آمار و احتمال مهندسی

نیم‌سال اول ۱۴۰۲-۱۴۰۱

دانشکده‌ی مهندسی کامپیوتر

تمرین سری ۴ | ویژگی‌های امید ریاضی، قضایای حد

نکات مهم

- پاسخ بخش تئوری را در قالب یک فایل pdf با اسم HW4_[STD-NUM] آپلود کنید.
- ددلاین تمرین ساعت ۲۳:۵۹ روز ۵ دی ۱۴۰۱ می‌باشد.
- سوالات ستاره‌دار، غیر تحویلی هستند و برای تمرین بیشتر قرار داده شده‌اند.

مسئله‌ی ۱. کوواریانس شرطی

کوواریانس شرطی (*Conditional Covariance*) دو متغیر تصادفی X و Y به صورت زیر تعریف می‌شود:

$$Cov(X,Y|Z) = \mathbb{E}[(X-\mathbb{E}[X|Z])(Y-\mathbb{E}[Y|Z])|Z]$$

با توجه به تعریف فوق موارد زیر را اثبات کنید:

الف)

$$Cov(X,Y|Z) = \mathbb{E}[XY|Z] - \mathbb{E}[X|Z]\mathbb{E}[Y|Z]$$

ب)

$$Cov(X,Y) = \mathbb{E}[Cov(X,Y|Z)] + Cov(\mathbb{E}[X|Z],\mathbb{E}[Y|Z])$$

ج)

$$Var(X) = \mathbb{E}[Var(X|Y)] + Var(\mathbb{E}[X|Y])$$

حل.

الف

$$Cov(X,Y|Z) = \mathbb{E}[(X-\mathbb{E}[X|Z])(Y-\mathbb{E}[Y|Z])|Z] =$$

$$\mathbb{E}[XY - Y\mathbb{E}[X|Z] - X\mathbb{E}[Y|Z] + \mathbb{E}[X|Z]\mathbb{E}[Y|Z]|Z] =$$

$$\mathbb{E}[XY|Z] - \mathbb{E}[Y\mathbb{E}[X|Z]|Z] - \mathbb{E}[X\mathbb{E}[Y|Z]|Z] + \mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]|Z] =$$

$$\mathbb{E}[XY|Z] - \mathbb{E}[Y|Z]\mathbb{E}[X|Z] - \mathbb{E}[Y|Z]\mathbb{E}[X|Z] + \mathbb{E}[Y|Z]\mathbb{E}[X|Z] =$$

$$\mathbb{E}[XY|Z] - \mathbb{E}[Y|Z]\mathbb{E}[X|Z]$$

دقت کنید که در عباراتی مثل $\mathbb{E}[Y\mathbb{E}[X|Z]|Z]$ می توانیم $\mathbb{E}[X|Z]$ را بیرون بیاریم، چون یک تابع بر حسب Z است و ما روی Z شرط داریم.

ب

از بخش الف کمک میگیریم

$$\mathbb{E}[Cov(X,Y|Z)] = \mathbb{E}[\mathbb{E}[XY|Z] - \mathbb{E}[Y|Z]\mathbb{E}[X|Z]] =$$

$$\mathbb{E}[\mathbb{E}[XY|Z]] - \mathbb{E}[\mathbb{E}[Y|Z]\mathbb{E}[X|Z]] = \mathbb{E}[XY] - \mathbb{E}[\mathbb{E}[Y|Z]\mathbb{E}[X|Z]]$$

