EXERCISE · SLIDE 6 · QUESTION

Probability Mass Function (PMF)

Which values belong to the event?

Errors $-1, 0, 2$ cm have masses $1/2, 1/4, 1/4$. What is $P(x \geq 0)$?

- **A** $1/4$: count only positive errors.
- **B** $1/2$: include zero and positive errors.
- **C** $2/3$: two of three possible values.

Event probability; PMF — [1] DL §3.3.1 · Course question

EXERCISE · SLIDE 6 · SOLUTION 1/3

Probability Mass Function (PMF)

Which values belong to the event?

Solution 1 / 3

Identify the event

x: error (cm); $p(a) = P(x = a)$: probability mass at value a.

---

1. x: error in cm; $p(a) = P(x = a)$: probability mass at a.
2. Possible values: $-1, 0, 2$; masses: $1/2, 1/4, 1/4$.
3. Nonnegative means $x \geq 0$, so include 0 and 2.

Event probability; PMF — [1] DL §3.3.1 · Course question

Probability Mass Function (PMF)

Which values belong to the event?

Solution 2 / 3

Add the included masses

x: error (cm); $p(a) = P(x = a)$: probability mass at value a.

1. $P(x \geq 0) = P(x = 0) + P(x = 2).$
2. $= 1/4 + 1/4 = 2/4 = 1/2.$
3. Check: $P(x < 0) + P(x \geq 0) = 1/2 + 1/2 = 1.$

Event probability; PMF — [1] DL §3.3.1 · Course question

EXERCISE · SLIDE 6 · SOLUTION 3/3

Probability Mass Function (PMF)

Which values belong to the event?

Solution 3 / 3

Choose B

x: error (cm); $p(a) = P(x = a)$: probability mass at value a.

---

1. A excludes zero and answers a different event.
2. B includes both nonnegative values and their masses.
3. C treats unequal masses as equal. Count probability, not values.

Event probability; PMF — [1] DL §3.3.1 · Course question

EXERCISE · SLIDE 6 · CHECK THE RESULT

Probability Mass Function (PMF)

Which values belong to the event?

Errors $-1, 0, 2$ cm have masses $1/2, 1/4, 1/4$. What is $P(x \geq 0)$?

A $1/4$: count only positive errors.

B $1/2$: include zero and positive errors. ✓

C $2/3$: two of three possible values.

> $P(x \geq 0) = p(0) + p(2) = 1/4 + 1/4 = 1/2$. Total mass is 1.

Event probability; PMF — [1] DL §3.3.1 · Course question

EXERCISE · SLIDE 9 · QUESTION

Probability Density Function (PDF)

Density is height; probability is area

x is uniform on $[0, .5]$ cm. Which statement is valid?

A. Density 2 cm^{-1}; $P(.1 \leq x \leq .3) = .4$.

B. Density 2 means event probability 2.

C. Probability is $.2$, the interval width alone.

Density; event probability — [1] DL §3.3.2 · Course question

EXERCISE · SLIDE 9 · SOLUTION 1/3

Probability Density Function (PDF)

Density is height; probability is area

Solution 1 / 3

Normalize the density

x: error (cm); h: uniform density; t: integration variable (cm).

---

1. x: error in cm; h: constant density on $[0, .5]$.
2. Total probability is area: $h(.5 - 0) = 1$.
3. $h = 1/.5 = 2\ \text{cm}^{-1}$.

Density; event probability — [1] DL §3.3.2 · Course question

Probability Density Function (PDF)

Density is height; probability is area

Solution 2 / 3

Integrate over the requested interval

x: error (cm); h: uniform density; t: integration variable (cm).

1. $P(.1 \le x \le .3) = \int_{.1}^{.3} h\, dt.$
2. $= 2(.3 - .1) = 2(.2) = .4.$
3. Density times width has units $\mathrm{cm}^{-1}\mathrm{cm} = 1.$

Density; event probability — [1] DL §3.3.2 · Course question

EXERCISE · SLIDE 9 · SOLUTION 3/3

Probability Density Function (PDF)

Density is height; probability is area

Solution 3 / 3

Choose A

x: error (cm); h: uniform density; t: integration variable (cm).

1. A gives both the normalized density and event probability.
2. B confuses density height with probability; probabilities cannot exceed 1.
3. C omits the height: width alone is not the area.

Density; event probability — [1] DL §3.3.2 · Course question

EXERCISE · SLIDE 9 · CHECK THE RESULT

Probability Density Function (PDF)

Density is height; probability is area

x is uniform on $[0, .5]$ cm. Which statement is valid?

A. Density 2 cm^{-1}; $P(.1 \le x \le .3) = .4$. ✓

B. Density 2 means event probability 2.

C. Probability is .2, the interval width alone.

> $h(.5) = 1 \Rightarrow h = 2$; mass $= h(.3 - .1) = .4$.

Density; event probability — [1] DL §3.3.2 · Course question

EXERCISE · SLIDE 11 · QUESTION

Moments

What does centering remove?

$\mathrm{Var}(x) = \mathbb{E}[(x-\mu)^2]$. Given $\mu = 1/4$ cm and $\mathbb{E}[x^2] = 5/4$ cm^2, choose the variance.

A $5/4$: the second moment is always variance.

B 1: subtract the mean, not its square.

C $19/16$: subtract the squared mean.

Centering; variance — [1] DL §3.8 · Course question

EXERCISE · SLIDE 11 · SOLUTION 1/3

Moments

What does centering remove?

Solution 1 / 3

Expand the centered square

x: error; $\mathbb{E}$: expectation; $\mu = \mathbb{E}[x]$: mean; Var: variance.

1. x: error; $\mu = \mathbb{E}[x] = 1/4$ cm.
2. $\mathbb{E}[x^2] = 5/4$ cm^2 is the second moment.
3. $(x - \mu)^2 = x^2 - 2\mu x + \mu^2$.

Centering; variance — [1] DL §3.8 · Course question

Moments

What does centering remove?

Solution 2 / 3

Take expectation, then substitute

x: error; $\mathbb{E}$: expectation; $\mu = \mathbb{E}[x]$: mean; Var: variance.

1. $\mathrm{Var}(x) = \mathbb{E}[x^2] - 2\mu\mathbb{E}[x] + \mu^2.$
2. $\mathbb{E}[x] = \mu \;\Rightarrow\; \mathrm{Var}(x) = \mathbb{E}[x^2] - \mu^2.$
3. $= 5/4 - (1/4)^2 = 20/16 - 1/16 = 19/16 \text{ cm}^2.$

Centering; variance — [1] DL §3.8 · Course question

EXERCISE · SLIDE 11 · SOLUTION 3/3

Moments

What does centering remove?

Solution 3 / 3

Choose C

x: error; $\mathbb{E}$: expectation; $\mu = \mathbb{E}[x]$: mean; Var: variance.

1. A leaves the squared mean inside the second moment.
2. B subtracts a length from a squared length; its units do not match.
3. C subtracts the squared mean and measures spread around the mean.

Centering; variance — [1] DL §3.8 · Course question

EXERCISE · SLIDE 11 · CHECK THE RESULT

Moments

What does centering remove?

$\mathrm{Var}(x) = \mathbb{E}[(x-\mu)^2]$. Given $\mu = 1/4$ cm and $\mathbb{E}[x^2] = 5/4$ cm^2, choose the variance.

A $5/4$: the second moment is always variance.

B 1: subtract the mean, not its square.

C $19/16$: subtract the squared mean. ✓

> $\mathrm{Var}(x) = \mathbb{E}[x^2] - \mu^2 = 5/4 - 1/16 = 19/16$ cm^2.

