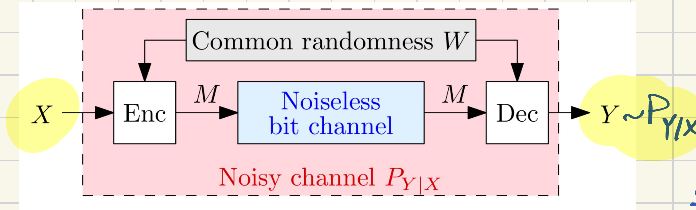
3.3 Exponential and Poisson Functional Representation

Recap:

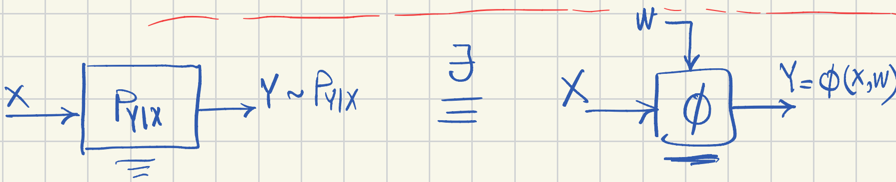
⓪ Channel simulation:



$L := E[|M|]$

goal: $L^* = \inf L$

①



Functional Representation Lemma

- $X \perp\!\!\!\perp W$
- $Y \sim P_{Y|X}$

② Using Functional Representation Lemma in channel simulation setup.

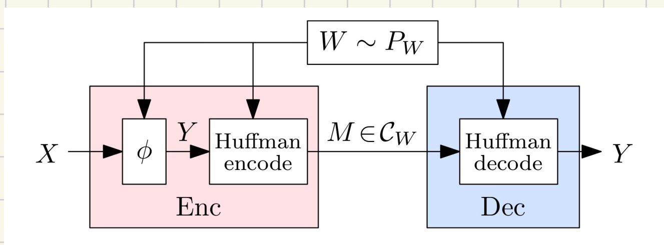


Figure 3.2: Functional representation and channel simulation.

$L := E[|M|]$

goal: $L^* := \inf E[|M|]$

Theorem 4:
$$I(X;Y) \leq L^* \leq I(X;Y) + \log_2\left(I(X;Y)+2\right) + 3$$

③
$$H(Y|W) \leq E\left[|M|\right] \leq H(Y|W) + 1$$

let $H^* := \inf H(Y|W)$ then

$$I(X;Y) \leq H^* \leq L^* \leq H^* + 1$$

④

Classical Rejection Sampling: $H(Y|W) \leq -\log_2 \gamma - \frac{1-\gamma}{\gamma}\log_2(1-\gamma)$

Greedy Rejection Sampling: $H(Y|W) \leq I(X;Y) + \log_2\left(I(X;Y) + \log_2 4e\right) + \log_2 4e$

Discrete $Y \in \{1, 2, \dots, n\}$: Gumbel-Max Trick

goal: to generate a sample Y.

In contrast to rejection sampling, instead of generating candidates $\bar{Y}_1, \bar{Y}_2, \cdots$ we generate time for each element of $\{1, 2, \dots, n\} \ni Y$.

- time $\equiv Z_i \sim \mathrm{Exp}(r_i)$

review of Exp R.V.:
- $Z \sim \mathrm{Exp}(r)$, $Z \in \mathbb{R}^+$, $\frac{Z}{a} \sim \mathrm{Exp}(ar)$ $\forall a > 0$
- $Z_1 \sim \mathrm{Exp}(r_1)$, $Z_2 \sim \mathrm{Exp}(r_2)$ $\Rightarrow \min(Z_1, Z_2) \sim \mathrm{Exp}(r_1 + r_2)$
- " , " $\Rightarrow P(Z_1 < Z_2) = \frac{r_1}{r_1 + r_2}$

- let times $Z_1, Z_2, \dots, Z_n \overset{\text{indep}}{\sim} \mathrm{Exp}(r_i)$ correspond to $\{1, 2, \dots, n\}$.
- What is $P\left(\operatorname{argmin}_i \frac{Z_i}{a_i/r_i} = y\right) = ?$

$$\frac{Z_i}{a_i/r_i} \sim \mathrm{Exp}\left(r_i \times \frac{a_i}{r_i}\right) = \mathrm{Exp}(a_i)$$

$$P\left(\operatorname{argmin}_i \frac{Z_i}{a_i/r_i} = y\right) = P\left(\frac{Z_y}{a_y/r_y} = \min_{i \in \{1,\dots,n\}} \left\{\frac{Z_i}{a_i/r_i}\right\}\right)$$

