INFORMATION-THEORETIC LIMITS OF RANDOMNESS GENERATION

Cheuk Ting Li

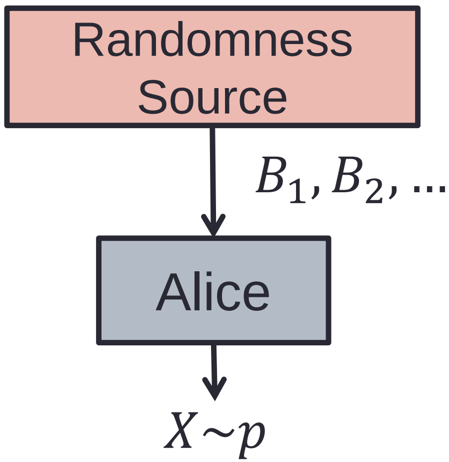
Department of Electrical Engineering, Stanford University

Randomness Generation

- Generating random variables from coin flips
- Applications:
  - Monte Carlo simulation
  - Randomized algorithms
  - Cryptography
- What is the minimum number of coin flips needed?

Outline

0. One-shot randomness generation



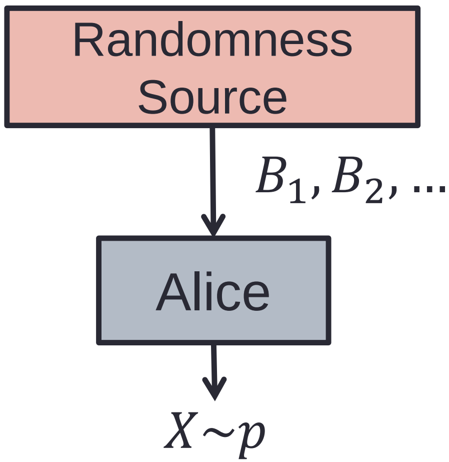
1. Distributed generation

2. Universal remote generation

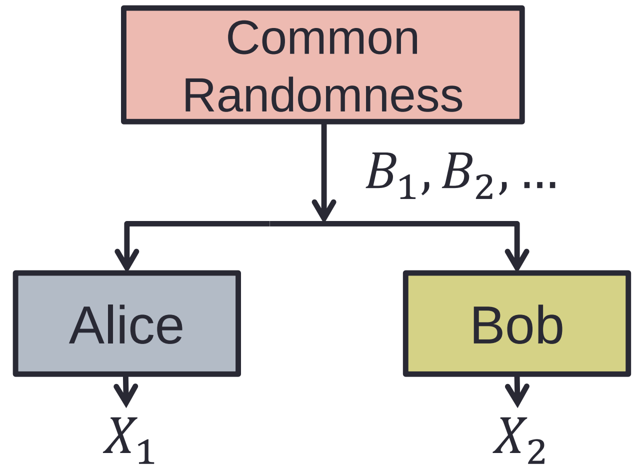
3. Channel simulation with Common Randomness

Outline

0. One-shot randomness generation



1. Distributed generation

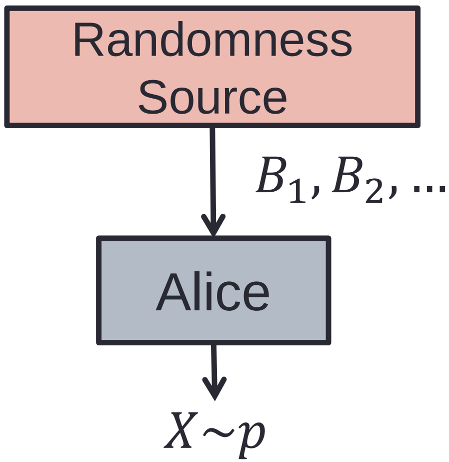


2. Universal remote generation

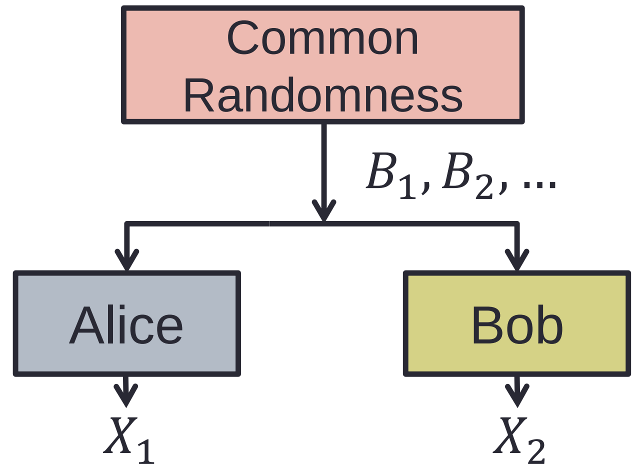
3. Channel simulation with Common Randomness

Outline

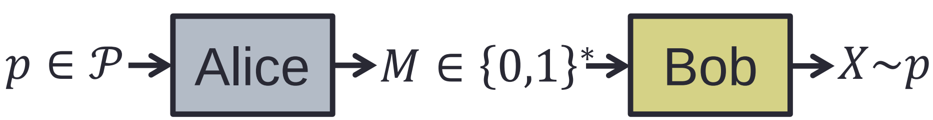
0. One-shot randomness generation



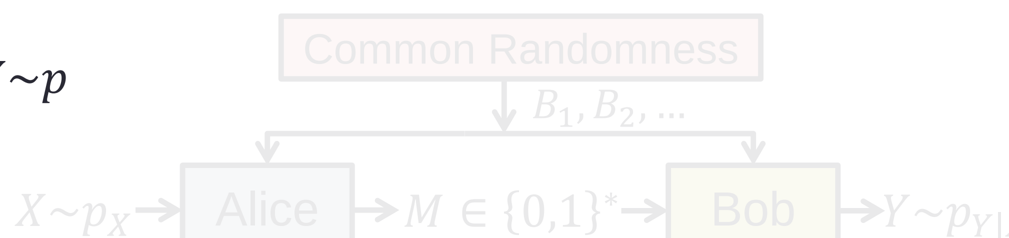
1. Distributed generation



2. Universal remote generation

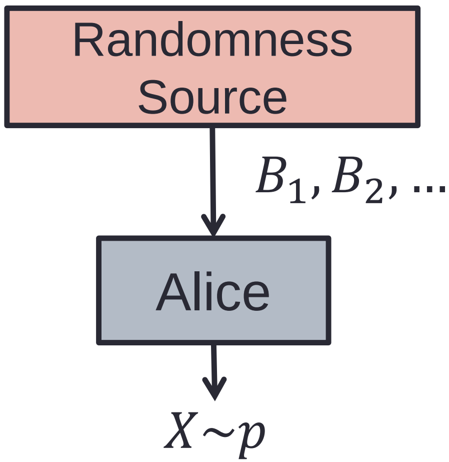


3. Channel simulation with Common Randomness

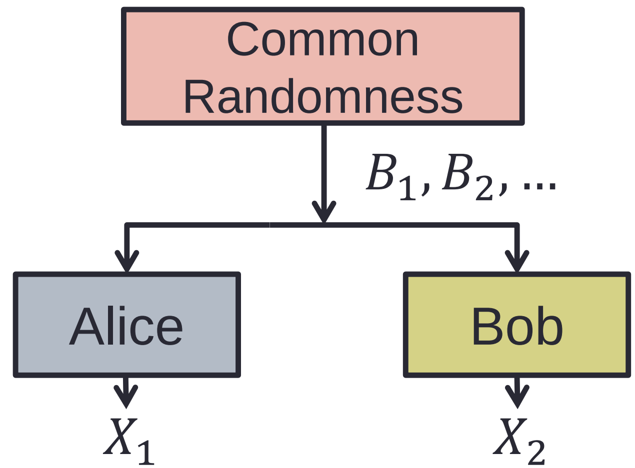


Outline

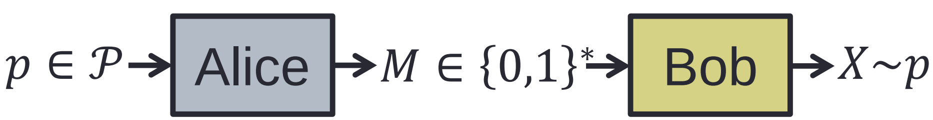
0. One-shot randomness generation



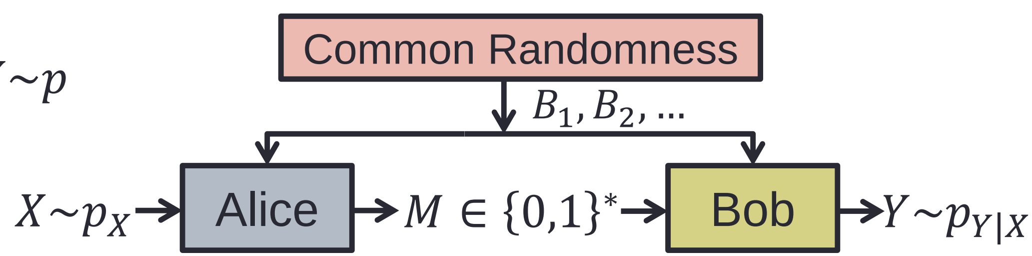
1. Distributed generation



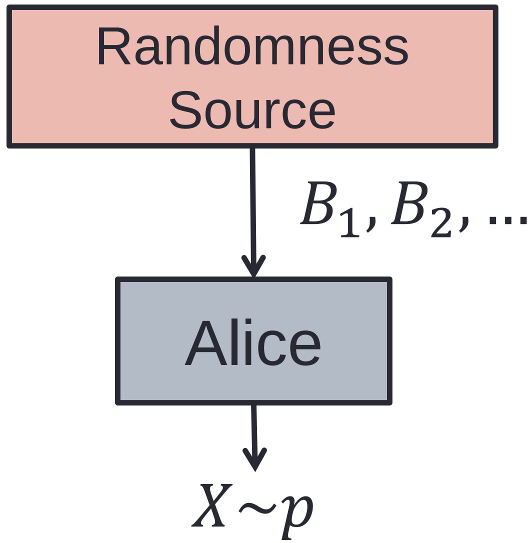
2. Universal remote generation



3. Channel simulation with Common Randomness



0. One-shot Randomness Generation



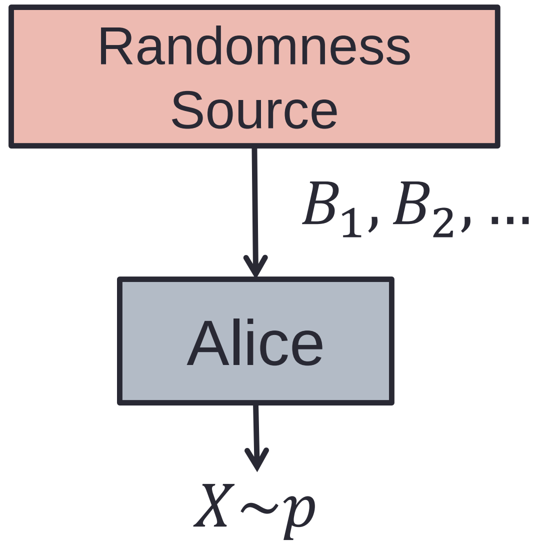
- Sequence of i.i.d. fair coin flips $B_1, B_2, \ldots \sim \mathrm{Bern}(1/2)$

1. At time i , Alice observes B_i, and
2. Either output X, or proceed to time $i+1$

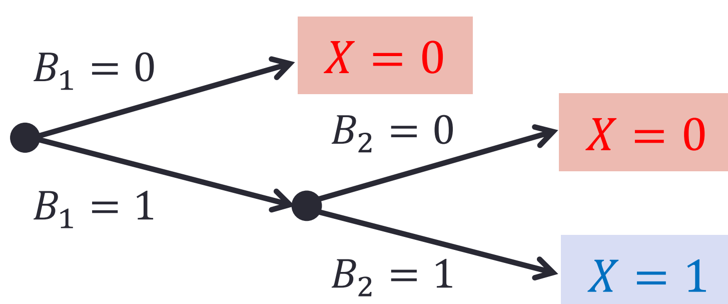
i.e., Alice decodes $B_1, B_2, \ldots$ using a prefix-free codebook

- Let L be the number of B_i bits observed
- What is the minimum $\mathrm{E}[L]$ to generate $X \sim p$?

0. One-shot Randomness Generation

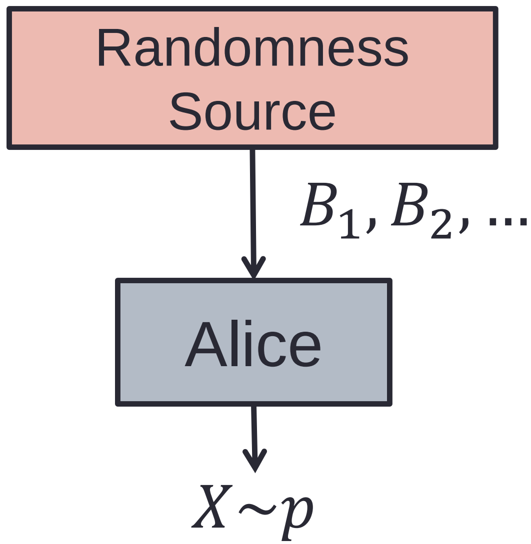


- E.g. $X \sim \mathrm{Bern}(1/4)$
- Flip twice, output 0 if $B_1 B_2 = 00{,}01{,}10$, output 1 if $B_1 B_2 = 11$
- Discrete distribution generating tree



