Conducting Meta-analysis under Confidence Distribution Framework Using gmeta Package in R

Jerry Q. Cheng

Robert Wood Johnson Medical School
Rutgers University

ASA/Bayer Statistics Workshop
November 11, 2016

Outline of Today's Presentation

- Motivation
- Confidence distribution (CD)
- Meta-analysis under CD combining framework
- Demonstrations of gmeta package
- Conclusion and future work

Fusion Learning/Meta-analysis - not just a Buzz-word

- Advances of technologies $\Rightarrow$ information explosion
  - data acquisition, storage and access
  - streaming media, cloud computing

  $\overset{need}{\Longrightarrow}$ efficient and meaningful processing

- Meta-analysis: combining results from separate studies
  - inferences from the combined results are typically more reliable than inferences from any single study
  - enormous literature: i.e. 281 references in a review by Sutton and Higgins (2008) on model based meta-analysis alone
  - important research: both in application and theoretical foundation

What is a Confidence Distribution?

- Characterization: a sample-dependent distribution function on parameter space that can represent confidence intervals of all levels for a parameter of interest.

- Θ - parameter space of the unknown parameter of interest θ, $\mathcal{X}$ - sample space corresponding to sample $\boldsymbol{x} = \{x_1, \ldots, x_n\}$

  DEFINITION: a function $H_n(\cdot) = H_n(\boldsymbol{x}, \cdot)$ on $\mathcal{X} \times \Theta \to [0, 1]$ is a CD for θ if

  1. for each given $\boldsymbol{x} \in \mathcal{X}$, $H_n(\cdot)$ is a continuous cumulative distribution function on Θ

  2. when $\theta = \theta_0$ (true), $H_n(\theta_0) \equiv H_n(\boldsymbol{x}, \theta_0) \sim U[0, 1]$

     $\Rightarrow$ provide confidence intervals of all levels for θ_0

$$\Pr(\theta_0 \leq H_n^{-1}(\alpha)) = \alpha \quad \text{for all } 0 \leq \alpha \leq 1$$

     – ensuring unbiasedness, consistency, efficiency

CD: the Past

- Long history
  - Fraser (2011) suggested the seed idea (alas, fiducial idea) can be traced back to Bayes (1763) and Fisher (1922).
  - the term "CD" appeared as early as 1937 in a letter from E.S. Pearson to W.S. Gosset (Student)
  - the first use of the term in a formal publication is Cox (1958).

- Little attention on CDs in the past
  - historic connection to fiducial distribution (considered as "Fisher's biggest blunder" (Efron 1998)
  - no important utilities of CDs in applications

CD: the Present

- Recent developments
  - entirely within the frequentist school
    - with no attempt to derive a new "paradox free" fiducial theory
  - focus on providing frequentist inference tools for problems
    - unavailable/unknown solutions

- Renewed interest
  - cluster of important publications:
    - Schweder & Hjort 2002, 2014; Singh, Xie & Strawderman 2001, 2005, 2007; Xie, Singh & Strawderman 2011, Xie & Singh 2013, Liu et al. 2014a,b, etc.
  - focus on developing new inference tools
    - Efron (1998): "... but here is a safe prediction for the 21st century: ... something like fiducial inference will play an important role ..."
    - Efron (2013): The CD development is "a grounding process" to help solve "perhaps the most important unresolved problem in statistical inference...."

CD - the Content

- Provide meaningful answers for many questions in statistical inference: e.g., point estimate, CI, p-values, etc

