Matching symmetries with variational inference

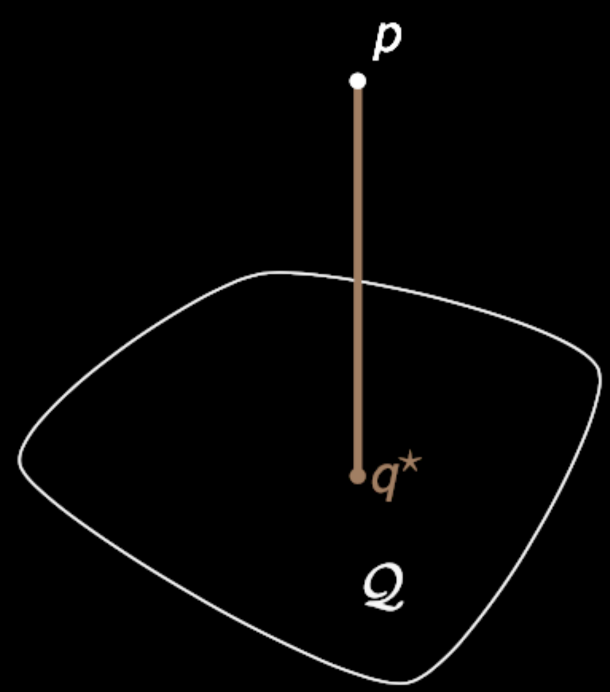
Charles Margossian
(Flatiron Institute)

Matching symmetries with variational inference

Variational Inference in location-scale families.
CM and Lawrence Saul
AISTATS 2025 (best paper award)

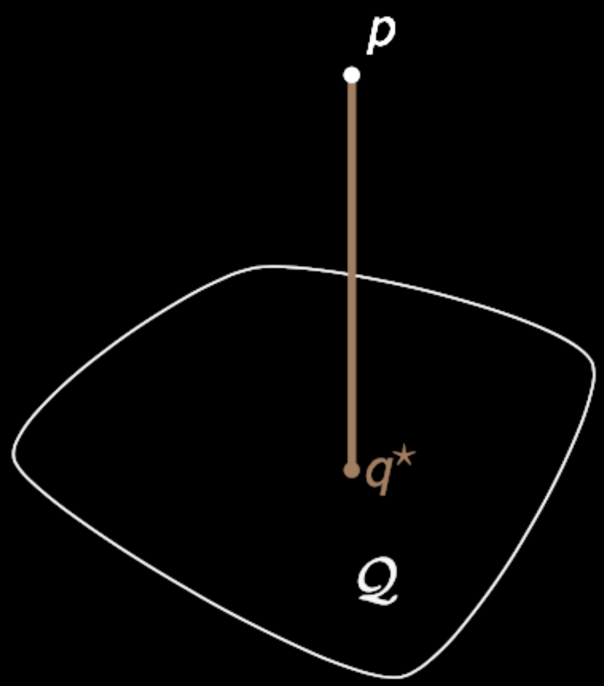
Charles Margossian
(Flatiron Institute)

Variational inference



$$q^\star = \textbf{argmin}_{q \in \mathcal{Q}}\ \textbf{KL}(q\,||\,p)$$

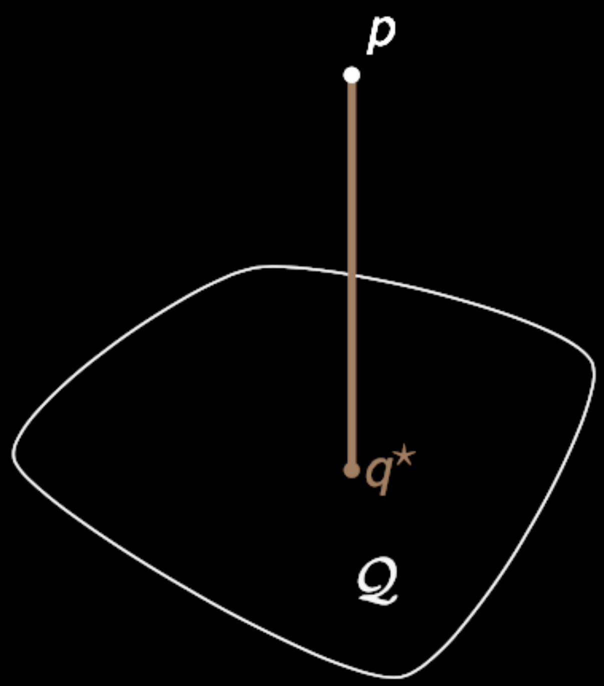
Variational inference



$$q^\star = \textbf{argmin}_{q \in \mathcal{Q}} \textbf{KL}(q \,||\, p)$$

$$\textbf{KL}(q \,||\, p) = \int [\log q(z) - \log p(z)] q(z) \textbf{d}z \,.$$

Variational inference

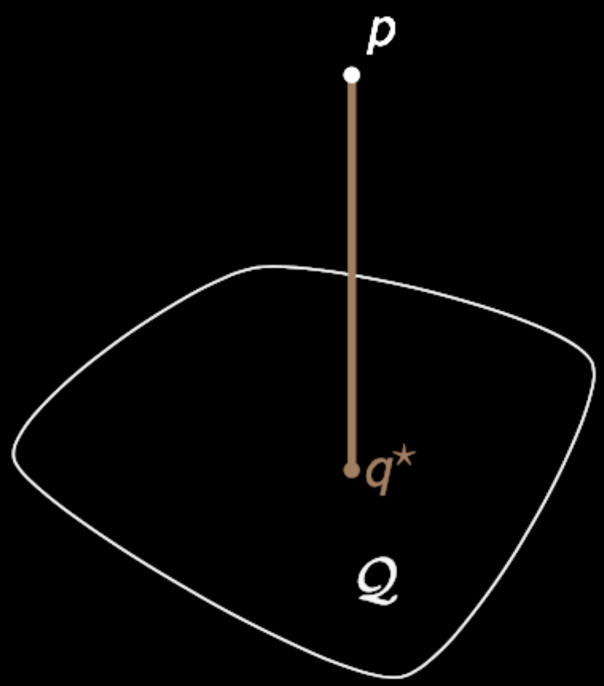


$$q^\star = \textbf{argmin}_{q \in \mathcal{Q}} \, \textbf{KL}(q \,||\, p)$$

$$\textbf{KL}(q \,||\, p) = \int [\log q(z) - \log p(z)] q(z) \mathbf{d}z \,.$$

In practice $p \notin \mathcal{Q}$, meaning $q^\star \neq p$

Variational inference



$$q^\star = \textbf{argmin}_{q \in \mathcal{Q}} \, \textbf{KL}(q \,||\, p)$$

$$\textbf{KL}(q \,||\, p) = \int [\log q(z) - \log p(z)] q(z) \mathbf{d}z \, .$$

In practice $p \notin \mathcal{Q}$, meaning $q^\star \neq p$ …so what?

❓ Can a misspecified q learn key properties* of p?

❓ Can a misspecified q learn key properties* of p?

*mean, median, correlations, covariances, quantiles, "credible" samples…

❓ Can a misspecified q learn key properties* of p?

*mean, median, correlations, covariances, quantiles, "credible" samples…

💡 Guarantees for VI when p and $\mathcal{Q}$ exhibit certain symmetries

❓ Can a misspecified q learn key properties* of p?

*mean, median, correlations, covariances, quantiles, "credible" samples…

💡 Guarantees for VI when p and $\mathcal{Q}$ exhibit certain symmetries

- *even symmetry* $\rightarrow$ q recovers the mean

❓ Can a misspecified q learn key properties* of p?

*mean, median, correlations, covariances, quantiles, "credible" samples…

💡 Guarantees for VI when p and $\mathcal{Q}$ exhibit certain symmetries

- *even symmetry* $\rightarrow$ q recovers the mean
- *elliptical symmetry* $\rightarrow$ q recovers the correlation matrix

❓ Can a misspecified q learn key properties* of p?

*mean, median, correlations, covariances, quantiles, "credible" samples…

💡 Guarantees for VI when p and $\mathcal{Q}$ exhibit certain symmetries

- *even symmetry* $\longrightarrow$ q recovers the mean
- *elliptical symmetry* $\longrightarrow$ q recovers the correlation matrix

✅ theory allows for severe misspecifications in $\mathcal{Q}$

❓ Can a misspecified q learn key properties* of p?

*mean, median, correlations, covariances, quantiles, "credible" samples…

💡 Guarantees for VI when p and $\mathcal{Q}$ exhibit certain symmetries

- *even symmetry* $\longrightarrow$ q recovers the mean
- *elliptical symmetry* $\longrightarrow$ q recovers the correlation matrix

✅ theory allows for severe misspecifications in $\mathcal{Q}$

✅ q has even and elliptical symmetry if it is location-scale

❓ Can a misspecified q learn key properties* of p?

*mean, median, correlations, covariances, quantiles, "credible" samples…

💡 Guarantees for VI when p and $\mathcal{Q}$ exhibit certain symmetries

- *even symmetry* $\longrightarrow$ q recovers the mean
- *elliptical symmetry* $\longrightarrow$ q recovers the correlation matrix

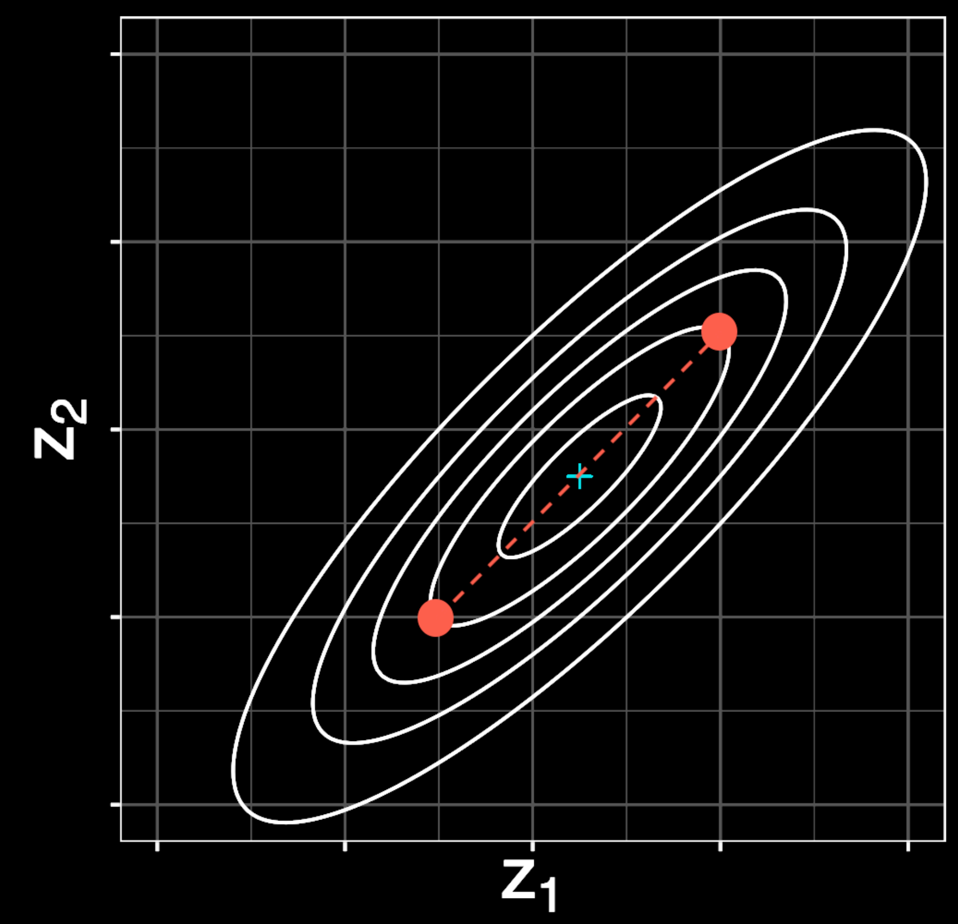
✅ theory allows for severe misspecifications in $\mathcal{Q}$

✅ q has even and elliptical symmetry if it is location-scale

🤔 When does p have the studied symmetries?

