Making decisions under uncertainty using Bayesian inference and Stan

Data Science In Insurance Conference, London

16 July 2018

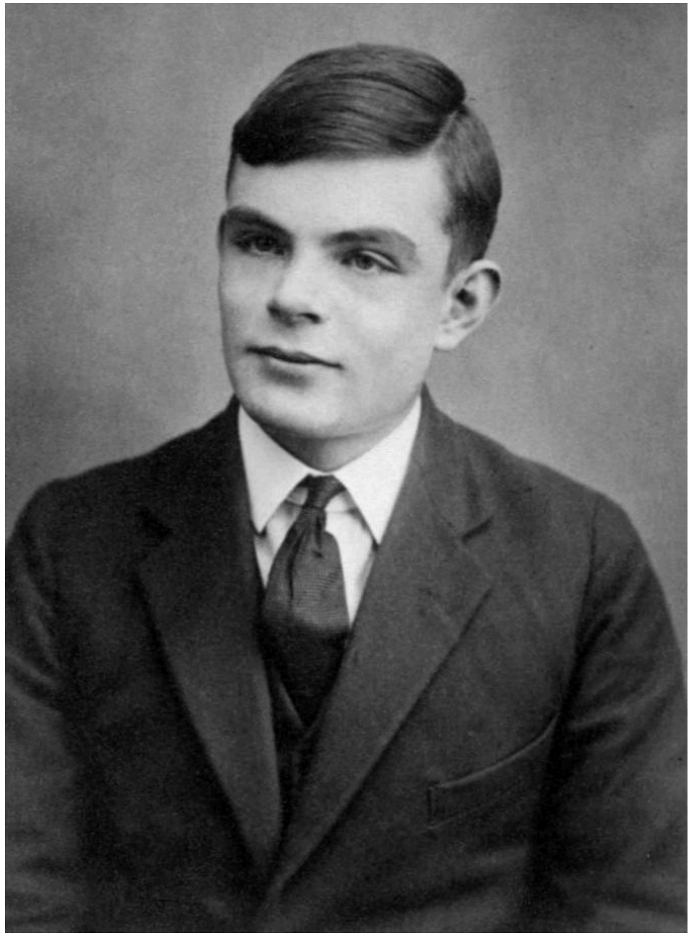
Eric Novik: eric@generable.com

Outline

- Intro to Decisions
  - Inference vs decision making
  - Bayesian loop
  - Bayesian Expected Loss
  - Decision Example
- Example: Book Pricing
  - Hierarchical model for pricing
  - Communicating risk and reward tradeoffs to business people

Some benefits of Bayesian approach

- Make rational decisions under uncertainty
- Express your beliefs about parameters **and** the data generating process
- Properly account for uncertainty at the individual and group level
- Do not collapse grouping variables
- Small data is fine if you have a strong model



Learning. To act.

