Week 4: The Normal Distribution

Recap

So far we have

- learned about maximising the likelihood
- estimated confidence intervals and standard errors

These are the basic tools we will use to fit and understand our models

This Week

- More than one datum: the probability
- Some data: Punxsutawney Phil & Groundhog Day
- The Normal Distribution
  - the log likelihood -Different Amounts of Data
- MLEs
  - The MLE for μ
  - The MLE for σ^2
  - The distribution of $\hat{\mu}$
- Modeling: Predicting the End of Winter, with a t-test

More than one datum

So far we have only used one data point. But what if we have more?

If we make one assumption, the maths is easy

More than one datum: the probability

If data are independent, then

$$Pr(X_1 \& X_2) = Pr(X_1)Pr(X_2)$$

So we can multiply the probabilities

In general, then

$$Pr(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} Pr(X_i)$$

More than one datum: the likelihood

The log-likelihood for the parameters (θ) given the data is

$$l(\theta|X_1, X_2, ..., X_n) = \sum_{i=1}^{n} \log(Pr(X_i|\theta))$$

So we just add the log-likelihoods together

- easier than multiplying!

Punxsutawney Phil & Groundhog Day

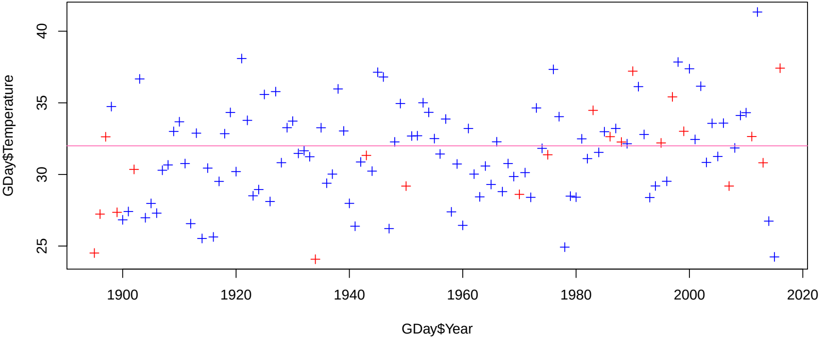
Here is some data on whether Punxsutawney Phil predicts another 6 weeks of winter (which he tries to do every Feb 2)

We will look at the average temperature for Feb/March in Pennsylvania in each year (web link: "https://www.math.ntnu.no/emner/ST2304/2019v/Week4/GroundhogDayta.csv")

```
GDay <- read.csv(file="https://www.math.ntnu.no/emner/ST23
```

Winter/Spring Temperatures

These are the mean February/March tempertures



Initially we want to summrise this distribution

The Normal Distribution

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The Normal Distribution: Exercises

In the module, go through the exercises in *The Normal Distribution: Exercises*

The Normal Distribution: the log likelihood

If we take logs of the density,

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

we get this:

$$l(\mu, \sigma^2|x) = -\frac{1}{2} log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$$

The likelihood for a single data point. For μ this is just a quadratic

The Likelihood for μ

The likelihood for n independent samples is the product of each likelihood:

$$L(\mu, \sigma | x_1, ..., x_n) = p(x_1|\mu, \sigma) p(x_2|\mu, \sigma) ... p(x_n|\mu, \sigma) = \prod_{i=1}^{n} p(x_i|\mu, \sigma)$$

This means that the log-likelihood is the sum of the likelihoods, the sum of quadratic terms:

$$\log L(\mu, \sigma | x_1, ..., x_n) = \sum_{i=1}^{n} l(x_i|\mu, \sigma) = C - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

