Stanford

ICME
INSTITUTE for COMPUTATIONAL & MATHEMATICAL ENGINEERING

Understanding Regression-adjusted Control Variate

Sobolev embedding and rare event

Joint work with Jose Blanchet, Haoxuan Chen, Lexing Ying

Yiping Lu yplu@stanford.edu
https://2prime.github.io/

ML Nowadays

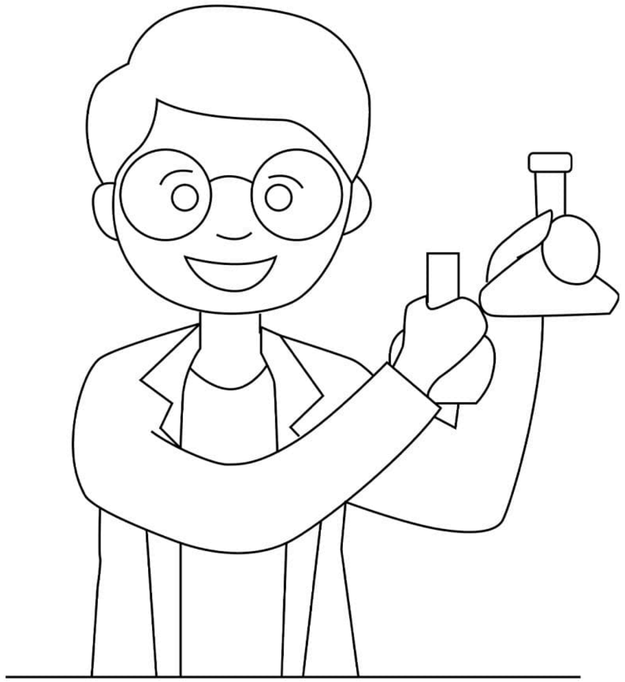
"entertainment, advising"

Midjourney

OpenAI ChatGPT4.0

《science》

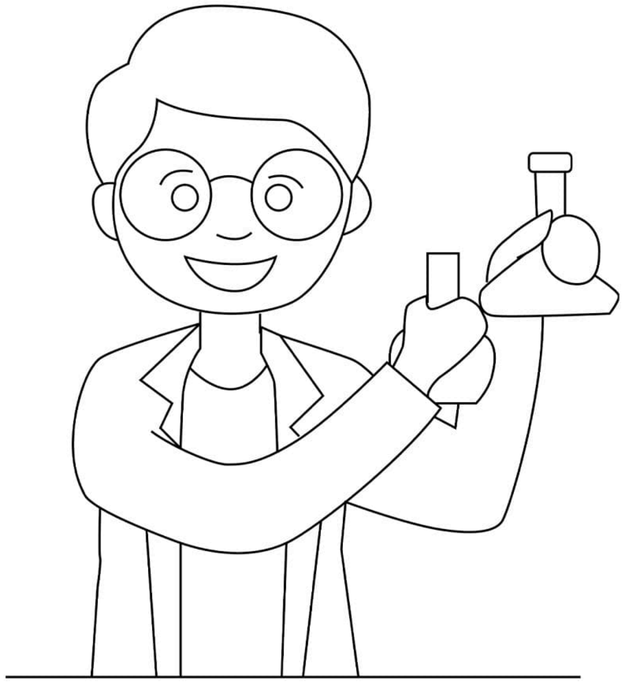
We want Guarantee!



Theorem If you randomly collect
( ) data, then you can achieve
( ) accuracy with your AI!

We want Guarantee!

Big constant



Theorem If you randomly collect
( ) data, then you can achieve
( ) accuracy with your AI!

Assumption 1. Xxx
Assumption 2. Xxx
Assumption 3. Xxx
Assumption 4. Xxx

ML for Science nowadays

Debaising ML for Science

You can prove theorem, but I still don't trust you!

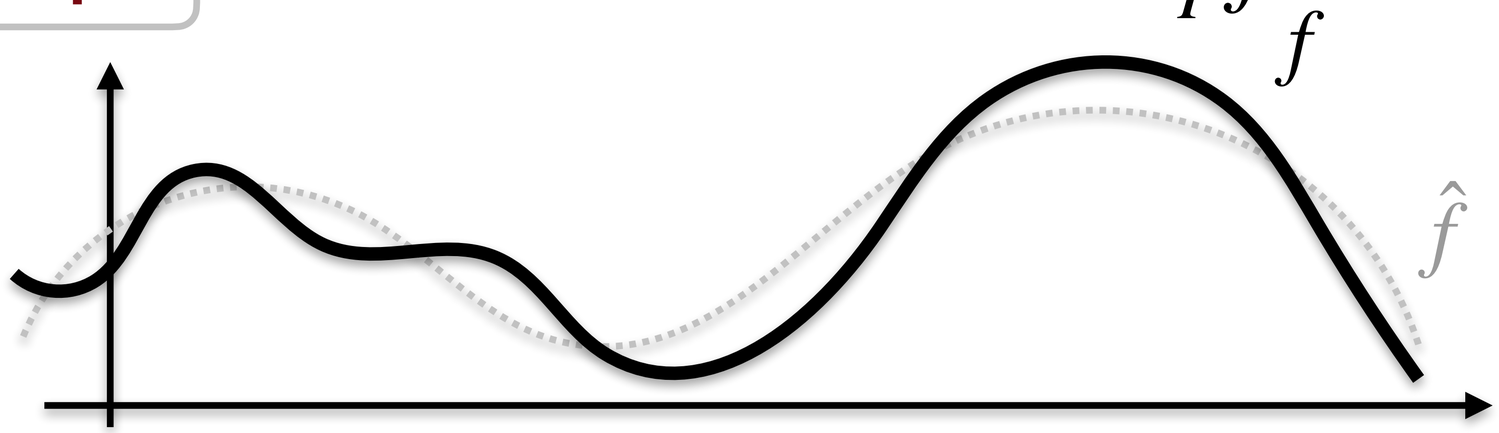
Can we debias ML estimator or use it in an unbiased way?

Debaising ML for Science

You can prove theorem, but I still don't trust you!

Can we debias ML estimator or use it in an unbiased way?

Example Monte Carlo Estimate $\mathbb{E}_P f$

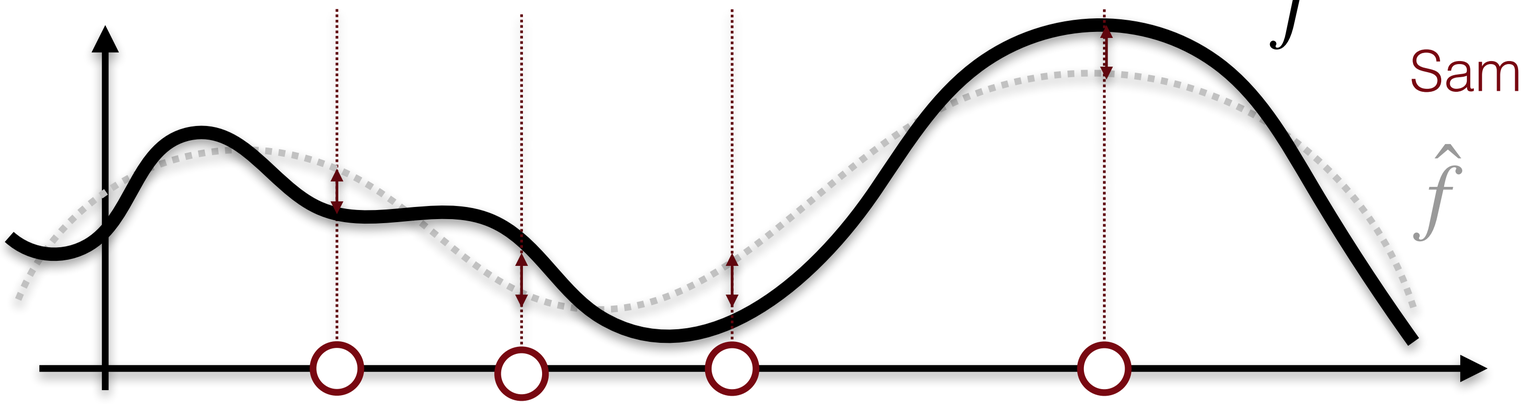


Debaising ML for Science

You can prove theorem, but I still don't trust you!

Can we debias ML estimator or use it in an unbiased way?

Example Monte Carlo Estimate $\mathbb{E}_P f$



Sample extra data to know $f - \hat{f}$

Debaising ML for Science

You can prove theorem, but I still don't trust you!

