

Tutorials

Diffusion decision model

Rescorla-wagner model

Rescorla-wagner diffusion decision model

Bonus!-Maximum Likelihood Estimation

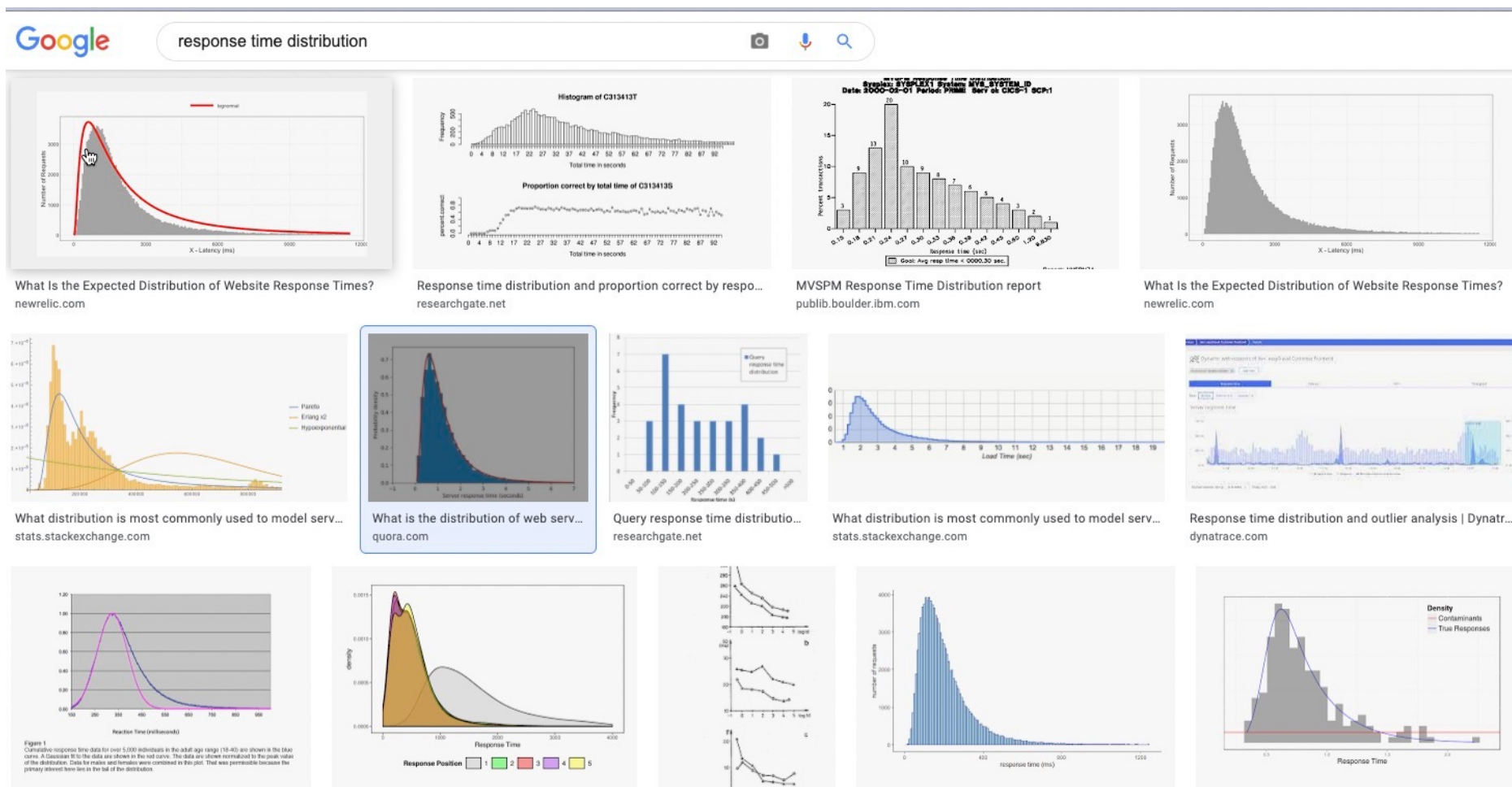
Xinyuan Yan

22th, May, 2021

xinyuanyan2016@gmail.com

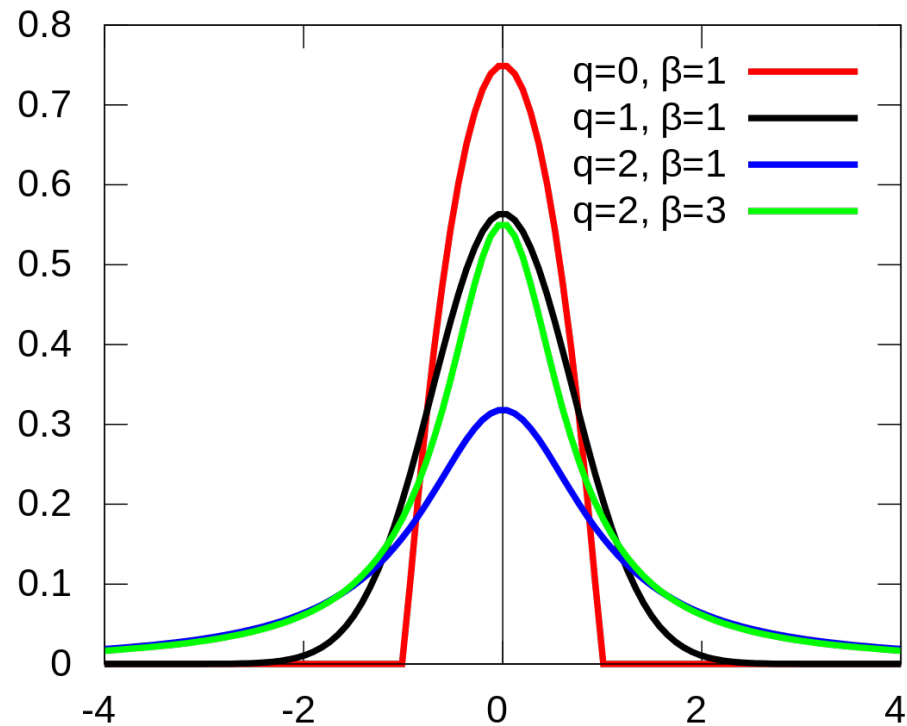
Diffusion decision model

Response time is a distribution! But what would be the best-fitted PDF?

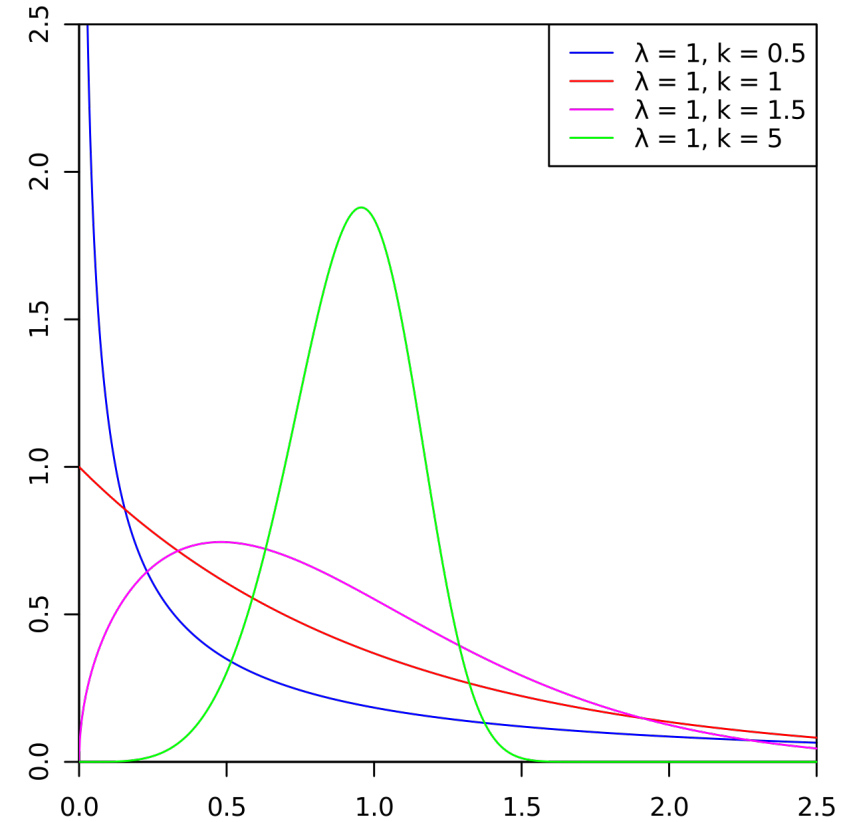


Response time is a distribution! But what would be the best-fitted PDF?

