Multiparameter Models - Normal Data

Statistics 220

Spring 2005

Normal Inference Models

Most analyzes we wish to perform involve multiple parameters

- $y_i \overset{iid}{\sim} N(\mu, \sigma^2)$

- Multiple Regression: $y_i | x_i \overset{ind}{\sim} N(x_i^t \beta, \sigma^2)$

- Logistic Regression: $y_i | x_i \overset{ind}{\sim} Bern(p_i)$ where $\mathrm{logit}(p_i) = \beta_0 + \beta_1 x_i$

In these cases we want to assume all of the parameters are unknown and want to perform inference on some or all of them.

An example of the case, where only some of them may be of interest is multiple regression. Usually only the regression parameters β are of interest. The measurement variance σ^2 is often considered as a nuisance parameter.

Lets consider the case with two parameters θ_1 and θ_2 and that only θ_1 is of interest. An example of this would be $N(\mu, \sigma^2)$ data where $\theta_1 = \mu$ and $\theta_2 = \sigma^2$.

Want to base our inference on $p(\theta_1|y)$. We can get at this a couple of ways. First we can start with the joint posterior

$$p(\theta_1, \theta_2|y) \propto p(y|\theta_1, \theta_2)p(\theta_1, \theta_2)$$

This gives

$$p(\theta_1|y) = \int p(\theta_1, \theta_2|y)d\theta_2$$

We can also get it by

$$p(\theta_1|y) = \int p(\theta_1|\theta_2, y)p(\theta_2|y)d\theta_2$$

This implies that distribution of θ_1 can be considered a mixture of the conditional distributions, averaged over the nuisance parameter.

Note that this marginal conditional distribution is often difficult to determine explicitly. Normally it needs to be examined by Monte Carlo methods.

Example: Normal Data

$$y_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$$

For a prior, lets assume that μ and σ^2 are independent and use the standard non-informative priors

$$p(\mu, \sigma^2) = p(\mu)p(\sigma^2) \propto \frac{1}{\sigma^2}$$

So the joint posterior satisfies

$$
\begin{aligned}
p(\mu, \sigma^2) &\propto \frac{1}{\sigma^2} \prod_{i=1}^{n} \frac{1}{\sigma} \exp\left(-\frac{1}{2\sigma^2} (y_i - \mu)^2 \right) \\
&= \frac{1}{\sigma^{n+2}} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \right] \right) \\
&= \frac{1}{\sigma^{n+2}} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2 \right] \right)
\end{aligned}
$$

where s^2 is the sample variance of the y_i's. Note that the sufficient statistics are $\bar{y}$ and s^2.

- The conditional distribution $p(\mu|\sigma^2, y)$

  Note that we have already derived this as this is just the fixed and known variance case. So

$$\mu|\sigma^2, y \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right)$$

  We can also get it by looking at the joint posterior. The only part that contains μ looks like

$$p(\mu|\sigma^2, y) \propto \exp\left(-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\right)$$

  which is proportional to a $N\left(\bar{y}, \frac{\sigma^2}{n}\right)$ density.

- The marginal posterior distribution $p(\sigma^2|y)$

  To get this, we must integrate μ out of the joint posterior.

$$\begin{aligned} p(\sigma^2|y) &\propto \int \frac{1}{\sigma^{n+2}} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right) d\mu \\ &= \frac{1}{\sigma^{n+2}} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \int \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right) d\mu \end{aligned}$$

  The piece left inside the integral is $\sqrt{2\pi\sigma^2/n}$ times the $N\left(\bar{y}, \frac{\sigma^2}{n}\right)$ density which gives

$$\begin{aligned} p(\sigma^2|y) &\propto \frac{1}{\sigma^{n+2}} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\ &\propto \frac{1}{(\sigma^2)^{(n+1)/2}} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \end{aligned}$$

Which is a scaled inverse-χ^2 density

$$\sigma^2|y \sim \text{Inv}-\chi^2(n-1, s^2)$$

A random variable $\theta \sim \text{Inv}-\chi^2(\nu, s^2)$ if

$$\frac{\nu s^2}{\theta} \sim \chi^2_\nu$$

Note that this result agrees with the standard frequentist result on the sample variance. However this shouldn't be surprising using the results on non-informative priors, particularly the result involving pivotal quantities.