"Regression-adjusted control variate"

Can we debias ML estimator or use it in an unbiased way?

Example Monte Carlo Estimate $\mathbb{E}_P f$

Step 1 Using half of the data to estimate $\hat{f}$

Step 2 $\mathbb{E}_P f = \mathbb{E}_P(\hat{f}) + \mathbb{E}_P(f - \hat{f})$

Low order term

“Modern” regression-adjusted cv

Trace estimation:

Hutch++ Lin 17 Numerische Mathematik Mewyer-Musco-Musco-Woodruff 20

Dimension Reduction:

Sobczyk and Luisier Neuips 22

Conformal Prediction:

Conformalized quantile regression Romano-Patterson-Candes Neurips 19

Gradient Estimation

Shi-Zhou-Hwang-Tisias-Mackey Neurips 22 outstanding paper

Causal Inference:

Double Robust estimation ….

Debaising ML for Science

Is this algorithm statistical optimal?

When this improves MC estimator?

Can we debias ML estimator or use it in an unbiased way?

Example Monte Carlo Estimate $\mathbb{E}_P f$

Step 1 Using half of the data to estimate $\hat{f}$

Step 2 $\mathbb{E}_P f = \mathbb{E}_P(\hat{f}) + \mathbb{E}_P(f - \hat{f})$

Low order term

Understanding this statistically…

Is this algorithm statistical optimal?

Why consider $q-$th moment?

When this improves MC estimator?

Can we debias ML estimator or use it in an unbiased way?

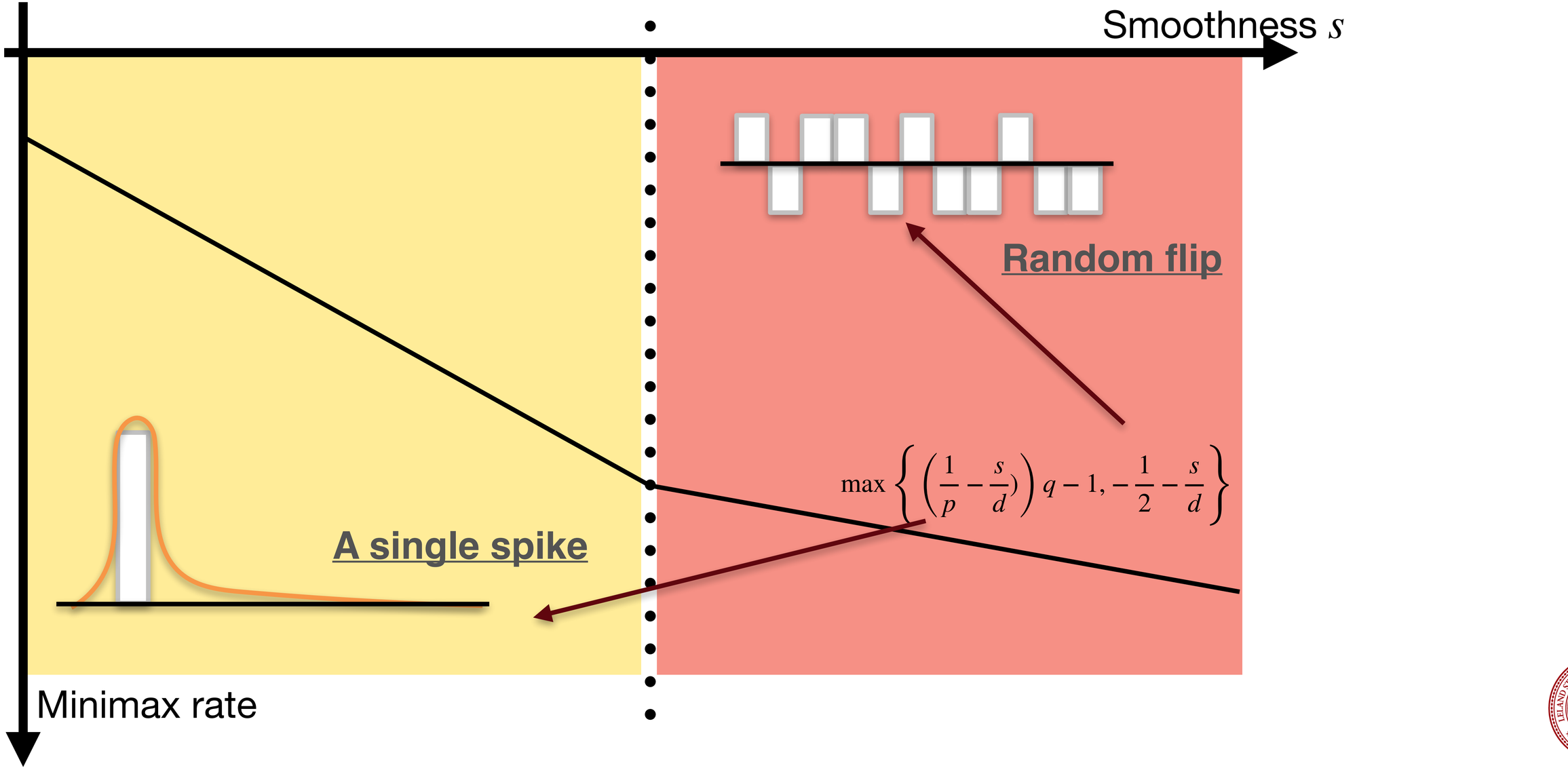
Example Monte Carlo Estimate $\mathbb{E}_P f$ $\mathbb{E}_P f^q, f \in W^{s,p}$

Step 1 Using half of the data to estimate $\hat{f}$

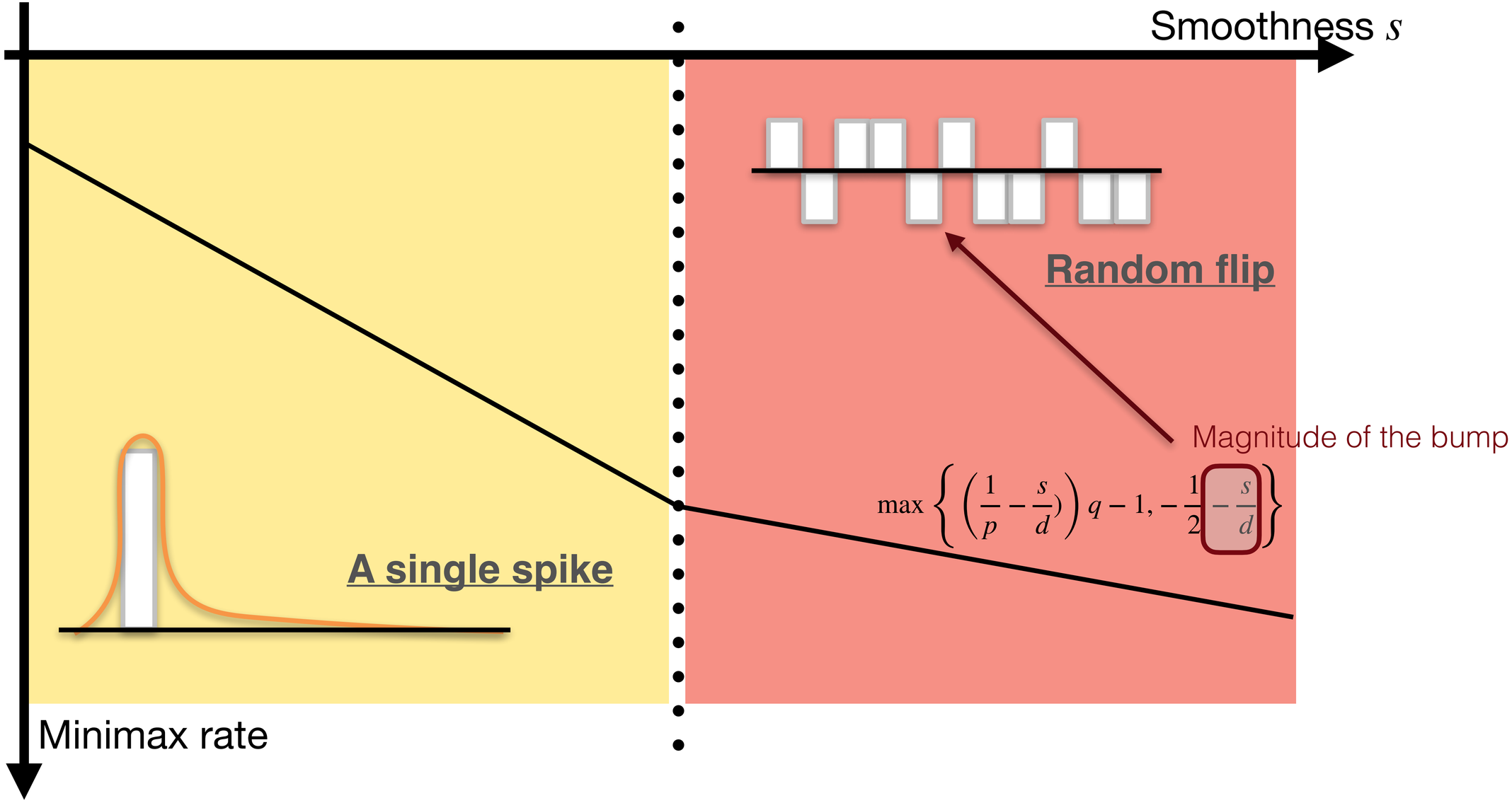
Step 2 $\mathbb{E}_P f^q = \mathbb{E}_P(\hat{f})^q + \mathbb{E}_P(f - \hat{f})^q$

