Data Assimilation
- A03. A Toy Model -

Shunji Kotsuki

Center for Environmental Remote Sensing / Institute of Advanced Academic Research

(shunji.kotsuki@chiba-u.jp)

CHIBA UNIVERSITY

Environmental Prediction Science Laboratory

DA Lectures A (Basic Course)

- (1) Introduction and NWP
- (2) Deterministic Chaos and Lorenz-96 model
- (3) A toy model and Bayesian estimation
- (4) Kalman Filter (KF)
- (5) 3D Variational Method (3DVAR)
- (6) Ensemble Kalman Filter (PO method)
- (7) Serial Ens. Square Root Filter (Serial EnSRF)
- (8) Local Ens. Transform Kalman Filter (LETKF)
- (9) Innovation Statistics
- (10) Adaptive Inflations
- (11) 4D Variational Method (4DVAR)

Today's Goal

- To understand minimum variance estimation
- To understand maximum likelihood estimation
- To understand assumptions in these estimations
- To understand Bayesian estimation

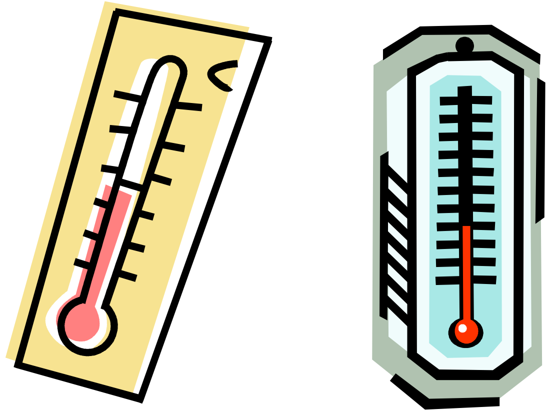
Data Assimilation
- a toy model -

Two Major Streams of DA

- **Minimum variance estimation**
  - Kalman filter (KF)
  - ensemble Kalman filter (EnKF)

- **Maximum likelihood estimation**
  - 3D variational (3DVAR)
  - 4D variational (4DVAR)
  - particle filter (PF)

A simple example: two thermometers



x_1=20℃
(less reliable)

x_2= 30℃
(more reliable)

Minimum Variance Estimation (最小分散推定)

Minimum Variance Estimation

forecast $x_1 = x^{tru} + \varepsilon_1$

observation $x_2 = x^{tru} + \varepsilon_2$

x^{tru}: truth

ε : random error

< · >: *expectation*

Assumption (1) : unbiased error

$$\langle x_1 \rangle = \langle x_2 \rangle = x^{tru} \quad \Leftrightarrow \quad \langle \varepsilon_1 \rangle = \langle \varepsilon_2 \rangle = 0$$

Assumption (2) : uncorrelated error

$$\langle \varepsilon_1 \cdot \varepsilon_2 \rangle = 0$$

Minimum Variance Estimation

forecast $x_1 = x^{tru} + \varepsilon_1$

observation $x_2 = x^{tru} + \varepsilon_2$

(1) unbias $\langle x_1 \rangle = \langle x_2 \rangle = x^{tru}$

(2) uncorr. $\langle \varepsilon_1 \varepsilon_2 \rangle = 0$

$x^a = \alpha x_1 + (1 - \alpha) x_2$ & minimize variance of analysis (a)

$$(\sigma^a)^2 = \langle (x^a - x^{tru})^2 \rangle = \langle (\alpha(x_1 - x^{tru}) + (1-\alpha)(x_2 - x^{tru}))^2 \rangle$$

$$= \alpha^2 \langle \varepsilon_1^2 \rangle + 2\alpha(1-\alpha)\langle \varepsilon_1 \varepsilon_2 \rangle + (1-\alpha)^2 \langle \varepsilon_2^2 \rangle$$

$$= \alpha^2 \sigma_1^2 + (1-\alpha)^2 \sigma_2^2$$

definition of variance

$$V(x) = E\left(\left(x - E(x)\right)^2\right)$$

$$\sigma^2 = \langle \varepsilon \cdot \varepsilon \rangle$$

σ : standard deviation

σ^2 : variance

Minimum Variance Estimation

Environmental Prediction Science Laboratory

forecast $x_1 = x^{tru} + \varepsilon_1$

observation $x_2 = x^{tru} + \varepsilon_2$

(1) unbias $\langle x_1 \rangle = \langle x_2 \rangle = x^{tru}$

(2) uncorr. $\langle \varepsilon_1 \varepsilon_2 \rangle = 0$

$$x^a = \alpha x_1 + (1-\alpha) x_2$$

$$(\sigma^a)^2 = \alpha^2 \sigma_1^2 + (1-\alpha)^2 \sigma_2^2$$

$$\frac{\partial (\sigma^a)^2}{\partial \alpha} = 2\alpha\sigma_1^2 - 2(1-\alpha)\sigma_2^2 = 0$$

$$\Leftrightarrow \alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

weighted average by variance (σ^2; =accuracy)

$$x^a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2$$

$$= \underbrace{x_1}_{\text{first guess}} + \underbrace{\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(x_2 - x_1)}_{\text{increment}}$$

Minimum Variance Estimation

Kalman filter

Predicted state estimate	$\mathbf{x}_t^b = M(\mathbf{x}_{t-1}^a)$
Predicted estimate covariance	$\mathbf{P}_t^b = \mathbf{M}\mathbf{P}_{t-1}^a\mathbf{M}^T + \mathbf{Q}$
Optimal Kalman gain	$\mathbf{K}_t = \mathbf{P}_t^b\mathbf{H}^T\left[\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R}\right]^{-1}$
Update estimate covariance	$\mathbf{P}_t^a = [\mathbf{I} - \mathbf{K}_t\mathbf{H}]\mathbf{P}_t^b$
Update state estimate	$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}_t(\mathbf{y}_t^o - H(\mathbf{x}_t^b))$

$M(\,)$: nonlinear model

M: Tangent Linear model

$H(\,)$: nonlinear obs. operator

H: Jacobian of $H()$

P: error covariance

Q: model error covariance

K: Kalman gain

R: obs. error covariance

b: background
o: observation
a: analysis

analysis equation

$$x^a = \underbrace{x_1}_{\text{first guess}} + \underbrace{\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(x_2 - x_1)}_{\text{increment}}$$

Maximum Likelihood Estimation
(最尤推定)

