The Likelihood Encoder for Lossy Source Compression

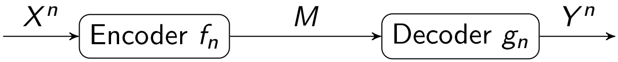
Eva Song, Paul Cuff, and H. Vincent Poor

Department of Electrical Engineering
Princeton University

July 3, 2014

What is a likelihood encoder?

- a stochastic source encoder: $f_n : \mathcal{X}^n \mapsto \mathcal{M}$



Given

- a codebook $\{y^n(m)\}_m$, $m \in [1 : 2^{nR}]$
- a joint distribution P_{XY}

the likelihood function for each codeword:

$$\mathcal{L}(m|x^n) \triangleq P_{X^n|Y^n}(x^n|y^n(m)) = \prod P_{X|Y}(x^n|y^n(m))$$

the likelihood encoder determines the message index according to:

$$P_{M|X^n}(m|x^n) = \frac{\mathcal{L}(m|x^n)}{\sum_{m' \in [1:2^{nR}]} \mathcal{L}(m'|x^n)} \propto \mathcal{L}(m|x^n).$$

What can a likelihood encoder do?

It can be used for achievability proofs of

- Point-to-point rate-distortion theory [Cuff & Song, ITW '13]
- Wyner-Ziv (this talk)
- Berger-Tung inner bound (this talk)
- Rate-distortion based secrecy systems (Schieler & Cuff, Satpathy & Cuff, ISIT '14 Mon)

Why likelihood encoder?

- Simpler proofs
  - no "error" events
  - soft-covering lemma
- Results hold for both discrete and continuous alphabets
- No need for Markov lemma, mutual packing lemma, ...

Warm-up I – total variation distance

$$\|P - Q\|_{TV} \triangleq \sup_{A \in \mathcal{F}} |P(A) - Q(A)|.$$

- **C**ontinuity: If $f(x)$ has bounded range with width $b \in \mathbb{R}^+$, then

$$\|P - Q\|_{TV} < \varepsilon \implies \left|\mathbb{E}_P[f(X)] - \mathbb{E}_Q[f(X)]\right| < \varepsilon b.$$

- **T**riangle inequality:

$$\|P - Q\|_{TV} \leq \|P - S\|_{TV} + \|S - Q\|_{TV}.$$

- **M**ultiplicative identity:

$$\|P_X P_{Y|X} - Q_X P_{Y|X}\|_{TV} = \|P_X - Q_X\|_{TV}.$$

- **J**oint-Marginal inequality:

$$\|P_X - Q_X\|_{TV} \leq \|P_{XY} - Q_{XY}\|_{TV}.$$

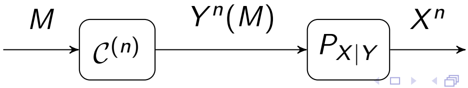
Warm-up II – **S**oft-**C**overing lemma

Lemma (Wyner '75; Han-Verdú '93; Cuff '12)

- *Given*
  1) P_{XY}
  2) random $\mathcal{C}^{(n)}$ *of sequences* $Y^n(m) \sim \prod_{t=1}^{n} P_Y(y_t)$,
  $m \in [1 : 2^{nR}]$,
- $\mathbf{P}_{X^n}$*: the output distribution induced by*
  1) $M \sim Unif[1 : 2^{nR}]$
  2) $M \rightarrow Y^n(M) \rightarrow \prod P_{X|Y}$

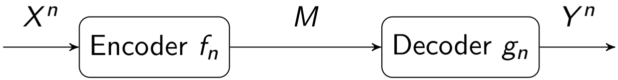
Then if $R > I(X;Y)$,

$$\mathbb{E}_{\mathcal{C}^n} \left\| \mathbf{P}_{X^n} - \prod_{t=1}^{n} P_X \right\|_{TV} \leq \epsilon_n \rightarrow_n 0.$$



Review – point-to-point rate distortion theory

- Distortion measure $d(\cdot,\cdot)$
  - $d(x^n, y^n) = \frac{1}{n}\sum_{t=1}^{n} d(x_t, y_t)$
  - $d_{max} = \max_{(x,y)\in\mathcal{X}\times\mathcal{Y}} d(x,y)$
- i.i.d. source sequence $X^n \sim \prod \overline{P}_X \triangleq \overline{P}_{X^n}$
- $\mathbb{E}[d(X^n, Y^n)] \leq_n D$



$$R(D) = \min_{\overline{P}_{Y|X}:\mathbb{E}[d(X,Y)]\leq D} I_{\overline{P}}(X;Y)$$