Low order term

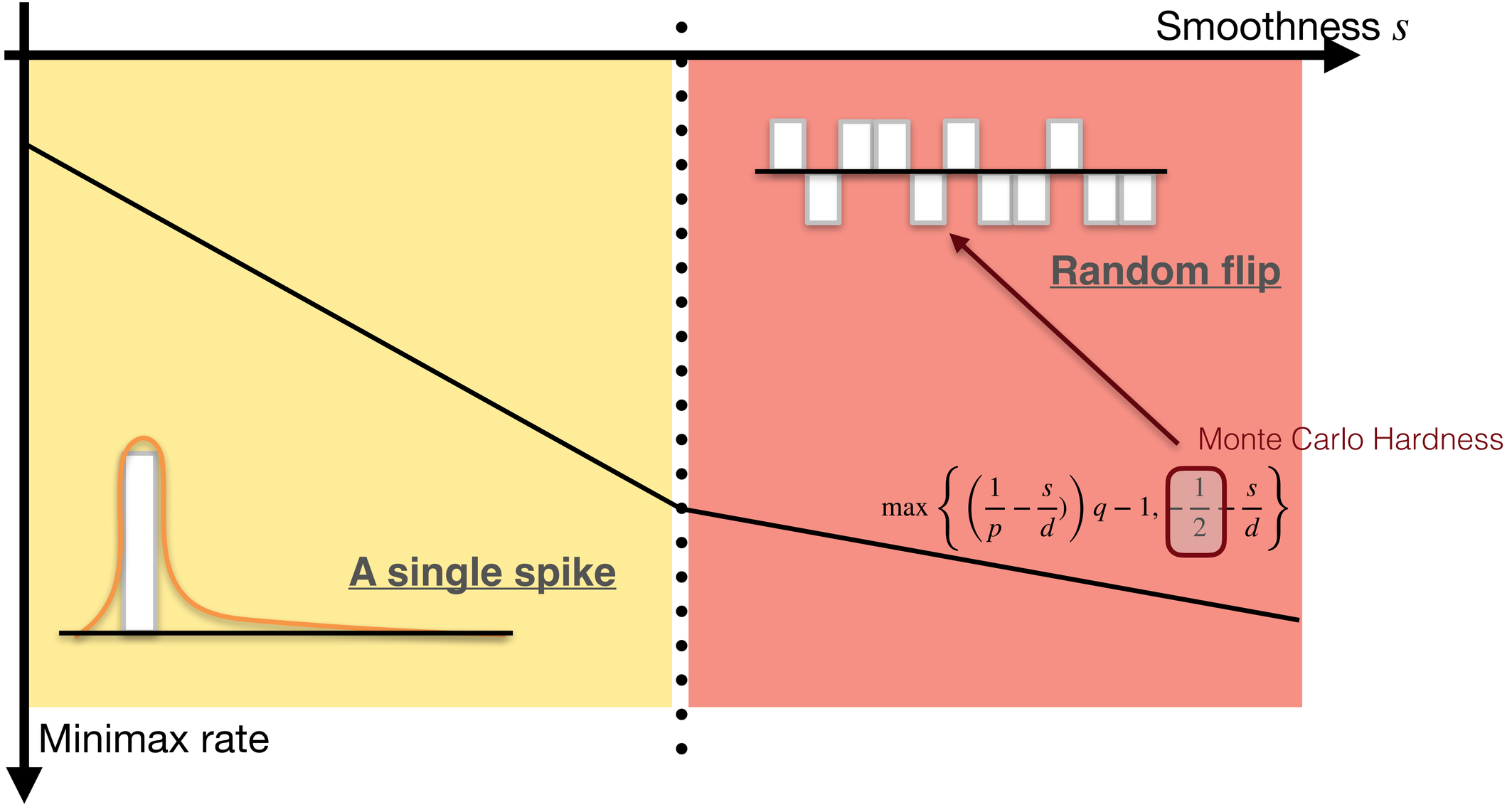
Setting the information theoretical limit



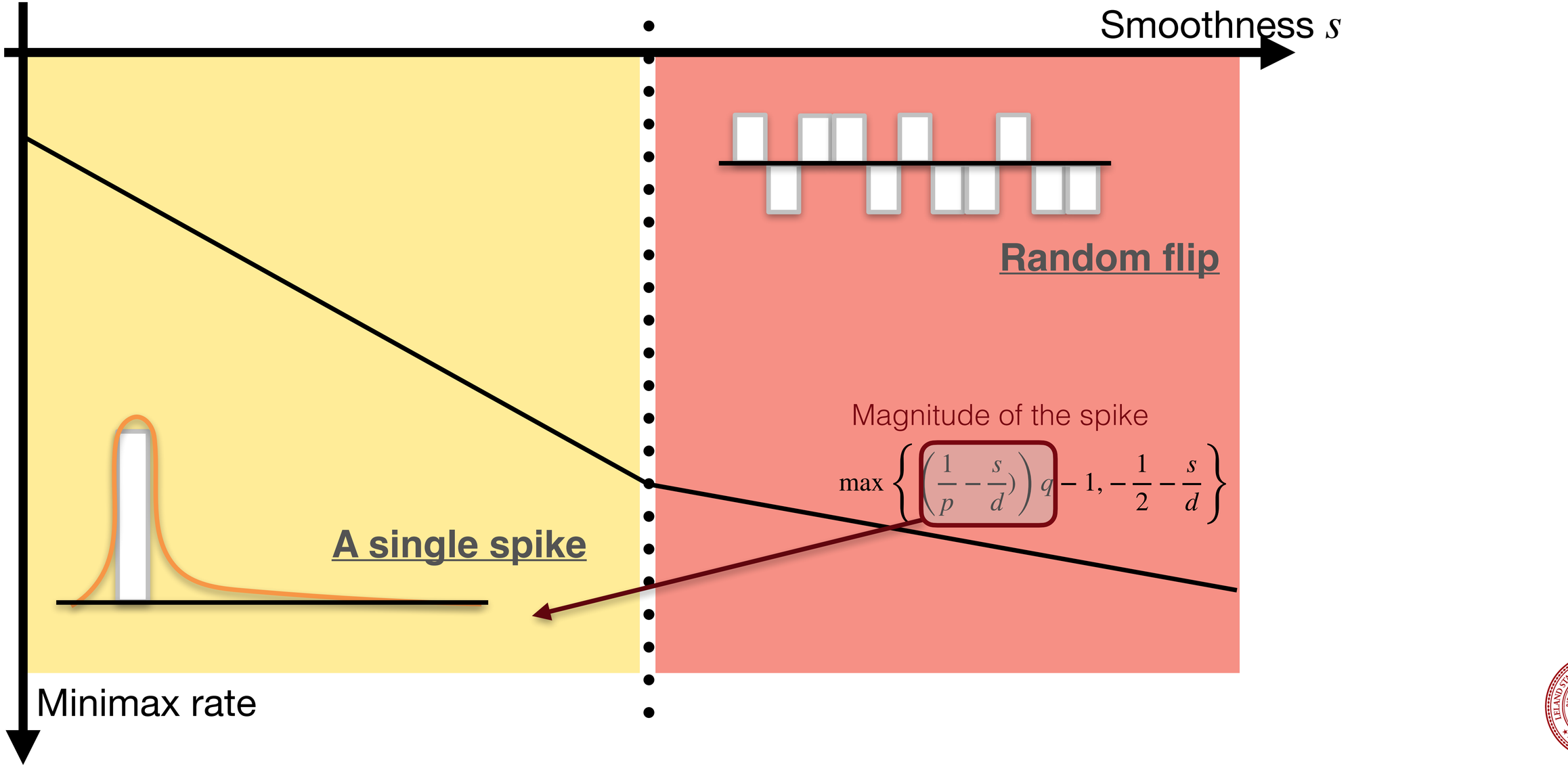
Setting the information theoretical limit



