Homework 1: Probabilistic ML

Marcus Hansen

November 2024

Exercise 1

Consider two boxes with white and black balls. Box 1 contains three black and five white balls and box 2 contains two black and five white balls. First a box is chosen at random with a prior probability p(box = 1) = p(box = 2) = 0.5, secondly a ball picked at random from that box. This ball turns out to be black. What is the posterior probability that this black ball came from box 1?

We want to calculate the posterior probability $p(\text{box} = 1 \mid \text{ball=black})$ using Bayes' theorem:

$$p(\text{box} = 1 \mid \text{ball=black}) = \frac{p(\text{ball=black} \mid \text{box} = 1) \cdot p(\text{box} = 1)}{p(\text{ball=black})}$$

The numerator consists of:

- **Likelihood:** $p(\text{ball=black} \mid \text{box} = 1) = \frac{3}{3+5} = \frac{3}{8}$
- **Prior:** $p(\text{box} = 1) = \frac{1}{2}$

The denominator is the marginal probability $p(\text{ball=black})$, which can be found using the law of total probability:

$$\begin{aligned} p(\text{ball=black}) &= p(\text{ball=black} \mid \text{box} = 1) \cdot p(\text{box} = 1) + p(\text{ball=black} \mid \text{box} = 2) \cdot p(\text{box} = 2) \\ &= \frac{3}{8} \cdot \frac{1}{2} + \frac{2}{7} \cdot \frac{1}{2} = \frac{37}{112} \end{aligned}$$

Then substitute the prior, likelihood, and marginal probability into Bayes' theorem:

$$p(\text{box} = 1 \mid \text{ball=black}) = \frac{\frac{3}{8} \cdot \frac{1}{2}}{\frac{37}{112}} \approx 0.568$$

So, the posterior probability that the black ball came from box 1 is approximately 0.568.

Exercise 2

The weather in Gothenburg can be summarized as follows: if it rains or snows one day, there is a 60% chance it will also rain or snow the following day; if it does not rain or snow one day, there is an 80% chance it will not rain or snow the following day either.

(i) Assuming that the prior probability it rained or snowed yesterday is 50%, what is the probability that it was raining or snowing yesterday given that it does not rain or snow today?

(ii) If the weather follows the same pattern as above, day after day, what is the probability that it will rain or snow on any day (based on an effectively infinite number of days of observing the weather)?

(iii) Use the result from part 2 above as a new prior probability of rain/snow yesterday and recompute the probability that it was raining/snowing yesterday given that it does not rain or snow today.

Definition: Here we define R as rain or snow for shorter notations and t is defined as today which means that t-1 is yesterday. From the exercise description we have:

$$p(weather_t = R|weather_{t-1} = R) = 0.6$$
$$p(weather_t = R^c|weather_{t-1} = R^c) = 0.8$$

Which implies:

$$p(weather_t = R^c|weather_{t-1} = R) = 0.4$$
$$p(weather_t = R|weather_{t-1} = R^c) = 0.2$$

as the sum of probability of R and R^c today, given that $weather_{t-1} = R^c$ needs to add up to 1, same argument goes for $weather_{t-1} = R$.

Answer (i):

From the description we are given the prior:

$$p(weather_{t-1} = R) = 0.5$$

and to are to calculate $p(weather_{t-1} = R|weather_t = R^c)$ which according to Bayes theorem is given by:

$$p(weather_{t-1} = R|weather_t = R^c) = \frac{p(weather_t = R^c|weather_{t-1} = R) \cdot p(weather_{t-1} = R)}{p(weather_t = R^c)} \tag{1}$$

in which $p(weather_t = R^c)$ is the only unknown. It is according to the law of total probability given by:

$$\begin{aligned} p(weather_t = R^c) &= p(weather_t = R^c|weather_{t-1} = R) \cdot p(weather_{t-1} = R) + \\ &\quad\; p(weather_t = R^c|weather_{t-1} = R^c) \cdot p(weather_{t-1} = R^c) \\ &= 0.4 \cdot 0.5 + 0.8 \cdot 0.5 \\ &= 0.6 \end{aligned} \tag{2}$$

Inserting the known probabilities into equation 1 we have:

$$p(weather_{t-1} = R|weather_t = R^c) = \frac{0.4 \cdot 0.5}{0.6} = \mathbf{\frac{1}{3}}$$

Answer (ii):

