

שיטות סטטיסטיות ונומריות בפיסיקה

סוכם על ידי חן אבינדב. מבוסס על ספר אמיתי:
R. J. Barlow, Statistics, Wiley & Sons, Chichester, 1989.
המספרים בסוגריים המרובעים מציינים את העמוד בספר בו מופיע הנושא הנדון.

הסתברות

מבוא והגדרות:

ההסתברות שמאורע מסויים יתרחש מסומנת ב- $\Pr(x)$, ומתקיים:

$$0 \leq \Pr(x) \leq 1$$

אם כל המאורעות האפשריים הם $\{x_i\}$, נדרוש:

$$\sum_i \Pr(x_i) = 1$$

בתנאי שאין חפיפה בין המאורעות השונים (הם mutually exclusive). כמו כן מתקיים:

$$\Pr(x_1 \text{ or } x_2) = \Pr(x_1) + \Pr(x_2) - \Pr(x_1 \text{ and } x_2)$$

כאשר אם אין חפיפה בין x_1 ו- x_2 האיבר האחרון מתאפס, כמובן.

אחת ההגדרות להסתברות היא כזו - אם מבצעים ניסיון מסויים n פעמים, ומאורע A מתרחש בסך הכל m פעמים, אז $\Pr(A)$ מוגדרת לפי:

$$\Pr(A) = \lim_{n \rightarrow \infty} \frac{m}{n}$$

הסתברות מותנית ומשפט בייס (Bayes' Theorem):

$P(A|B)$ היא ההסתברות שמאורע A יתרחש, בהינתן שמאורע B כבר התרחש. כך למשל, אם בוחרים יום באקראי אזי ההסתברות שמדובר ביום ראשון היא $1/7$. אך אם בוחרים יום באקראי וידוע לנו שזהו יום חול, ההסתברות היא $1/5$. מתוך שתי צורות הכתיבה השקולות:

$$\begin{aligned} \Pr(A \text{ and } B) &= \Pr(A|B)\Pr(B) \\ &= \Pr(B|A)\Pr(A) \end{aligned}$$

מקבלים את משפט בייס:

$$\Pr(A|B) = \Pr(B|A) \frac{P(A)}{P(B)}$$

ובהכללה, אם אוסף המאורעות $\{A_i\}$ הם mutually exclusive ומכסים ביחד את כל מרחב האפשרויות, מתקיים:

$$\Pr(B) = \sum_j \Pr(B|A_j)P(A_j)$$

ולכן ניתן לרשום:

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)P(A_i)}{\sum_j \Pr(B|A_j)P(A_j)}$$

קומבינטוריקה:

מספר הדרכים לסדר N עצמים, עם חשיבות לסדר שלהם:

$$N! = N \cdot (N-1) \cdots 2 \cdot 1 \quad (0! = 1)$$

מספר הדרכים לסדר r עצמים ב- n מקומות פנויים, עם חשיבות לסדר:

$${}_n P_r = n(n-1) \cdots (n-r+1) = \frac{n!}{(n-r)!}$$

דוגמה: כמה מילים בנות 4 אותיות ניתן לבנות בשפה העברית? $22!/18!$. כאן ה"מקומות" הם למעשה האותיות, וה"עצמים" הם אות ראשונה במילה, אות שנייה במילה, וכו'. מספר הדרכים לסדר את העצמים, ללא חשיבות לסדר שלהם:

$${}^n C_r = \frac{{}^n P_r}{r!} = \frac{n!}{r!(n-r)!} = \binom{n}{r}$$

דוגמה: כמה קבוצות של 4 אותיות ניתן ליצור בשפה העברית? $22!/(18!4!)$. ה"מקומות" הם שוב האותיות העומדות לרשותנו, אך ה"עצמים" הם המקומות שאנו מקצים בקבוצה, ללא חשיבות לסדר של כל אות בקבוצה. נוסחת סטירלינג:

$$n! \approx n^n e^{-n} \sqrt{2\pi n}$$

$$\ln n! \approx n \ln n - n$$

תיאור הנתונים

[6] ממוצע חשבוני (Average):

בהינתן אוסף נתונים עם N איברים:

$$\{x_1, x_2, x_3, \dots, x_N\}$$

הממוצע של x הוא:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

באופן דומה ניתן להגדיר ממוצע של כל פונקציה $f(x)$:

$$\bar{f} = \frac{1}{N} \sum_{i=1}^N f(x_i)$$

אם הנתונים הוכנסו לתאים (bins), כאשר התא j -שמתאים לערך x_j מכיל n_j איברים, הממוצעים הם:

$$\bar{x} = \frac{1}{N} \sum_j n_j x_j$$

$$\bar{f} = \frac{1}{N} \sum_j n_j f(x_j)$$

[8] שונות (Variance):

השונות של אוסף נתונים מוגדרת בתור:

$$V(x) = \frac{1}{N} \sum_i (x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$$

והיא מדד לרוחב של התפלגות הנתונים.

באופן דומה ניתן להגדיר שונות לכל פונקציה $f(x)$:

$$V(f) = \frac{1}{N} \sum_i (f(x_i) - \bar{f})^2$$

[9] סטיית תקן (Standard Deviation):

מוגדרת בתור שורש השונות:

$$\sigma = \sqrt{\frac{1}{N} \sum_i (x_i - \bar{x})^2} = \sqrt{\overline{x^2} - \bar{x}^2}$$

ברוב המקרים נוח יותר לעבוד עם סטיית התקן, משום שיש לה את אותן יחידות כמו לאוסף הנתונים עצמו.

הגדרה נוספת לסטיית התקן של אוסף נתונים היא:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_i (x_i - \bar{x})^2}$$

והיא 'נכונה' כאשר לא ידוע לנו μ , הממוצע של התפלגות האם (mean) ממנה מגיעים הנתונים. כאשר ערך זה כן ידוע לנו, נציבו בנוסחה הראשונה במקום הממוצע של אוסף הנתונים \bar{x} .

[13] רוחב בחצי הגובה (Full Width at Half Maximum):

סטיית התקן כפי שהיא מוגדרת לעיל יכולה להיות מושפעת מאוד מערכי קצה של ההתפלגות, ובמקרים כאלו מגדירים מדד נוסף זה לרוחב ההתפלגות. עבור התפלגות גאוסיאנית מתקיים:

$$FWHM = 2.35\sigma$$

[15] שונות משותפת (Covariance):

ניח כי כל איבר באוסף המדידות שלנו מורכב משני ערכים:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

ניתן לחשב את הממוצע וסטיית התקן של כל אחד מהמשתנים, x ו- y , בנפרד. אך מדד חשוב נוסף הוא השונות המשותפת של שני המשתנים, שמוגדרת בתור:

$$\begin{aligned} \text{cov}(x, y) &= \frac{1}{N} \sum_i (x_i - \bar{x})(y_i - \bar{y}) \\ &= \overline{(x - \bar{x})(y - \bar{y})} = \overline{xy} - \bar{x}\bar{y} \end{aligned}$$

אם ערכים של x שהם מעל לממוצע נוטים להופיע ביחד עם ערכים של y שגם הם מעל לממוצע (ולהיפך), המכפלה בתוך הסכום תהיה חיובית ובסך הכל השונות המשותפת תהיה גדולה מאפס. אם לעומת זאת, ערכים גדולים של x נוטים להופיע ביחד עם ערכים קטנים של y , המכפלה תהיה שלילית ובסך הכל השונות המשותפת תהיה קטנה מאפס. אם אין קשר בין שני המשתנים, יש סיכוי שווה לכך שערך גדול של x יופיע יחד עם ערך גדול או קטן של y , ובממוצע הסכום יתאפס ונקבל שאין שונות משותפת.

(דוגמה: השונות המשותפת של גובה ומשקל בקרב בני אדם תהיה חיובית, כי ככל שאדם גבוה יותר כך הוא גם שוקל יותר. באופן דומה השונות המשותפת של משקל וסיבולת לב-ריאה תהיה שלילית, ואילו השונות המשותפת של משקל ו-IQ תהיה כנראה אפס, משום שאין קשר בין שני הדברים.)

השונות המשותפת היא הכללה של השונות הרגילה, שכן מתקיים:

$$\text{cov}(x, x) = V(x)$$

[15] מתאם (Correlation):

גרסה שונה לשונות המשותפת היא מתאם, שמוגדרת בתור:

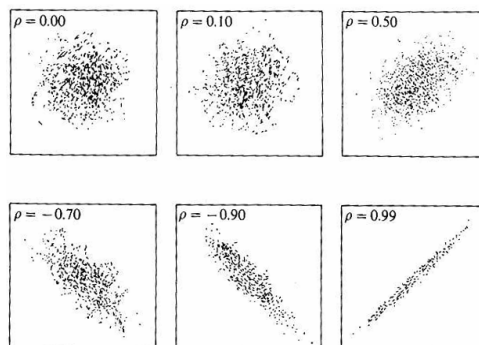
$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

זהו גודל חסר יחידות, הנע בין -1 ל- $+1$.

מתאם של 0 משמעותו שאין קשר בין שני המשתנים. מתאם חיובי/שלילי אומר שעבור ערך גבוה מהממוצע של x , סביר יותר שנקבל גם ערך גבוה/נמוך מהממוצע של y . אם המתאם הוא בדיוק

± 1 , הרי ששני המשתנים הם למעשה זהים, ובהינתן ערך x_i ניתן לקבוע במדויק את y_i .

דוגמה:



[17] מספר משתנים:

בהינתן אוסף נתונים שמורכב מכמה משתנים:

$$(x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)})$$

מגדירים את מטריצת השונות (Variance matrix):

$$V_{ij} = \text{cov}(x_{(i)}, x_{(j)})$$

(נקראת גם Covariance matrix או Error matrix). הערכים באלכסון הם השונות של כל משתנה בנפרד. כמו כן מוגדרת מטריצת המתאם (Correlation matrix):

$$\rho_{ij} = \frac{\text{cov}(x_{(i)}, x_{(j)})}{\sigma_i \sigma_j}$$

הערכים במטריצה זו נמצאים בתחום בין -1 ל- $+1$, כאשר ערכי האלכסון הם כולם 1.

