2. Obligatoriske Opgave

Anvendt Statistik - forårsemesteret 2001.

Følgende opgave er den sidste af to, som er obligatoriske for at kunne gå til eksamen i kurset *Anvendt Statistik*. Den vil blive udlevet onsdag d. 18. april 2001 efter dagens forelæsning, og en skriftelig besvarelse skal være undertegnede i hænde senest onsdag d. 2. maj 2001. Det er tilladt at arbejde sammen om opgaven, dog skal der afleveres separate løsninger. Til eksamen vil samarbejde dog ikke være tilladt.

Der ikke blot tillades men anbefales kraftigt at bruge computere og de programmer I har arbejdet med samt modifikationer deraf, og i visse opgaver vil brugen af computere være krævet.

God arbejdslyst, Troels

I – Fordelinger og sandsynlighed:

1.1 Lad x være fordelt efter sandsynlighedsfordelingen $f(x) = x^2$ i intervallet $[0, C]$.

- For hvilken værdi af C er sandsynlighedsfordelingen $f(x)$ normeret?
- Hvad er forventningsværdien af e^{-x}?
- Sammenlign denne med $e^{-E[x]}$.

1.2 En stråle af partikler indeholder 10^{-4} andele elektroner og resten fotoner. Partiklerne passerer gennem en detektor med to lag, som giver signalerne 0, 1 eller 2 alt efter hvor mange lag den passerende partikel detekteres i. Sandsynlighederne for disse signaler er for elektroner (e) og fotoner (γ) som følger:

$$\begin{array}{lcl} P(0|e) = 0.001 & \text{and} & P(0|\gamma) = 0.99899 \\ P(1|e) = 0.01 & & P(1|\gamma) = 0.001 \\ P(2|e) = 0.989 & & P(2|\gamma) = 0.00001 \end{array}$$

- Hvad er sandsynligheden for, at en partikel der detekteres i et lag, er en foton?
- Hvad er sandsynligheden for, at en partikel der detekteres i to lag, er en elektron?

1.3 Beregn middelværdi og spredning af følgende fordeling: $f(x) = \ln(x)$, $x \in [1, e]$.

II – Fejlberegning:

2.1 En studerende har målt lysets hastighed ti gange og fået følgende resultater:

Måling	1	2	3	4	5	6	7	8	9	10
Resultat (10^8 m/s)	3.61	2.00	3.90	2.23	2.32	2.48	2.43	3.86	4.43	3.78

- Hvad er gennemsnittet og usikkerheden på gennemsnittet?
- Hvis $m = 1.02$, $\sigma_m = 0.05$ og $\rho_{mc} = -0.2$, hvad er så fejlen på $E = mc^2$?

2.2 Lad $x = \ln(y)/\sqrt{z}$ og lad y, z være korrelerede med kovarians V_{yz} og have middelværdi μ_y, μ_z og spredning σ_y, σ_z, henholdsvis.

- Hvad er middelværdien μ_x og spredningen σ_x på x?

III – Monte Carlo: (Til denne del skal der bruges computere. Plots kan vedlægges opgavebesvarelsen).

3.1 Lad $f(x) = e^{-x^3+2x^2} - 1$ være proportional med en sandsynlighedsfordeling for $x \in [0, 2]$.

- Hvilken metode bør benyttes til at generere denne fordeling? Forklar.
- Lav en algoritme, der udfra en jævn fordeling i intervallet $[0, 1]$, genererer tal efter sandsynlighedsfordelingen $f(x)$.
- Bestem $\int_0^2 f(x)dx$ og dennes usikkerhed ved at benytte denne algoritme. Dermed kan $f(x)$ normeres.

3.2 Lav en Monte Carlo algoritme, som simulerer 500 kast med to terninger.

- Lav et histogram over frekvensen af kast for hver mulig sum.
- Plot den statistiske usikkerhed samt den forventede fordeling.
- Udregn χ^2 for overensstemmelsen mellem forventningen og data, og bestem sandsynligheden for dette χ^2.

Statistiske tests:

4.1 Ladede partikler, der går gennem en gas, producerer ionisation, mængden af hvilken afhænger af partiklens type. Antag at en statistik, t, baseret på mængden af ionisation er konstrueret således, at den er fordelt Gaussisk omkring 0 for elektroner og 2 for pioner, begge fordelinger med enhedsstandardafvigelser. En test bliver konstrueret til at udvælge elektroner ved at kræve at $t < 1$.

- Hvad er signifikansniveauet for denne test, dvs. ssh. for at acceptere en elektron?
- Hvad er testens evne til at fravælge pioner?
- Antag at en mængde partikler vides at bestå af 99 % pioner og 1 % elektroner. Hvad er så renheden af den mængde elektroner der udvælges af kriteriet $t < 1$?
- Hvis man kræver en mængde elektroner med renhed 95 %, hvilket kriterium for t skal man så kræve? Hvad er effektiviteten for udvælgelse af elektroner ved denne værdi af t?

Fitting data:

5.1 Et eksperiment har givet følgende resultat, hvor usikkerheden på y, σ_y, vurderes til 0.2:

x	y	x	y	x	y	x	y
0.5	2.17	3.0	3.73	5.5	4.63	8.0	6.58
1.0	2.61	3.5	3.64	6.0	4.93	8.5	6.92
1.5	2.93	4.0	3.66	6.5	5.44	9.0	7.04
2.0	2.87	4.5	4.04	7.0	5.58	9.5	7.77
2.5	2.92	5.0	4.43	7.5	5.88	10.0	8.26

- Antag en lineær sammenhæng mellem x og y, og lav et χ^2-fit til data.
- Beregn udfra χ^2 og antallet af frihedsgrader sandsynligheden for at opnå et lignende eller højere χ^2 givet den lineære hypotese. Er fittet godt?
- Forsøg med andre hypoteser for sammenhængen mellem x og y, og diskuter deres validitet.

Bonus opgave:

6.1 Betragt en binomial variable med sandsynlighed p og $q = 1-p$ for udfaldet af hvert forsøg, og konstruer asymmetrien $\alpha = p - q$. Hvis man vil måle en meget lille asymmetry, fx. 10^{-3}, hvor mange forsøg skal man så udføre, for at kunne give en 3σ udtalelse?

What we have to learn to do, we learn by doing. [Aristotle, 384-322 bc]