دقت کنید که از این موضوع استفاده کردیم که: $\mathbb{E}[\mathbb{E}[XY|Z]] = \mathbb{E}[XY]$

$$Cov(\mathbb{E}[X|Z], \mathbb{E}[Y|Z]) = \mathbb{E}[\mathbb{E}[Y|Z]\mathbb{E}[X|Z]] - \mathbb{E}[\mathbb{E}[X|Z]]\mathbb{E}[\mathbb{E}[Y|Z]] =$$

$$\mathbb{E}[\mathbb{E}[Y|Z]\mathbb{E}[X|Z]] - \mathbb{E}[X]\mathbb{E}[Y]$$

دقت کنید که از این موضوع استفاده کردیم که: $\mathbb{E}[\mathbb{E}[X|Z]] = \mathbb{E}[X]$
حالا این ۲ تا را جمع میزنیم:

$$\mathbb{E}[Cov(X,Y|Z)] + Cov(\mathbb{E}[X|Z], \mathbb{E}[Y|Z]) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = Cov(X,Y)$$

ج

$$\mathbb{E}[Var(X|Y)] = \mathbb{E}[\mathbb{E}[X^۲|Y] - \mathbb{E}[X|Y]^۲] = \mathbb{E}[\mathbb{E}[X^۲|Y]] - \mathbb{E}[\mathbb{E}[X|Y]^۲] = \mathbb{E}[X^۲] - \mathbb{E}[\mathbb{E}[X|Y]^۲]$$

$$Var(\mathbb{E}[X|Y]) = \mathbb{E}[\mathbb{E}[X|Y]^۲] - \mathbb{E}[\mathbb{E}[X|Y]]^۲ = \mathbb{E}[\mathbb{E}[X|Y]^۲] - \mathbb{E}[X]^۲$$

$$\rightarrow \mathbb{E}[Var(X|Y)] + Var(\mathbb{E}[X|Y]) = \mathbb{E}[X^۲] - \mathbb{E}[X]^۲ = Var(X)$$

$\triangleright$

مسئله‌ی ۲. بانک احتمالاتی

شعبه‌ی خیابان آزادی بانک ملت به مشتریان ایستاده در صف خدمات عابر بانک، یکی یکی، رسیدگی می‌کند. فرض کنید زمان ارائه خدمت مد نظر مشتری i را X_i بنامیم. این فرض را نیز در نظر بگیرید که داریم:

$$EX_i = ۲, \quad VAR(X_i) = ۱$$

همچنین در این مسئله فرض می‌کنیم که زمان رسیدگی به درخواست مشتریان از همدیگر مستقل می‌باشد. می‌دانیم ۵۰ نفر در صف این بانک ایستاده‌اند.Yرا مدت زمانی در نظر بگیرید که عابر بانک به درخواست تمامی این مشتریان رسیدگی می‌کند. $P(۹۰ < Y < ۱۱۰)$ را بیابید.

حل.

$$Y = X_۱ + X_۲ + ... + X_n$$

$n = ۵۰, EX_i = \mu = ۲, Var(X_i) = \sigma^۲ = ۱$
بنابراین

$$P(۹۰ < Y \leqslant ۱۱۰) = P(\frac{۹۰-n\mu}{\sqrt{n}\sigma} < \frac{Y-n\mu}{\sqrt{n}\mu} < \frac{۱۱۰-n\mu}{\sqrt{n}\sigma})$$

$$= P(\frac{۹۰-۱۰۰}{\sqrt{۵۰}} < \frac{Y-n\mu}{\sqrt{n}\sigma} < \frac{۱۱۰-۱۰۰}{\sqrt{۵۰}})$$

$$= P(-\sqrt{۲} < \frac{Y-n\mu}{\sqrt{n}\sigma} < \sqrt{۲})$$

$\triangleright$

مسئله‌ی ۳. توابع نزولی

الف

فرض کنید X یک متغیر تصادفی گسسته باشد که مقادیر ممکن برای آن برابر با $۱, ۲, ...$ می‌باشد. اگر $P(X = k)$ در $k = ۱, ۲, ...$ نزولی باشد، ثابت کنید:

$$P(X = k) \leqslant ۲\frac{E[X]}{k^۲}$$

ب

فرض کنید X یک متغیر تصادفی پیوسته نامنفی باشد که تابع چگالی احتمال آن، نزولی است. نشان دهید:

$$f(x) \leqslant ۲\frac{E[X]}{x^۲} \quad ; \quad \forall x > ۰$$

حل.

