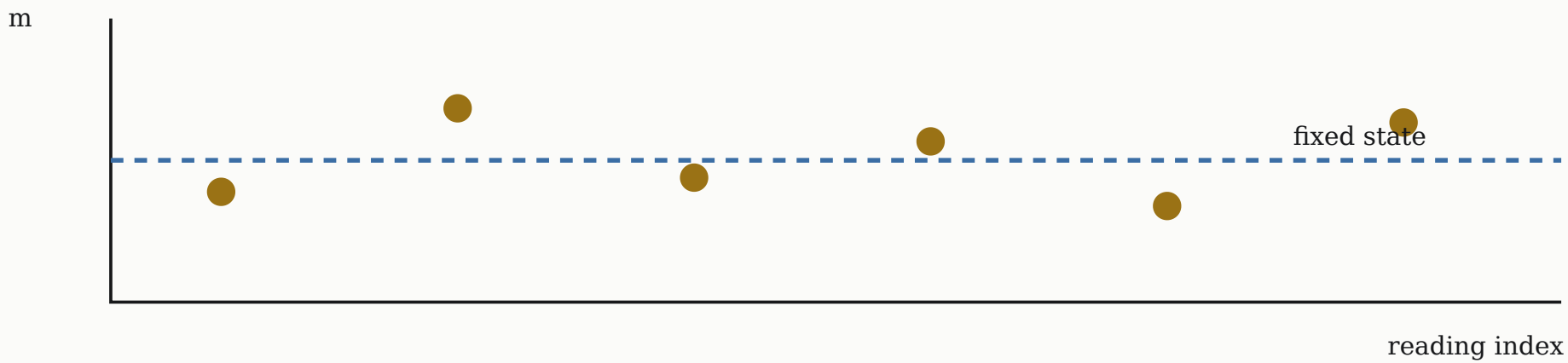
From Noisy Readings to Reliable Estimates

Which position should a noisy sensor report?



> One stationary object. Several different readings.

`state` · `measurements` · `inferred answer`

Report an estimate and the uncertainty around it.

LECTURE ROADMAP

From Readings to Estimates

How does the lecture fit together?

$$z = x + \varepsilon$$

- x unknown state (m)
- z observed reading (m)
- ε random error (m)

1. Model error	→	Assign probabilities	→	Describe spread
2. Use observations	→	Combine information	→	Update uncertainty
3. Estimate parameters	→	Compare methods	→	Point estimate
4. Report uncertainty	→	Compare intervals	→	Interpret 95%

> Estimate the fixed state from noisy readings.

Probability distribution; uncertainty — [1] DL §3.1 · Course concept map

Random variables

How does an outcome become a number?

$$x : \Omega \to \mathbb{R}, \qquad x(\omega) \in \{-1, 0, 2\}$$

- Ω sample space: all possible outcomes
- ω one outcome
- x random error here, not position (cm)

