STSCI 4780
Hierarchical/graphical models for measurement error, 2

Tom Loredo, CCAPS & SDS, Cornell University

Agenda

1. **Lec21: Density estimation with measurement error**

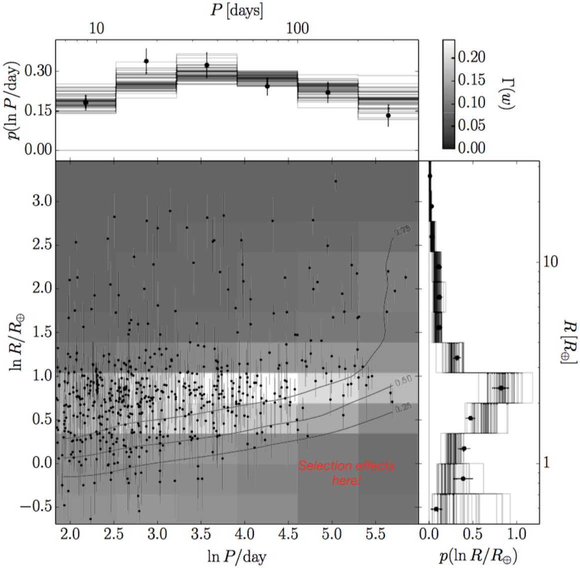
2. **Regression with measurement error**

TESS

- Transiting Exoplanet Survey Satellite (TESS)
- SpaceX Falcon 9 TESS mission launch and landing
- Falcon Heavy first landing

Exoplanets: Density estimation with measurement error

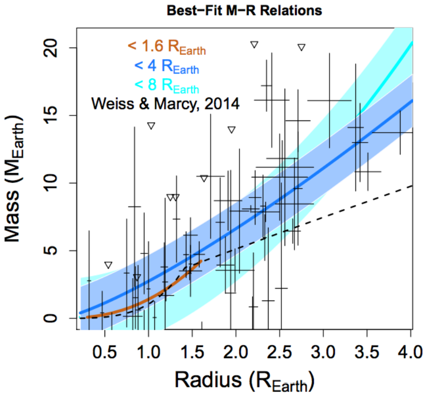
Exoplanet occurrence rate vs. orbital period, size



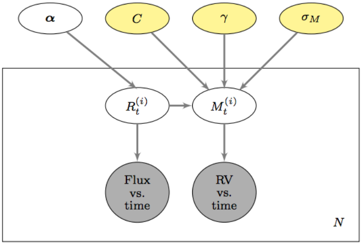
Foreman-Mackey[+] 2014

Exoplanets: Regression with measurement error

Exoplanet mass vs. radius → avg. density → composition

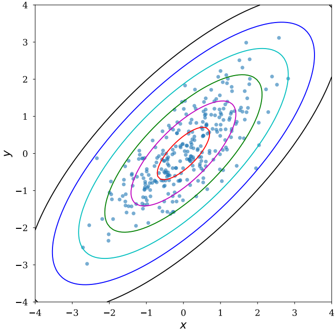


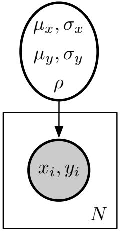
Wolfgang+ 2014



Lec18: Regression with precise predictor data

BVN density estimation

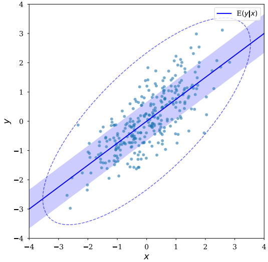


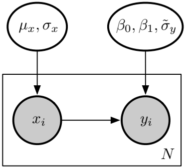


Joint: $p(\mu_x, \sigma_x, \mu_y, \sigma_y, \rho) \prod_{i=1}^{N} p(x_i, y_i | \mu_x, \sigma_x, \mu_y, \sigma_y, \rho)$

Inference: $p(\mu_x, \sigma_x, \mu_y, \sigma_y, \rho | \{x_i, y_i\}) \propto$ Joint

BVN regression





Regression focuses on $p(y|x)$:

$$\text{Joint:} \quad p(\mu_x, \sigma_x)\, p(\beta_0, \beta_1, \tilde{\sigma}_y) \prod_{i=1}^{N} p(x_i|\mu_x, \sigma_x)\, p(y_i|x_i, \beta_0, \beta_1, \tilde{\sigma}_y)$$

$$\text{Inference:} \quad p(\beta_0, \beta_1, \tilde{\sigma}_y|\{x_i, y_i\}) = \int d\mu_x \int d\sigma_x\, p(\mu_x, \sigma_x, \beta_0, \beta_1, \tilde{\sigma}_y|\{x_i, y_i\})$$

$$\propto p(\beta_0, \beta_1, \tilde{\sigma}_y) \prod_{i=1}^{N} p(y_i|x_i, \beta_0, \beta_1, \tilde{\sigma}_y)$$

