Why approximate inference?

Analytical inference

$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)}$$

- Easy for conjugate priors
- Hard otherwise

Example: $p(x|z) = \mathcal{N}(x|\mu(z), \sigma^2(z))$

sometimes, computing the posterior is intractable like this.
We want to APPROXIMATE the posterior so that it is easier for us, and that the approximation is acceptable.

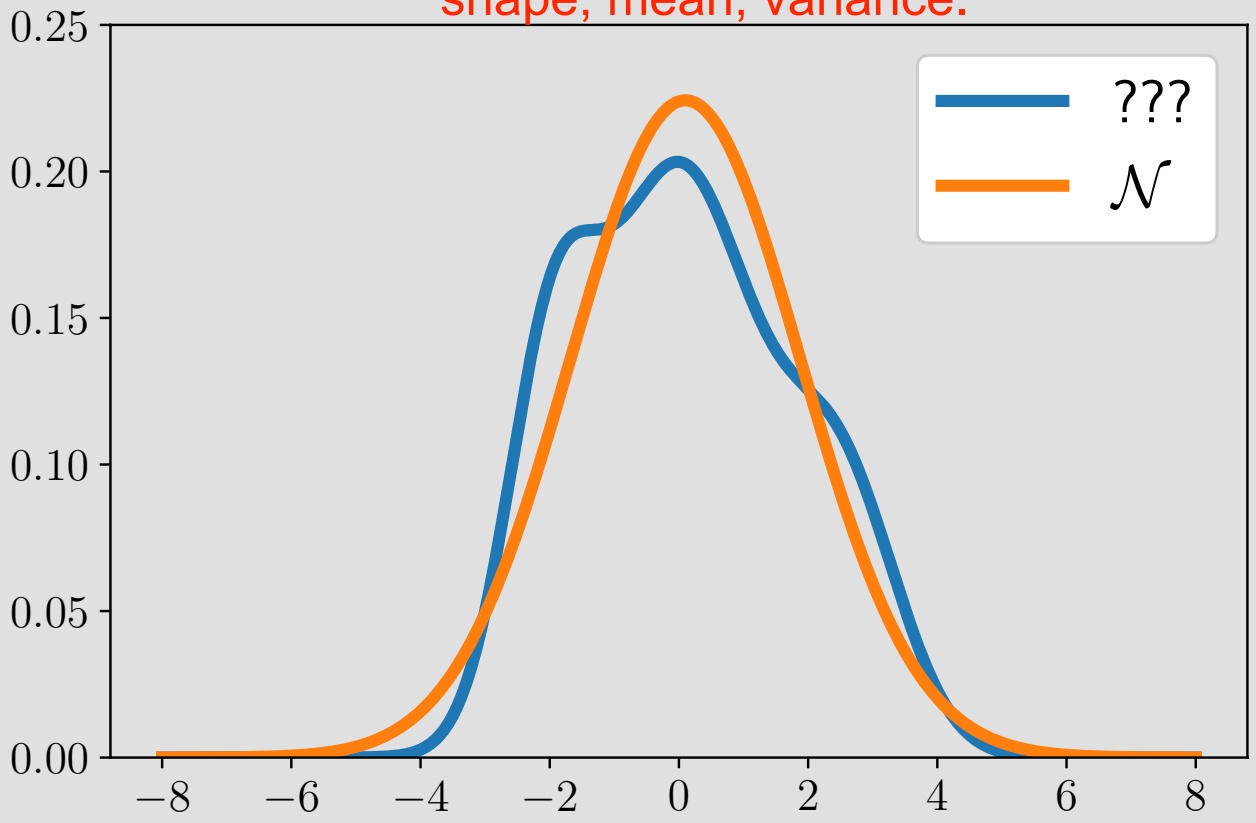
Neural networks

the likelihood here is a neural network.
Here, there is obviously no conjugacy.

