STA 602 - Intro to Bayesian Statistics

Lecture 9

Li Ma

Duke University

Gaussian model with unknown mean and unkown variance

- Sampling model for n readings given the mean θ and variance σ^2 is

$$X_1, X_2, \ldots, X_n \mid \theta, \sigma^2 \overset{\text{iid}}{\sim} \mathrm{N}(\theta, \sigma^2)$$

- Prior distribution for the mean μ

$$\theta \mid \sigma^2 \sim \mathrm{N}(\mu_0, \tau_0^2)$$

- Now let's assume that the sampling variance σ^2 is also unknown.
- So now we have a two parameter model (θ, σ^2).
- To complete Bayesian inference, we need to specify a joint prior for (θ, σ^2).

A conjugate prior

- If we let $\tau_0^2 = \sigma^2/\kappa_0$,

$$\theta \,|\, \sigma^2 \sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0),$$

then we have found the conditional posterior

$$\theta \,|\, \mathbf{x}, \sigma^2 \sim \mathrm{N}(\mu_n, \tau_n^2)$$

where

$$\mu_n = \frac{\kappa_0}{\kappa_n}\mu_0 + \frac{n}{\kappa_n}\bar{x}$$

where $\kappa_n = \kappa_0 + n$ and

$$\tau_n^2 = \sigma^2/\kappa_n.$$

$$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} = \frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2} = \frac{\kappa_n}{\sigma^2}$$

A conjugate prior

- There exists a marginal prior on σ^2 as well to make the joint prior on $\underline{(\theta, \sigma^2)}$ fully conjugate.

$p(\theta|\sigma^2) \cdot p(\sigma^2)$

$$\frac{1}{\sigma^2} \sim \text{Gamma}(\nu_0/2, \nu_0\sigma_0^2/2).$$

- Equivalently sometimes we say that σ^2 follows an *inverse-Gamma* (IG) prior

$$\sigma^2 \sim \text{IG}(\nu_0/2, \nu_0\sigma_0^2/2).$$

An equivalent parametrization

$\chi_\nu^2 = \nu \cdot [N(0,1)]^2$

- Recall that χ_ν^2 distribution (i.e., with ν degrees of freedom) is $\text{Gamma}(\nu/2, 1/2)$.
- So $\text{Gamma}(\nu_0/2, \nu_0\sigma_0^2/2)$ is a scaled $\chi_{\nu_0}^2$ distribution

equal in distribution

$\nu_0 \to \infty$ converge to 1.

$$\text{Gamma}(\nu_0/2, \nu_0\sigma_0^2/2) =_d \frac{\chi_{\nu_0}^2}{\nu_0\sigma_0^2} = \boxed{\frac{\chi_{\nu_0}^2}{\nu_0}} \cdot \frac{1}{\sigma_0^2}.$$

- It has mean $1/\sigma_0^2$, and its variance is smaller as the degrees of freedom ν_0 increases. (What happens when $\nu_0 \uparrow \infty$?)
- ν_0 is a *prior degrees of freedom (d.f.)*, which quantifies the strength of our prior knowledge for the variance.

The marginal posterior of σ^2

- One can find the marginal posterior of σ^2 to be

$$\frac{1}{\sigma^2} \mid \mathbf{x} \sim \text{Gamma}(\nu_n/2, \nu_n \sigma_n^2/2)$$

or

$$\sigma^2 \mid \mathbf{x} \sim \text{IG}(\nu_n/2, \nu_n \sigma_n^2/2)$$

where $\nu_n = \nu_0 + n$ (the posterior d.f.)

$$\sigma_n^2 = \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_n}(\bar{x}-\mu_0)^2\right]$$

$$= \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + \sum_{i=1}^{n}(x_i-\bar{x})^2 + \frac{\kappa_0}{\kappa_n}\cdot \underline{n(\bar{x}-\mu_0)^2}\right]$$

σ^2 plays two roles:
1) sample variances
2) prior variance of θ

- The posterior mean of σ^2 has contribution from three pieces
  - prior mean σ_0^2 (with weight proportional to prior d.f.)
  - sample variance $s^2 = \sum_i (x_i - \bar{x})^2/(n-1)$ (with weight proportional to $n-1$)
  - the deviation of sample mean $\bar{x}$ from its prior mean μ_0. (Because $\tau_0^2 = \sigma^2/\kappa_0$, σ^2 is tied to the variability of θ.)

