arXiv:2301.05459v1 [q-bio.PE] 13 Jan 2023

EWF：模拟 Wright–Fisher 扩散过程的精确路径

Jaromir Sant[1], Paul A. Jenkins[1,2,3], Jere Koskela[1], Dario Span`o[1]

统计系[1] 和计算机科学系[2]华威大学，考文垂 CV4 7AL，英国；艾伦·图灵研究所[3], 大英图书馆，伦敦 NW1 2DB，英国2023年1月16日

摘要

Wright–Fisher 扩散过程在群体遗传学中十分重要，可用于模拟选择、突变和遗传漂变等生物现象影响下等位基因频率随时间的演变。由于其转移密度的形式特殊，对该过程进行路径模拟颇具挑战。我们介绍 EWF，这是一种稳健高效的抽样器，可为扩散过程和扩散桥过程生成精确样本，并能处理包括具有频率依赖性的模型在内的各种选择模型。给定选择、突变和端点配置后，EWF 会根据相应 Wright–Fisher 过程的概率分布，返回指定采样时间上的样本。我们通过将输出与转移密度的近似结果进行比较，并采用柯尔莫哥洛夫–斯米尔诺夫检验和 QQ 图验证了输出。所有软件均可在 https://github.com/JaroSant/EWF 获取。

1 引言

Wright–Fisher 扩散是描述大型种群在随机交配以及存在突变和选择的情况下，等位基因频率随时间波动的核心模型。尽管该模型十分重要，但从计算角度来看，无论是在无选择的情况下（此时转移密度可表示为无穷级数展开），还是在非中性情形下（此时相应的无穷级数展开包含难以处理的项），都仍然难以应用。此外，在双等位基因模型中，扩散过程定义在有界区间 $[0,1]$ 上，因此，即使是欧拉–丸山格式这类简单的近似抽样技术，也需要经过复杂的修改才能符合其边界行为（Dangerfield 等人，2012）。文献中的现有方法通过结合离散化和数值近似来解决这一问题，例如，数值求解 Kolmogorov 后向方程

（Bollback 等，2008；Malaspinas 等，2012）、通过更易处理的过程进行近似（Mathieson 和 McVean，2013）、截断转移密度的谱展开（Steinr¨ucken 等，2016），以及使用黎曼和近似（Schraiber 等，2016）；这些方法都会引入难以量化的偏差。

在某些情况下，可以使用拒绝采样的精确采样程序。此类技术已扩展到 Wright–Fisher 扩散的某些变体：Jenkins 和 Span`o（2017年）表明，可以通过专为无穷级数设计的模拟技术，精确模拟中性 Wright–Fisher 扩散路径和扩散桥；还表明，中性路径是通过拒绝采样模拟非中性路径的自然提议机制。他们的研究假设突变参数严格为正，且扩散过程及扩散桥的端点均位于 $[0,1]$ 的内部。Griffiths 等（2018年）研究了起点和终点均为 0 的扩散桥，但起点、终点和参数的其他若干组合仍未涉及。此外，目前还没有能够实现所有相关情形的模拟软件包。

我们推出 EWF，这是一款 C++ 软件包，可对中性和非中性 Wright–Fisher 扩散过程进行精确抽样。该方法恰当处理所有类型的边界（进入边界、反射边界和吸收边界），涵盖范围广泛的选择模型，并允许任意端点，显著拓展了 Jenkins 和 Span`o（2017年）以及 Griffiths 等（2018年）的先前工作。相关的新理论细节见随附的补充材料。此外，与模拟路径上的误差会不断累积的欧拉–丸山法方案不同，EWF 在长时间范围内仍能保持准确性。

2 模型

考虑双等位基因非中性 Wright–Fisher 扩散 $(X_t)_{t\geq 0}$，其突变参数为 $\boldsymbol{\theta} = (\theta_1, \theta_2)$，该扩散过程由以下随机微分方程的解给出

$$\begin{aligned} dX_t = &\frac{1}{2}\left[\sigma X_t(1-X_t)\eta(X_t) - \theta_2 X_t + \theta_1(1-X_t)\right]dt \\ &+ \sqrt{X_t(1-X_t)}dW_t \end{aligned} \tag{1}$$

对于 $t \geq 0$ 且 $X_0 \in [0,1]$，以及 $\eta(x) = \sum_{i=0}^{n} a_i x^i$ 对于 n 有限（例如，对于基因选择 $\eta(x) = 1$，对于二倍体选择 $\eta(x) = h + x(1-2h)$，其中 h 为显性参数）。当突变参数 $\boldsymbol{\theta}$ 的各项均为正时，相应的中性（即 $\sigma = 0$）转移密度可以

分解为混合分布

$$p^{(\theta_1,\theta_2)}(x,y;t)=\sum_{m=0}^{\infty}q_m^{\theta}(t)\sum_{l=0}^{m}\mathrm{Bin}_{m,x}(l)\mathrm{Beta}_{\theta_1+l,\theta_2+m-l}(y),$$

其中，$(q_m^{\theta}(t))_{m\in\mathbb{N}}$ 是定义在整数上的分布，且 $\theta:=\theta_1+\theta_2$。这使得精确模拟成为可能（Jenkins 和 Span`o，2017年，第 2 节）。EWF 将这一方法扩展到了 $\theta_1=0$和/或 $\theta_2=0$ 情形，此时扩散过程几乎必然会在有限时间内于到达 0 和/或 1 时被吸收。

人们通常会关注一个在时间 t_0 出现的新生突变的演化，并在采样时间 $t>t_0$观察其在种群中的情况。如果 $\boldsymbol{\theta}=\mathbf{0}$，则需要对扩散过程在非吸收条件下进行条件化，以得到非退化的转移密度。所得密度见补充信息第 1 节（其中也包含相应的详细说明），其中还给出了 $\boldsymbol{\theta}=(0,\theta)$ 或 $\boldsymbol{\theta}=(\theta,0)$情况下相应的转移密度。

扩散桥的转移密度也可以用类似方法推导（见补充信息第 2 节）；而在存在选择时（即式 (1) 中的 $\sigma\neq0$ ），可以将中性路径作为候选路径，通过适当的拒绝方案进行模拟，从而得到相应非中性过程的样本（Jenkins 和 Span`o，2017年，第 5 节）。

3 方法

$p^{(\theta_1,\theta_2)}(x,y;\,t)$ 的表达式表明，可以通过以下步骤从转移密度中抽取样本：

1. 抽取 $M\sim\{q_m^{\theta}(t)\}_{m\in\mathbb{N}}$
2. 在给定 $M=m$ 的条件下，抽取 $L\sim\mathrm{Bin}(m,x)$
3. 在给定 $M=m$、$L=l$ 的条件下，抽取 $Y\sim$贝塔分布$(\theta_1+l,\theta_2+m-l)$

步骤 2 和 3 很简单。步骤 1 则更复杂，因为每个 $q_m^{\theta}(t)$ 都是一个无穷级数（参见补充信息第 3 节，其中我们扩展了该过程，以便在 $\boldsymbol{\theta}=\mathbf{0}$ 或 $\boldsymbol{\theta}=(0,\theta)$ 时生成样本）。

如果时间增量 t 较小，由于计算 $q_m^{\theta}(t)$时存在数值不稳定性，因此需要进行近似。对于较小的 t ，EWF 默认采用 $q_m^{\theta}(t)$的高斯近似（Griffiths，1984，定理 4）（$t\leq0.08$ ），在桥的后续时间增量低于某个阈值时也会使用类似的近似。详情请参见补充信息中的第 5 节。

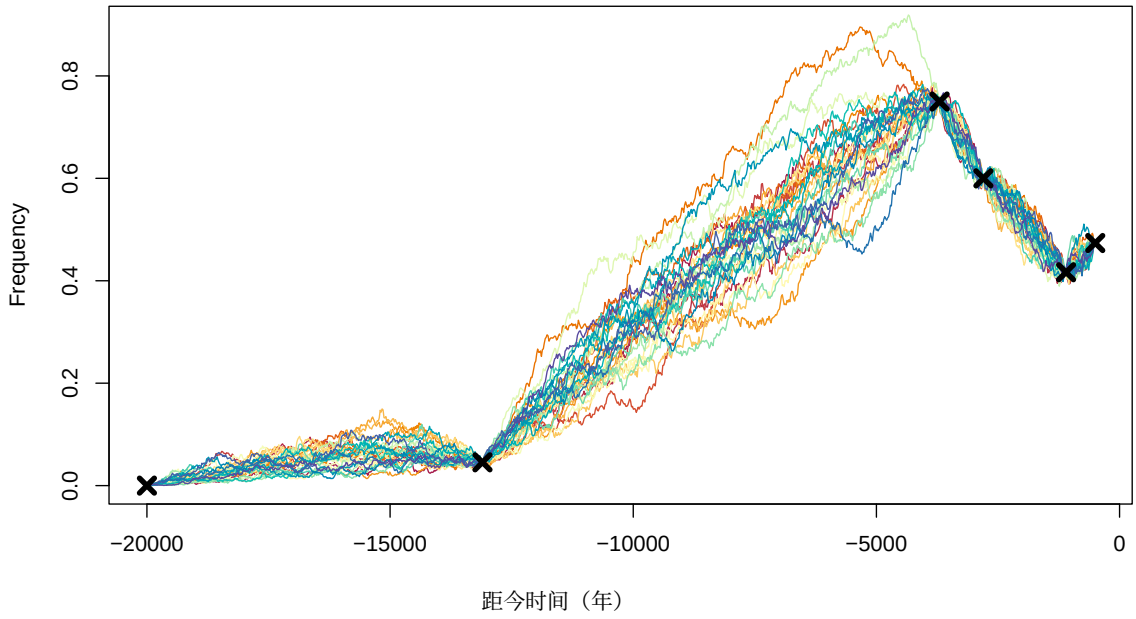


图 1：展示利用 EWF 模拟的马匹毛色数据的 30 条候选轨迹，数据见 Lud-wig 等（2009）（注意，假定观测到的频率（黑色十字）是对底层扩散过程的精确观测）。模拟使用推断出的选择系数 $s = 0.0007$，共识有效群体大小为 $N_e = 10{,}000$（Ludwig 等，2009；Malaspinas 等，2012；Schraiber 等，2016），得到 $\sigma = 2N_e s = 14$。我们使用了 $\boldsymbol{\theta} = \mathbf{0}$，世代时间为 5 年。

该实现经过了广泛测试，并通过 QQ 图和柯尔莫哥洛夫–斯米尔诺夫检验进行了验证（见补充信息第 7 节）。图 1 展示了一个示例。

4 讨论

EWF 提供了一种稳健、高效且精确的抽样程序，用于针对涵盖广泛选择机制、任意突变参数以及任意起点和终点的一大类 Wright–Fisher 扩散过程。该实现既可作为独立软件包使用，也可整合到基于时间序列等位基因频率数据的基于模拟的推断流程中。鉴于此类数据近年来日益丰富，这一实现尤其具有实用价值（Wutke 等，2016；Fages 等，2019）。

资助

本研究得到了英国工程与自然科学研究委员会和艾伦·图灵研究所的支持，资助编号为 EP/R044732/1、EP/V049208/1、EP/N510129/1。

补充信息

1 中性 Wright–Fisher 扩散过程的转移密度

考虑一个从任意初始点 $x \in [0,1]$ 出发的 Wright–Fisher 扩散过程，其中一个突变参数设为 0，例如 $\boldsymbol{\theta} = (0, \theta)$。在这种设定下，该扩散过程会存续至时刻 $T_0 := \inf\{t \geq 0 \colon X_t = 0\}$，此时它到达 0 并停留在那里。本节推导到达时间 T_0 允许在任意时刻发生时的转移密度，以及给定采样时间为 $\{t < T_0\}$时的转移密度。后一种情况稍难处理，因为必须纳入这一条件。

突变缺失的情形（即 $\boldsymbol{\theta} = \mathbf{0}$）也可用类似论证处理。此外，考虑对称映射 $x \mapsto 1 - x$ 并注意到所得过程仍是一个突变参数为 $\boldsymbol{\theta}' = (\theta_2, \theta_1)$、选择参数为 $\sigma' = -\sigma$的 Wright–Fisher 扩散过程，即可知情形 $\boldsymbol{\theta} = (\theta, 0)$可由情形 $\boldsymbol{\theta} = (0, \theta)$立即推出。

1.1 具有严格正突变的中性扩散

我们首先考虑 $\theta_1, \theta_2 > 0$，使得 0 和 1 都是非吸收边界。在这种情况下，转移密度可表示为（Griffiths（1979）；Tavaŕe（1984））：

$$p^{(\theta_1,\theta_2)}(x, y; t) = \sum_{m=0}^{\infty} q_m^{\theta}(t) \sum_{l=0}^{m} \mathcal{B}_{m,x}(l) \mathcal{D}_{\theta_1+l,\theta_2+m-l}(y), \tag{2}$$

