arXiv:2301.02149v1 [q-bio.QM] 5 Jan 2023

DNA定量的随机性

Abdoelnaser M Degoot 和 Wilfred Ndifon*非洲数学科学研究所、下一代爱因斯坦倡议，卢旺达

2023年1月6日

1 摘要

DNA定量的一种常见方法是反复进行DNA扩增循环。聚合酶链式反应（PCR）采用这种方法，但其输出会受到扩增噪声的干扰，因此难以准确反推输入DNA的量。常见的数学方法未能充分考虑这种噪声，因而容易产生误差。本文建立了一个简约的数学模型，用随机映射描述输入DNA如何产生实验输出，并以自然的方式纳入扩增噪声。我们利用该模型推导出定量循环的概率密度；定量循环是一项经常报告的实验输出，其概率密度可通过拟合数据来估计输入DNA。值得注意的是，模型预测，只有一个输入DNA分子的样本，其检测呈阳性的概率为 <4%，比标准PCR数据解读方法所假定的低 >25倍。我们提供了计算检出限和定量限的公式。这两项是DNA定量方法的重要性能指标，通常通过临时数学技术进行评估。我们的结果为严格分析DNA定量奠定了数学基础。

2 引言

基因组靶标的定量在生物学、生物技术和医学的众多应用中都备受关注，从确定个体的疾病状态，到检测基因表达谱在空间和时间上发生的细微变化（例如[1, 2]）。通常，这一过程通过将非 DNA 基因组靶标转化为 DNA 来实现，随后扩增 DNA 以进行定量。原则上，这使得即使数量很少的基因组靶标也

*通信地址：wndifon@aims.ac.za

能够得到准确测量。然而在实践中，DNA 扩增过程具有随机性，会产生含有噪声的结果。因此，准确测量需要对这种噪声有充分的定量认识。迄今为止，要做到这一点仍然颇具挑战。

一种具体且非常流行的 DNA 定量方法是实时聚合酶链式反应（PCR） [3, 4]。在 PCR 中，DNA分子以循环方式反复扩增。随着扩增进行，荧光标记核苷酸被掺入新形成的 DNA 分子中，使发出的总荧光增强。根据由此产生的荧光曲线，可以确定定量循环（记作 Cq 或Ct 值）；在该循环中，分子数量超过一个称为定量阈值的设定阈值。如果某个 PCR 反应的Ct 值小于或等于最大可能循环数，则该反应被视为阳性。尽管Ct 值只是对输入DNA分子数量的间接读数，但它往往是 PCR 实验唯一报告的结果。常规PCR的一种变体称为数字PCR [5]，它利用阳性反应所占的比例来估计输入DNA分子的数量。为此，该方法假设当且仅当反应中至少含有一个目标分子时，该反应才呈阳性。目前尚不清楚这一假设在什么条件下成立，以及何时必须舍弃该假设，转而采用更现实的替代方案。

本文介绍一种简约的数学模型，可用于分析 DNA 定量过程，并指导对实验输出的解读。我们以 PCR 为例，但我们的分析也适用于其他方法，例如 DNA 环介导等温扩增 [6]。实验表明，PCR 过程包含不同阶段，各阶段的 DNA 扩增效率各不相同。因此，我们构建了一个包含任意数量阶段的 PCR 过程数学模型，每个阶段都有各自的扩增效率。我们利用该模型得出以下结果：

- 我们推导出在任意时间t下，PCR 实验中分子数量的概率分布生成函数。我们还推导出此类实验产生的Ct 值的概率密度函数（pdf）、均值、方差和累积分布函数（cdf）。可以将 pdf 或 cdf 拟合到 PCR 数据，以估计输入 DNA 分子的数量。

- 在我们模型最简单的情形中——考虑扩增噪声但不考虑（上游）DNA 取样噪声的单相 PCR 模型——当$n \gg 1$时，均值Ct 值 $(\psi(x+1)-\psi(n))/r$可由 $\ln(x/n)/r$ [7] 很好地近似，其中n 为输入分子的数量，r （以底数为e 的尺度定义）为扩增效率，x 为定量阈值，$\psi(\cdot)$表示双伽马函数。

- 我们给出一种计算 PCR 实验检出限（LoD）的公式，

即在失败率不超过 α 的情况下能够检出的最少输入分子数。使用单相 PCR 模型，我们发现，当$\alpha = 0.05$ 时，仅考虑抽样噪声时确定的 LoD 为 3；同时考虑抽样噪声和扩增噪声（其中 r设为最大可能效率 m.p.e. 的 95%）时，LoD 则增至 ≈ 10 。随着r降低，LoD 也会增加；当效率为最大可能效率的 90% 时，LoD 翻倍至 ≈ 20 。这说明扩增效率对 LoD 有着不容忽视且显著的影响。

- 我们提出了一个计算 PCR 实验定量限（LoQ）的公式，即在给定精度水平和最大失败率的情况下，能够定量的最少分子数 α。与直觉相反，单相 PCR 模型预测，定量限与扩增效率无关。当 $\alpha = 0.05$ 时，定量限会从 10 增至 820：前者对应允许输入分子数与预期值之间最多相差两倍，后者则对应最多允许 10% 的偏差。这表明，测量 10 个或更少的分子时，误差在 95% 以上的情况下无法优于 2 倍。

- 该模型表明，解读数字 PCR 数据时常用的一项关键假设并不成立：只有一个输入分子的 PCR 实验总会产生阳性结果。即使将扩增效率设为较高的 95% m.p.e，该模型仍预测，此类实验产生阳性结果的概率为 $<4\%$。我们介绍了两种不同的方法，可用于根据数字 PCR 数据准确估计输入 DNA 分子的数量。

值得注意的是，此前已有尝试通过数学建模改进对PCR数据的解读。估算目标样本中DNA含量的经典方法，是将该样本产生的数据与含有已知或未知DNA含量的参考样本所获得的数据进行比较 [8]。由于已知DNA含量的参考样本必不可少，而其含量的测定本身也会受到实验误差的影响，因此，要准确实现目标样本中DNA的绝对定量颇具挑战。另一种方法是对仅由目标样本产生的PCR数据拟合数学模型，这些模型大多是现象学的 [4, 8, 9, 10, 11]。有关基于这种方法的各种方法的比较，请参见 [12] 。这些方法都未能充分说明扩增噪声如何影响PCR数据。

本文其余部分安排如下：我们在第 3.1 节概述模型结构，并在第 3.2 和 3.3 节介绍主要的数学结果。在第 3.4 节，我们应用这些结果计算检出限和定量限，并探讨扩增噪声如何使数字

PCR 数据的准确解读变得复杂。我们在第4节总结研究结果，并讨论我们的方法的其他应用。为提高可读性，数学证明和详细计算仅在附录（第5.1节）中给出。

3 结果

3.1 预备知识

我们将 PCR 过程建模为一个连续时间、离散状态马尔可夫跳跃过程 [13]，该过程演化至时间T。这种对 PCR 过程的表示基于以下事实：(1) PCR 反应的主要产物 DNA 分子是可计数的；(2) 给定反应的当前状态后，反应下一循环中发生的情况在条件上独立于过去发生的情况。我们选择将时间视为连续的（而非离散的），是因为实验测得的Ct 值为正实数。因此，反应速率以底数e 定义，而不是以底数 2 定义（离散时间 PCR 过程所预期的情况）；不过，这两种底数之间的转换十分简单。

我们将 PCR 过程的时间区间 $[0,T]$ 划分为p 个互不重叠的子区间I_i，每个子区间对应过程的一个不同阶段，并与概率状态转移速率r_i，$i = 1, 2, \ldots, p$相关联。这些转移率决定 DNA 扩增的效率。我们推导任意时间t时靶分子数量的概率生成函数 [14] 。利用该生成函数，我们推导相应的概率分布，尤其是Ct 值的概率密度函数（概率密度函数）。我们在两种不同情况下推导该概率密度函数，即

1. 当 PCR 过程的初始状态是确定性时，且 PCR 阶段长度和扩增效率已知；以及

2. 当初始状态服从泊松分布，且阶段长度和扩增效率已知时。

为阐释这些数学思想，我们将报告基于单阶段模型的计算和模拟。我们认为，这个模型较简单的实例足以分析各种现实世界中的PCR实验。原则上，每个 PCR 实验都可根据扩增速率分为以下三个阶段：指数前期，此时扩增速率低于指数增长；指数期；以及指数后期，此时 DNA 分子逐渐耗尽进一步扩增所需的试剂，扩增速率随之减慢。然而在实践中，PCR 实验通常的输出结果——Ct 值——在 PCR 过程进入指数期后便已确定，这意味着指数前期发生的动力学主要决定了这一特定结果。因此，为了理解

影响Ct值及其统计特性的因素，并评估 PCR 的相关运行特性，单阶段模型似乎已足够。因此，在适用的情况下，我们会着重说明数学方程在单阶段模型下的形式。此外，我们使用单阶段模型估计检出限和定量限（第 3.4 节），并将其用于评析解读数字 PCR 数据的标准方法（第3.5节）。

3.2 情形 1：初始状态为确定性的 PCR过程

3.2.1 分子数的概率生成函数

定理 1. 设 $\{X(t), t \in R\}$ 为一个具有 p 个阶段、可数状态空间 $S \subset \mathbb{N}^+$、各阶段转移率为 r_i、$i \in 1$、2、……、p 的连续时间马尔可夫过程，其状态转移概率为

$$P(X(t'+\Delta t)=x|X(t')=x') = \delta(x'-x+1)\sum_{i=1}^{p} r_i \mathbb{1}_{I_i}(t'), \tag{1}$$

其中，$\mathbb{1}$ 表示指示函数，$\delta(.)$ 表示克罗内克δ函数。若该过程以 n 个分子为初始状态，则时刻 $t \in I_k$、$k \le p$ 时存在的分子数的概率生成函数为

$$G(n,\vec{r},t,\vec{\tau};s) = \left[\frac{se^{-z}}{1-s(1-e^{-z})}\right]^n, \tag{2}$$

其中

$$z = r_k t + \sum_{i=1}^{k-1}(r_i - r_k)\tau_i, \tag{3}$$

I_i 表示第 i 个阶段，$\tau_i = |I_i|$、$i < k$ 表示其长度。

该定理的证明见第5.1.1节。现在我们将利用该定理推导在t时刻发现的分子数量的概率分布。

3.2.2 分子数量的概率分布

推论1. 在定理 *1* 所述的 *PCR*过程中，时刻 t 存在 x 个分子的概率 $\in I_k$ 由以下负二项分布给出：

$$P(x|n,\vec{r},t,\vec{\tau}) = \binom{x-1}{n-1} e^{-nz} \times (1-e^{-z})^{x-n}, \tag{4}$$

其中 z 由(3)给出。

该推论的证明见第5.1.2节。现在我们将利用该推论推导Ct值的概率密度函数、均值、方差和累积分布函数。

3.2.3 Ct值的概率密度函数、均值和方差

设t为定理 1 中所述 PCR过程的Ct值。根据定义，t是分子数量达到定量阈值时的时间，我们用x表示该阈值。设$t \in I_k$。在附录 [第5.1节.5],中，我们证明，给定$n, \vec{r} = (r_1, r_2, \ldots, r_{k-1})$和 $\vec{\tau} = (\tau_1, \tau_2, \ldots, \tau_{k-1})$后，$t$的概率密度函数具有如下形式：

$$P(t|n, \vec{r}, \vec{\tau}, x) = \frac{r_k e^{-nz}(1-e^{-z})^{x-n}}{B_\theta(n, x-n+1)}, \tag{5}$$

其中$B\theta(n, x-n+1)$是不完全贝塔函数，z由式（3）给出，并且

$$\theta = e^{-\sum_{i=1}^{k-1} r_i \tau_i}. \tag{6}$$

对于单阶段PCR过程，$\theta = 1$，因此概率密度函数为

$$P(t|n, r_1, x) = \frac{r_1 e^{-nz}(1-e^{-z})^{x-n}}{B(n, x-n+1)}. \tag{7}$$

Ct值的均值由下式给出（见第5.1.5节）

$$\mathbb{E}(t) = \sum_{i=1}^{k-1} \tau_i + \frac{\Gamma(n)^2 \theta^n \, {}_3\tilde{F}_2(n, n, n-x; n+1, n+1; \theta)}{r_k B_\theta(n, x-n+1)}, \tag{8}$$

其中 ${}_3\tilde{F}_2(n, n, n-x;\ n+1, n+1; \theta)$是正则化广义超几何函数。对于单相过程，均值由下式给出

$$\mathbb{E}(t) = \frac{\psi(x+1) - \psi(n)}{r_1}. \tag{9}$$

注意，当$n \gg 1$时，式（9）右侧可用 $\ln(x/n)/r_1$很好地近似。后一表达式通常用于近似Ct值的均值。例如， [7] 中使用该表达式根据数据估计PCR扩增效率。

Ct 值的方差为 $\mathrm{E}(t^2) - \mathrm{E}(t)^2$, 其中 $\mathrm{E}(t^2)$ 由式 (85) 给出。对于单相过程，方差为

$$\mathbf{Var}(t) = \frac{\psi_1(n) - \psi_1(x+1)}{r_1^2}, \tag{10}$$

其中 $\psi_1(\cdot)$ 是二阶多伽马函数（也称为三伽马函数）。

最后，Ct 值的累积分布函数为 [参见第5.1节。5]

$$F(t|n, \vec{r}, \vec{\tau}, x) = 1 - \frac{B_{e^{-z}}(n, x-n+1)}{B_\theta(n, x-n+1)}. \tag{11}$$

对于单相过程，其累积分布函数为

$$F(t|n,r_1,x) \quad = \quad 1-I_{e^{-r_1 t}}(n,x-n+1), \tag{12}$$

其中，$I_{e^{-r_1 t}}(n,x-n+1) = B_{e^{-r_t}}(n,x-n+1)/B(n,x-n+1)$ 是正则化不完全贝塔函数。从该累积分布函数中抽样相对简单：随机 Ct 值 t 可按如下方式获得：

$$t \quad = \quad -\frac{\ln I^{-1}_{1-u}(n,x-n+1)}{r_1}, \tag{13}$$

其中，u 从区间 $(0,1)$ 中均匀随机抽取，I^{-1}_{1-u} 是正则化不完全贝塔函数的反函数。若要找到与分位数$q \in (0,1)$ 对应的 Ct 值，只需用 q 替换 u。

要估计 n，可以将 t 的概率密度函数或累积分布函数拟合到数据。或者，也可以计算在给定 t 条件下 n 的后验密度。我们将在第5.1.5节中说明，对于单相过程，该后验密度为

$$P(n|r_1,t,x) \quad = \quad \frac{e^{-(n-1)r_1 t}(1-e^{-r_1 t})^{x-n}}{xB(n,x-n+1)}. \tag{14}$$

