Diffusion Modelと Linear Ballistic Accumulator

東京大学大学院　博士課程２年

分寺杏介

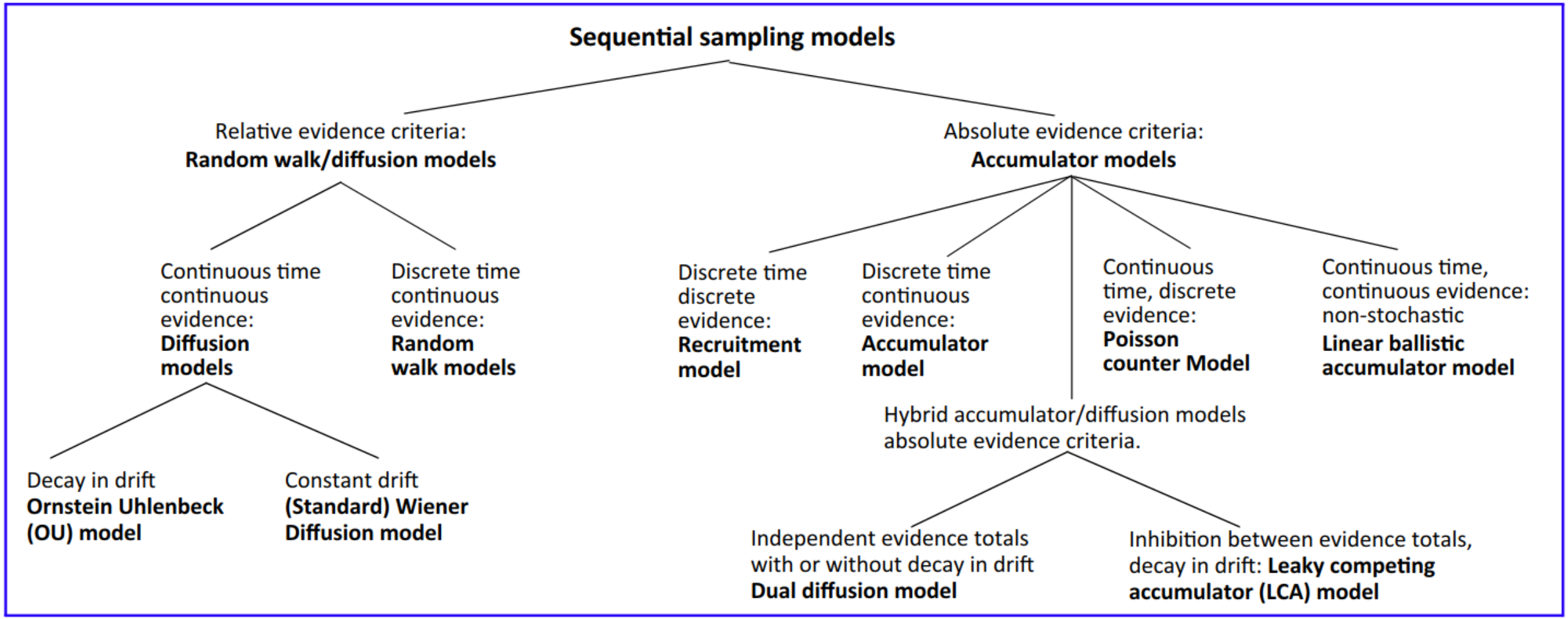
bunji@p.u-Tokyo.ac.jp

この報告では

- Diffusion ModelとLBAの元論文を可能なレベルで紹介します。
  - Applicationには個人的に興味が無いので飛ばしていきます。
- そこから派生したいくつかのモデルについても紹介します。
  - 「ほかにもこんなのあるよ」というのがあれば教えていただけると嬉しいです。
- 最後に私の研究について軽く紹介します。
  - どちらかというと理論寄りだと自覚しているので，「このモデルを使って実データを分析して解釈してみた」みたいなことは出てきません。

- 作り始めたのは10/4です。なので未完成です。
  - 図による解説がほとんどありません。頑張って喋ります。
  - 間違いや不足している点があったら指摘してください。修正して完成とします。

Sequential sampling models



1

Diffusion Modelをたずねて

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

1978年って

「ザ・ベストテン」放送開始

サザンオールスターズがメジャーデビュー「勝手にシンドバッド」

成田空港が開港

クラウンライターライオンズ▶西武ライオンズに

江川事件（空白の一日）

スペースインベーダー発売

（福岡は５月から翌年３月まで大渇水だったらしい）

Apple II の発売の翌年

はじめに

当時の認知心理学の状況

- 多くの実験に対する統一的な理論みたいなのが無かった
- みんなだいたい二項対立の構図で研究していた

例｜serial vs. parallel / storage vs. retrieval / semantic vs. episodic

そんなんで新しい理論も作れるかいや！

二項対立ばっかりしてるともっと大事なこと見逃すよ！

データの分布の性質だとか

ということで

この研究では，「記憶からの検索」に関する統一的な理論を提案します。

memory retrievalに関する様々なparadigmが提案されてきましたが…

- データによっては結果に一貫性がない
- paradigm間で互換性がない

ので，この研究で

- いろいろなparadigmで言われているいろいろな現象を全部説明する
- paradigmが変わっても使える

そんな理論を提案したいわけです。

1.1

理論の概念から

課題

【item recognition task】

- 記憶の中にsearch setがある（例｜覚えた複数の単語）
- プローブが提示される

→ search setと照合して「あったかなかったか」を答える的課題

  - 提示されると，search set内の各項目と比較する
  - 各比較がrandom walk (or diffusion process)であると考えられる
  - いずれか一つでもmatchしたら「あった」と答える

Figure 1

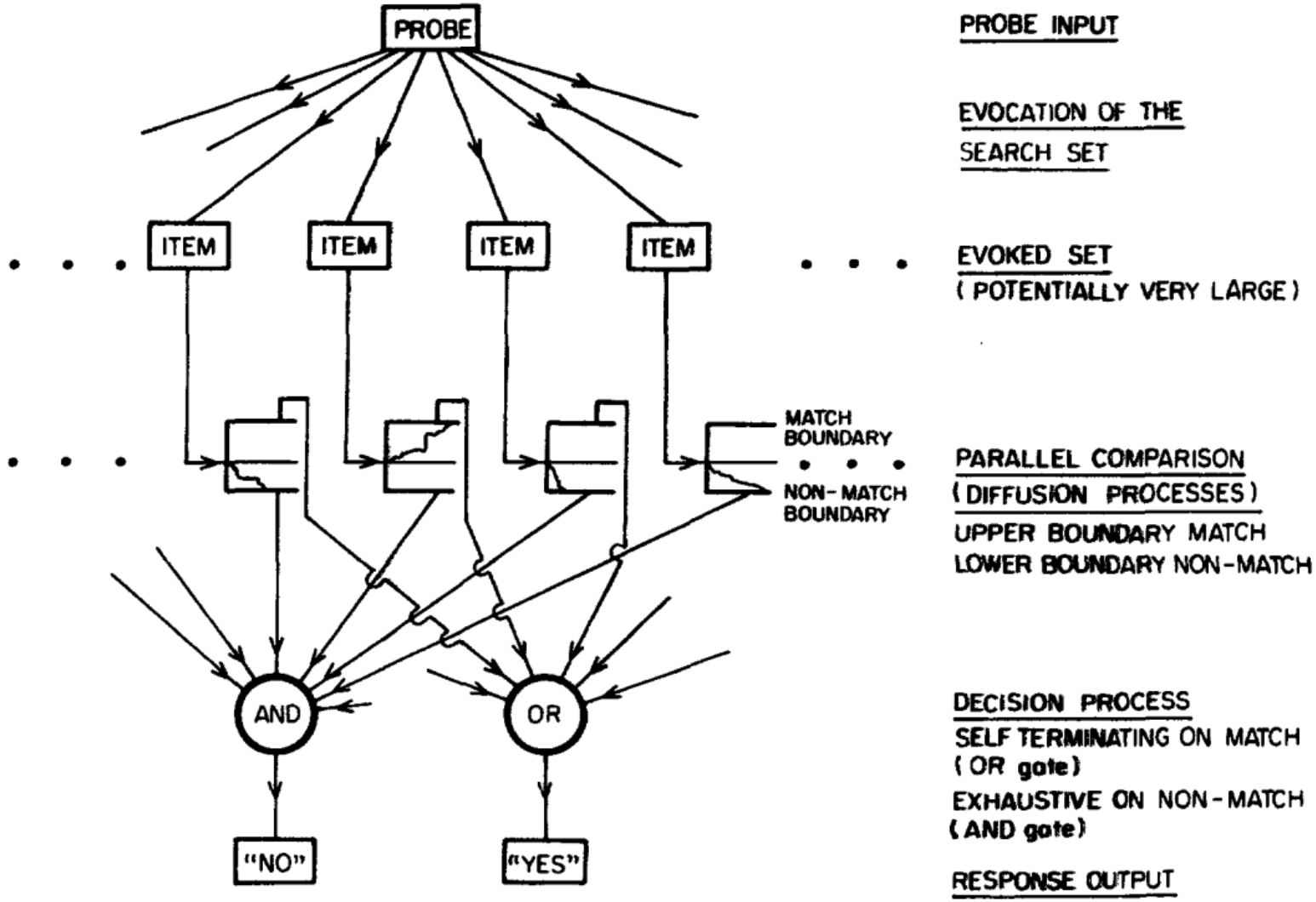


Figure 1. An overview of the item recognition model.

