

卷33： 概率统计

1. 两个独立的均匀分布的随机变量 $X_1, X_2 \sim U(0, 1)$ ，求 $\max(X_1, X_2) - \min(X_1, X_2)$ 的期望

- ✔ A. 1/3
- ✘ B. 1/6
- ✘ C. 1/9

解答： 等价于求解一个二重积分

$$2 \int_0^1 \int_{x_1}^1 (X_2 - X_1) dX_2 dX_1$$

最终结果为1/3。

2. 我们从一个有100000条样本的数据库中，有放回地随机抽取100000条样本。原数据库中大约有百分几的样本没有被抽中？

- ✘ A. 50%
- ✔ B. 37%
- ✘ C. 0%

解答： 这个问题是bootstrap范围的常考问题。拓展阅读：[自助法\(bootstrap\)的0.632是怎么来的？](#)

3. 有两个盒子（A和B），已知有1/3的概率两个盒子里都有球、1/3的概率只有一个盒子里有球、1/3的概率两个盒子里都没有球。现在你随机打开一个盒子（A或者B），该盒子里有球的概率是多大？

- ✘ A. 2/5
- ✔ B. 1/2
- ✘ C. 无法确定

解答：有四种可能，A和B都没有球，概率为1/3；A和B都有球，概率为1/3；A有球但B没有球，概率为 $x$ ；A没球但B有球，概率为1/3 -  $x$ 。这四种情况下，随便打开一个盒子，并且有球的概率为

$$\frac{1}{3}0 + \frac{1}{3}1 + x \times 0.5 + \left(\frac{1}{3} - x\right) \times 0.5 = 0.5$$

4. 对于一个线性回归模型（最小二乘法，没有正则项），如果对训练集中某个特征做了中心化处理（也就是 $X_i = X_i - \text{Mean}(X_i)$ ），那么和原模型相比，下列哪个说法是正确的？

- ✘ A. 新线性模型回归系数变了
- ✔ B. 新线性模型的截距变了

解答：可以原模型想象为

$$Y = aX + a_0$$

中心化之后的模型为

$$Y = b(X - \bar{X}) + b_0$$

要使得两者等价，必须满足

$$a = b$$

$$a_0 = -\bar{X}b + b_0$$

所以只有截距变了。

5. 有两个盒子（A和B），已知有1/3的概率两个盒子里都有球、1/3的概率只有一个盒子里有球、1/3的概率两个盒子里都没有球。现在你随机打开一个盒子，发现该盒子里有球，那么另外一个盒子里也有球的概率是多大？

- ✔ A. 无法确定
- ✘ B. 1/2
- ✘ C. 2/3

解答：盒子A有球为事件 $A = 1$ ，无球是 $A = 0$ ；盒子B有球为事件 $B = 1$ ，无球是 $B = 0$ 。那么

$$P(A = 0, B = 0) = 1/3, \quad P(A = 1, B = 1) = 1/3$$

以及

$$P(A = 1, B = 0) = x, \quad P(A = 0, B = 1) = 1/3 - x$$

其中 $x$ 为任意一个数。显然 $P(A = 1) = 1/3 + x$ ， $P(B = 1) = 2/3 - x$ 。我们要求解的是

$$\frac{P(A = 1|B = 1) + P(B = 1|A = 1)}{2} = \frac{1}{2} \left( \frac{P(A = 1, B = 1)}{P(A = 1)} + \frac{P(A = 1, B = 1)}{P(B = 1)} \right) = \frac{1}{2} \left( \frac{1}{1/3 + x} + \frac{1}{2/3 - x} \right)$$

显然，结果和 $x$ 的取值有关，所以无法确定。