حل نمونه‌سوالات نهایی

آمار و احتمال مهندسی

دانشکده مهندسی کامپیوتر ـ بهمن‌ماه ۱۴۰۲

مدرس: امیر نجفی

بخش اول

میان‌ترم دوم بهار ۱۴۰۰

سوال ۱ میان‌ترم دوم (۲ نمره):

فرض کنید n نمونه مستقل از توزیع نمایی با پارامتر یقینی ولی نامعلوم $\lambda > 0$ در دسترس باشند. این نمونه‌ها را $X_1, \ldots, X_n$ می‌نامیم.

الف) تخمین بیشینه درست‌نمایی از پارامتر λ را بدست آورید.

ب) تخمین‌گر بیشینه درست‌نمایی (MLE) برای میانگین این توزیع کدام است؟

جواب:

الف) مطابق روال عادی در تخمین بیشینه درست‌نمایی عمل خواهیم کرد. لذا، تابع log-likelihood برای n نمونه را مطابق زیر تشکیل خواهیم داد:

$$\mathrm{LLE}\left(\lambda; X_{1:n}\right) = n \log \lambda - \lambda \sum_{i=1}^{n} X_i \quad (\lambda > 0)$$

واضح است به ازای $\lambda \to \infty$ یا $\lambda \to 0$ مقدار تابع LLH به سمت منفی بی‌نهایت میل کرده و درست‌نمایی داده‌ها به سمت صفر می‌رود. لذا حتما حداقل یک بیشینه محلی در این میان اتفاق خواهد افتاد. با مشتق‌گیری و برابر صفر قرار دادن آن مشخص می‌شود که تنها یک بیشینه‌کننده وجود دارد که همان جواب است.

$$\frac{\partial}{\partial \lambda}\mathrm{LLE}\left(\lambda; X_{1:n}\right) = \frac{n}{\lambda} - \sum_{i=1}^{n} X_i = 0$$

در نتیجه خواهیم داشت:

$$\hat{\lambda}_{\mathrm{ML}} = \frac{n}{X_1 + \ldots + X_n}$$

ب) ویژگی خاص تخمین‌گر MLE که در کلاس بدان اشاره شد: در صورتیکه تخمین ML از پارامترهای پایه یک مدل آماری (منظور در اینجا همان توزیع است) در دست باشد، تخمین هر پارامتر دیگری از مدل که بتوان آن را به صورت تابعی از پارامترهای پایه نوشت سر راست خواهد بود:

$$\hat{\mu}_{\mathrm{ML}} = 1/\hat{\lambda}_{\mathrm{ML}} = \bar{X}_n$$

سوال ۲ میان‌ترم دوم (۴ نمره):

یک توزیع پارامتریزه شده با PMF برابر با $P_X\left(x; \theta_1, \theta_2, \theta_3\right)$ را به شکل زیر تعریف می‌کنیم:

$$P_X\left(x; \theta_{1:3}\right) = \begin{cases} 1/3 & x = \theta_1 \\ 1/3 & x = \theta_2 \\ 1/3 & x = \theta_3 \\ 0 & \text{o.w.} \end{cases}$$

که بردار پارامترهای $\theta = \left(\theta_1, \theta_2, \theta_3\right)$ یقینی ولی نامعلوم است. همچنین حتماً داریم $i \neq j \Rightarrow \theta_i \neq \theta_j$. فرض کنید n نمونه i.i.d. از این توزیع به صورت $X_1, X_2, \ldots, X_n$ در دسترس باشند. می‌خواهیم میانگین این توزیع را تخمین بزنیم:

الف) فرض کنید از تخمین‌گر میانگین نمونه‌ای $\hat{\mu} = \hat{\mu}\left(X_{1:n}\right) = \bar{X}_n$ (population mean یا sample mean) برای این کار استفاده شود. آیا این تخمین‌گر بدون گرایش (unbiased) است؟

ب) میانگین مربعات خطا، یعنی $\mathbb{E}\left[\left(\bar{X}_n-\mu\right)^2\right]$ را برای تخمین‌گر قسمت الف) محاسبه کنید. در اینجا مقصود از μ میانگین واقعی است.

فرض کنید به جای تخمین‌گر میانگین نمونه‌ای، از یک تخمین‌گر جدید به نام $\hat{\mu}_{\text{New}}=\hat{\mu}_{\text{New}}\left(X_{1:n}\right)$ استفاده شود. نحوه محاسبه مقدار این تخمین‌گر به صورت زیر است: ابتدا تمام **مقادیر یکتای** موجود بین $X_1, X_2, \ldots, X_n$ را پیدا کرده و سپس فقط میان این مقادیر یکتا میانگین گرفته می‌شود.

ج) آیا این تخمین‌گر بدون گرایش است؟ در صورتیکه بدون گرایش است نشان دهید که مقدار bias صفر می‌شود. در غیر این صورت مثال نقض بیاورید.

د) مانند قسمت ب)، میانگین مربعات خطای این تخمین‌گر را برای n دلخواه محاسبه کنید. نرخ کاهش خطای آن بر حسب n را با تخمین‌گر میانگین نمونه‌ای مقایسه کنید.

ه) (**امتیازی**) تخمین بیشینه درست‌نمایی برای میانگین این توزیع (یعنی $\hat{\mu}_{\text{ML}}=\hat{\mu}_{\text{ML}}\left(X_1, \ldots, X_n\right)$) را بدست آورید.

جواب:

الف) تخمین‌گر میانگین نمونه‌ای (sample mean یا population mean) همواره بدون گرایش (unbiased) است.

ب) در صورتیکه واریانس متغیر تصادفی X را با σ^2 نشان دهیم، واریانس $\bar{X}_n$ همواره $\frac{\sigma^2}{n}$ خواهد بود (البته در صورت مستقل و هم توزیع بودن نمونه‌های $X_{1:n}$). از طرفی، به دلیل بدون گرایش بودن $\bar{X}_n$، میانگین خطای خواسته شده برابر با همان واریانس $\bar{X}_n$ است. در نتیجه، داریم:

$$\mathbb{E}\left[\left(\bar{X}_n-\mu\right)^2\right]=\frac{\sigma^2}{n}=\frac{1}{3n}\sum_{i=1}^{3}\left(\theta_i-\bar{\theta}\right)^2$$

که در اینجا، $\bar{\theta}=\left(\theta_1+\theta_2+\theta_3\right)/3$ است. واضح است که در این سوال مقدار μ یا میانگین توزیع برابر با $\bar{\theta}$ است.