图1展示了输入分子数量和扩增效率不同时，单相概率密度函数的形状。为此，我们设定 $T = 35$ （现实世界中的PCR实验持续时间的一个常见上限）以及 $x = 2^T$；在理想扩增条件下，经过 T 个循环后，预期分子数即为该值（含有一个输入分子的样本在该分子得到完美扩增时，预计会在时间 $t \leq T$达到定量阈值 x）。我们将效率从最大可能效率的60%（相当于设定 $r_1 = 0.6 \times \ln 2$）变化到最大可能效率的100%。概率密度函数呈钟形，其位置和宽度取决于扩增效率和输入分子数量（图1左图）。扩增效率越高或输入分子数量越多，Ct 均值越小、方差越小，概率密度函数也越窄。相反，扩增效率越低或输入分子数量越少，Ct 均值越大、方差越大，概率密度函数也越宽（图1左图）。事实上，公式（5）预测，当扩增效率趋近于0时，概率密度函数将变为平坦分布，因为它会将每个 Ct 值都映射为0。

3.3 情形2：初始状态服从泊松分布的PCR过程

3.3.1 分子数的概率生成函数

定理 2. 设$\{X(t), t \in R\}$ 是定理 *1* 中描述的连续时间马尔可夫过程。若该过程的初始状态不是一个精确已知的输入*DNA*分子数，而是服从均值为 λ的泊松分布，那么其概率生成函数为

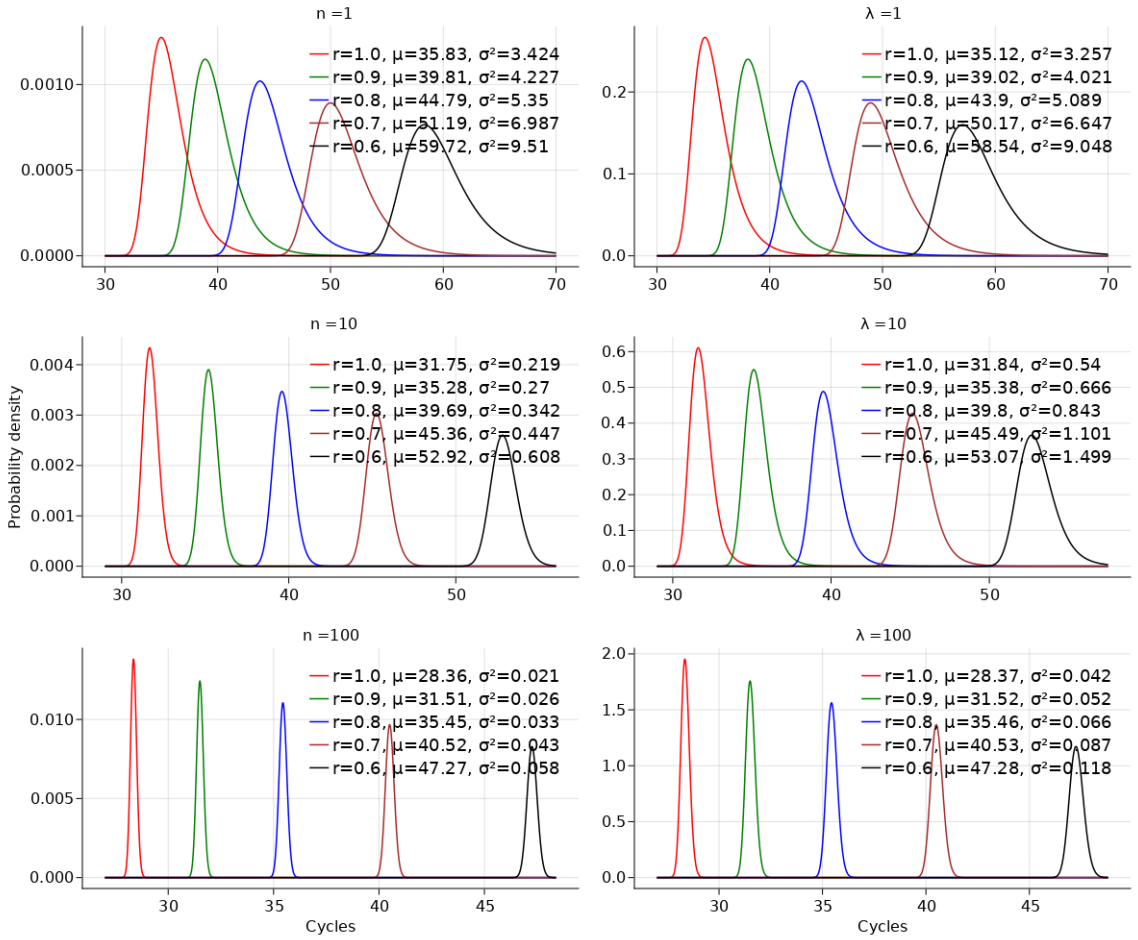


图1：单相过程的定量循环概率密度函数。考虑了初始状态为确定性（左图）或泊松分布（右图）的过程。前一种情况下，使用公式（5）计算概率密度函数；后一种情况下，使用方程（18）计算。定量阈值x设为 2^{35}。不同扩增效率r 对应的均值μ 和方差 σ^2 如图所示。为便于理解，r在我们的模型中以底数为e 的尺度定义，此处表示为其最大可能值 ln(2) 的百分比。

过程在未来时间 $t \in I_k$ 的状态由下式给出

$$G(\lambda, \vec{r}, t, \vec{\tau}; s) = e^{\frac{\lambda(s-1)}{1-s(1-e^{-z})}}, \tag{15}$$

其中 z 由(3)给出。

定理 2 的证明见第 5.1.3 节。现在我们利用定理 2 推导 PCR 过程中时间t所发现的分子数量的概率分布。

3.3.2 分子数量的概率分布

推论 2. 定理 *2* 所述 *PCR* 过程中，在循环 $t \in I_k$ 时存在 x 个分子的概率为：

$$P(x|\lambda, \vec{r}, t, \vec{\tau}) = e^{-\lambda}(1-e^{-z})^x \sum_{i=1}^{x} \frac{\binom{x-1}{i-1}}{i!} \left(\frac{\lambda e^{-z}}{1-e^{-z}} \right)^i, \tag{16}$$

其中 z 由(3)给出。

该推论的证明见第 5.1.4 节。值得注意的是，从证明中得出了以下组合三角形，它与著名的纳拉亚纳三角形 [15] 有关：

x					
1	1				
2	1	2			
3	1	6	6		
4	1	12	36	24	
5	1	20	120	240	120
	1	2	3	4	5
			k		.

该三角形的各项定义如下：

$$T(x,k) = \binom{x}{k-1}\binom{x-1}{k-1}(k-1)!, \; x \in \mathbb{Z}^+, k = 1, 2, ..., x, \tag{17}$$

统计通过复制随机选取的k个分子，得到x 个分子的方式数。$T(x,k)$ 与纳拉亚纳数$N(x,k)$ 的关系为

$$T(x,k) = k!\, N(x,k).$$

现在我们将使用这一推论推导Ct 值的概率密度函数，以及其均值、方差和累积分布函数。

3.3.3 Ct 值的概率密度函数、均值和方差

设t 为定理 2中所述PCR过程的Ct 值。如前所述，Ct值t 是该过程中DNA分子数量达到定量阈值的时间，我们将该阈值记为x。在附录 [第5.1节6], 中，我们将证明t 的概率密度函数为

$$P(t|\lambda,\vec{r},\vec{\tau},x) = \frac{r_k\lambda e^{-z}(1-e^{-z})^{x-1}\ {}_1F_1\left(1-x;2;\frac{-\lambda e^{-z}}{1-e^{-z}}\right)}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)}, \tag{18}$$

其中， θ由 (6) 给出，z由 (3) 给出，$1F_1(a;b;c)$ 是超几何函数（也称为第一类库默合流超几何函数）。

对于单相过程，其概率密度函数由 [参见第5.1节。6] 给出

$$P(t|\lambda,r_1,x) = \frac{r_1 x\lambda e^{-r_1 t}(1-e^{-r_1 t})^{x-1}\ {}_1F_1\left(1-x;2;\frac{-\lambda e^{-r_1 t}}{1-e^{-r_1 t}}\right)}{e^\lambda-1}. \tag{19}$$

Ct 值的均值由 [参见第5.1节。6] 给出

$$\mathbb{E}(t) = \frac{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}\left[r_k B_\theta(j,x-j+1)\sum_{i=1}^{k-1}\tau_i+\Gamma(j)^2\theta^j\ {}_3\tilde{F}_2(j,j,j-x;j+1,j+1;\theta)\right]}{r_k\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)}, \tag{20}$$

而二阶矩由 (119) 给出。

对于单相过程，其均值和方差分别由 [参见第5.1节。6] 给出

$$\mathbb{E}(t) = \frac{\psi(x+1)}{r_1}-\frac{\sum_{j=1}^{x}\frac{\lambda^j}{j!}\psi(j)}{r_1\left(e^\lambda-1\right)} \quad \text{and} \tag{21}$$

$$\mathbf{Var}(t) = \frac{\left(e^\lambda-1\right)\sum_{j=1}^{x}\frac{\lambda^j}{j!}\left[\psi_1(j)+\psi(j)^2\right]-\left[\sum_{j=1}^{x}\frac{\lambda^j}{j!}\psi(j)\right]^2}{\left(r_1(e^\lambda-1)\right)^2}-\frac{\psi_1(x+1)}{r_1^2}. \tag{22}$$

Ct值的累积分布函数由 [见第5.1节。6]给出

$$F(t|\lambda,\vec{r},\vec{\tau},x) = 1-\frac{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_{e^{-z}}(j,x-j+1)}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)}. \tag{23}$$

对于单相过程，其累积分布函数为

$$F(t|\lambda, r_1, x) \quad = \quad 1 - \frac{x \sum_{j=1}^{x} \frac{\binom{x-1}{j-1}\lambda^j}{j!} B_{e^{-r_1 t}}(j, x-j+1)}{e^{\lambda} - 1}. \tag{24}$$

要估计 λ，可以对t 的概率密度函数或累积分布函数进行数据拟合。或者，也可以计算给定t 时 λ 的后验密度。在第5.1.6节中，我们将展示，对于单相过程，该后验密度为

$$P(\lambda|r_1, t, x) = \frac{\lambda w \, {}_1F_1(1-x; 2; \frac{-\lambda w}{1-w})}{(e^{\lambda} - 1)(1-w) \sum_{j=1}^{x} \binom{x-1}{j-1} \left(\frac{w}{1-w}\right)^j \zeta(j+1)}, \tag{25}$$

其中，$w = e^{-r_1 t}$，而 $\zeta(\cdot)$ 表示黎曼ζ函数。

请注意，由于这些求和项来自对泊松分布求期望，公式（18）—（25）中的求和可以在任意 $j \gg \lambda$值处截断，而不会损失精度。

图1展示了不同 λ值和不同扩增效率下单相过程概率密度函数的形状。与初始状态确定的 PCR 过程（图1左图）一样，该概率密度函数也呈钟形（图1右图）。其位置和宽度由扩增效率和 λ共同决定。正如预期，较高的扩增效率或较大的 λ 值会产生较小的 Ct 均值、较小的方差和较窄的概率密度函数（图1右图）。相反，较低的扩增效率或较小的 λ 值会产生较大的 Ct 均值、较大的方差和较宽的概率密度函数。

3.4 检出限和定量限

现在，我们将从理论上说明，如何运用我们所构建的数学框架来实现某些在操作层面具有重要意义的目标。尤其是，人们通常希望量化特定 PCR方法（下文称为“PCR方案”）的检出限（LoD）。PCR方案的检出限是指其能够检测到的最少分子数，且失败率不超过预设阈值 α （α 也称为显著性水平）。通常，检出限较低的方案优于检出限较高的方案。理想情况下，检出限应等于或小于预期在所考虑样本中出现的输入DNA分子数。另一个重要的操作目标是确定PCR方案的定量限（LoQ），即在给定精度水平（此处用参数 β衡量）和给定的最大失败率 α的条件下，该方案能够估计的最少分子数。当需要检测目标分子数的 $\geq \beta$ 倍变化时，该方案的定量限应具有 $\leq \beta$ 的精密度。用于估计检出限和定量限的方法既繁琐 [16]，又往往依赖某些临时的 数学

近似 [16, 17], ，而我们希望通过建立并执行对估计问题的精确数学表述来避免使用这些近似。

我们从检出限估计问题开始。对于具有确定性初始状态的 PCR 过程，检出限可表示如下：

$$\begin{aligned} \mathrm{LoD} = \quad & \mathbf{min}\ n \\ & \text{s.t. } F(T|n,\vec{r},\vec{\tau},x) > 1-\alpha, \end{aligned} \tag{26}$$

其中，$F(T|n,\vec{r},\vec{\tau},x)$由式（11）给出，$T$是 PCR 过程的最大实际持续时间。对于初始状态服从泊松分布的过程，n 由 λ 替代。

在补充图5.1中，我们展示了单相过程中检出限如何随扩增效率变化，其中初始状态可以是确定性的，也可以服从泊松分布。前一种情况下，过程包含扩增噪声，但不包含抽样噪声；后一种情况下，过程同时包含抽样噪声和扩增噪声。作为比较，我们还展示了一个仅包含抽样噪声、并以泊松分布建模但不包含扩增噪声的过程中的检出限。仅存在抽样噪声时，检出限最低（检出限为 $=3$ 个分子）；同时存在抽样噪声和扩增噪声时，检出限最高（从最大可能效率的 100% 时的 6 个分子，到最大可能效率的 80% 时的 157 个分子）。仅包含扩增噪声而不包含抽样噪声的过程，其检出限介于两者之间。在这些计算示例中，用于估计定量限的方程参数完全已知，因此可以精确获知检出限。在实际应用中，参数值会存在不确定性，这会以可量化的方式使检出限估计值变得不确定。

现在我们将注意力转向一个问题：如何估计确定性初始状态下 PCR过程的定量限。假设该过程生成一个Ct 值t ，随后利用它得到输入分子数量的估计值，记为 $\hat{}n$。设n为实际输入分子数量。我们希望计算这样的概率：对于同一过程生成的任意数据t ，$\hat{}n$ 与n 的差异不超过β倍 $\beta \geq 1$。我们将定量限定义为使该概率超过$1-\alpha$的最小n 值。具体而言，定量限为

$$\begin{aligned} \mathrm{LoQ} = \quad & \mathbf{min}\ n \\ & \text{s.t. } P\left(\lfloor n/\beta \rfloor \leq \hat{n} \leq \lceil \beta n \rceil \mid n,\vec{r},\vec{\tau},x\right) > 1-\alpha, \end{aligned} \tag{27}$$

其中，$\lfloor v \rfloor$（分别为 $\lceil v \rceil$）表示小于（分别为大于）或等于v的最大（分别为最小）整数。

聚焦于单相过程，为了得到 $P(\lfloor n/\beta \rfloor \leq \hat{}n \leq \lceil \beta n \rceil \mid n,r_1,x)$，我们先对（14）右侧关于 t 求边缘化，然后将求和范围从 $\lfloor n/\beta \rfloor$ 取至 $\lceil \beta n \rceil$,

得出 [参见第5.1节。5]

$$P(\lfloor n/\beta \rfloor \le \hat{n} \le \lceil \beta n \rceil \mid n, r_1, x) = \sum_{\hat{n}=\lfloor n/\beta \rfloor}^{\lceil \beta n \rceil} \frac{B(\hat{n}+n-1, 2x-\hat{n}-n+1)}{xB(\hat{n}, x-\hat{n}+1)B(n, x-n+1)} = P(\lfloor n/\beta \rfloor \le \hat{n} \le \lceil \beta n \rceil \mid n, x). \quad (28)$$