Centering; variance — [1] DL §3.8 · Course question

EXERCISE · SLIDE 16 · QUESTION

Moments

Independent sources, dependent components

U, V are independent fair signs. Let $x = (U, 2U + V)^T$. What is $\mathrm{Cov}(x_1, x_2)$?

A 2: both components share U.

B 0: independent sources imply independent outputs.

C 1: doubling a coefficient has no effect.

Independence; covariance — [1] DL §3.7 · [1] DL §3.8 · Course question

EXERCISE · SLIDE 16 · SOLUTION 1/4

Moments

Independent sources, dependent components

Solution 1 / 4

Separate sources from components

U, V: independent fair signs; $x = (U, 2U + V)^T$; $\Sigma_{ij} = \mathrm{Cov}(x_i, x_j)$.

1. U, V: independent signs, each equally likely -1 or 1.
2. $\mathbb{E}[U] = \mathbb{E}[V] = (-1 + 1)/2 = 0$; hence $\mathbb{E}[x_1] = \mathbb{E}[x_2] = 0$.
3. $\mathbb{E}[U^2] = \mathbb{E}[V^2] = 1$; $\mathbb{E}[UV] = \mathbb{E}[U]\mathbb{E}[V] = 0$.

Independence; covariance — [1] DL §3.7 · [1] DL §3.8 · Course question

EXERCISE · SLIDE 16 · SOLUTION 2/4

Moments

Independent sources, dependent components

Solution 2 / 4

Compute cross-covariance

U, V: independent fair signs; $x = (U, 2U + V)^T$; $\Sigma_{ij} = \mathrm{Cov}(x_i, x_j)$.

1. $\Sigma_{12} = \mathrm{Cov}(x_1, x_2) = \mathbb{E}[x_1 x_2]$.
2. $= \mathbb{E}[U(2U + V)] = 2\mathbb{E}[U^2] + \mathbb{E}[UV]$.
3. $= 2(1) + 0 = 2$.

Independence; covariance — [1] DL §3.7 · [1] DL §3.8 · Course question

Moments

Independent sources, dependent components

Solution 3 / 4

Check the complete covariance

U, V: independent fair signs; $x = (U, 2U + V)^T$; $\Sigma_{ij} = \mathrm{Cov}(x_i, x_j)$.

1. $\Sigma_{11} = \mathbb{E}[U^2] = 1.$
2. $\Sigma_{22} = \mathbb{E}[(2U + V)^2] = 4(1) + 4(0) + 1 = 5.$
3. $\Sigma = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$; $\det \Sigma = 1(5) - 2(2) = 1.$

Independence; covariance — [1] DL §3.7 · [1] DL §3.8 · Course question

Moments

Independent sources, dependent components

Solution 4 / 4

Choose A

U, V: independent fair signs; $x = (U, 2U + V)^T$; $\Sigma_{ij} = \mathrm{Cov}(x_i, x_j)$.

1. A retains the shared source U.
2. B confuses independence of U and V with independence of their outputs.
3. C ignores the coefficient 2 in the shared contribution.

Independence; covariance — [1] DL §3.7 · [1] DL §3.8 · Course question

EXERCISE · SLIDE 16 · CHECK THE RESULT

Moments

Independent sources, dependent components

U, V are independent fair signs. Let $x = (U, 2U + V)^T$. What is $\mathrm{Cov}(x_1, x_2)$?

- **A** 2: both components share U. ✓
- **B** 0: independent sources imply independent outputs.
- **C** 1: doubling a coefficient has no effect.

> $\mathbb{E}[U(2U + V)] = 2$; $\Sigma = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$, $\det \Sigma = 1$.

Independence; covariance — [1] DL §3.7 · [1] DL §3.8 · Course question

EXERCISE · SLIDE 22 · QUESTION

Gaussian geometry

Which direction makes a point less surprising?

$\mu = 0$, $\Sigma = \mathrm{diag}(4, 1)$; $u = (2, 0)^T$, $v = (0, 2)^T$. Compare Gaussian densities.

A $p(u) = p(v)$: their Euclidean lengths agree.

B $p(u) > p(v)$: u is one standard deviation away.

C $p(u) < p(v)$: higher variance always lowers density.

Mahalanobis distance — [2] UBC Lec.18 p.5 · Course question

EXERCISE · SLIDE 22 · SOLUTION 1/3

Gaussian geometry

Which direction makes a point less surprising?

Solution 1 / 3

Use the covariance-scaled distance

μ: mean; Σ: covariance; p: density; d_M: covariance-scaled distance.

1. $\mu = 0$: mean; $\Sigma = \mathrm{diag}(4, 1)$: covariance.
2. $u = (2, 0)^T$, $v = (0, 2)^T$ are points in the same Gaussian.
3. $d_M^2(x) = (x - \mu)^T \Sigma^{-1} (x - \mu) = x_1^2/4 + x_2^2$.

Mahalanobis distance — [2] UBC Lec.18 p.5 · Course question

Gaussian geometry

Which direction makes a point less surprising?

Solution 2 / 3

Compare exponents

μ: mean; Σ: covariance; p: density; d_M: covariance-scaled distance.

1. $d_M^2(u) = 2^2/4 + 0^2 = 1$; $d_M^2(v) = 0^2/4 + 2^2 = 4$.
2. $p(x) = c\exp(-d_M^2(x)/2)$; the normalization c is shared.
3. $p(u)/p(v) = e^{-1/2}/e^{-4/2} = e^{3/2} > 1$.

Mahalanobis distance — [2] UBC Lec.18 p.5 · Course question

EXERCISE · SLIDE 22 · SOLUTION 3/3

Gaussian geometry

Which direction makes a point less surprising?

Solution 3 / 3

Choose B

μ: mean; Σ: covariance; p: density; d_M: covariance-scaled distance.

1. A uses ordinary distance and ignores different directional variances.
2. B: u is one standard deviation away; v is two.
3. C confuses changing covariance with comparing points in one fixed model.

Mahalanobis distance — [2] UBC Lec.18 p.5 · Course question

EXERCISE · SLIDE 22 · CHECK THE RESULT

Gaussian geometry

Which direction makes a point less surprising?

$\mu = 0$, $\Sigma = \mathrm{diag}(4, 1)$; $u = (2, 0)^T$, $v = (0, 2)^T$. Compare Gaussian densities.

A $p(u) = p(v)$: their Euclidean lengths agree.

B $p(u) > p(v)$: u is one standard deviation away. ✓

C $p(u) < p(v)$: higher variance always lowers density.

$d_M^2(u) = 1$, $d_M^2(v) = 4$; $p(u)/p(v) = e^{3/2} > 1$.

Mahalanobis distance — [2] UBC Lec.18 p.5 · Course question

EXERCISE · SLIDE 28 · QUESTION

Bayes

How does prevalence change the chance an alarm is true?

$P(H) = .1$, $P(+ \mid H) = .9$, $P(+ \mid \neg H) = .1$. What is $P(H \mid +)$?

A .9: the chance of a true event equals sensor sensitivity.

B .1: an observation cannot change the prior.

C .5: true and false positives contribute equally.

Prior probability; posterior probability — [1] DL §3.11 · Course question

EXERCISE · SLIDE 28 · SOLUTION 1/3

Bayes

How does prevalence change the chance an alarm is true?

Solution 1 / 3

Name the conditional populations

H: event; $\neg H$: no event; $+$: positive reading; P: probability.

