Radev, Stefan T., et al. "BayesFlow: Learning complex stochastic models with invertible neural networks." *IEEE transactions on neural networks and learning systems* 33.4 (2020): 1452-1466.
Orozco, Rafael, et al. "Adjoint operators enable fast and amortized machine learning based Bayesian uncertainty quantification." Medical Imaging 2023: Image Processing. Vol. 12464. SPIE, 2023.
Alsing, Justin, Benjamin Wandelt, and Stephen Feeney. "Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology." Monthly Notices of the Royal Astronomical Society 477.3 (2018): 2874-2885.

Summary statistic

physics-based

Assume Gaussian log-likelihood:

$$\log p(\mathbf{y} \mid \mathbf{x}) = \|\mathscr{H}(\mathbf{x}) - \mathbf{y}\|_2^2$$

Use *score* function at *fiducial* point $\bar{\mathbf{x}}$:

$$p(\mathbf{x} \mid \nabla_{\mathbf{x}} \log p(\mathbf{y} \mid \mathbf{x})\Big|_{\bar{\mathbf{x}}}) \approx p(\mathbf{x} \mid \mathbf{y})$$

in combination w/ learned *summary network* yields

$$\bar{\mathbf{y}} = h_\psi(\mathbf{J}_{\mathscr{H}}^\top(\bar{\mathbf{x}})(\mathscr{H}(\bar{\mathbf{x}}) - \mathbf{y}))$$

such that

$$p(\mathbf{x} \mid \bar{\mathbf{y}}) \approx p(\mathbf{x} \mid \mathbf{y})$$

Orozco, Rafael, et al. "Refining Amortized Posterior Approximations using Gradient-Based Summary Statistics." *arXiv preprint arXiv:2305.08733* (2023)
Siahkoohi, Ali, et al. "Reliable amortized variational inference with physics-based latent distribution correction." Geophysics 88.3 (2023): R297-R322.

Sampling
from the *posterior* distribution

After training on *in silico* pairs:

$$\{(\mathbf{x}^{(m)}, \bar{\mathbf{y}}^{(m)})\}_{m=1}^{M}$$

posterior samples conditioned on *summarized* field data $\bar{\mathbf{y}}^{\mathrm{o}}$ are drawn via

$$\mathbf{x} \sim p_{\hat{\phi}}^{-1}\left(\mathbf{x} \mid \bar{\mathbf{y}} = \bar{\mathbf{y}}^{\mathrm{o}}\right) \iff \mathbf{x} = f_{\hat{\phi}}^{-1}(\mathbf{z}; \bar{\mathbf{y}}^{\mathrm{o}}) \quad \text{with} \quad \mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$$

- CNF is *amortized*
- training *marginalizes* forward KL divergence $\mathbb{E}_{\mathbf{y}\sim p(\mathbf{y})}\left[\mathbb{KL}\left(p(\mathbf{x} \mid \mathbf{y}) \,||\, p_{\phi}(\mathbf{x})\right)\right]$