Deriving the marginal posterior of σ^2 (or $1/\sigma^2$)

- There are several ways to derive the marginal posterior.
- The most basic way is by marginalizing out θ in the joint posterior of $(\theta, 1/\sigma^2)$.
- For notational simplicity, let $\gamma = 1/\sigma^2$.
- By Bayes theorem,

$$p(\theta,\gamma|\mathbf{x}) \propto p(\theta|\gamma)p(\gamma)p(\mathbf{x}|\theta,\gamma)$$

$$\propto \underbrace{\gamma^{\frac{1}{2}} e^{-\frac{1}{2}\kappa_0\gamma(\theta-\mu_0)^2}}_{N(\mu_0,\, \kappa_0\cdot\gamma)} \cdot \underbrace{\gamma^{\frac{\nu_0}{2}-1} e^{-\frac{\nu_0\sigma_0^2}{2}\gamma}}_{\text{gamma}\left(\frac{\nu_0}{2},\, \frac{\nu_0}{2}\sigma_0^2\right)} \cdot \underbrace{\gamma^{\frac{n}{2}} e^{-\frac{1}{2}\gamma\sum_i(x_i-\theta)^2}}_{N(\theta,\, \frac{1}{\gamma})}$$

Note that the constant must not involve either θ or γ.

- Use the fact that

$$\sum_i (x_i-\theta)^2 = \sum_i (x_i-\bar{x})^2 + n(\bar{x}-\theta)^2 = (n-1)s^2 + n(\bar{x}-\theta)^2$$

$$\sum_i (x_i - \bar{x} + \bar{x} - \theta)^2 = \sum_i (x_i-\bar{x})^2 + \sum_i (\bar{x}-\theta)^2 + 2\sum_i \underbrace{(x_i-\bar{x})}_{0}(\bar{x}-\theta)$$

$$= (n-1)s^2 + n(\bar{x}-\theta)^2.$$

- We have

$$\int p(\theta,\gamma|\mathbf{x})\,d\theta \propto \gamma^{\frac{1}{2}} e^{-\frac{1}{2}\kappa_0\gamma(\theta-\mu_0)^2} \cdot \gamma^{\frac{\nu_0}{2}-1} e^{-\frac{\nu_0\sigma_0^2}{2}\gamma} \cdot \gamma^{\frac{n}{2}} e^{-\frac{1}{2}\gamma\cdot(n-1)s^2} \cdot e^{-\frac{1}{2}\gamma\cdot n(\bar{x}-\theta)^2}$$

$$\propto \int \underbrace{\gamma^{\frac{1}{2}} e^{-\frac{1}{2}[\kappa_0\gamma(\theta-\mu_0)^2+n\gamma(\bar{x}-\theta)^2]}}_{\text{Involves } \theta \text{ and } \gamma} \cdot \underbrace{\gamma^{\frac{\nu_0+n}{2}-1} e^{-\frac{\gamma}{2}[\nu_0\sigma_0^2+(n-1)s^2]}}_{\text{Involves only } \gamma}\,d\theta$$

- Now let's focus on the first term. The idea is to turn it into the form of

$$\gamma^{1/2} e^{-\frac{(\theta-A)^2}{2}\cdot B\gamma}$$

through completion of squares, because

$$\int \gamma^{1/2} e^{-\frac{(\theta-A)^2}{2}\cdot B\gamma} d\theta = \sqrt{2\pi/B} \quad \text{a constant not involving } \theta \text{ and } \gamma$$

from the fact that

$$\int \left(\frac{B\gamma}{2\pi}\right)^{1/2} e^{-\frac{1}{2}B\gamma(\theta-A)^2} = 1$$

Finding the values of A and B (which we have already done before!)