それまでのモデルの問題点

memory setにある単語(positive)のみがitem recognitionに関与する

→ いくつかの例で，negativeプローブによる影響が見られた

→ **negativeプローブもprocessに影響しうる，というモデルを作るべき**

このプロセスを音叉にたとえてみよう(resonance mataphor)

プローブの音叉を叩く▶memory setにある似た音叉が共鳴する

※音叉がpositiveかnegativeかは関係ない。ただその場にある音叉のうち似た音叉が共鳴するだけである。

comparison processに関わる情報

いろいろな先行研究があります。

- 類似度が上がるほど「はい」と回答する割合が高く，回答時間も短い
- 音や意味が似たnegative単語があると回答時間が長くなる
- negativeプローブがmemory setと数的に離れると回答時間は短くなる

➡ **probeとmemory setの類似度は，回答時間を決める主な要因**

実際には，もっと様々な情報がrecognition performanceに影響する

どうやってこれらをconceptualizeするのか？

relatedness

いろいろな情報があるが，ここでは**relatedness**という一つに代表させる

probeとmemory setの類似度

▶memory set内の項目ごとに異なる

数理モデルを構築する上で

「**relatednessは分散を持つ(正規分布に従う)**」という仮定が必要

- relatednessに分散を導入することで**「accuracyは漸近的に無限大にならない」**という性質を満たすことができる

Figure 2

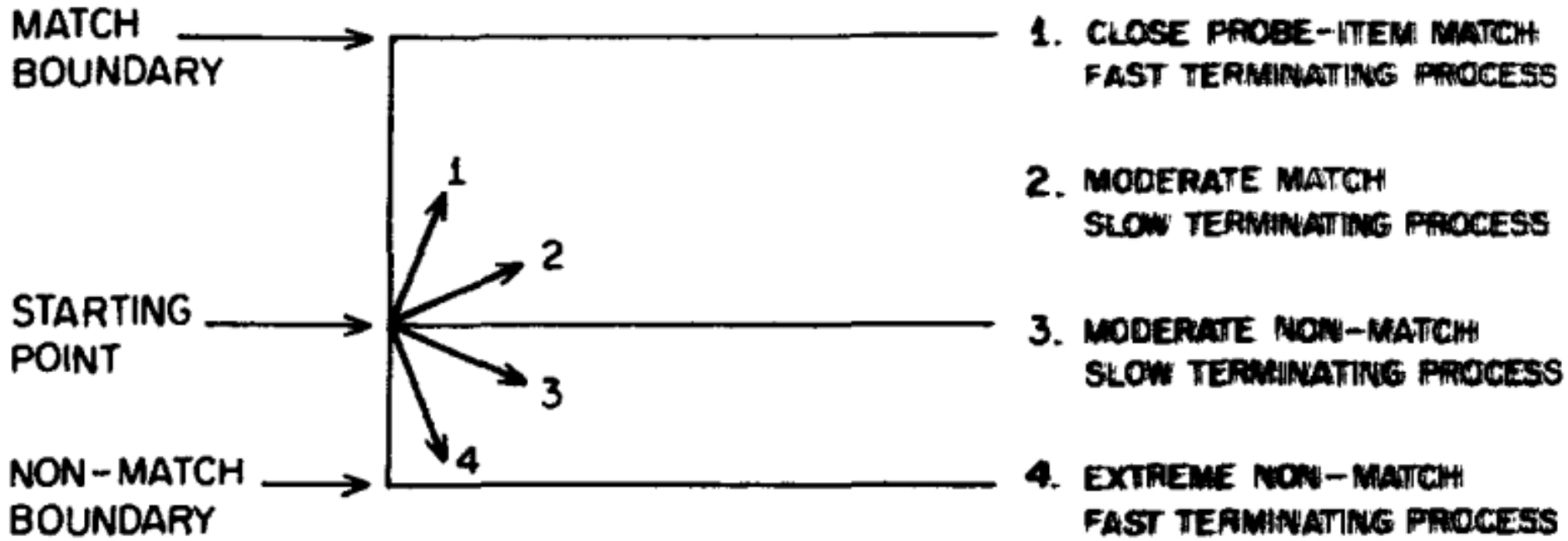


Figure 2. The relation between relatedness and amount of match in the diffusion random walk process. (Relatedness varies from high [Process 1] to low [Process 4].)

ここまで

２つの重要なポイントを示しました

1. **音叉のmetaphorを導入しました**

   probeとmemory setの交互作用を表現するために

   negativeプローブの特性による影響もアリとするために

2. **relatednessを導入しました**

   いろいろな情報を一次元的に表すために

Comparison Process

probeとmemory setを比較する過程

時間経過に応じてevidenceあるいはgoodness of matchを示すinformationが増加する

「特徴を一つ一つ比べていく」という感じ

最初はZから始まるとする（カウンターがある）

ある特徴がmatchした場合カウンターが１増える，non-matchの場合１減る

これを繰り返しカウンターがAに到達したら”match”と回答

0に到達したら”non-match”と回答

RelatednessはmatchするFeatureの割合で表される

2つのVariability

trial間 Variability

比較対象が異なれば，relatednessが異なるのは当然のこと

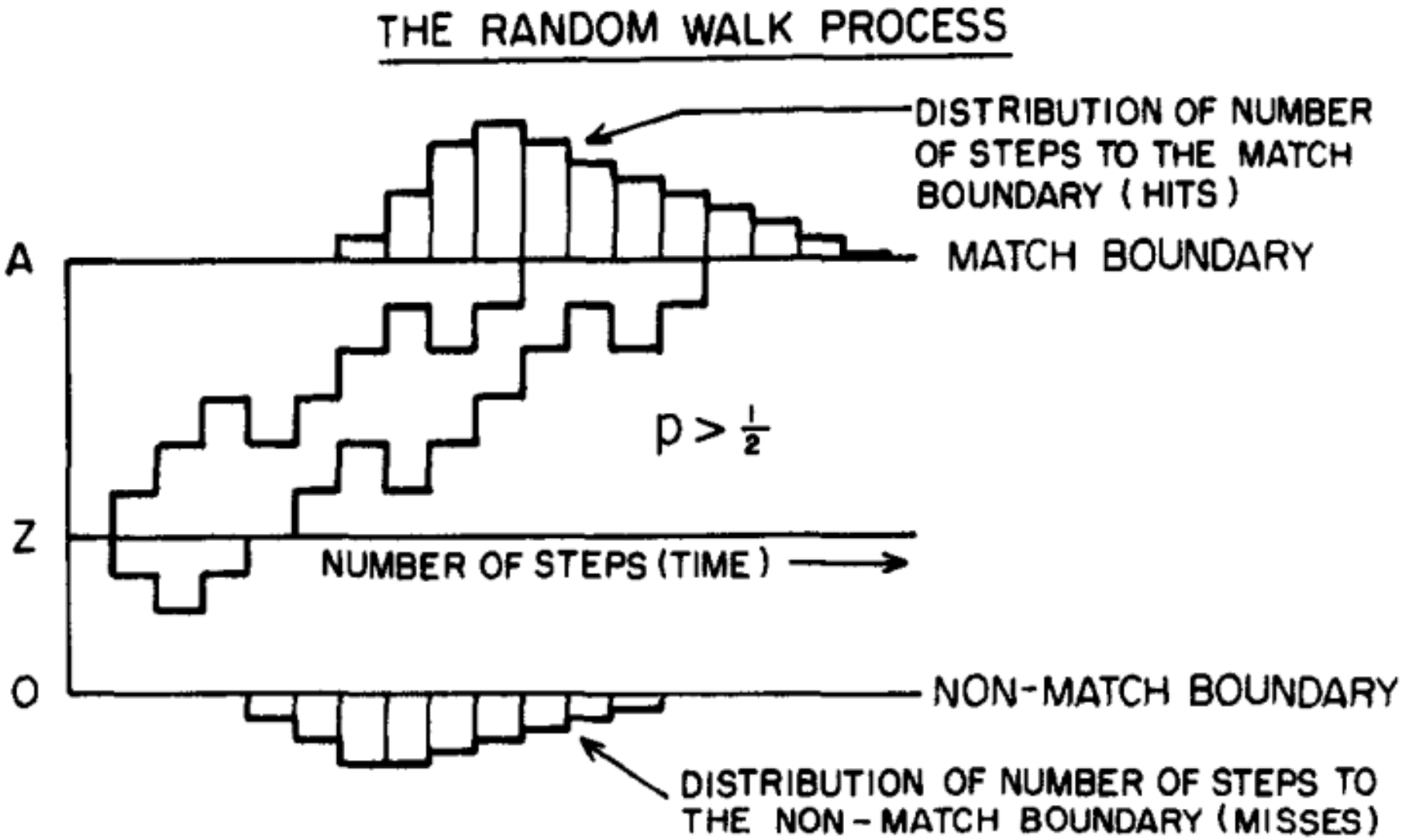
trial内 Variability

同じ比較であっても，比較される特徴の順序によって結果は変わる

先にmatchする特徴ばかり比較されれば，”match”と判断される

先にnon-matchする特徴ばかり比較されれば，”non-match”と判断される

Random walk process



Diffusion Process

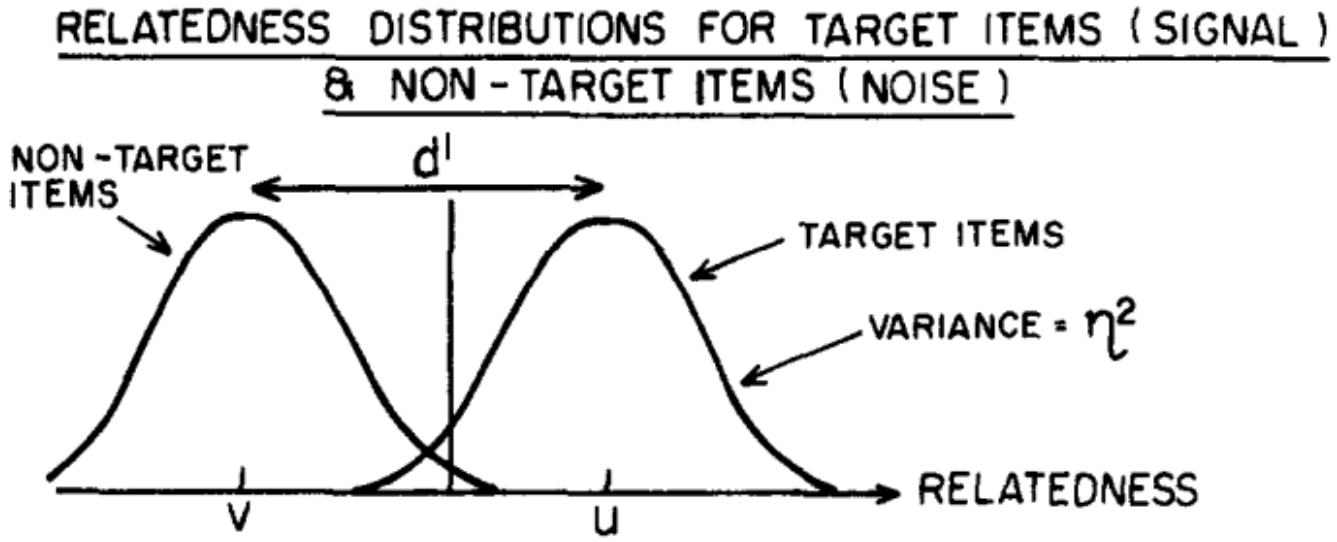
Diffusion Processはrandom walk processの連続量バージョンで表される