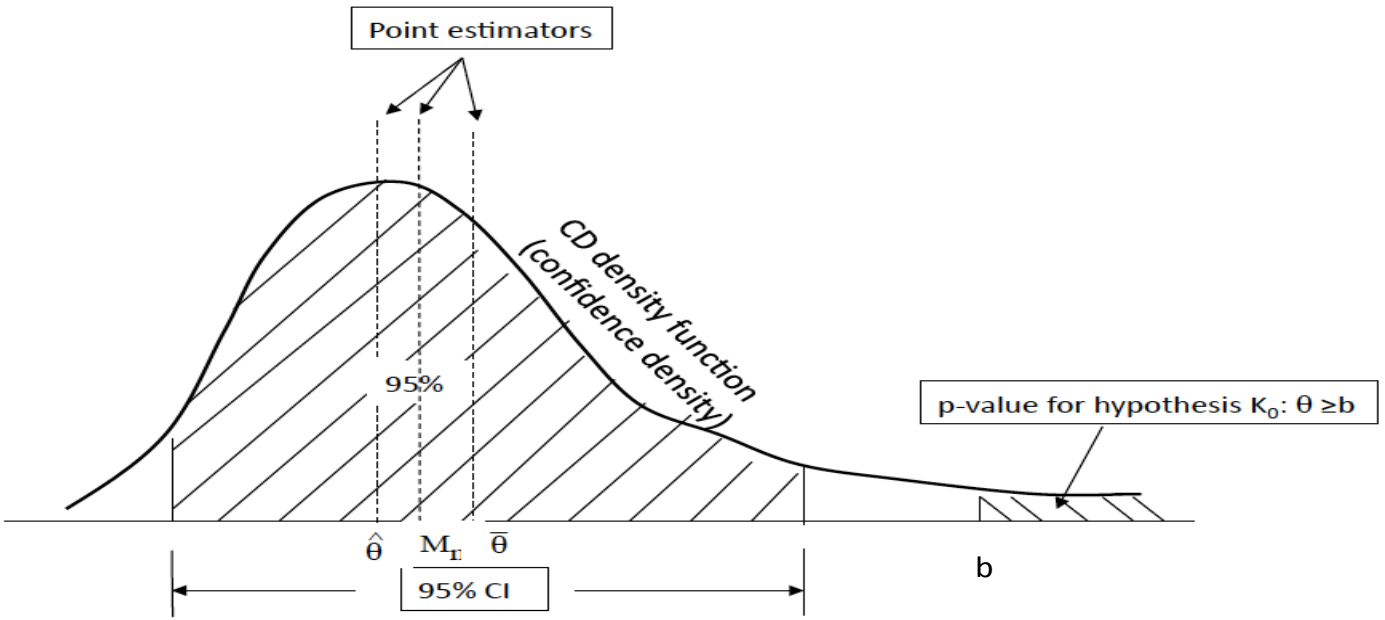


Figure 1. *The plot is a graphical illustration on making inference using a CD, including examples of point estimators (mode $\hat{\theta}$, median M_n and mean $\bar{\theta}$), a level 95% confidence interval and a one-sided p-value.*

Example 1: Normal CD function $N(\bar{x}, 1/n) \Longrightarrow$ (i) Point estimator: $\bar{x}$; (ii) level $(1-\alpha)100\%$ CI: $\bar{x} \pm \Phi^{-1}(1-\alpha/2)/\sqrt{n}$; (iii) p-value: $1-\Phi\left(\sqrt{n}(b-\bar{x})\right)$, etc.

(cf., Xie & Singh 2013; Singh et al. 2007)

CD - a statistical inference tool

- Statistical inference (parameter estimation)

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$
  - Point estimate:

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$
  - Point estimate: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$
  - Point estimate: $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$
  - Interval estimate:

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$
  - Point estimate: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Interval estimate: $(\bar{x}_n - \frac{1.96}{\sqrt{n}}, \bar{x}_n + \frac{1.96}{\sqrt{n}})$

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$
  - Point estimate: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Interval estimate: $(\bar{x}_n - \frac{1.96}{\sqrt{n}}, \bar{x}_n + \frac{1.96}{\sqrt{n}})$
  - Distribution estimate:

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$
  - Point estimate: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Interval estimate: $(\bar{x}_n - \frac{1.96}{\sqrt{n}}, \bar{x}_n + \frac{1.96}{\sqrt{n}})$
  - Distribution estimate: $N(\bar{x}_n, \frac{1}{n})$

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$
  - Point estimate: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Interval estimate: $(\bar{x}_n - \frac{1.96}{\sqrt{n}}, \bar{x}_n + \frac{1.96}{\sqrt{n}})$
  - Distribution estimate: $N(\bar{x}_n, \frac{1}{n})$

- The idea of the CD approach is to

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$
  - Point estimate: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Interval estimate: $(\bar{x}_n - \frac{1.96}{\sqrt{n}}, \bar{x}_n + \frac{1.96}{\sqrt{n}})$
  - Distribution estimate: $N(\bar{x}_n, \frac{1}{n})$

- The idea of the CD approach is to use a sample-dependent distribution (or density) function to

CD - a statistical inference tool

- Statistical inference (parameter estimation)
  - Point estimate
  - Interval estimate
  - Distribution estimate (e.g. confidence distribution)

- Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$
  - Point estimate: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Interval estimate: $(\bar{x}_n - \frac{1.96}{\sqrt{n}}, \bar{x}_n + \frac{1.96}{\sqrt{n}})$
  - Distribution estimate: $N(\bar{x}_n, \frac{1}{n})$

- The idea of the CD approach is to use a sample-dependent distribution (or density) function to estimate the parameter of interest.

Example 1: normal mean and variance

- assume sample $x_1, \ldots, x_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then

$$H_n(y) = F_{t_{n-1}}\left(\frac{y - \bar{x}}{s_n/\sqrt{n}}\right) \text{ is a CD for } \mu$$

  - $\bar{x}$: sample mean; s_n^2: sample variance
  - $F_{t_{n-1}}$: CDF of the Student t_{n-1}-distribution

** satisfying the two conditions for CD
(1). with a sample $\boldsymbol{x}$,

$$H_n(y) \to 0 \text{ as } y \to -\infty \text{ and } H_n(y) \to 1 \text{ as } y \to \infty$$

(2). with $y = \mu$,

$$\begin{aligned}\Pr(H_n(\mu) \le \alpha) &= \Pr(F_{t_{n-1}}\left(\frac{\mu - \bar{x}}{s_n/\sqrt{n}}\right) \le \alpha) \\ &= \Pr(T_{n-1} \le F_{t_{n-1}}^{-1}(\alpha)) \\ &= F_{t_{n-1}}(F_{t_{n-1}}^{-1}(\alpha)) = \alpha\end{aligned}$$

Example 2: CD from hypothesis test

- *p*-value for one-sided test: H_0: $\mu = \mu_0$ vs H_a: $\mu > \mu_0$

$$p(\mu_0) = P(\bar{X} > \bar{x}_n) = 1 - \Phi\left(\frac{\bar{x}_n - \mu_0}{1/\sqrt{n}}\right) = \Phi\left(\frac{\mu_0 - \bar{x}_n}{1/\sqrt{n}}\right)$$

