Bayesian Statistics

Various Bayesian Models

Brani Vidakovic

Professor

School of Industrial and Systems Engineering

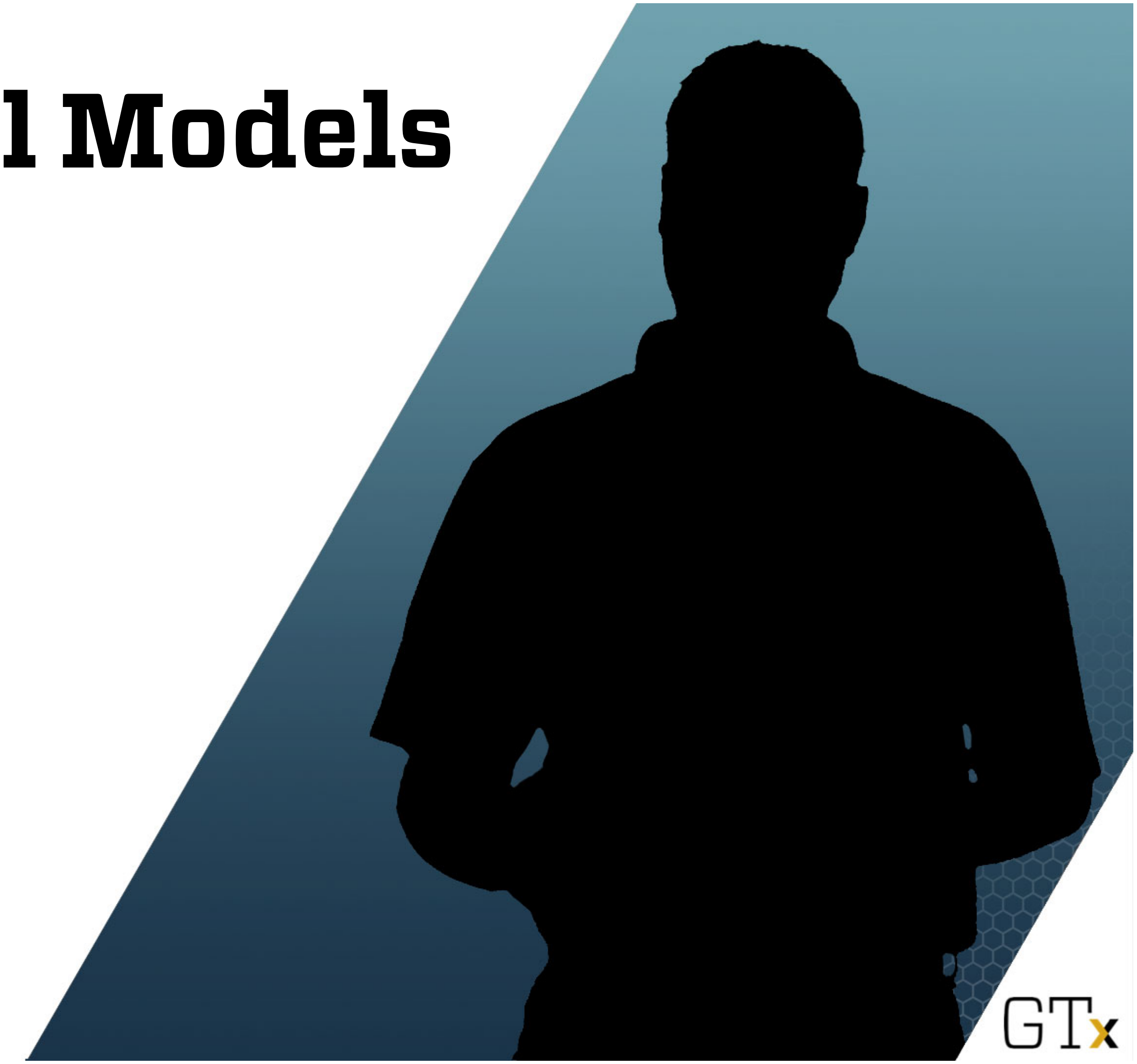
Before We Begin...

In this unit:

- Hierarchical Models
- Bayesian Linear Models
- Other Models

Hierarchical Models

- Why the hierarchy?
- Exchangeability
- Examples



Why the Hierarchy?

Prior $\pi(\theta) = \int \pi_1(\theta|\theta_1)\pi_2(\theta_1|\theta_2)\dots\pi_n(\theta_{n-1}|\theta_n)\pi_{n+1}(\theta_n)\mathrm{d}\theta_1\mathrm{d}\theta_2\dots\mathrm{d}\theta_n$

$$\left\{\begin{matrix} X|\theta \sim f(x|\theta) \\ \theta|\theta_1 \sim \pi_1(\theta|\theta_1) \\ \theta_1|\theta_2 \sim \pi_2(\theta_1|\theta_2) \\ \vdots \\ \theta_{n-1}|\theta_n \sim \pi_n(\theta_{n-1}|\theta_n) \\ \theta_n \sim \pi_{n+1}(\theta_n) \end{matrix}\right\} \Leftrightarrow \left\{\begin{matrix} X|\theta \sim f(x|\theta) \\ \theta \sim \pi(\theta) \end{matrix}\right\}$$

Why the Hierarchy?

