STATISTICS AND PROBABILITY
First Semester 2024-2025

Midterm Exam
Wednesday, October 2, 2024
A closed-book, two-hour exam

Part A (20%)

Do either one of the following two problems.

Problem A.1 (PI a1, a2, a3; 20%)

Concrete supply for a construction project comes from three batching plants, namely plant A, B, and C. The probabilities that the concrete comes from those plants are 30%, 45%, and 25%, respectively. The quality engineer of the construction project sees that concrete supply from each plant has a risk of being late. The probability of late supply from A, B, and C plants are 12%, 10%, and 7%, respectively.

a. Find the probability of late concrete supply to the project.
b. Suppose that a late supply has occurred. Find the probability that the supply came from batching plant B.

Answer

This is a probability problem. Let event S_i expresses a concrete supply from batching plant i ($i =$ A, B, C) and event L shows a late supply. We have

$\text{prob}(S_A) = 0.30$ and $\text{prob}(L|S_A) = 0.12$

$\text{prob}(S_B) = 0.45$ and $\text{prob}(L|S_B) = 0.10$

$\text{prob}(S_C) = 0.25$ and $\text{prob}(L|S_C) = 0.07$

a. The probability of late concrete supply to the project is the total probability of late supply from the three batching plants.

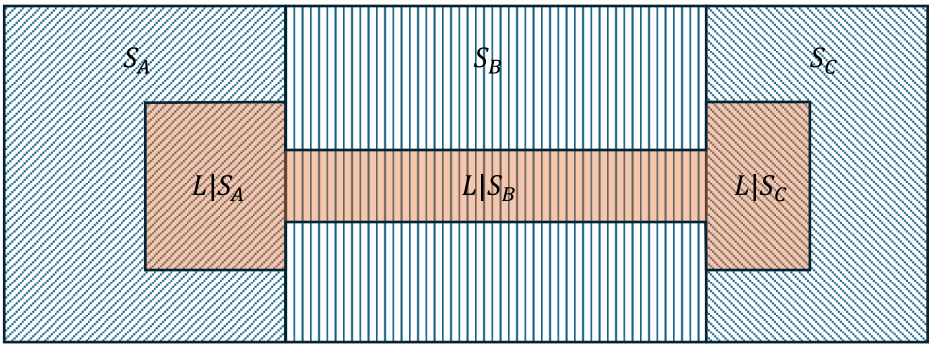
$$\text{prob}(L) = \sum_{i=A}^{C} \text{prob}(S_i)\ \text{prob}(L|S_i) = 0.30 \times 0.12 + 0.45 \times 0.10 + 0.25 \times 0.07 = 0.0985$$

b. The probability of concrete supply from batching plant B when a late has occurred can be explained by the Bayes theorem.

$$\text{prob}(S_B|L) = \frac{\text{prob}(L|S_B)}{\sum_{i=A}^{C} \text{prob}(S_i)\ \text{prob}(L|S_i)} = \frac{0.45 \times 0.10}{0.0985} = 0.4569$$

This type of problem can be expressed by a diagram (Venn diagram) as follows.

$S_A = 0.30$ $S_B = 0.45$ $S_C = 0.25$



Problem A.2 (PI a1, a2, a3; 20%)

The following data are the yield strength, f_y, in mega pascal (MPa) of one of the reinforcing steel columns of building columns.

472	431	455	431	425
415	430	427	437	421
429	432	431	425	419

a. Find the mean and the standard deviation of the sample.
b. Find the lower quartile and upper quartile of the sample.
c. The design yield strength of the reinforcing steel is defined at 420 MPa. Could we conclude that the steel column meets this design criterion? Give your argument to support your conclusion.

Answer

a. The mean, $\bar{f_y}$, and the standard deviation, s_{fy}, of the sample are as follows.

$$\bar{f_y} = \frac{1}{15}\sum_{i=1}^{15}(f_y)_i = \frac{6480}{15} = 432 \text{ MPa}$$

$$s_{fy} = \sqrt{\frac{\sum_{i=1}^{15}\left[(f_y)_i - \bar{f_y}\right]^2}{14}} = 14.32 \text{ MPa}$$

b. The sample shall be sorted from the lowest to the largest to get the lower and upper quartiles of the sample. The lower, or the first, quartile and the upper, or the third, quartile are as follows.

$$Q_1 = Q_{25\%} = 425 \text{ MPa}$$

$$Q_3 = Q_{75\%} = 431.5 \text{ MPa}$$

c. The design yield strength defines **the minimum value** of the yield strength. The steel yield strength shall not be lower than its design value. The sample shows that there are values below 420 MPa, i.e. 415 MPa and 419 MPa. Therefore, the sample does not meet the required yield strength. The quality engineer shall declare that the column does not comply with the design criteria.

