Inf2 - Foundations of Data Science: Regression and inference - Generalised linear models

THE UNIVERSITY of EDINBURGH informatics

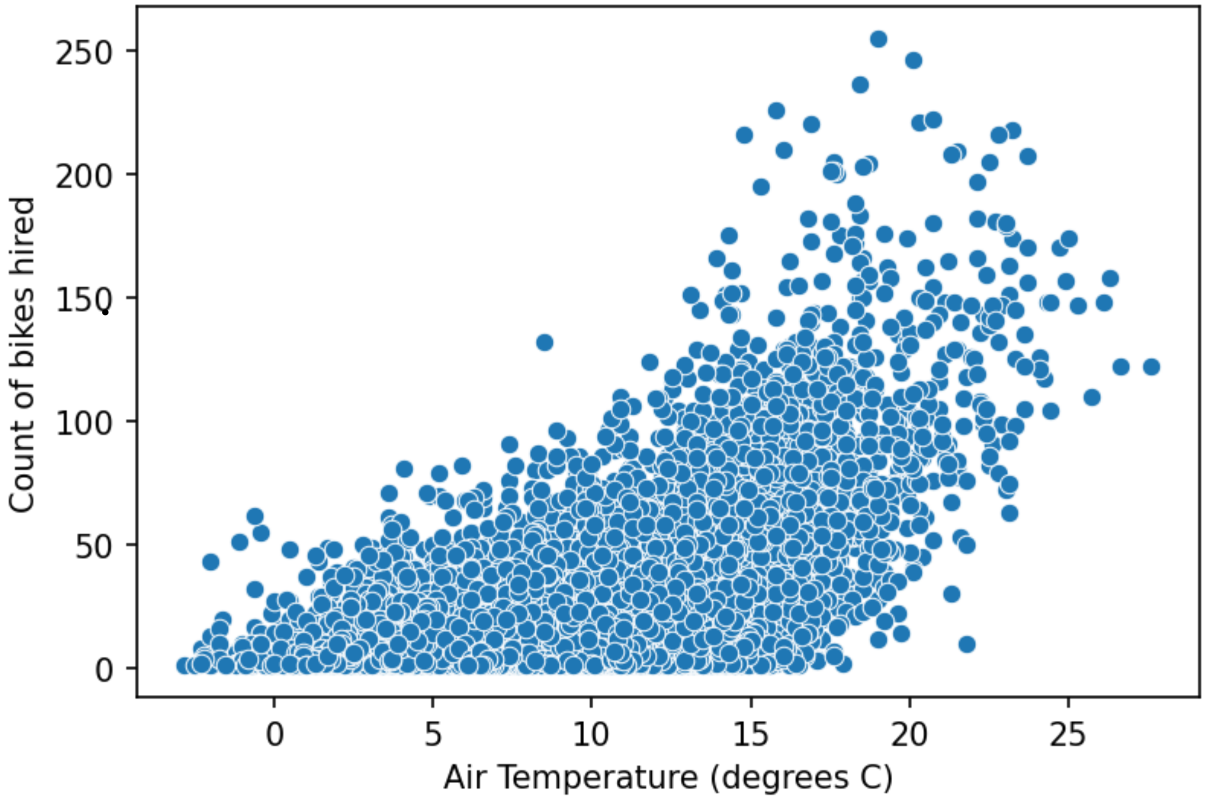
FOUNDATIONS OF DATA SCIENCE

We want to investigate the relationship between the number of bikes hired in an hour and the mean temperature during that hour

Is there a problem with using ordinary least squares linear regression to do this?

Image copyright Pashley Cycles

Data sources:
- Edinburgh Just Eat Bikes data 2020
- Edinburgh temperature observations, Met Office via MIDAS



Overview

Monday

1. The maximum likelihood principle
2. Application of max likelihood to a simple example
3. Application of max likelihood to linear regression

Today

0. Recap
1. Max likelihood with non-normal distributions
2. Poisson regression
3. Logistic regression and generalised linear models

Inf2 - Foundations of Data Science: Regression and inference - Recap of max likelihood applied to linear regression

THE UNIVERSITY of EDINBURGH
informatics

FOUNDATIONS OF DATA SCIENCE

Likelihood and log likelihood as a function of parameters

$P(Y = y_1, \ldots, y_n \mid \mu, \sigma^2)$

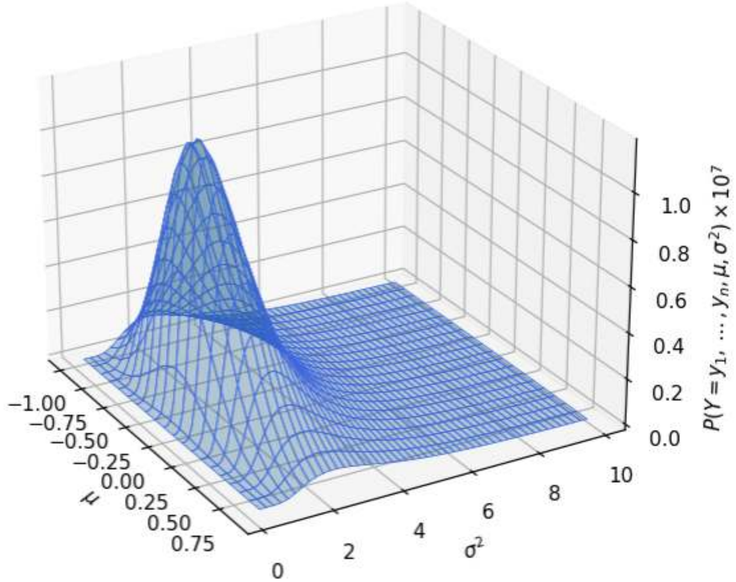
Data:

