

آمار و احتمالات مهندسی

تمرین هفتم - آمار

حسام و علیرضا

تاریخ تحویل ۱۴۰۱/۱۰/۲۲

سؤال ۱.

فرض کنید برنامه‌ای با $n = 100$ فایل کد داریم. بر اثر حمله‌ای که به سرور حاوی فایل‌های برنامه انجام شده، تعدادی باگ به هر فایل اضافه شده است. فرض کنید متغیر X_i تعداد باگ در فایل i ام است که از توزیع پواسون با میانگین ۱ پیروی می‌کند. همچنین تعداد باگ در هر فایل مستقل از فایل‌های دیگر است. احتمال اینکه تعداد کل باگ‌های اضافه شده به برنامه کمتر از ۹۰ باشد را تخمین بزنید.

سؤال ۲.

فرض کنید انتخابات شورای شهر پیش رو است که در آن ۲ کاندیدا با هم به رقابت خواهند پرداخت. یک شرکت نظرسنجی تصمیم گرفته با کمک شما نتیجه انتخابات را پیش‌بینی کند. این شرکت قصد دارد طوری نظرسنجی خود را انجام دهد که با ۹۶ درصد اطمینان، خطای پیش‌بینی کمتر از ۱ درصد باشد.

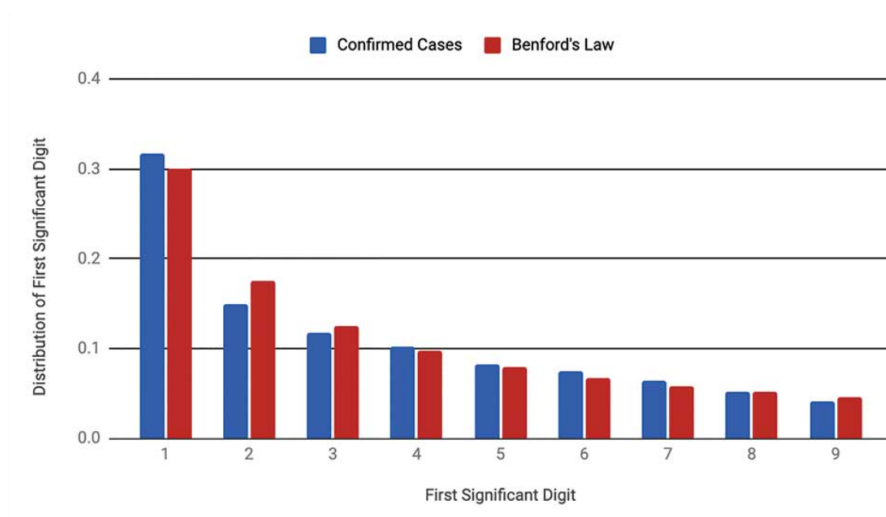
الف) حداقل از چند نفر باید نظرسنجی کنیم تا دقت خواسته شده توسط شرکت به دست آید؟

ب) اگر تعداد افراد واجد شرایط رای دادن در شهر کمتر از تعداد افرادی باشد که در قسمت الف به دست آمد، چه اتفاقی خواهد افتاد؟ این موضوع را تفسیر کنید.

سؤال ۳.

نمودار زیر، نشان‌دهنده توزیع پرارزش‌ترین رقم میزان مبتلایان روزانه به کووید ۱۹ در آمریکا از ژانویه تا اکتبر سال ۲۰۲۰ (به مدت ۲۹۰ روز) می‌باشد. مطابق نمودار، پرارزش‌ترین رقم تعداد افراد مبتلا شده در ماه‌های مختلف، از قانون بنفورد پیروی می‌کند! شما به عنوان متخصص آمار، با فرض صحیح بودن آمار ابتلای کووید در کشور آمریکا قصد دارید تا با استفاده از قانون بنفورد، صحت آمار دیتاست دیگری که به شما داده شده است را بدست آورید.