Definition
$p(z)$ is *even symmetric* about a point μ if

$$p(\mu - z) = p(\mu + z)\,.$$

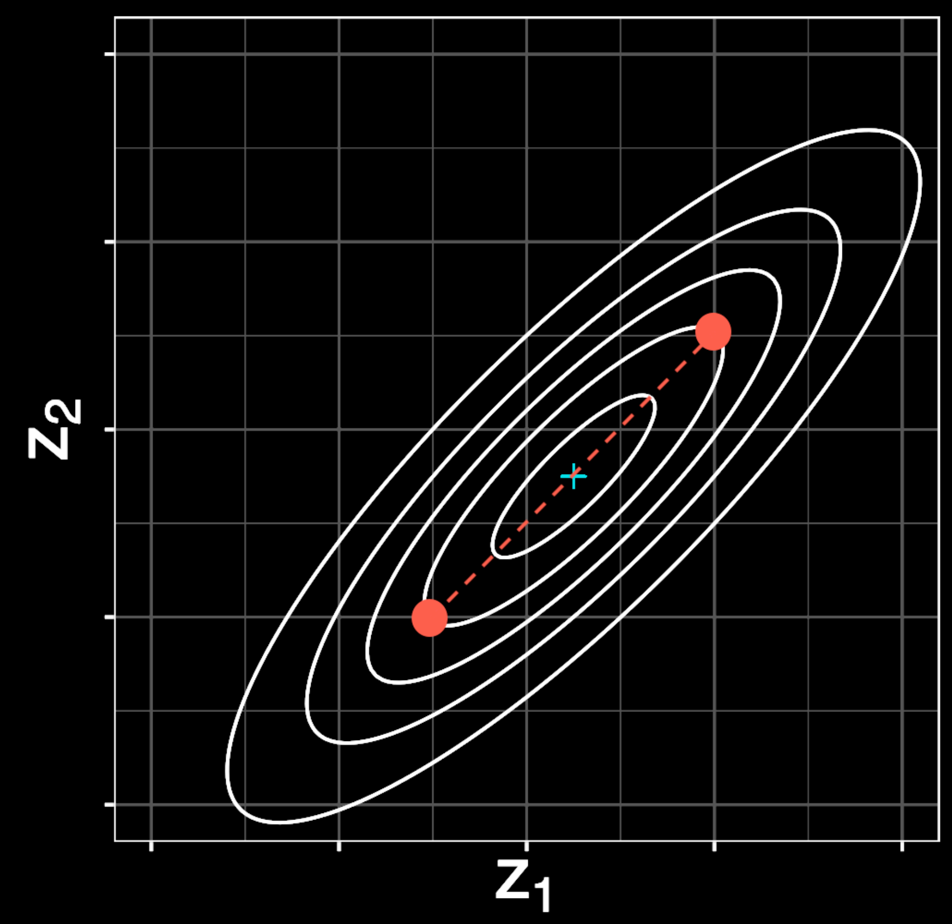


Definition

$p(z)$ is *even symmetric* about a point μ if

$$p(\mu - z) = p(\mu + z)\,.$$

Theorem



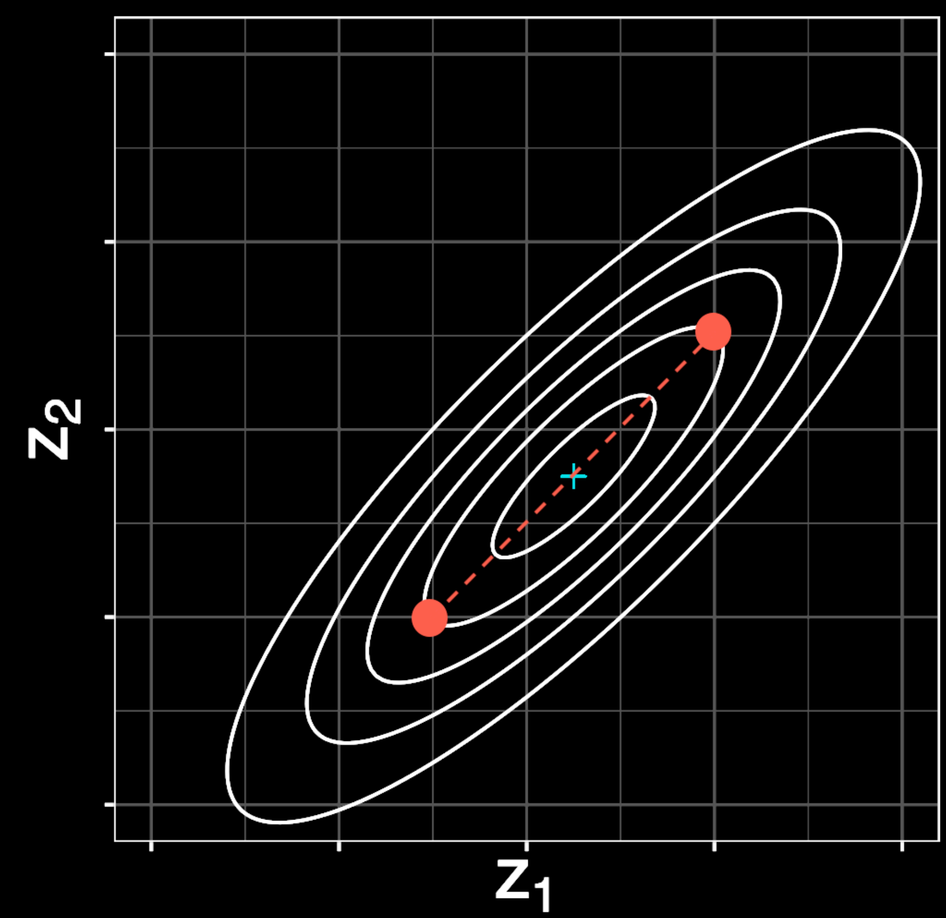
Definition

$p(z)$ is *even symmetric* about a point μ if

$$p(\mu - z) = p(\mu + z)\,.$$

Theorem

- Let $p(z)$ be even symmetric.



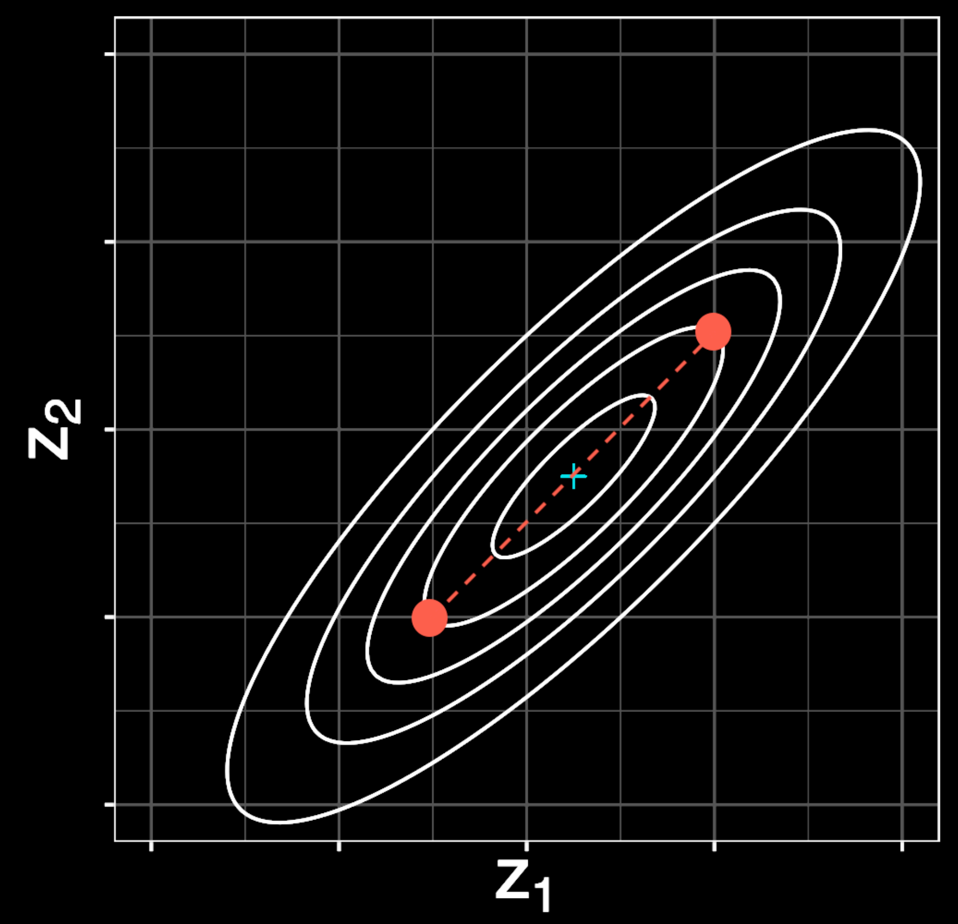
Definition

$p(z)$ is *even symmetric* about a point μ if

$$p(\mu - z) = p(\mu + z)\,.$$

Theorem

- Let $p(z)$ be even symmetric.
- Let $\mathcal{Q}$ be a family with location parameter and an even-symmetric based distribution q_0.



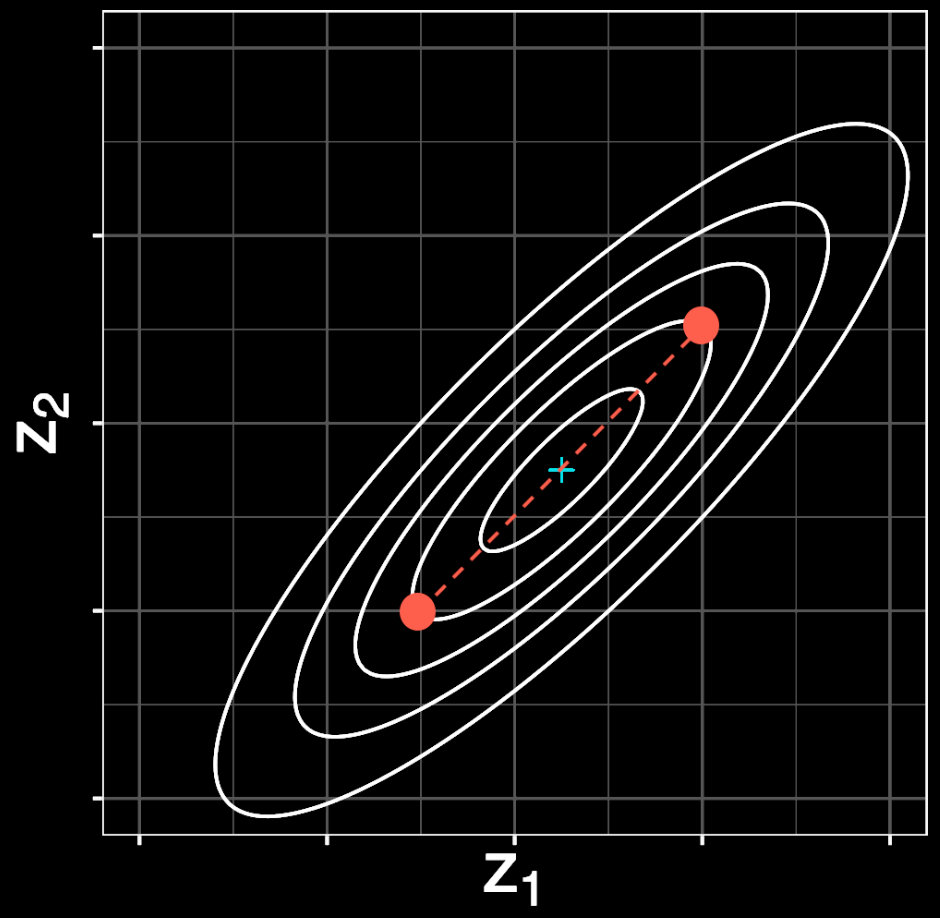
Definition

$p(z)$ is *even symmetric* about a point μ if

$$p(\mu - z) = p(\mu + z)\,.$$

Theorem

- Let $p(z)$ be even symmetric.
- Let $\mathcal{Q}$ be a family with location parameter and an even-symmetric based distribution q_0.
- Then, *under some regularity conditions*, minimizing $\mathbf{KL}(q||p)$, yields $\nu^\star = \mu$.



Definition

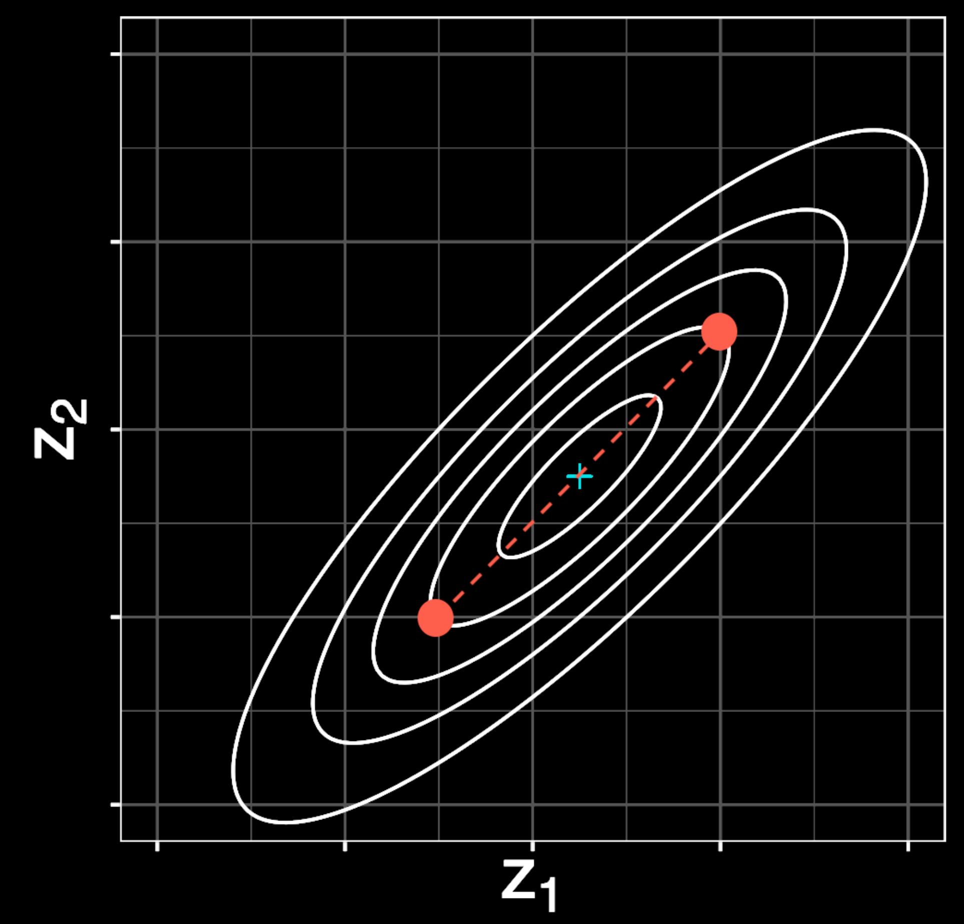
$p(z)$ is *even symmetric* about a point μ if

$$p(\mu - z) = p(\mu + z)\,.$$

Theorem

- Let $p(z)$ be even symmetric.
- Let $\mathcal{Q}$ be a family with location parameter and an even-symmetric based distribution q_0.
- Then, *under some regularity conditions*, minimizing $\mathbf{KL}(q||p)$, yields $\nu^\star = \mu$.

*If the mean exists, it gets recovered.



Definition

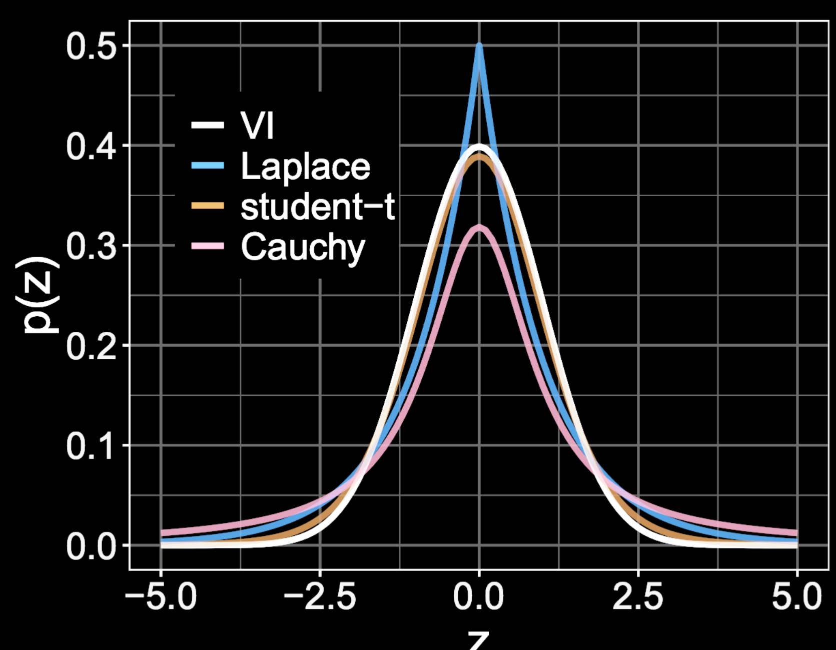
$p(z)$ is *even symmetric* about a point μ if

