نام کتاب

دکتر ...

(عضو هیأت علمی دانشگاه فردوسی مشهد)

دکتر ...

(عضو هیأت علمی دانشگاه فردوسی مشهد)

فهرست مطالب

فصل ۱

احتمالات

آزمایش تصادفی هر آزمایش که با اتفاق همراه بود، نتیجه آن به طور قطع مشخص نباشد ولی مجموعه تمام حالات آن قابل پیش‌بینی باشد را یک آزمایش تصادفی نامند.

فضای نمونه مجموعه تمام نتایج ممکن یک آزمایش تصادفی را فضای نمونه نامند و آن‌را با Ω نمایش می‌دهند.

مثال ۱.۱.۱. ۱- *میزان کریستال یک قطعه سنگ نازک:* $\Omega = R^+$

۲- *اندازه طول یک نمونه سنگ انتخاب شده،* $\Omega = R^+$

۳- *تعداد زمین لرزه‌هایی که سالانه در یک ناحیه رخ دهد.* $\Omega = \{۰, ۱, ۲, \cdots\}$

۴- *گروه خونی یک فرد:* $\{A, B, O, AB\}$

۵- *درصد شوری آب یک نمونه چاه:*

۶- *سطح آب پشت یک سد در لحظه* t.

فضای نمونه گسسته فضاهای نمونه متناهی یا شمارش‌پذیر ۳ و ۴ را گسسته نامند.

پیشامد تصادفی هر زیرمجموعه فضای نمونه را یک پیشامد تصادفی نامند، پیشامدهای تک عضوی را ساده، تهی($\emptyset$) را پیشامد محال، فضای نمونه (Ω) را پیشامد حتمی نامند.

پیشامدهای ناسازگار (مجزا = disjoint)

۱- دو پیشامد A و B مجزا هستند اگر $A \cap B = \emptyset$

۲- دنباله پیشامدهای $\{A_{i,j} \geq ۱\}$ را دوبه‌دو مجزا نامند اگر: $A_i \cap A_j = \emptyset \ , \forall i \neq j$

۲.۱ مدل احتمال

۱.۲.۱ مدل احتمال متناهی

اگر $\Omega = \{e_۱, e_۲, \cdots, e_n\}$ و $P_i = P[\{e_i\}]$ و $i = ۱, ۲, \cdots, n$ به قسمی که، $۰ \leq P_i \leq ۱$ و $\sum_{i=۱}^{n} P_i = ۱$، در این‌صورت برای هر پیشامد $A \subseteq \Omega$،

$$P(A) = \sum_{\{j, e_j \in A\}} P_j$$

۲.۲.۱ مدل احتمال یکنواخت

اگر $P_i = \frac{۱}{n}$ و $i = ۱, ۲, \cdots, n$، آن‌گاه

$$\begin{aligned} P(A) &= \frac{\text{تعداد اعضای } A}{\text{تعداد اعضای } \Omega} \\ &= \frac{n(A)}{n(\Omega)} = \frac{\text{تعداد کل حالات مساعد}}{\text{تعداد کل حالات ممکن}} \end{aligned}$$

مثال ۱.۲.۱. *۱-در ۲۰ نمونه انتخاب شده از یک نوع سنگ در ۸ تای آن‌ها کریستال مشاهده شده است، شانس وجود کریستال در هر نمونه سنگ چقدر است:*

$$P = \frac{۷}{۲۰} \ ;$$

۳.۲.۱ مدل احتمال نامتناهی

الف- اگر $\Omega = \underset{a,b\in R}{[a,b]}$ *آن‌گاه:*

$$\forall A \subseteq \Omega \ : \ P(A) = \frac{|A|}{|\Omega|}$$
$$= \frac{A \text{ طول}}{\Omega \text{ طول}} = \frac{|A|}{b-a}$$

ب- اگر $\Omega \subseteq R^{۲}$ *(*Ω *یک زیرمجموعه کران‌دار) آن‌گاه*

$$\forall A \subseteq \Omega \ : \ P(A) = \frac{A \text{ مساحت}}{\Omega \text{ مساحت}}$$

اصول موضوع کولموگوروف

فرض کنید Ω یک فضای نمونه و $\mathcal{F}$ گردایه‌ای از زیرمجموعه‌های Ω باشد در حالت خاص $\mathcal{F}$ می‌تواند مجموعه توان Ω باشد،در این‌صورت تابع مجموعه‌ای $P : \mathcal{F} \to R$ یک تابع احتمال است اگر در سه اصل زیر صدق کند

الف- $\forall A \subseteq \mathcal{F} : P(A) \geq ۰$

ب- $P(\Omega) = ۱$

ج- برای دنباله دوبه‌دو مجزای $\{A_j\}$،

$$P(\bigcup_{j=۱}^{+\infty} A_j) = \sum_{j=۱}^{+\infty} P(A_j) \ .$$

قضایای اساسی

اگر Ω یک فضای نمونه و $P(\cdot)$ یک تابع احتمال باشد آن‌گاه

الف- برای هر پیشامد A و B از Ω ($P(\emptyset) = ۰$)

۱- $P(A^c) = ۱ - P(A)$

۲- $A \subseteq B \Rightarrow P(A) \leq P(B)$

۳- $P(A-B) = P(A) - P(A \cap B) \quad (A - B \equiv A \cap B^c)$

۴- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

۵- اگر $A \cap B = \emptyset$ آن‌گاه: $P(A \cup B) = P(A) + P(B)$

۶- اگر پیشامدهای $A_۱, A_۲, \cdots, A_n$ دوبه‌دو مجزا باشند آن‌گاه

$$P(\bigcup_{j=۱}^{n} A_j) = \sum_{j=۱}^{n} P(A_j)$$

مثال ۱.۲.۲. ۱- *اگر میزان کریستال موجود در یک نمونه سنگ در بازه $\Omega = [۰, a]$ باشد، احتمال این‌که در یک نمونه بدست آمده اندازه کریستال در بازه $[c, d]$ باشد $(۰ < c < d < a)$*

$$P = \frac{d-c}{a} \; ,$$

۲- *درصد شوری یک چاه در بازه $[۰, ۰/۴]$ تعیین شده است، در یک نمونه انتخاب شده احتمال این‌که میزان شوری حداقل ۰/۳ باشد چقدر است:*

$$P = \frac{۰/۴ - ۰/۳}{۰/۴} = \frac{۱}{۴} \text{ یا } P = ۱ - \frac{۰/۳}{۰/۴} = ۱ - \frac{۳}{۴} = \frac{۱}{۴}$$

۳- *اگر A و B دو پیشامد از فضای نمونه Ω باشند به قسمی‌که $P(A) = P(B) = \frac{۱}{۳}$ و $P(A \cap B) = \frac{۱}{۱۰}$ مطلوبست محاسبه:*

$$P(A^c) \; , \; P(A \cup B) \; , \; P(A - B) \; , \; P(A \cup B^c) \; , \; P(A^c \cup B^c)$$

۴- *از یک معدل سنگ طلا، تعداد $n = ۲۰$ نمونه به تصادف انتخاب شده است، اگر ۴ تای آن‌ها، فاقد طلا، ۶ تای آن‌ها مقدار کمی طلا و ۱۰ تای آن‌ها مقدار زیادی طلا داشته باشند، و هرگاه از ۲۰ سنگ موجود ۵ نمونه به تصادف انتخاب شود، مطلوبست محاسبه احتمالات هریک از پیشامدهای زیر:*

الف- هر ۵ نمونه حاوی مقدار زیادی طلا باشند.

ب- *دقیقا ۳ نمونه طلای زیاد داشته باشند.*

ج- *حداقل ۳ نمونه طلای زیاد داشته باشند.*

د- *هر ۵ نمونه حاوی طلای زیاد یا کم باشند.*

توجه ۳.۲.۱. *برای حل مسائلی مشابه مسئله فوق نیاز به استفاده از قوانین شمارشی در مسائلی معروف به مسائل مهره و جعبه به‌صورت زیر می‌باشد.*

جعبه‌ای شامل N مهره از دو نوع (در حالت کلی از k نوع) است، n مهره به تصادف از این جعبه انتخاب می‌شود

الف- نمونه‌برداری با جایگذاری: $n(\Omega) = N^n = \underbrace{N \times N \times \cdots \times N}_{\text{بار } n}$

ب- نمونه‌برداری بدون جایگذاری و ترتیب مورد توجه: $n(\Omega) = (N)_n = \dfrac{N!}{(N-n)!} = P_N^k$

ج- نمونه‌برداری بدون جایگذاری مورد توجه نیست. $n(\Omega) = \dbinom{N}{n} = \dfrac{N!}{n!(N-n)!}$

اگر جعبه شامل دو نوع مهره (نوع I و II) باشد آن‌گاه

$$\begin{cases} N_1 \equiv \text{تعداد مهره‌های نوع I} \\ N_2 \equiv \text{تعداد مهره‌های نوع II} \end{cases}$$

الف) $P(\text{مهره از نوع I } k) = \dfrac{P_{N_1}^k \times P_{N-N_1}^{n-k}}{P_N^n}$

ب) $P(\text{مهره از نوع I } k) = \dfrac{\binom{N_1}{k} \times \binom{N-N_1}{n-k}}{\binom{N}{n}}$.

حل. $N = 20$ و $n = 5$: اگر نمونه‌ها بدون جایگذاری انتخاب شوند:

الف- بدون جایگذاری و بدون توجه به ترتیب، و $P(\text{هر ۵ نمونه دارای طلای زیاد}) = \dfrac{\binom{10}{5}}{\binom{20}{5}}$

ب) $P(\text{دقیقا ۳ نمونه طلای زیاد}) = \dfrac{\binom{10}{3} \times \binom{10}{2}}{\binom{20}{5}}$

ج) $\frac{\binom{10}{5}}{\binom{20}{5}} + \frac{\binom{10}{4}\binom{10}{1}}{\binom{20}{5}} + \frac{\binom{10}{3} \times \binom{10}{2}}{\binom{20}{5}} = P(\text{حداقل ۳ نمونه طلای زیاد})$

- اگر ۳ نمونه سنگ انتخاب شود احتمال این‌که اولی طلای زیاد، دومی فاقد طلا و سومی طلای کم داشته باشد:

$$P = \frac{P_{10}^{1} \times P_{4}^{1} \times P_{6}^{1}}{P_{20}^{3}} \ .$$

احتمال شرطی

احتمال شرطی یک مفهوم مفید و پرکاربرد در آمار است و در مسائل کاربردی نیز موارد زیادی پیش می‌آید که با داشتن اطلاعات جزئی در مورد نتیجه آزمایش می‌خواهیم احتمال رخ دادن یک پیشامد مورد مطالعه را بدست آوریم:

فرض کنید A و B دو پیشامد از فضای نمونه Ω باشند و $P(A) > 0$ و $P(B) > 0$ آن‌گاه:

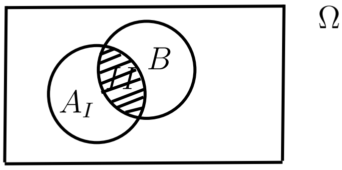
$$P(\text{رخ دادن } A \text{ به شرط این‌که } B \text{ قبلا رخ داده}) = P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (1)$$

$$P(\text{رخ دادن } B \text{ به شرط این‌که } A \text{ قبلا رخ داده}) = P(B|A) = \frac{P(A \cap B)}{P(A)} \ . \qquad (2)$$

قانون ضرب در احتمال

$$(1) \Rightarrow P(A \cap B) = P(A|B) \cdot P(B)$$

$$(2) \Rightarrow P(A \cap B) = P(B|A) \cdot P(A) \ .$$



قانون احتمال کل

۱)

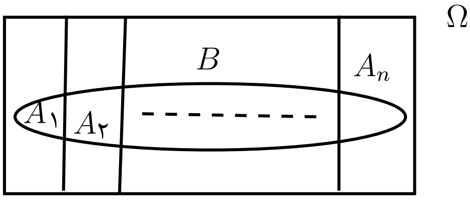
$$P(A) = \underbrace{P(A \cap B)}_{II} + \underbrace{P(A \cap B^c}_{I} = P(B).P(A|B) + P(B^c) \cdot P(A|B^c) \qquad (3)$$

اگر $A_1, A_2, \cdots, A_n$ پیشامدهای دوبه‌دو مجزا از Ω باشند به قسمی که $\Omega = \bigcup_{j=1}^{n} A_j$ در این‌صورت می‌گویند فضای Ω توسط پیشامدهای $A_1, \cdots, A_n$ افراز شده است:

برای هر پیشامد دلخواه دیگر B داریم:

۲)

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i) \ , \ P(A_i) > 0 \ , \qquad (4)$$



روابط (۳) و (۴) به قانون احتمال کل (مرکب) معروف هستند.