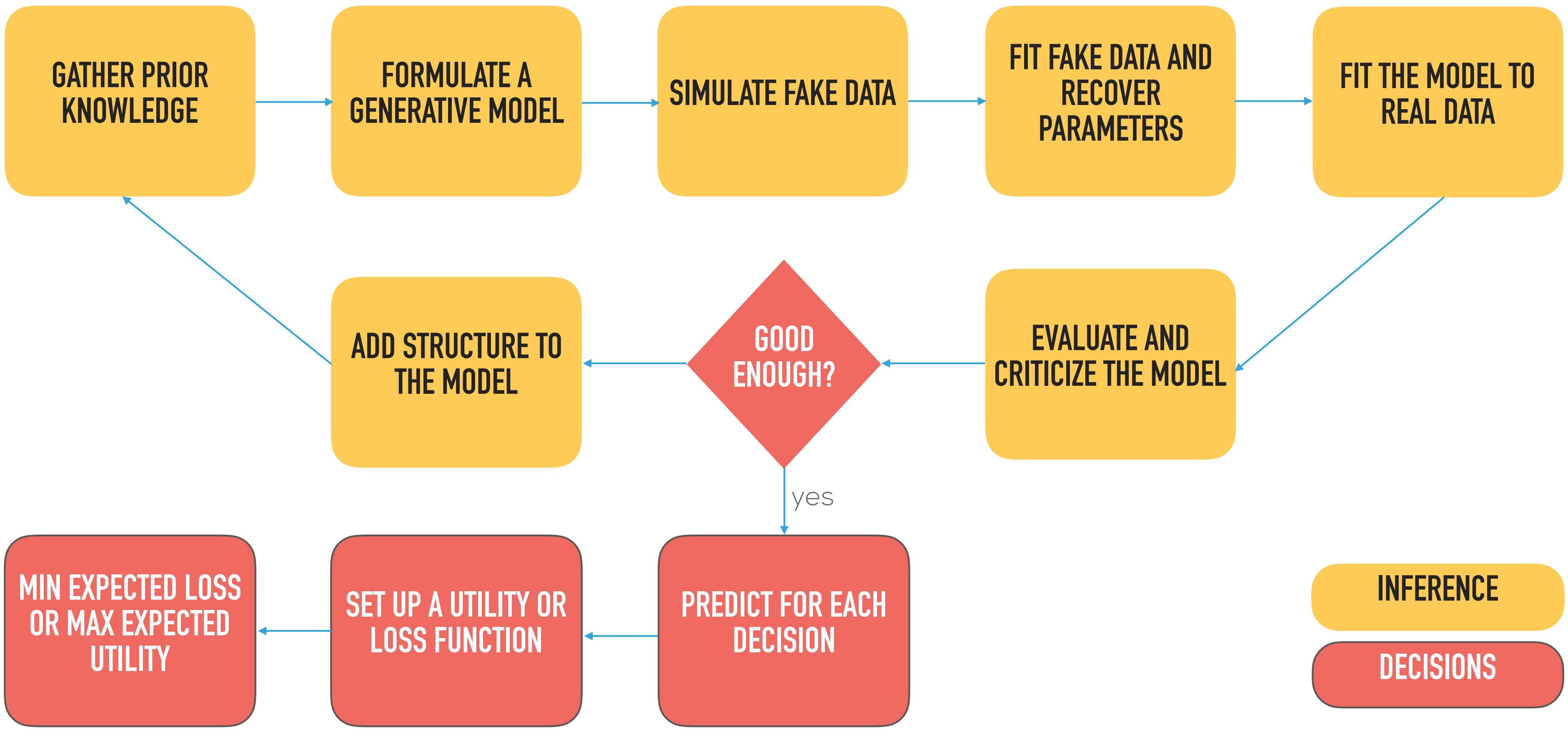
Communicating with stakeholders

- Before model model building

- During model building

- After the model had been fit

Enter the Bayesian loop




To Bet or Not to Bet
That is the question
Decisions Under Uncertainty

Motivation for using prior information

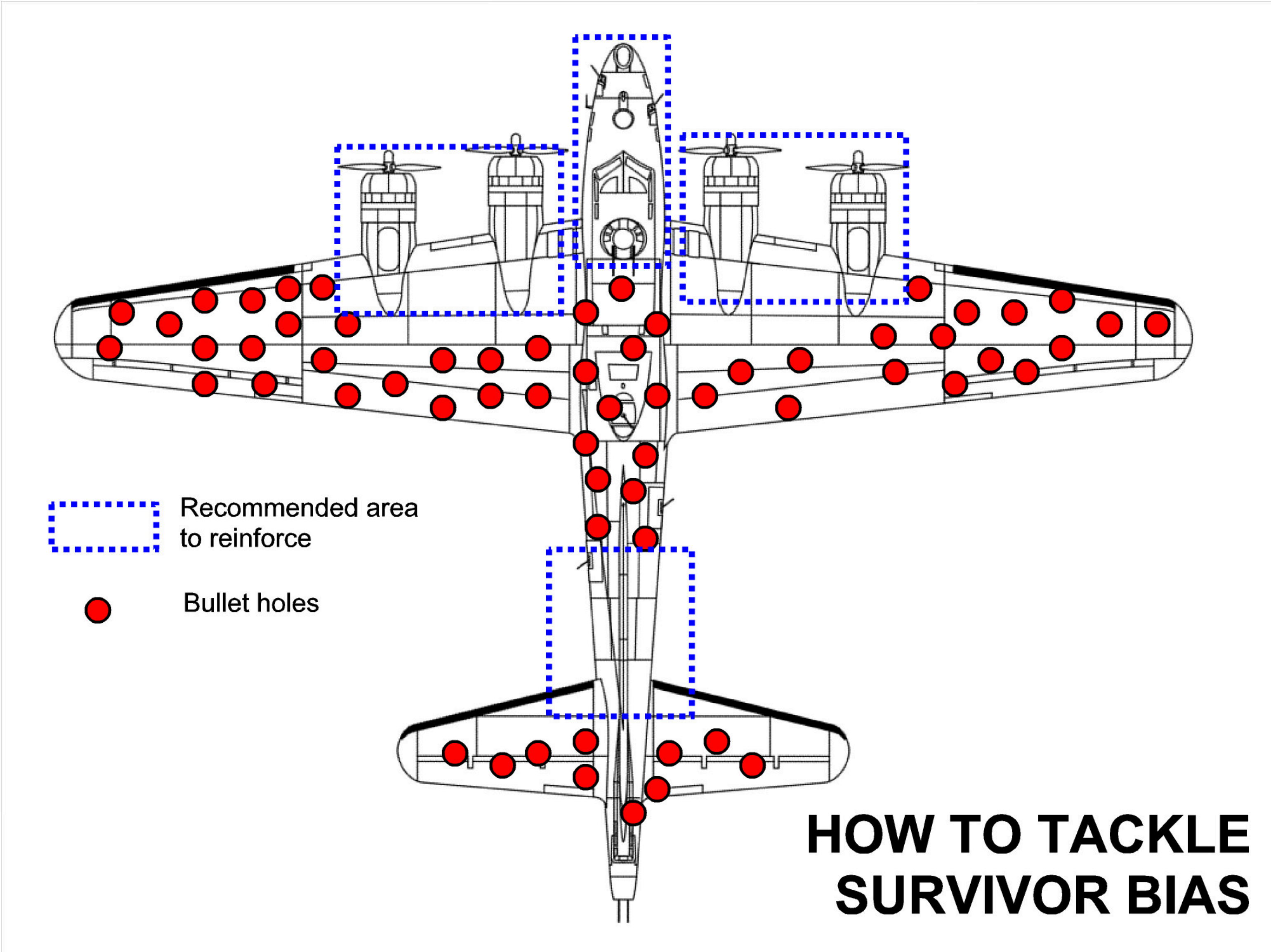
- From LJ Savage (1961)

1. A lady, who adds milk to her tea, claims to be able to tell whether the tea or the milk was poured into the cup first. In all of ten trials conducted to test this, she correctly determines which was poured first.
2. A music expert claims to be able to distinguish a page of Haydn score from a page of Mozart score. In ten trials conducted to test this, he makes a correct determination each time.
3. A drunken friend says he can predict the outcome of a flip of a fair coin. In ten trials conducted to test this, he is correct each time.

Bayesian expected loss

$$\rho(\pi, a) = E_{\theta|y}(L(\theta, a)) = \int_{\Theta} L(\theta, a)\pi(\theta|y)d\theta$$

- θ — Unknown parameter vector, state of the world
- y — Observed data, say outcome of an experiment
- a — Action to be taken (decision d(y), s.t. if Y = y, do a)
- $L(\theta, a)$ — Is the loss function

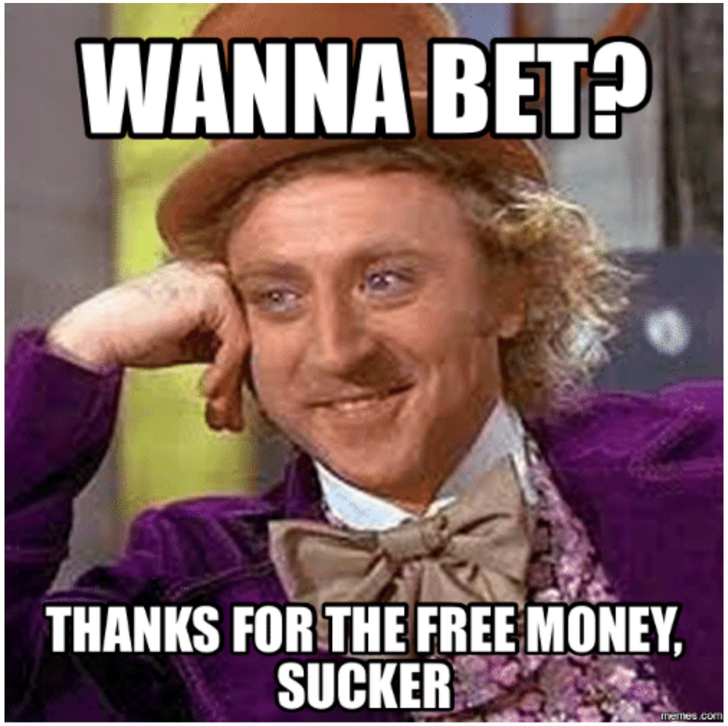

Recommended area
to reinforce
Bullet holes
HOW TO TACKLE
SURVIVOR BIAS

Simple decision problem

- You observe the following sequence of of draws from a Bernoulli process:

T, H, H, T, H

- If the process is biased towards Heads, you get $100
- If it is biased towards Tails, you loose $100
- It costs $30 to play the game
- Are you in?



Bayesian machinery

- The joint probability of data **y** and unknown parameter **theta**:

$$p(y, \theta) = p(y|\theta) * p(\theta)$$

$$p(y, \theta) = p(\theta|y) * p(y)$$

- The conditional probability of **theta** given **y**:

Likelihood Prior

$$p(\theta|y) = \frac{p(y|\theta) * p(\theta)}{p(y)} = \frac{p(y|\theta) * p(\theta)}{\int p(y, \theta) d\theta} = \frac{p(y|\theta) * p(\theta)}{\int p(y|\theta) * p(\theta) d\theta}$$

Marginal Likelihood

$$\propto p(y|\theta) * p(\theta)$$

Bernoulli Model

- If we model each occurrence as independent, the joint model can be written as:

Bernoulli Likelihood $p(y|\theta)$

$$p(y,\theta) = \prod_{n=1}^{N} \theta^{y_n} * (1-\theta)^{1-y_n} = \theta^{\sum_{n=1}^{N} y_n} * (1-\theta)^{\sum_{n=1}^{N}(1-y_n)}$$