$\mathrm{E}[L] = 3/2$

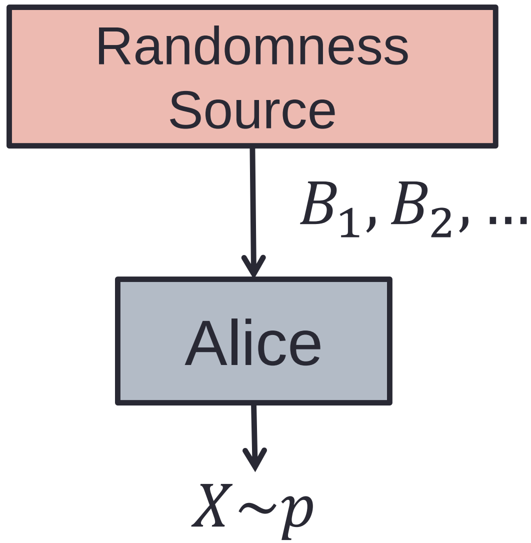
0. One-shot Randomness Generation



- Knuth-Yao (1976):

$$H(X) \leq \min \mathrm{E}[L] \leq H(X) + 2$$

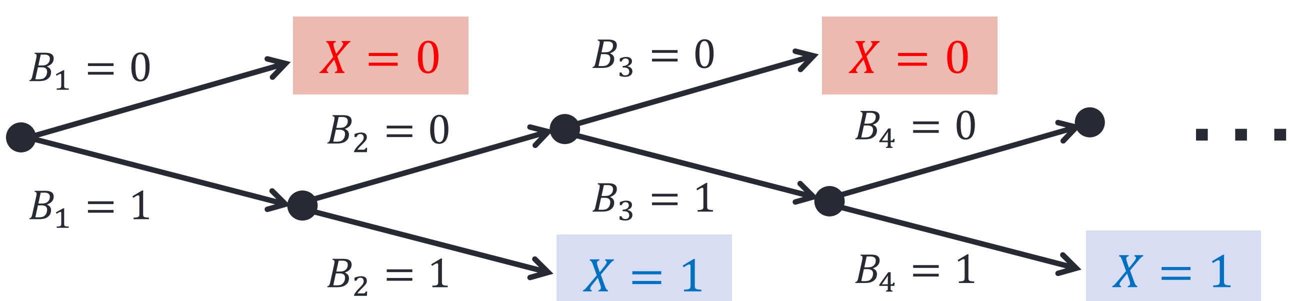
0. One-shot Randomness Generation



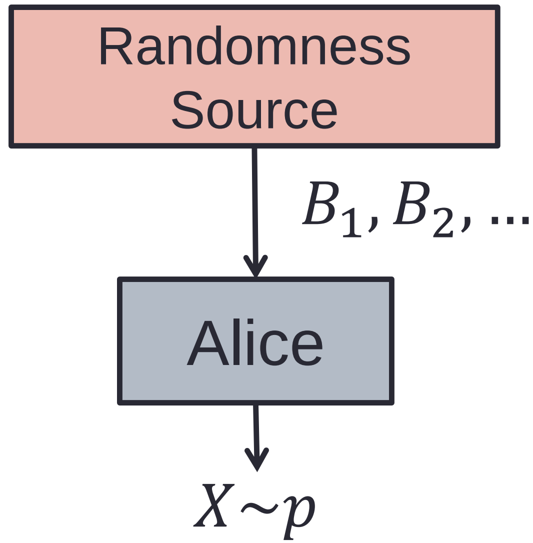
- Knuth-Yao (1976):

$$H(X) \leq \min \mathrm{E}[L] \leq H(X) + 2$$

- E.g. $X \sim \mathrm{Bern}(1/3)$



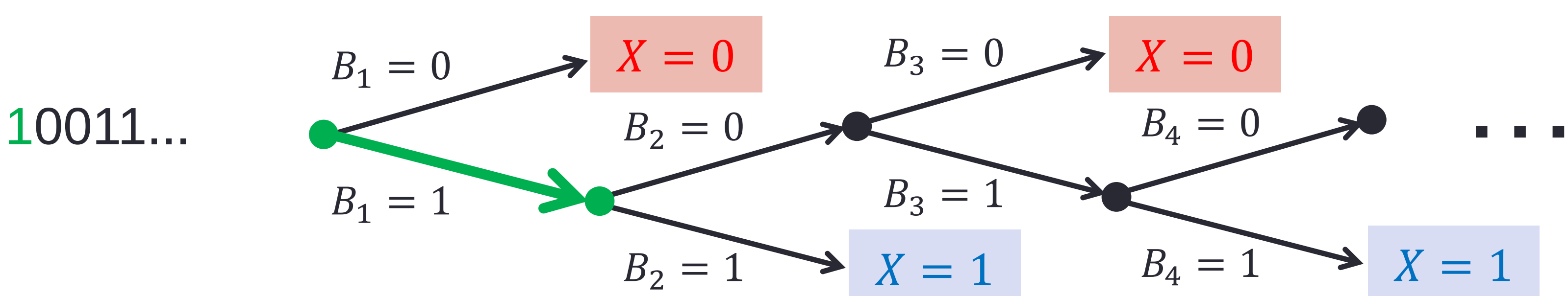
0. One-shot Randomness Generation



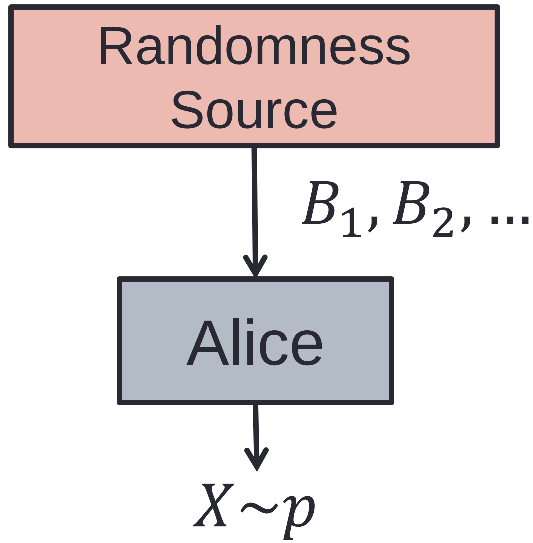
- Knuth-Yao (1976):

$$H(X) \leq \min \mathrm{E}[L] \leq H(X) + 2$$

- E.g. $X \sim \mathrm{Bern}(1/3)$



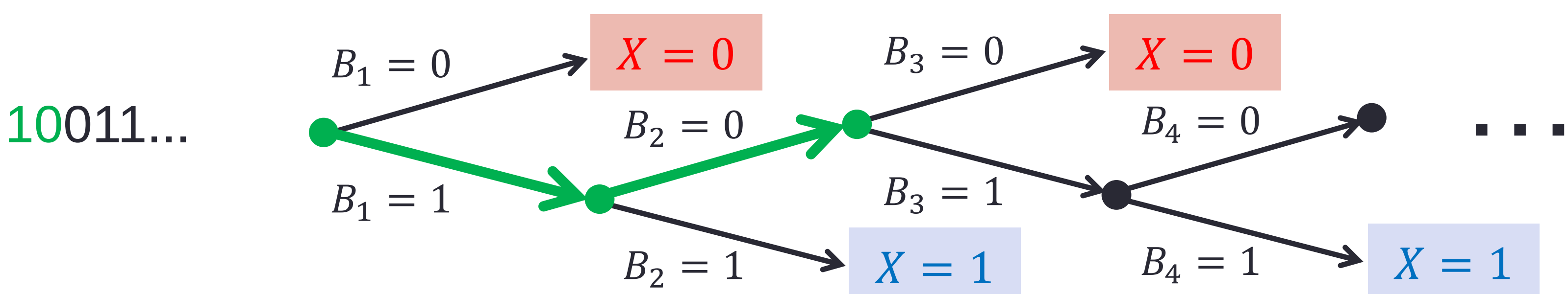
0. One-shot Randomness Generation



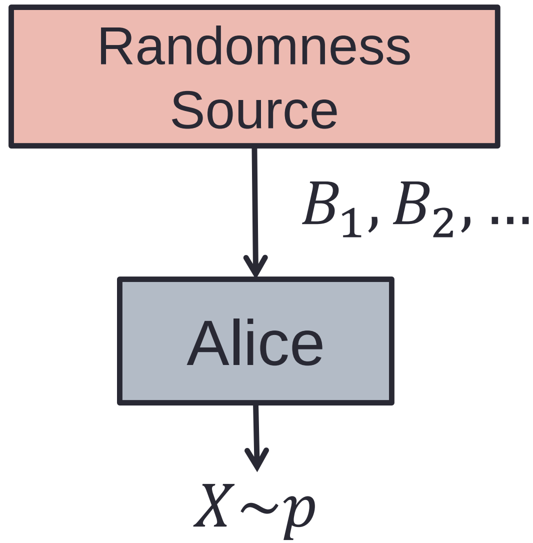
- Knuth-Yao (1976):

$$H(X) \leq \min \mathrm{E}[L] \leq H(X) + 2$$

- E.g. $X \sim \mathrm{Bern}(1/3)$



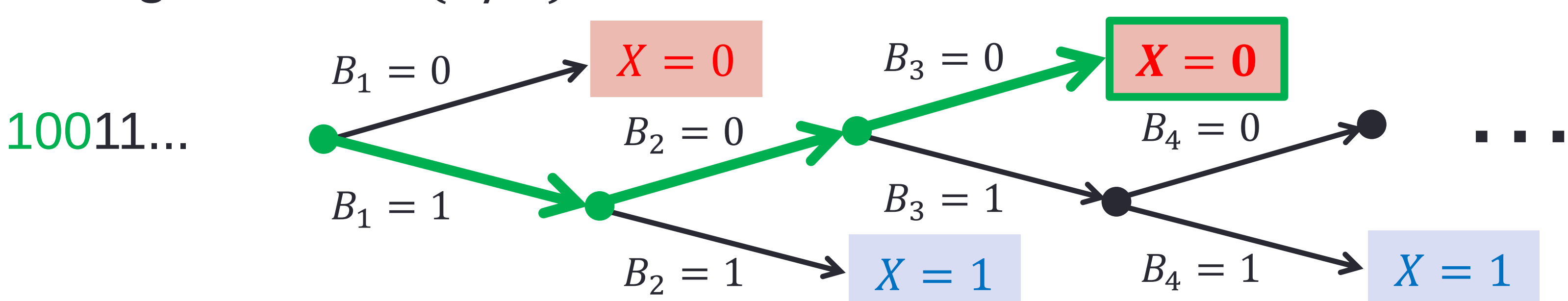
0. One-shot Randomness Generation



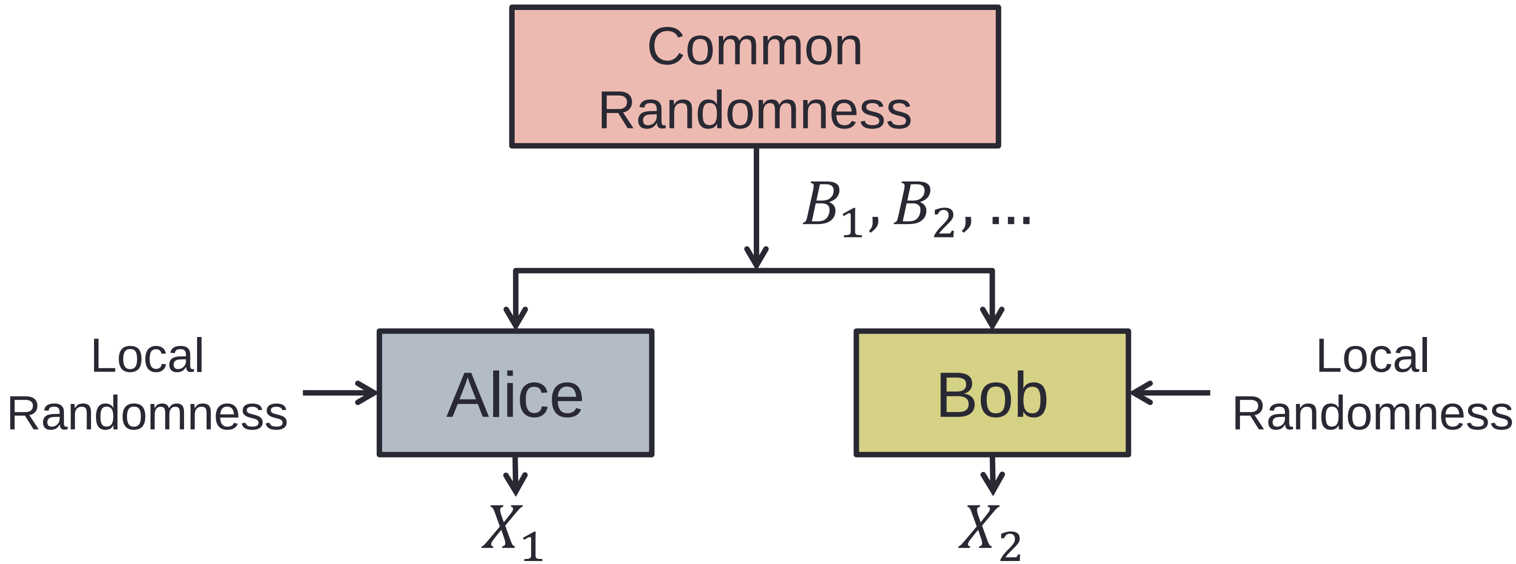
- Knuth-Yao (1976):

$$H(X) \le \min \mathrm{E}[L] \le H(X) + 2$$

- E.g. $X \sim \mathrm{Bern}(1/3)$

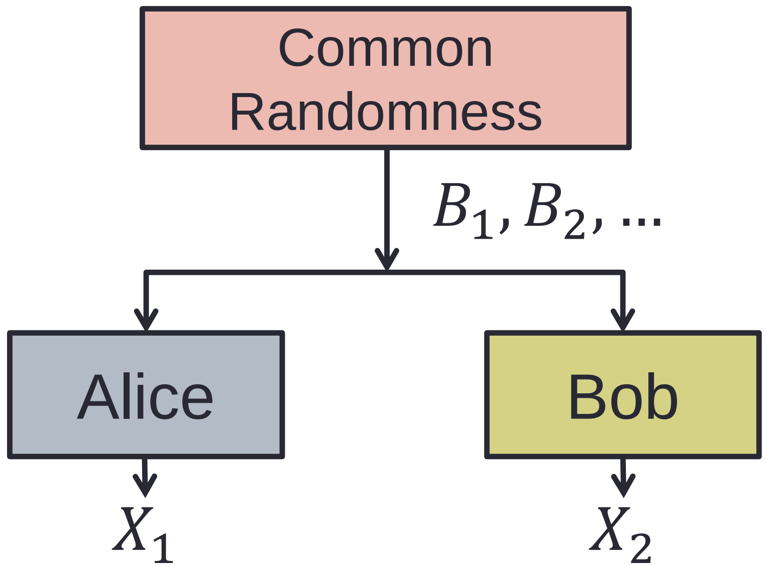


1. One-shot Distributed Generation



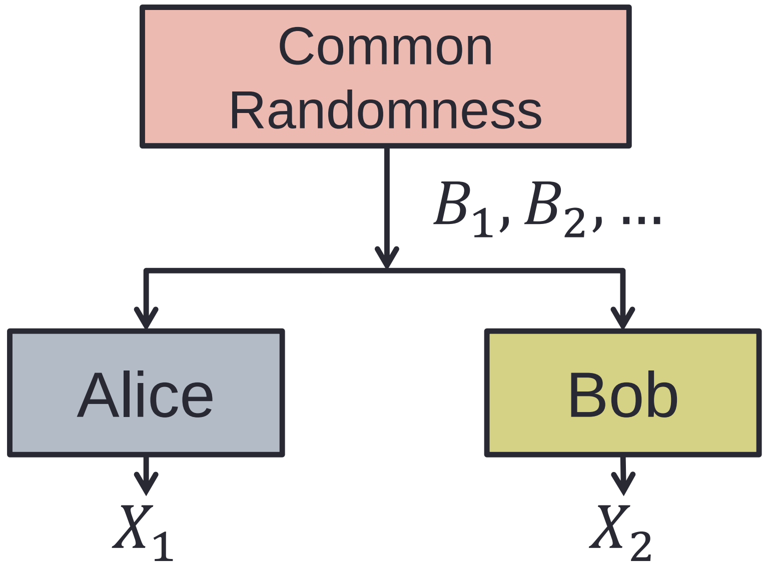
- Generalize to 2 random variables $(X_1, X_2) \sim p$
- 3 sequences of random bits:
  - Common randomness $B_1, B_2, \ldots$ to both Alice and Bob
  - Local randomness to Alice, and another to Bob

1. One-shot Distributed Generation



- Assume unlimited local randomness
- Alice and Bob must use the same DDG tree
  - Use the same number of common random bits
- Let L be number of common random bits B_i bits used
- What is the minimum $\mathrm{E}[L]$ to generate $(X_1, X_2) \sim p$?

