The waiting time for a second mutation

Rinaldo B. Schinazi

University of Colorado, Colorado Springs

The model

The appearance of cancer in a tissue is thought to be the result of two or more successive mutations. We propose a stochastic model that allows for an exact computation of the distribution of the waiting time for a second mutation. This models the time of appearance of the first cancerous cell in a tissue. Our model is an alternative to the Moran model with mutations.

We now describe our model. We are interested in the time it takes for a given organ to have a first cancerous cell. We assume that all cells are in one of three stages: healthy, pre-cancerous (i.e. type 1) and cancerous (i.e. type 2). We start the process with all cells healthy. As the cells divide pre-cancerous cells may appear due to a type 1 mutation on a healthy cell. A type 2 mutation on a pre-cancerous cell makes the cell cancerous.

The number of type 1 mutations is modeled by a Poisson process with rate $\mu_1 N$. We think of μ_1 as a mutation rate and N as a division rate.

Let $N_1(t)$ be the number of type 1 mutations that have occurred up to time t. Let $T_1 < T_2 < \ldots$ be the arrival times of this Poisson process. Given that $N_1(t) = k$ and that $T_1 = t_1$, $T_2 = t_2, \ldots, T_k = t_k$ let $S_1, S_2, \ldots, S_k$ be random variables with density

$$P(S_i > t | T_i = t_i) = \exp(-\mu_2(t - t_i)) \qquad \text{for}$$

and $i = 1, \ldots, k$.

We also assume that given $T_1 = t_1$, $T_2 = t_2 \ldots T_k = t_k$, the random variables $S_1, S_2, \ldots, S_k$ are independent.

The random variables S_i are the times when a type 2 mutation appears. The minimum of these times (i.e. the first time a type 2 mutation appears in the tissue) is denoted by τ_2.

Here is our main result.

Theorem 1. *Let τ_2 be the time for the first type 2 mutation to appear. Then,*

$$P(\tau_2 > t) = \exp[\mu_1 N t(-1 + \frac{1 - \exp(-t\mu_2)}{t\mu_2})].$$

Note that

$$P(\tau_2 > t) = \exp[\mu_1 t f(t\mu_2)]^N$$

where $f(x) = -1 + \frac{1-\exp(-x)}{x}$.

• Our model is equivalent to a model with N independent cells.

Observe also that the function f decreases from 0 to -1 for x in $[0, +\infty)$.

• μ_1 and μ_2 do not hold symmetric roles in the formula. A small μ_1 cannot be compensated by a large μ_2. A large μ_1 can be compensated by a small μ_2.

As a consequence of Theorem 1 we have the following limits. To compute these limits assume that μ_1 and μ_2 are functions of N.

- Assume that

$$\lim_{N\to\infty} \frac{\mu_2}{\mu_1 N} = \alpha \in (0, +\infty).$$

Then,

$$\lim_{N\to\infty} P(\mu_1 N\tau_2 > t) = \exp[t(-1 + \frac{1-\exp(-t\alpha)}{t\alpha})].$$

- Assume that

$$\lim_{N\to\infty} \frac{\mu_2}{\mu_1 N} = +\infty.$$

Then,

$$\lim_{N\to\infty} P(\mu_1 N\tau_2 > t) = \exp(-t).$$

- Assume that

$$\lim_{N\to\infty} \mu_1 N\mu_2 = \alpha \in (0, +\infty),$$

$$\lim_{N\to\infty} \mu_2 = 0.$$

Then,

$$\lim_{N\to\infty} P(\tau_2 > t) = \exp(-\frac{1}{2}\alpha t^2).$$

In particular, the distribution of τ_2 exhibits at least three different behaviors depending on the relative magnitude of μ_1, μ_2 and N. These limits show a number of similarities with limits found for the Moran model by Durrett, Schmidt and Schweinsberg (2009) and Schweinsberg (2008).

Note also that as t approaches 0 (for fixed μ_1, μ_2 and N)

$$P(\tau_2 \leq t) \sim \frac{1}{2}\mu_1\mu_2 N t^2.$$

This is consistent with the model of Armitage and Doll (1954).

Waiting times for successive mutations have been recently studied by several authors using the Moran model with mutations, see Iwasa et al. (2004) and (2005) and Wodarz and Komarova (2005). A more mathematical approach is taken by Durrett et al. (2009) and Schweinsberg (2008).

The Moran model assumes a fixed number N of cells. Each cell lives for a mean 1 exponential time and then is replaced by a new cell chosen at random from one of the N cells. Moreover, a healthy cell mutates into a precancerous cell at rate μ_1 and a precancerous mutates into a cancerous cell at rate μ_2. Each new cell has the same number of mutations as its parent.

An attractive feature of the Moran model is that it is defined as a cell based model. A drawback is that the analysis of the model is quite involved and the only results that can be hoped for are non rigorous approximations or limits for different configurations of N, μ_1 and μ_2. In contrast, our model gives an exact formula for the waiting time distribution.

Our model is not fundamentally different from the Moran model.

