

The normal-normal model

Giorgio Corani

Bayesian Data Analysis and Probabilistic
Programming

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

- 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:

$$E(Y) = Mode(Y) = \mu$$

$$Var(Y) = \sigma^2$$

$$SD(Y) = \sigma$$

Standard deviation σ

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

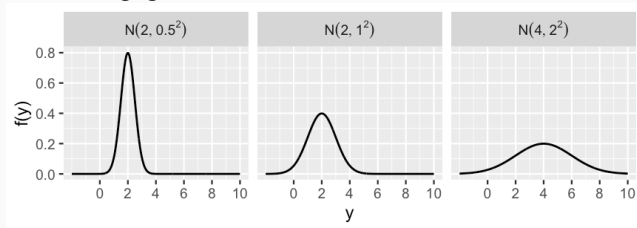
$$\mu \pm 2\sigma$$

- 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

- 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 cm³.
- Thus, the hippocampal volume is thought to vary, within the population, between 6 and 7 cm³.
- The average volume μ is thought to be between 6.4 and 6.6 cm³.

Normal prior

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

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

which in this example yields :

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

- μ' is our prior guess on the value of μ .
- σ_μ represents our uncertainty on the guess μ' .
- According to this prior, μ lies with 95% probability in (6.4, 6.6).
- We allow the volume to range over $(-\infty, \infty)$, but values beyond $\mu \pm 3\sigma$ are given negligible probability.

- There is no single right prior, but different reasonable priors.

- We now define a model for the distribution of the observations.
- We make a *second* assumption of normality.
- The hippocampal volumes observed in n subjects (y_1, y_2, \dots, y_n) are normally distributed $N(\mu, \sigma)$.

- μ is the mean volume in the population.
- σ expresses the spread of the measures within the population.
- We expect y to vary in (6-7); we interpret this interval as $\mu \pm 2\sigma$, hence it has length of 4σ .
- We thus set $\sigma=0.25$.

- We moreover assume the observations y_1, \dots, y_n to be *independent* samples from $N(\mu, \sigma)$.
- This is realistic: the measure y_i tells us nothing about the measure y_{i+1} (assuming they refer to different subjects)

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

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

- \vec{y} is the vector containing the measures y_1, \dots, y_n .

The Normal-Normal model

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

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

- We treat μ' , σ_μ and σ as fixed numbers.
- The only parameter of the model is μ .
- Later we will treat also σ as a parameter.

Your turn: normal likelihood functions

- For a Normal random sample $y_i \sim N(\mu, \sigma)$ with $\sigma=10$ we observe:

$$y_1, y_2, y_3 = (-4.3, 0.7, -19.4)$$

- Specify and plot the corresponding likelihood function of μ .

Conjugacy of the normal-normal model

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

$$\mu | \vec{y} \sim N \left(\underbrace{\mu' \frac{\sigma^2}{n\sigma_\mu^2 + \sigma^2} + \bar{y} \frac{n\sigma_\mu^2}{n\sigma_\mu^2 + \sigma^2}}_{\text{posterior mean}}, \underbrace{\frac{\sigma_\mu^2 \sigma^2}{n\sigma_\mu^2 + \sigma^2}}_{\text{posterior variance}} \right).$$

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

- Which is the posterior mean, if we did 5 measures with $\bar{y} = 6.7$?
- Which is the posterior mean, if we did 35 measures with $\bar{y} = 6.7$?

Your turn

- Let μ be the average 3 p.m. temperature in Lugano.
- Your friend's prior understanding is that μ is around 15 degrees Celsius, though might be anywhere between 5 and 25 degrees.
- To learn about μ , he will analyze 1000 days of temperature data.
- Letting y_i denote the 3 p.m. temperature on day i , they'll assume that daily temperatures vary Normally around μ with a standard deviation of 5 degrees.
- Formalize a normal-normal model.

Treating σ as a parameter

- A more sophisticated approach is to treat σ as a parameter.
- We assigning a prior to it; it should cover a wide range of plausible values for σ , leaving out however values that make no sense.
- In this case there is no closed-form expression of the posterior.

Half-normal distribution

- σ is strictly positive; a suitable prior is the *half-normal* distribution.
- The half-normal is a Gaussian restricted to positive values.
- Sample s from a half-normal are obtained by:
 - sampling from a normal distribution
 - applying the absolute value to the sampled values
 - $s \sim |N(0, \xi)|$, where ξ is the standard deviation of the underlying normal. It is referred to as the *scale* of the half-normal.

