



University of Tehran

آمار و احتمالات مهندسی

تمرین چهارم - متغیرهای تصادفی توأم توزیع شده

آرین و رضا

تاریخ تحویل ۱۴۰۰/۰۸/۲۳

سؤال ۱.

توزیع احتمال رشد جمعیت یک شهر خاص از تابع توزیع زیر پیروی می کند:

$$f(x, k, \theta) = \begin{cases} \frac{k \cdot \theta^k}{x^{k+1}} & \theta \leq x \\ 0.W. & \end{cases}$$

که در آن مقادیر k و θ دو ثابت مثبت هستند.

۱. ثابت کنید که این تابع یک تابع توزیع احتمال است.

۲. در صورتی که $k > 1$ امید ریاضی را محاسبه کنید.

۳. در صورتی که $k > 2$ مقدار واریانس را بیابید.

پاسخ.

۱. برای این مورد ابتدا باید ثابت کنیم که تابع همواره مثبت است و سپس باید ثابت کنیم مساحت زیر سطح نمودار برابر با ۱ است. با توجه به این که تابع برای مقادیر بزرگتر از صفر تعریف شده است و مقدار θ نیز همواره مثبت است، مقدار تابع در هر نقطه مثبت است. همچنین داریم:

$$\int_{\theta}^{\infty} f(x, k, \theta) dx = k\theta^k \int_{\theta}^{\infty} x^{-(k+1)} dx = -\frac{k\theta^k}{kx^k} \Big|_{\theta}^{\infty} = 1$$

۲. با توجه به رابطه‌ی محاسبه‌ی امید ریاضی داریم:

$$E(X) = \int_{\theta}^{\infty} x f(x, k, \theta) dx = k\theta^k \int_{\theta}^{\infty} x^{-k} dx = \frac{k\theta^k}{1-k} x^{1-k} \Big|_{\theta}^{\infty} = \frac{k\theta}{k-1}$$

۳. ابتدا گشتاور مرتبه دوم را محاسبه می‌کنیم و سپس به محاسبه‌ی واریانس می‌پردازیم:

$$E(X^2) = \int_{\theta}^{\infty} x^2 f(x, k, \theta) dx = k\theta^k \int_{\theta}^{\infty} x^{2-k} dx = \frac{k\theta^k}{2-k} x^{2-k} \Big|_{\theta}^{\infty} = \frac{k\theta^2}{k-2}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{k\theta^2}{k-2} - \left(\frac{k\theta}{k-1} \right)^2$$

سؤال ۲.

فرض کنید X متغیر تصادفی نشان‌دهنده‌ی تخمینی از مقدار هزینه‌ی یک فرد خاص (بر حسب میلیون تومان) در طول یک سال باشد:

۱. اگر تابع چگالی احتمال این متغیر تصادفی به صورت $f(x) = k(1 + \frac{x}{5})^{-\gamma}$ $x \geq 0$ باشد، مقدار k را بیابید.

۲. فرض کنید این فرد از یک شرکت بیمه استفاده می‌کند که برنامه‌ی پرداختی آن بدین شرح است:

برای هزینه‌های کمتر از ۵۰۰ هزار تومان چیزی نمی‌پردازد و برای هزینه‌های بیشتر از ۵۰۰ هزار تومان نیز تا ۸۰ درصد هزینه‌ی باقی‌مانده را می‌پردازد (یعنی ابتدا فرد ۵۰۰ هزار تومان را می‌پردازد و سپس بیمه ۸۰ درصد هزینه‌ی باقی‌مانده را می‌پردازد). همچنین در برنامه‌ی این بیمه بیشترین هزینه‌ای که فرد می‌پردازد نیز ۲ میلیون و ۵۰۰ هزار تومان است (یعنی در صورتی که هزینه‌ی درمانی بیشتر از این مقدار شد، فرد تنها همین مقدار ۲ میلیون و ۵۰۰ هزار تومان را می‌پردازد و سایر هزینه‌ها هر چقدر باشد توسط بیمه پرداخت می‌شود).

ابتدا تابع مقدار هزینه‌ی پرداخت شده توسط بیمه را بر حسب متغیر تصادفی X بنویسید و سپس امید ریاضی آن را محاسبه کنید.

