Analysing response time distributions with the ex-Gaussian and shifted Wald

New Tricks Internal Seminar

Mark Hurlstone
Lancaster University
https://mark-hurlstone.github.io

June 3, 2021

Outline

- Response time (RT) as a dependent measure
- Problems with conventional analysis of mean RT
- Probability functions for representing RT distributions:
  - ex-Gaussian
  - Shifted Wald
- How to fit probability functions to RT data:
  - conceptual overview
  - RT-Distrib-Fit program for MATLAB
- Application example
- Caveats regarding interpretation of distribution parameters

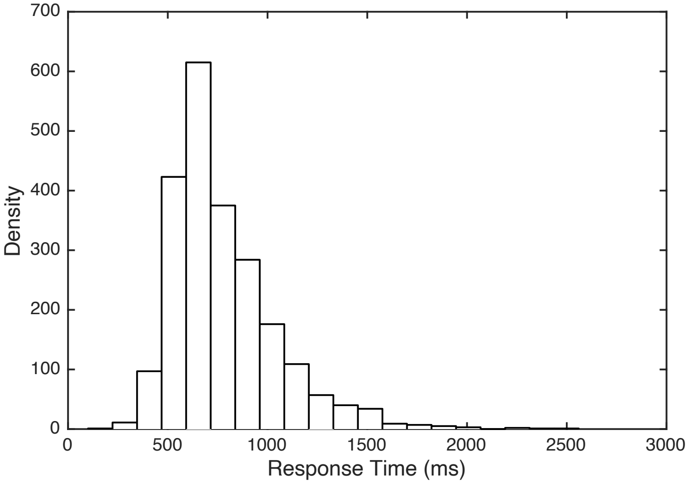
Response Times (RT)

- Response time (RT) is a popular dependent measure in cognitive psychology
- Long history and rich tradition (e.g., Donders, 1868; Luce, 1986; Ratcliff, 1978; Sternberg, 1969; Wundt, 1880)
- Spawned several sophisticated sequential sampling models of choice RT (e.g., Ratcliff, 1978; Brown & Heathcote, 2005, 2008; Logan et al., 2014; Usher & McClelland, 2001)
- Detailed analyses of empirical RTs provide powerful constraints for choosing between cognitive models (Farrell & Lewandowsky, 2004; Hurlstone & Hitch, 2015, 2018)

Mean RT (M_{RT})

- In empirical studies of RT, mean RT, M_{RT}, is the dominant measure of performance
- Faster M_{RT} in condition A than condition B implies more efficient cognitive processing in condition A
- Thus, M_{RT} is a measure of performance—lower M_{RT} implies better performance
- However, some researchers have abandoned this approach in favour of more detailed distributional analyses
- Motivated by many problems associated with analysis of M_{RT} (e.g., Heathcote et al., 1991)

Problems With M_{RT}: Skewed Data



Problems With M_{RT}: Skewed Data

- Two possible implications of skewed data (Heathcote et al., 1991):
  1. the cognitive process of interest yields skewed data
  2. the cognitive process of interest yields symmetrical data—skew reflects nuisance variables
- If (1) is true, then an analysis of distribution shape, not M_{RT}, is required
- If (2) is true, then nuisance scores must be removed
- Most researchers assume (2) is true and trim or transform their data

Problems With M_{RT}

1. Data trimming
2. Data transformation
3. Data representation

Problems With M_{RT}

1. Data trimming
2. Data transformation
3. Data representation

Problems With M_{RT}: Data Trimming

- Addressing skew by trimming data to eliminate extreme (presumed to be nuisance) values
  - removing trials with an RT above a fixed value
  - removing trials with an RT more than a fixed number of standard deviations (typically 2.5) from the mean
- Creates a distribution closer to normal
- It is reasonable to trim post-error trials and when a participant is known to have been distracted
- But trimming on basis of a trial's value is a brutal response to managing skew that "risks throwing the baby out with the bath water" (Heathcote et al., 1991, p.341)

