Skrivne og trykte hjelpemiddel samt kalkulator er tillate.
Ta med **all mellomrekning** som trengst for å grunngje svaret.

Oppgåve 1 . . . . *(15%)*

Me kastar ein balansert mynt fire gongar. Lat den stokastiske variabelen X vera talet på kron i dei fire kasta.

(a) Kva er utfallsrommet til X?

Solution: Utfallsrommet er $\{0, 1, 2, 3, 4\}$.

(b) Kva er sannsynsfordelinga til X?

Solution:

$$P(X = x) = \binom{4}{x}\frac{1}{16}, \quad x = 0, 1, 2, 3, 4$$

(c) Rekn ut forventningsverdien for X.

Solution:

$$\begin{aligned} E(x) &= \sum xp(x) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} \\ &= \frac{0 + 4 + 12 + 12 + 4}{16} = \frac{32}{16} = 2 \end{aligned} \tag{1}$$

Alternativt kan vi bruke formelen for forventningsverdi i en binomialfordeling:

$$\mu = n \cdot \pi = 4 \cdot 0{,}5 = 2 \tag{2}$$

(d) Rekn ut populasjonsvariansen for X.

Solution:

$$\begin{aligned} \mathsf{var}(X) &= E\big((X - \mu)^2\big) \\ &= \sum_{i=0}^{4} (x - 2)^2 \binom{4}{x}\frac{1}{16} \\ &= (-2)^2 \cdot \frac{1}{16} + (-1)^2 \cdot \frac{4}{16} + (0)^2 \cdot \frac{6}{16} + (+1)^2 \cdot \frac{4}{16} + (+2)^2 \cdot \frac{1}{16} \\ &= \frac{4 + 4 + 0 + 4 + 4}{16} = \frac{16}{16} = 1. \end{aligned}$$

Alternativt kan vi bruke formelen for varians i en binomialfordeling:

$$\mathsf{var}(X) = n \cdot \pi \cdot (1 - \pi) = 4 \cdot 0{,}5 \cdot (1 - 0{,}5) = 1$$

(e) Kva er den kumulative sannsynsfordelinga for X?

Solution:

x	$f(x)$	$F(x)$
0	1/16	1/16
1	4/16	5/16
2	6/16	11/16
3	4/16	15/16
4	1/16	1

Den fullstendige kumulative fordelingsfunksjonen er:

$$F(X) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/16 & \text{if } 0 \leq x < 1 \\ 5/16 & \text{if } 1 \leq x < 2 \\ 11/16 & \text{if } 2 \leq x < 3 \\ 15/16 & \text{if } 3 \leq x < 4 \\ 1 & \text{if } x \geq 4 \end{cases}$$

Vi gjentar eksperimentet fem ganger og får som resultat: 2; 1; 4; 2; 1.

(f) Rekn ut utvalsmiddelverdien.

Solution:

$$\bar{x} = \frac{2+1+4+2+1}{5} = 2$$

(g) Rekn ut utvalsvariansen.

Solution:

x	$x - \bar{x}$	$(x - \bar{x})^2$
2	0	0
1	-1	1
4	2	4
2	0	0
1	-1	1

$s^2 = \sum_x \frac{(x-\bar{x})^2}{n-1} = \frac{6}{5-1} = 3/2$

Oppgåve 2 . . . . *(4%)*

Bruk sannsynstabellane i boka og finn fylgjande:

(a) $P(T > 1{,}5)$ der T har Students t-fordeling med fem fridomsgradar.

Solution: $P(T > 1{,}5) = 0{,}097$

(b) Den kritiske verdien z slik at $P(Z < z) = 0{,}95$ der $Z \sim N(0,1)$ (standard normalfordeling).

Solution:

$$z \approx 1{,}645$$

Oppgåve 3 . . . . *(6%)*

Ein stokastisk variabel X er binomialfordelt med n forsøk og punktsannsyn $\pi = 0{,}3$. Svar på fylgjande spørsmål, og forklar korleis du kjem fram til svaret.

(a) Lat $n = 6$, og finn $P(X \leq 2)$.

