TMA4245 Statistikk
Eksamen august 2015

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

Vi tenkjer oss at total endring i global gjennomsnittstemperatur dei neste 50 åra kan skrivast som ein sum $W = X + Y$ hvor $X \sim N(0, \sigma_X^2)$ er endringa som følgjer av naturlege klimasvingingar og $Y \sim N(\mu, \sigma_Y^2)$ er endringa som følgjer av menneskeskapt CO_2-utslepp. Endringa som følgjer av naturlege svingingar har dermed varians σ_X^2 men er i forventing lik 0. Parameterane μ og σ_Y^2 representerer forventing og usikkerheit omkring effekten av menneskeskapt påverknad basert på klimamodellar. Vi går ut i frå at X og Y er uavhengige stokastiske variablar. Basert på at CO_2-utsleppa er på same nivå som i dag er det gitt at $\mu = 2°\text{C}$ og at $\sigma_Y = 1.5°\text{C}$. Vi går vidare ut i frå at $\sigma_X = 0.5°\text{C}$.

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

Vi tenkjer oss vidare at samfunnskostnadane knytta til øydeleggingar som følgje av klimaendringar vil vere minst dersom global gjennomsnittstemperatur er uforandra i framtida ($W = 0$) og at både ei auking og ein reduksjon i global gjennomsnittstemperatur vil vere kostbart. For å modellere dette tilnærmar vi dei totale årlege samfunnskostnadane som følgje av ei total temperaturendring W som eit 2. ordens Taylor-polynom $g(W) = aW^2$. Gå ut i frå at $a = 1.1 \cdot 10^{12}$ USD/$°C^2$.

b) Finn forventingsverdien til dei årlege samfunnskostnadane $g(W)$ som følgje av total endring i gjennomsnittstemperatur W uttrykt ved a, μ, σ_X og σ_Y.

Gå ut i frå at vi ved å halvere CO_2-utsleppa frå det nivået dei har i dag greier å halvere μ og σ_Y. Kor mykje blir då forventa samfunnskostnader redusert med som følgje av halverte utslepp? Avheng svaret av storleiken på dei naturlege klimasvingingane?

Oppgave 2

Ved verdsmeisterskap på enkeltdistansar på skøyter går kvar deltakar to 500 meterar, ein gong der deltakaren har indre bane i siste sving og ein gong der deltakaren har ytre bane i siste sving. Rekkjefølgja av deltakarane baserer seg på summen av tidene for kvar gong. Tilsvarande regel blir og brukt i olympiske leikar. Denne regelen blei innført frå og med verdsmeisterskapet på Hamar i 1995. Tidligare blei rekkjefølgja av deltakarne basert på kun ein 500 meter for kvar deltakar. Bakgrunnen for regelen om at kvar deltakar skal gå to gonger er at det kan vere ein fordel å ha siste ytre sidan deltakarane har stor fart i siste sving og i indre bane er krumminga større enn i ytre bane.

I eit meisterskap med n deltakarar, la Y_i og Z_i vere tidene som deltakar nummer i brukar på dei to 500 meterane med siste ytre og siste indre i gitt rekkjefølgje. La vidare X vere talet på deltakarar av dei n som har raskast tid på 500 meteren med siste ytre, dvs. X er talet på deltakarar som har $Y_i < Z_i$. Vi går ut i frå at X er binomisk fordelt, dvs. $P(X = x) = b(x; n, p)$ der $p = P(Y_i < Z_i)$.

a) Skriv opp dei føresetnadane som må vere oppfylte i situasjonen gitt ovanfor for at det skal vere korrekt at X er binomisk fordelt.

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

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

b) Skriv opp rimelegfunksjonen (likelihoodfunksjonen) for p og bruk denne til å vise at sannsynsmaksimeringsestimatoren for p blir

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

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

Vidare i denne oppgåva skal vi bruke resultata frå 500 meteren for menn i olympiske leikar i Sochi i Russland i februar 2014 til å vurdere om det er grunnlag for å hevde at det er ein fordel å gå siste ytre. Her var det $n = 39$ deltakarar som fullførte begge 500 meterane, og av desse var det $x = 24$ som hadde raskaste tida si på 500 meteren med siste ytre.

I dei vidare utrekningane kan du om nødvendig gjere approksimasjonar, men du må i så fall grunngje desse.

c) Formuler ein hypotesetest for situasjonen. Spesifiser H_0 og H_1, velg ein høveleg testobservator og utlei ein regel som kan brukast til å trekkje slutningar når signifikansnivået er $\alpha = 0.05$.

Kva blir konklusjonen av testen for resultata frå Sochi.

Rekn og ut p-verdien for testen basert på resultata frå Sochi.