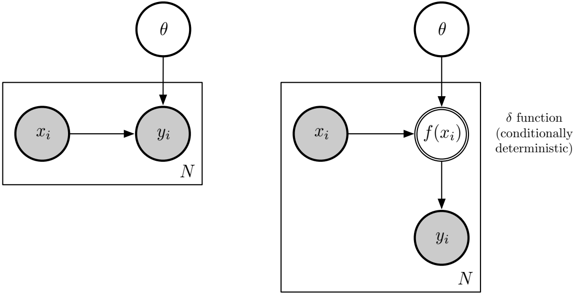
Note that $p(x_i|\cdots)$ drops out

Parametric regression

Infer θ determining the *conditional expectation*

$$\mathbb{E}(y_i|x_i,\theta) = f(x_i;\theta)$$

Regression function may be implicit (conditional expectation of y dist'n) or explicit ("true + error" or "typical + scatter")



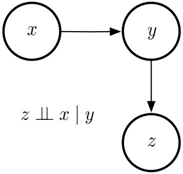
Often natural to express this via an additive error model:

$$y_i = f(x_i;\theta) + \epsilon_i; \qquad \mathbb{E}(\epsilon_i) = 0$$

Lec16: DAGs with missing edges

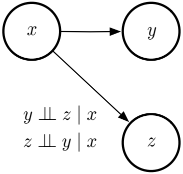
Conditional independence

$p(x)\,p(y|x)\,p(z|y)$



"Causal chain"

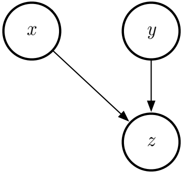
$p(x)\,p(y|x)\,p(z|x)$



"Common cause"

Conditional dependence

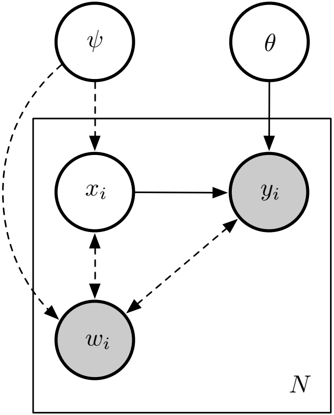
$p(x)\,p(y)\,p(z|x,y)$



"Multiple causes
or common effect"

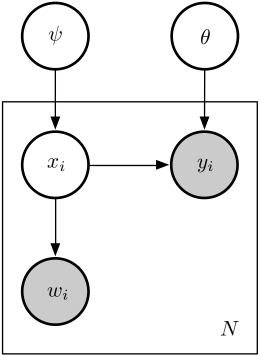
Regression with measurement error

Suppose x_i is not measured precisely—many variations!



See: *Measurement error in nonlinear models* (Carroll, Ruppert, Stefanski, Crainiceanu 2006) CRC Press

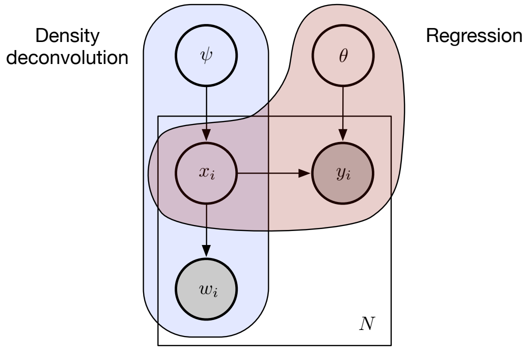
Classical measurement error



Prototype: Noisy measurement of a predictor

$$x_i \sim p(x_i|\psi)$$
$$w_i = x_i + \delta_i; \quad \delta_i \sim \mathcal{N}(0, \sigma_x^2)$$
$$y_i = f(x_i; \theta) + \epsilon_i; \quad \epsilon_i \sim \mathcal{N}(0, \sigma_y^2)$$

May have $\sigma_y \in \theta$ and/or $\sigma_x \in \psi$



Suppose we just want to learn θ, and ψ is *known*, so $p(x_i)$ is a fixed dist'n:

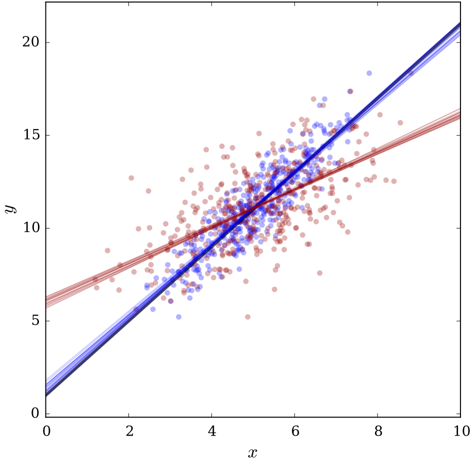
$$p(\theta|\{w_i, y_i\}) \propto \pi(\theta) \prod_i \int dx_i \, p(x_i) \, p(w_i|x_i) \, p(y_i|x_i, \theta)$$

Here *$p(x_i)$ does* not *drop out*; we must know or infer $p(x_i)$, and regression inference is now affected by the x_i marginal

Attenuation

Classical measurement error makes the noisy predictor measurement more dispersed than the true predictor values $\rightarrow$ tend to *under*estimate slopes

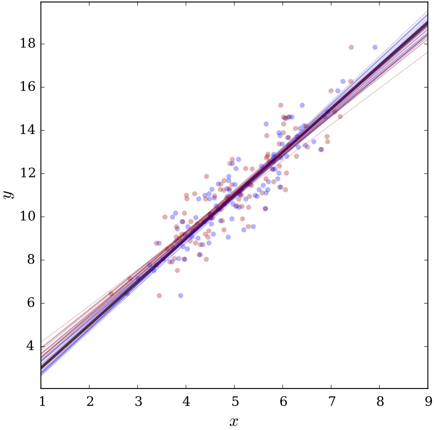
Example: $\sigma_x = 1$, $\sigma_w = 1 \quad \Rightarrow \quad \sqrt{\sigma_x^2 + \sigma_w^2} = \sigma_x \sqrt(2)$



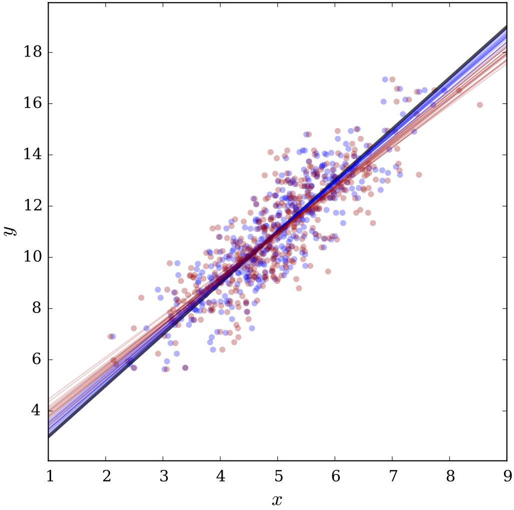
Measurement error does not "average out"

Small measurement error may produce a small effect on estimates, but the effect does not diminish as sample size grows → Ignoring measurement error often produces *inconsistent estimates*

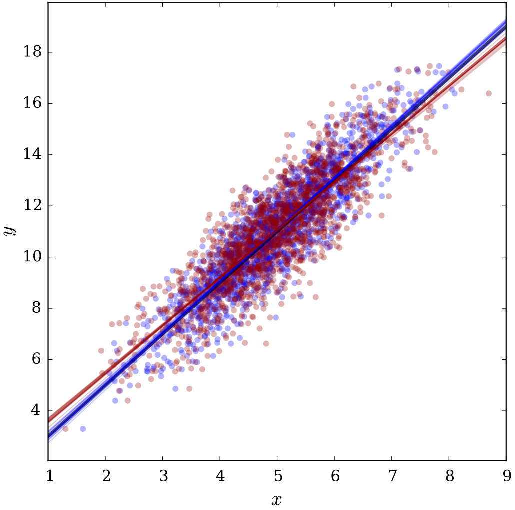
Example: $\sigma_x = 1$, $\sigma_w = 0.286$ $\quad\Rightarrow\quad$ $\sqrt{\sigma_x^2 + \sigma_w^2} = 1.04\sigma_x$
$N = 100$; Lines are posterior using x_is (blue) or w_is (red)



