



NTNU

Kunnskap for en bedre verden

Plenumsregning uke 37

13.09.23

Karina Lilleborge (vikar for Sara Martino)

Tema denne uka

- Kovarians og korrelasjon
- Forventningsverdi og varians til lineære funksjoner av stokastiske variabler
- Regneregler for forventningsverdi og varians
- Viktige diskrete og kontinuerlige fordelinger
 - Binomisk fordeling

Oppgave 1 (eksamen høst 2021 morgen)

1(d)

Innledning: La X og Y være to diskrete stokastiske variabler med simultanfordeling $P(X = x, Y = y)$ som gitt i tabellen under:

$x \backslash y$	0	1	2
0	0.22	0	0.31
1	0.32	0.15	0

Oppgave: Regn ut følgende størrelser. Angi svarene med tre desimaler etter komma.

- $P(X = 1) =$
- $P(Y = 0 | X = 0) =$
- $E(2XY) =$

Teorem: Setningen om total sannsynlighet

La A være en hendelse i et utfallsrom S . La $B_1, B_2, \dots, B_n$ være en partisjon av S , dvs $B_i \cap B_j = \emptyset$ for alle $i \neq j$ og $B_1 \cup B_2 \cup \dots \cup B_n = S$. Da har vi at

$$\begin{aligned} P(A) &= \sum_{i=1}^n P(A \cap B_i) \\ &= \sum_{i=1}^n P(A|B_i)P(B_i). \end{aligned}$$

Definisjon: Betinget sannsynlighet

La A og B være to hendelser i et utfallsrom S og anta at $P(A) > 0$. Den betingede sannsynligheten for B gitt A er da

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Temasider: <https://tma4245.math.ntnu.no/hendelser-og-sannsynlighet/betinget-sannsynlighet/setningen-om-total-sannsynlighet/>
<https://tma4245.math.ntnu.no/hendelser-og-sannsynlighet/betinget-sannsynlighet/>

Teorem: Forventningsverdi til en funksjon av flere stokastiske variabler

La $X_1, \dots, X_n$ være n stokastiske variabler med simultanfordeling $f(x_1, \dots, x_n)$, og la $Z = g(X_1, \dots, X_n)$ for en skalar funksjon $g(x_1, \dots, x_n)$, dvs $g(x_1, \dots, x_n) \in \mathbb{R}$. Hvis $X_1, \dots, X_n$ er diskrete stokastiske variabler har man da

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[g(X_1, \dots, X_n)] \\ &= \sum_{x_1} \cdots \sum_{x_n} g(x_1, \dots, x_n) f(x_1, \dots, x_n). \end{aligned}$$

Hvis $X_1, \dots, X_n$ er kontinuerlige stokastiske variabler har man

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[g(X_1, \dots, X_n)] = \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n. \end{aligned}$$

Temaside: <https://tma4245.math.ntnu.no/forventing-og-varians/forventningsverdi/forventningsverdi-funksjoner-av-stokastiske-variabler-egx/>

Teorem: Forventningsverdi til en funksjon av flere stokastiske variabler

La $X_1, \dots, X_n$ være n stokastiske variabler med simultanfordeling $f(x_1, \dots, x_n)$, og la $Z = g(X_1, \dots, X_n)$ for en skalar funksjon $g(x_1, \dots, x_n)$, dvs $g(x_1, \dots, x_n) \in \mathbb{R}$. Hvis $X_1, \dots, X_n$ er diskrete stokastiske variabler har man da

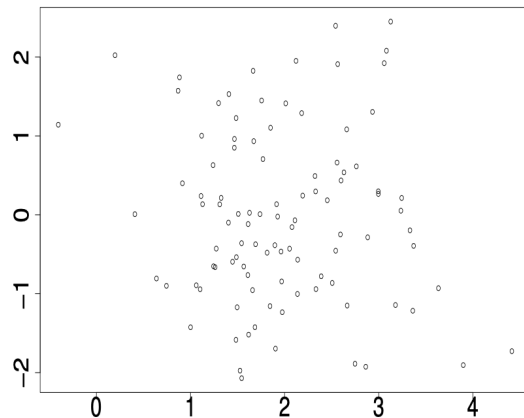
$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[g(X_1, \dots, X_n)] \\ &= \sum_{x_1} \cdots \sum_{x_n} g(x_1, \dots, x_n) f(x_1, \dots, x_n). \end{aligned}$$

