Multiparameter models

- Usually, we have models with many parameters, let's start with k=2.
  - **$\pi(\theta_1, \theta_2 | X) = \pi(X | \theta_1, \theta_2) \pi(\theta_1, \theta_2)$/const**
  - **$\pi(\theta_1, \theta_2)$ is joint prior. Often used:** $\pi(\theta_1)\, \pi(\theta_2)$ independent priors for each parameter.
  - **Prior could also be hierarchical:** $\pi(\theta_1 | \theta_2)\, \pi(\theta_2)$
  - $\pi(X | \theta_1, \theta_2)$ **could be e.g. $N(\mu,\sigma^2)$-model for X.**

  - **Marginal posterior density usually of interest:**

    $\pi(\theta_1 | X) = \int \pi(\theta_1, \theta_2 | X)\, d\theta_2$

    $= \int \pi(\theta_1 | \theta_2, X)\, \pi(\theta_2 | X)\, d\theta_2$

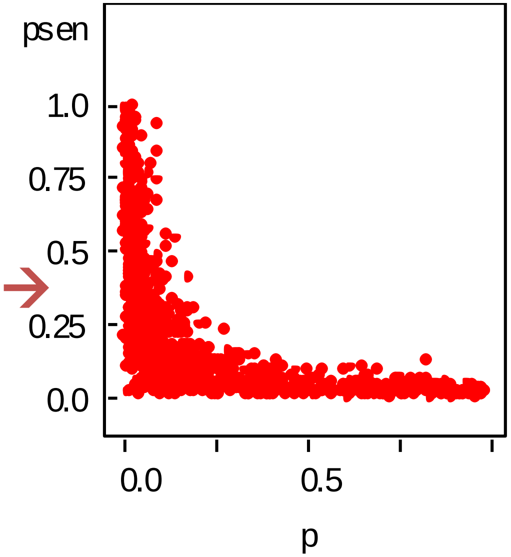
Multiparameter models

- The parameter of interest can be θ_1 while θ_2 is just a nuisance parameter.
  - **Example: diagnostic testing with sensitivity θ_2 :**
    X ~ Bin(N, $\theta_1 * \theta_2$)
  - Here, θ_1 is the unknown true prevalence, θ_2 is the unknown test sensitivity – for which we could have an informative prior, though.
  - We should take into account the uncertainty of both parameters jointly, given the data (and prior).
  - $\pi(\theta_1, \theta_2 \mid X) = \mathrm{Bin}(X \mid N, \theta_1\theta_2)\, \pi(\theta_1, \theta_2)$ /const

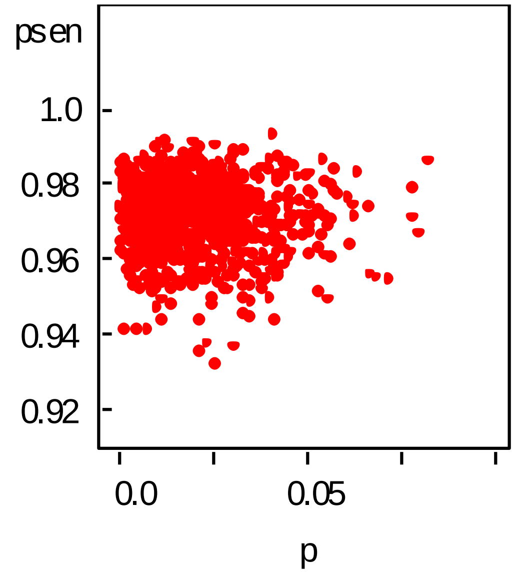
…Solving posterior becomes difficult, therefore try BUGS…

