



OPPGAVER I AUDITORIET ... 23/4

EMNE TMA4245 STATISTIKK
VÅREN 2008

Oppgave 1

$X_1, \dots, X_n$ er uavhengige og identisk fordelte TV med ST

$$f_X(x) = e^{-(x-\theta)}, \quad x \geq \theta > 0, \\ = 0, \text{ ellers} \quad (1)$$

- a) Vis at for $Y = \min\{X_1, \dots, X_n\}$, er $2n(Y - \theta)$ χ^2 -fordelt med 2 frihetsgrader.
- b) Vis at $V = 2 \sum_{j=1}^n (X_j - \theta)$ er χ^2 -fordelt med $2n$ frihetsgrader.
- c) Hypotesen $H_0 : \theta = 0$ testes mot den alternative hypotesen $H_1 : \theta > 0$. Følgende to tester er foreslått:

T_1 : Forkast H_0 når $Y \geq k_1$.

T_2 : Forkast H_0 når $\bar{X} \geq k_2$.

Signifikansnivå $\alpha = 0.05$ gjelder i begge testene.

1. Anta $n = 10$, og bestem k_1 og k_2 .
2. Hvor stor må n være for dersom T_1 skal ha teststyrke minst 0.95 i alternativet $\theta = 1.0$?
3. Hvor stor må n være for dersom T_2 skal ha teststyrke minst 0.95 i alternativet $\theta = 1.0$?
4. Hva synes resultatene i 2) og 3) å tyde på?

Oppgave 2

En fabrikk produserer en viss type maskinkomponenter. Tiden fra en komponent blir tatt i bruk til den bryter sammen for første gang, kaller vi levetiden for komponenten. Erfaring har vist at levetiden T , målt i uker, kan modelleres som en stokastisk variabel med ST:

$$f_T(t) = 2\lambda t e^{-\lambda t^2}, \quad t \geq 0, \\ f_T(t) = 0, \text{ ellers,} \quad (2)$$

der $\lambda > 0$ er en ukjent parameter.

- a) Bestem kumulativ fordelingsfunksjon (KF) $F_T(t)$ til T , og beregn $P(20 < T \leq 30)$ når $\lambda = 1.5 \cdot 10^{-3}$.
- b) Vis at $Z = 2\lambda T^2$ er χ^2 -fordelt med 2 frihetsgrader.
- c) Parameteren λ skal estimeres på basis av levetidene $T_1, \dots, T_n$ for $n > 2$ tilfeldig valgte komponenter. $T_1, \dots, T_n$ antas uavhengige og identisk fordelte med ST $f_T(t)$. Vis at sannsynlighetsmaksimerings-estimatoren (SME) for λ blir

$$\Lambda^* = \frac{n}{\sum_{i=1}^n T_i^2}. \quad (3)$$

- d) La X være χ^2 -fordelt med $2n$ frihetsgrader ($n > 2$). Vis at

$$E(X^{-1}) = \frac{1}{2(n-1)} \quad \text{og} \quad E(X^{-2}) = \frac{1}{4(n-1)(n-2)}. \quad (4)$$

Bruk dette til å undersøke om Λ^* er forventningsrett. Om nødvendig, korrigér for å få en estimator som er forventningsrett. Bestem denne estimatorens varians.

- e) Utled et $100(1 - \alpha)\%$ konfidensintervall for λ . Bestem intervallet numerisk når $\alpha = 0.05$, $n = 5$ og de observerte verdiene er

23.63 35.97 18.65 18.18 11.59

- f) En dag oppdages feil ved produksjonsprosessen. Komponentene testes ved at 5 komponenter velges tilfeldig fra dagsproduksjonen, og levetiden bestemmes for disse ved akselerert levetidstesting. Observasjonene brukes til å teste

$$H_0 : \lambda \leq \lambda_0 = 1.5 \cdot 10^{-3}$$

mot

$$H_1 : \lambda > \lambda_0 = 1.5 \cdot 10^{-3}$$

Foreslå en hypotesetest med signifikansnivå $\alpha = 0.05$. Konkluder når observasjonene er

12.06 18.02 19.86 16.60 9.36

Komponentene blir pakket i kasser med 5 komponenter i hver kasse. Fabrikken garanterer at alle komponentene i en kasse har levetid på minst a uker. Fabrikken får reklamasjon på en kasse hvis én eller flere av de 5 komponentene i kassen har levetid kortere enn a uker.