值得注意的是，虽然公式（28）同时取决于n和x，但它不取决于扩增效率r_1。在附录中，[见公式（130）]，，我们采用类似的步骤来得到$P\left(\lfloor \lambda/\beta \rfloor \le \hat{\lambda} \le \lceil \beta\lambda \rceil \mid \lambda, r_1, x\right)$，即对于由泊松分布的初始状态的单阶段PCR过程生成的任意数据t，估计值λ（记作 $\hat{\lambda}$）与实际值的差不超过一个倍数 β的概率。

设定 $\alpha = 0.05$，并允许 $\hat{n}$ 与n 之间的偏差最多为 10%（对应于设定 $\beta = 1.1$），得到的定量限为 820 个分子。允许偏差最高达 25%（对应于 $\beta = 1.25$）时，定量限降至 146 个分子；允许偏差增至 50%（对应于 β =1.5）时，定量限则降至 43 个分子。分析表明，在所考虑的 5% 失败率下，可以检测到 10 个输入分子，但误差至少为 ≈200%，即与n 相比存在 ≈ 2 倍偏差（对应于 $\beta = 2$）。由于这些计算未考虑抽样噪声，因此它们只能给出在所考虑的失败率和精度水平下可实现的定量限下限（β）。如前所述，在实际应用中，用于估计定量限的方程参数并非完全已知，这会决定估计的定量限所伴随的不确定性大小。

3.5 扩增噪声决定数字PCR的结果

如前所述，数字PCR是常规PCR的一种变体，旨在提高DNA定量的准确性。在数字PCR中，样品主混合液（含有数量未知的输入DNA分子以及DNA复制所需的所有试剂）被均匀分配到数百个（有时甚至数千个）物理分区中，这些分区可以是液滴或微孔 [5]。根据均值为 $\lambda = CV/D$ 的泊松分布，预计每个分区会接收到零个、一个或多个DNA分子，其中 C 是原始样品中的DNA浓度，V是分区体积，D 是制备主混合液时对样品采用的（已知）稀释因子。每个分区内独立且同时进行PCR反应，并识别出阳性分区。因此，所得数据——即每个PCR反应的阳性或阴性结果——是数字化的。解读这些数据的标准方法假设，接收到至少一个靶分子的分区会检测为阳性，其占比 $\hat{f}$ 近似为 $\hat{f} \approx 1 - e^{-\lambda}$，据此将 λ 估计为 $\hat{\lambda} = -\ln(1-\hat{f})$，然后用于估计 C。

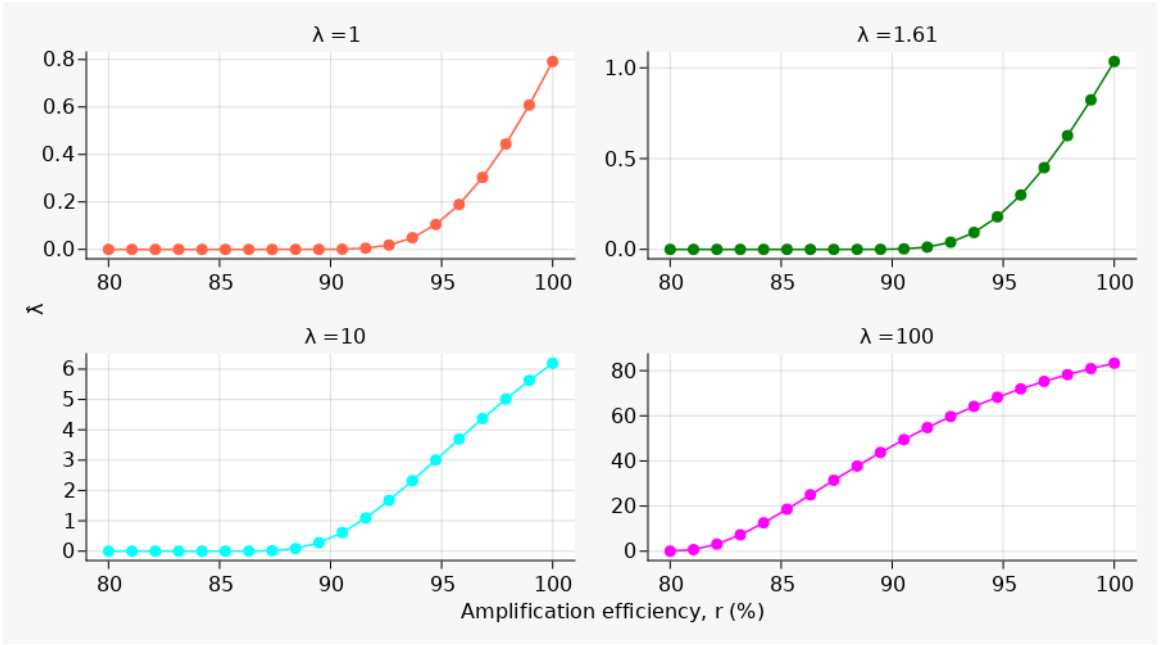


图2：使用解读数字 PCR 数据的标准方法会低估 λ 。我们改变了扩增效率和输入DNA量 λ ，并使用公式（29）估计 λ。所得估计值记为 $\hat{\lambda}$，并以最大可能效率的百分比为单位，绘制其与效率的关系图。当效率低于 95% 时，只能检测到极少量的输入DNA。当效率为 95% 时，检测到的量从 $\lambda = 1$时的 12% 到 $\lambda = 100$时的 69% 不等。效率达到 100% 时，检测到的量增加至 ≈80%。

然而，根据我们的模型，由于PCR扩增噪声的影响，认为阳性分区的占比等于分区接收到一个或多个靶分子的泊松概率，这一假设并不成立。事实上，将扩增效率设为相当高的95% m.p.e（即 $r_1 = 0.95 \ln 2$），并在方程（24）中使用 $T = 35, x = 2^T$，我们预测，仅含有一个分子的分区中有 <4% 会检测为阳性，这比泊松方法假设的结果小 >25 倍。在补充图5.2中，我们比较了使用我们的模型 [方程(24)] 计算的阳性分区占比与使用泊松方法计算的阳性占比，比较时采用了不同的 λ 值和不同的扩增效率。我们发现，当 λ值较小时（包括该方法预期能产生最精确 λ估计值时对应的数值1.61），泊松方法会高估阳性分区的占比。只有当 λ 值相对较大（10）时，我们才发现泊松方法对阳性分区占比的估计与使用我们的模型计算出的噪声校正期望值一致（补充图5.2）。

我们使用模型研究这种对阳性分区比例的高估会如何影响泊松方法得出的 λ 估计值（记为 $\hat{\lambda}$）的准确性

在公式（24）中令 $t = T$，我们得到

$$\begin{aligned}\hat{\lambda} &= -\ln\left(1-\hat{f}\right) \\ &= \ln\left(\frac{e^{\lambda}-1}{\sum_{j=1}^{x}\frac{\binom{x}{j}}{(j-1)!}\lambda^{j}B_{e^{-r_1 T}}(j,x-j+1)}\right).\end{aligned} \tag{29}$$

根据公式（29），$\hat{}\lambda$ 高度依赖于扩增效率r_1。当r_1 趋近于 0 时，$\hat{}\lambda$ 等于 0。随着r_1 增至其最大可能值 ln(2)，$\hat{}\lambda$也随之增加并趋近于 λ。图2展示了$\hat{}\lambda/\lambda$与$\hat{}$扩增效率之间的关系。值得注意的是，当效率低于最大可能效率的 95% 时，λ/λ 非常小，这表明泊松方法显著低估了 λ。当效率等于最大可能效率的 95% 时，$\hat{}\lambda/\lambda$ 从 $\approx 12\%$（在 $\lambda = 1$ 时）增至 $\approx 69\%$（在 $\lambda = 100$ 时）。将效率提高至最大可能值（最大可能效率的 100%）会使$\hat{}\lambda/\lambda$ 增至 $\approx 80\%$（在 $\lambda = 100$ 时）。这些结果表明，泊松方法预计会低估λ，因为该方法未考虑扩增噪声。事实上，实验数据表明，该方法倾向于严重低估输入DNA分子的数量（例如，参见 [19]中的补充表 6）。

4 讨论

DNA 定量实验的结果，包括基于聚合酶链式反应（PCR）的实验结果，往往会在同一实验的不同重复之间以及不同实验之间出现差异，因而难以解读，也限制了这些结果在其产生的特定情境之外的实用性。事实上，已知多种因素会导致 PCR 结果出现差异 [20, 21, 22]，包括 DNA 模板复杂程度的变化和反应环境中靶分子的随机分布；所用 PCR 仪的类型和缓冲液组分；PCR 三个热循环的持续时间和温度；寡核苷酸引物与靶 DNA 的结合动力学；以及 DNA 聚合酶和其他 PCR 试剂的稳定性。Taylor 等人 [22]回顾了 PCR 实验中差异的来源，并提出了一套逐步实施的流程，以尽量减少实践中的此类差异。

PCR 实验中常见的输出是定量循环（记作Ct 或Cq值），即DNA分子数量超过某个既定阈值时所处的PCR循环，该阈值称为定量阈值。Ct 值会随输入DNA分子数量和PCR 扩增效率而变化，而后者又会随上述实验变量而变化。理想情况下，可通过应用能考虑到潜在生成过程固有随机性的数学方法，对这些变化的输出进行解卷积，从而估计实验中最受关注的输入DNA分子数量，

该估计所依据的正是其背后的生成过程。

我们提出了一种用于DNA定量的数学建模方法，并将其内在随机性纳入考量。我们以PCR（应用最广泛的一类DNA定量过程）来阐释我们的数学思路；这些思路也适用于更广泛的一类此类过程（例如 [6]）。利用该模型，我们推导出了PCR过程中分子数的概率生成函数，其中输入分子数既可以是确定性的，也可以服从泊松分布；此外，我们还推导出了此类过程产生的Ct 值的概率密度函数（pdf）、均值、方差和累积分布函数（cdf）。与确定性情形不同，确定性情形下PCR的输出仅受扩增噪声影响；后一种情形下，输出还包含抽样噪声。我们为PCR过程的这些重要统计特性推导出的方程，揭示了Ct 值与其内在变量之间的函数关系；此前，这些关系只能通过经验方法获取。

为了阐释我们的数学思路，我们聚焦于单阶段PCR过程；在这一过程中，我们的建模结果具有相对简单的数学形式。我们发现，当输入DNA分子数n 较大时，关于均值Ct 值是简单对数函数这一常见假设成立。当n较小时，则需要进行修正。均值Ct 值的精确表达式为 $(\psi(x+1) - \psi(n))/r$，其中x 为定量阈值，r 为扩增效率，$\psi(\cdot)$表示第一多伽马函数。方差具有简洁的形式 $(\psi_1(n) - \psi_1(x+1))/r^2,$，其中 $\psi_1(\cdot)$表示二阶多伽马函数。因此，方差在很大程度上取决于扩增效率。图1展示了这一效应：随着扩增效率降低，Ct值的概率密度函数变得更宽。值得注意的是，在这个简单情形下，Ct 值的变异系数（即标准差与均值之比）不受扩增效率影响。

表征 PCR 过程性能的两个重要指标是检出限（LoD）和定量限（LoQ）。检出限是指在失败率不超过阈值α的情况下可检测到的最少分子数；定量限则是指在给定精度水平（即允许相对于真实值存在一定的最大倍数偏差）和给定最大失败率 α的条件下可定量的最少分子数。我们提供了计算检出限和定量限的数学公式。在单阶段PCR过程的背景下仔细考察这些公式后发现，扩增效率略微降低就可能导致检出限大幅增加。例如，当 $\alpha = 5\%$时，将效率从最大可能效率（m.p.e）的 95% 降至最大可能效率的 90%，会使检出限翻倍，从 $\approx$10个输入分子增至 $\approx$20 个分子。与检出限不同，我们发现定量限与效率无关。若允许n 与其估计值之间最多相差 2 倍，则定量限为 $\approx$10 个分子。将允许的倍数差异缩小至 10%，会使定量限增至 820

个分子。我们的方法可显著改进当前估计检出限和定量限的方法；这些方法既费时 [16] ，又常常依赖某些粗略的数学近似 [16, 17] ，而采用我们的方法即可避免这些近似。

此外，我们应用所提出的方法，揭示了扩增噪声对通过解读数字 PCR 数据的标准方法估计输入DNA分子期望数量 λ 所产生的影响。该方法的一个关键假设是，只要 PCR 反应中至少含有一个输入DNA分子，该反应就会呈阳性。我们证明，由于 PCR 扩增具有随机性，这一假设通常并不成立。当 λ 较小时，随机效应尤其显著，而这正是数字PCR优先采用的工作区间。在扩增效率高达 95% 最大可能效率时，我们发现，当 $\lambda = 1$ 时，按标准方法计算的数字PCR阳性反应比例与考虑扩增噪声后所得数值之比仅为 18.3%；当 $\lambda = 10$ 时，该比例增至 $\approx$100%（图5.2）。因此，我们发现，标准方法中的随机效应会导致 λ 被显著低估。事实上，在扩增效率高达 95% 最大可能效率时，预测标准方法会分别将 λ 低估 $\approx$8.1、 $\approx$3.1 和 $\approx$1.4 倍，其中 λ 分别等于 1、10 和 100（图2）。这与实证观察到的范围一致（例如 [19]）。

利用我们的数学方法，可以采用以下两种不同的方法，获得对 λ更准确的估计值。首先，如果能够获得阳性数字PCR反应的Ct 值，就可以对这些Ct 值拟合方程（19），采用基于似然的方法或贝叶斯统计方法，估计 λ 的最可能值，并给出其置信区间（或可信区间）。其次，如果只能获得各个反应的二元结果（即阳性或阴性），仍然可以利用这些结果的变异性来估计 λ。具体而言，假设有N个不同的反应，可将其随机分为若干组，每组包含N'个反应。假设每组中阳性反应数服从二项分布，则其一阶和二阶矩分别为$F(T)N'$和$F(T)N'(1 + F(T)(N' - 1))$，其中$F(T)$根据（24）计算。这些矩包含$F(T)$的两个自由参数（即 λ 和r）的信息，可轻易提取这些信息，以估计 λ并给出其置信区间（或可信区间）。

Ct值 是根据 DNA 分子在 PCR 扩增过程中产生的荧光曲线估计得出的。我们的数学分析可以直接扩展，以获得荧光强度$P(y)$随时间变化的概率密度，随后可将其拟合到荧光曲线上，作为估计输入DNA分子数量的另一种方法。利用概率论中的标准结果（例如，参见 [23]），可以根据 PCR过程中在时间t检测到的分子数量的累积分布函数$F_t(x)$ [（基于方程 (4) 或 (16) 计算得出）] 以及y 与x之间预期的线性关系 [24] ，推导出$P(y)$。具体而言，$P(y)$可

表示为

$$P_t(y) = \frac{1}{\alpha} h_t(g^{-1}(y)), \tag{30}$$

其中

$$h_t(x) = \frac{d}{dx} F_t(x), \tag{31}$$

$y = \alpha x + \beta = g(x)$，以及 α，$\beta > 0$。我们将在未来的一篇论文中详细探讨这种估算输入DNA分子数量的替代方法。

