

Oppgavesett SOK-2009

Øivind Schøyen

Høst 2023

1 Introduksjon

Velg en av oppgavene og løs alle deloppgavene. Du presenterer oppgaven din og intuisjonen bak som del av muntlig eksamen. Du vil også få en tilfeldig trukket annen oppgave en den du har løst som spørsmål på eksamen. Oppgavene vil være i delvis endret form. Oppgavene med stjerne på vil ikke bli trukket som oppgave på eksamen. Dersom du har løst oppgavene med stjerne vil du altså presentere den.

2 Oppgave 1

En diskret variabel X kan ta verdiene $x_i \in \{x_0, \dots, x_N\}$. Forventningen til X , $E[X]$, er summen av mulige utfall ganger sannsynligheten for utfallet:

$$E[X] = x_0 p(x_0) + x_1 p(x_1) + x_2 p(x_2) + \dots + x_N p(x_N) = \sum_{i=0}^N x_i p(X = x_i)$$

1. Bruk definisjonen til å finne forventningen av en variabel T som er lik antall prikker på en seks-sidet terning $t \in \{1, \dots, 6\}$ der alle utfall har samme sannsynlighet $p(X = x) = \frac{1}{6}$. Vi skriver dette, $E[T]$. Hva er sannsynligheten for å få den forventede verdien? Forklar svaret ditt.

2. Finn forventningen til summen av to seksidede terninger $2T$, $E[2T]$. Skriv svaret på formen av formelen over dvs. $E[2T] = \sum_{t=2}^{12} p(2T = t)t$. Hva er forskjellen mellom sannsynlighetene $p(2T = t)t$ og $p(T = t)$?

En sannsynlighetsfordeling for en diskret variabel som T gir sannsynligheten for å få de mulige verdiene i utfallrommet:

$$f(x) = P(X = x)$$

Den kumulative sannsynlighetsfordelingen gir sannsynligheten for at en variabel skal være mindre eller lik en gitt verdi x :

$$F(x) = P(X \leq x) = \sum_{i=0}^x f(x_i)$$

3. Hva er utfallrommet for $2T$ og utfallrommet for antall prikker på to separate terninger T, T ?

4. Hva er $f(5)$ for en eller to terninger? Hva er $F(3)$ for en eller to terninger?

5. Anta at en terning har $N > 3$ sider med $\{1, 2, 3, \dots, (N-1), N\}$ prikker på hver side. Hva er sannsynligheten for å få t prikker? Kan du skrive et uttrykk for forventet antall prikker?

6. Kan du finne sannsynligheten for at antall prikker på to terninger ligger over 2 eller under 12, det vil si $P(3 \leq 2T \leq 11) = F(11) - F(3) = \sum_{i=3}^{11} f(2T) - \sum_{i=0}^2 f(2T) = \sum_{i=3}^{11} f(2T)$.

3 Oppgave 2

Vi har en variabel X som har et utfallrom $x \in \{1, 2, 3\}$ der utfall har sannsynlighet $p(X = x) = \frac{x}{6}$ for alle utfall i utfallrommet.

1. Kan du tegne opp fordelingen til $f(x)$ for verdiene i utfallrommet? Hva kaller vi denne fordelingen?

2. Kan du tegne den kumulative sannsynlighetsfordelingen $F(X)$ for X og X^3 ?
3. Hva er forventningen $E[X]$? Hva er forventningen til $E[X^3]$?
4. Kan du regne ut korrelasjon, kovarians, varians og standardavik for (Y_1, Y_2) der $Y_1 = 3 + 5X$ og $Y_2 = 4 + 6X^3$ og (X, X^3) for verdiene $x \in \{1, 2, 3\}$. Kan du forklare svaret du får med egne ord?

4 Oppgave 3

Definer T som antall prikker på en seksidet terning og $2T$ som antall prikker på to sekssidede terninger.

