The drift-diffusion model of decision making

Ariel Zylberberg & Max J. Pensack
Columbia University

The drift-diffusion model of decision making

Ariel Zylberberg & Max J. Pensack
Columbia University

Code used in this tutorial:

https://github.com/arielzylberberg/CompPsychCourse_Zurich2022

The drift-diffusion model of decision making

Ariel Zylberberg & Max J. Pensack
Columbia University

Code used in this tutorial:

https://github.com/arielzylberberg/CompPsychCourse_Zurich2022

Please interrupt! Ask questions!

Speed, accuracy and difficulty

Speed, accuracy and difficulty

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

» Easy

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

- Easy
- Fast

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

- Easy
- Fast
- Consistent / accurate

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

» More difficult

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

- More difficult
- Slower

Speed, accuracy and difficulty

The garbage plate, Rochester, NY

- More difficult
- Slower
- Less consistent / accurate

Typical decision-making tasks

Typical decision-making tasks

Which snack do you prefer?

Typical decision-making tasks

Which snack do you prefer?

Word or non-word?

piggle

Typical decision-making tasks

Which snack do you prefer?

Word or non-word?

piggle

Which line is longer?

Typical decision-making tasks

Which snack do you prefer?

Word or non-word?

piggle

Which line is longer?

Which display has more dots?

Thought experiment: line length comparison

Model based on the Signal-detection Theory

Model based on the Signal-detection Theory

Model based on the Signal-detection Theory

Model based on the Signal-detection Theory

Model based on the Signal-detection Theory

Model based on the Signal-detection Theory

Model based on the Signal-detection Theory

Model based on the Signal-detection Theory

Model based on the Signal-detection Theory

Matlab exercise:

If the left stimulus is 25 cm, the right one is 35 cm,
and σ = 10 cm, what is the probability of getting the decision correct?

The difference between two normally-distributed samples is also normally distributed, with parameters:

$\mu_{diff} = \mu_{right} - \mu_{left}$, and $\sigma^2_{diff} = 2\sigma^2$

The difference between two normally-distributed samples is also normally distributed, with parameters:

$\mu_{diff} = \mu_{right} - \mu_{left}$, and $\sigma^2_{diff} = 2\sigma^2$

Evidence:

$$e \sim \mathcal{N}(\mu_{diff}, \sigma_{diff})$$

The difference between two normally-distributed samples is also normally distributed, with parameters:

$\mu_{diff} = \mu_{right} - \mu_{left}$, and $\sigma^2_{diff} = 2\sigma^2$

Evidence:

$$e \sim \mathcal{N}(\mu_{diff}, \sigma_{diff})$$

Decision rule:

$$\text{Choice} = \begin{cases} \text{Right}, & \text{if } e > 0 \\ \text{Left}, & \text{otherwise} \end{cases}$$

The difference between two normally-distributed samples is also normally distributed, with parameters:

$$\mu_{diff} = \mu_{right} - \mu_{left} \quad \text{, and} \quad \sigma^2_{diff} = 2\sigma^2$$

Evidence:

$$e \sim \mathcal{N}(\mu_{diff}, \sigma_{diff})$$

Decision rule:

$$\text{Choice} = \begin{cases} \text{Right}, & \text{if } e > 0 \\ \text{Left}, & \text{otherwise} \end{cases}$$

The difference between two normally-distributed samples is also normally distributed, with parameters:

$$\mu_{diff} = \mu_{right} - \mu_{left} \quad \text{, and} \quad \sigma^2_{diff} = 2\sigma^2$$

Evidence:

$$e \sim \mathcal{N}(\mu_{diff}, \sigma_{diff})$$

Decision rule:

$$\text{Choice} = \begin{cases} \text{Right}, & \text{if } e > 0 \\ \text{Left}, & \text{otherwise} \end{cases}$$

The difference between two normally-distributed samples is also normally distributed, with parameters:

$$\mu_{diff} = \mu_{right} - \mu_{left} \quad \text{, and} \quad \sigma^2_{diff} = 2\sigma^2$$

Evidence:

$$e \sim \mathcal{N}(\mu_{diff}, \sigma_{diff})$$

Decision rule:

$$\text{Choice} = \begin{cases} \text{Right}, & \text{if } e > 0 \\ \text{Left}, & \text{otherwise} \end{cases}$$

Signal-detection theory cannot explain why some decisions take longer than other ones

Random dot motion discrimination task

Random dot motion discrimination task

Random dot motion discrimination task

Random dot motion discrimination task

(More difficult)

Random dot motion discrimination task

(More difficult)

Random dot motion discrimination task

(More difficult)

Random dot motion discrimination task

(More difficult)

Speed-accuracy tradeoff

Hanks et al. eLife, 2014

Speed-accuracy tradeoff

Hanks et al. eLife, 2014

Speed-accuracy tradeoff

Hanks et al. eLife, 2014

Speed-accuracy tradeoff

Hanks et al. eLife, 2014

Mulder et al.
Att Percept Psych
2013

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

» Decisions may take a long time if high levels of accuracy are required.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

- Decisions may take a long time if high levels of accuracy are required.
- Decision rule is sub-optimal.

Opponent process theory of discrimination

Cartwright and Festinger (1943), “A quantitative theory of decision”.

- Decisions may take a long time if high levels of accuracy are required.
- Decision rule is sub-optimal.

