Stat 8931 (Aster Models)
Lecture Slides Deck 5

Charles J. Geyer

School of Statistics
University of Minnesota

June 7, 2015

Lande-Arnold Analysis

Lande and Arnold (*Evolution*, 1983) proved some theorems about natural and artificial selection and fitness landscapes and introduced a method of analysis that has become very widely used (Google Scholar says "Cited by 2858" as I write this).

The Lande-Arnold approach has been criticized (Mitchell-Olds and Shaw, *Evolution*, 1987). (Google Scholar says "Cited by 638" for that.) But the popularity of Lande-Arnold methodology continues.

The reason (IMHO) for the popularity is that simple statistical methodology — ordinary least squares (OLS) regression — is coupled with a sophisticated mathematical interpretation. Anyone can easily do the analysis, and the results are considered (by conventional wisdom) to be very important.

These slides follow a recent paper (Shaw and Geyer, *Evolution*, 2010) and its accompanying technical reports (U of M School of Statistics TR 670, TR 671 revised, TR 674 revised, and TR 675 revised). (Google Scholar only says "Cited by 20" for this paper.)

This shows how to combine the mathematics of Lande and Arnold with defensible statistical analysis (using aster models).

It avoids most (but not all) of the problematic aspects of Lande-Arnold methodology.

Phenotype and Genotype

All the biologists know, but statisticians may not, biologists distinguish between phenotype and genotype.

The **phenotype** of an organism is the "type" characterized by the organism's observable characteristics (morphology, behavior, whatever). The characteristics themselves are **phenotypic traits**.

The **genotype** of an organism is the "type" characterized by the organism's genetic material. The characteristics themselves are **genotypic traits** or just **genes**.

Throughout the twentieth century genotypes became closer and closer to being "observable" until now it only costs a few hundred dollars (and the time to send a tissue sample to a sequencing lab) to "observe" the genotype of an organism.

So phenotype and genotype are no longer divided by a philosophically clear line. Colloquially "phenotype" means observable by "ordinary" means. But this distinction is not philosophically clear either. For example, observe whether an adult mammal gets an upset stomach after drinking milk. That "phenotype" directly indicates the existence of a mutation in the enzyme lactase, which is part of the "genotype".

Nevertheless the distinction is usually clear enough in context and continues to be very useful.

The **fitness landscape for phenotypic traits** is the conditional expectation of (Darwinian) fitness given a vector of phenotypic traits

$$g(z) = E(w \mid z)$$

where w is observed fitness (number of offspring or the best surrogate thereof in the data) and z is the vector of phenotypic traits.

The **fitness landscape for genotypic traits** is the same except z is a vector of genotypic traits.

Another term for fitness landscape is **adaptive landscape**.

Yet another term is **individual selection surface**.

According Arnold (*Integrative and Comparative Biology*, 2003) the original fitness landscape idea is the genotypic one (Wright, *Proceedings of the Sixth International Congress on Genetics*, 1932). The phenotypic one came later (Simpson, *Tempo and Mode in Evolution*, 1944).

For the statisticians, who may not know, Sewall Wright and George Gaylord Simpson are two of the seminal figures of the "neo-Darwinian synthesis that merged Darwin's theory of evolution by natural selection with Mendelian genetics. They are more famous than any statistician except R. A. Fisher (who is most famous for also being one of the seminal figures of the neo-Darwinian synthesis — statistics was just a sideline for him).

The idea is that causation flows

$$genotype \longrightarrow phenotype \longrightarrow fitness$$

so the genotypic fitness landscape is more "fundamental" (the scare quotes are because this is simplistic).

Nevertheless, the phenotypic fitness landscape is more easily "observed" so that is what is most widely studied.

Lande-Arnold Analysis (cont.)

Both kinds of fitness landscape say important things about evolution.

Over the course of time natural selection tends to move a population toward a local maximum of the fitness landscape.

In particular, the change in phenotype expected under one generation of selection can be read off the phenotypic fitness landscape (and similarly with genotype replacing phenotype and genotypic replacing phenotypic).

Lande and Arnold established an important connection allowing (under certain assumptions) the change in *genotype* expected under one generation of selection can be read off the *phenotypic* fitness landscape! That is one reason for the importance of their methodology.

$$g(z) = E(w \mid z)$$

Any conditional expectation, thought of as a function of the variables "behind the bar" is called a **regression function** by statisticians. The fitness landscape for phenotypes is the regression function of fitness w on a vector z of phenotypic traits.

Everything statisticians know about regression functions can be applied to the problem. In particular, we are going to apply aster models. (So in this application, the connection between aster models and regression theory is very strong.)

But we are going to follow Lande and Arnold (1983) and our TR 670 rather than regression textbooks.

Lande and Arnold introduced β and γ as important concepts. TR 670 and the paper (Shaw and Geyer, 2010) show that actually these concepts split into four betas and three gammas.

Define

$$Q(\alpha, \beta, \gamma) = E\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)^2\}, \qquad (1)$$

where α is a scalar, β a vector, and γ a symmetric matrix. Let α_1 and β_1 be the values that minimize $Q(\alpha, \beta, 0)$ and let α_2, β_2, and γ_2 be the values that minimize $Q(\alpha, \beta, \gamma)$. Then

$$g_1(z) = \alpha_1 + z^T\beta_1 \qquad (2a)$$
$$g_2(z) = \alpha_2 + z^T\beta_2 + \tfrac{1}{2}z^T\gamma_2 z \qquad (2b)$$

are the **best linear approximation (BLA)** and the **best quadratic approximation (BQA)** to the fitness landscape $g(z)$, where "best" means minimizing (1).

$$Q(\alpha, \beta, \gamma) = E\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)^2\} \qquad (1)$$

The function (1) is well defined so long as second moments of w and fourth moments of z exist. Then (1) is convex quadratic in α, β, and γ, so minimizers always exist.

First considering α_1 and β_1, if there exist distinct minimizers of $Q(\alpha, \beta, 0)$, say with and without stars, then by convexity the whole line segment between them must have the same value, that is

$$Q\big(s\alpha_1^* + (1 - s)\alpha_1, s\beta_1^* + (1 - s)\beta_1, 0\big)$$

is a constant function of s.

$$Q\big(s\alpha_1^* + (1-s)\alpha_1, s\beta_1^* + (1-s)\beta_1, 0\big)$$
$$= E\big\{(w - \alpha_1 - z^T\beta_1 - s[\Delta\alpha + z^T\Delta\beta])^2\big\}$$
$$= Q\big(\alpha_1, \beta_1, 0\big) - 2sE\big\{(w - \alpha_1 - z^T\beta_1)(\Delta\alpha + z^T\Delta\beta)\big\}$$
$$+ s^2 E\big\{(\Delta\alpha + z^T\Delta\beta)^2\big\}$$

where $\Delta\alpha = \alpha_1^* - \alpha_1$ and $\Delta\beta = \beta_1^* - \beta_1$, and this can only be a constant function of s if its derivative with respect to s

$$-2sE\big\{(w - \alpha_1 - z^T\beta_1)(\Delta\alpha + z^T\Delta\beta)\big\} + s^2 E\big\{(\Delta\alpha + z^T\Delta\beta)^2\big\}$$

is zero for all s. And this happens if and only if both expectations in the formula above are zero.

And

$$E\{(\Delta\alpha + z^T\Delta\beta)^2\}$$

is zero if and only if

$$\Delta\alpha + z^T\Delta\beta = 0, \qquad \text{almost surely,}$$

that is, if and only if z is a degenerate random vector. But if that holds, then

$$E\{(w - \alpha_1 - z^T\beta_1)(\Delta\alpha + z^T\Delta\beta)\}$$

is zero too. Thus we have proved α_1 and β_1 are uniquely defined if and only if z is a nondegenerate random vector.

A similar analysis says that α_2, β_2, and γ_2, are uniquely defined if and only if there do not exist a scalar $\Delta\alpha$, a vector $\Delta\beta$, and a symmetric matrix $\Delta\gamma$, not all zero, such that

$$\Delta\alpha + z^T \Delta\beta + \tfrac{1}{2} z^T \Delta\gamma z = 0, \qquad \text{almost surely.}$$

This is a more general issue than what we have been calling degeneracy. This holds if and only if z is concentrated on a curved manifold having dimension lower than the dimension of z. We won't need that concept except right here. We can call that "curvilinear degeneracy". Using that notion, we have proved α_2, β_2, and γ_2 are uniquely defined if and only if z is not a curvilinear degenerate random vector.

Define

$$\beta_3 = E\{\nabla g(z)\} \tag{3a}$$

$$\gamma_3 = E\{\nabla^2 g(z)\} \tag{3b}$$

where (as always in these slides) $\nabla g(z)$ denotes the vector of partial derivatives $\partial g(z)/\partial z_i$ and $\nabla^2 g(z)$ denotes the matrix of second partial derivatives $\partial^2 g(z)/\partial z_i \partial z_j$.

Note that β_3 is not the gradient of the fitness landscape at any point (not $\nabla g(z)$ for any z). Rather it is the average gradient of the fitness landscape, averaged over all points. Similarly for γ_3.

Define

$$\beta_4 = \Sigma^{-1} \operatorname{cov}(w, z) \qquad (4a)$$

$$\gamma_4 = \Sigma^{-1} \operatorname{cov}(w, zz^T)\Sigma^{-1} \qquad (4b)$$

where $\Sigma = \operatorname{var}(z)$.

What are they? They are related to changes in phenotype or genotype from one generation of selection (more on this later).

One might wonder about the point of the preceding, but for the following.

Theorem (Lande-Arnold). If z is a multivariate normal random vector having mean vector zero and nonsingular variance matrix, and under no assumptions about w except that enough moments exist, then $\beta_1 = \beta_2 = \beta_3 = \beta_4$ and $\gamma_2 = \gamma_3 = \gamma_4$.

Comment 1. The equalities $\beta_1 = \beta_3 = \beta_4$ and $\gamma_2 = \gamma_3$ hold even if z is not centered at zero (more on this later).

Comment 2. The equality $\beta_1 = \beta_4$ holds without multivariate normality of z, only requiring enough moments of w and z.

The Lande-Arnold theorem is interesting because it says that (assuming multivariate normality of z) β and γ capture several different aspects of the fitness landscape.

- Linear and quadratic regression approximations (1st and 2nd concepts).
- Average first and second derivatives (3rd concepts).
- What? (4th concepts). Actually something about biology (more on this later).

Univariate normality can be assured: replace data by normal scores or an approximation thereto, for example, the R function qqnorm essentially does

```
plot(qnorm(ppoints(length(x))), sort(x))
```

so

```
qnorm(ppoints(length(x))[rank(x)]
```

transforms any continuous random variable (no ties!) to near perfect univariate normality.

But (a very big but) there is no such trick for multivariate normality.

Multivariate normality is impossible to transform to and also impossible to check.

