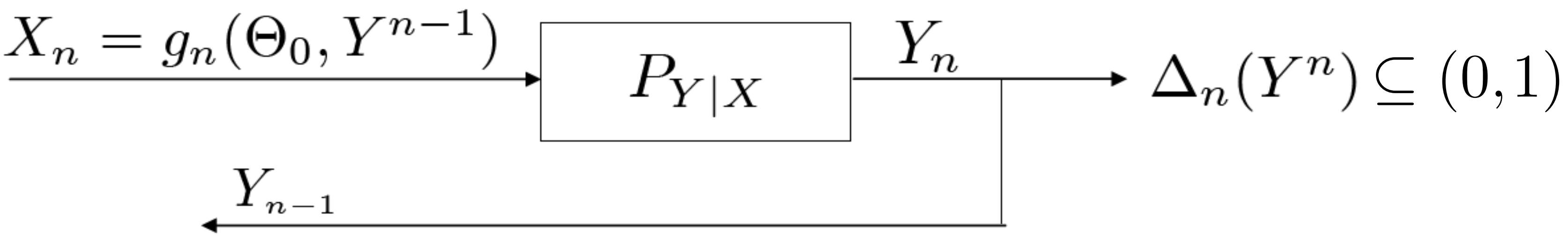
Posterior Matching Variants and Fixed-Point Elimination

Ofer Shayevitz, UCSD

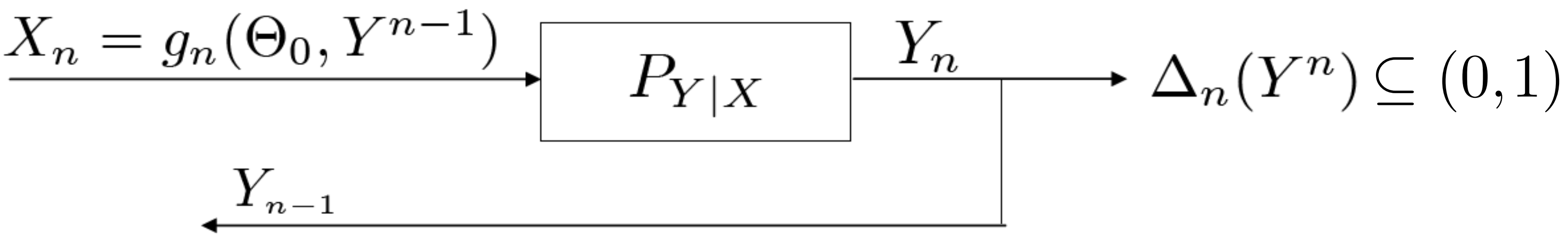
Allerton 2009

Feedback Communication Setting



- Message point $\Theta_0 \sim \mathrm{Unif}(0,1)$

Feedback Communication Setting



- Message point $\Theta_0 \sim \mathrm{Unif}(0,1)$
- Transmission scheme *achieves* a rate R if

$$P_{\Theta_0|Y^n}(\Delta_n|Y^n) \to 1\,, \qquad |\Delta_n(Y^n))| = O\left(2^{-nR}\right)$$

- Possibly under input constraints $\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n}\eta(X_k) \le \Gamma$ *a.s.*
- Implies standard achievability

The Posterior Matching (PM) Scheme *[Shayevitz & Feder '07,'08]*

- For any input/channel pair $(P_X, P_{Y|X})$

$$X_{n+1} = F_X^{-1} \circ F_{\Theta_0|Y^n}(\Theta_0|Y^n)$$

- Simple and sequential, achieves $R < I(X;Y)$ under general conditions, within input constraints encapsulated in P_X

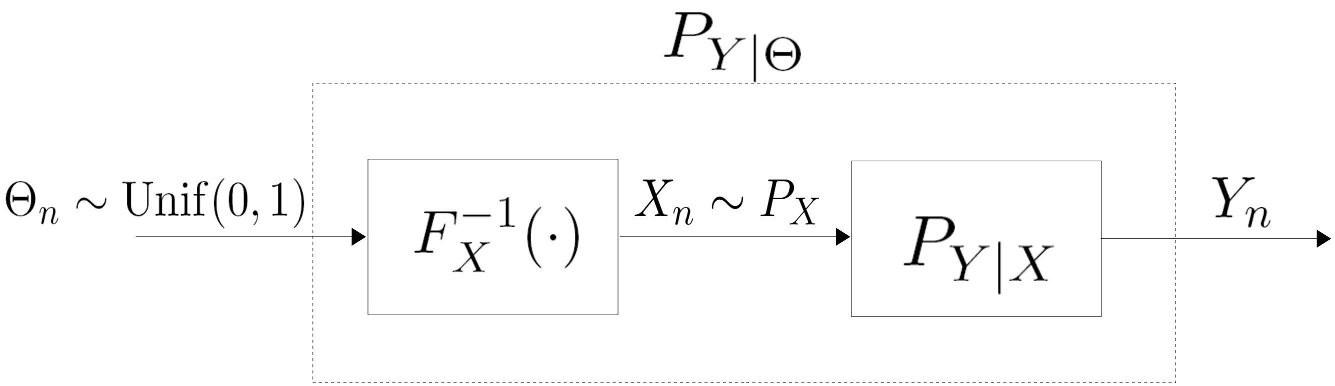
The Posterior Matching (PM) Scheme *[Shayevitz & Feder '07,'08]*

- For any input/channel pair $(P_X, P_{Y|X})$

$$X_{n+1} = F_X^{-1} \circ F_{\Theta_0|Y^n}(\Theta_0|Y^n)$$

- Simple and sequential, achieves $R < I(X;Y)$ under general conditions, within input constraints encapsulated in P_X
- Specifically, holds for (minus issues immediately discussed)
  - DMC
  - P_{XY} with a sufficiently smooth p.d.f.
- Horstein & Schalkwijk-Kailath are special cases
- Stochastic control formulation *[Coleman '09]*