Problems With M_{RT}

1. Data trimming
2. Data transformation
3. Data representation

Problems With M_{RT}

1. Data trimming
2. Data transformation
3. Data representation

Problems With M_{RT}: Data Transformation

- Addressing skew using a data transformation (e.g., logarithm of RT)
- Normalises the distribution by discounting extreme (presumed to be nuisance) values
- If skew is produced by a nuisance process, discounting must be done in proportion to the N data points produced by that process
- Failure to do so means rescaled data may misrepresent the cognitive process of interest
- Transformations are misleading and discard valuable information

Problems With M_{RT}

1. Data trimming
2. Data transformation
3. Data representation

Problems With M_{RT}

1. Data trimming
2. Data transformation
3. Data representation

Problems With M_{RT}: Data Representation

- When a distribution is skewed, the mean misrepresents central tendency—it gives extreme values too much weight
- A partial solution is to use median RT instead
- But when data are skewed, the mean, median, and mode do not converge—the concept of central tendency is ambiguous
- Central tendency is only meaningful for symmetrical distributions
- The analysis of means is misleading

Statistical Models

- The analysis of entire distributions of RT solves the problems with M_{RT}
- Preserves all information and provides clear description of behaviour
- Avoids mischaracterising central tendency
- Can detect changes across manipulations not possible with M_{RT} (e.g., an increase in skew or a shift in the distribution)
- But we need a statistical model to describe the distribution

Statistical Models

- The two statistical models that have proved most popular are:
  1. ex-Gaussian
  2. Shifted Wald

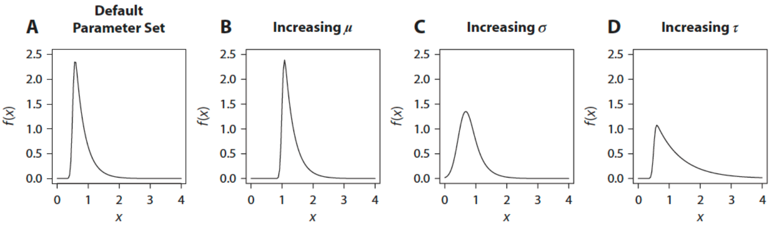
Statistical Models

- The two statistical models that have proved most popular are:
  1. ex-Gaussian
  2. Shifted Wald

ex-Gaussian

- Convolution of a Gaussian and an exponential distribution
- Three parameters:
  - μ and σ, the mean and standard deviation of the Gaussian component
  - τ, the mean of the exponential component
- Roughly, μ and σ reflect the leading edge of the distribution
- τ reflects the upper tail
- Has a positively skewed unimodal shape
- Provides excellent fit to RT distributions

ex-Gaussian



Taken from Matzke & Wagenmakers (2009)

ex-Gaussian

- The probability density function (henceforth, 'PDF') of the ex-Gaussian is given by:

$$f(x|\tau,\mu,\sigma) = \frac{1}{\tau}\exp\left(\frac{\mu}{\tau} + \frac{\sigma^2}{2\tau^2} - \frac{x}{\tau}\right)\Phi\left(\frac{x-\mu-\sigma^2/\tau}{\sigma}\right), \quad (1)$$

- where Φ is the cumulative density of the Gaussian component
- Its mean and variance are:

$$E(x) = \mu + \tau \quad (2)$$

and

$$\mathrm{Var}(x) = \sigma^2 + \tau^2 \quad (3)$$

ex-Gaussian

- Hohle (1965) proposed the ex-Gaussian reflects the duration of two successive components of cognitive processing
- Gaussian component reflects “the time required for organization and execution of the motor response” (transduction component)
- Exponential component reflects “the decision and perceptual portion of an RT” (decision component)
- This interpretation has been challenged repeatedly (see e.g., Luce, 1965; Matzke & Wagenmakers, 2009; McGill & Gibbon, 1965)

ex-Gaussian

- The ex-Gaussian does not have a plausible theoretical rationale
- The Gaussian component assigns positive probabilities to negative RTs
- It does not correspond to a plausible cognitive process model
- “Although the ex-Gaussian model describes RT data successfully, it does so without the benefit of an underlying theory” (Heathcote et al., 1991, p.346)