$X \leftarrow \theta \leftarrow \theta_1 \leftarrow \theta_2 \leftarrow \cdots \theta_{n-1} \leftarrow \theta_n$

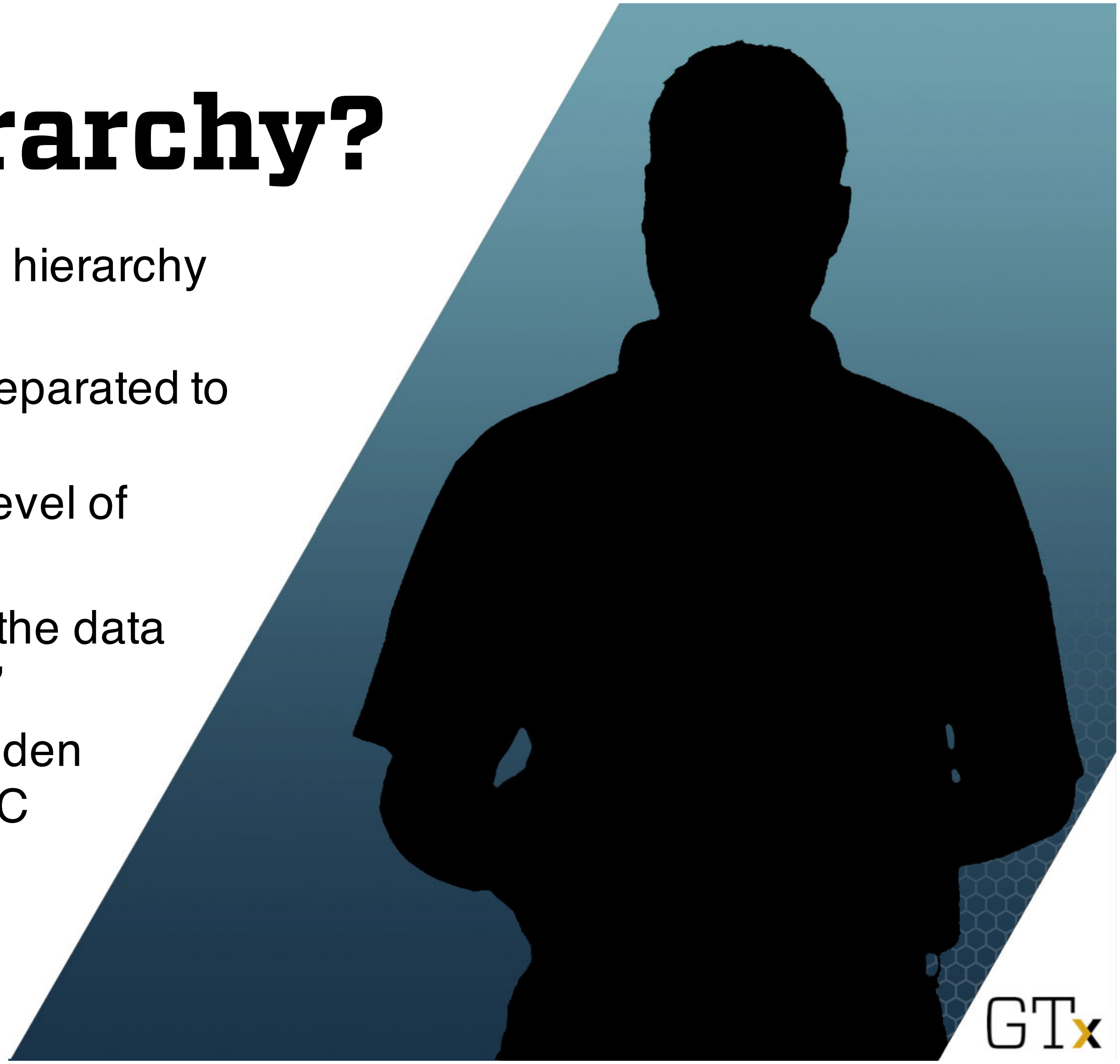
Joint distribution of $(X, \theta, \theta_1, \ldots, \theta_n)$

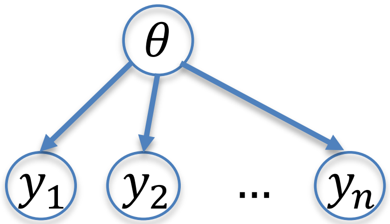
$$f(x, \theta, \theta_1, \ldots, \theta_n) \propto \begin{array}{c} f(x|\theta, \theta_1, \ldots, \theta_n)\times \\ \pi_1(\theta|\theta_1, \theta_2, \ldots, \theta_n)\times \\ \pi_2(\theta_1|\theta_2, \ldots, \theta_n)\times \\ \vdots \\ \pi_n(\theta_{n-1}|\theta_n)\times \\ \pi_{n+1}(\theta_n). \end{array}$$

$$f(x, \theta, \theta_1, \ldots, \theta_n) \propto f(x|\theta)\pi_1(\theta|\theta_1)\pi_2(\theta_1|\theta_2)\ldots\pi_n(\theta_{n-1}|\theta_n)\pi_{n+1}(\theta_n)$$

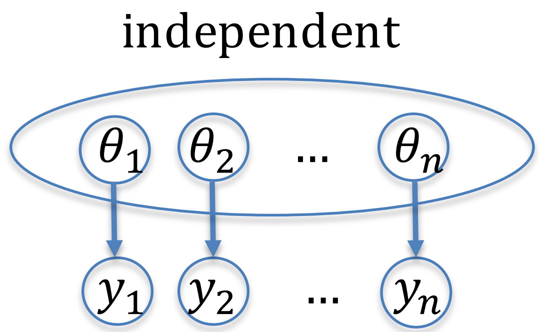
Why the Hierarchy?

