Properties of $\mu_X(s)$.

1. $\mu_X(s)$ is called the cumulant generating function.

$$\left. \frac{d^k}{ds^k} \mu_X(s) \right|_{s=0} = \begin{cases} 0 & k=0 \\ \mathbb{E}[X] & k=1 \\ \text{Var}[X] & k=2 \end{cases}$$

Proof: Do it yourself.

2. $\mu_X(s)$ is always convex.

Proof:

$$\mu_X(\lambda s_1 + (1-\lambda) s_2) = \ln \left\{ \mathbb{E} \left[ (e^{s_1 X})^\lambda (e^{s_2 X})^{1-\lambda} \right] \right\}$$

$$\leq \ln \left\{ (\mathbb{E} [ e^{s_1 X} ] )^\lambda (\mathbb{E} [ e^{s_2 X} ] )^{1-\lambda} \right\}$$

$$= \lambda \ln \mathbb{E} [ e^{s_1 X} ] + (1-\lambda) \ln \mathbb{E} [ e^{s_2 X} ]$$

$$= \lambda \mu_X(s_1) + (1-\lambda) \mu_X(s_2).$$

By Hölder's inequality:

$$\mathbb{E}[XY] \leq \mathbb{E}[X^p]^{\frac{1}{p}} \mathbb{E}[X^q]^{\frac{1}{q}}.$$
Examples
1. $X \sim N(0, 1)$, $\mu_X(s) = \frac{s^2}{2}$
2. $X \sim e^{-x}$, $x > 0$, $\mu_X(s) = \left\{ \begin{array}{ll} -\ln(1-s) & , \quad s < 1 \\ \infty & , \quad s \geq 1 \end{array} \right.$

The function
$$\mu_X^*(t) = \max_s \left\{ st - \mu_X(s) \right\}$$
is called the Legendre transform of $\mu_X$.

Properties of $\mu_X^*(t)$:

Let $\mu_X^{**}(s) = \max_t \left\{ st - \mu_X^*(t) \right\}$.

then

1. $\mu_X^*(t)$ is convex, regardless of $\mu_X(s)$.
2. $\mu_X^{**}(s)$ is convex, regardless of $\mu_X^*(t)$.
3. If $\mu_X(s)$ is convex, then $\mu_X^{**}(s) = \mu_X(s)$.

Pictorial Illustration

\[ \mu_X(s) \xrightarrow{*} \mu_X^*(t) \xrightarrow{*} \mu_X^{**}(s) \]
Pf of (1):

We want to show

\[ \mu_X^*(\alpha t_1 + (1-\alpha) t_2) \leq \alpha \mu_X^*(t_1) + (1-\alpha) \mu_X^*(t_2). \]

Note that

\[ \mu_X^*(\alpha t_1 + (1-\alpha) t_2) \]
\[ = \max_{s} \left\{ s \left( \alpha t_1 + (1-\alpha) t_2 \right) - \mu_X(s) \right\} \]
\[ = \max_{s} \left\{ \alpha (st_1 - \mu_X(s)) + (1-\alpha) (st_2 - \mu_X(s)) \right\} \]
\[ \leq \max_{s} \left\{ \alpha (st_1 - \mu_X(s)) \right\} + \max_{s} \left\{ (1-\alpha) (st_2 - \mu_X(s)) \right\} \]
\[ = \alpha \mu_X^*(t_1) + (1-\alpha) \mu_X^*(t_2). \]

where (a) holds because

\[ \max_{x} (f(x) + g(x)) \]
\[ \leq \max_{x} f(x) + \max_{x} g(x). \]

Pf of (2): follows the same argument of (1).

Pf of (3): For any \( s \), \( \mu_x^{**}(s) \) is defined as

\[ \mu_x^{**}(s) = \max_{\tau} \left\{ st - \mu_x^*(\tau) \right\}. \]

First, let \( t \) be the maximizer of

\[ t = \arg \max_{\tau} \left\{ st - \mu_x^*(\tau) \right\}. \]

Then,

\[ \mu_x^{**}(s) = st - \mu_x^*(t). \]
Since $t$ is the maximizer, we have
\[ \frac{d}{dt} (st - \mu^*_x(t)) \bigg|_{t=t} = 0 \]
\[ \implies s = \mu^*_x'(t) \quad \text{(i.e. $s$ is the slope of $\mu^*_x(t)$ at $t$)} \]
Therefore,
\[ \mu^*_x(s) = st - \mu^*_x(t) \]
\[ = \mu^*_x(t) + t - \mu^*_x(t) \]
\[ = \mu^*_x(t) + t - \max_s \{ st - \mu_x(s) \} \]
Now, let $\hat{s}$ be the maximizer of $st - \mu_x(s)$:
\[ \hat{s} = \arg\max_s \{ st - \mu_x(s) \} \]
Then,
\[ \max_s \{ st - \mu_x(s) \} = \hat{s} t - \mu_x(\hat{s}) \]
and
\[ \mu^*_x(s) = \mu^*_x(t) t - \hat{s} t + \mu_x(\hat{s}) \]
Now, if we can show that $\hat{s} = s$, and $\hat{s} = \mu^*_x(t)$, then we are done.
To this end, note that
\[ \mu^*_x'(t) = \frac{d}{dt} \mu^*_x(t) = \frac{d}{dt} \max_s \{ st - \mu_x(s) \} \]
\[ = \frac{d}{dt} (\hat{s} t - \mu_x(\hat{s})) = \hat{s} \]
and since $s = \mu^*_x(t)$, we also have $s = \hat{s}$. \[ \square \]
Pictorial Illustration

Case 1 Convex

\[ \mu_x(s) \uparrow \rightarrow \mu^*(t) \rightarrow \mu^{**}(s) \]

Case 2 Convex, Not differentiable

\[ \mu(s) \uparrow \rightarrow \mu^*(t) \rightarrow \mu^{**}(s) \]

\[ \mu_x^*(t) = \max_{s} (st - \mu_x(s)) \]

\[ \frac{d}{ds} = 0 \Rightarrow t = \mu'(s) \]

So for any \( t \in \text{dom}(\mu^*) \), \( t \) must be the slope of \( \mu(s) \).
At \( s_0 \), since there are multiple slopes in \( \mu(s) \) that can maximizes \( \{st - \mu(s)\} \), there is a range of \( t \)'s in \( \mu^*(t) \).
Case 3. Non-convex

Since \( \mu^*(t) \) is non-convex, then there must exist a line that connects \( s_a \) and \( s_b \). Let the slope of the line be \( t_0 \).

In \( \mu^*(t) \), since slope of \( \mu^*(t) \) is \( s \), and slope of \( \mu(s) \) is \( t \), we have two slopes \( s_a \) and \( s_b \) at \( t_0 \).

When taking \( \mu^* \), since \( \mu^*(t) \) is convex, \( \mu^{**}(s) \) must be the convex envelope of \( \mu(s) \).