UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1100 — Sannsynlighetsregning og statistisk modellering
Eksamensdag:	Tirsdag 31. mai 2016.
Tid for eksamen:	09.00 – 13.00.
Oppgavesettet er på 4 sider.	
Vedlegg:	Ingen
Tillatte hjelpemidler:	Godkjent kalkulator og formelsamling for STK1100/STK1110

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Nedenfor får du oppgitt følgende *øvre* kvantiler i standard normal fordelingen som kan brukes ulike steder i oppgavesettet.

α	0.100	0.050	0.025	0.010	0.005	0.0025
z_α	1.281	1.645	1.960	2.326	2.576	2.807

Oppgave 1

Vi vil i denne oppgaven studere Laplace-fordelingen, som er en kontinuerlig fordeling for en stokastisk variabel X definert ved

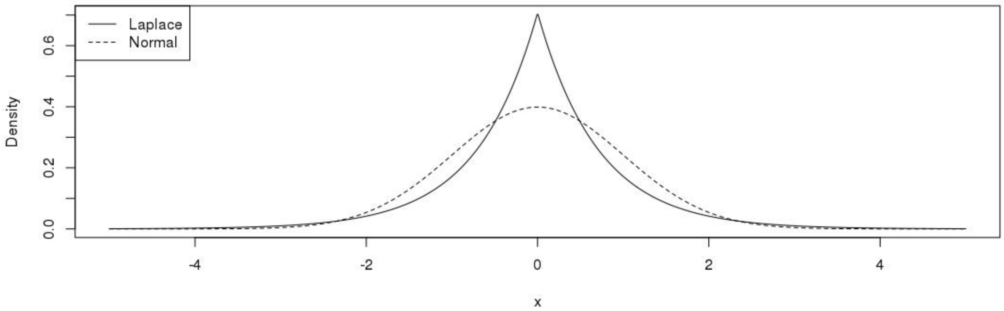
$$f_X(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}; \quad -\infty < x < \infty.$$

Man kan vise at den momentgenererende funksjonen for Laplace-fordelingen er gitt ved

$$M_X(t) = \frac{e^{\mu t}}{1 - b^2 t^2}$$

(dette behøver du ikke å vise).
Plottet på neste side viser Laplace-fordelingen for $\mu = 0$ og $b = 1/\sqrt{2}$ sammen med standard normalfordelingen.

(Fortsettes på side 2.)



(a) Vis at både forventning og median er μ for Laplace-fordelingen.

(b) Finn variansen $V(X)$ i Laplace-fordelingen.

(c) Anta Z_1 og Z_2 er uavhengige og begge eksponentielt fordelt med parameter λ. Vis at $\mu + Z_1 - Z_2$ er Laplace-fordelt.

(d) Foreslå en metode for å simulere (dvs generere på en datamaskin) en variabel fra Laplace-fordelingen.

(Du kan her anta at det finnes en metode for å generere en uniformt fordelt variabel).

Anta nå vi har observasjoner $x_1, ..., x_n$ som er observerte verdier av uavhengige identisk fordelte variabler $X_1, ..., X_n$ fra Laplace-fordelingen. Vi vil se på hvordan vi kan estimere μ. To mulige estimatorer er

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$\tilde{\mu} = \text{median}(X_1, ..., X_n)$$

Vi vil se på disse estimatorene basert på et datasett med $n = 49$, der dataene er angitt nedenfor.

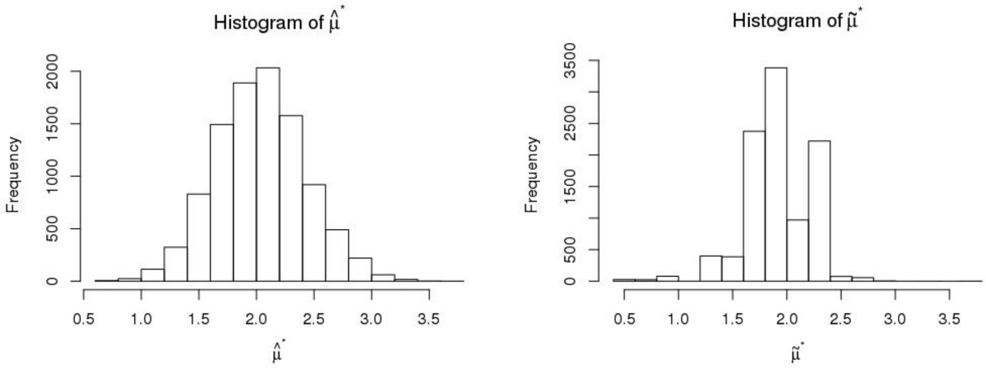
5.36	4.32	0.41	0.63	1.33	0.32	-3.13	2.35	5.35	3.71
6.40	-1.81	1.67	1.86	2.22	1.89	4.47	7.36	4.89	1.50
-0.54	0.60	2.99	-3.37	12.05	2.21	2.00	2.48	3.78	0.11
0.10	0.88	1.33	1.80	2.21	2.87	1.69	-0.44	2.61	0.48
2.75	-0.16	-3.63	5.97	3.61	-0.03	2.11	-0.27	2.33	

