蒙特卡洛法求误码率的准确度与仿真次数分析

519021911197 杜明俊 F1903404

【摘要】

通信信道中的比特因干扰而发生接收错误的概率，即误码率。通常以蒙特卡洛方法估计误码率，本质上是对 n 值很大的二项分布随机变量进行参数 p 的估计。本文研究了给定待估计参数的取值范围和最大误差，所需最小的蒙特卡洛试验次数。实验过程中利用频率替代法、最大似然估计法进行参数估计，并利用中心极限定律等检验参数区间。该研究结论不仅可以用在蒙特卡洛仿真估计误码率中，更可以运用在一切符合二项分布、泊松分布的随机变量蒙特卡洛法参数估计中，根据实际需求得到需要的最小仿真次数 n。

关键词：置信区间，中心极限定理，蒙特卡洛方法，仿真规模和误差

1. 引言

通信信道中的比特或码元，因为外界干扰而产生失真，在接收端会以一定概率产生接收错误，如发射比特为 1 而接收端判决为比特 0，该概率称为误码率，本文中以p_e表示。误码率由多种信道因素决定，现实环境中用理论推导求解往往会过于复杂，故采用蒙特卡洛法进行p_e的估计，通过在仿真的信道中传输 N 个比特并观察接收错误情况。[1] 蒙特卡洛方法以概率和统计理论方法为基础，能够真实地模拟实际物理过程，得到更符合实际的结果，且免去了复杂推导过程。[2] 蒙特卡洛方法求解精度与参数设定、仿真次数有关。但仿真次数过多会引起巨量的计算消耗，仿真次数过少使待估计参数无法有效收敛到真实值，求解精度降低。因此根据仿真建模的特性，找到给定误差范围内所需最小仿真次数是有必要的。

本文中参数p_e对应随机变量符合二项分布，故本质上是对任何符合二项分布的蒙特卡洛仿真进行探究。实验采用理论分析和 MATLAB 验证，目标是给定p_e范围，估计量$\widehat{p_e}$最大误差和仿真次数 n 数量级，求所需最小的仿真次数 n。

2. 问题建模

蒙特卡洛实验中，传输 n 个比特。传输第 i 个比特时的正确情况记为X_i，其中

$$X_i = \begin{cases} 1，发生错误 \\ 0，未发生错误 \end{cases}$$

$X_1, X_2 \dots X_n$独立同分布(i.i.d)，均满足$p = p_e$的 0-1 分布。传输结束后总错误比特数记为S_n,其中

$$S_n = \sum_{i=1}^{n} X_i$$

$$S_n \sim B(n, p_e)$$

通过已知参数n，S_n观测值k，估计参数p_e为$\widehat{p_e}$.

可以采用两种方法进行点估计，估计p_e:

1) 频率替代法

即$\widehat{p_e}=\frac{k}{n}$,

2) 最大似然估计法

即求

$$\arg\max_{\mathrm{p_e}} P(S_n=k)$$

其中$P(S_n=k)=C_n^k p_e^k(1-p_e)^{n-k}$

令$\frac{\partial}{\partial p_e}P(S_n=k)=0$, *得到*$\widehat{p_e}=\frac{k}{n}$

两种方法得到估计表达式相同。

现在给定p_e为$p_e=p_{e0}$，要满足估计量误差$\delta=\frac{\widehat{p_e}-p_{e0}}{p_{e0}}\le\delta_0$，求仿真需要的样本点数目 n 最小值。

3. 求解最小仿真次数

设 n, k 已知，构造枢轴量U

$$P(a<U<b)=P(p_{e0}-\delta_0 p_{e0}<\widehat{p_e}<p_{e0}+\delta_0 p_{e0})<\alpha$$

其中$\alpha>0.999$, 即有 99.9%以上的把握认为$\widehat{p_e}$误差$\delta\le\delta_0$。

n 的取值范围不同，构造枢轴量方式不同[3]

