Errors and uncertainty in variables – When to worry and when to Bayes?

Stefanie Muff

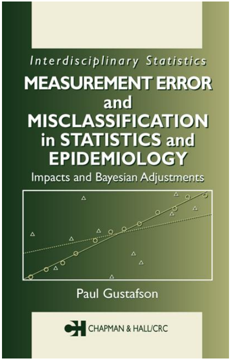
Errors-in-Variables Workshop Mainz
2. December 2016

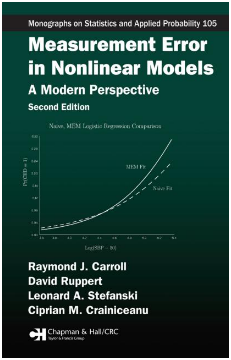
Sources of measurement uncertainty / measurement error (ME)

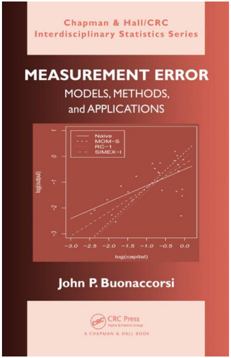
- **Measurement imprecision** in the field or in the lab (length, weight, blood pressure, etc.).
- Errors due to **incomplete** or **biased observations** (e.g., self-reported dietary aspects, health history).
- Biased observations due to **preferential sampling or repeated observations**.
- **Misclassification error** (e.g., exposure or disease classification).
- ...

"Error" or "uncertainty"?

- The existence, effects and treatment of ME has been discussed in the literature for more than a century (*e.g.* Pearson 1902, Wald 1940).
- A standard reference is Fuller (1987).
- More modern monographs are Gustafson (2004); Carroll et al. (2006); Buonaccorsi (2010); Yi (2016).







Why should ME not be ignored?

- It is a fundamental assumption that explanatory variables are measured or estimated without error, for instance for
  - the calculation of correlations.
  - linear, generalized linear and non-linear regressions and ANOVA.
  - survival analysis.
- Most other modelling assumptions are routinely checked.
- Violation of this assumption may lead to biased parameter estimates, altered standard errors and p-values, incorrect covariate importances, and to misleading conclusions.
- Even standard statistics textbooks do often not mention these problems.
- Interestingly, the topic of missing data has received considerable attention in the past decade – it is a special case of ME (or the other way round...)!

Example 1: Inbreeding in Alpine ibex

Goal: To quantify effect of inbreeding on the intrinsic population growth rate r_0 of 26 Alpine ibex populations.

(Bozzuto et al., 2016)

Analysis: A simple linear regression with $y_i = \log(r_0)_i$ as response

$$y_i = \beta_0 + \beta_x x_i + \mathbf{z}_i^\top \beta_z + \varepsilon_i \ ,$$

and erroneous measure of inbreeding $x_i = f_i$ for population i.

- If the estimated inbreeding values w_i are plugged in the regression, the naive estimate is

$$\hat{\beta}_x = -6.0 \ , 95\% \text{ CI: } [-11.2, -0.9] \ .$$

- If, however, the uncertainty estimate of w_i is included in an error model, the estimate is

$$\hat{\beta}_x = -10.6 \ , 95\% \text{ CI: } [-17.2, -4.5] \ .$$

$\rightarrow$ If the ME in w_i is not accounted for, the estimated influence of inbreeding on population growth is underestimated or attenuated.

Example 2: Framingham heart study

Goal: To investigate the influence of systolic blood pressure (SBP) on coronary heart disease from $n = 641$ males (Kannel et. al 1986).
Components:

Analysis:

- the error-prone covariate $x_i = \log(SBP - 50)$, measured twice.
- the error-free covariate $z_i \in \{0, 1\}$ indicating smoking status.
- response $y_i \in \{0, 1\}$ (diseased no/yes).
- **Logistic regression**

$$\eta_i = \text{logit}[Pr(y_i = 1)] = \beta_0 + \beta_x x_i + \beta_z z_i \ .$$

- **Naive** estimate: $\hat{\beta}_x = 1.66$, 95% CI: $[0.70, 2.63]$.
- ME-**adjusted**: $\hat{\beta}_x = 1.89$, 95% CI: $[0.79, 3.01]$.

Example 3: Miscounting error in a clinical trial

COPD: Chronic obstructive pulmonary disease

Exacerbation: A sudden worsening of symptoms that requires treatment with antibiotics, corticosteroids or hospitalization.

Goal: Investigate the effect of a pharmacotherapy vs placebo ($x_i \in \{0, 1\}$) on the number of exacerbations (y_i) of COPD patients (Calverley et al., 2007).

Analysis: Negative binomial regression with exacerbation numbers as outcome:

$$y_i \sim \text{NBin}\left(\exp(\log(t_i) + \beta_0 + x_i\beta_x + \mathbf{z}_i\boldsymbol{\beta}_z), \theta\right)$$

Study duration was 3 years. Additional covariates $\mathbf{z}_i$, t_i=actual time under treatment (offset).

Problem: Exacerbation numbers y_i are self-reported by the patients, and thus miscounted.

- In a separate study, Frei et al. (2016) investigated the error in the number of self-reported exacerbations for 409 patients during 3 years.

- Comparison between patient self-reports s_i and consensus classifications by a central adjudication committee, consisting of several experienced physicians ("gold standard", y_i).

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	127	24	5	4	2	2	1	0	0	0	0	0	0
1	26	40	5	2	1	3	0	0	0	0	0	0	0
2	9	17	10	4	2	1	0	0	0	0	0	0	0
3	3	6	7	10	2	3	2	1	0	0	0	0	0
4	1	7	3	6	2	3	2	1	0	0	0	1	0
5	0	3	5	4	0	4	1	1	0	0	0	0	0
6	0	2	4	1	6	1	2	0	0	0	0	0	0
7	0	2	2	0	2	0	0	0	0	0	0	0	0
8	0	0	0	2	2	0	1	2	1	0	0	0	1
9	2	0	0	1	0	0	0	1	1	0	0	0	0
10	...	...	...	...	...	...	...	...	...	...	...	...	...

Table : Self-reports (rows) vs. centrally adjudicated numbers (columns).

- The external validation data were used to estimate the parameters of a zero-inflated negative binomial error model:

$$s_i \mid y_i \sim \text{ZINB}\left(\gamma_0 + \gamma_1 y_i, p_i, \theta_E\right) \ .$$

- Modelling error accordingly, the actual treatment effect estimate increases:

  Naive rate ratio $\exp(\hat{\beta}_x) = 0.86$ (95% CI from 0.78 to 0.95)
  Corrected rate ratio $\exp(\hat{\beta}_x) = 0.80$ (95% CI from 0.68 to 0.93)
  (smaller=stronger)

Overview of error types

- Error in continuous vs error in categorical or count variables.
- Classical vs Berkson error.
- Differential vs non-differential error.
- Error in covariates vs error in the response.
- Error in linear regression vs error in a generalized linear (mixed) model.
- ...

Notation

- True response $\mathbf{y}$.
- True covariate that is subject to measurement error $\mathbf{x}$.
- The observed, erroneous proxy of $\mathbf{x}$ is denoted as $\mathbf{w}$.
- In the presence of reponse error, the observed, erroneous proxy of $\mathbf{y}$ is denoted as $\mathbf{s}$.
- Other covariates observed without error $\mathbf{z}$.

