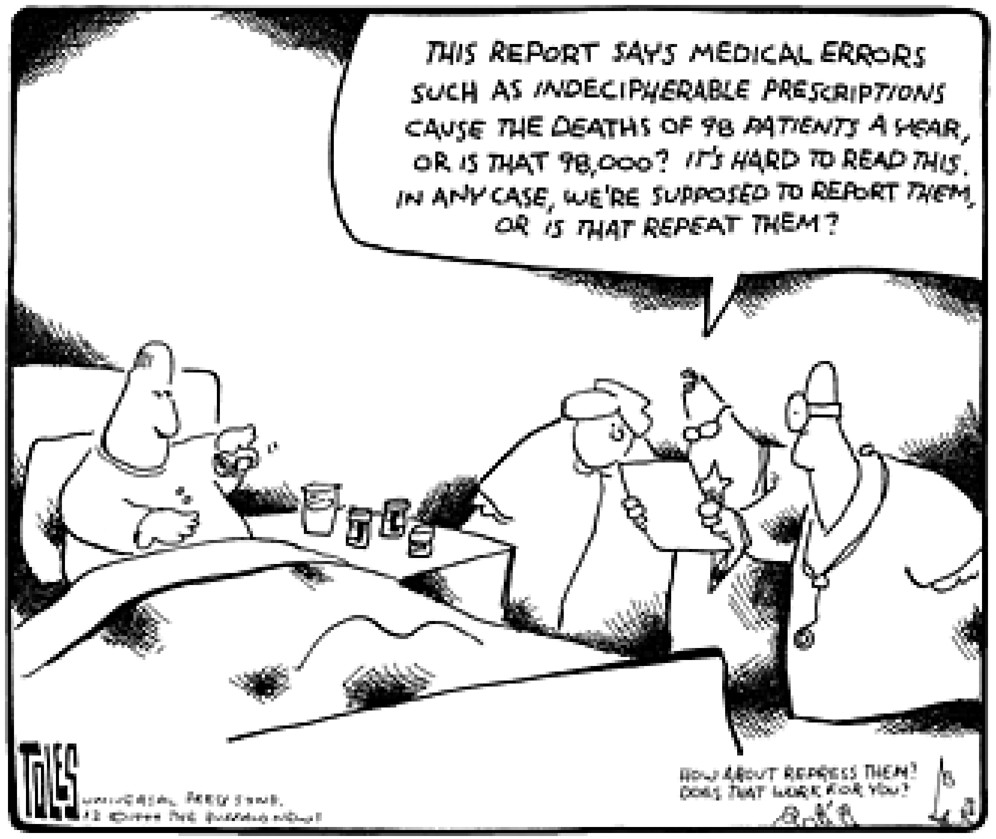


TOLES©1999 The Buffalo News. Reprinted with permission of UNIVERSAL PRESS SYNDICATE. All rights reserved.

Outline:

Basics

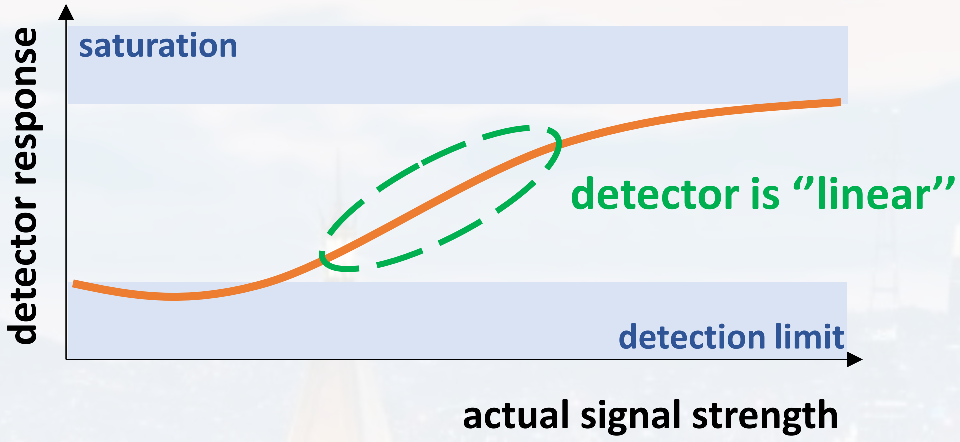
Most Common PDFs

- uniform
- binomial
- Poissonian
- Normal/Gaussian

Error Estimation

Bayesian Statistics

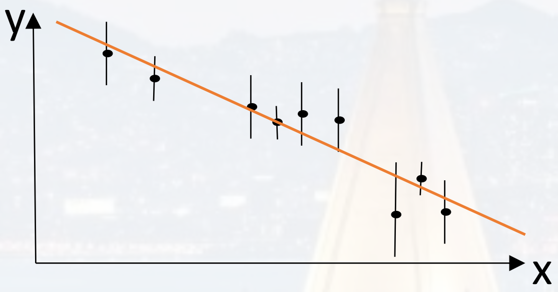
- **systematic errors:** calibration, non-linearity of the detector



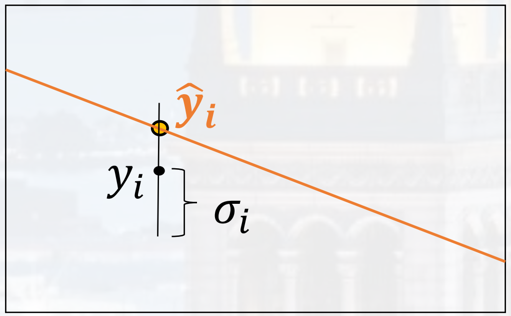
- **statistical errors:** limited precision, natural variance of the data
  → spread of the data around **an average value**

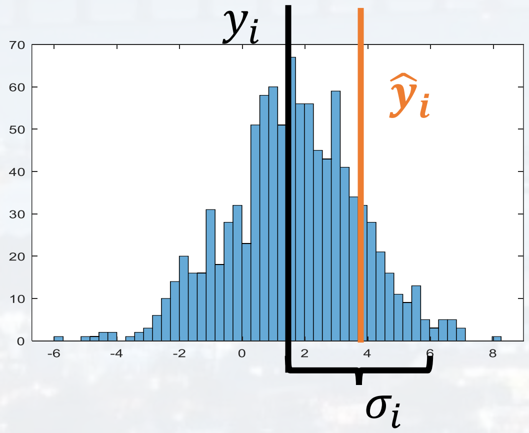
- **<u>statistical errors:</u>** limited precision, natural variance of the data
  → spread of the data around **an average value**

<u>assumption:</u> far from the detection limit and the saturation → the spread follows a **normal distribution**



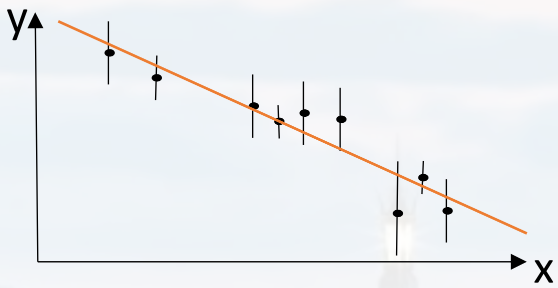
$$P(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\, exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]$$

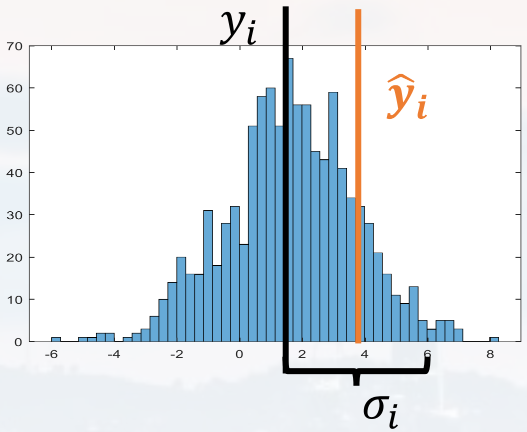




$$p_i(y_i|\hat{y}_i,\sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i}\, exp\left[-\frac{1}{2}\frac{(y_i-\hat{y}_i)^2}{{\sigma_i}^2}\right]$$

assumption: far from the detection limit and the saturation → the spread follows a **normal distribution**