ج) فرض کنید احتمال دیده شدن تنها مقدار θ_i در دنباله nتایی از نمونه‌های $X_{1:n}$ برابر با p_i باشد $(i=1,2,3)$. همچنین، فرض کنید احتمال دیده **نشدن** مقدار θ_i (ولی دیده شدن دو مقدار دیگر) در این دنباله نیز برابر با p_{-i} باشد. و در نهایت، احتمال دیده شدن هر سه مقدار یکتای $\theta_{1:3}$ در دنباله برابر با q فرض گردد. واضح است که داریم:

$$\sum_{i=1}^{3}p_i+\sum_{i=1}^{3}p_{-i}+q=1$$

از طرفی، بنا به تقارن داریم $p_i = p_j$ و $p_{-i} = p_{-j}$ برای تمامی $i, j \in \{1,2,3\}$.
لذا، مقدار گرایش (bias) به صورت زیر محاسبه می‌گردد:

$$\mathbb{E}\left(\hat{\mu}_{\text{new}} - \mu\right) = 0 \times q + \sum_{i=1}^{3} p_i \left(\theta_i - \bar{\theta}\right) + \sum_{i=1}^{3} p_{-i} \left(\frac{3\bar{\theta} - \theta_i}{2} - \bar{\theta}\right)$$

به دلیل تقارن‌های اشاره شده، به راحتی می‌توانید تحقیق کنید که مقدار عبارت بالا برابر با صفر بوده و لذا $\hat{\mu}_{\text{new}}$ بدون گرایش است.

د) مشابه با قسمت ب) و با استفاده از فرمول بدست آمده در قسمت ج) عمل میکنیم. می‌توان دید که میانگین مربعات خطا به صورت زیر نوشته خواهد شد:

$$\begin{aligned}
\mathbb{E}\left[\left(\hat{\mu}_{\text{new}} - \mu\right)^2\right] &= p_1 \sum_{i=1}^{3} \left(\theta_i - \bar{\theta}\right)^2 + p_{-1} \sum_{i=1}^{3} \left(\frac{3\bar{\theta} - \theta_i}{2} - \bar{\theta}\right)^2 \\
&= p_1 \sum_{i=1}^{3} \left(\theta_i - \bar{\theta}\right)^2 + \frac{p_{-1}}{4} \sum_{i=1}^{3} \left(\theta_i - \bar{\theta}\right)^2 \\
&= \left(p_1 + \frac{p_{-1}}{4}\right) \sum_{i=1}^{3} \left(\theta_i - \bar{\theta}\right)^2
\end{aligned}$$

در اینجا از تقارن $p_i = p_1$ و $p_{-i} = p_{-1}$ استفاده شده است.
تنها کار باقیمانده، محاسبه p_1 و p_{-1} است. با ابزارهای مقدماتی آمار و احتمال به راحتی می‌توان دید که برای $n \geq 2$ داریم:

$$p_1 = \left(\frac{1}{3}\right)^n \quad , \quad p_{-1} = \left(\frac{2}{3}\right)^n - 2\left(\frac{1}{3}\right)^n$$

که در مورد p_{-1} نحوه محاسبه به این صورت است که احتمال رشته‌ای را بسنجیم که شامل θ_1 نباشد، اما دو حالتی که رشته تماماً شامل θ_2 یا تماماً θ_3 باشد را دور میاندازیم.
لذا، نتیجه نهایی برای $n \geq 2$ بدین صورت خواهد شد:

$$\mathbb{E}\left[\left(\hat{\mu}_{\text{new}} - \mu\right)^2\right] = \left[\frac{1}{2}\left(\frac{1}{3}\right)^n + \frac{1}{4}\left(\frac{2}{3}\right)^n\right] \sum_{i=1}^{3} \left(\theta_i - \bar{\theta}\right)^2$$

برای حالت $n = 1$ نیز جواب واضح است و از نوشتن آن خودداری میکنیم. در صورتیکه این حالت در جواب برگه‌ها نوشته نشده بود نیز اشکالی ندارد.

نرخ کاهش خطای تخمین‌گر میانگین نمونه‌ای ب صورت $O\left(n^{-1}\right)$ (کاهش چندجمله‌ای) و نرخ کاهش خطای تخمین‌گر پیشنهادی $\hat{\mu}_{\text{new}}$ به صورت $O\left(e^{-n\log(3/2)}\right)$ (کاهش نمایی) است. این مثال نشان می‌دهد که میانگین نمونه‌ای همواره بهترین روش تخمین میانگین نیست.

ه) (**امتیازی**) تخمین‌گر بیشینه درست‌نمایی از بیشینه‌سازی تابع درست‌نمایی حاصل می‌گردد. برای تخمین MLE از میانگین، ابتدا لازم است که تخمین پارامترهای پایه $\theta_{1:3}$ را پیدا نموده و سپس در رابطه $\mu = \left(\theta_1+\theta_2+\theta_3\right)/3$ جایگذاری کنیم (خاصیت تخمین‌گرهای MLE). لذا، داریم:

$$\left(\hat{\theta}_1,\ldots,\hat{\theta}_3\right)_{\text{ML}} = \arg\max_{\theta_{1:3}} \prod_{i=1}^{n}\left(\frac{1}{3}\mathbf{1}_{X_i=\theta_1}+\frac{1}{3}\mathbf{1}_{X_i=\theta_2}+\frac{1}{3}\mathbf{1}_{X_i=\theta_3}\right)$$

البته شرط $\theta_1 \neq \theta_2 \neq \theta_3$ نیز لازم است در حین بهینه‌سازی لحاظ شود. چرا که از ویژگی‌های مدل آماری است. بیشینه‌سازی تابع فوق سرراست و از طریق مشتق‌گیری قابل انجام نیست. لذا می‌بایست استدلال نمود:

۱) در صورتیکه حداقل سه مقدار یکتا در میان نمونه‌های $X_{1:n}$ وجود داشته باشد، می‌توان $\theta_{1:3}$ را (به ترتیب دلخواه) برابر با هر سه‌تایی دلخواهی از آن مقادیر یکتا در نظر گرفت. در این صورت مقدار بیشینه درست‌نمایی همواره برابر با 3^{-n} و در غیر این صورت همواره برابر با صفر می‌گردد.
البته توجه کنید که در صورتیکه $X_{1:n}$ واقعا بر مبنای مدل آماری فوق تولید شده باشد نمی‌توان بیش از سه مقدار یکتا در میان آنان دید. همچنین، در این حالت جواب بیشینه درست‌نمایی (بدون احتساب جایگشت) یکتا خواهد بود.
در این صورت تخمین‌گر $\hat{\mu}_{\text{ML}}$ با تخمین‌گر بخش د) یکی می‌شود.