Normalized Channel and Recursive Representation



- PM scheme recursive representation *[Shayevitz & Feder '08]*

$$\Theta_1 = \Theta_0\,, \quad \Theta_{n+1} = F_{\Theta|Y}(\Theta_n|Y_n)$$

- $F_{\Theta|Y}(\cdot|\cdot)$ is called the *PM kernel*
- Posterior c.d.f. evolves by function composition with PM kernel

$$F_{\Theta_0|Y^n}(\cdot|Y^n) = F_{\Theta|Y}(\cdot|Y_n) \circ F_{\Theta_0|Y^{n-1}}(\cdot|Y^{n-1})$$

A Necessary Condition for Achievability

- The PM kernel has a fixed-point at θ_f if

$$F_{\Theta|Y}(\theta_f|y) = \theta_f\,, \quad \forall\, y \in \mathcal{Y}$$

- Claim: If the PM kernel has a fixed point, then the PM scheme is not ergodic and cannot achieve any positive rate

A Necessary Condition for Achievability

- The PM kernel has a fixed-point at θ_f if

$$F_{\Theta|Y}(\theta_f|y) = \theta_f \,, \quad \forall\, y \in \mathcal{Y}$$

- Claim: If the PM kernel has a fixed point, then the PM scheme is not ergodic and cannot achieve any positive rate

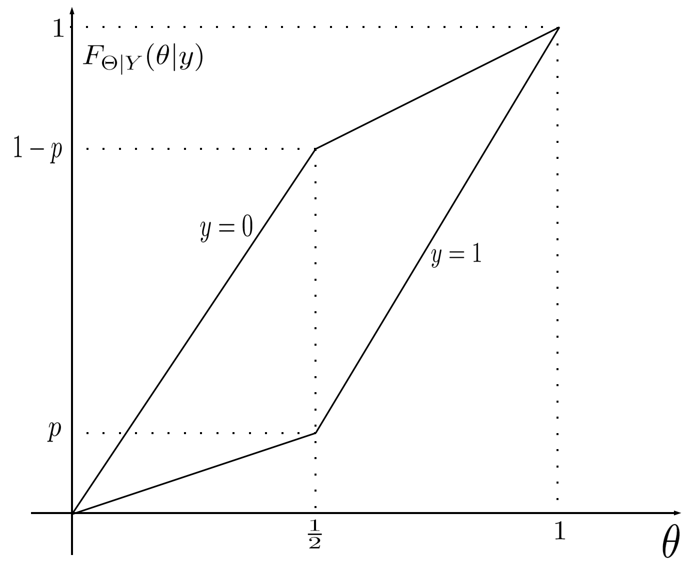
- For a DMC, maximal number of fixed-points is $|\mathcal{X}| - 2$
- No fixed point for a binary input alphabet

A Necessary Condition for Achievability

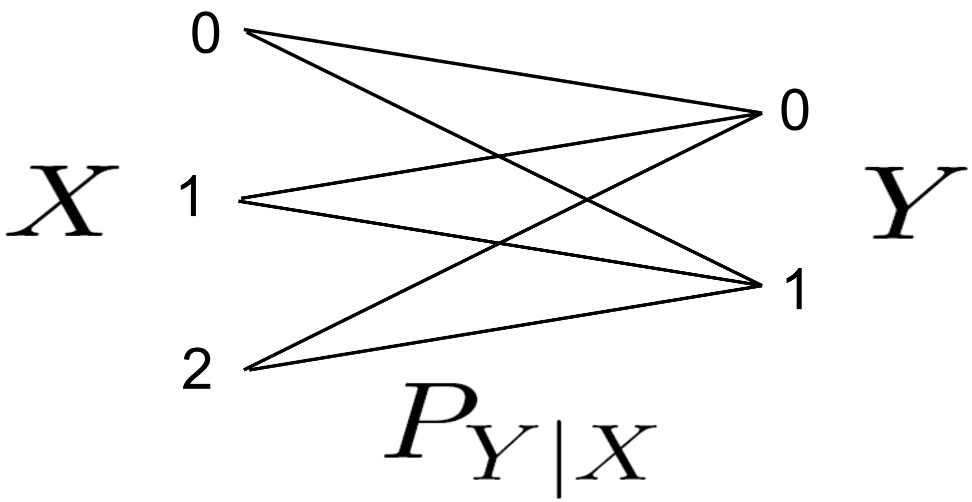
- The PM kernel has a fixed-point at θ_f if

$$F_{\Theta|Y}(\theta_f|y) = \theta_f \,, \quad \forall\, y \in \mathcal{Y}$$

- Claim: If the PM kernel has a fixed point, then the PM scheme is not ergodic and cannot achieve any positive rate

