UPPSALA UNIVERSITET
Matematiska institutionen
Fredrik Jonsson

PROV I MATEMATISK STATISTIK
Sannolikhet och Statistik, W m.fl.
5:e juni 2008

Kortfattade lösningar

1. (2p) Man finner att 15 av 36 fall är gynnsamma. Dvs. $\frac{15}{36} = \frac{5}{12} \approx 0.4167$.

   (3p) P(summa ≥ 8 givet att första är en femma). Andra måste vara större eller lika med 3, dvs. 4 av 6 fall är gynnsamma. Dvs. $\frac{4}{6} = \frac{2}{3} \approx 0.667$.

2. Låt X vara antalet krona vid n slantsinglingar. Vid varje singling är $P(\text{få krona}) = 1/2 = p$ och olika singlingar är oberoende, så $X \in \mathrm{Bin}\,(n, p)$.

   (2p) Låt först $n = 10$. P(lika många krona som klave)$=P(X = 5) = P(4 < X \leq 5) = 0.62305 - 0.37695 \approx 0.2461$ enligt Tabell 6.
   Alternativt kan sannolikhetsfunktionen för binomialfördelningen användas: $P(X = 5) = p_X(5) = \binom{10}{5}(\frac{1}{2})5(\frac{1}{2})^{10-5} = \binom{10}{5}(\frac{1}{2})^{10} = 0.24609...$

   (2p) Låt sedan $n = 20$. P(få bara krona eller bara klave) $= P(X = 20$ eller $X = 0) = (\frac{1}{2})^{20} + (\frac{1}{2})^{20} = (\frac{1}{2})^{19} \approx 1.907 \cdot 10^{-6}$.

   (2p) Låt slutligen $n = 100$. Eftersom $np(1-p) = 25 > 10$ så är $X \tilde{\in} N(50, 5)$ så P(lika många krona som klave)$=P(X = 50) = P(49 < X \leq 50) = P(49.5 < X \leq 50.5) = P(\frac{49.5-50}{5} < \frac{X-50}{5} \leq \frac{50.5-50}{5}) = P(-0.1 < \frac{X-50}{5} \leq 0.1) \approx \Phi(0.1) - \Phi(-0.1) = 2\Phi(0.1) - 1 = 2 \cdot 0.5398 - 1 = 0.0796$ enligt Tabell 1.
   Det går även att räkna exakt via sannolikhetsfunktionen för binomialfördelningen.

3. $x_1, \ldots x_7$ slumpmässigt stickprov från $N(\mu_1, \sigma)$ och $y_1, \ldots y_7$ slumpmässigt stickprov från $N(\mu_2, \sigma)$ (ej stickprov i par!), bilda 99% konfidensintervall för $\mu_1 - \mu_2$.

   (4p) $\bar{x} \approx 70.57$, $\bar{y} \approx 70.14$, $s_x \approx 0.2430$, $s_y \approx 0.2370$, $t_{0.005}(12) \approx 3.05$, $s_p \approx 0.240$ ger:
   $$I_{\mu_1-\mu_2} = \bar{x} - \bar{y} \pm t_{0.005}(12) s_p \sqrt{2/7} \approx 0.43 \pm 0.391 \approx [0.038, 0.82]$$

   (2p) Med konfidensgrad 99% är det statistiskt säkerhetsställt att A är en mer effektiv luftfuktare ($\mu_1 \leq \mu_2$ är osannolikt) och att den förväntade skillnaden i prestation (för denna typ av situation) befinner sig inom intervallet i (a).

4. X är $N(\mu, 10)$.

   (2p) Antag $\mu = 45$.
   $$\begin{aligned} P(X < 40) &= P\left(\frac{X-45}{10} < \frac{40-45}{10}\right) = \Phi(-0.5) = 1 - \Phi(0.5) \\ &\approx 1 - 0.6915 = 0.3085 \approx 0.31. \end{aligned}$$

   (2p) Tre upprepade försök (oberoende av varandra)
   $$P(\text{alla understiger}) = P(\text{den första understiger})^3 \approx 0.3085^3 \approx 0.029$$

   (2p) Antag $\mu = 42$.
   $$P(X < 40) = \Phi(-0.2) = 1 - \Phi(0.2) \approx 1 - 0.5793 = 0.4207$$
   $$P(\text{alla } n \text{ understiger}) \approx 0.4207^n$$
   Testa små värden på n eller lös ekvationen $0.4207^n = 0.01$ (lösning $n \approx 5.32$).
   Svar: $n = 6$.

5. (5p) Låt p vara den andelen angripna tomatplantor i hela tomatodlingen. Antag att X =antalet anpripna tomatplantor bland 225 undersökta är $\mathrm{Bin}\,(225, p)$. Skatta p med $p^* = \frac{X}{225}$. Det är då möjligt att ge ett approximativt (normalapproximation av binomialfördelning, centrala gränsvärdessatsen) 95% konfidensintervall för p:

$$I_p = p^* \pm \lambda_{0.025}\sqrt{p^*q^*/225}.$$

$\lambda_{0.025} \approx 1.96$, $p^* = 70/225 \approx 0.3111$, $\sqrt{p^*q^*/225} \approx 0.03086$ ger:

$$I_p \approx 0.311 \pm 0.060 \approx [0.25, 0.37].$$

6. (1p) $8/90 \approx 0.0889$.

   (2p) $p^* = \frac{X}{n}$ med $n = 90$ och $X \in \mathrm{Bin}\,(90, p)$.
   $E(p^*) = p$, $V(p^*) = p(1-p)/n$,

   (1p) Ja, ty p^* är en skattning av p och $E(p^*) = p$, oavsett värde på p.

   (2p) Den avtar lineärt i n mot 0. Tolkning: att skattningen blir mer och mer säker/precis ju fler kopieringstillfällen man undersöker samt att den asymptotiskt inte innehåller någon osäkerhet alls.

7. (2p) $y = \alpha + \beta u$ med $\alpha^* \approx 120.02$ och $\beta^* \approx -24.24$

   (2p) Väljer man konfidensgraden 95% blir resultatet det följande:

$$\begin{aligned} I_\beta &= \beta^* \pm t_{0.025}(22) s_r \sqrt{1/S_{uu}} \\ I_\alpha &= \alpha^* \pm t_{0.025}(22) s_r \sqrt{1/22 + (\bar{x})^2/S_{uu}} \end{aligned}$$

   med $s_r \approx 4.09$, $t_{0.025}(22) \approx 2.07$ ger

$$\begin{aligned} I_\beta &\approx -24.24 \pm 2.647 \approx [-26.89, -21.59] \\ I_\alpha &\approx 120.02 \pm 10.30 \approx [109.7, 130.3]. \end{aligned}$$

   Följande har använts:

$$\begin{aligned} S_{uy} &\approx -247.65 \\ S_{uu} &\approx 10.22 \\ S_{yy} &\approx 6368.8 \end{aligned}$$

   Kan vara värt att notera att värden som ges i uppgiften är väl grovt avrundade; skattade parametrar och intervall blir något annorlunda om man räknar mer precist med givna data (vilket troligen hänger samman med att u-värden blir ganska små efter logaritmering).

   (2p) $y_i = \alpha + \beta u_i + \varepsilon_i$, med ε_i oberoende $N(0, \sigma)$ för $i = 1, \ldots, 24$.
   Följaktligen modelleras relationen mellan y och x som

$$y = \alpha + \beta \ln x,$$

   samt oberoende, lika- och normalfördelade additivt slumpmässiga avvikelser vid varje observationstillfälle.