Data Science for Mathematicians
Exercises 3: Probabilistic Foundations of Modeling

Instructions

Answer all exercises completely. Show all working, justify your answers, and state any assumptions you make. For computational exercises, carry out all intermediate steps explicitly. For proof exercises, clearly identify which definitions and theorems you are applying.

Exercises

Exercise 1. (Measure Theory) Let $\Omega = \{a, b, c, d\}$. Determine whether the following collections of subsets are σ-algebras on Ω. If not, explain which axiom fails.

(a) $\mathcal{F}_1 = \{\emptyset, \{a\}, \{b, c, d\}, \Omega\}$

(b) $\mathcal{F}_2 = \{\emptyset, \{a, b\}, \{c, d\}, \{a, c\}, \Omega\}$

(c) $\mathcal{F}_3 = \{\emptyset, \{a, b\}, \{c, d\}, \{a, b, c, d\}\}$

Exercise 2. (Lebesgue Integral) Consider the function $f : [0, 2] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 3 & \text{if } x \in [0, 1) \cap \mathbb{Q}, \\ 1 & \text{if } x \in [0, 1) \setminus \mathbb{Q}, \\ 2 & \text{if } x \in [1, 2]. \end{cases}$$

Compute the Lebesgue integral $\int_{[0,2]} f \, d\lambda$, where λ is the Lebesgue measure. Is this function Riemann integrable? Justify your answer.

Exercise 3. (Probability Space Construction) A biased die is rolled, where the probability of rolling face k is proportional to k (i.e., $\mathbb{P}(\{k\}) \propto k$ for $k \in \{1, 2, 3, 4, 5, 6\}$).

(a) Determine the exact probability of each outcome.

(b) Define a random variable X that equals 1 if the outcome is prime and 0 otherwise. Find the PMF of X.

(c) Compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Exercise 4. (Gaussian Properties) Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Prove that the standardized variable $Z = \dfrac{X - \mu}{\sigma}$ follows a standard normal distribution $\mathcal{N}(0, 1)$. Then use this result to prove that for any $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = \mathbb{P}(-2 \leq Z \leq 2) \approx 0.9545.$$

Exercise 5. **(Expectation of Functions)** Let X be a continuous random variable with PDF $f_X(x) = 2x$ for $x \in [0,1]$ and $f_X(x) = 0$ otherwise.

(a) Verify that f_X is a valid PDF.
(b) Compute $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, and $\mathrm{Var}(X)$.
(c) Find $\mathbb{E}[e^X]$.
(d) If $Y = 3X^2 - 2X + 1$, find $\mathbb{E}[Y]$.

Exercise 6. **(Binomial Distribution)** In a clinical trial, a new drug has a 70% success rate. If 15 patients are treated independently:

(a) What is the probability that exactly 12 patients respond successfully?
(b) What is the probability that at least 10 patients respond successfully?
(c) Find the expected number of successes and the standard deviation.
(d) Using the normal approximation with continuity correction, estimate $\mathbb{P}(X \geq 10)$ and compare with the exact value.

Exercise 7. **(Variance of Sums)** Let $X_1, X_2, \ldots, X_n$ be independent random variables with $\mathbb{E}[X_i] = \mu_i$ and $\mathrm{Var}(X_i) = \sigma_i^2$. Prove that

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \sigma_i^2.$$

Then, show that if the X_i are identically distributed with common variance σ^2, the variance of the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is $\dfrac{\sigma^2}{n}$.

Exercise 8. **(Covariance Properties)** Let X and Y be random variables. Prove the following properties:

(a) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
(b) $\mathrm{Cov}(aX + b, cY + d) = ac \cdot \mathrm{Cov}(X, Y)$ for constants $a, b, c, d \in \mathbb{R}$.
(c) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.
(d) If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.

Exercise 9. **(MLE for Gaussian)** Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations from $\mathcal{N}(\mu, \sigma^2)$, where both μ and σ^2 are unknown.

(a) Write down the likelihood function $L(\mu, \sigma^2 | \mathbf{x})$.
(b) Derive the log-likelihood function.
(c) Show that the MLE for μ is $\hat{\mu}_{MLE} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
(d) Show that the MLE for σ^2 is $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.
(e) Why is the MLE for variance biased? What is the unbiased estimator?