So if z is multivariate, there are essentially 3 betas and 3 gammas, all different (we usually have $\beta_1 = \beta_4$, but no other equalities).

$$Q(\alpha, \beta, \gamma) = E\left\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)^2\right\} \tag{1}$$

Here we don't assume anything about the distribution of w and z except that the moments that appear in formulas exist.

Define $\eta = E(w)$ and $\mu = E(z)$. Then

$$\frac{\partial Q(\alpha, \beta, 0)}{\partial \alpha} = -2E\left\{w - \alpha - z^T\beta\right\}$$
$$\frac{\partial Q(\alpha, \beta, 0)}{\partial \beta_i} = -2E\left\{(w - \alpha - z^T\beta)z_i\right\}$$

Set equal to zero

$$\eta - \alpha - \mu^T\beta = 0 \tag{$*$}$$
$$E(wz) - \alpha\mu - \beta E(zz^T)\beta = 0 \tag{$**$}$$

$$\eta - \alpha - \mu^T \beta = 0 \qquad (*)$$
$$E(wz) - \alpha\mu - \beta E(zz^T)\beta = 0 \qquad (**)$$
$$\beta_4 = \Sigma^{-1} \operatorname{cov}(w, z) \qquad (4a)$$

Solving $(*)$ for α gives

$$\alpha = \eta - \mu^T \beta$$

and plugging this into $(**)$ gives

$$E(wz) - \eta\mu - \left[E(zz^T) - \mu\mu^T\right]\beta = \operatorname{cov}(w, z) - \Sigma\beta = 0$$

and solving for β gives the right-hand side of (4a).

Proof: $\beta_2 = \beta_4$ and $\gamma_2 = \gamma_4$

$$Q(\alpha, \beta, \gamma) = E\big\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)^2\big\} \qquad (1)$$

Now we do assume z is mean-zero multivariate normal.

$$\frac{\partial Q(\alpha, \beta, \gamma)}{\partial \alpha} = -2E\big\{w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z\big\}$$

$$\frac{\partial Q(\alpha, \beta, \gamma)}{\partial \beta_i} = -2E\big\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)z_i\big\}$$

$$\frac{\partial Q(\alpha, \beta, \gamma)}{\partial \gamma_{jk}} = -E\big\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)z_j z_k\big\}$$

Set equal to zero and solve.

$$E\{w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z\} = 0 \qquad (*)$$

$$E\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)z_i\} = 0 \qquad (**)$$

$$E\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)z_j z_k\} = 0 \qquad (***)$$

As before, let $E(w) = \eta$. Setting $(*)$ equal to zero gives

$$\eta - \alpha - \tfrac{1}{2}\operatorname{tr}(\gamma\Sigma) = 0$$

(the term involving first moments of z drops out) and solving for α gives

$$\alpha = \eta - \tfrac{1}{2}\operatorname{tr}(\gamma\Sigma)$$

$$E\left\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)z_i\right\} = 0 \qquad (**)$$

$$\beta_4 = \Sigma^{-1}\operatorname{cov}(w, z) \qquad (4a)$$

Setting $(**)$ equal to zero gives

$$E\left\{(w - z^T\beta)z_i\right\} = 0$$

(the terms involving only first and third moments of z drop out). Rewriting this in vector form gives

$$\operatorname{cov}(w, z) = E(zz^T)\beta = \Sigma\beta$$

and solving for β again gives the right-hand side of (4a).

$$E\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)z_j z_k\} = 0 \qquad (***)$$
$$\alpha = \eta - \tfrac{1}{2}\operatorname{tr}(\gamma\Sigma) \qquad (5)$$

In $(***)$ terms involving only third moments of z drop out giving

$$E\{(w - \alpha - \tfrac{1}{2}z^T\gamma z)z_j z_k\} = 0$$

and then plugging in our solution for α gives

$$E\{(w - \eta + \tfrac{1}{2}\operatorname{tr}(\gamma\Sigma) - \tfrac{1}{2}z^T\gamma z)z_j z_k\} = 0 \qquad (****)$$

We need the identity about fourth moments of the multivariate normal distribution

$$E(z_i z_j z_k z_l) = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk} \tag{6}$$

that can be looked up in multivariate analysis books or derived by differentiating the moment generating function. In it σ_{ij} are the components of the variance matrix Σ.

$$E\left\{(w - \eta + \tfrac{1}{2}\operatorname{tr}(\gamma\Sigma) - \tfrac{1}{2}z^T\gamma z)z_j z_k\right\} = 0 \qquad (****)$$

$$E(z_i z_j z_k z_l) = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk} \qquad (6)$$

By (6)

$$
\begin{aligned}
E\left\{(z^T\gamma z)z_j z_k\right\} &= \sum_i \sum_l \gamma_{il} E(z_i z_j z_k z_l) \\
&= \sum_i \sum_l \gamma_{il}\left[\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}\right] \\
&= 2(\Sigma\gamma\Sigma)_{jk} + \operatorname{tr}(\gamma\Sigma)\sigma_{jk}
\end{aligned}
$$

where the first term on the last line means the j, k term of the matrix $\Sigma\gamma\Sigma$.

Plugging this back into $(****)$ gives

$$E\left\{(w - \eta)z_j z_k\right\} - (\Sigma\gamma\Sigma)_{jk} = 0 \qquad (*****)$$

$$E\{(w - \eta)z_j z_k\} - (\Sigma\gamma\Sigma)_{jk} = 0 \qquad (*****)$$
$$\gamma_4 = \Sigma^{-1} \text{cov}(w, zz^T)\Sigma^{-1} \qquad (4b)$$

Rewriting $(*****)$ in vector form gives

$$\text{cov}(w, zz^T) = \Sigma\gamma\Sigma$$

and solving for γ gives the right-hand side of (4b).

The proof is done, but note that we used the multivariate normality assumption in a very strong way via the identity (6). In order that $\gamma_2 = \gamma_4$ we need the relation between the first four moments of z to be exactly that of the multivariate normal distribution. This is not something we are likely to achieve in any real data.

On the other hand the proof of $\beta_2 = \beta_4$ only required that z have mean zero and that the moments that appear in the equations exist.

This part of the proof uses integration by parts. First we give a quick sketch of the proof, then the details. Statisticians call this use of integration by parts Stein's lemma (Stein, *Annals of Statistics*, 1981). Lande and Arnold independently invented it.

$$\beta_3 = E\{\nabla g(z)\} \tag{3a}$$
$$\beta_4 = \Sigma^{-1} \text{cov}(w, z) \tag{4a}$$

Again z is multivariate normal with mean vector zero and non-singular variance matrix Σ. Let f denote the marginal PDF of z, and let g denote the fitness landscape. Then

$$\begin{aligned}
\beta_3 &= \int \nabla g(z) f(z)\, dz \\
&= -\int g(z) \nabla f(z)\, dz \\
&= \Sigma^{-1} \int z g(z) f(z)\, dz \\
&= \Sigma^{-1} \text{cov}\{g(z), z\}
\end{aligned}$$

which agrees with the right-hand side of (4a) by the iterated expectation theorem.

$$\gamma_3 = E\{\nabla^2 g(z)\} \tag{3b}$$

$$\gamma_4 = \Sigma^{-1} \operatorname{cov}(w, zz^T)\Sigma^{-1} \tag{4b}$$

Similarly,

$$
\begin{aligned}
\gamma_3 &= \int \nabla^2 g(z) f(z)\, dz \\
&= \int g(z) \nabla^2 f(z)\, dz \\
&= \int \left[\Sigma^{-1} zz^T \Sigma^{-1} - \Sigma^{-1}\right] g(z) f(z)\, dz \\
&= \Sigma^{-1} \operatorname{cov}\{g(z), zz^T\}\Sigma^{-1}
\end{aligned}
$$

which agrees with the right-hand side of (4b) by the iterated expectation theorem.

The integration by parts formula is

$$\int u\, dv = uv - \int v\, du.$$

It allows us to flip derivatives from one part of an integrand to another but there is the *uv* part to deal with.

The integration by parts is a single-integral technique. Here we have multiple integrals, so it must be applied to one of the multiple integrals. Look at one partial derivative.

$$\int_{-\infty}^{+\infty} \frac{\partial g(z)}{\partial z_i} f(z)\, dz_i = g(z)f(z)\Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} g(z)\frac{\partial f(z)}{\partial z_i}\, dz_i$$

We want the first term on the right-hand side to be zero, which it certainly will be if this term is integrable, that is, if $E\{g(z)\}$ exists.

Similarly,

$$\int_{-\infty}^{+\infty} \frac{\partial^2 g(z)}{\partial z_i \partial z_j} f(z) \, dz_i = \left. \frac{\partial g(z)}{\partial z_j} f(z) \right|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \frac{\partial g(z)}{\partial z_j} \frac{\partial f(z)}{\partial z_i} \, dz_i$$

and the first term on the right-hand side will be zero of $E\{\nabla g(z)\}$ exists.

And

$$-\int_{-\infty}^{+\infty} \frac{\partial g(z)}{\partial z_j} \frac{\partial f(z)}{\partial z_i}\, dz_i$$

$$= -g(z)\frac{\partial f(z)}{\partial z_i}\bigg|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} g(z)\frac{\partial^2 f(z)}{\partial z_i \partial z_j}\, dz_i$$

$$= -g(z)(\Sigma^{-1}z)_i f(z)\bigg|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} g(z)\frac{\partial^2 f(z)}{\partial z_i \partial z_j}\, dz_i$$

and the first term on the right-hand side will be zero if $E\{zg(z)\}$ exists.

The iterated expectation theorem says

$$E\{E(Y \mid X)\} = E(Y)$$

for any random vectors X and Y. We already used this in deriving

$$\mu_j = \xi_j \mu_{p(j)}$$

the relationship between conditional and unconditional mean value parameters in aster models.

Here we need it to prove

$$\text{cov}(w, z) = \text{cov}\{g(z), z\}$$
$$\text{cov}(w, zz^T) = \text{cov}\{g(z), zz^T\}$$

where, as always,

$$g(z) = E(w \mid z)$$

The two proofs have the same pattern

$$
\begin{aligned}
\text{cov}\{g(z), z\} &= E\{g(z)z\} - E\{g(z)\}E(z) \\
&= E\{E(w \mid z)z\} - E\{E(w \mid z)\}E(z) \\
&= E\{E(wz \mid z)\} - E\{E(w \mid z)\}E(z) \\
&= E(wz) - E(w)E(z) \\
&= \text{cov}(w, z)
\end{aligned}
$$

The middle step on the preceding slide also used

$$h(z)E(w \mid z) = E\{h(z)w \mid z\}$$

for any function h, that is, functions of the variables "behind the bar" are treated as constants and can be moved in and out of a conditional expectation.

That finishes the proof of the Lande-Arnold theorem.

Now for the criticism.

In getting $\gamma_2 = \gamma_4$ we used multivariate normality in a very strong way: the first four moments of z exactly match those of a multivariate normal distribution.