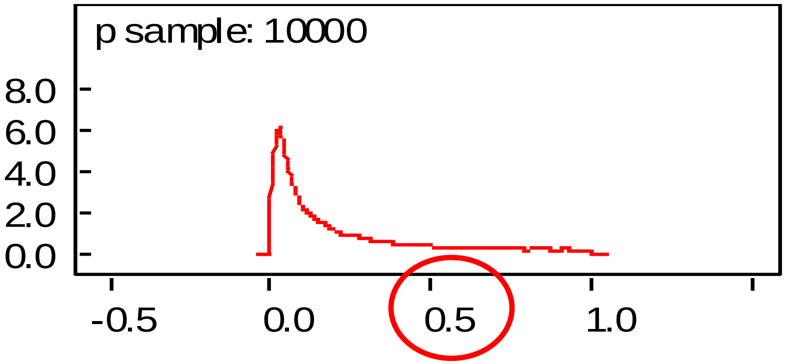
- **Assume we observed N=100, X=1.**

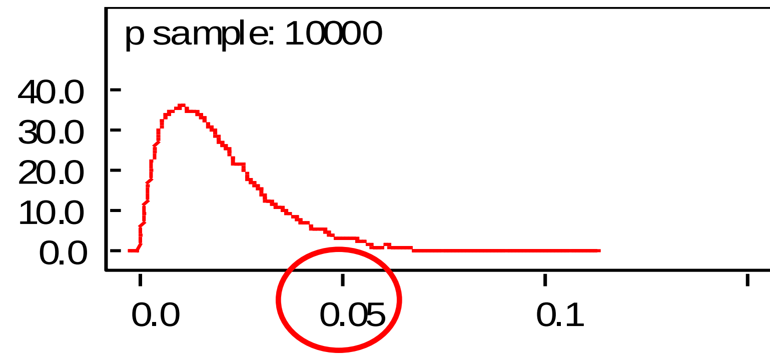
Without any prior knowledge of test sensitivity →