۲) در صورتیکه کمتر از سه مقدار یکتا در میان نمونه‌ها بود، مابقی θ_iها مقدار دلخواه می‌گیرند. در این حالت نیز، کماکان مقدار درست‌نمایی برابر با 3^{-n} و در غیر این صورت صفر خواهد شد. در این حالت، جواب بیشینه درست‌نمایی مقدار یکتا (حتی بدون احتساب جایگشت) نداشته و تخمین MLE از میانگین خوش‌تعریف نیست.

در نتیجه دقت کنید که تخمین‌گر MLE در این حالت همواره با تخمین‌گر $\hat{\mu}_{\text{new}}$ یکی **نیست**.

سوال ۴ میان‌ترم دوم (۴ نمره):

متغیرهای تصادفی $X_1, \ldots, X_n$ به صورت i.i.d. از چگالی احتمال $f_X(x)$ (و متناظراً توزیع انباشته احتمال $F_X(x)$) حاصل شده‌اند. فرض کنید که از این متغیرهای تصادفی نمونه‌گیری شده و آنان را بر حسب مقدارشان به صورت صعودی مرتب کرده باشیم.
مقدار **یکی مانده به بزرگترین** در این دنباله به دلیلی برای ما مهم است. توزیع آماری آن را بدست آورید.

جواب:

راه‌حل کاملا مشابه با روش اتخاذ شده برای محاسبه توزیع بیشینه در اسلایدهای درس (و همچنین کتاب) است. از میان n متغیر تصادفی نمونه‌گیری شده، ابتدا لازم است ببینیم که به چند طریق می‌توان آنان را به سه دسته ۱، ۱ و $n-2$ تایی تقسیم بندی نمود. اولی نماینده بزرگ‌ترین مقدار، دومی یکی مانده به بزرگ‌ترین (مثلاً با Y نشانش می‌دهیم) و مابقی سایرین را تشکیل خواهند داد. جواب مطابق با زیر

$$\frac{n!}{1!1!(n-2)!} = n\left(n-1\right)$$

برابر با $n\left(n-1\right)$ است.

حال لازم است اینگونه استدلال نمود: برای یک طول بی‌نهایت کوچک $h \to 0$ و y دلخواه خواهیم داشت:

$$\mathbb{P}\left(y \le Y \le y+h\right) \simeq n\left(n-1\right)\Big(1-F_X\left(y+h\right)\Big)\,\mathbb{P}\left(y \le X \le y+h\right)F_X^{n-2}\left(y\right)$$

در نتیجه، با تقسیم طرفین بر h و گرفتن حد به رابطه زیر خواهیم رسید:

$$f_Y\left(y\right) = n\left(n-1\right)f_X\left(y\right)F_X^{n-2}\left(y\right)\Big(1-F_X\left(y\right)\Big)$$

سوال ۶ میان‌ترم دوم (۳ نمره):

متغیرهای تصادفی X, Y با توزیع مشترک دانسته شده‌ای مفروض هستند. فرض کنید که قصد داریم متغیر تصادفی Y را به صورت یک ترکیب همگن درجه دو به صورت $\hat{Y} = aX^2 + bX$ از روی X تخمین بزنیم، به طوری که میانگین مربعات خطا کمینه شود.
الف) ضرایب مجهول a, b را بر حسب گشتاورهای مجزا و یا مشترک X, Y (یعنی $\mathbb{E}\left(X^nY^m\right),\ \ n,m \ge 0$) بدست آورید.

ب) در صورتیکه بخواهیم تخمین ناهمگن درجه دوم به شکل $\hat{Y} = aX^2 + bX + c$ داشته باشیم، ضرایب مجهول را دوباره محاسبه کنید.

جواب:

مشابه با آن‌چه در مورد تخمین MSE می‌دانیم عمل می‌کنیم.

الف) قرار است عبارت زیر بر حسب a, b کمینه گردد:

$$\mathbb{E}\left[\left(Y-\hat{Y}\right)^2\right] = \mathbb{E}\left[\left(Y-aX^2-bX\right)^2\right]$$

پس از مشتق‌گیری به دستگاه دو معادله و دو مجهول زیر می‌رسیم:

$$\frac{\partial}{\partial a}\mathbb{E}\left[\left(Y-aX^2-bX\right)^2\right] = -2\mathbb{E}\left[X^2\left(aX^2+bX-Y\right)\right] = 0$$

$$\frac{\partial}{\partial b}\mathbb{E}\left[\left(Y-aX^2-bX\right)^2\right] = -2\mathbb{E}\left[X\left(aX^2+bX-Y\right)\right] = 0$$

که به صورت ماتریسی به فرم زیر نوشته شده و حل می‌گردد:

$$\begin{bmatrix}\mu_{4,0} & \mu_{3,0}\\ \mu_{3,0} & \mu_{2,0}\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix} = \begin{bmatrix}\mu_{2,1}\\ \mu_{1,1}\end{bmatrix} \Rightarrow \begin{bmatrix}a\\ b\end{bmatrix} = \frac{1}{\mu_{4,0}\mu_{2,0}-\mu_{3,0}^2}\begin{bmatrix}\mu_{2,0} & -\mu_{3,0}\\ -\mu_{3,0} & \mu_{4,0}\end{bmatrix}\begin{bmatrix}\mu_{2,1}\\ \mu_{1,1}\end{bmatrix}$$

مقصود از $\mu_{m,n} = \mathbb{E}\left(X^nY^m\right)$ است. به قرار ساده‌تر، داریم:

$$a = \frac{\mu_{2,0}\mu_{2,1}-\mu_{3,0}\mu_{1,1}}{\mu_{4,0}\mu_{2,0}-\mu_{3,0}^2} \quad , \quad b = \frac{\mu_{4,0}\mu_{1,1}-\mu_{3,0}\mu_{2,1}}{\mu_{4,0}\mu_{2,0}-\mu_{3,0}^2}$$

