

Bayesian Learning

Lecture 4 - Prediction and Decision

Mattias Villani

Department of Statistics
Stockholm University



Lecture overview

- Prediction
- Decision making

- **Posterior predictive density** for future $\tilde{y}$ given observed $\mathbf{y} = (y_1, \dots, y_n)$

$$p(\tilde{y}|\mathbf{y}) = \int_{\theta} p(\tilde{y}|\theta, \mathbf{y})p(\theta|\mathbf{y})d\theta$$

- IID data:

$$p(\tilde{y}|\mathbf{y}) = \int_{\theta} p(\tilde{y}|\theta)p(\theta|\mathbf{y})d\theta$$

- **Parameter uncertainty** in $p(\tilde{y}|\mathbf{y})$ by **averaging over** $p(\theta|\mathbf{y})$.

Prediction - Normal data, known variance

- Under the uniform prior $p(\theta) \propto c$, then

$$p(\tilde{y}|\mathbf{y}) = \int_{\theta} p(\tilde{y}|\theta)p(\theta|\mathbf{y})d\theta$$

$$\theta|\mathbf{y} \sim N(\bar{y}, \sigma^2/n)$$

$$\tilde{y}|\theta \sim N(\theta, \sigma^2)$$

Simulation algorithm:

- 1 Generate a **posterior draw** of θ ($\theta^{(1)}$) from $N(\bar{y}, \sigma^2/n)$
- 2 Generate a **predictive draw** of $\tilde{y}$ ($\tilde{y}^{(1)}$) from $N(\theta^{(1)}, \sigma^2)$
- 3 Repeat Steps 1 and 2 N times to output:
 - ▶ Sequence of posterior draws: $\theta^{(1)}, \dots, \theta^{(N)}$
 - ▶ Sequence of predictive draws: $\tilde{y}^{(1)}, \dots, \tilde{y}^{(N)}$.

Predictive distribution - Normal model

- $\theta^{(1)} = \bar{y} + \varepsilon^{(1)}$, where $\varepsilon^{(1)} \sim N(0, \sigma^2/n)$. (Step 1).
- $\tilde{y}^{(1)} = \theta^{(1)} + v^{(1)}$, where $v^{(1)} \sim N(0, \sigma^2)$. (Step 2).
- $\tilde{y}^{(1)} = \bar{y} + \varepsilon^{(1)} + v^{(1)}$.
- $\varepsilon^{(1)}$ and $v^{(1)}$ are independent.
- The sum of two normal random variables is normal so

$$\begin{aligned} E(\tilde{y}|\mathbf{y}) &= \bar{y} \\ V(\tilde{y}|\mathbf{y}) &= \frac{\sigma^2}{n} + \sigma^2 = \sigma^2 \left(1 + \frac{1}{n}\right) \end{aligned}$$

$$\tilde{y}|\mathbf{y} \sim N\left[\bar{y}, \sigma^2 \left(1 + \frac{1}{n}\right)\right]$$

Iteration laws

- Expectation with respect to what? Explicit:

$$\mathbb{E}_{\theta|\mathbf{y}}(\theta) \equiv \int \theta p(\theta|\mathbf{y}) d\theta$$

- **Law of iterated expectation** and **Law of total variance**.

Iteration laws

Law of iterated expectation:

$$\mathbb{E}_X(X) = \mathbb{E}_Y(\mathbb{E}_{X|Y}(X))$$

Law of total variance:

$$\begin{aligned}\mathbb{V}_X(X) &= \mathbb{E}_Y(\mathbb{V}_{X|Y}(X)) \\ &\quad + \mathbb{V}_Y(\mathbb{E}_{X|Y}(X))\end{aligned}$$

Iteration laws for Bayes

Marginal posterior mean:

$$\mathbb{E}_{\theta_1|\mathbf{y}}(\theta_1) = \mathbb{E}_{\theta_2|\mathbf{y}}(\mathbb{E}_{\theta_1|\theta_2,\mathbf{y}}(\theta_1))$$

Marginal posterior variance:

$$\begin{aligned}\mathbb{V}_{\theta_1}(\theta_1) &= \mathbb{E}_{\theta_2|\mathbf{y}}(\mathbb{V}_{\theta_1|\theta_2,\mathbf{y}}(\theta_1)) \\ &\quad + \mathbb{V}_{\theta_2|\mathbf{y}}(\mathbb{E}_{\theta_1|\theta_2,\mathbf{y}}(\theta_1))\end{aligned}$$