Gaussian distribution?

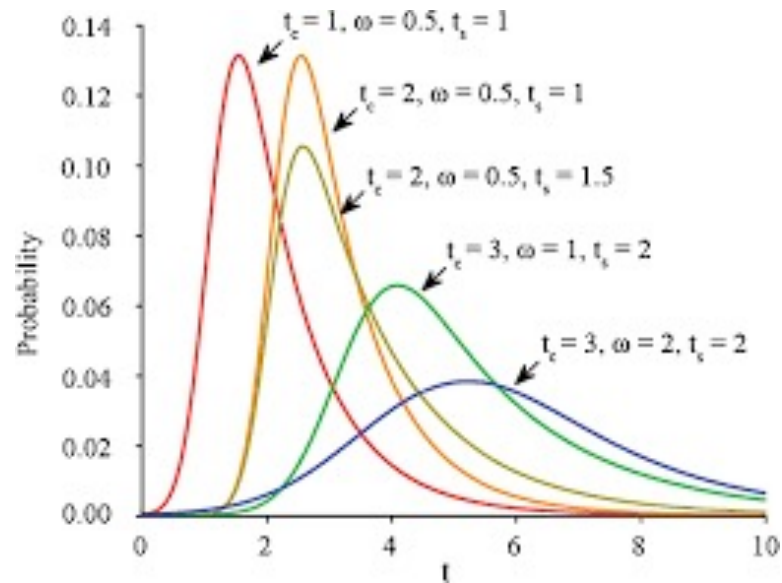


Weibull distribution?

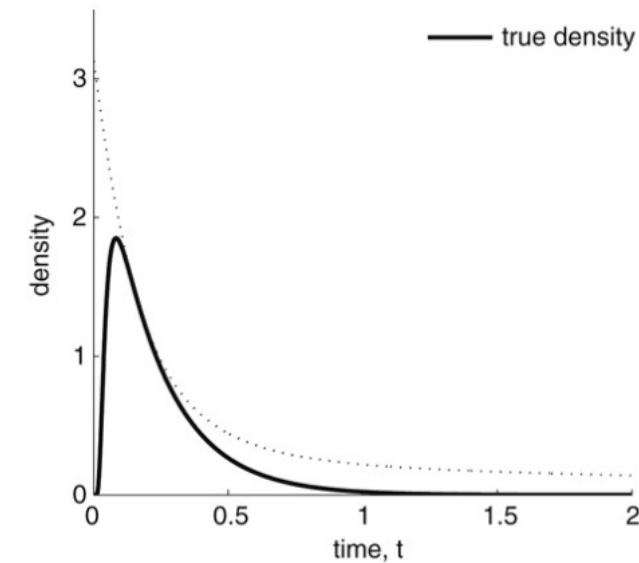


Response time is a distribution! But what would be the best-fitted PDF?

Ex-gaussian distribution



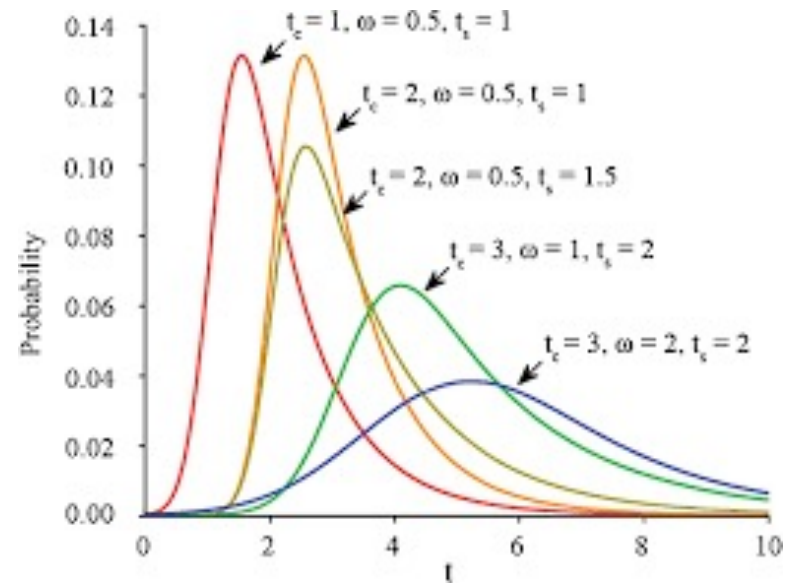
Wiener distribution



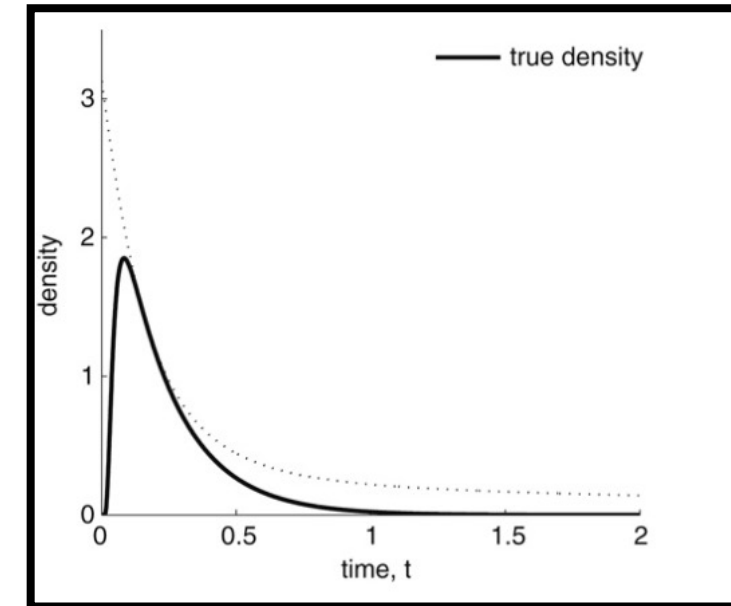
$$g(t | v, \eta^2, a, w) = \frac{1}{\sqrt{t^3 (1 + \eta^2 t)}} \times \exp \left[\frac{-v^2 t - 2vaw + \eta^2 (aw)^2}{2(1 + \eta^2 t)} \right] \times \sum_{j=0}^{\infty} (-1)^j r_j \phi \left(\frac{r_j}{\sqrt{t}} \right) \quad (1)$$

Response time is a distribution! But what would be the best-fitted PDF?