1. One-shot Distributed Generation

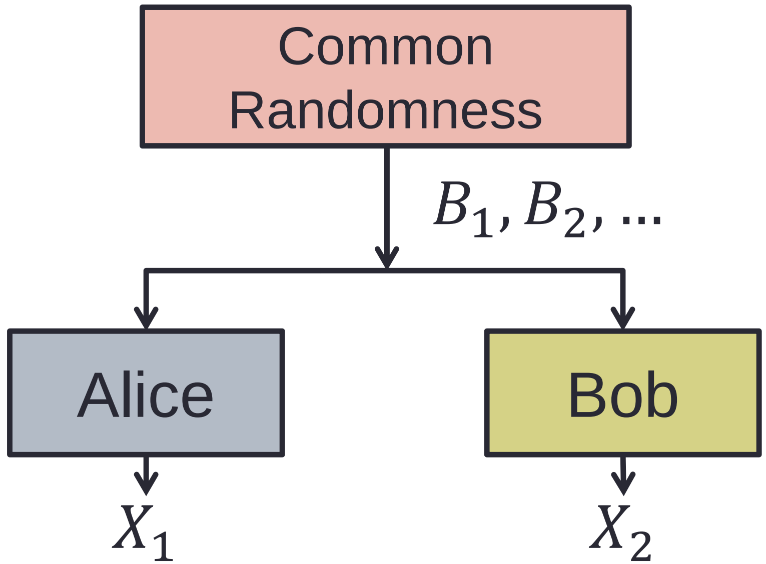


- Using Knuth-Yao, $G(X_1; X_2) \leq \min \mathrm{E}[L] \leq G(X_1; X_2) + 2$

  $G(X_1; X_2) = \min_{X_1 - W - X_2} H(W)$ common entropy [Kumar-Li-El Gamal 2014]

- Focus on bounding $G(X_1; X_2)$

1. One-shot Distributed Generation



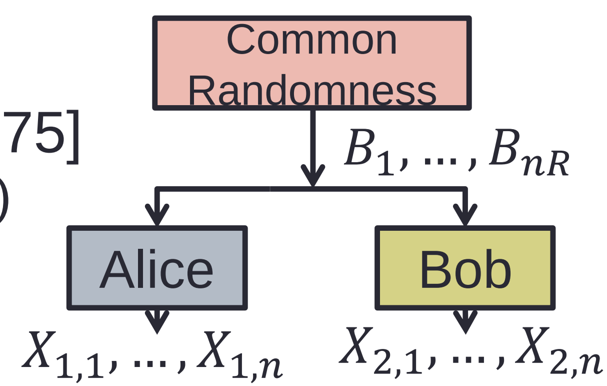
- For discrete (X_1, X_2),

$$I(X_1;X_2) \le J(X_1;X_2) \le G(X_1;X_2) \le \min\{H(X_1), H(X_2)\}$$

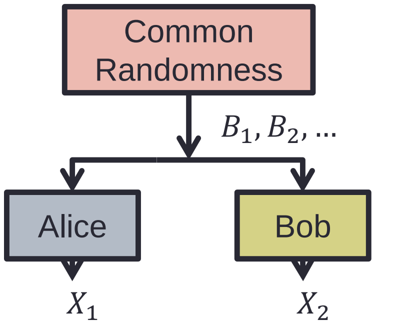
  - I : mutual information
  - J : Wyner common information [Wyner 1975]

$$J(X_1;X_2) = \min_{X_1 - V - X_2} I(X_1, X_2; V)$$

    - Asymptotic distributed generation with vanishing total variation distance



What about (X_1, X_2) continuous?



- Still have

$$I(X_1; X_2) \le J(X_1; X_2) \le G(X_1; X_2) \le ?$$

- But no general upper bound on G or J
- For example:

$$(X_1, X_2) \sim N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right), \ \ \rho < 1$$

  - $I = \frac{1}{2} \log\left(\frac{1}{1-\rho^2}\right)$
  - $J = \min_{X_1 - V - X_2} I(X_1, X_2; V) = \frac{1}{2} \log\left(\frac{1+\rho}{1-\rho}\right)$

    $V \sim N(0, \rho),\ X_i = V + Z_i,\ Z_i \sim N(0, 1-\rho)$
  - Is $G(X_1; X_2) = \min_{X_1 - W - X_2} H(W)$ also finite?

Main Result

- We show that for (X_1, X_2) with log-concave pdf f

Theorem [Li-El Gamal 2016]

$$I(X_1; X_2) \leq J(X_1; X_2) \leq G(X_1; X_2) \leq I(X_1; X_2) + 24$$

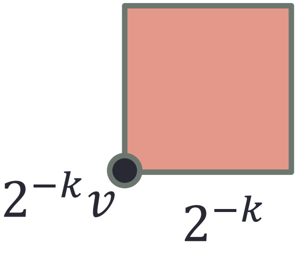
- Result extends to n continuous random variables

Outline of Proof

- (X_1, X_2) uniform over a set in $\mathbf{R}^2$
  - Construct W using dyadic decomposition scheme
  - Bound $H(W)$ by $I(X_1; X_2)$ using erosion entropy
- (X_1, X_2) general pdf
  - Apply scheme for uniform to hypograph of pdf
  - Bound $G(X_1; X_2)$ by $I(X_1; X_2)$ for log-concave pdf

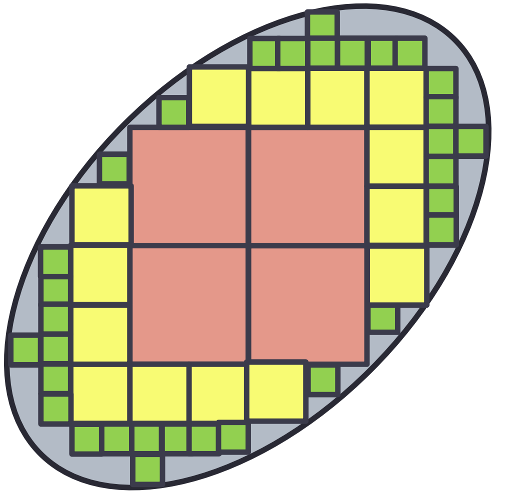
$(X_1, X_2) \sim \text{Unif}(A),\ A \subseteq \mathbf{R}^2$

- Scheme uses dyadic decomposition
- Dyadic square

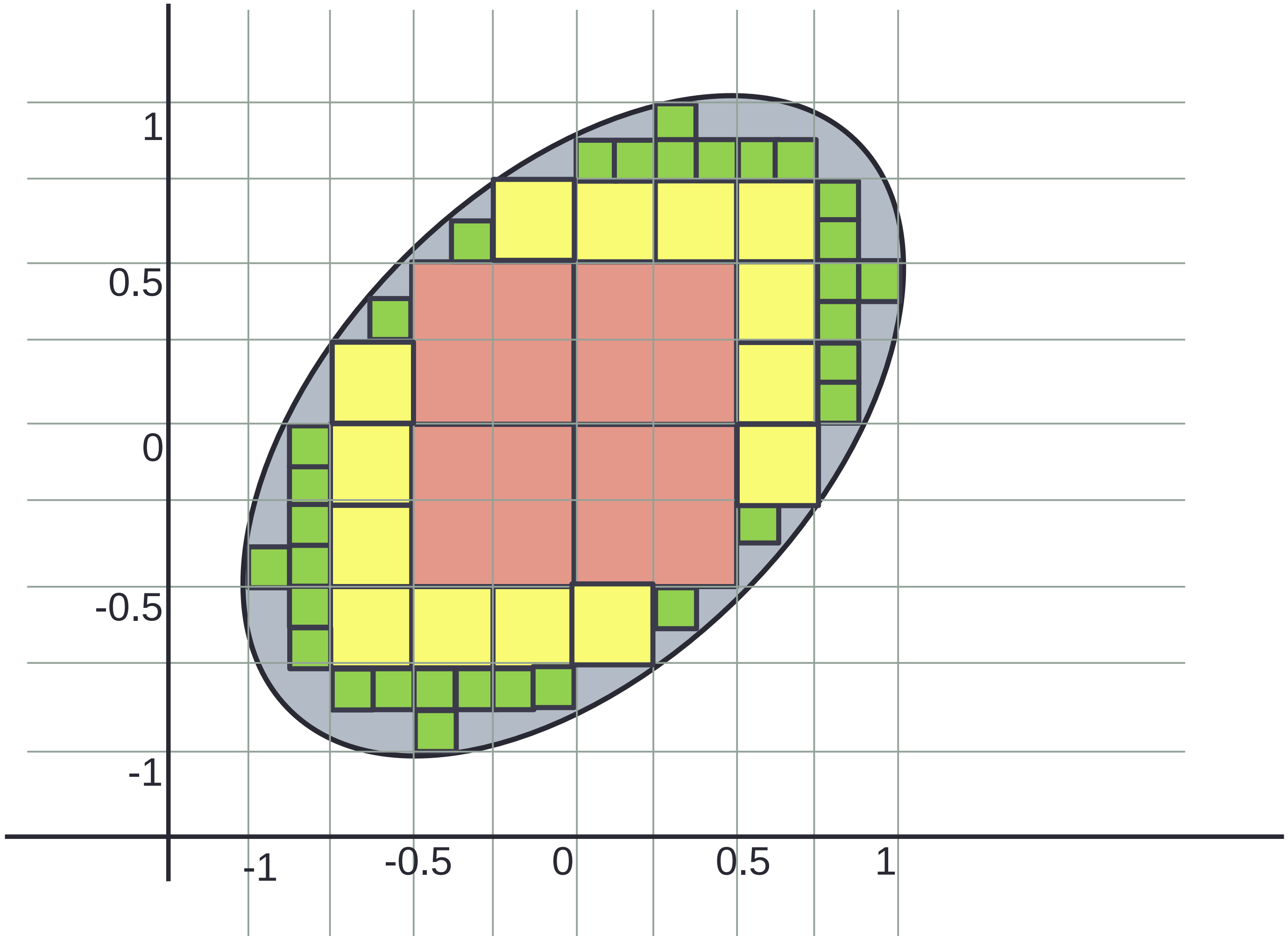


$k \in \mathbf{Z}$ and $v \in \mathbf{Z}^2$

- Partition A into largest possible dyadic squares

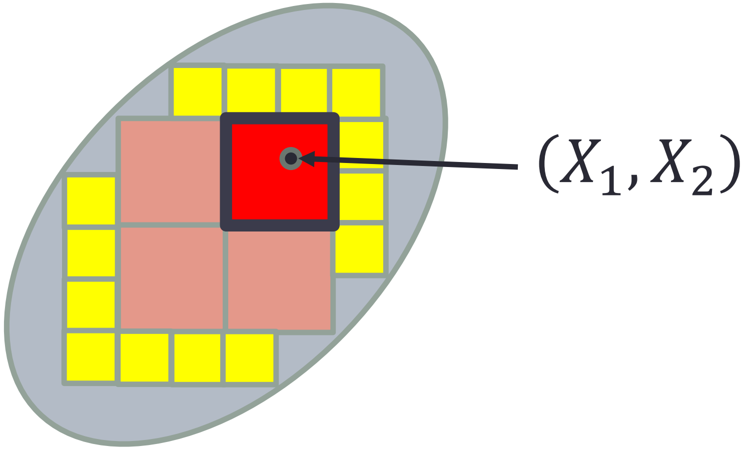


Dyadic Decomposition Example



Constructing W from Dyadic Decomposition

- W_{D} is the square containing (X_1, X_2)
- Conditioned on W_{D}, X_1 and X_2 are uniformly distributed over each square side, thus $X_1 - W_{\mathrm{D}} - X_2$

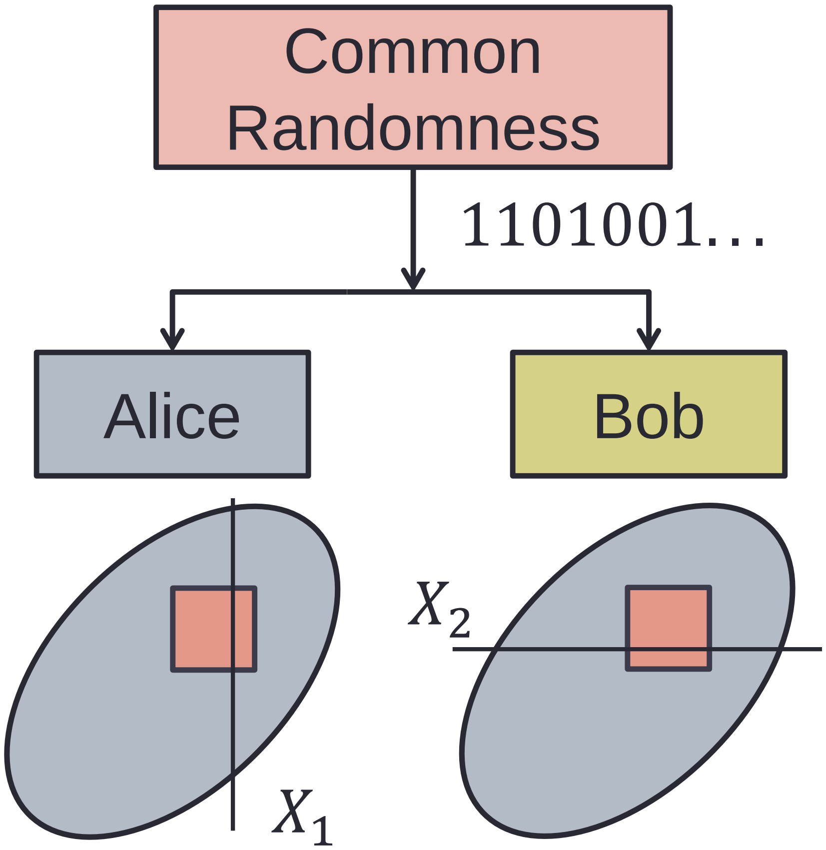


- Hence

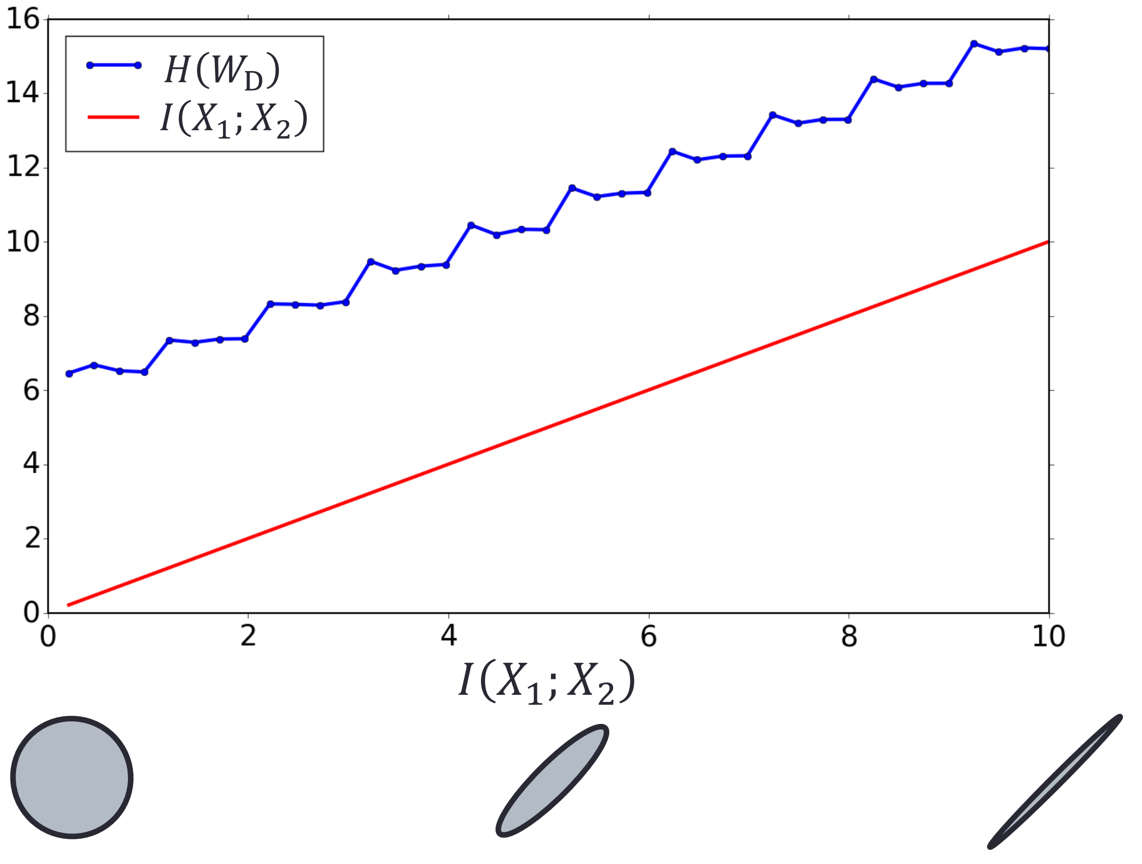
$$H(W_{\mathrm{D}}) \geq G(X_1; X_2) = \min_{X_1 - W - X_2} H(W)$$

Scheme for $(X_1, X_2) \sim \mathrm{Unif}(A)$

- Use Knuth-Yao to generate W_{D} using common random bits
- Generate X_1 or X_2 uniformly on each square side of W_{D}

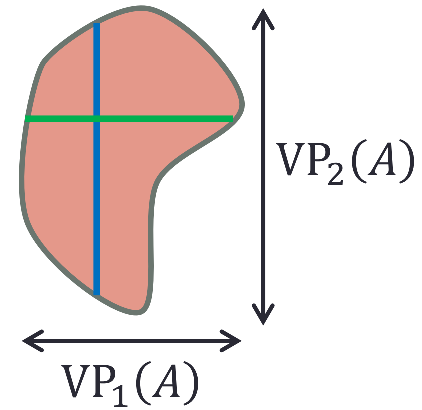


(X_1, X_2) Uniform over Ellipse



- Constant bound on gap?

