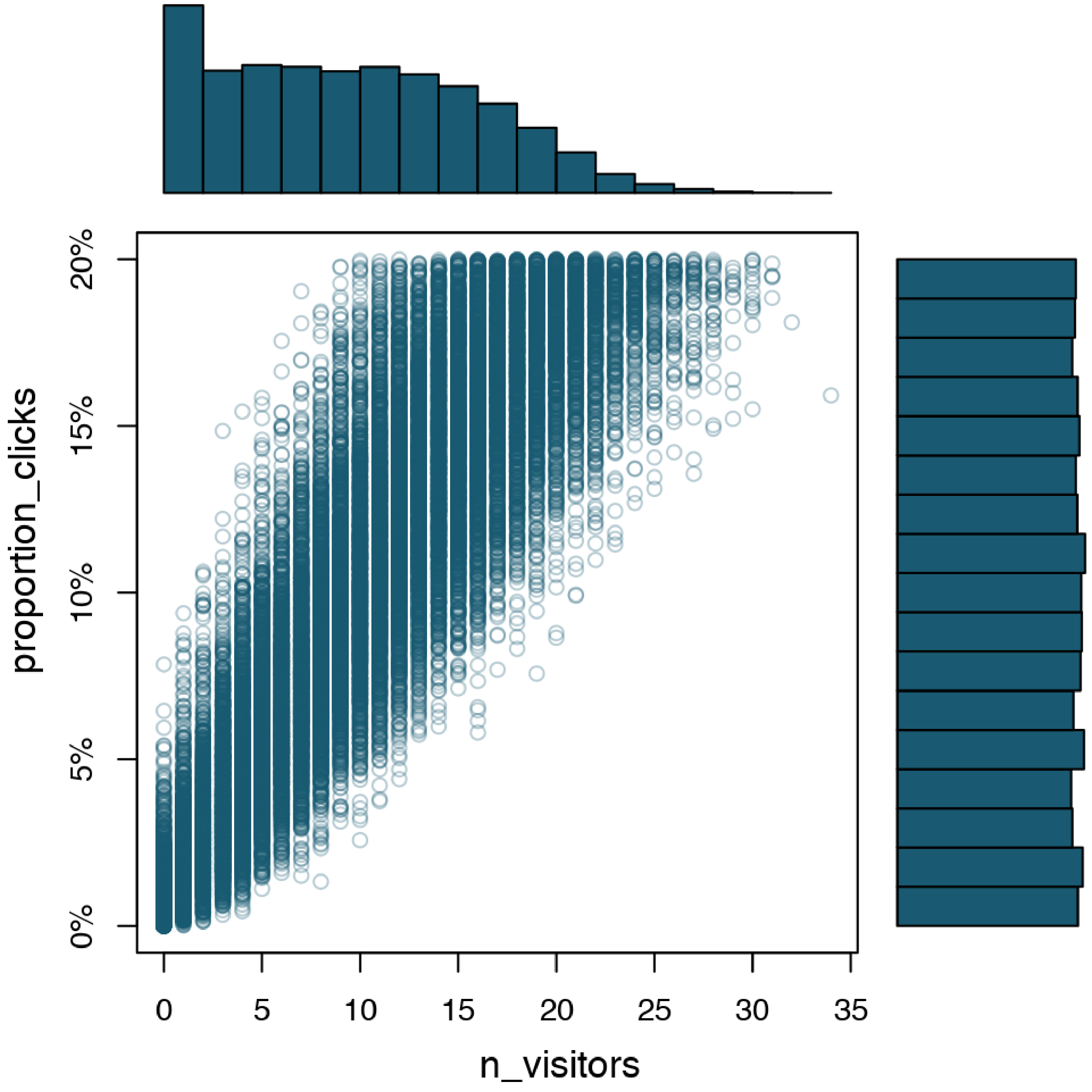
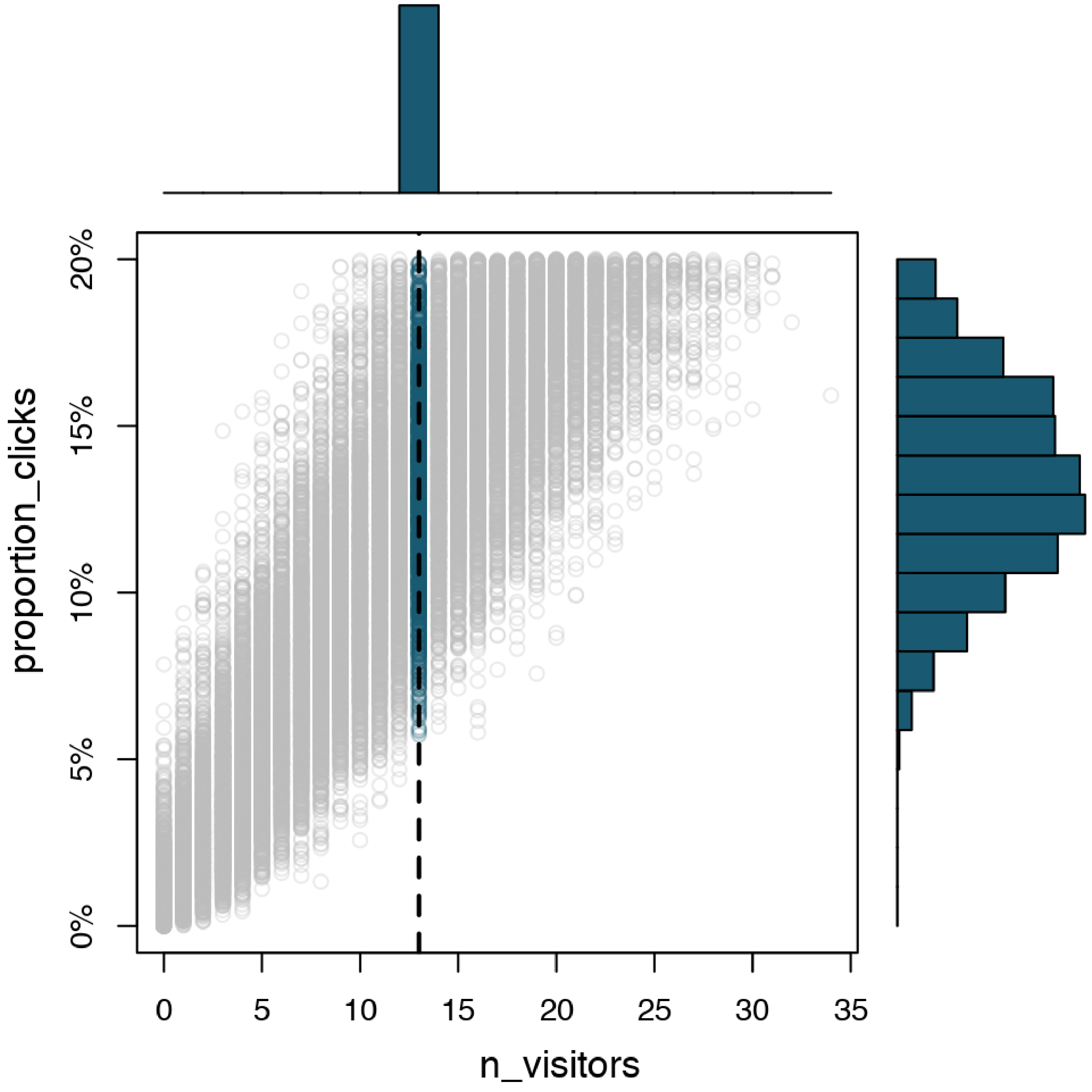
Probability rules