Outcome A $\to -1$ cm

Outcome B $\to 0$ cm

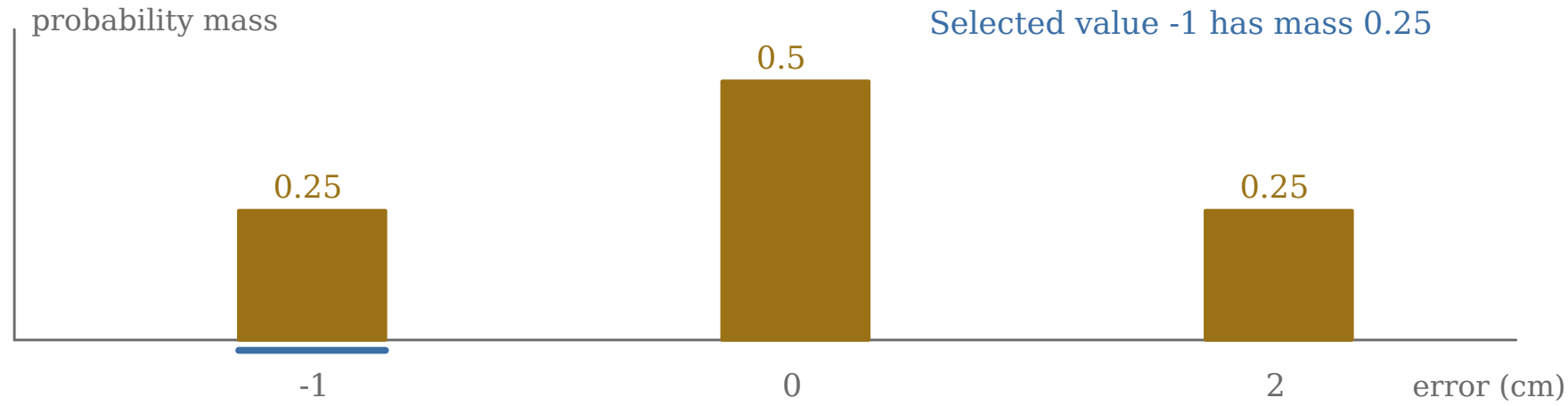
Outcome C $\to 2$ cm

> The rule is fixed; the outcome is uncertain.

Random variable; realization — [1] DL §3.2

Probability Mass Function (PMF)

How do probabilities attach to discrete values?



$$p(a) = P(x = a), \qquad p(a) \geq 0, \qquad \sum_{a} p(a) = 1$$

x random error (cm) a possible value (cm)

$p(a)$ probability of $x = a$ P event probability

Probability mass function (PMF) — [1] DL §3.3.1

Probability Mass Function (PMF)

What does the three-point error model predict?

Error values -1, 0, 2 cm; masses ¼, ½, ¼. Find nonnegative-error probability and mean.

$$P(x \geq 0) = \tfrac{1}{2} + \tfrac{1}{4} = \tfrac{3}{4}, \qquad \mu = (-1)\tfrac{1}{4} + 0\tfrac{1}{2} + 2\tfrac{1}{4} = \tfrac{1}{4}$$

- P event probability
- x error (cm)
- μ mean error (cm)

Discrete probability; expectation — [1] DL §3.3.1 · [1] DL §3.8 · Course example

CHOOSE, EXPLAIN, VOTE

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

Probability Density Function (PDF)

Why does a continuous event need an area?

$$p(t) \geq 0, \quad \int_{-\infty}^{\infty} p(t)\, dt = 1$$

$$P(a \leq x \leq b) = \int_{a}^{b} p(t)\, dt$$

- x random error (cm)
- a, b interval endpoints
- t integration variable
- $p(t)$ density (cm^{-1})

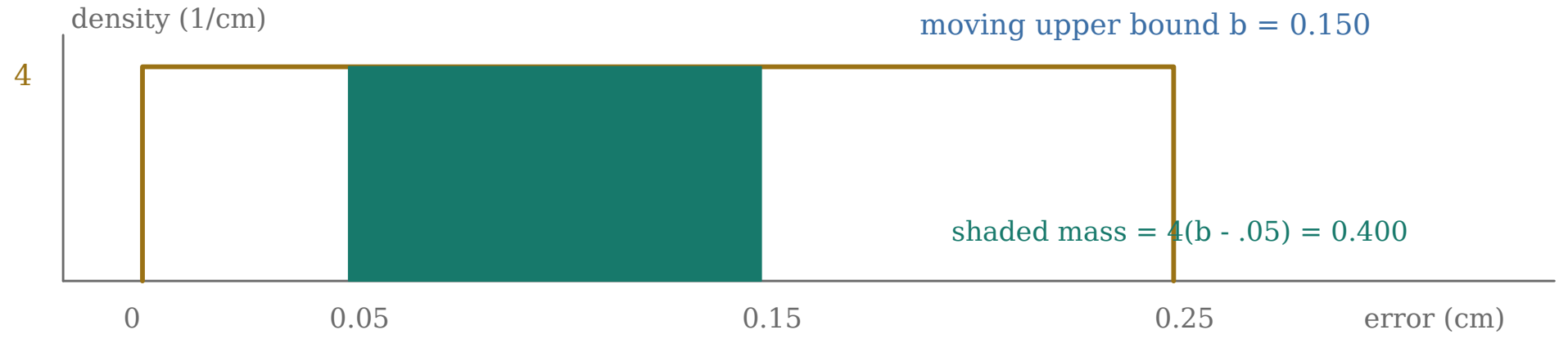
$$P(x = c) = \int_{c}^{c} p(t)\, dt = 0$$

- c fixed error value

Probability density function (PDF) — [1] DL §3.3.2

Probability Density Function (PDF)

Can a valid density be greater than one?