In getting $\beta_3 = \beta_4$ and $\gamma_3 = \gamma_4$ we used multivariate normality in a very strong way: the first two derivatives of the PDF of z exactly match on average those of a multivariate normal distribution.

Of course if the first derivatives exactly match at all points, then the distributions exactly match. But we don't need an exact match at all points, only an exact match when integrated against various functions. But that is not a "natural" condition. Nothing leads us to expect that except when the derivatives match pointwise.

Because exact multivariate normality of z is such a strong condition, we do not expect it to hold in applications.

Thus the conclusion of the Lande-Arnold theorem (all the betas are the same and so are all the gammas) won't hold either.

Thus there are different betas and gammas, and it behooves the scientist to know the differences.

On the other hand, the Lande-Arnold theorem does not assume anything about the distribution of fitness w (other than that enough moments exist).

In particular it does not assume w is normal.

Because of the way Lande and Arnold (1983) suggest doing statistical inference about w (more on this later), some people say they assume w is normal. But neither the definitions of the betas and gammas nor the Lande-Arnold theorem assume any such thing.

So what does all of this have to do with selection?

The difference between the average value of z before selection and "after selection but before reproduction" (quoting Lynch and Walsh, *Genetics and Analysis of Quantitative Traits*, 1998) is

$$E(w_r z) - E(z) = \text{cov}(w_r, z) \tag{7}$$

where

$$w_r = \frac{w}{E(w)}$$

is fitness reweighted to act like a probability density (integrates to one). The quantity w_r is called **relative fitness**.

Lynch and Walsh call (7) the Robertson-Price identity, and Lande and Arnold call it "a multivariate generalization of the results of Robertson (1966) and Price (1970, 1972)."

It is not clear what the phrase "after selection but before reproduction" can mean in natural populations or in experiments where observed fitness includes components of fitness involving reproduction.

But the mathematical meaning of the left hand side of (7) is clear: the difference between the weighted average of phenotype values, weighted according to relative fitness, and the unweighted average.

What Price (1972) calls "type I selection" is more general than (7), it being "a far broader problem category than one might at first assume" but "intended mainly for use in deriving general relations and constructing theories, and to clarify understanding of selection phenomena, rather than for numerical calculation" (both quotations from Price, 1972).

The reason for the discrepancy between the narrow applicability of the theory in Lande and Arnold (1983) and the broad applicability of Price (1972) is that the theory in Price (1972) is more general: (7) corresponds to (A 13) in Price (1972) but this is only a limited special case of his (A 11) which contains an extra term on the right-hand side.

$$E(w_r z) - E(z) = \text{cov}(w_r, z) \qquad (7)$$

Nevertheless, (7), even though it doesn't properly account for reproduction, is what everybody uses because it it the only thing simple enough use in calculations.

IMHO, the covariance on the right-hand side is a red herring. In general, for any random variable w_r and and random vector z we have

$$\text{cov}(w_r, z) = E(w_r z) - E(w_r)E(z)$$

and we only get (7) because we defined w_r so that $E(w_r) = 1$.

$$E(w_r z) - E(z) = \text{cov}(w_r, z) \tag{7}$$

Even if you like woofing about covariance, you have to be careful to say that the change of phenotypic mean "after selection but before reproduction" (or some other weasel wording) is the covariance of phenotype and **relative fitness** (not **fitness** itself).

$$\text{cov}(w_r, z) = E(w_r z) - E(z) \tag{7}$$

$$\beta_4 = \Sigma^{-1} \text{cov}(w, z) \tag{4a}$$

Notice that this important quantity appears in β_4 if we redefine fitness to be relative fitness.

The fact that $\beta_1 = \beta_4$ under very weak conditions (mere existence of moments) connects selection with least squares regression, because β_1 is the "true unknown vector of regression coefficients" in the linear regression of w on z (by definition).

Lande and Arnold then make the curious mistake of essentially saying correlation is causation (although they are careful to not quite say this).

Lande and Arnold do say

> $[\beta_1 = \beta_4]$ is a set of partial regression coefficients of relative fitness on the characters (Kendall and Stuart, 1973, eq. 27.42). Under quite general conditions, the method of least squares indicates that the element β_i gives the slope of the straight line that best describes the dependence of relative fitness on character z_i, after removing the residual effects of other characters on fitness (Kendall and Stuart, 1973, Ch. 27.15). There is no need to assume that the actual regressions of fitness on the characters are linear, or that the characters have a multivariate normal distribution. For this reason, the partial regression coefficients β provide a general solution to the problem of measuring the forces of directional selection acting directly on the characters.

For those who don't know, Kendall and Stuart is a multivolume tome on statistics that has gone through many editions and continues being revised by other authors now that the original ones are dead. It is considered quite authoritative by biologists, although curiously, I have never heard or read a Ph. D. statistician mentioning it.

And the 1973 edition of Kendall and Stuart does contain some language trying to justify the use of linear regression in situations that we now would consider inappropriate. Current editions no longer contain this language.

Before the smoothing revolution and the robustness revolution and the generalized linear models revolution and the regularization revolution and the bootstrap revolution, ordinary least squares was the only game in town and was used in many situations we would now consider very inappropriate (although we still teach it as a very important tool and still overuse it). And academics do apply their cleverness to trying to justify what they are doing (even when it is unjustifiable).

Let that be a lesson. It can happen to any of us. Defending the indefensible makes anyone — even those with the highest IQ, *especially* those with the highest IQ — stupid. All that IQ goes into making even more clever and convoluted arguments proving black is white.

To be fair to Lande and Arnold, (1) they didn't exactly say correlation is causation, (2) they were quoting statisticians, and (3) there is an important issue here to deal with.

When several traits are highly correlated, they tend to all change together. If selection is actually acting only on one of them, all will change. How can this be disentangled? Actually, it can't (correlation is not causation). But β does do the best job that can be done.

When $\beta_1 = \beta_3 = \beta_4$ (these equalities require z be multivariate normal), we get another interpretation of β. It is the average gradient of the fitness landscape.

Fisher (1918) invented quantitative genetics. Before this paper Darwinism and Mendelism were thought to be incompatible. Fisher showed they in fact work well together and so started "the modern synthesis" a. k. a. neo-Darwinism.

Quantitative genetics writes

$$z = \mu + x + e$$

where μ is an unknown parameter vector, x is the vector of
additive genetic effects, e is everything else (environmental,
non-additive genetic, and gene-environment interaction effects),
and x and e are assumed to be independent with

$$x \sim \text{Normal}(0, G)$$
$$e \sim \text{Normal}(0, E)$$

where G is the "G matrix" (additive genetic variance matrix) and E
is another variance matrix.

The additive genetic effect vector x cannot, of course be "genes" because (1) genes do not necessarily act additively — this is only part of the genetic effects — and (2) genes (or, to be more precise, mutations or alleles) are discrete but x is continuous. Sometimes x is referred to as a vector of *polygenes*, emphasizing that it models the cumulative effect of many genes.

From the theory of the multivariate normal distribution (found in textbooks on multivariate analysis)

$$E(x \mid z) = G\Sigma^{-1}(z - \mu) \qquad (*)$$

$$\operatorname{var}(x \mid z) = G - G\Sigma^{-1}G \qquad (**)$$

We also need a **very strong assumption**: x and w are conditionally independent given z. Equivalently, genotypic characters x influence fitness only through the values of the observed phenotypic characters z (not in any other way).

Then the difference of genotypic values before selection and after selection but before reproduction is

$$\begin{aligned}
E(wx) - E(x) &= E\{E(wx \mid z)\} \\
&= E\{E(w \mid z)E(x \mid z)\} \\
&= E\{E(w \mid z)G\Sigma^{-1}(z - \mu)\} \\
&= E\{E(wG\Sigma^{-1}(z - \mu) \mid z)\} \\
&= E\{wG\Sigma^{-1}(z - \mu)\} \\
&= G\Sigma^{-1}\operatorname{cov}(w, z) \\
&= G\beta_4
\end{aligned}$$

The second equality is the "very strong assumption". We are taking w to be w_r everywhere.

So $\beta_1 = \beta_4$ appears here too.

But what about the argument that β disentangles the effects of selection on the traits (to the extent that is does)? Shouldn't we apply that here too?

Yes. We should multiply by the analog of Σ^{-1} for genotypic effects, which is G^{-1}. But that is

$$G^{-1}G\beta_4 = \beta_4$$

so (assuming the "very strong assumption") β says the same thing about genotypes (at least additive genetic effects) as it does about phenotypes!

In fact this "very strong assumption" is very questionable in the real world (Rausher, *Evolution*, 1992).

This needs more research.

Now we follow Lande and Arnold in playing the same game with changes of variance instead of changes of means. Everywhere we are just going to write w instead of w_r. Define

$$\zeta = E(wz)$$

(mean "after selection but before reproduction"). Then

$$E\{w(z - \zeta)(z - \zeta)^T\}$$

is variance "after selection but before reproduction". Also define

$$s = \mathrm{cov}(w, z)$$

(Greek letters are parameters and so are s, G, and E).

$$\gamma_4 = \Sigma^{-1} \operatorname{cov}(w, zz^T) \Sigma^{-1} \qquad (4b)$$

$$
\begin{aligned}
E\{w(z-\zeta)(z-\zeta)^T\} &= E\{w(z-\mu-s)(z-\mu-s)^T\} \\
&= E\{w(z-\mu)(z-\mu)^T\} \\
&\quad - 2E\{w(z-\mu)\}s^T + ss^T \\
&= E\{w(z-\mu)(z-\mu)^T\} - ss^T
\end{aligned}
$$

and the change in variance is

$$
\begin{aligned}
E\{w(z-\zeta)(z-\zeta)^T\} - \operatorname{var}(z) &= E\{w(z-\mu)(z-\mu)^T\} - ss^T - \Sigma \\
&= \operatorname{cov}\{w, (z-\mu)(z-\mu)^T\} - ss^T
\end{aligned}
$$

To repeat, change in variance (due to selection)

$$\mathrm{cov}\{w, (z - \mu)(z - \mu)^T\} - ss^T$$

and

$$\gamma_4 = \Sigma^{-1} \mathrm{cov}(w, zz^T)\Sigma^{-1}$$

share the term $\mathrm{cov}(w, zz^T)$ when we assume $\mu = 0$ (or have arranged that by centering the phenotypic characters z).

But they aren't the same, so it is unclear what the relevance of this calculation is.

Quantitative Genetics (cont.)

Lande and Arnold also play the same game with change in variance of additive genetic effects under selection and obtain

$$G\Sigma^{-1}\big[\text{cov}\{w, (z - \mu)(z - \mu)^T\} - ss^T\big]\Sigma^{-1}G$$

and this equals

$$G\gamma_4 G - (G\Sigma^{-1}s)(G\Sigma^{-1}s)^T$$

when we assume (or arrange that) $\mu = 0$.

