Bayesian analysis of VARs, State-space models and DSGEs Part II: Prediction, Model inference, Decisions

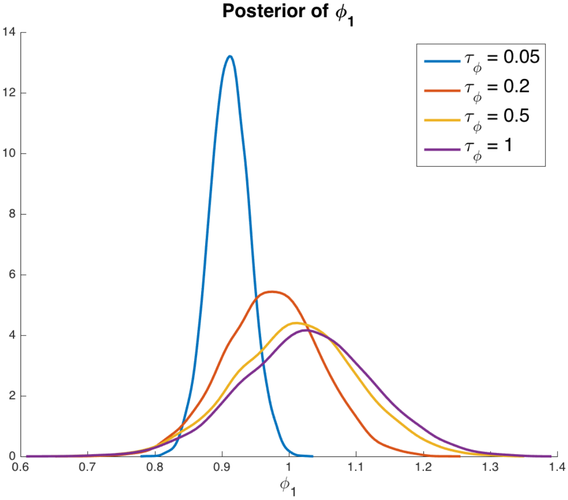
Mattias Villani

Division of Statistics and Machine Learning
Department of Computer and Information Science
Linköping University

Lecture overview

- **Bayesian prediction**
- **Model comparison**
- **Model evaluation**
- **Bayesian decision making**

Univariate AR(4) posterior
Foreign interest rate 1980Q2-2005Q4



Marginalization

- Wait! How could I plot $p(\phi_1|y_1, ..., y_T)$ in the AR(4). What happend to $\phi_2, \phi_3, \phi_4, c$ and σ^2?
- Example: **Regression model**:

$$\mathbf{y}|\mathbf{X}, \beta \sim N\left(\mathbf{X}\beta, \sigma^2 I_n\right).$$

- **Posterior** $p(\beta, \sigma^2|\mathbf{y}, \mathbf{X})$ is a $(k+1)$-dimensional posterior distribution. Hard to visualize!
- **Marginal posterior** of β_i

$$p(\beta_i|\mathbf{y}, \mathbf{X}) = \int_{\beta_{-i}} \int_{\sigma^2} p(\beta, \sigma^2|\mathbf{y}, \mathbf{X}) d\sigma^2 d\beta_{-i}$$

- Marginal posteriors are immediately available when we approximate the joint posterior by simulation (MCMC).

Prediction/Forecasting

- Example: **Regression model**:

$$\mathbf{y}|\mathbf{X},\beta \sim N\left(\mathbf{X}\beta, \sigma^2 I_n\right).$$

- **Posterior predictive distribution** for new observation $\tilde{y}$ given $\tilde{\mathbf{x}}$ and estimation sample $(\mathbf{y},\mathbf{X})$:

$$p(\tilde{y}|\tilde{\mathbf{x}},\mathbf{y},\mathbf{X}) = \int_\beta p(\tilde{y}|\tilde{\mathbf{x}},\beta)p(\beta|\mathbf{y},\mathbf{X})d\beta$$

- The **parameter uncertainty** is represented in $p(\tilde{y}|\tilde{\mathbf{x}},\mathbf{y},\mathbf{X})$ by **averaging over** posterior $p(\beta|\mathbf{y},\mathbf{X})$.
- It can be shown that $p(\tilde{y}|\tilde{\mathbf{x}},\mathbf{y},\mathbf{X})$ is a student-t density.
- When the integral cannot be computed analytically: simulation!

Example: Bayesian prediction in steady-state AR processes

- **Autoregressive process**

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

- Simulate a draw from $p(\phi_1, \phi_2, \ldots, \phi_p, \mu, \sigma | y)$ [**Gibbs sampling**, Part III]
  - Conditional on that draw $\theta^{(1)} = (\phi_1^{(1)}, \phi_2^{(1)}, \ldots, \phi_p^{(1)}, \mu^{(1)}, \sigma^{(1)})$, simulate
  - $\tilde{y}_{T+1} \sim p(y_{T+1} | y_T, y_{T-1}, \ldots, y_{T-p}, \theta^{(1)})$
  - $\tilde{y}_{T+2} \sim p(y_{T+2} | \tilde{y}_{T+1}, y_T, \ldots, y_{T-p}, \theta^{(1)})$
  - and so on.
- Repeat for new θ draws.