قانون بیز (Bayas): تحت شرایط بالا برای محاسبه هر پیشامد A_i زمانی که B رخ داده باشد از رابطه زیر که معروف به فرمول بیز است استفاده می‌شود.

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{P(B)} = \frac{P(A_k) \cdot P(B|A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)} \ , \ k = 1, 2, \cdots, n \ . \qquad (5)$$

مثال ۴.۲.۱. *تعداد ۱۵۰ نمونه سنگ مس بر اساس ناحیه برداشت شده و میزان عیار مس در جدول زیر خلاصه شده‌اند:*

سنگ	A	B	C	
عیار کم Low	۳۰	۱۹	۸	۵۷
عیار متوسط Mean	۱۰	۲۰	۱۷	۴۷
عیار زیاد: High	۱۵	۱۸	۱۳	۴۶
	۵۵	۵۷	۳۸	

نمونه‌ای به تصادف انتخاب شد.

الف- احتمال این‌که دارای عیار بالا باشد چقدر است؟

$$P(H) = P(A) \cdot P(H\big|A) + P(B) \cdot P(H\big|B) + P(C) \cdot P(H\big|C) = \frac{۴۶}{۱۵۰}$$

ب- نمونه بدست آمده دارای عیار بالا بوده است احتمال این‌که از ناحیه θ برداشت شده باشد چقدر است؟

$$P(A\big|H) = \frac{P(A) \cdot P(H\big|A)}{P(H)} = \frac{\frac{۵۵}{۱۵۰} \times \frac{۱۵}{۵۵}}{\frac{۴۶}{۱۵۰}} = \frac{۱۵}{۴۶} \, .$$

پیشامدهای مستقل

دو پیشامد A و B را از نظر آماری مستقل از هم نامند اگر رخ دادن هریک تأثیری بر رخ دادن دیگری نداشته باشد یعنی:

$$\begin{cases} P(A\big|B) = P(A) \\ P(B\big|A) = P(B) \end{cases} \Longrightarrow P(A \cap B) = P(A) \cdot P(B) \, .$$

n- پیشامد $A_۱, A_۲, \cdots, A_n$ را توأما مستقل نامند اگر برای هر m تا احتمال اشتراک برابر حاصل‌ضرب احتمالات آن‌ها باشد.

توجه ۵.۲.۱. *اگر A و B مستقل باشند آن‌گاه A و B^c، A^c و B، و A^c و B^c نیز مستقل از هم*

هستند و این نکته قابل تعمیم است:

$$\text{متمم پیشامد } A = A^c = \Omega - A \,.$$

فصل ۲

متغیر تصادفی، امیدریاضی، واریانس، کوواریانس

متغیر تصادفی متغیر تصادفی تابعی از Ω به R است که نتایج یک آزمایش تصادفی را به‌صورت کمی بیان می‌کند.

$$X : \underset{w\longmapsto X(w)=x}{\Omega \to R}$$

معمولا متغیرهای تصادفی را با حروف بزرگ X و Y و Z نمایش می‌دهند.

تکیه‌گاه برد هر متغیر تصادفی را تکیه‌گاه آن نامند.

مثال ۲.۱.۱. ۱- *میزان کاهش یا افزایش سطح آب پشت سد در هر روز*

۲- *تعداد زمین لرزه‌هایی که هر سال در یک ناحیه رخ می‌دهد.*

۳- *تعداد چاه‌های حفر شده در بین n چاه که به آب می‌رسد.*

متغیر تصادفی گسسته اگر برد یک متغیر تصادفی متناهی یا نامتناهی شمارش‌پذیر باشد آن‌را گسسته نامند.

- متغیرهای تصادفی از طریق اندازه‌گیری بدست می‌آیند. مثال ۱ یا ۴- درجه خلوص یک سنگ طلا

تابع احتمال اگر X یک متغیر تصادفی گسسته با تکیه‌گاه: $S_X = \{۱, ۲, ۳, \cdots, \}$ باشد آنگاه تابع $P(x) = P[X = x]\ ,\ x \in S_X$ را تابع احتمال X نامند، این تابع در شرایط زیر صدق می‌کند:

الف- $P(n) \geq ۰$

ب- $\sum_{x \in S_X} P(x) = ۱$.

تابع توزیع برای هر متغیر تصادفی X، تابع $\forall x \in R\ ,\ F(x) = P[X \leq x]$ که دارای ویژگی‌های زیر است را تابع توزیع متغیر تصادفی X نامند:

$$F : R \to [۰, ۱]$$

الف- $۰ \leq F(x) \leq ۱$

ب- $F(x)$ غیر نزولی است.

ج- $F(x)$ از راست پیوسته است.

تابع چگالی اگر X یک متغیر تصادفی پیوسته با تابع توزیع $F(x)$ باشد آنگاه $f(x) = F'(x)$ را تابع چگالی متغیر تصادفی X نامند، و علاوه بر این:

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad , \qquad f(x) = \frac{dF(x)}{dx} = F'(x)\ .$$

مثال ۲.۱.۲. *تعداد ۶ نمونه سنگ طلا و ۴ نمونه سنگ مس در جعبه‌ای قرار دارد، ۳ نمونه به تصادف از جعبه برداشت می‌شود. اگر X تعداد نمونه سنگ‌های طلاها در بین ۳ نمونه انتخاب شده باشد مطلوبست تعیین تابع احتمال X:*

X	۰	۱	۲	۳
$P(x)$	$\frac{\binom{۴}{۳}}{\binom{۱۰}{۳}}$	$\frac{\binom{۶}{۲} \times \binom{۴}{۲}}{\binom{۱۰}{۳}}$	$\frac{\binom{۶}{۳}\binom{۴}{۱}}{\binom{۱۰}{۳}}$	$\frac{\binom{۶}{۳}}{\binom{۱۰}{۳}}$

$$P[X=0]=\frac{\binom{4}{3}}{\binom{10}{3}}\ ,P(X=1)=\frac{\binom{6}{1}\binom{4}{2}}{\binom{10}{3}}$$
$$P[X=2]=\frac{\binom{6}{2}\binom{4}{1}}{\binom{10}{3}}\ ,\ P(X=3)=\frac{\binom{6}{3}}{\binom{10}{3}}$$

امید ریاضی امید ریاضی یا مقدار متوسط متغیر تصادفی X، کمیتی است که مرکز ثقل توزیع X را تعیین می‌کند با نماد $E(X)$ نشان داده می‌شود و به‌صورت زیر تعریف می‌شود.

$$E(X)=\begin{cases}\sum_{x\in S_X} xP(x) & X \text{ گسسته} \\ \int xf(x)dx & X \text{ پیوسته .}\end{cases}$$

ویژگی‌ها

۱- اگر $Y=aX+b$ آن‌گاه $E(Y)=aEX+b$ از این ویژگی برای داده‌های که تغییر مقیاس و تبدیل می‌شوند استفاده می‌شود.

۲- اگر $X_1,\cdots,X_n$ متغیرهای تصادفی و $a_1,\cdots,a_n$ اعداد حقیقی باشند آن‌گاه

$$E\big(\sum_{i=1}^{n} a_iX_i\big)=\sum_{i=1}^{n} a_iEX_i\ .$$

واریانس واریانس متغیر تصادفی X کمیتی است که میزان تراکم یا عدم تراکم توزیع حول مرکز ثقل $(E(X))$ را محاسبه و اشکال در متن می‌نماید با نماد $\mathrm{Var}(X)$ نمایش داده شده به‌صورت زیر تعریف می‌شود:

$$\sigma_X^2=\mathrm{Var}(X)=E(X-EX)^2=EX^2-E^2X$$

که در آن:

$$E(X^2)=\begin{cases}\sum_{x} x^2P(x) & X \text{ گسسته} \\ \int x^2f(x)dx & X \text{ پیوسته .}\end{cases}$$

ویژگی‌ها

۱- $\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$

۲- اگر $X_1, X_2, \cdots, X_n$ متغیرهای تصادفی مستقل و $a_1, \cdots, a_n$ ثابت‌های حقیقی باشند آن‌گاه:

$$\mathrm{Var}\Big(\sum_{i=1}^{n} a_i X_i\Big) = \sum_{i=1}^{n} a_i^2 \cdot \mathrm{Var}(X_i) \ .$$

انحراف معیار جذر واریانس است؛

$$\begin{cases} \sigma = \sqrt{\mathrm{Var}(X)} \\ \sigma_Y = \sqrt{a^2 \cdot \mathrm{Var}(X)} = |a| \cdot \sigma_X \quad \text{آن‌گاه } Y = aX+b \text{ اگر} \end{cases}$$

مثال ۳.۱.۲. *در یک منطقه تنها ۲۰٪ چاه‌های اکتشافی به آب می‌رسد. اگر ۵ چاه گمانه در این منطقه حفر شود و X تعداد چاه‌هایی باشد که به آب می‌رسند.* $S_X = \{0, 1, 2, 3, 4, 5\}$.