informationの蓄積の平均と分散が別々のパラメータで表される（正規分布）

※「特徴を比べる」モデルの場合，二項分布に従うため平均と分散が連動してしまう

drift rateはrelatednessと同じ値になる

類似度が高いプローブほど平均的に上に早く遷移する



Diffusion Process

パラメータは６つ

固定

① relatedness の分散｜η^2

② drift rate の分散｜s^2

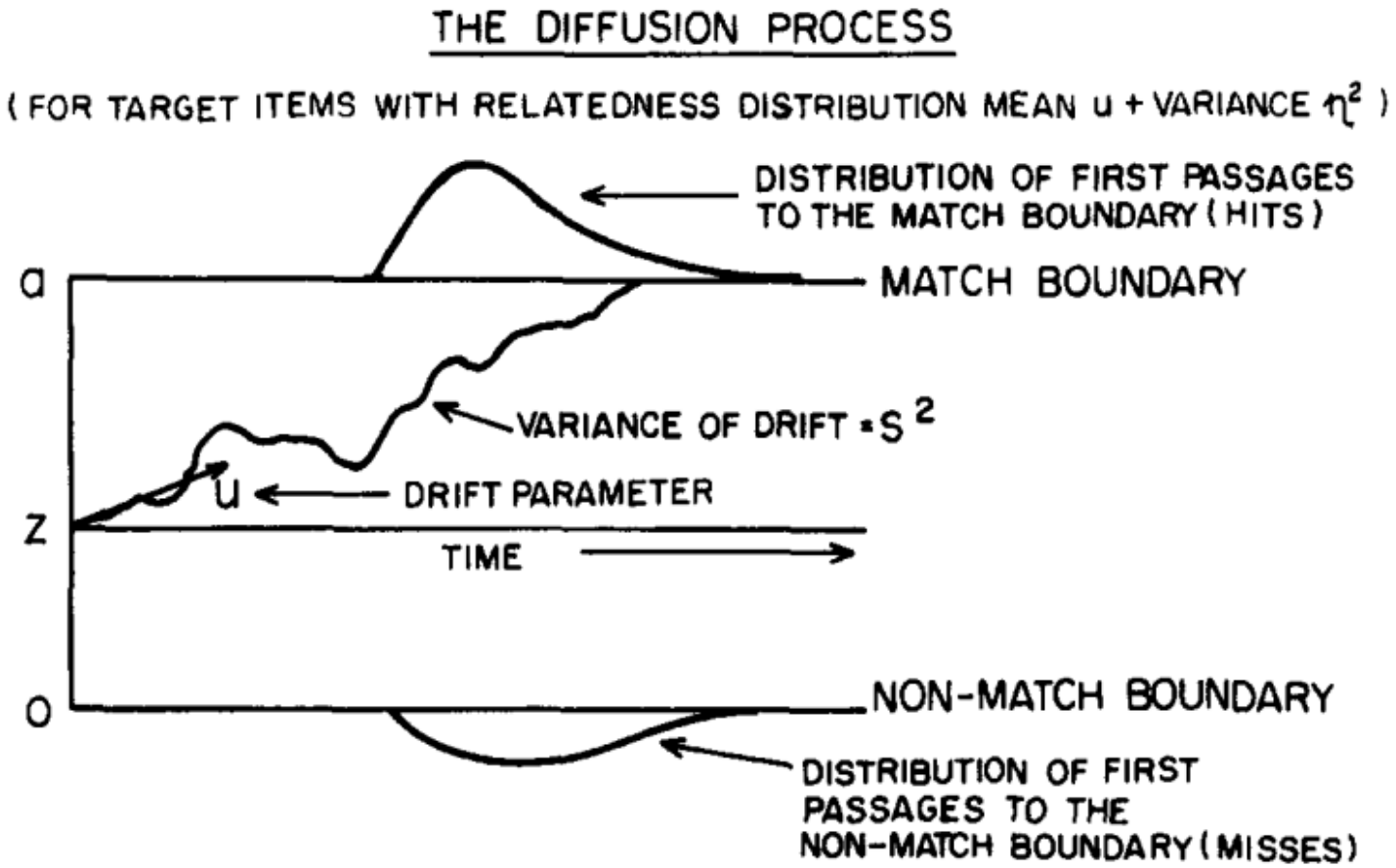
③ 上の閾値｜a

④ 開始地点｜z

⑤ match のときの drift rate｜u

⑥ non-match のときの drift rate｜v

＋ 符号化の時間｜T_{er}



Reaction Time distribution

これまでのモデルと一線を画するretrieval theoryの強みは

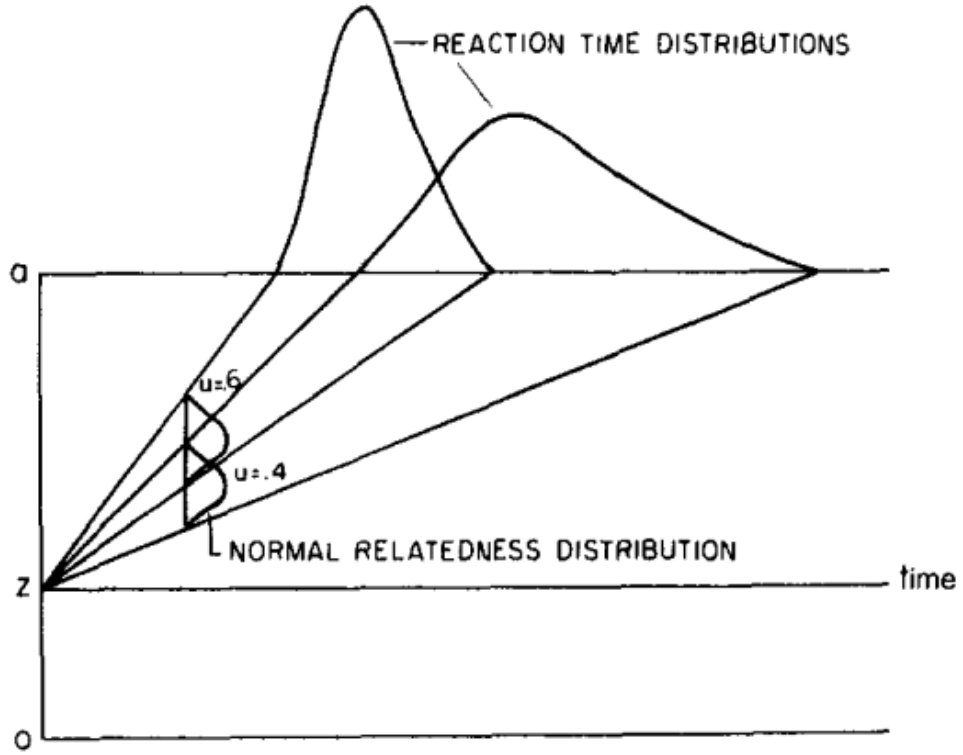
反応時間(RT)の分布の形を予測できるということ

仮に$s^2 = 0$（trial内分散がゼロ）だとすると…

drift rateは開始時点で$N(u, \eta^2)$からランダムに発生した値に固定されるので，RTの分布が描きやすい

- relatednessが高いほど，RTは早く分布の裾は軽い
- relatednessが低いほど，RTは遅く分布の裾も重い

いわゆる”Speed-accuracy tradeoff”については後ほど…



Decision Process

retrieval processでの答え方

はい　｜memory setのうち一つでもmatchしたら

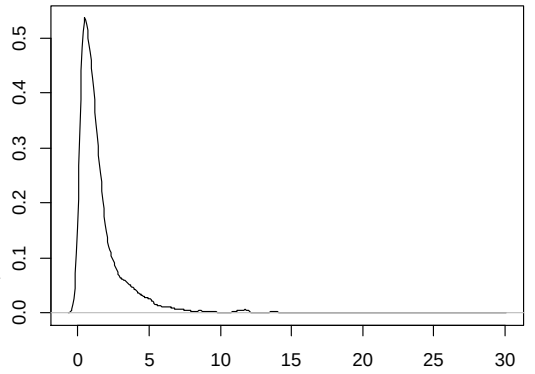
いいえ｜memory setが全てnon-matchだったら

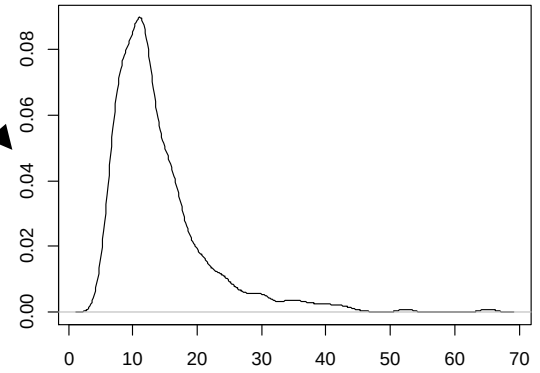
➡ memory set内のすべての単語との比較が並行している

- 各比較におけるRTの分布は，観察される実際のRTの分布と似てる
- 複数比較におけるRTの最大値の分布も，実際のRTの分布と似てる

つまり最後の比較が終わる▶「いいえ」と回答するまでのRTの分布

```
rlnorm(1000,0,1) %>% density %>% plot
```






```
rlnorm(100000,rnorm(100,0,0.2),1)
%>% matrix(ncol=100,byrow=T)
%>% apply(1,max) %>% density %>% plot
```

positiveプローブだけでなく，negativeプローブでも問題なくこのモデルはフィットする！

Decision Process

一般的に，”No”と回答するほうが”Yes”と比べて早い傾向がある

　これまでのモデルはここに対応できていないものもあった

Retrieval theoryなら無問題！

➡ もしstart pointが下のboundaryに近ければ，”No”のRTは”Yes”と比較して早くなるのは当然

※と元論文には書いてあるが，実際にはstart pointのvariabilityとtrial間のdrift rateの分散(η^2)によって初めてこの”fast error”を説明できる，というのが現在の一般的な認識

…start pointははじめから持っている「バイアス」みたいなものを表すので，これ自体を「errorのほうが早いか」みたいな意味で動かすのはどうなのか，という感じ？（よくわかってない）

Resonance Metaphor

memory set = 16のデータに対して，

relatednessの低い単語を追加しまくってみた

Correct Rejection

正答率もRTもほとんど変わらなかった

関連の低い単語がいくら増えたとしても

それらの影響はとても小さい

音叉みたいだね！

Table 1
Accuracy and Latency as a Function of the Number of Extra Fast Finishing Processes for Correct Rejections, with v = − .4, a = .12, z = .03, *and Search Set Size 16*

No. extra processes *m* (evoked set = *m* + 16)	No. standard deviations *r* (of the *m* process distribution) below the nontarget distribution	Proportion correct	Mean reaction time (in msec)
0	0	.839	358
100	2η	.838	363
500	3η	.838	362
5,000	4η	.838	358
25,000	5η	.838	363
100,000	5η	.835	358