Hvis $X_1, \dots, X_n$ er kontinuerlige stokastiske variabler har man

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[g(X_1, \dots, X_n)] = \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n. \end{aligned}$$

Temaside: <https://tma4245.math.ntnu.no/forventing-og-varians/forventningsverdi/forventningsverdi-funksjoner-av-stokastiske-variabler-egx/>

Oppgave 2



Figur 1: Parvise data fra en simultanfordeling.

Figur 1 viser parvise utfall av to stokastiske variabler X (horisontalt) og Y (vertikalt).

a) Hva vil du anslå at korrelasjonen mellom X og Y er? Gi en kort begrunnelse.

Forventningsverdien og standardavviket for hver av de to variablene har heltallsverdier. Anslå disse verdiene ved å studere Figur 1 visuelt.

Kovarians og korrelasjon (fra forelesning)

- ★ Forventningsverdi og varians sier noe om fordelingen til en stokastisk variabel
- ★ Vi trenger også en størrelse som sier noe om hvordan to SV samvarierer

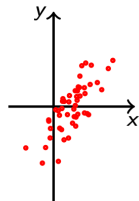
Definisjon (Kovarians)

Kovariansen mellom to SV X og Y er

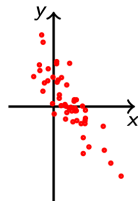
$$\sigma_{XY} = \text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)],$$

der $\mu_X = E[X]$ og $\mu_Y = E[Y]$.

- ★ Tolkning av (fortegnet til) kovarians
 - $\text{Cov}[X, Y] > 0$: Store verdier av X tenderer til å komme sammen med store verdier av Y , og små verdier av X tenderer til å komme sammen med små verdier av Y
 - $\text{Cov}[X, Y] < 0$: Store verdier av X tenderer til å komme sammen med små verdier av Y , og små verdier av X tenderer til å komme sammen med store verdier av Y



$\text{Cov}[X, Y] > 0$



$\text{Cov}[X, Y] < 0$

Definisjon (Korrelasjon)

Korrelasjonen mellom to stokastiske variabler X og Y er

$$\rho = \rho_{XY} = \text{Corr}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \cdot \text{Var}[Y]}}.$$

- ★ Merk: Fortegnet til $\text{Corr}[X, Y]$ er det samme som for $\text{Cov}[X, Y]$
- ★ Skal senere vise: X og Y uavhengige $\Rightarrow \text{Cov}[X, Y] = 0, \text{Corr}[X, Y] = 0$

Oppgave 3 – del 1 (av 2)

I et spill kaster hver deltager i hver omgang en terning 10 ganger. Spilleren får ett poeng hver gang han får to 6-ere etter hverandre. Dersom utfallet av en omgang er

1 5 6 6 6 4 2 3 6 6

får spilleren 3 poeng. Etter at en spiller er ferdig med sine terningkast, går turen videre til nestemann. La X_i være utfallet av terningkast nr. i , $i = 1, \dots, 10$.

For å telle antall poeng en spiller får i omgangen, Z , lar vi Y_j være lik 1 dersom spilleren får ett poeng etter kast nr. j , $j = 2, \dots, 10$, og lik 0 ellers. Altså:

$$Y_j = \begin{cases} 1, & \text{dersom } X_j = X_{j-1} = 6 \\ 0, & \text{ellers} \end{cases}$$

Vi har da sammenhengen $Z = \sum_{j=2}^{10} Y_j$.

- b) Er Y_2 og Y_3 avhengige eller uavhengige? (Begrunn svaret).
Er Y_2 og Y_4 avhengige eller uavhengige? (Begrunn svaret).
Avgjør om korrelasjonskoeffisienten mellom Y_2 og Y_3 er negativ, positiv eller null. (Begrunn svaret). Gjør det samme for korrelasjonskoeffisienten mellom Y_2 og Y_4 .