- System induced distribution:

$$\mathbf{P}_{X^nMY^n}(x^n, m, y^n) = \overline{P}_{X^n}(x^n)\mathbf{P}_{LE}(m|x^n)\mathbf{P}_D(y^n|m)$$

Recipe

- System induced distribution $\mathbf{P}$
- Idealized distribution $\mathbf{Q}$
- Distortion under $\mathbf{Q} \leftrightarrow$ single-letter distortion
- SC $\Rightarrow \|\mathbf{P} - \mathbf{Q}\|_{TV} \rightarrow_n 0$
- Distortion under $\mathbf{P} \rightarrow_n$ Distortion under $\mathbf{Q}$

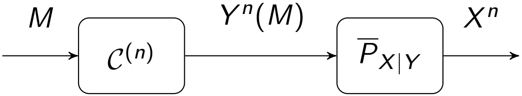
Achievability scheme

- Fix
  - D
  - $\overline{P}_{Y|X}$: $\mathbb{E}_{\overline{P}}[d(X;Y)] < D$
  - $R > I_{\overline{P}}(X;Y)$
- Codebook generation: Independently generate 2^{nR} sequences in $\mathcal{Y}^n$ according to $\prod_{t=1}^{n} \overline{P}_Y(y_t)$ and index them by $m \in [1:2^{nR}]$.
- Encoder: the likelihood encoder with

$$\mathcal{L}(m|x^n) = \overline{P}_{X^n|Y^n}(x^n|Y^n(m)).$$

- Decoder: codeword lookup $\mathbf{P}_D(y^n|m) = \mathbb{1}\{y^n = Y^n(m)\}$.

Analysis



- Idealized distribution

$$\mathbf{Q}_{X^nY^nM}(x^n, y^n, m) = \frac{1}{2^{nR}} \mathbb{1}\{y^n = Y^n(m)\} \prod_{t=1}^{n} \overline{P}_{X|Y}(x_t|y_t)$$

- Properties of $\mathbf{Q}$:
  - $\mathbf{Q}_{M|X^n}(m|x^n) = \mathbf{P}_{LE}(m|x^n)$
  - $\mathbf{Q}_{Y^n|M}(y^n|m) = \mathbf{P}_D(y^n|m)$
  - $\mathbb{E}_{\mathcal{C}^{(n)}}[\mathbf{Q}_{X^nY^n}(x^n, y^n)] = \overline{P}_{X^nY^n}(x^n, y^n)$
  - $\mathbb{E}_{\mathcal{C}^{(n)}}[\mathbb{E}_{\mathbf{Q}}[d(X^n, Y^n)]] = \mathbb{E}_{\overline{P}}[d(X, Y)]$

Analysis – continued

- **P** and **Q** are close in TV, $R > I_{\overline{P}}(X;Y)$

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|\mathbf{P}_{X^nY^n} - \mathbf{Q}_{X^nY^n}\|_{TV}\right] &\overset{(J)}{\leq} \mathbb{E}_{\mathcal{C}^{(n)}}\left[\|\mathbf{P}_{X^nY^nM} - \mathbf{Q}_{X^nY^nM}\|_{TV}\right] \\
&\overset{(M)}{=} \mathbb{E}_{\mathcal{C}^{(n)}}\left[\|\overline{P}_{X^n} - \mathbf{Q}_{X^n}\|_{TV}\right] \\
&\overset{SC}{\leq} \epsilon_n \to_n 0
\end{aligned}$$

- Bound the expected distortion under **P**

$$\begin{aligned}
& \mathbb{E}_{\mathcal{C}^{(n)}}\left[\mathbb{E}_{\mathbf{P}}[d(X^n,Y^n)]\right] - \mathbb{E}_{\mathcal{C}^{(n)}}\left[\mathbb{E}_{\mathbf{Q}}[d(X^n,Y^n)]\right] \\
\leq\ & \mathbb{E}_{\mathcal{C}^{(n)}}\left[\left|\mathbb{E}_{\mathbf{P}}[d(X^n,Y^n)] - \mathbb{E}_{\mathbf{Q}}[d(X^n,Y^n)]\right|\right] \\
\overset{(C)}{\leq}\ & d_{max}\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|\mathbf{P}_{X^nY^n} - \mathbf{Q}_{X^nY^n}\|_{TV}\right] \\
\leq\ & d_{max}\epsilon_n
\end{aligned}$$