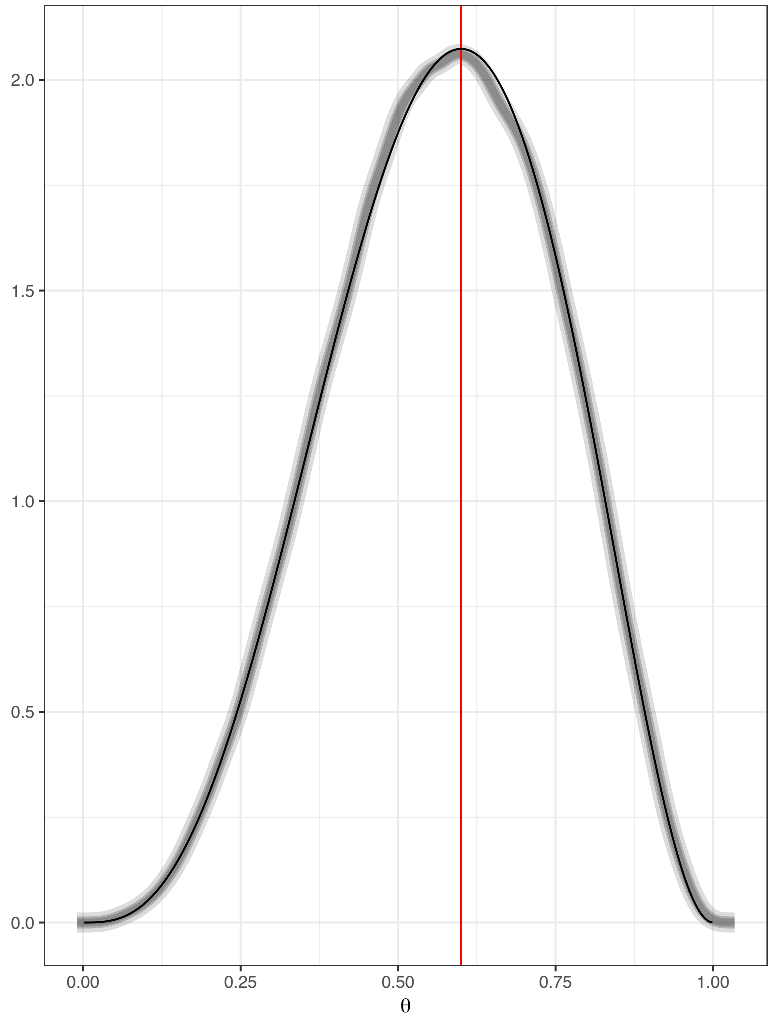
- What happened to the prior, $p(\theta)$
- On the log scale:

$$\log(p(y,\theta)) = \sum_{n=1}^{N} y_n * \log(\theta) + \sum_{n=1}^{N}(1-y_n) * \log(1-\theta)$$

```
data <- list(N = 5,
             y = c(0, 1, 1, 0, 1))

# log probability function
lp <- function(theta, data) {
  lp <- 0
  for (i in 1:data$N) {
    lp <- lp + log(theta) * data$y[i] +
      log(1 - theta) * (1 - data$y[i])
  }
  return(lp)
}
```

Bernoulli Model

```
# generate the parameter grid
theta <- seq(0.001, 0.999,
             length.out = 250)

# log p(theta | y)
log_likelihood <- lp(theta = theta, data)
log_prior <- log(dbeta(theta, 1, 1))
log_posterior <- log_likelihood + log_prior
posterior <- exp(log_posterior)

# normalize
posterior <- posterior / sum(posterior)

# sample from the posterior
post_draws <- sample(theta, size = 1e5,
                     replace = TRUE,
                     prob = posterior)
post_dens <- density(post_draws)
mle <- sum(data$y) / data$N
> mle
[1] 0.6

# conjugate prior / posterior
theta_conj <- dbeta(theta, 1 + 3, 1 + 5 - 3)
```



Same Model in Stan

```
data {
  int<lower=1> N;
  int<lower=0, upper=1> y[N];
}
parameters {
  real theta;
}
model {
  for (n in 1:N)
    target += y[n] * log(theta) +
                (1 - y[n]) * log(1 - theta);
}
```

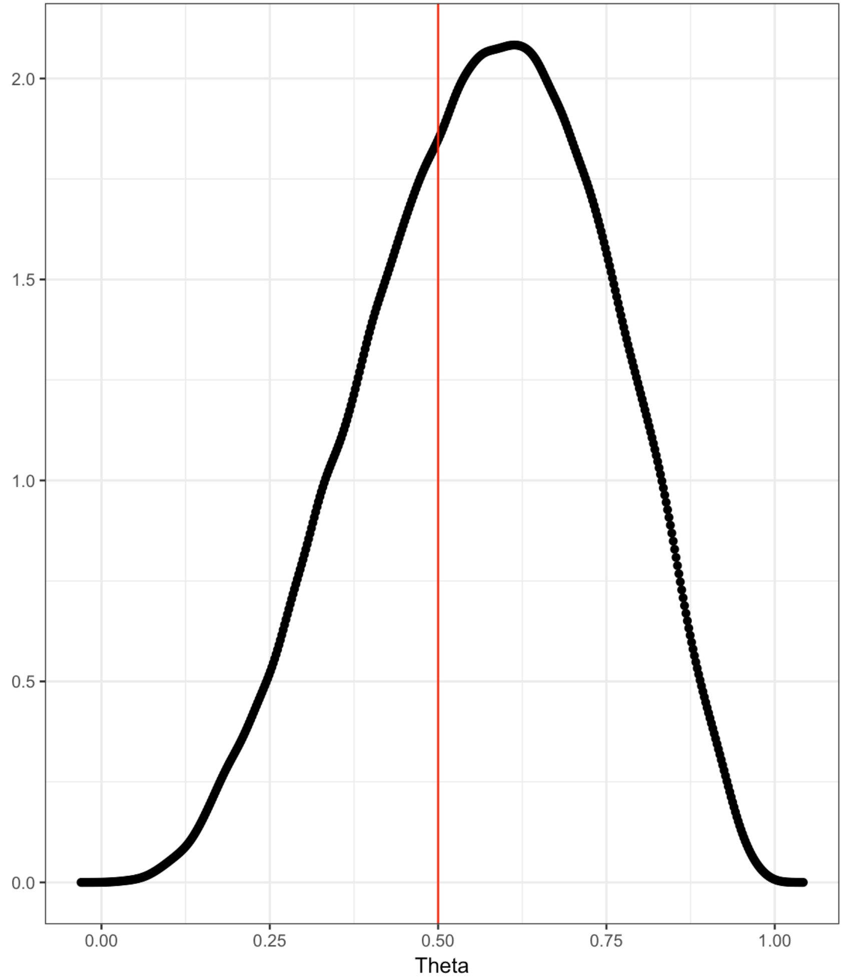
```
data {
  int<lower=1> N;
  int<lower=0, upper=1> y[N];
}
parameters {
  real<lower=0, upper=1> theta;
}
model {
  y ~ bernoulli(theta);
}
```

$$\log(p(y, \theta)) = \sum_{n=1}^{N} y_n * \log(\theta) + \sum_{n=1}^{N} (1 - y_n) * \log(1 - \theta)$$

Decision problem answer?

