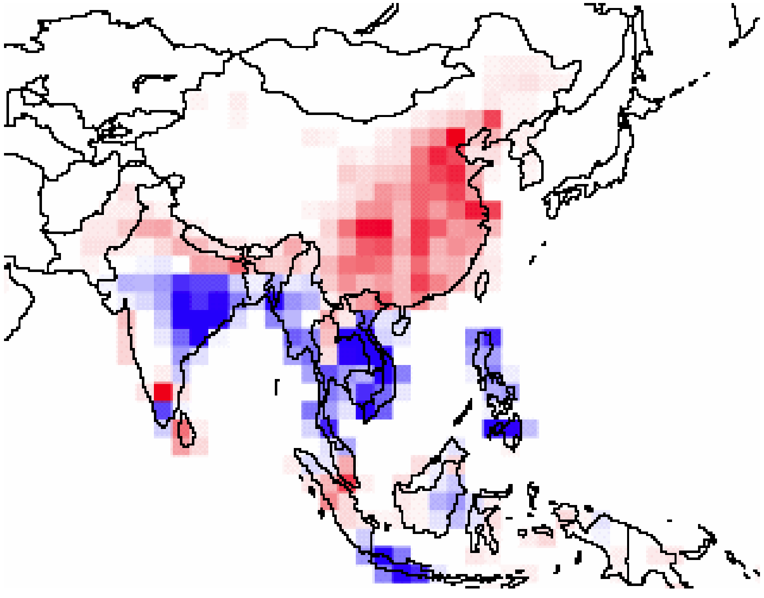
More Information! Conditional Probability and Inverse Methods

Peter Zoogman
Harvard Atmospheres Journal Club
April 12, 2012

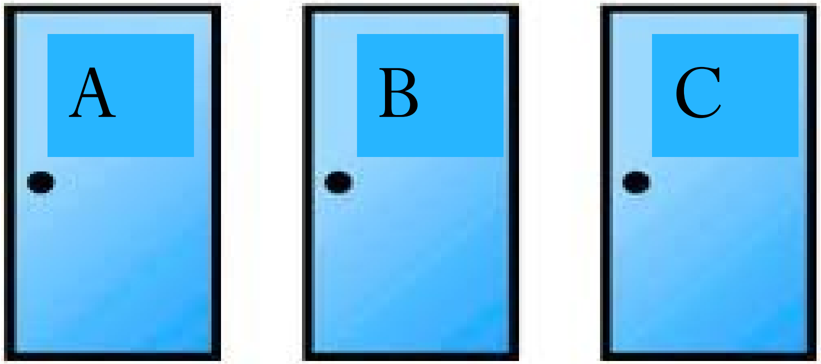

Bayesian Analysis – not very intuitive!

- Humans are bad at combining a priori information with new observations
- Conditional Probability examples
- 2-box inversion to constrain emissions
- Arellano et al paper on application of these methods

The Monty Hall Problem

You're given the choice of three doors:

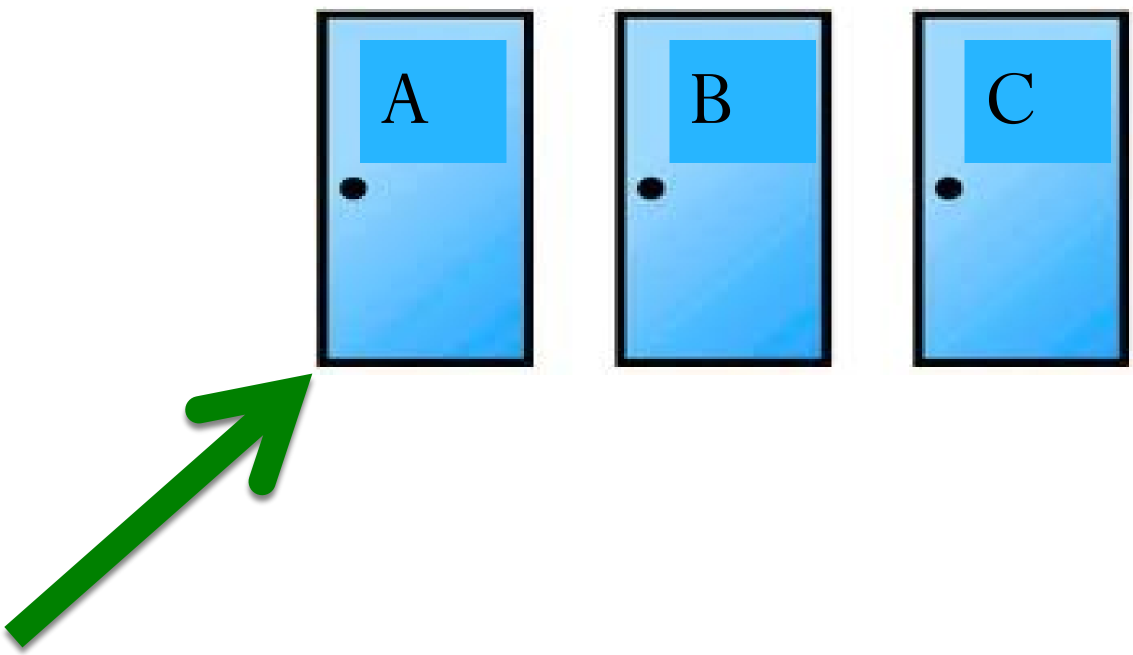
Behind one door is a car; behind the others, goats. You pick a door, say door A [but the door is not opened], and the host, who knows what's behind the doors, opens another door, say door B, which has a goat. He then says to you, "Do you want to pick door C?" Is it to your advantage to switch your choice?



