FUNDAMENTAL INFORMATION COMBINATION METHODS

INTRODUCTION

Purpose:

Describe and illustrate simple methods for combining information

Overview:

- **Classical Methods**
- **Basic Bayesian Methods**

RDMS EXAMPLE

GOAL: Estimate **R(t| q)** for motor component one (MC1).

R(t| q) = Pr (T $\geq$ t) is the reliability function, there are several choices.

θ = mean time to failure due to overheating of MC1

T = time to failure

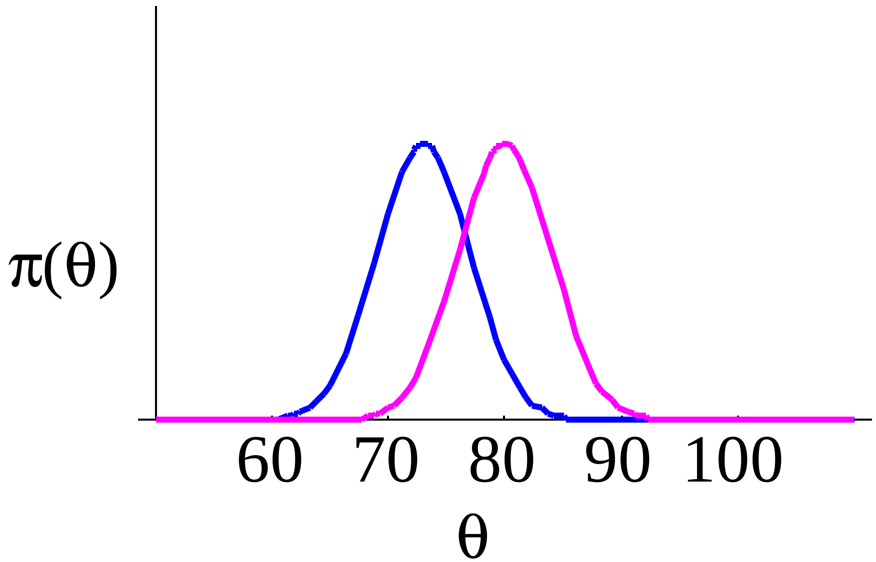
PROBLEM: Determine a value for q**.**

INFORMATION SOURCES FOR EXAMPLE

- 2 Experts
- 3 Computer Codes (similar system)
- 5 Sets of Data from Physical Experiments

EXPERT'S INFORMATION

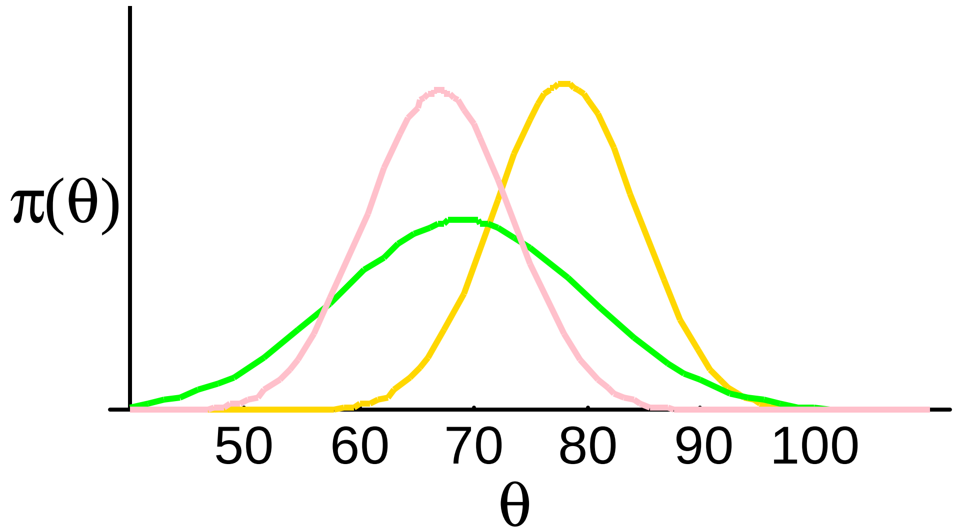
Suppose Jack and Jill are identified as experts due to their experience with MC1's use in previous systems. From these elicitations, distributions and point estimates for θ are obtained.



Jill:	mean = 80.0 standard deviation = 4.0
Jack:	mean = 73.0 standard deviation = 4.0