- Modeling requirements ask for hierarchy (e.g. Bayesian Meta-Analysis)
- The prior information can be separated to structural part and subjective/ noninformative part at higher level of hierarchy
- Robustness + objectivity: “Let the data “talk” about hyperparameters.”
- Computing issues (utilizing hidden mixtures, mixture priors, MCMC efficiency, etc.)

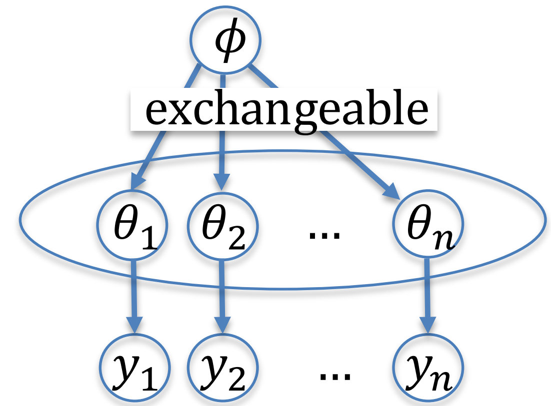




$$\prod_{i=1}^{n} f(y_i|\theta)\pi(\theta)$$



$$\prod_{i=1}^{n} f(y_i|\theta_i)\pi(\theta_i)$$



$$\pi(\phi)\prod_{i=1}^{n} f(y_i|\theta_i)\pi(\theta_i|\phi)$$

$\underline{Y_1, Y_2, \dots, Y_n}$ <u>exchangeable</u>

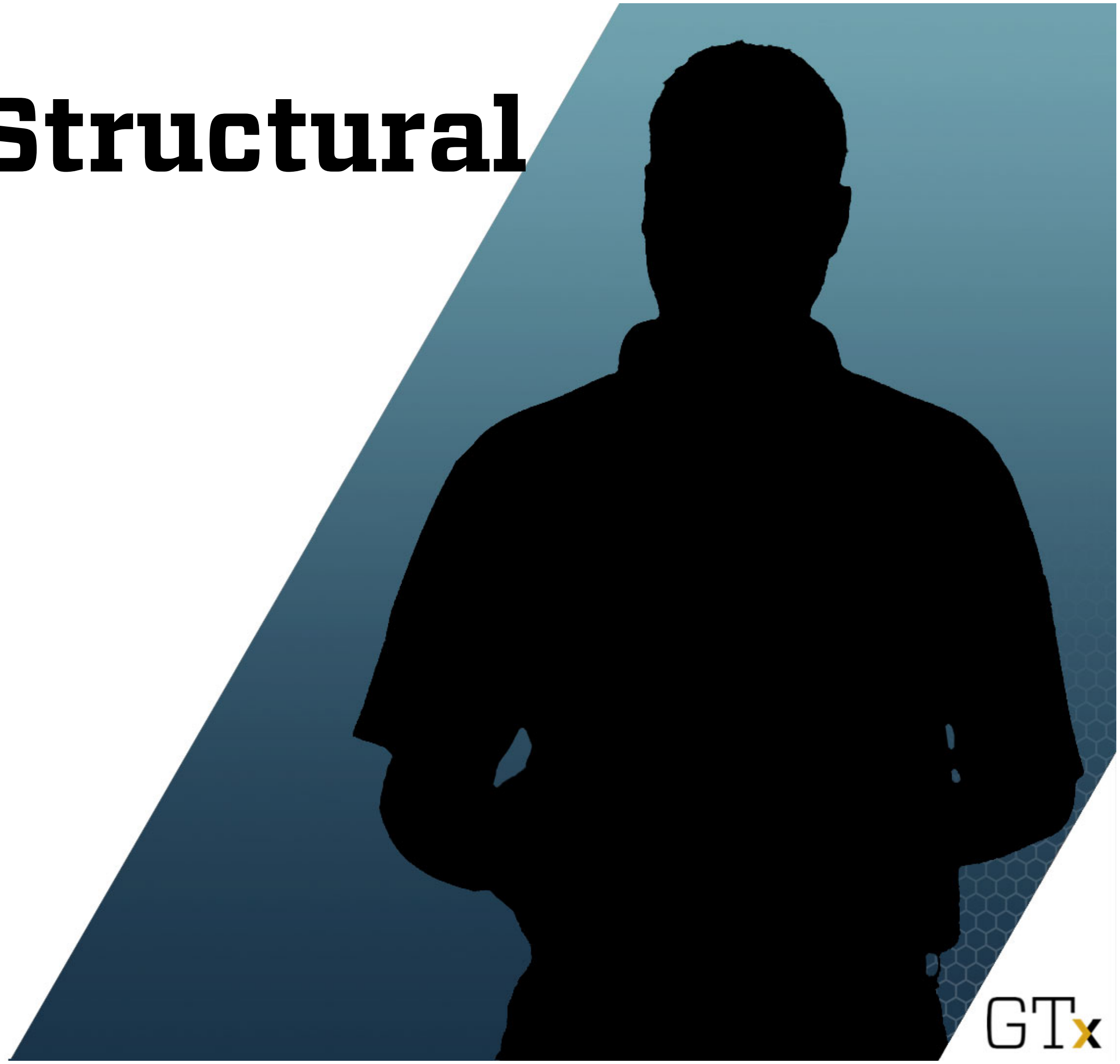
$\Leftrightarrow (Y_1, \dots, Y_n) \overset{d}{=} (Y_{\pi_1}, Y_{\pi_2}, \dots, Y_{\pi_n})$