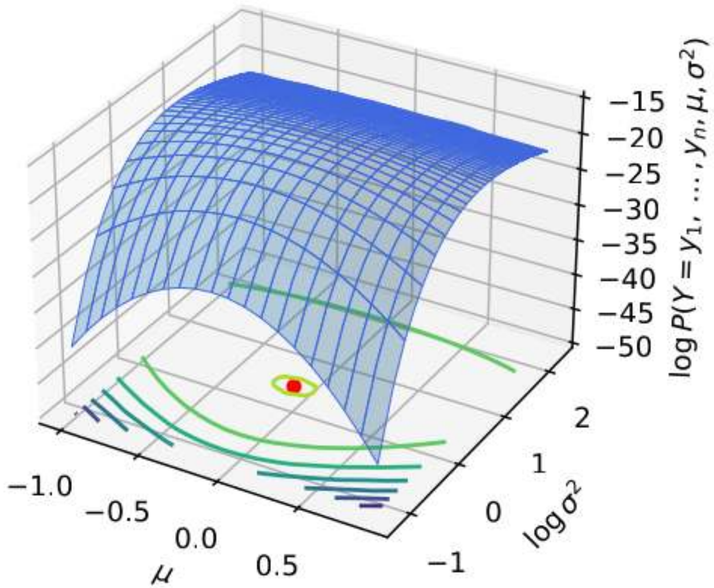
$y_1, \ldots, y_{10}$ drawn from

$\mathcal{N}(0, 1)$

	Data
y_1	1.624345
y_2	-0.611756
y_3	-0.528172
y_4	-1.072969
y_5	0.865408
y_6	-2.301539
y_7	1.744812
y_8	-0.761207
y_9	0.319039
y_{10}	-0.249370

Not mentioned: optimisation



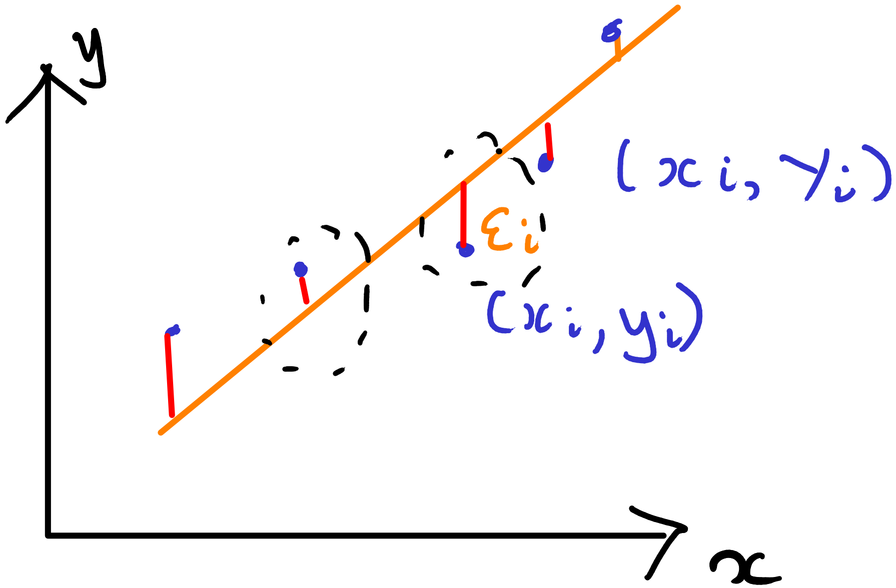


Application of max likelihood to linear regression

$$Y_i = \beta_0 + \beta_1 x_i + \underbrace{\varepsilon_i}_{\text{error term}}$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

residual



OR

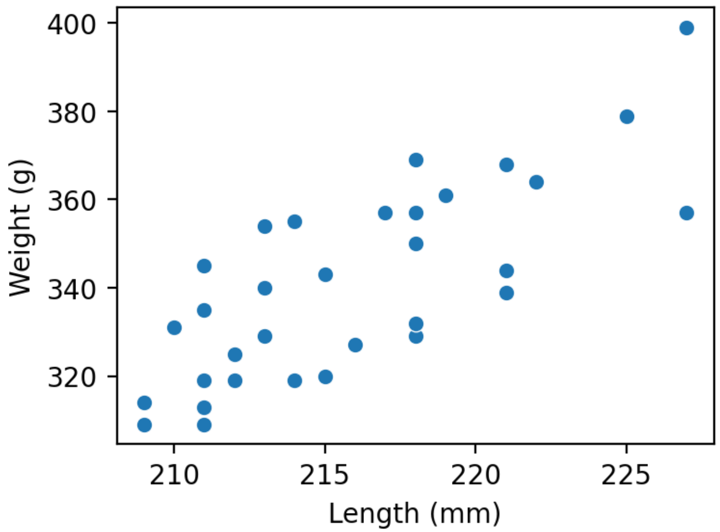
$$Y_i \sim N(\overbrace{\beta_0 + \beta_1 x_i}^{\mu}, \sigma^2)$$

$$\ln p(\underline{Y} = y_1, \dots, y_n ; x_1, \dots, x_n \mid \overbrace{\beta_0, \beta_1}^{\mu}, \sigma^2)$$