Maximum Likelihood Estimation

forecast $x_1 = x^{tru} + \varepsilon_1$

observation $x_2 = x^{tru} + \varepsilon_2$

(1) unbias $\langle x_1 \rangle = \langle x_2 \rangle = x^{tru}$

(2) uncorr. $\langle \varepsilon_1 \varepsilon_2 \rangle = 0$

Likelihood

Prior (uniform, i.e., no prior info)

$$p(x|x_{1,2}) = \frac{p(x_{1,2}|x)\,p(x)}{p(x_{1,2})}$$

Posterior

constant (since they are given)

Bayesian Estimates

$$maximize \quad p(x|x_{1,2}) \Leftrightarrow maximize \quad p(x_{1,2}|x)$$

$$\Leftrightarrow maximize \quad p(x_1|x) \cdot p(x_2|x)$$

to maximize likelihood

Maximum Likelihood Estimation

forecast $x_1 = x^{tru} + \varepsilon_1$

observation $x_2 = x^{tru} + \varepsilon_2$

(1) unbias $\langle x_1 \rangle = \langle x_2 \rangle = x^{tru}$

(2) uncorr. $\langle \varepsilon_1 \varepsilon_2 \rangle = 0$

$$maximize \quad p(x_1|x) \cdot p(x_2|x)$$

Suppose x_1 & x_2 follow Gaussian PDF $N(x, \sigma)$

$$p(x_i|x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x_i - x)^2}{2\sigma_i^2}\right]$$

$$maximize \quad p(x_1|x) \cdot p(x_2|x)$$

$$\Leftrightarrow maximize \quad \frac{1}{\sqrt{2\pi\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{(x_1 - x)^2}{2\sigma_1^2} - \frac{(x_2 - x)^2}{2\sigma_2^2}\right]$$

$$\Leftrightarrow minimize \quad J(x) = \frac{(x_1 - x)^2}{\sigma_1^2} + \frac{(x_2 - x)^2}{\sigma_2^2}$$

Maximum Likelihood Estimation

forecast $x_1 = x^{tru} + \varepsilon_1$ (1) unbias $\langle x_1 \rangle = \langle x_2 \rangle = x^{tru}$

observation $x_2 = x^{tru} + \varepsilon_2$ (2) uncorr. $\langle \varepsilon_1 \varepsilon_2 \rangle = 0$

$$minimize \quad J(x) = \frac{(x_1 - x)^2}{\sigma_1^2} + \frac{(x_2 - x)^2}{\sigma_2^2}$$

$$\frac{\partial J}{\partial x} = -2\frac{(x_1 - x)}{\sigma_1^2} - 2\frac{(x_2 - x)}{\sigma_2^2} = 0$$

analysis of maximum likelihood estimates

$$x^a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2$$

Summary

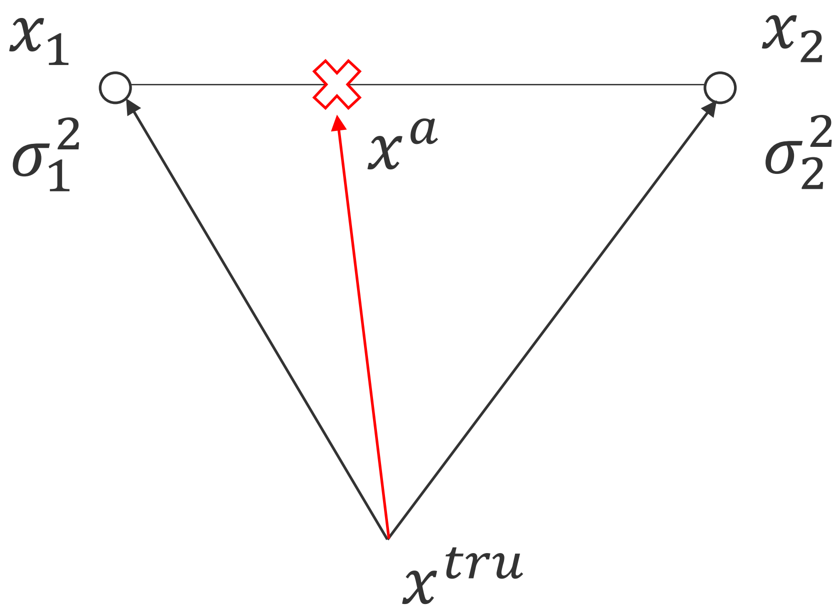
weighted average by variance (σ^2; =accuracy)

$$x^a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2$$

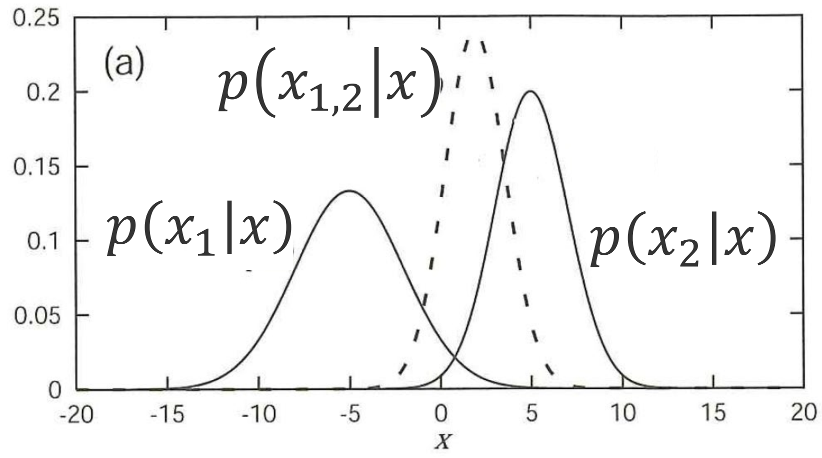
(1) unbias $\langle x_1 \rangle = \langle x_2 \rangle = x^{tru}$