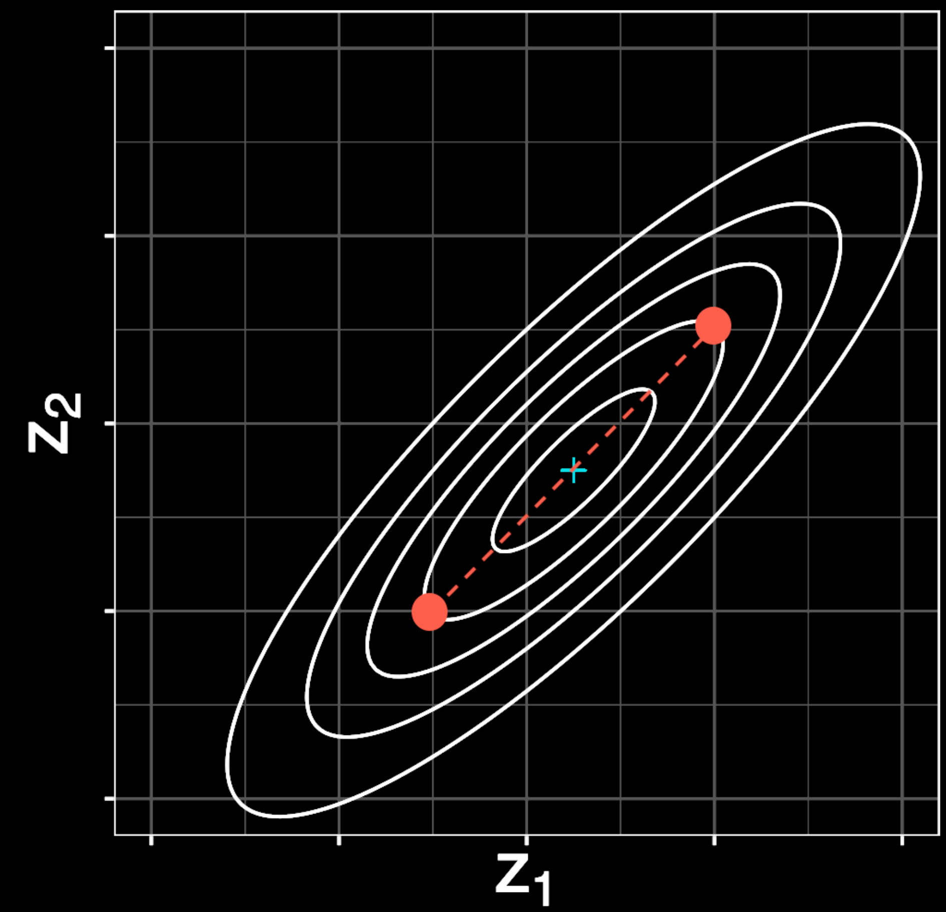
$$p(\mu - z) = p(\mu + z)\,.$$

Theorem

- Let $p(z)$ be even symmetric.
- Let $\mathcal{Q}$ be a family with location parameter and an even-symmetric based distribution q_0.
- Then, *under some regularity conditions*, minimizing $\mathbf{KL}(q\,||\,p)$, yields $\nu^\star = \mu$.

*If the mean exists, it gets recovered.





Definition

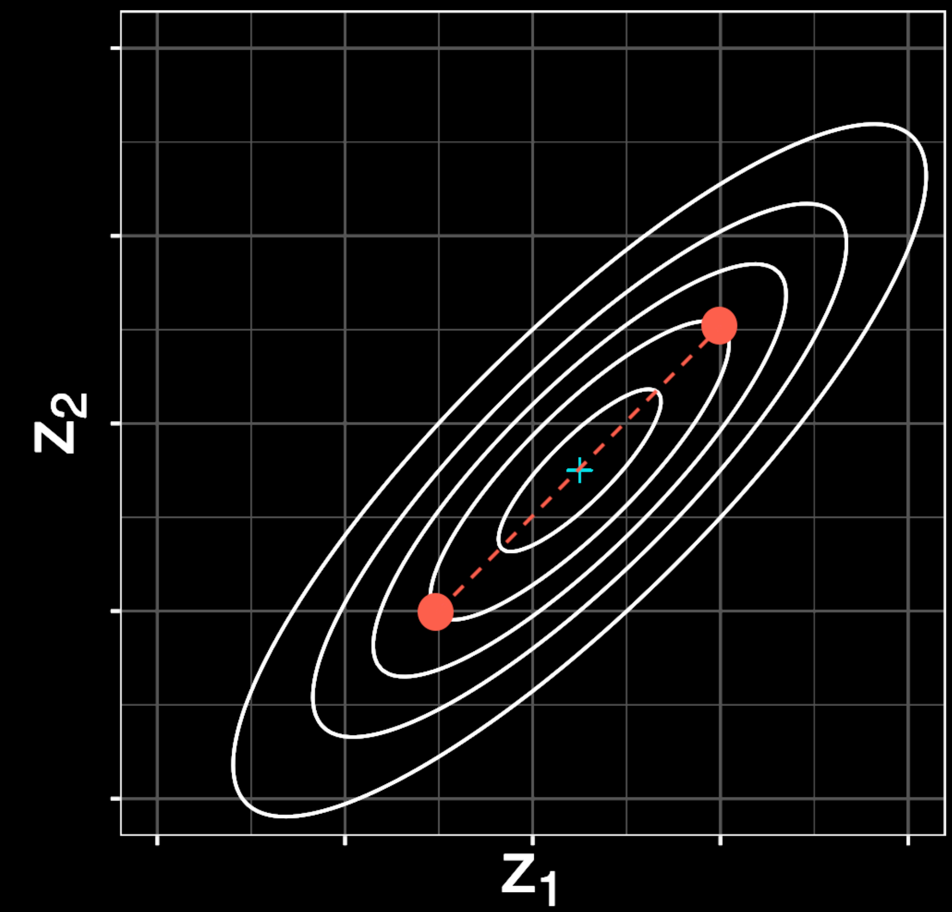
$p(z)$ is *even symmetric* about a point μ if

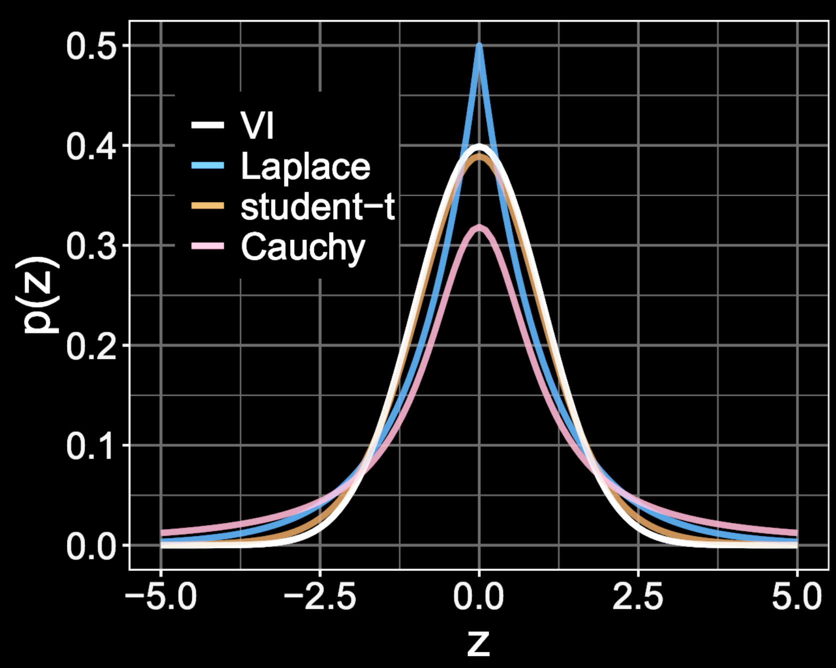
$$p(\mu - z) = p(\mu + z)\,.$$

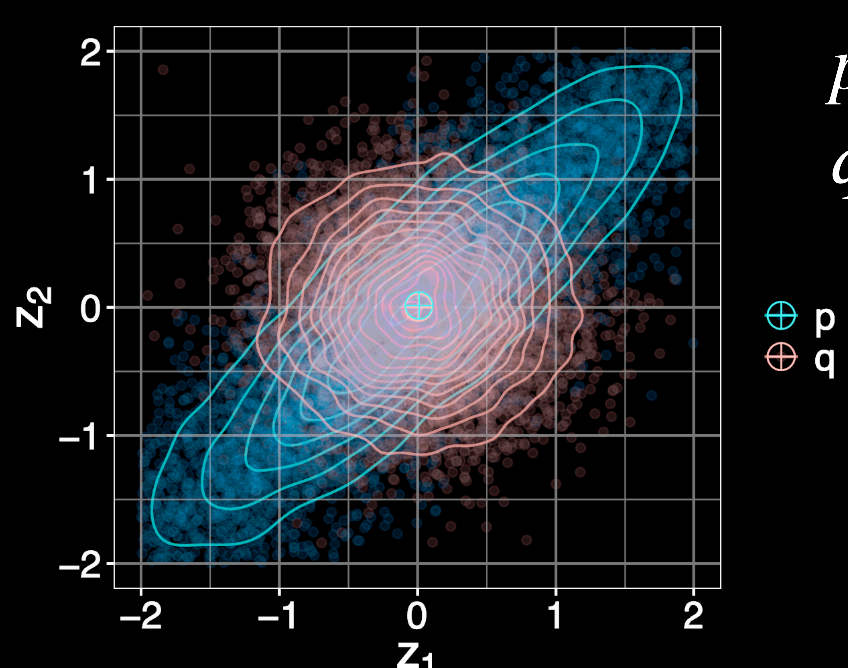
Theorem

- Let $p(z)$ be even symmetric.
- Let $\mathcal{Q}$ be a family with location parameter and an even-symmetric based distribution q_0.
- Then, *under some regularity conditions*, minimizing $\textbf{KL}(q\,||\,p)$, yields $\nu^\star = \mu$.

*If the mean exists, it gets recovered.







p = multivariate Student-t

q = factorized Gaussian

"under some regularity conditions…"

"under some regularity conditions…"

Condition
$\log p$ is concave on $\mathbb{R}^d$ and strictly concave on some open set of $\mathbb{R}^d$.

"under some regularity conditions…"

Condition
$\log p$ is concave on $\mathbb{R}^d$ and strictly concave on some open set of $\mathbb{R}^d$.

$\Rightarrow$ $\mathbf{KL}(q_\nu || p)$ is strictly convex in ν.

"under some regularity conditions…"

Condition
$\log p$ is concave on $\mathbb{R}^d$ and strictly concave on some open set of $\mathbb{R}^d$.

$\Rightarrow$ $\mathbf{KL}(q_\nu || p)$ is strictly convex in ν.

Without condition, $\nu^\star = \mu$ is still a stationary point of $\mathrm{KL}(q||p)$.

"under some regularity conditions…"

Condition

$\log p$ is concave on $\mathbb{R}^d$ and strictly concave on some open set of $\mathbb{R}^d$.

$\Rightarrow$ $\mathbf{KL}(q_\nu || p)$ is strictly convex in ν.

Without condition, $\nu^\star = \mu$ is still a stationary point of $\mathrm{KL}(q || p)$.

Example:

$$p(z) = \frac{1}{2}\mathrm{Normal}(-m, 1) + \frac{1}{2}\mathrm{Normal}(m, 1)$$
$$q(z) = \mathrm{Normal}(\nu, 1)$$

"under some regularity conditions…"

Condition
$\log p$ is concave on $\mathbb{R}^d$ and strictly concave on some open set of $\mathbb{R}^d$.

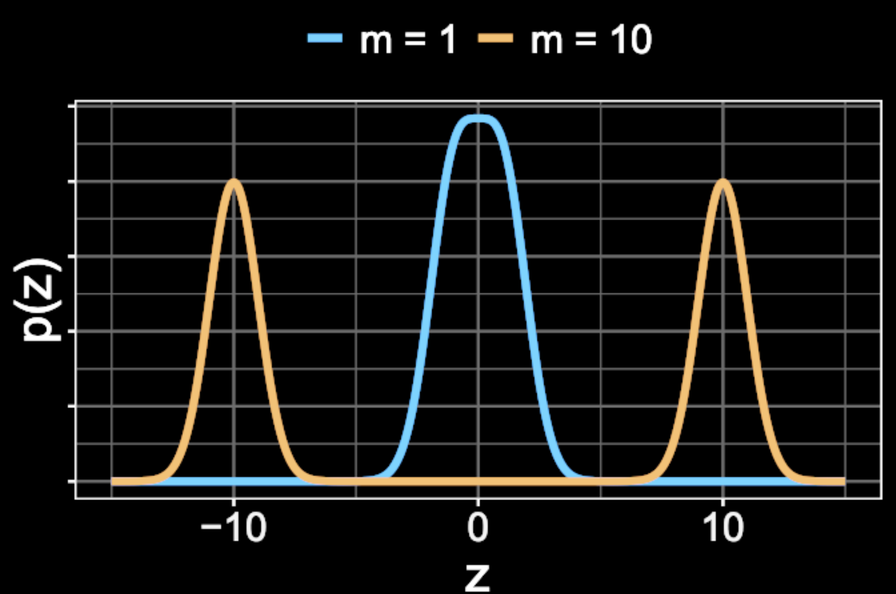
$\Rightarrow$ $\mathbf{KL}(q_\nu || p)$ is strictly convex in ν.

Without condition, $\nu^\star = \mu$ is still a stationary point of $\mathrm{KL}(q||p)$.

Example:

$$p(z) = \frac{1}{2}\mathrm{Normal}(-m, 1) + \frac{1}{2}\mathrm{Normal}(m, 1)$$

$$q(z) = \mathrm{Normal}(\nu, 1)$$



"under some regularity conditions…"

Condition
$\log p$ is concave on $\mathbb{R}^d$ and strictly concave on some open set of $\mathbb{R}^d$.

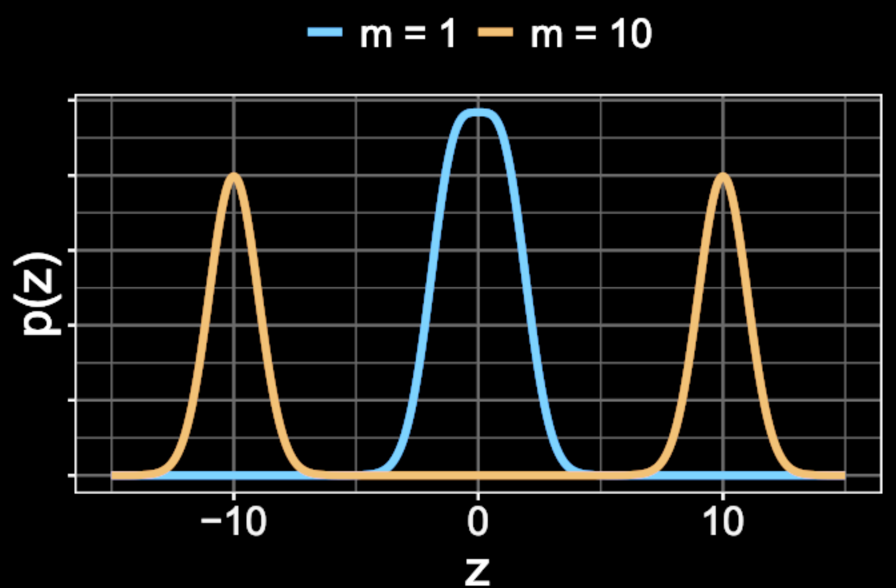
$\Rightarrow$ $\mathbf{KL}(q_\nu || p)$ is strictly convex in ν.