התפלגויות תאורטיות

[22] מבוא:

כאשר משתנה r יכול לקבל ערכים בדידים בלבד, מגדירים עבורו את התפלגות ההסתברות (Probability Distribution) לקבלת ערך מסויים של המשתנה r , במובן זה שההסתברות לקבל $r = r_i$ היא $P(r_i)$. מכאן מגדירים את ערך התצפית (Expectation Value) של r :

$$\langle r \rangle = \sum_r r P(r)$$

ובאופן דומה מגדירים את ערך התצפית לכל פונקציה $f(r)$:

$$\langle f \rangle = \sum_r f(r) P(r)$$

כאשר דוגמים ערכים של r מתוך ההתפלגות הנתונה, חוק המספרים הגדולים מבטיח כי:

$$\lim_{N \rightarrow \infty} \bar{f} = \langle f \rangle$$

כאשר N הוא מספר הנתונים במדגם.

ניתן לחבר ערכי תצפית, $\langle f + g \rangle = \langle f \rangle + \langle g \rangle$, ואולם באופן כללי לא מתקיים $\langle fg \rangle = \langle f \rangle \langle g \rangle$ (אלא אם הפונקציות f ו- g הן בלתי תלויות).

עבור משתנים רציפים x מגדירים פונקצית צפיפות הסתברות (Probability Density function). אם נסמנה בתור $P(x)$, ההסתברות לקבל ערך של x בקטע $(x, x + dx)$ היא $P(x) dx$. למשתנים כאלה מגדירים את ערכי התצפית באופן דומה אך מעט שונה:

$$\langle x \rangle = \int_{-\infty}^{\infty} x P(x) dx$$

$$\langle f \rangle = \int_{-\infty}^{\infty} f(x) P(x) dx$$

לפונקציה $P(x)$ מימדים של $1/x$.

[24] התפלגות בינומית (Binomial Distribution):

מתארת תהליכים עם מספר ידוע של ניסיונות זהים, שלכל אחד שתי תוצאות - הצלחה וכישלון. מסמנים את ההסתברות להצלחה בניסיון בודד בתור p , וההסתברות לכישלון $q = (1 - p)$. ההתפלגות מתארת את ההסתברות ל- r הצלחות מתוך n הניסיונות, והיא:

$$P(r; p, n) = p^r (1 - p)^{n-r} \frac{n!}{r!(n-r)!}$$

הסבר: הפקטור $p^r (1-p)^{n-r}$ מתאר את ההסתברות לקבל בדיוק r הצלחות (ולפיק, בדיוק $n-r$ כשלונות), ולכך עלינו להוסיף את מספר האפשרויות השונות לבחור באילו מתוך n הניסיונות נקבל את r הצלחות:

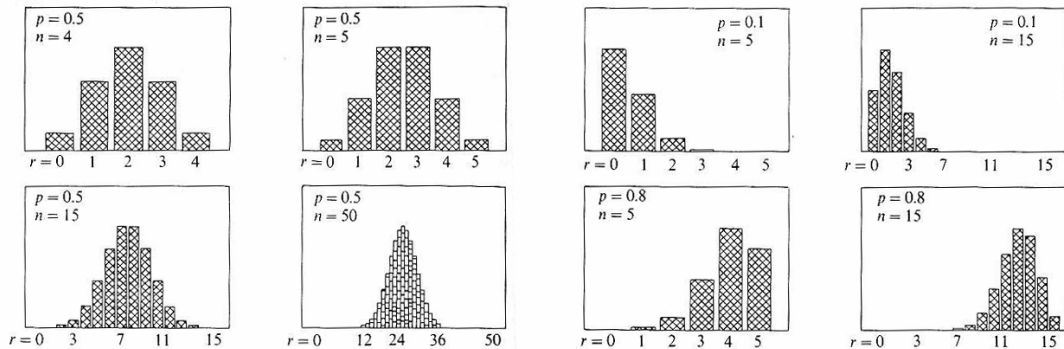
$${}_n C_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

מאפיינים חשובים של ההתפלגות הבינומית:

$$\langle r \rangle = np$$

$$V(r) = np(1-p)$$

$$\sigma = \sqrt{np(1-p)}$$



[28] התפלגות פואסון (Poisson Distribution):

מתארת מצבים שבהם סופרים אירועים דיסקרטיים, אך כאשר אין משמעות ל"לא-אירוע". זאת בניגוד להתפלגות בינומית, שם יש משמעות להצלחה ולכישלון. למשל, התפלגות זו מתאימה לתיאור מספר הטיפות בסופת גשמים.

ניתן לקבל את התפלגות פואסון מתוך ההתפלגות הבינומית, בגבול שבו מספר הניסיונות n שואף לאינסוף, ההסתברות להצלחה בודדת p שואף לאפס, אך המכפלה $\lambda = np$ נשמרת קבועה. בגבול זה מקבלים את הנוסחה של התפלגות פואסון:

$$P(r; \lambda) = \frac{e^{-\lambda} \lambda^r}{r!}$$

שמתארת את הסיכוי לצפות ב- r אירועים בפרק זמן כלשהו, כאשר בממוצע ישנם λ אירועים כאלו באותו פרק זמן.

מאפיינים חשובים של התפלגות פואסון:

$$\langle r \rangle = \lambda$$

$$V(r) = \lambda$$

$$\sigma = \sqrt{\lambda}$$

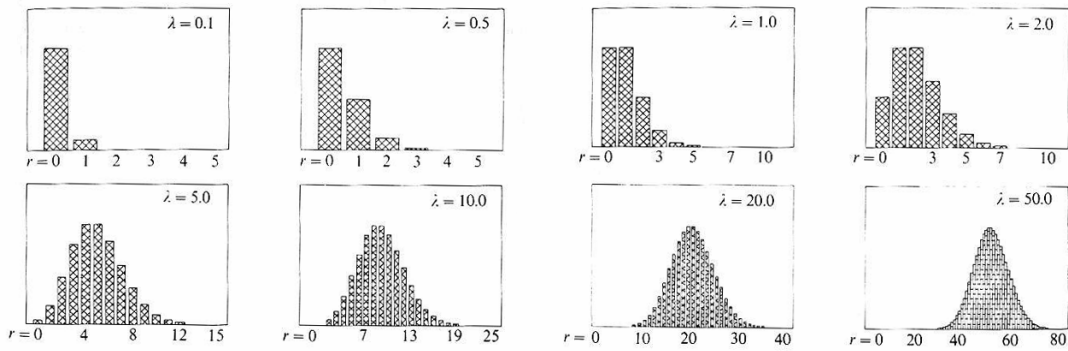
עבור $\lambda < 1.0$, התוצאה המסתברת ביותר היא $r = 0$. לערכים גדולים יותר נוצר שיא אחר, ואולם הוא תמיד נמוך מעט מהממוצע (λ) עצמו. בפרט נשים לב שעבור λ שלם, מתקבלת אותה הסתברות ל- $r = \lambda - 1$ ו- $r = \lambda$, והמקסימום נמצא ביניהם.

התפלגות פואסון משמשת קירוב טוב להתפלגות הבינומית, כאשר מספר הניסיונות הוא גדול מאוד ו/או כאשר ההסתברות להצלחה קטנה מאוד (או לחילופין, כאשר ההסתברות להצלחה קרובה מאוד לאחד, ואז כישלון ייחשב בעינינו כ"אירוע").

כאשר ייתכנו שני סוגי אירועים, a ו- b , עם ממוצעים λ_a ו- λ_b בהתאמה, מספר האירועים הכולל גם כן מתפלג פואסונית, עם ממוצע כולל של $\lambda = \lambda_a + \lambda_b$ (למעשה, ניתן להראות שהדבר נכון גם ליותר משני סוגי אירועים). כמו כן, בהינתן שנצפו r אירועים בסך הכל, ההסתברות ש- r_a

מתוכם היו אירועים מסוג a מתפלגת בינומית עם $p = \lambda_a / (\lambda_a + \lambda_b)$ ו- $n = r$.

דוגמאות: בעמוד הבא.

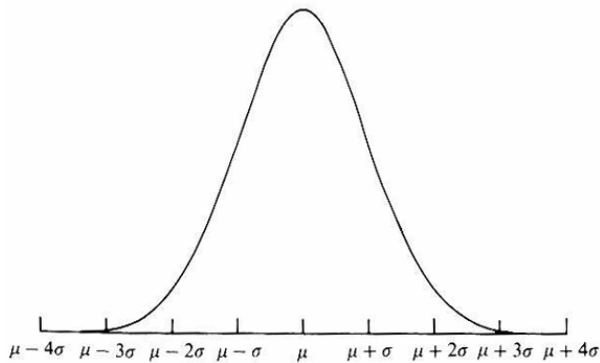


[34] התפלגות גאוסיאנית (Gaussian Distribution):

נקראת גם התפלגות נורמלית. היא מתוארת על ידי פונקציית צפיפות ההסתברות הבאה:

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

ההתפלגות ממורכזת סביב $x = \mu$ וסימטרית ביחס לנקודה זו. סטיית התקן שלה היא σ , וככל σ גדול/קטן יותר כך ההתפלגות רחבה/צרה יותר. בנקודות $x = \mu \pm \sigma$, הפונקציה $P(x)$ יורדת ל-0.61 מערכה המקסימלי. נקודות אלה הן גם נקודות הפיתול של הפונקציה.



נוח להגדיר את גאוסיאן היחידה, המתקבל על ידי החלפת משתנים $z = (x - \mu) / \sigma$

$$P(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

זוהי התפלגות עם תוחלת 0 וסטיית תקן 1. מאפיינים חשובים של ההתפלגות הנורמלית:

$$\langle x \rangle = \int_{-\infty}^{\infty} xP(x; \mu, \sigma) dx = \mu$$

$$V(x) = \int_{-\infty}^{\infty} (x - \mu)^2 P(x; \mu, \sigma) dx = \sigma^2$$

נהוג לדבר על אחוז השטח של הגאוסיאן שנמצא במרחק של כ-וכך σ מהממוצע μ . שטח זה מייצג, כמובן, את ההסתברות שמדידה תניב תוצאה בתחום הזה. מספרים נפוצים:

90% $\rightarrow 1.645\sigma$	$\sigma \rightarrow 68.27\%$
95% $\rightarrow 1.960\sigma$	$2\sigma \rightarrow 95.45\%$
99% $\rightarrow 2.576\sigma$	$3\sigma \rightarrow 99.73\%$
99.9% $\rightarrow 3.290\sigma$	$3.5\sigma \rightarrow 99.95\%$

ובאופן מעשי מתייחסים ל- 2σ כ-95%. טבלאות חישוב לגאוסיאן היחידה: בעמוד הבא.