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Rasmus Bååth
Data Scientist



20%
15%
10%
5%
0%
proportion_clicks
0
5
10
15
20
25
30
35
n_visitors


proportion_clicks
20%
15%
10%
5%
0%
0
5
10
15
20
25
30
35
n_visitors

Bad and good news

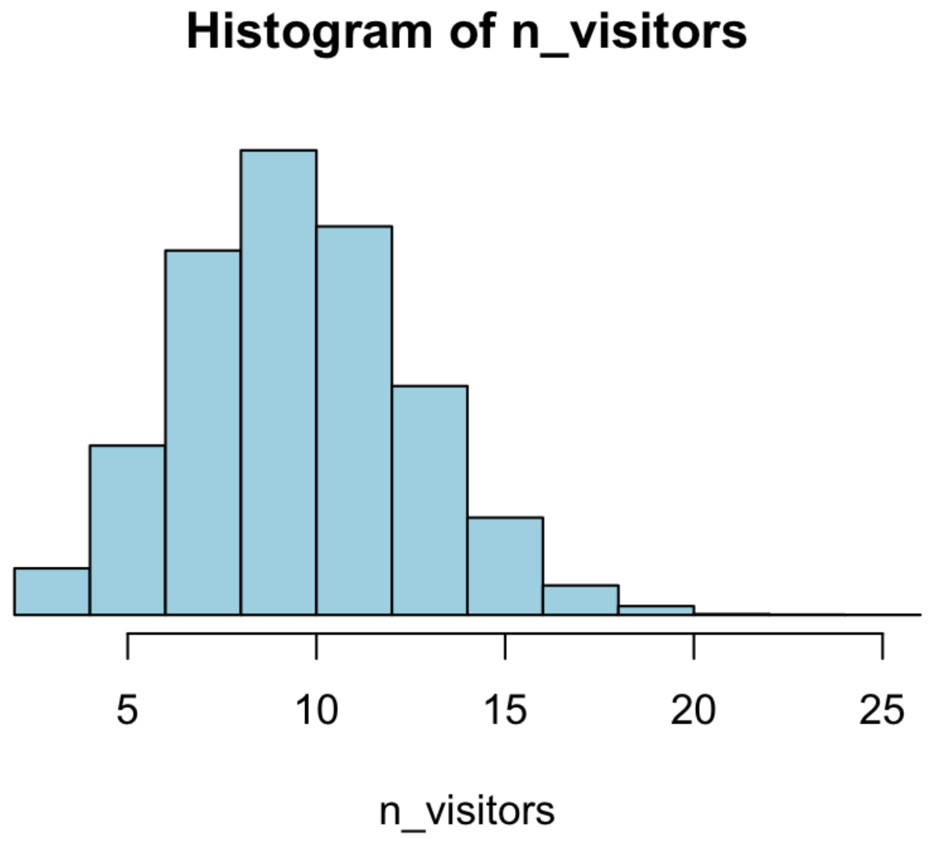
- Bad news
  - The computation method we've used scales horribly.
- Good news
  - Bayesian computation is a hot research topic.
  - There are many methods to fit Bayesian models more efficiently.
  - The result will be the same, you'll just get it faster.

Probability theory

- Probability
  - A number between 0 and 1.
  - A statement of certainty/uncertainty.
- Mathematical notation:
  - P(`n_visitors` = 13) is a probability
  - P(`n_visitors`) is a probability *distribution*
  - P(`n_visitors` = 13 | `prop_clicks` = 10%) is a *conditional* probability
  - P(`n_visitors` | `prop_clicks` = 10%) is a *conditional* probability *distribution*

P(n_visitors | prop_clicks = 10%)

```r
n_visitors <- rbinom(n = 10000, size = 100, prob = 0.1)
hist(n_visitors)
```



Manipulating probability

Manipulating probability

- The sum rule

Manipulating probability

- The sum rule
  - p(1 or 2 or 3)

Manipulating probability

- The sum rule
  - p(1 or 2 or 3) = 1/6 + 1/6 + 1/6 = 0.5

Manipulating probability

- The sum rule
    - p(1 or 2 or 3) = 1/6 + 1/6 + 1/6 = 0.5
- The product rule

Manipulating probability

- The sum rule
  - p(1 or 2 or 3) = 1/6 + 1/6 + 1/6 = 0.5
- The product rule
  - p(6 and 6)

Manipulating probability

- The sum rule
  - p(1 or 2 or 3) = 1/6 + 1/6 + 1/6 = 0.5
- The product rule
  - p(6 and 6) = 1/6 * 1/6 = 1 / 36 = 2.8%

Manipulating probability

- The sum rule
  - p(1 or 2 or 3) = 1/6 + 1/6 + 1/6 = 0.5
- The product rule
  - p(6 and 6) = 1/6 * 1/6 = 1 / 36 = 2.8%

Let's try out these rules!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

We can calculate!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Rasmus Bååth
Data Scientist

Simulation vs calculation

- Simulation using 'r'-functions, for example, `rbinom` and `rpois`
- Simulating P(`n_visitors` = 13 | `prob_success` = 10%)

```
n_visitors <- rbinom(n = 100000, size = 100, prob = 0.1)
sum(n_visitors == 13) / length(n_visitors)
```

```
0.074
```

- Calculation using the 'd'-functions, for example, `dbinom` and `dpois`
- Calculating P(`n_visitors` = 13 | `prob_success` = 10%)

```
dbinom(13, size = 100, prob = 0.1)
```

```
0.074
```

- Calculating P(`n_visitors` = 13 or `n_visitors` = 14 | `prob_success` = 10%)

```
dbinom(13, size = 100, prob = 0.1) + dbinom(14, size = 100, prob = 0.1)
```

```
0.126
```

- Calculating P(`n_visitors` | `prop_success` = 10%)

```
n_visitors = seq(0, 100, by = 1)
probability <- dbinom(n_visitors, size = 100, prob = 0.1)
```

```
n_visitors
```