Solution:

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{6}{0} \cdot 0{,}3^0 \cdot 0{,}7^6 + \binom{6}{1} \cdot 0{,}3^1 \cdot 0{,}7^5 + \binom{6}{2} \cdot 0{,}3^2 \cdot 0{,}7^4 \\ &= 0{,}117649 + 0{,}302526 + 0{,}324135 \\ &= 0{,}74431 \end{aligned} \tag{3}$$

(b) Lat $n = 180$, og finn $P(X \leq 60)$.

Solution: Fordi n er stor, kan me bruka normalfordelinga (sentralgrensesatsen). Me har $\mu = n\pi = 54$ og $\sigma^2 = n\pi(1 - \pi) = 37.8$.
Me normaliserer med formelen

$$Z = \frac{X - \mu}{\sigma}$$

Dette gjev

$$P(X \leq 60) = P(Z \leq \frac{60 - 54}{\sqrt{37.8}}) = P(Z \leq 0{,}9759) \approx 0{,}8365$$

Oppgåve 4 . . . . *(9%)*

Mange dataprogram treng tilfeldige tal.

(a) Gje tre døme på praktiske problem som krev programmering med slumptal.

Solution: T.d. online gambling, kryptografiske nøgler til nettbanker, simulering av økosystem

Ein vanleg måte å generera slike slumptal er den lineære kongruensgeneratoren

$$x_i = a \cdot x_{i-1} + c \pmod{m}.$$

(b) Slumptalsgeneratoren krev eit *frø* (*seed*). Kva meiner me med eit frø og korleis vert det brukt for å gje oss ein serie slumptal frå generatoren over?

Solution: Frøet er starttilstanden for generatoren, dvs. x_0 i dømet over. Kvart slumptal er ein funksjon av føregåande, og ein treng altso eit nullte tal for å kunne starta generatoren.

(c) Korleis ville du ha vald frø i eit praktisk system, og kvifor?

Solution: Ein vanleg teknikk er å bruka dei minst signifikante bitsa i klokketida. Det er enkelt å få til i praksis, relativt portabelt mellom platformar, og det er relativt uforutsigbart (tilfeldig).

Oppgåve 5 .... *(10%)*

Det er vanleg å testa kvaliteten på ein slumptalsgenerator ved hjelp av statistisk hypotesetesting. Lat $[x_1, x_2, \ldots, x_n]$ vera ein heiltalsfylgje frå slumptalsgeneratoren. For å testa at dei tre siste bitsa er uniformt fordelte kan me bruka χ^2-statistikken (kji-kvadrat-statistikken)

$$G = \sum_{i=0}^{7} \frac{(F_y - E_y)^2}{E_y}$$

der $y_i = x_i \mod 8$, og F_y er frekvensen av kvar mogleg verdi av y.

(a) Kva er nullhypotesen for testen?

> **Solution:** Nullhypotesen er at kvart slumptal er uniformt fordelt.
> (Hypotesen kan formulerast på fleire ulike måtar.)

(b) Kva fordeling har G under nullhypotesen? Forklar svaret. Hugs å spesifisera talet på fridomsgradar.

> **Solution:** χ^2-fordelt med sju fridomsgradar, siden n=8.

(c) Kva er den kritiske verdien for testen ved 5% signifikansnivå?

> **Solution:** Vi finner i tabellen for kji-kvadratfordelingen at den kritiske verdien er lik 14,067. D.v.s.: Om X er kji-kvadratfordelt med 7 frihetsgrader, er $P(X > 14{,}067) = 0{,}05$.

(d) Kva er forventa frekvens E_y i denne testen?

> **Solution:** $E_y = 1/8$ for alle y

(e) Forklar korleis me kan bruka χ^2-testen til å testa om påfylgjande verdiar i slumptalsfylgja er uavhengig fordelte.

> **Solution:** Testen over ser på tre bits frå eitt slumptal. Me kan like gjerne ta tre bits frå tre påfylgjande tal. Då treng me tre gongar so mange tal frå fylgja for å få same statistiske signifikans. Fylgja kan delast i triplar, der me tek t.d. minst signifikante bit og les det som tal frå 0 til 7 frå kvar trippel. Resten av testen går som over.