پاسخ.

۱. برای محاسبه‌ی k کافیست معادله‌ی $\int_{-\infty}^{\infty} f(x) dx = 1$ را حل کنیم تا به پاسخ برسیم:

$$\int_{-\infty}^{\infty} k \left(1 + \frac{x}{5}\right)^{-\gamma} dx = k \int_{-\infty}^{\infty} \left(1 + \frac{x}{5}\right)^{-\gamma} dx = \frac{5k}{\gamma} \int_1^{\infty} u^{-\gamma} du = -\frac{5k}{\gamma-1} u^{-\gamma+1} \Big|_1^{\infty} = 1$$

$$k = \frac{\gamma-1}{5}$$

۲. تابع مورد نظر به صورت زیر می‌شود:

$$h(x) = \begin{cases} 0 & x < 500, \\ \frac{4}{5}(x - 500) & 500 \leq x \leq 10500, \\ x - 2500 & x \geq 10500 \end{cases}$$

بنابراین کافیست امید ریاضی تابع جدید را محاسبه کنیم:

$$\begin{aligned} E(h(X)) &= \int_{-\infty}^{\infty} h(x) k \left(1 + \frac{x}{5}\right)^{-\gamma} dx \\ &= k \int_{500}^{10500} \frac{4}{5}(x - 500) \left(1 + \frac{x}{5}\right)^{-\gamma} dx + k \int_{10500}^{\infty} (x - 2500) \left(1 + \frac{x}{5}\right)^{-\gamma} dx \\ &= \gamma k \int_{201}^{4201} (2/5 u - 502/5) u^{-\gamma} du + \frac{5}{\gamma} \int_{4201}^{\infty} ((2/5 u - 2502/5) u^{-\gamma} du \\ &= \frac{1}{\gamma} \left\{ \frac{1}{5} \left(\frac{1}{201^{\gamma}} - \frac{1}{4201^{\gamma}} \right) - \frac{201}{\gamma} \left(\frac{1}{201^{\gamma}} - \frac{1}{4201^{\gamma}} \right) \right\} + \frac{1}{\gamma} \left\{ \frac{5}{4 \times 4201^{\gamma}} - \frac{2502}{24 \times 4201^{\gamma}} \right\} \end{aligned}$$

سؤال ۳.

رستوران زنجیره‌ای مک دونالد اخیراً ادعا کرده است که به منظور اهمیت داشتن سلامتی مشتریان از روغنی سالم‌تر برای پخت غذاهایش استفاده می‌کند اما از طرفی مزه‌ی غذاها چندان تغییر نکرده است به طوری که از بین هر ۱۰۰ نفری که یک غذای طبخ شده با آن روغن را می‌چشند ۹۷ نفر متوجه تغییر نمی‌شوند. با فرض درست بودن این ادعا در یک نمونه‌ی صدتایی از مشتریان موارد زیر را تخمین بزنید:

۱. احتمال این که حداقل ۴۰ درصد از افراد متوجه تفاوت مزه‌ی غذای طبخ شده با روغن جدید بشوند.

۲. احتمال این که حداکثر ۵ درصد متوجه این تفاوت بشوند.

پاسخ.

برای حل این سوال از تخمین توزیع دوجمله‌ای با استفاده از توزیع پواسون استفاده می‌کنیم چرا که مقدار n به اندازه‌ای نیست که بتوان با توزیع نرمال آن را تخمین زد (شرط $np(1-p) \geq 10$ برای این مسأله صادق نیست) ما با شرایط تخمین با توزیع پواسون همخوانی دارد در نتیجه با $\lambda = np$ آن را تخمین می‌زنیم.

با توجه به آن که از بین صد نفری که غذا را چشیده‌اند، تنها سه نفر متوجه تفاوت شده‌اند احتمال تشخیص تفاوت برابر است با $p = 0.03$. با توجه به این مورد داریم:

$$\lambda = np = 3$$

در نتیجه برای یافتن پاسخ دو پرسش مطرح شده باید از توزیع $Poi(3)$ استفاده کنیم.