其中，$\theta = |\boldsymbol{\theta}| = \theta_1 + \theta_2$，$\mathcal{B}_{m,\ x}(\cdot)$ 表示参数为 m 和 x 的二项概率质量函数，$\mathcal{D}_{\theta_1+l,\ \theta_2+m-l}(\cdot)$ 表示参数为 $\theta_1 + l$ 和 $\theta_2 + m - l$ 的贝塔概率密度函数，并且

$$q_m^{\theta}(t) := \sum_{k=m}^{\infty} (-1)^{k-m} \frac{\theta + 2k - 1}{k!(k-m)!} \frac{\Gamma(\theta + m + k - 1)}{\Gamma(\theta + m)} e^{\frac{-k(k+\theta-1)t}{2}},$$

其中 $\Gamma(\cdot)$ 表示伽马函数。我们指出，$\{q_m^\theta(t)\}_{m\in\mathbb{N}}$ 对应于 Kingman 合并过程中谱系数量的转移概率（该过程是 Wright–Fisher 扩散的矩对偶），因此 $q_m^\theta(t)$ 表示从时刻 0 的无限多个谱系出发时，有 m 个谱系存活至时刻 t 的概率。更多详情请参阅 Griffiths (1979)；Tavar´e (1984)。在式 (2) 左侧加入突变参数，明确表示转移密度对这些量的依赖性；不过，为了尽量简化表达，以下将省略这一记法，仅将扩散过程的转移密度写作 $p(x, y; t)$，具体考虑的突变机制则在外部指定。

1.2 带单侧突变的中性扩散

对于 $\boldsymbol{\theta} = (0, \theta)$，扩散过程在到达 0 时被吸收，转移密度可表示为

$$p(x,y;t) = \sum_{m=0}^{\infty} q_m^\theta(t) \left[\sum_{l=1}^{m} \mathcal{B}_{m,x}(l)\mathcal{D}_{l,\theta+m-l}(y) + (1-x)^m \delta_0(y) \right], \tag{3}$$

其中，$\delta_0(y)$ 表示位于 0 处的点质量，对应扩散过程在 0 处被吸收的情形。在这类情况下，我们对“密度”作适当的重新解释，即相对于一个支配测度而言，该测度包含勒贝格分量以及分别位于 0 和 1 处的原子。

若以事件 $\{t < T_0\}$ 为条件，根据标准条件概率，扩散过程在直到时间 t 之前未被吸收这一条件下的转移密度为

$$\tilde{p}(x,y;t) = \frac{p(x,y;t)}{\mathbb{P}_x\left[T_0 > t\right]},$$

对于 $y \in (0,1],$，我们使用记号 $\tilde{p}(\cdot,\cdot;\cdot)$，以明确表示这是条件扩散过程的转移密度。此外，我们有

$$\begin{aligned} \mathbb{P}_x\left[T_0 > t\right] &= \int_{(0,1]} p(x,u;t)du \\ &= \sum_{m=1}^{\infty} q_m^\theta(t) \sum_{l=1}^{m} \mathcal{B}_{m,x}(l), \end{aligned} \tag{4}$$

并且我们请注意，$m = 0$ 上述贡献缺失，因为相应的 ta

密度在 0 处坍缩为点质量。因此，对于 x， $y \in$ （$0,1]$ 我们有

$$\tilde{p}(x,y;t)=\sum_{m=1}^{\infty}\frac{q_m^{\theta}(t)}{\sum_{l=1}^{m}\mathcal{B}_{m,x}(l)\sum_{d=1}^{\infty}q_d^{\theta}(t)(1-(1-x)^d)}\mathcal{D}_{l,\theta+m-l}(y). \tag{5}$$

当 x 较小时，我们在 x 中得到以下首阶展开

$$p(x,y;t)=x\sum_{m=1}^{\infty}q_m^{\theta}(t)m(\theta+m-1)(1-y)^{\theta+m-2}+O(x^2), \tag{6}$$

并且进一步注意到，当 x 较小时，(4) 也是关于 x 的首阶项。因此，在 (5) 中取 $x\to 0$ 的极限，得到

$$\tilde{p}(0,y;t)=\sum_{m=1}^{\infty}\frac{mq_m^{\theta}(t)}{\sum_{d=1}^{\infty}dq_d^{\theta}(t)}\mathcal{D}_{1,\theta+m-1}(y). \tag{7}$$

综上可得，条件扩散过程的转移密度为

$$\tilde{p}(x,y;t)=\begin{cases}\displaystyle\sum_{m=1}^{\infty}\frac{mq_m^{\theta}(t)}{\sum_{d=1}^{\infty}dq_d^{\theta}(t)}\mathcal{D}_{1,\theta+m-1}(y) & x=0,\\[2ex] \displaystyle\sum_{m=1}^{\infty}\frac{q_m^{\theta}(t)\sum_{l=1}^{m}\mathcal{B}_{m,x}(l)}{\sum_{d=1}^{\infty}q_d^{\theta}(t)(1-(1-x)^d)}\mathcal{D}_{l,\theta+m-l}(y) & x\in(0,1].\end{cases} \tag{8}$$

需要指出的是，由于该扩散过程以不触及 0 为条件，因此其矩对偶金曼合并过程中始终至少有一条存续谱系，所以 m 的指标从 1 开始。

1.3 无突变扩散过程

若 $\boldsymbol{\theta}=\mathbf{0}$，则扩散过程在触及任一边界时都会被吸收，相应的转移密度为

$$p(x,y;t)=\sum_{m=2}^{\infty}q_m^{\theta}(t)\left[\sum_{l=1}^{m-1}\mathcal{B}_{m,x}(l)\mathcal{D}_{l,m-l}(y)+(1-x)^m\delta_0(y)+x^m\delta_1(y)\right], \tag{9}$$

将扩散过程限制在 $[0,1]$, 的内部区域内，并再次对所得分子和分母进行主阶分析，我们可以得出结论:

此时的转移密度为

$$\tilde{p}(x,y;t)=\begin{cases}\displaystyle\sum_{m=2}^{\infty}\frac{m q_m^0(t)}{\sum_{d=2}^{\infty} d q_d^0(t)}\mathcal{D}_{1,m-1}(y) & x=0,\\[2ex] \displaystyle\sum_{m=2}^{\infty}\frac{m q_m^0(t)}{\sum_{d=2}^{\infty} d q_d^0(t)}\mathcal{D}_{m-1,1}(y) & x=1,\\[2ex] \displaystyle\sum_{m=2}^{\infty}\frac{q_m^0(t)\sum_{l=1}^{m-1}\mathcal{B}_{m,x}(l)}{\sum_{d=2}^{\infty} q_d^0(t)(1-x^d-(1-x)^d)}\mathcal{D}_{l,m-l}(y) & x\in(0,1).\end{cases} \tag{10}$$

请注意，由于 $\boldsymbol{\theta}=\mathbf{0}$ 且我们以非吸收为条件，索引 m 和 d 现在必须从 2 开始。这是因为上述推导假设起点 x 位于 (0, 1) 内，随后从 (0, 1) 的内部将 x 趋近于相应的边界，这相当于从任意接近边界的位置开始扩散过程。因此，在任何时刻，种群中都有比例为 x 的一类个体，另有比例为 $1-x$ 的另一类个体，且两类都不会因突变而消失。

2 中性 Wright–Fisher 扩散桥的转移密度

现在，我们推导在时间 $s\in$从时间 0 时刻始于 x 、并在时间 t结束于 z 的 Wright–Fisher 扩散桥的分布中抽取的点 $y\in[0,1]$ 的密度。除了分别考虑每种突变机制外，我们还根据起点和终点 x与 z 所取的值进一步划分讨论。如同扩散过程的情形，我们推导 $\boldsymbol{\theta}=(0,\theta)$时的相关表达式，因为其他情形（$\boldsymbol{\theta}=(0,0)$ 或 $\boldsymbol{\theta}=(\theta,0)$）可用类似论证得出。我们还考虑以下两种情形：(i) 允许扩散桥在时间区间 $(0,t)$内的任意时刻被吸收；(ii) 扩散桥是条件非吸收的：对于所有 $s\in(0,t)$，$X_s\in(0,1)$。对于不同的端点组合，我们采用以下简写符号。

	$x=0$	$x=1$	$x\in(0,1)$
$z=0$	A1	B1	C1
$z\in(0,1)$	A2	B2	C2
$z=1$	A3	B3	C3

我们还在上述每种情形前加上一个字母，以作区分：$\boldsymbol{\theta}=\mathbf{0}$ （‘Z’表示零），$\boldsymbol{\theta}=(0,\theta)$（‘O’表示单侧），以及 $\boldsymbol{\theta}$ （各项均严格为正）（‘P’表示严格为正）。

在推导上述所有情形的转移密度之前，请注意，从时间 0 的 $x \in [0,1]$ 出发、在时间 t 结束于 $z \in [0,1]$、并在时间 s 采样的 Wright–Fisher 扩散桥，其转移密度可对 $y \in [0,1]$作如下因式分解：

$$p^{x,z;t}(y;s) = \frac{p(x,y;s)p(y,z;t-s)}{p(x,z;t)}, \tag{11}$$

同样，为简便起见，符号中省略了式 (11) 对突变参数的依赖。

2.1 单向突变 $\boldsymbol{\theta} = (0,\theta)$ 的中性扩散桥

首先注意，如果允许扩散桥在区间 $(0,t)$ 内的任意时刻在 0 处被吸收，那么值得关注的情况只有左端点 $x \in (0,1]$,；否则，该桥会一直停留在 0。此外，如果 $z \in (0,1]$, ，那么该桥不可能在时间区间 $(0,t)$ 内被吸收，因此这等价于以不发生吸收为条件（稍后将讨论这一点）。于是，我们取 $x \in (0,1)$ 和 $z = 0$，将式 (3) 代入式 (11)，并重新组合各项，得到

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^{\infty} q_d^\theta(t)(1-x)^d} \left[\sum_{l=1}^{m} \mathcal{B}_{m,x}(l) \frac{B(l,\theta+m-l+k)}{B(l,\theta+m-l)} \mathcal{D}_{l,\theta+m-l+k}(y) + (1-x)^m \delta_0(y) \right]. \tag{12}$$

其中 $B(\cdot,\cdot)$ 是贝塔函数。

为了推导 $z \in (0,1]$, 时的转移密度，我们首先指出，对扩散过程（或扩散桥）施加非吸收条件，是对该过程进行 h-变换的一种特殊情形（例如参见 Fitzsimmons 等人（1993）；Griffiths 等（2018））。此外，扩散桥在 h-变换下保持不变（参见 Griffiths 等（2018）中的公式（10）），因此，给定非吸收条件下的扩散桥分布，与相应未加条件的过程分布相同。因此，我们无须区分条件扩散桥与未加条件扩散桥的转移密度，并在全文中统一使用 $p^{x,\ z;t}(y;s)$。

将 (11) 在 $x, z \in (0,1]$ 下展开，得

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^{\infty} q_d^\theta(t) \sum_{f=1}^{d} \mathcal{B}_{d,x}(f) \mathcal{D}_{f,\theta+d-f}(z)} \times \sum_{l,j=1}^{m,k} \binom{k}{j} \frac{B(l+j, \theta+m-l+k-j)}{B(l, \theta+m-l)} \mathcal{D}_{j,\theta+k-j}(z) \mathcal{D}_{l+j,\theta+m-l+k-j}(y). \quad (13)$$

当 $x = 0$ 时，我们在上述分子和分母中都使用式 (6)，随后令 $x \to 0$，得到

$$\begin{aligned} p^{0,z;t}(y;s) &= \lim_{x\to 0} \left(\frac{x \sum_{m=1}^{\infty} q_m^\theta(s) m(\theta+m-1)(1-y)^{\theta+m-2} + o(x^2)}{x \sum_{d=1}^{\infty} q_d^\theta(t) d(\theta+d-1)(1-z)^{\theta+d-2} + o(x^2)} \right) p(y,z;t-s) \\ &= \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^{\infty} q_d^\theta(t) d(d+\theta-1)(1-z)^{\theta+d-2}} \\ &\quad \times \sum_{j=1}^{k} \binom{k}{j} \frac{B(j+1, \theta+m-1+k-j)}{B(1, \theta+m-1)} \mathcal{D}_{j,\theta+k-j}(z) \mathcal{D}_{j+1,\theta+m-1+k-j}(y). \end{aligned} \quad (14)$$

通过在 z 中取首阶项并令 $z \to 0$，还可以利用上述表达式推导出 $z = 0$ 时的表达式，得到

$$p^{0,0;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^{\infty} q_d^\theta(t) d(d+\theta-1)} \frac{m(m+\theta-1)k(k+\theta-1)}{(m+k+\theta-1)(m+k+\theta-2)} \mathcal{D}_{2,\theta+m-1+k-1}(y). \quad (15)$$

如前所述，情形 $\boldsymbol{\theta} = (\theta, 0)$ 可通过考虑对称映射 $x \mapsto 1-x$，由上述结果得到。

2.2 无突变的中性扩散桥

对于 $\boldsymbol{\theta} = \mathbf{0}$ 的情况，我们可以重复上述所有论证，从而得到：若 $x \in (0,1)$ 且 $z = 0$，则对于 $y \in [0,1)$，有

$$p^{x,0;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^0(s) q_k^0(t-s)}{\sum_{d=1}^{\infty} q_d^0(t)(1-x)^d} \left[\sum_{l=1}^{m-1} \mathcal{B}_{m,x}(l) \frac{B(l, m-l+k)}{B(l, m-l)} \mathcal{D}_{l,m-l+k}(y) + (1-x)^m \delta_0(y) \right] \quad (16)$$

而若 $x \in (0,1)$ 且 $z = 1$，则对于 $y \in (0, 1]$

$$p^{x,1;t}(y; s) = \sum_{m,k=1}^{\infty} \frac{q_m^0(s) q_k^0(t-s)}{\sum_{d=1}^{\infty} q_d^0(t) x^d} \left[\sum_{l=1}^{m-1} \mathcal{B}_{m,x}(l) \frac{B(l+k, m-l)}{B(l, m-l)} \mathcal{D}_{l+k,m-l}(y) + x^m \delta_1(y) \right] \quad (17)$$

注意，t 如果 $z = 0$，那么我们就不能有 $y = 1$；同理，如果 $z = 1$，y 就不能等于 0.