Bayesian model comparison [1]

- Consider two models for the data $\mathbf{y} = (y_1, ..., y_n)$: M_1 and M_2.
- **Estimated likelihoods** $p_1(\mathbf{y}|\hat{\theta}_1)$ and $p_2(\mathbf{y}|\hat{\theta}_2)$ can not be used directly for model comparison. Bigger models always win.
- Bayesian: the **marginal likelihood** for model M_k with parameters θ_k

$$p(\mathbf{y}|M_k) = \int p_k(\mathbf{y}|\theta_k) p_k(\theta_k) d\theta_k.$$

- θ_k is "removed" by the prior. **Not a magic bullet**. **Priors matter!**
- Often reported on log scale:
  - Strong evidence for M_1 if $3 < \ln p(y|M_1) - \ln p(y|M_2) \leq 5$
  - Very strong evidence for M_1 if $\ln p(y|M_1) - \ln p(y|M_2) > 5$.

DSGE EXAMPLE [2]

Parameter	Prior distribution			Posterior distributions									
				Instrument rule without policy break		Fixed exchange rate rule		Semi-fixed exchange rate rule		Instrument rule with policy break			
	Type	Mean	Std. dev./df	Median	Std.	Median	Std.	Median	Std.	UIP		Modified UIP	
										Median	Std.	Median	Std.
Calvo wages ξ_w	beta	0.750	0.050	0.751	0.047	0.518	0.041	0.669	0.046	0.743	0.049	0.752	0.049
Calvo domestic prices ξ_d	beta	0.750	0.050	0.862	0.046	0.852	0.048	0.885	0.027	0.868	0.044	0.838	0.044
Calvo import cons. prices $\xi_{m,c}$	beta	0.750	0.050	0.896	0.017	0.922	0.013	0.900	0.014	0.900	0.017	0.901	0.017
Calvo import inv. prices $\xi_{m,i}$	beta	0.750	0.050	0.946	0.010	0.948	0.008	0.943	0.007	0.946	0.010	0.944	0.010
Calvo export prices ξ_x	beta	0.750	0.050	0.868	0.021	0.870	0.016	0.874	0.020	0.869	0.021	0.883	0.020
Indexationwages κ_w	beta	0.500	0.150	0.290	0.098	0.238	0.086	0.287	0.098	0.292	0.100	0.313	0.103
Indexation prices κ_d	beta	0.500	0.150	0.213	0.059	0.163	0.069	0.194	0.052	0.212	0.061	0.218	0.061
⋮													
Output response $r_{y,1}$	normal	0.125	0.050	0.129	0.046			0.216	0.051	0.113	0.044	0.138	0.048
Diff. output response $r_{\Delta y,1}$	normal	0.063	0.050	0.152	0.036			0.142	0.050	0.127	0.041	0.120	0.046
Monetary policy shock $\sigma_{R,1}$	invgamma	0.150	2	0.249	0.024			2.335	0.778	0.398	0.060	0.398	0.066
Inflation target shock $\sigma_{\bar{\pi}^c,1}$	invgamma	0.050	2	0.116	0.041			0.083	0.054	0.148	0.067	0.248	0.085
Interest rate smoothing $\rho_{R,2}$	beta	0.800	0.050			0.884	0.018	0.864	0.021	0.896	0.018	0.874	0.022
Inflation response $r_{\pi,2}$	truncnormal	1.700	0.100			1.725	0.090	1.747	0.089	1.709	0.099	1.718	0.097
Diff. infl response $r_{\Delta\pi,2}$	normal	0.300	0.050			0.127	0.023	0.143	0.025	0.104	0.026	0.120	0.027
Real exch. rate response $r_{x,2}$	normal	0.000	0.050			0.022	0.019	−0.001	0.003	0.038	0.026	−0.023	0.020
Output response $r_{y,2}$	normal	0.125	0.050			0.269	0.040	0.274	0.039	0.107	0.041	0.106	0.041
Diff. output response $r_{\Delta y,2}$	normal	0.063	0.050			0.099	0.031	0.107	0.030	0.104	0.030	0.105	0.030
Monetarypolicy shock $\sigma_{R,2}$	invgamma	0.150	2			0.102	0.013	0.094	0.011	0.104	0.013	0.103	0.013
Inflation target shock $\sigma_{\bar{\pi}^c,2}$	invgamma	0.050	2			0.065	0.030	0.069	0.035	0.080	0.038	0.077	0.038
Log marginal likelihood				−2285.8		−2636.72		−2348.24		−2268.33		−2252.57	