Exercise 10. **(MLE for Exponential)** Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. from an exponential distribution with rate parameter $\lambda > 0$, so the PDF is $f(x|\lambda) = \lambda e^{-\lambda x}$ for $x \geq 0$.

(a) Write down the likelihood and log-likelihood functions.

(b) Derive the MLE $\hat{\lambda}_{MLE}$.

(c) If the observed data are $\{1.2, 0.8, 2.1, 1.5, 0.9\}$, compute $\hat{\lambda}_{MLE}$.

(d) Find $\mathbb{E}[\hat{\lambda}_{MLE}]$ and determine whether it is an unbiased estimator of λ.

Exercise 11. (Central Limit Theorem Application) A factory produces bolts with weights that have mean $\mu = 50$ grams and standard deviation $\sigma = 4$ grams (distribution unknown). A random sample of $n = 64$ bolts is selected.

(a) Using the CLT, what is the approximate distribution of the sample mean $\bar{X}$?

(b) Find the probability that the sample mean is between 49 and 51 grams.

(c) Find the value c such that $\mathbb{P}(\bar{X} > c) = 0.05$.

(d) How large should n be so that $\mathbb{P}(|\bar{X} - 50| < 0.5) \geq 0.99$?

Exercise 12. (Bayes' Theorem) A factory has three machines (A, B, C) producing items. Machine A produces 50% of items with 2% defect rate, Machine B produces 30% with 3% defect rate, and Machine C produces 20% with 5% defect rate.

(a) What is the probability that a randomly selected item is defective?

(b) If an item is found to be defective, what is the probability it came from each machine?

(c) Two items are selected and both are defective. What is the probability that both came from Machine C? (Assume independence.)

Exercise 13. (Bayesian Inference) A coin has unknown probability θ of landing heads. We use a Beta(2, 2) prior for θ, which has PDF $f(\theta) = 6\theta(1 - \theta)$ for $\theta \in [0, 1]$.

(a) Compute the prior mean and variance of θ.

(b) After observing 8 heads in 10 flips, derive the posterior distribution of θ.

(c) Compute the posterior mean and compare it to the MLE.

(d) Find the MAP estimate and show how it differs from both the posterior mean and MLE.

(e) Compute a 95% credible interval for θ (you may express this in terms of the Beta distribution quantiles).

Exercise 14. (Naive Bayes Classification) Consider a classification problem with two classes (C_1 and C_2) and two continuous features (x_1, x_2). The class priors are $\mathbb{P}(C_1) = 0.4$ and $\mathbb{P}(C_2) = 0.6$. Within each class, the features are assumed independent and normally distributed:

	μ_1	σ_1	μ_2	σ_2
Class C_1	2	1	5	2
Class C_2	4	1.5	3	1

(a) For a new observation $\mathbf{x} = (3, 4)$, compute the (unnormalized) posterior probability for each class.

(b) Classify the observation and compute the probability of correct classification.

(c) Find the decision boundary in the (x_1, x_2) plane (the set of points where both classes are equally likely).

Exercise 15. **(Comprehensive Problem)** Let $X_1, X_2, \ldots, X_n$ be i.i.d. Bernoulli(p) random variables representing whether each of n website visitors clicks an advertisement.

(a) Write the joint PMF of $(X_1, \ldots, X_n)$ and show it depends on the data only through $T = \sum_{i=1}^{n} X_i$.

(b) Show that $T \sim \text{Binomial}(n, p)$ and derive $\mathbb{E}[T]$ and $\text{Var}(T)$.

(c) Using MLE, show that $\hat{p} = T/n$ and prove that $\hat{p}$ is an unbiased estimator of p.

(d) Using Chebyshev's inequality, show that for any $\epsilon > 0$:

$$\mathbb{P}(|\hat{p} - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}.$$

(e) Apply the CLT to show that for large n, $\hat{p} \overset{\text{approx}}{\sim} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$.

(f) If we observe 45 clicks out of 500 visitors, construct an approximate 95% confidence interval for p.

