Generalized maximum likelihood estimates for exponential families

Imre Csiszár¹  František Matúš²

¹A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest
²Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague

March 6, 2007
Institute for Mathematics and its Applications
University of Minnesota
Logarithmic affinity
Exponential families
Maximizing likelihood
Generalized MLE
Divergence from EF
$P$ ... a probability measure (pm) on a finite set $\Omega$, 
$P$ ... a probability measure (pm) on a finite set $\Omega$, 
\[ \sum_{\omega \in \Omega} P(\omega) = 1, \text{ each } P(\omega) \text{ nonnegative.} \]
\( P \) ... a probability measure (pm) on a finite set \( \Omega \),
\[
\sum_{\omega \in \Omega} P(\omega) = 1, \text{ each } P(\omega) \text{ nonnegative.}
\]
\( s(P) = \{\omega \in \Omega: P(\omega) > 0\} \) ... the support of a pm \( P \),
$P$ ... a probability measure (pm) on a finite set $\Omega$,

$$\sum_{\omega \in \Omega} P(\omega) = 1,$$

each $P(\omega)$ nonnegative.

$s(P) = \{ \omega \in \Omega: P(\omega) > 0 \}$ ... the support of a pm $P$,

$P$ sits on $s(P)$.
A probability measure (pm) on a finite set \( \Omega \),
\[
\sum_{\omega \in \Omega} P(\omega) = 1, \text{ each } P(\omega) \text{ nonnegative}.
\]
\( s(P) = \{ \omega \in \Omega : P(\omega) > 0 \} \) ... the support of a pm \( P \),
\( P \) sits on \( s(P) \).

For pm’s \( P, Q \) on \( \Omega \) with \( A = s(P) \cap s(Q) \) nonempty and \( t \in \mathbb{R} \),
the log-affine combination of \( P \) and \( Q \) is the pm \( P^tQ^{1-t} \)
sitting on \( A \) and proportional to
\[
\omega \mapsto P(\omega)^t Q(\omega)^{1-t}
\]
A probability measure (pm) \( P \) on a finite set \( \Omega \),
\[
\sum_{\omega \in \Omega} P(\omega) = 1, \text{ each } P(\omega) \text{ nonnegative.}
\]
\( s(P) = \{\omega \in \Omega: P(\omega) > 0\} \) ... the support of a pm \( P \),
\( P \) sits on \( s(P) \).

For pm’s \( P, Q \) on \( \Omega \) with \( A = s(P) \cap s(Q) \) nonempty and \( t \in \mathbb{R} \),
the log-affine combination of \( P \) and \( Q \) is the pm \( P^t Q^{1-t} \)
sitting on \( A \) and proportional to
\[
\omega \mapsto P(\omega)^t Q(\omega)^{1-t}
\]

\[
\begin{array}{cccc}
\omega_1 & \omega_2 & \omega_3 & \omega_4 \\
\end{array}
\]
$P$ ... a probability measure (pm) on a finite set $\Omega$, 
\[ \sum_{\omega \in \Omega} P(\omega) = 1, \text{ each } P(\omega) \text{ nonnegative.} \]
$s(P) = \{\omega \in \Omega: P(\omega) > 0\}$ ... the support of a pm $P$, 
$P$ sits on $s(P)$.

For pm's $P, Q$ on $\Omega$ with $A = s(P) \cap s(Q)$ nonempty and $t \in \mathbb{R}$, 
the log-affine combination of $P$ and $Q$ is the pm 
\[ P^t Q^{1-t} \]
sitting on $A$ and proportional to 
\[ \omega \mapsto P(\omega)^t Q(\omega)^{1-t} \]

| $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_4$ |
\( P \) ... a probability measure (pm) on a finite set \( \Omega \),
\[
\sum_{\omega \in \Omega} P(\omega) = 1, \text{ each } P(\omega) \text{ nonnegative.}
\]
\( s(P) = \{ \omega \in \Omega : P(\omega) > 0 \} \) ... the support of a pm \( P \),
\( P \) sits on \( s(P) \).

For pm’s \( P, Q \) on \( \Omega \) with \( A = s(P) \cap s(Q) \) nonempty and \( t \in \mathbb{R} \),
the log-affine combination of \( P \) and \( Q \) is the pm \( P^t Q^{1-t} \)
sitting on \( A \) and proportional to
\[
\omega \mapsto P(\omega)^t Q(\omega)^{1-t}
\]

\[
\begin{array}{c|cccc}
\omega & \omega_1 & \omega_2 & \omega_3 & \omega_4 \\
\hline
P & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0
\end{array}
\]
A probability measure (pm) on a finite set $\Omega$, $\sum_{\omega \in \Omega} P(\omega) = 1$, each $P(\omega)$ nonnegative.

$s(P) = \{\omega \in \Omega : P(\omega) > 0\}$ ... the support of a pm $P$,

$P$ sits on $s(P)$.

For pm’s $P$, $Q$ on $\Omega$ with $A = s(P) \cap s(Q)$ nonempty and $t \in \mathbb{R}$, the log-affine combination of $P$ and $Q$ is the pm $P^t Q^{1-t}$ sitting on $A$ and proportional to

$$\omega \mapsto P(\omega)^t Q(\omega)^{1-t}$$

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>$\omega_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{2}$</td>
<td>$0$</td>
</tr>
<tr>
<td>$Q$</td>
<td>$0$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{5}{8}$</td>
</tr>
</tbody>
</table>
A probability measure (pm) on a finite set $\Omega$, $\sum_{\omega \in \Omega} P(\omega) = 1$, each $P(\omega)$ nonnegative.

$s(P) = \{\omega \in \Omega : P(\omega) > 0\}$ ... the support of a pm $P$, $P$ sits on $s(P)$.

For pm’s $P, Q$ on $\Omega$ with $A = s(P) \cap s(Q)$ nonempty and $t \in \mathbb{R}$, the log-affine combination of $P$ and $Q$ is the pm $P^t Q^{1-t}$ sitting on $A$ and proportional to

$$\omega \mapsto P(\omega)^t Q(\omega)^{1-t}$$

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>$\omega_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{2}$</td>
<td>$0$</td>
</tr>
<tr>
<td>$Q$</td>
<td>$0$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{5}{8}$</td>
</tr>
<tr>
<td>$P^tQ^{1-t}$</td>
<td>$0$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

$t = \frac{1}{2}$
$P$ ... a probability measure (pm) on a finite set $\Omega$,
\[ \sum_{\omega \in \Omega} P(\omega) = 1, \text{ each } P(\omega) \text{ nonnegative}. \]
$s(P) = \{\omega \in \Omega : P(\omega) > 0\}$ ... the support of a pm $P$,
$P$ sits on $s(P)$.

For pm’s $P, Q$ on $\Omega$ with $A = s(P) \cap s(Q)$ nonempty and $t \in \mathbb{R}$,
the log-affine combination of $P$ and $Q$ is the pm $P^t Q^{1-t}$
sitting on $A$ and proportional to
\[ \omega \mapsto P(\omega)^t Q(\omega)^{1-t} \]

... log-convex combinations if $0 \leq t \leq 1$
\[ \sum_{\omega \in \Omega} P(\omega)^t Q(\omega)^{1-t} \leq 1, \text{ tight if and only if } P = Q. \]
A family $\mathcal{P}$ of pm's on $\Omega$ is log-affine if it is closed to log-affine combinations.
A family $\mathcal{P}$ of pm’s on $\Omega$ is **log-affine** if it is closed to log-affine combinations.

The log-affine **envelope** of a family $\mathcal{P}$ is the inclusion smallest log-affine family that contains $\mathcal{P}$. 
A family $\mathcal{P}$ of pm’s on $\Omega$ is log-affine if it is closed to log-affine combinations.

The log-affine envelope of a family $\mathcal{P}$ is the inclusion smallest log-affine family that contains $\mathcal{P}$.

Binomial family $\mathcal{P} = \{Q_p : 0 < p < 1\}$
A family $\mathcal{P}$ of pm’s on $\Omega$ is log-affine if it is closed to log-affine combinations.

The log-affine envelope of a family $\mathcal{P}$ is the inclusion smallest log-affine family that contains $\mathcal{P}$.

Binomial family $\mathcal{P} = \{ Q_p : 0 < p < 1 \}$ of the pm’s $Q_p(\omega) = \binom{n}{\omega} p^\omega (1 - p)^{n-\omega}$ on $\Omega = \{0, 1, \ldots, n\}$:
A family $\mathcal{P}$ of pm’s on $\Omega$ is log-affine if it is closed to log-affine combinations.

The log-affine envelope of a family $\mathcal{P}$ is the inclusion smallest log-affine family that contains $\mathcal{P}$.

Binomial family $\mathcal{P} = \{ Q_p : 0 < p < 1 \}$ of the pm’s $Q_p(\omega) = \binom{n}{\omega} p^\omega (1 - p)^{n-\omega}$ on $\Omega = \{0, 1, \ldots, n\}$: the log-affine combination of $Q_p$ and $Q_q$ at $\omega \in \Omega$ is $\propto$

$$\left[ \binom{n}{\omega} p^\omega (1 - p)^{n-\omega} \right]^t \left[ \binom{n}{\omega} q^\omega (1 - q)^{n-\omega} \right]^{1-t},$$
A family \( \mathcal{P} \) of pm’s on \( \Omega \) is log-affine if it is closed to log-affine combinations.

The log-affine envelope of a family \( \mathcal{P} \) is the inclusion smallest log-affine family that contains \( \mathcal{P} \).

Binomial family \( \mathcal{P} = \{ Q_p : 0 < p < 1 \} \) of the pm’s \( Q_p(\omega) = \binom{n}{\omega} p^\omega (1 - p)^{n-\omega} \) on \( \Omega = \{0, 1, \ldots, n\} \): the log-affine combination of \( Q_p \) and \( Q_q \) at \( \omega \in \Omega \) is

\[
\left[ \binom{n}{\omega} p^\omega (1 - p)^{n-\omega} \right]^t \left[ \binom{n}{\omega} q^\omega (1 - q)^{n-\omega} \right]^{1-t},
\]

\( Q_t^t Q_q^{1-t} = Q_r \) with \( r = \frac{p^t q^{1-t}}{p^t q^{1-t} + (1-p)^t (1-q)^{1-t}} \).
A family $\mathcal{P}$ of pm's on $\Omega$ is \textbf{log-affine} if it is closed to log-affine combinations.

The log-affine \textbf{envelope} of a family $\mathcal{P}$ is the inclusion smallest log-affine family that contains $\mathcal{P}$.

Binomial family $\mathcal{P} = \{ Q_p : 0 < p < 1 \}$ of the pm's $Q_p(\omega) = \binom{n}{\omega} p^\omega (1 - p)^{n-\omega}$ on $\Omega = \{0, 1, \ldots, n\}$: the log-affine combination of $Q_p$ and $Q_q$ at $\omega \in \Omega$ is $\propto$

$$\left[ \binom{n}{\omega} p^\omega (1 - p)^{n-\omega} \right]^t \left[ \binom{n}{\omega} q^\omega (1 - q)^{n-\omega} \right]^{1-t},$$

$Q_p^t Q_q^{1-t} = Q_r$ with $r = \frac{p^t q^{1-t}}{p^t q^{1-t} + (1-p)^t (1-q)^{1-t}}$.

$\mathcal{P}$ is log-affine,
A family $\mathcal{P}$ of pm’s on $\Omega$ is log-affine if it is closed to log-affine combinations.

The log-affine envelope of a family $\mathcal{P}$ is the inclusion smallest log-affine family that contains $\mathcal{P}$.