The proof of Theorem 1

We have

$$P(\tau_2 > t | N_1(t) = k) =$$

$$P(S_1 > t, S_2 > t, \ldots, S_k > t | N_1(t) = k).$$

Let $T_1, T_2, \ldots, T_k$ be the arrival times of the Poisson process N_1. By definition of the random variables $S_1, S_2, \ldots, S_k$ we have for $i = 1, \ldots, k$

$$P(S_i > t | T_i = t_i) = \exp(-\mu_2(t - t_i)),$$

for $t > t_i$. Given $(T_1, T_2, \ldots, T_k)$ the random variables S_i are conditionally independent. Hence,

$$P(\tau_2 > t | N_1(t) = k) =$$

$$\int_{0<t_1<t_2<\ldots<t_k<t} \exp(-\mu_2(t-t_1)) \ldots \exp(-\mu_2(t-t_k)) f(t_1, t_2, \ldots, t_k) dt$$

where f is the density of the random vector $(T_1, T_2, \ldots, T_k)$ conditioned on $\{N_1(t) = k\}$.

A classical Poisson process result is that this conditional distribution is the order statistics distribution corresponding to k independent random variables uniformly distributed on $(0, t)$. Therefore,

$$P(\tau_2 > t | N_1(t) = k) =$$

$$\exp(-\mu_2 k t) \times$$

$$\int_{0<t_1<t_2<\ldots<t_k<t} \frac{t^k}{k!} \exp(\mu_2(t_1 + t_2 + \ldots + t_k) dt_1 \ldots dt_k.$$

In order to compute this integral we make the following remark. Let $U_1, U_2, \ldots, U_k$ be independent and uniformly distributed on $(0, t)$. Let $U_{(1)} < U_{(2)} < \ldots < U_{(k)}$ be the corresponding order statistics. We have that

$$E[\exp(\mu_2(U_{(1)} + U_{(2)} + \ldots + U_{(k)}))] =$$

$$\int_{0<u_1<u_2<\ldots<u_k<t} \frac{k!}{t^k} \exp(\mu_2(u_1 + u_2 + \ldots + u_k) du_1 \ldots du_k.$$

Observe that

$$U_{(1)} + U_{(2)} + \ldots + U_{(k)} = U_1 + U_2 + \ldots + U_k.$$

Hence,

$$E[\exp(\mu_2(U_{(1)} + U_{(2)} + \ldots + U_{(k)}))] = E[\exp(\mu_2 U_1)]^k.$$

It is easy to compute

$$E[\exp(\mu_2 U_1)] = \frac{1}{t\mu_2}(\exp(t\mu_2) - 1).$$

Therefore,

$$P(\tau_2 > t | N_1(t) = k) = \exp(-\mu_2 k t)[\frac{1}{t\mu_2}(\exp(t\mu_2) - 1)]^k =$$

$$[\frac{1}{t\mu_2}(1 - \exp(-t\mu_2))]^k.$$

Now,

$$P(\tau_2 > t) =$$

$$\sum_{k=0}^{\infty} P(N_1(t) = k) P(\tau_2 > t | N_1(t) = k) =$$

$$\sum_{k=0}^{\infty} \exp(-\mu_1 N t) \frac{(\mu_1 N t)^k}{k!} [\frac{1}{t\mu_2}(1 - \exp(-t\mu_2))]^k.$$

Summing the series yields

$$P(\tau_2 > t) = \exp(-\mu_1 N t) \exp[\frac{\mu_1 N}{\mu_2}(1 - \exp(-t\mu_2))].$$

This formula can be rewritten as

$$P(\tau_2 > t) = \exp[\mu_1 N t(-1 + \frac{1 - \exp(-t\mu_2)}{t\mu_2})].$$

The proof of Theorem 1 is complete.

References

P. Armitage P. and R. Doll (1954) The age distribution of cancer and a multistage theory of carcinogenesis. British Journal of cancer **8:** 1-12.

P. Armitage P. and R. Doll (1957) A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. British Journal of cancer **11:** 161-169.

R.N.Bhattacharya and E.C.Waymire (1990)*Stochastic processes with applications*, Wiley.

R. Durrett, D. Schmidt, and J. Schweinsberg (2009). A waiting time problem arising from the study of multi-stage carcinogenesis. Annals of Applied probability **19**, 676-718.

H.W. Hethcote and A.G. Knudson Jr (1978). Model for the incidence of embryonal cancers: Application to retinoblastoma. Proceedings of the National Academy of Sciences **75**, 2453-2457.

Y. Iwasa, F. Michor, N.L. Komarova, and M.A. Nowak (2005). Population genetics of tumor suppressor genes. J. Theoret. Biol. **233**, 15-23

Y. Iwasa, F. Michor, and M.A. Nowak (2004). Stochastic tunnels in evolutionary dynamics. Genetics **166**, 1571-1579.

A.G.Knudson (2001) Two genetic hits (more or less) to

cancer. Nature reviews Cancer **1:** 157-162.

A. G. Knudson (1971). Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA **68**, 820-823.

N.L. Komarova, A. Sengupta, and M.A. Nowak (2003). Mutation-selection networks of cancer initiation: Tumor suppressor genes and chromosomal insta- bility. J. Theoret. Biol. **223**, 433-450.

S.H. Moolgavkar and A.G. Knudson (1981). Mutation and cancer:a model for human carcinogenesis. J.Natl. Cancer. Inst **66**, 1037-1052.

S. H. Moolgavkar and E. G. Luebeck (1992). Multistage carcinogenesis: population-based model for colon cancer. J. Natl. Cancer Inst. **18**, 610-618.

J. Schweinsberg (2008). The waiting time for m mutations. Electron. J. Probab. **13**, 1442-1478.

D. Wodarz and N.L. Komarova (2005). *Computational Biology Of Cancer.* Lecture Notes And Mathematical Modeling. World Scientic, Singapore.