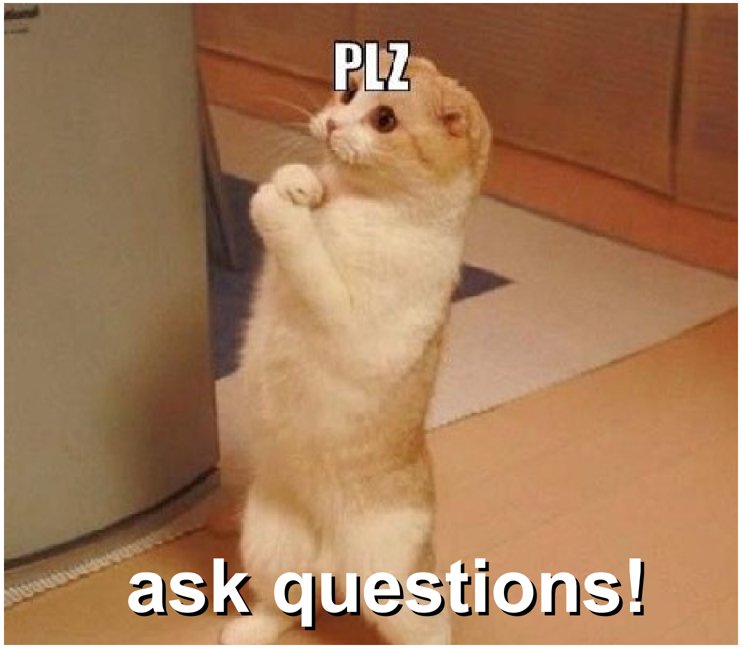
Data Mining & Machine Learning

Dipl.-Inf. Christoph Carl Kling

WeST
People and Knowledge Networks

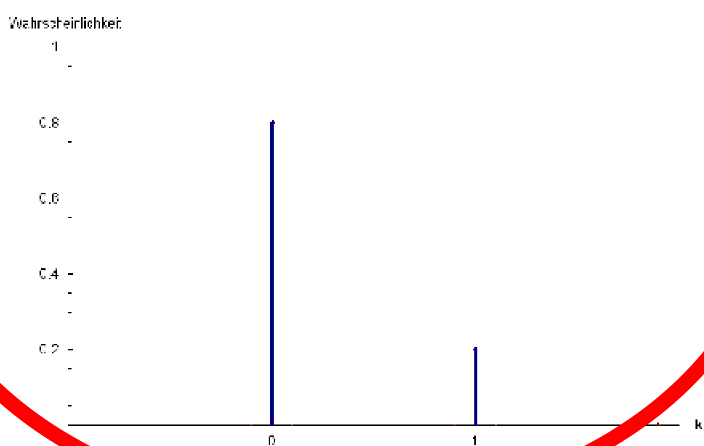


DM@C-Kling.de

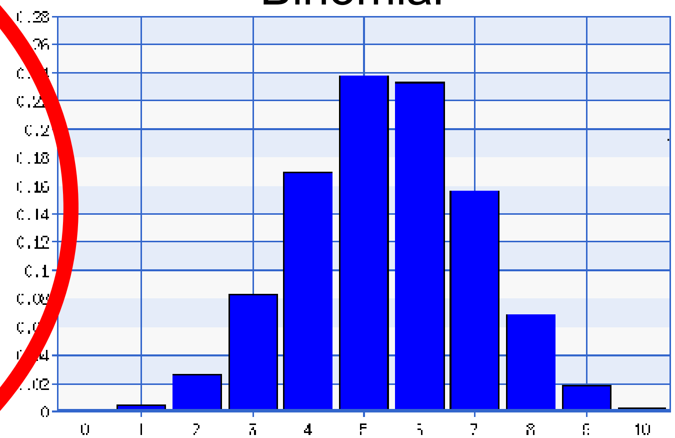
Probability Theory

	n = 1	n >= 1	n → ∞
k = 2	Bernoulli = Binomial for n = 1	Binomial	Gaussian
k > 2	Multinomial for n = 1	Multinomial	Mulivariate Gaussian