- The marginal posterior distribution $p(\mu|y)$

  Now that we have $p(\mu|\sigma^2, y)$ and $p(\sigma^2|y)$, inference on μ isn't difficult.

  One method is to use the Monte Carlo approach discussed earlier

  1. Sample σ_i^2 from $p(\sigma^2|y)$
  2. Sample μ_i from $p(\mu|\sigma_i^2, y)$

  Then $\mu_1, \ldots, \mu_m$ is a sample from $p(\mu|y)$.

  Note that in this case, it is actually possible to derive the exact density of $p(\mu|y)$.

  In this case

$$p(\mu|y) = \int p(\mu, \sigma^2|y) d\sigma^2$$

  is tractable. With the substitution $z = \frac{A}{2\sigma^2}$ where $A = (n-1)s^2 + n(\bar{y} - \mu)^2$, leaves a integral involving the gamma density (see the book, page 76).

Cranking though this leaves

$$p(\mu|y) \propto \frac{1}{\left[1 + \frac{n(\mu-\bar{y})^2}{(n-1)s^2}\right]^{n/2}}$$

a $t_{n-1}(\bar{y}, \frac{s^2}{n})$ density.

Or

$$\frac{\mu - \bar{y}}{s/\sqrt{n}}|y \sim t_{n-1}$$

which corresponds to the standard result used for inference on a population mean

$$\frac{\bar{y} - \mu}{s/\sqrt{n}}|\mu \sim t_{n-1}$$

- Posterior predictive distribution $p(\tilde{y}|y)$

$$\begin{aligned} p(\tilde{y}|y) &= \int\int p(\tilde{y}|\mu,\sigma^2,y)p(\mu,\sigma^2|y)d\mu d\sigma^2 \\ &= \int \left[\int p(\tilde{y}|\mu,\sigma^2,y)p(\mu|\sigma^2,y)d\mu\right] p(\sigma^2|y)d\sigma^2 \\ &= \int p(\tilde{y}|\sigma^2,y)p(\sigma^2|y)d\sigma^2 \end{aligned}$$

So we can figure this out by first integrating out μ, giving $p(\tilde{y}|\sigma^2,y)$. By tweaking the earlier stuff on the normal mean with a fixed variance, we know

$$\tilde{y}|\sigma^2,y \sim N(\bar{y},(1+\tfrac{1}{n})\sigma^2)$$

We can also show this by an equivalent method to the way we derived $\mu|\sigma^2,y \sim N(\bar{y},\frac{\sigma^2}{n})$