all expcept $\frac{Z_y}{a_y/r_y}$

$$= P\Big(\underbrace{\frac{Z_y}{a_y/r_y}}_{\sim \mathrm{Exp}(a_y)} \leq \underbrace{\min\left\{\frac{Z_1}{a_1/r_1}, \frac{Z_2}{a_2/r_2}, \dots, \frac{Z_n}{a_n/r_n}\right\}}_{\sim \mathrm{Exp}\left(\sum_{i \neq y} a_i\right)}\Big)$$

$$= \frac{a_y}{a_y + \sum_{i \neq y} a_i} = \frac{a_y}{\sum a_i}$$

$$P\left(\operatorname{argmin}_i \frac{Z_i}{a_i/r_i} = y\right) = a_y$$

- if $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ a probability vector, then $\operatorname{argmin}_i \frac{Z_i}{a_i/r_i} =: \phi(Z_i)$ gives a sample of the probability vector. $\Rightarrow$ $\boxed{\text{Gumbel-Max Trick}}$

Exponential Functional Representation Scheme

1. $Y \in \{1, 2, \ldots, n\}$

2. $W = (Z_1, Z_2, \ldots, Z_n)$ where $Z_y \sim \mathrm{Exp}(Q(y))$ with Q some reference dist.

3. $$Y = \underset{y}{\operatorname{argmin}} \frac{Z_y}{P_{Y|X}(y|x) / Q(y)} \quad \text{(given } X\text{)}$$

4. sort Z_y's and send Z_Y's rank $(:= K)$ to decoder.

   e.g. let $Y = 10$ and sorted times $Z_3, Z_{10}, Z_1, Z_7, Z_9, Z_4, \ldots$

   then $K = 2$.

5. The decoder finds Z_Y as the K^{th} element in the sorted times.

   And output its index Y.

Question: Why not send Y itself in the encoder?

Elias-delta code $\longrightarrow$

+ the smaller Z_y's are more likely to be chosen in the argmin.

Number	N	N+1	δ encoding
$1 = 2^0$	0	1	1
$2 = 2^1 + 0$	1	2	0 1 0 0
$3 = 2^1 + 1$	1	2	0 1 0 1
$4 = 2^2 + 0$	2	3	0 1 1 00
$5 = 2^2 + 1$	2	3	0 1 1 01
$6 = 2^2 + 2$	2	3	0 1 1 10
$7 = 2^2 + 3$	2	3	0 1 1 11
$8 = 2^3 + 0$	3	4	00 1 00 000
$9 = 2^3 + 1$	3	4	00 1 00 001
$10 = 2^3 + 2$	3	4	00 1 00 010
$11 = 2^3 + 3$	3	4	00 1 00 011
$12 = 2^3 + 4$	3	4	00 1 00 100
$13 = 2^3 + 5$	3	4	00 1 00 101
$14 = 2^3 + 6$	3	4	00 1 00 110
$15 = 2^3 + 7$	3	4	00 1 00 111
$16 = 2^4 + 0$	4	5	00 1 01 0000
$17 = 2^4 + 1$	4	5	00 1 01 0001

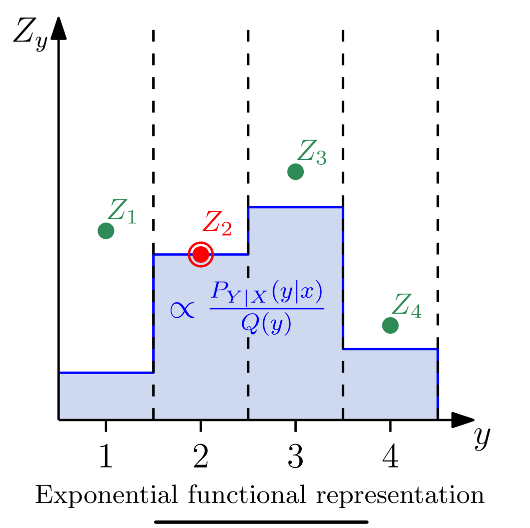


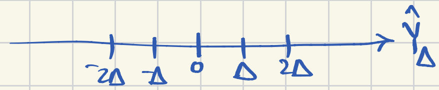
Figure 3.5: Left: an illustration of the exponential functional representation, where $\mathcal{Y} = \{1, 2, 3, 4\}$. For the argmin in (3.13), note that $\min_y Z_y/(P_{Y|X}(y|x)/Q(y)) = \min\{\gamma > 0 : \exists y : Z_y = \gamma P_{Y|X}(y|x)/Q(y)\}$, and hence the argmin in (3.13) can be regarded as having a shape $\gamma P_{Y|X}(y|x)/Q(y)$ (blue shape in the figure), which we keep "inflating" this shape by increasing γ until it hits the first point Z_y, and the point it hits corresponds to the Y selected in (3.13) ($Y = 2$ in the figure).

