Report on stay abroad at La Trobe University, Melbourne, Australia

a review of the course content

by Mikkel Meyer Andersen

January 8, 2010

$$\mathcal{K}_{\theta_0}^{-1}\left(\frac{T_{1:K} \pm z_{1-\frac{\alpha}{2}}}{\sqrt{N}}\right)$$

Advisor: Poul Svante Eriksen

Department of Mathematical Sciences, Aalborg University

Department of Mathematical Sciences
Fredrik Bajers Vej 7G
9220 Aalborg st
Denmark
Phone: +45 99 40 88 04
http://www.math.aau.dk

Title: A review of the course content

Semester: MAT5, from July 2009 to November 2009

Study abroad university: La Trobe University, Melbourne, Australia

Advisor: Poul Svante Eriksen

Circulation: 4
Pages: 65
Finished: January 8, 2010

Summary: This report contains a description of what material three Semester 2 courses held at La Trobe University, Melbourne, Australia have covered. The courses in question are STA4SI (Statistical Inference), STA4AMD (Analysis of Medical Data), and STA3AS (Applied Statistics). Among the topics that are described in depth are meta analysis and how to use variance stabilizing methods to get better meta analysis because of quicker convergence for non-normal data.

Mikkel Meyer Andersen, January 8th, 2010

Contents

Chapter 1

Courses

I've been enrolled in the following courses at La Trobe University, Melbourne, Australia (the descriptions is taken from [University, 2009a]):

STA4SI – Statistical Inference (taught by Paul Kabaila) This unit covers a selection of topics in classical statistical inference at the fourth year level. It consists of a selection of material from the following chapters of [Casella and Berger, 2002]: Chapter 6 (Principles of Data Reduction), Chapter 7 (Point Estimation), Chapter 8 (Hypothesis Testing), Chapter 9 (Interval Estimation) and Chapter 10 (Asymptotic Evaluations). A knowledge of this material is helpful in almost any statistical endeavour. References: Lecture material based on [Casella and Berger, 2002].

STA4AMD – Analysis of Medical Data (taught by Robert Staudte) This unit considers calibrating evidence in a test, variance stabilizing transformations, one- and two-sample Binomial models, evaluating and comparing Poisson rates, evidence in one and two-sample Welch t-tests and compensating for publication bias. The evidence obtained by variance stabilization will be the basis for confidence intervals for effects, which are demonstrably more accurate than those obtained by traditional large-sample methods. Further, variance stabilizations facilitates a meta-analysis of results from different studies. References: [Kulinskaya et al., 2008].

STA3AS – Applied Statistics (taught by Luke Prendergast and Andriy Olenko) This unit provides advanced-level introductions to the topics of sample surveys, multivariate analysis and time series analysis. These topics are very important in applied statistics. The unit also includes an introduction to statistical consulting. On successful completion of this unit, the student should have: 1) An understanding of the subtle difficulties encountered when analysing data sampled using simple random sampling. 2) A theoretical understanding of some common, yet powerful, statistical methods for the analysis of multivariate data. 3) An understanding of formulating, estimating and interpreting various linear time series models for empirical studies. 4) An understanding of conducting basic statistical inquiries with meaningful interpretation. Refer to [University, 2009b] for a unit guide. References: [Prendergast and Kabaila, 2009] (lecture notes based on [Rice, 2006], [Johnson and Wichern, 2001], and [Box et al., 1994]).

1.1 STA3AS – Applied Statistics

1.1.1 Studied topics

- Sample surveys
  - Simple random sampling
  - Stratified random sampling
  - Overview of other survey sampling schemes and potential improper uses (systematic, cluster, and unrepresentative sampling)
- The multivariate normal distribution
  - The connection with the χ^2_p-distribution
  - Solid ellipsoids and the squarred statistical distance $(\boldsymbol{X}-\boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{X}-\boldsymbol{\mu})$ of $\boldsymbol{X}$ from $\boldsymbol{\mu}$
  - Estimation of the mean vector and covariance matrix
  - Checking for multivariate normality (through univariate marginals, bivariate maginals etc.)
  - Testing the normal population mean (and how Hotelling's T^2-test is inferred from the t-test in the univariate case)
  - Simultaneous confidence statements: T^2 intervals inferred from the Maximization lemma in [Johnson and Wichern, 2001] stating that: For a positive definite $p \times p$ matrix B and $\boldsymbol{d}$ a given p-dimensional vector, let $f(\boldsymbol{x}) = \frac{\left(\boldsymbol{x}^\top \boldsymbol{d}\right)^2}{\boldsymbol{x}^\top B \boldsymbol{x}}$. Then it's true that $\max_{\boldsymbol{x}\neq\boldsymbol{0}} f(\boldsymbol{x}) = \boldsymbol{d}^\top B^{-1} \boldsymbol{d}$. It's also true that for $c \neq 0$ then $\boldsymbol{x} = cB^{-1}\boldsymbol{d}$ maximizes $f(x)$.
- Principal component analysis
  - Population principal components
  - Sample principal components
- Classification (only for two populations)
  - Known pdfs for the populations
    - ECM – Expected cost of misclassification
    - TPM – Total probability of misclassification (equivalent to equaling the costs of misclassification)
  - Multivariate normal distributed populations
    - Explicit ECM for both equal and unequal covariance matrices
  - Fisher's discriminant function for equal covariance matrices
  - Problems with classification rules, e.g. (estimation of) covariance matrix can be singular if the dimension of data equals or exceeds the number of observations
  - Use of principal component analysis together with classification in order to avoid singular (estimation of) covariance matrix
- Time series
  - First and second order autoregressive process, i.e. $AR(1)$ and $AR(2)$

  - Introduction to processes of arbitrary order: $AR(p)$, $MA(q)$ (moving average), $ARMA(p,q)$ (autoregressive moving average process), and $ARIMA(p,d,q)$ (autoregressive integrated moving average process), e.g. finding autocovariance functions for given processes and 1-, 2-, and 3-step forecasting

- One and a half weeks with statistical consulting (a very little theory on what is important when meeting a client, how to start, pitfalls etc., but mainly learning by doing though one mini-project/assignment)

Some general mathematical topics used and slightly studied (no proofs):

- The δ-method
- Lagranges multipliers
- Simple operation theory (used for the the delay operator in time series)

1.1.2 Assignments

One weekly assignment making 12 in total. Those were divided on 11 theoretical assignments (7 in sample surveys and multivariate analysis in addition to 4 in time series) and 1 in statistical consultancy (practical with another student so that we each others clients and statistician).

1.2 STA4SI – Statistical Inference

A continuation of STA3SI, which should be somewhat equivalent to the MAT4-course "Statistical Theory and Method" at AAU, but some overlap occured.

1.2.1 Studied topics

- Data reduction
  - By sufficiency (a bit of repitition of "Statistical Theory and Method")
  - By an ancillary statistic (a statistic that doesn't depend on the unknown parameter θ, whereas we in general want statistics to depend on θ as much as possible to make inference from it)
    * Motivated with Cox's example of mixtured normal distributions
    * Useful practical examples: linear regression (how conditioning on observations makes inference easier to deal with as opposed to treating observations as random variables) and how conditioning on the column sum in a 2×2 contingency table is "a kind of ancillary statistic" (rumor has it that this was Fisher's words) and can make inference easier
    * How the order of data reduction matters (through Cox's example of mixtured normal distributions)
  - By equivariance

- Exact and approximate confindence intervals and their properties such as coverage and risk functions
- The effect of preliminary model selection on confidence intervals studied through [Freeman, 1989] (”The performance of the two-stage analysis of two-treatment, two-period crossover trials“)
  - The confidence coefficient of the confidence interval after a preliminary test is horrible (an example is made for certain choices of parameters such as level of significans and the confidence coefficient is below 0.5)
  - Preliminary testing in this certain case with analysis of two-treatment, two-period crossover trials gives false security – the recommendation is: don't do it!
- Estimators
  - Maximum likelihood estimation and invariance property of MLEs (if $\hat{\theta}$ is a MLE of θ then, for any function $\tau(\theta)$, an MLE of $\tau(\theta)$ is $\tau\left(\hat{\theta}\right)$)
  - Methods of evaluating estimators: mean squarred error and risk functions (with a generic loss function measuring the distance between the estimator and the parameter of interest)
  - The Cramér-Rao Inequality
- Hypothesis testing basics
- The Neyman-Pearson Lemma
- Intersection-Union Tests
- The Probability Integral Transformation Theorem
- Confidence sets obtained by inverting a family of hypothesis tests
- Large-sample results and a asymptotic theory
  - Convergence in probability and distribution (and why the first implies the latter)
  - Slutsky's Theorem (without proof)
  - The δ-method / the delta-method and the flaws in the simple formulation (refer to the footnote in section 3.2.2)
  - Consistent estimators (and sufficient conditions)
  - Asymptotic variance, limiting variance and the inequality for unbiased estimators (when the limiting variance exist)
  - The optimality of MLE's in regards to asymptotic variance under certain regularity conditions

1.2.2 Assignments

Seven assignments in total.

1.3 STA4AMD – Analysis of Medical Data

1.3.1 Studied topics

- Asymptotic results (no proofs)
- Different approximate confidence intervals for the binomial distribution (among with the Agresti and Coull-correction)
- Traditional meta analysis (combining evidence from K studies)
  - Unequal fixed effects model (UFEM)
  - Fixed effects model (FEM)
  - Random effects model (REM)
  - How to choose a model e.g., Cochran's Q
- Calibration of evidence – refer to section 3.2 for a quick overview of some of the theory
  - What is evidence and what it is in regards to p-values
  - Motivating examples (taken from course material):
    - How much evidence is there in a p-value of 0.01, say, relative to 0.05? (refer to Table 3.1)
    - How small must a p-value be to represent twice as much evidence against the null hypothesis as 0.05?
    - The random p-value is a monotone function of a test statistic and hence contains the same evidence for the alternative. This is easily measured by transforming it to the normal location family, which serves as a convenient calibration scale.
- Vst's (variance stabilization transformations), why (e.g. to get known inverse variance weights in meta analysis and as a bonus often quick convergence to normal in distribution), and how to derive them (simple approach using the δ-method) – refer to section 3.2.2 for details
- Evidence in different statistics (t-statistics, simple Binomial and Poisson models, matched binomial pairs and some others)
- "New" meta analysis based on vst's (refer to section 3.2 for a quick overview of some of the theory)
  - Fixed effects model (FEM) (and an example with matched binomial pairs)
  - Random transformed effects model (and an example with matched binomial pairs)
- Confidence intervals for the risk difference of two binomials
  - The confidence interval in [Newcombe, 1998] (using Wilson's score intervals) for the risk difference
  - A vst for the risk difference
  - A quick comparing between these and the new exact confidence interval in [Chan and Zhang, 1999]

- Evaluating Poisson rates: transformed effect, evidence, and an example using the L. J. Bortkewitsch's horse-kicks data set
- Comparing Poisson rates: transformed effect and evidence for unconditional approach (inference of the rate difference) and conditional approach (inference of the rate ratio)
- Simple linear meta-regression

1.3.2 Assignments

Four assignments in total.

Chapter 2

Statistical inference

This chapter is where some of the material from the courses has been described in details.

2.1 CLT – central limit theorem

A theorem used throughout all the courses is the Central Limit Theorem (CLT). For i.i.d. random variables $X_1, \ldots, X_n$ with mean μ and variance $\sigma^2 > 0$ define

$$S_n := \sum_{i=1}^{n} X_i \quad \text{and} \quad \overline{X}_n = \frac{S_n}{n}.$$

Then CLT states that

$$Z_n \xrightarrow{D} N(0,1) \quad \text{and} \quad n \to \infty \quad \text{where} \quad Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

The theorem can be shown to be true for weaker requirements than i.i.d. random variables.

It's worth mentioning that this is a asymptotical result and in practice the convergence can be quite poor.

2.2 Confidence intervals

Confidence intervals are very important. They contain a lot more information than e.g. p-values. Confidence intervals can roughly be devided into two parts: exact and approximate confidence intervals. In order to make a proper distinction, a few defitions have be be made.

Let $\Theta \subseteq \mathbb{R}^d$ be the parameter space and $\theta \in \Theta$ the unknown parameter. Suppose that $\tau(\theta) \in \mathbb{R}$ is a scalar parameter of interest. Then a confidence interval for $\tau(\theta)$ is any pair of real-valued functions $L(X)$ and $U(X)$ satisfying $L(x) \leq U(x)$ for all data points x and such that if $X = x$ is observed, then the inference $L(x) \leq \tau(\theta) \leq U(x)$ is made.

The following two definitions are taken directly from the lecture material of STA4SI.

Definition 2.1 (Coverage probability). For the confidence interval $[L(X), U(X)]$ for $\tau(\theta)$, the coverage probability is the probability that the random interval $[L(X), U(X)]$ covers $\tau(\theta)$. In other words, the coverage probability is

$$P_\theta\left(\tau(\theta) \in [L(X), U(X)]\right). \qquad \square$$

Definition 2.2 (Confidence coefficient). For the confidence interval $[L(X), U(X)]$ for $\tau(\theta)$, the confidence coefficient is defined to be

$$\inf_{\theta \in \Theta} P_\theta\left(\tau(\theta) \in [L(X), U(X)]\right).$$

A confidence interval with confidence coefficient equal to some value, say $1-\alpha$, is called a $1-\alpha$ confidence interval. $\square$

With these definitions in mind, it possible to make a proper distinction between exact and approximate confidence intervals. Exact confidence intervals usually rely on the actually distribution of the data, and have a garuanteed confidence coefficient. On the other hand, approximate confidence intervals usually rely on approximate results (e.g. asymptotic normality, variance stabilization transformations or similar) and thus usually have a poor confidence coefficient. This means that for approximate confidence intervals the coverage probability is only as promised for a sample size $n \to \infty$, which practically is not true no matter how large n is. This means that in certain cases it's possible to end up with a quite poor coverage probability without knowing it.

In this section, I'll take a closer look at the mentioned topics and give some examples of how bad things can be.

2.2.1 Confidence intervals for the binomial distribution

Binomial random variables has a great practical importance. Because of this a lot of reseach in binomal confidence intervals has been performed. First I'll review some simple approximate confidence intervals, and afterwards I'll discuss exact confidence intervals.

Approximate confidence intervals

This section is based on the lecture material from STA4AMD.

If $X \sim \text{Binomial}(n, p)$, and $\hat{p} = \frac{X}{n}$, then for fixed p, CLT states that

$$\frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1) \quad \text{for } n \to \infty.$$

Hence a $(1-\alpha)100\%$ approximate confidence interval for p is

$$I(X) = \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}.$$

This is also called the Wald interval. Because of the CLT is asymptotical result, sufficiently big n is required. Because several problems arise near the boundary

of the parameter space, samples some people requires that $np(1-p) \geq 5$ for the approximate confidence interval to be sufficiently reliable.

The Wald interval is really poor, and one should avoid using it. To see why, define

$$A(p) = \{x \in \{0, 1, \ldots, n\} : p \in I(x)\}.$$

Then

$$\begin{aligned}
P_p\left(p \in I(X)\right) &= \sum_{x \in \mathcal{X}} P_p\left(p \in I(X) \mid X = x\right) P_p\left(X = x\right) \quad \text{(by the Law of Total Probability)} \\
&= \sum_{x \in A(p)} P_p\left(p \in I(X) \mid X = x\right) P_p\left(X = x\right) + \\
&\quad \sum_{x \in (A(p))^c} P_p\left(p \in I(X) \mid X = x\right) P_p\left(X = x\right) \\
&\quad \text{(because } \mathcal{X} = A(p) \cup (A(p))^c \text{ and trivially } A(p) \cap (A(p))^c = \varnothing\text{)} \\
&= \sum_{x \in A(p)} P_p\left(p \in I(X) \mid X = x\right) P_p\left(X = x\right) + \\
&\quad \sum_{x \notin A(p)} P_p\left(p \in I(X) \mid X = x\right) P_p\left(X = x\right) \\
&= \sum_{x \in A(p)} 1 \cdot P_p\left(X = x\right) + \sum_{x \notin A(p)} 0 \cdot P_p\left(X = x\right) \\
&= \sum_{x \in A(p)} P_p\left(X = x\right)
\end{aligned}$$

because $P_p\left(p \in I(X)\right) = 1$ when $X \in A(p)$ and $P_p\left(p \in I(X)\right) = 0$ when $X \notin A(p)$ by definition of $A(p)$. Hence $P_p\left(p \in I(X)\right)$ as a function of p is a discontinuous function. A terrifying characteristic about the Wald interval is, that for no matter what n is, we can select a $p \in (0, 1)$ such that the coverage probability comes arbitrary close to 0. This fact will now be proven.

We have to prove that for a given n, we can select $p \in (0, 1)$ and $\varepsilon > 0$ such that

$$P_p\left(p \in I(X)\right) < \varepsilon.$$

Let n we given. Then we have that

$$\begin{aligned}
P_p\left(p \in I(X)\right) &= \sum_{x \in A(p)} P_p\left(X = x\right) \\
&\leq \sum_{x \in \{1,2,\ldots,n-1\}} P_p\left(X = x\right) \quad \text{(because } A(p) \subseteq \{1, 2, \ldots, n-1\}\text{)} \\
&= 1 - \sum_{x \in \{0,n\}} P_p\left(X = x\right) \\
&= 1 - \left(\binom{n}{0} p^0 \left(1-p\right)^{n-0} + \binom{n}{n} p^n \left(1-p\right)^{n-n}\right) \\
&= 1 - \left(\left(1-p\right)^n + p^n\right) \quad \text{(because } \tbinom{n}{0} = \tbinom{n}{n} = 1\text{)} \\
&= 1 - \left(1-p\right)^n - p^n.
\end{aligned}$$

So for given n, we can choose ε and p such that

$$P_p\left(p \in I(X)\right) \leq 1 - \left(1-p\right)^n - p^n < \varepsilon,$$

i.e. when we have been given n and chosen an ε, we have to solve the inequality

$$1 - \left(1-p\right)^n - p^n < \varepsilon$$

for p.