Then using the same method we used to get $p(\mu|y)$, we get that

$$\tilde{y}|y \sim t_{n-1}(\bar{y}, (1 + \tfrac{1}{n})s^2)$$

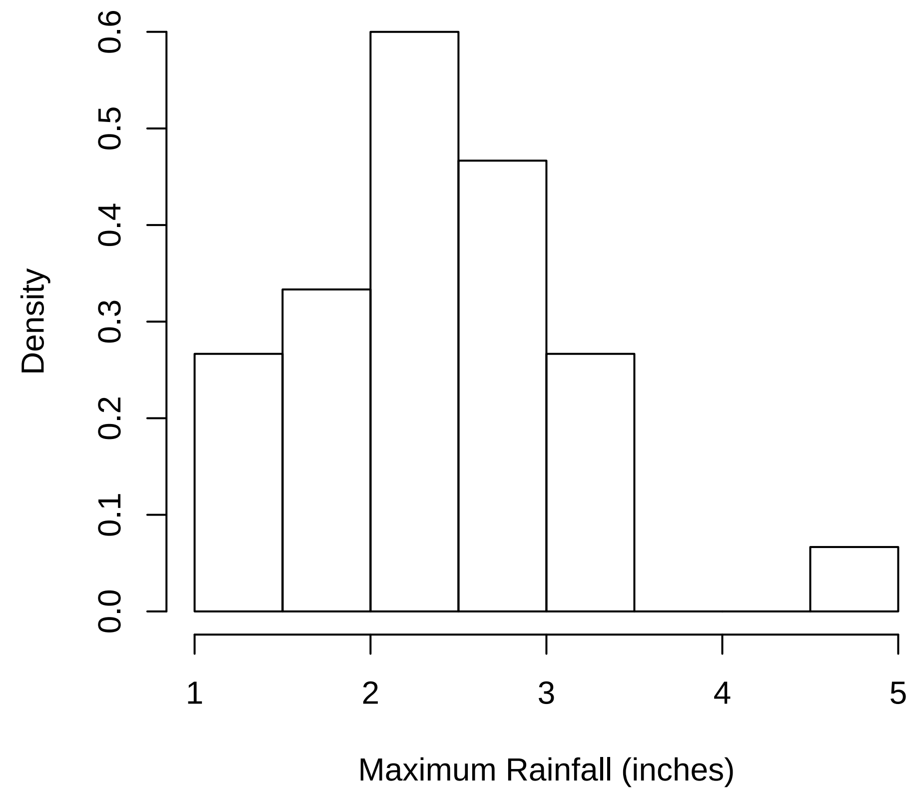
This matches with the prediction interval formula you would use in ANOVA (though it is something you don't see very often).

Example: South Bend Maximum Rainfall

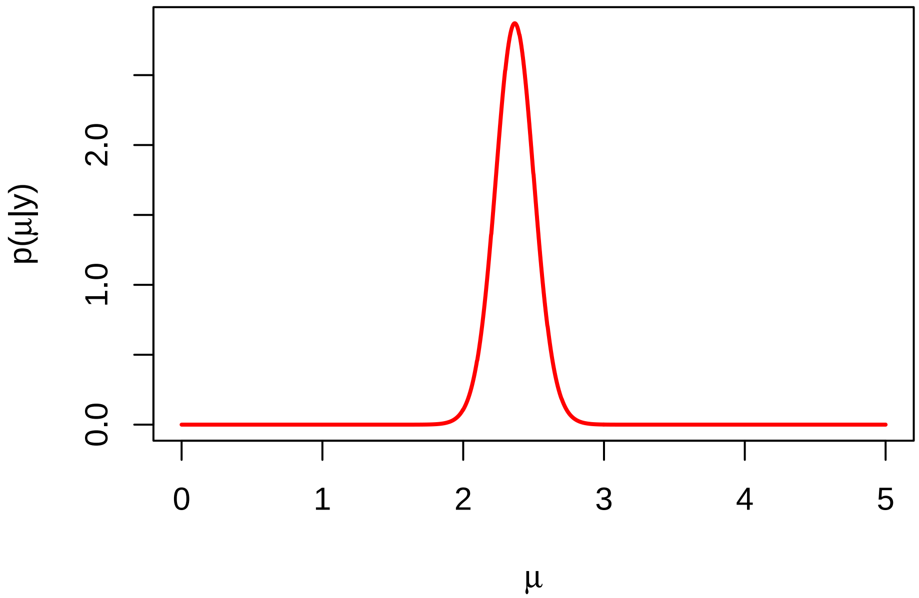
30 years (1941 - 1970) of maximum yearly rainfall at South Bend Indiana.

$$\bar{y} = 2.367 \quad s = 0.7545 \quad s^2 = 0.5692$$

Based on a normal scores plot, a normality assumption for this data doesn't seem unreasonable. This is a bit surprising as one would expect something closer to one of the three extreme value distributions (Gumbel, Frechet, or Weibull).



So the posterior distribution of the mean maximum rainfall is $t_{29}(2.367, 0.0190)$.



