Recap

We saw that in order to simulate a distribution P_V exactly, we need at LEAST $H(P_V)$ iid Ber($\frac{1}{2}$) bits, or equivalently a uniform r.v. $W \sim Unif(\{1, \dots, 2^{H(P_V)}\})$

Now we are relaxing the requirements in several aspects:

(i) We assume "P_V" from before is an n-fold product measure $P_V^{\otimes n}$

(ii) We no longer require exact simulation but only approximate simulation

(iii) We have access to a noisy communication channel that maps inputs from some alphabet $\mathcal{U}$ to the alphabet $\mathcal{V}$, which corresponds with supp(P_V). The channel is denoted by $P_{V|U}$.

Strategy: We will design a codebook $\mathcal{C}_n = \{u^n(w)\}_{w \in \mathcal{W}}$ such that when picking a codeword uniformly at random and transmitting it through the channel will produce a channel output sequence $V^n \sim Q_{V^n}^{(\mathcal{C}_n)}$ with the property that

$$\delta_{TV}\left(Q_{V^n}^{(\mathcal{C}_n)}, P_V^{\otimes n}\right) \xrightarrow{n \to \infty} 0$$

Assumption: We assume the following relation between $P_{V|U}$ and P_V: $\exists\, P_U \in \mathcal{P}(\mathcal{U}), \quad \sum_{u \in \mathcal{U}} P_U(u) P_{V|U}(\cdot|u) = P_V(\cdot)$

Diagram

randomly chosen deterministic u sequence

$$W \sim \mathrm{Unif}(\{1,2,\dots,2^{nR}\}) \Longrightarrow \boxed{\mathcal{C}_n} \xrightarrow{u^n(W)} \boxed{P_{V|U}^{\otimes n}} \longrightarrow V^n \sim Q^{(\mathcal{C}_n)}_{V^n} \overset{TV}{\approx} P_V^{\otimes n}$$

Goal

Remark: Taking P_V as $P_V^{\otimes n}$ in the exact simulation setup we know that roughly $H(P_V^{\otimes n}) = n \cdot H(P_V)$ bits suffice to simulate $P_V^{\otimes n}$ exactly (i.e. taking $R > H(P_V)$ is enough). But here we will see that in the approximate setup we can get away with smaller R values.

More formally, consider the induced distribution (by the codebook) that corresponds to the random experiment:

$$Q^{(\mathcal{C}_n)}_{W,U^n,V^n}(w,u^n,v^n) = \frac{1}{|\mathcal{W}|} \mathbb{1}_{\{u^n(w)=u^n\}} P_{V|U}^{\otimes n}(v^n|u^n)$$

