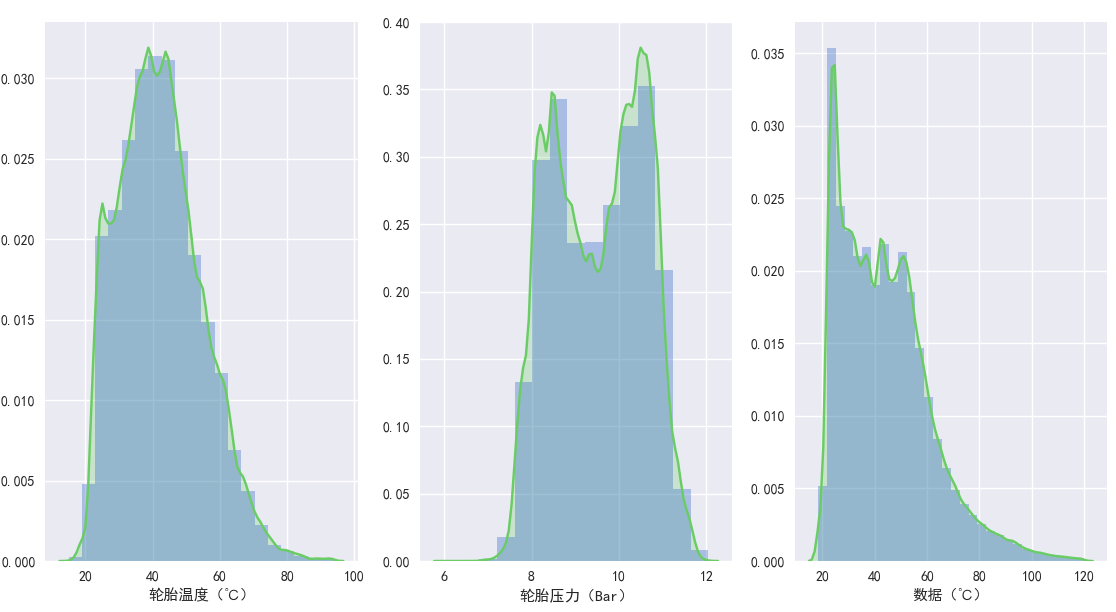
1、总体设计思路：

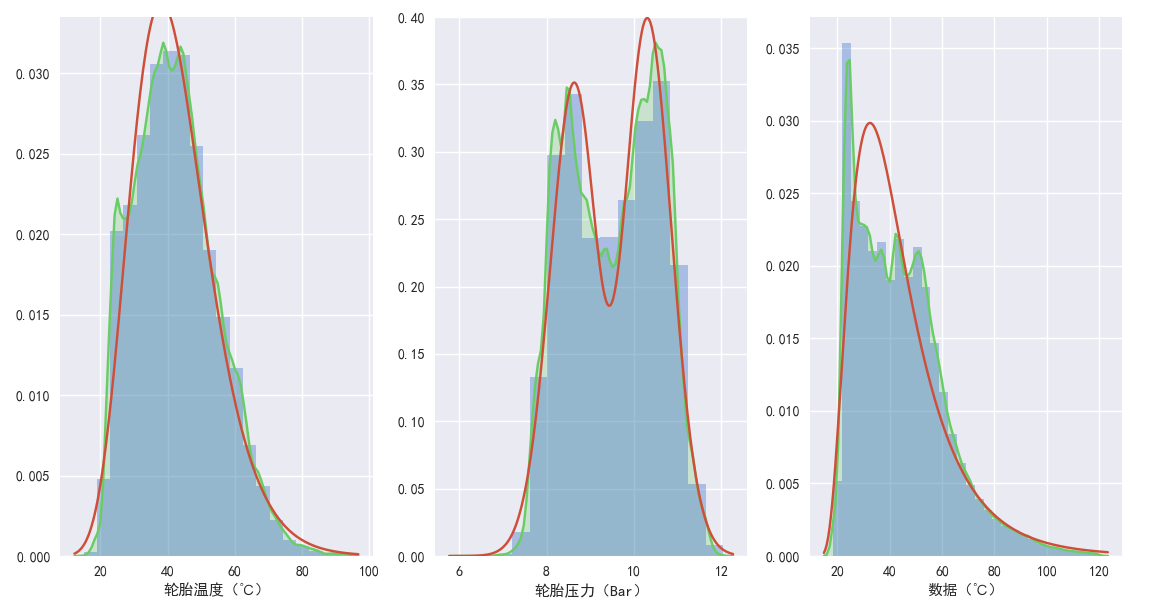
1. 取一部分数据（5.25至6.30日）构建训练数据集。
2. 运用核密度估计法绘制训练数据集的分布。用不同种类的曲线拟合，挑选拟合效果较理想的几种曲线类型，并记录曲线函数式结构参数。
3. 另取一部分数据（7.1-7.15日）构建测试集。
4. 同样用核密度估计法绘制测试数据集的分布，并与（2）中得到的各种拟合曲线对比，观察分析拟合曲线轮廓与测试数据分布是否一致。
5. 选取在训练集、测试集上表现最好的拟合曲线，基于该拟合函数选取阈值。