5 补充信息

5.1 附录

本节包含数学证明和详细计算，以支持第3节中呈现的结果。

5.1.1 定理 1 的证明

证明。 我们将对k使用数学归纳法来证明定理 1。

- $k = 1$：对应于单阶段过程的 Chapman–Kolmogorov 前向方程为：

$$\frac{\partial P(X = x, t | X = x', t')}{\partial t} = r_1(x-1)P(X = x-1, t | X = x', t') - r_1 x P(X = x, t | X = x', t'), \tag{32}$$

其中我们设$t = t' + \Delta t$，而r_1 是该过程的扩增效率。为简化记法，我们将$P(X = x, t | X = x', t')$简写为$P(x, t)$。

我们将使用一种称为概率生成函数（pgf）的强大组合工具来求解 (32) [14]。回想一下，$P(x,t)$的 pgf 定义如下：

$$G(s, t) = \sum_{x=0}^{\infty} s^x P(x, t),$$

其中s 是一个记账变量。

将（32）两边乘以s^x，并对x的所有可能取值求和，得到：

$$\begin{aligned}\sum_{x=0}^{\infty} s^x \frac{\partial P(x,t)}{\partial t} &= r_1 \sum_{x=0}^{\infty}(x-1)s^x P(x-1,t) - r_1 \sum_{x=0}^{\infty} x s^x P(x,t) \\ &= r_1 s^2 \sum_{x=0}^{\infty}(x-1)s^{x-2} P(x-1,t) - r_1 s \sum_{x=0}^{\infty} x s^{x-1} P(x,t) \\ &= = r_1 s \left[s \sum_{x=0}^{\infty}(x-1)s^{x-2} P(x-1,t) - \sum_{x=0}^{\infty} x s^{x-1} P(x,t) \right]. \end{aligned} \tag{33}$$

利用

$$\begin{aligned}\frac{\partial G(s,t)}{\partial s} &= \sum_{x=0}^{\infty} x s^{x-1} P(x,t) \text{ and} \\ \frac{\partial G(s,t)}{\partial t} &= \sum_{x=0}^{\infty} s^x \frac{\partial P(x,t)}{\partial t}, \end{aligned} \tag{34}$$

我们将（33）简化为

$$\frac{\partial G(s,t)}{\partial t} = r_1 s(s-1)\frac{\partial G(s,t)}{\partial s}, \tag{35}$$

这是关于$G(s,t)$的偏微分方程（pde）。

我们将用特征线法求解（35）。为此，我们定义新变量

$$u = u(s,t) \text{ and } v = v(s,t),$$

它将（35）转化为更简单的方程

$$\frac{\partial W(u,v)}{\partial u} + H(u,v)W(u,v) = F(u,v), \tag{36}$$

其解为

$$W(u,v) = e^{-\int H(u,v)du}\left[\int F(u,v)e^{\int H(u,v)du} + \Psi(v)\right],$$

其中

$$W(u,v) = G(s(u,v),t(u,v)).$$

这要求 $v(s,t) = c$，其中 c 是任意常数。由此得到的特征方程为

$$\frac{ds}{dt} = -r_1 s(s-1),$$

其解为

$$\frac{s-1}{s}e^{r_1 t} = c = v(s,t).$$

令$u(s,t) = t$，得到

$$\begin{aligned}\frac{\partial G}{\partial t} &= \frac{\partial W}{\partial t} = \frac{\partial W}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial W}{\partial v}\frac{\partial v}{\partial t} \\ &= \frac{\partial W}{\partial u} + \frac{r_1(s-1)}{s}e^{r_1 t}\frac{\partial W}{\partial v} \end{aligned} \tag{37}$$

and

$$\begin{aligned}\frac{\partial G}{\partial s} &= \frac{\partial W}{\partial u}\frac{\partial u}{\partial s} + \frac{\partial W}{\partial v}\frac{\partial v}{\partial s} \\ &= \frac{1}{s^2}e^{r_1 t}\frac{\partial W}{\partial v}. \end{aligned} \tag{38}$$

将 (37) 和 (38) 代入 (35)，得到

$$\frac{\partial W}{\partial u} = 0, \tag{39}$$

其形式与 (36) 相同。(39) 的解为

$$\begin{aligned} W(u,v) &= \Psi(v) \\ \Longrightarrow \quad G(s,t) &= \Psi\left(\frac{s-1}{s}e^{r_1 t}\right). \end{aligned} \tag{40}$$

如果过程开始时有n 个分子（$t = 0$），那么$p(x,0) = 1$ 当$x = n$ 时，且$p(x,0) = 0$ 在其他情况下。因此，

$$G(s,0) = \Psi\left(\frac{s-1}{s}\right) = \sum_{x=0}^{\infty} s^x P(x,0) = s^n. \tag{41}$$

在 (41) 中，$\Psi(y)$ 的自变量y 映射为 $(\frac{1}{1-y})^n$，~~这~~意味着

$$G(s,t) = \Psi\left(\frac{s-1}{s}e^{r_1 t}\right) = \left(\frac{1}{1-\frac{s-1}{s}e^{r_1 t}}\right)^n = \frac{s^n e^{-nr_1 t}}{[1-s(1-e^{-r_1 t})]^n}. \tag{42}$$

当$k = 1$时，方程 (42) 与 (2) 相符。

推论 3. 方程 (42) 求解了 (35)。

证明。(35) 右侧为

$$\begin{aligned}
\frac{\partial G}{\partial s} &= ns^{n-1}e^{-nr_1t}\left[1-s(1-e^{-r_1t})\right]^{-n} + ns^n e^{-nr_1t}\left[1-s(1-e^{-r_1t})\right]^{-(n+1)}\left(1-e^{-r_1t}\right) \\
&= \frac{ns^{n-1}e^{-nr_1t}}{[1-s(1-e^{-r_1t})]^n}\left[1+\frac{s(1-e^{-r_1t})}{1-s(1-e^{-r_1t})}\right] \\
&= \frac{ns^{n-1}e^{-nr_1t}}{[1-s(1-e^{-r_1t})]^{(n+1)}}\left[1-s(1-e^{-r_1t})+s(1-e^{-r_1t})\right] \\
&= \frac{ns^{n-1}e^{-nr_1t}}{[1-s(1-e^{-r_1t})]^{(n+1)}}, \qquad (43)
\end{aligned}$$

而左侧为

$$\begin{aligned}
\frac{\partial G}{\partial t} &= -nr_1 s^n e^{-nr_1t}\left[1-s(1-e^{-r_1t})\right]^{-n} + ns^{n+1}r_1 e^{-(n+1)r_1t}\left[1-s(1-e^{-r_1t})\right]^{-(n+1)} \\
&= \frac{nr_1 s^n e^{-nr_1t}}{[1-s(1-e^{-r_1t})]^n}\left[\frac{se^{-r_1t}}{1-s(1-e^{-r_1t})}-1\right] \\
&= \frac{nr_1 s^n e^{-nr_1t}}{[1-s(1-e^{-r_1t})]^{(n+1)}}\left[se^{-r_1t}-1+s-se^{-r_1t}\right] = \frac{nr_1 s^n e^{-nr_1t}(s-1)}{[1-s(1-e^{-r_1t})]^{(n+1)}} \\
&= r_1 s(s-1)\overbrace{\left[\frac{ns^{n-1}e^{-nr_1t}}{[1-s(1-e^{-r_1t})]^{(n+1)}}\right]}^{\text{this matches (43)}} = r_1 s(s-1)\frac{\partial G}{\partial s}. \qquad (44)
\end{aligned}$$

□

- $k = 2$:

  有两个放大阶段，速率分别为r_1 和r_2，。第一阶段从时间$t = 0$ 持续到$t = \tau_1$，第二阶段则从$t = \tau_1$ 持续到$t = \tau_1 + \tau_2$。在第二阶段，概率生成函数的通用函数形式与第一阶段完全相同，但初始条件不同。具体而言，我们有

$$G(s,t) = \Psi\left(\frac{s-1}{s}e^{r_2(t-\tau_1)}\right),$$

  在初始条件（时间为 $t = \tau_1$）下

$$G(s,\tau_1) = \Psi\left(\frac{s-1}{s}\right) = \frac{s^n e^{-nr_1\tau_1}}{[1-s(1-e^{-r_1\tau_1})]^n}.$$

使用与 $k = 1$ 时相同的步骤，我们得到

$$\begin{aligned} G(s,t) &= \Psi\left(\frac{s-1}{s}e^{r_2(t-\tau_1)}\right) \\ &= \frac{\left(\frac{1}{1-\frac{s-1}{s}e^{r_2(t-\tau_1)}}\right)^n e^{-nr_1\tau_1}}{\left[1-\left(\frac{1}{1-\frac{s-1}{s}e^{r_2(t-\tau_1)}}\right)(1-e^{-r_1\tau_1})\right]^n} \\ &= \frac{s^n e^{-n[r_2 t+(r_1-r_2)\tau_1]}}{\left[1-s\left(1-e^{-[r_2 t+(r_1-r_2)\tau_1]}\right)\right]^n}. \end{aligned} \tag{45}$$

当$k = 2$时，(45) 的右侧等于 (35)，正如预期的那样。

- 我们假设该命题对$t \in I_k$成立，即

$$G(s,t) = \left[\frac{se^{-z}}{1-s(1-e^{-z})}\right]^n,$$

其中$z = r_k t + \sum_{i=1}^{k-1}(r_i - r_k)\tau_i$，我们将证明该命题对$t \in I_{k+1}$成立。与之前一样，在$k+1$阶段，生成函数具有如下函数形式

$$G(s,t) = \Psi\left(\frac{s-1}{s}e^{r_{k+1}(t-\sum_{i=1}^{k}\tau_i)}\right).$$

在$t = \sum_{i=1}^{k}\tau_i$时，根据归纳步骤，我们有

$$G(s,t) = \Psi\left(\frac{s-1}{s}\right) = \left[\frac{se^{-z}}{1-s(1-e^{-z})}\right]^n.$$

使用与之前相同的论证，我们得到，对于$t \in I_{k+1}$,

$$\begin{aligned} G(s,t) &= \Psi\left(\frac{s-1}{s}e^{r_{k+1}(t-\sum_{i=1}^{k}\tau_i)}\right) \\ &= \left[\frac{\frac{1}{1-\frac{s-1}{s}e^{r_{k+1}(t-\sum_{i=1}^{k}\tau_i)}}e^{-z}}{1-\frac{1}{1-\frac{s-1}{s}e^{r_{k+1}(t-\sum_{i=1}^{k}\tau_i)}}(1-e^{-z})}\right]^n \\ &= \frac{s^n e^{-n\left[r_{k+1}t+\sum_{i=1}^{k}(r_i-r_k)\tau_i\right]}}{\left[1-s\left(1-e^{-\left[r_{k+1}t+\sum_{i=1}^{k}(r_i-r_k)\tau_i\right]}\right)\right]^n}, \end{aligned} \tag{46}$$

至此，定理 1 的证明结束。

□

5.1.2 推论1的证明

Proof. 根据定理 1，我们知道 $P(x|n,\vec{r},t,\vec{\tau})$ 的生成函数为 y

$$G(s,t)=\left[\frac{se^{-z}}{1-s(1-e^{-z})}\right]^n.$$

令 $p=e^{-z}$ 且 $q=1-p$。那么，

$$G(s,t)=\left[\frac{sp}{1-sq}\right]^n=(sp)^n[1-sq]^{-n}.$$

$P(x|n,\vec{r},t,\vec{\tau})$ 是 $G(s,t)$ 的幂级数展开中 s^x 的系数，表示为

$$G(s,t) = (sp)^n[1-sq]^{-n}=(sp)^n\sum_{i=0}\binom{-n}{i}(-sq)^i=(sp)^n\sum_{i=0}\binom{-n}{i}(-1)^i(sq)^i. \tag{47}$$

But

$$\begin{aligned}\binom{-n}{i} &= \frac{-n(-n-1)(-n-2)(-n-3)\ldots(-n-(i-2))(-n-(i-1))}{1.2.3.\ldots.(i-1)i}\\ &= (-1)^i\frac{n(n+1)(n+2)(n+3)\ldots(n+(i-2))(n+(i-1))}{i!}\\ &= (-1)^i\binom{n+i-1}{i}\\ \Longrightarrow\ (-1)^i\binom{-n}{i} &= \binom{n+i-1}{i}.\end{aligned} \tag{48}$$

将 (48) 代入 (47) 得

$$G(s,t) = (sp)^n\sum_{i=0}\binom{-n}{i}(-1)^i(sq)^i=(sp)^n\sum_{i=0}\binom{n+i-1}{i}(sq)^i. \tag{49}$$

令 $x=n+i$。那么，

$$G(s,t) = (sp)^n\sum_{x=n}^{\infty}\binom{x-1}{x-n}(sq)^{x-n}=\sum_{x=n}^{\infty}\binom{x-1}{x-n}p^nq^{x-n}s^x. \tag{50}$$

因此，在时间 t 时有 x 个分子的概率 $P(x,t)$ 为

$$P(x,t)=\binom{x-1}{x-n}p^nq^{x-n}=\binom{x-1}{n-1}e^{-nz}(1-e^{-z})^{x-n}. \tag{51}$$

□

Coro推论 4. 定理 *1* 给出的概率分布满足 *Chapman-Kolmogorov* *rov*

方程(32)。

证明。

$$\begin{aligned}
\frac{\partial P(x,t)}{\partial t} &= \binom{x-1}{x-n}\left[-nr_k e^{-z}(1-e^{-z})^{x-n} + r_k(x-n)e^{-2z}(1-e^{-z})^{x-n-1}\right] \\
&= r_k\binom{x-1}{x-n}e^{-z}(1-e^{-z})^{x-n}\left[-n+(x-n)\tfrac{e^{-z}}{1-e^{-z}}\right] \\
&= r_k\binom{x-1}{x-n}e^{-z}(1-e^{-z})^{x-n}\left[-n+(x-n)\tfrac{e^{-z}}{1-e^{-z}} - \tfrac{x-n}{1-e^{-z}} + \tfrac{x-n}{1-e^{-z}}\right] \\
&= r_k\binom{x-1}{x-n}e^{-z}(1-e^{-z})^{x-n}\left[\tfrac{x-n}{1-e^{-z}} - n + \tfrac{(x-n)}{1-e^{-z}}(e^{-z}-1)\right] \\
&= r_k\binom{x-1}{x-n}e^{-z}(1-e^{-z})^{x-n}\left[\tfrac{x-n}{1-e^{-z}} - x\right] \\
&= r_k\left[(x-n)\binom{x-1}{x-n}\right]e^{-z}(1-e^{-z})^{x-n-1} - r_k x\binom{x-1}{x-n}e^{-z}(1-e^{-z})^{x-n} \\
&= r_k(x-1)\left[\binom{x-2}{x-n-1}e^{-z}(1-e^{-z})^{x-n-1}\right] - r_k x\left[\binom{x-1}{x-n}e^{-z}(1-e^{-z})^{x-n}\right] \\
&= r_k(x-1)P(x-1,t) - r_k x P(x,t), \qquad (52)
\end{aligned}$$