- Let's complete the square for the exponent in the first term

$$
\begin{aligned}
&\kappa_0\gamma(\theta-\mu_0)^2+n\gamma(\bar{x}-\theta)^2\\
=&(\kappa_0+n)\gamma\left[\left(\theta-\frac{\kappa_0\mu_0+n\bar{x}}{\kappa_0+n}\right)^2-\left(\frac{\kappa_0\mu_0+n\bar{x}}{\kappa_0+n}\right)^2+\frac{\kappa_0\mu_0^2+n\bar{x}^2}{\kappa_0+n}\right]\\
=&(\kappa_0+n)\gamma\left[\left(\theta-\frac{\kappa_0\mu_0+n\bar{x}}{\kappa_0+n}\right)^2+\frac{\kappa_0 n}{(\kappa_0+n)^2}(\bar{x}-\mu_0)^2\right]\\
=&\kappa_n\gamma\left[(\theta-\mu_n)^2+\frac{\kappa_0 n}{\kappa_n^2}(\bar{x}-\mu_0)^2\right].
\end{aligned}
$$

- Remark: Previously when σ^2 is known, we have treated the second term, which does not involve θ as a "constant". Note that here we cannot do that because σ^2 is also an unknown parameter!

$$\int_{\Theta} \gamma^{1/2} \cdot e^{-\frac{1}{2}\kappa_n \cdot \gamma (\theta - \mu_n)^2} d\theta \propto 1 \text{ (Const. w.r.t. } \theta \text{ and } \gamma).$$

- Therefore

$$p(\theta,\gamma|\mathbf{x}) \propto \underbrace{\gamma^{\frac{1}{2}} e^{-\frac{1}{2}\overset{\kappa_n}{B}\gamma(\theta-\overset{\mu_n}{A})^2}}_{\text{Involves } \theta \text{ and } \gamma} \cdot \underbrace{\gamma^{\frac{\nu_0+n}{2}-1} e^{-\frac{\gamma}{2}[\nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{\kappa_n}(\bar{x}-\mu_0)^2]}}_{\text{Involves only } \gamma}$$

where

$$A = \mu_n = \frac{\kappa_0\mu_0 + n\bar{x}}{\kappa_0 + n} \quad \text{and} \quad B = \kappa_n = \kappa_0 + n.$$

- *Caveat*: Just because the first term looks like a Guassian density in θ, we still need to keep track of the γ in its normalizing constant!

- Moreover,

$$
\begin{aligned}
p(\gamma|\mathbf{x}) &= \int p(\theta,\gamma|\mathbf{x})d\theta \\
&\propto \sqrt{2\pi/B}\cdot\gamma^{\frac{\nu_0+n}{2}-1}e^{-\frac{\gamma}{2}[\nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{\kappa_n}(\bar{x}-\mu_0)^2]} \\
&\propto \gamma^{\frac{\nu_0+n}{2}-1}e^{-\frac{\gamma}{2}[\nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{\kappa_n}(\bar{x}-\mu_0)^2]} \\
&= \gamma^{\frac{\nu_n}{2}-1}e^{-\frac{\gamma\nu_n\sigma_n^2}{2}},
\end{aligned}
$$

  which is a Gamma$(\nu_n/2, \nu_n\sigma_n^2/2)$ where $\nu_n = \nu_0 + n$.
- Note that we never needed to know the value of A and B!

The marginal posterior of θ

- One can also find the marginal posterior of θ, $p(\theta \,|\, \mathbf{x})$, by integrating out σ^2 (or γ) in the joint posterior. It is a (scaled and shifted) t-distribution with ν_n degrees of freedom.
- Specifically,
$$\frac{\theta - \mu_n}{\sigma_n / \sqrt{\kappa_n}} \sim t_{\nu_n}.$$
- This can be done directly through integrating out γ.
- This time, we will need the values of A and B!

- To see this, note that by plugging in the values of A and B,

$$p(\theta,\gamma\,|\,\mathbf{x}) = \underbrace{p(\theta\,|\,\gamma,\mathbf{x})p(\gamma\,|\,\mathbf{x})}_{\text{use a short cut}} \propto \gamma^{\frac{\nu_n-1}{2}} e^{-\frac{1}{2}\gamma[\kappa_n(\theta-\mu_n)^2+\nu_n\sigma_n^2]}.$$

$\gamma^{\frac{1}{2}} \exp\{-\frac{\kappa_n\gamma}{2}(\theta-\mu_n)^2\} \times \gamma^{\frac{\nu_n}{2}-1} \exp\{-\nu_n\sigma_n^2\gamma\}$.