Part B (80%)

Do two out of four of the following problems.

Problem B.1 (PI a1, a2, a3; 40%)

The probability density function of the annual series of maximum daily rainfall at a weather station, H in millimeters, is expressed as follows.

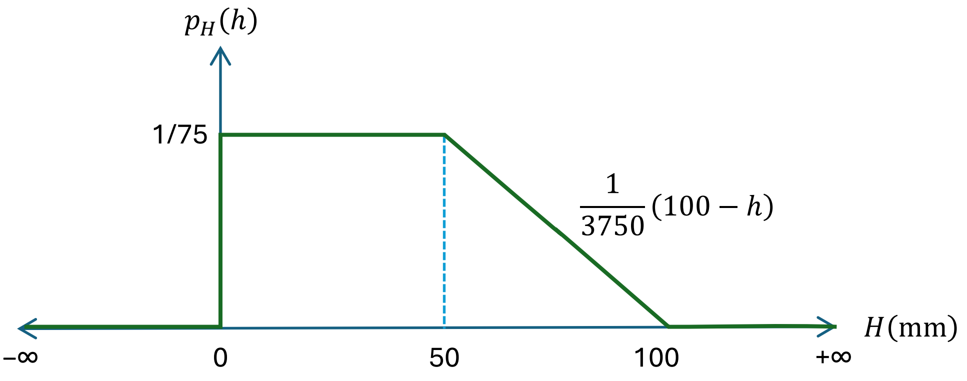
$$p_H(h) = \begin{cases} 0 & \text{if } h < 0 \\ \dfrac{1}{75} & \text{if } 0 < h < 50 \\ \dfrac{1}{3750}(100 - h) & \text{if } 50 < h < 100 \\ 0 & \text{if } h > 100 \end{cases}$$

a. Draw the probability density function curve.
b. Find and draw the cumulative distribution function curve.
c. Find prob(40 mm < H < 60 mm).
d. Find the median of the maximum daily rainfall.

Answer

a. The pdf curve of the annual maximum daily rainfall.

$$p_H(h) = \begin{cases} 0 & \text{if } h < 0 \\ 1/75 & \text{if } 0 < h < 50 \\ \dfrac{1}{3750}(100 - h) & \text{if } 50 < h < 100 \\ 0 & \text{if } h > 100 \end{cases}$$



b. The cumulative distribution function of the annual maximum daily rainfall.

- CDF for $-\infty < h < 0$

$$P_H(h) = \int_{-\infty}^{h} p_H(t)dt = \int_{-\infty}^{h} 0\ dt = C$$

Since $P_H(h) = 0$ for $h < 0$, thus $C = 0$

$$P_H(h) = 0$$

- CDF for $0 < h < 50$

$$P_H(h) = \int_0^h p_H(t)dt = \int_0^h \frac{1}{75}\,dt = \frac{h}{75} + C$$

Since $P_H(h) = 0$ for $h = 0$, thus $C = 0$

$$P_H(h) = \frac{h}{75}$$

- CDF for $50 < h < 100$

$$P_H(h) = \int_0^h p_H(t)dt = \int_0^h \frac{1}{3750}(100 - t)\,dt = \frac{1}{3750}\left(100h - \frac{h^2}{2} + C\right)$$

For $h = 50$, we have $P_H(50) = 50/75 = 2/3$; thus

$$P_H(h = 50) = \frac{1}{3750}\left(100 \times 50 - \frac{50 \times 50}{2} + C\right) = \frac{2}{3}$$

