

In [135...]

```

import jax.numpy as jnp
import numpy as np
import math
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.special import gammaln
from jax.scipy.special import beta as beta_func
from scipy.stats import beta as beta_dist

sns.set_theme(font_scale=1.25)

```

Group 13

- Mikkel Agerbo Essendrop, s194157
- Anton Mosquera Storgaard, s214611
- Gustav Hylsberg Jacobsen, s214374

All group members share equal responsibility for all three parts of this assignment. The work presented here represents a collective effort, and each member has contributed to understanding and completing the tasks.

Part 1

We record following data Customers: $N = 115$, Customers with purchase: $y = 4$, we want to deduce the probability that at least one of the next 20 Customers : $N^* = 20$ makes a purchase.

We define the following distributions

$$\theta \sim Beta(a_0, b_0) \quad (1)$$

$$y|\theta \sim Binomial(N, \theta) \quad (2)$$

we now that $a_0 = b_0 = 1$, so the prior is given as the $\theta \sim Uniform(0, 1)$

Task 1.1 Compute the prior mean of θ and provide a 95% credibility interval for the prior:

We could deduce that the prior is uniformly distributed where the mean is found by the following

$$\mathbb{E}(\theta) = \frac{a}{a+b} = \frac{1}{1+1} = \frac{1}{2}$$

When we are dealing with a uniform prior the credibility interval would simply equivalent to 95% of the possible outcomes within the distribution between 0 and 1 centered around the mean, thus the credibility interval becomes

$$(0.025, 0.975)$$

Task 1.2 Compute the posterior mean of θ and provide a 95% credibility interval for the posterior

We can start of by writing up our binomial likelihood PMF function (where $N = 115$ and $y = 4$)

$$p(y|\theta) = \binom{115}{4} p^4(1-p)^{115-4} \quad (3)$$

We know that the Beta distribution is conjugate on the likelihood which means our posterior distribution will follow the same distribution as our prior.

Our posterior is then given as

$$p(\theta|y, N) = \text{Beta}(a + y, b + N - y) \quad (4)$$

$$p(\theta|y, N) = \text{Beta}(5, 112) \quad (5)$$

We find the mean in the same way we found the mean for the Beta distribution in Task 1.1

$$\mathbb{E}(\theta|y, N) = \frac{5}{5+112} = \frac{5}{117} \approx 0.043$$

We use a snippet of the python code from handout 1 in order to determine the credibility interval

```
In [136...]: theta_lower, theta_upper = beta_dist.interval(95/100, a=5, b=112)
print(f'95%-credibility interval: [{theta_lower:.2f}:{theta_upper:.2f}]')
```

95%-credibility interval: [0.01;0.09]

So in conclusion the 95%-credibility interval for $(\theta|y, N)$ becomes: [0.01;0.09]

Task 1.3 Compute and plot the posterior predictive distribution for y^* given $y = 4$

We know that the posterior predictive distribution is given as

$$p(y^*|y) = \int p(y^*|\theta)p(\theta|y)d\theta \quad (6)$$

We choose to sample different values of θ from the posterior and then afterwards we sample new predictions y^* from the posterior predictive using the different values of θ

