Kong VT AM-mck1

matthew.c.keller

June 30 2020

1 Path Tracing Rules

(Note: I need to change the specific wording because this is currently plagiarized from the Cascade paper)

To derive the expected covariance between two variables, one identifies all pathways or 'chains' that start at the first variable and end at the second, such that:

1. A chain may begin either by tracing backwards against the direction of a single-headed arrow, or by crossing a double-headed arrow before tracing forward in the direction of single-headed arrow(s).

2. A chain changes direction at a double-headed arrow, and moves thereafter only in the direction of single-headed arrow(s)

3. Chains must cross exactly one double-headed arrow (which implies that no chain changes direction more than once)

4. No chain is counted twice; however, it is important to keep in mind that order matters, such that fwa is not counted the same as awf, even though they are algebraically equivalent. Similarly, for variances caused by two other variables that are correlated, the covariance is counted in both directions.

5. As with other paths, co-paths can only be traced once in any chain; however, once traversed, the four tracing rules described above are 'reset'. For example, a double-headed arrow would need to be crossed again.

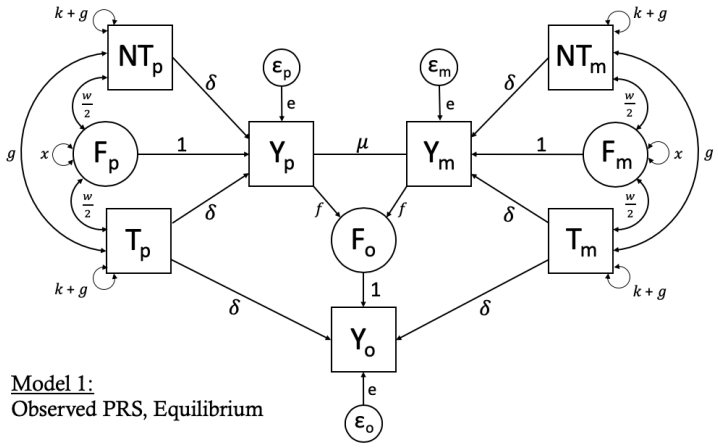
The expected covariance is found by multiplying the coefficients in the same manner, except that the goal is to find all possible, non-redundant chains that connect two variables and summing them. Variances are found in the same manner, except that the goal is to find all chains that begin at a variable and arrive back at that same variable.

2 Variable Definitions

Terms included in both Model 1 and 2
$T_{P|M}$: Transmitted observed additive genetic score (PRS) from father or mother
$NT_{P|M}$: Non transmitted PRS from father or mother
$F_{P|M|O}$: Familial shared environment factor of father, mother, or offspring
$f_{P|M}$: Effect size from paternal or maternal phenotype to F_O
σ^2: Equilibrium phenotypic variance
δ: path coefficient of haplotypic PRS to phenotype
k: variance of haplotypic PRS at time 0 (before AM); determined by how PRS is scaled
w: covariance between PRS and F
x: variance of F
g: covariance between haplotypic PRS's; $g > 0$ to the degree there is AM
μ: copath between mates

Terms included only in Model 2
$LNT_{P|M}$: Non transmitted latent additive genetic score from father or mother
$LT_{P|M}$: Transmitted latent additive genetic score from father or mother
a: path coefficient of haplotypic latent genetic score to Y; analogous to δ
j: variance of haplotypic latent genetic score at time 0; analogous to k
v: covariance between latent genetic score and F; analogous to w
h: covariance between haplotypic latent genetic scores; analogous to g
i: covariance between haplotypic latent genetic score and haplotypic PRS

3 Model 1: VT + equilibrium AM, no latent genes



At Equilibrium

$\Omega = cov(Y_{m|p}, [N]T_{m|p}) = \delta k + 2\delta g + \frac{1}{2}w$

$g = cov([N]T_{m|p}, [N]T_{m|p}) = \Omega^2\mu$

$\theta_{NT} = cov(Y_o, NT_m + NT_p) = 2cov(Y_o, NT_m) = 2(2\delta g + f\Omega(1+\sigma^2\mu)) = 4\delta g + w$
$\theta_{NT} = 2(\Omega - \delta k)$

