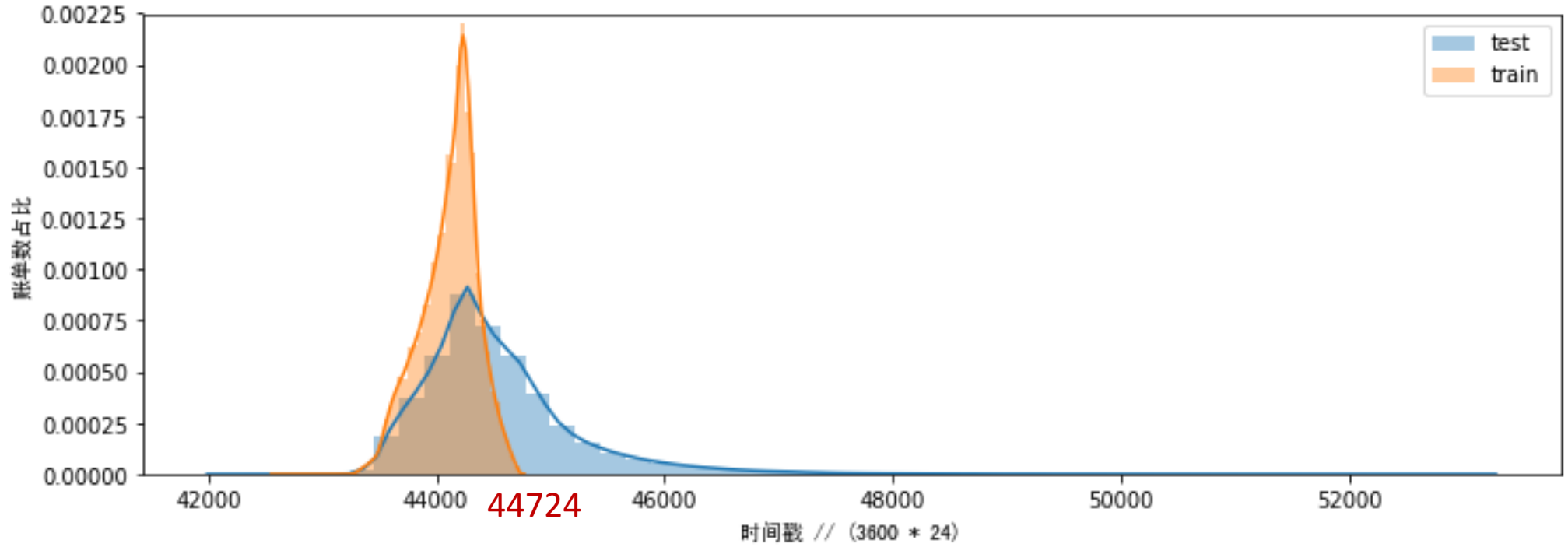
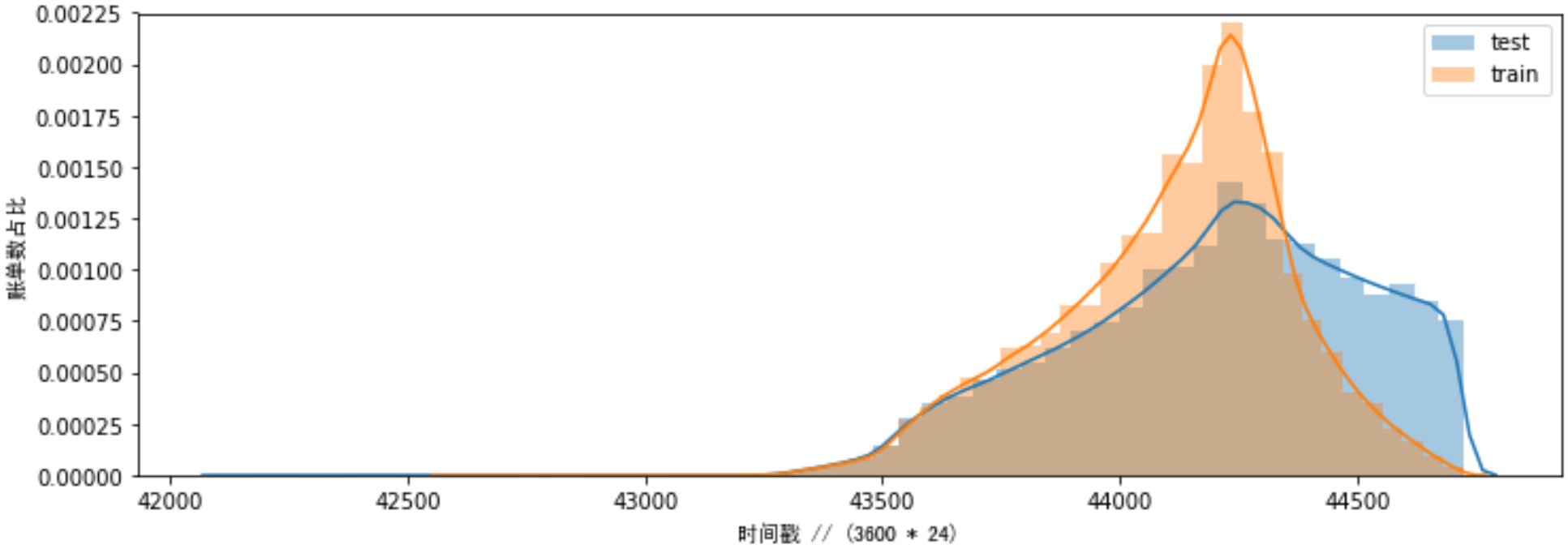
训练测试集对于时间（天）的账单数占比，可以看出**测试集给出账单的时间范围更广**。
区分位置是在第44724天。