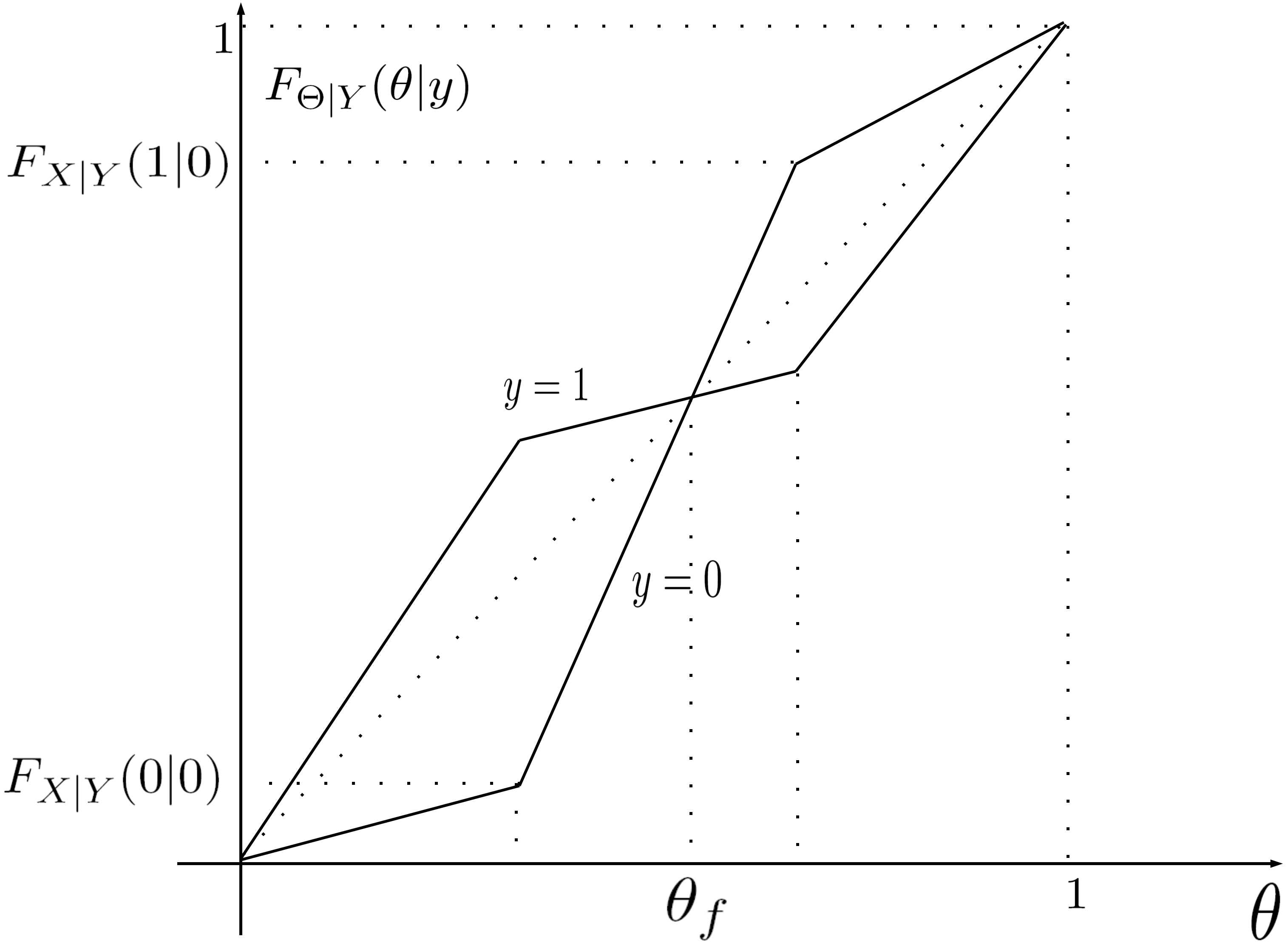


3×2 *Example*

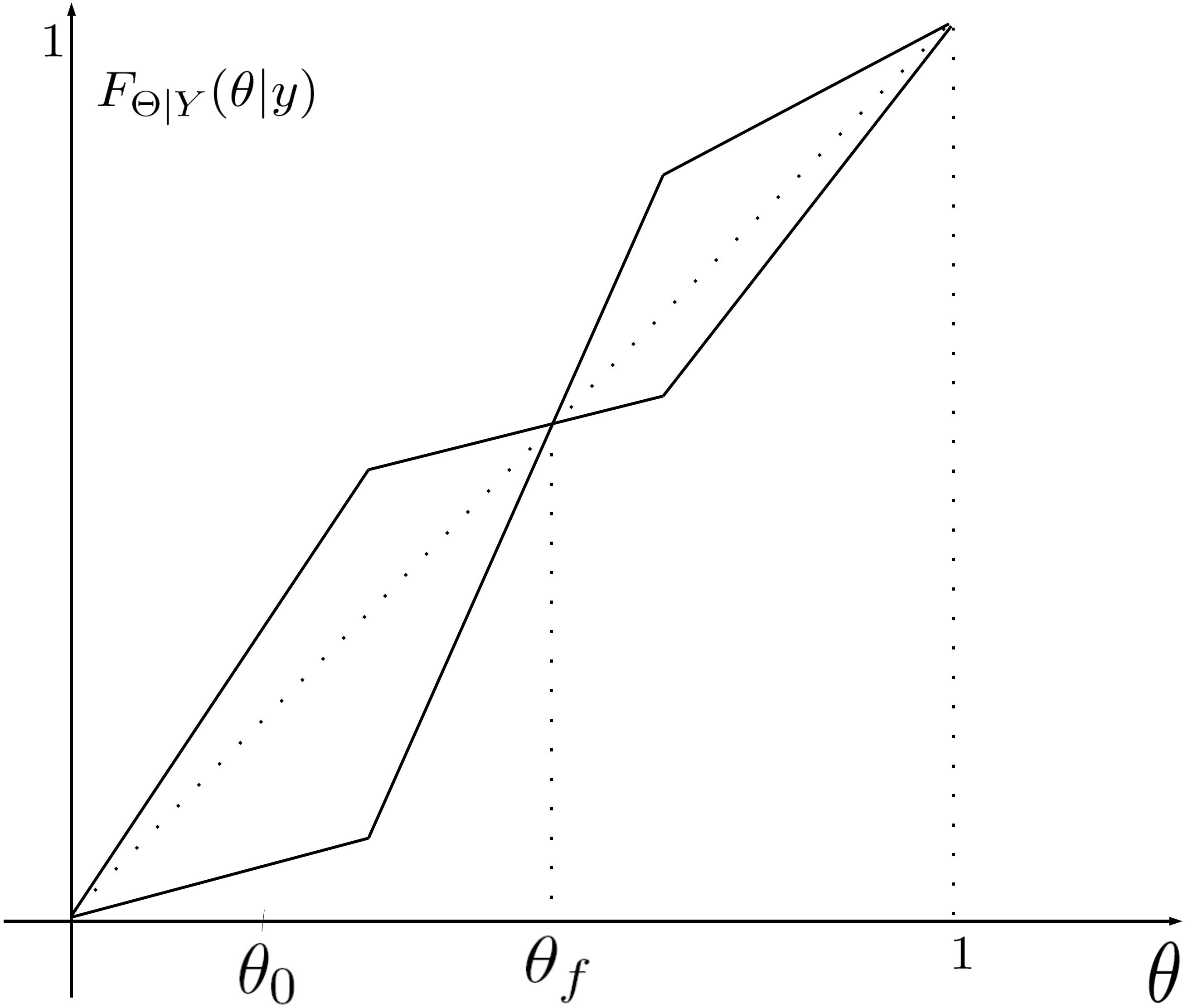


- PM kernel comprised of two quasi-linear functions
- Suppose there is a fixed point, i.e., $F_{\Theta|Y}(\theta_f|0) = F_{\Theta|Y}(\theta_f|1) = \theta_f$