Ex-gaussian distribution

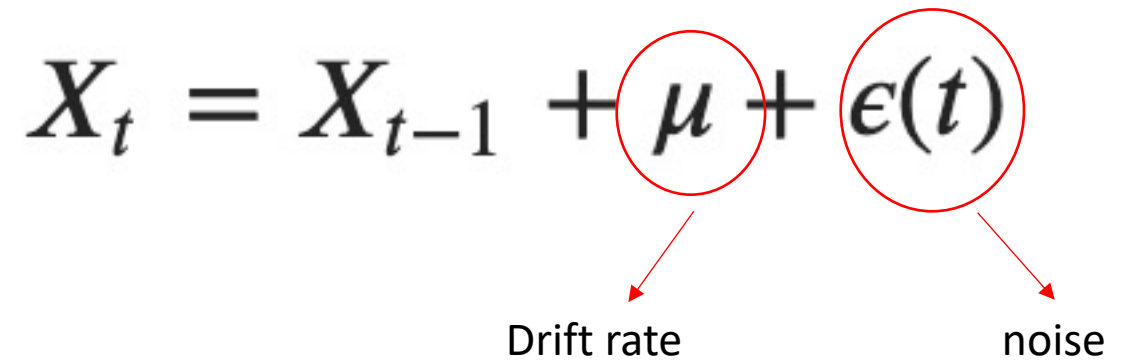


Wiener distribution



$$g(t | v, \eta^2, a, w) = \frac{1}{\sqrt{t^3 (1 + \eta^2 t)}} \times \exp \left[\frac{-v^2 t - 2vaw + \eta^2 (aw)^2}{2(1 + \eta^2 t)} \right] \times \sum_{j=0}^{\infty} (-1)^j r_j \phi \left(\frac{r_j}{\sqrt{t}} \right) \quad (1)$$

