University of Tehran

آمار و احتمالات مهندسی

تمرین پنجم - توزیع بتا

علی و کیمیا

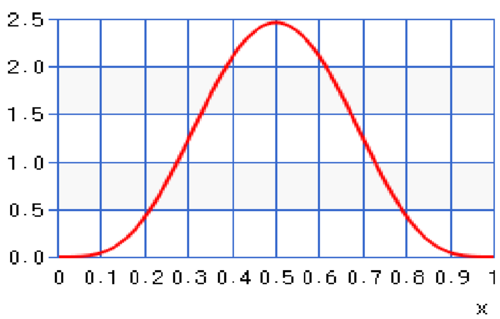
تاریخ تحویل ۱۴۰۰/۰۹/۱۵

سؤال ۱.

متغیر تصادفی X دارای تابع چگالی احتمال زیر است:

$$f_X(x) = \begin{cases} 0.25x^3 & , 0 < x < 1 \\ 0 & , \text{otherwise} \end{cases}$$

همچنین متغیر تصادفی Y دارای توزیع بتا با تابع چگالی احتمال زیر است،



اگر X و Y از یکدیگر مستقل باشند، $Cov(X + 0.25, XY)$ را محاسبه کنید.

پاسخ .

طبق نمودار، چون تابع چگالی احتمال بتا متقارن است در نتیجه الفا و بتا برابر هستند بنابراین داریم: $E(Y) = 0.5$

$$\begin{aligned} Cov(X + 0.25, XY) &= \text{Cov}(X, XY) \\ &= E(X \cdot XY) - E(X) \cdot E(XY) \\ &= E(X^2)E(Y) - (E(X))^2 E(Y) \\ &= 0.5 \cdot (E(X^2) - (E(X))^2) \\ &= 0.5 \cdot \text{Var}(X) \end{aligned}$$

سؤال ۲.

زوج (X, Y) مختصات یک نقطه بر روی دایره‌ی واحد هستند (یعنی $\{(x, y)|x^۲ + y^۲ \leqslant ۱\}$). مکان این نقطه به‌صورت تصادفی و با توزیع یکنواخت بر روی دایره‌ی واحد تعیین می‌شود. نشان دهید X و Y از هم مستقل نیستند اما $cov[X, Y] = ۰$.

پاسخ .

دو متغیر تصادفی X و Y مستقل‌اند اگر و تنها اگر متغیر (X, Y) توزیع توام تجمعیِ

$$F_{X,Y}(x, y) = F_X(x).F_Y(y)$$

داشته باشد. یا اگر چگالی توام موجود باشد:

$$f_{X,Y}(x, y) = f_X(x).f_Y(y)$$

برای محاسبه‌ی $f_{X,Y}(x, y)$ دقت به این نکته کافی‌ست که انتگرال آن باید روی سطح دایره‌ی واحد برابر ۱ باشد پس:

$$f_{X,Y}(x, y) = ۱/\pi$$

از طرفی به‌سادگی می‌توان نشان داد $f_X(x)$ در نقطه‌ی $x = ۰$ بیشینه است و $f_X(۱) = ۰$. زیرا

$$f_X(x) = \int_{-\sqrt{۱-x^۲}}^{\sqrt{۱-x^۲}} f_{X,Y}(x, y)\, dy$$
$$= \int_{-\sqrt{۱-x^۲}}^{\sqrt{۱-x^۲}} \frac{dy}{\pi}$$
$$= \frac{۲\times\sqrt{۱-x^۲}}{\pi}$$