The posterior mean and standard deviation are

$$E[\mu|y] = 2.367 \qquad \mathrm{SD}(\mu|y) = 0.143 = \frac{s}{\sqrt{n}}\sqrt{\frac{n-1}{n-3}}$$

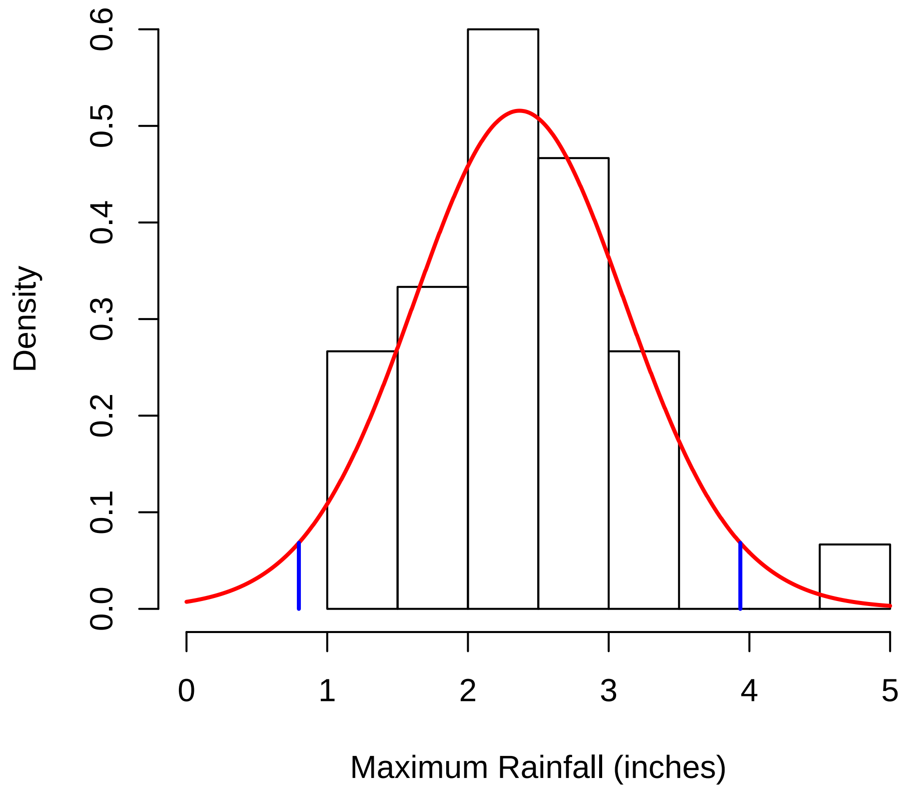
A $100(1-\alpha)\%$ central credibility interval for μ is given by

$$\bar{y} \pm t^* \frac{s}{\sqrt{n}}$$

where t^* is the $1 - \frac{\alpha}{2}$ quantile of the $t_{n-1}(0,1)$ distribution. These are what are tabled in every intro stat text. For the example, a 95% credibility for the mean maximum rainfall is

$$2.367 \pm 2.045 \frac{0.7545}{\sqrt{30}} = (2.085, 2.649)$$

The posterior predictive distribution of new observations of yearly maximum rainfall is $t_{29}(2.367, 0.558)$.



A 95% credibility interval for new observations is (0.798, 3.936). The is one observation outside this interval (4.69 in).

Inverse Gamma and Inverse Chi-square Distributions

- Gamma Distribution ($Gamma(\alpha, \beta)$)

$$p(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$$

$$E[y] = \frac{\alpha}{\beta} \qquad \mathrm{Var}(y) = \frac{\alpha}{\beta^2}$$

- Chi-square Distribution ($\chi^2_\nu = Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right)$)

$$p(y|\nu) = \frac{2^{\nu/2}}{\Gamma(\nu/2)} y^{\nu/2-1} e^{-y/2}$$

$$E[y] = \nu \qquad \mathrm{Var}(y) = 2\nu$$

- Inverse Gamma ($\mathrm{Inv-gamma}(\alpha, \beta)$)

