# Bayesian Learning Lecture 4 - Predictions and decisions

#### Bertil Wegmann

Department of Computer and Information Science Linköping University



#### Lecture overview

#### Prediction

- Normal model
- More complicated 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 y

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

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

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

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

#### Prediction - Normal data, known variance

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

$$p(\tilde{y}|y) = \int_{\theta} p(\tilde{y}|\theta) p(\theta|y) d\theta$$
$$\theta|y \sim N(\bar{y}, \sigma^{2}/n)$$
$$\tilde{y}|\theta \sim N(\theta, \sigma^{2})$$

#### Simulation algorithm:

- **I** 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:  $\theta^{(1)}, ...., \theta^{(N)}$
  - ▶ Sequence of predictive draws:  $\tilde{y}^{(1)}$ , ...,  $\tilde{y}^{(N)}$ .

#### Predictive distribution - Normal model

- lacksquare  $heta^{(1)} = \bar{y} + \varepsilon^{(1)}$ , 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).
- $\tilde{y}^{(1)} = \bar{y} + \varepsilon^{(1)} + v^{(1)}$ .
- lacksquare  $arepsilon^{(1)}$  and  $v^{(1)}$  are independent.
- The sum of two independent normal random variables is normal, so

$$\begin{split} E(\tilde{y}|\mathbf{y}) &= \bar{y} \\ V(\tilde{y}|\mathbf{y}) &= \frac{\sigma^2}{n} + \sigma^2 = \sigma^2 \left( 1 + \frac{1}{n} \right) \\ \tilde{y}|\mathbf{y} \sim N \left[ \bar{y}, \sigma^2 \left( 1 + \frac{1}{n} \right) \right] \end{split}$$

# Predictive distribution - Normal model and prior

- lacksquare  $heta^{(1)}=\mu_n+arepsilon^{(1)}, heta$  where  $arepsilon^{(1)}\sim extsf{N}(0, au_n^2).$  (Step 1).
- $\tilde{y}^{(1)} = \theta^{(1)} + v^{(1)}$ , where  $v^{(1)} \sim N(0, \sigma^2)$ . (Step 2).
- $\tilde{\mathbf{y}}^{(1)} = \mu_n + \varepsilon^{(1)} + v^{(1)}$ .
- lacksquare  $arepsilon^{(1)}$  and  $v^{(1)}$  are independent.
- The sum of two independent normal random variables is normal, so

$$E(\tilde{y}|y) = \mu_n$$

$$V(\tilde{y}|y) = \tau_n^2 + \sigma^2$$

$$\tilde{y}|y \sim N\left[\mu_n, \tau_n^2 + \sigma^2\right]$$

### Predictive distribution - Normal model and prior

The mean from the law of iterated expectations:

$$E(\tilde{y}|y) = E[E(\tilde{y}|\theta, y)|y] = E[\theta|y] = \mu_n$$

The variance from the law of total variance:

$$\begin{split} V(\tilde{y}|y) &= E[V(\tilde{y}|\theta,y)|y] + V[E(\tilde{y}|\theta,y)|y] \\ &= E\left[\sigma^2|y\right] + V\left[\theta|y\right] \\ &= \sigma^2 + \tau_n^2 \\ &= \text{(Population variance + Posterior variance of $\theta$)}. \end{split}$$

In summary:

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

## 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:

- Generate a **posterior draw** of  $\theta^{(1)} = (\phi_1^{(1)}, ..., \phi_p^{(1)}, \mu^{(1)}, \sigma^{(1)})$  from  $p(\phi_1, ..., \phi_p, \mu, \sigma|y_{1:T})$ .
- 2 Generate a predictive draw of future time series by:

1 
$$\tilde{y}_{T+1} \sim p\left(y_{T+1}|y_T, y_{T-1}, ..., y_{T-\rho}, \theta^{(1)}\right)$$

2 
$$\tilde{y}_{T+2} \sim p\left(y_{T+2}|\tilde{y}_{T+1}, y_T, ..., y_{T-p+1}, \theta^{(1)}\right)$$

3 
$$\tilde{y}_{T+3} \sim p\left(y_{T+3}|\tilde{y}_{T+2},\tilde{y}_{T+1},y_{T},...,y_{T-p+2},\theta^{(1)}\right)$$

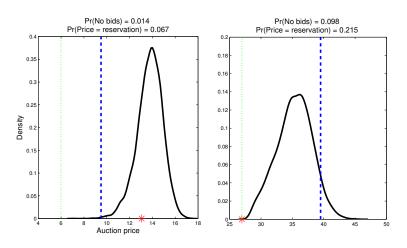
## Predicting auction prices on eBay

- 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 distibution of the price in a new auction:
- **I** 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,  $v^{(i)}$
- 4 Simulate all bids,  $b^{(i)}$ , conditional on the valuations
- **5** For  $b^{(i)}$ , 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  $L(a, \theta) = -U(a, \theta)$ .

Example: Umbrella 20 10
No umbrella 50 0

## Decision Theory, cont.

- lacksquare Example loss functions when both a and heta are continuous:
  - ▶ Linear:  $L(a, \theta) = |a \theta|$
  - **Quadratic**:  $L(a, \theta) = (a \theta)^2$
  - ► Lin-Lin

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

- Example:
  - ightharpoonup heta 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:

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

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

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

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

- Separation principle:
- I 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** ightarrow  $c_2/(c_1+c_2)$  quantile of the posterior