Eksamensdisposition - Kapitel 4

Søren Mulvad, rbn601

17. juni 2019

- **Chernoff bounds**
  - Bevis for case med $\mathbb{P}[X > (1+\delta)\mu]$
  - Bevis for case med $\mathbb{P}[X < (1-\delta)\mu]$
- **Set balancing**
  - Håndkørsel af algoritme
  - Kvalitet af resultat

Eksamensdisposition - Kapitel 4

Chernoff Bounds

Case for $\mathbb{P}[X > (1+\delta)\mu]$

Givet uafhængige Poisson trials $X_1, \ldots, X_n$, dvs. for $i = 1, \ldots, n$, lad X_i være en indikatorvariabel med $\mathbb{P}[X_i = 1] = p_i$ hvor $0 < p_i < 1$. Lad

$$\delta > 0 \qquad X = \sum_{i=1}^{n} X_i \qquad \mu = \mu_X = \sum_{i=1}^{n} p_i$$

Da gælder:

$$\mathbb{P}[X > (1+\delta)\mu] < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

Bevis

Lad $t > 0$ (fastlægges senere til $t = \ln(1+\delta)$). Da gælder:

$$\begin{aligned} \mathbb{P}[X > (1+\delta)\mu] &= \mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right] && (1) \\ &< \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}} && (2) \end{aligned}$$

I (1) opløfter vi e med vores udtryk, hvorved uligheden stadig gælder og i (2) bruger vi Markovs ulighed. Lad os nu regne videre på nævneren i (2):

$$\begin{aligned} \mathbb{E}\left[e^{tX}\right] &= \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] && (3) \\ &= \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] && (4) \\ &= \prod_{i=1}^{n} \left(p_i e^t + (1-p_i)e^0\right) && (5) \\ &= \prod_{i=1}^{n} \left(1 + p_i(e^t - 1)\right) && (6) \\ &\leq \prod_{i=1}^{n} e^{p_i(e^t-1)} && (7) \\ &= e^{\left(\sum_{i=1}^{n} p_i(e^t-1)\right)} && (8) \\ &= e^{\mu(e^t-1)} && (9) \end{aligned}$$

I (3) bruger vi definitionen for X, og bestemmer produktet i stedet for summen da den står i en potens.
I (4) benytter vi at vores stokastiske variable er uafhængige, og derfor kan vi flytte vores forventning ind i produktet.
I (5) benytter vi definitonen på forventningsværdien og egenskaberne for Poisson trials.
I (6) reducerer vi blot udtrykket.
I (7) benytter vi reglen $1 + x \leq e^x \ \forall x \in \mathbb{R}$.
I (8) har vi, at produktet af en masse udtryk med samme base svarer til summen af alle potenserne.
I (9) benytter vi vores definition af μ.

Indsætter vi denne værdi tilbage i (2) får vi:

$$\begin{aligned}
\frac{\mathbb{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}} &\leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}} && (10)\\
&= \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} && (11)\\
&= \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} && (12)
\end{aligned}$$

I (11) indsætter vi vores $t = \ln(1+\delta)$ så vores udtryk bliver pænt.
I (12) sætter vi blot μ uden for en parentes for brøken.

Case for $\mathbb{P}\left[X < (1-\delta)\mu\right]$

Givet uafhængige Poisson trials $X_1, \ldots, X_n$, dvs. for $i = 1, \ldots, n$, lad X_i være en indikatorvariabel med $\mathbb{P}\left[X_i = 1\right] = p_i$ hvor $0 < p_i < 1$. Lad

$$0 < \delta \leq 1 \qquad X = \sum_{i=1}^{n} X_i \qquad \mu = \mu_X = \sum_{i=1}^{n} p_i$$

Da gælder:

$$\mathbb{P}\left[X < (1-\delta)\mu\right] < e^{-\frac{\delta^2\mu}{2}} \tag{13}$$

Bevis

Lad $t > 0$ (fastlægges senere til $t = -\ln(1-\delta)$). Da gælder:

$$\begin{aligned}
\mathbb{P}\left[X < (1-\delta)\mu\right] &= \mathbb{P}\left[e^{-tX} > e^{-t(1-\delta)\mu}\right] && (14)\\
&= \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}
\end{aligned}$$