Assuming sensitivity is average 0.97, SD 0.01 →



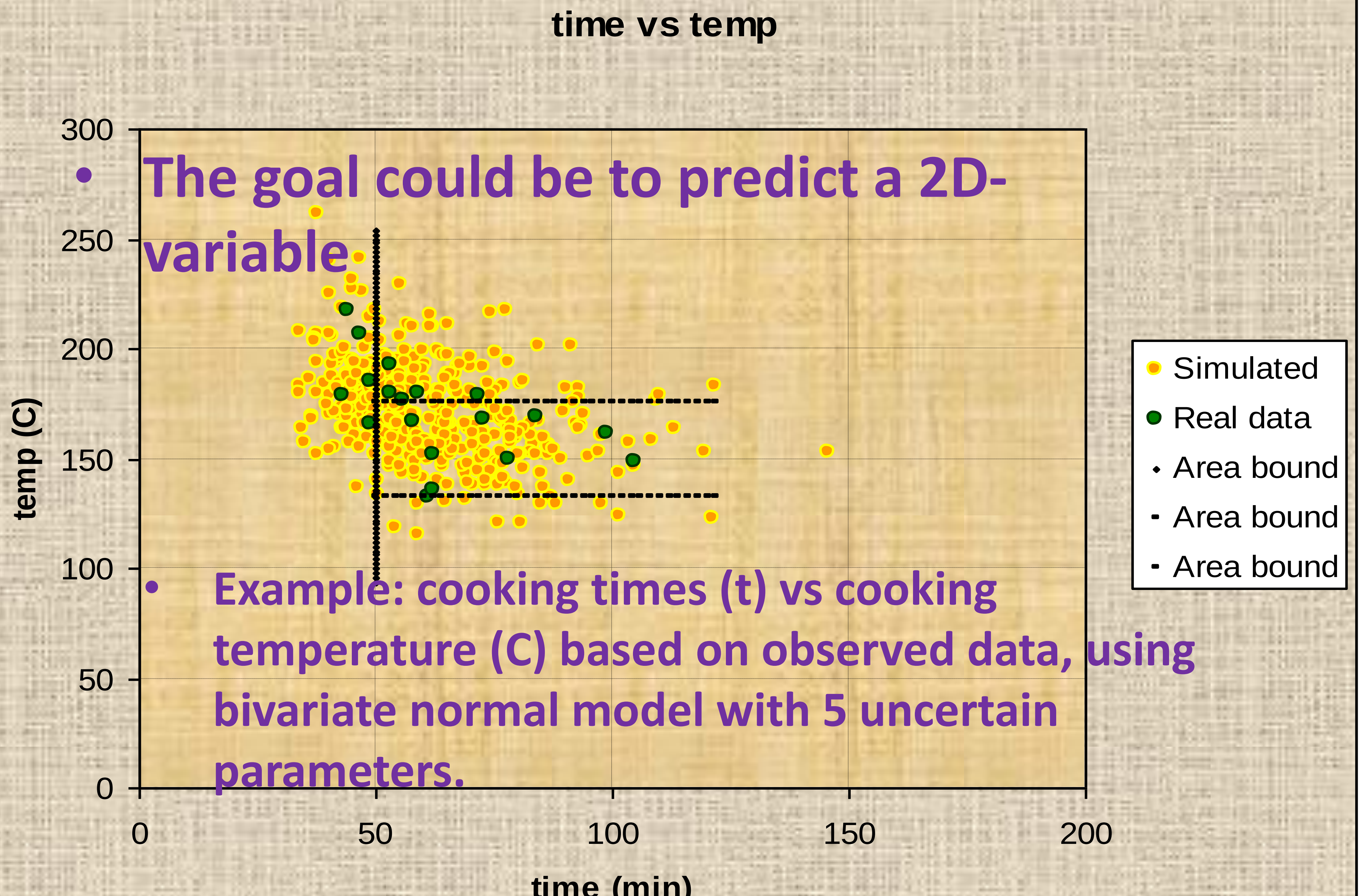




Multiparameter models

- The aim could also be to predict a multivariate response (as in correlated data models)
  - This requires several parameters in the model. $\pi(X_1,X_2 \mid \theta_1,\ldots,\theta_k)$
  - Posterior prediction $\pi(X_1^*,X_2^*|X_1,X_2)$ requires integration over all parameters
  - Then some more integration to get marginal predictive distributions $\pi(X_1^* \mid X_1,X_2)= \int\pi(X_1^*,X_2^*|X_1,X_2)dX_2^*$

- **The goal could be to predict a 2D-variable**



- **Example: cooking times (t) vs cooking temperature (C) based on observed data, using bivariate normal model with 5 uncertain parameters.**
- **Compute predictive distribution $\pi(t,C \mid data)$**

Identifiability and multiparameter models

- Parameters are unidentifiable (from data) if $P(X \mid \theta_1) = P(X \mid \theta_2)$, with $\theta_1 \neq \theta_2$
- Posterior result then depends solely on prior.
- Example: $X \sim N(\theta_1 + \theta_2, 1)$
  - All combinations with $\theta_1 + \theta_2 = c$ are equally probable, unless prior can make a difference.
  - Is the posterior a proper density?

  - Multiparameter models with insufficient data may lead to problems of identifiability. Useful to examine the likelihood function.

Multinomial model

- E.g. large bag of balls of k different colors. Pick N balls (with replacement)
- **$X_1,\ldots,X_k$ = number of balls of each color.**
- **$X_1+,\ldots,+X_k = N$**
- Vector X is multinomially distributed, given the true proportions $\theta_1,\ldots,\theta_k$.
- **Find out $\pi(\theta_1,\ldots,\theta_k \mid X)$**

Conjugate prior is possible!

Multinomial model

- This is a generalization of earlier inference problem with Binomial & Beta
- $\pi(\theta_1,\ldots,\theta_k) = \text{Dirichlet}(\alpha_1,\ldots,\alpha_k)$
- $\Sigma\theta_i = 1$
- Thanks to conjugate prior:

$$\pi(\theta_1,\ldots,\theta_k \mid X) = \text{Dirichlet}(\alpha_1+X_1,\ldots,\alpha_k+X_k)$$

