

# Models are about what changes, and what doesn't

---

Markus Gesmann

28 September 2018

# How do you motivate and build a model from first principles?

The purpose of most models is to understand change, and yet, considering what doesn't change and should be kept constant can be equally important.

Using differential equations to describe the data, I will show how various regression models can be motivated and demonstrate how limitations of generalised linear models can be overcome with a full Bayesian approach.

Along the way I will emphasise the value of domain knowledge and share my experience of engaging with domain experts.

# Why we build models?

- To test our understanding of the world
- To understand change
- To make decision

# Example problem: Selling ice cream



**Figure 1:** How many units of ice cream should I stock?

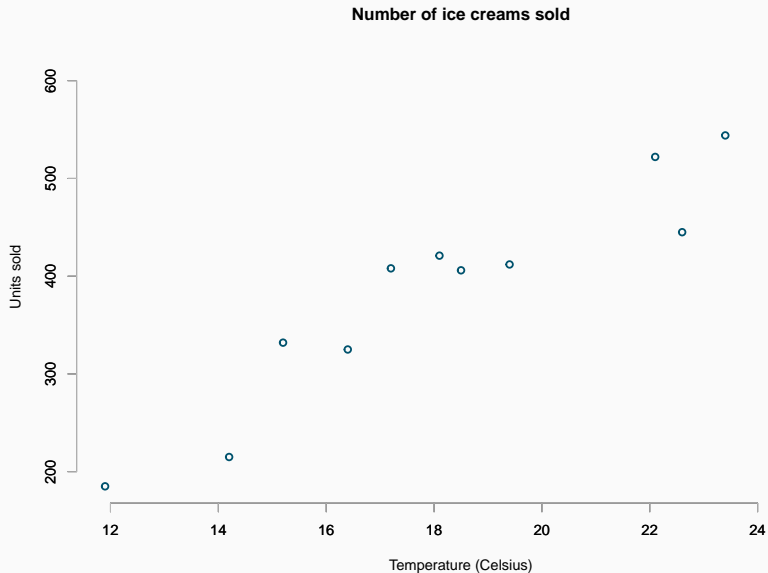
# The challenge

- Create a model that predicts the number of ice creams sold for different temperatures
- Ensure model behaves well when the temperature drops to  $0^{\circ}C$  and for a very hot summer's day at  $35^{\circ}C$ .
- Decide how much ice cream to stock

# Available data

```
icecream <- data.frame(  
  temp=c(11.9, 14.2, 15.2, 16.4, 17.2, 18.1,  
         18.5, 19.4, 22.1, 22.6, 23.4, 25.1),  
  units=c(185L, 215L, 332L, 325L, 408L, 421L,  
         406L, 412L, 522L, 445L, 544L, 614L)  
)
```

# Plot of icream data



# Model building

- Models are about what changes, and what doesn't. Some are useful.
- If something doesn't change, we shouldn't expect a variation of its measurement
- Change can be described with differential equations



**No change**

## Hot or cold, we assume constant demand

$$\frac{d\mu(t)}{dt} = 0, \quad \mu(0) = \mu_0$$

It's solution is  $\mu(t) = \mu_0$  and hence, we could assume for the observable data  $y(t)$ :

$$y(t) \sim \mathcal{N}(\mu_0, \sigma^2)$$

## No change: Critique

Of course, this can't be right. We would expect higher sales of ice cream for higher temperatures.

## Constant change

## Constant change: ODE

Regardless of the starting temperature, we assume constant increase in sales as the temperature increases by  $1^{\circ}\text{C}$ :

$$\frac{d\mu(t)}{dt} = b, \quad \mu(0) = a$$

Solving the ODE gives us the classic linear equation for the mean:

$$\mu(t) = a + b t$$

## Constant change: Constant variance

Now, since we are assuming constant change, I should also assume that the variance ( $\sigma^2$ ) around my data doesn't change either and hence, I could model my data as Gaussian again:

$$y(t) \sim \mathcal{N}(\mu(t), \sigma^2) = \mathcal{N}(a + b t, \sigma^2)$$

## Constant change: Critique

This model makes a little more sense, but it could also predict negative expected sales, if  $a$  is negative. In addition, generating data from this distribution will produce negative values occasional as well, even if  $a > 0$ .

## Constant sales growth rate



## Constant sales growth rate

Assuming a constant sales growth rate means that a  $1^{\circ}C$  increase in temperature is responded by the same proportional change in sales, i.e. for every  $1^{\circ}C$  increase the sales would increase by  $b\%$ .

Thus, we are moving from an additive to a multiplicative scale.

