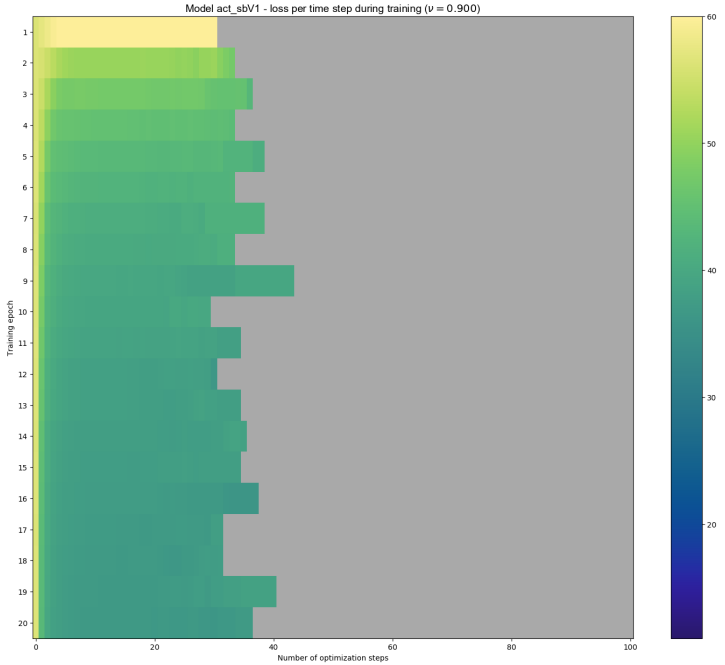
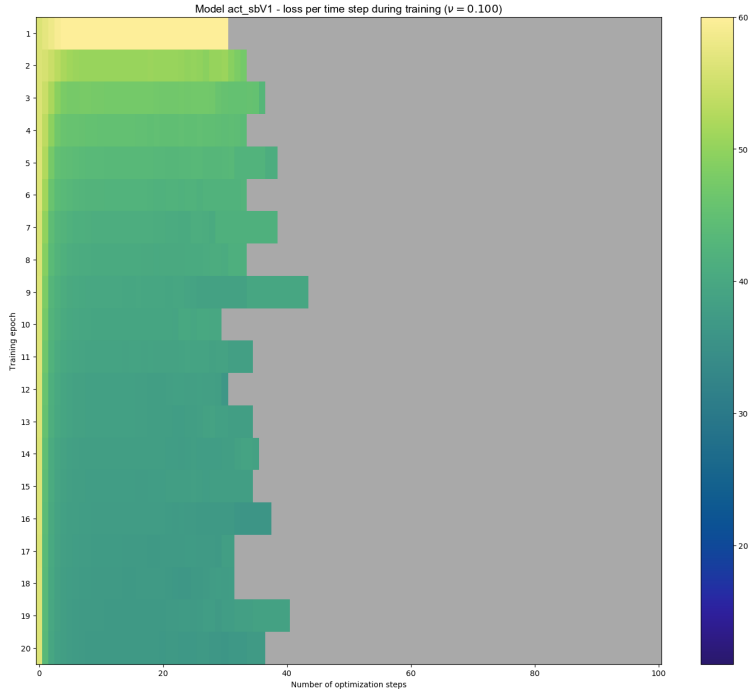
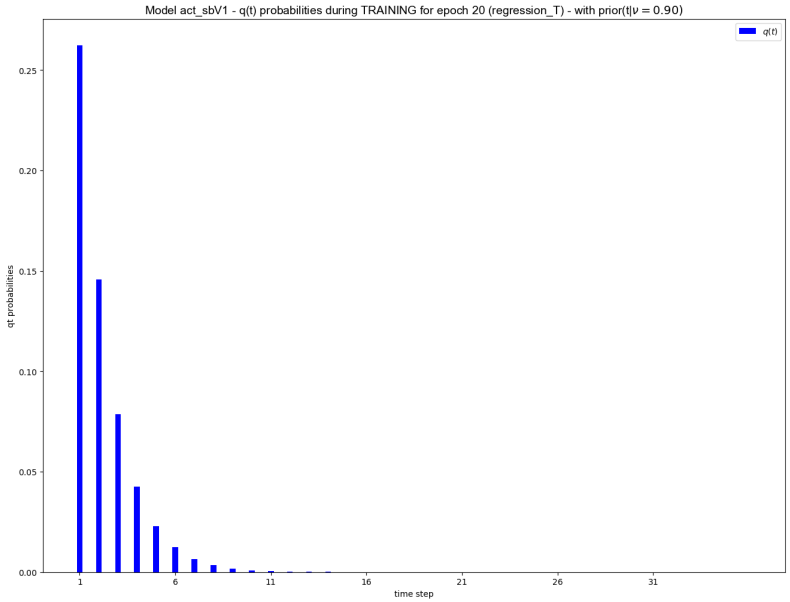
Illustrate *behavior* of ACT-SB models during training with different prior shape parameters (nu) - **step loss figures and q(t|x) during training**

