مباحث ویژه در سیستم‌های دیجیتال

نیم‌سال اول ۱۴۰۰-۱۴۰۱

مدرس: دکتر ایمان غلام‌پور

دانشکده مهندسی برق

تمرین سری سوم

نام و نام‌خانوادگی: امین کشیری

شماره دانشجویی: ۹۷۱۰۱۰۲۶

تمرین ۱

آ) ابتدا این فرمول را به صورت آشنایی در می‌آوریم:

$$\text{conv}(A \to B) = \frac{1-S(B)}{1-\text{conf}(A \to B)} = \frac{P(B^c)}{1-P(B|A)} = \frac{P(B^c)}{P(B^c|A)}$$

از طرفی، اگر A و B مستقل باشند، داریم:

$$P(B|A) = P(B) \Rightarrow P(B^c|A) = P(B^c)$$

حال با کمی دقت می‌توان فهمید که $P(B^c|A)$ برابر است با احتمال اشتباه بودن قانون [1]. با توجه به این معادلات، می‌توانیم بگوییم که $\text{conv}(A \to B)$ برابر است بااحتمال اشتباه بودن قانون وقتی که A و B مستقل‌اند تقسیم بر احتمال اشتباه بودنی که مشاهده کرده‌ایم. مثلا اگر conv برابر 1.5 باشد، یعنی اگر A و B مستقل بودند، این قانون ۵۰ درصد بیشتر اشتباه می‌کرد (که این یعنی نسبت به حالت استقلال، اشتباهات کمتری داشته‌ایم).

ب) conf و conv نامتقارن‌اند اما lift متقارن است.

برای lift داریم:

$$\text{lift}(A \to B) = \frac{\text{conf}(A \to B)}{S(B)} = \frac{P(B|A)}{P(B)} = \frac{P(B \cap A)}{P(B)P(A)} = \frac{\text{conf}(B \to A)}{S(A)} = \text{lift}(B \to A)$$

برای conf ، در حالت کلی به وضوح داریم که: $P(B|A) \neq P(A|B)$. برای conv ، با یک مثال نقض ثابت می‌کنیم که متقارن نیست. فرض کنید که $A = B^c$. در این صورت، داریم:

$$\text{conv}(A \to B) = \frac{P(B^c) \cdot P(A)}{P(B^c \cap A)} = P(A)$$
$$\text{conv}(B \to A) = \frac{P(A^c) \cdot P(B)}{P(A^c \cap B)} = P(B)$$

اما می‌توانیم B را طوری انتخاب کنیم که $P(A) = P(B)$.

پ) ضعف این تعریف آن است که در آن احتمال رخ دادن B در نظر گرفته نشده‌است. اگر بدون تغییر دادن A مجموعه‌ی B را بزرگ و بزرگ‌تر کنیم، $P(B|A)$ تنها می‌تواند بزرگ‌تر شود. اما این بزرگ شدن، به ما چیزی راجب جالب بودن این قانون نمی‌گوید. یعنی تنها با محتمل‌تر کردن نتیجه، احتمال رخ دادن قانون (که با conf نشان داده‌ایم) بیشتر می‌شود، که به ما نشان می‌دهد conf معیار خوبی برای اندازه‌گیری قدرت قانون $A \to B$ نیست.

تمرین ۲

این سوال را تنها به صورت احتمالاتی می‌توانیم پاسخ دهیم. یعنی می‌توانیم با احتمال خیلی خیلی بالا بگوییم که مجموعه‌های پرتکرار چه خواهند بود (با احتمال تقریبا ۱۰۰). برای این کار نیز، احتمال پر تکرار بودن یک مجموعه وقتی تعداد سبدها به سمت بی‌نهایت می‌رود را به دست می‌آوریم.

متغیر تصادفی X را به صورت زیر تعریف می‌کنیم:

$$X = S(A)$$

که یعنی X تعداد تکرار زیر مجموعه‌ی A است. حال فرض کنید:

$$A = \{a_i | i \in I\}$$

احتمال این که A در یک سبد ظاهر شود، برابر است با احتمال ظاهر شدن تمام اعضای آن. دقت کنید که وجود عناصر اضافه برای ما مشکلی ایجاد نمی‌کند. از طرفی طبق صورت سوال، ظاهر شدن یک کالا، مستقل از باقی کالاها است. در نتیجه احتمال این که زیر مجموعه‌ی A را در یک سبد ببینیم، برابر است با:

[1] Association Rule

$p_A = \Pi_{i \in I} p_i$

حال اگر به تعریف X دقت کنید، وقتی که n سبد داشته باشیم، داریم:

