mye reduseres forventede samfunnskostnader da som følge av halverte utslipp? Avhenger svaret av størrelsen på de naturlige klimasvingningene?

Oppgave 2

Ved verdensmesterskap på enkeltdistanser på skøyter går hver deltager to 500 meters løp, ett løp hvor deltageren har indre bane i siste sving og ett løp hvor deltageren har ytre bane i siste sving. Rekkefølgen av deltagerne baserer seg på summen av tidene på de to løpene. Tilsvarende regel benyttes også i olympiske leker. Denne regelen ble innført fra og med verdensmesterskapet på Hamar i 1995. Tidligere ble rekkefølgen av deltagerne basert på kun et løp for hver deltager. Bakgrunnen for regelen om at hver deltager skal gå to løp er at det kan være en fordel å ha siste ytre siden løperne har stor fart i siste sving og i indre bane er krumningen større enn i ytre bane.

I et mesterskap med n deltagere, la Y_i og Z_i betegne løpstidene for løper nummer i for løpene med henholdsvis siste ytre og siste indre. La videre X være antall av de n løperne som har sin raskeste løpstid i løpet med siste ytre, dvs. X er antall løpere som har $Y_i < Z_i$. Vi antar at X er binomisk fordelt, dvs. $P(X = x) = b(x; n, p)$ der $p = P(Y_i < Z_i)$.

a) Angi hvilke forutsetninger som må være oppfylt i situasjonen beskrevet over for at antagelsen om at X er binomisk fordelt skal være korrekt.

Dersom $n = 20$ og $p = 0.7$, finn sannsynlighetene

$$P(X \leq 10) \quad \text{og} \quad P(X \geq 8 | X \leq 10).$$

b) Skriv opp rimelighetsfunksjonen (likelihoodfunksjonen) for p og benytt denne til å vise at sannsynlighetsmaksimeringsestimatoren for p blir

$$\widehat{p} = \frac{X}{n}.$$

Vis at $\widehat{p}$ er en forventningsrett estimator for p og at $\text{Var}(\widehat{p}) = p(1-p)/n$.

Videre i denne oppgaven skal vi benytte resultater fra 500 meter for menn i olympiske leker i Sochi i Russland i februar 2014 til å vurdere om det er grunnlag for å hevde at det er en fordel å gå siste ytre. Her var det $n = 39$ løpere som fullførte begge løpene, og av disse var det $x = 24$ som hadde sin raskeste løpstid i løpet med siste ytre.

I de videre utregningene kan du om nødvendig gjøre approksimasjoner, men du må i så fall begrunne disse.

c) Formuler en hypotesetest for situasjonen. Spesifiser H_0 og H_1, velg en passende testobservator og utled en beslutningsregel når signifikansnivået er $\alpha = 0.05$.

Hva blir konklusjonen av testen for resultatene fra Sochi.

Regn også ut testens p-verdi basert på resultatene fra Sochi.

Vi skal i resten av denne oppgaven betegne testen du formulerte over som Test 1. Man kan formulere en alternativ test for samme situasjon, som vi skal kalle Test 2, ved å definere differansene

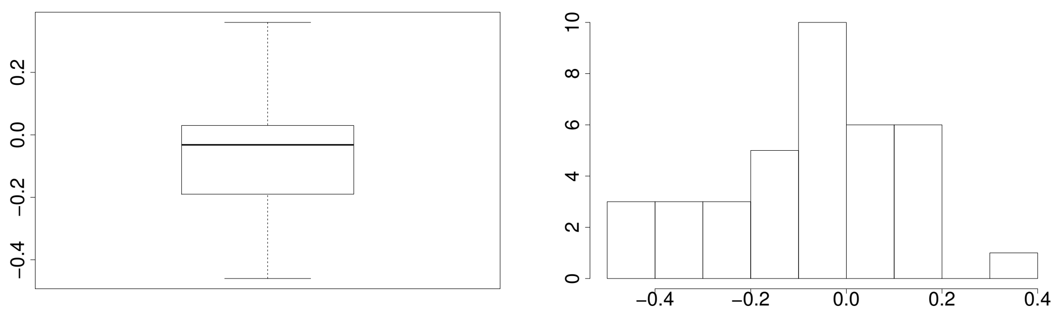
$$D_i = Y_i - Z_i \quad \text{for } i = 1, 2, \ldots, n,$$

der Y_i og Z_i altså er løpstiden for løper nummer i i løpene med henholdsvis siste ytre og siste indre. Man kan da lage en test ved å ta utgangspunkt i $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$.

d) Formuler en hypotesetest for situasjonen med utgangspunkt i $\bar{D}$. Spesifiser H_0 og H_1, velg en passende testobservator og utled en beslutningsregel når signifikansnivået er α.