- Finally,

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\mathbb{E}_{\mathbf{P}}[d(X^n,Y^n)]\right] \leq D + d_{max}\epsilon_n$$

Warm-up III–**A**pproximation lemma

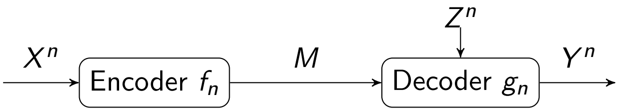
Lemma

For a distribution P_{UVX} *and* $0 < \varepsilon < 1$,

$$\mathbb{P}[U \neq V] \leq \varepsilon \quad \Rightarrow \quad \|P_{UX} - P_{VX}\|_{TV} \leq \varepsilon.$$

Wyner-Ziv

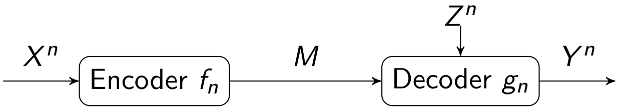
- i.i.d. source with correlated side info $X^n Z^n \sim \prod \overline{P}_{XZ}$
- $\mathbb{E}[d(X^n, Y^n)] \leq_n D$



$$R(D) = \min_{\overline{P}_{V|XZ} \in \mathcal{M}(D)} I_{\overline{P}}(X; V|Z)$$

$$\mathcal{M}(D) = \Big\{ \overline{P}_{V|XZ} : \quad V - X - Z, \quad |\mathcal{V}| \leq |\mathcal{X}| + 1, \text{ and there exists a function } \phi \text{ s.t. } \mathbb{E}\left[d(X,Y)\right] \leq D, Y \triangleq \phi(V,Z) \Big\}.$$

Wyner-Ziv



- System induced distribution:

$$
\begin{aligned}
& \mathbf{P}_{X^n Z^n M M' \hat{M}' Y^n}(x^n, z^n, m, m', \hat{m}', y^n) \\
= \; & \overline{P}_{X^n Z^n}(x^n, z^n)\mathbf{P}_{LE}(m, m'|x^n)\mathbf{P}_D(\hat{m}'|m, z^n)\mathbf{P}_\Phi(y^n|m, \hat{m}', z^n)
\end{aligned}
$$

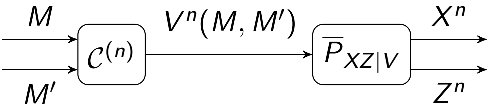
Achievability scheme

- Fix
  - D
  - $\overline{P}_{V|XZ} \in \mathcal{M}(D)$ with corresponding ϕ
  - $R + R' > I_{\overline{P}}(X;V)$, $R' < I_{\overline{P}}(V;Z)$
- Codebook generation: Independently generate $2^{n(R+R')}$ sequences in $\mathcal{V}^n$ according to $\prod_{i=1}^{n} \overline{P}_V(v_i)$ and index by $(m, m') \in [1 : 2^{nR}] \times [1 : 2^{nR'}]$.
- Encoder: the likelihood encoder with

$$\mathcal{L}(m, m'|x^n) = \overline{P}_{X^n|V^n}(x^n|V^n(m, m')).$$

- Decoder
  - $\mathbf{P}_D(\hat{m}'|m, z^n)$: a good channel decoder w.r.t.
    - ★ sub-codebook $\mathcal{C}^{(n)}(m) = \{v^n(m, a)\}_a$
    - ★ memoryless channel $\overline{P}_{Z|V}$
  - $\mathbf{P}_\Phi(y^n|m, \hat{m}', z^n) = \mathbb{1}\{y^n = \phi^n(V^n(m, \hat{m}'), z^n)\}$
    - ★ $\phi^n(v^n, z^n)$ is the concatenation $\{\phi(v_t, z_t)\}_{t=1}^{n}$

Analysis



- Idealized distribution

$$\mathbf{Q}_{X^nZ^nV^nMM'}(x^n, z^n, v^n, m, m') = \frac{1}{2^{n(R+R')}} \mathbb{1}\{v^n = V^n(m, m')\} \prod_{t=1}^{n} \overline{P}_{X|V}(x_t|v_t)\overline{P}_{Z|X}(z_t|x_t)$$