پس $f_X(۱) = ۰$ به همین صورت برای Y داریم: $f_Y(۱) = ۰$ پس:

$$f_X(۱) \cdot f_Y(۱) \neq f_{X,Y}(x, y)$$

پس X و Y مستقل نیستند. برای محاسبه‌ی $cov[X, Y]$ داریم:

$$cov[X, Y] = E[XY] - E[X]E[Y]$$

برای محاسبه‌ی $E[x]$ به این نکته توجه می‌کنیم که $f_X(x) = \frac{۲}{\pi}\sqrt{۱ - x^۲}$ نسبت به $x = ۰$ متقارن است پس $E[X] = ۰$. با یک مشاهده‌ی همانند می‌توان به‌سادگی دریافت که $E[XY]$ نیز برابر صفر است: انتگرال محاسبه‌ی میانگین XY روی سطح دایره‌ی واحد را می‌توان به چهار جمع‌وند برای چهار ربع محورهای مختصات تقسیم کرد.

$$g(x, y) = x \cdot y \cdot f_X(x) \cdot f_Y(y)$$
$$E[XY] = \int_{(۱)} g(x, y)dxdy + \int_{(۲)} g(x, y)dxdy + \int_{(۳)} g(x, y)dxdy + \int_{(۴)} g(x, y)dxdy$$

به سادگی می‌توان نشان داد انتگرال‌های (۱) و (۲) قرینه هستند. زیرا برای هر y مشخص، مقدار $g(x, y)$ روی محور x متقارن است. همچنین به شیوه‌ی همانندی می‌توان نشان داد انتگرال‌های (۳) و (۴) نیز قرینه‌اند و بنابراین $E[XY] = ۰$ و درنتیجه

$$cov[X, Y] = ۰ - ۰ = ۰$$

سؤال ۳.

شرکتی در دفترچه‌ی راهنمای محصولاتش طول عمر محصولات را با متغیر تصادفی $P(T \geq t) = e^{-t/۵}$ نشان می‌دهد که T نمایشگر تعداد سال‌هایی‌ست که محصول تا قبل از آنکه خراب شود به درستی کار می‌کند (عمر محصول). به عنوان مثال احتمال آنکه محصول بیشتر یا مساوی ۲ سال کار کند برابر است با: $P(T \geq ۲) = e^{-۲/۵} = ۰/۶۷۰۳$.

الف) محصولی ازین شرکت خریده‌اید و تا دو سال بدون پیش آمدن مشکلی از آن استفاده کرده‌اید. احتمال آنکه این محصول در سال سوم خراب شود چقدر است؟

ب) اگر محصول تا n سال بدون پیش آمدن مشکلی استفاده شده باشد، احتمال آنکه در سال $n + ۱$ام خراب شود چقدر است؟ (n یک عدد طبیعی است)

پاسخ .

الف) فرض کنید A پیشامد آن باشد که یک محصول خریداری شده در سال سوم خراب شود. همچنین، فرض کنید B پیشامد آن باشد که یک محصول خریداری شده، در دو سال خراب نشوند. ما به دنبال محاسبه $P(A|B)$ هستیم. برای محاسبه آن داریم:

$$P(B) = P(T \geq ۲) \\ = e^{-\frac{۲}{۵}}$$

همچنین داریم:

$$P(A) = P(۲ \leq T \leq ۳) \\ P(T \geq ۲) - P(T \geq ۳) \\ e^{-\frac{۲}{۵}} - e^{\frac{۳}{۵}}$$

از آنجایی که داریم: $A \subset B$، پس $A \cap B = A$. لذا خواهیم داشت:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \\ = \frac{P(A)}{P(B)} \\ = \frac{e^{-\frac{۲}{۵}} - e^{\frac{۳}{۵}}}{e^{-\frac{۲}{۵}}} \\ = ۰٫۱۸۱۳$$

ب) راه‌حلی که در بخش الف به کار گرفته‌ایم تکرار می‌کنیم با این تفاوت که در این بخش به دنبال جواب کلی مسئله هستیم پس نظیر سال ۲ در بخش قبلی مقدار n و نظیر سال سوم مقدار $n + ۱$ را قرار می‌دهیم. با تکرار قدم‌های قبلی در نهایت به این عبارت خواهیم رسید:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \\ = \frac{P(A)}{P(B)} \\ = \frac{e^{-\frac{n}{۵}} - e^{\frac{-n-۱}{۵}}}{e^{-\frac{n}{۵}}} \\ = \frac{e^{-\frac{n}{۵}}(۱-e^{-\frac{۱}{۵}})}{e^{-\frac{n}{۵}}} \\ = ۰٫۱۸۱۳$$

سؤال ۴.

