实验报告

对于不同的主题个数 K ，每个 topic 中频率最高的词：

- $K = 10$ 时，分别为：island, haki, luffy, manga, devil, pirates, grand, crew, luffy, island
- $K = 20$ 时，分别为：island, government, dressrosa, animals, piece, baroque, manga, haki, luffy, sea, navy, series, zou, luffy, devil, captain, crew, piece, grand, island
- $K = 30$ 时，分别为：pose, animals, piece, zou, luffy, devil, crew, luffy, smoker, straw, island, series, luffy, chopper, alabasta, dressrosa, fishman, pirates, luffy, piece, treasure, bur, franky, sea, manga, captain, war, piece, haki, ace
- $K = 50$ 时，分别为：luffy, franky, alabasta, pirates, zou, series, manga, pirates, devil, island, war, luffy, devil, ace, luffy, grand, haki, mountain, luffy, north, straw, sanji, grand, luffy, japanese, pirates, grand, crew, piece, mom, piece, captain, pose, dressrosa, crew, den, red, island, island, fishman, piece, baroque, sea, pirates, crew, manga, luffy, pirates, user, defeat

最佳 K 值

首先考察各个话题频率最高的词汇的重复率，$K = 50$ 时为 $19/50$，$K = 30$ 时为 $5/30$，$K = 20$ 时为 $2/20$，$K = 10$ 时为 $2/10$ 。重复率过高说明各个话题的区分度不够。

其次考察推断时，`inferGamma` 的最大值和次大值，二者相差越大，说明话题越具有区分度：

测试样例	K=10	K=20	K=30	K=50
样例一	11.85, 5.67	11.19, 6.01	10.30, 5.81	7.83, 6.72
样例二	8.65, 6.05	8.09, 6.41	8.03, 5.77	7.81, 5.38

综合来看，$K = 10$ 的效果是最佳的。