COMPUTER CODES

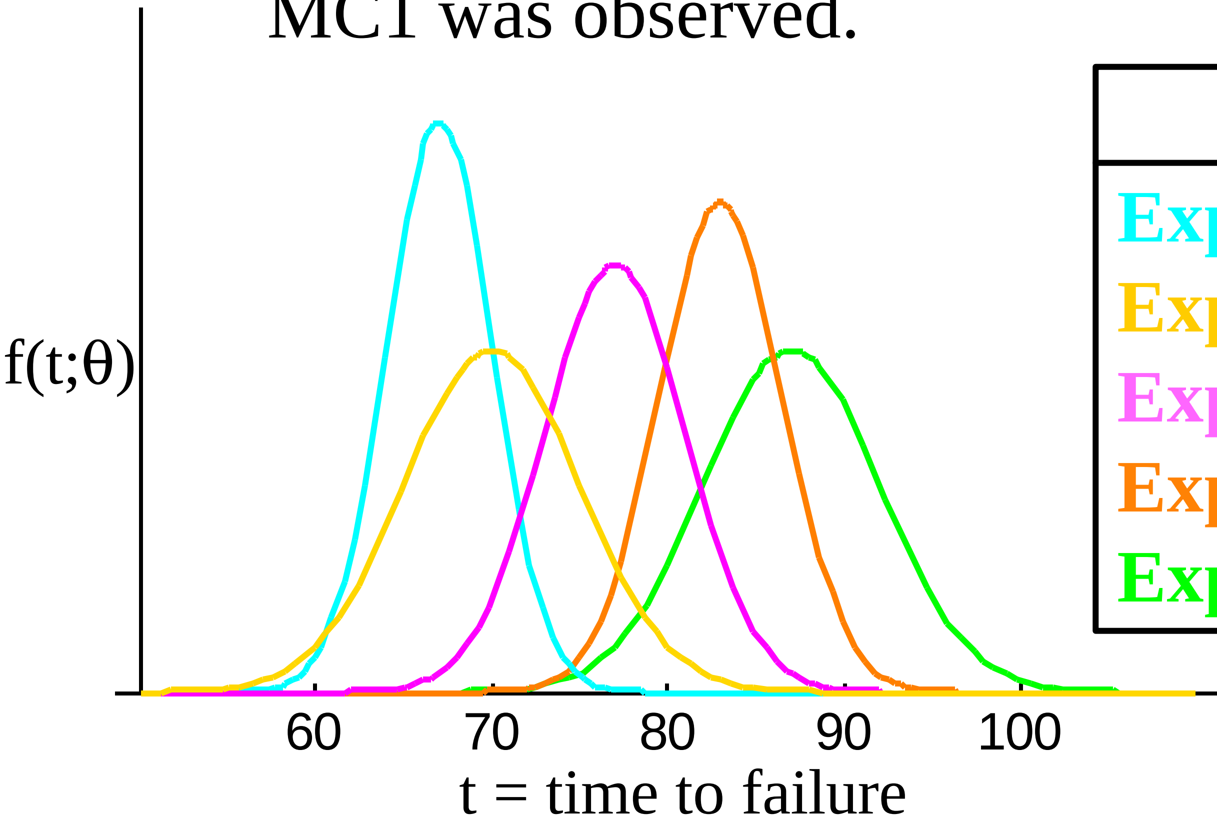
Down the hall in the computer lab, three computer models have been identified as being able to forecast distributions for θ.



Code 1: mean = 78.0	standard deviation = 6.3
Code 2: mean = 69.0	standard deviation = 10.8
Code 3: mean = 67.0	standard deviation = 6.5

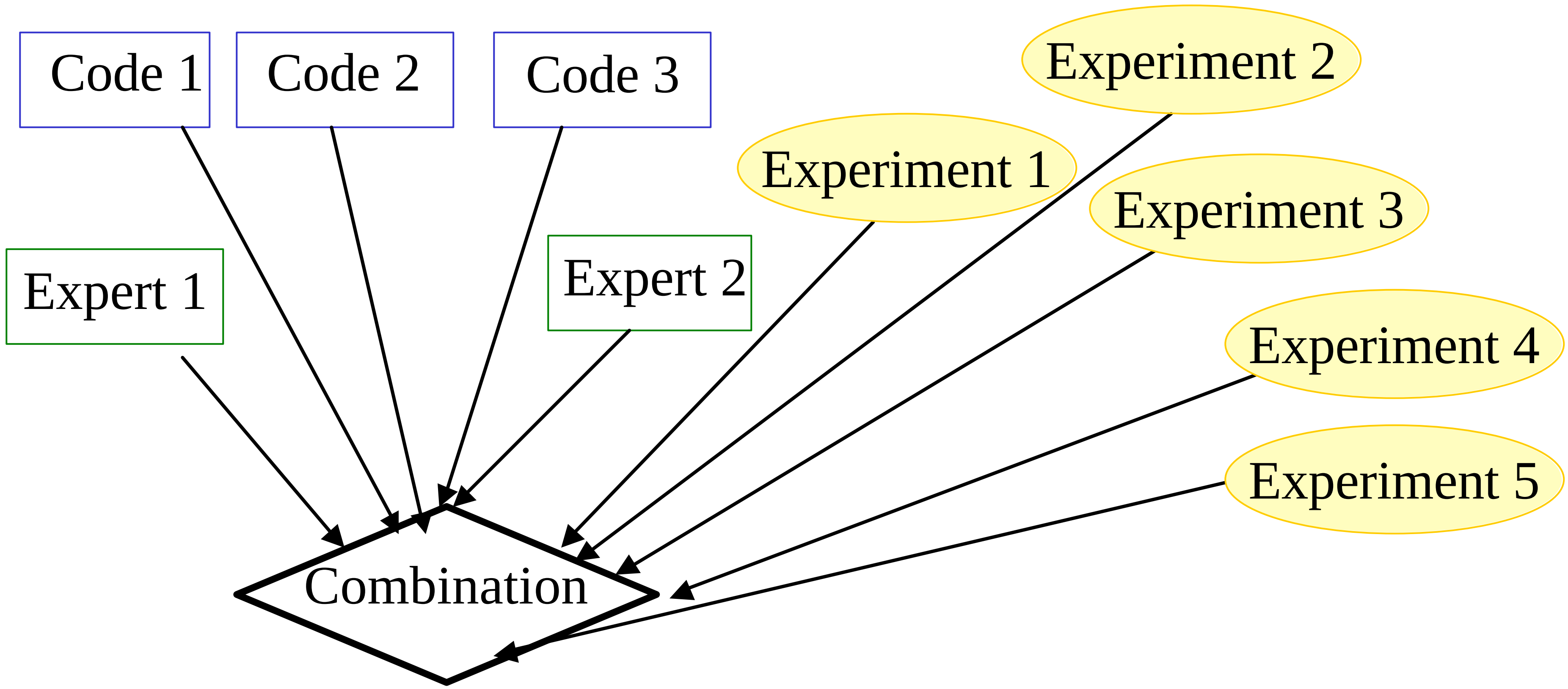
PHYSICAL EXPERIMENTS

In a lab across the street, physical experiments (heat stress) were performed on five different sets of motors. For each motor, the time to failure for MC1 was observed.



