In this lecture, we expose the technique of deriving concentration inequalities with the entropy tensorization inequality. The entropy tensorization inequality enables us to bound the entropy of a function of n variables by the sum of the n entropies of this function in terms of the individual variables. The second step of this technique uses the variational formulation of the n entropies to form a differential inequality that gives an upper bound of the log-Laplace transform of the function. We can subsequently use Markov inequality to get a deviation inequality involving this function.

Let $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ be a measurable space, and $u : \mathcal{X} \to \mathbb{R}^+$ a measurable function. The **entropy** of u with regard to $\mathbb{P}$ is defined as $\mathrm{Ent}_{\mathbb{P}}(u) \stackrel{\text{def.}}{=} \int u \log u d\mathbb{P} - \int u \cdot \left(\log\left(\int u d\mathbb{P}\right)\right) d\mathbb{P}$. If $\mathbb{Q}$ is another probability measure and $u = \frac{d\mathbb{Q}}{d\mathbb{P}}$, then $\mathrm{Ent}_{\mathbb{P}}(u) = \int \left(\log \frac{d\mathbb{Q}}{d\mathbb{P}}\right) d\mathbb{Q}$ is the **KL-divergence** between two probability measures $\mathbb{Q}$ and $\mathbb{P}$.

The following lemma gives variational formulations for the entropy.

Lemma 40.1.

$$\begin{aligned} Ent_{\mathbb{P}}(u) &= \inf\left\{\int (u \cdot (\log u - \log x) - (u - x))\, d\mathbb{P} : x \in \mathbb{R}^+\right\} \\ &= \sup\left\{\int (u \cdot g)\, d\mathbb{P} : \int \exp(g) d\mathbb{P} \le 1\right\}. \end{aligned}$$

Proof. For the first formulation, we define x pointsizely by $\frac{\partial}{\partial x}\int (u \cdot (\log u - \log x) - (u - x))\, d\mathbb{P} = 0$, and get $x = \int u d\mathbb{P} > 0$.

For the second formulation, the Laplacian corresponding to $\sup\left\{\int (u \cdot g)\, d\mathbb{P} : \int \exp(g) d\mathbb{P} \le 1\right\}$ is $\mathcal{L}(g, \lambda) = \int (ug)\, d\mathbb{P} - \lambda\left(\int \exp(g) d\mathbb{P} - 1\right)$. It is linear in λ and concave in g, thus $\sup_g \inf_{\lambda \ge 0} \mathcal{L} = \inf_{\lambda \ge 0} \sup_g \mathcal{L}$. Define g pointwisely by $\frac{\partial}{\partial g}\mathcal{L} = u - \lambda \exp(g) = 0$. Thus $g = \log \frac{u}{\lambda}$, and $\sup_g \mathcal{L} = \int \left(u \log \frac{u}{\lambda}\right) d\mathbb{P} - \int u d\mathbb{P} + \lambda$. We set $\frac{\partial}{\partial \lambda}\sup_g \mathcal{L} = -\frac{\int u d\mathbb{P}}{\lambda} + 1 = 0$, and get $\lambda = \int u d\mathbb{P}$. As a result, $\inf_\lambda \sup_g \mathcal{L} = \mathrm{Ent}_{\mathbb{P}}(u)$. ☐

Entropy $\mathrm{Ent}_{\mathbb{P}}(u)$ is a convex function of u for any probability measure $\mathbb{P}$, since

$$\begin{aligned} \mathrm{Ent}_{\mathbb{P}}\left(\sum \lambda_i u_i\right) &= \sup\left\{\int \left(\sum \lambda_i u_i \cdot g\right) d\mathbb{P} : \int \exp(g) d\mathbb{P} \le 1\right\} \\ &\le \sum \lambda_i \sup\left\{\int (u_i \cdot g_i)\, d\mathbb{P} : \int \exp(g_i) d\mathbb{P} \le 1\right\} \\ &= \sum \lambda_i \mathrm{Ent}_{\mathbb{P}}(u_i). \end{aligned}$$

Lemma 40.2. *[Tensorization of entropy]* $\mathcal{X} = (\mathcal{X}_1, \cdots, \mathcal{X}_n)$, $\mathbb{P}^n = \mathbb{P}_1 \times \cdots \times \mathbb{P}_n$, $u = u(x_1, \cdots, x_n)$, $Ent_{\mathbb{P}^n}(u) \le \int \left(\sum_{i=1}^n Ent_{\mathbb{P}_i}(u)\right) d\mathbb{P}^n$.