- Properties of $\mathbf{Q}$:
  - $\mathbf{Q}_{MM'|X^n}(m, m'|x^n) = \mathbf{P}_{LE}(m, m'|x^n)$
  - $\mathbb{E}_{\mathcal{C}^{(n)}}[\mathbf{Q}_{X^nZ^nV^n}(x^n, z^n, v^n)] = \overline{P}_{X^nZ^nV^n}(x^n, z^n, v^n)$

Analysis–continued

- Two distributions

$$\mathbf{Q}^{(1)}_{X^nZ^nMM'\hat{M}'Y^n}(x^n, z^n, m, m', \hat{m}', y^n)$$
$$\triangleq \mathbf{Q}_{X^nZ^nMM'}(x^n, z^n, m, m')\mathbf{P}_D(\hat{m}'|m, z^n)\mathbf{P}_\Phi(y^n|m, \hat{m}', z^n)$$

$$\mathbf{Q}^{(2)}_{X^nZ^nMM'\hat{M}'Y^n}(x^n, z^n, m, m', \hat{m}', y^n)$$
$$\triangleq \mathbf{Q}_{X^nZ^nMM'}(x^n, z^n, m, m')\mathbf{P}_D(\hat{m}'|m, z^n)\mathbf{P}_\Phi(y^n|m, m', z^n)$$

- $\mathbf{P}$ and $\mathbf{Q}^{(1)}$ are close in TV

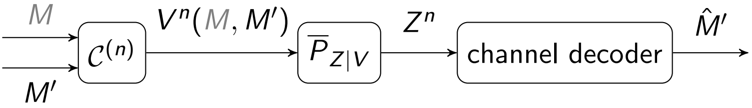
$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|\mathbf{P}_{X^nZ^nMM'\hat{M}'Y^n} - \mathbf{Q}^{(1)}_{X^nZ^nMM'\hat{M}'Y^n}\|_{TV}\right] \overset{SC}{\leq} \epsilon_n \to_n 0$$

- Distortion under $\mathbf{Q}^{(2)}$

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\mathbb{E}_{\mathbf{Q}^{(2)}}[d(X^n, Y^n)]\right] = \mathbb{E}_{\overline{P}}[d(X, Y)]$$

Analysis–continued

- $\mathbb{P}_{\mathbf{Q}^{(1)}}[\hat{M}' \neq M'] = \mathbb{P}_{\mathbf{Q}^{(2)}}[\hat{M}' \neq M']$
- Channel coding for fixed M

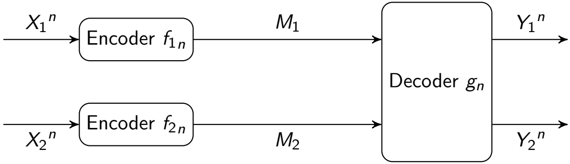


  - sub-codebook $\mathcal{C}^{(n)}(m) = \{v^n(m,a)\}_a$
  - memoryless channel $\overline{P}_{Z|V}$
  - $R' < I_{\overline{P}}(V;Z) \Rightarrow \mathbb{E}_{\mathcal{C}^{(n)}}\left[\mathbb{P}_{\mathbf{Q}^{(1)}}[M' \neq \hat{M}']\right] \leq \delta_n \to_n 0$
  - $\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|\mathbf{Q}^{(1)}_{X^nZ^nM\hat{M}'Y^n} - \mathbf{Q}^{(2)}_{X^nZ^nMM'Y^n}\|_{TV}\right] \overset{A}{\leq} \delta_n$

- $\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|\mathbf{P}_{X^nY^n} - \mathbf{Q}^{(2)}_{X^nY^n}\|_{TV}\right] \overset{(T)}{\leq} \epsilon_n + \delta_n$
- $\mathbb{E}_{\mathcal{C}^{(n)}}\left[\mathbb{E}_{\mathbf{P}}[d(X^n,Y^n)]\right] \overset{(C)}{\leq} \mathbb{E}_{\overline{P}}[d(X,Y)] + d_{max}(\epsilon_n + \delta_n)$