Oppgave 3 – del 2

I et spill kaster hver deltager i hver omgang en terning 10 ganger. Spilleren får ett poeng hver gang han får to 6-ere etter hverandre. Dersom utfallet av en omgang er

1 5 6 6 6 4 2 3 6 6

får spilleren 3 poeng. Etter at en spiller er ferdig med sine terningkast, går turen videre til nestemann. La X_i være utfallet av terningkast nr. i , $i = 1, \dots, 10$.

For å telle antall poeng en spiller får i omgangen, Z , lar vi Y_j være lik 1 dersom spilleren får ett poeng etter kast nr. j , $j = 2, \dots, 10$, og lik 0 ellers. Altså:

$$Y_j = \begin{cases} 1, & \text{dersom } X_j = X_{j-1} = 6 \\ 0, & \text{ellers} \end{cases}$$

Vi har da sammenhengen $Z = \sum_{j=2}^{10} Y_j$.

c) Beregn $\text{Cov}(Y_2, Y_3)$, $E(Z)$ og $\text{Var}(Z)$.

Oppgave 4 (fra eksamen høst 2021 kveld

3(a) Innledning: Det har brutt ut et virus i samfunnet. Myndighetene ønsker å finne andelen smittede i befolkningen på et tidspunkt ved å teste et tilfeldig utvalg av N personer.

La p være sannsynligheten for at en tilfeldig utvalgt person fra befolkningen er smittet. Anta at testen ikke er helt pålitelig, sannsynligheten for et "falskt negativt" resultat er 0.02. Dette betyr at sannsynligheten for at en syk person tester positiv er 0.98. Du kan anta at sannsynligheten for et "falskt positivt" resultat, det vil si at en frisk person tester positivt, er null.

La X være en stokastisk variabel som er lik antall av de N personene i utvalget som tester positivt.

Oppgave:

- Forklar hvorfor X er binomisk fordelt med parametre N og $0.98p$.
- Vis at $\hat{p} = \frac{X}{0.98N}$ er en forventningsrett estimator for p og regn ut standardavviket til denne estimatoren.

Oppgave 4 (fra eksamen høst 2021 kveld

3(a) Innledning: Det har brutt ut et virus i samfunnet. Myndighetene ønsker å finne andelen smittede i befolkningen på et tidspunkt ved å teste et tilfeldig utvalg av N personer.

La p være sannsynligheten for at en tilfeldig utvalgt person fra befolkningen er smittet. Anta at testen ikke er helt pålitelig, sannsynligheten for et "falskt negativt" resultat er 0.02. Dette betyr at sannsynligheten for at en syk person tester positiv er 0.98. Du kan anta at sannsynligheten for et "falskt positivt" resultat, det vil si at en frisk person tester positivt, er null.

La X være en stokastisk variabel som er lik antall av de N personene i utvalget som tester positivt.

Oppgave:

- Forklar hvorfor X er binomisk fordelt med parametre N og $0.98p$.
- $E[\hat{p}] = E[\frac{X}{0.98N}] = p$ og regn ut standardavviket til denne estimatoren.

Binomisk fordeling

Definisjon (Bernoulli forsøksrekke)

Anta at vi gjentar et stokastisk forsøk mange ganger. Vi sier da at vi har en Bernoulli forsøksrekke dersom

- ★ forsøket har to mulige utfall, som vi kaller henholdsvis «suksess» og «fiasko»,
- ★ utfallet av forsøkene er uavhengige,
- ★ sannsynligheten for suksess er lik (p) i alle forsøkene.

Definisjon (Binomisk fordeling)

Anta at vi har en Bernoulli forsøksrekke hvor vi gjentar et forsøk n ganger, og la X være antall forsøk som gir suksess. Vi sier da at X er binomisk fordelt med parametre n og p .