Wiener processes (Brownian motion), random walk

$$X_t = X_{t-1} + \mu + \epsilon(t)$$


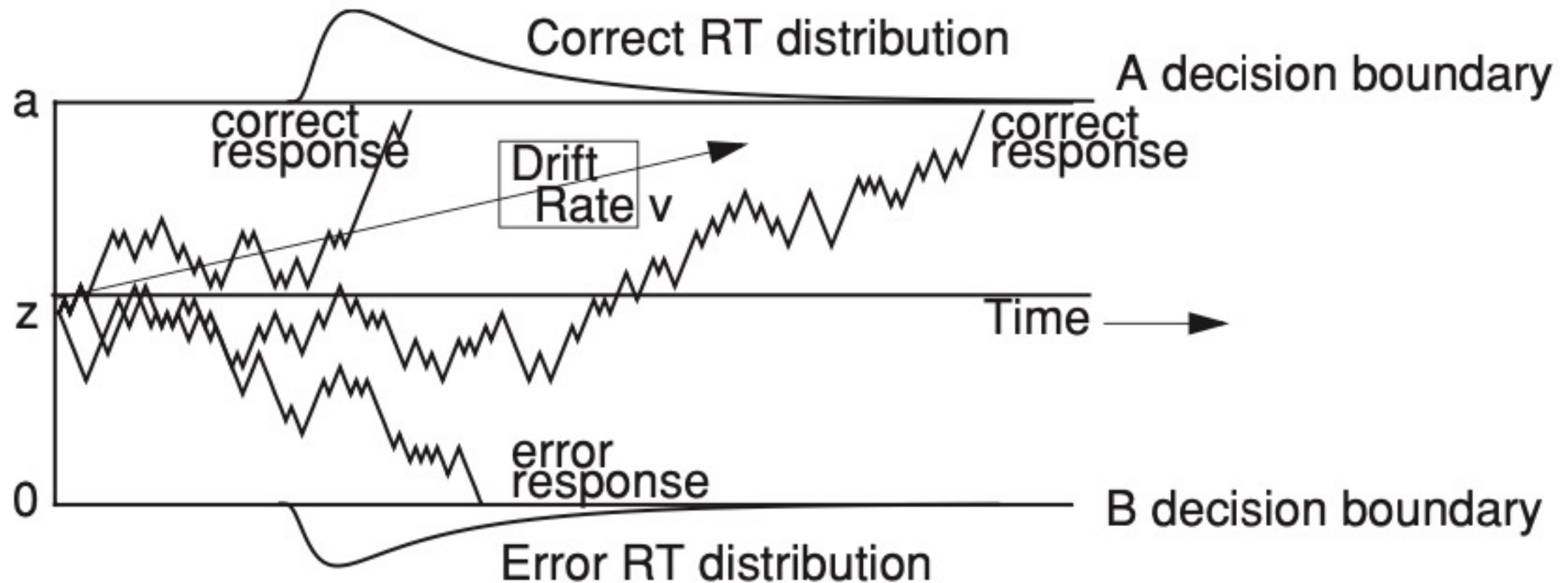
Drift rate

noise

$$\epsilon(t) \sim \text{Gaussian}(0, \sigma)$$

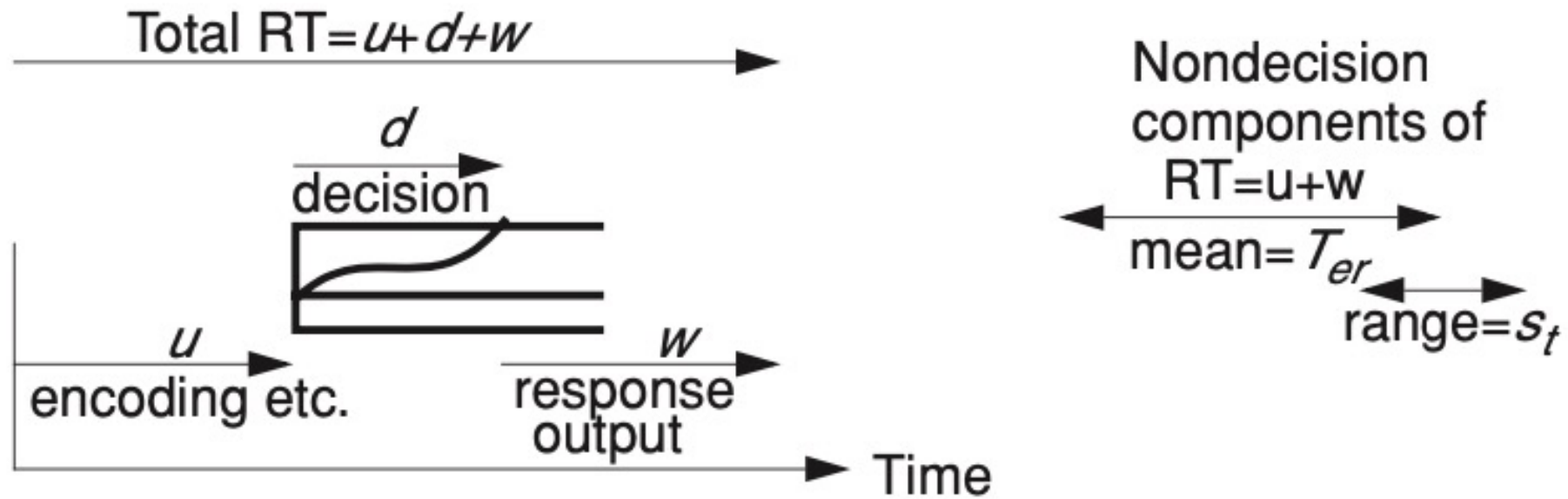
Diffusion decision model

Key free parameters in DDM



Diffusion decision model

Key free parameters in DDM



Key free parameters in DDM

d – *drift.rate*

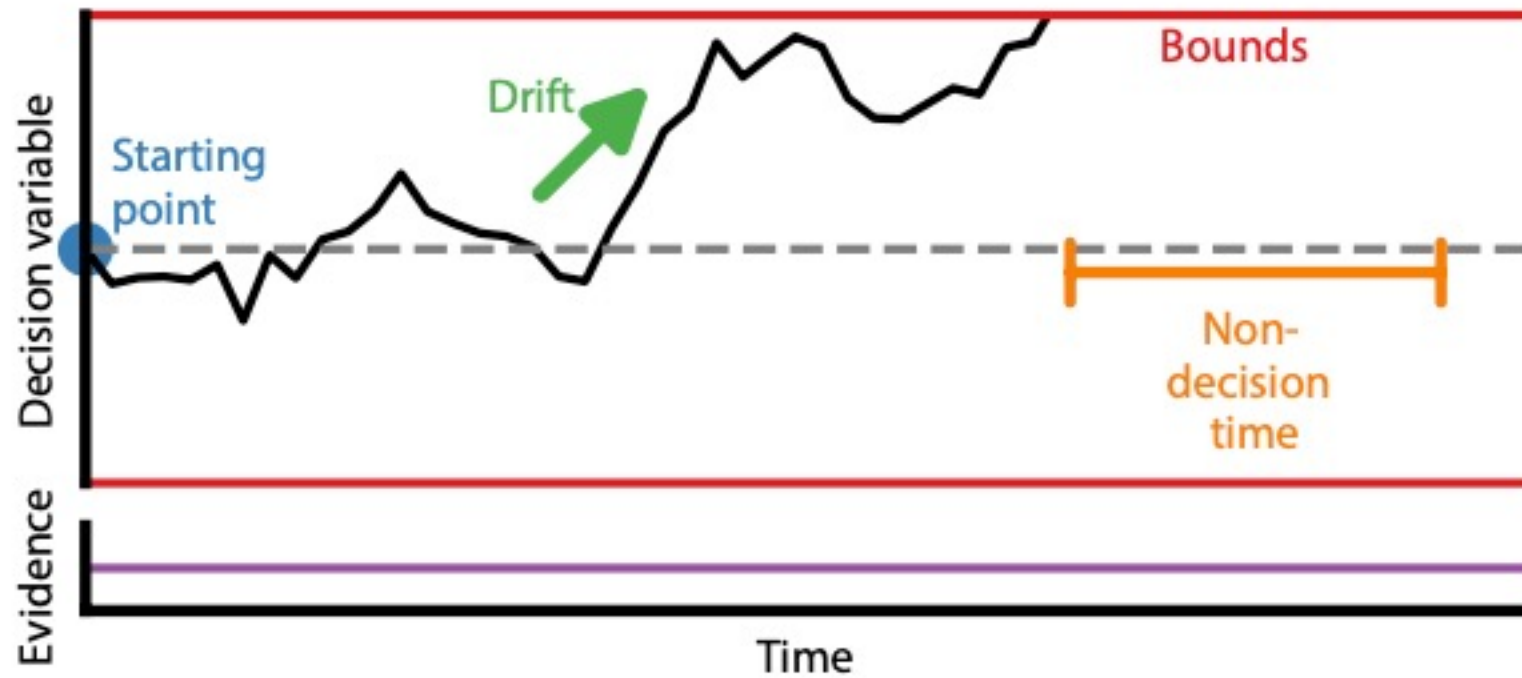
α – *boundary*

z – *startpoint*

T_{er} – *non.decision.time*

Simple DDM

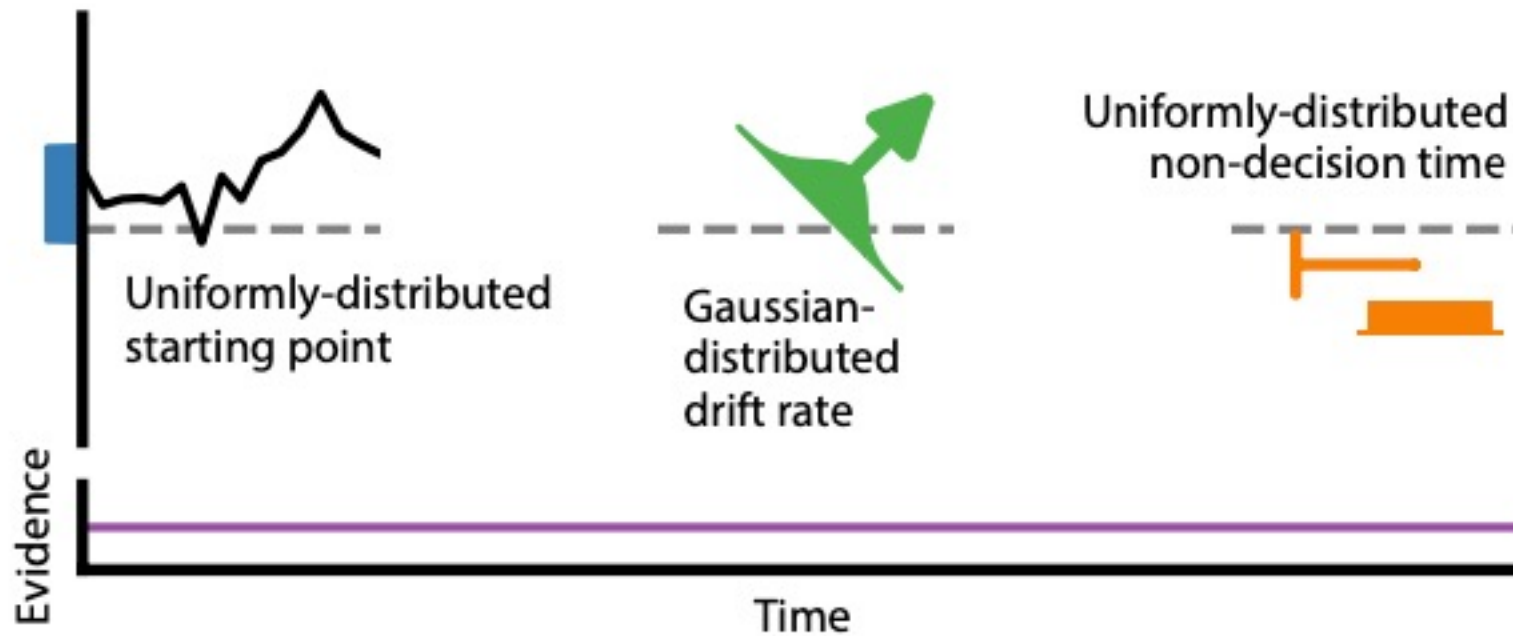
DDM