1) n>>100

根据 De Moivre-Laplace 中心极限定理

$$\lim_{n\to\infty}P\left(\frac{S_n-np_e}{\sqrt{np_e(1-p_e)}}\le x\right)=\Phi(x)$$

$$S_n\overset{近似}{\sim}N\big(np_e,np_e(1-p_e)\big)$$

令$U=\frac{n\widehat{p_e}-np_{e0}}{\sqrt{np_{e0}(1-p_{e0})}}$, *其中*$n\widehat{p_e}=S_n$，枢轴量服从 N(0,1)分布

$$\begin{aligned}P(p_{e0}-\delta_0 p_{e0}<\widehat{p_e}<p_{e0}+\delta_0 p_{e0})&=P\left(\frac{-n\delta_0 p_{e0}}{\sqrt{np_{e0}(1-p_{e0})}}<U<\frac{n\delta_0 p_{e0}}{\sqrt{np_{e0}(1-p_{e0})}}\right)\\&=2\Phi\left(\frac{n\delta_0 p_{e0}}{\sqrt{np_{e0}(1-p_{e0})}}\right)-1\\&=\alpha\end{aligned}$$

根据 α 求出标准正态分布函数值，并查表找到对应的 x。反解 n，得到

$$n=\frac{10.96(1-p_{e0})}{\delta_0^2 p_{e0}}$$

为方便使用，绘制了图 1 供查询不同误差要求、不同p_e下，所需最小仿真次数 n：

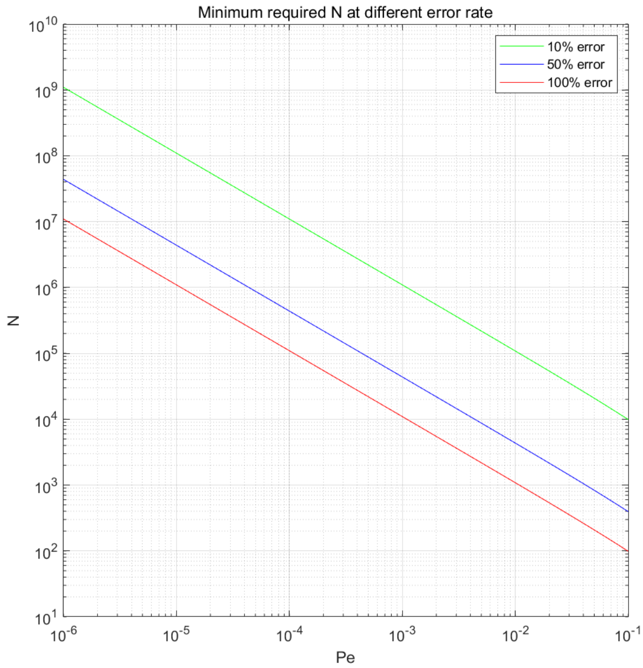

图 1 不同误差和 Pe 下最小仿真次数 N

2) 10<n<100

$$S_n \overset{近似}{\sim} P(np_e)$$

$$P(S_n < k) = \sum_{i=0}^{k} \exp(-\lambda)\frac{\lambda^k}{k!}$$

令$U = n\widehat{p_e} = S_n$，枢轴量服从泊松分布

$$P(p_{e0} - \delta_0 p_{e0} < \widehat{p_e} < p_{e0} + \delta_0 p_{e0})$$
$$= P(n(1-\delta_0)p_{e0} < U < n(1+\delta_0)p_{e0})$$
$$= \alpha$$

实际应用时，先假定一个 n，得到λ。查表可根据λ、$n(1 \pm \delta_0)p_{e0}$对应 k 值找到两个 F(k)，并相减。如果不符合α，改变 n 并继续上述过程，直到找到一组参数使得 F(k)之差最接近α。

3) n<10

$$S_n \sim B(n, p_e)$$

$$P(S_n < k) = \sum_{i=0}^{k} C_n^k p_e^k (1-p_e)^{n-k}$$

令$U = n\widehat{p_e} = S_n$，枢轴量服从二项分布

$$P(p_{e0} - \delta_0 p_{e0} < \widehat{p_e} < p_{e0} + \delta_0 p_{e0})$$
$$= P(n(1-\delta_0)p_{e0} < U < n(1+\delta_0)p_{e0})$$
$$= \sum_{i=0}^{k_1} C_n^{k_1} p_{e0}^{k_1} (1-p_{e0})^{n-k_1} - \sum_{i=0}^{k_0} C_n^{k_0} p_{e0}^{k_0} (1-p_{e0})^{n-k_0}$$
$$= \alpha$$

其中$k_1 = n(1+\delta_0)p_{e0}$, $k_0 = n(1-\delta_0)p_{e0}$. 实际计算时函数过于复杂，可以考虑牛顿切线法求根。