Statistical Models

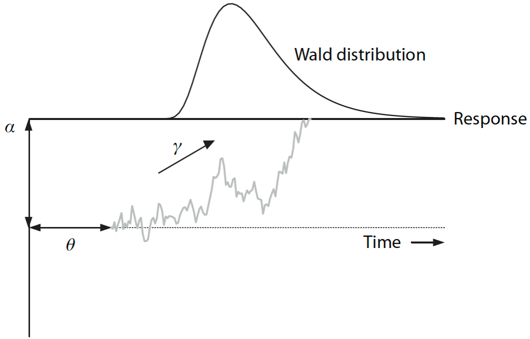
- The two statistical models that have proved most popular are:
  1. ex-Gaussian
  2. Shifted Wald

Statistical Models

- The two statistical models that have proved most popular are:
  1. ex-Gaussian
  2. Shifted Wald

Shifted Wald

- The Wald (1947) distribution is the finishing time distribution of a Wiener diffusion process towards a boundary

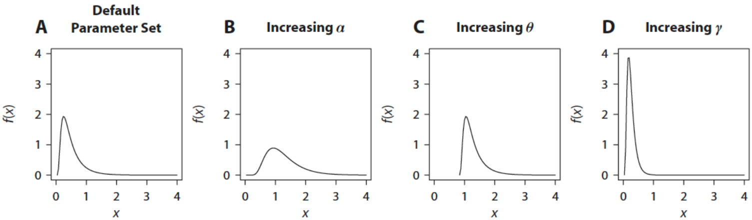


Taken from Matzke & Wagenmakers (2009)

Shifted Wald

- The Wald distribution has two parameters:
  - γ, reflecting the drift rate of the diffusion process
  - α, reflecting the separation between the diffusion starting point and boundary
- In the RT context, a third parameter, θ is included that shifts the location of the distribution
- Has a positively skewed unimodal shape
- Provides excellent fit to RT distributions

Shifted Wald



Taken from Matzke & Wagenmakers (2009)

Shifted Wald

- The PDF of the shifted Wald is given by:

$$f(x|\alpha,\theta,\gamma) = \frac{\alpha}{\sqrt{2\pi(x-\theta)^3}} \exp\left(-\frac{[\alpha-\gamma(x-\theta)]^2}{2(x-\theta)}\right), \quad (4)$$

- where $x > \theta$, its mean and variance are:

$$E(x) = \theta + \alpha/\gamma \quad (5)$$

and

$$\mathrm{Var}(x) = \alpha/\gamma^3 \quad (6)$$

Shifted Wald

- Shifted Wald has a cognitive interpretation
- People accumulate noisy information from the environment until a threshold amount is reached and a response initiated
- Drift rate γ reflects task difficulty or participant ability
- Response criterion α reflects response caution
- Shift parameter θ reflects nondecision time

Fitting the ex-Gaussian and Shifted Wald to RT Distributions

Next ...

- Guide to how to fit probability functions to RT distributions
- RT-Ditrib-FIT: a MATLAB toolbox for fitting the ex-Gaussian and shifted Wald
- https://github.com/mark-hurlstone/RT-Distrib-Fit
- R toolbox forthcoming

Fitting the ex-Gaussian and Shifted Wald to RT Distributions

- Fitting probability functions to RT distributions requires at least three functions:
  1. functions implementing the ex-Gaussian PDF and shifted Wald PDF
  2. a function implementing the computation of the objective function
  3. a search algorithm to find best-fitting parameter values

Fitting the ex-Gaussian and Shifted Wald to RT Distributions

- Fitting probability functions to RT distributions requires at least three functions:
  1. functions implementing the ex-Gaussian PDF and shifted Wald PDF
  2. a function implementing the computation of the objective function
  3. a search algorithm to find best-fitting parameter values