For disse dataene er $\hat{\mu} = 2.03$ mens $\tilde{\mu} = 1.895$.
Ikke-parametrisk Bootstrapping kan benyttes for å estimere standardfeil for de to estimatorene. En kjøring med $B = 10000$ Bootstrap-simuleringer ga

$$\overline{\hat{\mu}}^* = \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\hat{\mu}_b^* = 2.0353, \quad \hat{\sigma}_{\hat{\mu}} = 0.391$$
$$\overline{\tilde{\mu}}^* = \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\tilde{\mu}_b^* = 1.9155, \quad \hat{\sigma}_{\tilde{\mu}} = 0.283$$

(Fortsettes på side 3.)

der $\hat{\mu}_b^*$ og $\tilde{\mu}_b^*$ er Bootstrap-simuleringer av $\hat{\mu}$ og $\tilde{\mu}$, henholdsvis.
Plottet nedenfor viser også histogram av Bootstrap-simuleringene.



(e) Forklar hvordan estimatene for standardfeilene ovenfor er konstruert.

Diskuter hvorfor det kan være rimelig at standardfeilen for $\tilde{\mu}$ er mindre enn standardfeilen for $\hat{\mu}$ i dette tilfellet.

(f) Basert på en antagelse om at både $\hat{\mu}$ og $\tilde{\mu}$ er tilnærmet normalfordelte, *utled* 95% konfidensintervaller for μ basert på hver av de to ulike estimatorene. Bruk de tall som er oppgitt for å spesifisere de to intervaller du får.

Diskuter om de antagelser som ligger til grunn for utledningen av konfidensintervallene er rimelig for de to estimatorer.

Oppgave 2

En student besvarer en flervalgs oppgave som angir 5 mulige svaralternativer. Anta at sannsynligheten for at studenten vet svaret er r og at sannsynligheten for at studenten gjetter svaret er $1-r$. Anta også at hvis studenten gjetter, så er sannsynligheten for å gjette riktig svar 0.2. Vi antar også at hvis studenten vet svaret, så svarer hun korrekt.

(a) Vis at sannsynligheten for at en student svarer riktig på oppgaven er $r + (1 - r) \cdot 0.2$.

Hvis studenten svarer riktig, hva er sannsynligheten for at studenten visste det korrekte svaret?

Hva blir disse sannsynlighetene når $r = 0.6$?

Under midtveis-eksamen i STK1100 var det 238 besvarelser, hver med svar på 15 spørsmål, alle spørsmål hadde 5 alternativer. Av de totalt $238 * 15 = 3570$ svar var 2433 riktige.

(b) La X være antall riktige svar. Anta $X \sim \text{Binom}(3570, p)$.

Lag et estimat for p og et (tilnærmet) 99% konfidensintervall for p.

(Fortsettes på side 4.)

(c) Diskuter rimeligheten mhp den binomiske antagelsen i forrige delpunkt.

Vi vil nå konsentrere oss om den siste deloppgaven i midtveiseksamnen. La Y være antall som besvarte denne deloppgaven riktig. Vi vil anta $Y \sim \text{Binom}(238, q)$.
Av de 238 besvarelsene var det 57 som svarte riktig. Et (tilnærmet) 95% konfidensintervall for q blir $[0.185, 0.294]$ (dette behøver du ikke å vise).

(d) Bruk informasjonen ovenfor til å lage et tilnærmet 95% konfidensintervall for sannsynligheten r diskutert i begynnelsen av denne oppgaven.

Er det noen grunn til å tro studentene kunne denne oppgaven?

Diskuter også om antagelsene som ligge til grunn er rimelige her.

Oppgave 3

Vi skal i denne oppgaven se på en bivariat sannsynlighetsfordeling for to kontinuerlige tilfeldige variabler (X, Y) der sannsynlighetstettheten er gitt ved

$$f(x,y) = \begin{cases} x + y, & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{ellers} \end{cases}$$

(a) Finn de marginale tetthetsfunksjonene til X og Y.

(b) Finn $P(X \geq 0.5 | Y \geq 0.5)$.