  $y \sim \mathrm{Inv-gamma}(\alpha, \beta)$ if $\frac{1}{y} \sim Gamma(\alpha, \beta)$

$$p(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)} e^{-\beta/y}$$

  CDF: $P_{IG}(y, \alpha, \beta) = 1 - P_G\left(\frac{1}{y}, \alpha, \beta\right)$

  Quantile Function $P_{IG}^{-1}(p, \alpha, \beta) = \frac{1}{P_G^{-1}(1-p, \alpha, \beta)}$

  These are based on the fact that if $X = \frac{1}{Y}$

$$P[X \leq x] = P\left[Y \geq \frac{1}{x}\right] = 1 - P\left[Y \leq \frac{1}{x}\right]$$

$$E[y] = \frac{\beta}{\alpha - 1} \qquad \mathrm{Var}(y) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$$

- Inverse Chi-square ($\mathrm{Inv}-\chi^2_\nu$)

  $y \sim \mathrm{Inv}-\chi^2_\nu$ if $\frac{1}{y} \sim \chi^2_\nu$

  Note that $\mathrm{Inv}-\chi^2_\nu = \mathrm{Inv}-\mathrm{gamma}\left(\frac{\nu}{2}, \frac{1}{2}\right)$

$$p(y|\nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} y^{-(\nu/2+1)} e^{-1/2y}$$

  CDF: $P_{I\chi^2}(y, \nu) = 1 - P_{\chi^2}(\frac{1}{y}, \nu)$

  Quantile Function $P^{-1}_{I\chi^2}(p, \nu) = \frac{1}{P^{-1}_{\chi^2}(1-p,\nu)}$

$$E[y] = \frac{1}{\nu - 2} \qquad \mathrm{Var}(y) = \frac{2}{(\nu-2)^2(\nu-4)}$$

- Scaled Inverse Chi-square ($\mathrm{Inv}{-}\chi^2(\nu, s^2)$)

  $y \sim \mathrm{Inv}{-}\chi^2(\nu, s^2)$ if $\frac{\nu s^2}{y} \sim \chi^2_\nu$

  Note that $\mathrm{Inv}{-}\chi^2(\nu, s^2) = \mathrm{Inv}{-}\mathrm{gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}s^2\right)$

$$p(y|\nu) = \frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{\Gamma(\nu/2)} s^\nu y^{-(\nu/2+1)} e^{-\nu s^2/2y}$$

  CDF: $P_{I\chi^2}(y, \nu, s^2) = 1 - P_{\chi^2}(\frac{\nu s^2}{y}, \nu)$

  Quantile Function $P^{-1}_{I\chi^2}(p, \nu) = \frac{\nu s^2}{P^{-1}_{\chi^2}(1-p,\nu)}$

$$E[y] = \frac{\nu}{\nu - 2} s^2 \qquad \mathrm{Var}(y) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)} s^4 \qquad \mathrm{Mode}(y) = \frac{\nu}{\nu+2} s^2$$

  Note that this is a conjugate prior for the $N(\mu, \sigma^2)$ model with fixed μ.

