Random Restrictions on Boolean Functions with Small Influences

Ronen Eldan, Avi Wigderson, *Pei Wu*

Oct., 2022

Some background

Influences

$$f : \{0,1\}^n \to \{0,1\} .$$

Influences

$f : \{0,1\}^n \to \{0,1\}$.

Definition (influence and max influence)

$$\mathrm{Inf}_i[f] = \Pr_x[f(x) \neq f(x \oplus e_i)],$$

$$\mathrm{MaxInf}[f] = \max_{i \in [n]} \mathrm{Inf}_i(f) \, .$$

Examples

1. $\mathrm{MAJ_n} : x \mapsto \mathbf{I}_{\{|x|>n/2\}}$.

Examples

1. $\mathrm{MAJ_n} : x \mapsto \mathbf{I}_{\{|x|>n/2\}}$.

Fact. $\mathrm{MaxInf}(\mathrm{MAJ}_n) = \Theta\left(\frac{1}{\sqrt{n}}\right).$

Examples

1. $\mathrm{MAJ}_{\mathrm{n}} : x \mapsto \mathbf{I}_{\{|x|>n/2\}}$.

Fact. $\mathrm{MaxInf}(\mathrm{MAJ}_n) = \Theta\left(\frac{1}{\sqrt{n}}\right)$.

Proof:

$$\mathrm{Inf}_i(f) = \Pr\left[\left| |x| - \frac{n}{2} \right| \leq 1\right] \cdot \Theta(1) .$$

Examples

1. $\mathrm{MAJ_n} : x \mapsto \mathbf{I}_{\{|x|>n/2\}}$.

Fact. $\mathrm{MaxInf}(\mathrm{MAJ}_n) = \Theta\left(\frac{1}{\sqrt{n}}\right)$.

Proof:

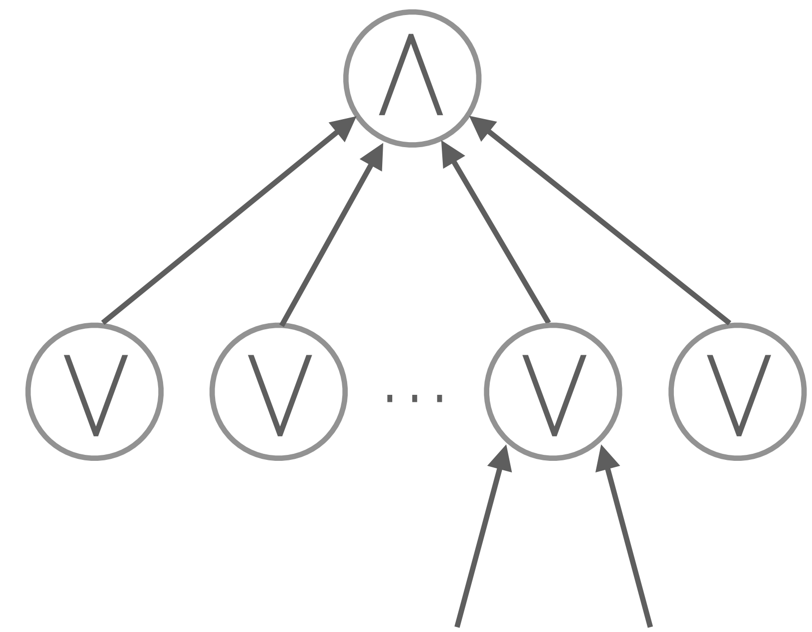
$$\mathrm{Inf}_i(f) = \Pr\left[\left| |x| - \frac{n}{2} \right| \leq 1\right] \cdot \Theta(1) .$$

e.g. $0101010101$0, any 0 is sensitive

Examples

2. $\mathrm{TRIBES_n}(x) = \bigwedge_{i=1}^{s} \bigvee_{j=1}^{w} (x_{i,j}),$

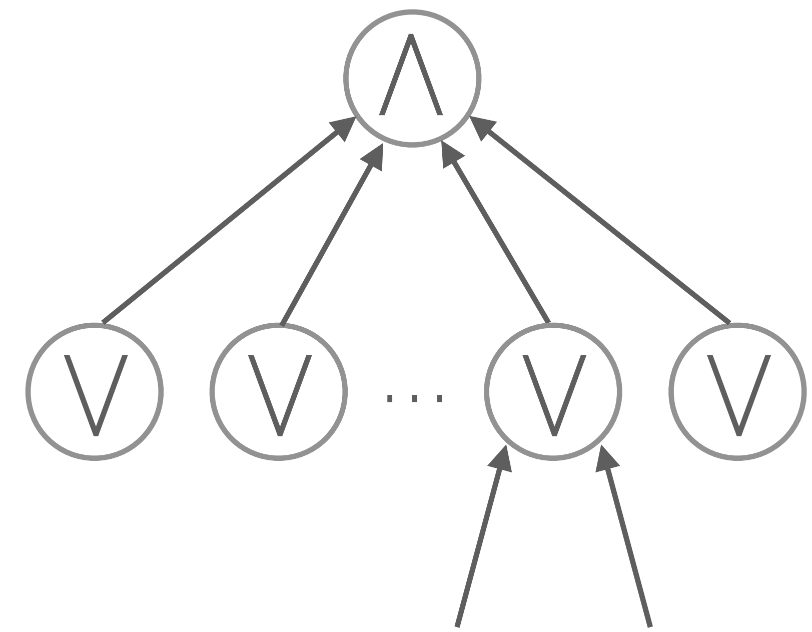
$$s = n/w, w \approx \log n - \log\log n.$$



Examples

2. $\mathrm{TRIBES_n}(x) = \bigwedge_{i=1}^{s} \bigvee_{j=1}^{w} (x_{i,j}),$

$$s = n/w, w \approx \log n - \log\log n.$$

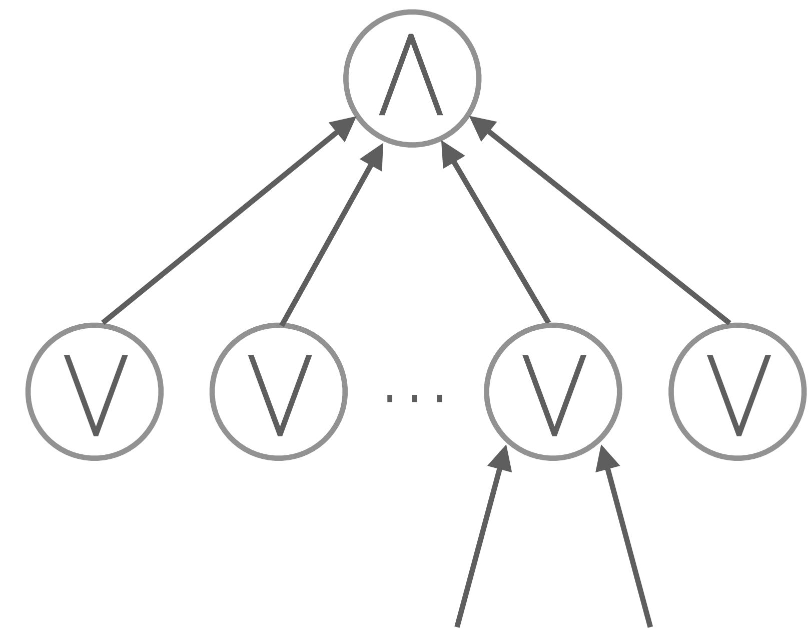


Fact. $\mathrm{MaxInf}(\mathrm{TRIBES}_n) = \Theta\left(\frac{\log n}{n}\right).$

Examples

2. $\mathrm{TRIBES_n}(x) = \bigwedge_{i=1}^{s} \bigvee_{j=1}^{w} (x_{i,j}),$

$s = n/w, w \approx \log n - \log\log n.$



Fact. $\mathrm{MaxInf}(\mathrm{TRIBES}_n) = \Theta\left(\frac{\log n}{n}\right).$

Proof sketch:

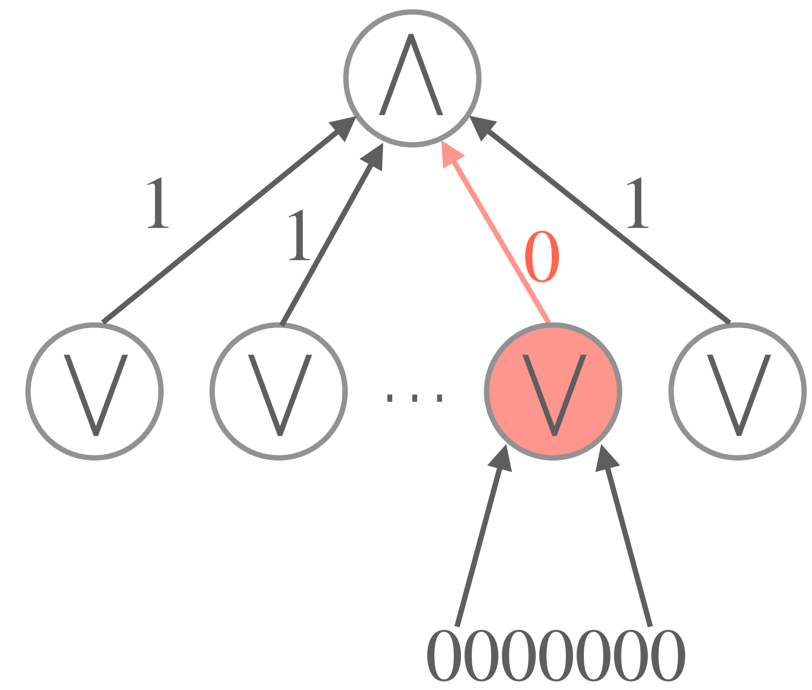
Suffices to check the case when $\mathrm{TRIBES}(x) = 0$.

With $\Theta(1)$ probability, exactly one OR gate evaluates to 0.

Examples

2. $\text{TRIBES}_{\text{n}}(x) = \bigwedge_{i=1}^{s} \bigvee_{j=1}^{w} (x_{i,j}),$

$$s = n/w, w \approx \log n - \log\log n.$$



Fact. $\text{MaxInf}(\text{TRIBES}_n) = \Theta\left(\frac{\log n}{n}\right).$

Proof sketch:

Suffices to check the case when $\text{TRIBES}(x) = 0$.

With $\Theta(1)$ probability, exactly one OR gate evaluates to 0.

Functions with small influences

Assume: $\mathrm{MaxInf}[f] = o(1), \mathrm{Var}[f] = \Omega(1)$.

Functions with small influences

Assume: $\mathrm{MaxInf}[f] = o(1), \mathrm{Var}[f] = \Omega(1)$.

Some motivations:

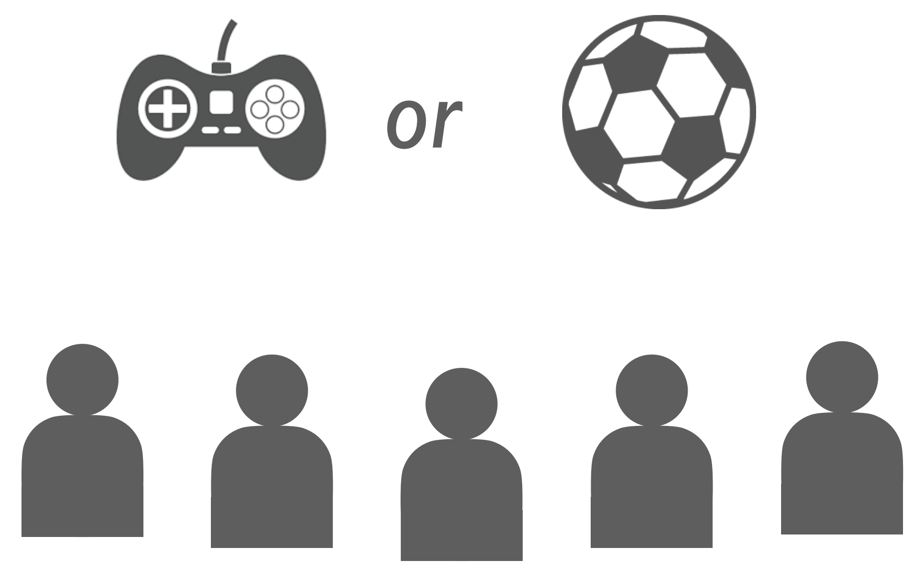
1. Social choice theory

Functions with small influences

Assume: $\mathrm{MaxInf}[f] = o(1), \mathrm{Var}[f] = \Omega(1)\,.$

Some motivations:

1. Social choice theory

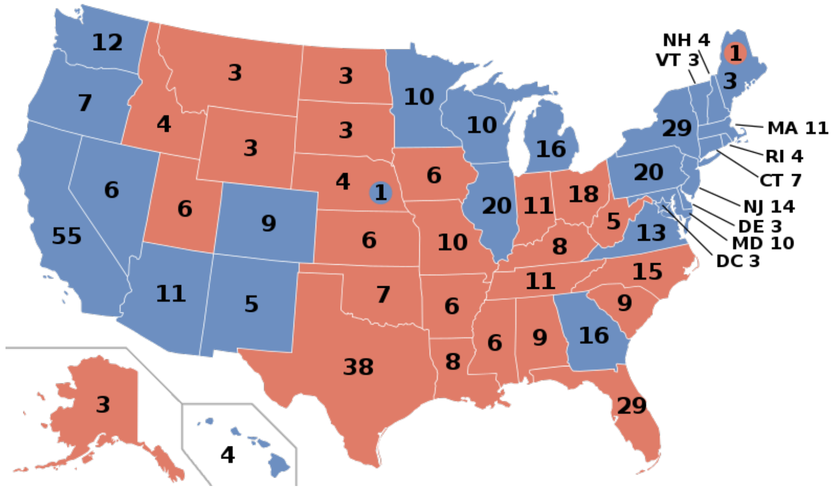


Functions with small influences

Assume: $\mathrm{MaxInf}[f] = o(1), \mathrm{Var}[f] = \Omega(1)$.

Some motivations:

1. Social choice theory

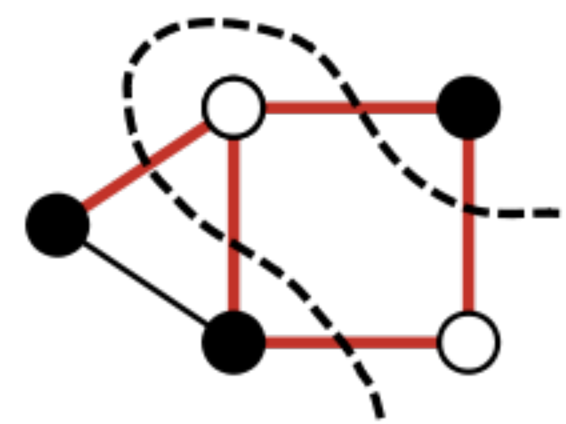


Functions with small influences

Assume: $\mathrm{MaxInf}[f] = o(1), \mathrm{Var}[f] = \Omega(1)$.

Some motivations:

1. Social choice theory
2. PCP/Hardness of approximation



Functions with small influences

Assume: $\mathrm{MaxInf}[f] = o(1), \mathrm{Var}[f] = \Omega(1)$.

Some motivations:

1. Social choice theory
2. PCP/Hardness of approximation
3. Invariance principle

$\sum X_i \approx$ ***Gaussian***

Random restrictions

$f : \{0,1\}^n \to \{0,1\}$.

Random restrictions

$f : \{0,1\}^n \to \{0,1\}$.

0	*	1	0	*	1	0	*	0	1	…	…

Random restrictions

$f : \{0,1\}^n \to \{0,1\}$.

0	*	1	0	*	1	0	*	0	1	…	…

p-random restriction $f|_{R_p}$: fix p random bits

Our results

Theorem (Eldan, Wigderson, W.)

For any $f:\{0,1\}^n \to \{0,1\}$, with $\mathrm{MaxInf}(f) = \tau = o(1)$, $\Omega(1)$ variance. Then for alive probability $\rho = \tilde{\Omega}\left(\frac{1}{\log 1/\tau}\right)$, we have

$$\Pr[f|_{R_{1-\rho}} \text{ is nonconstant}] = 1 - o(1)\,.$$

Our results

Theorem (Eldan, Wigderson, W.)

For any $f : \{0,1\}^n \to \{0,1\}$, with $\mathrm{MaxInf}(f) = \tau = o(1)$, $\Omega(1)$ variance. Then for alive probability $\rho = \tilde{\Omega}\left(\frac{1}{\log 1/\tau}\right)$, we have

$$\Pr[f|_{R_{1-\rho}} \text{ is nonconstant}] = 1 - o(1)\,.$$

Remarks:

1. *Optimality (w.r.t. ρ)*: the Tribes function.

Our results

Theorem (Eldan, Wigderson, W.)
For any $f: \{0,1\}^n \to \{0,1\}$, with $\mathrm{MaxInf}(f) = \tau = o(1)$, $\Omega(1)$ variance. Then for alive probability $\rho = \tilde{\Omega}\left(\frac{1}{\log 1/\tau}\right)$, we have

$$\Pr[f|_{R_{1-\rho}} \text{ is nonconstant}] = 1 - o(1) .$$

Remarks:

1. *Optimality (w.r.t. ρ)*: the Tribes function.

Proof. by Hastad's Switching Lemma

Our results

Theorem (Eldan, Wigderson, W.)