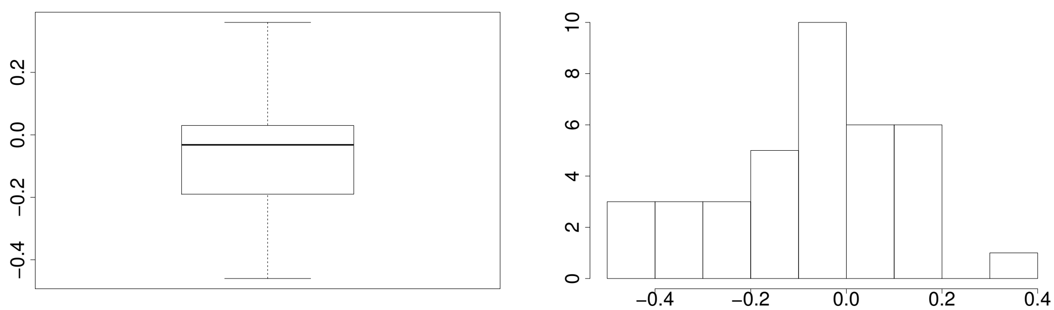
Vi skal i resten av denne oppgåva kalle testen du formulerte over som Test 1. Ein kan formulere ein alternativ test for same situasjon, som vi skal kalle Test 2, ved å definere differansane

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

der Y_i og Z_i som før er tida for deltakar nummer i i 500 meterane med siste ytre og siste indre i gitt rekkjefølgje. Ein kan då lage ein test ved å ta utgangspunkt i $\bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i$.

d) Formuler ein hypotesetest for situasjonen med utgangspunkt i $\bar{D}$. Spesifiser H_0 og H_1, velg ein passande testobservator og utlei en regel som kan brukast til å trekkje slutningar når signifikansnivået er α.

Av dei 39 deltakarane i Sochi var det to som datt i ein av 500 meterane. Dersom vi ser bort fra desse to gjev resultatene i Sochi $n = 37$, $\sum_{i=1}^{37} d_i = -2.654$ og $\sum_{i=1}^{37} d_i^2 = 1.552$. Kva blir då konklusjonen av Test 2 med resultata frå Sochi når $\alpha = 0.05$? Avrund om nødvendig til næraste fridomsgrad oppgjeve i tabellen.



Figur 1: Boxplott og histogram over differansene d_i for dei 37 deltakarene i 500 meter for menn i Sochi som ikkje datt. I boxplottet er differansene langs y-aksen, medan i histogrammet er differansene langs x-aksen.

Du får oppgjeve at teststyrken (sannsynet for å forkaste H_0 når H_1 er rett) for Test 2 når $n = 37$, $E(D_i) = -0.07$ og $\text{Var}(D_i) = 0.2^2$ er lik 0.67. Desse verdiane for $E(D_i)$ og $\text{Var}(D_i)$ er omlag kva ein får når ein estimerer basert på resultata av dei 37 deltakarane i Sochi som ikkje datt. Det kan også nemnast at med desse verdiane for $E(D_i)$ og $\text{Var}(D_i)$ blir $p = P(Y_i < Z_i) = 0.64$ dersom ein går ut i frå at D_i er normalfordelt.

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

Figur 1 viser boxplot og histogram over d_i for dei 37 deltakarane i Sochi som ikkje datt. Diskuter basert på konklusjonene du fann for Test 1 og Test 2, dei utrekna teststyrkane for desse testane, samt plotta i Figur 1, kva du totalt sett ville konkludert med i den aktuelle situasjonen. Dei to deltakarane i Sochi som datt, datt begge i den 500 meteren der dei hadde siste indre. Har dette noko å seie for konklusjonen din?

Oppgave 3

For å vurdere kor nøyaktig ei ny måleprosedyre er, gjer ein n målingar av same storleik. La $X_1, X_2, \ldots, X_n$ vere resultata av desse målingane, og gå ut i frå at desse er eit tilfeldig utval frå ei normalfordeling med forventing μ og standardavvik σ. Vi er her interesserte i verdien til σ, men vi skal gå ut i frå at verdien til μ og er ukjend.

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

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

Utlei og eit $(1-\alpha)100\,\%$ konfidensintervall for σ.

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

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

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 forventingsrett for σ. Føreslå eventuelt ein estimator for σ der forventingsfeilen er korrigert.

Ein estimator $\hat{\theta}$ som over- og underestimerer θ med like stort sannsyn er sagt å vere medianrett. For kontinuerleg fordelte medianrette estimatorar er derfor $P(\hat{\theta} \leq \theta) = 1/2$. Føreslå ein 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