Binomial family $\mathcal{P} = \{Q_p: 0 < p < 1\}$ of the pm’s $Q_p(\omega) = \binom{n}{\omega} p^\omega (1 - p)^{n-\omega}$ on $\Omega = \{0, 1, \ldots, n\}$: the log-affine combination of $Q_p$ and $Q_q$ at $\omega \in \Omega$ is

$$
\left[ \binom{n}{\omega} p^\omega (1 - p)^{n-\omega} \right]^t \left[ \binom{n}{\omega} q^\omega (1 - q)^{n-\omega} \right]^{1-t},
$$

$Q^t_p Q^{1-t}_q = Q_r$ with $r = \frac{p^t q^{1-t}}{p^t q^{1-t} + (1-p)^t (1-q)^{1-t}},$

$\mathcal{P}$ is log-affine,

$r$ ranges between 0 and 1 when $t \in \mathbb{R}$ and $p \neq q,$
A family $\mathcal{P}$ of pm’s on $\Omega$ is log-affine if it is closed to log-affine combinations.

The log-affine envelope of a family $\mathcal{P}$ is the inclusion smallest log-affine family that contains $\mathcal{P}$.

Binomial family $\mathcal{P} = \{Q_p : 0 < p < 1\}$ of the pm’s $Q_p(\omega) = \binom{n}{\omega} p^\omega (1-p)^{n-\omega}$ on $\Omega = \{0, 1, \ldots, n\}$: the log-affine combination of $Q_p$ and $Q_q$ at $\omega \in \Omega$ is $\propto$

$$\left[\binom{n}{\omega} p^\omega (1-p)^{n-\omega}\right]^t \left[\binom{n}{\omega} q^\omega (1-q)^{n-\omega}\right]^{1-t},$$

$Q_p^t Q_q^{1-t} = Q_r$ with $r = \frac{p^t q^{1-t}}{p^t q^{1-t} + (1-p)^t (1-q)^{1-t}},$

$\mathcal{P}$ is log-affine, $r$ ranges between 0 and 1 when $t \in \mathbb{R}$ and $p \neq q$, $\mathcal{P}$ equals the envelope of any two of its pm’s.
The restriction of a pm $P$ on $\Omega$ to $A \subseteq \Omega$

$$P^A(\omega) = \begin{cases} 
  P(\omega) & \omega \in A, \\
  0 & \text{otherwise}.
\end{cases}$$
The **restriction** of a pm $P$ on $\Omega$ to $A \subseteq \Omega$

$$P^A(\omega) = \begin{cases} P(\omega) & \omega \in A, \\ 0 & \text{otherwise}. \end{cases}$$

A partition $\pi$ of $\Omega$ is **sufficient** for a family $P$ of pm’s on $\Omega$ if

$$\dim \{ P^A : P \in P \} \leq 1$$

for any block $A \in \pi$. 
The restriction of a pm $P$ on $\Omega$ to $A \subseteq \Omega$

$$P^A(\omega) = \begin{cases} P(\omega) & \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$$

A partition $\pi$ of $\Omega$ is sufficient for a family $\mathcal{P}$ of pm’s on $\Omega$ if

$$\dim \{P^A : P \in \mathcal{P}\} \leq 1 \text{ for any block } A \in \pi.$$
The restriction of a pm $P$ on $\Omega$ to $A \subseteq \Omega$

$$P^A(\omega) = \begin{cases} P(\omega) & \omega \in A, \\ 0 & \text{otherwise}. \end{cases}$$

A partition $\pi$ of $\Omega$ is sufficient for a family $\mathcal{P}$ of pm’s on $\Omega$ if $\dim \{P^A: P \in \mathcal{P}\} \leq 1$ for any block $A \in \pi$.

$$\mathcal{P} = \{P_1, P_2, P_3\}$$
The restriction of a pm \( P \) on \( \Omega \) to \( A \subseteq \Omega \)

\[
P_A(\omega) = \begin{cases} 
P(\omega) & \omega \in A, \\ 0 & \text{otherwise.} \end{cases}
\]

A partition \( \pi \) of \( \Omega \) is sufficient for a family \( \mathcal{P} \) of pm’s on \( \Omega \) if

\[
\dim \{ P_A: P \in \mathcal{P} \} \leq 1 \text{ for any block } A \in \pi.
\]

\[
\mathcal{P} = \{ P_1, P_2, P_3 \} \\
\begin{array}{cccc}
\omega_1 & \omega_2 & \omega_3 & \omega_4
\end{array}
\]
The restriction of a pm $P$ on $\Omega$ to $A \subseteq \Omega$

$$P^A(\omega) = \begin{cases} P(\omega) & \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$$

A partition $\pi$ of $\Omega$ is sufficient for a family $\mathcal{P}$ of pm’s on $\Omega$ if
$$\dim \{ P^A : P \in \mathcal{P} \} \leq 1$$

for any block $A \in \pi$.

$\mathcal{P} = \{ P_1, P_2, P_3 \}$

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>$\omega_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>
The restriction of a pm $P$ on $\Omega$ to $A \subseteq \Omega$

$$P^A(\omega) = \begin{cases} P(\omega) & \omega \in A, \\ 0 & \text{otherwise}. \end{cases}$$

A partition $\pi$ of $\Omega$ is sufficient for a family $\mathcal{P}$ of pm’s on $\Omega$ if

$$\dim \{ P^A : P \in \mathcal{P} \} \leq 1$$

for any block $A \in \pi$.

$$\mathcal{P} = \{ P_1, P_2, P_3 \}$$

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>$\omega_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$P_2$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{0}{4}$</td>
<td>$\frac{0}{4}$</td>
</tr>
</tbody>
</table>
The restriction of a pm $\mathcal{P}$ on $\Omega$ to $A \subseteq \Omega$

$$P^A(\omega) = \begin{cases} P(\omega) & \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$$

A partition $\pi$ of $\Omega$ is sufficient for a family $\mathcal{P}$ of pm’s on $\Omega$ if $\dim \{P^A: P \in \mathcal{P}\} \leq 1$ for any block $A \in \pi$.

$$\mathcal{P} = \{P_1, P_2, P_3\}$$

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>$\omega_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$P_2$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$P_3$</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
</tbody>
</table>
The restriction of a pm $P$ on $\Omega$ to $A \subseteq \Omega$

$$P^A(\omega) = \begin{cases} 
  P(\omega) & \omega \in A, \\
  0 & \text{otherwise.}
\end{cases}$$

A partition $\pi$ of $\Omega$ is sufficient for a family $\mathcal{P}$ of pm’s on $\Omega$ if $\dim \{P^A : P \in \mathcal{P}\} \leq 1$ for any block $A \in \pi$.

$\mathcal{P} = \{P_1, P_2, P_3\}$

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>$\omega_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$P_2$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$P_3$</td>
<td>$0$</td>
<td>$0$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
</tbody>
</table>

$A_1$, $A_2$, $A_3$ sufficient
The restriction of a pm \( P \) on \( \Omega \) to \( A \subseteq \Omega \)

\[
P^A(\omega) = \begin{cases} 
    P(\omega) & \omega \in A, \\
    0 & \text{otherwise.}
\end{cases}
\]

A partition \( \pi \) of \( \Omega \) is sufficient for a family \( \mathcal{P} \) of pm’s on \( \Omega \) if

\[
\dim \{P^A : P \in \mathcal{P}\} \leq 1 \text{ for any block } A \in \pi.
\]

\[
\mathcal{P} = \{P_1, P_2, P_3\}
\]

<table>
<thead>
<tr>
<th></th>
<th>( \omega_1 )</th>
<th>( \omega_2 )</th>
<th>( \omega_3 )</th>
<th>( \omega_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_1 )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{4} )</td>
</tr>
<tr>
<td>( P_2 )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
<td>( 0 )</td>
<td>( 0 )</td>
</tr>
<tr>
<td>( P_3 )</td>
<td>( 0 )</td>
<td>( 0 )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
</tr>
</tbody>
</table>

\( A_1 \) \( A_2 \) \( A_3 \)

not sufficient
The restriction of a pm $P$ on $\Omega$ to $A \subseteq \Omega$

$$P^A(\omega) = \begin{cases} 
P(\omega) & \omega \in A, \\
0 & \text{otherwise.} \end{cases}$$

A partition $\pi$ of $\Omega$ is sufficient for a family $\mathcal{P}$ of pm’s on $\Omega$ if \( \dim \{ P^A : P \in \mathcal{P} \} \leq 1 \) for any block $A \in \pi$.

$$\mathcal{P} = \{ P_1, P_2, P_3 \}$$

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>$\omega_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$P_2$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$P_3$</td>
<td>$0$</td>
<td>$0$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
</tbody>
</table>

\( A_1 \) \( A_2 \)

minimal sufficient
The restriction of a pm \( P \) on \( \Omega \) to \( A \subseteq \Omega \)

\[
P^A(\omega) = \begin{cases} 
P(\omega) & \omega \in A, \\
0 & \text{otherwise}.
\end{cases}
\]

A partition \( \pi \) of \( \Omega \) is **sufficient** for a family \( \mathcal{P} \) of pm’s on \( \Omega \) if
\[
\dim \{ P^A : P \in \mathcal{P} \} \leq 1 \text{ for any block } A \in \pi.
\]

\[
\mathcal{P} = \{ P_1, P_2, P_3 \}
\]

<table>
<thead>
<tr>
<th></th>
<th>( \omega_1 )</th>
<th>( \omega_2 )</th>
<th>( \omega_3 )</th>
<th>( \omega_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_1 )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{4} )</td>
</tr>
<tr>
<td>( P_2 )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( P_3 )</td>
<td>0</td>
<td>0</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
</tr>
</tbody>
</table>

If sufficient for \( \mathcal{P} \) then sufficient also for its log-affine envelope.
\( \Pi \) ... a Markov kernel between finite sets \( \Omega, \Omega' \),
\( \Pi \) ... a Markov kernel between finite sets \( \Omega, \Omega' \),
\[
\sum_{\omega' \in \Omega'} \Pi(\omega, \omega') = 1, \text{ each } \Pi(\omega, \omega') \geq 0 \text{ nonnegative.}
\]
\( \Pi \) ... a Markov kernel between finite sets \( \Omega, \Omega' \),
\[
\sum_{\omega' \in \Omega'} \Pi(\omega, \omega') = 1, \text{ each } \Pi(\omega, \omega') \geq 0 \text{ nonnegative.}
\]
Pm’s \( P \) on \( \Omega \) transform to the pm’s \( P\Pi \) on \( \Omega' \) by \( \Pi \).
$\Pi$ ... a Markov kernel between finite sets $\Omega, \Omega'$,
\[
\sum_{\omega' \in \Omega'} \Pi(\omega, \omega') = 1, \text{ each } \Pi(\omega, \omega') \geq 0 \text{ nonnegative.}
\]
Pm’s $P$ on $\Omega$ transform to the pm’s $P\Pi$ on $\Omega'$ by $\Pi$.

If positive pm’s $P, Q$ are invariant to a Markov kernel $\Pi$ on $\Omega$
then their log-affine combinations are invariant to $\Pi$. 
$\Pi$ ... a Markov kernel between finite sets $\Omega, \Omega'$,

$$\sum_{\omega' \in \Omega'} \Pi(\omega, \omega') = 1,$$

each $\Pi(\omega, \omega') \geq 0$ nonnegative.

Pm's $P$ on $\Omega$ transform to the pm's $P \Pi$ on $\Omega'$ by $\Pi$.

If positive pm’s $P, Q$ are invariant to a Markov kernel $\Pi$ on $\Omega$
then their log-affine combinations are invariant to $\Pi$.

$$P_A = P^A / P(A)$$ ... the truncation of $P$ to $A$ with $P(A) > 0$, 

\[ \Pi \] ... a Markov kernel between finite sets \( \Omega, \Omega' \),
\[ \sum_{\omega' \in \Omega'} \Pi(\omega, \omega') = 1, \text{ each } \Pi(\omega, \omega') \geq 0 \text{ nonnegative}. \]
Pm’s \( P \) on \( \Omega \) transform to the pm’s \( P\Pi \) on \( \Omega' \) by \( \Pi \).

If positive pm’s \( P, Q \) are invariant to a Markov kernel \( \Pi \) on \( \Omega \) then their log-affine combinations are invariant to \( \Pi \).

\[ P_A = P^A / P(A) \] ... the truncation of \( P \) to \( A \) with \( P(A) > 0 \),
the normalized restriction.
\( \Pi \ldots \) a Markov kernel between finite sets \( \Omega, \Omega' \),
\[ \sum_{\omega' \in \Omega'} \Pi(\omega, \omega') = 1, \text{ each } \Pi(\omega, \omega') \geq 0 \text{ nonnegative.} \]

Pm’s \( P \) on \( \Omega \) transform to the pm’s \( P \Pi \) on \( \Omega' \) by \( \Pi \).