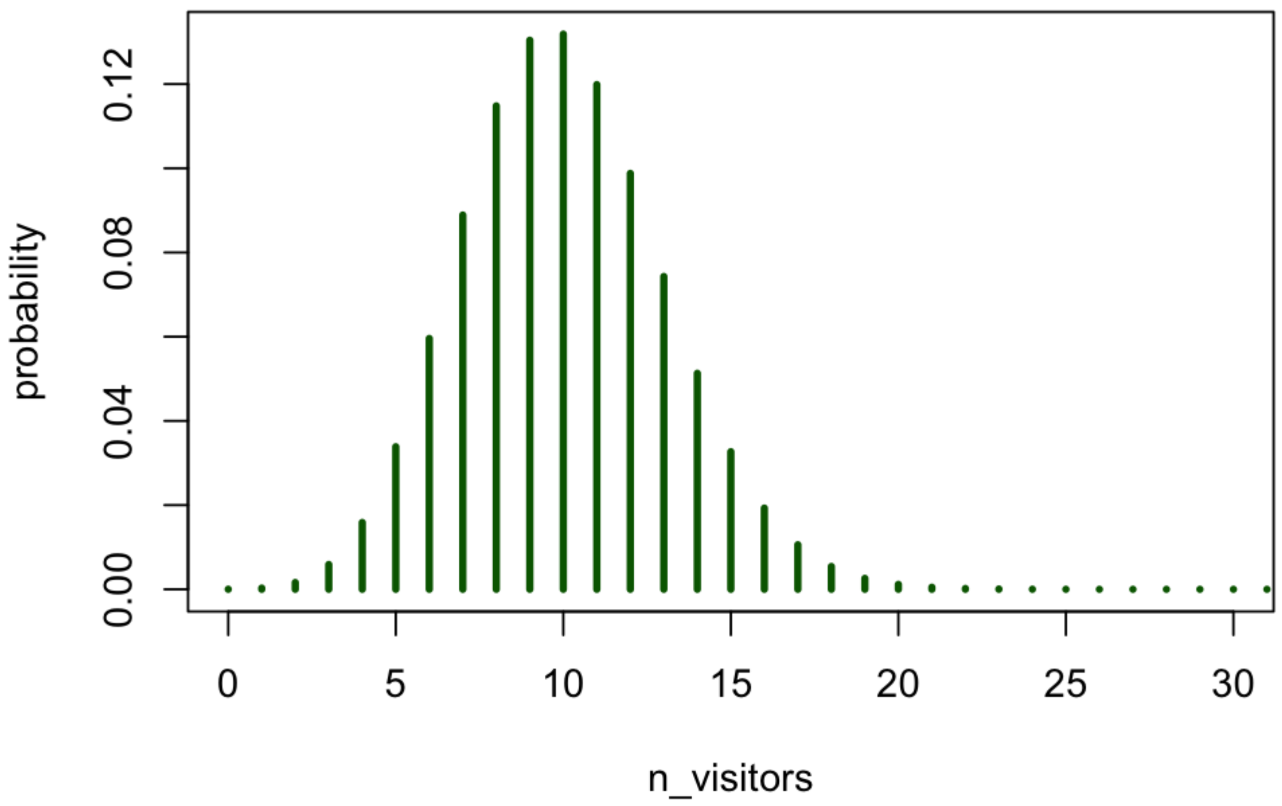
```
0 1 2 3 4 5 6 7 ...
```

```
probability
```

```
0.000 0.000 0.002 0.006 0.016 0.034 0.060 0.089 ...
```

Plotting a calculated distribution

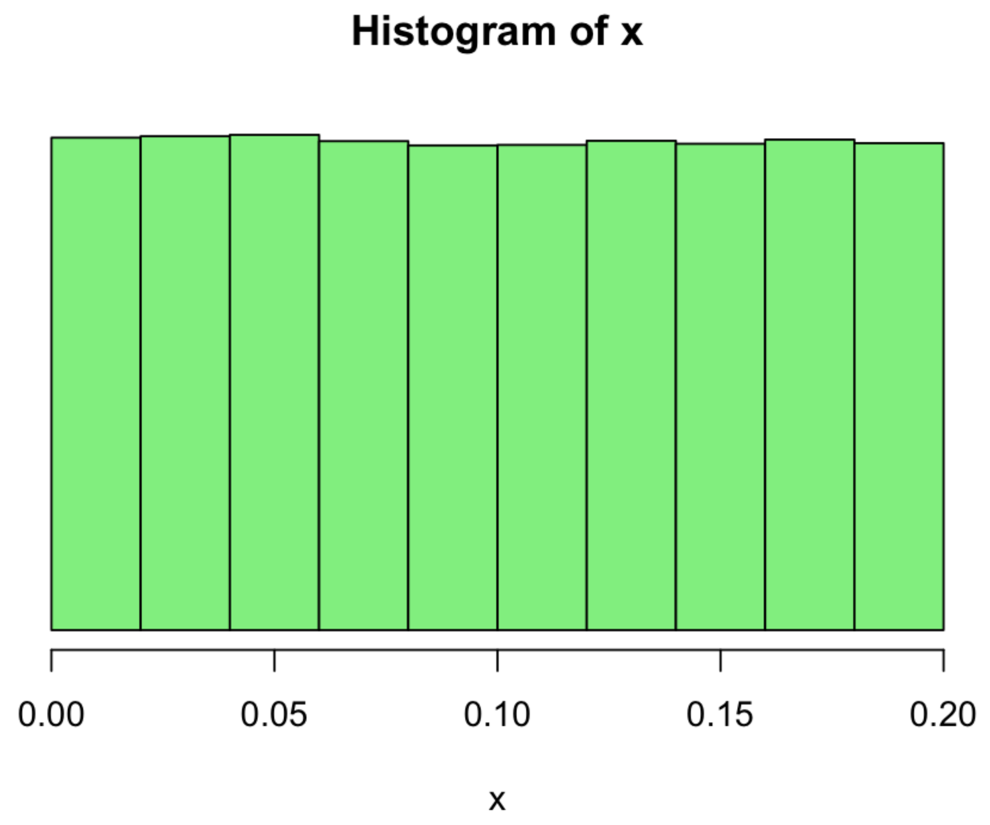
```
plot(n_visitors, probability, type = "h")
```



Continuous distributions

- The Uniform distribution

```
x <- runif(n = 100000, min = 0.0, max = 0.2)
hist(x)
```



Continuous distributions

- The Uniform distribution
  - The d-version of `runif` is `dunif` :

```
dunif(x = 0.12, min = 0.0, max = 0.2)
```

```
5
```

- Probability *density*: Kind of a relative probability

```
x = seq(0, 0.2, by=0.01)
dunif(x, min = 0.0, max = 0.2)
```

```
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
```