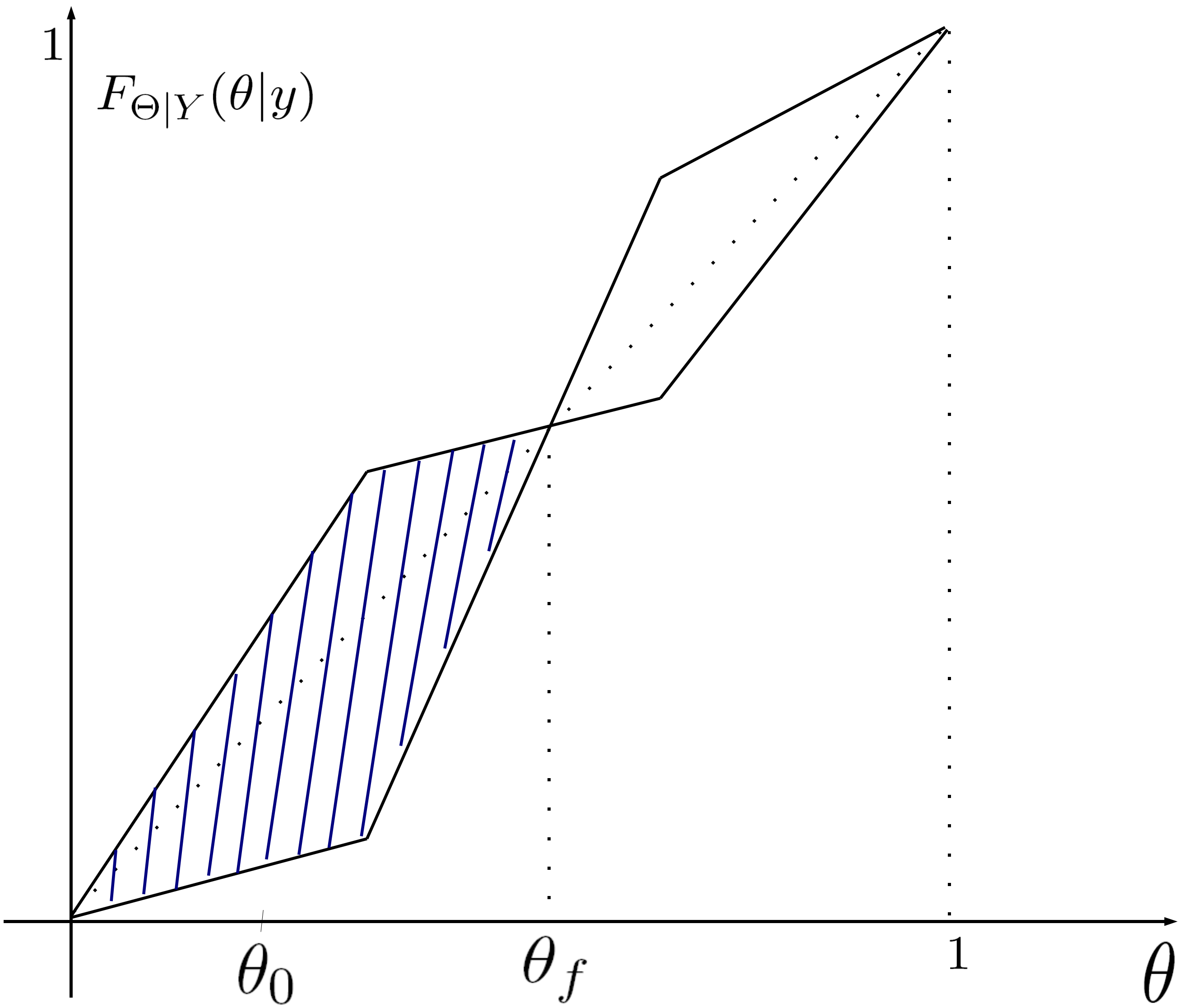
3×2 *Example*



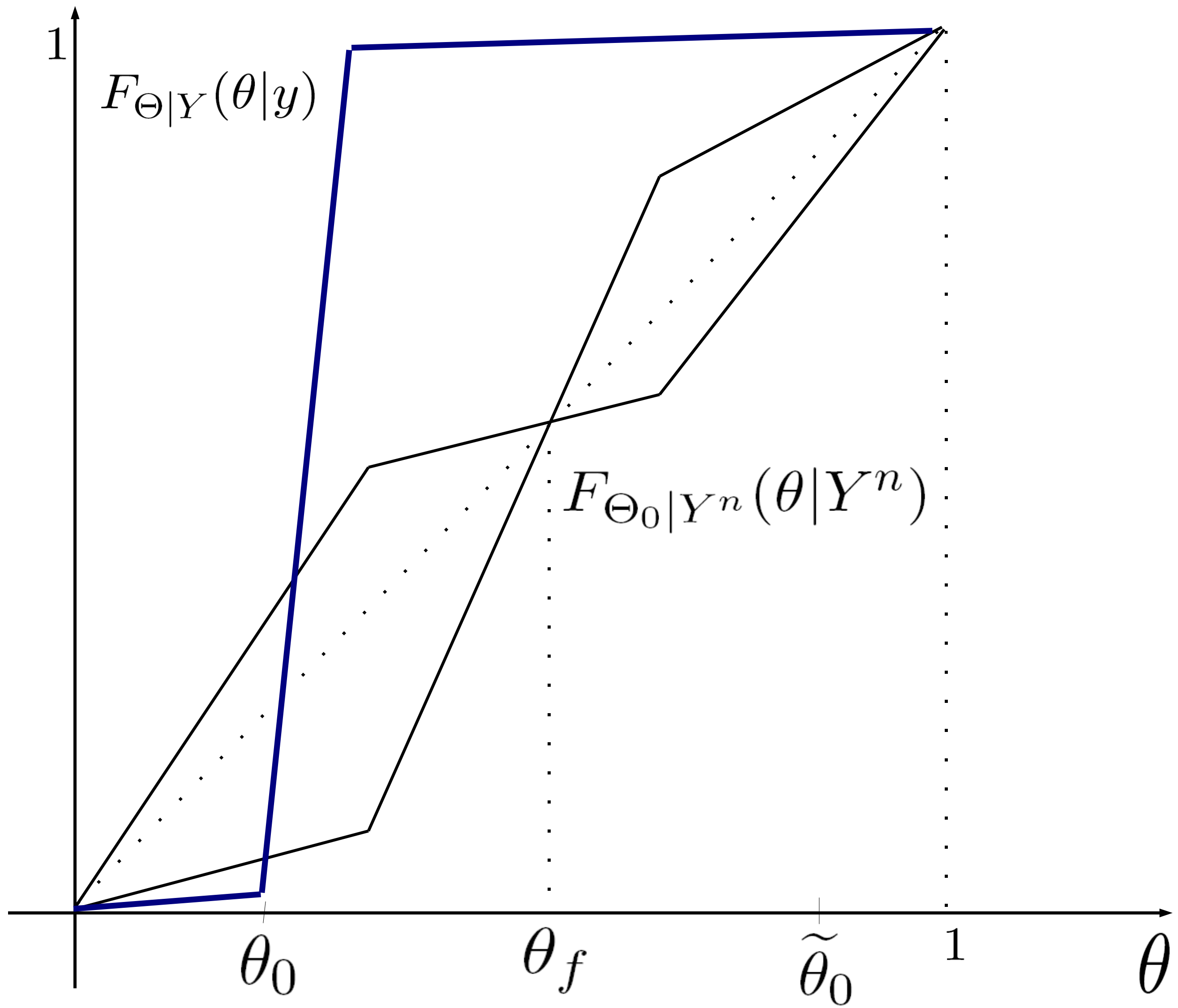
3×2 *Example*



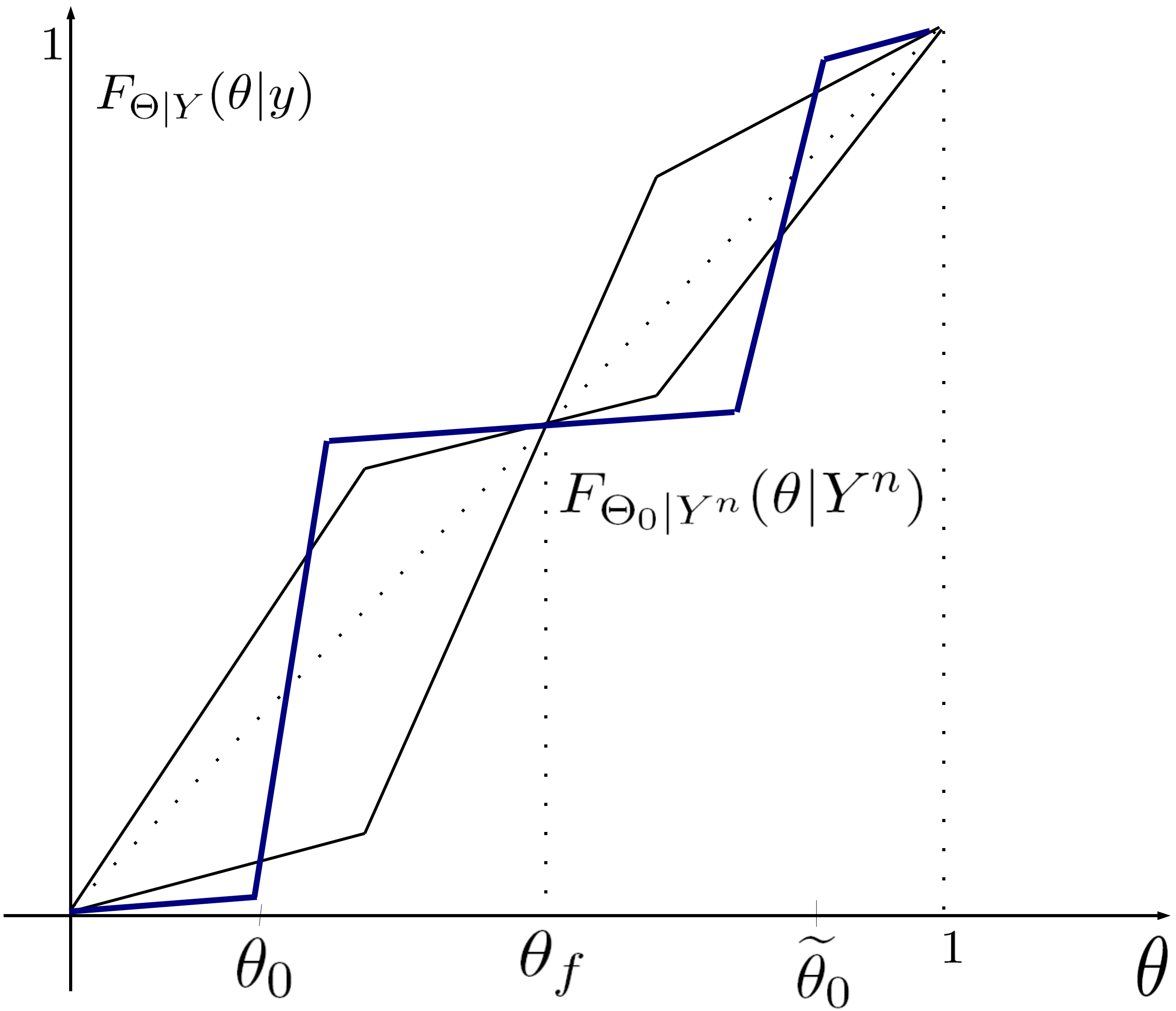
3×2 *Example*



3×2 *Example*

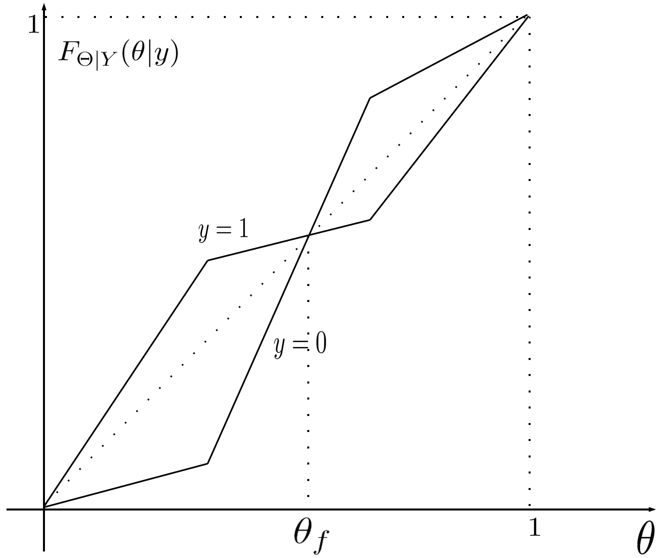


3×2 *Example*