$\theta_T = cov(Y_o, T_m + T_p) = 2cov(Y_o, T_m) = 2(\delta k + 2\delta g + \frac{1}{2}w) = 2\delta(k + 2g) + w$
$\theta_T = \theta_{NT} + 2\delta k$
$\theta_T = 2\Omega$

$\theta_T - \theta_{NT} = \delta 2k$

$w = cov(F_{p|m|o}, T_{m|p} + NT_{m|p}) = 2cov(F_{p|m|o}, [N]T_{m|p}) = 2(f\Omega + f\Omega\sigma^2\mu)$
$w = 2f\Omega(1 + \sigma^2\mu)$

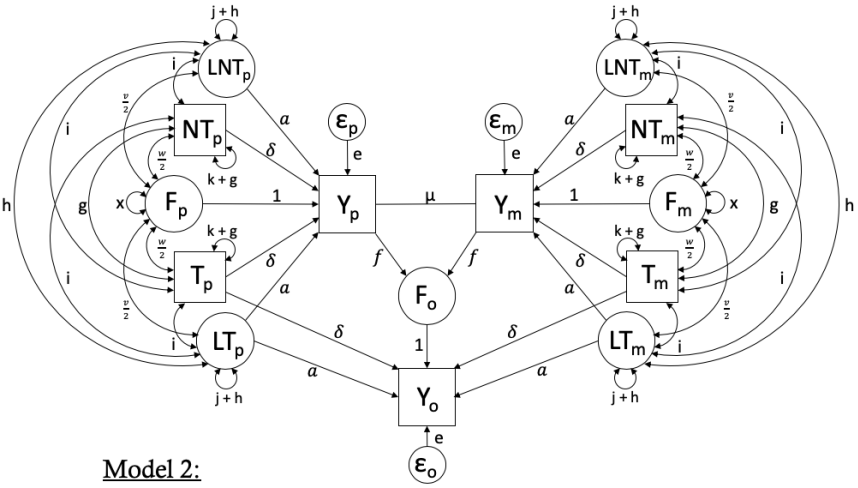
$x = var(F_{m|p|o}) = 2f^2\sigma^2(1 + \sigma^2\mu)$

$\sigma^2_{p|m|o} = 2\delta\Omega + \delta w + x + e^2$
$\sigma^2_{p|m|o} = 2\delta^2 k + 4\delta^2 g + 2\delta w + x + e^2$

Rules for k

1. If the PRS is standardized at time 0 (before AM), then $k = \frac{1}{2}$.
2. If the PRS is standardized at equilibrium to have a variance of 1, then $k = \frac{1}{2} - 2g$.
3. If the *haplotypic* PRS is scaled at equilibrium to have a variance of $\frac{1}{2}$, then $k = \frac{1}{2} - g$.

4 Model 2: VT + equilibrium AM, with latent genetics and observed parental phenotypes



Model 2:
Observed & Latent genetics, Equilibrium AM

At Equilibrium

$\Omega = cov(Y_{m|p}, [N]T_{m|p}) = \delta k + 2\delta g + 2ai + \frac{1}{2}w$

$\Gamma = cov(Y_{m|p}, L[N]T_{m|p}) = aj + 2ah + 2\delta i + \frac{1}{2}v$

$g = cov([N]T_{m|p}, [N]T_{m|p}) = \Omega^2\mu$

$h = cov(L[N]T_{m|p}, L[N]T_{m|p}) = \Gamma^2\mu$

$i = cov([N]T_{m|p}, L[N]T_{m|p}) = \Gamma\mu\Omega = \sqrt{gh}$

$\theta_{NT} = cov(Y_o, NT_m + NT_p) = w + 4\delta g + 4ai$
$\theta_{NT} = 2(\Omega - \delta k)$

$\theta_T = cov(Y_o, T_m + T_p) = 2\delta k + w + 4\delta g + 4ai$
$\theta_T = \theta_{NT} + 2\delta k$
$\theta_T = 2\Omega$