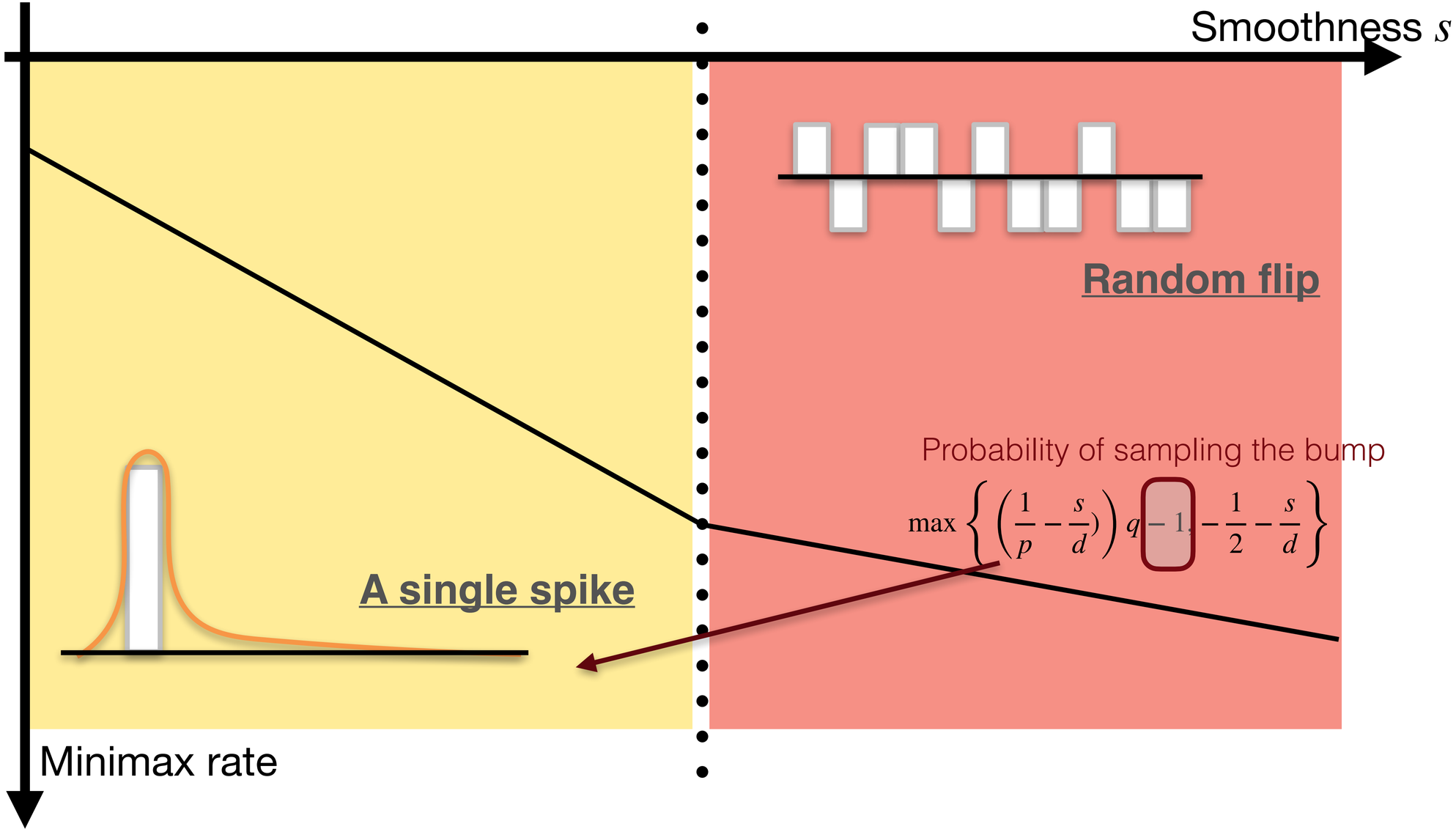
Setting the information theoretical limit



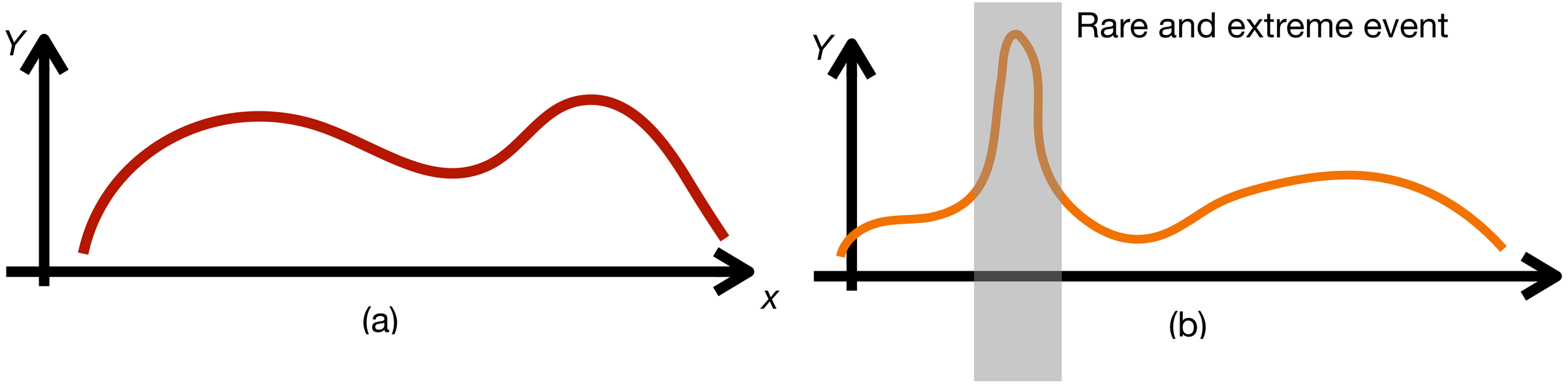
Setting the information theoretical limit



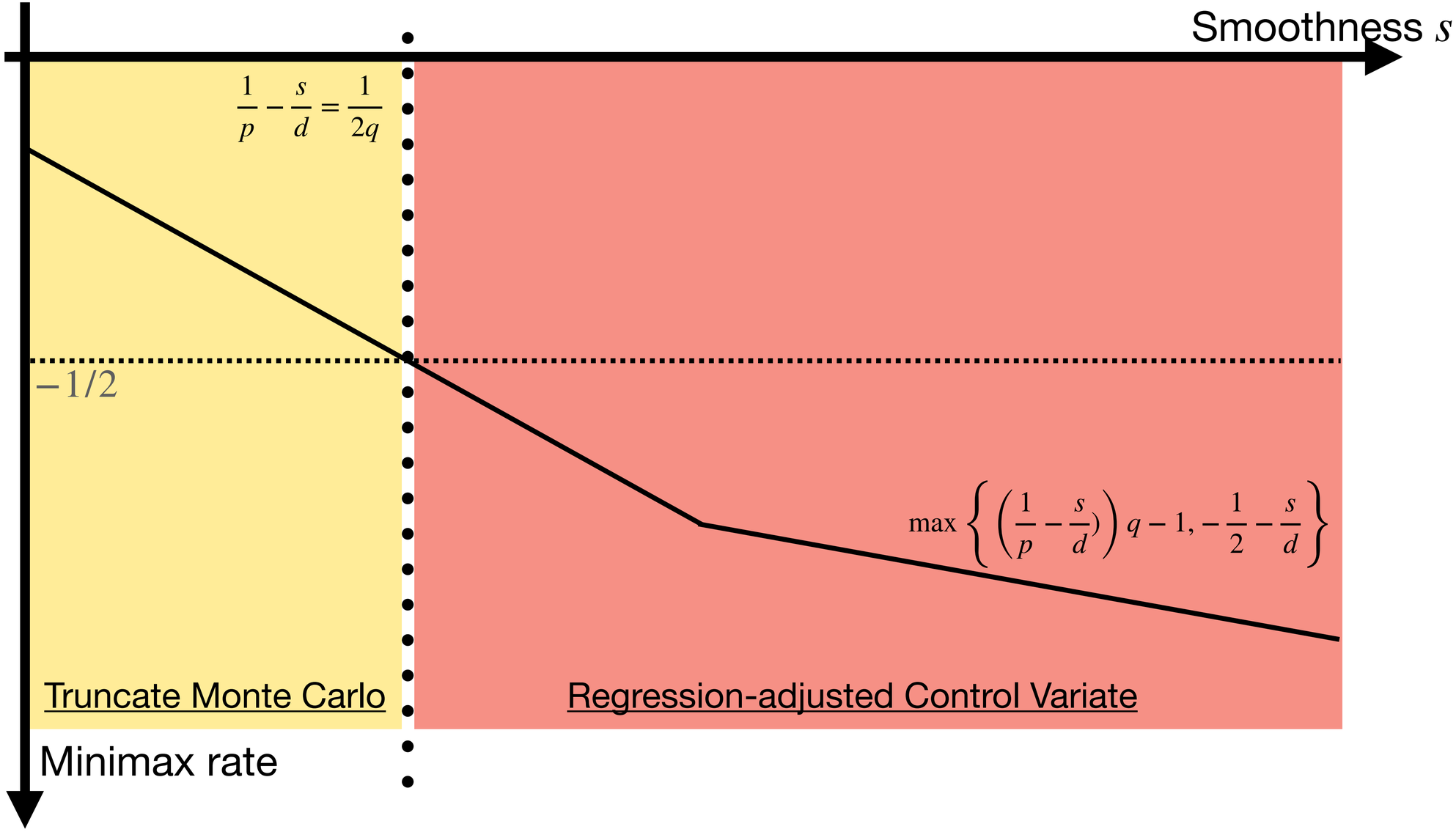
Setting the information theoretical limit