	mean	std
Exp1	87.0	5.0
Exp2	83.0	3.5
Exp3	67.0	3.0
Exp4	77.0	4.0
Exp5	70.0	5.0

INFORMATION SOURCE INTEGRATION



Final Integrated Estimate

APPROACHES FOR DETERMINING θ

- Classical Estimation
  - data are random
  - θ is fixed
  - the problem is to estimate θ
- Bayesian Prediction
  - data are fixed
  - θ is random
  - the problem is to use the distribution $\pi(\theta)$ to predict θ

These differences are subtle, but lead to two different approaches for determining θ

CLASSICAL ESTIMATION: BLUE

- E: Estimation
- L: Linear, a weighted average

$$\hat{\theta} = w_1\hat{\theta}_1 + w_2\hat{\theta}_2 + w_3\hat{\theta}_3 + \ldots$$

- U: Unbiased, correct on average, $\sum w_i = 1$
- B: Best, most precise, $\min_{w_i} \mathrm{Var}(\hat{\theta})$
  - w_i is inversely related to $\mathrm{Var}(\hat{\theta}_i)$
  - w_i, w_j are inversely related to $\mathrm{Correlation}(\hat{\theta}_i, \hat{\theta}_j)$

CLASSICAL ESTIMATION: EXPERT JUDGMENT

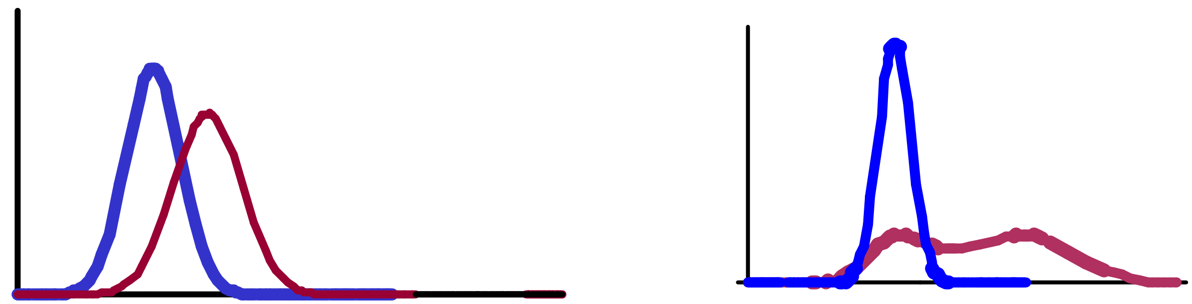
- The elicited information is taken as estimates of θ :
  - $\hat{\theta}_{\text{Jill}} = 80$ and $\hat{\theta}_{\text{Jack}} = 73$
  - STD($\hat{\theta}_{\text{Jill}}$) = 4 and STD($\hat{\theta}_{\text{Jack}}$) = 4
- An intuitive way to combine this information into a single estimate for θ is
  - $\hat{\theta} = .5(80) + .5(73) = 76.5$
  - STD($\hat{\theta}$) = sqrt($.5^2 * 4^2 + .5^2 * 4^2$) = 2.82
  - BLUE because the STDs are the same and the information is assumed independent.

COMPUTER CODES: SIMILAR SYSTEMS

- Similar system: a process distinctly different from the system under study (e.g., random variable $T \sim f(t;\theta)$), but expected to behave in a similar fashion
  - prototypes
  - components produced by the same design team
  - computer codes
- Assume the performance of the similar system is measured by $X \sim f(x;\delta)$

SIMILAR SYSTEMS

- What does it mean to be "similar"?
  - It does not mean that T and X are correlated.
- The distribution functions $f(t;\theta)$ and $f(x;\delta)$ are similar in form and location



- δ is treated as a surrogate for θ, with $\theta = \delta + \varepsilon$, where ε is random, with μ_ε and σ^2_ε , OR some other relationship between $f(t;\theta)$ and $f(x;\delta)$ must be assumed and modeled

COMPUTER CODES: SIMILAR SYSTEMS

- Computer code gives estimate of δ and $\mathrm{Var}(\hat{\delta}) + \sigma_{\varepsilon}^{2}$.
  This is the similar system information.

- Suppose there is no reason to believe δ is greater than or less than θ. This means $E(\varepsilon)=0$ and $\hat{\theta} = \hat{\delta}$.

- The variance estimate is

$$\mathrm{Var}(\hat{\theta}) = \widehat{\mathrm{Var}(\hat{\delta}) + \sigma^{2}_{\varepsilon}}.$$

- Now we are ready to combine the computer code information with the expert judgment data.

EXPERT JUDGEMENT + CODES

- We now have five $\hat{\theta}$'s and STD($\hat{\theta}$)'s.
- The BLUE for θ is a weighted average of the five with weights inversely proportional to the STDs.

$$\hat{\theta} = .34(80) + .34(73) + .14(78) + .05(69) + .13(67) = 75.12$$

$$\mathrm{Std}(\hat{\theta}) = 2.34$$

PHYSICAL EXPERIMENT DATA

- For a single experiment $\hat{\theta}$ and $\mathrm{Var}(\hat{\theta})$ are computed in a traditional fashion using maximum likelihood or method of moments estimation, e.g., $\hat{\theta} = \overline{T}$.
- If the experiments generated completely independent observations, the combined estimate would be obtained using weights that are a function of the individual variances (same as previous example).
- Let's suppose the experiments do not generate independent data. Now the weights for the BLUE for θ will depend both on the variances and the correlations between the experiments.