School of Statistics Technical Report 670 *Commentary on Lande-Arnold Analysis* (on the aster models web site) gives details.

But, again, it is unclear what the relevance of this calculation is.

It also depends on the "very strong assumption" that is very questionable.

On Not Centering Phenotypic Variables

$$Q(\alpha, \beta, \gamma) = E\{(w - \alpha - z^T\beta - \tfrac{1}{2}z^T\gamma z)^2\}, \qquad (1)$$

What if we don't assume (or arrange that) $\mu = 0$?

Keep $E(z) = 0$ but also define $y = z + \mu$.

Now define the analog of $Q(\alpha, \beta, 0)$ for y

$$\begin{aligned}
Q_1^*(\alpha^*, \beta^*) &= E\{(w - \alpha^* - (z + \mu)^T\beta^*\} \\
&= Q(\alpha^* + \mu^T\beta^*, \beta^*, 0)
\end{aligned}$$

The minimizers satisfy

$$\begin{aligned}
\alpha_1 &= \alpha_1^* + \mu^T\beta_1^* \\
\beta_1 &= \beta_1^*
\end{aligned}$$

In summary: not centering predictors does not affect β_1. (It does affect α_1 but this parameter has no biological importance.)

A similar argument (for once we spare you the details, which are in TR 670) says that not centering predictors does not affect γ_2 but does affect the lower-order parameters α_2 and β_2. In particular,

$$\beta_2^* = \beta_2 - \gamma_2 \mu$$

The notions of average gradient and average hessian are invariant. The fitness landscape for y is

$$g^*(y) = g(y - \mu)$$

and the chain rule gives

$$\nabla g^*(y) = \nabla g(y - \mu)$$

and

$$E\{\nabla g^*(y)\} = E\{\nabla g(y - \mu)\} = E\{\nabla g(z)\}$$

so nothing changes. And similarly for γ_3.

Still assuming $E(z) = 0$ and $y = z + \mu$, define

$$\beta_4^* = \Sigma^{-1} \operatorname{cov}(w, y)$$
$$\gamma_4^* = \Sigma^{-1} \operatorname{cov}(w, yy^T)\Sigma^{-1}$$

Then $\beta_4 = \beta_4^*$ because centering does not affect covariance.

But

$$\begin{aligned}
\operatorname{cov}(w, yy^T) &= \operatorname{cov}\{w, (z + \mu)(z + \mu)^T\} \\
&= \operatorname{cov}(w, zz^T) + 2\operatorname{cov}\{w, z\mu^T\} \\
&= \operatorname{cov}(w, zz^T) + 2\operatorname{cov}(w, z)\mu^T
\end{aligned}$$

so

$$\gamma_4^* = \gamma_4 + 2\beta_4\mu^T\Sigma^{-1}$$

Summary: not centering phenotypic variables affects β_2 and γ_4 (only).

TR 670 notes this fact about β_2 but not about γ_4. It should have noted this about γ_4.

If one uses γ_4 (which hardly anyone does), then one should center the predictors. If one uses β_2, then one should center the predictors. For the other betas and gammas, it doesn't matter.

Similarly, if one wants the interpretation of β as change of mean under selection and of γ as some part of change of variance under selection (but not all of it, so why would one want that?), one must use relative fitness w_r in place of actual fitness w.

The following calculations come from TR 670. There they are done for the multivariate case, but to keep things simple, we will only do the univariate case.

Assume z is exactly mean zero normal with nonzero variance σ^2.

For mathematical convenience, assume that the relative fitness landscape looks exactly like a normal distribution

$$g(z) = ce^{-(z-\eta)^2/2\lambda}$$

where c is a constant that makes $g(z)$ have expectation one.

Let f be the PDF of z so

$$f(z)g(z) = \frac{1}{\sqrt{2\pi}\sigma}e^{-z^2/2\sigma^2}ce^{-(z-\eta)^2/2\lambda}$$
$$= c^*e^{-(z-\zeta)^2/2\tau^2}$$

where c^* is another constant chosen to make this integrate to one (that is, $c^* = 1/\sqrt{2\pi}\tau = c/\sqrt{2\pi}\sigma$) and

$$\frac{1}{\tau^2} = \frac{1}{\sigma^2} + \frac{1}{\lambda}$$
$$\frac{\zeta}{\tau^2} = \frac{\eta}{\lambda}$$

In order for this to make sense we do not need $\lambda > 0$, only $\tau^2 > 0$.

$$\frac{1}{\tau^2} = \frac{1}{\sigma^2} + \frac{1}{\lambda}$$

$$\frac{\zeta}{\tau^2} = \frac{\eta}{\lambda}$$

So

$$\tau^2 = \frac{\lambda\sigma^2}{\lambda + \sigma^2}$$

and

$$\zeta = \frac{\eta\tau^2}{\lambda} = \frac{\eta\sigma^2}{\lambda + \sigma^2}$$

Now we calculate β_3 and γ_3 (and all of the betas and gammas are the same by the Lande-Arnold theorem)

$$\nabla g(z) = -g(z)(z - \eta)/\lambda$$
$$\nabla^2 g(z) = g(z)(z - \eta)^2/\lambda^2 - g(z)/\lambda$$

Hence

$$\beta = -(\zeta - \eta)/\lambda$$
$$\gamma = \tau^2/\lambda^2 + (\zeta - \eta)^2/\lambda^2 - 1/\lambda$$
$$= \frac{\tau^2 - \lambda}{\lambda^2} + \beta^2$$

Complicated!

$$\zeta = \frac{\eta\sigma^2}{\lambda + \sigma^2}$$

$$\begin{aligned}
\beta &= -\frac{\zeta - \eta}{\lambda} \\
&= \frac{\eta - \frac{\eta\sigma^2}{\lambda+\sigma^2}}{\lambda} \\
&= \frac{\eta(\lambda + \sigma^2) - \eta\sigma^2}{\lambda(\lambda + \sigma^2)} \\
&= \frac{\eta}{\lambda + \sigma^2}
\end{aligned}$$

$$\tau^2 = \frac{\lambda \sigma^2}{\lambda + \sigma^2}$$

$$
\begin{aligned}
\gamma &= \frac{\tau^2 - \lambda}{\lambda^2} + \beta^2 \\
&= \frac{\frac{\lambda \sigma^2}{\lambda + \sigma^2} - \lambda}{\lambda^2} + \beta^2 \\
&= \frac{\lambda \sigma^2 - \lambda^2 - \lambda \sigma^2}{\lambda^2 (\lambda + \sigma^2)} + \beta^2 \\
&= -\frac{1}{\lambda + \sigma^2} + \beta^2
\end{aligned}
$$

There is a large literature, reviewed in Kingsolver, et al. (*American Naturalist*, 2001), that interprets the signs of components of β as indicating the direction of selection and interprets the signs of diagonal components of γ as indicating stabilizing selection (negative curvature, so a peak, and selection moves toward the peak) or disruptive selection (positive curvature, so a trough, and selection moves away from the trough).

Mitchell-Olds and Shaw (*Evolution*, 1987) pointed out that extrapolating beyond the range of the data is a bad idea so when the peak or the trough of the best quadratic approximation occur outside the range of the observed data an interpretation of stabilizing or disruptive selection is unwarranted (or at best very very weakly supported).

$$\beta = \frac{\eta}{\lambda + \sigma^2}$$

$$\gamma = -\frac{1}{\lambda + \sigma^2} + \beta^2$$

In the toy model we are looking at here η is the peak of the fitness landscape if $\lambda > 0$ or the trough if $\lambda < 0$.

We see that β does have the same sign as η (because of the requirement that $\lambda + \sigma^2 > 0$ regardless of the sign of λ).

But, depending on the relative sizes of η and λ, the sign of γ need not be the same as the sign of λ.

$$\beta = \frac{\eta}{\lambda + \sigma^2}$$

$$\gamma = -\frac{1}{\lambda + \sigma^2} + \beta^2$$

Hence the sign of γ does not always reflect whether the fitness landscape has a peak or a trough *even if the peak or trough of the BQA is in the range of the data*.

Clearly $\gamma = 0$ when

$$\frac{1}{\lambda + \sigma^2} = \beta^2 = \frac{\eta^2}{(\lambda + \sigma^2)^2}$$

or

$$\lambda + \sigma^2 = \eta^2$$

Here are a few example plots taken from TR 670.

```
> v1 <- 1 # sigma squared
> mu2 <- 0 # eta
> v2 <- 2 # lambda

> beta <- mu2 / (v1 + v2)
> gamma <- beta^2 - 1 / (v1 + v2)
> alpha <- 1 - gamma * v1 / 2
> v3 <- 1 / (1 / v1 + 1 / v2)
> c2 <- sqrt(v1 / v3) *
+      exp(mu2^2 * (v2 - v3) / (2 * v2^2))
```

Then the following code makes the figure on the next slide

```
> zlim <- 3
> foo <- function(z) alpha + beta * z + gamma * z^2 / 2
> bar <- function(z) c2 * exp(- (z - mu2)^2 / (2 * v2))
> zz <- seq(-zlim, zlim, 0.01)
> ylim <- c(min(foo(zz), bar(zz)), max(foo(zz), bar(zz)))
> par(mar = c(5, 4, 0, 0) + 0.1)
> curve(foo, col = "magenta", ylab = "relative fitness",
+       xlab = "z", from = -3, to = 3, ylim = ylim, lwd = 2)
> curve(bar, col = "green3", add = TRUE, lwd = 2)
```
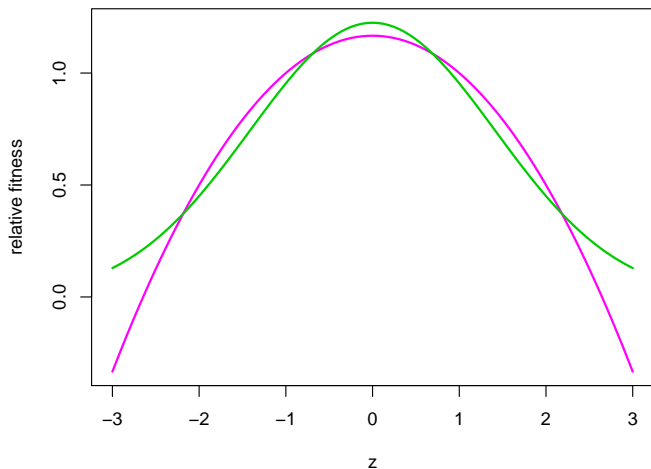
Figure : Fitness landscape (green) and its best quadratic approximation (magenta). $\sigma^2 = 1$, $\lambda = 2$, $\eta = 0$
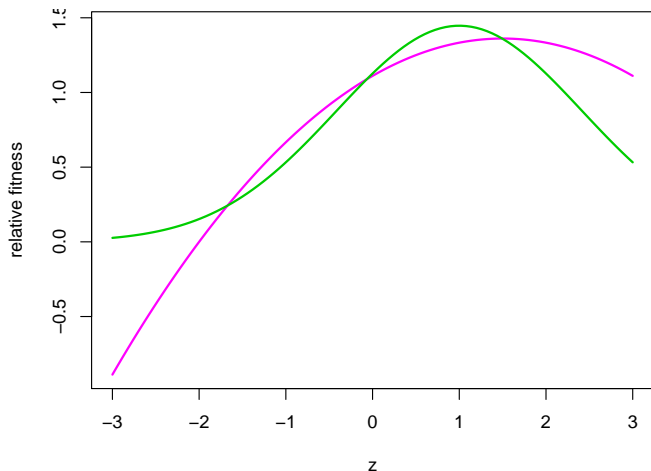
BQA versus Fitness Landscape (cont.)

