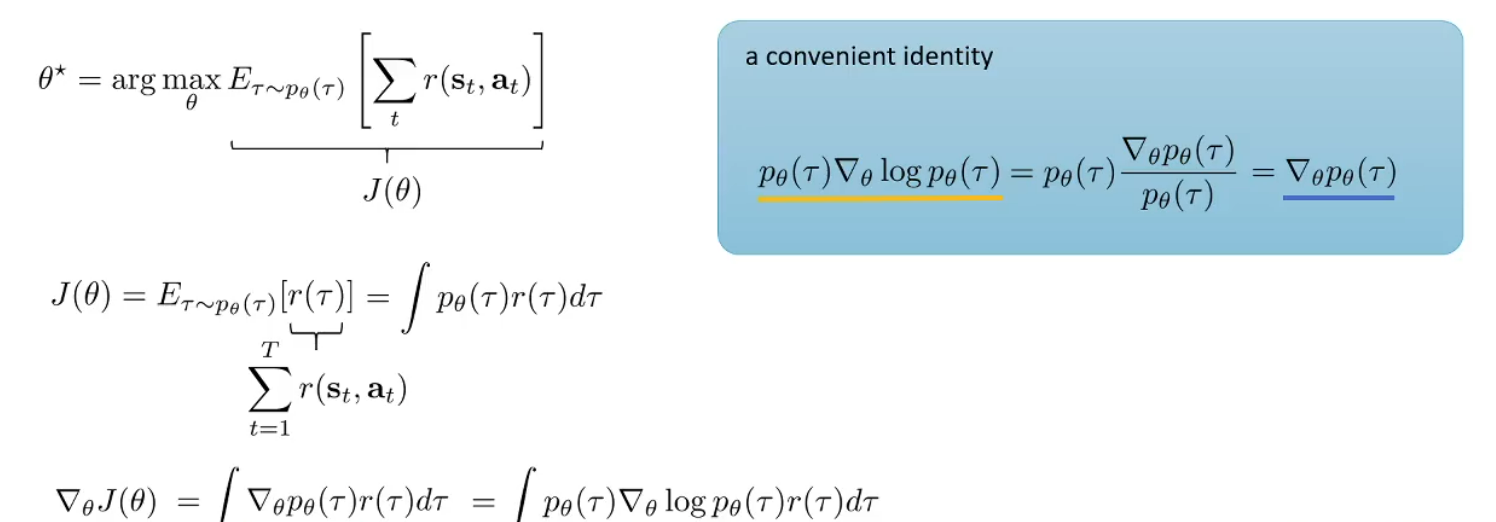
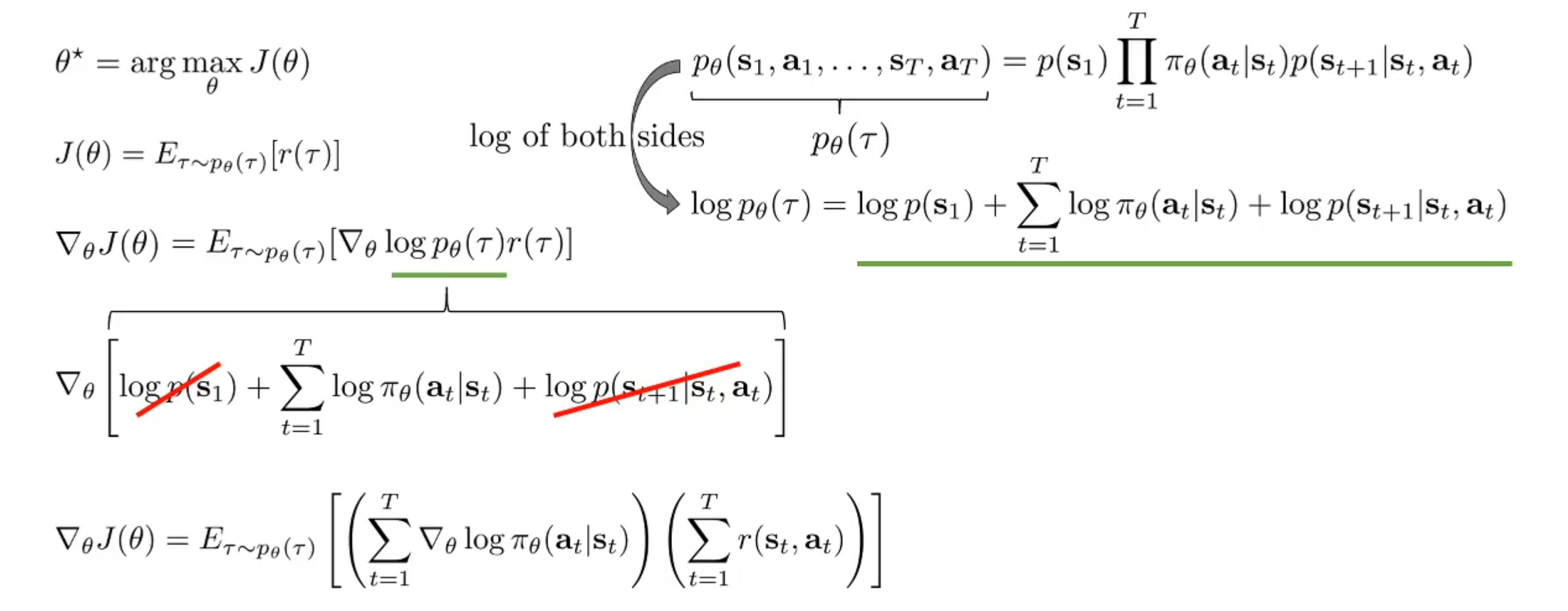