The Monty Hall Problem

You're given the choice of three doors:

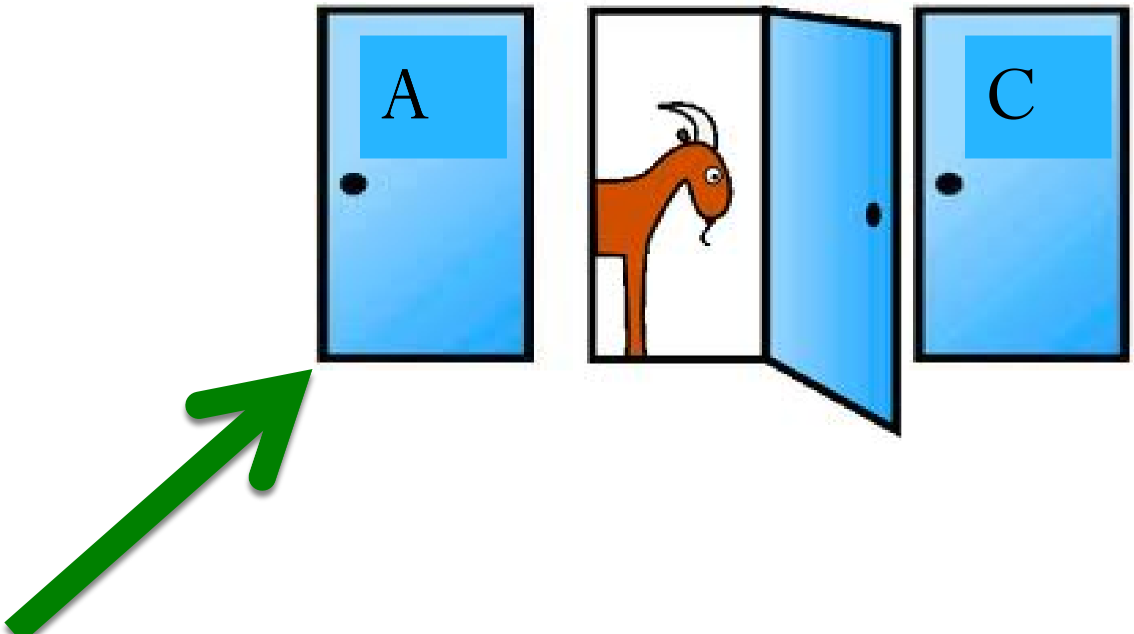
Behind one door is a car; behind the others, goats. You pick a door, say door A [but the door is not opened], and the host, who knows what's behind the doors, opens another door, say door B, which has a goat. He then says to you, "Do you want to pick door C?" Is it to your advantage to switch your choice?



The Monty Hall Problem

You're given the choice of three doors:

Behind one door is a car; behind the others, goats. You pick a door, say door A [but the door is not opened], and the host, who knows what's behind the doors, opens another door, say door B, which has a goat. He then says to you, "Do you want to pick door C?" Is it to your advantage to switch your choice?



Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(\text{car is behind A}|\text{door B opened}) = \frac{P(\text{door B opened}|\text{car is behind A})P(\text{car is behind A})}{P(\text{door B opened})}$$

$$P(\text{door B opened}|\text{car is behind A}) = 0.5$$

$$P(\text{car is behind A}|\text{door B opened}) = \frac{0.5 * 0.33}{0.5} = 0.33$$

Why is door C different?

Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(\text{car is behind A}|\text{door B opened}) = \frac{P(\text{door B opened}|\text{car is behind A})P(\text{car is behind A})}{P(\text{door B opened})}$$

$$P(\text{door B opened}|\text{car is behind A}) = 0.5$$

$$P(\text{car is behind A}|\text{door B opened}) = \frac{0.5 * 0.33}{0.5} = 0.33$$

Why is door C different?

$$P(\text{door B opened}|\text{car is behind C}) = 1$$

Updating an Initial Guess

- A cab was involved in a hit-and-run accident at night
  - 85% of the cabs in the city are Green and 15% are Blue
  - A witness identified the cab as Blue. Tests show that this witness correctly identifies the color of the cab 80% of the time

What is the probability the cab was involved in the accident was Blue?

Set-up from Thinking, Fast and Slow (Kahneman)

Combine observation with a priori information

$$P(\text{car is Blue} \mid \text{witnessed Blue}) = \frac{P(\text{witnessed Blue} \mid \text{car is Blue})P(\text{car is blue})}{P(\text{witnessed Blue})}$$