Predictive distribution - Normal model and prior

- Predictive distribution still normal (sum of normals is normal).
- Predictive mean conditional on θ is trivial:

$$E_{\tilde{y}|\theta}(\tilde{y}) = \theta$$

- “Remove the conditioning” on θ by averaging over posterior:

$$E(\tilde{y}|\mathbf{y}) = E_{\theta|\mathbf{y}}(\theta) = \mu_n \text{ (Posterior mean of } \theta\text{).}$$

- The predictive variance of $\tilde{y}$ by **law of total variance**

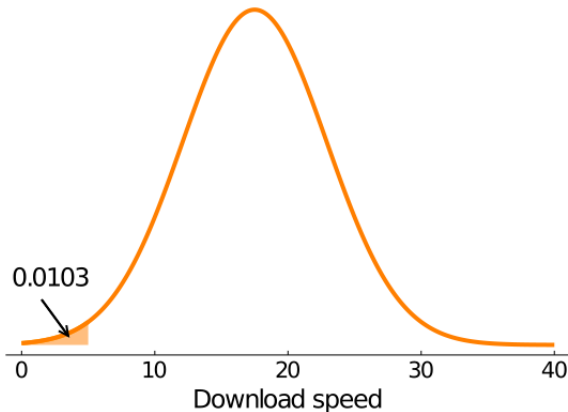
$$\begin{aligned} V(\tilde{y}|\mathbf{y}) &= E_{\theta|\mathbf{y}}[V_{\tilde{y}|\theta}(\tilde{y})] + V_{\theta|\mathbf{y}}[E_{\tilde{y}|\theta}(\tilde{y})] \\ &= E_{\theta|\mathbf{y}}(\sigma^2) + V_{\theta|\mathbf{y}}(\theta) \\ &= \sigma^2 + \tau_n^2 \end{aligned}$$

- So, predictive distribution is

$$\tilde{y}|\mathbf{y} \sim N(\mu_n, \sigma^2 + \tau_n^2).$$

Predictive distribution - Internet speed data

- My Netflix starts to buffer at speeds < 5 Mbit. 🤔



Bayesian prediction for time series

■ Autoregressive process

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \dots + \phi_p(y_{t-p} - \mu) + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$$

Predictive distribution - AR process.

Input: time series $\mathbf{y}_{1:T} = (y_1, \dots, y_T)$
number of predictive draws m .
forecast horizon h .

for i in $1:m$ **do**

$\mu, \phi_1, \dots, \phi_p, \sigma \leftarrow \text{rPOSTERIORAR}(\mathbf{y}_{1:T}, \text{PriorSettings})$

$\varepsilon_{T+1} \leftarrow \text{rNORM}(0, \sigma)$

$\tilde{y}_{T+1} \leftarrow \mu + \phi_1(y_T - \mu) + \dots + \phi_p(y_{T+1-p} - \mu) + \varepsilon_{T+1}$

$\varepsilon_{T+2} \leftarrow \text{rNORM}(0, \sigma)$

$\tilde{y}_{T+2} \leftarrow \mu + \phi_1(\tilde{y}_{T+1} - \mu) + \dots + \phi_p(y_{T+2-p} - \mu) + \varepsilon_{T+2}$

$\vdots$

$\varepsilon_{T+h} \leftarrow \text{rNORM}(0, \sigma)$

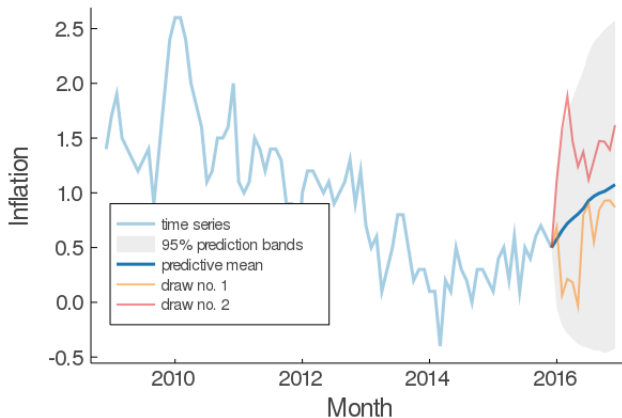
$\tilde{y}_{T+h} \leftarrow \mu + \phi_1(\tilde{y}_{T+h-1} - \mu) + \dots + \phi_p(\tilde{y}_{T+h-p} - \mu) + \varepsilon_{T+h}$