Bayesian model comparison

- The **Bayes factor**

$$B_{12}(y) = \frac{p(\mathbf{y}|M_1)}{p(\mathbf{y}|M_2)}.$$

- **Posterior model probabilities**

$$\underbrace{\Pr(M_k|\mathbf{y})}_{\text{posterior model prob.}} \propto \underbrace{p(\mathbf{y}|M_k)}_{\text{marginal likelihood}} \cdot \underbrace{\Pr(M_k)}_{\text{prior model prob.}}$$

where

$$p(\mathbf{y}|M_k) = \int p_k(\mathbf{y}|\theta_k)p_k(\theta_k)d\theta_k.$$

- Two different priors:
  - priors over the models $\Pr(M_k)$
  - prior $p_k(\theta_k)$ for the parameters θ_k within model M_k.

Model choice in multivariate time series [3]

- Multivariate time series

$$\mathbf{x}_t = \alpha\beta'\mathbf{z}_t + \Phi_1\mathbf{x}_{t-1} + \ldots\Phi_k\mathbf{x}_{t-k} + \Psi_1 + \Psi_2 t + \Psi_3 t^2 + \varepsilon_t$$

- Need to choose:
  - **Lag length**, $(k = 1, 2.., 4)$
  - **Trend model** $(s = 1, 2, ..., 5)$
  - **Long-run (cointegration) relations** $(r = 0, 1, 2, 3, 4)$.

The most probable (k, r, s) combinations in the Danish monetary data.

k	1	1	1	1	1	1	1	1	0	1
r	3	3	2	4	2	1	2	3	4	3
s	3	2	2	2	3	3	4	4	4	5
$p(k, r, s \mid y, x, z)$	.106	.093	.091	.060	.059	.055	.054	.049	.040	.038

Example: Geometric vs Poisson

- Model 1 - **Geometric** with Beta prior:
  - $y_1, ..., y_n | \theta_1 \sim Geo(\theta_1)$
  - $\theta_1 \sim Beta(\alpha_1, \beta_1)$
- Model 2 - **Poisson** with Gamma prior:
  - $y_1, ..., y_n | \theta_2 \sim Poisson(\theta_2)$
  - $\theta_2 \sim Gamma(\alpha_2, \beta_2)$
- Marginal likelihood for M_1

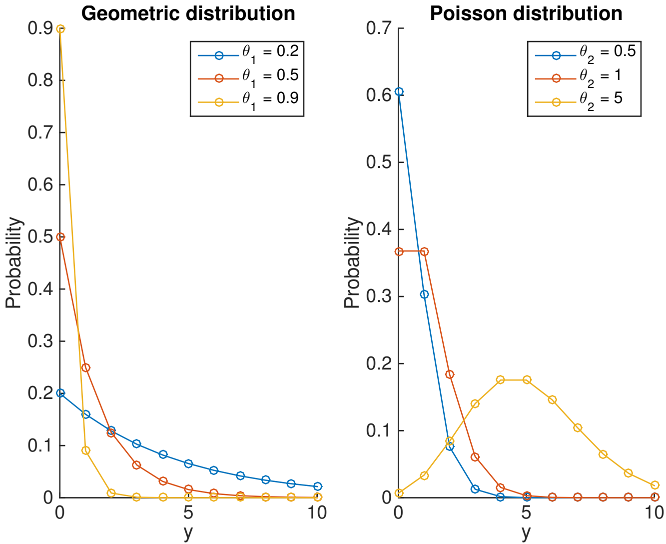
$$
\begin{aligned}
p_1(y_1, ..., y_n) &= \int p_1(y_1, ..., y_n | \theta_1) p(\theta_1) d\theta_1 \\
&= \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(n + \alpha_1)\Gamma(n\bar{y} + \beta_1)}{\Gamma(n + n\bar{y} + \alpha_1 + \beta_1)}
\end{aligned}
$$

