The normal-normal model

Giorgio Corani

Bayesian Data Analysis and Probabilistic Programming

Credits

- Chap. 5 of *Bayes Rules! An Introduction to Applied Bayesian Modeling*
  - https://www.bayesrulesbook.com/chapter-5.html

The Normal model

- Let Y be a continuous random variable which can take values in $(-\infty, \infty)$
- The variability of Y might be well represented by a Normal model

$$Y \sim N(\mu, \sigma^2)$$

The Normal model

- The Normal pdf is

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$$

* It has the following features:

$$E(Y) = Mode(Y) = \mu$$
$$Var(Y) = \sigma^2$$
$$SD(Y) = \sigma$$

Standard deviation σ

- σ provides a sense of scale for Y.
- Roughly 95% of Y values are within 2 standard deviations of μ:

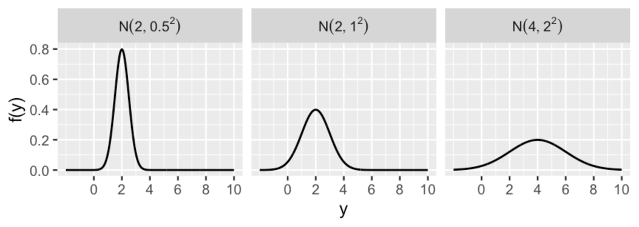
$$\mu \pm 2\sigma$$

- Roughly 99% of Y values are within 3 standard deviations of μ:

$$\mu \pm 3\sigma$$

The normal model

- The Normal model is bell-shaped and symmetric around μ.
- As σ gets larger, the pdf becomes more spread out.
- Though a Normal variable is defined in $(-\infty, \infty)$, the plausibility of values that are more than 3 standard deviations σ from the mean μ is negligible.



Example

- Assume we are interested in μ, the average volume (in cubic centimeters) of a specific part of the brain: the hippocampus. This is often researched in studies about the effect of concussions.
- In the general population, both halves of the hippocampus have a volume between 3.0 and 3.5 cubic centimeters.
- Thus, the total hippocampal volume of both sides of the brain is between 6 and 7 cm^3.

Prior

- Assuming it to be symmetrically distributed, we can summarize this information in our prior:

$$p(\mu) = N(6.5, 0.4)$$

or equivalently:

$$\mu \sim N(6.5, 0.4)$$

- We have set the mean of the normal at the midpoint of the interval (6-7).
- To tune the SD, we keep the $\mu \pm 2\sigma$ slightly larger than the our prior interval; this is a conservative choice and it is a good practice.

Prior

$$p(\mu) = N(6.5, 0.4)$$

- According to our prior information, μ (the mean volume in the population) is 6.5; our uncertainty is that however μ can generally lie in the interval 6.5 $\pm$ 2 ×0.4.
- We are assuming that our prior beliefs are appropriately represented by a Normal distribution.
- There is no single right prior, but multiple reasonable priors!

Likelihood

- Suppose that now we measure the hippocampal volumes of n=25 subjects.
- We make a *second* assumption of normality.
- The hippocampal volumes of our subjects $y_1, y_2, ..., y_n$, are independent and Normally distributed around the mean volume μ with standard deviation σ.

Likelihood

- σ expresses the spread of the hippocampal volumes within the population.
- Further research shows that most people have hippocampal volumes within 1.2 cm^3 of the average.
- We thus set σ= 0.6.

Likelihood

- For the moment we fix σ to a specific value. Later we will see how to set a prior also on σ.
- The dependence of y_i on the unknown mean μ is:
$$y_i \mid \mu \sim N(\mu, \sigma^2)$$
- The Normal model assumes that each subject's hippocampal volume can range from $-\infty$ to ∞. However, we're not too worried: it will put negligible weight on unreasonable values of hippocampal volume.

Likelihood

The joint pdf which describes the collective randomness in our n=25 subjects' hippocampal volumes, $(y_1, y_2, ..., y_n)$, is the product of the unique Normal pdfs $f(y_i \mid \mu)$:

$$f(\vec{y}|\mu) = \prod_{i=1}^{n} f(y_i|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i-\mu)^2}{2\sigma^2}\right].$$

The Normal-Normal Bayesian model

- Let μ be the unknown mean parameter and $(y_1, y_2, ..., y_n$ be an independent $N(\mu, \sigma^2)$ sample, where σ is known.
- The prior pdf of μ is also normal with fixed variance τ^2.

$$\mu \sim N(\theta, \tau^2)$$
$$y_i|\mu \overset{ind}{\sim} N(\mu, \sigma^2)$$

- Upon observing data $\rightarrow y_1, y_2, ..., y_n$ with mean $\bar{y}$, the posterior mean of μ is also Normal with updated parameters:

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

The Normal-Normal Bayesian model

- The posterior mean is a weighted average of the prior mean E(μ)=θ and the sample mean $\bar{y}$.
- The posterior variance is informed by the prior variability τ and variability in the data σ.
- As n increases, the posterior mean places less weight on the prior mean and more weight on sample mean.
- As n increases, the posterior variance decreases.