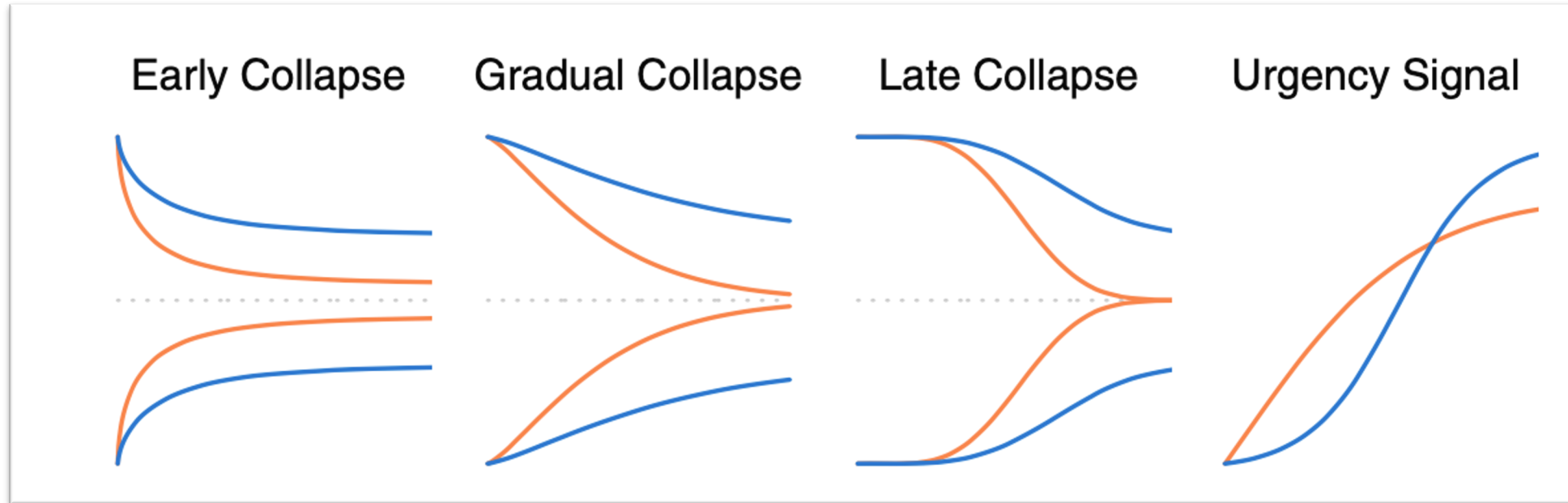
Diffusion decision model

Full DDM

Full DDM



Full DDM with time-varying boundaries



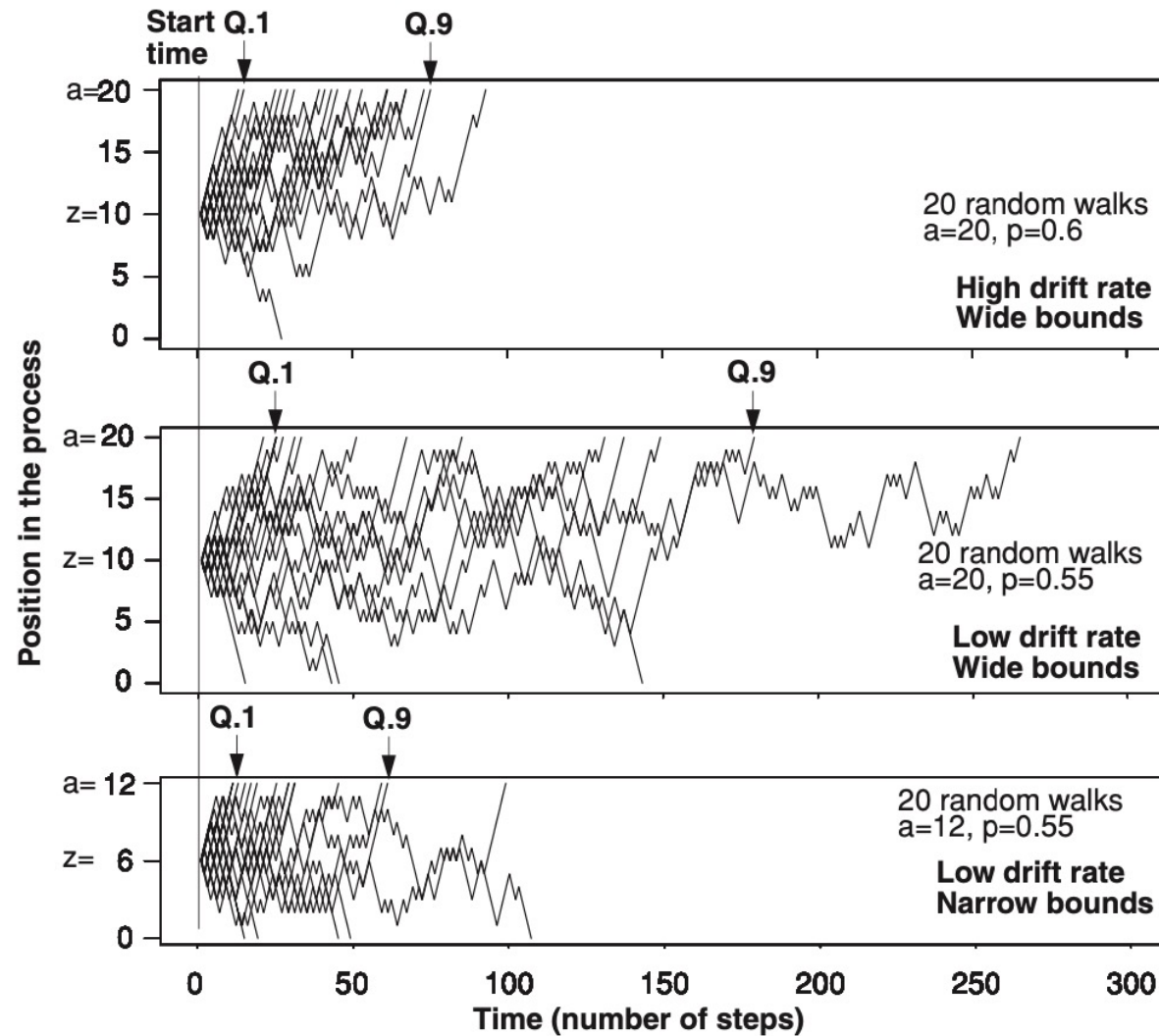
$$u(t) = a - \left(1 - \exp\left(-\left(\frac{t}{\lambda}\right)^k\right)\right)\left(\frac{1}{2}a - a'\right)$$

a' , asymptotic boundary setting

What can be explained by DDM?

Speed-accuracy trade-off

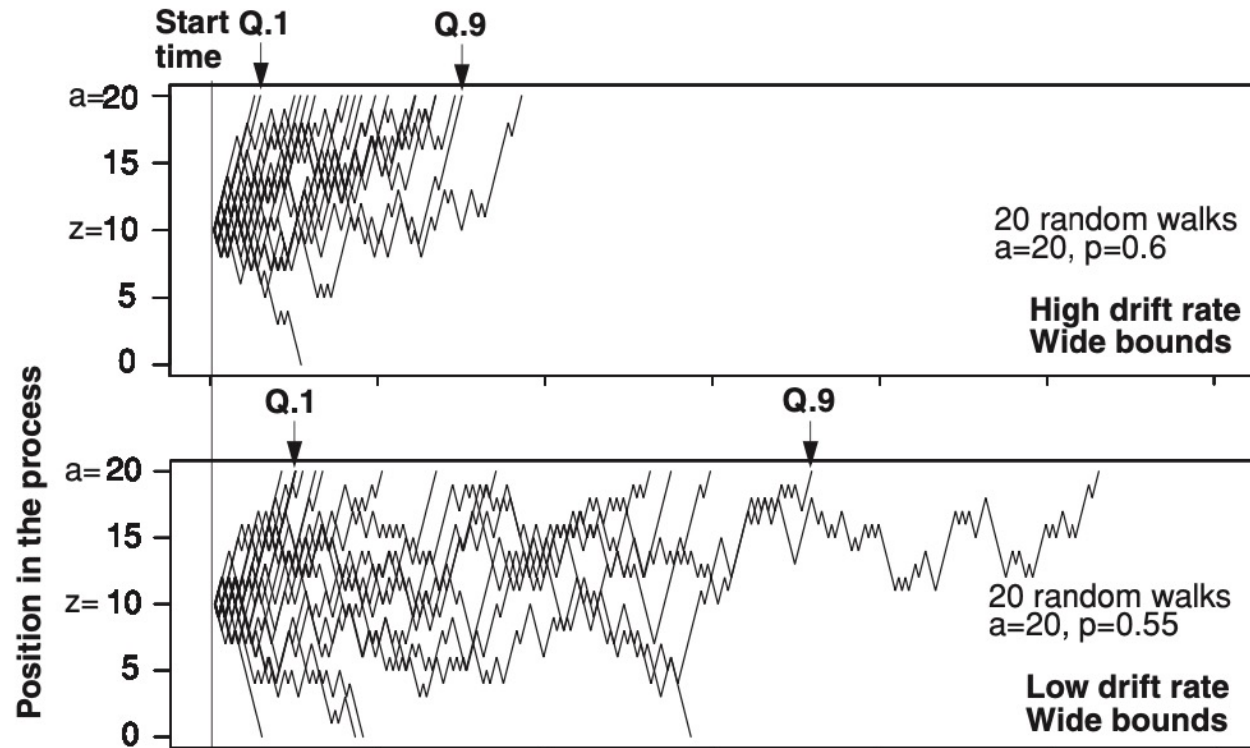
Increasing speed stress shortens mean RT but increases the proportion of errors



Q.1 and Q.9 refer to the .1 and .9 quantiles of the resulting sets of RTs.