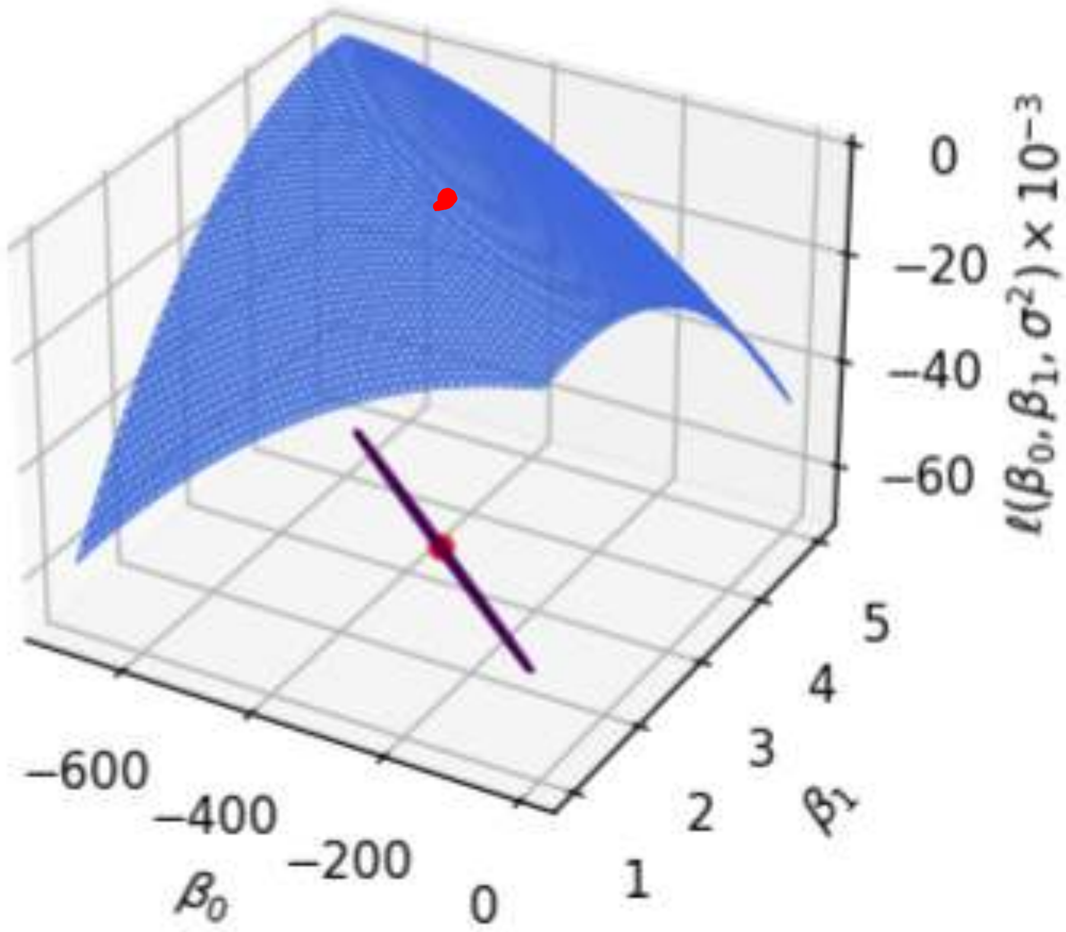
$$= \sum_{i=1}^{n} \left(-\frac{1}{2} \ln 2\pi\sigma^2 - \frac{1}{2} \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^2} \right)$$

Log likelihood of coefficients

Peter Trimming, Wikimedia Commons, CC BY 2.0



Data from Wauters and Dhondt 1989



Understanding more regression output

```
results = smf.ols('Weight ~ Length', data=datf).fit()
results.summary()
```

OLS Regression Results

Dep. Variable:	Weight	**R-squared:**	0.597
Model:	OLS	**Adj. R-squared:**	0.583
Method:	Least Squares	**F-statistic:**	44.37
Date:	Sun, 10 Jan 2021	**Prob (F-statistic):**	2.24e-07
Time:	21:08:04	**Log-Likelihood:**	-129.18
No. Observations:	32	**AIC:**	262.4
Df Residuals:	30	**BIC:**	265.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-382.7372	108.680	-3.522	0.001	-604.692	-160.783
Length	3.3515	0.503	6.661	0.000	2.324	4.379

Omnibus:	8.046	**Durbin-Watson:**	2.337
Prob(Omnibus):	0.018	**Jarque-Bera (JB):**	2.231
Skew:	0.092	**Prob(JB):**	0.328
Kurtosis:	1.720	**Cond. No.**	9.38e+03

Bigger better

Max log likelihood $\ln \hat{L}$

Akaike Information criterion

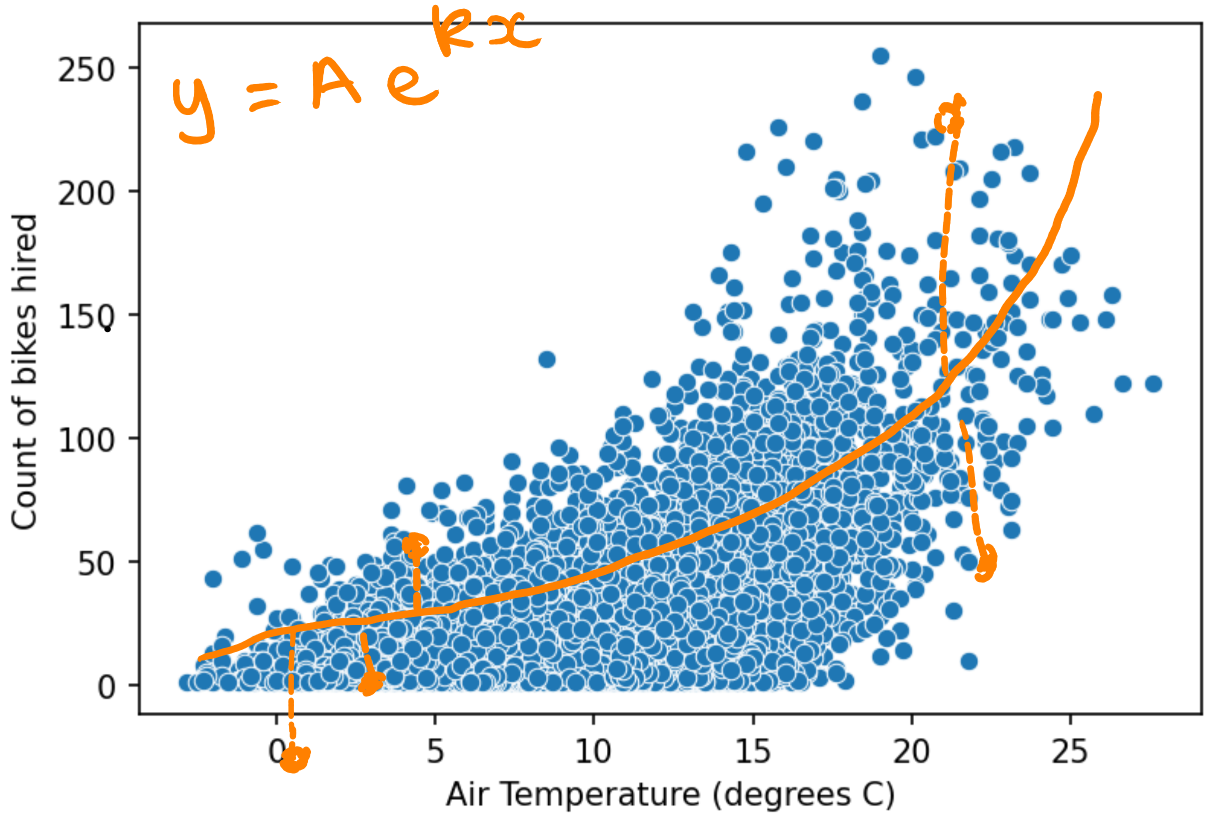
$$\underbrace{2k}_{\text{number of parameters}} - \underbrace{2\ln \hat{L}}_{\text{log likelihood}}$$