where $(\pi_1, \dots, \pi_n)$ is any <u>permutation</u> of $(1,2,\dots,n)$

- $(X,Y) \sim \mathrm{MVN}_2\left(\underset{\sim}{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \ \rho \in (-1,1)$

are <u>exchangeable</u> but not <u>independent</u>

Priors with Structural Information



- **Example**:

  $X|p \sim \mathrm{Bin}(n, p)$

  $p|k \sim \mathrm{Beta}(k, k), \quad k \in \mathbb{N}$

  $\mathrm{k}|\mathrm{r} \sim \mathrm{Geom}(r), \ p(k = i) = (1 - r)^{i-1} r, \ i = 1,2, \ldots \quad 0 < r < 1$

  $r \sim \mathrm{Beta}(2,2)$

$$[p|k] \times [k|r] \times [r] \propto \frac{1}{\mathrm{B}(k,k)} p^{k-1} (1-p)^{k-1} (1-r)^{k-1} r \frac{1}{\mathrm{B}(2,2)} r(1-r),$$

$$[p|k] \times [k] \propto \frac{\mathrm{B}(3, k+1)}{\mathrm{B}(k,k)\mathrm{B}(2,2)} p^{k-1} (1-p)^{k-1}$$

$$[p] \propto \sum_{k=1}^{\infty} \frac{\mathrm{B}(3, k+1)}{\mathrm{B}(k,k)\mathrm{B}(2,2)} p^{k-1} (1-p)^{k-1} =$$

$$\frac{2p^4(4a - 15) - 4p^3(4a - 15) + 2p^2(11a - 25) - 2p(7a - 10) + (3a - 3)}{20p^4(1-p)^4},$$

$p \in (0,1) \quad a = \sqrt{(2p-1)^2}$

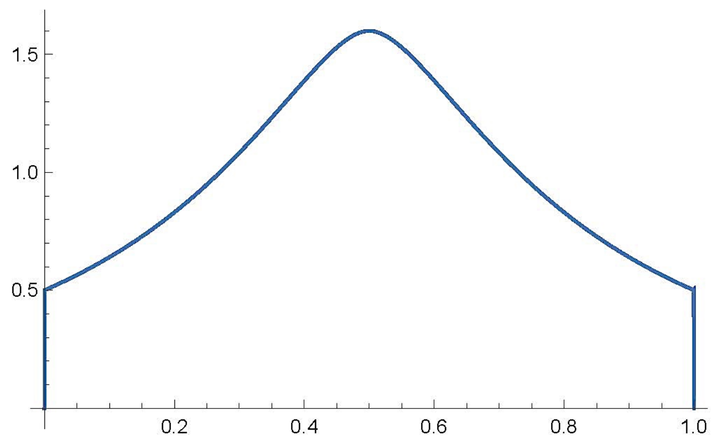
$X|p \sim \mathrm{Bin}(n,p)$: $\quad f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$

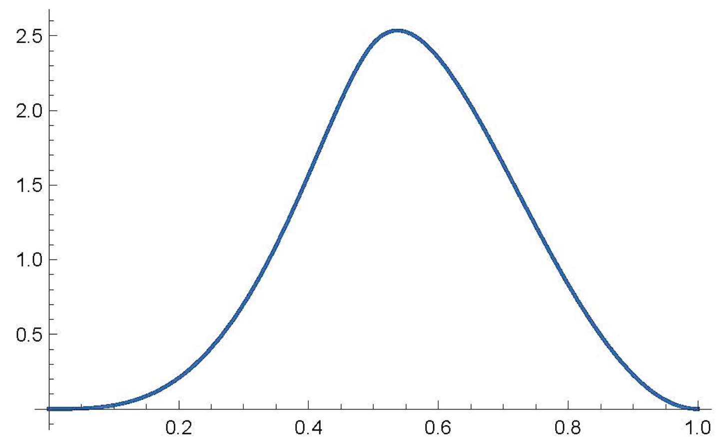
$p \sim \pi(p)$; **Assume** $\mathrm{X} = 3, n = 5$

$$\mathrm{E}^{\pi} p = \frac{1}{2} = 0.5$$

$$\hat{p} = \frac{X}{n} = \frac{3}{5} = 0.6$$

$$\mathrm{E}^{\pi(\cdot|X)} p = \frac{80 \log 2 - 55}{56 \log 2 - 38} = 0.553481$$





OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

basichierarhy

A SIMPLE HIERARCHICAL MODEL

Suppose that in a cohort of n people X of them have a particular condition. The likelihood is [X|p] ~ Bin(n,p), and population proportion p is the parameter of interest. You believe that the proportion is close to 1/2, but not quite sure. The appropriate prior on p would be Beta Be(k,k) for k positive integer; it is symmetric about k/(k + k)=1/2, Thus, [p|k] ~Be(k,k). You are reluctant tu specify k since this would specify the variance on p as 1/(8k +4) . Instead you place a hyperprior on k, as geometric with parameter r, and hyperprior on r as Be(2,2).