۱. احتمال این که حداقل ۴۰ درصد متوجه تغییر شوند برابر با $P(X \geq 40)$ است. داریم:

$$P(X \geq 40) = 1 - P(X < 40) = \sum_{i=0}^{39} \frac{3^i e^{-3}}{i!} = e^{-3} \sum_{i=0}^{39} \frac{3^i}{i!}$$

۲. احتمال این که حداکثر ۵ درصد متوجه این تفاوت بشوند برابر با $P(X \leq 5)$ است. داریم:

$$P(X \leq 5) = \sum_{i=0}^5 \frac{3^i e^{-3}}{i!} = e^{-3} \sum_{i=0}^5 \frac{3^i}{i!}$$

سؤال ۴.

فردی می‌خواهد یک بسته‌ی مکالمه‌ی تلفن همراه بخرد و با دو گزینه روبرو است. در گزینه‌ی اول به ازای هر دقیقه مکالمه باید ۱۰ تومان بپردازد و در گزینه‌ی دوم به ازای هر مدت زمان کمتر از ۲۰ دقیقه ۹۹ تومان و به ازای هر دقیقه‌ای که از ۲۰ دقیقه می‌گذرد نیز ۱۰ تومان اضافی باید بپردازد (یعنی در صورتی که مکالمه ۲۲ دقیقه طول بکشد باید ۱۱۹ تومان بپردازد). در صورتی که طول مکالمه‌ی این فرد از توزیع نمایی با پارامتر λ پیروی کند، مشخص کنید کدام گزینه برای وی مناسب‌تر است در صورتی که:

۱. میانگین زمان مکالمه‌ی وی ۱۰ دقیقه باشد.

۲. میانگین زمان مکالمه‌ی وی ۱۵ دقیقه باشد.

راهنمایی: برای مقایسه‌ی گزینه‌ها از امید ریاضی استفاده کنید.

پاسخ.

ابتدا موارد زیر را در نظر می‌گیریم:

X : مدت زمان مکالمه

Y_1 : متغیر تصادفی هزینه‌ی مکالمه با استفاده از بسته‌ی نخست

Y_2 : متغیر تصادفی هزینه‌ی مکالمه با استفاده از بسته‌ی دوم

در هر دو حالت داریم:

$$E(Y_1) = 10 \times E(X)$$

$$E(Y_2) = P(X \leq 20) \times E(Y_2|X \leq 20) + P(X > 20) \times E(Y_2|X > 20)$$

از طرفی در محاسبه‌ی $E(Y_2|X > 20)$ می‌دانیم که برای ۲۰ دقیقه‌ی نخست باید ۹۹ تومان بپردازیم و برای هر دقیقه‌ی بیش از آن ۱۰ تومان. همچنین از آنجایی که زمان مکالمه از توزیع نمایی پیروی می‌کند و این توزیع نیز بی‌حافظه است، توزیع مدت زمان مکالمه‌ی بیش از ۲۰ دقیقه با توزیع مدت زمان مکالمه‌ی اولیه برابر است و در نتیجه میانگین آن نیز ۱۰ دقیقه می‌شود و داریم:

$$E(Y_2|X > 20) = 99 + 10 \times E(X)$$

حال با جایگذاری مقادیر در هر دو حالت به پاسخ نهایی می‌رسیم.

۱. با توجه به این که در توزیع نمایی مقدار میانگین برابر با $\frac{1}{\lambda}$ است، در نتیجه مقدار λ برابر با $\frac{1}{10}$ می‌گردد. داریم:

$$E(Y_1) = 10 \times 10 = 100$$

$$\begin{aligned} E(Y_2) &= P(m \leq 20) \times 99 + (1 - P(m \leq 20)) \times (99 + 100) \\ &= P(m \leq 20) \times 99 + (1 - P(m \leq 20)) \times 199 \end{aligned}$$

در نتیجه:

$$E(Y_2) = (1 - e^{-0.1 \times 20}) \times 99 + e^{-0.1 \times 20} \times 199 = 99 + 100e^{-2} \simeq 112.53$$

بنابراین در حالتی که میانگین مدت زمان مکالمه ده دقیقه باشد میزان پرداختی به طور میانگین در حالت نخست کمتر است.