Error in continuous covariates

We then distinguish between two different ME processes:

1. **The classical ME model**
   $\mathbf{w} = \mathbf{x} + \mathbf{u}$

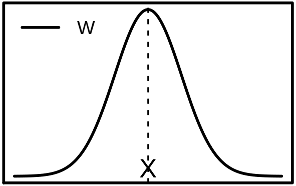
2. **The Berkson ME model**
   $\mathbf{x} = \mathbf{w} + \mathbf{u}$

The classical ME model

$\mathbf{x}$ is the correct but *unobserved* variable and $\mathbf{w}$ the observed proxy with error $\mathbf{u}$. Then

$$\begin{aligned} \mathbf{w} &= \mathbf{x} + \mathbf{u} \\ \mathbf{u} &\sim \mathrm{N}(\mathbf{0}, \sigma_u^2 \mathbf{D}) \quad , \end{aligned}$$

is the classical ME model.



Usually, $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ and $d_i \propto \sigma_u^2(x_i)$.

Assumption: $\mathbf{u}$ is independent of $\mathbf{x}$; error is non-differential.

Characteristics of classical ME

Or: How do I identify **classical** error/uncertainty in a variable?

- Usually, classical ME occurs in the context of measurements, e.g., in the field or in the lab.
- A typical characteristic is that

$$\sigma_w^2 = \sigma_x^2 + \sigma_u^2 \ ,$$

that is: the measured variable **w** is more variable than the true **x**.

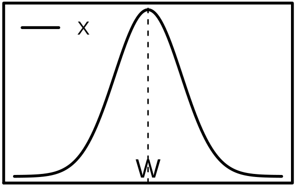
The Berkson ME model

Again, $\mathbf{x}$ is the correct but *unobserved* variable and $\mathbf{w}$ the observed proxy with error $\mathbf{u}$. Then

$$\begin{aligned} \mathbf{x} &= \mathbf{w} + \mathbf{u} \\ \mathbf{u} &\sim \text{N}(\mathbf{0}, \sigma_u^2 \mathbf{D}) \end{aligned}$$

is the Berkson ME model.

(Berkson, 1950)



Usually, $\mathbf{D} = \text{diag}(d_1, \ldots, d_n)$ and $d_i \propto \sigma_u^2(x_i)$.

Assumption: $\mathbf{u}$ is independent of $\mathbf{w}$; error is non-differential.

Characteristics of Berkson ME

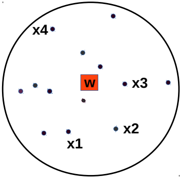
Or: How do I identify **Berkson** error/uncertainty in a variable?

- Berkson error can occur
  - in experimental settings (predefined fixed concentration or time interval).
  - when a variable is rounded.
  - in exposure models, e.g. in environmental or epidemiologic studies.

- A typical characteristic is that

$$\sigma_x^2 = \sigma_w^2 + \sigma_u^2 \ ,$$

meaning that the true variable **x** is more variable than the observed **w**.



Of course, more complicated error structures are possible. Examples include

- Classical error with dependencies on an error-free covariate **z** (Prentice et al., 2002)

$$w_i = \gamma_0 + \gamma_1 x_i + \gamma_2 z_i + \gamma_3 x_i z_i + u_i \ .$$

- Multiplicative error structures (additive on the log scale):

$$w_i = x_i \cdot u_i \quad \Rightarrow \quad \log(w_i) = \log(x_i) + \log(u_i)$$

- Berkson *and* classical error in the same covariate.

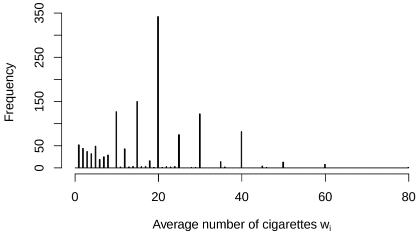
Error in binary/categorical covariables and counts

- Binary and categorical variables are particularly relevant in epidemiologic research → Misclassification.

  Misclassification matrix:

$$\Pi = \begin{pmatrix} 0.8 & 0.25 \\ 0.2 & 0.75 \end{pmatrix}$$

- Miscounting error: May occur in any count variable. Example: self-reported cigarette consumption in survival or epidemiologic studies.



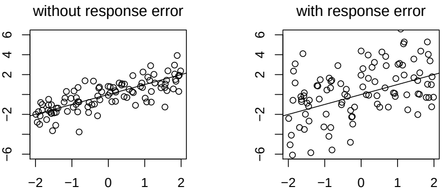
Error in the outcome of regression models

- Continuous error in a linear regression outcome.

  Note: In the case when the observed response

  $$s_i = y_i + v_i \quad v_i \sim \mathrm{N}(0, \sigma_v^2) \ ,$$

  the error variance is simply absorbed in the residual variance σ_ϵ^2.



- Misclassification of the (binary) outcome in logistic regression.
- Miscounting error of the outcome in Poisson regression.

Non-differential vs differential error

Non-differential error (Carroll et al., 2006):

> Non-differential ME occurs when **w** contains no information about **y** other than what is available in **x** and **z**.

Technically, this means that

$$\mathbf{y} \mid \{\mathbf{x}, \mathbf{z}, \mathbf{w}\} = \mathbf{y} \mid \{\mathbf{x}, \mathbf{z}\} \ .$$

Differential error

The error is differential otherwise.

> **Note:** In most error modelling approaches, the assumption is that the error is non-differential!

The effects of ME

1. The biasing effects of ME on the parameter estimates can be roughly categorized into
   - Attenuation: the slope parameters are underestimated.
   - Reverse attenuation: the slope parameters are overestimated.
2. ME leads to a loss of power for detecting signals.
3. ME masks imporant features of the data, making graphical model inspection difficult.

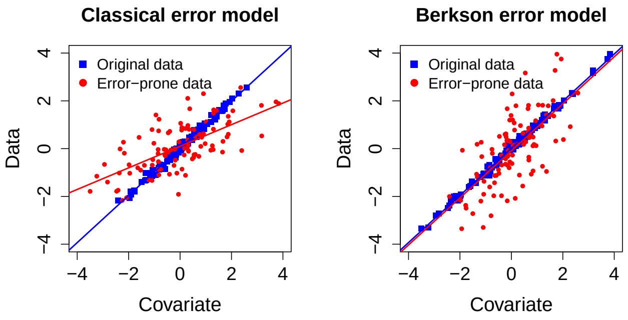
Carroll et al. (2006) call this the **"Triple Whammy of Measurement Error"**.

Effect of ME in linear regression

Find regression parameters β_0 and β_x for unobserved $\mathbf{x}$:

$$y_i = 1 \cdot {\color{red} x_i} + \epsilon_i, \quad \epsilon_i \sim \mathrm{N}(0, \sigma_\epsilon^2) \ .$$

Simulation: $n = 100$, $\sigma_\epsilon^2 = 1/100$, $\sigma_x^2 = \sigma_u^2 = 1$.



Bias in linear regression: Formulas

Let us look at the linear regression model

$$y_i = \beta_0 + \beta_x x_i + \mathbf{z}_i^\top \boldsymbol{\beta}_z + \epsilon_i$$

with error-prone x_i and error-free covariates $\mathbf{z}_i$.