***** What is the Bayes estimator of p if n=5 and X=3?**

***** What are the Bayes estimators of k and r.**

```
model{
x ~ dbin(p, n)
p ~ dbeta(k,k)
k ~ dgeom(r)
 # This works for OpenBUGS only. For WinBUGS
 #    instead of k ~ dgeom(r) one needs
 #    k1 ~ dnegbin(r,1)
 #    k <- k1+1
r ~ dbeta(2,2)
}
```

Delete

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

basichierarhy

***** What is the Bayes estimator of p if n=5 and X=3?**
***** What are the Bayes estimators of k and r.**

```
model{
x ~ dbin(p, n)
p ~ dbeta(k,k)
k ~ dgeom(r)
 # This works for OpenBUGS only. For WinBUGS
 #    instead of k ~ dgeom(r) one needs
 #    k1 ~ dnegbin(r,1)
 #    k <- k1+1
r ~ dbeta(2,2)
}
```

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
k	3.713	9.717	0.02344	1.0	2.0	17.0	1001	1100000
p	0.5536	0.1546	1.527E-4	0.2488	0.5523	0.8514	1001	1100000
r	0.474	0.2259	4.038E-4	0.07948	0.4682	0.8965	1001	1100000

DATA

```
list(n=5, x=3)
```

INITS

```
list(r=0.5, k=2, p=0.5)
```

Delete

Priors as (hidden) Mixtures

Priors as Mixtures

t_n – prior is a scale mixture of normals

tasmixture

BUGS **t as a scale (precision) mixture of normals**

Suppose (1) mu ~ dt(mu0, tau, df) -- mu0 location, tau --precision, df - shape/degrees of freedom, and
(2) mu ~ dnorm(mu0, prec), prec ~ dgamma(a,b) for a=df/2 and b=df/(2*tau).

```
model{
X ~ dnorm(mu1, tau1)
mu1 ~ dt(mu0, tau, df)
#
Y ~ dnorm(mu2, tau1)
mu2 ~ dnorm(mu0, prec)
prec ~ dgamma(a,b)
a <- df/2
b <- df/(2 * tau)
}
```

DATA

```
list(X=10, Y=10, df =6, tau=0.4, tau1=10, mu0=6)
```

INITS

```
list(mu1=1, mu2=1, prec=1)
```

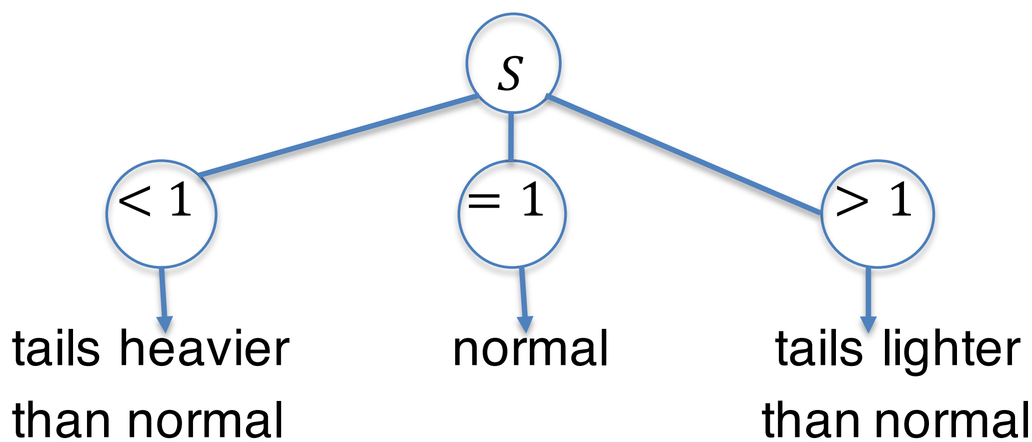
	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
mu1	9.91	0.3162	3.013E-4	9.29	9.91	10.53	1001	1000000
mu2	9.91	0.3161	3.143E-4	9.29	9.91	10.53	1001	1000000

- **Example:**

  Any symmetric unimodal distribution is a scale mixture of uniforms

$$y|\mu,\delta^2 \sim \mathrm{N}(\mu,\delta^2) \Leftrightarrow \begin{cases} y|\mu,\delta^2,d \sim \mathrm{U}\left(\mu-\sqrt{\delta^2 d},\mu+\sqrt{\delta^2 d}\right) \\ d \sim \mathrm{Ga}\left(\frac{3}{2},\frac{1}{2}\right) \end{cases}$$

$$d \sim \mathrm{Ga}\left(\frac{3}{2},\frac{S}{2}\right)$$



Jeremy's IQ

Suppose, as before $X \sim \mathrm{N}(\theta, \delta^2)$

$\theta \sim \mathrm{N}(\mu, \tau^2)$, δ^2, τ^2, μ **known**

$X = 98\,; \delta^2 = 80\,, \tau^2 = 120\,, \mu = 110$

$\Longrightarrow \hat{\theta} = 102.8$

$\theta \sim \mathrm{U}\big(\mu - \sqrt{\tau^2 d}\big), \mu + \sqrt{\tau^2 d}$

$d \sim \mathrm{Ga}\left(\frac{3}{2}, \frac{s}{2}\right)$

$s = 1,\ \theta \sim \mathrm{N}(\mu, \tau^2), \hat{\theta} = 102.8$

$s < 1,\ \theta \sim$ **heavy tailed**, $s = \frac{1}{2}, \hat{\theta} = 101.0$

$s > 1,\ \theta \sim$ **light tailed**, $s = 2, \hat{\theta} = 104.9$

OpenBUGS

File Edit Attributes Tools Info Model Inference Doodle Map PKBugs PKHelp Text Window Examples Manuals Help