If positive pm’s \( P, Q \) are invariant to a Markov kernel \( \Pi \) on \( \Omega \)
then their log-affine combinations are invariant to \( \Pi \).

\( P_A = P^A / P(A) \ldots \) the truncation of \( P \) to \( A \) with \( P(A) > 0 \),
the normalized restriction.

Truncations of log-aff comb’s equal log-aff comb’s of truncations.
\( \Pi \) ... a Markov kernel between finite sets \( \Omega, \Omega' \),
\[
\sum_{\omega' \in \Omega} \Pi(\omega, \omega') = 1, \text{ each } \Pi(\omega, \omega') \geq 0 \text{ nonnegative.}
\]
Pm’s \( P \) on \( \Omega \) transform to the pm’s \( P\Pi \) on \( \Omega' \) by \( \Pi \).

If positive pm’s \( P, Q \) are invariant to a Markov kernel \( \Pi \) on \( \Omega \)
then their log-affine combinations are invariant to \( \Pi \).

\[
P_A = P^A / P(A) \quad \text{... the truncation of } P \text{ to } A \text{ with } P(A) > 0,
\]
the normalized restriction.

Truncations of log-aff comb’s equal log-aff comb’s of truncations.

Chentsov, N.N. (1972,82):
\( \Pi \) ... a Markov kernel between finite sets \( \Omega, \Omega' \),
\[
\sum_{\omega' \in \Omega'} \Pi(\omega, \omega') = 1, \text{ each } \Pi(\omega, \omega') \geq 0 \text{ nonnegative.}
\]
Pm’s \( P \) on \( \Omega \) transform to the pm’s \( P\Pi \) on \( \Omega' \) by \( \Pi \).

If positive pm’s \( P, Q \) are invariant to a Markov kernel \( \Pi \) on \( \Omega \) then their log-affine combinations are invariant to \( \Pi \).

\[
P_A = P^A / P(A) \quad \text{... the truncation of } P \text{ to } A \text{ with } P(A) > 0,
\]
the normalized restriction.

Truncations of log-aff comb’s equal log-aff comb’s of truncations.

Chentsov, N.N. (1972,82):
geometry of pm’s, also differential
\( \Pi \) ... a Markov kernel between finite sets \( \Omega, \Omega' \),
\[
\sum_{\omega' \in \Omega'} \Pi(\omega, \omega') = 1, \text{ each } \Pi(\omega, \omega') \geq 0 \text{ nonnegative. }
\]
Pm’s \( P \) on \( \Omega \) transform to the pm’s \( P \Pi \) on \( \Omega' \) by \( \Pi \).

If positive pm’s \( P, Q \) are invariant to a Markov kernel \( \Pi \) on \( \Omega \)
then their log-affine combinations are invariant to \( \Pi \).

\[ P_A = P^A / P(A) \] ... the truncation of \( P \) to \( A \) with \( P(A) > 0 \),
the normalized restriction.

Truncations of log-aff comb’s equal log-aff comb’s of truncations.

Chentsov, N.N. (1972,82):
geometry of pm’s, also differential
categories of pm’s with Markov morphisms
Exponential family (\(\text{EF, full}\))

is the log-affine family of pm's sitting on the same set.
**Exponential family** ($\text{EF, full}$)

is the log-affine family of pm’s sitting on the same set.

Fischer, R.A. (1934); Darmois, G. (1935); Koopman, L.H. (1936);
Pitman, E.J.G. (1936); Chentsov, N.N. (1972,82);
Barndorff-Nielsen, O. (1978); Brown, L.D. (1986); Letac, G. (1992); ...
Exponential family \((\mathbb{E}F, \text{full})\)

is the log-affine family of pm’s sitting on the same set.

Fischer, R.A. (1934); Darmois, G. (1935); Koopman, L.H. (1936);
Pitman, E.J.G. (1936); Chentsov, N.N. (1972,82);
Barndorff-Nielsen, O. (1978); Brown, L.D. (1986); Letac, G. (1992); ...

\(\Omega = \Omega_1 \times \Omega_2\)
Exponential family \((\text{EF, full})\)

is the log-affine family of pm’s sitting on the same set.

\[
\Omega = \Omega_1 \times \Omega_2
\]

Fischer, R.A. (1934); Darmois, G. (1935); Koopman, L.H. (1936); Pitman, E.J.G. (1936); Chentsov, N.N. (1972,82); Barndorff-Nielsen, O. (1978); Brown, L.D. (1986); Letac, G. (1992); ...
Exponential family ($\text{EF, full}$)

is the log-affine family of pm’s sitting on the same set.

\[ \Omega = \Omega_1 \times \Omega_2 \]

\( \mathcal{P} \) the positive product

measures on \( \Omega \)

Fischer, R.A. (1934); Darmois, G. (1935); Koopman, L.H. (1936);
Pitman, E.J.G. (1936); Chentsov, N.N. (1972,82);
Barndorff-Nielsen, O. (1978); Brown, L.D. (1986); Letac, G. (1992); ...
Exponential family \((\text{EF, full})\)

is the log-affine family of pm’s sitting on the same set.

\[
\Omega = \Omega_1 \times \Omega_2
\]

\(\mathcal{P}\) the positive product measures on \(\Omega\)

\((\Omega_1 = \Omega_2 = \{0, 1\})\)
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set,
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set,
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set, consists of the log-affine combinations, proportional to

$$\omega \mapsto P_1^{t_1}(\omega) \cdot \ldots \cdot P_d^{t_d}(\omega) \cdot P_0^{1-t_1-\ldots-t_d}(\omega).$$
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set, consists of the log-affine combinations, proportional to

$$\omega \mapsto P_{1}^{t_{1}}(\omega) \cdot \ldots \cdot P_{d}^{t_{d}}(\omega) \cdot P_{0}^{1-t_{1}-\ldots-t_{d}}(\omega).$$

With the notation $\mu = P_0$
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set, consists of the log-affine combinations, proportional to

$$\omega \mapsto P_1^{t_1}(\omega) \cdot \ldots \cdot P_d^{t_d}(\omega) \cdot P_0^{1-t_1-\ldots-t_d}(\omega).$$

With the notation $\mu = P_0$
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set, consists of the log-affine combinations, proportional to

$$\omega \mapsto P_1^{t_1}(\omega) \cdot \ldots \cdot P_d^{t_d}(\omega) \cdot P_0^{1-t_1-\ldots-t_d}(\omega).$$

With the notation $\mu = P_0$ and $f_i = \ln \frac{P_i}{P_0}$,
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set, consists of the log-affine combinations, proportional to

$$\omega \mapsto P_1^{t_1}(\omega) \cdot \ldots \cdot P_d^{t_d}(\omega) \cdot P_0^{1-t_1-\ldots-t_d}(\omega).$$

With the notation $\mu = P_0$ and $f_i = \ln \frac{P_i}{P_0}$, this is

$$\omega \mapsto \exp \left[ t_1 f_1(\omega) + \ldots + t_d f_d(\omega) \right] \mu(\omega).$$
The log-affine envelope of $P_0, P_1, \ldots, P_d$, sitting on the same set, consists of the log-affine combinations, proportional to

$$
\omega \mapsto P_{1}^{t_1}(\omega) \cdot \ldots \cdot P_{d}^{t_d}(\omega) \cdot P_{0}^{1-t_1-\ldots-t_d}(\omega).
$$

With the notation $\mu = P_0$ and $f_i = \ln \frac{P_i}{P_0}$, this is

$$
\omega \mapsto \exp \left[ t_1 f_1(\omega) + \ldots + t_d f_d(\omega) \right] \mu(\omega)
$$

or

$$
\omega \mapsto e^{\langle \theta, f(\omega) \rangle} \mu(\omega).
$$
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set, consists of the log-affine combinations, proportional to

$$\omega \mapsto P_1^{t_1}(\omega) \cdot \ldots \cdot P_d^{t_d}(\omega) \cdot P_0^{1-t_1-\ldots-t_d}(\omega).$$

With the notation $\mu = P_0$ and $f_i = \ln \frac{P_i}{P_0}$, this is

$$\omega \mapsto \exp \left[ t_1 f_1(\omega) + \ldots + t_d f_d(\omega) \right] \mu(\omega)$$

or

$$\omega \mapsto e^{\langle \theta, f(\omega) \rangle} \mu(\omega).$$
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set, consists of the log-affine combinations, proportional to

$$\omega \mapsto P_1^{t_1}(\omega) \cdot \ldots \cdot P_d^{t_d}(\omega) \cdot P_0^{1-t_1-\ldots-t_d}(\omega).$$

With the notation $\mu = P_0$ and $f_i = \ln \frac{P_i}{P_0}$, this is

$$\omega \mapsto \exp \left[ t_1 f_1(\omega) + \ldots + t_d f_d(\omega) \right] \mu(\omega)$$

or

$$\omega \mapsto e^{\langle \theta, f(\omega) \rangle} \mu(\omega).$$

$$\theta = (t_1, \ldots, t_d) \ldots \text{the canonical parameter}$$
The log-affine envelope of \( P_0, P_1, \ldots P_d \), sitting on the same set, consists of the log-affine combinations, proportional to

\[
\omega \mapsto P_{1}^{t_1}(\omega) \cdot \ldots \cdot P_{d}^{t_d}(\omega) \cdot P_{0}^{1-t_1-\ldots-t_d}(\omega).
\]

With the notation \( \mu = P_0 \) and \( f_i = \ln \frac{P_i}{P_0} \), this is

\[
\omega \mapsto \exp \left[ t_1 f_1(\omega) + \ldots + t_d f_d(\omega) \right] \mu(\omega)
\]

or

\[
\omega \mapsto e^{\langle \theta, f(\omega) \rangle} \mu(\omega).
\]

\( \theta = (t_1, \ldots, t_d) \) ... the canonical parameter

\( f = (f_1, \ldots, f_d) \) ... the directional statistic
The log-affine envelope of $P_0, P_1, \ldots P_d$, sitting on the same set, consists of the log-affine combinations, proportional to

$$
\omega \mapsto P_1^{t_1}(\omega) \cdot \ldots \cdot P_d^{t_d}(\omega) \cdot P_0^{1-t_1-\ldots-t_d}(\omega).
$$

With the notation $\mu = P_0$ and $f_i = \ln \frac{P_i}{P_0}$, this is

$$
\omega \mapsto \exp \left[ t_1 f_1(\omega) + \ldots + t_d f_d(\omega) \right] \mu(\omega)
$$

or

$$
\omega \mapsto e^{\langle \theta, f(\omega) \rangle} \mu(\omega).
$$

$\theta = (t_1, \ldots, t_d)$ ... the canonical parameter

$f = (f_1, \ldots, f_d)$ ... the directional statistic

$\langle \cdot, \cdot \rangle$ ... the scalar product on $\mathbb{R}^d$
Hence, the full EF consists of the pm’s

\[ Q_{\mu,f,\theta}(\omega) = \exp \left[ \langle \theta, f(\omega) \rangle - \Lambda_{\mu,f}(\theta) \right] \cdot \mu(\omega), \quad \omega \in \Omega, \]
Hence, the full EF consists of the pm’s

\[ Q_{\mu, f, \theta}(\omega) = \exp \left[ \langle \theta, f(\omega) \rangle - \Lambda_{\mu, f}(\theta) \right] \cdot \mu(\omega), \quad \omega \in \Omega, \]

where \( \theta \in \mathbb{R}^d \) and \( \Lambda_{\mu, f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} \exp[\langle \theta, f(\omega) \rangle] \cdot \mu(\omega) \right]. \)
Hence, the full EF consists of the pm’s

\[ Q_{\mu,f,\theta}(\omega) = \exp \left[ \langle \theta, f(\omega) \rangle - \Lambda_{\mu,f}(\theta) \right] \cdot \mu(\omega), \quad \omega \in \Omega, \]

where \( \theta \in \mathbb{R}^d \) and \( \Lambda_{\mu,f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} \exp[\langle \theta, f(\omega) \rangle] \cdot \mu(\omega) \right]. \)

