1. Show that $(\mathbf{x}, \mathbf{y}) \in T^n_{[X,Y]\delta}$ and $(\mathbf{y}, \mathbf{z}) \in T^n_{[Y,Z]\delta}$ do not imply $(\mathbf{x}, \mathbf{z}) \in T^n_{[X,Z]\delta}$

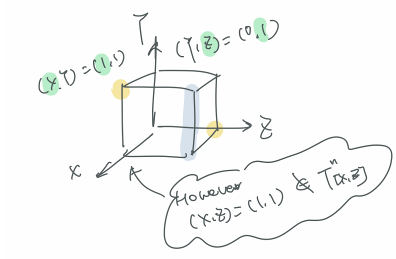

Figure 15: The illustrated random variables (X, Y, Z) gives a counter example.

2. Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$, where X_k are i.i.d. with generic random variable X. Prove that

$$\Pr\left\{\mathbf{X} \in T^n_{[X]\delta}\right\} \geq 1 - \frac{|\mathcal{X}|^3}{n\delta^2}$$

for any n and $\delta > 0$. This shows that $\Pr\left\{\mathbf{X} \in T^n_{[X]\delta}\right\} \to 1$ as $\delta \to 0$ and $n \to \infty$ if $\sqrt{n}\delta \to \infty$

Answer:

$$\begin{aligned}
\Pr\left\{\mathbf{X} \in T^n_{[X]\delta}\right\} &= \Pr\left\{\sum_x \left|\frac{1}{n}N(x;\mathbf{X}) - p(x)\right| \leq \delta\right\} \\
&= 1 - \Pr\left\{\sum_x \left|\frac{1}{n}N(x;\mathbf{X}) - p(x)\right| > \delta\right\} \\
&\geq 1 - \Pr\left\{\left|\frac{1}{n}N(x;\mathbf{X}) - p(x)\right| > \frac{\delta}{|\mathcal{X}|} \text{ for some } x \in \mathcal{X}\right\}
\end{aligned}$$

$$\begin{aligned}
\Pr\left\{\left|\frac{1}{n}N(x;\mathbf{X}) - p(x)\right| > \frac{\delta}{|\mathcal{X}|} \text{ for some } x\right\} &= \Pr\left\{\left|\frac{1}{n}\sum_{k=1}^{n} B_k(x) - p(x)\right| > \frac{\delta}{|\mathcal{X}|} \text{ for some } x\right\} \\
&= \Pr\left\{\bigcup_x \left\{\left|\frac{1}{n}\sum_{k=1}^{n} B_k(x) - p(x)\right| > \frac{\delta}{|\mathcal{X}|}\right\}\right\} \\
&\leq \sum_x \Pr\left\{\left|\frac{1}{n}\sum_{k=1}^{n} B_k(x) - p(x)\right| > \frac{\delta}{|\mathcal{X}|}\right\} \\
&\leq \sum_x \Pr\left\{|B_k(x) - p(x)| > \frac{\delta}{|\mathcal{X}|}\right\} \qquad \text{(Since i.i.d)} \\
&< \sum_x \frac{\sigma^2|\mathcal{X}|^2}{n\delta^2} \qquad \text{(Chebyshev's inequality)} \\
&< \frac{|\mathcal{X}|^3}{n\delta^2}
\end{aligned}$$

Using Chebyshev's inequality

$$\mathrm{P}\left(\left|\bar{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Therefore we proved

$$\Pr\left\{\mathbf{X} \in T^n_{[X]\delta}\right\} \geq 1 - \frac{|\mathcal{X}|^3}{n\delta^2}$$

4. Prove Proposition 6.13. Hint: Use the fact that if $(\mathbf{X}, \mathbf{Y}) \in T^n_{[XY]\delta}$, then $\mathbf{X} \in S^n_{[X]\delta}$

Proposition 6.13. With respect to a joint distribution $p(x, y)$ on $\mathcal{X} \times \mathcal{Y}$, for any $\delta > 0$

$$\Pr\left\{\mathbf{X} \in S^n_{[X]\delta}\right\} > 1 - \delta$$

for n sufficiently large.

Answer: By Theorem 6.7. If $(\mathbf{x}, \mathbf{y}) \in T^n_{[XY]\delta}$, then $\mathbf{x} \in T^n_{[X]\delta}$

$$\begin{aligned}\sum_x \left|\frac{1}{n}N(x; \mathbf{x}) - p(x)\right| &= \sum_x \left|\frac{1}{n}\sum_y N(x, y; \mathbf{x}, \mathbf{y}) - \sum_y p(x, y)\right| \\ &\sum_x \left|\sum_y \left(\frac{1}{n}N(x, y; \mathbf{x}, \mathbf{y}) - p(x, y)\right)\right| \\ &\leq \sum_x \sum_y \left|\frac{1}{n}N(x, y; \mathbf{x}, \mathbf{y}) - p(x, y)\right|\end{aligned}$$