smaller better

We want to investigate the relationship between the number of bikes hired in an hour and the mean temperature during that hour

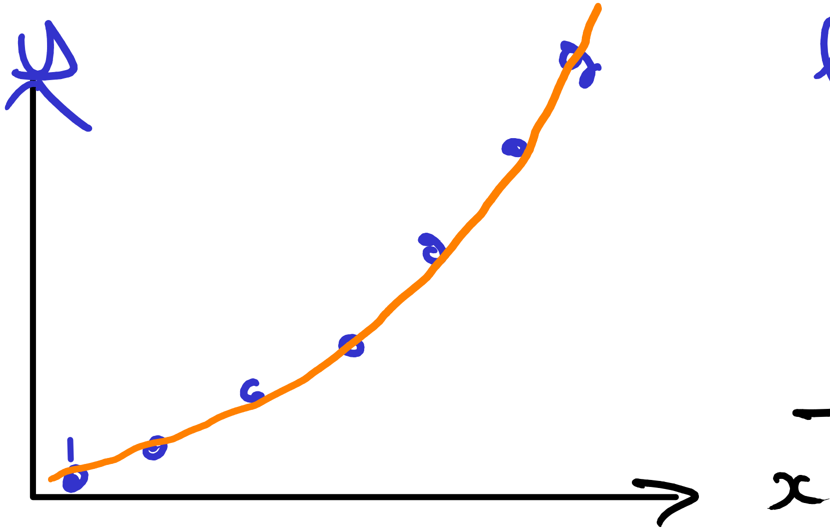
Is there a problem with using ordinary least squares linear regression to do this?

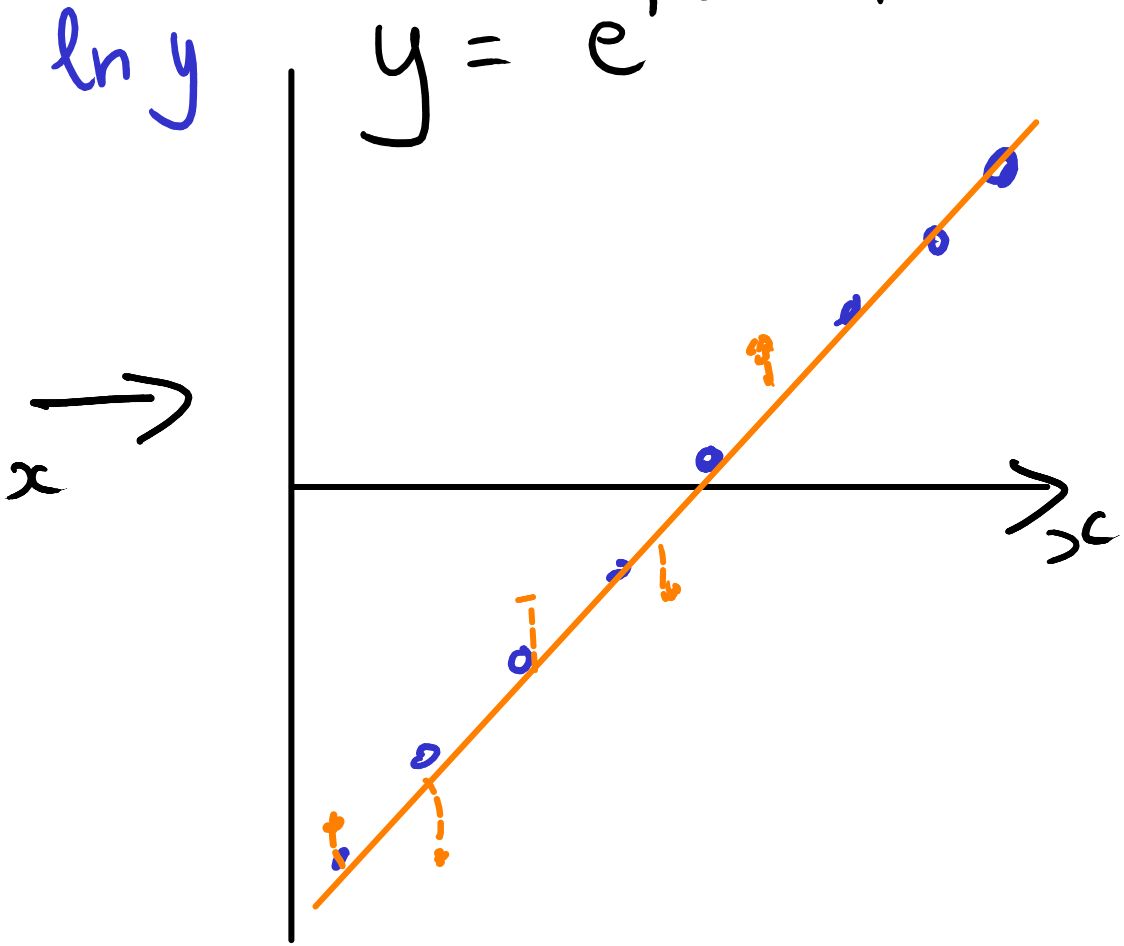
Are there any techniques described in the course so far that could fit the data?