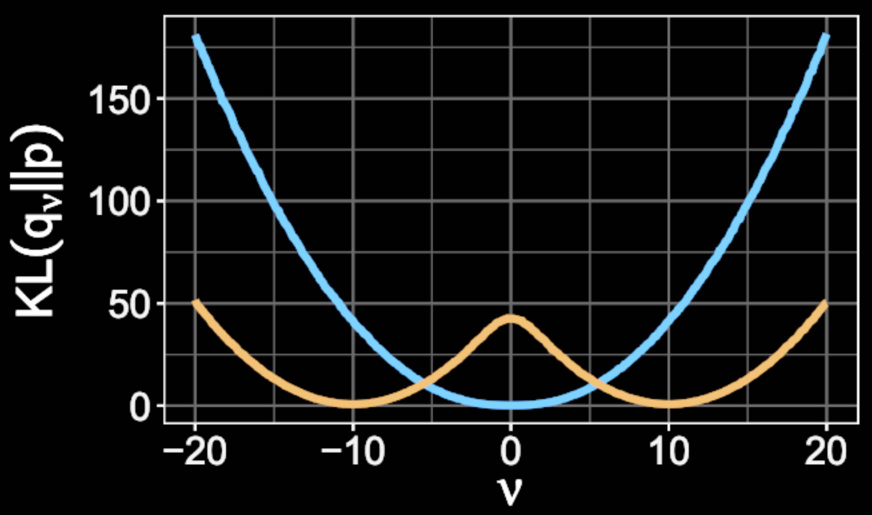
Without condition, $\nu^\star = \mu$ is still a stationary point of $\mathrm{KL}(q || p)$.

Example:

$$p(z) = \frac{1}{2}\mathrm{Normal}(-m, 1) + \frac{1}{2}\mathrm{Normal}(m, 1)$$

$$q(z) = \mathrm{Normal}(\nu, 1)$$



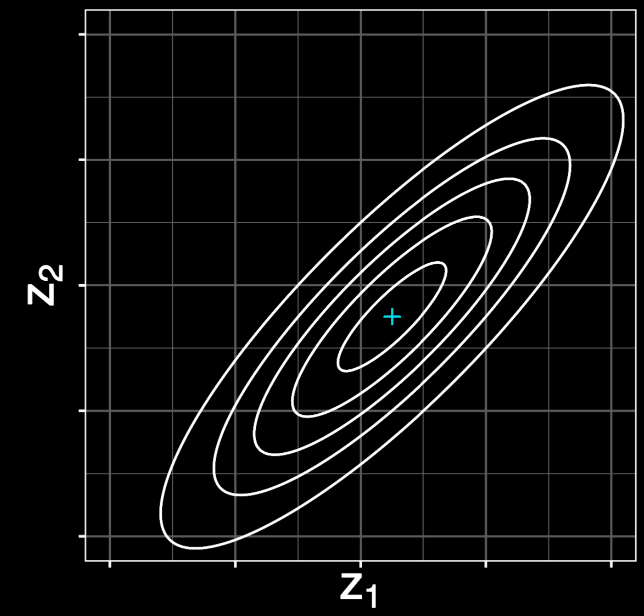


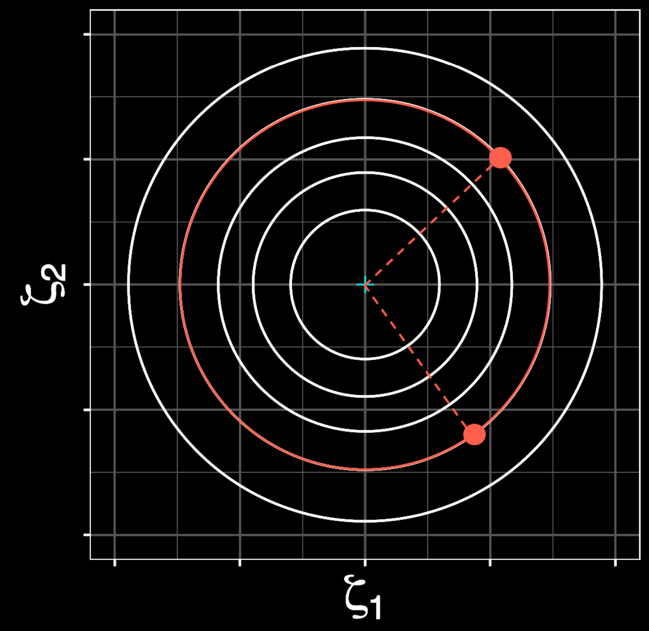
Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\ M$ s.t

$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

and $p(\zeta)$ is spherically symmetric.





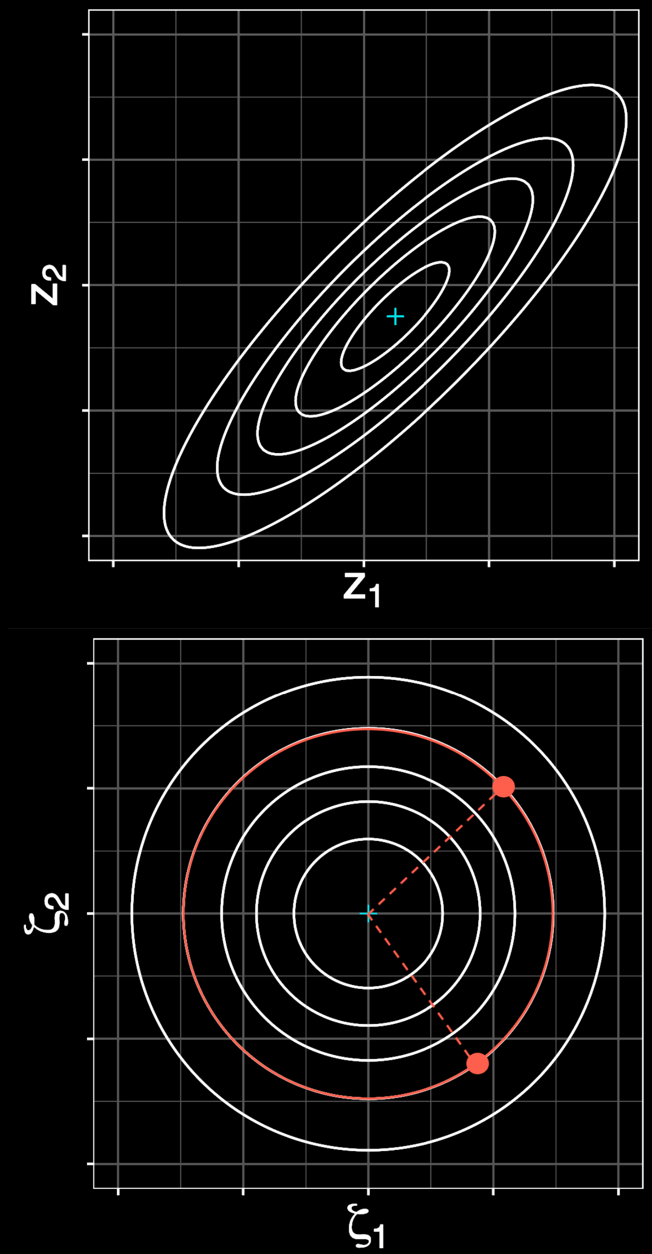
Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\ M$ s.t

$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

and $p(\zeta)$ is spherically symmetric.

*If M exists, a whole of family of conditioners exists.

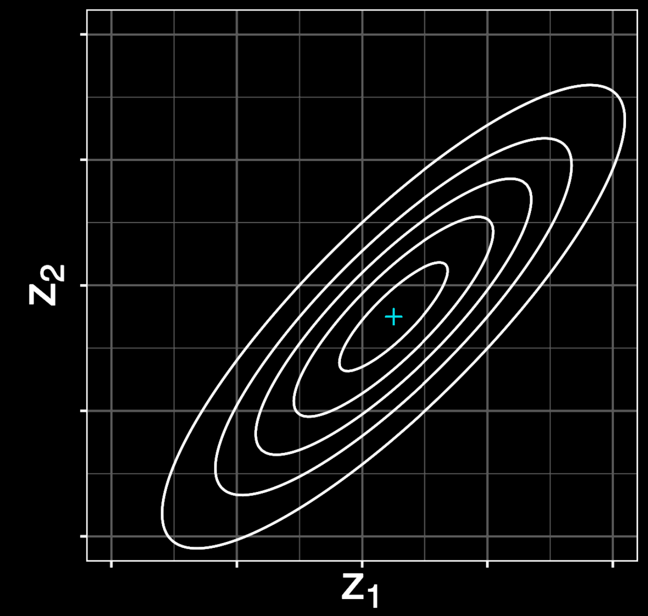


Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\ M$ s.t

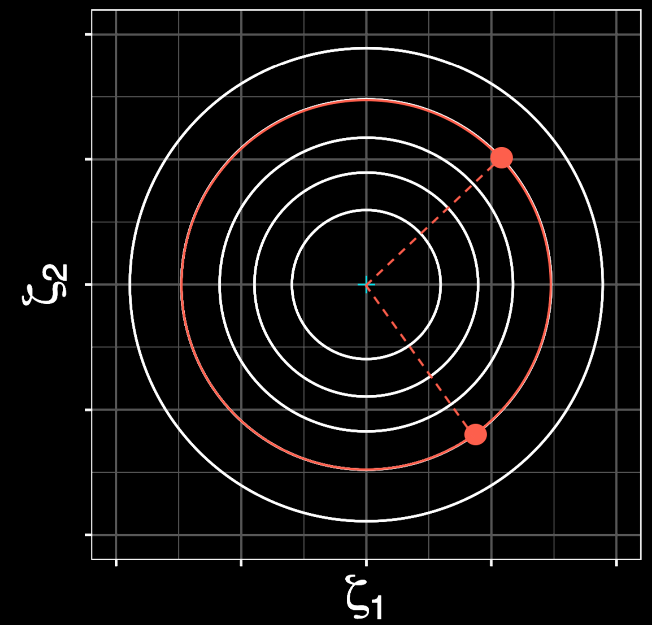
$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

and $p(\zeta)$ is spherically symmetric.



*If M exists, a whole of family of conditioners exists.

Theorem



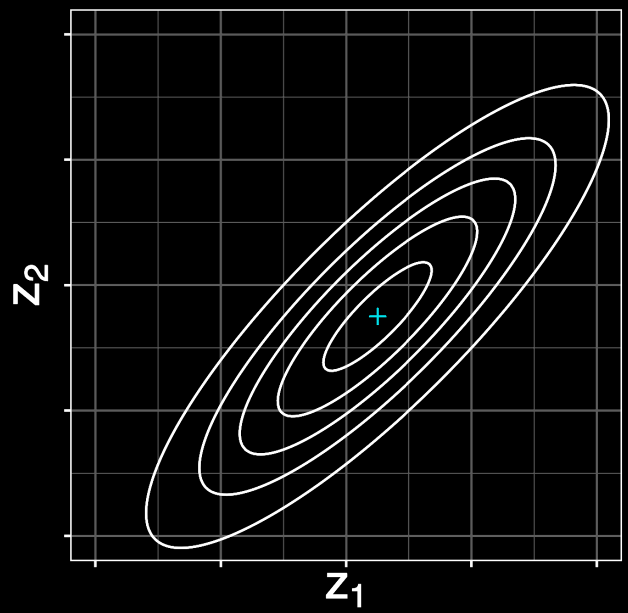
Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\, M$ s.t

$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

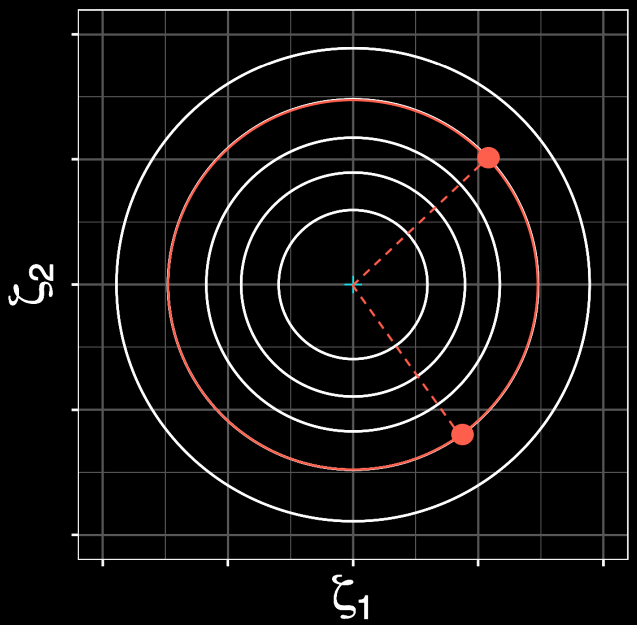
and $p(\zeta)$ is spherically symmetric.

*If M exists, a whole of family of conditioners exists.



Theorem

- Let $p(z)$ be elliptically symmetric.

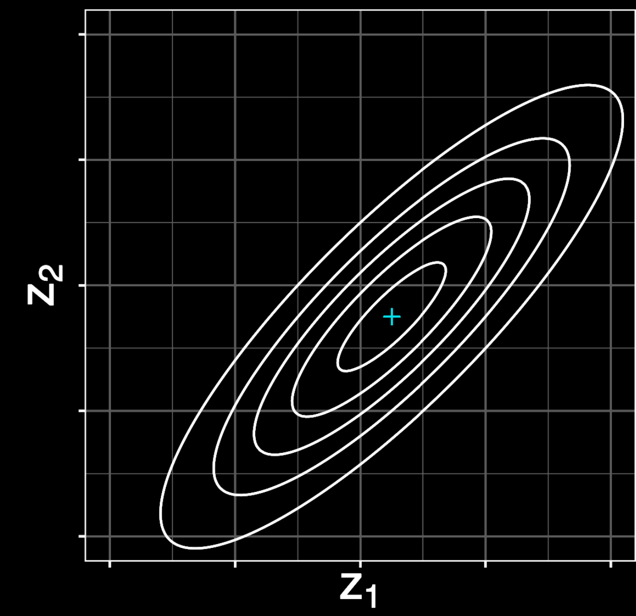


Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\ M$ s.t

$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

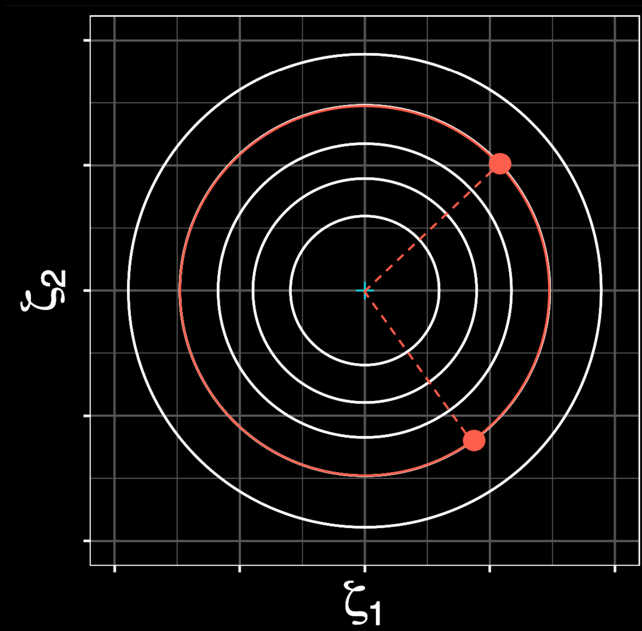
and $p(\zeta)$ is spherically symmetric.