Classical error:

When the observed covariate $w_i = x_i + u_i$, $u_i \sim \mathrm{N}(0, \sigma_u^2)$ is included instead of x_i, the estimated slope parameter is

$$\beta_x^\star = \lambda \cdot \beta_x \qquad \text{with} \qquad \lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \ ,$$

where $\lambda \leq 1$ is denoted as the attenuation factor.

Berkson error:

On the other hand, when the error is given as $x_i = w_i + u_i$, $u_i \sim \mathrm{N}(0, \sigma_u^2)$, then $\beta_x^\star = \beta_x$, that is, no bias is expected!

Attenuation or reverse attenuation?

In the presence of ME/uncertainty in a covariate, the β (slope) parameters are often underestimated ("dilution bias").

In (medical) studies, an implicit assumption is often that (treatment) effects are conservative.

However, this it **by no means always the case**!

Examples that may induce reverse attenuation:

- In the presence of collinear covariates (Freckleton, 2011).
- When the error is differential (Mwalili et al., 2008).
- In logistic regression, β_x may be attenuated *or* reversely attenuated when $\mathbf{x}$ is mismeasured (even for non-differential error)!
- ...

Reverse attenuation example 1: collinear covariates

Situation:

$$y_i = \beta_0 + \beta_x x_i + \beta_z z_i + \epsilon_i \ , \quad \text{Cov}(\mathbf{x}, \mathbf{z}) \neq 0 \ .$$

Then: Parameters β_z of covariate **z** measured without error may be biased by the error in **x**. Naive least-squares does not estimate β_z, but

$$\beta_z^\star = \beta_z + \beta_x(1-\lambda)\Gamma_z \ ,$$

where Γ_z is the slope of **z** when **x** is regressed on **z**, i.e., $\text{E}(\mathbf{x} \,|\, \mathbf{z}) = \Gamma_0 \mathbf{1} + \Gamma_z \mathbf{z}$.

$\rightarrow$ If $\text{Cov}(\mathbf{x}, \mathbf{z}) > 0$ then $\beta_z^\star > \beta_z$, thus reverse attenuation!

Reverse attenuation example 2: heteroscedastic error

Assume we have a linear regression model including a continuous error-prone covariate **x**, a binary covariate **z** $\in \{0,1\}$ indicating group membership (e.g., sex), and an interaction term **xz**:

$$y_i = \beta_0 + \beta_x x_i + \beta_z z_i + \beta_{xz} x_i z_i + \epsilon_i \ .$$

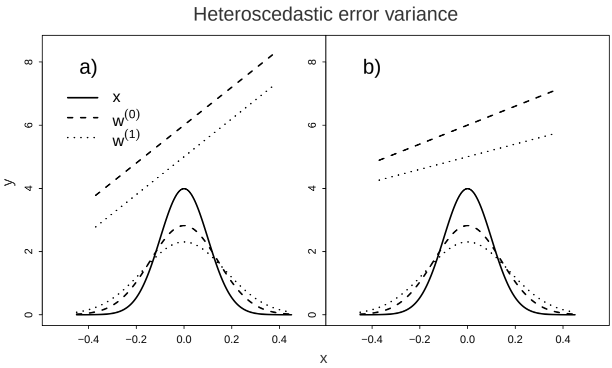
Further assume that the measurements of **x** are more precise for individuals in group 0 than in group 1, i.e., that the error variance depends on z_i.

Formally:

$$w_i = x_i + u_i \ , \quad \begin{cases} u_i \sim \mathrm{N}(0, \sigma_{u_0}^2), \ \text{if } s_i = 0 \ , \\ u_i \sim \mathrm{N}(0, \sigma_{u_1}^2), \ \text{if } s_i = 1 \ , \end{cases}$$

and $\sigma_{u_0}^2 < \sigma_{u_1}^2$.

Let the true interaction coefficient be $\beta_{xz} = 0$, but the error variance heteroscedastic with $\sigma^2_{u_0} < \sigma^2_{u_1}$.



a) When **x** is included in the regression, the regression lines $\mathbf{y} \sim \mathbf{x}$ for groups 0 and 1 are parallel; no interaction, $\hat{\beta}_{xz} = 0$.

b) When **w** is included in the regression, the non-parallel regression lines for $\mathbf{y} \sim \mathbf{w}$ indicate a spurious interaction, $\hat{\beta}_{xz} > 0$ (Muff and Keller, 2015).

Effect of misclassification or miscounting in the outcome

- A misclassified or miscounted outcome (e.g. in logistic or Poisson regression) typically induces attenuation of the regression parameters.
- However, if the error distribution in the outcome depends in some way on the covariates $\mathbf{z}$, anything can happen...

When do I have to worry?

Many applied scientists ask for guidelines to decide if the error they find in their data can be tolerated, and when it is substantial, so that error modelling is necessary.

Some thoughts from my side:

- If analytical formulas to calculate the bias exist, you should use them to obtain an estimate of the expected bias.
- Otherwise, simulations are often a good idea: generate error-free data and add error of the type you encounter in your case.
- There is no general rule about the error that can be tolerated – **this must depend on your situation** (e.g., clinical study vs explorative analysis)

A pragmatic check

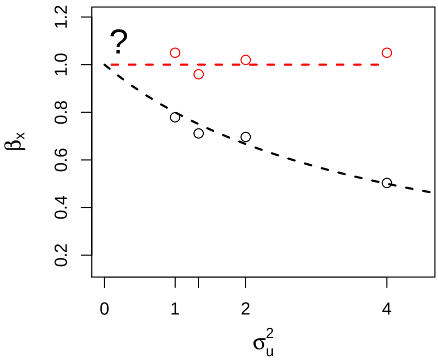
Assume your error-prone variable has been measured repeatedly. Then try the following:

1. Fit the model iteratively, each time including as variable only one single measurement.
2. Fit the model iteratively, each time including the average of two measurements.
3. Continue with 3, 4, ... measurements.
4. Finally, fit the model and include the average of all repeats.
5. Look at the trend of your estimates.

If there is a clear trend of your parameter estimates that worries you, error modelling might be worth.

Note: This simple check is similar in spirit to the SIMulation EXtrapolation (SIMEX) idea (Cook and Stefanski, 1994).

Example of the “pragmatic check” idea when 4 repeated measurements are available for a covariate:



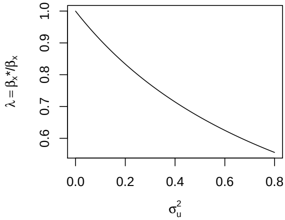
Analytic formulas: linear regression

For $y_i = \beta_0 + \beta_x x_i + \epsilon_i$ with error-prone covariate x_i and classical error such that $w_i = x_i + u_i$, $u_i \sim \text{N}(0, \sigma_u^2)$, the biased versions of the parameters are given as

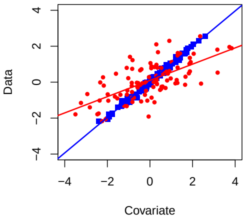
$$\beta_x^\star = \lambda \beta_x \quad \text{and} \quad \beta_0^\star = \beta_0 + (1-\lambda)\beta_x \mu_x \ ,$$

with $\lambda = \sigma_x^2 / (\sigma_x^2 + \sigma_u^2)$.