$$\ln y = \beta_0 + \beta_1 x$$

$$y = e^{\beta_0 + \beta_1 x}$$





Inf2 - Foundations of Data Science: Regression and inference - Max likelihood of univariate non-normal distributions

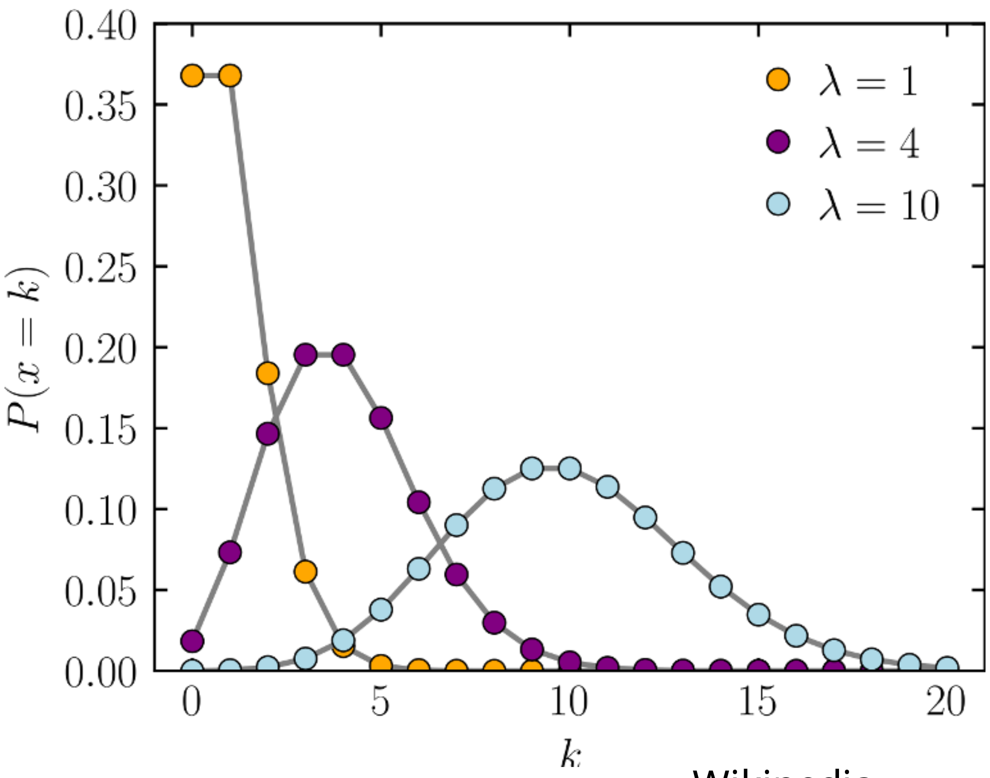
THE UNIVERSITY of EDINBURGH informatics

FOUNDATIONS OF DATA SCIENCE

Max likelihood for models other than the normal

We don't have to assume the data is normally distributed.

E.g. Poisson distribution



Wikipedia

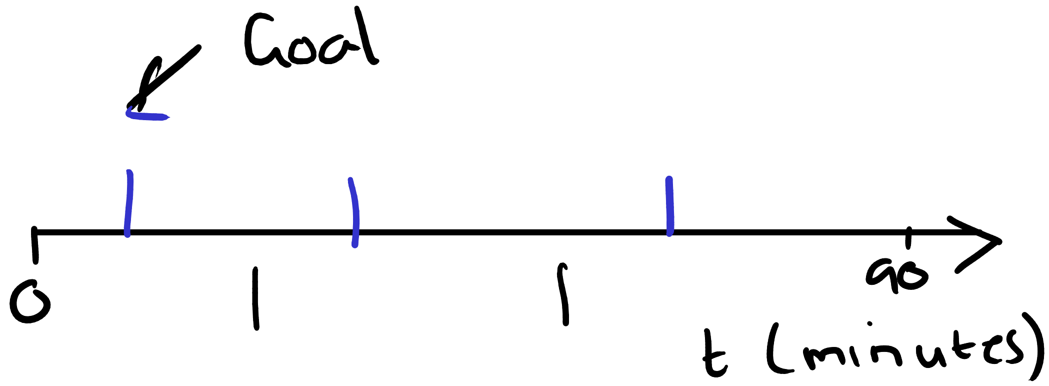
$$P(Y=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$k = 0, 1, 2, \ldots$

$E[Y] = \lambda = \mu$

$V[Y] = \lambda = \sigma^2$

E.g. Number of goals in World Cup football matches



Expected number of goals in a match $\lambda = 2.5$

$$P(Y=k) = \frac{2.5^k e^{-2.5}}{k!}$$

$$P(Y=0) = \frac{2.5^0 e^{-2.5}}{0!} = e^{-2.5} = 0.082$$

$$P(Y=1) = 0.205$$

$$P(Y=2) = 0.257$$

Number of deaths by horse kicks in the Prussian army

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
G	—	2	2	1	—	—	1	1	—	3	—	2	1	—	—	1	—	1	—	1
I	—	—	—	2	—	3	—	2	—	—	—	1	1	1	—	2	—	3	1	—
II	—	—	—	2	—	2	—	—	1	1	—	—	2	1	1	—	—	2	—	—
III	—	—	—	1	1	1	2	—	2	—	—	—	1	—	1	2	1	—	—	—
IV	—	1	—	1	1	1	1	—	—	—	—	1	—	—	—	—	1	1	—	—
V	—	—	—	—	2	1	—	—	1	—	—	1	—	1	1	1	1	1	1	—
VI	—	—	1	—	2	—	—	1	2	—	1	1	3	1	1	1	—	3	—	—
VII	1	—	1	—	—	—	1	—	1	1	—	—	2	—	—	2	1	—	2	—
VIII	1	—	—	—	1	—	—	1	—	—	—	—	1	—	—	—	1	1	—	1
IX	—	—	—	—	—	2	1	1	1	—	2	1	1	—	1	2	—	1	—	—
X	—	—	1	1	—	1	—	2	—	2	—	—	—	—	2	1	3	—	1	1
XI	—	—	—	—	2	4	—	1	3	—	1	1	1	1	2	1	3	1	3	1
XIV	1	1	2	1	1	3	—	4	—	1	—	3	2	1	—	2	1	1	—	—
XV	—	1	—	—	—	—	—	1	—	1	1	—	—	—	2	2	—	—	—	—