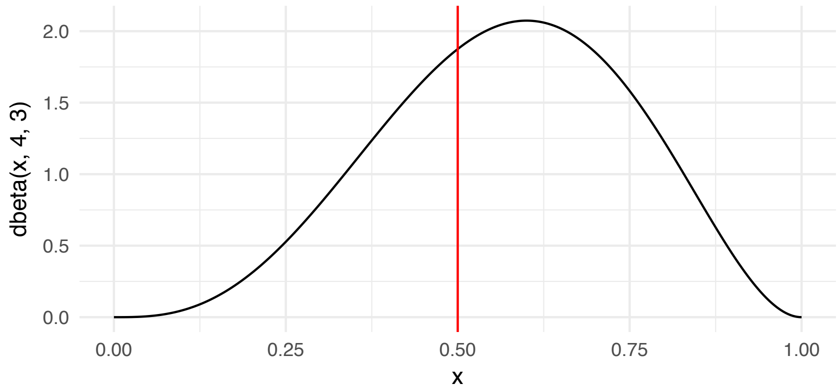
- If the coin is biased towards heads you get $100
- If not, you loose $100
- It costs $30 to play the game

```
> (p_heads_bias <- mean(post_draws > 0.5))
[1] 0.656346
> integrate(dbeta,
+           lower = 0.5, upper = 1,
+           shape1 = 4, shape2 = 3)
0.65625 with absolute error < 7.3e-15
> (p_tails_bias <- 1 - p_heads_bias)
[1] 0.343654
> p_heads_bias * 100 +
+   p_tails_bias * (-100) - 30
[1] 1.27
```

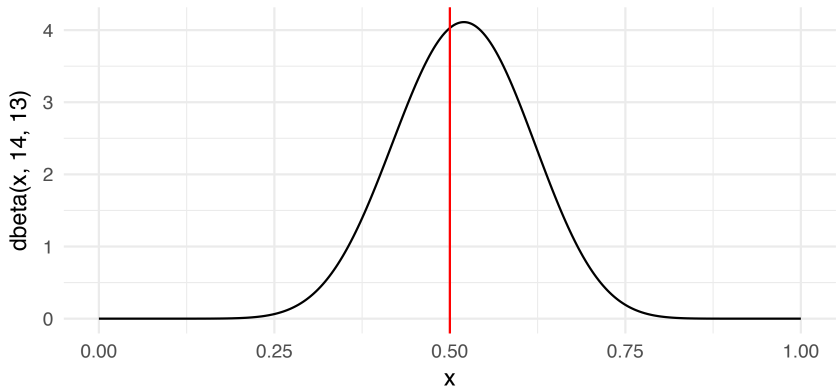


Changing the prior?

- You were told that the same machine was observed producing a sequence with 10 heads and 10 tails
- Would you still like to play the game?




```
> p_heads_bias <- integrate(dbeta,
+            lower = 0.5, upper = 1,
+            shape1 = 14, shape2 = 13)$value
> (p_tails_bias <- 1 - p_heads_bias)
[1] 0.42
> p_heads_bias * 100 +
+   p_tails_bias * (-100) - 30
[1] -14.50
```



Expected payoffs are not utilities. Paradox?

- A wealthy individual offers you the following bet:
- You flip a fair coin until you see Heads. For each round you will get 2^i dollars. How much are you willing to pay to play this game? If X is the total winnings:

$$E(X) = \frac{1}{2} * 2^1 + \frac{1}{4} * 2^2 + \frac{1}{8} * 2^3 + ... = 1 + 1 + 1 + ... = \infty$$

- Number of rounds N is the First Success distribution

$$N \sim FS(1/2)$$

$$E(N) = \frac{1}{p} = 2$$


Product Pricing
Decisions Under Uncertainty

Decision problem

- A publisher has thousands of books in the catalog
- Every year, hundreds of new books (products) are published
- There are also new authors, repeat authors, genres, seasonality, and so on
- **Publisher wants to maximize revenue, but not if it results in more than 10% loss in quantity sold**

Basic model for quantity sold

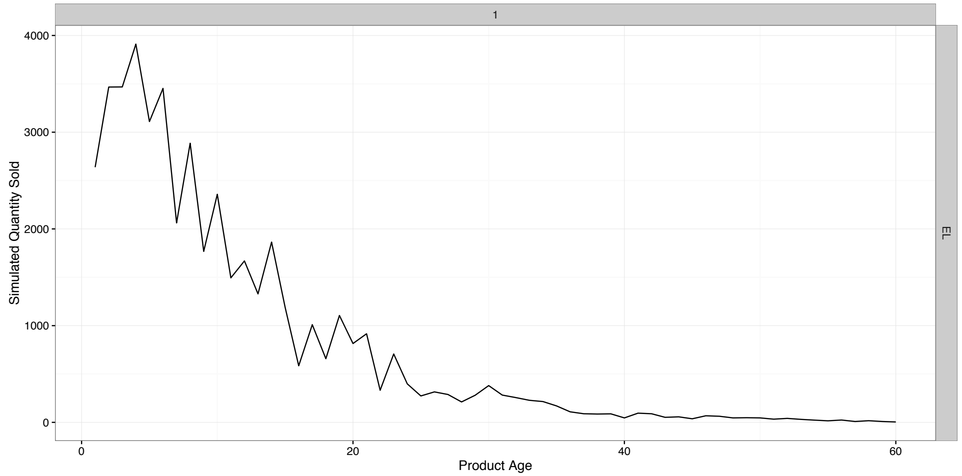
$$qty = f(price, price^2, product\_age, ...)$$

- For a Gaussian model, and one product:

$$qty_i \sim N(X_i\beta, \sigma^2)$$

- For products that sell thousands of units we would fit a log-log model
- For lower volume products that sometimes sell zero units, we fit a count model that does not force the mean to be equal to the variance

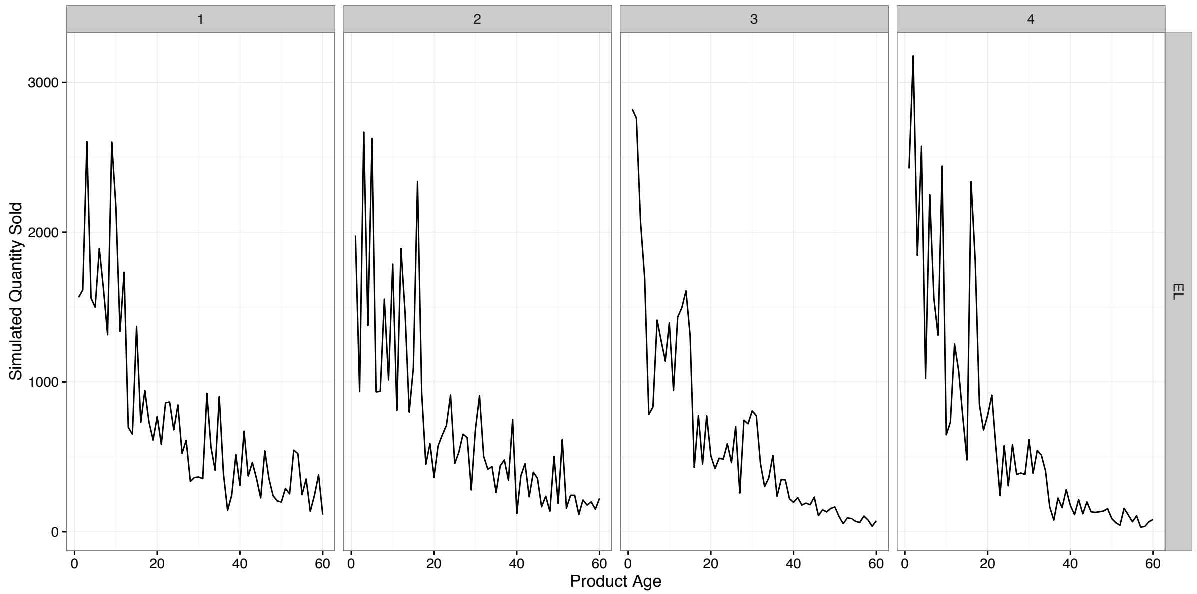
$$qty \sim NegativeBinomial(\mu, \phi)$$
$$\mu = exp(\alpha + \beta_1 * product\_age + \beta_2 * price + ...)$$
$$\sigma^2 = \mu + \mu^2/\phi$$




```
simd2 <- hir_data_sim(N_prod = 1,
                      global_intercept = 8.5,
                      theta = 10,
                      qty_process = "negbinom",
                      primary_price_process = "none",
                      …
                      linkinv = exp)
```

