**Comments on Diffusion equations vs Wright-Fisher discrete models**

Diffusion approximations, including those developed by Kimura, are powerful tools in population genetics, providing elegant analytical solutions for phenomena like allele fixation and time to fixation. However, their accuracy is **not uniform across all ranges of population size (N) and selection coefficient (s)**.

Here's a breakdown of their accuracy and limitations:

**When Diffusion Approximations are Generally Accurate:**

1. **Large Population Sizes (Large N):** Diffusion models treat allele frequency as a continuous variable and the changes in allele frequency as a continuous-time process. This approximation holds best when the population is large enough that allele frequency changes by small increments each generation. In large populations, the "steps" of the discrete Wright-Fisher model become sufficiently small that they resemble a continuous diffusion.
2. **Weak Selection (Small 's' such that Ns is not extremely large):** When selection is weak, the deterministic change in allele frequency per generation is small. This allows drift to remain a significant factor, and the continuous approximation of the stochastic process remains valid.
3. **Intermediate Allele Frequencies:** Diffusion approximations tend to be less accurate when allele frequencies are very close to 0 or 1. At these boundaries, the "continuity" assumption can break down because the allele count is an integer, and the probability of moving from 1 copy to 0 (loss) or from 2N−1 copies to 2N (fixation) is a discrete event.

**When Diffusion Approximations May Be Less Accurate (Limitations):**

1. **Small Population Sizes (Small N):** In very small populations, allele frequency changes are inherently discrete and large (e.g., a single allele loss or gain represents a significant frequency shift). The continuous approximation can be poor, and a discrete Markov chain approach (which underlies the Wright-Fisher simulation) is generally more accurate.
2. **Strong Selection (Large 's' or Large Ns):** When selection is very strong, the allele frequency can change rapidly and deterministically, often pushing the allele quickly towards fixation or loss. In these scenarios, drift plays a smaller role, and the continuous approximation of the stochastic process might not capture the rapid discrete jumps accurately. Some recent research suggests that diffusion theory *can* be applied in regimes of strong selection under certain conditions, but this often involves more complex forms of the diffusion equations.
3. **Initial Allele Frequencies Near Boundaries:** As mentioned, if the allele frequency starts very close to 0 or 1 (e.g., a single new mutation, p0​=1/(2N)), the behavior is heavily influenced by discrete sampling early on, which the continuous approximation may struggle to perfectly capture.
4. **Complex Population Structures or Dynamics:** Diffusion approximations are typically developed for ideal, well-mixed populations. For more complex scenarios like structured populations, varying population sizes, overlapping generations, or complex mating systems, the standard diffusion models might need significant modifications or might not be suitable, and individual-based simulations become more reliable.
5. **Time-Dependent Probabilities:** While diffusion theory can provide time-dependent probabilities of fixation or loss, numerically solving these equations can sometimes be challenging, and approximations might lead to inaccuracies, particularly at early time points or if the series convergence is slow.

In summary, diffusion approximations are incredibly valuable for providing analytical insights and general tendencies in population genetics, especially for large populations and weak selection. However, for precise numerical predictions across *all* parameter ranges, particularly in small populations or under strong selection, direct simulations like the Wright-Fisher model are often considered more accurate as they directly implement the underlying discrete stochastic process without continuous approximations.

Literature

Ragsdale AP, Moreau C, Gravel S. Genomic inference using diffusion models and the allele frequency spectrum. Curr Opin Genet Dev. 2018 Dec;53:140-147. doi: 10.1016/j.gde.2018.10.001.

Paula Tataru, Maria Simonsen, Thomas Bataillon, Asger Hobolth, Statistical Inference in the Wright–Fisher Model Using Allele Frequency Data, *Systematic Biology*, Volume 66, Issue 1, January 2017, Pages e30–e46, <https://doi.org/10.1093/sysbio/syw056>

<https://www.khanacademy.org/science/ap-biology/natural-selection/population-genetics/a/genetic-drift-founder-bottleneck>