$N = 400$



$N = 1600$

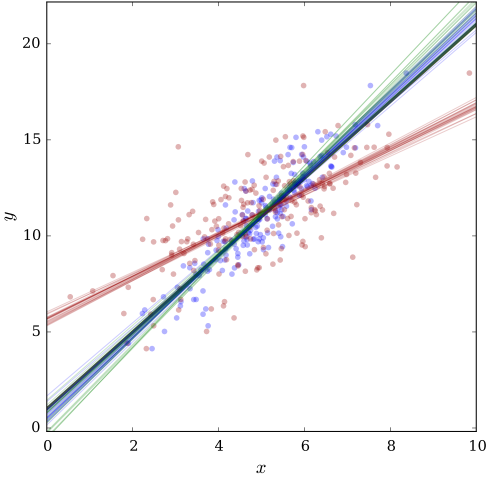


Latent variable model in Stan

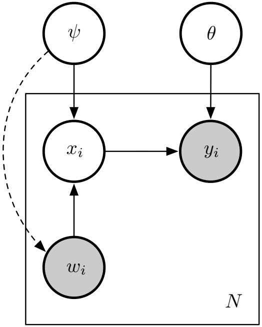
```
merr_code = """
data {
    int<lower=0> n; // number of samples
    real w[n]; // samples
    real y[n]; // samples
}

parameters {
    real beta_0;
    real beta_1;
    real x[n]; // latents
}

model {
    beta_0 ~ normal(0, 10.);  // prior is a wide normal
    beta_1 ~ normal(0, 10.);
    for (i in 1:n) {
        x[i] ~ normal(5., 1.);  // normal marginal for x
        w[i] ~ normal(x[i], %f);
        y[i] ~ normal(beta_0 + beta_1*x[i], 1.);
    }
}
""" % sig_err
```

Example with large measurement error: $\sigma_x = 1$, $\sigma_w = 1$

- Blue: Standard regression using (non-noisy) x_is
- Red: Standard regression using (noisy) w_is
- Green: Classical measurement error model using w_is



Berkson measurement error

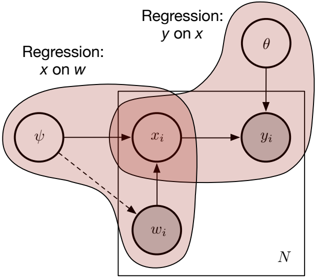


Prototype: Dose-response model

$$w_i \sim p(w_i|\psi) \quad \text{(dose)}$$
$$x_i \sim p(x_i|w_i, \psi)$$
$$x_i = g(w_i; \psi) + \delta_i; \quad \delta_i \sim \mathcal{N}(0, \sigma_x^2)$$
$$y_i = f(x_i; \theta) + \epsilon_i; \quad \epsilon_i \sim \mathcal{N}(0, \sigma_y^2)$$

Simplest: $x_i = w_i + \delta_i$,
e.g., received dose is not exactly administered dose

May have $\sigma_y \in \theta$ and/or $\sigma_x \in \psi$



Suppose we just want to learn θ, and ψ is *known* (or absent), so $p(w_i)$ and $p(x_i|w_i)$ are fixed (fully specified) dist'ns:

$$
\begin{aligned}
p(\theta|\{w_i, y_i\}) &\propto \pi(\theta) \prod_i \int dx_i \, p(w_i) \, p(x_i|w_i) \, p(y_i|x_i, \theta) \\
&\propto \pi(\theta) \left(\prod_i p(w_i) \right) \prod_i \int dx_i \, p(x_i|w_i) \, p(y_i|x_i, \theta)
\end{aligned}
$$

Here $p(x_i|w_i)$ *does* not *drop out*; we must know or infer it

But $p(w_i)$ **does** drop out (as long as $w_i \perp\!\!\!\perp \psi$)

Clustered data and mixed effects models

Examples of clustered data

- Subpopulations: Country → States → Counties
- Repeated measurements of related items
- Longitudinal studies of populations
- Meta-analysis—combining results of multiple studies

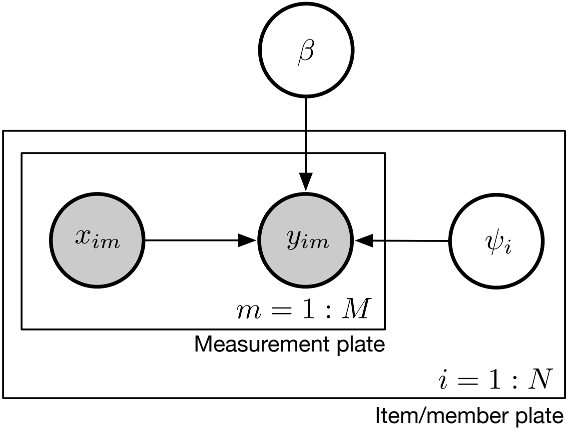
Mixed effects models

Model for measurement m (replication/time/study) on item i:

$$y_{im} = f(x_{im}; \beta) + g(x_{im}; \psi_i) + \epsilon_{im}$$

- β: *fixed (shared/bulk) effect*
  $f(x_{im})$ is like a shared *template*
- ψ_i: *random (individual/specific/peculiar) effect*
  $g(x_{im})$ is the individual departure from the template

DAG for mixed effects models



Mixed effects — Explicit shared/specific effects