Simulating Data

```
if (process == "normal") {
  data <- data %>%
    mutate(qty = linkinv(product_intercept + product_beta_time * days + product_beta_price * price +
                 error_sd * rnorm(sum(n)))) %>%
    mutate(qty = ifelse(qty <= 0, 0, round(qty)))
} else { # negative binomial
  data <- data %>%
    mutate(mu = linkinv(product_intercept + product_beta_time * day + product_beta_price * price)
           qty = MASS::rnegbin(n = sum(n), mu = mu, theta = theta))
}
```

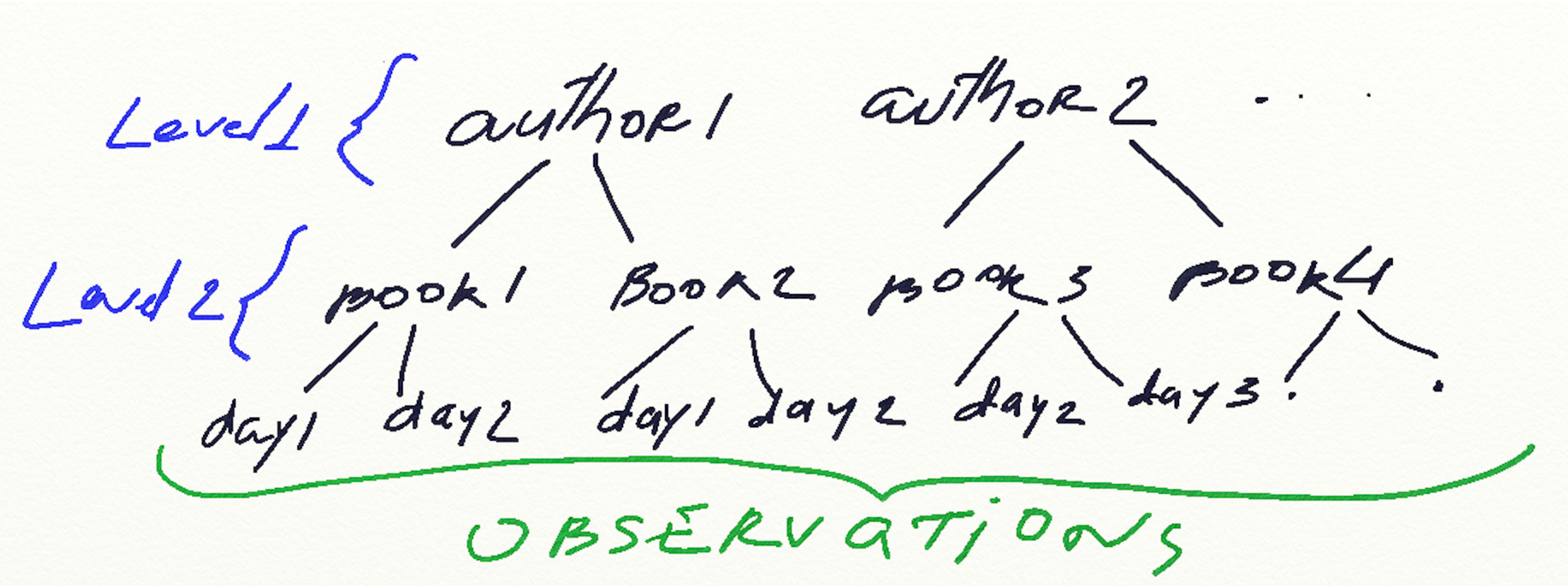


Baseline stan model for a single product

```
data {
  int<lower=0> N;
  int<lower=0> y[N];
  vector[N] t;
}
parameters {
  real alpha;         // overall mean
  real beta;          // time beta
  real<lower=0> phi; // dispersion
}
model {
  vector[N] eta;
  // linear predictor
  eta = alpha + t * beta;
  // priors
  alpha ~ normal(0, 10);
  phi ~ cauchy(0, 2.5);
  beta ~ normal(0, 1);
  // likelihood
  y ~ neg_binomial_2_log(eta, phi);
}
```

```
simd2_m2 <- stan('m2_self_stan_nbinom.stan'
                 data = list(N = nrow(simd2$data),
                             y = simd2$data$qty,
                             t = simd2$data$day),
                 control = list(stepsize = 0.01,
                                adapt_delta = 0.99),
                 cores = 4,
                 iter = 400)

# truth: alpha = 8.5, beta = -0.10, phi = 10
samples <- rstan::extract(simd2_m2,
                          pars = c('alpha',
                          'beta',
                          'phi'))

> lapply(samples, quantile)
$alpha
  0%  25%  50%  75% 100%
 8.3  8.4  8.5  8.6  8.8

$beta
    0%    25%    50%    75%   100%
-0.107 -0.102 -0.100 -0.099 -0.092

$phi
  0%  25%  50%  75% 100%
 6.2 10.1 11.5 13.0 24.1
```

Multiple products, authors, genres



Hierarchical pooling in one slide

Average sales across all books

Average sales for book j

Number of observations for book j

$$\widehat{\alpha}_j^{multilevel} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

Indexes books

Estimate of average sales for book j

Within-book variance

Variance among average sales of different books

Fitting multilevel models in rstanarm

```
install.packages("rstanarm")
```

“Fixed Effects”

```
fit <- stan_glmer.nb(qty ~ product_age + price + price_sqr +
          (1 + product_age + price + price_sqr | product),
                    algorithm = "sampling",
                    seed = 123,
                    cores = 4,
                    iter = 600,
                    data = data)
```