计算以非吸收为条件的转移密度，可以按照上文所示的单侧突变情形，采用相同的论证来完成。

三种突变机制下以条件化方式得到的扩散桥表达式如下（请回顾表2中的符号）。

2.3 当 $\boldsymbol{\theta}=\mathbf{0}$ 时的桥扩散转移密度

$$p^{x,z;t}(y;s)=\sum_{m,k=2}^{\infty}\frac{q_m^0(s)q_k^0(t-s)}{\sum_{d=2}^{\infty}q_d^0(t)d(d-1)}\frac{m(m-1)k(k-1)}{(m+k-1)(m+k-2)}\mathcal{D}_{2,m+k-2}(y) \tag{ZA1}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=2}^{\infty}\frac{q_m^0(s)q_k^0(t-s)}{\sum_{d=2}^{\infty}q_d^0(t)d\mathcal{D}_{1,d-1}(z)}m\sum_{j=1}^{k-1}\binom{k}{j}\frac{B(j+1,m-1+k-j)}{B(1,m-1)}\mathcal{D}_{j,k-j}(z)$$
$$\times\mathcal{D}_{j+1,m-1+k-j}(y) \tag{ZA2}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=2}^{\infty}\frac{q_m^0(s)q_k^0(t-s)}{2q_2^0(t)}m(m-1)k(k-1)B(k,m)\mathcal{D}_{k,m}(y) \tag{ZA3}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=2}^{\infty}\frac{q_m^0(s)q_k^0(t-s)}{2q_2^0(t)}m(m-1)k(k-1)B(m,k)\mathcal{D}_{m,k}(y) \tag{ZB1}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=2}^{\infty}\frac{q_m^0(s)q_k^0(t-s)}{\sum_{d=2}^{\infty}q_d^0(t)d\mathcal{D}_{d-1,1}}m\sum_{j=1}^{k-1}\binom{k}{j}\frac{B(m-1+j,k-j+1)}{B(m-1,1)}\mathcal{D}_{j,k-j}(y)$$
$$\times\mathcal{D}_{m-1+j,1+k-j}(y) \tag{ZB2}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=2}^{\infty}\frac{q_m^0(s)q_k^0(t-s)}{\sum_{d=2}^{\infty}q_d^0(t)d(d-1)}\frac{m(m-1)k(k-1)}{(m+k-1)(m+k-2)}\mathcal{D}_{m+k-2,2}(y) \tag{ZB3}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=2}^{\infty}\frac{q_m^0(s)q_k^0(t-s)}{\sum_{d=2}^{\infty}q_d^0(t)(d-1)\mathcal{B}_{d,x}(1)}\sum_{l=1}^{m-1}\mathcal{B}_{m,x}(l)k(k-1)\frac{B(l+1,m-l+k-1)}{B(l,m-l)}$$
$$\times\mathcal{D}_{l+1,m-l+k-1}(y) \tag{ZC1}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=2}^{\infty}\frac{q_m^0(s)q_k^0(t-s)}{\tilde{p}(x,z;t)}\sum_{l,j=1}^{m-1,k-1}\mathcal{B}_{m,x}(l)\binom{k}{j}\frac{B(l+j,m-l+k-j)}{B(l,m-l)}\mathcal{D}_{j,k-j}$$
$$\times\mathcal{D}_{l+j,m-l+k-j}(y) \tag{ZC2}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=2}^{\infty}\frac{q_m^0(s)q_k^0(t-s)}{\sum_{d=2}^{\infty}q_d^0(t)(d-1)\mathcal{B}_{d,x}(d-1)}\sum_{l=1}^{m-1}\mathcal{B}_{m,x}(l)k(k-1)\frac{B(l+k-1,m-l+1)}{B(l,m-l)}$$
$$\times\mathcal{D}_{l+k-1,m-l+1}(y) \tag{ZC3}$$

2.4 当 $\boldsymbol{\theta} = (0, \theta)$ 时的桥扩散转移密度

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^{\infty} q_d^\theta(t) d(d+\theta-1)} \frac{m(m+\theta-1)k(k+\theta-1)}{(m+k+\theta-1)(m+k+\theta-2)} \mathcal{D}_{2,\theta+m+k-2}(y) \quad OA1$$

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^{\infty} q_d^\theta(t) d\mathcal{D}_{1,\theta+d-1}(z)} m \sum_{j=1}^{k} \binom{k}{j} \frac{B(j+1, \theta+m-1+k-j)}{B(1, \theta+m-1)}$$
$$\times \mathcal{D}_{j,\ \theta+k-j}(z) \mathcal{D}_{j+1,\ \theta+m-1+k-j}(y) \quad OA2$$

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\theta q_1} m(m+\theta-1) \frac{B(k+1, \theta+m-1)}{B(k,\theta)} \mathcal{D}_{k+1,\theta+m-1}(y) \quad OA3$$

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\theta q_1} k(k+\theta-1) \frac{B(m+1, \theta+k-1)}{B(m,\theta)} \mathcal{D}_{m+1,\theta+k-1}(y) \quad OB1$$

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^{\infty} q_d^\theta(t) \mathcal{D}_{d,\theta}} \sum_{j=1}^{k} \binom{k}{j} \frac{B(m+j, \theta+k-j)}{B(m,\theta)} \mathcal{D}_{j,\theta+k-j}(z)$$
$$\times \mathcal{D}_{m+j,\theta+k-j}(y) \quad OB2$$

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^{\infty} q_d^\theta(t) \frac{1}{B(d,\theta)}} \frac{B(m+k,\theta)}{B(m,\theta)B(k,\theta)} \mathcal{D}_{m+k,\theta}(y) \quad OB3$$

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s) k(k+\theta-1)}{\sum_{d=1}^{\infty} q_d^\theta(t)(d+\theta-1)\mathcal{B}_{d,x}(1)} \sum_{l=1}^{m} \mathcal{B}_{m,x}(l) \frac{B(l+1, \theta+k-1+m-l)}{B(l, \theta+m-l)}$$
$$\times \mathcal{D}_{l+1,\theta+k-1+m-l}(y) \quad OC1$$

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\tilde{p}(x,z;t)} \sum_{l,j=1}^{m,k} \mathcal{B}_{m,x}(l) \binom{k}{j} \frac{B(l+j, \theta+m-l+k-j)}{B(l, \theta+m-l)} \mathcal{D}_{j,\theta+k-j}(z)$$
$$\times \mathcal{D}_{l+j,\theta+m-l+k-j} \quad OC2$$

$$p^{x,z;t}(y;s) = \sum_{m,k=1}^{\infty} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^{\infty} q_d^\theta(t) \frac{x^d}{B(d,\theta)}} \sum_{l=1}^{m} \mathcal{B}_{m,x}(l) \frac{B(l+k, \theta+m-l)}{B(l, \theta+m-l)B(k,\theta)} \mathcal{D}_{l+k,\theta+m-l}(y) \quad OC3$$

2.5 当 $\boldsymbol{\theta}=(\theta_1,\theta_2)$ 时的桥扩散转移密度

$$p^{x,z;t}(y;s)=\sum_{m,k=0}^{\infty}\frac{q_m^{\theta}(s)q_k^{\theta}(t-s)}{\sum_{d=0}^{\infty}q_d^{\theta}\frac{1}{B(\theta_1,\theta_2+d)}}\frac{B(\theta_1,\theta_2+m+k)}{B(\theta_1,\theta_2+m)B(\theta_1,\theta_2+k)}\mathcal{D}_{\theta_1,\theta_2+m+k}(y) \tag{PA1}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=0}^{\infty}\frac{q_m^{\theta}(s)q_k^{\theta}(t-s)}{\sum_{d=0}^{\infty}q_d^{\theta}\mathcal{D}_{\theta_1,\theta_2+d}(z)}\sum_{j=0}^{k}\binom{k}{j}\frac{B(\theta_1+j,\theta_2+m+k-j)}{B(\theta_1,\theta_2+m)}\mathcal{D}_{\theta_1+j,\theta_2+k-j}(z)$$
$$\times\,\mathcal{D}_{\theta_1+j,\theta_2+m+k-j}(y) \tag{PA2}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=0}^{\infty}\frac{q_m^{\theta}(s)q_k^{\theta}(t-s)}{q_0^{\theta}\frac{1}{B(\theta_1,\theta_2)}}\frac{B(\theta_1+k,\theta_2+m)}{B(\theta_1,\theta_2+m)B(\theta_1+k,\theta_2)}\mathcal{D}_{\theta_1+k,\theta_2+m}(y) \tag{PA3}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=0}^{\infty}\frac{q_m^{\theta}(s)q_k^{\theta}(t-s)}{q_0^{\theta}\frac{1}{B(\theta_1,\theta_2)}}\frac{B(\theta_1+m,\theta_2+k)}{B(\theta_1+m,\theta_2)B(\theta_1,\theta_2+k)}\mathcal{D}_{\theta_1+m,\theta_2+k}(y) \tag{PB1}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=0}^{\infty}\frac{q_m^{\theta}(s)q_k^{\theta}(t-s)}{\sum_{d=0}^{\infty}q_d^{\theta}\mathcal{D}_{\theta_1+d,\theta_2}(z)}\sum_{j=0}^{k}\binom{k}{j}\frac{B(\theta_1+m+j,\theta_2+k-j)}{B(\theta_1+m,\theta_2)}\mathcal{D}_{\theta_1+j,\theta_2+k-j}(z)$$
$$\times\,\mathcal{D}_{\theta_1+m+j,\theta_2+k-j}(y) \tag{PB2}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=0}^{\infty}\frac{q_m^{\theta}(s)q_k^{\theta}(t-s)}{\sum_{d=0}^{\infty}q_d^{\theta}\frac{1}{B(\theta_1+d,\theta_2)}}\frac{B(\theta_1+m+k,\theta_2)}{B(\theta_1+m,\theta_2)B(\theta_1+k,\theta_2)}\mathcal{D}_{\theta_1+m+k,\theta_2}(y) \tag{PB3}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=0}^{\infty}\frac{q_m^{\theta}(s)q_k^{\theta}(t-s)}{\sum_{d=0}^{\infty}q_d^{\theta}(t)\frac{(1-x)^d}{B(\theta_1,\theta_2+d)}}\sum_{l=0}^{m}\mathcal{B}_{m,x}(l)\frac{B(\theta_1+l,\theta_2+m-l+k)}{B(\theta_1,\theta_2+k)B(\theta_1+l,\theta_2+m-l)}$$
$$\times\,\mathcal{D}_{\theta_1+l,\theta_2+m-l+k}(y) \tag{PC1}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=0}^{\infty}\frac{q_m^{\theta}(s)q_k^{\theta}(t-s)}{\tilde{p}(x,z;t)}\sum_{l,j=0}^{m,k}\mathcal{B}_{m,x}(l)\binom{k}{j}\frac{B(\theta_1+l+j,\theta_2+m-l+k-j)}{B(\theta_1+l,\theta_2+m-l)}$$
$$\times\,\mathcal{D}_{\theta_1+j,\theta_2+k-j}(z)\mathcal{D}_{\theta_1+l+j,\theta_2+m-l+k-j}(y) \tag{PC2}$$

$$p^{x,z;t}(y;s)=\sum_{m,k=0}^{\infty}\frac{q_m^{\theta}(s)q_k^{\theta}(t-s)}{\sum_{d=0}^{\infty}q_d^{\theta}(t)\frac{x^d}{B(\theta_1+d,\theta_2)}}\sum_{l=0}^{m}\mathcal{B}_{m,x}(l)\frac{B(\theta_1+l+k,\theta_2+m-l)}{B(\theta_1+k,\theta_2)B(\theta_1+l,\theta_2+m-l)}$$
$$\times\,\mathcal{D}_{\theta_1+l+k,\theta_2+m-l}(y) \tag{PC3}$$