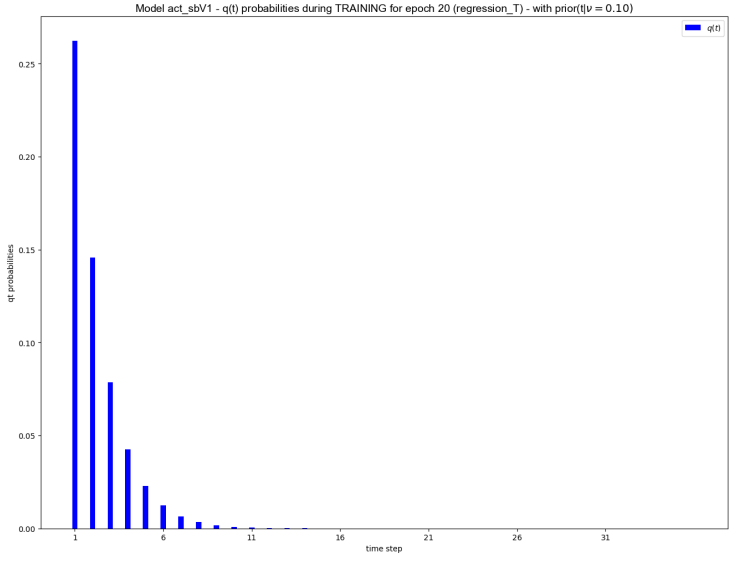
$\nu = 0.90$



$\nu = 0.1$

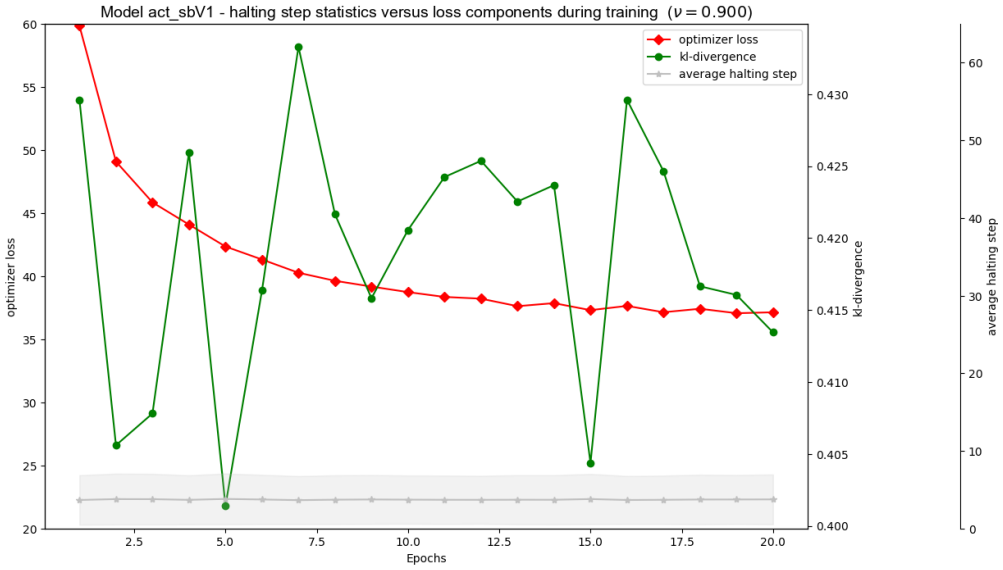






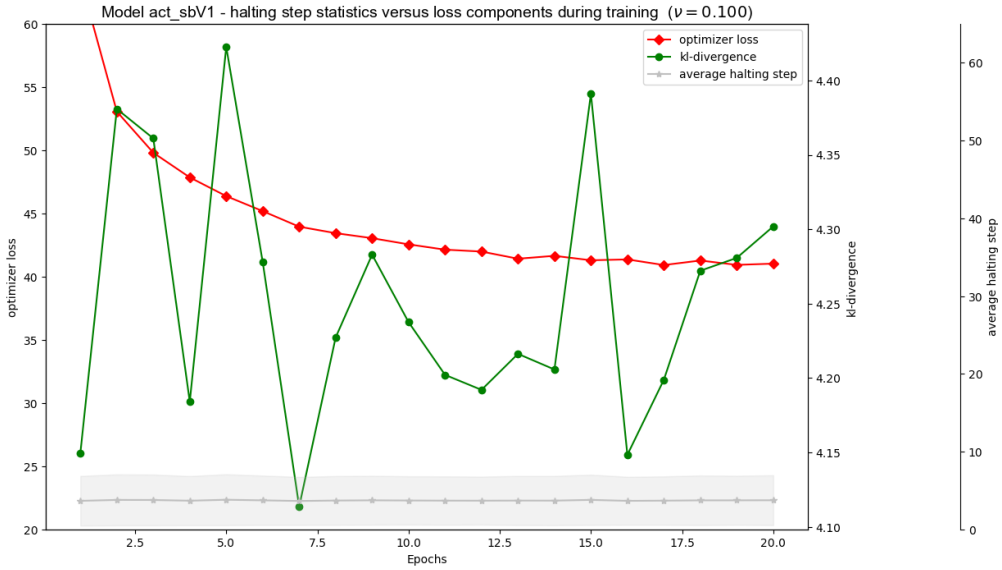
Illustrate *behavior* of ACT-SB models during training with different prior shape parameters (nu) - **optimizer loss - gradient statistics**

