# Applied Bayesian Modeling module 4: **Deriving a posterior, role of prior information**

Leontine Alkema, lalkema@umass.edu

Fall 2022

# Recap of module 3:

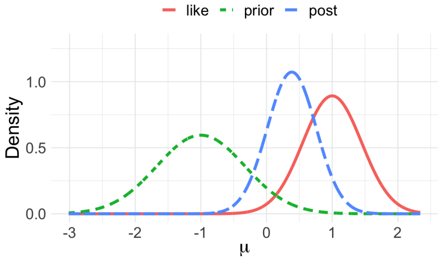
- Bayesian inference about a parameter $\mu$, using data $\boldsymbol{y}$:
  (1) relate data $\boldsymbol{y}$ to $\mu$ through a likelihood function $p(\boldsymbol{y}|\mu)$
  (2) set a prior distribution for $\mu$, $p(\mu)$
  (3) use Bayes' rule to get posterior distribution $p(\mu|\boldsymbol{y}) = \frac{p(\mu)p(\boldsymbol{y}|\mu)}{p(\boldsymbol{y})}$,
  (4) use the posterior $p(\mu|\boldsymbol{y})$ to provide summaries of interest, e.g. point estimates and uncertainty intervals, called credible intervals (CIs).
- Doing this for a normal-normal set-up:

$$\begin{aligned} y_i|\mu,\sigma^2 &\sim N(\mu,\sigma^2)(\text{independent});\ \mu \sim N(m_0, s_{\mu 0}^2); \\ \Rightarrow \mu|\boldsymbol{y},\sigma^2 &\sim N\left(\frac{m_0/s_{\mu 0}^2 + n\cdot\bar{y}/\sigma^2}{1/s_{\mu 0}^2 + n/\sigma^2}, \frac{1}{1/s_{\mu 0}^2 + n/\sigma^2}\right). \end{aligned}$$

- This module: derive the posterior, discuss the role of prior info (the module with the most equations)
- Note: conditioning on $\sigma$ is left out from equations to make it easier to focus on main points, $\sigma^2$ assumed known throughout.

# How did the normal posterior come about?

- Bayes' rule: $p(\mu|\boldsymbol{y}) = \frac{p(\mu)p(\boldsymbol{y}|\mu)}{p(\boldsymbol{y})}$
- Good news: To recognize the posterior as a normal distribution (or more generally, a specific parametric form for the density), we only have to consider the terms that include $\mu$: $p(\mu|\boldsymbol{y}) \propto p(\mu)p(\boldsymbol{y}|\mu)$




- $p(\mu|\boldsymbol{y})$ is a probability density function with $\int p(\mu|\boldsymbol{y})d\mu = 1$,
- what matters is how $p(\mu|\boldsymbol{y})$ depends on $\mu$
- $p(\boldsymbol{y})$ does NOT vary with $\mu$ and thus ends up being just a scaling factor in Bayes' rule.

## How did the normal posterior come about? (ctd)

- We work towards finding $m$ and $v$ in

$$\mu|\boldsymbol{y}, \sigma^2 \sim N(m, v),$$

so towards

$$p(\mu|\boldsymbol{y}) \quad \propto \quad \exp(-\frac{1}{2}f(\mu)),$$

where

$$f(\mu) = \frac{1}{v}(\mu - m)^2 \propto \frac{1}{v}(\mu^2 - 2m\mu).$$

- Re-arranging terms in the prior and likelihood will get you to that form, and thus $m$ and $v$
- We will use that summation in log-likelihood can be rewritten as follows:

$$\sum_{i=1}^{n}(y_i - \mu)^2 \quad \propto \quad \sum y_i \cdot \mu + n\mu^2 \propto n(\bar{y} - \mu)^2.$$

## Let's do it!

- Starting with Bayes' rule:

$$
\begin{aligned}
p(\mu|\boldsymbol{y}) &\propto p(\mu)p(\boldsymbol{y}|\mu) \\
&\propto \exp\left(\frac{-1}{2s_{\mu 0}^2}(\mu - m_0)^2\right) \cdot \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right), \\
&\propto \exp\left(\frac{-1}{2s_{\mu 0}^2}(\mu - m_0)^2\right) \cdot \exp\left(\frac{-n}{2\sigma^2}(\bar{y} - \mu)^2\right) \\
&\propto \exp\left(-\frac{1}{2}f(\mu)\right),
\end{aligned}
$$

where

$$
f(\mu) = 1/s_{\mu 0}^2(\mu^2 - 2 \cdot m_0 \cdot \mu) + n/\sigma^2(\mu^2 - 2 \cdot \bar{y} \cdot \mu).
$$

