# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK1100 — Sannsynlighetsregning og statistisk modellering

Eksamensdag: Fredag 3. juni 2022

Tid for eksamen: 09.00 – 13.00

Oppgavesettet er på 4 sider.

Vedlegg: Ingen

Tillatte hjelpemidler: Godkjent kalkulator<br>Formelsamling for STK1100

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

## Oppgave 1

Line har to sekssidede terninger. Terning 1 er en jukseterning, som er tyngre på den motsatte sida av seksersida, slik at sannsynligheten for å få en sekser når en kaster terningen er $\frac{1}{5}$. Terning 2 er en vanlig, balansert terning der alle sidene har lik sannsynlighet.

Line velger én av terningene helt tilfeldig (med sannsynlighet 0.5 for hver terning) og kaster den.

**a**

Hva er sannsynligheten for at det blir en sekser?

Du får ikke se resultatet av terningkastet, men du får vite at det *ikke* ble seks.

**b**

Hva er sannsynligheten for at det var Terning 1 som ble kastet?

## Oppgave 2

I et verneområde i Irland, har en har prøvd å måle bestanden til irske steinkobber (en type seldyr). La $x_1, \ldots, x_n$ være antall steinkobber som er telt i $n$ påfølgende sesonger. Disse antas å være observerte verdier av de stokastiske variablene $X_1, \ldots, X_n$, som er uavhengige og identisk fordelt. Siden dette er tellevariabler, er Poisson-fordelingen et naturlig valg, og det er også modellen vi antar i denne oppgaven, nærmere bestemt at alle $X_i$, $i = 1, \ldots, n$ er Poisson-fordelt med parameter $\lambda$. Punktsannsynligheten til

*(Fortsettes på side 2.)*

$X_i$ er da gitt ved

$$p(x) = \frac{\lambda^x}{x!}e^{-\lambda},\ x = 0, 1, 2, \ldots,$$

med momentgenererende funksjon

$$M_X(t) = e^{\lambda(e^t-1)}.$$

**a**

Vis at forventningen og variansen til $X_i$ begge er lik $\lambda$ ved å bruke den momentgenererende funksjonen.

La $Y$ være totalt antall steinkobber over alle sesongene, dvs. $Y = \sum_{i=1}^{n} X_i$.

**b**

Finn forventningen og variansen til $Y$.

**c**

Vis at også $Y$ er Poisson-fordelt.
(*Hint:* Her kan du også få bruk for den momentgenererende funksjonen fra tidligere i oppgaven.)

**d**

Sett opp likelihood- og log-likelihood-funksjonen, og vis at maksimum likelihood-estimatoren for $\lambda$ er $\widehat{\lambda}_{mle} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Begrunn at momentestimatoren $\widehat{\lambda}_{mom}$ er den samme som maksimum-likelihood estimatoren.

Du får oppgitt at det er telt steinkobber over $n = 43$ sesonger, med tilsvarende gjennomsnitt $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 20.37$.

**e**

Vis at standardfeilen $\sigma_{\widehat{\lambda}}$ til maksimum likelihood-estimatoren for $\widehat{\lambda}_{mle}$ (og momentestimatoren $\widehat{\lambda}_{mom}$) er $\sqrt{\lambda/n}$. Oppgi maksimum likelihood-estimatet for $\lambda$ basert på opplysningene over, og gi et estimat for standardfeilen.

Du får i tillegg oppgitt at den empiriske variansen i antall steinkobber per sesong er $s^2 = 448.14$.

**f**

Sett i lys av resultatet fra Oppgave a, syns du Poisson-fordelingen virker som en passende modell for steinkobbetellingene?

*(Fortsettes på side 3.)*

# Oppgave 3

La $X$ være en normalfordelt variabel med forventning $\mu$ og varians $\sigma^2$, som vi også kan skrive $X \sim N(\mu, \sigma^2)$, slik at tettheten til $X$ er gitt ved

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}.$$

**a**

Vis at fordelingen til $Y = \left(\frac{X-\mu}{\sigma}\right)^2$ er *gamma* $\left(\frac{1}{2}, 2\right)$ med formparameter $\alpha = \frac{1}{2}$ og skalaparameter $\beta = 2$, som er det samme som kjikvadratfordelingen med 1 frihetsgrad, dvs. at $Y$ har tettheten

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y > 0 \\ 0, & \text{ellers.} \end{cases}$$

Anta at du har observasjonene $x_1 \ldots, x_n$ av de stokastiske variablene $X_1, \ldots, X_n$, som er uavhengige og identisk normalfordelt med forventning $\mu$ og varians $\sigma^2$. Videre antar vi at $\mu$ er kjent, mens $\sigma^2$ er ukjent, slik at vi ønsker å estimere den. Nærmere bestemt, vil vi bruke den forventningsrette