Berger-Tung inner bound

- i.i.d. source sequences ${X_1}^n {X_2}^n \sim \prod \overline{P}_{X_1 X_2}$
- $\mathbb{E}[d_1({X_1}^n, {Y_1}^n)] \leq_n D_1$, $\mathbb{E}[d_2({X_2}^n, {Y_2}^n)] \leq_n D_2$



$$
\begin{aligned}
R_1 &> I_{\overline{P}}(X_1; U_1 | U_2), \\
R_2 &> I_{\overline{P}}(X_2; U_2 | U_1), \\
R_1 + R_2 &> I_{\overline{P}}(X_1, X_2; U_1, U_2),
\end{aligned}
$$

$\overline{P}_{U_1 U_2 | X_1 X_2} : U_1 - X_1 - X_2 - U_2$,
$\phi_k(\cdot, \cdot)$ s.t. $\mathbb{E}[d_k(X_k, Y_k)] \leq D_k$, where $Y_k \triangleq \phi_k(U_1, U_2), k = 1, 2$.

- not convex: can be improved through time-sharing
- sufficient to show for corner points:
  $C_1 \triangleq \left(I_{\overline{P}}(X_1; U_1), I_{\overline{P}}(X_2; U_2 | U_1)\right)$, $C_2 \triangleq \left(I_{\overline{P}}(X_1; U_1 | U_2), I_{\overline{P}}(X_2; U_2)\right)$

Achievability proof idea

- Demonstrate with $C_1 \triangleq \left(I_{\overline{P}}(X_1; U_1), I_{\overline{P}}(X_2; U_2|U_1)\right)$
- System induced distribution:
  $\mathbf{P}_{X_1{}^n X_2{}^n U_1{}^n M_1 M_2 M_2' \hat{M}_2' Y_1{}^n Y_2{}^n} = \overline{P}_{X_1{}^n X_2{}^n} \mathbf{P}_1 \mathbf{P}_2$
  - $\mathbf{P}_1(m_1, u_1{}^n | x_1{}^n) = \mathbf{P}_{LE1}(m_1|x_1{}^n)\mathbf{P}_{D1}(u_1{}^n|m_1)$
    - ★ same form as P2P
    - ★ $\mathbf{P}_{MY^n|X^n}(m, y^n|x^n) = \mathbf{P}_{LE}(m|x^n)\mathbf{P}_D(y^n|m)$
  - $\mathbf{P}_2(m_2, m_2', \hat{m}_2', y_1{}^n, y_2{}^n | x_2{}^n, u_1{}^n) =$
    $\mathbf{P}_{LE2}(m_2, m_2'|x_2{}^n)\mathbf{P}_D(\hat{m}_2'|m_2, u_1{}^n) \prod_{k=1,2} \mathbf{P}_{\Phi,k}(y_k{}^n|u_1{}^n, m_2, \hat{m}_2')$
    - ★ same form as WZ
    - ★ $\mathbf{P}_{MM'\hat{M}'Y^n|X^nZ^n}(m, m', \hat{m}', y^n|x^n, z^n) =$
      $\mathbf{P}_{LE}(m, m'|x^n)\mathbf{P}_D(\hat{m}'|m, z^n)\mathbf{P}_\Phi(y^n|z^n, m, \hat{m}')$
- Fix
  - D_1, D_2
  - $\overline{P}_{U_1U_2|X_1X_2}: U_1 - X_1 - X_2 - U_2$
    and ϕ_k s.t. $\mathbb{E}[d_k(X_k, Y_k)] \leq D_k$, where $Y_k \triangleq \phi_k(U_1, U_2)$
  - $R_1 > I_{\overline{P}}(X_1; U_1)$, $R_2 + R_2' > I_{\overline{P}}(X_2; U_2)$, $R_2' < I_{\overline{P}}(U_1; U_2)$

- Codebook generation:
  - $\mathcal{C}_1^{(n)}$: independently generate 2^{nR_1} sequences in ${\mathcal{U}_1}^n$ according to $\prod_{t=1}^n \overline{P}_{U_1}(u_{1\,t})$ and index them by $m_1 \in [1:2^{nR_1}]$
  - $\mathcal{C}_2^{(n)}$: independently generate $2^{n(R_2+R_2')}$ sequences in ${\mathcal{U}_2}^n$ according to $\prod_{t=1}^n \overline{P}_{U_2}(u_{2\,t})$ and index them by $(m_2, m_2') \in [1:2^{nR_2}] \times [1:2^{nR_2'}]$
- Encoders:
  - $\mathbf{P}_{LE1}(m_1|{x_1}^n)$: likelihood encoder with

$$\mathcal{L}(m_1|{x_1}^n) = \overline{P}_{{X_1}^n|{U_1}^n}({x_1}^n|{U_1}^n(m_1))$$