Solutions

Solution 1. (Measure Theory)

(a) $\mathcal{F}_1 = \{\emptyset, \{a\}, \{b,c,d\}, \Omega\}$ **is a σ-algebra**. We verify:

- Contains Ω: Yes, $\Omega = \{a,b,c,d\} \in \mathcal{F}_1$.
- Closed under complementation: $\{a\}^c = \{b,c,d\} \in \mathcal{F}_1$ and $\{b,c,d\}^c = \{a\} \in \mathcal{F}_1$. Also $\emptyset^c = \Omega$ and $\Omega^c = \emptyset$.
- Closed under countable unions: All possible unions of elements in $\mathcal{F}_1$ yield sets already in $\mathcal{F}_1$.

(b) $\mathcal{F}_2 = \{\emptyset, \{a,b\}, \{c,d\}, \{a,c\}, \Omega\}$ **is NOT a σ-algebra**.
It fails closure under complementation: $\{a,c\}^c = \{b,d\} \notin \mathcal{F}_2$.
Alternatively, it fails closure under union: $\{a,b\} \cup \{a,c\} = \{a,b,c\} \notin \mathcal{F}_2$.

(c) $\mathcal{F}_3 = \{\emptyset, \{a,b\}, \{c,d\}, \{a,b,c,d\}\}$ **is a σ-algebra**. This is the σ-algebra generated by the partition $\{\{a,b\},\{c,d\}\}$. We verify:

- $\Omega = \{a,b,c,d\} \in \mathcal{F}_3$.
- $\{a,b\}^c = \{c,d\} \in \mathcal{F}_3$ and vice versa.
- All unions remain in $\mathcal{F}_3$: $\{a,b\} \cup \{c,d\} = \Omega \in \mathcal{F}_3$.

Solution 2. (Lebesgue Integral)

The function f takes three values: 3 on $[0,1) \cap \mathbb{Q}$, 1 on $[0,1) \setminus \mathbb{Q}$, and 2 on $[1,2]$.

Computing the Lebesgue integral:

$$\begin{aligned}
\int_{[0,2]} f \, \mathrm{d}\lambda &= 3 \cdot \lambda([0,1) \cap \mathbb{Q}) + 1 \cdot \lambda([0,1) \setminus \mathbb{Q}) + 2 \cdot \lambda([1,2]) \\
&= 3 \cdot 0 + 1 \cdot 1 + 2 \cdot 1 \\
&= 0 + 1 + 2 = \boxed{3}
\end{aligned}$$

Here we used that $\lambda(\mathbb{Q} \cap [0,1)) = 0$ (rationals have measure zero) and $\lambda([0,1) \setminus \mathbb{Q}) = 1 - 0 = 1$.

Riemann integrability: The function is **NOT Riemann integrable** on $[0,1)$. For any partition of $[0,1)$, each subinterval contains both rationals and irrationals (both sets are dense in $\mathbb{R}$). Therefore, the upper Riemann sum over $[0,1)$ equals $3 \cdot 1 = 3$ while the lower sum equals $1 \cdot 1 = 1$. Since they don't converge to the same value, the Riemann integral doesn't exist on $[0,1)$.

Solution 3. (Probability Space Construction)

(a) Since $\mathbb{P}(\{k\}) \propto k$, we have $\mathbb{P}(\{k\}) = \frac{k}{C}$ where $C = 1+2+3+4+5+6 = 21$. Therefore:

$$\mathbb{P}(\{1\}) = \frac{1}{21}, \quad \mathbb{P}(\{2\}) = \frac{2}{21}, \quad \mathbb{P}(\{3\}) = \frac{3}{21} = \frac{1}{7},$$
$$\mathbb{P}(\{4\}) = \frac{4}{21}, \quad \mathbb{P}(\{5\}) = \frac{5}{21}, \quad \mathbb{P}(\{6\}) = \frac{6}{21} = \frac{2}{7}.$$

(b) The prime outcomes are $\{2,3,5\}$. The random variable X is defined as:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \in \{2,3,5\}, \\ 0 & \text{otherwise.} \end{cases}$$

The PMF of X is:

$$\mathbb{P}(X=1) = \mathbb{P}(\{2,3,5\}) = \frac{2+3+5}{21} = \frac{10}{21},$$
$$\mathbb{P}(X=0) = 1 - \frac{10}{21} = \frac{11}{21}.$$

