سوال ۱.

الف

با یک متغیر تصادفی به نام S کار می‌کنیم. تابع توزیع تجمعی (CDF) این متغیر تصادفی با $F_S(s)$ نمایش داده می‌شود. سپس می‌خواهیم $F_S(s)$ را به شکلی معادل با دیگر متغیرهای تصادفی توضیح دهیم.

$F_S(s)$ را می‌توان به صورت زیر نوشت:

$$F_S(s) = \begin{cases} s & ۰ \leq s < ۱ \\ ۲ - s & ۱ \leq s \leq ۲ \\ ۰ & \text{در غیر این صورت} \end{cases}$$

حال می‌توانیم به توضیح روابطی که در متن برقرار است، بپردازیم. ابتدا با استفاده از تعریف تابع توزیع تجمعی، معادله زیر را می‌توان نوشت:

$$F_S(s) = P(S < s)$$

سپس با استفاده از اعلانی که در متن داده شده، S را می‌توان به صورت جمع دو متغیر تصادفی X و Y نمایش داد:

$$S = X + Y$$

حال با جایگذاری معادله $S = X + Y$ در $P(S < s)$ و استفاده از قاعده توزیع شرطی، می‌توان نوشت:

$$P(S < s) = P(X + Y < s)$$

سپس، با استفاده از تعریف تابع توزیع تجمعی متغیر تصادفی X، $P(X < s - Y)$ را می‌توان به صورت زیر بازنویسی کرد:

$$P(X + Y < s) = P(X < s - Y)$$

با در نظر گرفتن اینکه مقدار Y ثابت است، می‌توانیم از رابطه $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid Y))$ استفاده کنیم و معادله بالا را به شکل زیر بازنویسی کنیم:

$$P(X < s - Y) = \mathbb{E}(P(X < s - Y \mid Y = y))$$

با استفاده از قاعده بازنویسی توزیع شرطی، می‌توانیم $P(X < s - Y \mid Y = y)$ را به شکل $F_X(s - y)$ بازنویسی کنیم. پس معادله بالا به شکل زیر ساده می‌شود:

$$\mathbb{E}(P(X < s - Y \mid Y = y)) = \mathbb{E}(F_X(s - y))$$

در اینجا، به دلیل محدود بودن بازه مقداردهی به s، یعنی $۰ \leq s \leq ۲$، می‌توانیم معادله بالا را به صورت یک انتگرال بر حسب y نوشت:

$$\mathbb{E}(F_X(s - y)) = \int_{-\infty}^{\infty} F_X(s - y) f_Y(y) dy$$

دقت کنید که $f_X(s-y)$ و $f_Y(y)$ تابع چگالی احتمال (PDF) متغیرهای تصادفی X و Y را نشان می‌دهند.

حال با توجه به بازه‌های مقداردهی به s، می‌توانیم به دو حالت متفاوت معادله بالا را مورد بررسی قرار دهیم. این دو حالت عبارتند از: $۰ \leq s \leq ۱$ و $۱ \leq s \leq ۲$.

حال، برای هر یک از این دو حالت، انتگرال را بر حسب y حل می‌کنیم.

در حالت اول، وقتی $۰ \leq s \leq ۱$، با توجه به بازه‌های مقداردهی به y و $s - y$، می‌توانیم انتگرال را به شکل زیر بنویسیم:

$$\int_{-\infty}^{\infty} F_X(s-y) f_Y(y) dy = \int_{۰}^{s} ۱ dy = s$$

حال، به حالت دوم، یعنی $۱ \leq s \leq ۲$، می‌پردازیم. در این حالت نیز، با توجه به بازه‌های مقداردهی به y و $s - y$، انتگرال به صورت زیر حل می‌شود:

$$\int_{-\infty}^{\infty} F_X(s-y) f_Y(y) dy = \int_{s-۱}^{۱} ۱ dy = ۲ - s$$

با توجه به نتایج به دست آمده در دو حالت، می‌توانیم تابع توزیع تجمعی $F_S(s)$ را به شکل زیر نوشت:

$$F_S(s) = \begin{cases} s & \text{برای } ۰ \leq s < ۱ \\ ۲ - s & \text{برای } ۱ \leq s \leq ۲ \\ ۰ & \text{در غیر این صورت} \end{cases}$$

با این توضیحات، می‌توانیم روشن‌تر درک کنیم که چگونه متغیرهای تصادفی X و Y منجر به متغیر تصادفی S می‌شوند و چگونه تابع توزیع تجمعی S به دست می‌آید.