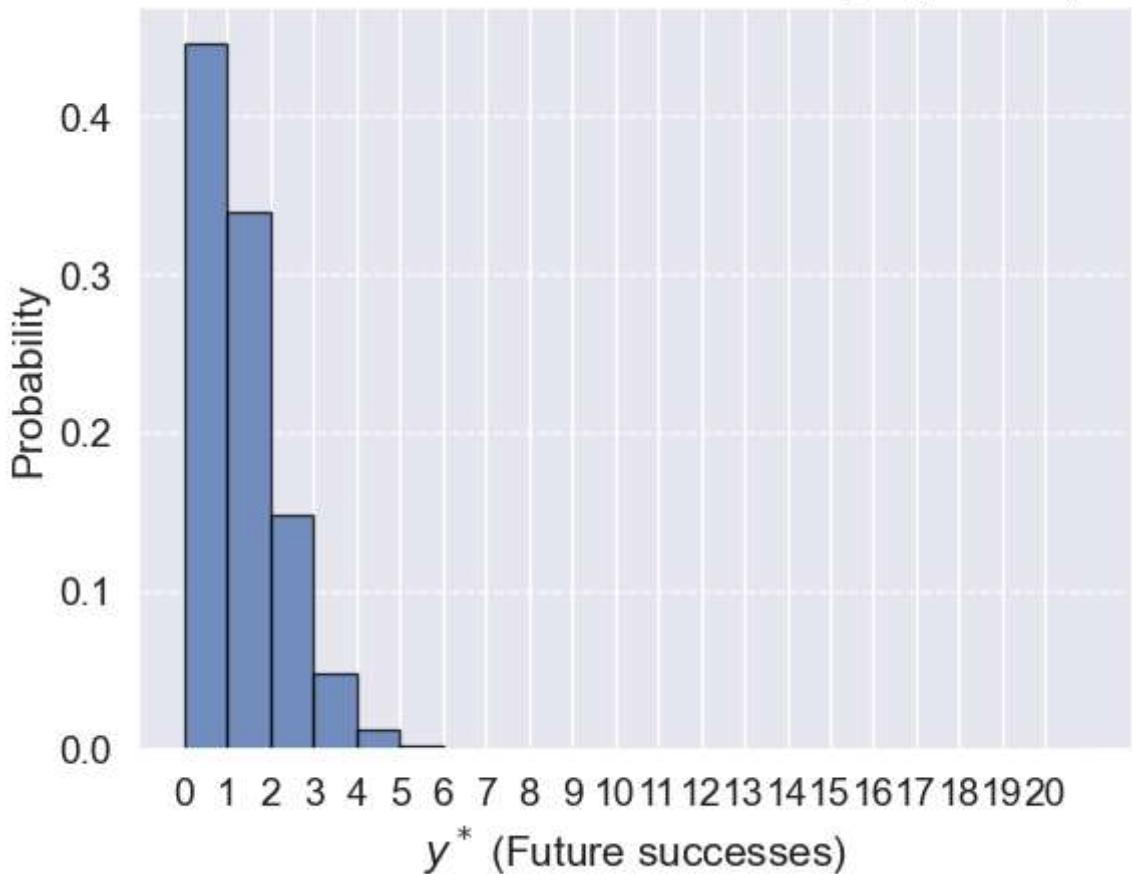
```
In [137...]: a, b = 5, 112 # Parameters of the posterior Beta distribution
N_star = 20 # Future trials

#We sample from the posterior distribution in order to visualize the posterior
#predictive with the samples we get.
theta_samples = np.random.beta(a, b, 100000)

#Once we have the posterior samples we can sample from the binomial predictive distribution.
y_star_samples = np.random.binomial(N_star, theta_samples)

# Step 3: Compute and plot the posterior predictive distribution
plt.hist(y_star_samples, bins=range(N_star+2), density=True, alpha=0.75, edgecolor='black')