- Hence

$$p(\theta\,|\,\mathbf{x}) = \int p(\theta,\gamma\,|\,\mathbf{x})d\gamma \propto \int \gamma^{\frac{\nu_n-1}{2}} e^{-\frac{1}{2}\gamma[\kappa_n(\theta-\mu_n)^2+\nu_n\sigma_n^2]} d\gamma$$

$$\propto \frac{\Gamma\left(\frac{\nu_n+1}{2}\right)}{\left\{\frac{1}{2}\left[\kappa_n(\theta-\mu_n)^2+\nu_n\sigma_n^2\right]\right\}^{\frac{\nu_n+1}{2}}}$$

$$\propto \left[1+\left(\frac{\theta-\mu_n}{\sigma_n/\sqrt{\kappa_n}}\right)^2\cdot\frac{1}{\nu_n}\right]^{-\frac{\nu_n+1}{2}}.$$

Now consider:

$\theta|\sigma^2 \sim N(\mu_0, \frac{\sigma^2}{\kappa_0})$

$1/\sigma^2 \sim \text{Gamma}(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2)$.

Then $\frac{\theta-\mu_0}{\sigma_0/\sqrt{\kappa_0}} \sim t(\nu_0)$

Strategy of draw t samples

σ^2 is latent variable

- Recall that the density of the t_ν distribution is

$$p(t) \propto (1+t^2/\nu)^{-\frac{\nu+1}{2}}$$

- So by change of variable, the posterior density of $t = \frac{\theta-\mu_n}{\sigma_n/\sqrt{\kappa_n}}$ is exactly proportional to the density of a t_n-distribution. Thus

$$\frac{\theta-\mu_n}{\sigma_n/\sqrt{\kappa_n}} \sim t_{\nu_n}.$$

t-distribution as a *scale-mixture* of normals

- More generally, a scale-mixture of Gaussian with IG prior on the variance leads to a t-distribution. That is, if

$$\begin{aligned} Y \mid b &\sim \mathrm{N}(a, b^2) \\ b^2 &\sim \mathrm{IG}(\nu/2, c^2\nu/2) \end{aligned}$$

  Then

$$\frac{Y-a}{c} \sim t_\nu.$$

- We just proved it on the previous slide! (Simply repeat by replacing μ_n with a, $1/\gamma$ with b^2, ν_n with ν, and c^2 with σ_n^2.)
- In particular, a t-distribution with ν d.f. is exactly the marginal distribution of a $\mathrm{Normal}(0, \sigma^2)$ and a $\mathrm{Gamma}(\nu/2, \nu/2)$ on $1/\sigma^2$.

$$\begin{aligned} t_\nu(\cdot) &= \int \mathrm{N}(\cdot \mid 0, \sigma^2) \times \mathrm{IG}(\sigma^2 \mid \nu/2, \nu/2) d\sigma^2 \\ &= \int \mathrm{N}(\cdot \mid 0, 1/\gamma) \times \mathrm{Gamma}(\gamma \mid \nu/2, \nu/2) d\gamma. \end{aligned}$$

Going the other way around knowing this property of t distributions

- Applying this to our posterior

$\theta \mid \sigma^2, x \sim N(\mu_n, \frac{\sigma^2}{\kappa_n})$

$1/\sigma^2 \mid x \sim \Gamma(\frac{\nu_n}{2}, \frac{\nu_n}{2}\sigma_n^2)$

$\frac{\kappa_n}{\sigma^2} \mid x \sim \Gamma(\frac{\nu_n}{2}, \frac{\nu_n}{2}\sigma_n^2/\kappa_n)$