Distribution Functions

- Functions are required for implementing the ex-Gaussian (equation 1), and shifted Wald (equation 4) PDFs
- RT-Ditrib-Fit contains two PDF functions:
  - f = exGaussPdf(parms,x)
  - f = shiftWaldPdf(parms,x)
- where f returns the PDF of the relevant distribution, parms is a vector of distribution paramater values (τ,μ,σ | α,θ,γ), and x is a data vector of empirical RTs

Fitting the ex-Gaussian and Shifted Wald to RT Distributions

- Fitting probability functions to RT distributions requires at least three functions:
  1. functions implementing the ex-Gaussian PDF and shifted Wald PDF
  2. a function implementing the computation of the objective function
  3. a search algorithm to find best-fitting parameter values

Fitting the ex-Gaussian and Shifted Wald to RT Distributions

- Fitting probability functions to RT distributions requires at least three functions:
  1. functions implementing the ex-Gaussian PDF and shifted Wald PDF
  2. a function implementing the computation of the objective function
  3. a search algorithm to find best-fitting parameter values

Objective Function

- An objective function is required that returns the goodness-of-fit of the theoretical PDF— given the supplied parameters—to the empirical data
- Several possibilities:
  - chi-square goodness-of-fit (Smith, 1995)
  - continuous maximum likelihood (Heathcote, 1991)
  - quantile maximum probability (Brown & Heathcote, 2003)
- RT-Distrib-Fit uses continuous maximum likelihood estimation

Likelihood Function

- Given a PDF $f(x|\theta)$ with k parameters, $\theta = [\theta_1, \theta_2, ..., \theta_k]$ and a set of data containing N observations, x_i, $i = 1, ..., N$, the likelihood function is:

$$L(\theta|x) = \prod_{i=1}^{N} f(x_i|\theta), \tag{7}$$

- where $\prod$ is the product operator
- **Problem:** can return values close to zero producing overflow errors

Log-Likelihood Function

- Overflow errors can be avoided by using the log of the likelihood
- Substitutes the sum operator with the product operator, which is less likely to produce overflow errors:

$$\ln L(\theta|x) = -\sum_{i=1}^{N} \ln\left[f(x_i|\theta)\right], \quad (8)$$

- where ln is the natural logarithm
- Search algorithms (next) typically use minimisation procedures, so it is customary to minimise the negative log-likelihood instead of maximising the log-likelihood

Log-Likelihood Function

- RT-Distrib-Fit computes the log-likelihood via the function, lnL = logMaxLikelihood(parms)
- Nested within the function wrapperLoopFmin, described next

Search Algorithm

- A search algorithm is needed that systematically adjusts the parameters of the to-be-fitted probability distribution to minimise the objective function (maximise the log-likelihood)
- The SIMPLEX algorithm (Nelder & Mead, 1965) is a robust and widely used parameter estimation method
- Invoked in MATLAB using the inbuilt fminsearch function (invoked via the optim function in R)
- RT-Distrib-Fit uses the function fminSearchBnd—version of SIMPLEX with reflection boundaries for to-be-estimated parameters

Search Algorithm

- SIMPLEX algorithm requires starting parameter values to initiate the search
- The closer these starting points are to the true parameter values, the better the performance of SIMPLEX
- Heuristic starting points are available for the ex-Gaussian (Lacouture & Cousineau, 2008)
- Sensible starting parameters can also be found for the shifted Wald (Heathcote, 2004)
- To avoid local minima problems, it is imperative that the search is conducted with multiple starting parameter values

Search Algorithm

- RT-Distrib-Fit contains a function that generates starting parameter values and reflection boundaries for parameters
- [startVec1,startVec2,startVec3,lB,uB] = ... genStartingParameters(data,chooseDistrib)
- where data is a data vector of empirical RTs and chooseDistrib is the distribution being fitted (0 = ex-Gaussian, 1 = shifted Wald)
- startvec1, startVec2, and startVec3 are vectors of starting parameter values (τ,μ,σ | α,θ,γ), and lB and UB are vectors of lower and upper boundaries on parameter values (τ,μ,σ | α,θ,γ)