Bound on $H(W_{\mathrm{D}})$



- If A orthogonally convex, i.e., intersection with axis-aligned lines are connected

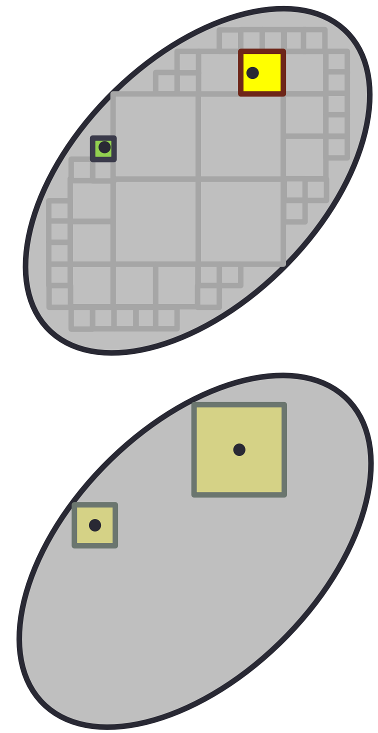
Proposition

$$G(X_1; X_2) \le H(W_{\mathrm{D}}) \le \log\left(\frac{\left(\mathrm{VP}_1(A) + \mathrm{VP}_2(A)\right)^2}{\mathrm{Vol}(A)}\right) + 4 + 2\log e$$

where $\mathrm{VP}_1(A) =$ length of projection of A onto x-axis

- Proof:
  - Bound $H(W_{\mathrm{D}})$ by erosion entropy
  - Bound erosion entropy by $\mathrm{VP}_1(A), \mathrm{VP}_2(A)$, $\mathrm{Vol}(A)$

Bounding $H(W_{\mathrm{D}})$ by Erosion Entropy



- $H(W_{\mathrm{D}}) = \mathrm{E}\left[-\log(L_{\mathrm{dy}}^2/\mathrm{Vol}(A))\right]$

  L_{dy}: side length of largest dyadic square $\ni (X_1, X_2)$

  $\Rightarrow \frac{1}{2}(H(W_{\mathrm{D}}) - \log \mathrm{Vol}(A)) = \mathrm{E}[-\log L_{\mathrm{dy}}]$

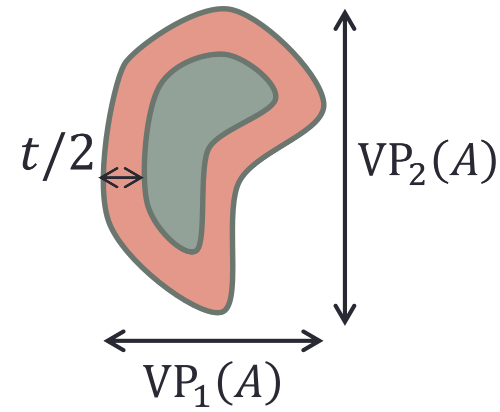
- Erosion entropy: $h_{\ominus B}(A) = \mathrm{E}[-\log L_{\mathrm{cen}}]$

L_{cen}: side length of largest square centered at (X_1, X_2)

Lemma

$$\frac{1}{2}(H(W_{\mathrm{D}}) - \log \mathrm{Vol}(A)) \leq h_{\ominus B}(A) + 2$$

Bounding Erosion Entropy



- When A orthogonally convex

$$\mathrm{P}\{L_{\mathrm{cen}} \leq t\} \leq t \cdot \frac{\mathrm{VP}_1(A) + \mathrm{VP}_2(A)}{\mathrm{Vol}(A)}$$

Lemma

$$h_{\ominus B}(A) = \mathrm{E}[-\log L_{\mathrm{cen}}] \leq \log\left(\frac{\mathrm{VP}_1(A) + \mathrm{VP}_2(A)}{\mathrm{Vol}(A)}\right) + \log e$$

- Substitute this into $\frac{1}{2}(H(W_{\mathrm{D}}) - \log \mathrm{Vol}(A)) \leq h_{\ominus B}(A) + 2$

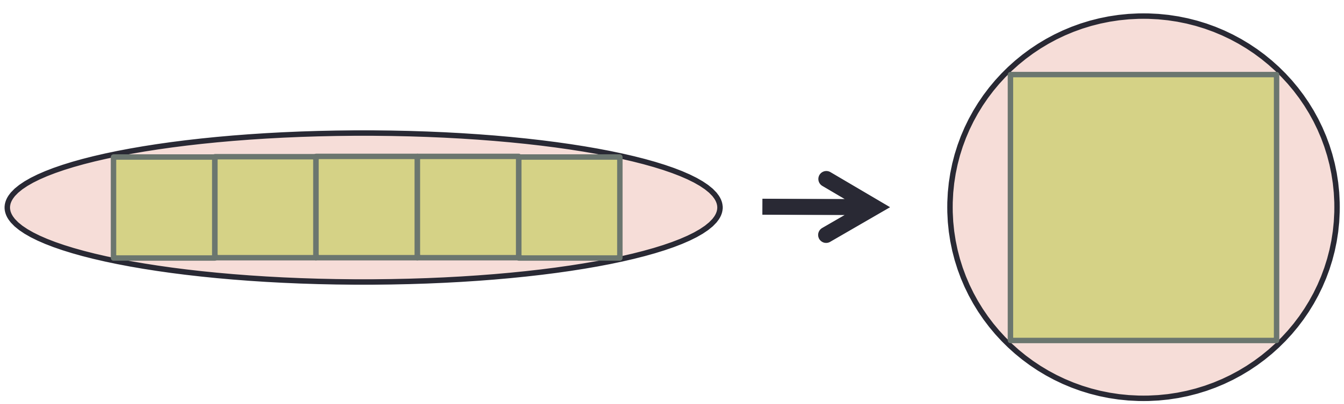
Gives proposition: $H(W_{\mathrm{D}}) \leq \log\left(\frac{(\mathrm{VP}_1(A)+\mathrm{VP}_2(A))^2}{\mathrm{Vol}(A)}\right) + 4 + 2\log e$

Scaling

$$H(W_{\mathrm{D}}) \leq \log\left(\frac{\left(\mathrm{VP}_1(A) + \mathrm{VP}_2(A)\right)^2}{\mathrm{Vol}(A)}\right) + 4 + 2\log e$$

- Bound depends on perimeter to area ratio
- A "flat" shape has high perimeter to area ratio and high $H(W_{\mathrm{D}})$
- Scale (X_1, X_2) to $(X_1', X_2') \sim \mathrm{Unif}(A')$ to make $\mathrm{VP}_1(A') = \mathrm{VP}_2(A')$:

$$H(W_{\mathrm{D}}') \leq \log\left(\frac{\mathrm{VP}_1(A) \cdot \mathrm{VP}_2(A)}{\mathrm{Vol}(A)}\right) + 6 + 2\log e$$



Bounding $H(W_{\mathrm{D}}')$ by $I(X_1;X_2)$

- We have

$$H(W_{\mathrm{D}}') \le \log\left(\frac{\mathrm{VP}_1(A)\cdot \mathrm{VP}_2(A)}{\mathrm{Vol}(A)}\right) + 6 + 2\log e$$

- Expanding $I(X_1;X_2) = h(X_1) + h(X_2) - \log\big(\mathrm{Vol}(A)\big)$
- Combining these two lines

$$H(W_{\mathrm{D}}') \le I(X_1;X_2) + \big(\log \mathrm{VP}_1(A) - h(X_1)\big) + \big(\log \mathrm{VP}_2(A) - h(X_2)\big) + 6 + 2\log e$$

- If $\mathrm{Unif}(A)$ log-concave $\Rightarrow$ A convex, marginal of X_i not too non-uniform

$$h(X_i) \approx \log \mathrm{VP}_i(A)\,,\, i = 1{,}2$$

- And we obtain constant gap between $H(W_{\mathrm{D}}')$ and $I(X_1;X_2)$

Bounding $H(W_{\mathrm{D}})$ for $(X_1, X_2) \sim \mathrm{Unif}(A)$

$H(W_{\mathrm{D}})$

$\lesssim$ — Erosion entropy bounds $H(W_{\mathrm{D}})$

$h_{\ominus B}(A)$

$\lesssim$ — If A orthogonally convex

$$\log\left(\frac{\left(\mathrm{VP}_1(A) + \mathrm{VP}_2(A)\right)^2}{\mathrm{Vol}(A)}\right)$$

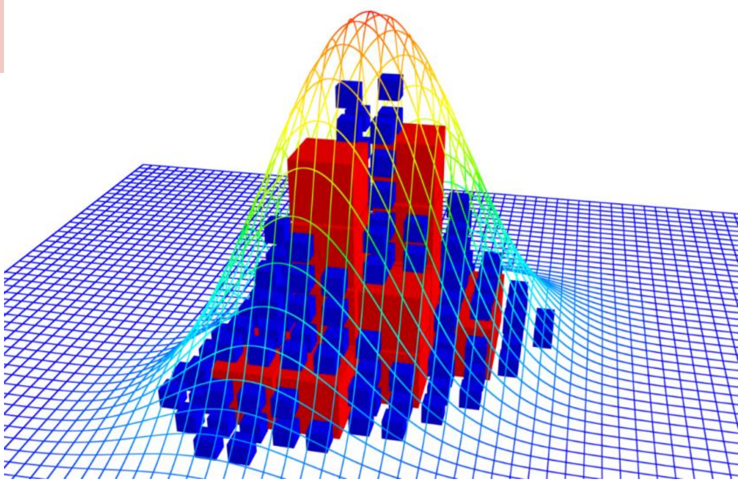
$\approx$ — After appropriate scaling

$$\log\left(\frac{\mathrm{VP}_1(A)\mathrm{VP}_2(A)}{\mathrm{Vol}(A)}\right)$$

$\approx$ — If $\mathrm{Unif}(A)$ log-concave $\Rightarrow A$ convex

$I(X_1; X_2)$ + const

Scheme for $(X_1, X_2) \sim f$



- Positive part of hypograph

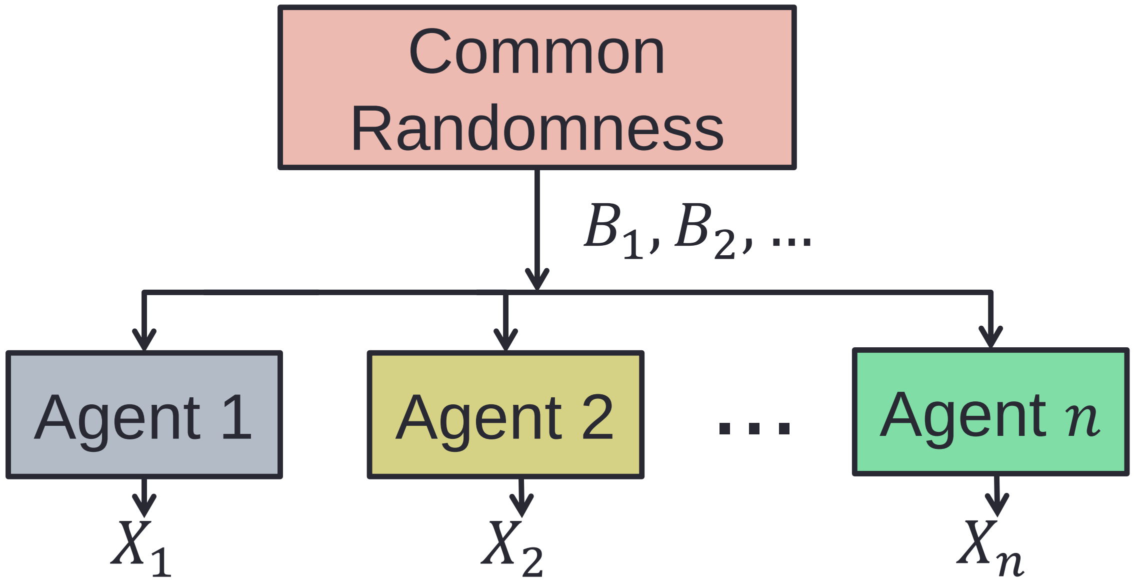
$$A = \{(x_1, x_2, z): 0 \leq z \leq f(x_1, x_2)\} \subseteq \mathbf{R}^3$$

- If we let $(X_1, X_2, Z) \sim \mathrm{Unif}(A)$, then $(X_1, X_2) \sim f$
- Apply dyadic decomposition for uniform case to $A \subseteq \mathbf{R}^3$