As an example, consider $n = 10$. Now say that we want a coverage probability below $1 - \alpha = 0.5$, i.e. very bad. Then we have the inequality

$$1 - (1-p)^{10} - p^{10} < 0.5$$

which we – by the use of Maple – find is true for

$$p < 0.066 \quad \text{or} \quad p > 0.934.$$

(The true results are $p < 0.06696700846\ldots$ and $p > 0.9330329915\ldots$, but instead of traditional rounding, the decimals are chosen to maintain the inequalities.)

To conclude the example, when $n = 10$ then by choosing $p \in (0\ ;\ 0.066) \cup (0.934\ ;\ 1)$ we get a coverage probability below 0.5.

Notice that the limits are symmetric, i.e. $0.066 = 1 - 0.934$, this is – by construction – generally true as well.

With a coverage probability below $1 - \alpha = 0.01$ we get

$$p < 0.001 \quad \text{or} \quad p > 0.999.$$

Another example is with $n = 783$ and a coverage probability below $1 - \alpha = 0.5$. Then the inequality is not solve symbolically, but instead one can use numerical solving (in this case Newton's Method). We then get

$$p < 0.00089 \quad \text{or} \quad p > 0.99911.$$

As demonstrated, larger n and smaller ε leads to p closer to 0 and 1.

As mentioned and showed, the convergence is especially poor near the boundary of the parameter space i.e., for p close to 0 and 1. In these cases it's highly likely that $x = 0$ or $x = 1$, such that $\hat{p} = 0$ or $\hat{p} = 1$, respectively. In these cases the Wald interval becomes a point estimate, namely $\hat{p}$. Because of this another estimate given by

$$\tilde{p} = \frac{X + c}{n + 2c} \quad \text{for} \quad c \in \mathbb{R}^+,$$

has been proposed. It can be interpreted as adding c heads and c tails such we move away from the boundary and the problematic degenerated approximate confidence interval. This estimate, $\tilde{p}$, is a generalization of the one Agresti and Coull proposed with $c = \frac{1}{2}$.

Confidence interval through use of a vst

Another way of constructing a confidence interval is to use a vst to obtain a approximate normal distributed statistic. With the same notation as earlier, let $X \sim \text{Binomial}(n, p)$ and $S_n = \frac{X}{n}$. Then $\mathbf{E}[S_n] = p$ and $\mathbf{Var}[S_n] = \frac{p(1-p)}{n}$. Thus $\mathbf{Var}[S_n] = g(\mathbf{E}[S_n])$ for $g(t) = \frac{t(1-t)}{n}$. Hence a simple approximate vst is

$$h(x) = \int^x (g(t))^{-\frac{1}{2}}\, dt = \int^x \left(\frac{n}{t(1-t)}\right)^{\frac{1}{2}} dt = 2\sqrt{n}\sin^{-1}\left(\sqrt{x}\right)$$

by the δ-method (refer to section 3.2.2 for details). Now $h(S_n)$ is approximately normal, hence the usual confidence interval can be used. Then by finding the inverse of the vst, an approximate confidence interval for p can be found. In [Kulinskaya et al., 2008, chapter 17] further details can be found, e.g. on how further refinements can improve the coverage of the interval.

Final remarks

For a comparison of different binomal confidence intervals, please refer to [Brown et al., 2001].

Because of the poor properties (typically poor coverage probability) of the approximate confidence intervals, it's desireable to find and use exact confidence intervals based on the exact distribution of X and not just a convergence in distribution to normality as n tends infinity. This is because these kinds of intervals are guaranteed to have confidence coefficient equal to a pre-specified value $1-\alpha$.

One way of finding exact confidence intervals is analysed in [Kabaila, 2005] and [Kabaila, 2008], but none of these are going to be analyzed in this project.

2.2.2 Risk functions for confidence intervals

There's several ways to measure the quality of a confidence interval. Besides the coverage, one can evaluate the so-called risk function.

Definition 2.3 (Risk function). Let X be a random variable whose distribution depends on the parameters $\theta \in \mathbb{R}^n$. Let $[L(X), U(X)]$ be a confidence interval for a scalar of interest $\tau(\theta) \in \mathbb{R}$. Then the risk function of this confidence interval is

$$R(\theta) = \mathbf{E}_\theta \left[U(X) - L(X) \right]. \qquad \square$$

That is the expected value of the length of the confidence interval.

Two $1-\alpha$ confidence interval can be compared by comparing their risk functions.

Example 2.4. This example has its starting point in a superficial example in the lecture notes to STA4SI. Suppose that $X \sim N(\theta, 1)$. The usual confidence interval is

$$I(x) = [x - z_{1-\frac{\alpha}{2}} \ ; \ x + z_{1-\frac{\alpha}{2}}].$$

The risk function for this interval is

$$R(\theta) = \mathbf{E}_\theta \left[U(X) - L(X) \right] = \mathbf{E}_\theta \left[\left(x + z_{1-\frac{\alpha}{2}} \right) - \left(x - z_{1-\frac{\alpha}{2}} \right) \right] = \mathbf{E}_\theta \left[2z_{1-\frac{\alpha}{2}} \right] = 2z_{1-\frac{\alpha}{2}}$$

which doesn't depend on θ. A different $1-\alpha$ confidence interval minimizing the expected length when $\theta = 0$ is proposed in [Pratt, 1961] as

$$\tilde{I}(\theta) = \begin{cases} [x - z_{1-\alpha} \ ; \ 0] & \text{if } x \leq -z_{1-\alpha} \\ [x - z_{1-\alpha} \ ; \ x + z_{1-\alpha}] & \text{if } -z_{1-\alpha} < x < z_{1-\alpha} \\ [0 \ ; \ x + z_{1-\alpha}] & \text{if } z_{1-\alpha} \leq x \end{cases}$$

It can be shown that both intervals are a $1-\alpha$ confidence intervals, i.e.

$$P_\theta \left(\theta \in I(X) \right) = P_\theta \left(\theta \in \tilde{I}(X) \right) = 1 - \alpha \quad \text{for all } \theta.$$

Denote the risk function for $\tilde{I}(\theta)$ as $\tilde{R}(\theta)$ defined by

$$\tilde{R}(\theta) = \mathbf{E}_\theta \left[\tilde{U}(X) - \tilde{L}(X) \right].$$

To compare the widths of the interval, define

$$f(\theta) = \frac{\tilde{R}(\theta)}{R(\theta)},$$

where we then are interested in extrema for $f(\theta)$. Actually $f(\theta)$ has one extremum, i.e. a minimum at $\theta = 0$ (this can be verified with various method). Now two different approaches will be demonstrated to find $f(0)$:

1. Simulate it with `R` and
2. use Maple to perform the calculations.

It's also possible to find a closed form expression by using the fact that

$$\frac{d}{dx}\varphi(x) = -x\varphi(x)$$

where $\varphi(x)$ is the pdf for the standard normal distribution.

Instead of doing that, I'll first simulate $f(\theta)$ with the following `R`-code:

```
Mdefault = 1000
alpha <- 0.05
z.a <- qnorm(1-alpha)
R <- 2*qnorm(1-alpha/2)

Rtilde <- function(theta, M=Mdefault)
{
  x <- rnorm(M, mean=theta, sd=1)
  l <- rep(0, M)
  u <- rep(0, M)

  for (i in 1:M)
  {
    if (x[i] <= -z.a)
    {
      l[i] <- x[i] - z.a
      next
    }

    if (z.a <= x[i])
    {
      u[i] <- x[i] + z.a
      next
    }

    l[i] <- x[i] - z.a
    u[i] <- x[i] + z.a
  }

  R.tilde <- mean(u-l)
}

f <- function(theta, M=Mdefault)
{
  Rtilde(theta, M) / R
}

f0 <- 0
thetas <- seq(-5, 5, 0.01)
```

```
thetas.len <- length(thetas)
ratio <- numeric(thetas.len)
for (i in 1:thetas.len)
{
  if (thetas[i] == 0)
  {
    f0 <- f(0)
    ratio[i] <- f0
  }

  else
  {
    ratio[i] <- f(thetas[i])
  }
}

plot(thetas, ratio, type="l", main="Relative risk functions",
  sub=paste("alpha = 0.05 and R.tilde(0)/R(0) =~", round(f0, 4)),
  xlab="theta", ylab="R.tilde(theta)/R(theta)")
abline(1, 0)
abline(f(0), 0, lty=3)

cat("Plot based on", Mdefault, "simulations.\n")
cat("f(0) =~", round(f0, 4), " based on", Mdefault, "simulations\n")
for (i in c(10, 100, 1000))
  cat("f(0) =~", round(f(0, Mdefault*i), 4), " based on", Mdefault*i, "simulations\n")
```

Producing figure 2.1 and this output (=~ means approximately):

```
Plot based on 1000 simulations.
f(0) =~ 0.8501  based on 1000 simulations
f(0) =~ 0.8495  based on 10000 simulations
f(0) =~ 0.85  based on 1e+05 simulations
f(0) =~ 0.8499  based on 1e+06 simulations
```

Remember that $\theta = 0$ is the one and only minimum of $f(\theta)$. With this result in mind, it's quite obvious that if θ is near 0, then $\tilde{I}(\theta)$ is a better confidence interval, but $f(\theta) \to \infty$ as $|\theta| \to \infty$.

Instead of relying on simulation, we can also just use Maple to evaluate $f(\theta)$ because

$$\begin{aligned}
\tilde{R}(\theta) &= \mathbf{E}_\theta\left[\tilde{U}(X)\right] - \mathbf{E}_\theta\left[\tilde{L}(X)\right] \\
&= \int_{-\infty}^{\infty} \tilde{U}(x)g(x)dx - \int_{-\infty}^{\infty} \tilde{L}(x)g(x)dx \\
&= \left(\int_{-\infty}^{-z_{1-\alpha}} 0 \times g(x)dx + \int_{-z_{1-\alpha}}^{z_{1-\alpha}} (x + z_{1-\alpha})\, g(x)dx + \int_{z_{1-\alpha}}^{\infty} (x + z_{1-\alpha})\, g(x)dx\right) - \\
&\quad \left(\int_{-\infty}^{-z_{1-\alpha}} (x - z_{1-\alpha})\, g(x)dx + \int_{-z_{1-\alpha}}^{z_{1-\alpha}} (x - z_{1-\alpha})\, g(x)dx + \int_{z_{1-\alpha}}^{\infty} 0 \times g(x)dx\right) \\
&= \int_{-z_{1-\alpha}}^{\infty} (x + z_{1-\alpha})\, g(x)dx - \int_{-\infty}^{z_{1-\alpha}} (x - z_{1-\alpha})\, g(x)dx
\end{aligned}$$

where $g(x)$ is the pdf for X. Then the following Maple-script produces figure 2.2 for $\alpha = 0.05$:

```
z := 1.644853626951472;
g :=
  proc (x) options operator, arrow;
    exp(-(1/2)*(x-t)^2)/sqrt(2*Pi)
  end proc;
Rtilde :=
  proc (t) options operator, arrow;
    int((x+z)*g(x), x = -z .. infinity)-(int((x-z)*g(x), x = -infinity .. z))
```

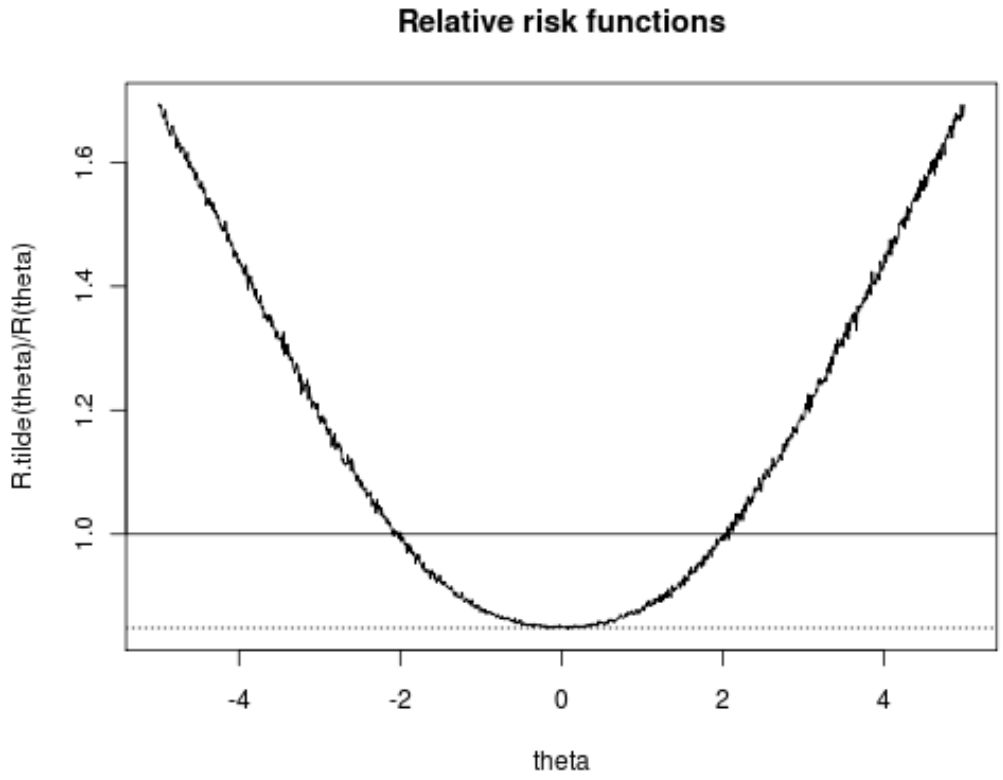

Figure 2.1: Relative risk functions $f(\theta) = \frac{\tilde{R}(\theta)}{R(\theta)}$ for $\alpha = 0.05$ and 1000 simulations. The exact value for $\theta = 0$ is $f(0) = 0.8499$.

```
  end proc;
plot([Rtilde(t)/(2*1.959963984540054), 1,
  eval(Rtilde(t)/(2*1.959963984540054), t = 0)], t = -5 .. 5)
```

Notice that the numerical values have been found with the following R-script:

```
options(digits=22)
qnorm(1-0.05)
[1] 1.644853626951472
qnorm(1-0.05/2)
[1] 1.959963984540054
```

It's also possible to calculate the exact minimum with this Maple-code:

```
eval(Rtilde(t)/(2*1.959963984540054), t = 0)
```

which gives

$$f(0) = 0.8498863239.$$

Notice how the methods yield the same result (if the number of simulated random variables is high enough)[1].

[1]The lecture notes from STA4SI in week 3 suggests that the minimum is 0.8485, but this result is invalid. This number appears because [Pratt, 1961, Table 1] states that the $(f(0))^2 = 0.72$, and the lecturer has then apparently just used it such that $f(0) = \sqrt{0.72} \approx 0.8485$, but this is clearly incorrect. Please notice that $0.8498863239^2 \approx 0.72$ so that results indeed is in coherence with the original article.

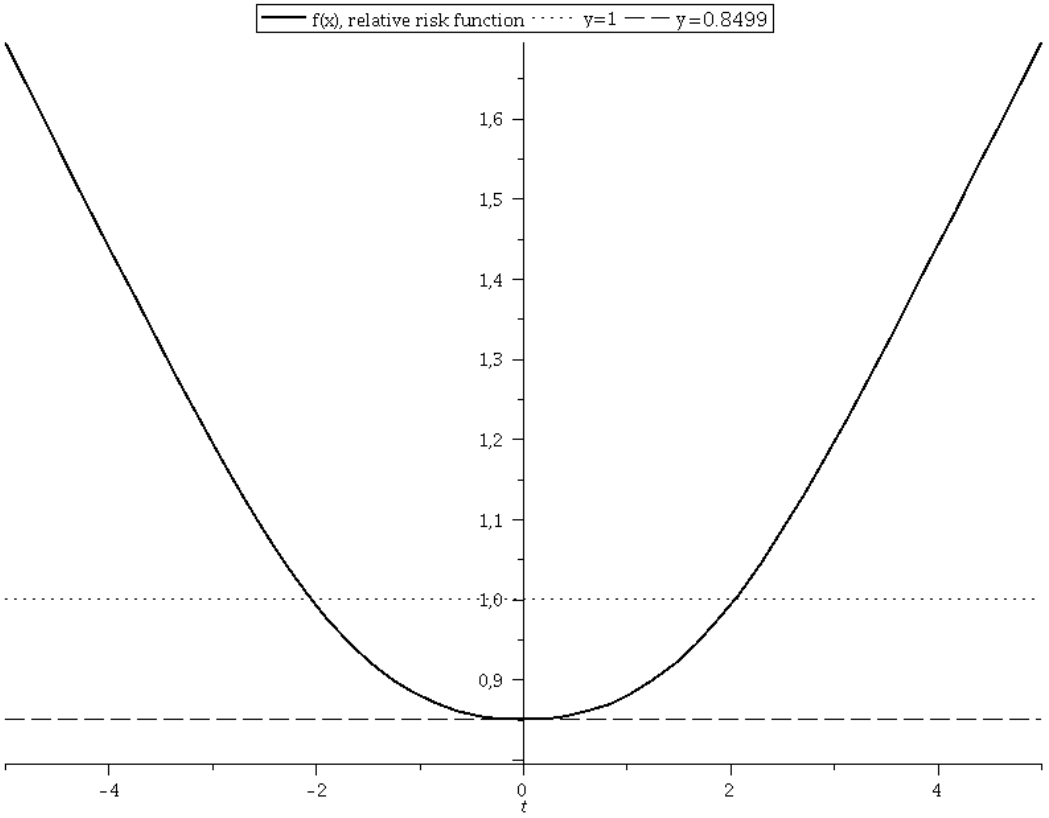

Figure 2.2: Relative risk functions $f(\theta) = \frac{\tilde{R}(\theta)}{R(\theta)}$ for $\alpha = 0.05$ made with Maple. The exact value for $\theta = 0$ is $f(0) = 0.8499$.