Bortkewitsch 1898

$$\underline{Y} = (0, 2, 2, 0, \ldots)$$

$y_1 \quad y_2 \quad \ldots \quad y_{280}$

$$n_k = \sum_{i=1}^{n} I(y_i = k)$$

k	n_k
0	144
1	91
2	32
3	11
4	2

Log likelihood calculation of Poisson distribution

Log likelihood $l = \ln P(\underline{Y} = y_1, \dots, y_n \mid \lambda)$

$$= \sum_{i=1}^{n} \ln P(Y = y_i)$$

$$= \sum_{i=1}^{n} \ln \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$$

$$= \sum_{i=1}^{n} \left(y_i \ln \lambda + (-\lambda) - \ln y_i! \right)$$

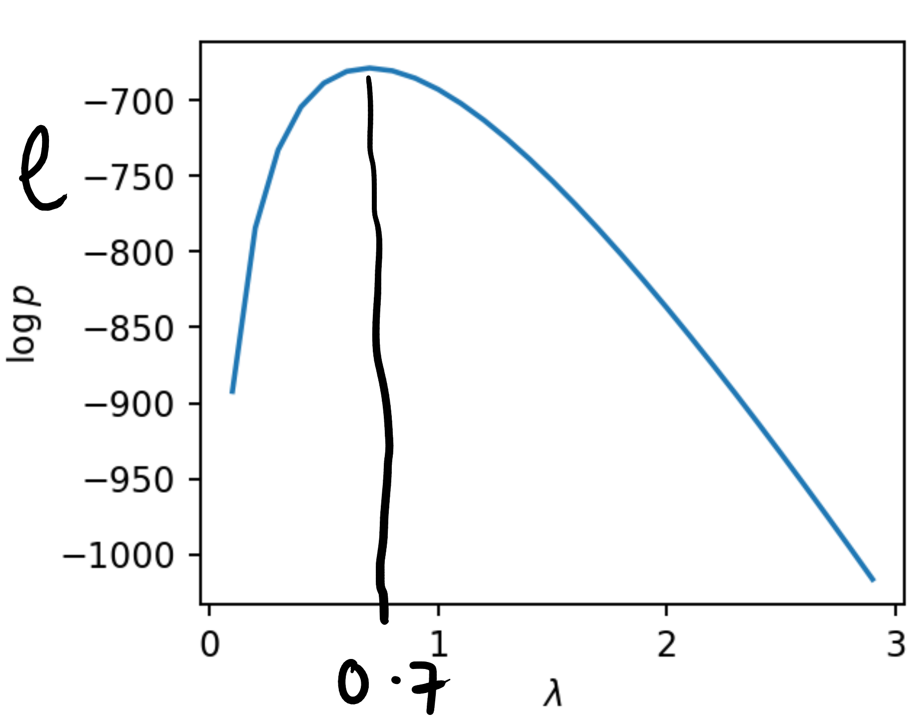
$$l(\lambda) = \ln \lambda \sum_{i=1}^{n} y_i - n\lambda - \sum_{i=1}^{n} \ln y_i!$$

$$\ell = \ln P(\underline{Y} = y_1, \ldots, y_n) = \ln \lambda \sum_{i=1}^{n} y_i - n\lambda - \sum_{i=1}^{n} \ln y_i!$$

