Thomas Bayes 1701-1761

Baysean Inference

https://www.youtube.com/watch?v=-e8wOcaascM

$$P(\theta|\text{data}) = [P(\text{data}|\theta) \times P(\theta)] \,/\, P(\text{data})$$

Z: 確率変数 (random variables)

$P(Z)$: 確率質量関数 (probability mass function)

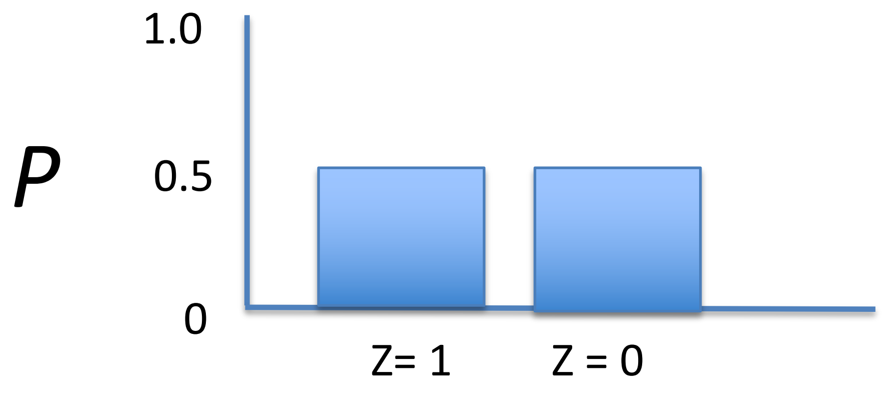
$$\sum_i P(Z_i) = 1$$

Z: 確率変数 (random variables)

P(Z) : 確率質量関数 (probability mass function)

