Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

# TMA4245 Statistikk
Vår 2017

**Innlevering 2**

Dette er den andre av to innleveringer i blokk 1. Denne øvingen skal oppsummere pensum forelest i uke 4-7. Øvingen handler om forventningsverdi, varians og diskrete og kontinuerlige sannsynlighetsfordelinger. Hver deloppgave teller 10 %.

## Oppgave 1

Høsten 2004 var det tellende skriftlig midtveiseksamen i Statistikk. Denne var gitt i form av en flervalgsoppgave (multiple choice) bestående av $n = 20$ spørsmål som alle har $m$ svaralternativer. Studentene måtte velge et svaralternativ for hvert spørsmål (det var således ikke lov å svare blankt på et spørsmål). For å bestå midtveiseksamen måtte minst 8 spørsmål være korrekt besvart.

Ole lurer på om han skal la være å lese til midtveiseksamen og heller velge tilfeldige svaralternativer på alle spørsmålene (han vil da ikke engang lese oppgaveteksten før han svarer). Før han bestemmer seg, ber han en studiekamerat regne ut hvor stor sannsynlighet han da har for å bestå midtveiseksamen.

La $X$ være antall korrekte svar Ole får på de $n = 20$ spørsmålene.

**a)** Forklar hvorfor det er rimelig å anta at $X$ er binomisk fordelt med $n = 20$ og $p = \frac{1}{m}$. (Det er ikke tilstrekkelig å skrive opp de generelle forutsetningene for en binomisk fordeling, betingelsene må relateres direkte til situasjonen som er beskrevet.)

Finn sannsynligheten for at Ole består midtveiseksamen hvis han velger å svare tilfeldig på alle spørsmålene, dvs. $P(X \geq 8)$, når antall svaralternativer er $m = 2$. Finn også $P(X \geq 8)$ for $m = 4$ og $m = 5$.

Hva blir forventet antall korrekte svar, dvs. $E(X)$, når $m = 2, 4, 5$?

Vi benytter videre at hvert spørsmål har $m = 5$ svaralternativer.

Ole bestemmer seg for å bruke midtsemesteruken til å jobbe med statistikk. Dagen før midtveiseksamen setter han seg ned og deler pensum inn i tre kategorier; de delene av pensum han synes han kan godt, de han kan middels godt, og de han kan dårlig.

Vi ser på et tilfeldig valgt spørsmål fra midtveiseksamen og definerer følgende fire hendelser:

$G$ : spørsmålet er fra den delen av pensum som Ole kan godt,
$M$ : spørsmålet er fra den delen av pensum som Ole kan middels godt,
$D$ : spørsmålet er fra den delen av pensum som Ole kan dårlig,
$K$ : Ole svarer korrekt på spørsmålet.

Ole setter opp følgende sannsynligheter:

$P(G) = 0.3$, $P(M) = 0.5$, $P(D) = 0.2$, $P(K|G) = 0.8$, $P(K|M) = 0.4$, $P(K|D) = 0.2$.

**b)** Vis de fire hendelsene i et venndiagram.

Hva er sannsynligheten for at Ole svarer korrekt på et tilfeldig valgt spørsmål, $P(K)$?

Gitt at Ole svarer korrekt på et tilfeldig valgt spørsmål, hva er da sannsynligheten for at spørsmålet er fra den delen av pensum som Ole kan dårlig, $P(D|K)$?

**Oppgave 2**

En forhandler kjøper inn et stort parti kulepenner og selger det videre til kunder. Han får klage på alle kulepenner som ikke virker. Anta først at antallet klager, $X$, er poissonfordelt med parameter $\lambda > 0$:

$$P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda} \quad ; x = 0, 1, 2, \ldots$$

**a)** Anta, kun i dette punktet, at $\lambda = 7$. Finn sannsynligheten for at forhandleren får en eller flere klager. Finn også sannsynligheten for at forhandleren får færre enn tre klager, gitt at han får en eller flere klager.

Leverandøren av kulepenner har tre fabrikker, benevnt $A$, $B$ og $C$, som produserer med ulik kvalitet. Forhandleren mottar varepartiet fra en av fabrikkene men han vet ikke fra hvilken. Dersom varepartiet kommer fra fabrikk $A$ vil antall klager være poissonfordelt med parameter $\lambda_A = 5$, hvis partiet er fra fabrikk $B$ og $C$ er klagefordelingsparameteren henholdsvis $\lambda_B = 15$ og $\lambda_C = 20$. Forhandleren antar i utgangspunktet at det er sannsynlighet $p_A = 0.5$ for at varepartiet kommer fra fabrikk $A$ og sannsynlighet $p_B = p_C = 0.25$ for hver av fabrikkene $B$ og $C$.

**b)** Utled et uttryk for sannsynlighetsfordelingen for antall klager forhandleren får.

**Oppgave 3**

På Botanisk forskningsstasjon er det plantet et felt med en sjelden grasart. På et bestemt tidspunkt er lengden $X$ målt i cm av et tilfeldig valgt grasstrå eksponentielt fordelt, dvs. at $X$ har sannsynlighetstetthet $f$ gitt ved at $f(x) = \frac{1}{\beta}e^{-x/\beta}$ for $x \geq 0$ og kumulativ fordelingsfunksjon $F$ gitt ved at $F(x) = 1 - e^{-x/\beta}$ for $x \geq 0$.