To answer ii) we want to calculate $p(weather_i = R)$ for any distant future day i. A starting point is to iteratively use the law of total probability in equation 2 for $p(weather_i = R)$. To investigate if we - given start a initial probability for $p(weather_0 = R)$ and $p(weather_0 = R^c)$, can find a value that $p(weather_i = R)$ converges to as $i \to \infty$. Using vectors and dot product we can formulate this problem as:

$$q_i = p_1 q_{i-1}$$

where

$$p_1 = \begin{bmatrix} p(weather_i = R|weather_{i-1} = R) & p(weather_i = R|weather_{i-1} = R^c) \end{bmatrix}$$

$$q_i = \begin{bmatrix} p(weather_{i-1} = R) & p(weather_{i-1} = R^c) \end{bmatrix}$$

This can be solved iteratively by calculating these for $i = 1, 2, ..., n$, where n is "big enough". **Or** one can identify that this is a Markov process where p_1 is the top row in the stochastic matrix P (see below) and q_i is a probability vector.

Given that this has been identified as a Markov process we can simplify the notations a bit and define the problem as:

$$P = \begin{bmatrix} P(R \to R) & P(R \to R^C) \\ P(R^C \to R) & P(R^C \to R^C) \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

where we want to find the steady state where $\pi = P\pi$. $\pi = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix}$ being the steady state vector of q_i when $i \to \infty$. We can then set up the system of equations as:

$$\begin{aligned} \pi_1 &= 0.6\pi_1 + 0.2\pi_2 \\ \pi_2 &= 0.4\pi_1 + 0.8\pi_2 \\ 1 &= \pi_1 + \pi_2 \end{aligned}$$

Solving them yields the steady state marginal likelihoods:

$$\pi = \begin{bmatrix} p(weather_\pi = R) \\ p(weather_\pi = R^c) \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \end{bmatrix} \tag{3}$$

meaning that at any given day i there's a third of a chance that it rains or snows. (Which seems way too low for Gothenburg.)

Answer (iii)

Given the new prior and (steady state) marginal likelihood (same as calculating the marginal likelihood again as it will just be another iteration, but since we are in steady state it will yield the same result and is therefore redundant), row one and two respectively in equation 3, we can calculate a new posterior, again using Bayes rule:

$$
\begin{aligned}
p(weather_{t-1} = R | weather_t = R^c) &= \frac{p(weather_t = R^c | weather_{t-1} = R) \cdot (weather_\pi = R)}{p(weather_\pi = R^c)} \\
&= \frac{\frac{2}{5} \cdot \frac{1}{3}}{2/3} = \mathbf{\frac{1}{5}}
\end{aligned}
$$

Exercise 3

Prove that the Beta distribution

$$\mathrm{Beta}(\mu; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}, \text{ where } \mu \in [0, 1], \tag{4}$$

$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx \tag{5}$$

is correctly normalized, i.e.,

$$\int_0^1 \mathrm{Beta}(\mu; a, b)d\mu = 1 \Longleftrightarrow \int_0^1 \mu^{a-1}(1-\mu)^{b-1}d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Answer We need to show that:

$$\int_0^1 \mathrm{Beta}(\mu; a, b)\, d\mu = 1,$$

Using equation (4):

$$\int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}\, d\mu = 1.$$

Since $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ is a constant.

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \mu^{a-1}(1-\mu)^{b-1}\, d\mu = 1.$$

Thus, it suffices to show that:

$$\int_0^1 \mu^{a-1}(1-\mu)^{b-1}\, d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

As

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = 1$$

From definition of $\Gamma(a)$ in equation (4)

$$
\begin{aligned}
\Gamma\left(a\right)\Gamma\left(b\right) &= \int_{u=0}^{\infty} e^{-u}u^{a-1}du \cdot \int_{v=0}^{\infty} e^{-v}v^{b-1}dv \\
&= \int_{v=0}^{\infty}\int_{u=0}^{\infty} e^{-u-v}u^{a-1}v^{b-1}dudv \\
\text{Doing the variable substitution} &\begin{cases} u = x\mu, \\ v = x(1-\mu) \end{cases} \\
\Gamma\left(a\right)\Gamma\left(b\right) &= \int_{x=0}^{\infty}\int_{\mu=0}^{1} e^{-x}(x\mu)^{a-1}(x(1-\mu))^{b-1}xd\mu dx \\
&= \int_{x=0}^{\infty} e^{-x}x^{a+b-1}dx \cdot \int_{\mu=0}^{1} \mu^{a-1}(1-\mu)^{b-1}d\mu \\
\text{From eq (5):} \qquad &= \Gamma\left(a+b\right) \cdot \int_{\mu=0}^{1} \mu^{a-1}(1-\mu)^{b-1}d\mu
\end{aligned}
$$