ONE-TAILED GAUSSIAN INTEGRAL
Giving the percentage probability that a point lies within the given number of σ to one side of the mean



TWO-TAILED GAUSSIAN INTEGRAL
Giving the percentage probability that a point lies within the given number of σ from the mean



$\frac{x-\mu}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	50.00	50.40	50.80	51.20	51.60	51.99	52.39	52.79	53.19	53.59
0.10	53.98	54.38	54.78	55.17	55.57	55.96	56.36	56.75	57.14	57.53
0.20	57.93	58.32	58.71	59.10	59.48	59.87	60.26	60.64	61.03	61.41
0.30	61.79	62.17	62.55	62.93	63.31	63.68	64.06	64.43	64.80	65.17
0.40	65.54	65.91	66.28	66.64	67.00	67.36	67.72	68.08	68.44	68.79
0.50	69.15	69.50	69.85	70.19	70.54	70.88	71.23	71.57	71.90	72.24
0.60	72.57	72.91	73.24	73.57	73.89	74.22	74.54	74.86	75.17	75.49
0.70	75.80	76.11	76.42	76.73	77.04	77.34	77.64	77.94	78.23	78.52
0.80	78.81	79.10	79.39	79.67	79.95	80.23	80.51	80.78	81.06	81.33
0.90	81.59	81.86	82.12	82.38	82.64	82.89	83.15	83.40	83.65	83.89
1.00	84.13	84.38	84.61	84.85	85.08	85.31	85.54	85.77	85.99	86.21
1.10	86.43	86.65	86.86	87.08	87.29	87.49	87.70	87.90	88.10	88.30
1.20	88.49	88.69	88.88	89.07	89.25	89.44	89.62	89.80	89.97	90.15
1.30	90.32	90.49	90.66	90.82	90.99	91.15	91.31	91.47	91.62	91.77
1.40	91.92	92.07	92.22	92.36	92.51	92.65	92.79	92.92	93.06	93.19
1.50	93.32	93.45	93.57	93.70	93.82	93.94	94.06	94.18	94.29	94.41
1.60	94.52	94.63	94.74	94.84	94.95	95.05	95.15	95.25	95.35	95.45
1.70	95.54	95.64	95.73	95.82	95.91	95.99	96.08	96.16	96.25	96.33
1.80	96.41	96.49	96.56	96.64	96.71	96.78	96.86	96.93	96.99	97.06
1.90	97.13	97.19	97.26	97.32	97.38	97.44	97.50	97.56	97.61	97.67
2.00	97.72	97.78	97.83	97.88	97.93	97.98	98.03	98.08	98.12	98.17
2.10	98.21	98.26	98.30	98.34	98.38	98.42	98.46	98.50	98.54	98.57
2.20	98.61	98.64	98.68	98.71	98.75	98.78	98.81	98.84	98.87	98.90
2.30	98.93	98.96	98.98	99.01	99.04	99.06	99.09	99.11	99.13	99.16
2.40	99.18	99.20	99.22	99.25	99.27	99.29	99.31	99.32	99.34	99.36
2.50	99.38	99.40	99.41	99.43	99.45	99.46	99.48	99.49	99.51	99.52
2.60	99.53	99.55	99.56	99.57	99.59	99.60	99.61	99.62	99.63	99.64
2.70	99.65	99.66	99.67	99.68	99.69	99.70	99.71	99.72	99.73	99.74
2.80	99.74	99.75	99.76	99.77	99.78	99.79	99.79	99.80	99.81	99.81
2.90	99.81	99.82	99.82	99.83	99.84	99.84	99.85	99.85	99.86	99.86
3.00	99.87	99.87	99.87	99.88	99.88	99.89	99.89	99.89	99.90	99.90
3.10	99.90	99.91	99.91	99.92	99.92	99.92	99.92	99.92	99.93	99.93
3.20	99.93	99.93	99.94	99.94	99.94	99.94	99.94	99.95	99.95	99.95
3.30	99.95	99.95	99.95	99.96	99.96	99.96	99.96	99.96	99.96	99.97
3.40	99.97	99.97	99.97	99.97	99.97	99.97	99.97	99.97	99.97	99.98
3.50	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98

$\frac{x-\mu}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.00	0.80	1.60	2.39	3.19	3.99	4.78	5.58	6.38	7.17
0.10	7.97	8.76	9.55	10.34	11.13	11.92	12.71	13.50	14.28	15.07
0.20	15.85	16.63	17.41	18.19	18.97	19.74	20.51	21.28	22.05	22.82
0.30	23.58	24.34	25.10	25.86	26.61	27.37	28.12	28.86	29.61	30.35
0.40	31.08	31.82	32.55	33.28	34.01	34.73	35.45	36.16	36.88	37.59
0.50	38.29	38.99	39.69	40.39	41.08	41.77	42.45	43.13	43.81	44.48
0.60	45.15	45.81	46.47	47.13	47.78	48.43	49.07	49.71	50.35	50.98
0.70	51.61	52.23	52.85	53.46	54.07	54.67	55.27	55.87	56.46	57.05
0.80	57.63	58.21	58.78	59.35	59.91	60.47	61.02	61.57	62.11	62.65
0.90	63.19	63.72	64.24	64.76	65.28	65.79	66.29	66.80	67.29	67.78
1.00	68.27	68.75	69.23	69.70	70.17	70.63	71.09	71.54	71.99	72.43
1.10	72.87	73.30	73.73	74.15	74.57	74.99	75.40	75.80	76.20	76.60
1.20	76.99	77.37	77.75	78.13	78.50	78.87	79.23	79.59	79.95	80.29
1.30	80.64	80.98	81.32	81.65	81.98	82.30	82.62	82.93	83.24	83.55
1.40	83.85	84.15	84.44	84.73	85.01	85.29	85.57	85.84	86.11	86.38
1.50	86.64	86.90	87.15	87.40	87.64	87.89	88.12	88.36	88.59	88.82
1.60	89.04	89.26	89.48	89.69	89.90	90.11	90.31	90.51	90.70	90.90
1.70	91.09	91.27	91.46	91.64	91.81	91.99	92.16	92.33	92.49	92.65
1.80	92.81	92.97	93.12	93.28	93.42	93.57	93.71	93.85	93.99	94.12
1.90	94.26	94.39	94.51	94.64	94.76	94.88	95.00	95.12	95.23	95.34
2.00	95.45	95.56	95.66	95.76	95.86	95.96	96.06	96.15	96.25	96.34
2.10	96.43	96.51	96.60	96.68	96.76	96.84	96.92	97.00	97.07	97.15
2.20	97.22	97.29	97.36	97.43	97.49	97.56	97.62	97.68	97.74	97.80
2.30	97.86	97.91	97.97	98.02	98.07	98.12	98.17	98.22	98.27	98.32
2.40	98.36	98.40	98.45	98.49	98.53	98.57	98.61	98.65	98.69	98.72
2.50	98.76	98.79	98.83	98.86	98.89	98.92	98.95	98.98	99.01	99.04
2.60	99.07	99.09	99.12	99.15	99.17	99.20	99.22	99.24	99.26	99.29
2.70	99.31	99.33	99.35	99.37	99.39	99.40	99.42	99.44	99.46	99.47
2.80	99.49	99.50	99.52	99.53	99.55	99.56	99.58	99.59	99.60	99.61
2.90	99.63	99.64	99.65	99.66	99.67	99.68	99.69	99.70	99.71	99.72
3.00	99.73	99.74	99.75	99.76	99.76	99.77	99.78	99.79	99.79	99.80
3.10	99.81	99.81	99.82	99.83	99.83	99.84	99.84	99.85	99.85	99.86
3.20	99.86	99.87	99.87	99.88	99.88	99.88	99.89	99.89	99.90	99.90
3.30	99.90	99.91	99.91	99.91	99.92	99.92	99.92	99.92	99.93	99.93
3.40	99.93	99.94	99.94	99.94	99.94	99.94	99.94	99.95	99.95	99.95
3.50	99.95	99.96	99.96	99.96	99.96	99.96	99.96	99.96	99.97	99.97

ניתן להוכיח כי ההתפלגות הבינומית שואפת לגאוסיאן, כאשר $\mu = np$ ו- $\mu = \sqrt{np(1-p)}$. עבור n קטן הדבר נכון בעיקר ל- $p \approx 0.5$, כאשר עבור ערכים קיצוניים יותר של p דרושים ערכים גדולים יותר של n . באופן דומה, גם ההתפלגות פואסון שואפת לגאוסיאן עבור λ גדול, כאשר $\mu = \lambda$ ו- $\sigma = \sqrt{\lambda}$. למעשה, כבר עבור $\lambda = 10$ התפלגות פואסון דומה מאוד לגאוסיאן. מגדירים גם גאוסיאן רב-מימדי, עבור:

$$\mathbf{x} = (x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)})$$

$$\boldsymbol{\mu} = (\mu_{(1)}, \mu_{(2)}, \mu_{(3)}, \dots, \mu_{(n)})$$

צפיפות ההסתברות היא פונקציה של הוקטור \mathbf{x} :

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{|\mathbf{V}|}} \exp\left[-\frac{1}{2}(\mathbf{x}^t - \boldsymbol{\mu}^t) \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

כאשר \mathbf{V} היא מטריצת השונות המשותפת לכל המשתנים.

[45] התפלגות אחידה (Uniform Distribution):

התפלגות זו קבועה ואחידה בתחום מסויים ושווה לאפס מחוצה לו, כלומר:

$$P(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

מאפיינים חשובים:

$$\langle x \rangle = \frac{a+b}{2}$$

$$V(x) = \frac{(b-a)^2}{12}$$

$$\sigma = \frac{b-a}{\sqrt{12}}$$

שגיאות

[49] משפט הגבול המרכזי (Central Limit Theorem):

הסכום X של N משתנים בלתי תלויים x_i , כאשר כל אחד מהם מגיע מתוך התפלגות עם ממוצע μ_i ושונות σ_i^2 מתפלג גאוסיאנית כאשר $N \rightarrow \infty$. כמו כן, ההתפלגות של X מקיימת:

$$\langle X \rangle = \sum_i \mu_i$$
$$V(X) = \sum_i \sigma_i^2$$

זו הסיבה ששגיאות בניסויים, שמגיעות מתוך הרבה מאוד גורמים בלתי תלויים, מתנהגות לכל צורך מעשה לפי התפלגות נורמלית.

