Customer Analytics

Lecture 1

TILBURG UNIVERSITY

Test & Roll

E-Beer

- E-Beer sells beer over the Internet and currently has about 50,000 customers
- A customer selects the favorite brand, pays, and within 1 hour the ordered amount of beer is delivered at the specified address
- To boost sales, E-Beer developed a catalog to send to their customers.
- Each catalog contains a flyer to remind customers of the offered service and a key ring with the name and web address of the company

Campaign costs

- Each mailing costs € 1.50

- Sending it to all customers would mean total costs of

$$€\ 1.50 \times 50000 = €\ 75000$$

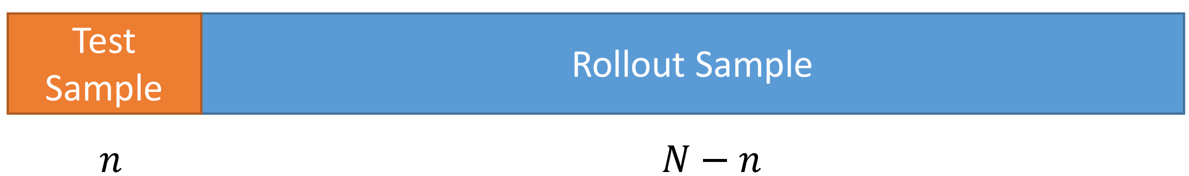
Is it worth it? Do benefits > costs?

The problem is that the **benefit** is uncertain!

Test & Roll experiments

Get some information on how effective your marketing is (the test) before you send it to everyone (the rollout)

1. Randomly select some subset of customers; call this test sample (size = n).
2. Send them mailing, collect & analyze responses
3. Use results to decide whether to send to the rest of the population (size $N - n$, rollout sample).



Results of test

- Assume the test sample size $n = 5000$. So, we randomly select 5000 customers and send them the mailing.

- Results of test mailing

  175 out of 5000 respond. So the estimated response rate: $\hat{p} = 175/5000$

  We assume the margin or profit per response is €50: $\mathrm{m} = 50$

So should we do the rollout? How much would we expect to make if we send to the rest (rollout sample)?

Expected rollout profits

Assuming they are like the test sample, (which they are if randomly sampled):

$$\mathrm{E}[\text{rollout profit}] = \underbrace{(N - n)}_{\textbf{\# rollout customers}} \underbrace{(m \cdot \hat{p} - c)}_{\textbf{profit per customer}}$$

$$= (45000)(50 \cdot 0.035 - 1.50)$$

$$= \mathbf{11250}$$

where

m is the margin (profit) per response (in euros)
$\hat{p}$ is the estimate of the response rate
c is the cost of marketing

Option value

- Therefore, because our expected rollout profit is positive, we roll it out to the rest of the sample.
- The test gives us the option – not the obligation – to rollout. We only roll out when:

$$E[\text{rollout profit}] > 0$$

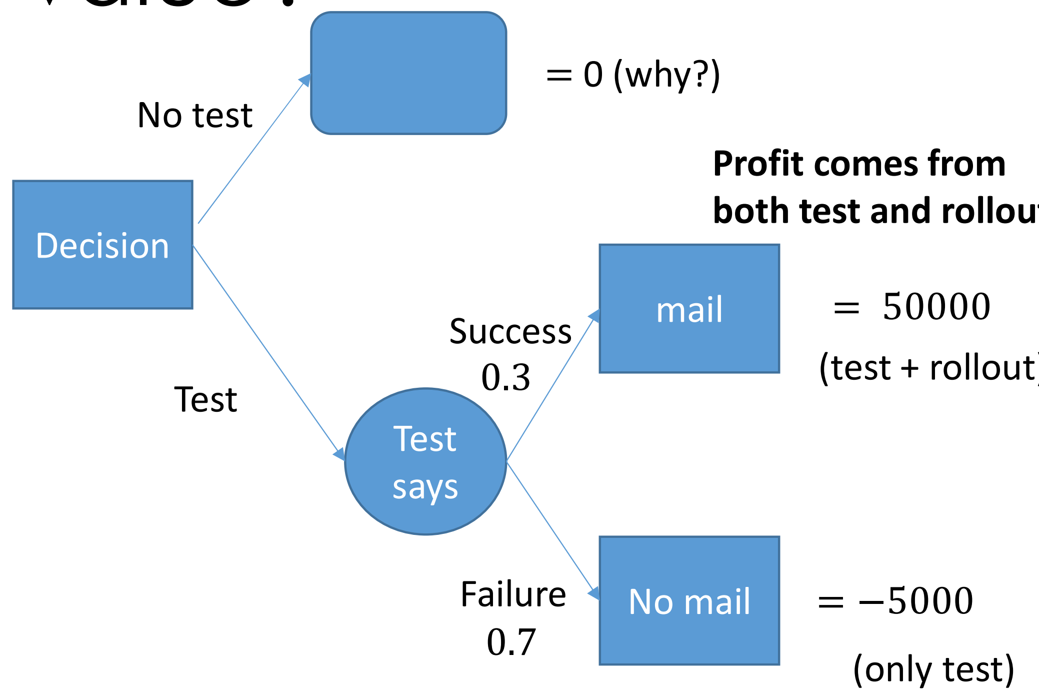
How big is the option value?

A1 Assume the test provides perfect information.