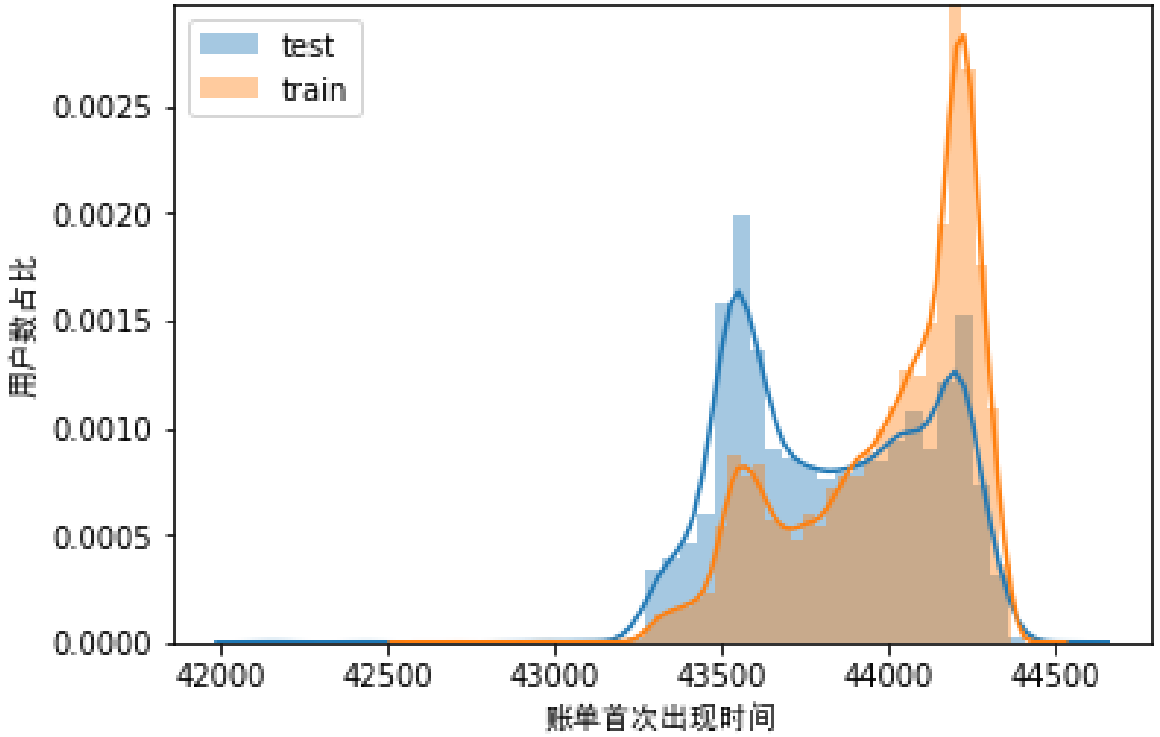
剔除测试集第44724天后的数据分布如下：可以看出直接剔除后，训练测试账单数关于时间的分布还是不一样。这种分布不同应该是因为**训练测试的用户性质不同**。而不仅仅是账单的截断。