Which then gives that

$$
\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_{\mu=0}^{1} \mu^{a-1}(1-\mu)^{b-1}d\mu
$$

And proves that

$$
\int_0^1 \mathrm{Beta}(\mu; a, b)\, d\mu = 1,
$$

Exercise 4

Consider a variable $x \in \{0, 1\}$ and $p(x = 1) = \mu$ representing the flipping of a coin. With 50% probability, we think the coin is fair, i.e., $\mu = 0.5$, and with 50% probability, we think that it is unfair, i.e., $\mu \neq 0.5$. We encode this prior belief with the following prior:

$$P(\mu) = \frac{1}{2}\text{Beta}(\mu; 1, 1) + \frac{1}{2}\delta(\mu - 0.5)$$

(i) Assume that we get one observation $x_1 = 1$. What is the posterior $p(\mu \mid x_1)$? In particular, how does the belief of the fairness of the coin change under this observation?

(ii) Assume that we get one additional observation $x_2 = 1$. What is the posterior $p(\mu \mid x_1, x_2)$? In particular, how does the belief of the fairness of the coin change under this observation?

(iii) Compute the probability of the coin being fair by defining an event *fair* with the prior probability $p(\text{fair}) = 0.5$. Compute $p(\text{fair} \mid x_1, x_2)$ using Bayes' theorem based on the observations $x_1 = 1$ and $x_2 = 1$.

Answer (i)

A single coin toss can be represented by a Bernoulli random variable (RV):

$$p(x \mid \mu) = \text{Bern}(x; \mu) = \mu^x (1 - \mu)^{1-x} \tag{6}$$

We observe that the first toss results in $x_1 = 1$, and we want to find the posterior distribution $p(\mu \mid x_1)$, which can be obtained through Bayes' Theorem:

$$p(\mu \mid x_1) = \frac{p(x_1 \mid \mu)p(\mu)}{p(x_1)} \tag{7}$$

The likelihood function $p(x_1 = 1 \mid \mu) = \mu$, (using equation 6). The prior distribution is given in the question. The marginal likelihood $p(x_1)$ is found by integrating over all values of μ, as shown below:

$$\begin{aligned}
p(x_1 = 1) &= \int_0^1 p(x_1 = 1 \mid \mu)\, p(\mu)\, d\mu \\
&= \int_0^1 \mu \left(\frac{1}{2}\delta(\mu - 0.5) + \frac{1}{2}\,\text{Beta}(\mu; 1, 1) \right) d\mu \\
&= \frac{1}{2} \int_0^1 \mu\, \delta(\mu - 0.5)\, d\mu + \frac{1}{2} \int_0^1 \mu\, \text{Beta}(\mu; 1, 1)\, d\mu
\end{aligned}$$

The first integral is nonzero only when $\mu = 0.5$, and the second integral can be solved using the fact that $\text{Beta}(\mu; 1, 1)$ is uniform over the interval $[0, 1]$.

$$p(x_1 = 1) = \frac{1}{2} \cdot 0.5 + \frac{1}{2} \int_0^1 \mu\, d\mu = 0.25 + 0.25 = 0.5$$

Now, substituting everything back into Bayes' Theorem (equation 7), we find the posterior distribution:

$$p(\mu \mid x_1 = 1) = \mu\,(\text{Beta}(\mu; 1, 1) + \delta(\mu - 0.5))$$

Our belief has shifted slightly more toward the unfair coin, which is now represented by the $\mu Beta(1, 1)$ distribution. The left plot shows our initial belief, where μ could take any value between 0 and 1. After observing the data, the distribution has shifted slightly toward an unfair coin (in favor of heads), as seen in the right plot. Our belief in the fair coin ($\mu = 0.5$) represented by the δ function remains the same at 0.5.