- Marginal likelihood for M_2

$$
p_2(y_1, ..., y_n) = \frac{\Gamma(n\bar{y} + \alpha_2)\beta_2^{\alpha_2}}{\Gamma(\alpha_2)(n + \beta_2)^{n\bar{y}+\alpha_2}} \frac{1}{\prod_{i=1}^{n} y_i!}
$$

VOID

Geometric and Poisson



VOID

Geometric vs Poisson, cont.

- Priors match prior predictive means:

$$E(y_i|M_1) = E(y_i|M_2) \iff \alpha_1\alpha_2 = \beta_1\beta_2$$

- **Data**: $y_1 = 0,\ y_2 = 0.$

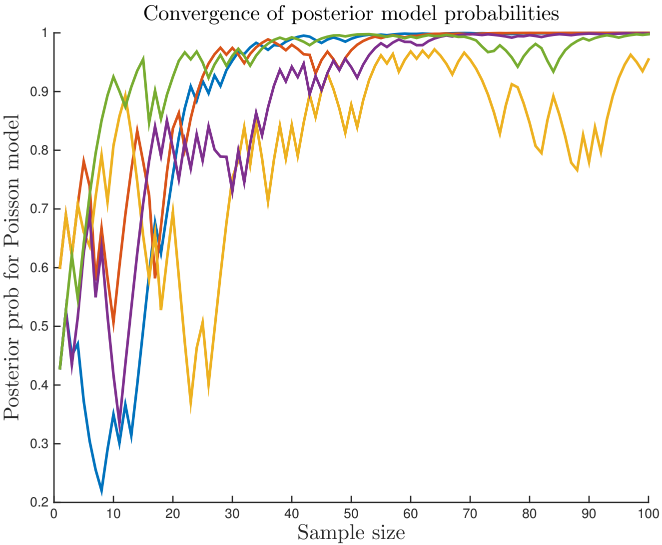
	$\alpha_1 = 1, \beta_1 = 2$	$\alpha_1 = 10, \beta_1 = 20$	$\alpha_1 = 100, \beta_1 = 200$
	$\alpha_2 = 2, \beta_2 = 1$	$\alpha_2 = 20, \beta_2 = 10$	$\alpha_2 = 200, \beta_2 = 100$
BF_{12}	1.5	4.54	5.87
$\Pr(M_1\|\mathbf{y})$	0.6	0.82	0.85
$\Pr(M_2\|\mathbf{y})$	0.4	0.18	0.15

- **Data**: $y_1 = 3,\ y_2 = 3.$

	$\alpha_1 = 1, \beta_1 = 2$	$\alpha_1 = 10, \beta_1 = 20$	$\alpha_1 = 100, \beta_1 = 200$
	$\alpha_2 = 2, \beta_2 = 1$	$\alpha_2 = 20, \beta_2 = 10$	$\alpha_2 = 200, \beta_2 = 100$
BF_{12}	0.26	0.29	0.30
$\Pr(M_1\|\mathbf{y})$	0.21	0.22	0.23
$\Pr(M_2\|\mathbf{y})$	0.79	0.78	0.77