Bernoulli = Binomial for n = 1

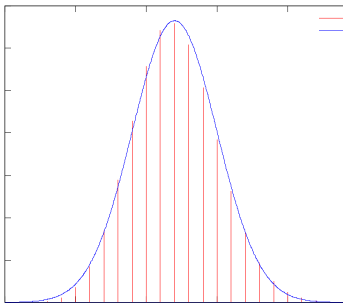


Binomial

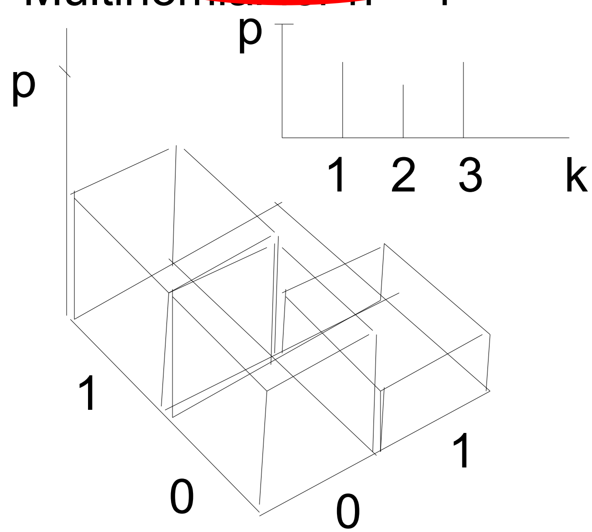


number of successes

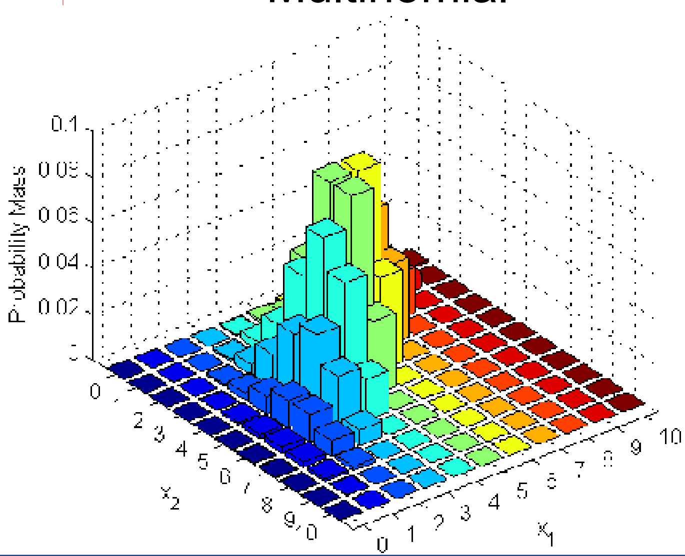
Gaussian



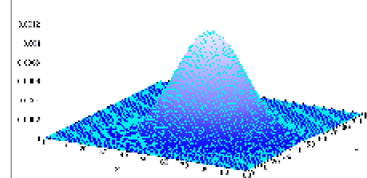
Multinomial for n = 1



Multinomial



Mulivariate Gaussian



Observations **c** (our Data)
Hidden (latent) parameter **p**

Example: tossing a coin: 2 x head, 0 x tail

$$c_i \in \{0, 1\}, i = 1, ..., n$$

tail head

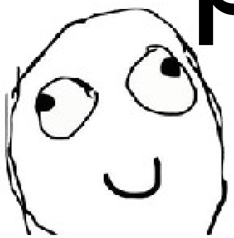
$$(c_1 = 1, \quad c_2 = 1)$$

$$p(p|\mathbf{c}) = \frac{p(\mathbf{c}|p) \cdot p(p)}{p(\mathbf{c})}$$

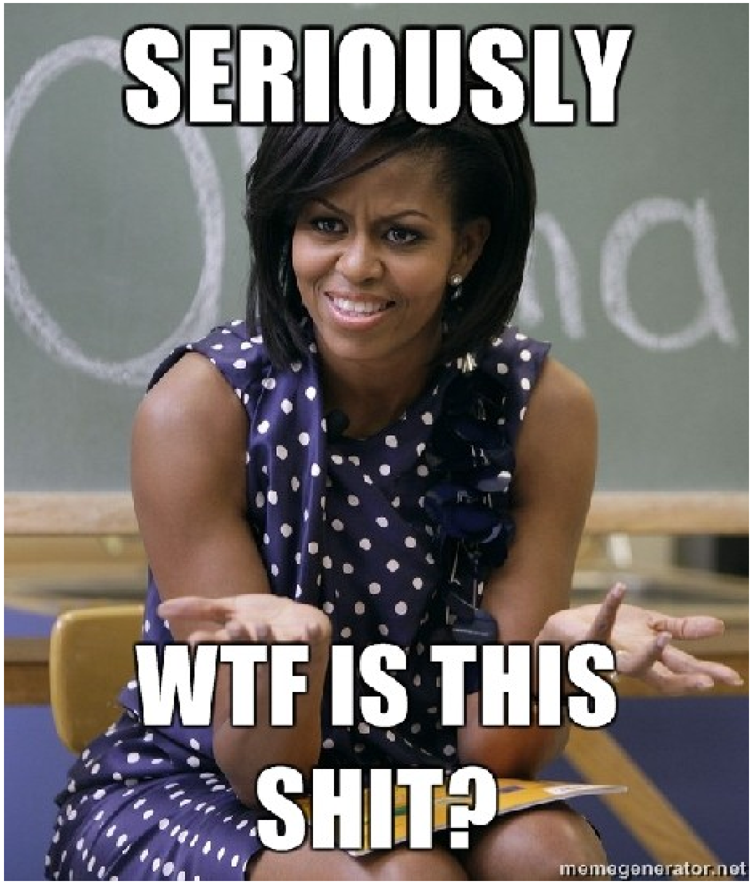
$$posterior = \frac{likelihood \cdot prior}{evidence}$$