On the other hand, starting with
Hence, the full EF consists of the pm’s

\[ Q_{\mu,f,\theta}(\omega) = \exp \left[ \langle \theta, f(\omega) \rangle - \Lambda_{\mu,f}(\theta) \right] \cdot \mu(\omega), \quad \omega \in \Omega, \]

where \( \theta \in \mathbb{R}^d \) and \( \Lambda_{\mu,f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} \exp[\langle \theta, f(\omega) \rangle] \cdot \mu(\omega) \right]. \)

On the other hand, starting with

a nonzero measure \( \mu \) on \( \Omega \).
Hence, the full EF consists of the pm’s

\[ Q_{\mu, f, \theta}(\omega) = \exp \left[ \langle \theta, f(\omega) \rangle - \Lambda_{\mu, f}(\theta) \right] \cdot \mu(\omega), \quad \omega \in \Omega, \]

where \( \theta \in \mathbb{R}^d \) and \( \Lambda_{\mu, f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} \exp[\langle \theta, f(\omega) \rangle] \cdot \mu(\omega) \right]. \)

On the other hand, starting with

- a nonzero measure \( \mu \) on \( \Omega \)
- and a directional statistic \( f : \Omega \to \mathbb{R}^d \),
Hence, the full $\text{EF}$ consists of the pm’s

$$Q_{\mu,f,\theta}(\omega) = \exp \left[ \langle \theta, f(\omega) \rangle - \Lambda_{\mu,f}(\theta) \right] \cdot \mu(\omega), \quad \omega \in \Omega,$$

where $\theta \in \mathbb{R}^d$ and $\Lambda_{\mu,f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} \exp[\langle \theta, f(\omega) \rangle] \cdot \mu(\omega) \right]$.

On the other hand, starting with

- a nonzero measure $\mu$ on $\Omega$
- and a directional statistic $f : \Omega \to \mathbb{R}^d$

$\mathcal{E}_{\mu,f} = \{Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d\}$ is log-affine, its pm’s sit on $s(\mu)$. 
Hence, the full \( \text{EF} \) consists of the pm’s

\[
Q_{\mu, f, \theta}(\omega) = \exp \left[ \langle \theta, f(\omega) \rangle - \Lambda_{\mu, f}(\theta) \right] \cdot \mu(\omega), \quad \omega \in \Omega,
\]

where \( \theta \in \mathbb{R}^d \) and \( \Lambda_{\mu, f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} \exp[\langle \theta, f(\omega) \rangle] \cdot \mu(\omega) \right] \).

On the other hand, starting with

- a nonzero measure \( \mu \) on \( \Omega \)
- and a directional statistic \( f: \Omega \to \mathbb{R}^d \),

\( \mathcal{E}_{\mu, f} = \{ Q_{\mu, f, \theta} : \theta \in \mathbb{R}^d \} \) is log-affine, its pm’s sit on \( s(\mu) \).

**Canonically convex \( \text{EF} \)**

\[
\{ Q_{\mu, f, \theta} : \theta \in \Theta \} \text{ for } \Theta \subseteq \mathbb{R}^d \text{ convex.}
\]
For $\Omega = \{0, 1, \ldots, n\}$, $\mu(\omega) = \binom{n}{\omega}$ and the embedding $f : \Omega \to \mathbb{R}$,
For $\Omega = \{0, 1, \ldots, n\}$, $\mu(\omega) = \binom{n}{\omega}$ and the embedding $f: \Omega \rightarrow \mathbb{R}$,
For \( \Omega = \{0, 1, \ldots, n\} \), \( \mu(\omega) = \binom{n}{\omega} \) and the embedding \( f: \Omega \to \mathbb{R} \),

\[
Q_{\mu, f, \theta}(\omega) = e^{\theta \omega - \Lambda_{\mu, f}(\theta)} \binom{n}{\omega}
\]

where

\[
\Lambda_{\mu, f}(\theta) = \ln \sum_{\omega=0}^{n} e^{\theta \omega} \binom{n}{\omega} = \ln (1 + e^{\theta})^n
\]
For \( \Omega = \{0, 1, \ldots, n\} \), \( \mu(\omega) = \binom{n}{\omega} \) and the embedding \( f: \Omega \rightarrow \mathbb{R} \),

\[
Q_{\mu,f,\theta}(\omega) = e^{\theta \omega - \Lambda_{\mu,f}(\theta)} \binom{n}{\omega}
\]

where

\[
\Lambda_{\mu,f}(\theta) = \ln \sum_{\omega=0}^{n} e^{\theta \omega} \binom{n}{\omega} = \ln \left(1 + e^{\theta}\right)^n
\]
For $\Omega = \{0, 1, \ldots, n\}$, $\mu(\omega) = \binom{n}{\omega}$ and the embedding $f : \Omega \to \mathbb{R}$, 

$$Q_{\mu, f, \theta}(\omega) = e^{\theta \omega - \Lambda_{\mu, f}(\theta)} \binom{n}{\omega}$$

where

$$\Lambda_{\mu, f}(\theta) = \ln \sum_{\omega=0}^{n} e^{\theta \omega} \binom{n}{\omega} = \ln (1 + e^{\theta})^n$$

$$Q_{\mu, f, \theta}(\omega) = \binom{n}{\omega} p^\omega (1 - p)^{n-\omega}$$

where $p = \frac{e^{\theta}}{1 + e^{\theta}}$. 

Definition
Coordinatization of an EF
Mean parametrization
The closure of EF
For $\Omega = \{0, 1, \ldots, n\}$, $\mu(\omega) = \binom{n}{\omega}$ and the embedding $f: \Omega \to \mathbb{R}$, 

$$Q_{\mu, f, \theta}(\omega) = e^{\theta \omega - \Lambda_{\mu, f}(\theta)} \binom{n}{\omega}$$

where

$$\Lambda_{\mu, f}(\theta) = \ln \sum_{\omega=0}^{n} e^{\theta \omega} \binom{n}{\omega} = \ln \left(1 + e^{\theta}\right)^n$$

$$Q_{\mu, f, \theta}(\omega) = \binom{n}{\omega} p^\omega (1 - p)^{n-\omega}$$

where $p = \frac{e^{\theta}}{1 + e^{\theta}}$.

$\mathcal{E}_{\mu, f}$ is Binomial family.
\[ \mu \ldots \text{nonzero measure on } \Omega \]
\[ \mu \quad \text{... nonzero measure on } \Omega \]
\[ f : \Omega \rightarrow \mathbb{R}^d \quad \text{... a directional statistic} \]
\( \mu \) ... nonzero measure on \( \Omega \)
\( f : \Omega \to \mathbb{R}^d \) ... a directional statistic
\( \mu_f \) ... the \( f \)-image of \( \mu \), a Borel pm on \( \mathbb{R}^d \)
\( \mu \) ... nonzero measure on \( \Omega \)

\( f: \Omega \to \mathbb{R}^d \) ... a directional statistic

\( \mu_f \) ... the \( f \)-image of \( \mu \), a Borel pm on \( \mathbb{R}^d \)

concentrated on \( f(s(\mu)) = \{f(\omega): \omega \in s(\mu)\} \)
\( \mu \) ... nonzero measure on \( \Omega \)

\( f : \Omega \to \mathbb{R}^d \) ... a directional statistic

\( \mu_f \) ... the \( f \)-image of \( \mu \), a Borel pm on \( \mathbb{R}^d \)

concentrated on \( f(s(\mu)) = \{f(\omega) : \omega \in s(\mu)\} \)

\( cs(\mu_f) \) ... the convex support of \( \mu_f \),
\( \mu \) ... nonzero measure on \( \Omega \)
\( f: \Omega \to \mathbb{R}^d \) ... a directional statistic
\( \mu_f \) ... the \( f \)-image of \( \mu \), a Borel pm on \( \mathbb{R}^d \)

concentrated on \( f(s(\mu)) = \{f(\omega): \omega \in s(\mu)\} \)
\( cs(\mu_f) \) ... the convex support of \( \mu_f \),
the convex hull of \( f(s(\mu)) \), a polytope
\[ \mu \ldots \text{nonzero measure on } \Omega \]
\[ f : \Omega \to \mathbb{R}^d \ldots \text{a directional statistic} \]
\[ \mu_f \ldots \text{the } f\text{-image of } \mu, \text{a Borel pm on } \mathbb{R}^d \]
\[ \text{concentrated on } f(s(\mu)) = \{ f(\omega) : \omega \in s(\mu) \} \]
\[ cs(\mu_f) \ldots \text{the convex support of } \mu_f, \]
\[ \text{the convex hull of } f(s(\mu)), \text{a polytope} \]
\[ ri(\mu_f) \ldots \text{the relative interior of the polytope} \]
µ ... nonzero measure on Ω
f : Ω → ℝ^d ... a directional statistic
µ_f ... the f-image of µ, a Borel pm on ℝ^d
   concentrated on f(s(µ)) = {f(ω): ω ∈ s(µ)}
 cs(µ_f) ... the convex support of µ_f,
   the convex hull of f(s(µ)), a polytope
ri(µ_f) ... the relative interior of the polytope

Taking the mean \( E_P f = \sum_{ω ∈ Ω} f(ω)P(ω) \) of f under P,
\( P \mapsto E_P f \), is a homeomorphism between \( E_{µ,f} \) and \( ri(µ_f) \).
\( \mu \) ... nonzero measure on \( \Omega \)

\( f : \Omega \rightarrow \mathbb{R}^d \) ... a directional statistic

\( \mu_f \) ... the \( f \)-image of \( \mu \), a Borel pm on \( \mathbb{R}^d \)

concentrated on \( f(s(\mu)) = \{ f(\omega) : \omega \in s(\mu) \} \)

\( cs(\mu_f) \) ... the convex support of \( \mu_f \),

the convex hull of \( f(s(\mu)) \), a polytope

\( ri(\mu_f) \) ... the relative interior of the polytope

Taking the mean \( E_P f = \sum_{\omega \in \Omega} f(\omega) P(\omega) \) of \( f \) under \( P \),

\( P \mapsto E_P f \), is a homeomorphism between \( \mathcal{E}_{\mu,f} \) and \( ri(\mu_f) \).
\( \mu \) ... nonzero measure on \( \Omega \)

\( f : \Omega \rightarrow \mathbb{R}^d \) ... a directional statistic

\( \mu_f \) ... the \( f \)-image of \( \mu \), a Borel pm on \( \mathbb{R}^d \)

concentrated on \( f(s(\mu)) = \{ f(\omega) : \omega \in s(\mu) \} \)

\( cs(\mu_f) \) ... the convex support of \( \mu_f \),

the convex hull of \( f(s(\mu)) \), a polytope

\( ri(\mu_f) \) ... the relative interior of the polytope

Taking the mean \( E_P f = \sum_{\omega \in \Omega} f(\omega)P(\omega) \) of \( f \) under \( P \),

\( P \mapsto E_P f \), is a homeomorphism between \( \mathcal{E}_{\mu,f} \) and \( ri(\mu_f) \).
Recall

\[ \Lambda_{\mu,f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} e^{\langle \theta, f(\omega) \rangle} \cdot \mu(\omega) \right] = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu_f(dx) \]
Recall

\[ \Lambda_{\mu,f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} e^{\theta, f(\omega)} \cdot \mu(\omega) \right] = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu_f(dx) \]

the log-Laplace transform of the Borel measure \( \mu_f \)
Recall

\[ \Lambda_{\mu,f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} e^{\langle \theta, f(\omega) \rangle} \cdot \mu(\omega) \right] = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu_f(dx) \]

the log-Laplace transform of the Borel measure \( \mu_f \)
(capumulant generating function)
Recall

\[ \Lambda_{\mu,f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} e^{\langle \theta, f(\omega) \rangle} \cdot \mu(\omega) \right] = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu_f(dx) \]

the log-Laplace transform of the Borel measure \( \mu_f \)
(cumulant generating function)
convex, lower-semicontinuous
Recall

\[ \Lambda_{\mu,f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} e^{\langle \theta, f(\omega) \rangle} \cdot \mu(\omega) \right] = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu_f(dx) \]

the log-Laplace transform of the Borel measure \( \mu_f \)

