

Oppgavesett nr. 3

“MAT110 Statistikk 1”, 2020

Oppgave 1: (Sport Management - binomisk fordeling , forventning , varians)

La oss forsette med oppgaven med fotballaget Tufte IL fra øving 2.
Vi definerer samme diskrete stokastiske variabel som tidligere, nemlig:

$X =$ antall mål som *Tufte IL* skårer i løpet av en tilfeldig valgt kamp

med en litt annen tilnærming enn i øving 2:

Istedet for å se på sannsynligheten for å score et gitt antall mål i en kamp, så ser vi denne gangen på sannsynligheten for et en målsjanse blir mål. Basert på **historiske data** viser det seg at 20 % av målsjansene resulterer i mål, dvs.:

$$p = 0.20 \quad (1)$$

For enkelhets skyld antar vi at *Tufte IL* produserer $n = 10$ målsjanser i løpet av en kamp, dvs.:

$$n = 10 \quad (2)$$

målsjanser i løpet av en kamp. Vi antar videre at alle målsjansene er uavhengige.¹



Figur 1: Tufte IL.

¹Man kan argumentere at flere målsjanser kommer ofte etter hverandre fordi laget er inne i en ”god periode”, men slike og lignende effekter ser vi altså bort fra i denne oppgaven.

a) Forklar hvorfor X er **binomisk** fordelt, dvs. forklar hvorfor $X \sim \text{Bin}[n, p]$.²

b) i) Hva betyr $E[X]$ på “godt norsk” i vårt tilfelle?³

ii) Finn $E[X]$.

c) i) Hva betyr $\text{Var}[X]$ på “godt norsk” i vårt tilfelle?

ii) Finn $\text{Var}[X]$.

d) i) Hva er sannsynligheten for at laget skårer minst 1 mål i en kamp?⁴

ii) Hva er sannsynligheten for at laget skårer minst 2 mål i en kamp?⁵

²Hvilke 4 kriterier må være oppfylt for et en forsøksserie skal være binomisk? Er disse oppfylt i vårt tilfelle?

³Dvs. gi en tolkning av $E[X]$.

⁴Bruk 4 desimalers nøyaktighet.

⁵Bruk 4 desimalers nøyaktighet.

For å forbedre lagets resultater vurderes to strategier:

- A) øke uttellingen på målsjansene, dvs. øke p til $p = 0.30$
(antall målsjanser er uforandret $n = 10$)
- B) øke antall produserte målsjansene, dvs. øke n til $n = 15$
(utellingsprosenten $p = 0.20$ er uforandret)

Vi skal nå gjøre en statistisk analyse for å se om den ene strategien er bedre enn den andre.

e) Regn ut og fyll inn forventningsverdiene $E[X]$ og standardavvikene $\sigma[X]$ for strategi A og B i tabellen i figur 2.

f) Sett ut fra et rent statistisk ståsted, er den ene strategien bedre enn den andre? ⁶

⁶Anta at *Tufte IL* har samme defensive styrke uansett valg av strategi. Den strategien som er best er den som gir størst forventet antall skårede mål og minst standardavvik.

	$E[X]$	$\sigma[X]$
Historiske data	2	1.26
Strategi A (øke "p")		
Strategi B (øke "n")		

Figur 2: Fyll inn for $E[X]$ og $\sigma[X]$ i tabellen og lever inn.



Oppgave 2: (petroleumslogistikk - Poisson-fordeling , sum av stokastiske variabler)

Anta at du jobber som logistikk-koordinator ved olje- og gass firmaet British Petroleum BP. En del av jobben din er å få oversikt over nestenulykker og andre uønskede “*hendelser*” som skjer på plattformer som BP opererer i.

Du ønsker å se nærmere på antall hendelser ved hjelp av statistikk. Siden hendelser skjer relativt sjelden og siden det er hensiktsmessig å se på antall hendelser per tidsenhet så foreslår du å bruke “*loven om sjeldne begivenheter*”, dvs. **Poisson**-fordelingen. Du definerer den stokastiske variabelen:

$$X = \text{antall hendelser per uke}$$

som vi altså antar er Poisson-fordelt:

$$X \sim \text{Poi}[\lambda] \quad (3)$$

Basert på erfaring vet man at det er ca. 0.6 hendelser per uke på BP-plattformen Valhall i Nord-sjøen, dvs.:

$$\lambda = 0.6 \quad (4)$$



Figur 3: Valhall.