---

1. H: event; $\neg H$: no event; $+$: positive test.
2. $P(H) = .1$, $P(\neg H) = 1 - .1 = .9$.
3. $P(+ \mid H) = .9$; $P(+ \mid \neg H) = .1$.

Prior probability; posterior probability — [1] DL §3.11 · Course question

Bayes

How does prevalence change the chance an alarm is true?

Solution 2 / 3

Find all positives, then normalize

H: event; $\neg H$: no event; $+$: positive reading; P: probability.

1. $P(H \cap +) = P(+ \mid H)P(H) = .9(.1) = .09.$
2. $P(\neg H \cap +) = P(+ \mid \neg H)P(\neg H) = .1(.9) = .09.$
3. $P(+) = .09 + .09 = .18$; $P(H \mid +) = .09/.18 = .5.$

Prior probability; posterior probability — [1] DL §3.11 · Course question

Bayes

How does prevalence change the chance an alarm is true?

Solution 3 / 3

Choose C

H: event; $\neg H$: no event; $+$: positive reading; P: probability.

1. A reverses the conditioning: sensitivity starts with event cases.
2. B uses the prior and ignores the positive observation.
3. C: true and false positives have equal probability mass.

Prior probability; posterior probability — [1] DL §3.11 · Course question

EXERCISE · SLIDE 28 · CHECK THE RESULT

Bayes

How does prevalence change the chance an alarm is true?

$P(H) = .1$, $P(+ \mid H) = .9$, $P(+ \mid \neg H) = .1$. What is $P(H \mid +)$?

A .9: the chance of a true event equals sensor sensitivity.

B .1: an observation cannot change the prior.

C .5: true and false positives contribute equally. ✓

$$P(H \mid +) = \frac{.9(.1)}{.9(.1) + .1(.9)} = .5.$$

Prior probability; posterior probability — [1] DL §3.11 · Course question

EXERCISE · SLIDE 34 · QUESTION

Gaussian conditioning

What does a noisier reading change?

Prior $\mathcal{N}(0, 4)$; $z = 3$ with independent Gaussian noise, variance R: $1 \to 4$. Compare posterior mean and variance.

A. Mean moves toward 0; posterior variance increases.

B. Mean moves toward 3; posterior variance decreases.

C. Only the mean changes; variance ignores sensor noise.

Gaussian conditioning — [2] UBC Lec.18 p.7 · Course question

Gaussian conditioning

What does a noisier reading change?

Solution 1 / 3

State the scalar measurement model

μ_0, P_0: prior mean/variance; μ_1, P_1: posterior mean/variance; K: measurement gain.

1. $x \sim \mathcal{N}(\mu_0, P_0)$ with $\mu_0 = 0$, $P_0 = 4$.
2. $z = x + \varepsilon = 3$; independent noise $\varepsilon \sim \mathcal{N}(0, R)$.
3. $K = P_0/(P_0 + R)$; $\mu_1 = \mu_0 + K(z - \mu_0)$; $P_1 = (1 - K)P_0$.

Gaussian conditioning — [2] UBC Lec.18 p.7 · Course question

Gaussian conditioning

What does a noisier reading change?

Solution 2 / 3

Compare the two noise variances

μ_0, P_0: prior mean/variance; μ_1, P_1: posterior mean/variance; K: measurement gain.

1. Old $R = 1$: $K = 4/5$, $\mu_1 = 0 + (4/5)3 = 2.4$, $P_1 = (1/5)4 = .8$.
2. New $R = 4$: $K = 4/8 = 1/2$, $\mu_1 = 0 + (1/2)3 = 1.5$.
3. New $P_1 = (1 - 1/2)4 = 2$; precision check: $P_1^{-1} = 1/4 + 1/4 = 1/2$.

Gaussian conditioning — [2] UBC Lec.18 p.7 · Course question

Gaussian conditioning

What does a noisier reading change?

Solution 3 / 3

Choose A

μ_0, P_0: prior mean/variance; μ_1, P_1: posterior mean/variance; K: measurement gain.

1. A: the mean is closer to 0 and posterior variance rises from .8 to 2.
2. B reverses both changes: a noisier reading carries less information.
3. C ignores R in the posterior variance. The sensor value z did not change.

EXERCISE · SLIDE 34 · CHECK THE RESULT

Gaussian conditioning

What does a noisier reading change?

Prior $\mathcal{N}(0, 4)$; $z = 3$ with independent Gaussian noise, variance R: $1 \to 4$. Compare posterior mean and variance.

- **A** Mean moves toward 0; posterior variance increases. ✓
- **B** Mean moves toward 3; posterior variance decreases.
- **C** Only the mean changes; variance ignores sensor noise.

$K = 4/(4+4) = 1/2$; $\mu_1 = 1.5$, $P_1 = 2$ (before: $2.4, .8$).

Gaussian conditioning — [2] UBC Lec.18 p.7 · Course question

EXERCISE · SLIDE 37 · QUESTION

Gaussian conditioning

Does the conditional variance change?

Gaussian $(a, b)^T$: $\mu = (1, 2)^T$, $\Sigma = \begin{bmatrix} 4 & 2 \\ 2 & 9 \end{bmatrix}$. Observe $b = -1$ instead of 5.

A. Both conditional mean and variance decrease.

B. Mean decreases; conditional variance stays fixed.

C. Neither changes because the joint model is fixed.

Conditional Gaussian distribution — [2] UBC Lec.18 p.7 · Course question

Gaussian conditioning

Does the conditional variance change?

Solution 1 / 4

Read the Gaussian blocks

a, b: scalar components; μ: means; Σ: covariance blocks; $a \mid b$: a conditioned on b.

1. a: unobserved scalar; b: observed scalar.
2. $\mu_a = 1$, $\mu_b = 2$; $\Sigma_{aa} = 4$, $\Sigma_{ab} = \Sigma_{ba} = 2$, $\Sigma_{bb} = 9$.
3. $\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(b - \mu_b)$.

Conditional Gaussian distribution — [2] UBC Lec.18 p.7 · Course question

Gaussian conditioning

Does the conditional variance change?

Solution 2 / 4

Update the center

a, b: scalar components; μ: means; Σ: covariance blocks; $a \mid b$: a conditioned on b.

1. Old $b = 5$: $\mu_{a|b} = 1 + (2/9)(5 - 2) = 1 + 2/3 = 5/3$.
2. New $b = -1$: $\mu_{a|b} = 1 + (2/9)(-1 - 2) = 1 - 2/3 = 1/3$.
3. The innovation $b - \mu_b$ changed; the mean decreases.

Conditional Gaussian distribution — [2] UBC Lec.18 p.7 · Course question

Gaussian conditioning

Does the conditional variance change?

Solution 3 / 4

Compute the unchanged variance

a, b: scalar components; μ: means; Σ: covariance blocks; $a \mid b$: a conditioned on b.

1. $\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}.$
2. $= 4 - 2(1/9)2 = 36/9 - 4/9 = 32/9.$
3. This expression contains covariance blocks, not the observed value b.

Conditional Gaussian distribution — [2] UBC Lec.18 p.7 · Course question

EXERCISE · SLIDE 37 · SOLUTION 4/4

Gaussian conditioning

Does the conditional variance change?

Solution 4 / 4

Choose B

a, b: scalar components; μ: means; Σ: covariance blocks; $a \mid b$: a conditioned on b.

---

1. A wrongly changes the width when only the observed value changes.
2. B changes the mean and keeps the conditional variance fixed.
3. C ignores the changed innovation. Constant width here uses joint Gaussianity.