A2 Test predicts

1. Success ($p = .05$)
   $\Rightarrow (m \cdot p - c) = 1.00$
2. Failure ($p = .01$)
   $\Rightarrow (m \cdot p - c) = -1.00$

A3 Success occurs 30% of the time



$$\text{Value of test} = E[\text{profit}|\text{test}] - E[\text{profit}|\text{no test}]$$

$$= 0.3 \cdot (50000) + 0.7 \cdot (-5000) - 0$$
$$= \mathbf{11500}$$

BKN 9.2

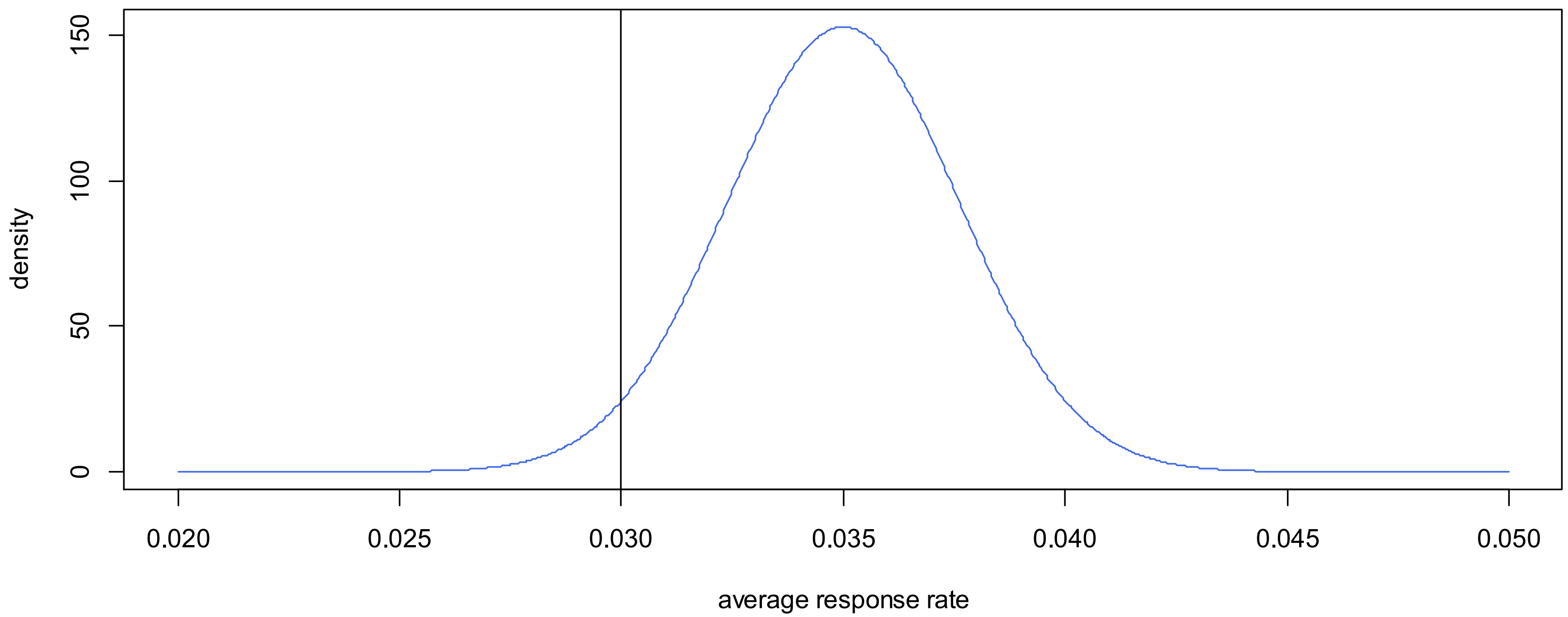
Uncertainty

The true unobserved population response rate is p

- What we observe: sample mean estimate, $\hat{p} = \frac{1}{n}\sum_i x_i$
- Its standard error, $se(p) = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{p(1-p)}{n}}$
- Central limit theorem. For large enough sample, distribution of sample mean is approximately normal

$$\hat{p} \sim N(\mathrm{p}, \mathrm{se}(\mathrm{p})^2)$$

What's the probability we make a mistake?