- a) Hva er sannsynligheten for at det skjer 1 hendelse i løpet av en uke, dvs. $P(X = 1)$? ⁷
- b) Hva er sannsynligheten for at det skjer **mer enn** 1 hendelse i løpet av en uke, dvs. $P(X > 1)$?

⁷Bruk 4 desimalers nøyaktighet.

La oss nå istedet se på antall hendelser i løpet av en periode på 2 uker.
I den sammenheng defineres den stokastiske variabelen Y :

$$Y = X_1 + X_2 \quad (5)$$

dvs. Y = antall hendelser i løpet av en periode på 2 uker.

c) Hva er forventet antall hendelser i løpet av 2 uker, dvs. $E[Y]$?

Anta at antall hendelser gitt en uke er uavhengig av antall hendelser en annen uke.

Man kan vise at summen av uavhengige Poisson-fordelinger også er -fordelt. ⁸

Det betyr at siden X_1 og X_2 er Poisson-fordelt, så er også summen av dem,
 $Y = X_1 + X_2$, Poisson-fordelt med forventning $E[Y] = \lambda_Y$: ⁹

$$Y \sim \text{Poi}[\lambda_Y] \quad (6)$$

d) Hva er sannsynligheten for at det skjer **mer enn** 1 hendelse i løpet av 2 uker,
dvs. hva er $P(Y > 1)$?

e) Hva er eller-sannsynligheten $P(X_1 > 1 \text{ eller } X_2 > 1)$? ¹⁰

⁸Du skal ikke vise dette. Bare ta dette for gitt.

⁹Dette gjelder kun for Poisson-fordelinger. Det gjelder ikke generelt. Det er ikke slik at f.eks. summen av binomial fordelinger er en ny binomial fordeling.

¹⁰Tips: bruk den generelle addisjonssetningen.

Dersom du har gjort oppgave **2d** og **2e** rett så innser du at:

$$\underbrace{P(Y > 1)}_{= 0.3374} > \underbrace{P(X_1 > 1 \text{ eller } X_2 > 1)}_{= 0.2289} \quad (7)$$

$$0.3374 > 0.2289 \quad (8)$$

Dette viser at:

sannsynligheten for at det skjer mer enn én hendelse i løpet av 2 uker er større enn sannsynligheten for at det skjer mer enn én hendelse i uke 1 ELLER mer enn én hendelse i uke 2.

Via komplementsetningen kan man vise at: ¹¹

$$P(Y > 1) = P(X_1 > 1 \text{ eller } X_2 > 1) + P(X_1 = 1, X_2 = 1) \quad (9)$$

- f)** Bruk lign.(9) til å argumentere for at det alltid er slik at $P(Y > 1) > P(X_1 > 1 \text{ eller } X_2 > 1)$ for Poisson-fordelinger.

■

¹¹Lign.(9) skal du ikke vise. Bare ta den for gitt.

Oppgave 3: (logistikk og økonomi - binomisk , sannsynlighet , forventning , varians)

Du har nylig blitt ansatt som “revenue-analytiker”, dvs. booking-ansvarlig, i SAS ved Kvernberget flyplass i Kristiansund. Denne jobben fikk du mye pga. din nylig avlagte bachelor i logistikk ved Høgskolen i Molde. Du skal tilnærme deg arbeidsoppgavene du har fått ved hjelp av statistikk.

Definer den stokastiske variabelen:

$$X = \text{antall personer som kjøper billett} \quad (10)$$

La videre:

$$n = \text{antall billetter som legges ut for salg for en bestemt flyavgang} \quad (11)$$

$$p = \begin{array}{l} \text{sannsynligheten for at en person kjøper en billett} \\ \text{og at personen faktisk møter opp til sin flyavgang} \end{array} \quad (12)$$



Figur 4: Kristiansund lufthavn, Kvernberget.

- a) Forklar hvorfor X er **binomisk** fordelt, dvs. forklar hvorfor $X \sim \text{Bin}[n, p]$.

Basert på historiske data viser det seg at “bare” 95 % av billettkjøperne møter opp, dvs. $p = 0.95$. Til Kvernberget har SAS satt inn Boeing 737-500 maskiner. Disse flyene har en kapasitet på $n = 120$ seter.

- b) i) Finn $E[X]$.