Maximum likelihood estimation (MLE)

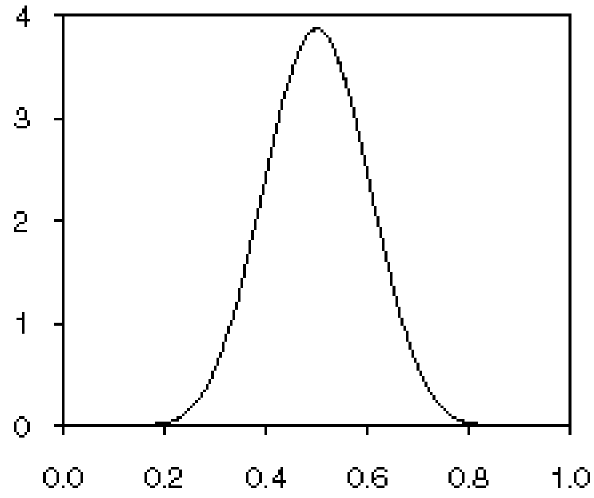
$p = 1.0$!



p = 1.0



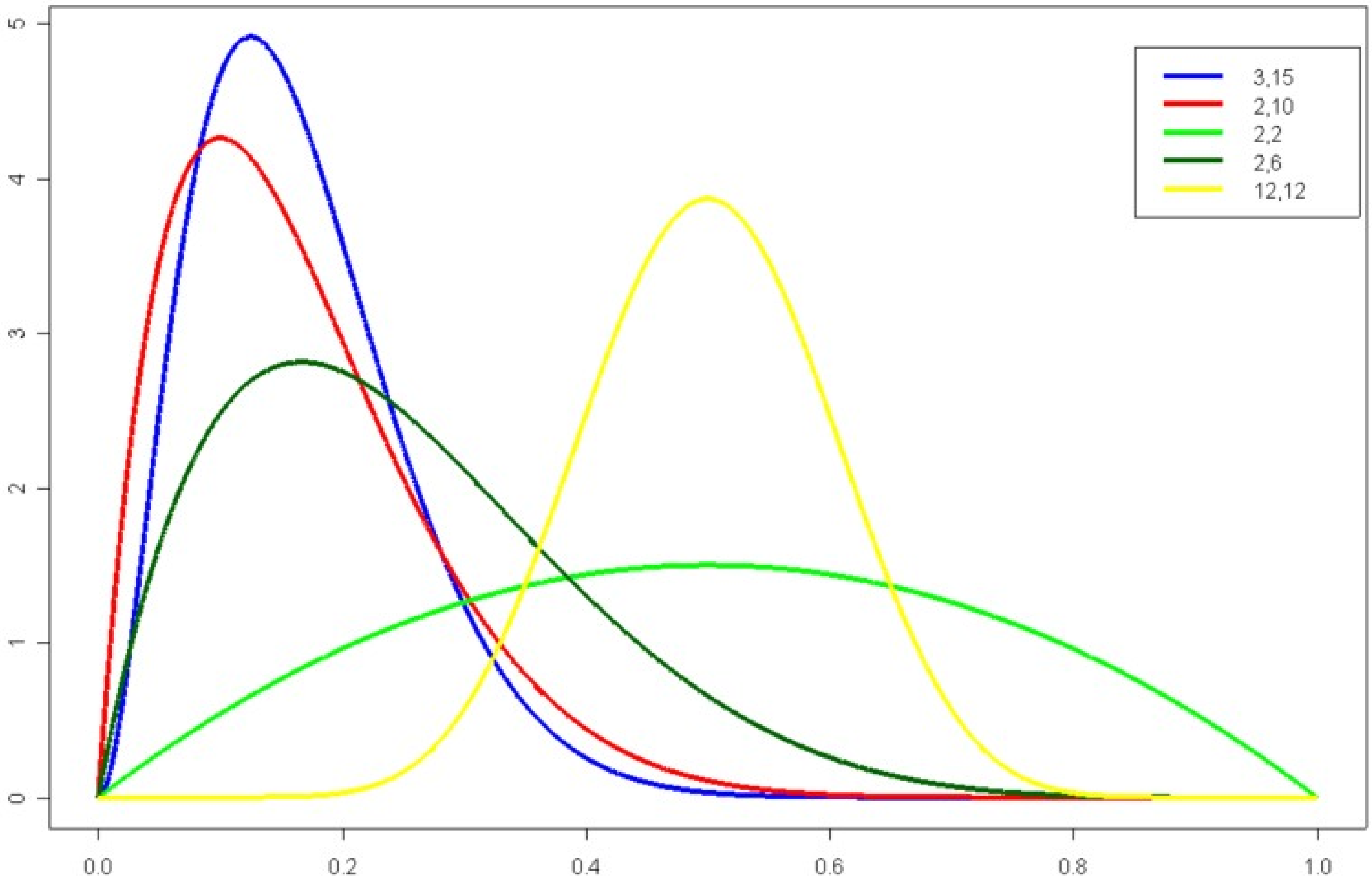
p more likely is close to 0.5!



