TILASTOTIETEEN JATKOKURSSI. 5+5 OP. 21.1.–6.5.2016. Luennoi yliopistonlehtori Pekka Pere.

1. välikokeen uusintakokeen 20.5.2016 ratkaisuehdotukset

1. Todennäköisyys, että tilintarkastaja löytää keksityn kuitin yhdestä kuittipinosta on 1/100. Todennäköisyys, että talousrikollinen jää kiinni on 0.634:

$$\begin{aligned}
\mathsf{P}(\text{talousrikollinen jää kiinni}) &= \mathsf{P}(\text{tilintarkastaja löytää ainakin yhden keksityn kuitin}) \\
&= 1 - \mathsf{P}(\text{tilintarkastaja ei löydä yhtään keksittyä kuittia}) \\
&= 1 - (1 - 1/100)^{100} \\
&\approx 1 - 0.366 \\
&\approx 0.634.
\end{aligned}$$

2.

a) Todennäköisyys, että selviää hengissä yhdestä pelistä venäläistä rulettia on 5/6. Todennäköisyys, että selviää hengissä kuudesta pelistä venäläistä rulettia on pelien riippumattomuuden perusteella

$$(\frac{5}{6})^6 \approx 0.33.$$

b) Revolverin laukeamisen todennäköisyys on 1/6, mikä on myös todennäköisyys, että pelaaja kuolee ensimmäisessä pelissä. Mikäli peli jatkuu pidempään viimeistä peliä, jossa revolveri laukeaa (satunnaismuuttujan Y realisaatio y), edeltää $y-1$ peliä, jossa revolveri ei laukea. Niiden kunkin todennäköisyys on 5/6. Todennäköisyys pelien lukumäärälle kunnes revolveri laukeaa on siten (jälleen pelien riippumattomuuden perusteella)

$$\mathsf{P}(Y = y) = (\frac{5}{6})^{y-1} \times \frac{1}{6},$$

jossa $y = 1, 2, \ldots$

Ylimääräistä kommentointia: Tällaista jakaumaa kutsutaan *geometriseksi jakaumaksi*. Se kuvaa todennäköisyyksiä tapahtumaketjussa, jossa "onnistumisia" seuraa "epäonnistuminen" (tai toisinpäin). Jakauman määrittelevien todennäköisyyksien ei tarvitse rajoittua arvoihin 5/6 ja 1/6.

c) Alla on kaksi vaihtoehtoista ratkaisua.

i) Vastaus saadaan kohdan b) pistetodennäköisyysfunktion avulla laskettavasta kertymäfunktiosta:

pelien lukumäärän pistetodennäköisyyksiä

1	2	3	4	5	6	7	8	9	10	11
0.167	0.139	0.116	0.096	0.080	0.067	0.056	0.046	0.039	0.032	0.027

pelien lukumäärän kertymäfunktion arvoja

1	2	3	4	5	6	7	8	9	10	11
0.167	0.306	0.421	0.518	0.598	0.665	0.721	0.767	0.806	0.838	0.865

Kertymäfunktiosta nähdään, että jo 4. pelin jälkeen pelaaja on kuollut todennäköisyydellä $0.518 > 1/2$.

ii) Ilmaistaan kuoleman todennäköisyys n pelin jälkeen komplementtitapahtumansa "elää n pelin jälkeen" avulla, ja merkitään kuoleman todennäköisyys yhtäsuureksi tai suuremmaksi kuin $1/2$:

$$1 - \mathsf{P}(\text{elää } n \text{ pelin jälkeen}) \geq \frac{1}{2}.$$

Sijoitetaan (kohdan a) mukaisesti) $\mathsf{P}(\text{elää } n \text{ pelin jälkeen}) = (5/6)^n$, logaritmoidaan yhtälö puolittain, käytetään sääntöä $\log x^n = n \log x$ ja ratkaistaan n:

$$\begin{aligned} 1-(5/6)^n &\geq \frac{1}{2} && \Leftrightarrow \\ (5/6)^n &\leq 1-1/2 && \Leftrightarrow \\ n\log(5/6) &\leq \log(1/2) && \Leftrightarrow \\ n &\geq \frac{\log(1/2)}{\log(5/6)} \\ &\approx 3.802. \end{aligned}$$