**a)** Anta (bare i dette punktet) at $\beta = 10$. Regn ut

$$P(X \leq 4), \quad P(X > 7) \quad \text{og} \quad P(X > 7 \mid X > 4).$$

En forsker hadde tenkt å måle lengden på et utvalg av grasstrå, men ved en misforståelse slår vaktmesteren graset samme dag som målingene skal gjøres. Forskeren vil da i stedet måle lengden Y av et utvalg av stråene som ligger i grasklipperens oppsamler. Alle stråene ble klipt i samme høyde $c$, og bare de som var høyere enn klippehøyden ble klipt. Gitt at et strå hadde lengde $X > c$, ligger det altså i oppsamleren, og har lengde $Y = X - c$.

**b)** Vis at $P(X > c) = e^{-c/\beta}$.

Finn $P(Y > y)$, der $y > 0$.

Hvilken kjent fordeling har $Y$?

## Oppgave 4

Ved et gartneri dyrkes hodekål. Vekten av den modne kålen, X, antas å være normalfordelt med forventningsverdi $\mu = 2$ kg og standardavvik $\sigma = 0.5$ kg.

**a)** Hvis en velger et kålhode tilfeldig, hva er sannsynligheten for at dette skal:
1) veie mindre enn 1.5 kg?
2) veie mellom 2 og 2.5 kg?

Kålhoder som veier mindre enn 1.5 kg oppfyller ikke kravet til klasse 1-kål.

**b)** Gitt at et kålhode oppfyller kravet (veier mer enn 1.5 kg), hva er sannsynligheten for at det veier mellom 2 og 2.5 kg?

Finn den kumulative fordelingen til vekten av et kålhode som oppfyller kravet til klasse 1? (Uttrykk svaret ved kumulativ fordelingsfunksjon for en normalfordelt variabel.)

## Oppgave 5

Selskapene Trysil Invest og Oppdal Invest er notert på børsen. Vi skal i denne oppgaven undersøke prisendringen de siste 120 dagene for disse selskapene. Vi antar at verdien til Trysil Invest ved børsens stengetid er $X_t$ for dag $t = 1, \ldots, 120$ og tilsvarende $Y_t$ for Oppdal Invest.

Filen `aksjekurser.txt`, som er tilgjengelig på kursets hjemmeside, inneholder observerte aksjekurser for Trysil Invest og Oppdal Invest de siste 120 dagene. Last ned filen og last den inn i Matlab, for eksempel ved:

```
aksjepriser = load('aksjekurser.txt');

Trysil_invest = aksjepriser(:,1);
Oppdal_invest = aksjepriser(:,2);
N = length(Trysil_invest);
```

**a)** Finn gjennomsnitt og standardavvik for de to aksjene. Plott de to aksjekursene i samme figur.

```
figure;
dag = 1:N;
plot(dag, Trysil_invest, 'k-');
hold on;
plot(dag, Oppdal_invest, 'k-.');
legend({'Trysil Invest', 'Oppdal Invest'}, 'Location','northwest')
```

Bruk figuren til å anslå om korrelasjonen mellom $X$ og $Y$ er negativ, omtrent null eller positiv. Begrunn svaret ditt.

I resten av oppgaven skal vi se på endring i pris fra dag til dag. Definer $\tilde{X}_t = X_{t+1} - X_t$ og $\tilde{Y}_t = Y_{t+1} - Y_t$ for $t = 1, \ldots 119$.

**b)** Du kan benytte kommandoene

```
diff_Trysil = diff(Trysil_invest);
diff_Oppdal = diff(Oppdal_invest);
```

i Matlab til å definere $\tilde{X}_t$ og $\tilde{Y}_t$. Funksjonen `histfit` plotter et histogram og en skalert sannsynlighetstetthet (fra en valgt sannsynlighetsfordelingsklasse) som skal passe best mulig til dataene. Bruk denne funksjonen til å lage et histogram med åtte søyler av endring i aksjepris for Oppdal Invest og en tilpasset normalfordeling i samme plott. Diskuter om det er rimelig å anta at endring i aksjekurs for Oppdal Invest er normalfordelt.

```
figure;
histfit(diff_Oppdal,8,'normal')
title('Oppdal Invest')
```

Benytt de observerte endringene i pris for Oppdal Invest til å anslå sannsynligheten for at aksjeprisen for Oppdal Invest øker fra en tilfeldig valgt dag til den påfølgende dagen.

```
sum(diff_Oppdal >= 0) / (N-1);
```

Gitt at prisen for Trysil Invest går ned fra en dag til den neste dagen, hva er sannsynligheten for at prisen til Oppdal Invest øker til den samme dagen?

```
(sum(diff_Oppdal >= 0 & diff_Trysil <= 0) / (N-1)) / (sum(diff_Trysil <0) / (N-1))
```

Forklar kommandoen over ved bruk av definisjonen av betinget sannsynlighet.

## Fasit

**2**. **a**) 0.9991, 0.0288

**3**. **a**) 0.33, 0.50, 0.74

**4**. **a**) $0.16, 0.34, 0.16$ **b**) 0.40