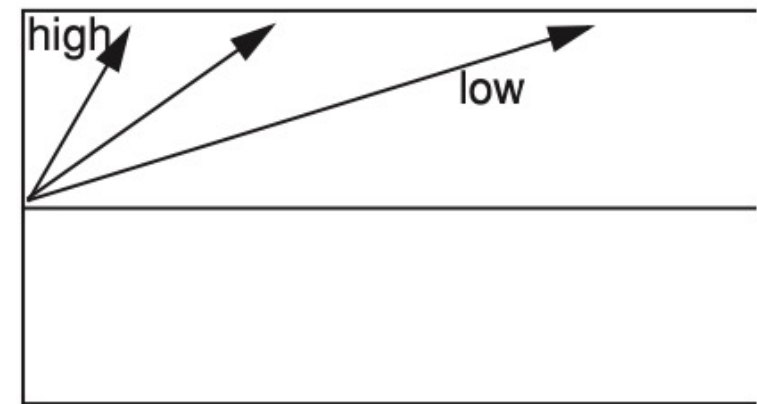
Ratcliff and McKoon, 2008, Neural computation

What can be explained by DDM?



Mean RT (easy) < Mean RT (hard)

Quality of evidence from the stimulus
Only **drift rate** varies



Ratcliff and McKoon, 2008, Neural computation

Q.1 and Q.9 refer to the .1 and .9 quantiles of the resulting sets of RTs.

Value-based DDM

Judgment and Decision Making, Vol. 5, No. 6, October 2010, pp. 437–449

The Drift Diffusion Model can account for the accuracy and reaction time of value-based choices under high and low time pressure

Milica Milosavljevic^{*1}, Jonathan Malmaud^{1,2*}, Alexander Huth^{1*},
Christof Koch^{1,2,3}, and Antonio Rangel^{1,4}

Value-based DDM



Value difference modulated drift rate

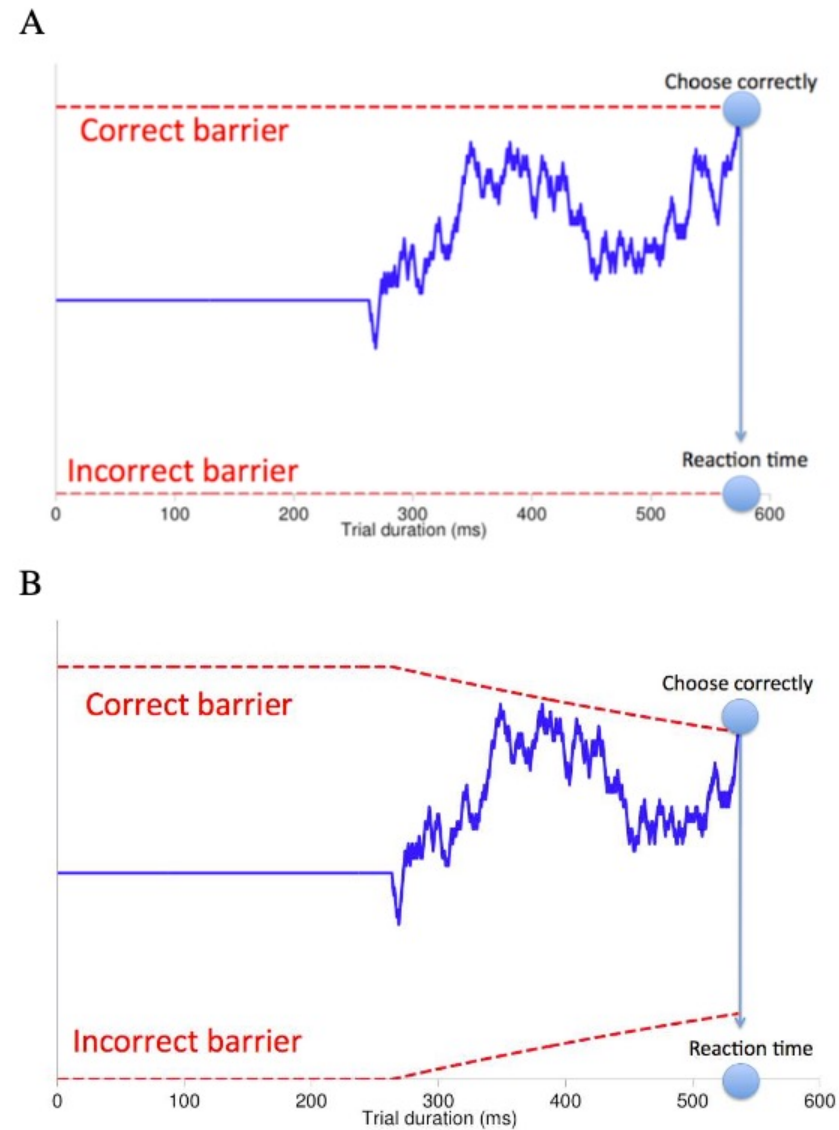
$$v = v_{mod} * (value_{hamburg} - value_{salad})$$