## Constant sales growth rate: ODE

We can describe the change in sales as:

$$\frac{d\mu(t)}{dt} = \mu(t) b, \quad \mu(0) = a$$

The solution to this ODE is exponential growth ( $b > 0$ ), or decay ( $b < 0$ ):

$$\mu(t) = a \exp(bt)$$

## Constant sales growth rate: Log-normal model

On a log-scale we have:

$$\log \mu(t) = \log a + b t$$

This is a linear model again. Thus, we can assume constant change and variance ( $\sigma^2$ ) on a log-scale. A natural choice for the data generating distribution would be:

$$\log y(t) \sim \mathcal{N}(\log a + b t, \sigma^2)$$

## Constant sales growth rate: Constant CV

This model assumes a constant coefficient of variation ( $CV$ ) on the original scale:

$$\begin{aligned} CV &= \frac{\sqrt{Var[y]}}{E[y]} \\ &= \frac{\sqrt{(e^{\sigma^2} - 1)e^{2\mu(t) + \sigma^2}}}{e^{\mu(t) + \sigma^2/2}} \\ &= \sqrt{e^{\sigma^2} - 1} \\ &= \text{const} \end{aligned}$$

## Constant sales growth rate: Critique

Although this model avoids negative sales, it does assume exponential growth as temperatures increase.

Perhaps, this a little too optimistic?

# Constant demand elasticity

## Constant demand elasticity

A constant demand elasticity of  $b = 60\%$  would mean that a 10% increase in the temperature would lead to a 6% increase in demand/ sales.

Thus, the relative change in sales is proportional to the relative change in temperature:

$$\frac{d\mu(t)}{\mu(t)} = b \frac{dt}{t}$$

## Constant demand elasticity: ODE

Re-arranging and adding an initial value gives us the following ODE:

$$\frac{d\mu(t)}{dt} = \frac{b}{t}\mu(t), \quad \mu(1) = a$$

It's solution is:

$$\mu(t) = \exp(a + b \log(t)) = \exp(a) \cdot t^b$$



## Constant demand elasticity: Log-log scale

Now take the log on both sides and we are back to linear regression equation:

$$\log(\mu(t)) = a + b \log(t)$$

Thus, we are assuming constant change on a log-log scale. We can assume a log-normal data generating distribution again, but this time we measure temperature on a log-scale as well.

$$\log y(t) \sim \mathcal{N}(a + b \log t, \sigma^2)$$

## Constant demand elasticity: Critique

This model, as well as the log-transformed model above, predicts ever increasing ice cream sales as temperature rises.

## Constant market size

## Constant market size: ODE

However, if I believe that my market/ sales opportunities are limited, i.e. we reach market saturation, then a logistic growth model would make more sense.

Let's assume as temperature rises, sales will increase exponential initially and then tail off to a saturation level  $M$ . The following ODE would describe this behaviour:

$$\frac{d\mu(t)}{dt} = b \mu(t) (1 - \mu(t)/M), \quad \mu(0) = \mu_0$$

## Constant market size: Integration

Integrating the ODE provides us with the logistic growth curve (often used to describe population growth):

$$\mu(t) = \frac{M}{1 + \frac{M - \mu_0}{\mu_0} e^{-bt}} = \frac{\mu_0 M e^{bt}}{\mu_0 (e^{bt} - 1) + M}$$

## Constant market size: Proportional growth

Setting  $M = 1$  gives me the proportion of maximum sales, and setting  $a := (1 - \mu_0)/\mu_0$  gives us:

$$\begin{aligned}\mu_p(t) &= \frac{1}{1 + \frac{1-\mu_0}{\mu_0}e^{-bt}} \\ &= \frac{1}{1 + e^{-(a+bt)}}$$

Nice, this is logistic regression. Defining  $\text{logistic}(u) := \frac{1}{1+e^{-u}}$ , we can see that the above has a linear expression in  $u$ .

## Constant market size: Logistic regression

Thus, if I set  $z(t) := y(t)/M$ , I could assume the Binomial distribution as the data generating process:

$$z(t) \sim \text{Bin}(M, \text{logistic}(\mu_p(t)))$$

But I don't know  $M$ . Thus, I have to provide a prior distribution, for example a Poisson distribution with parameter (my initial guess)  $M_0$

$$M \sim \text{Pois}(M_0)$$

# Changing gears: Non-linear growth curves



## Data generating distribution: Poisson

We could consider the Poisson distribution as the data generating distribution, with its mean and variance being described by the non-linear growth curve above (this is not a GLM, unlike the models so far):

$$y(t) \sim \text{Pois}(\mu(t)) = \text{Pois} \left( \frac{\mu_0 M e^{bt}}{\mu_0 (e^{bt} - 1) + M} \right)$$

The Poisson distribution has the property that the mean and variance are the same. If  $y \sim \text{Pois}(\lambda)$ , then  $E[y] = \text{Var}[y] = \lambda$ , but more importantly, we are generating integers, in line with the observations.