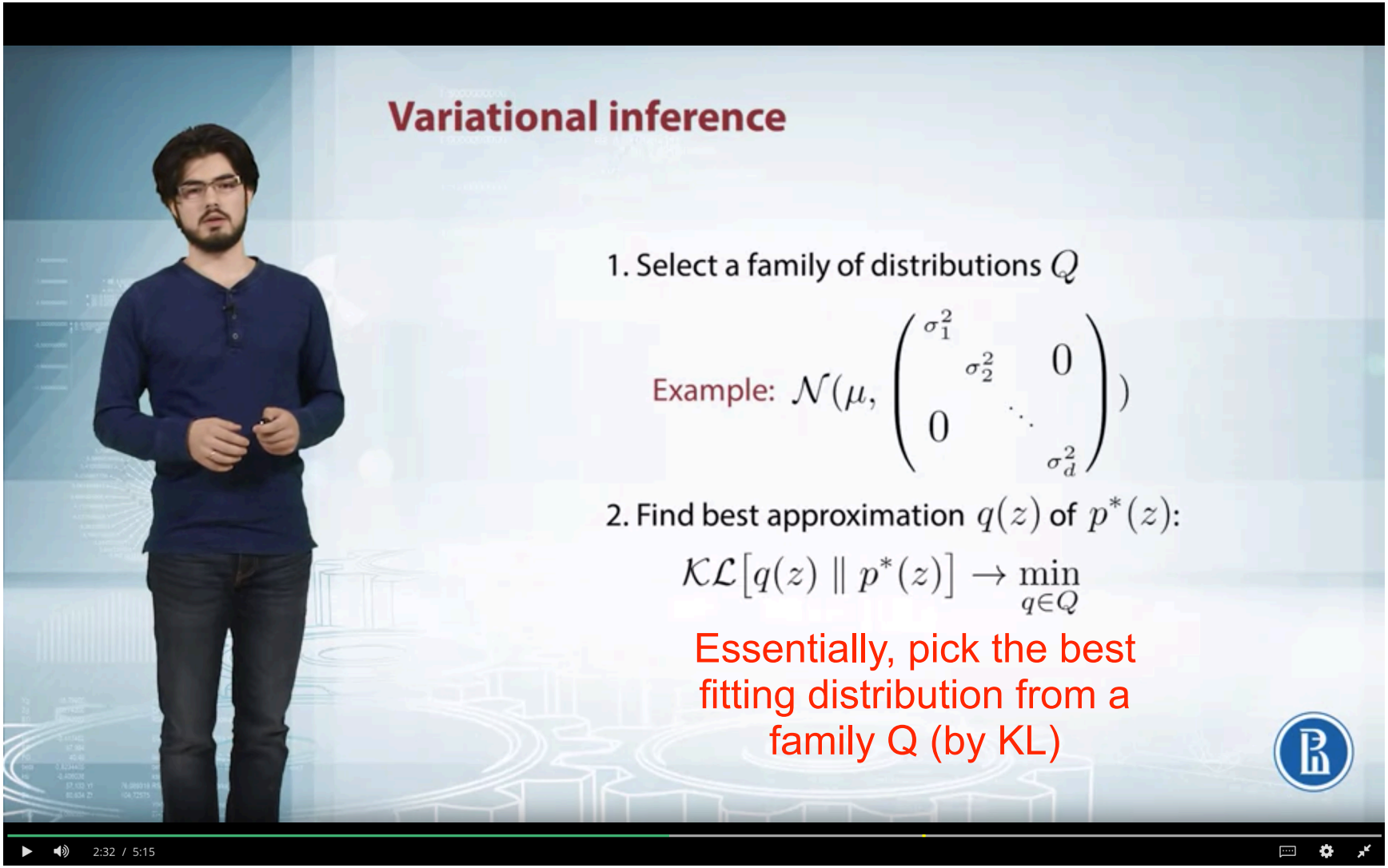
Do we need exact posterior?