- $\beta_x^\star$ decreases with increasing σ_u^2:



- β_0 is unbiased **if** the covariate x is centered.



Analytic formulas: other regression types

There is no general formula for all regression types...
Some simple cases:

- Berkson error in a continuous covariate of log-linear models (e.g., Poisson regression): All parameters unbiased, except $\beta_0^\star = \beta_0 + \sigma_u^2/2$.

- Berkson error in a continuous covariate of Probit regression ($\boldsymbol{\beta} = (\beta_0, \beta_1, ..)^\top$):
$$\boldsymbol{\beta}^\star = \boldsymbol{\beta} \cdot (1 + \sigma_u^2)^{-1/2} \ .$$

But generally, I recommend simulations to investigate potential effects.

Simulations or apps

Shiny app for some classial error in linear, logistic and Poisson regression:

▸ Classical error

▸ Berkson error

Caveats of error modelling

(Which might lead to the decision **not** to model the error.)

1. **Bias vs variance trade-off**:
   Error analysis leads to an estimate with higher variability / more uncertainty.

2. **Error analysis comes at a cost**:
   Additional (internal/external) data is needed to estimate the structure and parameters of the error model.

   Estimates from external validation data are assumed to be **transportable**, which is often not fulfilled.

   And, believe me, error modelling can be tedious!

3. **Loss of power:**
   Even when error is accounted for, power cannot be gained back.

In any case, assessing the biasing effect of the error, as well as error modelling, can be done only if the error structure (model) and the respective model parameters (e.g., error variances) are known!

Therefore: Information about the error mechanism is essential, and potential errors must be identified early in a study – ideally in the planning phase.

Error correction methods

Many different ME modelling approaches have been proposed:

- Method-of-moments correction
- Simulation extrapolation (SIMEX)
- (Quasi-) Likelihood approaches
- Multiple imputation
- Bayesian methods
- ...

Why Bayesian ME modelling?

1. **Incorporation of prior knowledge**:
   Most non-Bayesian approaches require the precise estimation of error model parameters (e.g., the error variance). Instead, Bayesian approaches naturally allow to incorporate prior uncertainty.

2. **Simple and general:**
   The formulation of Bayesian error models is usually straightforward (hierarchical modelling).

3. **Identifiability issues:**
   Most models with error components are nonidentifiable, e.g.:

$$w_i = x_i + u_i \quad \text{with} \quad \sigma_w^2 = \sigma_x^2 + \sigma_u^2 \ .$$

   The error variance σ_u^2 and the sampling variance σ_x^2 are *confounded*. However, Bayesian approaches allow to estimate the posterior distribution even if only crude information about σ_u^2 is available!
   $\rightarrow$ Partially identified models (Gustafson, 2005).

A word on (non)identfiability

The "Bayesian crank" can be turned even if a model is nonidentifiable (Gustafson, 2015).

All we need is a legitimate probability distribution as prior distribution.

A word on notation

In the Bayesian context, variances are often parameterized as precisions.

Thus from now on, we will use, e.g.

$$\begin{aligned}\tau_x &= 1/\sigma_x^2 \\ \tau_u &= 1/\sigma_u^2 \\ \tau_\epsilon &= 1/\sigma_\epsilon^2\end{aligned}$$

etc...

Bayesian error modelling steps

(Assuming that a regression model is given, and that stucture and severity of the error have been assessed.)

1. Formulate the error model.
2. Combine the regression and the error model into a hierarchical model.
3. Specify prior distributions for all parameters, in particular for the error model parameters.
4. Estimate the posterior distribution using MCMC or INLA.

Step 1: Formulate the error model

Remember the various error types and formulate a model that encodes for the relation between the true and the observed variable.

Examples:

- Continuous variables: $w_i = x_i + u_i$ or $x_i = w_i + u_i$, u_i homo- or heteroscedastic.
- Binary variables: $\Pr(w_i = 1 \mid x_i) = \frac{\exp(\alpha_0 + \alpha_x x_i)}{(1+\exp(\alpha_0+\alpha_x x_i))}$
- Count variables: True counts y_i vs obsered counts s_i

$$s_i \mid y_i \sim \text{ZINB}\left(\gamma_0 + \gamma_1 y_i, p_i, \theta_E\right) \ .$$

Step 2: Formulate a hierarchical model

The error model from step 1 is now combined with the regression model of interest. The hierarchy is given by (at least) two levels:

Regression model	(level 1)
Error model	(level 2)

As an example, for linear regression with classical, homoscedastic ME in $\mathbf{x}$, the hierarchical model is given as

$$\begin{aligned} y_i &= \beta_0 + \beta_x x_i + \beta_z z_i + \epsilon_i \ , \quad \boldsymbol{\epsilon} \sim \mathrm{N}(0, \tau_\epsilon \mathbf{I}) \\ w_i &= x_i + u_i \ , \quad \mathbf{u} \sim \mathrm{N}(0, \tau_u \mathbf{I}) \ . \end{aligned}$$

Step 3: Prior distributions

Prior specifications are needed for all unobserved variables.

In the example above, a model for **x** is needed, e.g., a so-called exposure model

$$x_i = \alpha_i + \alpha_z z_i + \epsilon_{xi} , \quad \boldsymbol{\epsilon}_x \sim \mathrm{N}(0, \tau_x \mathbf{I}) \ .$$

Moreover, priors are needed for $(\beta_0, \beta_x, \beta_z)$, and (α_0, α_z), as well as hyperpriors for τ_x, τ_u and τ_ϵ.

Step 4: Estimate the posterior distribution

Essentially two approaches:

- Markov chain Monte Carlo (MCMC) sampling
- Integrated nested Laplace approximations (INLA)

MCMC

- MCMC is very general, flexible and widely used.
- A first rush of ME modelling with MCMC in the 1990s (Stephens and Dellaportas, 1992; Richardson and Gilks, 1993).
- However, case-specific implementation may be challenging: need to specify full conditionals, sampling design, check mixing and convergence properties...
- Sampling can become rather time-consuming.
- Generic software such as `jags` (Plummer, 2003) or `Stan` (Carpenter et al., 2016) provide simple ways to perform MCMC sampling.

INLA

- INLA was introduced as a fast and accurate alternative to MCMC sampling (Rue et al., 2009).
- INLA is able to deal with latent Gaussian hierarchical models, consisting of three sub-models:
  - Observation model $\mathbf{y} \mid \mathbf{v}, \boldsymbol{\theta}_1$: Encodes information about data.
  - Latent model $\mathbf{v} \mid \boldsymbol{\theta}_2$: The unobserved process.
  - Hyperpriors for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$: Models for the hyperparameters in the observation and latent processes.
- It has been shown that error modelling for continuous covariates (classical and Berkson ME) is possible with INLA for generalized linear mixed models (GLMMS, Muff et al., 2015) and for survival models.
- Caveat: Misclassification error, response error in categorical and count outcomes.

