دانشکده‌ی مهندسی کامپیوتر

آمار و احتمال مهندسی	نیم‌سال دوم ۱۴۰۱-۱۴۰۰

تمرین سری ششم

مدرس: مهدی جعفری	موعد تحویل: ۲۷ خرداد ۱۴۰۱

سؤال ۱ با توجه به توزیع $Chi-squared$ درستی یا نادرستی عبارات زیر را مشخص نمائید و برای عبارات نادرست، توضیح مختصری ارائه کنید.

الف) با افزایش درجه آزادی، نمودار توزیع متقارن‌تر به نظر می‌رسد.

ب) انحراف معیار توزیع، دو برابر میانگین آن است.

ج) با درجه آزادی ۲۴، میانگین و میانه توزیع یکی خواهد بود.

د) با درجه آزادی ۲، توزیع نموداری شبیه به توزیع نمایی خواهد داشت.

حل:

الف) درست.

ب) نادرست. واریانس توزیع دو برابر میانگین آن است نه انحراف معیار.

ج) نادرست. میانه‌ی توزیع در این حالت برابر 23.33673 است ولی میانگین برابر 24 می‌باشد.

د) درست. اگر درجه‌ی آزادی برابر 2 باشد، تابع چگالی احتمال به صورت $\frac{1}{2}x^{-\frac{x}{2}}$ خواهد بود که همان تابع چگالی احتمال نمایی با پارامتر نرخ برابر $\frac{1}{2}$ است.

سؤال ۲ مسافری قصد سفر به یک کشور خارجی را دارد و می‌خواهد احتمال اینکه جرم چمدانش از ۳۲ کیلوگرم بیشتر شود را محاسبه کند. در صورتی که او ۳۰ بسته در چمدان داشته باشد و جرم هر بسته مستقل و از توزیعی یکنواخت بین ۱۰۰۰ تا ۱۱۰۰ گرم تبعیت کند، احتمال مورد نظر را محاسبه کنید. (احتمال مورد نظر را تا ۵ رقم اعشار محاسبه کنید.)

حل:

ابتدا میانگین و واریانس جرم بسته‌ها را محاسبه می‌کنیم:

$$E(X) = \frac{1000+1100}{2} = 1050, Var(X) = \frac{(1100-1000)^2}{12} = \frac{10000}{12}$$

حال طبق قضیه حد مرکزی خواهیم داشت:

$$P(\Sigma X_i > 32000) = P(\frac{\Sigma X_i - 30\times 1050}{\sqrt{30\times\frac{10000}{12}}} > \frac{32000-30\times 1050}{\sqrt{30\times\frac{10000}{12}}}) = P(Z > \sqrt{10}) = 1-\Phi(\sqrt{10}) = 0.00078$$

سؤال ۳ در یک منطقه‌ای ارتفاع رسیدن به آب از سطح زمین برای حفر چاه اندازه‌گیری شده است و برابر با ۸۰ متر است. با توجه به بارش‌های نامنظم و جنس خاک این منطقه، این ارتفاع ممکن است کم یا زیاد شود. در صورتی که X متغیری تصادفی و نشان‌دهنده تغییرات ارتفاع مورد نظر(رو به بالا(منفی) یا پایین(مثبت)) باشد و بدانیم $E(X) = -20$ باشد،به کمک نامساوی‌های احتمالاتی، حد بالایی برای $P(X \geq -10)$ ارائه کنید.

حل:

به وضوح باید از ناماسوی مارکو استفاده کنیم اما چون این نامساوی برای مقادیر مثبت تعریف می‌شود، تعریف می‌کنیم: $Y = X + 80$ بنابراین داریم:

$$E(Y) = E(X) + 80 = 60$$

$$P(X \geq -10) = P(Y \geq 70) \leq \frac{E(Y)}{70} = \frac{60}{70}$$

سؤال ۴ در صورتی که X از توزیع نمایی با میانگین $\frac{1}{\lambda}$ تبعیت کند، ثابت کنید:

$$P(X \geq \frac{0.2\lambda^2 + 1}{\lambda}) \leq \frac{25}{\lambda^4}$$

حل:

از نامساوی چبیشف استفاده می‌کنیم:

$$P(X \geq \frac{0.2\lambda^2 + 1}{\lambda}) = P(X - \frac{1}{\lambda} \geq \frac{\lambda}{5}) = P(|X - \frac{1}{\lambda}| \geq \frac{\lambda}{5}) \leq fracVar(X)(\frac{\lambda}{5})^2 = \frac{\frac{1}{\lambda^2}}{\frac{\lambda^2}{25}} = \frac{25}{\lambda^4}$$