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Weight of evidence e *in favor of* h_1 *over* h_2 =

$$\text{Log}\left[\frac{\textit{Probability of } e \textit{ if } h_1 \textit{ is true}}{\textit{Probability of } e \textit{ if } h_2 \textit{ is true}}\right]$$

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Weight of evidence e *in favor of* h_1 *over* h_2 =

$$\text{Log}\left[\frac{\textit{Probability of } e \textit{ if } h_1 \textit{ is true}}{\textit{Probability of } e \textit{ if } h_2 \textit{ is true}}\right]$$

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Weight of evidence e *in favor of* h_1 *over* h_2 =

$$\text{Log}\left[\frac{\textit{Probability of } e \textit{ if } h_1 \textit{ is true}}{\textit{Probability of } e \textit{ if } h_2 \textit{ is true}}\right]$$

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Weight of evidence e *in favor of* h_1 *over* h_2 =

$$\text{Log}\left[\frac{\textit{Probability of } e \textit{ if } h_1 \textit{ is true}}{\textit{Probability of } e \textit{ if } h_2 \textit{ is true}}\right]$$

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Weight of evidence e *in favor of* h_1 *over* h_2 =

$$\text{Log}\left[\frac{\textit{Probability of } e \textit{ if } h_1 \textit{ is true}}{\textit{Probability of } e \textit{ if } h_2 \textit{ is true}}\right]$$

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Weight of evidence e *in favor of* h_1 *over* h_2 =

$$\text{Log}\left[\frac{\textit{Probability of } e \textit{ if } h_1 \textit{ is true}}{\textit{Probability of } e \textit{ if } h_2 \textit{ is true}}\right]$$

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Weight of evidence e *in favor of* h_1 *over* h_2 =

$$\text{Log}\left[\frac{\textit{Probability of } e \textit{ if } h_1 \textit{ is true}}{\textit{Probability of } e \textit{ if } h_2 \textit{ is true}}\right]$$

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Weight of evidence e *in favor of* h_1 *over* h_2 =

$$\text{Log}\left[\frac{\textit{Probability of } e \textit{ if } h_1 \textit{ is true}}{\textit{Probability of } e \textit{ if } h_2 \textit{ is true}}\right]$$

$$B \approx \log\left(\frac{\text{acc}}{1-\text{acc}}\right)$$

bound height

required accuracy

Sequential probability ratio test (SPRT)

Wald & Wolfowitz (1948), A. Turing (194?)

Weight of evidence e *in favor of* h_1 *over* h_2 =

$$\text{Log}\left[\frac{\textit{Probability of } e \textit{ if } h_1 \textit{ is true}}{\textit{Probability of } e \textit{ if } h_2 \textit{ is true}}\right]$$

$$B \approx \log\left(\frac{\text{acc}}{1-\text{acc}}\right)$$

bound height

required accuracy

Many variants of evidence accumulation models for binary decisions

Ratcliff et al. TiCS 2016

Many variants of evidence accumulation models for binary decisions

Ratcliff et al. TiCS 2016

» Time (continuous or discrete)

Many variants of evidence accumulation models for binary decisions

Ratcliff et al. TiCS 2016

- Time (continuous or discrete)
- Momentary evidence (continuous or discrete)

Many variants of evidence accumulation models for binary decisions

Ratcliff et al. TiCS 2016

- Time (continuous or discrete)
- Momentary evidence (continuous or discrete)
- Perfect or “leaky” accumulation

Many variants of evidence accumulation models for binary decisions

Ratcliff et al. TiCS 2016

- Time (continuous or discrete)
- Momentary evidence (continuous or discrete)
- Perfect or “leaky” accumulation
- Other factors (number of accumulation processes, with or without inhibition, …)

Many variants of evidence accumulation models for binary decisions

Ratcliff et al. TiCS 2016

- Time (continuous or discrete)
- Momentary evidence (continuous or discrete)
- Perfect or “leaky” accumulation
- Other factors (number of accumulation processes, with or without inhibition, …)

Drift diffusion model

We will simulate this in Matlab

Roger Ratcliff, 1978

Drift diffusion model

We will simulate this in Matlab

» Decision based on the accumulation of samples of evidence arriving sequentially over time

Roger Ratcliff, 1978

Drift diffusion model

We will simulate this in Matlab

» Decision based on the accumulation of samples of evidence arriving sequentially over time

decision variable

$$x_{i+1} = x_i + e_i$$

i: time step

evidence

Roger Ratcliff, 1978

Drift diffusion model

We will simulate this in Matlab

» Decision based on the accumulation of samples of evidence arriving sequentially over time

decision variable

$$x_{i+1} = x_i + e_i$$

i: time step

evidence

» Evidence samples e_i are normally distributed

time step

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

drift coefficient

diffusion coefficient

Roger Ratcliff, 1978

Drift diffusion model

We will simulate this in Matlab

» Decision based on the accumulation of samples of evidence arriving sequentially over time

decision variable

$$x_{i+1} = x_i + e_i$$

evidence

i: time step

» Evidence samples e_i are normally distributed

time step

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

drift coefficient

diffusion coefficient

Roger Ratcliff, 1978

Drift diffusion model

We will simulate this in Matlab

» Decision based on the accumulation of samples of evidence arriving sequentially over time

decision variable

$$x_{i+1} = x_i + e_i$$

i: time step

evidence

» Evidence samples e_i are normally distributed

time step

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

drift coefficient

diffusion coefficient

» In the limit $\Delta t \to 0$, the decision process is described by a Wiener process with drift:

decision variable

$$\frac{dx(t)}{dt} = \underbrace{\mu}_{\text{signal}} + \underbrace{\sigma W(t)}_{\text{noise}}$$

Roger Ratcliff, 1978

Hands on: Simulate the drift-diffusion process (no bounds for now)

decision variable

$$x_{i+1} = x_i + e_i$$

i: time step

evidence

time step

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

drift coefficient

diffusion coefficient

Hands on: Simulate the drift-diffusion process (no bounds for now)

decision variable

$$x_{i+1} = x_i + e_i$$

evidence

i: time step

time step

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

drift coefficient

diffusion coefficient

Parameters:

- NumTrials = 5000
- NumTimeSteps = 1000
- $\Delta t = 0.001$ seconds
- $\sigma = 1$
- $\mu = 5$

Hands on: Simulate the drift-diffusion process (no bounds for now)

decision variable

$$x_{i+1} = x_i + e_i$$

evidence

i: time step

time step

$$e_i \sim \mathcal{N}(\mu\Delta t, \sigma^2\Delta t)$$

drift coefficient

diffusion coefficient

Parameters:

- NumTrials = 5000
- NumTimeSteps = 1000
- $\Delta t = 0.001$ seconds
- $\sigma = 1$
- $\mu = 5$

After t second of unstopped accumulation

$$E[x] = \mu t$$

$$Var[x] = \sigma^2 t$$

Simulate diffusion paths and verify this relationship

Hands on: Simulate the drift-diffusion process (no bounds for now)

decision variable

$$x_{i+1} = x_i + e_i$$

evidence

i: time step

time step

$$e_i \sim \mathcal{N}(\mu\Delta t, \sigma^2\Delta t)$$

drift coefficient

diffusion coefficient

Parameters:

- NumTrials = 5000
- NumTimeSteps = 1000
- $\Delta t = 0.001$ seconds
- $\sigma = 1$
- $\mu = 5$

Key Matlab code:

e = $\mu\Delta t$ + $\sigma\sqrt{\Delta t}$ randn(NumTrials, NumTimeSteps)

dv = cumsum(e,2)

Simulate diffusion paths and verify this relationship

The state of the decision variable remains Gaussian over time

$$E[x] = \mu t$$

$$Var[x] = \sigma^2 t$$

The state of the decision variable remains Gaussian over time

$$E[x] = \mu t$$

$$Var[x] = \sigma^2 t$$

No longer true when we add bounds…

The decision terminates when the accumulated evidence reaches a bound

Sample paths:

$$e_i = \mathcal{N}(\mu\Delta t, \sigma^2\Delta t)$$

$$x_{i+1} = x_i + e_i$$

$$x_0 = 0$$

The decision terminates when the accumulated evidence reaches a bound

Sample paths:

$$e_i = \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

$$x_{i+1} = x_i + e_i$$

$$x_0 = 0$$

The decision terminates when the accumulated evidence reaches a bound

Sample paths:

$$e_i = \mathcal{N}(\mu\Delta t, \sigma^2\Delta t)$$

$$x_{i+1} = x_i + e_i$$

$$x_0 = 0$$

The decision terminates when the accumulated evidence reaches a bound

Sample paths:

$$e_i = \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

$$x_{i+1} = x_i + e_i$$

$$x_0 = 0$$

The decision terminates when the accumulated evidence reaches a bound

Sample paths:

$$e_i = \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

$$x_{i+1} = x_i + e_i$$

$$x_0 = 0$$

The decision terminates when the accumulated evidence reaches a bound

Sample paths:

$$e_i = \mathcal{N}(\mu\Delta t, \sigma^2\Delta t)$$

$$x_{i+1} = x_i + e_i$$

$$x_0 = 0$$

The decision terminates when the accumulated evidence reaches a bound

Sample paths:

$$e_i = \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

$$x_{i+1} = x_i + e_i$$

$$x_0 = 0$$

Add bounds to the simulated trajectories

Code in "02_Simulate_diffusion_with_drift/main.m"

» Added symmetric bounds at: B = ± 1 and set $\mu = 1$ to have more errors

Add bounds to the simulated trajectories

Code in "02_Simulate_diffusion_with_drift/main.m"

» Added symmetric bounds at: B = ± 1 and set $\mu = 1$ to have more errors

Some trajectories

Add bounds to the simulated trajectories

Code in "02_Simulate_diffusion_with_drift/main.m"

» Added symmetric bounds at: B = ± 1 and set $\mu = 1$ to have more errors

t = 0.5 s

Some trajectories

The state of the DV is no longer Gaussian

Add bounds to the simulated trajectories

Code in "02_Simulate_diffusion_with_drift/main.m"

» Added symmetric bounds at: B = ± 1 and set $\mu = 1$ to have more errors

t = 0.5 s

Some trajectories

The state of the DV is no longer Gaussian

Correct and incorrect decisions have the same decision time

Add bounds to the simulated trajectories

Code in "02_Simulate_diffusion_with_drift/main.m"

» Added symmetric bounds at: B = ± 1 and set $\mu = 1$ to have more errors

Some trajectories

Response time distributions have long tails

Add bounds to the simulated trajectories

Code in "02_Simulate_diffusion_with_drift/main.m"

» Added symmetric bounds at: B = ± 1 and set $\mu = 1$ to have more errors

Some trajectories

Response time distributions have long tails

Model parameters, simplest version

Parameters in the 'simple' version:

Model parameters, simplest version

Parameters in the 'simple' version:

» Drift rate, μ

Model parameters, simplest version

Parameters in the 'simple' version:

» Drift rate, μ

Model parameters, simplest version

Parameters in the 'simple' version:

» Drift rate, μ

» Bound height, B

Model parameters, simplest version

Parameters in the 'simple' version:

» Drift rate, μ

» Bound height, B

Model parameters, simplest version

Parameters in the 'simple' version:

- Drift rate, μ
- Bound height, B
- Non-decision time, t_{nd}

Model parameters, simplest version

Parameters in the 'simple' version:

- Drift rate, μ
- Bound height, B
- Non-decision time, t_{nd}

Model parameters, simplest version

Parameters in the 'simple' version:

» Drift rate, μ

» Bound height, B

» Non-decision time, t_{nd}

Model parameters, simplest version

Parameters in the 'simple' version:

- Drift rate, μ
- Bound height, B
- Non-decision time, t_{nd}
- Diffusion coefficient, σ

Model parameters, simplest version

Parameters in the 'simple' version:

- Drift rate, μ
- Bound height, B
- Non-decision time, t_{nd}
- Diffusion coefficient, σ

Model parameters, simplest version

Parameters in the 'simple' version:

- Drift rate, μ
- Bound height, B
- Non-decision time, t_{nd}
- Diffusion coefficient, $\sigma = 1$

Model parameters, simplest version

Parameters in the 'simple' version:

- Drift rate, μ
- Bound height, B
- Non-decision time, t_{nd}
- Diffusion coefficient, $\sigma = 1 \times 2$

Model parameters, simplest version

Parameters in the 'simple' version:

- Drift rate, μ $\times$ **2**
- Bound height, B $\times$ **2**
- Non-decision time, t_{nd}
- Diffusion coefficient, $\sigma = 1$ $\times$ **2**

Model parameters, simplest version

Parameters in the 'simple' version:

- Drift rate, μ
- Bound height, B
- Non-decision time, t_{nd}
- Diffusion coefficient, $\sigma = 1$

Change in bound height

Change in bound height

Change in bound height

Change in bound height

Change in drift rate

Change in drift rate

Change in drift rate

Change in drift rate

Change in non-decision time

Change in non-decision time

Change in non-decision time

Analytic solution to the choice and response times

Probability of terminating at the upper bound:

$$p_+ = \frac{1}{1 + e^{-2\mu B}}$$

Analytic solution to the choice and response times

Probability of terminating at the upper bound:

$$p_+ = \frac{1}{1 + e^{-2\mu B}}$$

Analytic solution to the choice and response times

Probability of terminating at the upper bound:

$$p_+ = \frac{1}{1 + e^{-2\mu B}}$$

Average response time:

$$T = \frac{B}{\mu} \tanh(\mu B) + t_{nd}$$

Analytic solution to the choice and response times

Probability of terminating at the upper bound:

$$p_+ = \frac{1}{1 + e^{-2\mu B}}$$

Average response time:

$$T = \frac{B}{\mu} \tanh(\mu B) + t_{nd}$$

When $\mu = 0$: $T = B^2 + t_{nd}$

How to model an actual task?