[Farchione and Kabaila, 2008] goes futher and describes how to use prior knowledge, and limiting the error for θ far away from 0. □

2.3 Equivariance

Equivariance is a principle stating that a statician's inference for given data should not change when the data is expressed in a different but equivalent form.

This principle will now be illustrated with inference about a binomial distributed random variable.

Example 2.5 (Binomial Equivariance). This example is based on [Casella and Berger, 2002, Example 6.4.1]. Suppose that

$$X \sim \text{Binomial}\,(n, \theta)$$

and that we have a specific prodecure for obtaining an estimate $T(x)$ of θ for an observed $X = x$.

Now imagine that x and the specific prodecure for obtaining an estimate $T(x)$ of θ is provided to two scientists. Scientist 1 just uses $T(x)$, whereas scientist 2 instead constructs

$$Y \sim \text{Binomial}\,(n, 1 - \theta)$$

corrosponding to focusing on the number of failures of the n independent Bernoulli trials instead of the successes. Now scienist 2 uses $T(y)$ as an estimate of $1-\theta$. This estimate is equivalent to the estimate $1-T(y)=1-T(n-x)$ of θ.

The principle of equivariance states that it's reasonable to require that the two scientists obtain the same estimate for θ for all $x \in \{0,1,\ldots,n\}$ i.e.,

$$T(x) = 1 - T(n-x) \quad \text{for all } x \in \{0,1,\ldots,n\}.$$

The principle of equivariance states that we should only consider estimators satisfying this condition. This is a greatly simplification of the possible estimators.

For simplicify assume that $n = 2k$. Then an estimator requires specification of

$$T(0), T(1), \ldots, T(n),$$

but using the principle of equivariance it's only necessary to specify

$$T(0), T(1), \ldots, T(k)$$

because

$$\begin{aligned} T(n) &= 1 - T(n-n) = 1 - T(0) \\ T(n-1) &= 1 - T(1) \\ &\vdots \\ T(k+1) &= 1 - T(k-1). \end{aligned}$$

If for example we consider the generalized Agresti-Coull estimator of p given by

$$\tilde{T}(X) = \frac{X+c}{n+2c},$$

then it's easy to verify that

$$\begin{aligned} 1 - \tilde{T}(n-x) &= 1 - \frac{(n-x)+c}{n+2c} \\ &= \frac{(n+2c)-(n-x+c)}{n+2c} \\ &= \frac{n+2c-n+x-c}{n+2c} \\ &= \frac{x+c}{n+2c} \\ &= \tilde{T}(x) \end{aligned}$$

for all $x \in \{0,1,\ldots,n\}$, showing that it satisfies the principle of equivariance. □

Chapter 3

Meta analysis

This entire chapter is based on [Kulinskaya et al., 2008] and the lecture material from STA4AMD.

In this chapter some theory about meta analysis and neo meta analysis (using transformed effects) will be described. In chapter 4 some examples utilizing the different meta analysis approaches will be presented.

Meta analysis is a theory of how to combine studies to one conclusion using the information in the provided studies in the best possible way.

What we want to combine from each study is an effect, and all the studies estimate the same effect. Throughout this chapter it is assumed that K independent studies are to be combined. The estimated effect for the k'th study is denoted $\hat{\theta}_k$ and are based on N_k observations.

A key assumption is that the effect is asymptotic normal distributed. Non-trivial examples where this is the case is for two binomial samples (e.g. treatment and control group) with the effect is risk difference, log relative risk, or log odds ratio. Notice that relative risk and odds ratio are not asymptoticly normal without the logarithm transformation. This assumptions is denoted as

$$\hat{\theta}_k \sim AN\left(\theta_k, \frac{\sigma_k^2}{N_k}\right) \quad \text{for} \quad k = 1, 2, \ldots, K.$$

Effects for comparing normal samples can be mean difference $\mu_1 - \mu_2$ or Cohen's d defined as $d = \frac{\mu_1 - \mu_2}{\sigma}$ with σ^2 is assumed to be a common variance.

By combining the estimated effects one either hopes to be able to estimate a representative θ for all K studies, or maybe even for all studies of the same type (the latter is treated in section 3.1.3).

Also define z_β by $P\left(Z \leq z_\beta\right) = \beta$ for $Z \sim N(0, 1)$.

3.1 Traditional meta analysis

Often one operates with three different models, UFEM, FEM, and REF which will now be described.

Like inituiotion might tell us, each study has to be weightened when combining the studies. This is normally done using an inverse variance approach. This is perfectly reasonable when the variance is known, but that is often not the case. Therefore the weights w_k are estimated with the inverse estimate of the variances. This estimation of w_k turns out to be crucial in the quality of the confidence intervals for a combined "true" parameter in terms of coverage probability. This topic is investigated further in section 3.2.

3.1.1 Unequal fixed effects model (UFEM)

This model assumes that the the true effects are fixed and not necessarily equal, hence the name.

Because the effects are asymptotic normal, a large sample $100(1-\alpha)\%$ confidence interval for each θ_k is

$$\hat{\theta}_k \pm z_{1-\frac{\alpha}{2}} \sigma_k N_k^{-\frac{1}{2}}$$

if the variance is known, and if that is not the case, then

$$\hat{\theta}_k \pm z_{1-\frac{\alpha}{2}} \hat{\sigma}_k N_k^{-\frac{1}{2}}$$

where the variance is estimated by the usual estimate.

In order to combine the studies, each study is weigthed by

$$w_k = \frac{N_k}{\sigma_k^2}$$

to calculate a weighted effect

$$\theta_w = \frac{\sum_{k=1}^{K} w_k \theta_k}{W} \quad \text{for} \quad W = \sum_{k=1}^{K} w_k.$$

yielding

$$\hat{\theta}_w \sim AN\left(\theta_w, W^{-1}\right).$$

Now the a large sample $100(1-\alpha)\%$ confidence interval for θ_w is

$$\hat{\theta}_w \pm z_{1-\frac{\alpha}{2}} W^{-\frac{1}{2}}.$$

Because we seldom have the w_k's (and hence not θ_w nor W), these are estimated by their usual estimates. Thus the a large sample $100(1-\alpha)\%$ confidence interval for θ_w is

$$\hat{\theta}_{\hat{w}} \pm z_{1-\frac{\alpha}{2}} \hat{W}^{-\frac{1}{2}}.$$

Here a representative θ for all K studies is estimated.

3.1.2 (Equal) Fixed effects model (FEM)

This is a special case of UFEM with $\theta_k = \theta$ for all k. This directly results in

$$\theta_w = \frac{\sum_{k=1}^{K} w_k \theta_k}{\sum_{k=1}^{K} w_k} = \frac{\theta \sum_{k=1}^{K} w_k}{\sum_{k=1}^{K} w_k} = \theta,$$

and hence a large sample $100(1-\alpha)\%$ confidence interval for θ is

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} W^{-\frac{1}{2}} \quad \text{for} \quad \hat{\theta} = \frac{\sum_{k=1}^{K} w_k \hat{\theta}_k}{W}$$

if the weights are known, because this results in the smallest asymptotic variance among all unbiased linear combinations of the θ_k's (this can be shown using for instance Lagrange multipliers). If the weights are to be estimated, a large sample $100(1-\alpha)\%$ confidence interval for θ is

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \hat{W}^{-\frac{1}{2}} \quad \text{for} \quad \hat{\theta} = \frac{\sum_{k=1}^{K} \hat{w}_k \hat{\theta}_k}{\hat{W}}.$$

3.1.3 Random effects model (REM)

This model tries to estimate effects from all studies of the same type. It is assumed that the true effects θ_k are realizations of a random variable distributed as $N\left(\theta, \gamma^2\right)$ with both parameters unknown, where γ^2 interprets as an inter-study variance. Hence inference on θ can be interpreted as saying something about all such possible studies.

Because it is assumed that the K studies is a random sample from the possible studies, it is further assumed that the results for (U)FEM is condifional on the θ_k's, i.e.

$$\hat{\theta}_k | \theta_k \sim AN\left(\theta_k, \frac{\sigma_k^2}{N_k}\right) \quad \text{for} \quad k = 1, 2, \ldots, K.$$

Then the unconditional distribution is

$$\hat{\theta}_k \sim AN\left(\theta_k, \frac{\sigma_k^2}{N_k} + \gamma^2\right) \quad \text{for} \quad k = 1, 2, \ldots, K.$$

If the conditional distribution is exactly normal, then the unconditional distribution would be exactly normal as well. Hence the inverse variance weights are

$$w_k^* = \left(\frac{\sigma_k^2}{N_k} + \gamma^2\right)^{-1} = \left(w_k^{-1} + \gamma^2\right)^{-1}$$

used by the estimator

$$\theta^* = \frac{\sum_{k=1}^{K} w_k^* \hat{\theta}_k}{W^*} \sim AN\left(\theta, \left(W^*\right)^{-1}\right) \quad \text{for} \quad W^* = \sum_{k=1}^{k} w_k^*.$$

Because γ^2 is in general unknown, this has to be estimated. To do that, first define Cochran's Q as

$$Q = \sum_{k=1}^{K} w_k \left(\hat{\theta}_k - \hat{\theta}_w\right)^2 \quad \text{for} \quad \hat{\theta}_w = \frac{\sum_{k=1}^{K} w_k \hat{\theta}_k}{W}.$$

It is possible to show that under the null hypothesis of homogeneity, i.e. $\theta_k = \theta$ for all k, it is true that

$$Q \sim \chi_{K-1}^2$$

asymptotically, so that a level α-test rejects the null hypothesis of homogeneity when

$$Q \geq \chi^2_{K-1,1-\alpha}.$$

Using these facts, we can estimate γ^2. To do that, define

$$M_r = \sum_{k=1}^{k} w_k^r \quad \text{and} \quad a = M_1 - \frac{M_2}{M_1}$$

for the inverse variance weights $w_k = \frac{N_k}{\sigma_k^2}$ like before. Then it can be shown that

$$\mathbf{E}[Q] = K - 1 + a\gamma^2,$$

yielding the DerSimonian and Laird estimator given by

$$\hat{\gamma}^2_{DL} = \frac{\{Q - (K-1)\}^+}{a}$$

where

$$\{x\}^+ = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Using the $\hat{\gamma}^2_{DL}$ estimator, we are now able to estimate the weights

$$\hat{w}_k^* = \left(\hat{w}_k^{-1} + \hat{\gamma}^2_{DL}\right)^{-1}$$

and hence also the estimator

$$\hat{\theta}^* = \frac{\sum_{k=1}^{K} \hat{w}_k^* \hat{\theta}_k}{\hat{W}^*}$$

of θ and hence a large sample $100(1-\alpha)\%$ confidence interval for θ is

$$\theta^* \pm z_{1-\frac{\alpha}{2}} \left(W^*\right)^{-\frac{1}{2}}$$

estimated by

$$\hat{\theta}^* \pm z_{1-\frac{\alpha}{2}} \left(\hat{W}^*\right)^{-\frac{1}{2}}.$$

3.2 Neo meta analysis

Like described in section 3.1, the inverse variance weights are often estimated because the variance is unknown. This leads to quite poor intervals if the sample sizes are small because the results are only asymptotic. To avoid this problem, one approach is to get known variances, e.g. to get variance 1 such that the weights only depend on the known sample sizes for the studies. A way to do this is to transform the effects to evidence with vst's (variance stabilization transformations). Although this also has to be done approximate, [Kulinskaya et al., 2008, p. 128] states that it turns out to yield much better results than the traditional approach.

First the concept of evidence is described, and then how effects can be transformed to evidence, usually with vst's. This way of getting known variance usually gives the great bonus of giving fast converging (asymptotic) normality as well.

The theory described in this section will be used for some examples in chapter 4.

3.2.1 Evidence

[Kulinskaya et al., 2008] gives a lot of motivation of defining evidence. To get this background-material, please refer to [Kulinskaya et al., 2008].

A measure of evidence is defined as follows. Let θ be an unknown effect and imagine that we want to test $H_0 : \theta = 0$ against $H_1 : \theta > 0$. Furthermore, let S be a test statistic which rejects H_0 for large values of S. Then the evidence T should fulfill the following four properties (found in [Kulinskaya et al., 2008, p. 115]):

- E_1: The one-sided evidence T is monotonically increasing function of S
- E_2: The distribution of T is normal for all values of the unknown parameters
- E_3: $\mathbf{Var}[T] = 1$ for all values of the unknown parameters
- E_4: The expected evidence $\tau = \tau(\theta) = \mathbf{E}_\theta[T]$ is monotonically increasing in θ from $\tau(0) = 0$

Note that a measure of evidence will always have a standard error of 1. Also note that:

> *In general, properties E_2 fo E_4 will hold only approximately, but to a surprising degree, even for small sample sizes.*
>
> [Kulinskaya et al., 2008, p. 115]

The p-value for an observed $S = s$ is $p = P_0(S \geq s)$ where P_0 is the probability under the null hypothesis. For an observed value of a measure of evidence $T = t$, then the p-value can be computed by

$$p = P_0(T \geq t) = \Phi(-t)$$

because T is normal with mean value τ and variance 1, and at the null hypothesis $\tau = 0$ so that T is standard normal at the null hypothesis. This also means that a p-value can be converted to evidence as

$$p = 1 - \Phi(t) \quad \Leftrightarrow \quad \Phi(t) = 1 - p \quad \Leftrightarrow \quad t = t(p) = \Phi^{-1}(1 - p).$$

In table 3.1, which is a reproduction of [Kulinskaya et al., 2008, Table 16.1, p. 116], a comparison between p-values and values for evidence is shown. The last row in the table suggests that a p-value of 0.01 only represents $\sqrt{2}$ more evidence than a p-value of 0.05. To get twice as much evidence compared to a p-value of 0.05, a p-value of 0.0005 is required.

p	0.0005	0.001	0.01	0.02	0.025	0.05	0.1	0.1587
$t(p)$	3.291	3.090	2.326	2.054	1.960	1.645	1.276	1.000
$t(p)/t(0.05)$	2.000	1.879	1.414	1.248	1.192	1.000	0.779	0.608

Table 3.1: Selected values of p-values and comparitive values of evidence

Example 3.1 (Simple example of evidence in normal sample with known variance). This example is based on the one given in [Kulinskaya et al., 2008, p. 114-115]. Assume that n independent random samples $X_1, X_2, \ldots, X_n$ are taken with

$$X_i \sim N\left(\mu, \sigma_0^2\right) \quad \text{for} \quad i = 1, 2, \ldots, n$$

for unknown μ and known $\sigma_0^2 > 0$. Further let

$$\theta = \mu - \mu_0$$

be the unknown effect for a given $\mu_0 \in \mathbb{R}$. Now define the evidence $T = T(S)$ for $H_0 : \theta = 0$ against $H_1 : \theta > 0$ (corresponding to $\mu > \mu_0$) by

$$T = T(S) = \frac{\sqrt{n}\left(S - \mu_0\right)}{\sigma_0} = \frac{\sqrt{n}\left(\overline{X}_n - \mu_0\right)}{\sigma_0} \quad \text{for} \quad S = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

This is also called the Z-statistic. Now see that the evidence T satisfies all the four properties:

- E_1 (The one-sided evidence T is monotonically increasing function of S): Clearly statisfied
- E_2 (The distribution of T is normal for all values of the unknown parameters): Satisfied because S is normal and T is just a scaled and shifted compared to S, hence also normal
- E_3 ($\mathbf{Var}\left[T\right] = 1$ for all values of the unknown parameters): $\mathbf{Var}\left[T\right] = \frac{n}{\sigma_0^2}\mathbf{Var}\left[\overline{X}_n\right] = \frac{n}{\sigma_0^2}\frac{\sigma_0^2}{n} = 1$ for all values of μ
- E_4 (The expected evidence $\tau = \tau(\theta) = \mathbf{E}_\theta[T]$ is monotonically increasing in θ from $\tau(0) = 0$): $\tau = \mathbf{E}_\theta[T] = \frac{\sqrt{n}}{\sigma_0}\left(\mathbf{E}_\theta[\overline{X}_n] - \mu_0\right) = \frac{\sqrt{n}(\mu - \mu_0)}{\sigma_0} = \frac{\sqrt{n}\theta}{\sigma_0}$ is monotonically increasing in θ from $\tau(0) = 0$

Thus $T = T(S)$ is a measure of evidence.

As a simple example, choose $\mu_0 = \sigma_0 = 5$. For $n = 4$ and $\overline{X}_4 = 10$, the evidence against the null hypothesis against the favor of the positive alternative is $T = \frac{\sqrt{4}(10-5)}{5} = 2$ with standard error 1 corresponding to a p-value of $\Phi(-2) = 0.023$. For $n = 36$ and $\overline{X}_{36} = 10$, the evidence against the null hypothesis is 6 corresponding to a p-value of $\Phi(-6) = 9.87 \cdot 10^{-10}$. □

3.2.2 Variance stabilisation

In example 3.1 the Z-test statistic

$$Z_n = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

was used. Because the variance σ^2 was assumed known, this gave variance 1 for all unknown parameter values of μ. In general, forming Z_n for n independent observations X_i for $i = 1, 2, \ldots, n$ of any random variable with mean μ and variance σ^2, means that Z_n is approximately standard normal because of the central limit

theorem which was stated in section 2.1. For practical purposes, this is also the case although the variance is actually unknown and replaced by the usual estimator s^2 [Kulinskaya et al., 2008, p. 126].