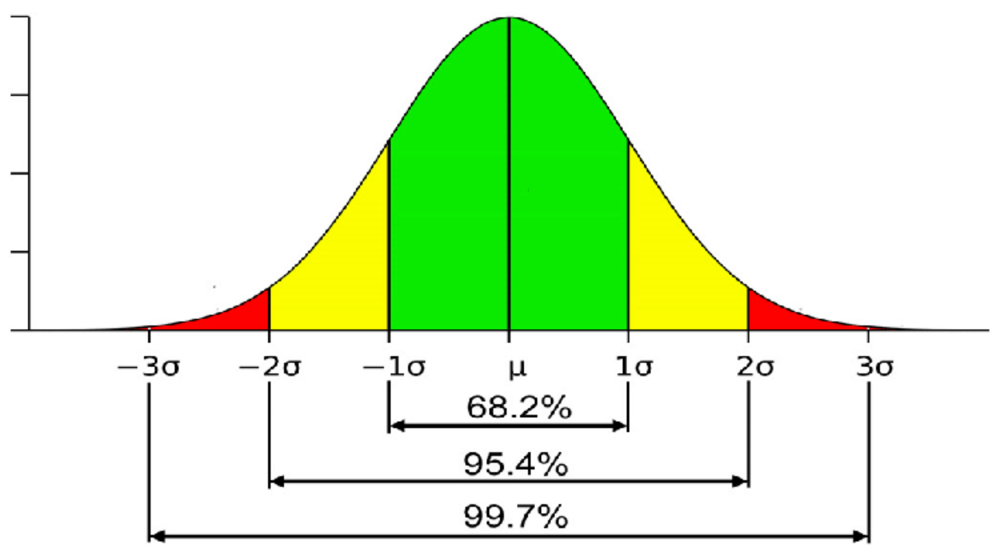
سؤال ۵ از آن‌جایی که به دوران شیرین امتحانات نزدیک می‌شویم و معمولا راه‌های جدیدی برای اتلاف وقت خود پیدا می‌کنیم، نشسته‌ایم و ۵۰۰ بار یک تاس سالم را پرتاپ کرده‌ایم. حال می‌خواهیم تعداد دفعاتی که تاس عدد ۴ آمده است را با دقت ۹۵ درصد تخمین بزنیم. بازه اطمینان مورد نظر را بیابید. (راهنمایی: از قانون ۶۸-۹۵-۹۹/۷ در توزیع نرمال استفاده کنید.)

حل:

می‌دانیم که این آزمایش از توزیع دو جمله‌ای پیروی می‌کند و داریم:

$$E(X) = np, Var(X) = npq$$

طبق قضیه حد مرکزی چون پرتاب‌ها مستقل و از یک توزیع یکسان بوده‌اند، میانگین آن‌ها به توزیع نرمال میل می‌کند. همچنین برای توزیع نرمال داریم:



بنابراین بازه اطمینان مورد نظر به این صورت محاسبه خواهد شد:

$$(np - 2 * \sqrt{npq}, np + 2 * \sqrt{npq}) = (\frac{500}{6} - 2 \times \sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}, \frac{500}{6} + 2 \times \sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}) \approx (67, 100)$$

سؤال ۶ دانشگاه برای یک مناسبت خاص می‌خواهد که به دانشجویان غذای رایگان بدهد و به همین منظور دو غذای چلو کباب سلطانی و ماکارونی را در نظر گرفته است. با فرض اینکه جمعا ۲۰۰۰ دانشجو در دانشگاه باشند و هر کدام به صورت کاملا تصادفی و مستقل غذای خود را در سلف انتخاب کنند، دانشگاه باید حداقل چه تعدادی از هر نوع غذا تهیه کند تا با احتمال حداقل ۹۸ درصد غذای مورد نظر برای دانشجویان وجود داشته باشد؟

حل:

محاسبات را برای یک نوع غذا انجام می‌دهیم چرا که برای دیگری نیز مشابه خواهد بود. متغیر تصادفی X_i را به این صورت تعریف می‌کنیم:

$$X_i = \begin{cases} 1 & kebab, p = 0.5 \\ 0 & pasta, p = 0.5 \end{cases}$$

حال X را تعداد دانشجویانی که چلو کباب سلطانی را انتخاب کرده‌اند تعریف می‌کنیم و داریم:

$$X = \Sigma X_i$$

$$E(X) = 2000 \times 0.5 = 1000, Var(X) = 2000 \times (0.5 - 0.25) = 500$$

حال مطلوب سوال را محاسبه می‌کنیم:

$$P(X < n) \geq 0.98 \Rightarrow P(\frac{X - 1000}{\sqrt{500}} < \frac{n - 1000}{\sqrt{500}}) \geq 0.98 \Rightarrow \Phi(\frac{n - 1000}{\sqrt{500}}) \geq 0.98 \Rightarrow \frac{n - 1000}{\sqrt{500}} \geq 2.05$$

$$\Rightarrow n \geq 1045.8$$

بنابراین از هر نوع غذا باید حداقل ۱۰۴۶ تا تهیه شود.

سؤال ۷ کارگاهی به دنبال تولید باتری‌های با کیفیت و طول عمر بالا است. آن‌ها برای اینکه بتوانند از کیفیت محصول خود مطمئن شوند ۱۰۰ باتری را به طور تصادفی انتخاب کرده و طول عمر آن‌ها را آزمایش می‌کنند. فرض کنید طول عمر باتری‌ها از توزیع نمایی پیروی میکند. اگر ۷۲ عدد از آن‌ها بیشتر از ۱۸۰ ساعت عمر کرده باشند با روش بیشینه‌ی درست نمایی مقدار λ را برای توزیع نمایی به دست بیاورید.