BLUE: EXPERT + CODE + EXPERIMENT

$$\hat{\Sigma}_{EJ} = \begin{bmatrix} 4^2 & 0 \\ & 4^2 \end{bmatrix} \qquad \hat{\Sigma}_{CODE} = \begin{bmatrix} 6.3^2 & 0 & 0 \\ & 10^2 & 0 \\ & & 6.5^2 \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{EJ} & \phi & \phi \\ & \hat{\Sigma}_{CODE} & \phi \\ & & \hat{\Sigma}_{EXP} \end{bmatrix}$$

$$\hat{\Sigma}_{EXP} = \begin{bmatrix} 5^2 & 5 \cdot 3.5 \cdot .33 & 5 \cdot 3 \cdot .19 & 5 \cdot 4 \cdot .37 & 5 \cdot 5 \cdot .37 \\ & 3.5^2 & 3.5 \cdot 3 \cdot .73 & 3.5 \cdot 4 \cdot .80 & 3.5 \cdot 5 \cdot .90 \\ & & 3^2 & 3 \cdot 4 \cdot .85 & 3 \cdot 5 \cdot .87 \\ & & & 4^2 & 4 \cdot 5 \cdot .85 \\ & & & & 5^2 \end{bmatrix}$$

$$\underline{\hat{\theta}}^T = (80, 73, 78, 69, 67, 87, 83, 67, 77, 70)$$

$$\underline{\hat{w}} = \left(\underline{1}^T \hat{\Sigma}^{-1} \underline{1}\right)^{-1} \underline{1}^T \hat{\Sigma}^{-1} = (.14, .14, .06, .02, .05, .04, .53, .65, -.07, -.55)$$

$$\hat{\theta} = \underline{\hat{w}}\,\underline{\hat{\theta}} = 77.06 \text{ and } \mathrm{Var}(\hat{\theta}) = \left(\underline{1}^T \hat{\Sigma}^{-1} \underline{1}\right)^{-1}, \ \mathrm{STD}(\hat{\theta}) = 1.49$$

CLASSICAL CRITIQUE

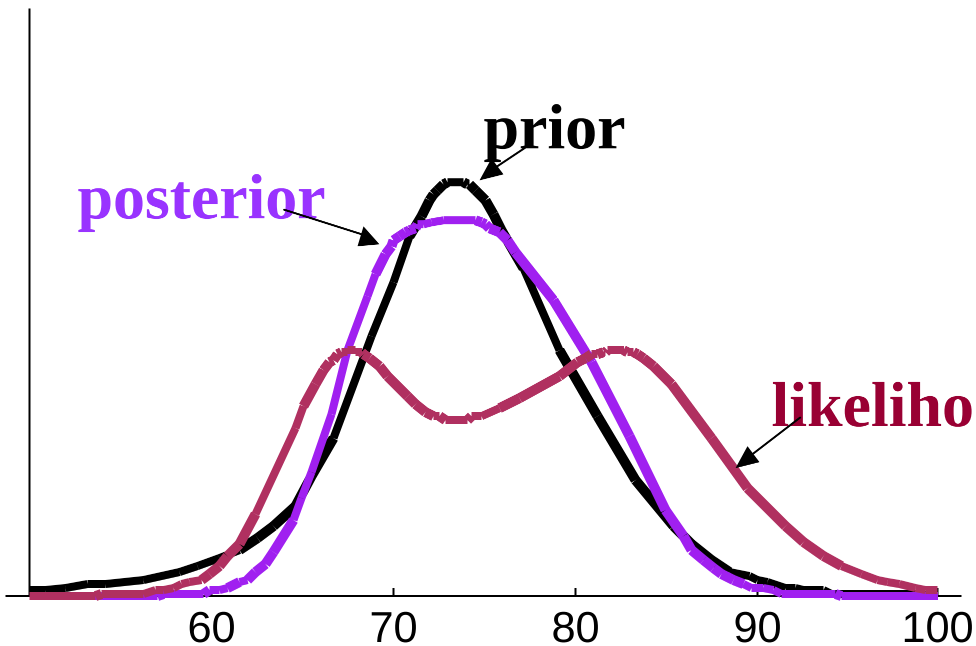
- Advantages
  - robust (distribution free)
  - computationally straightforward