UNIFORM ERROR

$$p(t) = 4 \;\; (0 \le t \le .25), \quad P(.05 \le x \le .15) = 4(.15 - .05) = .4$$

- t, x error (cm)
- p density (cm^{-1})
- P interval probability

Uniform density; normalization — [1] DL §3.3.2 · Course example

CHOOSE, EXPLAIN, VOTE

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

Moments

How do we average a function of a random value?

$$\mathbb{E}[g(x)] = \sum_{a} g(a)p(a)$$

- x discrete random variable
- a possible value
- g function being averaged
- p PMF

$$\mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(t)p(t)\,dt$$

- x continuous random variable
- t integration variable
- p density
- g assume E[|g(x)|] is finite

Expectation of a function — [1] DL §3.8 · Course derivation

CHOOSE, EXPLAIN, VOTE

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

Moments

What is the mean of a random column vector?

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \mu = \mathbb{E}[x] = \begin{bmatrix} \mathbb{E}[x_1] \\ \mathbb{E}[x_2] \end{bmatrix}$$

- x random column in R^2
- x_i component i
- μ mean column in R^2

> Average corresponding components; preserve the column shape.

Samples use superscripts: $x^{(s)}$. Components use subscripts: x_i.

Random vector; mean vector — [1] DL §3.8

DEFINITION AND REASONING

Moments

Why does covariance use an outer product?

$$\Sigma = \mathbb{E}[(x-\mu)(x-\mu)^T], \qquad \Sigma_{ij} = \mathbb{E}[(x_i-\mu_i)(x_j-\mu_j)]$$

x, μ random column and mean in R^{d}

Σ population covariance, d × d

i, j coordinate indices

> Diagonal: component variance. Off-diagonal: paired deviations.

Covariance matrix — [1] DL §3.8 · Course derivation

Moments

Why is covariance positive semidefinite?

$$a^T \Sigma a = \mathbb{E}[(a^T(x-\mu))^2] \geq 0$$

- a fixed direction column in R^{d}
- x, μ random column and population mean
- Σ population covariance: PSD means $a^T \Sigma a \geq 0$ for every a

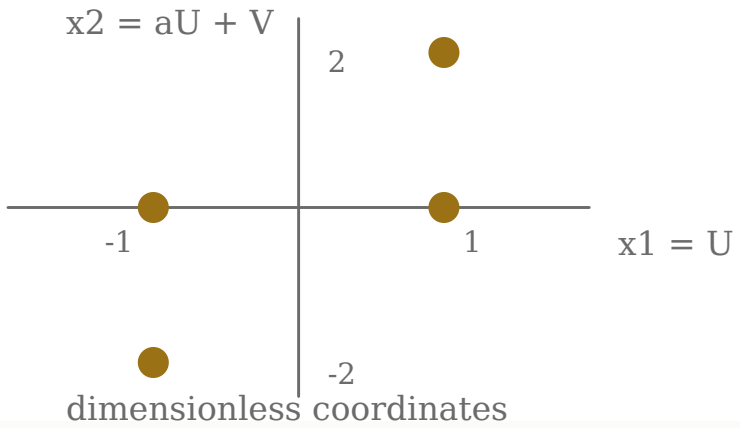
$$C = \frac{X_c^T X_c}{N-1} \quad (N > 1)$$

- C sample covariance estimate
- X_c N centered sample rows
- N number of samples

Positive semidefinite covariance — [1] DL §3.8 · Course derivation

Moments

How do independent signs create correlation?



a = 1.00

four equally likely outcomes

a: 0 to 1 adds the shared sign

Independent fair signs U,V: knowing one does not change the other's probabilities. Let $x = (U, U+V)^T$.