Proof. Proof by induction. When $n = 1$, the above inequality is trivially true. Suppose

$$\int u \log u d\mathbb{P}^n \le \int u d\mathbb{P}^n \log \int u d\mathbb{P}^n + \int \sum_{i=1}^n \mathrm{Ent}_{\mathbb{P}_i}(u) d\mathbb{P}^n.$$

Integrate with regard to $\mathbb{P}_{n+1}$,

$$
\begin{aligned}
& \int u \log u d\mathbb{P}^{n+1} \\
\leq \quad & \int \left(\overbrace{\int u d\mathbb{P}^n}^{v} \log \overbrace{\int u d\mathbb{P}^n}^{v} \right) d\mathbb{P}_{n+1} + \int \sum_{i=1}^{n} \mathrm{Ent}_{\mathbb{P}_i}(u) d\mathbb{P}^{n+1} \\
\underbrace{=}_{\text{definition of entropy}} \quad & \int \overbrace{\int u d\mathbb{P}^n}^{v} d\mathbb{P}_{n+1} \cdot \left(\log \int \overbrace{\int u d\mathbb{P}^n}^{v} d\mathbb{P}_{n+1} \right) + \mathrm{Ent}_{\mathbb{P}_{n+1}} \left(\overbrace{\int u d\mathbb{P}^n}^{v} \right) + \int \sum_{i=1}^{n} \mathrm{Ent}_{\mathbb{P}_i}(u) d\mathbb{P}^{n+1} \\
\underbrace{=}_{\text{Foubini's theorem}} \quad & \int u d\mathbb{P}^{n+1} \cdot \left(\log \int u d\mathbb{P}^{n+1} \right) + \mathrm{Ent}_{\mathbb{P}_{n+1}} \left(\int u d\mathbb{P}^n \right) + \int \sum_{i=1}^{n} \mathrm{Ent}_{\mathbb{P}_i}(u) d\mathbb{P}^{n+1} \\
\underbrace{\leq}_{\text{convexity of entropy}} \quad & \int u d\mathbb{P}^{n+1} \cdot \left(\log \int u d\mathbb{P}^{n+1} \right) + \int \mathrm{Ent}_{\mathbb{P}_{n+1}}(u) d\mathbb{P}^n + \int \sum_{i=1}^{n} \mathrm{Ent}_{\mathbb{P}_i}(u) d\mathbb{P}^{n+1} \\
= \quad & \int u d\mathbb{P}^{n+1} \cdot \left(\log \int u d\mathbb{P}^{n+1} \right) + \int \mathrm{Ent}_{\mathbb{P}_{n+1}}(u) d\mathbb{P}^{n+1} + \int \sum_{i=1}^{n} \mathrm{Ent}_{\mathbb{P}_i}(u) d\mathbb{P}^{n+1} \\
\leq \quad & \int u d\mathbb{P}^{n+1} \cdot \left(\log \int u d\mathbb{P}^{n+1} \right) + \int \sum_{i=1}^{n+1} \mathrm{Ent}_{\mathbb{P}_i}(u) d\mathbb{P}^{n+1}.
\end{aligned}
$$

By definition of entropy, $\mathrm{Ent}_{\mathbb{P}^{n+1}}(u) \leq \int \sum_{i=1}^{n+1} \mathrm{Ent}_{\mathbb{P}_i}(u) d\mathbb{P}^{n+1}$. $\square$

The tensorization of entropy lemma can be trivially applied to get the following tensorization of Laplace transform.

Theorem 40.3. *[Tensorization of Laplace transform] Let $x_1, \cdots, x_n$ be independent random variables and $x'_1, \cdots, x'_n$ their indepent copies, $Z = Z(x_1, \cdots, x_n)$, $Z^i = Z(x_1, \cdots, x_{i-1}, x'_i, x_{i+1}, \cdots, x_n)$, $\phi(x) = e^x - x - 1$, and $\psi(x) = \phi(x) + e^x \phi(-x) = x \cdot (e^x - 1)$, and I be the indicator function. Then*