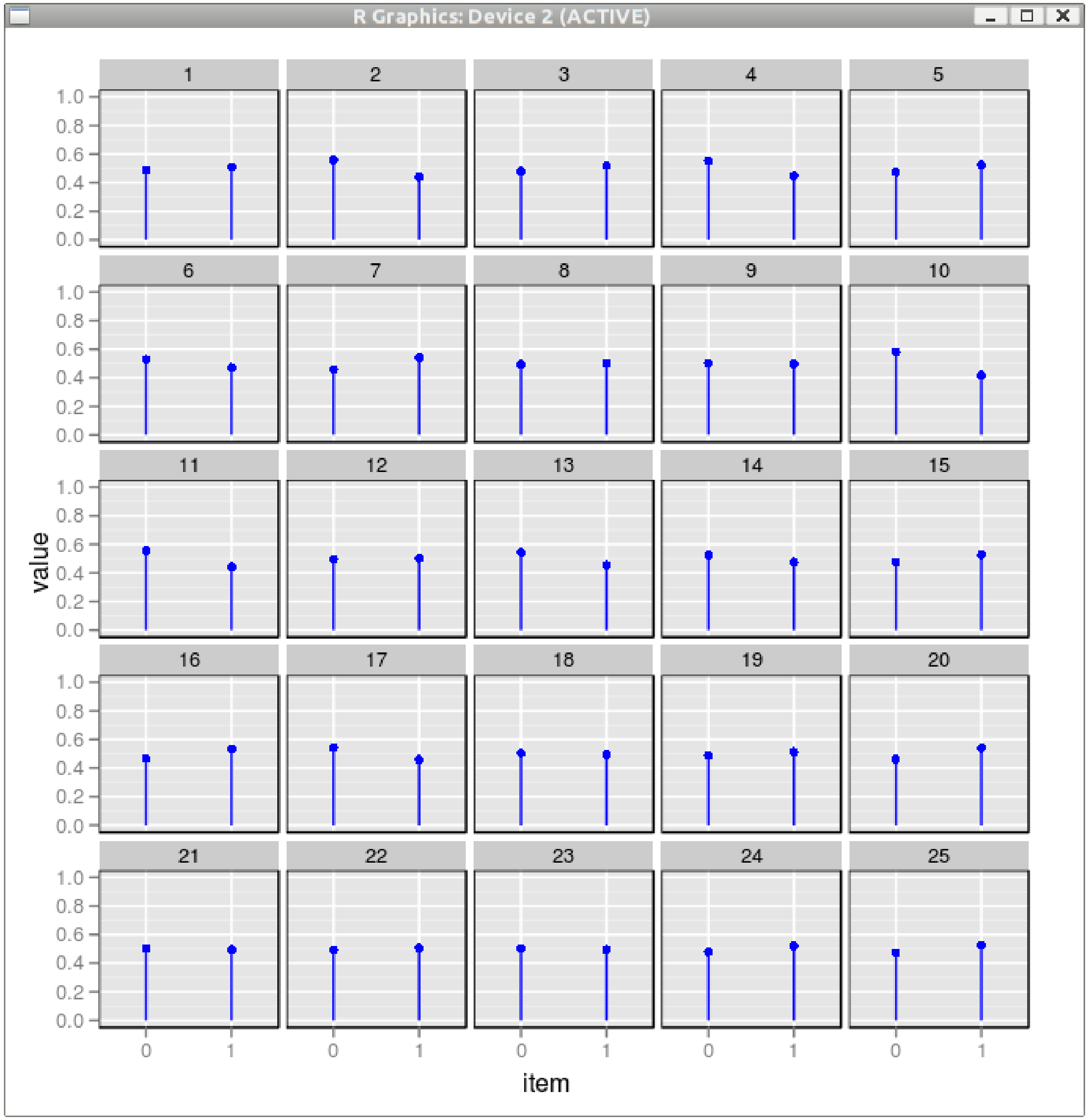
Prior probability $p(p) = Beta(\alpha, \beta)$