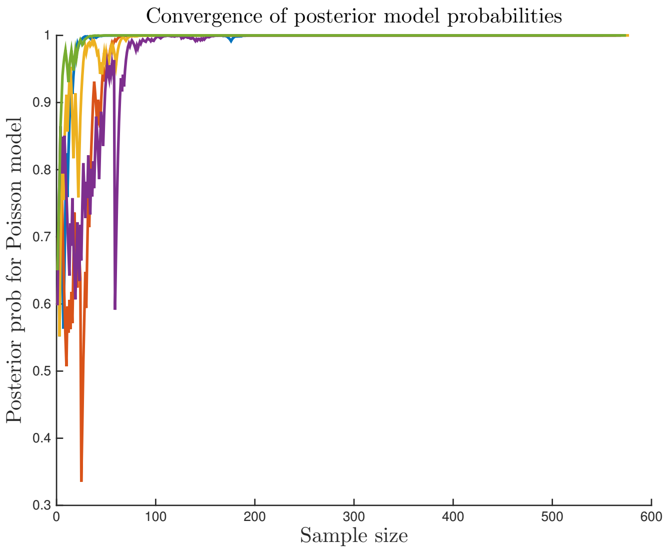
VOID

Geometric vs Poisson for Pois(1) data



VOID

Geometric vs Poisson for Pois(1) data



Properties of Bayesian model comparison

- **Consistency** when true model is in $\mathcal{M} = \{M_1, ..., M_K\}$

$$\Pr(M = M_{TRUE}|\mathbf{y}) \to 1 \quad \text{as} \quad n \to \infty$$

- "KL-consistency" when $M_{TRUE} \notin \mathcal{M}$

$$\Pr(M = M^*|\mathbf{y}) \to 1 \quad \text{as} \quad n \to \infty$$

  where M^* is the model that minimizes Kullback-Leibler distance between $p_M(\mathbf{y})$ and $p_{TRUE}(\mathbf{y})$.

1. Smaller models always win when priors are very vague.

- **Improper priors** can't be used for model comparison.

Marginal likelihood measures out-of-sample predictive performance

- The **marginal likelihood** can be decomposed as

$$p(y_1, ..., y_n) = p(y_1)p(y_2|y_1) \cdots p(y_n|y_1, y_2, ..., y_{n-1})$$

- If we assume that y_i is independent of $y_1, ..., y_{i-1}$ conditional on θ:

$$p(y_i|y_1, ..., y_{i-1}) = \int p(y_i|\theta)p(\theta|y_1, ..., y_{i-1})d\theta$$

- The prediction of y_1 is based on the prior of θ, and is therefore **sensitive to the prior**.
- The prediction of y_n uses almost all the data to infer θ. Very little influenced by the prior when n is not small.

Normal example

- **Model**: $y_1, ..., y_n | \theta \sim N(\theta, \sigma^2)$ with σ^2 known.
- **Prior**: $\theta | \sigma^2 \sim N(0, \kappa^2 \sigma^2)$.
- Intermediate posterior at time $i - 1$

$$\theta | y_1, ..., y_{i-1} \sim N\left[w_i(\kappa) \cdot \bar{y}_{i-1}, \frac{\sigma^2}{i - 1 + \kappa^{-2}} \right]$$

  where $w_i(\kappa) = \frac{i-1}{i-1+\kappa^{-2}}$.
- Predictive density at time $i - 1$

$$y_i | y_1, ..., y_{i-1} \sim N\left[w_i(\kappa) \cdot \bar{y}_{i-1}, \sigma^2 \left(1 + \frac{1}{i - 1 + \kappa^{-2}} \right) \right]$$

- Terms with i large: $y_i | y_1, ..., y_{i-1} \overset{approx}{\sim} N\left(\bar{y}_{i-1}, \sigma^2\right)$, not sensitive to κ
- For $i = 1$, $y_1 \sim N\left[0, \sigma^2 \left(1 + \frac{1}{\kappa^{-2}}\right)\right]$ can be very sensitive to κ.

Log Predictive Score - LPS [4, 5]

- To reduce sensitivity to the prior: sacrifice n^* observations to train the prior into a better posterior.
- Predictive density score: PS

$$PS(n^*) = p(y_{n^*+1}|y_1, ..., y_{n^*}) \cdots p(y_n|y_1, ..., y_{n-1})$$

- Usually report on log scale: **Log Predictive Score** (**LPS**).
- But which observations to train on (and which to test on)?
- Straightforward for time series.
- Cross-sectional data: **cross-validation**.

