高速串行链路误码率测试置信区间估计工程化分析方法

1、为什么误码率测试需要引入置信度作为衡量指标？

假设某高速串行链路系统真实的误码率 BER 为 p，但是在事实上，这个真实误码率参数 p 我们是事先无法准确预知的，也许永远无法准确获知客观的误码率 p。唯一的办法，通过观测高速串行链路系统一定数量的测试比特样本，来推测实际系统的误码率 BER 参数 p。该问题类似于，一个非常巨大的袋子里有数量庞大但不详的红色球与白色球，如果我们想知道从袋子中拿球能够选中红球的概率 P，因此，我们唯一观测手段就是从袋子中取得 N 个球，观察红球的个数 r，从这个观测数据来推测袋子中红球的概率 P。**这就是贝叶斯统计分析的参数估计，因为参数也是通过观测样本估计获得，由此获得的参数自身也就存在分布概率，也就是置信度。**

2、NRZ 误码率置信度计算方法

误码率测试，目前常用的方法，可以参考参考资料：

Calculating statistical confidence levels for error-probability estimatest, Justin Redd, Maxim,
http://www.lightwaveonline.com/articles/print/volume-17/issue-5/features/calculating-statistical-confidence-levels-for-error-probability-estimates-53462167.html

其原理大致如此，比特误码遵守二项式分布，可以得到

$$\mathrm{P}_n(k)=\binom{n}{k}p^k(1-p)^{n-k}=\frac{n!}{k!\,(n-k)!}p^k(1-p)^{n-k}$$

其中，n 为传输比特个数，k 为发生误码个数，p 为误码率

由于二项式公式有阶乘公式，计算比较复杂，在后续求解置信度的时候公式也难以化简。当 n 数量很大，p 数量比较小的时候，可以近似为泊松分布

近似泊松分布推导过程如下，

根据$1+\mathrm{x}<\mathrm{e}^x$,当 x≤0 时，x 趋向于 0,1+0=1=$e^0$=1

令 p=λ/n，n 趋于无穷时$\mathrm{P}_n(k)$的极限

$$\begin{aligned}\lim_{\mathrm{n}\to\infty}P(X=k)&=\lim_{\mathrm{n}\to\infty}\binom{n}{k}p^k(1-p)^{n-k}=\lim_{\mathrm{n}\to\infty}\frac{n!}{k!\,(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\&=\lim_{\mathrm{n}\to\infty}\left[\frac{n!}{n^k(n-k)!}\right]\left(\frac{\lambda^k}{k!}\right)\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}\\&=\lim_{\mathrm{n}\to\infty}\left[(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})\right]\left(\frac{\lambda^k}{k!}\right)\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}\end{aligned}$$

由于：

$$\lim_{\mathrm{n}\to\infty}\left[(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})\right]=1$$

根据$1+\mathrm{x}<\mathrm{e}^x$，当 x≤0 时，x 趋向于 0,1+0=1=$e^0$=1

$$\lim_{\mathrm{n}\to\infty}\left(1-\frac{\lambda}{n}\right)^n=e^{-\lambda}$$

$$\lim_{\mathrm{n}\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=1$$

所以，最终可以获得：

$$\lim_{n\to\infty} P(X=k)=\left(\frac{\lambda^k}{k!}\right)e^{-\lambda}$$

误码率测试，置信度测试样本最终公式为：

$$n=\frac{1}{BER}\left[-\ln(1-CL)+\ln(\sum_{k=0}^{E}\frac{(n*BER)^k}{k!})\right]$$

其中，n 为测试样本个数，E 为测试样本所允许发生最大的误码个数，CL 为置信度，BER 为测试误码率要求。通常，我们假设误码个数 E 为 0，避免求解右边式中$\ln(\sum_{k=0}^{E}\frac{(n*BER)^k}{k!})$的 n。

对于 NRZ 测试，通常是可以实现误码个数 E 为 0，因为 SNR 比较高，误码率量级很低，可以实现在置信度要求的测试时间内，没有任何误码 E=0 的观测样本。例如：对于 10G 速率信号，BER 要求是 1e-12，置信度要求 99.9%，那么测试时间要求为 690.78 秒，大约 11.5 分钟。

3、PAM4 误码率测试置信度计算

3.1 PAM4 误码率置信度计算方法

然而，对于 56G PAM4 测试来说，相对 NRZ 来说 TX SNR 损失了 9dB，在接收端通常 SNR 比较低，正常工作链路的误码率通常落在[1e-4~1e-9]区间。这对于泊松分布模型推导的置信度测试样本来说，是一个很坏的消息。以 56Gbps 速率、链路误码率取最好的数值 1e-9，测试 1 秒平均产生错误样本 E=56，这时无法忽略右侧$\ln(\sum_{k=0}^{E}\frac{(n*BER)^k}{k!})$。要求解 n 的话，测试 1s 时间就必须求解关于 n 的 56 次方程。随着测试时间增加，位置变量 n 阶数也变得更高，求解 n 会变得非常复杂、繁琐，该方法变得不适合工程应用。

不过，从错误样本数增多来看，总测试样本数量应当可以减少一些就能符合相同的置信度要求，但是没有定量分析的数据支撑，定性判断永远只能作为趋势性结论，也无法获得测试结果的置信区间。