الف

$$\begin{aligned}
E(X) &= \sum_{x=۱}^{\infty} xP(X = x) \\
&\geqslant \sum_{x=۱}^{k} xP(X = x) \\
&\geqslant \sum_{x=۱}^{k} xP(X = k) \\
&\geqslant P(X = k)\sum_{x=۱}^{k} x \\
&\geqslant P(X = k)\frac{k(k+۱)}{۲} \\
&\geqslant P(X = k)\frac{k^۲}{۲}
\end{aligned}$$

بنابراین در نهایت داریم:

$$P(X=k) \leqslant ۲\frac{E(X)}{k^{۲}}$$

ب

برای حالت پیوسته هم همانند قسمت قبل خواهد بود. ▷

مسئله‌ی ۴. خطای تقریبی

۵۰ عدد گرد شده و سپس با هم جمع شده اند. اگر خطای گرد کردن هر عدد با توزیع یکنواخت در بازه‌ی $(-۰/۵, ۰/۵)$ آنگاه احتمال این که تفاوت مجموع حاصل و مجموع اصلی این اعداد بیشتر از ۳ باشد را با استفاده از قانون حد مرکزی تقریب بزنید.

حل. فرض می‌کنیم X_i خطای گرد کردن عدد i ام باشد، در این صورت می‌خواهیم احتمال زیر را تقریب بزنیم:

$$\mathbb{P}(|\sum_{i=۱}^{۵۰} X_i| > ۳) \ = \ ۱ - \mathbb{P}(-۳ \leqslant \sum_{i=۱}^{۵۰} X_i \leqslant ۳)$$

داریم:

$$\sigma^{۲} = (X_i) = \frac{(۰/۵ - (-۰/۰۵))^{۲}}{۱۲} = \frac{۱}{۱۲}$$

$$\mu = \mathbf{E}[X_i] = \frac{-۰/۵ + ۰/۵}{۲} = ۰$$

بنابراین طبق قضیه‌ی حد مرکزی داریم:

$$\mathbb{P}(-۳ \leqslant \sum_{i=۱}^{۵۰} X_i \leqslant ۳) = \mathbb{P}(\frac{-۳-\mu}{\sigma\sqrt{۵۰}} \leqslant \sum_{i=۱}^{۵۰} X_i \leqslant \frac{۳-\mu}{\sigma\sqrt{۵۰}})$$

$$= \mathbb{P}(\frac{-۳}{\sqrt{\frac{۵۰}{۱۲}}} \leqslant \sum_{i=۱}^{۵۰} X_i \leqslant \frac{۳}{\sqrt{\frac{۵۰}{۱۲}}}) \approx \phi(\frac{۳}{\sqrt{\frac{۵۰}{۱۲}}}) - \phi(\frac{-۳}{\sqrt{\frac{۵۰}{۱۲}}})$$

$$\approx ۰/۹۲۷۹ - ۰/۰۷۲۱ = ۰/۸۵۵۸$$

در نتیجه داریم:

$$\mathbb{P}(|\sum_{i=۱}^{۵۰} X_i| > ۳) \approx ۱ - ۰/۸۵۵۸ = ۰/۱۴۴۲$$

▷

مسئله‌ی ۵. سرعت الگوریتم

فرض کنید می‌خواهید سرعت اجرای یک الگوریتم را اندازه بگیرید. می‌دانید که واریانس سرعت اجرا $\sigma^{۲} = ۴sec$ می‌باشد ولی می‌خواهید میانگین زمان اجرا را بدست بیاورد. برای این کار الگوریتم را چندین بار اجرا می‌کنید. چند بار باید الگوریتم را اجرا کنید که زمان اجرا برابر $t \pm ۰/۵$ را با اطمینان ٪۹۵ بدست بیاید.