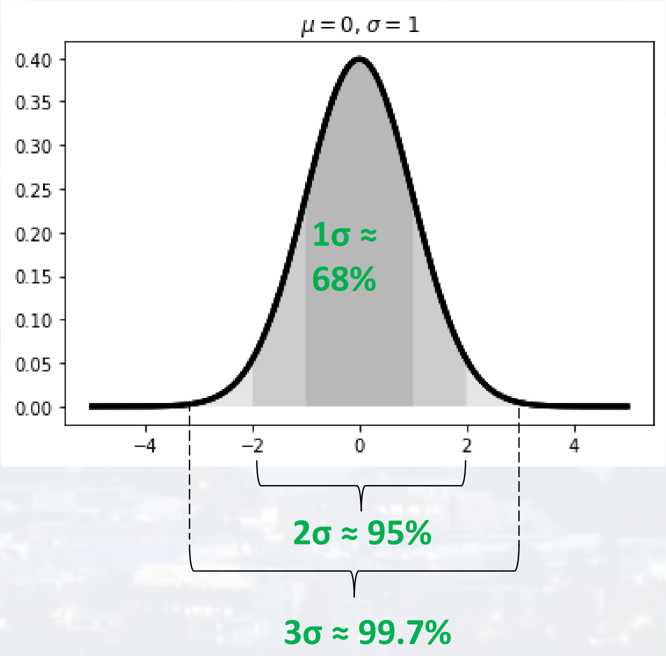


$$p_i(y_i|\hat{y}_i,\sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} exp\left[-\frac{1}{2}\frac{(y_i-\hat{y}_i)^2}{\sigma_i^2}\right]$$

for large **(> 50…100)** N **(number of data points)**:

≈ **2/3** of the data points should be consistent with the model within their **1σ** error bars

≈ **95%** of the data points should be consistent with the model within their **2σ** error bars

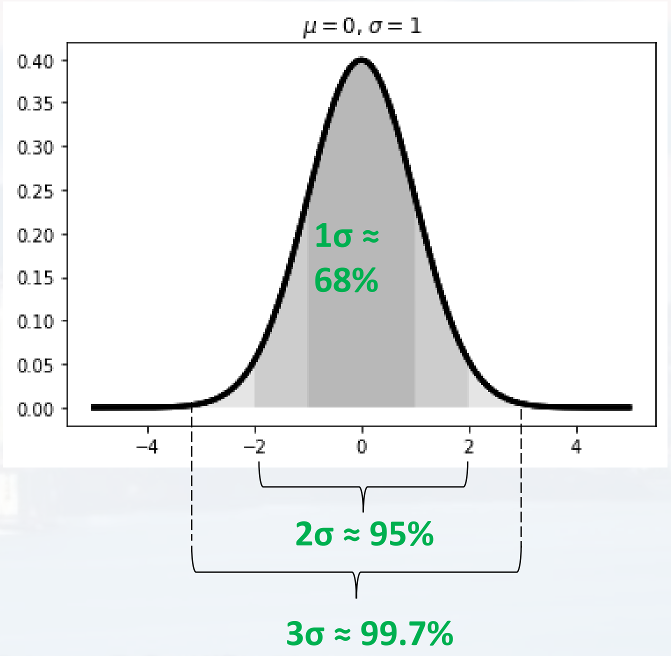


for large (> 50…100) N (number of data points):

≈ **2/3** of the data points should be consistent with the model within their **1σ** error bars

≈ **95%** of the data points should be consistent with the model within their **2σ** error bars

$$p_i(y_i|\hat{y}_i,\sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} exp\left[-\frac{1}{2}\frac{(y_i-\hat{y}_i)^2}{{\sigma_i}^2}\right]$$



$$\chi^2_{red} = \frac{1}{df}\sum_{i=1}^{N}\left(\frac{y_i - \hat{y}_i}{\sigma_i}\right)^2 \qquad df = N - p - 1$$

<u>rule of thumb:</u>

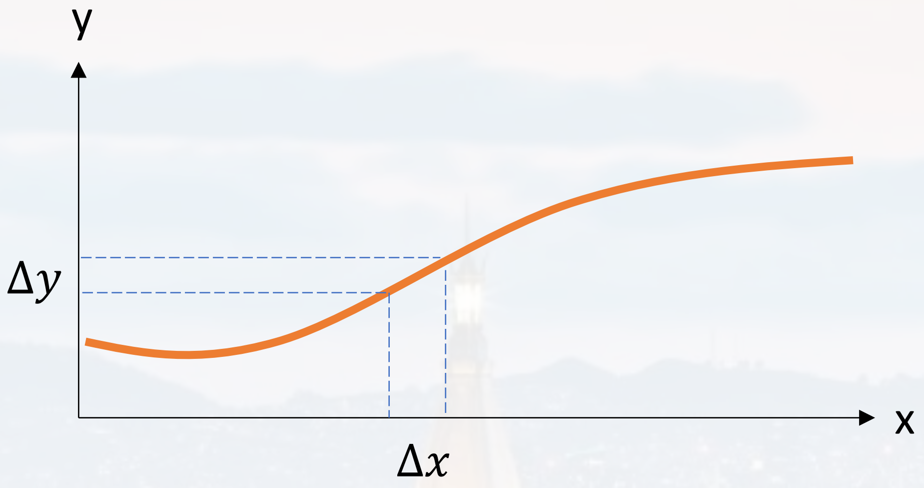
reduced χ^2 ≈

1.0 excellent fit
1.0…1.5 acceptable fit
1.5…1.7 bad fit
>2.0 not acceptable

<<1.0 suspicious, errors are overestimated!

y_i: measured value of data point
σ_i: statistical error of y_i (often aka ey_i)
$\hat{y}_i$: prediction by the model *after the fit*
N : number of data points
p: number of fit parameter

error propagation



$$\frac{dy}{dx} \approx \frac{\Delta y}{\Delta x} \qquad \text{for } \Delta x \ll \text{x}$$

$$\Delta x \left|\frac{dy}{dx}\right| \approx \Delta y$$

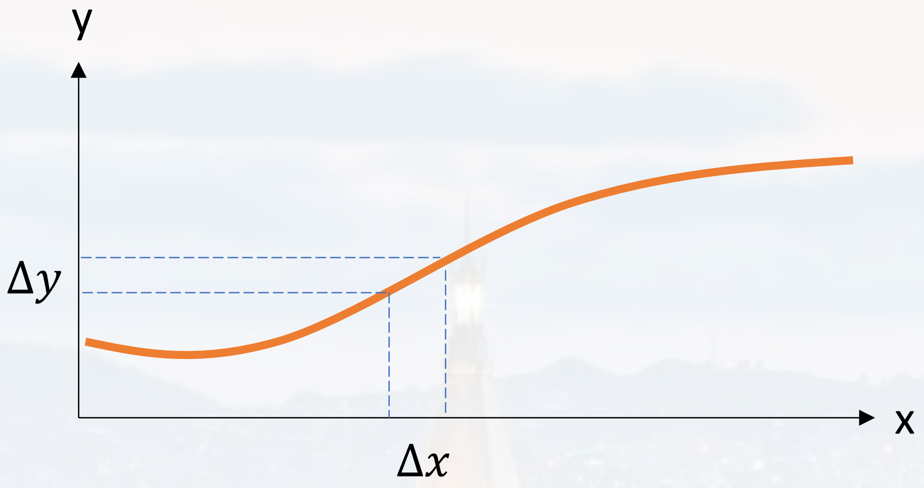
example:

$$V = \frac{4}{3}\pi\, r^3 \qquad \Delta V = ?\ \text{ given } \Delta r$$

$$\Delta V = \frac{dV}{dr}\Delta r = 4\,\pi\, r^2 \Delta r$$

$$\boxed{\frac{\Delta V}{\text{V}} = 3\,\frac{\Delta r}{r}} \qquad \Delta r, \Delta\text{V} \approx 1\sigma$$

