## 3.2 The Gibbs Conditioning Principle for Finite alphabets

Let $Y_1, Y_2, \dots$ be a sequence of i.i.d random variables with strictly positive law $\mu \in \mathcal{M}_1(E)$ and having values in the finite alphabet $E$.

Furthermore, let $X_k = f(Y_k)$ for some deterministic function $f : E \to \mathbb{R}$.

Given a set $A \subset \mathbb{R}$ and a constraint of the type $1/n \sum_{i=1}^{n} X_i \in A$, we are interested in the conditional law of $Y_1$ when $n$ is large.

That is, what are the limit points, as $n \to \infty$, of $\mu_n^*$ defined by

$$\mu_n^*(x) = P_\mu\left(Y_1 = x \mid 1/n \sum_{i=1}^{n} X_i \in A\right);$$

$x \in E$.

$$\hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \langle f, L_n^Y \rangle, \text{ where } f = (f(x))_{x \in E};$$

and note that under the conditioning $\hat{S}_n \in A$, $Y_j$ are identically distributed, although not independent.

Therefore, for every function $\phi : E \to \mathbb{R}$,

$$\langle \phi, \mu_n^* \rangle = \mathbb{E}[\phi(Y_1) \mid \hat{S}_n \in A] = \mathbb{E}[\phi(Y_2) \mid \hat{S}_n \in A]$$

$$= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \phi(Y_i) \mid \hat{S}_n \in A\right] = \mathbb{E}[\langle \phi, L_n^Y \rangle \mid \langle f, L_n^Y \rangle \in A],$$

where $\phi = (\phi(x))_{x\in E}$. Hence, with $\Gamma := \{\nu \in \mathcal{M}_1(E) : \langle f, \nu\rangle \in A\}$

$$\mu_n^* = \mathbb{E}[L_n^Y \mid L_n^Y \in \Gamma]$$

Pick a nonempty set $\Gamma$ for which

$$\inf_{\nu\in\Gamma^o} H(\nu|\mu) = \inf_{\nu\in\overline{\Gamma}} H(\nu|\mu) =: I_\Gamma \quad (*).$$

<u>Theorem 15: (Gibbs principle)</u>

Let $\mathcal{M} := \{\nu \in \overline{\Gamma} : H(\nu|\mu) = I_\Gamma\}$ (set of minimizers)

(a) All the limit points of $(\mu_n^*)_{n\geq 1}$ belong to $\overline{co}(\mathcal{M})$ – the closure of the convex hull of $\mathcal{M}$.

(b) When $\Gamma$ is convex with $\Gamma^o \neq \phi$, the set $\mathcal{M}$ consists of a single point to which $\mu_n^*$ converge as $n\to\infty$.

<u>Remarks</u>: (1) Assume $E_\mu = E$ (support set of the law $\mu$) and that $\Gamma$ is a convex subset of $\mathcal{M}_1(E)$ with $\Gamma^o \neq \phi$.
One can show that $\lim_{n\to\infty} 1/n \log P_\mu(L_n^Y \in \Gamma)$
$= -\inf_{\nu\in\Gamma} H(\nu|\mu) = -I_\Gamma$ and $I_\Gamma = H(\nu^*|\mu)$ for ~~some~~ a unique $\nu^* \in \overline{\Gamma^o}$ (Hint: continuity of $H(\cdot|\mu)$ on the compact set $\overline{\Gamma^o}$; $\Gamma \subset \overline{\Gamma^o}$, and prove $H(\cdot|\mu)$ strictly convex on $\mathcal{M}_1(E)$).

(2) Let $(Y_n)_{n\geq 1}$ be i.i.d Bernoulli $(\frac{1}{2})$ and $X_i = Y_i$, and take $\Gamma = [0,\alpha] \cup [1-\alpha, 1]$ for some small $\alpha > 0$.

$\mathcal{M}$ consists of the probability distributions $\{B(\alpha), B(1-\alpha)\}$, however, the symmetry of the problem implies that the only possible limit point of $(\mu_n^*)_{n\geq 1}$ is the probability distribution $B(1/2)$.

Hence, the limit distribution may be in $co(\mathcal{M})$ and not just in $\mathcal{M}$.

<u>Proof:</u> $\overline{\Gamma}$ is compact and thus $\mathcal{M}$ is non-empty.

For every $\mathcal{U} \subset \mathcal{M}_1(E)$,

$$\mathbb{E}[L_n^Y \mid L_n^Y \in \Gamma] - \mathbb{E}[L_n^Y \mid L_n^Y \in \mathcal{U} \cap \Gamma]$$

$$= P_\mu(L_n^Y \in U^c \mid L_n^Y \in \Gamma)\{\mathbb{E}[L_n^Y \mid L_n^Y \in \mathcal{U}^c \cap \Gamma]$$

$$- \mathbb{E}[L_n^Y \mid L_n^Y \in U \cap \Gamma]\}.$$

Recall that the conditional expectation $\mathbb{E}[L_n^Y \mid L_n^Y \in \mathcal{U} \cap \Gamma]$ ~~belongs to~~ belongs to $co(\mathcal{U})$, and that $\mu_n^* = \mathbb{E}[L_n^Y \mid L_n^Y \in \Gamma]$.