(cumulant generating function)

convex, lower-semicontinuous

The gradient at \( \theta \)

\[ \sum_{\omega \in \Omega} f(\omega) \cdot e^{\langle \theta, f(\omega) \rangle} \cdot \mu(\omega) \]

\[ \sum_{\omega \in \Omega} e^{\langle \theta, f(\omega) \rangle} \cdot \mu(\omega) \]

\[ = \sum_{\omega \in \Omega} f(\omega) \cdot Q_{\mu,f,\theta}(\omega) \]
Recall

\[ \Lambda_{\mu,f}(\theta) = \ln \left[ \sum_{\omega \in \Omega} e^{\langle \theta, f(\omega) \rangle} \cdot \mu(\omega) \right] = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu_f(dx) \]

the log-Laplace transform of the Borel measure \( \mu_f \)

(cumulant generating function)

convex, lower-semicontinuous

The gradient at \( \theta \)

\[ \sum_{\omega \in \Omega} f(\omega) \cdot e^{\langle \theta, f(\omega) \rangle} \cdot \mu(\omega) \]

\[ \sum_{\omega \in \Omega} e^{\langle \theta, f(\omega) \rangle} \cdot \mu(\omega) \]

\[ = \sum_{\omega \in \Omega} f(\omega) \cdot Q_{\mu,f,\theta}(\omega) \]

... the mean of \( f \) under \( Q_{\mu,f,\theta} \).
The closure $cl(\mathcal{E}_{\mu,f})$ of an EF in the topology of $\mathbb{R}^\Omega$ equals

$$\bigcup_F \mathcal{E}_{\mu^{f^{-1}(F)},f}$$

where the union is over the (nonempty) faces $F$ of $cs(\mu_f)$. 
The closure $cl(\mathcal{E}_{\mu,f})$ of an EF in the topology of $\mathbb{R}^\Omega$ equals

$$
\bigcup_F \mathcal{E}_{\mu^{f^{-1}(F)},f}
$$

where the union is over the (nonempty) faces $F$ of $cs(\mu_f)$.

$\mu^{f^{-1}(F)}$ ... the restriction of $\mu$ to $f^{-1}(F) \subseteq \Omega$
The closure \( cl(\mathcal{E}_\mu, f) \) of an \( \text{EF} \) in the topology of \( \mathbb{R}^\Omega \) equals

\[
\bigcup_F \mathcal{E}_{\mu^{f^{-1}(F)}, f}
\]

where the union is over the (nonempty) faces \( F \) of \( cs(\mu_f) \).

\( \mu^{f^{-1}(F)} \) ... the restriction of \( \mu \) to \( f^{-1}(F) \subset \Omega \)

\[ \supseteq: \lim_{n \to \infty} Q_{\mu, f, \theta + n\vartheta} = Q_{\mu^F, f, \theta} \text{ for some } F \]
The closure \( cl(\mathcal{E}_\mu, f) \) of an EF in the topology of \( \mathbb{R}^\Omega \) equals

\[
\bigcup_F \mathcal{E}_{\mu^{f^{-1}(F)}, f}
\]

where the union is over the (nonempty) faces \( F \) of \( cs(\mu_f) \).

\( \mu^{f^{-1}(F)} \) ... the restriction of \( \mu \) to \( f^{-1}(F) \subseteq \Omega \)

\( \supseteq: \lim_{n \to \infty} Q_{\mu, f, \theta + n\vartheta} = Q_{\mu^F, f, \theta} \) for some \( F \)

\( \subseteq: \) by the mean parameterizations in the union
The closure $cl(\mathcal{E}_{\mu,f})$ of an EF in the topology of $\mathbb{R}^\Omega$ equals

$$\bigcup_F \mathcal{E}_{\mu^{f^{-1}(F)},f}$$

where the union is over the (nonempty) faces $F$ of $cs(\mu_f)$

$\mu^{f^{-1}(F)}$ ... the restriction of $\mu$ to $f^{-1}(F) \subseteq \Omega$

$\supseteq: \lim_{n \to \infty} Q_{\mu,f,\theta+n\theta} = Q_{\mu^F,f,\theta}$ for some $F$

$\subseteq$: by the mean parameterizations in the union

Taking the mean of the statistic $f$, $P \mapsto E_P f$,

is a homeomorphism between $cl(\mathcal{E}_{\mu,f})$ and $cs(\mu_f)$;
the component $\mathcal{E}_{\mu^{f^{-1}(F)},f}$ corresponds to $ri(F)$. 
The closure $\text{cl}(\mathcal{E}_\mu, f)$ of an EF in the topology of $\mathbb{R}^\Omega$ equals

$$
\bigcup_F \mathcal{E}_{\mu^{f^{-1}(F)}, f}
$$

where the union is over the (nonempty) faces $F$ of $\text{cs}(\mu_f)$

$\mu^{f^{-1}(F)}$ ... the restriction of $\mu$ to $f^{-1}(F) \subseteq \Omega$

$\supseteq$: $\lim_{n \to \infty} Q_{\mu, f, \theta + n\vartheta} = Q_{\mu^F, f, \theta}$ for some $F$

$\subseteq$: by the mean parameterizations in the union

Taking the mean of the statistic $f$, $P \mapsto E_P f$,

is a homeomorphism between $\text{cl}(\mathcal{E}_\mu, f)$ and $\text{cs}(\mu_f)$;

the component $\mathcal{E}_{\mu^{f^{-1}(F)}, f}$ corresponds to $\text{ri}(F)$.

For $a \in \text{cs}(\mu_f)$ denote by $R^*_{\mu, f}(a)$

the unique pm $P$ of $\text{cl}(\mathcal{E}_\mu, f)$ such that $a = E_P f$. 
sample \((\omega^{(1)}, ..., \omega^{(n)})\), an \(n\)-tuple of elements of \(\Omega\)
sample \((\omega^{(1)}, \ldots, \omega^{(n)})\), an \(n\)-tuple of elements of \(\Omega\)

\text{pm } P \text{ on } \Omega
sample \((\omega^{(1)}, \ldots, \omega^{(n)})\), an \(n\)-tuple of elements of \(\Omega\)

a fit between the sample and the pm can be rated by

\[
\mathbb{P}^n(\omega^{(1)}, \ldots, \omega^{(n)}) = \prod_{i=1}^{n} P(\omega^{(i)})
\]
sample \((\omega^{(1)}, \ldots, \omega^{(n)})\), an \(n\)-tuple of elements of \(\Omega\)

pm \(P\) on \(\Omega\)

a fit between the sample and the pm can be rated by

\[
P^n(\omega^{(1)}, \ldots, \omega^{(n)}) = \prod_{i=1}^{n} P(\omega^{(i)})
\]

\(P \mapsto \prod_{i=1}^{n} P(\omega^{(i)})\) ... the likelihood function (fn) given the sample
sample \((\omega^{(1)}, \ldots, \omega^{(n)})\), an \(n\)-tuple of elements of \(\Omega\)

a fit between the sample and the pm can be rated by

\[
P^n(\omega^{(1)}, \ldots, \omega^{(n)}) = \prod_{i=1}^{n} P(\omega^{(i)})
\]

\(P \mapsto \prod_{i=1}^{n} P(\omega^{(i)})\) ... the likelihood function (fn) given the sample

**Maximum likelihood (ML) principle**

A maximizer of the likelihood function over a family \(P\) (ML estimate) provides the explanation of the sample.
sample \((\omega^{(1)}, \ldots, \omega^{(n)})\), an \(n\)-tuple of elements of \(\Omega\)

A fit between the sample and the pm can be rated by

\[
P^n(\omega^{(1)}, \ldots, \omega^{(n)}) = \prod_{i=1}^{n} P(\omega^{(i)})
\]

\[
P \mapsto \prod_{i=1}^{n} P(\omega^{(i)}) \quad \text{... the likelihood function (fn) given the sample}
\]

**Maximum likelihood (ML) principle**

A maximizer of the likelihood function over a family \(P\)

(ML estimate) provides the explanation of the sample.

Lambert (1760); Bernoulli (1777); Laplace (1781); Gauss (1809);
Pearson (1896); Fisher (1922); ...
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$. 
The likelihood fn has at most one maximizer over a log-convex \( \mathcal{P} \).

(up to the trivial cases when it is identically 0 on \( \mathcal{P} \))
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$.

(up to the trivial cases when it is identically 0 on $\mathcal{P}$)

If the likelihood fn at $P, Q \in \mathcal{P}$ equals $K > 0$ then
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$.

(up to the trivial cases when it is identically 0 on $\mathcal{P}$)

If the likelihood fn at $P, Q \in \mathcal{P}$ equals $K > 0$ then
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$.

(upto the trivial cases when it is identically 0 on $\mathcal{P}$)

If the likelihood fn at $P, Q \in \mathcal{P}$ equals $K > 0$ then $s(P)$ and $s(Q)$ contain $\{\omega^{(1)}, ..., \omega^{(n)}\}$,
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$.

(up to the trivial cases when it is identically 0 on $\mathcal{P}$)

If the likelihood fn at $P, Q \in \mathcal{P}$ equals $K > 0$ then
$s(P)$ and $s(Q)$ contain $\{\omega^{(1)}, ..., \omega^{(n)}\}$,
the log-convex combination $P^t Q^{1-t}$ makes sense,
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$.

(up to the trivial cases when it is identically 0 on $\mathcal{P}$)

If the likelihood fn at $P, Q \in \mathcal{P}$ equals $K > 0$ then
$s(P)$ and $s(Q)$ contain $\{\omega^{(1)}, ..., \omega^{(n)}\}$,
the log-convex combination $P^t Q^{1-t}$ makes sense,
belongs to $\mathcal{P}$
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$.

(up to the trivial cases when it is identically 0 on $\mathcal{P}$)

If the likelihood fn at $P, Q \in \mathcal{P}$ equals $K > 0$ then
$s(P)$ and $s(Q)$ contain $\{\omega^{(1)}, \ldots, \omega^{(n)}\}$,
the log-convex combination $P^t Q^{1-t}$ makes sense, belongs to $\mathcal{P}$ and

$$K = \left[ \prod_{i=1}^{n} P(\omega^{(i)}) \right]^t \left[ \prod_{i=1}^{n} Q(\omega^{(i)}) \right]^{1-t} \leq \prod_{i=1}^{n} P^t Q^{1-t}(\omega^{(i)})$$
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$.

(up to the trivial cases when it is identically 0 on $\mathcal{P}$)

If the likelihood fn at $P, Q \in \mathcal{P}$ equals $K > 0$ then $s(P)$ and $s(Q)$ contain $\{\omega^{(1)}, \ldots, \omega^{(n)}\}$, the log-convex combination $P^t Q^{1-t}$ makes sense, belongs to $\mathcal{P}$ and

$$K = \left[ \prod_{i=1}^{n} P(\omega^{(i)}) \right]^t \left[ \prod_{i=1}^{n} Q(\omega^{(i)}) \right]^{1-t} \leq \prod_{i=1}^{n} P^t Q^{1-t}(\omega^{(i)})$$

as the normalizing constant is $\geq 1$, tight iff $P = Q$. 
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$.

(up to the trivial cases when it is identically 0 on $\mathcal{P}$)

If the likelihood fn at $P, Q \in \mathcal{P}$ equals $K > 0$ then $s(P)$ and $s(Q)$ contain $\{\omega^{(1)}, ..., \omega^{(n)}\}$, the log-convex combination $P^t Q^{1-t}$ makes sense, belongs to $\mathcal{P}$ and

$$K = \left[ \prod_{i=1}^{n} P(\omega^{(i)}) \right]^t \left[ \prod_{i=1}^{n} Q(\omega^{(i)}) \right]^{1-t} \leq \prod_{i=1}^{n} P^t Q^{1-t}(\omega^{(i)})$$

as the normalizing constant is $\geq 1$, tight iff $P = Q$.