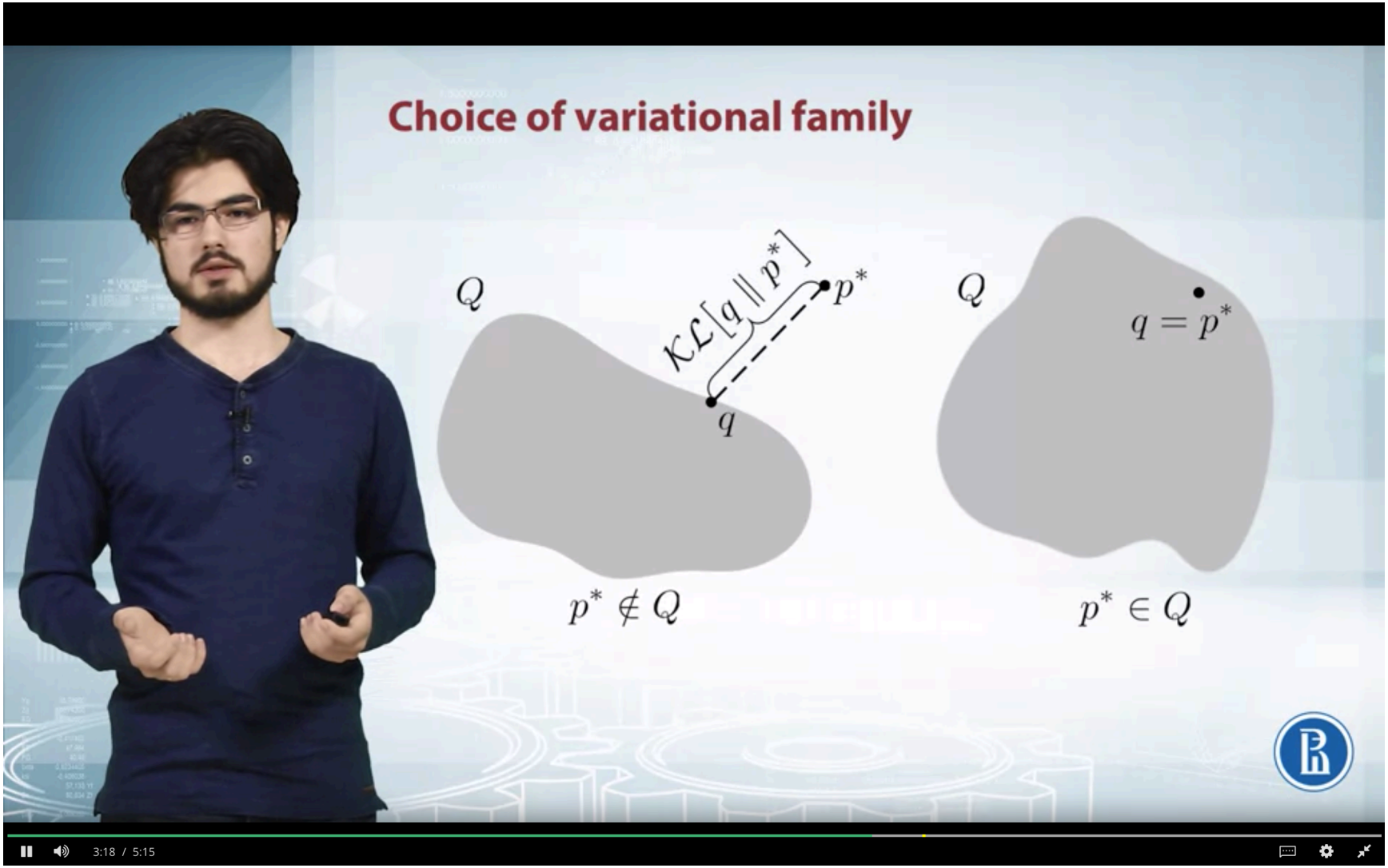
not really right? An approximation is good enough?

See the gaussian approximation below. It is a really good approximation. Matches shape, mean, variance.