ب) در این حالت یک معادله و یک متغیر c اضافه می‌شوند. اما متغیر اضافی همان ابتدای کار حل خواهد شد. می‌توان به راحتی دید که مقدار بهینه c بر حسب سایر متغیرها و گشتاورهای X, Y به صورت $c = \mathbb{E}Y - a\mathbb{E}X^2 - b\mathbb{E}X$ خواهد بود. با جایگذاری این مقدار در مسئله اولیه در حالت الف) خواهیم داشت:

$$\mathbb{E}\left[\left(\hat{Y}-Y\right)^2\right] = \mathbb{E}\left[\left(Y_c-aZ-bX_c\right)^2\right]$$

که در اینجا مقصود از $Z = X^2 - \mathbb{E}X^2$ است. همچنین مقصود از A_c نسخه مرکزی شده (centered) از متغیر تصادفی A است. یعنی $A_c = A - \mathbb{E}A$.

از اینجا به بعد مطابق با حالت الف) عمل می‌کنیم:

$$\frac{\partial}{\partial a}\mathbb{E}\left[\left(Y_c - aZ - bX_c\right)^2\right] = -2\mathbb{E}\left[Z\left(aZ + bX_c - Y_c\right)\right] = 0$$
$$\frac{\partial}{\partial b}\mathbb{E}\left[\left(Y_c - aZ - bX_c\right)^2\right] = -2\mathbb{E}\left[X_c\left(aZ + bX_c - Y_c\right)\right] = 0$$

جواب دستگاه بالا مشابه با جواب آخر قسمت الف) است، تنها لازم است که در فرمول‌های نهایی جایگذاری‌ها (replacement) زیر را انجام دهید.

$$\mu_{4,0} \to \mathbb{E}Z^2 = \mathbb{E}\left[\left(X^2 - \mathbb{E}X^2\right)^2\right]$$
$$\mu_{3,0} \to \mathbb{E}\left(ZX_c\right) = \mathbb{E}\left((X - \mathbb{E}X)\left(X^2 - \mathbb{E}X^2\right)\right)$$
$$\mu_{2,0} \to \mathbb{E}\left(X_c^2\right) = \mathrm{Var}(X)$$

$$\mu_{2,1} \to \mathbb{E}\left(ZY_c\right) = \mathbb{E}\left[\left(X^2 - \mathbb{E}X^2\right)(Y - \mathbb{E}Y)\right]$$
$$\mu_{1,1} \to \mathbb{E}\left(X_cY_c\right) = \mathrm{Cov}(X, Y)$$

بخش دوم
سوالات پایان‌ترم بهار ۱۴۰۰

سوال ۲ امتحان نهایی (۴ نمره):

یک فرستنده مخابراتی قصد ارسال n بیت اطلاعات به یک گیرنده را دارد. داده‌ها از طریق یک کانال نویزی فرستاده شده و هر بیتی که توسط فرستنده ارسال می‌گردد با احتمال $p = 1/2$ و مستقل از ارسال‌های قبلی دچار خطا می‌شود. به منظور اطمینان از صحت ارسال‌ها، در سمت فرستنده از یک کدگذار (coder) بهره می‌بریم. بدین شکل که به جای ارسال هر کدام از n بیت فوق، تعداد L بیت که به طریقی خاص انتخاب شده‌اند، به نمایندگی از آن ارسال می‌گردند. لذا، درنهایت به جای n بیت اولیه، تعداد کل nL بیت ارسال خواهند شد.
این کار قرار است با چسباندن اطلاعات زائد (Redundancy) تعداد ارسال‌ها را افزایش داده، اما در عوض احتمال بروز خطا در آنان را کاهش دهد. خاصیت هر بلوک L بیتی که نماینده یکی از n بیت اولیه می‌باشد این است که تنها در صورتی بیت اصلی در سمت گیرنده دچار خطا خواهد شد که تمامی L بیتی که آن را نمایندگی می‌کنند در کانال خراب شوند. و حتی اگر صرفاً یکی از آنان سالم به مقصد برسد، بیت اصلی به درستی decode می‌گردد. در صورتیکه علاقه داشتید بدانید این کار چگونه امکان‌پذیر است، درس «تئوری اطلاعات» را در ترم‌های آینده بگیرید.
در این سوال قصد داریم تا احتمال بروز حداقل یک خطا در ارسال n بیت اصلی را بدست بیاوریم. در آخر نشان خواهیم داد که در صورت $O\left(n \log_2 n\right)$ بار استفاده از کانال به جای n بار، احتمال انتقال کاملاً صحیح کل داده‌ها برای $n \gg 1$ به سمت ۱ میل خواهد نمود.

الف) فرض کنید که اصلاً از کدگذاری استفاده نمی‌شد و در ارسال هر یک از n بیت اصلی، صرفاً همان بیت و به همان شکل اصلی خود ارسال می‌گشت (به عبارتی، داشتیم $L = 1$). احتمال اینکه از میان n بیت اصلی، حداقل یکی دچار خطا شود چقدر است؟
(به اختصار استدلال کنید که با افزایش n این احتمال به سمت یک میل خواهد کرد.)

ب) در صورت کدگذاری با طول بلوک‌های کد L، احتمال اینکه حداقل یکی از n بیت اصلی با خطا در گیرنده decode شود را حساب کنید.

ج) نشان دهید در صورتیکه طول بلوک‌های کد را به صورت $L = (1+\varepsilon)\log_2 n$ انتخاب کنیم (به ازای هر $\varepsilon > 0$)، احتمال بروز خطا با افزایش n به سمت صفر میل خواهد کرد.

جواب:

الف) احتمال بروز حداقل یک خطا در ارسال برابر یک منهای احتمال سالم رسیدن همگی بیت‌هاست. از آنجا که ارسال‌ها مستقل از یکدیگر و هرکدام تنها با احتمال ۱/۲ سالم می‌رسند، پس داریم:

$$\mathbb{P}(\text{Error}) = 1 - \frac{1}{2^n}$$

که به وضوح با افزایش n به صورت نمایی به ۱ نزدیک می‌شود. یعنی برای n های بزرگ احتمالاً حداقل یک خطا خواهیم داشت.

