Mid Semester Exam Solutions
Statistical Foundations for ML

Solutions (September, 2025)

Information

- Total marks: 30
- Two sections: Section A (MCQ, 10 marks) and Section B (Numerical, 20 marks)

Section A: Multiple Choice (10 marks)

Correct options are indicated; brief justifications are provided for clarity.

Q1. Ans: (b) Mean=median=mode for Normal.

Q2. Ans: (b) For 3 flips, $\binom{3}{2}(1/2)^3 = 3/8$.

Q3. Ans: (c) Poisson models counts in a fixed interval with rate λ.

Q4. Ans: (b) $\mathbb{P}(X < Y) = \int_0^\infty \left(\int_0^y 2e^{-x}e^{-2y}\, dx \right) dy = \int_0^\infty 2(1-e^{-y})e^{-2y} dy = \frac{1}{3}$.

Q5. Ans: (a) Linearity: $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

Q6. Ans: (a) $Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.

Q7. Ans: (b) Median is robust to a single extreme value; mean, SD, range change.

Q8. Ans: (b) Mixture of two grades $\Rightarrow$ multi-modal likely.

Q9. Ans: (d) Market share of categories $\Rightarrow$ pie chart.

Q10. Ans: (c) Correlation -0.5 indicates a negative linear relationship; not causation.

Section B: Subjective (20 marks)

Q1. (2 marks) Given $\mathbb{P}(A_2 \cap B_1) = 0.10$, $\mathbb{P}(B_1) = 0.27$, $\mathbb{P}(A_2) = 0.27$.

$$\mathbb{P}(A_2 \mid B_1) = \frac{\mathbb{P}(A_2 \cap B_1)}{\mathbb{P}(B_1)} = \frac{0.10}{0.27} \approx 0.37037.$$

Independence would require $\mathbb{P}(A_2 \mid B_1) = \mathbb{P}(A_2) = 0.27$ (or $\mathbb{P}(A_2 \cap B_1) = 0.27 \times 0.27 = 0.0729$). Since $0.10 \neq 0.0729$, A_2 and B_1 are *not* independent.

Q2. (2 marks) For constants a, b, prove the formulas for expectation and variance.

Expectation:

$$\mathbb{E}[aX + b] = \sum_x (ax + b)\mathbb{P}(X = x) \quad \text{(discrete)}$$

Split the terms:

$$\mathbb{E}[aX + b] = a\sum_x x\mathbb{P}(X = x) + b\sum_x \mathbb{P}(X = x) = a\,\mathbb{E}[X] + b \cdot 1 = a\,\mathbb{E}[X] + b.$$

This file is meant for personal use by mepravintpatil@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

$$E[aX+b] = a\mathbb{E}[X] + b,$$

Variance:

$$\mathrm{Var}(X) = \mathbb{E}\big[(X-\mathbb{E}[X])^2\big].$$

$$\mathrm{Var}(aX+b) = \mathbb{E}\big[(aX+b-\mathbb{E}[aX+b])^2\big].$$

From above, $\mathbb{E}[aX+b] = a\mathbb{E}[X]+b$, so

$$aX+b-(a\mathbb{E}[X]+b) = a(X-\mathbb{E}[X]).$$

Thus

$$\mathrm{Var}(aX+b) = \mathbb{E}[(a(X-\mathbb{E}[X]))^2] = a^2\,\mathbb{E}[(X-\mathbb{E}[X])^2] = a^2\,\mathrm{Var}(X).$$

If $\mathrm{Var}(X)=\sigma^2$, then

$$\mathrm{Var}(aX+b) = a^2\sigma^2.$$

$$\boxed{\mathbb{E}[aX+b] = a\,\mathbb{E}[X]+b, \qquad \mathrm{Var}(aX+b) = a^2\sigma^2.}$$

Q3. (2 marks) Overbooking with $X \sim \mathrm{Bin}(n=20, p=0.8)$ and 16 seats.

$$\text{(a) } \mathbb{P}(\text{overbook}) = \mathbb{P}(X \geq 17) = \sum_{k=17}^{20} \binom{20}{k} 0.8^k 0.2^{20-k} \approx \boxed{0.41145}.$$

$$\text{(b) } \mathbb{P}(X \leq 15) = \sum_{k=0}^{15} \binom{20}{k} 0.8^k 0.2^{20-k} \approx \boxed{0.37035}.$$

Q4. (2 marks) If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{P}(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

where Φ is the standard normal cdf. Symmetry identity: $\Phi(-z) = 1-\Phi(z)$.