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2.$$

**b**

Vis at $n\frac{\hat{\sigma}^2}{\sigma^2}$ kan skrives som $\sum_{i=1}^n Y_i$ med $Y_i = \left(\frac{X_i-\mu}{\sigma}\right)^2$, og bruk det sammen med resultatet fra Oppgave a til å argumentere for at $n\frac{\hat{\sigma}^2}{\sigma^2} \sim$ *gamma* $\left(\frac{n}{2}, 2\right)$, som er det samme som kjikvadratfordelingen med $n$ frihetsgrader.
(*Hint:* Husk at fordelingen til en sum av uavhengige gamma$(\alpha_i, \beta)$-fordelte variabler med samme skalaparameter $\beta$ er gamma$(\sum_i \alpha_i, \beta)$-fordelt).

**c**

Bruk resultatene fra Oppgave b til å lage et 95% konfidensintervall for $\sigma^2$ når du får vite at $\chi^2_{n,0.975}$ og $\chi^2_{n,0.025}$ er henholdsvis 2.5%- og 97.5%-persentilene i kjikvadratfordelingen med $n$ frihetsgrader (som er det samme som *gamma* $\left(\frac{n}{2}, 2\right)$-fordelingen).

# Oppgave 4

De to variablene $X$ og $Y$ er binormalfordelt med parametere $\mu_1$, $\mu_2$, $\sigma_1$, $\sigma_2$ og $\rho$. De har dermed simultan sannsynlighetstetthet

$$f_{X,Y}(x, y) = $$
$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma^1}\right)^2 + \left(\frac{y-\mu_2}{\sigma^2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma^1}\right)\left(\frac{y-\mu_2}{\sigma^2}\right)\right)}, \quad (x, y) \in \mathbb{R}^2.$$

*(Fortsettes på side 4.)*

Her er $\mu_1 = \mathrm{E}(X)$, $\mu_2 = \mathrm{E}(Y)$, $\sigma_1^2 = \mathrm{V}(X)$, $\sigma_2^2 = \mathrm{V}(Y)$ og $\rho = \mathrm{Corr}(X, Y)$.

Et kjennetegn ved binormalfordelingen er at marginalfordelingen til både $X$ og $Y$ er normalfordelinger, nærmere bestemt er $X \sim N(\mu_1, \sigma_1^2)$ og $Y \sim N(\mu_2, \sigma_2^2)$ (dette kan du bruke uten å vise det).

Videre er det velkjent at dersom to stokastiske variabler $V$ og $W$ er uavhengige, så er korrelasjonen $\mathrm{Corr}(V, W)$ mellom dem lik 0. Det motsatte er generelt ikke tilfelle; det vil si at korrelasjonen $\mathrm{Corr}(V, W) = 0$, selv om $V$ og $W$ er avhengige. Binormalfordelingen er imidlertid et unntak, og er slik at dersom korrelasjonen er 0, så er de to variablene uavhengige.

**a**

Vis at når $\rho = 0$, dvs. at korrelasjonen mellom $X$ og $Y$ er lik 0, så er $X$ og $Y$ uavhengige.

Et annet kjennetegn ved den binormale fordelingen er at den betingede fordelingen til $Y$ gitt $X = x$ er normalfordelingen med forventning $\mu_2 + \rho\frac{\sigma_2}{\sigma_2}(x - \mu_1)$ og varians $\sigma_2^2(1 - \rho^2)$, dvs. at

$$[Y|X = x] \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_2}(x - \mu_1), \sigma_2^2(1 - \rho^2)\right). \tag{1}$$

En alternativ måte å modellere sammenhengen mellom $Y$ og $X$ på, betinget på en gitt verdi av $x$ av $X$, er en lineær regresjonsmodell, som altså her er en modell for $[Y|X = x]$, og er gitt ved

$$Y = \beta_0 + \beta_1 x + \epsilon, \quad \varepsilon \sim N(0, \psi^2).$$

**b**

Argumentér for at $[Y|X = x]$ i den lineære regresjonsmodellen over er normalfordelt og vis at

$$\mathrm{E}(Y|X = x) = \beta_0 + \beta_1 x \quad \text{og} \quad \mathrm{V}(Y|X = x) = \psi^2.$$

**c**

Forklar at den betingede fordelingen til $[Y|X = x]$ i den binormale fordelingen, altså modellen i (1), er en lineær regresjonsmodell, og oppgi hva $\beta_0$, $\beta_1$ og $\psi^2$ er i denne modellen.