UNIVERSITY of STIRLING

ACRG Applied Choice Research Group

Identifying consideration sets through search and an implementation using Shiny

Erlend Dancke Sandorf[a] Danny Campbell[a]

[a]Economics Division, Stirling Management School, University of Stirling

ENVECHO Verona 2019

The search for information - the search for alternatives

The role that information and search plays in choice has a long history in economics

- Search models – Simon (1955), Stiegler (1961), Weitzman (1979), Gabaix et al. (2006)
- Consideration set models – Richardson (1982), Roberts and Lattin (1991)
- Information acquisition models – Hausmann and Lage (2008), Chorus et al. (2013)

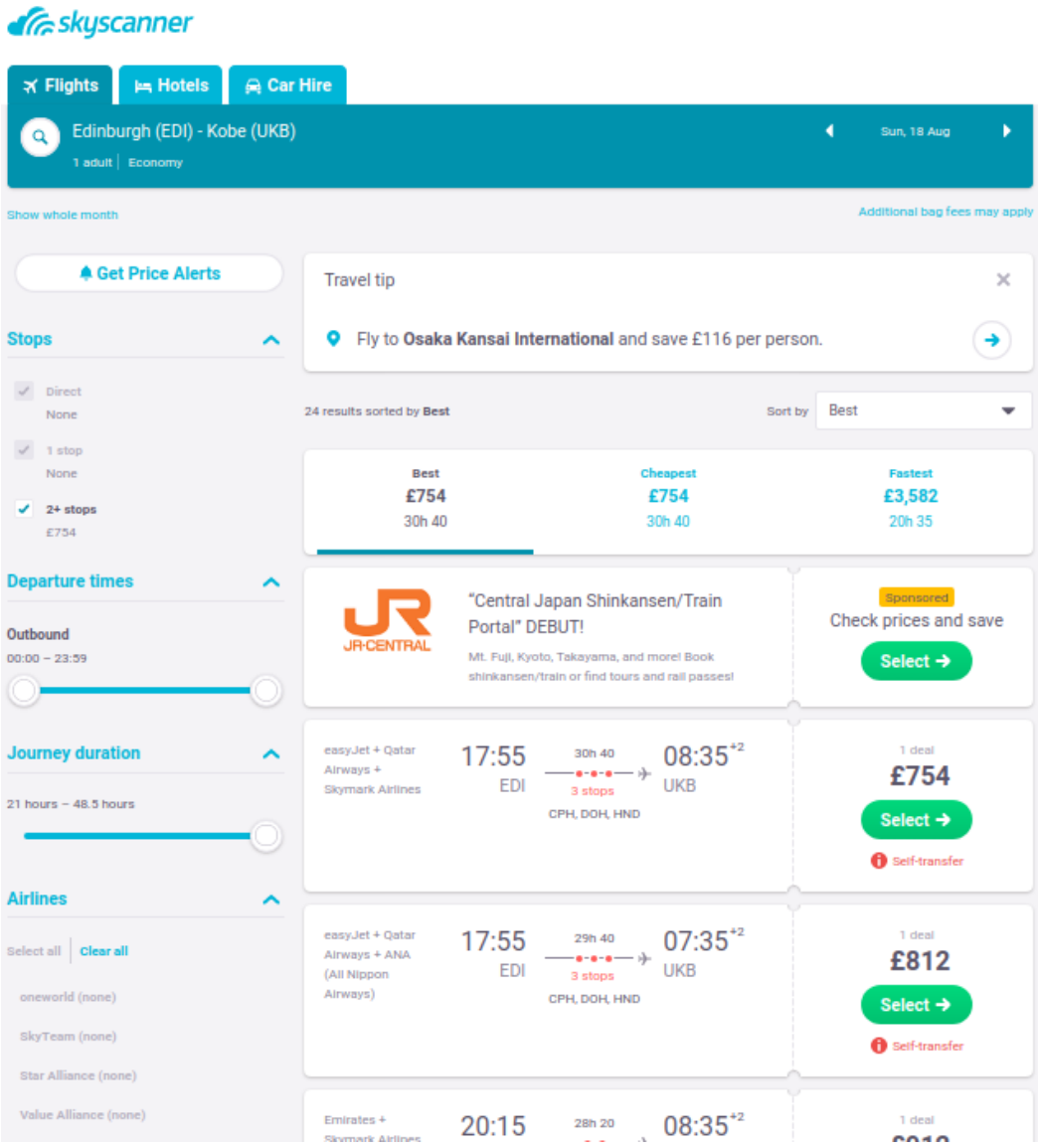
The role that information and search plays in choice has a long history in economics