For any $f : \{0,1\}^n \to \{0,1\}$, with $\mathrm{MaxInf}(f) = \tau = o(1)$, $\Omega(1)$ variance. Then for alive probability

$\rho = \tilde{\Omega}\left(\frac{1}{\log 1/\tau}\right)$, we have

$$\Pr[f|_{R_{1-\rho}} \text{ is nonconstant}] = 1 - o(1)\,.$$

$$\Pr[\mathrm{Var}[f|_R] \le p^{\tilde{\Theta}(1/\rho)}] \le p$$

Remarks:

1. *Optimality (w.r.t. ρ)*: the Tribes function.

Our results

Theorem (Eldan, Wigderson, W.)

For any $f: \{0,1\}^n \to \{0,1\}$, with $\mathrm{MaxInf}(f) = \tau = o(1)$, $\Omega(1)$ variance. Then for alive probability $\rho = \tilde{\Omega}\left(\frac{1}{\log 1/\tau}\right)$, we have

$$\Pr[f|_{R_{1-\rho}} \text{ is nonconstant}] = 1 - o(1)\,.$$

$$\Pr[\mathrm{Var}[f|_R] \le p^{\tilde{\Theta}(1/\rho)}] \le p$$

Remarks:

1. *Optimality (w.r.t.* ρ*)*: the Tribes function.
2. *Optimality (w.r.t.* Var*)*: the Majority function.

Our results

Theorem (Eldan, Wigderson, W.)
For any $f : \{0,1\}^n \to \{0,1\}$, with $\mathrm{MaxInf}(f) = \tau = o(1)$, $\Omega(1)$ variance. Then for alive probability $\rho = \tilde{\Omega}\left(\frac{1}{\log 1/\tau}\right)$, we have

$$\Pr[f|_{R_{1-\rho}} \text{ is nonconstant}] = 1 - o(1)\,.$$

Remarks (more):

- "It ain't over till it's over" conjecture! — Friedgut-Kalai '06

Our results

Theorem (Eldan, Wigderson, W.)
For any $f : \{0,1\}^n \to \{0,1\}$, with $\mathrm{MaxInf}(f) = \tau = o(1)$, $\Omega(1)$ variance. Then for alive probability $\rho = \tilde{\Omega}\left(\frac{1}{\log 1/\tau}\right)$, we have

$$\Pr[f|_{R_{1-\rho}} \text{ is nonconstant}] = 1 - o(1)\,.$$

Remarks (more):

- "It ain't over till it's over" conjecture! — Friedgut-Kalai '06
- Proved by Mossel-O'Donnell-Oleszkiewicz '10

Our results

Theorem (Eldan, Wigderson, W.)
For any $f : \{0,1\}^n \to \{0,1\}$, with $\mathrm{MaxInf}(f) = \tau = o(1)$, $\Omega(1)$ variance. Then for alive probability $\rho = \tilde{\Omega}\left(\frac{1}{\log 1/\tau}\right)$, we have

$$\Pr[f|_{R_{1-\rho}} \text{ is nonconstant}] = 1 - o(1).$$

Remarks (more):

- "It ain't over till it's over" conjecture! — Friedgut-Kalai '06
- Proved by Mossel-O'Donnell-Oleszkiewicz '10
- Koehler-Lifshitz-Minzer-Mossel have a different approach '22

One application

Corollary.
For any balanced function f, with $\tau = o(1)$ max influence, then

$$\Pr[\mathrm{bs}_f(x) \geq \tilde{\Omega}(\log(1/\tau))] = 1 - o(1)\,.$$

Sensitivity $s_f(x)$: number of sensitive bits

$$s_f(x) := |\{i : f(x) \neq f(x \oplus e_i)\}|\,.$$

Block sensitivity $\mathrm{bs}_f(x)$: max number of disjoint sensitive blocks

$$\mathrm{bs}_f(x) := \max |\{\text{disjoint } S_1, S_2, \ldots, S_k \subseteq [n] : f(x) \neq f(x \oplus 1_{S_i})\}|\,.$$

One application

Corollary.

For any balanced function f, with $\tau = o(1)$ max influence, then

$$\Pr[\mathrm{bs}_f(x) \geq \tilde{\Omega}(\log(1/\tau))] = 1 - o(1)\,.$$

One application

Corollary.
For any balanced function f, with $\tau = o(1)$ max influence, then

$$\Pr[\mathrm{bs}_f(x) \geq \tilde{\Omega}(\log(1/\tau))] = 1 - o(1)\,.$$

No such result holds for sensitivity.

One application

Corollary.
For any balanced function f, with $\tau = o(1)$ max influence, then

$$\Pr[\mathrm{bs}_f(x) \geq \tilde{\Omega}(\log(1/\tau))] = 1 - o(1)\,.$$

No such result holds for sensitivity.

Remark: By KKL inequality,

$$\mathbf{E}[\mathrm{bs}_f(x)] \geq \mathbf{E}[\mathrm{s}_f(x)] \geq \Omega(\log(1/\tau))\,.$$

One application

Corollary.
For any balanced function f, with $\tau = o(1)$ max influence, then

$$\Pr[\mathrm{bs}_f(x) \geq \tilde{\Omega}(\log(1/\tau))] = 1 - o(1)\,.$$

Proof sketch.

One application

Corollary.
For any balanced function f, with $\tau = o(1)$ max influence, then
$$\Pr[\mathrm{bs}_f(x) \geq \tilde{\Omega}(\log(1/\tau))] = 1 - o(1)\,.$$

Proof sketch.

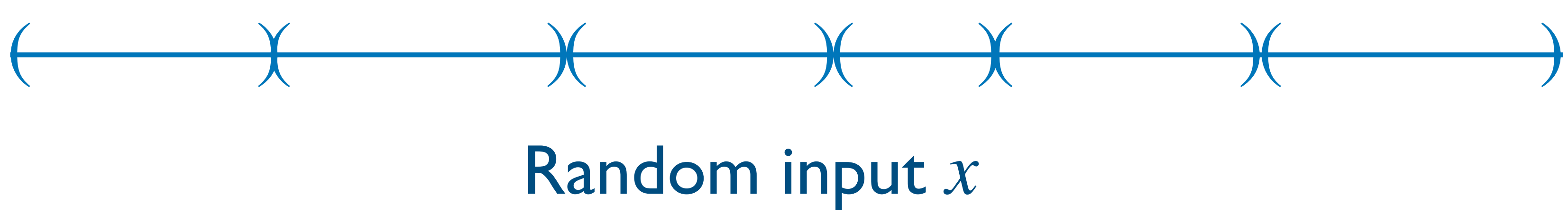
Random input x

One application

Corollary.
For any balanced function f, with $\tau = o(1)$ max influence, then

$$\Pr[\mathrm{bs}_f(x) \geq \tilde{\Omega}(\log(1/\tau))] = 1 - o(1)\,.$$

Proof sketch.



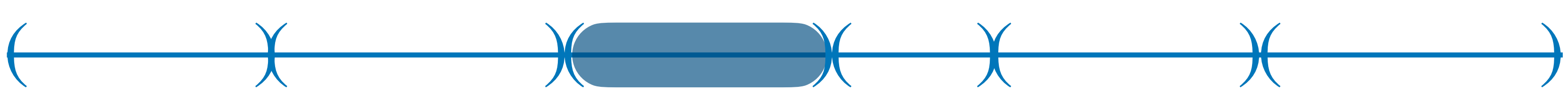
Partition [n] into $M = \tilde{O}(\log(1/\tau))$ random blocks,

One application

Corollary.
For any balanced function f, with $\tau = o(1)$ max influence, then

$$\Pr[\mathrm{bs}_f(x) \geq \tilde{\Omega}(\log(1/\tau))] = 1 - o(1)\,.$$

Proof sketch.



Random input x

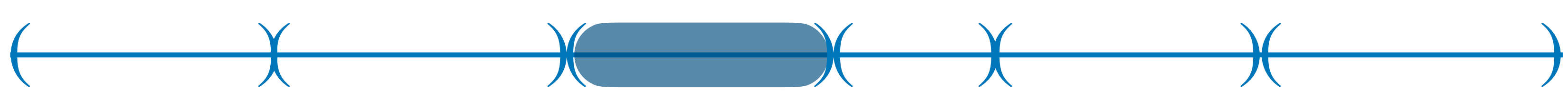
Partition [n] into $M = \tilde{O}(\log(1/\tau))$ random blocks,

One application

Corollary.
For any balanced function f, with $\tau = o(1)$ max influence, then

$$\Pr[\mathrm{bs}_f(x) \geq \tilde{\Omega}(\log(1/\tau))] = 1 - o(1)\,.$$

Proof sketch.



Random input x

Partition [n] into $M = \tilde{O}(\log(1/\tau))$ random blocks,

Any $M - 1$ blocks, induces a random restriction.

Proof of the main result

Two random processes

$f : \{-1,1\}^n \to \{0,1\}$ (by multilinear extension $f : \mathbb{R}^n \to \mathbb{R}$)

Two random processes

$f : \{-1,1\}^n \to \{0,1\}$ (by multilinear extension $f : \mathbb{R}^n \to \mathbb{R}$)

$f = 1 + x_1 + x_2 + x_1 x_2$

Two random processes

$f : \{-1,1\}^n \to \{0,1\}$ (by multilinear extension $f : \mathbb{R}^n \to \mathbb{R}$)

$f = 1 + x_1 + x_2 + x_1 x_2$

Uniform process

1. Random permutation π,
2. $X(0) = 0^n$,
3. $X_i(t) = X_i(t-1)$, for $i \neq \pi(t)$;
   $X_{\pi(t)}(t) \sim \{-1,1\}$, for $t = 1,2,\ldots,n$

Two random processes

$f : \{-1,1\}^n \to \{0,1\}$ (by multilinear extension $f : \mathbb{R}^n \to \mathbb{R}$)

$f = 1 + x_1 + x_2 + x_1 x_2$

Uniform process

1. Random permutation π,
2. $X(0) = 0^n$,
3. $X_i(t) = X_i(t-1)$, for $i \neq \pi(t)$;
   $X_{\pi(t)}(t) \sim \{-1,1\}$, for $t = 1,2,\ldots,n$

$X(0)=$ 0 0 0 0 0 0 0 0 ……

Two random processes

$f : \{-1,1\}^n \to \{0,1\}$ (by multilinear extension $f : \mathbb{R}^n \to \mathbb{R}$)

$f = 1 + x_1 + x_2 + x_1 x_2$

Uniform process

1. Random permutation π,
2. $X(0) = 0^n$,
3. $X_i(t) = X_i(t-1)$, for $i \neq \pi(t)$;
   $X_{\pi(t)}(t) \sim \{-1,1\}$, for $t = 1,2,\ldots,n$

$X(0)=$ 0 0 0 0 0 0 0 0 ……

$X(1)=$ 0 0 0 0 0 1 0 0 ……

Two random processes

$f: \{-1,1\}^n \to \{0,1\}$ (by multilinear extension $f: \mathbb{R}^n \to \mathbb{R}$)

$f = 1 + x_1 + x_2 + x_1x_2$

Uniform process

1. Random permutation π,
2. $X(0) = 0^n$,
3. $X_i(t) = X_i(t-1)$, for $i \neq \pi(t)$;
   $X_{\pi(t)}(t) \sim \{-1,1\}$, for $t = 1,2,\ldots,n$

$X(0)=$ 0 0 0 0 0 0 0 0 ……

$X(1)=$ 0 0 0 0 0 1 0 0 ……

$X(2)=$ 0 0 -1 0 0 1 0 0 ……

Two random processes

$f:\{-1,1\}^n \to \{0,1\}$ (by multilinear extension $f:\mathbb{R}^n \to \mathbb{R}$)

Uniform process

1. Random permutation π,
2. $X(0) = 0^n$,
3. $X_i(t) = X_i(t-1)$, for $i \neq \pi(t)$;
   $X_{\pi(t)}(t) \sim \{-1,1\}$, for $t = 1,2,\ldots,n$

Observation.

i. $X(t)$ induces a random restriction $f|_{R(t)}$,
ii. $X(n)$ is a uniformly random element from $\{-1,1\}^n$,
iii. $f(X(t)) = \mathbf{E}_{z\in\{-1,1\}^{n-t}}[f|_{R(t)}(z)]$ is a martingale.

Two random processes

$f : \{-1,1\}^n \to \{0,1\}$ (by multilinear extension $f : \mathbb{R}^n \to \mathbb{R}$)

Uniform process

1. Random permutation π,
2. $X(0) = 0^n$,
3. $X_i(t) = X_i(t-1)$, for $i \neq \pi(t)$;
   $X_{\pi(t)}(t) \sim \{-1,1\}$, for $t = 1,2,\ldots,n$

Observation.

i. $X(t)$ induces a random restriction $f|_{R(t)}$,

ii. $X(n)$ is a uniformly random element from $\{-1,1\}^n$,

iii. $f(X(t)) = \mathbf{E}_{z\in\{-1,1\}^{n-t}}[f|_{R(t)}(z)]$ is a martingale.

e.g. $f = 1 + x_1 + x_2 + x_1x_2$,
$f(0) = \mathbf{E}_{x\in\{-1,1\}^n}[f(x)]$

Two random processes

Conditioned process

1. Random permutation π,
2. $Y(0) = 0^n$,
3. for $t = 1,2,\ldots,n$, let $Y_i(t) = Y_i(t-1)$,

$$\Pr\{Y_{\pi(t)}(t) = \pm 1\} = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2f(Y(t-1))}.$$

Two random processes

Conditioned process

1. Random permutation π,
2. $Y(0) = 0^n$,
3. for $t = 1,2,\ldots,n$, let $Y_i(t) = Y_i(t-1)$,

$$\Pr\{Y_{\pi(t)}(t) = \pm 1\} = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2f(Y(t-1))} .$$

Fact. $Y(n)$ is a uniformly random element from $f^{-1}(1)$

Two random processes

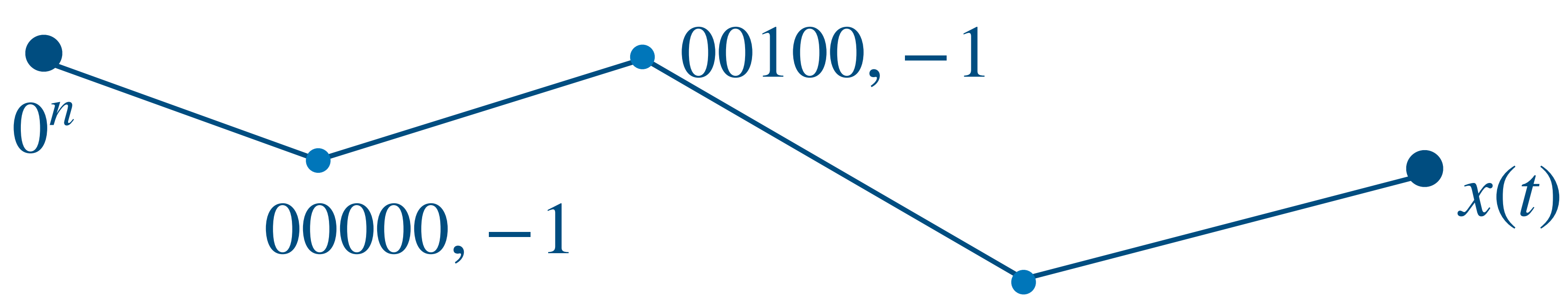
Conditioned process

1. Random permutation π,
2. $Y(0) = 0^n$,
3. for $t = 1,2,\ldots,n$, let $Y_i(t) = Y_i(t-1)$,

$$\Pr\{Y_{\pi(t)}(t) = \pm 1\} = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2f(Y(t-1))}.$$

Fact. Fix some some path $x(0), x(1), \ldots, x(t) \in \{-1,0,1\}^n$,

$$\frac{\Pr[\forall i \in [t], Y(t) = x(t)]}{\Pr[\forall i \in [t], X(t) = x(t)]} = \frac{f(x(t))}{f(0)}.$$



Two random processes

Conditioned process

1. Random permutation π,
2. $Y(0) = 0^n$,
3. for $t = 1,2,\ldots,n$, let $Y_i(t) = Y_i(t-1)$,

$$\Pr\{Y_{\pi(t)}(t) = \pm 1\} = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2f(Y(t-1))}.$$

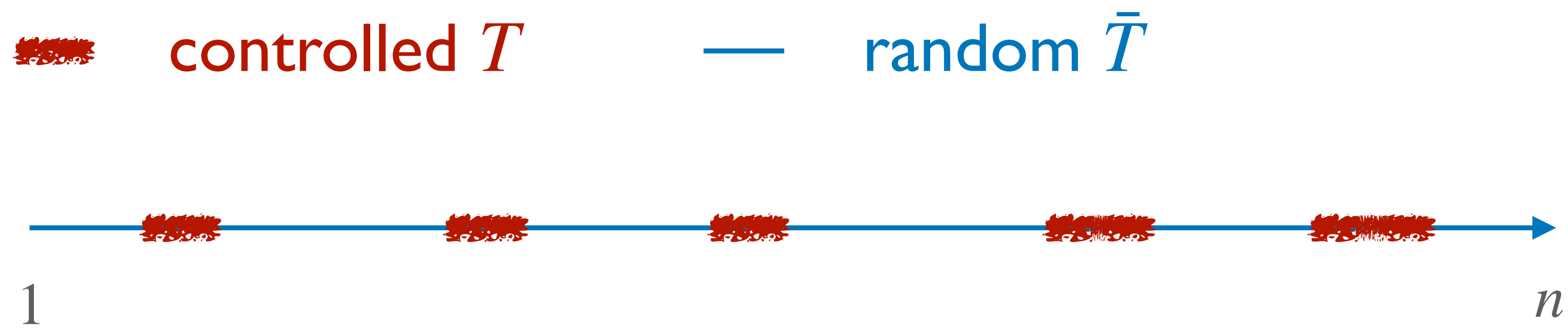
Fact. Fix some some path $x(0), x(1), \ldots, x(t) \in \{-1,0,1\}^n$,

$$\frac{\Pr[\forall i \in [t], Y(t) = x(t)]}{\Pr[\forall i \in [t], X(t) = x(t)]} = \frac{f(x(t))}{f(0)}.$$