- ii) Hva betyr $E[X]$ på “godt norsk” i vårt tilfelle? (Dvs. gi en tolkning av $E[X]$.)

- c) i) Finn $\text{Var}[X]$.

- ii) Hva betyr $\text{Var}[X]$ på “godt norsk” i vårt tilfelle? (Dvs. gi en tolkning av $\text{Var}[X]$.)

Anta at SAS tjener 800 NOK per passasjer på en reise Oslo-Kristiansund for alle passasjerer som møter opp. Dermed kan vi definere en diskret stokastisk variabel I som beskriver **inntekten** til SAS for vår flyavgang:

$$I = aX \quad (13)$$

hvor

$$a = 800 \text{ NOK} \quad (\text{"a" er bare et konstant tall}) \quad (14)$$

d) Hva er forventet inntekt til SAS dersom $n = 120$ billetter legges ut for salg?

For at flyavgangen skal gå med færrest mulig tomme seter blir fly ofte "overbooket" - det legges ut flere billetter til salgs enn det er plass til i flyet. Dette fordi SAS regner med at noen passasjerer ikke møter opp. Anta i resten av oppgaven at SAS legger 123 billetter ut til salg selv om flyet kun har 120 seter, dvs. en potensiell "overbooking" på 3 seter.

e) Hva er forventet inntekt til SAS dersom de legger ut $n = 123$ billetter for salg?

f) Hva er sannsynligheten for at det møter opp flere passasjerer enn flyet har kapasitet til?

Dersom en billettgjøper møter opp og ikke får plass fordi flyet er fullt så regner SAS med en utgift på 5000 NOK. Dermed kan vi definere en diskret stokastisk variabel C som beskriver **kostnaden** til SAS for vår flyavgang:

$$C = \begin{cases} b(X - 120) & , \quad X \geq 120 \\ 0 & , \quad X < 120 \end{cases} \quad (15)$$

hvor

$$b = 5000 \text{ NOK} \quad (\text{“b” er bare et konstant tall}) \quad (16)$$

g) Hva betyr $E[C]$ på “godt norsk” ? ¹²

h) Hva er forventet kostnad til SAS dersom de legger ut $n = 123$ billetter for salg? ¹³

i) Lønner det seg for SAS å “overbooke”? ¹⁴



¹²Dvs. gi en tolkning av $E[C]$ i lign.(15). Ën setning er nok.

¹³Dvs. finn $E[C]$, hvor C er gitt ved lign.(15). Denne oppgaven er litt vanskelig. Spør gjerne om hjelp dersom du står fast.

¹⁴Gi en kort begrunnelse for svaret med å sammenligne oppgavene **3d**, **3e** og **3h**.

Oppgave 4: (logistikk - Poisson-fordeling , forventning , varians , sannsynligheter)

Helse Møre og Romsdal (Helse M&R) har ansvaret for ambulansetjenesten i fylket. Logistikerne i Helse M&R ønsker å se nærmere på antall utrykninger og responstid for å fordele kapasiteten av utrykningsbiler. De bestemmer seg for å begrense studien til 3 kommuner: Eide, Fræna og Gjemnes.

Logistikerne i Helse M&R ønsker å modellere dynamikken ved hjelp av statistikk. De definerer derfor følgende stokastiske variabler:

$$X_E = \text{antall akutte utrykninger per uke i Eide} \quad (17)$$

$$X_F = \text{antall akutte utrykninger per uke i Fræna} \quad (18)$$

$$X_G = \text{antall akutte utrykninger per uke i Gjemnes} \quad (19)$$

Siden utrykninger skjer relativt sjelden og siden vi ser på antall utrykninger per tid, altså en rate, så foreslår de å bruke “[loven om sjeldne begivenheter](#)”, dvs. **Poisson**-fordelingen:

$$X_i \sim \text{Poi}[\lambda_i] \quad (20)$$

hvor $i = E, F, G$. Basert på erfaring fra tidligere år så vet Helse M&R følgende:

$$\lambda_E = 1.5 \quad (21)$$

$$\lambda_F = 2.3 \quad (22)$$

$$\lambda_G = 1.9 \quad (23)$$

Her er λ_i rater, dvs. antall akutte utrykninger per uke.



Figur 5: Ambulanse.

- a) Hva er sannsynligheten for at det skjer 2 **utrykninger** i løpet av en uke i Gjemnes?
- b) Hva er sannsynligheten for at det skjer **mer enn** 2 utrykninger i løpet av en uke?
Oppgitt: $P(X_G = 0) = 0.1496$ og $P(X_G = 1) = 0.2842$.

