

Anvendt Statistik

Opgave i statistiske grundbegreber

Følgende opgave er en gennemgang af de statistiske grundbegreber, som vi har gennemgået indtil nu. Den vil blive udlevet tirsdag d. 15. september 2009 efter dagens forelæsning og øvelser, og en skriftelig besvarelse skal være undertegnede i hænde senest fredag d. 18. september 2009. Det er tilladt at arbejde sammen om opgaven, dog skal der afleveres separate løsninger.

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 nødvendig.

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 middelværdien og spredningen på x ?

1.2 Lille Per går på casino og spiller på et nummer ad gangen ($p = 1/37$). Hvis han ikke snyder, hvad er så sandsynligheden for at han ud af 100 spil vil vinde mere end 3 gange?

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

II – Fejlregning:

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 $x = 2yz + z^2$. Hvis det er givet at $\sigma_y = 0.2$, $\sigma_z = 0.3$ og $V_{yz} = 0.6$, hvad er så usikkerheden på x ?

2.3 En studerende måler lydens hastighed tyve gange og konkluderer at spredningen på resultaterne er 12 m/s. Hvis fejlkilderne er mange og tilfældige, hvilken fordeling skulle målingerne så følge? Hvor mange målinger ville det kræve for, at usikkerheden på middelværdien når ned på 2 m/s?

2.4 Begyndelsesaktiviteten N_0 og levetiden τ af en radioaktiv kilde kendes med en relativ nøjagtighed på 1%. Ved estimering af aktiviteten $N = N_0 e^{t/\tau}$ vil usikkerheden i starten domineres af usikkerheden på N_0 og senere af usikkerheden på τ . For hvilken værdi af t/τ vil de to usikkerheder bidrage ligeligt til usikkerheden på N ?

III – Monte Carlo: (Til denne del anbefales brug af 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 og brug resultatet til at normere $f(x)$.

3.2 Lav en Monte Carlo algoritme, som simulerer 200 kast med to terninger (eller gør det i hånden!).

- 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 .

Estimators:

4.1 I tidens løb har flere grupper studerende hver for sig målt myonens levetid i kælderens på NBI. Deres resultater og estimerede fejl er anført nedenfor:

Gruppe	1	2	3	4	5	6	7	8	9	10
Resultat (μs)	1.82	1.95	1.46	2.12	2.09	1.70	1.93	1.87	2.25	2.16
Usikkerhed (μs)	0.06	0.09	0.12	0.13	0.24	0.11	0.07	0.10	0.21	0.14

- Udregn gennemsnit og spredning af disse målinger, samt deres χ^2 og sandsynligheden for et lignende eller højere χ^2 i både det uvægtede og det vægtede tilfælde.
- Er der en af målingerne, som ikke passer særlig godt? Kvantiser hvorfor!
- Gentag forrige udregning, hvor den mindst sandsynlige måling ikke medtages.
- Hvor godt stemmer resultaterne (uvægtede og vægtede) overens med τ_μ ?

Fitting data:

5.1 Et eksperiment har givet følgende resultat, hvor usikkerheden på y , σ_y , vurderes til 0.06:

x	y	x	y	x	y	x	y
-2.0	-.29	1.0	0.06	4.0	0.63	7.0	0.81
-1.0	-.19	2.0	0.33	5.0	0.89	8.0	1.04
0.0	0.04	3.0	0.57	6.0	0.80	9.0	0.94

- 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 Hvor mange fysikstuderende er der i Danmark?

- Uden anden information, hvilken usikkerhed ville du tilskrive dette antal?
- Hvis en journalist påstod, at antallet af fysikstuderende var steget/faldet med 3% i forhold til året forinden, hvad burde man så fortælle ham?