It is stated in [Kulinskaya et al., 2008, p. 126] that variance stabilization is about chosing $h_n(S_n)$ for a test statistic S_n, where h_n might depend on the sample size, such that

$$\mathbf{Var}\left[h_n(S_n)\right] \approx 1$$

where $\approx$ means approximately close to. Like [Kulinskaya et al., 2008, definition 17.1, p. 126], this means that $\mathbf{Var}\left[h_n(S_n)\right] = 1 + o\left(n^{-1}\right) = 1 + c_n$ where $nc_n \to 0$ as $n \to \infty$ where $\{c_n\}_{n\in\mathbb{N}}$ could depend on parameter values.

Simplifying with $X = S_n$, a simple way of actually finding such h_n is noting that the δ-method states[1]

$$\mathbf{Var}\left[h_n(X)\right] \approx \mathbf{Var}\left[X\right]\left(h_n'(X)\right)^2.$$

Then provided that it's possible to write $\mathbf{Var}\left[X\right] = g\left(\mathbf{E}\left[X\right]\right)$ for a known function g [Kulinskaya et al., 2008, p. 127], the vst can then be found as

$$h_n(x) = \int^x \left(g(t)\right)^{-\frac{1}{2}} dt$$

provided that the indefinite integral exists. Thus $h_n(x)$ is defined up to an additive constant. This constant can be chosen, and is often used to center the expected value of the evidence at 0 at the null hypothesis. Also, we have that

$$\mathbf{E}\left[h_n(X)\right] \approx h_n\left(\mathbf{E}\left[X\right]\right).$$

In practise, although the δ-method only gives approximations to vst's, this approach turns out to give rise to better confidence intervals than the traditional approach cf. [Kulinskaya et al., 2008, p. 128]. This is due to the fact that the vst found often converges very quickly. Unfortunately, although the method seems simple, it's not always easy to find the vst in practise, for example we can end up with a h_n that depends on unknown parameters.

3.2.3 The Key Inferential Function

In [Kulinskaya et al., 2008, section 17.2.2, p. 127] it's argumented that the four properties for a measure of evidence is statisfied by $h_n(S_n)$. At least if substituting estimates for unknown parameters still gives a stabilized variance. If that is the case, the arguments are as follows, directly taken from [Kulinskaya et al., 2008, section 17.2.2, p. 127] (still remembering that θ is the unknown effect in interest with the null hypothesis $\theta = \theta_0$ against the alternative $\theta > \theta_0$):

[1] At least the informal version of the δ-method states this whereas the formal statement of δ-method involves cdf's instead. For assume that $X_1, X_2, \ldots$ are i.i.d. $N(\mu, \sigma^2)$ where $\mu \neq 0$. Let $\overline{X}_n$ be the usual estimator of μ. Then it can be shown that using the simple δ-method, $\frac{1}{\overline{X}_n} \approx \frac{1}{\mu} - \frac{1}{\mu^2}\left(\overline{X}_n - \mu\right)$ and thus $\mathbf{E}\left[\frac{1}{\overline{X}_n}\right] \approx \frac{1}{\mu}$, but it can be shown that $\mathbf{E}\left[\frac{1}{\overline{X}_n}\right]$ is not even defined! In [Casella and Berger, 2002, Theorem 5.5.24] a correct version of the δ-method is states as: Suppose that $Y_1, Y_2, \ldots$ is sequence of random variables that satisfies $\sqrt{n}\left(\overline{Y}_n - \theta\right) \sim N(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exits and is non-zero. Then $\sqrt{n}\left(g\left(\overline{Y}_n\right) - g(\theta)\right) \to N\left(0, \left(g'(\theta)\right)^2\sigma^2\right)$ in distribution.

- E_3 (**Var** $[T] = 1$ for all values of the unknown parameters): This is exactly how $h_n(S_n)$ was chosen.
- E_1 (The one-sided evidence T is monotonically increasing function of S): It is often the case that $h_n(S_n)$ can be chosen to be monotonically increasing in its argument. In other words, $h_n(S_n)$ is still a valid test statistic.
- E_4 (The expected evidence $\tau = \tau(\theta) = \mathbf{E}_\theta[T]$ is monotonically increasing in θ from $\tau(0) = 0$): In many cases $\mathbf{E}[h_n(S_n)]$ is of the form $\sqrt{n}K(\theta)$ where K is a known monotonic increasing function. By substracting the known constant $\sqrt{n}K(\theta_0)$ we can ensure that $T_n = h_n(S_n) - \sqrt{n}K(\theta_0)$ will have a mean $\tau = \mathbf{E}[T_n]$ that satisfies E_4 as well as inheriting properties E_1 and E_3 from $h_n(S_n)$, because h_n is defined only up to an additive constant.
- E_2 (The distribution of T is normal for all values of the unknown parameters): It often turns out that this actually is the case, at least approximately and usually with quick convergence. But this indeed has to be checked.

Hence a measure of evidence is often chosen to be

$$T_n = h_n(S_n) - \sqrt{n}\mathcal{K}(\theta_0)$$

where $\mathbf{E}[h_n(S_n)]$ is of the form $\sqrt{n}\mathcal{K}(\theta)$.

Definition 3.2 (The Key Inferential Function). This is [Kulinskaya et al., 2008, definition 17.2, p. 127]. Let a statistical model and a measure of evidence T_n that satisfies properties E_1 to E_4 be given. Supposing further that its expected evidence $\tau = \mathbf{E}[T_n] \approx \sqrt{n}\mathcal{K}(\theta)$. Then $\mathcal{K}$ is called the Key Inferential Function or simply the Key for the statistical model. □

Note that the Key Inferential Function is free of the sample size n.

Like stated in [Kulinskaya et al., 2008, p. 128], the Key Inferential Function is involved in solving many routine problems, among others is finding a $100(1-\alpha)\%$ confidence interval for θ, which is then

$$\left[\mathcal{K}^{-1}\left(\frac{T_n - z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) \;;\; \mathcal{K}^{-1}\left(\frac{T_n + z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right)\right] \tag{3.1}$$

where $\mathcal{K}^{-1}$ denotes the inverse of the Key Inferential Function and $z_{1-\frac{\alpha}{2}}$ defined such that $P(Z \leq z_\alpha) = \alpha$ for $Z \sim N(0,1)$.

So to summarize: in traditional meta analysis one has two problems:

- asymptotic normality of the effect
  - showing that the effect indeed is asymptotic normal
  - (choosing an effect with) quick convergence in order to get good confidence intervals for even small sample sizes
- estimate the weights (under the inverse variance approach), where the usual estimate s^2 turns out not always to be that good for small sample sizes

These problems can somewhat be overcome with the concepts just described of transformed effects and evidence.

3.2.4 Unequal fixed transformed effects model

Refer to section 3.1.1 for the traditional UFEM-approach and to section 3.1.2 for the traditional FEM-approach.

Assume that the evidence T_k for $\theta > \theta_0$ is given for $k = 1, 2, \ldots, K$ such that

$$T_k \sim N(\tau_k, 1)$$

where

$$\tau_k = \sqrt{n_k}\kappa \quad \text{and} \quad \kappa = \mathcal{K}_{\theta_0}(\theta).$$

and $\mathcal{K}$ is the Key Inferential Function (refer to section 3.2.3 for details).

Now define the combined evidence for $\theta > \theta_0$ by

$$T_{1:K} = \frac{\sqrt{n_1}T_1 + \cdots + \sqrt{n_K}T_K}{\sqrt{N}} \quad \text{for} \quad N = \sum_{i=1}^{K} n_i.$$

Then

$$\mathbf{E}[T_{1:K}] \approx \sqrt{N}\kappa$$

such that a $100(1-\alpha)\%$ approximate confidence interval for κ is

$$\frac{T_{1:K} \pm z_{1-\frac{\alpha}{2}}}{\sqrt{N}}.$$

Now because $\kappa = \mathcal{K}_{\theta_0}(\theta)$ such that $\theta = \mathcal{K}_{\theta_0}^{-1}(\kappa)$, a $100(1-\alpha)\%$ approximate confidence interval for θ is

$$\left[\mathcal{K}_{\theta_0}^{-1}\left(\frac{T_{1:K} - z_{1-\frac{\alpha}{2}}}{\sqrt{N}}\right) \;;\; \mathcal{K}_{\theta_0}^{-1}\left(\frac{T_{1:K} + z_{1-\frac{\alpha}{2}}}{\sqrt{N}}\right)\right].$$

3.2.5 Random transformed effects model

Refer to section 3.1.3 for the traditional REM-approach.

Assume that $\kappa_1, \ldots, \kappa_K$ are a sample from a $N(\kappa, \gamma^2)$ where both κ and $\gamma > 0$ is unknown. Further, assume that each estimator

$$\hat{\kappa}_k = \frac{T_k}{\sqrt{n_k}}$$

has a conditional distribution, given κ_k, which is $N\left(\kappa_k, n_k^{-1}\right)$. This means that

$$\hat{\kappa}_k | \kappa_k \sim N\left(\kappa_k, n_k^{-1}\right).$$

Now because in general $\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]]$, we have the unconditional properties

$$\mathbf{E}[\hat{\kappa}_k] = \kappa \quad \text{and} \quad \mathbf{Var}[\hat{\kappa}_k] = n_k^{-1} + \gamma^2$$

which defines the random transformed effects model.

Let $\overline{\kappa}$ and s_κ^2 be the sample mean and sample variance of the $\hat{\kappa}_k$'s. If all n_k are equal, then

$$S_{K-1} = \frac{\sqrt{K}(\overline{\kappa} - 0)}{s_\kappa} \sim t_{K-1}(\lambda),$$

hence the evidence for $\kappa > 0$ is

$$T^*_{1:K} = \sqrt{2K}\sinh^{-1}\left(\frac{\overline{\kappa}}{\sqrt{2}s_\kappa}\right).$$

(Refer to [Kulinskaya et al., 2008] for derivation of the vst and the Key Inferential Function for the non-central t-distribution.)

A $100(1-\alpha)\%$ approximate confidence interval for κ is based on the Student t confidence interval given by

$$\left[\overline{\kappa} - t_{K-1,1-\frac{\alpha}{2}}\frac{s_\kappa}{\sqrt{K}} \;\; ; \;\; \overline{\kappa} + t_{K-1,1-\frac{\alpha}{2}}\frac{s_\kappa}{\sqrt{K}}\right].$$

Even if the n_k differ, then as long as they do not differ too much, the above results hold. [Kulinskaya et al., 2008, chapter 25] states this requirement as when

$$\gamma > 2s_{1/n_k},$$

where s^2_{1/n_k} is the sample variance of the n_k^{-1}'s, then the above result hold.

Chapter 4

Examples of variance stabilisation and meta analysis

This sections uses the theory from chapter 3.

4.1 Variance stabilising a non-central χ^2 distribution

This example is based on assigment 4 in STA4AMD.

Assume that $X \sim \chi^2_\nu(\lambda)$, i.e. chi-square distributed with ν degrees of freedom and non-centrality parameter $\lambda \geq 0$. Let ν be known (often it is just the sample size shifted by one) and λ unknown. Assume that we want to derive a measure of evidence for the alternative $\lambda > 0$ against the null hypothesis of $\lambda = 0$.

First we will have to derive a vst for X before we can find the measure of evidence.

We known that

$$\mathbf{E}[X] = \nu + \lambda.$$

Because ν is known,

$$\mathbf{Var}[X] = 2\nu + 4\lambda = 4\nu + 4\lambda - 2\nu = 4(\nu + \lambda) - 2\nu = 4\mathbf{E}[X] - 2\nu = g(\mathbf{E}[X])$$

for

$$g(t) = 4t - 2\nu.$$

Thus

$$\begin{aligned} h(X) &= \int^X (g(t))^{-\frac{1}{2}}\, dt \\ &= \int^X (4t - 2\nu)^{-\frac{1}{2}}\, dt \\ &= \frac{1}{2}(4X - 2\nu)^{\frac{1}{2}} \\ &= \frac{1}{2}\sqrt{4X - 2\nu} \quad \text{for} \quad X \geq \frac{\nu}{2} \end{aligned}$$

is a variance stabilizing transformation for X up to an additive constant.

At the null hypothesis $\lambda = 0$, so by the δ-method we have that

$$\begin{aligned}
\mathbf{E}_{\lambda=0}[h(X)] &\approx h\left(\mathbf{E}_{\lambda=0}[X]\right) \\
&= h\left(\nu + \lambda\right)\Big|_{\lambda=0} \\
&= \frac{1}{2}\sqrt{4\left(\nu+\lambda\right) - 2\nu}\Big|_{\lambda=0} \\
&= \frac{1}{2}\sqrt{2\nu + 4\lambda}\Big|_{\lambda=0} \\
&= \frac{1}{2}\sqrt{2\nu}.
\end{aligned}$$

For $x \geq \frac{\nu}{2}$ define

$$T_\nu = T_\nu(x) = h(x) - \mathbf{E}_{\lambda=0}[h(X)] \approx \frac{1}{2}\sqrt{4x - 2\nu} - \frac{1}{2}\sqrt{2\nu} = \frac{1}{2}\left(\sqrt{4x-2\nu} - \sqrt{2\nu}\right)$$

as the evidence for the alternative hypothesis $\lambda > 0$ to the null hypothesis $\lambda = 0$.

To check that the distribution of the evidence T_ν is indeed normal, the R-script in Listing 4.1 has been used. First 100 samples from $\chi^2_\nu(\lambda)$ are simulated. Then these are transformed to the centered evidence T_ν. Then normal Q-Q plots are made to check the asymptotic normality. This is done for different values of ν and λ. The normal Q-Q plots can be seen in Figure 4.1. In Table 4.1 a few summary statistics are presented.

```
M <- 100
decimals <- 2

ncchisq.evidence <- function(nu, lambda)
{
  X <- rchisq(M, df=nu, ncp=lambda)
  X <- X[X >= nu/2]
  T.nu <- 0.5 * ( sqrt(4*X - 2*nu) - sqrt(2*nu) )

  qqnorm(T.nu, main=paste("QQ plot: nu =", nu, "and lambda =", lambda))
  qqline(T.nu)

  cat("$",
    nu, "$ & $",
    lambda, "$ & $",
    round(var(X), decimals), "$ & $",
    round(var(T.nu), decimals), "$ & $",
    round(mean(T.nu), decimals), "$ \\\\\\hline\n", sep="")
}

lambdas <- c(5, 10, 20)
nus <- c(10, 50, 100, 1000)
plots <- length(lambdas) * length(nus)

cols <- 4
rows <- floor(sqrt(plots))
while (cols * rows < plots)
  rows <- rows + 1

png(file="../figures/ex-non-central-chisq-vst-sim.png",
  bg="white", width = 2000, height = 2000)
```

```
par(mfrow=c(rows, cols))
par(cex=1.3)

sink(file="../topics/meta-analysis-examples-ex-non-central-chisq-vst-sim-tab.tex")
cat("\\begin{center}\n")
cat("\\begin{tabular}{|r|r||r|r|r|r|}\\hline\n")
cat("$\\nu$ & $\\lambda$ & $\\hat{\\sigma}_X$ & ", sep="")
cat("$\\hat{\\sigma}_{T_{\\nu}}$ & $\\overline{T_{\\nu}}$ \\\\\\hline\n", sep="")
cat("\\\\[-1em]\\hline\n")

i <- 1
lambdas.len <- length(lambdas)
for (lambda in lambdas)
{
  for (nu in nus)
    ncchisq.evidence(nu, lambda)

  if (i < lambdas.len)
    cat("\\\\[-1em]\\hline\n")

  i <- i + 1
}

cat("\\end{tabular}\n")
cat("\\end{center}\n")
sink()

par(mfrow=c(1,1))
dev.off()
```

Listing 4.1: Code for simulating the distribution of T_ν

ν	λ	$\hat{\sigma}_X$	$\hat{\sigma}_{T_\nu}$	$\overline{T_\nu}$
10	5	41.16	0.99	0.99
50	5	124.17	1.07	0.4
100	5	238.56	1.06	0.43
1000	5	2028.35	1	0.02
10	10	79.53	1.32	1.54
50	10	137.11	0.93	0.93
100	10	263	1.14	0.71
1000	10	1605.22	0.79	0.27
10	20	122.48	1.16	2.81
50	20	180.32	1.06	1.56
100	20	307.59	1.07	1.31
1000	20	2086.88	1	0.54

Table 4.1: Table with summary statistics of simulating the transformed evidence

Instead of simulating, the distribution can be found by finding the pdf and then plotting that together with the one for the standard normal distribution. To do this, we use the theorem that if $W = g(V)$, g is monotonically increasing, and the pdf f_V for V is known, then the pdf f_W of W is

$$f_W(w) = \left| \frac{1}{g'\left(g^{-1}(w)\right)} \right| f_V\left(g^{-1}(w)\right).$$

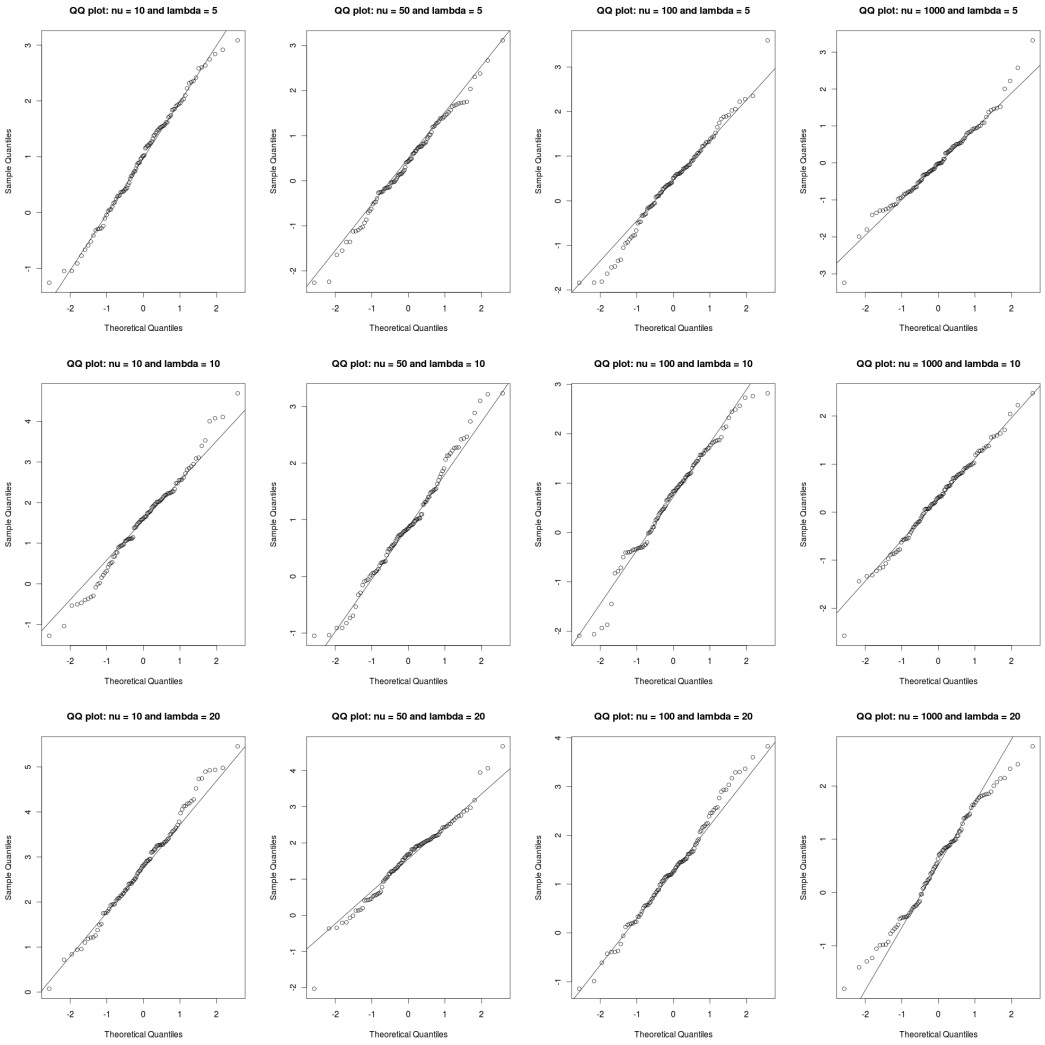

Figure 4.1: Normal Q-Q plots of the transformed evidence.