Back to Normal Inference Models

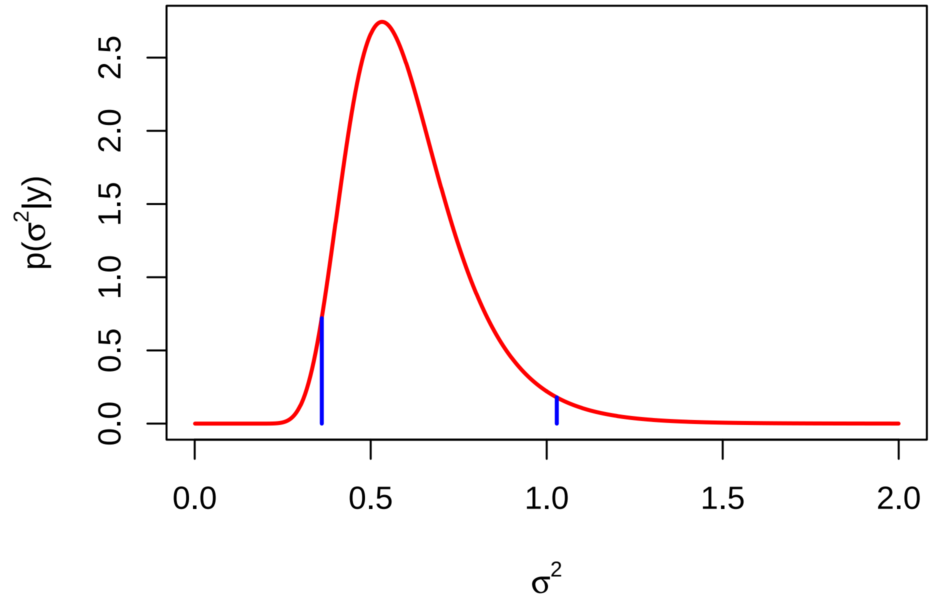
The marginal posterior distribution $p(\sigma^2|y)$

As discussed earlier,

$$\sigma^2|y \sim \text{Inv}-\chi^2(n-1, s^2)$$

For this example

$$\sigma^2|y \sim \text{Inv}-\chi^2(29, 0.5692)$$



$$E[\sigma^2|y] = 0.6114 \qquad \text{Var}(\sigma^2|y) = 0.2990 \qquad \text{SD}(\sigma^2|y) = 0.1729$$

A 95% central credibility interval is (0.3610, 1.0287)

Normal Inference Models - Informative Priors

- Conjugate Prior

  Want to find a prior such that $p(\mu, \sigma^2|y)$ is of the same form as $p(\mu, \sigma^2)$. One possibility is the two-stage prior

$$\begin{aligned} \mu|\sigma^2 &\sim N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) \\ \sigma^2 &\sim \text{Inv}-\chi^2(\nu_0, \sigma_0^2) \end{aligned}$$

  The joint density is

$$p(\mu, \sigma^2) \propto \frac{1}{\sigma}\frac{1}{(\sigma^2)^{\nu_0/2+1}} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2]\right)$$

This has been labelled as $\mathrm{N{-}Inv{-}}\chi^2(\mu_0, \frac{\sigma_0^2}{\kappa_0}; \nu_0, \sigma_0^2)$ distribution

One way of think of this prior is that we have κ_0 observations with an average of μ_0 and another $\nu_0 + 1$ with a sample variance of σ_0^2.

One important thing to note is that with this prior, μ and σ^2 are dependent (i.e. $p(\mu|\sigma^2)$ is a function of σ^2. Since this happens to be a conjugate prior, it also implies that μ and σ^2 are dependent aposteriori.

This has a different feel from the standard frequentist analysis where $\bar{y}$ and s^2 are independent.

The posterior density satisfies

$$\begin{aligned}
p(\mu, \sigma^2|y) &\propto \frac{1}{\sigma}\frac{1}{(\sigma^2)^{\nu_0/2+1}} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2]\right) \\
&\quad \times \frac{1}{(\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\right) \\
&\propto \frac{1}{\sigma}\frac{1}{(\sigma^2)^{\nu_n/2+1}} \exp\left(-\frac{1}{2\sigma^2}[\nu_n\sigma_n^2 + \kappa_n(\mu - \mu_n)^2]\right)
\end{aligned}$$

The posterior distribution is $\mathrm{N{-}Inv{-}}\chi^2(\mu_n, \frac{\sigma_n^2}{\kappa_n}; \nu_n, \sigma_n^2)$ where

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2
\end{aligned}$$

$p(\mu|\sigma^2, y)$:

By using that $p(\mu|\sigma^2, y) \propto p(\mu, \sigma^2|y)$, by collecting terms involving μ, we get

$$\mu|\sigma^2, y \sim N\left(\mu_n, \frac{\sigma}{\kappa_n}\right)$$

Note that that the mean and variance can be written as

$$\mu_n = \frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}} \qquad \sigma_n^2 = \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}$$

which matches with the fixed variance case discuss earlier (as it should).