$\nu = 0.90$

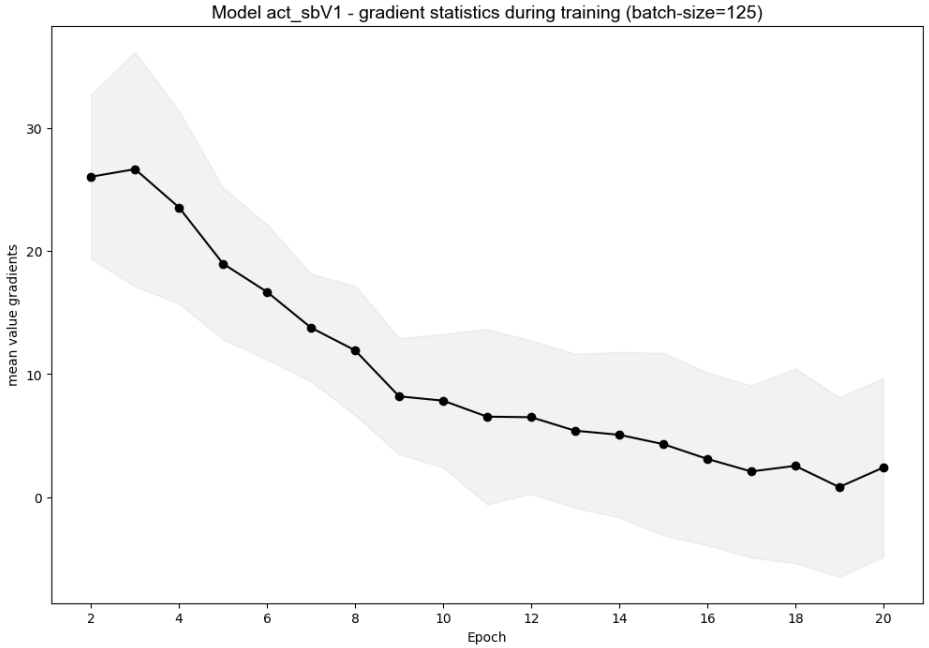


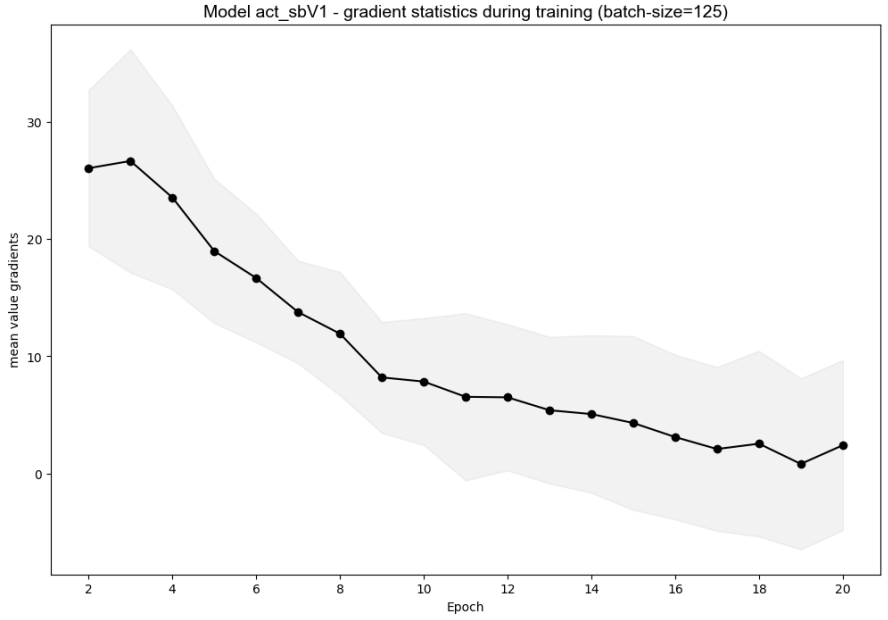
Halting step statistics final epoch: Range(1, 26) / mean=3.8 / stddev=3.2 / median=3.0

$\nu = 0.1$



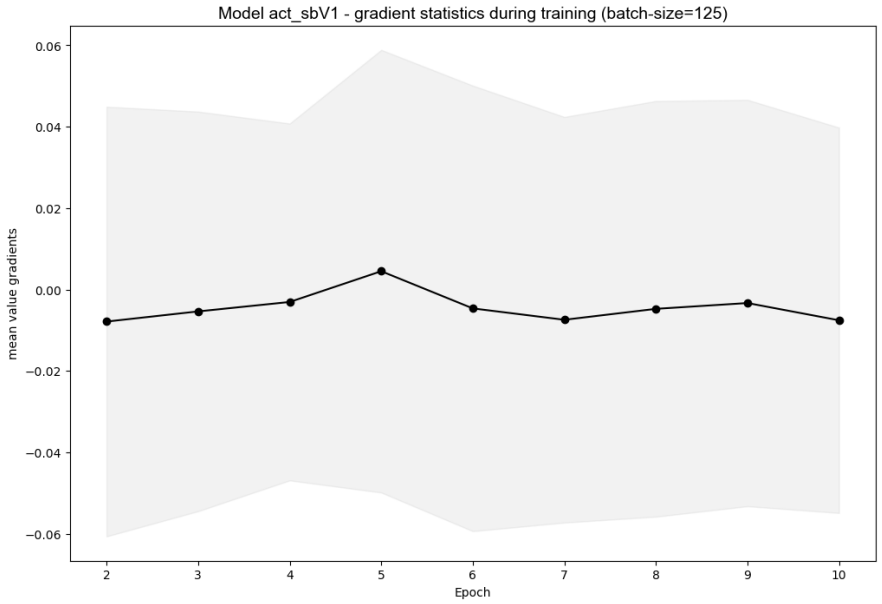
Halting step statistics final epoch: Range(1, 26) / mean=3.8 / stddev=3.2 / median=3.0