$$P(\hat{p} < 0.03) = .027$$

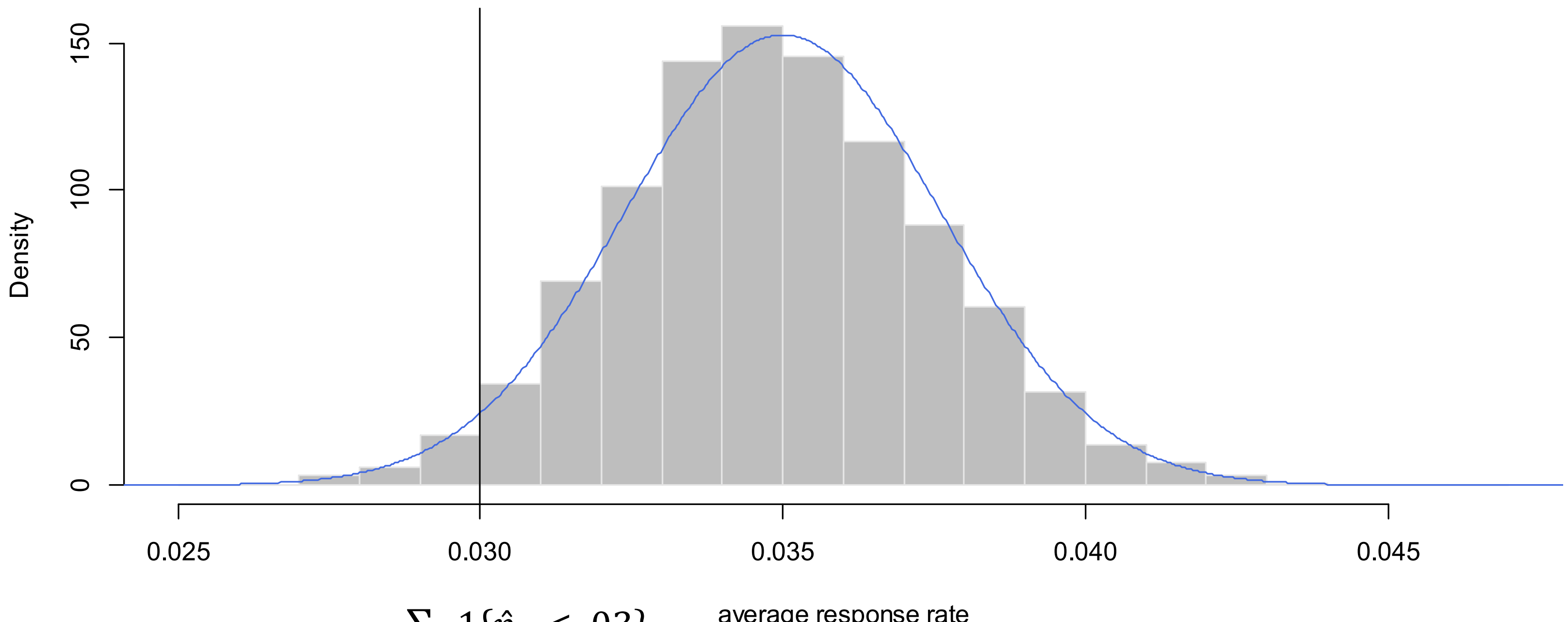
Bootstrap

Sample with replacement from the original sample, using the same sample size

For b = 1 … B bootstrap samples

1. Resample with replacement, $x_1^b, \ldots x_n^b$
2. Calculate estimate using this resample set, $\hat{p}_{\mathrm{b}} = \frac{1}{n}\sum_i x_i^b$

You now have a distribution $\hat{p}_1, \ldots \hat{p}_{\mathrm{B}}$.



$$\frac{\sum_b 1\{\hat{p}_b < .03\}}{B} = 0.0226$$

Bayesian approach

These are frequentist or classical approaches to quantifying uncertainty.

Before seeing the data, we may have some idea of what the response rate is:

- **Previous experience**: past campaigns' response rates range from about 0-10%, on average slightly below 3%
- **Or truly no idea:** every value is equally likely (**flat** or **diffuse** prior)

We summarize these ideas in a distribution called the **prior distribution**

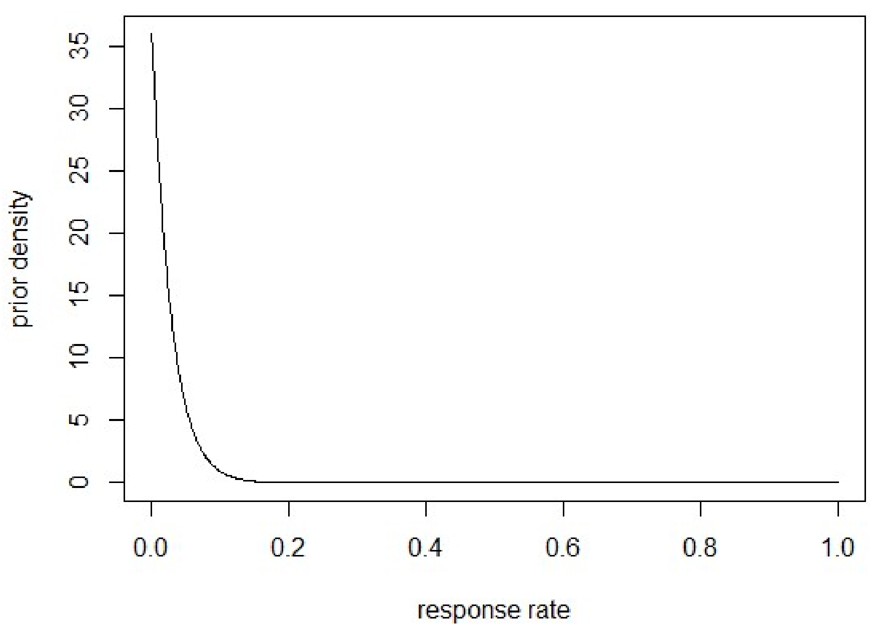
$$p \sim \mathrm{beta}(a, b)$$

$$\mathrm{f}(p) \propto p^{a-1}\,(1-p)^{b-1}$$

Prior density

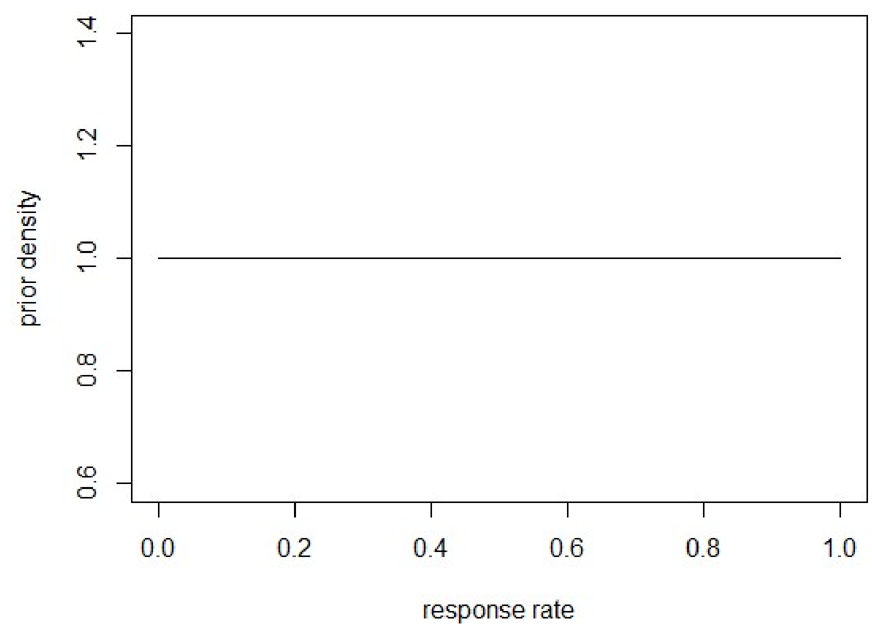
Previous experience

$$(a = 1, b = 36)$$



Flat or diffuse prior

$$(a = 1, b = 1)$$



Posterior distribution for the response rate

After we observe the response rate in the test, we can update our distribution. This is done via Bayes' rule