الف- انتظار می‌رود چند چاه به آب برسد. $P_X(k) = \binom{5}{k}(0.2)^K(0.8)^{5-k}$

ب- احتمال این‌که حداقل یک چاه به آب برسد چقدر است:

X	0	1	2	3	4	5
$P_X(k)$	$(0.8)^5$	$5 \times 0.2 \times (0.8)^4 = (0.8)^4$	$\binom{5}{2} \cdot (0.2)^2 \times (0.8)^3$	$\binom{5}{3}(0.2)^3 \cdot (0.8)^2$	$\binom{5}{4}(0.2)^4 \times 0.8$	$(0.2)^5$
$R \cdot P(k)$	0	$(0.8)^4$	$2 \times \binom{5}{2}(0.2)^2(0.8)^3$	$3 \times \binom{5}{3}(0.2)^3(0.8)^2$	$4 \times \binom{5}{4}(0.2)^4 \cdot 0.8$	$5 \times (0.2)^5$

$$E(X) = \sum_{k=0}^{5} kP(k) = 5 \times 0.2 = 1 \ , \ P(X \geq 1) = 1 - P(X < 1) = 1 - P(0) = 1 - (0.8)^5$$

کوواریانس

فرض کنید X و Y دو متغیر تصادفی باشند تابع توزیع توأم و تابع احتمال توأم X و Y به‌صورت زیر تعریف می‌شوند:

$$\forall x, y \in R : \ P(x,y) = P[X = x, Y = y]$$

تابع احتمال توأم

۱) $P(x,y) \geq 0 \ , \ \forall (x,y) \in S_X \times S_X$

۲) $\sum_{y \in S_Y} \sum_{x \in S_X} P(x,y) = 1$

۳) $P_X(n) = P(X = x) = \sum_{y \in S_Y} P(x,y)$

۴) $P_Y(y) = \sum_{X \in x} P(x,y)$

تابع توزیع توأم

$$F(x,y) = P[X \leq x \ , \ Y \leq y]\cdot \ , \ F_1(x) = F(x, +\infty) \ , \ F_2(y) = F(+\infty, y) \ .$$

اگر X و Y پیوسته باشند آن‌گاه تابع نامنفی $f(x,y)$ وجود دارد به قسمی که

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(t,s) dt ds \ , \ f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

$$1) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dx dy = 1 \ , \ f_X(x) = \int_{-\infty}^{+\infty} f(x,y) dy \ , \ f_X(y) = \int_{-\infty}^{+\infty} f(x,y) dx \ .$$

تعریف ۴.۱.۲. *دو متغیر تصادفی X و Y را مستقل آماری نامند اگر:*

$$F(x,y) = F_1(x) \cdot F_2(y) \ , \qquad \forall x, y \in k \ . \qquad (1)$$

در حالت پیوسته: $\quad (1) \iff f(x,y) = f_1(x) \cdot f_2(y) \quad \forall x, y \in R$

در حالت گسسته $\quad (1) \iff P(x,y) = P_1(x) \cdot P_2(y) \quad \forall x, y \in R \ .$

نکته ۵.۱.۲. *اگر $h : R^2 \to R$ و X و Y دو متغیر تصادفی باشند آن‌گاه $h(X,Y)$ یک متغیر*

تصادفی است و امید ریاضی آن به‌صورت زیر محاسبه می‌شود

$$E\big(h(X,Y)\big)=\begin{cases}\sum_y\sum_x h(x,y)\cdot P(x,y) & \text{در حالت گسسته}\\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)f(x,y)dxdy & \text{در حالت پیوسته .}\end{cases}$$

به ویژه اگر $h(X,Y)=X\cdot Y$ *باشد آن‌گاه*

$$E(X\cdot Y)=\begin{cases}\sum_y\sum_x xyP(x,y) & X \text{ و } Y \text{ گسسته}\\ \int\int xyf(x,y)dxdy & X \text{ و } Y \text{ پیوسته .}\end{cases}$$

تعریف ۶.۱.۲. *کواریانس دو متغیر تصادفی X و Y کمیتی است که نوع همبستگی بین دو متغیر X و Y را بیان می‌کند با نماد $\mathrm{Cov}(X,Y)$ نشان داده شده و به‌صورت زیر تعریف می‌شود:*

$$\begin{aligned}\mathrm{Cov}(X,Y)&=E[X-E(X)]\cdot[Y-EY]\\&=E(X\cdot Y)-E(X)\cdot E(Y)\ .\end{aligned}$$

۱- *مقادیر مثبت کواریانس بیان می‌کند که X و Y هماهنگ هستند و با افزایش یکی، دیگری نیز افزایش می‌یابد.*

۲- *مقادیر منفی کوواریانس بیان می‌کند که X و Y ناهماهنگ هستند و با افزایش یکی، دیگری کاهش می‌یابد.*

۳- *مقدار صفر بیان می‌کند که X و Y ناهمبسته هستند و ارتباط خطی با یکدیگر ندارند ولی ممکن است به‌صورت غیر خطی در ارتباط باشند.*

ضریب همبستگی کمیتی است که میزان و نوع ارتباط (هماهنگی) بین X و Y را اندازه‌گیری و بیان می‌نماید، و به‌صورت زیر تعریف می‌شود

فرض کنید X دارای $\mu_1=EX$ و $\sigma_1^2\mathrm{Var}(X)$ و Y دارای $\mu_2=EY$ و $\sigma_2^2=\mathrm{Var}(Y)$ باشد آن‌گاه $Z_1=\dfrac{X-\mu_1}{\sigma_1}$ و $Z_2=\dfrac{X-\mu_2}{\sigma_2}$ به ترتیب متغیرهای استاندارد شده X و Y هستند.

$$\rho(X,Y)=\mathrm{Cov}\Big(\frac{X-\mu_1}{\sigma_1}\ ,\ \frac{Y-\mu_2}{\sigma_2}\Big)=\frac{\mathrm{Cov}(X,Y)}{\sigma_1\cdot\sigma_2}\ .$$

* متغیرهای تصادفی استاندارد $Z_۱$ و $Z_۲$ دارای این ویژگی هستند که:

$$\begin{cases} EZ_۱ = EZ_۲ = ۰ \\ \mathrm{Var}(Z_۱) = \mathrm{Var}(Z_۲) = ۱ \end{cases}$$

ویژگی‌های ρ

۱- $-۱ \leq \rho \leq ۱$

۲- $\rho = ۰$ اگر و تنها اگر X و Y رابطه خطی نداشته باشند.

۳- $\rho = ۱ \Longleftrightarrow Y = aX + b\,, \quad a > ۰$

۴- $\rho = -۱ \Longleftrightarrow Y = aX + b\,, \quad a < ۰\ .$

۵- مقادیر مثبت بیان کننده وجود یک رابطه خطی بین X و Y است. با شیب مثبت.

۶- مقادیر منفی بیان کننده وجود یک رابطه خطی بین X و Y است با شیب منفی.

کمیت ρ یکی از ابزار تحلیل رگرسیون در تحلیل داده‌های آماری و مدل‌سازی می‌باشد (در بخش آمار و تحلیل رگرسیون با جزئیات کامل در مورد ρ بحث خواهد شد).

فصل ۳

توزیع‌های آماری

برخی از متغیر تصادفی از توزیع‌های خاصی پیروی می‌کنند، در این قسمت توزیع‌های خاص قابل کاربرد در عمق زمین را معرفی و مشخصه‌ها و ویژگی‌های آن‌ها ارائه می‌شود.

۱.۳ توزیع یکنواخت (Uniform-Distribution)

الف- یکنواخت گسسته: متغیر تصادفی X دارای توزیع یکنواخت گسسته است اگر:

$$P_X(k) = \begin{cases} \dfrac{۱}{N}\,, & k = ۱, ۲, \cdots, N \\ ۰ & o \cdot w \cdot \end{cases} \qquad X \sim U\{۱, ۲, \cdots, N\}$$

یعنی تمام مقادیر هم‌شانس باشند: در این‌صورت توزیع مشخصه‌های امید ریاضی و واریانس برابرند با: $\mathrm{Var}(X) = \dfrac{N^{۲} - ۱}{۱۲} \quad , \quad E(X) = \dfrac{۱ + N}{۲}$

ب- یکنواخت پیوسته:

$$f_X(x) = \begin{cases} \frac{۱}{b-a} & a < x < b \\ ۰ & o \cdot w \end{cases} \qquad X \sim U(a,b)$$

$$F(x) = \begin{cases} ۰ & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ ۱ & x \geq b \end{cases} \quad , E(X) = \frac{a+b}{۲} \ , \ \mathrm{Var}(X) = \frac{(a-b)^{۲}}{۱۲}$$

توجه ۱.۱.۳. *یکی از کاربردهای توزیع یکنواخت پیوسته، تولید اعداد تصادفی توسط کامپیوتر از هر توزیع پیوسته می‌باشد: زیرا اگر X یک متغیر تصادفی پیوسته با توزیع معلوم $F(x)$ باشد آن‌گاه $Y = F(X) \sim U(۰,۱)$. بنابراین اگر $u_۱ \in (۰,۱)$ یک عدد تصادفی تولید شده از بازه $(۰,۱)$ باشد آن‌گاه $x = F^{-۱}(u_۱)$ یک مشاهده شبیه‌سازی شده از متغیر تصادفی X با توزیع F خواهد بود.*

مثال ۲.۱.۳. *فرض کنید میزان طلای موجود در یک نوع سنگ طلا حداقل a و حداکثر b باشد و بدانیم توزیع میزان طلاها کاملا یکنواخت است، در این‌صورت $X \sim U(a,b)$ و برای محاسبه نسبت نمونه‌های با حداقل c مقدار طلا برابر است با:*

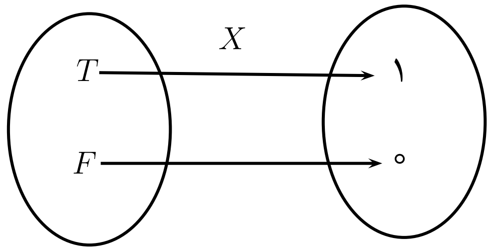
$$P[X > c] = \frac{b-c}{b-a} \ , \quad a < c < b \ .$$

۲.۳ توزیع برنولی Bernoulli Distribution

آزمایش تصادفی برنولی: هر آزمایش تصادفی که تنها دو نتیجه ممکن باشد را آزمایش برنولی نامند- یکی از نتایج را موفقیت یا پیروزی (T) و دیگری را شکست (F) نامیده و قرار می‌دهیم: $P[\{T\}] = p$ و $q = ۱ - p = P[\{F\}]$.

$$\text{فضای نمونه} = \{T, F\} \ ; \quad X : \underset{w \longmapsto X(w)}{\Omega \to R} \ , \begin{cases} X(T) = ۱ \ , \\ X(F) = ۰ \end{cases}$$

متغیر تصادفی X را متغیر تصادفی برنولی



نامند: $X(w) = \begin{cases} ۱ & w \equiv T \\ ۰ & w \equiv F \end{cases}$.

توزیع متناظر با متغیر تصادفی X را که به‌صورت زیر تعیین می‌شود توزیع برنولی با احتمال موفقیت P نامند.

X	۰	۱
$P_X(K)$	$P(۰) = q$	$P(۱) = p$

تابع احتمال X به‌صورت زیر قابل نمایش است:

$$P_X(k) = p^k(۱-p)^{۱-k}\ , \quad k = ۰, ۱, \quad X \sim Ber(p)\ .$$

$$EX = ۱ \times P(۱) + o \cdot P(۰) = p\ , \quad E(X^۲) = ۱^۲ \times P(۱) + o \cdot P(۰) = p$$

$$\mathrm{Var}(X) = EX^۲ - E^۲X = p - p^۲ = P(۱-p) = pq\ .$$

این توزیع سازنده توزیع‌های مهم آماری است $\begin{cases} X \sim Bar(p) \\ EX = p \\ Vv(X) = Pq \end{cases}$ که در ادامه ارائه می‌شود. چون در آمار با تعداد n–نمونه (متغیرهای $X_۱, X_۲, \cdots, X_n$) مواجهه هستیم بنابراین در n آزمایش مستقل برنولی با احتمال P، تعداد کل موفقیت‌ها یک متغیر تصادفی مفید و پرکاربرد در آمار می‌باشد.

چند نمونه از آزمایش‌های برنولی

۱- به آب رسیدن یا نرسیدن یک چاه گمانه،

۲- سالم بودن و معیوب بودن هر نمونه در یک خط تولید.

۳- زمین لرزه با شدت بیش از ۵ ریشتر و کمتر از آن،

۴- مهیب بودن و خفیف بودن هر سیلاب.

۵- میزان مس بالای ۸۰٪ و کمتر از آن در یک نمونه سنگ مس.

۳.۳ توزیع دو جمله‌ای (Binomical-Distribution)

تعداد موفقیت‌ها در n–آزمایش مستقل برنولی با احتمال موفقیت P یک متغیر تصادفی معروف به متغیر تصادفی دو جمله‌ای با تابع احتمال زیر است:

$$P[X=K]=\binom{n}{k}p^k(1-p)^{n-k}\ ,\ k=0,1,2,\cdots,n$$

نماد: $X\sim Bin(n,p)$

امید ریاضی: $E(X)=nP$ واریانس: $\mathrm{Var}=np\cdot q$.