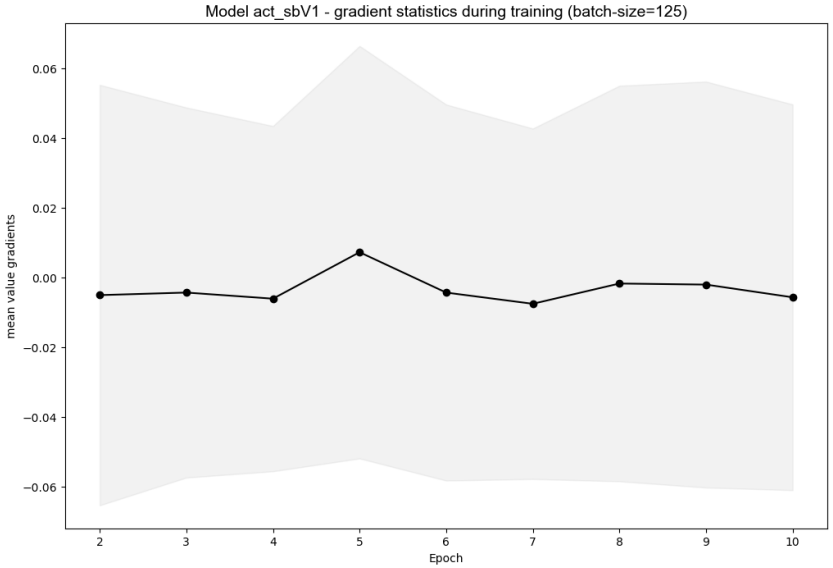


Illustrate *behavior* of ACT-SB models during training with different prior shape parameters (nu) - **gradients produced only by KL term**