コイン投げ (Coin throw)　Z = おもて: 1 or うら : 0

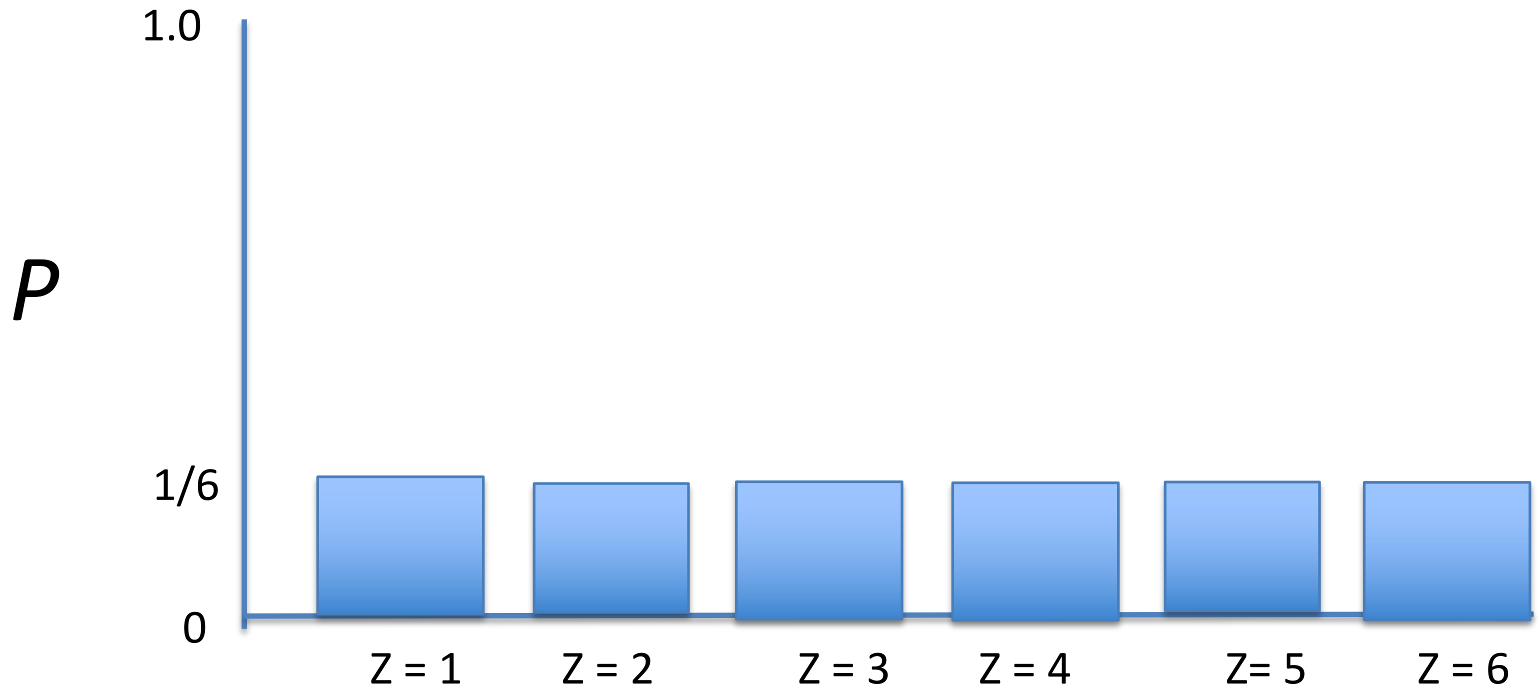
$P(1) = 0.5$, $P(0) = 0.5$



$P(Z{=}1) + P(Z{=}0) = 1$

サイコロ投げ（Dice throw)　Z = 1, 2, 3, 4, 5, 6

$P(1) = 1/6$, $P(2) = 1/6$, $P(1) = 1/6$, $P(2) = 1/6$, $P(1) = 1/6$, $P(2) = 1/6$,

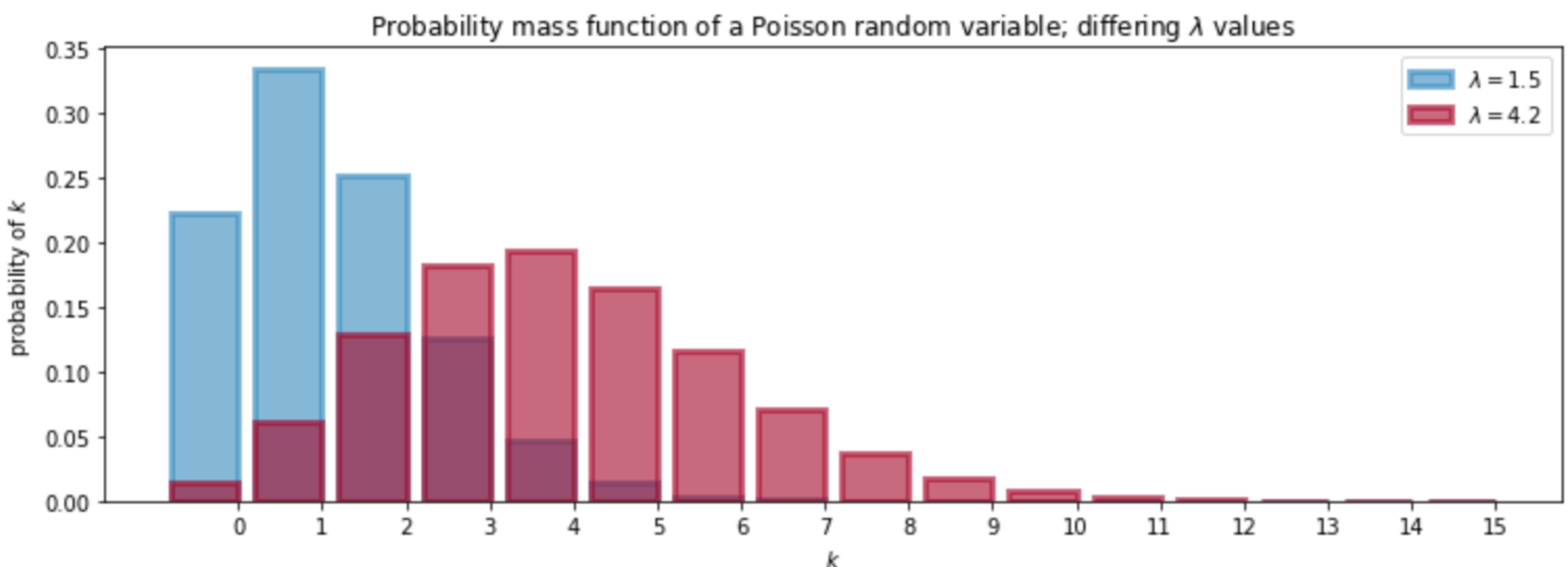


$$P(\mathrm{Z}=1) + P(\mathrm{Z}=2) + P(\mathrm{Z}=3) + P(\mathrm{Z}=4) + P(\mathrm{Z}=5) + P(\mathrm{Z}=6) = \sum_i P(\mathrm{Z} = i) = 1$$