ب

با استفاده از تعریف تابع چگالی احتمال مشروط، تابع چگالی احتمال شرطی $f_{X|S}(x|s)$ را به صورت زیر بیان می‌کنیم:

$$f_{X|S}(x|s) = \frac{f_{X,S}(x,s)}{f_S(s)} = \frac{f_{X,Y}(x, s-x)}{f_S(s)}$$

حال، با توجه به بازه مقداردهی به s که در متن اصلی ذکر شده است، تابع چگالی احتمال شرطی را در چند حالت متفاوت مورد بررسی قرار می‌دهیم. این حالات عبارتند از: $۰ \leq x \leq s < ۱$، $۰ \leq x \leq ۱ \leq s < ۲$ و سایر حالات.

در حالت اول، وقتی $۰ \leq x \leq s < ۱$ ، با توجه به بازه مقداردهی به x و s، تابع چگالی احتمال شرطی به صورت زیر بدست می‌آید:

$$f_{X|S}(x|s) = \begin{cases} \frac{f_{X,Y}(x,s-x)}{s} = \frac{s-x}{s} & \text{برای } ۰ \leq x \leq s < ۱ \end{cases}$$

در این حالت، تابع چگالی احتمال شرطی برابر با نسبت فاصله $s - x$ تا s به s است. یعنی مقدار X به مرور از ۰ به s افزایش می‌یابد و با افزایش x، احتمال کاهش می‌یابد.

حال، به حالت دوم، یعنی $۰ \leq x \leq ۱ \leq s < ۲$ می‌پردازیم. در این حالت، تابع چگالی احتمال شرطی به صورت زیر محاسبه می‌شود:

$$f_{X|S}(x|s) = \begin{cases} \frac{f_{X,Y}(x,s-x)}{۲-s} = ۱ & \text{برای } ۰ \leq x \leq ۱ \leq s < ۲, s - x < ۱ \\ \frac{f_{X,Y}(x,s-x)}{۲-s} = ۰ & \text{برای } ۰ \leq x \leq ۱ \leq s < ۲, s - x > ۱ \end{cases}$$

در این حالت، مقدار تابع چگالی احتمال شرطی برابر ۱ است زیرا X و Y به طور مستقل از هم توزیع شده‌اند و مقدار X در بازه $[۰, ۱]$ تعریف شده است. بنابراین، برای همه مقادیر x در این بازه وقتی که $s - x < ۱$ باشد، احتمال برابر ۱ است و در غیر این صورت، احتمال برابر صفر است.

در سایر حالات، تابع چگالی احتمال شرطی برابر با صفر است.

ج

برای محاسبه امید ریاضی X به شرط $S = ۰/۵$، از تابع چگالی احتمال شرطی استفاده می‌کنیم. طبق قسمت قبل، در این حالت، تابع چگالی احتمال شرطی برابر است با $\frac{s-x}{s}$.

با جایگذاری $s = ۰/۵$ و محاسبه امید ریاضی، داریم:

$$\mathbb{E}(X|S = ۰/۵) = \int_{۰}^{۰/۵} x \frac{۰/۵ - x}{۰/۵} dx$$

حال می‌توانیم با محاسبه این انتگرال، مقدار امید ریاضی X به شرط $S = ۰/۵$ را تقریباً محاسبه کنیم. این مقدار تقریباً برابر با ۰/۰۱۶۶۷ است.

د

برای محاسبه امید ریاضی M، ما نیاز داریم که به تمام حالت‌های محتمل که در بخش دوم بدست آورده‌ایم، امید ریاضی را محاسبه کنیم.