- Search models – Simon (1955), Stiegler (1961), Weitzman (1979), Gabaix et al. (2006)
- Consideration set models – Richardson (1982), Roberts and Lattin (1991)
- Information acquisition models – Hausmann and Lage (2008), Chorus et al. (2013)

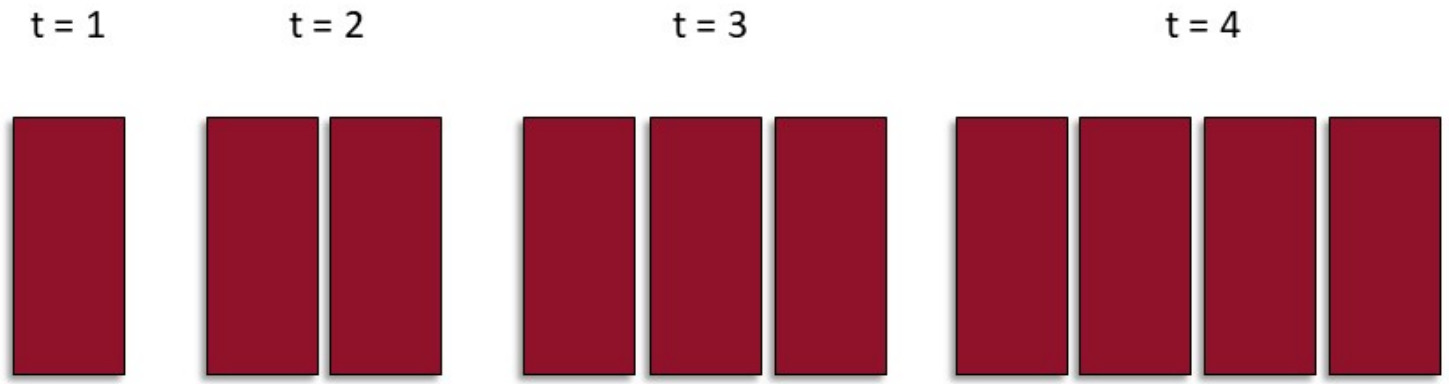
People search as long as the expected gain from search exceeds the marginal cost