这等于方程 (32) 的右侧。 □

5.1.3 定理 2 的证明

证明。我们通过对k进行数学归纳法来证明定理 (2)。

- $k = 1$: 只有一个阶段，其扩增效率为$r1$。回顾一下，动力学中$P(x,t)$所满足的 Chapman-Kolmogorov 前向方程由 (32) 给出，其初始条件为

$$P(x,0) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

使用与上述相同的论证，我们可以将$P(x,t)$的生成函数写为

$$G(s,t) = \Psi\left(\frac{s-1}{s}e^{r_1 t}\right), \qquad (53)$$

具有初始条件

$$G(s,0) = \Psi\left(\frac{s-1}{s}\right) = \sum_{x=0}^{\infty} s^x P(x,0) = e^{\lambda(s-1)}. \qquad (54)$$

因此，

$$\begin{aligned}
G(s,t) &= \Psi\left(\frac{s-1}{s}e^{r_1 t}\right) \\
&= e^{\lambda\left(\frac{1}{1-\frac{s-1}{s}e^{r_1 t}}-1\right)} \\
&= e^{\left(\frac{\lambda(s-1)}{1-s\left(1-e^{-r_1 t}\right)}\right)}. \qquad (55)
\end{aligned}$$

当$k = 1$时，方程 (55) 与 (15) 相符。

推论5. 由(55) 给出的生成函数满足方程(35)。

证明. 对s求右侧关于s的导数。 □

- $k = 2$:

有两个阶段，其扩增效率分别为r_1 和r_2，。第一阶段从时间$t = 0$ 持续到$t = \tau_1$，第二阶段从$t = \tau_1$ 持续到$t = \tau_1 + \tau_2$。与之前一样，对于$t \in I_2$，概率生成函数具有如下形式

$$G(s,t) = \Psi\left(\frac{s-1}{s}e^{r_2(t-\tau_1)}\right),$$

初始条件（在时间$t = t_1$）

$$G(s,\tau_1) = \Psi\left(\frac{s-1}{s}\right) = e^{\left[\frac{\lambda(s-1)}{1-s\left(1-e^{-r_1 t}\right)}\right]}.$$

因此，

$$\begin{aligned} G(s,t) &= \Psi\left(\frac{s-1}{s}e^{r_2(t-\tau_1)}\right) \\ &= e^{\left[\frac{\lambda\left(\frac{1}{1-\frac{s-1}{s}e^{r_2(t-\tau_1)}}-1\right)}{1-\frac{1}{1-\frac{s-1}{s}e^{r_2(t-\tau_1)}}\left(1-e^{-r_1 t}\right)}\right]} \\ &= e^{\left[\frac{\lambda(s-1)}{1-s\left(1-e^{-(r_2 t+(r_1-r_2)\tau_1)}\right)}\right]}. \end{aligned} \tag{56}$$

当$k = 2$时，(56) 的右边等于 (15)。

- 我们假设该命题对$t \in I_k$成立，即

$$G(s,t) = e^{\left[\frac{\lambda(s-1)}{1-s(1-e^{-z})}\right]},$$

其中$z = r_k t + \sum_{i=1}^{k-1}(r_i - r_k)\tau_i$，我们将证明该命题对$t \in I_{k+1}$成立。

在$k+1$阶段，概率生成函数的形式为

$$G(s,t) = \Psi\left(\frac{s-1}{s}e^{r_{k+1}(t-\sum_{i=1}^{k}\tau_i)}\right).$$

根据归纳步骤，初始条件（在$t = \sum_{i=1}^{k}\tau_i$时）为

$$G(s,t) = \Psi\left(\frac{s-1}{s}\right) = e^{\left[\frac{\lambda(s-1)}{1-s(1-e^{-z})}\right]}.$$

因此，对于$t \in I_{k+1}$，我们有

$$\begin{aligned} G(s,t) &= \Psi\left(\frac{s-1}{s}e^{r_{k+1}(t-\sum_{i=1}^{k}\tau_i)}\right) \\ &= e^{\left[\frac{\lambda(\frac{1}{1-\frac{s-1}{s}e^{r_{k+1}(t-\sum_{i=1}^{k}\tau_i)}}-1)}{1-\frac{1}{1-\frac{s-1}{s}e^{r_{k+1}(t-\sum_{i=1}^{k}\tau_i)}}(1-e^{-z})}\right]} \\ &= e^{\left[\frac{\lambda(s-1)}{1-s(1-e^{-z'})}\right]}, \end{aligned} \tag{57}$$

其中 $z' = r_{k+1}t + \sum_{i=1}^{k}(r_i - r_{k+1})\tau_i$，至此定理 2 的证明结束。

□

5.1.4 推论 2 的证明

证明。 回顾一下，$P(x|\lambda, \vec{r}, t, \vec{\tau})$ 是定理 2 中给出的概率生成函数的幂级数展开中 s^x 的系数，即

$$P(x|\lambda, \vec{r}, t, \vec{\tau}) = \frac{1}{x!}\frac{\partial^x G(s,t)}{\partial s^x}\Big|_{s=0}. \tag{58}$$

现在，让我们计算 $G(s,t)$ 关于 s 的偏导数。为简洁起见，我们设

$$a = e^{-z}, v = (1-e^{-z}) = (1-a), u = \lambda e^{-z} = a\lambda, \text{ and } Q(s,t) = 1 - s(1-e^{-z}) = 1 - sv.$$

注意到

$$G(s,t) = e^{\frac{\lambda(s-1)}{Q}}, Q(0,t) = 1, G(0,t) = e^{-\lambda}, \frac{\partial Q(s,t)}{\partial s} = \frac{\partial Q(s,t)}{\partial s}\Big|_{s=0} = -v$$

and

$$
\begin{aligned}
\frac{\partial}{\partial s}G(s,t) &= u\frac{G(s,t)}{Q^2} \implies \frac{\partial}{\partial s}G(s,t)\Big|_{s=0} = ue^{-\lambda},\\
\frac{\partial^2}{\partial s^2}G(s,t) &= \frac{ue^{-\lambda}}{Q^4}\left[\frac{Q^2uG(s,t)}{Q^2} + 2vQG(s,t)\right] = u\left[u+2vQ\right]\frac{G(s,t)}{Q^4}\\
&\implies \frac{\partial^2}{\partial s^2}G(s,t)\Big|_{s=0} = e^{-\lambda}\left[u^2+2uv\right],\\
\frac{\partial^3}{\partial s^3}G(s,t) &= \frac{u\left[u+2v\right]}{Q^8}\left[uG(s,t)(u+2vQ)Q^2 - 2v^2G(s,t)Q^4 + 4vG(s,t)(u+2vQ)Q^3\right]\\
&= u\frac{G(s,t)}{Q^6}\left[u^2+6uvQ+6v^2Q^2\right]\\
&\implies \frac{\partial^3}{\partial s^3}G(s,t)\Big|_{s=0} = e^{-\lambda}(u^3+6u^2v+6uv^2),\\
\frac{\partial^4}{\partial s^4}G(s,t) &= u\frac{G(s,t)}{Q^8}\left[u^3+12u^2vQ+36uv^2Q^2+24v^3Q^3\right]\\
&\implies \frac{\partial^4}{\partial s^4}G(s,t)\Big|_{s=0} = e^{-\lambda}\left[u^4+12u^3v+36u^2v^2+24uv^3\right]\\
\frac{\partial^5}{\partial s^5}G(s,t) &= u\frac{G(s,t)}{Q^8}\left[u^4+20u^3vQ+120u^2v^2Q^2+240uv^3Q^3+120v^4Q^4\right],\\
&\implies \frac{\partial^4}{\partial s^5}G(s,t)\Big|_{s=0} = e^{-\lambda}\left[u^5+20u^4v+120u^3v^2+240u^2v^3+120uv^4\right]. \qquad (59)
\end{aligned}
$$

仔细考察式 (59) 中项 u,v 和 uv 的幂次系数，可以得到以下组合三角形

x					
1	1				
2	1	2			
3	1	6	6		
4	1	12	36	24	
5	1	20	120	240	120
	1	2	3	4	5
			i		

特别地，该三角形的第 (x,i) 项为

$$
T(x,i) = \binom{x}{i-1}\binom{x-1}{i-1}(i-1)!, \text{ for } i = 1,2,\ldots x. \qquad (60)
$$

Thus

$$
\begin{aligned}
P(x|\lambda,\vec{r},t,\vec{\tau}) &= \frac{1}{x!}\frac{\partial^x}{\partial s^x}G(s,t)\Big|_{s=0} = \frac{e^{-\lambda}}{x!}\sum_{i=1}^{x}T(x,i) \\
&= \frac{e^{-\lambda}}{x!}\sum_{i=1}^{x}\binom{x}{i-1}\binom{x-1}{i-1}(i-1)!\,(\lambda e^{-z})^{x-i+1}\,(1-e^{-z})^{i-1} \\
&= e^{-\lambda}\sum_{i=1}^{x}\frac{1}{(x-i+1)!}\binom{x-1}{i-1}(\lambda e^{-z})^{x-i+1}\,(1-e^{-z})^{i-1}. \qquad (61)
\end{aligned}
$$

在 (61) 中令 $k = x - i + 1$，即可得到所需结果。 □

推论 6。 定理 *2* 中给出的概率分布满足（32）。

证明。 对方程（61）右侧关于t求导。 □

现在我们将推导Ct 值的概率密度函数（pdf）、均值、方差和累积分布函数（cdf）。我们将考虑两种不同的情形，即:

1. PCR 过程的初始状态是确定性的，且 PCR 阶段长度和扩增效率已知；以及

2. 初始状态服从泊松分布，且阶段长度和扩增效率已知。

5.1.5 情形 1：初始状态是确定性的，且阶段长度和扩增效率已知

- 概率密度函数的一般形式

  考虑定理 1 中描述的 PCR过程。该过程从n 个cDNA分子开始，在相应扩增效率 $\vec{r} = (r_1, r_2, \ldots, r_p)$下，最多经历$p$ 个连续阶段 $\{I_i\}$，各阶段长度分别为 $\vec{\tau} = (\tau_1, \tau_2, \ldots, \tau_p)$。根据定义，$Ct$ 值t 是分子数量达到定量阈值x时的时间。根据贝叶斯定理，t 的概率密度函数的一般表达式为

$$
P(t|n,\vec{r},\vec{\tau},x) = \frac{P(n,\vec{r},\vec{\tau},x|t)P(t)}{P(n,\vec{r},\vec{\tau},x)}. \qquad (62)
$$

  由于n 与 $\vec{r}$、 $\vec{\tau}$和t相互独立，且 $\vec{r}$ 也与t 以及 $\vec{\tau}$各项取值相互独立，因此我们将 (62) 右侧的分子改写为

如下：

$$\begin{aligned} P(n,\vec{r},\vec{\tau},x|t)P(t) &= P(x|n,\vec{r},\vec{\tau},t)P(n,\vec{r},\vec{\tau}|t)P(t) \\ &= P(x|n,\vec{r},\vec{\tau},t)P(n|\vec{r},\vec{\tau},t)P(\vec{r},\vec{\tau}|t)P(t) \\ &= P(x|n,\vec{r},\vec{\tau},t)P(n)P(\vec{r}|\vec{\tau},t)P(\vec{\tau}|t)P(t) \\ &= P(x|n,\vec{r},\vec{\tau},t)P(n)P(\vec{r})P(t|\vec{\tau})P(\vec{\tau}). \end{aligned} \tag{63}$$

同样，式（62）右侧的分母可简化为

$$\begin{aligned} P(n,\vec{r},\vec{\tau},x) &= \int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} P(n,\vec{r},\vec{\tau},x|t)P(t)dt \\ &= P(n)P(\vec{r})P(\vec{\tau})\int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} P(x|n,\vec{r},\vec{\tau},t)P(t|\vec{\tau})dt. \end{aligned} \tag{64}$$

因此，式（62）可改写为

$$P(t|n,\vec{r},\vec{\tau},x) = \frac{P(x|n,\vec{r},\vec{\tau},t)P(t|\vec{\tau})}{\int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} P(x|n,\vec{r},\vec{\tau},t)P(t|\vec{\tau})dt}. \tag{65}$$

我们将推导具有任意阶段数Ct 值的概率密度函数、均值、方差和累积分布函数p，适用于$t \in I_k, k \leq p$的 PCR 过程。不失一般性，我们假设。我们将假设给定 $\vec{\tau}$ 时t 服从均匀先验密度。正如我们稍后将要展示的，这一假设所得到的结果与假设采用 Jeffreys 先验 [25] 时非常相似。对于单阶段 PCR 过程，如果无法从一般结果中直接得出相应结果，我们将单独陈述。

我们首先考虑中间阶段长度已知的情况，这些长度记录在向量 $\vec{\tau}$ 中。例如，当我们有兴趣根据数据估计这些阶段的长度时，这种处理就很有用。然后，我们将说明如何将 $\vec{\tau}$ 从概率密度函数中边缘化。

- 概率密度函数

利用方程（65）给出的后验密度以及推论1中给出的似然函数，我们得到概率密度函数的以下函数形式：

$$P(t|n,\vec{r},\vec{\tau},x) \propto e^{-nz}(1-e^{-z})^{x-n}, \tag{66}$$

其中

$$z = r_k t + \sum_{i=1}^{k-1}(r_i - r_k)\tau_i, \tag{67}$$

$\vec{r} = (r_1, r_2, \ldots, r_k)$ 是扩增效率向量，$\vec{\tau} = (\tau_1, \tau_2, \ldots, \tau_k)$ 是一个

阶段长度向量，并且我们对t采用了均匀先验。

归一化常数为

$$C = \int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} e^{-nz}(1-e^{-z})^{x-n}dt. \tag{68}$$

令 $w = e^{-z}$. 则，

$$\begin{aligned} C &= \frac{1}{r_k}\int_0^{\theta} w^{n-1}(1-w)^{x-n}dw \\ &= \frac{B_\theta(n, x-n+1)}{r_k}, \end{aligned} \tag{69}$$

其中

$$\theta = e^{-\sum_{i=1}^{k-1} r_i\tau_i}. \tag{70}$$

因此，概率密度函数为

$$P(t|n,\vec{r},\vec{\tau},x) = \frac{r_k e^{-nz}(1-e^{-z})^{x-n}}{B_\theta(n,x-n+1)}. \tag{71}$$

对于单相过程， $\theta = 1$，因此概率密度函数简化为

$$P(t|n,r_1,x) = \frac{r_1 e^{-nr_1 t}\left(1-e^{-r_1 t}\right)^{x-n}}{B(n,x-n+1)}. \tag{72}$$

请注意，在某些情况下（例如，不关心各个 PCR 扩增阶段的长度时），将 $\vec{\tau}$ 从 $P(t|n,\vec{r},\vec{\tau},x)$ 中边缘化可能会很有用。这可以利用以下事实来实现：

$$\begin{aligned} P(t|n,\vec{r},x) &= \int_\Omega P(t,\vec{\tau}|n,\vec{r},x)d\vec{\tau} \\ &= \int_\Omega P(t|n,\vec{r},\vec{\tau},x)P(\vec{\tau}|n,\vec{r},x)d\vec{\tau}, \end{aligned} \tag{73}$$

