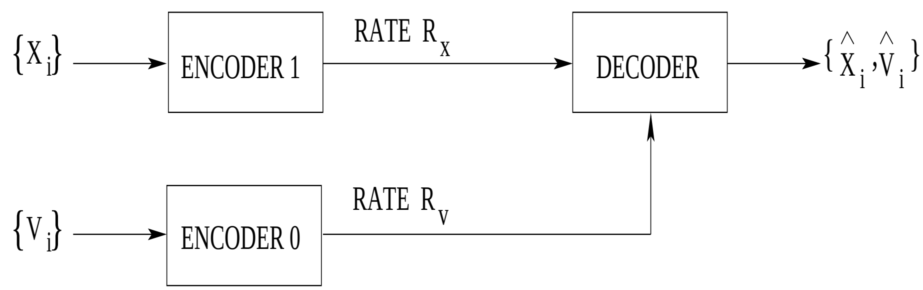
Source coding with side information at the decoder (Wyner-Ziv coding)

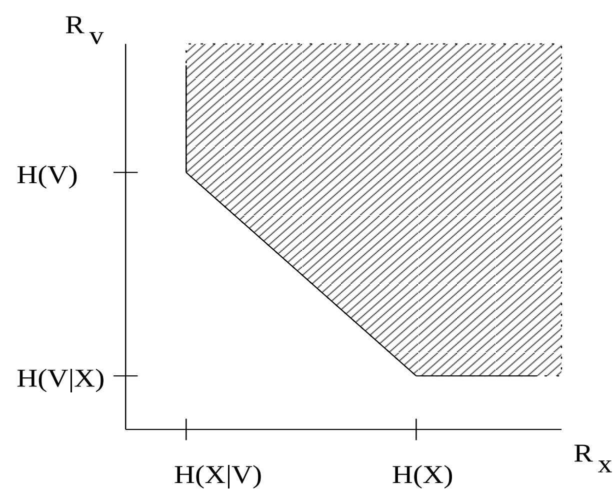
Feb 13, 2003

SLEPIAN-WOLF RESULT



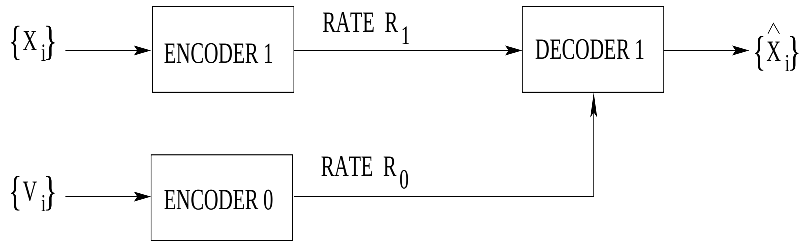
Problem: Determine $\mathcal{R}$, the set of all achievable rate pairs.

Result:



- $\mathcal{R} = \{(R_x, R_v) \mid R_x \geq H(X|V),\ R_v \geq H(V|X),\ R_x + R_v \geq H(X,V)\}$

Source coding with side information



- Goal: Design encoders E_0, E_1 to encode X with the optimal rate, and decoder D_1 to decode $\hat{X}$ with arbitrarily low probability of error.
- $\{V_k\}$ is statistically dependent on $\{X_k\}$.
- Rate of E_i is R_i bits/source-symbol.
- A rate pair (R_0, R_1) is achievable if ∃ encoders E_0, E_1 (with parameters R_0, R_1) and a decoder D_1 that can reproduce X with arbitrarily high reliability – i.e., if ∃ mappings

Encoder 0: $$f_0 : \mathcal{V}^n \to \{0, 1, \ldots, 2^{nR_0} - 1\},$$

Encoder 1: $$f_1 : \mathcal{X}^n \to \{0, 1, \ldots, 2^{nR_1} - 1\},$$

Decoder 1: $$g_1 : \{0, 1, \ldots, 2^{nR_0} - 1\} \times \{0, 1, \ldots, 2^{nR_1} - 1\} \to \mathcal{X}^n$$

such that the reconstruction error $\Delta = \frac{1}{n} E[d_H(\mathbf{X}^n, \hat{\mathbf{X}}^n)] \le \epsilon$,

- Problem: Determine $\mathcal{R}$, the set of all achievable rate pairs.
- Result: If $Q(x,v) = Pr\{X = x, V = v\}$, then (R_0, R_1) is achievable if and only if:

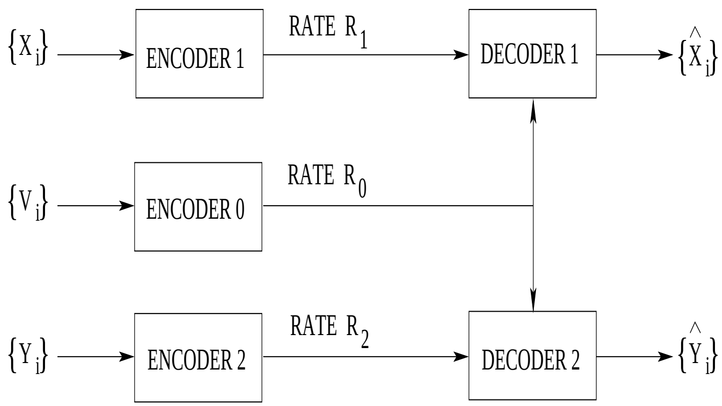
$$R_1 \geq H(X|W) \text{ and } R_0 \geq I(V;W)$$

for some auxiliary random variable W satisfying

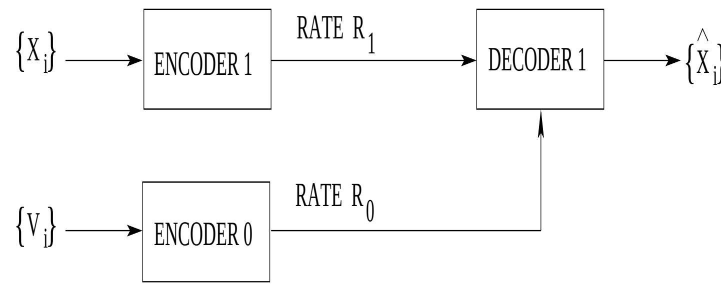
$$\text{(a)} \sum_{w \in \mathcal{W}} p(x,v,w) = Q(x,v), \quad \text{(b)}\ p(x,v,w) = Q(x,v)p_t(w|v)$$

[A. D. Wyner, "On source coding with side information at the decoder," *IEEE Transactions on Information Theory*, May, 1975.]

ZERO-ERROR SOURCE CODING WITH SIDE INFORMATION



$Y \equiv 0 \Rightarrow$



- Result: If $Q(x,y,v) = Pr\{X = x, Y = y, V = v\}$, then (R_0, R_1, R_2) is achievable if and only if: $R_1 \geq H(X|W), R_2 \geq H(Y|W),$ and $R_0 \geq I(V;W)$

  for some auxiliary random variable W satisfying

  (a) $\sum_{w \in \mathcal{W}} p(x,y,v,w) = Q(x,y,v),$ (b) $p(x,y,v,w) = Q(x,y,v)p_t(w|v)$

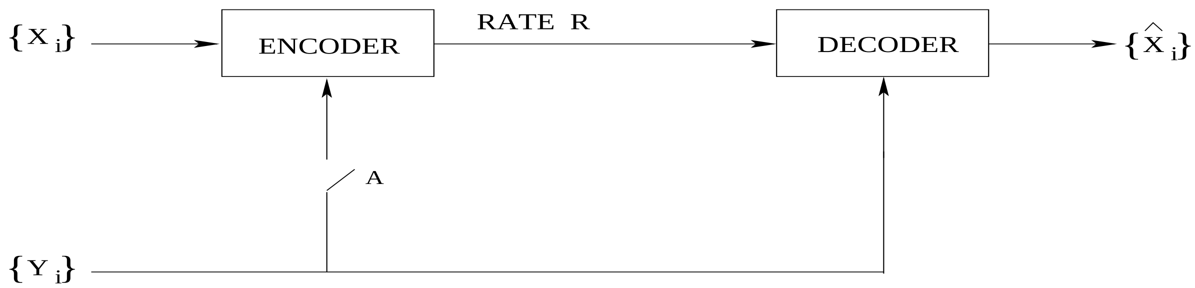
- Let $\mathcal{P}$ be the family of probability mass functions $p(x,y,v,w)$ satisfying (a) and (b).

- The set of achievable rate triplets is:

$$\mathcal{R} = \left\{ (R_0, R_1, R_2) \ : \ R_0 \geq I(V;W), R_1 \geq H(X|W), R_2 \geq H(Y|W), p \in \mathcal{P} \right\}^{[c]}$$

$^{[c]}$ denotes closure.