In many (if not most) choice situations, options are evaluated sequentially

In many (if not most) choice situations, options are evaluated sequentially



This means that consideration sets grow sequentially with each period of search

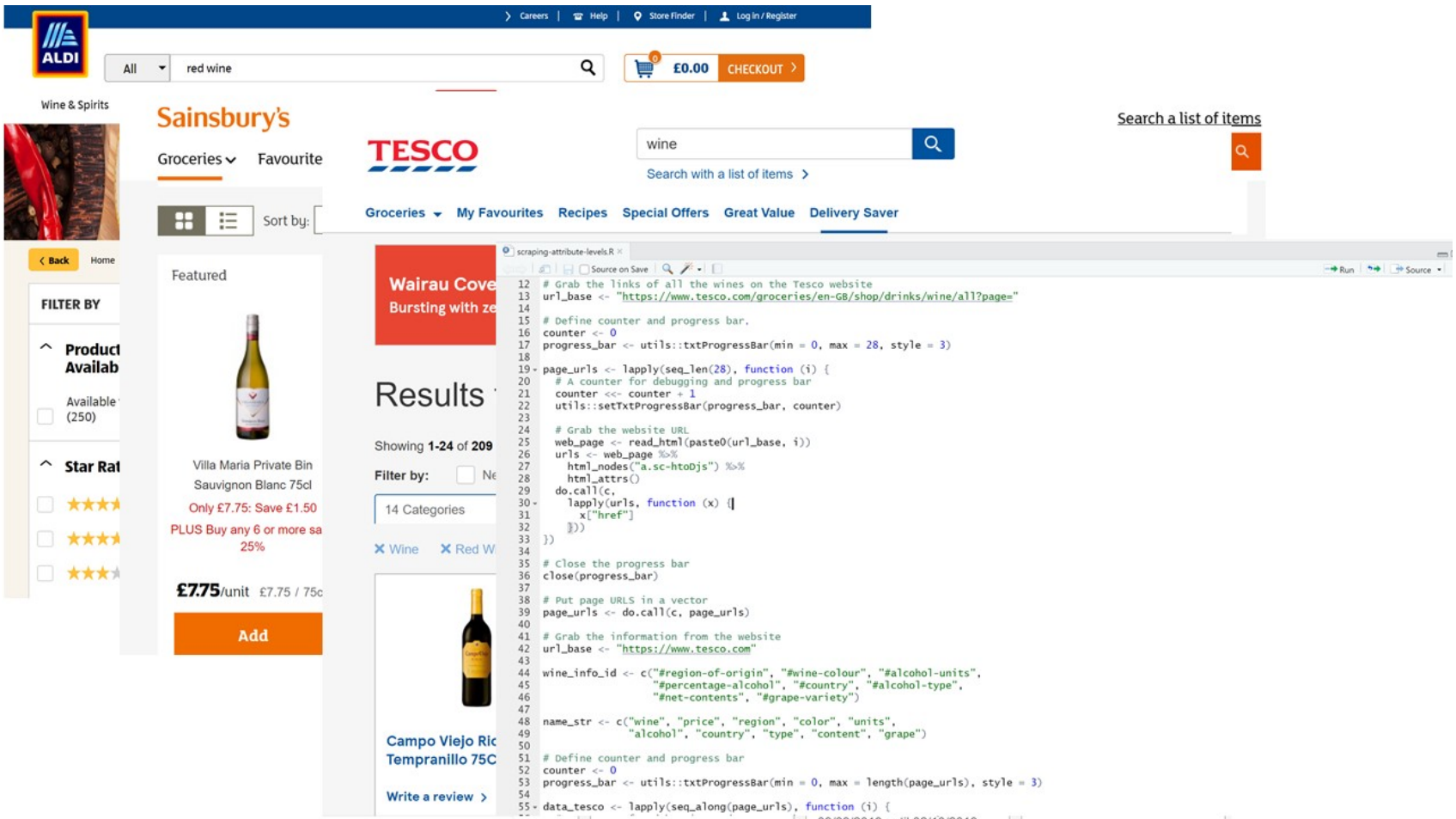


Gathering data to test hypotheses with respect to search and satisficing

A good that is familiar, reasonably priced and induces people to search

Selected attributes: Country of Origin, Color, Alcohol Content, Grape Variety, Characteristic, Organic and Price