$$\frac{d\ell}{d\lambda} = 0$$

$$\vdots$$

$$\Rightarrow \lambda_{MLE} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

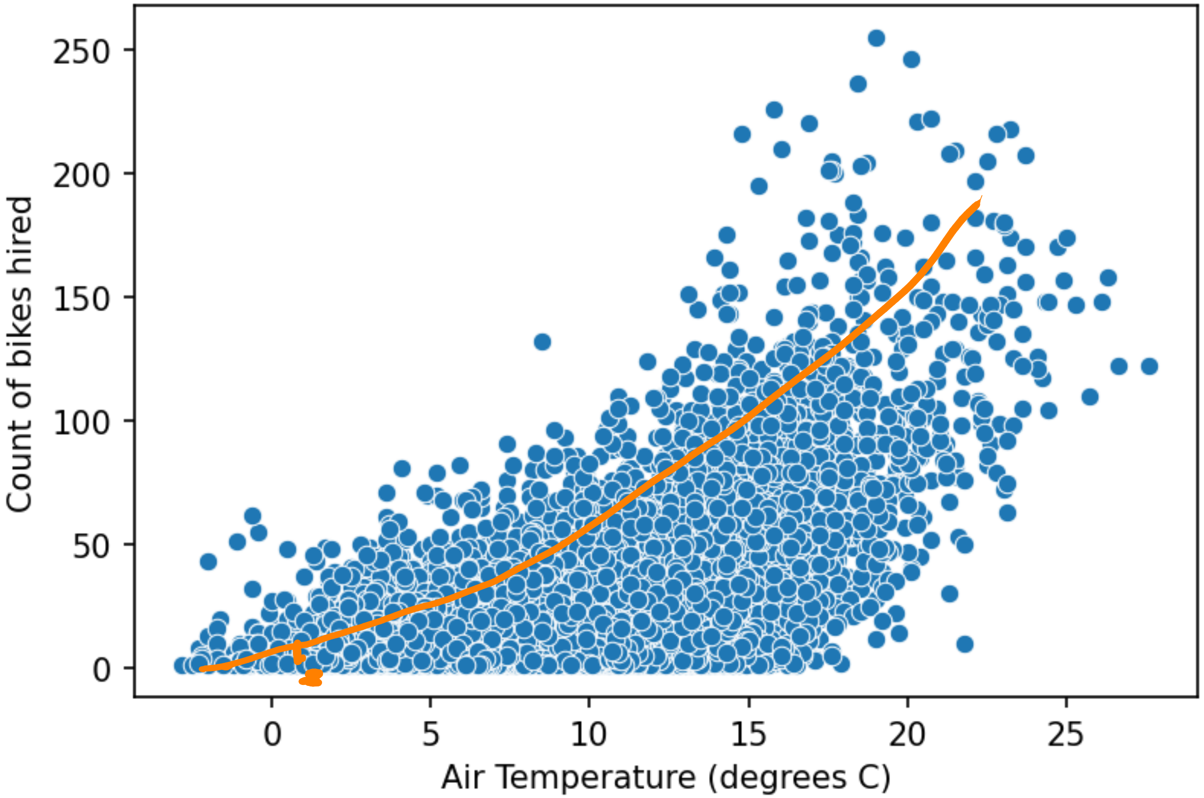


Inf2 - Foundations of Data Science: Regression and inference - Poisson regression

THE UNIVERSITY of EDINBURGH
informatics

FOUNDATIONS OF DATA SCIENCE

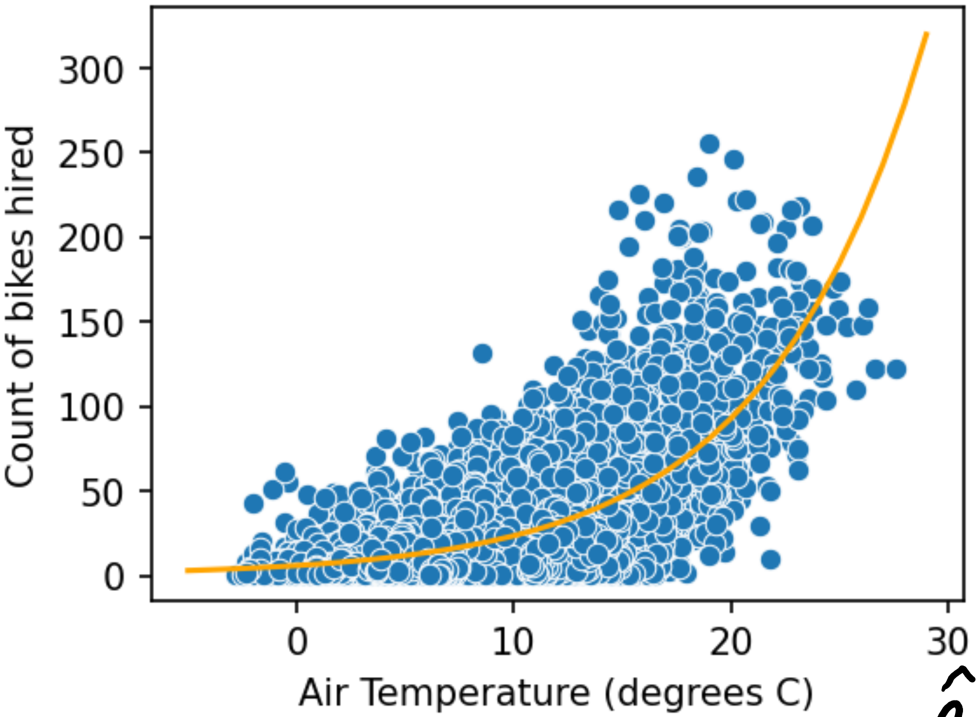
Poisson regression



$$Y_i \sim \text{Poisson}\left(\underbrace{e^{\beta_0 + \beta_1 x_i}}_{\lambda}\right)$$

$$\ln \lambda = \beta_0 + \beta_1 x$$

Results with statsmodels GLM



Generalized Linear Model Regression Results

Dep. Variable:	count	**No. Observations:**	8301
Model:	GLM	**Df Residuals:**	8299
Model Family:	Poisson	**Df Model:**	1
Link Function:	Log	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-84533.
Date:	Wed, 01 Mar 2023	**Deviance:**	1.3111e+05
Time:	06:46:41	**Pearson chi2:**	1.40e+05
No. Iterations:	5	**Pseudo R-squ. (CS):**	1.000
Covariance Type:	nonrobust		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	1.7861	0.006	304.092	0.000	1.775	1.798
air_temperature	0.1373	0.000	323.057	0.000	0.136	0.138

$\hat{\beta}_0$

$\hat{\beta}_1$

$$\ln \lambda = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\lambda = e^{\hat{\beta}_0 + \hat{\beta}_1 x} = e^{\hat{\beta}_0} \underbrace{e^{\hat{\beta}_1 x}}$$

$$e^{0.1373} = 1.14$$

Poisson regression

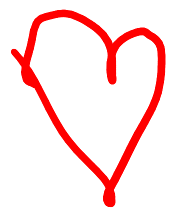
$$l = \ln P(\underline{Y} = y_1, \dots, y_n)$$

$$l(\beta_0, \beta_1) = \sum_{i=1}^{n} (\beta_0 + \beta_1 x_i) y_i - \sum_{i=1}^{n} e^{\beta_0 + \beta_1 x_i} - \sum_{i=1}^{n} \ln y_i!$$