$\Rightarrow$ The induced output distribution is

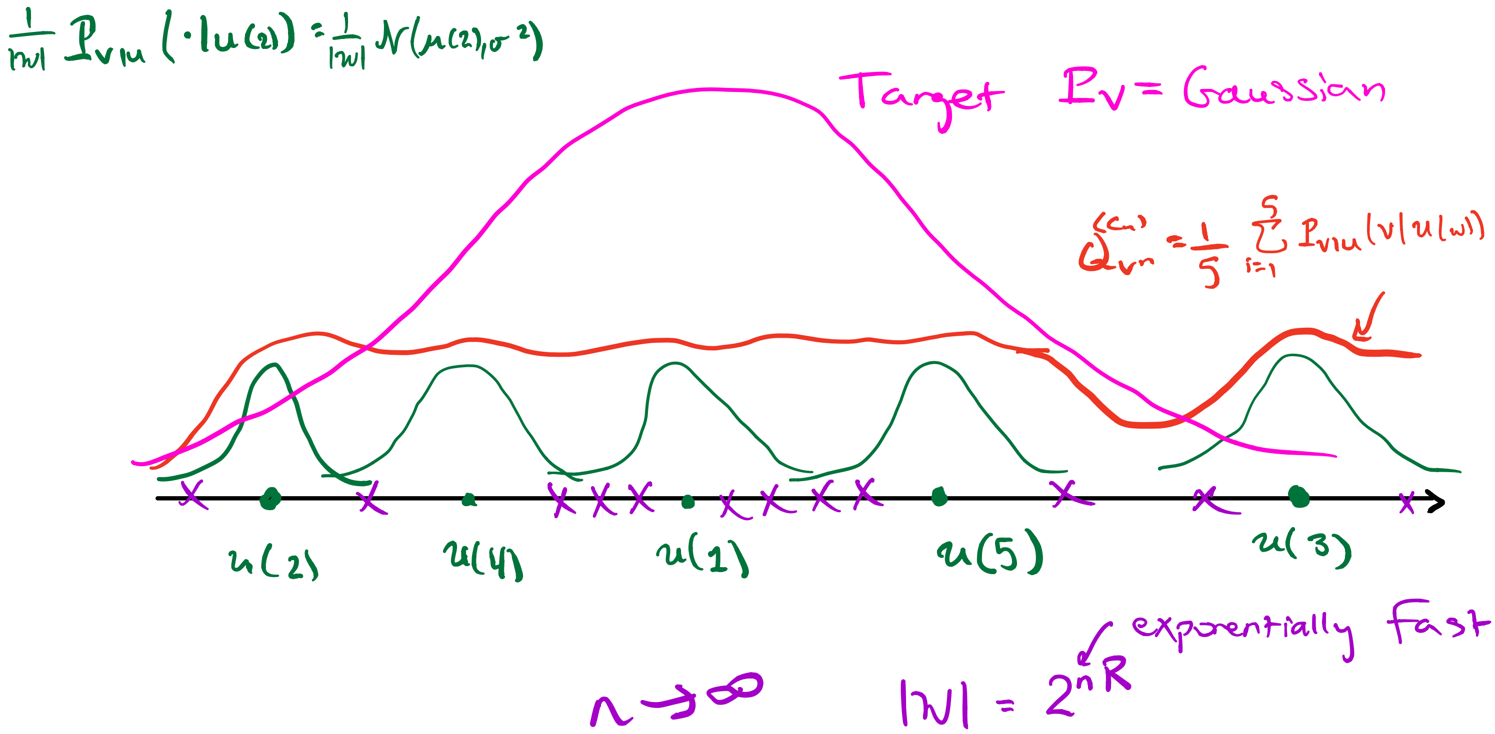
$$Q^{(\mathcal{C}_n)}_{V^n}(v^n) = \sum_{w\in\mathcal{W}} \sum_{u\in\mathcal{U}} Q_{W,U^n,V^n}^{(\mathcal{C}_n)}(w,u^n,v^n) = \frac{1}{|\mathcal{W}|}\sum_{w\in\mathcal{W}} P_{V|U}^{\otimes n}(v^n|u^n(w))$$

This is a mixture of conditional output distributions given each codeword.

$$\boxed{\text{Goal: Set } R \text{ and design } \mathcal{C}_n \text{ such that } \delta_{TV}(Q^{(\mathcal{C}_n)}_{V^n}, P_V^{\otimes n}) \xrightarrow{n\to\infty} 0}$$

What's going on here? Say $V_i = u_i + Z_i$, $Z_i \overset{iid}{\sim} \mathcal{N}(0,\sigma^2)$, is an AWGN channel. Assume $n=1$ for illustrative purposes and say that we want to simulate a Gaussian P_V. (Not formally correct but helps intuition)

The induced $Q^{(C_n)}_{V^n}$ distribution is thus a Gaussian mixture.



Clearly, the above is a bad approximation but the more well selected codewords we have and the larger the dimension becomes, the closer our mixture $Q^{(C_n)}_{V^n}$ and $P_V^{\otimes n}$ would be to one another.

How do we select / design the codebook? Seems the choice of $\mathcal{C}_n$ is quite crucial and we need to be careful about it.

However, it turns out that we can be pretty careless about how we pick it, essentially an iid random coding ensemble suffices.

$$\left[\begin{array}{l} \mathcal{C}_n \text{ iid} \qquad \mathbb{E}_{\mathcal{C}_n}\left[\delta_{TV}\left(Q_{V^n}^{(\mathcal{C}_n)}, P_V^{\otimes n}\right)\right] \xrightarrow{n\to\infty} 0 \\ \qquad \Rightarrow \exists\, \mathcal{C}_n \text{ deterministic w/ } \delta_{TV}\left(Q_{V^n}^{(\mathcal{C}_n)}, P_V^{\otimes n}\right) \xrightarrow{n\to\infty} 0 \end{array}\right]$$ what we will do