General Y : Poisson Functional Representation (assumed $Y \in \{1, \dots, n\}$)

Intuition: Limiting argument of Exponential Functional Representation

- $Y \in \mathbb{R}$ and Q a continuous distribution

Quantized Y : $\hat{Y}_\Delta = \Delta \left\lfloor \frac{Y}{\Delta} + \frac{1}{2} \right\rfloor \in \Delta \mathbb{Z}$



$\hat{Y}_\Delta \sim \hat{P}_\Delta$ when $Y \sim P$

$\hat{Q}_\Delta$

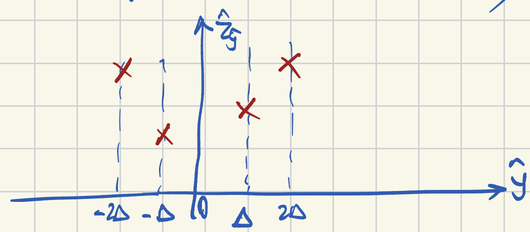
$\hat{Z}_{\hat{y}} \sim \mathrm{Exp}(\hat{Q}_\Delta(y))$

the rule: $\underset{\hat{y} \in \Delta\mathbb{Z}}{\mathrm{argmin}} \dfrac{\hat{Z}_{\hat{y}}}{\hat{P}_\Delta(\hat{y}) / \hat{Q}_\Delta(\hat{y})} \xrightarrow{\Delta \to 0} \underset{y \in \mathbb{R}}{\mathrm{argmin}} \dfrac{Z_y}{(dP/dQ)(y)}$?

ratio of densities (Radon Nikodym derivative)

- How about Z_y ? limit of stochastic process $\hat{Z}_{\hat{y}} \sim \mathrm{Exp}(\hat{Q}_\Delta(\hat{y}))$

plot of $(\hat{y}, \hat{Z}_{\hat{y}}) \in \mathbb{R}^2$



each line: only one point $\Rightarrow$ density of point in each line $\xrightarrow{\Delta \to 0} 0$

$E[\hat{Z}_{\hat{y}}] = \dfrac{1}{\hat{Q}_\Delta(\hat{y})} \xrightarrow{\Delta \to 0} \infty$

$\overset{?}{\Rightarrow} (\hat{y}, \hat{Z}_{\hat{y}})$ has a limiting distribution.

Sort pairs $(\hat{y}, \hat{Z}_{\hat{y}})$ in $\hat{Z}_{\hat{y}}$ and call it $(\bar{Y}_i, T_i)$ s.t. $T_1 \leq T_2 \leq T_3 \leq \dots$

- $T_1 = \min_{\hat{y}} \hat{Z}_{\hat{y}} = \mathrm{Exp}\left(\sum_{\hat{y}} Q(\hat{y})\right) = \mathrm{Exp}(1)$
- Consider waiting time $T_2 - T_1 \sim \mathrm{Exp}\left(\sum_{\hat{y} \neq \bar{y}_1} Q(\hat{y})\right) = \mathrm{Exp}(1 - \hat{Q}_\Delta(\bar{y}_1)) \approx \mathrm{Exp}(1)$
  (conditioned on $(\bar{Y}_1, T_1) = (\bar{y}_1, t_1)$)

$\vdots$

$$\left.\begin{aligned} &T_1, T_2 - T_1, T_3 - T_2, T_4 - T_3, \dots \overset{iid}{\approx} \mathrm{Exp}(1) \\ &\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \dots \overset{iid}{\approx} Q(y) \end{aligned}\right\} \text{Poisson Process.}$$