(2) uncorr. $\langle \varepsilon_1 \varepsilon_2 \rangle = 0$

minimum variance estimates



maximum likelihood estimates



(3) Gaussian error PDF

Summary

Extension to multi-dims problems

The two thermometers' example

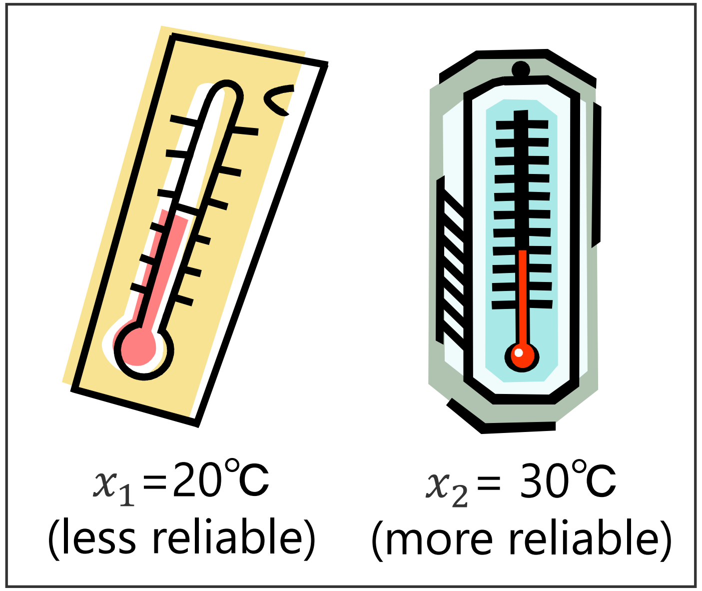
$$p_A(T) \propto \exp\left[-\frac{(T-T_A)^2}{2\sigma_A^2}\right]$$

$$p_B(T) \propto \exp\left[-\frac{(T-T_B)^2}{2\sigma_B^2}\right]$$

$$T^* = \frac{\sigma_B^2 T_A + \sigma_A^2 T_B}{\sigma_A^2 + \sigma_B^2}$$

$$= T_A + \frac{\sigma_A^2 T_B}{\sigma_A^2 + \sigma_B^2}(T_B - T_A)$$

weighted average



A simple example: two thermometers

Extension to multi-dims problems

The two thermometers' example

$$p_A(T) \propto \exp\left[-\frac{(T-T_A)^2}{2\sigma_A^2}\right]$$

$$p_B(T) \propto \exp\left[-\frac{(T-T_B)^2}{2\sigma_B^2}\right]$$

$$T^* = \frac{\sigma_B^2 T_A + \sigma_A^2 T_B}{\sigma_A^2 + \sigma_B^2}$$

$$= T_A + \frac{\sigma_A^2 T_B}{\sigma_A^2 + \sigma_B^2}(T_B - T_A)$$

weighted average

?

Uncertainty (reliability) of model forecasts

Uncertainty (reliability) of observations

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{P}_t^b\mathbf{H}^T\left[\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R}\right]^{-1}\left(\mathbf{y}_t^o - H(\mathbf{x}_t^b)\right)$$

x: state
y: observation

P: state error covariance
R: obs. error covariance
H: obs. operator

b: background
a: analysis
o: observation

Kalman Filter

Prediction (state)

$$\mathbf{x}_t^b = M(\mathbf{x}_{t-1}^a)$$

Prediction (error covariance)

$$\mathbf{P}_t^b = \mathbf{M}\mathbf{P}_{t-1}^a\mathbf{M}^T + \mathbf{Q}$$

Kalman gain

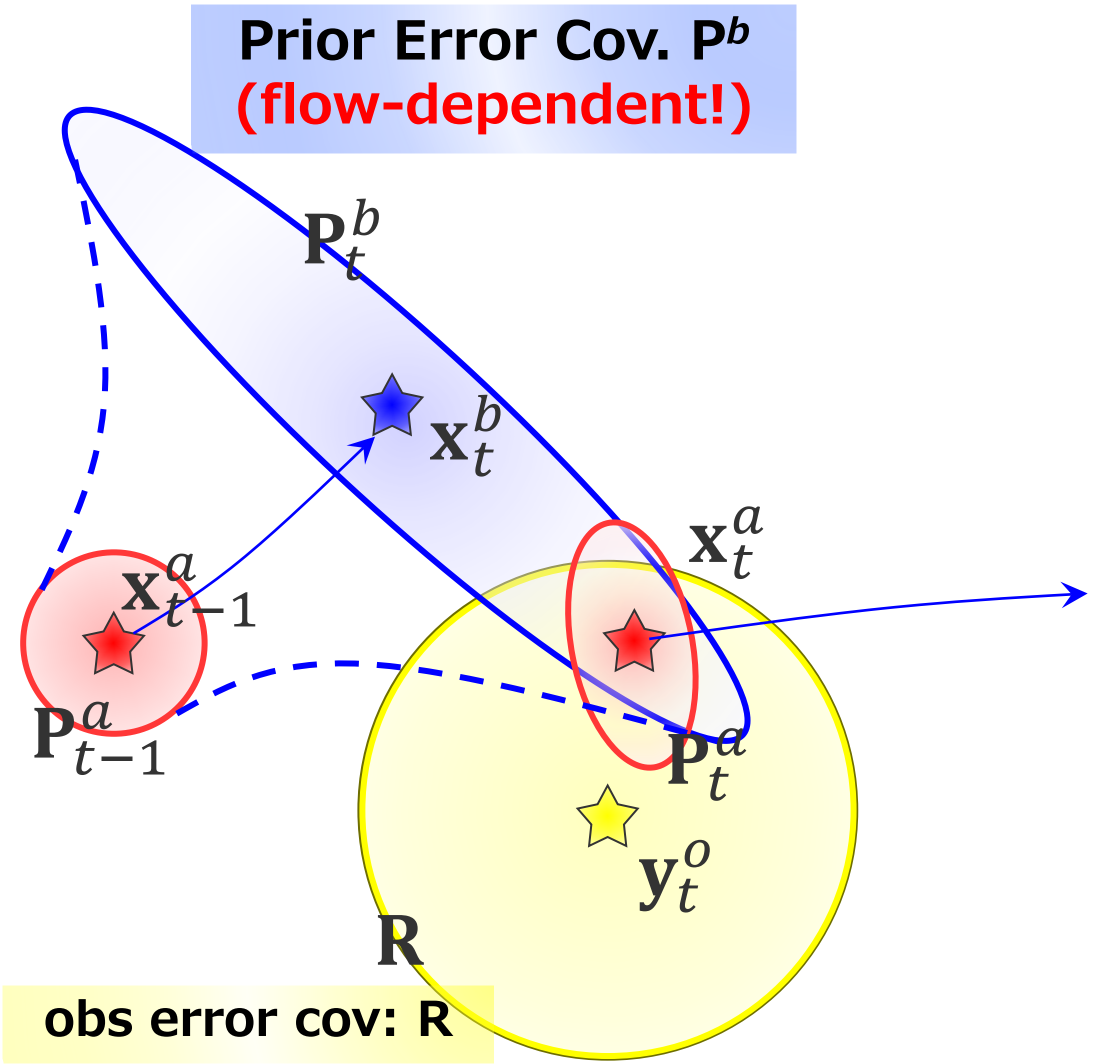
$$\mathbf{K}_t = \mathbf{P}_t^b\mathbf{H}^T\left[\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R}\right]^{-1}$$