$p-$value function $p(y) = \Phi\left(\frac{y - \bar{x}_n}{1/\sqrt{n}}\right)$ is a CD for μ

1. with a sample $\boldsymbol{x}$,

$$p(y) \to 0 \text{ as } y \to -\infty \text{ and } p(y) \to 1 \text{ as } y \to \infty$$

2. with $y = \mu_0$, $\Pr(p(\mu_0) \leq \alpha) = \Pr(\Phi\left(\frac{\mu_0 - \bar{x}_n}{1/\sqrt{n}}\right) \leq \alpha) = \alpha$

- suppose n $= 100$ and we observe $\bar{x}_n = 0.3$

μ_0	$-\infty$	-0.1	0	0.1	0.1355	0.2	0.3	0.496
p-value	0	0.00002	0.0014	0.022	0.05	0.16	0.5	0.975

Example 3: CD from normalizing likelihood

- Likelihood function

$$L(\mu|x_1,\ldots,x_n) = \prod f(x_i|\mu) = Ce^{-\frac{1}{2}(x_i-\mu)^2}$$
$$= Ce^{-\frac{n}{2}(\bar{x}_n-\mu)^2-\frac{1}{2}\sum(x_i-\bar{x}_n)^2}$$

- Normalized with respect to μ:

$$\frac{L(\mu|data)}{\int L(\mu|data)d\mu} = \cdots = \frac{1}{\sqrt{2\pi/n}}e^{-\frac{n}{2}(\bar{x}_n-\mu)^2}$$

  is a CD for μ

CD: Unifying Concept for Inference

- Wide range of examples (Xie and Singh 2013) :
  - bootstrap distribution
  - p-value functions
  - Bayesian posteriors
  - empirical likelihood
  - .....

- New methodology developments: e.g. Rutgers group
  - new prediction approaches
  - new testing methods
  - new simulation schemes
  - combining information from diverse sources through combining CDs (meta analysis, split & conquer, etc.)

Meta-analysis through Confidence Distributions

- Why combine CDs in meta-analysis?
  - informative: much more than a single point or an interval
  - broad: covering a broad range of examples
  - supported by statistical theory: e.g., ensuring frequentist coverage, etc

- Key points/steps:
  - summarize relevant data information using a CD in each study
  - synthesize information from diverse sources/studies via combination of the CDs from these studies

General framework on Combining CDs

- Suppose there are k independent studies to estimate θ - a common parameter of interest.
  - data $\boldsymbol{x_i}$ of the ith study $\overset{construct}{\Longrightarrow}$ CD - $H_i(\cdot) = H_i(\boldsymbol{x_i}, \cdot)$ for θ

- A general recipe for combining these k independent CDs:

$$H^{(c)}(\theta) = \Pr\{g_c(U_1, \ldots, U_k) \leq g_c(H_1(\theta), \ldots, H_k(\theta))\} \quad (1)$$

$H^{(c)}(\cdot)$:
  - a combined CD for θ the common true parameter
  - contains information from all k samples

$g_c(u_1, \ldots, u_k)$:
  - a non-decreasing and continuous function in each coordinate
    - $U_1, \ldots, U_k \overset{iid}{\sim} U[0, 1]$

Special Choices of $g(\cdot)$ Function

- Choice 1:

$$g_c(u_1, \ldots, u_k) = w_1 a_0(\mu_1) + \ldots + w_k a_0(\mu_k) \tag{2}$$

  - $a_0(\cdot)$: continuous and monotonic function
    - often $a_0(\cdot) = F_0^{-1}(\cdot)$ where $F_0(\cdot)$ is a CDF (i.e. $a_0 = \Phi^{-1}$)
  - w_i: generic weights (≥ 0) for the combination ($\sum w_i > 0$)
    - fixed: improve the efficiency of combination
    - adaptive: robust combination

- Choice 2:

$$g_c(u_1, \ldots, u_k) = w_1 \mu_1 + \ldots + w_k \mu_k \tag{3}$$

Case 1: Fisher p-value combination

- Conduct k clinical trials to study $H_0 : \theta \geq s$ vs $H_1 : \theta < s$
  - the p-value is a function of $s \Rightarrow$ p-value function
  - let $p_i(\cdot)$ be the p-value function for ith study $\Rightarrow$ a CD function

- Set $w_i = 1$ and $a_0(\cdot) = \log(\cdot)$ in CD combining recipe:

$$g_c(u_1, \ldots, u_k) = w_1 a_0(\mu_1) + \ldots + w_k a_0(\mu_k)$$

$$\begin{aligned}
\therefore H^{(c)}(s) &= \Pr\{g_c(U_1, \ldots, U_k) \leq g_c(p_1(s), \ldots, p_k(s))\} \\
&= \Pr\{\log U_1 + \ldots + \log U_k \leq \log p_1(s) + \ldots + \log p_k(s)\} \\
&= \Pr\{-2 \sum_{i=1}^{k} \log U_i \geq -2 \sum_{i=1}^{k} \log p_i(s)\} \\
&= \Pr\{\chi^2_{2k} \geq -2 \sum_{i=1}^{k} \log p_i(s)\} \qquad \Rightarrow \text{Fisher's method}
\end{aligned}$$