Figure : Fitness landscape (green) and its best quadratic approximation (magenta). $\sigma^2 = 1$, $\lambda = 2$, $\eta = 1$
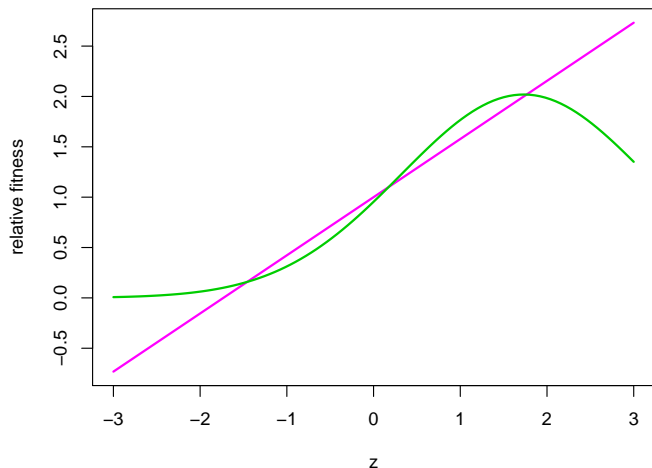
Figure : Fitness landscape (green) and its best quadratic approximation (magenta). $\sigma^2 = 1$, $\lambda = 2$, $\eta = 1.732$ ($\sqrt{\lambda + \sigma^2}$)
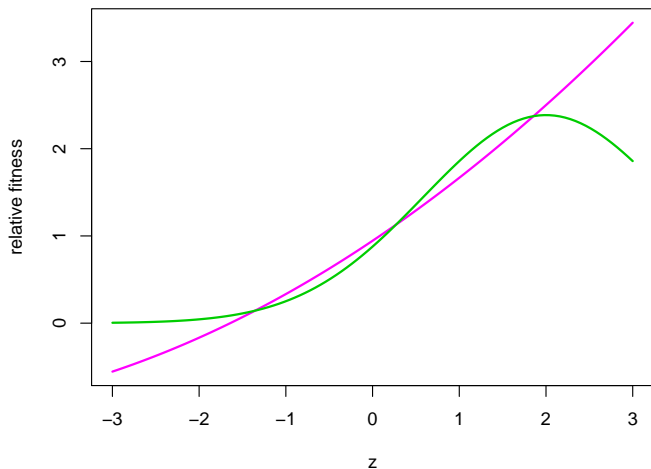
BQA versus Fitness Landscape (cont.)

Figure : Fitness landscape (green) and its best quadratic approximation (magenta). $\sigma^2 = 1$, $\lambda = 2$, $\eta = 2$

Summary of Lande-Arnold Theory

β, especially $\beta_1 = \beta_4$ has a strong biological interpretation. We call that "Lande-Arnold beta".

The connection of γ (any of the gammas) to biology is much weaker.

The connection of the sign of components of γ to whether the fitness landscape has a peak or not is also weak.

The sign of components of γ can be misleading not only when the stationary point is outside the range of the data but when it is quite a bit inside the range of the data — within 1.75 standard deviations from the mean in our example, but this depends on the exact shape of the fitness landscape and could be even worse for other examples.

OLS Estimation

So far we have no complaints about the applied probability in Lande and Arnold (1983). All of their theorems about betas and gammas are correct. (They would not use the plural because they work under the assumption that z is mean zero multivariate normal so all of the betas are the same as are all of the gammas.)

Note that there has been no statistical inference so far in this deck. It is all applied probability.

Now we turn to the examples in Lande and Arnold (1983), which do do statistical inference (incorrectly).

They propose to estimate β_1 and the BLA by ordinary least squares (OLS) linear regression of relative fitness w on the phenotypic trait vector z.

They propose to estimate γ_2 and the BQA by OLS quadratic regression of relative fitness w on the phenotypic trait vector z.

Up to a certain point, there is nothing wrong with OLS estimation.

By the Gauss-Markov theorem OLS estimators are best linear unbiased estimators (BLUE) of their expectations (conditional on the predictors). That is, BLUE requires that we treat z and fixed and w as random.

Thus $\hat{\alpha}_1$ and $\hat{\beta}_1$ are BLUE of α_1 and β_1 and, moreover, if

$$g_1(z) = \alpha_1 + \beta_1 z$$
$$\hat{g}_1(z) = \hat{\alpha}_1 + \hat{\beta}_1 z$$

so g_1 is the BLA of the fitness landscape, then $\hat{g}_1(z)$ is the BLUE of $g_1(z)$ for each z, and all of this holds simultaneously.

Similarly, $\hat{\alpha}_2$ and $\hat{\beta}_2$ and $\hat{\gamma}_2$ are BLUE of α_2 and β_2 and γ_2 and the OLS quadratic regression estimator of the fitness landscape is the BLUE of the BQA.

Since division is not a linear operation, OLS estimates are BLUE only applies to α_1, β_1, α_2, β_2, γ_2, the BLA, and the BQA if we use actual fitness rather than relative fitness (it is the operation of dividing actual fitness by mean fitness to get relative finitness that isn't linear so we can't have the L in BLUE and the Gauss-Markov theorem does not apply).

But that is as far as it goes.

The t and F statistics printed out by OLS regression software require that the conditional distribution of the response given the predictors be homoscedastic mean zero normal.

Fitness is far from normal, much less homoscedastic normal.

- never negative
- large atom at zero
- multimodal if multiple breeding seasons
- standard deviation varies with conditional expectation of response (rather than constant)

Thus using the t and F statistics for statistical inference is bogus. But that is what Lande and Arnold (1983) did, and that is what most of the literature following has done.

Mitchell-Olds and Shaw (1987) pointed out the bogosity, but the practice continued.

We, of course, recommend using aster models for life history analysis in general and estimation of fitness in particular.

But, before we can do that, there is one more issue that remains.

Aster models work on the canonical parameter scale. If we use a quadratic function of z on the canonical parameter scale, what happens on the mean value parameter scale? Does that make sense?

Yes. Because of the multivariate monotonicity property.

For this argument we follow the appendix of Shaw and Geyer (2010).

Multivariate Monotonicity

Suppose we have two kinds of predictors: the vector z of phenotypic traits and another vector x. And suppose we model the unconditional canonical parameter as follows

$$\varphi_j(x, z) = \begin{cases} a_j(x) + q(z), & j \in F \\ a_j(x), & j \notin F \end{cases}$$

where q is a linear function of z or a quadratic function of z and F is the set of nodes whose sum is deemed observed fitness.

Now consider the difference between two individuals that differ in z values but not x values.

$$\varphi_j(x, z) - \varphi_j(x, z') = \begin{cases} q(z) - q(z'), & j \in F \\ 0, & j \notin F \end{cases}$$

Multivariate Monotonicity (cont.)

Now the multivariate monotonicity property says

$$\sum_{j \in J} \left[\varphi_j(x, z) - \varphi_j(x, z') \right] \left[\mu_j(x, z) - \mu_j(x, z') \right] > 0$$

where, as usual, the mus are unconditional mean value parameters. But using what was derived on the preceding slide, this simplifies to

$$\left[q(z) - q(z') \right] \sum_{j \in F} \left[\mu_j(x, z) - \mu_j(x, z') \right] > 0$$

hence we have

$$\sum_{j \in F} \mu_j(x, z) < \sum_{j \in F} \mu_j(x, z') \qquad \text{if and only if} \qquad q(z) < q(z')$$

(fitness on the canonical parameter scale is a monotone function of fitness on the mean value parameter scale and vice versa).

Because individuals are independent, this result applies whether F is the set of fitness nodes for just one individual or for all individuals, so long as only predictor values for an individual enter the regression function for that individual.

This means a contour plot for the fitness landscape would have the same contours on either the canonical or the mean value parameter scale (the numbers attached to the contours would differ, but the same lines would be contours of both).

Since gradients are perpendicular to contours, the gradient vectors of both functions evaluated at the same z point in the same direction (the lengths differ but not the directions).

We use simulated data from Shaw and Geyer (2010). The reason for the simulated data was to illustrate many points that could not be illustrated with real data that existed at the time.
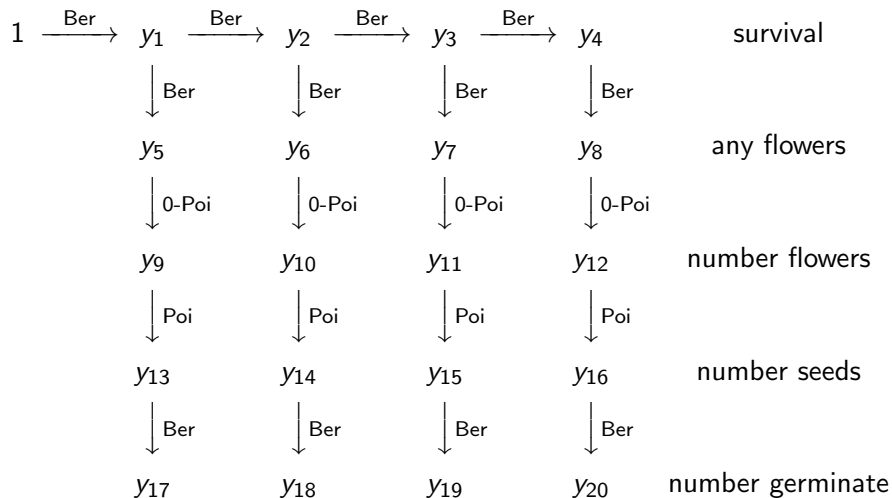
Data (cont.)

Stanton-Geddes, et al. (*PLoS One*, 2012) have data that is near to this design. In particular, their graph goes to germinated seeds.

But the graph doesn't cover multiple years because the organism is an annual plant (*Chamaecrista fasciculata*).

This graph has interesting features not present in other examples.

- Not all predecessors are Bernoulli.
- Not all data measured on one individual (germinated seeds are from the individual but not that individual).

Data (cont.)

