1.

Sigmoid $= \sigma(x) = \dfrac{1}{1+e^{-x}}$

$$\sigma(0) = \frac{1}{2}$$

$\dfrac{d}{dx}$ of sigmoid $= \sigma'(x) = \sigma(x)(1-\sigma(x))$

$$\sigma'(0) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Maclauren $= f(\varepsilon) = f(0) + f'(0)\varepsilon$

$$\sigma'(\varepsilon) = \frac{1}{4}$$

Tanh $= \dfrac{d}{dx} \tanh x = 1 - \tanh^2 x$

$$\tanh(0) = 0$$

$\dfrac{d}{dx}$ - $\tanh'(0) = 1 - \tanh^2(0) = 1$

Maclarin - $f(\varepsilon) = f(0) + f'(0)\varepsilon$

$$\tanh(\varepsilon) \approx \varepsilon$$

$\dfrac{d}{dx}$ Mac - $\tanh'(\varepsilon) \approx 1$

2. σ approx $(\varepsilon) \leq \tanh$ approx (ε)

$$\frac{1}{2} + \frac{1}{4}\varepsilon \leq \varepsilon$$

$$\frac{1}{2} \leq \varepsilon - \frac{1}{4}\varepsilon$$

$$\frac{1}{2} \leq \frac{3}{4}\varepsilon$$

$$\varepsilon \geq \frac{2}{3}$$

$$\frac{1}{2} + \frac{1}{4}\varepsilon \leq \varepsilon \rightarrow \varepsilon \geq \frac{2}{3}$$

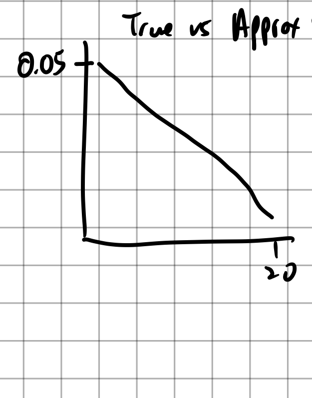
$$\varepsilon \in \left[\frac{2}{3}, \infty\right)$$

3. Both approx ratio $p(\varepsilon)^T$ & True ratio $p(\varepsilon)^T$ exhibit exponential decay as T increase.

True ratio $p(\varepsilon)^T = \left(\dfrac{\sigma'(\varepsilon)}{\tanh'(\varepsilon)}\right)^T$

Approximate ratio: $p_{approx}(\varepsilon)^T = (0.25)^T$

$\varepsilon = 0.05$, $T = 1$ to 20



4. Because backprog through time multiplies the derivatives across many steps, sigmoid causes gradient to vanish rapidly.

But tanh with derivative close to 1 will make gradients much longer.

$\therefore$ Tanh slow downs vanish-gradient problem & allows RNNs to learn dependencies over more time steps