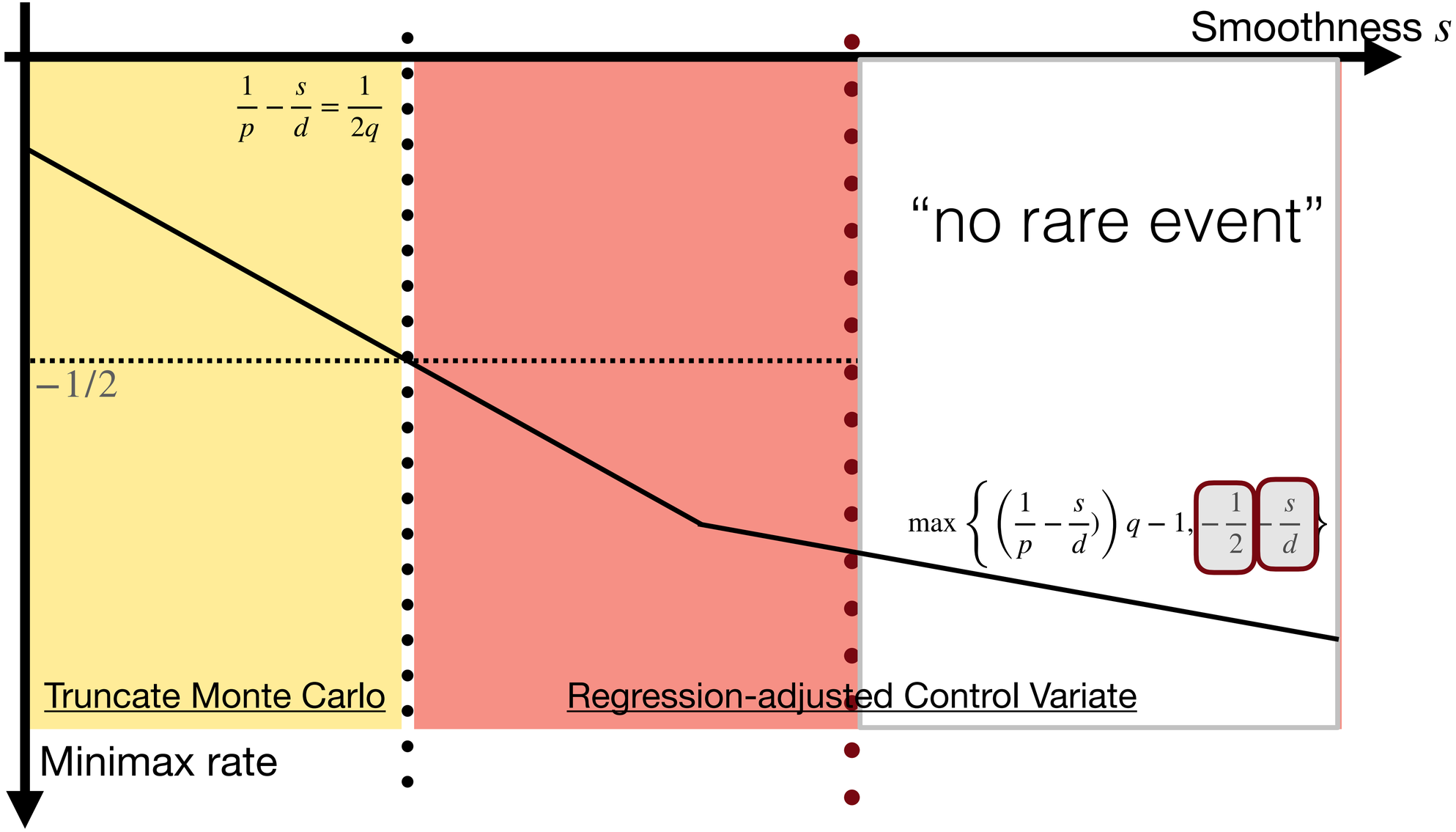
Rare Event and Smoothness…



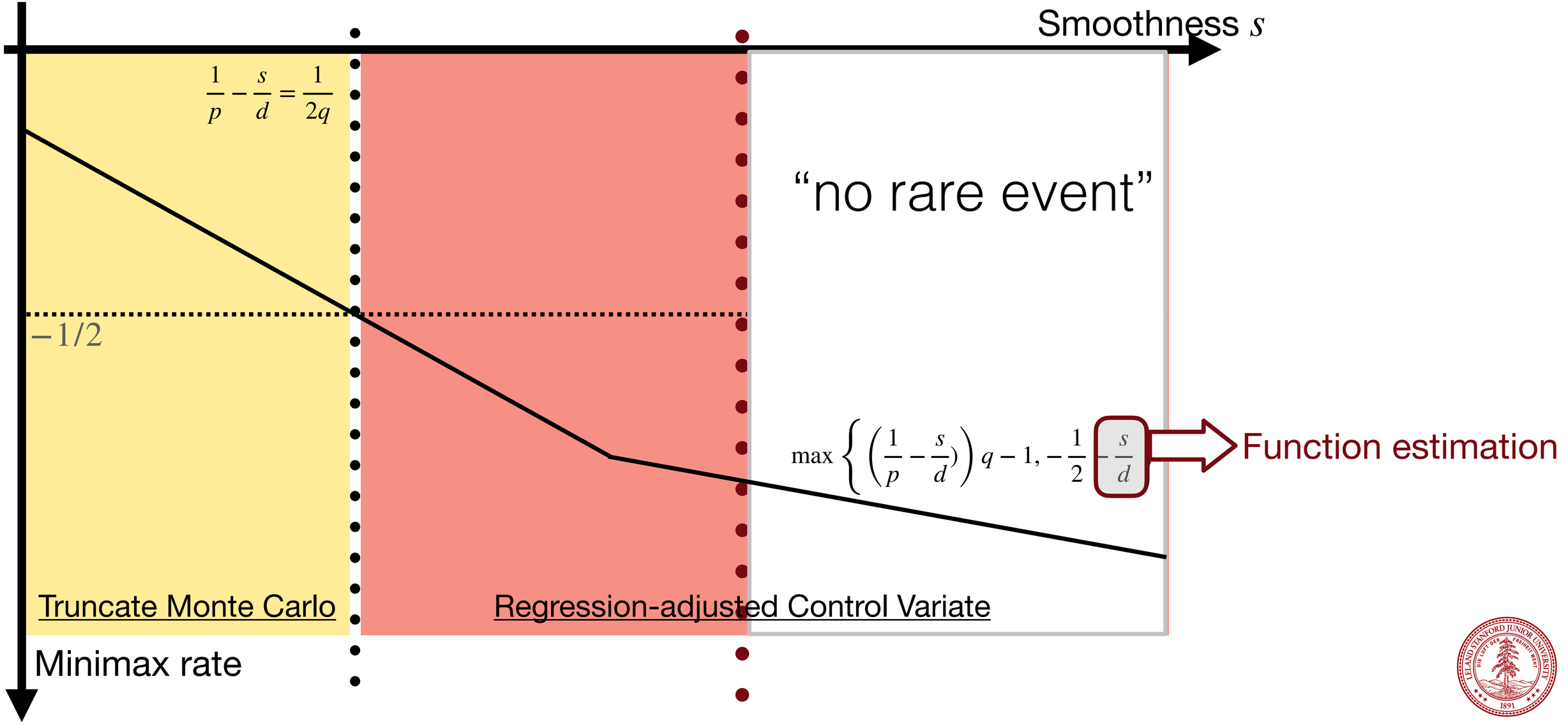
When the control variate helps



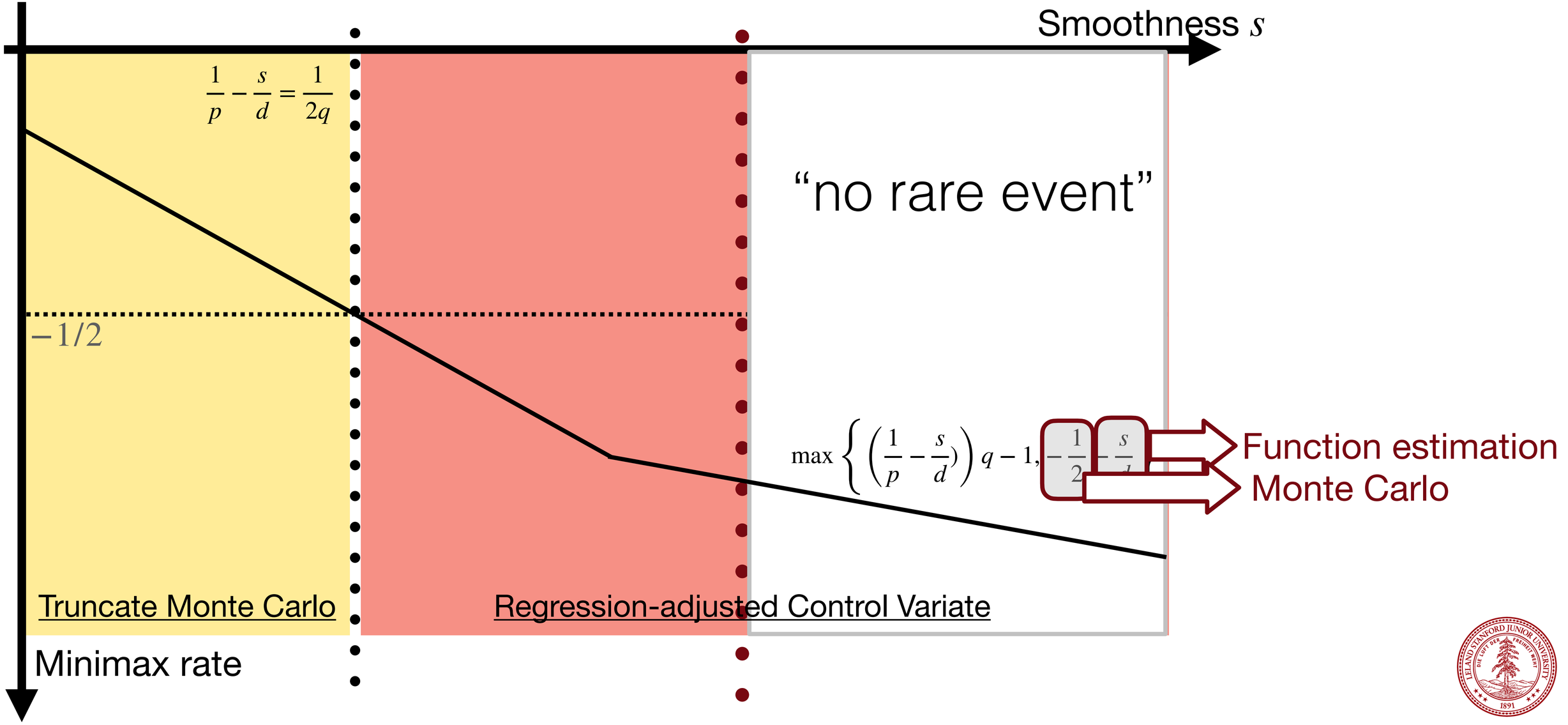
When the control variate helps



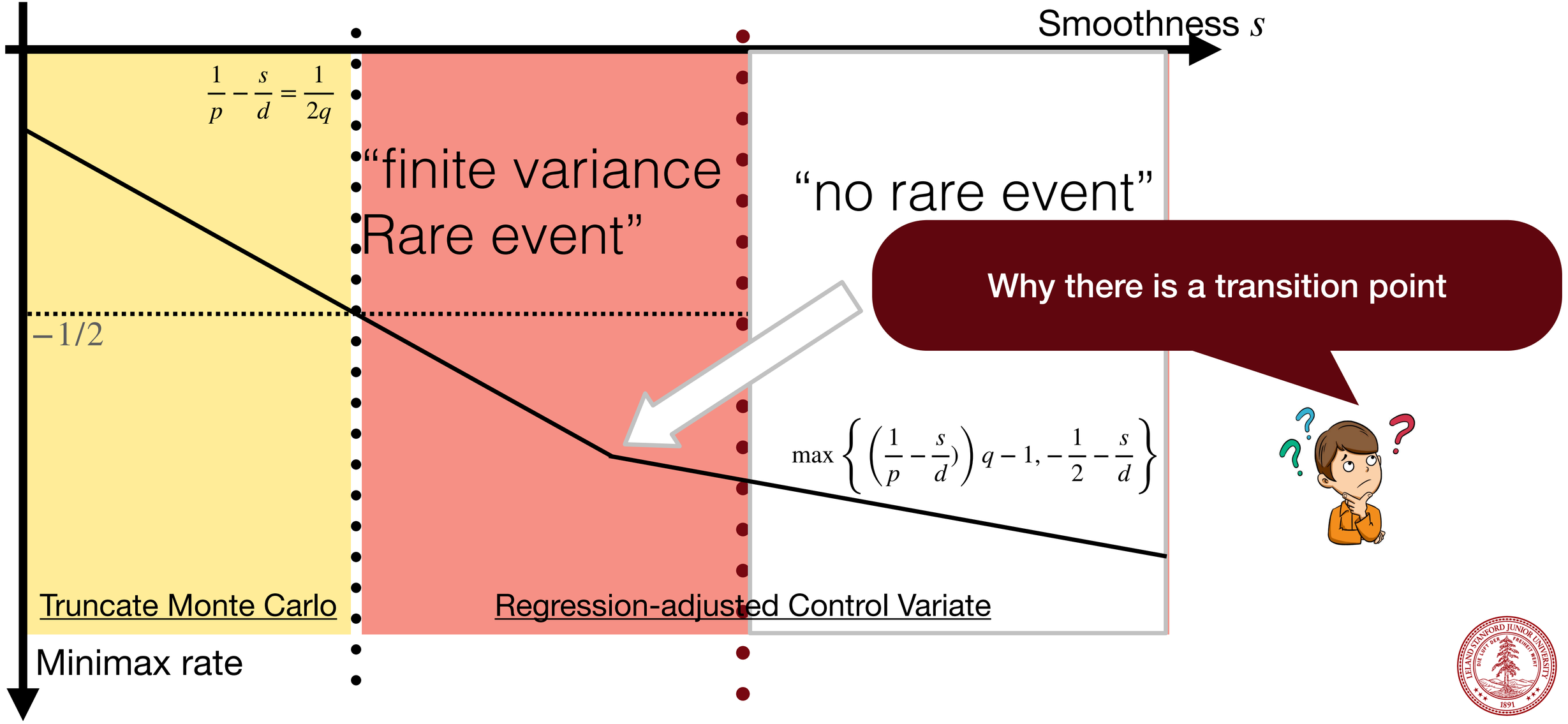
When the control variate helps



When the control variate helps



When the control variate helps



Semi-parametric efficiency…

Example Monte Carlo Estimate $\mathbb{E}_P f$ $\mathbb{E}_P f^q, f \in W^{s,p}$

Step 1 Using half of the data to estimate $\hat{f}$

Step 2 $\mathbb{E}_P f^q = \mathbb{E}_P(\hat{f})^q + \mathbb{E}_P(f^q - \hat{f}^q)$

Low order term

$f^{q-1}(f - \hat{f}) + (f - \hat{f})^q$

"influnce function" (gradient)