Model averaging

- Let γ be a quanitity with an interpretation which stays the same across the two models.
- Example: Prediction $\gamma = (y_{T+1}, ..., y_{T+h})'$.
- The marginal posterior distribution of γ reads

$$p(\gamma|\mathbf{y}) = p(M_1|\mathbf{y})p_1(\gamma|\mathbf{y}) + p(M_2|\mathbf{y})p_2(\gamma|\mathbf{y}),$$

  where $p_k(\gamma|\mathbf{y})$ is the marginal posterior of γ conditional on model k.
- Predictive distribution includes **three sources of uncertainty**:
  - **Future errors**/disturbances (e.g. the ε's in a regression)
  - **Parameter uncertainty** (the predictive distribution has the parameters integrated out by their posteriors)
  - **Model uncertainty** (by model averaging)

DECOMPOSE PREDICTION UNCERTAINTY - DSGE [6]

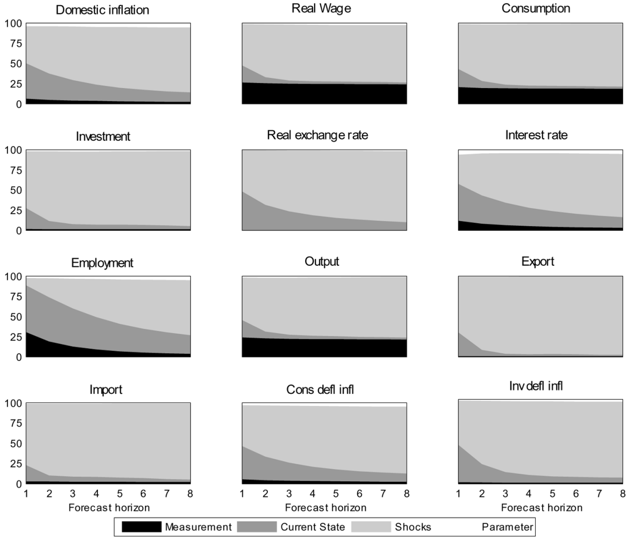


FIGURE 4 Decomposition of the forecast uncertainty. The subgraphs display the relative contribution to the predictive variances of the observed variables at different forecast horizons.

Bayesian forecast averaging

- Available: forecasts $\hat{x}^{(1)}_{t+h|t}, ..., \hat{x}^{(k)}_{t+h|t}$ from k different institutes/models.
- How to **combine the forecasts** to a single forecast?
- Bayesian solution assuming

$$\hat{\mathbf{x}}^{(1)}_{t+h|t} \sim N\left(x_t \mathbf{1}, \Sigma\right)$$

  where Σ describes the covariance between institutes' forecasts.
- **Optimal forecast combination** [7, 8]:

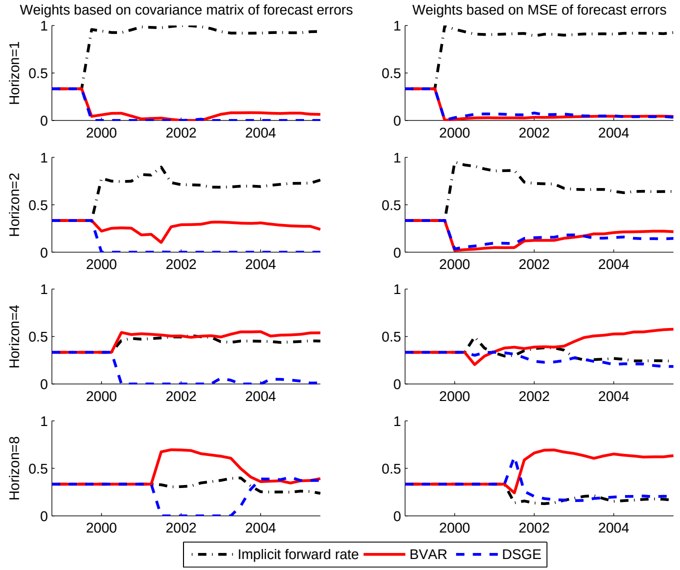
$$\sum_{j=1}^{k} w_{jt} \hat{x}^{(j)}_{t+h|t}$$