  - fact: $U_1, \ldots, U_k \overset{iid}{\sim} U[0,1] \Rightarrow -2 \sum_{i=1}^{k} \log U_i \sim \chi^2_{2k}$

Case 2: Stouffer (Normal) p-value Combination

- Use $g_c(u_1, \ldots, u_k) = \Phi^{-1}(u_1) + \ldots + \Phi^{-1}(u_k)$

$$\therefore H^{(c)}(s) = P\{g_c(U_1, \ldots, U_k) \leq g_c(p_1(s), \ldots, p_k(s))\}$$

$$= \Pr\left\{ \sum_{i=1}^{k} \Phi^{-1}(U_i) \leq \sum_{i=1}^{k} \Phi^{-1}(p_i(s)) \right\}$$

$$= \Pr\left\{ Z \leq \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \Phi^{-1}(p_i(s)) \right\} \quad (Z \sim N(0,1))$$

$$= \Phi\left\{ \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \Phi^{-1}(p_i(s)) \right\} \qquad \Rightarrow \text{Stouffer Method}$$

  - fact: $\Phi^{-1}(U_i) \sim N(0,1) \Rightarrow \sum \Phi^{-1}(U_i) \sim N(0,k)$

Case 3: Fixed-effects Meta-Analysis

- Let Y_i be a summary statistic with known variance s_i^2

$$Y_i \overset{iid}{\sim} N(\theta, s_i^2) \quad \text{for } i = 1, ..., k$$

  - CD function $H_i(\theta) = \Phi[(\theta - Y_i)/s_i]$

- Use $g_c(u_1, \ldots, u_k) = \Phi^{-1}(u_1)/s_1 + \ldots + \Phi^{-1}(u_k)/s_k$

$$\therefore H^{(c)}(\theta) = P\{g_c(U_1, \ldots, U_k) \leq g_c(H_1(\theta), \ldots, H_k(\theta))\}$$

$$= P\Big\{ \sum_{i=1}^{k} \Phi^{-1}(U_i)/s_i \leq \sum_{i=1}^{k} \Phi^{-1}(\Phi[(\theta - Y_i)/s_i])/s_i \Big\}$$

$$= \Phi\left(\left(\sum_{i=1}^{k} \frac{1}{s_i^2} \right)^{1/2} (\theta - \hat{\theta}_c) \right) \quad \Rightarrow \text{Normand 1999}$$

  - $\hat{\theta}_c = (\sum_{i=1}^{k} Y_i/s_i^2)/(\sum_{i=1}^{k} 1/s_i^2)$ - weighted average

** estimate θ by a normal CD $N(\hat{\theta}_c, (\sum_{i=1}^{k} 1/s_i^2)^{-1})$

Unifying Framework for Combining Information

A unifying platform for almost all existing methods of information combination

Classical approaches of combining p-values (from Marden, 1991)	Fisher method Stouffer (normal) method Tippett (min) method Max method Sum method
Model-based meta-analysis approaches (from Normand, 1999, Table IV)	Fixed-effects model: MLE method Fixed-effects model: Bayesian method Random-effects model: Method of moment Random-effects model: REML method Random-effects model: Bayesian method (normal prior on θ and fixed τ)
	(Xie, Singh & Strawderman, 2011)
Others (2×2 tables)	Tian et al (2009, *Biostat.*) approach of combining intervals (risk difference) Mantel-Haenszel (MH) method (odds ratio) Peto method (odds-ratio)
	(Yang et al, 2013; Yang, 2013)
Combining functions	Multiplication of likelihood functions Bayesian formula
	(Singh, Xie & Strawdermna 2005; Xie et al 2013)

Case 4: robust meta-analysis of large studies

- Setting: k studies of the same treatment
  - sample sizes $n_1, n_2, \ldots, n_k$ - all very large
  - true value of parameter $\theta_i^{(0)}$ for ith study (extension of fixed-effects models)

$$Y_i \overset{ind}{\sim} N(\theta_i^0, s_i^2) \quad \text{for } i = 1, 2, \ldots, k$$

  - parameter of interest $\theta_0 = median\{\theta_1^{(0)}, \ldots, \theta_k^{(0)}\}$
    - assume majority of studies with θ_0, others are outlying studies

- Key ideas (Xie et al 2011):
  - use asymptotic CD - $\Phi\left(\frac{\theta - \bar{Y}_i}{s_i/\sqrt{n}}\right)$ for each study
  - down-weight studies far from the majorities by using adaptive weights

Case 5: robust meta-analysis of large # of studies

- Setting: true parameters of studies come from a population with a mixture of two components (mixture mean θ_*):
  - a regular (or "good") part with mean of θ_0
  - a contaminated part (with a different distribution)

- Xie et al 2011 proposed a combined CD function:
  - using $H_i(\theta)$ - a CD function for θ_* in the mixed distribution

$$H^{(c)}(\theta) = \Phi\Big(\sum_{i=1}^{k} w_i (H_i(\theta) - 0.5) \Big/ \sqrt{k \sum_{i=1}^{k} w_i^2 / 12} \Big)$$

  - median of $H^{(c)}$: a robust and consistent estimator of θ_* distribution
    - protection against model mis-specification
  - extension of random-effects meta-analysis

Case 6: Exact Meta-Analysis for 2×2 Table

	events	**non-events**	**total**
treatment (drug)	x_i	$n_i - x_i$	n_i
control (placebo)	y_i	$m_i - y_i$	m_i