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

U, V independent, equally likely ± 1

x, μ random column and mean

Σ population covariance

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

CHOOSE, EXPLAIN, VOTE

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

DEFINITION AND REASONING

Moments

Does zero covariance guarantee independence?

x is uniform on $\{-1,0,1\}$; $y=x^2$. Is y independent of x?

$$\mathrm{Cov}(x,y)=0, \qquad P(y=0 \mid x=0)=1 \neq \tfrac{1}{3}=P(y=0)$$

x, y	random scalars
P	event probability
Cov	population covariance

Uncorrelatedness; independence — [1] DL §3.8 · [2] UBC Lec.18 p.8 · Course example

Gaussian geometry

What defines a multivariate Gaussian density?

$$p(x) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp(-d_M^2/2)$$

- x, μ evaluation point and mean in R^d
- d dimension
- $\Sigma \succ 0$ positive variance in every nonzero direction
- $|\Sigma|$ determinant
- p density
- d_M^2 squared Mahalanobis distance

$$d_M^2 = (x-\mu)^T \Sigma^{-1} (x-\mu)$$

- Σ^{-1} inverse covariance
- $x - \mu$ centered displacement

Multivariate Gaussian distribution — [1] DL §3.9.3

Gaussian geometry

Change basis, then standardize

Derivation

Derive the mean and covariance

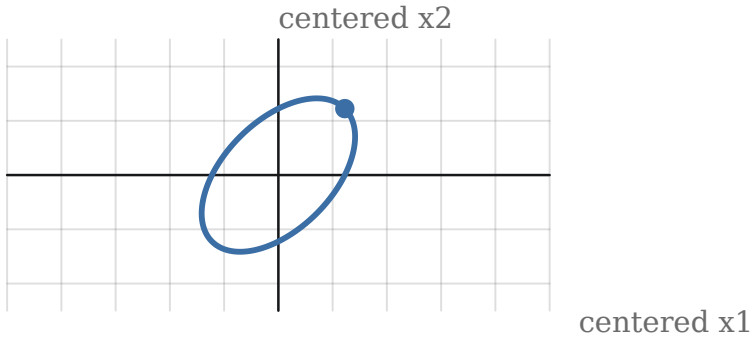
$\mu = \mathbb{E}[x]$; $\Sigma = \mathrm{Cov}(x) = Q\Lambda Q^T \succ 0$; $Q^TQ = I$; Λ positive diagonal. Fixed whitening map $W = \Lambda^{-1/2}Q^T$; $z = W(x-\mu)$; I: identity.

1. $\mathbb{E}[z] = W(\mathbb{E}[x] - \mu) = 0$ by linearity.
2. $\mathrm{Cov}(z) = \mathbb{E}[zz^T] = W\,\mathbb{E}[(x-\mu)(x-\mu)^T]\,W^T = W\Sigma W^T$.
3. $= \Lambda^{-1/2}Q^T(Q\Lambda Q^T)Q\Lambda^{-1/2} = \Lambda^{-1/2}\Lambda\Lambda^{-1/2} = I$.
4. Fixed matrices pass through expectation; $Q^TQ = I$ cancels the basis changes.

Whitening; eigendecomposition — [2] UBC Lec.18 p.5 · Course derivation

Gaussian geometry

A contour radius becomes an ellipse axis



unit Mahalanobis contour

eigenvalues 3 and 1

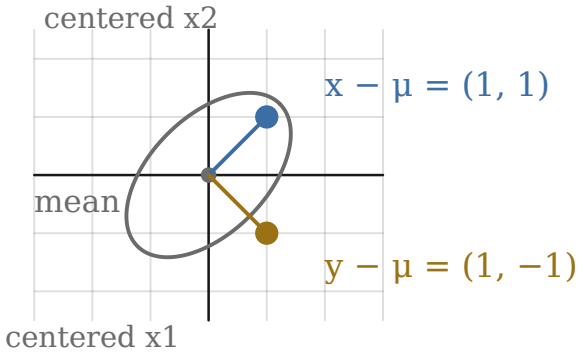
$$x = \mu + Q\Lambda^{1/2} z, \qquad \|z\| = c$$

c standardized radius; q_i: unit eigenvectors (columns of Q); $\sqrt{\lambda_i}$: axis scale. Plot of $x - \mu$: $c = 1$, $\lambda = (3, 1)$, $q_1 = (1, 1)^T/\sqrt{2}$, $q_2 = (1, -1)^T/\sqrt{2}$.