We have that

$$y = T_\nu = T_\nu(x) = \frac{1}{2}\left(\sqrt{4x - 2\nu} - \sqrt{2\nu}\right)$$

is a monotonic function of x. With a little algebra, the inverse is found to be

$$T_\nu^{-1}(y) = \frac{1}{4}\left(\left(2y + \sqrt{2\nu}\right)^2 + 2\nu\right).$$

First notice that the pdf f_X of X, i.e. the pdf for a non-central χ^2 distributed random variable with ν degrees of freedom and non-centrality parameter λ, is given by

$$f_X(x) = \sum_{i=0}^{\infty} \frac{\exp\left(-\frac{\lambda}{2}\right)\left(\frac{\lambda}{2}\right)^i}{i!} f_{V_{\nu+2i}}(x)$$

where $V_{\nu+2i} \sim \chi^2_{\nu+2i}$. Thus

$$f_{T_\nu}(y) = \left|\frac{1}{T'_\nu\left(T_\nu^{-1}(y)\right)}\right| f_X\left(T_\nu^{-1}(y)\right).$$

By using R, the dchisq-command can be used instead of the explicit infinite series for $f_X(x)$. We can then compare the pdf's for the standard normal distribution and $f_{T_\nu}(y)$ with the R-script in Listing 4.2. The plots can be found in Figure 4.2. Notice how well the pdf's match and that the pdf of the transformed evidence is centered at the null hypothesis of $\lambda = 0$.

```
pdf.T <- function(y, nu, lambda)
{
  T.diff <- function(x) (4*x - 2*nu)^(-1/2)
  T.inv <- function(y) 0.25 * ( (2*y + sqrt(2*nu))^2 + 2*nu)
  abs(1 / ( T.diff( T.inv(y) ) )) * dchisq( T.inv(y), df=nu, ncp=lambda )
}

plot.pdf.T <-function(nu, lambda)
{
  curve(dnorm(x), ylab="Probability density", xlim=c(-2, 2))
  curve(pdf.T(x, nu, lambda), add=T, lty="dashed")
  legend("topright", c("pdf for N(0, 1)", "pdf for T_nu"),
    lty=c("solid", "dashed"))
}

lambdas <- c(0, 5, 10)
nus <- c(10, 50, 100, 1000)
plots <- length(lambdas) * length(nus)

cols <- 4
rows <- floor(sqrt(plots))
while (cols * rows < plots)
  rows <- rows + 1

png(file="../figures/ex-non-central-chisq-vst-pdf.png",
  bg="white", width = 2000, height = 2000)
par(mfrow=c(rows, cols))
par(cex=1.3)

for (lambda in lambdas)
  for (nu in nus)
    plot.pdf.T(nu, lambda)

par(mfrow=c(1,1))
dev.off()
```

Listing 4.2: Code for plotting the pdf of the transformed evidence

Based on the tables and these plots, it is not unfair to claim that T_ν is approximate standard normal at the null hypothesis, and also normal with variance 1 elsewhere.

4.2 Evidence in the t-statistic

In [Kulinskaya et al., 2008, chapter 20] evidence in the t-statistic is defined and derived. Here a simple walk-through of some of the results are presented.

Let

$$X_i \sim N\left(\mu, \sigma^2\right) \quad \text{for} \quad i = 1, 2, \ldots, n$$

for unknown μ and σ^2 and X_i independent of X_j for $i \neq j$.

Assume we have the null hypothesis of $\mu = \mu_0$ against the alternative $\mu > \mu_0$ for some specified μ_0. This corresponds to finding the evidence of a positive effect

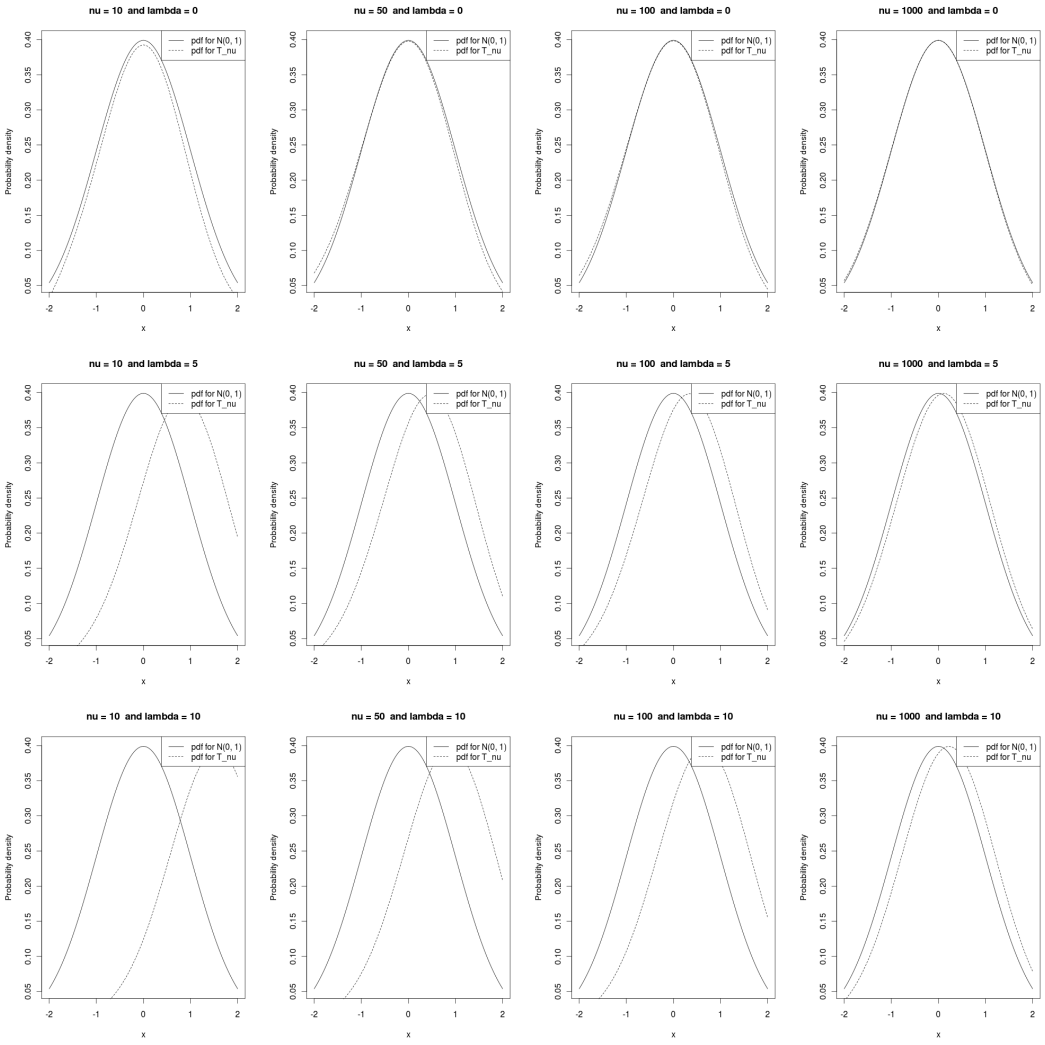

Figure 4.2: Plots of the pdf of the transformed evidence T_n against the standard normal pdf.

$\mu - \mu_0 > 0$ for a given μ_0. Equivalently we can define the effect as the standardised one, i.e.

$$\theta = \frac{\mu - \mu_0}{\sigma}$$

such that the null hypothesis $H_0 : \theta = \theta_0 = 0$ against $H_1 : \theta > \theta_0 = 0$.

We have the t-statistic given by

$$t_\nu = \frac{\sqrt{n}\left(\overline{X}_n - \mu_0\right)}{s_n} \sim t_{n-1}(\lambda),$$

i.e. non-central t distributed with $\nu = n - 1$ degrees of freedom and non-centrality parameter $\lambda = \sqrt{n}\theta$. $\overline{X}_n$ is the usual estimator of μ and s_n is the usual unbiased estimator of σ.

Azorin (supervised by Professor David Cox) published his thesis [Azorin, 1953] where he derived a vst for a $t_\nu(\lambda)$ distributed statistic, namely

$$h_\nu(x) = \sqrt{2\nu}\sinh^{-1}\left(\frac{x}{\sqrt{2\nu}}\right).$$

How this can be derived is described in [Kulinskaya et al., 2008, p. 160-161]. Observe that

$$h_\nu\left(t_\nu\right)=\sqrt{2\nu}\sinh^{-1}\left(\frac{\frac{\sqrt{n}\left(\overline{X}_n-\mu_0\right)}{s_n}}{\sqrt{2\nu}}\right)=\sqrt{2(n-1)}\sinh^{-1}\left(\frac{\frac{\sqrt{n}\left(\overline{X}_n-\mu_0\right)}{s_n}}{\sqrt{2(n-1)}}\right)$$

which simplifies if replacing $\nu=n-1$ with n. With this minor correction, we can define the evidence as

$$T=h_\nu\left(t_\nu\right)\approx\sqrt{2n}\sinh^{-1}\left(\frac{\left(\overline{X}_n-\mu_0\right)}{\sqrt{2}s_n}\right)\sim AN\left(\tau,1\right)$$

By multiplying a finite sample correction of $\frac{n-1.7}{n-1}$, [Kulinskaya et al., 2008, (20.5), p. 161] claims that the approximation gets better in the tails.

The expected evidence can be found to be

$$\tau=\mathbf{E}\left[T\right]\approx\sqrt{n}\mathcal{K}(\theta)$$

for the Key Inferential Functions (refer to section 3.2.3) given by

$$\mathcal{K}(x)=\sqrt{2}\sinh^{-1}\left(\frac{x}{\sqrt{2}}\right).$$

4.2.1 Power of a level α-test

A level α-test rejects the null hypothesis $\theta=\theta_0=0$ when $T\geq z_{1-\alpha}$. To find the power of such test, denote with $\beta(\theta_1)$ the probability of falsely accepting the null hypothesis when $\theta_1\neq\theta_0$ is the true alternative. Then the power of detecting $\theta_1>0$ is

$$\begin{aligned}1-\beta(\theta_1)=&1-P_{\theta_1}\left(T<z_{1-\alpha}\right)\\=&P_{\theta_1}\left(T\geq z_{1-\alpha}\right)\\=&P_{\theta_1}\left(T-\tau\geq z_{1-\alpha}-\tau\right)\\=&1-P_{\theta_1}\left(T-\tau\leq z_{1-\alpha}-\tau\right)\\=&1-\Phi\left(z_{1-\alpha}-\tau\right)\\=&\Phi\left(\tau-z_{1-\alpha}\right).\end{aligned}$$

4.2.2 Choosing sample size

Assume that we want to obtain a specified amount of expected evidence $\tau=\sqrt{n}\mathcal{K}\left(\theta_1\right)$ for an actually existing effect θ_1. To find the required sample size, one just solves $\tau=\sqrt{n}\mathcal{K}\left(\theta_1\right)$ for n such that

$$n=\left(\frac{\tau}{\mathcal{K}\left(\theta_1\right)}\right)^2.$$

4.2.3 Confidence interval

Because the set-up is general, the confidence interval for θ in (3.1) can be used with

$$\mathcal{K}^{-1}(y) = \frac{\exp\left(y/\sqrt{2}\right) - \exp\left(-y/\sqrt{2}\right)}{\sqrt{2}}$$

in this particular case.

4.3 Overview of evidence for statistics

This section merely plays the role of showing a few other evidence transformations for other distributions.

For one-sample t-tests we have

$$T = \sqrt{2\nu}\sinh^{-1}\left(\frac{t_\nu}{\sqrt{2\nu}}\right).$$

For one-sample Binomial tests we have

$$T = 2\sqrt{n}\left(\sin^{-1}\left(\sqrt{\tilde{p}}\right) - \sin^{-1}\left(\sqrt{p_0}\right)\right)$$

where p_0 is the null hypothesis and $\tilde{p}$ is an estimate of p (or with an ad-hoc correction applied like described in section 2.2.1).

For $X \sim \chi^2(\lambda)$ statistics, the evidence is

$$T = \left(X - \frac{\nu}{2}\right)^{\frac{1}{2}} - \left(\frac{\nu}{2}\right)^{\frac{1}{2}}.$$

Evidence in other statistics is also possible. Refer to [Kulinskaya et al., 2008] for details.

4.4 Example of a meta analysis

4.4.1 Drop in systolic blood pressure

This example is based on an assignment in STA4AMD. A lot of the theory will be applied on a data set, although not all the theory will be covered thoroughly, refer to [Kulinskaya et al., 2008] for further details. The R-code used to produce the results can be found in appendix A.

Summary statistics from $K = 7$ studies in [Mulrow et al., 2004] can be found in Table 1 in the first seven columns (with k as the first column). The last four columns are found as follows.

The estimated effect $\hat{\theta}_k$ is found as the difference between the patient groups, such that

$$\hat{\theta}_k = \overline{y}_k - \overline{x}_k.$$

	Control group			Treatment group							
k	m_k	$\overline{x}_k$	s_{1k}	n_k	$\overline{y}_k$	s_{2k}	N_k	$\hat{\theta}_k$	$s_{pool,k}$	$t_{pool,k}$	T_k
1	24	0.20	13.80	27	−4.80	13.80	51	−5.00	13.80	−1.29	−1.29
2	18	7.40	8.10	20	13.30	8.10	38	5.90	8.10	2.24	2.22
3	64	4.00	15.70	66	11.00	17.10	130	7.00	16.43	2.43	2.42
4	9	−3.00	13.50	10	4.00	15.30	19	7.00	14.48	1.05	1.05
5	25	15.00	16.50	24	8.00	20.40	49	−7.00	18.51	−1.32	−1.32
6	5	2.50	5.10	5	9.80	7.10	10	7.30	6.18	1.87	1.82
7	14	9.90	6.40	19	12.50	6.30	33	2.60	6.34	1.16	1.16

Table 4.2: Summary statistics from the seven studies in [Mulrow et al., 2004]. In each study the sample mean $\overline{x}_k$ gives the average drop in systolic blood pressure with corresponding standard deviation s_{1k} for a control group of m_k patients. $\overline{y}_k$ is the average drop with corresponding standard deviation s_{2k} for a group of n_k patients following a weight reducing diet. N_k is the total number of patiens in the k'th study, i.e. $N_k = m_k + n_k$.

The pooled estimate $s_{pool,k}$ of standard deviation is found by

$$s_{pool,k} = \sqrt{\frac{(m_k - 1)s_{1k}^2 + (n_k - 1)s_{2k}^2}{m_k + n_k - 2}}$$

because $s_{1k} \approx s_{2k}$ in the studies. (The case with differing sample sizes is described briefly in [Kulinskaya et al., 2008, p. 32].) Thus $s_{pool,k}$ can be seen as an estimator of some common unknown standard deviation σ.

The two-sample pooled t-statistic $t_{pool,k}$ is defined as

$$t_{pool,k} = \frac{\sqrt{m_k n_k}\theta_k}{\sqrt{N_k}s_{pool,k}}$$

such that $t_{pool,k} \sim t_{\nu_k}(\lambda)$ with $\nu_k = N_k - 2$ and the non-centrality parameter λ as described in [Kulinskaya et al., 2008, p. 32] (where an explicit λ also can be found).

The evidence T_k for a positive effect $\theta_k > 0$ is found by

$$T_k = \sqrt{N_k} \cdot \mathcal{K}\left(\frac{t_{pool,k}}{\sqrt{N_k}}\right) = \sqrt{N_k} \cdot \mathcal{K}\left(\frac{\sqrt{\frac{n_k}{N_k}\left(1 - \frac{n_k}{N_k}\right)}\,(\overline{y}_k - \overline{x}_k)}{s_{pool,k}}\right)$$

where $\mathcal{K}(x) = \sqrt{2} \cdot \sinh^{-1}\left(x/\sqrt{2}\right)$ (refer to section 4.2, [Kulinskaya et al., 2008, chap. 3], or [Kulinskaya et al., 2008, chap. 20] where the latter contains an elaborate theoretical argument).