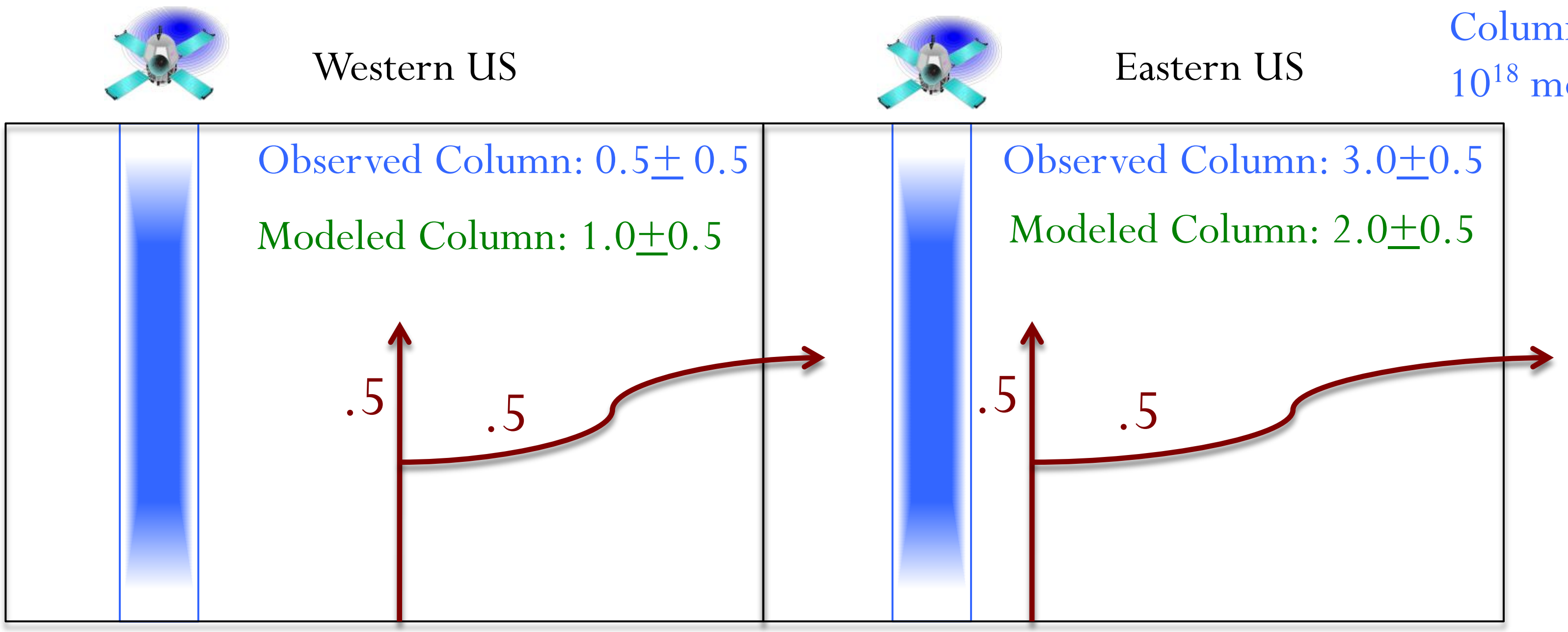
$$P(\text{car is blue}) = 0.15$$

$$P(\text{witnessed Blue} \mid \text{car is Blue}) = 0.8$$

$$P(\text{witnessed Blue}) = 0.8 * 0.15 + 0.2 * 0.85 = 0.29$$

$$P(\text{car is Blue} \mid \text{witnessed Blue}) = \frac{0.8 * 0.15}{0.29} = 0.41$$

Constraining Emissions



Column Units: 10^{18} molec/cm^2

Emissions = 1 ± 1

Emissions = 1 ± 1

Emissions units 100 Tg/yr

Modeled Column ~ Western Emissions

Modeled Column ~ Western Emissions + Eastern Emissions

Bayesian Inversion

$$P(X(\text{emissions})\,|\,Y(\text{observations})) = \frac{P(Y\,|\,X)P(X)}{P(Y)}$$

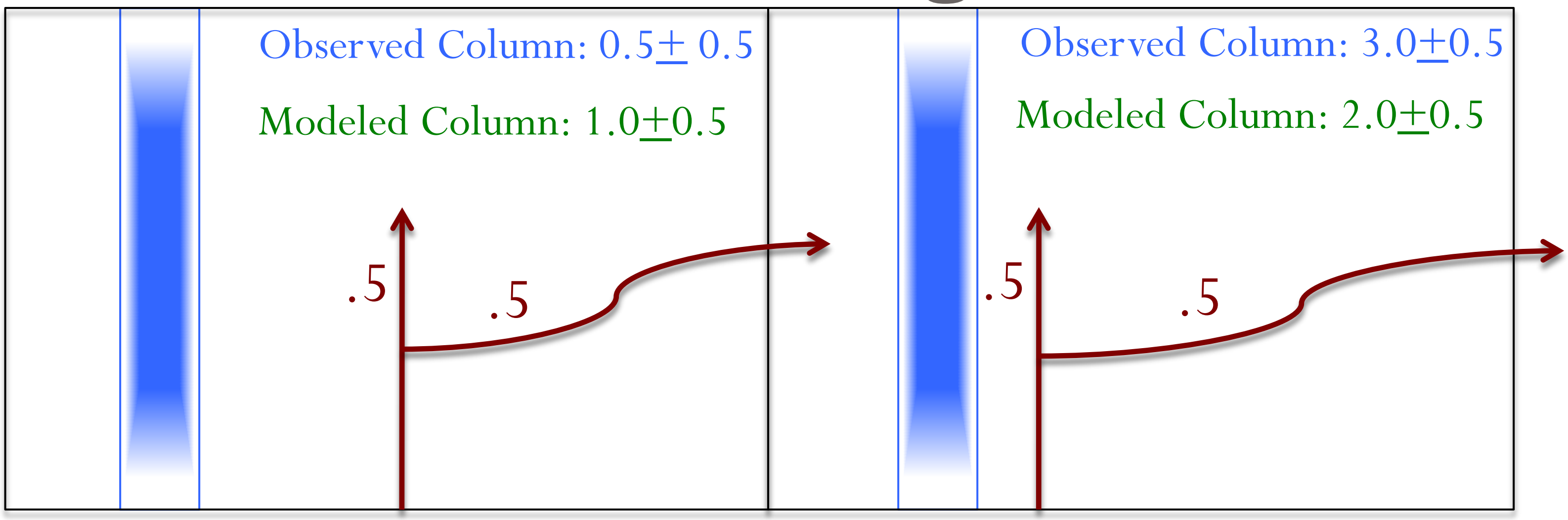
$$\mathbf{K}^{\mathrm{T}}\mathbf{S}_{\varepsilon}^{-1}(\mathbf{K}\hat{\mathbf{x}}-\mathbf{y})+\mathbf{S}_{a}^{-1}(\hat{\mathbf{x}}-\mathbf{x}_{\mathbf{a}})=0$$

$$\mathbf{y}=\mathbf{K}\mathbf{x}=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\begin{pmatrix}x_1\\ x_1+x_2\end{pmatrix}$$

$$\hat{\mathbf{x}}=\mathbf{x}_{\mathbf{a}}+\mathbf{G}(\mathbf{y}-\mathbf{K}\hat{\mathbf{x}}_{a})$$

$$\mathbf{G}=\mathbf{S}_{a}\mathbf{K}^{\mathrm{T}}\left(\mathbf{K}\;\mathbf{S}_{a}\mathbf{K}^{\mathrm{T}}+\mathbf{S}_{\varepsilon}\right)^{-1}$$

