Eksamen 26. mai 2017 **IR201812 Statistikk og Simulering**
Eksamenstid 4 timar *Nynorsk*

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?
(b) Kva er sannsynsfordelinga til X?
(c) Rekn ut forventningsverdien for X.
(d) Rekn ut populasjonsvariansen for X.
(e) Kva er den kumulative sannsynsfordelinga for X?

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

(f) Rekn ut utvalsmiddelverdien.
(g) Rekn ut utvalsvariansen.

Oppgåve 2 .... *(4%)*

Bruk sannsynstabellane i boka og finn fylgjande:

(a) $P(T > 1{,}5)$ der T har Students t-fordeling med fem fridomsgradar.
(b) Den kritiske verdien z slik at $P(Z < z) = 0{,}95$ der $Z \sim N(0, 1)$ (standard normalfordeling).

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 $P(X \leq 2)$.
(b) Lat $n = 180$, og finn $P(X \leq 60)$.

Oppgåve 4 .... *(9%)*

Mange dataprogram treng tilfeldige tal.

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

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

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?

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

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

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

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

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

(b) reflekterende

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

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.

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.

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

(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.

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

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.

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

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

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

(e) Rekn ut korrelasjonskoeffisienten mellom X og Y.

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.

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

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

(d) Kva klasser i klassediagrammet ditt er agentar?

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

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