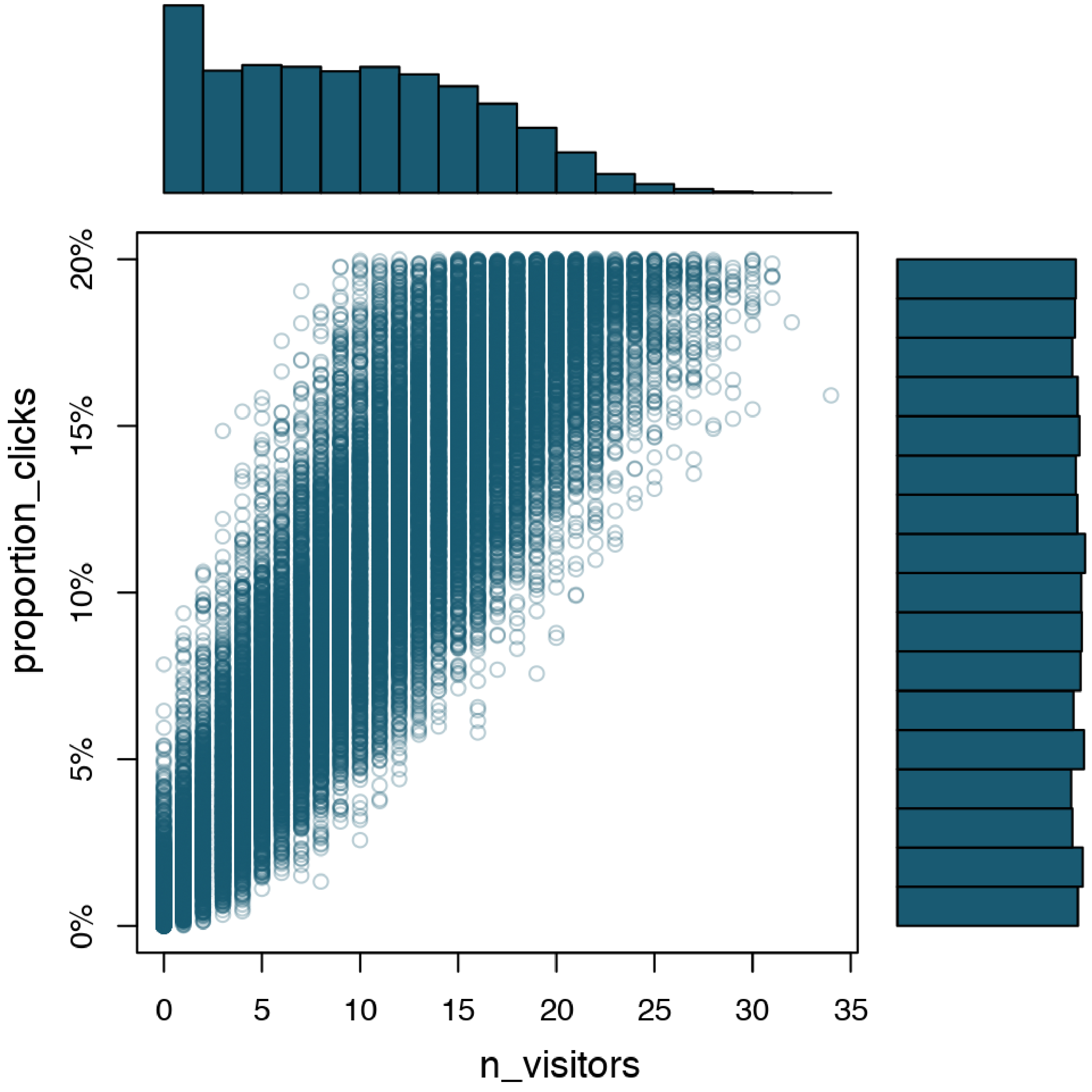
Try this out!

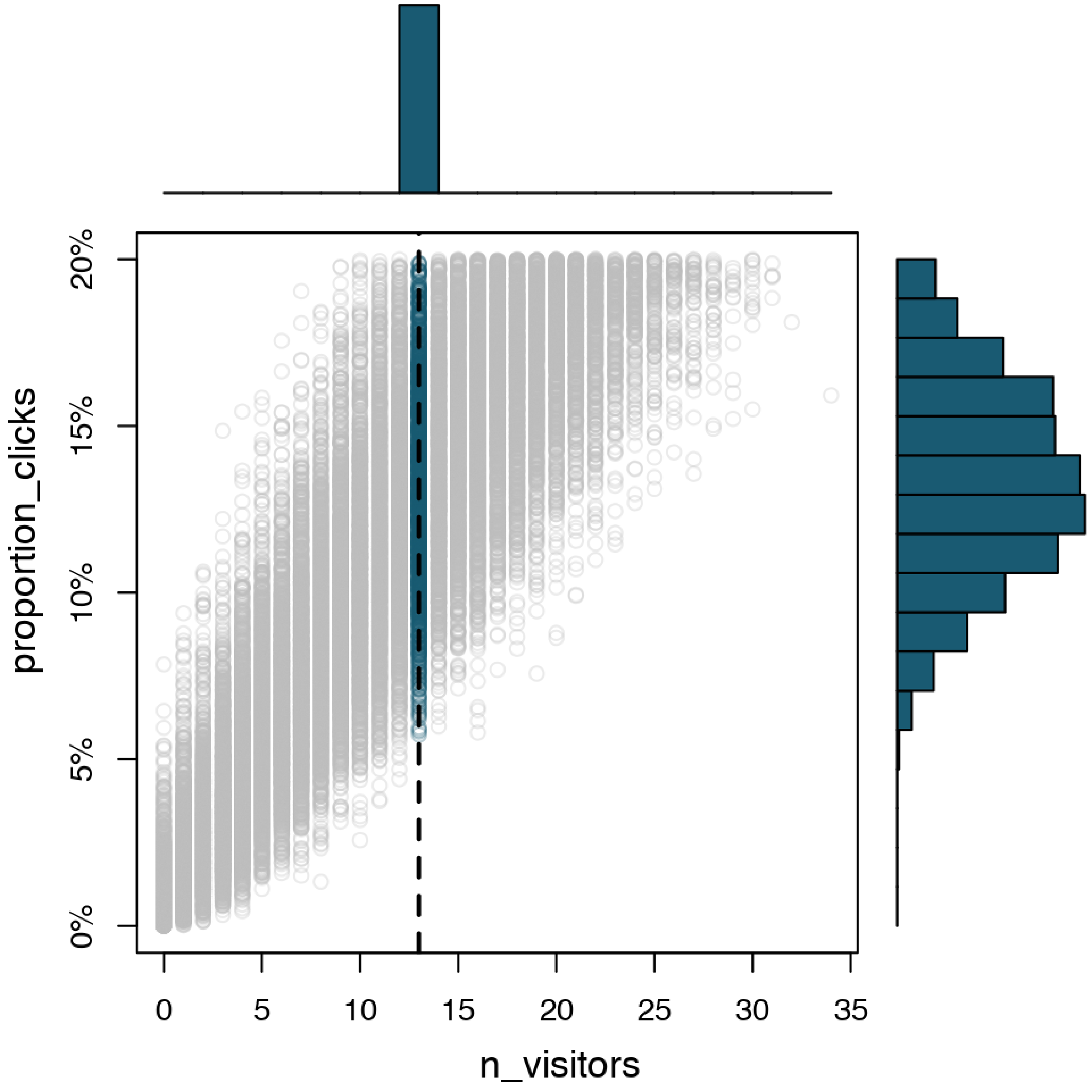
FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Bayesian calculation

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Rasmus Bååth
Data Scientist


proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors


proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors

Bayesian inference by calculation

Bayesian inference by calculation

```
n_ads_shown <- 100
```

Bayesian inference by calculation

```
n_ads_shown <- 100
n_visitors
proportion_clicks
```

Bayesian inference by calculation

```
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks
```

Bayesian inference by calculation

```
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
```

Bayesian inference by calculation

```
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
```

```
proportion_clicks n_visitors
             0.04         38
             0.11         93
             0.16        100
             0.67         98
             0.96          3
             0.48         73
             0.14         13
              ...        ...
```

Bayesian inference by calculation

```
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
proportion_clicks <- runif(n_samples, min = 0.0, max = 0.2)
```

```
proportion_clicks n_visitors
             0.04         38
             0.11         93
             0.16        100
             0.67         98
             0.96          3
             0.48         73
             0.14         13
             ...         ...
```