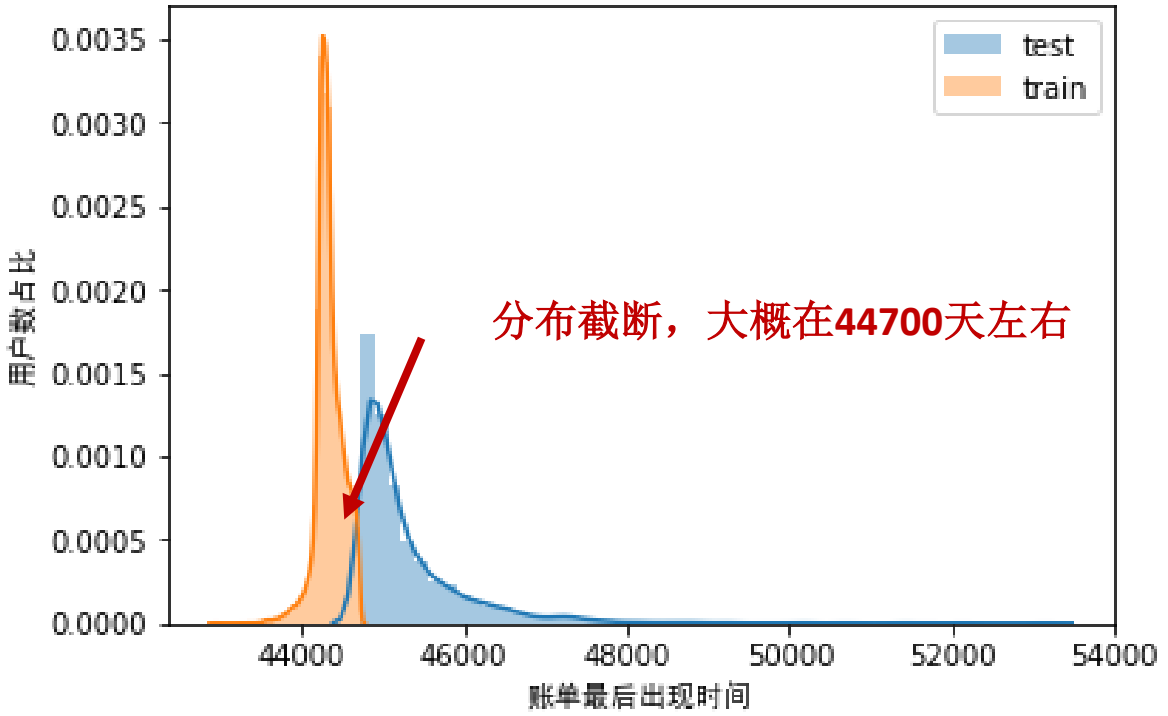
比如说，训练集中的用户 都是相对较老的用户，而测试的用户覆盖了新老用户。



如果用账单首次和最后出现时间来表征**用户加入和流失的时间**，可以看出，所有用户都是在第44500天前加入的，**测试集老用户占比更多**。测试集用户流失时间也相对推后，说明测试集用户是相对更稳定的用户。

另外，训练集用户流失时间分布出现截断。这个位置大概也是44700天左右。所以，**训练集用户是被采样过的，可能是根据流失时间采样的**，所以训练集账单在44724天后就没有了。





可以看出，关于加入时间的用户数分布呈现了双峰。
猜测: 所有用户由两类用户组成，假设为用户组A，组B，则：

1. 组A用户加入时间比组B更早
2. 组A和组B应该由一个可以明显区分的类别，比如通过不同渠道加入的用户，比如通过线上/线下活动加入的用户。
3. 训练测试按照这个类别做了采样。