حل. در واقع می‌خواهیم مقدار n را در معادله زیر حساب کنیم:(فرض کنید میانگین واقعی t است)

$$۰/۹۵ = P(-۰/۵ \leqslant \frac{\sum_{i=۱}^{n} X_i}{n} - t \leqslant ۰/۵)$$

حال به استفاده از قضیه حد مرکزی حساب می‌کنیم:

$$\begin{aligned}
۰/۹۵ &= P(-۰/۵ \leqslant \frac{\sum_{i=۱}^{n} X_i - nt}{n} \leqslant ۰/۵) \\
&= P(-\frac{۰/۵\sqrt{n}}{۲} \leqslant \frac{\sum_{i=۱}^{n} X_i - nt}{۲\sqrt{n}} \leqslant \frac{۰/۵\sqrt{n}}{۲}) \\
&= P(-\frac{۰/۵\sqrt{n}}{۲} \leqslant Z \leqslant \frac{۰/۵\sqrt{n}}{۲}) \\
۰/۹۵ &= \phi(\frac{\sqrt{n}}{۴}) - \phi(-\frac{\sqrt{n}}{۴}) \\
&= ۲\phi(\frac{\sqrt{n}}{۴}) - ۱ \\
۱/۹۵ &= ۲\phi(\frac{\sqrt{n}}{۴}) \\
۱/۹۶ &= \frac{\sqrt{n}}{۴} \\
n &= ۶۱/۴
\end{aligned}$$

$\triangleright$

مسئله‌ی ۶. نامساوی کانتلی

فرض کنید X یک متغیر تصادفی باشد با شرط $\mathbb{E}[X] = ۰$. برای هر عدد حقیقی مثبت مثل a گزاره زیر را اثبات کنید:

$$\mathbb{P}(X \geqslant a) \leqslant \frac{Var(X)}{Var(X) + a^۲}$$

حل. فرض کنید t یک عدد دلخواه نامنفی باشد.

$$\mathbb{P}(X \geqslant a) = \mathbb{P}(X + t \geqslant a + t) \leqslant \mathbb{P}((X+t)^۲ \geqslant (a+t)^۲) \leqslant \frac{\mathbb{E}[(X+t)^۲]}{(a+t)^۲} = \frac{\mathbb{E}[X^۲] + t^۲}{(a+t)^۲} = \frac{VAR(X) + t^۲}{(a+t)^۲}$$

دقت کنید که آخرین نامساوی بالا، مشابه نامساوی $Chebyshev$ بدست می آید.
حالا می خواهیم t را طوری مقدار بدهیم که این عبارت را از $\frac{Var(X)}{Var(X)+a^۲}$ کمتر مساوی کند:

$$\frac{VAR(x)+t^۲}{(a+t)^۲} \leqslant \frac{Var(X)}{Var(X)+a^۲} \Leftrightarrow (VAR(x) + t^۲)(Var(X) + a^۲) \leqslant Var(X)(a+t)^۲$$

$$\Leftrightarrow VAR(x)^۲ + t^۲a^۲ \leqslant ۲atVar(X) \to (VAR(x) - ta)^۲ \leqslant ۰$$

$$\to t = \frac{Var(X)}{a}$$

پس اگر قرار دهیم $t = \frac{Var(X)}{a}$ نتیجه میگریم که

$$\mathbb{P}(X \geqslant a) \leqslant \frac{VAR(X)+t^{۲}}{(a+t)^{۲}} = \frac{VAR(X)+(\frac{Var(X)}{a})^{۲}}{(a+(\frac{Var(X)}{a}))^{۲}} = \frac{Var(X)}{Var(X)+a^{۲}}$$

▷

مسئله‌ی ۷. قانون حد مرکزی پیچیده*

فرض کنید f تابع چگالی احتمالی غیر منفی باشد. می‌دانیم:

$$\int_{\mathbb{R}} f(x)dx = ۱, \quad \int_{\mathbb{R}} xf(x)dx = ۰, \quad \int_{\mathbb{R}} x^{۲}f(x)dx = ۲۵$$