- Disadvantages
  - sub-optimal use of information

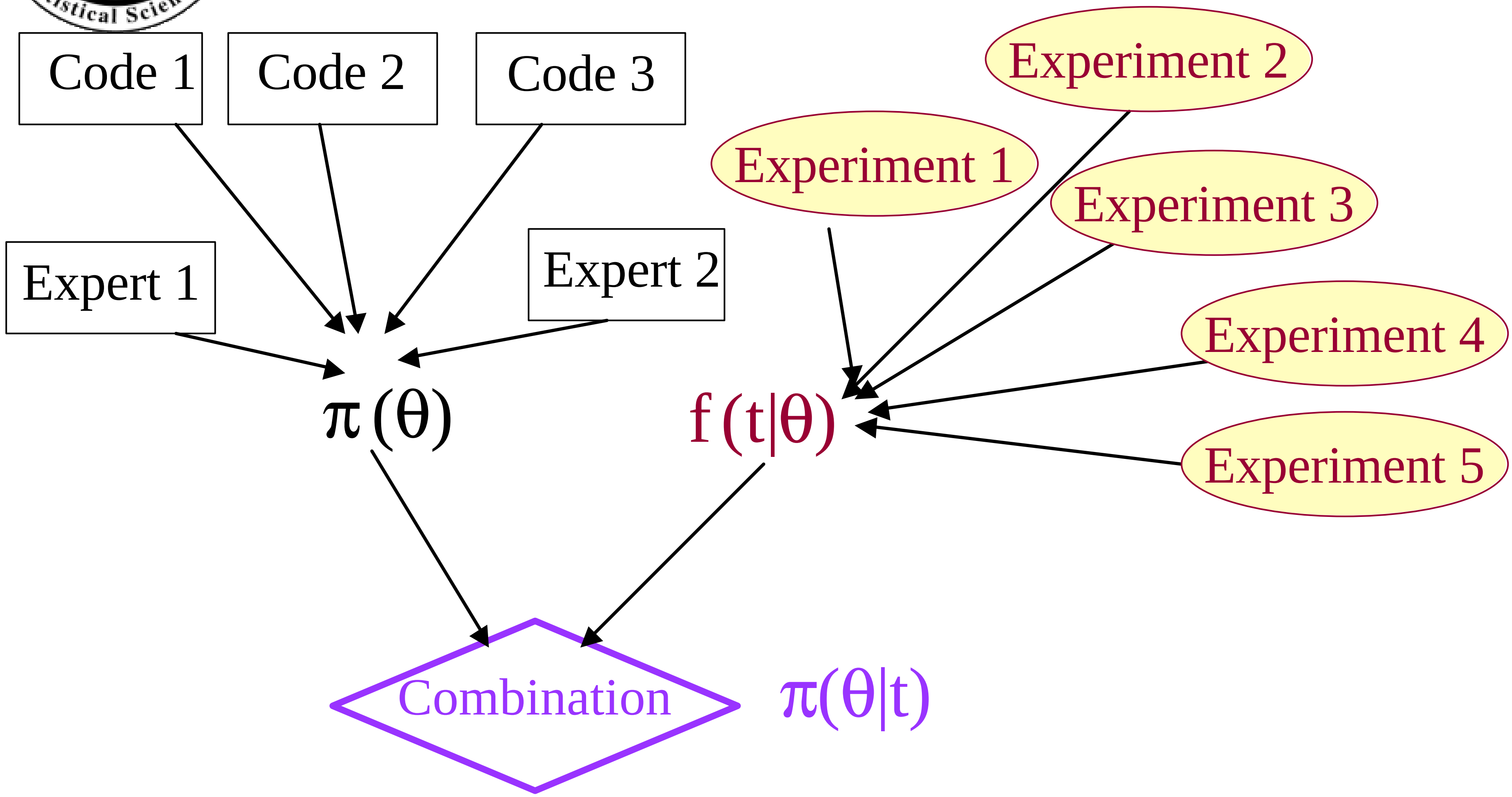
BAYESIAN PREDICTION

$$\pi(\theta|\text{data}) \propto f(\text{data}|\theta) * \pi(\theta)$$

Posterior ← **Likelihood** * **Prior**



INFORMATION SOURCE INTEGRATION



Final Integrated Distribution

FINDING THE PRIOR $\pi(\theta)$

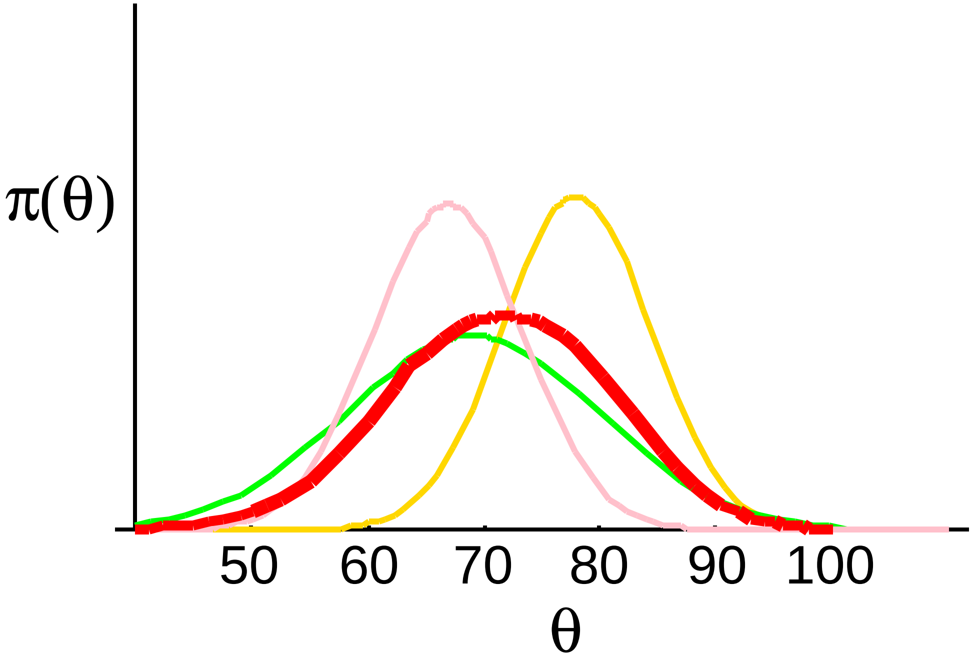
Use general mixture distribution weighting formula

$$\pi(\theta) = w_1 \cdot \pi_1(\theta) + w_2 \cdot \pi_2(\theta) + w_3 \cdot \pi_3(\theta) + \ldots.$$