- Conventional meta-analysis - some with zero cell count
  - exclusion of tables with zero cell count
  - continuity correction
  - large sample approximation for distribution of summary statistic

- Exact meta-analysis (Liu et al 2014)
  - conduct an exact test study for each of the studies
  - obtain associated p-values
  - choose $a_0(\cdot)$ and weights to get a combined CD

  ** achieve asymptotical efficiency & enhance finite sample efficiency

Computing Package - gmeta

- R package for CD combining (Yang, Cheng and Xie 2015)
  - mimicing the structure of the glm() function in R
    - "family" in glm() $\Leftrightarrow$ "gc" in gmeta()
    - "link" in glm() $\Leftrightarrow$ "weights" in gmeta()
  - a plot function to visualize individual and combined CDs through extended forest plots

- Functionalities:
  - classical p-value combination methods (such as methods of Fisher, Stouffer, Tip- pett, etc.)
  - meta-anlaysis with fixed-effects and random-effects models
  - robust meta-analysis
  - exact meta-analysis

Computing Package - gmeta

- Available at:
  `http://cran.r-project.org/web/packages/gmeta`

- CRAN Page:

gmeta: Meta-Analysis via a Unified Framework of Confidence Distribution

An implementation of an all-in-one function for a wide range of meta-analysis problems. It contains a single function gmeta() that unifies all standard meta-analysis methods and also several newly developed ones under a framework of combining confidence distributions (CDs). Specifically, the package can perform classical p-value combination methods (such as methods of Fisher, Stouffer, Tippett, etc.), fit meta-analysis fixed-effect and random-effects models, and synthesizes 2x2 tables. Furthermore, it can perform robust meta-analysis, which provides protection against model-misspecifications, and limits the impact of any unknown outlying studies. In addition, the package implements two exact meta-analysis methods from synthesizing 2x2 tables with rare events (e.g., zero total event). A plot function to visualize individual and combined CDs through extended forest plots is also available.

Version:	2.2-6
Depends:	stats, BiasedUrn, binom
Published:	2016-02-23
Author:	Guang Yang, Jerry Q. Cheng, and Minge Xie
Maintainer:	Guang Yang <gyang.rutgers at gmail.com>
License:	GPL-2 | GPL-3 [expanded from: GPL (≥ 2)]
NeedsCompilation:	yes
In views:	MetaAnalysis
CRAN checks:	gmeta results

Arguments of gmeta()

```
gmeta <- function(gmi, gmi.type = c('pivot', 'cd', 'pvalue', '2x2'),
    method = c('fixed-mle', 'fixed-robust1', 'fixed-robust2',
    'fixed-robust2(sqrt12)', 'random-mm', 'random-reml', 'random-tau2',
    'random-robust1', 'random-robust2', 'random-robust2(sqrt12)',
    'fisher', 'normal', 'stouffer', 'min', 'tippett', 'max', 'sum',
    'MH', 'Mantel-Haenszel', 'Peto', 'exact1', 'exact2'),
    linkfunc = c('inverse-normal-cdf', 'inverse-laplace-cdf'),
    weight = NULL, study.names = NULL, gmo.xgrid = NULL, ci.level = 0.95,
    tau2 = NULL, mc.iteration = 10000, eta = 'Inf', verbose = FALSE,
    report.error = FALSE)
```

- key arguments
  - method - assumption for a meta-analysis model
  - linkfunc - how to handle information from individual studies
  - other
    - gmi - summary statistics
    - gmo.xgrid - range and grid points to evaluate a combined CD

Demonstration of gmeta - the Ulcer Data

- 41 randomized trials of a treatment for stomach ulcers (Efron 1996)

```
> library(gmeta)
> data(ulcer); ulcer.o <- as.matrix(ulcer)  # original data
> ulcer.o
```

	TrtEvent	TrtNonevent	CtrlEvent	CtrlNonevent
[1,]	7	8	11	2
[2,]	8	11	8	8
[3,]	5	29	4	35
[4,]	7	29	4	27
[5,]	3	9	0	12
[6,]	4	3	4	0
[7,]	4	13	13	11
...				
...				
[36,]	10	30	12	8
[37,]	3	13	2	14
[38,]	4	30	5	14
[39,]	7	31	15	22
[40,]	0	34	34	0
[41,]	0	9	0	16

Example 1: *p*-value combination

- make continuity correction
- calculate log odds ratios and their standard deviations
- calculate p-value for all studies

```
> # impute 0.5 to zero events
> ulcer <- ifelse(ulcer.o==0, 0.5, ulcer.o)

> # summary statistics
> ulcer.theta <- log( (ulcer[,1]*ulcer[,4]) / (ulcer[,2]*ulcer[,3]) )
> ulcer.sigma <- sqrt(1/ulcer[,1] + 1/ulcer[,2] + 1/ulcer[,3] + 1/ulcer[,4])

> # p-values from individual studies for H0: LOR >=0 vs. Ha: LOR < 0
> ulcer.pvalues <- 1 - pnorm(0, mean=ulcer.theta, sd=ulcer.sigma)

> ulcer.pvalues
[1] 2.364514e-02 3.204119e-01 7.170408e-01 7.631117e-01 9.045478e-01
[6] 1.435593e-01 2.786246e-02 2.963888e-04 2.499260e-01 1.713687e-03
........
........
[36] 5.038550e-03 6.858163e-01 9.294398e-02 1.960676e-02 1.402044e-05
[41] 6.108883e-01
```