$\frac{\sigma^2}{\kappa_n} \mid x \sim \Gamma^{-1}(\frac{\nu_n}{2}, \frac{\nu_n}{2}\frac{\sigma_n^2}{\kappa_n})$

$$
\begin{aligned}
p(\theta|\mathbf{x}) &= \int \underbrace{p(\theta|\gamma,\mathbf{x})}_{\mathrm{N}(\mu_n, \tau_n^2 = \frac{1}{\kappa_n \gamma})} \underbrace{p(\gamma|\mathbf{x})}_{\mathrm{Gamma}(\nu_n/2, \nu_n\sigma_n^2/2)} d\gamma \\
&= \int \underbrace{p(\theta|\tau_n^2,\mathbf{x})}_{\mathrm{N}(\mu_n,\tau_n^2)} \underbrace{p(\tau_n^2|\mathbf{x})}_{\mathrm{IG}\left(\frac{\nu_n}{2}, \frac{\nu_n\sigma_n^2/\kappa_n}{2}\right)} d(\tau_n^2).
\end{aligned}
$$

- By setting $a = \mu_n$, $b = \tau_n$, and $c = \sigma_n/\sqrt{\kappa_n}$, we have

$$
\frac{\theta - a}{c} = \frac{\theta - \mu_n}{\sigma_n/\sqrt{\kappa_n}} \sim t_{\nu_n}.
$$

Monte Carlo for t-distributions as a scale-mixture of Gaussian

- We can use this fact to draw Monte Carlo samples from a t-distribution with a shift a and a scaled parameter c?
- For $i = 1, 2, \ldots, S$,
  - First draw $\sigma^{2\,(i)} \sim \mathrm{IG}(\nu/2, \nu/2)$.
  - Then draw $t^{(i)} \,|\, \sigma^{(i)} \sim \mathrm{N}(0, \sigma^{2\,(i)})$ and set $\theta^{(i)} = a + ct^{(i)}$.
- The last step is equivalent to drawing $\theta^{(i)} \,|\, \sigma^{(i)} \sim \mathrm{N}(a, c^2\sigma^{2\,(i)})$.

- Online animation for the special case $a = 0$ and $c = 1$: http://www.sumsar.net/blog/2013/12/t-as-a-mixture-of-normals/

Predictive distribution

- One can also show that the predictive distribution for a new observation x_{new} is also a (scaled and shifted) t distribution with ν_n degrees of freedom.
- Hint:

$$
\begin{aligned}
p(x_{new}|\mathbf{x}) &= \int\int p(x_{new}|\theta,\gamma)p(\theta,\gamma|\mathbf{x})d\mu d\gamma \\
&= \int \underbrace{\int p(x_{new}|\theta,\gamma)p(\theta|\gamma,\mathbf{x})d\theta}_{\mathrm{N}(\mu_n,\tau_n^2+\sigma^2=(1/\kappa_n+1)/\gamma)} \quad \underbrace{p(\gamma|\mathbf{x})}_{\mathrm{Gamma}(\nu_n/2,\nu_n\sigma_n^2/2)} \quad d\gamma
\end{aligned}
$$

and therefore

$$\frac{x_{new}-\mu_n}{\sigma_n\sqrt{1+1/\kappa_n}} \sim t_{\nu_n}.$$

Monte Carlo sampling from the posterior

- We want to draw S samples form the joint posterior (θ, σ^2) given the data $\mathbf{x}$.

$$(\theta^{(1)}, \sigma^{2\,(1)}), (\theta^{(2)}, \sigma^{2\,(2)}), \ldots, (\theta^{(S)}, \sigma^{2\,(S)}).$$

- For $i = 1, 2, \ldots, S$, we proceed in two steps
  - Draw a sample of $\sigma^{2\,(i)}$ or equivalently $\gamma^{(i)} = 1/\sigma^{2(i)}$ from the marginal posterior $p(\sigma^2|\mathbf{x})$ or $p(\gamma|\mathbf{x})$.
  - Then draw $\theta^{(i)}$ from the conditional distribution of $p(\theta|\sigma^{2\,(i)}, \mathbf{x})$.

