۱. (۳ نمره) فرض کنید $Y_1, \dots, Y_n$ نمونه‌ای تصادفی از یک توزیع پواسون با پارامتر λ باشند. تنها اطلاعاتی که از Y_iها در اختیار داریم، صفر یا غیرصفر بودن آنهاست. یعنی نمونه $X_1, \dots, X_n$ را در اختیار داریم به طوری که:

$$X_i = \begin{cases} 1 & \text{if } \; Y_i = 0 \\ 0 & \text{if } \; Y_i \neq 0 \end{cases}$$

الف) یک تخمینگر ML برای پارامتر λ بر حسب X_iها پیدا کنید.

$$P(Y_i = 0) = e^{-\lambda}\, \lambda^0/0!$$

بنابراین X_iها توزیع برنولی با احتمال $p = e^{-\lambda}$ دارند:

$$f_{X_i}(x_i|\lambda) = \left(e^{-\lambda}\right)^{x_i}\left(1-e^{-\lambda}\right)^{1-x_i}$$

$$L(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n}\left(e^{-\lambda}\right)^{x_i}\left(1-e^{-\lambda}\right)^{1-x_i} = e^{-\lambda\sum_{i=1}^{n} x_i}\,\left(1-e^{-\lambda}\right)^{\sum_{i=1}^{n}(1-x_i)}$$

$$LL(\lambda) = -\lambda\left(\sum_{i=1}^{n} x_i\right) + \left(\sum_{i=1}^{n}(1-x_i)\right)\ln\left(1-e^{-\lambda}\right)$$

$$\frac{\partial LL(\lambda)}{\partial \lambda} = -\left(\sum_{i=1}^{n} x_i\right) + \left(\sum_{i=1}^{n}(1-x_i)\right)\left(\frac{e^{-\lambda}}{1-e^{-\lambda}}\right) = 0$$

$$-\left(1-e^{-\lambda}\right)\left(\sum_{i=1}^{n} x_i\right) + \left(\sum_{i=1}^{n}(1-x_i)\right)e^{-\lambda} = 0$$

$$\sum_{i=1}^{n}\left(-x_i + e^{-\lambda}x_i + e^{-\lambda} - e^{-\lambda}x_i\right) = 0 \to ne^{-\lambda} = \sum_{i=1}^{n} x_i$$

$$\Rightarrow \hat{\lambda}_{ML} = -\ln\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

ب) مقدار $E[Y_1|\sum_{i=1}^{n} Y_i]$ چقدر است؟

متغیر تصادفی $S = \sum_{i=1}^{n} Y_i$ را در نظر بگیرید. می‌دانیم: $E[S|S] = S$

از طرفی با توجه به یکسان بودن توزیع Y_iها: $E[Y_1|S] = E[Y_i|S]$ برای هر مقدار i.

$$E[S|S] = E[\textstyle\sum_{i=1}^{n} Y_i\,|S] = \displaystyle\sum_{i=1}^{n} E[Y_i|S] = n\,E[Y_1|S]$$

$$\to E[Y_1|S] = \frac{S}{n} \to \; E[Y_1|\textstyle\sum_{i=1}^{n} Y_i] = \dfrac{\sum_{i=1}^{n} Y_i}{n} = \bar{Y}$$

۲. کلاسی ۲۰ نفره در دانشکده فنی از ۵ دانشجوی برق، ۵ دانشجوی کامپیوتر، ۵ دانشجوی مکانیک، و ۵ دانشجوی صنایع تشکیل شده است. استاد این کلاس دانشجویان را به صورت تصادفی به چهار گروه ۵ نفره تقسیم می‌کند. فرض کنید N_i تعداد دانشجویان کامپیوتر در گروه i-ام باشد.

الف) $\mathrm{Cov}(N_1, N_2)$ را حساب کنید.