Value difference modulated boundary

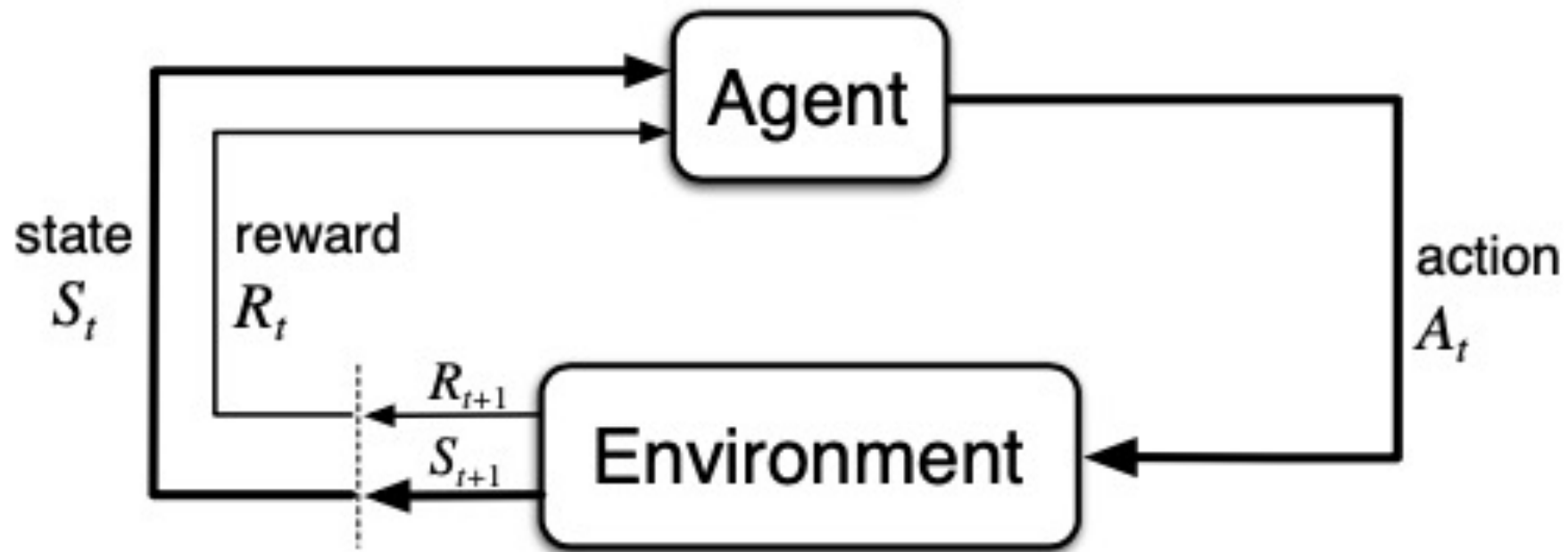
$$\alpha(t) = e^{-rt}$$

$$r \in \mathbb{R}^+$$

Value-based DDM



RL? RL is life! Markov Decision Process



General picture about reinforcement learning-MDP

RL? RL is life! Markov Decision Process

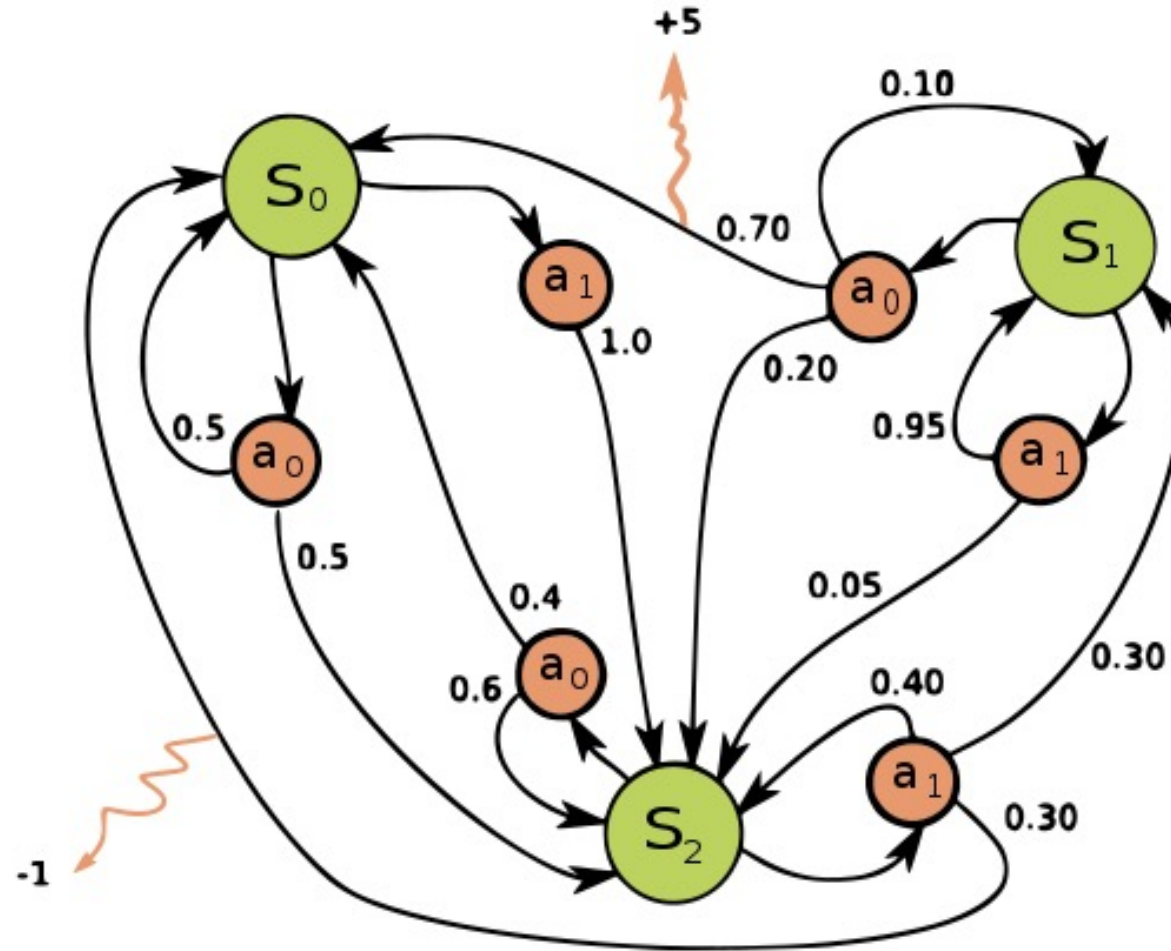


Figure from wikipedia

Rescorla-Wagner Theory (1972)

Simple "delta-rule" model

If stimulus present $S = 1$ ($S=0$ if not present)

$$w = w + \text{epsilon} * S * \text{delta}$$

$$\text{delta} = r - wS$$

r – reinforcer

S-CS

epsilon -learning rate

delta -Prediction error

Classical conditioning



CS: Conditioned Stimulus (S)

Pair stimulus (bell, light)

US: Unconditioned stimulus (reinforcement r)

...with significant event (food, shock)

CR: Conditioned response

Measure anticipatory behavior (salivation, freezing)

Now, You can answer! How animals learn from errors?

Now, You can answer! How human/animals learn from errors?