[51] מדידות חוזרות:

נניח שמבצעים מספר מדידות בלתי תלויות של אותו גודל באותה שיטת מדידה, כך שהתוחלת והשונות של כל המדידות הן זהות. מתוך משפט הגבול המרכזי, עם $\mu_i = \mu$ ו- $\sigma_i = \sigma$, מקבלים:

$$\langle X \rangle = \sum_i \mu = N\mu$$
$$V(X) = \sum_i \sigma^2 = N\sigma^2$$

ובמונחים של הממוצע של המדידות, $\bar{x} = X/N$,

$$\langle \bar{x} \rangle = \mu$$

$$V(\bar{x}) = \frac{\sigma^2}{N}$$

השגיאה על הממוצע היא לפיכך σ/\sqrt{N} , וזהו מדד לכמה טוב אנו יודעים את התוחלת (μ) של התפלגות האם שממנה נלקחו המדידות. בפרט היא דועכת כמו $1/\sqrt{N}$, כאשר N הוא מספר המדידות.

[53] שקלול מדידות שונות:

נניח כעת כי ביצענו מספר ניסויים שונים למדידת אותו גודל. כלומר, לכל התוצאות אותה תוחלת $\mu_i = \mu$, אך סטיית התקן של כל תוצאה, μ_i , שונה. נרצה לבצע מיצוע של התוצאות שקיבלנו, אך עם חשיבות גדולה יותר לתוצאות עם סטיית תקן קטנה יותר. הנוסחה המדוייקת היא:

$$\bar{x} = \frac{\sum_i x_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2}$$

השונות של הממוצע הנ"ל (או: ריבוע השגיאה עליו) נתונה על ידי:

$$V(\bar{x}) = \frac{1}{\sum_i 1 / \sigma_i^2}$$

[55] חיבור שגיאות:

השגיאה על ערך הפונקציה $f(x_1, x_2, \dots, x_n)$, כאשר המשתנים x_i הם בלתי תלויים וידועים בדיוק של σ_i , נתונה על ידי:

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x_1} \right)^2 \sigma_1^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \sigma_2^2 + \dots + \left(\frac{\partial f}{\partial x_n} \right)^2 \sigma_n^2$$

באופן הכללי ביותר, ניתן להתייחס ל- m פונקציות f_1, f_2, \dots, f_m ב- n משתנים x_1, x_2, \dots, x_n , לאו דווקא בלתי תלויים. נרצה לדעת, בהינתן השונות של כל משתנה בנפרד וכן המתאם בין כל זוג משתנים, מהי השונות של כל פונקציה בנפרד ומהו המתאם בין כל שתיים מהן. התוצאה: בעמוד הבא.

$$V(f_i) = \sum_j \left(\frac{\partial f_i}{\partial x_j} \right)^2 V(x_j) + \sum_j \sum_{k \neq j} \left(\frac{\partial f_i}{\partial x_j} \right) \left(\frac{\partial f_i}{\partial x_k} \right) \text{cov}(x_j, x_k)$$

$$\text{cov}(f_k, f_\ell) = \sum_i \sum_j \left(\frac{\partial f_k}{\partial x_i} \right) \left(\frac{\partial f_\ell}{\partial x_j} \right) \text{cov}(x_i, x_j)$$

ניתן לרשום תוצאה זו בצורה מטריצית:

$$\mathbf{V}_f = \mathbf{G} \mathbf{V}_x \mathbf{G}^t$$

כאשר $\mathbf{V}_x, \mathbf{V}_f$ הן מטריצות השונות של הוקטורים:

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

$$\mathbf{f} = (f_1, f_2, \dots, f_m)$$

מסדר $n \times n, m \times m$ בהתאמה, ו- \mathbf{G} היא מטריצת הנגזרות המוגדרת לפי $G_{ki} = \partial f_k / \partial x_i$, או:

$$\mathbf{G} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

המטריצה \mathbf{V}_f נותנת לנו מידע הן לגבי השונות של כל פונקציה f_i בנפרד, והן לגבי המתאם בין כל זוג פונקציות f_i, f_j . התוצאה תלויה בשונות ובמתאמים של ה- x_i השונים.

[61] שגיאות שיטתיות:

על מנת לטפל בשגיאות שיטתיות, כל שעלינו לדעת הוא כיצד לשלב אותן במטריצת השונות \mathbf{V}_x . לצורך כך ניקח דוגמה עם שני משתנים בלבד, x_1 ו- x_2 , שלכל אחד מהם שגיאה סטטיסטית בשיעור σ_1, σ_2 , ולשניהם אותה שגיאה שיטתית S . במקרה כזה נקבל:

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 + S^2 & S^2 \\ S^2 & \sigma_2^2 + S^2 \end{pmatrix}$$

ומכאן ניתן להמשיך בטיפול הרגיל של שגיאות. כך למשל, השגיאה על ההפרש תהיה:

$$\sigma_{\text{diff}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 + S^2 & S^2 \\ S^2 & \sigma_2^2 + S^2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \sigma_1^2 + \sigma_2^2$$

כלומר, השגיאה השיטתית, שהייתה זהה בשתי המדידות, התבטלה.

מעריכים

[69] תכונות המעריך:

בהגדרה הכללית ביותר, מעריך (estimator) הוא התוצאה של תהליך כלשהו שמבצעים על אוסף מדידות, שבסופו מקבלים ערך מספרי למאפיין כלשהו של התפלגות האם ממנה הגיעו המדידות. אם הגודל שאותו אנו מנסים למדוד הוא a , נסמן את המעריך שלו בתור \hat{a} . מעריך ייקרא עקבי (consistent) אם מתקיים:

$$\lim_{N \rightarrow \infty} \hat{a} = a$$

כאשר N הוא מספר המדידות שעל פיהם אנו קובעים המעריך. כלומר, נרצה שככל שניקח יותר מדידות, כך המעריך יתקרב לערך האמיתי שאותו אנו רוצים למדוד. מעריך ייקרא בלתי-משוחד (unbiased) אם מתקיים:

$$\langle \hat{a} \rangle = a$$

כלומר, נרצה שערך התוחלת של המעריך יהיה הערך האמיתי שלו; שבממוצע, המעריך שנקבל מתוך המדידות יהיה לפעמים גדול יותר ולפעמים קטן יותר מהגודל שאותו אנו מודדים, אך שלא תהיה סטייה קבועה לכיוון מסויים.

מעריך ייקרא יעיל (efficient) אם השונות שלו קטנה. נרצה שבמדידה ספציפית, הערך שנקבל עבור המעריך יהיה קרוב ככל הניתן לערך התוחלת שלו, ועל כל נבחר מעריך עם התפלגות צרה ככל שניתן. מעריך עקבי, בלתי-משוחד ויעיל נקרא מעריך טוב.

[71] פונקציית ההיתכנות (Likelihood Function):

נניח כי לוקחים מדידות מתוך התפלגות אם שמתוארת על ידי הפונקציה $P(x; a)$, כאשר a הוא מאפיין כלשהו של ההתפלגות שאנו מעוניינים למדוד. הפונקציה:

$$L(x_1, x_2, \dots, x_N; a) = P(x_1; a)P(x_2; a) \cdots P(x_N; a) = \prod_{i=1}^N P(x_i; a)$$

מתארת את ההסתברות שאם ניקח דגימה של N מדידות מתוך התפלגות האם הזו, נקבל בדיוק את אוסף הנתונים $\{x_1, x_2, \dots, x_N\}$.

לפיכך ניתן לחשב את ערך התצפית לכל פונקציה של המשתנים $\{x_i\}$ בעזרת:

$$\langle f(x_1, \dots, x_N) \rangle = \int \cdots \int f(x_1, \dots, x_N) L(x_1, \dots, x_N; a) dx_1, \dots, dx_N$$

או בקיצור:

$$\langle f(X) \rangle = \int f(X) L(X; a) dX$$

ובפרט:

$$\langle \hat{a} \rangle = \int \hat{a} L dX, \quad \langle \hat{a}^2 \rangle = \int \hat{a}^2 L dX$$

ניתן להוכיח שעל כל מעריך בלתי-משוחד קיים חסם תחתון (Minimum Variance Bound):

$$\text{MVB} = \frac{1}{\langle (d \ln L / da)^2 \rangle} = \frac{-1}{\langle (d^2 \ln L / da^2) \rangle}$$

כך ש- $V(\hat{a}) \geq \text{MVB}$. היעילות (efficiency) של מעריך מוגדרת בתור:

$$\text{MVB} / V(\hat{a})$$

ומעריך נקרא יעיל כאשר היעילות שלו היא 1, כלומר השונות שלו היא המינימלית.

[76] דוגמאות למעריכים:

מעריך טוב לממוצע האמיתי של התפלגות האם (mean) הוא הממוצע של המדידות (average):

$$\hat{\mu} = \bar{x}$$

$$V(\hat{\mu}) = \frac{\sigma^2}{N}$$

כאשר σ היא סטיית התקן של התפלגות האם. מעריך זה הוא אינו בהכרח היעיל ביותר, פרט במקרה של התפלגות נורמלית, שם ניתן להוכיח כי אכן $\text{MVB} = \sigma^2 / N$. ניתן למצוא מעריך גם לשונות של התפלגות האם. ראשית, כאשר μ ידוע:

$$\hat{V}(x) = \frac{1}{N} \sum_i (x_i - \mu)^2$$

$$V(\hat{V}(x)) = \frac{1}{N} \left[\langle (x - \mu)^4 \rangle - \langle (x - \mu)^2 \rangle^2 \right]$$

וכאשר μ אינו ידוע, משתמשים ב- \bar{x} במקומו:

$$\hat{V}(x) = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

$$V(\hat{V}(x)) = \frac{1}{N-1} \left[\langle (x - \bar{x})^4 \rangle - \langle (x - \bar{x})^2 \rangle^2 \right]$$

התיקון (על שם Bessel) למכנה מוודא שהמעריך לא יהיה משוחד. להתפלגות נורמלית מקבלים:

$$V(\hat{V}(x)) = \frac{2\sigma^4}{N^2}(N-1) \approx \frac{2\sigma^4}{N}$$

ושבו, אם מתחשבים בתיקון בֶּסֶל מחליפים את המכנה ב- $(N-1)$.