Gaussian density contour — [2] UBC Lec.18 p.5 · Course derivation

Gaussian geometry

Equal lengths can have unequal surprise



gray ellipse: $d_m = 1$

both arrows: Euclidean length $\sqrt{2}$

$$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mu = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Σ covariance; μ mean; compare $x = (2, 0)^T$ and $y = (2, -2)^T$.

$$d_M^2(x) = \tfrac{2}{3}, \qquad d_M^2(y) = 2$$

d_M covariance-scaled distance.

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

CHOOSE, EXPLAIN, VOTE

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.

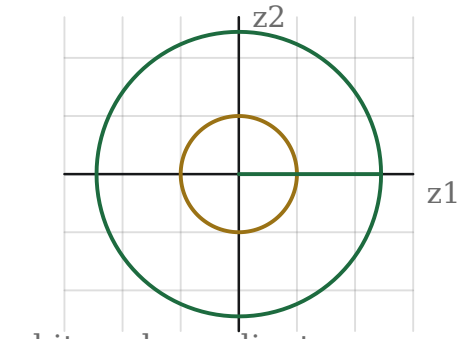
B $p(u) > p(v)$: u is one standard deviation away.

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

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

Gaussian geometry

A unit ellipse contains less than half



c = 1: mass 39.3%

c = 2.448: mass 95%

original semiaxes: 4.24 and 2.45

circle radius = Mahalanobis distance

$$P(d_M^2 \le c^2) = 1 - e^{-c^2/2} \quad (d = 2)$$

c contour radius; $d = 2$ dimension; the formula assumes a nonsingular Gaussian. Example: $\lambda_1 = 3, \lambda_2 = 1$.

$$c_{.95} = \sqrt{-2 \log .05} \approx 2.44775$$

$c_{.95}$ radius enclosing 95%.

Chi-square distribution; Gaussian probability region — [3] NIST §1.3.6.6.6 · [3] NIST χ^2 table · Course derivation

Bayes

Which population does a positive select?

joint events	positive +	negative −
event H	H and +	H and −
no event	not H and +	

$$P(H \mid +) = \frac{P(H \cap +)}{P(+)}$$

$$P(+ \mid H) = \frac{P(H \cap +)}{P(H)}$$

H: event; $+$: positive reading; $\cap$: both occur. Each denominator must be positive.

Conditional probability — [1] DL §3.4–3.5 · Course example

Bayes

Write the joint event in two ways

$$P(H \mid +) = \frac{P(+ \mid H)P(H)}{P(+)}$$

$P(H)$ prior; $P(+ \mid H)$ likelihood; $P(+) > 0$ evidence. H: event; $+$: positive reading.

Bayes' rule — [1] DL §3.11 · Course derivation

Bayes

Normalize over both sources of positives

$$P(+) = P(+ \mid H)P(H) + P(+ \mid \neg H)P(\neg H)$$

$\neg H$ complement; the two joint events are disjoint and exhaustive. H: event; $+$: positive reading.

Law of total probability; evidence — [1] DL §3.4–3.5 · [1] DL §3.11 · Course derivation

LECTURE 8

Bayes

Why can most positive alarms be false?

joint events	positive +	negative −
event H	90	10
no event	990	8910

$$P(H) = .01, \quad P(+ \mid H) = .9, \quad P(+ \mid \neg H) = .1$$

10,000 model trials: 100 event cases and 9,900 non-event cases.

$$P(H \mid +) = \frac{90}{90 + 990} = \frac{1}{12}$$

H event; $+$ positive reading.