ب) جواب مشابه با بخش قبل است، با این تفاوت که این بار احتمال وقوع خطا دیگر ۱/۲ نیست. بلکه خطا در ارسال هر بیت زمانی رخ می‌دهد که تمامی L بیت نماینده آن دچار خطا شوند. پس احتمال سالم رسیدن هر بیت، مستقل از سایرین، برابر با $1 - 2^{-L}$ است که مشابه با استدلال بخش الف) و با توجه به استقلال در ارسال‌ها بدست آمده است. لذا، رابطه نهایی خواسته شده در بخش ب) برابر است با:

$$\mathbb{P}(\text{Error}) = 1 - \left(1 - \frac{1}{2^L}\right)^n$$

ج) صرفا می‌بایست $L = (1+\varepsilon)\log_2 n$ را در رابطه بخش دوم جایگذاری کنیم:

$$\lim_{n\to\infty} \mathbb{P}(\text{Error}) = \lim_{n\to\infty} 1 - \left(1 - \frac{1}{2^L}\right)^n = \lim_{n\to\infty} 1 - e^{-n2^{-L}} = \lim_{n\to\infty} 1 - e^{-\frac{n}{n^{1+\varepsilon}}} = \lim_{n\to\infty} 1 - e^{-n^{-\varepsilon}} = 0$$

که به ازای هر $\varepsilon > 0$، احتمال حتی یک ارسال خطادار با افزایش n به سمت صفر میل خواهد کرد.

سوال ۳ امتحان نهایی (۴ نمره):

یک سکه تصادفی با احتمال شیر یا خط نامعلوم را n بار به صورت مستقل پرتاب کرده و مشاهده می‌کنیم که k بار شیر ظاهر می‌شود.

الف) تخمین MLE از احتمال شیر آمدن سکه را محاسبه کنید.

حال با سازنده سکه ملاقات داشته و اطلاعاتی در مورد سکه کسب می‌کنیم. سازنده سکه می‌گوید که سکه‌های ساخت او به یک سمت بایاس دارند. وی معتقد است که احتمال شیر آمدن سکه‌هایش با احتمال α بین صفر تا ۱/۲ (یعنی در بازه

$[0,0.5]$)، و با احتمال $1-\alpha$ بین ۱/۲ تا ۱ (یعنی بازه $[0.5,1]$)) است. همچنین، در بازه صفر تا ۱/۲ میان مقادیر برتری نسبت به یکدیگر وجود ندارد. در بازه ۱/۲ تا ۱ نیز به همین شکل، برتری بین مقادیر نیست.
(بدون کاستن از کلیت مسئله، فرض کنید که $\alpha > 0.5$)

ب) با استفاده از این اطلاعات پیشین، تخمین MAP از احتمال شیر آمدن سکه را دوباره محاسبه کنید.

جواب:

الف) توزیع تعداد شیرها به شرط دانستن احتمال آنان p یک توزیع دوجمله‌ای است. با توجه به این واقعیت، ابتدا تابع درست‌نمایی را تشکیل داده و سپس آن را بیشینه می‌کنیم:

$$\mathbb{P}\left(k\,|\,p\right) = \binom{n}{k} p^k \left(1-p\right)^{n-k}, \quad \log \mathbb{P}\left(k\,|\,p\right) = k \log p + \left(n-k\right) \log\left(1-p\right) + \text{const}$$

برای بیشینه‌سازی، ابتدا لازم است ببینیم که به ازای $p \to 0$ و $p \to 1$ مقدار درست‌نمایی به سمت صفر میل میکند. لذا در این بین می‌بایست یک بیشینه وجود داشته باشد. مقدار بیشینه کننده p^* را با مشتق‌گیری از تابع log-likelihood بدست می‌آوریم:

$$\frac{\partial}{\partial p} LL\left(p;k,n\right) = \frac{k}{p} - \frac{n-k}{1-p} = 0 \quad \Rightarrow \quad p^* = \frac{k}{n}$$

ب) به منظور یافتن تخمین بیشینه توزیع پسین (MAP) ابتدا لازم است که یک توزیع پیشین بر حسب اطلاعات داده شده مدلسازی شود. بر مبنای اطلاعات داده شده، تابع توزیع پیشین می‌بایست از دو بخش یکنواخت تشکیل شده باشد. بخش اول بین صفر و نیم (با احتمال مجموع α و لذا چگالی احتمال 2α) و بخش دوم بین نیم و ۱ با چگالی احتمال $2\left(1-\alpha\right)$. یعنی:

$$f_p\left(p\right) = \begin{cases} 2\alpha & p \le 1/2 \\ 2\left(1-\alpha\right) & p > 1/2 \end{cases}$$

حال بیشینه توزیع پسین که مطابق با زیر است را محاسبه می‌کنیم:

$$f\left(p\,|\,k\right) \propto \mathbb{P}\left(k\,|\,p,n\right) f_p\left(p\right) = p^k \left(1-p\right)^{n-k} f_p\left(p\right)$$

به دلیل خوشرفتار نبودن توزیع پیشین، بیشینه‌سازی را نمی‌توان با مشتق‌گیری انجام داد. از طرفی، از حل قسمت الف) می‌دانیم که بیشینه بخش درست‌نمایی از توزیع پسین در $p = k/n$ رخ می‌دهد. برای بیشینه‌سازی لازم است حالت‌بندی کنیم:

- (حالت اول) فرض کنید که داشته باشیم $k/n \le 1/2$. در این صورت، تخمین MAP همواره همان تخمین MLE یعنی k/n می‌شود. چرا که مقدار $p^* = k/n$ هم بخش درست‌نمایی $\mathbb{P}\left(k\,|\,p,n\right)$ را بیشینه می‌سازد، و هم بخش توزیع پیشین $f_p\left(p\right)$ (دقت کنید که فرض کرده بودیم $\alpha > 1-\alpha$).

