LÖSNINGAR TILL TENTAMEN 2018-10-31
Sannolikhetsteori och statistik TMS137

Examinator:
J. Björnberg

1. (a) (i) sannolikheten för KANNA är $\boxed{\frac{2}{8}(\frac{3}{8})^4 \approx 0.00494}$.

(ii) antal sätt att få rätt bokstäver är $\frac{5!}{2!2!} = 30$, så sannolikheten blir

$$\boxed{30\tfrac{2}{8}(\tfrac{3}{8})^4 \approx 0.148}.$$

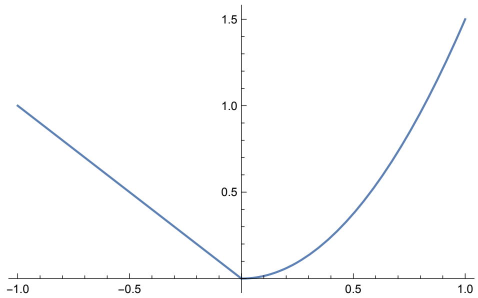
(b) (i) nu blir sannolikheten för KANNA $\boxed{\frac{2}{8}\frac{3}{7}\frac{3}{6}\frac{2}{5}\frac{2}{4} = \frac{3}{280} \approx 0.0107}$.

(ii) antal sätt att få rätt bokstäver är fortfarande $\frac{5!}{2!2!} = 30$ och alla varianter är lika sannolika, så svaret blir

$$\boxed{30\tfrac{3}{280} \approx 0.321}.$$

Alternativt kan man använda g/m där $m = \binom{8}{5}$ och $g = 2 \cdot 3 \cdot 3$ vilket ger samma svar.

2. Täthetsfunktionen har grafen



(a)

$$E(X) = \int_{-1}^{1} tf(t)dt = \int_{-1}^{0} (-t^2)dt + \int_{0}^{1} \tfrac{3}{2}t^3 dt = \boxed{\tfrac{1}{24} \approx 0.0417}.$$

(b)

$$\boxed{F(t) = \int_{-\infty}^{t} f(s)ds = \begin{cases} 0, & \text{om } t \leq -1, \\ \frac{1}{2}(1-t^2), & \text{om } -1 < t \leq 0, \\ \frac{1}{2}(1+t^3), & \text{om } 0 < t \leq 1, \\ 1, & \text{om } t > 1. \end{cases}}$$

(c) Undre kvartilen $x_{0.75}$ ges av $F(x_{0.75}) = 0.25$ och övre kvartilen $x_{0.25}$ av $F(x_{0.25}) = 0.75$. Eftersom $F(0) = 0.5$ så gäller att $x_{0.75} < 0$ och $x_{0.25} > 0$. Lösning av $\frac{1}{2}(1-t^2) = \frac{1}{4}$ ger $\boxed{x_{0.75} = -1/\sqrt{2} \approx -0.707}$. Lösning av $\frac{1}{2}(1+t^3) = \frac{3}{4}$ ger $\boxed{x_{0.25} = 1/2^{1/3} \approx 0.794}$.

(d)

$$E(g(X)) = \int_{-1}^{1/2} tf(t)dt + 0 = \int_{-1}^{0} (-t^2)dt + \int_{0}^{1/2} \tfrac{3}{2}t^3 dt$$
$$= \boxed{-\tfrac{119}{384} \approx -0.31}.$$

3. (a) $P(\text{två kast}) = P(X \le 3) = \boxed{\frac{1}{2}}$.

   (b)

$$
\begin{aligned}
P(Y=1) &= P(X \le 3, Y=1) = \tfrac{1}{2}\tfrac{1}{6} = \boxed{\tfrac{1}{12} \approx 0.083}. \\
P(Y=6) &= P(X \ge 4, Y=6) + P(X \le 3, Y=6) \\
&= P(X=6) + P(X \le 3, Y=6) = \tfrac{1}{6} + \tfrac{1}{2}\tfrac{1}{6} = \boxed{\tfrac{1}{4}}.
\end{aligned}
$$

   (c)

$$
\begin{aligned}
P(X=Y) &= P(X \le 3, Y=X) + P(X \ge 4) \\
&= P(X=1, Y=1) + P(X=2, Y=1) + P(X=3, Y=1) + \tfrac{1}{2} \\
&= \tfrac{1}{2} + 3(\tfrac{1}{6})^2 = \boxed{\tfrac{7}{12} \approx 0.583}.
\end{aligned}
$$

   (d) $E(Y) = 1 \cdot \frac{1}{12} + 2 \cdot \frac{1}{12} + 3 \cdot \frac{1}{12} + 4 \cdot \frac{1}{4} + 5 \cdot \frac{1}{4} + 6 \cdot \frac{1}{4} = \boxed{\frac{17}{4} = 4.25}$.

4. (a) Låt X_i beteckna antal vinster under vecka i. Då gäller att $X_i \sim$Bin(300, p), med väntevärde $\boxed{E(X_i) = 300p = 0.03}$.

   (b) Låt $X = X_1 + X_2 + \cdots + X_{52}$ vara totalt antal vinster under ett år. Då gäller att $X \sim$Bin($300 \cdot 52$, p)$\approx$Po(1.56). Så

$$
P(X \ge 3) = 1 - P(X \le 2) \approx 1 - e^{-1.56}(1 + 1.56 + \tfrac{1}{2}(1.56)^2)\boxed{\approx 0.206}.
$$

   (c) λ ger snittantal vinster per tidsenhet. Så man ska ta $\boxed{\lambda = 0.03}$.