- That is, for $i = 1, 2, \ldots, S$,

$$\sigma^{2\,(i)} \sim \mathrm{IG}(\nu_n, \nu_n \sigma_n^2)$$
$$\theta^{(i)} \,|\, \sigma^{2\,(i)} \sim \mathrm{N}(\mu_n, \sigma^{2\,(i)}/\kappa_n).$$

Example: Air pollutant

- Suppose you decided to adopt the above conjugate prior.
- Your prior belief on the mean θ is equivalent to about $\kappa_0 = 5$ measurements. That is, $\tau_0 = \sigma^2/5$.
- Regarding the variance σ^2 of the device, however, suppose the device, you know that the precision (i.e., $1/\sigma^2$) is about $1/4$ and so $\sigma_0^2 = 4$ but you are unsure about the quality of the device, so have large uncertainty in its precision. For example, $\nu_0 = 1$.
- Suppose you recorded 10 readings

$$\mathbf{x} = (104, 105, 103, 102, 105, 107, 106, 104, 103, 106)$$

```r
x <- c(104,105,103,102,105,107,106,104,103,106) # the data
n <- length(x) # sample size

mu.0 <- 100   # prior mean for mu
kappa.0 <- 5  # prior sample size
nu.0 <- 1     # prior degrees of freedom
sigma2.0 <- 4
s2 <- var(x)  # sample variance of data


kappa.n <- kappa.0 + n
mu.n <- (kappa.0*mu.0 + sum(x))/kappa.n
nu.n <- nu.0 + n
sigma2.n <- (nu.0*sigma2.0 + (n-1)*s2 +
       kappa.0*n/kappa.n*(mean(x)-mu.0)^2)/nu.n
```

```r
print(mean(x))
```

```
## [1] 104.5
```

```r
print(mu.n)
```

```
## [1] 103
```

```r
print(kappa.n)
```

```
## [1] 15
```

```r
print(nu.n)
```

```
## [1] 11
```

```r
print(sigma2.n)
```

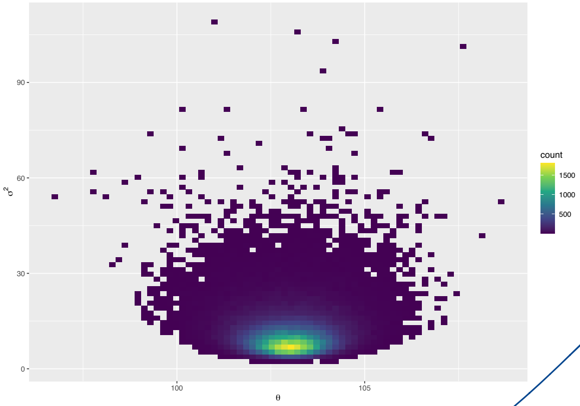
```
## [1] 8.545455
```

Draw Monte Carlo samples

```r
S=100000
sigma2.mc <- 1/rgamma(S,shape=nu.n/2,rate=nu.n*sigma2.n/2)
theta.mc <- rnorm(S,mean=mu.n,sd=sqrt(sigma2.mc/kappa.n))
```

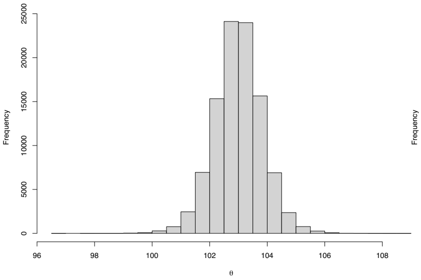
2D histogram of joint samples

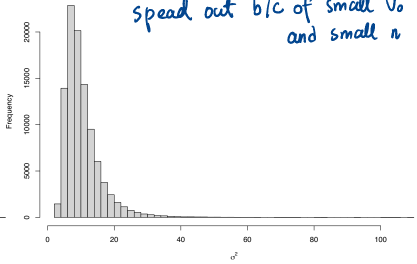
```
post.samples <- data.frame(theta=theta.mc,
                           sigma2=sigma2.mc)
ggplot(post.samples, aes(x=theta, y=sigma2) ) +
  labs(x=expression(theta),y=expression(sigma^2)) +
  geom_bin2d(bins=70) +
  scale_fill_continuous(type = "viridis")
```



Marginal histograms

```r
hist(theta.mc,breaks=20,main="",xlab=expression(theta))
hist(sigma2.mc, breaks=50, main="",xlab=expression(sigma^2))
```





spead out b/c of small v_0 and small n

$t(v_n) = t(11)$
low degree of freedom.