ポアソン分布 (Poisson Distribution)

10Kgのパン生地に400粒のレーズンをいれてこねた後、0.1Kgずつのパンに分けて焼いた時、パン一個あたりのレーズンの数の分布は？　　$\lambda = 4$

$$P(Z=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots, \quad \lambda \in \mathbb{R}_{>0}$$



競馬の例 （Horse Race）:　馬A と 馬B の勝敗の割合

A

B

	0 負 (lose)	1 勝 (win)
0 負 (lose)	30/100	10/100
1 勝 (win)	10/100	50/100

$$\sum_{i=0}^{1}\sum_{j=0}^{1} \mathrm{P}(\mathrm{Z_A} = i, \mathrm{Z_B} = j) = 30/100 + 10/100 + 10/100 + 50/100 = 1$$

Lambert, B. A Student’s Guide to Bayesian Statistics, 2018より引用

競馬の例 （Horse Race）: 周辺分布 (Marginal Distribution)

A

B

	0 負 (lose)	1 勝 (win)	$P(Z_B)$
0 負 (lose)	30/100	10/100	40/100
1 勝 (win)	10/100	50/100	60/100
$P(Z_A)$	40/100	60/100	周辺確率

$P(Z_A=0) = P(Z_A=0, Z_B=0) + P(Z_A=0, Z_B=0) = 30/100+10/100 = 40/100$

条件付き確率（Conditional Probability)

P(A|B）＝P(A,B)/P(B)

条件付確率　= 同時確率/周辺確率

Aが勝った場合に、Bも勝つ確率

$\mathbf{P(Z_B=1 | Z_A=1)} = P(Z_A=1, Z_B=1) / P(Z_A=1)$

$= P(Z_A=1, Z_B=1) / P(Z_A=1, Z_B=0) + P(Z_A=1, Z_B=1)$

$= (50/100) / [(10/100) + (50/100)] = 5/6$

A

B	0 負 (lose)	1 勝 (win)
0 負 (lose)	30/100	10/100
1 勝 (win)	10/100	50/100

—>

Bが勝った場合に
Aが勝つ確率は？

Baysean Inference

https://www.youtube.com/watch?v=-e8wOcaascM

Thomas Bayes 1701-1761

$$P(\theta|data) = [P(data|\theta) \times P(\theta)] / P(data)$$

ベイズの定理 （Bayes‘s rule)

$P(X|Y) = [P(Y|X) \times P(X)] / P(Y)$

1/ $P(X|Y) = P(X,Y)/P(Y)$

2/ $P(Y|X) = P(X,Y)/P(X)$

3/ $P(X|Y) \times P(Y) = P(X,Y) = P(Y|X) \times P(X)$

$\rightarrow$ $P(X|Y) = [P(Y|X) \times P(X)] / P(Y)$

条件付き確率（Conditional Probability)