If we observe s responses in n observations

$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$

$$\text{f}(p) \propto p^{a+x-1}\,(1-p)^{b+n-s-1}$$

$$p \sim \text{beta}(a+s, b+n-s)$$

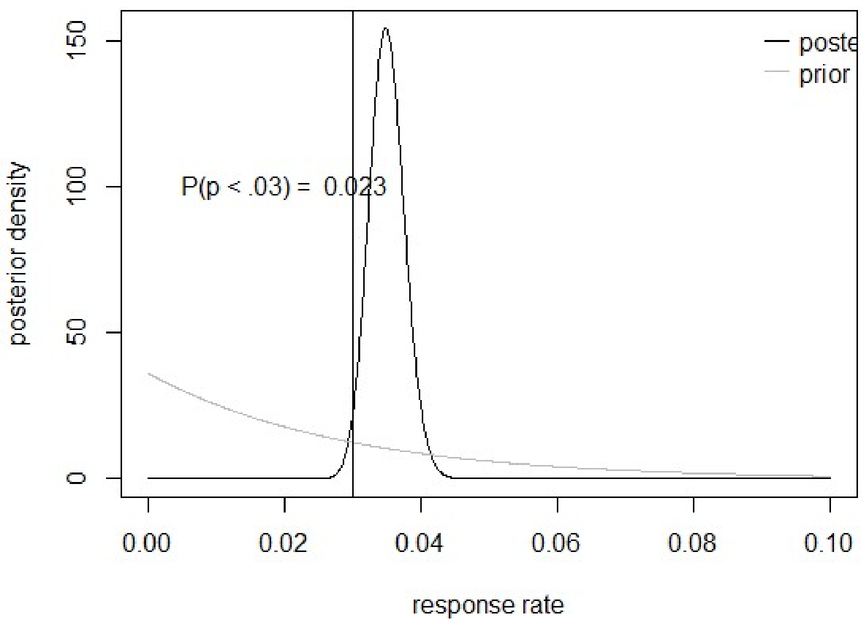
We can use these distributions to calculate e.g., the probability that we make a mistake, e.g. $P(p < .03)$

This is called the Beta-Binomial model.