plt.xlabel(r'$y^*$ (Future successes)')
plt.ylabel('Probability')
plt.title(f'Posterior Predictive Distribution for $y^*$ (N* = {N_star})')
plt.xticks(range(N_star+1))
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```

Posterior Predictive Distribution for y^* ($N^* = 20$)



We can also compute it analytically. Here $a = y + a_0$ and $b = N - y - b_0$. This can be done computing the integral:

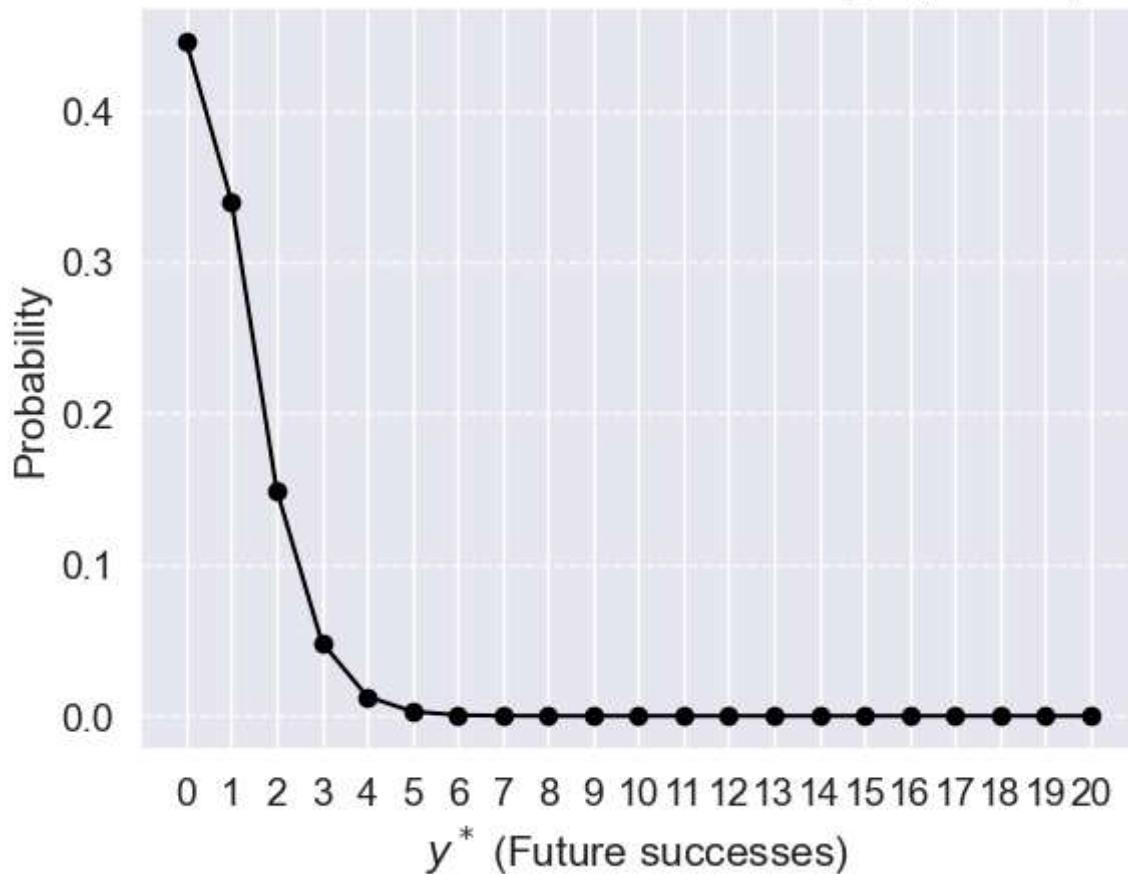
$$p(y^*|y) = \int p(y^*|\theta)p(\theta|y)d\theta = \binom{N^*}{k} \frac{B(a+k, b+N^*-k)}{B(a,b)} = \binom{20}{5} \frac{B(5+k, 132-k)}{B(5, 112)} \quad (7)$$