结论：

蒙特卡洛法求解p_e,所需仿真次数 n：

- 先根据其他已知条件，或采用方法 1)进行p_e范围判断，至少确定数量级。
- 无论p_e多大，允许 n>>100，均可采用正态分布近似计算，n 取值参考 1)
- 目标p_e预计<0.01：必须采用正态分布近似计算，n 取值参考 1)
- 要求 n<100，且$0.1 < np_e < 20$：采用泊松分布近似计算，n 取值参考 2)．其中后者应该用$(1-p_e)$代替p_e进行计算
- 目标$0.1 < p_e < 0.9$，要求 n<10：采用二项分布直接计算，n 取值参考 3)
- 当要求 n<100 时，应考虑减小置信度α

4. MATLAB 实验验证

在 matlab 中生成符合二项分布的 0,1 序列，其中每一位符合$p = p_e$的 0-1 分布，序列的长度代表仿真次数。最终求序列的均值作为$\widehat{p_e}$.进行如下两组实验，验证先前结论

- 实验一：设置$p_e = 1 \times 10^{-5}$，通常误码率≤这个数量级。要求仿真得到$\widehat{p_e}$误差 10%以内

根据结论，必须用方法 1）.计算得到 n=109,598,904。设置 20 组仿真，每组取 11e7 个样本点，即每个序列长度 n=11e7。观察 20 组得到的$\widehat{p_e}$分布。