Theorem

$$I(X_1; X_2) \leq J(X_1; X_2) \leq G(X_1; X_2) \leq I(X_1; X_2) + 24$$

Generalization to n agents



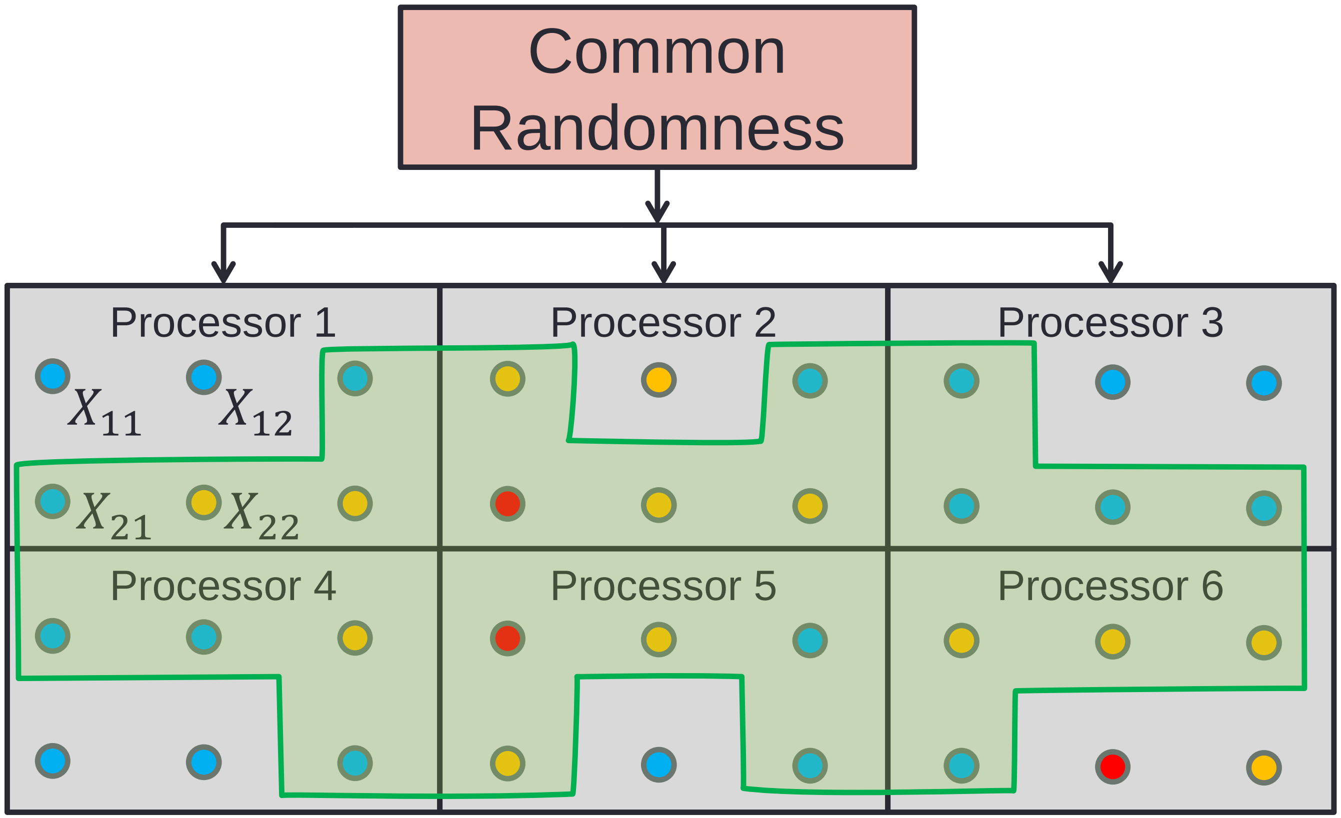
Theorem

$$I_D \leq J \leq G(X_1; \cdots; X_n) \leq I_D + n^2 \log e + 9n \log n$$

- I_D is Dual total correlation, a generalization of I

$$I_D(X_1; \cdots; X_n) = h(X_1, \dots, X_n) - \sum_{i=1}^{n} h(X_i | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$$

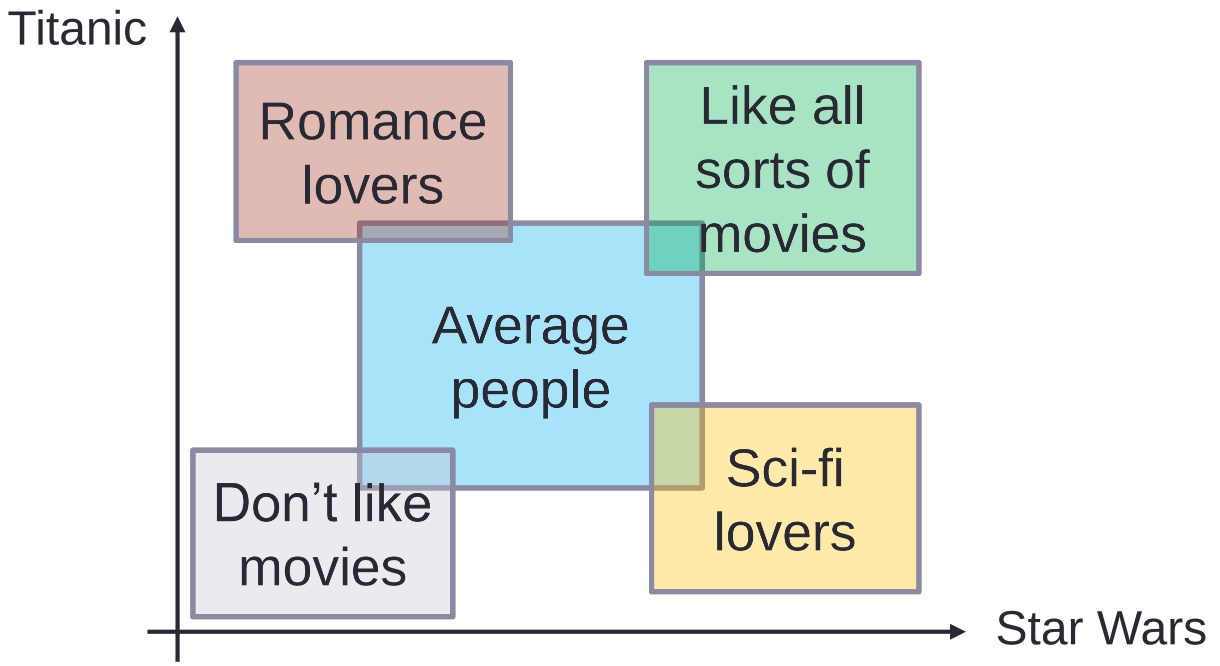
Distributed Computer Simulation



- Simulation of heat distribution - X_{ij} is a Markov random field
- Temperatures in blocks are dependent continuous RVs
- Distributed generation algorithm to generate boundary

Latent Variable Model

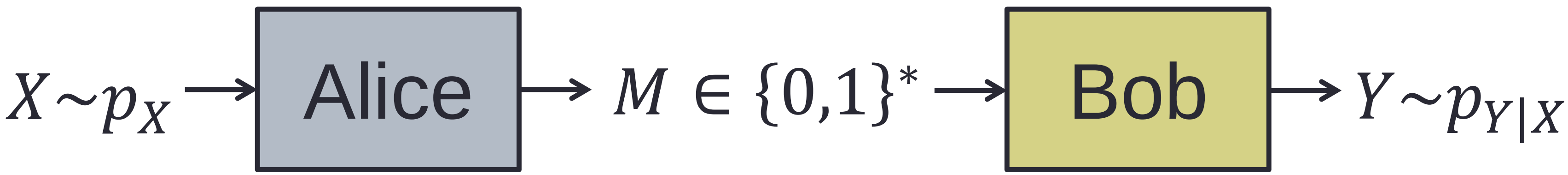
- Observed variables $X_1, \ldots, X_n$
- Latent variable W
- Within each class $W = w$, $X_1, \ldots, X_n$ are independent
- E.g. X_1 = user's score for *Star Wars*,
  X_2 = user's score for *Titanic*



Latent Variable Model

- Minimize number of classes (cardinality of W)
  - Nonnegative rank $= \min_{X_1 \perp \cdots \perp X_n | W} |\mathcal{W}|$
  - Nonnegative matrix/tensor factorization
  - If $X_1, \ldots, X_n$ continuous, cannot be exact in general
- Minimize entropy $H(W)$
  - $G(X_1; \ldots; X_n) = \min_{X_1 \perp \cdots \perp X_n | W} H(W)$
  - Can be exact even if $X_1, \ldots, X_n$ continuous

One-shot Channel Simulation



- Considered in Steiner 2000, Harsha et al. 2010
- Alice observes $X \sim p_X$, sends prefix-free codeword M
- Bob generates an instance $Y \sim p_{Y|X}$
- If M is the Huffman codeword of W:

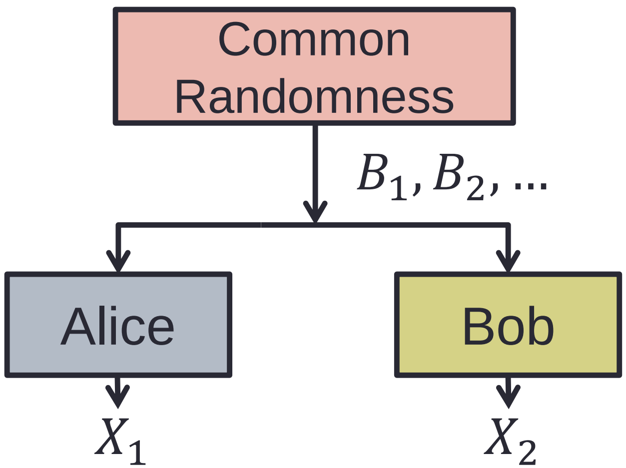
$$G(X;Y) = \min_{X-W-Y} H(W) \leq \min \mathrm{E}[L(M)] \leq G(X;Y) + 1$$

- If (X,Y) log-concave, then $\min \mathrm{E}[L(M)] \leq I(X;Y) + 25$

Summary

1. Distributed generation

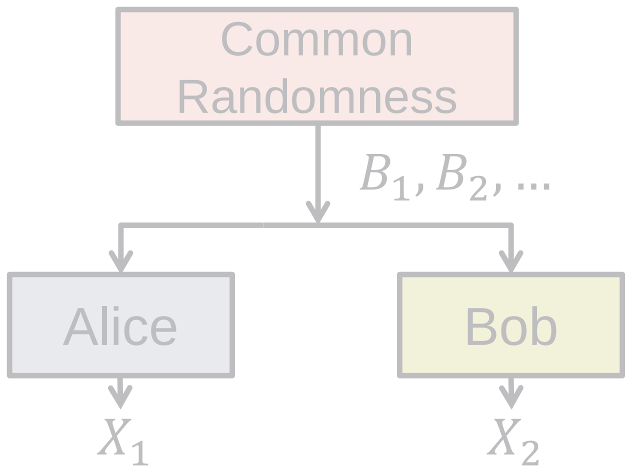
Avg #bits $\leq I(X_1; X_2) + 26$ for log-concave



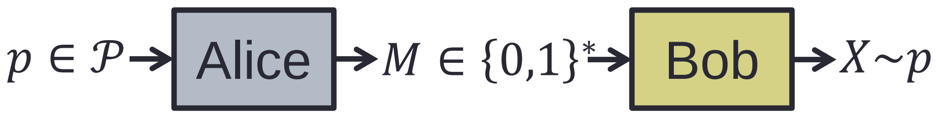
Outline

1. Distributed generation

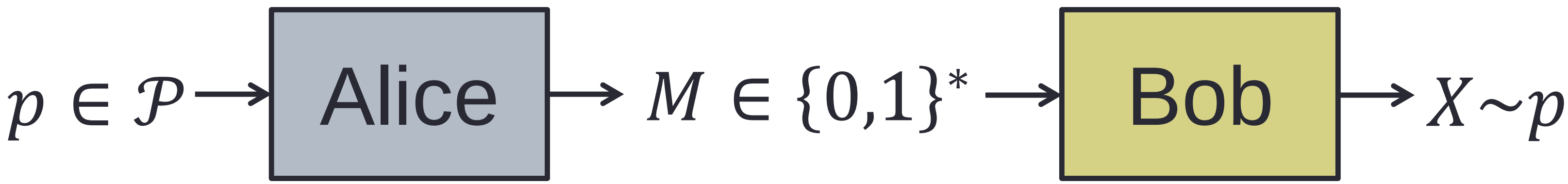
Avg #bits $\leq I(X_1; X_2) + 26$ for log-concave



2. Universal remote generation

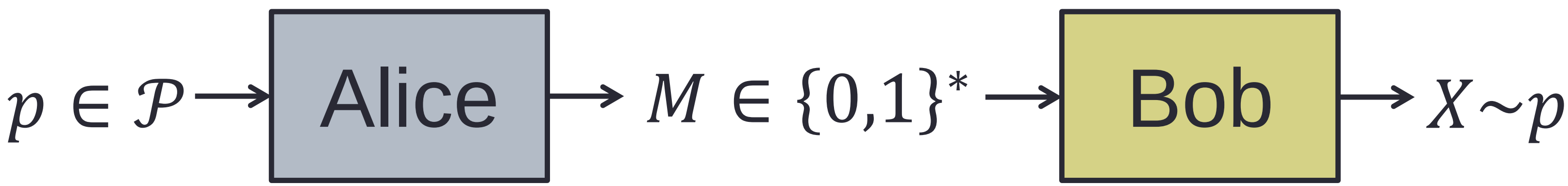


Remote Generation



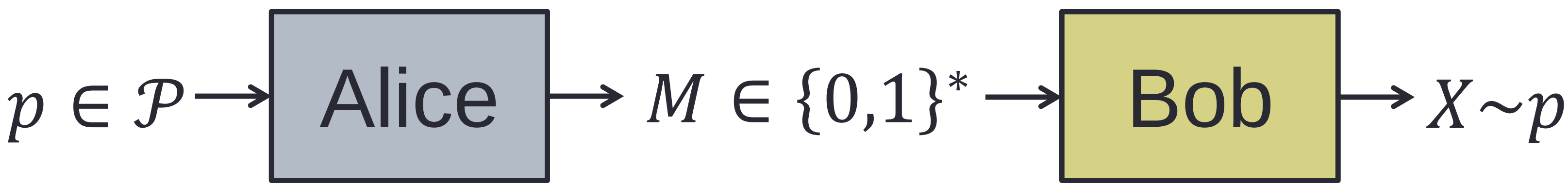
- Set of distributions $\mathcal{P}$ (over discrete/continuous)
- Alice observes arbitrary $p \in \mathcal{P}$,sends prefix-free codeword M
- Bob generates an instance $X \sim p$
- Find scheme with expected codeword length $\mathrm{E}_p[L(M)] < \infty$

Case 1: $\mathcal{P}$ = set of pmfs over integers



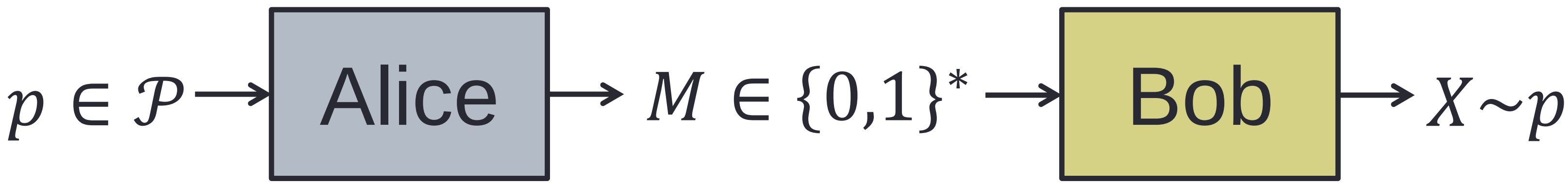
- Generate-compress
  1. Alice generates $X \sim p$
  2. Alice encodes X using universal code over integers
  3. Bob recovers X from M
- Stochastic encoder, deterministic decoder

Case 2: $|\mathcal{P}|$ finite / countable



- Compress-generate
  1. Alice encodes p using universal code over integers
  2. Bob recovers p from M
  3. Bob generates $X \sim p$
- Deterministic encoder, stochastic decoder

$\mathcal{P}$ = all continuous distributions

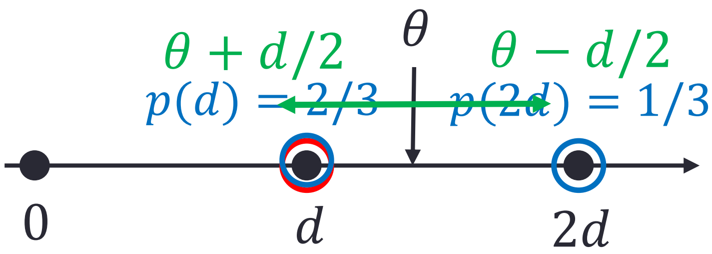


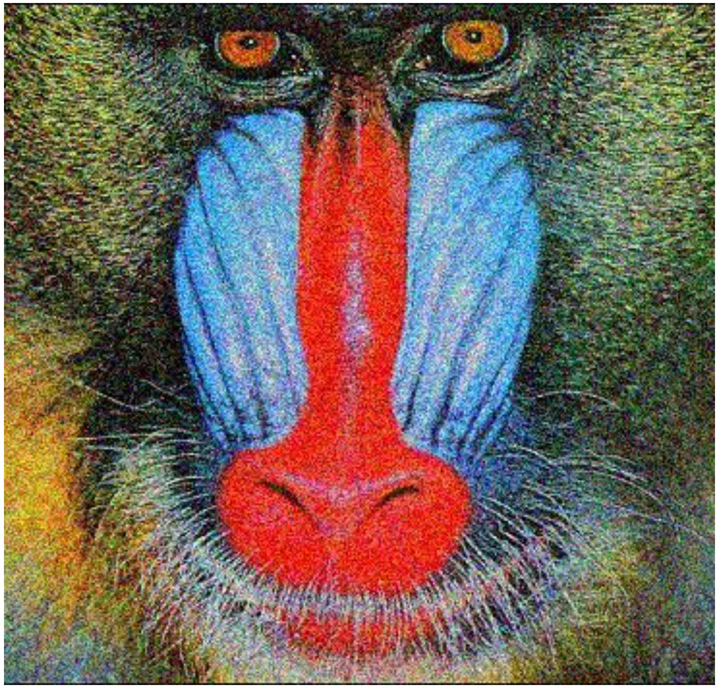
- Support not countable, $\mathcal{P}$ not countable
- We devise universal scheme
- Uses both stochastic encoder and decoder