   (d) Mars består av vecka 1 till och med 4, april vecka 5 till och med 8, och maj vecka 9 till och med 12.

$$
P(N(4) \ge 1 \mid N(12) = 3) = 1 - P(N(4) = 0 \mid N(12) = 3),
$$

   där, med $\lambda = 0.03$,

$$
\begin{aligned}
P(N(4) = 0 \mid N(12) = 3) &= \frac{P(N(4) = 0, N(12) - N(4) = 3)}{P(N(12) = 3)} \\
&= \frac{e^{-4\lambda} \cdot e^{-8\lambda}(8\lambda)^3/3!}{e^{-12\lambda}(12\lambda)^3/3!} \\
&= \left(\frac{8\lambda}{12\lambda}\right)^3 = (\tfrac{2}{3})^3 \approx 0.296.
\end{aligned}
$$

   Så svaret är $\boxed{1 - (\frac{2}{3})^3 \approx 0.704}$.

5. Låt X_i vara leveranstiden vid leverans nr i, och låt Y_i vara motsvarande pris för leveransen. Så X_i har fördelning Exp($\lambda = \frac{1}{15}$) och

$$
Y_i = \begin{cases} 50, & \text{om } X_i \le 10, \\ 20, & \text{om } 10 < X_i \le 20, \\ 0, & \text{om } X_i > 20. \end{cases}
$$

   (a) Här gäller att

$$
Y_i = \begin{cases} 50 & \text{med sannolikhet } P(X_i \le 10) = 1 - e^{-10/15} = 0.487, \\ 20 & \text{med sannolikhet } P(10 < X_i \le 20) = 0.249, \\ 0 & \text{med sannolikhet } P(X_i > 20) = e^{-20/15} = 0.264. \end{cases}
$$

Så

$$\mu = E(Y_i) = 50 \cdot 0.487 + 20 \cdot 0.249 = \boxed{29.33}$$

och $E(Y_i^2) = 1317.1$, så $V(Y_i) = 456.851$ vilket ger $\sigma = D(Y_i) = \boxed{21.37}$.

(b) Vi ska beräkna med $n = 1000$:

$$P\Big(\sum_{i=1}^{n} Y_i \geq 30000\Big) = P\Big(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(Y_i - \mu) \geq \frac{30000 - n\mu}{\sigma\sqrt{n}}\Big).$$

Vi använder CGS för att approximera med normalfördelningen samt

$$\frac{30000 - n\mu}{\sigma\sqrt{n}} = 0.99.$$

Svaret blir

$$\boxed{\Phi(0.99) = 0.8389}.$$

6. Här ska man använda stickprov i par. Om x_i respektive y_i betecknar vinsterna för `PointBlank` respektive `MarginReaper 2.0`, och $z_i = x_i - y_i$ betecknar skillnaderna, får vi:

Börs:	NASDAQ	Stockholm	Euronext	TMX	KASE
z_i:	769.5	888.829	933	598.5	372.5

Medelvärde och standardavvikelse blir: $\bar{z} = 712.5$ och $s = 229.9$. Konfidensintervall tas fram med t-fördelningens kvantil $t_{0.005}(5-1) = 4.60$. Konfidensintervallet för skillnaden Δ blir

$$I_\Delta = (\bar{z} \pm t_{0.005}(4)s/\sqrt{5}) = (239.4, 1185.6).$$

Då konfidensintervallet inte innehåller värdet 0 så drar vi slutsatsen att det **är** en signifikant skillnad mellan algoritmerna.

7. Bevisbördan ligger hos husägarna, och testet ska vara ensidigt. Vi ställer upp hypoteserna: H_0: $\mu \geq \mu_0 = 300$ mot H_1: $\mu < 300$ och använder testvariabeln

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

För signifikansnivå 5% ska vi använda $-\lambda_{0.05} = -1.645$ som gräns för testet.

Observerat värde av testvariabeln blir

$$z = \frac{295 - 300}{15/\sqrt{16}} = \boxed{-1.33}.$$

Slutsatsen är att vi **inte** kan förkasta H_0.

För styrkan om sanna $\mu = 290$ använder vi att

$$Z_0 = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}}$$

där den första termen är N(0,1) och den andra är -2.67. Styrkan är sannolikheten att H_0 förkastas, dvs

$$h = 1 - \beta = P_{\mu=290}(Z_0 < -1.645) = \Phi(2.67 - 1.645) = \Phi(1.02) = \boxed{0.846}.$$

8. (a) För Bin(6, 0.5) gäller att $p_i = \binom{6}{i}(\frac{1}{2})^i$. Utvidga tabellen genom att fylla i $E_i = 1000p_i$:

$i =$	0	1	2	3	4	5	6
$O_i =$	10	75	225	431	190	62	7
$E_i =$	15.6	93.8	234	312.5	234	93.8	15.6

Med 5% signifikansnivå blir gränsen $\chi^2_{0.05}(6) = 12.6$. Observerat värde på testvariabeln är

$$q = \frac{(10 - 15.6)^2}{15.6} + \frac{(75 - 93.8)^2}{93.8} + \cdots = 74.86.$$

Eftersom $74.86 > 12.6$ blir slutsatsen att man bör förkasta H_0.

(b) När p ska skattas använder vi att Bin(6, p) har väntevärde $\mu = 6p$ och skattar μ med medelvärdet

$$\mu^\star = \overline{x} = \frac{1 \cdot 75 + 2 \cdot 225 + 3 \cdot 431 + 4 \cdot 190 + 5 \cdot 62 + 6 \cdot 7}{1000} = 2.93.$$

Detta ger $p^\star = \mu^\star/6 = 2.93/6 = \boxed{0.488}$.