# Re-arranging terms, toward $f(\mu) \propto \frac{1}{v}(\mu^2 - 2m\mu)$

- We have

$$\begin{aligned} f(\mu) &= 1/s_{\mu 0}^2(\mu^2 - 2 \cdot m_0 \cdot \mu) + n/\sigma^2(\mu^2 - 2 \cdot \bar{y} \cdot \mu) \\ &\propto (1/s_{\mu 0}^2 + n/\sigma^2)\mu^2 - 2 \cdot (1/s_{\mu 0}^2 m_0 + n/\sigma^2 \bar{y}) \cdot \mu, \end{aligned}$$

- Let $v = (1/s_{\mu 0}^2 + n/\sigma^2)^{-1}$, then

$$f(\mu) \propto 1/v(\mu^2 - v \cdot 2 \cdot (1/s_{\mu 0}^2 m_0 + n/\sigma^2 \bar{y})\mu).$$

- Comparing this to $f(\mu) \propto \frac{1}{v}(\mu^2 - 2m\mu)$, we recognize that

$$\mu | \boldsymbol{y} \sim N\left( \frac{m_0/s_{\mu 0}^2 + n \cdot \bar{y}/\sigma^2}{1/s_{\mu 0}^2 + n/\sigma^2}, \frac{1}{1/s_{\mu 0}^2 + n/\sigma^2} \right).$$

# Summary so far

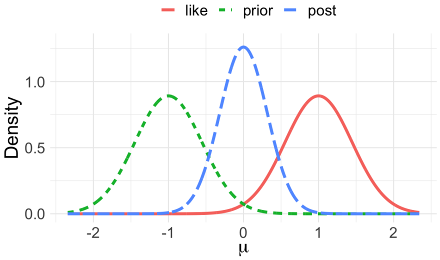
- Bayes' rule simplifies to: $p(\mu|\boldsymbol{y}) \propto p(\mu)p(\boldsymbol{y}|\mu)$
- We derived the posterior in the normal-normal set-up:

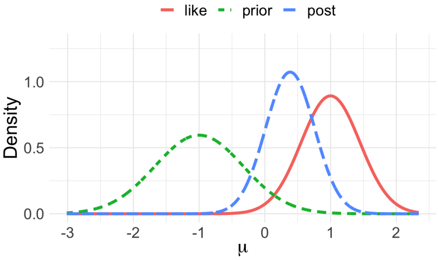
$$\begin{aligned} y_i|\mu,\sigma^2 &\sim N(\mu,\sigma^2)(\text{independent}); \\ \mu &\sim N(m_0, s^2_{\mu 0}); \\ \Rightarrow \mu|\boldsymbol{y},\sigma^2 &\sim N\left(\frac{m_0/s^2_{\mu 0} + n\cdot\bar{y}/\sigma^2}{1/s^2_{\mu 0} + n/\sigma^2}, \frac{1}{1/s^2_{\mu 0} + n/\sigma^2}\right). \end{aligned}$$

- Now let's examine the expressions for the posterior mean and variance

# Posterior variance

- Interpretation is easiest in terms of precision $= 1/\text{variance}$
- Posterior precision $= 1/s_{\mu 0}^2 + n/\sigma^2$;
  This is the sum of the prior precision and the precision of $\bar{y}$
- Examples: same likelihood with $\bar{y} = 1$; prior differs



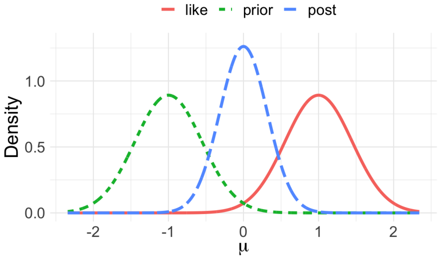



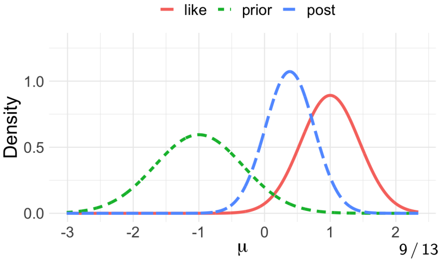
# Posterior mean

- Rewrite as weighted combination of prior mean and $\bar{y}$:

$$E(\mu|\boldsymbol{y}) = \frac{m_0/s^2_{\mu 0} + n \cdot \bar{y}/\sigma^2}{1/s^2_{\mu 0} + n/\sigma^2} = \frac{1/s^2_{\mu 0}}{1/s^2_{\mu 0} + n/\sigma^2} m_0 + \frac{n/\sigma^2}{1/s^2_{\mu 0} + n/\sigma^2} \bar{y},$$

  with the weight assigned to the prior mean (sample mean) given by the share of prior (sample mean) precision out of the overall precision.
- Examples: same likelihood with $\bar{y} = 1$; prior differs