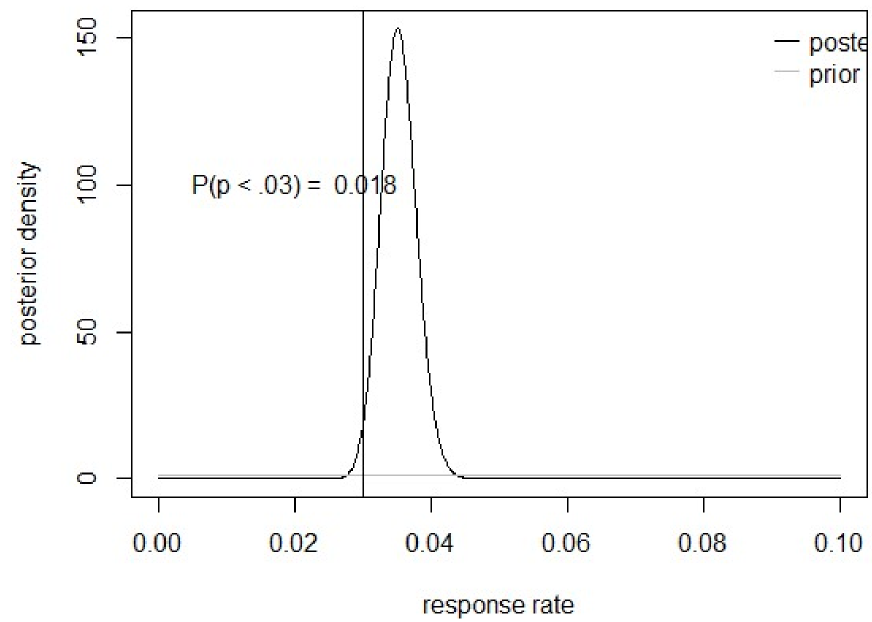
Posterior density

$$p \sim \mathrm{beta}(a + s, b + n - s)$$

$(a = 1 + 175, b = 36 + 5000 - 175)$



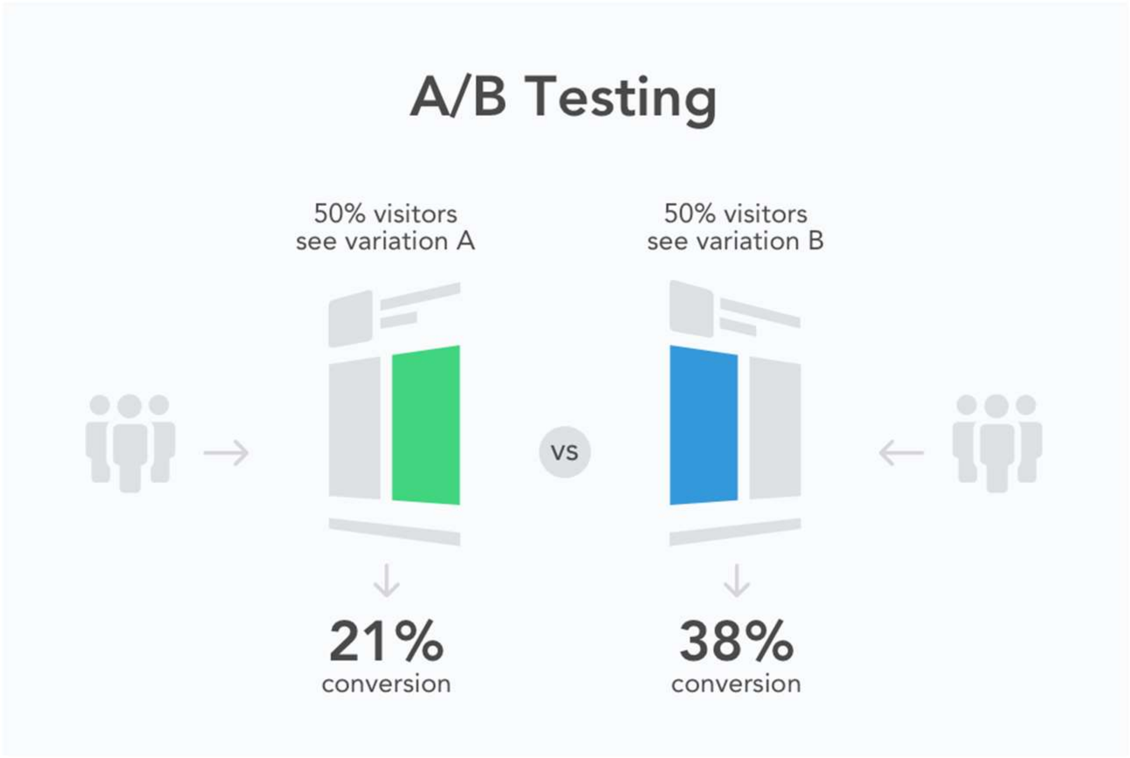
$(a = 1 + 175, b = 1 + 5000 - 175)$



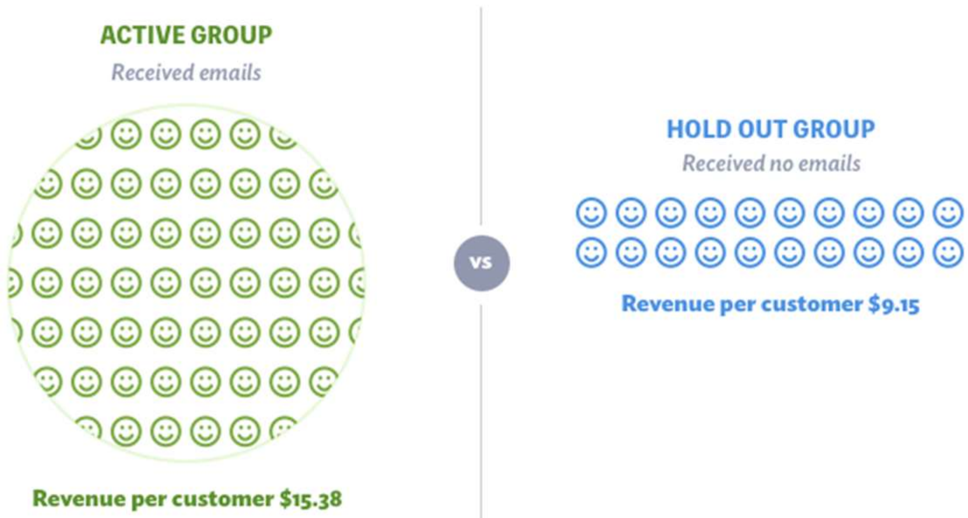
The more data you have, the less the prior matters

Tilburg University

You can use this to compare two versions



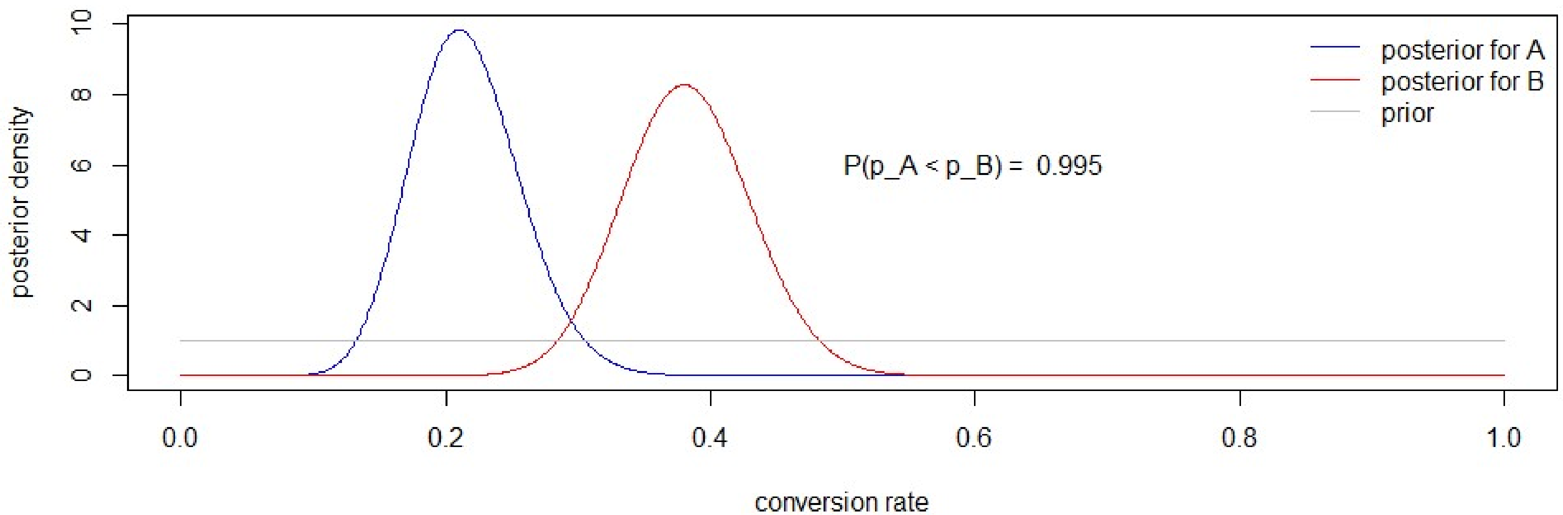
Holdout tests



$$p \sim \text{beta}(a + s, b + n - s)$$

$$p_A \sim \text{beta}(a = 1 + 26, b = 1 + 100 - 26)$$

$$p_B \sim \text{beta}(a = 1 + 38, b = 1 + 100 - 38)$$



Example

You collect data from an A/B test comparing the time users spend on your website for two versions of the homepage. A summary of the data looks like this:

Ver.	Viewers	Avg. Time	St. Dev.
A	500	5.40	1.97
B	500	5.44	2.11

It looks like B keeps users on the site longer, but how sure are we that B produces longer visits on average? We've only seen 500 visitors in each group.