- (حالت دوم) فرض کنید که داشته باشیم $k/n > 1/2$. در مورد تابع درست‌نمایی می‌دانیم که مقدار آن با افزایش p از صفر به سمت k/n افزایشی است و همچنین داریم $\mathbb{P}\left(p = k/n\,|\,k\right) > \mathbb{P}\left(p = 1/2\,|\,k\right)$. اما در نقطه

$p = 1/2$، مقدار توزیع پیشین یک افت ناگهانی داشته و مقدار آن $(1 - \alpha)/\alpha$ برابر می‌شود و سپس ثابت می‌ماند. در این صورت لازم است که چک شود که شرط زیر برقرار است یا نه؟

$$(1-\alpha)\left(\frac{k}{n}\right)^{k}\left(1-\frac{k}{n}\right)^{n-k} > \alpha\left(\frac{1}{2}\right)^{k}\left(1-\frac{1}{2}\right)^{n-k} \quad \text{or} \quad \left(\frac{k}{n}\right)^{k}\left(1-\frac{k}{n}\right)^{n-k} > \left(\frac{\alpha}{1-\alpha}\right)\frac{1}{2^{n}}$$

در صورتیکه برقرار بود، تخمین احتمال شیر آمدن کماکان همان k/n می‌شود چرا که شواهد درست‌نمایی توانسته بر اطلاعات پیشین فائق بیاید. در غیر این صورت، تخمین MAP برابر با $p^* = 1/2$ می‌شود.
به صورت خلاصه، می‌توان نوشت:

$$p^*_{\text{MAP}} = \begin{cases} \frac{k}{n} & k/n < 1/2 \quad \text{or} \quad \left(\frac{k}{n}\right)^{k}\left(1-\frac{k}{n}\right)^{n-k} > \left(\frac{\alpha}{1-\alpha}\right)\frac{1}{2^{n}} \\ \frac{1}{2} & o.w. \end{cases}$$

سوال ۴ امتحان نهایی (۴ نمره):

یک سازمان دولتی از شاخص عملکردی خود راضی نیست، و قصد دارد آن را بهبود ببخشد. شاخص عملکردی کل سازمان برابر با میانگین آماری شاخص عملکردی یکایک نیروهایش است، که می‌توان آنان را متغیرهای تصادفی هم‌توزیع ولی مستقل در نظر گرفت. سازمان یک بازرس انتخاب کرده تا نحوه انجام کار نیروهایش را بررسی کرده و شاخص عملکردی تعدادی از آنان را محاسبه کند.
بازرس تعداد ۱۰۰ نیرو را به صورت تصادفی انتخاب کرده و پس از بررسی سوابق کاری یکساله آنان شاخص‌های عملکردیشان را محاسبه نموده است. وی فرض کرده است که توزیع شاخص فوق برای هر یک از نیروها یک متغیر تصادفی گاوسی است. میانگین شاخص‌های عملکردی برای این جمعیت برابر با ۵۴٪ شده. همچنین بازرس انحراف معیار ۵٪ را کران بالایی برای انحراف معیار هر یک از ۱۰۰ اندازه‌گیری خود دانسته است.
سازمان به منظور بهبود عملکرد اخیراً تعدادی از مدیران اجرایی خود را تغییر داده. مدتی پس از این اتفاق، بازرس دوباره مشغول به کار شده و این بار ۱۰ نیرو را به صورت تصادفی انتخاب و میانگین شاخص عملکردی این جمعیت نمونه را محاسبه کرده است. عدد بدست آمده این بار به ۵۷٪ افزایش پیدا کرده، و بازرس به دلیل دقت پایین‌تر در اثر کمبود وقت، این دفعه انحراف معیار ۸٪ را کران بالایی برای انحراف معیار اندازه‌گیری‌هایش گزارش کرده. (دقت کنید که کران‌های بالای انحراف معیارها، یعنی ۵٪ و ۸٪، تخمین‌های شخصی خود بازرس هستند و مستقل از دادها مقداردهی شده‌اند. همچنین وی کماکان فرض کرده که توزیع شاخص‌ها گاوسی است).

حال، سازمان شما را به عنوان یک آماردان استخدام می‌کند و از شما می‌پرسد که آیا افزایش ۳ درصدی در شاخص عملکردی پس از تغییر مدیران اجرایی به لحاظ آماری «معنی‌دار» هست یا خیر؟
الف) فرضیه‌های H_0 و H_1 را در این حالت تعریف کرده و یک آماره مناسب برای انجام کار پیشنهاد دهید. همچنین توضیح دهید که آزمون فرضی که طراحی کرده‌اید دقیق (exact) است یا نادقیق.

ب) «میزان معنی‌دار بودن» افزایش شاخص عملکردی را بیابید. برای انجام این کار لازم است یک شاخص شناخته شده مانند p-value را محاسبه کنید. (برای محاسبه دم یا tail نمودارها می‌توانید از اینترنت یا زبان‌های برنامه‌نویسی آماری مانند R استفاده کنید (**امتحانات در آن سال مجازی بوده**). همچنین برخی نامساوی‌ها در بخش راهنمایی‌ها نیز آمده که قابل استفاده هستند.)

فرض کنید که سازمان پس از مطالعه گزارش شما، اعلام می‌کند که p-value گزارش شده به اندازه کافی پایین نیست و لذا افزایش شاخص عملکردی سازمان هنوز معنی‌دار نشده است.

ج) اگر می‌توانستید اندازه فقط یکی از جمعیت‌های نمونه ۱۰۰ نفره (قبل از تعویض مدیران) و یا ۱۰ نفره (بعد از آن) را به اندازه ۲۰ نفر افزایش داد تا مقدار p-value به بیشترین مقدار کمتر شود، شما پیشنهاد افزایش کدامیک را می‌دادید؟

جواب:

الف) قرار است ببینیم تغییر میانگین یک توزیع تصادفی معنی‌دار بوده یا نه. به عبارت دیگر، فرض کنید که شاخص‌های عملکردی نیروهای سازمان پیش از تغییر مدیران از توزیع f_1 با میانگین μ_1 و واریانس σ_1^2 پیروی کرده (در اینجا واریانس σ_1^2 مجموع واریانس ذاتی شاخص‌های عملکردی و دقت بازرس در اندازه‌گیری آنان است). این شاخص‌ها پس از تعویض مدیران از توزیع f_2 با میانگین و واریانس به ترتیب μ_2 و σ_2^2 پیروی می‌کنند.
سوال اصلی این است که آیا $\mu_1 = \mu_2$ یا خیر؟ لذا مطابق زیر المان‌های آزمون فرض را تعریف می‌کنیم:

- فرضیه H_0: میانگین‌ها برابرند، یعنی داریم $\mu_1 = \mu_2$.
- فرضیه H_1: میانگین‌ها برابر نیستند. یعنی داریم $\mu_1 < \mu_2$ یا $\mu_2 < \mu_1$.
- از آنجا برای انحراف معیارهای σ_1, σ_2 کران‌های بالای مستقل از داده داریم، می‌توانیم از آزمون آماری z-test استفاده کنیم. در صورتیکه مقادیر واریانس از روی داده‌ها محاسبه شده بود (و نه اطلاعات پیشین مستقل از دادگان) مجبور به استفاده از t-test بودیم.

