

The two connections we will establish

1. Starting from a fixed density p , its exponentially tilted densities form a **natural exponential family**.
2. If p is already a member of a natural exponential family, an admissible tilt stays in the **same family**.

The second result explains when a familiar sampling method can be reused.

original natural parameter θ_0 $\longrightarrow$ tilted natural parameter $\theta_0 + \theta$.

Review: exponential families

A parametric family has an exponential family representation if

$$p(x; \lambda) = h(x) \exp\left(\eta(\lambda)^\top T(x) - A(\lambda)\right).$$

- $h(x) \geq 0$ is a scalar base factor.
- $\eta(\lambda)$ and $T(x)$ have matching dimensions.
- $A(\lambda)$ is scalar and normalises the density:

$$\int p(x; \lambda) dx = 1.$$

We use a common support and a fixed underlying measure.

The natural exponential family special case

Take the natural parameter to be the parameter itself, and the statistic to be x :

$$\eta(\vartheta) = \vartheta, \quad T(x) = x.$$

Then

Natural exponential family

$$p(x; \vartheta) = h(x) \exp\left(\vartheta^\top x - A(\vartheta)\right).$$

Here x and ϑ have the same dimension.

In a general exponential family, $T(x)$ need not equal x .

The log-normaliser and its domain

The normalising function is

$$A(\vartheta) = \log \int h(x) e^{\vartheta^\top x} dx.$$

It is defined on the natural parameter domain

$$\mathcal{D} = \left\{ \vartheta : 0 < \int h(x) e^{\vartheta^\top x} dx < \infty \right\}.$$

Every parameter value used must lie in the admissible domain.

For discrete distributions, use sums in place of integrals.

Recall: exponential tilting

For a fixed density $p(x)$, define

$$q(x; \theta) = \frac{e^{\theta^\top x} p(x)}{C(\theta)}, \quad C(\theta) = \int e^{\theta^\top x} p(x) dx.$$

The tilting parameter θ is admissible when $C(\theta) < \infty$.

Since the integrand is positive wherever $p(x) > 0$, $C(\theta) > 0$.

$$\frac{1}{C(\theta)} = \exp(-\log C(\theta)).$$

Result 1: tilted densities form a natural family

Rewrite the tilted density as

$$q(x; \theta) = p(x) \exp\left(\theta^\top x - \log C(\theta)\right).$$

Natural exponential family component	Tilted family
Base factor	$p(x)$
Statistic	x
Natural parameter	θ
Log-normalising function	$\log C(\theta)$

The original density need not belong to a familiar natural family.

Why remaining in the same family matters

Importance sampling requires us to generate observations from q .

- A valid expression for a density does not supply a sampling method.
- Tilting a familiar density can produce an unfamiliar proposal.
- If tilting preserves a tractable family, its sampler may be reused.

We now prove when exponential tilting preserves the family.

A family and one fixed member

The family is

$$p(x; \vartheta) = h(x)e^{\vartheta^\top x - A(\vartheta)}, \quad \vartheta \in \mathcal{D}.$$

Fix $\theta_0 \in \mathcal{D}$ to obtain the original density:

$$p(x) = h(x)e^{\theta_0^\top x - A(\theta_0)}.$$

- An exponential distribution at a specified rate is one family member.
- A normal family with fixed variance and varying mean is a natural exponential family in x .
- With both mean and variance varying, the usual normal exponential family uses $(x, x^2)^\top$.

Keep the three parameters distinct

Symbol	Role
θ_0	Fixed natural parameter of the original density
θ	Tilting parameter
$\theta' = \theta_0 + \theta$	Natural parameter after tilting

We require

$$\theta_0 + \theta \in \mathcal{D} \iff C(\theta) < \infty.$$

$C(\theta)$ is defined relative to p ; its dependence on fixed θ_0 is suppressed.

The normalisation identity to keep in mind

For every admissible $\vartheta \in \mathcal{D}$,

$$\int h(x) \exp(\vartheta^\top x - A(\vartheta)) \, dx = 1.$$

The term $A(\vartheta)$ does not depend on x , so

$$e^{-A(\vartheta)} \int h(x) e^{\vartheta^\top x} \, dx = 1.$$

Therefore,

Family normalisation identity

$$\int h(x) e^{\vartheta^\top x} \, dx = e^{A(\vartheta)}.$$

Result 2: exponential tilting shifts the parameter

Suppose

$$p(x) = h(x)e^{\theta_0^\top x - A(\theta_0)}, \quad \theta_0 \in \mathcal{D}.$$

For a tilt with $\theta_0 + \theta \in \mathcal{D}$, we will show that

Closure under admissible exponential tilting

$$q(x; \theta) = h(x) \exp\left((\theta_0 + \theta)^\top x - A(\theta_0 + \theta)\right).$$

The tilted density has the same h and A , with a shifted natural parameter.