[81] שיטת Maximum Likelihood:

בהינתן מדגם נתונים $\{x_1, x_2, \dots, x_N\}$, המעריך \hat{a} הוא הערך של a שמביא למקסימום את L :

$$L(x_1, x_2, \dots, x_N; a) = \prod_{i=1}^N P(x_i; a)$$

כלומר, הערך של a עבורו ההסתברות שנקבל דווקא את אוסף הנתונים שקיבלנו, היא הגבוהה ביותר. בדרך כלל נוח יותר למצוא את המקסימום של $\ln L$:

$$\ln L(x_1, x_2, \dots, x_N; a) = \sum_{i=1}^N P(x_i; a)$$

ניתן לפתור את הבעיה נומרית, אם הנתונים ידועים לנו, או באופן אנליטי לפי המשוואה:

$$\left. \frac{d \ln L}{da} \right|_{a=\hat{a}} = 0$$

ואז הפיתרון יהיה נכון לכל אוסף נתונים שיקרה בדרכנו. מעריכים שמחושבים בשיטה זו הם בדרך כלל עקביים, אך משוחדים. השיחוד מגיע מתוך מאפיין חשוב ושימושי של מעריכים אלו, והוא האינוריאנטיות שלהם:

$$\widehat{f(a)} = f(\hat{a})$$

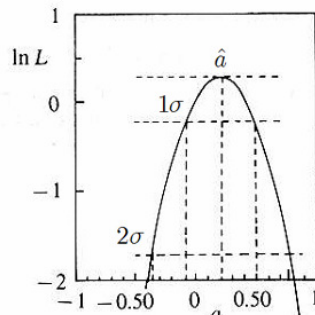
כלומר, אם רוצים להעריך פונקציה כלשהי של a , ניתן להעריך את a בעצמו ואז להפעיל את הפונקציה. מעריכים בשיטות אחרות לא בהכרח מקיימים זאת. עם זאת, בגבול של $N \rightarrow \infty$, השיחוד של המעריך נעלם, וכמו כן המעריך הוא יעיל, כלומר $V(\hat{a}) = \text{MVB}$, ובפרט מתקיים:

$$[V(\hat{a})]^{-1} = \left\langle \frac{d^2 \ln L}{da^2} \right\rangle = \left. \frac{d^2 \ln L}{da^2} \right|_{a=\hat{a}}$$

יתרה מכך, ניתן לקבל את השגיאה על המעריך באמצעות שרטוט גרף של $\ln L$ כפונקציה של a , ומציאת הנקודות שבהם הפונקציה יורדת ב-0.5 מערכה המקסימלי (אבסולוטית; זו לא הנקודה שבה הפונקציה יורדת ל-1/2 מהמקסימום!). בגבול של N גדול, מתפלג נורמלית, ולכן שתי הנקודות הללו יהיו סימטריות. עבור N קטן זה לא בהכרח מתקיים, ונקבל שהגרף יהיה לא סימטרי. במקרה כזה מצטטים את ערך המעריך באופן הבא:

$$a = 0.21^{+0.29}_{-0.27}$$

כמו כן ניתן למצוא את גבול ה- 2σ בנקודות שבהן $\ln L$ יורד ב-2.0 מערכה המקסימלי, וגבול ה- 3σ בנקודות שבהן הגרף יורד ב-4.5. כאשר L אינו מתפלג נורמלית, לא בהכרח נקבל שנקודת ה- 2σ היא במרחק כפול מהשיא מאשר נקודת גבול ה- σ . דוגמה:



ניתן להשתמש בשיטה זו על מנת להעריך את ערכם של מספר פרמטרים $\{a_1, a_2, \dots, a_n\}$. במקרה כזה עלינו לפתור בו-זמנית את מערכת המשוואות:

$$\frac{\partial \ln L(x_1, \dots, x_N; a_1, \dots, a_n)}{\partial a_j} = 0$$

לכל $j = 1, \dots, n$.

מטריצת השונות המשותפת לכל המעריכים ניתנת בצורה הופכית על ידי:

$$\text{cov}^{-1}(a_i, a_j) = \left\langle \frac{\partial \ln L}{\partial a_i} \frac{\partial \ln L}{\partial a_j} \right\rangle = - \left\langle \frac{\partial^2 \ln L}{\partial a_i \partial a_j} \right\rangle = - \frac{\partial^2 \ln L}{\partial a_i \partial a_j} \Big|_{a=\hat{a}}$$

כלומר, מטריצת השונות המשותפת היא ההופכית למטריצת הנגזרות. לא מתקיים:

~~$$\text{cov}(a_i, a_j) = \left\langle \frac{\partial \ln L}{\partial a_i} \frac{\partial \ln L}{\partial a_j} \right\rangle^{-1}$$~~

[93] דגימת שכבות (Stratified Sampling):

נראה כיצד ניתן להקטין את השגיאה על מעריכים בעזרת חלוקה נכונה של מדגם הנתונים. נניח כי אוכלוסיית המדגם מגיעה משתי התפלגויות שונות, $P_1(x)$ ו- $P_2(x)$, שנועות סביב μ_1 ו- μ_2 בהתאמה, ומכל התפלגות אנו לוקחים נתונים באחוזים של f_1 ו- f_2 (כך ש- $f_1 + f_2 = 1$). הממוצע של ההתפלגות המשוקללת יהיה $\mu = f_1\mu_1 + f_2\mu_2$, עם שונות של:

$$\begin{aligned} V(x) &= \int [f_1 P_1(x) + f_2 P_2(x)] (x - \mu)^2 dx \\ &= f_1 V_1(x) + f_2 V_2(x) + f_1 f_2 (\mu_1 - \mu_2)^2 \end{aligned}$$

ניתן להיפטר מהאיבר האחרון על ידי בחירה מושכלת של f_1, f_2 , ובכך להקטין את השגיאה על המעריך שנמצא עבור μ (שהיא, כזכור, $(V(x)/N)^{1/2}$). נשים לב ששיטה זו יעילה אך ורק

כאשר ההפרש $(\mu_1 - \mu_2)$ הוא משמעותי, כלומר כאשר שתי האוכלוסיות שונות באופן מהותי. השיטה טובה למספר כלשהו של אוכלוסיות. נניח שלוקחים m_i מדידות מכל התפלגות, כל אחת עם ממוצע μ_i ושונות σ_i^2 (זו השונות של ההתפלגות, ולא השגיאה על ידיעת μ_i !), כאשר ידוע לנו שבאוכלוסיית המטרה, שאותה אנו מנסים להעריך, התפלגויות אלה מופיעות באחוזים של f_i . המעריך על הממוצע באוכלוסיית המטרה הוא:

$$\mu = f_1\mu_1 + f_2\mu_2 + f_3\mu_3 + \dots$$

והשגיאה על ההערכה הזו נתונה על ידי:

$$\sigma^2 = \frac{f_1^2 \sigma_1^2}{m_1} + \frac{f_2^2 \sigma_2^2}{m_2} + \frac{f_3^2 \sigma_3^2}{m_3} + \dots$$

נדגיש שוב את ההבדל בין σ_i^2 , שהיא השונות של ההתפלגות ה- i , לבין σ^2 , שהיא השגיאה על המעריך שלנו עבור μ באוכלוסיית היעד. נרצה לבחור m_i כאלה שיביאו למינימום את השגיאה הנ"ל, ובעזרת כופלי לגרנז' (תחת האילוץ $\sum_i m_i = N$), מקבלים:

$$m_i \propto f_i \sigma_i$$

עם אותו קבוע פרופורציה לכל i , כלומר:

$$\frac{m_1}{f_1 \sigma_1} = \frac{m_2}{f_2 \sigma_2} = \frac{m_3}{f_3 \sigma_3} = \dots$$

ריבועים מינימליים

[97] מבוא:

נניח כי נתון אוסף מדידות מהצורה (x_i, y_i) , כאשר x_i ידוע בוודאות ו- y_i נמדד עם שגיאה σ_i . אנו משערים כי מתקיים הקשר $y = f(x; a)$, כאשר a הוא הפרמטר אותו אנו מעוניינים להעריך מתוך המדידות. בהנחה שהשגיאה על המדידות היא נורמלית, ניתן לרשום:

$$P(y_i; a) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-[y_i - f(x_i; a)]^2 / 2\sigma_i^2}$$

מתוך שיטת ה-Maximum Likelihood מגלים שהמעריך הוא הערך של a שעבורו הגודל:

$$\chi^2 = \sum_{i=1}^N \left[\frac{y_i - f(x_i; a)}{\sigma_i} \right]^2$$

הוא מינימלי, ומכאן השם (Least Squares). עלינו לפתור $d\chi^2 / da = 0$, ובאופן מפורש:

$$\sum_i \frac{1}{\sigma_i^2} \frac{df(x_i; a)}{da} [y_i - f(x_i; a)] = 0$$

ממשוואה זו מקבלים את \hat{a} כפונקציה של y_1, \dots, y_N , ומכיוון שהשגיאה על כל אחד מהם ידועה, ניתן מכאן לחשב גם את השגיאה על \hat{a} בעזרת חיבור שגיאות רגיל.

[99] התאמה לקו ישר (Straight Line Fit):

התאמה של אוסף נתונים לקו ישר $y = mx + c$ בשיטה זו נותנת את המעריכים הבאים:

$$\hat{m} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}$$

$$\hat{c} = \bar{y} - \hat{m}\bar{x}$$

עם שגיאות בשיעור של:

$$\sigma_m^2 = V(\hat{m}) = \frac{\sigma^2}{N(\overline{x^2} - \bar{x}^2)}$$

$$\sigma_c^2 = V(\hat{c}) = \frac{\sigma^2 \bar{x}^2}{N(\overline{x^2} - \bar{x}^2)}$$

כאשר σ היא השגיאה על ערכי y_i . השונות המשותפת היא:

$$\text{cov}(\hat{m}, \hat{c}) = -\frac{\sigma^2 \bar{x}}{N(\overline{x^2} - \bar{x}^2)}$$

$$\rho_{\hat{m}, \hat{c}} = -\frac{\bar{x}}{\sqrt{\overline{x^2} - \bar{x}^2}}$$

ו- χ^2 של ההתאמה הוא:

$$\chi^2 = N \frac{V(y)}{\sigma^2} (1 - \rho_{x,y}^2)$$

כאשר $V(y), \rho_{x,y}$ מוגדרים על אוסף המדידות $\{y_i\}$. בפרט, $V(y) \neq \sigma^2$! הנוסחה האחרונה שימושית, משום שהיא מאפשרת לנו לדעת מהו χ^2 של ההתאמה לקו ישר מבלי לחשב אותה במפורש.