Bayesian inference by calculation

```
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
```

```
proportion_clicks n_visitors prior
             0.04         38     5
             0.11         93     5
             0.16        100     5
             0.67         98     0
             0.96          3     0
             0.48         73     0
             0.14         13     5
             ...          ...   ...
```

Bayesian inference by calculation

```r
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
n_visitors <- rbinom(n = n_samples, size = n_ads_shown,
                     prob = proportion_clicks)
```

```
proportion_clicks n_visitors prior
             0.04         38     5
             0.11         93     5
             0.16        100     5
             0.67         98     0
             0.96          3     0
             0.48         73     0
             0.14         13     5
             ...         ...    ...
```

Bayesian inference by calculation

```r
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(pars$n_visitors, size = n_ads_shown,
                          prob = pars$proportion_clicks)
```

```
proportion_clicks n_visitors prior    likelihood
             0.04         38     5  3.409439e-27
             0.11         93     5  5.006969e-80
             0.16        100     5  2.582250e-80
             0.67         98     0  4.863666e-15
             0.96          3     0 3.592054e-131
             0.48         73     0  2.215148e-07
             0.14         13     5  1.129620e-01
             ...         ...   ...           ...
```

Bayesian inference by calculation

```
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(pars$n_visitors, size = n_ads_shown,
                          prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
```

```
proportion_clicks n_visitors prior    likelihood  probability
             0.04         38     5  3.409439e-27 1.704720e-26
             0.11         93     5  5.006969e-80 2.503485e-79
             0.16        100     5  2.582250e-80 1.291125e-79
             0.67         98     0  4.863666e-15 0.000000e+00
             0.96          3     0 3.592054e-131 0.000000e+00
             0.48         73     0  2.215148e-07 0.000000e+00
             0.14         13     5  1.129620e-01 5.648101e-01
             ...         ...   ...           ...          ...
```

Bayesian inference by calculation

```
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(pars$n_visitors, size = n_ads_shown,
                          prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
sum(pars$probability)
```

```
105
```

```
proportion_clicks n_visitors prior    likelihood  probability
             0.04         38     5  3.409439e-27 1.704720e-26
             0.11         93     5  5.006969e-80 2.503485e-79
             0.16        100     5  2.582250e-80 1.291125e-79
             0.67         98     0  4.863666e-15 0.000000e+00
             0.96          3     0 3.592054e-131 0.000000e+00
             0.48         73     0  2.215148e-07 0.000000e+00
             0.14         13     5  1.129620e-01 5.648101e-01
```

Bayesian inference by calculation

```
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(pars$n_visitors, size = n_ads_shown,
                          prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

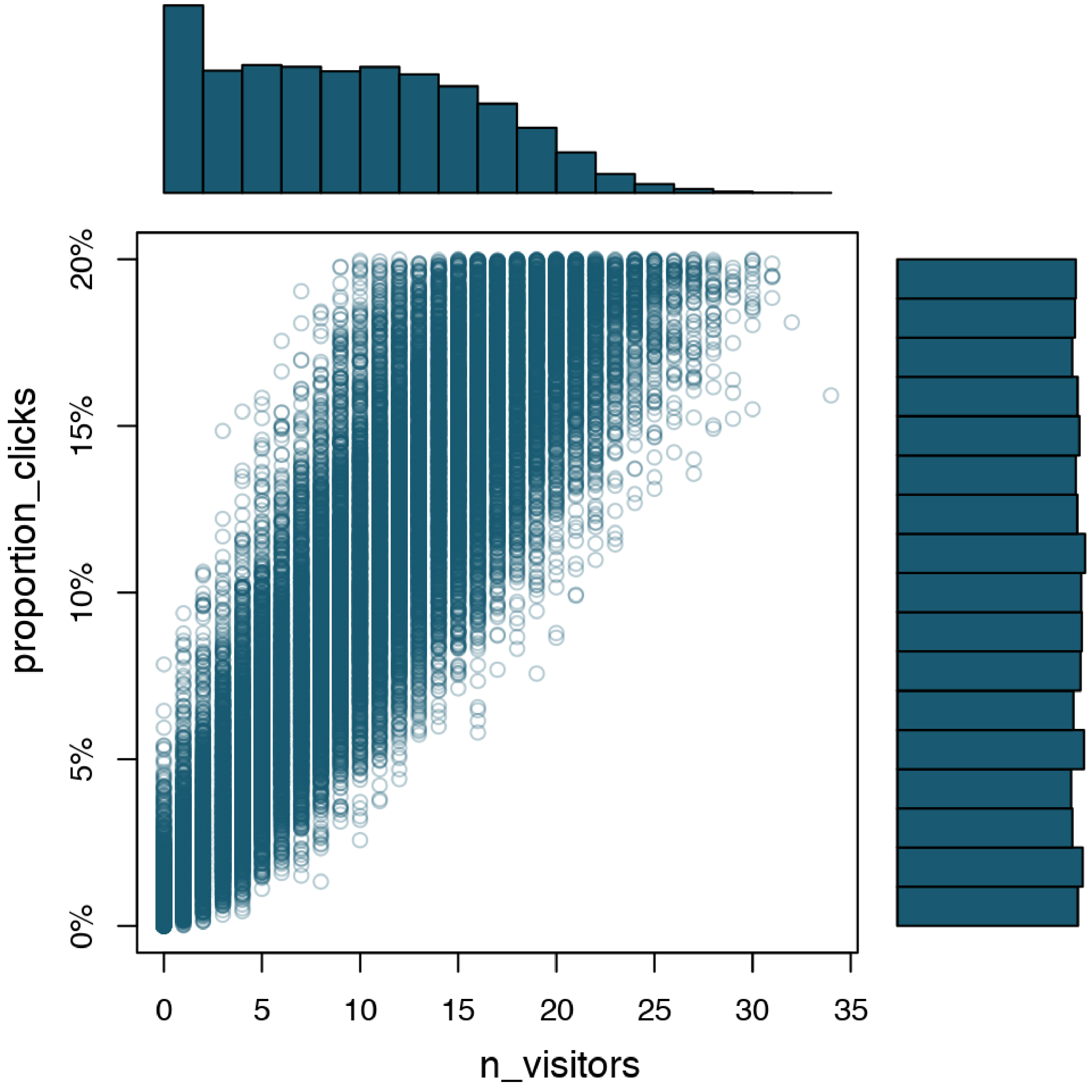
```
proportion_clicks n_visitors prior    likelihood  probability
             0.04         38     5  3.409439e-27 1.623542e-28
             0.11         93     5  5.006969e-80 2.384271e-81
             0.16        100     5  2.582250e-80 1.229643e-81
             0.67         98     0  4.863666e-15 0.000000e+00
             0.96          3     0 3.592054e-131 0.000000e+00
             0.48         73     0  2.215148e-07 0.000000e+00
             0.14         13     5  1.129620e-01 5.379144e-03
              ...        ...   ...           ...          ...
```

Bayesian inference by calculation

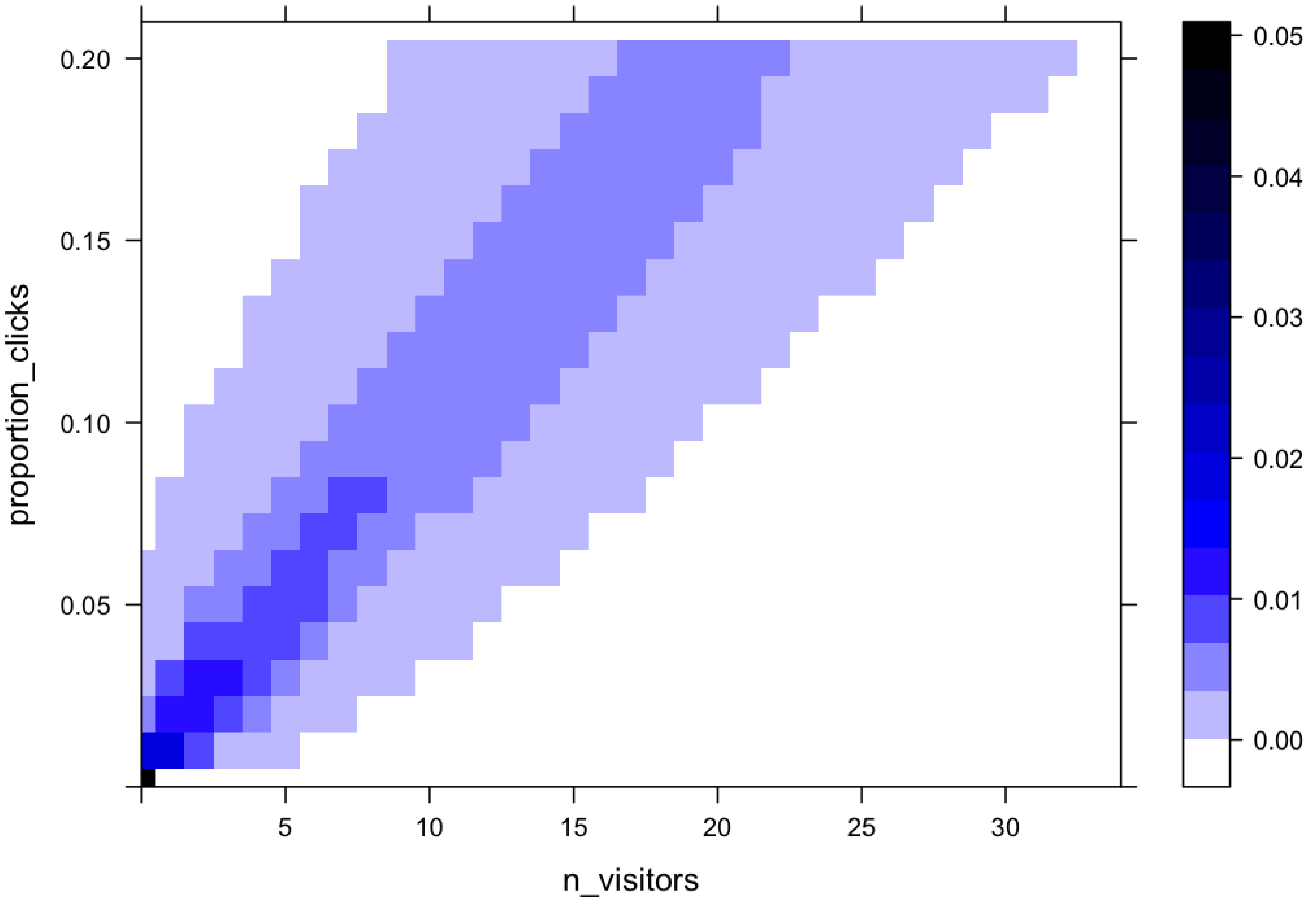
```
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(pars$n_visitors, size = n_ads_shown,
                          prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
sum(pars$probability)
```