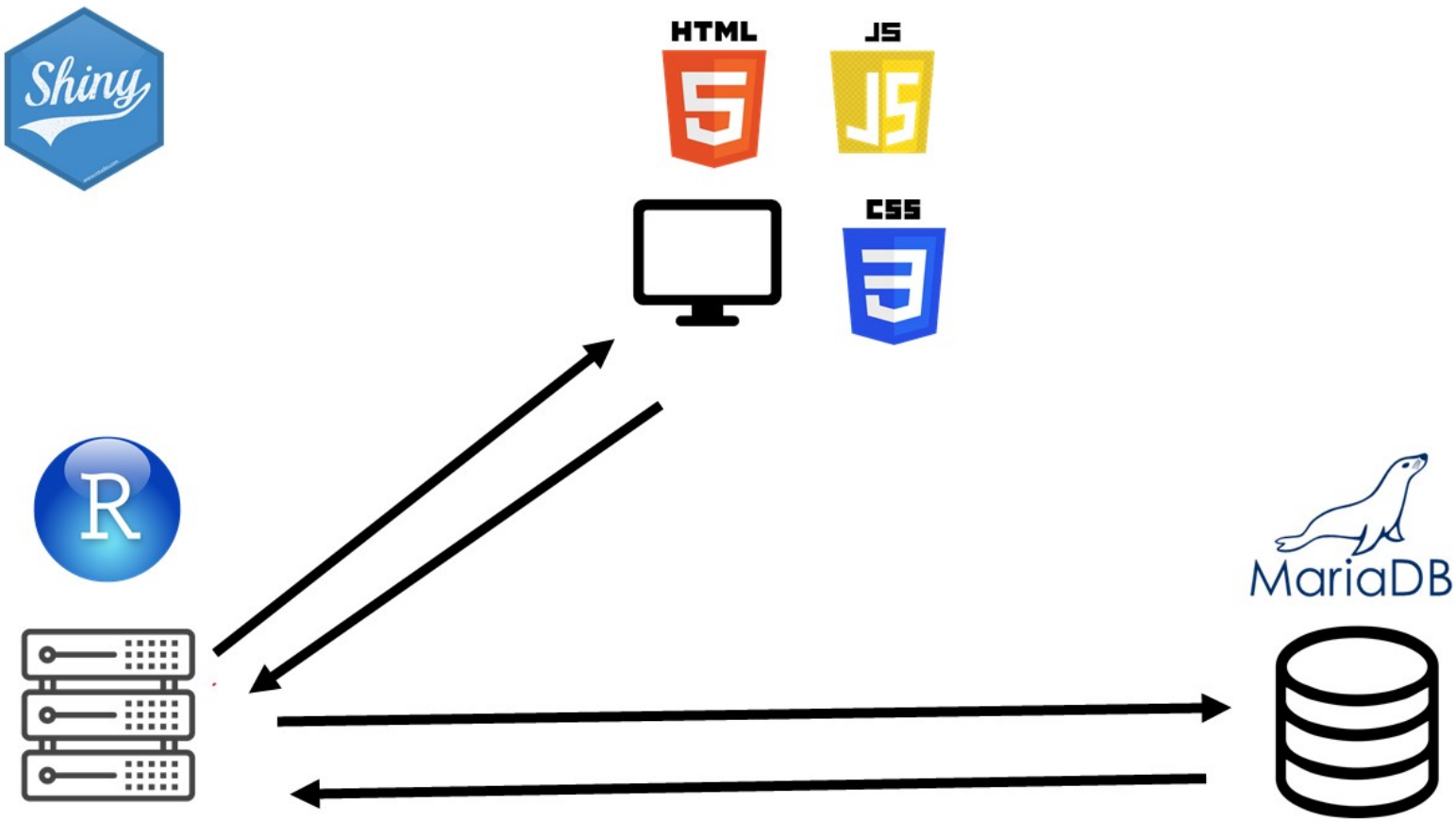
We used a web scraping algorithm to gather information on our attributes



