Tilleggsoppgaver for STK1100

Matematisk institutt

Tilleggsoppgave 1
I en eske er det to pengestykker. Det ene er normalt, mens det andre har krone på begge sider. Du trekker tilfeldig ett pengestykke og kaster det tre ganger.

(*a*) Hva er sannsynligheten for at du får krone i alle de tre kastene?

(*b*) Anta at du fikk krone i alle de tre kastene. Hva er sannsynligheten for at du har kastet med det normale pengestykket?

Tilleggsoppgave 2 (Utfordring - poker)
I tradisjonell poker (five card draw") får en spiller delt ut fem av kortstokkens 52 kort. En slik samling av fem kort kaller vi en pokerhånd". En pokerhånds verdi avhenger av hvor ofte den opptrer. Her følger en beskrivelse av ulike pokerhender:

- Royal straight flush: De fem høyeste kortene i samme farge, dvs. ess-K-D-Kn-10.
- Straight flush: Fem kort i rekkefølge i samme farge, f.eks. 6-7-8-9-10.
- Fire like: Fire kort med samme verdi, f.eks. fire 5-ere.
- Fullt hus: Tre kort av samme verdi (tre like) sammen med et par (to like) av en annen verdi.
- Flush: Fem kort i samme farge uansett verdi.
- Straight: Fem kort i rekkefølge uansett farge.
- Tre like: Tre kort med samme verdi.
- To par: To kort (par) av en verdi sammen med et par av en annen verdi.
- Ett par: To kort (par) av en verdi.

Hvis en pokerhånd passer til to eller flere av beskrivelsene, er det den mest verdifulle, (dvs. den øverste på lista) som gjelder. For eksempel regnes en pokerhånd med tre kort av en verdi og to av en annen som fullt hus og ikke som tre like, to par eller ett par. Videre er det slik at ess gjelder både som laveste (1) og høyeste (14) kort i straight og straight flush.

(*a*) På hvor mange måter kan en pokerspiller få

(*i*) royal straight flush

(*ii*) straight flush

(*iii*) fire like

(*iv*) fullt hus

(*v*) flush

(*vi*) straight

(*vii*) tre like

(*viii*) to par

(*ix*) ett par

(*b*) Hva er sannsynligheten for å få delt ut hver av pokerhendene i oppgave (*a*)? Vi forutsetter at kortstokken er stokket veldig godt.

Tilleggsoppgave 3 (Utfordring - kortbunker)
På bordet ligger det to bunker med kort. I bunke A er det 4 røde og 8 svarte kort, mens det i bunke B er 6 røde og 6 svarte kort. Kortene har baksiden opp. Du velger tilfeldig én av bunkene og trekker to kort fra den. Det viser seg at begge kortene du trekker er svarte.

Du skal nå trekke ett kort til, enten fra den samme bunken som du trakk de to kortene fra eller fra den andre bunken.

Hvilken bunke bør du trekke fra for å få størst mulig sannsynlighet for å få et rødt kort?

Tilleggsoppgave 4 (Matlab oppgave)
På kursets semesterside er det en link til data om levetidsfordelingen for norske menn og kvinner. Her finner du filen `doedssans.txt` som gir ettårige dødssannsynligheter (i promille) for norske menn og kvinner basert på statistikk for perioden 2009-2013. Tabellen angir

$$q_x = P(X = x | X \geq x)$$

for $x = 0, 1, ...$ Basert på dette kan en bruke produktsetningen til å vise at

$$\begin{aligned} F(0) &= q_0 \\ F(x) &= P(X \leq x) = 1 - (1 - q_0)(1 - q_1) \cdots (1 - q_x) \qquad\qquad x > 0 \end{aligned}$$

og dermed at

$$\begin{aligned} p(0) &= F(0) \\ p(x) &= P(X = x) = F(x) - F(x - 1) \end{aligned}$$

MATLAB kommander for å beregne disse punktsannsynlighetene for menn er gitt i filen `levetidsfordeling_menn.m` som er tilgjengelig fra "matlab-kode" siden under kursets hjemmeside.

(*a*) Plott de ettårige dødssannsynlighetene for kvinner og menn i samme figur og diskuter hvordan dødeligheten varierer med alderen for kvinner og menn. (For å få dødssannsynlighetene for menn og kvinner i samme figur, plotter du først for menn slik det ble vist på forelesningen. Så gir du kommandoen `hold on` og plotter for kvinnene. Så gir du kommandoen `hold off` slik at du får en ny figur neste gang du plotter.)

(*b*) La X være levealderen til en tilfeldig valgt norsk mann og la Y være levealderen til en tilfeldig valgt norsk kvinne. Regn ut den kumulative fordelingsfunksjonen $F(x) = P(X \leq x)$ for levealderen til menn og den kumulative fordelingsfunksjonen $G(y) = P(Y \leq y)$ for levealderen til kvinner og tegn dem i samme figur. Bruk de kumulative fordelingsfunksjonene til å finne sannsynligheten for at en mann/kvinne blir minst 60 år, minst 70 år og minst 80 år.

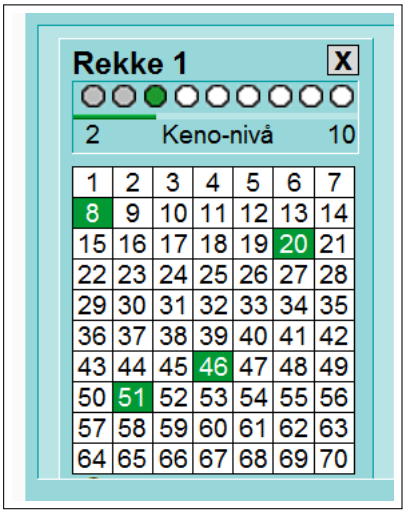
(*c*) Medianalderen $\tilde{\mu}_X$ for menn er den minste verdien av x som er slik at $F(x) \geq 0.5$ og medianalderen $\tilde{\mu}_T$ for kvinner er den minste verdien av y som er slik at $G(y) \geq 0.5$. Bestem medianalderen for menn og kvinner. (Hvis en vil være mer nøyaktig, kan en bestemme medianalderen ved interpolasjon.)

(*d*) Bestem punktsannsynligheten $p_x(x) = P(X = x)$ for menn og $p(y) = P(Y = y)$ for kvinner og tegn dem opp i hver sin figur.

(*e*) Bestem forventet levealder $E(X)$ for menn og $E(Y)$ for kvinner. Sammenlign forventet levealder med medianalderen.