```
1
```

```
proportion_clicks n_visitors prior    likelihood  probability
             0.04         38     5  3.409439e-27 1.623542e-28
             0.11         93     5  5.006969e-80 2.384271e-81
             0.16        100     5  2.582250e-80 1.229643e-81
             0.67         98     0  4.863666e-15 0.000000e+00
             0.96          3     0 3.592054e-131 0.000000e+00
             0.48         73     0  2.215148e-07 0.000000e+00
```



proportion_clicks
n_visitors
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35



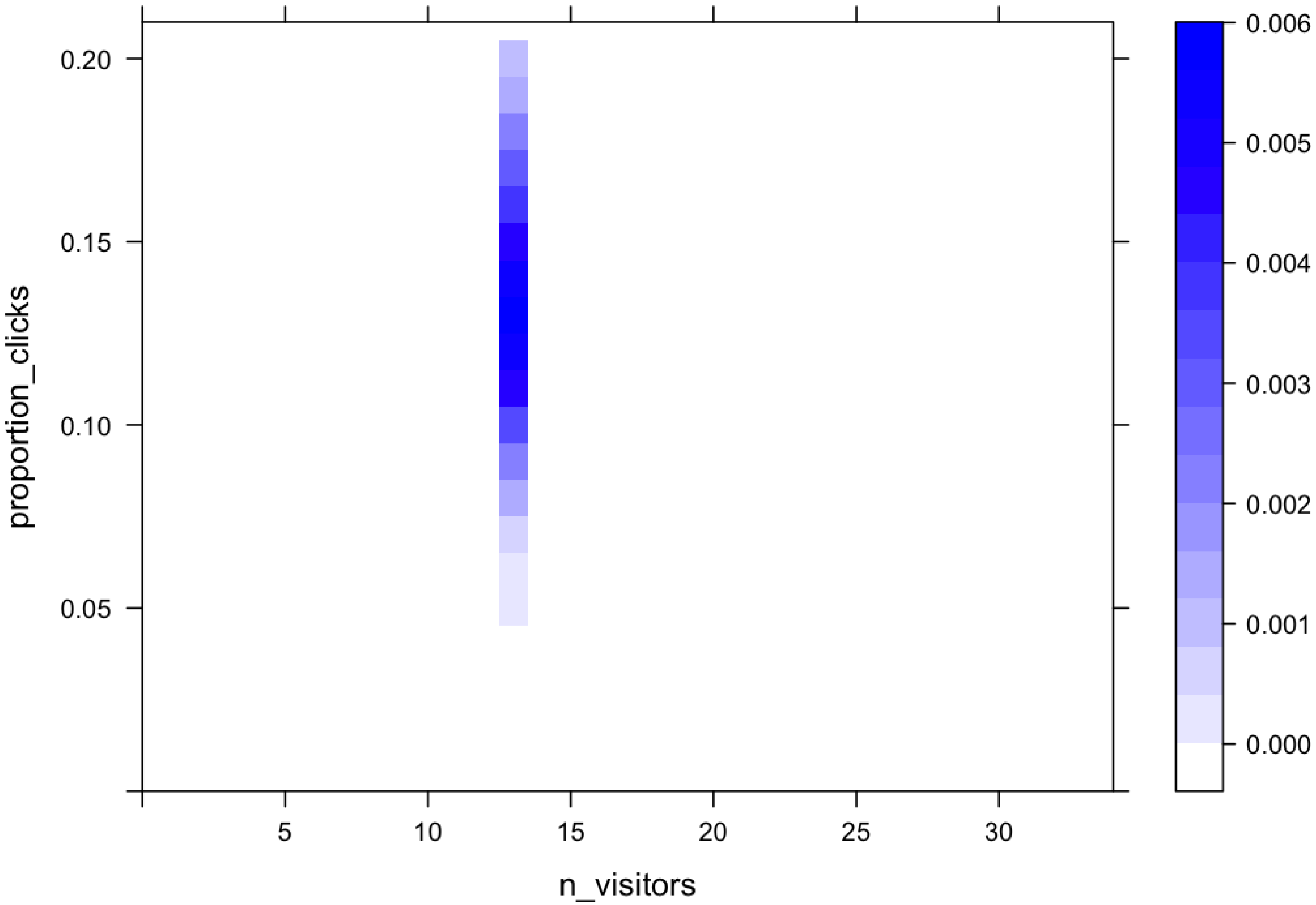
proportion_clicks
0.20
0.15
0.10
0.05
5
10
15
20
25
30
n_visitors
0.05
0.04
0.03
0.02
0.01
0.00

