Exact inference for the OR

	Lung cancer		
Smoker	Cases	Controls	Total
Yes	688	650	1338
No	21	59	80
	709	709	1418

- X the number of smokers for the cases
- Y the number of smokers for the controls
- Calculate an exact CI for the odds ratio
- Have to eliminate a nuisance parameter

Notation

- $\text{logit}(p) = \log\{p/(1-p)\}$ is the **log-odds**
- Differences in $\text{logit}s$ are log-odds *ratios*
- $\text{logit}\{P(\text{Smoker} \mid \text{Case})\} = \delta$
  - ▶
- $\text{logit}\{P(\text{Smoker} \mid \text{Control})\} = \delta + \theta$
  - ▶
- θ is the log-odds ratio
- δ is the nuisance parameter

Notation

- X is binomial with n_1 trials and success probability $e^{\delta}/(1+e^{\delta})$
- Y is binomial with n_2 trials and success probability $e^{\delta+\theta}/(1+e^{\delta+\theta})$

$$P(X = x) = \begin{pmatrix} n_1 \\ x \end{pmatrix} e^{x\delta} \left\{ \frac{1}{1 + e^{\delta}} \right\}^{n_1}$$

$$P(Y = z - x) = \begin{pmatrix} n_2 \\ z - x \end{pmatrix} e^{(z-x)\delta + (z-x)\theta} \left\{ \frac{1}{1 + e^{\delta+\theta}} \right\}^{n_2}$$

Non-central hypergeometric distribution

$$P(X = x \mid X + Y = z; \theta) = \frac{\binom{n_1}{x}\binom{n_2}{z-x} e^{x\theta}}{\sum_u \binom{n_1}{u}\binom{n_2}{z-u} e^{u\theta}}$$

- θ is the log odds ratio
- This distribution is used to calculate exact hypothesis tests for $H_0 : \theta = \theta_0$
- Inverting exact tests yields exact confidence intervals for the odds ratio
- Simplifies to the hypergeometric distribution for $\theta = 0$

In R

```
x <- matrix(c(688, 650, 21, 59), 2, byrow = TRUE)
fisher.test(x)
```

```
Result
95 percent confidence interval:
 1.755611 5.210711
```

CI is for the odds ratio, e^{θ} (not the log odds ratio, θ)

Conditional likelihood

Notice that

$$P(X = x \mid X + Y = z; \theta) = \frac{\binom{n_1}{x}\binom{n_2}{z-x} e^{x\theta}}{\sum_u \binom{n_1}{u}\binom{n_2}{z-u} e^{u\theta}}$$

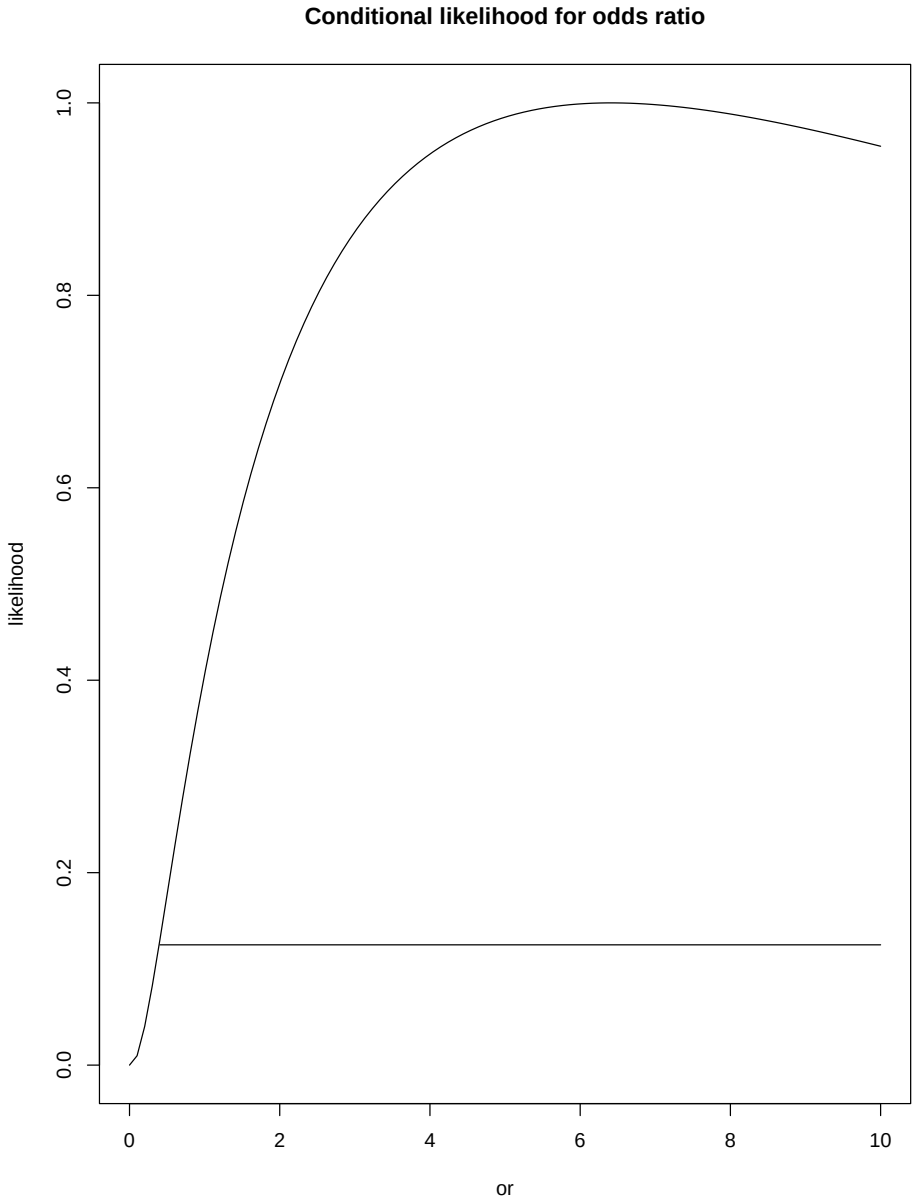
is a conditional likelihood that depends only on θ

The function `ORlikelihood` on the course web site works for smaller data sets

Some R code

```
#Fisher's data
dat <- ORlikelihood(matrix(c(3, 1, 1, 3), 2),
                    seq(0,10, length = 100))
plot(dat$or, dat$like,
     type = "l",
     xlab = "or",
     ylab = "likelihood")
lines(range(dat$or[dat$like > 1/8]), c(1/8,1/8))
title("Conditional likelihood for odds ratio")
```

The function `ORlikelihood` is on the course web site


Conditional likelihood for odds ratio
likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
or