مثال ۱.۳.۳. *تعداد ۱۰ نمونه سنگ برای بررسی وجود فسیل به تصادف انتخاب و آزمایش می‌شود اگر تنها ۶۰ درصد این نمونه‌ها شانس وجود فسیل باشد.*

الف- احتمال وجود فسیل در حداقل ۲ تای آن‌ها چقدر است؟

ب- انتظار می‌رود در چند نمونه فسیل باشد: $X\sim Bin(10,0/6)$

$$P[X=k]=\binom{10}{k}(0/6)^k(0/4)^{10-k}\ ,k=0,1,2,\cdots,10\ .$$

ج- *در مثال، آزمایش به این‌صورت اجرا می‌شود که نمونه‌ها به تصادف انتخاب شده و تا پیدا کردن اولین سنگ فسیلی آزمایش می‌شود، به طور متوسط چند نمونه بایستی برداشت شود؟*

۴.۳ توزیع هندسی (Geometry-Disk)

در یک دنباله از آزمایش‌های مستقل برنولی با احتمال موفقیت P، تعداد آزمایش‌های لازم برای کسب اولین موفقیت (X) یا تعداد آزمایش‌های لازم قبل از اولین موفقیت (تعداد شکست‌ها قبل از اولین موفقیت) (Y)، متغیرهای تصادفی با توزیع‌های زیر و هردو معروف به توزیع هندسی هستند:

$$P[X=K]=Pq^{k-1}\ ,k=1,2,\cdots, \qquad X\sim G(p)\ . \tag{1}$$
$$E(X)=\frac{1}{P}\ ,\quad \mathrm{Var}(X)=\frac{1}{p}\Big(\frac{1}{p}-1\Big)\ .$$
$$P[Y=R]=Pq^{K}\ ,R=0,1,2,\cdots \qquad Y\sim G(p)\ . \tag{2}$$
$$E(X)=\frac{q}{p}\ ,\qquad \mathrm{Var}(X)=$$

مثال ۳.۴.۱. ۱- *تعداد زمین لرزه‌ها قبل از رخ دادن اولین زمین لرزه شدید در یک ناحیه:*

فرض کنید $p=P($*رخ دادن یک زمین لرزه شدید*$)$ ، $q=P($*رخ دادن یک زمین لرزه خفیف*$)$

۲- *تعداد چاه‌های حفاری شده برای رسیدن به اولین چاهی که به آب می‌رسد.*

$$p=P(\text{به آب رسیدن یک چاه})$$

۳- *تعداد سیلاب‌هایی که قبل از اولین سیلاب مهیب رخ می‌دهد.*

$$p=P[\text{رخ دادن یک سیلاب مهیب}]\ .$$

۵.۳ توزیع پواسون (Poisson-Dist)

توزیع پواسون که یکی از مفیدترین و پرکاربردترین توزیع‌های آماری در علوم ارتباطات و علوم زمین آمار می‌باشد از دیدگاه تئوری به‌صورت زیر بدست می‌آید:

تقریب دو جمله‌ای به پواسون

اگر X یک متغیر تصادفی دو جمله‌ای ($X \sim Bin(n,p)$) باشد به قسمی که $p \to ۰$ و $n \to \infty$ طوری‌که $n\,p \underset{n\to\infty}{\to} \lambda$ آن‌گاه:

$$\lim_{n\to\infty} P[X=K] = \lim_{n\to\infty} \binom{n}{k} p^k q^{n-k} = \frac{e^{-\lambda}\lambda^k}{k!}\ , \quad k = ۰, ۱, ۲, \cdots$$

در مسائل کاربردی استفاده از تقریب فوق زمانی مجاز است که ($p \leq ۰/۱$ و $np \leq ۵$). تابع $\begin{cases} p_k = \dfrac{e^{-\lambda}\lambda^k}{k!}\ ,\ k = ۰, ۱, \cdots \\ \lambda > ۰ \end{cases}$ یک تابع احتمال است، این مسئله توسط ریاضی‌دان فرانسوی دنیس-پواسون اثبات و توزیع متناظر نیز با نام وی ثبت شده است.نماد:

$$\begin{cases} X \sim Po(\lambda) \\ E(X) = \lambda \\ \mathrm{Var}(X) = \lambda \end{cases}$$

نکته ۱.۵.۳. *تنها توزیعی از بین تمام توزیع‌های شناخته شده که میانگین و واریانس برابر هستند.*

درکاربردها، پدیده‌هایی که از توزیع پواسون پیروی می‌کنند بایستی در اصول پواسون که به‌صورت زیر معرفی می‌شود صدق کنند:

اصول پواسون

۱- احتمال وقوع یک واقعه در بازه زمانی کوتاه با طول بازه متناسب باشد.

۲- احتمال وقوع بیش از یک واقعه در بازه زمانی کوتاه تقریبا صفر باشد.

۳- وقوع یا عدم وقوع اتفاقات در فواصل مجزا مستقل از هم باشند.

در نتیجه هر آزمایش تصادفی در سه اصل فوق صدق کند یک آزمایش تصادفی پواسون نامیده می‌شود،

تعریف ۲.۵.۳. *تعداد وقایع یا اتفاقات که در واحد زمان رخ می‌دهد یک پدیده پواسون و متغیر تصادفی متناظر را متغیر تصادفی پواسون نامند. بدیهی است که λ بیان کننده متوسط تعداد وقایع در واحد زمان می‌باشد.*

مثال ۳.۵.۳. ۱- *تعداد سیلاب‌های مهیب در ناحیه‌ای از توزیع پواسون با متوسط (نرخ) $\lambda = ۲$. در هر دوره ۵۰ ساله پیروی می‌کند:*

الف- $P(X = ۲)$

ب- $P(X \geq ۲)$

ج- احتمال رخ دادن حداقل ۱ سیلاب مهیب در ۲۰ سال چقدر است؟

د- احتمال وقوع دقیقا یک سیلاب مهیب در سه دوره از ۵ دوره ۵۰ ساله آینده چقدر است.

۲- *تعداد زمین لرزه‌ها در یک ناحیه از فرآیند پواسون با نرخ ۳ در هر ۱۰ سال پیروی می‌کند، ۴۰ درصد زمین لرزه‌ها شدید و دارای قدرت بیش از ۵ ریشتر هستند*

الف- احتمال اینکه از زمان رخ دادن اولین زمین لرزه شدید بیشتر از ۵ سال بگذرد چقدر است؟

ب- احتمال رخ دادن دقیقا دو زمین لرزه شده در یک دوره ۱۰ ساله چقدر است؟

فرآیند پواسون

فرض کنید برای هر $t \geq ۰$، $N(t)$ بیان کننده تعداد وقایع یک پدیده پواسون باشد که تا لحظه t رخ می‌دهد و نرخ وقایع λ (متوسط تعداد وقایع در واحد زمان) در این‌صورت برای هر $t \geq ۰$، $N(t)$ یک متغیر تصادفی پواسون با متوسط λt می‌باشد: $P[N(t) = k] = \dfrac{e^{-\lambda t}(\lambda t)^k}{k!}\ ,\ k = ۰, ۱, ۲, \cdots$

در این‌صورت رده $\{N(t), t \geq ۰\}$ را یک فرآیند تصادفی پواسون با نرخ λ نامند.
مشخصه‌های $N(t)$:

$$1)\ \ EN(t) = \mathrm{Var}\big(N(t)\big) = \lambda t$$

$$2)\ \ \mathrm{Cov}\big(N(t), N(s)\big) = \lambda \cdot \min\{t, s\}\ ,\quad \forall t, s \geq 0\ .$$

ویژگی‌های $N(t)$

۱- مجموع دو فرآیند پواسون مستقل یک فرآیند پواسون است:

$$\begin{cases} N_۱(t) \sim P(\lambda, t) \\ N_۲(t) \sim P(\lambda_۲ t) \end{cases} \qquad N_۱(t) \perp N_۲(t) \Rightarrow N_۱(t) + N_۲(t) \sim P\big((\lambda_۱ + \lambda_۲)t\big)\ .$$

۲- برای هر $۰ \leq t \leq s$:

$$P[N(t) = K \big| N(s) = n] = \binom{n}{k} \left(\frac{t}{s}\right)^k \left(۱ - \frac{t}{s}\right)^{n-k}\ ,\quad k = ۰, ۱, \cdots, n$$

یعنی: $۰ < t < s \quad N(t) \big| N(s) = n \sim Bin\big(n, \frac{t}{s}\big)$

۳- تفکیک یک فرآیند پواسون به دو فرآیند پواسون مستقل:

یعنی اگر $N(t) \sim P(\lambda t)$ و هر واقعه فرآیند $N(t)$ با احتمال P گزارش و با احتمال q گزارش نشود و این گزارش‌ها و عدم گزارش‌ها مستقل از هم رخ دهد آن‌گاه

$$N(t) = N_۱(t) + N_۲(t)$$

که در آن:

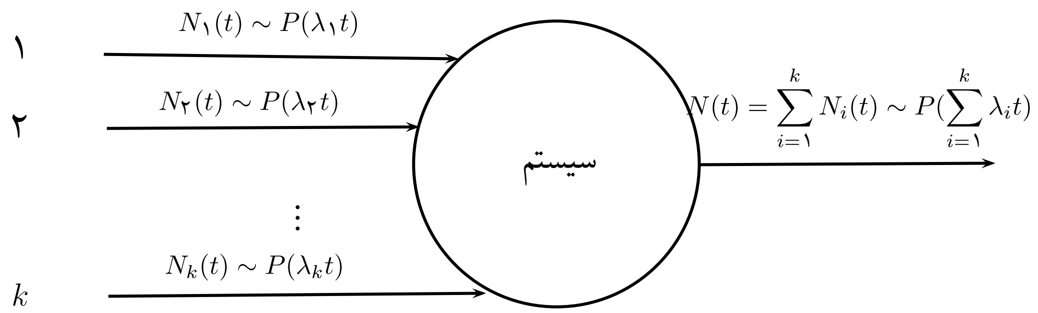
$N_۱(t) \sim P(\lambda p t) \equiv$ تعداد وقایع گزارش شده تا لحظه t

$N_۲(t) \sim P(\lambda q t) \equiv$ تعداد وقایع گزارش نشده تا لحظه t

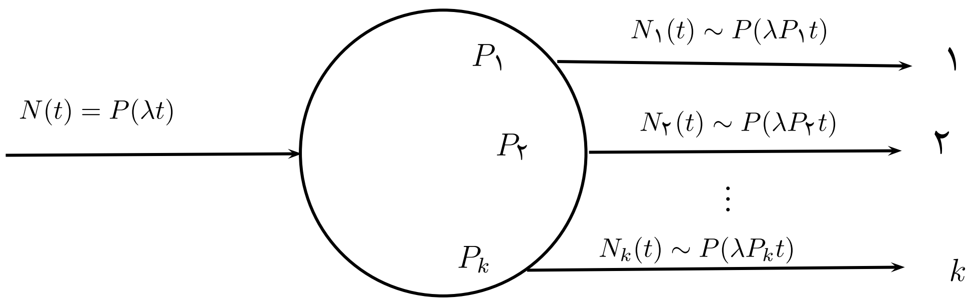
مثال ۴.۵.۳. *مثال ۲ مربوط به زمین‌لرزه‌ها را در نظر بگیرید، که در آن نسبت زمین‌لرزه‌های شدید ۰/۴ است. اگر $N_۱(t)$ را تعداد زمین‌لرزه‌های شدید و $N_۲(t)$ را تعداد زمین‌لرزه‌های خفیف در نظر بگیریم در این صورت:*

$$N_۱(t) \sim P(۰/۴\lambda t), \qquad N_۲(t) \sim P(۰/۶\lambda t)$$

۱. *ترکیب پوآسون‌های مستقل*



۲. *تجزیه یک فرآیند به فرآیندهای پوآسون مستقل از هم*



که در آن p_i نسبت خروجی از مسیر i و $i = ۱, ۲, \ldots, ۱۲$ است و $\sum_{i=۱}^{k} p_i = ۱$.