*If M exists, a whole of family of conditioners exists.

Theorem

- Let $p(z)$ be elliptically symmetric.
- Let $\mathcal{Q}$ be a family with location-scale parameter (ν, S), with a spherically symmetric base distribution q_0.

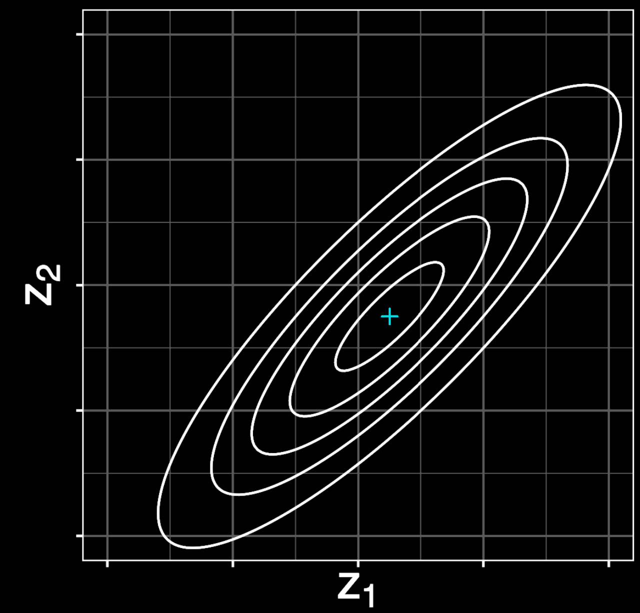


Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\, M$ s.t

$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

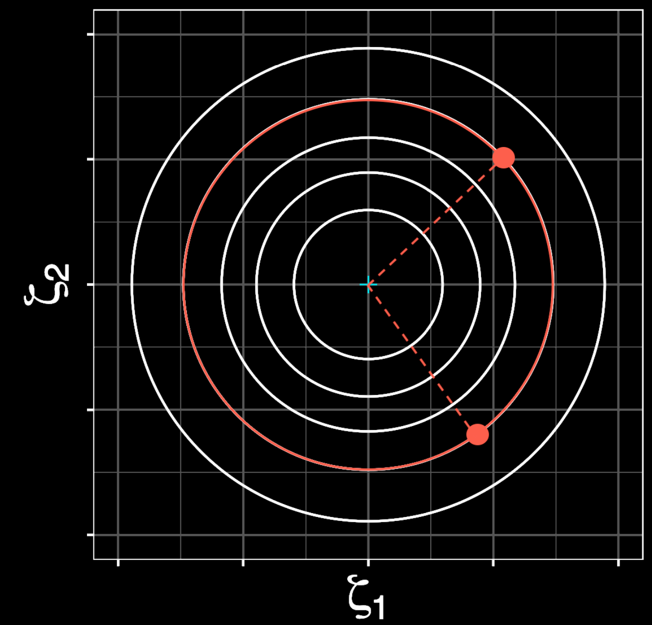
and $p(\zeta)$ is spherically symmetric.



*If M exists, a whole of family of conditioners exists.

Theorem

- Let $p(z)$ be elliptically symmetric.
- Let $\mathcal{Q}$ be a family with location-scale parameter (ν, S), with a spherically symmetric base distribution q_0.
- Then, under some regularity conditions, a minimizer of $\mathbf{KL}(q\,||\,p)$ yields $S^\star = \gamma^2 M$ for some $\gamma \in \mathbb{R}$.

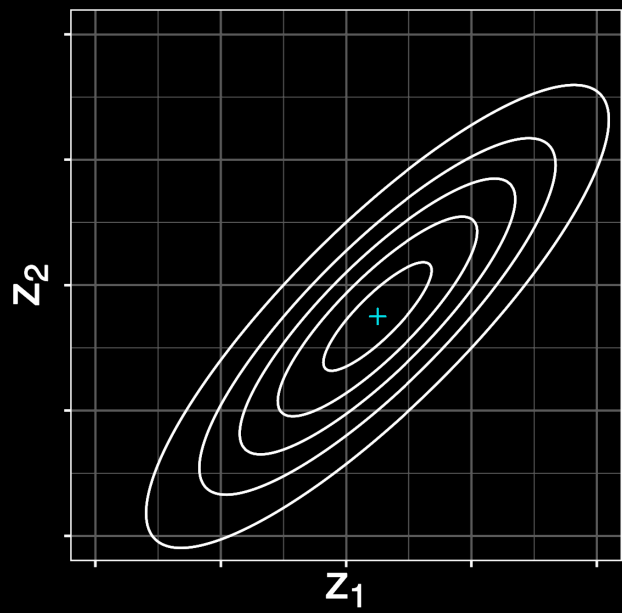


Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\ M$ s.t

$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

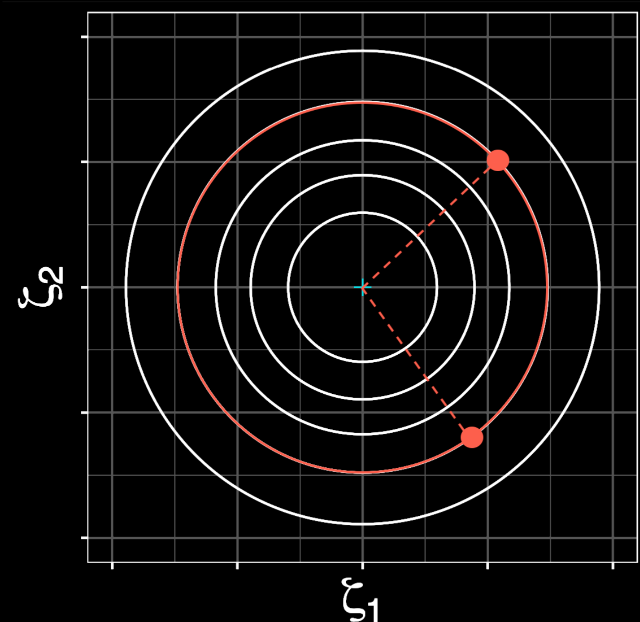
and $p(\zeta)$ is spherically symmetric.

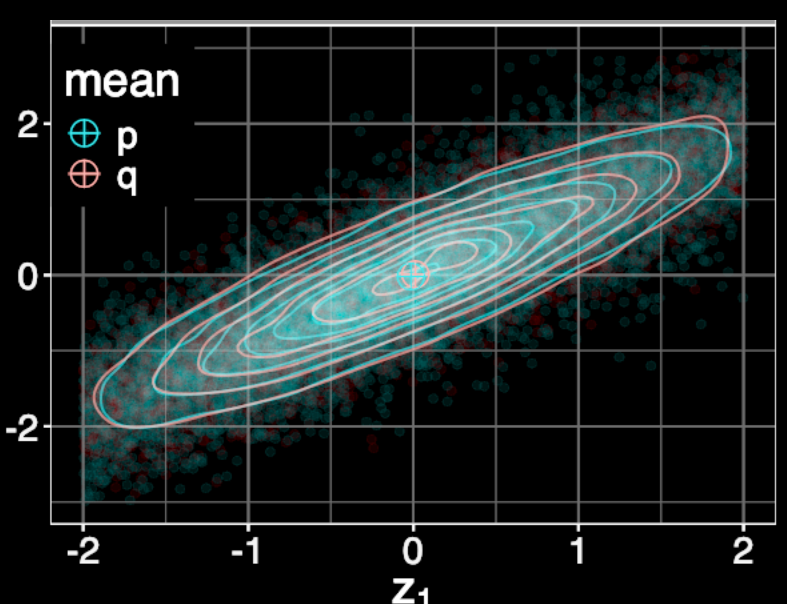


*If M exists, a whole of family of conditioners exists.

Theorem

- Let $p(z)$ be elliptically symmetric.
- Let $\mathcal{Q}$ be a family with location-scale parameter (ν, S), with a spherically symmetric base distribution q_0.
- Then, under some regularity conditions, a minimizer of $\mathbf{KL}(q||p)$ yields $S^\star = \gamma^2 M$ for some $\gamma \in \mathbb{R}$.





$p =$ multivariate Student-t

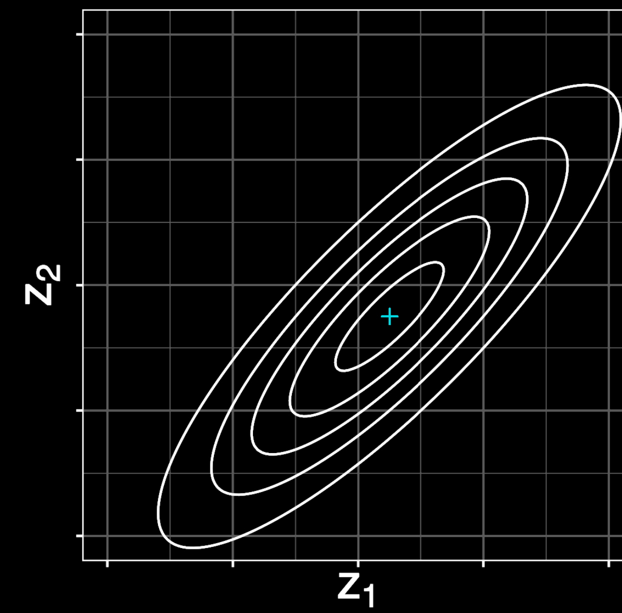
$q =$ multivariate Gaussian

Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\ M$ s.t

$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

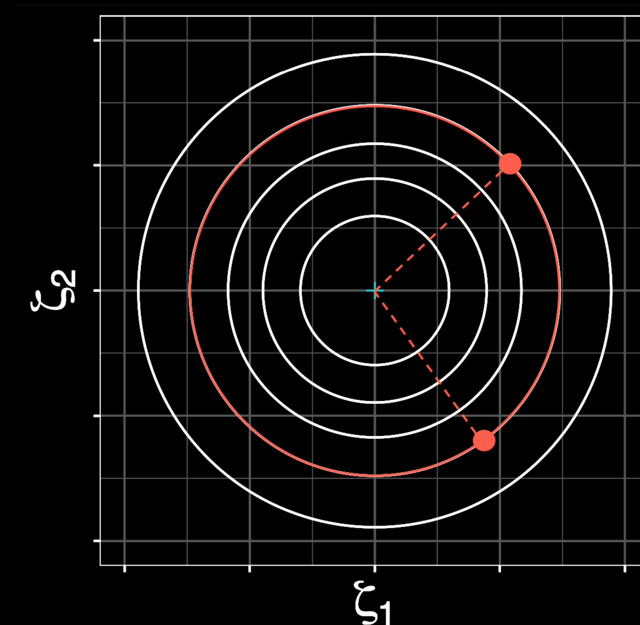
and $p(\zeta)$ is spherically symmetric.

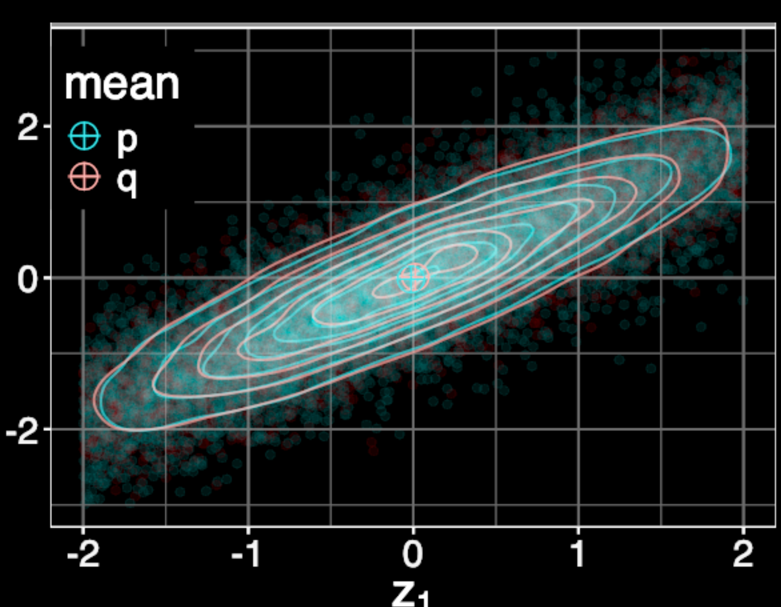


*If M exists, a whole of family of conditioners exists.

Theorem

- Let $p(z)$ be elliptically symmetric.
- Let $\mathcal{Q}$ be a family with location-scale parameter (ν, S), with a spherically symmetric base distribution q_0.
- Then, under some regularity conditions, a minimizer of $\mathbf{KL}(q\,||\,p)$ yields $S^\star = \gamma^2 M$ for some $\gamma \in \mathbb{R}$.





Which statistics do we recover?

p = multivariate Student-t

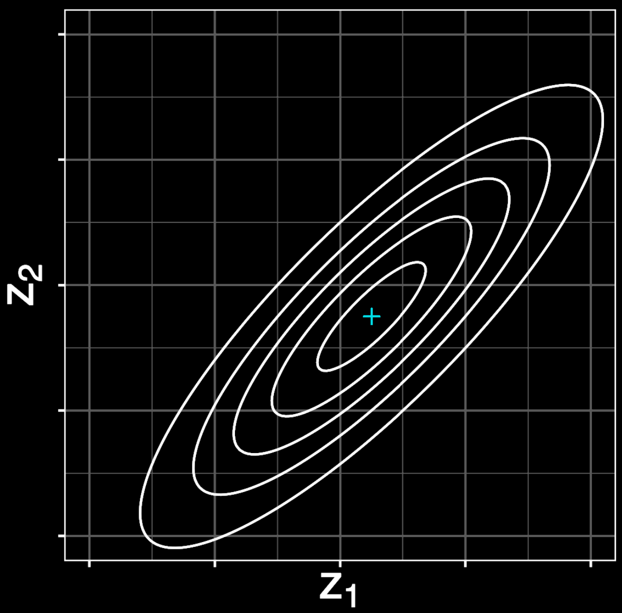
q = multivariate Gaussian

Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\, M$ s.t

$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

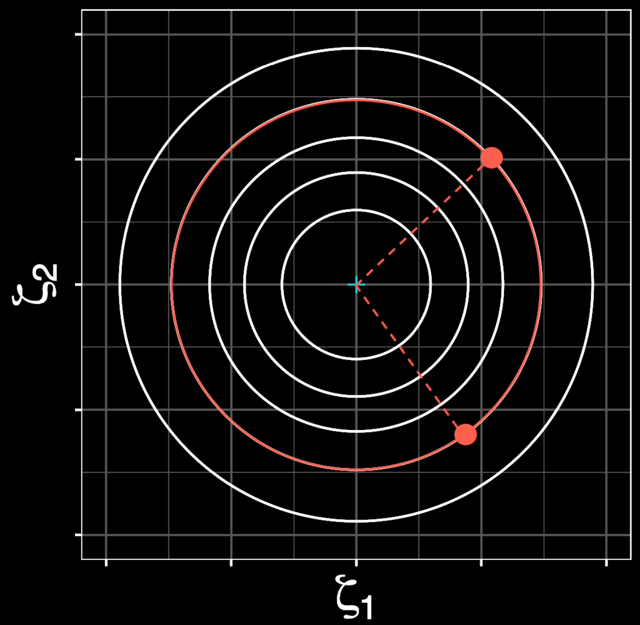
and $p(\zeta)$ is spherically symmetric.

