Allometry – some measurement-theoretical thoughts
Classic (Huxley) allometry

• Two phenotypic characters, x and y, can typically be described by a power-law:

\[ y = ax^b \]

(For \( b=1 \), x and y are proportional (no allometry), while \( b \neq 1 \), this describes interesting (allometric) relationships.)

• Three types of allometric relationships:
  1. Dynamic allometry: Description of how one trait grows compared to another, either for an individual or as a mean.
  2. Static allometry: Describes how one trait compares to another for a given developmental stage for a set of organisms in a population.
  3. Evolutionary allometry: Describes how one trait compares to another in a set of species belonging to a clade.
Naïve measurement-theoretical justification for classic allometry

- Searching for a simple relationship between two phenotypic characters, x and y. Note that x and y must both be positive.

- The simplest relationship imaginable is the linear one, $y = a + bx$. But this can not work in general ($a \neq 0$), as there is either a positive y that yields a negative x or vice versa. The case $a = 0$ has too little description power, as one only has the option of describing proportionality.

- Solution, transform x and y so that the transformed characters can move between – infinity and infinity. Then assign a linear relationship.

- Simplest and most intuitive such transformation: $x' = \log(x)$, $y' = \log(y)$: $y' = a' + bx' \Rightarrow y = ax^b$ where $a = \exp(a')$
What if part (y) is a part of the whole (x)?

- Sometimes it may be interesting to compare a part to the whole. For instance:
  a) Brain mass to body mass
  b) Head height to total height (or length in most animals)
  c) Lung volume to chest cavity volume
- Can we still use the power-law relationship?
- Quick answer: No.

$$\forall y = ax^b \mid b \neq 1 \exists x \mid y > x$$

- i.e., for any allometric relationship, there is a case where the power-law predicts that the part becomes larger than the whole.

For $b > 1$, this happens for $x > a^{1/(b-1)}$

For $b < 1$, this happens for $x < a^{1/(b-1)}$
How to deal with a part (y) compared to a whole (x)?

1. Don’t do this comparison. Compare y to \( z = x - y \) instead and be happy with that. Unsatisfactory!

2. Pretend this is not a problem, and use the Huxley relationship anyhow.
   - This is done, I’m told! Requires that the variation in x and y is not too great. But the allometric parameters do not describe the distributional properties of neither x nor y!

3. Use the same kind of naïve measurement-theoretical approach as for the Huxley relationship.

4. Use the Huxley relationship to derive \( y = f(x) \).

Since 3 and 4 are competing approaches, a statistical test should be devised.
A statistical detour on allometry

In order to deal with competing allometric models, statistical versions are needed (for likelihood-based inference).

The statistical qualities of an allometric relationship ought to be interesting in itself. We would not be here discussing allometry if humans were on the allometric brain vs body size line!

The methods used for the standard power-law uses normal models on the log-log-scale as basis for the inference.

For two different parts, y and z, it should be interesting to see how z relates to y and vice versa. A joint distribution will do the trick and describe what is and is not a reasonable value for y alone, z alone and y and z together.
A joint distribution perspective on standard (Huxley) allometry

The bi-variate normal distribution is the simplest joint distribution describing a linear relationship:

\[
f(u,v) = \frac{1}{2\pi \sigma_u \sigma_v \sqrt{1-\rho^2}} \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \frac{(u-\mu_u)^2}{\sigma_u^2} + \frac{(v-\mu_v)^2}{\sigma_v^2} - 2\rho \frac{(u-\mu_u)(v-\mu_v)}{\sigma_u \sigma_v} \right] \right)
\]

In this example, the distributional density is shown with colors. 10000 samples are also plotted. Parameters:

\(\mu_u = 2, \mu_v = 1, \sigma_u = 0.2, \sigma_v = 0.3, \rho = 0.98\)

This could thus be used to model \(u = \log(y)\) and \(v = \log(z)\). Note that

\[
f(v | u) \sim N(\mu_v + \rho \frac{\sigma_v}{\sigma_u} (u-\mu_u), (1-\rho^2)\sigma_v^2) \Rightarrow v | u = c + bu + s_v \epsilon \quad (\text{where } \epsilon \sim N(0,1))
\]