Beta distribution

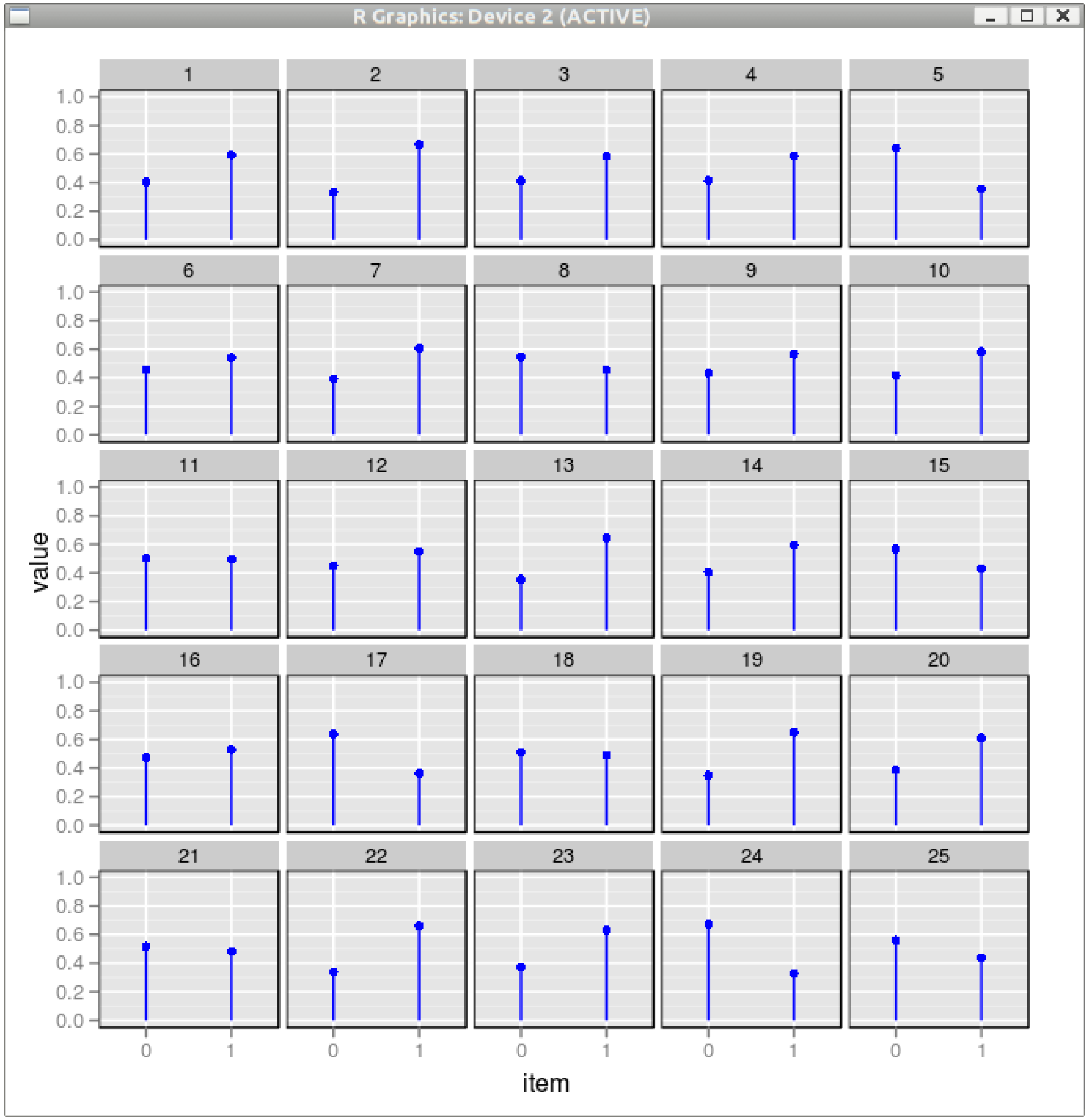
Density of $Beta(\alpha, \beta)$



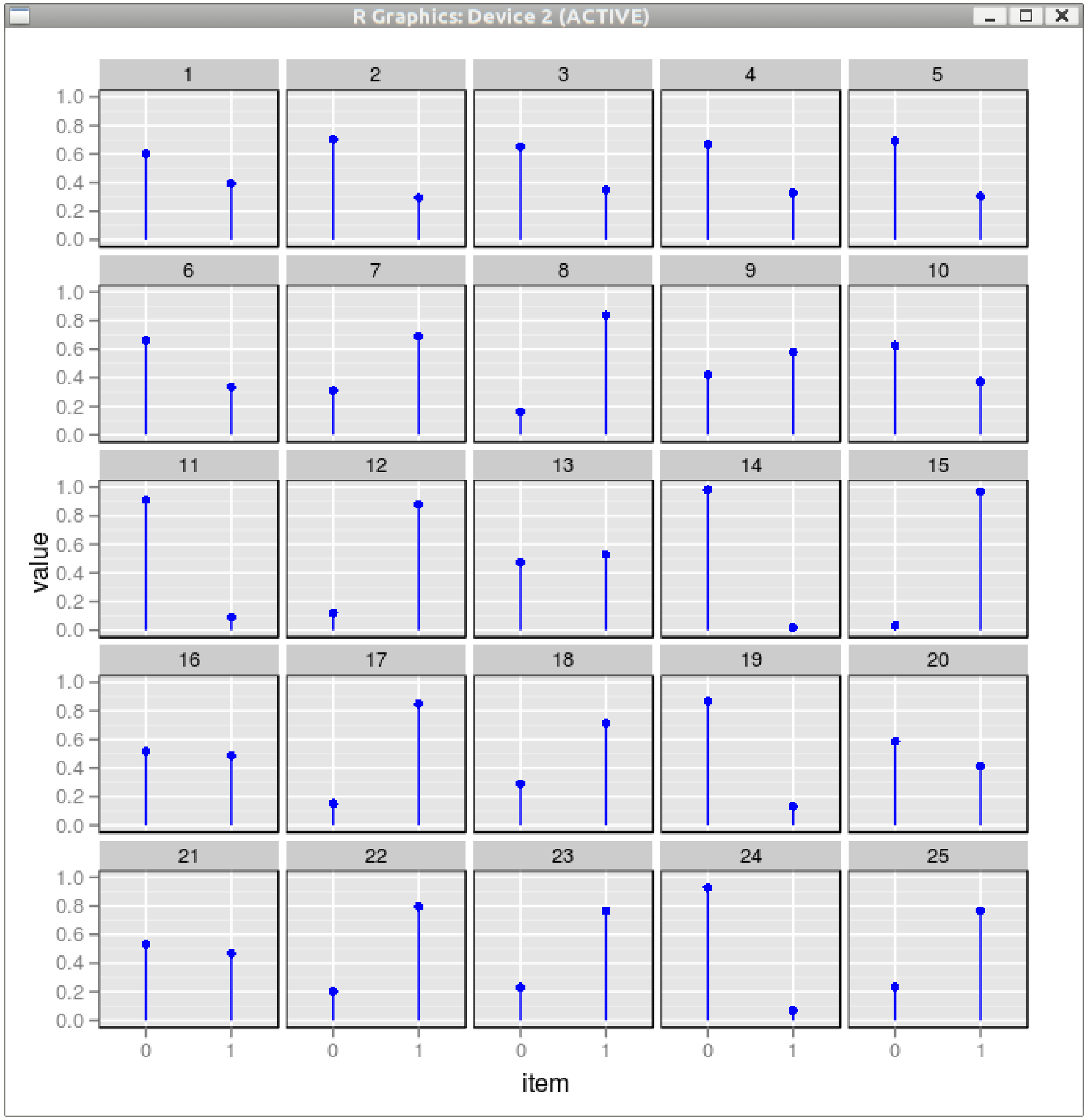
Beta(100,100)


