(Jain, Radhakrishnan, Sen '03)

Rejection Sampling. [stat intuition]

Alice $p(x)$ ⟶ Bob $q(x)$ (known to both)

$D(p\|q) \leq$ small

Alice and Bob wish to sample $X \leftarrow p$ such that Alice communicates $\approx O(D(p\|q))$ bits.

NB: $q(x)$ should have a fair amount of mass where $p(x)$ is substantial.

Say $\exists$ Good $\subseteq [k]$ s.t. $p(\text{Good}) = 1-\varepsilon$,

$\text{supp}(p) = \text{supp}(q) = [k]$ $\quad \forall x \in \text{Good}, p(x) \leq 2^a \cdot q(x)$

Protocol: Bob samples iid samples from q

$X_1, X_2, \ldots$ Alice sends an integer $R \geq 0$

to Bob and Bob outputs X_R ($R=0$: fail).

Hope: Get R small.

For the moment suppose we can use public randomness.

Let $X \sim q$. Defined correlated $Z \in [k] \cup \{0, *\}$.

Def: $p'(x) \triangleq \begin{cases} p(x) & \text{if } x \in \text{Good} \\ 0 & \text{else.} \end{cases}$

$$\Pr(Z=j \mid X=i) = \begin{cases} \dfrac{p'(i)}{q(i)\,2^{a}} \to \gamma_i & j = i \\ 0 & j \neq i, *, 0 \\ \beta(1-\gamma_i) & j = 0 \\ \text{remaining} & j = * \end{cases}$$

$$\beta \triangleq \frac{\varepsilon\, 2^{-a}}{1-(1-\varepsilon)\,2^{-a}}.$$

$\xrightarrow{(i.i.d)}$

$$\begin{pmatrix} X_1 & X_2 & X_3 & \cdots \\ Z_1 & Z_2 & Z_3 \checkmark & \\ * & * & & \end{pmatrix}$$

$$R = \min \{ i : Z_i \neq * \}$$

R geometric with mean $\Pr(Z \neq *)$. $\quad p'(i)\,2^{-a}$

$$\Pr(Z \neq *) = \sum_{i \in [k]} q(i) \left(\gamma_i + \beta(1-\gamma_i) \right) = \beta + (1-\beta) \sum_i q(i)\,\gamma_i$$

$$= \beta + (1-\beta)(1-\varepsilon)\,2^{-a} = 2^{-a}.$$

$\Rightarrow \mathbb{E}(R) = 2^q$.

~~NB: Either $Y = X_R$ or $Y = 0$.~~

Def: $Z \triangleq Z_R$.

NB: either $Z = X_R$ or $Z = 0$.

Distribution of Z?

$$i \geq 0 \Rightarrow Pr(Z=i) = \sum_{r \geq 0} Pr(R=r) \cdot Pr(Z_r = i \mid R=r)$$

$$\overset{\text{independence}}{=} \sum_r Pr(R=r) \cdot Pr(Z_r = i \mid Z_r \neq \star)$$

$$= \sum_r Pr(R=r) \cdot \frac{Pr(Z_r = i)}{Pr(Z_r \neq \star)}$$

If $i \geq 1$,

$$= \sum_r Pr(R=r) \frac{Pr(X_r = i)\, Pr(Z_r = i \mid X_r = i)}{Pr(Z_r \neq \star)}$$

$$= \sum_r Pr(R=r) \frac{q(i)\, \gamma_i}{2^{-q}}$$

$$= p'(i) = \begin{cases} p(i) & i \in \text{Good} \\ 0 & \text{else} \end{cases}$$