Example 1: *p*-value combination - Stouffer & Tippet

```
> ## apply Stouffer (Normal) method ##
> gmo.pvalue1 <-  gmeta(ulcer.pvalues, gmi.type='pvalue', method='normal')
> print(gmo.pvalue1)

------------------------------------------------------------
        P-value combination through CD-Framework
Combine Method:    normal
Combined p-value:  1.10779e-16
------------------------------------------------------------


> ## apply Tippet method ##
> gmo.pvalue2 <-  gmeta(ulcer.pvalues, gmi.type='pvalue', method='tippet')
> print(gmo.pvalue2)

------------------------------------------------------------
        P-value combination through CD-Framework
Combine Method:    tippett
Combined p-value:  0.0005746768
------------------------------------------------------------
```

Example 1: *p*-value Combination - all 5 methods

```
> ## Table of the results from all p-value combination methods
> pvalue.combine.methods = c('fisher', 'stouffer', 'tippet', 'max', 'sum')
> combined.pvalue.vector = rep(NA, 5)
> for ( i in 1:5 ) {
>     combined.pvalue.vector[i] <- gmeta(ulcer.pvalues, gmi.type='pvalue',
>         method=pvalue.combine.methods[i])$cmbd.pvalue
> }
> mthds <- 'method'
> pvlus <- 'p-value'
> for ( i in 1:5) {
>     mthds = paste(mthds, pvalue.combine.methods[i], sep='\t\t& ')
>     pvlus = paste(pvlus, combined.pvalue.vector[i], sep='\t& ')
> }
```

p value combination results:

Method	Fisher	Stouffer	Tippet	Max	Sum
p-value	2.1684e-19	1.1078e-16	5.7468e-04	1.6357e-02	4.8591e-09

Model-based Meta-analysis

- *method* argument:
  - fixed-effects: method='fixed-mle'
  - random-effects:
    - method='random-mm' (DL estimator)
    - method='random-reml' (REML estimator)

- *linkfunc* argument - specifying $a_0(\cdot)$ function
  - default value - 'inverse-normal-cdf': $a_0(\cdot) = \Phi^{-1}(\cdot)$
  - other choice - 'inverse-laplace-cdf':

- *weight* argument - specifying study-specific weights
  - default NULL - assigning weights depending on linkfunc
    - inverse standard deviation weights when linkfunc='inverse-normal-cdf'
    - all ones when linkfunc='inverse-laplace-cdf'

Example 2: Fixed-effects Meta-analysis

```
> ## use summary statistics ##
> ulcer.theta <- log( (ulcer[,1]*ulcer[,4]) / (ulcer[,2]*ulcer[,3]) )
> ulcer.sigma <- sqrt(1/ulcer[,1] + 1/ulcer[,2] + 1/ulcer[,3] + 1/ulcer[,4])
> ulcer.pivots <- data.frame(mns=ulcer.theta, sds=ulcer.sigma)
> ulcer.pivots

-------------------------------------------
          mns       sds
1  -1.8382795 0.9266964
2  -0.3184537 0.6825753
3   0.4111958 0.7162780
....
....
40 -8.4390154 2.0146522
41  0.5753641 2.0429418
-------------------------------------------

> ## fixed-effect model - use default linkfunc='inverse-normal-cdf' ##
> ##    for Fisher efficiency                                        ##
> gmo.fx1 <- gmeta(ulcer.pivots, method='fixed-mle',
>         gmo.xgrid=seq(-20,20,by=0.001))
> summary(gmo.fx1)
```

Example 2: Fixed-effects Meta-analysis - cont'd

```
-------------------------------------------------------------------
Model-Based Meta-Analysis through CD-Framework

Call:
gmeta.default(gmi = ulcer.pivots, method = "fixed-mle",
    gmo.xgrid=seq(-20, 20, by=0.001))

Summary of Combined CD:
             mean        median      stddev      ci.lower    ci.upper
Combined CD -0.8875844  -0.8875844  0.1255535   -1.133665   -0.6415031
-------------------------------------------------------------------


> ## using DE link for combining with default weight - Bahadur efficiency ##
> gmo.fx2 <- gmeta(ulcer.pivots, method='fixed-mle',
>         + linkfunc='inverse-laplace-cdf', gmo.xgrid=seq(-20,20,by=0.001))
> summary(gmo.fx2)

-------------------------------------------------------------------
Summary of Combined CD:
             mean        median      stddev      ci.lower    ci.upper
Combined CD -1.063764   -1.064124   0.1061674   -1.271884   -0.8534409
-------------------------------------------------------------------
```

Example 3: Random-effects Meta-analysis

```
> ## random-effects meta-analysis, DL estimator for tau2 ##
> gmo.rd1 <- gmeta(ulcer.pivots, method='random-mm',
>     gmo.xgrid=seq(-20,20,by=0.001))
> summary(gmo.rd1)

-----------------------------------------------------------------------
                    Summary of Combined CD:
             mean         median      stddev       ci.lower    ci.upper
Combined CD -1.097574   -1.097574   0.2103224   -1.509798   -0.6853496
-----------------------------------------------------------------------


> ## random-effects meta-analysis, REML estimator for tau2
> gmo.rd2  <- gmeta(ulcer.pivots, method='random-reml',
>     gmo.xgrid=seq(-20,20,by=0.001))
> summary(gmo.rd2)

-----------------------------------------------------------------------
Summary of Combined CD:
             mean         median      stddev       ci.lower    ci.upper
Combined CD -1.091384   -1.091384   0.2069317   -1.496963   -0.6858053
-----------------------------------------------------------------------
```

