考虑对于一个状态s，所有状态行动价值函数都相等，即:

 for some 

令任意的值估计，其在整体意义上是无偏的，即满足：



但是这并不完全正确，因此有：

 for some 

其中，表示状态s下动作的数目.

则：



这个下界是紧的.在相同条件下，双Q-learning估计的绝对误差下界为零.

证明如下：

定义每个动作的误差为：



假定存在集合，使得.令为的正向误差(包含等于0的情况)集合，集合大小为n，令为的严格负向误差集合，集合大小为m-n(因此).如果,且 ，这与矛盾.因此.所以，,又因为，因此我们可以得到.这也暗含了.则：



原论文中说此处用到了赫尔德不等式(Hölder inequality)，感觉此处只是运用了**放缩**。

由此我们可以得到:



最后一个不等式的推导用到了，则。

与是矛盾的，因此假设不成立，所以对于满足上述条件的成立下面的不等式:



我们可以检验下界是紧的(能取到)，令for ，，可以验证其满足，并且.

对于Double Q-learning 的紧界是0.这可以有如下推导得到，因为我们有：



并且

 for 

满足定理的条件.

此外如果我们有，那么误差为0.剩余的行动值函数,for ，可以是任意的。