$$\left(w_{1t}, w_{2t}, ..., w_{kt}\right) = \frac{\mathbf{1}'\tilde{\Sigma}_t^{-1}}{\mathbf{1}'\tilde{\Sigma}_t^{-1}\mathbf{1}}$$

$$\tilde{\Sigma}_t = \frac{v}{t+v}\Sigma_0 + \frac{t}{t+v}\hat{\Sigma}_t,$$

  assuming the inverse Wishart prior $\Sigma_t \sim IW(\Sigma_0, v)$.

Bayesian forecast averaging



VOID

Computing the marginal likelihood

- Usually difficult to evaluate the integral

$$p(\mathbf{y}) = \int p(\mathbf{y}|\theta)p(\theta)d\theta = E_{p(\theta)}[p(\mathbf{y}|\theta)].$$

- Draw from the prior $\theta^{(1)}, ..., \theta^{(N)}$ and use the Monte Carlo estimate

$$\hat{p}(\mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N} p(\mathbf{y}|\theta^{(i)}).$$

  Unstable if the posterior is somewhat different from the prior.
- **Importance sampling**. Let $\theta^{(1)}, ..., \theta^{(N)}$ be iid draws from $g(\theta)$.

$$\int p(\mathbf{y}|\theta)p(\theta)d\theta = \int \frac{p(\mathbf{y}|\theta)p(\theta)}{g(\theta)}g(\theta)d\theta \approx N^{-1}\sum_{i=1}^{N} \frac{p(\mathbf{y}|\theta^{(i)})p(\theta^{(i)})}{g(\theta^{(i)})}$$

- **Modified Harmonic mean**: $g(\theta) = N(\tilde{\theta}, \tilde{\Sigma}) \cdot I_c(\theta)$, where $\tilde{\theta}$ and $\tilde{\Sigma}$ is the posterior mean and covariance matrix estimated from an MCMC chain, and $I_c(\theta) = 1$ if $(\theta - \tilde{\theta})'\tilde{\Sigma}^{-1}(\theta - \tilde{\theta}) \leq c$.

Approximate marginal likelihoods

- **The Laplace approximation**:

$$\ln \hat{p}(\mathbf{y}) = \ln p(\mathbf{y}|\hat{\theta}) + \ln p(\hat{\theta}) + \frac{1}{2}\ln|\Sigma| + \frac{p}{2}\ln(2\pi),$$

  where $\Sigma = -H^{-1}$ and p is the number of unrestricted parameters in the model.

- Note that $\hat{\theta}$ and H can be obtained with **numerical optimization** with BFGS update of Hessian.

- The **BIC approximation**

$$\ln \hat{p}(\mathbf{y}) = \ln p(\mathbf{y}|\hat{\theta}) + \ln p(\hat{\theta}) - \frac{p}{2}\ln n.$$

Posterior predictive analysis

- If $p(y|\theta)$ is a 'good' model, then the data actually observed should not differ 'too much' from simulated data from $p(y|\theta)$.

- Bayesian: simulate data from the **posterior predictive distribution** [9]:

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)d\theta.$$

- Difficult to compare y and y^{rep} because of dimensionality.
- Solution: compare **low-dimensional statistic** $T(y, \theta)$ to $T(y^{rep}, \theta)$.
- Evaluates the full probability model consisting of both the likelihood *and* prior distribution.

Posterior predictive analysis, cont.