Tilleggsoppgave 5 (Utfordring)

Keno er et er et kjent pengespill. I Norge er det Norsk Tipping som er arrangør av spillet.
På Kenokupongen er tallene fra 1 til 70 gitt. Når du spiller én rekke Keno, avgjør du først om du vil krysse av for 2, 3, 4, 5, 6, 7, 8, 9 eller 10 tall på kupongen. Hvis du krysser av for to tall, spiller du på Keno-nivå 2, hvis du krysser av for tre tall spiller du på Keno-nivå 3, osv.
Til høyre ser du et eksempel på en utfylt kupong på nivå 10.
Vi vil i denne oppgaven tenke oss at du spiller én rekke på nivå 10. For det betaler du 10 kroner.
Ved Keno-trekningen velges det tilfeldig ut 20 tall blant tallene fra 1 til 70.



Figur 1: Keno

Tabellen nedenfor viser hvordan utbetalingen avhenger av hvor mange av tallene dine som blir trukket ut.

Keno-nivå	Antall rette	Odds	Premie ved rekkepris: 10,-
10	10	200 000	2 000 000,-
	9	5 000	50 000,-
	8	200	2 000,-
	7	20	200,-
	6	4	40,-
	5	1	10,-
	0	1	10,-

La X være den utbetalingen (premien) du får.

(*a*) Skriv opp de mulige verdiene for X.

(*b*) Bestem punktsannsynligheten for hver av de mulige verdiene for X.

(*c*) Hva er sannsynligheten for at du får utbetalt minst 200 kroner?

Hva er sannsynligheten for at du får utbetalt 10 kroner?

(*d*) Bestem forventningsverdien.

(*e*) Hvor mye vil du i gjennomsnitt tape/vinne per spill hvis du mange ganger spiller én rekke på Keno-nivå 10?

Tilleggsoppgave 6 (Utfordring - Entydighet av momentgenererende funksjoner)
Anta først vi har to Bernoulli variable X_1 og X_2 der

$$p_1 = P(X_1 = 1) = 1 - P(X_1 = 0)$$
$$p_2 = P(X_2 = 1) = 1 - P(X_2 = 0)$$

Anta videre at $M_{X_i}(t)$ er den momentgenererende funksjonen til X_i for $i = 1, 2$. Vi antar begge funksjonene eksisterer for $t \in (-t_0, t_0)$.

(*a*) Vis at hvis $M_{X_1}(t) = M_{X_2}(t)$ for $t \in (-t_0, t_0)$ så må $p_1 = p_2$

Se så på to generelle diskrete stokastiske variable Y_1 og Y_2 definert over utfallsrommene $\mathcal{D}_1$ og $\mathcal{D}_2$ med tilhørende momentgenererende funksjoner $M_{Y_1}(t)$ og $M_{Y_2}(t)$.

(*b*) Argumenter for hvorfor vi alltid kan lage et felles utfallsrom $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ for de to variablene Y_1 og Y_2.

(*c*) Anta nå $M_{Y_1}(t) = M_{Y_2}(t)$ for $t \in (-t_0, t_0)$. Vis at da må $P(Y_1 = y) = P(Y_2 = y)$ for alle $y \in \mathcal{D}$.

Hint: Definer

$$X_i = \begin{cases} 1 & \text{hvis } Y_i = y \\ 0 & \text{ellers} \end{cases}$$

for $i = 1, 2$ og bruk de tidligere resultater.

Tilleggsoppgave 7
I denne oppgaven skal vi bruke MATLAB til å simulere kast med to terninger. Vi vil først se på den stokastiske variabelen

$X =$ "sum antall øyne"

og undersøke hvordan det går med gjennomsnittet av X-verdiene når vi kaster mange ganger.

Vi kan simulere $n = 500$ kast med to terninger ved kommandoene:

```
n = 500
t1 = randi(6, [1 n]);
t2 = randi(6, [1 n]);
```

Vektorene `t1` og `t2` vil da gi antall øyne for hver av de to terningene i de n kastene.

Vi kan så beregne summen av antall øyne i de N kastene ved kommandoen:

```
x = t1+t2
```

Vi er interessert i å studere hvordan gjennomsnittet av sum antall øyne endrer seg etter hvert som vi kaster. Etter n kast er gjennomsnittet lik summen av X-verdiene for de n første kastene dividert med n. Vi kan bestemme gjennomsnittet som en funksjon av n ved kommandoene:

```
gjsnX=cumsum(x)./[1:n];
```

Vi kan så plotte gjennomsnittet etter n kast som en funksjon av n og sette skalaen for aksene i plottet ved kommandoene:

```
plot(n,gjsnX);
axis([1 n 2 12]);
```

(a) Utfør kommandoene ovenfor. Pass på at du forstår hva hver av kommandoene gjør!

Av det plottet du får, vil det se ut som om gjennomsnittet av X-verdiene nærmer seg en bestemt verdi. Hvilken verdi er det?

Vi ser så på den stokastiske variabelen $Y =$"største antall øyne". Du kan finne verdien av denne for de n terningkastene ved kommandoen:

```
y = max(t1,t2)
```

(b) Gjenta kommandoene ovenfor for Y. Hvordan går det med gjennomsnittet av Y-verdiene når n øker?

(*c*) Bestem $P(Y = y)$ for $y = 1, 2, 3, 4, 5, 6$ og bruke denne til å finne $E(Y)$. Forklar resultatet i punkt (*b*) ut fra denne forventningsverdien.

Tilleggsoppgave 8 (Utfordring)
I september 1990 stod det et leserbrev i spalten "Ask Marilyn" i det amerikanske bladet "Parade Magazin". Leseren spør:

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

Redaktøren av spalten, Marilyn vos Savant, anbefalte deltakeren i showet å bytte dør. Da ville sannsynligheten for å vinne bilen bli 2/3, mens sannsynligheten for å vinne bare ville være 1/3 hvis deltakeren holdt fast ved dør nummer 1. Dette svaret førte til en storm av protester, også fra fagfolk, som mente at sannsynligheten for å vinne bilen ville være 50% uansett om deltakeren byttet dør eller ikke.

(*a*) Gjør deg opp en intuitiv mening om hvem som har rett, Marilyn vos Savant eller hennes kritikere.

(*b*) Klarer du å gi et formelt argument for din oppfatning?

Tilleggsoppgave 9
Få tak i en tippekupong. På baksiden av denne er det gitt en "systemoversikt", det vil si en oversikt over hvor mange rekker en tipper på en systemkupong med et visst antall hel- og halvgarderinger. Forklar hvordan disse tallene er beregnet.

Tilleggsoppgave 10
Få tak i en vanlig lottokupong (ikke Vikinglotto). På denne er det gitt "kontrollfelt for systemkupong", det vil si en oversikt over hvor mange lottorekker en tipper på en systemkupong hvor det krysses av mer enn sju tall. Forklar hvordan disse tallene er beregnet.

