

Allowed tools: Calculator, one handwritten double-sided A4 cheat sheet

Forbidden tools: Everything else

1. Priors [10 points]

- (i) State Bayes' theorem in its general form for a parameter θ and data y . Name and explain every component (prior, likelihood, marginal likelihood, posterior).
- (ii) Explain how Bayesian inference differs from frequentist inference in terms of (a) the treatment of parameters, (b) the role of prior information, and (c) the interpretation of probability.
- (iii) Let $Y_i \mid \lambda \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$, $i = 1, \dots, n$, and suppose $\lambda \sim \text{Gamma}(\alpha, \beta)$ (shape–rate parameterisation, so $p(\lambda) \propto \lambda^{\alpha-1}e^{-\beta\lambda}$). Derive the conjugate posterior distribution $p(\lambda \mid y_1, \dots, y_n)$. State the updated hyperparameters explicitly.

Priorit

- (i) Kirjoita Bayesin lause yleisessä muodossaan parametrille θ ja datalle y . Nimeä ja selitä jokainen komponentti (priori, uskottavuus, marginaaliuskottavuus, posteriori).
- (ii) Selitä, miten bayesilainen päättely eroaa frekventistisestä päättelystä (a) parameterien käsittelyssä, (b) prioritiedon roolissa ja (c) todennäköisyyden tulkitsemisessa.
- (iii) Olkoon $Y_i \mid \lambda \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$, $i = 1, \dots, n$, ja oletetaan $\lambda \sim \text{Gamma}(\alpha, \beta)$ (muota-nopeus-parametrisointi, joten $p(\lambda) \propto \lambda^{\alpha-1}e^{-\beta\lambda}$). Johda konjugaatiposteriorijakauma $p(\lambda \mid y_1, \dots, y_n)$. Ilmoita päivitetyt hyperparametrit eksplisiittisesti.

2. Regression [10 points]

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with μ *known*.

- (i) Write down the log-likelihood $\ell(\sigma^2; x_1, \dots, x_n)$.
- (ii) Derive the maximum-likelihood estimator $\hat{\sigma}_{\text{MLE}}^2$ by differentiating with respect to σ^2 and solving the score equation.

Regressio

Olkoon $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, missä μ on *tunnettu*.

- (i) Kirjoita log-uskottavuusfunktio $\ell(\sigma^2; x_1, \dots, x_n)$.
- (ii) Johda suurimman uskottavuuden estimaattori $\hat{\sigma}_{\text{MLE}}^2$ derivoimalla σ^2 :n suhteen ja ratkaisemalla score-yhtälö.

3. MC & MCMC [10 points]

Suppose you want to sample from the posterior

$$p(\theta | y) \propto \exp\left(-\frac{1}{2.4}(\theta - 3)^2\right) \cdot \exp\left(-\frac{1}{2}(y - \theta)^2\right), \quad y = 5.$$

- (i) Identify the prior $p(\theta)$ and the likelihood $p(y | \theta)$. State their distributions and parameters.
- (ii) Derive analytically the posterior mean and variance (use the formula for conjugate Gaussian–Gaussian updating).
- (iii) Describe in words the Random-Walk Metropolis–Hastings (RWMH) algorithm. Explain why the normalising constant cancels in the acceptance ratio and state the desired acceptance rate and the reasoning behind it.

MC & MCMC

Oletetaan, että haluat ottaa otoksen posteriorista

$$p(\theta | y) \propto \exp\left(-\frac{1}{2.4}(\theta - 3)^2\right) \cdot \exp\left(-\frac{1}{2}(y - \theta)^2\right), \quad y = 5.$$

- (i) Määritä priorijakauma $p(\theta)$ ja uskottavuus $p(y | \theta)$. Ilmoita niiden jakaumat ja parametrit.
- (ii) Johda analyttisesti posteriorin odotusarvo ja varianssi (käytä kaavaa konjugaatti-Gauss-Gauss-päivitykselle).
- (iii) Kuvaile sanallisesti satunnaiskulku-Metropolis-Hastings-algoritmi (RWMH). Selitä, miksi normalisointivakio supistuu pois hyväksymissuhteesta, ja ilmoita tavoiteltu hyväksymisaste sekä sen perustelu.

4. Model Selection [10 points]

A researcher fits four regression models to a dataset with $n = 30$ observations. The table below provides the number of free parameters k and the OLS coefficient estimates and $-2\hat{\ell}$ for each model.