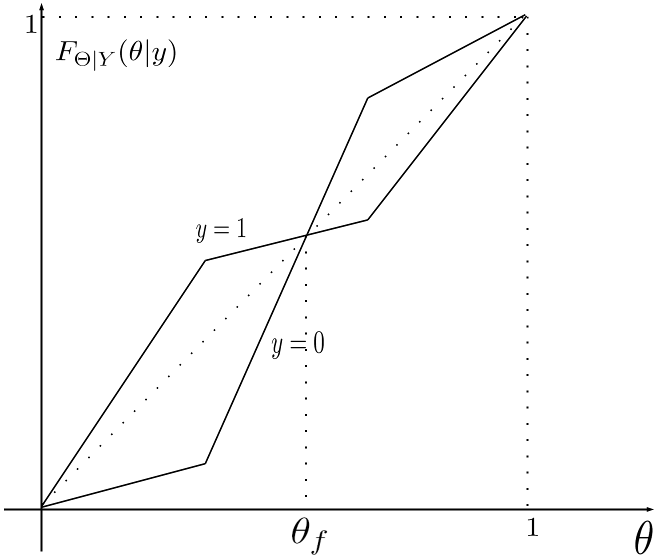


3×2 *Example – What can we do?*

- Two invariant intervals
- Idea I:
  - Decode two messages
  - Resolve (e.g. via repetition)



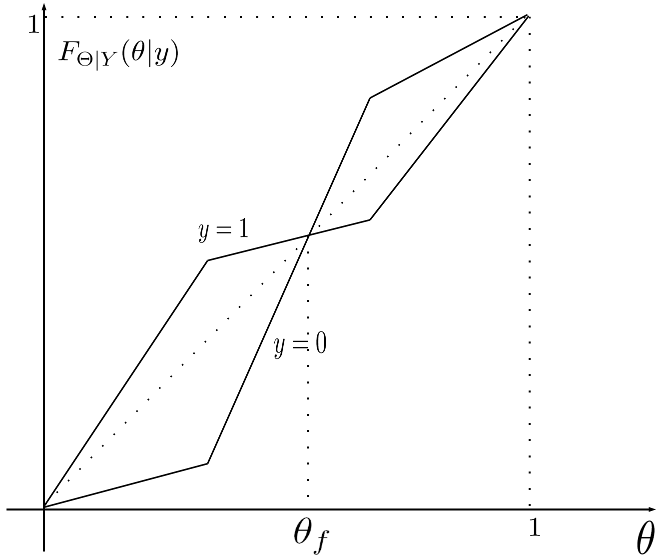
3×2 *Example – What can we do?*