Need to link parameters of the task to those of the model

Momentary evidence in the DDM

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

time step

drift coefficient

diffusion coefficient

How to model an actual task?

Need to link parameters of the task to those of the model

Momentary evidence in the DDM

time step

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

drift coefficient

diffusion coefficient

Make the drift and diffusion coefficients a function of the coherence

$$\mu = f(C)$$

Motion coherence

$$\sigma^2 = g(C)$$

How to model an actual task?

Need to link parameters of the task to those of the model

Momentary evidence in the DDM

time step

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

drift coefficient

diffusion coefficient

Make the drift and diffusion coefficients a function of the coherence

$$\mu = f(C)$$

Motion coherence

$$\sigma^2 = g(C)$$

$$\mu = \kappa \times C$$

Linear

How to model an actual task?

Need to link parameters of the task to those of the model

Momentary evidence in the DDM

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

time step

drift coefficient

diffusion coefficient

Make the drift and diffusion coefficients a function of the coherence

$$\mu = f(C)$$

$$\sigma^2 = g(C)$$

Motion coherence

$$\mu = \kappa \times C$$

Linear

$$\sigma^2 = 1$$

Constant

How to model an actual task?

Need to link parameters of the task to those of the model

Momentary evidence in the DDM

time step

$$e_i \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$$

drift coefficient

diffusion coefficient

Make the drift and diffusion coefficients a function of the coherence

$$\mu = f(C)$$

$$\sigma^2 = g(C)$$

Motion coherence

$\mu = \kappa \times C$

Linear

$\sigma^2 = 1$

Constant

Are these assumptions sensible?

The source of the momentary motion evidence in the random dot motion (RDM) task

Mean and variance of the momentary evidence

Mean and variance of the momentary evidence

Mean and variance of the momentary evidence

Mean and variance of the momentary evidence

Mean and variance of the momentary evidence

Momentary evidence:
$\langle \text{pref.} \rangle - \langle \text{non pref} \rangle$

Mean and variance of the momentary evidence

Mean and variance of the momentary evidence

- To model data from the RDM task, it is reasonable to assume that the mean of the momentary evidence is a linear function of stimulus strength

- To model data from the RDM task, it is reasonable to assume that the mean of the momentary evidence is a linear function of stimulus strength
- Scaling the variance of the momentary evidence with stimulus strength usually leads to (slightly) better fits than assuming that the variance is constant (e.g., Fetsch et al. Neuron 2014)

Modeling choice biases

When the momentary evidence is just noise, the “simple” DDM predicts that both choice options are equally likely

Modeling choice biases

When the momentary evidence is just noise, the “simple” DDM predicts that both choice options are equally likely

Probability of terminating at the upper bound:

$$p_+ = \frac{1}{1 + e^{-2\mu B}}$$

Modeling choice biases

When the momentary evidence is just noise, the “simple” DDM predicts that both choice options are equally likely

Probability of terminating at the upper bound:

$$p_+ = \frac{1}{1 + e^{-2\mu B}}\bigg|_{\mu=0} = \frac{1}{2}$$

Modeling choice biases

When the momentary evidence is just noise, the “simple” DDM predicts that both choice options are equally likely

Probability of terminating at the upper bound:

$$p_+ = \frac{1}{1 + e^{-2\mu B}}\Bigg|_{\mu=0} = \frac{1}{2}$$

Yet, sometimes decision makers have a systematic bias favoring one of the choice alternatives

Modeling choice biases

Modeling choice biases

$$p_{\text{rightward}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{coh})}}$$

Modeling choice biases

$$p_{\text{rightward}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{coh})}}$$

≠0

Modeling choice biases

$$p_{\text{rightward}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{coh})}}$$

≠0

Bias in the DDM model:

- Offset to the starting point
- Offset to the drift rate

Offset in starting point

decision variable

$$x_{i+1} = x_i + e_i$$

evidence

i: time step

$$x_0 = 0$$

Offset in starting point

decision variable

$$x_{i+1} = x_i + e_i$$

evidence

i: time step

$x_0 \neq 0$

Offset in starting point

decision variable

$$x_{i+1} = x_i + e_i$$

evidence

i: time step

$x_0 \neq 0$

Offset in starting point

decision variable

$$x_{i+1} = x_i + e_i$$

evidence

i: time step

$x_0 \neq 0$

Offset in starting point

decision variable

$$x_{i+1} = x_i + e_i$$

evidence

i: time step

$x_0 \neq 0$

Offset in drift rate

Offset in drift rate

$$e_i = \mathcal{N}(\mu\Delta t + \mu_0\Delta t\,, \sigma^2\Delta t)$$

$$x_{i+1} = x_i + e_i$$

$$x_0 = 0$$

Offset in drift rate

$$e_i = \mathcal{N}(\mu\Delta t + \mu_0\Delta t, \sigma^2\Delta t)$$

$$x_{i+1} = x_i + e_i$$

$$x_0 = 0$$

Questions?

Questions?

Questions?

Take a break

Towards data fitting…

1. How to predict choice & RT from a set of model parameters?
2. What to minimize/maximize when fitting parameters?
3. What fitting algorithm to use to search over the parameter space?