Hvor vi gør stort set det samme som før, men havde den lidt anderledes start i (14) og derudover brugte $t = -\ln(1-\delta)$. Lad os nu regne videre på dette:

$$\begin{aligned}
\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} &= \left(\frac{e^{-\delta}}{e^{(1-\delta)\ln(1-\delta)}}\right)^{\mu} && (15)\\
&< \left(\frac{e^{-\delta}}{e^{-\delta+\frac{1}{2}\delta^2}}\right)^{\mu} && (16)\\
&= e^{-\frac{1}{2}\delta^2\mu} && (17)
\end{aligned}$$

I (15) omskriver vi blot for at få noget med e i nævneren. I (16) benytter vi McLaurin expansion, som medfører at $(1-\delta)\ln(1-\delta) > -\delta + \frac{1}{2}\delta^2$. Og endelig i (17) er det blot simple potensregneregler.

McLaurin Expansion

Givet en funktion $f(x) : \mathbb{R} \to \mathbb{R}$, som er vilkårligt ofte differentiabel i $x = 0$, da gælder:

$$f(x) = \sum_{i=0}^{n} \frac{f^{(i)}(0)}{i!} x^i$$

hvor f.eks. $f^{(3)}(0)$ betegner $f'''(0)$.

For vores konkrete case ovenfor betyder det:

$$(1-\delta)\ln(1-\delta) = -\delta + \frac{1}{2}\delta^2 + \frac{1}{6}\delta^3 + \frac{1}{12}\delta^4 + \frac{1}{20}\delta^5 + \cdots > -\delta + \frac{1}{2}\delta^2$$

Set Balancing

Problem

Givet en $n \times n$ 0/1-matrix $\mathbf{A}$, find vektor $\mathbf{b} \in \{-1, 1\}^n$ så $||\mathbf{Ab}||_\infty$ (udtales infinity-normen) bliver så lille som mulig. Eksempelvis kunne vi have:

$$\underbrace{\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{b}} = \underbrace{\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{Ab}} \qquad \rightarrow \qquad ||\mathbf{Ab}||_\infty = 2$$

$$\underbrace{\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}}_{\mathbf{b}'} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}}_{\mathbf{Ab}'} \qquad \rightarrow \qquad ||\mathbf{Ab}'||_\infty = 1$$

Her er f.eks. $||\mathbf{Ab}||_\infty = 2$ da den er defineret som den numerisk største værdi i vektor $\mathbf{Ab}$.

Man kan f.eks. forstå problemet som at matrix $\mathbf{A}$ i søjlerne har en række individer, og i rækkerne har en værdi for om besidder en bestemt egenskab eller ej. Så vil vi gerne dele individerne i to grupper, som skal sikre så jævn fordeling af egenskaberne som muligt. $||\mathbf{Ab}||_\infty$ er da et udtryk for den egenskab der er mest ujævnt fordelt.

Algoritme

Vælg indgang i i $\mathbf{b}$ til -1 med sandsynlighed $1/2$ og ellers 1 for alle i uafhængigt.

Analyse af sandsynlighed for $||\mathbf{Ab}||_\infty \leq 2\sqrt{2n \ln n}$

Betragt række j i $\mathbf{A}$, $\mathbf{A}_j$. Antag at $\mathbf{A}_j$ er på følgende form, som er en gyldig måde at anskue det på så længe vi kun kigger på én række, da vi i vores analyse bare kunne flytte alle 1-taller i $\mathbf{A}_j$ til venstre og så have de resterende 0-taller til højre:

$$\mathbf{A}_j = [\, \underbrace{1 \,\dots\, 1}_{k} \; \underbrace{0 \,\dots\, 0}_{n-k} \,]$$

Alle 0-taller kan vi da ignorere, da de ikke vil bidrage til summen i vores analyse.
For $i = 1, \dots, k$, lad da indikatorvariablen X_i være

$$X_i = \begin{cases} 1 & \text{hvis } b_i = -1 \\ 0 & \text{ellers (dvs. } b_i = 1) \end{cases}$$

Lad nu

$$X = \sum_{i=1}^{k} X_i \qquad \mu_X = k \cdot \mathbb{P}[X_i = 1] = \frac{k}{2}$$

Vi får $\mathbf{A}_j\mathbf{b} = 0$ når de første k indgange i $\mathbf{b}$ er præcis ligeligt fordelte mellem at være -1 og 1:

$$\mathbf{b} = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \left.\vphantom{\begin{matrix} -1 \\ -1 \\ 1 \\ 1 \end{matrix}}\right\} k$$

