STK 1110 høsten 2004
Oblig I

Dette er oppgavesettet for første obligatoriske innleveringsprosjekt for kurset STK 1110. Det legges ut **mandag 20/ix/4**, og en bunke blir også tatt med på forelesningene onsdag og fredag den uken. Leveringsfrist er **fredag 1/x/4**, til instituttkontoret ved Matematisk institutt, senest kl. 14:30.

Du kan levere håndskrevet eller tekstbehandlet besvarelse. Skriv navnet ditt øverst til høyre på første side. Der du bruker MINITAB eller MATLAB må utskrifter og plott legges ved, eller klebes inn i besvarelsen. Disse softwarepakkene skal anvendes i noe større grad siden i kurset, for eksempel i forbindelse med Oblig II (som vil ha leveringsfrist fredag 12/xi/4).

Oppgave 1

VI STARTER MED VERDENS VIKTIGSTE modell, der data $X_1, \ldots, X_n$ er uavhengige fra samme normalfordeling (μ, σ^2).

(a) Hva blir sannsynlighetsfordelingene til

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{og} \quad s = \Big[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\Big]^{1/2},$$

empirisk gjennomsnitt og standardavvik? [Man kan bla i boken her, og reutledning av resultater der trenges ikke.] Anta at de sanne verdiene av μ og σ er lik henholdvis 4.0 og 1.5, og at $n = 10$. Sett opp eksplisitte formler for sannsynlighetstetthetene for $\bar{X}$ og s, og plott disse [helst ved hjelp av softwarepakke].

(b) Hvordan forandres sannsynlighetstetthetene over, når antall datapunkter øker fra $n = 10$ til $n = 20$? Lag figurer som illustrerer dette.

(c) De ti datapunktene [i mitt lille simulerte eksempel] er

$$5.288,\ 5.051,\ 2.727,\ 3.092,\ 4.188,\ 5.020,\ 6.093,\ 5.437,\ 4.186,\ 5.618,$$

hvor man finner $\bar{X} = 4.670$ og $s = 1.100$. Sett opp konfidensintervall, med konfidensgrad 95%, for μ og for σ. Forklar også hvordan disse forandrer seg når n økes fra $n = 10$ til $n = 20$.

Oppgave 2

GOTT WÜRFELT NICHT, mente Einstein, og dessuten er ikke alle terninger rettferdige. Anta at en terning har sannsynlighetene

$$p_1 = \tfrac{1}{6} - \theta, \quad p_2 = \tfrac{1}{6}, \quad p_3 = \tfrac{1}{6}, \quad p_4 = \tfrac{1}{6}, \quad p_5 = \tfrac{1}{6}, \quad p_6 = \tfrac{1}{6} + \theta,$$

for sine seks forskjellige utfall, der θ er en ukjent parameter.

(a) Hvordan vil du fortolke parameteren θ? Hva er dens naturlige parameterområde?

(b) La X være resultatet av et tilfeldig terningkast. På hvilken måte adskiller forventningsverdien for X seg fra det den skal være for en rettferdig terning?

(c) Terningen kastes $n = 600$ ganger, resulterende i Y_1 1-ere, Y_2 2-ere, osv. Forklar hvorfor likelihoodfunksjonen blir proporsjonal med

$$L(\theta) = p_1^{Y_1} p_2^{Y_2} \cdots p_6^{Y_6}.$$

Finn en formel for maximum likelihood-estimatoren (rimelighetsfunksjonsmaksimeringsestimatoren) $\widehat{\theta}$ for θ.

(d) Anvend generelle resultater fra pensum til å sette opp grensefordelingen for $\widehat{\theta}$, passende normert.

(e) Anta at eksperimentet gir resultatene $85, 103, 100, 95, 102, 115$ for antall 1-ere, antall 2-ere, osv. Lag et konfidensintervall for θ med konfidensgrad tilnærmet 90%.

(f) Er terningen rettferdig? Sett opp en presis hypotese, test den, og formuler en konklusjon.

Oppgave 3

Anta at observasjoner $X_1, X_2, \ldots, X_n$ er uavhengige og at de alle kommer fra sannsynlighetsfordelingen med tetthet

$$f(x) = 12x^2(1-x) \quad \text{for } x \in (0,1).$$

(a) Lag en skisse av tettheten. Vis at forventningen og variansen blir henholdsvis

$$\mathrm{E}X = 3/5 \quad \text{og} \quad \mathrm{Var}\, X = 1/25.$$

(b) La $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ være gjennomsnittet av de n observasjonene. Bruk Tsjebysjofs ulikhet (неравенство Чебышёва) til å sette opp en begrensning for sannsynligheten for at $\bar{X}_n$ skal være mer enn 0.10 fra 0.60. Forklar hvorfor $\bar{X}_n$ vil konvergere i sannsynlighet mot 0.60 når n øker.

(c) Anta i dette punktet at $n = 100$. Bruk sentralgrenseteoremet (the central limit theorem) til å lage et intervall av typen $[a, b] = [0.60 - \delta, 0.60 + \delta]$ som med sannsynlighet tilnærmet 90% vil fange den stokastiske $\bar{X}_n = \bar{X}_{100}$.

(d) Hvor stor må n være for at intervallet $[a, b] = [0.58, 0.62]$ skal inneholde $\bar{X}_n$, med sannsynlighet minst 90%?

Oppgave 4

I EN STOR NOK FLOSSHATT ligger det et ukjent antall N like store lapper, nummerert fra 1 til N. Ti ganger på rad trekker du en lapp fra hatten, helt tilfeldig, med tilbakelegging, og får svarene

$$939,\ 169,\ 754,\ 843,\ 74,\ 849,\ 843,\ 150,\ 598,\ 437.$$

Den statistiske oppgaven går ut på å anslå N, helst supplert med et estimat for dets usikkerhet.

(a) Vi kaller resultatene av de $n = 10$ operasjonene $X_1, \ldots, X_{10}$, og antar at alle utfall er like sannsynlige for hver av dem, altså at

$$\mathrm{P}\{X_i = x\} = \frac{1}{N} \quad \text{for } x = 1, 2, \ldots, N.$$

Vis nå at

$$\mathrm{E} X_i = \frac{N+1}{2} \quad \text{og} \quad \mathrm{Var}\, X_i = \frac{N^2 - 1}{12}.$$

Her bør du bruke følgende klassiske formler:

$$1 + 2 + \cdots + m = \frac{m(m+1)}{2}, \quad 1^2 + 2^2 + \cdots + m^2 = \frac{m(m+1)(2m+1)}{6}.$$

(b) Vis at momentestimatoren (altså den estimator som fremkommer ved å anvende momentmetoden) blir

$$\widehat{N} = 2\bar{X}_n - 1.$$

Beregn estimatet, og anslå dets standardavvik.

(c) Vis så at maximum likelihood-estimatoren blir

$$N^* = \max\{X_1, \ldots, X_n\}.$$

Beregn også dette estimatet.

(d) Som en sannsynlighetsteoretisk øvelse, vis at N^* er konsistent for N, når n, antallet ganger man trekker en lapp, vokser.

(e) Anta for illustrasjonens skyld at den virkelige N er 1066. Hvilken av de to estimatorene er da egentlig best, målt ut fra bruttovariansene $\mathrm{E}(\widehat{N} - N)^2$ og $\mathrm{E}(N^* - N)^2$? Generaliser gjerne det du måtte finne frem til.

Nils Lid Hjort