مسابقه‌ای در دانشگاه در حال برگزاریست. در جریان این مسابقه، تعداد زیادی کوپن که روی هریک از آن‌ها زوج مرتب (x, y) نوشته شده، در جعبه‌ای بزرگ قرار داده شده‌اند. بازیکنان باید در زمان مسابقه حداکثر تعداد از کوپن‌هایی که مجموع x و y یکسان دارند جمع‌آوری کنند و در نهایت برنده‌ی بازی شرکت‌کننده‌ایست که بیشترین تعداد کوپن با مجموع یکسان را گرد آورده باشد. در صورتی که شما قصد شرکت در این مسابقه را داشته باشید و بدانید که x و y مستقلند و توزیع x و y های داخل جعبه به صورت زیر است:

$$f_X(x) = \begin{cases} exp(-x) & x \in R_X \\ ۰ & otherwise \end{cases} \quad R_X = [۰, \infty)$$

$$f_Y(y) = \begin{cases} exp(-y) & y \in R_Y \\ ۰ & otherwise \end{cases} \quad R_Y = [۰, \infty)$$

آنگاه از نظر شما جمع‌آوری کوپن‌هایی با کدام حاصل‌جمع x و y هوشمندانه‌تر است؟

پاسخ .

برای آنکه بتوانیم در این بازی هوشمندانه‌ترین انتخاب را داشته باشیم، باید روی مجموعی از x و y سرمایه‌گذاری کنیم که احتمال مشاهده شدن آن از دیگر حالت‌های ممکن بیشتر است. بدین منظور لازم است با استفاده از توزیع دو متغیر تصادفی مستقل x و y، توزیع متغیر تصادفی z که همان مجموع $x + y$ را نشان میدهد، محاسبه کنیم. نحوه‌ی محاسبه‌ی توزیع متغیر تصادفی z در زیر نشان داده شده است.

$$Z = X + Y \\ R_Z = [۰, \infty)$$

به ازای $z \in R_Z$ داریم:

$$f_Z(Z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy \\ = \int_{۰}^{\infty} f_X(z-y) exp(-y)\, dy$$

$$
\begin{aligned}
&= \int_0^{\infty} exp(-(z-y))1_{z-y\geq 0}.exp(-y)\,dy \\
&= \int_0^{\infty} exp(-z+y))1_{y\leq z}exp(-y)\,dy \\
&= \int_0^{z} exp(-z+y))exp(-y)\,dy \\
&= exp(-z)\int_0^{z} dy \\
&= zexp(-z)
\end{aligned}
$$

پس تابع چگالی احتمال متغیر Z به شکل زیر خواهد بود:

$$f_Z(z) = \begin{cases} zexp(-z) & z \in R_Z \\ 0 & otherwise \end{cases}$$

حال برای پیدا کردن Zای که احتمال وقوع آن از بقیه‌ی حالت‌های ممکن بیشتر است، باید نقطه‌ی ماکسیموم در توزیع چگالی احتمالی Z را پیدا کنیم. بدین منظور باید از این توزیع مشتق گرفته و مشتق آن را برابر با صفر قرار دهیم.

$$
\begin{aligned}
f'_Z(z) &= exp(-z) - zexp(-z) \\
&= -(z-1)exp(-z)
\end{aligned}
$$

این مشتق در نقطه‌ی $z = 1$ برابر با صفر می‌شود. بنابراین در انتخاب تصادفی کوپن‌ها از درون جعبه احتمال مشاهده‌ی x و y ای که مجموعشان برابر با ۱ باشد بیشتر از مقادیر مجموعی دیگر است و بهتر است به دنبال جمع‌آوری هر چه بیشتر کوپن‌هایی با مجموع x و y ۱ باشیم.

