由于在其他条件不变的情况下真实的回报奖励一定随着生存时间的变大而变大。因此对于生存时间是一个分布的情况而言，在使用和均值一样的策略下（即固定batch）其reward一定小于其最大值（或者是99%？）所对应的的固定时间的reward。因此该分布的reward一定位于一个区间之内，该区间的上界对应以最大值为固定生存时间的reward，其下界为其最小值对应的reward。同时可以看出所用均值也位于该区间内，因此均值和实际分布的reward的差值的绝对值小于区间上界和下界之差。因此真实的分布对应的reward小于均值对应的reward加上两界之差并大于均值减去两界之差。即求出该分布对应的reward范围。同时应该可以看出该区间随着生存时间的变化。