Picking $P_U \in \mathcal{P}(\mathcal{U})$ with $\sum_{u\in\mathcal{U}} P_U(u) P_{V|U}(\cdot|u) = P_V(\cdot)$ and drawing $\mathcal{C}_n$ iid according to P_U, i.e., $\mathcal{C}_n = \{u^n(w)\}_{w\in\mathcal{W}}$ with $u^n(w) \sim P_U^{\otimes n}$ and $(u^n(1), u^n(2), \ldots, u^n(2^{nR}))$ are mutually independent, we are extremely likely to get a good simulation codebook out of it so long that we set R correctly!

<u>Remark</u>: Another way to understand the approximate solution problem is using the "sparse mixture"

$$Q_{V^n}^{(\mathcal{C}_n)}(\cdot) = \frac{1}{|\mathcal{W}|}\sum_{w\in\mathcal{W}} P_{V|U}^{\otimes n}(\cdot|u^n(w))$$

to approximate the "dense mixture"

$$P_V^{\otimes n} = \sum_{u^n\in\mathcal{U}^n} P_U^{\otimes n}(u^n) P_{V|U}^{\otimes n}(\cdot|u^n)$$

Since $|\mathcal{W}| = 2^{nR}$ and $|\mathcal{U}^n| = |\mathcal{U}|^n = 2^{n \log|\mathcal{U}|}$, so long as $R < \log|\mathcal{U}|$ we see why "sparse" and "dense" terminology above makes sense.

Thus we expect $R < H(\text{Unif}(\mathcal{U}))$ for the above perspective to make sense.

As we will see next, we can get way below $\log|\mathcal{U}|$; in fact $R > I(U;V)$ will suffice and of course

$$I(U;V) \leq H(U) \leq \log|\mathcal{U}|$$

The next result, known in the IT literature as the "soft-covering lemma" (also sometimes referred to as "channel resolvability"), makes the above intuition and statements precise.

<u>**Theorem:**</u> Fix $P_U \in \mathcal{P}(\mathcal{U})$, $P_V \in \mathcal{P}(\mathcal{V})$ and let $P_{V|U}$ be a transition kernel from $(\mathcal{U}, 2^{\mathcal{U}})$ to $(\mathcal{V}, 2^{\mathcal{V}})$, with $\sum_{u \in \mathcal{U}} P_U(u) P_{V|U}(\cdot|u) = P_V(\cdot)$. Let $C_n := \{U^n(w)\}_{w \in \mathcal{W}}$, $\mathcal{W} = \{1, \ldots, 2^{nR}\}$, where $U^n(w) \sim P_U^{\otimes n}$ is iid across different $w \in \mathcal{W}$ values.

If $R > I(U;V)$ where the underlying distribution is $P_U P_{V|U}$, then

$$\mathbb{E}_{C_n}\left[\delta_{TV}\left(Q_{V^n}^{(C_n)}, P_V^{\otimes n}\right)\right] \xrightarrow{n \to \infty} 0$$

In fact, as a consequence we have that there exists a deterministic $\mathcal{C}_n = \{u^n(w)\}_{w\in\mathcal{W}}$, $u^n(w) \in \mathcal{U}^n$, such that

$$\delta_{TV}\left(Q_{V^n}^{(\mathcal{C}_n)}, P_V^{\otimes n}\right) \xrightarrow{n\to\infty} 0$$

In both the above statements, convergence happens exponentially fast, i.e., $\exists \gamma > 0$ such that for any n large enough, we have

$$\mathbb{E}_{\mathcal{C}_n}\left[\delta_{TV}\left(Q_{V^n}^{(\mathcal{C}_n)}, P_V^{\otimes n}\right)\right] \le e^{-\gamma n}$$

$$\delta_{TV}\left(Q_{V^n}^{(\mathcal{C}_n)}, P_V^{\otimes n}\right) \le e^{-\gamma n}$$