Tilleggsoppgave 11
Nedenfor er det gitt en tabell over dødssannsynligheter for norske kvinner i alder 75-80 år (basert på dødelighetstabellen for Norge for 2002):

Alder	Dødssannsynligheter pr 1000 kvinner
75	26.17
76	28.45
77	30.12
78	37.74
79	41.57
80	44.11

(*a*) Bestem sannsynligheten for at en 75 år gammel kvinne skal bli minst 80 år.

(*b*) Bestem sannsynligheten for at en 75 år gammel kvinne skal bli nøyaktig 80 år, dvs. oppleve sin åttiende, men ikke sin åttiførste fødselsdag.

Tilleggsoppgave 12
HIV-tester er, som de fleste diagnostiske metoder, ikke helt pålitelige. For en HIV-test har vi at:

- Hvis en person er smittet av HIV-viruset, vil testen avsløre dette i 98% av tilfellene.
- Hvis en person ikke er smittet av HIV-viruset, vil undersøkelsen likevel indikere at personen er smittet, dvs. avgi falsk alarm, i 0.2% av tilfellene.

Vi antar først at vi tester personer i en høyrisikogruppe hvor 10% av individene er HIV-smittet. Et tilfeldig valgt individ fra denne gruppen blir testet.

(*a*) Hva er sannsynligheten for at testen vil indikere at personen er HIV-smittet?

(*b*) Anta at testen indikerer at personen er HIV-smittet. Hva er (den betingede) sannsynligheten for at han/hun virkelig er smittet?

Vi antar så at vi tester personer i en lavrisikogruppe hvor bare 1 av 10000 er HIV-smittet. Et tilfeldig valgt individ fra denne gruppen blir testet og testen indikerer at han/hun er smittet.

(*c*) Hva er (den betingede) sannsynligheten for at han/hun virkelig er HIV-smittet?

Tilleggsoppgave 13
(*a*) Vi skal simulere myntkast i MATLAB og se hvor ofte vi får kron. Vi gjør dette ved å trekke tall fra mengden $\{1, 2\}$ med like stor sannsynlighet for hver, og la 1 være mynt og 2 være kron. Sannsynligheten for kron er altså $p = \frac{1}{2}$.

I MATLAB kan vi simulere m myntkast ved å bruke

```
Y = unidrnd(2, [1,m]);
```

Det første argumentet, `2`, er antall mulige utfall. Det andre argumentet, `[1,m]`, angir størrelsen på resultatet `Y` - i dette tilfellet en vektor med `m` elementer (egentlig en matrise av dimensjon `1` $\times$ `m`).

Bruk kommandoen som i eksemplet over og simuler 100 myntkast. Vis resultatet i kommandovinduet.

(*b*) Hvert 2-tall tilsvarer et myntkast som ga kron. Fremfor å regne på en vektor som inneholder 1- og 2-tall skal vi heller regne på en vektor som inneholder 0- og 1-verdier hvor 1 indikerer at myntkastet ga kron. En slik vektor kan lages ved

```
X = (Y == 2);
```

Hvorfor blir dette riktig? Sammenlign verdiene i vektorene `Y` og `X`.

(c) Bruk

```
k = sum(X)
```

til å telle opp hvor mange kast som ga kron. Et anslag av p vil være antall kron dividert på antall kast. Lag et slikt anslag. Ble det et godt anslag?

(d) Resultatene fra de `i` første kastene kan vi få frem ved å bruke `X(1:i)`. I uttrykket `X(1:i)` er `1:i` en vektor med alle tallene 1,2,...,`i`. Med andre ord, `X(1:i)` plukker ut de `i` første elementene i `X`.

Vi kan bruke dette til å lage et anslag av p basert på de `i` første kastene, med

```
for i=1:m
q(i) = sum(X(1:i))/i;
end
```

Bruk dette til å lage et anslag av p basert på 10, 30 og 100 myntkast. Hvordan forventer du at anslagene blir (dårligere, bedre, uforandret) etter hvert som antallet kast øker?

Stemte anslagene med dine forventninger?

(e) For-løkken ovenfor kan unngås ved å bruke vektorielle opperasjoner og ferdige funksjoner. Prøv ut kommandoene

```
K = cumsum(X);
M = 1:m;
Q = K ./M;
```

og sjekk at du får de samme svar som ovenfor.

(f) Å se hva som skjer med anslaget ettersom antallet kast øker kan være enklere dersom vi plotter anslaget, `Q`, som en funksjon av antall myntkast, `M`. Et slikt plott kan lages med

```
plot(M,Q)
```

Blir anslagene av p bedre ettersom antall myntkast øker?

(g) I stedet for 100 myntkast, simuler nå 10000 myntkast og lag et tilsvarende plott av anslaget som en funksjon av antall myntkast. Hvordan blir anslaget nå ettersom antall kast øker?

Tilleggsoppgave 14 (Matlab oppgave)
Du vil i denne oppgaven få bruk for kommandoen `binopdf` for å beregne binomiske sannsynligheter. Gi kommandoen `help binopdf` for å få nærmere informasjon om kommandoen.

Vi skal bruke følgende datasett fra 1800-tallet i Saxen i Tyskland om kjønnsfordelingen i 6115 tolvbarnsfamilier:

Antall gutter	Antall jenter	Antall familier
0	12	3
1	11	24
2	10	104
3	9	286
4	8	670
5	7	1033
6	6	1343
7	5	1112
8	4	829
9	3	478
10	2	181
11	1	45
12	0	7

Betrakt en tilfeldig valgt tolvbarnsfamilie (fra Saxen i forrige århundre), og la X være antall gutter i denne. Vi vil diskutere mulige sannsynlighetsmodeller for X, dvs. spesifikasjoner av punktsannsynlighetene $p(x) = P(X = x)$ for $x = 0, 1, 2, \ldots, 12$.

(a) En mulig sannsynlighetsmodell er å anta at X er binomisk fordelt med $n = 12$ og $p = 0.50$. Hvilke forutsetninger bygger en slik modell på?

Beregn punktsannsynlighetene $p(x) = P(X = x)$ for $x = 0, 1, 2, \ldots, 12$ under de gitte forutsetningene.

Det er kjent at det systematisk blir født flere gutter enn jenter. Den relative frekvensen av guttefødsler i de 6115 tolvbarnsfamiliene var 0.52. (Kontroller dette!)

(b) En alternativ sannsynlighetsmodell til den i punkt (a) er å anta at X er binomisk fordelt med $n = 12$ og $p = 0.52$. Hvilke forutsetninger bygger en slik modell på?