Jeremymus

BUGS

Jeremy with Normal Prior as a Scale Mixture of Uniforms

Any Symmetric Unimodal distribution is a Scale Mixture of Uniforms

theta|mu, tau^2 ~ N(mu, tau^2)
is the same as
theta|mu, tau^2, d ~ U(mu - tau * d^0.5, mu + tau * d^0.5)
d ~ Ga(3/2, 1/2)
If
d ~ Ga(3/2, s/2), s<1 prior is heavy tails,
for s>1 light tails. Heavy/light compared to normal.

Jeremy: y=98, sigma^2=80, tau^2=120, mu=110

```
model{
y  ~ dnorm(theta, precy)   #precy = 1/sigma^2 = 1/80
lb <- mu - sqrt(d/precth)    #precth = 1/tau^2 = 1/120
ub <- mu + sqrt(d/precth)
theta ~ dunif(lb, ub)
#s <- 1         #normal
#s <- 1/2      #tails heavier than normal
s <- 2         #tails lighter than normal

beta <- s/2
d ~ dgamma(1.5, beta)
}
```

OpenBUGS version 3.2.3 rev 1012

Jeremymus

DATA
list(y = 98, precy = 0.0125, precth = 0.0083333, mu = 110)

INITS
list(theta=100, d=2)

s=1

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
d	2.828	2.233	0.008332	0.2259	2.276	8.576	1001	300000
theta	102.8	6.928	0.01805	89.27	102.8	116.4	1001	300000

s = 1/2 (heavy tail)

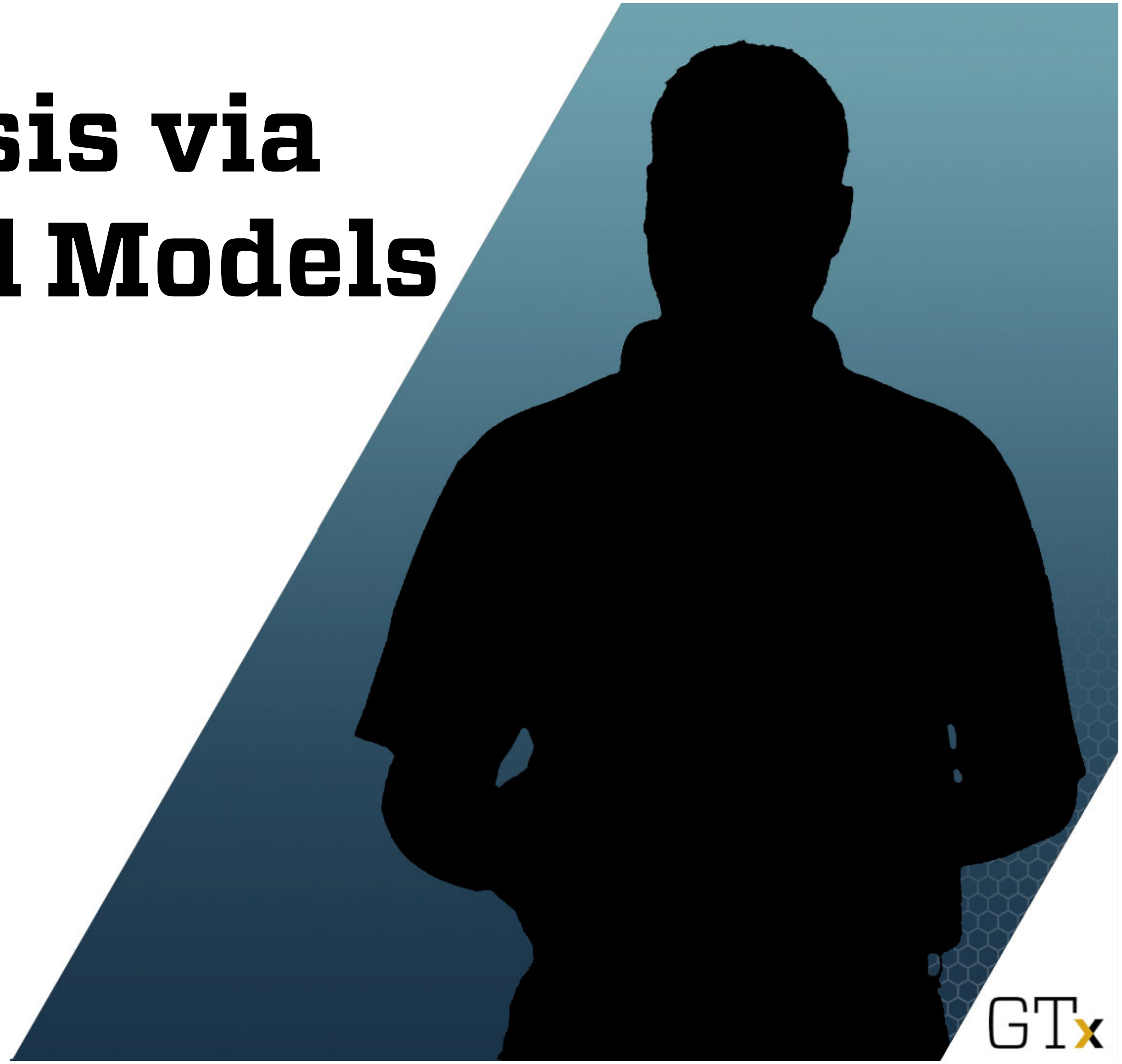
	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
d	5.163	4.196	0.01406	0.4123	4.074	16.15	1001	300000
theta	101.0	7.757	0.01756	85.81	101.0	116.2	1001	300000