Application 1: Lossy Compression / Dither

- How to compress θ?
- Quantize to the nearest multiple of d
- Dithering: quantize at random such that mean is θ
- Remote generation: generate

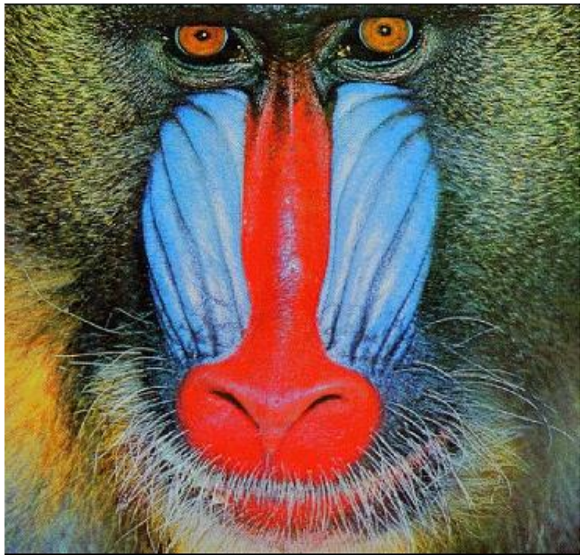
$$X \sim \text{Unif}(\theta - d/2, \theta + d/2)$$





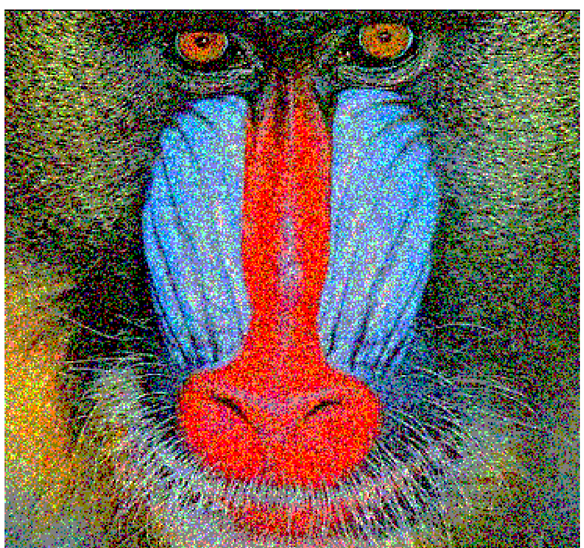
Application 1: Lossy Compression / Dither

- Original

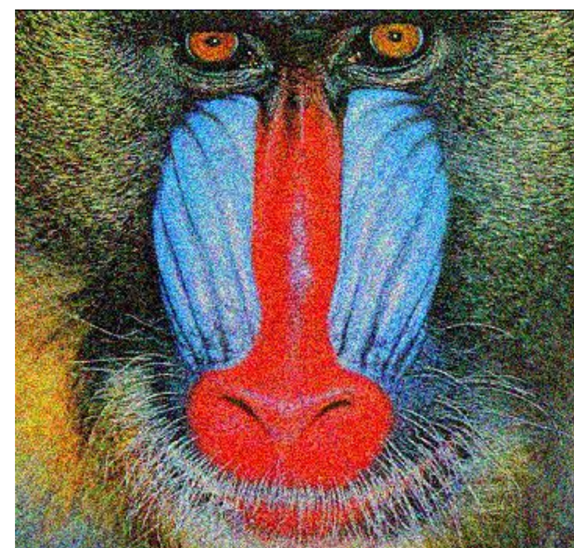


- Quantize

- Dither



- Remote generation



Application 2: Simulation of Bell State

- Pair of qubits $|\Phi^+\rangle = (|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B)/\sqrt{2}$
- Alice measures in direction θ_A, Bob in θ_B
- $Y_A, Y_B \in \{\pm 1\}$, $p_{Y_A}(1) = p_{Y_B}(1) = \frac{1}{2}$, $\mathrm{E}[Y_A Y_B] = -\cos(\theta_A - \theta_B)$
- Alice sends codeword M to Bob to simulate Bell state
- Let $X \in [0,2\pi]$, $f(x|y_A;\theta_A) = \frac{1}{2}\max\{\cos(y_A(x-\theta_A)), 0\}$,

  $Y_B = -\mathrm{sgn}(\cos(X - \theta_B))$

Application 3: Mixed Strategy with Helper

- Payoff $g(X,\theta)$ depends on decision X and unknown θ
- Minimax strategy – random X to maximize $\inf_{\theta} \mathrm{E}[g(X,\theta)]$
- $\theta = (\theta_1, \theta_2)$, Alice knows θ_1, sends W to Bob to generate X
- Use scheme to remote generate $\mathrm{argmax}_{f_X} \inf_{\theta_2} \mathrm{E}[g(X,\theta_1,\theta_2)]$
- E.g. $g(x,\theta) = e^{2\theta - x}$ if $x \geq \theta$, $g(x,\theta) = 0$ otherwise
  - Alice knows $\theta \geq a$, optimal strategy $f_X(x;a) = e^{-(x-a)}$, $x \geq a$

Main Result – Bounded Support

- For quasiconcave distributions over $[0,1]^n$:

Theorem [Li-El Gamal 2016]

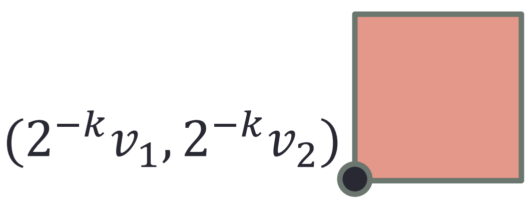
$$\mathrm{E}_p[L(M)] \le n(\log(\sup f(x)) + \log n + \log e + 2)$$
$$+2\log(\log(\sup f(x)) + \log n + \log e + 3) + 1$$

- Can extend to $\mathcal{P}$ = continuous distributions over $\mathbf{R}^n$
- Simulation of Bell state:

  $\mathrm{E}[L(M)] \le 8.96$, compared to 20 in [Massar et.al. 2001]

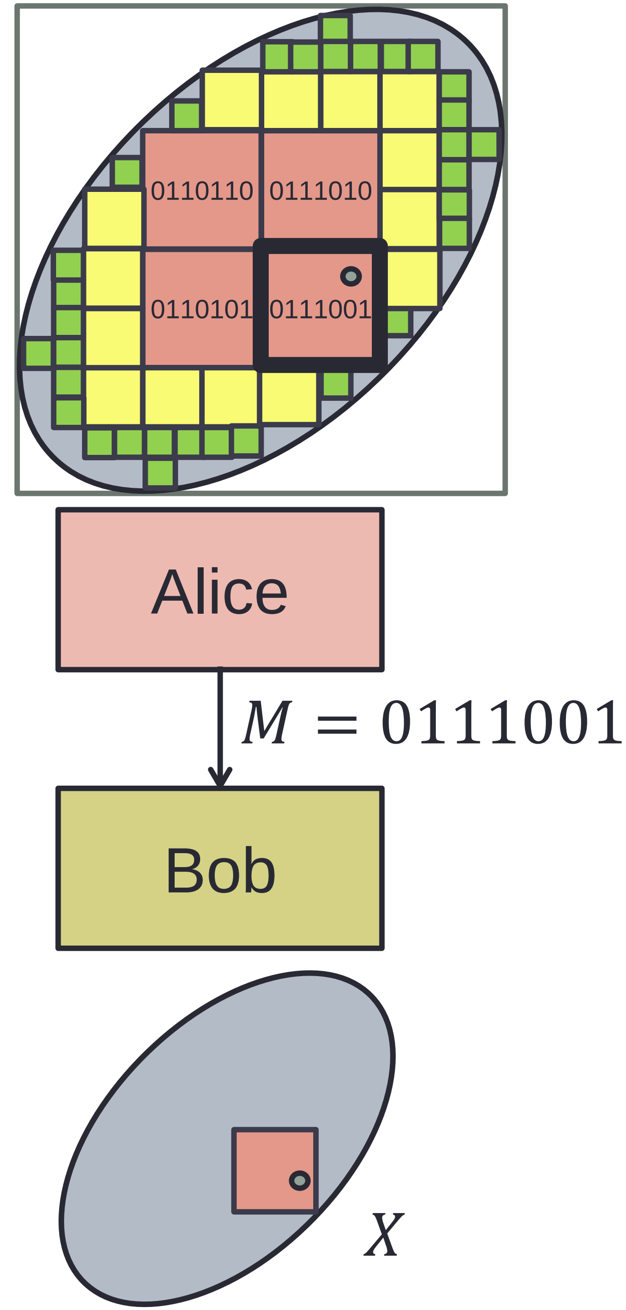
Scheme for $(X_1, X_2) \sim \mathrm{Unif}(A)$

- Dyadic decomposition of A
- Alice
  - Generate random point in A
  - Find square containing point:



$k, v_1, v_2 \in \mathbf{Z}$

  - Encode k, v_1, v_2 into M and send
- Bob
  - Recover the square
  - Generate X uniformly over square



Encoding the Dyadic Squares

- Elias gamma code [Elias 1975] of $a \geq 1$
  - Let i be the number of bits in the binary representation of a
  - Code for a: $i - 1$ zeros followed by binary representation of a
  - E.g. $a = 9$, binary representation is 1001, code is 0001001
  - Codeword length $\leq 2\log(a) + 1$

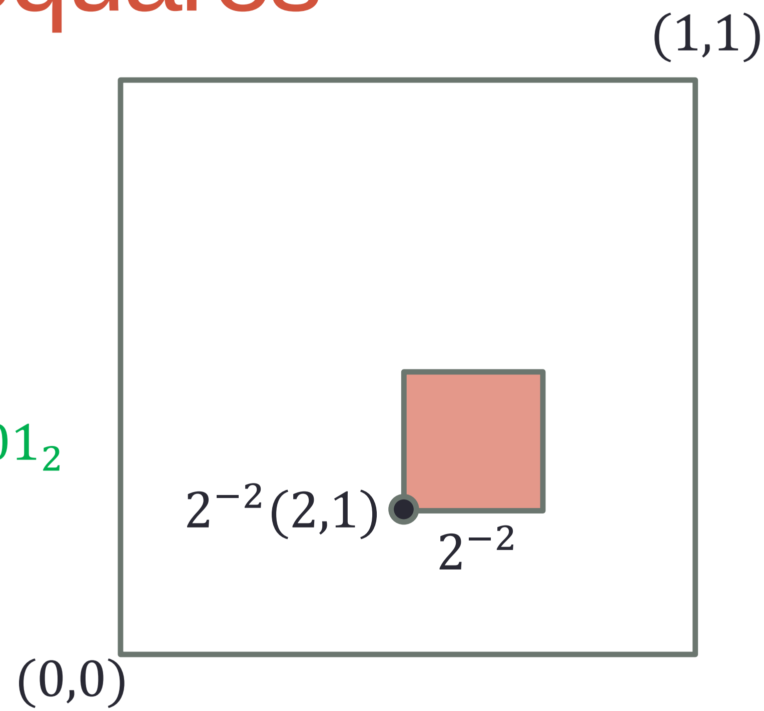
Encoding the Dyadic Squares

- E.g. $n = 2$, $k = 2$, $v = (2, 1)$
- $M = 0111001$

k -bit binary of $v_2 = 1 = 01_2$

k -bit binary of $v_1 = 2 = 10_2$

Elias gamma code of $k + 1$
$k + 1 = 3 = 11_2$



- $L(M) \leq nk + 2\log(k+1) + 1$
- $\mathrm{E}[L(M)] \leq n\mathrm{E}[k] + 2\log(\mathrm{E}[k] + 1) + 1$
- $\mathrm{E}[k] = \mathrm{E}\left[-\log L_{\mathrm{dy}}\right] \leq h_{\ominus B}(A) + 2$

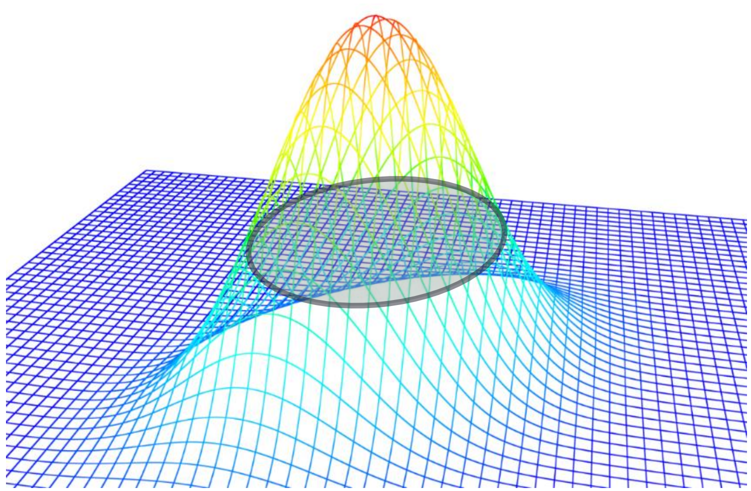
Scheme for $(X_1, X_2) \sim f$

- Positive part of hypograph

$$\mathrm{hyp}_+(f) = \{(x, z): 0 \le z \le f(x)\} \subseteq \mathbf{R}^{n+1}$$

- If we let $(X, Z) \sim \mathrm{Unif}(\mathrm{hyp}_+(f))$, then $X \sim f$
- Alice generate Z, apply uniform scheme on

$$L_z^+(f) = \{x: f(x) \ge z\}$$



Comparing Schemes

	Distributed Generation	Remote Generation
Reason for using dyadic decomposition	X_1, X_2 conditionally indep. given the square containing (X_1, X_2)	All continuous distributions can be expressed as mixtures of uniform distributions over dyadic squares

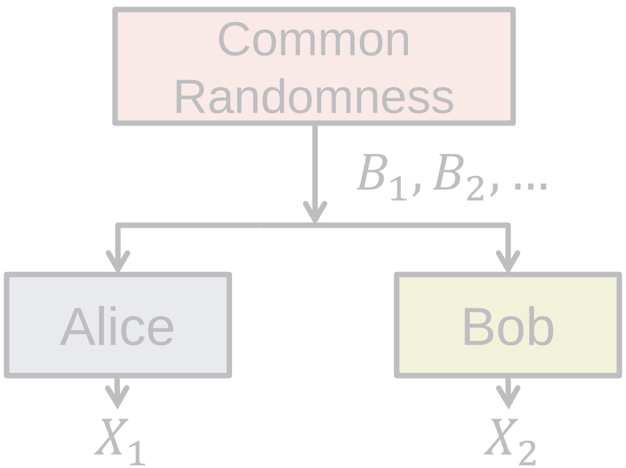
Comparing Schemes

	Distributed Generation	Remote Generation
Reason for using dyadic decomposition	X_1, X_2 conditionally indep. given the square containing (X_1, X_2)	All continuous distributions can be expressed as mixtures of uniform distributions over dyadic squares
Representing dyadic squares	Knuth-Yao	Universal code