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

CHOOSE, EXPLAIN, VOTE

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

Break

Pause: probability, density, and conditioning

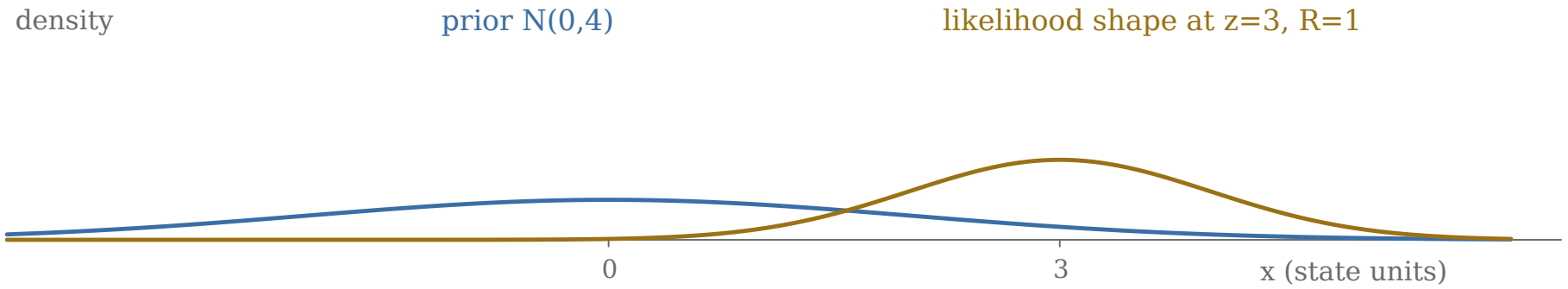
$$P(X \in A), \qquad p(x), \qquad P(H \mid +)$$

X random quantity; x evaluation value; A event; p density. H: event of interest; $+$: positive reading. Conditioning changes the reference population.

Event probability; density; conditional probability — [1] DL §3.3.2 · [1] DL §3.4–3.5 · Course recap

Gaussian conditioning

A numeric reading updates a curve



$$x \sim \mathcal{N}(\mu_0, P_0), \quad z = x + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, R)$$

x: state; z: reading; $P_0, R > 0$: prior/noise variances; ε independent of x. $p(x)$: prior; $p(z \mid x)$: likelihood at fixed z; $p(x \mid z)$: posterior.

Gaussian prior; Gaussian likelihood — [2] UBC Lec.18 p.7

Gaussian conditioning

Add precisions, then complete the square

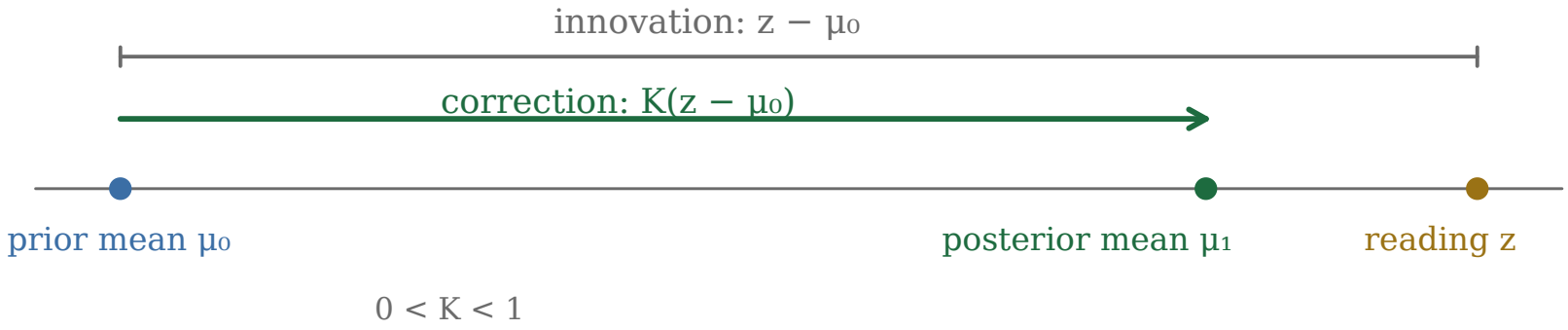
$$J(x) = \frac{(x-\mu_0)^2}{2P_0} + \frac{(z-x)^2}{2R}, \quad P_1^{-1} = P_0^{-1} + R^{-1}, \quad \mu_1 = P_1(\mu_0/P_0 + z/R)$$

J: negative log posterior, up to a constant. Independent Gaussian prior/noise; $P_0, R > 0$. x: state; z: reading. μ_0, P_0: prior mean/variance; R: noise variance; μ_1, P_1: posterior mean/variance.