R Graphics: Device 2 (ACTIVE)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
1.0
0.8
0.6
0.4
0.2
0.0
value
0
1
item

Beta(10,10)

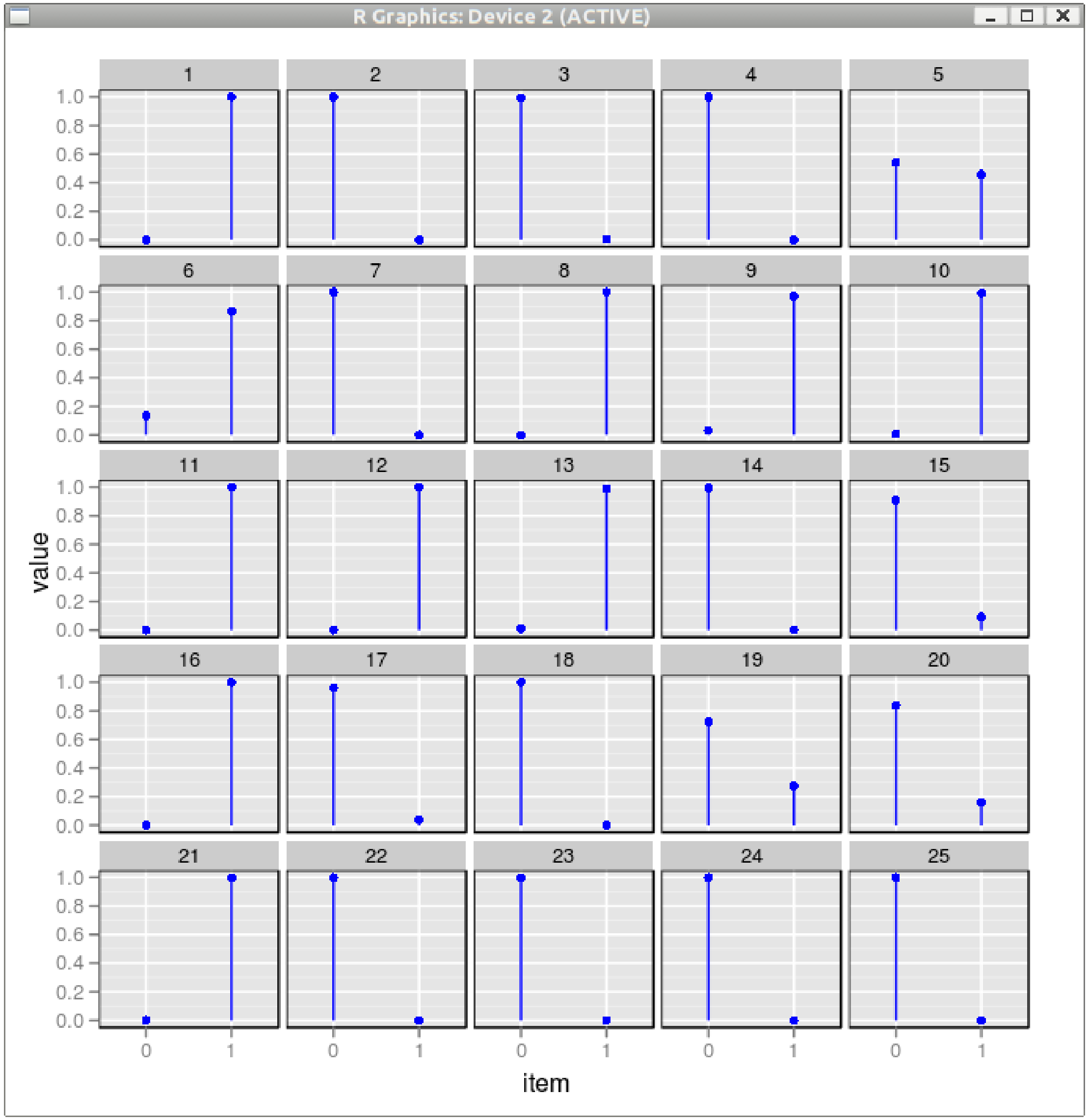

R Graphics: Device 2 (ACTIVE)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
value
0.0
0.2
0.4
0.6
0.8
1.0
item
0
1

Beta(1,1)



R Graphics: Device 2 (ACTIVE)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
value
1.0
0.8
0.6
0.4
0.2
0.0
0
1
item