حال مقدار زیر را پیدا کنید: (علامت * به معنای Convolution است)

$$\lim_{n\to\infty}\int_{۱}^{\sqrt{n}} \underbrace{f*f*\cdots*f(x)}_{\text{n-۱ بار}}\,dx$$

حل. ابتدا جمع دو متغیر تصادفی را بدست می‌اوریم. فرض کنید برای Z جمع دو متغیر تصادفی X و Y است، $Z = X + Y$ حال بدست می‌اید.

$$\begin{aligned}
F_Z(z) &= P(Z \leqslant z) = P(X+Y \leqslant z) = \\
&= \int_{-\infty}^{\infty} P(X \leqslant z-y|Y=y)f_Y(y)dy \\
&= \int_{-\infty}^{\infty} P(X \leqslant z-y)f_Y(y)dy \\
f_Z(z) = \frac{\partial F_Z(z)}{\partial z} &= \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy \\
&= f*g(z)
\end{aligned}$$

این فرمول را می‌توانیم به جمع چندین متغیر تصادفی بسط دهیم. حال طبق فرمول به دست آمده به حل می‌پردازیم. فرض کنید $\{X_i\}_i$ دنباله‌ای از متغیرهای تصادفی مستقل با تابع چگالی احتمال $f(x)$ باشد. طبق فرمول امید ریاضی بدست می‌اوریم که: $E[X_i] = ۰$ و می‌دانیم که $E[X_i^{۲}] = ۲۵$ پس واریانس بدست می‌اید:

$$Var(X_i) = E[X_i^{۲}] - E[X_i]^{۲} = ۲۵$$

با استفاده از فرمول حساب شده داریم:

$$\begin{aligned}
&\lim_{n\to\infty}\int_{۱}^{\sqrt{n}} f*f*\cdots*f(x)dx \\
=&\lim_{n\to\infty} P(۱ \leqslant \sum_{k-۱}^{n} X_k \leqslant \sqrt{n}) \\
=&\lim_{n\to\infty} P(\frac{۱}{۵\sqrt{n}} \leqslant \frac{\sum_{k-۱}^{n} X_k - ۰}{۵\sqrt{n}} \leqslant \frac{۱}{۵}) \\
=&P(۰ \leqslant Z \leqslant \frac{۱}{۵})
\end{aligned}$$

و $Z \sim N(۰,۱)$ پس جواب برابر است با: $\phi(\frac{۱}{۵}) - (۱ - \phi(۰))$ ▷

مسئله‌ی ۸. حاج حسین پفکی*

حسین کوپن‌هایی که در بسته‌های پفک هستند را جمع‌آوری می‌کند. n نوع کوپن متفاوت وجود دارند و هر بسته‌ی پفک با احتمال یکسانی حاوی یکی از انواع این کوپن‌هاست. فرض کنید T تعداد بسته‌هایی باشد که حسین باز می‌کند تا تمام n نوع کوپن را پیدا کند. نشان دهید داریم:

$$\mathbb{P}(T \geqslant ۲n \cdot (۱ + \ln n)) \leqslant \frac{\pi^۲/۶}{(۱ + \ln n)^۲}$$

راهنمایی:

$$\sum_{i=۱}^{n} \frac{۱}{i^۲} \leqslant \frac{\pi^۲}{۶}$$

حل. فرض کنید t_i تعداد بسته‌هایی باشند که حسین بعد از دیدن $i - ۱$ امین کوپن منحصر به فرد باز می‌کند تا به i امین کوپن جدید برسد. در این صورت داریم:

$$T = t_۱ + \cdots + t_n$$

می‌توان دید که در صورتی که i نوع کوپن مختلف دیده باشیم، احتمال دیدن یک نوع جدید برابر است با $\frac{n-i+۱}{n}$ بنابراین $t_i \sim Geom(\frac{n-i+۱}{n})$. اکنون با توجه به خطی بودن امیدریاضی داریم:

$$\begin{aligned} \mathbf{E}[T] = \mathbf{E}[t_۱] + \mathbf{E}[t_۲] + \cdots + \mathbf{E}[t_n] \ &= \ \frac{n}{n} + \frac{n-۱}{n} + \cdots + \frac{۱}{n} \\ &= n \cdot (\frac{۱}{۱} + \frac{۱}{۲} + \cdots + \frac{۱}{n}) \leqslant n \cdot (۱ + \ln n) \end{aligned}$$