其中，Ω 是 $\vec{\tau}$ 的定义域。

此外，请注意，基于 Jeffreys [25] 提出的一种生成对重参数化保持不变的先验分布的方法，t 的先验分布还可以用以下形式表示：

$$p(t|\vec{\tau}) \propto \sqrt{|I(t|\vec{\tau})|}, \tag{74}$$

其中$I(t|\vec{\tau})$ 是似然函数的费 Fisher 信息，其定义如下：

$$\begin{aligned} I(t|\vec{\tau}) &= \mathbb{E}_X\left[\left(\frac{\partial}{\partial t}\ln P(x|n,\vec{r},\vec{\tau},x)\right)^2\right] \\ &= \mathbb{E}_X\left[\frac{r_k^2\left(w^2x^2-2nwx+n^2\right)}{(1-w)^2}\right] \\ &= \frac{r_k^2w^2}{(1-w)^2}\mathbb{E}_X\left[x^2\right]-\frac{2nr_k^2w}{(1-w)^2}\mathbb{E}_X\left[x\right]+\frac{n^2r_k^2}{(1-w)^2} \\ &= \frac{r_k^2}{(1-w)^2}\left[w^2\sum_{j=1}^{\infty}x^2\binom{x-1}{j-1}w^n(1-w)^{x-n}-2nw\sum_{j=1}^{\infty}x\binom{x-1}{j-1}w^n(1-w)^{x-n}+n^2\right], \end{aligned} \tag{75}$$

其中 $w = e^{-z}$。

注意到

$$\begin{aligned} \sum_{x=1}^{\infty}x\binom{x-1}{n-1}w^n(1-w)^{x-n} &= \sum_{x=1}^{\infty}\frac{x!}{(x-n)!(n-1)!}w^n(1-w)^{x-n} \\ &= \sum_{y=2}^{\infty}\frac{(y-1)!}{(y-m)!(m-2)!}w^{m-1}(1-w)^{y-m} \\ &= \frac{m-1}{w}\sum_{y=1}^{\infty}\frac{(y-1)!}{(y-m)!(m-1)!}w^m(1-w)^{y-m} \\ &= \frac{m-1}{w} \\ &= \frac{n}{w} \end{aligned} \tag{76}$$

and

$$\begin{aligned}
\sum_{x=1}^{\infty} x^2 \binom{x-1}{n-1} w^n (1-w)^{x-n} &= \sum_{x=1}^{\infty} \frac{x!x}{(x-n)!(n-1)!} w^n (1-w)^{x-n} \\
&= \sum_{y=2}^{\infty} \frac{(y-1)!(y-1)}{(y-m)!(m-2)!} w^{m-1} (1-w)^{y-m} \\
&= \sum_{y=1}^{\infty} \frac{(y-1)!y}{(y-m)!(m-2)!} w^{m-1} (1-w)^{y-m} - \frac{m-1}{w} \\
&= \sum_{y'=2}^{\infty} \frac{(y'-1)!}{(y'-m')!(m'-3)!} w^{m'-2} (1-w)^{y'-m'} - \frac{m-1}{w} \\
&= \frac{(m'-1)(m'-2)}{w^2} \sum_{y'=1}^{\infty} \frac{(y'-1)!}{(y'-m')!(m'-1)!} w^{m'} (1-w)^{y'-m'} - \frac{m-1}{w} \\
&= \frac{(m'-1)(m'-2)}{w^2} - \frac{m-1}{w} \\
&= \frac{n(n+1)}{w^2} - \frac{n}{w},
\end{aligned} \tag{77}$$

其中 $m = n+1, m' = m+1, y = x+1, y' = y+1$。

将 (76) 和 (77) 代入 (75)，我们得到

$$\begin{aligned}
I(t|\vec{\tau}) &= \frac{n r_k^2}{1-w} \\
&\Longrightarrow \quad p(t|\vec{\tau}) \propto \frac{1}{\sqrt{1-w}} = \frac{1}{\sqrt{1-e^{-z}}}.
\end{aligned} \tag{78}$$

使用这个先验分布，并按照之前推导 (71) 时所用的步骤，我们得到概率密度函数为

$$\begin{aligned}
P(t|n, \vec{r}, \vec{\tau}, x) &\propto e^{-nz} (1-e^{-z})^{x-n-1/2} \\
\Longrightarrow P(t|n, \vec{r}, \vec{\tau}, x) &= \frac{r_k e^{-nz} (1-e^{-z})^{x-n-1/2}}{B_\theta(n, x-n+1/2)},
\end{aligned} \tag{79}$$

其形式与 (71) 相似。

为简单起见，我们将继续对t使用均匀先验。

- 均值
  均值Ct 值为

$$\mathbb{E}(t) = \frac{r_k \int_{\sum_i^{k-1} \tau_i}^{\infty} t e^{-nz} (1-e^{-z})^{x-n} dt}{B_\theta(n, x-n+1)}, \tag{80}$$

其中z 由式（67）给出。

令 $w = e^{-z}$. 那么，

$$\begin{aligned}
\mathbb{E}(t) &= \frac{r_k \int_\theta^0 \left[-\frac{(\ln w - \ln \theta')}{r_k} \right] w^n (1-w)^{x-n} \left(-\frac{dw}{r_k w} \right)}{B_\theta(n, x-n+1)} \\
&= \frac{\int_\theta^0 (\ln w - \ln \theta') \, w^{n-1} (1-w)^{x-n} dw}{r_k B_\theta(n, x-n+1)} \\
&= \frac{\ln \theta' \int_0^\theta w^{n-1}(1-w)^{x-n} dw - \int_0^\theta \ln w \, w^{n-1}(1-w)^{x-n} dw}{r_k B_\theta(n, x-n+1)} \\
&= \frac{\ln \theta'}{r_k} - \frac{\left(\frac{\partial}{\partial n} + \frac{\partial}{\partial x} \right) B_\theta(n, x-n+1)}{r_k B_\theta(n, x-n+1)} \\
&= \frac{\ln \frac{\theta'}{\theta}}{r_k} + \frac{\Gamma(n)^2 \theta^n \, {}_3\tilde{F}_2(n, n, n-x; n+1, n+1; \theta)}{r_k B_\theta(n, x-n+1)} \\
&= \sum_{i=1}^{k-1} \tau_i + \frac{\Gamma(n)^2 \theta^n \, {}_3\tilde{F}_2(n, n, n-x; n+1, n+1; \theta)}{r_k B_\theta(n, x-n+1)},
\end{aligned} \tag{81}$$

其中 θ 由式（70）给出， $\psi(\cdot)$ 是第一多伽马函数（也称为双伽马函数），并且

$$\theta' = \theta e^{r_k \sum_{i=1}^{k-1} \tau_i}. \tag{82}$$

注意，对于单相过程， $\theta = \theta' = 1$。在这种情况下，利用

$${}_3\tilde{F}_2(n, n, n-x; n+1, n+1; 1) = \frac{n\Gamma(x-n+1)\left[\psi(x+1) - \psi(n)\right]}{\Gamma(n+1)\Gamma(x+1)},$$

我们可以得到，均值Ct 为

$$\mathbb{E}(t) = \frac{\psi(x+1) - \psi(n)}{r_1}. \tag{83}$$

- 方差

  Ct 的方差由 $\mathrm{E}(t^2)$ $-\mathrm{E}(t)^2$, 给出，其中

$$\mathbb{E}(t^2) = \frac{r_k \int_{\sum_{i=1}^{k-1} \tau_i}^\infty t^2 e^{-nz} (1-e^{-z})^{x-n} dt}{B_\theta(n, x-n+1)}, \tag{84}$$

并且z 由（67）给出。

令$w = e^{-z}$。那么，

$$\begin{aligned}
\mathbb{E}(t^2) &= \frac{r_k \int_\theta^0 \left[\frac{\ln w - \ln \theta'}{r_k}\right]^2 w^n (1-w)^{x-n} \left(-\frac{dw}{r_k w}\right)}{B_\theta(n, x-n+1)} \\
&= \frac{\int_0^\theta (\ln w - \ln \theta')^2 w^{n-1}(1-w)^{x-n} dw}{{r_k}^2 B_\theta(n, x-n+1)} \\
&= \frac{\int_0^\theta (\ln w)^2 w^{n-1}(1-w)^{x-n} dw - 2\ln\theta' \int_0^\theta \ln w\, w^{n-1}(1-w)^{x-n} dw}{{r_k}^2 B_\theta(n, x-n+1)} + \\
&\quad \frac{(\ln\theta')^2 \int_0^\theta w^{n-1}(1-w)^{x-n} dw}{{r_k}^2 B_\theta(n, x-n+1)} \\
&= \frac{\left(\frac{\partial^2}{\partial n^2} + 2\frac{\partial^2}{\partial n \partial x} + \frac{\partial^2}{\partial x^2} - 2\ln\theta'\left(\frac{\partial}{\partial n} + \frac{\partial}{\partial x}\right)\right) B_\theta(n, x-n+1)}{{r_k}^2 B_\theta(n, x-n+1)} + \frac{(\ln\theta')^2}{{r_k}^2} \\
&= \frac{\left(\frac{\partial^2}{\partial n^2} + 2\frac{\partial^2}{\partial n \partial x} + \frac{\partial^2}{\partial x^2}\right) B_\theta(n, x-n+1)}{{r_k}^2 B_\theta(n, x-n+1)} + \frac{2\ln\theta' \Gamma(n)^2 \theta^n\, {}_3\tilde{F}_2(n, n, n-x; n+1, n+1; \theta)}{r_k^2 B_\theta(n, x-n+1)} + \\
&\quad \frac{\ln\theta' \ln\frac{\theta'}{\theta^2}}{r_k^2} \\
&= \frac{\left(\frac{\partial^2}{\partial n^2} + 2\frac{\partial^2}{\partial n \partial x} + \frac{\partial^2}{\partial x^2}\right) B_\theta(n, x-n+1)}{{r_k}^2 B_\theta(n, x-n+1)} + \frac{2\ln\theta' \Gamma(n)^2 \theta^n\, {}_3\tilde{F}_2(n, n, n-x; n+1, n+1; \theta)}{r_k^2 B_\theta(n, x-n+1)} + \\
&\quad \left[\sum_{i=1}^{k-1} \tau_i\right]^2 - \left[\sum_{i=1}^{k-1} \frac{r_i \tau_i}{r_k}\right]^2,
\end{aligned} \tag{85}$$

其中 θ 由式 (70) 给出， θ' 由式 (82) 给出。

对于单相过程，Ct 值的二阶矩为

$$\begin{aligned}
\mathbb{E}(t^2) &= \frac{\left(\frac{\partial^2}{\partial n^2} + 2\frac{\partial^2}{\partial n \partial x} + \frac{\partial^2}{\partial x^2}\right) B(n, x-n+1)}{{r_k}^2 B(n, x-n+1)} \\
&= \frac{\psi_1(n) - \psi_1(x+1) + [\psi(x+1) - \psi(n)]^2}{{r_k}^2},
\end{aligned} \tag{86}$$

其中， $\psi_1(\cdot)$ 是二阶多伽马函数（也称为三伽马函数）。因此，方差为

$$\mathbf{Var}(t) = \frac{\psi_1(n) - \psi_1(x+1)}{{r_k}^2}. \tag{87}$$

- 累积分布函数

Ct 值的累积分布函数为

$$F(t|n,\vec{r},\vec{\tau},x) = \frac{r_k}{B_\theta(n,x-n+1)}\int_{\sum_{i=1}^{k-1}\tau_i}^{t} e^{-nz'}(1-e^{-z'})^{x-n}ds, \tag{88}$$

其中 $z' = r_k s + \sum_{i=1}^{k-1}(r_i - r_k)\tau_i$

令 $w = e^{-z'}$. 则,

$$\begin{aligned} F(t|n,\vec{r},\vec{\tau},x) &= \frac{\int_{e^{-z}}^{\theta} w^{n-1}(1-w)^{x-n}dw}{B_\theta(n,x-n+1)} \\ &= \frac{B_\theta(n,x-n+1) - B_{e^{-z}}(n,x-n+1)}{B_\theta(n,x-n+1)} \\ &= 1 - \frac{B_{e^{-z}}(n,x-n+1)}{B_\theta(n,x-n+1)}, \end{aligned} \tag{89}$$

其中 θ 由 70 给出。

对于单相过程，累积分布函数为

$$F(t|n,r_1,x) = 1 - I_{e^{-r_1 t}}(n,x-n+1), \tag{90}$$

其中 $I_{e^{-r_1 t}}(n,x-n+1) = \frac{B_{e^{-r_1 t}}(n,x-n+1)}{B(n,x-n+1)}$ 是正则化不完全贝塔函数。由于累积分布函数具有解析闭式表达式，我们可以使用高效的反变换法生成随机样本Ct 值，如下所示:

$$t = -\frac{\ln I^{-1}_{1-u}(n,x-n+1)}{r_1}, \tag{91}$$

其中 u 是从区间 (0,1) 均匀随机抽取的实数，且I^{-1}_{1-u} 是正则化不完全贝塔函数的反函数。要找到与分位数$q \in (0,1)$对应的Ct 值，只需将方程 (91) 中的u 替换为q。

- n 的概率分布

最后，我们推导单阶段 PCR 过程中的 n 的概率分布，记作 $P(n|r_1,t,x)$。该分布可用于根据测得的Ct 值估计n。正如正文所示，该分布还可用于计算 PCR 检测的检出限和定量限。以下步骤也可用于推导具有任意阶段数的 PCR 过程的$P(n|\vec{r},t,\vec{\tau})$，但无法像单阶段情形那样得到闭式结果。

根据贝叶斯定理，我们有

$$P(n|r_1,t,x) \propto \frac{w^n(1-w)^{x-n}}{B(n,x-n+1)}, \tag{92}$$

其中

$$w = e^{-r_1 t}. \tag{93}$$

归一化常数为

$$\begin{aligned} C &= \sum_{n=1}^{x} \frac{w^n (1-w)^{x-n}}{B(n, x-n+1)} \\ &= \sum_{n=1}^{x} \frac{x!\, w^n (1-w)^{x-n}}{(n-1)!\,(x-n)!}. \end{aligned} \tag{94}$$

设 $m = n-1$。则,

$$\begin{aligned} C &= \sum_{m=0}^{x-1} \frac{x!\, w^{m+1} (1-w)^{x-1-m}}{m!\,(x-1-m)!} \\ &= xw \sum_{m=0}^{x-1} \binom{x-1}{m} w^m (1-w)^{x-1-m} \\ &= xw. \end{aligned} \tag{95}$$

因此，我们有

$$P(n|r_1, t, x) = \frac{w^{n-1}(1-w)^{x-n}}{xB(n, x-n+1)}. \tag{96}$$

假设t 是由含有n 个输入分子的 PCR 过程产生的Ct 值。若在公式（96）中用$\hat{}n$ 替换 n ，那么该方程将给出：根据数据t，$\hat{}n$ 被作为n 的估计值所得出的概率。计算根据任何可能由含n个输入分子的 PCR 过程生成的数据t ，$\hat{}n$ 被作为n 的估计值所得出的概率很有用，例如可用于确定定量限。这个概率由下式给出

$$\begin{aligned} P(\hat{n}|n, r_1, x) &= \int_0^\infty P(\hat{n}, t|n, r_1, x)\, dt \\ &= \int_0^\infty P(\hat{n}|n, r_1, t, x)\, P(t|n, r_1, x)\, dt \\ &= \int_0^\infty P(\hat{n}|r_1, t, x)\, P(t|n, r_1, x)\, dt \\ &= r_1 \int_0^\infty \frac{w^{\hat{n}-1}(1-w)^{x-\hat{n}}}{xB(\hat{n}, x-\hat{n}+1)} \frac{w^n (1-w)^{x-n}}{B(n, x-n+1)} dt \\ &= \int_0^1 \frac{w^{\hat{n}+n-2}(1-w)^{2x-m-n}}{xB(\hat{n}, x-\hat{n}+1)B(n, x-n+1)} dw \\ &= \frac{B(\hat{n}+n-1, 2x-\hat{n}-n+1)}{xB(\hat{n}, x-\hat{n}+1)B(n, x-n+1)} = P(\hat{n}|n, x), \end{aligned} \tag{97}$$