כעת, בהינתן X כלשהו, ניתן להעריך את $y(X)$ באופן הבא:

$$Y = y(X) = \hat{m}X + \hat{c}$$

השגיאה על ערך זה היא, מתוך חיבור שגיאות:

$$V(Y) = X^2 V(\hat{m}) + V(\hat{c}) + 2X \text{cov}(\hat{m}, \hat{c})$$

ניתן להיפטר מהשונות המשותפת של \hat{m} ו- \hat{c} בעזרת בחירה כזו של מערכת הצירים, כך ש- \bar{x} יתאפס. באופן שקול, ניתן לבצע את ההתאמה לקו הישר הנתון על ידי:

$$y = \hat{m}(x - \bar{x}) + \hat{c}'$$

במקרה כזה, המתאם של \hat{m} עם \hat{c}' נעלם, והשגיאה על \hat{c}' היא פשוט σ / \sqrt{N} .

כאשר ישנה שגיאה שיטתית S על ערכי $\{y_i\}$, היא אינה משפיעה כלל על המעריך לשיפוע \hat{m} ,

אך היא כן משפיעה על נקודת החיתוך \hat{c} , ולשונות של מעריך זה מתווסף איבר S^2 .

[106] התפלגות χ^2 :

כזכור, הגודל χ^2 מוגדר לפי:

$$\chi^2 = \sum_{i=1}^N \left[\frac{y_i - f(x_i; a)}{\sigma_i} \right]^2 = \sum_{i=1}^N \left(\frac{y_i^{\text{actual}} - y_i^{\text{ideal}}}{\text{expected error}} \right)^2$$

באופן כללי, ככל ש- χ^2 קטן יותר, כך ההתאמה לפונקציה טובה יותר. ואולם גם ערכים קטנים מדי של χ^2 לא רצויים, משום שמשמעות הדבר היא שהפרזנו בהערכת השגיאות σ_i . ההתפלגות המדויקת של χ^2 נתונה על ידי:

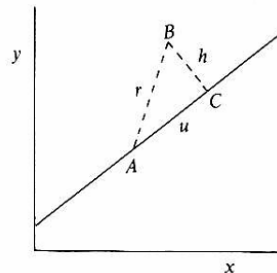
$$P(\chi^2; n) = \frac{2^{-n/2}}{\Gamma(n/2)} \chi^{n-2} e^{-\chi^2/2}$$

זו צפיפות ההסתברות לקבלת ערך מסויים של χ^2 , והיא תלויה במספר דרגות החופש n , ששווה למספר הנתונים N פחות מספר הפרמטרים שאותם אנו מעריכים (כלומר, שבעזרתם הבאנו למינימום את χ^2).

להתפלגות הזו ממוצע n עם שונות של $2n$. לכן, אנו נצפה ש- $\chi^2 \approx n$, ואם קיבלנו ערך גדול בהרבה יש מקום לחשד שההתאמה לא טובה. ההתפלגות הזו אומנם שואפת לגאוסיאן, אך רק עבור n מאוד גדולים. לכן, עבור $n \geq 30$ משתמשים בדרך כלל בפרמטר $\sqrt{2\chi^2}$, שדומה יותר לגאוסיאן עם ממוצע $\sqrt{2n-1}$ ושונות של 1.

[109] שגיאות בשני הצירים:

תסריט אפשרי נוסף הוא שקיימת שגיאה מדידה גם על ערכי x_i וגם על y_i . בתור דוגמה נתייחס למקרה שבו השגיאות הללו זהות, σ , ונעשית התאמה לקו ישר.



אם B היא הנקודה שנמדדה בניסוי, צפיפות ההסתברות שהיא "שייכת" בעצם לנקודה A היא:

$$\begin{aligned} P(A \rightarrow B) &= \frac{1}{2\pi\sigma^2} e^{-(x_A - x_B)^2 / 2\sigma^2} e^{-(y_A - y_B)^2 / 2\sigma^2} \\ &= \frac{1}{2\pi\sigma^2} e^{-u^2 / 2\sigma^2} e^{-h^2 / 2\sigma^2} \end{aligned}$$

וההסתברות שהנקודה B שייכת בכלל לקו הישר הזה, נתונה על ידי:

$$P = \int_{-\infty}^{\infty} P(A \rightarrow B) du = \frac{1}{\sigma\sqrt{2\pi}} e^{-h^2 / 2\sigma^2}$$

h הוא מרחק הנקודה שנמדדה מהישר, ולכן באופן מעשי אנו רואים כי ההסתברות שהנקודה שייכת לקו ישר ספציפי, תלויה רק במרחק שלה ממנו. הנקודה C היא גם הנקודה ש- B שייכת לה בהסתברות הכי גבוהה. מפורשות, המרחק h_i של נקודה (x_i, y_i) מישר $y = mx + c$ הוא:

$$h_i = \frac{y - mx_i - c}{\sqrt{1 + m^2}}$$

ולכן:

$$\chi^2 \propto \sum_i \frac{(y - mx_i - c)^2}{1 + m^2}$$

גזירה ביחס ל- c והשוואה לאפס נותנת את המעריך, כצפוי:

$$\hat{c} = \bar{y} - \hat{m}\bar{x}$$

והמעריך \hat{m} מתקבל מתוך גזירה לפי m :

$$\hat{m} = \frac{\sigma_y}{\sigma_x} \left(A \pm \sqrt{A^2 + 1} \right), \quad A = \frac{\sigma_x^2 V(y) - \sigma_y^2 V(x)}{2\sigma_x \sigma_y \text{cov}(x, y)}$$

באופן משעשע, שני הפתרונות נותנים את הקו הישר הטוב ביותר ואת הקו הרע ביותר (אין דבר כזה קו מכוער). בוחרים את הפיתרון עם סימן הפלוס/מינוס כאשר $\text{cov}(x, y)$ הוא חיובי/שלילי.

רמת וודאות

[125] מבוא:

רמת הוודאות (Confidence Level), כשמה כן היא, מתארת באיזו וודאות אנו יכולים לומר משהו על התוצאות שקיבלנו. למשל, אם ידוע לנו שהמשקל של אוסף ארגזים מסויים מתפלג נורמלית עם ממוצע μ וסטטיית תקן σ , נוכל לומר שב-68% מהמקרים, משקלו של ארגז יהיה בין $\mu - \sigma$ ל- $\mu + \sigma$. ערכים נפוצים שנהוג לציין עבור התפלגות גאוסיאנית:

$$68\%(1\sigma), \quad 95.4\%(2\sigma), \quad 90\%(1.64\sigma), \quad 95\%(1.96\sigma), \quad 99\%(2.58\sigma)$$

להתפלגויות אחרות ניתן לבחור את התחום במספר דרכים שונות. לאחר שבחרנו את רמת הוודאות הדרושה לנו, C , הגבולות x_- ו- x_+ צריכים לקיים:

$$\Pr(x_- \leq x \leq x_+) = \int_{x_-}^{x_+} P(x) dx = C$$

ועל מנת להגדיר לגמרי את הגבולות, יש לבחור באפשרות הרצויה:

1. תחום סימטרי - כך שמתקיים $(x_+ - \mu) = (\mu - x_-)$.
2. תחום מינימלי - כך שהגודל $(x_+ - x_-)$ הוא הקטן ביותר תחת האילוץ לעיל.
3. תחום מרכזי - כך שההסתברויות להיות מעל ומתחת לתחום שוות, כלומר:

$$\int_{-\infty}^{x_-} P(x) dx = \int_{x_+}^{\infty} P(x) dx = \frac{1-C}{2}$$

בדרך כלל בוחרים באפשרות האחרונה, ואולם לכל התפלגות סימטרית (ובכלל זה, הגאוסיאן) שלושת ההגדרות שקולות לחלוטין.

ניתן להגדיר גם רמת וודאות ביחס לגבול עליון או תחתון, שמוגדרים לפי:

$$\Pr(x < x_+) = \int_{-\infty}^{x_+} P(x) dx = C$$

$$\Pr(x > x_-) = \int_{x_-}^{\infty} P(x) dx = C$$

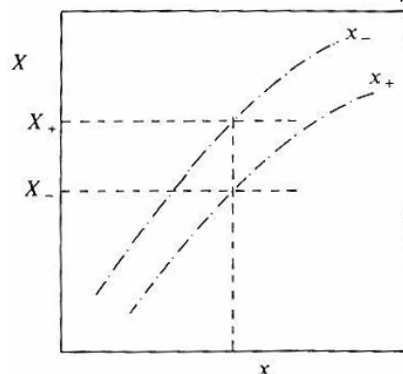
כל זאת עבור משתנים רציפים. כאשר מדובר בגדלים דיסקרטיים, האינטגרלים מוחלפים בסכומים ויש לוודא שערכי הקצה נכללים בסכום כאשר מדובר בתחום $(x_- \leq x \leq x_+)$ ולא נכללים בו כאשר מגדירים גבולות בלבד $(x > x_-$ או $x < x_+)$.

[127] רמת וודאות למעריכים:

נניח כי ביצענו מדידה של גודל מסויים, וקיבלנו תוצאה של x עם שגיאה בשיעור σ . אין זה נכון לומר כי "בוודאות של 68%, הערך האמיתי של הגודל הנמדד הוא בין $x - \sigma$ ל- $x + \sigma$ ", משום שהצהרה כזו עלולה לגרום לבעיות אם, למשל, $x \approx \sigma$, ואז קיים סיכוי שהגודל שניסינו למדוד הוא שלילי. ואם רצינו למדוד מסה, כמובן שהדבר לא הגיוני.