Epäyhtälön suunta säilyy kolmannella rivillä logaritmoinnin jälkeen, koska logaritmifunktio on aidosti kasvava ($\log x$ suurenee x:n suuretessa). Epäyhtälön suunta vaihtuu neljännellä rivillä, kun epäyhtälö jaetaan puolittain negatiivisella luvulla $\log \frac{5}{6}$. (Log-funktio saa negatiivisia arvoja, kun sen argumentti on nollan ja yhden välillä.) Tulos on, että 4. pelin jälkeen todennäköisyys, että pelaaja on kuollut, on suurempi kuin $1/2$.

Ylimääräistä kommentointia: Sijoittamalla tavalla *i)* laskettu todennäköisyys kuolemalle (suuremmalla tarkkuudella) neljännessä pelissä tavan *ii)* ensimmäistä epäyhtälöä vastaavaan yhtälöön saadaan n:n ratkaisuksi tasan 4:

$$\begin{aligned} 1-(5/6)^n &= 0.51774691358025 && \Leftrightarrow \\ (5/6)^n &= 1-0.51774691358025 && \Leftrightarrow \\ n\log(5/6) &= \log 0.48225308641975 && \Leftrightarrow \\ n &= \frac{\log 0.48225308641975}{\log(5/6)} \\ &= 4. \end{aligned}$$

3. Binomijakauman normaalijakauma-approksimaatio on (yhden) peukalosäännön mukaan käyttökelpoinen, kun $n\pi > 5$ ja $n(1-\pi) > 5$, jossa n on havaintojen lukumäärä ja π ja $1-\pi$ ovat vaihtoehtoisten tapahtumien todennäköisyydet. Tehtävän tilanteessa $n = 54$ ja oletuksen mukaan $\pi = 1 - \pi = 0.5$, joten

$$n\pi = n(1-\pi) = 54 \times 0.5 = 27 > 5.$$

Peukalosäännön mukaan normaalijakauma-approksimaatio on käytettävissä. Merkitään Y:llä satunnaismuuttujaa "päätösten lukumäärä, joissa lapset määrätään isälle asumaan". Tehtävän tilanteessa odotusarvo $\mathsf{E}(Y) = n\pi = 27$ ja varianssi $\mathsf{V}(Y) = n\pi(1-\pi) = 54 \times 0.5 \times 0.5 = 13.5$. Näin ollen normaalijakauma-approksimaation mukaan todennäköisyys, että satunnaisessa 54:n havainnon aineistossa lapset määrätään isälle asumaan päätöksistä 14:ssä tai pienemmässä lukumäärässä on

$$\begin{aligned}
\mathsf{P}(Y \le 14) &= \mathsf{P}\left(\frac{Y-27}{\sqrt{13.5}} \le \frac{14-27}{\sqrt{13.5}}\right) &&\approx \mathsf{P}\left(Z \le \frac{14-27}{\sqrt{13.5}}\right) \\
&= \Phi\left(\frac{14-27}{\sqrt{13.5}}\right) &&\approx \Phi\left(-3.538\right) \\
&\approx 0.0002.
\end{aligned}$$

Yllä Z on standardinormaalijakautunut satunnaismuuttuja ja $\Phi(z)$ on sen kertymäfunktio. Standardinormaalijakauman kertymäfunktion arvo 0.0002 on saatu R-käskyllä

```
pnorm(-3.538)
```

Todennäköisyys on noin kaksi kymmenestuhannesosaa.

Ylimääräistä kommentointia. R:n käsky `pbinom(14,54,0.5)` antaa eksaktiksi vastaukseksi (neljän desimaalin tarkkuudella) 0.0003:

$$\mathsf{P}(Y \le 14) = \sum_{k=0}^{14} \binom{54}{k} 0.50^k \times 0.50^{54-k} = 0.0003.$$

Mikäli hovioikeudenkäynnissä eri sukupuolta olevilla vanhemmilla olisi oletetulla tavalla yhtä suuri todennäköisyys (0.5) saada lapsi asumaan luokseen, olisi tutkimuslaitoksen aineisto erittäin poikkeuksellinen. Oletuksen $\pi = 0.5$ paikkansapitävyyttä on syytä epäillä.