Hierarchical model for classical ME in INLA

- Observation model
  - **Regression model**: $p(\mathbf{y} \mid \mathbf{x}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\theta}_1)$

$$\mathrm{E}(\mathbf{y}) = h^1(\beta_0 + \beta_x \mathbf{x} + \mathbf{z}^\top \boldsymbol{\beta}_z)$$

  - **Error model**: $p(\mathbf{w} \mid \mathbf{x}, \boldsymbol{\theta}_2)$

$$\mathbf{w} = \mathbf{x} + \mathbf{u} \ , \qquad \mathbf{u} \sim \mathrm{N}(\mathbf{0}, \tau_u \mathbf{D})$$

- Latent model for $\mathbf{v} = (\beta_0, \boldsymbol{\beta}_z^\top, \alpha_0, \boldsymbol{\alpha}_z^\top, \mathbf{x}^\top)^\top$
  - **Exposure model** for $\mathbf{x}$: $p(\mathbf{x} \mid \boldsymbol{\theta}_2)$

$$\mathbf{x} = \alpha_0 + \mathbf{z}^\top \boldsymbol{\alpha}_z + \boldsymbol{\varepsilon}_x, \quad \boldsymbol{\varepsilon}_x \sim \mathrm{N}(\mathbf{0}, \tau_x \mathbf{I})$$

  - Independent Gaussian priors for $(\beta_0, \boldsymbol{\beta}_z^\top, \alpha_0, \boldsymbol{\alpha}_z^\top)$

- Hyperpriors $p(\boldsymbol{\theta}_1)$, $p(\boldsymbol{\theta}_2)$ with $\boldsymbol{\theta}_2 = (\beta_x, \tau_u, \tau_x)^\top$

[1] monotonic inverse link function, **y** of exp. family form

Joint model formulation for classical ME:

$$\begin{aligned}
\mathsf{E}(\mathbf{y}) &= h(\beta_0 + \beta_x \mathbf{x} + \mathbf{z}^\top \boldsymbol{\beta}_z) \ , & & \\
\mathbf{0} &= -\mathbf{x} + \alpha_0 + \mathbf{z}^\top \boldsymbol{\alpha}_z + \boldsymbol{\epsilon}_x \ , & \boldsymbol{\epsilon}_x &\sim \mathsf{N}(\mathbf{0}, \tau_x \mathbf{I}) \ , \\
\mathbf{w} &= \mathbf{x} + \mathbf{u} \ , & \mathbf{u} &\sim \mathsf{N}(\mathbf{0}, \tau_u \mathbf{D}) \ .
\end{aligned}$$

Challenges:

- $\mathbf{x}$ appears in all three levels of the model, with and without multiplication by β_x.
- Different likelihood functions are involved.

$$\underbrace{\begin{bmatrix}
y_1 & \texttt{NA} & \texttt{NA} \\
\vdots & \vdots & \vdots \\
y_n & \texttt{NA} & \texttt{NA} \\
\texttt{NA} & 0 & \texttt{NA} \\
\vdots & \vdots & \vdots \\
\texttt{NA} & 0 & \texttt{NA} \\
\texttt{NA} & \texttt{NA} & w_{1.} \\
\vdots & \vdots & \vdots \\
\texttt{NA} & \texttt{NA} & w_{n.}
\end{bmatrix}}_{\texttt{Y}} = \ldots$$

Definition in `r-inla`

Note the application of the **"copy"** function

```
> library(INLA)
> formula <- Y ~  f(beta.x, copy = "idx.x",
+     hyper = list(beta = list(param = prior.beta, fixed = FALSE))) +
+   f(idx.x, weight.x, model = "iid", values = 1:n,
+     hyper = list(prec = list(initial = -15, fixed = TRUE))) +
+   beta.0 - 1 + beta.z + alpha.0 + alpha.z
```

Note the definition of three likelihood functions

```
> r <- inla(formula, Ntrials = Ntrials, data = data,
+   family = c("binomial", "gaussian", "gaussian"),
+   control.family = list(
+     list(hyper = list()),
+     list(hyper = list(
+           param = prior.prec.x,
+           fixed = FALSE)),
+     list(hyper = list(
+           param = prior.prec.u,
+           fixed = FALSE))),
+   control.fixed = list(
+     mean = prior.beta[1],
+     prec = prior.beta[2])
+   )
```

The `mec` model

- If $\mathbf{x}$ is assumed independent of other covariates, a simplified model can be formulated:

$$\mathbf{x} = \alpha_0 + \boldsymbol{\epsilon}_x \ , \quad \boldsymbol{\epsilon}_x \sim \mathrm{N}(\mathbf{0}, \tau_x \mathbf{I}) \ .$$

- For this case we implemented a model termed "`mec`".

- Technically, this is done by directly formulating a latent model for $\boldsymbol{\nu} = \beta_x \mathbf{x}$. The model has four hyperparameters: β_x, τ_x, τ_u, α_0.

Hierarchical model for the Berkson ME

- Observation model
  - **Regression model**: $\text{p}(\mathbf{y} \,|\, \mathbf{x}, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\theta}_1)$

$$\text{E}(\mathbf{y}) = h(\beta_0 + \beta_x \mathbf{x} + \mathbf{z}^\top \boldsymbol{\beta}_z)$$

- Latent model for $\mathbf{v} = (\beta_0, \boldsymbol{\beta}_z^\top, \mathbf{x}^\top)^\top$
  - **Error model**: $\text{p}(\mathbf{x} \,|\, \boldsymbol{\theta}_2)$

$$\mathbf{x} = \mathbf{w} + \mathbf{u} \ , \qquad \mathbf{u} \sim \text{N}(0, \tau_u \mathbf{D})$$

  - Independent Gaussian priors for $(\beta_0, \boldsymbol{\beta}_z^\top)$

- Hyperpriors $\text{p}(\boldsymbol{\theta}_1)$, $\text{p}(\boldsymbol{\theta}_2)$ with $\boldsymbol{\theta}_2 = (\beta_x, \tau_u)^\top$

Joint model formulation for Berkson ME in INLA

$$\begin{aligned} \mathrm{E}(\mathbf{y}) &= h(\beta_0 + \beta_x \mathbf{x} + \mathbf{z}^\top \boldsymbol{\beta}_z) \ , \\ -\mathbf{w} &= -\mathbf{x} + \mathbf{u} \ , \qquad\qquad \mathbf{u} \sim \mathrm{N}(0, \tau_u \mathbf{D}) \ . \end{aligned}$$

- Things are easier here because the latent model for **x** is the same as the error model:

$$\mathbf{x} \,|\, \mathbf{w}, \boldsymbol{\theta} \sim \mathrm{N}(\mathbf{w}, \tau_u \mathbf{D}) \ .$$

- Directly formulate a model termed "`meb`" with two hyperparameters β_x, τ_u by reparameterizing $\boldsymbol{\nu} = \beta_x \mathbf{x}$:

$$\boldsymbol{\nu} \,|\, \mathbf{w}, \boldsymbol{\theta} \ \sim \ \mathrm{N}\left(\beta_x \mathbf{w}, \ \frac{\tau_u}{\beta_x^2}\mathbf{D}\right) \ .$$

Example 1: Inbreeding in Alpine ibex

Remember:

- A simple linear regression with $y_i = \log(r_0)_i$ as response

$$y_i = \beta_0 + \beta_x x_i + \mathbf{z}_i^\top \beta_z + \varepsilon_i \ ,$$

  and erroneous measure of inbreeding $x_i = f_i$ for population i.