```r
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(pars$n_visitors, size = n_ads_shown,
                          prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

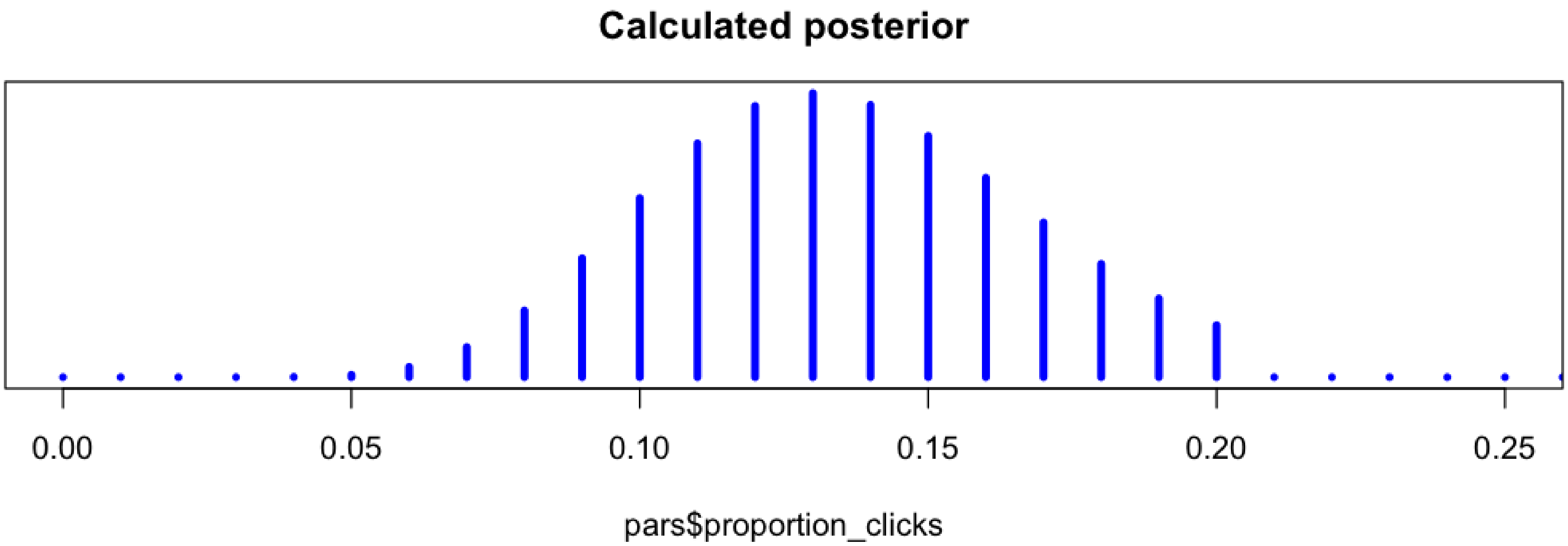
```
proportion_clicks n_visitors prior    likelihood  probability
             0.04         38     5  3.409439e-27 1.623542e-28
             0.11         93     5  5.006969e-80 2.384271e-81
             0.16        100     5  2.582250e-80 1.229643e-81
             0.67         98     0  4.863666e-15 0.000000e+00
             0.96          3     0 3.592054e-131 0.000000e+00
             0.48         73     0  2.215148e-07 0.000000e+00
             0.14         13     5  1.129620e-01 5.379144e-03
              ...        ...   ...           ...          ...
```

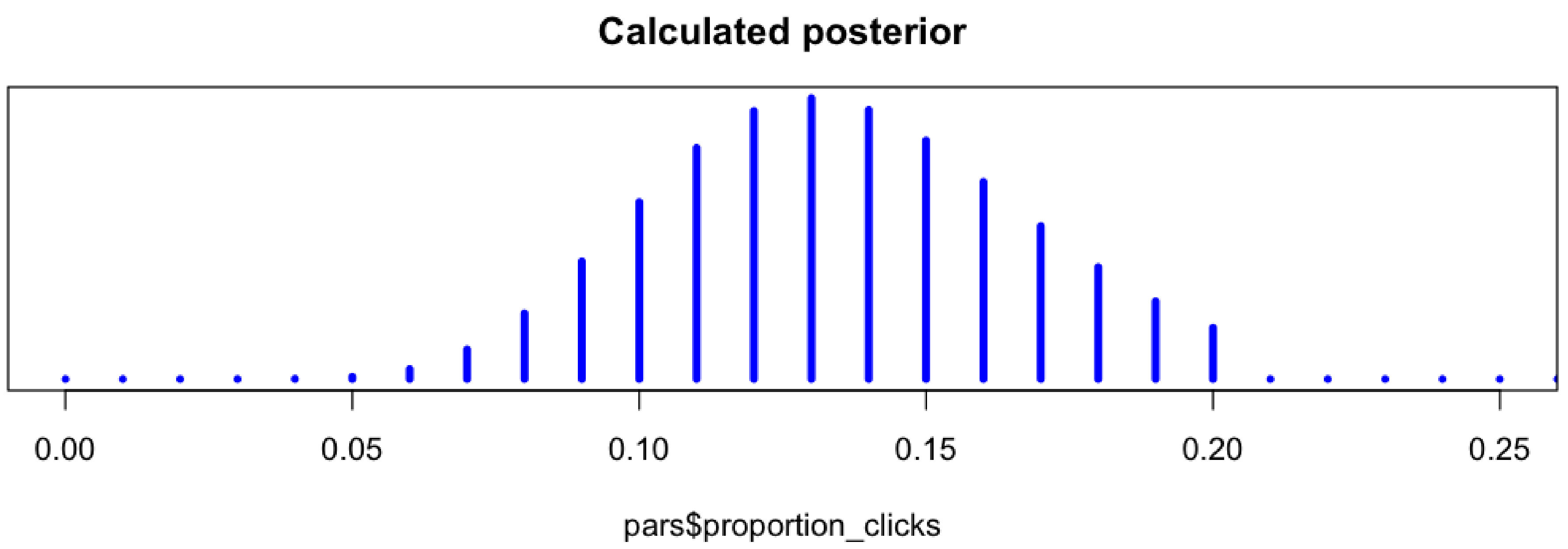
```r
n_ads_shown <- 100
n_visitors <- seq(0, 100, by = 1)
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks,
                    n_visitors = n_visitors)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(pars$n_visitors, size = n_ads_shown,
                          prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
pars <- pars[pars$n_visitors == 13, ]
pars$probability <- pars$probability / sum(pars$probability)
```