Search Algorithm

- Yields three starting values for each distribution parameter
- Starting values and reflection boundaries are input arguments to another function wrapperLoopFmin
- bestX = wrapper-LoopFmin(parms,data,startVec1,startVec2,startVec3,lB,uB, chooseDistrib)
- Runs SIMPLEX with 27 different starting parameter combinations
- Returns bestX, a vector of the best-fitting parameter estimates (τ,μ,σ | α,θ,γ)

Using RT-Distrib-Fit

- A front-end script, rtDistribFitScript, controls the fitting
- To-be-fitted data should be stored as text files in the RT-Distrib-Fit MATLAB directory
- Naming convention: Participant_1.txt, Participant_2.txt, Participant_3.txt ...
- Each row represents an RT, each column represents a condition
- Choose what distribution you want to fit by setting the parameter chooseDistrib (0 = exGaussian, 1 = shifted Wald) then hit F5 to run

Using RT-Distrib-Fit

1. Retrieves data for participant p and condition c
2. Sort's the data and removes any missing values (coded as NaN in data files)
3. Generates starting parameter values based on the participant's data
4. Fits the data using 27 different starting points
5. Records best-fitting parameters, lnL, χ^2, and KS tests of data RT distributions
6. Iterate until all participant data has been fit
7. z-Transform participant RTs, rescale, then fit group distribution (see Rouder, 2014; cf. Vincent averaging)
8. Results written to text files (last row is group fit)
9. Generate histogram plots with best-fitting PDF overlaid

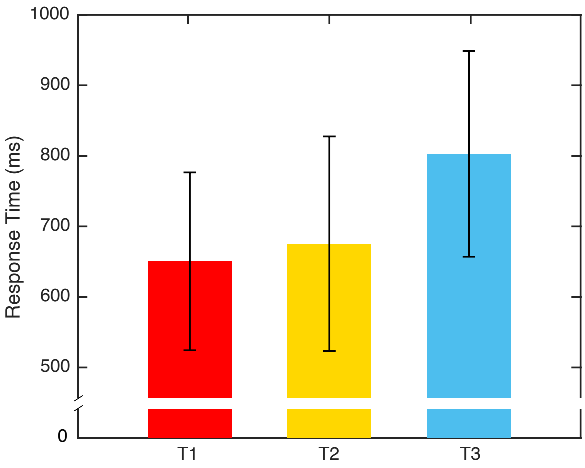
Some Considerations

- You need at least 100 RT observations per condition to obtain stable maximum likelihood estimates (Heathcote et al., 1991)
- Ignore fits to individual participants if there are less than 100 observations each—use group fits instead
- Reminder: these are contained in the final row of the output files

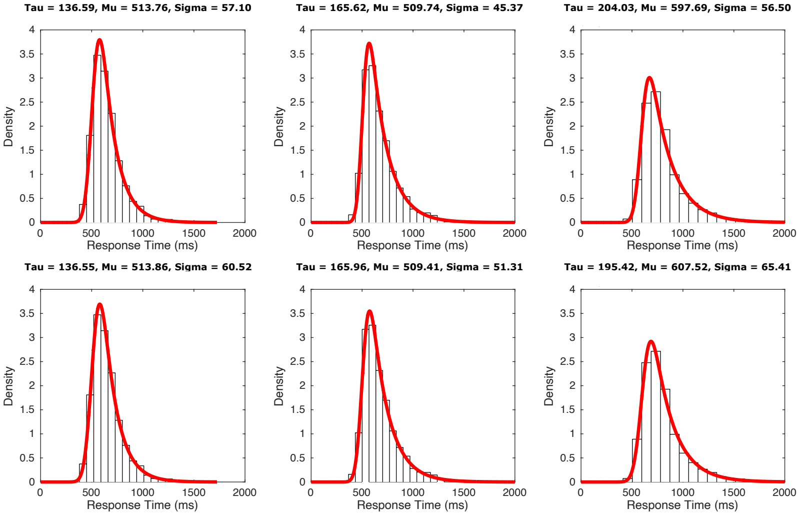
Application Example

- Artificial RT data set generated using shifted log-normal distribution (for lack of simple RT data set)
- Another distribution that provides an excellent fit to empirical RT distributions (Ratcliff & Murdock, 1976)
- 15 artificial participants, 3 treatments, 150 RTs each
- Uniform random sampling of parameters (μ, σ, and θ) with different expected values across treatments
- Fit ex-Gaussian and shifted Wald to resulting RT distributions