متغیر شاخص S_i برابر با 1 است اگر دانشجوی کامپیوتری i-ام در گروه اول قرار گیرد: $N_1 = \sum_{i=1}^{5} S_i$

متغیر شاخص T_i برابر با 1 است اگر دانشجوی کامپیوتری i-ام در گروه دوم قرار گیرد: $N_2 = \sum_{i=1}^{5} T_i$

$$E[N_1N_2] = E\left[\left(\sum_{i=1}^{5} S_i\right)\left(\sum_{j=1}^{5} T_j\right)\right] = \sum_{i=1}^{5}\sum_{j=1}^{5} E[S_iT_j] = \sum_{i=1}^{5}\sum_{j=1}^{5} \Pr(S_i = 1, T_j = 1)$$

از آنجا که یک دانشجو نمی‌تواند در دو گروه قرار بگیرد $\Pr(S_i = 1, T_j = 1)$ در حالت $i = j$ برابر صفر است. در حالت $i \neq j$ داریم:

$$\Pr\left(S_i = 1, T_j = 1\right) = P(S_i = 1)P\left(T_j = 1\middle|S_i = 1\right)$$

احتمال قرارگیری دانشجوی کامپیوتری i-ام در هر یک از چهار گروه یکسان است بنابراین: $P(S_i = 1) = 1/4$

اگر دانشجوی i-ام در گروه اول قرار داشته باشد، برای دانشجوی j-ام ۱۹ جای خالی باقی است (۴ جا در گروه اول و ۵ جا در هر یک از سه گروه دیگر). از این ۱۹ جا، ۵ تا متعلق به گروه دوم است، بنابراین: $P\left(T_j = 1\middle|S_i = 1\right) = 5/19$

از ۲۵ زوج (i,j)، ۲۰ زوج غیریکسان هستند بنابراین:

$$E[N_1N_2] = 20 \times \frac{1}{4} \times \frac{5}{19} = \frac{25}{19}$$

$$E[N_1] = \sum_{i=1}^{5} E[S_i] = \frac{5}{4} = E[N_2]$$

$$\rightarrow Cov(N_1, N_2) = \frac{25}{19} - \frac{25}{16}$$

ب) ضریب همبستگی بین $N_1 + N_2$ و $N_3 + N_4$ چقدر است؟

متغیر تصادفی $A = N_1 + N_2$ را تعریف می‌کنیم. واضح است که: $N_3 + N_4 = 5 - A$ بنابراین:

$$\rho(A, 5 - A) = \rho(A, -A) = -1$$

۳. فرض کنید شما در یک بازی شانسی شرکت کرده‌اید که در هر دست آن با احتمال p_1 برنده می‌شوید، با احتمال p_2 بازنده می‌شوید، و با احتمال p_3 بازی به تساوی می‌انجامد. در هر دست که برنده شوید یک دلار جایزه می‌گیرید، و در صورت باخت یک دلار جریمه می‌شوید. در حالت تساوی پولی جابجا نمی‌شود. فرض کنید متغیر تصادفی X نمایشگر میزان برد شما بعد از n دست شرکت در این بازی باشد.

الف) تابع مولد گشتاور X را محاسبه کنید.

اگر X_i میزان برد در دست i-ام باشد:

$$\Pr\{X_i = 1\} = p_1 \ , \Pr\{X_i = -1\} = p_2 \ , \Pr\{X_i = 0\} = p_3$$
$$p_1 + p_2 + p_3 = 1$$
$$\phi_{X_i}(s) = E[e^{sX_i}] = p_1e^s + p_2e^{-s} + p_3$$

از آنجا که $X = \sum_{i=1}^{n} X_i$ و X_iها مستقل از هم هستند:

$$\phi_X(s) = \left[\phi_{X_i}(s)\right]^n = (p_1e^s + p_2e^{-s} + p_3)^n$$

ب) میانگین و واریانس X را به کمک قسمت الف به دست آورید.

$$\phi'(s) = (p_1e^s - p_2e^{-s})n(p_1e^s + p_2e^{-s} + p_3)^{n-1} \rightarrow \ \phi'(0) = n(p_1 - p_2)$$
$$\phi''(s) = (p_1e^s + p_2e^{-s})n(p_1e^s + p_2e^{-s} + p_3)^{n-1} + (p_1e^s - p_2e^{-s})n(n - 1)(p_1e^s - p_2e^{-s})(p_1e^s + p_2e^{-s} + p_3)^{n-2}$$
$$\rightarrow \ \phi''(0) = n(p_1 + p_2) + n(n-1)(p_1 - p_2)^2$$
$$E[X] = \phi'(0) = n(p_1 - p_2)$$
$$Var(X) = E[X^2] - E[X]^2 = n(p_1 + p_2) + (n^2 - n)(p_1 - p_2)^2 - n^2(p_1 - p_2)^2$$

$$\rightarrow Var(X) = n(p_1 + p_2 - (p_1 - p_2)^2)$$

۴. (۳ نمره) متغیرهای تصادفی X و Y دارای تابع چگالی احتمال مشترک زیر هستند:

$$f_{XY}(x,y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

الف) تابع چگالی شرطی $f_Y(y|x)$ را به دست آورید.