$$
\begin{aligned}
\mathbb{E}\left(e^{\lambda Z} \cdot \lambda Z\right) - \mathbb{E}e^{\lambda Z} \cdot \log \mathbb{E}e^{\lambda Z} \quad & \leq \quad \mathbb{E}_{x_1, \cdots, x_n, x'_1, \cdots, x'_n} e^{\lambda Z} \sum_{i=1}^{n} \phi\left(-\lambda(Z - Z^i)\right) \\
\mathbb{E}\left(e^{\lambda Z} \cdot \lambda Z\right) - \mathbb{E}e^{\lambda Z} \cdot \log \mathbb{E}e^{\lambda Z} \quad & \leq \quad \mathbb{E}_{x_1, \cdots, x_n, x'_1, \cdots, x'_n} e^{\lambda Z} \sum_{i=1}^{n} \psi\left(-\lambda(Z - Z^i)\right) \cdot I(Z \geq Z^i).
\end{aligned}
$$

Proof. Let $u = \exp(\lambda Z)$ where $\lambda \in \mathbb{R}$, and apply the tensorization of entropy lemma,

$$\begin{aligned}
& \underbrace{\mathbb{E}\left(e^{\lambda Z} \cdot \lambda Z\right) - \mathbb{E}e^{\lambda Z} \cdot \log \mathbb{E}e^{\lambda Z}}_{\mathrm{Ent}_{\mathbb{P}^n} \log u} \\
\leq \; & \mathbb{E} \sum_{i=1}^{n} \mathrm{Ent}_{\mathbb{P}_i} e^{\lambda Z} \\
\underbrace{=}_{\text{variational formulation}} \; & \mathbb{E} \sum_{i=1}^{n} \inf \left\{ \int \left(e^{\lambda Z}(\lambda Z - \lambda x) - (e^{\lambda Z} - e^{\lambda x})\right) d\mathbb{P}_i : x \in \mathbb{R}^+ \right\} \\
\leq \; & \mathbb{E} \sum_{i=1}^{n} \mathbb{E}_{x_i x_i'} \left(e^{\lambda Z}(\lambda Z - \lambda Z^i) - (e^{\lambda Z} - e^{\lambda Z^i}) \right) \\
= \; & \mathbb{E} \sum_{i=1}^{n} \mathbb{E}_{x_i x_i'} e^{\lambda Z} \left(e^{-\lambda(Z - Z^i)} - \left(-\lambda \cdot \left(Z - Z^i\right)\right) - 1 \right) \\
= \; & \mathbb{E}_{x_1, \cdots, x_n, x_1', \cdots, x_n'} e^{\lambda Z} \sum_{i=1}^{n} \phi\left(-\lambda \cdot \left(Z - Z^i\right)\right).
\end{aligned}$$

Moreover,

$$\begin{aligned}
& \mathbb{E}e^{\lambda Z} \sum_{i=1}^{n} \phi\left(-\lambda \cdot \left(Z - Z^i\right)\right) \\
= \; & \mathbb{E} \sum_{i=1}^{n} e^{\lambda Z} \phi\left(-\lambda \cdot \left(Z - Z^i\right)\right) \cdot \left(\underbrace{I\left(Z \geq Z^i\right)}_{\mathbf{I}} + \underbrace{I\left(Z^i \geq Z\right)}_{\mathbf{II}} \right) \\
= \; & \mathbb{E} \sum_{i=1}^{n} \left(\underbrace{e^{\lambda Z^i} \phi\left(-\lambda \cdot \left(Z^i - Z\right)\right) \cdot I\left(Z \geq Z^i\right)}_{\text{switch } Z \text{ and } Z^i \text{ in } \mathbf{II}} + \underbrace{e^{\lambda Z} \phi\left(-\lambda \cdot \left(Z - Z^i\right)\right) \cdot I\left(Z \geq Z^i\right)}_{\mathbf{I}} \right) \\
= \; & \mathbb{E} \sum_{i=1}^{n} e^{\lambda Z} \cdot I\left(Z \geq Z^i\right) \cdot \left(\underbrace{e^{\lambda(Z^i - Z)} \cdot \phi\left(-\lambda \cdot \left(Z^i - Z\right)\right)}_{\mathbf{II}} + \underbrace{\phi\left(-\lambda \cdot \left(Z - Z^i\right)\right)}_{\mathbf{I}} \right) \\
= \; & \mathbb{E} \sum_{i=1}^{n} e^{\lambda Z} \cdot I\left(Z \geq Z^i\right) \cdot \psi\left(-\lambda \cdot \left(Z^i - Z\right)\right).
\end{aligned}$$

□