- g) Hvis $\lambda = 1.5 \cdot 10^{-3}$, hvor stor kan a velges for at sannsynligheten for reklamasjon skal være høyst 0.05? Du kan anta uavhengige levetider.
- h) La a og λ være som i punkt g), og anta at i et visst tidsrom selges 1000 kasser. La U være antall kasser som det blir reklamert på. Hvilke forutsetninger må en gjøre for at U skal være binomisk fordelt? Anta at disse forutsetningene er oppfylt og bestem $P(U \leq 60)$ ved å bruke tilnærming til normalfordeling.
- i) Finn fordelingen til $V = \min\{T_1, \dots, T_5\}$.
- j) Fabrikken får en kasse i retur der alle komponentene fungerer etter 2 ukers bruk. Kassen selges på nytt med samme garanti som før. Bestem sannsynligheten for reklamasjon når a og λ er som i punkt g).

Oppgave 3

Hyppigheten av større ulykker ved gruvedrift skal undersøkes.

La λ = gjennomsnittlig antall ulykker pr. timeverk. λ antas ukjent (men $\lambda > 0$).

La X_i = antall større ulykker ved gruve nr. i , $i = 1, \dots, n$. X_i antas Poisson-fordelt med parameter $a_i\lambda$, der a_i = totalt antall utførte timeverk ved gruve nr. i . $X_1, \dots, X_n$ antas å være uavhengige tilfeldige variabler.

- a) Beregn sannsynligheten p_1 for at det i gruve nr. 1 inntreffer minst 2 større ulykker når $a_1 = 1.5 \cdot 10^5$ og det antas at $\lambda = 10^{-5}$ ulykker pr. timeverk. Hva er denne sannsynligheten (p'_1) gitt at det har inntruffet minst en større ulykke?
- b) Vis at for hver $i = 1, \dots, n$, så er $\tilde{\Lambda}_i = X_i/a_i$ en forventningsrett estimator for λ . Hva blir variansen til $\tilde{\Lambda}_i$?

- c) Et mulig forslag til estimator for λ basert på data fra de n gruvene er

$$\hat{\Lambda} = \frac{1}{n} \sum_{i=1}^n \tilde{\Lambda}_i$$

Vis at $\hat{\Lambda}$ er forventningsrett, og finn variansen.

- d) Vis at

$$\hat{\hat{\Lambda}} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n a_i}$$

er en forventningsrett estimator for λ , og finn variansen.

Beregn verdiene til $\hat{\Lambda}$ og $\hat{\hat{\Lambda}}$ når $n = 5$ og de innsamlede data er

Tabell 1:

Gruve nr. i	$a_i \cdot 10^{-5}$	x_i
1	1.50	1
2	2.72	4
3	0.87	0
4	3.28	2
5	0.63	1

Hvilken av de to estimatorene ville du foretrekke i denne situasjonen? Begrunn svaret.

- e) Anta at data blir innsamlet bare fra gruver der det har vært minst en større ulykke. La Y betegne antall større ulykker ved en slik gruve, og la $a =$ antall utførte timeverk i gruva. Vis at

$$P(Y = y) = \frac{(a\lambda)^y}{y!} \frac{1}{e^{a\lambda} - 1}, \quad y = 1, 2, \dots$$

Beregn $E[Y]$.

- f) La $Y_1, \dots, Y_m$ være antall større ulykker ved m gruver som angitt i punkt e). Anta at antall utførte timeverk er a ved hver av de m gruvene. Anta videre at $Y_1, \dots, Y_m$ er uavhengige og identisk fordelte med fordeling angitt i punkt e). Utled en ligning som sannsynlighetsmaksimerings-estimatoren Λ^* for λ må tilfredsstille når den baseres på observasjoner av $Y_1, \dots, Y_m$. Dvs., anta et tilfeldig utvalg $y_1, \dots, y_m$ med observasjoner. Finn et tilnærmet punktestimat λ^* for Λ^* når $m = 4$, $a = 2 \cdot 10^5$, $y_1 = 1, y_2 = 2, y_3 = 1, y_4 = 4$.

Oppgitt: Den eneste positive løsningen z_0 av ligningen $2e^{-z} + z - 2 = 0$ er $z_0 \approx 1.59$.