error propagation



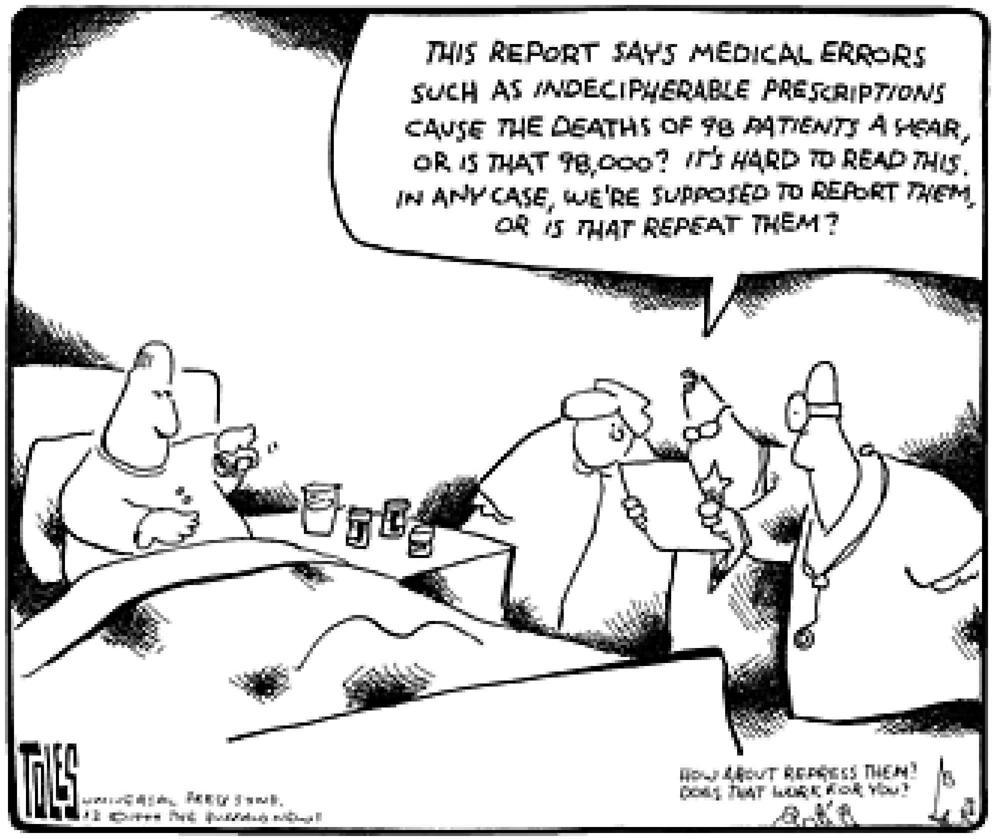
general:

$$\Delta f(max) = \sum_{i=1}^{I} \left|\frac{\partial f}{\partial x_i}\right| \Delta x_i$$ **maximum** error estimation

if x_i do **not correlate**, i. e. are **mutually independent**:

$$\Delta f^2 = \sum_{i=1}^{I} \left|\frac{\partial f}{\partial x_i}\right|^2 (\Delta x_i)^2$$

Note: $\Delta f(max)^2 > \Delta f^2$ because of the mixed terms $\left|\frac{\partial f}{\partial x_i}\right| \left|\frac{\partial f}{\partial x_j}\right| \Delta x_i \Delta x_j$ in $\Delta f(max)^2$



Outline:

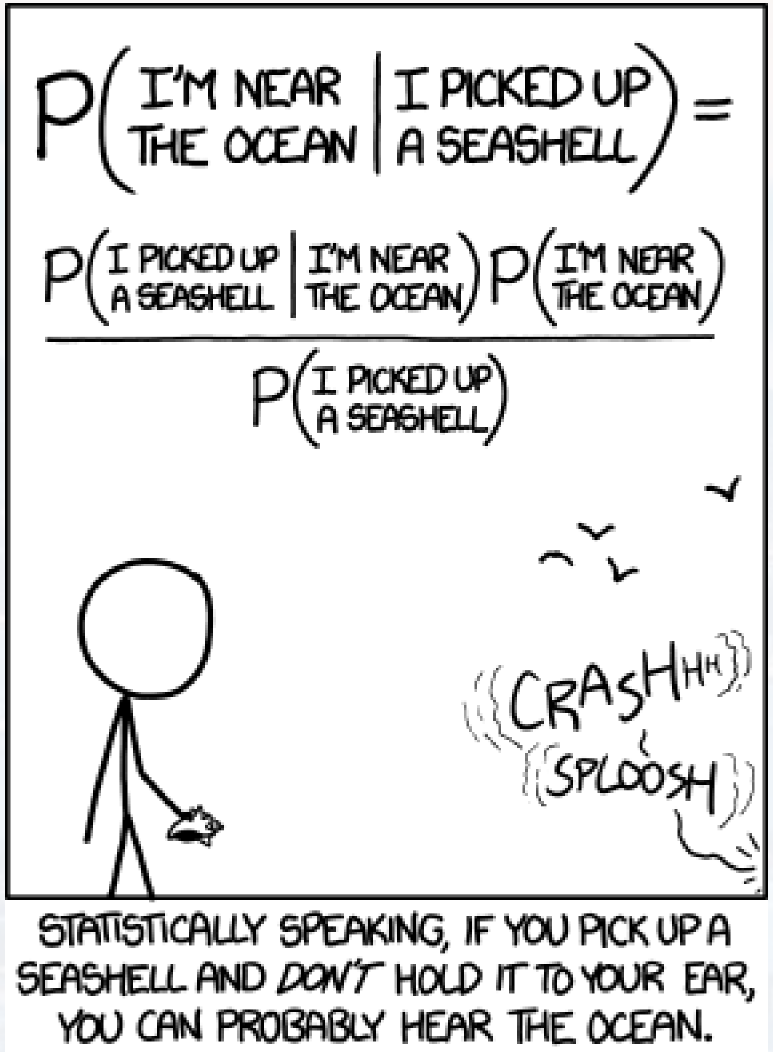
Basics

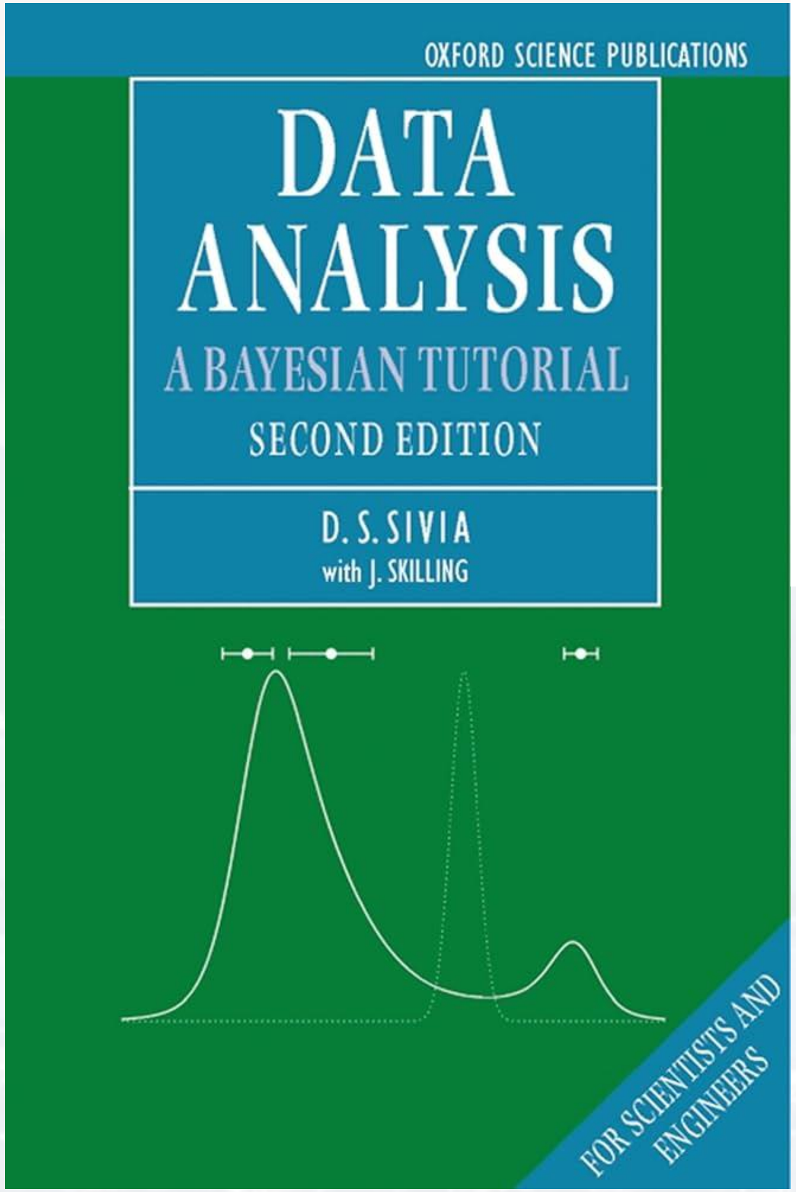
Most Common PDFs

- uniform
- binomial
- Poissonian
- Normal/Gaussian

Error Estimation

Bayesian Statistics


P(I'M NEAR THE OCEAN | I PICKED UP A SEASHELL) =
P(I PICKED UP A SEASHELL | I'M NEAR THE OCEAN) P(I'M NEAR THE OCEAN)
P(I PICKED UP A SEASHELL)
CRASHHH
SPLOOSH
STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND DON'T HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.