Conditional Gaussian distribution — [2] UBC Lec.18 p.7 · Course question

EXERCISE · SLIDE 37 · CHECK THE RESULT

Gaussian conditioning

Does the conditional variance change?

Gaussian $(a, b)^T$: $\mu = (1, 2)^T$, $\Sigma = \begin{bmatrix} 4 & 2 \\ 2 & 9 \end{bmatrix}$. Observe $b = -1$ instead of 5.

A. Both conditional mean and variance decrease.

B. Mean decreases; conditional variance stays fixed. ✓

C. Neither changes because the joint model is fixed.

$\mu_{a|b} = 1 + \frac{2}{9}(-1 - 2) = \frac{1}{3}$; $\Sigma_{a|b} = 4 - \frac{4}{9} = \frac{32}{9}$.

Conditional Gaussian distribution — [2] UBC Lec.18 p.7 · Course question

EXERCISE · SLIDE 44 · QUESTION

Estimation

What makes the MAP estimate unique?

$A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$, $b = (1, 2)^T$; $J = \frac{1}{2}\|b - A\theta\|^2 + \frac{1}{2}\|\theta\|^2$.

A. The prior makes the two data columns independent.

B. Gaussian noise makes every data fit unique.

C. The prior adds curvature along the data null space.

MAP estimate; positive definiteness — [1] DL §5.6.1 · Course question

Estimation

What makes the MAP estimate unique?

Solution 1 / 5

Identify the data null space

θ: parameter column; A: data matrix; b: data column; $r = b - A\theta$; I: 2×2 identity.

1. $\theta = (\theta_1, \theta_2)^T$: parameters; $b = (1, 2)^T$: readings.
2. $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$; $r = b - A\theta$: residual.
3. $A\theta = (\theta_1 + \theta_2)(1, 2)^T$; zero residual requires $\theta_1 + \theta_2 = 1$.

MAP estimate; positive definiteness — [1] DL §5.6.1 · Course question

Estimation

What makes the MAP estimate unique?

Solution 2 / 5

Add the stated prior penalty

θ: parameter column; A: data matrix; b: data column; $r = b - A\theta$; I: 2×2 identity.

1. J: scaled negative-log posterior; $J = \frac{1}{2}\|r\|^2 + \frac{1}{2}\|\theta\|^2$, with $\lambda^2 = \sigma^2/\tau^2 = 1$.
2. $\nabla J = A^T(A\theta - b) + \theta = 0$.
3. $A^T A = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}$, $A^T b = (5, 5)^T$; hence $(A^T A + I)\theta = A^T b$.

MAP estimate; positive definiteness — [1] DL §5.6.1 · Course question

Estimation

What makes the MAP estimate unique?

Solution 3 / 5

Solve the two equations

θ: parameter column; A: data matrix; b: data column; $r = b - A\theta$; I: 2×2 identity.

1. $6\theta_1 + 5\theta_2 = 5$; $5\theta_1 + 6\theta_2 = 5$.
2. Subtract: $\theta_1 - \theta_2 = 0$; substitute: $11\theta_1 = 5$.
3. $\hat{\theta}_{MAP} = (5/11, 5/11)^T$; $r = (1/11, 2/11)^T$.

MAP estimate; positive definiteness — [1] DL §5.6.1 · Course question

Estimation

What makes the MAP estimate unique?

Solution 4 / 5

Why the minimizer is unique

θ: parameter column; A: data matrix; b: data column; $r = b - A\theta$; I: 2×2 identity.

1. For any nonzero column v, $v^T(A^TA + I)v = \|Av\|^2 + \|v\|^2 > 0$.
2. For the data-null direction $v = (1, -1)^T$, $Av = 0$ but $\|v\|^2 = 2$.
3. Thus the total objective has positive curvature even where the data loss is flat.

MAP estimate; positive definiteness — [1] DL §5.6.1 · Course question

Estimation

What makes the MAP estimate unique?

Solution 5 / 5

Choose C

θ: parameter column; A: data matrix; b: data column; $r = b - A\theta$; I: 2×2 identity.

1. A is false: A still has identical columns and rank 1.
2. B is false: Gaussian noise alone does not identify their separate coefficients.
3. C is correct: the stated Gaussian prior supplies curvature in the null direction.

MAP estimate; positive definiteness — [1] DL §5.6.1 · Course question

EXERCISE · SLIDE 44 · CHECK THE RESULT

Estimation

What makes the MAP estimate unique?

$A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$, $b = (1,2)^T$; $J = \frac{1}{2}\|b - A\theta\|^2 + \frac{1}{2}\|\theta\|^2$.

A. The prior makes the two data columns independent.

B. Gaussian noise makes every data fit unique.

C. The prior adds curvature along the data null space. ✓

> Least squares: $\theta_1 + \theta_2 = 1$. MAP: $(A^TA + I)\theta = A^Tb$ gives $\theta_1 = \theta_2 = 5/11$.

MAP estimate; positive definiteness — [1] DL §5.6.1 · Course question

EXERCISE · SLIDE 45 · QUESTION

Integrated board

Calibration: choose the likelihood

$b = (1, 2, 3)^T$, $b_i = \theta + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, 1)$. Which data objective fits this model?

A $\frac{1}{2}\sum_i (b_i - \theta)^2$

B $\sum_i |b_i - \theta|$

C $\frac{1}{2}\sum_i (b_i - \theta)^2 + \frac{1}{2}\theta^2$

Gaussian likelihood; maximum likelihood — [1] DL §5.5.1 · Course example

EXERCISE · SLIDE 45 · SOLUTION 1/4

Integrated board

Calibration: choose the likelihood

Solution 1 / 4

Define the calibration model

b_i: reading i; θ: offset; ε_i: noise; J_{data}: data objective; C: constant in θ.

1. $b = (1, 2, 3)^T$: observed readings; θ: unknown common offset.
2. $b_i = \theta + \varepsilon_i$; independent $\varepsilon_i \sim \mathcal{N}(0, 1)$.
3. $p(b_i \mid \theta) = (2\pi)^{-1/2} \exp[-(b_i - \theta)^2/2]$.

Gaussian likelihood; maximum likelihood — [1] DL §5.5.1 · Course example

Integrated board

Calibration: choose the likelihood

Solution 2 / 4

From likelihood to objective

b_i: reading i; θ: offset; ε_i: noise; J_{data}: data objective; C: constant in θ.

1. Independence: $p(b \mid \theta) = \prod_{i=1}^{3} p(b_i \mid \theta)$.
2. $-\log p(b \mid \theta) = C + \frac{1}{2}\sum_{i=1}^{3}(b_i - \theta)^2$.
3. C is independent of θ; minimizing leaves $J_{data} = \frac{1}{2}\sum_i (b_i - \theta)^2$.

Gaussian likelihood; maximum likelihood — [1] DL §5.5.1 · Course example

Integrated board

Calibration: choose the likelihood

Solution 3 / 4

Expand every square

b_i: reading i; θ: offset; ε_i: noise; J_{data}: data objective; C: constant in θ.

1. $2J_{data} = (1-\theta)^2 + (2-\theta)^2 + (3-\theta)^2.$
2. $= (1 - 2\theta + \theta^2) + (4 - 4\theta + \theta^2) + (9 - 6\theta + \theta^2).$
3. $= 3\theta^2 - 12\theta + 14.$

Gaussian likelihood; maximum likelihood — [1] DL §5.5.1 · Course example

EXERCISE · SLIDE 45 · SOLUTION 4/4

Integrated board

Calibration: choose the likelihood

Solution 4 / 4

Minimize and choose A

b_i: reading i; θ: offset; ε_i: noise; J_{data}: data objective; C: constant in θ.