其中，利用（96），我们假设在给定t的条件下，$\hat{n}$ 与n 条件独立。值得注意的是，（97）不依赖于r_1。

5.1.6 情形2：初始状态服从泊松分布，且阶段长度和扩增效率已给定

- 概率密度函数的一般形式

设t 为定理 2中所述PCR过程的Ct 值。该过程始于数量服从泊松分布、均值为 λ的输入DNA分子，这些分子在长度为 $\vec{\tau}=(\tau_1,\tau_2,\ldots,\tau_p)$的最多$p$ 个不同阶段中进行复制，各阶段的扩增效率为 $\vec{r}=(r_1,r_2,\ldots,r_p)$。如前所述，$t$ 是分子数量达到定量阈值的时间，我们将该阈值记为x。将t 的概率密度函数记为$P(t|\lambda,\vec{r},\vec{\tau},x)$。根据贝叶斯定理，有

$$P(t|\lambda,\vec{r},\vec{\tau},x) = \frac{P(\lambda,\vec{r},\vec{\tau},x|t)P(t)}{P(\lambda,\vec{r},\vec{\tau},x)} \tag{98}$$

然而，λ 与 $\vec{r}$、$\vec{\tau}$和t相互独立，而 $\vec{r}$ 也与t 以及 $\vec{\tau}$各项所取的确切值相互独立。因此，按照我们此前推导（65）式时所采用的相同步骤，可得

$$P(t|\lambda,\vec{r},\vec{\tau},x) = \frac{P(x|\lambda,\vec{r},t,\vec{\tau})P(t|\vec{\tau})}{\int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} P(x|\lambda,\vec{r},t,\vec{\tau})P(t|\vec{\tau})dt}. \tag{99}$$

我们将假设$t \in I_k$（不失一般性），以推导概率密度函数、均值、方差和累积分布函数；随后，我们将在$t \in I_1$这一便于说明的情形下，给出结果所呈现的函数形式。与前面一样，为简单起见，我们将对t使用均匀先验密度。

- 概率密度函数

我们利用方程（99）给出的一般表达式，结合定理中给出的分子数的概率分布，推导Ct 值t 的概率密度函数

2 作为似然。具体而言，

$$
\begin{aligned}
P(t|\lambda,\vec{r},\vec{\tau},x) &\propto (1-e^{-z})^x \sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{\lambda e^{-z}}{1-e^{-z}}\right)^j && (100)\\
&= (1-e^{-z})^x \sum_{j=0}^{x-1} \frac{\binom{x-1}{j}}{(j+1)!}\left(\frac{\lambda e^{-z}}{1-e^{-z}}\right)^{j+1} && (101)\\
&\overset{\text{since } \binom{x}{j}=0 \text{ for } j>x}{=} (1-e^{-z})^x \sum_{j=0}^{\infty} \frac{\binom{x-1}{j}}{(j+1)!}\left(\frac{\lambda e^{-z}}{1-e^{-z}}\right)^{j+1} && (102)\\
&= \lambda e^{-z}(1-e^{-z})^{x-1} \sum_{j=0}^{\infty} \frac{(x-1)(x-2)\cdots(x-j)}{(j+1)!\,j!}\left(\frac{\lambda e^{-z}}{1-e^{-z}}\right)^j && (103)\\
&= \lambda e^{-z}(1-e^{-z})^{x-1} \sum_{j=0}^{\infty} \frac{(1-x)(2-x)\cdots(j-x)}{(j+1)!\,j!}\left(\frac{-\lambda e^{-z}}{1-e^{-z}}\right)^j && (104)\\
&= \lambda e^{-z}(1-e^{-z})^{x-1} \sum_{j=0}^{\infty} \frac{(1-x)_j}{(2)_j}\frac{\left(\frac{-\lambda e^{-z}}{1-e^{-z}}\right)^j}{j!} && (105)\\
&= \lambda e^{-z}(1-e^{-z})^{x-1}\, {}_1F_1\left(1-x;2;\frac{-\lambda e^{-z}}{1-e^{-z}}\right), && (106)
\end{aligned}
$$

其中z 由（67）给出，$1F_1$ 是超几何函数（也称为第一类库默合流超几何函数），定义如下

$$
{}_1F_1\left(1-x;2;\frac{-\lambda e^{-r_1 t}}{1-e^{-r_1 t}}\right) = \sum_{j=0}^{\infty} \frac{(1-x)_j}{(2)_j}\frac{\left[\frac{-\lambda e^{-r_1 t}}{1-e^{-r_1 t}}\right]^j}{j!},
$$

且 $(\alpha)_j$表示上升阶乘，即 $(\alpha)_j = \alpha(\alpha+1)(\alpha+2)\ldots(\alpha+j-1)$，其中 $(\alpha)_0 = 1$。归一化常数为

$$
C = \sum_{j=1}^{x} \frac{\binom{x-1}{j-1}\lambda^j}{j!} \int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} e^{-jz}(1-e^{-z})^{x-j}dt. \qquad (107)
$$

令 $w = e^{-z}$。则，

$$
\begin{aligned}
C &= \frac{1}{r_k}\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}\lambda^j}{j!} \int_0^{\theta} w^{j-1}(1-w)^{x-j}dw\\
&= \frac{1}{r_k}\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}\lambda^j}{j!} B_\theta(j, x-j+1), && (108)
\end{aligned}
$$

其中 θ 由式（70）给出。

因此，概率密度函数为

$$\begin{aligned} P(t|\lambda,\vec{r},\vec{\tau},x) &= \frac{r_k(1-e^{-z})^x \sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{\lambda e^{-z}}{1-e^{-z}}\right)^j}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)} \\ &= \frac{r_k\lambda e^{-z}(1-e^{-z})^{x-1}\,{}_1F_1\left(1-x,2,\frac{-\lambda e^{-z}}{1-e^{-z}}\right)}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)}, \end{aligned} \tag{109}$$

回想一下，对于单相过程， $\theta = 1$，因此我们有

$$\begin{aligned} \sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B(j,x-j+1) &= \sum_{j=1}^{\infty}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B(j,x-j+1) \\ &= \sum_{j=0}^{\infty}\frac{\binom{x-1}{j}\lambda^{j+1}}{(j+1)!}B(j+1,x-j) \\ &= \sum_{j=0}^{\infty}\frac{\lambda^{j+1}}{(j+1)!} \\ &= \frac{e^\lambda-1}{x}, \end{aligned} \tag{110}$$

这意味着概率密度函数为

$$P(t|\lambda,r_1,x) = \frac{r_1 x\lambda e^{-r_1 t}(1-e^{-r_1 t})^{x-1}\,{}_1F_1\left(1-x,2,\frac{-\lambda e^{-r_1 t}}{1-e^{-r_1 t}}\right)}{e^\lambda-1}. \tag{111}$$

请注意，在某些情况下（例如，当不关心各个 PCR 扩增阶段的长度时），对 $P(t|\lambda,\vec{r},\vec{\tau},x)$ 中的 $\vec{\tau}$ 进行边缘化可能会有用。利用以下事实即可实现这一点：

$$\begin{aligned} P(t|\lambda,\vec{r},x) &= \int_\Omega P(t,\vec{\tau}|\lambda,\vec{r},x)d\vec{\tau} \\ &= \int_\Omega P(t|\lambda,\vec{r},\vec{\tau},x)P(\vec{\tau}|\lambda,\vec{r},x)d\vec{\tau}, \end{aligned} \tag{112}$$

其中，Ω 是 $\vec{\tau}$ 的定义域。

- 均值

均值Ct 值为

$$\mathbb{E}(t) = \frac{r_k}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)}\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}\overbrace{\int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} te^{-jz}(1-e^{-z})^{x-j}dt}^{(D)}, \tag{113}$$

其中z 由（67）给出。

令 $w = e^{-z}$。那么，

$$D \overset{\text{see (81)}}{=} \frac{B_\theta(j,x-j+1)\sum_{i=1}^{k-1}\tau_i}{r_k} + \frac{\Gamma(j)^2\theta^j\,{}_3\tilde{F}_2(j,j,j-x;j+1,j+1;\theta)}{r_k^2}. \tag{114}$$

因此

$$\mathbb{E}(t) = \frac{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}\left[r_k B_\theta(j,x-j+1)\sum_{i=1}^{k-1}\tau_i + \Gamma(j)^2\theta^j\,{}_3\tilde{F}_2(j,j,j-x;j+1,j+1;\theta)\right]}{r_k\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)}. \tag{115}$$

回想一下，对于单相过程，$\theta = \theta' = 1$，所以

$$\mathbb{E}(t) = \frac{\psi(x+1)}{r_1} - \frac{\sum_{j=1}^{x}\frac{\lambda^j}{j!}\psi(j)}{r_1\left(e^\lambda - 1\right)}. \tag{116}$$

- 方差

  方差为 $\mathrm{E}(t^2) - \mathrm{E}(t)^2$, 其中

$$\mathbb{E}(t^2) = \frac{r_k}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)}\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}\overbrace{\int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} t^2e^{-jz}(1-e^{-z})^{x-j}dt}^{(D)}, \tag{117}$$

其中z 由（67）给出。

令$w = e^{-z}$. 那么,

$$D \overset{\text{see (85)}}{=} \frac{\left(\frac{\partial^2}{\partial n^2} + 2\frac{\partial^2}{\partial n \partial x} + \frac{\partial^2}{\partial x^2}\right) B_\theta(n, x-n+1)}{{r_k}^3} + \frac{2\ln\theta' \Gamma(n)^2 \theta^n \, {}_3\tilde{F}_2(n,n,n-x;n+1,n+1;\theta)}{r_k^3} + B_\theta(j, x-j+1)\left(\frac{\left(\sum_{i=1}^{k-1}\tau_i\right)^2 - \left(\sum_{i=1}^{k-1}\frac{r_i\tau_i}{r_k}\right)^2}{r_k}\right) \tag{118}$$

其中 θ' 由式 (82) 给出。

将式 (118) 代入式 (117)，得到

$$\mathbb{E}(t^2) = \frac{1}{r_k^2 \sum_{j=1}^{x} \frac{\binom{x-1}{j-1}\lambda^j}{j!} B_\theta(j, x-j+1)} \sum_{j=1}^{x} \frac{\binom{x-1}{j-1}\lambda^j}{j!} \left[\left(\frac{\partial^2}{\partial n^2} + 2\frac{\partial^2}{\partial n \partial x} + \frac{\partial^2}{\partial x^2}\right) B_\theta(n, x-n+1) + 2\ln\theta' \Gamma(n)^2 \theta^n \, {}_3\tilde{F}_2(n,n,n-x;n+1,n+1;\theta) + r_k^2 B_\theta(j, x-j+1)\left(\left(\sum_{i=1}^{k-1}\tau_i\right)^2 - \left(\sum_{i=1}^{k-1}\frac{r_i\tau_i}{r_k}\right)^2\right)\right]. \tag{119}$$

对于单相过程，方差为

$$\mathbf{Var}(t) = \frac{(e^\lambda - 1)\sum_{j=1}^{x} \frac{\lambda^j}{j!}\left[\psi_1(j) + \psi(j)^2\right] - \left(\sum_{j=1}^{x} \frac{\lambda^j}{j!}\psi(j)\right)^2}{\left(r_1(e^\lambda - 1)\right)^2} - \frac{\psi_1(x+1)}{r_1^2}. \tag{120}$$

- 累积分布函数

Ct 值的累积分布函数为

$$F(t|\lambda, \vec{r}, \vec{\tau}, x) = \frac{r_k \sum_{j=1}^{x} \frac{\binom{x-1}{j-1}\lambda^j}{j!} \int_{\sum_{i=1}^{k-1}\tau_i}^{t} e^{-jz'}(1-e^{-z'})^{x-j} ds}{\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}\lambda^j}{j!} B_\theta(j, x-j+1)}, \tag{121}$$

其中 $z' = r_k s + \sum_{i=1}^{k-1}(r_i - r_k)\tau_i$。

令 $w = e^{-z'}$。则，

$$\begin{aligned} F(t|\lambda,\vec{r},\vec{\tau},x) &= \frac{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}\int_{e^{-z}}^{\theta} w^{j-1}(1-w)^{x-j}dw}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)} \\ &= \frac{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}\left[\int_0^{\theta} w^{j-1}(1-w)^{x-j}dw - \int_0^{e^{-z}} w^{j-1}(1-w)^{x-j}dw\right]}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)} \\ &= \frac{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}\left[B_\theta(j,x-j+1) - B_{e^{-z}}(j,x-j+1)\right]}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)} \\ &= 1 - \frac{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_{e^{-z}}(j,x-j+1)}{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_\theta(j,x-j+1)}, \end{aligned} \tag{122}$$

其中 θ 由式（70）给出。

对于单相过程，利用式（110）简化累积分布函数，得到

$$F(t|\lambda,r_1,x) = 1 - \frac{x\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}\lambda^j}{j!}B_{e^{-r_1 t}}(j,x-j+1)}{e^{\lambda}-1}. \tag{123}$$

- λ的概率密度
  最后，我们推导单相过程中 λ的概率密度，$P(\lambda|r_1,t,x)$。该密度可用于根据测得的Ct值估计 λ，并计算PCR过程的检出限和定量限。下文所述步骤也可用于推导具有任意数量阶段的PCR过程的$P(\lambda|\vec{r},t,\vec{\tau})$。

根据贝叶斯定理，我们有

$$P(\lambda|r_1,t,x) \propto \frac{\sum_{j=1}^{x}\frac{\binom{x-1}{j-1}}{j!}\left(\frac{\lambda w}{1-w}\right)^j}{e^{\lambda}-1}, \tag{124}$$

其中$w = e^{-r_1 t}$。

归一化常数为

$$\begin{aligned} C &= \int_0^\infty \frac{\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{\lambda w}{1-w}\right)^j}{e^\lambda - 1} d\lambda \\ &= \sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{w}{1-w}\right)^j \int_0^\infty \frac{\lambda^j}{e^\lambda - 1} d\lambda \\ &= \sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{w}{1-w}\right)^j \Gamma(j+1)\zeta(j+1) \\ &= \sum_{j=1}^{x} \binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j \zeta(j+1), \end{aligned} \tag{125}$$

其中 $\zeta(j)$是黎曼ζ函数。

因此， λ 的概率密度为

$$\begin{aligned} P(\lambda|r_1,t,x) &= \frac{\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{\lambda w}{1-w}\right)^j}{(e^\lambda - 1)\sum_{j=1}^{x} \binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j \zeta(j+1)} \\ &= \frac{\lambda w \, {}_1F_1(1-x,2,\frac{-\lambda w}{1-w})}{(e^\lambda - 1)(1-w)\sum_{j=1}^{x} \binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j \zeta(j+1)}. \end{aligned} \tag{126}$$