Lossy source coding with side information



- Goal: Design an encoder E_c that encodes X at the optimal rate and a decoder D_c that reconstructs $\hat{X}$ within an average distortion d, given side information Y.
- Y is statistically dependent on X.
- Rate of E_c is R bits/source-symbol.
- A pair (R, d) is achievable if $\exists$ an encoder E_c of rate R and a decoder D_c that can reconstruct X within an average distortion d – i.e., $\exists$ mappings

Encoder: $$f_E : \mathcal{X}^n \rightarrow \{0, 1, \ldots, 2^{n(R)} - 1\},$$

Decoder: $$f_D : \mathcal{Y}^n \times \{0, 1, \ldots, 2^{n(R)} - 1\} \rightarrow \hat{\mathcal{X}}^n,$$

such that: $$\lim \sup_{n \rightarrow \infty} E[D(\mathbf{X}^n, f_D(\mathbf{Y}^n, f_E(\mathbf{X}^n)))] \leq d$$

- Let $R_{X|Y}(d)$ be the smallest allowable rate if Y is available to both E_c and D_c.
- Let $R_Y^*(d)$ be the smallest allowable rate if Y is available only to D_c.
- Problem: Determine the set of achievable pairs (R,d).
- **Theorem 1** *If (X,Y) are drawn i.i.d. according to $Q(x,y)$=Pr$\{X=x,Y=y\}$, and $D:\mathcal{X}\times\hat{\mathcal{X}}\rightarrow R^+$, $(D(x^n,\hat{x}^n)=\frac{1}{n}\sum_i D(x_i,\hat{x}_i))$, then*

$$R_Y^*(d)=\min_{p(w|x)}\min_f(I(X;W)-I(Y;W)) \quad (1)$$

where W is an auxiliary random variable satisfying

$$\text{(a)} \sum_{w\in\mathcal{W}} p(x,y,w)=Q(x,y), \quad \text{(b)}\; p(x,y,w)=Q(x,y)p_t(w|x)$$

The minimization in (1) is over all $f:\mathcal{Y}\times\mathcal{W}\rightarrow\hat{\mathcal{X}}$ and $p(w|x)$, $|\mathcal{W}|\le|\mathcal{X}|+1$, such that $E[D(x,\hat{x})]=\sum_x\sum_w\sum_y Q(x,y)p(w|x)D(x,f(y,w))\le d$.

- $R_Y^*(d)\ge R_{X|Y}(d)$

  (For $d=0$, $R_Y^*(0)=R_{X|Y}(0)=H(X|Y)\leftarrow$ (Slepian-Wolf))

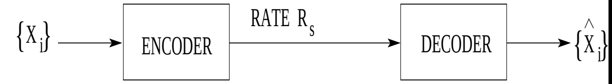
- If X and Y are jointly Gaussian, then $R_Y^*(d)=R_{X|Y}(d)$

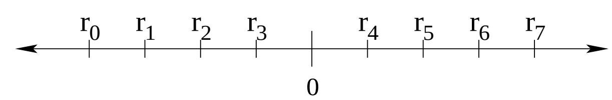
[A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, Jan, 1976.]

EXAMPLE: SOURCE CODING WITH SIDE INFORMATION

Without Side Information

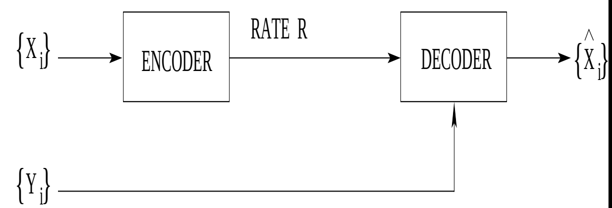
- X, Y – continuous valued RVs.
- Quantize x using 8-level quantizer. $w = Q_8(x)$.
- Encoder: Send index i of the region containing x
  $\Rightarrow R_s = 3$ bits/sample.
- Decoder: $\hat{x} = \arg\min_{a \in \mathbb{R}} E\Big[D(x,a) \mid X \in \Gamma_i\Big]$



