Øving 4 for FY1004, høsten 2007

Kvantemekanikk er en sannsynlighetsteori, og denne øvingen er et lynkurs i sannsynlighetsregning. Her vil vi ta for oss et eksempel på en *diskret* sannsynlighetsfordeling og et eksempel på en *kontinuerlig* sannsynlighetsfordeling.

Diskret sannsynlighetsfordeling

Terningkast er eksempel på en diskret sannsynlighetsfordeling. Sannsynligheten for et gitt resultat n av et kast kaller vi $P(n)$. En sannsynlighet $P(n)$ er alltid positiv, hvis den ikke er null. Hvis ingen jukser med terningen, eller med kastingen av terningen, kan vi si på forhånd at de seks mulige resultatene, enten $n = 1$, $n = 2$, $n = 3$, $n = 4$, $n = 5$ eller $n = 6$, er like sannsynlige. Summen av alle sannsynlighetene er pr. definisjon lik 1, og når alle sannsynlighetene er like, er $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

Vi definerer *forventningsverdien* av n som

$$\langle n \rangle = \sum_n P(n)\, n = \sum_{n=1}^{6} \frac{1}{6}\, n = \frac{1}{6}\,(1+2+3+4+5+6) = \frac{21}{6} = 3{,}5\ .$$

Navnet «forventningsverdi» høres merkelig ut, fordi vi aldri ville forvente å få resultatet 3,5 når vi kaster terning. Det er heller slik at om vi gjør mange kast, forventer vi å få en *middelverdi* som ligger i nærheten av 3,5.

Vi definerer *variansen* av n som

$$\mathrm{var}(n) = \left\langle (n - \langle n \rangle)^2 \right\rangle = \sum_n P(n)\ (n - \langle n \rangle)^2\ . \tag{1}$$

Variansen er positiv pr. definisjon, og kvadratroten av variansen er et mål på hvor mye vi kan forvente at resultatet av et kast avviker fra forventningsverdien 3,5.

Generelt definerer vi det k-te *momentet* til sannsynlighetsfordelingen som

$$\langle n^k \rangle = \sum_n P(n)\, n^k\ ,$$

og det k-te *sentrale momentet* som

$$\left\langle (n - \langle n \rangle)^k \right\rangle = \sum_n P(n)\, (n - \langle n \rangle)^k\ .$$

a) Utled en ekvivalent formel for variansen,

$$\mathrm{var}(n) = \langle n^2 \rangle - \langle n \rangle^2\ . \tag{2}$$

Beregn variansen i eksemplet vårt med terningkast. Bruk enten definisjonsligningen (1) eller den alternative formelen (2), eventuelt begge for å se at de gir samme resultat.

b) Ti terningkast gir resultatene $n_1 = 5$, $n_2 = 3$, $n_3 = 3$, $n_4 = 2$, $n_5 = 4$, $n_6 = 4$, $n_7 = 3$, $n_8 = 3$, $n_9 = 6$, $n_{10} = 4$. Beregn *middelverdien*, definert slik, med $N = 10$:

$$\overline{n} = \frac{1}{N} \sum_{i=1}^{N} n_i\ ,$$

og *standardavviket* s, definert ved at

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (n_i - \overline{n})^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left({n_i}^2 - \overline{n}^2 \right) = \frac{1}{N-1} \left[\left(\sum_{i=1}^{N} {n_i}^2 \right) - N\,\overline{n}^2 \right] . \tag{3}$$

Hva er *forventningsverdien av middelverdien*? Jo, den er

$$\langle \overline{n} \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle n_i \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle n \rangle = \langle n \rangle .$$

Fordi de N terningkastene er uavhengige, har hver n_i forventningsverdien $\langle n \rangle = 3{,}5$.

Når vi gjør N terningkast og regner ut middelverdien $\overline{n}$, får vi som regel et svar som avviker fra 3,5. Spørsmålet blir da om avviket er større eller mindre enn det burde være?

For å kunne vurdere det, bør vi kjenne *variansen av middelverdien*,

$$\mathrm{var}(\overline{n}) = \langle \overline{n}^{\,2} \rangle - \langle \overline{n} \rangle^2 = \langle \overline{n}^{\,2} \rangle - \langle n \rangle^2 .$$

c) Vis at

$$\langle \overline{n}^{\,2} \rangle = \frac{N-1}{N} \langle n \rangle^2 + \frac{1}{N} \langle n^2 \rangle , \tag{4}$$

og følgelig

$$\mathrm{var}(\overline{n}) = \frac{1}{N} \left(\langle n^2 \rangle - \langle n \rangle^2 \right) = \frac{1}{N} \mathrm{var}(n) . \tag{5}$$

Hint:

$$\langle \overline{n}^{\,2} \rangle = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle n_i n_j \rangle .$$

For $i = j$ er $\langle n_i n_j \rangle = \langle {n_i}^2 \rangle = \langle n^2 \rangle$.
For $i \neq j$ er terningkastene nr. i og nr. j uavhengige, slik at $\langle n_i n_j \rangle = \langle n_i \rangle \langle n_j \rangle = \langle n \rangle^2$.