```
In [138]: def binomial_beta(k, a, b, N):
    return math.comb(N, k) * beta_func(a + k, b + N - k) / beta_func(a, b)

a, b = 5, 112
N_star = 20
k_values = np.arange(0, N_star+1)
post_pred_values = [binomial_beta(k, a, b, N_star) for k in k_values]

plt.plot(k_values, post_pred_values, 'o-', color='black')
plt.xlabel(r'$y^*$ (Future successes)')
plt.ylabel('Probability')
plt.title(f'Posterior Predictive Distribution for $y^*$ ($N^* = {N_star}$)')
plt.xticks(k_values)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```

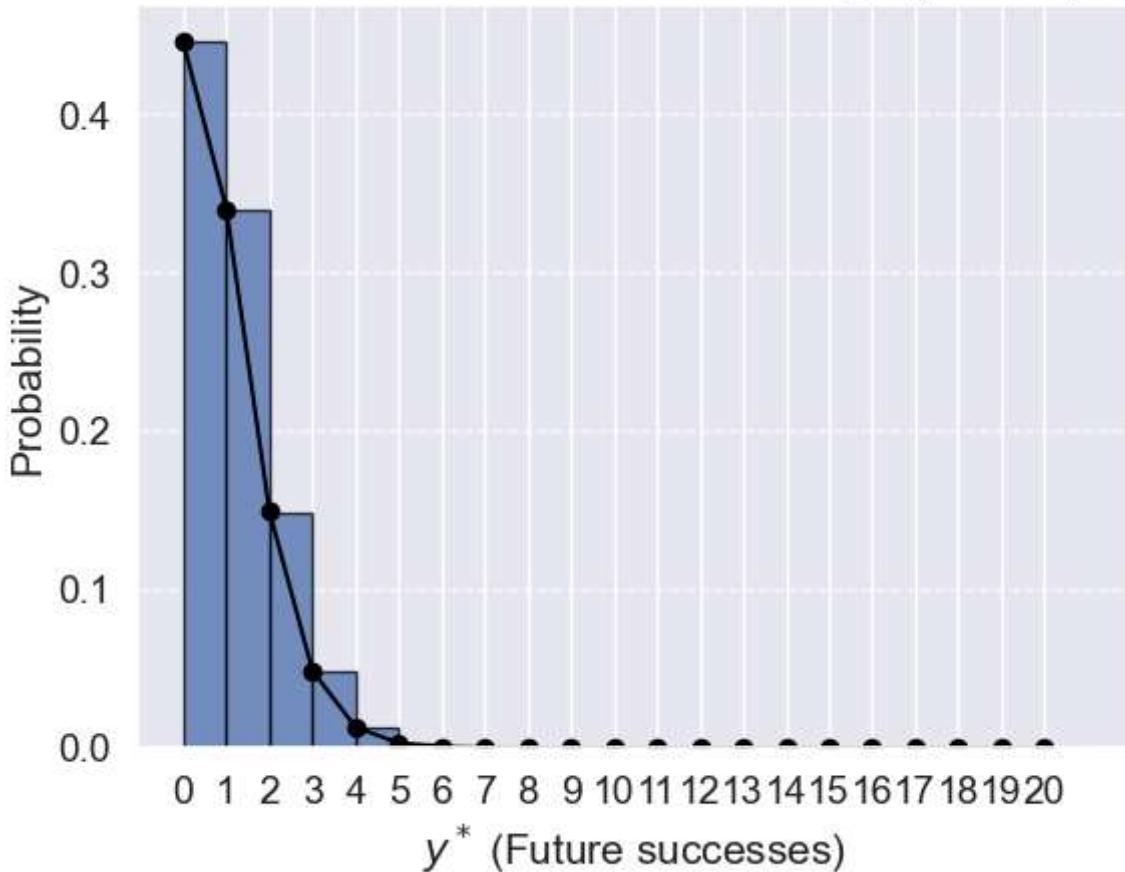
Posterior Predictive Distribution for y^* ($N^* = 20$)



The results line up nicely:

```
In [139]: plt.plot(range(N_star+1), post_pred_values, 'o-', color='black')
plt.hist(y_star_samples, bins=range(N_star + 2), density=True, alpha=0.75, edgecolor="black")
plt.xlabel(r"$y^*$ (Future successes)")
plt.ylabel("Probability")
plt.title(f"Posterior Predictive Distribution for $y^*$ (N* = {N_star})")
plt.xticks(range(N_star + 1))
plt.grid(axis="y", linestyle="--", alpha=0.7)
plt.show()
```

Posterior Predictive Distribution for y^* ($N^* = 20$)



Task 1.4 What is the posterior predictive probability that at least one of the next $N^* = 20$ will make a purchase?

We simply negate the question and says what is the probability that no customers made a purchase.

$$p(y^* \geq 1 | N^* = 20) = 1 - p(y^* = 0 | N^* = 20)$$

```
In [140]: print(f'Probability of at least one future success: {1 - np.mean(post_pred_values[0]):.3f}')
```

Probability of at least one future success: 0.554

This says the probability that at least one made a purchase is around 55.4%

1.5 Compute mean and variance of the posterior predictive distribution for y^*

We know that the posterior predictive distribution is distributed as follows

$$p(y^*) \sim BetaBinomial(N^*, a, b)$$

where the mean and variance is given as

$$\mathbb{E}(y^*) = N^* \frac{a}{a+b} \tag{8}$$

$$\sigma^2 = N^* \frac{ab(a+b+N^*)}{(a+b)^2(a+b+1)} \tag{9}$$

We plugin our values

$$\mathbb{E}(y^*) = 20 \cdot \frac{5}{5+112} \approx 0.85 \quad (10)$$

$$\sigma^2 = 20 \cdot \frac{5 \cdot 112(5+112+20)}{(5+112)^2(5+112+1)} \approx 0.95 \quad (11)$$

We verify the results by taking the mean and variance of our sampled values.

```
In [141]: print(f"Mean of posterior predictive: {np.mean(y_star_samples):.2f}")
print(f"Variance of posterior predictive: {np.var(y_star_samples):.2f}")
```

Mean of posterior predictive: 0.85
 Variance of posterior predictive: 0.95

So the mean of posterior predictive is around: 0.85 and the variance of posterior predictive is around: 0.95

Part 2

In part 2 we will be working with Linear Gaussian systems.