آماره پیشنهادی بدین قرار است: باید آماره‌ای پیشنهاد داده شود که تحت فرضیه H_0 توزیع مشخصی داشته باشد. فرض کنید که $X_1, \ldots, X_{100}$ نمونه‌های i.i.d. متناظر با توزیع اولیه f_1، و نمونه‌های $Y_1, \ldots, Y_{10}$ مواردی مشابه و متناظر با توزیع f_2 باشند. فرض کنید که $\bar{X}_{100}$ و $\bar{Y}_{10}$ به ترتیب میانگین‌های نمونه‌ای متناظر با Xها و Yها باشند.
تحت فرضیه H_0 داریم:

$$\mathbb{E}\bar{X}_{100} = \mu_1 = \mu_2 \quad , \quad \mathrm{Var}\left(\bar{X}_{100}\right) = \frac{\sigma_1^2}{100} \leq \frac{5 \times 5}{100}$$

$$\mathbb{E}\bar{Y}_{10} = \mu_2 = \mu_1 \quad , \quad \operatorname{Var}\left(\bar{Y}_{10}\right) = \frac{\sigma_2^2}{10} \leq \frac{8 \times 8}{10}$$

با استفاده از فرض گاوسی بودن نمونه‌ها برای $\bar{X}_{100}$ و $\bar{Y}_{10}$، و همچنین استقلال این دو از یکدیگر داریم:

$$Z \triangleq \frac{\bar{X}_{100} - \bar{Y}_{10}}{\sqrt{\frac{5^2}{100} + \frac{8^2}{10}}} \sim \mathcal{N}\left(0,\sigma^2\right), \quad \sigma < 1$$

که می‌بایست روی مقدار z-value فوق آزمون زده شود. در ضمن، به دلیل استفاده از تقریب گاوسی، آزمون فرض طراحی شده دقیق (exact) نیست و مقدار p-value به صورت دقیق محاسبه یا کراندار نمی‌گردد. مگر آنکه فرض کنیم خود توزیع‌های f_1 و f_2 گاوسی بوده باشند.

ب) به منظور چک کردن برقراری فرض باطل یا فرضیه جایگزین، می‌توان از یک آزمون دوطرفه ساده استفاده کرد. در این صورت لازم است تا با فرض $Z \sim \mathcal{N}\left(0,\sigma^2\right)$، احتمال زیر را محاسبه نمود:

$$\text{p} - \text{value} = \mathbb{P}\left(|Z| > \frac{57 - 54}{\sqrt{\frac{25}{100} + \frac{64}{10}}}\right) \leq 0.245$$

که علامت کوچک‌تر یا مساوی به دلیل این است که واریانس z-value که با σ^2 نشان داده شده است با کران بالای خود یعنی ۱ تقریب خورده است.

ج) مقدار p-value برابر با 0.245 به وضوح بسیار بزرگ است و نشان می‌دهد افزایش شاخص عملکردی به لحاظ آماری معنی‌دار نشده است. به منظور بهبود z-value، با فرض ثابت ماندن درصدهای بدست آمده برای میانگین‌ها (یعنی ۵۷٪ و ۵۴٪) لازم است تا مخرج آن که مشتمل بر انحراف معیارهاست تا جای ممکن کوچک شود.

در صورت اضافه نمودن ۲۰ نفر جدید به جمعیت اولیه، مخرج کسر برابر با $\sqrt{\dfrac{25}{120} + \dfrac{64}{10}}$ و در صورت اضافه نمودن آنان به جمعیت دوم مخرج z-value برابر با $\sqrt{\dfrac{25}{100} + \dfrac{64}{30}}$ می‌گردد. واضح است که حالت دوم باعث افزایش چشمگیرتر مقدار z-value و لذا کاهش مقدار p-value خواهد گردید. مقدار کران بالای p-value در این حالت دوم برابر با 0.052 خواهد شد که کران بسیار بهتری است.

سوال ۵ (۴ نمره):

ملکول DNA در بدن موجودات زنده یک دنباله طولانی به طول N و متشکل از چهار الفبای اصلی A,C,G و T است. در اینجا، الفبای A,C,G,T بیانگر حروف ابتدایی از نام ۴ نوع منحصر به فرد از زیرملکولهایی به نام نوکلئوتید هستند.
فرض کنید که بتوانیم رشته DNA را به صورت تصادفی مدل‌سازی کنیم: فرض کنید هر یک از N حرف رشته مستقل از سایرین و با احتمال‌های یکسان $1/4$ بتواند هر یک از اعضای الفبای A,C,G,T باشد (دقت کنید که در واقعیت اینطور نیست!). در این صورت، رشته DNA مورد بحث در این سوال، یک نمونه (یا realization) از این توزیع خواهد بود.

الف) قصد داریم بدانیم که یک زیررشته (substring) فرضی و ساختگی خاص به طول L (برای مثال، یک زیر رشته ۱۰۰تایی مانند $ACCGTATT...GCC$) با چه احتمالی در حداقل یک جا از رشته DNA دیده خواهد شد. یک کران بالا برای احتمال خواسته شده بیابید.
(راهنمایی: می‌توانید از کران اجتماع یا Union Bound استفاده کنید)

ب) یک کران پایین نیز برای احتمال قسمت الف) پیدا کنید.
(راهنمایی: رشته DNA را به زیررشته‌های بدون همپوشانی با طول L تقسیم کنید)

ج) نشان دهید که در صورت انتخاب $L \geq (1+\varepsilon)\log_4 N$ احتمال دیده شدن یک زیررشته پیش فرض و ساختگی در رشته تصادفی DNA با افزایش N به سمت صفر میل می‌کند (به ازای هر $\epsilon > 0$). همچنین، در صورت انتخاب $L \leq (1-\varepsilon)\log_4 N$ احتمال مشاهده هر زیررشته ساختگی دلخواهی در رشته اصلی به سمت ۱ خواهد رفت.