Summary

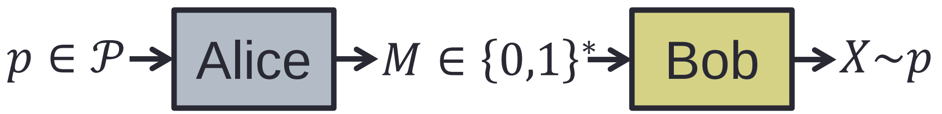
1. Distributed generation

Avg #bits $\leq I(X_1; X_2) + 26$ for log-concave



2. Universal remote generation

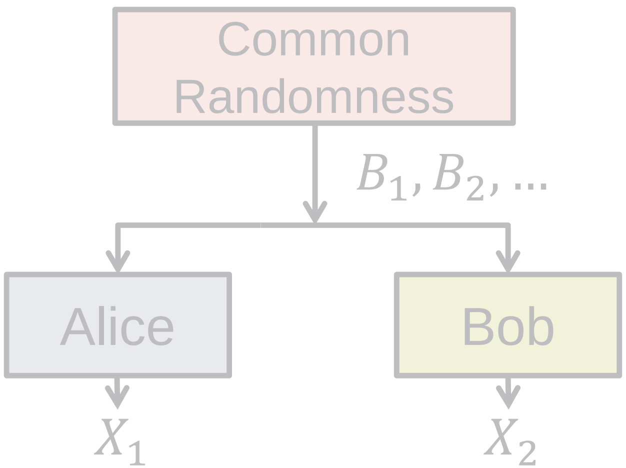
Scheme for any continuous distribution p



Outline

1. Distributed generation

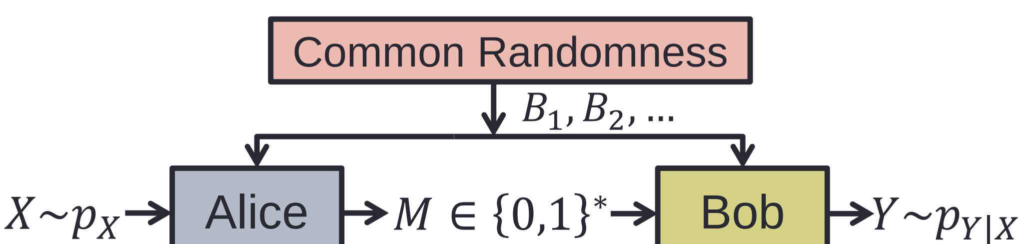
Avg #bits $\leq I(X_1; X_2) + 26$ for log-concave



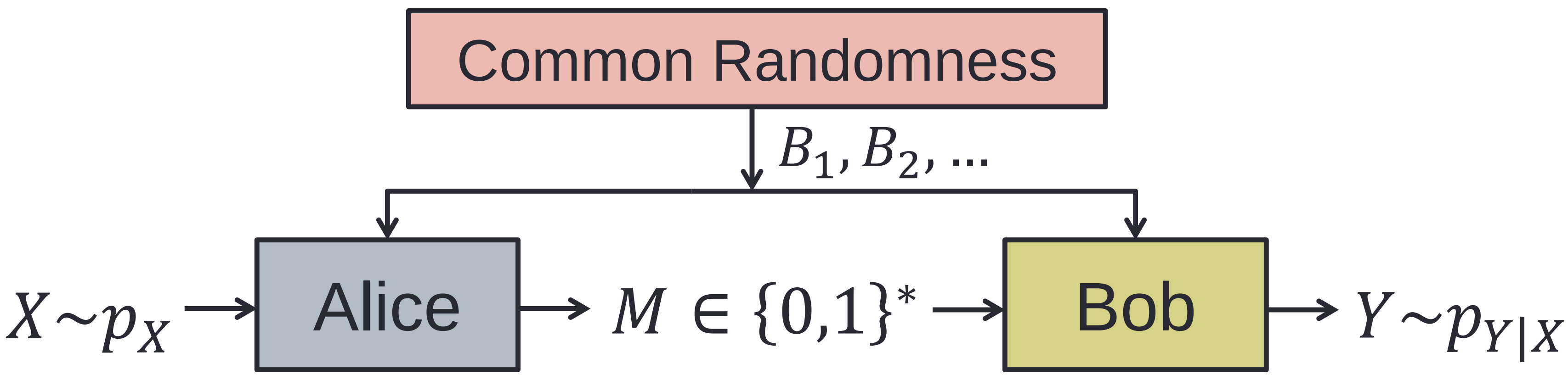
2. Universal remote generation

Scheme for any continuous distribution p

3. Channel simulation with Common Randomness



3. Channel Simulation with Common Randomness



- Unlimited common randomness $B_1, B_2, \ldots$
- Harsha et. al. (2010) showed that for discrete X, Y

$$I(X;Y) \leq \min \mathrm{E}[L(M)] \leq I(X;Y) + 2\log(I(X;Y) + 1) + c$$

- Rejection sampling
- We strengthen it to general X, Y

$$\min \mathrm{E}[L(M)] \leq I(X;Y) + \log(I(X;Y) + 1) + 5$$

Strong Functional Representation Lemma

Functional representation lemma:
For any X, Y, there exists Z indep. of X such that Y is a function of X, Z

- Applications in multi-user information theory
  - Broadcast channel [Hajek-Pursley 1979]
  - Multiple access channel with cribbing encoders [Willems-van der Meulen 1985]

Strong Functional Representation Lemma

Strong functional representation lemma [Li-El Gamal 2017]: For any X, Y, there exists Z indep. of X such that Y is a function of X, Z,

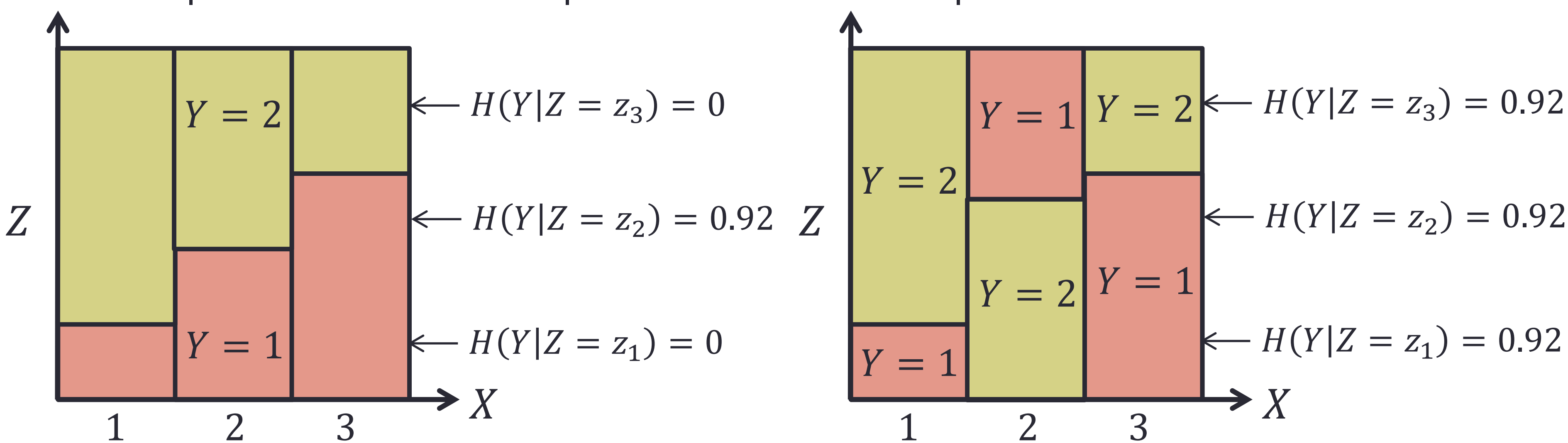
$$H(Y|Z) \leq I(X;Y) + \log(I(X;Y) + 1) + 4$$

- Applications
  - Tighter bound for channel simulation with common randomness
  - One-shot lossy source coding
  - Simple proof for Gelfand-Pinsker Theorem
  - Other coding theorems
- Exists examples where SFRL is tight within 5 bits

Exponential Functional Representation

- E.g. : $Y \in \{1,2\}$, $X \sim \mathrm{Unif}\{1,2,3\}$,

$$p_{Y|X}(1,1) = 0.2,\ p_{Y|X}(1,2) = 0.4\ ,\ p_{Y|X}(1,3) = 0.6$$



- In general: $Y \in \{1, \dots, k\}$, $Z_1, \dots, Z_k$ i.i.d. $\mathrm{Exp}(1)$,

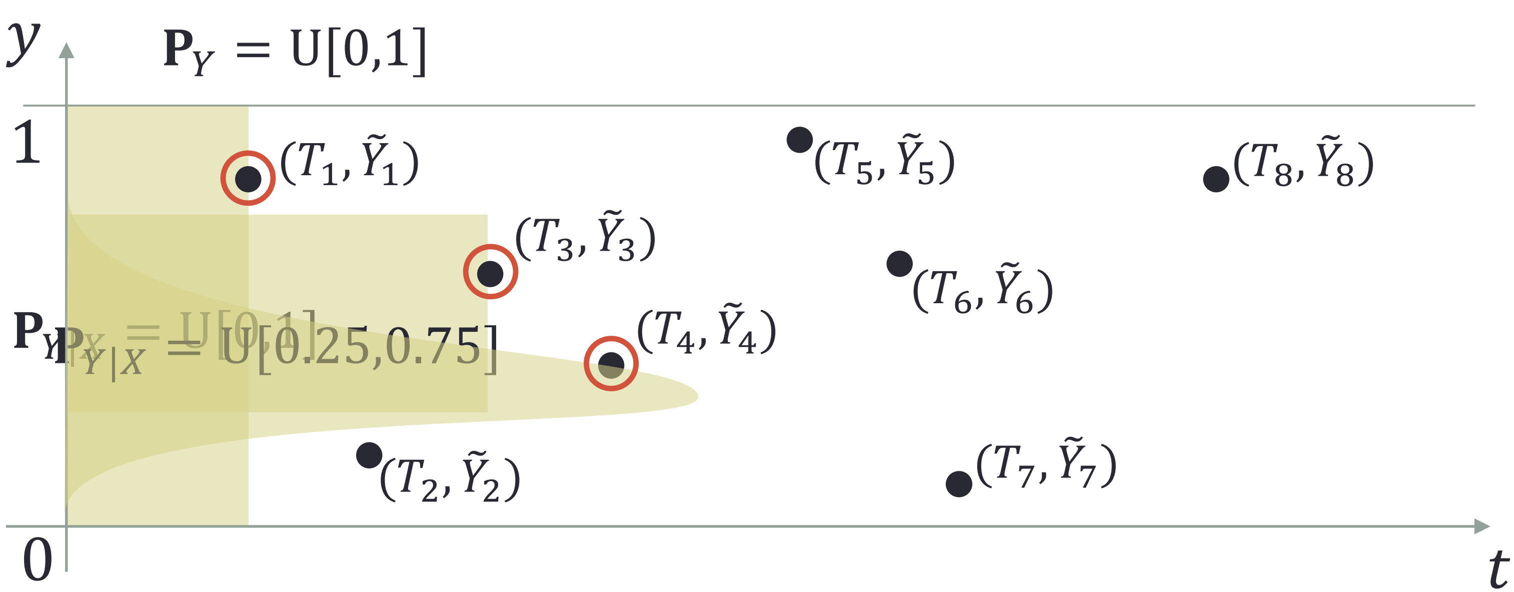
$$Y = \arg\min_y \frac{Z_y}{p_{Y|X}(y|X)}$$

Poisson Functional Representation

- Poisson process $0 \le T_1 \le T_2 \le \cdots$ ($T_i - T_{i-1}$ i.i.d. $\mathrm{Exp}(1)$)
- Marks $\tilde{Y}_1, \tilde{Y}_2, \ldots$ i.i.d. $\mathbf{P}_Y$, take $Z = \{T_i, Y_i\}$

$$K(X,Z) = \arg\min_i T_i \cdot \frac{d\mathbf{P}_Y}{d\mathbf{P}_{Y|X}(\cdot\,|X)}(\tilde{Y}_i), \qquad Y(X,Z) = \tilde{Y}_{K(X,Z)}$$

- E.g. $\mathbf{P}_Y = \mathrm{U}[0,1]$, $K = \arg\min_i \frac{T_i}{f_{Y|X}(\tilde{Y}_i|X)}$



Proof of SFRL

- Poisson process $0 \le T_1 \le T_2 \le \cdots, \tilde{Y}_1, \tilde{Y}_2, \ldots$ i.i.d. $\mathbf{P}_Y$

$$K = \arg\min_i T_i \cdot \frac{d\mathbf{P}_Y}{d\mathbf{P}_{Y|X}(\cdot\,|X)}\left(\tilde{Y}_i\right), \qquad Y = \tilde{Y}_K$$

- Can show $\mathrm{E}[\log K\,|X = x] \le D(\mathbf{P}_{Y|X}(\cdot\,|x)\|\mathbf{P}_Y) + 1.54$
- $\mathrm{E}[\log K] \le I(X;Y) + 1.54$
- By max entropy distribution for fixed $\mathrm{E}[\log K]$,

$$H(K) \le \mathrm{E}[\log K] + \log(\mathrm{E}[\log K] + 1) + 1$$

- Since Y is a function of $Z = \{\tilde{Y}_i, T_i\}$ and K, $H(Y|Z) \le H(K)$

One-shot Variable-length Lossy Source Coding

- Encode source $X \sim p_X$ into prefix-free $M \in \{0,1\}^*$
- Decode M to recover Y with distortion $d(X,Y) \geq 0$
- Trade-off avg length $\bar{R} = \mathrm{E}[L(M)]$, avg distortion $\mathrm{E}[d(X,Y)] \leq D$

Theorem [Li-El Gamal 2017]

$(\bar{R}, D)$ achievable if

$$\bar{R} > R(D) + \log(R(D) + 1) + 6,$$

where $R(D) = \min_{\mathrm{E}[d(X,Y)] \leq D} I(X;Y)$ is rate-distortion function

- Posner-Rodemich 1971: epsilon entropy
- Kostina-Polyanskiy-Verdu 2015: consider $\mathrm{P}(d(X,Y) \geq D)$