$p(\sigma^2|y)$:

$$\sigma^2|y \sim \mathrm{Inv}{-}\chi^2(\nu_n, \sigma_n^2)$$

This can be seen by the same way $p(\sigma^2|y)$ was shown in the non-informative prior case or by recognizing the $\mathrm{N}{-}\mathrm{Inv}{-}\chi^2$ form of the joint density.

$p(\mu|y)$:

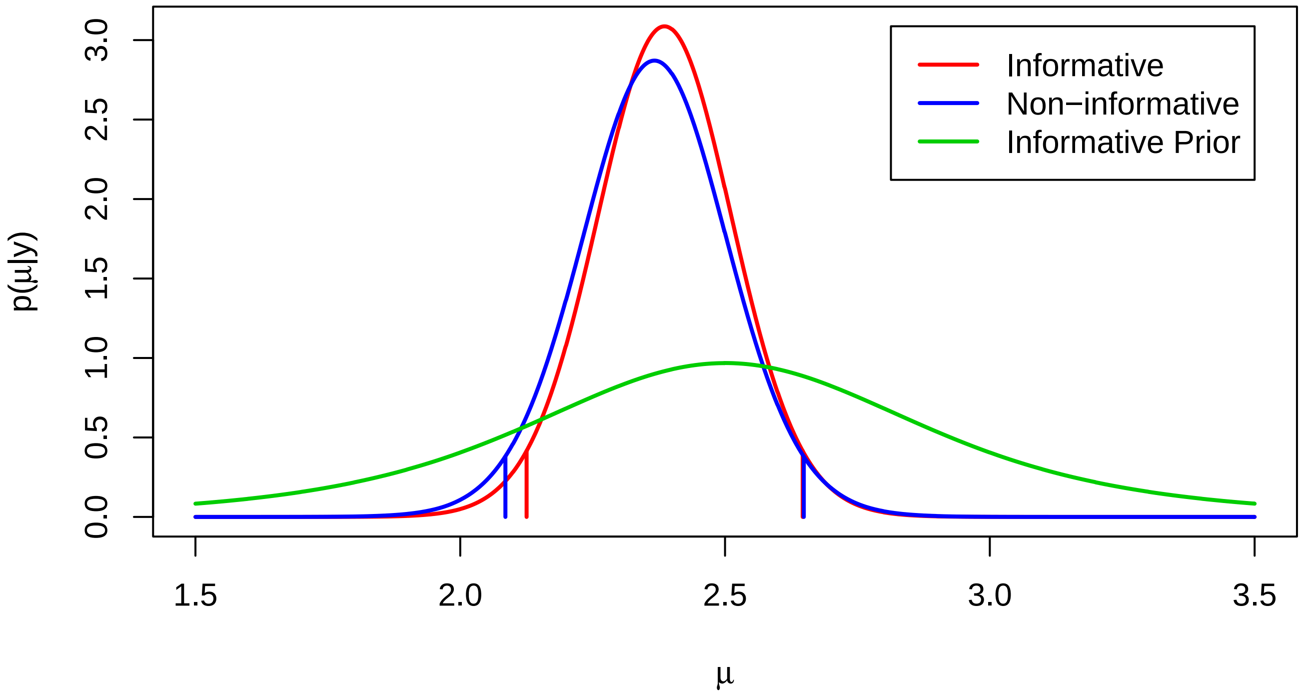
As mentioned before, this can be determined by simulation. However in this case an exact answer can be determined by integrating out σ^2 from the joint density (as in the non-informative case), we get

$$\mu|y \sim t_{v_n}\left(\mu_n, \frac{\sigma_n^2}{\kappa_n}\right)$$

Example analysis:

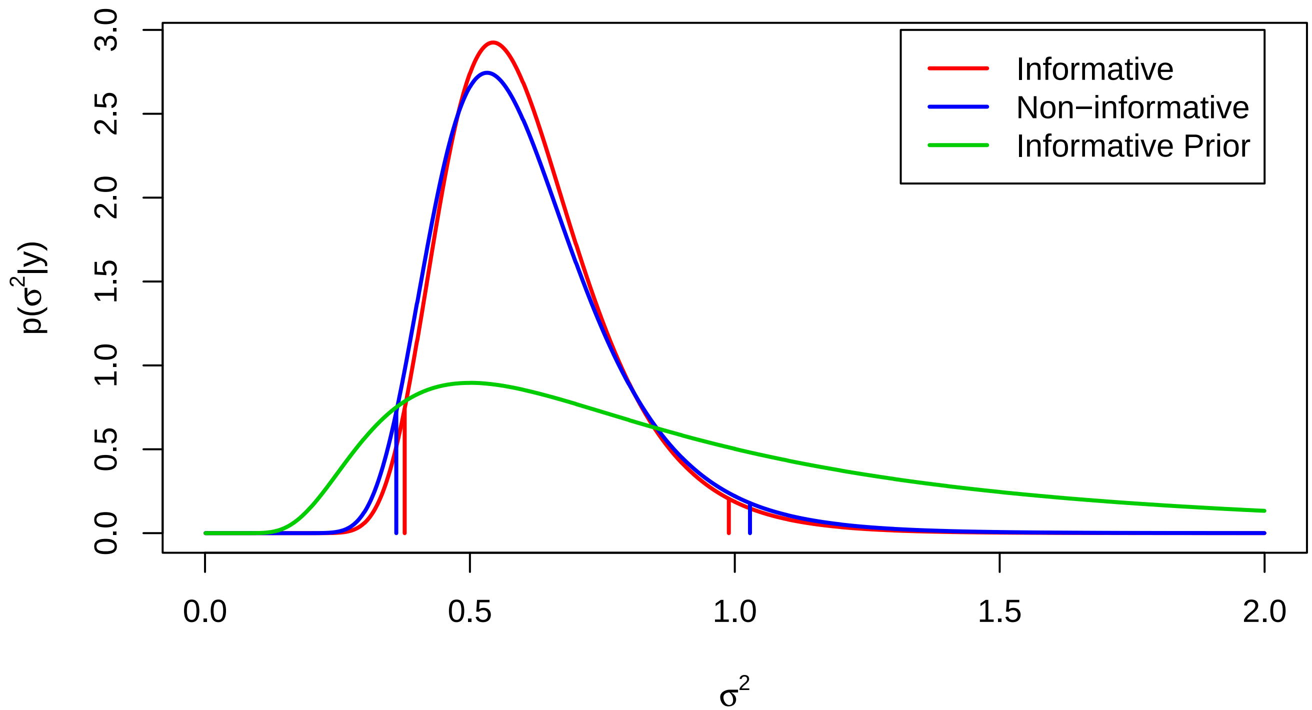
Prior choice: Let's assume that maximum rainfall should fall between 1 and 4 inches and that the distribution should be roughly symmetric. This implies that $\mu_0 = 2.5$. Also lets assume that this is based on $\kappa_0 = 5$ observations. For σ, let $\sigma_0^2 = 0.75$ and $\nu_0 = 4$ (Why these, I don't know. We need numbers).

$\mu|y \sim t_{35}(2.386, 0.0165)$:



Prior	$E[\mu\|y]$	$\mathrm{SD}(\mu\|y)$	95% Cred. Int.
Non-informative	2.367	0.143	(2.085, 2.649)
Informative	2.386	0.132	(2.125, 2.647)

$$\sigma^2 | y \sim \text{Inv}-\chi^2(34, 0.5760)$$



Prior	$E[\sigma^2 \mid y]$	$\text{SD}(\sigma^2 \mid y)$	95% Cred. Int.
Non-informative	0.6114	0.1729	(0.3610, 1.0287)
Informative	0.6120	0.1580	(0.3768, 0.9887)