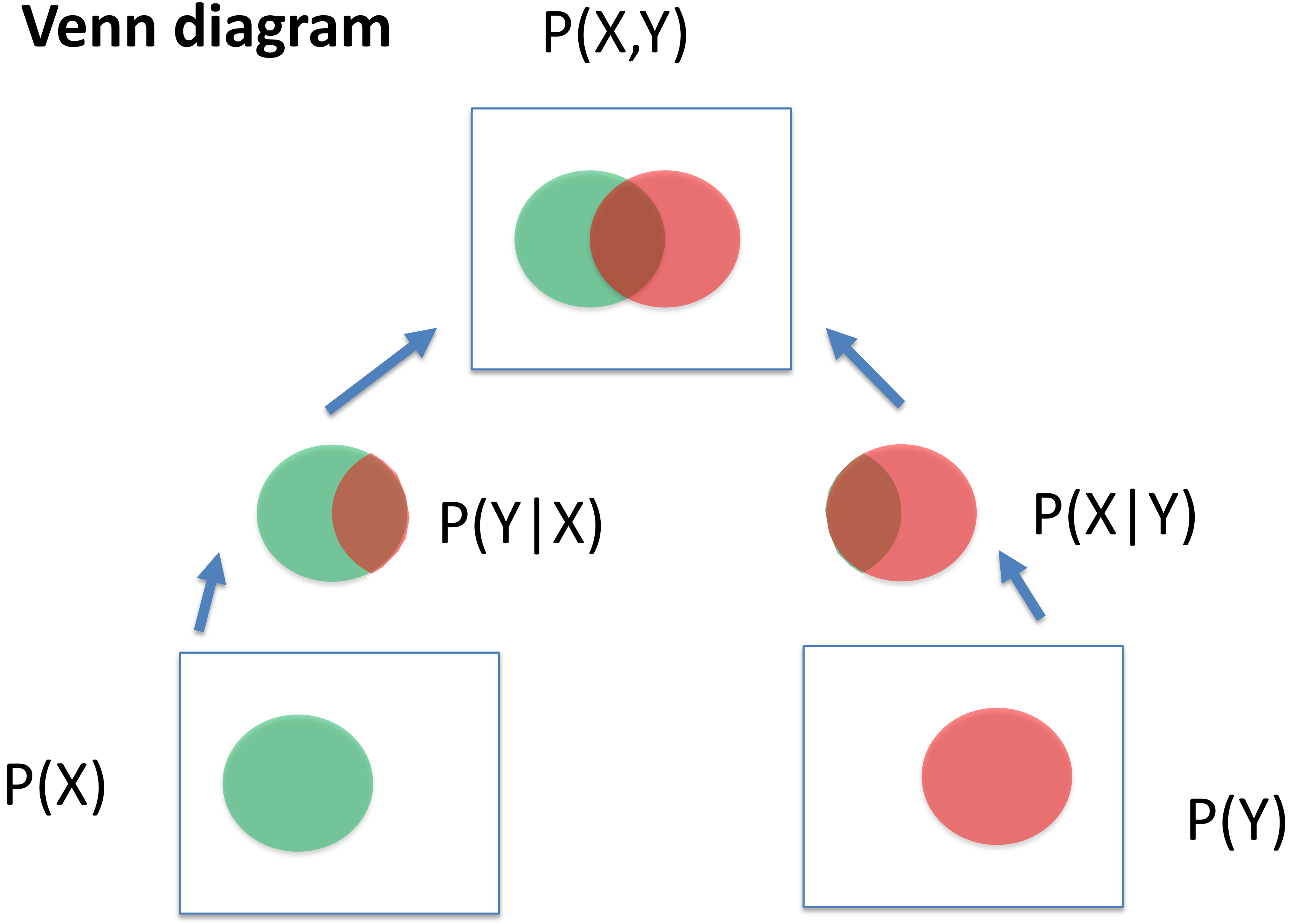
P(X|Y）＝P(X,Y)/P(Y)

条件付確率　= 同時確率/周辺確率

Conditional　　simultaneous (joint) probalibity

Probability　　　/ marginal probability

Venn diagram



$$P(X|Y) \times P(Y) = P(X,Y) = P(Y|X) \times P(X)$$

$$\rightarrow \quad P(X|Y) = [P(Y|X) \times P(X)] \,/\, P(Y)$$

Use of Bayes theorem **(modified from Lambert, 2018)**

Example of Colon Cancer

1. Out of all men aged 40, 1% have colon cancer
2. Screen of colon cancer: 80% of men with colon cancer show positive
3. 10% of men without colon cancer show screen positive

Question: What is the percentage of colon cancer if screen result is positive?

P(cancer) = 0.01
P(Sc+|cancer) = 0.8
P(Sc+|no cancer) = 0.1 Question: **P(cancer|Sc+)**

P(cancer|Sc+) x P(Sc+) = P(cancer, Sc+) = P(Sc+|cancer) x P(cancer)

P(cancer|Sc+) = [P(Sc|cancer) x P(cancer)] / P(Sc+)
= (0.8 x 0.01) /0.11 = 0.08

P(Sc+) = P(Sc+, cancer) + P(Sc+, no cancer)
= P(Sc+|cancer) x P(cancer) + P(Sc+|no cancer) x P(no cancer)

= 0.8 x 0.01 + 0.1 x 0.99 = 0.11

Baysean Inference

https://www.youtube.com/watch?v=-e8wOcaascM

Thomas Bayes 1701-1761

$$P(\theta|data) = [P(data|\theta) \times P(\theta)] / P(data)$$

Posterior = Likelihood x Prior / P(data)

Likelihood (尤度)