Random Seed

1 Core per Chain

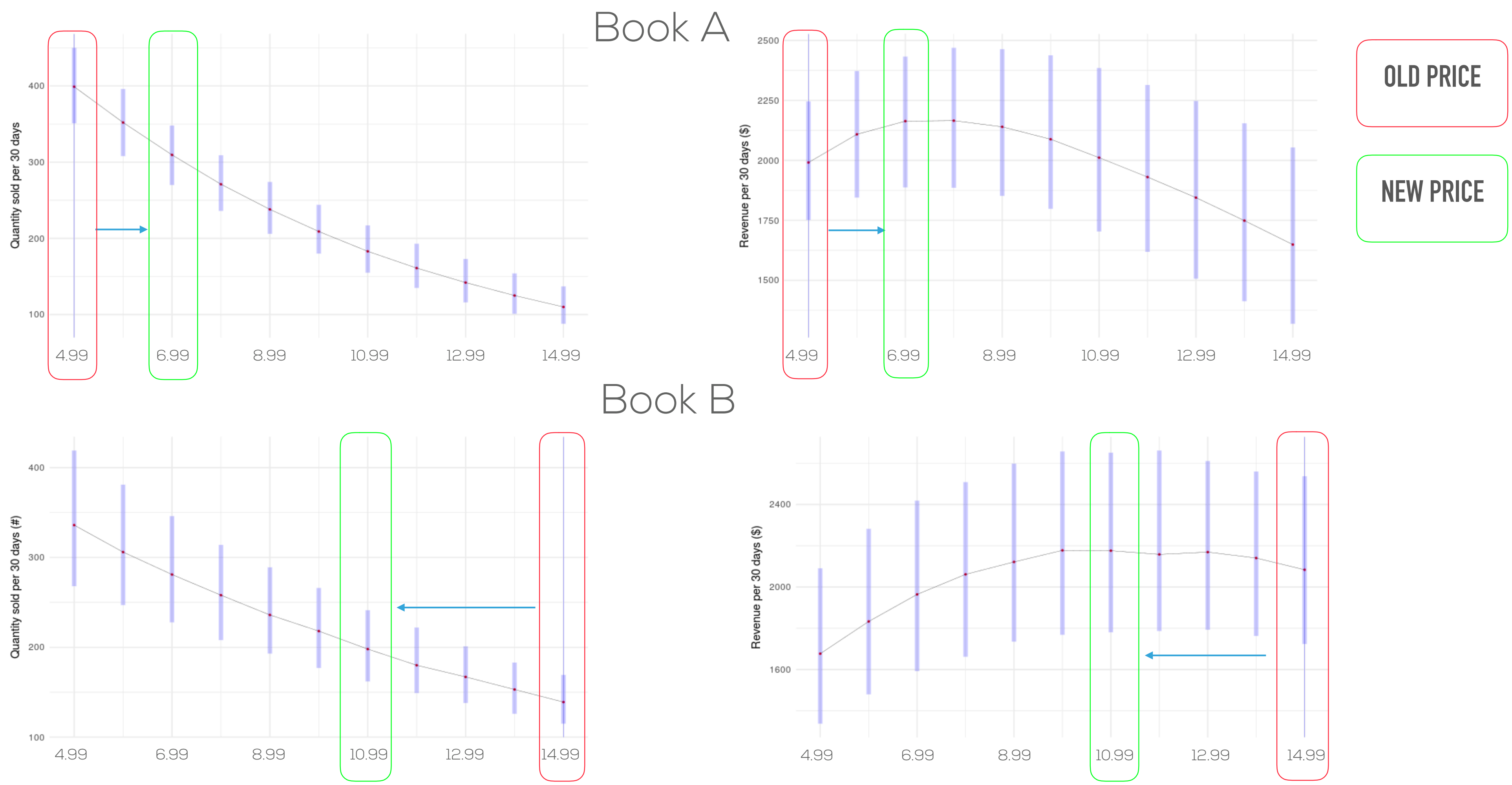
Varying Intercepts

Varying Slopes

Fit using MCMC

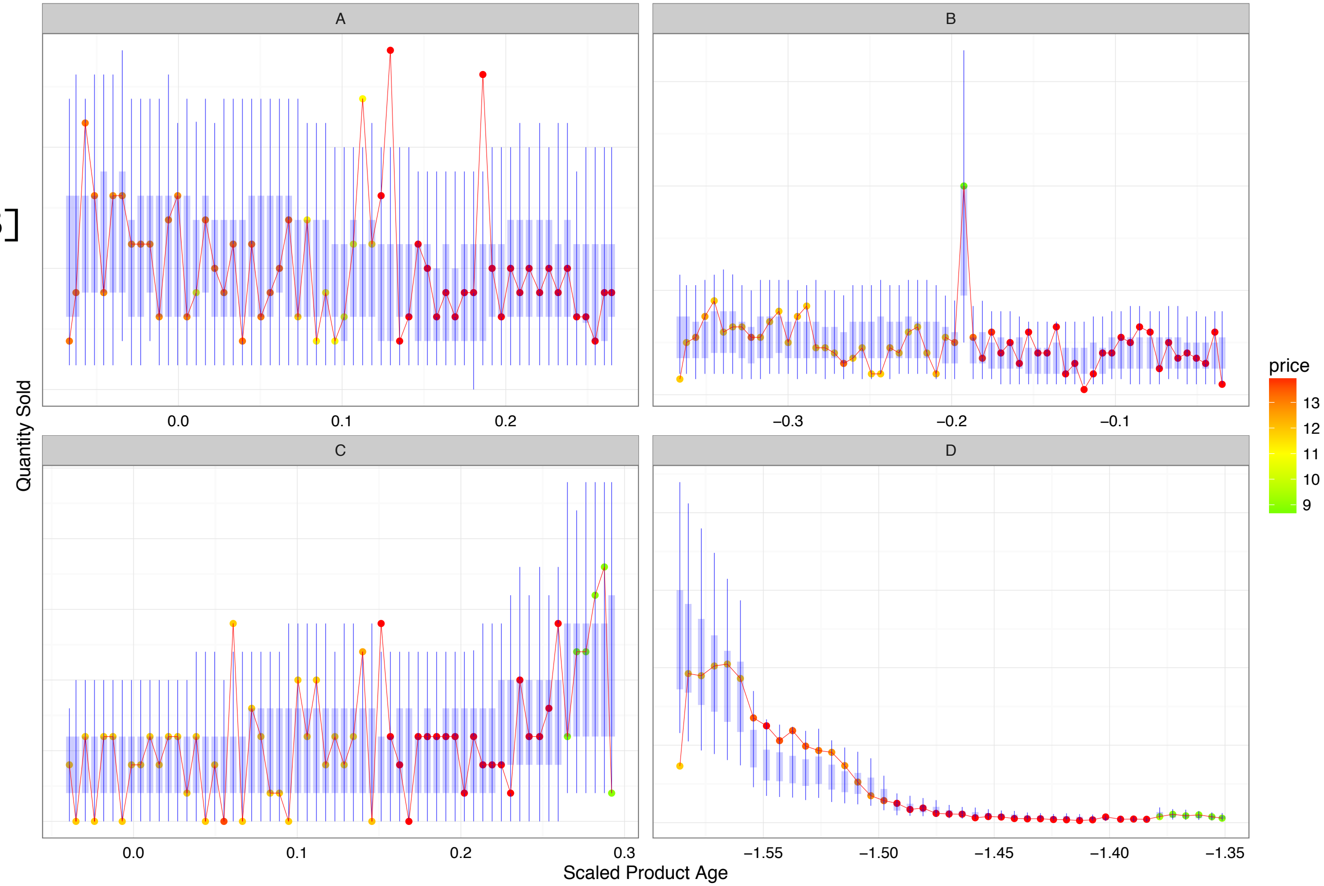
Number of Iterations per Chain

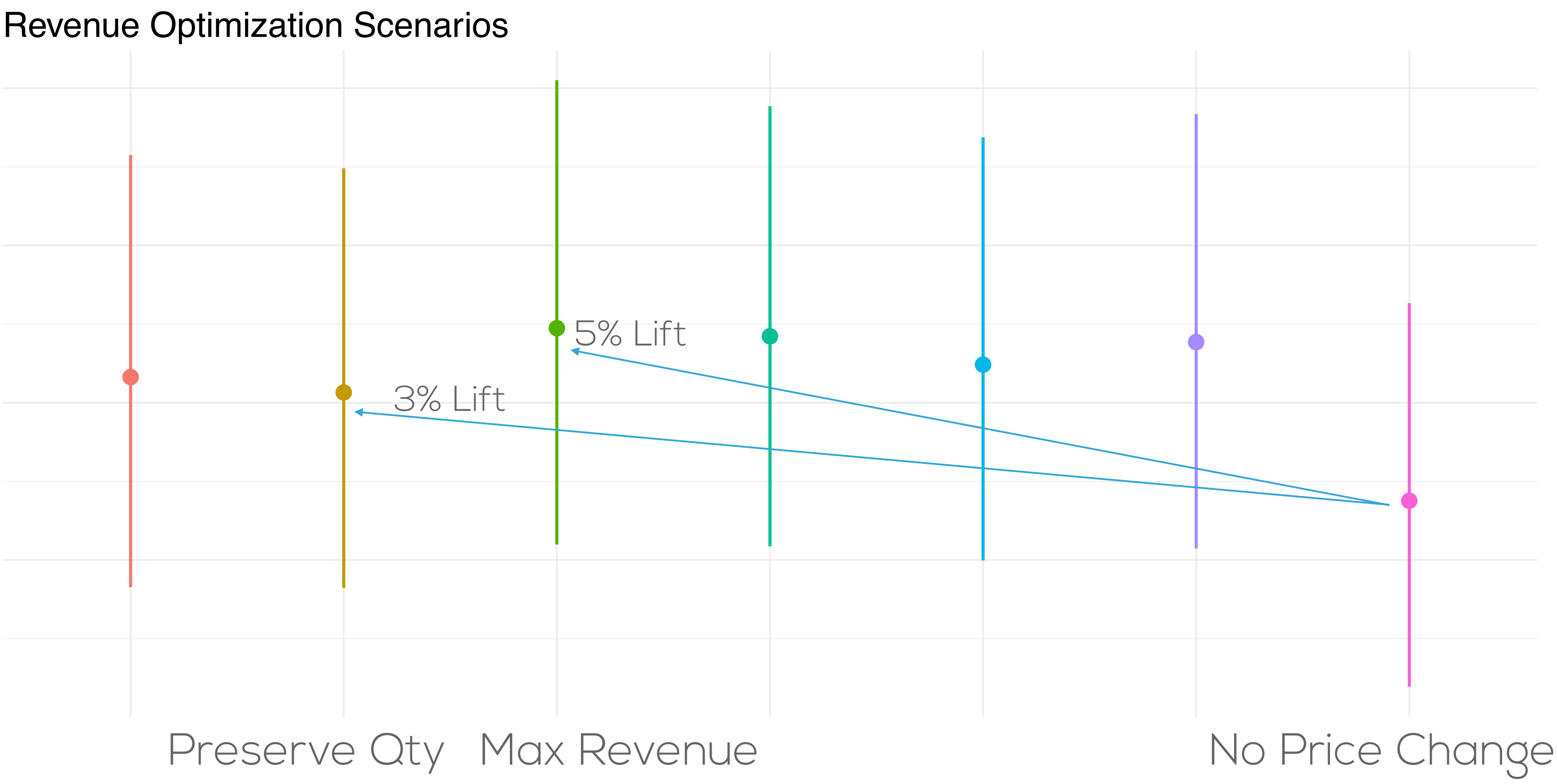
Demand elasticity vs revenue



Calibration Checks

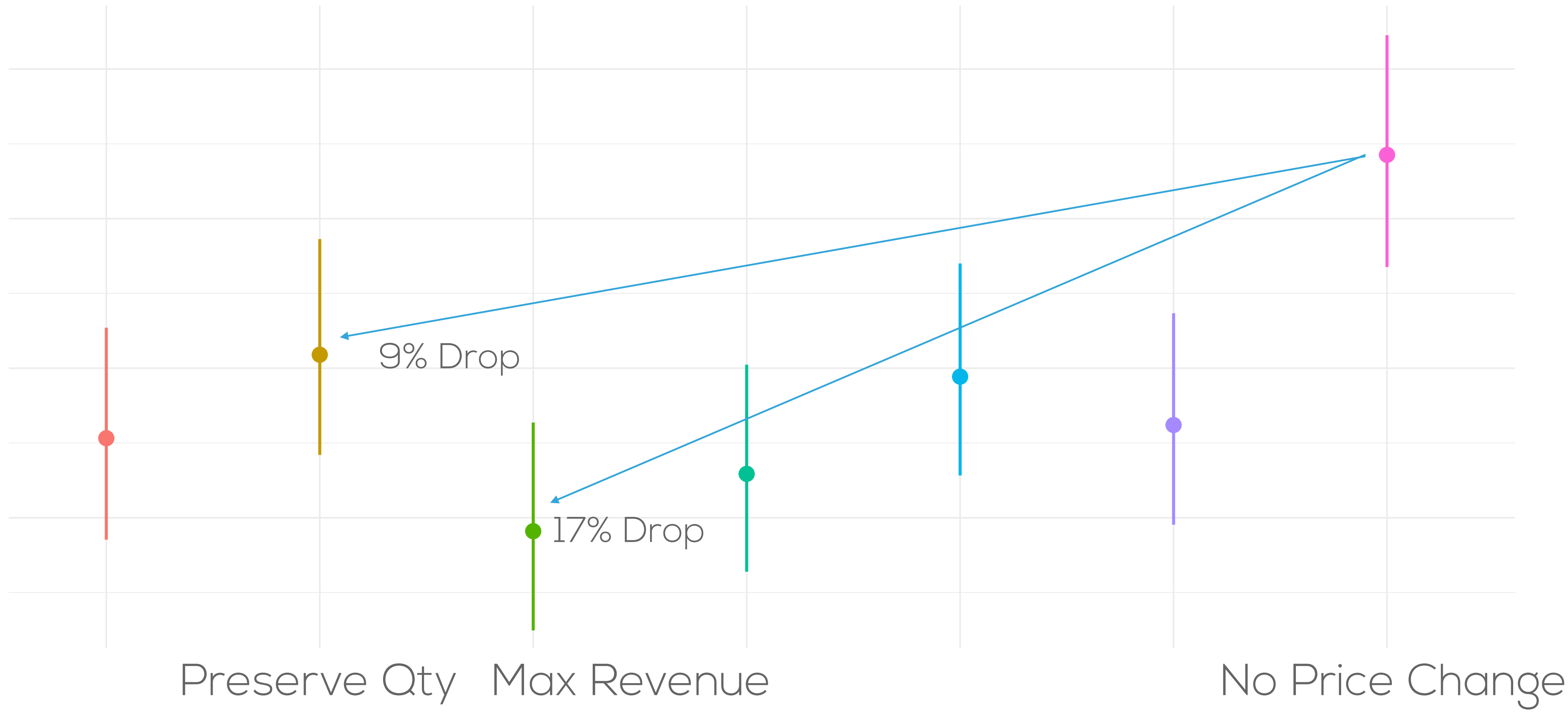
```
> check_calib(d)
      in_90     in_50
1 0.9573893 0.7125305
> check_calib(d, TRUE)
Source: local data frame [203 x 3]

   prod_key     in_90     in_50
      (dbl)     (dbl)     (dbl)
1           0.9333333 0.7166667
2           0.9500000 0.8333333
3           0.9833333 0.8500000
4           0.9666667 0.6500000
5           0.9666667 0.7000000
6           0.9833333 0.8833333
7           0.9666667 0.6833333
8           1.0000000 0.7666667
9           0.8833333 0.6166667
10          0.9500000 0.8500000
..        ...       ...       ...
```