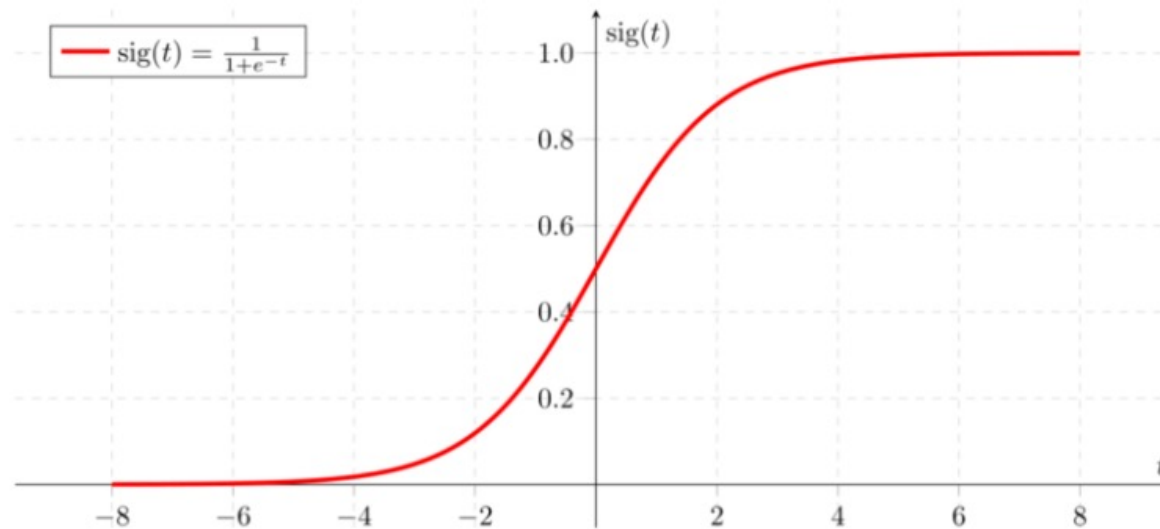
$$Q_t = Q_{t-1} + \alpha * PE$$

$$PE = f_t - Q_{t-1}$$

Since you've acquired the value of each potential choices
trial by trial

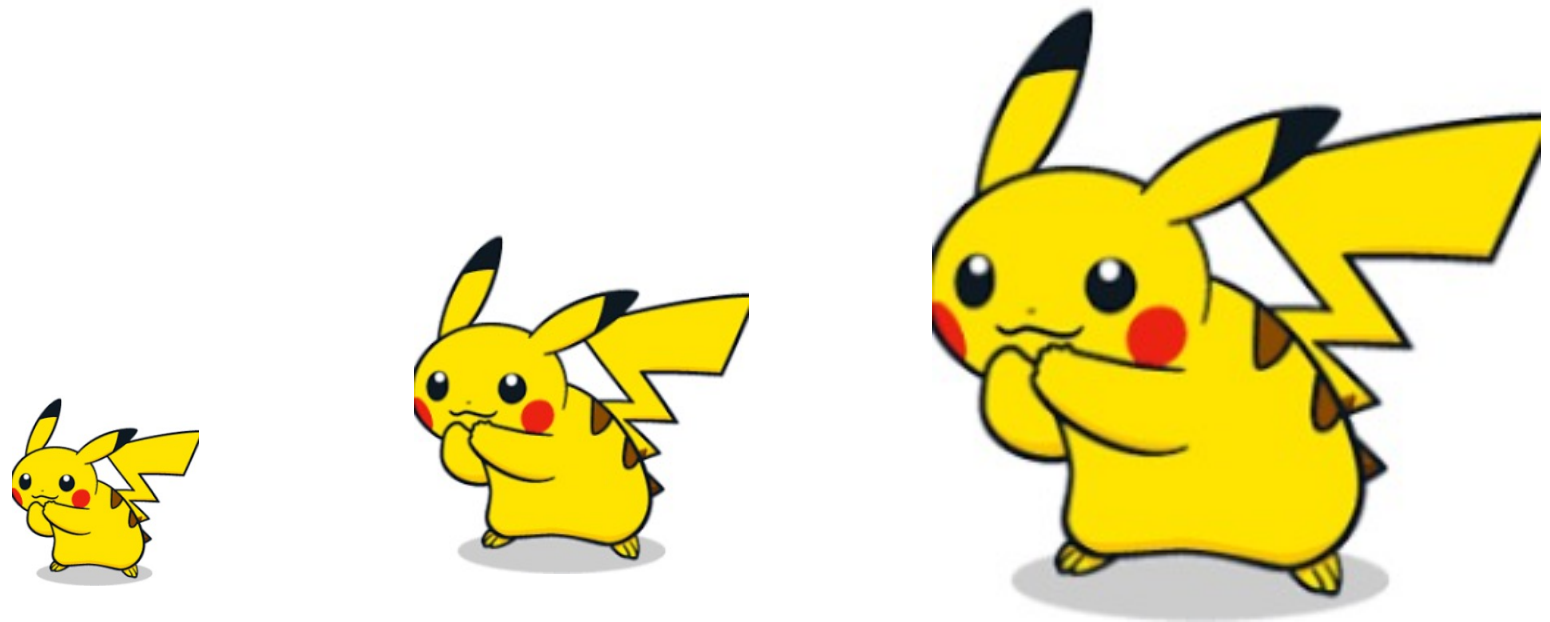
Then, how do you make your decisions?

softmax function, sigmoid function



$$P(t) = \frac{\exp^{Q_1\beta}}{\exp^{Q_1\beta} + \exp^{Q_2\beta}}$$

Please imagine their combination before my next slide!



$$v = v_{mod} * (value_{correct} - value_{incorrect})$$

$$alpha(t) = e^{-rt}$$

$$Wiener(RT|v, alpha, z, T_{er})$$

e.g.

Maximum Likelihood Estimation

Maximum Likelihood Estimation

Let's start from the most simple linear regression

$$Y_n = \theta * X_n + \epsilon_n$$

$$\min_{\theta} \frac{1}{N} \sum_{n=1}^N (y_n - \theta x_n)^2$$

Fitting the linear model by MSE (mean-squared error)

$$\frac{d}{d\theta} \frac{1}{N} \sum_{i=1}^N (y_i - \theta x_i)^2 = 0$$

$$\frac{1}{N} \sum_{i=1}^N -2x_i (y_i - \theta x_i) = 0$$

$$\hat{\theta} = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}$$


$$\hat{\theta} = \frac{\vec{x}^\top \vec{y}}{\vec{x}^\top \vec{x}}$$

Fitting the by MLE

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$Y_n = \theta * X_n + \epsilon_n$$

$$\epsilon \sim \mathcal{N}(0, 1).$$


$$p(y|x, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta x)^2}$$

$$\mathcal{L}(\theta|X, Y) = \prod_{i=1}^N \mathcal{L}(\theta|x_i, y_i)$$

$$\log \mathcal{L}(\theta|X, Y) = \sum_{i=1}^N \log \mathcal{L}(\theta|x_i, y_i)$$

Fitting the by MLE

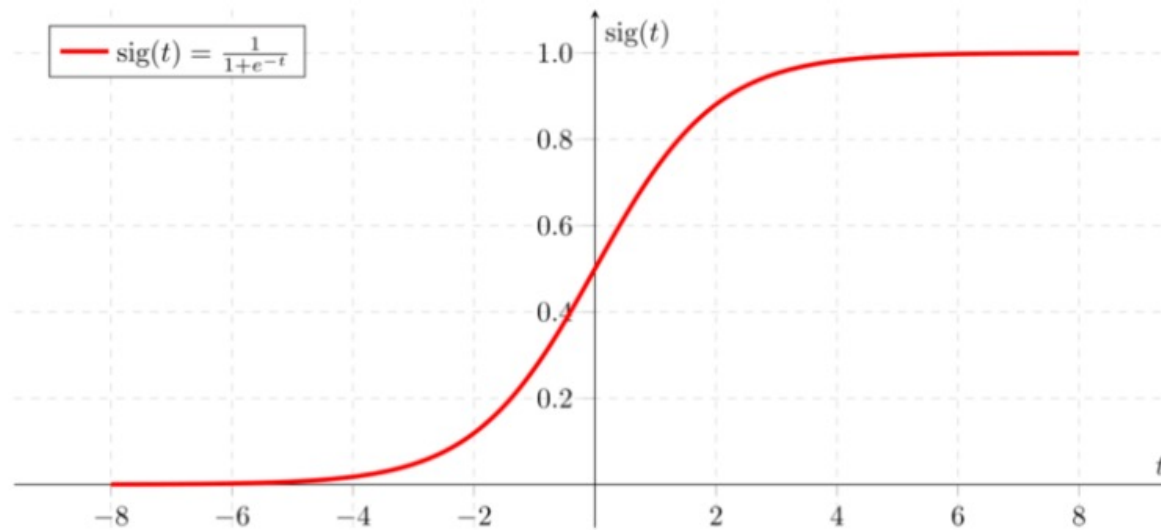
$$\hat{\theta}_{\text{MLE}} = \operatorname{argmax}_{\theta} \left[-\frac{N}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \theta x_i)^2 \right]$$

$$\text{MLE} === \text{MSE}$$

In this example, MLE has already included the cost function

Fitting optimal parameters in Rescorla-Wagner model by MLE

$$F(x) = P(X \leq x) = \frac{1}{1 + e^{-(x-\mu)/\gamma}}$$



<https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>

Figure

Fitting optimal parameters in Rescorla-Wagner model by MLE

$$P(choice = 1|value.difference) = F(value.difference)$$

$$P(choice = 0|value.difference) = 1 - F(value.difference)$$

Likelihood function

$$P(choice = choice_i|value.difference) = Bernoulli.function$$

$$= F(value.difference)^{choice_i} * [1 - F(value.difference)^{1-choice_i}]$$

Papers you may read

0_REFERENCE	12:45 PM	--	Folder
ddm_great_tool	9:44 AM	--	Folder
2013_HDDM.pdf	2021/4/21	2.5 MB	PDF
2020_elife_A flexible framework for simulating and fitting generalized drift-diffusion models.pdf	2021/4/17	987 KB	PDF
ddm_math	2021/5/19	--	Folder
2009_jmp_Fast and accurate calculations for first-passage times in Wiener diffusion models.pdf	2021/5/19	2.1 MB	PDF
2012_JMP_The_first-passage_time_distribution for the diffusion model with variable drift.pdf	2021/5/19	450 KB	PDF
ddm_must_read	9:45 AM	--	Folder
2008_neural_computation_The Diffusion Decision Model- Theory and Data for Two-Choice Decision Tasks.pdf	Yesterday	578 KB	PDF
2010_The Drift Diffusion Model can account for the accur...value-based choices under high and low time pressure.pdf	Yesterday	792 KB	PDF
2015_JN_Revisiting the Evidence for Collapsing Boundaries and Urgency Signals in Perceptual Decision-Making.pdf	Yesterday	1.9 MB	PDF
2016_annual_review_psy_Sequential Sampling Models in...euroscience- Advantages, Applications, and Extensions.pdf	2021/5/20	1.1 MB	PDF