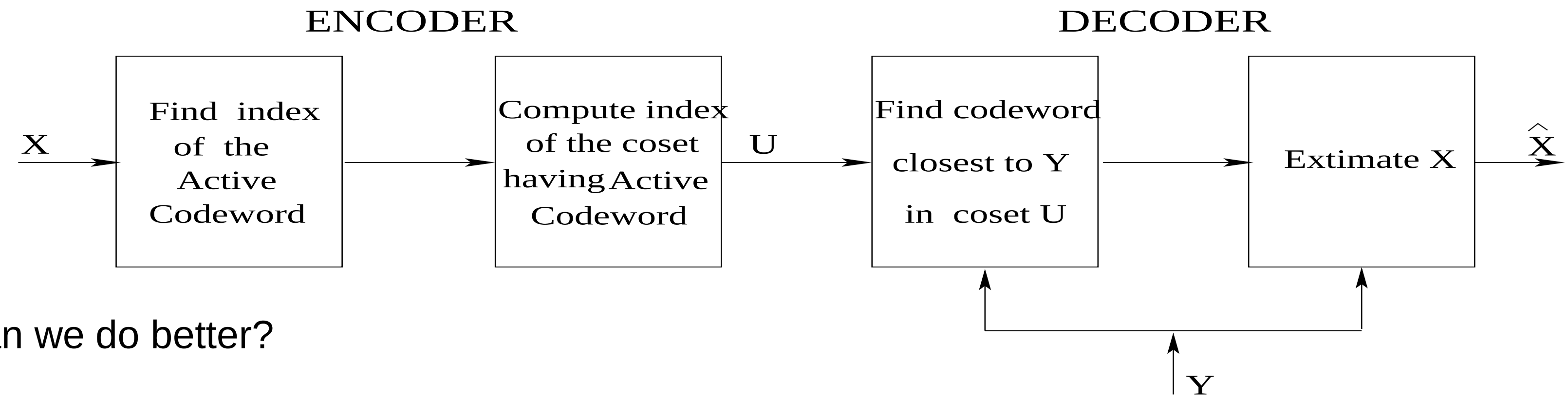


With Side Information

- Case 1: Encoder is the same.
- Decoder: $\hat{x} = \arg\min_{a \in \mathbb{R}} E\Big[D(x,a) \mid X \in \Gamma_i, Y = y\Big]$
- Case 2: $W_c = \{r_0, \ldots, r_1\}$
- $\mathbb{C} = \{r_0, r_2, r_4, r_6\}$, rate of $\mathbb{C}$ is $R_c = 2$ bits/sample.
- Encoder: Send index j of the coset containing
  $w = Q_8(x)$. Rate is $R = R_s - R_c = 1$ bit/sample.
- Decoder: Look for the most likely w_i in coset j
  i.e., $Y = y$ is the output of channel $P(Y|W)$.
  Estimate: $\hat{x} = \arg\min_{a \in \mathbb{R}} E\Big[D(x,a) \mid X \in \Gamma_i, Y = y\Big]$



- Case 3: $\mathbb{C} = \{r_0, r_4\}$, rate of $\mathbb{C}$ is $R_c = 1$ bits/sample.
- Encoder: Send index j of the coset containing $w = Q_8(x)$. Rate is $R = R_s - R_c = 2$ bit/sample.
- Rate of the channel code reduced, distance of the channel code $\mathbb{C}$ increased.



Can we do better?

- Block encoding, vector quantizers or scalar quantizers with memory.
- design a channel code $\mathbb{C}$ for the fictitious channel $P(Y|W)$ such that $\mathbb{C}$ partitions the set of codewords W_c into disjoint cosets.
- $R_c \leq I(Y;W)$. (From channel coding theorem)
- $R_s \geq I(X;W)$. (From rate distortion theorem)
- $\Rightarrow R = R_s - R_c \geq (I(X;W) - I(Y;W))$

[S. S. Pradhan, *Distributed Source Coding Using Syndromes (DISCUS)*. PhD thesis, University of California, Berkeley, 2001.]

THEOREM 1: PROOF OF CONVERSE

[T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Series, 1991.]

Lemma 2 (Convexity): $R_Y^*(d)$ is a non-increasing convex function in d.

Proof: $R_Y^*(d)$ monotonic since the domain in (1) increases with d.