1. How to predict choice & RT from a set of model parameters?

1. How to predict choice & RT from a set of model parameters?

» Simulations of diffusion paths,

1. How to predict choice & RT from a set of model parameters?

- Simulations of diffusion paths,
- Analytical solutions to the choice and response time functions, for averages or full distributions,

1. How to predict choice & RT from a set of model parameters?

- Simulations of diffusion paths,
- Analytical solutions to the choice and response time functions, for averages or full distributions,
- Numerical solutions (propagate over time a probability distribution for the state of the decision variable)

Example: analytic solution to the choice and response time averages

Probability of terminating at the upper bound:

$$p_+ = \frac{1}{1 + e^{-2\mu B}}$$

Average response time:

$$T = \frac{B}{\mu} \tanh(\mu B) + t_{nd}$$

When $\mu = 0$: $T = B^2 + t_{nd}$

Example: analytic solution to the choice and response time averages

We saw this slide already…

Probability of terminating at the upper bound:

$$p_+ = \frac{1}{1 + e^{-2\mu B}}$$

Average response time:

$$T = \frac{B}{\mu} \tanh(\mu B) + t_{nd}$$

When $\mu = 0$: $T = B^2 + t_{nd}$

2. What to minimize/maximize when fitting parameters?

- Minimize the sum of squared errors of the mean RTs
- Max. the likelihood of the mean RTs
- Max. the likelihood of the single-trial RTs and choices

Minimize the sum of squared errors

Single trials

Minimize the sum of squared errors

Single trials

Average per coherence

Minimize the sum of squared errors

Single trials

Average per coherence

Model prediction

Minimize the sum of squared errors

Single trials

Average per coherence

Model prediction

$$\text{error} = \sum_{coh} \left(E[\text{RTdata}|\text{coh}] - E[\text{RTmodel}|\text{coh}] \right)^2$$

3. What fitting algorithm to use to search over the parameter space?

3. What fitting algorithm to use to search over the parameter space?

- fminsearch (Nelder–Mead simplex method)
- fmincon
- Bayesian Adaptive Direct Search (BADS)
  - Luigi Acerbi (https://github.com/lacerbi/bads)

3. What fitting algorithm to use to search over the parameter space?

- fminsearch (Nelder–Mead simplex method)
- fmincon
- Bayesian Adaptive Direct Search (BADS)
  - Luigi Acerbi (https://github.com/lacerbi/bads)

fminsearch

#useBADS

https://github.com/lacerbi/bads
https://github.com/lacerbi/optimviz

Slide designed by Wei Ji Ma

fminsearch

#useBADS

https://github.com/lacerbi/bads
https://github.com/lacerbi/optimviz

Slide designed by Wei Ji Ma

BADS

#useBADS

https://github.com/lacerbi/bads
https://github.com/lacerbi/optimviz

Slide designed by Wei Ji Ma

BADS

#useBADS

https://github.com/lacerbi/bads
https://github.com/lacerbi/optimviz

Slide designed by Wei Ji Ma

Let’s try it!

1. How to predict choice & RT from a set of model parameters?

   Using analytic solutions to the choice and RT averages

2. What to minimize/maximize when fitting parameters?

   Minimize the sum of squared errors of the mean RTs

3. What fitting algorithm to use to search over the parameter space?

   BADS

Model fitting

Code in “03_Fit_means/main.m”

test_data.mat

Motion Coherence (%)	Choice	RT (s)	Correct
-3.2	0	2.32	1
-12.8	0	0.76	1
0.0	0	1.31	0
51.2	1	0.39	1
-51.2	0	0.38	1
-25.6	0	0.55	1
-12.8	0	0.52	1
0.0	0	1.40	0
51.2	1	0.31	1
-3.2	0	2.05	1

Model fitting

Code in “03_Fit_means/main.m”

test_data.mat

Motion Coherence (%)	Choice	RT (s)	Correct
-3.2	0	2.32	1
-12.8	0	0.76	1
0.0	0	1.31	0
51.2	1	0.39	1
-51.2	0	0.38	1
-25.6	0	0.55	1
-12.8	0	0.52	1
0.0	0	1.40	0
51.2	1	0.31	1
-3.2	0	2.05	1

Why not fit the data using the sum of squared errors?

- There can be a big mismatch in the RT distributions between model and data
- Not straightforward to combine choice and response times data

Better approach: Maximum Likelihood

Better approach: Maximum Likelihood

Better approach: Maximum Likelihood

data points

$$LL = \sum_i log(p(x^{(i)} \mid \theta))$$

model parameters

Better approach: Maximum Likelihood

data points

$$LL = \sum_i log(p(x^{(i)} \mid \theta))$$

model parameters

Maximum Likelihood for choice & RT data

model parameters

$$\hat{\theta}_{\text{mle}} = \arg\max_{\theta} \left(\sum_{i=1}^{n_{tr}} \log \left(p(\text{RT}^{(i)}, \text{choice}^{(i)} | c^{(i)}, \theta) \right) \right)$$

motion coherence

Maximum Likelihood for choice & RT data

model parameters

$$\hat{\theta}_{\text{mle}} = \arg\max_{\theta} \left(\sum_{i=1}^{n_{tr}} \log \left(\underbrace{p(\text{RT}^{(i)}, \text{choice}^{(i)} | c^{(i)}, \theta)}_{\text{We need to be able to evaluate this p.d.f.}} \right) \right)$$

For ‘flat’ bounds, there are analytic solutions to the choice and response time distributions

Probability of terminating at the lower bound at time t:

$$p(t, \text{left}|\mu, B, \sigma, x_0) = p_-(\mu, B, \sigma, x_0) - \frac{\pi\sigma^2}{4B^2} e^{-\frac{(B+x_0)\mu}{\sigma^2}} \times \sum_{k=1}^{\infty} \frac{2k\ \sin(\frac{k\pi(B+x_0)}{2B}) e^{-\frac{1}{2}(\frac{\mu^2}{\sigma^2} + \frac{\pi^2 k^2 \sigma^2}{4B^2})t}}{\left(\frac{\mu^2}{\sigma^2} + \frac{\pi^2 k^2 \sigma^2}{4B^2}\right)}$$

, where

$$p_-(\mu, B, \sigma, x_0) = \frac{e^{-(4\mu B/\sigma^2)} - e^{-(2\mu(B+x_0)/\sigma^2)}}{e^{-(4\mu B/\sigma^2)} - 1}$$