Model	k	$\hat{\beta}$ (selected coeff.)	$-2\hat{\ell}$
M_1	2	(1.2, 0.8)	48.3
M_2	3	(1.1, 0.9, 0.3)	44.7
M_3	4	(1.0, 0.8, 0.4, 0.1)	43.1
M_4	5	(0.9, 0.7, 0.5, 0.2, 0.1)	42.8

- (i) Compute BIC for each model.
- (ii) Compute the pairwise Bayes Factors $B_{j1} = p(y | M_j)/p(y | M_1)$ using the BIC approximation $\log B_{ij} \approx -\frac{1}{2}(\text{BIC}_i - \text{BIC}_j)$, and interpret each one using Jeffreys' scale.
- (iii) Compute the BIC weights $w_j = e^{-\frac{1}{2}\Delta_j} / \sum_k e^{-\frac{1}{2}\Delta_k}$ where $\Delta_j = \text{BIC}_j - \min_k \text{BIC}_k$. Provide the Bayesian Model Averaging (BMA) estimate of the selected coefficient β_1 using $\hat{\beta}_1^{\text{BMA}} = \sum_j w_j \hat{\beta}_{1,j}$.

Mallinvalinta

Tutkija sovittaa neljä regressiomallia aineistoon, jossa on $n = 30$ havaintoa. Alla olevassa taulukossa on esitetty vapaiden parametrien lukumäärä k sekä OLS-kertoimien estimaatit ja $-2\hat{\ell}$ jokaiselle mallille.

Malli	k	$\hat{\beta}$ (valittu kerroin)	$-2\hat{\ell}$
M_1	2	(1.2, 0.8)	48.3
M_2	3	(1.1, 0.9, 0.3)	44.7
M_3	4	(1.0, 0.8, 0.4, 0.1)	43.1
M_4	5	(0.9, 0.7, 0.5, 0.2, 0.1)	42.8

- (i) Laske BIC jokaiselle mallille.
- (ii) Laske pareittaiset Bayesin kertoimet $B_{j1} = p(y | M_j)/p(y | M_1)$ käyttäen BIC-approksimaatiota $\log B_{ij} \approx -\frac{1}{2}(\text{BIC}_i - \text{BIC}_j)$, ja tulkitse jokainen Jeffreysin asteikon avulla.
- (iii) Laske BIC-painot $w_j = e^{-\frac{1}{2}\Delta_j} / \sum_k e^{-\frac{1}{2}\Delta_k}$ missä $\Delta_j = \text{BIC}_j - \min_k \text{BIC}_k$. Anna bayesilaisen mallikeskiarvoistuksen (BMA) estimaatti valituille kertoimille β_1 käyttäen $\hat{\beta}_1^{\text{BMA}} = \sum_j w_j \hat{\beta}_{1,j}$.

5. Time Series Analysis [10 points]

You are given ACF and PACF descriptions for four different time series. For each one, state the $\text{ARMA}(p, q)$ model you would choose and justify your answer.

- (a) ACF: cuts off after lag 2; PACF: decays exponentially (alternating signs).
- (b) ACF: decays exponentially to zero; PACF: cuts off after lag 1.
- (c) ACF: cuts off after lag 1; PACF: decays exponentially (positive).
- (d) ACF: both decay exponentially; PACF: both decay exponentially.

Additionally: when would you extend an ARMA model to an ARIMA or SARIMA model? How do you choose the integration order d and the seasonal period s ?

Aikasarja-analyysi

Sinulle on annettu ACF ja PACF kuvaukset neljälle eri aikasarjalle. Ilmoita kullekin sarjalle valitsemasi $\text{ARMA}(p, q)$ -malli ja perustele vastauksesi.

- (a) ACF: katkeaa viiveen 2 jälkeen; PACF: vaimenee eksponentiaalisesti (vuorottelevat merkit).
- (b) ACF: vaimenee eksponentiaalisesti nollaan; PACF: katkeaa viiveen 1 jälkeen.
- (c) ACF: katkeaa viiveen 1 jälkeen; PACF: vaimenee eksponentiaalisesti (positiiviset arvot).
- (d) ACF: molemmat vaimenevat eksponentiaalisesti; PACF: molemmat vaimenevat eksponentiaalisesti.

Lisäksi: milloin ARMA-malli tulisi laajentaa ARIMA- tai SARIMA-malliksi? Miten valitaan integraatioaste d ja kausittainen jakso s ?

