Exponential Family of Distributions

- Parametric density functions

$$P_G(\boldsymbol{x}|\boldsymbol{\theta}) = e^{\boldsymbol{\theta}\cdot\boldsymbol{x} - G(\boldsymbol{\theta})}\, P_0(\boldsymbol{x})$$

- $\boldsymbol{\theta}$ and $\boldsymbol{x}$ vectors in R^d
- Cumulant function $G(\boldsymbol{\theta})$ assures normalization

$$G(\boldsymbol{\theta}) = \ln \int e^{\boldsymbol{\theta}\cdot\boldsymbol{x}}\, P_0(\boldsymbol{x})\, d\boldsymbol{x}$$

- $G(\boldsymbol{\theta})$ is convex function on convex set $\boldsymbol{\Theta} \subseteq R^d$
- G characterizes members of the family
- $\boldsymbol{\theta}$ is *natural* parameter

- Expectation parameter

$$\boldsymbol{\mu} = \int_{\boldsymbol{x}} \boldsymbol{x} P_G(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = E_{\boldsymbol{\theta}}(\boldsymbol{x}) = g(\boldsymbol{\theta})$$

  where $g(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta})$

- Second convex function $F(\boldsymbol{\mu})$ on space $g(\boldsymbol{\Theta})$

$$F(\boldsymbol{\mu}) = \boldsymbol{\theta} \cdot \boldsymbol{\mu} - G(\boldsymbol{\theta})$$

- $G(\boldsymbol{\theta})$ and $F(\boldsymbol{\mu})$ are *convex conjugate* functions
- Let $f(\boldsymbol{\mu}) = \nabla_{\boldsymbol{\mu}} F(\boldsymbol{\mu})$
- $f(\boldsymbol{\mu}) = g^{-1}(\boldsymbol{\mu})$

Primal & Dual Parameters

natural paramater		expectation parameter
$\boldsymbol{\theta}$	$\overset{g}{\longrightarrow}$ $\underset{f}{\longleftarrow}$	$\boldsymbol{\mu}$
$G(\boldsymbol{\theta})$		$F(\boldsymbol{\mu})$

- $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ are dual parameters
- Parameter transformations
  $g(\boldsymbol{\theta}) = \boldsymbol{\mu}$ and $f(\boldsymbol{\mu}) = \boldsymbol{\theta}$

[A,BN]

Gaussian (unit variance)

$$\begin{aligned} P(\boldsymbol{x}|\boldsymbol{\theta}) &\sim e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{x})^2} \\ &= e^{\boldsymbol{\theta}\cdot\boldsymbol{x}-\frac{1}{2}\boldsymbol{\theta}^2} e^{\frac{1}{2}\boldsymbol{x}^2} \end{aligned}$$

Cumulant function: $G(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^2$

Parameter transformations:

$g(\boldsymbol{\theta}) = \boldsymbol{\theta} = \boldsymbol{\mu}$ and $f(\boldsymbol{\mu}) = \boldsymbol{\mu} = \boldsymbol{\theta}$

Dual convex function: $$\begin{aligned} F(\boldsymbol{\mu}) &= \boldsymbol{\theta}\cdot\boldsymbol{\mu} - G(\boldsymbol{\theta}) \\ &= \frac{1}{2}\boldsymbol{\mu}^2 \end{aligned}$$

Square loss: $L_t(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2$

Bernoulli

Examples x_t are coin flips in $\{0,1\}$

$$P(x|\mu) = \mu^x (1-\mu)^{1-x}$$

μ is the probability (expectation) of 1

Natural parameter: $\theta = \ln \frac{\mu}{1-\mu}$

$$P(x|\theta) = \exp\left(\theta x - \ln(1+e^\theta)\right)$$

Cumulant function: $G(\theta) = \ln(1+e^\theta)$

Parameter transformations:

$$\mu = g(\theta) = \frac{e^\theta}{1+e^\theta} \text{ and } \theta = f(\mu) = \ln \frac{\mu}{1-\mu}$$

Dual function: $F(\mu) = \mu \ln \mu + (1-\mu) \ln(1-\mu)$

$$\begin{aligned} \text{Log loss:} \quad L_t(\theta) &= -x_t\theta + ln(1+e^\theta) \\ &= -x_t \ln \mu - (1-x_t)\ln(1-\mu) \end{aligned}$$