In practice, we can calculate this with `dnorm(..., log=TRUE)`

Our Task

Estimate the parameters of this distribution

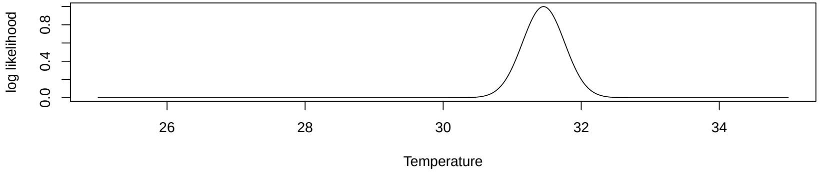
- estimate $\hat{\mu}$ and $\hat{\sigma}^2$

In practice, $\hat{\mu}$ is more important, because we will be modelling μ as a function of different effects

Finding the Estimate

We can simulate data and calculate the likelihood for different values of the mean(we will fix the standard deviation for now)

```
CalcNormLh <- function(mu, sigma, data) {
  lhood <- sum(dnorm(data, mean=mu, sd=sigma, log=TRUE))
  lhood
}
Means <- seq(25, 35, length=500)
sdTemp <- sd(GDay$Temperature)
lhoods <- sapply(Means, CalcNormLh, sigma=sdTemp,
                 data=GDay$Temperature)
plot(Means, exp(lhoods-max(lhoods)), type="l",
     xlab="Temperature", ylab = "log likelihood")
```



The MLE for μ

We could try simulating & finding the best value, or we could try numerically maximising this. But we can get an analytic solution

(this is one reason why the normal distribution is so nice - the maths is relatively easy)

The MLE for μ

We can differentiate the log-likelihood w.r.t μ, and set this to zero

$$0 = \frac{1}{2\sigma^2}\left(2\sum_{i=1}^{n} x_i - 2n\mu\right)$$

Then re-arrange, and the MLE is

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$

The sample mean!

The MLE for σ^2

This is usually less important. We are generally not interested in the standard deviation, but it is a parameter of the distribution, so it has to be estimated. What we do is differentiate w.r.t σ^2, set to zero, re-arrange, and get

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

For details, you can do it yourself or see https://www.statlect.com/fundamentals-of-statistics/normal-distribution-maximum-likelihood

Comments

The estimate $\hat{\mu}$ is just the sample mean, and $\hat{\sigma}^2$ is just the sample variance

- the whole distribution can be summarised by these two statistics

$\hat{\sigma}^2$ has n as a denominator, not $n-1$

- because we assume the MLE for $\hat{\mu}$: using (n-1) is better because it takes into account the uncertainty

Exercises

Confidence Intervals is the place to be for exercises on the MLE for the mean and its confidence interval

The distribution of $\hat{\mu}$

We can look at the distribution of $\hat{\mu}$, and (for example) estimate confidence intervals:

The distribution of $\hat{\mu}$

If the data are normally distributed, the distribution of $\hat{\mu}$ is a t-distribution.

- we can use `dt`, `pt` etc.

The parameters:

- mean
- standard error (= standard deviation/*sqrtn*)
- degrees of freedom ($= n - 1$)

```
dt(32, mean(GDay$Temperature), sd(GDay$Temperature)/sqrt(le
   df=length(GDay$Temperature)-1)
```

```
## [1] 0.1690379
```

If we have enough data, this distribution look like a normal distribution

Your work

Answer the exercises in *The distribution of* $\hat{\mu}$

and then *How the amount of data affects confidence intervals* to look at how the amount of data affects confidence intervals

Next: something useful...

So far we have been learning about statistical inference and statisical programming. Now we can start to use this in modelling.