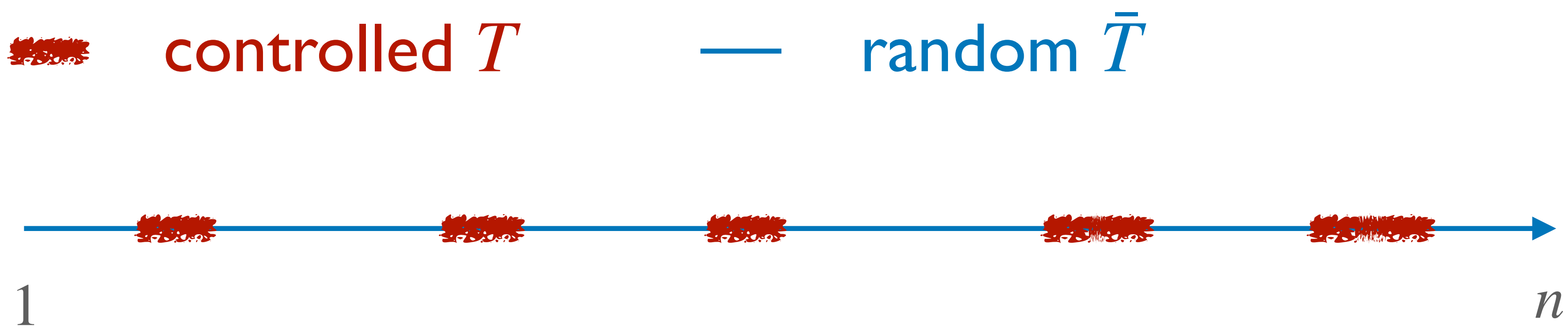
Pf. $\text{lhs} = \prod_{i=1}^{t} 2 \frac{f(x(i-1) + x_{\pi(i)} e_{\pi(i)})}{2f(x(i-1))} = \frac{f(x(t))}{f(0)}.$

Controlled Process

Controlled process



Controlled Process



Controlled process

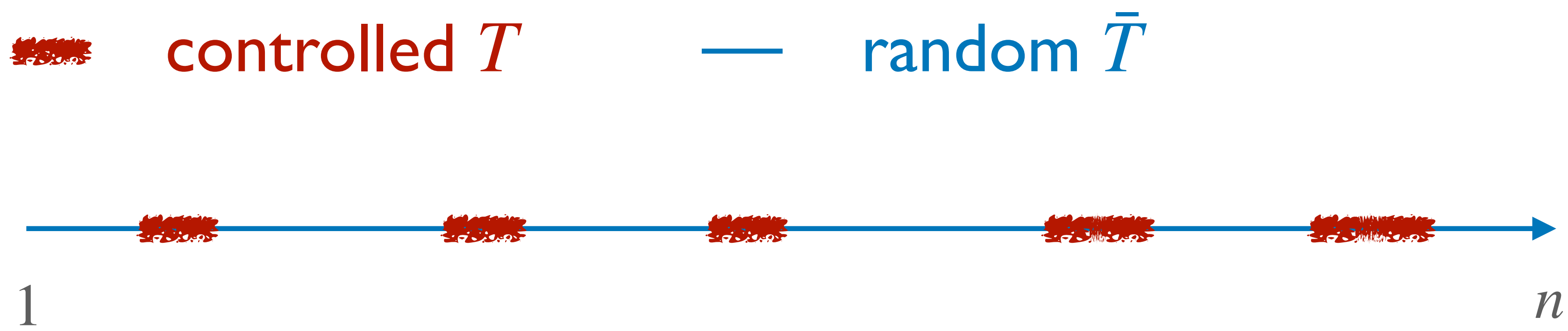
THE GAME: expose the variables in a random order:

(w.p. $1 - \epsilon$) expose the variable uniformly randomly;

(w.p. ϵ) player gets to decide.

Controlled Process

Controlled process



THE GAME: expose the variables in a random order:

(w.p. $1-\epsilon$) expose the variable uniformly randomly;

(w.p. ϵ) player gets to decide.

THE GOAL: sample $x \in f^{-1}(1)$ uniformly at random.

Controlled Process

controlled T — random $\bar{T}$

Controlled process

THE GAME: expose the variables in a random order:

(w.p. $1-\epsilon$) expose the variable uniformly randomly;

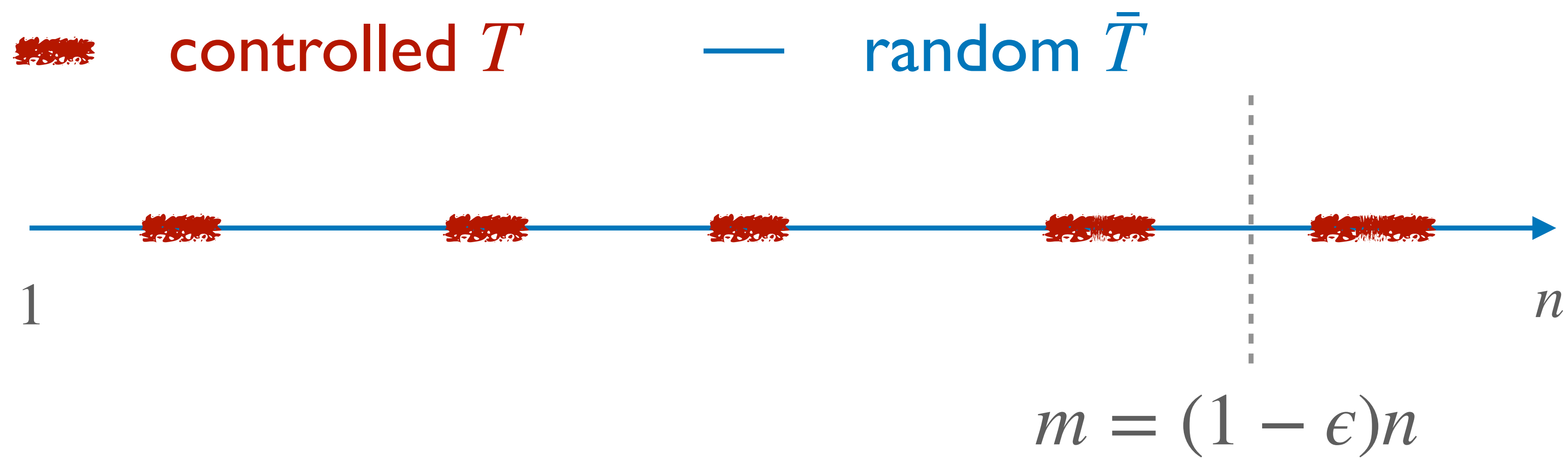
(w.p. ϵ) player gets to decide.

THE GOAL: sample $x \in f^{-1}(1)$ uniformly at random.

Mission impossible!

Controlled Process

Controlled process



THE GAME: expose the variables in a random order:

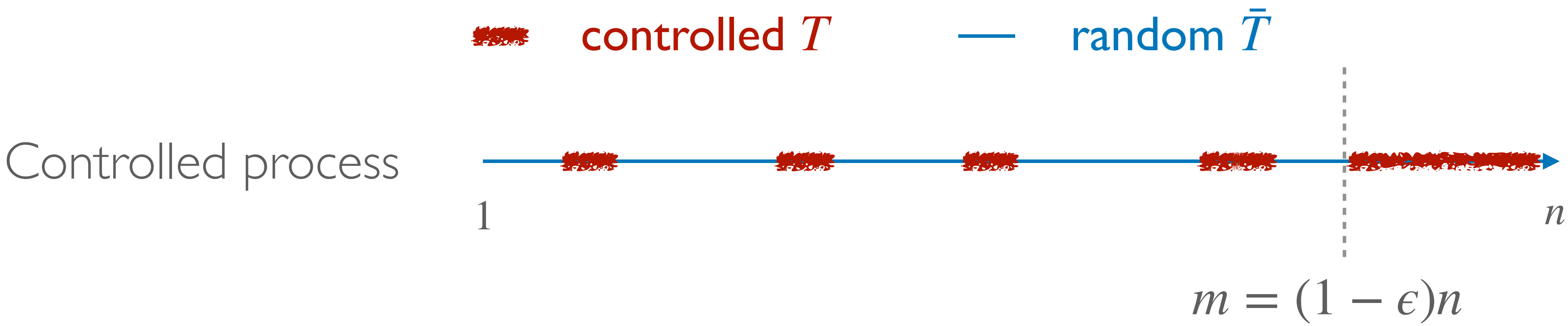
(w.p. $1-\epsilon$) expose the variable uniformly randomly;

(w.p. ϵ) player gets to decide.

THE GOAL: sample $x \in f^{-1}(1)$ uniformly at random.

Mission impossible!

Controlled Process



THE GAME: expose the variables in a random order:

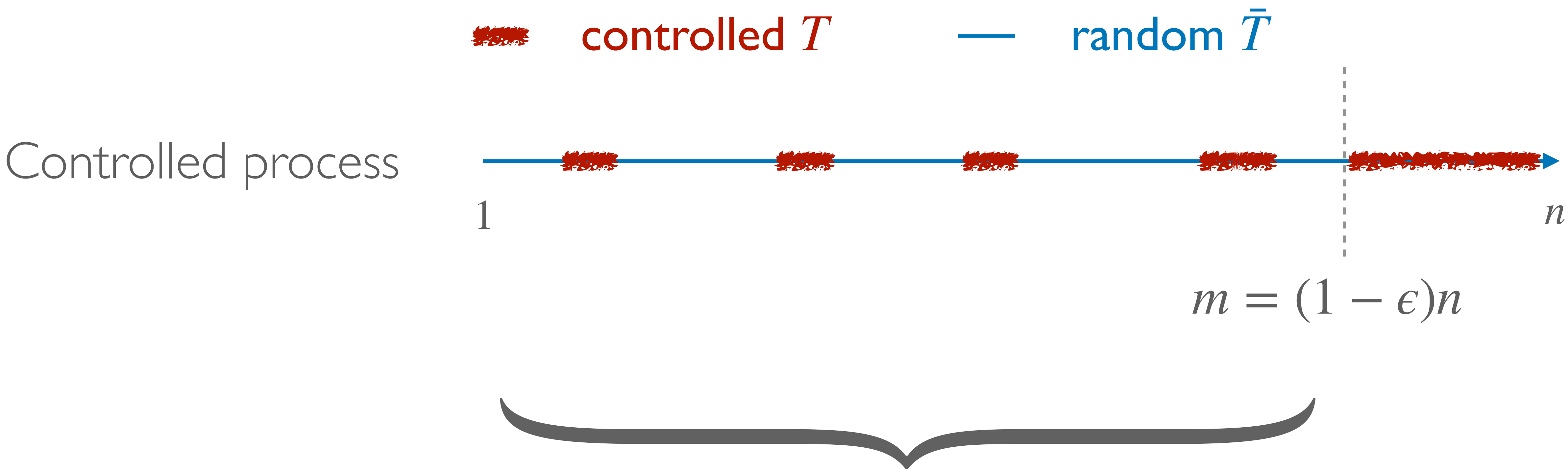
(w.p. $1 - \epsilon$) expose the variable uniformly randomly;

(w.p. ϵ) player gets to decide;

(after m) player gets to decide.

THE GOAL: sample $x \in f^{-1}(1)$ uniformly at random.

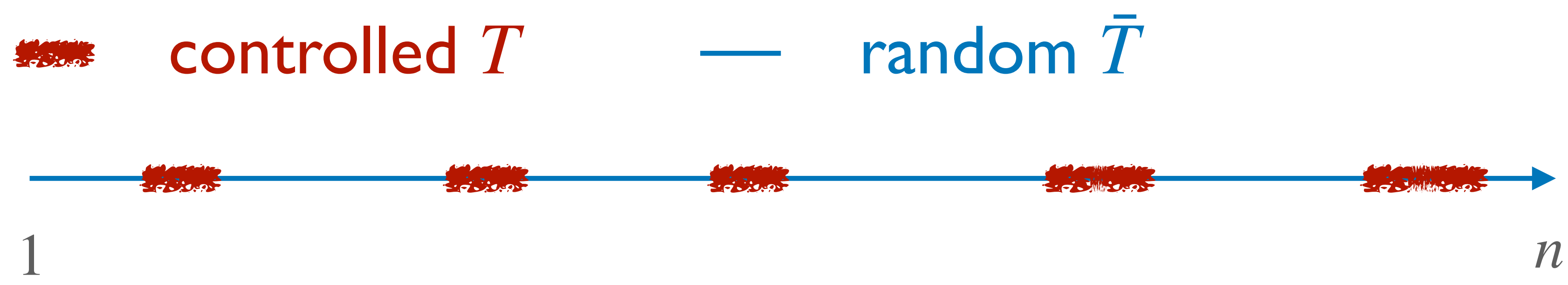
Controlled Process



the blue variables = random restriction

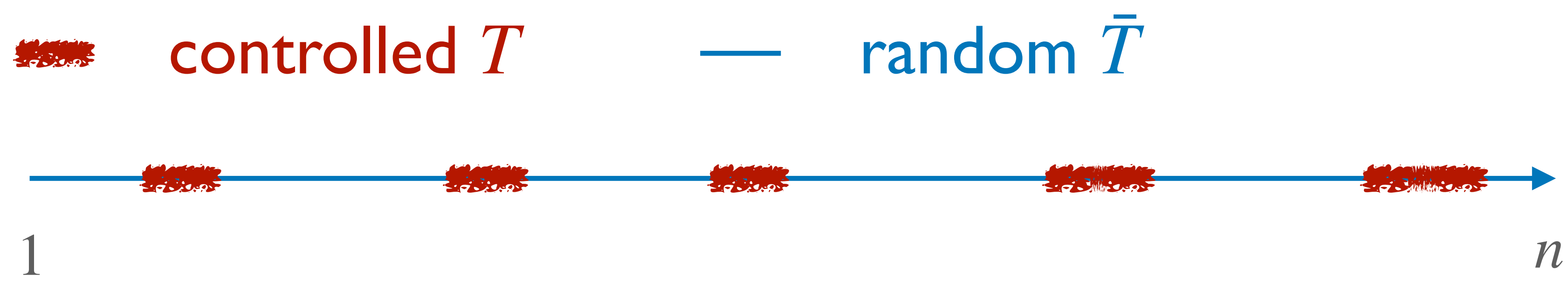
Controlled Process

Controlled process



Controlled Process

Controlled process



$T \subseteq [n]$, a random $(1 - \epsilon)$-set

(Random coordinate $t \notin T$)

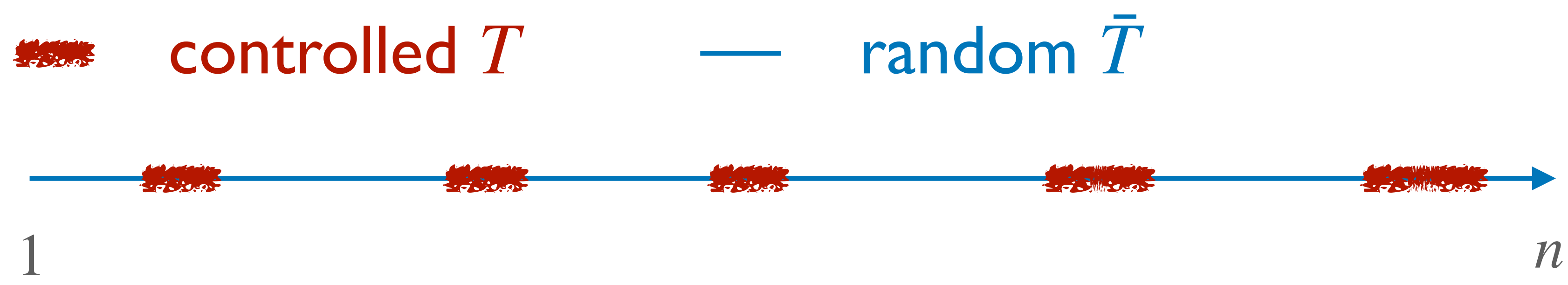
$$\Pr[Y_{\pi(t)}(t) = \pm 1] = 0.5$$

(Controlled coordinate $t \in T$)

$$\Pr[Y_{\pi(t)}(t) = \pm 1] = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2\epsilon f(Y(t-1))}.$$

Controlled Process

Controlled process



$T \subseteq [n]$, a random $(1-\epsilon)$-set

(Random coordinate $t \notin T$)

$$\Pr[Y_{\pi(t)}(t) = \pm 1] = 0.5$$

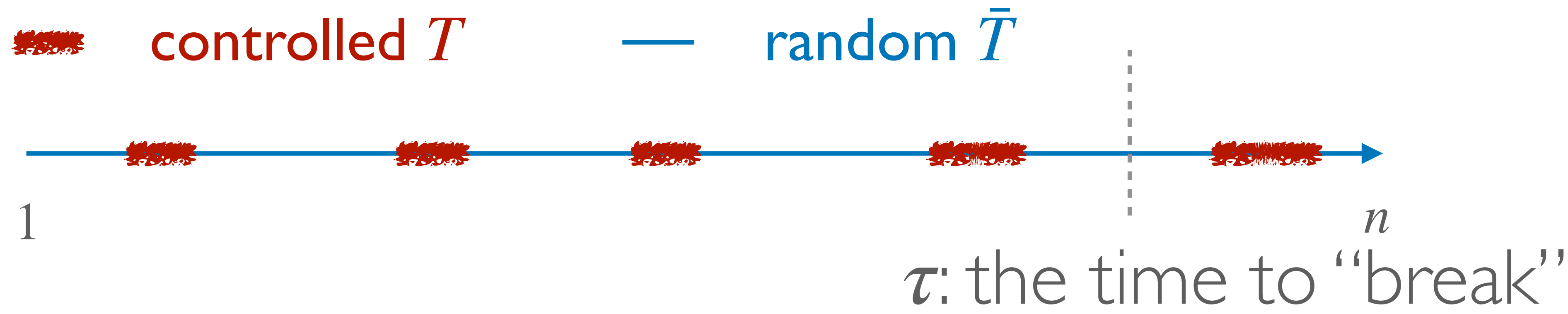
(Controlled coordinate $t \in T$)

$$\Pr[Y_{\pi(t)}(t) = \pm 1] = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2\epsilon f(Y(t-1))}.$$

Fact. As long as $\epsilon f(Y(t-1)) \geq |\partial_{\pi(t)} f(Y(t-1))|$, then the controlled process simulates the conditioned process.