3 抽样方案

现在详细说明如何从上述转移密度中抽取样本，分别针对扩散过程和扩散桥的情形。

3.1 从扩散过程的概率律中抽样

注意，我们只需考虑 $\boldsymbol{\theta} = (0, \theta)$ 和 $\boldsymbol{\theta} = 0$ 这两种情形，因为 $\boldsymbol{\theta} = (\theta_1, \theta_2)$ 的情形已由 Jenkins 和 Span`o (2017年) 讨论。此外， $x \in [0, 1)$ 的转移密度 (3)、(8)、(9) 和 (10) 相似，因此可以采用几乎相同的抽样方案。为此，我们将注意力限制在 $\boldsymbol{\theta} = (0, \theta)$ 的情形，并从针对 (3) 的抽样方案开始。

在这种情形下，可以轻松调整 Jenkins 和 Span`o (2017年) 中的算法1，从 (3) 中抽样：

1. 抽取 $M \sim \{q_m^{\theta}(t)\}_{m \in \mathbb{N}}$,
2. 在给定 $M = m$ 的条件下，采样 $L \sim \text{Bin}(m, x)$,
3. 如果 $L = 0$，则返回 0；否则，抽取 $Y \sim \text{Beta}(l, \theta + m - l)$。

对 Jenkins 和 Span`o（2017年）中的算法1，唯一的修改是第 3 步的采样过程，其中结果 $L = 0$ 表示扩散过程在采样时间之前被吸收的事件。类似的策略也可用于从式（9）中抽样；此外，如果 $L = m$，则在第 3 步返回 $Y = 1$。

对于以非吸收为条件的扩散过程，式（8）右端的两个表达式都是贝塔分布的混合分布，其权重在 $\mathbb{N}$ 上构成一个概率质量函数（pmf）。当起始点 x 设为 0 时，可以通过以下方式从相应扩散过程在时间 t 的分布中抽取样本 Y：

1. 抽取 $M \sim \left\{ \frac{m q_m^0(t)}{\sum_{d=1}^{\infty} d q_d^0(t)} \right\}_{m \in \mathbb{N}}$
2. 在给定 $M = m$ 的条件下，抽取 $Y \sim$ 贝塔分布$(1, m - 1)$

第 2 步很直接，而对于第 1 步，可以采用“交错级数技巧”——该技术要求能够获得分子和分母各项的单调上界和下界序列对，并且这两组序列都收敛于其精确值。对于分子而言，这一点是显然的（Jenkins 和 Span`o（2017年）中的命题1）；而对于分母，我们对 Jenkins 和 Span`o（2017年）中命题3所给出的论证稍作修改（见第

5 获取更多详情）。

可以采用类似的抽样方案，从以 $x \in (0,1]$，开始的扩散过程的分布中抽取样本；同样，可以为分子和分母构造适当的单调上界和下界（更多详情见第 5 节）。

上述方法可以复现并适当调整，从而从（10）中抽取样本；此外，还需要一种方案来处理 $x = 1$ 这种情况。

3.2 从扩散桥的分布中抽样

我们再次从考虑这样一种情况开始：允许桥在时间区间 $(0, t)$ 内被边界吸收，且突变参数为 $\boldsymbol{\theta} = (0, \theta)$。

要从（12）中抽样，我们采用与上文扩散过程情形所示方法类似的方法。请注意，我们只需关注 $z = 0$ 的情况；否则，桥在时间区间 $(0, t)$ 内不可能被吸收，因此等同于以非吸收为条件（稍后将给出相应的抽样方案）。注意，（12）的右端可以看作贝塔分布的混合，其混合权重为

$$p_{m,k,l} := \begin{cases} \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^\infty q_d^\theta(t)(1-x)^d} \mathcal{B}_{m,x}(l) \frac{B(l,\theta+m-l+k)}{B(l,\theta+m-l)} \mathcal{D}_{l,\theta+m-l+k}(y) & m,k \in \mathbb{N}, l \in \{1,\dots,m\} \\ \frac{q_m^\theta(s) q_k^\theta(t-s)}{\sum_{d=1}^\infty q_d^\theta(t)(1-x)^d} (1-x)^m & m,k \in \mathbb{N}, l = 0 \\ 0 & \text{otherwise} \end{cases}$$

在 $\mathbb{N}^3$ 的子空间上定义一个概率质量函数（更多详情参见 Jenkins 和 Span`o，2017年，第 3 节）。可以分别为 $\{q_m^\theta(s)\}_{m\in\mathbb{N}}$、$\{q_k^\theta(t-s)\}_{k\in\mathbb{N}}$ 和 $\sum_{d=1}^\infty q_d^\theta(t)(1-x)^d$ 构造单调上界和下界（关于最后一个量的完整详情，参见第 5 节），随后将它们组合起来，得到关于 $\{p_{m,k,l}\}_{m,\ k,\ l\in\mathbb{N}}$ 的单调上界和下界。因此，交错级数技巧可用于抽取一个样本 $(M, K, L) \sim \{p_{m,k,l}\}_{m,\ k,\ l,\ \in\mathbb{N}}$，我们据此抽取相应的扩散桥点：

1. 采样（M、K、L）$\sim \{p_{m,k,l}\}_{m,k,l,\in\mathbb{N}}$
2. 如果 $L = 0$，则返回 $Y = 0$，否则返回 $Y \sim$ 贝塔分布$(l, \theta + m - l)$。

通过对称论证，可以为 $\boldsymbol{\theta} = (\theta, 0)$ 推导出类似的方案；而对于$\boldsymbol{\theta} = \mathbf{0}$ ，则可重复上述过程，唯一显著的区别是：如果 $L = m$，那么该例程会在步骤 2 中返回 $Y = 1$ 。

现在我们转而讨论扩散桥在时间区间 $(0, t)$ 内未被吸收这一条件下的情形。Griffiths 等（2018）中的推论 2 指出，突变参数为 $\boldsymbol{\theta} = \mathbf{0}$ 或 $\boldsymbol{\theta} = (0, \theta)$ 的 Wright–Fisher 扩散桥，与突变参数分别为 $\boldsymbol{\theta} = (2, 2)$ 或 $\boldsymbol{\theta} = (2, \theta)$ 的 Wright–Fisher 桥在分布上相同。因此，从现在起，我们将只关注 $\theta_1, \theta_2 > 0$ 的情形。

我们采用的策略与上文提出的策略非常接近，并基于 (Jenkins and Span`o, 2017, 第 3 节) 中的方法。与无条件桥的情形一样，扩散桥密度 (PA1)–(PC3) 可视为贝塔分布的混合，其中混合权重现在定义了 $\mathbb{N}^4$ 子空间上的一个概率质量函数，其具体形式取决于所考虑的密度。

由于扩散桥在时间反演下保持不变，从 x 出发，在时间 s 到达 y ，随后于时间 t 终止于 z 的扩散桥，与从 z出发，在时间 $t - s$ 到达 y ，并最终到达 x （时间为 t）的扩散桥具有相同的分布。结合对称性论证，我们只需采用四种不同的方案，便可从各种转移密度 (PA1)–(PC3) 中抽样，具体归纳如下：

1. 起点与终点相同（即方程 (PA1) 和 (PB3)）。

2. 起点与终点是相对的边界点（即方程 (PA3) 和 (PB1)）。

3. z 位于 $[0, 1]$, 的内部，且起点位于其中一个边界点（即方程 (PA2)、(PB2)、(PC1) 和 (PC3)——注意，对于 (PC1) 和 (PC3)，我们利用了时间反演）。

4. 起点和终点均位于 $[0, 1]$ 的内部（即方程 (PC2)）。

利用上述分组，接下来只需证明所得的四种不同转移密度由混合权重 $\{p_{m,k,l,j}\}_{m、\ k、\ l、\ j \in \mathbb{N}}$ 构成，且可以为其求得单调的上界和下界序列。同样，分子的这些量很容易构造，而分母则在第 5 节中处理（通过适当修改 Jenkins 和 Span`o（2017年）中的命题 4）。

我们指出，对于扩散过程和扩散桥这两种情况，在计算小时间增量下概率$\{q_m^\theta(t)\}_{m\in\mathbb{N}}$无穷级数表示中的各项时都会出现数值不稳定性，因此需要对这些量进行近似处理。更多详情请参阅第6节。

4 非中性路径的模拟

正如（Jenkins 和 Span`o，2017年，第 5 节）所指出的，中性 Wright–Fisher 过程可作为合适拒绝采样器中的提议分布，从而精确抽取非中性过程的样本。为完整起见，我们在此简要概述。

记 $\mathbb{WF}_{\sigma,\boldsymbol{\theta}}^{x_0}$ 为由主论文中方程（1）所给随机微分方程的解 $X^T:=(X_t)_{t=0}^T$ 诱导的概率测度，该测度定义在从 $[0,T]$ 到 $[0,1]$ 的连续函数空间上，其中时间 T固定；记 $\mathbb{WF}_{0,\boldsymbol{\theta}}^{x_0}$ 为相应的中性概率测度。这两个概率测度之间的 Radon–Nikodym导数为

$$\frac{d\mathbb{WF}_{\sigma,\boldsymbol{\theta}}^{x_0}}{d\mathbb{WF}_{0,\boldsymbol{\theta}}^{x_0}}(X^T)\propto\exp\left\{\tilde{A}(X_T)-\tilde{A}^+\right\}\exp\left\{-\int_0^T\left(\varphi(X_s)-\varphi^-\right)dt\right\}\tag{18}$$

其中 $\tilde{A}(x):=(\sigma/2)\int_0^x\eta(z)dz$，且对于任意 $x\in[0,1]$，$\tilde{A}(x)\le\tilde{A}^+$，并且

$$\varphi(x):=\frac{\sigma}{4}\left[(-\theta_2x+\theta_1(1-x))\,\eta(x)+x(1-x)\left(\frac{\sigma}{2}\eta^2(x)+\eta'(x)\right)\right].\tag{19}$$

注意到 $\varphi(x)$ 是关于 x 的多项式（因为 $\eta(x)$ 是多项式），因此我们总能找到 φ^- 和 φ^+，使得在 $[0,1]$，上 $\varphi^-\le\varphi(x)\le\varphi^+$；同理，$\tilde{A}(x)\le\tilde{A}^+$。（18）式右端的第一项可视为一次简单的 $e^{\tilde{A}_{(X_T)}-\tilde{A}^+}$ 抛硬币，而第二项恰好是单位速率泊松点过程 $\Phi=\{(t_i,\omega_i)\}_{i\in\mathbb{N}}$ 在 $[0,T]\times[0,\infty)$ 上的所有点均落在映射 $t\mapsto\varphi(x)-\varphi^-$ 的上图集中的概率。此外，由于 $\varphi(x)\le\varphi^+$，我们可以对 Φ 进行稀疏化，在 $[0,T]\times[0,\varphi^+-\varphi^-]$ 上得到一个泊松点过程，从而模拟一个发生概率等于（18）式右端的事件。

这样，只需先模拟相应的泊松点过程，再根据该泊松点过程返回的时间戳，从中性 Wright–Fisher 过程中生成样本，检查生成的点是否全部位于相应区域内，最后进行一次简单的$e^{\tilde{A}(X_T)-\tilde{A}^+}$抛硬币，即可返回非中性 Wright–Fisher 过程的精确路径。

为了计算 $\tilde{A}^{+}$、φ^{-} 和 φ^{+}，使用了一个多项式类（该类配有由 Bill Hallahan[1]开发的、实现 Jenkins–Traub算法的求根算法）。虽然该例程的实现应适用于任意次数的多项式，但为确保代码能在合理时间内返回可靠的输出，只允许次数不超过 25 的多项式$\eta(x)$。

5 新分母的单调上界和下界

本节将说明，无论是扩散过程（方程（8）和（10）），还是扩散桥（方程（12）、（16）和（17），以及方程（PA1）至（PC3））的转移密度，其分母都允许构造单调序列的上界和下界。比较（8）和（10），以及（12）、（16）和（17）后可以看出，我们只需考虑分母 $\sum_{d=2}^{\infty} q_d^0(t)(1-x^d-(1-x)^d)$，因为其他量的证明采用几乎相同的论证即可得出。