Q5. (3 marks) Data: $(x,y) \in \{(6,80),(7,60),(8,70),(9,40),(10,0)\}$.

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}} = \boxed{-0.9}.$$

Interpretation: strong negative linear relationship. *Duplication:* repeating each paired observation multiplies both numerator and denominator sums by the same positive factor; r is unchanged. Hence the correlation *does not change.*

Q6. (3 marks) Machines M_1, M_2, M_3 produce $30\%, 45\%, 25\%$ with defect rates $0.02, 0.03, 0.05$.

$$\text{(a) } \mathbb{P}(D) = 0.3(0.02) + 0.45(0.03) + 0.25(0.05) = \boxed{0.032}.$$

$$\text{(b) } \mathbb{P}(M_2 \mid D) = \frac{0.45 \cdot 0.03}{0.032} = \boxed{0.421875}.$$

$$\text{(c) } \mathbb{P}(M_1 \mid D) = 0.1875,\ \mathbb{P}(M_2 \mid D) = 0.421875,\ \mathbb{P}(M_3 \mid D) = 0.390625.$$

Most likely: $\boxed{M_2}$.

Q7. (3 marks) Box with 7 red, 4 blue; two draws without replacement.

$$\text{(a) } \mathbb{P}(\text{both red}) = \frac{7}{11}\cdot\frac{6}{10} = \frac{21}{55} \approx \boxed{0.3818}.$$

$$\text{(b) } \mathbb{P}(\text{second red} \mid \text{first red}) = \frac{6}{10} = \boxed{0.6}.$$

$$\text{(c) } \mathbb{P}(\text{second red} \mid \text{first blue}) = \frac{7}{10} = \boxed{0.7}.$$

This file is meant for personal use by mepravintpatil@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Q8. (3 marks) Let $Y \sim \mathcal{N}(\mu, \sigma)$ and define base $B = |Y|$, height $H = 3|Y|$.

Note on wording: The prompt mentions constructing a *triangle* with base B and height H, but then calls A the area of a *rectangle*. We therefore show both interpretations.

Case 1: Triangle (likely intent).

$$A = \tfrac{1}{2} \cdot B \cdot H = \tfrac{1}{2}\,|Y| \cdot (3|Y|) = \tfrac{3}{2}\,|Y|^2.$$

Take expectation:

$$\mathbb{E}[A] = \tfrac{3}{2}\,\mathbb{E}[|Y|^2].$$

Since $Y^2 = |Y|^2$, we can drop the absolute value:

$$\mathbb{E}[|Y|^2] = \mathbb{E}[Y^2].$$

Since,

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}Y)^2$$

Now,

$$\mathbb{E}[Y^2] = \mathrm{Var}(Y) + (\mathbb{E}Y)^2 = \sigma^2 + \mu^2.$$

Therefore,

$$\mathbb{E}[A] = \tfrac{3}{2}(\mu^2 + \sigma^2).$$

Case 2: Rectangle (literal reading).

$$A = B \cdot H = |Y| \cdot (3|Y|) = 3|Y|^2 = 3Y^2.$$

So,

$$\mathbb{E}[A] = 3\,\mathbb{E}[Y^2] = 3(\mu^2 + \sigma^2).$$

Final Answer:

$$\boxed{\mathbb{E}[A] = \tfrac{3}{2}(\mu^2 + \sigma^2) \quad \text{(triangle)}} \qquad \text{or} \qquad \boxed{\mathbb{E}[A] = 3(\mu^2 + \sigma^2) \quad \text{(rectangle)}}$$

depending on the intended interpretation of the problem statement.

This file is meant for personal use by mepravintpatil@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.