Definer den stokastiske variabelen:

$$Y = X_E + X_F + X_G \quad (24)$$

dvs. Y = antall akutte utrykninger i Eide, Fræna og Gjemnes til sammen per uke.

- c) Hva er forventet antall akutte utrykninger i Eide, Fræna og Gjemnes tilsammen per uke? ¹⁵

¹⁵Finn $E[Y]$. Bruk gjerne en av regnereglene i kompendiet / formelsamlingen.

Anta at de stokastiske variablene X_E , X_F og X_G er uavhengige.

Man kan vise at summen av uavhengige Poisson-fordelinger også er Poisson-fordelt.¹⁶

Det betyr at siden X_E , X_F og X_G er uavhengige og Poisson-fordelt, så er også summen av dem, $Y = X_E + X_F + X_G$, Poisson-fordelt med forventning $E[Y] = \lambda_Y$:

$$Y \sim \text{Poi}[\lambda_Y] \quad (25)$$

- d) Hva er sannsynligheten for at det skjer **mer enn** 2 utrykninger i Eide, Fræna og Eide tilsammen per uke, $P(Y > 2)$?

■

¹⁶Dette skal du ikke vise. Bare ta det for gitt.

Oppgave 5: (medisin - Bernoulli-fordelingen og kobling med binomial-fordelingen)

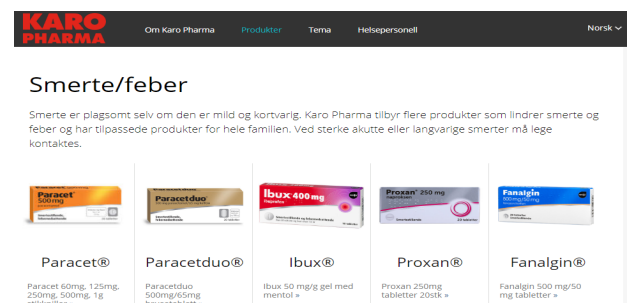
Weifa AS er en norsk legemiddelbedrift som i 2017 ble oppkjøpt av den svenske giganten *Karo Pharma*. Et av satsingsområdene til Weifa er reseptfri medisin mot smerte/feber, se figur 6.

Anta at Karo Pharma har utviklet en ny type smertelindrende medisin som skal gi **bedre effekt** enn det vanlige produktet paracet. For å kunne bruke dette i markedsføringen av produktet, må dem forsikre seg at dette stemmer, med en viss sannsynlighet.

Anta videre at Karo Pharma kan måle i hvilken grad kroppen til en forsøksperson kan oppta de viktigste smertelindrende stoffene. Kroppens evne til å absorbere medisinen vil derfor være avgjørende for hvordan medisinen fungerer.

Eksperiment:

De gir tabletten til en vilkårlig forsøksperson og måler opptaksnivået. Dersom opptaksnivået er større enn paracet sitt, så er forsøket en suksess. Hvis ikke er det fiasko.



Figur 6: Karo Pharma.

Ja/nei eksperimenter, dvs. eksperimenter med kun *to* utfall, kalles et *Bernoulli-forsøk*. For Bernoulli-forsøk defineres alltid følgende stokastiske variabel:

$$X = \begin{cases} 1 & , \text{ dersom forsøket gir } \underline{\text{suksess}} \\ 0 & , \text{ dersom forsøket gir } \underline{\text{fiasko}} \end{cases} \quad (26)$$

La suksess-sannsynligheten være gitt ved p , dvs.:

$$P(X = 1) = p \quad (\text{suksess-sannsynlighet}) \quad (27)$$

Ved komplementsetningen er da fiasko-sannsynligheten gitt ved $1 - p$, dvs.:¹⁷

$$P(X = 0) = 1 - p \quad (\text{fiasko-sannsynlighet}) \quad (33)$$

Den stokastiske variabelen X er dermed diskret med to mulige verdier 1 eller 0. Fordelingen til X kalles ”*Bernoulli-fordeling*”, og vi skriver:

$$X \sim \text{Ber}[p] \quad (34)$$

hvor p da er suksess-sannsynligheten.