Continuous data

Suppose the data are continuous, like minutes or profits, instead of binary {0,1} data like responses or clicks.

Likelihood (s is known)

$$y_i \sim \mathrm{N}(m, s^2)$$

Prior

$$m \sim \mathrm{N}(\mu_0, \sigma_0^2)$$

Posterior (*)

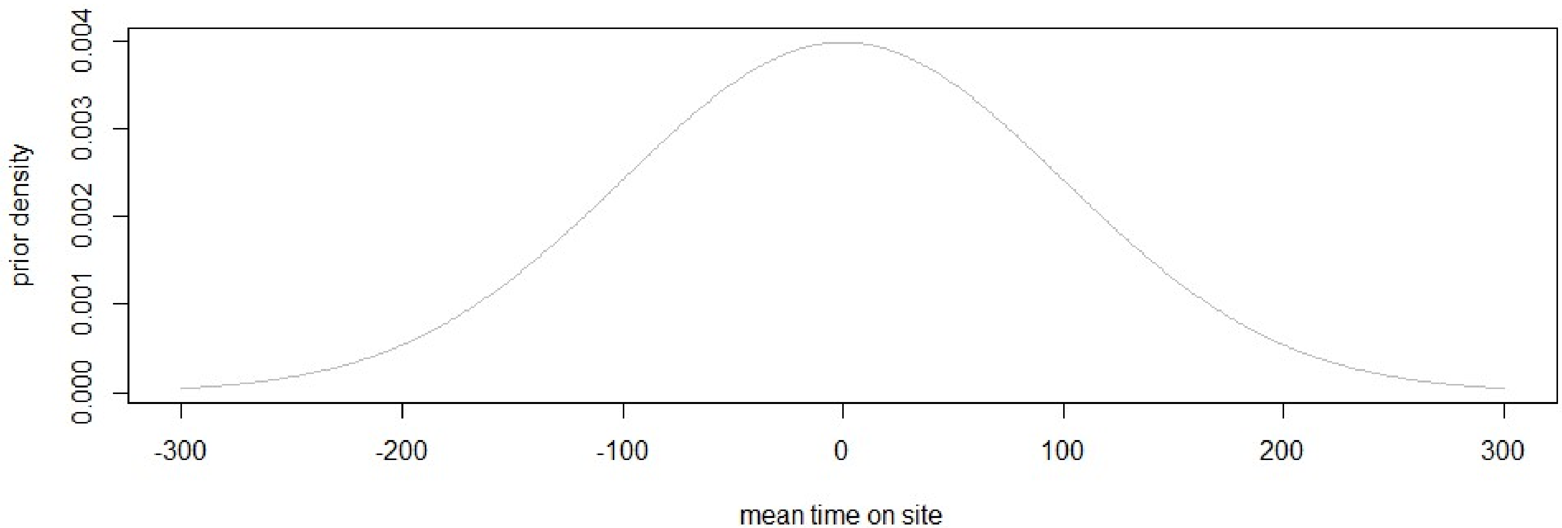
$$m \sim \mathrm{N}(\mu, \sigma^2)$$

$$\sigma = \sqrt{\left(\frac{1}{\sigma_0^2} + \frac{n}{s^2}\right)^{-1}} \qquad \mu = \sigma^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\,\bar{y}}{s^2}\right)$$

Prior

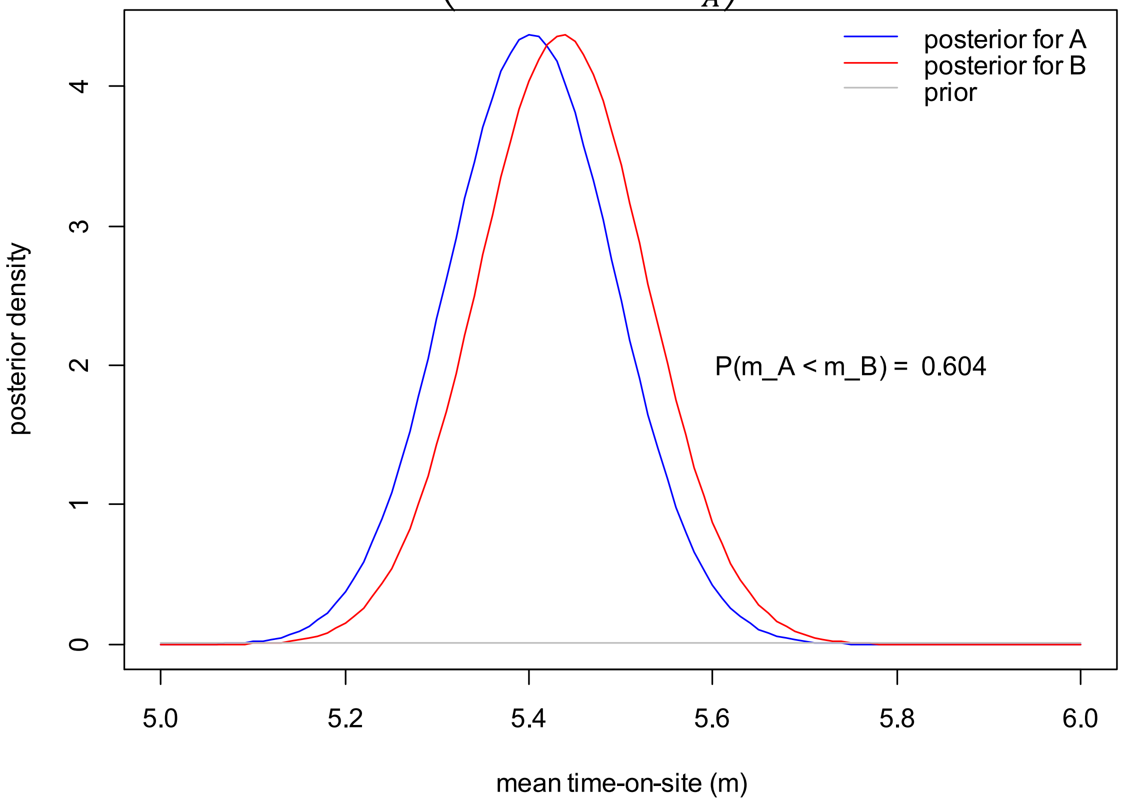
Both groups have the same diffuse prior, mean time on site (m)