---

1. $J'_{data} = 3\theta - 6 = 0 \Rightarrow \hat{\theta}_{MLE} = 2;\ J''_{data} = 3 > 0.$
2. A follows from independent Gaussian noise.
3. B uses absolute residuals, a different noise model. C already adds a prior penalty.

Gaussian likelihood; maximum likelihood — [1] DL §5.5.1 · Course example

EXERCISE · SLIDE 45 · CHECK THE RESULT

Integrated board

Calibration: choose the likelihood

$b = (1, 2, 3)^T$, $b_i = \theta + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, 1)$. Which data objective fits this model?

- **A** $\frac{1}{2}\sum_i (b_i - \theta)^2$ ✓
- **B** $\sum_i |b_i - \theta|$
- **C** $\frac{1}{2}\sum_i (b_i - \theta)^2 + \frac{1}{2}\theta^2$

> $-\log p(b \mid \theta) = C + \frac{1}{2}\sum_i (b_i - \theta)^2$; $\hat{\theta}_{MLE} = \bar{b} = 2$.

Gaussian likelihood; maximum likelihood — [1] DL §5.5.1 · Course example

EXERCISE · SLIDE 46 · QUESTION

Integrated board

Calibration: combine data and prior

Readings $(1, 2, 3)$; independent normal noise, variance 1; prior $\theta \sim \mathcal{N}(0, 1)$. Find posterior mean and precision.

A. Center stays at 2; prior changes only uncertainty.

B. Center moves toward 0; precision becomes 4.

C. Center moves toward 0; variances add to 4.

Gaussian posterior; MAP estimate — [2] UBC Lec.18 p.7 · [1] DL §5.6.1 · Course example

EXERCISE · SLIDE 46 · SOLUTION 1/4

Integrated board

Calibration: combine data and prior

Solution 1 / 4

Keep the data and state the prior

θ: offset; J_{post}: posterior objective; μ_1, P_1: posterior mean/variance; precision $= 1/P_1$.

1. $b = (1, 2, 3)^T$; independent noise variance 1; θ: offset.
2. $J_{data} = \frac{1}{2}(3\theta^2 - 12\theta + 14)$; $\theta \sim \mathcal{N}(0, 1)$.
3. Negative log prior adds $\theta^2/2$, up to a constant.

Gaussian posterior; MAP estimate — [2] UBC Lec.18 p.7 · [1] DL §5.6.1 · Course example

Integrated board

Calibration: combine data and prior

Solution 2 / 4

Collect the posterior quadratic

θ: offset; J_{post}: posterior objective; μ_1, P_1: posterior mean/variance; precision $= 1/P_1$.

1. $J_{post} = \frac{1}{2}(3\theta^2 - 12\theta + 14) + \frac{1}{2}\theta^2.$
2. $= 2\theta^2 - 6\theta + 7 = 2(\theta^2 - 3\theta) + 7.$
3. $\theta^2 - 3\theta = (\theta - 3/2)^2 - 9/4.$

Gaussian posterior; MAP estimate — [2] UBC Lec.18 p.7 · [1] DL §5.6.1 · Course example

EXERCISE · SLIDE 46 · SOLUTION 3/4

Integrated board

Calibration: combine data and prior

Solution 3 / 4

Complete the square and read variance

θ: offset; J_{post}: posterior objective; μ_1, P_1: posterior mean/variance; precision $= 1/P_1$.

1. $J_{post} = 2(\theta - 3/2)^2 - 9/2 + 7 = 2(\theta - 1.5)^2 + 2.5.$
2. Gaussian quadratic: $(\theta - \mu_1)^2/(2P_1)$; hence $\mu_1 = 1.5$, $1/(2P_1) = 2$.
3. $P_1 = 1/4$; precision $P_1^{-1} = 4$; $\theta \mid b \sim \mathcal{N}(1.5, 1/4)$.

Gaussian posterior; MAP estimate — [2] UBC Lec.18 p.7 · [1] DL §5.6.1 · Course example

EXERCISE · SLIDE 46 · SOLUTION 4/4

Integrated board

Calibration: combine data and prior

Solution 4 / 4

Choose B and check the weights

θ: offset; J_{post}: posterior objective; μ_1, P_1: posterior mean/variance; precision $= 1/P_1$.

1. Data precision 3 plus prior precision 1 gives 4.
2. Posterior mean $= (3 \cdot 2 + 1 \cdot 0)/4 = 1.5$; MAP equals this Gaussian mean.
3. B is correct. A ignores the prior shift. C adds variances instead of precisions.

Gaussian posterior; MAP estimate — [2] UBC Lec.18 p.7 · [1] DL §5.6.1 · Course example

EXERCISE · SLIDE 46 · CHECK THE RESULT

Integrated board

Calibration: combine data and prior

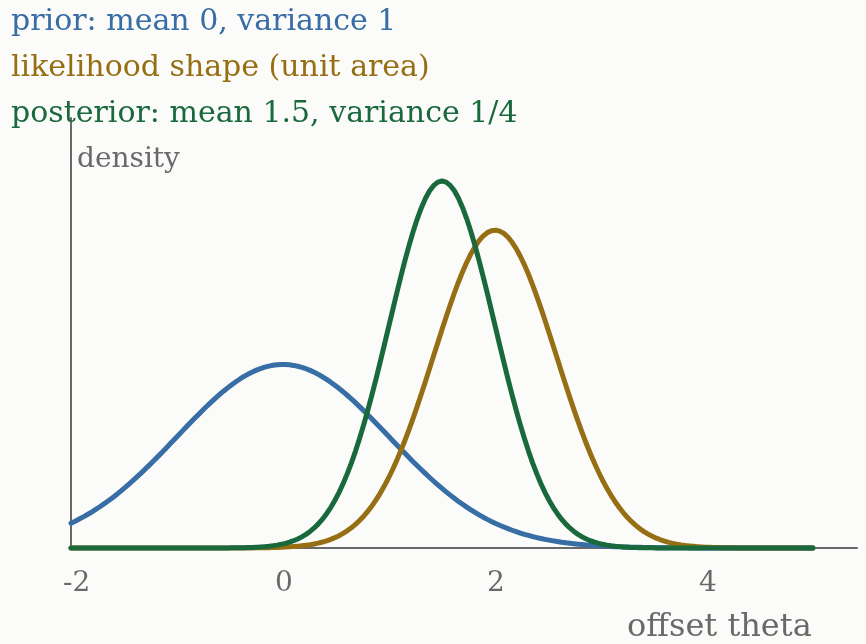
Readings $(1, 2, 3)$; independent normal noise, variance 1; prior $\theta \sim \mathcal{N}(0, 1)$. Find posterior mean and precision.

A. Center stays at 2; prior changes only uncertainty.

B. Center moves toward 0; precision becomes 4. ✓

C. Center moves toward 0; variances add to 4.

$J_{post} = 2(\theta - 1.5)^2 + 2.5$; $\theta \mid b \sim \mathcal{N}(1.5, 1/4)$.



Gaussian posterior; MAP estimate — [2] UBC Lec.18 p.7 · [1] DL §5.6.1 · Course example

EXERCISE · SLIDE 47 · QUESTION

Intervals

Calibration: confidence coverage

Three iid $B_i \mid \theta \sim \mathcal{N}(\theta, 1)$; $\bar{B}$: mean. $C(B) = \bar{B} \pm 1.96/\sqrt{3}$. Interpret 95% confidence.