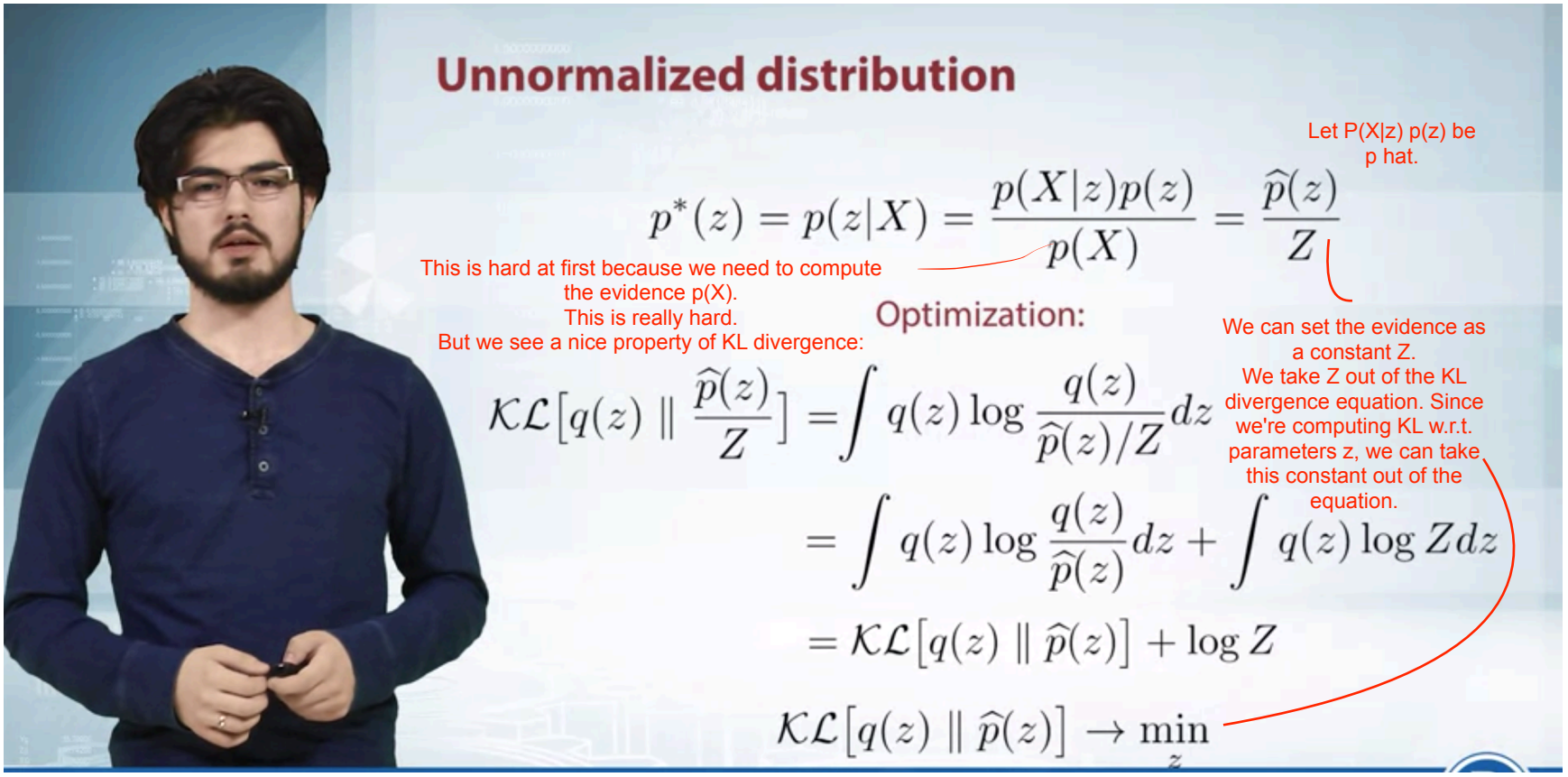


This week we'll be studying how to approximate the posterior.


Variational inference
1. Select a family of distributions Q
Example: $\mathcal{N}(\mu, \begin{pmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \\ & & & \sigma_d^2 \end{pmatrix})$
2. Find best approximation $q(z)$ of $p^*(z)$:
$\mathcal{KL}\left[q(z) \parallel p^*(z)\right] \to \min_{q \in Q}$
Essentially, pick the best fitting distribution from a family Q (by KL)
2:32 / 5:15



If Q is small, the true posterior will not lie in it.
If Q is sufficiently large, we'll get a good posterior.
If Q is larger, then it is harder to compute the variational inference. Search space too big, presumably?


Unnormalized distribution
$$p^*(z) = p(z|X) = \frac{p(X|z)p(z)}{p(X)} = \frac{\widehat{p}(z)}{Z}$$
Let P(X|z) p(z) be p hat.
This is hard at first because we need to compute the evidence p(X).
This is really hard.
But we see a nice property of KL divergence:
Optimization:
We can set the evidence as a constant Z.
We take Z out of the KL divergence equation. Since we're computing KL w.r.t. parameters z, we can take this constant out of the equation.
$$\mathcal{KL}\left[q(z) \parallel \frac{\widehat{p}(z)}{Z}\right] = \int q(z) \log \frac{q(z)}{\widehat{p}(z)/Z} dz$$
$$= \int q(z) \log \frac{q(z)}{\widehat{p}(z)} dz + \int q(z) \log Z dz$$
$$= \mathcal{KL}\left[q(z) \parallel \widehat{p}(z)\right] + \log Z$$
$$\mathcal{KL}\left[q(z) \parallel \widehat{p}(z)\right] \to \min_z$$