Speed-Accuracy tradeoff

多くのモデルでは**正答率とRTが無関係である**という設定をおいている

…実際には「早く回答するためには精度を犠牲にする」ことは知られているのに

【Wickelgen(1977)によるSpeed-Accuracy tradeoffの分析方法】

a. 教示を与えて違いを見る（「早く答えて」「正確に答えて」）

b. 制限時間を設ける

c. RTを区切って，各区分ごとの正答率を比較する

➡ retrieval theoryにおいて，aとbがどのように表現されるかを見ます

※cはちょっと他と違ったりするらしく無視されました。

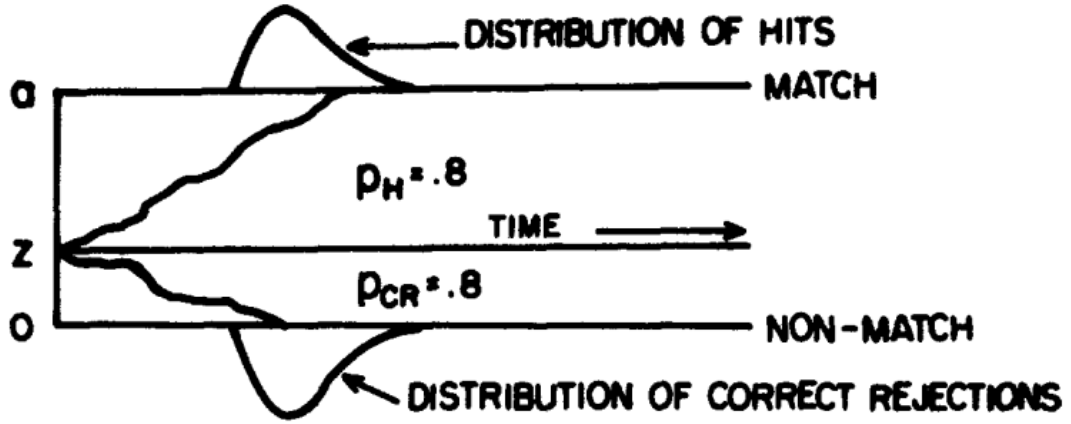
a. Information-controlled processing

教示によって**boundaryの位置が変わる**（このさきzの位置は変わらない）

もともとはこんなboundaryを上下に持っている

※p_H, p_{CR}はそれぞれHit, Correct Rejection率

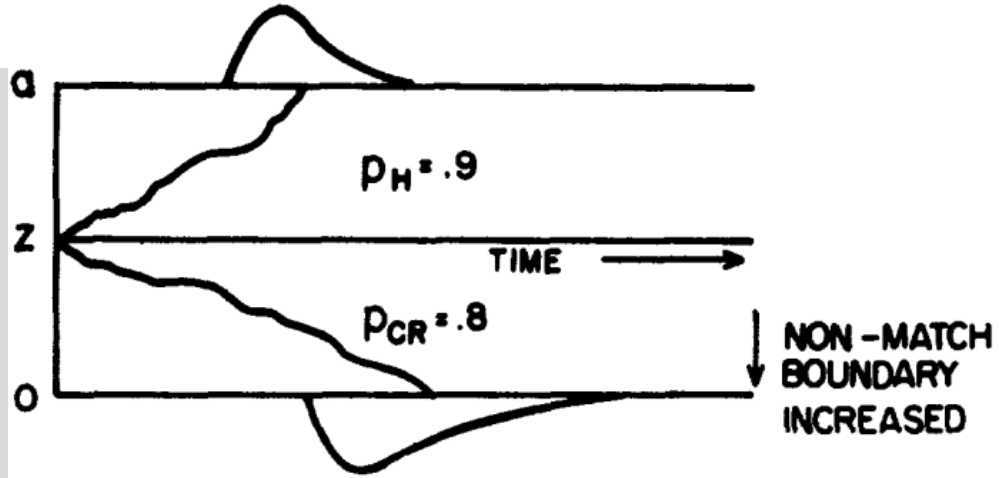
つまりそれぞれ上の正答率，下の正答率



下のboundaryが広がると…

- 下のRTが長くなる，分布が広がる
- p_Hが高くなる＝間違えて”No”と答える確率が下がる

(本当はp_{CR}が少し下がるが微々たるもの)



a. Information-controlled processing

教示によって**boundaryの位置が変わる**（このさきzの位置は変わらない）

もともとはこんなboundaryを上下に持っている

※p_H, p_{CR}はそれぞれHit, Correct Rejection率

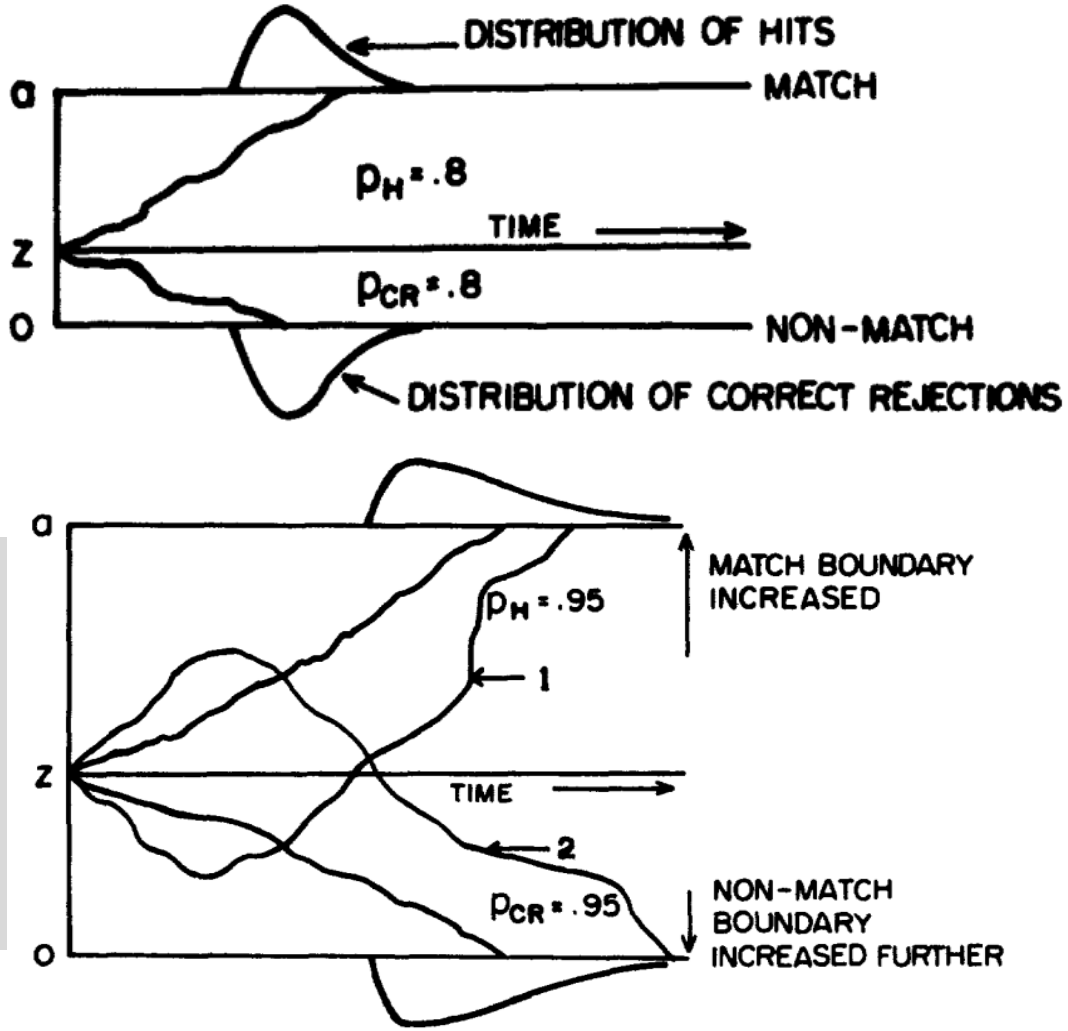
つまりそれぞれ上の正答率，下の正答率

上下のboundaryがどちらも広がると…

- どちらもRTが長くなる，分布が広がる
- p_H, p_{CR}が両方とも高くなる

＝間違いが減って精度が上がる

※右図1,2は「boundaryが広がったことでなくなったミス」



b. Time-controlled processing

制限時間が来たときに…

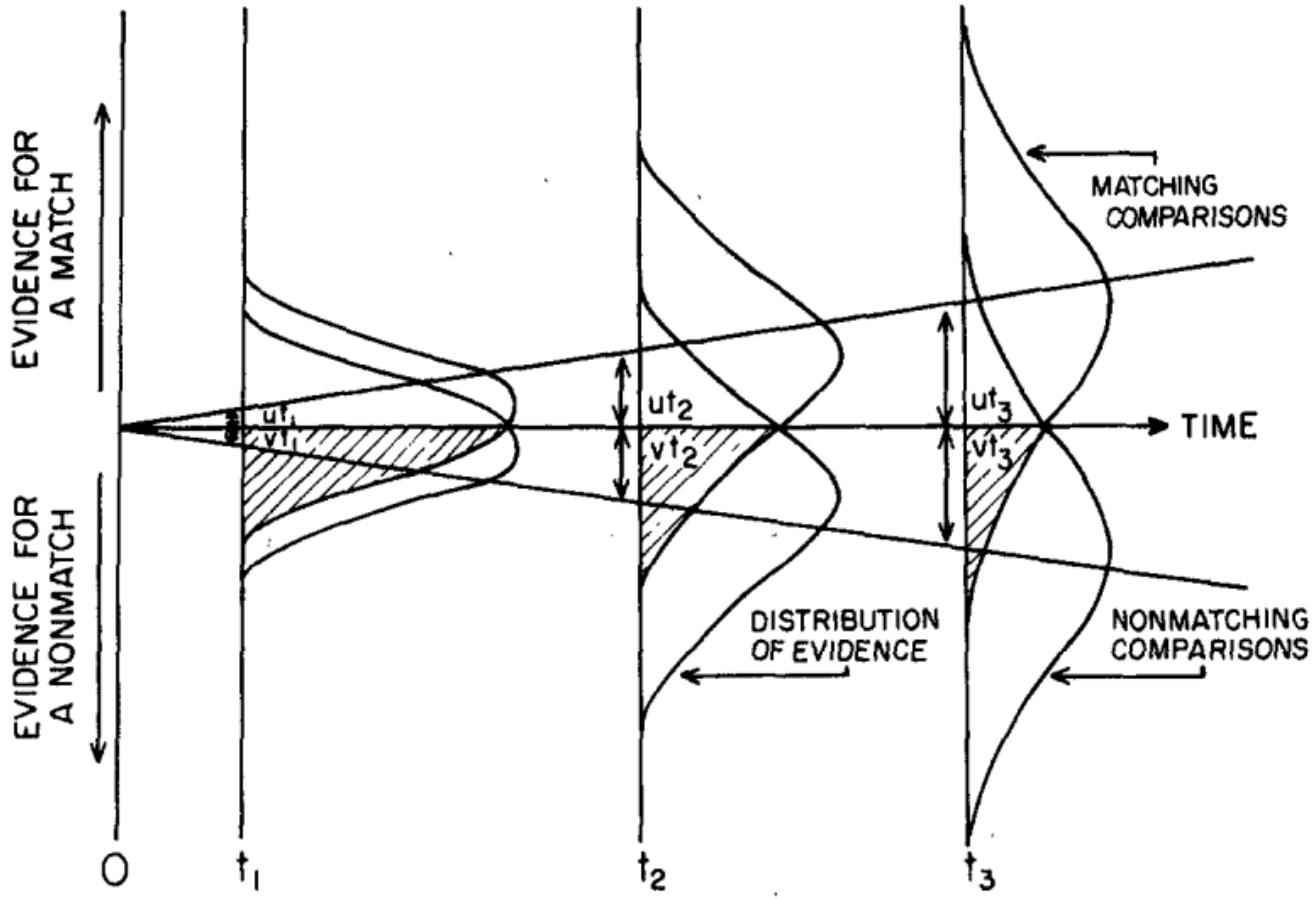
start pointより上にいる ▶ “Yes”

start pointより下にいる ▶ “No”