Controlled Process

Controlled process



$T \subseteq [n]$, a random $(1-\epsilon)$-set

(Random coordinate $t \notin T$)

$$\Pr[Y_{\pi(t)}(t) = \pm 1] = 0.5$$

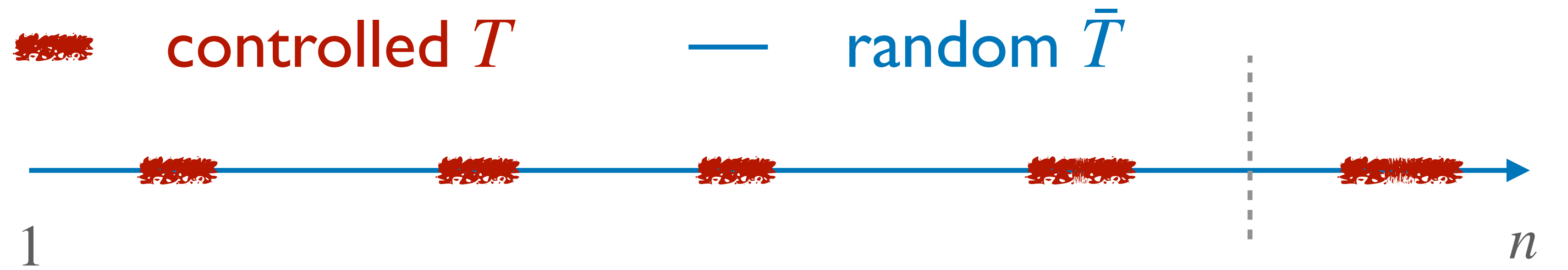
(Controlled coordinate $t \in T$)

$$\Pr[Y_{\pi(t)}(t) = \pm 1] = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2\epsilon f(Y(t-1))}.$$

Fact. As long as $\epsilon f(Y(t-1)) \geq |\partial_{\pi(t)} f(Y(t-1))|$, then the controlled process simulates the conditioned process.

Controlled Process

Controlled process



τ: the time to "break"

in particular, the first time t when
$$\max_i |\partial_i f(Y(t-1))| > \epsilon\delta,$$
$$f(Y(t-1)) < \delta$$

$T \subseteq [n]$, a random $(1-\epsilon)$-set

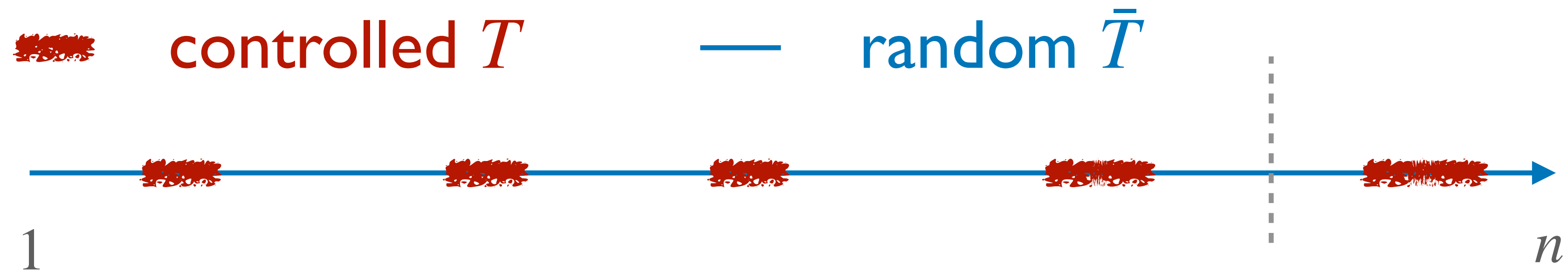
(Random coordinate $t \notin T$)
$$\Pr[Y_{\pi(t)}(t) = \pm 1] = 0.5$$

(Controlled coordinate $t \in T$)
$$\Pr[Y_{\pi(t)}(t) = \pm 1] = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2\epsilon f(Y(t-1))}.$$

Fact. As long as $\epsilon f(Y(t-1)) \geq |\partial_{\pi(t)} f(Y(t-1))|$, then the controlled process simulates the conditioned process.

Controlled Process

Controlled process



τ: the time to "break"

in particular, the first time t when

$\max_i |\partial_i f(Y(t-1))| > \epsilon\delta$,

$f(Y(t-1)) < \delta$

$T \subseteq [n]$, a random $(1-\epsilon)$-set

(Random coordinate $t \notin T$)

$$\Pr[Y_{\pi(t)}(t) = \pm 1] = 0.5$$

(Controlled coordinate $t \in T$)

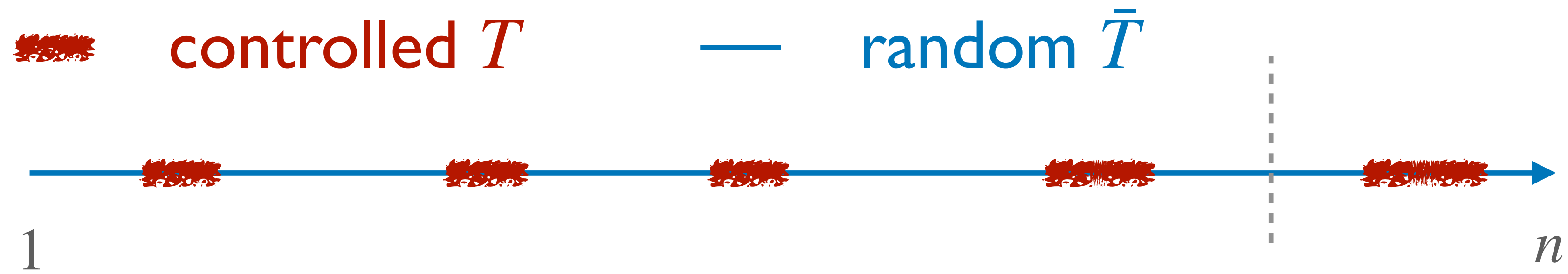
$$\Pr[Y_{\pi(t)}(t) = \pm 1] = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2\epsilon f(Y(t-1))}.$$

(After τ)

$$\Pr[Y_{\pi(t)}(t) = \pm 1] = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2 f(Y(t-1))}.$$

Controlled Process

Controlled process



τ: the time to "break"

in particular, the first time t when

$$\max_i |\partial_i f(Y(t-1))| > \epsilon\delta,$$

$$f(Y(t-1)) < \delta$$

$T \subseteq [n]$, a random $(1-\epsilon)$-set

(Random coordinate $t \notin T$)

$$\Pr[Y_{\pi(t)}(t) = \pm 1] = 0.5$$

(Controlled coordinate $t \in T$)

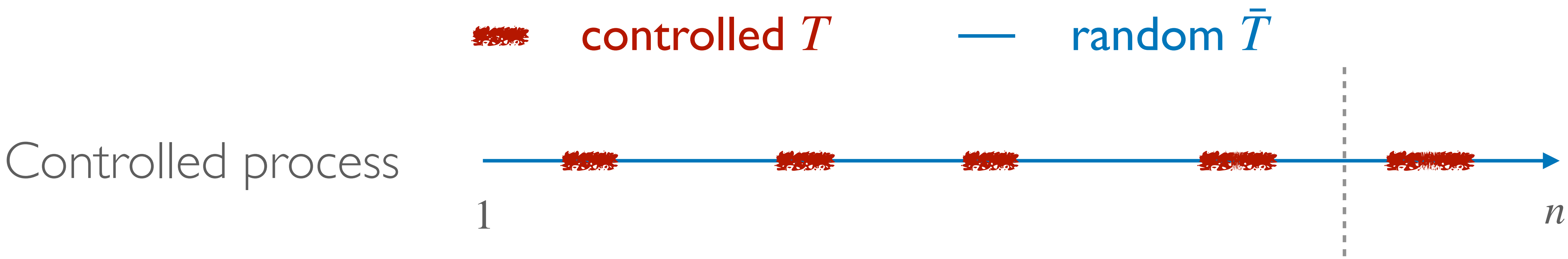
$$\Pr[Y_{\pi(t)}(t) = \pm 1] = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2\epsilon f(Y(t-1))}.$$

(After τ)

$$\Pr[Y_{\pi(t)}(t) = \pm 1] = \frac{1}{2} \pm \frac{\partial_{\pi(t)} f(Y(t-1))}{2f(Y(t-1))}.$$

Goal accomplished, as long as $\tau > m$!

Analysis



τ: the time to "break"

in particular, the first time t when

$$\max_i |\partial_i f(Y(t-1))| > \epsilon\delta,$$

$$f(Y(t-1)) < \delta$$

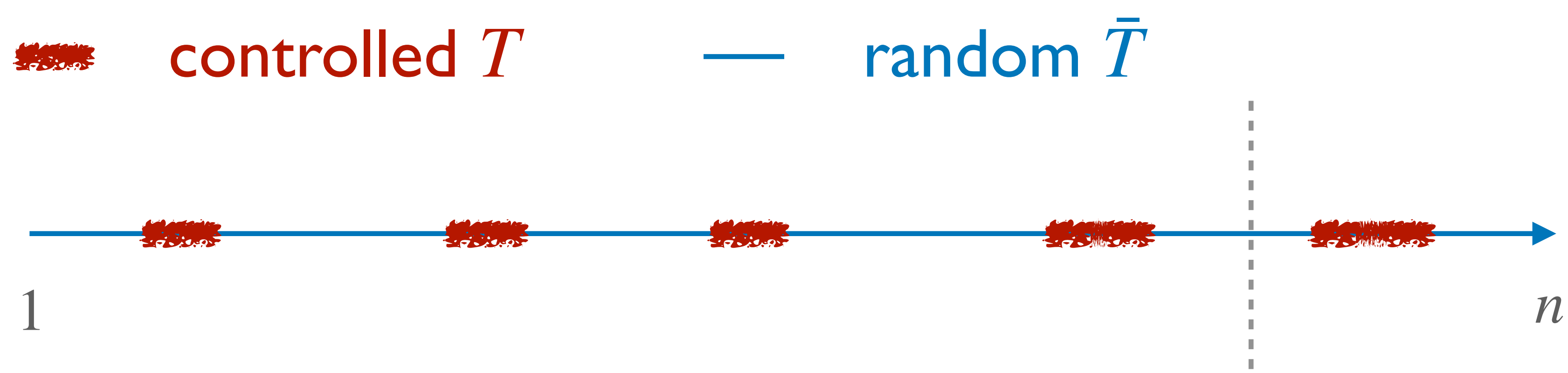
$$\tau = \min\{\tau_1, \tau_2\},$$

$$\tau_1 = \min_t\{f(Y(t-1)) < \delta\},$$

$$\tau_2 = \min_t\{\max_i |\partial_i f(Y(t-1))| > \epsilon\delta\}$$

Analysis

Controlled process



τ: the time to "break"

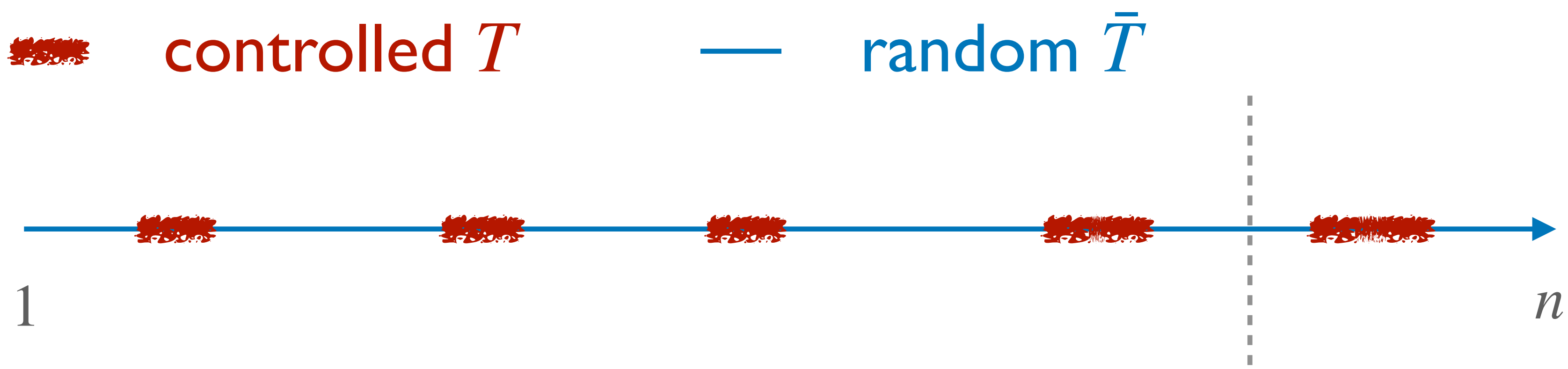
in particular, the first time t when
$$\max_i |\partial_i f(Y(t-1))| > \epsilon\delta,$$
$$f(Y(t-1)) < \delta$$

$$\tau = \min\{\tau_1, \tau_2\},$$
$$\tau_1 = \min_t\{f(Y(t-1)) < \delta\},$$
$$\tau_2 = \min_t\{\max_i |\partial_i f(Y(t-1))| > \epsilon\delta\}$$

Lemma 1. w.h.p. $\tau_1 > (1-\epsilon)n$

Analysis

Controlled process



τ: the time to "break"

in particular, the first time t when
$\max_i |\partial_i f(Y(t-1))| > \epsilon\delta$,
$f(Y(t-1)) < \delta$

$$\tau = \min\{\tau_1, \tau_2\},$$
$$\tau_1 = \min_t\{f(Y(t-1)) < \delta\},$$
$$\tau_2 = \min_t\{\max_i |\partial_i f(Y(t-1))| > \epsilon\delta\}$$

Lemma 1. w.h.p. $\tau_1 > (1-\epsilon)n$

Lemma 2. w.h.p. $\tau_2 > (1-\epsilon)n$

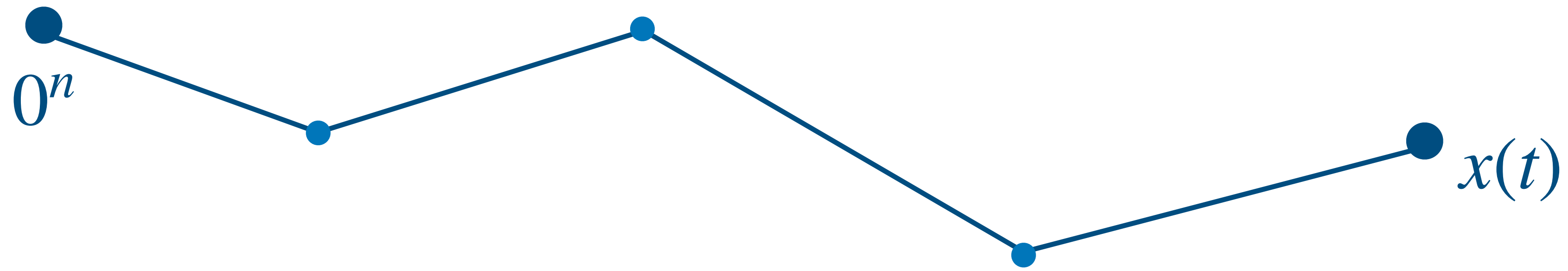
Mean remains large

$\tau_1 = \min_t \{f(Y(t-1)) < \delta\}$

Lemma 1. w.h.p. $\tau_1 > (1-\epsilon)n$

Proof. $\Pr_Y[\tau_1 \leq (1-\epsilon)n] \leq \Pr_X[\tau_1 \leq (1-\epsilon)n] \cdot \dfrac{\delta}{f(0)}$

$$\leq \frac{\delta}{f(0)}\,.$$



Partial derivatives remain small

$$\tau_2 = \min_t \{ \max_i |\partial_i f(Y(t-1))| > \epsilon\delta \}$$

Lemma 2. w.h.p. $\tau_2 > (1-\epsilon)n$

Proof. $\Pr_Y[\tau_2 \le (1-\epsilon)n] \le \Pr_X[\tau_2 \le (1-\epsilon)n] \cdot \dfrac{1}{f(0)}$.