Hva er *forventningsverdien av standardavviket*? I følge den siste formelen i ligning (3) er

$$\langle s^2 \rangle = \frac{1}{N-1} \left[\left(\sum_{i=1}^{N} \langle {n_i}^2 \rangle \right) - N \langle \overline{n}^2 \rangle \right] .$$

Her er $\langle {n_i}^2 \rangle = \langle n^2 \rangle$, igjen fordi hvert terningkast er uavhengig av alle de andre. I tillegg bruker vi ligning (4), og får at

$$\langle s^2 \rangle = \frac{1}{N-1} \left[N \langle n^2 \rangle - (N-1) \langle n \rangle^2 - \langle n^2 \rangle \right] = \langle n^2 \rangle - \langle n \rangle^2 = \mathrm{var}(n) .$$

Hvis vi gjør et eksperiment der vi ikke kjenner sannsynlighetsfordelingen på forhånd, for eksempel hvis vi kaster terning og tviler på at sannsynlighetene faktisk er like store, så kjenner vi hverken forventningsverdien $\langle n \rangle$ eller variansen $\mathrm{var}(n)$. Det beste vi kan gjøre i så fall, er å gjenta eksperimentet et passende antall N ganger. Så kan vi bruke middelverdien $\overline{n}$ som et *estimat* av den ukjente forventningsverdien $\langle n \rangle$. Og vi kan bruke s^2, kvadratet av standardavviket s, som et estimat av den ukjente variansen $\mathrm{var}(n)$.

Ligning (5) er et nyttig resultat. Den forteller hvordan vi kan forbedre et måleresultat som har statistiske feil ved å gjøre flere målinger og beregne middelverdien. Siden usikkerheten er kvadratroten av variansen, ser vi at den statistiske usikkerheten i en middelverdi av N målinger er $1/\sqrt{N}$ av usikkerheten i en enkelt måling. For eksempel må vi gjøre fire ganger så mange målinger om vi vil halvere usikkerheten. Og vi må gjøre hundre ganger så mange målinger om vi vil redusere usikkerheten med en faktor ti.

d) Kan vi stole på den terningen som kastes ti ganger og gir resultatene under punkt b)?

   Vurder det ved å sammenligne middelverdien $\overline{n}$ med forventningsverdien 3,5 for en pålitelig terning. Husk at standardavviket på middelverdien er $s/\sqrt{10}$.

   Hva blir din konklusjon?

Kontinuerlig sannsynlighetsfordeling

Hvis x er en variabel som kan ta kontinuerlige verdier, så er sannsynligheten for at x skal ta en helt bestemt verdi, alltid lik null, fordi det fins så uendelig mange mulige verdier å velge mellom. Men vi kan spørre etter sannsynligheten for å finne en verdi av x i et infinitesimalt intervall mellom x og $x + \mathrm{d}x$. Den er gitt av en *sannsynlighetstetthet* $f(x)$ som

$$\mathrm{d}P = f(x)\,\mathrm{d}x\ .$$

Sannsynligheter er alltid positive, det vil si at $f(x) \geq 0$. Dessuten er summen av alle sannsynlighetene lik 1, det vil si at

$$\int_{-\infty}^{\infty} \mathrm{d}x\ f(x) = 1\ .$$

Forventningsverdien defineres på tilsvarende måte for en kontinuerlig variabel x som for en diskret variabel n, nemlig:

$$\langle x \rangle = \int_{-\infty}^{\infty} \mathrm{d}x\ f(x)\,x\ .$$

Variansen av x defineres som

$$\mathrm{var}(x) = \left\langle (x - \langle x \rangle)^2 \right\rangle = \int_{-\infty}^{\infty} \mathrm{d}x\ f(x)\ (x - \langle x \rangle)^2\ .$$

Fremdeles gjelder formelen $\mathrm{var}(x) = \langle x^2 \rangle - \langle x \rangle^2$, og beviset er det samme som før.

e) Et enkelt eksempel på en kontinuerlig sannsynlighetsfordeling er når x tillates å variere for eksempel mellom 0 og 1, slik at alle verdier i dette intervallet er «like sannsynlige». En slik sannsynlighetsfordeling kalles *uniform* i intervallet $[0, 1]$, den har $f(x) = 1$ for $x \in [0, 1]$, mens $f(x) = 0$ for $x < 0$ og $x > 1$.

   Beregn forventningsverdi og varians for denne fordelingen.

f) Konkret kan vi lage en slik sannsynlighetsfordeling for eksempel ved å starte en stoppeklokke, la den gå noen sekunder, stoppe den igjen og lese av brøkdelen av et sekund.

   Her er en serie på ti tall mellom 0 og 1 generert med stoppeklokke:
   0,97; 0,95; 0,64; 0,45; 0,00; 0,87; 0,73; 0,73; 0,83; 0,16.

   Beregn middelverdi og standardavvik for denne tallserien, samt standardavviket på middelverdien.

   Stemmer middelverdien overens med forventningsverdien for den uniforme sannsynlighetsfordelingen?