۲. در این حالت نیز داریم:

$$\lambda = \frac{1}{15} \simeq 0.06$$

در نتیجه:

$$E(Y_1) = 10 \times 15 = 150$$

$$\begin{aligned} E(Y_2) &= P(m \leq 20) \times 99 + (1 - P(m \leq 20)) \times (99 + 150) = \\ &= (1 - e^{-\frac{20}{15}}) \times 99 + e^{-\frac{20}{15}} \times 249 = 99 + 150e^{-\frac{4}{3}} \simeq 138.53 \end{aligned}$$

در نتیجه در این حالت که میانگین زمان مکالمه یک ربع است، میزان پرداختی در حالت دوم به طور میانگین کمتر است.

سؤال ۵.

یک بوم‌شناس می‌خواهد در یک محوطه نمونه‌برداری دایروی با توزیع یکنواخت نقطه‌ای را مشخص کند. فرض کنید $(x, y) \sim (X, Y)$ مختصات این نقطه باشد. اگر مرکز دایره در $(0, 0)$ بوده و شعاع دایره نیز R باشد،

الف) احتمال آنکه نقطه انتخاب شده در فاصله $\frac{R}{\sqrt{2}}$ از مرکز دایره باشد چقدر است؟

ب) احتمال آنکه هم y و هم x هر دو از 0 حداکثر $\frac{R}{\sqrt{2}}$ فاصله داشته باشد چقدر است؟

پ) تابع چگالی احتمال حاشیه‌ای X و Y چیست؟

ت) آیا این دو متغیر مستقلند؟

پاسخ.

با توجه به توضیحات داده شده تابع چگالی احتمال دو متغیر X و Y به شکل زیر است.

$$f(x, y) = \begin{cases} \frac{1}{\pi R^2} & x^2 + y^2 \leq R^2 \\ 0 & \text{otherwise} \end{cases}$$

الف) برای پاسخ به این سوال کافیت از این تابع در فضای دایره به شعاع $\frac{R}{\sqrt{2}}$ انتگرال بگیریم. لذا A در اینجا فضای یک دایره به شعاع $\frac{R}{\sqrt{2}}$ است.

$$P[(X, Y) \in A] = \iint_A f(x, y) dx dy = \iint_A \frac{1}{\pi R^2} dx dy = \frac{1}{\pi R^2} \pi \left(\frac{R}{\sqrt{2}}\right)^2 = \frac{1}{4}$$

ب) از آنجا که هر دو متغیر x و y می‌توانند از $-\frac{R}{\sqrt{2}}$ تا $\frac{R}{\sqrt{2}}$ حرکت کنند، اشتراک این دو فضا یک مربع به ضلع $\sqrt{2}R$ می‌شود. لذا A در اینجا مربعی به ضلع $\sqrt{2}R$ است.

$$P[(X, Y) \in A] = \iint_A f(x, y) dx dy = \iint_A \frac{1}{\pi R^2} dx dy = \frac{1}{\pi R^2} (\sqrt{2}R)^2 = \frac{2}{\pi}$$

پ) حدود x و y را درون این دایره بدست می‌آوریم.

$$x^2 + y^2 = R^2$$

لذا x از $-\sqrt{R^2 - y^2}$ تا $\sqrt{R^2 - y^2}$ و y از $-\sqrt{R^2 - x^2}$ تا $\sqrt{R^2 - x^2}$ است.

$$f_X(x) = \begin{cases} \int_{-\sqrt{R^2 - x^2}}^{\sqrt{R^2 - x^2}} \frac{1}{\pi R^2} dy = \frac{1}{\pi R^2} * 2\sqrt{R^2 - x^2} & -R \leq x \leq R \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \int_{-\sqrt{R^2 - y^2}}^{\sqrt{R^2 - y^2}} \frac{1}{\pi R^2} dx = \frac{1}{\pi R^2} * 2\sqrt{R^2 - y^2} & -R \leq y \leq R \\ 0 & \text{otherwise} \end{cases}$$

ت) برای آنکه این دو متغیر مستقل باشند لازم است تا pdf مشترک آن دو با حاصل ضرب pdf هر یک از آن دو متغیر باشد که اینطور نیست. پس مستقل نیستند.