- **Marginal densities easy, if $\pi(\theta \mid X)=\text{Dir}(\alpha)$, then**

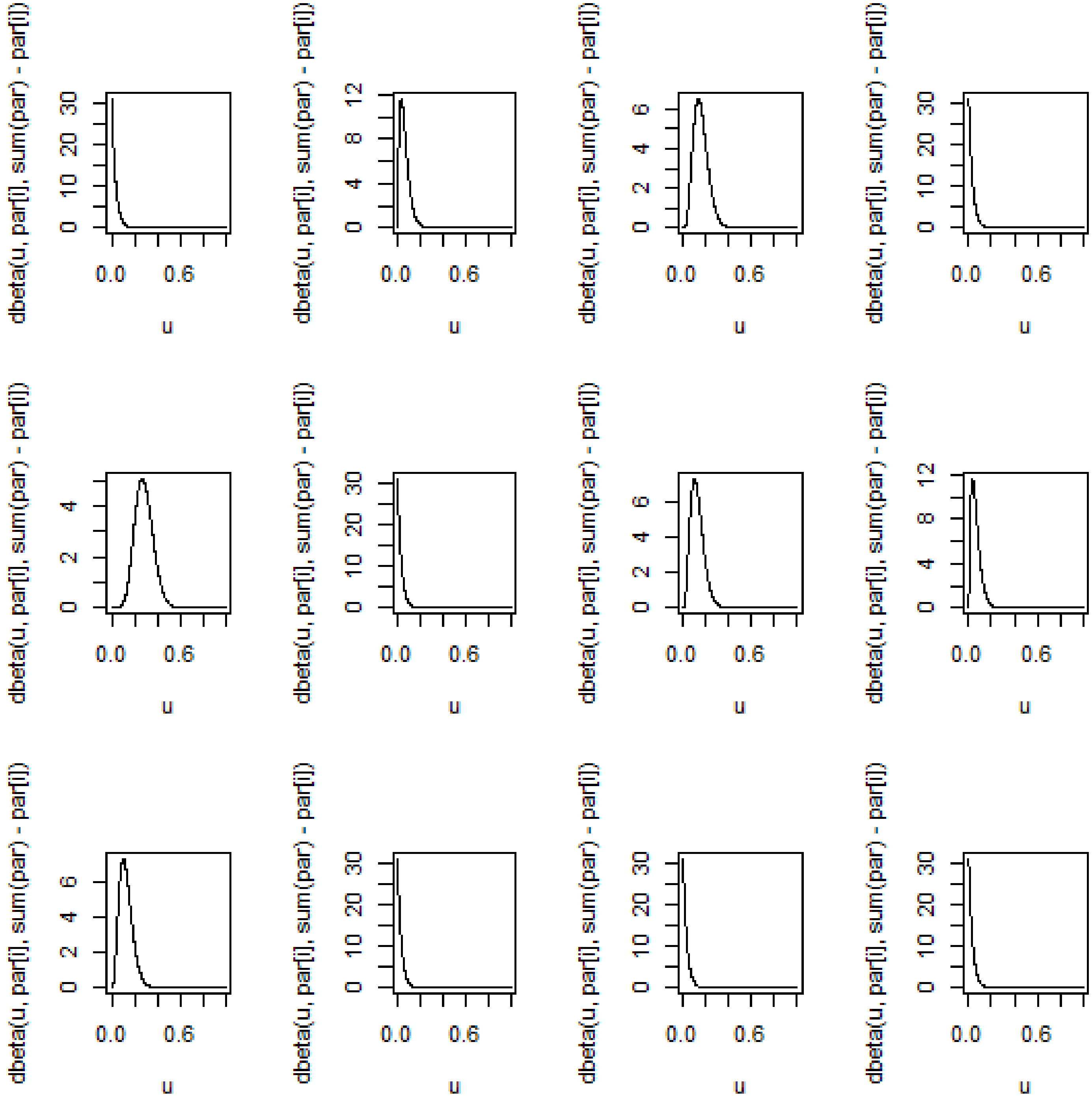
$$\pi(\theta_i \mid X) = \text{Beta}(\alpha_i, \Sigma\alpha_j - \alpha_i)$$

Multinomial model

- Example: there are 12 subtypes of bacteria. In a sample of 20, we observed the following numbers of each type:
- X=(0,1,4,0,8,0,3,1,3,0,0,0)
- $\boldsymbol{\pi(\theta_1,\ldots,\theta_k\,|X) = \mathrm{Dir}(\alpha_1+X_1,\ldots,\alpha_k+X_k) =}$

$\mathrm{Dir}(\alpha_1+0,\ \alpha_2+1,\ \alpha_3+4,\ \alpha_4+0,\ \alpha_5+8,\ \alpha_6+0,\ \alpha_7+3,\ \alpha_8+1,\ \alpha_9+3,\ \alpha_{10}+0,\ \alpha_{11}+0,\ \alpha_{12}+0)$

- Note the 'prior data sample' n=12 in the Dir(1,…,1) prior.