  - $\mathbf{P}_{LE2}(m_2, m_2'|{x_2}^n)$: likelihood encoder with

$$\mathcal{L}(m_2, m_2'|{x_2}^n) = \overline{P}_{{X_2}^n|{U_2}^n}({x_2}^n|{U_2}^n(m_2, {m_2}'))$$

- Decoder:
  - $\mathbf{P}_{D1}({u_1}^n|m_1) = \mathbb{1}\{{u_1}^n = {U_1}^n(m_1)\}$, $\mathcal{C}_1^{(n)}$ codeword lookup
  - $\mathbf{P}_D({\hat{m}_2}'|m_2, {u_1}^n)$: a good channel decoder w.r.t.
    - ★ sub-codebook $\mathcal{C}_2^{(n)}(m_2) = \{{u_2}^n(m_2, a)\}_a$
    - ★ memoryless channel $\overline{P}_{U_1|U_2}$
  - $\mathbf{P}_{\Phi,k}({y_k}^n|{u_1}^n, m_2, \hat{m}_2') = \mathbb{1}\{{y_k}^n = {\phi_k}^n({u_1}^n, {U_2}^n(m_2, \hat{m}_2'))\}$
    - ★ ${\phi_k}^n({u_1}^n, {u_2}^n)$ is the concatenation $\{\phi_k(u_{1\,t}, u_{2\,t})\}_{t=1}^n$

Analysis–key steps

- $\mathbf{Q}_{1\,X_1{}^nX_2{}^nU_1{}^nM_1M_2M_2'\hat{M}_2'Y_1{}^nY_2{}^n} = \mathbf{Q}_{1\,M_1U_1{}^nX_1{}^nX_2{}^n}\mathbf{P}_2$
  - Idealized distribution:

$$\mathbf{Q}_1 = \frac{1}{2^{nR_1}}\mathbb{1}\{u_1{}^n = U_1{}^n(m_1)\}\overline{P}_{X_1{}^n|U_1{}^n}(x_1{}^n|u_1{}^n)\overline{P}_{X_2{}^n|X_1{}^n}(x_2{}^n|x_1{}^n)$$

  - $\mathbb{E}_{\mathcal{C}_1^{(n)}}[\|\mathbf{Q}_1 - \mathbf{P}\|_{TV}] \overset{SC}{\leq} \epsilon_{1n}$
- $\mathbf{Q}_1^* = \mathbb{E}_{\mathcal{C}_1^{(n)}}\left[\mathbf{Q}_{1\,X_1{}^nX_2{}^nU_1{}^nM_2M_2'\hat{M}_2'Y_1{}^nY_2{}^n}\right] = \overline{P}_{X_1{}^nX_2{}^nU_1{}^n}\mathbf{P}_2$
  - $\mathbb{E}_{\mathcal{C}_1^{(n)}}[\mathbb{E}_{\mathbf{P}}[d_k(X_k{}^n, Y_k{}^n)]] \leq \mathbb{E}_{\mathbf{Q}_1^*}[d_k(X_k{}^n, Y_k{}^n)] + d_{k\,max}\epsilon_{1n}$
- Treat $\mathbf{Q}_1^*$ as "**P** in WZ"
  - SC+channel coding
  - $\mathbb{E}_{\mathcal{C}_2^{(n)}}\left[\mathbb{E}_{\mathbf{Q}_1^*}[d_k(X_k{}^n, Y_k{}^n)]\right] \leq \mathbb{E}_{\overline{P}}[d_k(X_k, Y_k)] + d_{k\,max}(\epsilon_{2n} + \delta_n)$
- Finally,

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}_2^{(n)}}\left[\mathbb{E}_{\mathcal{C}_1^{(n)}}[\mathbb{E}_{\mathbf{P}}[d_k(X_k{}^n, Y_k{}^n)]]\right] \\ \leq\; &\mathbb{E}_{\overline{P}}[d_k(X_k, Y_k)] + d_{k\,max}\epsilon_{1n} + d_{k\,max}(\epsilon_{2n} + \delta_n)\end{aligned}$$

Summary

- Tool: Soft-covering lemma
- Likelihood encoder for achievability proofs of
  - P2P
  - Wyner-Ziv
  - Berger-Tung inner bound
- Joint-typicality encoder: Markov lemma, subtleties in decoding bin index, etc...