$$\Rightarrow C = \frac{2 \times 3750}{3} - 5000 + 1250 = -1250$$

Therefore,

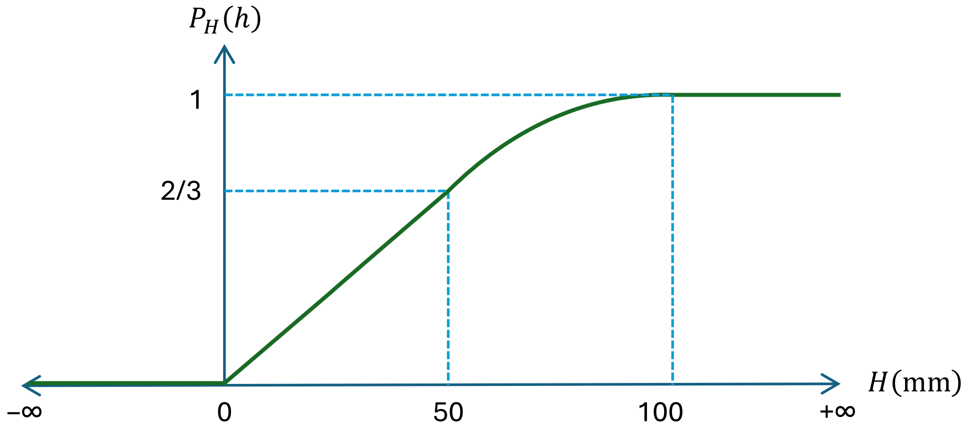
$$P_H(h) = \frac{1}{3750}\left(100h - \frac{h^2}{2} - 1250\right)$$

- CDF for $h > 100$

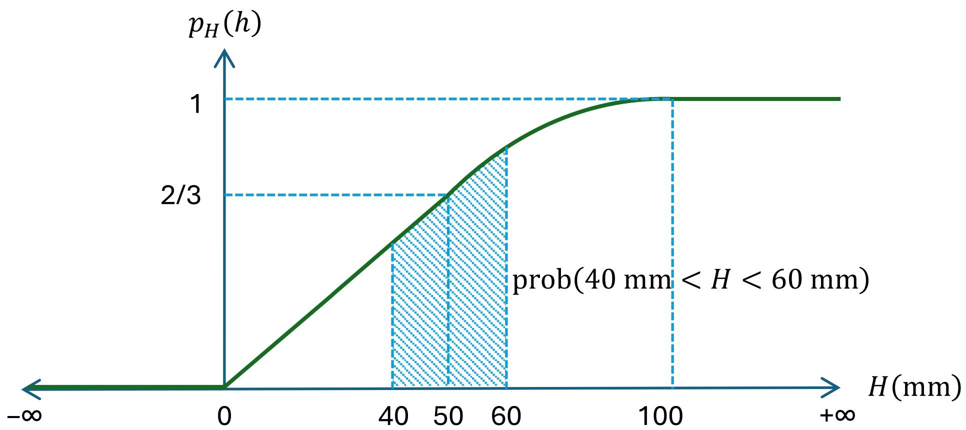
$$P_H(h) = 1$$

Below is the complete CDF.

$$P_H(h) = \begin{cases} 0 & \text{if } h < 0 \\ \dfrac{h}{75} & \text{if } 0 < h < 50 \\ \dfrac{1}{3750}\left(100h - \dfrac{h^2}{2} - 1250\right) & \text{if } 50 < h < 100 \\ 1 & \text{if } h > 100 \end{cases}$$



c. prob(40 mm $< H <$ 60 mm).



prob(40 mm < H < 60 mm) = area under the pdf curve from h = 40 mm to h = 60 mm.

$$\text{prob}(40\text{ mm} < H < 60\text{ mm}) = P_H(60) - P_H(40) = \frac{1}{3750}\left(100 \times 60 - \frac{60^2}{2} - 1250\right) - \frac{40}{75}$$
$$= \frac{2950}{3750} - \frac{40}{75} = \frac{2950 - 2000}{3750} = \frac{950}{3750} = \frac{19}{75}$$

d. The median of the annual maximum daily rainfall.

The sample median, H_m, is the score at the middle of a sorted data series, which is the score such that half of the scores lie on either side of the median. In a continuous variable, the median can be found from

$$\int_{-\infty}^{H_m} p_H(h)dh = 0.5 \implies \text{prob}(H < H_m) = P_H(H_m) = 0.5$$

Looking at the CDF curve, we expect that the median lies in the range of 0 to 50 mm.

$$P_H(H_m) = \frac{H_m}{75} = 0.5 \implies H_m = 0.5 \times 75 = 37.5 \text{ mm}$$

Problem B.2 (PI a1, a2, a3; 40%)

The table below shows the frequency of daily air temperature in centigrade and relative air humidity in percent at a weather station.