مثال ۵.۵.۳. *در مثال مربوط به تعداد سیلاب‌ها، اگر p نسبت تعداد سیلاب‌های مهیب باشد آن‌گاه*

$N_۱(t) \sim P(\lambda pt) =$ تعداد سیلاب‌های مهیب

و $N_۲(t) \sim P(\lambda qt) =$ تعداد سیلاب‌های ضعیف.

در صورتی‌که بدانیم $N(t) \sim P(\lambda t)$ و فرض استقلال آماری بین مهیب یا ضعیف بودن سیلاب‌ها برقرار باشد.

مثال ۶.۵.۳. *در مسئله تعداد تصادفات یک تقاطع (بزرگراه) اگر p نسبت تصادفات شدید منجر به مرگ و q نسبت تصادفات جزئی باشد، آنگاه*

$N_1(t) \sim P(\lambda pt) =$ *تعداد تصادفات شدید تا لحظه* t

$N_2(t) \sim P(\lambda qt) =$ *تعداد تصادفات خفیف تا لحظه* t

$N(t) \sim P(\lambda t) =$ *تعداد کل تصادفات.*

توجه ۷.۵.۳. *معمولا، تعداد تصادفات، تعداد سیلاب‌ها و تعداد زمین‌لرزه‌های پدیده‌های پوآسون هستند و در اصول پوآسون صدق می‌کنند.*

توجه ۸.۵.۳. *در ویژگی تفکیک یک فرآیند پوآسون $N(t)$ به دو پوآسون مستقل $N_1(t)$ و $N_2(t)$ نتایج زیر برقرار است:*

۱.

$$N_1(t)|N(t) = n \sim Bin(n, p)$$

۲.

$$N_2(t)|N(t) = n \sim Bin(n, q)$$

توزیع نمایی

متغیر تصادفی X دارای توزیع نمایی است اگر تابع چگالی آن به صورت $f(x) = \lambda e^{-\lambda x},\ \ x \geqslant 0,\ \ \lambda > 0$ باشد:

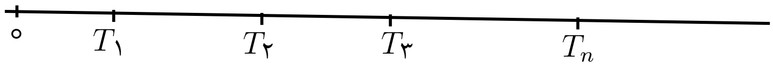
نماد: $X \sim Exp(\lambda)$

میانگین و واریانس: $E(X) = \frac{1}{\lambda},\quad EX^2 = \frac{1}{\lambda^2}$

توزیع نمایی یک ارتباط مستقیم با فرآیند پوآسون دارد:

قضیه ۹.۵.۳. *فرآیند شمارشی $\{N(t)\}$ یک فرآیند پوآسون با نرخ λ است اگر و تنها اگر فواصل زمانی بین وقایع متوالی آن متغیرهای تصادفی مستقل نمایی با متوسط $\frac{1}{\lambda}$ باشند:*

محور زمان:



T_i: زمان رخداد i امین واقعه فرآیند $N(t)$ و $N(t)$ تعداد وقایع تا لحظه t را نشان می‌دهد:

اگر $i \geqslant 1$,$X_i = T_i - T_{i-1}$ و $T_0 = 0$ آن‌گاه

$$X_i \overset{iid}{\sim} Exp(\lambda) \Leftrightarrow N(t) \sim P_0(\lambda t).$$

$$(iid : Identically\ Independent\ Distribution,)$$

در تاثیر فوق T_iها نیز متغیرهای تصادفی هستند که دارای توزیع خاص هستند به نام توزیع گاما:

$$f_{T_i}(x) = \frac{\lambda^i x^{i-1} e^{-\lambda x}}{(i-1)!}, \quad T_i \sim \Gamma(i, \lambda)$$

$$T_i = \sum_{j=1}^{i} X_i, \qquad X_j \overset{iid}{\sim} Exp(\lambda)$$

یعنی مجموع متغیرهای تصادفی مستقل نمایی، یک متغیر تصادفی با توزیع گاما است.

توزیع نرمال

یکی از پرکاربردترین توزیع‌های آماری در آمار و علوم مختلف توزیع نرمال یا گاوسی است. متغیر تصادفی z دارای توزیع نرمال استاندارد ایت اگر تابع چگالی آن به صورت زیر باشد:

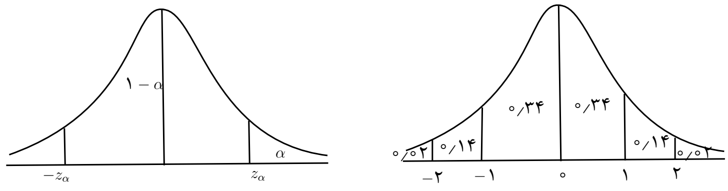
$$f_z(n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in R$$

$$X \sim N(0, 1), \quad EX = 0, \quad Vw(X) = 1$$

نمودار تابع چگالی

(۱)

$$P[Z > Z_\alpha] = \alpha, \quad P[Z < Z_\alpha] = 1 - \alpha$$



(ب) (آ)

۲)

$$P[Z < -Z_\alpha] = P[Z > Z_\alpha] = \alpha$$

۳)

$$P[Z < Z_\alpha] = P[Z > -Z_\alpha] = 1 - \alpha$$

۴)

$$P[|Z| < Z_\alpha] = 1 - 2\alpha$$

۵)

$$\forall Z_0 \in R: \quad P[|Z| < Z_0] = P[-Z_0 < Z < Z_0] = 2P[Z < Z_0] - 1$$

روابط فوق در محاسبه احتمال‌های $P[a < Z < b]$، $P[Z > b]$، $P[Z < a]$ با استفاده از جدول نرمال استاندارد مفید هستند:

تعریف ۱۰.۵.۳. *متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 است اگر تابع چگالی آن به صورت زیر باشد:*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in R \quad \mu \in R \quad \sigma > 0$$
$$E(X) = \mu, \quad Va(X) = \sigma^2, \quad X \sim N(\mu, \sigma^2)$$

قضیه ۱۱.۵.۳. *اگر $X \sim N(\mu, \sigma^2)$ باشد آن‌گاه $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$ برای محاسبه احتمالات مربوط به متغیر تصادفی نرمال غیر استاندارد، قضیه فوق مفید می‌باشد.*

برخی ویژگی‌های توزیع نرمال

۱. اگر X_1 و X_2 و ... و X_n متغیرهای تصادفی مستقل نرمال باشند آن‌گاه:
$X_i \sim N(\mu_i, \sigma_i^2) \;\; i = 1, 2, \ldots, n$

(الف)

$$\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$$

(ب)

$$\sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$$

که در آن a_1 و a_2 و ... و a_n راقب‌های حقیقی هستند.

۲. اگر $X_n \sim Bin(n,p)$ و $p > 0/1$ و $npq \geqslant 10$ آن‌گاه

تقریب دو جمله‌ای به نرمال:

$$Z_n = \frac{X_n - np}{\sqrt{npq}} \longrightarrow N(0, 1)$$

قضیه حد مرکزی: اگر $\{X_n, n \geqslant 1\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با $\mu = EX$ و $\sigma^2 = Var(X) < \infty$ آن‌گاه:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \underset{n\to\infty}{\longrightarrow} N(0, 1)$$

قضیه فوق یک قضیه بسیار مفید در تحلیل‌های آماری می‌باشد و بیان می‌کند اگر تعداد نمونه‌هاذ به اندازه کافی بزرگ باشند $(n \geqslant 30)$ محقق مجاز است توزیع جامعه را نرمال در نظر بگیرید.

فصل ۴

برآورد و آزمون فرض

۱.۴ توزیع‌های نمونه‌ای

n متغیر تصادفی $X_1, X_2, ..., X_n$ تشکیل یک نمونه تصادفی می‌دهند اگر مستقل و همتوزیع باشند:
از همتوزیع بودن داریم:

$$\begin{cases} EX_1 = EX_2 = ... = EX_n = \mu \\ \mathrm{Var}(X_1) = \mathrm{Var}(X_2) = ... = \mathrm{Var}(X_n) = \sigma^2 \end{cases}$$

الف- میانگین نمونه:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

ب- واریانس نمونه:

$$S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\}$$

$$S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\}$$

S_2^2 را واریانس تجربی نمونه نامند، تفاوت این دو واریانس در ادامه بیان خواهد شد.

۱.۱.۴ توزیع نرمال

اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از جامعه‌ای با میانگین μ و واریانس $\sigma^2 < \infty$ باشد آنگاه

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \underset{n\to\infty}{\longrightarrow} N(0, 1)$$

حد فوق به قضیه حد مرکزی معروف است که نقش مهمی در تحلیل‌های آماری دارد. و بیان می کند که میانگین نمونه به طور جانبی دارای توزیع نرمال است.
به این دلیل در تحلیل‌های آماری وقتی حجم نمونه به اندازه کافی بزرگ باشد $(n \geqslant 30)$ توزیع جامعه را نرمال در نظر می گیرند.

۲.۴

۱.۲.۴ توزیع تی- استودنت T-Student

اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از جامعه نرمال $\mu = EX$ و $\sigma^2 = \mathrm{Var} < \infty$ باشد و آنگاه $Z_n = \dfrac{\sqrt{n}(\bar{X} - \mu)}{S}$ برای هر n : دارای توزیع به نام توزیع تی-استودنت با پارامتر(درجه آزادی) $n - 1$ می باشد:

$$T_n = Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim T(n - 1)$$

از ویژگی‌های این توزیع این است که مشابه توزیع نرمال کاملا متقارن حول صفر بوده و علاوه بر این $ET_n = 0, \mathrm{Var}(T_n) =$ مقدار $t_\alpha(k)$ که به چندک $100(1-\alpha)\%$ توزیع معروف است با استفاده از جدول توزیع - T بدست می آید و در تحلیل‌ها از این مقادیر استفاده می شود.

موارد کاربرد توزیع

برآورد فاصله‌ای μ -برآورد فاصله‌ای $\mu_۲ - \mu_۱$ -آزمون فرض در مورد μ و $\mu_۲ - \mu_۱$ و آزمون فرض نمونه‌های زوج-برآورد و آزمون فرض ضرایب رگرسیون خطی.
÷ که در آن $\mu_۱$ و $\mu_۲$ میانگین‌های دو جامعه نرمال مستقل هستند:

نکته ۱.۲.۴. *اگر n به اندازه کافی بزرگ باشد آن‌گاه $t_\alpha(k) \cong Z_\alpha$ این نکته را با مراجعه به جداول توزیع‌های نرمال و $T-S$ می‌توان مشاهده نمود که*

$$\forall n \geqslant ۳۰ : \quad t_\alpha(n) \cong Z_\alpha$$

۳.۴

۱.۳.۴ توزیع خی دو $\chi^۲$

اگر $X_۱, X_۲, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^۲)$ -نمونه تصادفی نرمال با میانگین μ و واریانس $\sigma^۲$ باشد آن‌گاه

$$\chi = \frac{(n-۱)S^۲}{\sigma^۲} = \frac{(n-۱)S^۲}{\sigma^۲} = \frac{\sum_{i=۱}^{n}(X_i - \bar{x})^۲}{\sigma^۲} \sim \chi^۲(n-۱)$$

نماد $\chi^۲(n-۱)$ را برای توزیع خی دو با درجه آزادی $(n-۱)$ به کار می‌برند. این توزیع نامتقارن و چندک‌های آن نیز از طریق جدول توزیع خی -دو قابل دسترس هستند.
مقادیر $\chi^۲_\alpha(n-۱)$ و $\chi^۲_{۱-\alpha}(n-۱)$ از طریق جدول قابل دسترس هستند.