Thus this description is equivalent to describing the distribution of \(u\) and the regression of \(v\) on \(u\). The regression of \(v\) on \(u\) contains the classic power-law allometric parameters, \(a = e^c\) and \(b\).
A joint distribution perspective on standard (Huxley) allometry

\[ f(u, v) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-\rho^2}} \exp\left( -\frac{1}{2(1-\rho^2)} \left[ \frac{(u-\mu_u)^2}{\sigma_u^2} + \frac{(v-\mu_v)^2}{\sigma_v^2} - 2\rho \frac{(u-\mu_u)(v-\mu_v)}{\sigma_u\sigma_v} \right] \right) \]

Since \( f(v \mid u) \sim N(\mu_v + \rho \frac{\sigma_v}{\sigma_u} (u-\mu_u), (1-\rho^2)\sigma_v^2) \)
we can find

\[ \text{mean}(v \mid u) = \text{median}(v \mid u) = \mu_v + \rho \frac{\sigma_v}{\sigma_u} (u-\mu_u), \]

Similarly, since

\( f(u \mid v) \sim N(\mu_u + \rho \frac{\sigma_u}{\sigma_v} (v-\mu_v), (1-\rho^2)\sigma_u^2) \)
we can find

\[ \text{mean}(u \mid v) = \text{median}(u \mid v) = \mu_u + \rho \frac{\sigma_u}{\sigma_v} (v-\mu_v) \]

We have a two-way regression-description of \( u \) and \( v \) and thus also the untransformed phenotypes, \( y \) and \( z \).
A joint distribution perspective on standard (Huxley) allometry (2)

\[ f(u, v) = \frac{1}{2\pi \sigma_u \sigma_v \sqrt{1 - \rho^2}} \exp \left( -\frac{1}{2(1 - \rho^2)} \left[ \frac{(u - \mu_u)^2}{\sigma_u^2} + \frac{(v - \mu_v)^2}{\sigma_v^2} - 2\rho \frac{(u - \mu_u)(v - \mu_v)}{\sigma_u \sigma_v} \right] \right) \]

The slope \( \sigma_v / \sigma_u \) describes the principle component, \( \rho \sigma_v / \sigma_u \) describes the slope of the regression of \( v \) on \( u \) and \( \rho \sigma_u / \sigma_v \) describes the slope of the regression of \( u \) on \( v \). Please note that the slope of \( v \) on \( u \) is not the inverse of the slope of \( u \) on \( v \)!

Thus a regressed relationship of \( u \) and \( v \) using \( u \) as the covariate and using \( v \) as the covariate will yield two different results. PCR will yield yet another. But all these ways of viewing a dataset are described in a joint distribution!

It would be nice for a developmental process that would yield a bi-variate normal distribution for static allometry! Dynamic allometry either describes a single individual or an average in a population, and may not so readily be assigned a joint distribution description. More on this later.
A joint distribution perspective on standard (Huxley) allometry on the original scale

If we are to compare different models, it’s best to agree on which data are to be compared. I propose working on the original scale. If we start with $u$ and $v$ bi-variate normal and $y = e^u$, $z = e^v$, then

$$f(y, z) = \frac{1}{2\pi \sigma_u \sigma_v yz\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \frac{(\log(y) - \mu_u)^2}{\sigma_u^2} + \frac{(\log(z) - \mu_v)^2}{\sigma_v^2} - 2\rho \frac{(\log(y) - \mu_u)(\log(z) - \mu_v)}{\sigma_u \sigma_v} \right]\right)$$

Distributional density is shown with colors. 10000 samples are also plotted. Parameters:

$\mu_u = 2$, $\mu_v = 1$, $\sigma_u = 0.2$, $\sigma_v = 0.3$, $\rho = 0.98$

$z$ conditioned on $y$ is described by:

$$z \mid y \sim \log N(\mu_v + \rho \frac{\sigma_v}{\sigma_u} (\log(y) - \mu_u), (1-\rho^2)\sigma_v^2)$$