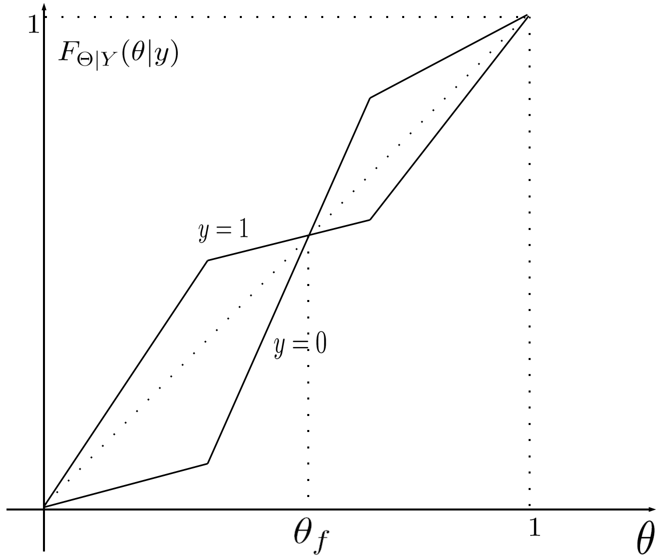
- Two invariant intervals
- Idea I:
  - Decode two messages
  - Resolve (e.g. via repetition)
- Problems:
  - Rate per invariant interval can be $< I(X;Y)$
  - Non-ergodic chain, input constraints not necessarily satisfied