Av de 39 løperne i Sochi var det to som falt i et av sine løp. Dersom vi ser bort fra disse to løperne gir resultatene i Sochi $n = 37$, $\sum_{i=1}^{37} d_i = -2.654$ og $\sum_{i=1}^{37} d_i^2 = 1.552$. Hva blir da konklusjonen av Test 2 med resultatene fra Sochi når $\alpha = 0.05$? Avrund om nødvendig til nærmeste frihetsgrad oppgitt i tabell.

Det oppgis at teststyrken (sannsynligheten for å forkaste H_0 når H_1 er riktig) for Test 2 når $n = 37$, $E(D_i) = -0.07$ og $\text{Var}(D_i) = 0.2^2$ er lik 0.67. Disse verdiene for $E(D_i)$ og $\text{Var}(D_i)$ er cirka hva man får når man estimerer basert på resultatene av de 37 løperne i Sochi som ikke falt. Det kan også nevnes at med disse verdiene for $E(D_i)$ og $\text{Var}(D_i)$ blir $p = P(Y_i < Z_i) = 0.64$ hvis man antar at D_i er normalfordelt.



Figur 1: Boxplott og histogram over differansene d_i for de 37 løperne i 500 meter for menn i Sochi som ikke falt. I boxplottet er differansene langs y-aksen, mens i histogrammet er differansene langs x-aksen.

e) Finn teststyrken for Test 1 når $n = 39$ og $p = 0.64$.

Figur 1 viser boxplot og histogram over d_i for de 37 løperne i Sochi som ikke falt. Diskuter basert på konklusjonene du fant for Test 1 og Test 2, de utregnede teststyrkene for disse testene, samt plottene i Figur 1, hva du totalt sett ville konkludert med i den aktuelle situasjonen. De to løperne i Sochi som falt, falt begge i løpet hvor de hadde siste indre, har dette noen betydning for din konklusjon?

Oppgave 3

For å vurdere nøyaktigheten av en ny måleprosedyre gjør man n målinger av samme størrelse. La $X_1, X_2, \ldots, X_n$ betegne resultatene av disse målingene, og anta at disse er et tilfeldig utvalg fra en normalfordeling med forventning μ og standardavvik σ. Vi er her interessert i verdien til σ, men vi skal anta at verdien til μ også er ukjent.

La $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Fra pensum er det da kjent at $S^2(n-1)/\sigma^2$ er kji-kvadratfordelt med $n-1$ frihetsgrader.

a) Utled et $(1-\alpha)100\,\%$ konfidensintervall for σ^2.

Utled også et $(1-\alpha)100\,\%$ konfidensintervall for σ.

b) Vis at dersom en stokastisk variabel Y er kji-kvadratfordelt med v frihetsgrader, dvs. har sannsynlighetstetthet

$$f(y) = \frac{1}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})} y^{\frac{v}{2}-1} e^{-\frac{y}{2}},$$

så er

$$E(\sqrt{Y}) = \frac{\sqrt{2}\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})}.$$

c) Bruk resultatet i forrige punkt til å undersøke om

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

er forventningsrett for σ. Foreslå eventuelt en estimator for σ hvor forventningsfeilen er korrigert.

En estimator $\hat{\theta}$ som over- og underestimerer θ med like stor sannsynlighet er såkalt medianrett. For kontinuerlig fordelte medianrette estimatorer er derfor $P(\hat{\theta} \leq \theta) = 1/2$. Foreslå en medianrett estimator for σ.

Fasit

1. **a**) 0.023,0.029 **b**) $5.11 \cdot 10^{12}$

2. **a**) 0.048, 0.98 **c**) Ikke forkast H_0, 0.075 **d**) Forkast H_0 **e**) 0.545