$$m_A, m_B \sim \mathrm{N}(\mu_0 = 0, \sigma_0^2 = 100^2)$$



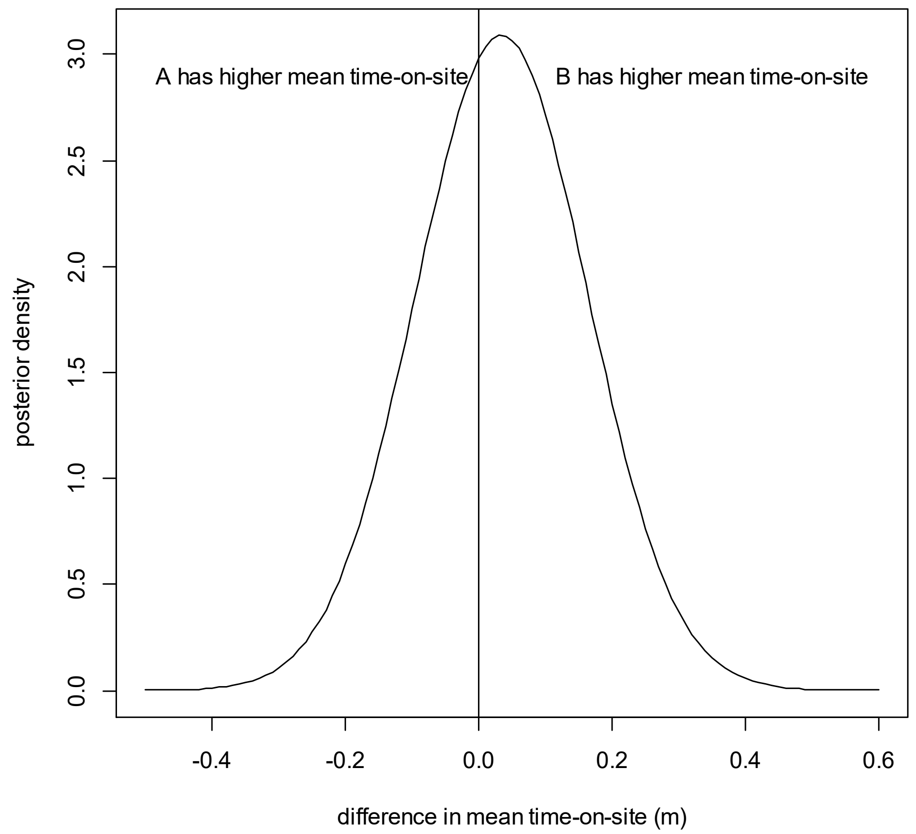
Posteriors

$$m_A \sim \mathrm{N}\left(\mu \approx \bar{y}_A, \sigma^2 \approx \frac{s^2}{n_A}\right)$$



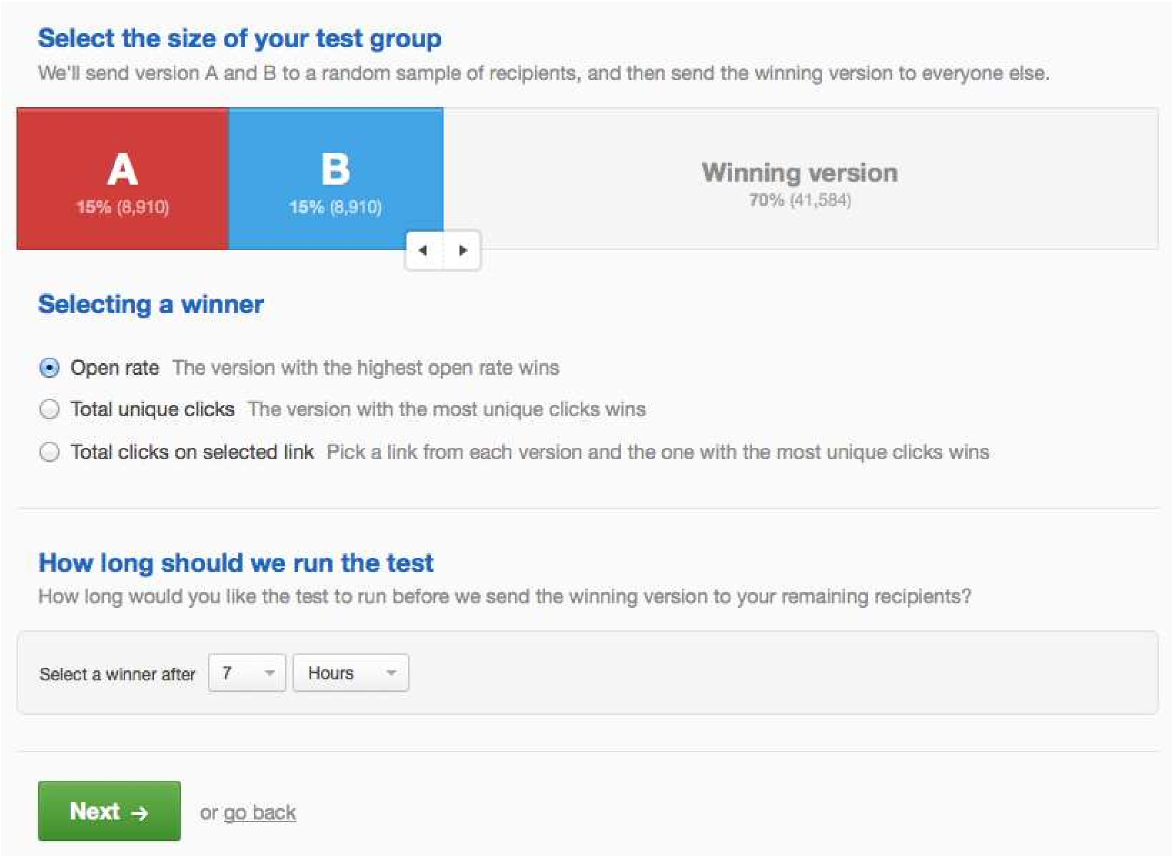
The posterior is the likelihood because the prior is so diffuse.