Optimise β_0 & β_1

To my Valentine, Poisson Regression

Roses are red



Violets are blue

Some things aren't normal

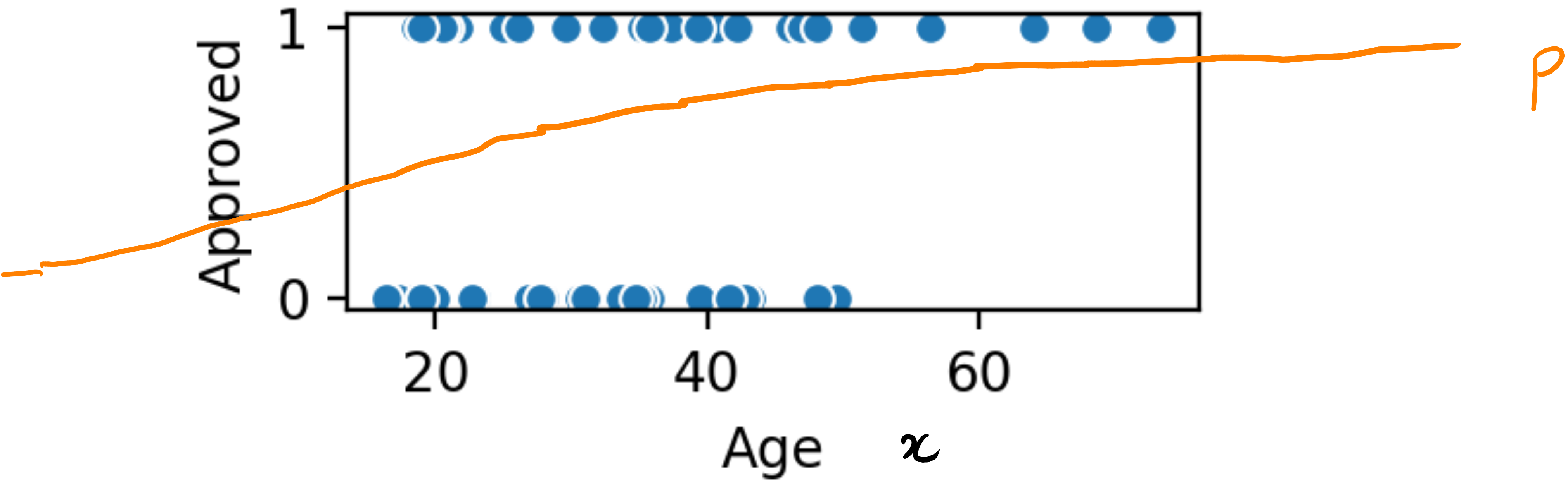
and nor are you

Inf2 - Foundations of Data Science: Regression and inference - Logistic regression and generalised linear models

THE UNIVERSITY of EDINBURGH
informatics

FOUNDATIONS OF DATA SCIENCE

Excercise



What distribution would we use to model the data here?

Bernoulli - Parameter p

How would the parameter of that distribution depend on x (Age)?

Logistic function

$p = \text{Logistic}(\beta_0 + \beta_1 x)$

Generalised linear models (GLMs)

	Distribution	Link function
linear regression	Normal	$\mu = \beta_0 + \beta_1 x \quad , \sigma^2$
Poisson regression	Poisson	$\ln \lambda = \beta_0 + \beta_1 x$
logistic regression	Bernoulli	$\ln \dfrac{p}{1-p} = \beta_0 + \beta_1 x$

Link functions

Expected value $\mu = E(Y|x)$ of a Bernoulli dist is p

" " $\mu = E(Y|x)$ " " Poisson dist is λ

In general the link function is denoted $g(\mu)$ where $\mu = E(Y|x)$ for that distribution:

$$g(\mu) = \beta_0 + \beta_1 x$$

To make predictions, we invert the link function:

$$\mu = g^{-1}(\beta_0 + \beta_1 x)$$

Inf2 - Foundations of Data Science: Regression and inference - And finally...

THE UNIVERSITY of EDINBURGH
informatics

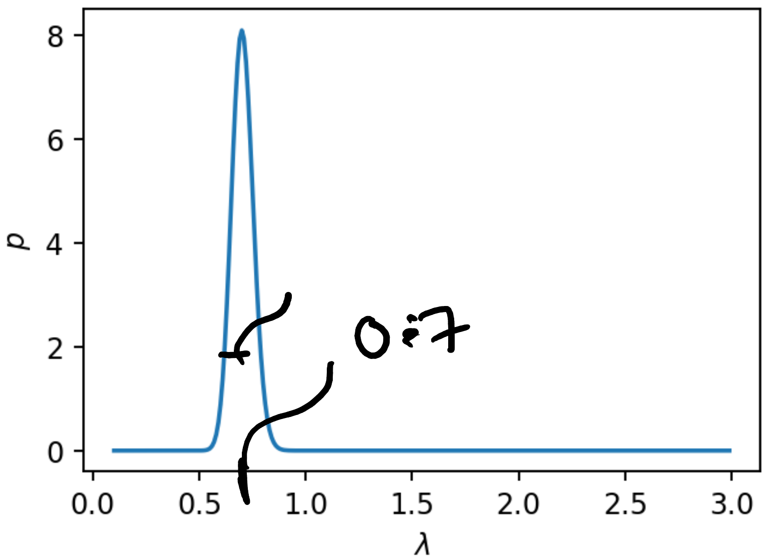
FOUNDATIONS OF DATA SCIENCE

Max likelihood -> Bayesian Inference

Bayes Theorem:

$$\underbrace{P(\vartheta \mid \underline{Y} = y)}_{\text{Posterior}} = \frac{\overbrace{P(\underline{Y} = y \mid \vartheta)}^{\text{Likelihood}} \, \overbrace{p(\vartheta)}^{\text{Prior}}}{\underbrace{P(Y = y)}_{\text{Evidence}}}$$

Horsekick posterior



$$P(Y = y) = \int_{-\infty}^{\infty} P(Y = y \mid \vartheta)\, p(\vartheta)\, d\vartheta$$

Summary

Motivated the probabalistic basis of inference using max likelihood

Important: think of what distribution should describe the data

Links to future courses:
- MLG (derivation of standard ML methods)
- ATML (new in 25-26: cutting-edge machine learning)
- MCI (Causal inference)