With a stronger prior on σ^2 (e.g., $\nu_0 = 100$)

```r
nu.0 <- 100     # prior degrees of freedom
nu.n <- nu.0 + n
sigma2.n <- (nu.0*sigma2.0 + (n-1)*s2 +
       kappa.0*n/kappa.n*(mean(x)-mu.0)^2)/nu.n
S=100000
sigma2.mc <- 1/rgamma(S,shape=nu.n/2,rate=nu.n*sigma2.n/2)
theta.mc <- rnorm(S,mean=mu.n,sd=sqrt(sigma2.mc/kappa.n))
```

Posterior parameters

```r
print(mean(x))
```

```
## [1] 104.5
```

```r
print(mu.n)
```

```
## [1] 103
```

```r
print(kappa.n)
```

```
## [1] 15
```

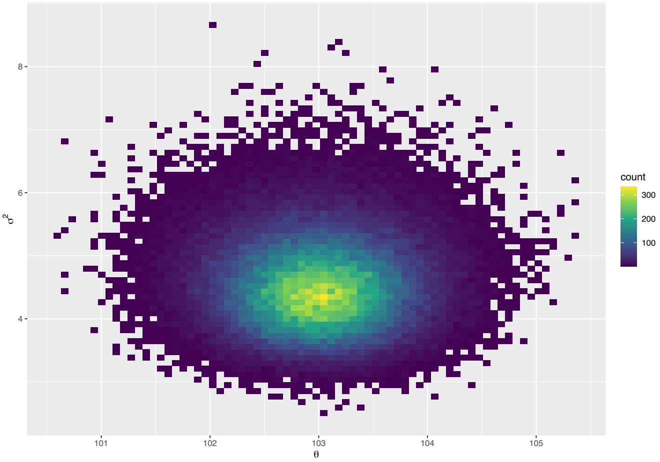
```r
print(nu.n)
```

```
## [1] 110
```

```r
print(sigma2.n)
```

```
## [1] 4.454545
```

Joint histogram



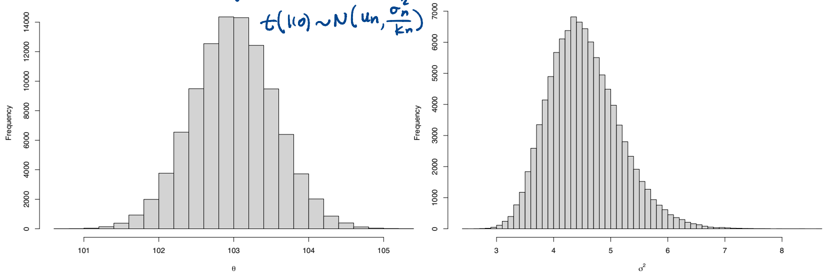
Marginal histograms

more tight, $v_n = v_0 + n = 100 + 10 = 110$.

essentially a gaussian

$t(110) \sim N(u_n, \frac{\sigma_n^2}{k_n})$

more concentrated



- Notice the change in spread of the marginal posterior of θ.
- Notice the change in spread and shape for σ^2.

A non-informative prior

- Suppose we have little prior knowledge about θ and σ^2.
- Use a vague prior by letting $\kappa_0 \to 0$ and $\nu_0 \to 0$.
  - $\kappa_0 \to 0$ leads to

$$p(\theta|\sigma^2) \propto 1.$$

$\sigma^2 \sim \Gamma^{-1}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right)$

  - $\nu_0 \to 0$ leads to

$$p(\sigma^2) \propto 1/\sigma^2.$$

$p(\sigma^2) \propto (\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left\{-\frac{\frac{\nu_0\sigma_0^2}{2}}{\sigma^2}\right\} \xrightarrow{\nu_0 \to 0} (\sigma^2)^{-1}$

- Together, we have

$$p(\theta, \sigma^2) \propto 1/\sigma^2.$$

This is an *improper* prior as it integrates to infinity over $-\infty < \theta < \infty$ and $\sigma^2 > 0$.
- Nevertheless, this prior leads to a proper posterior.