This means that median$(z \mid y) = ay^b$ and mean$(z \mid y) = a' y^b$, where

$$a = e^{\mu_v - \rho \frac{\sigma_v}{\sigma_u} \mu_u}, \quad a' = e^{\mu_v - \rho \frac{\sigma_v}{\sigma_u} (1-\rho^2)\sigma_v^2 \mu_u}, \quad b = \rho \frac{\sigma_v}{\sigma_u}$$
How to deal with a part (y) compared to a whole (x) – solution 3 (naïve measurement-theory)

Use the same kind of naïve measurement-theoretical approach as for the Huxley relationship.

I.e. we want transformations of x and y so that the transformed quantities can range over the whole real axis. Then we can assign a linear relationship.
Solution 3 (naïve measurement-theory) – the functional relationship

Restrictions: $x > 0$, $0 < y < x$.

If we look at $x$ and $y/x$: $x > 0$, $0 < y/x < 1$

The logarithm is the standard transformation for positive quantities, $u = \log(x)$. The logit-transformation, $v = \logit(y/x) = \log((y/x)/(1-y/x))$, is the standard transformation for quantities between 0 and 1.

$v = \log(a) + (b-1)u$ yields (linear components chosen so as to make the result as easy as possible):

$$y = \frac{ax^b}{1 + ax^{b-1}}$$

This is a power-law when $ax^{b-1} \ll 1$.

For $ax^{b-1} \to \infty$, $y \to x$ from below. This happens for $x \to 0$ when $b < 1$ and for $x \to \infty$ when $b > 1$.
Solution 3 – The shape of the functional relationship

\[ y = \frac{ax^b}{1 + ax^{b-1}} \]

Case: \(a=0.5, b=3.5\). Note how the curvature changes as the curve starts approaching \(x\) when \(x\) increases.

Case: \(a=4, b=0.1\). Note how the curve approaches \(x\) when \(x\) decreases.
Solution 3 – The joint distribution

Starting with \( u = \log(x) \), \( v = \text{logit}(y/x) \), a linear joint distribution is assigned using the bi-variate normal distribution. This yields:

\[
\begin{align*}
  f(x, y) &= \frac{I(y < x)}{2\pi \sigma_u \sigma_v (x - y) y \sqrt{1 - \rho^2}} \\
  \exp &\left\{- \frac{1}{2(1 - \rho^2)} \left[ \frac{(\log(x) - \mu_u)^2}{\sigma_u^2} + \frac{(\log(y/(x - y)) - \mu_v)^2}{\sigma_v^2} - 2\rho \frac{(\log(x) - \mu_u)(\log(y/(x - y)) - \mu_v)}{\sigma_u \sigma_v} \right] \right\}
\end{align*}
\]

Distributional density is shown with colors. 10000 samples are also plotted. Parameters:
\( \mu_u = 3, \mu_v = -0.3, \sigma_u = 0.3, \sigma_v = 0.8, \rho = 0.98 \)

Note that the distribution is zero for \( y > x \), which is exactly as is wanted.

When \( \rho \to 1 \), the functional relationship should be regained. For \( \rho < 1 \), extracting mean and median \( y \) given \( x \) can possibly be analytically impossible. However, using that the median of an increasing transformation is the transformation of the median:

\[
\text{median}(y \mid x) = \frac{ax^b}{1 + ax^{b-1}}
\]
Solution 3 – the problem

While this approach yields results that adheres to the restriction $y < x$, it clashes with the allometric relationship between to different parts, $z = ay^b$. To see this, let $z = x - y$ and

$$y = \frac{ax^b}{1 + ax^{b-1}}$$

As a function of $x$, $z$ will have the same type of relationship with $x$ as $y$ has:

$$z = \frac{\frac{1}{a} x^{2-b}}{1 + \frac{1}{a} x^{(2-b)-1}} = \frac{a' x^{b'}}{1 + a' x^{b'-1}}$$

where $a' = \frac{1}{a}$ and $b' = 2 - b$

However, solving $z$ as a function of $x$, on the log-scale, one gets:

$$\log(z) = \log(y) - \log(a) - b \log(y + z)$$

which is not on the same form as the Huxley relationship,

$$\log(z) = \log(a) + b \log(y)$$

Thus, the logit-solution clashes with the standard allometric relationship. (Meaning that solution 3 and 4 are different). This may be a fairly unwanted property of this proposed allometric relationship!
How to deal with a part \((y)\) compared to a whole \((x)\) – solution 4 (starting with Huxley)