We sample alternatives according to how common they are in the store

	A	B	C	D	E	F	G	H
1	country	color	alcohol	grape	characteristic	organic	price	weight
2	New Zealand	White	11.50%	Pinot Grigio	2 - Semi dry	No	8	3.02E-16
3	South Africa	Red	13%	Tempranillo	5 - Full bodied	Yes	4.5	2.39E-15
4	Chile	Red	14%	Malbec	5 - Full bodied	Yes	5.5	5.26E-16
5	New Zealand	RosÃ©	10.50%	Merlot	1 - Dry	Yes	5.5	2.62E-17
6	Italy	White	14.50%	Blend	1 - Dry	Yes	15	1.63E-16
7	South Africa	Red	11%	Malbec	5 - Full bodied	Yes	5.5	3.56E-16
8	USA	RosÃ©	12%	Pinot Noir	3 - Medium dry	No	5.75	2.83E-16
9	Italy	RosÃ©	14%	Tempranillo	4 - Semi sweet	Yes	9.5	7.08E-17
10	Australia	Red	11.50%	Blend	1 - Light bodied	Yes	7.25	1.12E-15
11	New Zealand	RosÃ©	13%	Cabernet Sauvignon	3 - Medium dry	No	16.5	6.16E-17
12	Italy	Red	12%	Blend	1 - Light bodied	Yes	8.5	6.13E-16
13	France	White	10%	Sauvignon Blanc	3 - Medium dry	No	5	4.96E-16
14	USA	White	12%	Pinot Grigio	2 - Semi dry	Yes	8.75	4.44E-16
15	New Zealand	RosÃ©	12%	Pinot Noir	4 - Semi sweet	Yes	5	3.84E-16
16	New Zealand	White	13%	Pinot Grigio	1 - Dry	Yes	17.5	2.89E-16
17	South Africa	Red	12.50%	Merlot	4 - Full bodied	No	6.5	1.07E-15
18	Italy	RosÃ©	14%	Merlot	3 - Medium dry	Yes	6.25	7.08E-17
19	New Zealand	Red	15%	Tempranillo	4 - Full bodied	Yes	11	2.19E-17
20	Italy	RosÃ©	12%	Pinot Noir	4 - Semi sweet	No	5.5	3.67E-16

We implement the survey using Shiny and host it on shinyapps.io



An example using Shiny

UNIVERSITY of STIRLING

Question 1: Please choose your most preferred wine

	None of these wines	Wine 1	Wine 2	Wine 3
Country		Australia	Chile	Australia
Color		White	White	Red
Alcohol		12%	13.5%	12.5%
Grape		Chenin Blanc	Blend	Tempranillo
Characteristic		1 - Dry	3 - Medium dry	5 - Full bodied
Organic		No	Yes	Yes
Price		5	9	13
I choose	○	○	○	○

www.acrg.site

Economics Division

Stirling Management School

University of Stirling

Scotland

Econometric model

Utility can be described by a separable and additive utility function

$$u_{nis} = \beta x_{nis} + \varepsilon_{nis}$$

U_{nis} Utility

β Vector of parameters to be estimated

X_{nis} Levels of the attributes

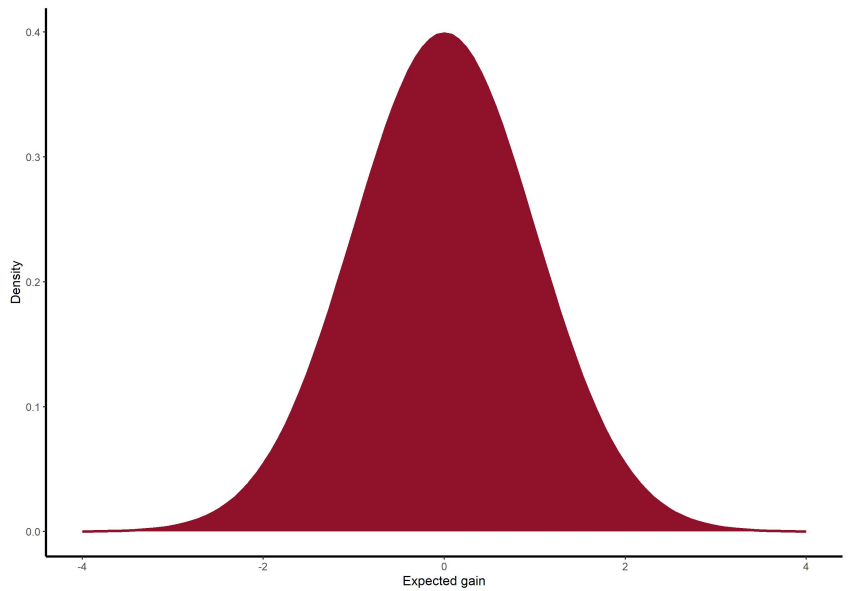
ε Type I Extreme value distributed error term with variance $\pi^2/6$