Next, we will start with a simple model, but put it in the context of maximum likelihood

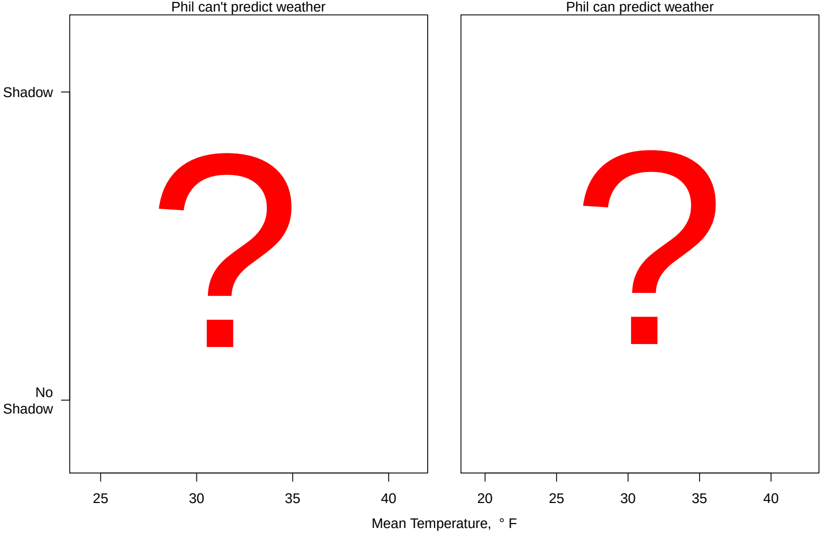
Predicting the End of Winter

If Punxsutawney Phil sees his shadow, there will be 6 more weeks of winter

If he is good at predicting winter, we should see lower average temperatures in the 2 months after the prediction

The Modelling

First, what would we expect if Punxsutawney Phil can predict winter, and if he cannot



The Model

The question is about the mean temperature: is there a clear difference when Phil sees his shadow or not?

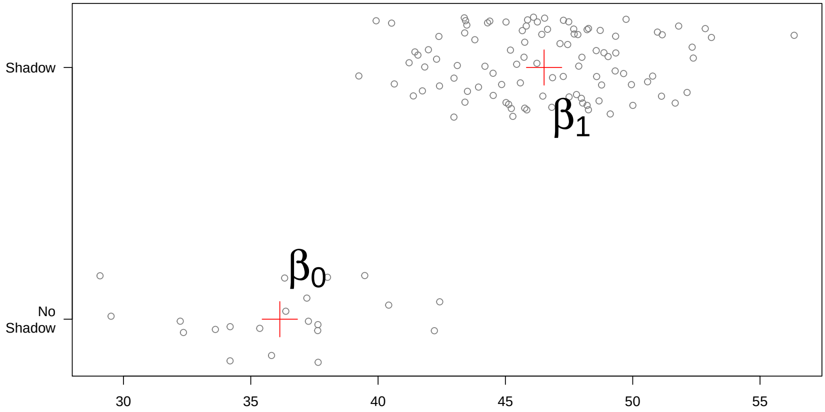
So, we have to build a model where there are different mean temparatures when he does see his shadow, and when he does not.

We can do this in a few ways. For all we assume $y_i \sim N(\mu_i, \sigma^2)$ (i.e. the data follow a normal distribution with the same variance)

We also will define a variable X_i: $X_i = 1$ if Phil saw his shadow (and hence predicted winter),$X_i = 0$ if he did not

The Model - First way

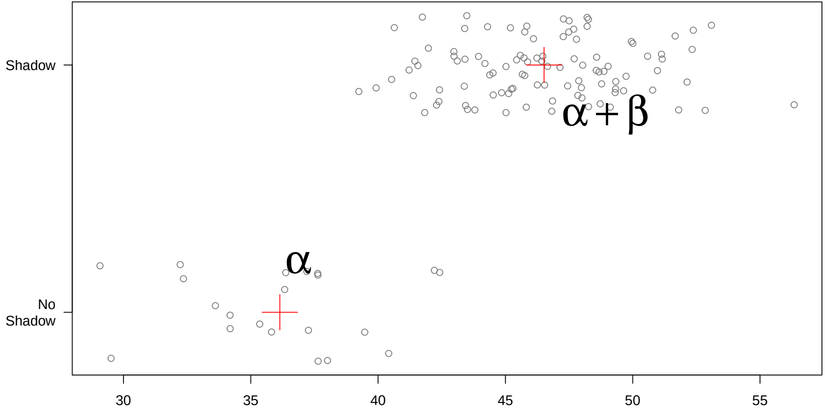
$$\mu_i = \begin{cases} \beta_0 & \text{if } X_i = 0 \\ \beta_1 & \text{if } X_i = 1 \end{cases}$$