Laskut ovat analogisia luentomonisteen (10.5.-versio) sivun 56 esimerkissä tehtyjen kanssa. Huomionarvoista on, että hovioikeuksissa isät saivat lapset asumaan prosentuaalisesti vielä harvemmin kuin käräjäoikeuksissa (26 vs. 74 % ja 30 vs. 70 %). Silti todennäköisyys isien juttujen havaittuun voittoprosenttiosuuteen on suurempi hovioikeus- kuin käräjäoikeusaineistossa. Selitys on, että hovioikeusaineistossa päätösten lukumäärän varianssi on suurempi pienemmän havaintomäärän (118 *vs.* 54) takia. Odotusarvosta poikkeavien tapahtumien todennäköisyys kasvaa tällöin.

Havaintojen riippumattomuus ei ole selviö. Ei liene mahdotonta esimerkiksi, että sama tuomari olisi tehnyt useampia päätöksiä. Toisaalta jos oikeus toimii yleisen oikeustajun mukaisesti, ei tuomarin henkilöllisyydellä saisi olla vaikutusta päätökseen, jolloin tuomarin henkilöllisyys ei vaikuttaisi päätösten riippumattomuuteen.

4.

a) Tehtävänä on laskea todennäköisyys $\mathsf{P}(I|K)$. Se saadaan Bayesin kaavasta, joka kääntää tapahtuman ja ehdon järjestyksen todennäköisyyttä laskettaessa:

$$\begin{aligned}\mathsf{P}(I|K) &= \frac{\mathsf{P}(K|I)\mathsf{P}(I)}{\mathsf{P}(K|I)\mathsf{P}(I)+\mathsf{P}(K|I^c)\mathsf{P}(I^c)} \\ &= \frac{0.5\times 0.0001}{0.5\times 0.0001+0.5\times 0.9999} \\ &= \frac{1}{1+\frac{0.9999}{0.0001}} \\ &= \frac{1}{1+9999} \\ &= \frac{1}{10000}.\end{aligned}$$

Koska ihmeestä oikein tai väärin kertominen on yhtä todennäköistä, ei siitä kertomisella ole informaatioarvoa. Ihmeen todennäköisyys on kertomisen jälkeenkin alkuperäinen 0.0001.

b) Nyt

$$\begin{aligned}\mathsf{P}(I|K) &= \frac{\mathsf{P}(K|I)\mathsf{P}(I)}{\mathsf{P}(K|I)\mathsf{P}(I)+\mathsf{P}(K|I^c)\mathsf{P}(I^c)} \\ &= \frac{1\times 0.0001}{1\times 0.0001+0\times 0.9999} \\ &= 1.\end{aligned}$$

Kun kertojan tieto on täydellisen luotettavaa, muuttuu ihmeen todennäköisyys yhdeksi. Voidaan olla varmoja, että ihme on tapahtunut.

c) Nyt

$$\begin{aligned}\mathsf{P}(I|K) &= \frac{\mathsf{P}(K|I)\mathsf{P}(I)}{\mathsf{P}(K|I)\mathsf{P}(I)+\mathsf{P}(K|I^c)\mathsf{P}(I^c)} \\ &= \frac{1\times 0.0001}{1\times 0.0001+0,01\times 0.9999} \\ &= \frac{1}{1+\frac{0.01\times 0.9999}{0.0001}} \\ &= \frac{1}{1+100\times 0.9999} \\ &= \frac{1}{1+99.99} \\ &\approx 0.0099.\end{aligned}$$

Luotettava kertominen oleellisesti satakertaisti ihmeen todennäköisyyden. Ihmeen todennäköisyys on silti pieni noin 0.01. Vaikka kertominen satakertaistaa ihmeen todennäköisyyden, on ihmeen alkuperäinen todennäköisyys niin pieni, että edes luotettava kertominen ei tee ihmettä uskottavaksi.