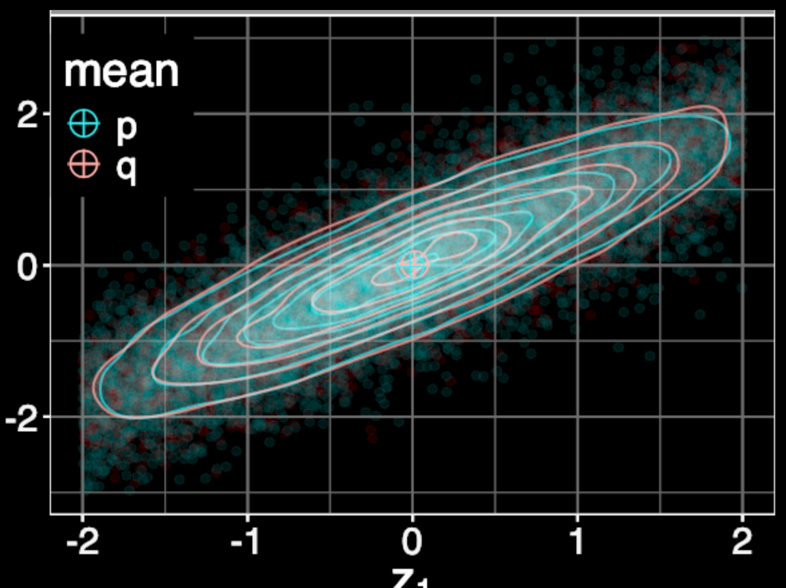


*If M exists, a whole of family of conditioners exists.

Theorem

- Let $p(z)$ be elliptically symmetric.
- Let $\mathcal{Q}$ be a family with location-scale parameter (ν, S), with a spherically symmetric base distribution q_0.
- Then, under some regularity conditions, a minimizer of $\mathbf{KL}(q||p)$ yields $S^\star = \gamma^2 M$ for some $\gamma \in \mathbb{R}$.





p = multivariate Student-t

q = multivariate Gaussian

Which statistics do we recover?

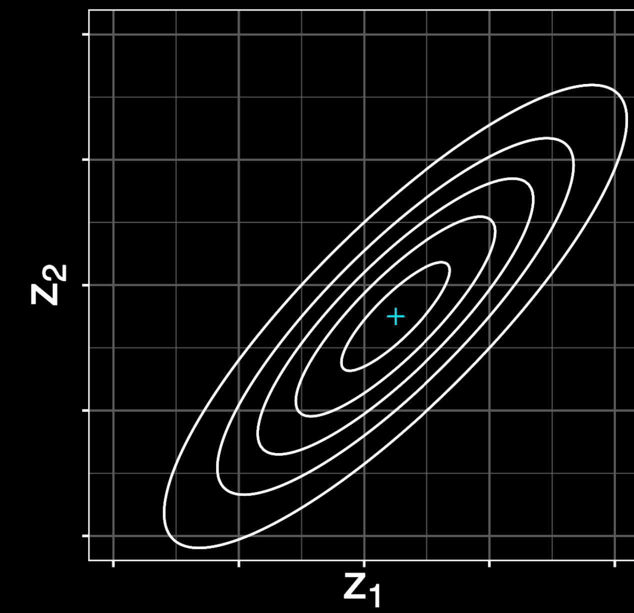
❌ covariance matrix, scale matrix

Definition

$p(z)$ is *elliptically symmetric* about a point μ if $\exists\, M$ s.t

$$\zeta = M^{-\frac{1}{2}}(z - \mu)$$

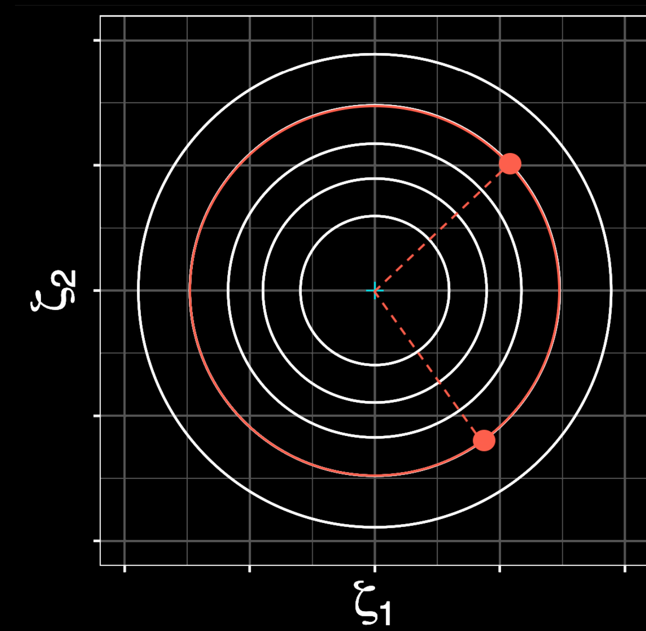
and $p(\zeta)$ is spherically symmetric.

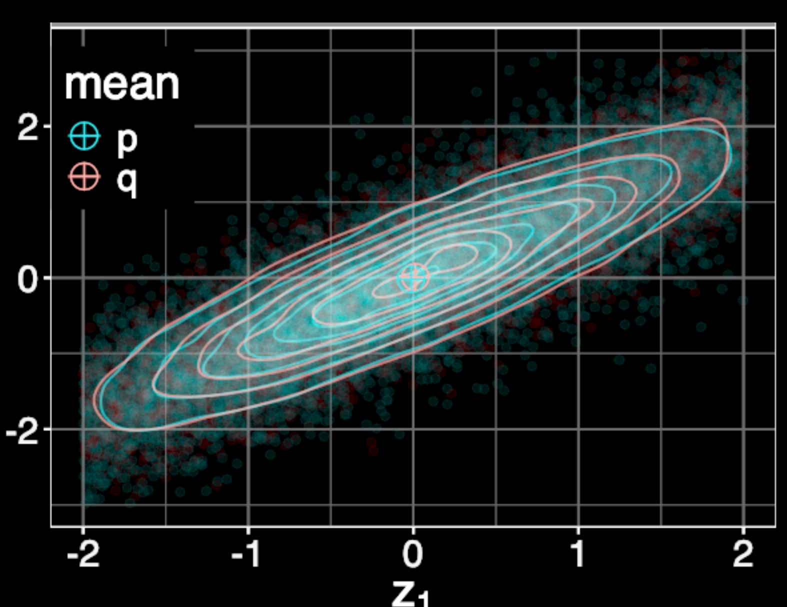


*If M exists, a whole of family of conditioners exists.

Theorem

- Let $p(z)$ be elliptically symmetric.
- Let $\mathcal{Q}$ be a family with location-scale parameter (ν, S), with a spherically symmetric base distribution q_0.
- Then, under some regularity conditions, a minimizer of $\mathbf{KL}(q\,||\,p)$ yields $S^\star = \gamma^2 M$ for some $\gamma \in \mathbb{R}$.





p = multivariate Student-t

q = multivariate Gaussian

Which statistics do we recover?

❌ covariance matrix, scale matrix

✅ correlation matrix

🤔 When does p have the studied symmetries?

🤔 When does p have the studied symmetries?

In Bayesian setting:

$$p(z) = \pi(z \mid x_{1:N}) \propto \pi(z)\ \pi(x_{1:N} \mid z)$$

🤔 When does p have the studied symmetries?

In Bayesian setting:

$$p(z) = \pi(z \mid x_{1:N}) \propto \pi(z)\ \pi(x_{1:N} \mid z)$$

Sparse data regime: prior $\pi(z)$ dominates.

🤔 When does p have the studied symmetries?

In Bayesian setting:

$$p(z) = \pi(z \mid x_{1:N}) \propto \pi(z)\ \pi(x_{1:N} \mid z)$$

Sparse data regime: prior $\pi(z)$ dominates.

Large data regime: likelihood $\pi(x_{1:N} \mid z)$ dominates.

🤔 When does p have the studied symmetries?

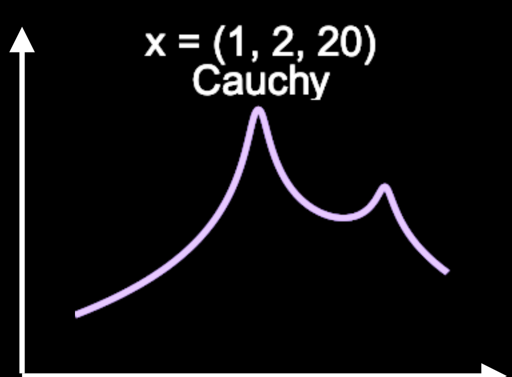
In Bayesian setting:

$$p(z) = \pi(z \mid x_{1:N}) \propto \pi(z)\ \pi(x_{1:N} \mid z)$$

Sparse data regime: prior $\pi(z)$ dominates.

Large data regime: likelihood $\pi(x_{1:N} \mid z)$ dominates.

$\log \pi(x \mid z = \nu)$



ν

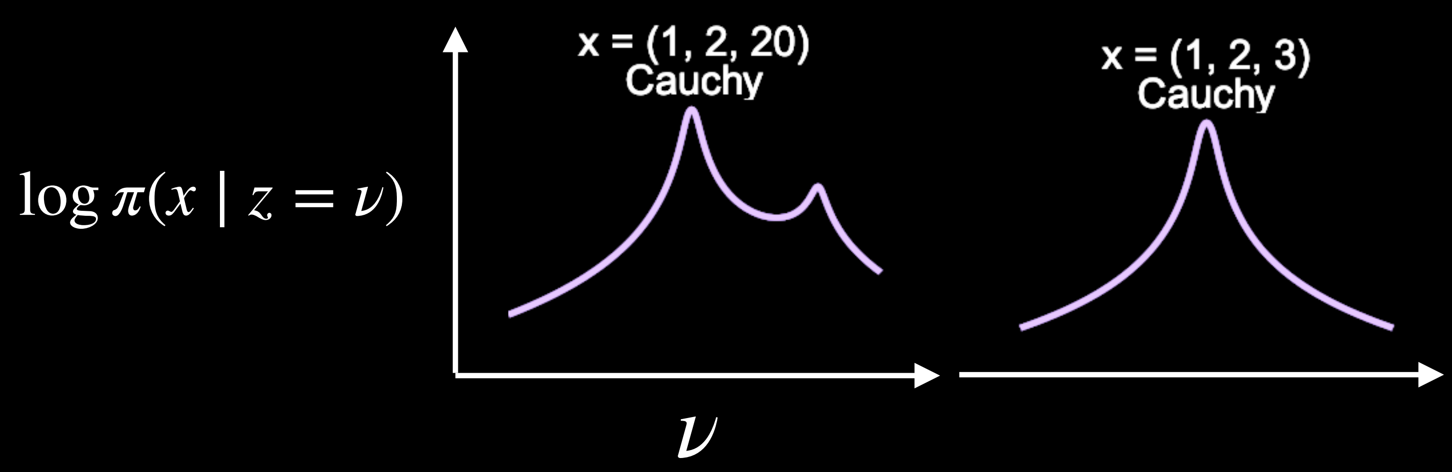
🤔 When does p have the studied symmetries?

In Bayesian setting:

$$p(z) = \pi(z \mid x_{1:N}) \propto \pi(z)\ \pi(x_{1:N} \mid z)$$

Sparse data regime: prior $\pi(z)$ dominates.

Large data regime: likelihood $\pi(x_{1:N} \mid z)$ dominates.



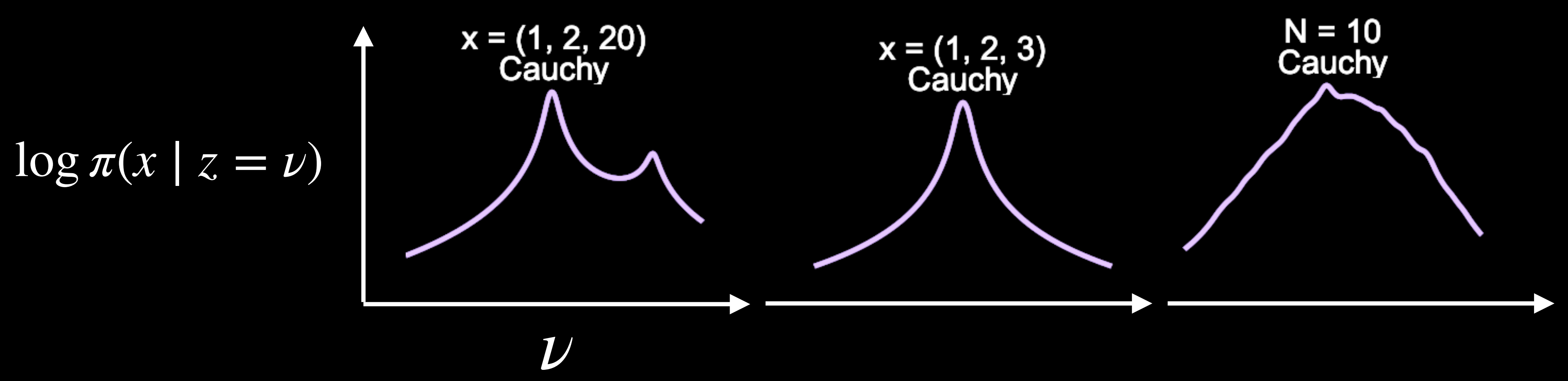
🤔 When does p have the studied symmetries?

In Bayesian setting:

$$p(z) = \pi(z \mid x_{1:N}) \propto \pi(z)\ \pi(x_{1:N} \mid z)$$

Sparse data regime: prior $\pi(z)$ dominates.

Large data regime: likelihood $\pi(x_{1:N} \mid z)$ dominates.



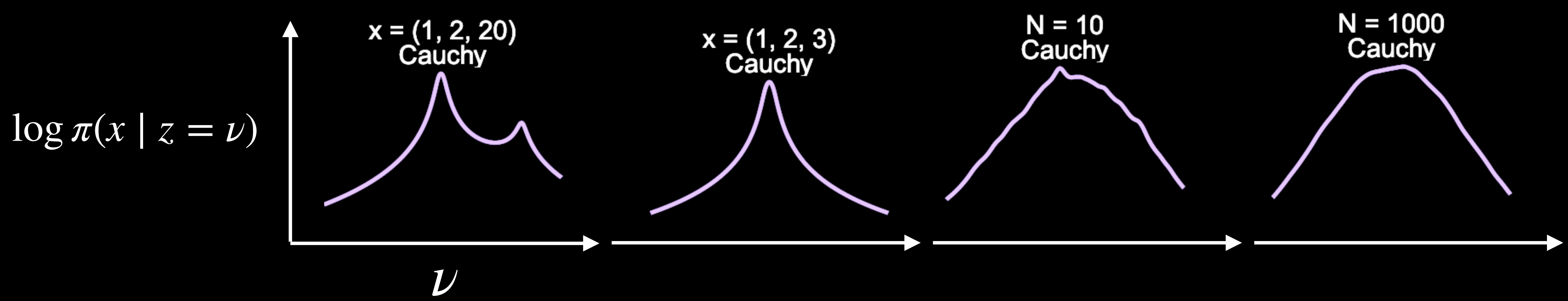
🤔 When does p have the studied symmetries?