Head (表）　　Tail (裏）

Coin throw (コイン投げ）

P(H) = θ = 0.5

Throw coin twice (２回コイン投げをする時の結果）

1st throw １回目　　2nd throw 2回目

P(H, H) = θ^2 ＝0.5 x 0.5 = 0.25

Coin throw: x 2　(2回のコイン投げ）

H H

H T

T T

$= \theta^2 = 0.5 \times 0.5$
$= 0.25$

T H

$= (1-\theta)^2 = 0.5 \times 0.5$
$= 0.25$

$= 2 \times \theta \times (1-\theta) = 2 \times 0.5 \times 0.5$
$= 0.5$

Binomial distribution (二項分布）

n: trial number (試行回数）
k: number of success (成功回数）

$P(X=k\,|\,\theta) = {}_nC_k\ \theta^k(1-\theta)^{n-k}$

$P(X=0) = \theta^0(1-\theta)^2 = (1-0.5)^2 = 0.25$
$P(X=1) = 2x\ \theta\ x\ (1-\theta) = 2x(0.5)x(1-0.5) = 0.5$
$P(X=2) = \theta^2(1-\theta)^0 = (0.5)^2 = 0.25$

尤度 （Likelihood)

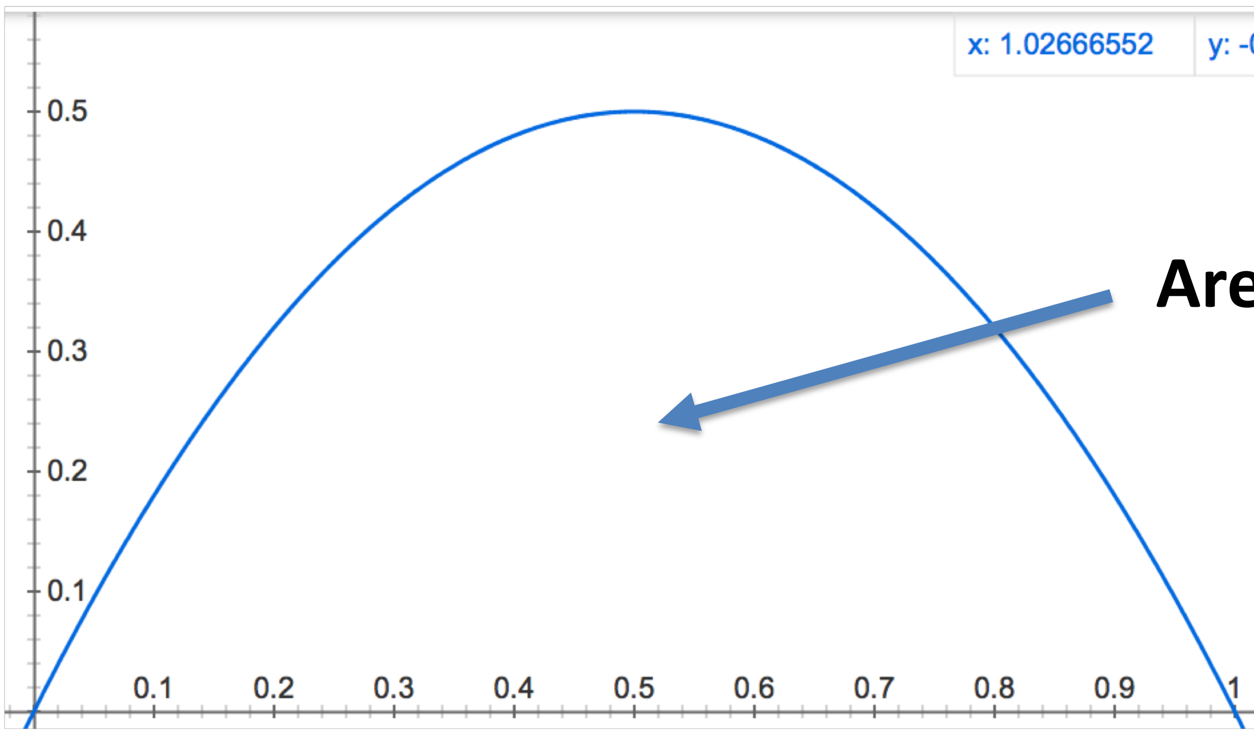
$$L(\theta|\text{data}) = P(\text{data}|\theta)$$