Partial derivatives remain small

$$\tau_2 = \min_t \{\max_i |\partial_i f(Y(t-1))| > \epsilon\delta\}$$

Lemma 2. w.h.p. $\tau_2 > (1-\epsilon)n$

Proof. $\Pr_Y[\tau_2 \leq (1-\epsilon)n] \leq \Pr_X[\tau_2 \leq (1-\epsilon)n] \cdot \dfrac{1}{f(0)}$.

Switch to continuous process

Continuous uniform process

Sample $\eta(i) \in [0,1], i = 1,2,\ldots,n$,
Each $Z_i(t)$ is 0 until $t = \eta(i)$, set $Z_i(t) \sim \{-1,1\}$.

Switch to continuous process

Continuous uniform process

Sample $\eta(i) \in [0,1], i = 1,2,\ldots,n$,
Each $Z_i(t)$ is 0 until $t = \eta(i)$, set $Z_i(t) \sim \{-1,1\}$.

$$\tau_3 = \min_t\{\max_i |\partial_i f(Z(t))| > \epsilon\beta\}$$

Lemma 3. w.h.p. $\tau_3 > (1-\epsilon)$

Partial derivatives under uniform process

$$\tau_3 = \min_t \{ \max_i |\partial_i f(Z(t))| > \epsilon\delta \}$$

Lemma 3. w.h.p. $\tau_3 > (1 - \epsilon)$

Partial derivatives under uniform process

$$\tau_3 = \min_t \{ \max_i |\partial_i f(Z(t))| > \epsilon\delta \}$$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1 - \epsilon)$

Partial derivatives under uniform process

$$\tau_3 = \min_t \{ \max_i | \partial_i f(Z(t)) | > \epsilon\delta \}$$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1 - \epsilon)$

Proof.

$$\Pr\left[\sup_{0 \le s \le t} \beta(s) \ge \theta\right] \le \Pr\left[\sup_{0 \le s \le t} \sum_{i=1}^{n} | \partial_i f(Z(s)) |^{2+\eta} \ge \theta^{2+\eta}\right]$$

Partial derivatives under uniform process

$$\tau_3 = \min_t \{ \max_i |\partial_i f(Z(t))| > \epsilon\delta \}$$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1-\epsilon)$

Proof.

$$\Pr\left[\sup_{0\le s\le t} \beta(s) \ge \theta\right] \le \Pr\left[\sup_{0\le s\le t} \sum_{i=1}^{n} |\partial_i f(Z(s))|^{2+\eta} \ge \theta^{2+\eta}\right]$$

$$\le \theta^{-2-\eta} \sum_i \mathbf{E}[\,|\partial_i f(Z(t))|^{2+\eta}]$$

Partial derivatives under uniform process

$\tau_3 = \min_t\{\max_i |\partial_i f(Z(t))| > \epsilon\delta\}$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1-\epsilon)$

Proof.

$$\Pr\left[\sup_{0\le s\le t} \beta(s) \ge \theta\right] \le \Pr\left[\sup_{0\le s\le t} \sum_{i=1}^{n} |\partial_i f(Z(s))|^{2+\eta} \ge \theta^{2+\eta}\right]$$

$$\le \theta^{-2-\eta} \sum_i \mathbf{E}[|\partial_i f(Z(t))|^{2+\eta}]$$

$\partial_i f(Z(t))$ is a martingale.

Partial derivatives under uniform process

$$\tau_3 = \min_t \{ \max_i |\partial_i f(Z(t))| > \epsilon\delta \}$$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1 - \epsilon)$

Proof.

$$\Pr\left[\sup_{0\le s\le t} \beta(s) \ge \theta\right] \le \Pr\left[\sup_{0\le s\le t} \sum_{i=1}^{n} |\partial_i f(Z(s))|^{2+\eta} \ge \theta^{2+\eta}\right]$$

$$\le \theta^{-2-\eta} \sum_i \mathbf{E}[|\partial_i f(Z(t))|^{2+\eta}]$$

$$\le \theta^{-2-\eta} \sum_i (\mathbf{E}[\partial_i f(Z(T))^2])^{1+\eta/2}$$

A hypercontractivity inequality for random restrictions

Partial derivatives under uniform process

Theorem (Hypercontractive inequality for random restriction).

For any multilinear function $f : [-1,1]^n \to \mathbb{R}$, and $0 \le t \le T \le 1$. Then, for $\eta \le T - t$,

$$\mathbf{E}[\,|f(Z(t))|^{2+\eta}\,]^{\frac{1}{2+\eta}} \le \mathbf{E}[f(Z(T))^2]^{\frac{1}{2}},$$

Partial derivatives under uniform process

Theorem (Hypercontractive inequality for random restriction).

For any multilinear function $f : [-1,1]^n \to \mathbb{R}$, and $0 \le t \le T \le 1$. Then, for $\eta \le T - t$,

$$\mathbf{E}[\,|f(Z(t))|^{2+\eta}\,]^{\frac{1}{2+\eta}} \le \mathbf{E}[f(Z(T))^2]^{\frac{1}{2}},$$

c.f. the standard HC inequality

$$\mathbf{E}[(\mathbf{T}_{\epsilon(\eta)} f(x)^{2+\eta}])^{\frac{1}{2+\eta}} \le \mathbf{E}[f(x)^2]^{\frac{1}{2}}$$

Partial derivatives under uniform process

$$\tau_3 = \min_t \{ \max_i |\partial_i f(Z(t))| > \epsilon\delta \}$$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1-\epsilon)$

Proof.

$$\Pr\left[\sup_{0\le s\le t} \beta(s) \ge \theta\right] \le \Pr\left[\sup_{0\le s\le t} \sum_{i=1}^{n} |\partial_i f(Z(s))|^{2+\eta} \ge \theta^{2+\eta}\right]$$

$$\le \theta^{-2-\eta} \sum_i \mathbf{E}[\,|\partial_i f(Z(t))|^{2+\eta}]$$

$$\le \theta^{-2-\eta} \sum_i (\mathbf{E}[\partial_i f(Z(T))^2])^{1+\eta/2}$$

$$\le \theta^{-2-\eta} \sum_i \mathrm{Inf}_i(f)^{\eta/2} \mathbf{E}[\partial_i f(Z(T))^2]$$

Partial derivatives under uniform process

$$\tau_3 = \min_t \{ \max_i |\partial_i f(Z(t))| > \epsilon\delta \}$$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1 - \epsilon)$

Proof.

$$\Pr\left[\sup_{0\le s\le t} \beta(s) \ge \theta\right] \le \Pr\left[\sup_{0\le s\le t} \sum_{i=1}^{n} |\partial_i f(Z(s))|^{2+\eta} \ge \theta^{2+\eta}\right]$$

$$\le \theta^{-2-\eta} \sum_i \mathbf{E}[\,|\partial_i f(Z(t))|^{2+\eta}]$$

$$\le \theta^{-2-\eta} \sum_i (\mathbf{E}[\partial_i f(Z(T))^2])^{1+\eta/2}$$

$$\le \theta^{-2-\eta} \sum_i \mathrm{Inf}_i(f)^{\eta/2} \mathbf{E}[\partial_i f(Z(T))^2]$$

Proposition. $\mathbf{E}[\partial_i f(Z(t))^2] \le \mathrm{Inf}_i(f)$

Partial derivatives under uniform process

$$\tau_3 = \min_t \{ \max_i |\partial_i f(Z(t))| > \epsilon\delta \}$$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1-\epsilon)$

Proof.

$$\Pr\left[\sup_{0\le s\le t} \beta(s) \ge \theta\right] \le \Pr\left[\sup_{0\le s\le t} \sum_{i=1}^{n} |\partial_i f(Z(s))|^{2+\epsilon} \ge \theta^{2+\eta}\right]$$

$$\le \theta^{-2-\eta} \sum_i \mathbf{E}[|\partial_i f(Z(t))|^{2+\eta}]$$

$$\le \theta^{-2-\eta} \sum_i (\mathbf{E}[\partial_i f(Z(T))^2])^{1+\eta/2}$$

$$\le \theta^{-2-\eta} \sum_i \mathrm{Inf}_i(f)^{\eta/2} \mathbf{E}[\partial_i f(Z(T))^2]$$

$$\le \theta^{-2-\eta} \frac{\mathrm{MaxInf}(f)^{\eta/2}}{1-T}$$

Partial derivatives under uniform process

$$\tau_3 = \min_t \{ \max_i |\partial_i f(Z(t))| > \epsilon\delta \}$$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1-\epsilon)$

Proof.

$$\Pr\left[\sup_{0\le s\le t} \beta(s) \ge \theta\right] \le \Pr\left[\sup_{0\le s\le t} \sum_{i=1}^{n} |\partial_i f(Z(s))|^{2+\epsilon} \ge \theta^{2+\eta}\right]$$

$$\le \theta^{-2-\eta} \sum_i \mathbf{E}[|\partial_i f(Z(t))|^{2+\eta}]$$

$$\le \theta^{-2-\eta} \sum_i (\mathbf{E}[\partial_i f(Z(T))^2])^{1+\eta/2}$$

$$\le \theta^{-2-\eta} \sum_i \mathrm{Inf}_i(f)^{\eta/2} \mathbf{E}[\partial_i f(Z(T))^2]$$

$$\le \theta^{-2-\eta} \frac{\mathrm{MaxInf}(f)^{\eta/2}}{1-T}$$

Proposition. $\displaystyle\sum_{i=1}^{n} \mathbf{E}[\partial_i f(Z(T))^2] \le \frac{1}{1-T}$

Partial derivatives under uniform process

$$\tau_3 = \min_t \{\max_i |\partial_i f(Z(t))| > \epsilon\delta\}$$

$\beta(t)$

Lemma 3. w.h.p. $\tau_3 > (1-\epsilon)$

Proof.

$$\Pr\left[\sup_{0\le s\le t} \beta(s) \ge \theta\right] \le \Pr\left[\sup_{0\le s\le t} \sum_{i=1}^{n} |\partial_i f(Z(s))|^{2+\epsilon} \ge \theta^{2+\eta}\right]$$

$$\le \theta^{-2-\eta} \sum_i \mathbf{E}[|\partial_i f(Z(t))|^{2+\eta}]$$

$$\le \theta^{-2-\eta} \sum_i (\mathbf{E}[\partial_i f(Z(T))^2])^{1+\eta/2}$$

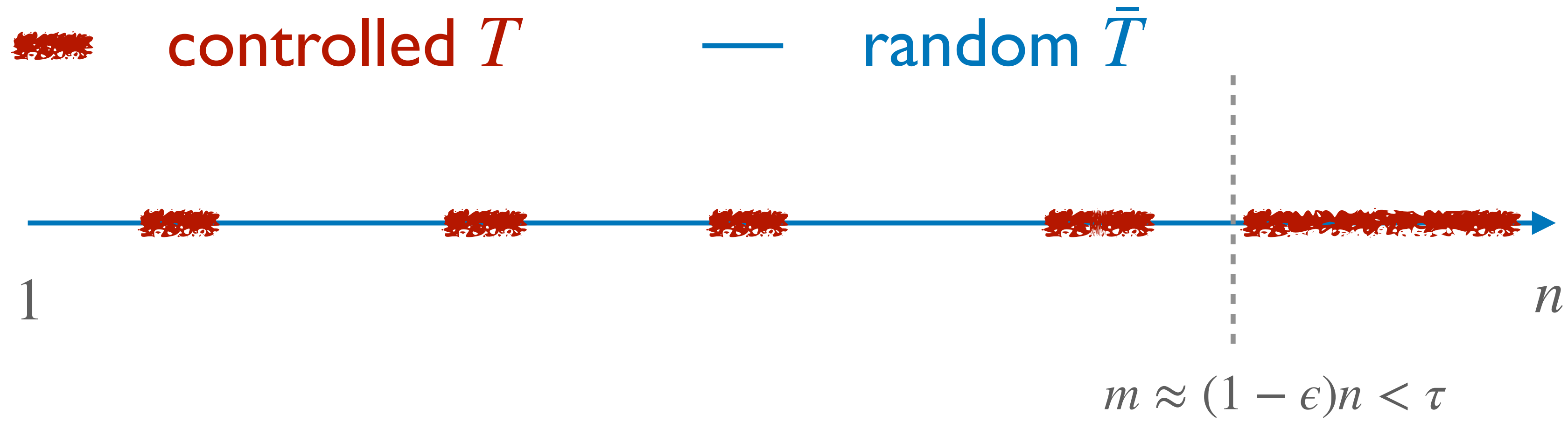
$$\le \theta^{-2-\eta} \sum_i \mathrm{Inf}_i(f)^{\eta/2} \mathbf{E}[\partial_i f(Z(T))^2]$$

$$\le \theta^{-2-\eta} \frac{\mathrm{MaxInf}(f)^{\eta/2}}{1-T} \le 2\theta^{-3+t/2} \frac{\mathrm{MaxInf}(f)^{(1-t)/4}}{1-t}$$