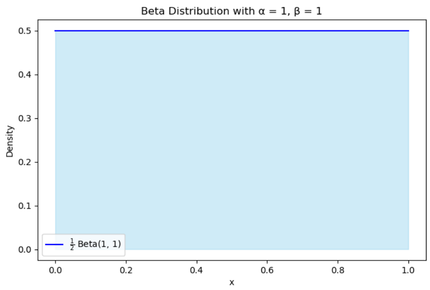


Figure 1: $0.5Beta(1, 1)$

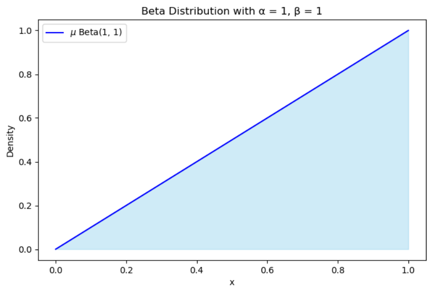


Figure 2: $\mu Beta(1, 1)$

Answer (ii)

In this question, we observe another toss $x_2 = 1$ and we want to calculate the posterior $p(\mu|x_1, x_2)$. The calculations are very similar to the last sub-question.

The likelihood function $p(x_1 = 1, x_2 = 1 \mid \mu) = \mu^2$, (using equation 6). The prior distribution is given in the question. The marginal likelihood $p(x_1, x_2)$ is found by integrating over all values of μ, as shown below:

$$\begin{aligned}
p(x_1 = 1, x_2 = 1) &= \int_0^1 p(x_1 = 1, x_2 = 1 \mid \mu)\, p(\mu)\, d\mu \\
&= \int_0^1 \mu^2 \left(\frac{1}{2}\delta(\mu - 0.5) + \frac{1}{2}\,\text{Beta}(\mu; 1, 1) \right) d\mu \\
&= \frac{1}{2} \int_0^1 \mu^2\, \delta(\mu - 0.5)\, d\mu + \frac{1}{2} \int_0^1 \mu^2\, \text{Beta}(\mu; 1, 1)\, d\mu \\
&= \frac{1}{2} \cdot 0.5^2 + \frac{1}{2} \int_0^1 \mu^2\, d\mu = \frac{1}{8} + \frac{1}{6} = \frac{7}{24}
\end{aligned}$$

Here we used the same reasoning as before. That $Beta(1, 1)$ is uniform over $[0, 1]$ and that $\delta(\mu - 0.5)$ is only non-zero when $\mu = 0.5$.

Now, substituting everything back into Bayes' Theorem, we find the posterior distribution:

$$p(\mu \mid x_1 = 1, x_2 = 1) = \left(\frac{24\mu^2}{14} \text{Beta}(\mu; 1, 1) + \frac{24\mu^2}{14} \delta(\mu - 0.5) \right)$$

After two tosses, where we got heads both times, our belief in the fair coin has decreased, reducing the factor in front of δ, ($\frac{24*0.25}{14} = 0.429$), and our belief in an unfair coin has increased, the constant in front of Beta. The left plot below shows our initial prior and the right plot shows how our belief in the unfair coin has shifted towards a more unfair coin (in favor of heads).

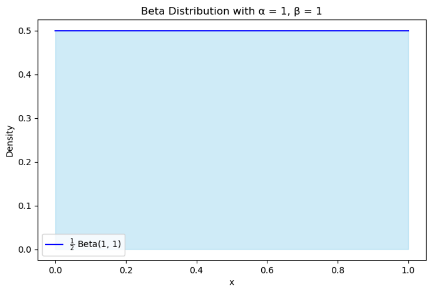


Figure 3: $0.5 Beta(1, 1)$

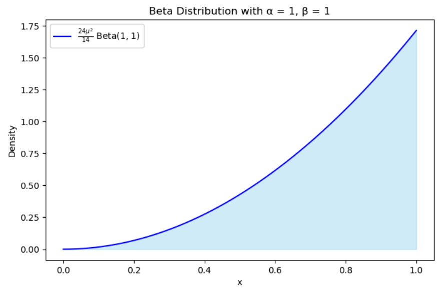


Figure 4: $\frac{24\mu^2}{14} Beta(1, 1)$

Answer (iii)

In this question, we want to compute the posterior probability $p(fair \mid x1, x2)$. This can also be done with Bayes theorem:

$$p(fair \mid x_1, x_2) = \frac{p(x_1, x_2 \mid fair) p(fair)}{p(x_1, x_2)} \tag{8}$$