Weighting Schemes

- Equal Weights
- Expert Supplied Weights
- Weights Based on Inverse Variance

CODE ESTIMATES



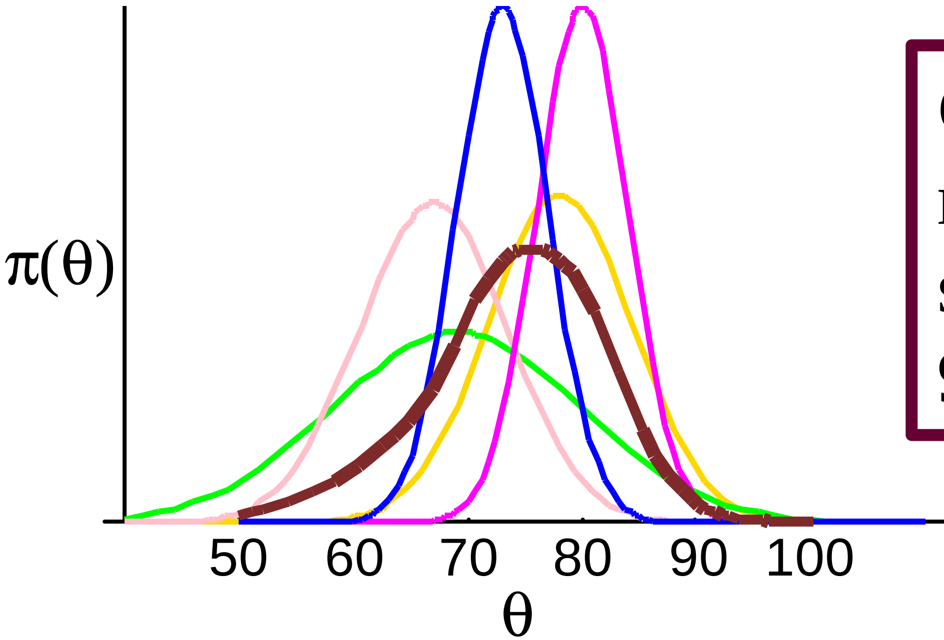
Equal weights combination

$1/3\ \pi_1(\theta) + 1/3\ \pi_2(\theta) + 1/3\ \pi_3(\theta)$

Combined mean=71.3

Combined standard deviation=9.4

CODE + EXPERTS' ESTIMATES



Combined Estimates:
mean=73.9
standard deviation=8.1
95% interval [55.7, 87.6]

Expert supplied weights combination:

$1/6\ \pi_1(\theta)+1/6\ \pi_2(\theta)+1/6\ \pi_3(\theta)+1/4\ \pi_4(\theta_4)+1/4\ \pi_5(\theta_5)$

$\{w_1 = w_2 = w_3 = 1/6\ ;\ w_{expert1} = w_{expert2} = 0.25\}$

ALTERNATE CODE + EXPERTS COMBINATION

Weights inversely proportional to variances and account for distances from overall mean

$\{w_1, w_2, w_3, w_4, w_5\}$

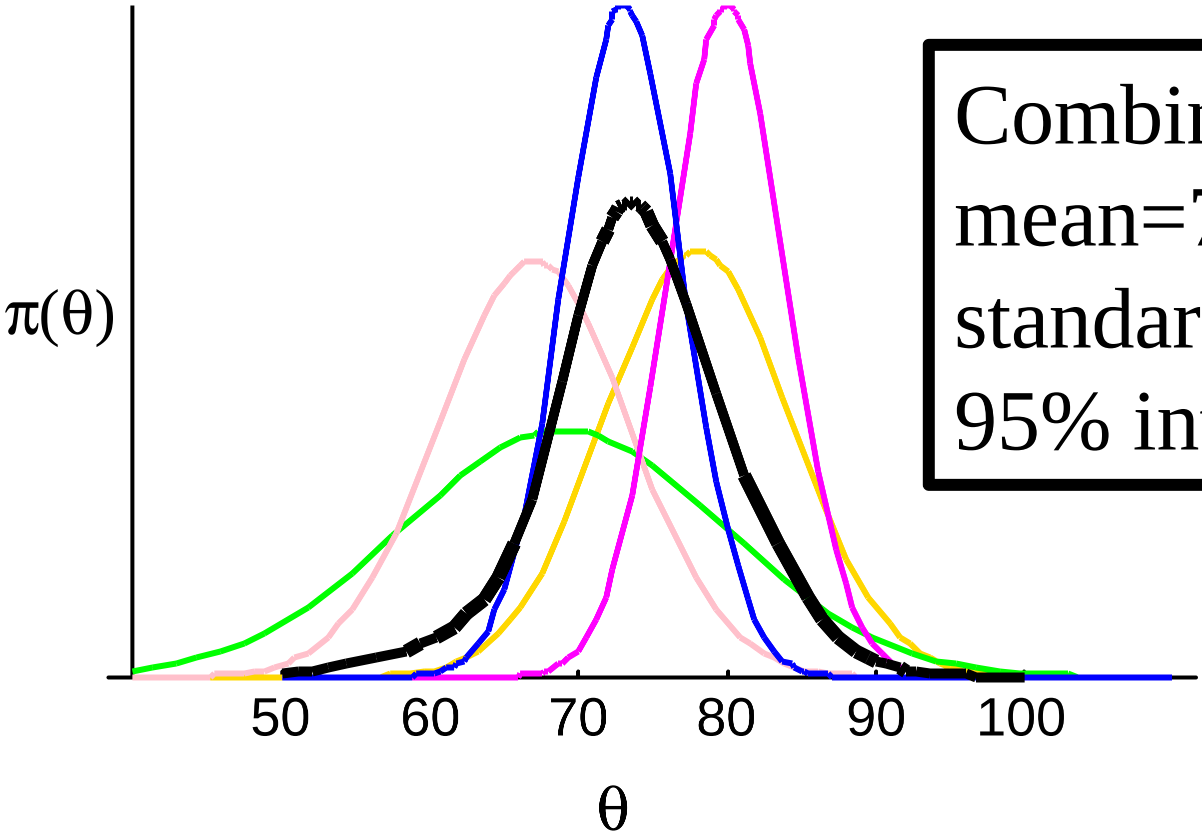
Experts Codes

For each information source i, $i = \boxed{1,2}\boxed{3,4,5}$, let mean = m_i and standard deviation = s_i. Then,

$$\overline{m} = \sum_{i=1}^{5} \frac{m_i}{5}, \ \text{IMS} = \sum_{i=1}^{5} 1/[(m_i - \overline{m})^2 - s_i^2],$$

$$w_i = \frac{1/[(m_i - \overline{m})^2 - s_i^2]}{\text{IMS}}, \ \text{and} \sum w_i = 1.$$