Beregn punktsannsynlighetene $p(x) = P(X = x)$ for $x = 0, 1, 2, \ldots, 12$ under de gitte forutsetningene.

(c) Beregn den relative frekvensene av familier med x gutter for $x = 0, 1, 2, \ldots, 12$. Sammenlign med punktsannsynlighetene i (a) og (b).

(d) Syns du at sannsynlighetsmodellene i punktene (a) og/eller (b) gir en rimelig god beskrivelse av virkeligheten? Har du noen tanker om hva eventuelle avvik kan komme av?

Tilleggsoppgave 15
Betrakt situasjonen hvor vi gjør en rekke forsøk etter hverandre, og hvert forsøk registrerer om en begivenhet A inntreffer eller ikke .Vi skal i denne oppgaven både betrakte situasjonen hvor forsøkene er avhengige og hvor de er uavhengige.

Konkret vil vi betrakte en modell for avhengighet (en såkalt Markov-modell) som er slik at de betingede sannsynlighetene for resultatet av i-te forsøk, gitt de tidligere forsøkende, bare avhenger av resultatet av det $i-1$-te forsøk. Hvis A_i angir at A inntreffer i i-te forsøk vil vi videre anta at $P(A_1) = 0.5$ og $P(A_i|A_{i-1}) = P(A_i'|A_{i-1}') = q$ for en q mellom 0 og 1.

(a) Vis at vi (ubetinget) har $P(A_i) = 0.5$ for alle i.

(b) Forklar hvorfor $q = 0.5$ svarer til at A_i-ene er uavhengige.

Vi kan simulere situasjonen over. Last ned MATLAB programmet `markov.m`. Vi kan simulere $n = 100$ forsøk for sitasjonen over, f.eks med $q = 0.3$, ved MATLAB kommandoene

```
n = 100;
q = 0.3;
res = markov(n,q);
```

Vektoren `res` vil inneholde $n = 100$ verdier av 0-ere og 1-ere. Her svarer 0-erne til at A ikke inntraff og 1-erne til at A inntraff.

(c) Utfør kommandoene over og studer den sekvensen du får. Legg spesielt merke til hvor lange "følgene" er, dvs. hvor mange ganger etter hverandre A inntreffer eller A' inntreffer.

(d) Gjenta kommandoene med $q = 0.5$ og $q = 0.7$. Studer sekvensene du nå får. Hvordan avhenger resultatet av verdien av q?

(e) Nedenfor er det gitt to sekvenser som hver gir resultatet av $n = 100$ forsøk. Den ene sekvensen er et resultat av uavhengige forsøk, mens for den andre er forsøkene avhengige. Avgjør hvilke av de to sekvensene som er et resultat av avhengige forsøk og finn ut om q er større enn eller mindre enn 0.5.

```
01010101010010010101010101010001101010101010101010
10101010101010101010101001011010101010101010101010

00001011001001010110110110111100110011010111100011
11100111101101101110010000101110000011110111110011
```

Tilleggsoppgave 16 (Matlab oppgave)
Du vil i denne oppgave få bruk for kommandoen `poisspdf` for å beregne sannsynligheter i Poissonfordelingen.

En av de tidlige studiene av radioaktivitet, var Rutherford og Geigers studie i 1910 av α-stråling. Nær en plutoniumskilde satte de opp en liten skjerm. De registrerte så hvor mange α-partikler som traff skjermen i løpet av 8 minutter. Dette forsøket ble gjentatt mange ganger, slik at de til sammen fikk data fra 2608 åtte-minutters perioder.

Dataene ble som gitt i tabellen:

Antall α-partikler	0	1	2	3	4	5	6	7	8	9	10	≥ 11
Antall perioder	57	203	383	525	532	408	273	139	45	27	10	6

(a) Bestem gjennomsnittlig antall α-partikler per periode. (I 6 perioder var det 11 α-partikler eller flere. Du kan regne som om det var nøaktig 11 α-partikler i disse periodene.)

Sett λ lik gjennomsnittet du fant i punkt (a).

(b) Undersøk hvor godt en Poissonfordeling med parameter λ beskriver dataene.

Tilleggsoppgave 17

I januar 1992 hadde Dagsrevyen et stort oppslag om at det i Sømna kommune på Helgelandskysten hadde vært unormalt mange tilfelle av hjernesvulst i 1991. Dette ble satt i sammenheng med Tsjernobyl-ulykken noen år tidligere. Saken ble fulgt opp i avisene. Dagbladets overskrift over hele forsiden dagen etter lød "Bygd i kreftsjokk". Og rett etter skrev samme avis "Politikere er skremt. Krever Stortings-orientering om krefttilfellene i Sømna".

Bakgrunnen for saken var at det i 1991 ble registrert 3 tilfelle av hjernesvulst i Sømna. Dette er mange for en så liten kommune. Etter Kreftregisterets statistikk skulle Sømna i gjennomsnitt bare få 0.16 tilfeller av hjernesvulst hvert år hvis kreftrisikoen der var som ellers i Norge.

Det ble satt i gang større undersøkelser i Sømna. Men etter en tid ble det konkludert med at radioaktivitet ikke var en årsak til krefttilfellene, og at det hele nok var et resultat av tilfeldigheter. For eksempel stod følgende i Aftenposten i oktober 1992: "Tilfellet Sømna ser foreløpig ut som en ren tilfeldighet, sier Frøydis Langmark, Kreftregisterets leder."

Vi skal i denne oppgaven se nærmere på Sømna-saken.

(a) La X være antall tilfelle av hjernesvulst i løpet av ett år for en kommune av Sømnas størrelse og med samme alders- og kjønnsfordeling som Sømna. Vi antar at X er Poissonfordelt med parameter 0.16 slik det vil være hvis kreftrisikoen i kommunen er som ellers i landet. Bestem sannsynligheten for at X skal være minst tre. Hva betyr denne sannsynligheten i forhold til Sømna-saken?

Før en trekker forhastede konklusjoner av resultatet i punkt a, må en tenke over hvorfor tallene fra Sømna vakte oppsikt. Dette skjedde nettopp fordi det ble registrert uvanlig mange tilfelle av hjernesvulst det året. Alle de kommunene der det ikke skjer noe oppsiktsvekkende, er det ingen som bryr seg om. I stedet for sannsynligheten i punkt a, er det derfor mer relevant å beregne sannsynligheten for at vi en gang i blant i en eller annen kommune vil observere noe så påfallende som det som skjedde i Sømna i 1991.

La oss som et regneeksempel tenke oss 50 kommuner av samme størrelse og med samme alders- og kjønnssammensetning som Sømna, og anta at hver av de observeres over 10 år. Tilsammen blir dette 500 "kommuneår".