Constraining Emissions



Emissions = 1 ± 1 Emissions = 1 ± 1

$$\mathbf{S}_a = \begin{pmatrix} 1^2 & 0 \\ 0 & 1^2 \end{pmatrix}$$

$$\mathbf{S}_\varepsilon = \begin{pmatrix} 2*0.5^2 & 0 \\ 0 & 2*0.5^2 \end{pmatrix}$$

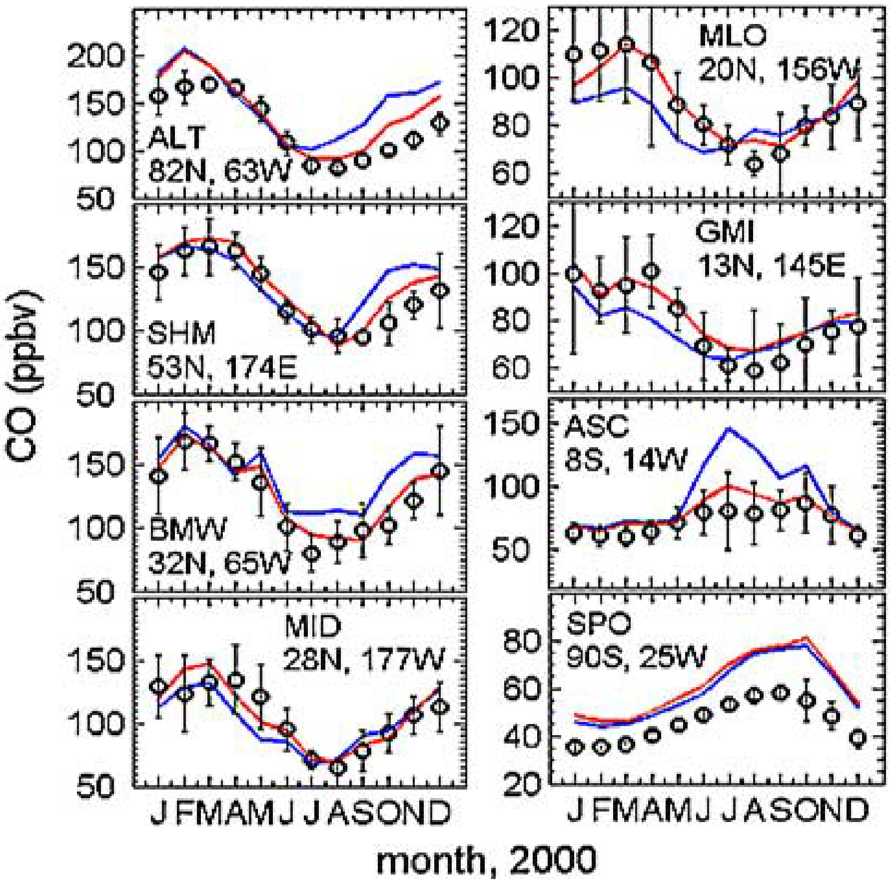
$$\mathbf{G} = \mathbf{S}_a\mathbf{K}^T\left(\mathbf{K}\ \mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\varepsilon\right)^{-1} = \begin{pmatrix} .54 & .18 \\ -.36 & .54 \end{pmatrix}$$

$$(\mathbf{y} - \mathbf{K}\hat{\mathbf{x}}_a) = \begin{pmatrix} .5 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -.5 \\ 1 \end{pmatrix}$$

$$\hat{\mathbf{x}} = \mathbf{x}_\mathbf{a} + \mathbf{G}(\mathbf{y} - \mathbf{K}\hat{\mathbf{x}}_a) = \begin{pmatrix} .91 \\ 1.72 \end{pmatrix}$$

Estimating CO Sources

- Why CO?
  - Strong link between free troposphere and regional emissions
  - Relatively easy to observe from space
- Important to compare results using optimized sources to independent data sets
- CO can be used to optimize emissions/concentrations of other species
  - CO_2, ozone



[Arellano et al 2004]