OXFORD SCIENCE PUBLICATIONS
DATA ANALYSIS
A BAYESIAN TUTORIAL
SECOND EDITION
D. S. SIVIA
with J. SKILLING
FOR SCIENTISTS AND ENGINEERS

$P(A \cap B)$ probability ***P*** that the events ***A*** **and** ***B*** occur

so far: A and B were independent $P(A \cap B) = P(A)P(B) = P(B)P(A)$

now: **conditional probabilities** | “given” or “under the condition”

Thomas Bayes (1701 - 1761)

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

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

Bayes Theorem

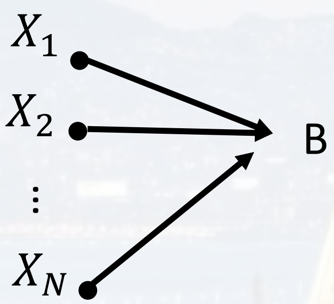
posterior $$\boldsymbol{P(A|B)} = \frac{P(B|A)\boldsymbol{P(A)}}{P(B)}$$ **prior**

$$\boxed{P(A|B)P(B) = P(B|A)\mathrm{P(A)}}$$

Bayes Theorem

Thomas Bayes (1701 - 1761)

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



$$P(B) = \sum_{n=1}^{N} P(B|X_n)P(X_n)$$

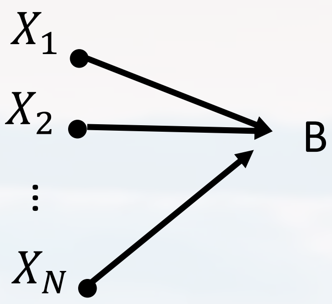
$$P(B) = \int P(B|X)P(X)\,dX$$

marginalization

example:

model: M
data: D

$$P(D|M) = \int P(D|all\ model\ param, M)\ P(all\ model\ param|M)\ \ d\ all\ model\ param$$



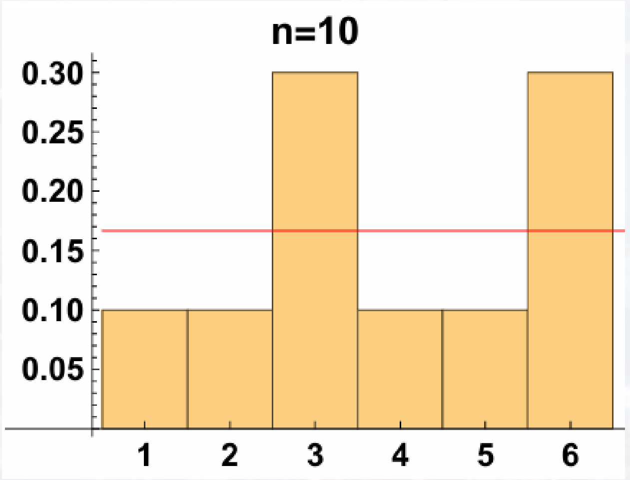
$$P(B) = \sum_{n=1}^{N} P(B|X_n)P(X_n)$$

$$P(B) = \int P(B|X)P(X)\,dX$$

marginalization

Thomas Bayes (1701 - 1761)

for a normal distribution $\mathcal{N}(\mu, \sigma)$

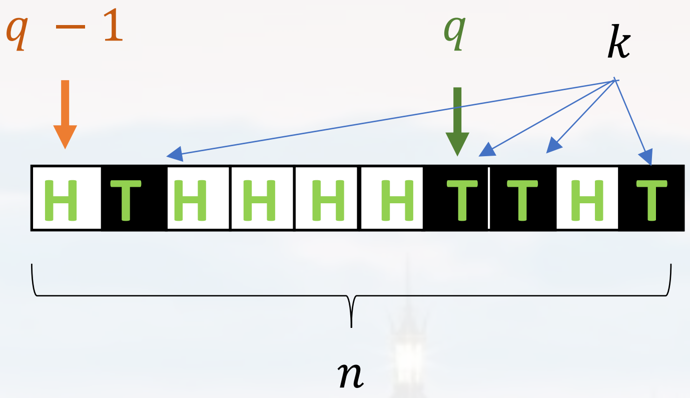


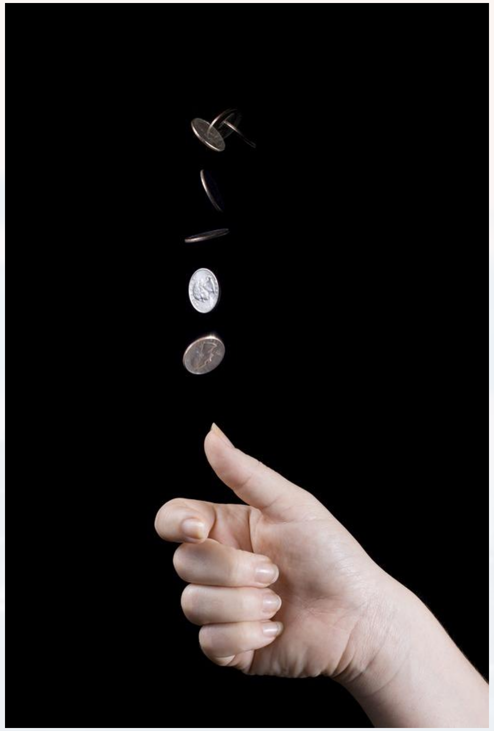
$\sigma = 2, \mu = 3.5$

$\sigma = 2, \mu = 5.0$

$\sigma = 1.5, \mu = 3.5$

$\sigma = 7.0, \mu = 1.0$





fair coin? q = 0.5 ???

probability of having **a** sequence of **k tails** and **n-k heads**

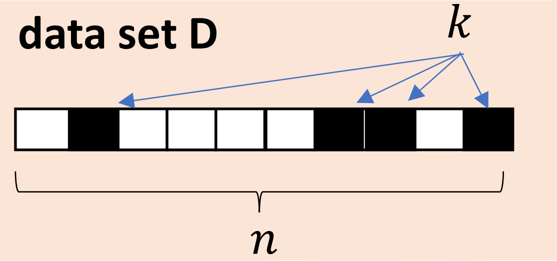
$$p_{tot} = \prod_i {q_i}^{n_i} = q^k\,(1-q)^{n-k}$$