(b) Hva er sannsynligheten for at vi i minst ett av de 500 "kommuneårene" vil observere minst tre tilfelle av hjernesvulst. Hva betyr denne sannsynligheten i forhold til Sømna-saken?

Tilleggsoppgave 18 (Matlab oppgave)
Før du begynner på denne oppgaven må du ha skaffet deg notatet «MATLAB for STK1100» som er tilgjengelig under MATLAB linken fra hjemmesiden til STK1100.

(a) Et menneskes vitalkapasitet er den totale mengde luft en klarer å blåse ut etter å ha fylt lungene maksimalt. I en undersøkelse ble vitalkapasiteten målt for 12 voksne personer med følgende resultat (i liter):

3.9 5.6 4.1 4.2 4.0 3.6 5.9 4.5 3.6 5.0 2.9 4.3

Les disse dataene inn i MATLAB enten ved å skrive dem rett inn i en variabel, eller ved å skrive dataene inn i en tekstfil som leses inn i en variabel med kommandoen `load`.

(b) Beregn gjennomsnitt, empirisk median, empirisk standardavvik og kvartildifferanse til dataene i punkt (a). Pass på at du forstår hva MATLAB har beregnet!

(c) Lag et histogram av dataene i punkt (a). Hvis du har dataene liggende i variabelen `x`, kan du lage histogram med `histogram(x)`.

MATLAB bruker alltid ti (like brede) intervaller med mindre man selv spesifiserer et annet antall intervaller. Man kan sette antall intervaller ved å bruke `histogram(x,m)`, hvor `m` er antall intervaller. Forsøk med forskjellige verdier for `m` og se hvordan histogrammet endrer seg. Hvilken verdi for `m` synes du er best?

(d) Enkelte ganger kan det være mer hensiktsmessig å angi selve intervallene framfor å angi antall intervaller. Dette gjøres med `histc`. Siden `histc` kan ikke plotte selv, må vi først bruke `histc` for å beregne hvor mange observasjoner som faller i hvert intervall, og så bruke `bar` for å plotte.

Her følger et eksempel hvor det er brukt to intervaller – ett fra 2–4 og ett fra 4–6.

```
edges = [2, 4, 6];          % definer intervallene
k = histc(x, edges);        % teller opp antall i hvert intervall
bar(edges, k, 'histc');     % plotter histogrammet
figure(gcf);                % faar fram figurvinduet
```

(Den siste linjen kan brukes når figurvinduet er gjemt bak et annet vindu. Kommandoen `gcf` står for «get current figure», og den returnerer nummeret på figuren

som det plottes i for øyeblikket, mens `figure` sørger for å bringe den angitte figuren fram.)

Ta utgangspunkt i eksemplet over. Bruk nå intervallene 2–3, 3–4, 4–5 og 5–6, og plott histogrammet.

(*e*) Lag endelig et normert histogram. At et histogram er normert vil si at arealet under histogrammet er lik én. Skalér verdiene i `k` som nødvendig og plott på nytt.

Tilleggsoppgave 19

I tabellen nedenfor er det gitt en oversikt over fødsler, tvillingfødsler og trillingfødsler i Norge i femårsperioder fra 1951-1955 til 2006-2010.

Tabell 1: Fødte i Norge i ulike tidsperioder[a]

År	Levendefødte			Flerfødsler		
Årsgjennomsnitt	I alt	Gutter	Jenter	I alt	Tvillingfødsler	Trillingfødsler[b]
1951-55	62478	32182	30296	796	787	9
1956-60	63021	32374	30647	738	731	7
1961-65	63989	32992	30997	708	700	8
1966-70	66697	34368	32329	670	663	7
1971-75	61393	31487	29906	572	568	4
1976-80	51744	26619	25125	498	494	4
1981-85	50660	26030	24629	503	495	8
1986-90	56862	29154	27708	653	634	19
1991-95	60196	30993	29202	845	821	24
1996-00	59522	30598	29043	981	957	24
2001-05	56459	28925	27534	1051	1034	17
2006-10	60150	30885	29265	1036	1021	15

[a] Tabellen er en redigert versjon av Tabell 71 i Statistisk årbok 2013 (`www.ssb.no/aarbok`).

[b] Medregnet firling- og femlingfødsler.

(*a*) Beregn de relative frekvensene av tvillingfødsel for de tolv femårsperiodene. Hvordan varierer disse? (I tabellen er det oppgitt antall fødte barn, ikke antall svangerskap. Siden forskjellen på antall svangerskap og antall fødte barn er liten, kan du se bort fra dette problemet.)

(*b*) Gjenta punkt (*a*) for trillingfødsler.

(*c*) Kunstig befruktning, hvor flere befruktede egg settes inn i kvinnens livmor, ble innført i Norge på 1980-tallet. Diskuter resultatene i punktene (*a*) og (*b*) i lys av dette.

(*d*) På grunn av forholdet nevnt i punkt (*c*), kan vi ikke bruke dataene for alle femårsperiodene til å bestemme sannsynligheten for "naturlig" tvilling- og trillingfødsel. Forklar hvordan du kan bruke de relative frekvensene for de sju første femårsperiodene

til å bestemme sannsynligheten for "naturlig" tvilling- og trillingfødsel. Hva blir sannsynlighetene?

Tilleggsoppgave 20
La Z_1 og Z_2 være uavhengige normalfordelte variable med forventning 0 og varians 1. Definer $X_1 = \mu_1 + \sigma_1 Z_1$ og $X_2 = \mu_2 + \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_2)$ for konstanter μ_1, μ_2 og σ_1, $\sigma_2 > 0$ og $-1 < \rho < 1$. Da er (X_1, X_2) bivariate normal fordelt med tetthetsfunskjon $f(x_1, x_2)$ som gitt på side 258 i Devore and Berk [2012].

(*a*) Vis at $\rho = 0$ impliserer at X_1 og X_2 er uavhengige.

(*b*) Anta nå$E(X_1 = E(X_2) = 0$, $\mathsf{Var}(X_1) = \mathsf{Var}(X_2) = 1$ og $\mathsf{Corr}(X_1, X_2) = \rho$. Den simultane tetthetsfunkskjon til (X_1, X_2) er da gitt ved

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{-x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)} \right\}$$

Vis ved å integrere ut x_1 (eller x_2) at marginalfordelingen til X_1 (eller X_2) er normal.

(*c*) Vis at hvis $\mathsf{Var}(X_1) = \mathsf{Var}(X_2)$, så er $Y_1 = X_1 + X_2$ uavhengig av $Y_2 = X_1 - X_2$.

Hint: Bruk strukturen og resultatet fra deloppgave (*a*).

