

Appendix 2: tail behaviour of quadratic forms in Gaussian vectors

There is no analytic expression for $f(s)$ as defined in (A.1.2), but it has a simple tail behaviour as s approaches its rightmost endpoint s_0 . We have seen above that, even when \mathbf{S} or \mathbf{M} are positive semi-definite (the most general case), the system can always be reduced to positive definite of dimension $m = \text{rank}(\mathbf{S}\mathbf{M})$. Therefore we can always apply JASCHKE et al.'s (2004) asymptotic approximation (eq. 3.23 and 3.24 p.261 of their paper with all $\lambda_i < 0$) in our context. In what follows we will denote results derived from this tail approximation by a star *. Applying the tail approximations shows that, to the leading order in $(s - s_0)$, $f(s)$ approaches

$$f(s) \underset{s \rightarrow s_0}{\approx} f^*(s) = d (s_0 - s)^{\frac{m}{2}-1}, \quad (\text{A.2.1})$$

where

$$d = \frac{e^{-\frac{1}{2} \sum_{i=1}^m \left(\frac{\delta_i}{\lambda_i}\right)^2}}{\Gamma\left(\frac{m}{2}\right) \prod_{i=1}^m \sqrt{|\lambda_i|}} \quad (\text{A.2.2})$$

is a constant depending on the λ_i and δ_j . Here we have assumed that all non-zero eigenvalues have multiplicity of 1 (i.e. they are distinct for each of m traits). With arbitrary multiplicity, the only change is in the expression of d in (A.2.2) (for details, see JASCHKE *et al.* 2004). From the above tail approximation one easily retrieves the proportion of beneficial mutations

$$p_b \underset{s_0 \rightarrow 0}{\approx} p_b^* = \int_{s=0}^{s_0} f^*(s) ds = 2d \frac{s_0^{m/2}}{m} \quad (\text{A.2.3})$$

and their pdf $f_b(s_b)$, provided that s_0 is close to 0, so that all $s_b > 0$ are also close to 0:

$$f_b(s_b) \underset{s_0 \rightarrow 0}{\approx} f_b^*(s_b) = \frac{f^*(s_b)}{p_b^*} = \frac{m}{2s_0} \left(1 - \frac{s_b}{s_0}\right)^{\frac{m}{2}-1}. \quad (\text{A.2.4})$$

This is equivalent to stating that the approximate distribution of s_b/s_0 , when s_0 is small, is a beta with shape parameters 1 and $m/2$:

$$\frac{s_b}{s_0} \sim \text{Beta}\left(1, \frac{m}{2}\right). \quad (\text{A.2.5})$$

The cumulative distribution function (cdf) of the beta distribution above ($F_\beta(x)$) is approximately equal to that of the exponential distribution (with rate $m/2$) as m gets large:

$$F_\beta\left(x = \frac{s_b}{s_0}\right) = 1 - (1-x)^{m/2} \underset{m \rightarrow \infty}{\rightarrow} 1 - e^{-m x/2}, \quad (\text{A.2.6})$$

which is why the classic FM (many independent traits, i.e. m large) yields an exponential distribution of beneficial effects, consistent with the MLM.

Finally, note that the same kind of tail behaviour as in eq. (A.2.1) is obtained with the displaced gamma approximation defined in eq. (A.1.3):

$$f_{\Gamma}(s) \underset{s \rightarrow s_0}{\approx} d'(s_0 - s)^{\beta-1}. \quad (\text{A.2.7})$$

where $d' = \alpha^{-\beta} / \Gamma(\beta)$ is a constant. However, while the exact $f(s)$ and the gamma approximation $f_{\Gamma}(s)$ have the same tail behaviour *qualitatively* (compare eqs. (A.2.1) and (A.2.7)), they differ *quantitatively*, as $d \neq d'$ and $m/2 \neq \beta$. Therefore, while the displaced gamma approximation is fairly accurate for the whole distribution of s (**Supplementary Figure 1**), it is less so for the subset of beneficial mutations $f_b(s_b)$, when s_0 is small, in which case the beta approximation in eqs (A.2.4) and (A.2.5) is the most accurate. As s_0 gets large, the displaced gamma approximation will become the most accurate for both $f(s)$ and $f_b(s_b)$. Notably, as s_0 gets close to 0, $\beta \rightarrow \beta_0 = n_e/2$, (see eq. A.1.2 and MARTIN and LENORMAND 2006), so that the discrepancy between the two tail behaviours ((A.2.1) vs. (A.2.7)) depends on the difference between the “effective number of traits” n_e and the “dimensionality” m , as discussed in the main text.

References:

- JASCHKE, S., C. KLUPPELBERG and A. LINDNER, 2004 Asymptotic behavior of tails and quantiles of quadratic forms of Gaussian vectors. *Journal of Multivariate Analysis* **88**: 252-273.
- MARTIN, G., and T. LENORMAND, 2006 A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* **60**: 893-907.