- The error in x_i is assumed to be classical: $w_i = x_i + u_i$, and w_i was estimated from a separate analysis providing an error precision $\hat{\tau}_u(x_i)$ for each population ($\rightarrow$ heteroscedastic error model).

INLA applicable?

Yes!

Step 1: Formulate the error model (classical heteroscedastic error model)

Step 2: Formulate the hierarchical model:

$$\begin{aligned} \mathbf{y} \,|\, \mathbf{x} &\sim \mathrm{N}(\beta_0 + \beta_x \mathbf{x} + \mathbf{z}\boldsymbol{\beta}_{\mathbf{z}}, \tau_\epsilon \mathbf{I}), \\ \mathbf{w} \,|\, \mathbf{x} &\sim \mathrm{N}(\mathbf{x}, \tau_u \mathbf{D}) \ , \end{aligned}$$

with $\mathbf{y}$ the intrinsic growth rate and $\mathbf{x}$ the inbreeding coefficient.

Step 3: Prior distributions.

- Assume $\mathbf{x}$ to be independent of other covariates:

$$\mathbf{x} \sim \mathrm{N}(\alpha_0 \mathbf{1}, \tau_x \mathbf{I}) \ .$$

- $\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 10^{-4}\mathbf{I})$ and $\boldsymbol{\alpha} \sim \mathrm{N}(\mathbf{0}, 10^{-4}\mathbf{I})$
- Hyperpriors for τ_x, τ_u, τ_ϵ are motivated by expert knowledge.

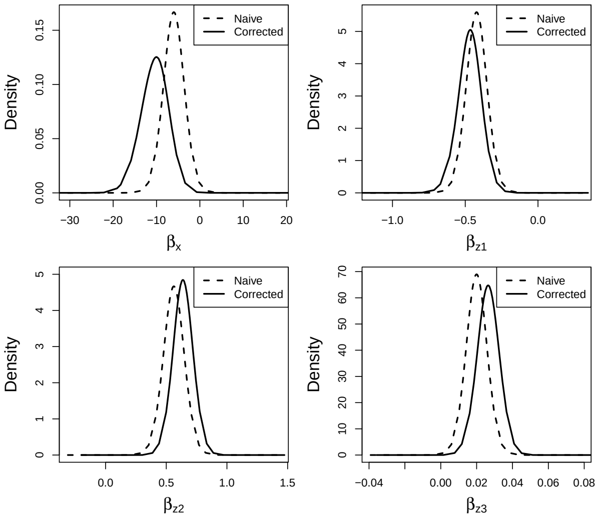
Step 4: Estimate posterior distributions with `r-inla`:

```
> formula <- y ~ f(w, model = "mec", scale = error.prec, values=w, hyper = list(
+   beta = list(param = prior.beta, fixed = FALSE),
+   prec.u = list(param = prior.prec.u, fixed = FALSE),
+   prec.x = list(param = prior.prec.x, fixed = FALSE),
+   mean.x = list(initial = 0, fixed = TRUE))
+ ) + z1 + z2 + z3 + z4

> r <- inla(formula, data = data.frame(y, w, z1, z2, z3, z4, error.prec),
+   family = "gaussian",
+   control.family = list(
+     hyper = list(prec = list(param = prior.prec.y, fixed = FALSE)
+     )),
+   control.fixed = list(
+     mean.intercept = prior.beta[1],
+     prec.intercept = prior.beta[2]
+   )
+ )
```

(For more details, please consult the Supp. Mat. of Muff et al. (2015), or the examples on the `r-inla` website at www.r-inla.org.)

Posterior distribution of β_x and $\boldsymbol{\beta}_z$, naive and error-corrected estimates:



Example 2: Framingham heart study

Remember:

A binary regression model

$$\eta_i = \text{logit}[Pr(y_i = 1)] = \beta_0 + \beta_x x_i + \beta_x z_i$$

with systolic blood pressure as error-prone covariate $x_i = \log(SBP - 50)$, and response $y_i \in \{0, 1\}$ (diseased no/yes).

INLA applicable? Yes!

Step 1: Classical, homoscedastic error model $w_i = x_i + u_i$ with $u_i \sim \mathrm{N}(0, \tau_u)$, each individual measured twice.

Step 2: Formulate the hierarchical model:

$$\begin{aligned}\mathrm{logit}\left[\Pr(\mathbf{y}=1)\right] &= \beta_0 + \beta_x \mathbf{x} + \beta_z \mathbf{z} \ , \\ \mathbf{w}_j \,|\, \mathbf{x} &\sim \mathrm{N}\left(\mathbf{x}, \tau_u \mathbf{I}\right) \ , j = 1, 2.\end{aligned}$$

Step 3: Prior distributions

- Assume $\mathbf{x}$ to depend on smoking status:

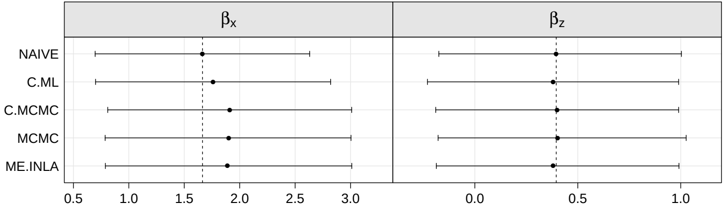
$$\mathbf{x} \,|\, \mathbf{z} \sim \mathcal{N}(\alpha_0 \mathbf{1} + \alpha_z \mathbf{z}, \tau_x \mathbf{I}) \ .$$

- $\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 10^{-2}\mathbf{I})$ and $\alpha_0, \alpha_1 \sim \mathrm{N}(0, 1)$.
- Hyperpriors for τ_x and τ_u are motivated by expert knowledge.

Step 4: Estimate posterior marginals with `r-inla`.

The example is also available on the `r-inla` website at www.r-inla.org.

Posterior distributions:



Example 3: Miscounting error in a clinical trial

Remember:

- Count outcome that was modelled with a negative binomial regression model, including x_i=treatment of patient i and other error-free covariates $\mathbf{z}_i$.
- The outcome is miscounted, that is, not y_i was observed, but some self-reported values s_i instead.
- An external validation study gave information on the error structure and error parameters (Frei et al., 2016).

INLA applicable? **No!** The hierarchical model is not latent Gaussian...

Step 1: Miscounting error according to a zero-inflated negative binomial model:

$$s_i \mid y_i \sim \text{ZINB}\left(\gamma_0 + \gamma_1 y_i, p_i, \theta_E\right) \ , \tag{1}$$

with $\text{logit}(p_i) = \delta_0 + \delta_1 \text{I}(y_i > 0)$, where y_i is unobserved.

Step 2: Combine the above error model with the regression model to a hierarchical model:

$$y_i \sim \text{Po}\left(\exp(\log(t_i) + \beta_0 + x_i\beta_x + \mathbf{z}_i\boldsymbol{\beta}_z)\right) \ .$$

Note that a Poisson regression model is used now.

Assumption: All extra-variability and zero-inflation in the measured response is attributed to the miscounting process.