1. Et av utfallsrommene T og $2T$ har uniform sannsynlighet det andre har ikke uniform sannsynlighetsfordeling. Forklar hva det betyr og hvorfor det er slik. Tegn opp grafen for $f(T), f(2T)$ og den kumulative sannsynlighetsfordelingen $F(2T), F(T)$?
2. Kan du regne ut korrelasjon, kovarians, varians og standardavik for T og $2T$?
3. Kan du lage en grafisk fremstilling fra to utfallsrom til T til ett utfallsrom med $2T$. Tips er å bruke piler.

Dersom to hendelser A og B er uavhengige hendelser holder det at $P(AB) = P(A)P(B)$. Les som og, altså settet av A og B .

4. For to terningkast, hva er sannsynligheten for å observere at to terninger er like, $t_1 = t_2$ definer dette som $P(A)$? Hva er sannsynligheten for å observere at summen er mellom 7 og 10 definer dette som $P(B)$? Hva er sannsynligheten for å observere at summen er 2, 7 eller 8 definer dette som $P(C)$. Hva er sannsynligheten for at alle tre A , B og C skal skje i.e $P(A \cap B \cap C)$? Hva er sannsynligheten for $P(A \cap B)$? Kan du forklare hvorfor du får svarene?

5 Oppgave 4

Anta at du observerer følgende tre data sett d_i^j : $d_i^1 \in \{1, 2, 4, 5\}$, $d_i^2 \in \{5, 6\}$, $d_i^3 \in \{6, 8, 4, 5\}$

1. Hvilke av datasettene har positiv sannsynlighet for å være generert av en seksidet terning og hvilke kan være fra to sekssidede terninger? Forklar svaret ditt.
2. Regn ut gjennomsnitt \bar{d}_i og standardaviket, SD_{d_i} for alle tre utvalgene i.e. $j \in 1, 2, 3$.
3. Regn ut forventning og standardaviket for antall prikker på en seksidet terning, T og summen av prikker på to sekssidede terninger $2T$.
4. Hva er forventningen til to terninger kastet to ganger i.e. $2 \times E[2T]$? Hva er forventningen til en terning kastet to ganger $2 \times E[T]$?
5. La oss si at du har en hypotese om at d_i^j er generert av en terning (T). Hva er sannsynligheten for å observere d_i^2 dersom data er generert av en terning? Hva er sannsynligheten for å observere d_i^2 dersom data er generert av to terninger?

6 Oppgave 5*

Anta at du har to sett med to sekssidede terninger. Noen andre kaster terningene og skriver ned summen av prikker fra begge terningene; T_1^1 og T_2^2 . Du får ikke se terningene men får vite at summen fra sett en er større en fra sett to dvs. $T_2^1 > T_2^2$. La oss si at du tar ut to terninger en fra sett en t_1^1 og en fra sett to t_2^2 .

1. Er forventningsverdien lik for disse to terningene? Forklar svaret ditt.
2. Tegn en graf med mulige utfall for T_1 og T_2 . Hvilke utfall er mulige og hvilke er ikke mulige gitt $T_2^1 > T_2^2$? Er det uniform sannsynlighet for utfallene?
3. Bruk R og regn ut forventningsverdien til de to terningene.

Tips: 1. Formelen for trekkning av X mulige r ganger med tilbakelegging er X^r . For fire seksidede terninger altså $X = 6^r = 4$. Sjekk også <https://tma4245.math.ntnu.no/hendelser-og-sannsynlighet/uniform-sannsynlighetsmodell/kombinatorikk-ordnet-utvalg-trekning-med-tilbakelegging/> eller <https://www.hackmath.net/en/calculator/combinations-and-permutations>.

Lag 36x36 vektorer med et tall for hver terning, det vil si $[(1,1)(1,1)], [(1,1)(1,2)], [(1,1)(1,3)], \dots, [(6,6)(6,6)]$. Dette er utfallsrommet for to sett med to terninger. Du må muligens bruke en spesial komando for å få til dette datasettet.