موارد کاربرد

برآورد فاصله‌ای و آزمون فرض در مورد واریانس و انحراف معیار جامعه.
-آزمون نیکویی برازش-آزمون ناپارامتری براساس جداول توافقی

۲.۳.۴ توزیع فیشر-F

اگر $X_۱, X_۲, ..., X_n$ یک نمونه تصادفی از جامعه $N(\mu_۱, \sigma_۱^۲)$ و $Y_۱, Y_۲, ..., Y_m$ یک نمونه تصادفی از جامعه $N(\mu_۲, \sigma_۲^۲)$ و دو جامعه مستقل از هم باشند و علاوه بر این:

$$S_۱^۲ = \frac{۱}{n-۱}\sum_{i=۱}^{n}(X_i - \bar{X})^۲$$

$$S_۲^۲ = \frac{۱}{m-۱}\sum_{i=۱}^{m}(Y_i - \bar{Y})^۲ \Rightarrow F = \frac{(n-۱)S_۱^۲}{(m-۱)S_۲^۲} \times \frac{\sigma_۲^۲}{\sigma_۱^۲} \sim F(n-۱, m-۱).$$

یعنی نسبت واریانس‌های نمونه در دو نمونه تصادفی از دو جامعه نرمال مستقل یک متغیر تصادفی با توزیع فیشر است. چندک‌های این توزیع یعنی $F_\alpha(k_۱, k_۲)$ و $F_{۱-\alpha}(k_۱, k_۲)$ که : $F_{۱-\alpha}(k_۱, k_۲) = \frac{۱}{F_\alpha(k_۲, k_۱)}$ از طریق جدول توزیع فیشر قابل دسترس هستند.

نکته ۱.۳.۴. *این توزیع نیز نامتقارن است.*

موارد کاربرد

مقایسه واریانس‌های دو جامعه نرمال مستقل- تحلیل رگرسیون - تحلیل طرح‌های آماری

برآورد نقطه ای(Point-Estimation)

فرض کنید $X_۱, X_۲, ..., X_n$ یک نمونه تصادفی از جامعه مورد مطالعه با پارامتر مجهول مورد بررسی θ باشد. اگر برای مقادیر مشاهده شده $x_۱, x_۲, ..., x_n$ تابع $T : R^n \to R$ موجود باشد به قسمی که $\theta = T(x_۱, x_۲, ..., x_n)$ یک تخمین برای θ باشد. آن‌گاه $T(x_۱, x_۲, ..., x_n)$ را یک برآورد کننده نقطه‌ای برای θ نامند.

مثال ۲.۳.۴. *الف- اگر $\theta = \mu$ آن‌گاه $\bar{X} = \frac{۱}{n}\sum_{i=۱}^{n} X_i$ و اگر $\theta = \sigma^۲$ باشد آن‌گاه*

$T_1 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ و $T_2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ برآوردکننده‌های نقطه‌ای برای μ و σ^2 هستند.

روشهای برآورد

۱- روش گشتاوری: گشتاور k ام نمونه $T = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ یک برآوردکننده نقطه‌ای برای گشتاور k ام جامعه یعنی $M_k = EX^k$ است.

۲- روش درستنمایی ماکزیمم:
در این روش براورد کننده پارامتر θ برمبنای نمونه‌ای بدست می‌آید که در بین تمام نمونه‌ها محتمل‌ترین باشد. یعنی θ را با حل یک معادله ریاضی طوری می‌یابند که وقتی $x_1, x_2, ..., x_n$ مشاهده شده باشند تابع چگالی ماکزیمم مقدار را دارا باشد.

ویژگیهای یک برآورد کننده

۱- ناا ریبی: یعنی متوسط مقادیر $T(\underset{\sim}{X})$ برابر مقدار واقعی پارامتر مجهول مورد بررسی θ باشد:
$ET(\underset{\sim}{X}) = \theta$

۲- کمترین واریانس: از بین دو برآورد کننده T_1 و T_2 برای θ برآوردکننده ای دقیق تر(خطای کمتر) است که واریانس کمتری داشته باشد. یعنی مقادیر آن دارای نوسان کمتری باشند.

۳- کمترین مربعات خطا:
اگر دو برآوردکننده ناا ریب نباشند آن‌گاه برای مقایسه آنها از میانگین مربعات خطا: $MSE(T) = E(T - \theta)^2$ استفاده می شود در این حالت برآوردکننده‌ای دقیق است که MSE آن کوچکتر است.

۴- سازگاری: برآوردکننده $T_n(X)$ را سازگار نامند اگر برای مقادیر بزرگ n

$$T_n(\underset{\sim}{X}) \cong \theta$$

مثال ۳.۳.۴. -برآوردکننده $T = \bar{X}$ برای θ نااریب است.

-برآوردکننده $T_1 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ برای σ^2 نااریب است

-برآوردکننده $T_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ برای σ^2 نااریب است و علاوه بر این $ET_2 = \frac{n-1}{n}\sigma^2 = (1 - \frac{1}{n})\sigma^2 < \sigma^2$ و برای مقادیر بزرگ n : $ET_2 \cong \sigma^2$

برآورد فاصله ای (Interral Estimation)

چون در برآورد نقطه‌ای میزان خطای برآورد قابل محاسبه نمی باشد، فاصله $(T(\underset{\sim}{X}) - E, T(\underset{\sim}{X}) + E)$ به مرکزیت برآورد نقطه‌ای $\hat{\theta} = T(\underset{\sim}{X})$ را طوری تعیین می‌کنیم که با احتمال $B^{(1-\alpha)}$ مقدار θ را در برداشته باشد. در این صورت $L = 2E$ بعنوان یک معیار دقت برآورد استفاده می شود: بدیهی است که هرچه L کوچکتر باشد برآوردکننده $T(\underset{\sim}{X})$ از دقت بیشتری برخوردار است.

نکته ۴.۳.۴. در تعیین برآورد فاصله‌ای θ توزیع $T(X_1, ..., X_n)$ و α نقش اصلی را دارند.

برآورد فاصله‌ای μ

اگر $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2).$ آن‌گاه:

الف- اگر $\sigma = \sigma_0$: ازاینکه $Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma_0} \sim N(0, 1)$ فاصله $(\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}})$

یک برآوردکننده فاصله‌ای با ضریب اطمینان $1 - \alpha$ برای μ است.

فاصله فوق را بازه اطمینان $(1-\alpha)100\%$ برای μ نامند.

تعیین حجم نمونه

حجم نمونه لازم برای برآورد μ بشرط اینکه میزان دقت حداکثر d باشد، چقدر است؟

$$l = ۲Z_{\frac{\alpha}{۲}}\frac{\sigma_{\circ}}{\sqrt{n}} \leqslant d \Rightarrow n \geqslant \frac{۴Z^{۲}_{\frac{\alpha}{۲}}\sigma^{۲}_{\circ}}{d^{۲}}$$

ب- اگر σ مجهول باشد آن‌گاه از اینکه $T = \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim T(n-۱)$ آن‌گاه برآوردکننده فاصله‌ای با ضریب اطمینان $(۱-\alpha)$ برای μ برابر است با: $(\bar{X} - t_{\frac{\alpha}{۲}(n-۱)}\frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{۲}(n-۱)}\frac{S}{\sqrt{n}})$

حجم نمونه لازم برای اینکه طول بازه حداکثر d باشد:

۱- یک نمونه مقدماتی به حجم $n_{\circ}$ انتخاب کرده از رابطه $n_{۱} = \left[\frac{۴t^{۲}_{\frac{\alpha}{۲}}(n_{\circ}-۱)S^{۲}_{\circ}}{d^{۲}}\right]$ را بدست آورده و سپس قرار می دهیم $n = max\{n_{\circ}, n_{۱}\}$

نکته ۵.۳.۴. *اگر جامعه غیر نرمال ولی $n \geqslant ۳۰$ نتایج فوق معتبر هستند.*

برآورد فاصله ای $\mu_{۱} - \mu_{۲}$

فرض کنید:

۱- $X_{۱}, X_{۲}, ..., X_{n}$ یک نمونه تصادفی از جامعه نرمال : $N(\mu_{۱}, \sigma^{۲}_{۱})$

۲- $Y_{۱}, Y_{۲}, ..., Y_{n}$ یک نمونه تصادفی از جامعه نرمال : $N(\mu_{۲}, \sigma^{۲}_{۲})$

۳- دو جامعه مستقل از هم باشند.

هدف: *مقایسه دو جامعه از طریق مقایسه میانگین‌ها*

الف- $\sigma_{۱}$ و $\sigma_{۲}$ معلوم: در این حالت $\widehat{\mu_{۱} - \mu_{۲}} = \hat{\mu_{۱}} - \hat{\mu_{۲}} = \bar{X} - \bar{Y}$ و

$$\frac{\bar{X} - \bar{Y} - (\mu_{۱} - \mu_{۲})}{\sqrt{\frac{\sigma^{۲}_{۱}}{n} + \frac{\sigma^{۲}_{۲}}{m}}} \sim N(\circ, ۱)$$

ب- $\sigma_1 = \sigma_2 = \sigma$ و σ مجهول : در این حالت $\hat{\sigma}^2 = S_P^2$

$$S_P^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{m+n-2}, T = \frac{\bar{X} - \bar{Y} - (M_1 - M_2)}{S_P\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim T(m+n-2)$$

و در نتیجه:

$$\mu_1 - \mu_2 : \bar{X} - \bar{Y} \mp t_{\frac{\alpha}{2}}(k)\frac{S_P}{\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

ج- $\sigma_2 \neq \sigma_1$ و هر دو مجهول : در این حالت برای $m, n \geqslant 30$ برآورد فاصله‌ای مجانبی به صورت زیر داریم:

$$\mu_1 - \mu_2 : \bar{X} - \bar{Y} \mp Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}$$

برآورد فاصله‌ای σ, σ^2

اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از $N(\mu, \sigma^2)$ باشد آن‌گاه از اینکه

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

داریم:

$$\sigma^2 \in \Big(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\Big)$$

و برای σ داریم:

$$\sigma \in \Big(\sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}\Big)$$

آزمون فرضیه های آماری Statistrical Hypothesis Test

در این قسمت می‌خواهیم با استفاده از اطلاعات حاصل از نمونه در مورد یک ادعا تصمیم‌گیری نماییم.

فرضیه آماری

هر ادعا یا پیشنهاد در مورد پارامتر مجهول یا توزیع جامعه آماری را یک فرضیه آماری نامند.
عموما با دو نوع فرضیه آماری مواجه هستیم: یکی از فرضیه‌ها را با H_0 (فرضیه صفر) و دیگری را با H_1 (فرضیه مقابل) نشان می دهند.

