

HW #2 solutions (2026)

3.3 Define $E[X] = \mu$ and $V[X] = \sigma^2$,
then for $Y = aX + b$ we have:

$$\begin{cases} E[Y] = aE[X] + b = a\mu + b, \\ V[Y] = a^2V[X] = a^2\sigma^2. \end{cases}$$

Next, $E[XY] = E[aX^2 + bX] = a(\mu^2 + \sigma^2) + b\mu.$

$$\begin{aligned} \text{Cov}[X, Y] &= E[XY] - E[X]E[Y] = \\ &= a(\mu^2 + \sigma^2) + b\mu - \mu(a\mu + b) = \underline{\underline{a\sigma^2}}. \end{aligned}$$

Finally,

$$\begin{aligned} \rho &= \frac{\text{Cov}[X, Y]}{\sqrt{V[X]V[Y]}} = \frac{a\sigma^2}{\sqrt{\sigma^2}\sqrt{a^2\sigma^2}} = \\ &= \frac{a}{|a|}. \end{aligned}$$

Thus, $\rho = 1$ if $\underset{\substack{\uparrow \\ a = |a|}}{a} > 0$ and $\rho = -1$ if $\underset{\substack{\uparrow \\ a = -|a|}}{a} < 0$.

3.6 Since Σ is square and symmetric, we can write

$$\Sigma = U \Lambda U^T,$$

"diag($\lambda_1, \dots, \lambda_d$)"

$$\begin{aligned} \text{Then } \Sigma^{-1} &= (U^T)^{-1} \Lambda^{-1} U^{-1} = U \Lambda^{-1} U^T = \\ &= \sum_{i=1}^d \frac{1}{\lambda_i} \vec{u}_i \cdot \vec{u}_i^T. \end{aligned}$$

$$\text{Next, } (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) =$$

$$= \sum_{i=1}^d \frac{1}{\lambda_i} (\vec{x} - \vec{\mu}) \cdot \vec{u}_i \times \vec{u}_i \cdot (\vec{x} - \vec{\mu}) \quad (\equiv)$$

$$\text{Define } y_i = \vec{u}_i \cdot (\vec{x} - \vec{\mu}), \text{ or}$$

$$\underbrace{\vec{y}}_{\substack{\text{d-dim} \\ \text{vector}}} = U(\vec{x} - \vec{\mu}) \Rightarrow \vec{x} = U^T \vec{y} + \vec{\mu}.$$

$$(\equiv) \sum_{i=1}^d \frac{y_i^2}{\lambda_i}.$$

The Jacobian is given by

$$J_{ij} = \frac{\partial x_i}{\partial y_j} = (U^T)_{ij} = u_{ji}.$$

$$\text{Thus, } J = U^T \Rightarrow \det(J) = 1.$$

Finally, we obtain:

$$\begin{aligned} \int d\vec{x} \, e^{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})} &= \\ = \int dy_1 \dots dy_d \underbrace{|\mathbf{J}|}_{=1} e^{-\frac{1}{2} \sum_{i=1}^d \frac{y_i^2}{\lambda_i}} &= \\ = \prod_{i=1}^d \sqrt{2\pi\lambda_i} = (2\pi)^{d/2} \underbrace{|\Sigma|}_{\det(\Sigma)}^{1/2} &= \end{aligned}$$

4.3 We want an interval that satisfies:

$$(*) \quad p(l \leq \mu_n \leq u | D) \geq 0.95, \text{ where}$$

$$\begin{cases} l = \mu_n - 1.96\sigma_n, \\ u = \mu_n + 1.96\sigma_n. \end{cases}$$

Moreover, $u - l = 1$:

$$2 \times 1.96\sigma_n = 1, \text{ or}$$

$$\sigma_n^2 = \frac{1}{4(1.96)^2} \equiv \underbrace{\alpha}_{\text{const}}.$$

Recall that

$$\sigma_n^2 = \frac{\sigma^2}{n + \left(\frac{\sigma}{\sigma_0}\right)^2} \text{ in the posterior,}$$

$$\text{s.t.} \quad n = \frac{\sigma^2}{\alpha} - \left(\frac{\sigma}{\sigma_0}\right)^2 \approx 61 \text{ samples.}$$
$$\begin{cases} \sigma^2 = 4, \\ \sigma_0^2 = 9 \end{cases}$$

So, need $n \geq 62$ samples to satisfy Eq. (*).