## Test model with prior predictive distribution

So, let's generate some data to check if this approach makes sense in the temperature range from  $-5^{\circ}C$  to  $35^{\circ}C$ .

Let's put priors over the parameters:

$$\mu_0 \sim \mathcal{N}(10, 1), \quad \text{ice creams sold at } 0^{\circ}C$$

$$b \sim \mathcal{N}(0.2, 0.02^2), \quad \text{initial growth rate}$$

$$M \sim \mathcal{N}(800, 40^2), \quad \text{total market size}$$

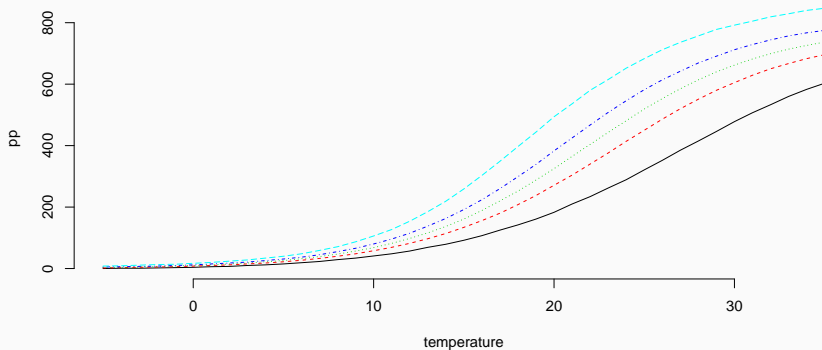
# Generate prior predictive output

With those preparations done I can simulate data:

```
temperature <- c(-5:35)
n <- length(temperature)
nSim <- 10000
set.seed(1234)
# Priors
mu0 = rep(rnorm(nSim, 10, 1), each=n)
b = rep(rnorm(nSim, 0.2, 0.02), each=n)
M = rep(rnorm(nSim, 800, 40), each=n)
lambda = matrix(
  M/(1 + (M - mu0)/mu0 * exp(-b * rep(temperature, nSim))),
  ncol=n)
y_hat_p <- matrix(rpois(nSim*n, lambda), nrow=n)
```

# Prior predictive plot

**Poisson percentiles:**  
2.5%, 25%, 50%, 75%, 97.5%



This looks reasonable, but the interquartile range looks quite narrow.

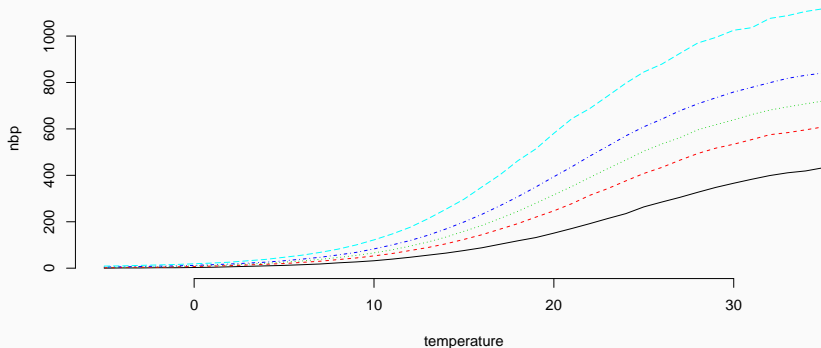
## Allow for over-dispersion: Negative-Binomial

Hence, I want to consider the **Negative-Binomial distribution**, which has an additional shape parameter  $\theta$  to allow for over-dispersion, i.e. the variance can be greater than the mean.

A Negative Binomial distribution with parameters  $\mu$  and  $\theta$  has mean  $\mu$  and variance  $\mu + \mu^2/\theta$ . With  $\theta \rightarrow \infty$  we are back to a Poisson distribution.

# Negative-Binomial prior predictive, $\theta = 20$

**Negative Binomial percentiles:**  
2.5%, 25%, 50%, 75%, 97.5%



This looks a little more realistic and adds more flexibility to my model.

# Fitting the non-linear growth curve model

Finally, I am getting to a stage where it makes sense to make use of my data.

Using `brms` I can fit the non-linear logistic growth curve with a Negative-Binomial data generating distribution.