A. This realized interval has 95% posterior mass.

B. 95% of future individual readings lie inside it.

C. Repeated intervals cover fixed θ 95% of the time.

Confidence interval; coverage probability — [3] NIST §7.1.4 · Course example

Intervals

Calibration: confidence coverage

Solution 1 / 4

Define the random interval

B: random data; $\bar{B}$: sample mean; θ: fixed offset; $C(B)$: confidence interval; Z: standardized mean.

1. θ: fixed true offset; B_i: random reading before observation.
2. $B_i \mid \theta \sim \mathcal{N}(\theta, 1)$ independently; $\bar{B} = (B_1 + B_2 + B_3)/3$.
3. $C(B)$: confidence interval computed anew for each random dataset.

Confidence interval; coverage probability — [3] NIST §7.1.4 · Course example

Intervals

Calibration: confidence coverage

Solution 2 / 4

Standardize the sample mean

B: random data; $\bar{B}$: sample mean; θ: fixed offset; $C(B)$: confidence interval; Z: standardized mean.

1. $\mathbb{E}_\theta[\bar{B}] = \theta$; $\mathrm{Var}_\theta(\bar{B}) = (1+1+1)/9 = 1/3$.
2. $Z = (\bar{B} - \theta)/(1/\sqrt{3}) \sim \mathcal{N}(0,1)$.
3. $P_\theta(-1.96 \le Z \le 1.96) \approx .95$.

Confidence interval; coverage probability — [3] NIST §7.1.4 · Course example

Intervals

Calibration: confidence coverage

Solution 3 / 4

Rearrange to isolate the fixed offset

B: random data; $\bar{B}$: sample mean; θ: fixed offset; $C(B)$: confidence interval; Z: standardized mean.

1. $-1.96/\sqrt{3} \leq \bar{B} - \theta \leq 1.96/\sqrt{3}$.
2. $\bar{B} - 1.96/\sqrt{3} \leq \theta \leq \bar{B} + 1.96/\sqrt{3}$.
3. Thus $C(B) = \bar{B} \pm 1.96/\sqrt{3}$ and $P_\theta(\theta \in C(B)) \approx .95$.

Confidence interval; coverage probability — [3] NIST §7.1.4 · Course example

EXERCISE · SLIDE 47 · SOLUTION 4/4

Intervals

Calibration: confidence coverage

Solution 4 / 4

Choose C

B: random data; $\bar{B}$: sample mean; θ: fixed offset; $C(B)$: confidence interval; Z: standardized mean.

1. C: repeated random intervals cover the same fixed offset about 95% of the time.
2. A needs a posterior distribution; confidence alone provides none.
3. B concerns individual future readings, whose noise variance is 1, not 1/3.

Confidence interval; coverage probability — [3] NIST §7.1.4 · Course example

EXERCISE · SLIDE 47 · CHECK THE RESULT

Intervals

Calibration: confidence coverage

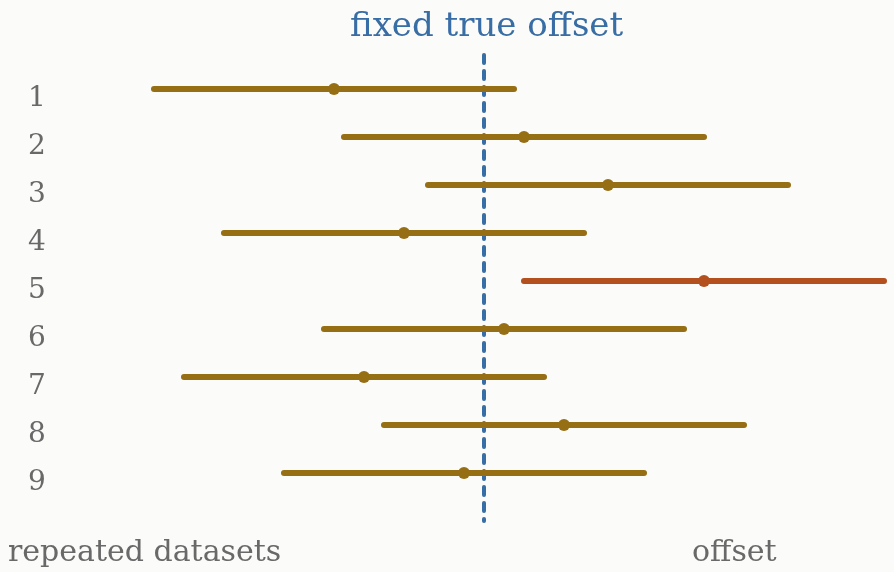
Three iid $B_i \mid \theta \sim \mathcal{N}(\theta, 1)$; $\bar{B}$: mean. $C(B) = \bar{B} \pm 1.96/\sqrt{3}$. Interpret 95% confidence.

A. This realized interval has 95% posterior mass.

B. 95% of future individual readings lie inside it.

C. Repeated intervals cover fixed θ 95% of the time. ✓

$\sqrt{3}(\bar{B} - \theta) \sim \mathcal{N}(0, 1)$; $P_\theta(\theta \in C(B)) \approx .95$.



Illustration, not measured coverage

Confidence interval; coverage probability — [3] NIST §7.1.4 · Course example

EXERCISE · SLIDE 48 · QUESTION

Intervals

Calibration: what does credibility describe?

Observed $b = (1, 2, 3)^T$; $\theta \mid b \sim \mathcal{N}(1.5, .25)$. A 95% credible interval describes which probability?

A. 95% posterior mass for θ, with b fixed.

B. 95% coverage for every fixed true offset.

C. 95% probability for a future noisy reading.

Posterior credible interval — [4] MIT 18.05 §2 · Course example

Intervals

Calibration: what does credibility describe?

Solution 1 / 3

Fix the data and use the posterior

b: fixed data; θ: offset; μ_1, P_1: posterior mean/variance; $I(b)$: credible interval; Z: standardized offset.

1. $b = (1, 2, 3)^T$: observed data, now fixed; θ: offset.
2. $\theta \mid b \sim \mathcal{N}(\mu_1, P_1)$ with $\mu_1 = 1.5$, $P_1 = .25$.
3. Posterior standard deviation $\sqrt{P_1} = \sqrt{.25} = .5$.

Posterior credible interval — [4] MIT 18.05 §2 · Course example

Intervals

Calibration: what does credibility describe?

Solution 2 / 3

Standardize, then undo the scaling

b: fixed data; θ: offset; μ_1, P_1: posterior mean/variance; $I(b)$: credible interval; Z: standardized offset.

1. $Z = (\theta - 1.5)/.5 \mid b \sim \mathcal{N}(0,1).$
2. $P(-1.96 \le Z \le 1.96 \mid b) \approx .95.$
3. $1.5 - 1.96(.5) \le \theta \le 1.5 + 1.96(.5)$ gives $I(b) = [.52, 2.48].$

Posterior credible interval — [4] MIT 18.05 §2 · Course example

Intervals

Calibration: what does credibility describe?

Solution 3 / 3

Choose A

b: fixed data; θ: offset; μ_1, P_1: posterior mean/variance; $I(b)$: credible interval; Z: standardized offset.

1. A: $P(\theta \in I(b) \mid b) \approx .95$ under this posterior.
2. B is a repeated-sampling claim for every fixed offset; it does not follow here.
3. C includes new measurement noise; this interval describes the offset only.

Posterior credible interval — [4] MIT 18.05 §2 · Course example

EXERCISE · SLIDE 48 · CHECK THE RESULT

Intervals

Calibration: what does credibility describe?

Observed $b = (1, 2, 3)^T$; $\theta \mid b \sim \mathcal{N}(1.5, .25)$. A 95% credible interval describes which probability?