# If helpful, an informal way to see how we arrive at the posterior

- Posterior $\propto$ prior $\cdot$ likelihood
- What shows up in the posterior after the rewrite of the likelihood?

$$p(\mu|\boldsymbol{y}) \quad \propto \quad \exp\left(\frac{-1}{2s_{\mu 0}^2}(\mu - m_0)^2\right) \cdot \exp\left(\frac{-n}{2\sigma^2}(\mu - \bar{y})^2\right).$$

- The likelihood contribution to the posterior is equivalent to a contribution of "$\mu \sim N(\bar{y}, \sigma^2/n)$"
- So posterior can be considered as arising from product of $\mu \sim N(m_0, s_{\mu 0}^2)$ and $\mu \sim N(\bar{y}, \sigma^2/n)$
- Posterior precision $= 1/s_{\mu 0}^2 + n/\sigma^2$, sum of the prior precision and precision coming from contribution of the likelihood
- Posterior mean $=$ precision-weighted means

# Summary (ctd)

- Bayes' rule simplifies to: $p(\mu|\boldsymbol{y}) \propto p(\mu)p(\boldsymbol{y}|\mu)$
- We derived the posterior in the normal-normal set-up:

$$\begin{aligned} y_i|\mu,\sigma^2 &\sim N(\mu,\sigma^2)\text{(independent)}; && (1)\\ \mu &\sim N(m_0, s_{\mu 0}^2); && (2)\\ \Rightarrow \mu|\boldsymbol{y},\sigma^2 &\sim N\left(\frac{m_0/s_{\mu 0}^2 + n\cdot\bar{y}/\sigma^2}{1/s_{\mu 0}^2 + n/\sigma^2}, \frac{1}{1/s_{\mu 0}^2 + n/\sigma^2}\right). && (3) \end{aligned}$$

- Posterior mean and variance determined by prior and sample mean and variance

## Sampling properties of Bayesian estimators

- A Bayesian point estimate for $\mu$ is given by $E(\mu|\boldsymbol{y})$.
- What are the sampling properties of this estimator, and how do they compare to those of the maximum likelihood estimator $\bar{y}$?
  - Estimator $=$ $E(\mu|\boldsymbol{y})$ as a function of the yet-to-be-observed data
  - Sampling properties $=$ behavior under hypothetically repeatable surveys or experiments
- Consider bias and mean squared error (MSE):
  - Bias $=$ expected value of estimator - (unknown) true value
  - MSE $=$ squared difference between expected value and the (unknown) true value; MSE $=$ bias$^2$ $+$ variance of estimator
- In a HW (optional) extra credit question, we show that in a normal-normal setting
  - Bayes estimator is biased if prior mean $\neq$ true value, $\bar{y}$ is unbiased
  - But prior information can result in smaller MSE for Bayes estimator, compared to $\bar{y}$

$\Rightarrow$ Illustration of bias-variance trade off of using prior information

## Summary and outlook

- Bayes' rule simplifies to: $p(\mu|\boldsymbol{y}) \propto p(\mu)p(\boldsymbol{y}|\mu)$
- We derived the posterior in the normal-normal set-up:

$$
\begin{aligned}
y_i|\mu,\sigma^2 &\sim N(\mu,\sigma^2)\text{(independent)};\\
\mu &\sim N(m_0, s_{\mu 0}^2);\\
\Rightarrow \mu|\boldsymbol{y},\sigma^2 &\sim N\left(\frac{m_0/s_{\mu 0}^2 + n\cdot\bar{y}/\sigma^2}{1/s_{\mu 0}^2 + n/\sigma^2}, \frac{1}{1/s_{\mu 0}^2 + n/\sigma^2}\right),
\end{aligned}
$$

  and discussed how the posterior depends on prior information and data.
- Next: sampling-based approach to do inference in settings where posterior is available in closed form