我们再次强调（如第 3 节所述），对于扩散桥，只需考虑（PA1）、（PA2）、（PA3）和（PC2），便能模拟任意 Wright–Fisher扩散桥过程的样本。此外，注意到（PA3）的分母为 $q_0^{\theta}(t)$，其单调上界和下界可直接得到；而（PC2）恰好是 Jenkins 和 Span`o（2017年）中命题3所涵盖的情形。因此，接下来需要为以下各项寻找单调收敛的上界和下界序列：

$$\sum_{d=0}^{\infty} q_d^{\theta}(t)d, \tag{20}$$

$$\sum_{d=0}^{\infty} q_d^{\theta}(t)(1-x^d-(1-x)^d), \tag{21}$$

$$\sum_{d=0}^{\infty} q_d^{\theta}(t)\frac{1}{B(\theta_1+d,\theta_2)}, \tag{22}$$

$$\sum_{d=0}^{\infty} q_d^{0}(t)\frac{z^{\theta_1+d-1}(1-z)^{\theta_2-1}}{B(\theta_1+d,\theta_2)}. \tag{23}$$

此外，根据 Griffiths 等（2018）中的方程（5），式（20）可通过祖先过程下降阶乘矩的解析表达式得到所需的单调界（详见 Griffiths 等（2018）中的定理5及前面的段落）。

[1]https://www.codeproject.com/Articles/674149/A-Real-Polynomial-Class-with-Root-Finder

5.1 对(21)的计算

处理式（21）需要更多工作；我们首先注意到（$1-x^m-(1-x)^m$）$=\sum_{l=1}^{m-1}\binom{m}{l}x^l(1-x)^{m-l}$。我们可以修改 Jenkins 和 Span`o（2017）中引理1的论证，从而推导出对于 $L_m\sim \text{Bin}(m,x)$，有

$$\sum_{l=1}^{m}\mathbb{P}\left[L_{m+1}=l\right]\leq(x+2)\sum_{l=1}^{m-1}\mathbb{P}\left[L_m=l\right]. \tag{24}$$

为此，注意到对于 $l\leq\lfloor mx\rfloor$

$$\mathbb{P}\left[L_{m+1}=l\right]=\frac{m+1}{m+1-l}(1-x)\mathbb{P}\left[L_m=l\right]\leq\mathbb{P}\left[L_m=l\right], \tag{25}$$

其中，在最后一个不等式中，我们用到了这样一个事实：$l\leq\lfloor mx\rfloor\leq mx$。当 $l\geq\lfloor mx\rfloor$时，我们有

$$\mathbb{P}\left[L_{m+1}=l+1\right]=\frac{m+1}{l+1}x\mathbb{P}\left[L_m=l\right]\leq(x+1)\mathbb{P}\left[L_m=l\right] \tag{26}$$

注意到当 $mx>1$、$\frac{m+1}{l+1}\leq\frac{m+1}{mx}\leq1+\frac{1}{x}$ 时，而对于 $mx\leq1$、$\frac{m+1}{l+1}\leq m+1\leq1+\frac{1}{x}$ 时则有。将(25)和(26)相加（并注意到在 $\lfloor mx\rfloor$处发生了重复计数），即可得到结果。有了这一结果，我们可以应用 Jenkins 和 Span`o（2017年）中的命题3，这次令 $c_{k,\ m}:=b_k^{(t,\theta)}(m)$ $\sum_{l=1}^{m-1}\mathbb{P}[L_m=l]$，并将 $K^{(x,z)}$ 替换为 $x+2$。

5.2 对(23)的计算

首先注意到

$$\begin{aligned}&\sum_{m=1}^{\infty}q_m^0(t)\frac{z^{\theta_1+m-1}(1-z)^{\theta_2-1}}{B(\theta_1+m,\theta_2)}\\&=\sum_{m=1}^{\infty}\left(\sum_{k=m}^{\infty}(-1)^{k-m}\frac{\theta+2k-1}{m!(k-m)!}\frac{(\theta+k+m-2)!}{(\theta+m-1)!}e^{-\frac{k(\theta+k-1)t}{2}}\right)\frac{z^{\theta_1+m-1}(1-z)^{\theta_2-1}}{B(\theta_1+m,\theta_2)},\end{aligned}$$

并注意到，内层求和中的各项（不包括因子 $(-1)^{k-m}$）对应于 Jenkins 和 Span`o (2017年) 中命题1所定义的项 $b_k^{(t,\theta)}(m)$。令$c_{k,\ m}:=b_k^{(t,\theta)}(m)z^{\theta_1+m-1}(1-z)^{\theta_2-1}/B(\theta_1+m,\theta_2)$，并注意到我们可以将上式改写为 $\sum_{i=0}^{\infty}(-1)^i d_i$

with

$$d_{2m} = \sum_{j=0}^{m} c_{m+j,m-j}, \qquad d_{2m+1} = \sum_{j=0}^{m} c_{m+1+j,m-j}. \tag{27}$$

固定 $\varepsilon \in (0,1)$，记为

$$E^t := \inf\left\{m \geq 0 : 2j \geq C^t_{m-j} \text{ for all } j = 0, \ldots, m\right\}, \tag{28}$$

$$D^{t,\theta}_{\varepsilon} := \inf\left\{k \geq \left(\frac{1}{t} - \frac{\theta+1}{2}\right) \vee 0 : (\theta+2k+1)e^{-\frac{(\theta+2k)t}{2}} < 1-\varepsilon\right\}. \tag{29}$$

Jenkins 和 Span`o（2017年）中的命题3，可针对我们在此考虑的情形改述如下。

命题 5.1。对于所有 $m > D^t, {}^{\varepsilon\theta} \vee E^t \vee \lfloor {}_{\varepsilon}{}^{\theta+2}_{\theta_1+1()} - 1 \rfloor$,

$$d_{2m+2} < d_{2m+1} < d_{2m}. \tag{30}$$

证明。证明过程与 Jenkins 和 Span`o（2017年）中的相同。由于 $m > E^t$, $2j \geq C^t_{m-j}$, 因此根据 Jenkins 和 Span`o（2017年）中的命题1, $b_{m+j+1}(m-j) < b_{m+j}(m-j)$。将不等式两边乘以 $z^{\theta_1+m-1}(\quad {}^{1-z\ \theta_2-1}_{B\ \theta_1+m\ \theta_2})(,\ )$ 并对 j 求和，得到

$$d_{2m+1} = \sum_{j=0}^{m} c_{m+j+1,m-j} < \sum_{j=0}^{m} c_{m+j,m-j} = d_{2m}.$$

上述推理还可得出

$$\sum_{j=1}^{m} c_{m+j+2,m-j} < \sum_{j=1}^{m} c_{m+j+1,m-j},$$

结合 $c_{m+1,\ m+1} + c_{m+2,\ m} < c_{m+1,\ m}$ （仍有待证明），即可得到所需的 $d_{2m+2} < d_{2m+1}$。现在注意到

$$\frac{c_{k+1,m}}{c_{k,m}} = \frac{b^{(t,\theta)}_{k+1}(m)}{b^{(t,\theta)}_{k}(m)} = \frac{\theta+m+k-1}{k-m+1}\frac{\theta+2k+1}{\theta+2k-1}e^{-\frac{(\theta+2k)t}{2}} \leq (\theta+2k+1)e^{-\frac{(\theta+2k)t}{2}},$$

设置 $k = m+1$ 并注意到 $m > D^t_{\varepsilon}$, 可得 $c_{m+2,m} < (1-\varepsilon)c_{m+1,m}$。类似地

$$\frac{c_{m+1,m+1}}{c_{m+1,m}} = \frac{\theta+2m}{(m+1)(\theta+m)}z\frac{B(\theta_1+m,\theta_2)}{B(\theta_1+m+1,\theta_2)} \leq \frac{\theta+2}{(m+1)(\theta_1+1)} < \varepsilon$$

如果 $m > \lfloor {}_{\varepsilon}{}^{\theta+2}_{\theta_1+1()} - 1 \rfloor$。结果由此得出。 □

5.3 式（22）的计算

上述相同的论证同样适用（省略 $z^{\theta_1+d-1}(1-z)^{\theta_2-1}$ 的存在，这会使证明略微简化）。

6 小时间近似与实现

每当模拟的时间增量变得过小时，计算量 $q_m^\theta(t)$ 的贡献时就会出现数值不稳定性。因此（如 Jenkins 和 Span`o（2017年）所做的那样），有必要采用适当的近似方法，利用 $q_m^\theta(t)$ 的小时间渐近结果。Griffiths（1984）的定理 4 表明，当 $t \to 0$ 时，合并过程的祖先块计数过程可以很好地近似为均值为

$$\mu = \frac{2\eta}{t}, \qquad \text{where } \eta = \begin{cases} 1 & \beta = 0 \\ \frac{\beta}{e^\beta - 1} & \beta \neq 0 \end{cases}, \qquad \text{and } \beta = \frac{(\theta-1)t}{2},$$

且方差为

$$\sigma^2 = \begin{cases} \frac{2}{3t} & \beta = 0 \\ \frac{2\eta}{t}\left(\frac{\eta+\beta}{\beta}\right)^2 \left(1 + \frac{\eta}{\eta+\beta} - 2\eta\right) & \beta \neq 0 \end{cases}$$

（请注意，Griffiths（1984）的定理 4 中漏掉了一个因子 β^{-2}）。因此，每当时间增量 t 低于特定阈值 ε_G 时，EWF 就会采用上述高斯近似，将概率 $q_m^\theta(t)$ 替换为相应的（经过适当舍入的）高斯近似值。在 EWF 的当前实现中，经过大量测试后将阈值 ε_G 设为 0.08，因为研究发现，这一截断值能够在保持精度（仅在必要时采用近似）与实现的稳健性和效率之间取得适当的平衡。

对于扩散桥的情形，我们对 $q_m^\theta(s)$ 和 $q_k^\theta(t-s)$ 都采用类似的近似，但还引入了一个额外阈值 $\varepsilon_D < \varepsilon_G$；低于该阈值时，我们通过从扩散的概率分布中抽样，来近似从扩散桥的概率分布中抽样。这是因为上文给出的高斯近似均值 μ 与时间增量 t 成反比。因此，只要时间增量 s 或 $t-s$ 中任一个较小，概率质量函数$\{p_{m,k,l,j}\}_{m,\ k,\ l,\ j\in\mathbb{N}}$ 就会在 $\mathbb{N}^4$ 上变得极为分散，导致数值过小而损失精度，同时运行时间也会变得难以接受；即便采用上文展示的高斯

使用了近似方法。

在此类情况下（即 $s < \varepsilon_D$ 或 $t - s < \varepsilon_D$），EWF 会先模拟一个从 x 出发、在时间 s 处取样的相应 Wright–Fisher 扩散样本，计算生成的样本 Y' 与起点 x 之间的增量，并将其叠加到左右端点 x 和 z 之间的线性插值上，以生成所需的样本 Y。所采用的线性插值会明确使用时间增量 s 和 $t - s$，以确保返回的样本 Y 来自一个从 x 出发并在 z 处结束的扩散桥，同时通过适当机制确保输出始终位于区间 $[0, 1]$ 内。当 $s \in [\varepsilon_D, \varepsilon_G)$ 或 $t - s \in [\varepsilon_D, \varepsilon_G)$ 时，相应时间区间内的 $\{q_i^\theta\}_{i \in \mathbb{N}}$ 会采用上述详细说明的（取整）高斯近似；对于超过 ε_G 的时间增量，则采用标准抽样方案。经过大量测试后，我们为 ε_D 选定了 0.008 这一阈值，使 EWF 的实现既稳健又高效，并且除非不使用此类近似会导致运行时间难以接受，否则不会采用这些近似。需要指出的是，如有需要，可以通过 myHelpers.h 头文件中 Options 类的 g1984 字段（用于 ε_G）和 bridgethreshold 字段（用于 ε_D）修改这两个阈值（不过我们不建议这样做）。

7 输出验证

我们通过针对各种不同情形生成 10,000 个样本，并随后使用柯尔莫哥洛夫–斯米尔诺夫检验和 QQ 图，将其与截断后的转移密度进行比较，从而验证输出。需要指出的是，这里仅展示中性输出，因为非中性输出采用了与 Jenkins 和 Span`o (2017年) 中相同的拒绝抽样过程生成。

为说明转移密度是如何截断的，考虑情形 (PC2)：它涉及对四个指标求和，其中两个指标的取值范围是无限的。通过使用迭代方案，找到了这四个指标对应的众数，并计算了其对给定点 $y \in [0, 1]$ 处密度的贡献。随后，将 (PC2) 的分母计算至机器精度，并将所得分母、众数对密度的贡献以及一个容差参数相乘，据此选取适当的截断级别。其他所有扩散过程和扩散桥情形也采用了类似的截断方法。