The corresponding posterior

- The posterior parameters are

$$
\begin{aligned}
\kappa_n &= \kappa_0 + n = n \\
\mu_n &= \frac{\kappa_0}{\kappa_n}\mu_0 + \frac{n}{\kappa_n}\bar{x} = \bar{x} \\
\tau_n^2 &= \sigma^2/\kappa_n = \sigma^2/n \\
\nu_n &= \nu_0 + n = n \\
\sigma_n^2 &= \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_n}(\bar{x}-\mu_0)^2\right] \\
&= \frac{(n-1)s^2}{n} = \frac{1}{n}\sum_i (x_i - \bar{x})^2.
\end{aligned}
$$

- Hence we have

$$
\begin{aligned}
\sigma^2 \mid \mathbf{x} &\sim \mathrm{IG}(n/2, (n-1)s^2/2) \\
\theta \mid \sigma^2, \mathbf{x} &\sim \mathrm{N}(\bar{x}, \sigma^2/n).
\end{aligned}
$$

Example: Air pollutant

```r
print(mean(x))
```

```
## [1] 104.5
```

```r
print(mu.n)
```

```
## [1] 104.5
```

```r
print(kappa.n)
```

```
## [1] 10
```

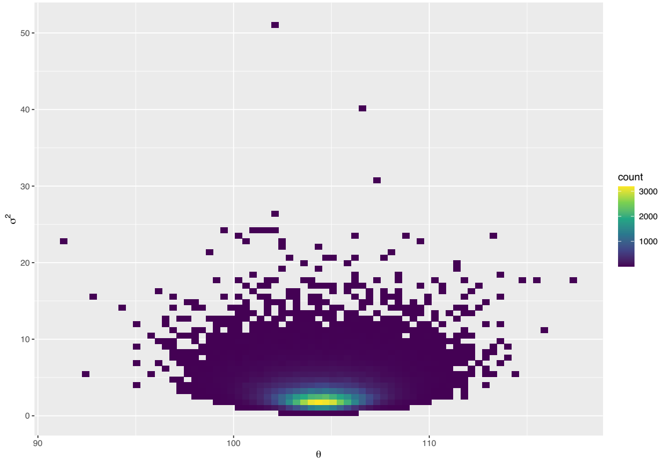
```r
print(nu.n)
```

```
## [1] 10
```

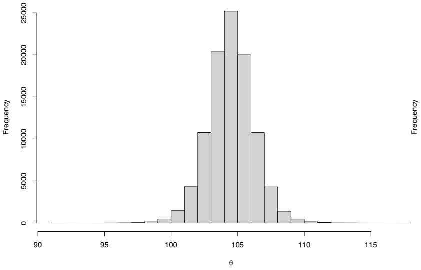
```r
print(sigma2.n)
```

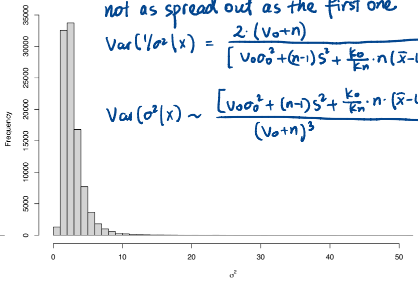
```
## [1] 2.25
```

Joint histogram



Marginal histograms





not as spread out as the first one

$$\mathrm{Var}(1/\sigma^2 \mid x) = \frac{2\cdot(\nu_0+n)}{\left[\nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0}{\kappa_n}\cdot n(\bar{x}-\mu_0)^2\right]^2}$$

$$\mathrm{Var}(\sigma^2 \mid x) \sim \frac{\left[\nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0}{\kappa_n}\cdot n\cdot(\bar{x}-\mu_0)^2\right]^2}{(\nu_0+n)^3}$$

- Notice the change in spread of the marginal posterior of θ.
- Notice the change in spread and shape for σ^2.

- Question: Why is the marginal posterior of σ^2 less spread out than the first case with $\nu_0 = 1$ and $\kappa_0 = 5$?
- Hint: Note the third term that contributes to σ_n^2.

$$\sigma_n^2 = \frac{1}{\nu_n}\left[\cancel{\nu_0\sigma_0^2} + (n-1)s^2 + \cancel{\frac{\kappa_0 n}{\kappa_n}(\bar{x}-\mu_0)^2}\right].$$