		Air temperature (°C)					
		22—24	24—26	26—28	28—30	30—32	32—34
Air humidity (%)	0—20	2	4	6	2	2	1
	20—40	4	8	12	30	6	9
	40—60	5	15	30	60	30	20
	60—80	3	7	9	25	17	11
	80—100	1	0	2	12	8	3

a. Find the joint probability density function of the two random variables.
b. Find the marginal pdf and cdf of the air temperature.
c. Find the marginal pdf and cdf of the air humidity.
d. Find the probability of the air temperature being in the range of 28°C to 30°C.
e. Find the probability of the air temperature being in the range of 28°C to 30°C when the air relative humidity is in the range of 60% to 80%.

Answer

This is a bivariate random variable problem. Let T and H are the air temperature and the air relative humidity, respectively. The probabilities of this bivariate random variable are easily calculated in tables.

Frequency table.

		Air temperature (°C)						Σ
		22—24	24—26	26—28	28—30	30—32	32—34	
Air humidity (%)	0—20	2	4	6	2	2	1	17
	20—40	4	8	12	30	6	9	69
	40—60	5	15	30	60	30	20	160
	60—80	3	7	9	25	17	11	72
	80—100	1	0	2	12	8	3	26
	Σ	15	34	59	129	63	44	344

Joint probability, marginal pdf and cdf of air temperature, marginal pdf and cdf of air humidity.

		Air temperature (°C)						$p_H(h)$	$P_H(h)$
		22—24	24—26	26—28	28—30	30—32	32—34		
Air humidity (%)	0—20	0.01	0.01	0.02	0.01	0.01	0.00	0.05	0.05
	20—40	0.01	0.02	0.03	0.09	0.02	0.03	0.20	0.25
	40—60	0.01	0.04	0.09	0.17	0.09	0.06	0.47	0.72
	60—80	0.01	0.02	0.03	0.07	0.05	0.03	0.21	0.92
	80—100	0.00	0.00	0.01	0.03	0.02	0.01	0.08	1.00
	$p_T(t) \rightarrow$	0.04	0.10	0.17	0.38	0.18	0.13	1	---
	$P_T(t) \rightarrow$	0.04	0.14	0.31	0.69	0.87	1.00	---	---

a. The joint probability density function of the air temperature, *T*, and the air relative humidity, *H*, are expressed in red color numbers in the table.
b. The marginal pdf and cdf of the air temperature are presented with cyan color numbers in the table.
c. The marginal pdf and cdf of the air humidity are presented with greed color numbers in the table.
d. The probability of the air temperature being in the range of 28°C to 30°C.

$$\text{prob}(28°\text{C} < T < 30°\text{C}) = 0.38$$

e. The probability of the air temperature being in the range of 28°C to 30°C when the air relative humidity is in the range of 60% to 80%.

$$\text{prob}(28°\text{C} < T < 30°\text{C} | 60\% < H < 80\%) = \frac{0.07}{0.21} = 0.33$$

Problem B.3 (PI a1, a2, a3; 40%)

A flood early warning in a river relies on measured water level at a cross section. It is known that the instrument has an error probability of 1/1000.

a. Find the probability of no more than 3 errors in 1000 measurements. Use Poisson distribution.
b. Redo the above problem by applying binomial distribution.
c. Consider problems a and b. Which approach is more suitable to predict the probability of the event?

Answer

a. The probability of no more than 3 errors in 1000 experiments.

The cumulative distribution function of Poisson processes reads

$$F_X(x;\lambda) = \sum_{i=0}^{x} \frac{\lambda^i e^{-\lambda}}{i!}$$

where $\lambda = np$ and $x = 0,1,2,\ldots$

In the above problem, we have $p = 0.001$, $x = 3$, and $n = 1000$. We have from there $\lambda = np = 1$. Therefore, the probability of no more than 3 errors in 1000 experiments is

$$F_X(3;1) = \sum_{i=0}^{3} \frac{1^i e^{-1}}{i!} = \frac{1^0 e^{-1}}{0!} + \frac{1^1 e^{-1}}{1!} + \frac{1^2 e^{-1}}{2!} + \frac{1^3 e^{-1}}{3!} = \frac{e^{-1}}{1} + \frac{e^{-1}}{1} + \frac{e^{-1}}{2} + \frac{e^{-1}}{6} = 0.981$$