s=2 (light tail)

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
d	1.509	1.191	0.004587	0.117	1.217	4.571	1001	300000
theta	104.9	5.876	0.01513	93.37	104.8	116.4	1001	300000

Meta-analysis via Hierarchical Models

Rats Example

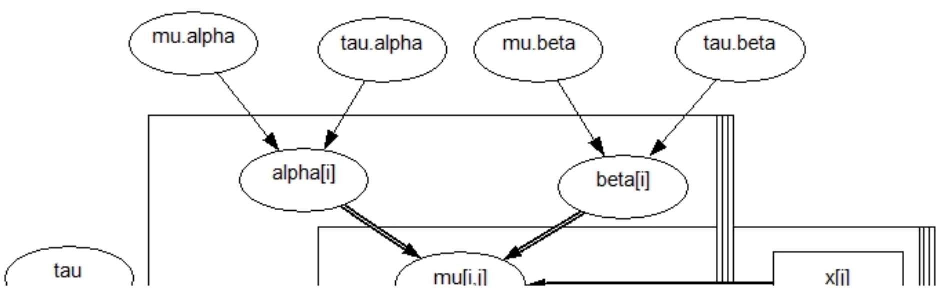


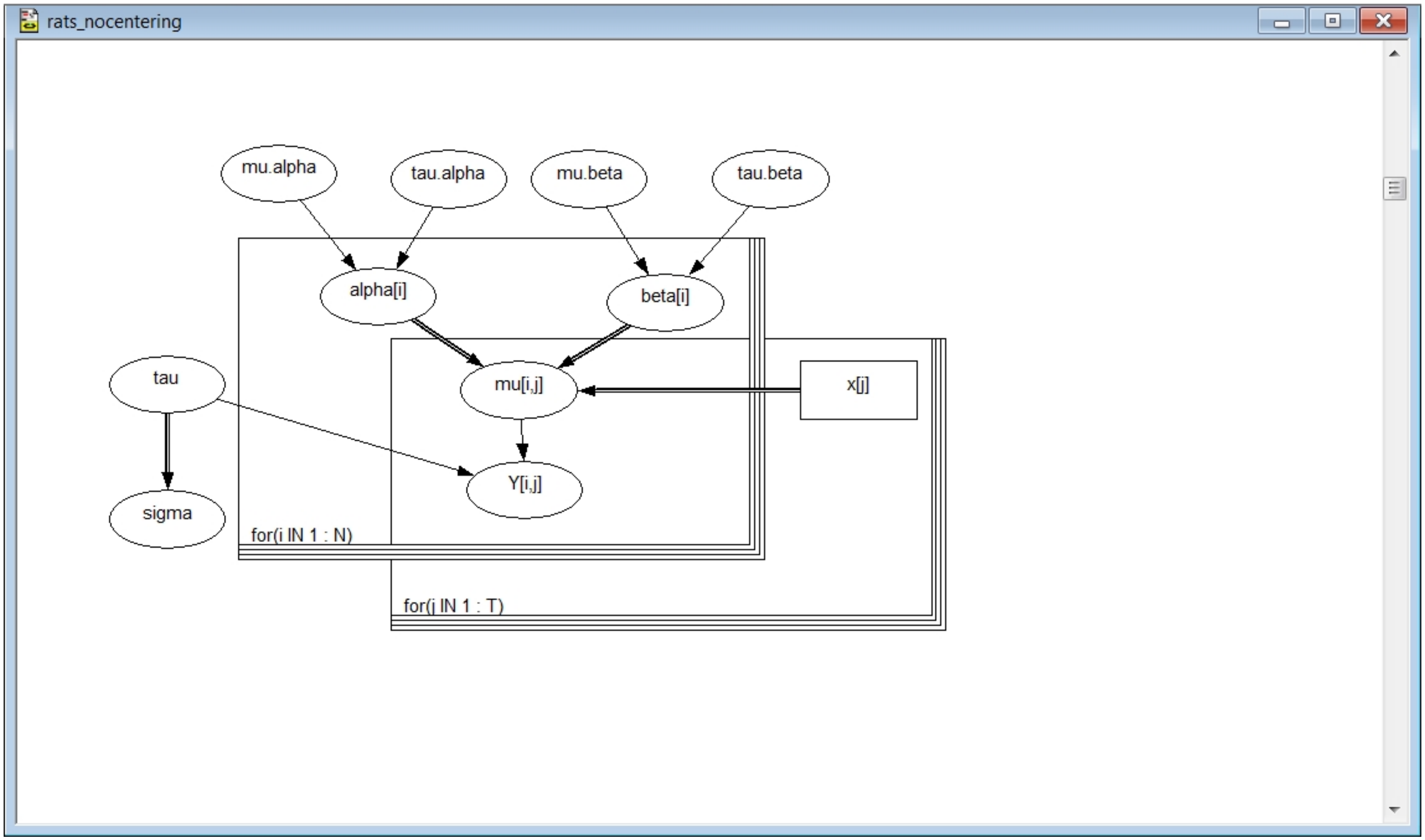
rats_nocentering

RATS

This example is taken from Gelfand *et al* (1990), and concerns 30 young rats whose weights were measured weekly for five weeks. Part of the data is shown below, where Y_{ij} is the weight of the ith rat measured at age x_j.

	Weights Y_{ij} of rat i on day x_j				
	$x_j = 8$	15	22	29	36
Rat 1	151	199	246	283	320
Rat 2	145	199	249	293	354
....					
Rat 30	153	200	244	286	324




rats_nocentering
mu.alpha
tau.alpha
mu.beta
tau.beta
alpha[i]
beta[i]
tau
mu[i,j]
x[j]
Y[i,j]
sigma
for(i IN 1 : N)
for(j IN 1 : T)

rats_nocentering

```
model
{
for( i in 1 : N ) {
for( j in 1 : T ) {
Y[i , j] ~ dnorm(mu[i , j],tau)
mu[i , j] <- alpha[i] + beta[i] *  x[j]
}
alpha[i] ~ dnorm(mu.alpha, tau.alpha)
beta[i] ~ dnorm(mu.beta,tau.beta)
}
tau ~ dgamma(0.001,0.001)
sigma <- 1 / sqrt(tau)
mu.alpha  ~ dnorm(0.0,1.0E-6)
tau.alpha ~ dgamma(0.001,0.001)
mu.beta  ~ dnorm(0.0,1.0E-6)
tau.beta  ~ dgamma(0.001,0.001)
}
```

➔ **DATA** ←

|

INITS

```
list(alpha = c(100, 100, 100, 100, 100,   100, 100, 100, 100, 100,   100, 100, 100, 100, 100,
        100, 100, 100, 100, 100,   100, 100, 100, 100, 100,   100, 100, 100, 100, 100),
beta  = c(6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6),
mu.alpha = 100, mu.beta = 6,  tau = 1, tau.alpha = 1, tau.beta = 1)
```

rats_nocentering

```
⇨

list(x = c(8.0, 15.0, 22.0, 29.0, 36.0),  N = 30, T = 5,
                    Y = structure(
        .Data =   c(151, 199, 246, 283, 320,
                      145, 199, 249, 293, 354,
                      147, 214, 263, 312, 328,
                      155, 200, 237, 272, 297,
                      135, 188, 230, 280, 323,
                      159, 210, 252, 298, 331,
