Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Kontaktperson:
Håvard Rue 92600021

Eksamen i ST1201/ST6201 Statistiske metoder

Fredag 3.desember 2010
Tid: 09:00–13:00

Tillatte hjelpemidler:

- Tabeller og formler i statistikk (Tapir forlag)
- Gyldig kalkulator

Alle svar skal begrunnes.

Notasjon: «log» er den naturlige logaritmen.

You may answer in English or Norwegian.
Du kan besvare enten på engelsk eller norsk (begge målføre).

Sensurfrist: mandag 3.januar 2011

Oppgave 1

La X være Binomisk fordelt: $X \sim$ Binomial$(n = 300,\ p)$.

a) Finn et tilnærmet 90% konfidensintervall for p når vi observerer $x = 75$.

b) Den parameteren vi er interessert i er ikke p, men θ, hvor

$$p = \frac{\exp(\theta)}{1 + \exp(\theta)}.$$

Finn (ved hjelp av **a)**) et tilnærmet 90% konfidensintervall for θ når vi observerer $x = 75$.

Oppgave 2

La

$$X_1, X_2, \ldots, X_n, X_{n+1}$$

være et tilfeldig utvalg av størrelse $n+1$ fra $N(\mu, \sigma^2)$, der både μ og σ^2 er ukjent.

La

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

og

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

Vi skal i denne oppgaven lage et prediksjonsintervall for X_{n+1} basert på $X_1, \ldots, X_n$.

a) Vis at

$$\frac{X_{n+1} - \overline{X}}{S\sqrt{1 + 1/n}}$$

er Student-t fordelt med $n-1$ frihetsgrader.

b) Finn k slik at

$$\text{Prob}(\overline{X} - kS < X_{n+1} < \overline{X} + kS) = 0.80$$

(Det observerte intervallet $(\overline{x} - ks, \overline{x} + ks)$ er et 80% prediksjonsintervall for X_{n+1}.)

Oppgave 3

La

$$X = (X_1, X_2, X_3) \sim \text{Multinomial}(N,\ p = (p_1, p_2, p_3))$$

dvs

$$\text{Prob}(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!\ x_2!\ x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

hvor $x_1 + x_2 + x_3 = N$ og $p_1 + p_2 + p_3 = 1$.

En modell fra genetikk, antar at

$$p_1 = q^2, \qquad p_2 = 2q(1-q), \qquad p_3 = (1-q)^2$$

a) Finn *Maximum likelihood estimator* (MLE) for q, $\widehat{q}$.

Cramér-Rao's ulikhet sier at (med notasjon fra lærebok)

$$\text{Var}(\widehat{\theta}) \geq \left\{ -n\text{E}\left[\frac{\partial^2 \log f_Y(Y;\theta)}{\partial \theta^2} \right] \right\}^{-1}.$$

b) Gjør rede for

1. hva dette resultatet betyr og kan brukes til, samt
2. hvilke forutsetninger som ligger til grunn for dette resultatet.

c) Bruk Cramér-Rao's ulikhet til å finne en nedre grense for variansen til en forventningsrett estimator for q.

d) Under generelle antakelser vil MLE for q, når $N \to \infty$, være Normalfordelt, forventningsrett, og ha varians lik Cramér-Rao's nedre grense.

Bruk dette resultatet til å konstruere et tilnærmet 95% konfidensintervall for q.

e) Gjør rede for hvordan du vil undersøke om modellen fra genetikk er urimelig utifra dine observasjoner.

Oppgave 4

Du har fått i oppgave å konstruere en hypotesetest som forkaster H_0 på nivå 5%, slik at

$$\text{Prob}(\text{Forkast } H_0 \mid H_0) = 5\%. \tag{1}$$

Du vil forkaste H_0 hvis observert verdi er stor nok, og fordi observasjonen(e) er diskret har du

$$\text{Prob}(X \geq 100 \mid H_0) = 4\%$$

mens

$$\text{Prob}(X \geq 99 \mid H_0) = 6\%,$$

dvs, du klarer ikke å tilfredsstille ligning (1).

a) Vis at du kan modifisere testen din slik at ligning (1) er oppfylt, dersom du har en (rettferdig) mynt tilgjengelig.

Diskuter resultatet.