Gaussian posterior; precision — [2] UBC Lec.18 p.7 · [1] DL §3.9.3 · Course derivation

Gaussian conditioning

The gain weights the innovation



$$\mu_1 = \frac{R\mu_0 + P_0 z}{P_0 + R} = \mu_0 + K(z - \mu_0), \quad K = \frac{P_0}{P_0 + R}, \quad P_1 = (1 - K)P_0$$

μ_0, P_0: prior mean/variance; z: reading; R: noise variance. μ_1, P_1: posterior mean/variance. Innovation $z - \mu_0$; gain $0 < K < 1$.

Kalman gain; measurement innovation — [5] Welch–Bishop §1 · [2] UBC Lec.18 p.7 · Course derivation

Gaussian conditioning

Fuse a prior and a precise reading



$$\mu_0 = 0, \quad P_0 = 4, \quad z = 3, \quad R = 1$$

μ_0: prior mean; P_0: prior variance; z: reading; R: independent noise variance.

$$K = \tfrac{4}{5}, \qquad \mu_1 = 2.4, \qquad P_1 = .8$$

K gain; μ_1, P_1 posterior mean, variance.

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

CHOOSE, EXPLAIN, VOTE

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

A correlated block predicts another block

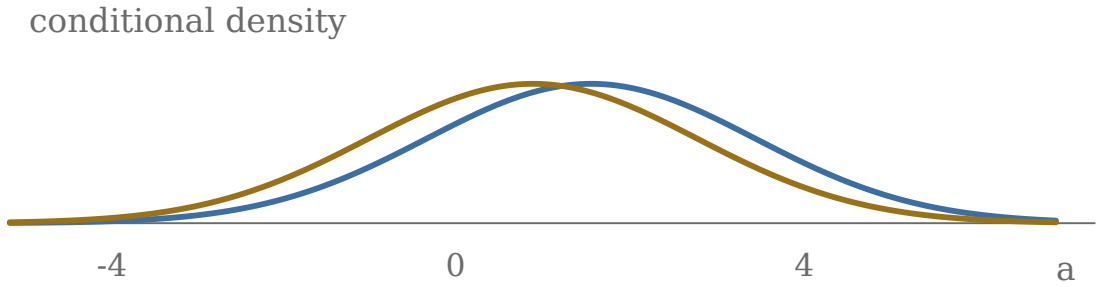
$$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(b - \mu_b)$$

Jointly Gaussian $a \in \mathbb{R}^p$, $b \in \mathbb{R}^q$; observe b. μ_a, μ_b: means; Σ_{bb}: invertible covariance of b. Cross-covariance: $\Sigma_{ab} = \mathbb{E}[(a - \mu_a)(b - \mu_b)^T]$.

Conditional Gaussian mean — [2] UBC Lec.18 p.7

Gaussian conditioning

Conditioning removes predictable variance



b = 5: mean 5/3

b = 2: mean 1

both variances 32/9

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$

a: unobserved block; b: observed block. Σ_{aa}: covariance of a; $\Sigma_{a|b}$: covariance after observing b. Joint Gaussian, $\Sigma_{bb} \succ 0$, $\Sigma_{ba} = \Sigma_{ab}^T$. Plot: $\mu = (1, 2)^T$, $\Sigma = \begin{bmatrix} 4 & 2 \\ 2 & 9 \end{bmatrix}$.

Conditional Gaussian covariance — [2] UBC Lec.18 p.7

CHOOSE, EXPLAIN, VOTE

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

Estimation

Maximum Likelihood Estimation (MLE)

$$b = A\theta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I)$$

$$L(\theta; b) = p(b \mid \theta), \quad \hat{\theta}_{MLE} \in \operatorname*{argmax}_{\theta} L(\theta; b)$$

$b \in \mathbb{R}^N$: fixed data; $A \in \mathbb{R}^{N \times p}$: known; $\theta \in \mathbb{R}^p$: unknown; $\sigma^2 > 0$: known variance. L: likelihood; argmax: maximizer(s).