Analysis (state)

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}_t(\mathbf{y}_t^o - H(\mathbf{x}_t^b))$$

Analysis (error covariance)

$$\mathbf{P}_t^a = [\mathbf{I} - \mathbf{K}\mathbf{H}]\mathbf{P}_t^b$$

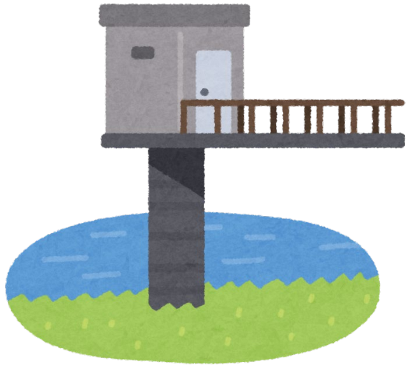


Summary

- **Minimum variance estimation suppose**
  - unbias
  - uncorrelated error
- **Maximum likelihood estimation suppose**
  - unbias
  - uncorrelated error
  - Gaussian error PDF
- **These solutions are identical**
  - when errors are Gaussian
  - Namely, minimum variance estimation gives optimal analysis following Bayesian theory w/ Gaussian errors
  - (細かいが大事) 最小分散推定 (KF & EnKF)は、誤差のガウス分布性を仮定しない。我々が信頼するのは最尤推定で、最小分散推定と最尤推定は、誤差がガウス分布の時に一致する。だから、誤差のガウス分布性はKF & EnKFにも望ましいのだ。

何故、観測を真としないか？

- **データ同化のコンセプト**
  - 観測よりも真値に近い値を求まる
- **よくある疑問**
  - 観測ってかなり精度が高いハズ
  - なんで観測にも誤差があると考えるのか？
- **(1) 劣決定問題 (観測 << モデルの変数)**
  - 一意に解が求まらない
    - → L2正則化で (a) 解を一意に、(b) 自信過剰の抑制
  - L2正則化と、「観測も誤差がある」は、数学的に同値
    - 伊藤・藤井 (2020; ながれ)
- **(2) 観測誤差 ≠ 計測誤差**
  - (a) 計測誤差: あらゆる計測は誤差を含む
  - (b) モデル表現性誤差: 自然現象の非完全な数値モデル化
  - (c) 代表性誤差: 或る観測の空間代表制
  - データ同化は数値モデルの予測が改善することを目的とする



e.g. 水位観測

On observation error

$$\boldsymbol{\varepsilon}^o \equiv \mathbf{y}^o - H^{model}(\mathbf{x}^{tru}) \qquad \langle \boldsymbol{\varepsilon}^o \rangle = 0$$

$$= \underbrace{\mathbf{y}^o - \mathbf{y}^{tru}}_{\text{measurement error}} + \underbrace{\mathbf{y}^{tru} - H^{tru}(\mathbf{x}^{tru})}_{\text{representativeness error}} + \underbrace{H^{tru}(\mathbf{x}^{tru}) - H^{model}(\mathbf{x}^{tru})}_{\text{error in obs. operator}}$$