سؤال ۶.

دو تاپیست با یکدیگر بر سر سریع‌تر تاپ کردن مسابقه می‌دهند. فرض کنید متغیر X و متغیر Y به ترتیب نمایان‌گر تعداد اشتباهات تایی این دو تاپیست باشد. اگر اشتباهات این دو تاپیست به یکدیگر ارتباطی نداشته و توزیع احتمالی این دو متغیر نیز توزیع پواسون با پارامتر μ_1 و μ_2 باشند:

الف) تابع احتمال مشترک این دو متغیر چیست؟

ب) احتمال آنکه حداکثر یک اشتباه در مجموع هر دو متن تایپ شده رخ داده باشد چقدر است؟

پ) یک عبارت کلی برای احتمال آنکه مجموع تعداد ایرادات تایپی در متن تایپ شده توسط هر دو تایپیست برابر عدد k باشد بنویسید.

پاسخ.

الف) از آنجا که دو متغیر مستقلند، کافیت pmf این دو را در هم ضرب کنیم تا pmf مشترک آن دو بدست آید.

$$P_{XY}(x, y) = \begin{cases} p_X(x) \times p_Y(y) = e^{-\mu_1} \frac{\mu_1^x}{x!} \times e^{-\mu_2} \frac{\mu_2^y}{y!} = e^{-\mu_1 - \mu_2} \frac{\mu_1^x \mu_2^y}{x! y!} & x, y \in \mathbb{N}, \\ \text{otherwise} \end{cases}$$

ب) کافیت این عبارت را حساب کنیم. $P(X + Y \leq 1) = p(0, 0) + p(1, 0) + p(0, 1)$ با جایگذاری اعداد در تابع بدست آمده در بخش الف حاصل برابر می شود با

$$e^{-\mu_1 - \mu_2} (1 + \mu_1 + \mu_2)$$

پ) $P(X + Y = k) = \sum_{m=0}^k P(X = m, Y = k - m)$ باقی را با جواب بخش های قبل جاگذاری و محاسبه می کنیم.

$$\begin{aligned} \sum_{m=0}^k e^{-\mu_1 - \mu_2} \frac{\mu_1^m \mu_2^{k-m}}{m! (k-m)!} &= \sum_{m=0}^k \frac{k!}{k!} e^{-\mu_1 - \mu_2} \frac{\mu_1^m \mu_2^{k-m}}{m! (k-m)!} = e^{-\mu_1 - \mu_2} \times \frac{1}{k!} \times \sum_{m=0}^k \mu_1^m \mu_2^{k-m} \times \binom{k}{m} \\ &= e^{-\mu_1 - \mu_2} \times \frac{1}{k!} \times (\mu_1 + \mu_2)^k \end{aligned}$$

سؤال ۷.

میانگین و انحراف معیار نشست سه پایه شرقی، میانی و غربی یک پل براساس جدول زیر است (اعداد برحسب سانی متر هستند). می توان پذیرفت که نشست ها توزیع نرمال دارند و مستقل از هم هستند.

پایه	میانگین	انحراف معیار
شرقی	۱/۰	۰/۳
میانی	۱/۵	۰/۵
غربی	۱/۰	۰/۳

آ) احتمال این که مقدار نشست بیشینه از ۲ سانی متر بیشتر شود چقدر است؟ (اگر مقدار نشست حداقل یکی از پایه های پل بیشتر از ۲ سانی متر شود، آنگاه کل پل هم بیشتر از ۲ سانی متر نشست خواهد داشت).

ب) مقدار نشست مجاز پایه میانی پل را چنان تعیین کنید که احتمال افزایش نشست این پایه از مقدار مجاز تعیین شده به ۰/۰۰۲ محدود گردد.

پاسخ.