7.1 非吸收条件下的扩散过程

使用 9 组不同的参数设置生成样本，涉及起始点 $x \in \{0, 0.5, 1\}$、采样时间 $t \in \{0.01, 0.05, 0.5\}$ 和突变参数 $\boldsymbol{\theta} = (0, 1)$。下方绘出了输出结果，首先展示 $x = 0$ 时的情形；沿着图从左向右，采样时间增量 t 逐渐增加。下方所有的柯尔莫哥洛夫–斯米尔诺夫检验和 QQ 图均证实，输出确实来自正确的分布。

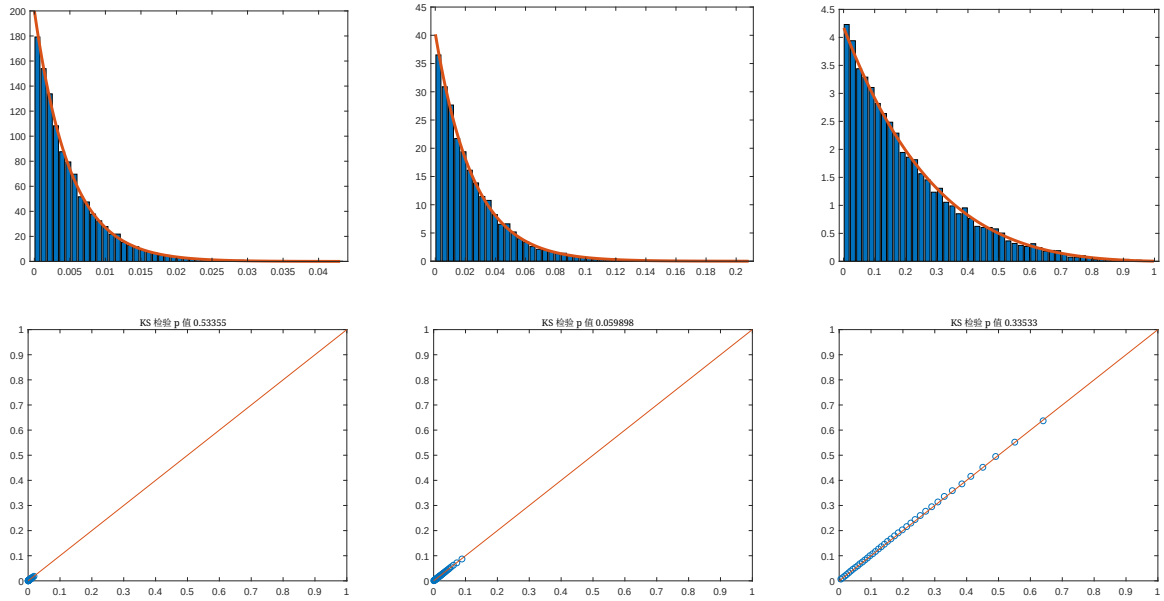


图 2: （上排）：由条件于非吸收的 Wright–Fisher 扩散过程的分布生成的 10,000 个样本的直方图，该过程在时间 0 时从 $x = 0$ 出发，分别在时间 $t = 0.01$、0.05、0.5 采样。截断转移密度以红色绘出。（下排）：对应样本的 QQ 图，图上方标出了柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

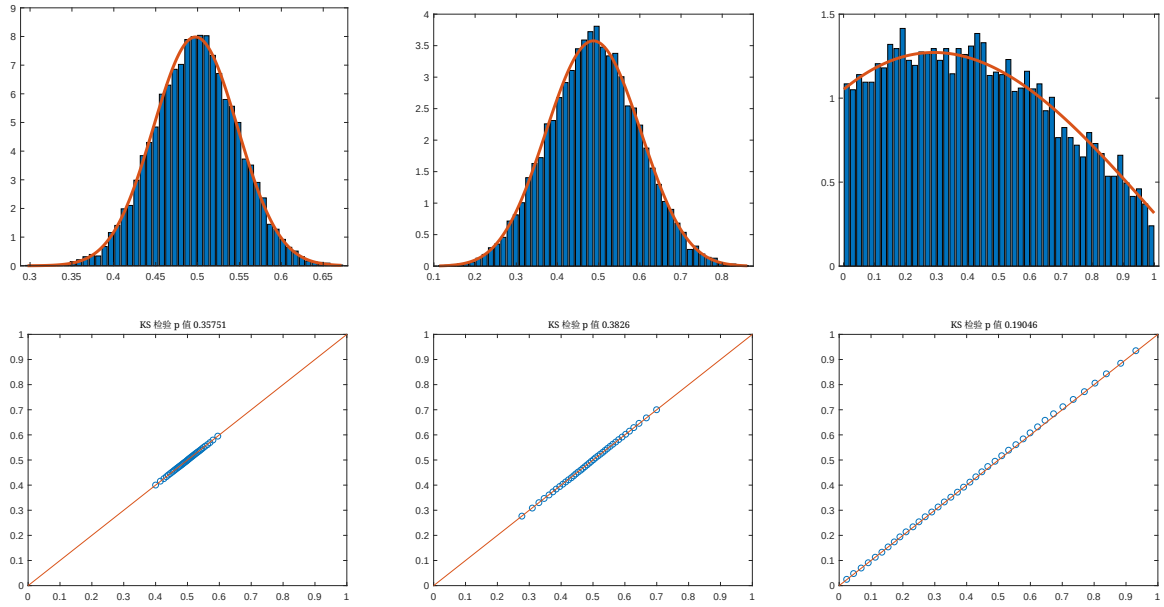


图3：（上排）：从以 $x = 0.5$ 为初始值、在时刻 0 开始的、条件为非吸收的 Wright–Fisher 扩散过程的概率分布中生成的 10,000 个样本的直方图，分别在时刻 $t = 0.01$、0.05、0.5 进行采样。截断转移密度以红色绘出。（下排）：对应样本的 QQ 图，图上方标出了柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

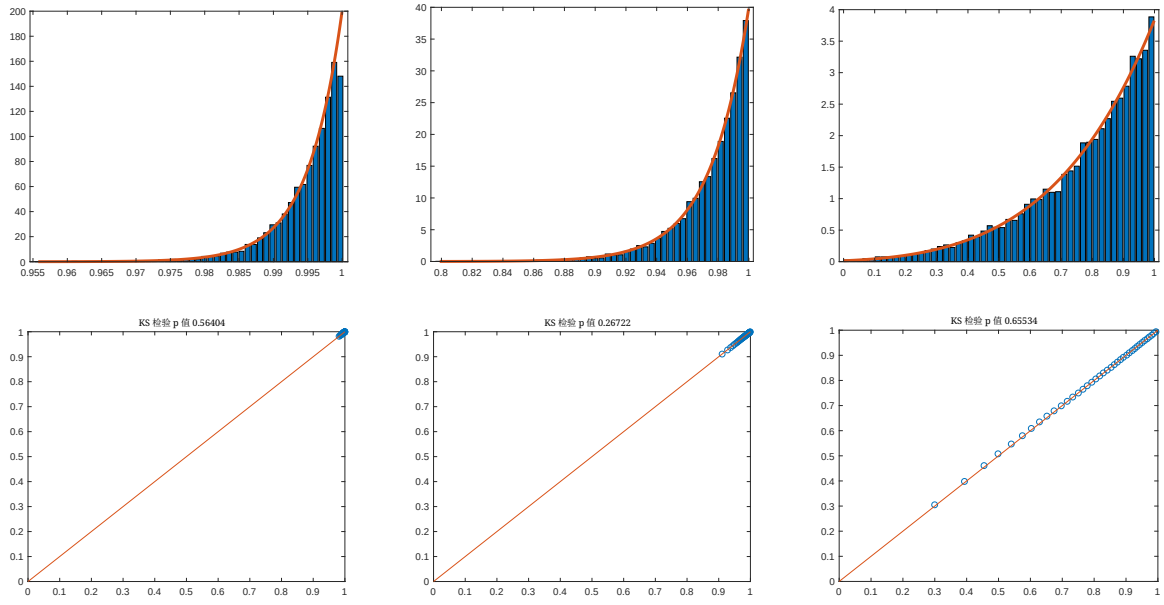


图4：（上排）：从给定非吸收条件的 Wright– Fisher 扩散过程的分布生成的 10,000 个样本的直方图，该过程在时间 0 从 $x = 1$ 出发，分别在时间$t = 0.01$、0.05、0.5 采样。截断转移密度以红色绘出。（下排）：对应样本的 QQ 图，图上方标出了柯尔莫尔戈洛夫–斯米尔诺夫检验返回的 p-值。

7.2 无条件扩散

在允许扩散过程在边界处被吸收的情况下，我们针对以下情形进行了模拟：起点 $x \in \{0.25, 0.5, 0.75\}$、采样时间 $t \in \{0.05, 0.25, 0.5\}$ 和突变参数 $\boldsymbol{\theta} = \mathbf{0}$。下表报告了在任一边界处被吸收的概率 $\widehat{\ }$，其中 $\mathbb{P}$ 表示该概率的经验估计，而 $\mathbb{P}$ 是通过计算转移密度在边界处的截断值获得的理论值。所有估计概率均与相应的理论值一致；此外，QQ 图和柯尔莫哥洛夫–斯米尔诺夫检验也都证实，生成的样本来自正确的分布。

$x = 0.25$	$\widehat{\mathbb{P}}$[在 0]处被吸收	$\mathbb{P}$[在 0]处被吸收	$\widehat{\mathbb{P}}$[在 1]处被吸收	$\mathbb{P}$[在 1]处被吸收
$t = 0.05$	0	1.51641e-5	0	8.92526e-26
$t = 0.25$	0.1025	0.101181	1e-4	9.79038e-5
$t = 0.5$	0.2923	0.302098	0.0074	0.0077254

$\widehat{\ }$表1：从$x = 0.25$开始的扩散过程的经验（$\mathbb{P}$）和理论（$\mathbb{P}$）吸收概率。

$x = 0.5$	$\widehat{\mathbb{P}}$[吸收于 0]	$\mathbb{P}$[吸收于 0]	$\widehat{\mathbb{P}}$[在 1]处被吸收	$\mathbb{P}$[在 1]处被吸收
$t = 0.05$	0	9.81343e-9	0	9.81343e-9
$t = 0.25$	0.0066	0.00569842	0.0064	0.00569842
$t = 0.5$	0.0687	0.066694	0.065	0.066694

$\widehat{\ }$表2：从$x = 0.5$开始的扩散过程的经验（$\mathbb{P}$）和理论（$\mathbb{P}$）吸收概率。

$x = 0.75$	$\widehat{\mathbb{P}}$[在 0]处被吸收	$\mathbb{P}$[在 0]处被吸收	$\widehat{\mathbb{P}}$[在 1]处被吸收	$\mathbb{P}$[在 1]处被吸收
$t = 0.05$	0	8.92526e-26	0	1.51641e-5
$t = 0.25$	1e-4	9.79038e-5	0.0986	0.101181
$t = 0.5$	0.0099	0.0077254	0.2979	0.302098

$\widehat{\ }$表 3：从$x = 0.75$开始的扩散过程的经验吸收概率（$\mathbb{P}$）和理论吸收概率（$\mathbb{P}$）。

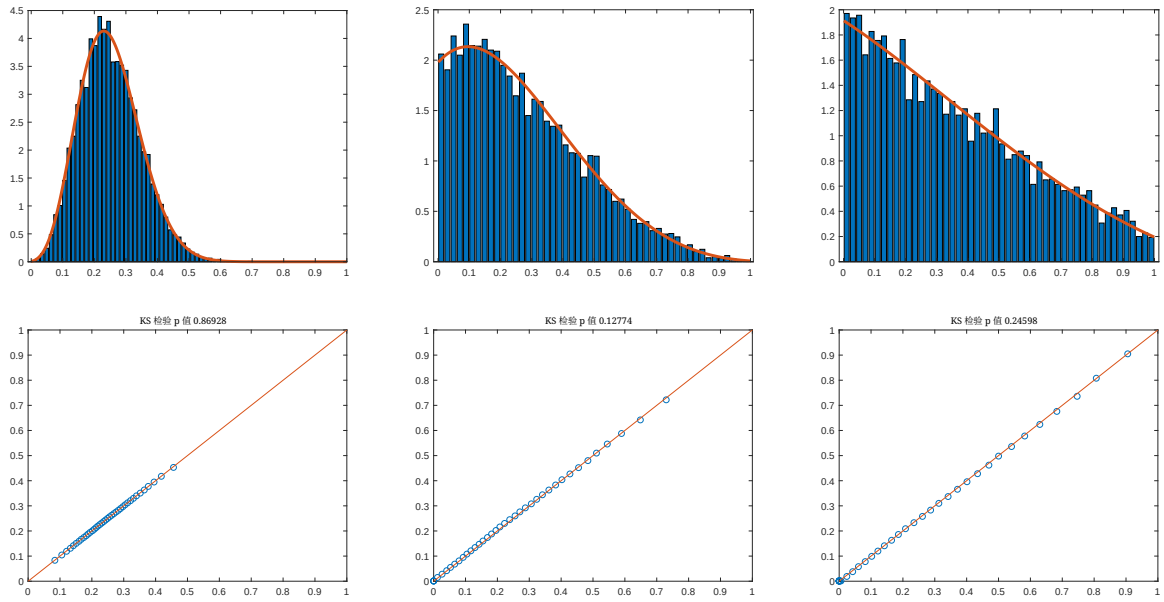


图5：（顶行）：从在时刻 0 以 $x = 0.25$ 为初始值的 Wright–Fisher 扩散过程的分布中生成的 10,000 个样本的直方图，分别在时刻 $t = 0.05$、0.25、0.5 进行采样，并允许该过程在边界处被吸收。请注意，上述直方图不包含等于 0 或 1 的样本，但这些样本的相对频率可通过表1中的经验概率获得。截断转移密度以红色绘制。（底行）：相应样本的 QQ 图，图上方报告了柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