b. Applying binomial distribution to the above problem, we have

$$F_X(x;n,p) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}$$

$$F_X(3;1000{,}0.001)$$
$$= \frac{0!}{(1000-0)!\,0!}0.001^0 0.999^{1000} + \frac{1!}{(1000-1)!\,1!}0.001^1 0.999^{999} + \frac{2!}{(1000-2)!\,2!}0.001^2 0.999^{998} + \frac{3!}{(1000-3)!\,3!}0.001^3 0.999^{997} = 0.981$$

c. The two approaches to estimate the probability of the event produce the same result. This is true for large number of experiments or observations. The binomial distribution, nevertheless, may not be possible to calculate by using hand-held calculator when we must calculate factorial of large numbers.

Problem B.4 (PI a1, a2, a3; 40%)

Laboratory test data show that the compressive strength, f'_c, of concrete samples is normally distributed whose mean and standard deviation are 27 MPa and 2.5 MPa, respectively.

a. Find the probability that a concrete sample has a compressive strength between 25 MPa and 28 MPa, $\text{prob}\left(25\text{ MPa} < f'_c < 28\text{ MPa}\right)$.
b. SNI 2847:2019 dictates that the minimum compressive strength of concrete as building material is 21 MPa. Does the sample meet the standard?

Answer

a. Probability of a normally distributed random variable.

$$z_{25} = \frac{25-27}{2.5} = -0.8$$

$$z_{28} = \frac{28-27}{2.5} = 0.4$$

$$\text{prob}\left(25\text{ MPa} < f'_c < 28\text{ MPa}\right) = \text{prob}\left(f'_c < 28\text{ MPa}\right) - \text{prob}\left(f'_c < 25\text{ MPa}\right)$$
$$= \text{prob}(Z < 0.4) - \text{prob}(Z < -0.8) = 0.6554\text{-}0.2119 = 0.4435$$

b. The minimum compressive strength 21 MPa has a probability of occurrence as follows.

$$z_{21} = \frac{21-27}{2.5} = -2.4$$

$$\text{prob}\left(f'_c < 21\text{ MPa}\right) = \text{prob}(Z < -2.4) = 0.0082$$

$$\Rightarrow \text{prob}\left(f'_c > 21\text{ MPa}\right) = 1 - \text{prob}(Z < -2.4) = 1 - 0.0082 = 0.9918$$

There is 99% of possibility that the concrete sample have compressive strength above the minimum value of 21 MPa. One might, therefore, conclude that the concrete meets the SNI. We shall, however, infer that the sample does not meet the SNI by interpreting that the code (SNI) states **the minimum value**, which must be met without any exception whatsoever. The confusion arises from the statement that the concrete sample is normally distributed. The concrete compressive strength should follow distribution that is bounded or asymptotic on the minimum side.

Notes

Sample mean

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

Standard deviation of sample

$$s_X = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}}$$

Binomial distribution

$$f_X(x; n, p) = \binom{n}{x}\ p^x (1-p)^{n-x}, \quad x = 0,\ 1,\ 2,\ \ldots,\ n$$

$$F_X(x; n, p) = \sum_{i=0}^{x} f_X(i; n, p) =$$

$$= \sum_{i=0}^{x}\binom{n}{i}\ p^i (1-p)^{n-i}\,,\ x = 0,1,2,\ldots,n$$

Poisson distribution

$$f_X(x; \lambda) = \frac{\lambda^x\ e^{-\lambda}}{x!}, x = 0,1,2,\ldots.\ \text{ and }\ \lambda = np > 0$$

$$F_X(x; \lambda) = \sum_{i=0}^{x}\frac{\lambda^i\ e^{-\lambda}}{i!}$$

Total probability

$$\text{prob}(A) = \sum_{i=1}^{n}\text{prob}(B_i)\ \text{prob}(A|B_i)$$

Bayes theorem

$$\text{prob}(B_j|A) = \frac{\text{prob}(B_j)\ \text{prob}(A|B_j)}{\sum_{i=1}^{n}\text{prob}(B_i)\ \text{prob}(A|B_i)}$$

CDF *vs* PDF

$$P_H(h) = \int_{t_1}^{t_2} p_H(t)\ dt$$

Standard normal distribution

$$Z_x = \frac{X-\mu}{\sigma}$$

$$\text{prob}(Z > z) = 1 - \text{prob}(Z < z)$$

-ist-