measurement error

representativeness error

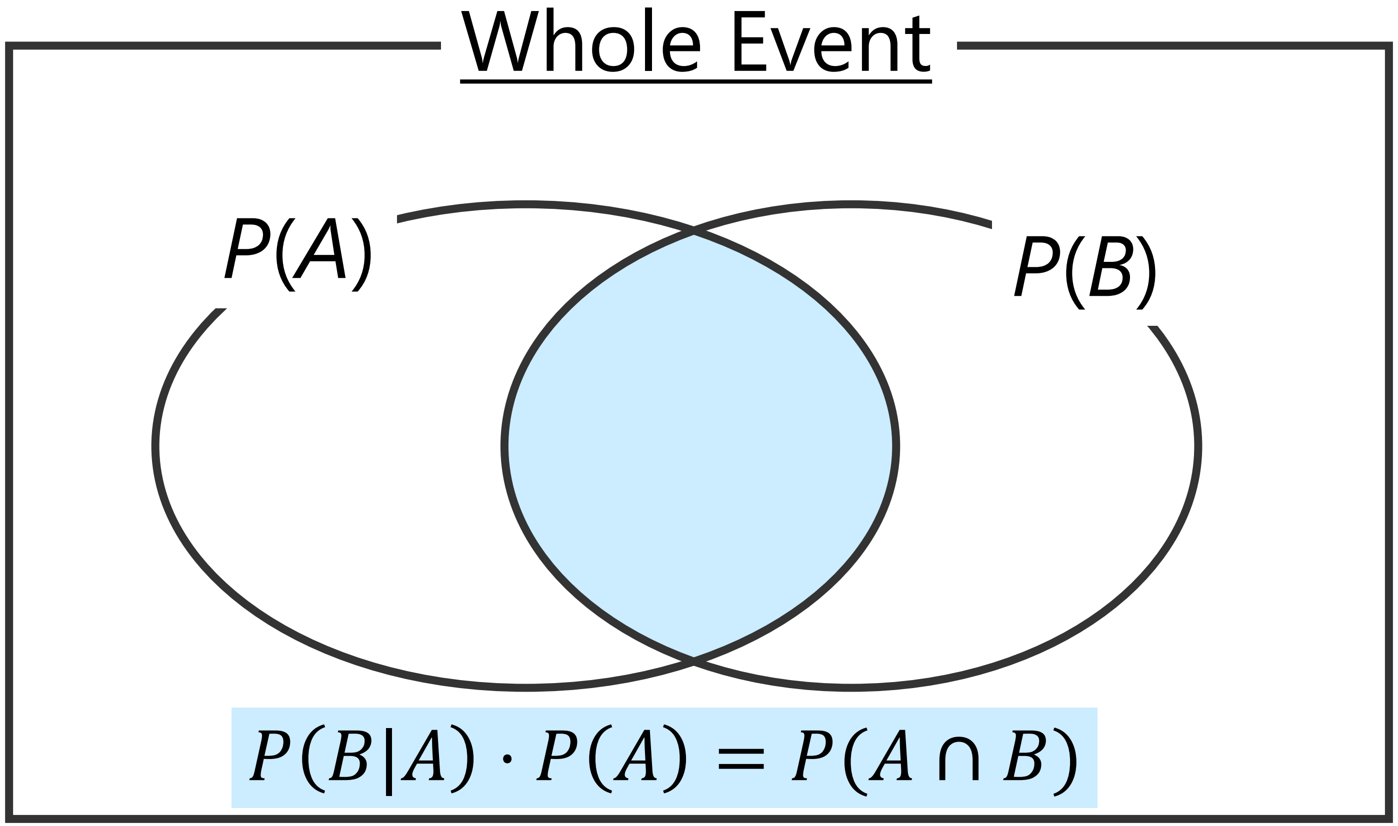
error in obs. operator

Bayesian Estimation

Bayesian Theorem

Posterior, Likelihood, Prior, Obs

$$\underbrace{P(B|A)}_{\text{Posterior}} = \frac{\overbrace{P(A|B)}^{\text{Likelihood}}\,\overbrace{P(B)}^{\text{Prior}}}{\underbrace{P(A)}_{\text{Obs}}}$$



$$P(B|A) \cdot P(A) = P(A \cap B)$$

Bayesian Theorem: an example

An example: virus infection (e.g. COVID19) and inspection

- virus rate is 0.005 (0.5 %)
- inspection to people with virus ➔ gives positive (+) w/ 80 %
- inspection to healthy people ➔ gives negative (-) w/ 90 %

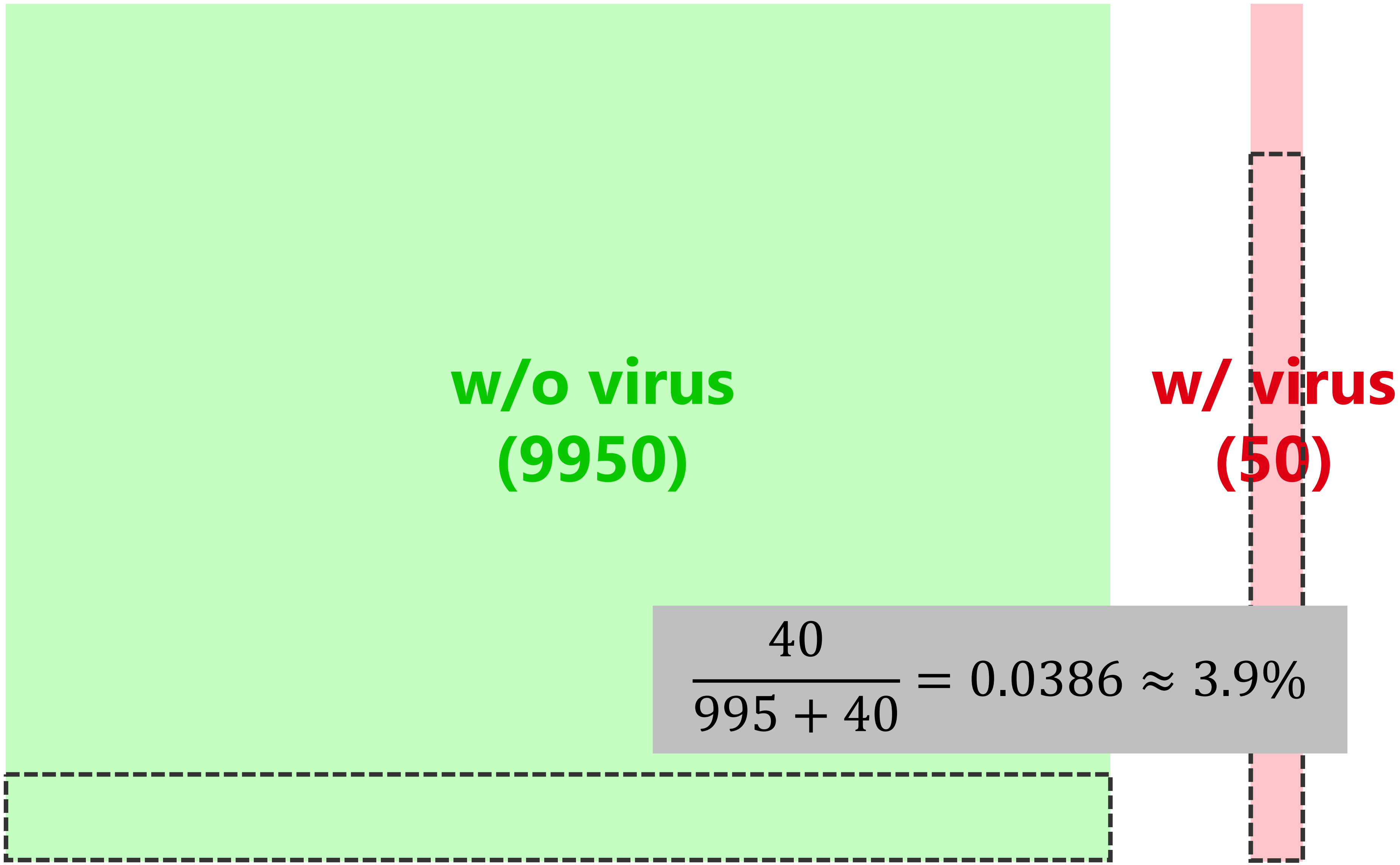
Now, you have a **positive** result by the inspection!!!
➔ The percentage of having virus is only **about 3.9 %.**

Interpretation by cases

$$\frac{40}{995+40} = 0.0386 \approx 3.9\%$$

Total	Virus Rate	Joint	Inspection Result
10000	50 (virus)	x0.8 = 40	**Positive** (correct)
		x0.2 = 10	Negative (incorrect)
	9950 (healthy)	x0.9 = 8955	Negative (correct)
		x0.1 = 995	**Positive** (incorrect)

Intuitive Interpretation



$$\frac{40}{995 + 40} = 0.0386 \approx 3.9\%$$

incorrect positive (偽陽性) 10% ➔ 995

correct positive (陽性) 80% ➔ 40

Bayesian Theorem: an example

An example: virus infection (e.g. COVID19) and inspection

- virus rate is 0.005 (0.5 %)
- inspection to people with virus ➔ gives positive (+) w/ 80 %
- inspection to healthy people ➔ gives negative (-) w/ 90 %

Now, you have a **positive** result by the inspection!!!