- **A** 95% posterior mass for θ, with b fixed. ✓
- **B** 95% coverage for every fixed true offset.
- **C** 95% probability for a future noisy reading.

> $I(b) = 1.5 \pm 1.96(.5)$; $P(\theta \in I(b) \mid b) \approx .95$.

Posterior credible interval — [4] MIT 18.05 §2 · Course example

EXERCISE · SLIDE 49 · QUESTION

Intervals

Calibration: why do the intervals differ?

Prior $\theta \sim \mathcal{N}(0,1)$. Confidence $C(b) = 2 \pm 1.96/\sqrt{3}$; credible $I(b) = 1.5 \pm 1.96(.5)$. Why different?

A. Narrower credible interval means better coverage.

B. The zero-centered prior adds information.

C. The physical sensor became less noisy.

Confidence interval; credible interval — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6.1 · Course comparison

Intervals

Calibration: why do the intervals differ?

Solution 1 / 3

Identify the two information sources

C: confidence interval; I: credible interval; $\bar{b}$: sample mean; precision: inverse variance.

1. $b = (1, 2, 3)^T$; $\bar{b} = 2$; noise variance 1; prior $\mathcal{N}(0, 1)$.
2. C: confidence interval; sampling standard error $1/\sqrt{3}$.
3. I: credible interval; posterior mean $(3 \cdot 2 + 1 \cdot 0)/4 = 1.5$, variance $1/4$.
4. Data precision $= 1/(1/3) = 3$; prior precision $= 1/1 = 1$.

Confidence interval; credible interval — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6.1 · Course comparison

Intervals

Calibration: why do the intervals differ?

Solution 2 / 3

Compute and compare endpoints

C: confidence interval; I: credible interval; $\bar{b}$: sample mean; precision: inverse variance.

1. $C = 2 \pm 1.96/\sqrt{3} \approx 2 \pm 1.1316 = [.8684, 3.1316]$.
2. $I = 1.5 \pm 1.96\sqrt{1/4} = 1.5 \pm .98 = [.52, 2.48]$.
3. $1.5 < 2$ and $.98 < 1.1316$: the prior shifts the center and reduces posterior width.

Confidence interval; credible interval — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6.1 · Course comparison

Intervals

Calibration: why do the intervals differ?

Solution 3 / 3

Choose B

C: confidence interval; I: credible interval; $\bar{b}$: sample mean; precision: inverse variance.

---

1. B: one unit of prior precision adds to three units from the data.
2. A: narrower does not prove better repeated-sampling coverage.
3. C: the physical noise variance remains 1; only the information used changed.

Confidence interval; credible interval — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6.1 · Course comparison

EXERCISE · SLIDE 49 · CHECK THE RESULT

Intervals

Calibration: why do the intervals differ?

Prior $\theta \sim \mathcal{N}(0,1)$. Confidence $C(b) = 2 \pm 1.96/\sqrt{3}$; credible $I(b) = 1.5 \pm 1.96(.5)$. Why different?

A. Narrower credible interval means better coverage.

B. The zero-centered prior adds information. ✓

C. The physical sensor became less noisy.

$C \approx [.8684, 3.1316]$; $I \approx [.5200, 2.4800]$. Prior precision 1 adds to data precision 3.

Confidence interval; credible interval — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6.1 · Course comparison

EXERCISE · SLIDE 50 · QUESTION

Transfer

Calibration: new data, new prior

Readings $b = (2, 3)^T$ estimate offset θ; iid noise $\mathcal{N}(0, 1)$; prior $\theta \sim \mathcal{N}(4, 1)$. Which is valid?

A. The prior changes the data-only MLE.

B. The prior center can be ignored in MAP.

C. MAP moves above the data mean; MLE stays fixed.

MLE; MAP; confidence and credible intervals — [1] DL §§5.5–5.6 · [3] NIST §7.1.4 · [4] MIT 18.05 §2 · Course question

Transfer

Calibration: new data, new prior

Solution 1 / 5

Set up the new data-only problem

θ: offset; J: objective; μ_1, P_1: posterior mean/variance; C: confidence; I: credible interval.

---

1. $b = (2,3)^T$; θ: offset; independent noise variance 1.
2. $J_{data} = \frac{1}{2}[(2-\theta)^2 + (3-\theta)^2]$.
3. $= \frac{1}{2}[(4 - 4\theta + \theta^2) + (9 - 6\theta + \theta^2)] = \frac{1}{2}(2\theta^2 - 10\theta + 13)$.

MLE; MAP; confidence and credible intervals — [1] DL §§5.5–5.6 · [3] NIST §7.1.4 · [4] MIT 18.05 §2 · Course question

Transfer

Calibration: new data, new prior

Solution 2 / 5

Minimize, then add the new prior

θ: offset; J: objective; μ_1, P_1: posterior mean/variance; C: confidence; I: credible interval.

1. $J'_{data} = 2\theta - 5 = 0 \Rightarrow \hat{\theta}_{MLE} = 2.5$; $J''_{data} = 2 > 0$.
2. Prior $\mathcal{N}(4, 1)$ adds $\frac{1}{2}(\theta - 4)^2 = \frac{1}{2}(\theta^2 - 8\theta + 16)$.
3. $J_{post} = \frac{1}{2}(3\theta^2 - 18\theta + 29)$.

MLE; MAP; confidence and credible intervals — [1] DL §§5.5–5.6 · [3] NIST §7.1.4 · [4] MIT 18.05 §2 · Course question

Transfer

Calibration: new data, new prior

Solution 3 / 5

Complete the posterior square

θ: offset; J: objective; μ_1, P_1: posterior mean/variance; C: confidence; I: credible interval.

---

1. $3\theta^2 - 18\theta + 29 = 3(\theta^2 - 6\theta) + 29.$
2. $= 3[(\theta - 3)^2 - 9] + 29 = 3(\theta - 3)^2 + 2.$
3. $J_{post} = \frac{3}{2}(\theta - 3)^2 + 1;\ \mu_1 = 3,\ 1/(2P_1) = 3/2 \Rightarrow P_1 = 1/3.$

MLE; MAP; confidence and credible intervals — [1] DL §§5.5–5.6 · [3] NIST §7.1.4 · [4] MIT 18.05 §2 · Course question

Transfer

Calibration: new data, new prior

Solution 4 / 5

Compute both interval constructions

θ: offset; J: objective; μ_1, P_1: posterior mean/variance; C: confidence; I: credible interval.

---

1. C: confidence; mean 2.5, sampling standard error $1/\sqrt{2}$.
2. $C = 2.5 \pm 1.96/\sqrt{2} \approx [1.1141, 3.8859]$.
3. I: credible; posterior mean 3, standard deviation $1/\sqrt{3}$; $I = 3 \pm 1.96/\sqrt{3} \approx [1.8684, 4.1316]$.

MLE; MAP; confidence and credible intervals — [1] DL §§5.5–5.6 · [3] NIST §7.1.4 · [4] MIT 18.05 §2 · Course question

Transfer

Calibration: new data, new prior

Solution 5 / 5

Choose C

θ: offset; J: objective; μ_1, P_1: posterior mean/variance; C: confidence; I: credible interval.

1. C: $\hat{\theta}_{MAP} = 3 > 2.5$, while the data-only MLE remains 2.5.
2. A incorrectly puts prior information into the data-only MLE.
3. B drops the prior center 4 and its linear term. Confidence and credibility retain different meanings.