If $\mathcal{P}$ is log-affine (log-convex) then $cl(\mathcal{P})$ has the same property.
The likelihood fn has at most one maximizer over a log-convex $\mathcal{P}$.

(up to the trivial cases when it is identically 0 on $\mathcal{P}$)

If the likelihood fn at $P, Q \in \mathcal{P}$ equals $K > 0$ then $s(P)$ and $s(Q)$ contain $\{\omega^{(1)}, ..., \omega^{(n)}\}$, the log-convex combination $P^t Q^{1-t}$ makes sense, belongs to $\mathcal{P}$ and

$$K = \left[ \prod_{i=1}^{n} P(\omega^{(i)}) \right]^t \left[ \prod_{i=1}^{n} Q(\omega^{(i)}) \right]^{1-t} \leq \prod_{i=1}^{n} P^t Q^{1-t}(\omega^{(i)})$$

as the normalizing constant is $\geq 1$, tight iff $P = Q$.

If $\mathcal{P}$ is log-affine (log-convex) then $cl(\mathcal{P})$ has the same property.

The ML estimate in any closed log-convex set exists and is unique.
For $\mathcal{P}$ equal to the EF $\mathcal{E}_{\mu,f} = \{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \}$,
For $\mathcal{P}$ equal to the EF $\mathcal{E}_{\mu,f} = \{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \}$, the fit between the sample $\omega^{(1)}, \ldots, \omega^{(n)}$ and $Q_{\mu,f,\theta}$ is rated by
For $\mathcal{P}$ equal to the EF $\mathcal{E}_{\mu,f} = \{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \}$, the fit between the sample $\omega^{(1)}, \ldots, \omega^{(n)}$ and $Q_{\mu,f,\theta}$ is rated by

$$\prod_{i=1}^{n} Q_{\mu,f,\theta}(\omega^{(i)}) = \prod_{i=1}^{n} \exp \left[ \langle \theta, f(\omega^{(i)}) \rangle - A_{\mu,f}(\theta) \right] \cdot \mu(\omega^{(i)}) .$$
For $\mathcal{P}$ equal to the EF $\mathcal{E}_{\mu,f} = \{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \}$, the fit between the sample $\omega^{(1)}, \ldots, \omega^{(n)}$ and $Q_{\mu,f,\theta}$ is rated by

$$\prod_{i=1}^{n} Q_{\mu,f,\theta}(\omega^{(i)}) = \prod_{i=1}^{n} \exp \left[ \langle \theta, f(\omega^{(i)}) \rangle - \Lambda_{\mu,f}(\theta) \right] \cdot \mu(\omega^{(i)}).$$

To maximize over $\theta$,
For \( \mathcal{P} \) equal to the \( \text{EF} \) \( \mathcal{E}_{\mu,f} = \{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \} \),
the fit between the sample \( \omega^{(1)}, \ldots, \omega^{(n)} \) and \( Q_{\mu,f,\theta} \) is rated by

\[
\prod_{i=1}^{n} Q_{\mu,f,\theta}(\omega^{(i)}) = \prod_{i=1}^{n} \exp \left[ \langle \theta, f(\omega^{(i)}) \rangle - \Lambda_{\mu,f}(\theta) \right] \cdot \mu(\omega^{(i)}).
\]

To maximize over \( \theta \), disregard \( \mu(\omega^{(i)}) \),
For $\mathcal{P}$ equal to the EF $\mathcal{E}_{\mu,f} = \{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \}$, the fit between the sample $\omega^{(1)}, ..., \omega^{(n)}$ and $Q_{\mu,f,\theta}$ is rated by

$$\prod_{i=1}^{n} Q_{\mu,f,\theta}(\omega^{(i)}) = \prod_{i=1}^{n} \exp \left[ \langle \theta, f(\omega^{(i)}) \rangle - \Lambda_{\mu,f}(\theta) \right] \cdot \mu(\omega^{(i)}).$$

To maximize over $\theta$, disregard $\mu(\omega^{(i)})$, take ln,
For $\mathcal{P}$ equal to the EF $\mathcal{E}_{\mu,f} = \{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \}$, the fit between the sample $\omega^{(1)}, \ldots, \omega^{(n)}$ and $Q_{\mu,f,\theta}$ is rated by

$$
\prod_{i=1}^{n} Q_{\mu,f,\theta}(\omega^{(i)}) = \prod_{i=1}^{n} \exp \left[ \langle \theta, f(\omega^{(i)}) \rangle - A_{\mu,f}(\theta) \right] \cdot \mu(\omega^{(i)}) .
$$

To maximize over $\theta$, disregard $\mu(\omega^{(i)})$, take ln, and divide by $n$: 
For $\mathcal{P}$ equal to the EF $\mathcal{E}_\mu,f = \{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \}$, the fit between the sample $\omega^{(1)}, \ldots, \omega^{(n)}$ and $Q_{\mu,f,\theta}$ is rated by

$$\prod_{i=1}^{n} Q_{\mu,f,\theta}(\omega^{(i)}) = \prod_{i=1}^{n} \exp \left[ \langle \theta, f(\omega^{(i)}) \rangle - A_{\mu,f}(\theta) \right] \cdot \mu(\omega^{(i)}).$$

To maximize over $\theta$, disregard $\mu(\omega^{(i)})$, take ln, and divide by $n$: a parametric variant of the normalized log-likelihood function

$$\theta \mapsto \langle \theta, a_f \rangle - A_{\mu,f}(\theta)$$
For $\mathcal{P}$ equal to the EF $\mathcal{E}_{\mu,f} = \{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \}$, the fit between the sample $\omega^{(1)}, ..., \omega^{(n)}$ and $Q_{\mu,f,\theta}$ is rated by

$$\prod_{i=1}^{n} Q_{\mu,f,\theta}(\omega^{(i)}) = \prod_{i=1}^{n} \exp \left[ \langle \theta, f(\omega^{(i)}) \rangle - \Lambda_{\mu,f}(\theta) \right] \cdot \mu(\omega^{(i)}).$$

To maximize over $\theta$, disregard $\mu(\omega^{(i)})$, take ln, and divide by $n$: a parametric variant of the normalized log-likelihood function

$$\theta \mapsto \langle \theta, a_f \rangle - \Lambda_{\mu,f}(\theta)$$

where $a_f = \frac{1}{n} \sum_{i=1}^{n} f(\omega^{(i)})$ is the empirical mean of $f$. 
For $\mathcal{P}$ equal to the EF $\mathcal{E}_{\mu,f} = \left\{ Q_{\mu,f,\theta} : \theta \in \mathbb{R}^d \right\}$, the fit between the sample $\omega^{(1)}, \ldots, \omega^{(n)}$ and $Q_{\mu,f,\theta}$ is rated by

$$\prod_{i=1}^{n} Q_{\mu,f,\theta}(\omega^{(i)}) = \prod_{i=1}^{n} \exp \left[ \langle \theta, f(\omega^{(i)}) \rangle - \Lambda_{\mu,f}(\theta) \right] \cdot \mu(\omega^{(i)}).$$

To maximize over $\theta$, disregard $\mu(\omega^{(i)})$, take ln, and divide by $n$: a parametric variant of the normalized log-likelihood function

$$\theta \mapsto \langle \theta, a_f \rangle - \Lambda_{\mu,f}(\theta)$$

where $a_f = \frac{1}{n} \sum_{i=1}^{n} f(\omega^{(i)})$ is the empirical mean of $f$.

A maximizer $\theta^*$ exists if and only if $a_f \in ri(\mu_f)$, in which case $a_f$ equals the $Q_{\mu,f,\theta^*}$-mean of $f$. The original likelihood fn has the unique maximizer

$$Q_{\mu,f,\theta^*} = R_{\mu,f}^*(a_f).$$
The MLE in \( cl(\mathcal{E}_{\mu,f}) \) from the sample with the empirical mean \( a_f \) equals \( R^*_{\mu,f}(a_f) \).
The MLE in $\text{cl}(E_{\mu,f})$ from the sample with the empirical mean $a_f$ equals $R^*_{\mu,f}(a_f)$. 
The MLE in $cl(\mathcal{E}_{\mu,f})$ from the sample with the empirical mean $a_f$ equals $R^*_{\mu,f}(a_f)$.

There is a unique face $F$ of $cs(\mu_f)$ such that $a_f \in ri(F)$,
The MLE in $cl(\mathcal{E}_{\mu,f})$ from the sample with the empirical mean $a_f$ equals $R^*_{\mu,f}(a_f)$.

There is a unique face $F$ of $cs(\mu_f)$ such that $a_f \in ri(F)$, then the MLE in $\mathcal{E}_{\mu^f-1(F),f}$ exists uniquely
The MLE in $\text{cl}(\mathcal{E}_{\mu,f})$ from the sample with the empirical mean $a_f$ equals $R^{\ast}_{\mu,f}(a_f)$.

There is a unique face $F$ of $cs(\mu_f)$ such that $a_f \in ri(F)$, then the MLE in $\mathcal{E}_{\mu_f^{-1}(F),f}$ exists uniquely and equals $R^{\ast}_{\mu_f^{-1}(F),f}(a_f)$. 
The MLE in \( cl(\mathcal{E}_{\mu, f}) \) from the sample with the empirical mean \( a_f \) equals \( R^*_{\mu, f}(a_f) \).

There is a unique face \( F \) of \( cs(\mu_f) \) such that \( a_f \in ri(F) \), then the MLE in \( \mathcal{E}_{\mu_f^{-1}(F), f} \) exists uniquely and equals \( R^*_{\mu_f^{-1}(F), f}(a_f) \) which coincides with \( R^*_{\mu, f}(a_f) \).
The (full, standard) exponential family $\mathcal{E}$
The (full, standard) **exponential family** $\mathcal{E}$
determined by a nonzero Borel measure $\mu$ on $\mathbb{R}^d$
The (full, standard) exponential family $\mathcal{E}$ determined by a nonzero Borel measure $\mu$ on $\mathbb{R}^d$ consists of the pm’s $Q_\theta$ with $\mu$-densities
The (full, standard) exponential family $\mathcal{E}$
determined by a nonzero Borel measure $\mu$ on $\mathbb{R}^d$
consists of the pm’s $Q_\theta$ with $\mu$-densities

$$\frac{dQ_\theta}{d\mu}(x) = \exp \left[ \langle \theta, x \rangle - \Lambda(\theta) \right]$$
The (full, standard) **exponential family** $\mathcal{E}$
determined by a nonzero Borel measure $\mu$ on $\mathbb{R}^d$
consists of the pm’s $Q_\theta$ with $\mu$-densities

$$
\frac{dQ_\theta}{d\mu}(x) = \exp \left[ \langle \theta, x \rangle - \Lambda(\theta) \right]
$$

where

$$
\Lambda(\theta) = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu(dx)
$$
The (full, standard) exponential family $\mathcal{E}$ determined by a nonzero Borel measure $\mu$ on $\mathbb{R}^d$ consists of the pm's $Q_\theta$ with $\mu$-densities

$$\frac{dQ_\theta}{d\mu}(x) = \exp \left[ \langle \theta, x \rangle - \Lambda(\theta) \right]$$

where

$$\Lambda(\theta) = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu(dx)$$

is the log-Laplace transform of $\mu$. 