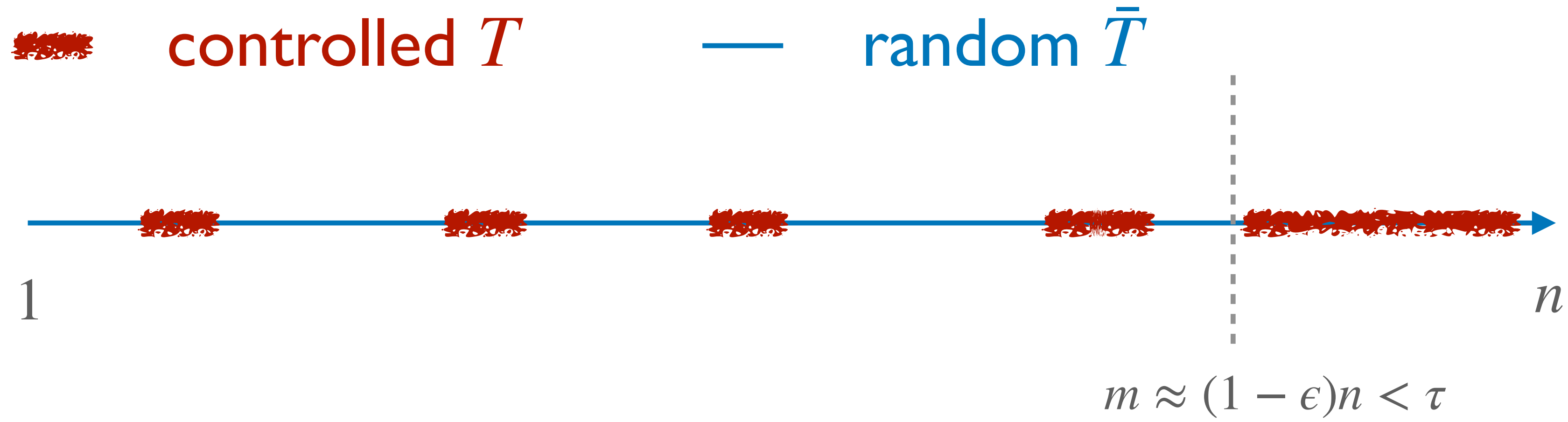
set $T = (1+t)/2, \eta = T - t$

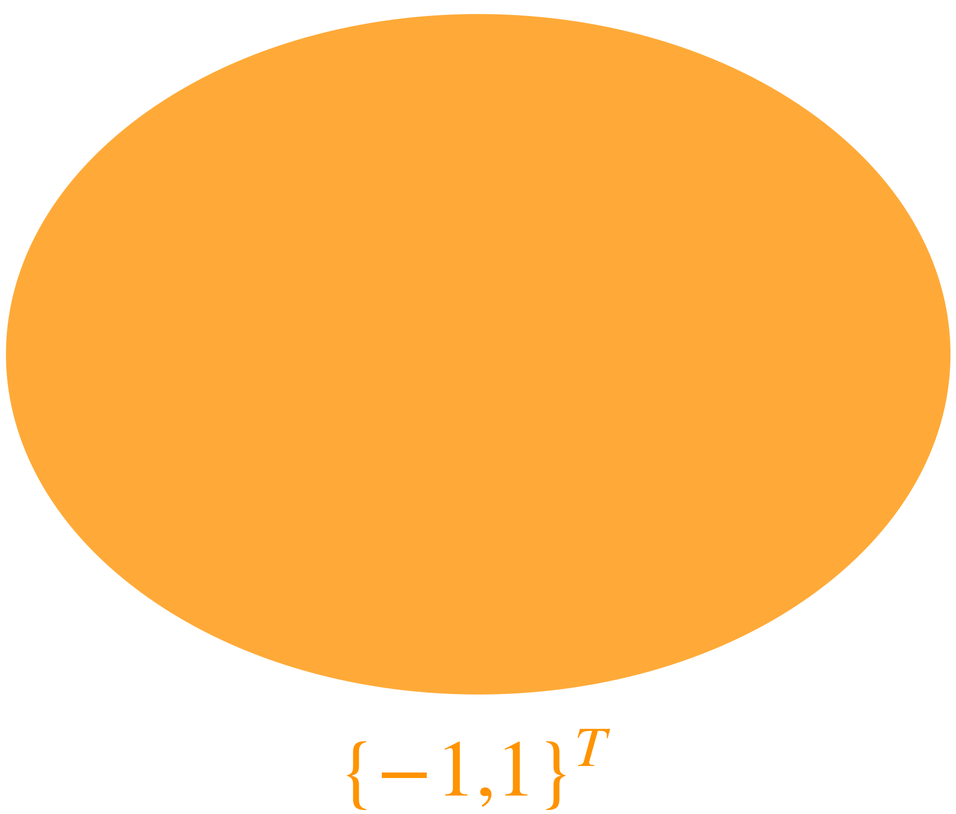
Bound the variance

Entropy comparison

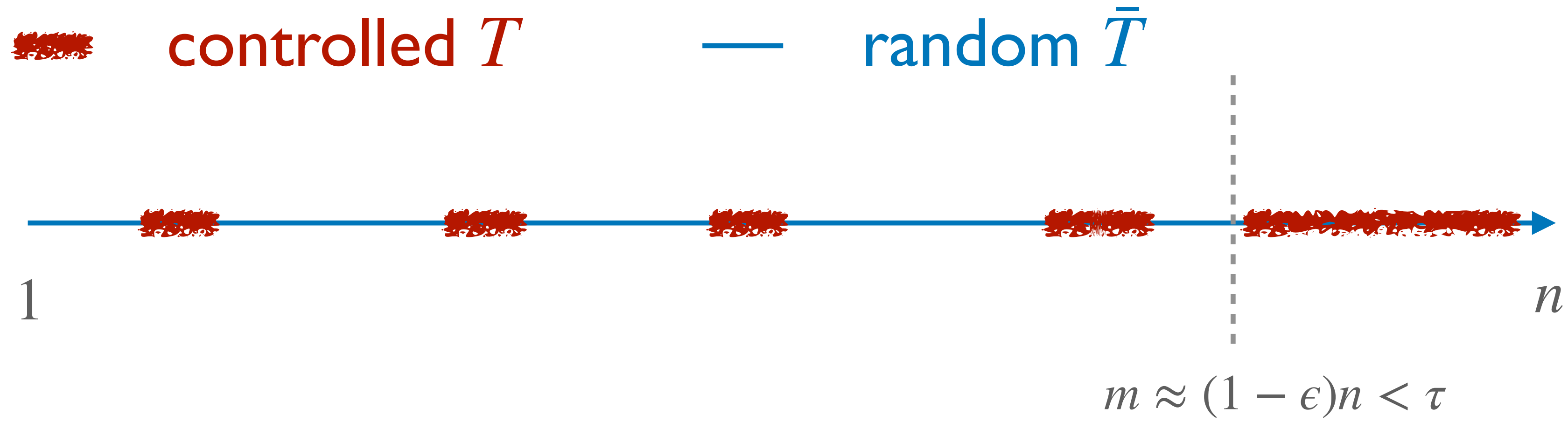


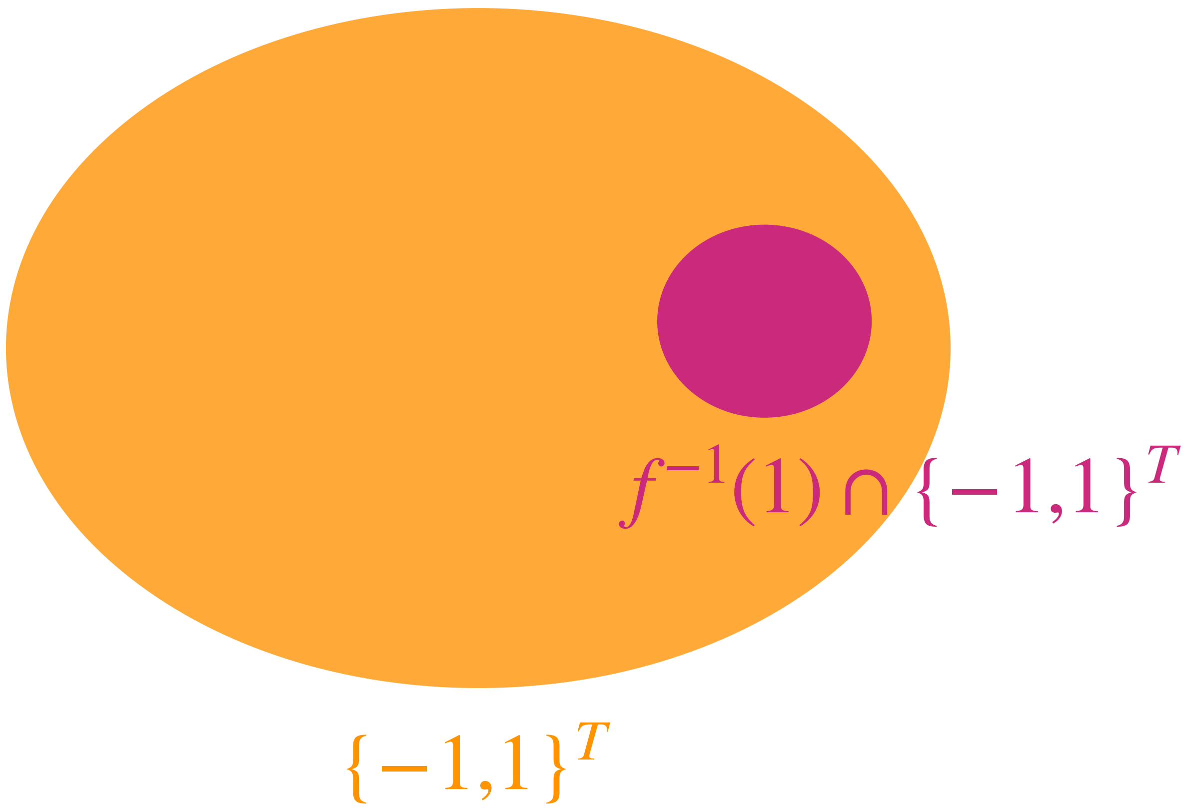
Entropy comparison



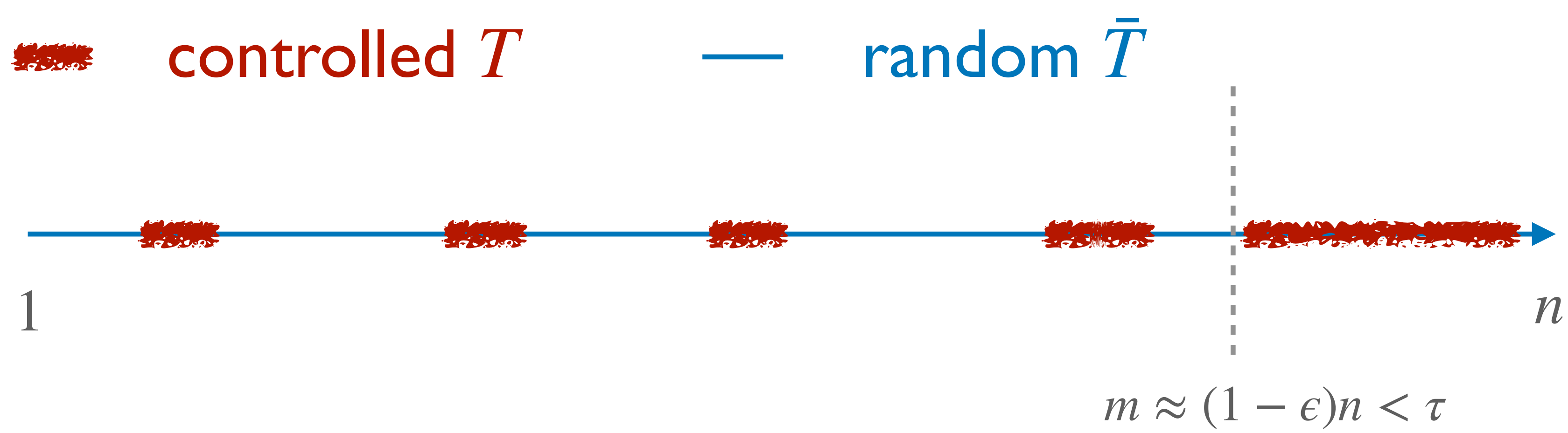


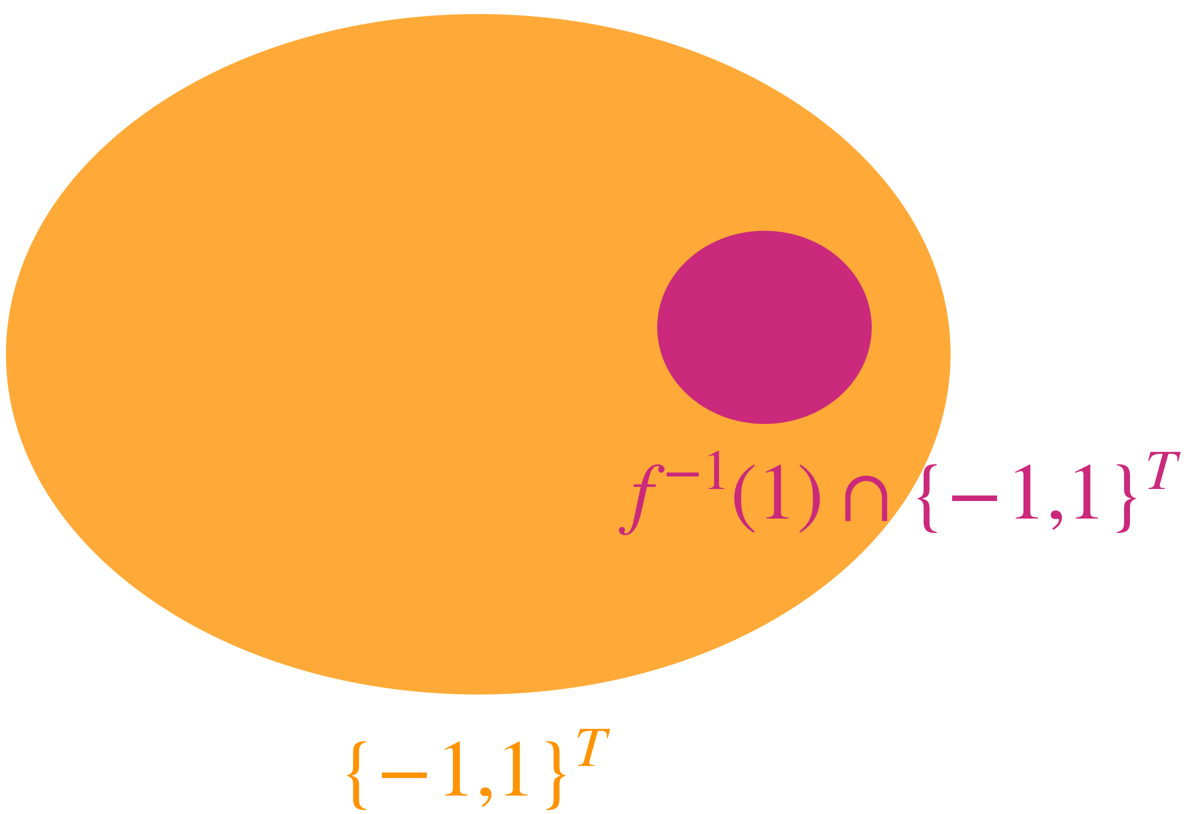
Entropy comparison





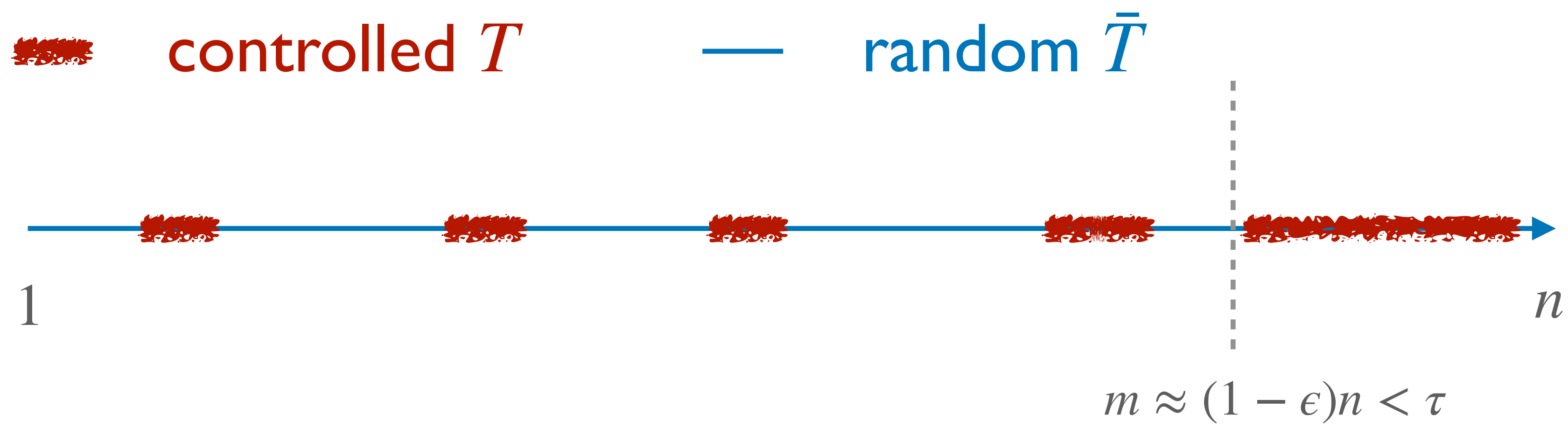
Entropy comparison

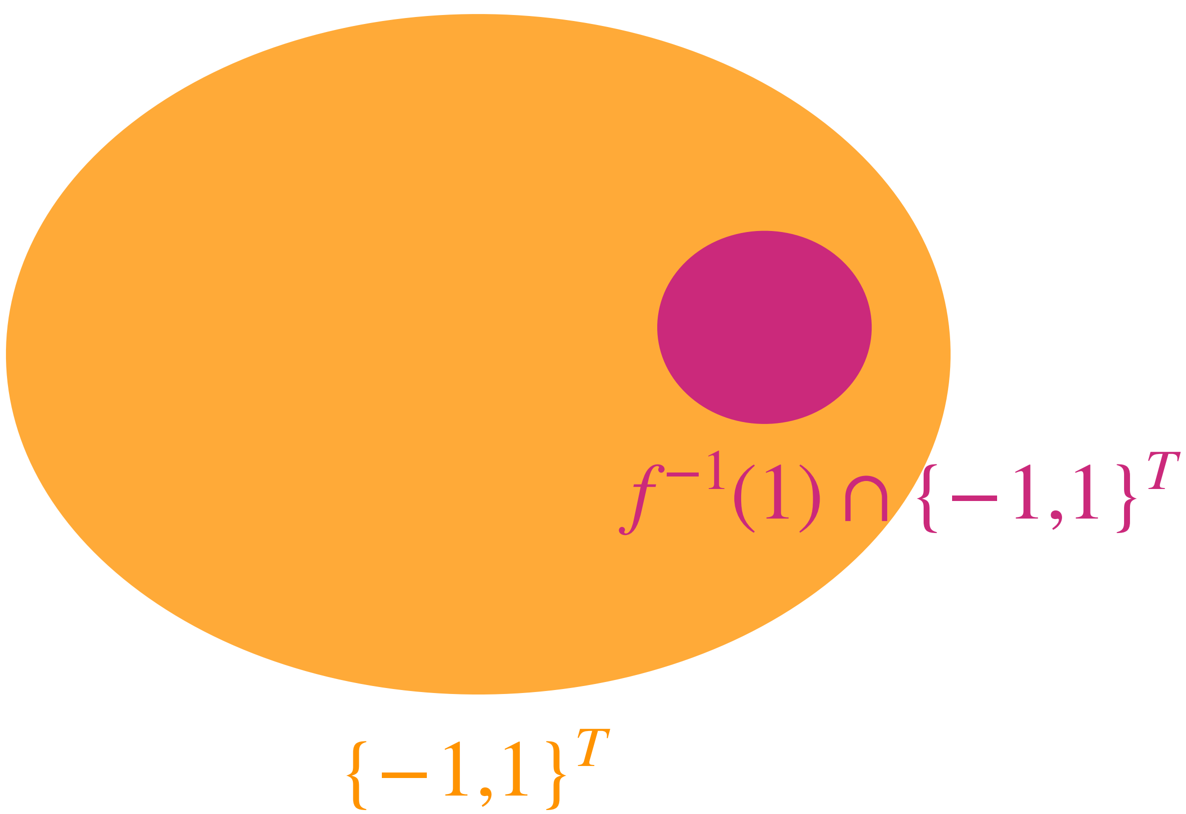




Strategy: to bound the KL-divergence between $Y(n)$ and $X(n)$ given π, T, and the restriction.