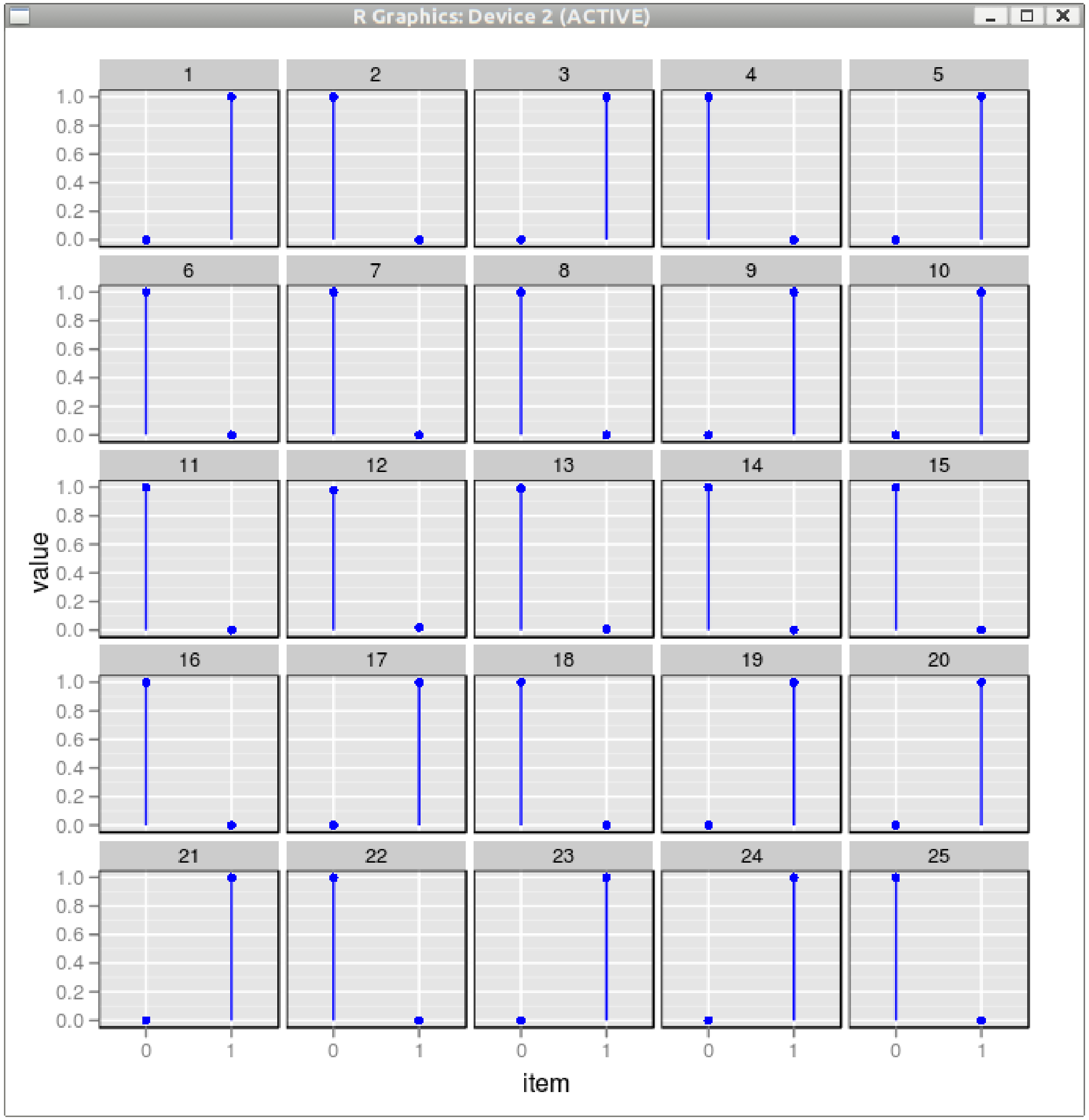
Beta(0.1,0.1)



R Graphics: Device 2 (ACTIVE)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
value
1.0
0.8
0.6
0.4
0.2
0.0
0
1
item

Beta distribution

Beta(0.01,0.01)



Maximum a posteriori estimation (MAP)

Bayesian inference

Maximum a posteriori estimation (MAP)

Bayesian inference

Lineare Regression

y = Größe	x1 = Geschlecht	x2 = Gewicht
168	1	65
172	0	80
164	1	52
187	0	120
194	0	90