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

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

- With:

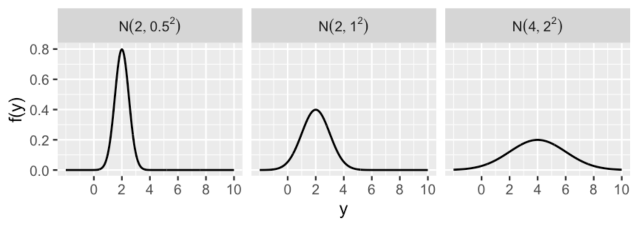
$$\begin{aligned} E(Y) &= Mode(Y) = \mu \\ Var(Y) &= \sigma^2 \\ SD(Y) &= \sigma \end{aligned}$$

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

- The volume of the hippocampus (a part of the brain) is 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 cm^3.
- Thus, the hippocampal volume is thought to vary, within the population, between 6 and 7 cm^3.
- The average volume μ is thought to be between 6.4 and 6.6 cm^3.

Normal prior

- Assuming symmetry, we formalize our prior information about μ as:

$$\mu \sim N(\mu', \sigma_\mu)$$
$$\mu \sim N(6.5, 0.05)$$

- σ_μ represents our prior uncertainty on the value of μ.
- μ' is our prior guess on the value of μ.
- According to this prior, there is about 95% probability of μ lying in (6.4, 6.6).
- There is generally no single right prior, but multiple reasonable priors.

Normal likelihood

- We now define a model for the distribution of the observations.
- We make a *second* assumption of normality.
- The observed volumes $y_1, y_2, ..., y_n$, are independent and normally distributed around μ, with standard deviation σ.
- σ expresses the spread of the hippocampal volumes within the population.
- As we expect y to vary between roughly 6 and 7, we set σ=0.25 (we interpret the interval as $\mu \pm 2\sigma$, hence it has length of 4σ).

Independence

- We assume the observations $y_1, ..., y_n$ to be *independent* samples from $N(\mu, \sigma)$.

Likelihood

Assuming independence, the joint pdf of the n measures ($y_1, y_2, ..., y_n$), is the product of the unique Normal pdfs $f(y_i \mid \mu)$:

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

- $\vec{y}$ is the vector containing the measures $y_1, ...y_n$.
- Theoretically, the normal model lets each hippocampal volume range from $-\infty$ to ∞. However it will assign negligible weight to values which are beyond $\mu \pm 3\sigma$.

The Normal-Normal model

$$\mu \sim N(\mu', \sigma_\mu)$$
$$\vec{y} \sim N(\mu, \sigma)$$

- For the moment we assume σ to be known and fixed. Later we will express our prior uncertainty also about it.

Conjugacy

- Denote the sample mean as $\bar{y} = \frac{1}{n} \sum_i y_i$.
- The posterior density of μ is normal with updated parameters:

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

Posterior mean

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

- The posterior mean is a weighted average of the prior mean μ' and the sample mean $\bar{y}$.
- As n increases, the posterior mean converges to $\bar{y}$.
- As n increases, the posterior variance decreases.

Your turn

- Assume that the sample of n measures has mean $\bar{y} = 6.7$.
- Which is the posterior mean?

What if σ is unknown?

- A more sophisticated approach is to treat σ as a parameter, by assigning a prior to it and make inference about it, rather than keeping it fixed.
- In this case there is no closed-form expression of the posterior.

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
  - applying the absolute value to the sampled values.

The half-normal distribution

- The HN pdf is asymmetric and right-skewed pdf.
- It has long tails which are much larger than the median.
- The prior should cover a wide range of plausible values for σ, leaving out however values that make no sense.

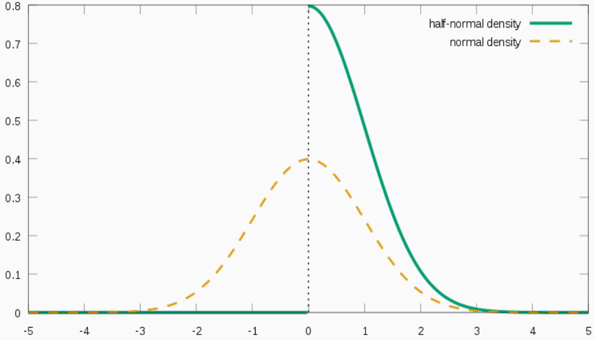
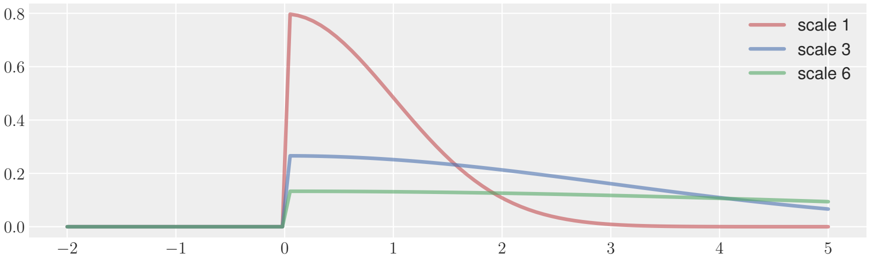


Figure 1: from wikipedia

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

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      8.796142
## std       6.675414
## min       0.009495
## 25%       3.304355
## 50%       7.504637
## 75%      12.704201
## max      35.607275
```

The probabilistic model

$\mu \sim N(\mu_\mu, \sigma_\mu)$	prior beliefs about μ
$\sigma \sim$ Half-Normal(σ_σ)	prior beliefs about σ
$y \sim \mathcal{N}(\mu, \sigma)$	the observation are affected by a noise with standard

- 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     27.042865
## std      19.776180
## min       0.086329
## 25%      11.900625
## 50%      22.551814
## 75%      39.137954
## max     104.828953
```

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.