PRIORS & BAYES

Bayes’ theorem

$$p(\theta \mid y) = \frac{p(y \mid \theta) p(\theta)}{p(y)}, \quad p(y) = \int p(y \mid \theta) p(\theta) \, \mathrm{d}\theta$$

Beta–Binomial conjugacy

$$\begin{aligned} \theta &\sim \text{Beta}(\alpha, \beta), \quad Y \mid \theta \sim \text{Bin}(n, \theta) \\ p(\theta \mid y) &= \text{Beta}(\alpha + y, \beta + n - y) \\ \mathbb{E}[\theta \mid y] &= \frac{\alpha + y}{\alpha + \beta + n} \end{aligned}$$

Poisson–Gamma conjugacy

$$\begin{aligned} \lambda &\sim \text{Ga}(\alpha, \beta), \quad Y_i \mid \lambda \stackrel{\text{iid}}{\sim} \text{Poi}(\lambda) \\ p(\lambda \mid \mathbf{y}) &= \text{Ga}(\alpha + \sum y_i, \beta + n) \\ \mathbb{E}[\lambda \mid \mathbf{y}] &= \frac{\alpha + \sum y_i}{\beta + n} \end{aligned}$$

Gaussian–Gaussian conjugacy

$$\begin{aligned} \theta &\sim \mathcal{N}(\mu_0, \tau_0^2), \quad y_i \mid \theta \stackrel{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2) \\ \sigma_n^2 &= \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \right)^{-1}, \quad \mu_n = \sigma_n^2 \left(\frac{\mu_0}{\tau_0^2} + \frac{\sum y_i}{\sigma^2} \right) \\ \theta \mid \mathbf{y} &\sim \mathcal{N}(\mu_n, \sigma_n^2) \end{aligned}$$

Fisher information & Jeffreys prior

$$\mathcal{I}(\theta) = -\mathbb{E} \left[\frac{\partial^2 \ln p(Y \mid \theta)}{\partial \theta^2} \right] = \mathbb{E} \left[\left(\frac{\partial \ln p}{\partial \theta} \right)^2 \right]$$

Bernoulli: $\mathcal{I}(\theta) = \frac{1}{\theta(1-\theta)}$
 $p_J(\theta) \propto \sqrt{\mathcal{I}(\theta)}$

Bernoulli/Binomial Jeffreys prior: $\text{Beta}(\frac{1}{2}, \frac{1}{2})$

Beta distribution

$$\begin{aligned} p(\theta) &= \frac{\theta^{\alpha-1} (1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \quad \mathbb{E}[\theta] = \frac{\alpha}{\alpha + \beta} \\ \text{Var}(\theta) &= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \end{aligned}$$

Gamma distribution (shape–rate)

$$p(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \mathbb{E}[\lambda] = \frac{\alpha}{\beta}, \quad \text{Var}(\lambda) = \frac{\alpha}{\beta^2}$$

REGRESSION

Log-likelihood

$$\ell(\theta) = \log(p(y \mid \theta))$$

OLS simple regression

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

OLS matrix form

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

Sums of squares

$$\begin{aligned} \text{SSE} &= \sum (y_i - \hat{y}_i)^2, \quad \text{SST} = \sum (y_i - \bar{y})^2 \\ \text{SSR} &= \text{SST} - \text{SSE}, \quad R^2 = 1 - \frac{\text{SSE}}{\text{SST}} = \frac{\text{SSR}}{\text{SST}} \end{aligned}$$

Bayesian linear regression

Prior: $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$

$$\boldsymbol{\Sigma}_n^{-1} = \boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X}$$

$$\boldsymbol{\mu}_n = \boldsymbol{\Sigma}_n \left(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + \frac{\mathbf{X}^\top \mathbf{y}}{\sigma^2} \right)$$

OLS = Bayesian posterior mean when $\boldsymbol{\Sigma}_0^{-1} \rightarrow \mathbf{0}$

MC & MCMC

Stationary distribution

$$\pi P = \pi, \quad \sum_i \pi_i = 1$$

Detailed balance (reversibility)

$$\pi_i P_{ij} = \pi_j P_{ji} \quad \forall i, j$$

MH acceptance ratio

$$\alpha(\theta^*, \theta^{(t)}) = \min \left(1, \frac{\tilde{\pi}(\theta^*) q(\theta^{(t)} \mid \theta^*)}{\tilde{\pi}(\theta^{(t)}) q(\theta^* \mid \theta^{(t)})} \right)$$

RWMH (q symmetric): $q(\theta^* \mid \theta^{(t)}) = q(\theta^{(t)} \mid \theta^*)$, so

$$\alpha = \min \left(1, \frac{\tilde{\pi}(\theta^*)}{\tilde{\pi}(\theta^{(t)})} \right)$$

RWMH proposal

$$\theta^* = \theta^{(t)} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \tau^2)$$

Accept if $u < \alpha$, $u \sim \text{Unif}(0, 1)$; else $\theta^{(t+1)} = \theta^{(t)}$.
Optimal acceptance rate: $\approx 44\%$ (univariate).