(14)

if $i=0 \Rightarrow$

$$\Pr(Z_R=0) = \sum_{r\geq 0} \frac{\Pr(Z_r=0)}{\Pr(Z_r\neq *)}$$

$$= \sum_{r\geq 0} \Pr(R=r)\ \frac{\Pr(Z_r=0)}{\Pr(Z_r\neq *)}$$

$$= \sum_{r} \frac{\Pr(R=r)}{2^{-a}} \cdot \sum_{j\in[k]} \Pr(X_r=j)\ \Pr(Z_r=0 \mid X_r=j)$$

$$= \sum_{r} \frac{\Pr(R=r)}{2^{-a}} \left(\sum_j q(j) \cdot \frac{\beta(1-\gamma_j)}{2^{-a}} \right)$$

$$\frac{\varepsilon}{1-(1-\varepsilon)2^{-a}} \sum_j q(j) \cdot \left(1 - \frac{p'(j)2^{-a}}{q(j)}\right)$$

$$= \sum_r \Pr(R=r)\cdot \varepsilon$$

$$= \varepsilon.$$

$$1 - \underbrace{\sum_j p'(j)2^{-a}}_{(1-\varepsilon)}$$

Substate Theorem:

Thm: $p(x)$ and $q(x)$ on $[k]$, $\mathcal{D}(p\|q) = a$, $\forall r \geq 1$:

$$\text{Good} \triangleq \left\{ i \in [k] : p(i) \leq 2^{r(a+1)} q(i) \right\}.$$

$$\Rightarrow p(\text{Good}) \geq 1 - \frac{1}{r}.$$

Proof: $\text{Bad} \triangleq [k] - \text{Good}$.

$$P_r(\text{Good}) \log \frac{P(\text{Good})}{q(\text{Good})} + P_r(\text{Bad}) \log \frac{P(\text{Bad})}{q(\text{Bad})} \leq \mathcal{D}(p\|q) = a$$

why:

$$\left(\sum p_i\right) \log \frac{\sum p_i}{\sum q_i} \leq \sum \left(p_i \log \frac{p_i}{q_i}\right)$$

Because the inequality is invariant under scaling of the $q_i \Rightarrow$ wleg $\sum p_i = \sum q_i$ and we reduce to positivity of $\mathcal{D}(\|)$.

$$\forall i \in \text{Bad}, \frac{p(i)}{q(i)} > 2^{r(a+1)}$$

$$\Rightarrow p(\text{Bad}) > 2^{r(a+1)} q(\text{Bad})$$

$$P(\text{Good}) \log \frac{P(\text{Good})}{q(\text{Good})} \geq p(\text{Good}) \log p(\text{Good}) \geq -1$$

$$\Rightarrow \left[x \in [0,1] \Rightarrow x \log x \geq -\frac{\log e}{e} > -1 \right]$$

$$\Rightarrow a \geq -1 + P_0(Bad) \log \left(\frac{P(Bad)}{q(Bad)} \right) \geq 2^{r(a+1)}$$

$$\Rightarrow a+1 \geq P(Bad) \cdot r(a+1)$$

$$\Rightarrow P(Bad) \leq \frac{1}{r}.$$

□ (Substack)

~~Corollary: Distributions $q, P_1, \ldots, P_N$ on $[k]$~~

~~$\exists S \left(P_i \| q \right) = q_i \quad \forall \varepsilon \in (0,1]$~~

Compression of Protocols:

Starting point: 1-way, private coins error δ.

Alice X —M→ Bob Y $\quad (X,Y) \sim \mu$.

Suppose $IC^{ext} = I(X;M) \leq a$. We want to construct a protocol with $O(a)$ communication and error $\leq \delta + \varepsilon$ on μ.

$I(X;M) \leq a$ → conditional dist. of message.

$$I(X;M) = \mathop{\mathbb{E}}_{x \leftarrow \mu} D(M_x \| M) \leq a.$$

~~$\Rightarrow \Pr_x \left(\mathbb{E} D(M_x \| M) > \frac{a}{\varepsilon} \right) < \varepsilon$.~~

(7)

Apply substate lemma on $p \leftarrow M_X$ and $q \leftarrow M$.

set $Pr(Good) = \varepsilon \Rightarrow \forall i \in Good,\ M_X(i) \leq 2^{\frac{D(M_X \| M)+1}{\varepsilon}} M(i)$.