Oppgave 4

I en produksjonsprosess for syntetiske fibre til bruk i tekstiler inngår 'krymping' av fibre i en trykk-koker. For å finne ut hvilken innflytelse temperaturen har på krympingsgraden, utføres krymping av fiberprøver ved 11 forskjellige temperaturer, og krympingsgraden X_j registreres (i %) for hver temperatur t_j , $j = 1, \dots, 11$. Resultatene viser seg å bli:

Fiberprøve nr. j	Temperatur i °C	Krympingsgrad X_j i %
1	120	3.37
2	123	3.68
3	126	3.98
4	129	4.35
5	132	4.70
6	135	5.05
7	138	5.70
8	141	6.02
9	144	6.38
10	147	6.74
11	150	7.23

$$\sum_{j=1}^{11} X_j = 57.20, \quad \sum_{j=1}^{11} (X_j - \bar{X})^2 = 16.95$$

der $\bar{X} = \sum_{j=1}^{11} X_j / 11$. De tilfeldige variablene $X_1, \dots, X_{11}$ antas uavhengige og normalfordelte med samme ukjente varians σ^2 og forventningsverdier:

$$E[X_j] = \alpha + \beta(t_j - \bar{t}); \quad j = 1, \dots, 11, \quad \text{der } \bar{t} = \frac{1}{11} \sum_{j=1}^{11} t_j.$$

- a) Sett opp minste kvadratsums-estimatorer for α og β . Kall dem henholdsvis A og B . Tegn opp regresjonslinjen på grunnlag av de registrerte data, .
- b) Bruk de registrerte data til å teste hypotesen

$$H_0 : \sigma^2 = \sigma_0^2 = 0.02,$$

mot

$$H_1 : \sigma^2 > \sigma_0^2,$$

på signifikansnivå $\alpha = 0.05$. For å foreta testen, trenger du en testobservator. Basér ditt valg av testobservator på følgende opplysning: Den tilfeldige variabelen

$$V = \sum_{j=1}^{11} \frac{(X_j - A - B(t_j - \bar{t}))^2}{\sigma^2},$$

er χ^2 -fordelt med 9 frihetsgrader.

- c) Hva er teststyrken når $\sigma^2 = 0.09$?

Oppgave 5

Når en partikkel beveger seg i en akselerator, vil den i løpet av en tid t tilbaklegge en strekning $z = vt + \frac{1}{2}\theta t^2$ (rettlinjet bevegelse), der v er partikkelens utgangsfart, og θ er den akselerasjon partikelen oppnår. I første omgang gjøres n observasjoner av tid og strekning med utgangsfart null. Observasjonene av tilbakelagt strekning modelleres som utfall av de tilfeldige variablene $Z_1, \dots, Z_n$, som antas å være uavhengige normalfordelte variabler med forventningsverdi $E[Z_j] = \frac{1}{2}\theta t_j^2$, og varians $\text{Var}[Z_j] = \sigma_0^2$, $j = 1, \dots, n$.

De valgte måletidene $t_1, \dots, t_n$ er faste konstanter. Standardavviket σ_0 antas kjent med verdi 0.070. θ er partikkelens ukjente akselerasjon som skal estimeres på grunnlag av målingene.

- a) Vis at sannsynlighetsmaksimerings-estimatoren (SME) Θ^* for θ basert på $Z_1, \dots, Z_n$ er gitt ved formelen,

$$\Theta^* = \frac{2 \sum_{j=1}^n t_j^2 Z_j}{\sum_{j=1}^n t_j^4}.$$

Regn ut et estimat basert på observasjonene gitt i tabellen nedenfor.

- b) Vis at Θ^* er forventningsrett, og bestem variansen.
Forklar hvorfor Θ^* er normalfordelt.
- c) Utled et uttrykk for et 90% konfidensintervall for θ basert på $Z_1, \dots, Z_n$. Angi dette numerisk ut fra observasjonene nedenfor.
- d) Tyder målingene på at $\theta > 0.5000$?

Formulér problemet som en hypotesetest. Gjennomfør testingen med observasjonene nedenfor. Bruk signifikansnivå 0.05. Skissér teststyrken som funksjon av alternativ verdi for θ .

Det utføres nå én måling av tilbakelagt strekning for partikkelen i en tid t_0 med ukjent utgangsfart v . Måleresultatet W for strekningen modelleres som en normalfordelt variabel med forventningsverdi $E[W] = vt_0 + \frac{1}{2}\theta t_0^2$ og varians σ_0^2 . Variablene $Z_1, \dots, Z_n, W$ er uavhengige.

- e) Bestem SME for v og θ basert på $Z_1, \dots, Z_n$, og W . Er de forventningsrette? Hva er deres varianser?

Observasjoner: ($n = 5$, $\sigma_0 = 0.070$)

j	1	2	3	4	5
t_j	5.0	5.0	6.0	7.0	7.0
Z_j	6.22	6.20	8.93	12.27	12.39