مطابق با قانون بنفورد، به احتمال 0.47 می‌بایست پرارزش‌ترین رقم تعداد مبتلایان در هر روز برابر با عدد ۱ یا ۲ باشد. یک دیتاست از میزان مبتلایان در جهان در یک بازه ۴۳ روزه به شما داده شده است. در دیتاست داده شده، در ۱۲ روز از ۴۳ روز، پرارزش‌ترین رقم تعداد مبتلایان برابر با ۱ یا ۲ می‌باشد. احتمال رخ دادن حالت ذکر شده برای پرارزش‌ترین رقم داده‌ها را در دیتاست داده شده بدست آورید، سپس با توجه به قانون بنفورد، صحت دیتاست داده شده را بررسی کنید. (راهنمایی: از توزیع دوجمله‌ای استفاده کنید. برای بررسی صحت دیتاست داده شده، کافی است تا درصد اختلاف حالت ذکر شده را با پیش‌بینی قانون بنفورد مقایسه کنید).



سؤال ۴.

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی n تایی از توزیعی با چگالی احتمال زیر باشد.

الف) برآورد گشتاوری پارامتر θ را به دست آورید. سپس واریانس این توزیع را برحسب پارامتر تخمین زده شده بنویسید.

ب) همچنین برای نمونه‌ی ۱۰ تایی به دست آمده از توزیع زیر، پارامتر به دست آمده توسط تخمین گر Maximum-Likelihood را محاسبه کنید.

$$f_X(x) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & o.w. \end{cases}$$

$$Sample = \{0.710, 0.211, 0.469, 0.367, 0.599, 0.951, 0.604, 0.991, 0.155, 0.215\}$$

سؤال ۵.

توزیع مقابل را در نظر بگیرید.

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|x-\mu|}{\sigma}}$$

با استفاده از روش گشتاور، امید ریاضی و واریانس توزیع فوق را بدست آورید.

سؤال ۶.

یک تحلیل گر مسابقات بوکس، قصد دارد تا احتمال برخورد موفق ضربات یک بوکسور را بسنجد. او ابتدا با بررسی مسابقات آن بوکسور، نمونه‌ای ۲۰۰ تایی از ضربات او در مسابقات مختلف را بدست می‌آورد. در این نمونه، ۱۳۰ ضربه از ۲۰۰ ضربه به صورت کامل با حریف برخورد کرده است و انحراف معیار نمونه برابر با $\sqrt{2}$ می‌باشد. (ضربات بوکسور نسبت به هم مستقل هستند) مربی او ادعا دارد که بوکسور در هر ۲۰۰ ضربه، حداقل ۱۵۰ ضربه را با موفقیت به حریف وارد می‌کند، درستی ادعای وی را بررسی کنید.

سؤال ۷.

علی در تابستان گذشته در زمینه بلاکچین مطالعاتی را انجام داده است و هم‌اکنون قصد دارد تا با استفاده از شبکه بیت کوین، مقداری پول را به حساب یکی از دوستانش بریزد. او هنگام ثبت تراکنش با هزینه‌ای به نام کارمزد تراکنش روبرو شده است. کارمزد تراکنش یک مقدار نسبی است که هر کاربر بابت ثبت شدن تراکنش خود می‌بایست پرداخت کند. اگر مقدار کارمزد تراکنش بسیار کم باشد، تراکنش ثبت نشده و اگر بیشتر از حد معمول باشد، با وجود ثبت شدن تراکنش کاربر پول اضافه‌ای را از دست می‌دهد و ضرر می‌کند. از آنجایی که کارمزد تراکنش یک مبلغ نسبی است، علی با جمع‌آوری آخرین کارمزدهای پرداخت شده در شبکه‌ی بیت کوین، سعی دارد تا هزینه تقریبی کارمزد مورد نیاز را تخمین بزند.

علی ۱۰۰ کارمزد را بررسی کرده و میانگین آن‌ها برابر با $1/08$ بدست آمد. انحراف معیار تمامی کارمزدهای شبکه طبق آمارها برابر با $2/3$ می‌باشد.

الف) بازه اطمینان ۹۵ و ۹۹ درصد را برای میزان کارمزد بهینه‌ای که علی باید پردازد را بدست آورید.

ب) عرض بازه‌های بدست آمده در قسمت الف را با یکدیگر مقایسه کنید و ارتباط میان میزان دقت و عرض بازه را مشخص کنید. شرح دهید که آیا همواره دقت بیشتر در بازه اطمینان برای ما مناسب است یا خیر.