Ratcliff 1978, Feller 1968

Maximum Likelihood for choice & RT data

model parameters

$$\hat{\theta}_{\text{mle}} = \arg\max_{\theta} \left(\sum_{i=1}^{n_{tr}} \log \left(p(\text{RT}^{(i)}, \text{choice}^{(i)} | c^{(i)}, \theta) \right) \right)$$

motion coherence

Maximum Likelihood for choice & RT data

model parameters

$$\hat{\theta}_{\text{mle}} = \arg\max_{\theta} \left(\sum_{i=1}^{n_{tr}} \log \left(p(\text{RT}^{(i)}, \text{choice}^{(i)} | c^{(i)}, \theta) \right) \right)$$

motion coherence

Let’s try Maximum-likelihood fitting; compare to fitting means

Code in “03_Fit_means/main.m”

Let's try Maximum-likelihood fitting; compare to fitting means

Code in "03_Fit_means/main.m"

Toolboxes for fitting the DDM with flat bounds

1. HDDM

   http://ski.clps.brown.edu/hddm_docs/

2. DMAT

   https://ppw.kuleuven.be/okp/software/dmat/

3. fast-dm

   https://www.psychologie.uni-heidelberg.de/ae/meth/fast-dm/

4. EZ

   https://www.ejwagenmakers.com/2007/EZ.pdf

Sometimes it is useful to represent information about choices and response time distributions in a single plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Quantile probability plot

Problem with simple DDM: symmetric RT distributions

Problem with simple DDM: symmetric RT distributions

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

Think about it for a few minutes'…

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Pr = $ppqppp$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Pr = $ppqppp$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Pr = $ppqppp$

Pr(mirror) = $qqpqqq$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Pr = $ppqppp$

Pr(mirror) = $qqpqqq$

$$\frac{\text{Pr}}{\text{Pr(mirror)}} = \frac{pppp}{qqqq} = \left(\frac{p}{1-p}\right)^{\gamma}$$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Pr = $ppqppp$

Pr(mirror) = $qqpqqq$

$$\frac{\text{Pr}}{\text{Pr(mirror)}} = \frac{pppp}{qqqq} = \left(\frac{p}{1-p}\right)^{\gamma} \quad =4$$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Pr = $ppqppp$

Pr(mirror) = $qqpqqq$

$$\frac{\text{Pr}}{\text{Pr(mirror)}} = \frac{pppp}{qqqq} = \left(\frac{p}{1-p}\right)^{\gamma} \quad =4$$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Pr = $qpqpppqppqqpqppqpp$

Pr(mirror) = $pqpqqqpqqppqpqqpqq$

$$\frac{\text{Pr}}{\text{Pr(mirror)}} = \frac{pppp}{qqqq} = \left(\frac{p}{1-p}\right)^{\gamma} \quad =4$$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Pr = $qpqpppqppqqpqppqpp$

Pr(mirror) = $pqpqqqpqqppqpqqpqq$

$$\frac{\text{Pr}}{\text{Pr(mirror)}} = \frac{pppp}{qqqq} = \left(\frac{p}{1-p}\right)^{\gamma} \quad =4$$

$$P(t, \text{right}) = aP(t, \text{left})$$

Why flat bounds lead to identical RT distributions for correct and incorrect decisions?

↗ with probability p

↘ with probability $q = 1 - p$

Pr = $qpqpppqppqqpqppqpp$

Pr(mirror) = $pqpqqqpqqppqpqqpqq$

$$\frac{\text{Pr}}{\text{Pr(mirror)}} = \frac{pppp}{qqqq} = \left(\frac{p}{1-p}\right)^{\gamma} \quad =4$$

After normalization:

$$P(t\,|\,\text{right}) = P(t\,|\,\text{left})$$

Problem with simple DDM: symmetric RT distributions

Problem with simple DDM: symmetric RT distributions

How to “patch” the simple model to make errors slower than correct responses?

Inter-trial variability in drift-rate across trials

Ratcliff & McKoon (2008)

Inter-trial variability in drift-rate across trials

Ratcliff & McKoon (2008)

Inter-trial variability in drift-rate across trials

Ratcliff & McKoon (2008)

Inter-trial variability in drift-rate across trials

Drift-rates closer to zero → Slower responses

Drift-rates closer to zero → Higher error rates

Ratcliff & McKoon (2008)

Collapsing bounds

MN Shadlen et al. (2006)

Collapsing bounds

MN Shadlen et al. (2006)

Collapsing bounds

MN Shadlen et al. (2006)

Collapsing bounds

MN Shadlen et al. (2006)

Collapsing bounds

MN Shadlen et al. (2006)

The patch is not really a patch…

Collapsing bounds maximize reward-rate when difficulty varies across trials

Frazier & Yu, NiPS 2008
Drugowitsch et al., JNeuro 2012

Collapsing bounds maximize reward-rate when difficulty varies across trials

Frazier & Yu, NiPS 2008
Drugowitsch et al., JNeuro 2012

Collapsing bounds maximize reward-rate when difficulty varies across trials

Frazier & Yu, NiPS 2008
Drugowitsch et al., JNeuro 2012

Collapsing bounds maximize reward-rate when difficulty varies across trials

Frazier & Yu, NiPS 2008
Drugowitsch et al., JNeuro 2012

Collapsing bounds maximize reward-rate when difficulty varies across trials

Frazier & Yu, NiPS 2008
Drugowitsch et al., JNeuro 2012

Some bound shapes…

$$B(t) = \frac{B_0}{1 + e^{a(t-d)}}$$

$$B(t) = \begin{cases} B_0, & \text{if } t < d \\ B_0 - a(t-d)^2 & \text{otherwise} \end{cases}$$

$$B(t) = \begin{cases} B_0, & \text{if } t < d \\ B_0 e^{-a*(t-d)}, & \text{otherwise} \end{cases}$$

See expand_bounds.m

Issue: collapsing bounds do not usually admit analytical solutions to the choice and response time distributions

We have seen this slide before…

How to predict choice & RT from a set of model parameters?

- Simulations of diffusion paths,
- Analytical solutions to the choice and response time functions, for averages or full distributions,
- Numerical solutions (propagate over time a probability distribution for the state of the decision variable)

We have seen this slide before…

How to predict choice & RT from a set of model parameters?