3×2 *Example – What can we do?*

- Two invariant intervals
- Idea II:
  - Map message to "best" interval
  - $R \geq I(X;Y)$

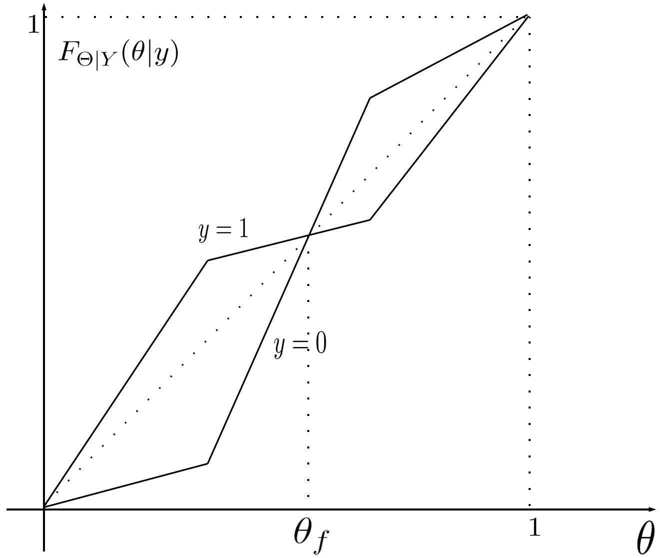


3×2 *Example – What can we do?*



- Two invariant intervals
- Idea II:
  - Map message to "best" interval
  - $R \geq I(X;Y)$
- Equivalent to PM with a different (binary) input distribution
- Related to the observation that when $|\mathcal{X}| > |\mathcal{Y}|$, using $|\mathcal{Y}|$ input symbols suffices to achieve the unconstrained capacity *[Shannon '57]*

3×2 *Example – What can we do?*



- Two invariant intervals
- Idea II:
  - Map message to ”best“ interval
  - $R \geq I(X;Y)$
- Equivalent to PM with a different (binary) input distribution
- Related to the observation that when $|\mathcal{X}| > |\mathcal{Y}|$, using $|\mathcal{Y}|$ input symbols suffices to achieve the unconstrained capacity *[Shannon '57]*
- Problem: Input constraints not satisfied

3×2 *Example – How Common?*

- Choose P_{XY} uniformly over the simplex
- Let $F_{\Theta|Y}$ be the corresponding PM kernel
- How likely is a fixed-point for $F_{\Theta|Y}$?

3×2 ***Example – How Common?***

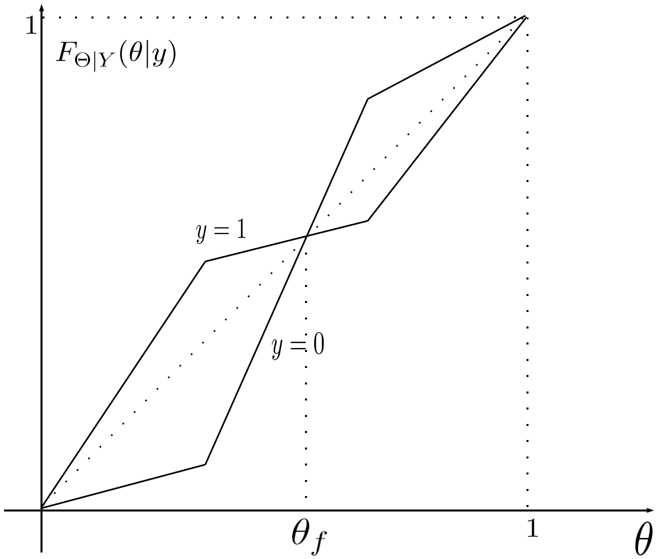
- Choose P_{XY} uniformly over the simplex
- Let $F_{\Theta|Y}$ be the corresponding PM kernel
- How likely is a fixed-point for $F_{\Theta|Y}$?
- Claim: $\mathbb{P}\left(F_{\Theta|Y} \text{ has a fixed point}\right) = \frac{1}{3}$

3×2 Example – How Common?

- Proof:
  - Define $\delta_k = P_{X|Y}(k|1) - P_{X|Y}(k|0)$, $k = 0, 1, 2$
  - $\sum \delta_k = 0$
  - Suppose $|\delta_0| \leq |\delta_2| \leq |\delta_1|$, hence $\delta_1 = -(\delta_0 + \delta_2)$
  - The c.d.f's must intersect

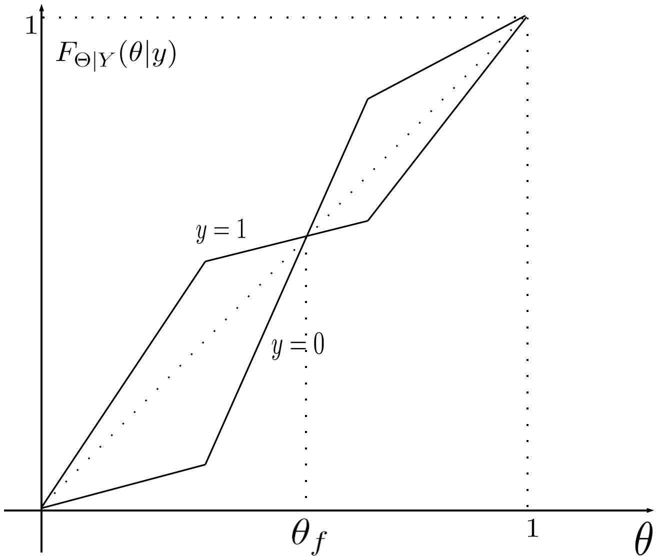
3×2 *Example – How Common?*

- Proof:
  - Define $\delta_k = P_{X|Y}(k|1) - P_{X|Y}(k|0)$, $k = 0, 1, 2$
  - $\sum \delta_k = 0$
  - Suppose $|\delta_0| \leq |\delta_2| \leq |\delta_1|$, hence $\delta_1 = -(\delta_0 + \delta_2)$
  - The c.d.f's must intersect