MLE; MAP; confidence and credible intervals — [1] DL §§5.5–5.6 · [3] NIST §7.1.4 · [4] MIT 18.05 §2 · Course question

EXERCISE · SLIDE 50 · CHECK THE RESULT

Transfer

Calibration: new data, new prior

Readings $b = (2, 3)^T$ estimate offset θ; iid noise $\mathcal{N}(0, 1)$; prior $\theta \sim \mathcal{N}(4, 1)$. Which is valid?

A The prior changes the data-only MLE.

B The prior center can be ignored in MAP.

C MAP moves above the data mean; MLE stays fixed. ✓

> $\hat{\theta}_{MLE} = 2.5$, $\theta \mid b \sim \mathcal{N}(3, 1/3)$. $C = 2.5 \pm 1.96/\sqrt{2}$; $I = 3 \pm 1.96/\sqrt{3}$.

MLE; MAP; confidence and credible intervals — [1] DL §§5.5–5.6 · [3] NIST §7.1.4 · [4] MIT 18.05 §2 · Course question

EXERCISE · SLIDE 51 · QUESTION

Intervals

What does the sensor promise?

A specification says “±2 cm, 95%.” Which conclusion is justified without another definition?

A. It is a confidence interval for the offset.

B. It is a credible interval for the offset.

C. None: ask what quantity and probability model it describes.

Confidence; credibility; predictive uncertainty — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6, Eq.5.68 · Course question

EXERCISE · SLIDE 51 · SOLUTION 1/3

Intervals

What does the sensor promise?

Solution 1 / 3

Identify what the specification omits

$C(B)$: confidence interval; $I(b)$: credible interval; ε_{new}: future-reading error.

1. The statement “± 2 cm, 95%” supplies a width and a percentage.
2. θ: offset; B: repeated random dataset; b: observed dataset.
3. No target quantity or probability construction is specified.

Confidence; credibility; predictive uncertainty — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6, Eq.5.68 · Course question

EXERCISE · SLIDE 51 · SOLUTION 2/3

Intervals

What does the sensor promise?

Solution 2 / 3

Compare three possible meanings

$C(B)$: confidence interval; $I(b)$: credible interval; ε_{new}: future-reading error.

1. Confidence: $P_\theta(\theta \in C(B)) \approx .95$ for a stated repeated-data procedure.
2. Credibility: $P(\theta \in I(b) \mid b) \approx .95$ under a stated posterior.
3. Future-error bound: $P(|\varepsilon_{new}| \leq 2\,\text{cm}) \approx .95$ under a stated error model.

Confidence; credibility; predictive uncertainty — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6, Eq.5.68 · Course question

EXERCISE · SLIDE 51 · SOLUTION 3/3

Intervals

What does the sensor promise?

Solution 3 / 3

Choose C

$C(B)$: confidence interval; $I(b)$: credible interval; ε_{new}: future-reading error.

1. A assumes an unstated repeated-sampling procedure.
2. B assumes an unstated prior and posterior.
3. C: ask which quantity is bounded and which model or procedure supplies 95%.

Confidence; credibility; predictive uncertainty — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6, Eq.5.68 · Course question

EXERCISE · SLIDE 51 · CHECK THE RESULT

Intervals

What does the sensor promise?

A specification says “±2 cm, 95%.” Which conclusion is justified without another definition?

A. It is a confidence interval for the offset.

B. It is a credible interval for the offset.

C. None: ask what quantity and probability model it describes. ✓

> Offset uncertainty, future-reading error and repeated coverage are different claims.

Confidence; credibility; predictive uncertainty — [3] NIST §7.1.4 · [4] MIT 18.05 §2 · [1] DL §5.6, Eq.5.68 · Course question

EXERCISE · SLIDE 52 · QUESTION

Synthesis

Which modeling bridge is valid?

Zero-mean iid Gaussian noise; isotropic Gaussian prior. Which chain gives the negative-log posterior?

A. Independence gives squares; Gaussian noise gives sums.

B. Gaussian noise gives squares; independence gives sums.

C. A Gaussian prior removes the data residual term.

Gaussian likelihood; independence; MAP regularization — [1] DL §§5.5–5.6 · [1] DL §3.7 · Course question

Synthesis

Which modeling bridge is valid?

Solution 1 / 4

State the model and symbols

i: reading index; I: identity matrix; C: constant in θ; J: negative-log posterior, up to constants.

1. $b = A\theta + \varepsilon$: data column b, design matrix A, parameter column θ.
2. ε_i are independent $\mathcal{N}(0, \sigma^2)$ noises; $r = b - A\theta$: residual.
3. Prior $\theta \sim \mathcal{N}(\mu_0, \tau^2 I)$; μ_0: prior mean; $\sigma^2, \tau^2 > 0$: variances.

Gaussian likelihood; independence; MAP regularization — [1] DL §§5.5–5.6 · [1] DL §3.7 · Course question

Synthesis

Which modeling bridge is valid?

Solution 2 / 4

Separate the square from the sum

i: reading index; I: identity matrix; C: constant in θ; J: negative-log posterior, up to constants.

1. One Gaussian factor: $p(b_i \mid \theta) = (2\pi\sigma^2)^{-1/2} e^{-r_i^2/(2\sigma^2)}$.
2. Independence gives $p(b \mid \theta) = \prod_i p(b_i \mid \theta)$.
3. $-\log p(b \mid \theta) = C + \sum_i r_i^2/(2\sigma^2) = C + \|b - A\theta\|^2/(2\sigma^2)$.

Gaussian likelihood; independence; MAP regularization — [1] DL §§5.5–5.6 · [1] DL §3.7 · Course question

Synthesis

Which modeling bridge is valid?

Solution 3 / 4

Add the prior through Bayes rule

i: reading index; I: identity matrix; C: constant in θ; J: negative-log posterior, up to constants.

1. $p(\theta \mid b) \propto p(b \mid \theta)p(\theta)$; logarithms turn this product into a sum.
2. $J = \|b - A\theta\|^2/(2\sigma^2) + \|\theta - \mu_0\|^2/(2\tau^2)$.
3. Multiply all of J by σ^2: $\frac{1}{2}\|r\|^2 + \frac{\lambda^2}{2}\|\theta - \mu_0\|^2$, where $\lambda^2 = \sigma^2/\tau^2$.

Gaussian likelihood; independence; MAP regularization — [1] DL §§5.5–5.6 · [1] DL §3.7 · Course question

EXERCISE · SLIDE 52 · SOLUTION 4/4

Synthesis

Which modeling bridge is valid?

Solution 4 / 4

Choose B

i: reading index; I: identity matrix; C: constant in θ; J: negative-log posterior, up to constants.

1. B: the Gaussian exponent supplies squares; independence and logarithms supply the sum.
2. A reverses the two reasons.
3. C discards the likelihood. A prior adds a penalty while retaining the residual term.

Gaussian likelihood; independence; MAP regularization — [1] DL §§5.5–5.6 · [1] DL §3.7 · Course question

EXERCISE · SLIDE 52 · CHECK THE RESULT

Synthesis

Which modeling bridge is valid?

Zero-mean iid Gaussian noise; isotropic Gaussian prior. Which chain gives the negative-log posterior?

A. Independence gives squares; Gaussian noise gives sums.

B. Gaussian noise gives squares; independence gives sums. ✓

C. A Gaussian prior removes the data residual term.

$$-\log p(\theta \mid b) = C + \|b - A\theta\|^2/(2\sigma^2) + \|\theta - \mu_0\|^2/(2\tau^2).$$

Gaussian likelihood; independence; MAP regularization — [1] DL §§5.5–5.6 · [1] DL §3.7 · Course question