زیرا:

$$\frac{۱}{۲} + \cdots + \frac{۱}{n} \leqslant \int_{۱}^{n} \frac{dx}{x} = \ln n$$

و برای واریانس داریم(با استفاده از مستقل بودن t_i ها):

$$\begin{aligned} \mathbf{Var}(T) &= \mathbf{Var}(t_۱) + \mathbf{Var}(t_۲) + \cdots + \mathbf{Var}(t_n) \\ &= \frac{۱ - p_۱}{p_۱} + \frac{۱ - p_۲}{p_۲} + \cdots + \frac{۱ - p_n}{p_n} \\ &\leqslant \frac{۱}{p_۱^۲} + \frac{۱}{p_۲^۲} + \cdots + \frac{۱}{p_n^۲} \\ &= n^۲ \cdot (\frac{۱}{۱^۲} + \frac{۱}{۲^۲} + \cdots + \frac{۱}{n^۲}) \leqslant \frac{n^۲\pi^۲}{۶} \end{aligned}$$

اکنون از نابرابری چبیشف استفاده می‌کنیم:

$$\begin{aligned} &\mathbb{P}(|T - n \cdot (۱ + \ln n)| \geqslant n \cdot (۱ + \ln n)) \leqslant \frac{n^۲\pi^۲/۶}{(n \cdot (۱ + \ln n))^۲} \\ &\to \mathbb{P}(T \geqslant ۲n \cdot (۱ + \ln n)) \leqslant \frac{\pi^۲/۶}{(۱ + \ln n)^۲} \end{aligned}$$

▷

مسئله‌ی ۹. پواسون زنجیرزن*

با استفاده از زنجیره‌ای از متغیرهای تصادفی پواسون عبارت زیر را اثبات کنید.

$$\lim_{n \to \infty} e^{-n} \sum_{k=۰}^{n} \frac{n^k}{k!} = \frac{۱}{۲}$$

حل. می‌دانیم حاصل جمع تعداد زیادی متغیر تصادفی iid، متغیر تصادفی‌ای از توزیع نرمال خواهد بود.

$$\lambda = ۱ \quad ، \quad X_i \sim Pois(۱) \quad ، \quad Z = \sum_{i=۱}^{n} X_i \sim \mathcal{N}(n, n)$$

از طرفی می‌دانیم حاصل جمع تعدادی متغیر پواسون مستقل نیز متغیری پواسون است.

$$Z = \sum_{i=۱}^{n} X_i \sim Pois(n)$$

سپس، مقدار $P[Z \leqslant n]$ را یک بار با استفاده از توزیع پواسون و باری دیگر به کمک توزیع نرمال محاسبه می‌کنیم:

$$Z \sim Pois(n) \rightarrow P[Z \leqslant n] = \sum_{i=۰}^{n} \frac{e^{-n} n^i}{i!}$$

از آنجایی که در اینجا n میانگن توزیع نرمال Z است، بنابرین احتمال تجمعی آن برابر $\frac{۱}{۲}$ خواهد بود:

$$Z \sim \mathcal{N}(n, n) \rightarrow P[Z \leqslant n] = \frac{۱}{۲}$$

بنابرین،

$$\lim_{n \to \infty} \sum_{i=۰}^{n} \frac{e^{-n} n^i}{i!} = \frac{۱}{۲}$$

▷

موفق باشید :)