به نام خدا

استنباط آماری

تمرین دوم

محمد ناصری

۸۱۰۱۰۰۴۸۶

بهار ۱۴۰۱

سوال اول

الف)

برای محاسبه همبستگی از معیار correlation که بر حسب کواریانس بدست می‌اید استفاده میکنیم.

$$X \sim N(\mu_1, \sigma_1), Y \sim N(\mu_2, \sigma_2)$$
$$Correlation(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sigma_1 \sigma_2} = \frac{E(XY) - \mu_1 \mu_2}{\sigma_1 \sigma_2}$$

ب)

برای اینکه یک متغیر تصادفی نسبت به خودش مستقل باشد (یعنی X نسبت به X مستقل باشد) با توجه به تعریف استقلال، رابطه زیر را خواهیم داشت:

$$P(X \in B) = P(X \in B, X \in B) = P(X \in B). P(X \in B) = [P(X \in B)]^2$$

تنها مقادیری که در رابطه بالا صدق میکند مقادیر $\{0, 1\}$ میباشند. از این رو شرط اینکه یک متغیر نسبت به خودش مستقل باشد در وهله اول Constant بودن مقدار احتمال و برابر بودن این مقدار با $\{0, 1\}$ است.

ج)

برای متغیر Z یک توزیع یونیفرم با مقدار (۱) در نظر میگیریم در این توزیع امید ریاضی یا همان میانگین برابر یک میشود. برای Z یک توزیع یونیفرم با مقدار ۰ در نظر میگیریم که در حداکثر دو حالت دچار ناپیوستگی بوده و مجموع این دو مقدار برابر ۵۰۰۰ است. آنگاه در این توزیع میانگین یا همان امید ریاضی برابر ۵۰ بوده و از طرفی مقادیر توزیع اول در حداقل 98% مواقع از توزیع دوم بیشتر است.

سوال دوم

الف)

در این حالت که k<N ارور سیستماتیک رخ نداده چون اگر رخ داده بود همه رای به محکوم بودن میدادند.با فرض اینکه G نشاندهنده گناهکار بودن و X تعداد قاضی‌هایی است که رای به محکوم بودن داده‌اند، داریم:

$$P(G|X=k)=\frac{P(X=k|G)P(G)}{P(X=k)}=\frac{c^k(1-c)^{n-k}*p}{c^k(1-c)^{n-k}p+(1-p)(1-v)^{n-k}v^k}$$

ب)

در این مورد ۲ حالت محتمل است: اول اینکه همه رای به محکومیت داده باشند که با F نشان میدهیم و دوم اینکه خطای سیستماتیک رخ داده باشد که با E نمایش میدهیم پس خواهیم داشت:

$$P(G|U)=\frac{P(U|G)P(G)}{P(U)}=\frac{P(U|G)p}{P(U|G)p+P(U|G^c)(1-p)}$$

با استفاده از قانون احتمال کل خواهیم داشت:

$$P(U|G)=P(U|G\cap E)P(E|G)+P(U|G\cap E^c)P(E^c|G)=c^n(1-s)+s$$
$$P(U|G^c)=P(U|G^c\cap E)P(E|G^c)+P(U|G^c\cap E^c)P(E^c|G^c)=w^n(1-s)+s$$

و در نهایت خواهیم داشت:

$$P(G|U)=\frac{P(U|G)P(G)}{P(U)}=\frac{P(U|G)p}{P(U|G)p+P(U|G^c)(1-p)}$$
$$=\frac{(c^n(1-s)+s)*p}{(c^n(1-s)+s)*p+(w^n(1-s)+s)(1-p)}$$

ج)

خیر، با افزایش مقدار n احتمال وقوع خطای سیستماتیک افزایش پیدا میکند و با وقوع این خطا نظرات قاضی‌ها هیچ اطلاعات مفیدی در بر ندارد و مشخص نمیکنند که متهم مجرم است یا نه.

سوال سوم

در تئوری احتمالات و آمار، **توزیع پواسون** یک توزیع احتمال گسسته است که احتمال وقوع تعداد معینی از رویدادها را در بازه زمانی یا مکانی ثابتی بیان می کند، در صورتی که این رویدادها با یک نرخ میانگین ثابت دیده شده و مستقل از زمان رویداد قبل رخ دهد.

در مثال ذکر شده ایمیل‌ها از یکدیگر مستقل هستند. یعنی احتمال رسیدن یک ایمیل بر احتمال رسیدن ایمیل دیگر تاثیری ندارد. همچنین یک بازه زمانی مشخص شده و نرخ ثابتی نیز در نظر گرفته شده فلذا میتوان از توزیه پواسون استفاده کرد.