The data are in the data set sim in the aster package.

```
> library(aster)
> data(sim)
> ls()

[1] "beta.true" "fam"       "ladata"    "mu.true"
[5] "phi.true"  "pred"      "redata"    "theta.true"
[9] "vars"

> vars

 [1] "isurv1" "isurv2" "isurv3" "isurv4" "iflow1"
 [6] "iflow2" "iflow3" "iflow4" "nflow1" "nflow2"
[11] "nflow3" "nflow4" "nseed1" "nseed2" "nseed3"
[16] "nseed4" "ngerm1" "ngerm2" "ngerm3" "ngerm4"

> names(redata)

[1] "z1"   "z2"   "varb" "resp" "id"   "root"
```

Model Fitting

Before we fit models we need to explain why we do not need a predictor named `fit`.

```
> fit <- grepl("ngerm", as.character(redata$varb))
> identical(fit, redata$z1 != 0)

[1] TRUE

> identical(fit, redata$z2 != 0)

[1] TRUE
```

The z1 in `redata` is already `fit` : z1. Similarly for z2.

```
> out2 <- aster(resp ~ varb + 0 + z1 + z2 + I(z1^2)
+      + I(z1*z2) + I(z2^2),
+      pred, fam, varb, id, root, data = redata)
> summary(out2)

Call:
aster.formula(formula = resp ~ varb + 0 + z1 + z2 + I(z1^2)
    I(z1 * z2) + I(z2^2), pred = pred, fam = fam, varvar =
    idvar = id, root = root, data = redata)

            Estimate Std. Error z value Pr(>|z|)
varbiflow1 -3.444251   0.180123 -19.122  < 2e-16 ***
varbiflow2 -3.064152   0.203311 -15.071  < 2e-16 ***
varbiflow3 -3.207467   0.218952 -14.649  < 2e-16 ***
varbiflow4 -3.284180   0.236597 -13.881  < 2e-16 ***
varbisurv1 -0.065167   0.160348  -0.406  0.68444
varbisurv2 -0.700847   0.225747  -3.105  0.00191 **
varbisurv3 -0.094013   0.275111  -0.342  0.73256
```

Model Fitting (cont.)

```
> out1 <- aster(resp ~ varb + 0 + z1 + z2,
+     pred, fam, varb, id, root, data = redata)
> out0 <- aster(resp ~ varb + 0,
+     pred, fam, varb, id, root, data = redata)
> anova(out0, out1, out2)

Analysis of Deviance Table

Model 1: resp ~ varb + 0
Model 2: resp ~ varb + 0 + z1 + z2
Model 3: resp ~ varb + 0 + z1 + z2 + I(z1^2) + I(z1 * z2) +
  Model Df Model Dev Df Deviance P(>|Chi|)
1     20     84703
2     22     84959  2  255.900 < 2.2e-16 ***
3     25     84975  3   16.359  0.000957 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Fitting (cont.)

```
> sout2 <- summary(out2)
> coef2 <- sout2$coefficients
> outies <- grepl("varb", rownames(coef2))
> printCoefmat(coef2[! outies, ])

            Estimate Std. Error z value  Pr(>|z|)
z1         0.1469497  0.0136949 10.7302 < 2.2e-16 ***
z2        -0.0205982  0.0098424 -2.0928  0.036366 *
I(z1^2)   -0.0278071  0.0095082 -2.9245  0.003450 **
I(z1 * z2) 0.0237128  0.0123521  1.9197  0.054891 .
I(z2^2)   -0.0179859  0.0065363 -2.7517  0.005929 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Normally, one doesn't need the stuff on the previous page. That was only because `summary(sout2)` ran off the slide.

IMHO, it doesn't make sense to look at P-values (or stars) for individual components. β is a vector and γ is a matrix. The `anova` command we did tests whether γ is statistically significantly different from zero (collectively) and similarly for β.

If we did want to do some "stargazing" we should Bonferroni correct for five tests (multiply all the P-values by 5). Then we see that the second component of β is not statistically significantly different from zero and so is the off-diagonal component of γ. But they are both marginally statistically significant when not Bonferroni corrected. So why did we want to open that can of worms in the first place?

There is a paper (Blows and Brooks, *American Naturalist*, 2003) whose main point (?) is that one should not leave out "interaction terms" like I(z1 * z2). Apparently many researchers had made this mistake.

That "mistake" never even occurred to us because we don't think of such terms as "interactions" but as the off-diagonal terms of γ. We are trying to fit a general quadratic function, and a general quadratic function has such terms. If we left them out, we wouldn't be doing (the aster competitor of) Lande-Arnold analysis.

The cure for this kind of "mistake" is to wash your mouth out with soap every time you say the word "interaction" (in a statistical context).

So what is this general quadratic function?

```
> coef2 <- out2$coefficients
> b <- coef2[c("z1", "z2")]
> C <- matrix(NA, 2, 2)
> C[1, 1] <- coef2["I(z1^2)"]
> C[1, 2] <- C[2, 1] <- coef2["I(z1 * z2)"] / 2
> C[2, 2] <- coef2["I(z2^2)"]
```

The divide by 2 is because in

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

$$= \sum_{i=1}^{n} a_{ii} x_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij} x_i x_j$$

each off-diagonal term appears in the sum multiplied by 2 when we don't include each term twice ($a_{ij} x_i x_j$ and $a_{ji} x_i x_j$).

So where does the maximum of

$$f(x) = b^T x + x^T C x$$

occur? Where the first derivative

$$\nabla f(x) = b + 2Cx$$

is zero.

```
> m <- solve(- 2 * C, b)
> m

[1] 3.335738 1.626314
```

Now we want to make a plot of the fitness landscape. We start with a scatterplot of z1 and z2 so we can see what range of values we want.

The following code makes the figure on the next slide

```
> z1 <- redata$z1[as.character(redata$varb) == "ngerm1"]
> z2 <- redata$z2[as.character(redata$varb) == "ngerm1"]
> xlim <- range(z1, m[1])
> plot(z1, z2, xlim = xlim)
> points(m[1], m[2], col = "red", pch = 19)
> ufoo <- par("usr")
```
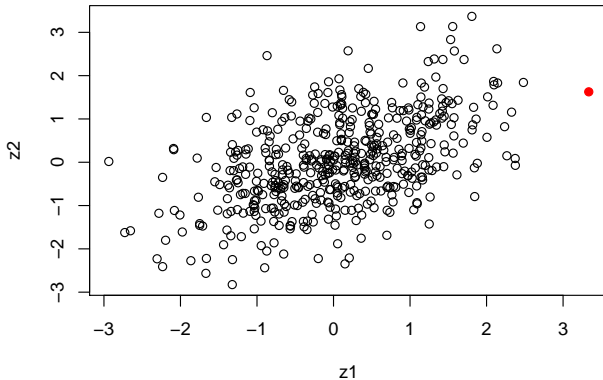
Figure : Scatterplot of observed phenotypic trait values. Solid red dot is maximum of fitness landscape.

Now we start following Section 4.1 of TR 669.

First make a grid of points (the contour function needs values on a grid). We go from one edge of the plot to the other, the coordinates of the edges being in ufoo which came from par("usr")

```
> nx <- 101
> ny <- 101
> xfoo <- seq(ufoo[1], ufoo[2], length = nx)
> yfoo <- seq(ufoo[3], ufoo[4], length = ny)
```

Then we make a data frame like `redata` with these predictor values

```
> xx <- outer(xfoo, yfoo^0)
> yy <- outer(xfoo^0, yfoo)
> xx <- as.vector(xx)
> yy <- as.vector(yy)
> nn <- length(xx)
> foo <- rep(1, nn)
> bar <- list(z1 = xx, z2 = yy, root = foo)
> for (lab in levels(redata$varb)) {
+     bar[[lab]] <- foo
+     ##### response doesn't matter for prediction #####
+ }
```

```
> bar <- as.data.frame(bar)
> rebar <- reshape(bar, varying = list(vars),
+     direction = "long", timevar = "varb",
+     times = as.factor(vars), v.names = "resp")
```

We also have to zero out some non-bottom-layer elements of z1
and z2, just like with the "real" (simulated) data

```
> barlayer <- substr(as.character(rebar$varb), 1, 5)
> rebar$z1[barlayer != "ngerm"] <- 0
> rebar$z2[barlayer != "ngerm"] <- 0
```

Then we predict at the new points.

```
> pbar <- predict(out2, newdata = rebar, varvar = varb,
+      idvar = id, root = root)
> pbar <- matrix(pbar, nrow = nrow(bar))
> pbar <- pbar[ , grep("germ", vars)]
> zz <- apply(pbar, 1, sum)
> zz <- matrix(zz, nx, ny)
```

Then the following code makes the figure on the next slide

```
> plot(z1, z2, pch = 20, xlim = xlim)
> contour(xfoo, yfoo, zz, add = TRUE, col = "blue",
+      labcex = 1, lwd = 2)
> points(m[1], m[2], col = "red", pch = 19)
```

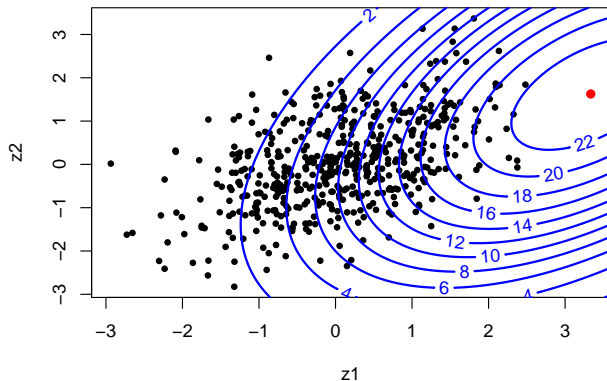Figure : Scatterplot of z1 versus z2 with contours of the fitness landscape as estimated by the aster model (blue). Solid red dot is location of maximum.

Following Mitchell-Olds and Shaw (1987) we should realize that the maximum of the (estimated, not true) fitness landscape is outside the range of the data and say that selection is mostly directional and that the evidence for stabilizing selection is weak (despite statistically significant regression coefficients for predictor components quadratic in z).

Still, the evidence, such as it is, weakly favors stabilizing selection.

There is no evidence for disruptive selection (more on this later).

Also recall that the the sign of the second component of $\hat{\beta}$ was negative, but it looks like selection should push the population up (not down) and to the right in the figure.

This is because we did not center the phenotypic variables so $\beta_1 \neq \beta_2$. If we want to interpret the signs of the components of β we have to do something different (more on this later).

There is, of course, no reason to stop at functions quadratic on the canonical parameter scale when fitting fitness landscapes.

Because quadratic on the fitness landscape maps into possible mean values, it is less objectionable than quadratic OLS, but it is still true that the fitness landscape can be more complicated than that.

If we add cubic terms, that would be 6 more parameters. We could also try regression splines (with fixed knot positions) or smoothing splines (with many more knot positions and regularization). Neither spline method is built into the aster package. So either would be a lot of hand work. But they could be done.