The likelihood function $p(x_1, x_2 \mid fair(\mu = 0.5)) = 0.5^2 = 0.25$ and $p(x_1, x_2 \mid unfair(\mu \neq 0.5)) = \mu^2$, (using equation 6). The prior is given in the question, $p(fair) = p(\mu = 0.5) = 0.5$ which also means that $p(unfair) = p(\mu \neq 0.5) = 0.5$. The marginal likelihood $p(x_1, x_2)$ is found by integrating over all values of μ, as shown below:

$$\begin{aligned}
p(x_1 = 1, x_2 = 1) &= \int_0^1 p(x_1 = 1, x_2 = 1 \mid \mu)\, p(\mu)\, d\mu \\
&= \int_0^1 p(x_1 = 1, x_2 = 1 \mid fair)\, p(fair) + p(x_1 = 1, x_2 = 1 \mid unfair)\, p(unfair) \\
&= \int_0^1 (p(x_1 = 1, x_2 = 1 \mid \mu = 0.5)\, p(\mu = 0.5) + p(x_1 = 1, x_2 = 1 \mid \mu \neq 0.5)\, p(\neq 0.5))\, d\mu \\
&= \int_0^1 \left(0.25 * 0.5 + \mu^2 * 0.5\right) d\mu = \left(0.25 * 0.5 + \frac{1}{3} * 0.5\right) = \frac{7}{24}
\end{aligned}$$

Now, substituting everything back into Bayes' Theorem (equation 8), we find the posterior distribution:

$$p(fair \mid x_1, x_2) = \frac{0.25 * 0.5}{\frac{7}{24}} = \frac{3}{7} \approx 0.429$$

This result is reasonable because we tossed the coin twice and observed $x_1 = 1$ and $x_2 = 1$. Intuitively, $p(\text{fair} \mid x_1, x_2)$ should decrease, i.e. be below 0.5 based on the observed data. Our result of 0.429 is indeed below 0.5, which we consider reasonable.

Exercise 5

1. Theoretical Questions:

Derive the likelihood of observing a dataset $x = \{x_1, x_2, \ldots, x_n\}$ where each x_i follows a Poisson distribution with unknown rate parameter λ.

From the assignment description we know that $x_i \sim \text{Pois}(\lambda)$ meaning that the likelihood of observing an x_i in the view of the parameter λ is given by:

$$p(x_i|\lambda) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

With the assumption that all x_i's are independent we have the likelihood of observing the all the data x in given the parameter λ as:

$$p(x|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} \cdot e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} \cdot e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \tag{9}$$

Assume a Gamma prior on λ with shape parameter α and rate parameter β. Using Bayes' theorem, derive the posterior distribution for λ after observing the dataset x.

As per the assignment description we assume a Gamma prior on λ with shape parameter α and rate parameter β which yields:

$$p(\lambda; \alpha, \beta) = p(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda \in (0, \infty) \tag{10}$$

We can then use Bayes theorem to derive the posterior distribution for λ after observing the dataset x:

$$p(\lambda|x) = \frac{p(x|\lambda) \cdot p(\lambda)}{p(x)} \tag{11}$$

where the marginal likelihood $p(x)$ is the only unknown. Per definition the marginal likelyhood is given by:

$$p(x) = \int_{-\infty}^{\infty} p(x,\lambda)d\lambda = \int_{-\infty}^{\infty} p(x|\lambda)p(\lambda)d\lambda = \big\{\text{using eq. 9 \& 10}\big\} =$$

$$= \int_0^{\infty} \frac{\lambda^{\sum_{i=1}^n x_i} \cdot e^{-n\lambda}}{\prod_{i=1}^n x_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda = \Big\{S = \sum_{i=1}^n x_i\Big\} =$$

$$= \frac{\beta^\alpha}{\prod_{i=1}^n x_i\, \Gamma(\alpha)} \int_0^{\infty} \lambda^{S+\alpha-1} e^{-(n+\beta)\lambda} d\lambda = \Big\{ \int_0^{\infty} x^n e^{-ax}\, dx = \frac{\Gamma(n+1)}{a^{n+1}} (\text{wikipedia}) \Big\} =$$

$$= \frac{\beta^\alpha}{\prod_{i=1}^n x_i\, \Gamma(\alpha)} \cdot \frac{\Gamma(S+\alpha)}{(n+\beta)^{S+\alpha}} \tag{12}$$

Inserting expression 12 into equation 11 together with expression 9 and 10 we get the posterior of λ given observation of the data x as:

$$p(\lambda|x) = \frac{\frac{\lambda^S \cdot e^{-n\lambda}}{\prod_{i=1}^n x_i!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}}{\frac{\beta^\alpha}{\prod_{i=1}^n x_i\, \Gamma(\alpha)} \cdot \frac{\Gamma(S+\alpha)}{(n+\beta)^{S+\alpha}}}$$

$$= \frac{(n+\beta)^{S+\alpha}}{\Gamma(S+\alpha)} \lambda^{S+\alpha-1} e^{-(n+\beta)\lambda} \tag{13}$$

Show that this posterior distribution is also a Gamma distribution, and specify the updated shape and rate parameters of the posterior.