Hypothesis testing

A hypothesis tests of $\theta_k = 0$ against $\theta_k > 0$ at level $\alpha = 0.05$ can be made using $t_{pool,k}$. Because large values of $t_{pool,k}$ favor the alternative, this is simply done by

finding the quantiles $t_{\nu_k,1-\alpha}$ corresponding to $1 - 0.05 = 0.95$, such that

$$\begin{aligned}
t_{N_1-2,0.95} &= 1.677 > t_{pool,1} = -1.29 \\
t_{N_2-2,0.95} &= 1.688 \leq t_{pool,2} = 2.24 \\
t_{N_3-2,0.95} &= 1.657 \leq t_{pool,3} = 2.43 \\
t_{N_4-2,0.95} &= 1.740 > t_{pool,4} = 1.05 \\
t_{N_5-2,0.95} &= 1.678 > t_{pool,5} = -1.32 \\
t_{N_6-2,0.95} &= 1.860 \leq t_{pool,6} = 1.87 \\
t_{N_7-2,0.95} &= 1.696 > t_{pool,7} = 1.16
\end{aligned}$$

This means that study 2, 3, and 6 reject the null hypothesis $\theta_k = 0$ in favour of $\theta_k > 0$ at significance level $\alpha = 0.05$.

Instead of using a hypothesis test, we can look at the evidence T_k. If we as usual let $T = 1.645$ correspond to weak evidence, then study 2, 3, and 6 would have weak evidence for the alternative because $T_2 = 2.22 \geq 1.645$, $T_3 = 2.42 \geq 1.645$, and $T_6 = 1.82 \geq 1.645$. This is in correspondence with the previous result.

Assuming fixed effects model

If we assume a fixed effects model, i.e. $\theta_k = \theta$ for all k, we can find a 95% confidence interval for this common θ using the traditional meta-analysis with inverse variance weights approach like described in section 3.1.2.

From [Kulinskaya et al., 2008, p. 83] we have that

$$\hat{w}_k = \frac{m_k n_k}{N_k s^2_{pool,k}}$$

resulting in

$$\hat{W} = \sum_{k=1}^{7} \hat{w}_k = 0.6556.$$

Note that the weights are slightly different than described in 3.1.1 because we now have both a control and a treatment group.

The point estimate of θ is

$$\hat{\theta} = \frac{\sum_{k=1}^{7} \hat{w}_k \hat{\theta}_k}{\hat{W}} = 3.459.$$

We then have that an approximate $100(1-\alpha)\% = 95\%$ confidence interval for θ given by

$$[\hat{\theta} - z_{1-\frac{\alpha}{2}} \hat{W}^{-\frac{1}{2}} \; ; \; \hat{\theta} + z_{1-\frac{\alpha}{2}} \hat{W}^{-\frac{1}{2}}] = [3.459 - 2.420 \; ; \; 3.459 + 2.420] = [1.038 \; ; \; 5.879].$$

Assuming random effects model

With the weights just found we can compute Cochran's Q as defined in section 3.1.3 and test at level 0.05 for heterogeneity of effects.

We have that

$$Q = \sum_{k=1}^{7} \hat{w}_k \left(\hat{\theta}_k - \hat{\theta}\right)^2 = 12.450$$

and because

$$\chi^2_{7-1,0.95} = 12.591,$$

hence at level 0.05, then because

$$12.450 = Q < \chi^2_{7-1,0.95} = 12.591,$$

we cannot reject the null hypothesis of homogeneity. But the margin is very small.

Although Cochran's Q doesn't support it, assume a random effects model. Thus $\theta_k \sim AN(\theta, \gamma^2)$ and θ_i and θ_j are independent for $i \neq j$ and $\hat{\theta}_k|\theta_k \sim AN(\theta_k, w_k^{-1})$. Now it's possible to find the DerSimonian and Laird estimate $\hat{\gamma}^2_{DL}$ of γ^2 as described in section 3.1.3 and then afterwards a confidence interval for θ.

First

$$\hat{\gamma}^2_{DL} = \frac{\{Q - (K-1)\}^+}{a} = \frac{6.450}{0.524} = 12.297$$

with $a = M_1 - \frac{M_2}{M_1}$ and $M_r = \sum_{k=1}^{7} w_k^r$.

Now the new weights are found by

$$\hat{w}_k^* = \left(\hat{\gamma}^2_{DL} + \hat{w}_k^{-1}\right)^{-1}$$

resulting in

$$\hat{W}^* = \sum_{k=1}^{7} \hat{w}_k^* = 0.273.$$

The point estimate for θ is

$$\hat{\theta}^* = \frac{\sum_{k=1}^{7} \hat{w}_k^* \hat{\theta}_k}{\hat{W}^*} = 3.026.$$

We then have that an approximate $100(1-\alpha)\% = 95\%$ confidence interval for θ given by

$$[\hat{\theta}^* - z_{1-\frac{\alpha}{2}} \left(\hat{W}^*\right)^{-\frac{1}{2}} \; ; \; \hat{\theta}^* + z_{1-\frac{\alpha}{2}} \left(\hat{W}^*\right)^{-\frac{1}{2}}] = [3.026 - 3.746 \; ; \; 3.026 + 3.746] = [-0.720 \; ; \; 6.771].$$

This seems reasonable that the interval gets wider than when assuming fixed effects model because study 1 and 5 has notable different values of $\hat{\theta}_k$ than the rest.

Using transformed effects

First define the standardized effect

$$\delta_k = \frac{\sqrt{q_k(1-q_k)}\theta_k}{\sigma_k} \quad \text{for} \quad q_k = \frac{n_k}{N_k}$$

estimated by substituting in the usual estimates and note that

$$t_{pool,k} = \sqrt{N_k}\hat{\delta}_k \quad (\text{because } \tfrac{m_k}{N_k} = \tfrac{N_k - n_k}{N_k} = 1 - q).$$

Because we have already calculated $t_{pool,k}$, it's easily found that

$$\begin{aligned}
\hat{\delta}_1 &= -0.181 \\
\hat{\delta}_2 &= 0.364 \\
\hat{\delta}_3 &= 0.213 \\
\hat{\delta}_4 &= 0.241 \\
\hat{\delta}_5 &= -0.189 \\
\hat{\delta}_6 &= 0.590 \\
\hat{\delta}_7 &= 0.203.
\end{aligned}$$

Assuming that

$$\mathbf{Var}\left[x_{ik}\right] = \mathbf{Var}\left[y_{jk}\right] = \sigma_k^2 \quad \text{for all } i = 1, 2, \ldots, m_k \text{ and } j = 1, 2, \ldots, n_k$$

such that

$$\mathbf{Var}\left[\overline{x}_k\right] = m_k^{-2} \sum_{i=1}^{m_k} \sigma_k^2 = \frac{\sigma_k^2}{m_k} \quad \text{and} \quad \mathbf{Var}\left[\overline{y}_k\right] = \frac{\sigma_k^2}{n_k}$$

because of independence, we also note that

$$\begin{aligned}
\mathbf{Var}\left[\hat{\theta}_k\right] &= \mathbf{Var}\left[\overline{y}_k - \overline{x}_k\right] \\
&= \mathbf{Var}\left[\overline{y}_k\right] + \mathbf{Var}\left[\overline{x}_k\right] \quad \text{(because of independence)} \\
&= \frac{\sigma_k^2}{m_k} + \frac{\sigma_k^2}{n_k} \\
&= \frac{n_k\sigma_k^2 + m_k\sigma_k^2}{m_k n_k} \\
&= \frac{\sigma_k^2\left(n_k + m_k\right)}{m_k n_k} \\
&= \frac{\sigma_k^2 N_k}{m_k n_k} \\
&= \sigma_k^2 \left(\frac{m_k n_k}{N_k}\right)^{-1} \\
&= \sigma_k^2 \left(N_k \frac{m_k n_k}{N_k N_k}\right)^{-1} \\
&= \sigma_k^2 \left(N_k q\left(1-q\right)\right)^{-1} \quad \text{(because } m_k/N_k = 1-q \text{ and } n_k/N_k = q\text{)} \\
&= \frac{\sigma_k^2}{N_k q\left(1-q\right)}.
\end{aligned}$$

Now we can find the transformed (standardized) effects, again with

$$\mathcal{K}(\delta) = \sqrt{2} \cdot \sinh^{-1}\left(\frac{\delta}{\sqrt{2}}\right)$$

as the Key Inferential Function yielding

$$\begin{aligned}
\hat{\kappa}_1 &= \mathcal{K}\left(\hat{\delta}_1\right) = -0.180 \\
\hat{\kappa}_2 &= \mathcal{K}\left(\hat{\delta}_2\right) = 0.360 \\
\hat{\kappa}_3 &= \mathcal{K}\left(\hat{\delta}_3\right) = 0.212 \\
\hat{\kappa}_4 &= \mathcal{K}\left(\hat{\delta}_4\right) = 0.240 \\
\hat{\kappa}_5 &= \mathcal{K}\left(\hat{\delta}_5\right) = -0.188 \\
\hat{\kappa}_6 &= \mathcal{K}\left(\hat{\delta}_6\right) = 0.575 \\
\hat{\kappa}_7 &= \mathcal{K}\left(\hat{\delta}_7\right) = 0.202.
\end{aligned}$$

We can now obtain a $100(1-\alpha)\% = 95\%$ confidence interval for κ using the combined evidence because

$$T_k = \sqrt{N_k}\mathcal{K}\left(\hat{\delta}_k\right) = \sqrt{N_k}\hat{\kappa}_k,$$

which has already been calculated earlier. The combined evidence assuming equal effect size model is defined in section 3.2.4 yielding

$$T_{1:K} = \frac{\sum_{k=1}^{K}\sqrt{N_k}T_k}{\sqrt{\sum_{k=1}^{K}N_k}} = 2.192$$

such that an approximate $100(1-\alpha)\% = 95\%$ confidence interval for κ is

$$\left[\frac{T_{1:K} - z_{1-\frac{\alpha}{2}}}{\sqrt{\sum_{k=1}^{K}N_k}}\ ;\ \frac{T_{1:K} + z_{1-\frac{\alpha}{2}}}{\sqrt{\sum_{k=1}^{K}N_k}}\right] = [0.013\ ;\ 0.229].$$

To obtain a approximate $100(1-\alpha)\% = 95\%$ confidence interval for δ, we apply the inverse transformation

$$\delta = \mathcal{K}^{-1}(\kappa) = \sqrt{2}\cdot\sinh\left(\frac{\kappa}{\sqrt{2}}\right)$$

on the endpoints, such that

$$[\mathcal{K}^{-1}(0.013)\ ;\ \mathcal{K}^{-1}(0.229)] = [0.013\ ;\ 0.230]$$

is a approximate $100(1-\alpha)\% = 95\%$ confidence interval for δ. This is almost the same as for κ, because $\mathcal{K}^{-1}(\cdot)$ behaves almost like the identity near origin, see e.g. [Kulinskaya et al., 2008, fig. 3.1, p. 25].

We can now compute Cochrans Q^* for the transformed effects. This is simply done using [Kulinskaya et al., 2008, p. 213] stating that

$$Q^* = \sum_{k=1}^{7} N_k\left(\hat{\kappa}_k - \hat{\bar{\kappa}}\right)^2 = 16.067 \quad \text{for} \quad \hat{\bar{\kappa}} = \frac{\sum_{k=1}^{7}N_k\hat{\kappa}_k}{\sum_{k=1}^{7}N_k}.$$

It further states that the evidence in Q^* for heterogeneity of the κ_k's is

$$T_{Q^*} = h_{K-1}(Q^*)$$

with $h_\nu(S)$ defined in [Kulinskaya et al., 2008, (22.1), p. 185] as

$$T_\nu = h_\nu(S) = \begin{cases} +\sqrt{S - m_\nu/2} - \sqrt{m_\nu/2} & \text{for } S \geq m_\nu \\ -\sqrt{S^* - m_\nu/2} + \sqrt{m_\nu/2} & \text{for } S < m_\nu \end{cases}$$

for $S^* = F_\nu^{-1}(1 - F_\nu(S))$ where F_ν is the cdf of the central χ^2 distribution with ν degrees of freedom and m_ν is the median of this distribution.

To find the this, we evaluate the value for the specific case. First notice that the median of the central $\chi^2_{K-1} = \chi^2_6$ distribution is estimated to be

$$m_{K-1} = m_6 = 5.347$$

found by the `R`-script

```
median(rchisq(100000000, 6))
```

Then because $16.067 = Q^* \geq m_6 = 5.347$, we have that

$$T_{Q^*} = h_6(Q^*) = \sqrt{Q^* - m_6/2} - \sqrt{m_6/2} = 5.964.$$

In other words, the evidence for heterogeneity of the κ_k's is 5.964 ± 1, which can be interpreted as at least strong evidence using the scale proposed by [Kulinskaya et al., 2008].

Now we assume a random transformed effects model. Notice that

$$2s_{1/N_k} = 0.062,$$

so even though γ is unknown, by [Kulinskaya et al., 2008, p. 226] we can proceed as if $N_i = N_j$ for all $i, j = 1, 2, \ldots, 7$, that is by using a Student's t-interval and thus avoiding estimating γ. This means that an approximate $100(1-\alpha)\% = 95\%$ confidence interval for κ is

$$\left[\hat{\bar{\kappa}} - t_{K-1,1-\frac{\alpha}{2}} \frac{\hat{s}_\kappa}{\sqrt{K}} \; ; \; \hat{\bar{\kappa}} + t_{K-1,1-\frac{\alpha}{2}} \frac{\hat{s}_\kappa}{\sqrt{K}}\right] = [0.174 - 0.255 \; ; \; 0.174 + 0.255] = [-0.081 \; ; \; 0.430]$$

where $\hat{s}^2_\kappa$ are the usual sample variance estimate of the $\hat{\kappa}$'s. To obtain an interval for δ, this is transformed with

$$\delta = \mathcal{K}^{-1}(\kappa) = \sqrt{2} \cdot \sinh\left(\frac{\kappa}{\sqrt{2}}\right)$$

like in part h), such that an approximate $100(1-\alpha)\% = 95\%$ confidence interval for δ is

$$\left[\mathcal{K}^{-1}(-0.081) \; ; \; \mathcal{K}^{-1}(0.430)\right] = [-0.081 \; ; \; 0.437]$$

which again is almost the same because the interval is near origin.

4.4.2 Meta-regression: Vaccination for the prevention of tuberculosis

If the data consists of explanatory covariates besides the actual effects, it's possible to make a meta-regression. When the theory of transforming effects is introduced, it's quite straight forward to take the step and make a meta-regression.

The data

This example based on a tutorial and a lecture in STA4AMD. It illustrates both how meta-analysis works and what to watch out for when performing a meta-regression. Refer to appendix B for the R-script used for this example.

The data is from 13 randomized controlled trials reported in [Colditz et al., 1994] and is shown in table 4.3. Each trial compares a group vaccinated by Bacillus Calmette-Guerin (BCG) vaccine for the prevention of tuberculosis against a non-vaccinated group, i.e. a control and a treatment group just like in the example in section 4.4.1.

The efficacy of the vaccine was suspected to depend on the distance from the equator, hence the distance was included in the data set. This example illustrates how the relationship between the efficacy of the vaccine and the distance from the equator can be modelled with a meta-regression.

Study 9 was carried out on the opposite side of the equator than the other studies, but because it's only the distance from the equator and not the sign of it that is of our interest, the sign in study 9 has been dropped.

The distance from the equator is provided through the latitude, which was centered by subtracting its mean, which was 33.46.

		Vaccinated (T)		Not vaccinated (C)								
k	L_k	$D_{T,k}$	$ND_{T,k}$	$D_{C,k}$	$ND_{C,k}$	r_k	m_k	n_k	$\tilde{p}_k$	q_k	$T_{cond,k}$	κ_k
1	44	4	119	11	128	139	15	262	0.722	0.531	1.547	0.399
2	55	6	300	29	274	303	35	609	0.822	0.498	4.164	0.704
3	42	3	228	11	209	220	14	451	0.771	0.488	2.236	0.598
4	52	62	13536	248	12619	12867	310	26465	0.799	0.486	11.785	0.669
5	13	33	5036	47	5761	5808	80	10877	0.587	0.534	0.950	0.106
6	44	180	1361	372	1079	1451	552	2992	0.674	0.485	9.042	0.385
7	19	8	2537	10	619	629	18	3174	0.553	0.198	3.203	0.755
8	13	505	87886	499	87892	88391	1004	176782	0.497	0.500	-0.189	-0.006
9	27	29	7470	45	7232	7277	74	14776	0.607	0.492	1.985	0.231
10	42	17	1699	65	1600	1665	82	3381	0.790	0.492	5.740	0.634
11	18	186	50448	141	27197	27338	327	77972	0.431	0.351	2.996	0.166
12	33	5	2493	3	2338	2341	8	4839	0.386	0.484	-0.560	-0.198
13	33	27	16886	29	17825	17854	56	34767	0.518	0.514	0.061	0.008

Table 4.3: $D_{T,k}$ is the number of vaccinated people who caught the disease and $ND_{T,k}$ is the number of vaccinated people who caught the disease. The capital T in subscript stands for treatment, i.e. the vaccinated. Similarly, the capital C in subscript stands for control, i.e. the not vaccinated. The number of people in the control group is $r_k = D_{C,k} + ND_{C,k}$. The number of people getting the disease is $m_k = D_{T,k} + D_{C,k}$. The total number of people in study k is $n_k = D_{T,k} + ND_{T,k} + D_{C,k} + ND_{C,k}$. The proportion of people getting the disease that were not vaccinated is estimated by $\tilde{p}_k = \frac{D_{C,k}+\frac{3}{8}}{m_k+\frac{3}{4}}$. Refer to section 2.2.1 for details about this ad-hoc correction of the proportion estimate. The proportion of people in study k who were not vaccinated is $q_k = \frac{r_k}{n_k}$. The conditional evidence for the alternative $\tilde{p}_k > q_k$ in study k, is denoted $T_{cond,k}$. The corresponding transformed effect is κ_k.