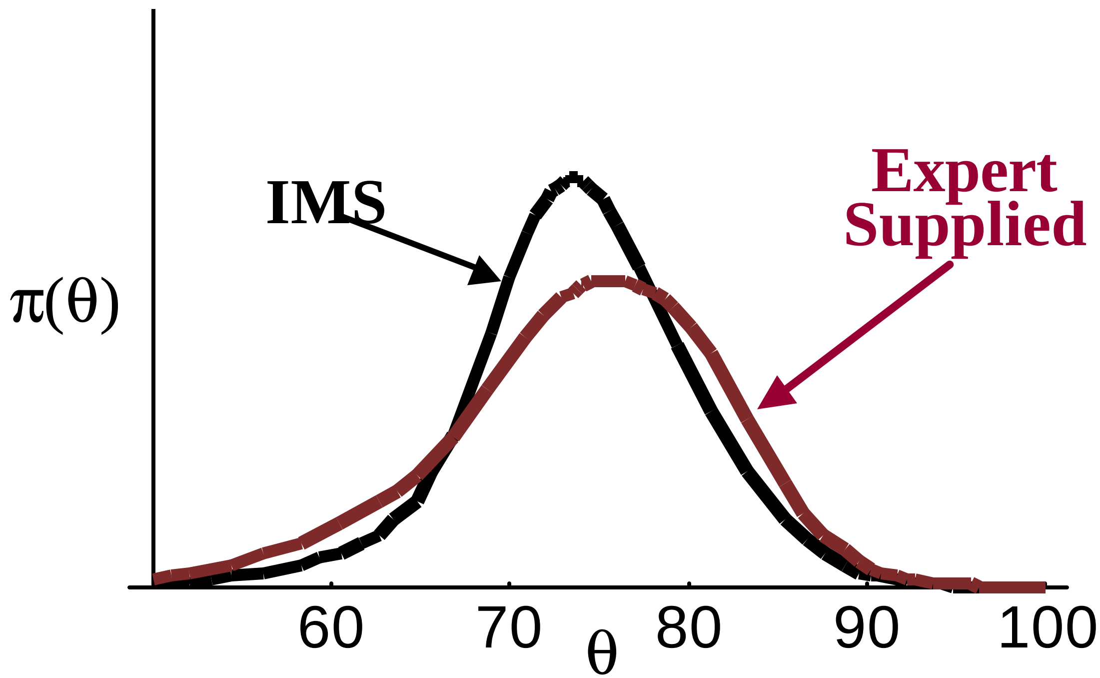
ALTERNATE CODE + EXPERTS COMBINATION



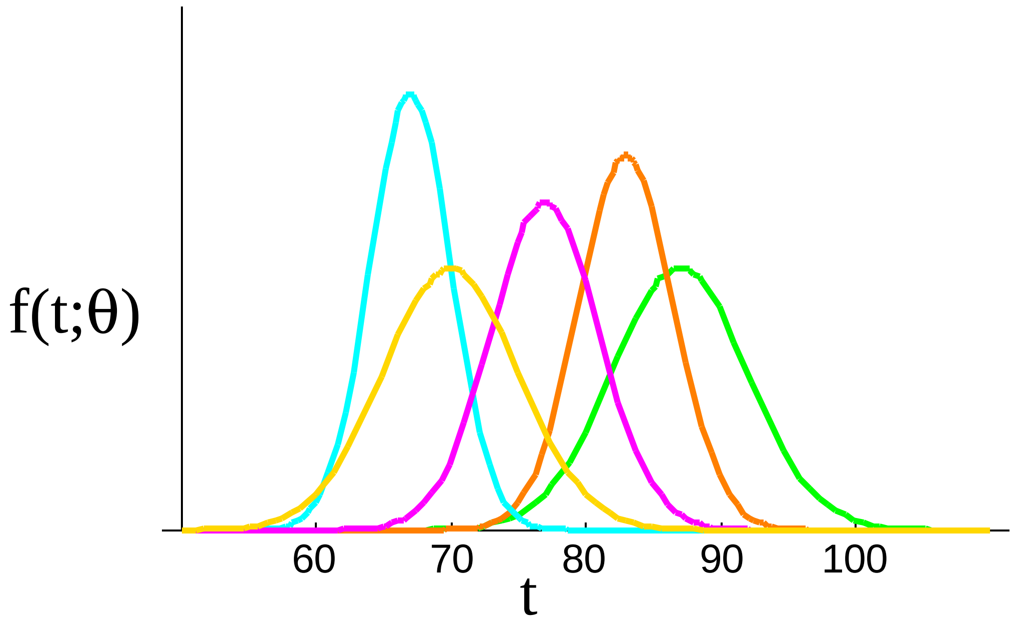
Combined Estimates:
mean=73.85
standard deviation=6.6
95% interval [59.4, 86.4]

IMS WEIGHTS VS EXPERT SUPPLIED

mean=73.85	mean=73.9
standard deviation=6.6	standard deviation=8.1
95% interval [59.4, 86.4]	95% interval [55.7, 87.6]