2、基于训练集获得拟合曲线

数据集包含3类数据（轮胎温度、轮胎压力、轮毂温度），依次对应下图3个子图。蓝色对应直方图，绿色曲线是由核密度估计法得到的概率密度曲线，各自分布如下。



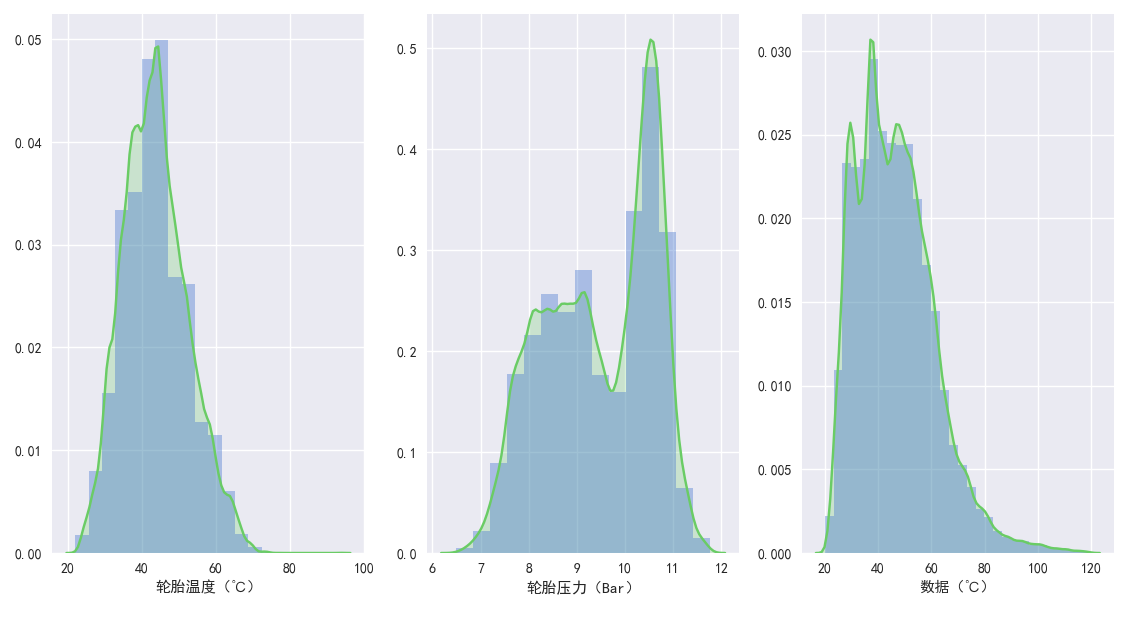
基于数据分布，选用不同类型的曲线拟合训练数据及分布。经过尝试，轮胎温度、轮毂温度数据分布可以用近20种曲线拟合获得理想效果，这里只列举lognorm（对数正态分布曲线）的拟合效果；轮胎压力数据呈现双峰分布，使用两个norm（正态分布）叠加获得较好效果。拟合效果如下。



