

# Bayesian Statistics I

## Lecture 4 - Predictions

Mattias Villani

Department of Statistics  
Stockholm University

Department of Computer and Information Science  
Linköping University



# Lecture overview

## ■ Prediction

- ▶ Normal model
- ▶ More complex examples

## ■ Decision theory

- ▶ The elements of a decision problem
- ▶ The Bayesian way
- ▶ Point estimation as a decision problem

# Prediction/Forecasting

- **Posterior predictive density** for future  $\tilde{y}$  given observed  $\mathbf{y}$

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

- If  $p(\tilde{y}|\theta, \mathbf{y}) = p(\tilde{y}|\theta)$  [not true for time series], then

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

- **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

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

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

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

# 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  $\theta$  ( $\theta^{(1)}$ ) 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

$$E(\tilde{y}|\mathbf{y}) = \bar{y}$$

$$V(\tilde{y}|\mathbf{y}) = \frac{\sigma^2}{n} + \sigma^2 = \sigma^2 \left(1 + \frac{1}{n}\right)$$

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

# Predictive distribution - Normal model and prior

- Easy to see that the predictive distribution is normal.
- The mean

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

and then remove the conditioning on  $\theta$  by averaging over  $\theta$

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

- The predictive variance of  $\tilde{y}$  (total variance formula):

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

- In summary:

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

# Bayesian prediction for time series

## ■ Autoregressive process

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

**Simulation algorithm.** Repeat  $N$  times:

- 1 Generate a **posterior draw** of  $\theta^{(1)} = (\phi_1^{(1)}, \dots, \phi_p^{(1)}, \mu^{(1)}, \sigma^{(1)})$  from  $p(\phi_1, \dots, \phi_p, \mu, \sigma | \mathbf{y}_{1:T})$ .
- 2 Generate a **predictive draw** of future time series by:
  - 1  $\tilde{y}_{T+1} \sim p(y_{T+1} | y_T, y_{T-1}, \dots, y_{T-p}, \theta^{(1)})$
  - 2  $\tilde{y}_{T+2} \sim p(y_{T+2} | \tilde{y}_{T+1}, y_T, \dots, y_{T-p}, \theta^{(1)})$
  - 3  $\tilde{y}_{T+3} \sim p(y_{T+3} | \tilde{y}_{T+2}, \tilde{y}_{T+1}, y_T, \dots, y_{T-p}, \theta^{(1)})$
  - 4 ...



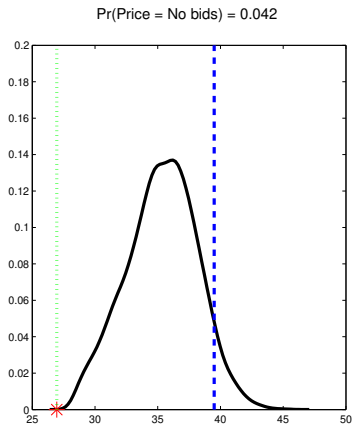
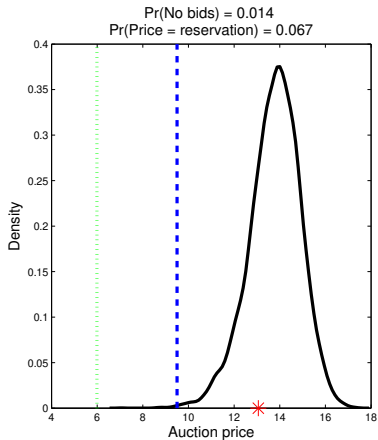
# Predicting auction prices on eBay

- Problem: **Predicting the auctioned price** in eBay coin auctions.
- **Data**: Bid from 1000 auctions on eBay.
  - ▶ The highest bid is not observed.
  - ▶ The lowest bids are also not observed because of the seller's reservation price.
- **Covariates**: auction-specific, e.g. Book value from catalog, seller's reservation price, quality of sold object, rating of seller, powerseller, verified seller ID etc
- Buyers are **strategic**. Their bids does not fully reflect their valuation. **Game theory**. Very complicated likelihood.

# Simulating auction prices on eBay

- Simulate from **posterior predictive distribution** of the **price** in a new auction:
  - 1 Simulate a draw  $\theta^{(i)}$  from the posterior  $p(\theta|\text{historical bids})$
  - 2 Simulate the number of bidders conditional on  $\theta^{(i)}$  (Poisson)
  - 3 Simulate the bidders' valuations,  $\mathbf{v}^{(i)}$
  - 4 Simulate all bids,  $\mathbf{b}^{(i)}$ , conditional on the valuations
  - 5 For  $\mathbf{b}^{(i)}$ , return the next to largest bid (proxy bidding).

# Predicting auction prices on eBay



# Decision Theory

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

$$U(a, \theta)$$

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

- Loss table:

	$\theta_1$	$\theta_2$
$a_1$	$L(a_1, \theta_1)$	$L(a_1, \theta_2)$
$a_2$	$L(a_2, \theta_1)$	$L(a_2, \theta_2)$

- Example:

	Rainy	Sunny
Umbrella	20	10
No umbrella	50	0

# Decision Theory, cont.

- Example **loss functions** when both  $a$  and  $\theta$  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:

- ▶  $\theta$  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 decision

- Ad hoc decision rules: *Minimax*. *Minimax-regret* ...
- **Bayesian theory**: maximize the **posterior expected utility**:

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

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 do inference,  $p(\theta|y)$
- 2 then form utility  $U(a, \theta)$  and finally
- 3 choose action  $a$  that maximizes  $E_{p(\theta|y)}[U(a, \theta)]$ .

# 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** →  $c_2 / (c_1 + c_2)$  quantile of the posterior