Here, because we know the simulation truth was quadratic on the canonical parameter scale, we don't bother.

There is a lot about model selection in Shaw and Geyer (2010). They use branch and bound to fit all subsets and AIC, AICc, or BIC to select models.

But this does not scale to more than about 10 phenotypic predictor variables.

The "bandwagon of the aughts" in statistics (machine learning, small n large p, model selection, "big data", yadda, yadda, yadda) is all about doing this in ways that scale.

So is "dimension reduction" (a misnomer because it isn't all of dimension reduction) intensely studied by Dennis Cook and his students and co-authors.

Either of these could be applied to aster models, but haven't so far.

The dataset `sim` in the `aster` package also contains data for Lande-Arnold analysis, the data frame `ladata`.

This is the same data as `redata` but `ladata` has only one component of the response vector which is the same as observed fitness in `redata`.

We start with estimating the BQA (best quadratic approximation).

```
> lout2 <- lm(y ~ z1 + z2 + I(z1^2) + I(z1*z2) + I(z2^2),
+     data = ladata)
> summary(lout2)

Call:
lm(formula = y ~ z1 + z2 + I(z1^2) + I(z1 * z2) + I(z2^2),

Residuals:
    Min      1Q  Median      3Q     Max
-17.6321 -4.5203 -0.9696  3.7392 25.8501

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.6740     0.4449  17.250  < 2e-16 ***
z1            5.6369     0.3589  15.704  < 2e-16 ***
z2           -0.7077     0.3472  -2.038  0.04207 *
I(z1^2)       0.8166     0.3046   2.681  0.00758 **
I(z1 * z2)    0.3549     0.4389   0.809  0.41909
```

Lande-Arnold Analysis: BQA (cont.)

```
> summary(lout2)

Call:
lm(formula = y ~ z1 + z2 + I(z1^2) + I(z1 * z2) + I(z2^2), data = ladata)

Residuals:
     Min      1Q  Median      3Q     Max
-17.6321  -4.5203  -0.9696   3.7392  25.8501

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.6740     0.4449  17.250  < 2e-16 ***
z1            5.6369     0.3589  15.704  < 2e-16 ***
z2           -0.7077     0.3472  -2.038  0.04207 *
I(z1^2)       0.8166     0.3046   2.681  0.00758 **
I(z1 * z2)    0.3549     0.4389   0.809  0.41909
I(z2^2)      -0.7045     0.2739  -2.572  0.01040 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.912 on 494 degrees of freedom
Multiple R-squared:  0.3855,	Adjusted R-squared:  0.3793
F-statistic: 61.98 on 5 and 494 DF,  p-value: < 2.2e-16
```

Find the point where the BQA achieves its maximum.

```
> coef2 <- lout2$coefficients
> b <- coef2[c("z1", "z2")]
> C <- matrix(NA, 2, 2)
> C[1, 1] <- coef2["I(z1^2)"]
> C[1, 2] <- C[2, 1] <- coef2["I(z1 * z2)"] / 2
> C[2, 2] <- coef2["I(z2^2)"]
> m <- solve(- 2 * C, b)
> m

[1] -3.16901 -1.30045
```

The following code makes the figure on the next slide

```
> xlim <- range(z1, m[1])
> plot(z1, z2, xlim = xlim)
> points(m[1], m[2], col = "red", pch = 19)
> ufoo <- par("usr")
```
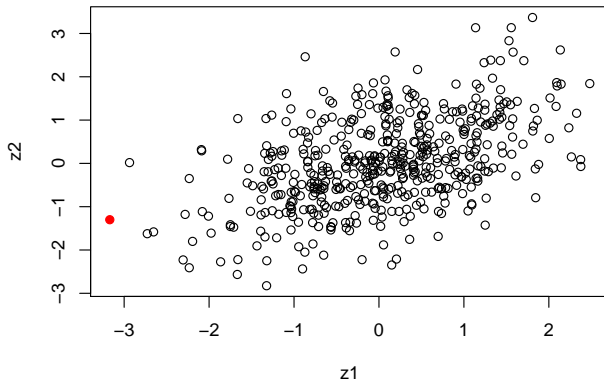
Figure : Scatterplot of observed phenotypic trait values. Solid red dot is stationary point of BQA (best quadratic approximation of fitness landscape).

Oops! It looks like the "maximum" of the BQA isn't a maximum.

A point where the first derivative of a function is zero is called a **stationary point**. It may be a maximum, a minimum, or neither.

In two dimensions it is called a **saddle point** if the curvature is positive in one direction and negative in the other. That appears to be what we have here (from looking at the coefficients in the fit).

Continuing as before (except we don't need to reshape the data)

```
> xfoo <- seq(ufoo[1], ufoo[2], length = nx)
> yfoo <- seq(ufoo[3], ufoo[4], length = ny)
> xx <- outer(xfoo, yfoo^0)
> yy <- outer(xfoo^0, yfoo)
> xx <- as.vector(xx)
> yy <- as.vector(yy)
> nn <- length(xx)
> foo <- rep(1, nn)
> bar <- data.frame(z1 = xx, z2 = yy)
> zz <- predict(lout2, newdata = bar)
> zz <- matrix(zz, nx, ny)
```

Then the following code makes the figure on the next slide

```
> plot(z1, z2, pch = 20, xlim = xlim)
> contour(xfoo, yfoo, zz, add = TRUE, col = "blue",
+     labcex = 1, lwd = 2)
> points(m[1], m[2], col = "red", pch = 19)
```
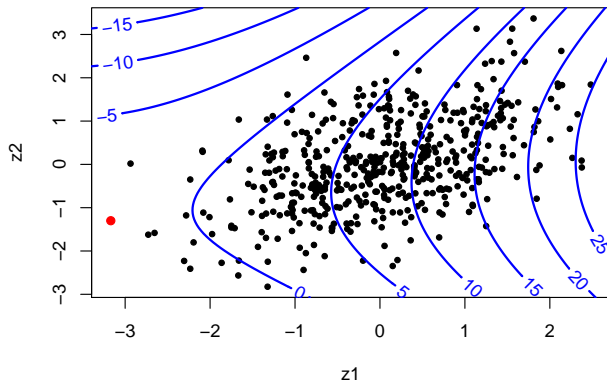
Figure : Scatterplot of z1 versus z2 with contours of the BQA (best quadratic approximation of the fitness landscape) as estimated by OLS (blue). Solid red dot is location of stationary point.

The BQA does a terrible job.

- It goes negative (negative fitness is impossible).
- It has a saddle point rather than a maximum (disagreeing with the simulation truth).

This is a bit unfair in one sense because we deliberately picked an example where the BQA would be bad. As our one-dimensional analysis shows the BQA will always be bad when the maximum of the fitness landscape is at the edge of the data. The BQA will be better when the maximum is near the middle of the data.

Also if we follow Mitchell-Olds and Shaw (1987) and say this is primarily directional selection, then both figures get this much right. It is only the quadratic part of the BQA that is wrong.

Now we tackle Lande-Arnold beta.

First we define relative fitness, then do linear OLS.

```
> w <- ladata$y / mean(ladata$y)
> bout <- lm(w ~ z1 + z2)
```

We do not need to center the predictors because that would only change the (Intercept) coefficient. Not the coefficients for z1 and z2.

Lande-Arnold Analysis: Beta (cont.)

```
> summary(bout)

Call:
lm(formula = w ~ z1 + z2)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8315 -0.5906 -0.1215  0.5026  3.4714

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.96693    0.03850  25.116  <2e-16 ***
z1           0.69969    0.04449  15.729  <2e-16 ***
z2          -0.10355    0.04250  -2.436  0.0152 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Great! We have got a good (almost BLUE, "almost" because of the division by mean fitness) point estimate of beta.

But the standard errors and *P*-values are completely bogus.

How can we do a valid analysis using aster models?

We cannot do anything with the canonical parameter estimates we already have. They are on the wrong scale!

Recall that beta was originally defined at the slope of the BLA defined for *expected values*.

So here is the idea: we apply OLS regression to the mean value parameters (μ) from the aster model. Because of the observed = expected property of exponential families and because both OLS and aster are exponential families. OLS and aster will give exactly the same estimates!

But since the aster model is a defensible statistical model, we get defensible standard errors using the delta method.

In order to have the observed = expected property it is not necessary that the aster model have exactly the same predictors as the OLS model.

It is enough that the aster model have the same predictors ((Intercept), z1, and z2) as the OLS model plus some more (varb, I(z1^2), I(z2^2), and I(z1 * z2)).

The components of the submodel canonical statistic for the shared predictors will match and have the same mean value parameter estimates (observed = expected). The rest don't matter.

First, let us convince ourselves, that we do indeed get exactly the same beta using the aster model.

```
> pout2 <- predict(out2)
> pout2 <- matrix(pout2, nrow = nrow(out2$x),
+     ncol = ncol(out2$x))
> colnames(pout2) <- colnames(out2$x)
> mufit <- pout2[ , grep("germ", colnames(pout2))]
> mufit <- rowSums(mufit)
> wmu <- mufit / mean(mufit)
> wmout <- lm(wmu ~ z1 + z2)
> all.equal(bout$coefficients, wmout$coefficients)

[1] TRUE
```

In order to apply the delta method we need an explicit formula for β as a function of μ.

It is simpler if we use the identity $\beta_1 = \beta_4$ and

$$\beta_4 = \Sigma^{-1} \operatorname{cov}(w, z)$$

```
> z1 <- z1 - mean(z1)
> z2 <- z2 - mean(z2)
> z <- cbind(z1, z2, deparse.level = 0)
> zvarinv <- solve(t(z) %*% z / nrow(z))
> zwcov <- t(z) %*% cbind(mufit) / nrow(z)
> mybeta <- as.numeric(zvarinv %*% zwcov) / mean(mufit)
> mybeta.too <- as.numeric(bout$coefficients[-1])
> all.equal(mybeta, mybeta.too)

[1] TRUE
```

Lande-Arnold Analysis: Beta (cont.)

Now we want to do the same predictions almost entirely with
predict.aster in order to get standard errors.

```
> mu <- predict(out2)
> amat <- matrix(0, nrow = length(mufit), ncol = length(mu)
> blank <- matrix(0, nrow = nrow(pout2), ncol = ncol(pout2)
> blank.idx <- grep("germ", colnames(pout2))
> for (i in 1:nrow(amat)) {
+     boom <- blank
+     boom[i, blank.idx] <- 1
+     amat[i, ] <- as.vector(boom)
+ }
> all.equal(mufit, as.vector(amat %*% mu))

[1] TRUE
```

```
> bmat <- zvarinv %*% t(z) %*% amat / nrow(z)
> all.equal(mybeta, as.numeric(bmat %*% mu) /
+     mean(as.numeric(amat %*% mu)))

[1] TRUE

> cmat <- apply(amat, 2, sum) / nrow(z)
> cmat <- rbind(cmat)
> all.equal(mybeta, as.numeric(bmat %*% mu) /
+     as.numeric(cmat %*% mu))

[1] TRUE
```

```
> dmat <- rbind(bmat, cmat, deparse.level = 0)
> all.equal(mybeta, as.numeric(dmat %*% mu)[1:2] /
+     as.numeric(dmat %*% mu)[3])

[1] TRUE
```

So now we have slowly — building bits up piece by piece — gotten so that we produce directly all of the pieces for making beta with one matrix multiplication by dmat.

Now because of PBD we have to make `dmat` a 3-way array before supplying it to `predict.aster`.