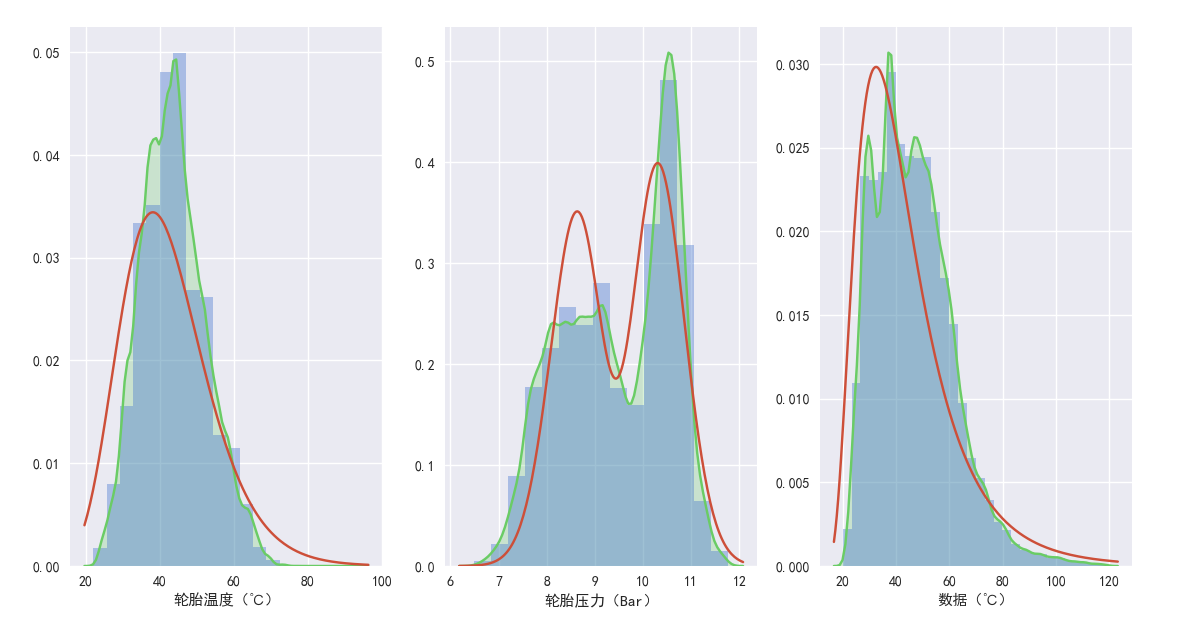
记录下这种拟合函数返回的结构参数。

3、基于测试集验证拟合曲线，选取最优拟合曲线

测试集3种类型数据分布如下。



调用2返回的拟合函数结构参数，对比测试数据集分布与拟合曲线形状是否吻合，效果如下图。经过测试集验证，选取表现最优的拟合曲线类型。



4、选取阈值

如经过步骤3，从多种拟合曲线中选出最优拟合曲线为lognorm（假设轮胎温度、轮毂温度都是），即可基于该拟合函数选取阈值。

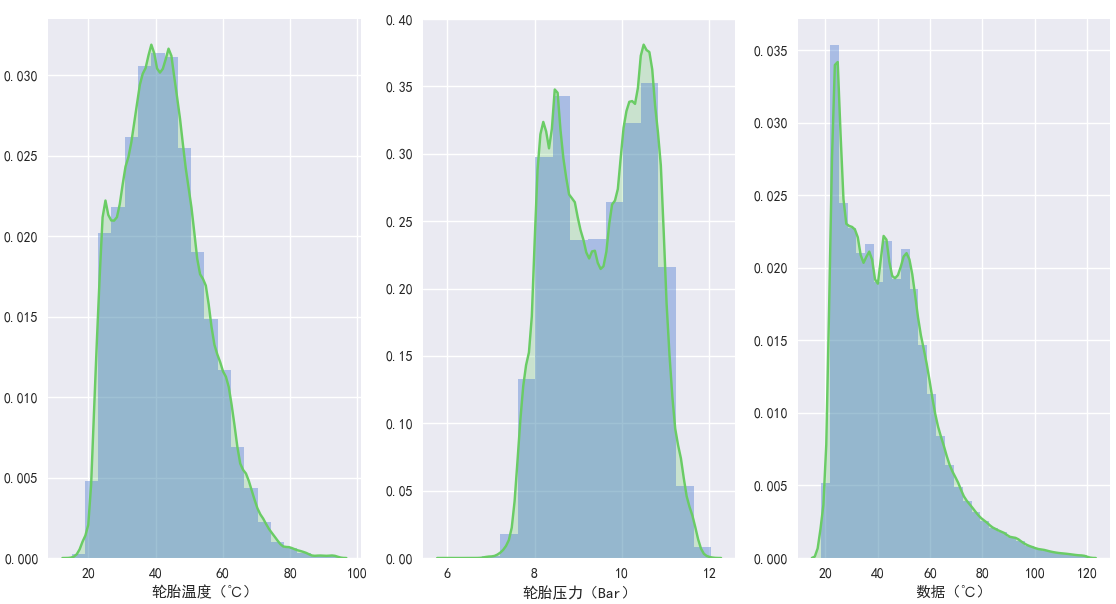
|  |  |  |  |
| --- | --- | --- | --- |
|  | 0.05（低于报警） | 0.5 | 0.95（高于报警） |
| 胎温 | 24.63652658 | 40.87216493 | 64.85362744 |
| 胎压 | 7.90928363 | 9.62444938 | 11.0762791 |
| 毂温 | 22.75479318 | 39.62522014 | 80.35222621 |