طبق بخش ب، حالاتی که برای محاسبه امید ریاضی در نظر گرفته می‌شوند عبارتند از:

$$\mathbb{E}(M) = \begin{cases} \int_{۰}^{s} \frac{s-x}{s} dx & \text{برای } ۰ \leq x \leq s < ۱ \\ \int_{۰}^{s} ۱ dx = s & \text{برای } ۰ \leq x \leq ۱ \leq s < ۲, s - x < ۱ \\ ۰ & \text{در غیر این صورت} \end{cases}$$

حال، با استفاده از فرمول انتگرالی امید ریاضی و با محاسبه این انتگرال‌ها، می‌توانیم مقدار امید ریاضی M را بدست آوریم.

سوال ۲. در ابتدا، برای محاسبه مقدار میانگین کل امتیازی که می‌توانیم حساب کنیم، از یک فرمول استفاده می‌کنیم. این مقدار برابر است با جمع مقادیری که در هر مرحله به ازای آنها میانگین را محاسبه می‌کنیم. فرمول به صورت زیر است:

$$\sum ۲^{i-۱} \mathbb{E}(X_i)$$

در این فرمول، i نشان‌دهنده‌ی شماره‌ی مرحله است و $\mathbb{E}(X_i)$ میانگین در آن مرحله را نشان می‌دهد.

حال برای محاسبه‌ی احتمال درست برای حدسی که در مرحله i برای برنده می‌زنیم، نکته‌ای وجود دارد که باید آن را مدنظر قرار دهیم. در درخت برنده‌ها، برای برنده شدن در مرحله i، باید تمام مراحل قبلی را نیز به درستی حدس بزنیم.

بنابراین، احتمال درست برای حدس زدن مرحله i، بستگی به این دارد که آیا تمام مراحل قبلی را به درستی حدس زده‌ایم یا خیر.

پس به صورت کلی داریم برای مثال داریم:

$$Points = \sum ۲^{i-۱}۲^{n-۲i}$$

سوال ۳. با توجه به اینکه اطلاعات خاصی از توزیع این متغیرها در اختیار نداریم، برای حدس زدن که W از Z بیشتر است یا خیر، می‌توانیم استراتژی ساده‌ای را پیش بگیریم: به صورت تصادفی حدس بزنیم.

یعنی در هر بار، با یک احتمال نیمه به نیمه، حدس بزنیم که W بیشتر از Z است یا خیر. با این روش، احتمال اینکه حدس ما درست باشد حدود ۱/۲ است.

برای اینکه احتمال موفقیت این استراتژی برابر با $\epsilon + ۱/۲$ شود، اگر ϵ را از مجموعه اعداد حقیقی کوچکتر از ۱/۲ در نظر بگیریم، می‌توانیم یک ترفند استفاده کنیم: اگر W و Z هر دو عدد حقیقی باشند و متفاوت باشند (که به این معنی است که یکی بیشتر و دیگری کمتر است)، اگر فرض کنیم W بیشتر از Z است، در ϵ درصد موارد، ما به جای حدس $W < Z$ ، حدس $W > Z$ را انتخاب می‌کنیم. این روش موجب می‌شود احتمال موفقیت ما در حدس زدن درست برابر با $\epsilon + ۱/۲$ شود.

به عبارت دیگر، اگر X یک متغیر تصادفی با توزیع برنولی با پارامتر ϵ باشد، استراتژی ما به این صورت است که اگر $X = ۱$ ، ما حدس می‌زنیم $W > Z$ و اگر $X = ۰$ ، حدس می‌زنیم $W < Z$. از آنجا که X با احتمال ϵ برابر با ۱ است، این استراتژی می‌تواند احتمال موفقیت ما را به $\epsilon + \frac{۱}{۲}$ برساند.

توجه داشته باشید که این استراتژی فقط در صورتی کاربرد دارد که ϵ کوچکتر از $\frac{۱}{۲}$ باشد. اگر ϵ بزرگتر از $\frac{۱}{۲}$ باشد، استراتژی مناسبی وجود ندارد که احتمال موفقیت ما را به $\epsilon + \frac{۱}{۲}$ برساند، زیرا در حالت کلی و با داشتن دانسته‌های فعلی، احتمال موفقیت ما هرگز نمی‌تواند بیشتر از $\frac{۱}{۲}$ باشد.