# Fitting the model with brms

```
library(rstan)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
library(brms)
mdl <- brm(
  bf(units ~ M / (1 + (M - mu0) / mu0 * exp(-b * temp))),
  mu0 ~ 1, b ~ 1, M ~ 1, nl = TRUE),
  data = icecream,
  family = brmsfamily("negbinomial", "identity"),
  prior = c(prior(normal(10, 100), nlpar = "mu0", lb=0),
            prior(normal(0.2, 1), nlpar = "b", lb=0),
            prior(normal(800, 400), nlpar = "M", lb=0)),
  seed = 1234, control = list(adapt_delta=0.9))
```



# Model output

Family: negbinomial

Links: mu = identity; shape = identity

Formula: units ~ M/(1 + (M - mu0)/mu0 \* exp(-b \* temp))

mu0 ~ 1

b ~ 1

M ~ 1

Data: icecream (Number of observations: 12)

Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;

total post-warmup samples = 4000

Population-Level Effects:

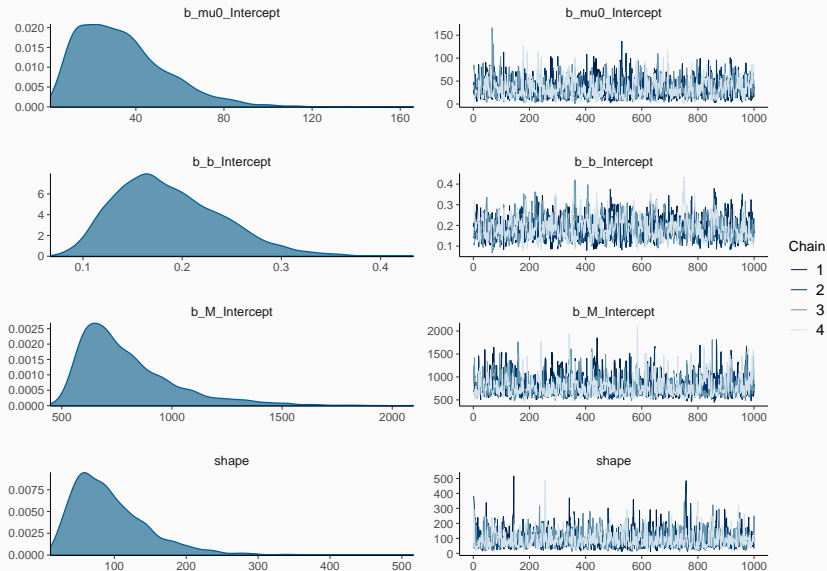
	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
mu0_Intercept	33.68	20.01	6.21	81.11	1011	1.00
b_Intercept	0.19	0.05	0.10	0.30	887	1.00
M_Intercept	794.45	215.38	536.00	1361.55	1036	1.00

Family Specific Parameters:

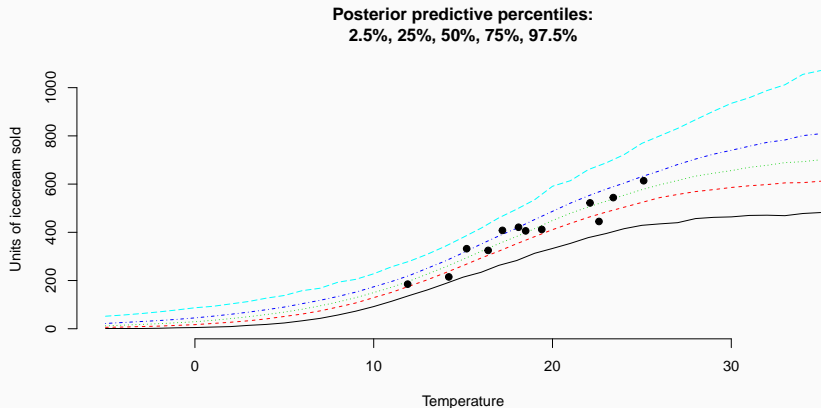
	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
shape	94.78	54.40	24.53	228.93	1417	1.00

Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample

# Posterior parameter plots



# Posterior predictive plot



Hurray! This looks really good.

## Non-linear growth curve model: Critique

I am happy with the model and its narrative. The parameters are easy interpretable, the curve can be explained and the model generates positive integers, in-line with the data.

I can use the distribution output to decide how much ice cream I want to stock.

Of course, there will be many other factors that influence the sales of ice cream, but I am confident that the temperature is the main driver.

# Conclusions

# What I learned

Approaching the model build by thinking about how the sale of ice cream will change, as the temperature changes, and considering what to keep constant has led me down the road of differential equations.

With a small data set like this, starting from a simple model and reasoning any additional complexity has helped me to develop a better understanding of what a reasonable underlying data generating process might be.

Thanks to [Stan](#) and [brms](#), I was freed from the constraints under GLMs and I could select my own non-linear model for the expected sales and distribution family.

## Further readings and more details

### Books:

- Deep Simplicity: Bringing Order to Chaos and Complexity, *John Gribbin*
- Data Analysis Using Regression and Multilevel/Hierarchical Models, *Andrew Gelman and Jennifer Hill*

### On my blog:

- Generalised Linear Models in R
- Principled Bayesian Workflow
- Models are about what changes, and what doesn't
- Use domain knowledge to review prior distributions

**The End. Thanks.**