Step 3: Priors:

- Use a normal prior on $(\gamma_0, \gamma_1, \delta_0, \delta_1) \sim \mathsf{N}(\hat{\boldsymbol{\alpha}}, \hat{\Sigma})$ with parameters from the fit of external validation data to the ZINB error model (1):

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}} = (\hat{\gamma}_0, \hat{\gamma}_1, \hat{\delta}_0, \hat{\delta}_1) &= (0.753, 0.966, 0.151, -3.174) \\
\hat{\Sigma} &= \begin{pmatrix} 0.020 & -0.007 & 0.033 & -0.019 \\ -0.007 & 0.007 & -0.011 & 0.018 \\ 0.033 & -0.011 & 0.122 & -0.094 \\ -0.019 & 0.018 & -0.094 & 0.401 \end{pmatrix}
\end{aligned}
$$

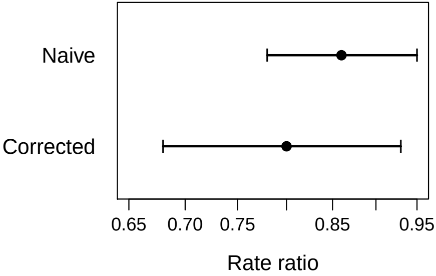
- In addition: $\hat{\theta}_E = 6.09$ with $\mathsf{se}(\hat{\theta}_E) = 2.03$, thus a log-normal prior $\theta_E \sim \mathsf{LN}(\log(6.09), 0.33^2)$ was used.
- Independent $\mathcal{N}(0, 10^{-2})$ priors on $\boldsymbol{\beta}$.

Step 4: Estimate posterior marginals using `r-jags`. Example jags code using fixed error model parameters $\boldsymbol{\alpha}$:

```
> model
> {
+ for (i in 1:Nobservations)
+ {
+   # Response model for true response; reduced model for illustration
+   Y.true[i] ~ dpois(exp(beta[1]+X[i]*beta[2] + loge[i]))
+
+   # Error model
+   Y.report[i] ~ dnegbin(thetaE/(thetaE + mu1[i]),thetaE)
+   mu1[i] <- mu2[i] * x[i] + 1E-09
+
+   mu2[i] <- alpha1[1] + alpha1[2]*Y.true[i]
+
+   x[i] ~ dbern(1-pro[i])
+   logit(pro[i]) <- LP[i]
+   LP[i] <- alpha1[3] + alpha1[4]*YY[i]
+   YY[i] <- Y.true[i]>0
+ }
+
+ # Priors:
+ for (i in 1:nbetas){beta[i]~dnorm(0,1.0E-2)}
+ Log_thetaE ~ dnorm(log(6.09),1/0.33^2)
+ thetaE <- exp(Log_thetaE)
+ }
```

- Two parallel MCMC chains with 25'000 iterations each and a burn-in of 5'000 iterations were run to sample from the posterior distribution.
- Computation time roughly 1 hour (on a slow remote environment).
- Convergence was checked visually.

Naive ML results and posterior 95% credible interval for the rate ratio $\exp(\hat{\beta}_x)$ of the treatment effect:



$\rightarrow$ The treatment effect was clearly underestimated in the naive analysis!

A word on transportability

Problem: Using data from an external validation study may lead to a prior-data-conflict $\rightarrow$ violation of the *transportability assumption*.

Idea: Adaptive weighting of the priors, using the recently suggested adaptive prior weighting approach by Held and Sauter (2016). Multiply the covariance matrix from the validation data with an unknown scalar $g > 0$, leading to the prior

$$\boldsymbol{\alpha} \,|\, g \quad \sim \quad \mathsf{N}(\hat{\boldsymbol{\alpha}}, g\hat{\boldsymbol{\Sigma}}) \ ,$$

with hyperprior

$$t \quad = \quad \frac{g}{g+1} \sim \mathsf{U}(0,1) \ .$$

This allows to weight the error model priors $\hat{\boldsymbol{\alpha}}$ with $w = 1/g$.

Some (frequent) questions:

1. “I think I have error in my variables, but I don’t know its structure and parameters. Can I do something?”

2. “Is it sometimes better to ignore the error, that is, not to model it?”

Some (frequent) questions:

1. "I think I have error in my variables, but I don't know its structure and parameters. Can I do something?"

2. "Is it sometimes better to ignore the error, that is, not to model it?"

1. The short answer is: No.
   But at least you could check the effects of potential errors, e.g. via simulations.

Some (frequent) questions:

1. "I think I have error in my variables, but I don't know its structure and parameters. Can I do something?"
2. "Is it sometimes better to ignore the error, that is, not to model it?"

1. The short answer is: No.
   But at least you could check the effects of potential errors, e.g. via simulations.
2. Yes, absolutely! If the error is "neglectable", error modelling introduces additional uncertainty (bias-variance-tradeoff).
   Moreover, if you don't know your error structure, better don't do anything: You could make the bias worse.

Summary

- Uncertainty and error in covariates and response variables has various effects (not just bias).
- There are many different error mechanisms.
- Error modelling is only possible when error structure and model parameters are (approximately) known.
- "When to worry?" depends on many aspects, especially on the context.
  $\rightarrow$ A pragmatic way to answer the question is by simulations.
- Bayesian approaches are particularly useful for error modelling.
  $\rightarrow$ MCMC or INLA.

Thank you for your attention!

References:

Bozzuto, C., I. Biebach, S. Muff, A. R. Ives, and L. F. Keller (2016). Inbreeding reduces long-term population growth rates of reintroduced Alpine ibex. In preparation.

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications*. Boca Raton, FL: CRC Press.

Calverley, P. M., J. A. Anderson, B. Celli, G. T. Ferguson, C. Jenkins, P. W. Jones, J. C. Yates, and J. Vestbo (2007). Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *New England Journal of Medicine 356*, 775–789.

Carroll, R., D. Ruppert, L. Stefanski, and C. Crainiceanu (2006). *Measurement Error in Nonlinear Models: A Modern Perspective* (2 ed.). Boca Raton: Chapman & Hall.

Cook, J. and L. Stefanski (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association 89*, 1314–1328.

Freckleton, R. P. (2011). Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology 65*, 91–101.

Frei, A., L. Siebeling, C. Wolters, L. Held, P. Muggensturm, A. Strassmann, M. Zoller, G. ter Riet, and M. Puhan (2016). The inaccuracy of patient recall for COPD exacerbation rate estimation and its implications: Results from central adjudication. *CHEST 150*(4), 860–868.

Fuller, W. A. (1987). *Measurement Error Models*. New York: John Wiley & Sons.

Gustafson, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Boca Raton: Chapman & Hall/CRC.

Gustafson, P. (2005). On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. *Statistical Science 20*, 111–140.

Gustafson, P. (2015). *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data*. Boca Raton: Chapman & Hall/CRC.

Held, L. and R. Sauter (2016). Adaptive prior weighting in generalized linear models. *Biometrics*. Early View at http://dx.doi.org/10.1111/biom.12541.

Muff, S. and L. F. Keller (2015). Reverse attenuation in interaction terms due to covariate error. *Biometrical Journal 57*, 1068–1083.

Muff, S., A. Riebler, L. Held, H. Rue, and P. Saner (2015). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 64*, 231–252.

Mwalili, S., E. Lesaffre, and D. Declerck (2008). The zero-inflated negative binomial regression model with correction for misclassification: An example in caries research. *Statistical Methods in Medical Research 17*, 123–139.

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In K. Hornik, F. Leisch, and A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna.