מה כן נוכל לומר? נניח כי ניסינו למדוד את הגודל X . אזי קיימת התפלגות מסויימת $P(x; X)$ שנותנת את ההסתברות למדידת ערך של x , בהנחה שהגודל שאנו מודדים הוא למעשה X . אם מדובר במדידה רגילה במעבדה עם שגיאה σ , הפונקציה תהיה גאוסיאן עם ממוצע X ושונות של σ^2 ; אם מדובר במספר אירועים שנצפה, זו תהיה התפלגות פואסון עם אותו ממוצע, X . באופן כללי, לפונקציה יהיה שיא בסביבת $x = X$, והיא תדעך לאפס משני צידיה.

מתוך הפונקציה הזו ניתן להגדיר תחום ערכים ברמת וודאות רצויה (נניח 90%), כך שעבור ערך מסוים של X , מדידת x תניב ערך שנמצא בין x_- ל- x_+ בהסתברות של 90%. לכל ערך שונה של X נקבל תחום אחר, כך שניתן לחשוב על הגבולות x_+, x_- כעל פונקציות של X . נוח להשתמש בגרף שבנוי באופן הבא:



את הגרף הנ"ל יוצרים מתוך שרטוט של העקומים x_+ ו- x_- כפונקציה של X , רק מתוך ידיעת פונקציות ההתפלגות $P(x; X)$. לאחר ביצוע המדידה יש לנו ערך של x ביד, וממנו מקבלים את הגדלים X_+ ו- X_- , כפי שמתואר בתרשים.

מהי משמעות הערכים הללו? בהינתן מדידה x , מתוך העקום של x_- מוצאים את X_+ , שעבורו x -ה- x_- שנמדד הוא x_- . זהו הגבול העליון על X , במובן זה שאם הערך האמיתי של X הוא גדול ממנו, ההסתברות לקבל במדידה את x או מספר נמוך יותר, היא 5%. באופן דומה, X_- , שנקבע מתוך העקום של x_+ , הוא הגבול התחתון על X , שכן אם הערך האמיתי שלו הוא קטן מ- X_- , ההסתברות לקבל במדידה x או יותר היא 5%.

לפיכך, אנו אומרים כי התחום של X הוא בין X_- ל- X_+ ברמת וודאות של 90%. מהצהרה אחת כזו לא ניתן ללמוד דבר - ייתכן התוצאה שקיבלנו היא בכל זאת לא נכונה. אך אם ניקח הרהב מאוד הצהרות כאלו, ברמת וודאות של 90% כל אחת, הרי ש-90% מהן יהיו נכונות. עבור התפלגות נורמלית, העקומים הם למעשה ישרים הנתונים על ידי:

$$x_{\pm} = X \pm n\sigma$$

כאשר n תלוי ברמת הוודאות הרצויה. עבור 68% ניקח $n = 1$, ל-90% ניקח $n = 1.64$ וכו'. התחום המתקבל הוא $X_{\pm} = x \pm n\sigma$, כאשר x הוא ערך המדידה שביצענו.

[130] מדידת גודל חסום:

כאשר מודדים, למשל, מסה של גוף מסויים, ברור לחלוטין כי התוצאה צריכה להיות חיובית - אין משמעות למסה שלילית. בתור דוגמה, נניח כי לרשותנו עומד גוף עם מסה של 0.1g (שידועה לנו מגורמים אחרים), אנו מודדים את המסה שלו בעזרת משקל עם שגיאת מדידה של 0.2g, ומעוניינים לתת תוצאה ברמת וודאות של 95.4%, כלומר 2σ . אם נמדוד ערך x למסה, נאמר שהתחום הוא $x \pm 0.4g$.

ישנו סיכוי של 2.3% שהמדידה תניב תוצאה של 0.5g ומעלה. במקרה כזה, התחום שנקבל לא כולל בתוכו את המסה האמיתית של הגוף, וההצהרה שניתן תהיה שקרית. אפשרות זו מתקבלת על הדעת ברמת הוודאות שבחרנו, ואין מה לעשות נגדה.

אם נמדוד מסה בין 0.4g ל-0.5g נקבל תחום דומה, אך הפעם הוא יכלול את המסה האמיתית של הגוף, 0.1g. כלומר, ההצהרה שלנו תהיה מהימנה. אם נמדוד ערכים מעט נמוכים יותר, כמו למשל 0.3g, נקבל שהתחום הוא בין -0.1g ל-0.7g, ובמקרה כזה נוכל לשנות את הגבול התחתון ל-0.0g מבלי לשנות באופן מהותי את התוצאות. ההצהרה שלנו עדיין תהיה נכונה. אם המדידה שלנו תרד מ-0.3g, הגבול העליון יהיה מתחת לערך האמיתי של המסה, ושוב ההצהרות שלנו יהיו שקריות. הסיכוי הנמוך, 2.3%, שזה יקרה מקובל עלינו.

ואולם, אם במדידה נקבל ערך של -0.5g, התחום שניתן יהיה בין -0.9g ל-0.1g. תוצאה זו לא נכונה בעליל. בחרנו רמת וודאות של 95.4%, לכן אנו מקבלים על עצמנו שב-4.6% מהמקרים ההצהרות שלנו יהיה שקריות. אך יהיה זה טיפשי להצהיר שהמסה היא שלילית, כאשר ברור שזו אינה אפשרות מציאותית.

על מנת לפתור בעיה זו פונים לסטטיסטיקה בייסיאנית. לפי משפט בייס, מתקיים:

$$p(\text{theory} | \text{result}) = \frac{p(\text{result} | \text{theory})}{p(\text{result})} p(\text{theory})$$

כאשר $p(\text{theory} | \text{result})$ היא ההסתברות שהתיאוריה נכונה בהינתן התוצאה שקיבלנו, בעוד ש- $p(\text{result} | \text{theory})$ היא ההסתברות לקבלת תוצאה מסוימת בהינתן שהתיאוריה נכונה. במקרה שלנו, בהנחה ששייאת המדידה מתפלגת נורמלית, נוכל לרשום:

$$p(\text{result} | \text{theory}) = p(x | X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-X)^2/2\sigma^2}$$

כאשר x הוא הערך שמדדנו, ו- X הוא הערך האמיתי של הגודל אותו אנו מודדים. ניתן להתעלם מ- $p(\text{result})$ במכנה, משום שבסופו של דבר ננרמל את הפונקציה. ומה לגבי $p(\text{theory})$? בהיעדר ידע מוקדם על הגודל X , לא ניתן לומר דבר פרט לכך שכל הערכים אפשריים באותה מידה. כלומר, $p(X)$ נתונה על ידי התפלגות אחידה על כל הציר. במקרה כזה נקבל:

$$p(\text{theory} | \text{result}) = p(X | x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-X)^2/2\sigma^2}$$

וניתן להמשיך ולקבל גרף ל- x_{\pm} כפונקציה של X וכו'. למעשה נקבל בדיוק אותה תוצאה שכבר הראנו לעיל.

אך מדידת מסה היא בדיוק דוגמה למקרה שבו יש לנו ידע מוקדם על X - אנו יודעים בוודאות שהוא חיובי. ולכן ההתפלגות $p(X)$ אינה אחידה לגמרי, אלא רק בחלק החיובי של הציר (סוג של פונקצית מדרגה מנורמלת). במקרה כזה נקבל, לאחר נרמול:

$$p(X | x) = \frac{e^{-(x-X)^2/2\sigma^2}}{\int_0^{\infty} e^{-(x-X')^2/2\sigma^2} dX'} \quad (x > 0)$$

דוגמה: במדידת מסה מקבלים תוצאה של $-0.5 \pm 0.2g$. מתוך טבלה מקבלים שהאינטגרל במכנה הוא 0.0062 (השטח מתחת לגאוסיאן שנמצא מעבר לנקודת ה- 2.5σ). באותה טבלה מחפשים את הנקודה שהחל ממנה מתקבל 10% מהמספר הנ"ל, כלומר 0.0006, ומקבלים שזה קורה ב- 3.24σ . לפיכך הגבול העליון, ברמת וודאות של 90%, הוא:
 $-0.5g + 3.24 \times 0.2g = 0.15g$

[132] רמת וודאות להתפלגות בינומית:

הערך שמודדים הוא r , מספר ההצלחות, ומטבעו הוא דיסקרטי. אם נרצה לקבל תחום ערכים על R (הגודל האמיתי) ברמת וודאות של 95%, למשל, נחפש ערכים של r_+ ו- r_- כך שיתקיים:

$$\sum_{r=0}^{r_+} P(r; R) = 0.975, \quad \sum_{r=r_-}^n P(r; R) = 0.975$$

אך סביר להניח שלא נוכל למצוא r_+, r_- שלם שמקיימים בדיוק זאת, ולכן נדרוש:

$$\sum_{r=0}^{r_+} P(r; R) \geq 0.975, \quad \sum_{r=r_-}^n P(r; R) \geq 0.975$$

ואז מובטחת לנו רמת וודאות של 95% או יותר, אך בוודאי לא פחות. באופן דומה ניתן למצוא את התחום על p , ההסתברות להצלחה בניסיון בודד. אם ביצענו n ניסויים, מתוכם m הסתיימו בהצלחה, נחפש p_+ ו- p_- שיקיימו:

$$\sum_{r=m+1}^n P(r; p_+, n) = 0.975, \quad \sum_{r=0}^{m-1} P(r; p_-, n) = 0.975$$

[133] רמת וודאות להתפלגות פואסון:

נניח כי בניסוי נצפו n אירועים בפרק זמן מסויים, ואנו רוצים להעריך את המספר הממוצע האמיתי של האירועים באותו פרק זמן, N . ברמת וודאות של 90%, הגבול העליון על מספר זה נתון על ידי N_+ שמקיים את אחת מבין המשוואות (השקולות):

$$\sum_{r=n+1}^{\infty} P(r; N_+) = 0.90 \quad \text{or} \quad \sum_{r=0}^n P(r; N_+) = 0.10$$

המשמעות - אם הערך האמיתי של N הוא N_+ , הסיכוי לקבל n כזה או נמוך יותר הוא 10% בלבד. אם N הוא גדול אף יותר, הסיכוי לכך פוחת עוד. לכן נאמר כי ברמת וודאות של 90%, N הוא לא יותר גדול מ- N_+ , וההצהרה כזו תהיה נכונה במוצע ב-9 מקרים מתוך 10.