MODEL SELECTION

Gaussian log-likelihood (up to constant)

$$\hat{\ell} = -\frac{n}{2} \ln \left(\frac{\text{RSS}}{n} \right) \quad \Rightarrow \quad -2\hat{\ell} = n \ln \left(\frac{\text{RSS}}{n} \right)$$

Full constant: $-2\hat{\ell} = n \ln \left(\frac{\text{RSS}}{n} \right) + n(1 + \ln 2\pi)$

Information criteria

$$\begin{aligned} \text{AIC} &= -2\hat{\ell} + 2k \\ \text{BIC} &= -2\hat{\ell} + k \ln n \\ \text{AIC}_c &= \text{AIC} + \frac{2k(k+1)}{n-k-1} \end{aligned}$$

Use AIC_c when $n/k < 40$.

$$\Delta_j = \text{IC}_j - \min_k \text{IC}_k$$

BIC $\approx$ log marginal likelihood

$$\ln p(\mathbf{y} \mid M_j) \approx -\frac{1}{2} \text{BIC}_j$$

Bayes factors (BIC approximation)

$$B_{ij} = \frac{p(\mathbf{y} \mid M_i)}{p(\mathbf{y} \mid M_j)} \approx e^{-(\text{BIC}_i - \text{BIC}_j)/2}$$

BIC weights & BMA

$$w_j = \frac{e^{-\Delta_j/2}}{\sum_k e^{-\Delta_k/2}}, \quad \Delta_j = \text{BIC}_j - \min_k \text{BIC}_k$$

$$\hat{\theta}^{\text{BMA}} = \sum_j w_j \hat{\theta}_j$$

TIME SERIES

Weak stationarity conditions

- $\mathbb{E}[X_t] = \mu < \infty$ (constant)
- $\text{Var}(X_t) < \infty$ (constant)
- $\text{Cov}(X_t, X_{t-h}) = \gamma(h)$ depends only on h

Autocorrelation function (ACF)

$$\rho(h) = \gamma(h)/\gamma(0), \quad h \geq 0$$

AR(1): $X_t = \phi X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$

Stationary iff $|\phi| < 1$ (root of $1 - \phi z$ outside unit circle).

General AR(p): $\Phi(B)X_t = \varepsilon_t$

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

Stationary iff all roots of $\phi(z) = 0$ lie outside unit circle.

MA(q): $X_t = \Theta(B)\varepsilon_t$

$$\Theta(B) = 1 - \theta_1 B - \dots - \theta_p B^p$$

ACF/PACF identification

Process	ACF	PACF
AR(p)	Tails off	Cuts off lag p
MA(q)	Cuts off lag q	Tails off
ARMA(p, q)	Tails off	Tails off

ARIMA(p, d, q)

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t$$

d : number of regular differences for stationarity.

SARIMA(p, d, q)(P, D, Q) $_s$

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^D X_t = \Theta(B^s)\theta(B)\varepsilon_t$$

s : seasonal period. D : seasonal differences.

MARKOV CHAINS

n -step transition

$$P^{(n)} = P^n, \quad P_{ij}^{(n)} = \Pr(X_n = j \mid X_0 = i)$$

Irreducibility

All states communicate: $\forall i, j \exists n : P_{ij}^n > 0$.

Sufficient: all $P_{ij} > 0$ (positive matrix).

Period

$$d(i) = \gcd\{n \geq 1 : P_{ii}^n > 0\}$$

Aperiodic: $d(i) = 1$. Sufficient: $P_{ii} > 0$ for some i .

Ergodic theorem

Irreducible + aperiodic $\Rightarrow$ unique π ; $P^n \rightarrow \mathbf{1}\pi^\top$ as $n \rightarrow \infty$.

DISTRIBUTIONS (REFERENCE)

Normal $\mathcal{N}(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

Poisson(λ)

$$p(k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad \mathbb{E}[X] = \text{Var}(X) = \lambda$$

Binomial(n, θ)

$$p(k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}, \quad \mathbb{E}[X] = n\theta,$$

$$\text{Var}(X) = n\theta(1 - \theta)$$

Useful identities

$$\ln \prod_i a_i = \sum_i \ln a_i, \quad \frac{\partial}{\partial \sigma^2} \ln \sigma^2 = \frac{1}{\sigma^2}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\frac{\partial}{\partial \sigma^2} \frac{1}{\sigma^2} = -\frac{1}{\sigma^4}, \quad \sum_{i=1}^n (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$$

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$$