Cramér from Sanov

Proposition 0.1. *Let $\mathcal{X} \subset \mathbb{R}$ be compact, and $\nu \in \mathcal{P}(\mathcal{X})$. Then:*

$$\inf_{\substack{\rho \in \mathcal{P}(\mathcal{X}): \\ \int x\rho(dx) = y}} \mathcal{H}(\rho \mid \nu) = \sup_{\lambda \in \mathbb{R}} y\lambda - \log \int e^{x\lambda}\, \nu(dx). \tag{1}$$

Proof by large deviations. Take iid RV's $X_1, X_2, \ldots$, and let $\phi : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ be the continuous bounded function $\phi(\rho) := \int x\, \rho(dx)$, so that $\phi(L^n) = \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n} S_n$, the empirical average. On the one hand, by Sanov's Theorem the empirical measure $L^n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ satisfies an LDP with good RF $\rho \mapsto \mathcal{H}(\rho \mid \nu)$, and so by the Contraction Principle $\frac{1}{n}S_n$ satisfies an LDP with good RF given by the left-hand side of (1). On the other hand, by Cramér's Theorem, $\frac{1}{n}S_n$ satisfies an LDP with good RF given by the right-hand side of (1). The statement follows by uniqueness of rate functionals. □

Proof by variational calculus. The left-hand side of (1) is given as a contrained minimisation problem in the Banach space of bounded signed measures, under the contraints that $\rho \geq 0$, $\int x\rho(dx) = y$ and $\rho(\mathcal{X}) = 1$. We may neglect the first constraint, since $\mathcal{H}(\rho \mid \nu)$ (and its derivative) automatically blows up when $\rho \ngeq 0$. The other constraints are $\mathbb{R}$-valued equalities, which we pair with two Lagrange multipliers $\lambda_1, \lambda_2 \in \mathbb{R}$. By compactness, lower semicontinuity and boundedness from below of the relative entropy, there exists a minimiser, say $\bar\rho$. Under sufficient regularity of this minimiser (which needs to be shown by an approximation argument), the Kuhn-Tucker Theorem says that the following Gateaux derivative must be zero in the minimiser $\bar\rho$, i.e.:

$$0 = D_\rho\Big[\mathcal{H}(\bar\rho \mid \nu) + \lambda_1\big(y - \textstyle\int x\, \bar\rho(dx)\big) + \lambda_2\big(1 - \textstyle\int \bar\rho(dx)\big)\Big](x) = \log\Big(\frac{d\bar\rho}{d\nu}(x)\Big) - \lambda_1 x - \lambda_2,$$

and so

$$\bar\rho(dx) = \frac{e^{\lambda_1 x}\, \nu(dx)}{\phi}, \qquad \phi := e^{-\lambda_2}. \tag{2}$$

The two Lagrange multipliers λ_1, λ_2 must be chosen such that the two contraints are satisfied, from which we deduce that

$$\phi = \phi(\lambda_1) = \int e^{\lambda_1 x}\, \nu(dx), \qquad \text{(the MGF from Cramér!)},$$
$$y = \int x\bar\rho(dx) = \frac{\int x e^{\lambda_1 x}\, \nu(dx)}{\phi(\lambda_1)} = \frac{\phi'(\lambda_1)}{\phi(\lambda_1)}.$$

From the last line it follows that λ_1 is optimal in $\sup_{\lambda \in \mathbb{R}} \lambda y - \log\phi(\lambda)$. Plugging in the optimal measure (2) in the relative entropy yields:

$$\inf_{\substack{\rho \in \mathcal{P}(\mathcal{X}): \\ \int x\rho(dx) = y}} \mathcal{H}(\rho \mid \nu) = \mathcal{H}(\bar\rho \mid \nu) = \lambda_1 \frac{\phi'(\lambda_1)}{\phi(\lambda_1)} - \log\phi(\lambda_1) = \lambda_1 y - \log\phi(\lambda_1) = \sup_{\lambda \in \mathbb{R}} \lambda y - \log\phi(\lambda).$$

□