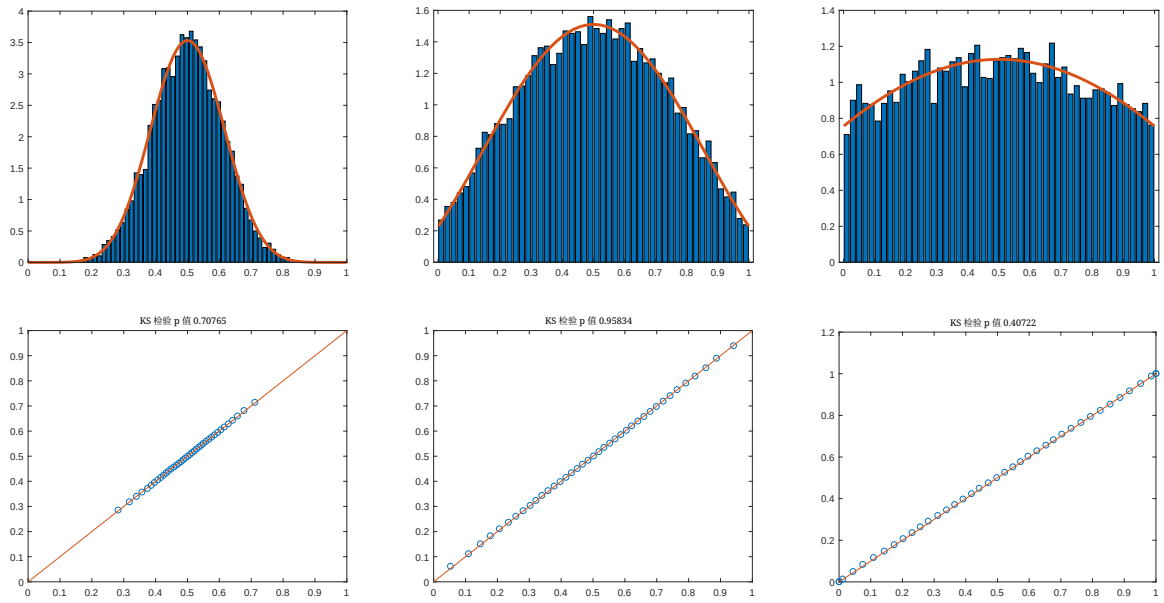


图6：（上排）：从初始时刻 $x = 0.5$ 的 Wright–Fisher 扩散过程的分布中生成的 10,000 个样本在 $t = 0.05$、0.25、0.5 时刻（依次对应）的直方图，过程中允许样本被边界吸收。请注意，上述直方图不包含等于 0 或 1 的样本，但其相对频率可从表2中的经验概率得知。截断转移密度以红色绘出。（下排）：对应样本的 QQ 图，图上方标出了柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

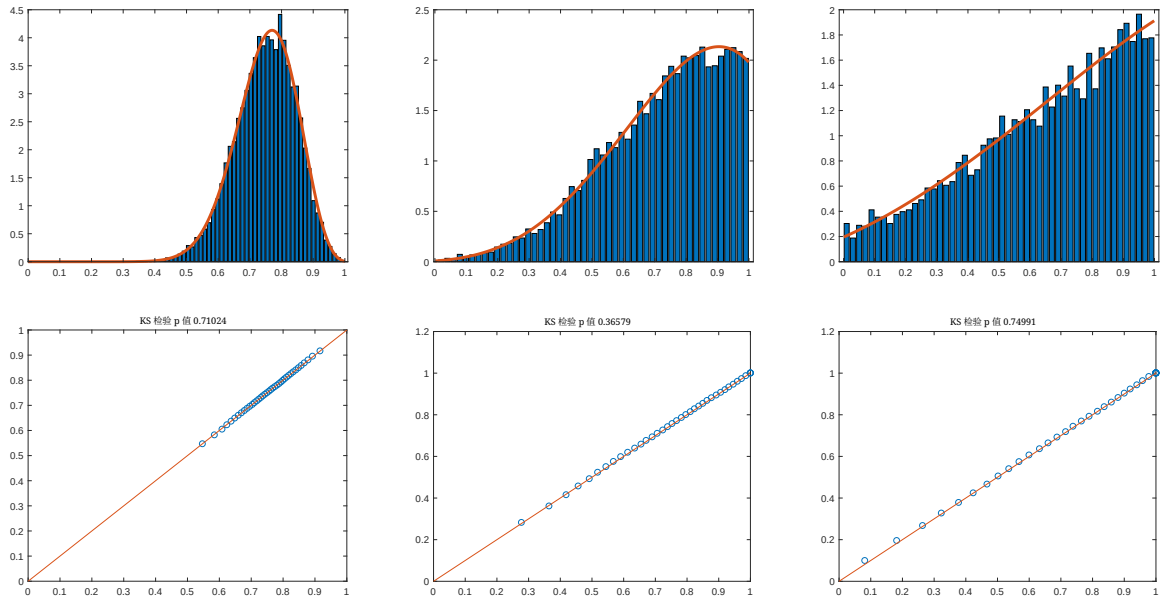


图 7：（上排）：从一个初始时刻 0 取值为 $x = 0.75$ 的 Wright–Fisher 扩散的分布中生成 10,000 个样本，并分别在时刻 $t = 0.05$、0.25 和 0.5 进行采样所得的直方图；该过程允许在边界处被吸收。请注意，以上直方图不包含取值为 0 或 1 的样本，但这些样本的相对频率可通过表 3 中给出的经验概率获知。截断转移密度以红色绘出。（下排）：对应样本的 QQ 图，图上方标出了柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

7.3 在非吸收条件下的扩散桥

为验证扩散桥模拟，我们选择模拟以下三种扩散桥的抽样结果：

	(t_0, x_0)	(t_1, x_1)	(t_2, x_2)	(t_3, x_3)
桥 1	(0,0)	(0.05,0.1)	(0.1,0.25)	
桥 2	(0.2,0.1)	(0.3,0.3)	(0.4,0.4)	(0.5,0.5)
桥 3	(0,1)	(0.5,0.95)		

表 4：所模拟的三种不同桥的左右端点，其中 (t_0, x_0) 表示桥的起始时间 t_0 和起始点 x_0，(t_1, x_1) 表示扩散桥的第二个观测时间和观测点，依此类推。

我们进一步考虑了每个桥的以下采样时间：

	s_1	s_2	s_3
桥 1	0.025	0.065	0.085
桥 2	0.25	0.35	0.45
桥 3	0.1	0.25	0.3

表 5：所考虑的三种不同扩散桥的采样时间。

生成的输出绘制如下，从桥 1 开始，采样时间 s_i 从左到右递增。同样，所有输出都强烈表明该方法返回的是从所需目标分布中抽取的样本。

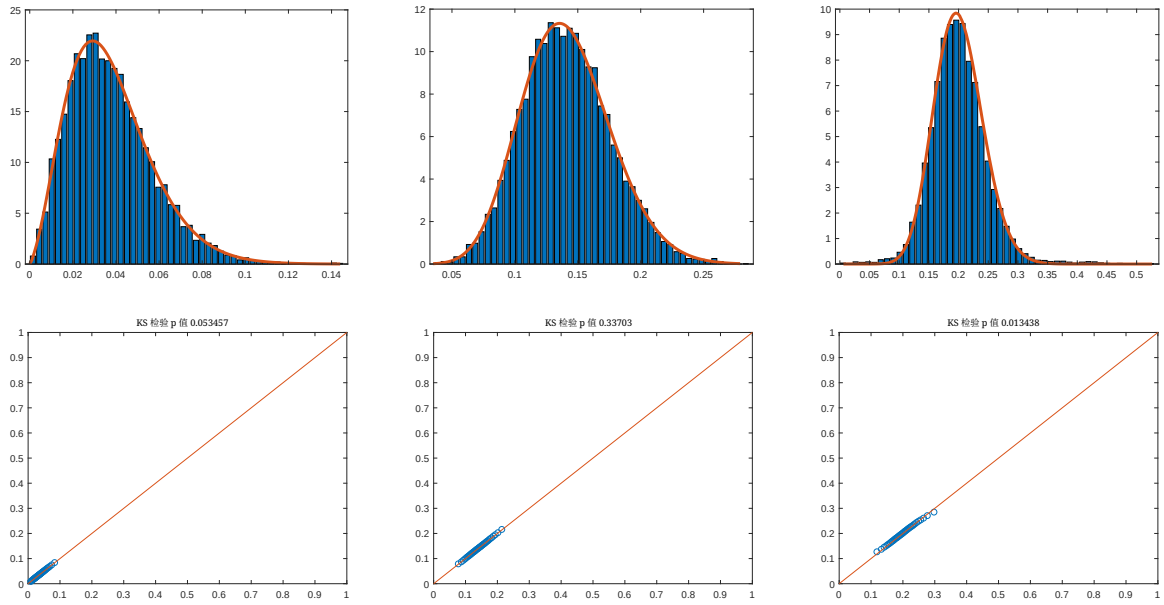


图 8：（上排）：从表 4 中赖特–费希尔扩散桥“桥 1”的分布生成的 10,000 个样本的直方图，采样时间为表 5 对应行中给出的时间。截断转移密度以红色绘出。（下排）：对应样本的 QQ 图，图上方标注了柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

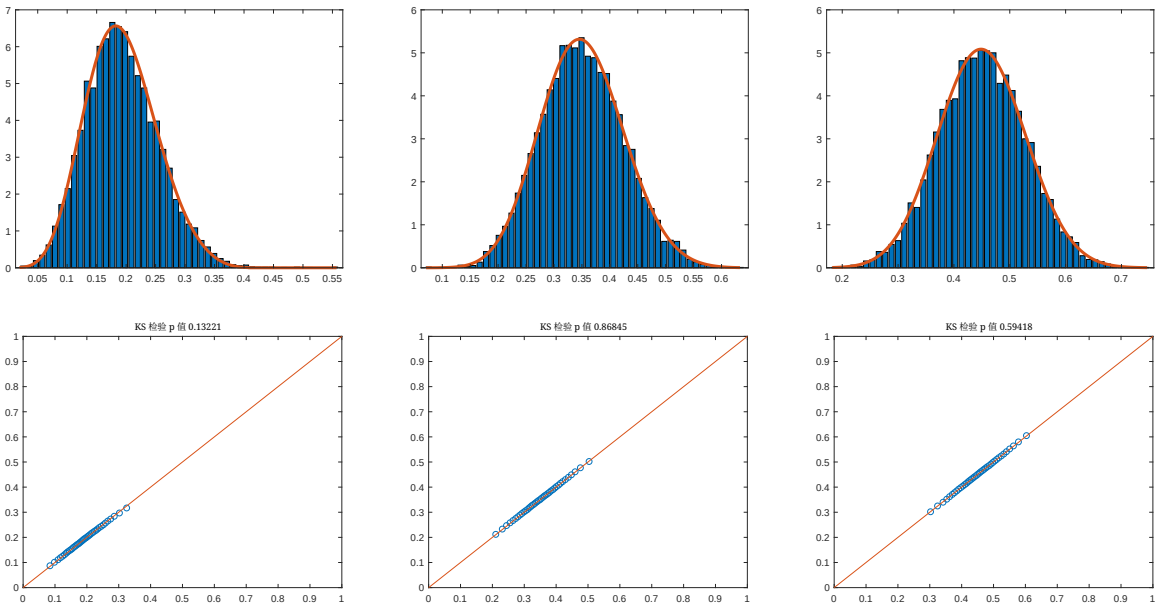


图 9：（上排）：根据上文表 4 所示的“桥 2”生成的 10,000 个样本的直方图，这些样本取自 Wright–Fisher 扩散桥的分布，并在表 5 对应行给出的时刻进行采样。截断转移密度以红色绘出。（下排）：对应样本的 QQ 图，图上方标出了柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