➔ The percentage of having virus is only **about 3.9 %.**

Posterior　Likelihood　Prior

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

Obs

Prior: Prob. of virus

Obs: Prob. of positive

Likelihood: Prob. of positive given virus

Posterior: Prob. of virus given positive

$$P(B|A) = \frac{0.8 \times 0.005}{(0.005 \times 0.8 + 0.995 \times 0.1)} = 0.0386 \approx 3.9\%$$

Then,,,, so what?

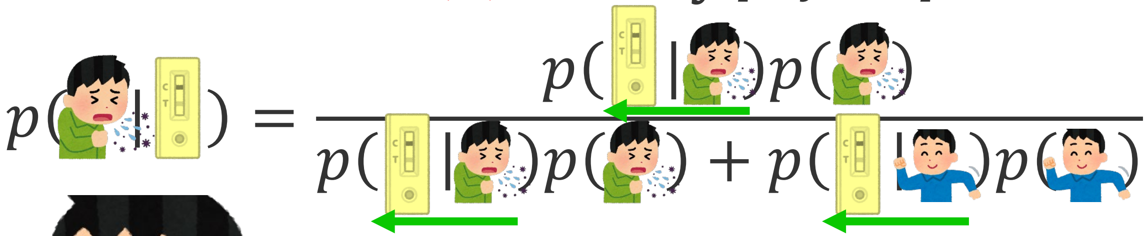
Bayesian Theorem

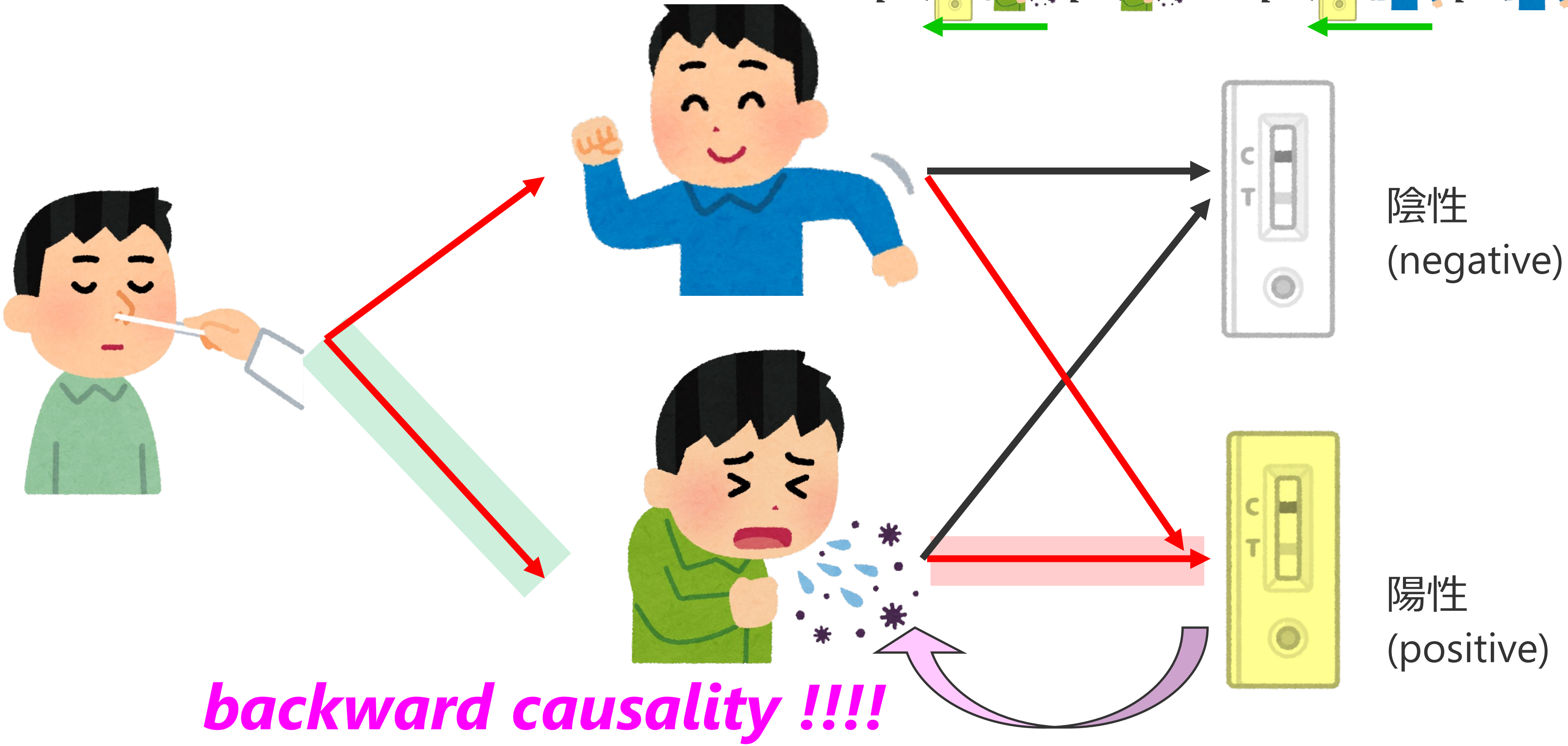
$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

←: forward

←: backward (結果➔原因)