(c) Computing the moments:

$$\mathbb{E}[X] = 0 \cdot \frac{11}{21} + 1 \cdot \frac{10}{21} = \frac{10}{21} \approx 0.476.$$
$$\mathbb{E}[X^2] = 0^2 \cdot \frac{11}{21} + 1^2 \cdot \frac{10}{21} = \frac{10}{21}.$$
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{10}{21} - \frac{100}{441} = \frac{210-100}{441} = \frac{110}{441} \approx 0.249.$$

Solution 4. (Gaussian Properties)

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

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with PDF $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

For the transformation $Z = g(X) = \frac{X-\mu}{\sigma}$, the inverse is $X = h(Z) = \sigma Z + \mu$, and $\frac{dh}{dZ} = \sigma$.

Using the change of variables formula:

$$\begin{aligned}
f_Z(z) &= f_X(h(z)) \cdot \left|\frac{dh}{dz}\right| = f_X(\sigma z + \mu) \cdot \sigma \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}\right) \cdot \sigma \\
&= \frac{\sigma}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\sigma^2 z^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).
\end{aligned}$$

This is exactly the PDF of $\mathcal{N}(0,1)$. $\square$

Application:

$$\begin{aligned}
\mathbb{P}(\mu - 2\sigma \le X \le \mu + 2\sigma) &= \mathbb{P}\left(\frac{\mu - 2\sigma - \mu}{\sigma} \le Z \le \frac{\mu + 2\sigma - \mu}{\sigma}\right) \\
&= \mathbb{P}(-2 \le Z \le 2) \\
&= \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \\
&\approx 2(0.9772) - 1 = 0.9545.
\end{aligned}$$

Solution 5. (Expectation of Functions)

(a) **Verification:** We check non-negativity and normalization.

- Non-negativity: $f_X(x) = 2x \ge 0$ for $x \in [0,1]$.
- Normalization: $\int_0^1 2x\,dx = [x^2]_0^1 = 1.$ ✓

(b) **Computing moments:**

$$\mathbb{E}[X] = \int_0^1 x \cdot 2x\, dx = 2\int_0^1 x^2\, dx = 2 \cdot \frac{x^3}{3}\Big|_0^1 = \frac{2}{3}.$$

$$\mathbb{E}[X^2] = \int_0^1 x^2 \cdot 2x\, dx = 2\int_0^1 x^3\, dx = 2 \cdot \frac{x^4}{4}\Big|_0^1 = \frac{1}{2}.$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{2} - \frac{4}{9} = \frac{9-8}{18} = \frac{1}{18}.$$

(c) **Computing $\mathbb{E}[e^X]$:**

$$\mathbb{E}[e^X] = \int_0^1 e^x \cdot 2x\, dx = 2\int_0^1 xe^x\, dx.$$

Using integration by parts with $u = x$, $dv = e^x dx$:

$$\int_0^1 xe^x\, dx = [xe^x]_0^1 - \int_0^1 e^x\, dx = e - [e^x]_0^1 = e - (e-1) = 1.$$

Therefore, $\mathbb{E}[e^X] = 2 \cdot 1 = 2$.

(d) **Computing $\mathbb{E}[Y]$:** For $Y = 3X^2 - 2X + 1$, using linearity:

$$\mathbb{E}[Y] = 3\mathbb{E}[X^2] - 2\mathbb{E}[X] + 1 = 3 \cdot \frac{1}{2} - 2 \cdot \frac{2}{3} + 1 = \frac{3}{2} - \frac{4}{3} + 1 = \frac{9-8+6}{6} = \frac{7}{6}.$$

Solution 6. (Binomial Distribution)

Let $X \sim \text{Binomial}(15, 0.7)$.

(a) $\mathbb{P}(X = 12) = \binom{15}{12}(0.7)^{12}(0.3)^3 = 455 \cdot 0.01384 \cdot 0.027 \approx \boxed{0.170}$.