λ 取值介于a 和b 之间的概率为

$$\begin{aligned} P(a \le \lambda \le b \,|\, r_1,t,x) &= \frac{\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{w}{1-w}\right)^j \int_a^b \frac{s^j}{e^s-1} ds}{\sum_{j=1}^{x} \binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j \zeta(j+1)} \\ &= \frac{\sum_{j=1}^{x} \binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j [\zeta_b(j+1) - \zeta_a(j+1)]}{\sum_{j=1}^{x} \binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j \zeta(j+1)}, \end{aligned} \tag{127}$$

其中 $\zeta_\lambda(\cdot)$是不完全黎曼ζ函数。

由此可得， λ 的累积分布函数为

$$F(\lambda \,|\, r_1,t,x) = \frac{\sum_{j=1}^{x} \binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j \zeta_\lambda(j+1)}{\sum_{j=1}^{x} \binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j \zeta(j+1)}. \tag{128}$$

如前文讨论$P(n|\vec{r},t,\vec{\tau},x)$ 时所述，方程（126）可以作如下解读：假设某个 PCR过程产生了一个Ct 值t ，其输入分子的期望数量为 λ。如果将 λ 替换为估计值 $\hat{\lambda}$，那么（126）给出的就是

$\hat{\lambda}$的似然。在实际应用中（例如确定定量限），计算这样一个概率很有用：对于任何由输入分子的期望数量为 λ 的 PCR过程所能产生的数据t，$\hat{\lambda}$ 将作为 λ 的估计值被得到。该概率为

$$
\begin{aligned}
P\left(\hat{\lambda}|\lambda, r_1, x\right) &= \int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} P\left(\hat{\lambda}, t|\lambda, r_1, x\right) dt \\
&= \int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} P\left(\hat{\lambda}|\lambda, r_1, t, x\right) P\left(t|\lambda, r_1, x\right) dt \\
&= \int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} P\left(\hat{\lambda}|r_1, t, x\right) P\left(t|\lambda, r_1, x\right) dt \\
&= \frac{r_1 x}{\left(e^{\lambda}-1\right)\left(e^{\hat{\lambda}}-1\right)} \int_{\sum_{i=1}^{k-1}\tau_i}^{\infty} \frac{(1-w)^x \left(\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{\lambda w}{1-w}\right)^j\right)\left(\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{\hat{\lambda} w}{1-w}\right)^j\right)}{\sum_{j=1}^{x}\binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j \zeta(j+1)} dt \\
&= \frac{x}{\left(e^{\lambda}-1\right)\left(e^{\hat{\lambda}}-1\right)} \int_{0}^{\theta} \frac{(1-w)^x \left(\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{\lambda w}{1-w}\right)^j\right)\left(\sum_{j=1}^{x} \frac{\binom{x-1}{j-1}}{j!}\left(\frac{\hat{\lambda} w}{1-w}\right)^j\right)}{w\sum_{j=1}^{x}\binom{x-1}{j-1}\left(\frac{w}{1-w}\right)^j \zeta(j+1)} dw,
\end{aligned}
\tag{129}
$$

其中，θ 由式 (70) 给出，并且我们假设在给定t的条件下，$\hat{\lambda}$ 与 λ 条件独立。

由此可得，$\hat{\lambda}$ 取值介于a 和b之间的、与t无关的概率为

$$
P(a \le \hat{\lambda} \le b \mid \lambda, r_1, x) = \int_a^b P\left(\hat{\lambda}|\lambda, r_1, x\right) d\hat{\lambda}. \tag{130}
$$

5.2 补充图

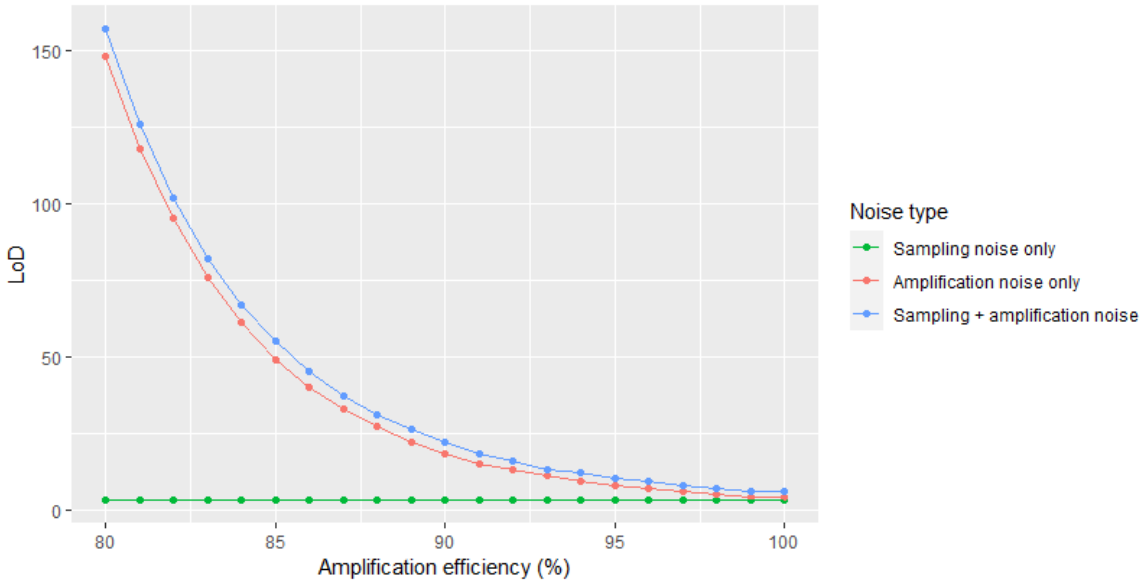


图 5.1：单相过程的检出限。计算检出限时，分别仅考虑抽样噪声（绿色实线）、仅考虑扩增噪声（红色实线），或同时考虑抽样噪声和扩增噪声（蓝色实线）。随后，以扩增效率为横轴绘制检出限，扩增效率以百分比表示，并采用以 2 为底的尺度。仅考虑抽样噪声时，检出限为 3；而同时考虑两种噪声时，检出限最高。在后一种情况下，当效率仅为 80% 时，检出限为 157；当效率为 100% 时，检出限为 6。图中显示，检出限高度依赖于效率。

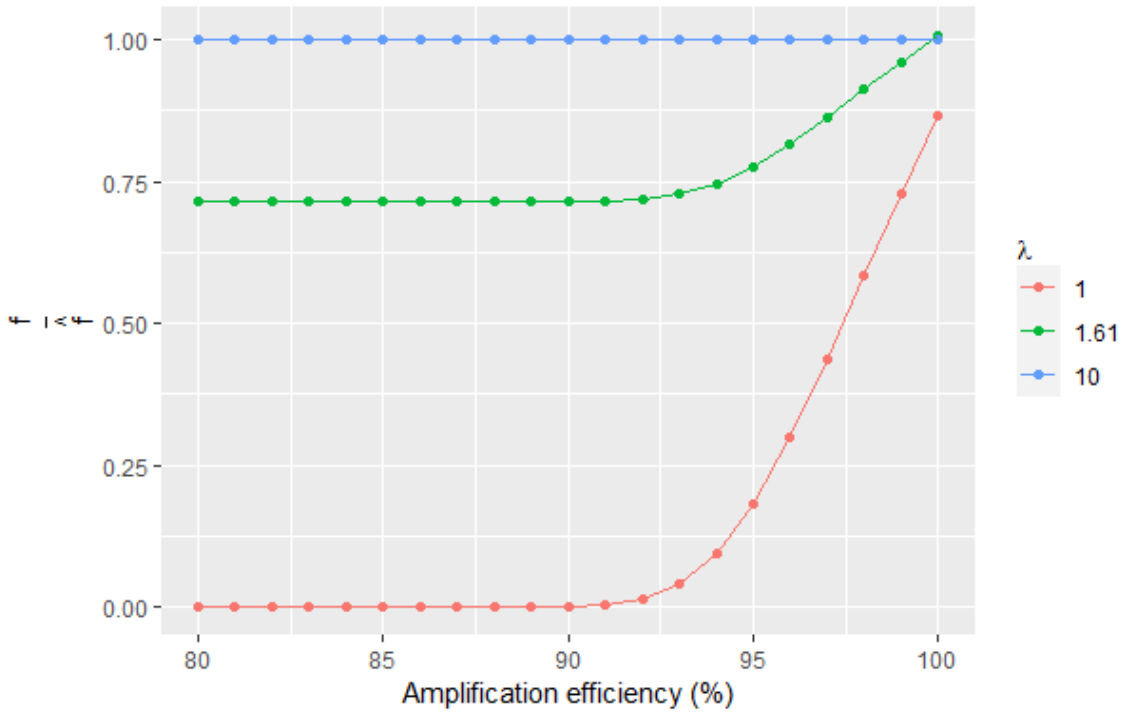


图5.2：数字PCR中阳性分区比例的期望值与估计值之比。对于不同的输入分子期望数量 λ，使用公式（24）计算数字PCR分区中预期呈阳性的比例，并将其除以基于泊松分布的标准估计值（即 $1-e^{-\lambda}$）。然后将结果与扩增效率的关系绘制成图。计算中使用的效率始终以 e 为底的尺度表示；为便于理解，将其转换为以 2 为底的尺度，并以百分比表示。

参考文献

[1] Roman W¨olfel、Victor M Corman、Wolfgang Guggemos、Michael Seil maier、Sabine Zange、Marcel A M¨uller、Daniela Niemeyer、Terry C Jones、Patrick Vollmar、Camilla Rothe 等。住院患者新冠肺炎的病毒学评估。《自然》，581(7809):465–469，2020。[2] Alice M Wang、Michael V Doyle 和 David F Mark。通过聚合酶链式反应定量信使 RNA。《美国国家科学院院刊》，86(24): 9717–9721，1989。[3] Randall K Saiki、Stephen Scharf、Fred Faloona、Kary B Mullis、Glenn T Horn、Henry A Erlich 和 Norman Arnheim。通过酶促扩增 β-珠蛋白基因组序列及限制性位点分析诊断镰状细胞性贫血。《科学》，230 (4732):1350–1354，1985。[4] Russell Higuchi、Carita Fockler、Gavin Dollinger 和 Robert Watson。动力学 PCR 分析：DNA 扩增反应的实时监测。《*Bio/technology*》，11(9): 1026–1030，1993。[5] Phenix-Lan Quan、Martin Sauzade 和 Eric Brouzes。数字 PCR：技术综述。《传感器》，18(4):1271，2018。[6] Tsugunori Notomi、Hiroto Okayama、Harumi Masubuchi、Toshihiro Yonekawa、Keiko Watanabe、Nobuyuki A mino 和 Tetsu Hase。DNA 的环介导等温扩增。《核酸研究》，28(12):e63–e63，2000。[7] Leon Mutesa、Pacifique Ndishimye、Yvan Butera、Jacob Souopgui、Annette Uwineza、Robert Rutayisire、Ella Larissa Ndoricimpaye、Emile Musoni、Nadine Rujeni、Thierry Nyatanyi 等。用于低流行率下识别 SARS-CoV-2 的混合检测策略。《自然》，589(7841):276–280，2021。[8] Kenneth J Livak 和 Thomas D Schmittgen。使用实时定量 PCR 和 2- $\delta\delta$ct 方法分析相对基因表达数据。*Methods*，25(4):402–408，2001。[9] Robert M Dorazio 和 Margaret E Hunter。数字聚合酶链式反应（数字 PCR）实验分析与设计的统计模型。《分析化学》，87(21):10886–10893，2015。

[10] Robert G Rutledge 和 Don Stewart。用于聚合酶链式反应的基于动力学的 S 形模型及其在高通量绝对实时定量 PCR 中的应用。《*BMC* 生物技术》，8(1):1–28，2008。

[11] Andrej-Nikolai Spiess、Caroline Feig 和 Christian Ritz。通过引入不对称性参数，实现对实时 PCR 数据的高精度 S 形拟合。*BMC* 生物信息学，9(1):1–12，2008。[12] Jan M Ruijter、Michael W Pfaffl、Sheng Zhao、Andrej N Spiess、Gregory Boggy、Jochen Blom、Robert G Rutledge、Davide Sisti、Antoon Lievens、Katleen De Preter 等。评估 qPCR 曲线分析方法，以可靠地发现生物标志物：偏倚、分辨率、精密度及其影响。*Methods*，59(1):32–46，2013。[13] C W Gardiner。物理、化学与自然科学随机方法手册。施普林格，德国柏林，2004。[14] H A Wilf。生成函数学。学术出版社，费城，美国，1994。[15] Neil JA Sloane 等。在线整数数列百科全书，2003。[16] Amin Forootan、Robert Sj¨oback、Jens Bj¨orkman、Bj¨orn Sj¨ogreen、Lucas Linz 和 Mikael Kubista。定量实时 PCR（qPCR）中检出限和定量限的确定方法。生物分子检测与定量，12：1–6，2017。[17] Sabine Nutz、Katharina D¨oll 和 Petr Karlovsky。通过受试者工作特征曲线分析确定实时 PCR 中的定量限：在轮枝镰刀菌和增殖镰刀菌 qPCR 检测中的应用。分析与生物分析化学，401(2):717–726，2011。[18] Nivedita Majumdar、Thomas Wessel 和 Jeffrey Marks。数字PCR建模，以实现最大灵敏度、动态范围和测量精密度。*PLOS ONE*，10(3): e0118833，2015。[19] Christopher M Hindson、John R Chevillet、Hilary A Briggs、Emily N Gallichotte、Ingrid K Ruf、Benjamin J Hindson、Robert L Vessella 和 Muneesh Tewari。微滴数字 PCR 与模拟实时 PCR 的绝对定量比较。*Nature Methods*，10 (10):1003–1005，2013。[20] Willard M Freeman、Stephen J Walker 和 Kent E Vrana。定量 RT-PCR：误区与潜力。*BioTechniques*，26(1):112–125，1999。[21] N. Chester 和 D.R Marshak。二甲基亚砜介导的引物 Tm 值降低：分析复性温度在聚合酶链式反应中作用的方法。分析生物化学，209:284–290，1993。doi：10.1006/abio.1993.1121。[22] Sean C Taylor、Katia Nadeau、Meysam Abbasi、Claude Lachance、Marie Nguyen 和 Joshua Fenrich。终极 qPCR 实验：首次即获得可发表、可重复的数据。*Trendsin Biotechnology*，37(7):761–774，2019。

[23] Kyle Siegrist。随机变量变换。https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/03:_Distributions/3.07:_Transformations_of_Random_Variables，2022。访问日期：2022-12-30。

[24] Kalpana Rengarajan、Stephen M Cristol、Milan Mehta 和 John M Nickerson。技术简报：使用荧光测定法定量DNA浓度：荧光团比较。*MolecularVision*，8:416–421，2002。

[25] Harold Jeffreys。估计问题中先验概率的一种不变形式。伦敦皇家学会会刊：*A*辑，数学与物理科学，186(1007):453–461，1946。