# ldp-cramer

zhaof17

May 2021

## 1 Cramér theorem for discrete random variable

Using Sanov's theorem to prove Cramér theorem for discrete random variables.

Suppose $X_i$ are i.i.d. distribution ($\sim P_X$) from $\Omega \to \mathcal{X} \subset \mathbb{R}$ where $|\mathcal{X}| < \infty$. Denote $x_{\min} = \min \mathcal{X}$ and $x_{\max} = \max \mathcal{X}$. The log-MGF for $X$ is defined as $\psi_X(t) = \log M_X(t) = \log \mathbb{E}[\exp(tX)]$. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. For any set $A \in \mathbb{R}$,

$$-\inf_{t \in A^o} I(t) \leq \liminf_{n\to\infty} \frac{1}{n} \log P(\bar{X}_n \in A) \leq \limsup_{n\to\infty} \frac{1}{n} \log P(\bar{X}_n \in A) \leq -\inf_{t\in A} I(t) \tag{1}$$

where $A^o$ is the interior of $A$. The rate function $I(t)$ is continuous at $t \in [x_1, x_M]$ and satisfies

$$I(t) = \psi_X^*(t) := \sup_{\lambda\in\mathbb{R}} \{\lambda t - \psi_X(\lambda)\} \tag{2}$$

That is, $I(t)$ is equal to the conjugate transpose of $\psi_X(t)$.

*Proof.* From Sanov's theorem, we have

$$\begin{aligned} -\inf_{Q_X \in \Gamma^o} D(Q_X||P_X) &\leq \liminf_{n\to\infty} \frac{1}{n} \log P(\hat{P}_{X^n} \in \Gamma) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log P(\hat{P}_{X^n} \in \Gamma) \leq -\inf_{Q_X\in\Gamma} D(Q_X||P_X) \end{aligned} \tag{3}$$

where $\Gamma := \{Q_X : \mathbb{E}_{Q_X}[X] \in A\}$. It is obvious that $\bar{X}_n = \mathbb{E}_{\hat{P}_{X^n}}[X]$. Therefore $\hat{P}_{X^n} \in \Gamma \iff \bar{X} \in A$. Besides, we define $I(t) = \inf_{Q_X:\mathbb{E}_{Q_X}[X]=t} D(Q_X||P_X)$. Then $\inf_{Q_X\in\Gamma} D(Q_X||P_X) = \inf_{t\in A} I(t)$ and $\inf_{Q_X\in\Gamma^o} D(Q_X||P_X) = \inf_{t\in A^o} I(t)$ (The open property of $\Gamma$ and $A$ are equivalent). Therefore, we have shown (1). Next we show that $I(t) = \psi_X^*(t)$, which is established by Lagrange duality. The Lagrange function for $I(t)$ is $L(Q_X, \lambda, \mu) = D(Q_X||P_X) - \lambda(\mathbb{E}_{Q_X}[X] - t) - \mu(\mathbb{E}_{Q_X}[1] - 1)$. After minimize $L(Q_X, \lambda, \mu)$ over $Q_X$, we get $Q_X(x) = P_X(x)\exp(\lambda X - \psi_X(\lambda))$ and $\inf_{Q_X} L(Q_X, \lambda, \mu) = \lambda t - \psi_X(\lambda)$. By the weak duality property we have $\psi_X^*(t) \leq I(t)$. On the other hand, for given $t$, when $\lambda$ satisfies $t = \psi_X'(\lambda)$ and $Q_X(x) = P_X(x)\exp(\lambda X - \psi_X(\lambda))$ simultaneously, $\psi_X^*(t) = \lambda t - \psi_X(\lambda) = L(Q_X, \lambda, \mu)$. Under such circumstance, we also have $\mathbb{E}_{Q_X}[X] = t$ and $\mathbb{E}_{Q_X}[1] = 1$. Therefore, $L(Q_X, \lambda, \mu) = D(Q_X||P_X) \geq I(t)$.

That is, $\psi_X^*(t) \geq I(t)$. In conclusion, $I(t) = \psi_X^*(t)$ when $t = \psi_X'(\lambda)$ has solution in $\mathbb{R}$.

If $t = x_1$, we can show that $I(x_1) = \log \frac{1}{P_X(x_1)}$ since $Q_X$ has probability 1 at $x_1$. On the other hand, $\psi_X^*(x_1) = \sup_\lambda \log \frac{e^{\lambda x_1}}{\sum_{x\in\mathcal{X}} P(x)\exp(\lambda x)} = \lim_{\lambda\to\infty} \log \frac{e^{\lambda x_1}}{\sum_{x\in\mathcal{X}} P(x)\exp(\lambda x)} = \log \frac{1}{P_X(x_1)}$. The case for $t = x_M$ can be shown similarly. Therefore, at the end point we also have $I(t) = \psi^*(t)$. □

# 2 Convex property of log-MGF and its conjugate

1. log-MGF $\psi_X(t)$ is a convex function
2. $\psi_X^*(t)$ is a convex function

*Proof.* To show that log-MGF is convex, we show $\psi_X''(t) > 0$. Let $P_X^{(t)}(x) = P_X(x)\exp(tx - \psi_X(t))$, which is a probability distribution. Then $\psi_X''(t) = \mathrm{Var}_{P_X^{(t)}}[X] > 0$. To show $\psi_X^*(t)$ is a convex function, we use the definition of convex functions. That is, we verify $\psi_X^*(\theta t_1 + (1-\theta)t_2) \leq \theta\psi_X^*(t_1) + (1-\theta)\psi_X^*(t_2)$. □

# 3 Using Cramér's theorem to prove Strong Law of Large Number

Suppose $X_1, \ldots, X_n$ are i.i.d. sampled (same with $X$) and $\mathbb{E}[X]$ exists, then $X_n \xrightarrow{a.s.} \mathbb{E}[X]$.

*Proof.* For any $\epsilon > 0$, we define $E_n = \{w \| \bar{X}_n(w) - \mathbb{E}[X]| \geq \epsilon\}$. By Cramér's theorem, the series sum $\sum_{n=1}^{+\infty} P(E_n)$ converges since $P(E_n)$ decays exponentially fast. Then by Borel Cantelli Lemma, $P(\cap_{k=1}^{+\infty} \cup_{n\geq k} \{w \| \bar{X}_n(w) - \mathbb{E}[X]| \geq \epsilon\}) = 0$ for any $\epsilon > 0$, which is equivalent to $P(\lim_{n\to\infty} \bar{X}_n = \mathbb{E}[X]) = 1$. □