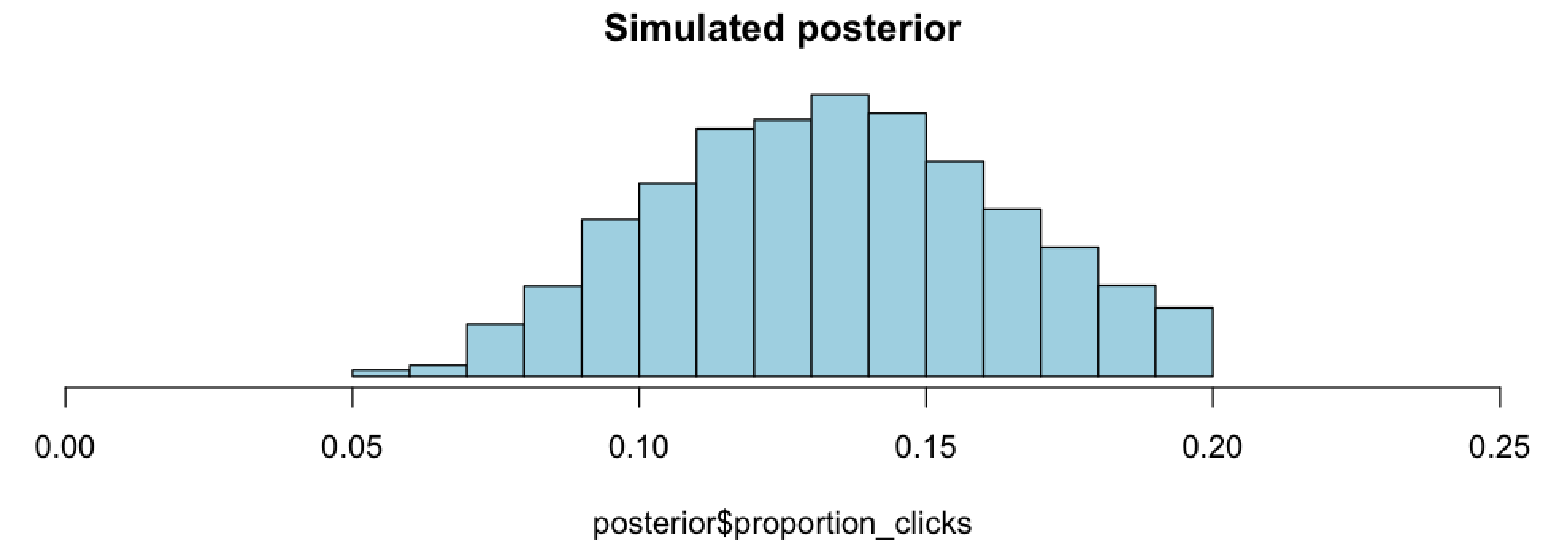
```
proportion_clicks n_visitors prior    likelihood  probability
             0.04         13     5  1.368611e-04 0.0001428716
             0.14         13     5  1.129620e-01 0.1179229621
             0.19         13     5  3.265098e-02 0.0340849069
             0.39         13     0  7.234996e-09 0.0000000000
             0.59         13     0  1.531703e-21 0.0000000000
             0.79         13     0  3.582066e-45 0.0000000000
             0.94         13     0  1.591196e-91 0.0000000000
              ...        ...   ...           ...          ...
```



0.20
0.15
0.10
0.05
proportion_clicks
5
10
15
20
25
30
n_visitors
0.006
0.005
0.004
0.003
0.002
0.001
0.000


Calculated posterior
0.00
0.05
0.10
0.15
0.20
0.25
pars$proportion_clicks


Calculated posterior
0.00
0.05
0.10
0.15
0.20
0.25
pars$proportion_clicks


Simulated posterior
0.00
0.05
0.10
0.15
0.20
0.25
posterior$proportion_clicks

Calculate for yourself!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Bayes' theorem

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Rasmus Bååth
Data Scientist

This is Bayes' theorem!

```
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

This is Bayes' theorem!

```
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

$$\mathrm{P}(\theta|D)$$

The probability of **different parameter values**
given some **data**

This is Bayes' theorem!

```
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

$$P(\theta|D) = P(D|\theta)$$

The probability of **different parameter values**
given some **data**
= equals =
The likelihood: The (relative) probability of the data
given different parameter values

This is Bayes' theorem!

```
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

$$\mathrm{P}(\theta|D) = \frac{\mathrm{P}(D|\theta) \times \mathrm{P}(\theta)}{}$$

The probability of **different parameter values**
given some **data**
= equals =
The likelihood: The (relative) probability of the data
given different parameter values
× times ×
The prior: The probability of different parameters
before seeing the data

This is Bayes' theorem!

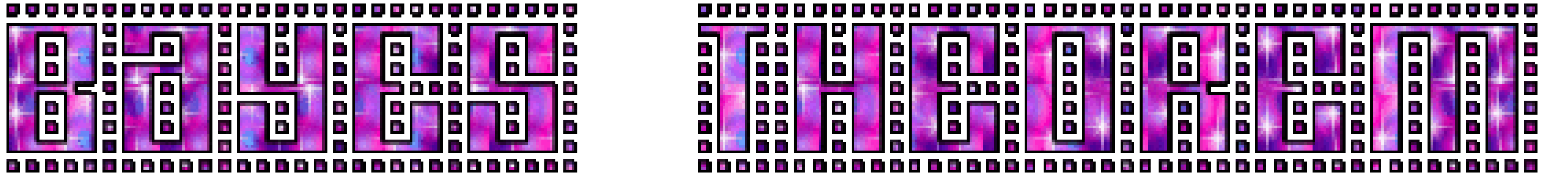
```
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{\sum P(D|\theta) \times P(\theta)}$$

The probability of **different parameter values**
given some **data**
= equals =
The likelihood: The (relative) probability of the data
given different parameter values
× times ×
The prior: The probability of different parameters
before seeing the data
/ divided by /
The total sum of the likelihood weighted by the prior.

This is Bayes' theorem!

$$\mathrm{P}(\theta|D) = \frac{\mathrm{P}(D|\theta) \times \mathrm{P}(\theta)}{\sum \mathrm{P}(D|\theta) \times \mathrm{P}(\theta)}$$



Grid approximation

- Define a grid over all the parameter combinations you need to evaluate.
- Approximate as it's often impossible try all parameter combinations.
- (There are many more algorithms to fit Bayesian models, some more efficient than others...)

A mathematical notation for models

$$n_{\text{ads}} = 100$$

$$p_{\text{clicks}} \sim \text{Uniform}(0.0, 0.2)$$

$$n_{\text{visitors}} \sim \text{Binomial}(n_{\text{ads}}, p_{\text{clicks}})$$

Up next: More parameters, more data!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R