(*d*) Bestem c (en konstant) slik at $\mathsf{Var}(X_1 + cX_2)$ blir minst mulig.

(*e*) La $X \sim N(\mu, \sigma^2)$ og $(Y|X = x) \sim N(x, \tau^2)$ Forklar hvorfor Y er normalfordelt og bruk reglene om total forventning og varians på side 261 i Devore and Berk [2012] til å finne forventning og varians for Y.

Hint: Bruk simultanfordelingen til (X, Y) for å finne fordelingen til Y.

(*f*) La Z_1 og Z_2 være uavhengige normalfordelte variable med forventning 0 og varians 1. Definer $X_1 = Z_1$ og $X_2 = \rho Z_1 + \sqrt{1-\rho^2} Z_2$ og vis at $\mathsf{Corr}(X_1^2, X_2^2) = \rho^2$.

Hint: Bruk (eller vis) at $EZ_1 Z_2 = EZ_1^3 Z_2 = 0$ og $EZ_i^4 = 3$ for $i = 1, 2$

Tilleggsoppgave 21
Denne oppgaven viser hvordan en med utgangspunkt i uniformt fordelte variable kan generere eksponensielt fordelte variable, og hvordan en fra disse igjen kan generere en Poissonprosess (se side 149-151 i Devore and Berk [2012]).

(*a*) Trekk 1000 variable fra den uniforme fordelingen $U(0, 1)$ med kommandoen

```
U = unifrnd(0,1,[1,1000]);
```

og lag et histogram over variablene. Kommenter histogrammets utseende.

(*b*) Vis at hvis U er fordelt $U(0,1)$, så vil $X = -\log(1-U)/\lambda$ være eksponensielt fordelt med parameter λ. Gitt vektoren `U` med uniformt fordelte variable fra forrige punkt, kan du derfor få eksponensielt fordelte variable med $\lambda = 1$ ved

```
X = -log(1-U);
```

Lag et histogram for `X` og kommenter histogrammets utseende.

(*c*) Fra eksponsialfordelte variable kan vi generere en Poisson-prosess ved å la de eksponensielt fordelte variablene være ventetidene mellom begivenheter i Poisson-prosessen (dette er en egenskap ved Poisson prosessen som vi vil ta som gitt her).

Du kan derfor generere en Poisson-prosess med $\lambda = 1$ ved

```
T = cumsum(X);
```

der `T` nå inneholder tidene for de 1000 første begivenhetene. Overbevis deg selv om at dette blir riktig.

(*d*) Vi skal nå se på hvor mange begivenheter i Poisson-prosessen som faller i gitte intervaller. Vi deler opp tiden fra 0 til 800 i 200 like lange intervaller med

```
t = 0:4:800;
```

Antall begivenheter som inntraff i hvert intervall kan nå telles opp med

```
N = histc(T,t);
```

(Vi begrenser oss til tiden frem til 800 for å sikre oss at Poisson-prosessen vi har simulert ikke har stoppet opp før.)

(*e*) Vi teller opp hvor mange intervaller som hadde ingen begivenheter, hvor mange intervaller som hadde én begivenhet, osv. opp til ti begivenheter, og plotter ved kommandoene

```
x=0:10;
y = histc(N,x);
bar(x,y);
```

Langs x-aksen har vi antall begivenheter og langs y-aksen antall intervaller.

(*f*) Vi vil til slutt sammenligne de relative frekvensene for den simulerte prosessen med punktsannsynlighetene i en Poisson-fordeling med parameter lik 4.

Lag en vektor med relative frekvenser og en annen vektor som inneholder punktsannsynlighetene med

```
relfr = y/200;
punktsanns = poisspdf(x,4);
```

Hvordan er samsvaret mellom verdiene i de to vektorene?

Det kan være enklere å sjekke samsvaret ved å plotte de relative frekvensene og punktsannsynlighetene i det samme plottet. For å få det til må begge vektorene være kolonne-vektorer og 'limes' sammen i en matrise med to kolonner. Dette gjøres med

```
x = x(:);                      %gjor om til kolonne-vektor
relfr = relfr(:);              %gjor om til kolonne-vektor
punktsanns=punktsanns(:);      %gjor om til kolonne-vektor
sanns = [relfr,punktsanns];  % lag en matrise med to kolonner
bar(x,sanns);
```

Tilleggsoppgave 22
På grunnlag av Statistisk Sentralbyrås dødelighetsstatistikk (for 1993) har vi at sannsynligheten er:

- 98.6% for at en 50 år gammel kvinne skal oppleve sin 55 årsdag
- 97.9% for at en 55 år gammel kvinne skal oppleve sin 60 årsdag
- 96.5% for at en 60 år gammel kvinne skal oppleve sin 65 årsdag

(a) Hva er sannsynligheten for at en 50 år gammel kvinne skal bli minst 65 år?

En 50 år gammel kvinne ønsker å kjøpe en pensjonsforsikring som gir henne en utbetaling på 100 000 kroner hvis hun fyller 65 år. Hvis hun dør før fylte 65 år utbetales intet.

(b) Forsikringsselskapet benytter en rentefot på 3% p.a. Forklar hvorfor nåverdien av pensjonsutbetalingen er $100\,000/1.03^{15}$ kroner hvis hun blir minst 65 år og 0 kroner hvis hun dør før hun fyller 65 år. Nåverdien er altså $100\,000\,X/1.03^{15}$ kroner, hvor X er lik 1 hvis hun blir minst 65 år og X er lik 0 ellers. (Vink: Husk at nåverdien av 100 000 kroner er det beløpet vi må sette i banken i dag for å få 100 000 kroner om 15 år når vi beregner renter og renters rente med 3% p.a.)

(c) Bestem $E(X)$ og forventet nåverdi av pensjonsutbetalingen. (Vink: Husk resultatet i punkt a.)

Anta at kvinnen betaler pensjonsforsikringen kontant på sin femtiårsdag.

d) Forklar hvorfor forventet nåverdi av pensjonsutbetalingen er en "rettferdig premie" (= pris) når vi ser bort fra selskapets omkostninger og fortjeneste. (Vink: Tenk deg at selskapet tegner mange pensjonsforsikringer av den typen vi har beskrevet. Hva blir gjennomsnittlig inn- og utbetaling pr. polise?)

Tilleggsoppgave 23
La $U \sim U(0,1)$ være uniform på $(0,1)$. Hvis X har sannsynlighetsfordelings funksjon F med entydig invers F^{-1} (slik at $F^{-1}(F(x)) = x$), vis at $Y = F^{-1}(U)$ har samme fordeling som X.

Hint: Hvorfor er det nok å vise at $P(Y \leq x) = P(X \leq x)$?