Sjekk kode på: <https://stackoverflow.com/questions/45878448/creating-sample-space-in-r>

2. Fjern alle vektorer der terning sett en er større en terningsett to dvs. der $[(1,1) < (1,2)]$.

3. Du skal nå sitte igjen med litt mindre en halvparten av de $[36 \times 36]$ utfallene. (Disse oppfyller at summen av terningkast på to er større en en $T_2^1 > T_2^2$.)

4. Kalkuler forventningen til den første av terningen. Alle utfall er like sannsynlige så dette er lett.

7 Oppgave 6

Du får i oppgave å lage et spill til kasino, Hver gang spillerne får et tall x på to terninger, i.e. $T_2 = x$ tjener spilleren en bonus (B) på 10 000 kroner.

1. En formel for forventet premie B om det bare er et utfall som gir premie er $E[B] = p(B)B$; kan du sette opp denne formelen dersom to eller N utfall gir premie?

2. Hva er minste spillet kan koste samtidig som kasino ikke taper penger. (Kasinoet har en kjempe stor egenkapital og kan låne så mye de vil i banken).

3. Kan du sette opp en B for få et eller flere tall, en pris for deltakelse D for å spille slik at spillerene vinner med sannsynlighet: A. $\frac{1}{36}$ B. $\frac{2}{36}$ C. $\frac{1}{2}$.

4.1. En gruppe kunder har tapt mange penger og er lei seg. De klager og sier det er urettferdig at kasinoet ditt tar mer betalt for å spille en forventet utbetaling. Hvordan forklarer du dette til kundene?

4.2 Eierne av kasinoet er lei seg fordi banken tar høyere rente av lånet hans en det banken betaler i innskuddsrente. Hvordan forklarer du til eieren at banken må ta rente.

En gruppe på 5 studenter $i \in \{1, 2, 3, 4, 5\}$ går på kasino. Vi skriver poengsummen i statistikkfaget deres på skalaen $p_i \in [1, 100]$ og gevinsten på kasino $[-100, 100]$. Data er som følger, $\{s_i, p_i\}_i$: $\{10, -50\}_1, \{78, -100\}_2, \{98, 100\}_3, \{60, 50\}_4, \{30, 25\}_5$.

5. Kan du sett opp gjennomsnitt og median for p_i og s_i . Er det noe type tall i utvalget? Sett også opp varians, standardavvik for s_i, p_i og korrelasjon og kovarians mellom s_i, p_i .

8 Oppgave 7*

Anta at noen foreslår at du får en ”ny” måte å organisere samfunnet ditt på. Vi definerer suksessen for et samfunn l med organisering i som $S_l^i \in (0, 1)$. Du observer følgende data for systemet som er innført i 7 samfunn som $S_{i=C}, [0, 011917, 0, 021975, 0, 041989, 0, 031953, 0, 00009, 0, 000004, 0, 1]$. Du finner et datasett som har kodet graden av historisk C mellom $(0, 1)$ for 88 land over 72 år. Korrelasjonen fra det historiske datasettet viser en signifikant og sterk negativ korrelasjon $\rho_{C,S} = -0,91989114$. For systemet dere allerede har er gjennomsnittet mellom 45 land $\overline{S^{i=D}} = 0,89$ med minimums observasjon 0,25 og maksimums observasjon 0,99.

1* Anbefaler du å prøve det nye systemet? Hvorfor?

2* Den som foreslår systemet nekter å diskutere antakelsene for teorien hans. Han er forøvrig glad i statistikk og sannsynlighetsregning. Hvilke egenskaper ved data tilsier at du nåde konklusjonen din? Hvordan forklarer du anbefalingen din?

3*. Kan du forklare korrelasjonskoeffesient og hvilke format data som ga $\rho_{C,S} = -0,91989114$ må ha hatt?

4*. Kan du bruke R til å genere et datasett som har egenskapene til det historiske datasettet data settet med 45 land.