Here, we start with the Huxley relationship for one part \((z)\) as a function of the rest:
\[
z = x - y = ay^b.
\]

We now want \(y\) and \(z\) as a function of \(x\).

\[
x - y = ay^b \quad \Rightarrow \quad x = y(1 + ay^{b-1})
\]

The right hand term is a \(b\)-polynomial in \(y\). So, solving \(y\) as a function of \(x\) will in general not be analytically possible. (It’s possible numerically, though.)
The analytical relationship only describes $y$ and $x$ when there is no noise. Thus getting an analytical expression for that relationship may not be so important. As it turned out, an analytical expression of $(x,y)$ can still be found.

Start with $u = \log(x-y)$ and $v = \log(y)$. $(u,v)$ bi-variate normal will then express the usual (Huxley) allometric relationship between $y$ and $z=x-y$ in a joint distribution form. Re-parametrizing to $y = e^u$ and $x = y + z = e^u + e^v$, we get:

$$
\begin{align*}
  f(x, y) &= \frac{I(y < x)}{2\pi \sigma_u \sigma_v (x-y) y \sqrt{1-\rho^2}} \times \\
  \exp &\left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(\log(x-y) - \mu_u)^2}{\sigma_u^2} + \frac{(\log(y) - \mu_v)^2}{\sigma_v^2} - 2\rho \frac{(\log(x-y) - \mu_u)(\log(y) - \mu_v)}{\sigma_u \sigma_v} \right] \right\}
\end{align*}
$$

Distributional density is shown with colors. 10000 samples are also plotted. Parameters:

$\mu_u = 0, \mu_v = -1, \sigma_u = 0.3, \sigma_v = 0.8, \rho = 0.98$

Note that also this distribution is zero for $y > x$. Extracting mean and median $y$ given $x$ analytically, may not be possible. Since the functional form in the absence of noise is not analytically available, then neither should the mean and median.
Comparing the solutions

• Solution 2-4 have now each been given a joint distribution on the original scale. Comparing with data should thus be possible.
• Numerical tests with artificial, large (N=10000) datasets, suggest that a test based on comparing the maximum likelihood can successfully distinguish between datasets generated by the different models. Bayesian model comparison is also expected to work.
• Note that information criteria will boil down to comparing ML-based likelihoods, as the number of parameters are the same in all 3 models.
A classic (?) developmental description of the Huxley relationship

• If one starts with the developmental model:
  \[ \dot{y}(t) = \Gamma_y h(t) y(t), \quad \dot{z}(t) = \Gamma_z h(t) z(t) \]
where \( h(t) \) is a common growth factor. Then a log-transform, \( u = \log(y) \), \( v = \log(z) \), yields:

\[ \dot{u}(t) = \Gamma_y h(t), \quad \dot{v}(t) = \Gamma_z h(t) \]

• Let \( H(t) = \int h(u)du \)
  Then

\[ u(t) = u_0 + \Gamma_y H(t), \]
\[ v(t) = v_0 + \Gamma_z H(t) = (v_0 - \frac{\Gamma_z}{\Gamma_y} u_0) + \frac{\Gamma_z}{\Gamma_y} u(t) = \log(a) + bu(t) \]

• This will trace a power-law, \( z = ay^b \) for dynamic allometry, as developmental time, \( t \), progresses.