Here, $\langle \cdot, \cdot \rangle$ denotes the inner product.
The (full, standard) exponential family $\mathcal{E}$ determined by a nonzero Borel measure $\mu$ on $\mathbb{R}^d$ consists of the pm’s $Q_\theta$ with $\mu$-densities

$$
\frac{dQ_\theta}{d\mu}(x) = \exp \left[ \langle \theta, x \rangle - \Lambda(\theta) \right]
$$

where

$$
\Lambda(\theta) = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu(dx)
$$

is the log-Laplace transform of $\mu$ and $\theta$ ranges over the effective domain of $\Lambda$

$$
dom(\Lambda) = \{ \theta : \Lambda(\theta) < +\infty \}.
$$
The (full, standard) exponential family $\mathcal{E}$
determined by a nonzero Borel measure $\mu$ on $\mathbb{R}^d$
consists of the pm’s $Q_{\theta}$ with $\mu$-densities

$$
\frac{dQ_{\theta}}{d\mu}(x) = \exp \left[ \langle \theta, x \rangle - \Lambda(\theta) \right]
$$

where

$$
\Lambda(\theta) = \ln \int_{\mathbb{R}^d} e^{\langle \theta, x \rangle} \mu(dx)
$$

is the log-Laplace transform of $\mu$

and $\theta$ ranges over the effective domain of $\Lambda$

$$
\text{dom}(\Lambda) = \{ \theta : \Lambda(\theta) < +\infty \}.
$$

$\mathcal{E}_\Xi = \{ Q_\theta : \theta \in \Xi \}$ where $\Xi \subseteq \text{dom}(\Lambda)$ is convex.
The likelihood, given the data $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$ w.r.t. $Q_\theta$

$$\frac{dQ_\theta}{d\mu}(x^{(1)}) \ldots \frac{dQ_\theta}{d\mu}(x^{(n)}) = \exp[\langle \theta, na \rangle - n\Lambda(\theta)]$$
The likelihood, given the data \( x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d \) w.r.t. \( Q_\theta \)

\[
\frac{dQ_\theta}{d\mu} (x^{(1)}) \ldots \frac{dQ_\theta}{d\mu} (x^{(n)}) = \exp[\langle \theta, na \rangle - n\Lambda(\theta)]
\]

where \( a = \frac{1}{n} \sum_{i=1}^{n} x^{(i)} \) is the empirical mean.
The likelihood, given the data $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$ w.r.t. $Q_\theta$

$$\frac{dQ_\theta}{d\mu}(x^{(1)}) \ldots \frac{dQ_\theta}{d\mu}(x^{(n)}) = \exp[\langle \theta, na \rangle - n\Lambda(\theta)]$$

where $a = \frac{1}{n} \sum_{i=1}^{n} x^{(i)}$ is the empirical mean.

The maximization of the normalized log-likelihood means

$$\Psi^*(a) = \Psi_{\mu, \Xi}^*(a) = \sup_{\theta \in \Xi} [\langle \theta, a \rangle - \Lambda(\theta)].$$
The likelihood, given the data $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$ w.r.t. $Q_\theta$

$$
\frac{dQ_\theta}{d\mu}(x^{(1)}) \ldots \frac{dQ_\theta}{d\mu}(x^{(n)}) = \exp[\langle \theta, na \rangle - n\Lambda(\theta)]
$$

where $a = \frac{1}{n} \sum_{i=1}^{n} x^{(i)}$ is the empirical mean.

The maximization of the normalized log-likelihood means

$$
\Psi^*(a) = \Psi^*_{\mu,\Xi}(a) = \sup_{\theta \in \Xi} \left[ \langle \theta, a \rangle - \Lambda(\theta) \right].
$$

If $a$ is the mean of some pm $Q_{\theta^*}$ with $\theta^* \in \Xi$ then

$$
\Psi^*(a) - \left[ \langle \theta, a \rangle - \Lambda(\theta) \right] = D(Q_{\theta^*} \| Q_\theta), \quad \theta \in \Xi.
$$
The likelihood, given the data $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$ w.r.t. $Q_\theta$

$$
\frac{dQ_\theta}{d\mu} (x^{(1)}) \ldots \frac{dQ_\theta}{d\mu} (x^{(n)}) = \exp[\langle \theta, na \rangle - n\Lambda(\theta)]
$$

where $a = \frac{1}{n} \sum_{i=1}^{n} x^{(i)}$ is the empirical mean.

The maximization of the normalized log-likelihood means

$$
\Psi^*(a) = \Psi_{\mu,\Xi}^*(a) = \sup_{\theta \in \Xi} \left[ \langle \theta, a \rangle - \Lambda(\theta) \right].
$$

If $a$ is the mean of some pm $Q_{\theta^*}$ with $\theta^* \in \Xi$ then

$$
\Psi^*(a) - \left[ \langle \theta, a \rangle - \Lambda(\theta) \right] = D(Q_{\theta^*} \| Q_\theta) , \quad \theta \in \Xi.
$$

using the relative entropy

$$
D(P \| Q) = \begin{cases} \int_{\mathbb{R}^d} \ln \frac{dP}{dQ} dP & \text{if } P \ll Q \\ +\infty & \text{otherwise.} \end{cases}
$$
If $\Psi^*(a)$ is finite then a unique pm $R^*_{\mu,\Xi}(a)$ exists such that

$$
\Psi^*(a) - \left[ \langle \theta, a \rangle - \Lambda(\theta) \right] \geq D(R^*_{\mu,\Xi}(a)\| Q_\theta), \quad \theta \in \Xi.
$$

(IEEE Trans. IT, June 2003)
If $\Psi^*(a)$ is finite then a unique pm $R^*_\mu,\Xi(a)$ exists such that

$$\Psi^*(a) - [\langle \theta, a \rangle - \Lambda(\theta)] \geq D(R^*_\mu,\Xi(a)\|Q_\theta), \quad \theta \in \Xi.$$
If $\Psi^*(a)$ is finite then a unique pm $R^*_{\mu,\Xi}(a)$ exists such that

$$\Psi^*(a) - [\langle \theta, a \rangle - \Lambda(\theta)] \geq D(R^*_{\mu,\Xi}(a) \| Q_\theta), \quad \theta \in \Xi.$$

(a nonconstructive existence proof extends to families of infinite dimension)
If $\Psi^*(a)$ is finite then a unique pm $R_{\mu,\Xi}^*(a)$ exists such that

$$
\Psi^*(a) - \left[ \langle \theta, a \rangle - \Lambda(\theta) \right] \geq D(R_{\mu,\Xi}^*(a)\|Q_\theta), \quad \theta \in \Xi.
$$

(a nonconstructive existence proof extends to families of infinite dimension)

The pm $R^*(a) = R_{\mu,\Xi}^*(a)$ is called generalized MLE for $\mathcal{E}_\Xi$. 

(IEEE Trans. IT, June 2003)
If $\Psi^*(a)$ is finite then a unique pm $R^*_{\mu,\Xi}(a)$ exists such that

$$\Psi^*(a) - \left[ \langle \theta, a \rangle - \Lambda(\theta) \right] \geq D(R^*_{\mu,\Xi}(a) \parallel Q_\theta), \quad \theta \in \Xi.$$ 

(a nonconstructive existence proof extends to families of infinite dimension)

The pm $R^*(a) = R^*_{\mu,\Xi}(a)$ is called generalized MLE for $\mathcal{E}_\Xi$.

If a sequence $\theta_n$ in $\Xi$ satisfies $\langle \theta_n, a \rangle - \Lambda(\theta_n) \rightarrow \Psi^*(a)$
then $Q_{\theta_n} \rightarrow R^*(a)$ in the variation distance.
If $\Psi^*(a)$ is finite then a unique pm $R^*_{\mu,\Xi}(a)$ exists such that

$$\Psi^*(a) - [\langle \theta, a \rangle - \Lambda(\theta)] \geq D(R^*_{\mu,\Xi}(a)\| Q_{\theta}) , \quad \theta \in \Xi .$$

(a nonconstructive existence proof extends to families of infinite dimension)

The pm $R^*(a) = R^*_{\mu,\Xi}(a)$ is called generalized MLE for $\mathcal{E}_{\Xi}$.

If a sequence $\theta_n$ in $\Xi$ satisfies $\langle \theta_n, a \rangle - \Lambda(\theta_n) \to \Psi^*(a)$ then $Q_{\theta_n} \to R^*(a)$ in the variation distance.

The GMLE belongs to $cl_v(\mathcal{E}_{\Xi})$, the closure in variation distance (Annals of Probab. 2005).
Theorem

\[ \text{dom}(\Psi^*) = \text{cc}(\mu) + \bar{\text{bar}}(\Xi) \]
Theorem

\[ \text{dom}(\Psi^*) = \text{cc}(\mu) + \text{bar}(\Xi) \]
### Theorem

\[
\text{dom}(\Psi^*) = \text{cc}(\mu) + \text{bar}(\Xi)
\]

- \(\text{cc}(\mu)\) ... the convex core of \(\mu\)
- \(\text{bar}(\Xi)\) ... the barrier cone of \(\Xi\)

(a special convex subset of \(\text{cs}(\mu)\),
containing its relative interior \(\text{ri}(\mu)\))
**Theorem**

\[ \text{dom}(\Psi^*) = \text{cc}(\mu) + \text{bar}(\Xi) \]

- \( \text{cc}(\mu) \) ... the **convex core** of \( \mu \)
  - (a special convex subset of \( \text{cs}(\mu) \), containing its relative interior \( \text{ri}(\mu) \))

- \( \text{bar}(\Xi) \) ... the **barrier cone** of \( \Xi \).
Theorem

\[ \text{dom}(\Psi^*) = \text{cc}(\mu) + \text{bar}(\Xi) \]

- \( \text{cc}(\mu) \) ... the \textit{convex core} of \( \mu \)
  (a special convex subset of \( \text{cs}(\mu) \), containing its relative interior \( \text{ri}(\mu) \))

- \( \text{bar}(\Xi) \) ... the \textit{barrier cone} of \( \Xi \).

Even the instance \( \Xi = \text{dom}(\Lambda) \) gives a new formula for \( \text{dom}(\Lambda^*) \).
Theorem

\[ \text{dom}(\Psi^*) = cc(\mu) + \text{bar}(\Xi) \]

\(cc(\mu)\) ... the convex core of \(\mu\)

(a special convex subset of \(cs(\mu)\),

containing its relative interior \(ri(\mu)\))

\(\text{bar}(\Xi)\) ... the barrier cone of \(\Xi\).

Even the instance \(\Xi = \text{dom}(\Lambda)\) gives a new formula for \(\text{dom}(\Lambda^*)\).

Since no regularity conditions are imposed

the classical convex analysis of MLE’s has to be revisited

\[ \theta^* = \theta^*_{\mu,\Xi} : \, ri(\mu) + \text{bar}(\Xi) \rightarrow \text{dom}(\Lambda) \]
Theorem

\[ \text{dom}(\Psi^*) = \text{cc}(\mu) + \text{bar}(\Xi) \]

\( \text{cc}(\mu) \) ... the **convex core** of \( \mu \)

(a special convex subset of \( \text{cs}(\mu) \), containing its relative interior \( \text{ri}(\mu) \))

\( \text{bar}(\Xi) \) ... the **barrier cone** of \( \Xi \).

Even the instance \( \Xi = \text{dom}(\Lambda) \) gives a new formula for \( \text{dom}(\Lambda^*) \).

Since no regularity conditions are imposed

the classical convex analysis of MLE’s has to be revisited

\[ \theta^* = \theta^*_{\mu,\Xi} : \text{ri}(\mu) + \text{bar}(\Xi) \to \text{dom}(\Lambda) \]

to cover the cases when \( \mathcal{E}_\Xi \) is overparameterized

or \( a \) is out of the affine hull of \( \text{cs}(\mu) \).
Theorem

For $a \in ri(\mu) + \text{bar}(\Xi)$, the GMLE $R^*(a)$ equals $Q_{\theta^*}(a) \in \mathcal{E}$. 
Theorem

For $a \in ri(\mu) + bar(\Xi)$, the GMLE $R^*(a)$ equals $Q_{\theta^*}(a) \in \mathcal{E}$. 
Theorem

For $a \in ri(\mu) + bar(\Xi)$, the GMLE $R^*(a)$ equals $Q_{\theta^*}(a) \in \mathcal{E}$.

... this is a revised MLE.
For $a \in ri(\mu) + \overline{\Xi}$, the GMLE $R^*(a)$ equals $Q_{\theta^*}(a) \in \mathcal{E}$.

... this is a revised MLE.

