582421 Satunnaisalgoritmit (syksy 2005)

1. välikoe 17.10., ratkaisuja (JK)

1. Harjoituskerran 2 tehtävä 2 (ja kurssikirjan tehtävä 2.16).

   Arvostelu: Kumpikin kohta 4 pistettä. Kohdassa (a) täydet pisteet edellyttivät selityksiä siitä, mistä eri todennäköisyydet ja odotusarvot tulevat; pelkkä kaavojen luettelu ei riitä. Kohdassa (b) sai yhden pisteen oikeasta ideasta (jaetaan heittojono lohkoihin), toisen pisteen kysytyn todennäköisyyden tarkasta lausekkeesta ja loput pisteet tämän arvioinnista.

2. Kurssikirja, sivut 64 ja 66.

   Arvostelu: Kumpikin kohta 4 pistettä. Arvostelussa on kiinnitetty huomiota perusidean lisäksi myös todistuksen täsmällisyyteen. Tyypillinen virhe kohdassa (b) oli tehdä kaksi toisensa kumoavaa merkkivirhettä, mistä on vähennetty yksi piste.

3. Olkoon X_i lokeroon i tulevien pallojen lukumäärä (kun myös ylivuotaneet pallot lasketaan normaalisti). Merkitään $Z_i = 1$, jos $X_i > 2$, ja $Z_i = 0$ muuten. Siis $Z = \sum_{i=1}^{n} Z_i$.

   (a) Koska X_i noudattaa kullakin i binomijakaumaa parametreilla n ja $1/n$, saadaan

$$\begin{aligned}
\Pr(X_i = 0) &= \left(1 - \frac{1}{n}\right)^n \\
\Pr(X_i = 1) &= n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1} \\
\Pr(X_i = 2) &= \frac{n(n-1)}{2}\left(\frac{1}{n}\right)^2\left(1 - \frac{1}{n}\right)^{n-2} = \frac{1}{2}\left(1 - \frac{1}{n}\right)^{n-1}
\end{aligned}$$

   Siis

$$\Pr(Z_i = 1) = 1 - \Pr(X_i \le 2) = 1 - \left(1 - \frac{1}{n}\right)^{n-1}\left(1 - \frac{1}{n} + 1 + \frac{1}{2}\right) = 1 - \left(\frac{5}{2} - \frac{1}{n}\right)\left(1 - \frac{1}{n}\right)^{n-1},$$

   joten

$$\mathbf{E}[Z] = \sum_{i=1}^{n} \mathbf{E}[Z_i] = n - \left(\frac{5n}{2} - 1\right)\left(1 - \frac{1}{n}\right)^{n-1}.$$

   Suurilla n voidaan approksimoida $(1 - 1/n)^n \approx \mathrm{e}^{-1}$ ja $5n/2 - 1 \approx 5n/2$, jolloin saadaan

$$\mathbf{E}[Z] \approx n\left(1 - \frac{5}{2\mathrm{e}}\right).$$

   (Tämä on noin 0,08n.)

   (b) Olkoot Y_i, $i = 1, \ldots, n$, toisistaan riippumattomia Poisson(1)-muuttujia. Merkitään $Z'_i = 1$, jos $Y_i > 2$, ja $Z'_i = 0$ muuten, ja määritellään $Z' = \sum_{i=1}^{n} Z'_i$. Satunnaismuuttuja Z' on siis Poisson-approksimaatio satunnaismuuttujalle Z.

   Poisson-jakauman määritelmästä nähdään, että

$$\Pr(Z'_i = 0) = \Pr(Y_i = 0) + \Pr(Y_i = 1) + \Pr(Y_i = 2) = \mathrm{e}^{-1}\left(\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!}\right) = \frac{5}{2\mathrm{e}}.$$

   Siis Z' noudattaa binomijakaumaa parametrein n ja $1 - 5/2\mathrm{e}$. Perusidea on rajata $\Pr(Z' < \frac{1}{2}\mathbf{E}[Z'])$ käyttämällä Chernoffin rajaa ja sitten käyttää tunnettuja Poisson-approksimaation tarkkuutta koskevia tuloksia.

