Løsning til øving 4 for FY1004, høsten 2007

a) Vi definerer *variansen* av n som

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

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.

At summen av sannsynlighetene er 1, betyr at forventningsverdien av 1 er 1, altså: $\langle 1 \rangle = 1$. Videre bruker vi at forventningsverdien er en lineær funksjon av argumentet. Det vil si: hvis a og b er konstanter, og f og g er statistiske variable, så er

$$\langle af + bg \rangle = a \langle f \rangle + b \langle g \rangle\ .$$

Det følger at

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

For terningkastene gjelder at $\langle n \rangle = 21/6 = 7/2$, $\langle n^2 \rangle = 91/6$, og

$$\mathrm{var}(n) = \langle n^2 \rangle - \langle n \rangle^2 = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12} = 3 - \frac{1}{12} = 2{,}91667\ .$$

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

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

Middelverdi:

$$\overline{n} = \frac{37}{10} = 3{,}7\ .$$

Standardavvik:

$$s^2 = \frac{121}{90} = 1{,}3444\ , \qquad s = \sqrt{s^2} = 1{,}1595\ .$$

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 $\langle n_i n_j \rangle = \langle n_i \rangle \langle n_j \rangle = \langle n \rangle^2$.

Hintet gir ganske direkte at

$$\langle \overline{n}^{\,2} \rangle = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle n_i n_j \rangle = \frac{N-1}{N} \langle n \rangle^2 + \frac{1}{N} \langle n^2 \rangle = \left(1 - \frac{1}{N} \right) \langle n \rangle^2 + \frac{1}{N} \langle n^2 \rangle \;.$$

I dobbeltsummen over i og j har vi nemlig N^2 ledd tilsammen, derav N ledd med $i = j$ og $N^2 - N = N(N-1)$ ledd med $i \neq j$.

Innsatt i formelen $\mathrm{var}(\overline{n}) = \langle \overline{n}^{\,2} \rangle - \langle n \rangle^2$ gir det at

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

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?

Vi fant en middelverdi for ti kast på $\overline{n} = 3{,}7$, som er 0,2 større enn forventingsverdien 3,5.

Vi fant dessuten et standardavvik på $s = 1{,}16$, slik at standardavviket på middelverdien skulle være $s/\sqrt{10} = 0{,}37$. Avviket mellom middelverdien og forventningsverdien er altså litt mer enn et halvt standardavvik: $0{,}2/0{,}37 = 0{,}55$. Ethvert avvik mindre enn to standardavvik er innenfor det akseptable, så terningen besto denne testen.

Merk at vi finner et standardavvik i kvadrat, $s^2 = 1{,}34$, som er mindre enn halvparten av variansen for en sannsynlighetsfordeling med like sannsynligheter, $\mathrm{var}(n) = 2{,}92$. Om et slikt avvik er innenfor det akseptable, eller ikke, kan vi ikke si før vi har beregnet variansen av s^2. Det er ingen ende på hvilke forventningsverdier og varianser vi kunne regne ut, når vi først er i gang!

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.

Forventningsverdi:

$$\langle x \rangle = \int_{-\infty}^{\infty} \mathrm{d}x\ f(x)\ x = \int_0^1 \mathrm{d}x\ x = \left.\frac{x^2}{2}\right|_{x=0}^{1} = \frac{1}{2}\ .$$

Selvfølgelig, hva ellers? 1/2 er midt mellom yttergrensene 0 og 1.

Varians:

$$\begin{aligned} \mathrm{var}(x) &= \int_{-\infty}^{\infty} \mathrm{d}x\ f(x) \left(x - \frac{1}{2}\right)^2 = \int_0^1 \mathrm{d}x\ \left(x^2 - x + \frac{1}{4}\right) \\ &= \left.\left(\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right)\right|_{x=0}^{1} = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}\ . \end{aligned}$$

Dette resultatet kan forresten sammenlignes med punkt a), om vi strekker intervallet $[0, 1]$ så det blir 6 ganger så langt. La f.eks. $f(x) = 1/6$ for $x \in [0{,}5\, , \, 6{,}5]$. Det gir forventningsverdi 3,5, det samme som for terningkast. Variansen blir $6^2 = 36$ ganger så stor for en uniform fordeling i et 6 ganger så langt intervall, altså 36/12, svært nær de 35/12 vi fant for terningkastene.

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.

Here 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?

Middelverdi:

$$\overline{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 0{,}6330\ .$$

Standardavvik i kvadrat:

$$s^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \overline{x})^2 = 0{,}1095\ .$$

Dette tallet skal sammenlignes med den teoretiske variansen $\mathrm{var}(x) = 1/12 = 0{,}0833$. Igjen mangler vi en beregning av variansen av s^2, men ≪på øyemål≫ kan vi kanskje si oss fornøyd med overensstemmelsen?

Standardavviket på serien av ti tall er $s = \sqrt{s^2} = 0{,}3310$, og standardavviket på middelverdien er $s/\sqrt{10} = 0{,}1047 = 0{,}10$. Avviket mellom middelverdien 0,63 og forventningsverdien 0,5 er 0,13, altså 1,3 standardavvik, som er godt innenfor det akseptable.