In Bayesian setting:

$$p(z) = \pi(z \mid x_{1:N}) \propto \pi(z)\ \pi(x_{1:N} \mid z)$$

Sparse data regime: prior $\pi(z)$ dominates.

Large data regime: likelihood $\pi(x_{1:N} \mid z)$ dominates.



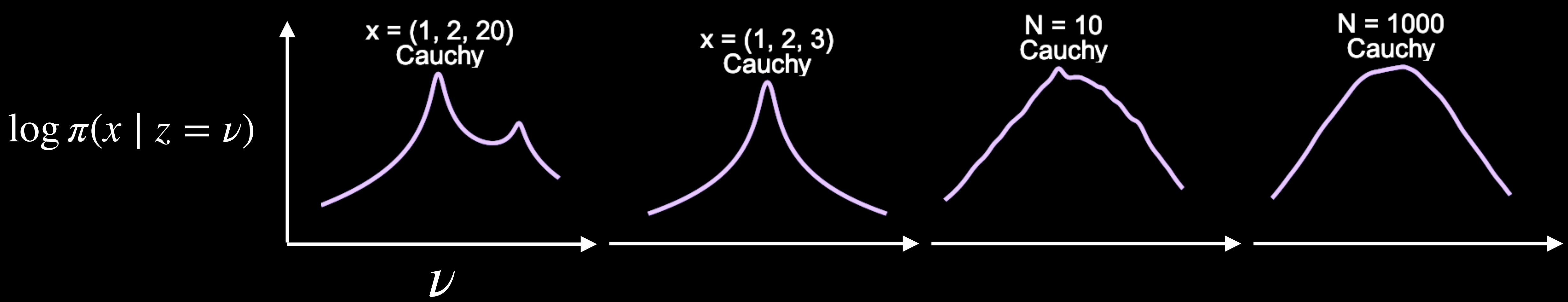
🤔 When does p have the studied symmetries?

In Bayesian setting:

$$p(z) = \pi(z \mid x_{1:N}) \propto \pi(z)\ \pi(x_{1:N} \mid z)$$

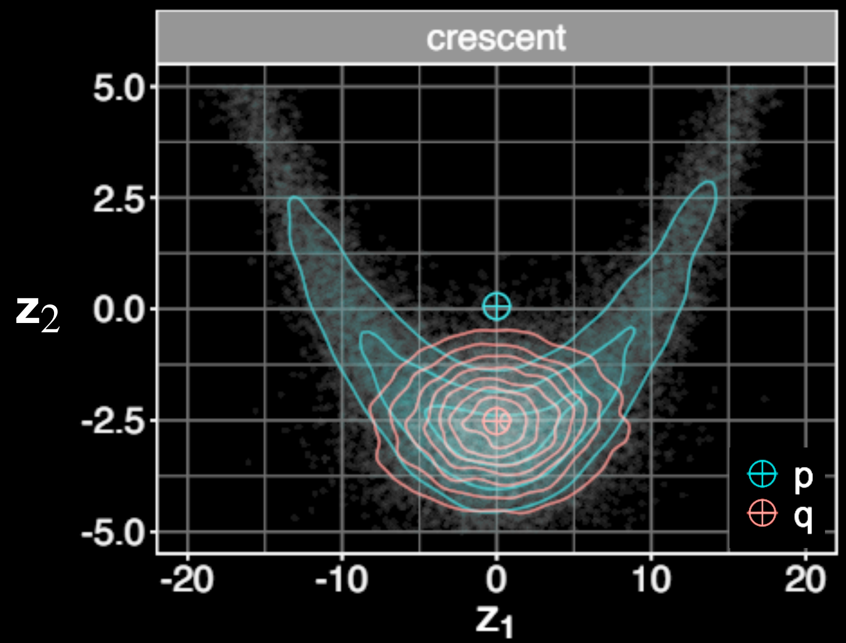
Sparse data regime: prior $\pi(z)$ dominates.

Large data regime: likelihood $\pi(x_{1:N} \mid z)$ dominates.

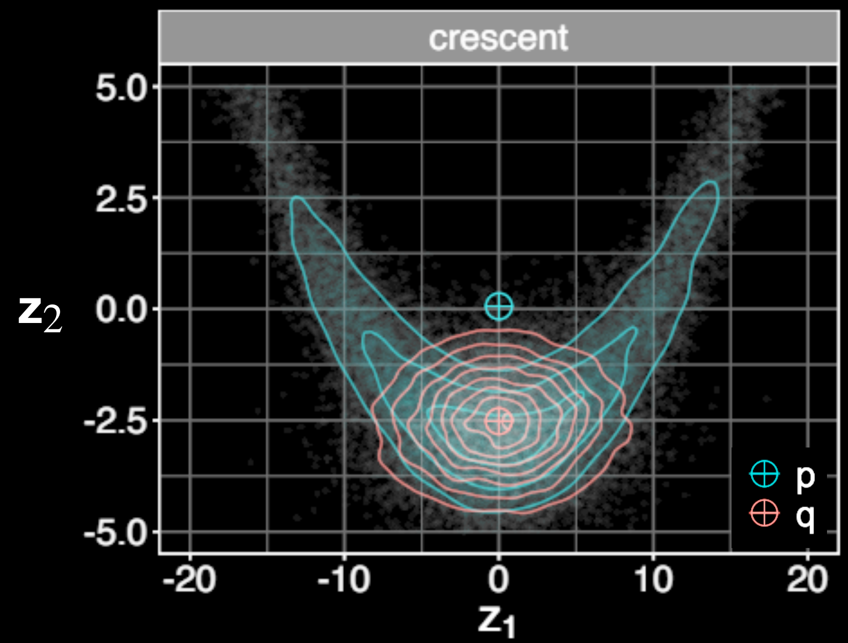


***Bernstein-von Mise:** as $N \to \infty$, $\pi(z \mid x_{1:N})$ becomes Gaussian (which has even and elliptical symmetry)

Targets with non-ideal symmetries

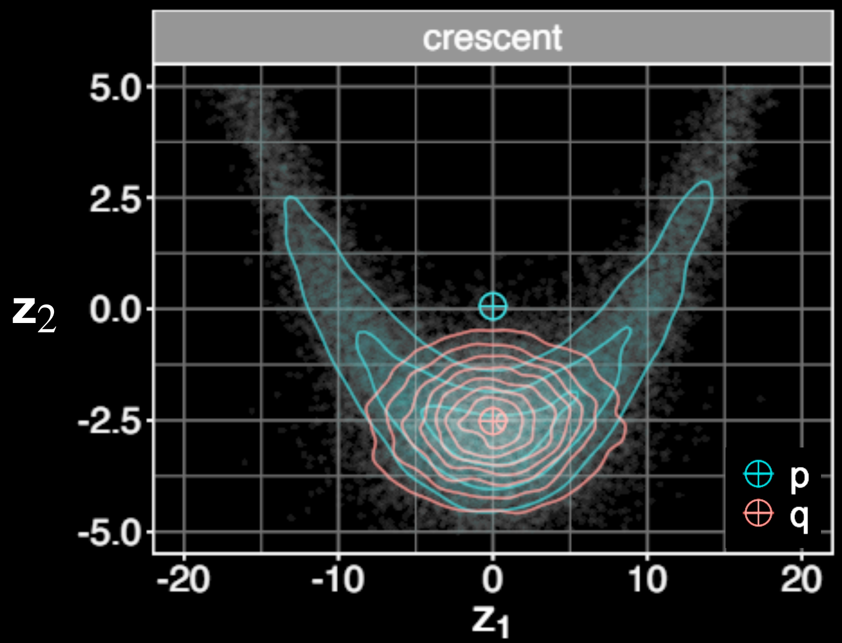


Targets with non-ideal symmetries



Violation of symmetry

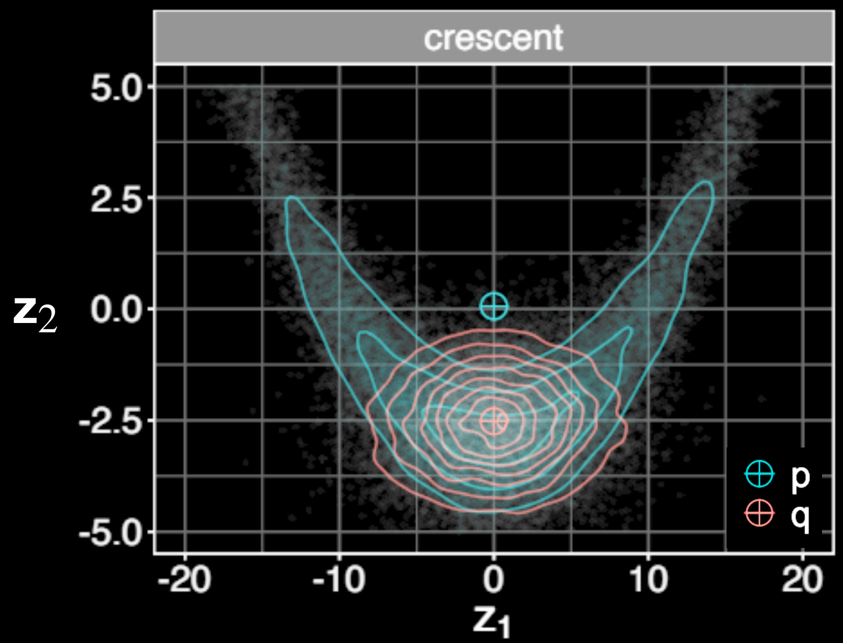
Targets with non-ideal symmetries



Violation of symmetry

1. $z \sim p(z)$

Targets with non-ideal symmetries

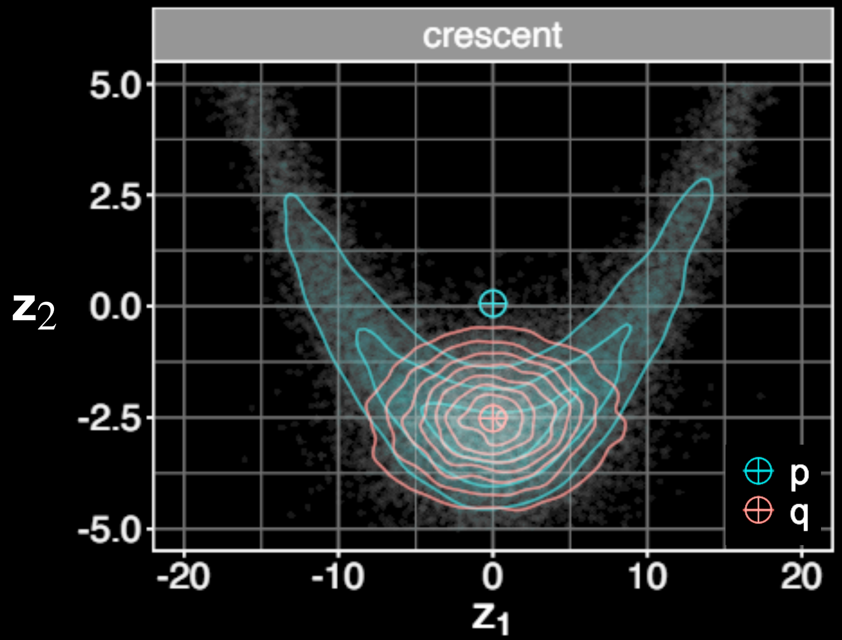


Violation of symmetry

1. $z \sim p(z)$

2. $z' = 2\mathbb{E}(z) - z$

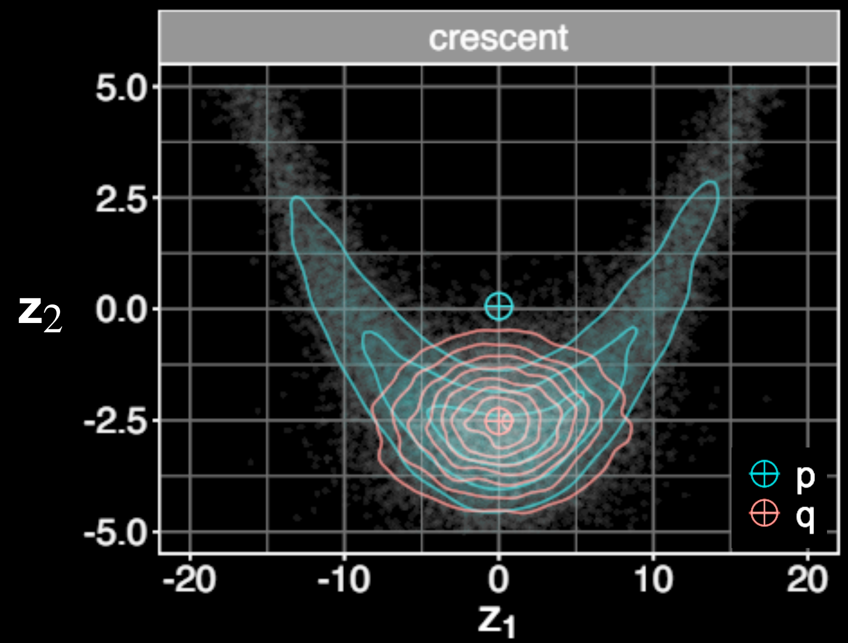
Targets with non-ideal symmetries



Violation of symmetry

1. $z \sim p(z)$
2. $z' = 2\mathbb{E}(z) - z$
3. $\varepsilon(z) = \left| \dfrac{\log \pi(z \mid x) - \log \pi(z' \mid x)}{\log \pi(z \mid x)} \right|$