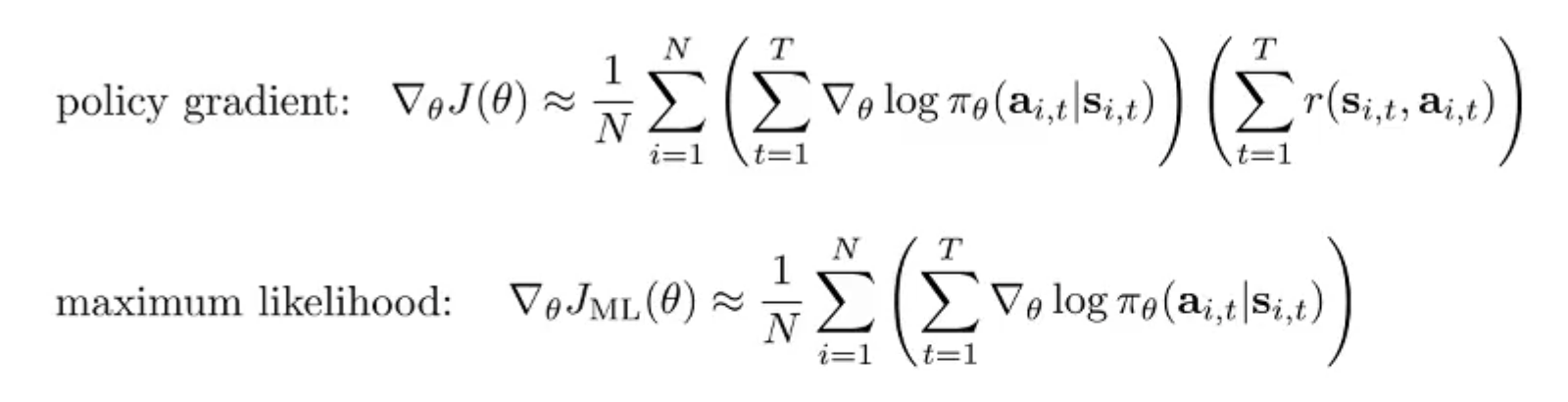
* **Lecture 5 Policy Gradient**