$$P(X=k|\theta) = {}_nC_k\, \theta^k(1-\theta)^{n-k}$$

$$n = 2$$

$$L(\theta \mid X = 1) = P(X = 1|\theta) = 2\theta \times (1-\theta)$$

Likelihood



Area = 1/3

θ

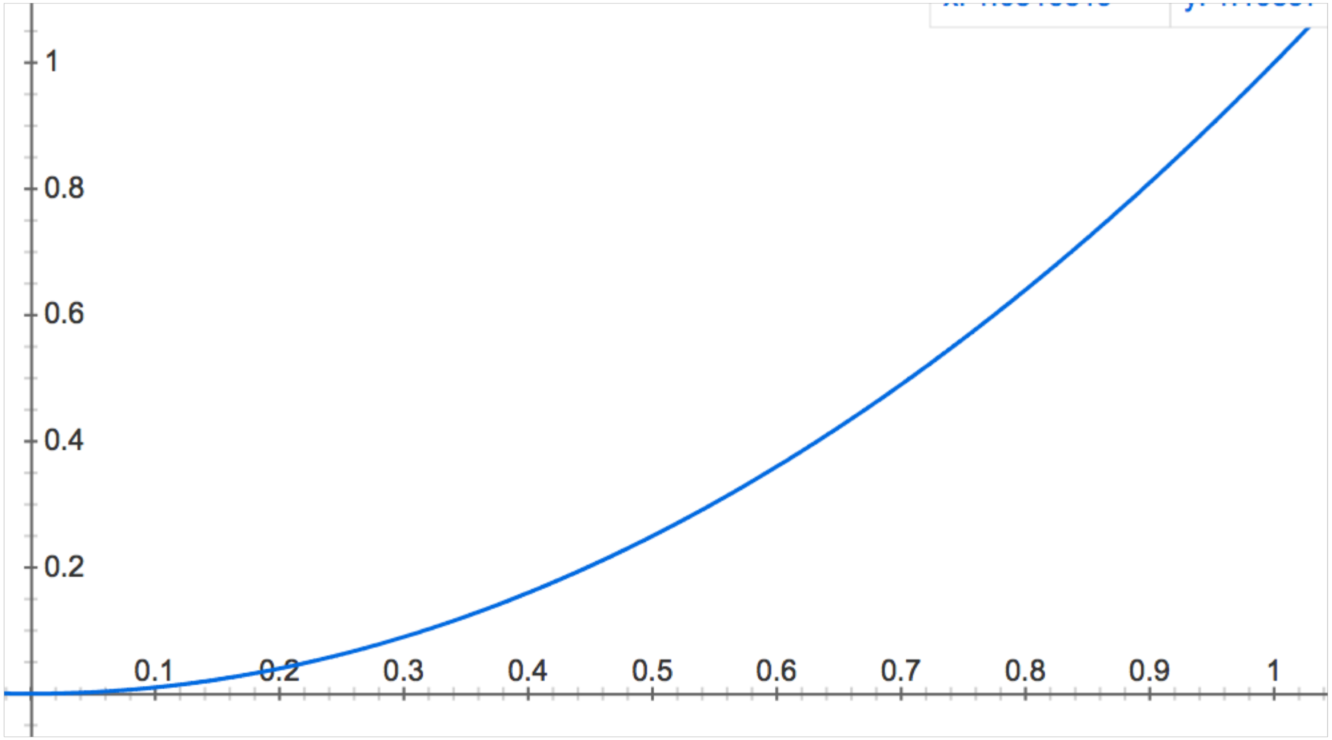
Maximum Likelihood (最尤推定値)

