Statistical Analysis of Hairpins and BasePairs in RNA Secondary Structures

AJ BU , Manuel KAURES, and Doron ZEILBERGER

In memory of Kequan Ding (?-2025)

Abstract: We find precise asymptotic expressions for the expectations, variances, covariance, and quite a few further mixed moments for number of hairpins and basepairs in RNA secondary structure and give convincing evidence that the central-scaled distribution of the pair of random variables (hairpins,basepairs) tends in distribution to the bi-variate normal distribution with correlation $\sqrt{5\sqrt{5}-11}/2 = 0.2123322205\ldots$

Theorem: Let X_n and Z_n be the discrete random variable, defined on the set of RNA secondary structures. We have the following asmptotic expressions

$$E[X_n] = (1-\frac{2}{5}\sqrt{5})n \cdot \Big(1+(\frac{7}{4}+\frac{11}{20}\sqrt{5})n^{-1}+(\frac{243}{160}+\frac{99}{160}\sqrt{5})n^{-2}+(\frac{339}{160}+\frac{1029}{800}\sqrt{5})n^{-3}$$
$$+(\frac{11917563}{819200}+\frac{5400687}{819200}\sqrt{5})n^{-4}+\mathrm{O}(n^{-5})\Big)$$
$$E[Z_n] = (\frac{1}{2}-\frac{1}{10}\sqrt{5})n \cdot \Big(1+(-\frac{5}{8}-\frac{13}{40}\sqrt{5})n^{-1}+(\frac{21}{320}+\frac{21}{320}\sqrt{5})n^{-2}$$
$$+(-\frac{93}{320}-\frac{177}{1600}\sqrt{5})n^{-3}+(-\frac{13887249}{1638400}-\frac{6272793}{1638400}\sqrt{5})n^{-4}+\mathrm{O}(n^{-5})\Big)$$
$$Var[X_n] = (2-\frac{22}{25}\sqrt{5})n \cdot \Big(1+(\frac{1}{16}+\frac{23}{80}\sqrt{5})n^{-1}+(-\frac{651}{160}-\frac{177}{64}\sqrt{5})n^{-2}$$
$$+(-\frac{1208783}{81920}-\frac{3216693}{409600}\sqrt{5})n^{-3}+\mathrm{O}(n^{-4})\Big)$$
$$Var[Z_n] = (0+\frac{1}{50}\sqrt{5})n \cdot \Big(1+(1+\frac{1}{10}\sqrt{5})n^{-1}-\frac{261}{160}n^{-2}$$
$$+(-\frac{27179}{2560}-\frac{915879}{204800}\sqrt{5})n^{-3}+\mathrm{O}(n^{-4})\Big)$$
$$Corr[X_n,Z_n] = \frac{1}{2}\sqrt{5\sqrt{5}-11} \cdot \Big(1+(\frac{15}{32}+\frac{13}{32}\sqrt{5})n^{-1}+(\frac{4469}{1024}+\frac{1345}{1024}\sqrt{5})n^{-2}$$
$$+(\frac{1766711}{32768}+\frac{801983}{32768}\sqrt{5})n^{-3}+\mathrm{O}(n^{-4})\Big)$$

We have lots of evidence that the following conjecture is true, and one of us (DZ) is offering a donation of \$100 to the OEIS for its rigorous proof

Conjecture: The centralized-scaled version of the pair (X_n, Z_n), namely

$$\left(\frac{X_n - E[X_n]}{\sqrt{Var(X_n)}}, \frac{Z_n - Z[X_n]}{\sqrt{Var(Z_n)}}\right) \quad ,$$

tends, in distribution, to the bi-variate normal distribution with covariance $c = \frac{\sqrt{5\sqrt{5}-11}}{2}$, whose probability density function (pdf) is:

$$\frac{e^{-\frac{1}{2}x^2-\frac{1}{2}y^2+cxy}\sqrt{-c^2+1}}{2\pi} \quad .$$

References

[Sl] Neil A. J. Sloane, *The On-Line Encyclopedia of Integer Sequences (OEIS)*, `https://oeis.org/`.

AJ Bu, Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA.
Email: `ab1854 at math dot rutgers dot edu`

Manuel Kauers, Institute for Algebra, Johannes Kepler Universität, Altenbergerstraße 69 A-4040 Linz, Austria Email: `manuel dot kauers at jku dot at`

Doron Zeilberger, Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA.
Email: `ab1854 at math dot rutgers dot edu`

Written: **Jan. 30, 2026**.