آ) برای محاسبه احتمال خواسته شده در این قسمت ابتدا احتمال اینکه مقدار نشست هر یک از پایه ها بیشتر از ۲ سانی متر شود را حساب می کنیم:

$$A = \text{شرقی} \Rightarrow P(X > 2) = 1 - P\left(\underbrace{\frac{X-1}{0.3}}_{=Z} \leq \frac{2-1}{0.3}\right) = 1 - 0.9996 = 0.0004$$

$$B = \text{میانی} \Rightarrow P(X > 2) = 1 - P\left(\underbrace{\frac{X - 1/5}{0/5}}_{=Z} \leq \frac{2 - 1/5}{0/5}\right) = 1 - 0/8413 = 0/1587$$

$$C = \text{غربی} \Rightarrow P(X > 2) = 1 - P\left(\underbrace{\frac{X - 1}{0/3}}_{=Z} \leq \frac{2 - 1}{0/3}\right) = 1 - 0/9996 = 0/0004$$

حال با توجه به خواسته سوال اجتماع رخ دادن هریک از پیشامدهای به دست آمده را حساب می کنیم. نکته حائز اهمیت این است که در صورت سوال ذکر شده این پیشامدها مستقل از یکدیگر هستند:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= (0/0004) + (0/1587) + (0/0004) - (0/0004 \times 0/1587) - (0/0004 \times 0/0004) - (0/1587 \times 0/0004) + (0/0004 \times 0/1587 \times 0/0004) \end{aligned}$$

احتمال اینکه مقدار نشست بیشینه پل بیشتر از ۲ سانتی متر شود، تقریباً ۰/۱۵۹۴ است.

(ب)

$$P\left(Z \geq \frac{X - 1/5}{0/5}\right) = 0/002 \Rightarrow P\left(Z \leq \frac{X - 1/5}{0/5}\right) = 0/998 \xrightarrow{Z\text{-table}} \frac{X - 1/5}{0/5} = 2/88 \Rightarrow X = 2/94$$

اگر مقدار نشست مجاز پایه پل ۲/۹۴ باشد. آنگاه احتمال افزایش نشست این پایه از مقدار مجاز ۰/۰۰۲ خواهد بود.

سؤال ۸.

اختیاری: این تمرین تحویل ندارد. در صورت تمایل برای بحث و گفتگو پیرامون این تمرین، با ایمیل behzad.shayegh@ut.ac.ir در ارتباط باشید.

بخش هفتم سری تمرینات کامپیوتری با موضوع مدل بیزی با دیگر توزیع ها را می توانید از طریق این لینک^۱ دریافت کنید.

پاسخ.

تکمیل شده ی فایل صورت سوال از طریق این لینک^۲ در دسترس است.

سؤال ۹.

سری هشتم تمرینات کامپیوتری با موضوع تخمین توزیع دوجمله ای به کمک توزیع های نرمال و پواسون را می توانید از طریق این لینک^۳ دریافت کنید.

- یک کپی از فایل مذکور با نام CA4_S8_SID در گوگل درایو خود تهیه کنید.
- در فایل خود بخش هایی که به وسیله مستطیل مشخص شده اند را با کدهای مناسب جایگزین کنید. در تکمیل کدها، از حلقه های تکرار استفاده نکنید.
- پاسخ سوالاتی که در متن فارسی مطرح و بدون پاسخ رها شده اند را در همان سلول اضافه کنید.

¹https://colab.research.google.com/drive/1opodKXSJkV_I_P0tsvLDkr14KH_ZmhVISb?usp=sharing

²<https://colab.research.google.com/drive/1OEqxEO81Mn8PLyW3gaqxsZ88d9oIgwpl1?usp=sharing>

³<https://colab.research.google.com/drive/12fmQ6VyKGFYQV3SM4mzZp91P3-t1b004?usp=sharing>

- فایل کد خود را با ایمیل `gelamm۷۶@gmail.com` با دسترسی Edit به اشتراک بگذارید.
 - لینک فایل پاسخ خود را در بخش متنی جایگاه آپلود این تمرین در سامانه ایلرن قرار دهید.
- هرگونه انتقاد، پیشنهاد یا نکته جانبی را می‌توانید از طریق یک سلول متنی در ابتدای فایل (قبل از سرفصل اصلی) به ما منتقل کنید.

پاسخ.

تکمیل شده‌ی فایل صورت سوال از طریق این لینک^۴ در دسترس است.

^۴<https://colab.research.google.com/drive/1h5n-inAS3FW76qPjp6BjrnLSbkl-SH5x?usp=sharing>