3×2 *Example – How Common?*

- Proof:
  - Define $\delta_k = P_{X|Y}(k|1) - P_{X|Y}(k|0)$, $k = 0, 1, 2$
  - $\sum \delta_k = 0$
  - Suppose $|\delta_0| \leq |\delta_2| \leq |\delta_1|$, hence $\delta_1 = -(\delta_0 + \delta_2)$
  - The c.d.f's must intersect
  - $\mathbb{E}(F_{\Theta|Y}(\theta|Y)) = F_{\Theta}(\theta) = \theta$
  - Intersection is a fixed point



3×2 *Example – How Common?*

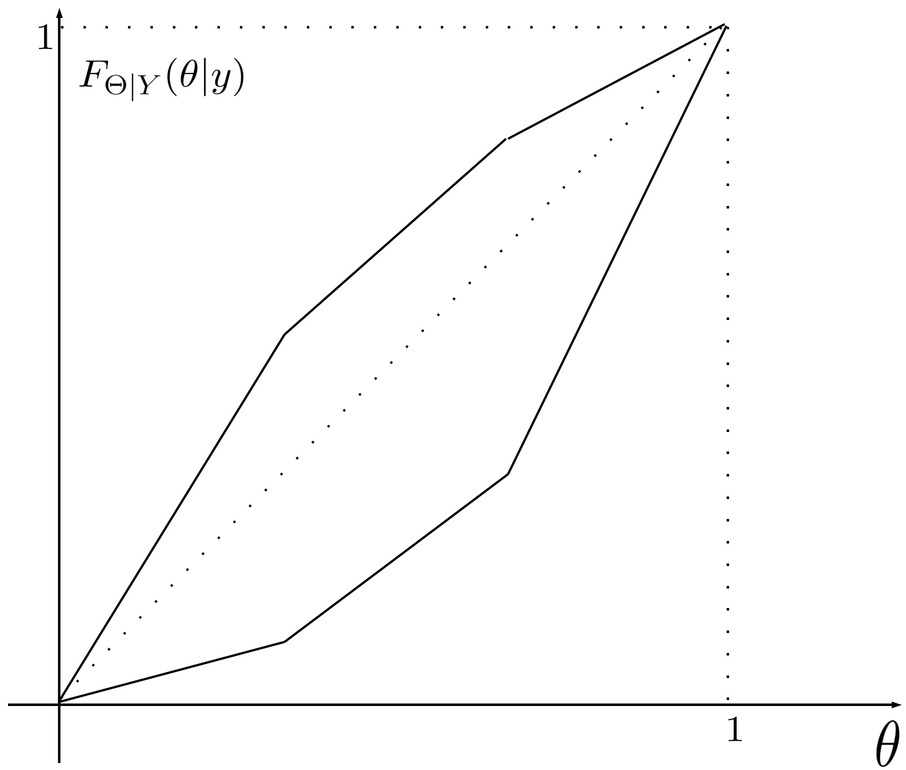
- Proof:
  - Define $\delta_k = P_{X|Y}(k|1) - P_{X|Y}(k|0)$, $k = 0, 1, 2$
  - $\sum \delta_k = 0$
  - Suppose $|\delta_0| \leq |\delta_2| \leq |\delta_1|$, hence $\delta_1 = -(\delta_0 + \delta_2)$
  - The c.d.f's must intersect
  - $\mathbb{E}(F_{\Theta|Y}(\theta|Y)) = F_\Theta(\theta) = \theta$
  - Intersection is a fixed point
  - Can always permute/relabel the input so that δ_k satisfies this
  - Fixed point obtained also for mirror case
  - 2 out of 6 permutations $\Rightarrow$ fixed point
  - Result follows from symmetry

A Systematic Solution

- Note the glass is $\frac{2}{3}$ full..
- Relabeling the input s.t. $\delta_0 \geq \delta_1 \geq \delta_2$

A Systematic Solution

- Note the glass is $\frac{2}{3}$ full..
- Relabeling the input s.t. $\delta_0 \geq \delta_1 \geq \delta_2 \;\Rightarrow\;$ no fixed points!



A Systematic Solution

- Note the glass is $\frac{2}{3}$ full..
- Relabeling the input s.t. $\delta_0 \geq \delta_1 \geq \delta_2 \ \Rightarrow$ no fixed points!
- For any DMC with $I(X;Y) > 0$
  - $\exists\, y_1, y_2 \in \mathcal{Y}$ s.t. $P_{X|Y}(\cdot|y_1) \neq P_{X|Y}(\cdot|y_2)$.
  - Find an input permutation σ sorting the δ_k, for which

$$F_{\sigma(X)|Y}(\cdot|y_1) < F_{\sigma(X)|Y}(\cdot|y_2)$$

  - Consider PM for the *equivalent* input/channel pair $(P_{\sigma(X)}, P_{Y|\sigma(X)})$
  - No fixed points, achieves $I(X;Y)$ within input constraints

The General Case

- Beyond permutations: A bijective function $\mu : (0,1) \mapsto (0,1)$ is called *uniformity preserving* if

$$\Theta \sim \mathrm{Unif}(0,1) \Rightarrow \mu(\Theta) \sim \mathrm{Unif}(0,1)$$

- The μ-variant of the PM scheme is

$$X_{n+1} = F_X^{-1} \circ \mu \circ F_{\Theta_0|Y^n}(\Theta_0|Y^n)$$

- μ-variant kernel $\mu \circ F_{\mu^{-1}(\Theta)|Y}(\cdot|y) \circ \mu^{-1}$
- μ eliminates fixed points, $I(X;Y)$ achieved under input constraints
- For DMC, μ was a permutation of intervals that correspond to the discrete input alphabet

Summary

- With fixed points, PM fails to achieve capacity
- PM variants can eliminate fixed points, achieving capacity
- Many variants achieving capacity, some “closer” to having a fixed-point than others
- What is the “best” variant, minimizing error probability for a given input/channel pair?