Posterior



A has higher mean time-on-site
B has higher mean time-on-site
posterior density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-0.4
-0.2
0.0
0.2
0.4
0.6
difference in mean time-on-site (m)

Typical A/B email test setup screen



How big should the test be?

The goal is to maximize expected profit from **both test and rollout stages.** (like our option value calculation earlier)

$$\max_{n_A, n_B} E\left[\text{Profit}_{\text{test}} + \text{Profit}_{\text{rollout}}\right]$$

n_A $\quad$ n_B $\quad$ $N - n_a - n_b$

Large test means you have a low rollout error (low risk)

Large test means people will get worse treatment (opportunity cost)

Feit and Berman (2019)

Decision rule

The profit maximizing decision rule is to choose the treatment with the higher posterior mean after observing the test

Calculate posterior mean

Test A	Test B

n_A n_B → m_A, m_B

choose A if $m_A > m_B$, otherwise B

Optimal sample size

Assumption: Normal-normal model for A & B, and they have identical priors

$$y_A \sim N(m_A, s^2),\ y_B \sim N(m_B, s^2) \text{ and } m_A, m_B \sim N(\mu, \sigma^2)$$

$$n_A^* = n_B^* = \sqrt{\frac{N}{4}\left(\frac{s}{\sigma}\right)^2 + \left(\frac{3}{4}\left(\frac{s}{\sigma}\right)^2\right)^2} - \frac{3}{4}\left(\frac{s}{\sigma}\right)^2$$

where

N is the total size of the customer base (+)
s is the standard deviation of the response (+)
σ is the standard deviation on the prior on the mean response for both groups in the normal-normal model: a prior range of effect sizes (-)

Feit and Berman (2019)

Example

Imagine your customer base is $N = 100000.$ In previous website tests, the mean conversion rate was distributed normally with mean $\mu = 0.68$ and standard deviation $\sigma = 0.03.$

We can approximate the response standard deviation $s = \sqrt{\mu\,(1-\mu)}.$

$$n_A^* = n_B^* = 2284$$

Your test size is $n_A^* + n_B^* = 4568.$ At the end of your test you roll out whichever version has a greater posterior mean conversion rate. If both have the same prior, then it's just the estimated conversion rate.

Extensions

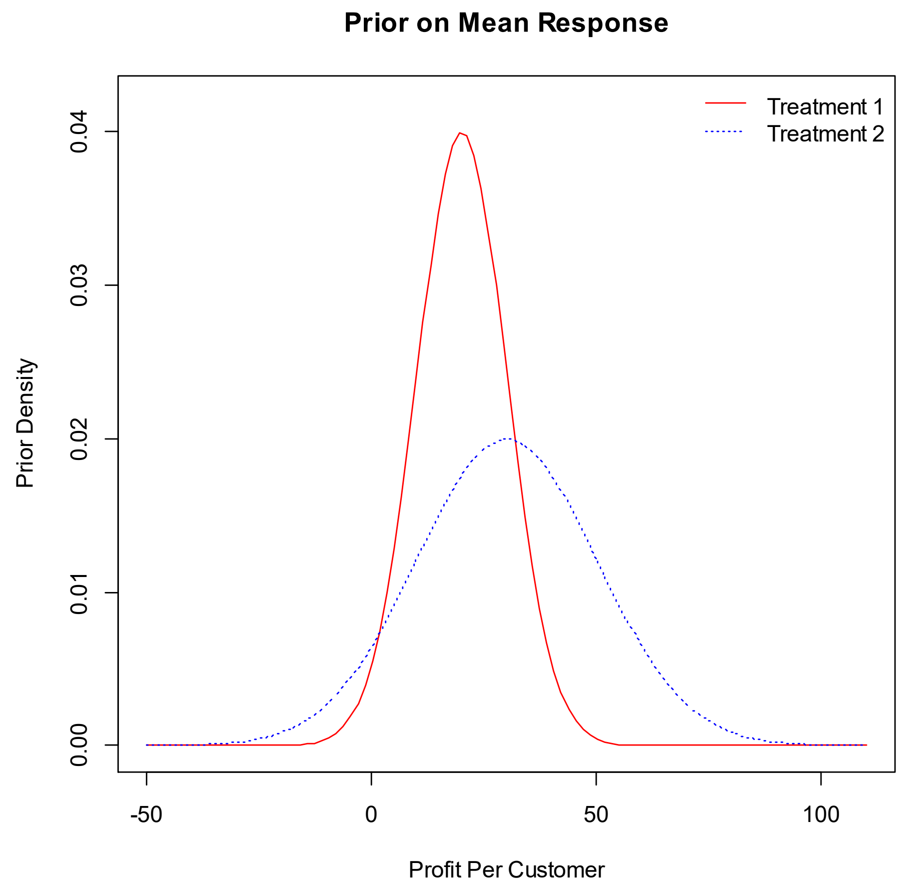
The two treatments come from different priors (e.g., holdout test)

$y_1 \sim N(m_1, 100^2)$, $y_2 \sim N(m_2, 200^2)$,

$m_1 \sim N(9, 10^2)$, $m_2 \sim N(30, 20^2)$

No closed-form solution, but can numerically solve

$$n_1^* = 642$$
$$n_2^* = 2284$$



We expect Treatment 2 to be better, so we allocate more tests.