Note that the relative risk for getting the disease when being unvaccinated is

$$\frac{D_{C,k}/ND_{C,k}}{D_{T,k}/ND_{T,k}}.$$

Conditional approach to make inference about risk ratio

Before continuing with the actual analysis, a small digression will be made. If we have two binomial samples with rates p_1 and p_2, then $\Delta = p_2 - p_1$ is called the risk difference and $\rho = p_2/p_1$ is called the risk ratio or the relative risk. These are not easily analysed (at least not yielding satisfying results because of complicated distributions or slow convergence), as oppose to the example with drop in systolic blood pressure in section 4.4.1 that also compared two samples (control and treatment). This is due the distribution of the data.

One way of making inference of the risk difference and ratio is to use the fact that it can be shown that the Poisson(np) approximates Binomial (n, p) well for large n and small p.

Applying variance stabilizing transformation to the individual samples for sufficiently large samples and the risks not too small will make it possible to calculate evidence for $\delta > 0$, where δ is the standardized effect. The problem is that this standardized effect cannot be rewritten as a function of the relative difference or risk alone. This means that the calculated measure of evidence is not particularly useful for constructing confidence intervals.

In order to make inference about the risk ratio, which is what we need here, we condition on the observed total – in this case $m_k = D_{T,k} + D_{C,k}$. Because of this, the name of this method is refered to as the conditional approach.

Assume that we have two-sample data modelled by Poisson distributions with unknown rates p_1 and p_2 denoted

$$X_{t_1} \sim \text{Poisson}\,(p_1 t_1) \quad \text{and} \quad Y_{t_2} \sim \text{Poisson}\,(p_2 t_2)$$

and let

$$\rho = \frac{p_2}{p_1}.$$

In particular we want to be able to find the evidence for the alternative $\rho > 1$ against the null $\rho = 1$ and also a confidence interval for ρ.

It can be shown that the conditional distribution of Y_{t2} given $X_{t_1} + Y_{t_2} = m$ is binomial with parameters m and

$$p = \frac{t_2 p_2}{t_1 p_1 + t_2 p_2} = \left(1 + \left(q^{-1} - 1\right)\rho^{-1}\right)^{-1} \quad \text{for} \quad q = \frac{t_2}{t_1 + t_2}$$

and ρ defined as earlier. In other words, with p defined as above, we have that

$$Y_{t_2|m} \sim \text{Binomial}\,(m, p)\,.$$

If we look at p as a function of ρ, then p is monotone increasing with inverse function

$$\rho = \frac{q^{-1} - 1}{p^{-1} - 1}. \tag{4.1}$$

Now the hypothesis of $\rho = 1$ against $\rho > 1$ is equivalent to one of $p = q$ against $p > q$.

Large values of $Y_{t_2|m}$ is supporting the alternative $p > q$, so the traditional conditional test is carried out for an observed $y = Y_{t_2|m}$ by computing and evaluating the p-value

$$P\left(Y_{t_2|m} \geq y | p = q\right).$$

The conditional evidence T_{cond} for the alternative $p > q$ (and equivalent $\rho > 1$) is obtained by applying the standard vst for binomial distributed variables, such that

$$T_{cond} = 2\sqrt{m}\left(\sin^{-1}\left(\sqrt{\tilde{p}}\right) - \sin^{-1}\left(\sqrt{q}\right)\right) \quad \text{for} \quad \tilde{p} = \frac{Y_{t_2|m} + \frac{3}{8}}{m + \frac{3}{4}}.$$

It follows that T_{cond} is approximately normal with variance 1 and mean

$$\mathbf{E}\left[T_{cond}\right] \approx \sqrt{m}\mathcal{K}\left(p\right) \quad \text{for} \quad \mathcal{K}\left(p\right) = 2\left(\sin^{-1}\left(\sqrt{p}\right) - \sin^{-1}\left(\sqrt{q}\right)\right)$$

Because $p(\rho)$ is monotone increasing in the risk ratio ρ we can interpret T_{cond} as the conditional evidence for the alternative $\rho > 1$ against the null $\rho = 1$. Thus a nominal 95% confidence interval for a transformed effect $\kappa = \mathcal{K}(p)$ is given by

$$\left[L\ ;\ U\right] = m^{-\frac{1}{2}}\left[T_{cond} - z_{0.975}\ ;\ T_{cond} + z_{0.975}\right]$$

where z_α is defined such that $P\left(Z \leq z_\alpha\right) = \alpha$ for $Z \sim N(0,1)$. So if h denotes the inverse of $\mathcal{K}\left(p\left(\rho\right)\right)$, then a nominal 95% confidence interval for ρ is given by

$$\left[h(L)\ ;\ h(U)\right].$$

Using the theory on the data

The conditional evidence for the alternative $\tilde{p}_k > q_k$ in study k, denoted $T_{cond,k}$, has to be calculated. This hypothesis reflects what is interesting in this case, namely whether the proportion of people getting the disease that were not vaccinated of all diseased, i.e.

$$\frac{D_{C,k}}{D_{T,k} + D_{C,k}}$$

estimated by $\tilde{p}_k$, equals the proportion of people in study k who were not vaccinated, i.e.

$$\frac{D_{C,k} + ND_{C,k}}{D_{T,k} + ND_{T,k} + D_{C,k} + ND_{C,k}} = q_k.$$

Then the conditional evidence for the alternative $\tilde{p}_k > q_k$ in study k is given by

$$T_{cond,k} = 2\sqrt{m_k}\left(\sin^{-1}\left(\sqrt{\tilde{p}_k}\right) - \sin^{-1}\left(\sqrt{q_k}\right)\right).$$

We then also have the transformed effects immidiately, because

$$\kappa_k = \frac{T_{cond,k}}{\sqrt{m_k}}.$$

These transformed effects have mean $\overline{\kappa} = 0.342$ and standard deviation $s_\kappa = 0.315$. Further, the weighted mean $\overline{\kappa}_w = 0.232$. Using these we can calculate Q^* just like in the example with the drop in systolic blood pressure in section 4.4.1, such that

$$Q^* = \sum_{k=1}^{13} m_k \left(\kappa_k - \overline{\kappa}_w\right)^2 = 164.28.$$

Now because

$$\chi^2_{12,0.95} = 21.03 < Q^*,$$

we adopt a random transformed effects model.

As described in section 3.2.5, the combined evidence for $\kappa > 0$ is

$$T^*_{1:K} = \sqrt{2K}\text{sinh}^{-1}\left(\frac{\overline{\kappa}}{\sqrt{2}s_\kappa}\right) = 3.61,$$

which is moderate using the scale proposed by [Kulinskaya et al., 2008].

A 95% Student's t confidence interval for κ is

$$\overline{\kappa} \pm t_{12,0.975}\frac{s_\kappa}{\sqrt{13}} = [0.152\ ;\ 0.532]\,.$$

Using the inverse of the transformation $K_q(p) = 2\left(\sin^{-1}\left(\sqrt{p}\right) - \sin^{-1}\left(\sqrt{q}\right)\right)$ assuming $q = 0.5$, the above interval for κ can be transformed into a 95% confidence interval for p. First we find the inverse, that is for $\kappa = K_q(p)$, we have

$$\begin{aligned}
\kappa &= 2\left(\sin^{-1}\left(\sqrt{p}\right) - \sin^{-1}\left(\sqrt{\frac{1}{2}}\right)\right) = 2\left(\sin^{-1}\left(\sqrt{p}\right) - \frac{\pi}{4}\right) \\
&\Updownarrow \\
\sin^{-1}\left(\sqrt{p}\right) &= \frac{\kappa}{2} + \frac{\pi}{4} \\
&\Updownarrow \\
p &= \sin^2\left(\frac{\kappa}{2} + \frac{\pi}{4}\right).
\end{aligned}$$

By using this inverse, found to be

$$p(\kappa) = \sin^2\left(\frac{\kappa}{2} + \frac{\pi}{4}\right), \tag{4.2}$$

the 95% confidence interval for p is

$$[0.576\ ;\ 0.754]\,.$$

The $q = 0.5$ is fixed to that value for all the studies is a requirement for this back-transformation to work, although it might not be the best choice, in this case it is convenient and seems representative as well.

This confidence interval for p can now be transformed to one for ρ using (4.1). In this case a 95% confidence interval for the relative risk ρ is

$$[1.358\ ;\ 3.062]\,.$$

To perform the actual meta regression, we use R with κ_k as a response and the latitude L_k as the covariate. Consider the following R-output:

```
> summary(lm(kappa~Lat,weights=m))

Call:
lm(formula = kappa ~ Lat, weights = m)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2449 -0.7463  0.4495  1.3643  2.7585

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.171305   0.062863  -2.725   0.0198 *
Lat          0.014530   0.001967   7.387 1.38e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.583 on 11 degrees of freedom
Multiple R-squared: 0.8322,Adjusted R-squared: 0.817
F-statistic: 54.57 on 1 and 11 DF,  p-value: 1.382e-05
```

This gives the relationship

$$\kappa_k(L_k) = -0.1713 + 0.0145 L_k.$$

This can be used to estimate the relative risk using L_k instead of κ_k.

Using (4.1) and (4.2) we have the relative risk ρ as a function of a transformed effect κ given by

$$\begin{aligned}
\rho(\kappa) &= \left.\frac{q^{-1}-1}{(p(\kappa))^{-1}-1}\right|_{q=\frac{1}{2}} \\
&= \frac{1}{(p(\kappa))^{-1}-1} \\
&= \frac{p(\kappa)}{1-p(\kappa)} \\
&= \frac{\sin^2\left(\frac{\kappa}{2}+\frac{\pi}{4}\right)}{1-\sin^2\left(\frac{\kappa}{2}+\frac{\pi}{4}\right)} \\
&= \tan^2\left(\frac{\kappa}{2}+\frac{\pi}{4}\right)
\end{aligned}$$

It is obviuos that $\rho(\kappa)$ is not linear over the range of κ_k's, but $\log\rho(\kappa)$ is nearly linear for $\kappa \in [-1,1]$ as seen in figure 4.3.

Using L_k, we can then estimate the log relative risk using the regression model from earlier, such that

$$\log\rho = \log\left(\rho\left(\kappa_k(L_k)\right)\right),$$

which is plotted in figure 4.4.

As a final remark, a few things have to be mentioned. First of all, the issue with selecting a representing a common q should not be neglected. Although $q = \frac{1}{2}$ is

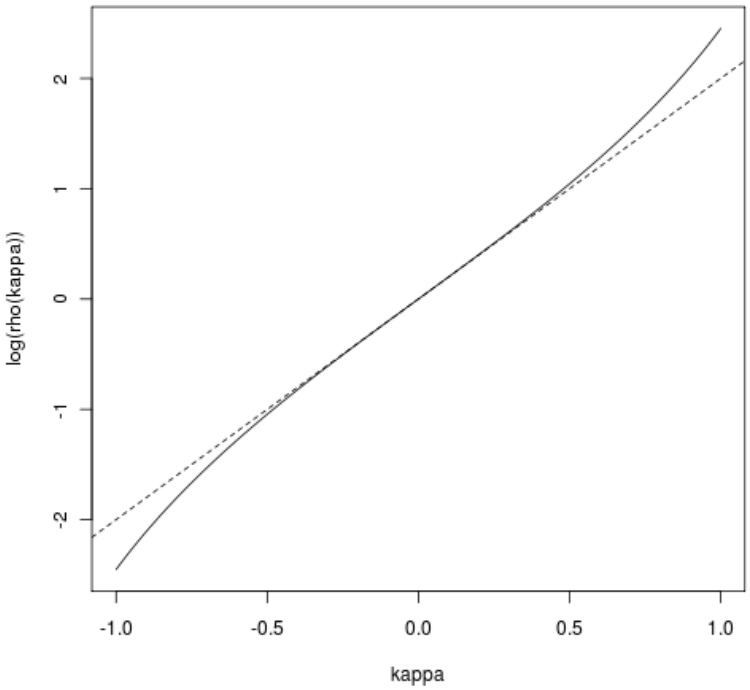

Figure 4.3: $\log(\rho(\kappa))$ is linear for κ between -1 and 1.

convenient, this is it necessarily the best choice. Second of all, instead of just using

$$\tilde{p}_k = \frac{D_{C,k} + \frac{3}{8}}{m_k + \frac{3}{4}}$$

it might be relevant to try with other constants in the corrections, e.g.

$$\tilde{p}_k = \frac{D_{C,k} + \frac{1}{2}}{m_k + 1}.$$

Refer to section 2.2.1 for details about this ad-hoc correction by Agresti and Coull of the proportion estimate.

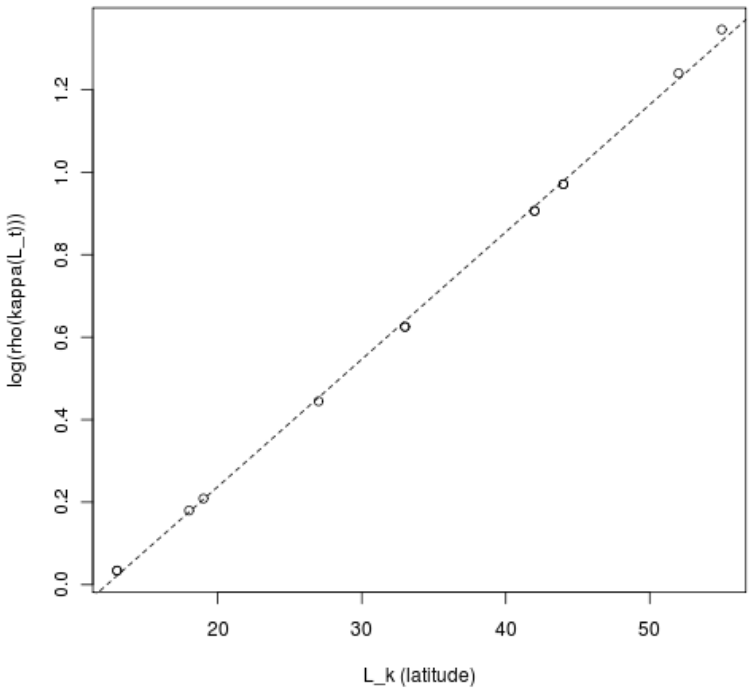

Figure 4.4: The log relative risk as a function of the latitude. The coefficients to the straight line is found using the R-command `lm`. Refer to appendix B for the R-script used for this example.

Chapter 5

Topics studied besides the course material

In this chapter a few topics are described on a very introductionary level. Some details are not included and the arguments might not be rigorous. This chapter is included as a suggestion to which topics one might look at as an extension to the topics went through in the course material. These topics have not been taught in the course, but have been suggested by lectures as follow-up topics for interested students.

5.1 Kernel principal component analysis

This section is based on [Schölkopf et al., 1996] and [Schölkopf et al., 1997]. Prior knowledge of traditional PCA (principal component analysis) is assumed. KPCA (kernel principal component analysis) is a way of tweaking PCA in the sense that it's often possible to reduce the required number of dimensions to separate the data even further than possible with PCA.

But by increasing the dimensionality we increase the calculations exponentially. This is called the curse of dimensionality, and is by [Wikipedia, 2009] which is adapted from an example by R. E. Bellman, explained to be

> *For example, 100 evenly-spaced sample points suffice to sample a unit interval with no more than 0.01 distance between points; an equivalent sampling of a 10-dimensional unit hypercube with a lattice with a spacing of 0.01 between adjacent points would require 1020 sample points: thus, in some sense, the 10-dimensional hypercube can be said to be a factor of 1018 "larger" than the unit interval.*
>
> [Wikipedia, 2009]

This is of course really bad, which is why it might be worth the effort to reduce the number of required dimensions as much as possible.

The use of KPCA will be motivated through an example. Consider figure 5.1. Suppose that it's the first two SPC (sample principal components). The coloring de-

notes group membership. As seen it's impossible to reduce to fewer dimensions without being unable to separate the groups.

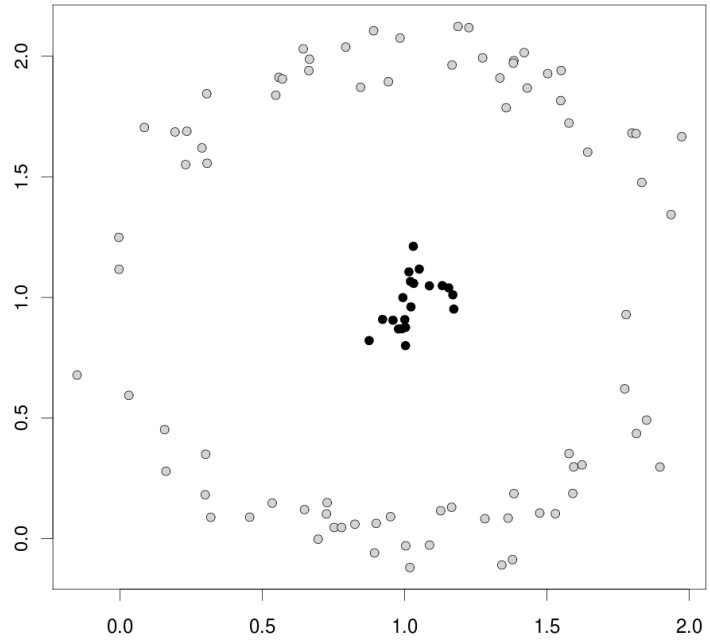

Figure 5.1: Lineary inseparable groups. The groupings are marked with different color.

