Anvendt Statistik

Opgaver i pensum for anvendt statistik

Følgende er et opgavesæt i det pensum, som vi har gennemgået indtil nu. Det vil blive udlevet tirsdag d. 13. oktober 2009, og en skriftelig besvarelse skal være undertegnede i hænde senest mandag d. 19. oktober 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, ikke mindst opgave 4, hvor I skal hente det aktuelle elektroniske datasæt.

God arbejdslyst, Troels

I – Fordelinger og sandsynlighed:

1.1 Lille Per slår 13 gange med en almindelig terning og får 7 seksere. Hvad er sandsynligheden for en sådan hændelse eller noget mere ekstremt? Har Lille Per snydt?

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:

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

- 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 Lad x være uniformt fordelt i intervallet $[\alpha, \beta]$, hvor $0 < \alpha < \beta$.

- Hvad er forventningsværdien og variansen af $1/x$?
- Sammenlign forventningsværdien $E[1/x]$ med $1/E[x]$.

1.3 Beregn middelværdi og spredning af følgende fordelinger:

- $f(x) = \frac{1}{2}\sin(x)$, $x \in [0, \pi]$.
- $f(x) = \frac{1}{2}e^{-|x-2|}$, $x \in [-\infty, \infty]$.

II – Fejlberegning:

2.1 Tennisbolde skal officielt veje mellem 56.0 og 59.4 gram. For hvilken middelværdi og usikkerhed vil 90% af tennisbolde ligge indenfor disse krav?
Hvis en serv har en hastighed på 73 m/s med en usikkerhed på 2%, hvad er så boldens kinetiske energi hvis der ingen korrelation er? Og hvis korrelationen er $\rho = -0.7$?

2.2 Hvis $\theta = 0.54 \pm 0.02$, hvad er så fejlen på $\cos\theta$, $\sin\theta$, $\tan\theta$? Hvad hvis $\theta = 1.54 \pm 0.02$?

2.3 Snell's lov siger at $n_1 \sin\theta_1 = n_2 \sin\theta_2$. Find n_2 og dens fejl udfra målingerne:

$$\theta_1 = (22.03 \pm 0.2)^\circ \qquad \theta_2 = (14.45 \pm 0.2)^\circ \qquad n_1 = 1.0000$$

III – Monte Carlo: (Til denne del anbefales brug af computere. Plots kan vedlægges opgavebesvarelsen).

3.1 Lad $f(x) = \frac{1}{2\sqrt{x}}$ være sandsynlighedsfordeling for $x \in [0, 1]$.

- Hvilken metode bør benyttes til at generere denne fordeling? Forklar.
- Lav en algoritme, der udfra en jævn fordeling i intervallet $[0, 1]$, genererer 1000 tal efter sandsynlighedsfordelingen $f(x)$. Udregn gennemsnittet af disse tal samt usikkerheden på dette gennemsnittet. Sammenlign med den analytiske værdi for gennemsnittet.

Estimators:

4.1 Betragt datasættet om tibetanske kraniers størrelser, som findes på:

`http://www.nbi.dk/~petersen/Teaching/Stat2009/Data_TibetanSkulls.txt`

- Hvor stor er middelværdien og spredningen for kranier af type A og type B for hver af de fem variable? Og hvor stor er seperationen (i.e. $\mu_A - \mu_B / \sqrt{\sigma_A^2 + \sigma_B^2}$) mellem type A og type B for hver af de fem variable?
- Hvor stor er den linære korrelation mellem de forskellige variable for hver type kranie?
- Hvor godt kan du separere de to typer af kranier ved hjælp af en Fisher discriminant, henholdsvis uden og med korrelationer inkluderit?

4.2 Betragt datasættet om lederes levetid efter indsættelse, som findes på:

`http://www.nbi.dk/~petersen/Teaching/Stat2009/Data_LifetimeLeaders.txt`

- Er levetiden lige lang for hver af de tre kategorier af ledere?
- Er der med tiden nogen ændring i levetiden for nogen af lederne?

Fitting data:

5.1 Et eksperiment har givet følgende resultat, hvor usikkerheden på y, σ_y, vurderes til at være kvadratroden af antallet samt en systematisk fejl på 2.0 lagt til i kvadratur:

x	y	x	y	x	y	x	y
0.05	13.8	0.45	9.9	0.85	12.8	1.25	12.1
0.15	7.4	0.55	19.6	0.95	8.8	1.35	9.8
0.25	6.5	0.65	23.3	1.05	7.9	1.45	3.4
0.35	7.1	0.75	19.2	1.15	14.6	1.55	11.1

- Antag en lineær sammenhæng mellem x og y, og lav et χ^2-fit til data. Er fittet godt?
- Hvis man forventer et signal et sted mellem 0.6 og 0.8, hvad ville du så fitte med? Diskuter fittets validitet og signalets signifikans.

Those who are good at archery learnt from the bow and not from Yi the Archer. Those who know how to manage boats learnt from boats and not from Wo [the legendary boatman]. Those who can think learnt for themselves and not from the sages.

[Guan Yin, 8th Century]