سوال ۴.

الف

برای اینکه تخمین‌گر MLE (تخمین‌گر بیشینه درستنمایی) برای مقدار θ را بیابیم، می‌خواهیم درستنمایی داده‌ها را بیشینه کنیم.

فرض کنیم که $Y_۱, Y_۲, \ldots, Y_n$ نمونه‌هایی از توزیع $\theta + x$ باشند که دریافت کرده‌ایم. چون x نویز گاوسی با میانگین صفر و واریانس ۱ است، پس $Y = \theta + x$ نیز یک توزیع گاوسی با میانگین θ و واریانس ۱ دارد.

درست‌نمایی (likelihood) این داده‌ها با توجه به مقدار θ به صورت زیر است:

$$L(\theta; Y_۱, Y_۲, \ldots, Y_n) = f(Y_۱, Y_۲, \ldots, Y_n; \theta)$$

$$= f(Y_۱; \theta) \cdot f(Y_۲; \theta) \cdot \ldots \cdot f(Y_n; \theta)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-(Y_1-\theta)^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(Y_2-\theta)^2/2} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi}} e^{-(Y_n-\theta)^2/2}$$

$$= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(Y_i-\theta)^2}$$

برای بیشینه کردن این تابع درستنمایی، می‌توانیم به جای آن لگاریتم درستنمایی را بیشینه کنیم:

$$l(\theta) = \log(L(\theta; Y_1, Y_2, \ldots, Y_n))$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(Y_i - \theta)^2$$

برای بیشینه‌سازی $l(\theta)$ نیاز است که مشتق آن نسبت به θ را برابر صفر قرار دهیم:

$$\frac{dl(\theta)}{d\theta} = \sum_{i=1}^{n}(Y_i - \theta) = 0$$

از این معادله به دست می‌آید که:

$$\hat{\theta}_{MLE} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

پس تخمین‌گر MLE برای θ میانگین نمونه‌ها است.

ب

حالا برای بررسی consistent بودن (یا همگرایی) تخمین‌گر MLE برای θ، باید نشان دهیم که هنگامی که تعداد نمونه‌ها (n) به سمت بی‌نهایت میل می‌کند، تخمین‌گر MLE به مقدار واقعی θ همگرا می‌شود. به عبارت دیگر:

با فرض $\hat{\theta}_{MLE,n}$ به عنوان تخمین‌گر MLE برای θ بر اساس n نمونه، می‌خواهیم نشان دهیم که $\hat{\theta}_{MLE,n}$ به θ همگرا می‌شود هنگامی که $n \to \infty$.

از قانون اعداد بزرگ می‌دانیم که اگر $Y_1, Y_2, \ldots, Y_n$ متغیرهای تصادفی i.i.d (independent and identically distributed) باشند و $\mathbb{E}[Y_i] = \mu$ ، آنگاه میانگین نمونه $\bar{Y_n} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ به μ همگرا می‌شود هنگامی که $n \to \infty$.

در این مورد، $Y_i = \theta + x$ و $x \sim N(0, 1)$ ، بنابراین $\mathbb{E}[Y_i] = \mathbb{E}[\theta + x] = \theta + \mathbb{E}[x] = \theta$ زیرا $\mathbb{E}[x] = 0$ است.

پس بر اساس قانون اعداد بزرگ، $\hat{\theta}_{MLE,n} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ به θ همگرا می‌شود هنگامی که $n \to \infty$.

بنابراین، تخمین‌گر MLE برای θ در این مورد consistent است.

ج

ما با یک مسئله‌ی استنتاج بیزی روبرو هستیم، که در آن مقدار θ یکی از دو مقدار ممکن ۱ یا ـ۱ است و توزیع prior آن یک توزیع برنولی با پارامتر p است.