سؤال ۵.

متغیر تصادفی X دارای تابع چگالی احتمال زیر است:

$$f_X(x) = \frac{\lambda}{2}e^{-\lambda|x|}$$

الف) تابع مولد گشتاور X را بیابید. (بازه‌ای را که این تابع متناهی می‌شود را مشخص کنید.)

ب) میانگین و واریانس X را با استفاده از تابع مولد گشتاور بیابید.

ج) U و Z متغیرهای تصادفی i.i.d با توزیع نمایی هستند. تابع مولد گشتاور متغیر تصادفی Y را به‌طوری‌که $Y = U - Z$ باشد، بدست آورید. سپس، با تحلیل این تابع، رابطه بین X و Y را مشخص کنید.

پاسخ .

الف)

$$M_X(t) = \int e^{tx}\tfrac{\lambda}{2}e^{-\lambda|x|}dx = \tfrac{\lambda}{2}(\int_0^{\infty} e^{(t-\lambda)x}dx + \int_{-\infty}^{0} e^{(t+\lambda)x}dx) = \tfrac{\lambda}{2}(-\tfrac{1}{t-\lambda} + \tfrac{1}{t+\lambda}) = \tfrac{\lambda^2}{\lambda^2-t^2}$$

طبق بازه بندی انتگرال اول و دوم داریم: $\lambda - t > 0$ و $\lambda + t > 0$

پس تابع مولد در بازه $\lambda > |t|$ متناهی خواهد بود.

ب)

$$E(X) = M'_X(0) = 0, Var(X) = M''_X(0) = \tfrac{2}{\lambda^2}$$

ج) $M_Y(t) = E[e^{t(U-V)}] = E[e^{tU}].E[e^{-tV}] = \frac{1}{1-t}.\frac{1}{1+t} = \frac{1}{1-t^2}$ تابع مولد متغیر تصادفی Y مشابه تابع مولد متغیر تصادفی X است به طوری که $\lambda = 1$. پس X، Y توزیع یکسانی دارند.

سؤال ۶.

اگر $X_1, X_2, ..., X_n$ متغیرهای تصادفی مستقل با توزیع پیوسته و نامعلوم F و $Y_1, Y_2, ..., Y_m$ متغیرهای تصادفی مستقل با توزیع پیوسته و نامعلوم G باشند، $n + m$ متغیر را مرتب کرده و تعریف می‌کنیم:

$$I_i = \begin{cases} 1, & \text{اگر } i \text{امین متغیر مرتب شده در } n+m \text{ متغیر از متغیر نوع } X \text{ باشد} \\ 0, & \text{در غیر این صورت} \end{cases} \tag{1}$$

متغیر تصادفی $R = \sum_{i=1}^{m+n} iI_i$ مجموع رتبه های X بوده و پایه و اساس یك روش استاندارد آماری برای آزمون یکسان بودن توزیع های F و G است. این آزمون فرض $F = G$ را وقتي كه R نه خیلی بزرگ و نه خیلی کوچك است می‌پذیرد. اگر فرض تساوی دو توزیع درست باشد، میانگین و واریانس R را محاسبه کنید.

پاسخ .