Proof: substitute and combine the linear terms

Substitute the expression for p into the tilted density:

$$q(x; \theta) = \frac{h(x)e^{\theta_0^\top x - A(\theta_0)}e^{\theta^\top x}}{C(\theta)}.$$

Combine the terms linear in x :

$$q(x; \theta) = \frac{h(x)e^{(\theta_0 + \theta)^\top x}}{e^{A(\theta_0)}C(\theta)}.$$

Write $C(\theta) = e^{\log C(\theta)}$:

$$q(x; \theta) = h(x) \exp\left((\theta_0 + \theta)^\top x - A(\theta_0) - \log C(\theta)\right).$$

Proof: what still needs to be shown?

Set $\theta' = \theta_0 + \theta$. Then

$$q(x; \theta) = h(x) \exp\left(\theta'^T x - [A(\theta_0) + \log C(\theta)]\right).$$

The original family at parameter θ' has density

$$p(x; \theta') = h(x) \exp\left(\theta'^T x - A(\theta')\right).$$

The remaining step

$$A(\theta_0) + \log C(\theta) \stackrel{?}{=} A(\theta').$$

Proof: name the candidate log-normaliser

Define

$$B(\theta, \theta_0) = A(\theta_0) + \log C(\theta).$$

The tilted density is now

$$q(x; \theta) = h(x)e^{\theta'^T x - B(\theta, \theta_0)}.$$

- $B(\theta, \theta_0)$ is constant with respect to x .
- Our goal is to show $B(\theta, \theta_0) = A(\theta')$.

Normalisation will identify this constant.

Proof: first use of normalisation

The tilted density q integrates to one:

$$1 = \int h(x) e^{\theta'^{\top} x - B(\theta, \theta_0)} dx.$$

Take the factor independent of x outside the integral:

$$1 = e^{-B(\theta, \theta_0)} \int h(x) e^{\theta'^{\top} x} dx.$$

Hence

Normalisation of the tilted density

$$e^{B(\theta, \theta_0)} = \int h(x) e^{\theta'^{\top} x} dx.$$

Proof: second use of normalisation

Since $\theta' \in \mathcal{D}$, the original family also has a density at θ' :

$$\int h(x) e^{\theta'^T x - A(\theta')} dx = 1.$$

Apply the family normalisation identity:

Normalisation of the family member at θ'

$$e^{A(\theta')} = \int h(x) e^{\theta'^T x} dx.$$

Both normalisation statements contain exactly the same integral.

Proof: identify the normaliser and conclude

Comparing the two expressions for the same integral gives

$$e^{B(\theta, \theta_0)} = e^{A(\theta')}.$$

Take logarithms:

$$B(\theta, \theta_0) = A(\theta').$$

Substitute this into q :

The tilted member of the same family

$$q(x; \theta) = h(x) \exp\left(\theta'^T x - A(\theta')\right), \quad \theta' = \theta_0 + \theta.$$

The proof also identifies the tilting normaliser

We have established

$$A(\theta_0) + \log C(\theta) = A(\theta_0 + \theta).$$

Therefore,

Normaliser in terms of the original family

$$\begin{aligned}\log C(\theta) &= A(\theta_0 + \theta) - A(\theta_0), \\ C(\theta) &= \exp(A(\theta_0 + \theta) - A(\theta_0)).\end{aligned}$$

These identities hold whenever $\theta_0 + \theta \in \mathcal{D}$.

What this means for importance sampling

Suppose a sampler is available for this family at the required parameter values.

1. Express the original density using natural parameter θ_0 .
2. Choose a tilt θ satisfying $\theta_0 + \theta \in \mathcal{D}$.
3. Reuse the family-specific sampler at parameter $\theta_0 + \theta$.

The proposal remains in the same distribution family.

What the result does and does not guarantee

The theorem guarantees preservation of the family under admissible tilting.

For a useful importance sampling proposal, we still need:

- A tractable sampler at the shifted parameter; natural exponential family membership alone does not guarantee this.
- A shift inside the admissible domain $\mathcal{D}$.
- A suitable choice of θ , for example through the variance minimisation discussed previously.

Take-home points

1. Tilting any fixed density gives a natural exponential family:

$$q(x; \theta) = p(x)e^{\theta^\top x - \log C(\theta)}.$$

2. Tilting a natural exponential family member preserves its family:

$$\theta_0 \longmapsto \theta_0 + \theta, \quad \theta_0 + \theta \in \mathcal{D}.$$

3. The full proof uses normalisation **twice** to show

$$A(\theta_0) + \log C(\theta) = A(\theta_0 + \theta).$$

A usable family sampler makes this closure property computationally valuable.