                      141, 189, 231, 275, 305,
                      159, 201, 248, 297, 338,
                      177, 236, 285, 350, 376,
                      134, 182, 220, 260, 296,
                      160, 208, 261, 313, 352,
                      143, 188, 220, 273, 314,
                      154, 200, 244, 289, 325,
                      171, 221, 270, 326, 358,
                      163, 216, 242, 281, 312,
                      160, 207, 248, 288, 324,
                      142, 187, 234, 280, 316,
                      156, 203, 243, 283, 317,
                      157, 212, 259, 307, 336,
                      152, 203, 246, 286, 321,
                      154, 205, 253, 298, 334,
                      139, 190, 225, 267, 302,
                      146, 191, 229, 272, 302,
                      157, 211, 250, 285, 323,
                      132, 185, 237, 286, 331,
                      160, 207, 257, 303, 345,
                      169, 216, 261, 295, 333
```

rats_nocentering

```
139, 190, 225, 267, 302,
146, 191, 229, 272, 302,
157, 211, 250, 285, 323,
132, 185, 237, 286, 331,
160, 207, 257, 303, 345,
169, 216, 261, 295, 333,
157, 205, 248, 289, 316,
137, 180, 219, 258, 291,
153, 200, 244, 286, 324),
.Dim = c(30,5)))
```

INITS

```
list(alpha = c(100, 100, 100, 100, 100,   100, 100, 100, 100, 100,   100, 100, 100, 100, 100,
         100, 100, 100, 100, 100,   100, 100, 100, 100, 100,   100, 100, 100, 100, 100),
beta  = c(6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6),
mu.alpha = 100, mu.beta = 6,  tau = 1, tau.alpha = 1, tau.beta = 1)
```

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
mu.alpha	106.6	2.308	0.01523	102.0	106.6	111.1	1001	100000
mu.beta	6.185	0.1056	6.625E-4	5.977	6.185	6.392	1001	100000
sigma	6.15	0.4677	0.003131	5.314	6.121	7.148	1001	100000
tau	0.0269	0.004047	2.674E-5	0.01957	0.02669	0.03541	1001	100000
tau.alpha	0.01013	0.00406	3.91E-5	0.004623	0.009401	0.01995	1001	100000
tau.beta	4.328	1.495	0.01078	2.104	4.101	7.882	1001	100000