نکته ۶.۳.۴. *ادعایی‌که به دنبال رد یا اثبات آن هستیم را بعنوان فرض H_1 در نظرگرفته و فرضیه صفر H_0 نقیض فرضیه H_1 خواهد بود.*

مثال ۷.۳.۴. *اگر هدف مطالعه میانگین یک جامعه آماری باشد*

$$\begin{cases} H_0 : \mu = 100 \rightarrow \text{یک فرض ساده} \\ H_1 : \mu \neq 100 \rightarrow \text{یک فرض مرکب دو طرفه} \end{cases}$$

انواع خطا:

خطای نوع I: رد H_0 به اشتباه $=$ قبول H_1 به اشتباه

$$\alpha = P(\text{خطای } \boldsymbol{I}) = P[RH_0 | H_0 \text{درست}]$$

خطای نوع II: رد H_1 به اشتباه $=$ قبول H_0 به اشتباه

$$\beta = P(\text{خطای } \boldsymbol{II}) = P[AH_0 | H_0 \text{غلط}] = P[RH_1 | H_1 \text{درست}]$$

آزمون آماری: *قاعده یا معیاری‌که با خلاصه‌کردن مشاهدات حاصل از نمونه حکم به تایید یا رد فرض H_0 می دهد.*

نکته ۸.۳.۴. *α را سطح آزمون نامیده و با معلوم بودن α وتوزیع برآوردکننده نقطه‌ای پارامتر θ آزمون آماری کاملا مشخص می شود.*

برآورد نسبت جامعه

فرض کنید p نسبت جامعه (نسبت افرادی که دارای ویژگی خاصی هستند) باشد. و

$$X_i = \begin{cases} ۱ & \text{عضو } i \text{ ام نمونه درگروه ویژه باشد} \\ ۰ & \text{در غیر این صورت} \end{cases}$$

در این صورت $\hat{p} = \bar{X} = \frac{۱}{n}\sum_{i=۱}^{n} X_i$ و برای مقادیر بزرگ n برآورد فاصله‌ای مجانبی p برابر است با :

$$p = \hat{p} \mp Z_{\frac{\alpha}{۲}} \sqrt{\frac{pq}{n}} \quad or \quad \hat{p} \mp Z_{\frac{\alpha}{۲}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

آزمون میانگین جامعه نرمال

فرض کنید $X_۱, X_۲, ..., X_n$ یک نمونه تصادفی از $N(\mu, \sigma^۲)$ باشد، در این بخش هدف آزمون هر یک از فرضیه‌های زیر است:

۱)

$$\begin{cases} H_۰ : \mu = \mu_۰ \\ H_۱ : \mu \neq \mu_۰ \end{cases}$$

۲)

$$\begin{cases} H_۰ : \mu = \mu_۰ \\ H_۱ : \mu > \mu_۰ \end{cases}$$

۳)

$$\begin{cases} H_۰ : \mu = \mu_۰ \\ H_۱ : \mu < \mu_۰ \end{cases}$$

الف) $\sigma = \sigma_0$ معلوم: از اینکه $Z_0 = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma_0} \sim N(0, 1)$ و $\hat{\mu} = \bar{X}$ آزمون‌های زیر برای هر یک از فرضیه‌های فوق با معلوم بودن α بدست می‌آید.

۱)

$$RH_0 \Leftrightarrow |Z_0| > Z_{\frac{\alpha}{2}} \quad or \quad p-value < \frac{\alpha}{2}$$
$$p-value = P[Z > |Z_0|] < \frac{\alpha}{2}$$

۲)

$$RH_0 \Leftrightarrow Z_0 > Z_{\alpha} \quad or \quad p-value < \alpha$$
$$p-value = P[Z > |Z_0|]$$

۳)

$$RH_0 \Leftrightarrow Z_0 < -Z_{\alpha} \quad or \quad p-value < \alpha$$
$$p-value = P[Z - |Z_0|]$$

ب- σ مجهول: در این حالت از $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{i-\bar{X}})^2$ استفاده کرده و چون $T_0 = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim T(n-1)$ با معلوم بودن α داریم:

۱)

$$RH_0 \Leftrightarrow |t_0| > Z_{\frac{\alpha}{2}}(n-1) \quad or \quad p-value = P[T(n-1) > |t_0|] < \frac{\alpha}{2}$$

۲)

$$RH_0 \Leftrightarrow t_0 > t_\alpha(n-1) \quad or \quad p-value < \alpha$$

۳)

$$RH_0 \Leftrightarrow t_0 < -t_\alpha(n-1) \quad or \quad p-value < \alpha$$

نکته ۹.۳.۴. *هرگاه توزیع جامعه آماری غیر نرمال باشد و $n \geqslant 30$ آزمون‌های فوق معتبر هستند.*

آزمون واریانس

فرضیه‌های مورد آزمون عبارتند از:

$$1)H_0 : \sigma = \sigma_0 \quad against \quad H_1 : \begin{cases} 1)\sigma > \sigma_0 \\ 2)\sigma < \sigma_0 \\ 3)\sigma \neq \sigma_0 \end{cases}$$

۱)

$$RH_0 \Leftrightarrow X = \frac{(n-1)S^2}{\sigma_0^2} > \chi^2_\alpha(n-1)$$

$$x_0 = \frac{(n-1)S^2}{\sigma_0^2} \quad or \quad p-value = P[\chi(n-1) > x_0] < \alpha$$

۲)

$$RH_\circ \Leftrightarrow X = \frac{(n-1)S^2}{\sigma_\circ^2} < \chi^2_{1-\alpha}(n-1)$$
$$or \quad p-value = P[\chi^2(n-1) > x_\circ] < \alpha$$

۳)

$$RH_\circ \Leftrightarrow X > \chi^2_\alpha(n-1) \quad or \quad < \chi^2_{1-\alpha}(n-1)$$
$$p-value = min\{P[X > x_\circ], P[X < x_\circ]\} < \frac{\alpha}{2}$$

آزمون فرضیه مقایسه دو جامعه نرمال

برای مقایسه دو جامعه نرمال مستقل از هم براساس اطلاعات حاصل از دو نمونه تصادفی از هر جامعه (نمونه۱ :حجم n نمونه۲ :حجم m) فرضیه‌های آماری زیر را در سطح معلوم α آزمون می‌کنیم:

$$1)H_\circ : \mu_1 = \mu_2 \quad against \quad : \begin{cases} 1)\mu_1 > \mu_2 \\ 2)\mu_1 < \mu_2 \\ 3)\mu_1 \neq \mu_2 \end{cases}$$

الف- $\sigma_1 = \sigma_2 = \sigma$ و σ مجهول: در این حالت $\hat{\sigma} = S_P^2$ از اینکه $T_{H_\circ} = \dfrac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim$ *$T(k), k = m+n-1$ آزمون‌های زیر بدست می‌آید:*

۱)

$$RH_\circ \Leftrightarrow t_\circ = \frac{\bar{x}-\bar{y}}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} > t_\alpha(k) \quad or \quad p-v = P[T(k) > t_\circ] < \alpha$$

۲)

$$RH_\circ \Leftrightarrow t_\circ = \frac{\bar{x}-\bar{y}}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} < -t_\alpha(k) \quad or \quad p-v = P[T(k) < t_\circ] < \alpha$$

۳)

$$RH_\circ \Leftrightarrow |t_\circ| > t_{\frac{\alpha}{2}}(k) \quad or \quad p-v = P[T(k) > |t_\circ|] < \frac{\alpha}{2}$$

ب- σ_1, σ_2 معلوم :

در این حالت از کمیت $Z_{H_\circ} = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}$ و از اینکه $Z_{H_\circ} \sim N(0,1)$ است استفاده می شود در نتیجه کافی است در آزمون‌های قسمت الف Z_α جایگزین $t_\alpha(k)$ و $Z_{\frac{\alpha}{2}}$ جایگزین $t_{\frac{\alpha}{2}}(k)$ شود.

ج- اگر $\sigma_1 \neq \sigma_2$ و هر دو مجهول:

در این حالت اگر $n, m \geqslant 30$ از کمیت $Z_{H_\circ} = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}}$ و از اینکه $Z_{H_\circ} \to N(0,1)$ استفاده کرده همچنین به جای $t_\alpha(k)$ و $t_{\frac{\alpha}{2}}(k)$ از Z_α و $Z_{\frac{\alpha}{2}}$ استفاده می شود.

فرضیه همگنی واریانسها

قبل از انجام هر یک آزمون‌های فوق لازم است فرضیه همگنی واریانسها $H_0 : \sigma_1 = \sigma_2$ آزمون شود تا در صورت پذیرش H_0 به سادگی از آزمون قسمت الف استفاده شود. روش انجام آزمون در نرم افزار $SPSS$ توضیح داده شده ولی بدون نرم افزار نیز می توان انجام داد: چون در اینجا فرضیه H_0 مورد توجه است:

$$AH_0 \Leftrightarrow 1 \in \Big(\frac{1}{F_\alpha(m-1,n-1)} \times \frac{S_2^2}{S_1^2} \quad F_\alpha(n-1,m-1) \times \frac{S_2^2}{S_1^2}\Big)$$

آزمون -تی(T-test)برای نمونه های زوج

در بسیاری ازکاربردهای آمار نمونه‌ها به صورت زوج بدست می‌آیند در این حالت دو نمونه وابسته هستند و هر عضو نمونه متناظر با عضو دیگر نمونه هستند، بنابراین:

$$(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$$

یکی از حوزه‌های پرکاربرد این آزمون در پزشکی می باشد، زمانی‌که هدف بررسی تاثیر یک دارو در بهبود یک مریضی خاص است و افراد نمونه بصورت گروه شاهد و گروه تحت کنترل مورد مطالعه قرار می‌گیرند.

$$H_0 : \text{تیمار موثر نیست} : \mu_D \quad against \quad H_1 : \begin{cases} 1)\mu_D > 0 \\ 2)\mu_D < 0 \\ 3)\mu_D \neq 0 \end{cases}$$

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i, \quad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$$

در این حالت:

$$T = \frac{\sqrt{n}\bar{D}}{S_D} \sim T(n-1)$$

befor	after	Difference
نمونه۱	نمونه۲	$D_i = X_i - Y_i$
X_1	Y_1	$D_1 = X_1 - Y_1$
X_2	Y_2	
$\vdots$	$\vdots$	$\vdots$
X_n	Y_n	$D_n = X_n - Y_n$

۱)

$$RH_0 \Leftrightarrow t_0 = \frac{\sqrt{n}\bar{d}}{S_D} > t_\alpha(n-1) \quad or \quad p-v < \alpha$$

۲)

$$RH_0 \Leftrightarrow t_0 = \frac{\sqrt{n}\bar{d}}{S_D} < -t_\alpha(n-1) \quad or \quad p-v < \alpha$$

۳-

$$RH_0 \Leftrightarrow |t_0| > t_{\frac{\alpha}{2}} \quad or \quad p-v = P[T(n-1) > |t_0|] < \frac{\alpha}{2}$$

فصل ۵

تمرین

۱- میزان شوری آب (درصد) یک چاه گمانه کمیت تصادفی با توزیع یکنواخت در بازه $[10, 30]$ است. تعداد $n = 10$ نمونه آب به تصادف انتخاب شده است.