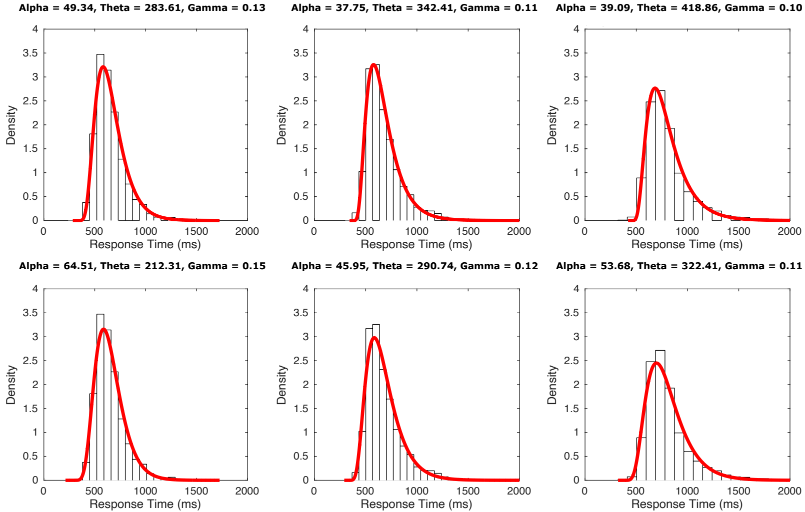
M_{RT} By Treatment



PDF Histograms With Fitted ex-Gaussian Functions



PDF Histograms With Fitted Shifted Wald Functions



Chi-Square Goodness-of-Fits

	Distribution					
	ex-Gaussian			Shifted Wald		
Fit Type	T_1	T_2	T_3	T_1	T_2	T_3
Individuals	**18.76**	**14.49**	27.47	20.99	16.45	**21.97**
Group	**19.55**	**12.22**	**30.60**	68.88	72.33	99.24

Summary

- Fits of ex-Gaussian and shifted Wald show the (hypothetical) manipulation caused an increase in skew of RT distribution, but not a shift in location
- Not discernible from analysis of M_{RT}
- Both ex-Gaussian and shifted Wald provided excellent fits to individual participant RT distributions
- ex-Gaussian also provided an excellent fit to group data, whereas shifted Wald performed less well
- Parameter averaging recommended for both distributions (shifted Wald perhaps more so) where possible (cf. Rouder & Speckman, 2004)

How To Report a Distributional Analysis

- Conventional analysis of M_{RT}
- Distributional analysis:
  - KS tests of empirical RT distributions with example density histograms for one or more participants
  - Group density histograms per condition, with best fitting probability function overlaid
  - Table or plot of estimated distribution parameters by condition
  - Inferential statistics (e.g., ANOVA) performed on distribution parameters (for fits to individual participants)
- See Heathcote et al. (1991) for guidelines and an example

Caveats

- Should you interpret changes in distribution parameters in terms of components of cognitive processing?
- Matzke and Wagenmakers (2009) fit the Ratcliff (1978) diffusion model to ex-Gaussian and shifted Wald probability distributions
- Diffusion model contains parameters that are known to map onto specific cognitive processes
- If ex-Gaussian and shifted Wald parameters represent components of cognitive processing, they should relate to parameters of the diffusion model

Caveats

- Matzke and Wagenmakers (2009) find no one-to-one mapping of diffusion model parameter estimates with ex-Gaussian and shifted Wald parameter values
- *We conclude that researchers should resist the temptation to interpret changes in the ex-Gaussian and shifted Wald parameters in terms of cognitive processes* (Matzke & Wagenmakers, 2009, p.798)
- Use these distributions as descriptive, rather than inferential, tools

Fin!

- Thanks for listening!

Recommended Reading and References

- To be added shortly ...