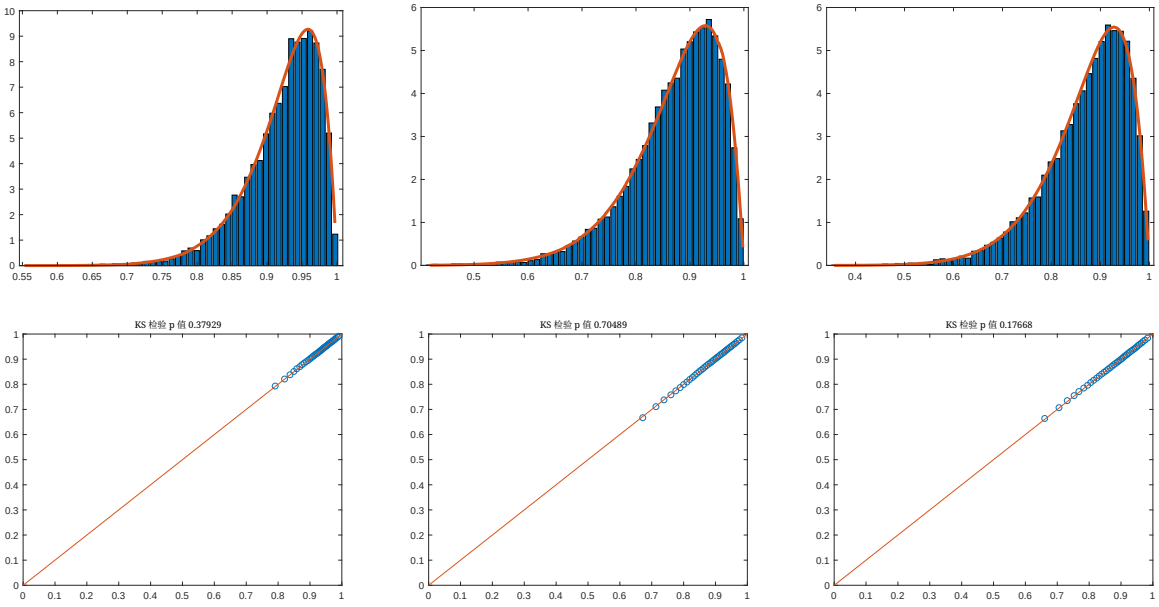


图 10：（顶行）：根据上文表 4 中给出的 Wright–Fisher 扩散过程“桥 3”的扩散桥分布生成的 10,000 个样本的直方图，样本取自表 5 相应行中给出的时间点。截断转移密度以红色绘制。（底行）：相应样本的 QQ 图，图上方标出了上述柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

7.4 无条件桥

当扩散桥允许在边界被吸收且 $\boldsymbol{\theta} = \mathbf{0}$ 时，我们只需考虑 $z \in \{0, 1\}$ 的情况。为此，我们考虑了以下两种设定：

	(t_0, x_0)	(t_1, x_1)
桥 1	(0,0.25)	(0.3,1)
桥 2	(0,0.5)	(0.5,0)

表 6：所模拟的三种不同桥的左右端点，其中 (t_0, x_0) 表示桥的起始时间 t_0 和起始点 x_0，(t_1, x_1) 表示扩散桥的第二个观测时间和点，依此类推。

我们还考虑了以下采样时间：

	s_1	s_2	s_3
桥 1	0.05	0.15	0.25
桥 2	0.05	0.25	0.45

表 7：所考虑的两种不同扩散桥的采样时间。

与扩散过程的情形一样，我们在下表$\widehat{}$中报告边界处的吸收概率，其中，$\mathbb{P}$ 再次表示该量的经验估计，而 $\mathbb{P}$ 则是通过计算边界处转移密度的截断得到的理论值。

桥 1	$\widehat{\mathbb{P}}$[在 1]处吸收	$\mathbb{P}$[在 1]处被吸收
$s = 0.05$	0	3.900485e-16
$s = 0.15$	7e-4	6.752749e-4
$s = 0.25$	0.2331	0.234209

$\widehat{}$表 8：从$x = 0.25$ 开始并在 $z = 1$结束的扩散过程的经验（$\mathbb{P}$）和理论（$\mathbb{P}$）吸收概率。

桥 2	$\widehat{\mathbb{P}}$[在 0]处吸收	$\mathbb{P}$[在 0]处吸收
$s = 0.05$	0	2.418920e-10
$s = 0.25$	0.0881	0.085472
$s = 0.45$	0.7634	0.765359

表 9：从$x = 0.5$ 开始并在 $x = 0$结束的扩散过程的经验（$\widehat{\mathbb{P}}$）和理论（$\mathbb{P}$）吸收概率。

生成的输出如下图所示，从桥 1 开始，采样时间 s 从左向右递增。以上所有图表、检验和概率都证实，我们抽取的样本来自所需的分布。

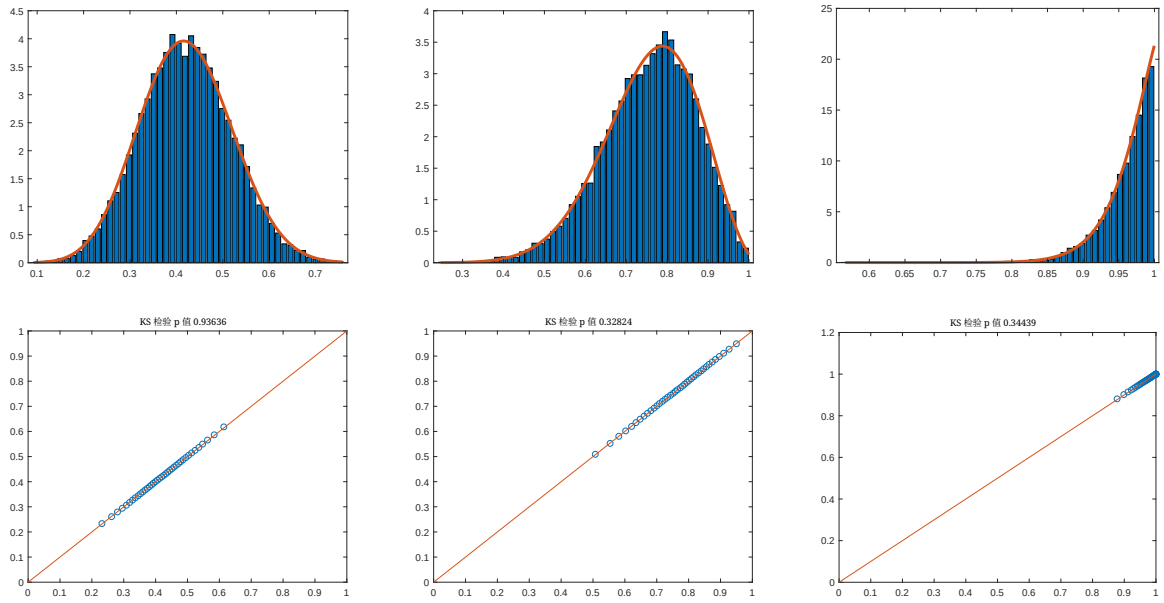


图 11：（上排）：根据表 6 所示的 Wright–Fisher 扩散桥「桥 1」（允许在 1 处被吸收）的分布生成的 10,000 个样本的直方图，采样时间为表 7 对应行中给出的时间。请注意，上述图中未包含等于 1 的样本，但其相对频率可在表 8 中查到。截断转移密度以红色绘出。（下排）：对应样本的 QQ 图，图上方标出了柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

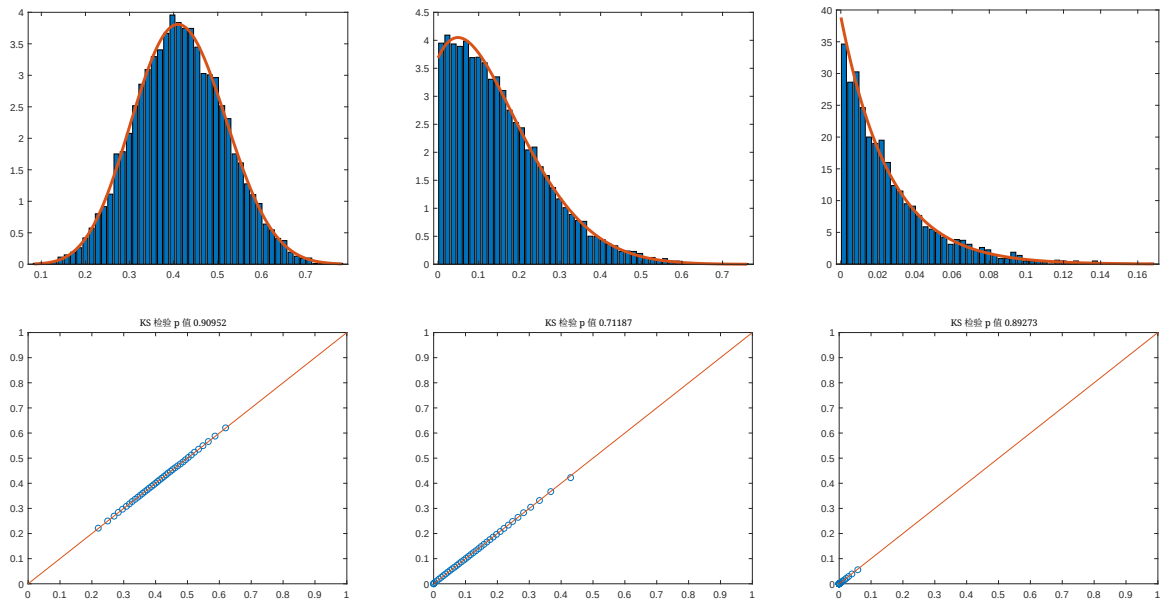


图 12：（上排）：直方图，展示根据表 6 所示的 Wright–Fisher 扩散桥“桥 2”（允许在 0 处被吸收）的分布生成的 10,000 个样本；样本取自表 7 对应行给出的时刻。请注意，上述图中未包含等于 0 的样本，但其相对频率可在表 9 中查到。截断转移密度以红色绘出。（下排）：对应样本的 QQ 图，图上方标出了柯尔莫哥洛夫–斯米尔诺夫检验返回的 p 值。

7.5 非中性扩散过程与扩散桥

非中性的 Wright–Fisher 路径可以通过中性路径与适当的泊松点过程耦合来生成（如第 4 节所述）。这项技术由 Jenkins 和 Span`o（2017年）提出，并用于当前版本的 EWF（未作任何改动），以从扩散过程及扩散桥的规律中生成非中性样本。因此，尽管 EWF 确实支持在非常广泛的一类选择机制下生成非中性样本（相关操作说明可在相应的配置文件中找到），我们仍省略生成的输出。

参考文献

Bollback, J. P. 等（2008）。基于时间序列等位基因频率数据估计 $2N_e s$。《遗传学》，179(1)，497–502。

Dangerfield, C. E. 等（2012）。一种用于含突变的 Wright– Fisher 模型的保持边界的数值算法。《BIT 数值数学》，52(2)，283–304。

Fages, A. 等（2019）。通过大规模古代基因组时间序列追踪五千年来的马匹管理。《细胞》，177，1419–1435.e31。

Fitzsimmons, P. 等（1993）。马尔可夫桥：构造、Palm解释与拼接。载于《随机过程研讨会, 1992》，第101–134页。施普林格。Griffiths, R.（1979）。多等位基因扩散模型的转移密度展开。《应用概率进展》，11(2)，310–325。Griffiths, R. C.（1984）。渐近谱系分布。《数学生物学杂志》，21(1)，67–75。Griffiths, R. C. 等（2018）。Wright–Fisher 扩散桥。《理论种群生物学》，122，67–77。Jenkins, P. A. 和 Span`o, D.（2017）。Wright–Fisher 扩散的精确模拟。《应用概率年刊》，27(3)，1478–1509。Ludwig, A. 等（2009）。马驯化初期的毛色变异。《科学》，324(5926)，485–485。

Malaspinas, A.-S. 等（2012）。根据时间序列数据估计等位基因年龄和选择系数。《遗传学》，192(2), 599–607。

Mathieson, I. 和 McVean, G.（2013）。根据等位基因频率时间序列数据估计空间结构化种群中的选择系数。《遗传学》，193(3), 973–984。

Schraiber, J. G. 等（2016）。根据等位基因频率时间序列进行自然选择的贝叶斯推断。《遗传学》，203(1), 493–511。

Steinr¨ucken, M. 等（2016）。SpectralTDF：具有时变选择参数、突变率和有效种群大小的扩散过程的转移密度。《生物信息学》，32(5), 795–797。

Tavar´e, S.（1984）。谱系与系谱过程及其在种群遗传学模型中的应用。《理论种群生物学》，26(2), 119–164。

Wutke, S. 等人（2016）。马匹的斑点表型在中世纪失去了吸引力。《科学报告》，6。