(b) $\mathbb{P}(X \geq 10) = \sum_{k=10}^{15} \binom{15}{k}(0.7)^k(0.3)^{15-k}$.
Computing each term:

$$\mathbb{P}(X = 10) \approx 0.206, \quad \mathbb{P}(X = 11) \approx 0.219, \quad \mathbb{P}(X = 12) \approx 0.170,$$
$$\mathbb{P}(X = 13) \approx 0.092, \quad \mathbb{P}(X = 14) \approx 0.031, \quad \mathbb{P}(X = 15) \approx 0.005.$$

Therefore, $\mathbb{P}(X \geq 10) \approx \boxed{0.722}$.

(c) $\mathbb{E}[X] = np = 15 \cdot 0.7 = \boxed{10.5}$.
$\sigma = \sqrt{np(1-p)} = \sqrt{15 \cdot 0.7 \cdot 0.3} = \sqrt{3.15} \approx \boxed{1.775}$.

(d) **Normal approximation with continuity correction:**
$X \overset{\text{approx}}{\sim} \mathcal{N}(10.5, 3.15)$. With continuity correction:

$$\mathbb{P}(X \geq 10) \approx \mathbb{P}(X > 9.5) = \mathbb{P}\left(Z > \frac{9.5 - 10.5}{1.775}\right) = \mathbb{P}(Z > -0.563).$$

$$= 1 - \Phi(-0.563) = \Phi(0.563) \approx 0.713.$$

This is close to the exact value of 0.722, with error about 1.2%.

Solution 7. (Variance of Sums)

Proof: Let $S = \sum_{i=1}^n X_i$ and $\mu_S = \mathbb{E}[S] = \sum_{i=1}^n \mu_i$ (by linearity).

$$\begin{aligned}
\operatorname{Var}(S) = \mathbb{E}[(S-\mu_S)^2] &= \mathbb{E}\left[\left(\sum_{i=1}^n (X_i-\mu_i)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n (X_i-\mu_i)^2 + 2\sum_{i<j}(X_i-\mu_i)(X_j-\mu_j)\right] \\
&= \sum_{i=1}^n \mathbb{E}[(X_i-\mu_i)^2] + 2\sum_{i<j}\mathbb{E}[(X_i-\mu_i)(X_j-\mu_j)] \\
&= \sum_{i=1}^n \operatorname{Var}(X_i) + 2\sum_{i<j}\operatorname{Cov}(X_i, X_j).
\end{aligned}$$

Since the X_i are independent, $\operatorname{Cov}(X_i, X_j) = 0$ for $i \neq j$. Therefore:

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \sigma_i^2. \quad \square$$

Variance of sample mean: If all $\sigma_i^2 = \sigma^2$, then:

$$\operatorname{Var}(\bar{X}) = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}. \quad \square$$

Solution 8. (Covariance Properties)

(a) $\operatorname{Cov}(X, X) = \mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])] = \mathbb{E}[(X-\mathbb{E}[X])^2] = \operatorname{Var}(X). \quad \square$

(b) Let $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$. Then $\mathbb{E}[aX+b] = a\mu_X + b$ and $\mathbb{E}[cY+d] = c\mu_Y + d$.

$$\begin{aligned}
\operatorname{Cov}(aX+b, cY+d) &= \mathbb{E}[(aX+b-a\mu_X-b)(cY+d-c\mu_Y-d)] \\
&= \mathbb{E}[a(X-\mu_X)\cdot c(Y-\mu_Y)] \\
&= ac\cdot\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = ac\cdot\operatorname{Cov}(X,Y). \quad \square
\end{aligned}$$

(c)

$$\begin{aligned}
\operatorname{Var}(X+Y) &= \mathbb{E}[(X+Y-\mathbb{E}[X]-\mathbb{E}[Y])^2] \\
&= \mathbb{E}[((X-\mu_X)+(Y-\mu_Y))^2] \\
&= \mathbb{E}[(X-\mu_X)^2] + \mathbb{E}[(Y-\mu_Y)^2] + 2\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y). \quad \square
\end{aligned}$$

(d) If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. Therefore:

$$\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = 0. \quad \square$$

Solution 9. (MLE for Gaussian)

(a) **Likelihood:**

$$L(\mu, \sigma^2|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right).$$

(b) **Log-likelihood:**

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

(c) **MLE for μ:** Taking derivative w.r.t. μ:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - n\mu\right).$$