- For distortions d_1,d_2, let f_1, W_1 and f_2, W_2 be the parameters that achieve the minima in (1).
- Let $Q = \begin{cases} 1 & \text{with prob } \lambda \\ 2 & \text{with prob } 1-\lambda \end{cases}$
- Let $W = (W_Q, Q)$
- d=$E[d(X,\hat{X})] = \lambda E[d(X, f_1(Y, W_1))] + (1-\lambda)E[d(X, f_2(Y, W_2))] = \lambda d_1 + (1-\lambda)d_2$

$$
\begin{aligned}
I(X;W) - I(Y;W) &= H(X) - H(X|W_Q, Q) - H(Y) + H(Y|W_Q, Q) \\
&= H(X) - \lambda H(X|W_1) - (1-\lambda)H(X|W_2) + \cdots \\
&= \lambda\Big(I(X;W_1) - I(Y;W_1)\Big) + (1-\lambda)\Big(I(X;W_2) - I(Y;W_2)\Big)
\end{aligned}
$$

- $R_Y^*(d) = \min(I(X;U) - I(Y;U)) \leq I(X;W) - I(Y;W)$
$= \lambda R_Y^*(d_1) + (1-\lambda)R_Y^*(d_2)$ □

Converse Theorem: *Suppose* $\exists\ f_n : \mathcal{X}^n \rightarrow \{1, \ldots, 2^{nR}\}$ *and* $g_n : \mathcal{Y}^n \times \{1, \ldots, 2^{nR}\} \rightarrow \hat{\mathcal{X}}^n$ *such that* $E[d(X^n, g_n(Y^n, f_n(X^n)))] \leq d$*, then* $R \geq R_Y^*(d)$.

Proof: Let $T = f_n(X^n)$ and let $W_i = (T, Y^{i-1}, Y_{i+1}^n)$.

$$nR \geq H(T) \geq H(T|Y^n) \geq I(X^n; T|Y^n) = \sum_{i=1}^{n} I(X_i; T|Y^n, X^{i-1})$$

$$= \sum_i (H(X_i|Y^n, X^{i-1}) - H(X_i|T, Y^n, X^{i-1})) = \sum_i (H(X_i|Y_i) - H(X_i|T, Y^n, X^{i-1}))$$

$$\geq \sum_i (H(X_i|Y_i) - H(X_i|T, Y^n)) = \sum_i (H(X_i|Y_i) - H(X_i|W_i, Y_i))$$

$$= \sum_i I(X_i; W_i|Y_i) = \sum_i (H(W_i|Y_i) - H(W_i|X_i, Y_i))$$

$$= \sum_i (H(W_i|Y_i) - H(W_i|X_i)) = \sum_i (H(W_i) - H(W_i|X_i) - (H(W_i) - H(W_i|Y_i)))$$

$$= \sum_i (I(X_i; W_i) - I(Y_i; W_i))$$

$$\geq \sum_i R_Y^*(E[d(X_i, g_{ni}(Y_i, W_i))]) = n\frac{1}{n}\sum_i R_Y^*(E[d(X_i, g_{ni}(Y_i, W_i))])$$

$$\geq nR_Y^*(E[\sum_i \frac{1}{n} d(X_i, g_{ni}(Y_i, W_i))]) = nR_Y^*(d)$$

(where $\hat{X}^n = g_n(Y^n, f_n(X^n)) = g_n(Y_i, W_i)$, $\hat{X}_i = g_{ni}(Y_i, W_i)) \Rightarrow R \geq R_Y^*(d)$

DUALITY BETWEEN CHANNEL CAPACITY AND RATE DISTORTION

Channel Coding Theorem: All rates below capacity C are achievable. Specifically, $\forall \epsilon > 0$ and rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$.

Conversely, any sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \to 0$ must have $R \leq C$.

Rate Distortion Theorem: All rates above rate distortion function $R(D)$ are achievable. Specifically, $\forall \epsilon > 0$ and rate $R > R(D)$, there exists a sequence of $(2^{nR}, n)$ rate distortion codes with average distortion $\leq D$.

Conversely, any sequence of $(2^{nR}, n)$ rate distortion codes with average distortion $\leq D$ must have $R \geq R(D)$.