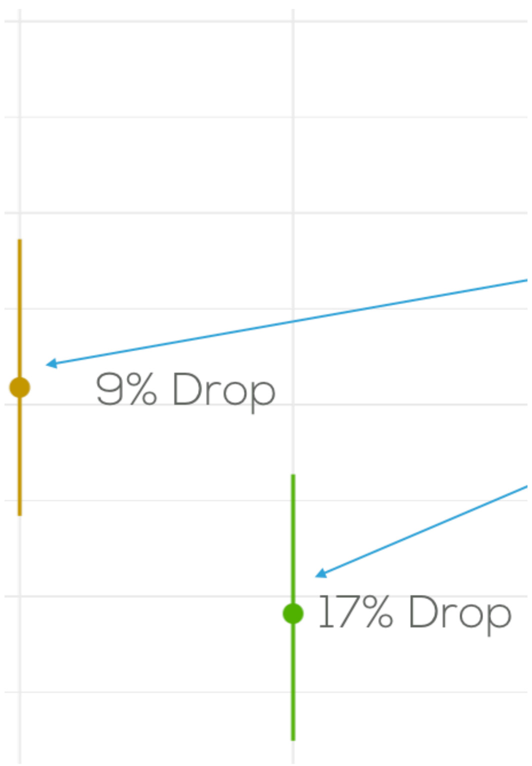




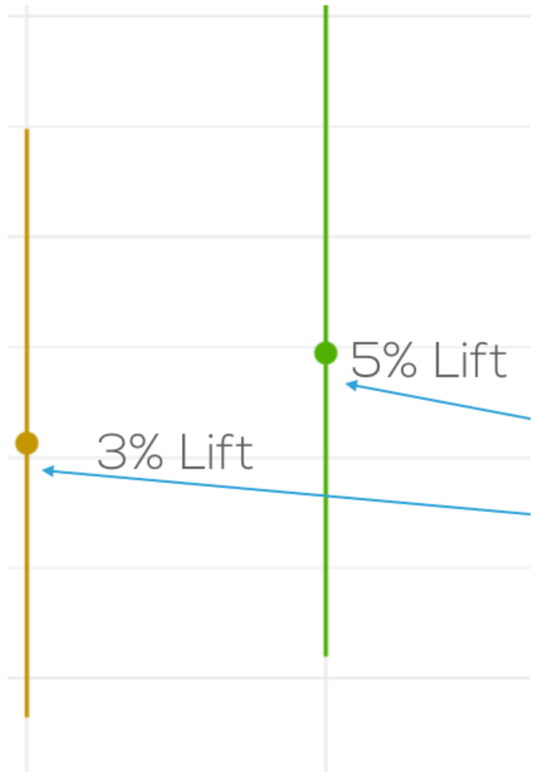
Revenue Optimization Scenarios
5% Lift
3% Lift
Preserve Qty
Max Revenue
No Price Change

Revenue Optimization Scenarios (Qty)





Quantity Drop



Revenue Lift

Socializing your models

$$C(S_t, t) = S_t N(d_1) - K\, e^{-r(T-t)} N(d_2)$$

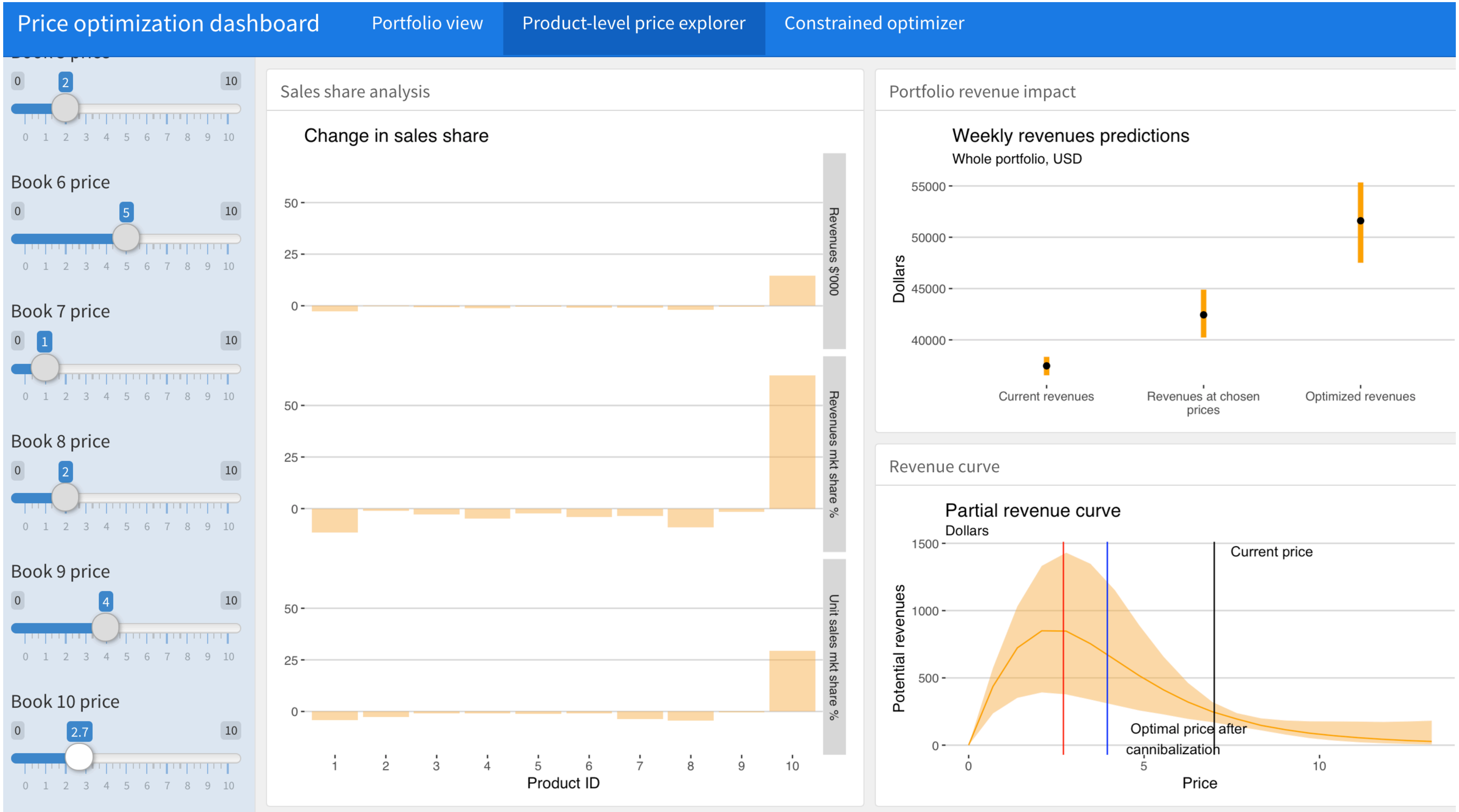
where,

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}}\, dy$$

$$d_1 = \frac{\log\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = \frac{\log\left(\frac{S_t}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$

K: **Option exercise price at maturity**


Price optimization dashboard
Portfolio view
Product-level price explorer
Constrained optimizer
Book 6 price
Book 7 price
Book 8 price
Book 9 price
Book 10 price
Sales share analysis
Change in sales share
Revenues $'000
Revenues mkt share %
Unit sales mkt share %
Product ID
Portfolio revenue impact
Weekly revenues predictions
Whole portfolio, USD
Dollars
Current revenues
Revenues at chosen prices
Optimized revenues
Revenue curve
Partial revenue curve
Dollars
Potential revenues
Current price
Optimal price after cannibalization
Price

Decision Analysis References

- Introductory
  - Understanding Uncertainty, Dennis Lindley, 2006, Chapter 10: Decision Analysis
  - Some Class-Participation Demonstration for Decision Theory and Bayesian Statistics, Andrew Gelman
- Classical (with Bayesian Flavor)
  - Statistical Decision Theory and Bayesian Analysis, James Berger, 1985
- Gelmanese
  - Bayesian Data Analysis, Gelman et. al, Chapter 22: Decision Analysis
  - Analysis of Local Decisions Using Hierarchical Modeling, Applied to Home Radon Measurement and Remediation, Lin et al., 1999, Statistical Science

Thank You!

- eric@generable.com
- @ericnovik
- www.linkedin.com/in/enovik