PHYSICAL EXPERIMENTS: $f(t|\theta)$

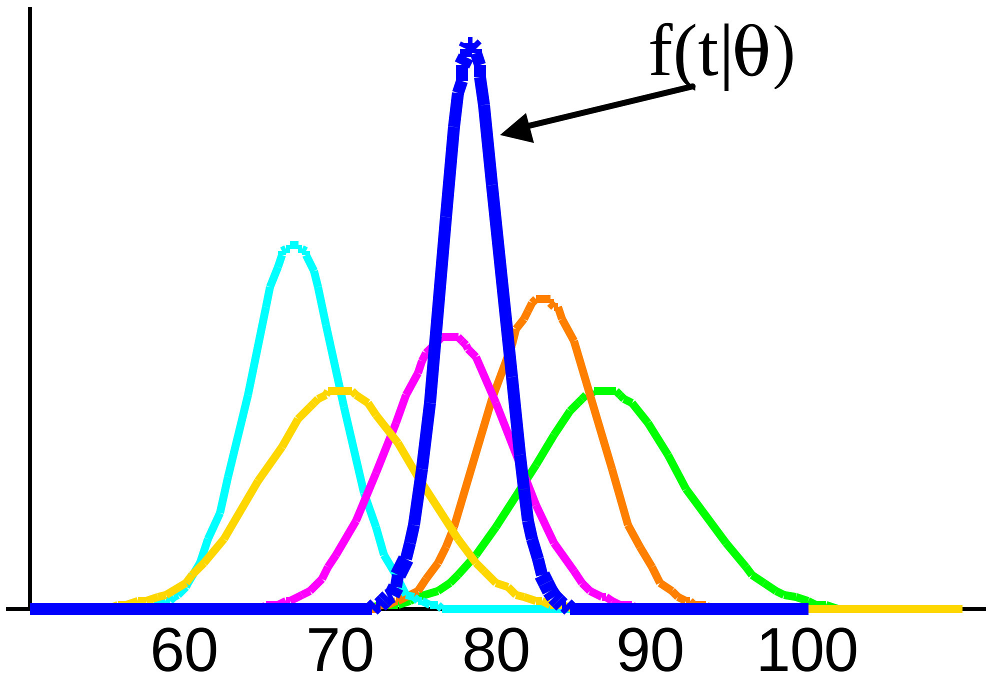


To build a likelihood, $f(t|\theta)$, from this data we need some assumptions:

- Across the five experiments, $T \sim (\theta,\Sigma)$
- Σ must be estimated or predicted via some prior

For this example we will assume $T \sim MVN(\theta,\hat{\Sigma}_{EXP})$

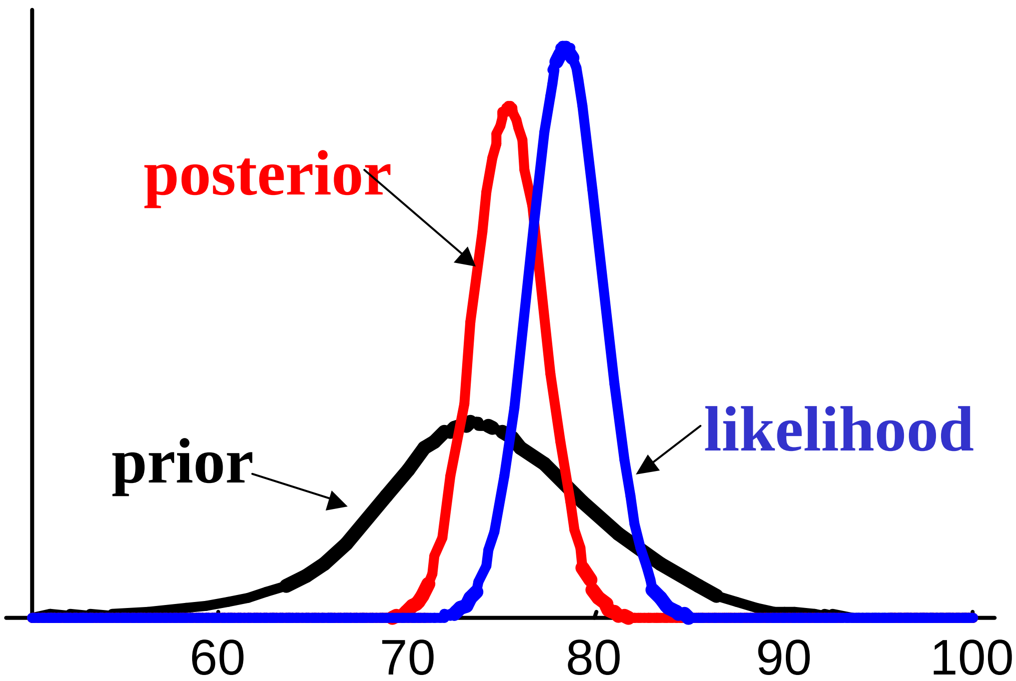
LIKELIHOOD RESULTS



mean=78.4 standard deviation=1.9

BAYESIAN COMBINATION

Example with IMS weighted Priors (codes + experts), and MVN Likelihood model (physical experiments).



BAYESIAN CRITIQUE

- Advantages
  - optimally combines information
  - naturally accommodates expert judgement and information updating

- Disadvantages
  - specifications of priors can be difficult (**sensitivity analyses recommended**)
  - computationally complex

CONCLUSIONS

- Bayesian and classical methods have more similarities than differences
- The methods should not produce wildly different results
- Computing both is a good check for
  - specification/computational errors
  - sensitivities
- Weights are selected via theory or elicited from experts --- theory is not w/o assumptions
- In practice, we would also put distributions on the weights