If we transform the data from $\mathbb{R}^2$ to $\mathbb{R}^3$ with the transformation

$$\Phi(x_1, x_2) = (x_1, x_2, x_1^2 + x_2^2)$$

we get figure 5.2. As we can see it is now possible to separate the data with a plane.

The idea with KPCA is to map the points into a feature space in a possibly non-linear way (or at least it's not required to be linear) in order to reduce the required number of dimensions. In the previous example the feature space was $\mathbb{R}^3$.

It luckily turns out that for certain transformations, it's actually not required to perform calculations in the feature space as long as the transformation can be expression solely by dot products. Refer to [Schölkopf et al., 1996] and [Schölkopf et al., 1997] for details.

5.2 Sliced inverse regression

The sliced inverse regression (SIR) part is based on [Gentle et al., 2004].

PCA is a dimension reduction method for multidimensional data where the variables are all of the same type. It might however be necessary to reduce the number of dimensions of multidimensional data where one of the variables is a response and the others are explanatory variables. Sliced inverse regression (SIR) is a method for reducing the dimensions of the explanatory variables.

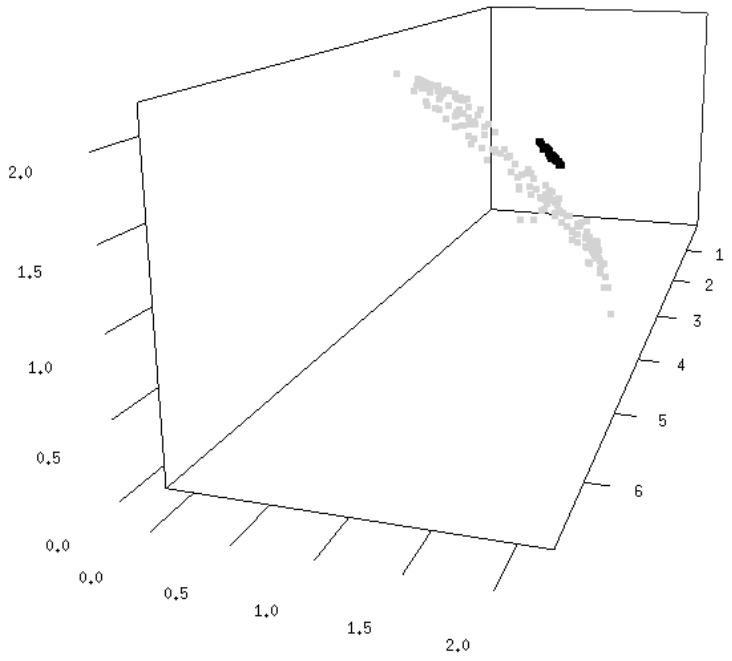

Figure 5.2: Points in figure 5.1 mapped to $\mathbb{R}^3$ in order to make them lineary separable.

There are several ways to try to reduce the number of explanatory variables. One conceptually simple approach is to make variable selection. Another is to project the explanatory variables on a lower dimensional space that estimates the responce variable well. SIR is a way of doing the latter with a linear projection. This means that SIR actually finds linear combinations of the explanatory variables.

Let

$$y = f\left(\boldsymbol{\beta}_1^\top \boldsymbol{x}, \boldsymbol{\beta}_2^\top \boldsymbol{x}, \ldots, \boldsymbol{\beta}_K^\top \boldsymbol{x}\right) + \boldsymbol{\varepsilon}$$

where $\boldsymbol{x} \in \mathbb{R}^p$ is the vector of explanatory variables; $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_K \in \mathbb{R}^p$ are unknown vectors; $\boldsymbol{\varepsilon}$ is independent of $\boldsymbol{x}$; and f is an arbitrary unknown function on $\mathbb{R}^K$. Note that f is not necessarily a linear function and that $K < p$ for the problem to make sense.

What SIR does is to estimate $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_K$ such that the model holds. If this is succesful, then we have $\boldsymbol{\beta}_1^\top \boldsymbol{x}, \ldots, \boldsymbol{\beta}_K^\top \boldsymbol{x}$ as explanatory variables, i.e. K variables instead of p.

In [Gentle et al., 2004] several algorithms to estimate the $\boldsymbol{\beta}_k$ can be found. One of the algorithms, SIR1, is based on using $\mathbf{E}[\boldsymbol{x}|\boldsymbol{y}]$. Another of the algorithms, SIR2, is instead based on using $\mathbf{E}[\mathbf{Cov}[\boldsymbol{x}|\boldsymbol{y}]]$.

Bibliography

P. Azorin. Sobre la distribución no central I,II. *Trabajos de Estadistica*, 4:173–198 and 307–337, 1953. 38

George E. P. Box, Gwilym M. Jenkins, and Gregory. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, 3 edition, 1994. 7

Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta. Interval Estimation for a Binomial Proportion. *Statistical Science*, 16(2):101–133, 2001. URL http://www-stat.wharton.upenn.edu/~tcai/paper/Binomial-StatSci.pdf. 17

George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Press, 2002. ISBN 978-0534243128. 7, 21, 29

I. S. F. Chan and Z. Zhang. Confidence intervals for the difference of two binomial proportions. *Biometrics*, 55:1202–1209, 1999. 11

G.A. Colditz, T.F. Brewer, C. S. Berkey, M. E. Wilson, E. Burdick, H. V. Fineberg, and F. Mosteller. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *The Journal of the American Medical Assiciation*, 271:698–702, 1994. 47

David Farchione and Paul Kabaila. Confidence intervals for the normal mean utilizing prior information. *Statistics & Probability Letters*, 78, 2008. URL http://linkinghub.elsevier.com/retrieve/pii/S0167715207003793. 21

P. R. Freeman. The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Statistics in Medicine*, 8:1421–1432, 1989. URL http://doi.wiley.com/10.1002/sim.4780081202. 10

James E. Gentle, Wolfgang Härdle, and Yuichi Mori. *Handbook of Computational Statistics: Concepts and Methods*. Springer, 2004. ISBN 3540404643. 56, 57

R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, 5 edition, 2001. 7, 8

Paul Kabaila. Computation of exact confidence intervals from discrete data using studentized test statistics. *Statisticsand Computing*, 15:71–78, 2005. URL http://portal.acm.org/citation.cfm?id=1050643. 17

Paul Kabaila. Statistical properties of exact confidence intervals from discrete data using studentized test statistics. *Statistics & Probability Letters*, 78:720–727, 2008. URL http://linkinghub.elsevier.com/retrieve/pii/S0167715207003239. 17

Elena Kulinskaya, Stephan Morgenthaler, and Robert G. Staudte. *Meta Analysis – A Guide to Calibrating and Combining Statistical Evidence*. Wiley Series in Probability

and Statistics, 2008. ISBN 9780470028643. 7, 17, 23, 26, 27, 28, 29, 30, 32, 37, 39, 40, 41, 42, 45, 46, 51

C. Mulrow, E. Chiquette, L. Angel, J. Cornell., C. Summerbell, B. Anagnosetelis, M. Brand, and R. Grimm. Dieting to reduce body weight for controlling hypertension in adults. *The Cochran Library*, 2004. 40, 41

R. G. Newcombe. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*, 17:873–890, 1998. 11

J.W. Pratt. Length of confidence intervals. *Journal of the American Statistical Association*, 56:549–567, 1961. URL http://www.jstor.org/stable/2282079. 17, 20

Luke Prendergast and Paul Kabaila. *STA3AS Unit Text*. La Trobe University, 2009. 7

J. A. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, 3 edition, 2006. 7

Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. Technical Report 44, Max Planck Institut fur biologische Kybernetik, 1996. URL http://www.face-rec.org/algorithms/Kernel/kernelPCA_scholkopf.pdf. 55, 56

Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. In *ICANN*, pages 583–588, 1997. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.128.7613. 55, 56

La Trobe University. Maths and Stats: Student Handbook 2009. Internet: http://www.latrobe.edu.au/mathstats/current-students/download-documents/StudentHandbook2009.pdf, 2009a. 7

La Trobe University. Maths and Stats: STA3AS Unit Guide. Internet: http://www.latrobe.edu.au/mathstats/undergraduates/unit-outlines-qa/unit-outlines-2009/Unit-Outline-STA3AS-2009.pdf, 2009b. 7

Wikipedia. Curse of dimensionality. Internet: http://en.wikipedia.org/wiki/Curse_of_dimensionality, 2009. 55

Appendix A

R-code to example with drop in systolic blood pressure

The following R-code has been used to produce the results in section 4.4.1 regarding the example with drop in systolic blood pressure.

```
KeyFunction <- function(x)
  return(sqrt(2) * asinh(x / sqrt(2)))

KeyFunctionInv <- function(y)
  return(sqrt(2) * sinh(y / sqrt(2)))

data <- matrix(c(24, 0.2, 13.8, 27, -4.8, 13.8), ncol=6)
data <- rbind(data, c(18, 7.4, 8.1, 20, 13.3, 8.1))
data <- rbind(data, c(64, 4.0, 15.7, 66, 11.0, 17.1))
data <- rbind(data, c(9, -3.0, 13.5, 10, 4.0, 15.3))
data <- rbind(data, c(25, 15.0, 16.5, 24, 8.0, 20.4))
data <- rbind(data, c(5,2.5, 5.1, 5, 9.8, 7.1))
data <- rbind(data, c(14, 9.9, 6.4, 19, 12.5, 6.3))

K <- nrow(data)

data <- cbind(data, NA, NA, NA, NA, NA)

                # 1     2          3      4     5
colnames(data) <- c("m.k", "x.mean.k", "s.1k", "n.k", "y.mean.k",
                  "s.2k", "N.k", "theta.k", "s.pool.k", "t.pool.k", "T.k")
                # 6      7     8         9          10         11

# N.k
data[, 7] <- data[, 1] + data[, 4]

# theta.k
data[, 8] <- data[, 5] - data[, 2]

# s.pool.k
data[, 9] <- ((data[, 1] - 1) * data[, 3]^2 + (data[, 4] - 1) * data[, 6]^2) /
            (data[, 1] + data[, 4] - 2)
data[, 9] <- sqrt(data[, 9])

# t.pool.k
data[, 10] <- (sqrt(data[, 1] * data[, 4]) * data[, 8]) /
              (sqrt(data[, 1] + data[, 4]) * data[, 9])
# T.k
q <- data[, 4] / data[, 7]
```

```
data[, 11] <- sqrt(data[, 7]) *
  KeyFunction(data[, 9]^(-1) * sqrt(q * (1 - q)) * data[, 8])

rownames(data) <- 1:nrow(data)

# part b)
cat("--------- part b) ---------\n")
qt(0.95, data[, 7] - 2)

# part d)
cat("--------- part d) ---------\n")
w.k <- (data[, 1] * data[, 4]) / (data[, 7] * data[, 9]^2)
W <- sum(w.k)
theta <- sum(w.k * data[, 8]) / W
c <- qnorm(0.975) * W^(-1/2)
cat("W =", W, "\n")
cat("theta =", theta, "\n")
cat("c =", c, "\n")
cat("[", theta - c, ",", theta + c, "]\n")
rm(c)

# part e)
cat("--------- part e) ---------\n")
Q <- sum(w.k * (data[, 8] - theta)^2)
cat("Q =", Q, "\n")
cat("chi^2_{K-1,0.95} = chi^2_{", K, "-1,0.95} = ",
  qchisq(0.95, K-1), "\n", sep="")

# part f)
cat("--------- part f) ---------\n")
M1 <- sum(w.k)
M2 <- sum(w.k^2)
a <- M1 - M2/M1
b <- Q - (K - 1)
gamma.sq.DL <- ifelse(b < 0, 0, b) / a

w.k.star <- (gamma.sq.DL + w.k^(-1))^(-1)
W.star <- sum(w.k.star)
theta.star <- sum(w.k.star * data[, 8]) / W.star
c <- qnorm(0.975) * W.star^(-1/2)
cat("Q - (K - 1) =", Q - (K - 1), "\n")
cat("a =", a, "\n")
cat("gamma^2_{DL} =", gamma.sq.DL, "\n")
cat("W* =", W.star, "\n")
cat("theta* =", theta.star, "\n")
cat("c =", c, "\n")
cat("[", theta.star - c, ",", theta.star + c, "]\n")
rm(c)

# part g)
cat("--------- part g) ---------\n")
delta <- sqrt(q*(1-q)) * data[, 8] / data[, 9]
cat("delta:\n")
print(delta)
cat("\n")

# part h)
cat("--------- part h) ---------\n")
kappa <- KeyFunction(delta)
cat("kappa:\n")
print(kappa)
cat("\n")

```

```
N <- sum(data[, 7])
T.comb <- sum(sqrt(data[, 7]) * data[, 11]) / sqrt(N)
cat("T_{1:K} =", T.comb, "\n")
cat("kappa in [", (T.comb - qnorm(0.975)) / sqrt(N) , ",",
  (T.comb + qnorm(0.975)) / sqrt(N), "]\n")
cat("delta in [", KeyFunctionInv((T.comb - qnorm(0.975)) / sqrt(N)) , ",",
  KeyFunctionInv((T.comb + qnorm(0.975)) / sqrt(N)), "]\n")

# part i)
cat("--------- part i) ---------\n")
kappa.mean <- sum(data[, 7] * kappa) / N
Q.star <- sum(data[, 7] * (kappa - kappa.mean)^2)
cat("Q* =", Q.star, "\n")
m6 <- 5.347
T.Q.star <- sqrt(Q.star + m6/2) + sqrt(m6/2)
cat("T_{Q*} =", T.Q.star, "\n")

cat("2*s_{1/n_k} =", 2*sd(1/data[, 7]), "\n")

kappa.sd <- sd(kappa)
c <- qt(0.975, K-1) * (kappa.sd / sqrt(K))
cat("kappa.mean =", kappa.mean, "\n")
cat("qt(0.975, K-1) * (kappa.sd / sqrt(K)) =", c, "\n")
cat("kappa in [", kappa.mean - c, ",", kappa.mean + c, "]\n")
cat("delta in [", KeyFunctionInv(kappa.mean - c), ",",
  KeyFunctionInv(kappa.mean + c), "]\n")
```

Appendix B

R-code to meta-regression example

The following R-code has been used to produce the results in section 4.4.2 regarding the example with meta-regression.

```
Lat <- c(44,55,42,52,13,44,19,13,27,42,18,33,33)
vacD <- c(4,6,3,62,33,180,8, 505, 29, 17, 186, 5, 27)
vacND <- c(119,300,228,13536, 5036, 1361,2537,87886,7470,1699,50448,2493,16886)
notvacD <- c(11,29,11,248,47,372,10,499,45,65,141,3,29)
notvacND <- c(128,274,209,12619,5761,1079,619,87892,7232,1600,27197,2338,17825)
vac <- vacD+vacND
notvac <- notvacD+notvacND
n <- notvac+vac
m <- vacD+notvacD
ptilde <- (notvacD+3/8)/(m+3/4)
q <- notvac/n
Tcond<- 2*sqrt(m)*(asin(sqrt(ptilde))-asin(sqrt(q)))
round(cbind(vacD,vacND,vac,notvacD,notvacND,notvac,m,n,ptilde,q,Tcond),digits=3)

K <- length(m)

# Transformed effects
kappa <- Tcond / sqrt(m)
cat("kappa: ")
print(round(kappa, 3))

kappa.mean <- mean(kappa)
kappa.sd <- sd(kappa)
kappa.w <- sum(m * kappa) / sum(m)
cat("kappa.mean     =", kappa.mean, "\n")
cat("kappa.sd       =", kappa.sd, "\n")
cat("kappa.w        =", kappa.w, "\n")

Qstar <- sum(m * (kappa - kappa.w)^2)
cat("Qstar          =", Qstar, "\n")
cat("chi^2_{K-1,0.95} =", qchisq(0.95, K-1), "\n")

Tstar <- sqrt(2*K) * asinh(kappa.mean / (sqrt(2) * sd(kappa)))
cat("Tstar          =", Tstar, "\n")

L <- kappa.mean - qt(0.975, K-1) * sd(kappa) / sqrt(K)
U <- kappa.mean + qt(0.975, K-1) * sd(kappa) / sqrt(K)
int <- c(L, U)
cat("95% t confidence interval for kappa: ")
print(int)

# K_q(p) = 2 * (asin(sqrt(p)) - asin(sqrt(q)))
```

```
# Finding the inverse yields:
K.q.p.inv <- function(x) (sin(0.5*x + asin(sqrt(0.5))))^2
int.p <- K.q.p.inv(int)
cat("95% t confidence interval for p: ")
print(int.p)

p.inv <- function(p) (0.5^(-1) - 1) / (p^(-1) - 1)
int.rho <- p.inv(int.p)
cat("95% t confidence interval for rho, the relative risk: ")
print(int.rho)

print(summary(lm(kappa~Lat,weights=m)))


rho <- function(k) (tan(k/2 + pi/4))^2
kappa.vec <- seq(-1, 1, 0.01)
png(file="meta-regression-linear.png", bg="white", width = 500, height = 500)
plot(kappa.vec, log(rho(kappa.vec)), type="l", xlab="kappa", ylab="log(rho(kappa))")
abline(a=0, b=2, lty='dashed')
dev.off()

# We want to estimate rho based on the kappas,
# but using the latitude as a covariate. So
# instead of using kappa, we use
# kappa = -0.1718 + 0.014528 * latitude
kappa.lin <- function(lat) -0.1718 + 0.014528 * lat
log.rho <- log(rho(kappa.lin(Lat)))

print(summary(lm(log.rho ~ Lat)))

png(file="meta-regression.png", bg="white", width = 500, height = 500)
plot(Lat, log.rho, xlab="L_k (latitude)", ylab="log(rho(kappa(L_t)))")
abline(a=-0.3796731, b=0.0308677, lty='dashed')
dev.off()
```