- Simulations of diffusion paths,
- Analytical solutions to the choice and response time functions, for averages or full distributions,
- Numerical solutions (propagate over time a probability distribution for the state of the decision variable)

Fokker-Planck equation

Partial differential equation that describes the time evolution of the probability density function of the decision variable x:

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \cancel{\frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}}$$

Fokker-Planck intuition

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \cancel{\frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}}$$

$$-\mu \frac{\partial p(x,t)}{\partial x}$$

Fokker-Planck intuition

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \cancel{\frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}}$$

$$-\mu \frac{\partial p(x,t)}{\partial x}$$

Fokker-Planck intuition

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = \cancel{-\mu \frac{\partial p(x,t)}{\partial x}} + \frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = \cancel{-\mu \frac{\partial p(x,t)}{\partial x}} + \frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial^2 x}$$

$$\frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = \cancel{-\mu \frac{\partial p(x,t)}{\partial x}} + \frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial^2 x}$$

$$\frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = \cancel{-\mu \frac{\partial p(x,t)}{\partial x}} + \frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial^2 x}$$

$$\frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = \cancel{-\mu \frac{\partial p(x,t)}{\partial x}} + \frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial^2 x}$$

$$\frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Numerical approximations to the FP equation

Finite difference methods: approximate a differential equation by a difference equation

Numerical approximations to the FP equation

Finite difference methods: approximate a differential equation by a difference equation

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Numerical approximations to the FP equation

Finite difference methods: approximate a differential equation by a difference equation

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Numerical approximations to the FP equation

Finite difference methods: approximate a differential equation by a difference equation

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Accumulated evidence (index j)

Time (index n)

Approximation to the Fokker-Planck equation (one of many possible ones…):

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = -\mu \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} + \frac{\sigma^2}{2} \frac{u_{j-1}^n - 2u_j^n + u_{j+1}^n}{\Delta x^2}$$

Numerical approximations to the FP equation

Finite difference methods: approximate a differential equation by a difference equation

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Accumulated evidence (index j)

Time (index n)

Approximation to the Fokker-Planck equation (one of many possible ones…):

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = -\mu \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} + \frac{\sigma^2}{2} \frac{u_{j-1}^n - 2u_j^n + u_{j+1}^n}{\Delta x^2}$$

Numerical approximations to the FP equation

Finite difference methods: approximate a differential equation by a difference equation

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Accumulated evidence (index j)

Time (index n)

Approximation to the Fokker-Planck equation (one of many possible ones…):

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = -\mu \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} + \frac{\sigma^2}{2} \frac{u_{j-1}^n - 2u_j^n + u_{j+1}^n}{\Delta x^2}$$

Numerical approximations to the FP equation

$$\begin{pmatrix} \\ U^{n+1} \\ \\ \end{pmatrix} = \begin{pmatrix} \\ \text{TRI-DIAGONAL} \\ \text{MATRIX} \\ \\ \end{pmatrix} \begin{pmatrix} \\ U^{n} \\ \\ \end{pmatrix} \quad \text{, where } U^n = \begin{pmatrix} u_1^n \\ u_2^n \\ \vdots \\ u_j^n \\ \vdots \end{pmatrix}$$

Numerical approximations to the FP equation

$$\begin{pmatrix} \\ U^{n+1} \\ \\ \end{pmatrix} = \begin{pmatrix} \\ \text{TRI-DIAGONAL} \\ \text{MATRIX} \\ \\ \end{pmatrix} \begin{pmatrix} \\ U^{n} \\ \\ \end{pmatrix} \quad \text{, where } U^n = \begin{pmatrix} u_1^n \\ u_2^n \\ \vdots \\ u_j^n \\ \vdots \end{pmatrix}$$

Can be implemented with a few lines of Matlab code:

```
% prepare tri-diagonal matrix
temp = repmat([(-a/(2*dx)+D/(dx^2)),-2*D/(dx^2),(a/(2*dx)+D/dx^2)],nx,1);
A = spdiags(temp,-1:1,nx,nx);

for i = 2:nt % iterate over time
    u = dt*A*u + u; % update u
    U(:,i) = u; % save
end
```

Fokker-Planck intuition

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Fokker-Planck intuition

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

With bounds

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

With bounds

Fokker Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = -\mu \frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x,t)}{\partial^2 x}$$

Model fitting

model parameters

$$\hat{\theta}_{\text{mle}} = \arg\max_{\theta} \left(\sum_{i=1}^{n_{tr}} \log \left(p(\text{RT}^{(i)}, \text{choice}^{(i)} | c^{(i)}, \theta) \right) \right)$$

motion coherence

Let’s try it!

1. How to predict choice & RT from a set of model parameters?

   ~~Using analytic solutions to the choice and RT averages~~

   Using numerical approximations to the Fokker Planck equation

2. What to minimize/maximize when fitting parameters?

   ~~Minimize the sum of squared errors~~

   Maximize the likelihood of the parameters given the single-trial choice and RT data

3. What fitting algorithm to use to search over the parameter space?

   BADS

Let’s try it!

1. How to predict choice & RT from a set of model parameters?

   ~~Using analytic solutions to the choice and RT averages~~

   Using numerical approximations to the Fokker Planck equation

2. What to minimize/maximize when fitting parameters?

   ~~Minimize the sum of squared errors~~

   Maximize the likelihood of the parameters given the single-trial choice and RT data

3. What fitting algorithm to use to search over the parameter space?

   BADS

Go to “04_Fit_single_trials/main.m”…

Exercise: fit the data from Roitman & Shadlen JNeurosci 2002

You should have obtained something like this…

Best fitting parameters:

$$\kappa = 12.77$$
$$\mu_{nd} = 0.13$$
$$\sigma_{nd} = 0.13$$
$$B_0 = 1.59$$
$$a = 4$$
$$d = 0.66$$

Validating the model fitting method

Validating the model fitting method

» Fit model to experimental data

Validating the model fitting method

- Fit model to experimental data
- Simulate data with the fitted parameters

Validating the model fitting method

- Fit model to experimental data
- Simulate data with the fitted parameters
- Fit the simulated data and see if the optimization procedure recovers the parameters used to generate the data

Validating the model fitting method

- Fit model to experimental data
- Simulate data with the fitted parameters
- Fit the simulated data and see if the optimization procedure recovers the parameters used to generate the data

Validating the model fitting method

- Fit model to experimental data
- Simulate data with the fitted parameters
- Fit the simulated data and see if the optimization procedure recovers the parameters used to generate the data

Exercise: compare fit parameter across datasets

- In “data/sim_dat_extrema.mat” there are two sets of behavioral data (sim(1) and sim(2)).
- These may correspond to, for instance, two sets of participants performing the same task (e.g., patients and controls).
- Fit the DDM model to these datasets.
- Discuss what parameters changed across the two groups.