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$$

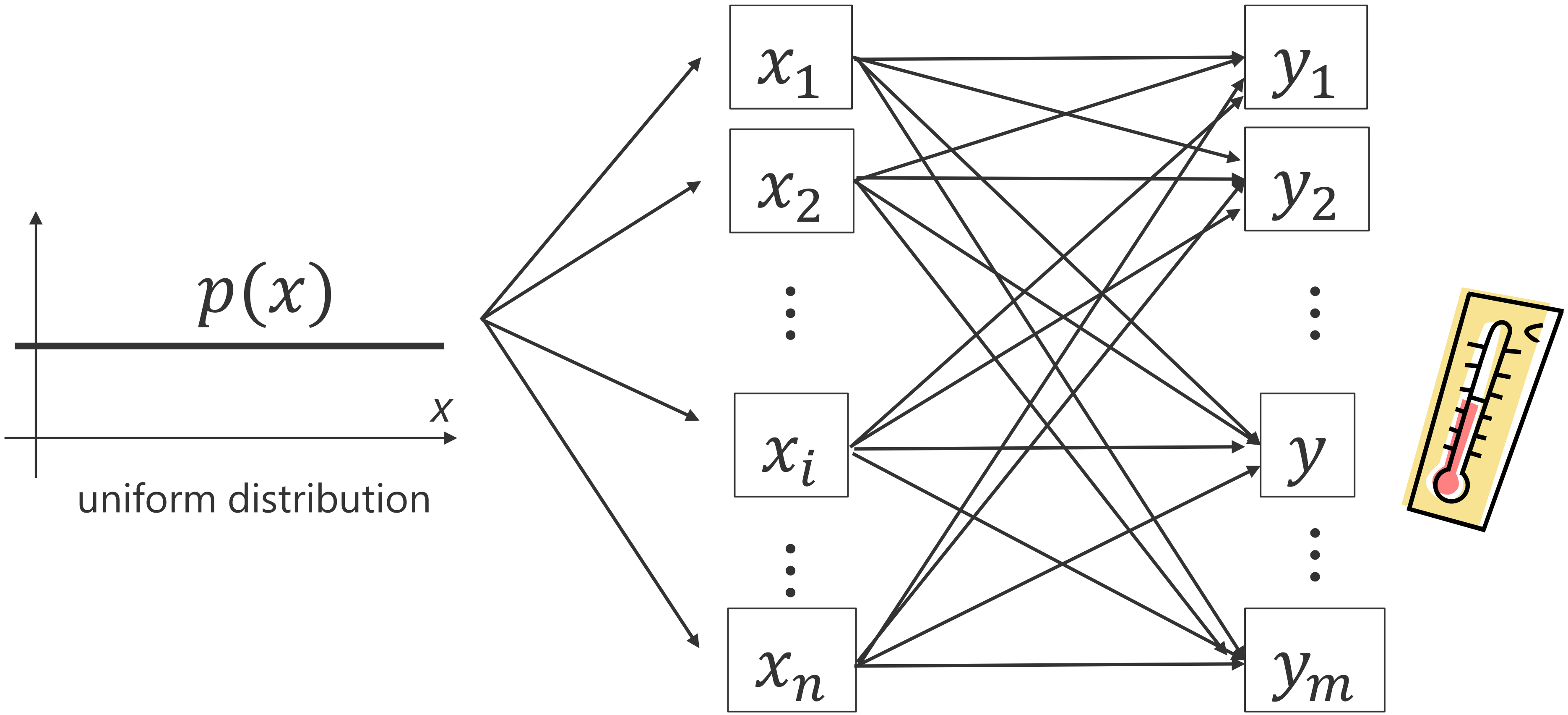




Bayesian Estimation

Bayesian Theorem (discrete)

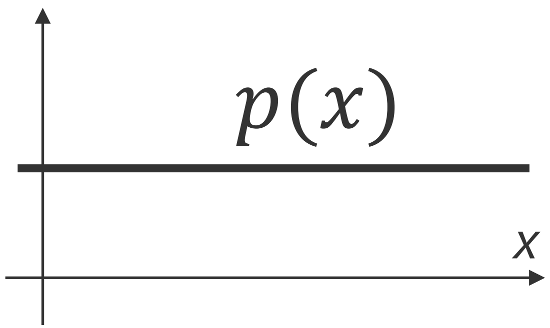
$$p(x_i|y) = \frac{p(y|x_i)p(x_i)}{p(y)} = \frac{p(y|x_i)p(x_i)}{\sum_{k=1}^{n} p(y|x_k)p(x_k)}$$



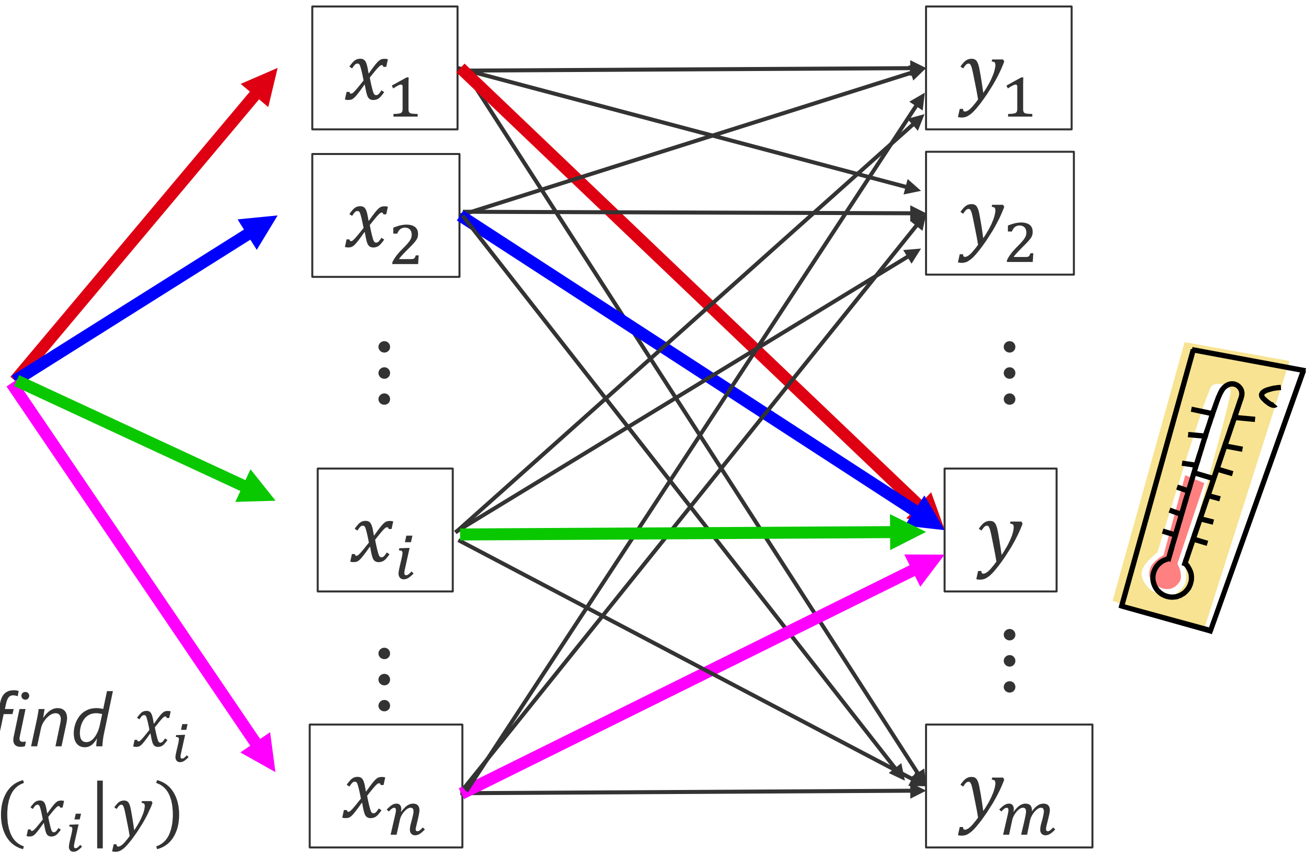
Bayesian Estimation

Bayesian Theorem (discrete)