時間t_1の時点では，match課題に対して”No”する確率が高め（網掛け）

時間がt_3まで行くと，match課題に”No”する確率はだいぶ小さくなっている

制限時間が短いほど精度が低下している



1.2

数理モデル

Diffusion Process

パラメータは~~6つ~~ 5つ

① relatednessの分散｜η^2

② drift rateの分散｜s^2

③ 上の閾値｜a

④ 開始地点｜z

⑤ matchのときのdrift rate｜u

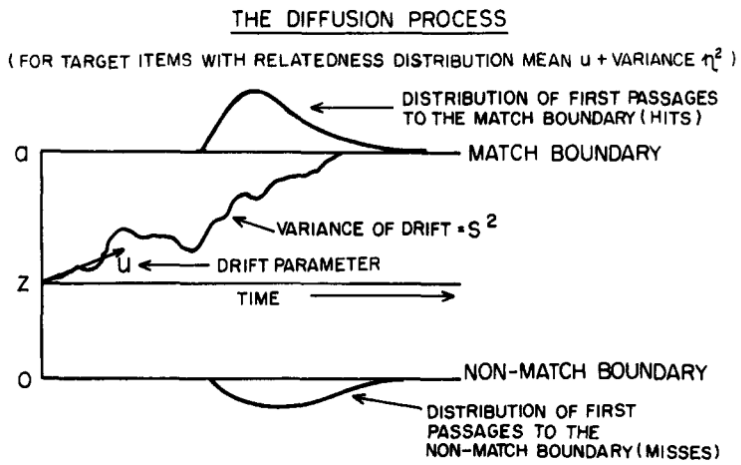
⑥ non-matchのときのdrift rate｜v

＋ 符号化の時間｜T_{er}

推定の際には固定される

課題ではなく教示や制限時間などで規定される

課題によって規定される



ここからはAppendixに従って式を追っていきます

ギャンブラーの破産 (gambler's ruin)問題

- 確率pで勝利することができ，確率$q(=1-p)$で負けてしまう
- 勝つと相手から１ドルもらえて，負けると１ドル払う
- ２人は合計でAドル持っていて，一方ははじめZドル持っていた
- 一方が総取りしたら（Aドルか０になったら）ゲーム終了

次スライドで解が出ますが

「勝率が50%未満だと，（漸近的に）ほぼ確実に破産するよ」

という例としてよく取り上げられるものです。

ここで「Zから開始して最終的に破産する確率」をq_Zとおいてみると

$$q_Z = pq_{Z+1} + qq_{Z-1}, \qquad \text{(A1)}$$

１回目に勝ったが，
その後破産する確率

１回目に負けて，
その後破産する確率

この漸化式 (A1)を解くと

$$q_Z = \begin{cases} \dfrac{\left(\frac{q}{p}\right)^A - \left(\frac{q}{p}\right)^Z}{\left(\frac{q}{p}\right)^A - 1}, & \text{when} \quad q \neq p \\ 1 - \dfrac{Z}{A}, & \text{when} \quad q = p. \end{cases} \qquad \text{(A3)}$$

同様に「Zから開始してn回目に破産する確率」を$g_{z,n}$とおくと，

$$g_{z,n+1} = pg_{z+1,n} + qg_{z-1,n}. \qquad \text{(A4)}$$

１回目に勝ったが，その後
ちょうどn回で破産する確率

１回目に負けて，
その後ちょうどn回で破産する確率

この漸化式(A4)を解くと

$$g_{z,n} = \frac{2^{n+1}}{A} p^{(n-z)/2} q^{(n+z)/2} \times \sum_{k<A/2} \cos^{n-1}\frac{\pi k}{A} \sin\frac{\pi k}{A} \sin\frac{\pi Zk}{A}. \qquad \text{(A6)}$$

※この導出はFeller (1968) An Introduction to Probability Theory and Its Applications, Vol. 1にて

(A3)と(A6)は当然ひっくり返すことが可能

pとqを入れ替えて，Zと$A-Z$も入れ替えれば**勝利する確率**として見ることができる

Diffusion process

Random walkは「特徴を一つずつ比べる」的モデルだったが，

個人(Ratcliff)的にはdiscreteではなくcontinuousと考えたい

1. information accumulationは連続的だろうと思ってるから
2. drift rateの平均と分散がindependentになるから
3. 離散変数の総和を取るより，連続変数の数値積分をするほうが，計算が楽だから

Diffusion processはRandom walkを連続的に拡張したもの

１ステップの大きさ（δ）を限りなくゼロに近づける

単位時間あたりのステップ数（r）が無限にあると考える

Diffusion process

$(p - q)r\delta \to \xi, \quad 4pqr\delta^2 \to s^2.$ (A7) という極値をとるようにしてあげると，

(A3)についての極限は

drift rateとか
relatednessとか

$$\gamma_-(\xi) = \frac{e^{-(2\xi a/s^2)} - e^{-(2\xi z/s^2)}}{e^{-(2\xi a/s^2)} - 1}, \qquad \text{(A8)}$$

(A6)についての極限は

よくWiener分布と言われるやつ

$$g_-(t, \xi) = \frac{\pi s^2}{a^2} e^{-(z\xi/s^2)} \times \sum_{k=1}^{\infty} k \sin\left(\frac{\pi z k}{a}\right) e^{-\frac{1}{2}(\xi^2/s^2 + \pi^2 k^2 s^2/a^2)t}. \qquad \text{(A9)}$$

ちなみに

先ほどと同じ感じでξを$-\xi$に置き換えて，zを$a-z$に置き換えると$\gamma_+(\xi)$や$g_+(t,\xi)$が得られるが，$\gamma_-(\xi)+\gamma_+(\xi)=1$を満たすことを考えると

$\gamma_+(\xi)=\frac{e^{-(2\xi z/s^2)}-1}{e^{-(2\xi a/s^2)}-1}$となる。ここで，$\left(z=\frac{\mathrm{a}}{2},s=1\right)$とおくと

$$\gamma_+(\xi)=\frac{e^{-\xi a}-1}{e^{-2\xi a}-1}=\frac{e^{\xi a}-e^{2\xi a}}{1-e^{2\xi a}}=\frac{e^{\xi a}(1-e^{\xi a})}{(1+e^{\xi a})(1-e^{\xi a})}=\frac{e^{\xi a}}{1+e^{\xi a}}$$

ここで$\xi=\theta-b$とおくと，まごうことなき2PLM $\frac{e^{a(\theta-b)}}{1+e^{a(\theta-b)}}$

こうしてDiffusion-IRTが生まれました

Diffusion process

【注意】

$g_+(t,\xi)$や$g_-(t,\xi)$は，コレ自体が確率分布ではない

すべてのtについて積分したものは１ではなく$\gamma_+(\xi)$になるため

➡ 確率分布にするためには$\frac{g_+(t,\xi)}{\gamma_+(\xi)}$および$\frac{g_-(t,\xi)}{\gamma_-(\xi)}$とすればよい

Decision process

memory set内の一つ一つの単語とプローブの比較がdiffusion process

⟶すべてのdiffusion processの総意としてのDecision processを考える

ある比較iについてnon-matchとなる累積確率分布は $$G_-(t, \xi) = \int_0^t g_-(t', \xi)dt'$$

実際に”No”と反応する（までのRTの累積）確率は

$$\gamma_-(\xi) = \prod_{i=1}^{n} \gamma_{i-}(\xi), \qquad G_{\max}(t, \xi) = \prod_{i=1}^{n} G_{i-}(t, \xi)$$

疲れました

このあと論文では

- 他のモデルとの関係
- 有名な実験パラダイムにこの理論を当てはめてみた
- neural network, semantic memory model, propositional modelsとの関係

と続きますが，気になった人は読んでみてください

1.3

Diffusion modelの仲間たち

EZ-diffusion

Wagenmakers, E. J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22.

Diffusion modelはすごいけど，あまり使われていなかったらしい

- パラメータの数が多すぎる
- 正答と誤答の両方のRT分布がないとうまくいかない▶結構データが難しい

  特に認知心理の実験とかでは極端に誤答が少なかったりする

➡ **心理学的に本当に必要なパラメータに絞って簡略化します**

v ｜与えられる情報の蓄積量

a ｜回答の慎重さなど

T_{er} ｜符号化の時間

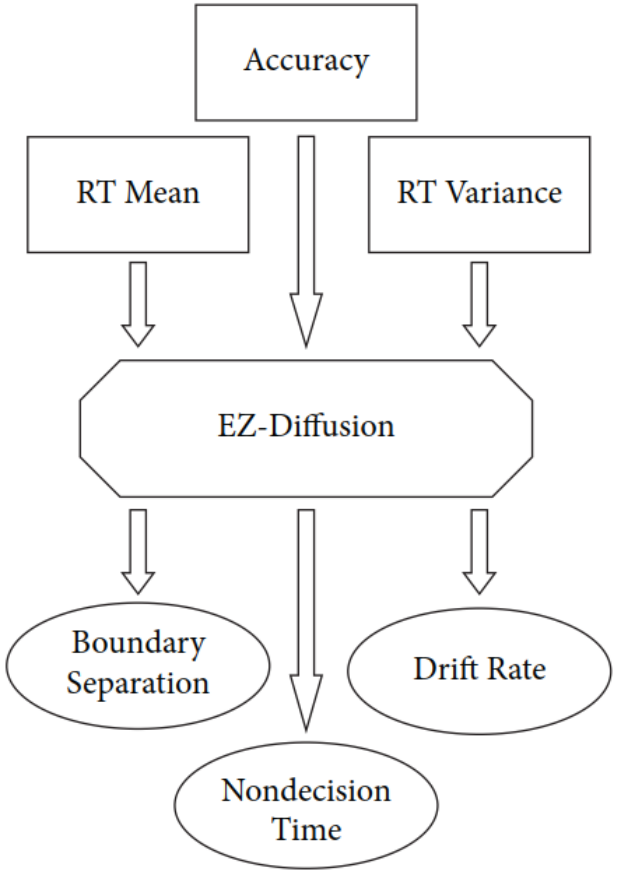
trial間で必ず同じ値になる（分散η^2などがゼロ）

※s^2はゼロになったわけではなく，スケーリングパラメータなので任意の値にfixされている

start pointはちょうど中間にある（$z = a/2$）

EZ-diffusion

RTも，分布をすべて使うのではなく「正答時RTの平均，分散」のみを使う



「正答率」「正答時RTの平均」「正答時RTの分散」
の３つをつかって３つのパラメータを求める

パラメータを「推定」する必要はなく，ただ
連立方程式をとけば良い

あとはざっと紹介

途中で刺激が変化するような課題の場合

drift rateがある時間で変化する

時間によるプレッシャーがある場合

boundaryがどんどん狭くなっていく

２択ではなく３択以上にしたい場合

各選択肢がそれぞれdiffusion processに従い，最初に閾値についた反応をする

▶他の選択肢による抑制があったりなかったり

最近の動向を知りたければ
Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281.