یعنی

$$P(\theta = ۱) = p$$

و

$$P(\theta = -۱) = ۱ - p$$

با فرض Y به عنوان مشاهده شده، می‌خواهیم تخمین‌گر بیزی θ را با استفاده از قاعده بیز پیدا کنیم. بر اساس قاعده بیز،

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

از آنجا که $P(Y)$ برای تمام θها یکسان است، می‌توانیم از آن صرف‌نظر کنیم و فقط به صورت نسبی توزیع‌ها را بررسی کنیم. این به ما اجازه می‌دهد تا θ را انتخاب کنیم به طوری که $P(Y|\theta)P(\theta)$ را بیشینه کند.

برای $\theta = ۱$،

$$P(Y|\theta = ۱)P(\theta = ۱) = \frac{۱}{\sqrt{۲\pi}} e^{-(Y-۱)^۲/۲} p$$

و برای $\theta = -۱$،

$$P(Y|\theta = -۱)P(\theta = -۱) = \frac{۱}{\sqrt{۲\pi}} e^{-(Y+۱)^۲/۲} (۱ - p)$$

حالا تخمین‌گر بیزی θ را می‌توانیم به این شکل بیان کنیم:

$$\hat{\theta}_{Bayes} = \begin{cases} ۱ & \text{if} \frac{۱}{\sqrt{۲\pi}} e^{-(Y-۱)^۲/۲} p > \frac{۱}{\sqrt{۲\pi}} e^{-(Y+۱)^۲/۲} (۱ - p) \\ -۱ & \text{otherwise} \end{cases}$$

د

با استفاده از تخمین‌گر LMMSE داریم:

$$\alpha_i = \frac{Cov(\theta, y_i)}{Var(Y_i)}$$

سوال ۵.

الف

$$L_x(c) = |x-c|$$

$$\rightarrow \int_c^\infty f_x(x)(x-c)\,dx + \int_{-\infty}^c f_x(x)(c-x)\,dx$$

$$= -(1 - F_x(c)) + F_x(c) = 2F_x(c) - 1 = 0$$

$$\rightarrow c = F_x^{-1}(\frac{1}{2})$$

ب

$$L_x(c) = (x-c)^2$$

$$\rightarrow \int_{-\infty}^\infty (x-c)^2\,dx$$

$$= 2c - 2E[x] = 0 \rightarrow c = E[x]$$

ج

برای این بخش نیز داریم:

$$\int_{-\infty}^c f_x(x)k(c-x)\,dx + \int_c^\infty f_x(x)(x-c)\,dx = 0$$

$$\rightarrow F_x(c) = \frac{1}{k+1}$$

$$c = F_x^{-1}\frac{1}{k+1}$$

سوال ۶. در اینجا از آزمون t ولچ برای مقایسه دو گروه استفاده می‌کنیم. فرمول آزمون t ولچ به صورت زیر است:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

در اینجا، $\bar{X_۱}$ و $\bar{X_۲}$ به ترتیب میانگین دو گروه را نشان می‌دهند. همچنین $\sigma_۱^۲$ و $\sigma_۲^۲$ واریانس دو گروه و $N_۱$ و $N_۲$ تعداد نمونه‌ها در هر گروه را نشان می‌دهند.

با استفاده از این فرمول و با داشتن مقادیر میانگین‌ها ($X_۱ = ۴۴$ و $X_۲ = ۵۷$) و واریانس‌ها ($\sigma_۱^۲ = ۸۲/۵$ و $\sigma_۲^۲ = ۱۵۴/۳$) و تعداد نمونه‌ها ($N_۱ = ۵$ و $N_۲ = ۷$)، مقدار t را محاسبه می‌کنیم:

$$t = \frac{۵۷ - ۴۴}{\sqrt{\frac{۸۲/۵}{۵} + \frac{۱۵۴/۳}{۷}}} = ۲/۰۹۴$$

همچنین در اینجا درجه آزادی (degrees of freedom) را نیز محاسبه می‌کنیم که به صورت زیر است:

$$df = N_۱ + N_۲ - ۲ = ۷ + ۵ - ۲ = ۱۰$$

سپس برای محاسبه‌ی مقدار p-value از زبان R و پکیج p-test استفاده می‌کنیم. برای محاسبه p-value ، از تابع pt در R استفاده می‌کنیم. مقدار p-value برابر است با دو برابر احتمالی که t استخراج شده از توزیع t با درجه آزادی df بزرگتر یا مساوی t محاسبه شده باشد:

p-value=2*pt(2.094, df = 10, lower=FALSE) 0.063p-value =

2*pt(2.094, df = 10, lower = FALSE) 0.063

سپس با مقایسه p-value با سطح اهمیت $\alpha = ۰/۰۵$ ، اگر p-value بیشتر از α باشد، نتیجه می‌گیریم که نتایج آماری معنادار نیست و تفاوت آماری معناداری وجود ندارد. در اینجا، زیرا $۰/۰۶۳ > ۰/۰۵$ است، پس می‌توان نتیجه گرفت که تفاوت بین دو گروه معنادار نیست.

سوال ۷. احتمال وقوع α_i که از توزیع نرمال با میانگین صفر و واریانس σ پیروی می‌کند، برابر با احتمال وقوع $X = y_i - ax_i - b$ است. در واقع، ما به دنبال محاسبه $P(X = y_i - ax_i - b)$ هستیم، که در آن X از توزیع نرمال با میانگین صفر و واریانس σ است.

سپس، تابع لگاریتم درست‌نمایی (log-likelihood) را بررسی می‌کنیم. تابع لگاریتم درست‌نمایی برای مقادیر $(x_۱, y_۱), (x_۲, y_۲), \ldots, (x_n, y_n)$ و پارامترهای a و b به صورت زیر تعریف می‌شود:

$$\mathcal{L}((x_۱, y_۱), (x_۲, y_۲), \ldots, (x_n, y_n); a, b) = \mathcal{L}(y_۱ - ax_۱ - b, y_۲ - ax_۲ - b, \ldots, y_n - ax_n - b; a, b)$$

سپس این تابع لگاریتم درست‌نمایی را ساده می‌کنیم و به صورت زیر نوشته می‌شود:

$$\mathcal{L} = (\frac{۱}{\sqrt{۲\pi\sigma}})^n e^{\frac{-۱}{۲\sigma}\sum_{i=۱}^{n}(y_i - ax_i - b)^۲}$$

با اعمال لگاریتم بر روی تابع لگاریتم درست‌نمایی، می‌توان آن را به شکل زیر نوشت:

$$\ln(\mathcal{L}) = \frac{-n}{۲}\ln(۲\pi\sigma) + \frac{-۱}{۲\sigma}\sum_{i=۱}^{n}(y_i - ax_i - b)^۲$$

سپس با مشتق‌گیری از این تابع لگاریتم درست‌نمایی نسبت به پارامترهای a و b، داریم:

$$\frac{d}{da}(\ln(\mathcal{L})) = \frac{۱}{۲\sigma}\sum_{i=۱}^{n} ۲x_i(y_i - ax_i - b) = ۰$$

$$\frac{d}{db}(\ln(\mathcal{L})) = \frac{۱}{۲\sigma}\sum_{i=۱}^{n}۲(y_i - ax_i - b) = ۰$$

با حل این دو معادله، به مقادیر a و b برسیم. از معادله دوم می‌توان نتیجه گرفت که باید برابر باشد:

$$\bar{y} - a\bar{x} = b$$

که در آن $\bar{y}$ و $\bar{x}$ به ترتیب نشان دهنده میانگین مقادیر y_i و x_i هستند.

سپس با جایگذاری این رابطه در عبارت اول، به عبارت زیر می‌رسیم:

$$\sum_{i=۱}^{n} x_i(y_i - ax_i - \bar{y} + a\bar{x}) = \sum_{i=۱}^{n} x_iy_i - ax_i^۲ - x_i\bar{y} + ax_i\bar{x} = \sum_{i=۱}^{n}(x_iy_i - x_i\bar{y}) - \sum_{i=۱}^{n}(ax_i^۲ - ax_i\bar{x}) = ۰$$

بنابراین، می‌توان a را به صورت زیر محاسبه کرد:

$$a = \frac{\sum_{i=۱}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=۱}^{n}(x_i - \bar{x})^۲}$$

همچنین a و b برابر هستند پس برای جواب هر دو بخش سوال داریم این را.

موفق باشید.