dbeta(u, par[i], sum(par) - par[i])
u
0.0
0.6
0
10
20
30
4
8
12
2
6

Normal model N(X|μ,σ)

- **Take a look at the easy cases first:**
- **π(μ|X,σ) and π(σ|X,μ)**
- Convenient notation: precision $\tau=1/\sigma^2$ this parameterization is also used in BUGS with normal densities.
- Conjugate prior for μ is $N(\mu_0,\sigma_0)$

$$\pi(\mu \mid \mu_0, \tau_0) = \exp(-0.5\tau_0(\mu - \mu_0)^2)/c$$

- Assume first a single observation X_i:

$$\pi(X_i \mid \mu, \tau) = \exp(-0.5\tau(X_i - \mu)^2)/c$$

Normal model N(X|μ,σ)

- **Posterior for μ is then**

$$\pi(\mu \mid X_i, \tau, \mu_0, \tau_0) = \exp(-0.5(\tau_0(\mu-\mu_0)^2 + \tau(X_i-\mu)^2))/c$$

$$= N\left(\frac{n_0\mu_0 + X_i}{n_0+1}, \frac{\sigma^2}{n_0+1}\right)$$

- Use ’completing a square’ –technique.
- Here $n_0 = \tau_0/\tau$ can be interpreted as ’prior sample size’.
- **Posterior mean is: $w\mu_0 + (1-w)X_i$, with: $w=\tau_0/(\tau_0+\tau)$**

Normal model N(X|$\boldsymbol{\mu}$,$\boldsymbol{\sigma}$)

- With several measurements $X_1,\ldots,X_N$, we can write the likelihood as (using sufficient statistics)

$$\pi(\overline{X} \mid \mu, \sigma) = N(\overline{X} \mid \mu, \sigma^2 / N)$$

- Similar to previous example, the posterior of μ is

$$N\left(\frac{n_0\mu_0 + \overline{X}}{n_0 + 1}, \frac{\sigma^2 / N}{n_0 + 1}\right)$$

- Here $n_0 = \tau_0 / (N\tau)$

Normal model N(X|μ,σ)

- Posterior mean and variance can also be expressed as

$$E(\mu \mid \overline{X}) = \frac{\dfrac{\mu_0}{\sigma_0^2} + \dfrac{N\overline{X}}{\sigma^2}}{\dfrac{1}{\sigma_0^2} + \dfrac{N}{\sigma^2}} \qquad V(\mu \mid \overline{X}) = \frac{1}{\dfrac{1}{\sigma_0^2} + \dfrac{N}{\sigma^2}}$$

- Note what happens when N→0, or N→∞ ?

Normal model N(X|**μ,σ**)

- **Another possibility: improper prior** $\pi(\mu) \propto 1$
- The posterior is proper density, and

$$\pi(\mu \mid \bar{X}) = N(\bar{X}, \sigma^2 / N)$$

- Compare with non-bayesian statistics, where the inference is based on

$$\pi(\bar{X} \mid \mu) = N(\mu, \sigma^2 / N)$$

- These are like mirror images…

Normal model N(X|**μ,σ**)

- **$\pi(\sigma|X,\mu)$?**
- Assume observations $X_1,\ldots,X_N$, set $\tau=\sigma^{-2}$

$$\pi(X \mid \mu,\sigma) \propto \sigma^{-N} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i-\mu)^2)$$

$$= (\sigma^2)^{-N/2} \exp(-\frac{N}{2\sigma^2} s_0^2) \qquad = \qquad \boxed{\tau^{N/2} \exp(-\frac{N\tau}{2} s_0^2)}$$

- Here $s_0^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2$
- Conjugate prior for τ? ….gamma(α,β)

Normal model N(X|$\boldsymbol{\mu}$,$\boldsymbol{\sigma}$)

- Following from Bayes theorem, the posterior $\pi(\tau|X,\mu)$ is proportional to

$$\tau^{N/2}\exp(-\frac{N\tau}{2}s_0^2)\times\tau^{\alpha-1}\exp(-\beta\tau)$$

$$=\tau^{N/2+\alpha-1}\exp(-(\frac{N}{2}s_0^2+\beta)\tau)$$

- This is recognized to be gamma($N/2+\alpha, Ns_0^2/2+\beta$)
- Uninformative prior $\alpha \to 0$, $\beta \to 0$.