The Normal-Normal Bayesian model

- Assume that the sample of n measures has mean $\bar{y} = 6.7$.
- The posterior pdf of μ is:

conti da rifare

$$\mu|\vec{y} \ \sim \ N\left(6.5 \cdot \frac{0.5^2}{25 \cdot 0.4^2 + 0.5^2} + 5.735 \cdot \frac{25 \cdot 0.4^2}{25 \cdot 0.4^2 + 0.5^2}, \ \frac{0.4^2 \cdot 0.5^2}{25 \cdot 0.4^2 + 0.5^2}\right).$$

What if σ is unknown?

Prior distribution of σ

- σ is strictly positive; a suitable prior is the *half-normal* distribution.
- The half-normal is a Gaussian restricted to positive values.
- You sample from a half-normal by sampling from a normal distribution and rejecting the negative values (or applying the absolute value to the sampled values).

The half-normal distribution

- The half-normal prior is a suitable choice for the standard deviation.

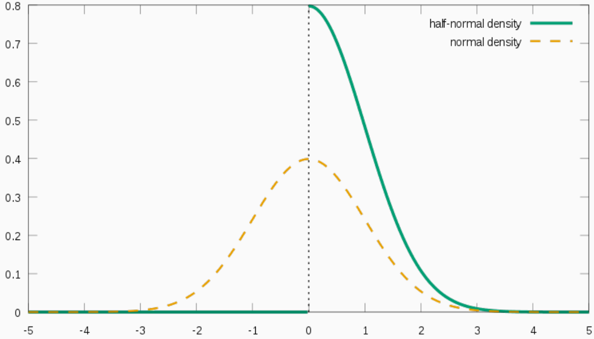
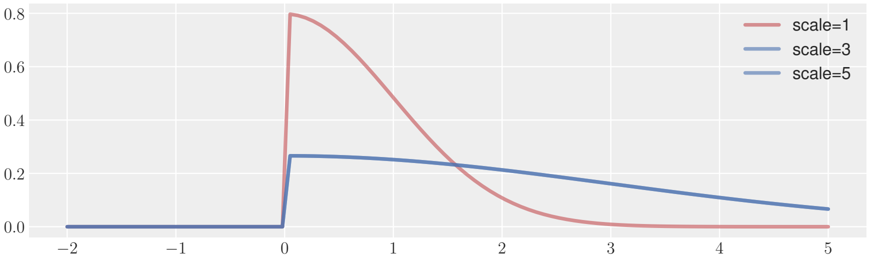


Figure 1: from wikipedia

The half-normal distribution

- The half-normal pdf is characterized by a scale parameter.



Tuning the half-normal distribution

- You can tune the HN prior distribution by matching its median with a plausible value of σ
- For instance we think a plausible value for the standard deviation of the noise is 7.5.
- with 95% probability the measures are lie in an interval of +- 15 around the mean.
- Of course, we are uncertain about this statament.
- Perhaps, with 95% probability the measures lie in an interval of +- 30, in which case the standard deviation of the noise is around 10.

Tuning the half-normal distribution

- The HN pdf is asymmetric right-skewed pdf; it has long tails which are much larger than the median.

- It thus offer a broad range of plausible values.
- The prior should cover a wide range of plausible values for σ, leaving out however values that make no sense.

Tuning the half-normal distribution

- The halfnormal distribution has been obtained by trying different scale parameters.
- Notice the long tails of the distribution, which allows to model to correct if our prior median guess (7.5) is underestimated.

```
pd.DataFrame(halfnorm.rvs(size=1000, scale=11)).describe()
```

```
##                  0
## count  1000.000000
## mean      8.575935
## std       6.442664
## min       0.043815
## 25%       3.464549
## 50%       7.353622
## 75%      12.104215
## max      37.500225
```

The probabilistic model

$$
\begin{aligned}
\mu &\sim N(\mu_\mu, \sigma_\mu) && \text{prior beliefs about } \mu \\
\sigma &\sim \text{Half-Normal}(\sigma_\sigma) && \text{prior beliefs about } \sigma \\
y &\sim \mathcal{N}(\mu, \sigma) && \text{the observation are affected by a noise with standard}
\end{aligned}
$$

- We cannot treat this model analytically, as the prior are no longer conjugates.
- We will implement it later via probabilistic programming.

Conceptual exercise

- Try to define the priors of a probabilistic model which represents the distribution of height of adult males in Switzerland

Population of Swiss adult males: $p(\mu)$

- The mean height of the population could be 175, though this is uncertain. Keeping our prior broad, the mean height of the population lies with 99% probability between 160 and 190 cm.
  - $\mu \sim \mathcal{N}(175, 5)$

Population of Swiss adult males: $p(\sigma)$

- We shall now assign a prior to σ. We assume that within the whole population the height varies with 99% probability between 100 and 250.
- Hence the corresponding value of the standard deviation is (250-100)/6 = 25.
- Notice the broad but sensible range.
- A half-normal distribution with scale 35 has roughly this median:
  - $\sigma \sim$ Half-Normal(35)

Population of Swiss adult males: $p(\sigma)$

```
pd.DataFrame(halfnorm.rvs(size=1000, scale=35)).describe()
```

```
##                  0
## count  1000.000000
## mean     28.351994
## std      21.145388
## min       0.261992
## 25%      12.035998
## 50%      24.165987
## 75%      39.688392
## max     116.359938
```

Likelihood

- The likelihood $y \sim \mathcal{N}(\mu, \sigma)$ requires no parameter specification.
- We are assuming that the measures are normally distributed around the mean.
- Moreover we assume that the measures are i.i.d.