probability of having **any** sequence of **k tails** and **n-k heads**

$$\boxed{P(k|q,n) = \binom{n}{k}\, q^k\,(1-q)^{n-k}}$$

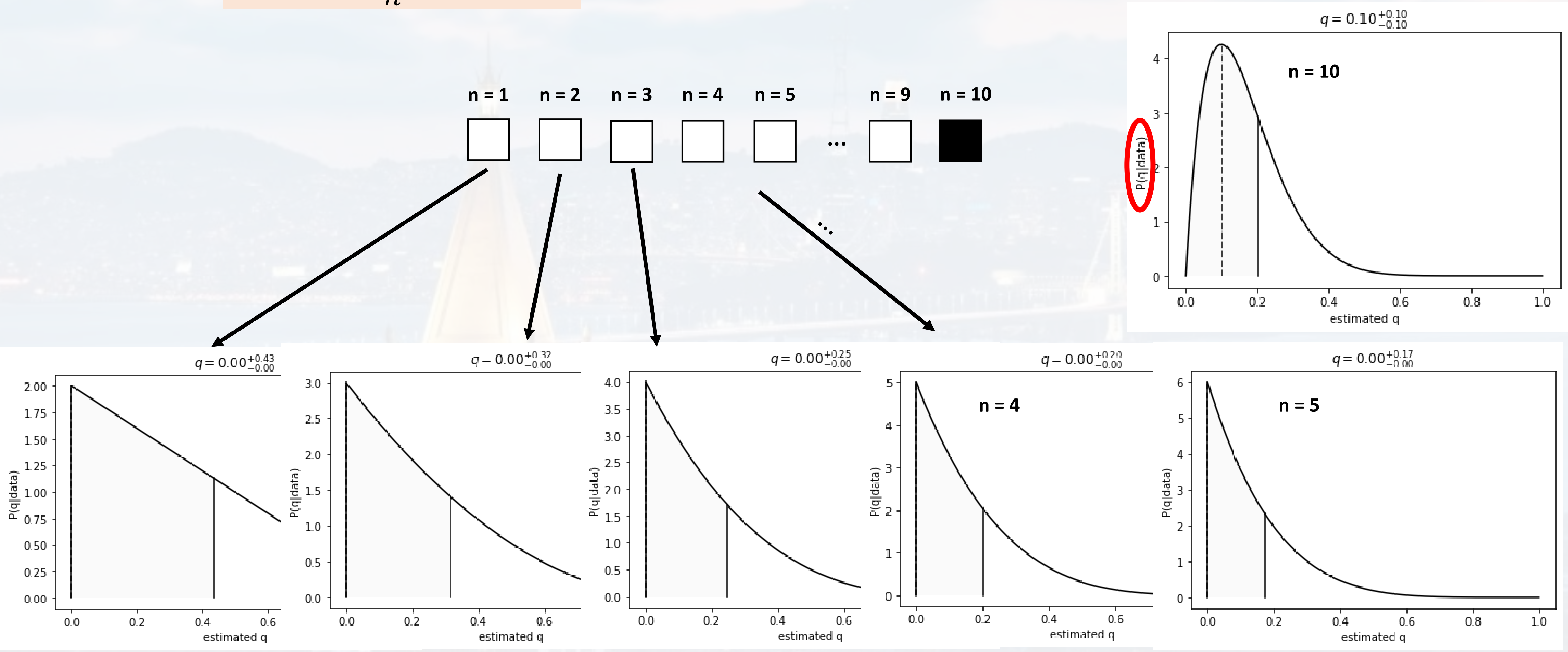
binomial distribution

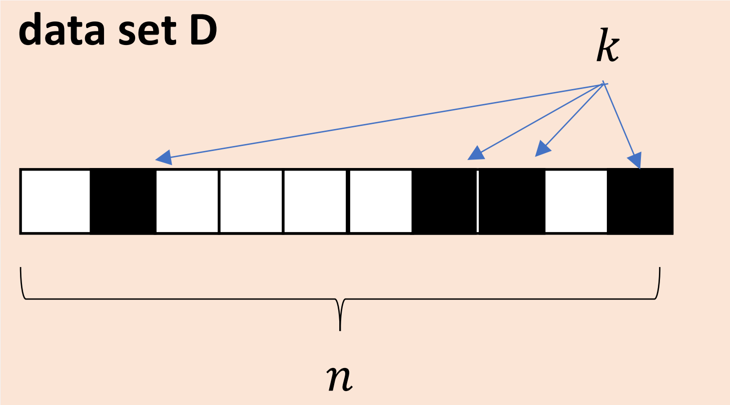
$$\frac{n!}{k!(n-k)!} =: \binom{n}{k}$$ **``n choose k''**



$q = ?$

goal:
- **P(q|D) Parameter Estimation**
- the larger ***D***, the more certain ***q***
  → learning





$\boldsymbol{q} = ?$

goal:
- **P(q|D)**
- the larger ***D***, the more certain ***q***
  → learning

$$P(k|n,q) = \binom{n}{k} q^k (1-q)^{n-k}$$

Bayes' theorem:

likelihood function **(here: binomial)**

$$P(q|data\ set) = \frac{\boxed{P(data\ set|q)}\boxed{P(q)}}{\boxed{P(data\ set)}}$$

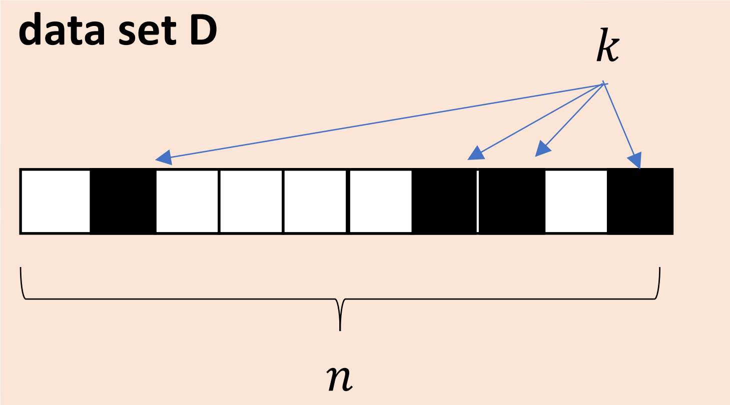
prior

evidence **(const wrt q)**

$$= \frac{\binom{n}{k} q^k (1-q)^{n-k}}{P(D)} P(q)$$

$$\sim q^k (1-q)^{n-k}\ P(q)$$

$P(D)$ and $\binom{n}{k}$ are no functions of q



$\boldsymbol{q} = ?$

goal:
- **P(q|D)**
- the larger ***D***, the more certain ***q***
  → learning

$$P(k|n,q) = \binom{n}{k} q^k (1-q)^{n-k}$$

$$P(q|data\ set) = \frac{P(data\ set|q)P(q)}{P(data\ set)}$$

$$= \frac{\binom{n}{k} q^k (1-q)^{n-k}}{P(D)} P(q)$$

$$\sim q^k (1-q)^{n-k}\ P(q)$$

max. entropy: $\mathrm{P}(q) = const$
if no prior information about q

$$\sim q^k (1-q)^{n-k}$$

$$\boxed{P(q|data\ set) = \frac{q^k (1-q)^{n-k}}{\int_0^1 q^k (1-q)^{n-k}\ dq}}$$

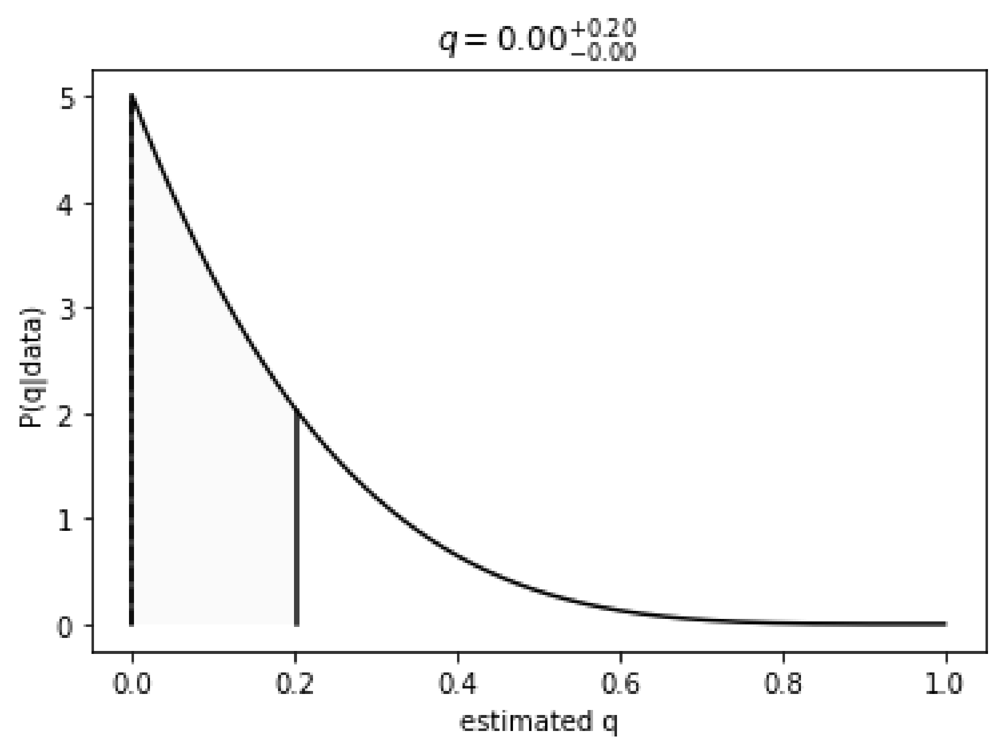
check out bayesian_bino.py

```
n1 = 4
k1 = np.random.binomial(n1, 0.25)
```

creating artificial data set
note: in reality ***q*** is unknown!