The possible gain from search is the difference between any alternative and the current best

$$g = u - u_{\max}$$

The value of all possible gains is the area under the "gain" curve

$$\bar{g} = \int_{-\infty}^{+\infty} gP(g)\mathrm{d}g,$$

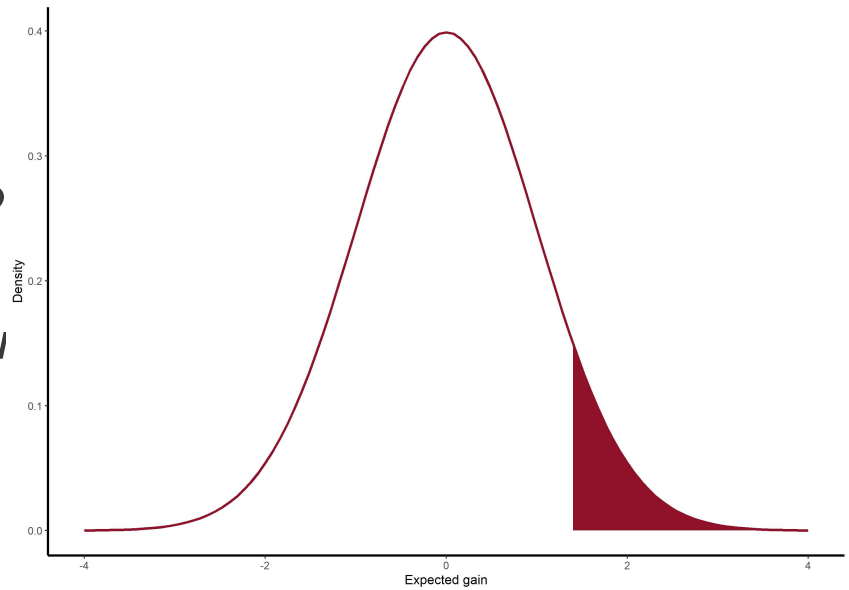


With recall you cannot lose utility by searching for another alternative

$$g = \begin{cases} u - u_{\max} & \text{if } u \geq u_{\max} \\ 0 & \text{if } u \leq u_{\max} \end{cases}$$

The gain from searching is the area under the "gain" curve above the current best

$$
\begin{aligned}
\bar{g} &= \int_{u_{\max}}^{+\infty} (u - u_{\max})\, \phi(u) \mathrm{d}u \\
&= \int_{u_{\max}}^{+\infty} u \phi(u) \mathrm{d}u - \int_{u_{\max}}^{+\infty} u_{\max} \phi \\
&= \phi(u_{\max}) - u_{\max} \int_{u_{\max}}^{+\infty} \phi(u) \mathrm{d}u
\end{aligned}
$$



where

$$u_{\max} = (U_{\max} - \mu_t) / \sigma_t$$

An individual will search as long as expected gains are higher than the marginal cost of searching

$$\bar{G} - \bar{c} > 0$$

where

$$\bar{G} = \bar{g}\sigma$$

i.e. the non-standardized gain to be compared with the marginal cost of search $\bar{c}$, e.g. time, money, cognitive cost of maintaining a consideration set

The log likelihood function

$$P(i_s|C_{ns}) = \prod_{t=1}^{T=J} \left[\frac{\exp(\beta x_{nis})}{\sum_{j \in C_{ns}^t} \exp(\beta x_{njs})} \right]^{I_t}$$

where

$$I_t = \begin{cases} 1 & \text{if } \bar{G}_t - \bar{c}_t < 0 \quad t = t* \\ 0 & \text{if } \bar{G}_t - \bar{c}_t \geq 0 \quad \forall \quad t \neq t^* \end{cases}$$

and t^* is the first time the condition is TRUE.