Semi-parametric efficiency…

Example Monte Carlo Estimate $\mathbb{E}_P f$ $\mathbb{E}_P f^q, f \in W^{s,p}$

Step 1 Using half of the data to estimate $\hat{f}$

Step 2 $\mathbb{E}_P f^q = \mathbb{E}_P(\hat{f})^q + \mathbb{E}_P(f^q - \hat{f}^q)$

Low order term

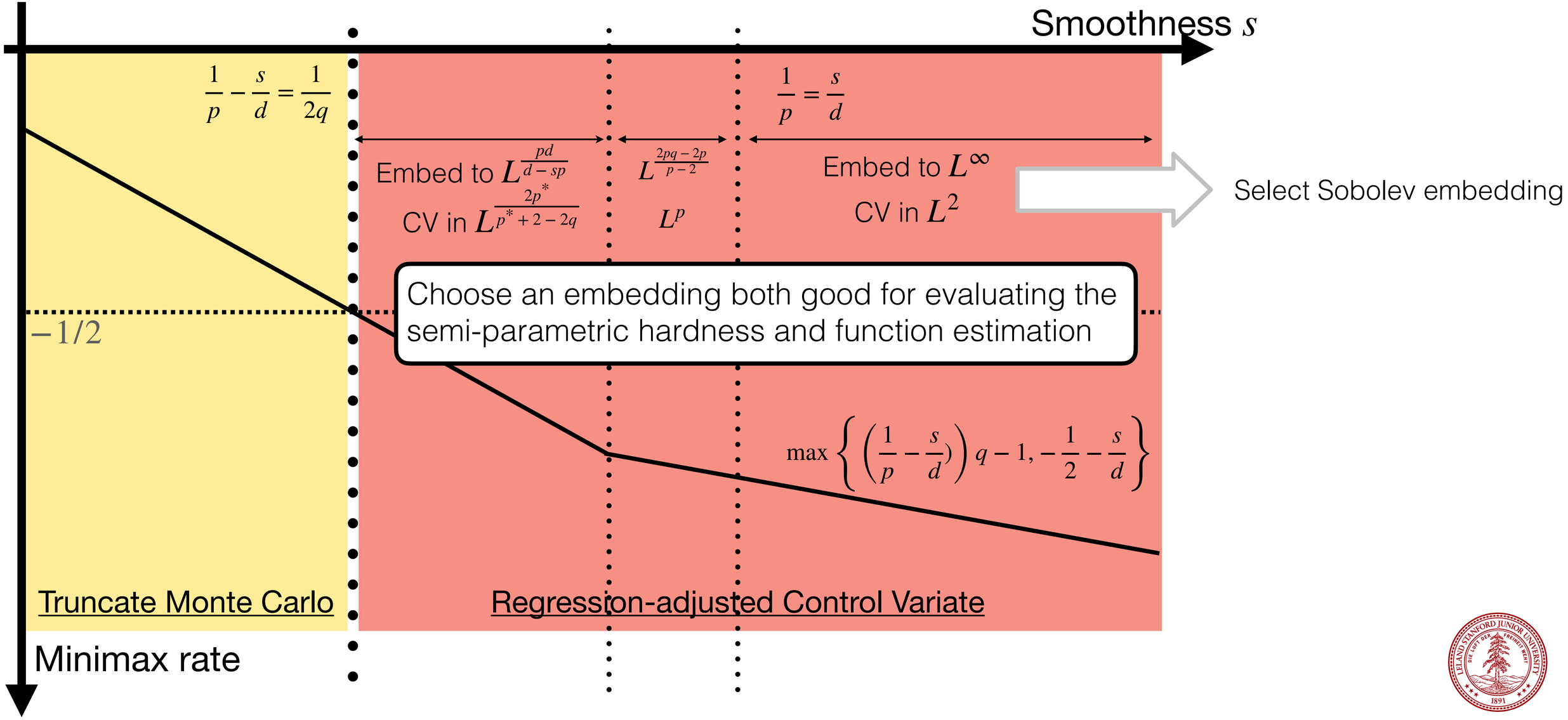
$f^{q-1}(f - \hat{f}) + (f - \hat{f})^q$

"influnce function" (gradient)

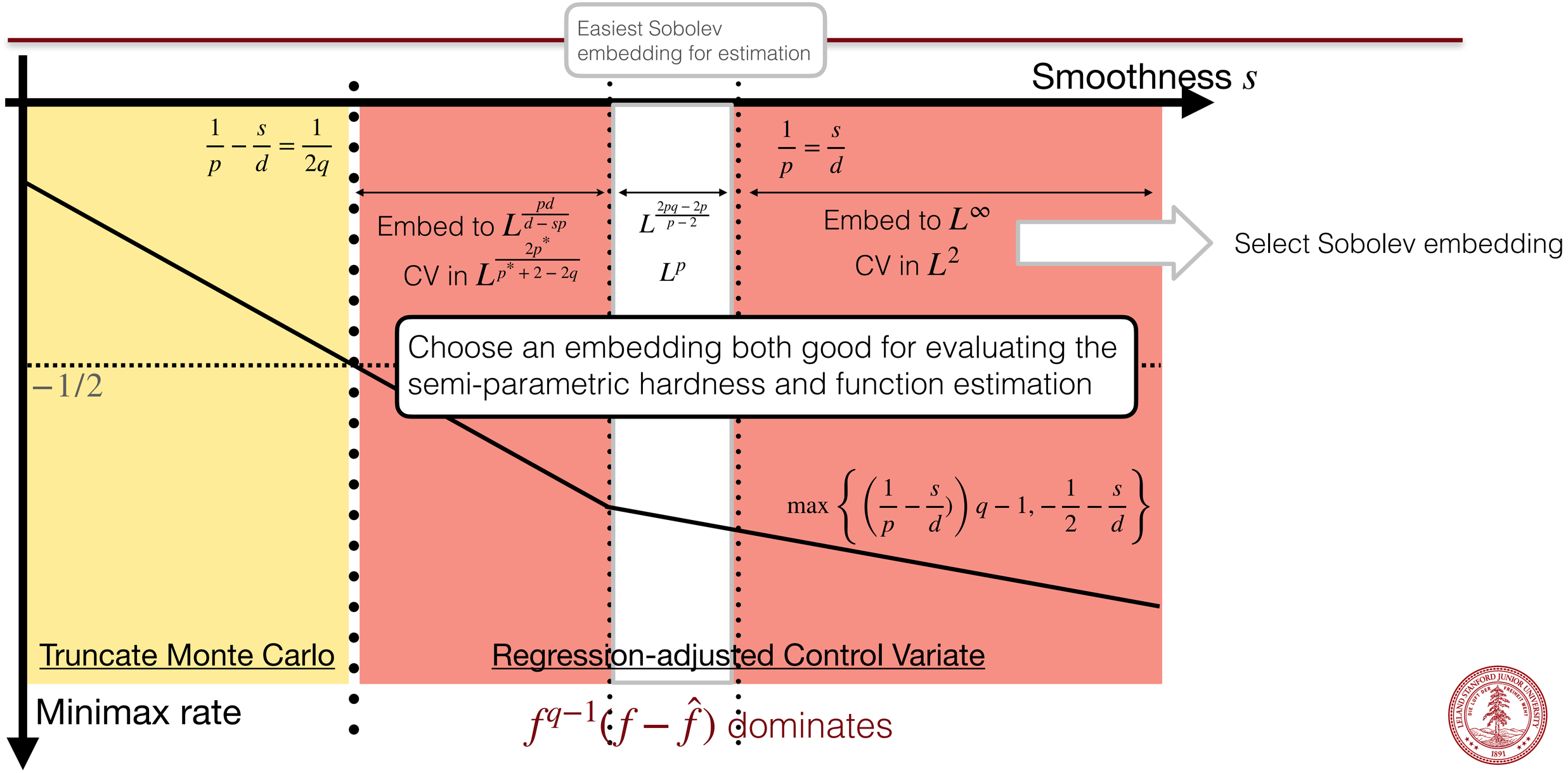
Embed f^{q-1} and $f - \hat{f}$ into "dual" space