برای محاسبه میانگین زمان صرف شده برای پاسخ به ایمیل ها از قانون Iterated Expectation استفاده میکنیم. این قانون به صورت زیر تعریف میشود:

$$E[X] = \sum_{i=1}^{n} E[X|A_i]P(A_i)$$

برای مساله مذکور خواهیم داشت:

$$E[spent\ time] = E[spent\ time \mid personal]P(personal) + E[spent\ time \mid work]P(work)$$

که با جاگذاری مقادیر بدست می‌آید:

$$E[spent\ time] = t\lambda * \mu_p * (1-p) + t\lambda * \mu_w * p$$

سوال چهارم

برای حل این مساله ابتدا W_i را بعنوان یک شاخص تعریف میکنیم که نشان‌دهنده این است که عنصر i ام از دنباله یک عنصر مینیمال محلی هست یا خیر (اگر مینیمال باشد مقدار ۱ میگیرد). حال خواهیم داشت:

$$E[N_{min}] = E[W_1 + W_2 + \cdots + W_n]$$

همینطور میدانیم:

$$E[W_1 + W_2 + \cdots + W_n] = E[W_1] + E[W_2] + \cdots E[W_n]$$

حال بر اساس symmetry داریم:

حالتهای کناری: $E[W_1] = E[W_n]$

حالتهای وسط: $E[W_j], j != 1,n$

از میان تمام حالتهای وسطی با در نظر گرفتن ۳تایی‌های کنار هم (برای چیدن هر ۳تایی !3 حالت داریم) فقط موارد زیر هستند که برای ما پاسخ هستند(بافرض مینیمم بودن b بین ۳ عدد): (a,b,c), (c,b,a) پس احتمال برای هر ۳ تایی برابر است با 2/3!=1/3 از این رو داریم:

$$E\left[W_j\right] = (n-2) * \frac{1}{3}$$

همینطور برای حالتهای کناری داریم که به احتمال ۱/۲ هر کدام میتوانند مینیمم محلی باشند پس در نهایت داریم:

$$E[N] = (n-2) * \frac{1}{3} + \frac{1}{2} + \frac{1}{2} = \frac{n+1}{3}$$

سوال پنجم

اینکه نفر آخر سرجای خود بنشیند منوط به این است که صندلی نفر اول زودتر از اینکه نفر آخر وارد شود پر شده باشد. از این رو میتوان دید دیگری نسبت به مساله داشت.

میتوان از این دید به این مساله نکاه کرد که هر تماشاگر هنگامی که به سراغ صندلی خود میرود اگر با صندلی پر مواجه شود فرد مورد نظر را از جا بلند کرده و آن فرد به صورت تصادفی به صندلی دیگری میرود. این بدین معنیست که در این دید به مساله، تماشاگر اول به طور مداوم جابجا میشود. در این حالت 148 تماشاگر دیگر با هر ترتیب سر جای خود قرار میگیرند و سوال به این تبدیل میشود که آیا تماشاگر اول سرجای خود مینشیند یا سرجای نفر ۱۵۰ام که این احتمال برابر ۱/۲ است.

سوال ششم

الف)

برای حل این سوال ابتدا حالت‌های عادی را در نظر میگیریم.

برای ۲ فرزند حالت‌های جنسیت عبارتند از ۴ حالت پپ و پد و دپ و دد

همچنین برای روزهای هفته نیز داریم: ش ی د س چ پ ج

حال با این فرضیات و داشته‌های ذکر شده برای حالات کلی خواهیم داشت:

$$Total\ Cases = 2 * 7 * 2 * 7$$

در حالتی که ما هیچ اطلاعاتی نداریم دختر یا پسر بودن از احتمال ۰.۵ برخوردار است ولی در حالتی که مساله بیان کرده خواهیم داشت:

$$Available\ Cases = (1 * 1 * 2 * 7) + (2 * 7 * 1 * 1) = 28 \ \rightarrow 28 - 1 = 27$$

منهای یک بخاطر حالت تکراری است که محاسبه میکنیم (هردو پسر و روز یکسان)

از این تعداد حالات، ۱۳ حالت شامل حالاتی است که فرزند دوم نیز پسر است:

$$Target\ Cases = (1 * 1 * 1 * 7) + (1 * 7 * 1 * 1) = 14 \ \rightarrow 14 - 1 = 13$$

پس اینکه بدانیم حداقل یکی از فرزندان پسر بوده و در ۳شنبه متولد شده بر احتمال پسر بودن فرزند دوم تاثیرگذار است.

ب)

در این حالت برخلاف بالا میدانیم دقیقا یکی از فرزندان پسر بوده و در ۳شنبه متولد شده است. در این حالت فرزند دیگر میتواند پسر یا دختر باشد ولی نمیتواند متولد ۳شنبه باشد. از این رو داریم:

$$Available\ Cases = (1 * 1 * 2 * 6) + (2 * 6 * 1 * 1) = 24$$

از این تعداد حالات، ۱۲ حالت مطلوب ماست:

$$Target\ Cases = (1 * 1 * 1 * 6) + (1 * 6 * 1 * 1) = 12$$

فلذا در این حالت احتمال به همان ۰.۵ میرسد.