Altså får vi:

$$
\begin{gathered}
\mathbb{P}\left[\mathbf{A}_j\mathbf{b} = 0\right] = \mathbb{P}\left[X = \frac{k}{2}\right] \\
\mathbb{P}\left[\mathbf{A}_j\mathbf{b} = 2\right] = \mathbb{P}\left[X = \frac{k}{2} - 1\right] \\
\vdots \\
\mathbb{P}\left[\mathbf{A}_j\mathbf{b} = 2c\right] = \mathbb{P}\left[X = \frac{k}{2} - c\right] \\
\Updownarrow \\
\mathbb{P}\left[\mathbf{A}_j\mathbf{b} > 2c\right] = \mathbb{P}\left[X < \frac{k}{2} - c\right] \qquad (18)
\end{gathered}
$$

Da hvis vi gør c-værdien meget stor, så betyder det at vi har meget få -1'ere, dvs. $\mathbf{A}_j\mathbf{b}$ bliver stor.

Nu benytter vi (13), hvor der skal gælde $0 < \delta \leq 1$:

$$
\mathbb{P}\left[X < (1-\delta)\mu\right] < e^{-\frac{\delta^2\mu}{2}} = e^{-\frac{\delta^2 k}{4}} = \frac{1}{n^2} \tag{19}
$$

Når vi vælger $\delta^2 = \frac{8\ln n}{k} \implies \delta = 2\sqrt{\frac{2\ln n}{k}}$.

Skriver vi videre på venstresiden af (19) får vi:

$$
\begin{aligned}
\mathbb{P}\left[X < (1-\delta)\mu\right] &= \mathbb{P}\left[X < \frac{k}{2} - \delta\mu\right] && (20) \\
&= \mathbb{P}\left[\mathbf{A}_j\mathbf{b} > 2\delta\mu\right] && (21) \\
&= \mathbb{P}\left[\mathbf{A}_j\mathbf{b} > k\delta\right] \\
&= \mathbb{P}\left[\mathbf{A}_j\mathbf{b} > 2\sqrt{2k\ln n}\right] && (22) \\
&\geq \mathbb{P}\left[\mathbf{A}_j\mathbf{b} > 2\sqrt{2n\ln n}\right] && (23)
\end{aligned}
$$

I (20) ganger vi parentesen ud og indsætter vores værdi for μ på leddet uden δ.
I (21) benytter vi hvad vi fandt frem til i (18).
I (22) indsætter vi vores værdier for k og δ.
I (23) benytter vi, at der ikke kan være flere 1-taller end længden af vektor $\mathbf{b}$, altså $k \leq n$. Herved bliver uligheden sværere at opfylde, så sandsynligheden falder.

Hermed har vi altså vist, at for en række $\mathbf{A}_j$ har vi følgende, hvor medfører pilen gælder ved symmetri:

$$
\mathbb{P}\left[\mathbf{A}_j\mathbf{b} > 2\sqrt{2n\ln n}\right] < \frac{1}{n^2} \iff \mathbb{P}\left[\mathbf{A}_j\mathbf{b} < -2\sqrt{2n\ln n}\right] < \frac{1}{n^2}
$$

Vi kan nu bruge union bound til at få:

$$
\mathbb{P}\left[|\mathbf{A}_j\mathbf{b}| > 2\sqrt{2n\ln n}\right] < \frac{2}{n^2}
$$

Går vi nu videre fra bare at kigge på en række $\mathbf{A}_i$ til hele matrix $\mathbf{A}$ får vi, igen ved union bound, at:

$$
\mathbb{P}\left[||\mathbf{A}\mathbf{b}||_\infty > 2\sqrt{2n\ln n}\right] = \mathbb{P}\left[\bigcup_{j=1}^{n}\left\{|\mathbf{A}_j\mathbf{b}| > 2\sqrt{2n\ln n}\right\}\right] \leq n \cdot \frac{2}{n^2} = \frac{2}{n}
$$

Altså har vi nu vist, at med sandsynlighed $\geq 1 - \frac{2}{n}$ gælder $||\mathbf{A}\mathbf{b}||_\infty \leq 2\sqrt{2n\ln n}$.