¹⁷Komplementsetningen:

$$P(\text{suksess}) + P(\overline{\text{suksess}}) = 1 \quad (28)$$

$$P(\text{suksess}) + P(\text{fiasko}) = 1 \quad (29)$$

$$P(X = 1) + P(X = 0) = 1 \quad (30)$$

altså

$$\underline{P(X = 0)} = 1 - \overbrace{P(X = 1)}^{= p} \quad (31)$$

$$= \underline{1 - p} \quad (32)$$

- a)** Anta $p = 0.8$.
Tegn opp sannsynlighetsfordelingen $P(X = x)$ for X .¹⁸

Anta p er ukjent i oppgave **5b**, **5c** og **5d** nedenfor.

- b)** Regn ut forventningsverdien av X , dvs. $E[X]$.¹⁹
- c)** Regn ut variansen av X , dvs. $Var[X]$.
- d)** Regn ut standardavviket av X , dvs. $\sigma[X]$.

¹⁸Dvs. tegn opp et stolpediagram.

¹⁹Siden p er ukjent så får man ikke et tallsvar for $E[X]$, men $E[X]$ uttrykt i p .

Karo Pharma ønsker å teste ut medisinen på 10 forsøkspersoner. Anta uavhengige forsøk. La X_i være den stokastiske variabelen som beskriver Bernoulli-forsøk nr. i , dvs.:

$$X_i = \begin{cases} 1 & , \text{ dersom forsøket gir } \underline{\text{suksess}} \text{ for pasient nr. } i \\ 0 & , \text{ dersom forsøket gir } \underline{\text{flasko}} \text{ for pasient nr. } i \end{cases} \quad (35)$$

hvor $i = 1, 2, 3, \dots, 10$.

Anta alle X_i 'ene er Bernoulli-fordelt, dvs. $X_i \sim \text{Ber}[p]$ for alle $i = 1, 2, 3, \dots, 10$.

Her er suksess-sannsynligheten p er ukjent, men den samme for alle 10 forsøkene.

Den stokastiske variabelen:

$$Y = X_1 + X_2 + X_3 \dots + X_{10} \quad (36)$$

beskriver da antall pasienter hvor den nye medisinen gir bedre effekt enn det vanlige produktet paracet.

e) Forklar med ord hvorfor Y er binomisk fordelt, dvs. forklar hvorfor $Y \sim \text{Bin}[n = 10, p]$.

I oppgave 5e argumenterte vi for at Y er binomisk fordelt.

Dermed kan vi konkludere med at summen av uavhengige stokastiske variabler som er Bernoulli-fordelt, er binomisk fordelt:

$$\underbrace{Y}_{\text{Bin}[n,p]} = \underbrace{X_1 + X_2 + X_3 \dots + X_n}_{\text{Ber}[p]} \quad (37)$$

Fordelingen til den binomisk variabelen Y er da:

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y} \quad (38)$$

hvor

$$n = \text{antall pasienter (antall "forsøk")} \quad (39)$$

$$p = \text{suksess-sannsynligheten} \quad (40)$$

$$y = \text{antall pasienter (av totalt } n \text{ pasienter)} \quad (41)$$

som har bedre effekt av medisinen enn paracet

- f) Karo Pharma ønsker å finne sannsynligheten for at $y = 9$ av $n = 10$ pasienter gir suksess, dvs. sannsynligheten for at 9 av 10 pasienter så gir Karo Pharma sin medisin **bedre effekt** enn det vanlige produktet paracet.

I denne deloppgaven skal vi bevise lign.(38) for det aktuelle tilfellet:

Vis at sannsynligheten for at $y = 9$ av $n = 10$ pasientene gir suksess, er gitt ved følgende binomiske fordeling: ²⁰

$$P(Y = 9) = \binom{10}{9} p^9 (1-p)^1 \quad (42)$$

²⁰Det er ikke meningen at du bare skal sette $y = 9$ og $n = 10$ inn i lign.(38). Du skal bevise lign.(42) fra grunnen av. Denne oppgaven er lang og litt vanskelig. Se gjerne eksempel på side 152 i kompendiet.

- g)** Kravet som Weifa stiller seg selv er at 10 av 10 Bernoulli-forsøk blir en suksess med 80 % sannsynlighet. Dette kravet kan formuleres som:

$$P(Y = 10) = 0.80 \quad (43)$$

Hva må p være for at kravet skal være oppfylt?

- h)** Tolk størrelsen p som du regnet ut i oppgave **5g**.