Entropy comparison

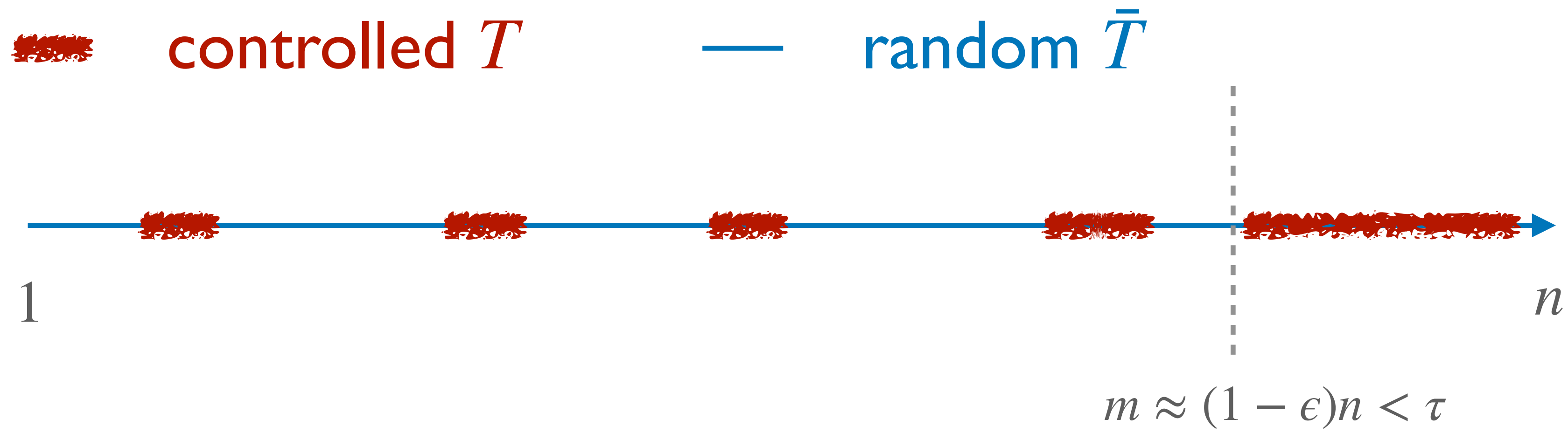




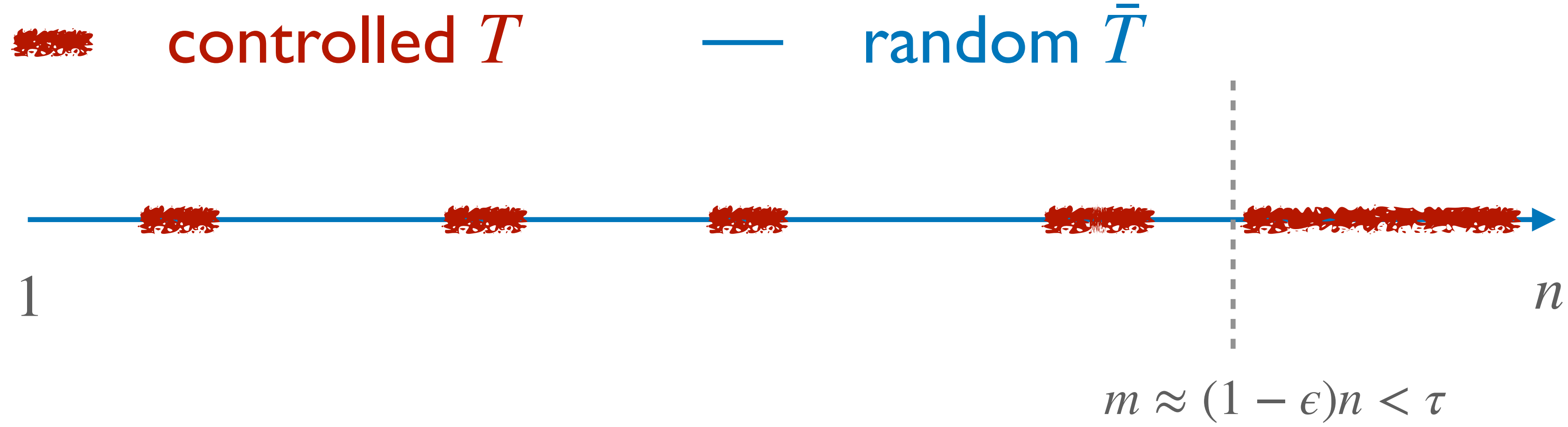
Strategy: to bound the KL-divergence between $Y(n)$ and $X(n)$ given π, T, and the restriction.

$$|T| - \log |f^{-1}(1) \cap \{-1,1\}^T|$$

KL-divergence

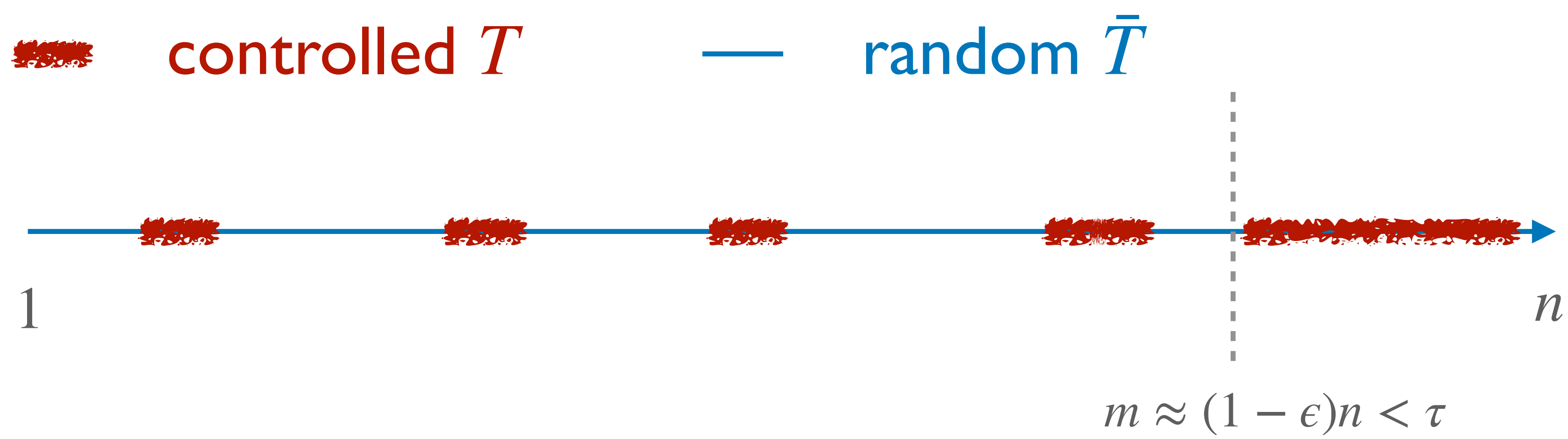


KL-divergence



$\mathscr{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

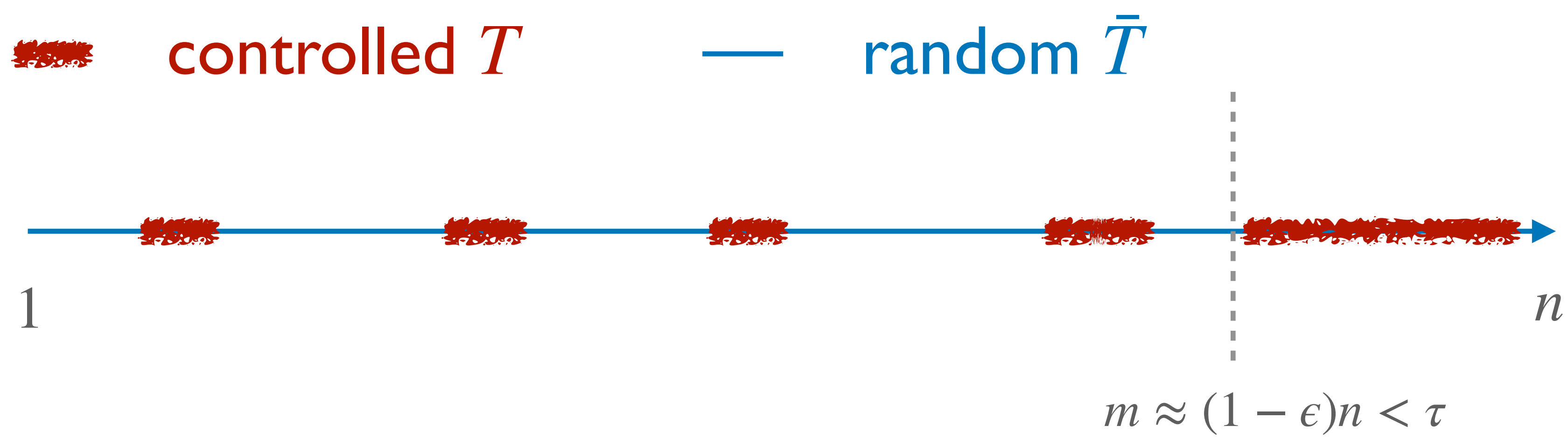
KL-divergence



$\mathcal{G}_m = (\pi, T, x(m)|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

$\mathbf{E}_{\mathcal{G}_m}[\mathrm{KL}(Y(n) \mid \mathcal{G}_m \,\|\, X(n) \mid \mathcal{G}_m)]$

KL-divergence



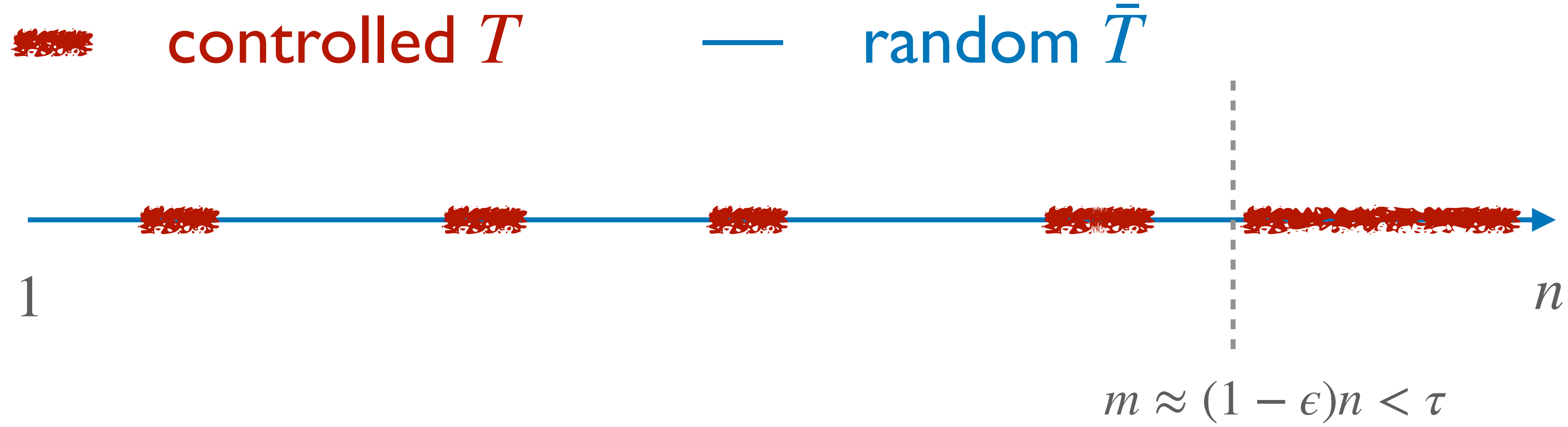
$\mathcal{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

$$\mathbf{E}_{\mathcal{G}_m}[\mathrm{KL}(Y(n) \mid \mathcal{G}_m \,\|\, X(n) \mid \mathcal{G}_m)]$$

$$\leq \sum_{t=1}^{m} \mathbf{E}[\mathbb{1}_{\{t \in T\}} \mathrm{KL}((Y_{\pi(t)}(t) \mid \mathcal{G}_m, Y(t-1)) \,\|\, (X_{\pi(t)}(t) \mid \mathcal{G}_m, Y(t-1)))]$$

$$+\mathbf{E}[\mathrm{KL}((Y(n) \mid \mathcal{G}_m, Y(m)) \,\|\, (X(n) \mid \mathcal{G}_m, Y(m)))]$$

KL-divergence



$\mathcal{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

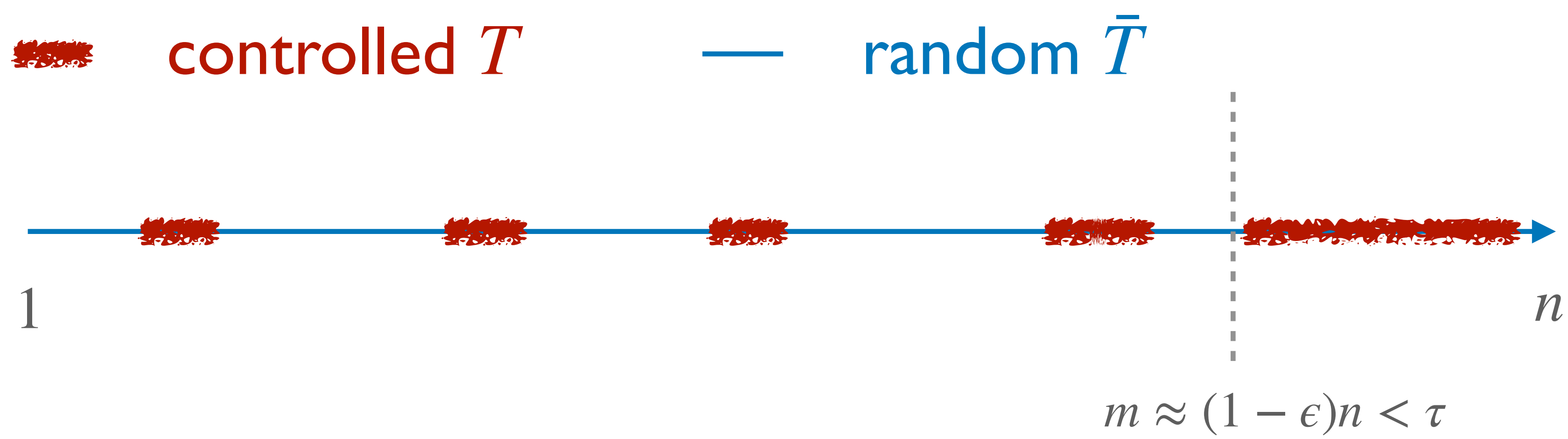
$$\mathbf{E}_{\mathcal{G}_m}[\mathrm{KL}(Y(n) \mid \mathcal{G}_m \,\|\, X(n) \mid \mathcal{G}_m)]$$

$$\leq \sum_{t=1}^{m} \mathbf{E}[\mathbb{1}_{\{t \in T\}} \mathrm{KL}((Y_{\pi(t)}(t) \mid \mathcal{G}_m, Y(t-1)) \,\|\, (X_{\pi(t)}(t) \mid \mathcal{G}_m, Y(t-1)))]$$

$$+\mathbf{E}[\mathrm{KL}((Y(n) \mid \mathcal{G}_m, Y(m)) \,\|\, (X(n) \mid \mathcal{G}_m, Y(m)))]$$

$$\leq \log \frac{1}{\delta}$$

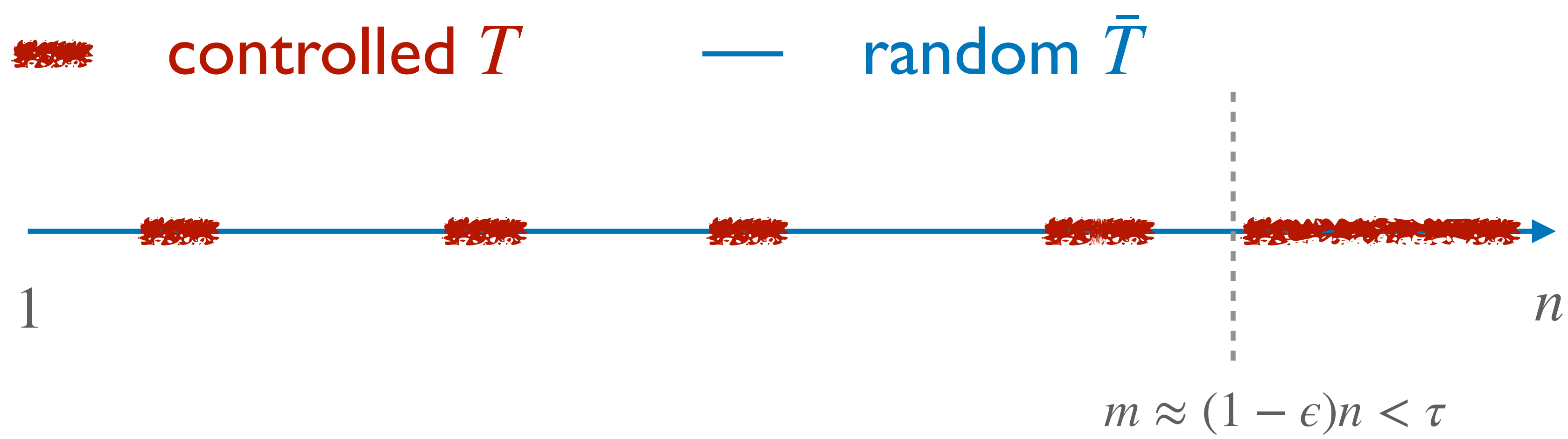
KL-divergence



$\mathcal{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

$$\mathbf{E}_{\mathcal{G}_m}[\mathrm{KL}(Y(n) \mid \mathcal{G}_m \,\|\, X(n) \mid \mathcal{G}_m)]$$
$$\leq \sum_{t=1}^{m} \mathbf{E}[\mathbb{1}_{\{t \in T\}} \mathrm{KL}((Y_{\pi(t)}(t) \mid \mathcal{G}_m, Y(t-1)) \,\|\, (X_{\pi(t)}(t) \mid \mathcal{G}_m, Y(t-1)))] + \log \frac{1}{\delta}$$