The half-normal distribution

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

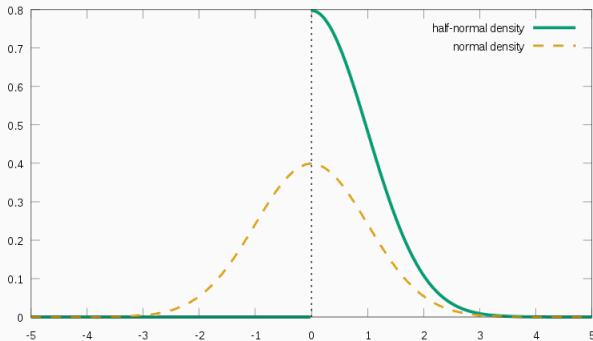
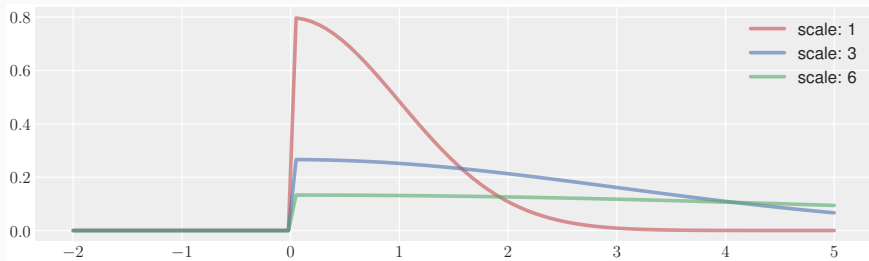


Figure 1: from wikipedia

Effect of the scale parameter

- The half-normal pdf is characterized by a scale parameter (the standard deviation of the underlying normal).



Tuning the half-normal distribution

- You can tune the scale of the HN by considering a plausible value of σ , and choose the scale so that it is close to the median of the HN.
- E.g., assume a plausible value of σ is 7.5.
- With 95% probability the measures are lie in an interval of ± 15 around the mean.
- But we are uncertain about this statement, as the interval could be well of ± 30 .

Tuning the half-normal distribution

- We try different scales, until the median is about 7.5.
- 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    8.965259
## std     6.679884
## min     0.002167
## 25%     3.713599
## 50%     7.845716
## 75%    12.530563
## max    39.893970
```

Probabilistic model with σ as parameter

$\mu \sim N(\mu_\mu, \sigma_\mu)$ prior beliefs about μ

$\sigma \sim \text{Half-Normal}(\sigma_\sigma)$ prior beliefs about σ

$y \sim N(\mu, \sigma)$ the observation are normally distributed σ

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

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

- The mean height of the population could be 175.
- Keeping our prior broad, we state the mean height of the population to lie with 99% probability between 160 and 190 cm (the 99% interval roughly corresponds to $\mu \pm 3\sigma$).
 - $\mu \sim N(175, 5)$

- We shall now assign a prior to σ . Within the population, we assume the height to lie with 99% probability between 100 and 250 (broad but realistic range).
- Hence the corresponding value of the standard deviation is $(250-100)/6 = 25$.

Tuning the half-normal

- A half-normal distribution with scale 35 has roughly median 25:

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

```
##              0
## count  1000.000000
## mean    27.837160
## std     21.099978
## min      0.036791
## 25%     11.594838
## 50%     23.323530
## 75%     39.355773
## max     117.077531
```

Likelihood (distribution of the data)

- Under the assumption of normality and independence, the likelihood is:

$$y \sim \mathcal{N}(\mu, \sigma)$$

- No further specification is required.

The resulting model

$$\mu \sim N(175, 5)$$

$$\sigma \sim \text{half-normal}(35)$$

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

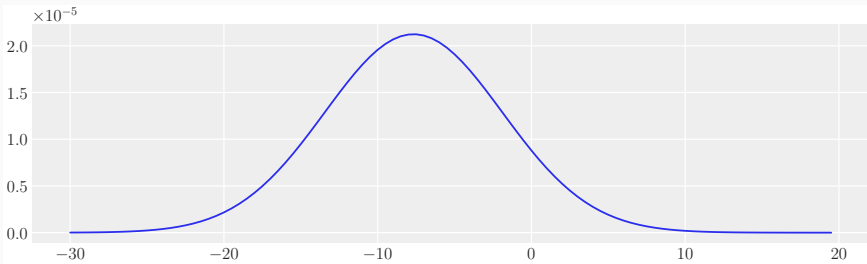
Solution for the exercise on normal likelihood functions

- Compute the likelihood as a function of μ for a normal sample with $\sigma=10$, given the observations:

$$y_1, y_2, y_3 = (-4.3, 0.7, -19.4)$$

Solution

```
#based on the observations, plausible values of mu  
#range between -30 and 20.  
mu = np.arange(-30, 20, 0.5)  
sigma = 10  
  
#a likelihood value for each value of mu  
lik = norm.pdf(-4.3, loc=mu, scale=sigma)  
  
#under independence, the likelihood of each observation multiplies  
lik = lik * norm.pdf(0.7, loc=mu, scale=sigma)  
lik = lik * norm.pdf(-19.4, loc=mu, scale=sigma)  
  
plt.figure(figsize=(10, 3))  
plt.plot(mu, lik)
```



- The values in the previous slide are numerically small. With more data, and more likelihood multiplication, it will become numerically untractable.
- For this reason it is numerically better to work with the log of the likelihood (log-likelihood).