Normal model N(X|$\boldsymbol{\mu},\boldsymbol{\sigma}$)

- **Joint density $\pi(\mu,\sigma|X)$?**
- Assume observations $X_1,\ldots,X_N$
- Several options:

1. conjugate 2D prior $\pi(\mu,\sigma)=\pi(\mu|\sigma)\pi(\sigma)$
2. independent priors $\pi(\mu,\sigma)=\pi(\mu)\pi(\sigma)$
3. improper prior $\pi(\mu,\tau)\propto 1/\tau$

more challenging to solve posterior…

Normal model $N(X|\mu,\sigma)$

- **Difficulties**:

1. conjugate 2D prior $\pi(\mu,\sigma)=\pi(\mu|\sigma)\pi(\sigma)$

Not very practical to express prior of μ, conditionally on σ.

This would be:

$\tau \sim$ Gamma(a/2,b/2) and

$\mu \mid \tau \sim N(\mu_0, 1/\omega_0\tau)$, known as

'normal-gamma', or 'normal inv-gamma' when parameter is σ^2.

Normal model $N(X|\boldsymbol{\mu},\boldsymbol{\sigma})$

- **Difficulties**:

2. independent priors $\pi(\mu)$, $\pi(\sigma)$

Not possible to choose so that posterior could be solved in any familiar form.

Normal model N(X|**μ,σ**)

- **Difficulties**:

3. Improper prior $\pi(\mu,\tau) \propto 1/\tau$

same as $\pi(\mu,\sigma^2) \propto 1/\sigma^2$

same as $\pi(\mu,\log(\sigma)) \propto 1$

Posterior can be solved by factorization

$\pi(\mu,\sigma^2|X) = \pi(\mu|\sigma^2,X)\pi(\sigma^2|X)$

…we already have solved the first part before.

Normal model $N(X|\boldsymbol{\mu},\boldsymbol{\sigma})$

- The second part is $\pi(\sigma^2|X)$

= Scaled-Inverse-χ^2(n-1,s)

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2$$

- Or: $\pi(\tau|X)=\pi(1/\sigma^2|X)$

= Gamma((n-1)/2,(n-1)s^2/2)

- Full joint density can thus be written as a *product of two known densities.*
  - Convenient for Monte Carlo simulations. (draw σ^2, then μ conditionally on σ^2)
  - Also, can solve $\pi(\sigma^2\,|\mu,X)$, useful for Gibbs sampling.

Working out $\pi(\sigma^2|X)$

- First, write $\pi(\mu,\sigma^2 | X_1,\ldots,X_n)$ in the form:

prior $\pi(\sigma^2)$ **likelihood**

$$\pi(\mu,\sigma^2 \mid X) \propto \sigma^{-2}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i-\mu)^2\right)$$

Manipulation

$$= \sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{X}-\mu)^2]\right)$$

where: $s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})^2$

- Then, integrate over μ to get marginal density for σ^2.

Working out $\pi(\sigma^2|X)$

- Solving $\pi(\sigma^2|X)$: integrate the joint density $\pi(\sigma^2,\mu|X)$ over μ.

$$\pi(\sigma^2 \mid X) \propto \int_{-\infty}^{\infty} \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{X}-\mu)^2]\right) d\mu$$

$$= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \times \int_{-\infty}^{\infty} \exp\left(-\frac{n}{2\sigma^2}(\bar{X}-\mu)^2\right) d\mu$$

$$= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \times \sqrt{2\pi\sigma^2 / n}$$

$$\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)$$

= Scaled-Inverse-χ^2(n-1,s)-distribution.

For $\tau=1/\sigma^2$: this is Gamma((n-1)/2,(n-1)s^2/2)

Working out $\pi(\sigma^2 | X)$

- That required a few steps and manipulations…
- The lesson was to:
  - See what kind of tricks and techniques are needed for exact solutions.
  - See why and how the seemingly simple principle of Bayes theorem leads to increasingly complicated math which has been a major obstacle in practical Bayesian applications in the past.
  - See usefulness of Monte Carlo methods and WinBUGS/OpenBUGS in practical computations.