$$f_Y(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{e^{-y}}{\int_x^{\infty} e^{-y}\,dy} = \frac{e^{-y}}{e^{-x}} = e^{x-y} \quad ,y > x$$

ب) امید ریاضی $E[Y|X]$ را محاسبه کنید.

$$E[Y|X=x] = \int_x^{\infty} y.e^{x-y}dy = e^x(-y.e^{-y} - e^{-y})|_x^{\infty} = e^x\big(0 - (-(x+1)e^{-x})\big) = x+1$$

$$\rightarrow E[Y|X] = X+1$$

پ) تابع چگالی احتمال متغیر تصادفی $Z = Y/X$ را به دست آورید.

از حل دستگاه زیر داریم:

$$z = y/x \ , \qquad w = x$$

$$x_1 = w \ ,y_1 = zw \ \rightarrow 0 < w < zw < \infty \rightarrow z > 1$$

$$J(x,y) = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} -y/x^2 & 1/x \\ 1 & 0 \end{vmatrix} = -\frac{1}{x}$$

$$\Rightarrow f_{ZW}(z,w) = \frac{f_{XY}(w\,,zw)}{|-1/w|} = we^{-zw}$$

$$f_Z(z) = \int_0^{\infty} we^{-zw}\,dw = -\frac{w}{z}e^{-zw} - \frac{1}{z^2}e^{-zw}\ |_0^{\infty} = 0 - \left(-\frac{1}{z^2}\right) = \frac{1}{z^2} \quad , \qquad z > 1$$

۵. (۳ نمره) یک برند سیگار در تبلیغات خود بیان می‌کند متوسط نیکوتین موجود در سیگارهای آن ۱/۵ میلی‌گرم است که در مقایسه با سایر برندها بسیار کمتر است. یک گروه پژوهشی مستقل با آزمایش ۱۰۰ نخ از سیگارهای این برند، می‌خواهد مشخص کند آیا ادعای این شرکت صحیح است یا مقدار نیکوتین سیگارهای آنها بیشتر از میزان ادعا شده است. با فرض این که انحراف معیار مقدار نیکوتین موجود در سیگارهای این برند ۰/۲ میلی‌گرم است، به سوالات زیر پاسخ دهید.

الف) اگر p-value در آزمون فرض اجرا شده توسط این گروه برابر با ۱/۶ درصد باشد، میانگین نمونه مورد استفاده آن‌ها چقدر بوده است؟

ب) یک بازه اطمینان ۹۹٪ برای میانگین نیکوتین موجود در سیگارهای این برند پیدا کنید و آن را تفسیر نمایید.

الف)

$$H_0: \mu = 1.5$$

$$H_A: \mu > 1.5$$

میانگین نمونه مورد استفاده = A

$$p - value = P(\bar{X} > A | H_0) = P\left(Z > \frac{A - 1.5}{\frac{0.2}{\sqrt{100}}} \right) = \Pr\big(Z > 50(A - 1.5)\big) = 0.016$$

$$P\big(Z < 50(A - 1.5)\big) = 0.984$$

طبق جدوا نرمال استاندارد نقطه مورد نظر برابر است با 2.15 در نتیجه:

$$50(A - 1.5) = 2.15 \rightarrow \; A = 1.5 + \frac{2.15}{50} = 1.543$$

ب)

$$\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

$$1 - \alpha = 0.99 \rightarrow 1 - \frac{\alpha}{2} = 0.995 \rightarrow \; z_{1-\frac{\alpha}{2}} = 2.58$$

$$1.543 - \frac{0.2}{10} \times 2.58 < \mu < 1.543 + \frac{0.2}{10} \times 2.58$$

$$1.4914 < \mu < 1.5946$$

میانگین واقعی نیکوتین موجود در سیگارهای این برند با احتمال ۹۹٪ داخل بازه (1.4914,1.5946)میلی‌گرم قرار دارد.