Setting to zero: $\sum_{i=1}^{n} x_i = n\hat{\mu}$, so $\hat{\mu}_{MLE} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. $\square$

(d) **MLE for σ^2:** Taking derivative w.r.t. σ^2:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Setting to zero and substituting $\hat{\mu} = \bar{x}$:

$$\frac{n}{2\hat{\sigma}^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2(\hat{\sigma}^2)^2} \implies \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2. \quad \square$$

(e) **Bias:** The MLE for variance is biased because:

$$\mathbb{E}[\hat{\sigma}^2_{MLE}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2.$$

The unbiased estimator is $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$.

Solution 10. (MLE for Exponential)

(a) **Likelihood:**

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \exp\left(-\lambda\sum_{i=1}^{n} x_i\right).$$

Log-likelihood:

$$\ell(\lambda) = n\log\lambda - \lambda\sum_{i=1}^{n} x_i.$$

(b) **MLE:** Taking derivative:

$$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i.$$

Setting to zero: $\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$.

(c) **Numerical computation:** With data $\{1.2, 0.8, 2.1, 1.5, 0.9\}$:

$$\bar{x} = \frac{1.2 + 0.8 + 2.1 + 1.5 + 0.9}{5} = \frac{6.5}{5} = 1.3.$$

Therefore, $\hat{\lambda}_{MLE} = \frac{1}{1.3} \approx 0.769$.

(d) **Bias analysis:** The MLE $\hat{\lambda} = 1/\bar{X}$ is **biased**.
For $X_i \sim \text{Exponential}(\lambda)$, we have $\bar{X} \sim \text{Gamma}(n, n\lambda)$.
Using properties of the inverse: $\mathbb{E}[1/\bar{X}] = \frac{n\lambda}{n-1}$ for $n > 1$.
Thus $\mathbb{E}[\hat{\lambda}_{MLE}] = \frac{n}{n-1}\lambda \neq \lambda$, so the estimator is biased.
An unbiased estimator would be $\tilde{\lambda} = \frac{n-1}{n\bar{x}} = \frac{n-1}{\sum_{i=1}^{n} x_i}$.

Solution 11. (Central Limit Theorem Application)

(a) By CLT, $\bar{X} \overset{\text{approx}}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) = \mathcal{N}\left(50, \frac{16}{64}\right) = \mathcal{N}(50, 0.25)$.
The standard error is $\sigma_{\bar{X}} = \frac{4}{\sqrt{64}} = 0.5$ grams.

(b)

$$\begin{aligned}\mathbb{P}(49 < \bar{X} < 51) &= \mathbb{P}\left(\frac{49 - 50}{0.5} < Z < \frac{51 - 50}{0.5}\right) \\ &= \mathbb{P}(-2 < Z < 2) = 2\Phi(2) - 1 \approx 0.9545.\end{aligned}$$

(c) We need $\mathbb{P}(\bar{X} > c) = 0.05$, so $\mathbb{P}(\bar{X} \leq c) = 0.95$.

$$\mathbb{P}\left(Z \leq \frac{c - 50}{0.5}\right) = 0.95 \implies \frac{c - 50}{0.5} = 1.645.$$

Therefore, $c = 50 + 0.5 \times 1.645 = 50.82$ grams.

(d) We need $\mathbb{P}(|\bar{X} - 50| < 0.5) \geq 0.99$.

$$\mathbb{P}\left(-\frac{0.5}{4/\sqrt{n}} < Z < \frac{0.5}{4/\sqrt{n}}\right) \geq 0.99.$$

$$2\Phi\left(\frac{0.5\sqrt{n}}{4}\right) - 1 \geq 0.99 \implies \Phi\left(\frac{\sqrt{n}}{8}\right) \geq 0.995.$$

From standard normal tables, $\Phi^{-1}(0.995) \approx 2.576$.

$$\frac{\sqrt{n}}{8} \geq 2.576 \implies \sqrt{n} \geq 20.61 \implies n \geq 424.7.$$

Therefore, $n \geq 425$ bolts are needed.