Therefore,

$$d_V(\mu_n^*, co(\mathcal{U})) \leq P_\mu(L_n^Y \in \mathcal{U}^c \mid L_n^Y \in \Gamma) d_V(\mathbb{E}[L_n^Y \mid L_n^Y \in \mathcal{U}^c \cap \Gamma],$$

$$\mathbb{E}[L_n^Y \mid L_n^Y \in \mathcal{U} \cap \Gamma]) \leq P_\mu(L_n^Y \in \mathcal{U}^c \mid L_n^Y \in \Gamma), \text{ as } d_V(\cdot,\cdot) \leq 1.$$

$\mathcal{M}^\delta := \{\nu \in \mathcal{M}_1(E) : d_V(\nu, \mathcal{M}) < \delta\}$. We are going to prove below that for every $\delta > 0$,

$$(+) \quad \lim_{n\to\infty} P_\mu (L_n^Y \in \mathcal{M}^\delta \mid L_n^Y \in \Gamma) = 1$$

Using the previous estimate and (+) we get $d_V(\mu_n^*, co(\mathcal{M}^\delta)) \to 0$ as $n \to \infty$. Since $d_V$ is convex, each point in $co(\mathcal{M}^\delta)$ is within variational distance $\delta$ of some point in $co(\mathcal{M})$.
Thus we conclude with (a) as $\delta > 0$ was arbitrary.
We are left to show (+):
Sanov's theorem gives

$$I_\Gamma = -\lim_{n\to\infty} \frac{1}{n} \log P_\mu (L_n^Y \in \Gamma)$$

and

$$\limsup_{n\to\infty} 1/n \log P_\mu (L_n^Y \in (\mathcal{M}^\delta)^c \cap \Gamma) \leq - \inf_{\nu \in (\mathcal{M}^\delta)^c \cap \Gamma} H(\nu|\mu)$$

$$\leq - \inf_{\nu \in (\mathcal{M}^\delta)^c \cap \overline{\Gamma}} H(\nu|\mu)$$

. $(\mathcal{M}^\delta)^c \cap \overline{\Gamma}$ are compact sets as $\mathcal{M}^\delta$ are open sets. Thus, for some $\overline{\nu} \in (\mathcal{M}^\delta)^c \cap \overline{\Gamma}$,

$$\inf_{\nu \in (\mathcal{M}^\delta)^c \cap \overline{\Gamma}} H(\nu|\mu) = H(\overline{\nu}|\mu) > I_\Gamma .$$

Now

$$\limsup_{n\to\infty} 1/n \log P_\mu (L_n^Y \in (\mathcal{M}^\delta)^c \mid L_n^Y \in \Gamma) =$$

$$= \limsup_{n\to\infty} \left\{ \frac{1}{n} \log P_\mu(L_n^Y \in (\mathcal{M}^\delta)^c \cap \Gamma) - \frac{1}{n} \log P_\mu(L_n^Y \in \Gamma) \right\} < 0.$$

■

The preceding theorem holds whenever $\Gamma$ satisfies (*).
The particular set $\{\nu : \langle f, \nu\rangle \in A\}$ has an important significance in statistical mechanics because it represents an energy-like constraint.
Let $A \subset \mathbb{R}$ be nonempty, convex, open subset of $K$, the interval supporting $(f(x))_{x\in E}$, that is w.l.o.g. $f(1) \le \cdots \le f(N)$, $E = \{1, \dots, N\}$, $K = [f(1), f(N)]$.

Then the unique limit of $\mu_n^*$ is of the form

$$\nu_\lambda(x) = \mu(x)\, e^{\lambda f(x) - \Lambda(\lambda)}$$

for some chosen $\lambda \in \mathbb{R}$, which is called the Gibbs parameter associated with $A$. We shall elaborate on this a bit and provide some hints for the proof of this assertion.
$\Gamma = \{\nu : \langle f, \nu\rangle \in A\}$; $\mathcal{M} = \{\nu \in \bar{\Gamma} : H(\nu|\mu) = I_\Gamma\}$.
$A \subset \mathbb{R}$ open implies that $\Gamma$ is open. For every $\lambda \in \mathbb{R}$, let
$\Lambda(\lambda) = \log \sum_{x\in E} \mu(x)\, e^{\lambda f(x)}$. For any $\nu \in \mathcal{M}_1(E)$ and

every $\lambda \in \mathbb{R}$, Jensen's inequality implies that

$$\Lambda(\lambda) \ge \sum_{x\in E} \nu(x) \log \frac{\mu(x)\, e^{\lambda f(x)}}{\nu(x)} = \lambda\langle f, \nu\rangle - H(\nu|\mu),$$

with equality for $\nu_\lambda \in \mathcal{M}_1(E)$ with $\nu_\lambda(x) = \mu(x) e^{\lambda f(x) - \Lambda(\lambda)}$.

Thus,

$$\Lambda(\lambda) \geq \lambda \langle f, \nu \rangle - H(\nu | \mu),$$

$$H(\nu|\mu) \geq \lambda \langle f, \nu \rangle - \Lambda(\lambda) \quad \text{and}$$

$$\lambda x - \Lambda(\lambda) \leq \inf_{\substack{\nu \in \mathcal{M}_1(E): \\ \langle f, \nu \rangle = x}} H(\nu | \mu) = I(x), \text{ where}$$

$$I(x) = \sup_{\lambda \in \mathbb{R}} \{ \lambda x - \Lambda(\lambda) \}$$ is the rate function of

Cramér's theorem, where $\lim_{n \to \infty} \frac{1}{n} P_\mu(\hat{S}_n \in A) = -\inf_{x \in A} I(x)$.

The function $\Lambda(\lambda)$ is differentiable with $\Lambda'(\lambda) = \langle f, \nu_\lambda \rangle$.