

Øving 6

Oppgave 1:

I mange matvarer finnes det ofte en viss mengde uønskede bakterier, og for høye konsentrasjoner av slike bakterier kan være uheldig for konsumenten. Vi skal se på forekomsten av en bestemt type uønskede bakterier i en bestemt matvare. Studier har vist at antall bakterier, Y , i en prøve på t milligram (mg) av matvaren kan antas å være Poissonfordelt med forventning λt .

En kritisk grense for bakteriantallet i matvaren antas å være 12 bakterier per mg. Dersom antall bakterier per mg er større enn 12, kan dette være uheldig for konsumenten av matvaren.

Anta i første omgang at det er kjent at forventet antall bakterier per mg matvare er 5, dvs $\lambda = 5$.

- a) Finn sannsynligheten for å finne mer enn 12 bakterier i en matvareprøve på 1 mg.
Finn sannsynligheten for å finne mer enn 6 bakterier i en matvareprøve på 0.5 mg.
Dersom man tar 10 uavhengige prøver på 1 mg, hva er sannsynligheten for at minst en av de 10 prøvene inneholder mer enn 12 bakterier?

I praksis vil λ , forventet antall bakterier per mg matvare, være ukjent. Basert på uavhengige målinger $Y_1, \dots, Y_n$ av antall bakterier i n prøver på henholdsvis $t_1, \dots, t_n$ mg ønsker vi å estimere λ .

- b) Vis at sannsynlighetsmaksimeringsestimatoren (SME) for λ er

$$\hat{\lambda} = \frac{Y_1 + \dots + Y_n}{t_1 + \dots + t_n}$$

Regn ut estimatorens forventning og varians.

For å sikre at grenseverdien 12 bakterier per mg ikke overskrides for ofte anbefaler Næringsmiddeltilsynet at forventet antall bakterier per mg matvare ikke bør overskride 6, dvs $\lambda \leq 6$.

- c) Vi ønsker for et bestemt matvareparti å undersøke om det er grunn til å påstå at Næringsmiddeltilsynets anbefaling overskrides.
Formuler dette som en hypotesetest.

Du kan i resten av dette punktet bruke uten bevis at en Poissonfordelt variabel med forventningsverdi større enn 15 er tilnærmet normalfordelt. Forklar hvorfor observatoren

$$\frac{\hat{\lambda} - 6}{\sqrt{6 / \sum_{i=1}^n t_i}}$$

er tilnærmet normalfordelt under nullhypotesen dersom $\sum_{i=1}^n t_i > 2.5$.
Utfør hypotesetesten på (tilnærmet) 5% nivå når vi fra $n = 10$ uavhengige målinger fra matvarepartiet har fått dataene gitt i tabellen under.

måling i	1	2	3	4	5	6	7	8	9	10
y_i	6	4	9	7	5	8	9	3	7	7
t_i	0.8	0.7	1.3	1.1	1.2	0.9	1.0	0.8	0.8	1.4

Det oppgis at $\sum_{i=1}^{10} y_i = 65$ og $\sum_{i=1}^{10} t_i = 10$.

Oppgave 2:

For skipsfart, fiske, offshore oljevirksomhet, andre marine konstruksjoner, etc er belastningen av bølger noe man må ta hensyn til. Spesielt er man typisk interessert i å si noe om den maksimale bølgehøyden man kan regne med å bli eksponert for i løpet av en tidsperiode (f.eks. den såkalte 100-årsbølgen).

Vi skal her se på en situasjon hvor det i Nordsjøen skal utføres en arbeidsoppgave som det tar 3 timer å gjennomføre og som krever at maksimal bølgehøyden mens arbeidet pågår ikke må overskride 4.0 meter.

For å avgjøre om det er forsvarlig å sette i gang med arbeidet gjør man n målinger av bølgehøyde, $X_1, \dots, X_n$. Vi antar at bølgeforholdene er konstante i det tidsrommet målingene tas slik at bølgehøyden kan betraktes som stokastiske variable fra samme fordeling, og at denne fordelingen er gitt ved

$$f_X(x) = \frac{2x}{\theta} e^{-x^2/\theta}, \quad x > 0.$$

Denne fordelingen kalles en *Rayleigh*-fordeling.

- a) Anta i dette punktet at $\theta = 3$. Hva er sannsynligheten for at en tilfeldig valgt bølge har en bølgehøyde på mer enn 4 meter?
Hva er sannsynligheten p for at to etterfølgende bølger begge skal ha en bølgehøyde på mer enn 4 meter dersom vi antar at de er uavhengige?
Hvordan vil det påvirke verdien til p dersom høyden til to etterfølgende bølger er positivt korrelert? (Begrunn om p vil øke, avta eller være uendret.)

I praksis er parameteren θ ukjent og skal estimeres fra målingene $X_1, \dots, X_n$, som vi nå antar er uavhengige.

- b) Vis ved hjelp av transformasjonsformelen at $Z = 2X^2/\theta$ er χ^2_2 -fordelt.
c) Vis at sannsynlighetsmaksimeringsestimatoren (SME) for θ , $\hat{\theta}$, er gitt ved

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i^2$$

Finn $E(\hat{\theta})$ og $\text{Var}(\hat{\theta})$. (Hint, resultatet fra b) kan komme til nytte.)

- d) Ta utgangspunkt i resultatet fra punkt b) og finn hva fordelingen til $\frac{2n\hat{\theta}}{\theta}$ blir. Utled et $(1-\alpha)100\%$ konfidensintervall for θ . Hva blir verdien på konfidensintervallet når $\alpha = 0.05$ og man fra observerte data har informasjonen

$$n = 50, \quad \frac{1}{n} \sum_{i=1}^n X_i = 0.911, \quad \frac{1}{n} \sum_{i=1}^n X_i^2 = 1.047$$

La $Y = \max(X_1, \dots, X_m)$ være den høyeste av m antatt uavhengige bølgehøyder i det kommende 3-timers intervallet.

- e) Finn et uttrykk for den kumulative fordelingsfunksjonen til Y , $F_Y(y)$.
Bruk dette uttrykket til å vise at dersom y_k er den verdien som er slik at

$$P(Y > y_k) = \frac{1}{100}$$

så er

$$y_k = \sqrt{-\theta \ln \left[1 - \left(1 - \frac{1}{100} \right)^{1/m} \right]}.$$

Anta at $m = 900$. Man starter arbeidsoppgaven dersom målingene gir grunnlag for å hevde at $y_k < 4$.

- f) Gir målingene rapportert i punkt d) grunnlag for å hevde at $y_k < 4$? Formuler problemstillingen som en hypotesetest og utfør testen på 5% nivå.

Noen fasitsvar:

1: a) 0.0020, 0.0142, 0.020; b) $E(\hat{\lambda}) = \lambda$ og $\text{Var}(\hat{\lambda}) = \lambda / \sum_{i=1}^n t_i$;

c) $H_0: \lambda = 6$ mot $H_1: \lambda > 6$. Ikke forkast H_0 ;

2: a) 0.0048 og 0.000023; c) θ og θ^2/n ; d) [0.81, 1.41] e) $[1 - e^{y^2/\theta}]^m$ f) forkast