• A way to describe static allometry is to assume the there are individual random lags and increases in the common growth driver, \( h(t) \). But that will describe perfect correlation between \( u(t) \) and \( v(t) \). Not realistic, I would think.
A stochastified developmental description of the Huxley relationship

- Assume stochastic contributions are on the log-scale. (i.e. incremental normal contributions are proportional to the scale and the process itself is not allowed to get negative values.) Use a minimal stochastic differential equation expansion of the differential equation:

\[ du(t) = \Gamma_y h(t) dt + s_y dB_t^{(y)}, \quad dv(t) = \Gamma_z h(t) dt + s_z dB_t^{(z)} \]

- Assume further that the developmental common factor traces an underlying developmental plan, \( g(h) \), linearly but with stochastic contributions:

\[ dh(t) = -\alpha(h(t) - g(t)) dt + s_h dB_t^{(h)} \]

- This means that \( h(t) \) can be written as \( h(t) = h_{\text{det}}(t) + h_{\text{OU}}(t) \), where

\[ h_{\text{det}}(t) = e^{-\alpha t} E(h_0) + \int_0^t e^{-\alpha(t-s)} g(s) ds \]

and \( h_{\text{OU}}(t) \) is an Ornstein-Uhlenbeck (mean-reverting) process with drift \( \alpha \) and diffusion \( s_h \) (and mean=0).
A stochastified developmental description of the Huxley relationship (2)

• The developmental common factor, $h(t)$, is thus a sum of a deterministic function and an OU process.

• In order to solve for $u(t)$ and $v(t)$ we need to do time-integrals on $h(t)$.

• Theory: the time-integral of an OU process is a random walk for sufficiently large time-scales ($t>>1/\alpha$).

• Thus:

$$u(t) = u_0 + \Gamma_y \int_0^t h_{\text{det}}(s) ds + \Gamma_y s_h B(t)^{(h)} + s_y B(t)^{(y)}$$

$$v(t) = v_0 + \Gamma_z \int_0^t h_{\text{det}}(s) ds + \Gamma_z s_h B(t)^{(h)} + s_z B(t)^{(z)}$$

where $B(t) \sim N(\mu=0, \sigma^2=t)$ is a random walk. Thus if $(u_0,v_0)$ are bi-variate normally distributed (or variance considered vanishingly small) then $(u(t),v(t))$ will also be bi-variate normally distributed. Meaning that $(y(t),z(t))$ will be in a joint distribution with a power-law description of conditional means and medians. Note: dynamic power-law allometry could be broken due to increasing variance?
A stochastified developmental description of the Huxley relationship (4) – variants

• First a quick reminder of the stochastic differential equations going into the developmental model:

\[ du(t) = \Gamma_y h(t) dt + s_y dB_t^{(y)} \]
\[ dv(t) = \Gamma_z h(t) dt + s_z dB_t^{(z)} \]
\[ dh(t) = -\alpha (h(t) - g(t)) dt + s_h dB_t^{(h)} \]

• If the diffusion terms were allowed to vary in time, the contributions to \( h(t), u(t) \) and \( v(t) \) should still be normal. We can even described this using the planned growth: \( s_h \rightarrow s_{h}(t) \equiv s'_{h}(g(t)) \)

• This would complicate the modelling, but ought to yield the same result, bivariate normal \((u(t),v(t))\), standard power-law for \((y(t),z(t))\) and solution 4 for \( y(t) \) vs \( x(t) \).

• However, if we allow \( s_h \rightarrow s'_{h}(h(t)) \) we get a feedback from previous stochasticity that can not be expected to yield normal \( h(t), u(t) \) or \( v(t) \).

• Discussion point: Is this more or less realistic?
A stochastified developmental description of the Huxley relationship (5) – more variants

\[
du(t) = \Gamma_y h(t) dt + s_y dB_t^{(y)}, \quad dv(t) = \Gamma_z h(t) dt + s_z dB_t^{(z)} \\
dh(t) = -\alpha(h(t) - g(t)) dt + s_h dB_t^{(h)}
\]

• The part-specific growth constants, \( \Gamma_y \) and \( \Gamma_z \), may also have some time-dependency.
• If this time-dependency is deterministic, Huxley allometry will be retained for the static allometry, but broken for the dynamic allometry.
• If this is stochastic (linear tracing SDE) but \( h(t) = g(t) \) is deterministic, I would expect the same, except that there will be no correlation for static allometry!
• If either \( \Gamma_y \) or \( \Gamma_z \) is stochastic, and also \( h(t) \) is so, then I so no way to get a resulting normal model for \( u(t) \) and \( v(t) \).