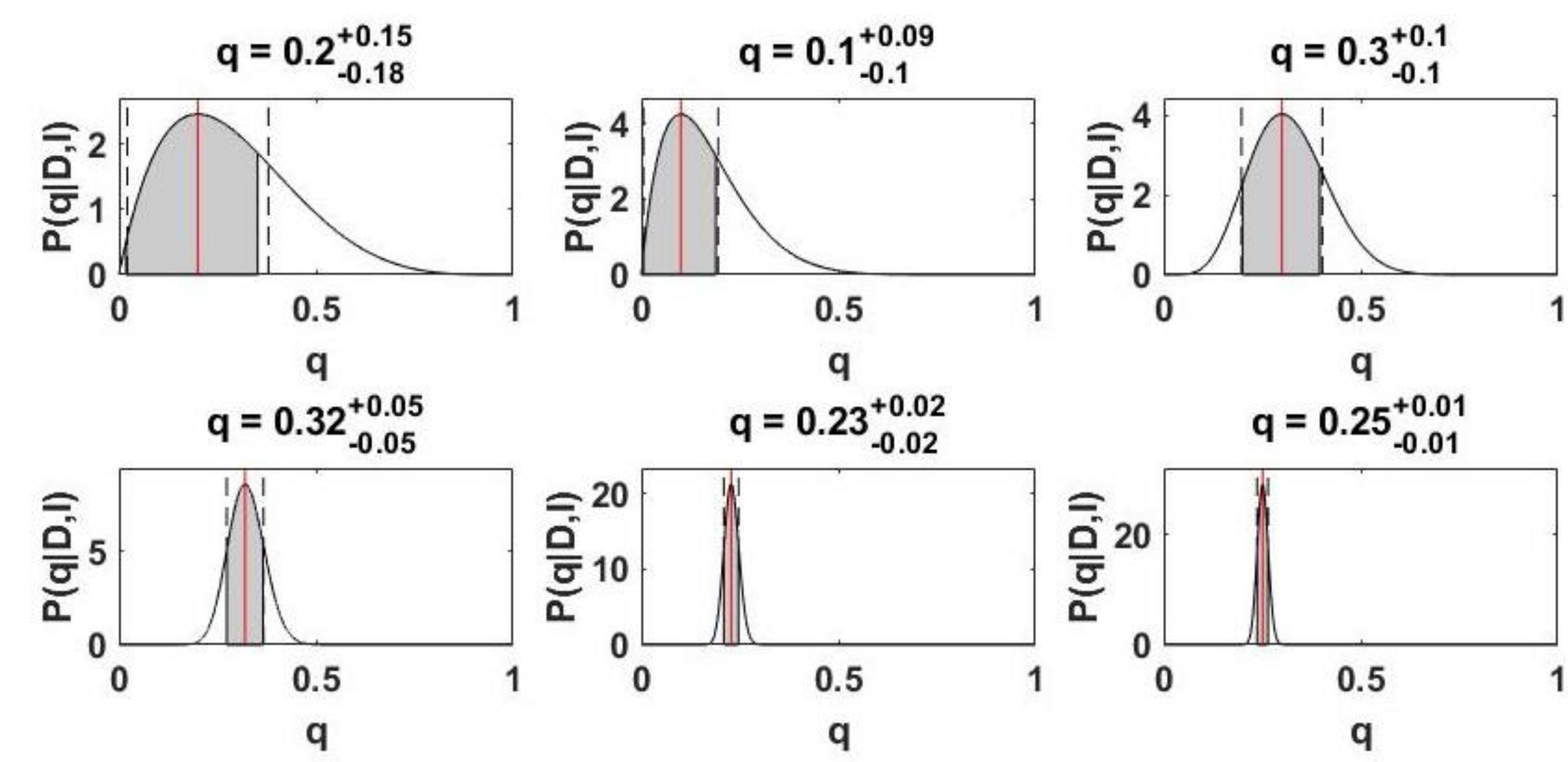
```
[q1, b, _] = bayesian_bino(n1, k1)
```

$$P(q|data\ set) = \frac{q^k(1-q)^{n-k}}{\int_0^1 q^k(1-q)^{n-k}\ dq}$$



check out `bayesian_bino.py`

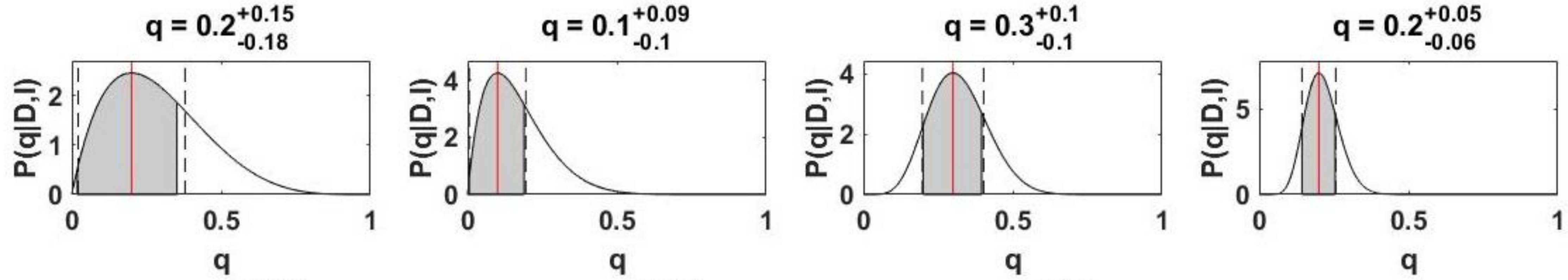
$$P(q|data\ set) = \frac{q^k(1-q)^{n-k}}{\int_0^1 q^k(1-q)^{n-k}\ dq}$$

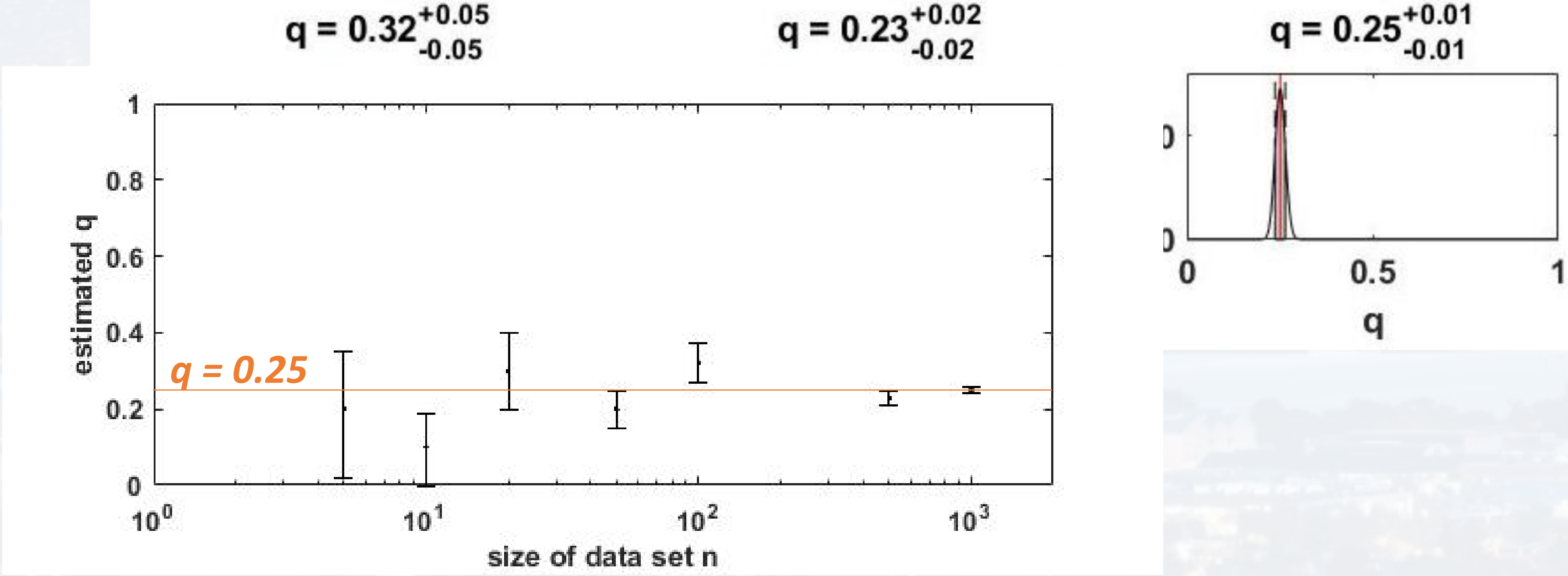


n	estimated q
5	$0.2^{+0.15}_{-0.18}$
10	$0.1^{+0.09}_{-0.1}$
20	$0.3^{+0.1}_{-0.1}$
50	$0.2^{+0.05}_{-0.06}$
100	$0.32^{+0.05}_{-0.05}$
500	$0.23^{+0.02}_{-0.02}$
1,000	$0.25^{+0.01}_{-0.01}$
infinity	0.25

check out `bayesian_bino.py`

$$P(q|data\ set) = \frac{q^k(1-q)^{n-k}}{\int_0^1 q^k(1-q)^{n-k}\ dq}$$





n	estimated q
5	$0.2^{+0.15}_{-0.18}$
10	$0.1^{+0.09}_{-0.1}$
20	$0.3^{+0.1}_{-0.1}$
50	$0.2^{+0.05}_{-0.06}$
100	$0.32^{+0.05}_{-0.05}$
500	$0.23^{+0.02}_{-0.02}$
1,000	$0.25^{+0.01}_{-0.01}$
infinity	0.25