Tilleggsoppgave 24
Anta X og Y er to diskrete stokastiske variable med simultan punktsannsynlighet gitt ved

$$p(x,y) = \begin{cases} \frac{x+y}{32} & x = 1,2, y = 1,2,3,4 \\ 0 & \text{ellers} \end{cases}$$

(*a*) Da er $P(X = 2)$ lik
A: $\frac{7}{16}$ B: $\frac{1}{2}$ C: $\frac{9}{16}$ D: $\frac{2+y}{32}$ E: $\frac{1+y}{32}$

(*b*) $P(Y = 2X)$ er
A: $\frac{1}{4}$ B: $\frac{5}{32}$ C: $\frac{1}{2}$ D: $\frac{3}{32}$ E: $\frac{9}{32}$

(*c*) Sannsynligheten for at $X \leq Y$ er
A: $\frac{1}{2}$ B: $\frac{3}{32}$ C: $\frac{29}{32}$ D: $\frac{3}{16}$ E: $\frac{3}{8}$

Tilleggsoppgave 25
Anta X og Y er to kontinuerlig stokastiske variable med simultan sannsynlighetstetthet 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*) Den marginale fordeling til X er
A: 1 B: $x + \frac{1}{2}$ C: $\frac{1}{2}$ D: $x + \frac{1}{2}x^2$ E: $y + \frac{1}{2}$

(*b*) $E(Y)$ er
A: $\frac{1}{3}$ B: $\frac{x}{2}$ C: $\frac{7}{12}$ D: $\frac{3}{8}$ E: y

(*c*) Sannsynligheten for at $X + Y > 1$ er
A: $-\frac{1}{2}$ B: $\frac{5}{8}$ C: $\frac{2}{3}$ D: $\frac{3}{4}$ E: $\frac{3}{8}$

(*d*) $E(X + Y)$ er
A: 1 B: 4 C: $\frac{7}{10}$ D: $\frac{7}{6}$ E: y

Tilleggsoppgave 26
Anta X og Y er to kontinuerlig stokastiske variable med simultan sannsynlighetstetthet gitt ved

$$f(x,y) = \begin{cases} kxy^2 & 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0 & \text{ellers} \end{cases}$$

(*a*) Da er k lik
A: $\frac{5}{16}$ B: $\frac{3}{16}$ C: $\frac{1}{16}$ D: $\frac{1}{4}$ E: $\frac{7}{16}$

(*b*) Den marginale fordeling til X er
A: 1 B: $x + \frac{1}{2}$ C: $\frac{1}{2}$ D: $x + \frac{1}{2}x^2$ E: $y + \frac{1}{2}$

(*c*) $E(Y)$ er
A: 1 B: $\frac{5}{12}$ C: $\frac{3}{2}$ D: $\frac{7}{12}$ E: y

Tilleggsoppgave 27 (Eksempel 7.8)
Betrakt eksempel 7.8 fra Devore and Berk [2012] hvor vi nå vil gjøre et simuleringseksperiment for å studere de ulike estimatorer når vi antall X_i-ene kommer fra forskjellige fordelinger.

- Hvis $\sigma = 1$ for den normale fordeling, hvilken verdi må vi ha for c i den uniforme fordeling slik at variansen blir like i de to fordelinger?
- Vis at hvis $U \sim \text{Uniform}(0, 1)$ så er $X = \theta + \tan(\pi(U - 0.5))$ Cauchy fordelt.
- Bruk MATLAB skriptet `example7_8.m` (som er tilgjengelig på hjemmesiden) til å simulere for de ulike fordelinger. Kommenter resultatene.

Tilleggsoppgave 28
En regnskapsfører ønsker å forenkle bokføring ved å runde av beløpene til nærmeste heltall; for eksempel blir 99.51 og 100.48 begge ført som 100 kroner. Vi skal undersøke den totale feilen for 100 beløp. En måte å løse dette på er ved å modellere avrundingsfeilen som en stokastisk variabel. Vi skal anta at hver av de 100 feilene, la oss si $X_1, ..., X_{100}$, er uavhengig og uniform fordelt på $[-0.5, 0.5]$.

(*a*) Finn forventning og varians til en X_i.

(*b*) Bruk Chebyshev's ulikhet til å beregne en øvre grense for sannsynligheten $P\{|X_1 + \cdots + X_{100}| > 10\}$ at den totale feilen skal overskride 10 kroner.

Tilleggsoppgave 29
I et gitt land vil en andel p av velgerne stemme på kandidat G og en andel $1-p$ stemme på kandidaten B. I en meningsmåling blir flere velgere spurt om hvordan de skal stemme. La $X_i = 1$ hvis person nummer i stemmer på kandidat G og 0 ellers. En modell for målingene er at folk blir valgt ut og intervjuet slik $X_1, X_2, ...,$ er uavhengige og har en Bernoulli fordeling med parameter p.

(*a*) Vi skal bruke $\overline{X}_n$ for å predikere p. Hvis vi bruker Chebyshev's ulikhet, hvor mange må vi spørre (dvs hvor stor må n være) slik at vi kan med sannsynlighet minst 0.90 si at $\overline{X}_n$ er innenfor ± 0.2 av sann p.

Hint: Løs dette for $p = 1/2$ og bruk at $p(1-p) \leq 1/4$ for alle $0 \leq p \leq 1$.

(*b*) Svar på oppgave (*a*) men finn n slik at $\overline{X}_n$ er innenfor ± 0.1 av sann p.

(*c*) Svar på oppgave (*a*) men vi skal nå oppgi svaret med sannsynlighet minst 0.95.

(*d*) Hvis $p > 1/2$ så vinner kandidat G og hvis $\overline{X}_n > 1/2$ predikerer vi at G vil vinne. Finn det minste utvalget n slik at sannsynligheten for at prediksjonene blir riktig er minst 0.90 hvis sann $p = 0.6$.

Tilleggsoppgave 30
Vi skal her vise en mer generell utgave av store talls lov (the law of large numbers). La $X_1X_2, ...$ være en sekvens av uavhengige stokastike variable med $E(X_i) = \mu_i$ (som alle er begrenset) og $\mathsf{Var}(X_i) = \sigma_i^2$, for $i = 1, 2, ...$ Anta at $0 \leq \sigma_i^2 \leq M \leq \infty$, for all i, og la a være et vilkårlig positivt tall.