   Jotta päästään soveltamaan kirjan lausetta 5.10, pitää määritellä sopiva f. Olkoon A kohdassa (a) laskettu tarkka odotusarvo $\mathbf{E}[Z]$. Asetetaan $f(x_1, \ldots, x_n) = 1$, jos ehto $x_i \ge 3$

pätee alle $A/2$ eri indeksillä $i \in \{1, \ldots, n\}$; muuten $f(x_1, \ldots, x_n) = 0$. Tämän määrittelyn motivaatio on, että nyt

$$\mathbf{E}[f(X_1, \ldots, X_n)] = \Pr(Z < A/2) = \Pr(Z < \tfrac{1}{2}\mathbf{E}[Z])$$

ja

$$\mathbf{E}[f(Y_1, \ldots, Y_n)] = \Pr(Z' < A/2).$$

Koska selvästi $\mathbf{E}[f(X_1, \ldots, X_n)]$ pienenee, jos pallojen lukumäärää kasvatetaan, kirjan lauseen 5.10 ehdot ovat voimassa, ja

$$\Pr(Z < \tfrac{1}{2}\mathbf{E}[Z]) = \mathbf{E}[f(X_1, \ldots, X_n)] \le 2\mathbf{E}[f(Y_1, \ldots, Y_n)] = 2\Pr(Z' < \tfrac{1}{2}A).$$

Jos halutaan olla tarkkoja, todennäköisyyden $\Pr(Z' < A/2)$ arvioimisessa pitää vielä ottaa huomioon, että $A = \mathbf{E}[Z]$ ei ole täsmälleen sama kuin $n(1 - 5/2\mathrm{e}) = \mathbf{E}[Z']$, vaan ainoastaan lähestyy tätä raja-arvona. Olkoon n_0 sellainen, että $A \le \frac{3}{2}\mathbf{E}[Z']$, kun $n \ge n_0$. Tällöin kun $n \ge n_0$, pätee

$$\Pr(Z < \tfrac{1}{2}\mathbf{E}[Z]) \le 2\Pr(Z' < \tfrac{1}{2}A) \le 2\Pr(Z' < \tfrac{3}{4}\mathbf{E}[Z']).$$

Koska Z' on binomijakautunut, siihen voidaan soveltaa esim. edellä tehtävässä 2 (b) esiintyvää rajaa valitsemalla $\delta = 1/4$:

$$\begin{aligned}
\Pr(Z' < \tfrac{3}{4}\mathbf{E}[Z']) &\le \left(\frac{\mathrm{e}^{-1/4}}{(3/4)^{3/4}}\right)^{(1-5/2\mathrm{e})n} \\
&= \left(\frac{64}{27\mathrm{e}}\right)^{(1/4)\cdot(1-5/2\mathrm{e})n} \\
&= \exp\left(\left(\ln\frac{64}{27\mathrm{e}}\right)\frac{1}{4}\left(1 - \frac{5}{2\mathrm{e}}\right)n\right).
\end{aligned}$$

Koska $27\mathrm{e} > 64$, eksponentti on negatiivinen, eli saadaan haluttua muotoa oleva raja

$$\Pr(Z \le \tfrac{1}{2}\mathbf{E}[Z]) \le \mathrm{e}^{-cn}$$

kun $n \ge n_0$, missä $c > 0$.

Arvostelu: 2 pistettä kohdasta (a), 6 pistettä kohdasta (b).

Tyypillisesti kohta (a) oli osatty hyvin.

Kohdassa (b) testattiin nimenomaan Poisson-approksimaation käyttöä, joten yritykset soveltaa Chernoffin rajoja suoraan muuttujaan Z antoivat nolla pistettä. Yhden pisteen sai, jos osasi jotenkin järkevällä tavalla yhdistää tämän Poisson-approksimaatioon. Kolme pistettä tuli Poisson-approksimaation soveltamisen perusajatuksesta: määritellään approksimoiva muuttuja (edellä Z') johon voidaan soveltaa Chernoffin rajaa. Viimeiset kaksi pistettä tuli todistuksen teknisistä yksityiskohdista. (Täysiin pisteisiinkään ei vaadittu aivan niin yksityiskohtaista tarkastelua kuin edellä on annettu.)