Example 4: Robust Meta-analysis

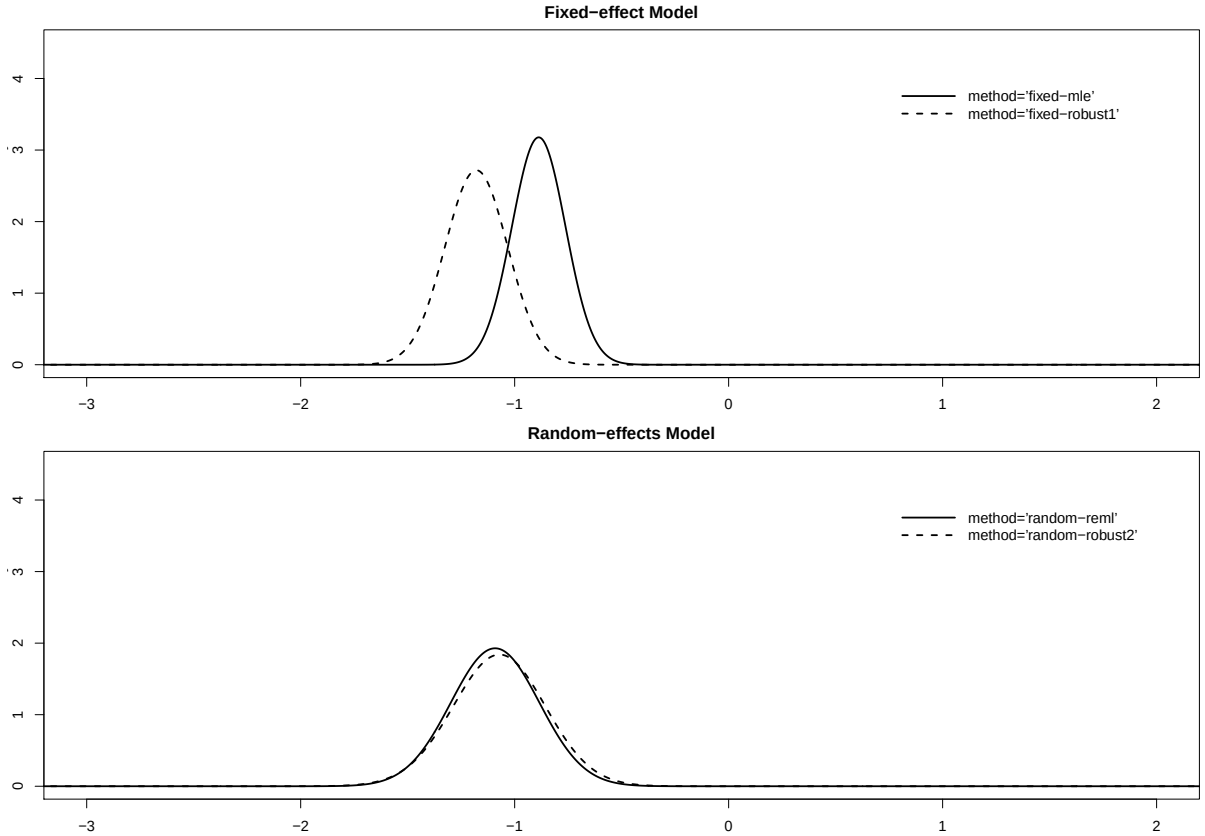
- *method* ='fixed-robust1': the large number of studies case

```
> gmo.robust1 <- gmeta(ulcer.pivots, method='fixed-robust1',
>     gmo.xgrid=seq(-20,20,by=0.001))
> summary(gmo.robust1)

-------------------------------------------------------------------
                    Summary of Combined CD:
            mean        median      stddev      ci.lower    ci.upper
Combined CD -1.179134   -1.179134   0.1467666   -1.466792   -0.8914757
-------------------------------------------------------------------
```

- *method* ='random-robust2': the large sample size case

```
> gmo.robust2 <- gmeta(ulcer.pivots, method='random-robust2',
>     gmo.xgrid=seq(-20,20,by=0.001))
> summary(gmo.robust2)

-------------------------------------------------------------------
Summary of Combined CD:
            mean        median      stddev      ci.lower    ci.upper
Combined CD -1.073812   -1.072808   0.2170723   -1.505937   -0.647466
-------------------------------------------------------------------
```

Example 4: Robust Meta-analysis - no outlying studies



- slight wider combined CDs from robust methods
- trade efficiency for robustness

Example 4: Robust Meta-analysis - robustness in play

- construct a contaminated data set with outlying studies
  - multiplying $\hat{\theta}$ in studies-05,13,14,22,35,41 by 10
- use random-effects meta-analysis and robust meta-analysis on
  - original data
  - contaminated data
- estimated θ and its 95% confidence interval

	original data	**contaminated data**
random-effects meta-analysis	-1.090(-1.497, -0.686)	-0.099(-0.504, 0.307)
robust meta-analysis	-1.074(-1.506, -0.648)	-1.010(-1.499, -0.509)