$\theta_T - \theta_{NT} = \delta 2k$

$\theta_{LNT} = cov(Y_o, LNT_m + LNT_p) = v + 4ah + 4\delta i$

$\theta_{LT} = cov(Y_o, LT_m + LT_p) = 2aj + v + 4ah + 4\delta i$
$\theta_{LT} = 2aj + \theta_{LNT}$

$w = cov(F_{p|m|o}, T_{m|p} + NT_{m|p}) = 2f\Omega(1 + \sigma^2\mu)$

$v = cov(F_{p|m|o}, LT_{m|p} + LNT_{m|p}) = 2f\Gamma(1 + \sigma^2\mu)$

$x = var(F_{m|p|o}) = 2f^2\sigma^2(1 + \sigma^2\mu)$

$\sigma^2_{p|m|o} = 2a\Gamma + 2\delta\Omega + av + \delta w + x + e^2$
$\sigma^2_{p|m|o} = 2a^2 j + 4a^2 h + 2av + 2\delta^2 k + 4\delta^2 g + 2\delta w + 8a\delta i + x + e^2$

$cov(Y_o, Y_p) = cov(Y_o, Y_m) = a\Gamma + \delta\Omega + f\sigma^2 + a\Gamma\mu\sigma^2 + \delta\Omega\mu\sigma^2 + f\sigma^2\mu\sigma^2$
$cov(Y_o, Y_p) = cov(Y_o, Y_m) = (a\Gamma + \delta\Omega + f\sigma^2)(1 + \sigma^2\mu)$

Fitting Model 2

Model 2 has only one additional informative statistic (the covariance between parents and offspring) but 6 new parameters: Γ, i, j, a, v and h. Clearly this is not an identified model. However, as explained below, only one of these parameters (a) actually needs to be estimated (it uses the covariance between parent and offspring to do so), whereas the rest of these parameters are either constants, constraints implied by the model, or constraints that stem from reasonable assumptions. In particular:

1 of these is an estimate, as noted above:
a

1 of these is a constant:
j, which is the assumed variance of the haplotypic latent genetic score at time 0 (before AM) and depends on how the PRS is scaled, as explained below

2 of these are constraints that are implied by the structural model:
$\Gamma = cov(Y_{m|p}, L[N]T_{m|p}) = aj + 2ah + 2\delta i + \frac{1}{2}v$
$i = cov([N]T_{m|p}, L[N]T_{m|p}) = \Gamma\mu\Omega = \sqrt{gh}$

and finally, 2 of these require assumptions that lead to constraints (both of these have been verified via simulation):
$\frac{v}{a} = \frac{w}{\delta}$, which states the ratio of G-E covariance to direct genetic effects is assumed to be the same for observed as latent genetic effects. This leads to the constraint that $v = \frac{wa}{\delta}$
$\frac{g}{\delta^2} = \frac{h}{a^2}$, which states that the increase in additive genetic variance due to AM for some subset of the genes is assumed to be proportionate to the additive genetic variance at time 0 for that subset of genes. For our model, the two

subsets of genes are those that make up the PRS and those that make up the latent genetic score. This leads to the constraint that $h = \frac{ga^2}{\delta^2}$

Rules for k and j

Unlike k, which depends upon how the PRS is scaled, j can take any arbitrary value and the model will fit. A natural choice might be to have $j = 1$. However, a simpler choice, and the one that we have assumed in our path diagram and math above, is to have j be defined analogously to k. This keeps the expectations for parameters related to the latent and observed genetic scores consistent with one another and thereby greatly simplifies the math. In particular:

1. If the PRS is standardized at time 0 (before AM), then $k = j = \frac{1}{2}$.
2. If the PRS is standardized at equilibrium to have a variance of 1, then $k = \frac{1}{2} - 2g$ and $j = \frac{1}{2} - 2h$.
3. If the *haplotypic* PRS is scaled at equilibrium to have a variance of $\frac{1}{2}$, then $k = \frac{1}{2} - g$ and $j = \frac{1}{2} - h$.