The Model - Second way

$$\mu_i = \begin{cases} \alpha & \text{if } X_i = 0 \\ \alpha + \beta & \text{if } X_i = 1 \end{cases}$$

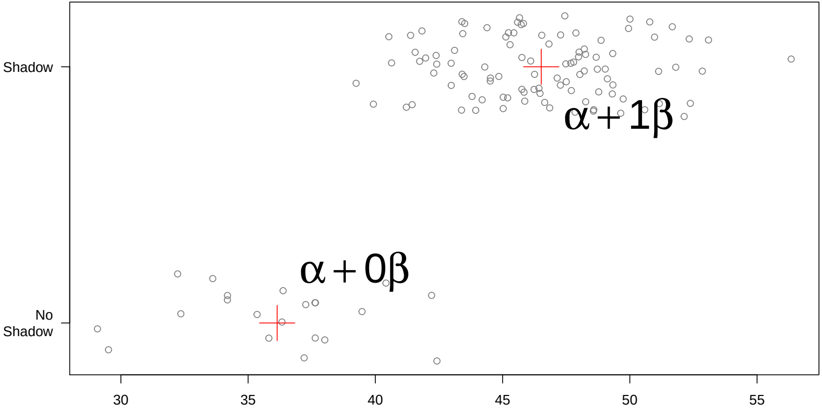
So the difference is β, and this is what we are interested in



The Model - Third way

$$\mu_i = \alpha + \beta X_i$$

The difference is β (because X_i can only be 0 or 1). This approach is the easiest to extend to more complex models



Calculating the Likelihood

The log-likelihood is not too difficult to write down

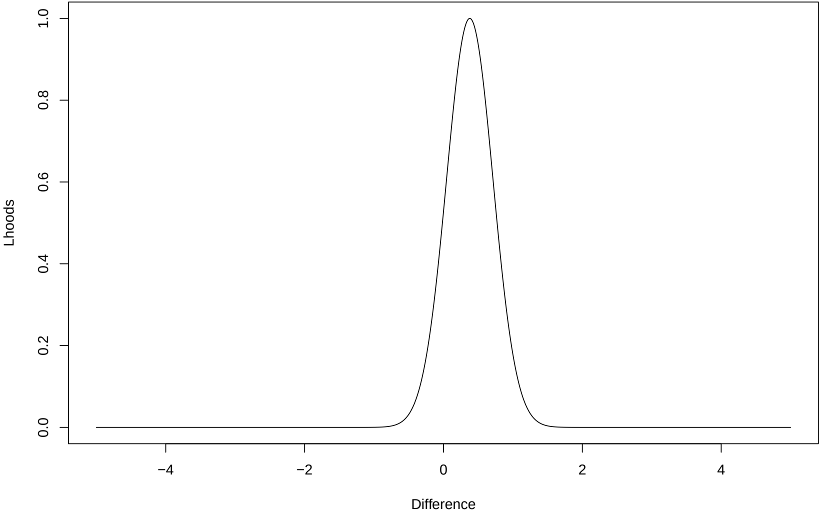
$$\log L(\mu_i, \sigma | x_1, ..., x_n) = C - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu_i)^2$$

where mu_i is described in the previous slide.

We now have 3 parameters (two for the means, and the standard deviation).

Calculate the likelihood for the difference

We can look at the likelihood for the difference, using the MLEs for the other parameters



Exercise

Do the exercise in *Calculating the Likelihood* to estimate the mle for teh difference in temperatures, and the confidence inteval.

Then look at the *In Practice* section on the R functions to do this, and do those exercises

Can Punxsutawney Phil predict winter?