Comparing 13 and 10 we identify that the posterior is Gamma distributed and we can rewrite it as:

$$p(\lambda|x) = \frac{\beta'^\alpha}{\Gamma(\alpha')} \lambda^{\alpha'-1} e^{-\beta'\lambda} \tag{14}$$

with new rate and shape parameters:

$$\begin{aligned} \alpha' &= \alpha + S = \alpha + \sum_{i=1}^n x_i \\ \beta' &= \beta + n \end{aligned} \tag{15}$$

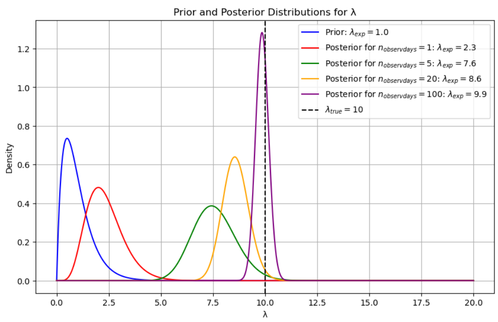
We think that one could say that the Gamma distribution serves as a conjugate prior for the Poisson likelihood, as it appears to yield a posterior distribution that is also Gamma. This enables straightforward updating of posteriors as new data x_{i+1} comes in, where the shape and rate parameters of the posterior could be directly calculated.

2. Programming Simulation:

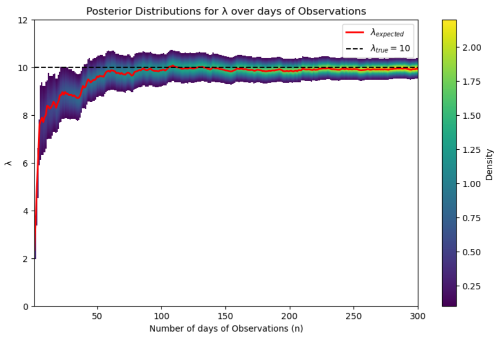
- Generate a synthetic dataset by simulating counts from a Poisson distribution with a true rate parameter λ_{true}.
- Assume a Gamma prior on λ with initial parameters $\alpha = 2$ and $\beta = 2$. After observing the synthetic data, update the prior parameters to obtain the posterior distribution.
- Write a Python function that, given data and prior parameters α and β, calculates the updated posterior parameters for λ.

3. Visualization:

- Plot the prior and posterior distributions for λ, clearly showing how observing the data has shifted your belief about the rate parameter.
- Compare the posterior distribution to the true value of λ_{true} used to generate the data. As you increase the number of observations, observe how the posterior distribution becomes more concentrated around λ_{true}.



(a) Gamma PDFs over different numbers of observed days.



(b) Expected λ while increasing the numbers of observed days. The density shows the PDF of the gamma function as number of days increases. The density is cut when the probability of the pdf is below 0.01.

Figure 5: Gamma distribution of the number of patients given different amounts of observed days.

4. Analysis:

Explain the impact of observing more data on the posterior distribution. How does the posterior distribution become more concentrated as the sample size grows?

As more data are observed, the posterior distribution gets closer to the true λ_{true}. It also becomes more concentrated (decreasing the variance of the gamma distribution), as seen in Figure 5.

However, this is only generally true, as the variable is random, which means that sometimes increasing the number of observed days can lead to an estimated λ_{exp} that is further from λ_{true}. Nevertheless, as $n \to \infty$, we have $\lambda_{\text{exp}} \to \lambda_{\text{true}}$, and the variance decreases.

The reason the variance decreases with larger n is related to how the α and β parameters of the gamma function increase, affecting the variance of the gamma distribution.

The variance of a gamma distribution is defined by

$$\sigma^2 = \frac{\alpha}{\beta^2}$$

Update rule of α, β as per equation 15.

$$\begin{aligned}
n &= \text{number of observed days} \\
x_i &= \text{number of patients observed day i} \\
\alpha_{updated} &= \alpha_{old} + \sum_{i=1}^{n} x_i, \text{Where } \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = \lambda, \text{Then } \alpha_{updated} = \alpha_{old} + \lambda n \\
\beta_{updated} &= \beta_{old} + n
\end{aligned}$$

So for large n

$$\alpha_{old} + \lambda n << (\beta_{old} + n)^2$$

which is why with growing n observations β^2 grows faster than α, reducing the variance of the gamma distribution.