Solution 12. (Bayes' Theorem)

(a) **Probability of defective item:**

$$\begin{aligned}\mathbb{P}(\text{Def}) &= \mathbb{P}(\text{Def}|A)\mathbb{P}(A) + \mathbb{P}(\text{Def}|B)\mathbb{P}(B) + \mathbb{P}(\text{Def}|C)\mathbb{P}(C) \\ &= 0.02 \times 0.50 + 0.03 \times 0.30 + 0.05 \times 0.20 \\ &= 0.010 + 0.009 + 0.010 = 0.029 = 2.9\%.\end{aligned}$$

(b) **Posterior probabilities:**

$$\mathbb{P}(A|\text{Def}) = \frac{\mathbb{P}(\text{Def}|A)\mathbb{P}(A)}{\mathbb{P}(\text{Def})} = \frac{0.02 \times 0.50}{0.029} = \frac{0.010}{0.029} \approx 0.345,$$

$$\mathbb{P}(B|\text{Def}) = \frac{0.03 \times 0.30}{0.029} = \frac{0.009}{0.029} \approx 0.310,$$

$$\mathbb{P}(C|\text{Def}) = \frac{0.05 \times 0.20}{0.029} = \frac{0.010}{0.029} \approx 0.345.$$

Note: These sum to 1, as expected.

(c) **Both defective items from Machine C:**
Using independence, if two items are selected and both are defective:

$$\mathbb{P}(\text{both from } C|\text{both defective}) = [\mathbb{P}(C|\text{Def})]^2 \approx (0.345)^2 \approx 0.119.$$

Solution 13. (Bayesian Inference)

(a) **Prior moments:** For $\theta \sim \text{Beta}(2, 2)$:

$$\mathbb{E}[\theta] = \frac{\alpha}{\alpha + \beta} = \frac{2}{4} = 0.5.$$

$$\text{Var}(\theta) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{2 \cdot 2}{16 \cdot 5} = \frac{4}{80} = 0.05.$$

(b) **Posterior distribution:** With $h = 8$ heads and $t = 2$ tails from $n = 10$ flips:

$$\theta|D \sim \text{Beta}(\alpha + h, \beta + t) = \text{Beta}(2 + 8, 2 + 2) = \text{Beta}(10, 4).$$

(c) **Posterior mean:**

$$\mathbb{E}[\theta|D] = \frac{10}{10 + 4} = \frac{10}{14} = \frac{5}{7} \approx 0.714.$$

MLE: $\hat{\theta}_{MLE} = \frac{h}{n} = \frac{8}{10} = 0.8$.
The posterior mean (0.714) is pulled toward the prior mean (0.5) relative to the MLE (0.8).

(d) **MAP estimate:** For $\text{Beta}(\alpha, \beta)$, the mode (MAP) is:

$$\hat{\theta}_{MAP} = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{10 - 1}{10 + 4 - 2} = \frac{9}{12} = 0.75.$$

Summary: MLE = 0.8, MAP = 0.75, Posterior Mean = 0.714. The prior information pulls estimates toward 0.5.

(e) **95% Credible interval:** The interval is $(q_{0.025}, q_{0.975})$ where q_p denotes the p-th quantile of Beta(10, 4).
Using Beta distribution tables or software:

$$q_{0.025} \approx 0.491, \quad q_{0.975} \approx 0.897.$$

The 95% credible interval is approximately $(0.491, 0.897)$.

Solution 14. (Naive Bayes Classification)

(a) **Computing posteriors for $\mathbf{x} = (3, 4)$:**
For class C_1: $\mu_1 = 2, \sigma_1 = 1$ for x_1 and $\mu_2 = 5, \sigma_2 = 2$ for x_2.

$$\mathbb{P}(x_1 = 3|C_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(3-2)^2}{2}\right) = \frac{1}{\sqrt{2\pi}} e^{-0.5} \approx 0.242,$$
$$\mathbb{P}(x_2 = 4|C_1) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{(4-5)^2}{8}\right) = \frac{1}{2\sqrt{2\pi}} e^{-0.125} \approx 0.176.$$

Unnormalized posterior: $\mathbb{P}(C_1) \cdot \mathbb{P}(x_1|C_1) \cdot \mathbb{P}(x_2|C_1) = 0.4 \times 0.242 \times 0.176 \approx 0.0170$.
For class C_2: $\mu_1 = 4, \sigma_1 = 1.5$ and $\mu_2 = 3, \sigma_2 = 1$.

$$\mathbb{P}(x_1 = 3|C_2) = \frac{1}{1.5\sqrt{2\pi}} \exp\left(-\frac{(3-4)^2}{4.5}\right) \approx 0.213,$$
$$\mathbb{P}(x_2 = 4|C_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(4-3)^2}{2}\right) \approx 0.242.$$

Unnormalized posterior: $0.6 \times 0.213 \times 0.242 \approx 0.0309$.