Proof of Achievability of CCT

- Fix $p(x)$. Let $R < C = \max_{p(x)} I(X;Y)$
- Generate a $(2^{nR}, n)$ code at random $\sim p(x)$.
- Encoder: If message is $w \in \{1, \ldots, 2^{nR}\}$, send codeword $X^n(w)$.
- Decoder: Receive Y^n and decode $\hat{w}$ if $\exists$ a unique $\hat{w}$ such that $(X^n(\hat{w}), Y^n) \in A_\epsilon^{(n)}$.
- Prob of error calculation
  - E1: there is no $X^n(\hat{w}$ that is jointly typical with Y^n. This prob. is very small, say $P_{E_1} < \epsilon$.
  - E2: there is more than codeword that is jointly typical with Y^n. This prob is $P_{E_2} \leq 2^{-n(I(X;Y)-3\epsilon)} 2^{nR}$.
  - $P_{E_2} \to 0$ if $R < I(X;Y) - 3\epsilon$.

Proof of Achievability of RDT

- Fix $p(\hat{x}|x)$. Let $R > R(d) = \min_{p(\hat{x}|x):Ed(X,\hat{X}) \leq d} I(X;\hat{X})$
- Generate a $(2^{nR}, n)$ RD code at random $\sim p(\hat{x})$.
- Encoder: If X^n is the message, look for a $w \in \{1, \ldots, 2^{nR}\}$ such that $(X^n, \hat{X}^n(w)) \in A_{d,\epsilon}^{(n)}$. If there is no such w send $w = 1$, else send the smallest w.
- Decoder: Estimate is $\hat{X}^n(w)$.
- Average distortion calculation
  - E1: If there is at least one w such that $(\hat{X}^n(w), X^n) \in A_{d,\epsilon}^{(n)}$, then the average distortion is $D1 \leq d + \epsilon$.
  - E2: If there is no such w, then $D2 \leq P_e d_{max}$
  - $P_e \to 0$ if $R > I(X;\hat{X}) + 3\epsilon$.

Strong Typicality and Markov Lemma

- A sequence $x^n \in \mathcal{X}^n$ is ϵ-strongly typical if:
  (1) $\forall a \in \mathcal{X}$, $\left|\frac{1}{n}N(a|x^n) - p(a)\right| < \frac{\epsilon}{|\mathcal{X}|}$, (2) $N(a|x^n) = 0$ if $p(a) = 0$ (where $N(a|x^n)$ is the number of occurrences of a in x^n).
- $A_\epsilon^{*(n)}$ is the set of ϵ-strongly typical sequences x^n.
- A pair (x^n, y^n) is ϵ-strongly typical if:
  (1) $\forall a \in \mathcal{X}, b \in \mathcal{Y}$, $\left|\frac{1}{n}N(a,b|x^n,y^n) - p(a,b)\right| < \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|}$,
  (2) $N(a,b|x^n,y^n) = 0$ if $p(a,b) = 0$ (where $N(a,b|x^n,y^n)$ is the number of occurrences of the pair (a,b) in (x^n,y^n)).
- $A_\epsilon^{*(n)}(X,Y)$ is the set of ϵ-strongly typical sequence pairs (x^n,y^n).
- **Lemma 3:** Pr$(A_\epsilon^{*(n)}) \to 1$ as $n \to \infty$.
- **Lemma 4:** Let $Y_1, \ldots, Y_n$ be drawn $\sim \prod p(y)$. For $x^n \in A_\epsilon^{*(n)}$, we have

$$2^{-n(I(X;Y)+\epsilon_1)} \leq \Pr((x^n, Y^n) \in A_\epsilon^{*(n)}) \leq 2^{-n(I(X;Y)-\epsilon_1)}$$

- $A_\epsilon^{*(n)}(X,Y,Z)$ is the set of ϵ-strongly typical sequence triplets (x^n,y^n,z^n).
- If $(x^n,y^n,z^n) \in A_\epsilon^{*(n)} \Rightarrow (x^n,y^n) \in A_\epsilon^{*(n)}$ and $(y^n,z^n) \in A_\epsilon^{*(n)}$

- Converse is not necessarily true.
- **Markov Lemma 5:** If $X \rightarrow Y \rightarrow Z$, i.e., $p(x, y, z) = p(x, y)p(z|y)$. If for a given $(y^n, z^n) \in A_\epsilon^{*(n)}$, X^n is drawn $\sim \prod_i p(x_i|y_i)$, then
  Pr$\{(X^n, y^n, z^n) \in A_\epsilon^{*(n)}(X, Y, Z)\} > 1 - \epsilon$ for n sufficiently large.