多肢選択ならもっと簡単なモデルが既に提案されているよ！

56.2. Wiener Diffusion Model Distributions

Probability Density Function

If $\alpha \in \mathbb{R}^+$, $\tau \in \mathbb{R}^+$, $\beta \in [0,1]$ and $\delta \in \mathbb{R}$, then for $y \in (0,\tau)$,

$$\text{Wiener}(y|\alpha,\tau,\beta,\delta) = \frac{\alpha}{(y-\tau)^3/2} \exp\left(-\delta\alpha\beta - \frac{\delta^2(y-\tau)}{2}\right) \sum_{k=-\infty}^{\infty} (2k+\beta)\phi\left(\frac{2k+\alpha\beta}{\sqrt{y-\tau}}\right)$$

where $\phi(x)$ denotes the standard normal density function (Blurton et al., 2012).

Sampling Statement

`y ~ wiener(alpha, tau, beta, delta);`
Increment log probability with `wiener_lpdf(y | alpha, tau, beta, delta)`, dropping constant additive terms; Section 5.3 explains sampling statements.

2

LBAって何よ

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178.

自信満々

We hope to advance this research effort with a new theory of choice RT—the linear ballistic accumulator (LBA). The LBA is simpler than the leading models in the field, and yet it accommodates all the important empirical phenomena. The LBA also comes with detailed but simple analytic solutions, making it easy to apply. The model's success in accounting for empirical phenomena is surprising, given it shares essential properties with older models that have proven inadequate. The LBA uses linear, independent response

【LBAのいいところ】

- simpler than the leading models in the field
- it accommodates all the important empirical phenomena
- comes with detailed but simple analytic solutions, making it easy to apply

そんなわけで

The model's success ...(略)... is surprisingなのです。

2.1

LBAの紹介

なにはともあれ

従来のモデル

Sequential Samplingと呼ばれる一連の枠組みでは

- 刺激が提示される▶二択で回答を求められる
- 情報の蓄積量が時間ごとに異なる
- 閾値に到達したら反応が生じる
- プロセスは**確率的**である

  このおかげでRTが分布を描けるようになり，全く同じ刺激でも毎回異なるRTが生じる

RTのすべての性質を表現しているモデルたちは，情報蓄積のプロセスに加え何かしらを付け加えることでこれを実現している

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*.

２つの選択肢のdrift rateが時間で減衰しながら閾値を目指す

【追加されているもの】

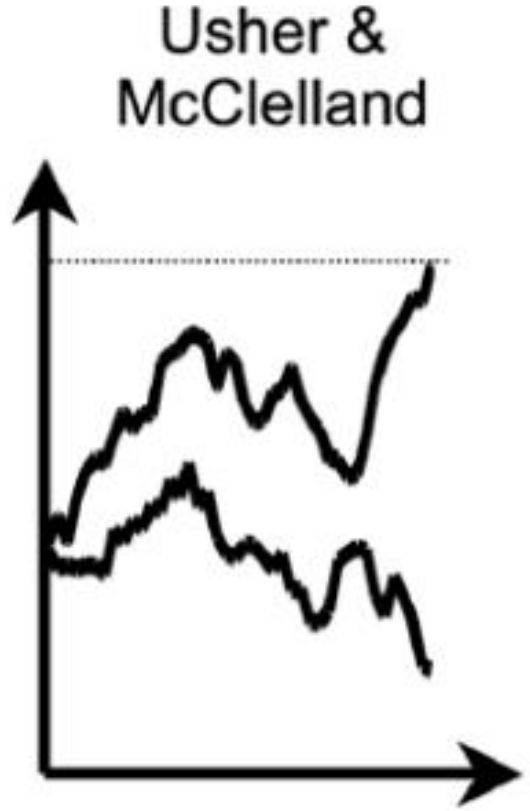
２つのVariability

- start pointが各選択肢ごとに，trial間で異なる
- drift rateも，trial間で異なる（ηのこと）

２つのnonlinear process

- passive decay of accumulated evidence
- response competition

▶時間ごとに減衰する２つが競争する



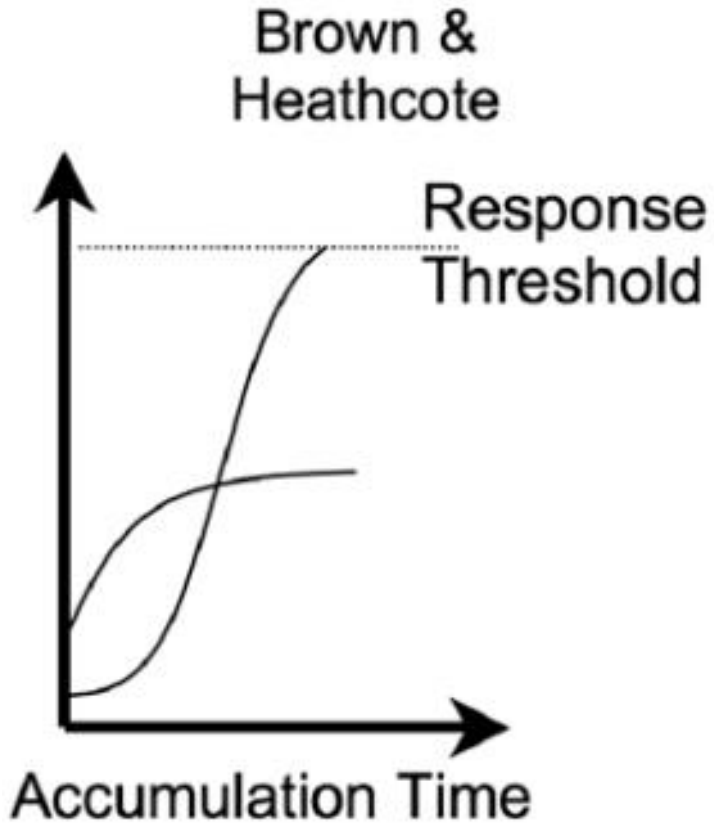
Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*(1), 117–128.