$\nu = 0.90$

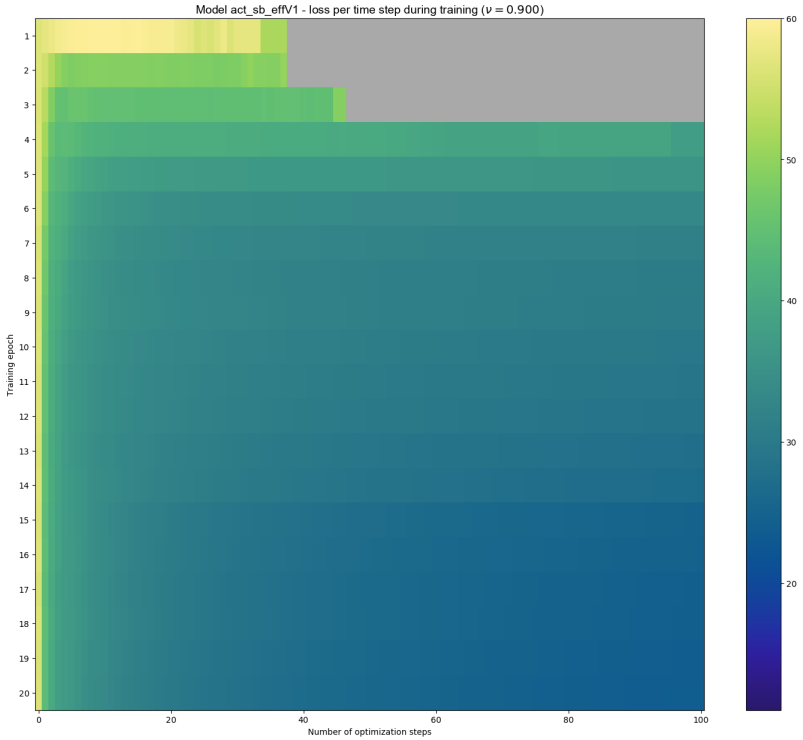


$\nu = 0.1$

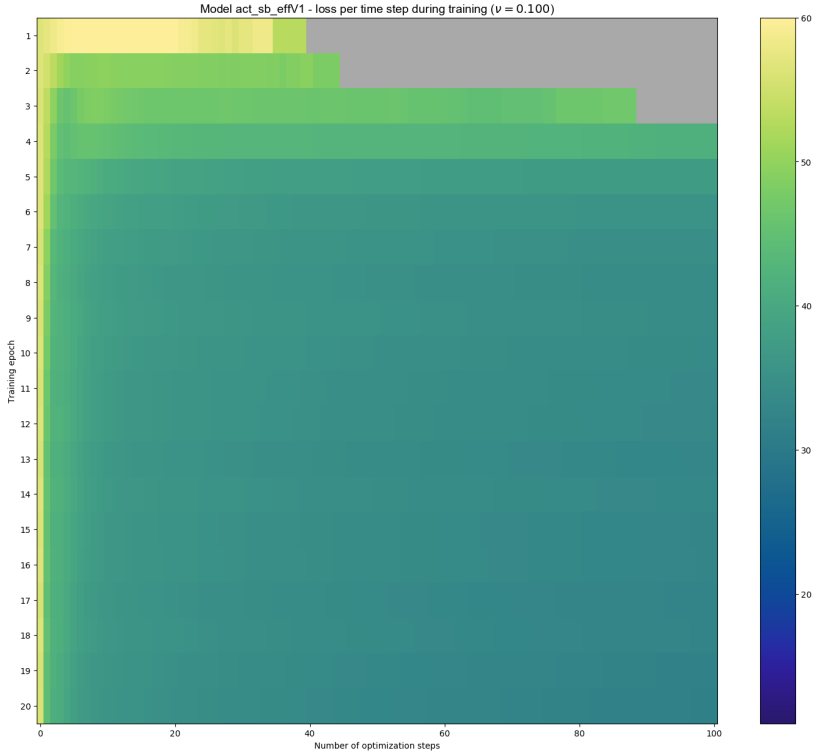


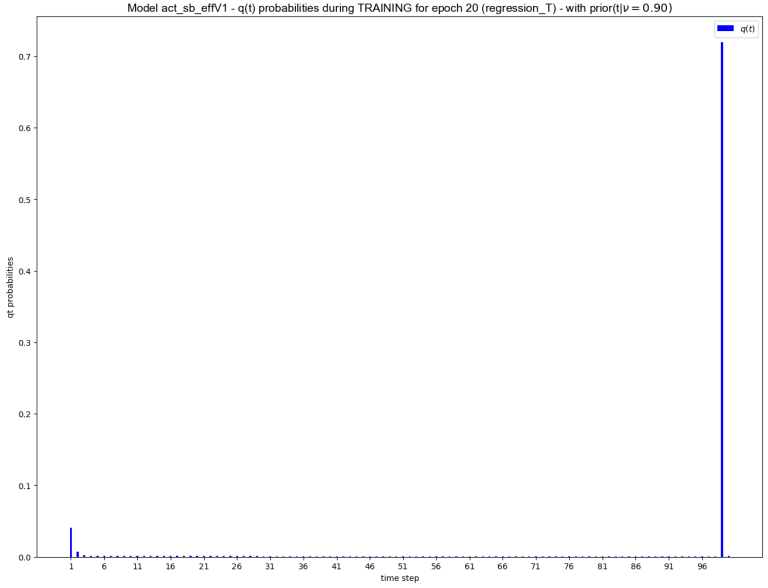
Illustrate *behavior* of **ACT-SB-ExT** models during training with different prior shape parameters (nu) - **step loss figures and q(t|x) during training**