د) (**امتیازی**) نشان‌دهید احتمال مشاهده زیررشته‌های خودتکرارشونده با طول $L \geq (2+\varepsilon)\log_4 N$ با افزایش N به سمت صفر میل خواهد کرد. رشته‌های خودتکرارشونده زیررشته‌هایی از DNA هستند که حداقل در یک جای دیگر از رشته نیز دوباره دیده می‌شوند.
(برای این بخش لازم است توضیحات مفصل ارائه دهید و صرف نوشتن روابط کافی نیست).

جواب:

الف) احتمال خواسته شده را با P نشان می‌دهیم. فرض کنید که رشته فرض مورد نظر را با S نشان دهیم. رخدادهای $A_1, A_2, \ldots, A_{N-L+1}$ را به صورت زیر تعریف می‌کنیم:

- رخداد A_i زمانی رخ می‌دهد که زیررشته شروع شده از اندیس i ام به طول L، تماماً با زیررشته پیش‌فرض S یکی باشد.؟

زیررشته پیش‌فرض S تصادفی نیست، ولی رشته اصلی به صورت تصادفی با احتمال‌های برابر ۱/۴ برای هر چهار عضو الفبا مدلسازی شده است. لذا احتمال یکی بودن هر یک از زیررشته‌های مورد نظر رخدادهای A_i با رشته ثابت S برابر با 4^{-L} خواهد بود. در اینجا از مستقل بودن اعضای رشته DNA نیز استفاده شده است.

مابقی ساده است. کافی است بنویسیم:

$$P = \mathbb{P}\left(A_1 \cup \ldots \cup A_{N-L+1}\right) \le \sum_{i=1}^{N-L+1} \mathbb{P}\left(A_i\right) = (N-L+1)\,4^{-L} \le N\left(\frac{1}{4}\right)^L$$

ب) دلیل استفاده از کران اجتماع در قسمت قبل این واقعیت بوده که رخدادهای A_i به دلیل همپوشانی‌های بالقوه از یکدیگر مستقل نبوده و امکان محاسبه اشتراک آنان بسیار سخت است. برای حل این مشکل، از کران پایین زیر استفاده می‌کنیم:

$$P = \mathbb{P}\left(A_1 \cup \ldots \cup A_{N-L+1}\right) \ge \mathbb{P}\left(A_1 \cup A_{L+1} \cup A_{2L+1} \cup \ldots \cup A_{L\lfloor N/L\rfloor - L+1}\right)$$

که معادل این است که رشته DNA به تعداد $\left\lfloor \frac{N}{L} \right\rfloor$ زیررشته با طول L تقسیم شده است که هیچ همپوشانی با یکدیگر ندارند. لذا رخدادهای متناظر با آنان از یکدیگر مستقل هستند. پس می‌توان نوشت:

$$\begin{aligned} P \ge \mathbb{P}\left(A_1 \cup A_{L+1} \cup A_{2L+1} \cup \ldots \cup A_{L\lfloor N/L\rfloor - L+1}\right) &= 1 - \mathbb{P}\left(A_1^c \cap A_{L+1}^c \cap \ldots \cap A_{L\lfloor N/L\rfloor - L+1}^c\right) \\ &= 1 - \prod_{i=0}^{\lfloor N/L\rfloor - 1} \mathbb{P}\left(A_{iL+1}^c\right) \end{aligned}$$

لذا خواهیم داشت:

$$P \ge 1 - \left(1 - \frac{1}{4^L}\right)^{\left\lfloor \frac{N}{L} \right\rfloor}$$

ج) کافی است که مقادیر اشاره شده برای L را در کران‌های بالا و پایین احتمال که در قسمت‌های الف) و ب) بدست آورده‌ایم جایگذاری کنیم. در مورد $L \ge (1+\varepsilon)\log_4 N$ داریم:

$$\lim_{N\to\infty} P \le \lim_{N\to\infty} N \frac{1}{4^{(1+\varepsilon)\log_4 N}} = \lim_{N\to\infty} N^{-\varepsilon} = 0$$

و در مورد $L \le (1-\varepsilon)\log_4 N$

$$\lim_{N\to\infty} P \ge \lim_{N\to\infty} 1 - \left(1 - \frac{1}{4^L}\right)^{\left\lfloor \frac{N}{L} \right\rfloor} = \lim_{N\to\infty} 1 - e^{-\frac{N}{L4^L}} \ge \lim_{N\to\infty} 1 - e^{-\frac{N}{N^{1-\varepsilon}\log N}} = \lim_{N\to\infty} 1 - e^{-\frac{N^\varepsilon}{\log N}} = 1$$

د) (**امتیازی**) رخداد A_{ij} را برای $i, j = 1,2,\ldots, N-L+1$ (با فرض $i \ne j$) به این صورت تعریف می‌کنیم که زیررشته به طول L که از اندیس i شروع شده، با زیررشته‌ای با طول مشابه که از اندیس j شروع شده تماماً یکی باشد. در این صورت می‌توان نوشت:

$$\mathbb{P}\left(A_{ij}\right)=\prod_{k=0}^{L-1}\mathbb{P}\left(X_{i+k}=X_{j+k}\right)=\prod_{k=0}^{L-1}\left[\sum_{x\in\{A,C,G,T\}}\mathbb{P}\left(X_{i+k}=x\right)\mathbb{P}\left(X_{j+k}=x\right)\right]=4^{-L}$$

مابقی کار ساده بوده و دوباره لازم است که یک کران اجتماع ساده در نظر گرفته شود:

$$\mathbb{P}\left(\text{self}-\text{repeat}\right)=\mathbb{P}\left(\bigcup_{i,j}A_{i,j}\right)\leq\binom{N-L+1}{2}4^{-L}\leq N^{2}4^{-L}$$

و در آخر با جایگذاری مقدار L مشابه با قسمت ج) خواهیم داشت:

$$\lim_{N\to\infty}\mathbb{P}\left(\text{self}-\text{repeat}\right)\leq\lim_{N\to\infty}N^{2}\frac{1}{4^{(2+\varepsilon)\log_{4}N}}=\lim_{N\to\infty}\frac{N^{2}}{N^{2}}N^{-\varepsilon}=0$$