以上表中结果因使用的拟合曲线类型而不同，但各方法的结果都比较接近。

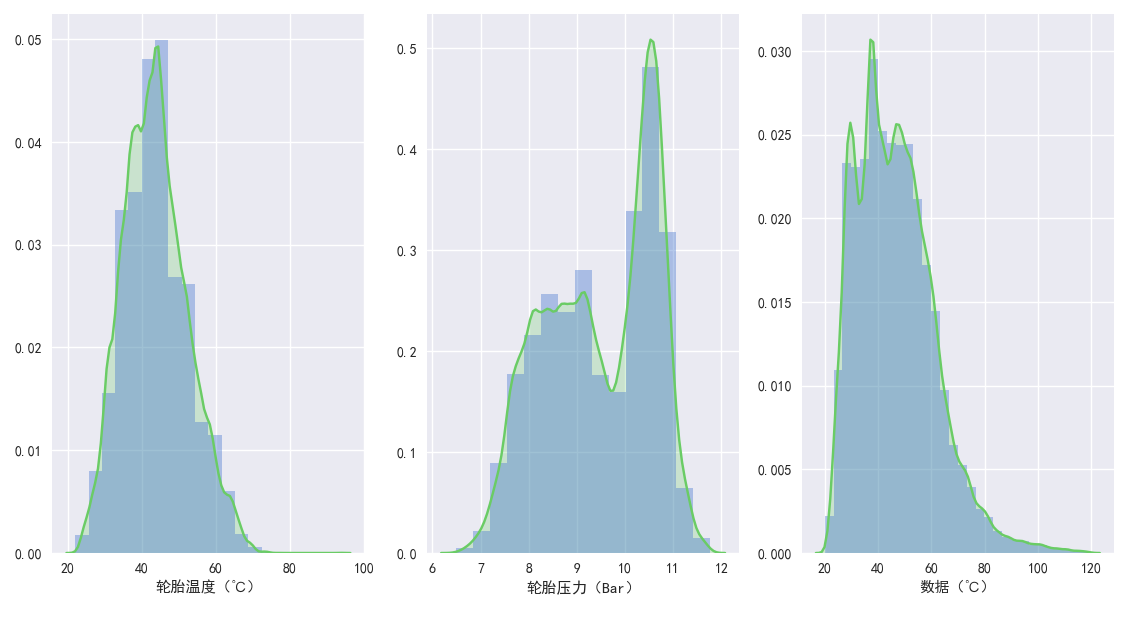
主要问题：

训练数据集与测试数据集分布不一致，轮胎压力数据尤为突出。

（训练集3类数据分布）



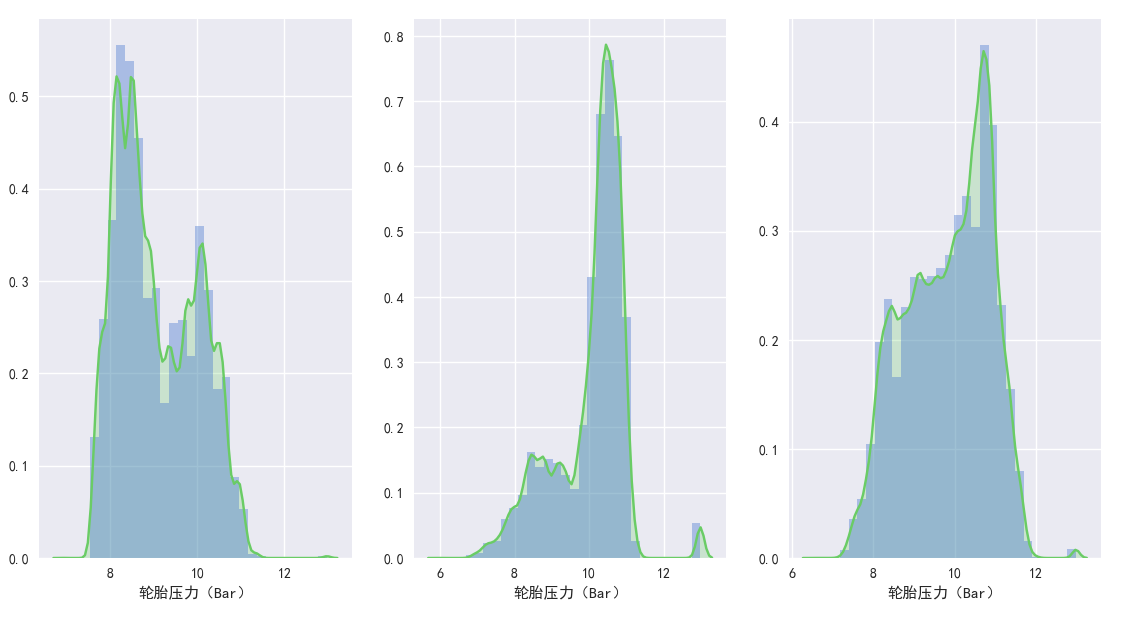
（测试集3类数据分布）



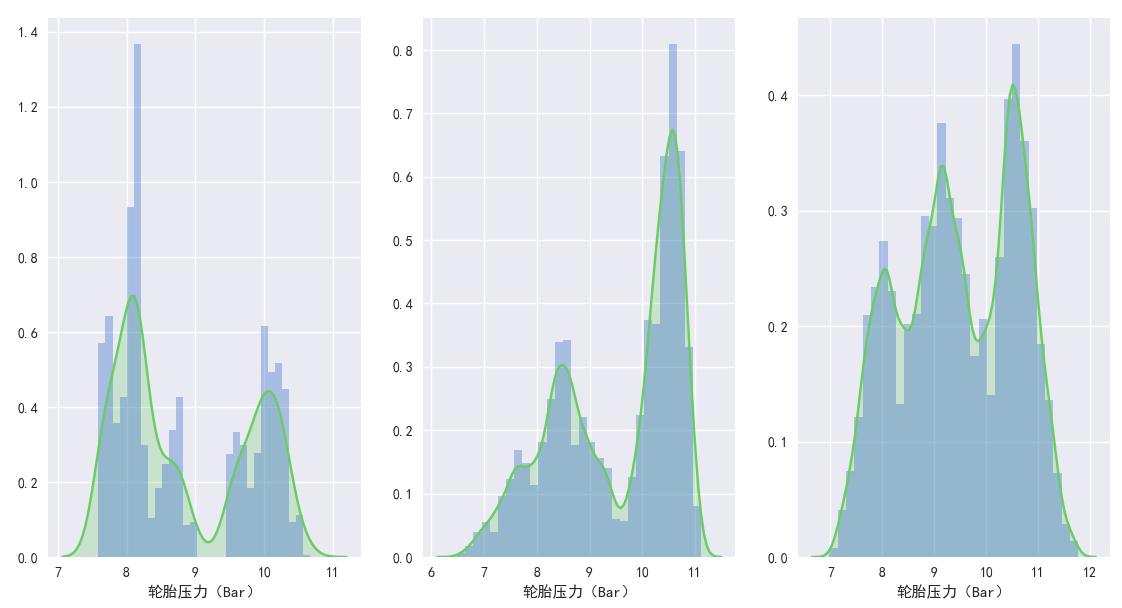
原因分析：

数据丰富性不够，样本和车辆密切相关，尤其是轮胎压力。轮胎压力一般和车辆是否运货直接相关，满载车辆交空车时胎压明显增大。根据实际3辆车的数据情况，有一辆车的胎压数据整体偏大，可能对应时段满载运货；而另外两辆车也都有各自特点。基于此，目前实验数据来源只有这三辆车，可能存在样本不够丰富的问题，整体收个体状态影响较大。根据各辆车的数据分类统计对比，确实存在这样的问题，详细如下。

训练数据集中的三辆车（依次203、207、208）轮胎压力分布，个体差异明显。



测试数据集中的三辆车（依次203、207、208）轮胎压力分布。



不同车辆轮胎压力的详细数值统计如下，绿框对应一辆车。其中测试数据集中的203车数据量非常少，该车可能停运，直接导致两个数据集分布不一致。其他两类数据也存在于车辆相关的情况，相对不明显。