PROOF OF ACHIEVABILITY

- Fix $p(w|x)$ and $f(w,y)$ that achieves equality in (1) for $Ed(X,\hat{X}) \leq d$.
- Calculate $p(w) = \sum_x p(w|x)p(x)$. Suppose $R_2 > R_Y^*(d)$.
- Codebook:
  - Let $R_1 = I(W;X) + \epsilon$. Generate 2^{nR_1} codewords $W^n(s) \sim \prod_{i=1}^n p(w_i)$, $s \in \{1, \ldots, 2^{nR_1}\}$.
  - Let $R_2 = I(X;W) - I(Y;W) + 5\epsilon$. Randomly assign $s \in \{1, \ldots, 2^{nR_1}\}$ to one of 2^{nR_2} bins.
- Encoder: If message is X^n, look for a s such that $(X^n, W^n(s)) \in A_\epsilon^{*(n)}$. If there is no such s, set $s = 1$, else choose the smallest s. Send index of bin i containing s.
- Decoder: Look for a s in bin i such that $(W^n(s), Y^n) \in A_\epsilon^{*(n)}$. If there is a unique such s, then $\hat{X}^n$ is found by $\hat{X}_i = f(W_i, Y_i)$, else $\hat{X}^n = \hat{x}^n$ arbitrary.
- Probability of error and average distortion analysis
  - E_1: $(X^n, Y^n) \notin A_\epsilon^{*(n)}$. $P_{E1} < \epsilon$, $D1 < \epsilon d_{max}$
  - E_2: X^n is typical, but there is no s such that $(X^n, W^n(s)) \in A_\epsilon^{*(n)}$. $P_{E_2} \to 0$ if $R_1 > I(X;W)$ (Proof of the rate distortion theorem)

- E_3: $(X^n, W^n(s)) \in A_\epsilon^{*(n)}$ but $(W^n(s), Y^n) \notin A_\epsilon^{*(n)}$.
  P_{E_3} is small (Markov Lemma).

- E_4: There is another s' in bin i such that $(W^n(s'), Y^n) \in A_\epsilon^{*(n)}$.
  $P_{E_4} \leq 2^{n(R1-R2)} 2^{-n(I(Y;W)-3\epsilon)}$. (Proof of the channel capacity theorem)

- $P_{E_4} \to 0$ since $R1 - R2 < I(Y;W) - 3\epsilon$.

- If s is decoded correctly, then $(X^n, W^n(s)) \in A_\epsilon^{*(n)}$.
  Since $P_{E1} < \epsilon$, we can assume $(X^n, Y^n) \in A_\epsilon^{*(n)}$.
  Then, Markov Lemma $\Rightarrow$ $(X^n, W^n(s), Y^n) \in A_\epsilon^{*(n)}$.
  $\Rightarrow$ empirical joint distribution is close to $p(x,y)p(w|x)$,
  $\Rightarrow$ $(X^n, \hat{X}^n)$ will have a joint distribution that is close to the distribution achieving distortion d.

$$\sum_{x,\hat{x}} h(x,\hat{x})\, d(x,\hat{x}) \cong \sum_{x} \sum_{y,w: f(w,y)=\hat{x}} p(x,y)p(w|x)\, d(x, f(w,y)) \leq d$$

Proof of Achievability

Wyner-Ziv argument

- **Lemma A** Let X, Y, W, f be as before. For $\epsilon_0 > 0$ and large n_0, there exists a code (n_0, M_0, Δ_0) defined by (F_E^0, F_D^0) such that:

$$\Delta_0 \leq d + \epsilon_0, \;\; \frac{1}{n} H(T|Y^{n_0}) \leq R_0 + \epsilon_0$$

  where $T = F_E^0(X^{n_0})$, $R_0 = I(X;W) - I(Y;W)$.

- **Lemma B** (Slepian-Wolf) Let (F_E^0, F_D^0) be a (n_0, M_0, Δ_0) code. Then if $R_0 = \frac{1}{n} H(T|Y^{n_0})$ (where $T = F_E^0(X^{n_0})$), then for a sufficiently large n_1 and $\delta > 0$ there exists a code (n, M, Δ) such that

$$n = n_0 n_1, \; M \leq 2^{n_1(n_0 R_0 + \delta)} \leq 2^{n(R_0+\delta)}, \text{ and } \Delta \leq \Delta_0 + \delta$$

- Lemma A and Lemma B $\Rightarrow$ Prove achievability of RDT with Side Information.