$X \sim \mathbf{B}(n, p_A)$

یعنی X یک متغیر دوجمله‌ای است. مجموعه‌ی A پر تکرار است اگر و تنها اگر $S(A) > \frac{n}{10}$. با استفاده از قانون چبیشف شرط پر تکرار بودن مجموعه‌ها را می‌یابیم:

$$P(X \leq \mu - c) \leq P(|X - \mu| > c) \leq \frac{\sigma^2}{c^2} \xrightarrow{c=\mu-\frac{n}{10}} P(X \leq \frac{n}{10}) \leq \frac{np_A(1-p_A)}{(\mu-\frac{n}{10})^2}$$

با توجه به تعریف X داریم:

$E(X) = p_A n$

حال اگر در نامعادلات بالا n را به بی‌نهایت میل دهیم، این احتمال به صفر میل می‌کند. یعنی وقتی n به بی‌نهایت میل می‌کند، احتمال این که سبد پرتکرار نشود به صفر میل می‌کند. دقت کنید که در نامعادلات بالا، شرط مثبت بودن c یا به عبارتی $\mu \geq \frac{n}{10}$ وجود داشته است، که این خود معادل است با $p_A > \frac{1}{10}$. به صورت مشابه می‌توانیم نشان دهیم:

$$P(X \geq \mu + c) \leq P(|X - \mu| > c) \leq \frac{\sigma^2}{c^2} \xrightarrow{c=\frac{n}{10}-\mu} P(X \geq \frac{n}{10}) \leq \frac{np_A(1-p_A)}{(\frac{n}{10}-\mu)^2}$$

و این یعنی در این حالت، که $p_A < \frac{1}{10}$ است، احتمال پر تکرار شدن به صفر میل می‌کند. در نتیجه برای این که سبدی پرتکرار شود، باید داشته باشیم $p_A \geq 0.1$ (دقت کنید که اگر $p_A = 0.1$ آنگاه سبد ما روی مرز قرار دارد و پرتکرار بودن یا نبودن آن را نمی‌توانیم تضمین کنیم). طبق فرض سوال داریم: $p_i = \frac{1}{i}$ پس باید تمامی حالاتی را برای I به دست آوریم که $p_A = \Pi_{i \in I} p_i \geq 0.1$ که این اتفاق وقتی رخ می‌دهد که:

$I_1 = \{1, 2, 3\} \Rightarrow p_A = \frac{1}{6} \geq 0.1$
$I_2 = \{1, 2\} \Rightarrow p_A = \frac{1}{2} \geq 0.1$
$I_3 = \{2, 3\} \Rightarrow p_A = \frac{1}{6} \geq 0.1$
$I_4 = \{1, 3\} \Rightarrow p_A = \frac{1}{3} \geq 0.1$
$I_5 = \{2, 4\} \Rightarrow p_A = \frac{1}{8} \geq 0.1$
$I_6 = \{2, 5\} \Rightarrow p_A = \frac{1}{10} \geq 0.1$
$I_7 = \{1, 4\} \Rightarrow p_A = \frac{1}{4} \geq 0.1$
$I_8 = \{1, 5\} \Rightarrow p_A = \frac{1}{5} \geq 0.1$
$I_9 = \{1, 6\} \Rightarrow p_A = \frac{1}{6} \geq 0.1$
$I_{10} = \{1, 7\} \Rightarrow p_A = \frac{1}{7} \geq 0.1$
$I_{11} = \{1, 8\} \Rightarrow p_A = \frac{1}{8} \geq 0.1$
$I_{12} = \{1, 9\} \Rightarrow p_A = \frac{1}{9} \geq 0.1$
$I_{13} = \{1, 10\} \Rightarrow p_A = \frac{1}{10} \geq 0.1$
$I_{14} = \{1\} \Rightarrow p_A = \frac{1}{1} \geq 0.1$
$I_{15} = \{2\} \Rightarrow p_A = \frac{1}{2} \geq 0.1$
$I_{16} = \{3\} \Rightarrow p_A = \frac{1}{3} \geq 0.1$
$I_{17} = \{4\} \Rightarrow p_A = \frac{1}{4} \geq 0.1$
$I_{18} = \{5\} \Rightarrow p_A = \frac{1}{5} \geq 0.1$
$I_{19} = \{6\} \Rightarrow p_A = \frac{1}{6} \geq 0.1$
$I_{20} = \{7\} \Rightarrow p_A = \frac{1}{7} \geq 0.1$
$I_{21} = \{8\} \Rightarrow p_A = \frac{1}{8} \geq 0.1$
$I_{22} = \{9\} \Rightarrow p_A = \frac{1}{9} \geq 0.1$
$I_{23} = \{10\} \Rightarrow p_A = \frac{1}{10} \geq 0.1$

پس مجموعه‌های پرتکرار به این صورت به دست می‌آیند.

تمرین ۳

توضیحات و نتایج کدها به صورت کامل در jupyter-notebook آمده است.