$P(X=k|\theta) = {}_nC_k\, \theta^k(1-\theta)^{n-k}$

$n = 2$

$L(\theta \mid X = 2) = P(X = 2 | \theta) = \theta^2$

Likelihood



θ

Thomas Bayes 1701-1761

Baysean Inference

https://www.youtube.com/watch?v=-e8wOcaascM

$$P(\theta | data) = [P(data | \theta) \times P(\theta)] / P(data)$$

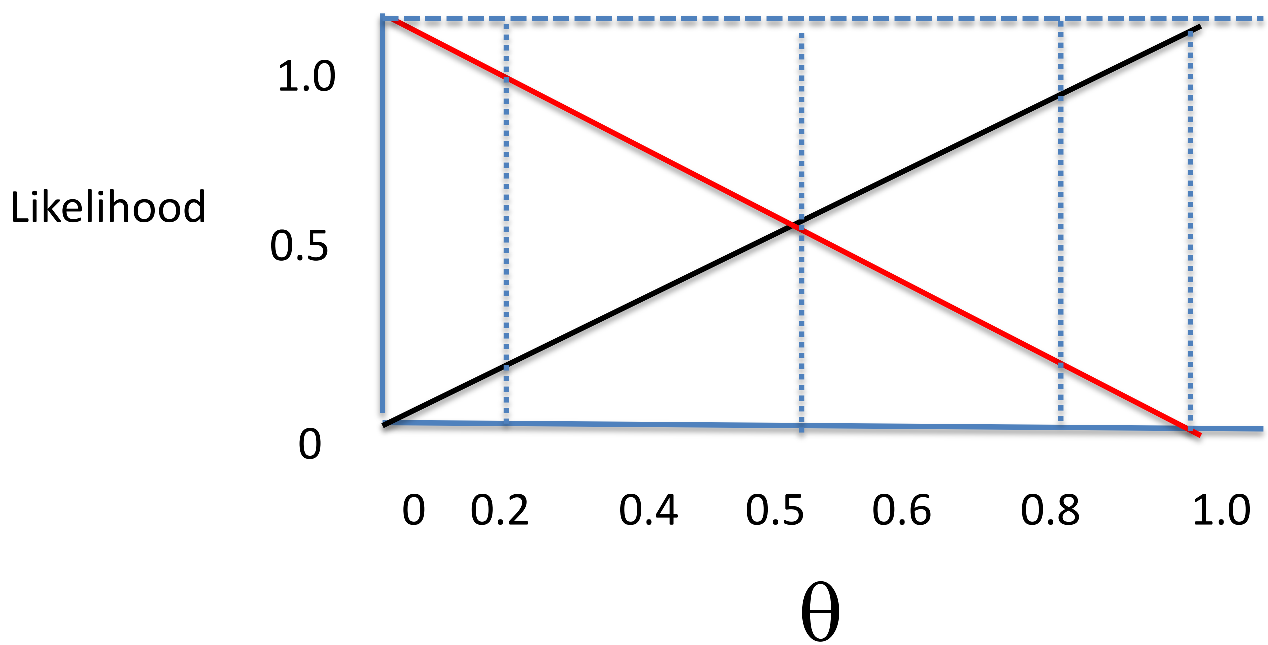
Posterior = Likelihood x Prior / P(data)

Likelihood (尤度)

X = 0 No Disease
X = 1 Disease

Goal: estimate the probability θ that a randomly chosen person has a disease

$P(X=0|\theta) = 1 - \theta$
$P(X=1|\theta) = \theta$



Example of Bayes Inference (from Dobson and Barnett, 2008)

Infection by *Schistosoma japonicum*

θ *:* Infection rate

H0: Infection is not endemic ($\theta <= 0.5$)
H1: Infection is endemic ($\theta > 0.5$)

After a visit to a village, the investigator thinks he is 80% sure that infection is endemic.

Sampling stool from 10 individuals of the village 7 had *S. japonicum*

$$P\ (y|\theta) = \mathrm{Bin}(10,\ \theta) = {}_{10}C_7 \times \theta^7 \times (1-\theta)^3$$

θ	Hypothesis	$P(\theta)$ Prior	$P(y\|\theta)$ Likelihood	$P(y\|\theta)$x $P(\theta)$	$P(\theta\|y)$ Posterior
0.0	H_0	0.0333	0.000	0.000	0.000
0.1	H_0	0.0333	0.000	0.000	0.000
0.2	H_0	0.0333	0.001	0.000	0.000
0.3	H_0	0.0333	0.009	0.000	0.002
0.4	H_0	0.0333	0.043	0.001	0.011
0.5	H_0	0.0333	0.117	0.004	0.032
Sum		0.2			0.046
0.6	H_1	0.16	0.215	0.034	0.277
0.7	H_1	0.16	0.267	0.043	0.344
0.8	H_1	0.16	0.201	0.032	0.260
0.9	H_1	0.16	0.057	0.009	0.074
1.0	H_1	0.16	0.000	0.000	0.000
Sum		0.8		0.124	0.954

ベイズ推定　(Bayes inference)

$$P(\theta|data) = [P(data|\theta) \times P(\theta)] / P(data)$$

θ　:　**確率分布のパラメータ**

(例 :　Poisson分布の λ)

ベイズの定理の応用例