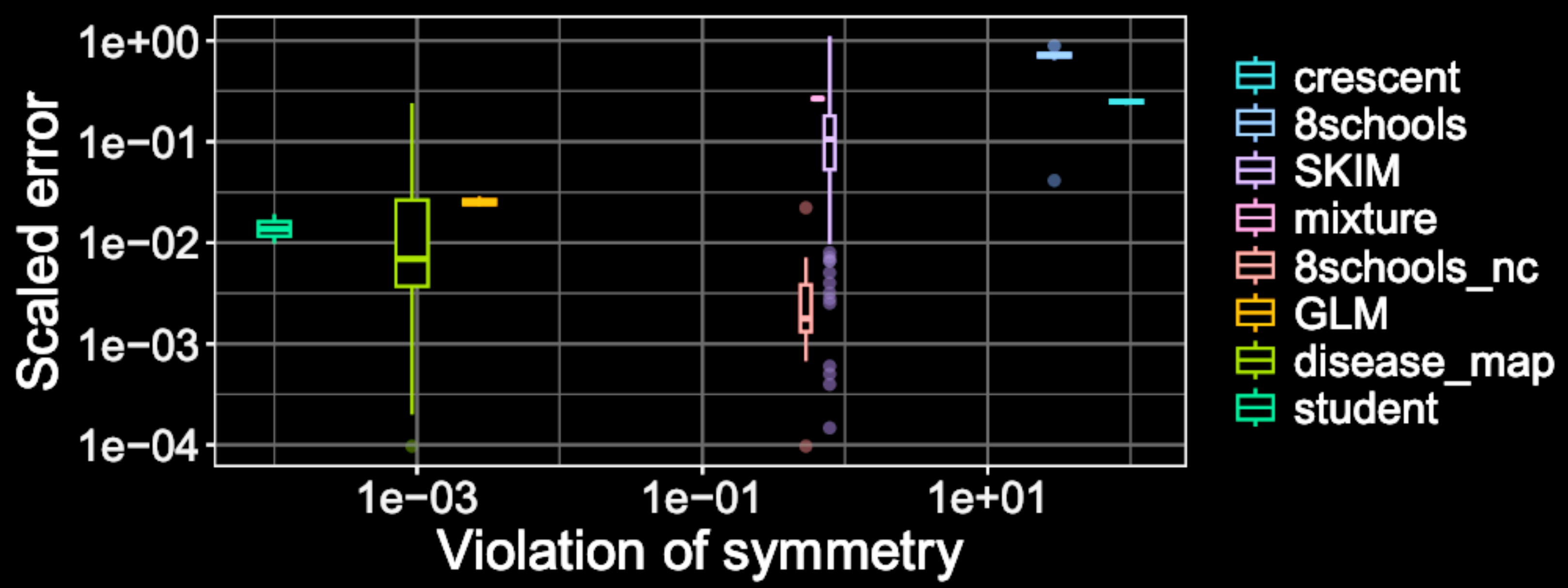
Targets with non-ideal symmetries



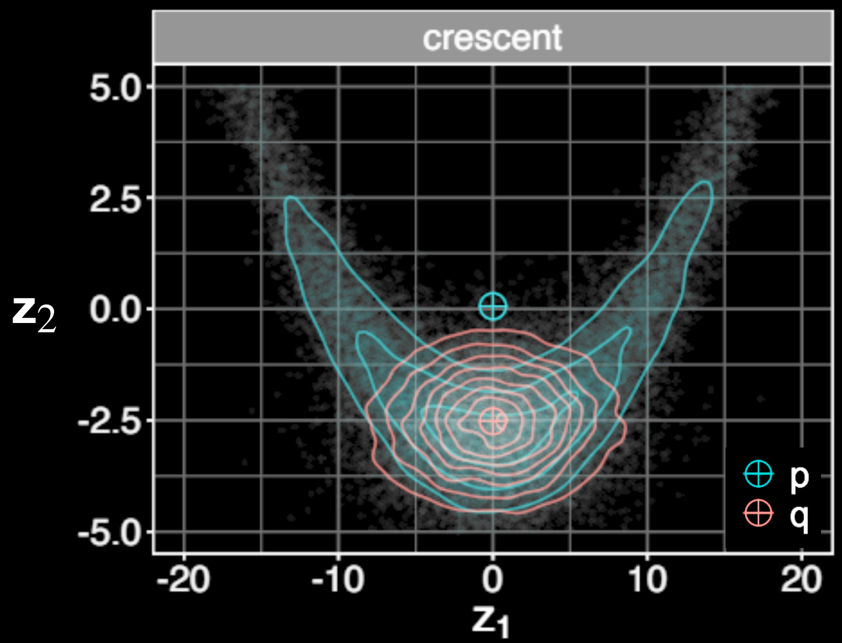
Violation of symmetry

1. $z \sim p(z)$
2. $z' = 2\mathbb{E}(z) - z$
3. $\varepsilon(z) = \left| \dfrac{\log \pi(z \mid x) - \log \pi(z' \mid x)}{\log \pi(z \mid x)} \right|$

(Scaled) error for estimates of mean across all dimensions



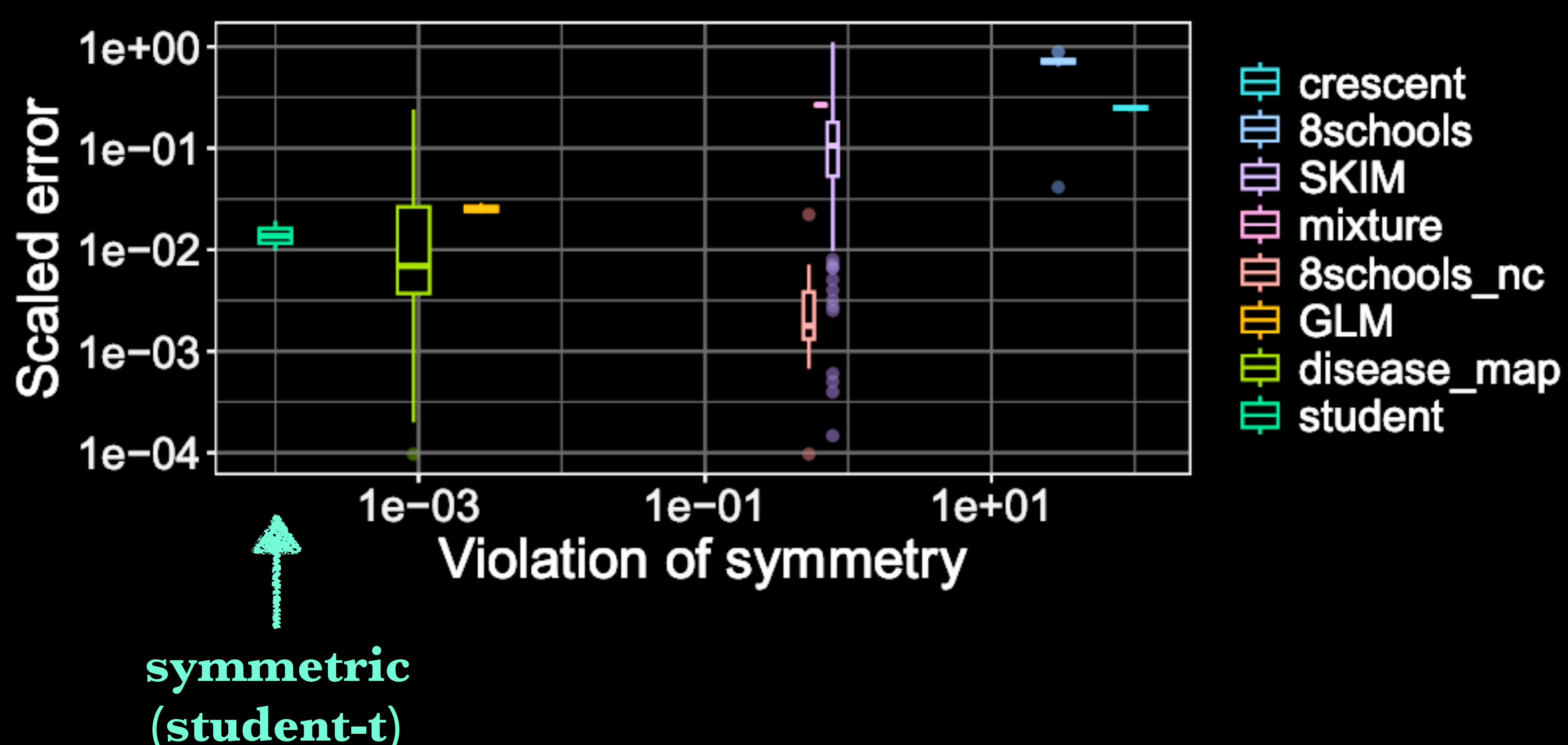
Targets with non-ideal symmetries



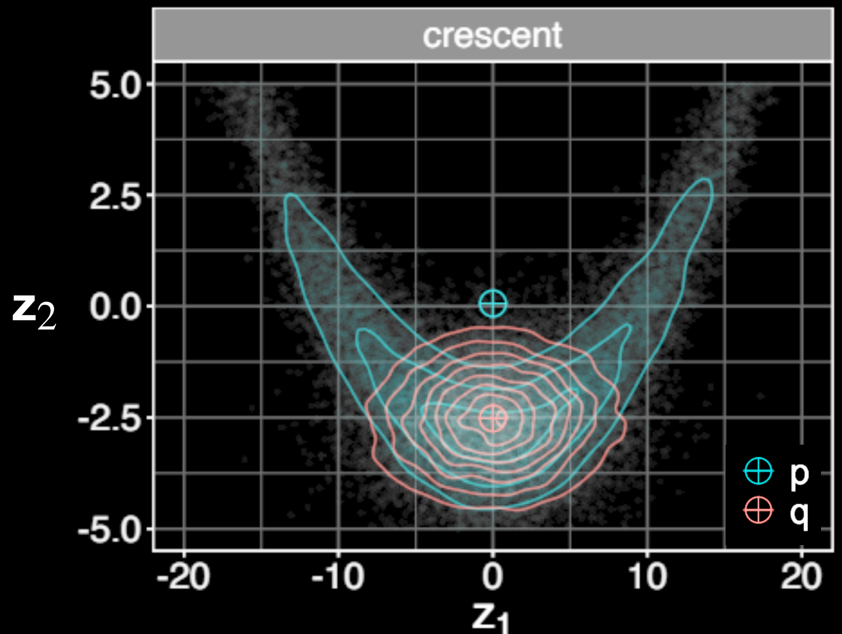
Violation of symmetry

1. $z \sim p(z)$
2. $z' = 2\mathbb{E}(z) - z$
3. $\varepsilon(z) = \left| \dfrac{\log \pi(z \mid x) - \log \pi(z' \mid x)}{\log \pi(z \mid x)} \right|$

(Scaled) error for estimates of mean across all dimensions



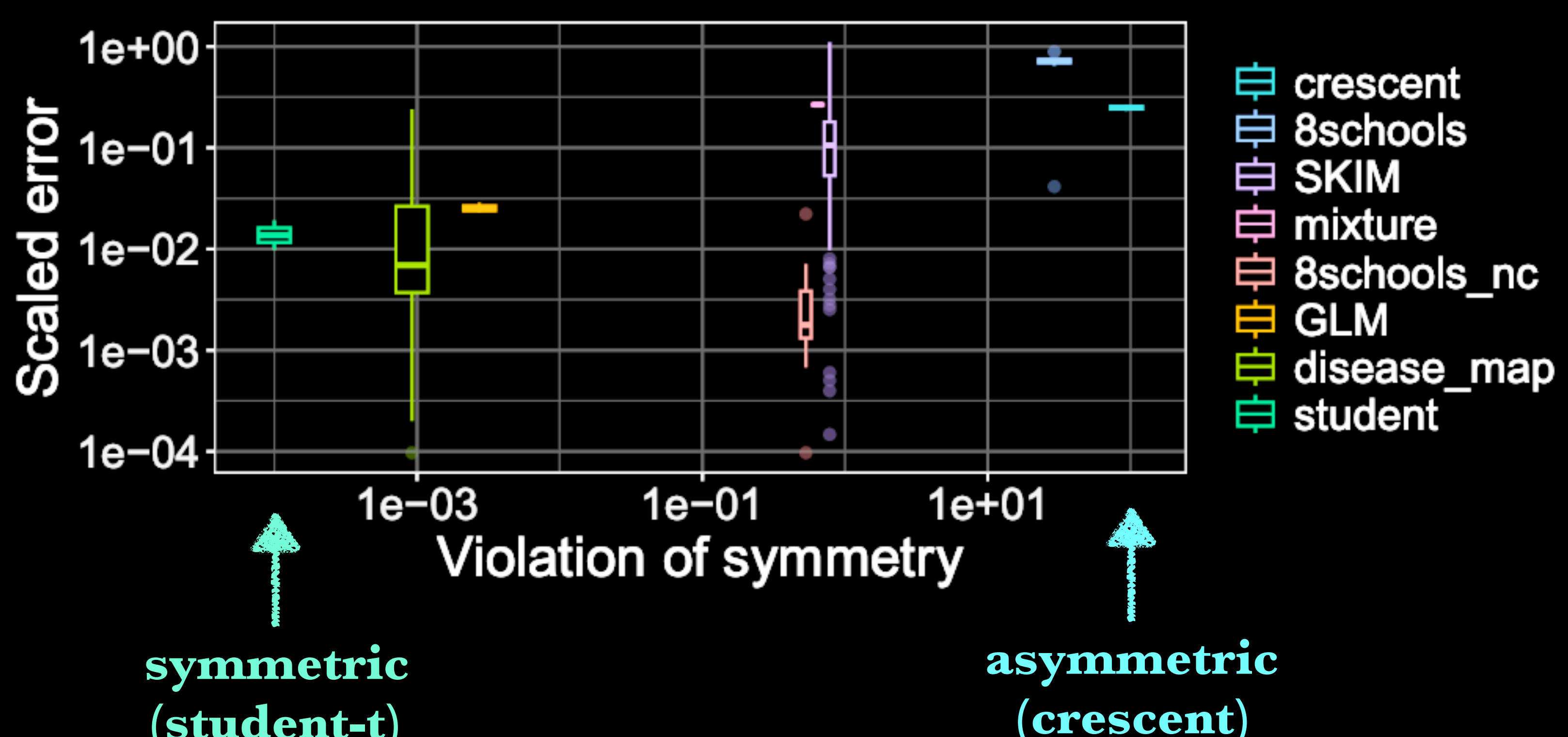
Targets with non-ideal symmetries



Violation of symmetry

1. $z \sim p(z)$
2. $z' = 2\mathbb{E}(z) - z$
3. $\varepsilon(z) = \left| \dfrac{\log \pi(z \mid x) - \log \pi(z' \mid x)}{\log \pi(z \mid x)} \right|$

(Scaled) error for estimates of mean across all dimensions



Why the next paper will be more interesting

Why the next paper will be more interesting

🖌️ **A more nuanced theory:** error bound based on violation of symmetry

Why the next paper will be more interesting

🖌️ **A more nuanced theory:** error bound based on violation of symmetry
Follow example of previous work on pre-asymptotic
regimes of VI [e.g. Katsevich and Rigollet'24]

Why the next paper will be more interesting

🖌️ **A more nuanced theory:** error bound based on violation of symmetry
Follow example of previous work on pre-asymptotic
regimes of VI [e.g. Katsevich and Rigollet'24]

🔎 **An actionable diagnostic:**
Find measures of asymmetry we can use to bound error in practice.

Why the next paper will be more interesting

🖌️ **A more nuanced theory:** error bound based on violation of symmetry
Follow example of previous work on pre-asymptotic regimes of VI [e.g. Katsevich and Rigollet'24]

🔎 **An actionable diagnostic:**
Find measures of asymmetry we can use to bound error in practice.

💭 **Conjecture:**
For each symmetry, there exists a statistics that VI recovers.