Of course, Bayesian Parameter Estimation works with **any other pdf**

goal: - **P(q|D)**
- the larger ***D***, the more certain ***q***
→ learning

likelihood function

$$P(q|data\ set) = \frac{\boxed{P(data\ set|q)}\boxed{P(q)}}{\boxed{P(data\ set)}}$$

prior

evidence **(const wrt q)**

<u>What is the average number of WhatsUp messages I get every day?</u>

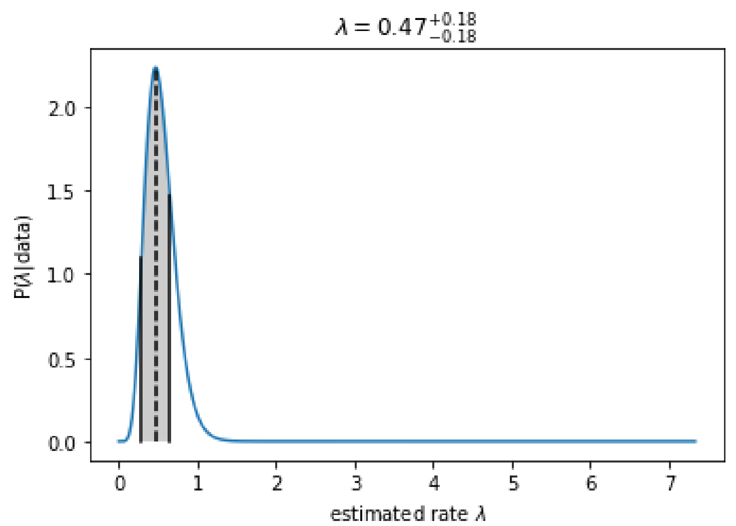
Mon:	5
Tue:	7
Wed:	1
Thu:	3
Fri:	9
Sat:	2
Sun:	5

event - has no duration
- is rare

→ Poissonian

$$P(k|\lambda) = \frac{\lambda^k}{k!}\ e^{-\lambda}$$

```
data = np.random.poisson(lam = 0.4, 15)
poissfit(data)
```



Of course, Bayesian Parameter Estimation works with **any other pdf**

goal: - **P(q|D)**
- the larger ***D***, the more certain ***q***
→ learning

<u>What is the average number of WhatsUp messages I get every day?</u>

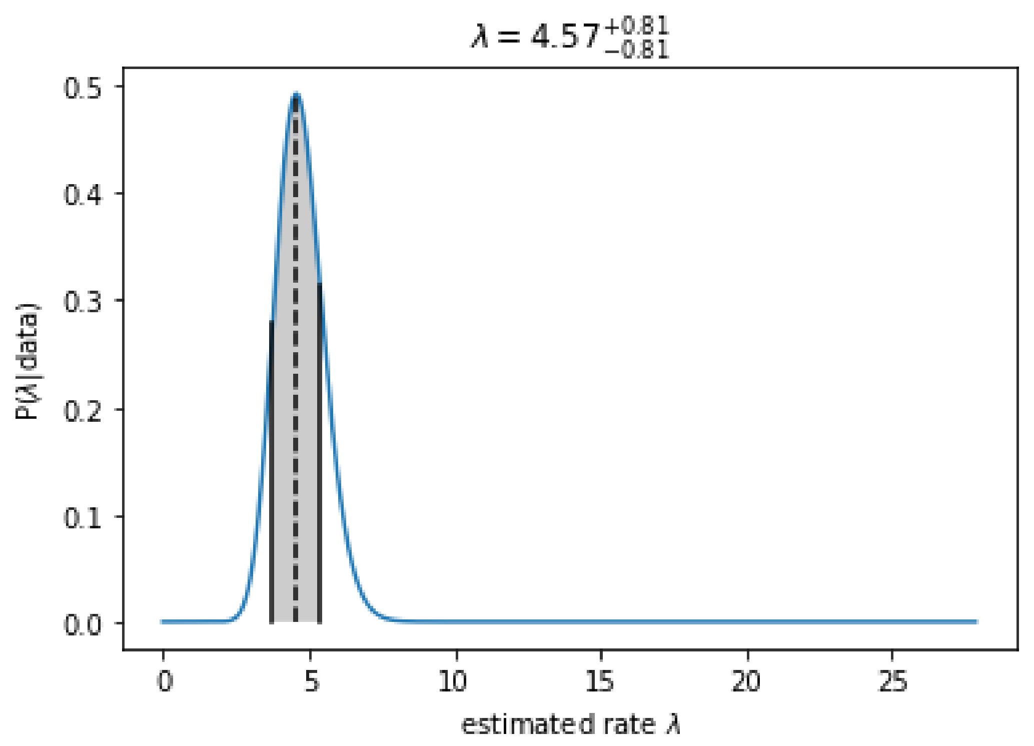
Mon:	5
Tue:	7
Wed:	1
Thu:	3
Fri:	9
Sat:	2
Sun:	5

event - has no duration
- is rare

→ Poissonian

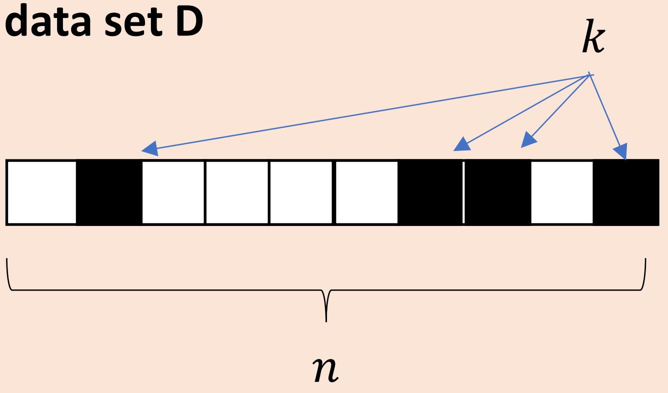
$$P(k|\lambda) = \frac{\lambda^k}{k!}\, e^{-\lambda}$$

```
poissfit([5, 7, 1, 3, 9, 2, 5])
```



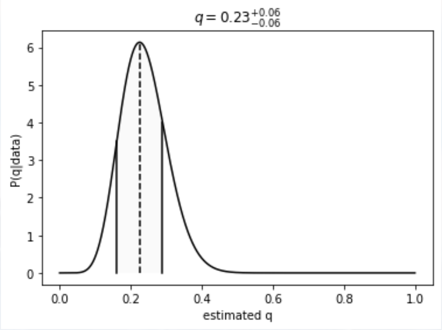
What if there is new data?

data set D

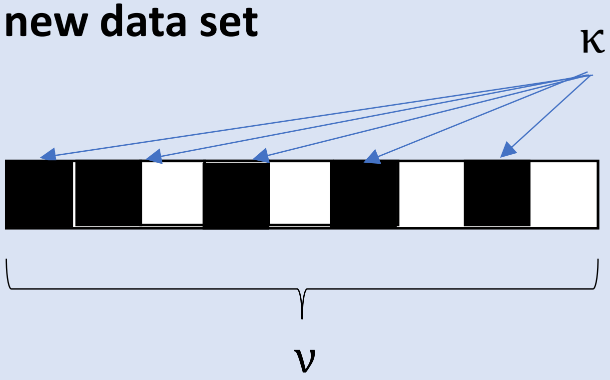


$\cancel{q = ?}$

$$\boxed{P(q|data\ set) = \frac{q^k(1-q)^{n-k}}{\int_0^1 q^k(1-q)^{n-k}\ dq}}$$



new data set



if there **is** prior information ***I*** about *q*:

$$P(q|new\ data\ set, I) = \frac{P(new\ data\ set|q, I)\ \boldsymbol{P(q, I)}}{P(new\ data\ set)}$$

What if there is new data?

$$P(q|data\ set) = \frac{q^k(1-q)^{n-k}}{\int_0^1 q^k(1-q)^{n-k}\ dq}$$

$$P(q|new\ data\ set, I) = \frac{P(new\ data\ set|q, I)\ \boldsymbol{P(q, I)}}{P(new\ data\ set)}$$

$$= \frac{q^{\kappa}(1-q)^{\nu-\kappa}\ q^k(1-q)^{n-k}}{\int_0^1 q^{\kappa}(1-q)^{\nu-\kappa}\ q^k(1-q)^{n-k}\ dq}$$

$$= \frac{q^{k+\kappa}(1-q)^{\nu-\kappa+n-k}}{\int_0^1 q^{k+\kappa}(1-q)^{\nu-\kappa+n-k}\ dq}$$

$$= \frac{q^{k+\alpha-1}(1-q)^{n-k+\beta-1}}{\boxed{\int_0^1 q^{k+\alpha-1}(1-q)^{n-k+\beta-1}\ dq}}$$

often: $\kappa = \alpha - 1$

$\beta = \nu - \kappa - 1$

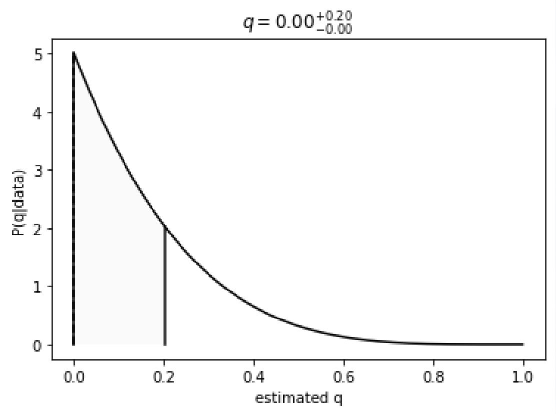
Beta function

What if there is new data?