By Weak Law of Large Number, as $n \to \infty$,

$$\Pr\left\{\left|\frac{1}{n}N(x, y; \mathbf{x}, \mathbf{y}) - p(x, y)\right| > \frac{\delta}{|\mathcal{X}||\mathcal{Y}|}\right\} < \frac{\delta}{|\mathcal{X}||\mathcal{Y}|}$$

Then

$$\sum_x \left|\frac{1}{n}N(x; \mathbf{x}) - p(x)\right| < \delta$$

$$\Pr\left\{\mathbf{X} \in S^n_{[X]\delta}\right\} > 1 - \delta$$

6. Let p be any probability distribution over a finite set $\mathcal{X}$ and η be a real number in $(0, 1)$. Prove that for any subset A of $\mathcal{X}^n$ with $p^n(A) \geq \eta$,

$$\left|A \cap T^n_{[X]\delta}\right| \geq 2^{n\left(H(p) - \delta'\right)}$$

where $\delta' \to 0$ as $\delta \to 0$ and $n \to \infty$

Answer: Recall for $\mathbf{x} \in T_{[X]\delta}$,

$$2^{-n(H(X)+\delta)} \leq p(\mathbf{x}) \leq 2^{-n(H(X)-\delta)}$$

Then for $\mathbf{y} \in A \cap T_{[X]\delta}$, we can write $p(\mathbf{y}) = \eta p(\mathbf{x})$
(Note that given in question, $p(A) \geq \eta$, for simplicity, we fix $p(A) = \eta, \eta \in (0, 1)$)
By using the upper bound inequality $p(\mathbf{x}) \leq 2^{-n(H(X)-\delta)}$

$$p(\mathbf{y}) \leq \eta 2^{-n(H(X)-\delta)} = 2^{-n(\frac{\log(\eta)}{n} + H(X) - \delta)}$$

As $n \to \infty$, $\frac{\log(\eta)}{n} + H(X) - \delta \to H(X) - \delta$, because $\lim_{n\to\infty} \frac{\log(\eta)}{n} = 0$ Therefore

$$\left|A \cap T^n_{[X]\delta}\right| 2^{-n(H(X)-\delta')} \geq \Pr\{A \cap T^n_{[X]\delta}\} \geq 1 - \delta$$

and

$$\left|A \cap T^n_{[X]\delta}\right| \geq 2^{n(H(X)-\delta')}$$

where $\delta' \to 0$ as $\delta \to 0$ and $n \to \infty$

In the following problems, for a sequence $\mathbf{x} \in \mathcal{X}^n$, let $q_{\mathbf{x}}$ be the empirical distribution of $\mathbf{x}$, i.e., $q_{\mathbf{x}}(x) = n^{-1}N(x;\mathbf{x})$ for all $x \in \mathcal{X}$. Similarly, for a pair of sequences $(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, let $q_{\mathbf{x},\mathbf{y}}$ be the joint empirical distribution of $(\mathbf{x},\mathbf{y})$, i.e., $q_{\mathbf{x},\mathbf{y}}(x,y) = n^{-1}N(x,y;\mathbf{x},\mathbf{y})$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$

7. Alternative definition of strong typicality. Show that (6.1) is equivalent to

$$V\left(q_{\mathbf{x}}, p\right) \leq \delta$$

where $V(\cdot,\cdot)$ denotes the variational distance. Thus strong typicality can be regarded as requiring the empirical distribution of a sequence to be close to the probability distribution of the generic random variable in variational distance. Also compare the result here with the alternative definition of weak typicality (Problem 5 in Chapter 5).

Answer:

$$V\left(q_{\mathbf{x}}, p\right) = \sum_{x\in\mathcal{X}} \left|q_{\mathbf{x}} - p(x)\right| = \sum_{x\in\mathcal{X}} \left|\frac{1}{n}N(x;\mathbf{x}) - p(x)\right| \leq \delta$$

We can see that strong typicality is stronger than weak in the sense that weak typicality only requires the closeness in entropy. Note that $d(q_x, p) = \left|D\left(q_{\mathbf{x}}\|p\right) + H\left(q_{\mathbf{x}}\right) - H(p)\right| = 0$ when $q_x = p$

8. The empirical distribution $q_{\mathbf{x}}$ of the sequence $\mathbf{x}$ is also called the type of x. Assuming that $\mathcal{X}$ is finite, show that there are a total of $\binom{n+|\mathcal{X}|-1}{n}$ distinct types $q_{\mathbf{x}}$. Hint: There are $\binom{a+b-1}{a}$ ways to distribute a identical balls in b boxes.

Answer: The original problem may be reformulated as arranging $k-1$ bars and the n balls, by selecting n positions for balls out of $n+k-1$ locations.

$$\underbrace{***}_{n \text{ balls}} \quad \underbrace{[||||||]}_{k-1 \text{ bars}}$$

Directly apply this idea to get emperical distribution q_x, treat as for assigning each sample x into $|\mathcal{X}|$ boxes. Which gives the result

$$\binom{n+|\mathcal{X}|-1}{n}$$