(*a*) Bruk Chebyshev's ulikhet til å vise at

$$P(\overline{X}_n - \tfrac{1}{n}\sum_{i=1}^{n}\mu_i \geq a) \leq \frac{1}{n^2a^2}\sum_{i=1}^{n}\sigma_i^2.$$

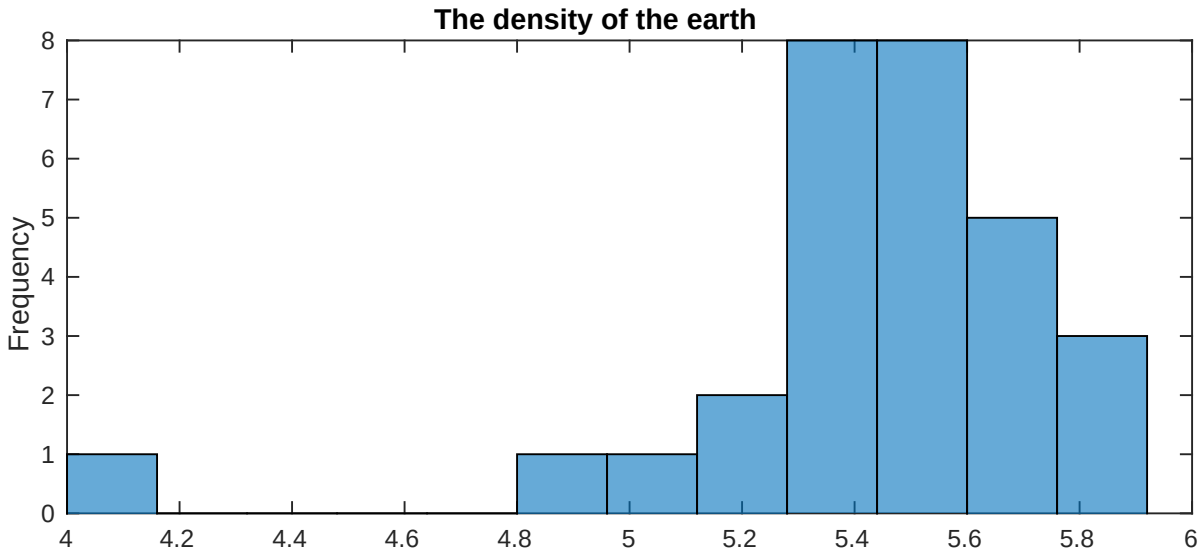
Husk: Chebyshev's ulikhet forteller oss at for enhver stokastisk variabel Y (med begrenset forventning) og konstant $a > 0$ så er $P(|Y - E(Y)| \geq a) \leq \mathsf{Var}(Y)/a^2$.

(*b*) Bruk (*a*) til å argumentere for at

$$\lim_{n\to\infty} P\left\{|\overline{X}_n - \tfrac{1}{n}\sum_{i=1}^{n}\mu_i| \geq a\right\} = 0$$

og sjekk at "klassisk" store talls lov [på side 303 i Devore and Berk, 2012] er et spesialtilfelle.

Tilleggsoppgave 31
I 1798 utførte Henry Cavendish 29 eksperimenter med en "torsion balance" for å estimere tettheten til jordkloden, her rapportert som et multiplum av tettheten til vann, se figur nedenfor.



Hvis vi lar ρ_{earth} være "sann" tetthet har vi følgende to naturlige estimatorer:

$$\hat{\rho}_{\text{earth}} = \overline{x}_n = 5.42 \text{ og } \tilde{\rho}_{\text{earth}} = \tilde{x}_n = 5.46,$$

hvor $x_1, ..., x_{29}$ er de 29 målingene gjort av Henry Cavendish, se `Cavenish_density.m` og der $\overline{x}_n$ og $\tilde{x}_n$ er gjennomsnitt og median, henholdsvis.

(a) Basert på histogrammet i figuren ovenfor, hvilken metode foretrekker du?

(b) Bruk koden i `Cavendish_density.m` til å utføre ikke-parametrisk bootstrapping for å undersøke standardavviket og forventningsskjevheten for de to estimatorene.

(c) Gjenta punkt (b), men bruk denne gangen parametrisk bootstrapping (se punkt (c)) i `Cavendish_density.m`) hvor du antar at eksprimentet er normalfordelt. Sammenlign og kommenter forskjeller, se spesielt også på forskjellen til histogrammene.

Tilleggsoppgave 32
Overlevelsestiden til 9 mus er gitt nedenfor. Datasettet er hentet fra Efron and Tibshirani [1993].

52 104 146 10 51 30 40 27 46

La X_i være i-te overlevelsestid og anta $X_1, \ldots, X_9$ er uavhengige identisk fordelte med sannsynlighetfordeling F. Vi er interessert i både forventning og median for fordelingen F. De naturlige estimater er empirisk gjennomsnitt og median:

$$\overline{x} = 56.22, \quad \text{median}(x_1, \ldots, x_9) = 46.0.$$

Det empiriske standardavviket til $x_1, \ldots, x_9$ er 42.48 som indikerer en stor variabilitet i dataene. Vi er interessert i egenskapene til disse estimatorene.

(*a*) Utfør ikke-parametrisk Bootstrapping til å finne standard feil og forventningsskjevhet for de to estimatorene. Kommenter resultatene, spesielt at det ser ut som forventningsskevheten til gjennomsnittet er svært lavt.

Hint: Se på kommandoene for Bootstrapping i ekstranotatet om boostrapping og stokastisk simulering.

(*b*) Lag histogram av dine simuleringer. Kommenter. Prøv spesielt å forklare den merkelige formen på histogrammet relatert til medianen.

(*c*) Anta nå at X_i ene er log-normal fordelte. Utfør en parametrisk Bootstrapping. Kommenter forskjeller fra resultatene du har fått tidligere.

Tilleggsoppgave 33
La $X_1, ..., X_n$ være uavhengige kontinuerlige tilfeldige variable med CDF $F(x)$. Den *empirisk kumulative fordelingsfunksjon* er gitt ved

$$\widehat{F}(x) = \frac{1}{n}(\#x_i \leq x) \tag{1}$$

(*a*) Vis at $\widehat{F}(x)$ er en forventningsrett estimator for $F(x)$ for alle verdier av x.

(*b*) Finn $V(\widehat{F}(x))$ og $\sigma_{\widehat{F}(x)}$. Bruk dette til å vise at $\widehat{F}(x)$ er en konsistent estimator for $F(x)$.

(*c*) Forklar hvorfor $\widehat{F}(x)$ er en diskret fordeling og vis at hvis $X^* \sim \widehat{F}(x)$ så er $P(X^* = x_i) = \frac{1}{n}$ for $i = 1, ..., n$.

(*d*) Diskuter styrker og svakheter ved $\widehat{F}(x)$.

Referanser

Jay L Devore and Kenneth N Berk. *Modern Mathematical Statistics with Applications*. New York, NY: Springer New York, 2012.

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.