Oppgåve 6 .... *(6%)*

I mange formar for simulering, vert landskapet modellert som eit rutenett eller raster (*grid*). Ein spesiell situasjon oppstår ved yttergrensa i landskapet. Dette kan ein handtera på ulike måtar. Forklar kort (gjerne supplert med ei teikning) kva me meiner med kvar av dei fylgjande randvilkåra:

(a) periodisk

> **Solution:** Periodiske randvilkår modellerer verda som ei kule eller ein torus (smultring). Alle rutane har naboar på alle kantar. Randrutane har ein nabo på motsett side av landskapet.

(b) reflekterende

> **Solution:** Her har ogso alle rutane naboar på alle kantar. Når ein går over kanten, sprett ein i staden tilbake i motsett retning. Randruten har altso samme rute som nabo på to sider.

(c) Kva slags randvilkår er best egna for simulering av korleis fisk sym i eit akvarium? Forklar svaret.

> **Solution:** Periodiske randvilkår er best egna for å simulera ein liten del av eit stort og relative einsarta område. Her har me eit lite og innelukka område. Reflekterande randvilkår er eit høveleg val, sidan det truleg er naturleg for fisken å snu når han treff kanten. Ein kan òg vurdera å modellera veggen som ein vegg, som fisken faktisk er medviten om, men det krev ein meir avansert fiskeoppførsel.

Oppgåve 7 .... *(5%)*

Av 15000 gongar me køyrer ei rovdyr-byttedyrsimulering, døyr byttedyrpopulasjonen ut 10 gongar. Har me statistisk grunnlag for å seia at sannsynet for at byttedyrpopulasjonen døyr ut er mindre enn 0,1 %? Gjennomfør hypotesetesten på 5 % signifikansnivå.

> **Solution:** Vi gjennomfører en ensidig hypotesetest.
>
> $$\begin{aligned} H_0 &: \pi = 0{,}1\% \\ H_1 &: \pi < 0{,}1\% \\ z_{obs} &= \frac{x_{obs} - n \cdot \pi_0}{\sqrt{n \cdot \pi_0 (1 - \pi_0)}} = \frac{10 - 15000 \cdot 0{,}1\%}{\sqrt{15000 \cdot 0{,}1\% \cdot 0,999}} = \frac{-5}{3{,}87} = -1{,}292 \\ z_{kritisk} &= -1.645 < -1{,}292 = z_{obs} \end{aligned} \tag{4}$$
>
> Derfor kan vi ikke forkaste H_0, og beholder vi $H_0 : \pi = 0{,}1$

Oppgåve 8 .... *(12%)*

Ein *random walker* flyttar seg på eit 1D raster. Lat den stokastiske variabelen X være forflytningen til partikkelen i løpet av et tidssteg. Sannsynlighetsfordelingen for X er:

$$P(X = -2) = P(X = 2) = 0,1 \tag{5}$$

$$P(X = -1) = P(X = 1) = 0,4. \tag{6}$$

$$\tag{7}$$

(a) Rekn ut populasjonsstandardavviket for posisjonen etter eitt tidssteg.

> **Solution:**
>
> $$\begin{aligned} \mu &= \sum x \cdot p(x) = -2 \cdot 0{,}1 + 2 \cdot 0{,}1 + -1 \cdot 0{,}4 + 1 \cdot 0{,}4 = 0 \\ \sigma^2 &= \sum (x - \mu)^2 \cdot p(x) \\ &= (-2 - 0)^2 \cdot 0{,}1 + (2 - 0)^2 \cdot 0{,}1 + (-1 - 0)^2 \cdot 0{,}4 + (1 - 0)^2 \cdot 0{,}4 \\ &= 0{,}4 + 0{,}4 + 0{,}4 + 0{,}4 = 1{,}6 \end{aligned} \tag{8}$$
>
> Vi finner $\sigma = \sqrt{1{,}6} = 1{,}265$

(b) Rekn ut populasjonsstandardavviket for posisjonen etter 100 tidssteg.

Solution:

$$\begin{aligned}\sigma_{100}^2 =& 100 \cdot \sigma_1^2 = 100 \cdot 1{,}6 = 160 \\ \Rightarrow \sigma_{100} =& \sqrt{160} = 12{,}65. \\ (\text{og } \mu_{100} =& 100 \cdot \mu_1 = 100 \cdot 0 = 0)\end{aligned} \tag{9}$$

(c) Me er interesserte i den eksakte sannsynsfordelinga for posisjonen til partikkelen etter 100 tidssteg. Forklar korleis me kan finna denne sannsynsfordelinga ved hjelp av ei datamaskin.