حل:

احتمال اینکه یک باتری بیشتر از ۱۸۰ ساعت عمر کند بر اساس توزیع نمایی برابر است با:

$$p = \int_{180}^{\infty} \lambda e^{-\lambda x} dx = e^{-180\lambda} => \lambda = -\frac{\ln p}{180}$$

از طرفی احتمال اینکه k باتری از n باتری بیشتر از ۱۸۰ ساعت عمر کنند برابر است با:

$$p^K (1-p)^{n-k}$$

$$=> L = k \ln p + (n - K) \ln (1 - p)$$

$$\frac{dL}{dp} = \frac{k}{p} - \frac{n-k}{1-p} = 0 => \hat{p} = \frac{k}{n}$$

پس نتیجه می‌گیریم:

$$\lambda = \frac{\ln \frac{72}{100}}{180} = 1.82 * 10^{-3}$$

سؤال ۸ تعدادی داده که از توزیع هندسی بر حسب p پیروی می‌کنند داده شده است. با توجه به اینکه پارامتر p خود از توزیع بتا با تابع چگالی احتمال $f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ پیروی می‌کند، به سوالات زیر پاسخ دهید:

الف) توزیع پسین را برای پارامتر p محاسبه کنید و بگویید چه ویژگی‌ای دارد.

ب) تخمینگر MAP را برای p محاسبه کنید.

حل:

الف)

$$P(p|X) = \frac{P(X|p)P(p)}{\int_p P(X|p)P(p)\,dp} = \frac{\prod_{i=1}^n (1-p)^{X_i-1} p \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}}{C(X)} =$$

$$\frac{(1-p)^{\sum_{i=1}^n X_i - n+\beta-1} p^{n+\alpha-1}}{C(X)B(\alpha,\beta)} \sim Beta(n+\alpha, \sum_{i=1}^n X_i - n + \beta)$$

همانطور که در روابط بالا قابل مشاهده است، توزیع‌های بتا و هندسی نسبت به یکدیگر $conjugate$ هستند. یا به عبارتی توزیع بتا برای پارامتر p در توزیع هندسی یک $prior-conjugate$ است.

ب) از عبارت محاسبه شده در قسمت قبل پس از حذف مقادیر غیرمرتبط با p ، لگاریتم می‌گیریم و سپس مشتق آن را نسبت به p محاسبه می‌کنیم و مساوی صفر قرار می‌دهیم:

$$\hat{p}_{MAP} = argmax(n+\alpha-1)\ln p + (\sum_{i=1}^n X_i - n + \beta - 1)\ln 1 - p$$

$$\frac{n+\alpha-1}{p} - \frac{\sum_{i=1}^n X_i - n + \beta - 1}{1-p} = 0 \Rightarrow \hat{p}_{MAP} = \frac{n+\alpha-1}{\sum_{i=1}^n X_i + \beta + \alpha - 2}$$

سؤال ۹ با توجه به N جفت داده مستقل (x_i, y_i) معادله خط عبوری از نقاط را به صورت $Y = \beta_0 + \beta_1 X + \epsilon$ تخمین زده‌ایم به صورتی که $\epsilon \sim N(0,1)$. با استفاده از کمترین خطای مربعات، ضرایب خواسته شده را محاسبه کنید.

حل:

$$MSE(Y, \hat{Y}) = \frac{1}{N}\sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2$$

برای یافتن هر پارامتر، نسبت به آن مشتق گرفته و برابر ۰ قرار می‌دهیم.

$$\beta_0 : \frac{1}{N}\sum_{i=1}^N 2\times(y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow \beta_0 = \frac{1}{N}\sum_{i=1}^N y_i - \beta_1 x_i$$

$$\beta_1 : \frac{1}{N}\sum_{i=1}^N 2x_i(y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow \sum_{i=1}^N x_i(y_i - \frac{1}{N}(\sum_{i=1}^N y_i - \beta_1 x_i) - \beta_1 x_i) = 0$$

$$\Rightarrow \sum_{i=1}^N x_i y_i - \beta_1 \sum_{i=1}^N x_i^2 - \frac{1}{N}(\sum_{i=1}^N x_i)(\sum_{i=1}^N y_i) + \frac{\beta_1}{N}(\sum_{i=1}^N x_i)^2 = 0$$

$$\Rightarrow \beta_1 = \frac{\frac{1}{N}(\sum_{i=1}^N x_i)(\sum_{i=1}^N y_i) - \sum_{i=1}^N x_i y_i}{\frac{1}{N}(\sum_{i=1}^N x_i)^2 - \sum_{i=1}^N x_i^2}.$$

$$\Rightarrow \beta_0 = \sum_{i=1}^N y_i - \frac{\frac{1}{N}(\sum_{i=1}^N x_i)(\sum_{i=1}^N y_i) - \sum_{i=1}^N x_i y_i}{\frac{1}{N}(\sum_{i=1}^N x_i)^2 - \sum_{i=1}^N x_i^2} x_i.$$

موفق باشید