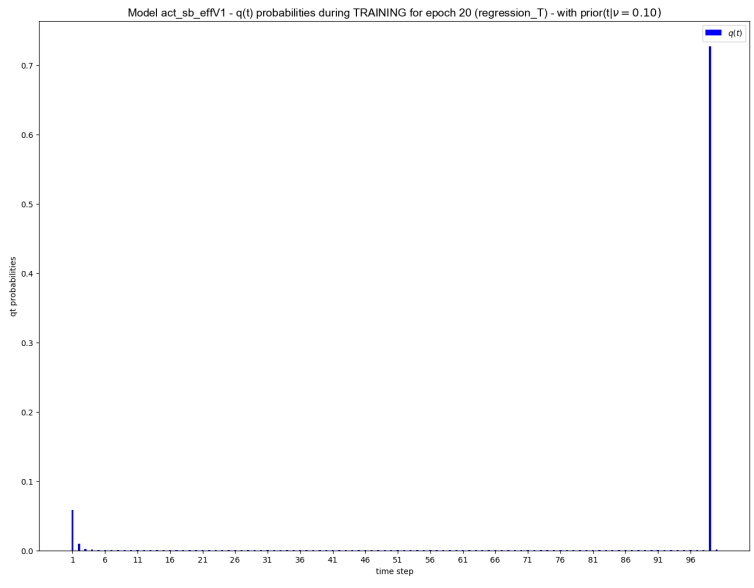
$\nu = 0.90$



$\nu = 0.1$

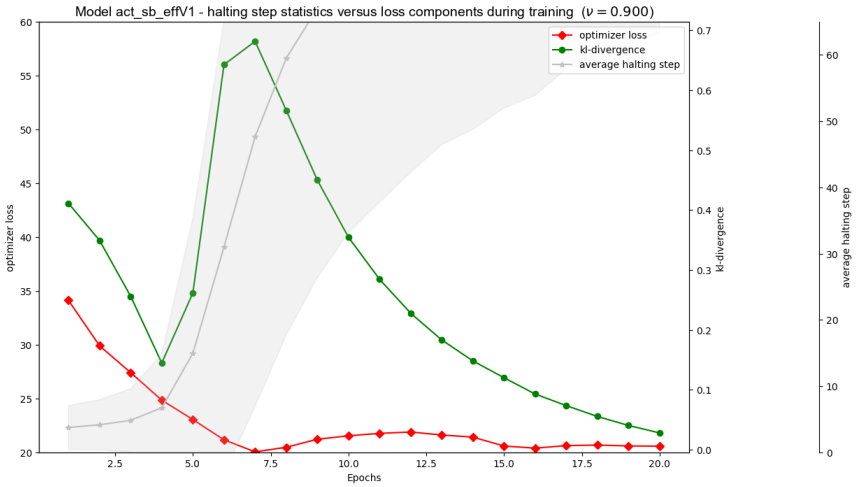




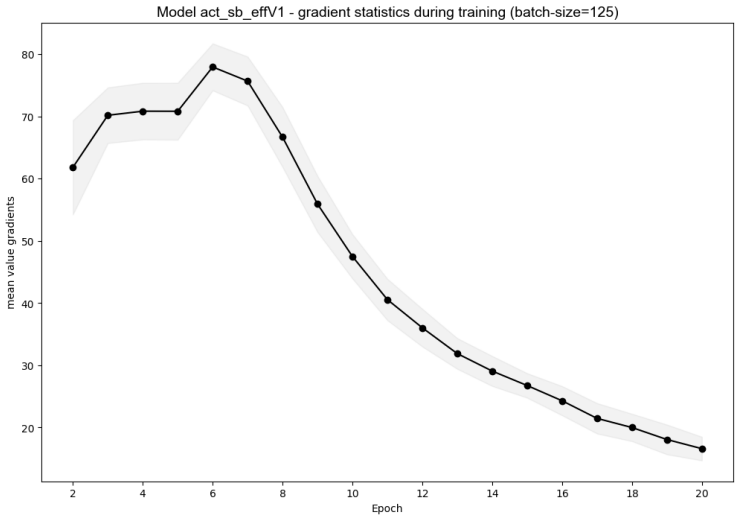


Illustrate *behavior* of **ACT-SB-ExT** models during training with different prior shape parameters (nu) - **optimizer loss - gradient statistics**