If $\Psi^*_{\mu,\Xi}(a)$ is finite then 
the GMLE $R^*_{\mu,\Xi}(a)$ equals the GMLE $R^*_{\nu,\Xi}(a)$
where $\nu$ is the restriction of $\mu$ to $cl(G)$
for a special face $G = G^*_{\mu,\Xi}(a)$ of $cc(\mu)$
and $R^*_{\nu,\Xi}(a)$ obtains by the revisited MLE.
Theorem

For $a \in ri(\mu) + \text{bar}(\Xi)$, the GMLE $R^*(a)$ equals $Q_{\theta^*}(a) \in \mathcal{E}$.

... this is a revised MLE.

Theorem

If $\Psi^*_{\mu,\Xi}(a)$ is finite then
the GMLE $R^*_{\mu,\Xi}(a)$ equals the GMLE $R^*_{\nu,\Xi}(a)$
where $\nu$ is the restriction of $\mu$ to $\text{cl}(G)$
for a special face $G = G^*_{\mu,\Xi}(a)$ of $\text{cc}(\mu)$
and $R^*_{\nu,\Xi}(a)$ obtains by the revisited MLE.
Theorem

For \( a \in ri(\mu) + \text{bar}(\Xi) \), the GMLE \( R^*(a) \) equals \( Q_{\theta^*}(a) \in \mathcal{E} \).

... this is a revised MLE.

Theorem

If \( \Psi^*_{\mu,\Xi}(a) \) is finite then

the GMLE \( R^*_{\mu,\Xi}(a) \) equals the GMLE \( R^*_{\nu,\Xi}(a) \)

where \( \nu \) is the restriction of \( \mu \) to \( cl(G) \)

for a special face \( G = G^*_{\mu,\Xi}(a) \) of \( cc(\mu) \)

and \( R^*_{\nu,\Xi}(a) \) obtains by the revisited MLE.

(a proof by induction on the dimension of \( \text{aff}(\mu) \))
\( R^*: a \mapsto R^*(a), \text{ on } dom(\Psi^*) \)
\[ R^*: a \mapsto R^*(a), \text{ on } dom(\Psi^*) \]

The range of \( R^* \) consists of the pm’s \( P \in cl_v(\mathcal{E}_\Xi) \) with means. (assuming \( \Xi \) intersects the interior of \( dom(\Lambda) \))
$R^*: a \mapsto R^*(a)$, on $\text{dom}(Ψ^*)$

The range of $R^*$ consists of the pm’s $P \in cl_\nu(\mathcal{E}_\Xi)$ with means.
(assuming $\Xi$ intersects the interior of $\text{dom}(Λ)$)

The inverse image $\{a: R^*(a) = P\}$ is a shifted cone.
(not necessarily convex)
$R^*: a \mapsto R^*(a)$, on $\text{dom}(\Psi^*)$

The range of $R^*$ consists of the pm’s $P \in \text{cl}_{\nu}(\mathcal{E}_\Xi)$ with means.

(assuming $\Xi$ intersects the interior of $\text{dom}(\Lambda)$)

The inverse image $\{a: R^*(a) = P\}$ is a shifted cone.

(not necessarily convex)

The GMLE mapping is continuous, assuming

$\text{dom}(\Psi^*)$ has the topology of the graph of $\Psi^*$

$\text{cl}_{\nu}(\mathcal{E}_\Xi)$ has the topology of variation distance.
$R^*: a \mapsto R^*(a)$, on $\text{dom}(\Psi^*)$

The range of $R^*$ consists of the pm’s $P \in \text{cl}_\nu(E_{\Xi})$ with means.

(assuming $\Xi$ intersects the interior of $\text{dom}(\Lambda)$)

The inverse image $\{a: R^*(a) = P\}$ is a shifted cone.

(not necessarily convex)

The GMLE mapping is continuous, assuming

$\text{dom}(\Psi^*)$ has the topology of the graph of $\Psi^*$

$\text{cl}_\nu(E_{\Xi})$ has the topology of variation distance.

If MLE in $\text{cl}_\nu(E_{\Xi})$ exists then it coincides with the GMLE for $E_{\Xi}$. 
The **Fenchel conjugate** of the log-Laplace transform of $\mu_f$

\[
\Lambda_{\mu,f}^*(a) = \sup_{\theta \in \mathbb{R}^d} \left[ \langle \theta, a \rangle - \Lambda_{\mu,f}(\theta) \right], \quad a \in \mathbb{R}^d,
\]

is finite if and only if $a \in \text{cs}(\mu_f)$. 

For the binomial family, $\Lambda_{\mu,f}^*(\epsilon) = \epsilon \ln \epsilon + \epsilon \left[ -1 - \ln n \right] + o(\epsilon)$. 

For $\epsilon > 0$ small

\[
\Lambda_{\mu,f}^*(\epsilon) = \epsilon \ln \epsilon + \epsilon \left[ -1 - \ln n \right] + o(\epsilon).
\]
The Fenchel conjugate of the log-Laplace transform of $\mu_f$

$$\Lambda_{\mu,f}^*(a) = \sup_{\theta \in \mathbb{R}^d} \left[ \langle \theta, a \rangle - \Lambda_{\mu,f}(\theta) \right], \quad a \in \mathbb{R}^d,$$

is finite if and only if $a \in cs(\mu_f)$. 
The **Fenchel conjugate** of the log-Laplace transform of $\mu_f$

$$\Lambda^*_{\mu,f}(a) = \sup_{\theta \in \mathbb{R}^d} \left[ \langle \theta, a \rangle - \Lambda_{\mu,f}(\theta) \right], \quad a \in \mathbb{R}^d,$$

is finite if and only if $a \in cs(\mu_f)$.

For the binomial family, $\Lambda^*_{\mu,f}$ is finite on $[0, n]$
(can be computed explicitly)
The **Fenchel conjugate** of the log-Laplace transform of $\mu_f$

$$
\Lambda_{\mu,f}^*(a) = \sup_{\theta \in \mathbb{R}^d} \left[ \langle \theta, a \rangle - \Lambda_{\mu,f}(\theta) \right], \quad a \in \mathbb{R}^d,
$$

is finite if and only if $a \in cs(\mu_f)$.

For the binomial family, $\Lambda_{\mu,f}^*$ is finite on $[0, n]$

(can be computed explicitly)

For $\varepsilon > 0$ small

$$
\Lambda^*(\varepsilon) = \varepsilon \ln \varepsilon + \varepsilon[-1 - \ln n] + o(\varepsilon).
$$
For the family of the positive product measures on $\Omega = \{0, 1\}^2$, 

...
For the family of the positive product measures on $\Omega = \{0, 1\}^2$, the conjugate is finite on a square.
For the family of the positive product measures on $\Omega = \{0, 1\}^2$, the conjugate is finite on a square

By (FM 2007), starting at any boundary point $a$ and moving inside,

$$\Lambda^*(a + \varepsilon (b - a)) = \Lambda^*(a) + C_1 \cdot \varepsilon \ln \varepsilon + C_2 \cdot \varepsilon + o(\varepsilon)$$

where the constants $C_1$, $C_2$ can be explicitly constructed.
The divergence of a pm $P$ from a family $\mathcal{E} = \mathcal{E}_{\mu,f}$

$$D(P\|\mathcal{E}) = \inf_{\theta \in \mathbb{R}^d} D(P\|Q_\theta).$$
The divergence of a pm $P$ from a family $\mathcal{E} = \mathcal{E}_{\mu, f}$

$$D(P \| \mathcal{E}) = \inf_{\theta \in \mathbb{R}^d} D(P \| Q_\theta).$$
\[ D(P \parallel \mathcal{E}_{\mu, f}) = \inf_{\theta \in \mathbb{R}^d} \sum_{\omega \in s(P)} \left[ \ln \frac{P(\omega)}{\mu(\omega)} - \ln \frac{Q_{\mu, f, \theta}(\omega)}{\mu(\omega)} \right] P(\omega) \]
\[ D(P \| \mathcal{E}_{\mu,f}) = \inf_{\theta \in \mathbb{R}^d} \sum_{\omega \in s(P)} \left[ \ln \frac{P(\omega)}{\mu(\omega)} - \ln \frac{Q_{\mu,f,\theta}(\omega)}{\mu(\omega)} \right] P(\omega) \]

\[ = D(P \| \mu) + \inf_{\theta \in \mathbb{R}^d} \sum_{\omega \in s(P)} \left[ - \ln e^{\langle \theta, f(\omega) \rangle - \Lambda(\theta)} \right] P(\omega) \]
\[ D(P\|\mathcal{E}_\mu,f) = \inf_{\theta \in \mathbb{R}^d} \sum_{\omega \in \mathcal{S}(P)} \left[ \ln \frac{P(\omega)}{\mu(\omega)} - \ln \frac{Q_{\mu,f,\theta}(\omega)}{\mu(\omega)} \right] P(\omega) \]

\[ = D(P\|\mu) + \inf_{\theta \in \mathbb{R}^d} \sum_{\omega \in \mathcal{S}(P)} \left[ - \ln e^{\langle \theta, f(\omega) \rangle - \Lambda(\theta)} \right] P(\omega) \]

\[ = D(P\|\mu) - \sup_{\theta \in \mathbb{R}^d} \left[ \langle \theta, \sum_{\omega \in \mathcal{S}(P)} f(\omega) P(\omega) \rangle - \Lambda(\theta) \right] \]
\begin{align*}
D(P\|\mathcal{E}_{\mu,f}) &= \inf_{\theta \in \mathbb{R}^d} \sum_{\omega \in s(P)} \left[ \ln \frac{P(\omega)}{\mu(\omega)} - \ln \frac{Q_{\mu,f,\theta}(\omega)}{\mu(\omega)} \right] P(\omega) \\
&= D(P\|\mu) + \inf_{\theta \in \mathbb{R}^d} \sum_{\omega \in s(P)} \left[ - \ln e^{\langle \theta, f(\omega) \rangle - \Lambda(\theta)} \right] P(\omega) \\
&= D(P\|\mu) - \sup_{\theta \in \mathbb{R}^d} \left[ \langle \theta, \sum_{\omega \in s(P)} f(\omega)P(\omega) \rangle - \Lambda(\theta) \right] \\
&= D(P\|\mu) - \Lambda^*(E_Pf) \quad \text{where } E_Pf = \sum f(\omega)P(\omega) \quad \text{is the } P\text{-mean of } f.
\end{align*}
Let $D(P \| \mathcal{E}_{\mu, f})$ be defined as:

$$D(P \| \mathcal{E}_{\mu, f}) = \inf_{\theta \in \mathbb{R}^d} \sum_{\omega \in s(P)} \left[ \ln \frac{P(\omega)}{\mu(\omega)} - \ln \frac{Q_{\mu, f, \theta}(\omega)}{\mu(\omega)} \right] P(\omega)$$

Then:

$$= D(P \| \mu) + \inf_{\theta \in \mathbb{R}^d} \sum_{\omega \in s(P)} \left[ - \ln e^{\langle \theta, f(\omega) \rangle} - \Lambda(\theta) \right] P(\omega)$$

$$= D(P \| \mu) - \sup_{\theta \in \mathbb{R}^d} \left[ \langle \theta, \sum_{\omega \in s(P)} f(\omega) P(\omega) \rangle - \Lambda(\theta) \right]$$

$$= D(P \| \mu) - \Lambda^*(E_P f)$$

where $E_P f = \sum f(\omega) P(\omega)$ is the $P$-mean of $f$.

... difference of two convex functions
Nihat Ay’s ideas and results (Annals of Probab. 2002)

Maximize $D(\cdot\|\mathcal{E})$. This has nice interpretations.
First order optimality conditions for a pm $P$ to be a maximizer
when $E_P f$ is inside the polytope $cs(\mu)$. 
Nihat Ay’s ideas and results (Annals of Probab. 2002)

Maximize $D(\cdot \| \mathcal{E})$. This has nice interpretations.
First order optimality conditions for a pm $P$ to be a maximizer when $E_P f$ is inside the polytope $cs(\mu)$.

FM 2007

All directional derivatives of $D(\cdot \| \mathcal{E})$ at any pm $P$.
All first order optimality conditions.