Example 5: Meta-analysis of 2×2 Tables

```
> ## data format (x, n, y, m) for 2 by 2 tables ##
> ulcer.2x2 <- cbind(ulcer.o[,1], ulcer.o[,1]+ulcer.o[,2],
>    ulcer.o[,3], ulcer.o[,3]+ulcer.o[,4])
```

- *method* ='MH' - Mantel-Haenszel method; Result: OR

```
> gmo.MH <- gmeta(ulcer.2x2dt, gmi.type='2x2', method='MH',
    gmo.xgrid=seq(-20,20,by=0.001))
> summary(gmo.MH)
-------------------------------------------------------------------
Summary of Combined CD:
             mean        median      stddev      ci.lower    ci.upper
Combined CD 0.3370037   0.3370037   0.03746689  0.26357     0.4104375
-------------------------------------------------------------------
```

- *method* ='Peto' - Peto method; Result: LOR

```
> gmo.Pt <- gmeta(ulcer.2x2d, gmi.type='2x2', method='Peto',
>    gmo.xgrid=seq(-20,20,by=0.001))
> summary(gmo.Pt)
-------------------------------------------------------------------
Summary of Combined CD:
             mean        median      stddev      ci.lower    ci.upper
Combined CD -1.132064   -1.132064   0.1093015   -1.346291   -0.9178366
-------------------------------------------------------------------
```

Example 5: Meta-analysis of 2×2 Tables

- *method* ='exact1' - Liu et al. 2015; Result: LOR
  - use significant functions based on mid-p adaption of Fisher's exact test for individual CD
  - adjustment for improvement in efficiency

```
> gm0.exact1 <- gmeta(ulcer.2x2d, gmi.type='2x2', method='exact1',
>     gmo.xgrid=seq(-20,20,by=0.001))
> summary(gmo.exact1)

-------------------------------------------------------------------
Summary of Combined CD:
             mean        median      stddev      ci.lower    ci.upper
Combined CD -1.239827   -1.240056   0.1270008   -1.485981   -0.9882136
-------------------------------------------------------------------
```

Concluding Remarks

- Meta-analysis (or fusion learning more generally) using CDs
  - provides an effective way to combine information from diverse sources
  - explores heterogeneous data

- This talk provided an illustration of some examples with gmeta demonstrations
  - p-value combining
  - model-based fixed-effects meta-analysis
  - model-based random-effects meta-analysis
  - robust meta-analysis
  - exact meta-analysis

Related Research: current and forthcoming

- Expand gmeta package
  - meta-analysis with fixed, unknown, and study-specific parameters (Clagett et al 2014)
    - no need to assume either fixed-effects or random-effects model
    - uses a novel resampling method via CDs the study-level parameters for inference
    - non-parametric approach
    - expected data of release: end of 2016
  - meta-analysis of heterogeneous studies (Liu et al 2015)
    - estimable parameters different from one study to another
    - use some mapping functions to link estimable parameters in different studies with the common parameter of interest
    - expected date of release: spring of 2017

Related Research: current and forthcoming - cont'd

- Analysis of recurrent event data (i.e. hospital readmission, relapse of a disease, etc)
  - data sources:
    - universal billing data from NJ hospitals from 1986 to 2015
    - NJ death records
    - (partial) NJ cancer register
    - .....
  - computing difficulties $\Rightarrow$ split-conquer-combine approach
    - use CD combining framework to aggregate results

- Individualized medicine studies of cardiovascular disease (i.e. heart failure)
  - build survival models for each patient
  - borrow strength from models of similar patients then use CD for combination

Acknowledgement

- Collaborators
  - John B. Kostis, Regina Liu, Minge Xie, Guang Yang

- Research grant support
  - NSF-SES1449958 and NSF-CNS1514224

Thank You!

Selected References

1. Cheng, J.Q., Liu, R., M. Xie, M. (2016) "Fusion learning," to appear in StatRefOnline.
2. Claggett, B., Xie, M., and Tian, L. (2014) "Meta-Analysis with Fixed, Unknown, Study-Specific Parameters," JASA, 109, 1660-1671.
3. Liu, D., Liu, R., and Xie, M. (2014) "Exact Meta-Analysis Approach for Discrete Data and its Application to 2 $\times$ 2 Tables With Rare Events," JASA, 109, 1450-1465.
4. Singh, K., Xie, M., and Strawderman, W.E. (2005) "Combining Information from Independent Sources through Confidence Distribution," The Annals of Statistics, 33, 159-183.
5. Schweder, T. and Hjort, N.L. (2002) "Confidence and Likelihood," Scandinavian Journal of Statistics, 309-332.
6. Whitehead, A. (2002) "Meta-Analysis of Controlled Clinical Trials, " London:Wiley.
7. Xie, M., Singh, K. and Strawderman, W.E. (2011) "Confidence Distribution and a Unifying Framework for Meta-Analysis," JASA, 106, 320-333.
8. Yang, G., Cheng, J.Q., and Xie, M. (2015) "gmeta: Meta-Aanalysis via a Unified Framework of Confidence Distribution," Available at https://cran.r-project.org/web/packages/gmeta/index.html.