<u>Remark:</u> Recall that the approximate distribution strategy is

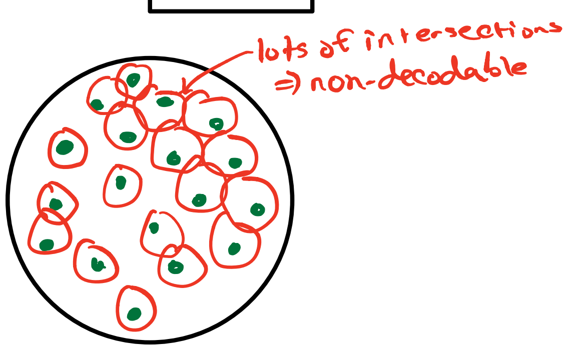
$$W\sim \mathrm{Unif}(\mathcal{W}) \longrightarrow \boxed{\mathcal{C}_n} \xrightarrow{u^n(W)} \boxed{P_{V|U}^{\otimes n}} \longrightarrow V^n \sim Q_{V^n}^{(\mathcal{C}_n)} \quad \longleftarrow \text{Have}$$

$R > I(U;V)$

Another way to understand the target distribution $P_V^{\otimes n}$ in the context of the above diagram is as the output of the channel when it is fed with an iid $U^n \sim P_U^{\otimes n}$:

$$U^n \sim P_U^{\otimes n} \longrightarrow \boxed{P_{V|U}^{\otimes n}} \longrightarrow V^n \sim P_V^{\otimes n} \quad \longleftarrow \text{Want}$$

$R > I(U;V)$



Discussion Section

$$R > I(U;V) \Rightarrow Q_V^{(C_n)} \approx P_V^{\otimes n}$$

Wyner (1975) $\mathbb{E}_{C_n} \frac{1}{n} D_{KL}\left(Q_{V^n}^{(C_n)} \| P_V^{\otimes n}\right) \xrightarrow{n\to\infty} 0$

Han-Verdu (1993): $\mathbb{E}_{C_n} \delta_{TV}\left(Q_{V^n}^{(C_n)}, P_V^{\otimes n}\right) \xrightarrow[n\to\infty]{} 0$

also provided a CONVERSE (Resolvability)

$\Uparrow$ Pinsker's Inequality

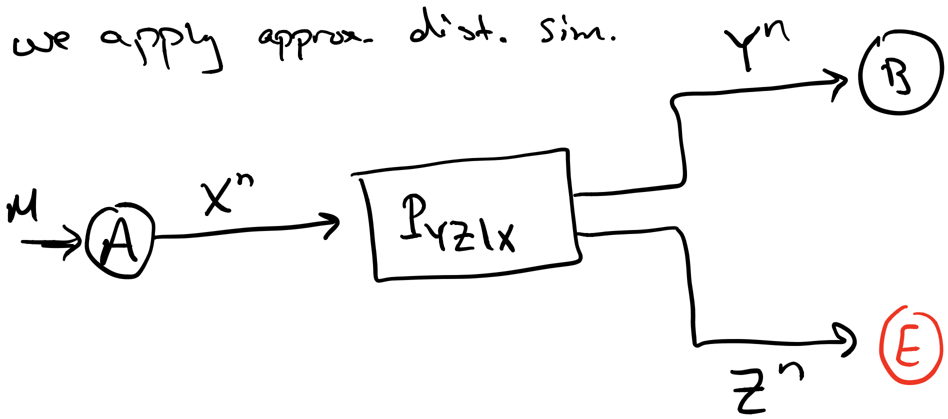
Hou-Kramer (2014): $\mathbb{E}_{C_n}\left[D_{KL}\left(Q_{V^n}^{(C_n)} \| P_V^{\otimes n}\right)\right] \xrightarrow{n\to\infty} 0$

Goldfeld-Cuff-Permuter (2017):

$$\mathbb{P}_{C_n}\left(D_{KL}\left(Q_{V^n}^{(C_n)} \| P_V^{\otimes n}\right) > e^{-\gamma_1 n}\right) \le e^{-e^{-\gamma_2 n}}$$

$$\Rightarrow \mathbb{P}\left(D_{KL}\left(Q_{V^n}^{(C_n)} \| P_V^{\otimes n}\right) \le e^{-\gamma_1 n}\right) \geq 1 - e^{-e^{-\gamma_2 n}}$$

How we apply approx. dist. sim.



$\Rightarrow \hat{m}$ ✓ w.h.p.

want

$I(M; Z^n)$ is small