$$p(x_i|y) = \frac{p(y|x_i)p(x_i)}{p(y)} = \frac{p(y|x_i)p(x_i)}{\sum_{k=1}^{n} p(y|x_k)p(x_k)}$$



uniform distribution



We would like to find x_i that maximizes $p(x_i|y)$

Bayesian Estimation

Bayesian Theorem (discrete)

$$p(x_i|y) = \frac{p(y|x_i)p(x_i)}{p(y)} = \frac{p(y|x_i)p(x_i)}{\sum_{k=1}^{n} p(y|x_k)p(x_k)}$$

Bayesian Theorem (general)

Likelihood — Prior (uniform)

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$$

Posterior: $p(x|y)$

Likelihood: $p(y|x)$; Prior (uniform): $p(x)$

constant (i.e., not a func. of x): $\int p(y|x)p(x)dx$

$$\left(= \frac{p(y|x)p(x)}{\int p(y|\theta)p(\theta)d\theta} \right)$$

We would like to find x that maximizes $p(x|y)$

$$maximize\ p(x|y)$$
$$\Leftrightarrow\ maximize\ p(y|x)$$

Maximum Likelihood Estimation

forecast	$x_1 = x^{tru} + \varepsilon_1$	(1) unbias	$\langle x_1 \rangle = \langle x_2 \rangle = x^{tru}$
observation	$x_2 = x^{tru} + \varepsilon_2$	(2) uncorr.	$\langle \varepsilon_1 \varepsilon_2 \rangle = 0$

Likelihood — Prior (uniform, i.e., no prior info)

$$p(x|x_{1,2}) = \frac{p(x_{1,2}|x)\,p(x)}{p(x_{1,2})}$$

Posterior

$p(x_{1,2})$: constant (since they are given)

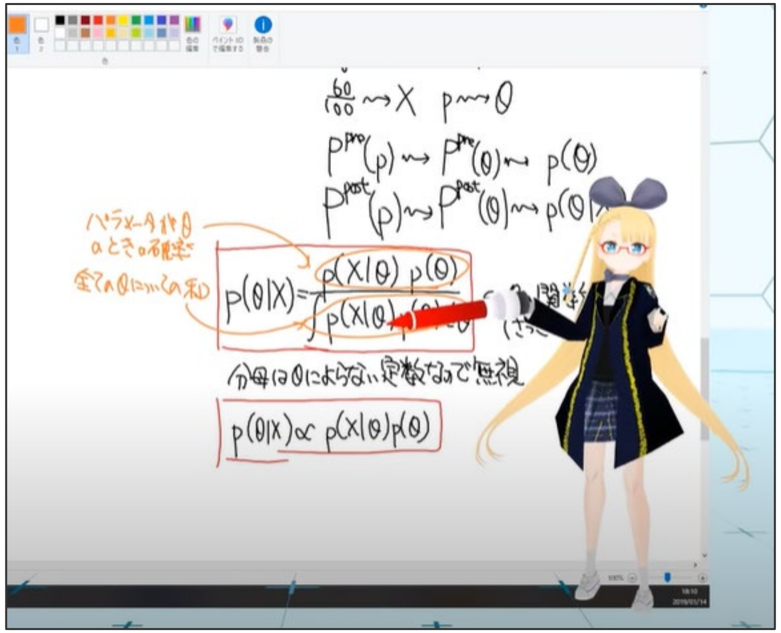
Bayesian Estimates

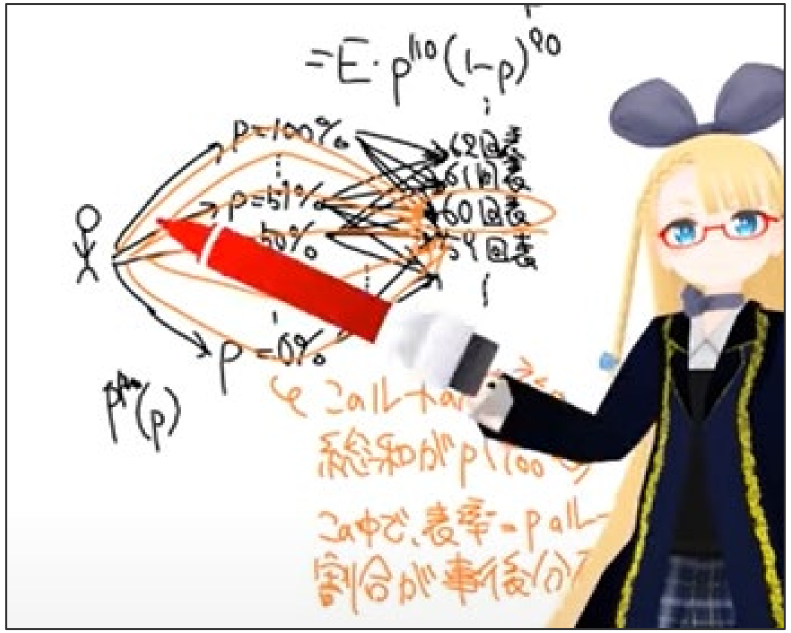
$$maximize \quad p(x|x_{1,2}) \Leftrightarrow maximize \quad p(x_{1,2}|x)$$

$$\Leftrightarrow maximize \quad p(x_1|x) \cdot p(x_2|x)$$

to maximize likelihood

Recommendations (Jpn)





Thank you for your attention!

Presented by Shunji Kotsuki
(shunji.kotsuki@chiba-u.jp)

Further information is available at
https://kotsuki-lab.com/

Environmental Prediction Science Laboratory