Now apply Restriction Sampling

Using public randomness, Alice sends a sample of M_X via integer R s.t. $E(R) \leq 2^{\frac{D(M_X \| M)+1}{\varepsilon}}$.

$\Rightarrow E(\#bits) = E(\log R) \leq \log E(R) \leq \frac{D(M_X \| M)+1}{\varepsilon}$

Protocol fails if $R=0$, $Pr(R=0) \leq \varepsilon$.

Conditioned on $R \neq 0$, Alice and Bob sample same point in $Good_X$ according to M_X, and the effect of the original protocol is simulated.

$\Rightarrow E_X(\#bits) \leq E_X(D(M_X \| M)+1)/\varepsilon \leq \frac{a+1}{\varepsilon}$

Markov: $Pr(\#bits > \frac{(a+1)}{\varepsilon^2}) \leq \varepsilon$.

Fail if the communication is too long $\Rightarrow$ error $\leq 2\varepsilon + \delta$.

Finally, fix public randomness to get a deterministic protocol.

Multiple (k) rounds.

Start with $\Pi'_{k+1} := \Pi$. Inductively, construct Π'_i from Π'_{i+1}, where Π'_i behaves like Π up to round $(i-1)$, and compresses round $i \to k$.

Suppose the i^{th} message in Π'_i is (to be) sent by Alice.

We compress the i^{th} round of Π'_{i+1} as follows.

$M :=$ transcript for rounds $1 \to i$.

round $1 \to i-1$ ← → round i.

$M = (M_1, M_2)$

$$I(X,Y;M) = I(X,Y;M_1) + \mathop{\mathbb{E}}_{m_1 \leftarrow M_1} I(X,Y;M_2 \mid M_1 = m_1)$$

$$= I(X,Y;M_1) + \mathop{\mathbb{E}}_{M_1,X,Y} \mathcal{D}\left(M_2^{xym_1} \,\|\, M_2^{m_1}\right).$$

($M_2^{xym_1}$: cond. dist. of M_2)

NB: $M_2^{xym_1} = M_2^{xm_1}$ (indep. of y).

Def: $a_i \triangleq \mathop{\mathbb{E}}_{M,XY} \mathcal{D}\left(M_2^{xym_1} \,\|\, M_2^{m_1}\right)$.

We use fresh public randomness for each round, and also for each choice of m_1.

NB: Both parties know m_1.

As before, we use rejection sampling to sample from $M_2^{x m_1}$. But we set $\Pr\left(\text{Good}_{x m_1} \in M_2^{x m_1}\right) \geq 1 - \varepsilon/k$.

$$\mathbb{E}\ \#\text{bits sent} \leq \frac{k}{\varepsilon} \mathcal{D}\left(M_2^{xym_1} \,||\, M_2^{m_1}\right)$$

$$\Rightarrow \begin{cases} \mathop{\mathbb{E}}_{X,Y}\ \#\text{bits sent} \leq O\left(\frac{k}{\varepsilon}(a_i + 1)\right) \\ \Pr(\text{failure in round } i) \leq \varepsilon/k. \end{cases}$$

$$\star \quad \sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} I\left(X,Y ; M_2^{(i)} \,\middle|\, M_1^{(i)}\right)$$

$$\overset{\text{Chain rule}}{=} I(X,Y; M) \leq a.$$

$$\Rightarrow \mathop{\mathbb{E}}_{X,Y}\left(\text{total \# bits in } k \text{ rounds}\right) = O\left(\frac{k}{\varepsilon}(a+k)\right)$$

Use Markov to restrict total # bits to $O\left(\frac{k}{\varepsilon^2}(a+k)\right)$

and error prob $\leq \delta + 2\varepsilon$.

□