How to select the sobolev emebedding

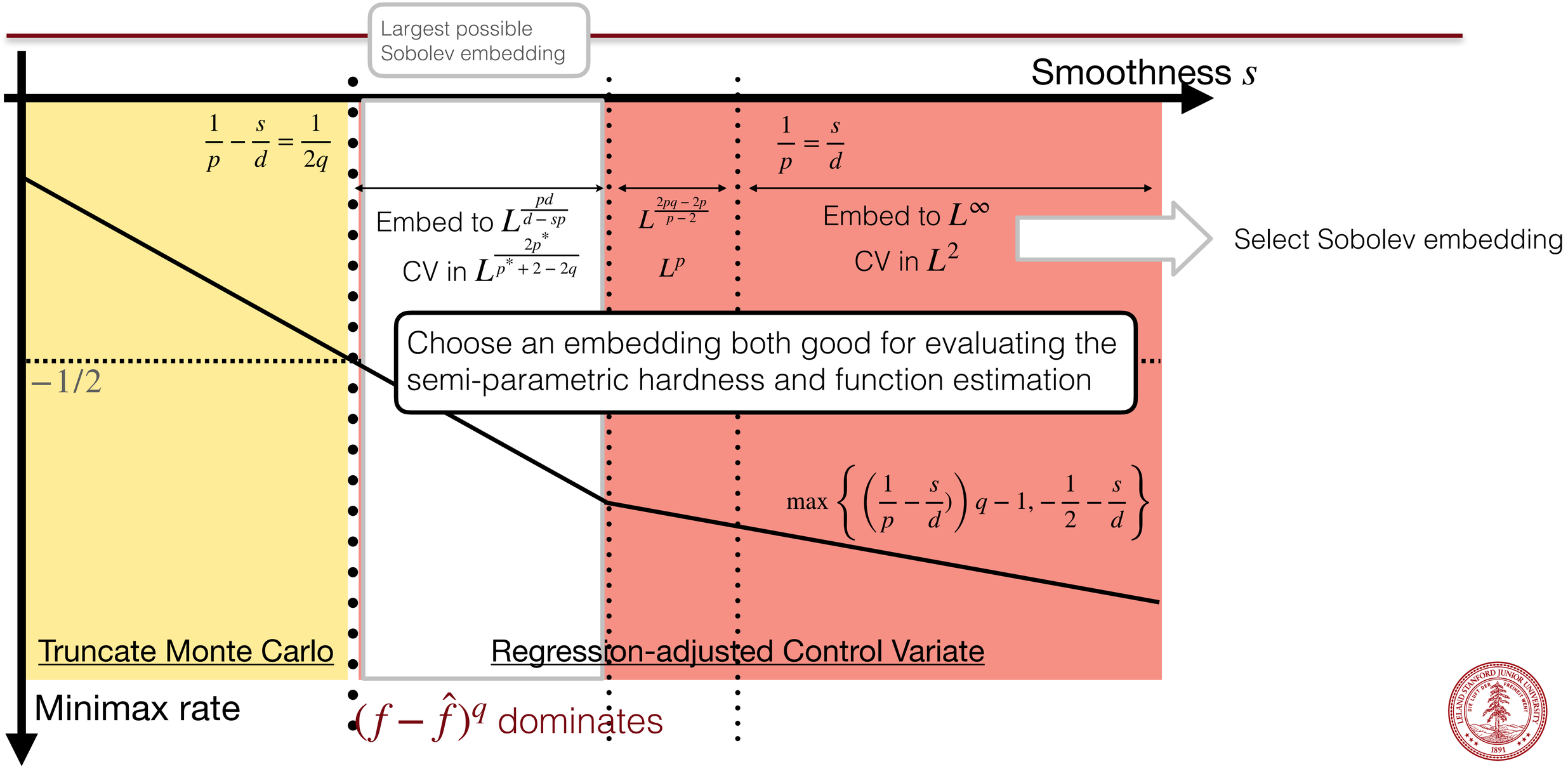
Tricky part of the Proof:select embedding



Tricky part of the Proof:select embedding

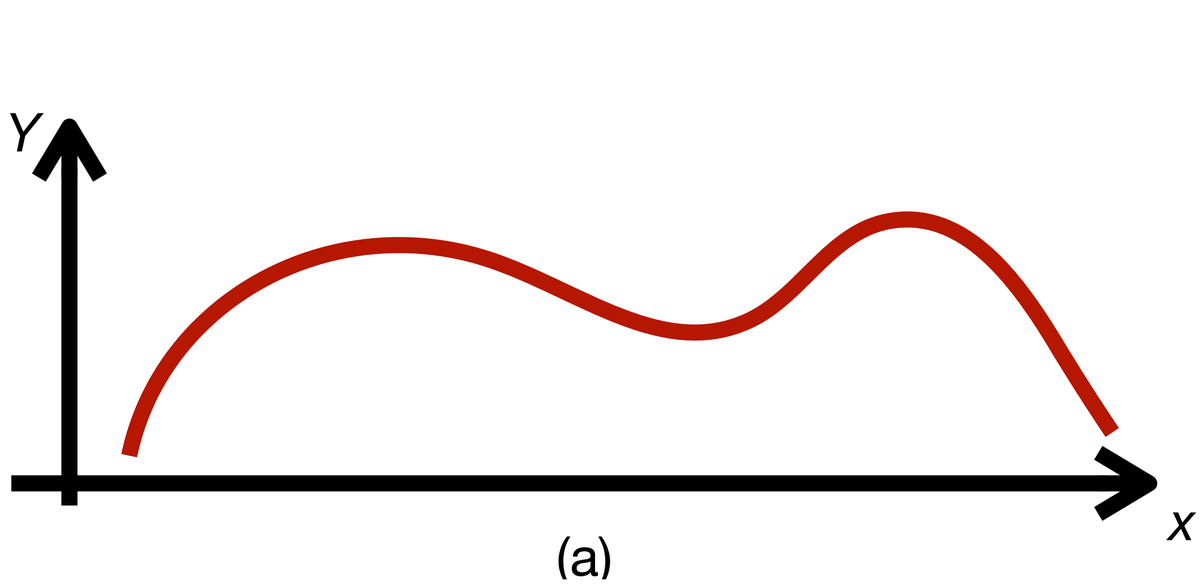


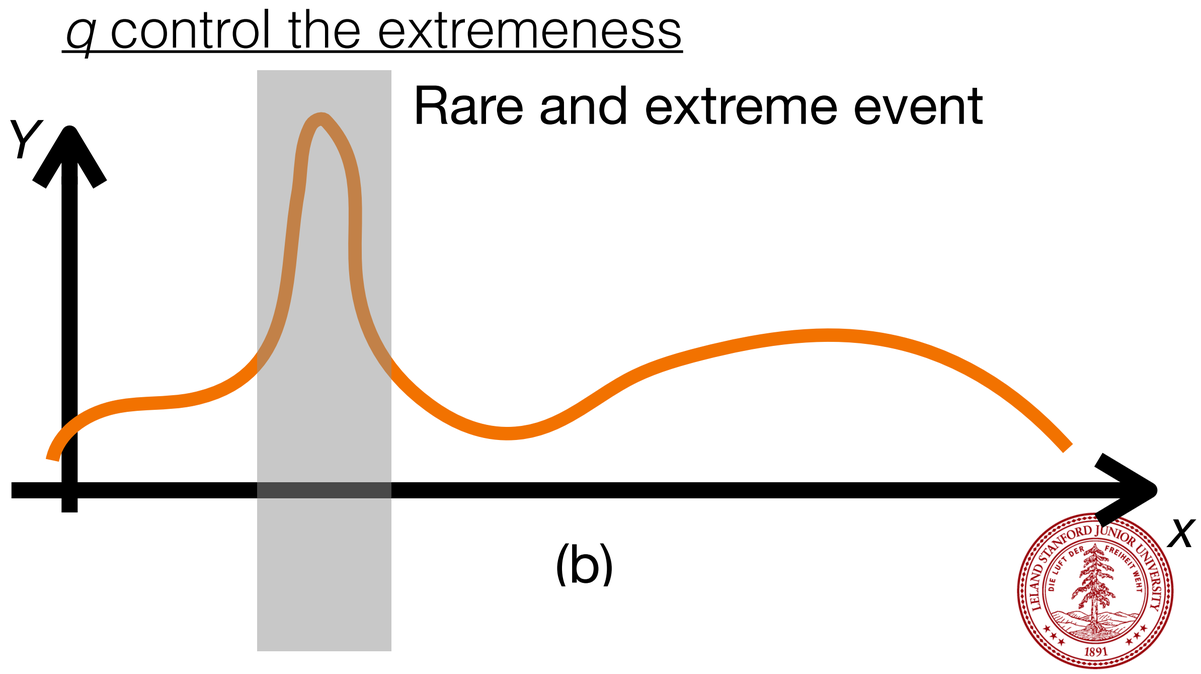
Tricky part of the Proof:select embedding



Take home message

a) Statistical optimal regression is the optimal control variate
b) It helps only if there isn't a hard to simulate (infinite variance) Rare and extreme event





Contact: yplu@stanford.edu