Additional Information (optional)

Properties of $\mu_X(s)$.

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

$$\frac{d^k}{ds^k}\mu_X(s)\Big|_{s=0} = \begin{cases} 0 & , \quad k=0 \\ \mathbb{E}[X] & , \quad k=1 \\ \mathrm{Var}[X] & , \quad k=2 \end{cases}$$

Pf: Do it yourself.

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

Pf: $\mu_X(\lambda s_1 + (1-\lambda)s_2)$

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

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

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

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

By Hölder's inequality:

$$\mathbb{E}[XY] \leq \mathbb{E}[X^p]^{1/p}\,\mathbb{E}[X^q]^{1/q}.$$

Examples

1. $X \sim N(0,1)$, $\mu_X(s) = \frac{s^2}{2}$

2. $X \sim e^{-x}$, $x>0$, $\mu_X(s) = \begin{cases} -\ln(1-s) & , s<1 \\ \infty & , s \geq 1 \end{cases}$

The function

$$\mu_X^*(t) = \max_s \{ st - \mu_X(s) \}$$

is called the <u>Legender transform</u> of μ_X.

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

~~(1)~~ Let $\mu_X^{**}(s) = \max_t \{ st - \mu_X^*(t) \}$.

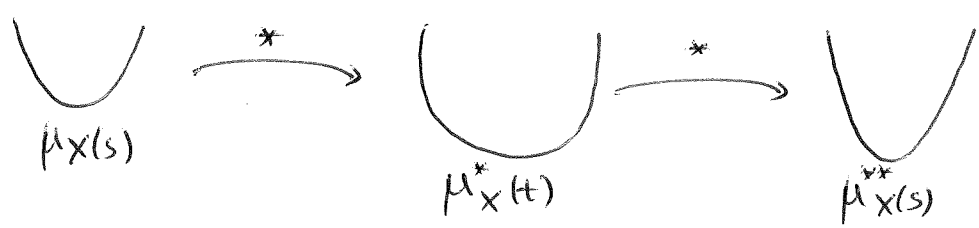
then

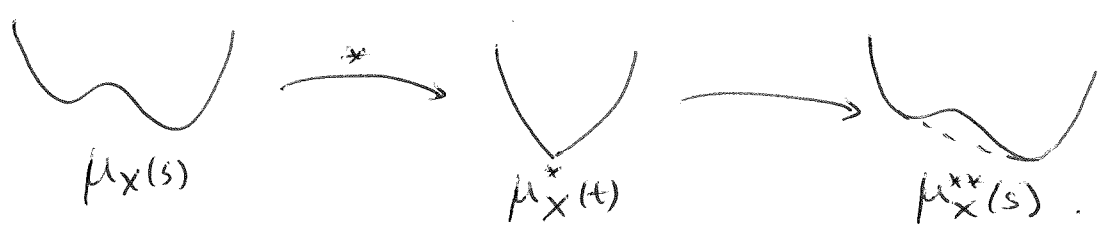
(1) $\mu_X^*(t)$ is convex, regardless of $\mu_X(s)$.

(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





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

$$\begin{aligned}
&\mu_X^*(\alpha t_1 + (1-\alpha) t_2) \\
&= \max_s \{ s(\alpha t_1 + (1-\alpha)t_2) - \mu_X(s) \} \\
&= \max_s \{ \alpha(st_1 - \mu_X(s)) + (1-\alpha)(st_2 - \mu_X(s)) \} \\
&\overset{(a)}{\leq} \max_s \{ \alpha(st_1 - \mu_X(s)) \} + \max_s \{ (1-\alpha)(st_2 - \mu_X(s)) \} \\
&= \alpha \mu_X^*(t_1) + (1-\alpha)\mu_X^*(t_2).
\end{aligned}$$

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

$$\mu_X^{**}(s) = \max_\tau \{ s\tau - \mu_X^*(\tau) \}.$$

First, let t be the maximizer of

$$t = \operatorname{argmax}_\tau \{ s\tau - \mu_X^*(\tau) \}.$$

Then,

$$\mu_X^{**}(s) = st - \mu_X^*(t).$$

Since t is the maximizer, we have

$$\frac{d}{d\tau}\left(s\tau - \mu_X^*(\tau)\right)\Big|_{\tau=t} = 0$$

$$\Longrightarrow \qquad s = \mu_X^{*\prime}(t) \qquad (\text{i.e. } s = \text{slope of } \mu_X^*(t) \text{ at } t)$$

Therefore,

$$\begin{aligned}\mu_X^{**}(s) &= st - \mu_X^*(t) \\ &= \mu_X^{*\prime}(t)\, t - \mu_X^*(t) \\ &= \mu_X^{*\prime}(t)\, t - \max_{\xi}\{\xi t - \mu_X(\xi)\}\end{aligned}$$

Now, let $\hat{s}$ be the maximizer of $\xi t - \mu_X(\xi)$:

$$\hat{s} = \operatorname{argmax}_{\xi}\{\xi t - \mu_X(\xi)\}.$$

Then, $\max_{\xi}\{\xi t - \mu_X(\xi)\} = \hat{s}t - \mu_X(\hat{s})$, and

$$\mu_X^{**}(s) = \mu_X^{*\prime}(t)\, t - \hat{s}t + \mu_X(\hat{s}).$$

Now, if we can show that $\hat{s} = s$, ~~then~~, and $\hat{s} = \mu^{*\prime}(t)$, then we are done.