What I expect from dynamic allometry if \( \Gamma_z \) is constant while \( \Gamma_z \) changes from one fixed value to another at a given time.
Partition scaling

If we make a standard relationship description for one part vs another or for one part vs a larger part containing it, it can not be universal unless

a) All separate parts relate to one another with the same description. So, \((y_1,y_2), (y_1,z), (y_2,z), (y,z),(y_1,y_2+z)\) should be described using the same mathematical or distributional apparatus.

b) (alt) All parts contained in larger parts relate to one another with the same description. So \((y,x),(z,x),(y_1,y),(y_2,y),(y_2,y_2+z)\) should be described using the same mathematical or distributional apparatus.
Partition scaling – how do the solutions stack up?

Solution 3:
- Functional relationship: If \( y \) relates to \( x \) using the functional relationship
  \[
  y = \frac{ax^b}{1 + ax^{b-1}}
  \]
  and \( y_1 \) relates to \( y \) the same way, then \( y_1 \) relates to \( x \) with a more complicated relationship. Requirement b fails.
- Distributional relationship: If \( x \) is log-normal, then \( y \) is not. But \( y \) is required to be log-normal when related to \( y_1 \) or \( y_2 \). Requirement b fails again.

Solution 4:
- Functional relationship: If \( y_1 = az^b \) and \( y_2 = a'z^{b'} \) then \( y = y_1 + y_2 = az^b + a'z^{b'} \), which does not follow the functional form.
- Distributional form: If \( y_1 \) and \( y_2 \) are log-normally distributed, then \( y \) is not and vice versa.
Partition scaling – a way out?

If solution 3 and 4 fails, what’s left?

The distributional descriptions both failed because if two parts, \( y_1 \) and \( y_2 \), are described with a given distributional family, then \( y = y_1 + y_2 \) will be described with a different distribution. But \( y \) is again a part of a bigger scheme, and ought to be described by the same distributional family.

Def: A stable distribution is one where if \( y_1 \) and \( y_2 \) independent are from the same distributional family then \( ay_1 + by_2 = cy + d \), where also \( y \) is from the same distributional family. Four such known families: normal, Cauchy, Levy and gamma with fixed scale (which seems weird to require). Only Levy and gamma works for positive variables.

Problem 1: The Levy distribution is extremely heavy-tailed. Not realistic?
Problem 2: \( y_1 \) and \( y_2 \) are *not* independent. \( \Rightarrow \) More or less possible distributions?
Problem 3: No standard dependency between Levy or gamma distributions?
Problem 4: A process having the Levy (or gamma with fixed scale) distribution as a result, could suggest the nature of the dependency. But what kind of process is that?
Questions to the audience

• How big a problem is the lack of partition scaling? Should we solve this and if so, how?
• Can solution 3 to the part-as-a-function-of-the-whole-problem, be dismissed entirely?
• Are there datasets that can resolve between solution 3 and 4?
• Are there datasets that can show that the Huxley model applied directly to these problems (solution 2) is inappropriate?
• In the developmental model
  \[ du(t) = \Gamma_y h(t)dt + s_y dB_t^{(y)}, \quad dv(t) = \Gamma_z h(t)dt + s_z dB_t^{(z)} \]
  \[ dh(t) = -\alpha(h(t) - g(t))dt + s_h dB_t^{(h)} \]
  ➢ Are there factors here that ought to be time-dependent, that aren’t?
  ➢ If yes, is the time dependency deterministic or stochastic?
  ➢ Does this stochasticity depend on stochastic processes?
  ➢ If stochastic, are the stochastic contributions small but steady (white-noise contributions as in SDEs) or seldom but sudden (jump-processes) or a combination?
  ➢ What’s the best compromise between realism and tractability?
• What about evolutionary allometry? Are there process models resulting in allometry there?

Lots of data and/or great variance in y/x is expected to be necessary.