LCAを簡略化したモデル

LCAが持っていた４つの追加要素をすべて保持

drift rateのtrial内分散(s)を取っ払った

※毎時間ごとに$N(0, \sigma_I)$から発生したノイズだけdrift rateが変動するので真っ直ぐではなく曲線になっている



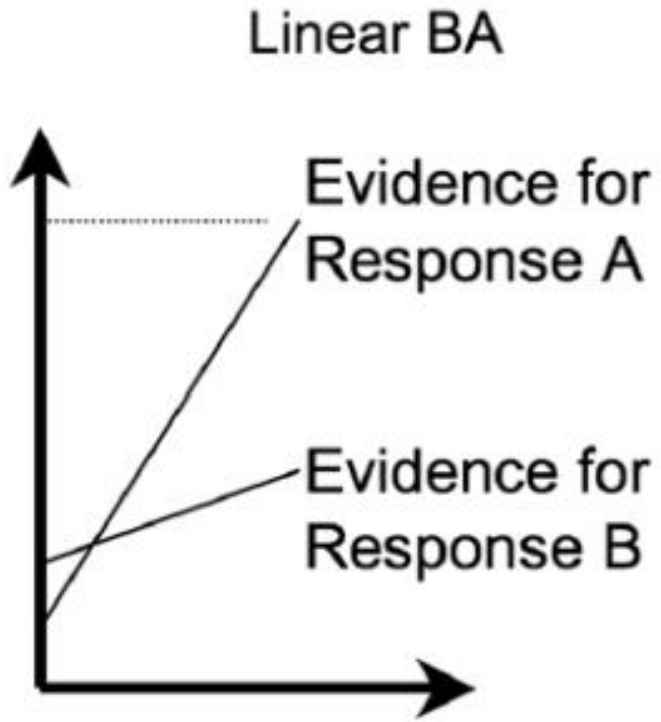
LBA

BAをさらに線形にしてより簡略化を目指したモデル

一度drift rateが決まったら突っ走るだけ

independentかつlinearなaccumulatorというのは珍しい仮定

➡ RTに関するすべての情報を表現することはできている



２つの大きなメリット

- RTと正答率の分布が解析的に出せる（わりとかんたんに）
- independentなので選択肢数が３以上でもいくつでも問題なし

2.2

Simpler models of RT

単純化するモデルたち

EZ-diffusion

すごいシンプル，なにせtrial内分散しか残っていないので

➡ **RTに関するすべての特性を表現できるわけではない**

例えばEZ-diffusionではcorrectとerrorにおけるRT分布は同じである

そもそも提案者らも”process model”よりは”descriptive model”として提案

分布ではなく，それを要約したものについてのモデルであるということ

つまりRTのすべてを表現するつもりもなく「とにかく単純なモデル」を作った

LBAはシンプルだがcompleteなモデルだからすごい！

LATER；variable criterion；random-ray models

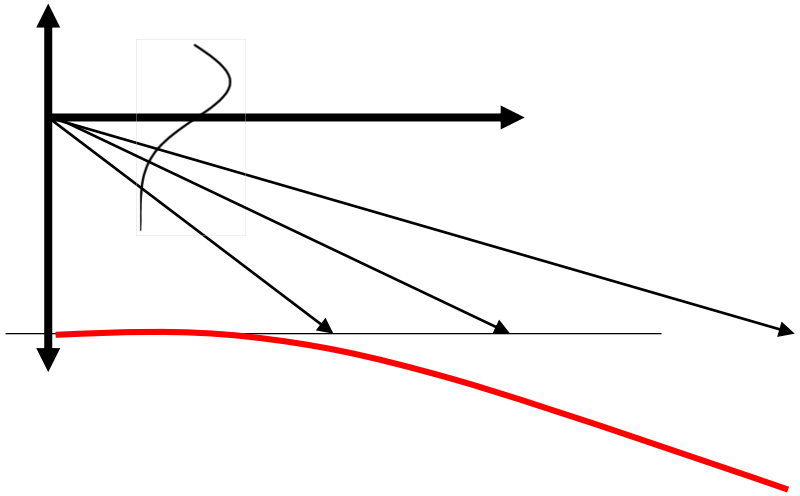
random walkをベースにしているがtrial内分散（s）がない

共通して抱えている問題点が

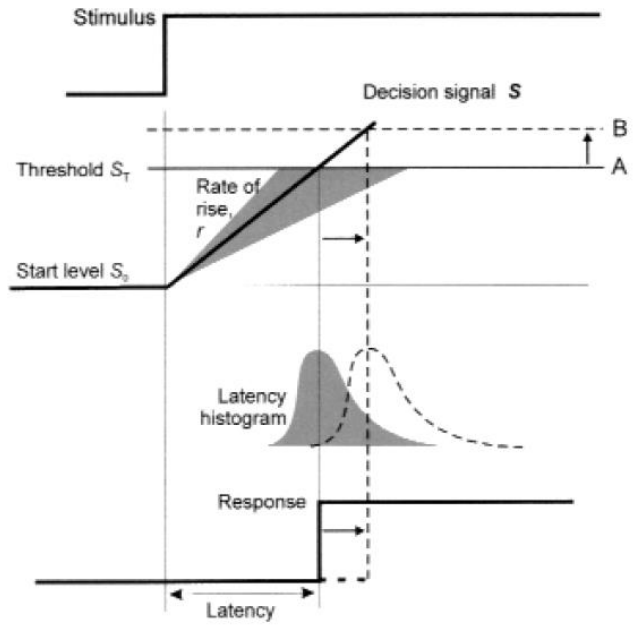
誤答反応RTが負に歪んでいる（裾が左に長い）ということ

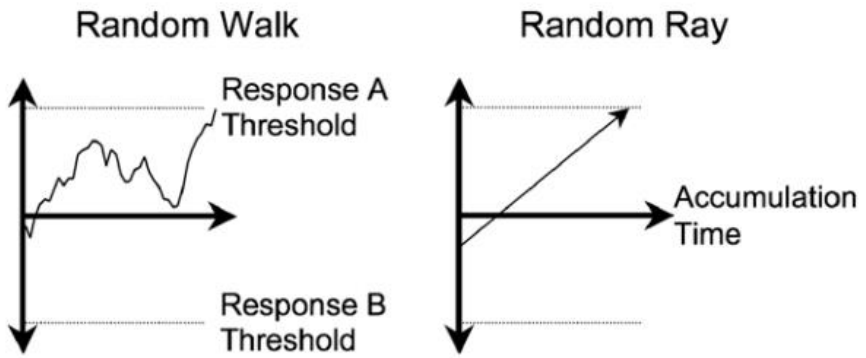
現実では，RTは常に右に長い裾を持っているはず

…drift rateの平均が正のとき，誤答反応が生じるdrift rateの側には分布のピークが来ない



LATER
Reddi & Carpenter (2000)





やっぱりLBAはすごいという話

ただ単にtrial内分散を取り除くと，誤答RT分布がおかしくなる

が，LBAは**選択肢ごとにaccumulatorを置く**ことで解決した

➡ そのうえで「accumulatorが下に行かない」とすれば，分布は同じに

ここまでのsimplificationはいわば「オッカムの剃刀」であった

できる限り単純化することを目指すが，データの現象を説明できなくなるほど削ぎ落としたら良くないわけです

LBAの”simplicity”

一概には言えないが「パラメータ数」は一つの指標

K個の刺激があるときに，RTの特性をすべて説明するのに必要な数は…

Diffusion modelはk+5か6個

BAはk+6個

LBAはk+4個！

※具体的に各モデルで何を計上してるかは書いてありません。たぶんkがdrift rateだと思いますが，それ以外は何だろうか…？

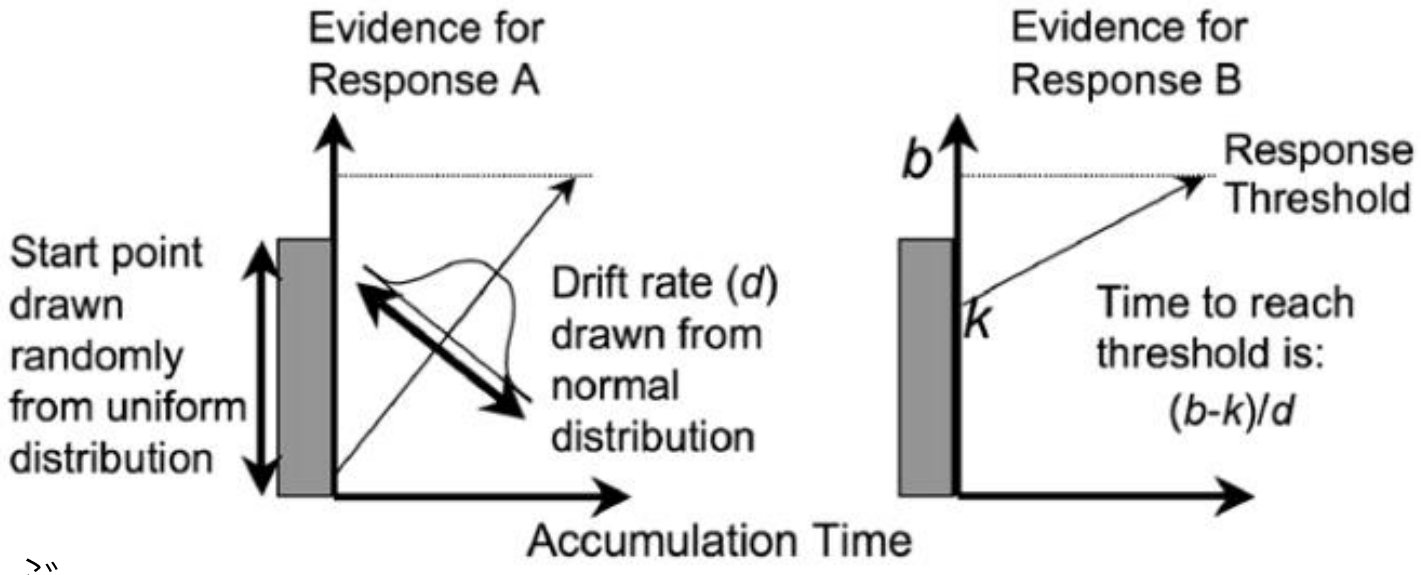
2.3

The linear ballistic accumulator

モデルの全貌が明らかに

これがLBAだ！

一様分布に基づく
基点から同時に出発
↓
時間がたつと
情報が蓄積される
↓
最初に閾値にたどり着いた選択肢をえらぶ



※LBAの模式図ではよく一つのグラフでいくつかの線が競争してる的な図を見ることが多いですが，そもそもindependent accumulatorなので，このように別々に書いても何ら構わない。
物によってはstart pointの上限がaccumulatorごとに異なる場合もあり，そういったときはこうしたほうが見やすいかも。

これがLBAのパラメータ だ！

重要なパラメータ

Drift rate (d)

…平均的な移動方向（選択肢ごと）

Boundary (b)

…閾値の高さ

広いと「慎重に」狭いと「素早く」

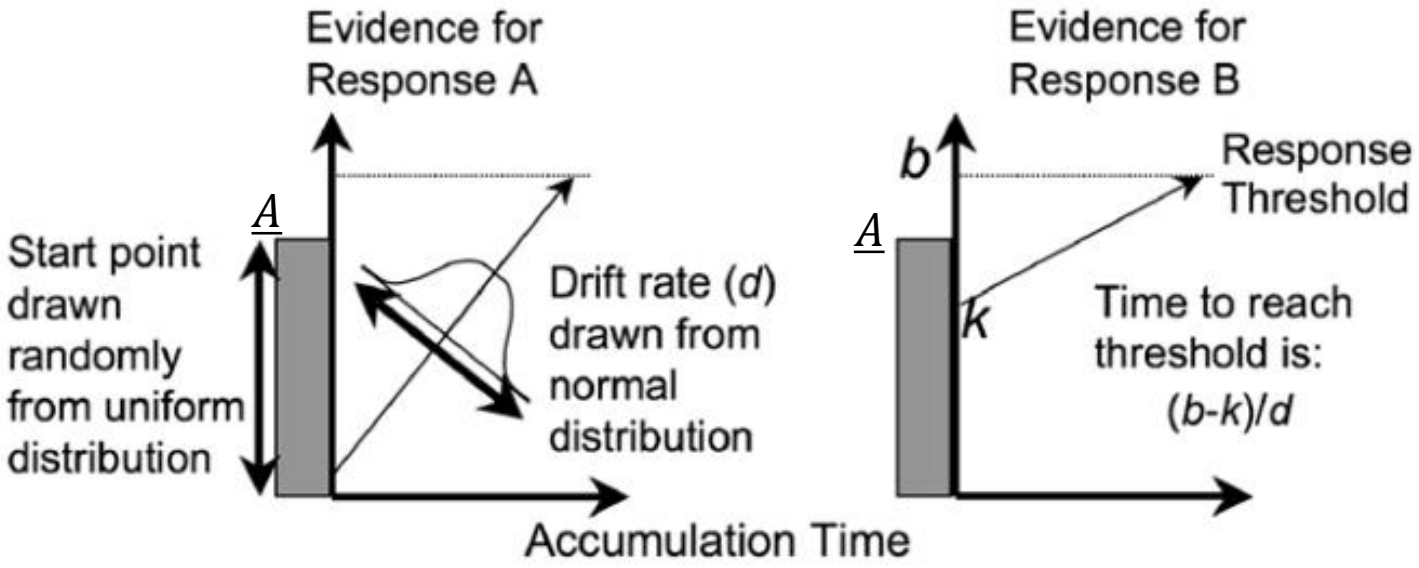
Start Point (A)

…開始点の高さ

高いと「素早く」低いと「慎重に」

Non-decision Time (t_0)

…無反応時間



式の導出

start pointは$U[0, A]$からランダムに選ばれた値

⟹ start pointからboundaryまでの距離は$U[b-A, b]$　▶pとおく

また，drift rateは$N(d = v, s^2)$からランダムに選ばれた値　▶qとおく

「時間tまでに選択肢１がboundaryに到達している確率」は

$$
\begin{aligned}
F_1(t) &= \text{prob}\left(\frac{p}{q} < t\right) \\
&= \text{prob}(p < qt) \quad \# \text{ assumes } q > 0. \\
&= \int_{-\infty}^{\infty} U(u|b-A, b)\phi(u|tv_1, ts)\,\mathrm{d}u \\
&= \int_{b-A}^{b} \frac{u-b+A}{A}\phi(u|tv_1, ts)\,\mathrm{d}u + 1 - \Phi(b|tv_1, ts)
\end{aligned}
$$