(b) **Classification:** Since $0.0309 > 0.0170$, classify as C_2.
Normalized posterior:

$$\mathbb{P}(C_2|\mathbf{x}) = \frac{0.0309}{0.0170 + 0.0309} = \frac{0.0309}{0.0479} \approx \boxed{0.645}.$$

(c) **Decision boundary:** The boundary occurs where:

$$\mathbb{P}(C_1)\mathbb{P}(x_1|C_1)\mathbb{P}(x_2|C_1) = \mathbb{P}(C_2)\mathbb{P}(x_1|C_2)\mathbb{P}(x_2|C_2).$$

Taking logarithms and simplifying (after substantial algebra):

$$\log(0.4) - \frac{(x_1-2)^2}{2} - \frac{(x_2-5)^2}{8} - \log(2) = \log(0.6) - \frac{(x_1-4)^2}{4.5} - \frac{(x_2-3)^2}{2} - \log(1.5).$$

This simplifies to a quadratic equation in x_1 and x_2:

$$0.278x_1^2 - 0.125x_2^2 - 0.778x_1 + 1.75x_2 + C = 0,$$

where C is a constant determined by the remaining terms. This is a hyperbola in the (x_1, x_2) plane.

Solution 15. (Comprehensive Problem)

(a) **Joint PMF:**

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i} = p^T(1-p)^{n-T},$$

where $T = \sum_{i=1}^n x_i$. The joint PMF depends only on T, not on the individual values. $\square$

(b) **Distribution of T:** By definition, $T \sim \text{Binomial}(n, p)$.

$$\mathbb{E}[T] = np, \qquad \text{Var}(T) = np(1-p). \quad \square$$

(c) **MLE and unbiasedness:** The log-likelihood is:

$$\ell(p) = T \log p + (n - T) \log(1 - p).$$

Taking derivative: $\frac{d\ell}{dp} = \frac{T}{p} - \frac{n-T}{1-p} = 0$. Solving: $\hat{p} = T/n$.
Unbiasedness: $\mathbb{E}[\hat{p}] = \mathbb{E}[T/n] = \frac{1}{n}\mathbb{E}[T] = \frac{np}{n} = p$. $\square$

(d) **Chebyshev bound:**

$$\operatorname{Var}(\hat{p}) = \operatorname{Var}(T/n) = \frac{1}{n^2}\operatorname{Var}(T) = \frac{p(1-p)}{n}.$$

By Chebyshev: $\mathbb{P}(|\hat{p} - p| \geq \epsilon) \leq \frac{\operatorname{Var}(\hat{p})}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}$.
Since $p(1-p) \leq \frac{1}{4}$ for all $p \in [0, 1]$ (maximum at $p = 0.5$):

$$\mathbb{P}(|\hat{p} - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}. \quad \square$$

(e) **CLT application:** By CLT, for large n:

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Therefore, $\hat{p} \overset{\text{approx}}{\sim} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$. $\square$

(f) **95% Confidence interval:** With $T = 45$ and $n = 500$:

$$\hat{p} = \frac{45}{500} = 0.09.$$

Using the approximate variance $\hat{p}(1 - \hat{p})/n = \frac{0.09 \times 0.91}{500} = 0.0001638$:

$$\text{SE} = \sqrt{0.0001638} \approx 0.0128.$$

The 95% CI is:

$$\hat{p} \pm 1.96 \times \text{SE} = 0.09 \pm 1.96 \times 0.0128 = 0.09 \pm 0.025.$$

The 95% confidence interval is approximately $(0.065, 0.115)$ or $(6.5\%, 11.5\%)$.