- **Algorithm** for simulating from the posterior predictive density $p[T(y^{rep})|y]$:

1 Draw a $\theta^{(1)}$ from the posterior $p(\theta|y)$.

2 Simulate a data-replicate $y^{(1)}$ from $p(y^{rep}|\theta^{(1)})$.

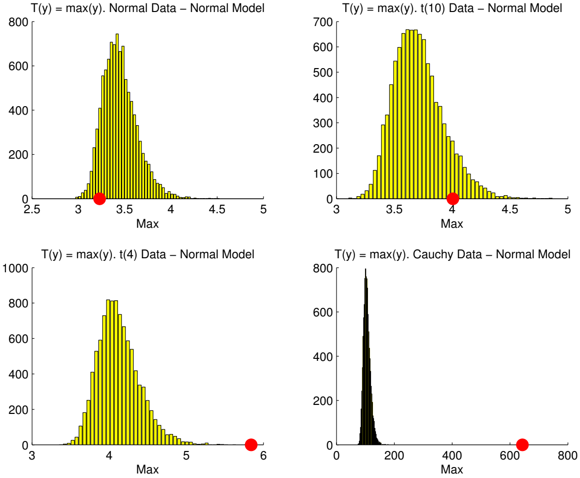
3 Compute $T(y^{(1)})$.

4 Repeat steps 1-3 a large number of times to obtain a sample from $T(y^{rep})$.

- We may now compare the observed statistic $T(y)$ with the distribution of $T(y^{rep})$.
- Posterior predictive p-value: $\Pr[T(y^{rep}) \geq T(y)]$
- Informal **graphical analysis**.

Posterior predictive analysis - Normal model, max statistic

- Model: $y_1, \ldots, y_n \overset{iid}{\sim} N(\mu, \sigma^2)$. $T(y) = \max_i |y_i|$.



Decision Theory

- Let θ 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 turns out to be θ gives **utility**

$$U(a, \theta)$$

- Utility table:

	θ_1	θ_2
a_1	$U(a_1, \theta_1)$	$U(a_1, \theta_2)$
a_2	$U(a_2, \theta_1)$	$U(a_2, \theta_2)$

- Example:

	Rainy	Sunny
Umbrella	50	70
No umbrella	0	100

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 decision

- Ad hoc decision rules:
  - *Minimax*. Choose the decision that minimizes the maximum loss.
  - *Minimax-regret* ... bla bla bla ...
- Bayesian **theory**: Just maximize the **posterior expected utility**:

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

  where $E_{p(\theta|y)}$ denotes the posterior expectation.
- Using simulated draws $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)}$ from $p(\theta|\mathbf{y})$:

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

- **Separation principle**:

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

- R. E. Kass and A. E. Raftery, "Bayes factors," *Journal of the american statistical association*, vol. 90, no. 430, pp. 773–795, 1995.
- M. Adolfson, S. Laséen, J. Lindé, and M. Villani, "Evaluating an estimated new keynesian small open economy model," *Journal of Economic Dynamics and Control*, vol. 32, no. 8, pp. 2690–2721, 2008.
- J. Corander and M. Villani, "Bayesian assessment of dimensionality in reduced rank regression," *Statistica Neerlandica*, vol. 58, no. 3, pp. 255–270, 2004.
- M. Villani, R. Kohn, and P. Giordani, "Regression density estimation using smooth adaptive gaussian mixtures," *Journal of Econometrics*, vol. 153, no. 2, pp. 155–173, 2009.
- M. Villani, R. Kohn, and D. J. Nott, "Generalized smooth finite mixtures," *Journal of Econometrics*, vol. 171, no. 2, pp. 121–133, 2012.

- M. Adolfson, J. Lindé, and M. Villani, "Forecasting performance of an open economy dsge model," *Econometric Reviews*, vol. 26, no. 2-4, pp. 289–328, 2007.
- R. L. Winkler, "Combining probability distributions from dependent information sources," *Management Science*, vol. 27, no. 4, pp. 479–488, 1981.
- M. Adolfson, M. Andersson, J. Linde, M. Villani, and A. Vredin, "Modern forecasting methods in action: Improving macroeconomic analyses at central banks," *International Journal of Central Banking*, vol. 3, no. 4, pp. 111–144.
- A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*, vol. 2.
  Taylor & Francis, 2014.