```
> d3way <- array(as.vector(t(dmat)),
+     dim = c(dim(out2$modmat)[1:2], nrow(dmat)))
> dout <- predict(out2, amat = d3way, se.fit = TRUE)
> all.equal(mybeta, dout$fit[1:2] / dout$fit[3])

[1] TRUE
```

We are still not completely out of the woods.

If we denote the R object dout$fit as the vector ζ in mathematical notation, then

$$\beta_i = \zeta_i/\zeta_3, \qquad i = 1, 2.$$

This is a non-linear transformation $\zeta \to \beta$ that has Jacobian matrix

$$\begin{pmatrix} 1/\zeta_3 & 0 & -\zeta_1/\zeta_3^2 \\ 0 & 1/\zeta_3 & -\zeta_2/\zeta_3^2 \end{pmatrix}$$

$$\begin{pmatrix} 1/\zeta_3 & 0 & -\zeta_1/\zeta_3^2 \\ 0 & 1/\zeta_3 & -\zeta_2/\zeta_3^2 \end{pmatrix}$$

This matrix is constructed in R as

```
> zeta1 <- dout$fit[1]
> zeta2 <- dout$fit[2]
> zeta3 <- dout$fit[3]
> jacobian <- rbind(c( 1 / zeta3, 0, - zeta1 / zeta3^2 ),
+                   c( 0, 1 / zeta3, - zeta2 / zeta3^2 ))
```

Because of this nonlinearity, which arises from the fact that we measure actual fitness not relative fitness, hence must estimate relative fitness by dividing actual fitness by mean fitness, OLS would not calculate correct standard errors even if actual fitness were homoscedastic normal.

Also because of this nonlinearity, OLS estimates of Lande-Arnold beta (which agree exactly with aster estimates) are not actually BLUE.

Our next step takes this issue into account correctly.

Lande-Arnold Analysis: Beta (cont.)

The function predict.aster does not give the full variance matrix for its "predictions". It does, however, give a component gradient which is (in this case) $\partial\zeta/\partial\beta$ and can be used to calculate the asymptotic variance matrix for ζ

```
> dvar <- dout$gradient %*% solve(out2$fisher) %*%
+     t(dout$gradient)
> all.equal(dout$se.fit, sqrt(diag(dvar)))

[1] TRUE

> print(dvar)

            [,1]          [,2]          [,3]
[1,]  0.11074988 -0.049378460  0.051190895
[2,] -0.04937846  0.099797054 -0.008332417
[3,]  0.05119090 -0.008332417  0.092359357
```

Lande-Arnold Analysis: Beta (cont.)

Then the delta method is applied to get the asymptotic
variance-covariance matrix for β

```
> bvar <- jacobian %*% dvar %*% t(jacobian)
> print(bvar)

              [,1]            [,2]
[1,]  0.0012646355 -0.0006739173
[2,] -0.0006739173  0.0014855497
```

The diagonal elements give standard errors for beta

```
> foo <- cbind(mybeta, sqrt(diag(bvar)))
> colnames(foo) <- c("beta", "se(beta)")
> print(foo)

            beta   se(beta)
[1,]  0.6996907 0.03556171
[2,] -0.1035472 0.03854283
```

```
> sbout <- summary(bout)
> printCoefmat(sbout$coefficients, signif.stars = FALSE)

            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.966933   0.038499 25.1156 < 2e-16
z1          0.699691   0.044485 15.7286 < 2e-16
z2         -0.103547   0.042501 -2.4363  0.01519

> print(foo)

           beta   se(beta)
[1,]  0.6996907 0.03556171
[2,] -0.1035472 0.03854283
```

Correct standard errors (derived from the aster model) are smaller than the putative (and erroneous) standard errors derived from OLS.

But one example does not mean that OLS produces "conservative" standard errors. Don't say that without a proof (which does not exist).

Anyway, why would one want bogusly overwide confidence intervals?

We can also make an elliptical confidence region that accounts for the correlation of the components of β.

Let B denote the matrix that is `bvar` in R.

Then we know

$$(\hat{\beta} - \beta)^T B^{-1} (\hat{\beta} - \beta) \approx \mathsf{ChiSq}(2)$$

so our $100(1 - \alpha)\%$ confidence region should be the set of β that satisfy

$$(\hat{\beta} - \beta)^T B^{-1} (\hat{\beta} - \beta) < c_\alpha$$

where c_α is the upper α critical value of the relevant chi-square distribution.

$$(\hat{\beta} - \beta)^T B^{-1}(\hat{\beta} - \beta) < c_\alpha$$

So how do we draw this ellipse? An ellipse is a linear transformation of a circle.

The linear transformation we want is represented by the matrix $B^{1/2}$, the symmetric square root of B.

We have

$$\|B^{-1/2}(\hat{\beta} - \beta)\|^2 = c_\alpha$$

when

$$\beta = \hat{\beta} + \sqrt{c_\alpha}B^{1/2}u$$

and u runs around the unit circle.

Lande-Arnold Analysis: Beta (cont.)

```
> fred <- eigen(bvar, symmetric = TRUE)
> sally <- fred$vectors %*% diag(sqrt(fred$values)) %*%
+     t(fred$vectors)
> u1 <- cos(seq(0, 2 * pi, length = 101))
> u2 <- sin(seq(0, 2 * pi, length = 101))
> u <- rbind(u1, u2)
> jane <- sqrt(qchisq(0.95, 2)) * sally %*% u
> jill <- sqrt(qchisq(0.50, 2)) * sally %*% u
```

Now `jane` is a matrix with two columns giving the two coordinates
running around the ellipse. Similarly for `jill`.

Then the following code makes the figure on the next slide

```
> plot(mybeta[1] + jane[1, ], mybeta[2] + jane[2, ],
+     type = "l", xlab = expression(beta[1]),
+     ylab = expression(beta[2]))
> points(mybeta[1], mybeta[2], pch = 20)
> lines(mybeta[1] + jill[1, ], mybeta[2] + jill[2, ],
+     lty = "dashed")
```
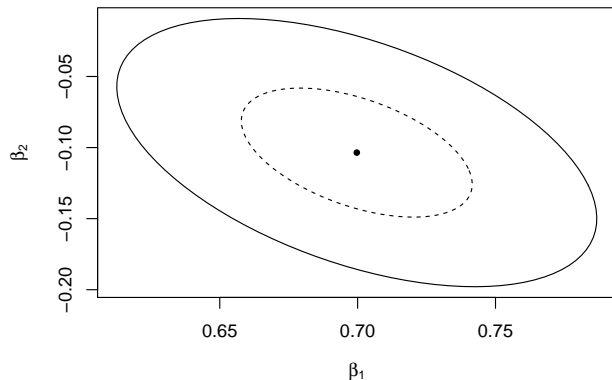
Figure : Confidence ellipse for beta. Solid curve is boundary of 95% confidence region, dashed curve is boundary of 50% confidence region, solid dot is location of MLE for beta.

In the figure the 95% confidence ellipse does not cross either coordinate axis. This says β_1 is statistically significantly greater than zero and β_2 is statistically significantly less than zero (at the 0.05 significance level), even accounting for doing two tests.

This last section of the deck is an overview of how Lande-Arnold analysis and it aster model competition differ and how they agree.

The agreement starts with aster analysis accepting all of the definitions in Lande and Arnold (1983) and the theorem.

The agreement continues with both using exponential family models and with the saturated model canonical statistic y for the Lande Arnold (observed fitness) being a linear function of the saturated model canonical statistic y for the aster model (usually, observed fitness is a sum of some components of the graphical model, also called components of fitness).

The agreement continues with both using linear (in z) or quadratic (in z) submodels.

The agreement continues with the submodel canonical statistic $M_{\mathsf{LA}}^T y_{\mathsf{LA}}$ for Lande-Arnold analysis (either linear or quadratic) being a subvector of the submodel canonical statistic $M_{\mathsf{ast}}^T y_{\mathsf{ast}}$ for aster analysis (either linear or quadratic).

The agreement continues with the observed $=$ expected property of maximum likelihood in exponential families matching these submodel mean value parameters to their observed values (submodel canonical statistics).

The agreement continues with both producing the exact same point estimate of Lande-Arnold beta (actually $\beta_1 = \beta_4$ if multivariate normality of z is not assumed).

The agreement would continue with both producing the exact same point estimate of Lande-Arnold gamma (actually γ_2) if we thought there was any point to estimating γ, which we do not.

And both methods agree that these estimates are BLUE if actual fitness is used and "almost" BLUE if relative fitness is used.

That is the end of the agreement.

The disagreement starts with aster analysis using a defensible statistical model for fitness, whereas Lande-Arnold analysis "assumes" the conditional distribution of fitness given phenotypic traits is homoscedastic normal. ("Assumes" is in scare quotes, because they don't actually state such nonsense, although all their P-values or confidence intervals depend on it.)

P-values and confidence intervals produced by Lande-Arnold analysis are indefensible.

P-values and confidence intervals produced by aster analysis are defensible (the "usual" asymptotics of maximum likelihood plus the delta method). They can be improved by the parametric bootstrap if necessary (more on this in a later slide deck).

The disagreement continues with aster models being quadratic on the unconditional canonical parameter scale (φ is quadratic in z), whereas must model on the unconditional mean value parameter scale (μ is quadratic in z). The latter is because $\varphi = \mu$ for linear models.

Thus we say Lande-Arnold analysis can only estimate the BQA, whereas aster analysis can model the actual fitness landscape.

Of course, the true unknown fitness landscape can be more complicated than quadratic in z on the unconditional canonical parameter scale. But one can add more terms to the aster model formula without losing any of the agreements with Lande-Arnold analysis.

It should be clear from the example that Lande-Arnold analysis only estimates the BQA when (1) its estimate goes negative, which the true fitness landscape cannot and (2) it has a saddle point, when the true fitness landscape (which we know because this is a simulation — the simulation truth parameter values, not mentioned in this slide deck, are in the sim dataset) has a peak.

The disagreement continues with Lande-Arnold analysis using signs of diagonal components of γ or better yet signs of eigenvalues of γ to indicate stabilizing (selection moves toward a peak) or disruptive (selection moves away from a trough), whereas aster analysis does not use γ at all but instead uses the quadratic form on the unconditional canonical parameter scale (whose components are aster model regression coefficients).