Fits of the DDM to the 1st dataset

Best fitting parameters:

$$\kappa = 24.34$$
$$\mu_{nd} = 0.35$$
$$\sigma_{nd} = 0.01$$
$$B_0 = 0.65$$
$$a = 3.95$$
$$d = 1.26$$

Fits of the DDM to the 2nd dataset

Best fitting parameters:

$\kappa = 24.34$	21
$\mu_{nd} = 0.35$	0.33
$\sigma_{nd} = 0.01$	0.01
$B_0 = 0.65$	2.43
$a = 3.95$	0.25
$d = 1.26$	-2.82

The data was generated with the ‘extrema’ model; only B_0 changed across conditions

Cartwright and Festinger 1943

The data was generated with the ‘extrema’ model; only B_0 changed across conditions

Cartwright and Festinger 1943

True difference between conditions:

The data was generated with the ‘extrema’ model; only B_0 changed across conditions

Cartwright and Festinger 1943

True difference between conditions:

- Good fits from a drift-diffusion model are not enough to conclude that decision makers accumulate evidence over time (alternative models could fit just as well)

The data was generated with the ‘extrema’ model; only B_0 changed across conditions

Cartwright and Festinger 1943

True difference between conditions:

- Good fits from a drift-diffusion model are not enough to conclude that decision makers accumulate evidence over time (alternative models could fit just as well)
- Need for formal model comparison, after showing that the models are distinguishable in synthetic data

The data was generated with the ‘extrema’ model; only B_0 changed across conditions

Cartwright and Festinger 1943

True difference between conditions:

- Good fits from a drift-diffusion model are not enough to conclude that decision makers accumulate evidence over time (alternative models could fit just as well)
- Need for formal model comparison, after showing that the models are distinguishable in synthetic data
- Still, there will be a blindspot (due to models not considered)

The data was generated with the ‘extrema’ model; only B_0 changed across conditions

Cartwright and Festinger 1943

True difference between conditions:

- Good fits from a drift-diffusion model are not enough to conclude that decision makers accumulate evidence over time (alternative models could fit just as well)
- Need for formal model comparison, after showing that the models are distinguishable in synthetic data
- Still, there will be a blindspot (due to models not considered)
- Careful when interpreting fitted parameters

How to compare DDM models with different number of parameters?

Visual inspection

Fit to Roitman's data as in folder "05_Fitting_exercise_roitman"

Fit to Roitman's data with an extra parameter (bias on drift rate)

Likelihood-based metrics

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)
- Marginal likelihoods (model evidence, Bayes' factors)
- Deviance Information Criterion (DIC)
- Watanabe-Akaike Information criterion
- Leave-one-out or K-fold cross-validation
- …

Akaike information criterion (AIC)

Log likelihood of the model

$$\text{AIC} = -2LL + 2K$$

Number of model parameters

, and choose the model with the smallest AIC.

$$\text{AIC}_{\text{model no bias}} = -2 \times (-27251) + 2 \times 6 = 54514$$

$$\text{AIC}_{\text{model bias}} = -2 \times (-27235) + 2 \times 7 = 54484$$

Akaike information criterion (AIC)

Log likelihood of the model

$$\text{AIC} = -2LL + 2K$$

Number of model parameters

, and choose the model with the smallest AIC.

$$\text{AIC}_{\text{model no bias}} = -2 \times (-27251) + 2 \times 6 = 54514$$

$$\text{AIC}_{\text{model bias}} = -2 \times (-27235) + 2 \times 7 = 54484$$

Bayesian information criterion (BIC)

Log likelihood of the model

$$\text{BIC} = -2LL + K \times log(N)$$

Number of observations

Number of model parameters

, and choose the model with the smallest BIC.

$$\text{BIC}_{\text{model no bias}} = -2 \times (-27251) + 6 \times log(3534) = 54551$$

$$\text{BIC}_{\text{model bias}} = -2 \times (-27235) + 7 \times log(3534) = 54527$$

Bayesian information criterion (BIC)

Log likelihood of the model

$$\text{BIC} = -2LL + K \times log(N)$$

Number of observations

Number of model parameters

, and choose the model with the smallest BIC.

$$\text{BIC}_{\text{model no bias}} = -2 \times (-27251) + 6 \times log(3534) = 54551$$

$$\text{BIC}_{\text{model bias}} = -2 \times (-27235) + 7 \times log(3534) = 54527$$

Summary

Summary

Background

Summary

Background

Early sequential sampling model

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters
- The quantile-probability plot

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters
- The quantile-probability plot
- Differences in RT between error and correct responses

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters
- The quantile-probability plot
- Differences in RT between error and correct responses

Model fitting

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters
- The quantile-probability plot
- Differences in RT between error and correct responses

Model fitting

- Probability distributions over choice and RT

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters
- The quantile-probability plot
- Differences in RT between error and correct responses

Model fitting

- Probability distributions over choice and RT
- Parameter optimization

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters
- The quantile-probability plot
- Differences in RT between error and correct responses

Model fitting

- Probability distributions over choice and RT
- Parameter optimization
- Parameter recovery

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters
- The quantile-probability plot
- Differences in RT between error and correct responses

Model fitting

- Probability distributions over choice and RT
- Parameter optimization
- Parameter recovery
- Model mimicry

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters
- The quantile-probability plot
- Differences in RT between error and correct responses

Model fitting

- Probability distributions over choice and RT
- Parameter optimization
- Parameter recovery
- Model mimicry
- Model comparison using BIC and AIC

Summary

Background

- Early sequential sampling model
- Sequential probability ratio test

The drift-diffusion model

- Influence of the different parameters
- The quantile-probability plot
- Differences in RT between error and correct responses

Model fitting

- Probability distributions over choice and RT
- Parameter optimization
- Parameter recovery
- Model mimicry
- Model comparison using BIC and AIC