Proof

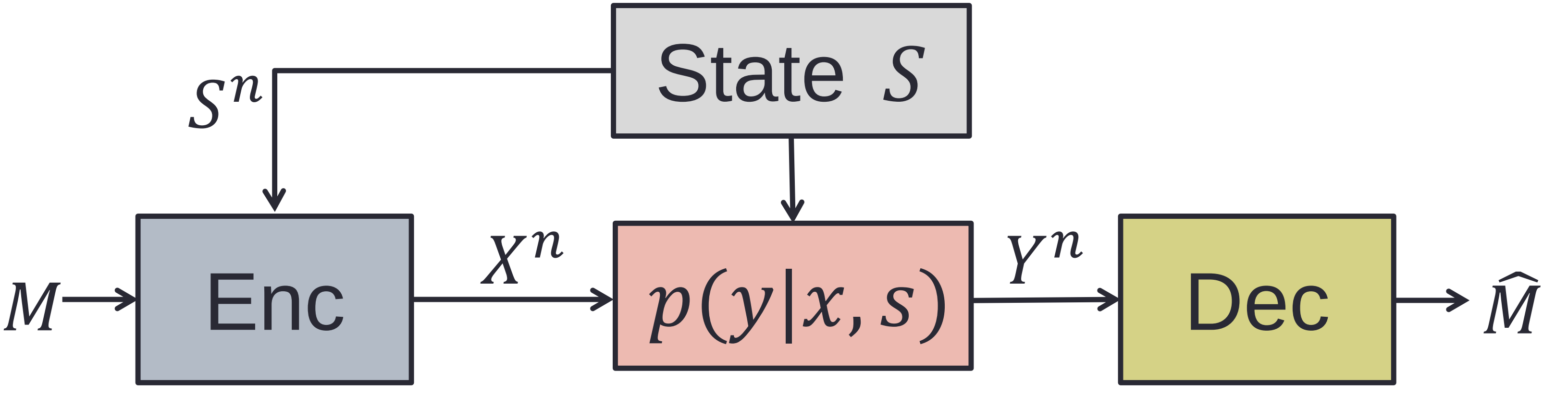
- Let Y attain $R(D) = \min_{\mathrm{E}[d(X,Y)] \leq D} I(X;Y)$
- SFRL: there exists Z indep. of X, and Y is a fcn of X, Z,

$$H(Y|Z) \leq R(D) + \log(R(D) + 1) + 4$$

- Find z with small $H(Y|Z = z)$ (avg length of Huffman code) and small avg distortion $\mathrm{E}[d(X,Y)|Z = z] \leq D$
- Carathéodory theorem: $\tilde{Z}$ mixture of z_1, z_2 can give small avg length and distortion

$$H\left(Y\middle|\tilde{Z}\right) \leq R(D) + \log(R(D) + 1) + 5$$

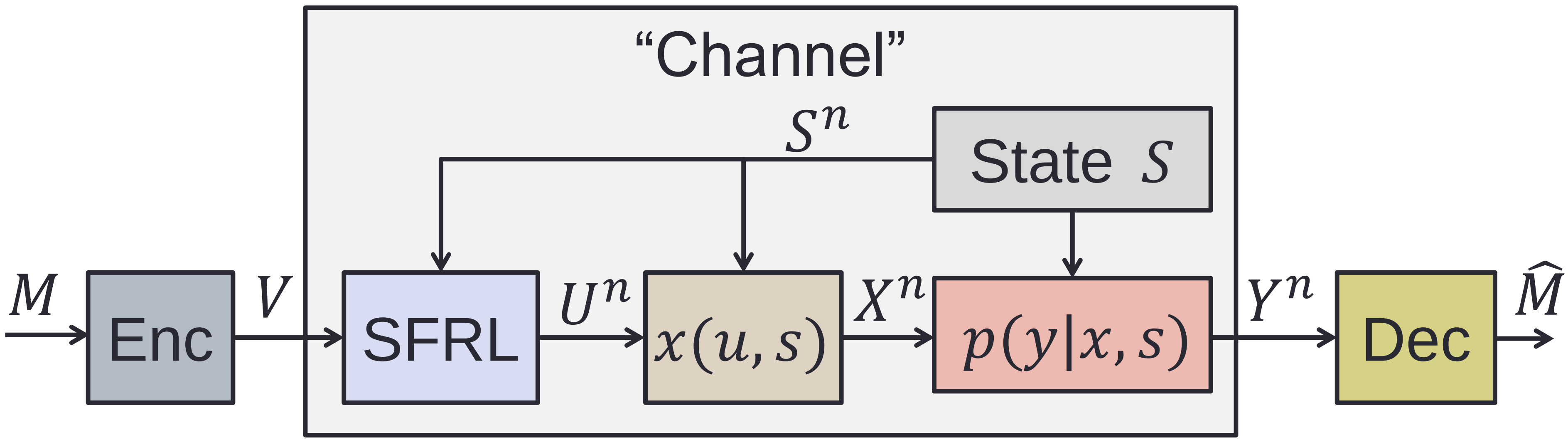
Gelfand-Pinsker Theorem



- State noncausally available at encoder

$$C = \max_{p_{U|S,X}(u,s)} \left(I(U;Y) - I(U;S) \right)$$

Gelfand-Pinsker Theorem

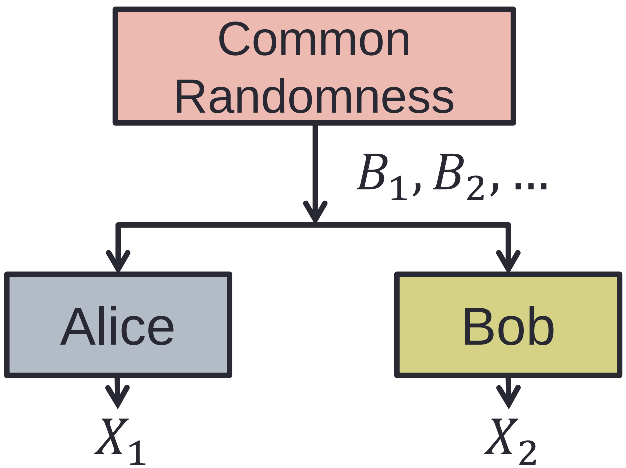


- Let U, $x(u,s)$ attain $C = \max_{p_{U|S}, x(u,s)} (I(U;Y) - I(U;S))$
- SFRL: there exists V indep. of S^n, and U^n is a fcn of S^n, V,
$$H(U^n|V) = nI(U;S) + o(n)$$
- $I(V;Y^n) \geq I(U^n;Y^n) - H(U^n|V) = nC - o(n)$
- Treat $V \to Y^n$ as channel and apply channel coding

Conclusion

1. Distributed generation

Avg #bits $\leq I(X_1;X_2) + 26$ for log-concave



2. Universal remote generation

Scheme for any continuous distribution p

3. Channel simulation with Common Randomness

$\mathrm{E}[L(M)] \leq I(X;Y) + \log(I(X;Y) + 1) + 5$

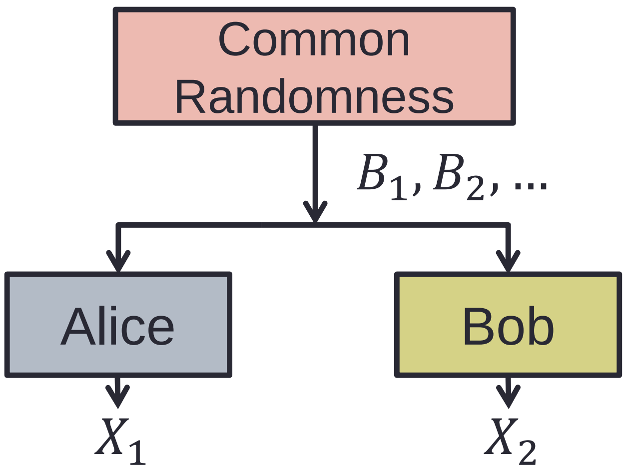
Strong functional representation lemma

- One-shot lossy source coding
  $\bar{R} > R(D) + \log(R(D) + 1) + 6$
- Simple proof for Gelfand-Pinsker theorem

Conclusion

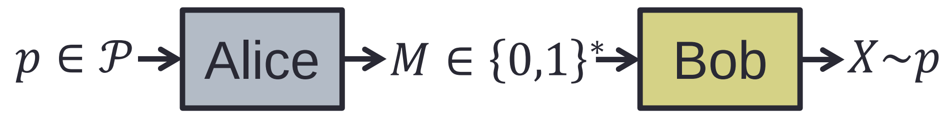
1. Distributed generation

Avg #bits $\leq I(X_1; X_2) + 26$ for log-concave



2. Universal remote generation

Scheme for any continuous distribution p



3. Channel simulation with Common Randomness

$\mathrm{E}[L(M)] \leq I(X;Y) + \log(I(X;Y)+1) + 5$

Strong functional representation lemma

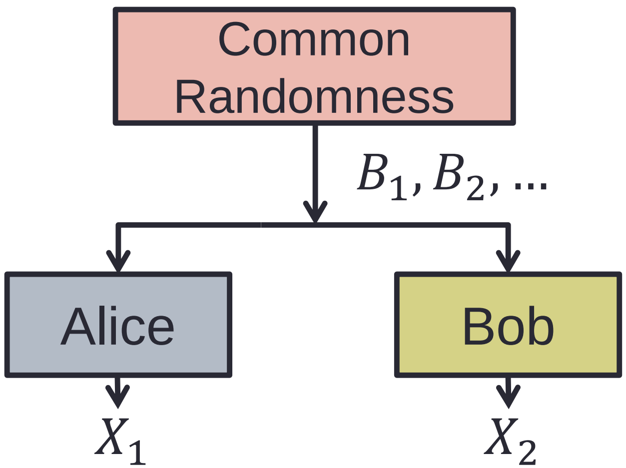
- One-shot lossy source coding

  $\bar{R} > R(D) + \log(R(D)+1) + 6$
- Simple proof for Gelfand-Pinsker theorem

Conclusion

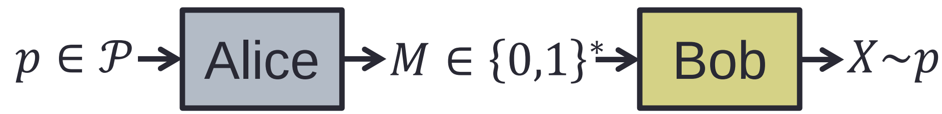
1. Distributed generation

Avg #bits $\leq I(X_1; X_2) + 26$ for log-concave



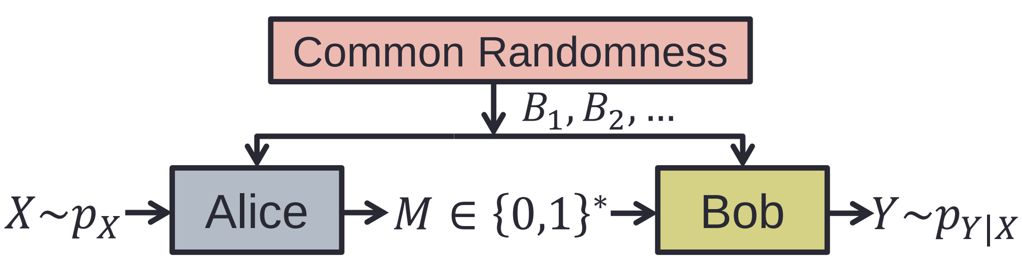
2. Universal remote generation

Scheme for any continuous distribution p



3. Channel simulation with Common Randomness

$\mathrm{E}[L(M)] \leq I(X;Y) + \log(I(X;Y) + 1) + 5$



Strong functional representation lemma

- One-shot lossy source coding

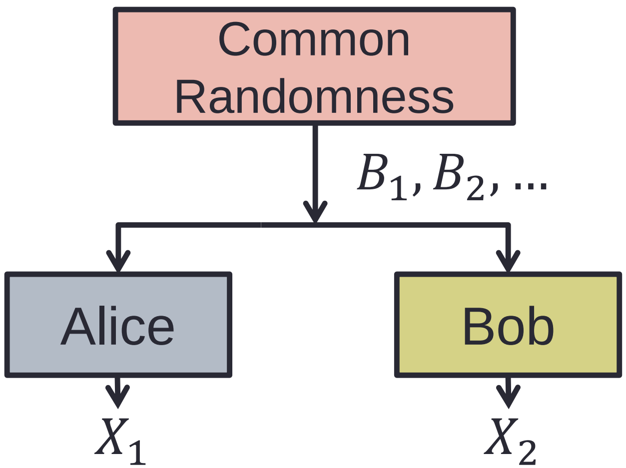
  $\bar{R} > R(D) + \log(R(D) + 1) + 6$

- Simple proof for Gelfand-Pinsker theorem

Conclusion

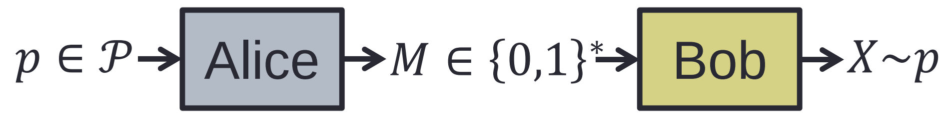
1. Distributed generation

Avg #bits $\le I(X_1; X_2) + 26$ for log-concave



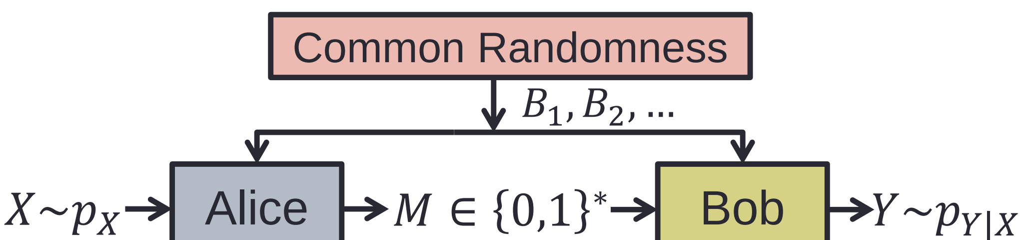
2. Universal remote generation

Scheme for any continuous distribution p



3. Channel simulation with Common Randomness

$\mathrm{E}[L(M)] \le I(X;Y) + \log(I(X;Y) + 1) + 5$



Strong functional representation lemma

- One-shot lossy source coding

  $\bar{R} > R(D) + \log(R(D) + 1) + 6$
- Simple proof for Gelfand-Pinsker theorem

References

- G. R. Kumar, C. T. Li, and A. El Gamal, “Exact common information,” in Proc. IEEE Int. Symp. Info. Theory 2014, pp. 161–165.
- C. T. Li and A. El Gamal, “Distributed simulation of continuous random variables,” in Proc. IEEE Int. Symp. Info. Theory 2016, pp. 565–569.
- C. T. Li and A. El Gamal, “A universal coding scheme for remote generation of continuous random variables,” Info. Theory Workshop 2016.
- C. T. Li and A. El Gamal, “Strong functional representation lemma and applications to coding theorems,” IEEE Int. Symp. Info. Theory 2017.
- C. T. Li and A. El Gamal, “Extended Gray-Wyner System with Complementary Causal Side Information,” IEEE Int. Symp. Info. Theory 2017.
- C. T. Li and A. El Gamal, “An efficient feedback coding scheme with low error probability for discrete memoryless channels,” IEEE Trans. Info. Theory, vol. 61, no.6, pp. 2953-2963, 2015.
- A. D. Wyner, “The common information of two dependent random variables,” IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 163–179, Mar. 1975.
- P. Cuff, H. Permuter, and T. M. Cover, “Coordination capacity,” IEEE Trans. Info. Theory, vol. 56, no. 9, pp. 4181–4206, 2010.
- T. S. Han, “Nonnegative entropy measures of multivariate symmetric correlations,” Information and Control, vol. 36, pp. 133–156, 1978.
- P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, “The communication complexity of correlation,” IEEE Trans. Info. Theory, vol. 56, no. 1, pp. 438–449, Jan 2010.
- B. Hajek and M. Pursley, “Evaluation of an achievable rate region for the broadcast channel,” IEEE Transactions on Information Theory, vol. 25, no. 1, pp. 36–46, Jan 1979.
- F. Willems and E. van der Meulen, “The discrete memoryless multiple-access channel with cribbing encoders,” IEEE Transactions on Information Theory, vol. 31, no. 3, pp. 313–327, May 1985.
- E. C. Posner and E. R. Rodemich, “Epsilon entropy and data compression,” The Annals of Mathematical Statistics, pp. 2079–2125, 1971.
- V. Kostina, Y. Polyanskiy, and S. Verdú, “Variable-length compression allowing errors,” IEEE Transactions on Information Theory, vol. 61, no. 8, pp. 4316–4330, 2015.
- S. I. Gelfand and M. S. Pinsker, “Coding for channel with random parameters,” Probl. Control Inf. Theory, vol. 9, no. 1, pp. 19–31, 1980.