Solution: Vi bruker en rekursiv algoritme, hvor vi hele tiden bruker sannsynlighetsfordelingen etter i steg for å finne sannsynlighetsfordelingen etter $i+1$ steg. I begynnelsen ($t = 0$) er sannsynligheten lik 1 på posisjon 0 (om det er der partikkelen begynner), og sannsynligheten er lik 0 på alle andre posisjoner. På $t = 1$ blir sannsynlighetsfordelingen: $P(X = -2) = P(X = 2) = 0,1$ og $P(X = -1) = P(X = 1) = 0,4$. Generelt bruker vi: $p(x,t) = 0{,}1 \cdot p(x-2, t-1) + 0{,}4 \cdot p(x-1, t-1) + 0{,}4 \cdot p(x+1, t-1) + 0{,}1 \cdot p(x+2, t-1)$.

(d) Forklar korleis me kan generera eit utval med ti observasjonar av posisjonen etter 100 tidssteg.

Solution: Vi kjører ti simuleringer med lengde 100.

Oppgåve 9 .... *(15%)*

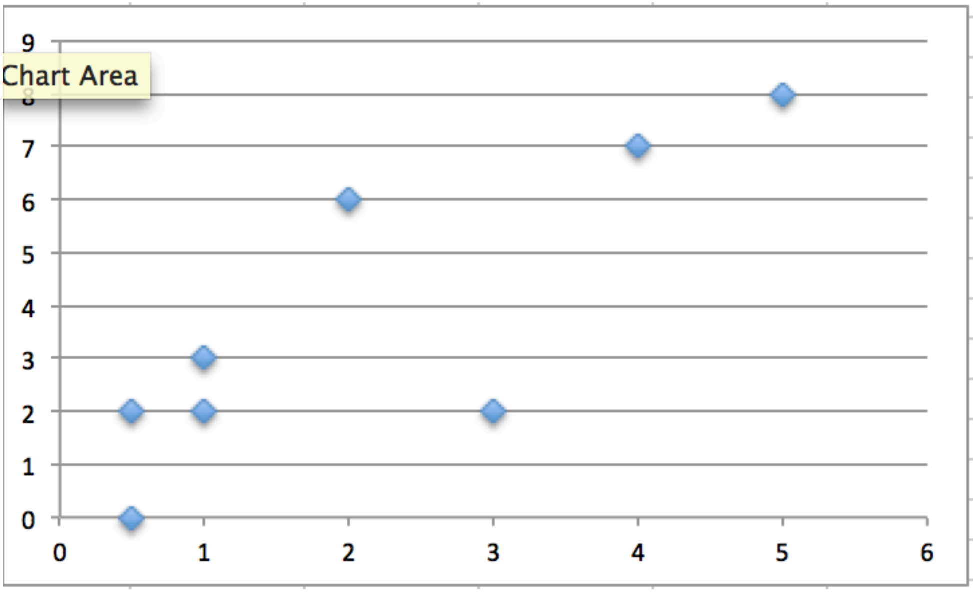
Eit maskinlæringsprogram får ei stor datamengd som inndata. Målet er å studera avhengigheiter mellom ulike variablar. Eit tilfeldig utval av datasettet for variablane X og Y er gjeve i tabellen under.

X	Y
2	6
3	2
5	8
4	7
0,5	0
1	3
0,5	2
1	2

For datasettet gjeld:

$$\sum x = 17$$
$$\sum y = 30$$
$$\sum x^2 = 56,5$$
$$\sum y^2 = 170$$
$$\sum xy = 92$$

(a) Skisser datapunkta i eit plott.



Solution:

(b) Er der korrelasjon mellom X og Y? Forklar svaret.

Solution: Det ser slik ut fra plottet, siden datapunktene ligger rundt en linje med stigningstall ulik 0 eller ∞. Men vi må huske at dette bare er et utvalg fra populasjonen, så vi kan ikke si med sikkerhet at X og Y er korrelerte.

(c) Kva kan me seia om avhenget mellom X og Y? Forklar svaret.

Solution: Siden X og Y er korrelerte, sier vi at det er en avhengighet mellom disse variablene. Men det betyr ikke automatisk at det finnes en *logisk* avhengighet mellom X og Y. F. eks. kan både X og Y være styrt av en tredje variabel.