Prentice, R. L., E. Sugar, C. Y. Wang, M. Neuhouser, and P. Patterson (2002). Research strategies and the use of nutrient biomarkers in studies of diet and chronic disease. *Public Health Nutrition 5*(6A), 977–984.

Richardson, S. and W. Gilks (1993). Conditional independence models for epidemiological studies with covariate measurement error. *Statistics in Medicine 12*.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological) 71*, 319–392.

Stephens, D. and P. Dellaportas (1992). Bayesian analysis of generalised linear models with covariate measurement error. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*. Oxford Univ Press.

Yi, G. Y. (2016). *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*. New York: Springer.

Appendix

Defining a joint model

Challenge:

$\mathbf{x}$ appears in different levels of the model (either with β_x or without).

Idea within INLA:

Create an **almost identical copy** $\mathbf{x}^\star$ for $\beta\mathbf{x}$ and extend **the latent model** to $\mathbf{x}_c = (\mathbf{x}, \mathbf{x}^\star)$, with $\pi(\mathbf{x}_c) = \mathrm{p}(\mathbf{x})\,\mathrm{p}(\mathbf{x}^\star \mid \mathbf{x})$, and

$$\mathrm{p}(\mathbf{x}^\star \mid \mathbf{x}, \tau) \propto \exp\left(-\frac{\tau}{2}(\mathbf{x}^\star - \mathbf{x})^\top(\mathbf{x}^\star - \mathbf{x})\right),$$

with precision τ fixed to some large value.

Defining a joint model

Challenge:

$\mathbf{x}$ appears in different levels of the model (either with β_x or without).

Idea within INLA:

Create an **almost identical copy** $\mathbf{x}^\star$ for $\beta\mathbf{x}$ and extend **the latent model** to $\mathbf{x}_c = (\mathbf{x}, \mathbf{x}^\star)$, with $\pi(\mathbf{x}_c) = \mathrm{p}(\mathbf{x})\,\mathrm{p}(\mathbf{x}^\star \,|\, \mathbf{x})$, and

$$\mathrm{p}(\mathbf{x}^\star \,|\, \mathbf{x}, \tau, \psi) \propto \exp\left(-\frac{\tau}{2}(\mathbf{x}^\star - \psi\mathbf{x})^\top(\mathbf{x}^\star - \psi\mathbf{x})\right),$$

with precision τ fixed to some large value. The copied model may contain an unknown scale parameter ψ, which represents here β_x.

The `mec` model

- Let us consider the simplified model without exposure model, i.e.,

$$\boldsymbol{\eta} = \beta_x \mathbf{x} \ ,$$
$$\mathbf{w} = \mathbf{x} + \mathbf{u} \ ,$$
$$\mathbf{x} = \alpha_0 + \boldsymbol{\epsilon}_x \ ,$$

  with $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \tau_u \mathbf{D})$ and $\boldsymbol{\epsilon}_x \sim \mathcal{N}(\mathbf{0}, \tau_x \mathbf{I})$.

- To be tractable by INLA, $\mathbf{x}$ must be representable as a Gaussian model.

The `mec` model

The **posterior distribution** of $\mathbf{x}$ and $\boldsymbol{\theta}$ is

$$\mathrm{p}(\mathbf{x}, \boldsymbol{\theta} \,|\, \mathbf{y}, \mathbf{w}) \;\propto\; \mathrm{p}(\boldsymbol{\theta}) \; \underbrace{\mathrm{p}(\mathbf{x} \,|\, \boldsymbol{\theta}) \;\; \mathrm{p}(\mathbf{w} \,|\, \mathbf{x}, \boldsymbol{\theta})}_{\mathrm{p}(\mathbf{x} \,|\, \mathbf{w}, \boldsymbol{\theta})\,\mathrm{p}(\mathbf{w} \,|\, \boldsymbol{\theta})} \; \mathrm{p}(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\theta})$$

Thus, **$\mathbf{x}$ only enters in one term** (apart from the likelihood) and can be treated as an ordinary **latent Gaussian model**:

$$\begin{aligned} \mathrm{p}(\mathbf{x} \,|\, \mathbf{w}, \boldsymbol{\theta}) &\propto \mathrm{p}(\mathbf{x} \,|\, \boldsymbol{\theta})\, \mathrm{p}(\mathbf{w} \,|\, \mathbf{x}, \boldsymbol{\theta}) \\ &\propto \exp\left(-\frac{\tau_x}{2}(\mathbf{x} - \alpha_0\mathbf{1})^\top(\mathbf{x} - \alpha_0\mathbf{1}) - \frac{\tau_u}{2}(\mathbf{x} - \mathbf{w})^\top\mathbf{D}(\mathbf{x} - \mathbf{w})\right) \;. \end{aligned}$$

Combining the quadratic forms gives

$$\mathbf{x} \,|\, \mathbf{w}, \boldsymbol{\theta} \;\sim\; \mathcal{N}\left[(\tau_x\alpha_0\mathbf{1} + \tau_u\mathbf{D}\mathbf{w})(\tau_x\mathbf{I} + \tau_u\mathbf{D})^{-1} \;,\; \tau_x\mathbf{I} + \tau_u\mathbf{D}\right].$$

The `mec` model

A more convenient model formulation is achieved by setting

$$\beta_x \mathbf{x} \rightarrow \boldsymbol{\nu}.$$

Then

$$\boldsymbol{\nu} \,|\, \mathbf{w}, \boldsymbol{\theta} \;\sim\; \mathcal{N}\left(\beta_x(\tau_x \alpha_0 \mathbf{1} + \tau_u \mathbf{D}\mathbf{w})(\tau_x \mathbf{I} + \tau_u \mathbf{D})^{-1}, \; \frac{\tau_x \mathbf{I} + \tau_u \mathbf{D}}{\beta_x^2}\right) \;.$$

This model is termed "`mec`" within the `R`-package `r-INLA`. Its hyperparameters are β_x, τ_x, τ_u, α_0.

Note that now both β_x and α_0 are considered as hyperparameters.

The `meb` model

- Let us consider the simplified model without covariates:

$$\mathrm{E}(\mathbf{y}) = \beta_x \mathbf{x} \ ,$$
$$\mathbf{x} = \mathbf{w} + \mathbf{u} \ , \qquad \mathbf{u} \sim \mathcal{N}(0, \tau_u \mathbf{D}) \ .$$

- The latent model $\mathbf{x}|\mathbf{w}, \boldsymbol{\theta}$ now corresponds to the error model.
- It is thus straightforward to calculate the posterior distribution

$$\mathrm{p}(\mathbf{x}, \boldsymbol{\theta} \,|\, \mathbf{y}, \mathbf{w}) \ \propto \ \mathrm{p}(\boldsymbol{\theta}) \ \mathrm{p}(\mathbf{x} \,|\, \mathbf{w}, \boldsymbol{\theta}) \ \mathrm{p}(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\theta}) \ .$$

- Using the reparameterization $\boldsymbol{\nu} = \beta_x \mathbf{x}$ leads to

$$\boldsymbol{\nu} \,|\, \mathbf{w}, \boldsymbol{\theta} \ \sim \ \mathcal{N}\left(\beta_x \mathbf{w}, \ \frac{\tau_u}{\beta_x^2} \mathbf{D}\right) \ .$$