Deriving Cramér type bound directly : (1)

$(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is the space. $\alpha \in M_1(\mathbb{R}^d)$ is a fixed distribution so that $\alpha(e_i) = p_i$ with $p_i > 0$ for $i = 1, \dots, d$ & $\sum_{i=1}^{d} p_i = 1$. $e_i = (0, \dots, 1, 0 \dots)$ (i^{th})

Given such $\alpha \in M_1(\mathbb{R}^d)$, we can have iid RV's (X_n) taking values in $\mathbb{R}^d$ with distribution μ_n of $\frac{X_1 + \dots + X_n}{n}$. This, written in a formula means

$$\mu_n\left(\left(\frac{n_1}{n}, \frac{n_2}{n}, \dots, \frac{n_d}{n}\right)\right) = \frac{n!}{n_1! \cdots n_d!} p_1^{n_1} \cdots p_d^{n_d}$$

where $n_1 + n_2 + \dots + n_d = n$ & $0 \le n_i$ $i = 1 \dots d$

For simplicity, take $d = 2$.

Then $\mu_n\left(\left(\frac{k}{n}, \frac{n-k}{n}\right)\right) = \frac{n!}{k!\,(n-k)!} p_1^k p_2^{n-k}$ $\quad 0 \le k \le n$.

Stirling $\Rightarrow$ $n! \simeq n^n e^{-n} \sqrt{2\pi n}$ as $n \to \infty$.

Now, if we impose k to grow like, $\boxed{k \simeq yn}$ as $n \to \infty$, for some $0 < y < 1$, then

$$k! \simeq (yn)^{yn} e^{-yn} \sqrt{2\pi yn}$$ and similarly

$$(n-k)! \simeq (n(1-y))^{n(1-y)} e^{-n(1-y)} \sqrt{2\pi n(1-y)}.$$

This means, as $n \to \infty$

$$\frac{n!}{k!\,(n-k)!}\, p_1^{k}\, p_2^{n-k}$$

$$\simeq \frac{n^n e^{-n} \sqrt{2\pi n}}{(ny)^{ny} e^{-ny} \sqrt{2\pi n y}\; (n(1-y))^{n(1-y)} e^{-n(1-y)} \sqrt{2\pi n(1-y)}} \times p_1^{ny}\, p_2^{n(1-y)}$$

$$\Rightarrow \lim_{n\to\infty} \frac{1}{n} \log \mu_n\left(\left(\frac{k}{n}, \frac{n-k}{n}\right)\right)$$

$$= -y\log y - (1-y)\log(1-y) + y \log p_1 + (1-y)\log p_2$$

$$= y \log \frac{p_1}{y} + (1-y) \log \frac{p_2}{1-y}.$$

ie $\lim_{n\to\infty} \frac{1}{n} \log \mu_n(\bar{y}) = -\Lambda^*(\bar{y})$

where for $\bar{y} = (y_1, \ldots, y_d)$

$$\Lambda^*(\bar{y}) := \begin{cases} \sum_{i=1}^{2} y_i \log \frac{y_i}{p_i} & \text{if } y_i \geq 0 \text{ \& } \sum_{i=1}^{2} y_i = 1 \\ \infty & \text{else} \end{cases}$$

What did we learn from this boring calculation ?

In other words, μ_n satisfies a LDP with rate function $\Lambda^*(y)$ which is the relative entropy of the Binomial $(y, 1-y)$ with respect to Binomial $(p, 1-p)$.

This is not a coincidence. Now consider any $d \in \mathbb{N}$. Then

$$\mu_n\left(\left(\frac{n_1}{n}, \dots, \frac{n_d}{n}\right)\right) = \frac{n!}{n_1! \dots n_d!} p_1^{n_1} \dots p_d^{n_d}.$$

Again assuming $n_i \simeq n y_i$ with $\sum_{i=1}^{d} y_i = 1$ $0 \leq y_i \leq 1$

we can check, using stirling like before,

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(\bar{y}) = -\Lambda^*(\bar{y}) \qquad \bar{y} = (y_1 \dots y_d)$$

where
$$\Lambda^*(\bar{y}) = \begin{cases} \sum_{i=1}^{d} y_i \log \frac{y_i}{p_i} & \text{if } \sum_{i=1}^{d} y_i = 1 \ \& \ 0 \leq y_i \leq 1 \\ \infty & \text{else} \end{cases}$$

ie rate function is the relative entropy of the multinomial $(y_1, \dots, y_d)$ with respect to the multinomial $(p_1, \dots, p_d)$.

This calculation again tells us something:

Recall that ~~any probab~~ (from your first course in probability) any probability distribution can be approximated by a probability distribution which is concentrated on a finite set (it's finite dimensional cylinder sets!). The "empirical distribution" from a sample of size n (think of $\frac{\delta_{x_1} + \cdots + \delta_{x_n}}{n}$) will have then a multinomial distribution (Why? Think about an iid structure!).

Just like before, if $\alpha \in M_1(\mathbb{R}^d)$ was the original distribution, then the probability that the empirical distribution $\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ of n independent observations from a common distribution α is close to another distribution β should satisfy

$$\mathbb{P}\left(\frac{1}{n}\sum \delta_{x_i} \sim \beta\right) \simeq \exp\left\{-n\, H(\beta\,;\alpha)\right\} \quad n \to \infty$$

where
$$H(\beta,\alpha) = \int \log \frac{d\beta}{d\alpha}\, d\beta$$

(by Radon-Nikodym derivative formula for change of variables)

$$= \int \frac{d\beta}{d\alpha} \log \frac{d\beta}{d\alpha}\, d\alpha$$

What is this? Recall Sanov's theorem proven in the lecture!