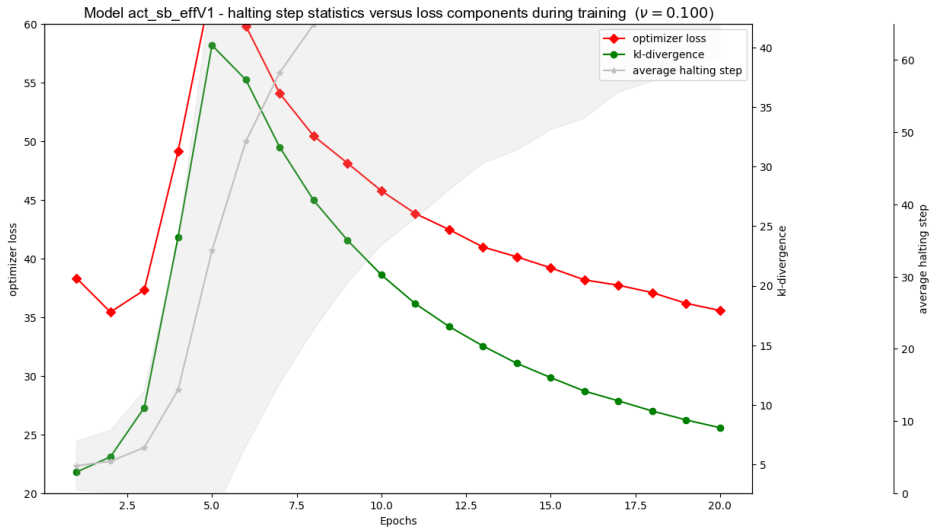
$\nu = 0.90$



Halting step statistics final epoch: Range(1, 100) / mean=90.0 / stddev=26.7 / median=100.0



$\nu = 0.1$



Halting step statistics final epoch: Range(1, 100) / mean=89.1 / stddev=28.7 / median=100.0