Poisson

Examples x_t are natural numbers in $\{0, 1, \dots\}$

$$P(x|\mu) = \frac{e^{-\mu}\mu^x}{x!}$$

μ is expectation of x

Natural parameter: $\theta = \ln \mu$

$$P(x|\theta) = \exp\left(\theta x - e^{\theta}\right)\frac{1}{x!}$$

Cumulant function: $G(\theta) = e^{\theta}$

Parameter transformations:

$$\mu = g(\theta) = e^{\theta} \text{ and } \theta = f(\mu) = \ln \mu$$

Dual function: $F(\mu) = \mu \ln \mu - \mu$

$$\begin{aligned}\text{Loss:}\quad L_t(\theta) &= -x_t\theta + e^{\theta} + \ln x_t! \\ &= -x_t \ln \mu + \mu + \ln x_t!\end{aligned}$$

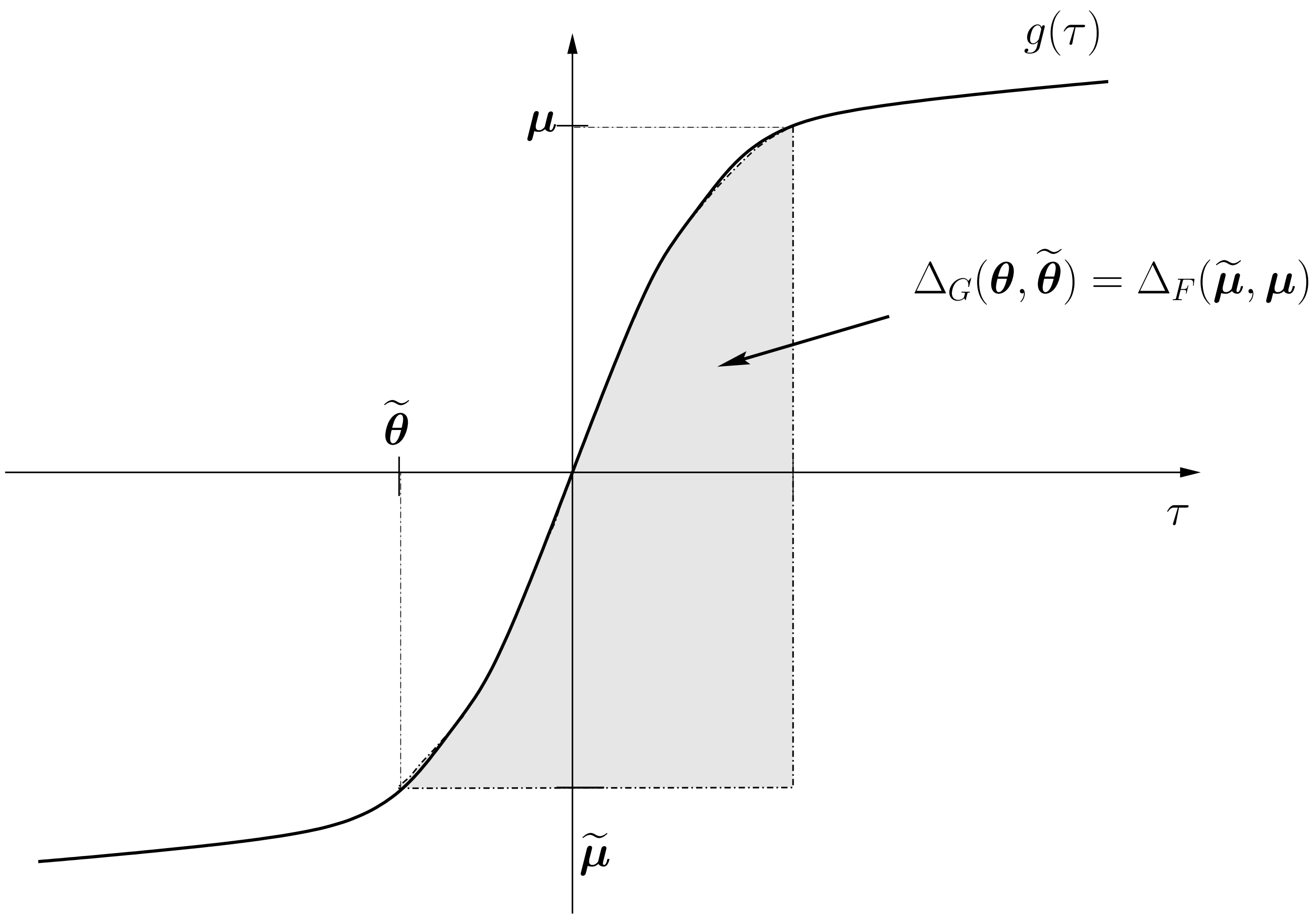
Bregman Div. as Rel. Ent. between Distributions