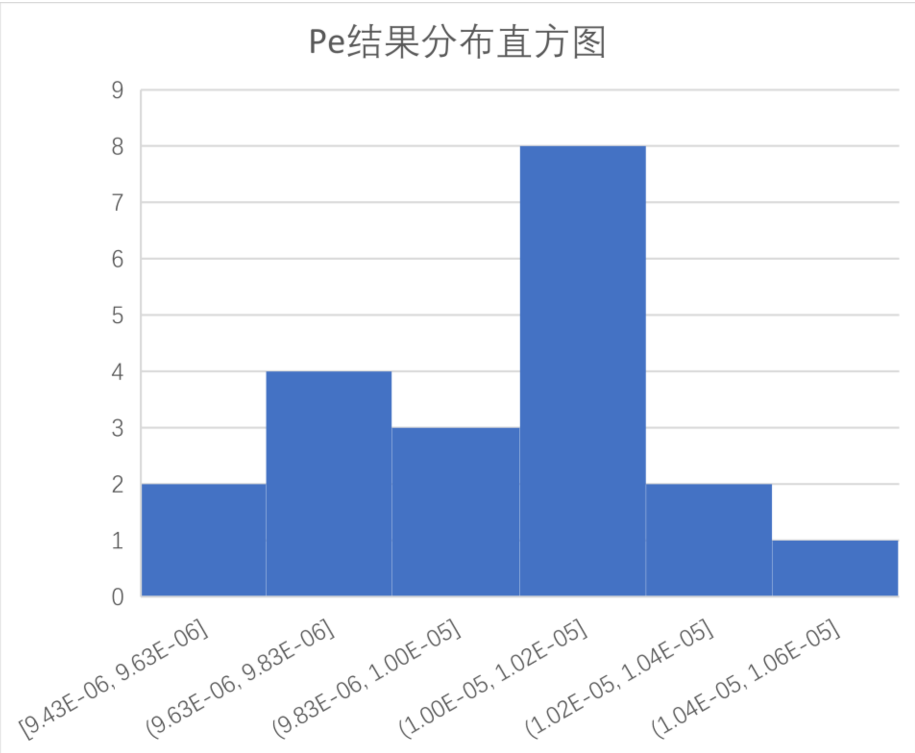

图 2 $\widehat{p}_e$结果分布直方图

分析结果，可以得到全部结果误差在 10%内。

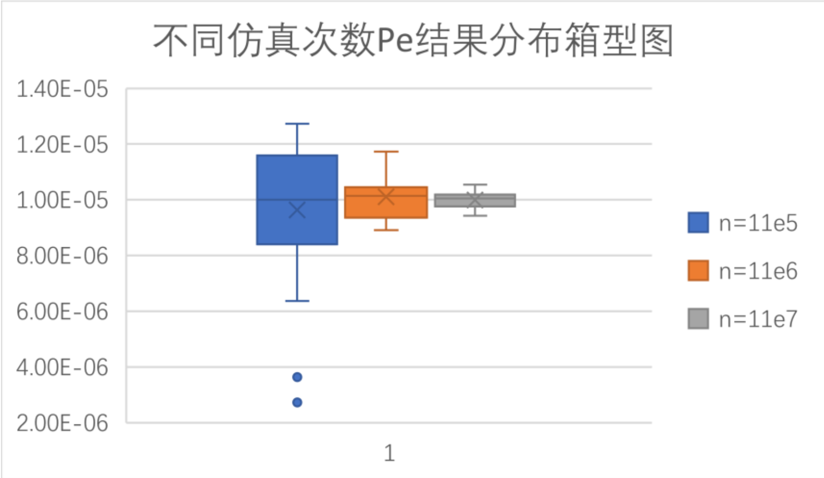

图 3 不同仿真次数 Pe 结果分布箱型图

改变仿真次数数量级，可以看到只有仿真次数≥11e7，才能完全保证 10%以内误差

- 实验二：设置$p_e = 0.2$，要求仿真得到$\widehat{p_e}$误差 10%以内

根据结论用方法 1）计算得到 n=4384

n<100 时，收敛到 10%误差可能性很小。但是可以计算不同 n 误差在 10%之内的置信度，并取置信度最大的 n。方法 2）计算得到

表 1 不同仿真次数下结果落在误差范围内置信度

仿真次数	置信度
100	0.3392
90	0.3556
80	0.2918
70	0.2050

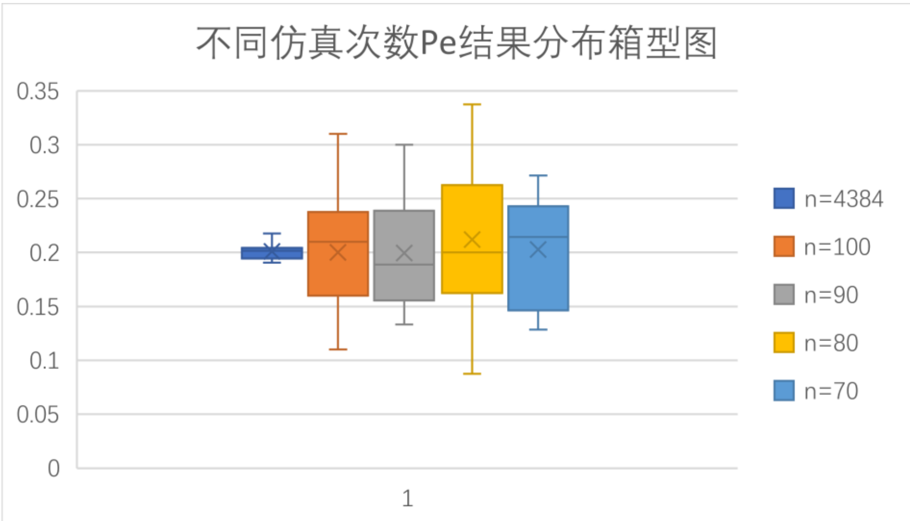

图 4 不同仿真次数 Pe 结果分布箱型图

分析得到，n=90 时样本点向 p=0.2 聚合程度更高。因此在限制了仿真次数的条件下，可以根据上述结论选择 n=90，可以获得相对更可靠的结果。

5. 结果分析与推广

以上研究了蒙特卡洛方法求误码率保证结果在一定误差范围内所需的最小仿真规模 n，并将结论推广到一切符合二项分布的情况。实验验证了上述结论的正确性。实际仿真时，计算机性能、仿真复杂度可能会限制仿真次数。此时可以根据本文结论，在 n=10-100 区间内根据 2) 3)找到最优仿真次数 n。

除了求复杂信道下的误码率，其他抽样检测场景同样也可运用本文结论。例如生产一批产品，产品次品率为固定值 p，通过抽样检测方式估计该值。但抽样检测的样本数不能过多，尤其是对一些小规模、一次性产品而言。根据本文的 2) 3)结论，可以找到最合适的抽样个数 n，来减小次品率估计值和真值之间的误差。

6. 引用列表

[1] Weiyao Lin, Principles of Communications(PPT slide)

[2] 百度百科.蒙特卡罗模拟 [EB/OL]https://baike.baidu.com/item/%E8%92%99%E7%89%B9%E5%8D%A1%E7%BD%97%E6%A8%A1%E6%8B%9F/5160083

[3] 卫淑芝,熊德文,皮玲等. 概率论与数理统计[M]北京:高等教育出版社,2020:158-159