(d) Bruk minste kvadratsum-metoden for å finna den beste rette lina for å beskriva datasettet.

Solution:

$$\begin{aligned} b =& \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \\ =& \frac{8 \cdot 92 - 17 \cdot 30}{8 \cdot 56{,}5 - 17^2} = 1{,}39 \\ a =& \bar{y} - b \cdot \bar{x} = \frac{30}{8} - 1{,}39 \cdot \frac{17}{8} = 0{,}79 \end{aligned} \tag{10}$$

Vi finner derfor regresjonslinja $y = 0{,}79 + 1{,}39 \cdot x$

(e) Rekn ut korrelasjonskoeffisienten mellom X og Y.

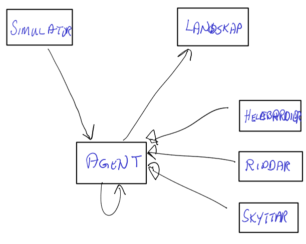
Solution:

$$\begin{aligned} S_{xx} =& \sum x^2 - n\bar{x}^2 = 56{,}5 - 8 \cdot (\frac{17}{8})^2 = 20{,}375 \\ S_{yy} =& \sum y^2 - n\bar{y}^2 = 170 - 8 \cdot (\frac{30}{8})^2 = 57{,}5 \\ S_{xy} =& \sum xy - n\bar{x}\bar{y} = 92 - 8 \cdot \frac{17}{8} \cdot \frac{30}{8} = 28{,}25 \\ r =& \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{28{,}25}{\sqrt{20{,}375 \cdot 57{,}5}} = 0{,}825 \end{aligned} \tag{11}$$

Oppgåve 10 . . . . *(18%)*

Tenk deg at me vil simulera eit feltslag mellom to hærar vha. ein agentbasert modell. Kvar hær har 2-3 ulike typar soldatar (t.d. riddar, fotsoldat med helebard, bogeskyttar) med ulike eigenskapar. Slagmarken kan me modellera som eit rutenett (raster, *grid*).

(a) Teikn ei skisse til klassediagram for ein *enkel* simulator.



Solution:
Det meste vesentlege i svaret er at ein har ei simulatorklasse og ein generisk agentklasse. Kvar soldattype må vera ein underklasse av agent. I tillegg er det naturleg å ha ei klasse for slagmarken. Dersom ein ynskjer visualisering, er det ein separat klasse. Prosjektilar (pilar) kan vera ein agent, men treng ikkje vera det; det avheng av detaljnivået i simuleringa.

(b) Forklar kort funksjon og formål for kvar av klassene i skissa di.

Solution: Agent-klassen er ei abstrakt klasse. Alle agentane er ansvarlege for sin eigen oppførsel.

Grid definerer landskapet og held styr på kvar agentane er. Eit grid-objekt kan òg fortelja agentane kva dei ser.
Viz representerer all den grafiske visualiseringa av simuleringa.
Simulator har ansvaret for å instantiera modellen og styrer klokka for å fortelja agentane når dei skal handla.

(c) Drøft: kva klasser er spesielle for feltslagmodellen, og kva kan gjenbrukast på andre problem?

Solution: Alle soldattypane er spesielle. Oppførselen må modellerast ut frå kva soldatar gjer i eit slag.
Simulatorklassa er openbert gjenbrukbar.
Der skal ikkje liggja noko domenespesifikk kode der. Lands

(d) Kva klasser i klassediagrammet ditt er agentar?

Solution: Alle klassane som arvar agent er agentar.

(e) Kva eigenskapar (tilstand og oppførsel) er felles for alle agentane i systemet?

Solution:

(a) Dei har ein plass i landskapen.

(b) Dei handlar kvar runde.

(c) Dei ser etter mål dei kan skada.

(d) Dei slåss, skadar andre agentar og vert skada.

(f) Gje to døme på at ulike agenttypar i modellen din har (eller kan ha) ulike eigenskapar.

Solution: Bogeskyttaren ser etter mål på avstand, og skadar dei, og vil freista å halda avstand.
Helebardsoldatane og riddarane vil nærma seg måla og skada dei på kloss hold.
Riddarane flyttar seg raskare enn andre.