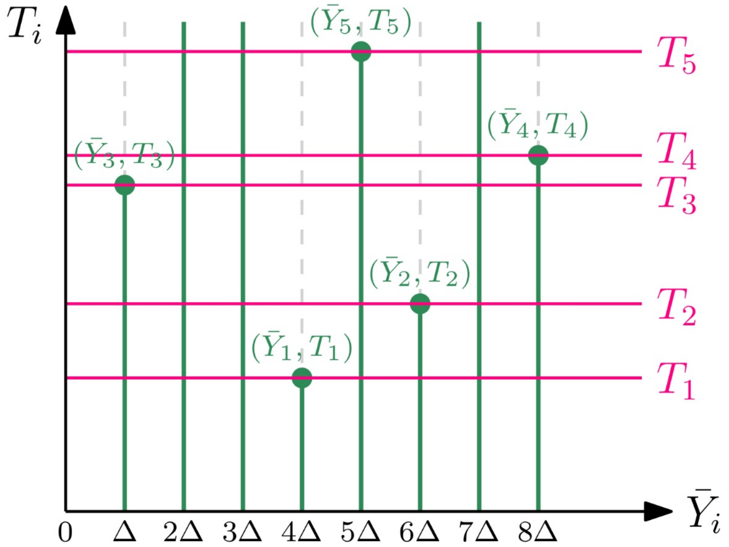


Figure 3.6: An illustration of the points $(\hat{y}, \hat{Z}_{\hat{y}})$ where $\hat{y} \in \Delta\mathbb{Z}$, $\hat{Z}_{\hat{y}} \sim \mathrm{Exp}(\hat{Q}_\Delta(\hat{y}))$, sorted in ascending order of the second coordinate to become $(\bar{Y}_i, T_i)$ with $T_1 \leq T_2 \leq \cdots$. The arrival time along each vertical line at $\hat{y}$ is $\mathrm{Exp}(\hat{Q}_\Delta(\hat{y}))$. At the beginning, we are waiting for any one of all the lines, and the waiting time is $T_1 \sim \mathrm{Exp}(\sum_{\hat{y}} \hat{Q}_\Delta(\hat{y})) = \mathrm{Exp}(1)$. After the first arrival, we are waiting for all except one line, and hence $T_2 - T_1$ should be approximately $\mathrm{Exp}(1)$ as well.

$\longrightarrow$ Another viewpoint: point process of pairs $(\bar{Y}_i, T_i)$

Thm 13. (Strong Functional Representation Lemma)

$\forall P_X, P_{Y|X} \quad \exists \quad (P_W, \phi) \quad$ s.t. $\quad X \perp\!\!\!\perp W \quad$ and $\quad Y = \phi(X, W) \sim P_{Y|X}$

such that

$$H(Y|W) \leq I(X;Y) + \log_2\big(I(X;Y) + 2\big) + 2 .$$

proof. First the following Lemma:

Lem 12. (Poisson Functional Representation)

For $P_Y \ll Q_Y$ and $g(y) := \frac{dP_Y}{dQ_Y}(y)$, let $0 \leq T_1 \leq T_2 < \cdots$ be a poission point process of rate 1; Also $\bar{Y}_1, \bar{Y}_2, \ldots \overset{iid}{\sim} Q_Y$ and $Y = \bar{Y}_K$, where $K := \underset{i}{\operatorname{argmin}} \frac{T_i}{g(\bar{Y}_i)}$.

Then $Y \sim P_Y$. And we have

$$E[\log_2 K] \leq D_{KL}(P_Y \| Q_Y) + 1 .$$

let $Q = P_Y$ (marginal of $P_X P_{Y|X}$).

let $T_1, T_2, \ldots \sim PP(1)$ and $\bar{Y}_1, \bar{Y}_2, \ldots \overset{iid}{\sim} Q$.

let $g(y|x) := \frac{dP_{Y|X}(\cdot|x)}{dP_Y}(y)$.

then $Y = \bar{Y}_K$ where $K := \underset{i}{\operatorname{argmin}} \frac{T_i}{g(\bar{Y}_i|X)}$

By (Lem 12) : $\left(E[\log_2 K \mid X=x]\right) \leq \left(D_{KL}(P_{Y|X}(\cdot|x) \| P_Y) + 1\right)$

$E_x[\cdot]$ (E_x ... E_x

$$E[\log_2 K] \leq I(X;Y) + 1$$

let $\underline{W} = \{(\bar{Y}_1, T_1), (\bar{Y}_2, T_2), \ldots\}$

$$H(Y|W) \leq \boxed{H(K|W)} \leq H(K)$$

how?

$$\boxed{\begin{aligned} H(Y,K|W) &= H(K|W) \\ H(Y,K|W) &\geq H(Y|W) \end{aligned}} + \underbrace{H(Y|K,W)}_{\text{zero}}$$

$$\boxed{H(K) \leq E[\log_2 K] + \log_2(E[\log K] + 1) + 1}$$

For any RV $K \in \{1, 2, \ldots\}$

$$\leq I(X;Y) + 1 + \log_2(I(X;Y) + 2) + 1$$

Compare:

Classical Rejection Sampling: $H(Y|W) \leq -\log_2 \gamma - \frac{1-\gamma}{\gamma} \log_2(1-\gamma)$

Greedy Rejection Sampling: $H(Y|W) \leq I(X;Y) + \log_2(I(X;Y) + \log_2 4e) + \log_2 4e$

Poisson Functional Rep. Sampling Scheme: $H(Y|W) \leq I(X;Y) + \log_2(I(X;Y) + 2) + 2$