式の導出

そんな感じでひたすら式変形していくと，気づいたら

$$F_i(t) = 1 + \frac{b - A - tv_i}{A}\Phi\left(\frac{b - A - tv_i}{ts}\right) + \frac{tv_i - b}{A}\Phi\left(\frac{b - tv_i}{ts}\right) + \frac{ts}{A}\phi\left(\frac{b - A - tv_i}{ts}\right) - \frac{ts}{A}\phi\left(\frac{b - tv_i}{ts}\right)$$

これを時間tによって微分して式変形しまくると

$$f_i(t) = \frac{1}{A}\left[-v_i\Phi\left(\frac{b - A - tv_i}{ts}\right) + s\phi\left(\frac{b - A - tv_i}{ts}\right) + v_i\Phi\left(\frac{b - tv_i}{ts}\right) - s\phi\left(\frac{b - tv_i}{ts}\right)\right]$$

先程のは「ある選択肢がboundaryに到達する」に関する分布

※boundaryに到達しないこともあり得るので確率分布ではない

興味があるのは「ある選択肢が**<u>最初に</u>**boundaryに到達する」の分布

$$\mathrm{PDF}_i(t) = f_i(t) \prod_{j \neq i} \left(1 - F_j(t)\right)$$

$\mathrm{PDF}_i(t) =$ 時間tでiが閾値にたどり着く確率 $f_i(t)$ × 時間tまでi以外が閾値にたどり着かない確率 $\prod_{j \neq i} \left(1 - F_j(t)\right)$

LBAの仮定

「少なくとも１つは正のdrift rate（実現値が）である」

正である限りは時間さえかければboundaryに到達するが，負だと一生到達しない

実際に反応が観測されているということは，１つは正じゃないとおかしい

drift rateは$N(v, s^2)$から発生しているので，確率$\Phi(-\frac{v}{s})$で負になる

よって，すべてが負になる確率は$\prod_{n=1}^{N} \Phi(-\frac{v_n}{s})$

尤度などを計算する際には，$\frac{\text{尤度}}{\left(1-\prod_{n=1}^{N} \Phi\left(-\frac{v_n}{s}\right)\right)}$として調整する

※そもそもdrift rateが対数正規分布やガンマ分布から発生すると考えることも可能

Terry, A., Marley, A. A. J., Barnwal, A., Wagenmakers, E. J., Heathcote, A., & Brown, S. D. (2015). Generalising the drift rate distribution for linear ballistic accumulators. *Journal of Mathematical Psychology*, *68–69*, 49–58.

Fast and slow error

問題が簡単なとき，「早く回答して」と教示があるとき▶誤答の方が早い

難しいとき，「正確に回答して」と教示があるとき▶誤答の方が遅い

誤答の方が早い／遅いの両現象を説明するためには２つのvariabilityが必要

- start point valiability ($U[0, A]$)
- between-trial drift rate variability (s)

Fast error

「なるべく早く答えてください」と言われたとき

➡boundaryはstart pointの上限に近づく

① start pointの上の方から始まった場合

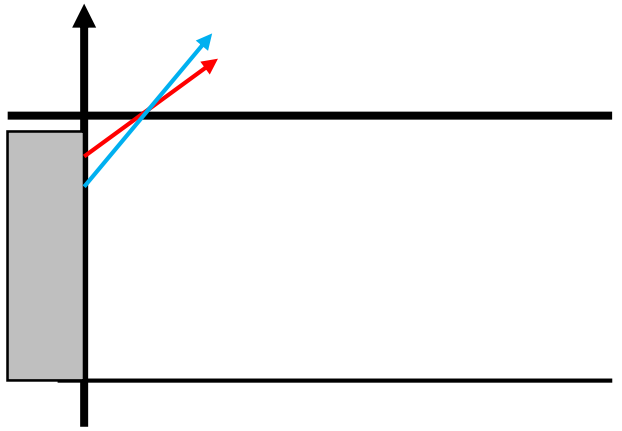
drift rateの違いよりも「どちらが上から始まったか」で決まる

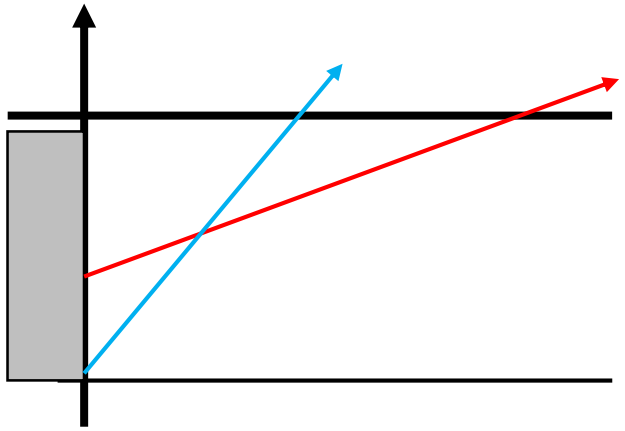
➡早いRTについては結果はランダムに近い

② start pointの下の方から始まった場合

drift rateの違いの分だけ”correct”が先に到達する確率が高い

➡遅いRTについては正答率が高い





まとめると「早いときの方がerrorが多い」ことになる

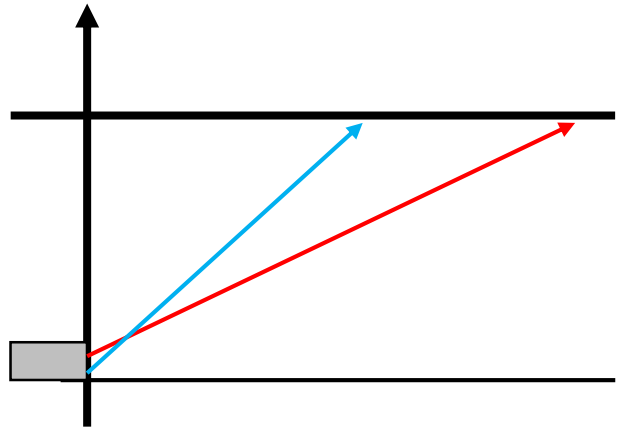
Slow error

「なるべく正確に答えてください」と言われたとき

➡ boundaryはstart pointの上限から離れる

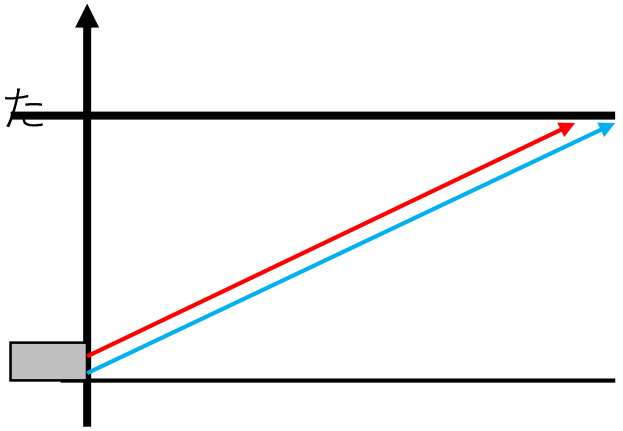
① 正解するとき

大抵はdrift rateが高い”correct”が先に到達する



② 誤答するとき

正規分布から発生するので，たまたま”correct”のdrift rateが低かった



結果的に「errorのほうがRTが長い」ことになる

2.4

Limitations of the model

完璧なんてないよ

1. 神経生理学の視点からツッコミ

ベースモデルのLCAを含めて，一般的なsequential sampling modelsは神経生理学的にも根拠があった

drift rateがnoisyであるというのがニューロンの「発火率」とか「膜電位」と合致

➡ **LBAはtrial内分散を排除した，という点は叩かれるかもしれない**

【数ある反論のうち２つ】

① そもそも科学のモデルって現象のSimplificationじゃないですか

② 神経生理学の世界ではtrial内分散を排除したモデルが使われてますよ

神経生理学の世界で認められているんだから，心理学の世界では「神経生理学的におかしい」とか言ってくれるなよ

2. 特殊な実験パラダイム

Linear accumulatorを使っているため，nonlinear な刺激には合わないかも

例｜ランダムドットキネマトグラム（moment-to-momentに変動）

➡ smoothingされてconstantなaccumulationになっている可能性

Independent accumulatorを使っているので，dependentには合わないかも

例｜時間ごとに減衰decayする，accumulator同士が抑制し合う

➡ 別の要因で原因を説明できることが多々ある

前半の刺激の影響が弱いというときに

- 時間ごとに減衰してるから弱くなっているんだ
- 最初はどうやらまだ集中してないようだ

stanだと

Behav Res (2017) 49:863–886
DOI 10.3758/s13428-016-0746-9

CrossMark

Bayesian inference with Stan: A tutorial on adding custom distributions

Jeffrey Annis[1] · Brent J. Miller[1] · Thomas J. Palmeri[1]

```
22  real lba_cdf(real t, real b, real A, real v, real s){
23
24      real b_A_tv;
25      real b_tv;
26      real ts;
27      real term_1;
28      real term_2;
29      real term_3;
30      real term_4;
31      real cdf;
32
33      b_A_tv <- b - A - t*v;
34      b_tv <- b - t*v;
35      ts <- t*s;
36      term_1 <- b_A_tv/A * Phi(b_A_tv/ts);
37      term_2 <- b_tv/A   * Phi(b_tv/ts);
38      term_3 <- ts/A     * exp(normal_log(b_A_tv/ts,0,1));
39      term_4 <- ts/A     * exp(normal_log(b_tv/ts,0,1));
40      cdf <- 1 + term_1 - term_2 + term_3 - term_4;
41
42      return cdf;
43
44    }
```


```
22  real lba_cdf(real t, real b, real A, real v, real s){
23
24      real b_A_tv;
25      real b_tv;
26      real ts;
27      real term_1;
28      real term_2;
29      real term_3;
30      real term_4;
31      real cdf;
32
33      b_A_tv <- b - A - t*v;
34      b_tv <- b - t*v;
35      ts <- t*s;
36      term_1 <- b_A_tv/A * Phi(b_A_tv/ts);
37      term_2 <- b_tv/A   * Phi(b_tv/ts);
38      term_3 <- ts/A     * exp(normal_log(b_A_tv/ts,0,1));
39      term_4 <- ts/A     * exp(normal_log(b_tv/ts,0,1));
40      cdf <- 1 + term_1 - term_2 + term_3 - term_4;
41
42      return cdf;
43
44    }
```