اگر فرض تساوی دو توزیع درست باشد هر كدام از $n+m$ متغیر تصادفي براي قرار گرفتن در مكان i داراي شانس مساوي با احتمال $\left(\frac{1}{m+n}\right)$ است. پس احتمال اینکه در مکان iام متغیر تصادفي X ظاهر شود، $\left(\frac{n}{m+n}\right)$ است.

$$E(R) = E\Big(\sum_{i=1}^{m+n} iI_i\Big) = \sum_{i=1}^{m+n} i\Big(\frac{n}{n+m}\Big) = \frac{n}{n+m}\Big(\frac{(n+m+1)(n+m)}{2}\Big)$$
$$= \frac{n(m+n+1)}{2}$$

$$Var(I_i) = \frac{n}{m+n}\Big(1 - \frac{n}{m+n}\Big) = \frac{mn}{(n+m)^2}$$

و براي $i \neq j$:

$$Cov(I_i, I_j) = \Big(\frac{n}{m+n}\Big)\Big(\frac{n-1}{m+n-1}\Big) - \Big(\frac{n}{m+n}\Big)^2 = \frac{-mn}{(m+n)^2(m+n-1)}$$

$$Var(R) = \sum_{i=1}^{m+n} i^2 \frac{mn}{(m+n)^2} + 2\sum_{i<j} \frac{-mn}{(m+n)^2(m+n-1)} ij$$
$$= \frac{mn(m+n+1)}{12}$$

سؤال ۷.

احتمالا با محصولات M&M آشنایی دارید. اسمارتیز آبی M&M در سال ۱۹۹۵ معرفی شد. پیش از آن، ترکیب رنگی اسمارتیزها در یک بسته به صورت زیر بود:

٪۳۰ قهوه‌ای، ٪۲۰ زرد، ٪۲۰ قرمز، ٪۱۰ سبز، ٪۱۰ نارنجی و ٪۱۰ نسکافه‌ای

پس از معرفی اسمارتیز آبی رنگ، ترکیب رنگ‌های داخل بسته به صورت زیر تغییر کرد:

٪۲۴ آبی، ٪۲۰ سبز، ٪۱۶ نارنجی، ٪۱۴ زرد، ٪۱۳ قرمز و ٪۱۳ قهوه‌ای

فرض کنید دوست شما یک بسته M&M مربوط به سال۱۹۹۴ و بسته‌ی دیگری مربوط به سال ۱۹۹۶ دارد، اما شما نمی‌دانید که هر بسته متعلق به چه سالی است. او از هر بسته یک اسمارتیز به شما می‌دهد که یکی سبز و دیگری زرد رنگ می‌باشد. احتمال آن‌که اسمارتیز زرد از درون بسته‌ی متعلق به سال ۱۹۹۴ بیرون آمده باشد را بدست آورید. (فرض کنید این‌که دوست شما بسته‌هایی با این قدمت را چگونه بدست آورده است تاثیری در حل شما ندارد!)

پاسخ .

دو فرضیه‌ی زیر را در نظر میگیریم:
A : بسته‌ی ۱ مربوط به سال ۱۹۹۴ و بسته‌ی ۲ مربوط به سال ۱۹۹۶ باشد.
B : بسته‌ی ۲ مربوط به سال ۱۹۹۴ و بسته‌ی ۱ مربوط به سال ۱۹۹۶ باشد.
و میدانیم که P(A) = P(B) = ۰.۵ . حال E را نیز اینگونه تعریف می‌کنیم:
E : زرد از بسته‌ی ۱ و سبز از بسته‌ی ۲ باشد.
حال دو احتمال شرطی زیر را با ضرب کردن احتمال‌های اسمارتیز M&M محاسبه می‌کنیم:

$$P(E|A) = (۰٫۲)(۰٫۲) = ۰٫۰۴$$
$$P(E|B) = (۰٫۱)(۰٫۱۴) = ۰٫۰۱۴$$

برای مثال، P(E|B) حاصل ضرب احتمال زرد بودن اسمارتیز در سال ۱۹۹۶ در سبز بودن اسمارتیز در سال ۱۹۹۴ است. حال با اضافه کردن احتمالات به قضیه‌ی بیز میتوان P(A|E) را محاسبه کرد.

$$P(A|E) = \frac{۴۰}{۵۴} \approx ۰٫۷۴$$