[2] Lemma Two: differentiating $\text{quantize}_k(r_{i_\ell})$ yields 1 under condition of Expectation of a Straight-Through Estimator.

Chromium OSX
https://github.com/ppwwyyxx/tensorpack/issues/31#issuecomment-254919972

(1) Given $\text{quantize}_k(r_{i_\ell}) = (1 / (2^k-1)) * \text{round}((2^k - 1)r_{i_\ell})$.
(2) Consider $\sim\text{quantize}_k(r_{i_\ell}) \coloneqq (1 / (2^k-1)) * ((2^k - 1)r_{i_\ell}) = r_{i_\ell}$.
(3) $\sim\text{quantize}'_k(r_{i_\ell}) = 1$.
(4) Observation: in the absence of a *round()* function the two scale factors cancel and the derivative is 1. Can we make the *round()* function "disappear" in some justified fashion?
(5) Note that the *round(r_{i_ℓ})* function adds some number $n \in [-0.5, 0.5]$ to r_{i_ℓ} to round r_{i_ℓ} to the nearest integer.
(6) If $n = 0$, then $\text{quantize}'_k(r_{i_\ell}) = \sim\text{quantize}'_k(r_{i_\ell}) = 1$.
(7) if $n \neq 0$, then our outer scale factor of $1/(2^k-1)$ is technically multiplying the result of $(2^k-1)(r_{i_\ell} + n)$, *and thus the scale factors do not cancel.*
(8) Note that the probability of n == 0 is low, and so any individual call to quantize$_k$() is unlikely to result in a case where we can ignore the round() function.
(8) However, under assumption that n is drawn uniformly and at random from [-0.5, 0.5], the expectation $\mathbf{E}(n) = 0$.
(9) *∴ under expectation $\mathbb{E}(n) = 0$, $\text{quantize}'_k(r_{i_\ell}) = \sim\text{quantize}'_k(r_{i_\ell}) = 1$.**

[2] Lemma Two: differentiating $\text{quantize}_k(r_{i_\ell})$ yields 1 under condition of Expectation of a Straight-Through Estimator.

Firefox 49.0.2 OSX
https://github.com/ppwwyyxx/tensorpack/issues/31#issuecomment-254919972

(1) Given $\text{quantize}_k(r_{i_\ell}) = (1 / (2^k-1)) * \text{round}((2^k - 1)r_{i_\ell})$.
(2) Consider $\sim\text{quantize}_k(r_{i_\ell}) \coloneqq (1 / (2^k-1)) * ((2^k - 1)r_{i_\ell}) = r_{i_\ell}$.
(3) $\sim\text{quantize}'_k(r_{i_\ell}) = 1$.
(4) Observation: in the absence of a *round()* function the two scale factors cancel and the derivative is 1. Can we make the *round()* function "disappear" in some justified fashion?
(5) Note that the *round(r_{i_ℓ})* function adds some number $n \in [-0.5, 0.5]$ to r_{i_ℓ} to round r_{i_ℓ} to the nearest integer.
(6) If $n = 0$, then $\text{quantize}'_k(r_{i_\ell}) = \sim\text{quantize}'_k(r_{i_\ell}) = 1$.
(7) if $n \neq 0$, then our outer scale factor of $1/(2^k-1)$ is technically multiplying the result of $(2^k-1)(r_{i_\ell} + n)$, *and thus the scale factors do not cancel.*
(8) Note that the probability of n == 0 is low, and so any individual call to quantize$_k$() is unlikely to result in a case where we can ignore the round() function.
(8) However, under assumption that n is drawn uniformly and at random from [-0.5, 0.5], the expectation $\boxed{\substack{\text{01D}\\\text{404}}}(n) = 0$.
(9) *∴ under expectation $\boxed{\substack{\text{01D}\\\text{404}}}(n) = 0$, $\text{quantize}'_k(r_{i_\ell}) = \sim\text{quantize}'_k(r_{i_\ell}) = 1$.**