4.5

(a) Consider $p(x, y | \theta_1, \theta_2)$:

	$y=0$	$y=1$	
$x=0$	$(1-\theta_1)\theta_2$	$(1-\theta_1)(1-\theta_2)$	$\begin{cases} x=0 : \text{tails} \\ x=1 : \text{heads} \end{cases}$
$x=1$	$\theta_1(1-\theta_2)$	$\theta_1\theta_2$	

(b) Since $p(x, y | \theta_1, \theta_2) = p(y | x, \theta_2) p(x | \theta_1)$,

$$\log p(\mathcal{D} | \vec{\theta}) = \sum_{i=1}^N \log p(x_i | \theta_1) + \sum_{i=1}^N \log p(y_i | x_i, \theta_2),$$

(θ_1, θ_2)

where $N=7$ is the # of measurements.

$$\text{MLE : } \begin{cases} \hat{\theta}_1 = \frac{N(x=1)}{N} = \frac{4}{7}, \\ \hat{\theta}_2 = \frac{N(x=y)}{N} = \frac{4}{7}. \end{cases} \quad \begin{matrix} \nwarrow \text{given the data} \\ \nwarrow \vec{x}, \vec{y} \end{matrix}$$

Finally, $p(\mathcal{D} | \hat{\theta}, M_2) = \left(\frac{4}{7}\right)^{N(x=1)} \left(\frac{3}{7}\right)^{N(x=0)} \times$

$\uparrow$
cur 2-param model

$$\times \left(\frac{4}{7}\right)^{N(x=y)} \left(\frac{3}{7}\right)^{N(x \neq y)} = \left(\frac{4}{7}\right)^8 \left(\frac{3}{7}\right)^6.$$

(c) Joint counts:
 $N(x, y)$

	$y=0$	$y=1$
$x=0$	2	1
$x=1$	2	2

$\uparrow$
sum to $N=7$

Then $\hat{\theta} = (\hat{\theta}_{\underset{\substack{\uparrow \uparrow \\ x \ y}}{0,0}}, \hat{\theta}_{0,1}, \hat{\theta}_{1,0}, \hat{\theta}_{1,1}) = (\frac{2}{7}, \frac{1}{7}, \frac{2}{7}, \frac{2}{7})$.

Likelihood $p(\mathcal{D} | \hat{\theta}, M_4) = \hat{\theta}_{0,0}^{N(x=0,y=0)} \hat{\theta}_{0,1}^{N(x=0,y=1)} \times \hat{\theta}_{1,0}^{N(x=1,y=0)} \hat{\theta}_{1,1}^{N(x=1,y=1)}$
 $\uparrow$
 4-param model

$$= \left(\frac{2}{7}\right)^6 \left(\frac{1}{7}\right)^1$$

(d) We need to compare

$$L(m) = \sum_{i=1}^N \log p(x_i, y_i | m, \underbrace{\hat{\theta}(\mathcal{D}_{-i})}_{\substack{\text{MLE computed on} \\ \text{data with the} \\ \text{measurement } i \\ \text{excluded}}})$$

leave-one-out cross-validated log L

Here, $m = M_2$ or M_4 .

Since $p(x_7, y_7 | M_4, \hat{\theta}(\mathcal{D}_{-7})) = 0$,

$$L(M_4) = -\infty$$

$L(M_2)$ remains finite and $\sqrt{M_2}$ will be preferred over M_4 .

$$(e) \text{ BIC}(M_2) = 8 \log\left(\frac{4}{7}\right) + 6 \log\left(\frac{3}{7}\right) - \underbrace{\frac{2}{2} \log 7}_{\substack{\text{complexity} \\ \text{penalty}}} = -11.51$$

$$\text{BIC}(M_4) = 6 \log\left(\frac{2}{7}\right) + 1 \log\left(\frac{1}{7}\right) - \frac{3}{2} \log 7 = -12.38$$

3 b/c 4 $\hat{\theta}_{x,y}$ prms sum up to 1

So, BIC also prefers M_2 .