الف- احتمال این‌که میزان شوری ۵ نمونه در بازه $[10, 20]$ باشد چقد راست؟

ب- وقتی نمونه‌ها به‌طور متوالی برداشت می‌شوند، احتمال این‌که ۵امین نمونه اولین نمونه‌ای با میزان شوری کمتر از ۲۰ {در بازه $[10, 20]$} باشد چقدر است؟

ج- اگر تعداد نمونه‌ها $n = 100$ باشد آن‌گاه:

-احتمال تقریبی این‌که حداقل ۲ نمونه دارای میزان شوری در بازه $[10, 11]$ باشند چقدر است؟

-احتمال تقریبی این‌که میزان شوری حداقل ۴۵ نمونه در بازه $[10, 20]$ باشد چقدر است؟

۲- تعداد زمین‌لرزه‌ها در یک ناحیه از فرآیند پوآسون با نرخ $\lambda = 4$ در هر دوره ۲۰ ساله پیروی می‌کند. ۵۰ درصد زمین‌لرزه‌ها شدی هستند. مطلوب است احتمال این‌که

الف- در ۱۰ سال آینده ۲ زمین‌لرزه رخ دهد؟

ب- اولین زمین‌لرزه پس از ۵ سال رخ دهد؟

ج- اولین زمین‌لرزه شدید پس از ۱۰ سال رخ دهد؟

د- اگر در ۱۰ سال آینده ۴ زمین‌لرزه رخ دهد، ۲تای آن‌ها شدید باشد؟

۳- میزان کریستال یک نوع سنگ کمیت تصادفی با توزیع نرمال، میانگین ۶۰ (درصد)، و انحراف معیار ۲ می‌باشد. یک پژوهشگر نمونه‌هایی از این نوع سنگ جمع‌آوری کرده است.

الف- چه نسبتی از این نمونه‌ها دارای کریستال بیشتر از ۵۸ می‌باشند؟

ب- چه نسبتی از این نمونه‌ها دارای کریستال در بازه [۵۸, ۶۲] می‌باشند؟

ج- میزان کریستال ۹۵ درصد نمونه‌ها حداکثر چقدر است؟

$$P[X=k]=\binom{n}{k}p^k q^{n-k}, k=0,1,2,\ldots,n \qquad X\sim Bin(n,p)$$

$$P[X=k]=\frac{e^{-\lambda}\lambda^k}{k!}, k=0,1,2,\ldots \qquad X\sim P_\circ(\lambda)$$

$$P[N(t)=k]=\frac{e^{-\lambda t}(\lambda t)^k}{k!}, k=0,1,\ldots \qquad N(t)\sim P_\circ(\lambda t)$$

$$P[X=k]=Pq^{k-1}, k=1,2,\ldots \qquad X\sim G(P)$$

$$f(x)=\lambda e^{-\lambda x}, x\geqslant 0 \qquad X\sim Exp(\lambda)$$

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

$$Z=\frac{X-\mu}{\sigma}\sim N(0,1)$$

$$If \quad X\sim Ex=\mu,\sigma^2=var(X)\Rightarrow Z=\frac{X-\mu}{\sigma}$$

۴- ناحیه‌ای به سه بلوک A_1, A_2, A_3 تقسیم شده است. نسبت مساحت هر بلوک به ترتیب ۰٫۵, ۰٫۲, ۰٫۳ است. نسبت معینی از این بلوک‌ها توسط کارشناسان برای اکتشاف یک ماده معدنی به کار برده می‌شود که این نسبت‌ها به ترتیب عبارتند از ۰٫۴, ۰٫۳, ۰٫۱.

الف- چه نسبتی از کل ناحیه تحت اکتشاف قرار خواهد گرفت.

ب- اگر نسبت کل ناحیه تحت اکتشاف ۰٫۶ تعیین شده باشد چه نسبتی از هر بلوک تحت اکتشاف قرار خواهد گرفت.

۵- در یک ناحیه که مشخص شده است که تنها ۲۰ درصد چاه‌های حفاری شده به آب می‌رسد تعداد ۵ چاه آزمایشی حفر خواهد شد.

الف- احتمال با شکست روبرو شدن طرح حفر چاه چقدر است.

ب- احتمال به آب رسیدن لااقل یکی از چاه‌های حفاری شده چقدر است.

۶- تعداد سیلاب‌ها در ناحیه‌ای از توزیع پوآسون با میانگین ۵ در دوره ۲۰ ساله پیروی می‌کند.

الف- احتمال رخ دادن حداقل ۲ سیلاب در دوره بیست ساله چقدر است.

ب- احتمال رخ دادن دقیقا ۲ سیلاب در دوره ده ساله چقدر است.

ج- اگر ۳۰ درصد سیلاب‌ها مهیب باشند

۱- احتمال رخ دادن ۲ سیلاب مهیب در دوره ۳۰ ساله چقدر است.

۲- احتمال این‌که در ۲ دوره از ۵ دوره ۲۰ ساله آینده سیلاب مهیب رخ ندهد چقدر است.

۷- تعداد زمین‌لرزه‌ها در یک ناحیه از توزیع پوآسون با نرخ ۶ زمین‌لرزه در دوره زمانی ۴۰ ساله پیروی می‌کند.

الف- احتمال این‌که فاصله زمانی بین دو زمین‌لرزه بیش از ۲۰ سال باشد چقدر است.

ب- اگر ۲۰ درصد این زمین‌لرزه‌ها شدید باشد احتمال رخ دادن یک زمین‌لرزه شدید در یک دوره ۴۰ ساله در صورتی‌که کلا ۴ زمین‌لرزه در این دوره رخ داده باشد چقدر است.

۸- از بین ۴ دانشجوی سال سوم و ۶ دانشجوی سال چهارم و گروهی شامل ۳ دانشجو به تصادف انتخاب می‌شود متغیرهای تصادفی X را تعداد دانشجویان سال سوم و Y را تعداد دانشجویان سال چهارم در نظر بگیرید. مطلوب است محاسبه

الف- تابع احتمال توام (X, Y).

ب- $P[X < Y]$

ج- ضریب همبستگی (X, Y).

۹- ۵ نمونه ماسه سنگ و ۴ نمونه سنگ آهک برای شمارش تعداد لایه‌های آن جمع‌آوری شده است. ۳ نمونه سنگ از بین ۹ نمونه فوق به تصادف انتخاب می‌شود.

الف- احتمال این‌که ۲ نمونه سنگ آهکی باشد چقدر است؟

ب- اگر نمونه‌ها یکی یکی انتخاب شوند احتمال این‌که سومی ماسه‌سنگ باشد چقدر است؟

۱۰- در یک برنامه اکتشافی تعداد $n = ۱۰$ چاه‌گمانه برای رسیدن به آب حفر شده است. اگر احتمال رسیدن آب در این ناحیه ۰٫۶ باشد.

الف- احتمال این که طرح با شکست روبرو نشود چقدر است؟

ب- احتمال این‌که حداکثر ۴ چاه به آب برسد چقدر است؟

۱۱- تعداد سیلاب‌ها در یک ناحیه از فرآیند پوآسون با نرخ $\lambda = ۴$ در هر دوره ۶۰ ساله پیروی می‌کند اگر ۷۰ درصد سیلاب‌ها مهیب و ویران‌گر باشند مطلوب است محاسبه هر یک از احتمالات زیر

الف- فاصله زمانی بین دو سیلاب مهیب بیش از ۲۰ سال باشد.

ب- حداکثر ۳ سیلاب مهیب در یک دوره ۱۰ ساله رخ دهد.

ج- اگر در ۴۰ سال آینده ۵ سیلاب رخ دهد احتمال این‌که حداقل ۲تای آن‌ها شدید باشد چقدر است؟

۱۲- مشخص شده است که تنها ۱۲ درصد نمونه‌هایی که از معدن طلا استخراج می‌شوند مقدار طلای بسیار کمی دارند: اگر $n = ۱۰۰$ نمونه به تصادف از این معدن انتخاب شده باشد. احتمال این‌که

الف- دقیقا ۵تای آن‌ها طلای بسیار کمی داشته باشند چقدر است؟ (احتمال تقریبی)

ب- حداکثر ۱۰ تای آن‌ها طلای بسیار کمی داشته باشند چقدر است؟ (احتمال تقریبی)

۱۳- میزان کوارتز یک نوع سنگ آذرین از توزیع نرمال با میانگین ۳۰ (درصد) و انحراف معیار ۳ پیروی می‌کند.

الف- چه نسبتی از این سنگ‌ها دارای کوارتز بیش از ۲۷ درصد هستند؟

ب- مقدار انحراف معیار چقدر باشد تا فقط ۱۰ درصد نمونه‌ها میزان کوارتز کمتر از ۹ درصد داشته باشند:

مقادیر جدول نرمال:

$$P[Z<۱]\cong ۰/۸۴,\quad P[Z<۱/۲۸]\cong ۰/۹$$

$$P[Z<-۰/۱۵]\cong ۰/۴۴,\quad P[Z<-۰/۱۹]\cong ۰/۴۲$$

$$P[Z<۱/۳۱]\cong ۰/۹۱,\quad P[Z<-۰/۳۱]\cong ۰/۳۸۵$$

$$P[Z<۰/۷]=۰/۷۵۸,\quad P[Z<۰/۶]=۰/۷۲۶$$

$$P[Z<۰/۷۳]\cong ۰/۷۶,\quad P[Z<۰/۶۳]=۰/۷۳$$

$$P[X=K]=\binom{n}{k}p^kq^{n-k}, k=۰,۱,۲,\dots,n \quad X\sim Bin(n,p)$$

$$P[X=K]=\frac{e^{-\lambda}\lambda^k}{k!} \qquad X\sim P_{\circ}(\lambda)$$

$$f(x)=\frac{۱}{\sqrt{۲\pi}\sigma}e^{-\frac{(x-\mu)^۲}{۲\sigma^۲}} \qquad X\sim N(\mu,\sigma^۲)$$

۱۴- یک معدن را در نظر بگیرید که تنها ۳۰٪ نمونه‌های استخراج شده دارای مرغوبیت لازم برای استخراج یک ماده معدنی هستند:

الف- اگر تعداد ۱۰ نمونه به تصادف انتخاب شود مطلوب است احتمال این‌که

۱- حداقل شامل دو نمونه‌ی خوب باشد.

۲- اگر میزان ماده معدنی هر نمونه یک متغیر تصادفی یکنواخت در بازه (a,b) باشد چه نسبتی از این نمونه‌ها شامل ماده معدنی بیش از $\frac{a+b}{۲}$ خواهند بود.

۳- احتمال این‌که حداقل ۵ نمونه بیش از $\frac{a+b}{۲}$ ماده معدنی داشته باشند چقدر است؟

ب- اگر تعداد $n = ۱۰۰$ نمونه به تصادف انتخاب شود مطلوب است محاسبه احتمال تقریبی در هر یک از سوالات فوق

ج- اگر نمونه‌برداری را ادامه دهیم تا اولین نمونه با ماده معدنی شن از $\frac{a+b}{۲}$ بدست آید، به‌طور متوسط چند نمونه لازم است.

۱۵- تعداد زمین‌لرزه‌ها در یک ناحیه از فرآیند پوآسون با نرخ $\lambda = ۵$ در یک دوره ۳۰ ساله پیروی می‌کنند، بر اساس اطلاعات اشکال در متن مشخص شده است که ۴۰ درصد این زمین‌لرزه‌ها شدید هستند مطلوب است:

الف- احتمال این‌که اولین زمین‌لرزه شدید بعد از ۲۰ سال رخ دهد چقدر است.

ب- احتمال این‌که چهارمین زمین‌لرزه شدید بعد از ۲۵ سال رخ دهد چقدر است.

ج- اگر در یک دوره ۳۰ کلا ۱۰ زمین‌لرزه رخ دهد احتمال این‌که حداقل ۲تای آن‌ها در ۵ سال اول رخ دهند چقدر است.

پیشنهادات و انتقادات خود در مورد این راهنمای کوتاه را می‌توانید به رایانامه xepersian@mail.com ارسال کنید. با توجه به این‌که آماده‌سازی این مجموعه زمان زیادی برده و حاصل چند سال تجربه دوستان شماست، با در نظر گرفتن اهداف کسانی که در آماده‌سازی آن به هر نحو تلاش داشته‌اند. انتظار می‌رود که شما دوست عزیز فقط در جهت انجام رساله‌های شخصی خویش از آن استفاده نمایید.