- ★ Punktsannsynlighet

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \dots, n$$

Teorem: Forventningsverdi og varians

La X være binomisk fordelt med parametre n og p . Da er forventningsverdi og varians til X gitt ved

$$E[X] = np \quad \text{og} \quad \text{Var}[X] = np(1-p).$$

Oppgave 5 (fra eksamen vår 2019)

Oppgave 1 Elektrisk sparkesykkel

Et sykehus som behandler pasienter med bruddskader, ønsker å studere sammenhengen mellom bruk av elektrisk sparkesykkel og bruddskader. De registrerer derfor om hver pasient med bruddskader fikk skaden ved bruk av elektrisk sparkesykkel. I 2018 behandlet sykehuset n pasienter med bruddskader. Anta videre at sannsynligheten for at en tilfeldig valgt pasient med bruddskade hadde vært utsatt for en ulykke ved bruk av elektrisk sparkesykkel er p .

La X være antall pasienter med bruddskader som var involvert i ei ulykke ved bruk av elektrisk sparkesykkel. Da er X binomisk fordelt med n forsøk og konstant suksesssannsynlighet p .

a) Anta bare i dette punktet at $n = 17$ og $p = 0.2$.

Finn sannsynligheten for at nøyaktig 4 av pasientene med bruddskader var involvert i en ulykke med elektrisk sparkesykkel.

Finn sannsynligheten for at minst 4 av pasientene med bruddskader var involvert i en ulykke med elektrisk sparkesykkel.

Gitt at minst 4 av pasientene med bruddskader var involvert i en ulykke med elektrisk sparkesykkel, finn sannsynligheten for at minst 6 av pasientene med bruddskader var involvert i en ulykke med elektrisk sparkesykkel.

Oppgave 6 – del 1(av 2) (fra eksamen høst 2015)

Rakk ikke gå igjennom denne oppgaven i timen,
det er mulig å spørre på forum eller statistikklab
om du vil ha veiledning!

Oppgave 3 Robotkommunikasjon

To roboter kommuniserer ved å sende binære signaler til hverandre. Dessverre er ikke overføringen av signaler perfekt: Det er en sannsynlighet for at hvis 0 er sendt fra den ene roboten så mottas 1 av den andre roboten, og hvis 1 er sendt fra den ene roboten mottas 0 av den andre roboten.

- a) La X være en diskret stokastisk variabel som angir signalet som er sendt, og la Y være en diskret stokastisk variabel som angir signalet som er mottatt. Mulige verdier for X og Y er 0 og 1. Anta at $P(X = 1) = 0.2$. Anta videre at kommunikasjonen er slik at $P(Y = 1 \mid X = 1) = P(Y = 0 \mid X = 0) = 0.9$.

Bruk loven om total sannsynlighet til å regne ut $P(Y = 1)$.

Bruk Bayes' regel til å regne ut sannsynligheten $P(X = 1 \mid Y = 1)$.

Oppgave 6 – del 2(av 2)

(fra eksamen høst 2015) (ekstraoppgave)

Vi ser videre på signaler av lengde 5. La X_i være en diskret stokastisk variabel som angir signalet som er sendt i posisjon i , og Y_i en diskret stokastiske variabel som angir signalet som er mottatt i posisjon i . Her er $i = 1, 2, \dots, 5$, og de mulige verdiene til X_i og Y_i er 0 og 1. Anta at signalet i hver posisjon i overføres korrekt med sannsynlighet $P(Y_i = 1 \mid X_i = 1) = P(Y_i = 0 \mid X_i = 0) = 0.9$ for alle $i = 1, 2, \dots, 5$. Den stokastiske variabelen Y_i er kun avhengig av X_i og ikke av andre Y_j eller X_j , for $j \neq i$.

Nøtt:

- b) Hva er sannsynligheten for at et utsendt signal blir korrekt mottatt i minst 4 av de 5 posisjonene?

I en situasjon har vi at signalene $(0, 0, 0, 0, 0)$ og $(1, 1, 1, 1, 1)$ begge har sannsynlighet 0.35 for å bli utsendt, mens hver av de resterende 30 mulige signalene har sannsynlighet 0.01. Dersom mottatt signal er $(0, 0, 0, 0, 0)$. Hva er sannsynligheten for at utsendt signal er $(0, 0, 0, 0, 0)$?