Let $P(\boldsymbol{x}|\boldsymbol{\theta})$ and $P(\boldsymbol{x}|\widetilde{\boldsymbol{\theta}})$ denote two distributions with cumulant function G

$$
\begin{aligned}
\Delta_G(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) &= \int_{\boldsymbol{x}} P_G(\boldsymbol{x}|\boldsymbol{\theta}) \ln \frac{P_G(\boldsymbol{x}|\boldsymbol{\theta})}{P_G(\boldsymbol{x}|\widetilde{\boldsymbol{\theta}})} d\boldsymbol{x} \\
&= G(\widetilde{\boldsymbol{\theta}}) - G(\boldsymbol{\theta}) - (\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \cdot \boldsymbol{\mu} \\
&\overset{F(\boldsymbol{\mu})=\boldsymbol{\theta}\cdot\boldsymbol{\mu}-G(\boldsymbol{\theta})}{=} F(\boldsymbol{\mu}) - F(\widetilde{\boldsymbol{\mu}}) - (\boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}}) \cdot \widetilde{\boldsymbol{\theta}} \\
&= \Delta_F(\boldsymbol{\mu}, \widetilde{\boldsymbol{\mu}})
\end{aligned}
$$

[A,BN,AW]

Area unchanged When Slide Flipped



$$
\begin{aligned}
\Delta_G(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) &= G(\boldsymbol{\theta}) - G(\widetilde{\boldsymbol{\theta}}) - (\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}) \cdot g(\widetilde{\boldsymbol{\theta}}) \\
&= \int_{\widetilde{\boldsymbol{\theta}}}^{\boldsymbol{\theta}} (g(\boldsymbol{\tau}) - g(\widetilde{\boldsymbol{\theta}})) \cdot d\boldsymbol{\tau} \\
&\overset{\text{flip}}{=} \int_{\boldsymbol{\mu}}^{\widetilde{\boldsymbol{\mu}}} (f(\boldsymbol{\sigma}) - f(\boldsymbol{\mu})) \cdot d\boldsymbol{\sigma} \\
&= F(\widetilde{\boldsymbol{\mu}}) - F(\boldsymbol{\mu}) - (\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu}) \cdot f(\boldsymbol{\mu}) \\
&= \Delta_F(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\mu})
\end{aligned}
$$

Dual divergence for Bernoulli

$$G(\boldsymbol{\theta}) = \ln(1 + e^{\theta}) \qquad F(\mu) = \mu \ln \mu + (1 - \mu) \ln(1 - \mu)$$

$$g(\theta) = \tfrac{e^{\theta}}{1+e^{\theta}} = \mu \qquad f(\mu) = \ln \tfrac{\mu}{1-\mu} = \theta$$

$$\Delta_G(\widetilde{\theta}, \theta) = \ln(1 + e^{\widetilde{\theta}}) - \ln(1 + e^{\theta}) - (\widetilde{\theta} - \theta)\frac{e^{\theta}}{1 + e^{\theta}}$$

$$\Delta_F(\mu, \widetilde{\mu}) = \mu \ln \frac{\mu}{\widetilde{\mu}} + (1 - \mu) \ln \frac{1 - \mu}{1 - \widetilde{\mu}}$$

Binary relative entropy

Dual divergence for Poisson

$$G(\boldsymbol{\theta}) = e^{\theta} \qquad F(\mu) = \mu \ln \mu - \mu$$

$$g(\theta) = e^{\theta} = \mu \quad f(\mu) = \ln \mu = \theta$$

$$\Delta_G(\widetilde{\theta}, \theta) = e^{\widetilde{\theta}} - e^{\theta} - (\widetilde{\theta} - \theta)e^{\theta}$$

$$\Delta_F(\mu, \widetilde{\mu}) = \mu \ln \frac{\mu}{\widetilde{\mu}} + \widetilde{\mu} - \mu$$

Unnormalized relative entropy