为了解决上述方程 n 求解问题，需要重新调整解决该问题的思路。根据中心极限定理，满足相同分布的大量随机样本，其最终趋近高斯分布模型。

$$\mathrm{Gaussian}(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-u}{\sigma}\right)^2}$$

对于每个误码发生概率都满足二项分布，假设误码之间不相关，误码个数 k 的期望值为 np，由于样本数量 n 为一常数，那么 r/n 的期望值为误码发生概率 p。因此，对参数 p 的估计是无偏估计。

那么，我们需要知道参数 p 的标准差，如下式表示

$$\sigma=\sqrt{\frac{p(1-p)}{n}}$$

其中，$p=\frac{k}{n}$作为输入

因此，对应 N%双置信区间，可以采用如下公式

$$\mathrm{BER} = \frac{k}{n} \pm Q * \sqrt{\frac{\frac{k}{n}(1-\frac{k}{n})}{n}}$$

$$\mathrm{Q} = \sqrt{2} * erfinv(N\%)$$

因此，对于 56G 速率 PAM4 信号，链路 RAW BER（FEC 之前）测试结果为 1e-6，那么需要检验一下 5 分钟测试结果，其 BER 测试结果落在 99.9%置信区间的精度范围 1e-6 ± 8.03e-10。当测试结果为 1e-9，测试 5 分钟，置信度要求为 99.9%置信区间的精度范围 1e-9 ± 2.54e-11。其精度结果，依然可以接受。

从测试结果来看，PAM4 由于误码率要求比较宽泛，所以测试时间可以大大压缩，测试置信区间依然可以接受。

3.2 考虑突发误码条件下 PAM4 误码率置信度修正计算方法

以上分析方法，二项分布模型条件，假设每个误码事件本身是独立的。然而，由于 DFE 导致的突发误码特征，使得第一个比特发生误码之后，后续比特码流发生误码事件与第一个比特发生误码事件是相关的。

假设，观测到 k 个误码，其中总共有 M 次突发误码，第 i 个突发误码集合发生概率为 $Pb_{rll,i}$,设 r_i 为第 i 个突发误码的等效错误比特个数：

$$(p_r)^{r_i} = P_{rll,i}\ , \mathrm{i} = 1\mathrm{toM}$$

$$r_i = \frac{\ln(P_{rll,i})}{\ln(p_r)}, \mathrm{i} = 1\mathrm{toM}$$

其中，$p_r = \frac{M}{n}$，表示突发误码第一个比特发生的概率。

$$\mathrm{p} = \sum_{i=1}^{M} \frac{r_i}{n}$$

当采用 ONE-TAP DFE 突发误码公式时，

$$P_{rll,i} = P_r * (P_b)^{rll_i - 1}$$

其中，P_b为突发误码传递概率，因此

$$r_i = (rll_i - 1) * \frac{\ln(P_b)}{\ln(p_r)} + 1, \mathrm{i} = 1\mathrm{toM}$$

但是，根据上式细化分析突发误码情况是有一定的难度的，需要事先获知突发误码次数 M，以及每比特误码传递概率，确定统计意义上的各个突发误码长度内每个突发误码的平均个数。在测试环节，以目前 SerDes 的 BIST 测试功能要具体获知这些信息，存在很大困难。

由于 3.1 分析的置信度区间量级精度都比较高，可以考虑采用简化 WORST CASE 处理方法，把每次突发误码可以只看成一个比特错误，突发误码后续比特误码递推概率P_b=1（实际上不可能，PAM4 最大传递概率为 0.75），也就是和第一个比特误码完全相关，第 i 个突发误码的等效错误比特个数 r_i=1，p 公式简化为：

$$\mathrm{p} = \frac{M}{n}$$

当假设高速链路突发误码平均长度为 100，那么，当忽略后面错误比特的数量对 BER 贡献，相同测试时间内，PAM4 链路误码率 p 将由[1e-4 1e-9]区间降低到[1e-6 1e-11]。含突发误码 1e-4 误码率置信区间对应无突发误码 1e-6 置信区间，因此等效精度为 1e-6± 8.03e-10。

含突发误码 1e-7 误码率置信区间对应无突发误码 1e-9 置信区间，1e-9 ± 2.54e-11。

在相同条件下，当测试结果为 1e-9，测试 5 分钟，含突发误码 1e-9 误码率置信区间对应无突发误码 1e-11 置信区间的为 1e-11 ± 2.54e-12。其置信误差比较接近链路原始误码率，有点大，可以适当延长测试时间，例如测试 10 个小时，置信度要求为 99.9%置信区间的等效精度范围为 1e-11 ± 2.32e-13。

因此，PAM4 高速链路在考虑突发误码情况下，有如下两点结论：

1、 如果原始链路误码率落在[1e-4 1e-7]范围，5 分钟测试置信区间范围还是满足 99.9%置信度应用要求的。

2、 如果原始链路误码率落在[1e-7 1e-9]误码范围，需要适当加长测试时间以保证测试精度。当然，PAM4 链路原始误码率落在[1e-7 1e-9]误码范围，PAM4 链路来说 SNR 已经非常高了，该链路通常也不属于系统关键链路（误码率临界链路），即使测试时间不足影响了测试精度，通常也不会影响系统高速链路可行性的判断结论。