1. The mathematical derivation process

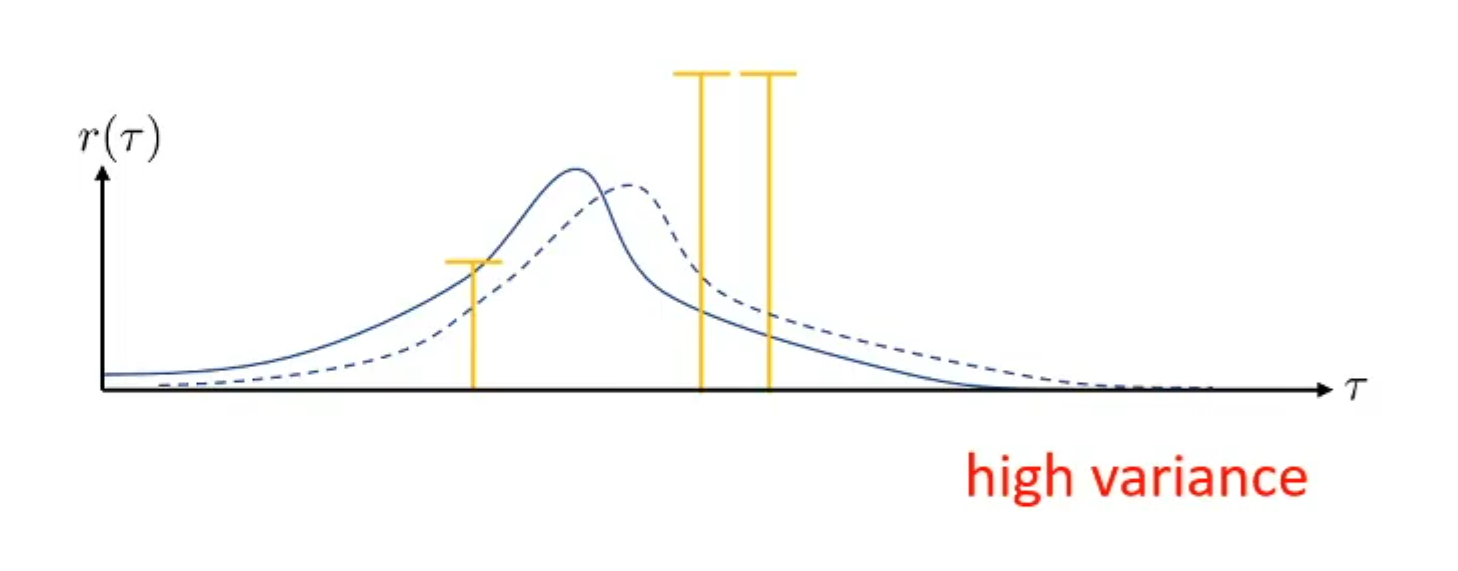




1. It is like the weighted maximum likelihood.

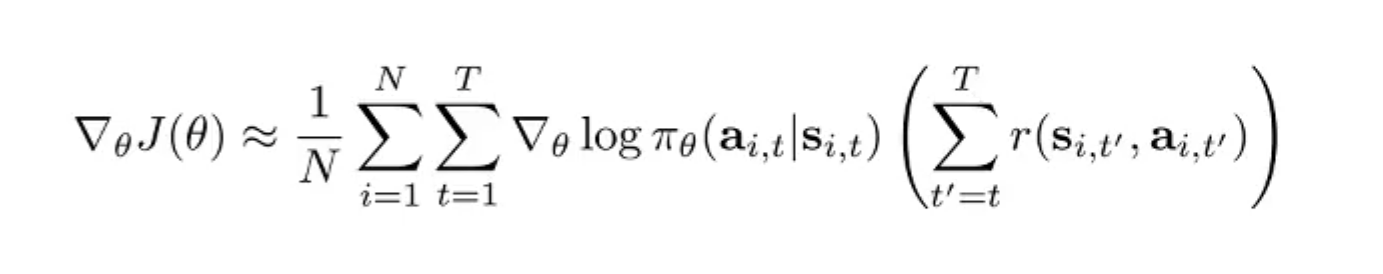


1. High variance problem

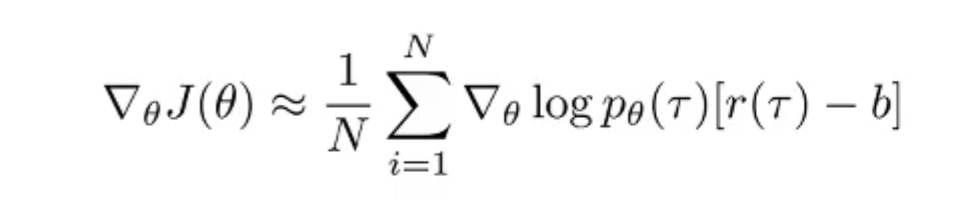


Solutions to high variance problem:

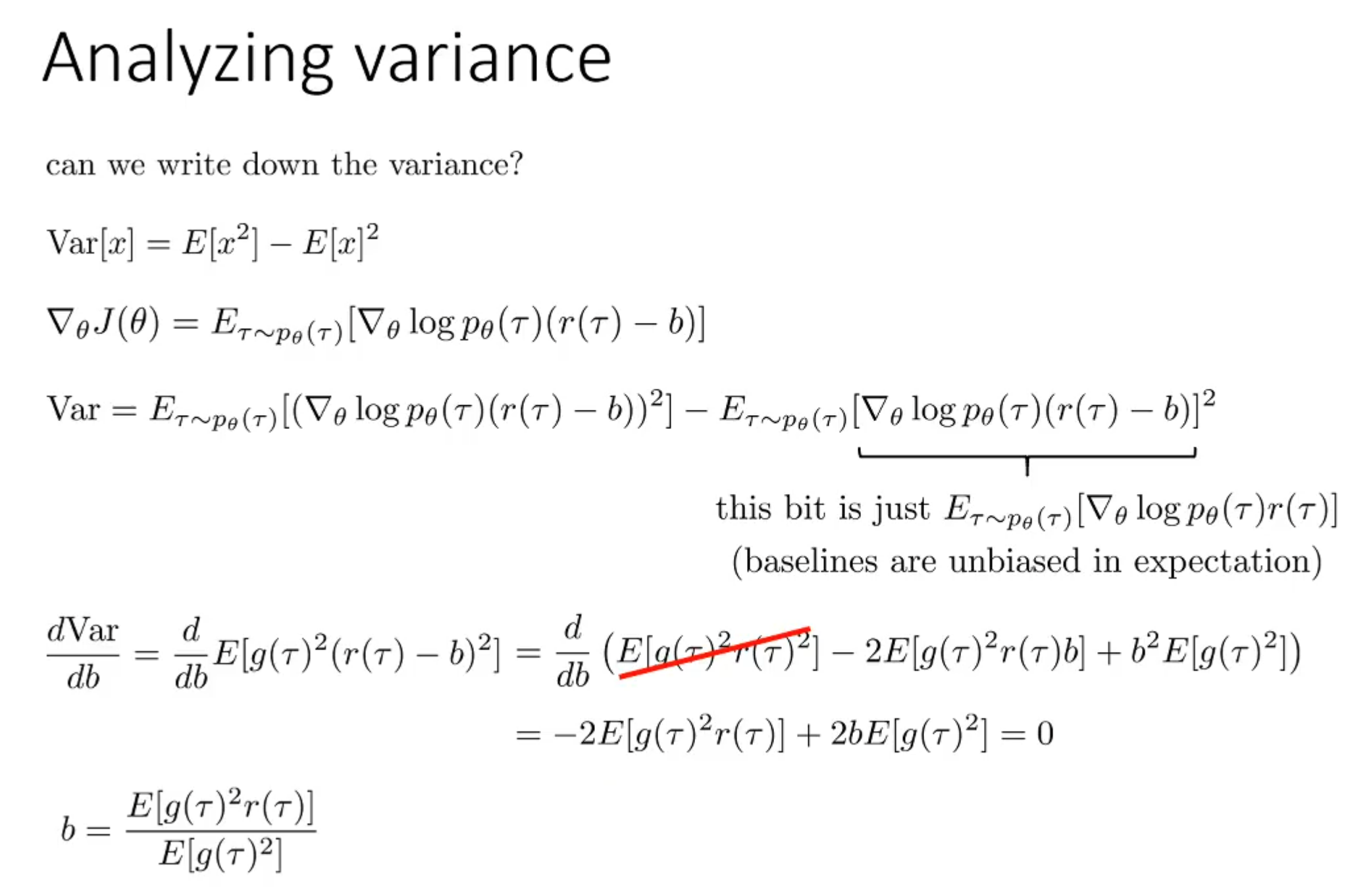
1. Future does not affect the past



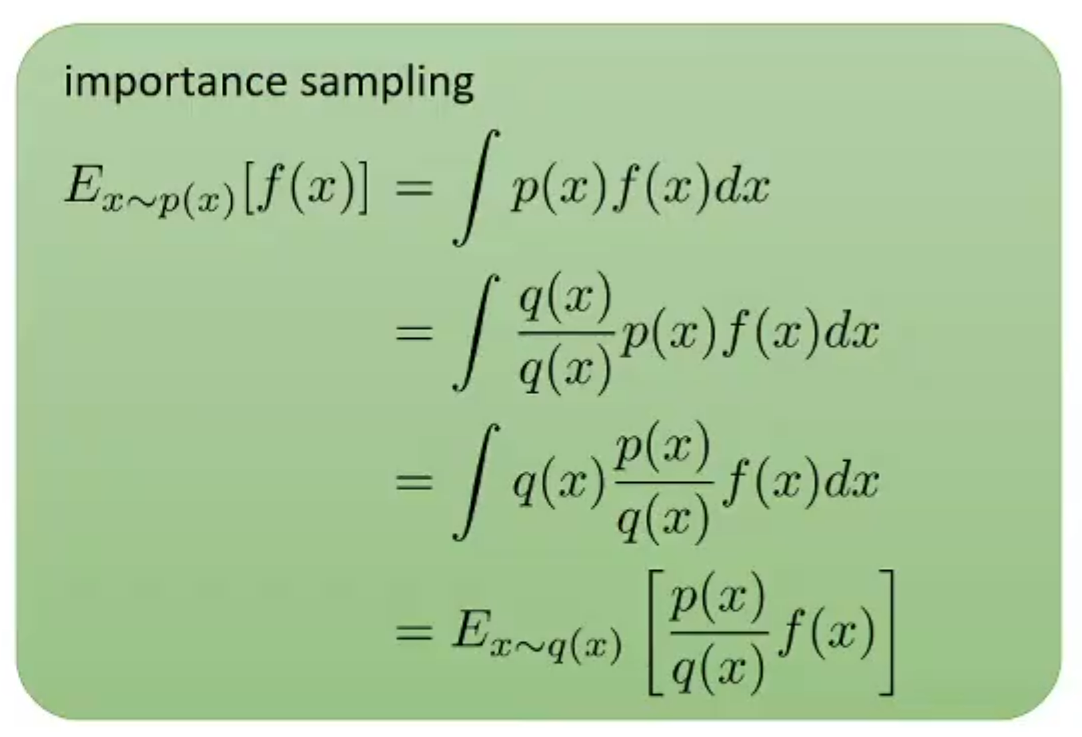
1. Add baselines

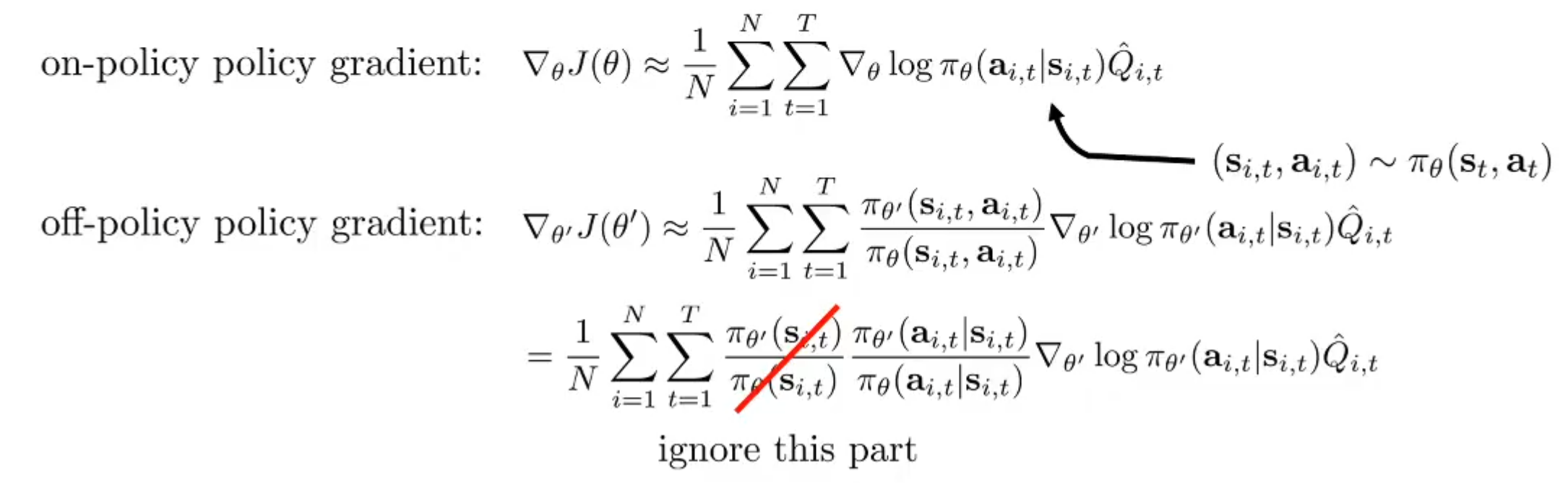


And it is worth to note that adding baselines does not affect the original derivation process, it is an unbiased estimate.



1. From on-policy to off-policy (Important Sampling)





1. An interesting problem at training process