KL-divergence



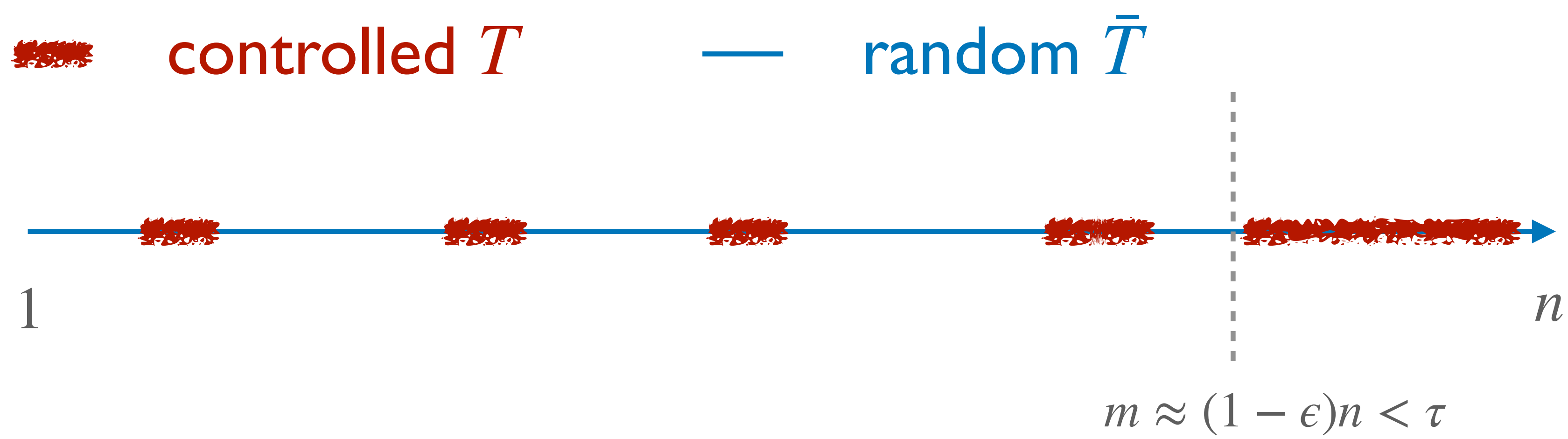
$\mathcal{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

$$\mathbf{E}_{\mathcal{G}_m}[\mathrm{KL}(Y(n) \mid \mathcal{G}_m \,\|\, X(n) \mid \mathcal{G}_m)]$$

$$\leq \sum_{t=1}^{m} \mathbf{E}[\mathbb{I}_{\{t \in T\}} \mathrm{KL}((Y_{\pi(t)}(t) \mid \mathcal{G}_m, Y(t-1)) \,\|\, (X_{\pi(t)}(t) \mid \mathcal{G}_m, Y(t-1)))] + \log \frac{1}{\delta}$$

$$\leq \sum_{t=1}^{m} \frac{\epsilon}{n-t+1} \sum_{i: Y_i(t-1)=0} \left(\frac{\partial_i f(Y(t-1))}{2\epsilon f(Y(t-1))} \right)^2 + \log \frac{1}{\delta}$$

KL-divergence

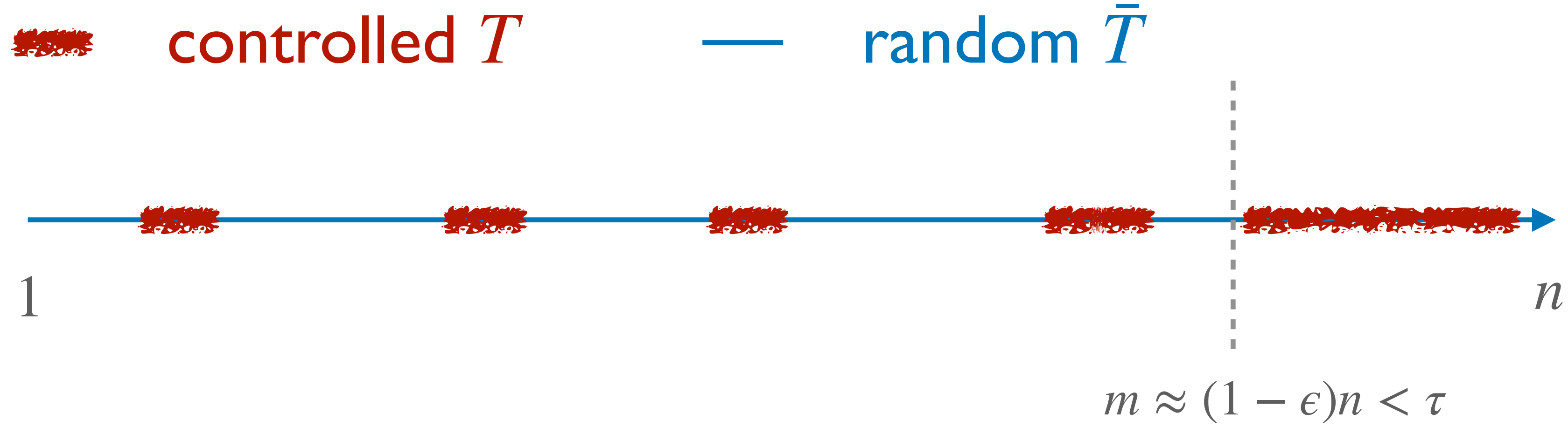


$\mathcal{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

$$\mathbf{E}_{\mathcal{G}_m}[\mathrm{KL}(Y(n) \mid \mathcal{G}_m \,\|\, X(n) \mid \mathcal{G}_m)]$$

$$\leq \sum_{t=1}^{m} \frac{\epsilon}{n-t+1} \sum_{i:Y_i(t-1)=0} \left(\frac{\partial_i f(Y(t-1))}{2\epsilon f(Y(t-1))} \right)^2 + \log \frac{1}{\delta}$$

KL-divergence



$\mathcal{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

$$\mathbf{E}_{\mathcal{G}_m}[\mathrm{KL}(Y(n) \mid \mathcal{G}_m \,\|\, X(n) \mid \mathcal{G}_m)]$$

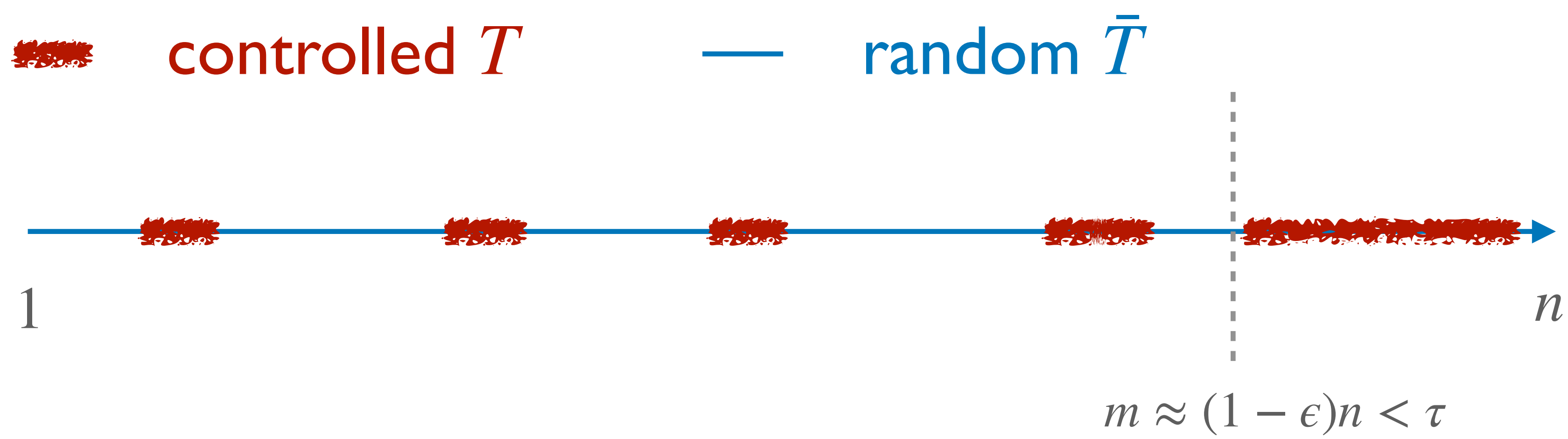
$$\leq \sum_{t=1}^{m} \frac{\epsilon}{n-t+1} \sum_{i:Y_i(t-1)=0} \left(\frac{\partial_i f(Y(t-1))}{2\epsilon f(Y(t-1))} \right)^2 + \log \frac{1}{\delta}$$

Theorem ([Talagrand 96])

For any $g : \{-1,1\}^n \to [0,1]$, we have

$$\sum \partial_i g(0)^2 \leq C g(0)^2 \log \frac{e}{g(0)}.$$

KL-divergence



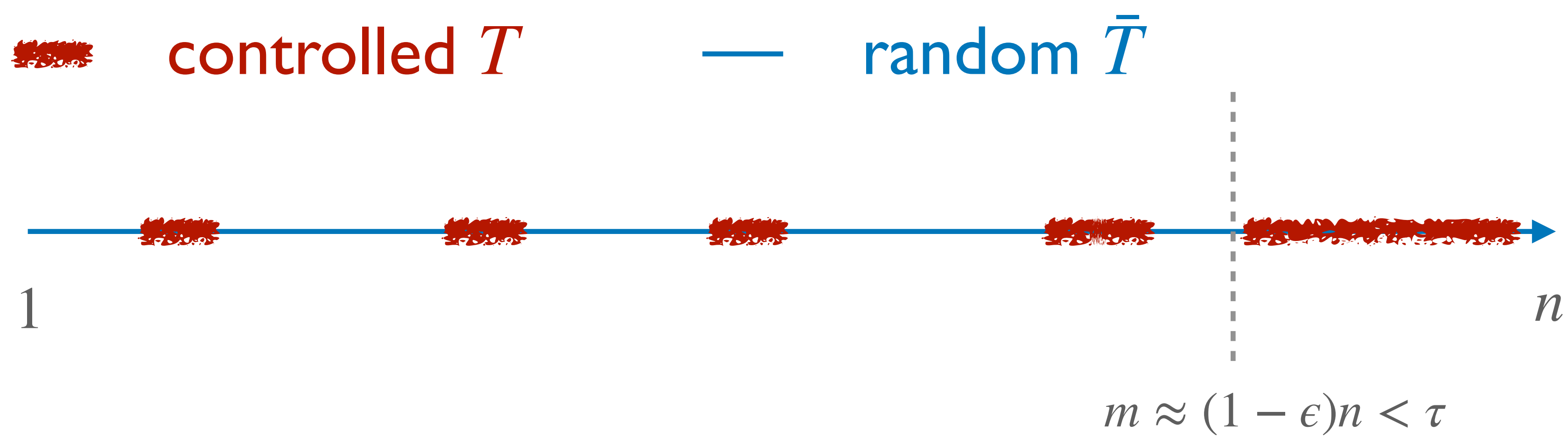
$\mathcal{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

$$\mathbf{E}_{\mathcal{G}_m}[\mathrm{KL}(Y(n) \mid \mathcal{G}_m \,\|\, X(n) \mid \mathcal{G}_m)]$$

$$\leq \sum_{t=1}^{m} \frac{\epsilon}{n-t+1} \sum_{i:Y_i(t-1)=0} \left(\frac{\partial_i f(Y(t-1))}{2\epsilon f(Y(t-1))} \right)^2 + \log \frac{1}{\delta}$$

$$\leq \sum_{t=1}^{m} \frac{C}{4\epsilon(n-t+1)} \log \frac{e}{f(Y(t-1))} + \log \frac{1}{\delta}$$

KL-divergence

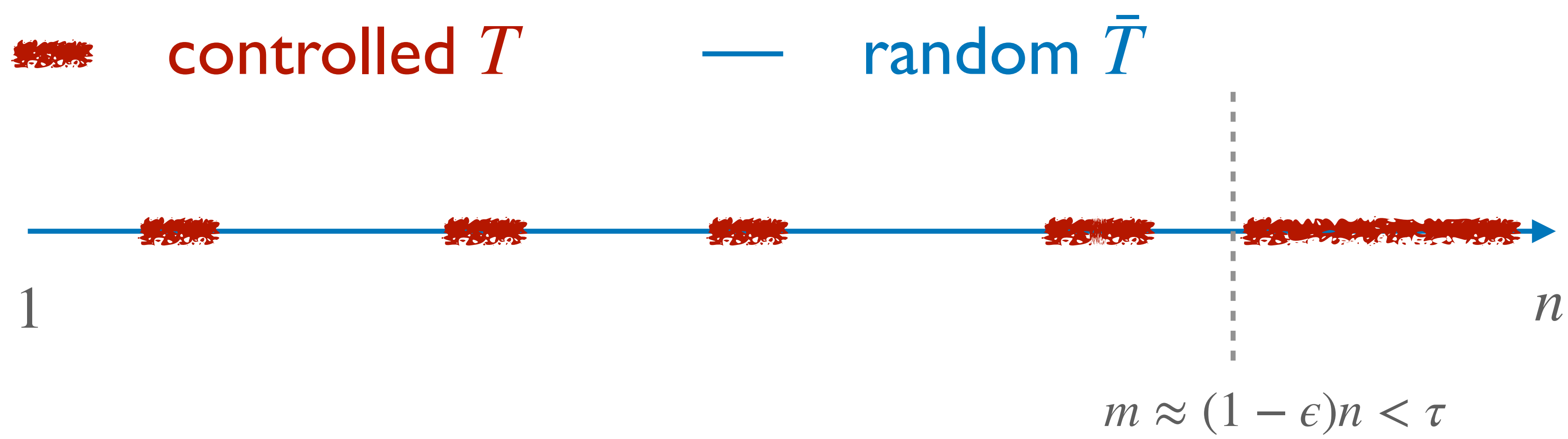


$\mathscr{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

$$\mathbf{E}_{\mathscr{G}_m}[\mathrm{KL}(Y(n) \mid \mathscr{G}_m \,\|\, X(n) \mid \mathscr{G}_m)]$$

$$\leq \sum_{t=1}^{m} \frac{C}{4\epsilon(n-t+1)} \log \frac{e}{f(Y(t-1))} + \log \frac{1}{\delta}$$

KL-divergence



$\mathscr{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

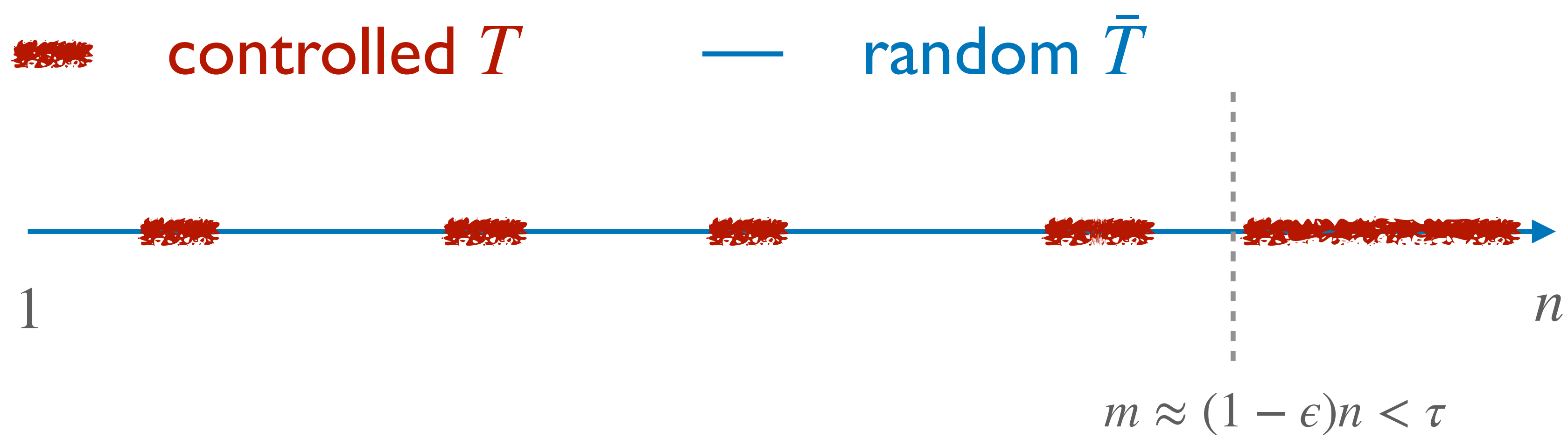
$$\mathbf{E}_{\mathscr{G}_m}[\mathrm{KL}(Y(n) \mid \mathscr{G}_m \,\|\, X(n) \mid \mathscr{G}_m)]$$

$$\leq \sum_{t=1}^{m} \frac{C}{4\epsilon(n-t+1)} \log \frac{e}{f(Y(t-1))} + \log \frac{1}{\delta}$$

$$\leq O\left(\frac{1}{\epsilon} \log \frac{1}{\delta}\right) \cdot \sum_{t=1}^{m} \frac{1}{n-t+1}$$

$$= O\left(\frac{1}{\epsilon} \log \frac{1}{\delta} \log \frac{n-m}{n}\right).$$

KL-divergence



$\mathcal{G}_m = (\pi, T, x(m)\,|_{\bar{T}})$: all the information we know about coordinates come before time m on $\bar{T}$.

$$\mathbf{E}_{\mathcal{G}_m}[\mathrm{KL}(Y(n) \mid \mathcal{G}_m \,\|\, X(n) \mid \mathcal{G}_m)]$$

$$\leq \sum_{t=1}^{m} \frac{C}{4\epsilon(n-t+1)} \log \frac{e}{f(Y(t-1))} + \log \frac{1}{\delta}$$

$$\leq O\left(\frac{1}{\epsilon} \log \frac{1}{\delta}\right) \cdot \sum_{t=1}^{m} \frac{1}{n-t+1}$$

$$= O\left(\frac{1}{\epsilon} \log \frac{1}{\delta} \log \frac{n-m}{n}\right).$$

Thank you!