באופן דומה ניתן לקבל את הגבול התחתון N_- על N ברמת וודאות דומה:

$$\sum_{r=0}^{n-1} P(r; N_-) = 0.90 \quad \text{or} \quad \sum_{r=n}^{\infty} P(r; N_-) = 0.10$$

תוצאות מוכנות מראש עבור $n = 0, \dots, 10$ נתונות בטבלה הבאה:

	Upper			Lower		
	90%	95%	99%	90%	95%	99%
$n = 0$	2.30	3.00	4.61	—	—	—
1	3.89	4.74	6.64	0.11	0.05	0.01
2	5.32	6.30	8.41	0.53	0.36	0.15
3	6.68	7.75	10.05	1.10	0.82	0.44
4	7.99	9.15	11.60	1.74	1.37	0.82
5	9.27	10.51	13.11	2.43	1.97	1.28
6	10.53	11.84	14.57	3.15	2.61	1.79
7	11.77	13.15	16.00	3.89	3.29	2.33
8	12.99	14.43	17.40	4.66	3.98	2.91
9	14.21	15.71	18.78	5.43	4.70	3.51
10	15.41	16.96	20.14	6.22	5.43	4.13

קבלת החלטות

חסרים:

עמודים 141 עד 149, כולל.

[150] מבחן χ^2 :

כזכור, χ^2 מוגדר לפי:

$$\chi^2 = \sum_{i=1}^N \left[\frac{y_i - f(x_i; a)}{\sigma_i} \right]^2 = \sum_{i=1}^N \left(\frac{y_i^{\text{actual}} - y_i^{\text{ideal}}}{\text{expected error}} \right)^2$$

וההתפלגות של χ^2 נתונה על ידי:

$$P(\chi^2; n) = \frac{2^{-n/2}}{\Gamma(n/2)} \chi^{n-2} e^{-\chi^2/2}$$

כאשר n הוא במספר דרגות החופש, ששווה למספר הנקודות באוסף הנתונים, N , פחות מספר הפרמטרים שאותם אנו מעריכים (כלומר, שבעזרתם הבאנו למינימום את χ^2 ; לדוגמה, כאשר עושים התאמה לקו ישר מעריכים שני פרמטרים - שיפוע ונק' חיתוך). מבחן זה מתאים גם כאשר אנו מסיקים את צורת הפונקציה מתוך הנתונים (ואז מספר דרגות החופש קטן ממספר הנתונים), אך גם כאשר אנו רוצים לבדוק התאמה לנתונים של פונקציה שנבחרה מראש (ואז $n = N$).

הצפי הוא לקבל $\chi^2 \approx n$. זאת משום שאם הפונקציה מתאימה לנתונים, הסטיות במונה יהיו בערך σ_i , ובסך הכל נקבל תרומה של 1 מכל נקודה באוסף הנתונים. באופן כמותי, את ההתאמה של הפונקציה $f(x; a)$ לנתונים מעריכים בעזרת האינטגרל:

$$\text{Prob}(\chi^2; n) = \int_{\chi^2}^{\infty} P(\chi'^2; n) d\chi'^2$$

גודל זה נקרא "הסתברות- χ^2 ", והוא מייצג את הסיכוי שהפונקציה שבאמת מתארת את הנתונים, תניב ערך של χ^2 שגדול או שווה לערך שקיבלנו מתוך הניסוי. הסתברות- χ^2 נמוכה (כלומר, ערך גבוה של χ^2) בדרך כלל מרמזת על כך שהפונקציה לא מתאימה לנתונים שלנו, או שהערכנו שהשיאות הסטטיסטיות (σ_i) קטנות מכפי שהן באמת. מצד שני, הסתברות- χ^2 ששואפת לאחד (χ^2 נמוך מאוד) היא גם חשודה, וייתכן כי הפרזנו בגודל השיאות.

הטבלה הבאה נותנת את הערכים הקריטיים של χ^2 עבור כמה ערכים של דרגות חופש:

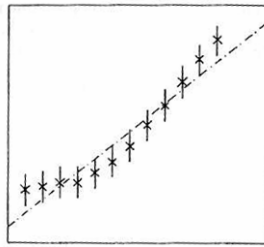
CRITICAL χ^2 VALUES

	$P = 10\%$	$= 5\%$	$= 2\%$	$= 1\%$
$n = 1$	2.71	3.84	5.41	6.63
2	4.61	5.99	7.82	9.21
3	6.25	7.82	9.84	11.34
4	7.78	9.49	11.67	13.28
5	9.24	11.07	13.39	15.09
6	10.64	12.59	15.03	16.81
7	12.02	14.07	16.62	18.47
8	13.36	15.51	18.17	20.09
9	14.68	16.92	19.68	21.67
10	15.99	18.31	21.16	23.21
11	17.27	19.68	22.62	24.72
12	18.55	21.03	24.05	26.22
13	19.81	22.36	25.47	27.69
14	21.06	23.68	26.87	29.14
15	22.31	25.00	28.26	30.58
16	23.54	26.30	29.63	32.00
17	24.77	27.59	31.00	33.41
18	25.99	28.87	32.35	34.81
19	27.20	30.14	33.69	36.19
20	28.41	31.41	35.02	37.57
21	29.62	32.67	36.34	38.93
22	30.81	33.92	37.66	40.29
23	32.01	35.17	38.97	41.64
24	33.20	36.42	40.27	42.98
25	34.38	37.65	41.57	44.31
26	35.56	38.89	42.86	45.64
27	36.74	40.11	44.14	46.96
28	37.92	41.34	45.42	48.28
29	39.09	42.56	46.69	49.59
30	40.26	43.77	47.96	50.89

למשל, עבור $n = 10$ דרגות חופש, ההסתברות שנקבל χ^2 גדול מ-15.99 היא 10% בלבד. לערכים גדולים יותר של n ניתן להיעזר בכך ש- $\sqrt{2\chi^2}$ מתפלג גאוסיאנית עם ממוצע $\sqrt{2n-1}$ וסטיית תקן של 1, ולהשתמש בטבלת הערכים של ההתפלגות הנורמלית. ניתן להשתמש במבחן זה גם עבור היסטוגרמות. y_i הם מספר האירועים בכל עמודה, x_i הוא הערך הממוצע של העמודה, ו- $f(x_i; a)$ הוא מספר האירועים הצפוי בעמודה. מכיוון שההתפלגות פואסונית, השיאות הן $\sigma_i = \sqrt{f(x_i; a)}$. מספר דרגות החופש הוא מספר הבינים פחות מספר הפרמטרים שאנו מעריכים (לא לשכוח לספור גם את פרמטר הנרמול של הפונקציה, שכן צריך להתקיים $\sum_i f(x_i; a) = N$ כאשר N הוא המספר הכולל של האירועים).

[153] מבחן הסדרות (Run Test):

נתבונן בהתאמה של הנקודות הבאות לקו ישר:



למרות שהנקודות נמצאות קרוב יחסית לקו הישר, ברור שזו לא התאמה טובה ויש להשתמש בפולינום ממעלה גבוהה יותר. כיצד ניתן לכמת את התחושה הזו? בציוור לעיל 12 נקודות, מתוכן חצי נמצאות מעל לקו הישר וחצי מתחת. בסך הכל קיימים שלושה "סדרות" - 3 נקודות מעל לקו, 6 נקודות מתחת, ושוב 3 נקודות מעל. זהו מספר נמוך של סדרות, בהתחשב במספר הנקודות הכולל. היינו מצפים שעבור התאמה טובה יותר, הנקודות מעל ומתחת לגרף לאו דווקא יגיעו ביחד, ונקבל יותר סדרות קצרות יותר. למעשה, ניתן להראות שההתפלגות עבור r , מספר הסדרות, היא כזו כך שמתקיים:

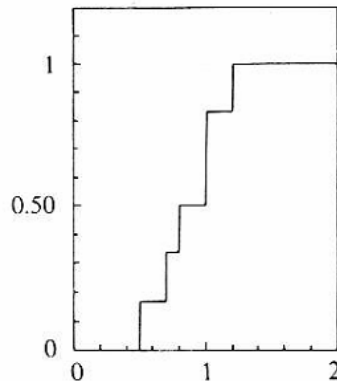
$$\langle r \rangle = 1 + \frac{2N_A N_B}{N}$$

$$V(r) = \frac{2N_A N_B (2N_A N_B - N)}{N^2 (N - 1)}$$

N הוא מספר הנקודות הכולל, ו- N_A ו- N_B הם מספר הנקודות מעל ומתחת לגרף, בהתאמה.

[155] מבחן קולמוגורוב (Kolmogorov):

שימושי כאשר אוסף הנתונים שעומד לרשותנו קטן יחסית, כך שקשה לבנות היסטוגרמה מתאימה ולהשתמש במבחן χ^2 . לצורך המבחן בונים גרף של ההתפלגות המצטברת (מנורמלת) $\text{cum}(x)$. למשל:



גרף זה מתאים לנקודות $\{0.5, 0.7, 0.8, 1.0, 1.0, 1.2\}$. בנוסף מציירים גרף של $\text{cum}(P)$, ההסתברות המצטברת של ההתפלגות P :

$$\text{cum}(P)(x) = \int_{-\infty}^x P(x') dx'$$

מחפשים את המרחק המקסימלי בין שני הגרפים ומחשבים את הגודל:

$$d = \sqrt{N} \cdot \max_x |\text{cum}(x) - \text{cum}(P)|$$

כאשר N הוא מספר הנקודות באוסף הנתונים. ערכים חשובים של d כוללים:

Critical value	Significance
1.63	1%
1.36	5%
1.22	10%
1.07	20%

[156] **השוואת שתי תוצאות:**

בהרבה מקרים נרצה להחליט האם, למשל, שתי מדידות שביצענו בניסויים שונים, מייצגות את אותו גודל פיסיקלי. בד"כ ניתן להניח שהתוצאות מתפלגות נורמלית, עם איזו סטיית תקן ידועה:

$$X \pm \sigma_X, \quad Y \pm \sigma_Y$$

הדבר שקול לשאלה האם $X - Y = 0$. השונות על הגודל הנ"ל היא:

$$V(X - Y) = \sigma_X^2 + \sigma_Y^2$$

נתעניין במרחק של $(X - Y)$ מהאפס ביחידות של סטיית התקן, כלומר:

$$\eta = \frac{|X - Y|}{\sqrt{\sigma_X^2 + \sigma_Y^2}}$$

ומכאן ניתן למצוא את רמת הוודאות שבה תיעשה ההחלטה לפי טבלאות הגאוסיאן.