Gaussian linear regression; maximum likelihood — [1] DL §5.5.1

Estimation

Gaussian products become a squared loss

$$-\log p(b \mid \theta) = \frac{N}{2}\log(2\pi\sigma^2) + \frac{\|b - A\theta\|^2}{2\sigma^2}$$

A known; b observed; θ unknown; independent errors, each $\mathcal{N}(0, \sigma^2)$, $\sigma^2 > 0$.

Negative log-likelihood; least squares — [1] DL §5.5.1 · Course derivation

Estimation

The familiar normal equation returns

$$A^T A\hat{\theta} = A^T b, \qquad A^T(b - A\hat{\theta}) = 0$$

$A \in \mathbb{R}^{N\times p}$: known design; $b \in \mathbb{R}^N$: data; $\theta \in \mathbb{R}^p$: parameters. Independent equal-variance Gaussian errors make $\hat{\theta}$ the MLE. Full column rank gives uniqueness; residual is orthogonal to columns.

Normal equations; least squares — [1] DL §5.1.4 · Course derivation

Estimation

Fit three readings with two coefficients

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$\theta = (\text{intercept}, \text{slope})^T$; $r = b - A\hat{\theta}$.

$$\hat{\theta} = \begin{pmatrix} 7/6 \\ 1/2 \end{pmatrix}, \quad r = \begin{bmatrix} -1/6 \\ 1/3 \\ -1/6 \end{bmatrix}$$

$\hat{\theta}$ estimate; r observed minus fitted.

Least-squares estimate — [1] DL §5.1.4 · Course example

Estimation

Maximum a Posteriori Estimation (MAP)

$$J(\theta) = \tfrac{1}{2}\|b - A\theta\|^2 + \tfrac{\lambda^2}{2}\|\theta - \mu_0\|^2, \quad \lambda^2 = \sigma^2/\tau^2$$

$\theta \sim \mathcal{N}(\mu_0, \tau^2 I)$ isotropic prior: variance τ^2 in every unit direction; σ^2 noise variance; λ^2 relative strength.

Maximum a posteriori (MAP); quadratic regularization — [1] DL §5.6.1 · Course derivation

Estimation

The prior changes the normal equation

$$(A^TA + \lambda^2 I)\hat{\theta} = A^Tb + \lambda^2\mu_0$$

Example: $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$, $b = (1, 2, 2)^T$, $\mu_0 = 0$, $\lambda^2 = 1$.

$$\hat{\theta}_{MAP} = \begin{pmatrix} 4/5 \\ 3/5 \end{pmatrix}, \quad \|b - A\hat{\theta}\|^2 = \tfrac{2}{5}$$

$\hat{\theta}_{MAP} = (\text{intercept}, \text{slope})^T$: posterior mode; $\|b - A\hat{\theta}\|^2$: squared residual norm.

MAP normal equations; ridge regularization — [1] DL §5.6.1 · [1] DL §5.1.4 · Course derivation

CHOOSE, EXPLAIN, VOTE

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

CHOOSE, EXPLAIN, VOTE

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

CHOOSE, EXPLAIN, VOTE

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

CHOOSE, EXPLAIN, VOTE

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

CHOOSE, EXPLAIN, VOTE

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

CHOOSE, EXPLAIN, VOTE

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

CHOOSE, EXPLAIN, VOTE

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

CHOOSE, EXPLAIN, VOTE

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

CHOOSE, EXPLAIN, VOTE

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

References

Sources and optional extensions

[1] Goodfellow et al. Deep Learning. MIT Press, 2016.

[2] UBC CS340. Multivariate Gaussian Distributions and LDA. 2011.

[3] NIST/SEMATECH. e-Handbook of Statistical Methods.

[4] Orloff & Bloom. Probability Intervals. MIT 18.05, 2022.

[5] Welch & Bishop. An Introduction to the Kalman Filter. 2004.