To this end, note that

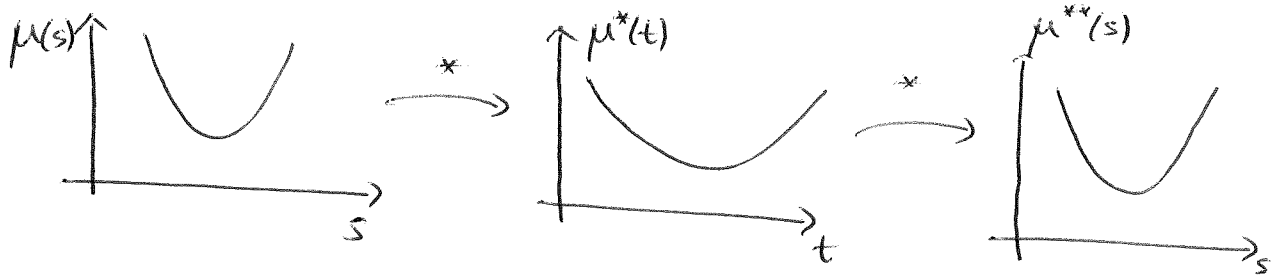
$$\begin{aligned}\mu_X^{*\prime}(t) = \frac{d}{dt}\mu_X^*(t) &= \frac{d}{dt}\max_{\xi}\{\xi t - \mu_X(\xi)\} \\ &= \frac{d}{dt}\left(\hat{s}t - \mu_X(\hat{s})\right) = \hat{s}.\end{aligned}$$

and since $s = \mu_X^{*\prime}(t)$, we also have $s = \hat{s}$.

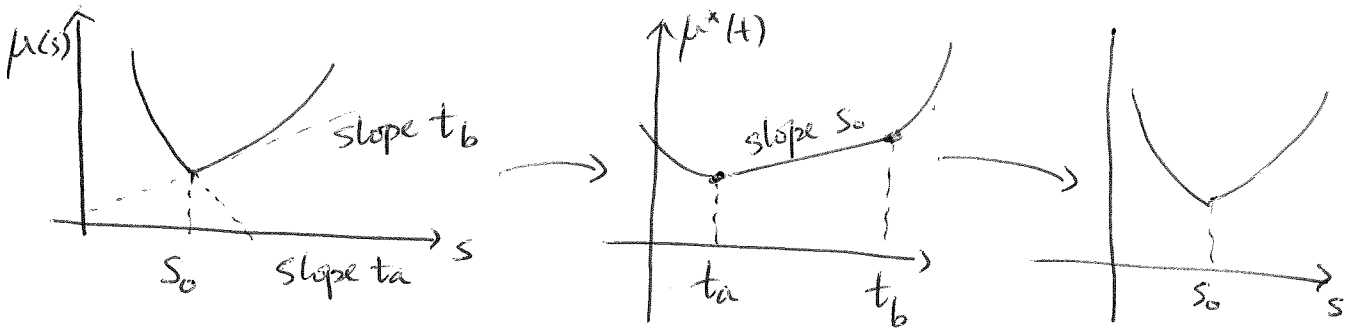
$\blacksquare$

Pictorial Illustration

Case 1 Convex



Case 2 Convex, Not differentiable

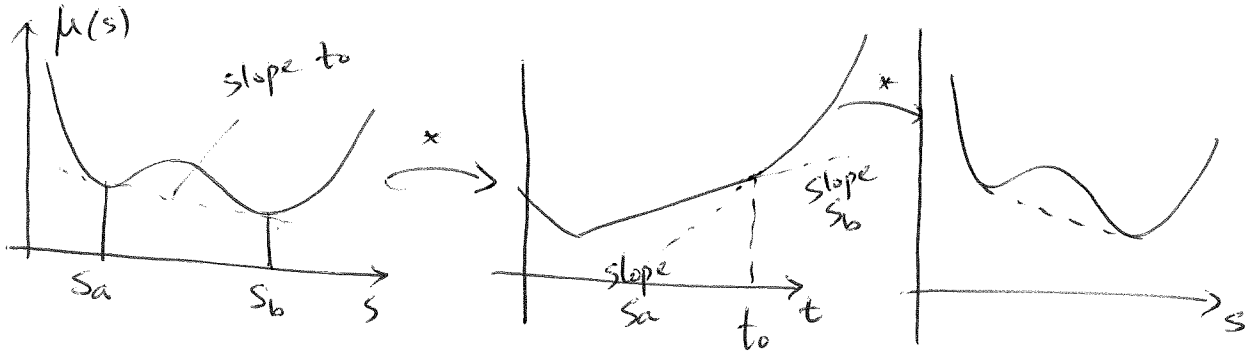


$$\mu_X^*(t) = \max_s \left(st - \mu_X(s) \right)$$

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

So for any $t \in \mathrm{dom}(\mu^*)$, 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 $t = \mu^*(s)$~~

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