end

Output: m draws from the joint predictive density:

$$p(\tilde{y}_{T+1}, \dots, \tilde{y}_{T+h} | \mathbf{y}_{1:T}).$$

Bayesian prediction of Swedish inflation



Decision problems

- Let θ be an **unknown quantity**. **State of nature**.
 - ▶ Future inflation
 - ▶ Global temperature
 - ▶ Disease.
- Let $a \in \mathcal{A}$ be an **action**.
 - ▶ Interest rate
 - ▶ Energy tax
 - ▶ Surgery.
- Choosing action a when state of nature is θ gives **utility**

$$U(a, \theta)$$

- Alternatively **loss** $L(a, \theta) = -U(a, \theta)$.

Decision tables - when both a and θ are discrete

Decision table

	θ_1	θ_2	$\dots$	θ_K
a_1	$u(a_1, \theta_1)$	$u(a_1, \theta_2)$	$\dots$	$u(a_1, \theta_K)$
a_2	$u(a_2, \theta_1)$	$u(a_2, \theta_2)$	$\dots$	$u(a_2, \theta_K)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
a_J	$u(a_J, \theta_1)$	$u(a_J, \theta_2)$	$\dots$	$u(a_J, \theta_K)$

The eternal umbrella decision:

	Rain	Sun
No umbrella	-50	50
Umbrella	10	30

Decision Theory

- Example **loss functions** when both a and θ are continuous:

- ▶ **Linear:** $L(a, \theta) = |a - \theta|$
- ▶ **Quadratic:** $L(a, \theta) = (a - \theta)^2$
- ▶ **Lin-Lin:**

$$L(a, \theta) = \begin{cases} c_1 \cdot |a - \theta| & \text{if } a \leq \theta \\ c_2 \cdot |a - \theta| & \text{if } a > \theta \end{cases}$$

- Example:

- ▶ θ is the number of items demanded of a product
- ▶ a is the number of items in stock
- ▶ Utility

$$U(a, \theta) = \begin{cases} p \cdot \theta - c_1(a - \theta) & \text{if } a > \theta \text{ [too much stock]} \\ p \cdot a - c_2(\theta - a)^2 & \text{if } a \leq \theta \text{ [too little stock]} \end{cases}$$

Optimal decisions

■ Ad hoc decision rules:

- ▶ *Minimax*. Minimizes the maximum loss.
- ▶ *Minimax-regret*
- ▶ ... 

■ **Bayesian theory**: maximize **posterior expected utility**

$$a_{bayes} = \operatorname{argmax}_{a \in \mathcal{A}} E_{p(\theta|y)}[U(a, \theta)],$$

where $E_{p(\theta|y)}$ denotes the posterior expectation.

■ Using simulated draws $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}$ from $p(\theta|y)$:

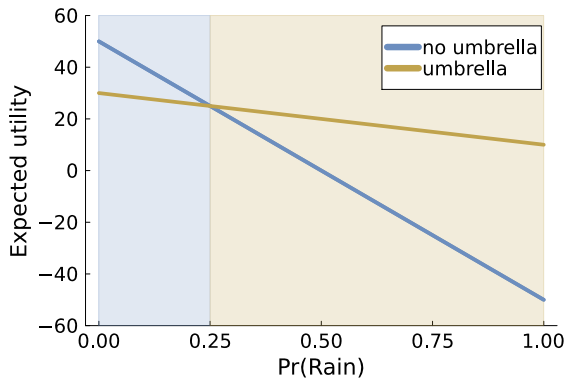
$$E_{p(\theta|y)}[U(a, \theta)] \approx N^{-1} \sum_{i=1}^N U(a, \theta^{(i)})$$

■ **Separation principle**:

- 1 First obtain $p(\theta|y)$
- 2 then form $U(a, \theta)$ and finally
- 3 choose a that maximizes $E_{p(\theta|y)}[U(a, \theta)]$.

The umbrella decision

	Rain	Sun
No umbrella	-50	50
Umbrella	10	30



Choosing a point estimate is a decision

- Choosing a **point estimator** is a decision problem.
- Which to choose: posterior median, mean or mode?
- It depends on your loss function:
 - ▶ **Linear loss** → Posterior median
 - ▶ **Quadratic loss** → Posterior mean
 - ▶ **Zero-one loss** → Posterior mode
 - ▶ **Lin-Lin loss** → $c_1/(c_1 + c_2)$ quantile of the posterior

The umbrella decision