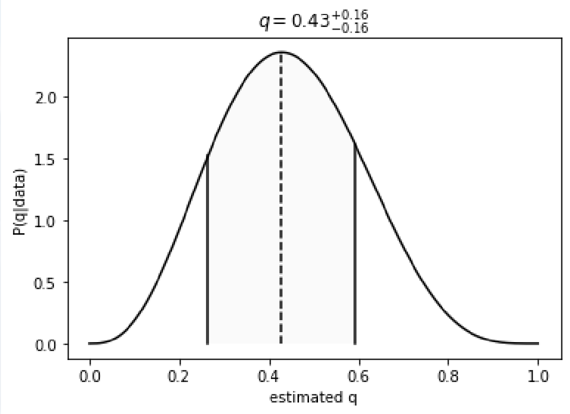
$$P(q|new\ data\ set, I) = \frac{q^{\kappa}(1-q)^{\nu-\kappa}\ q^{k}(1-q)^{n-k}}{\int_0^1 q^{\kappa}(1-q)^{\nu-\kappa}\ q^{k}(1-q)^{n-k}\ dq}$$

```
n1 = 4
k1 = np.random.binomial(n1, q = 0.2)
[_, _, Prior] = bayesian_bino(n1, k1)
```

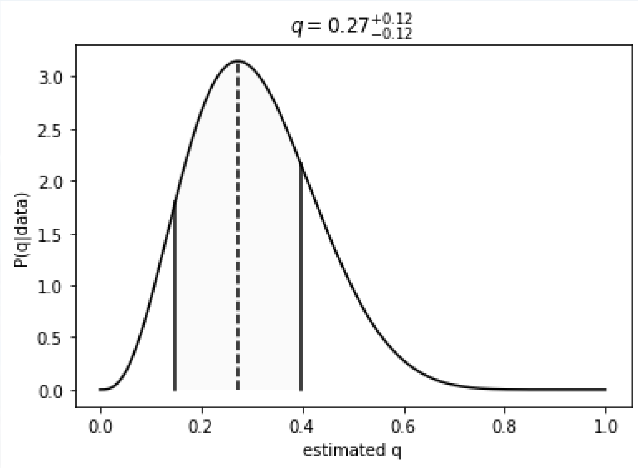
```
n2 = 7
k2 = np.random.binomial(n2, q = 0.2)
[_, _, _] = bayesian_bino(n2, k2)
```



$$P(q, I) = \frac{q^{k}(1-q)^{n-k}}{\int_0^1 q^{k}(1-q)^{n-k}\ dq}$$



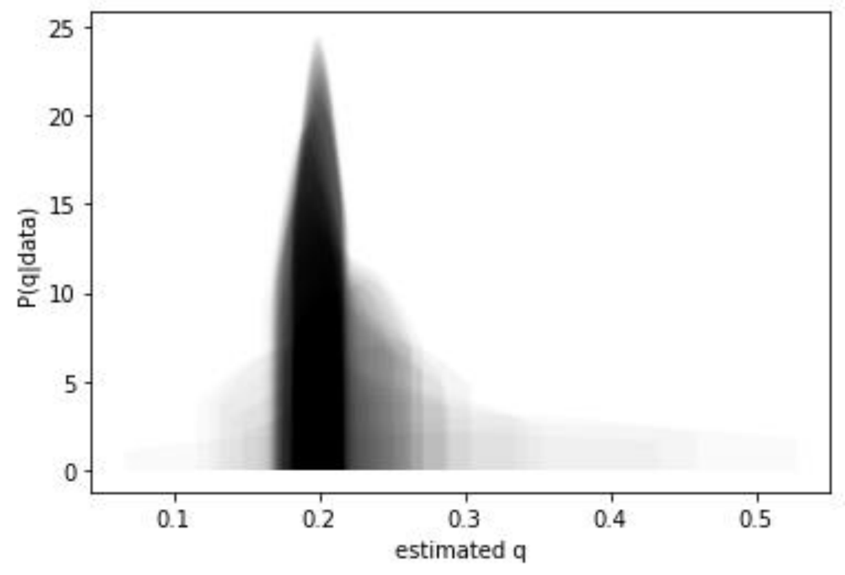

```
[_ , _, _] = bayesian_bino(n2, k2, Prior = Prior)
```



What if there is new data?

$$P(q|new\ data\ set, I) = \frac{q^{\kappa}(1-q)^{\nu-\kappa}\ q^{k}(1-q)^{n-k}}{\int_0^1 q^{\kappa}(1-q)^{\nu-\kappa}\ q^{k}(1-q)^{n-k}\ dq}$$

The **posterior** from the **previous** experiment is the **prior** of the **next** experiment



→ we become more certain about the model parameters
→ learning!

→ see e.g. **V**ariational **A**uto **E**ncoders

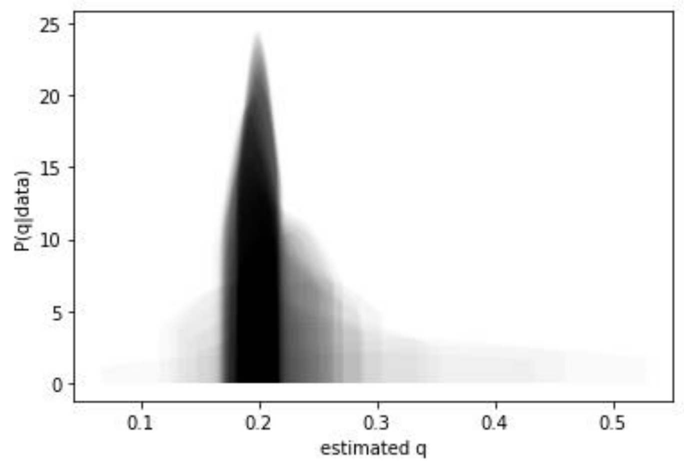
2D images → 3D objects

credit: StableAI

What if there is new data?

$$P(q|new\ data\ set, I) = \frac{q^{\kappa}(1-q)^{\nu-\kappa}\ q^{k}(1-q)^{n-k}}{\int_0^1 q^{\kappa}(1-q)^{\nu-\kappa}\ q^{k}(1-q)^{n-k}\ dq}$$

The **posterior** from the **previous** experiment is the **prior** of the **next** experiment

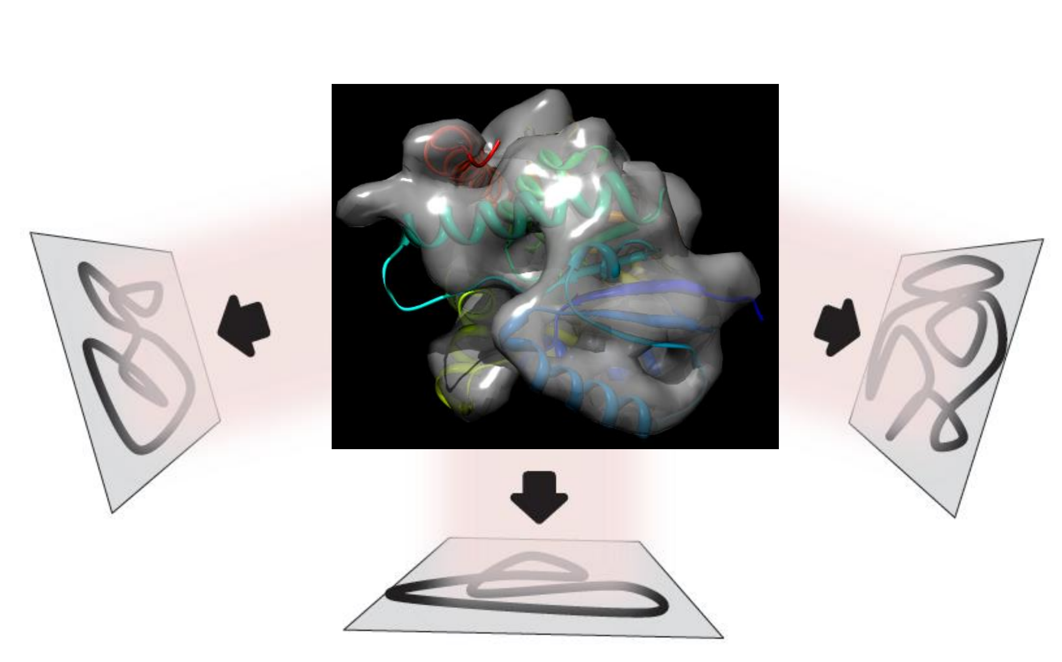


→ we become more certain about the model parameters
→ learning!

→ see e.g. **V**ariational **A**uto **E**ncoders

2D images → 3D objects

Cryo – EM: 3D structure from 2D images/projections



(image courtesy: Thomas Becker, GC LMU Munich)

Thank you very much for your attention!