We let $z_1, z_2 \in \mathbb{R}^2$, and $y \in \mathbb{R}$ be random variables and we consider the following system

$$z_1 \sim \mathcal{N}(0, vI) \quad (1)$$

$$z_2|z_1 \sim \mathcal{N}(z_1, vI) \quad (2)$$

$$y|z_2 \sim \mathcal{N}(a^T z_2, \sigma^2) \quad (3)$$

where $a \in \mathbb{R}^2$ is constant. The joint distribution of (z_1, z_2, y) is given by

$$p(y, z_1, z_2) = p(y|z_2)p(z_2|z_1)p(z_1) \quad (4)$$

Let's try and put some words on the information provided here. For instance, we can see that the distribution of $z_2|z_1$ is conditionally dependent on z_1 .

We also see that y must be conditionally independent of z_1 since if z_2 is known, then z_1 is not needed, and thus we have independence. This allows us to write:

$$p(y|z_1, z_2) = p(y|z_2) \quad (5)$$

Task 2.1 Determine the distribution $p(y)$.

We start by wanting to compute $p(z_2)$ first, as this allows us to use the following equation:

$$p(y) = \int p(y|z_2)p(z_2)dz_2 \quad (6)$$

We make use of both the Conditional rule and the Sum rule where we have the following

$$P(b) = \int p(a, b)da \quad p(a|b) = \frac{p(a, b)}{p(b)} \quad (7)$$

This means that we can do this manipulation

$$p(z_2|z_1) = \frac{p(z_2, z_1)}{p(z_1)} \quad p(z_2, z_1) = p(z_2|z_1)p(z_1) \quad (8)$$

This expression can be substituted into the sum rule, and now we want to marginalize out the z_1 factor so we are left with the z_2 distribution

$$p(z_2) = \int p(z_2|z_1)p(z_1)dz_1 \quad (9)$$

Once we are here we can utilize the equations for bayes rules for gaussian given in Murphy (3.37)

$$\begin{aligned} p(z_2) &\sim \mathcal{N}(0 + z_1, vI + vI) \\ &\sim \mathcal{N}(z_1, 2vI) \\ &\sim \mathcal{N}(0, 2vI) \end{aligned} \quad (10)$$

We now complete our equation for $p(y)$ and marginalize out z_2 this time:

$$p(y) = \int p(y|z_2)p(z_2)dz_2 \quad (12)$$

However we see again that y is a linear transformation of a gaussian variable plus gaussian noise. This gives: $y \sim a^T z_2 + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma^2)$

It is easy to see that the mean of y becomes 0 as $E[y] = E[a^T z_2] + E[\epsilon] = a^T E[z_2] + 0 = 0$.

With the law of total variance (murphy (2.45)) we get the variance of y as:

$$Var(y) = Var(a^T z_2) + \sigma^2 \quad (13)$$

Then:

$$Var(a^T z_2) = a^T Cov(z_2)a = a^T(2vI)a = 2v(a^T a) \quad (14)$$

Therefore:

$$Var(y) = 2v(a^T a) + \sigma^2 \quad (15)$$

This gives us the marginal distribution of y as:

$$p(y) \sim \mathcal{N}(0, 2v(a^T a) + \sigma^2) \quad (16)$$

Task 2.2 Determine the distribution $p(y, z_2|z_1)$

We use the rules of conditional independence to determine this joint distribution. Lets start by using the chain rule of probability so we can rewrite

$$p(y, z_2|z_1) = p(y|z_2, z_1)p(z_2|z_1) \quad (12)$$

We remember our deduction in equation (5) and our expression becomes

$$p(y, z_2|z_1) = p(y|z_2)p(z_2|z_1) \quad (13)$$

When we multiply two gaussians we get a two-dimensional gaussian distribution, so our mean vector will consist of the means of y and z_2 , and our covariance matrix include the covariance between and the individual variances.

$$\mathcal{N} \left(\begin{bmatrix} \mathbb{E}(y|z_1) \\ \mathbb{E}(z_2) \end{bmatrix}, \begin{bmatrix} \sigma^2(y|z_1) & \Sigma(y, z_2) \\ \Sigma(y, z_2) & \sigma^2(z_2|z_1) \end{bmatrix} \right) \quad (14)$$

Here we see that the only value we are missing in order to plug in our distribution is the covariance. Covariance between y and z_2 is given by the following equation

$$\begin{aligned} \Sigma(y, z_2) &= \mathbb{E}[(y - \mathbb{E}[y|z_1])(z_2 - \mathbb{E}[z_2|z_1])^T] \\ &= \mathbb{E}[(a^T z_2 - a^T z_1)(z_2 - z_1)^T] \\ &= a^T \cdot \mathbb{E}[(z_2 - z_1)(z_2 - z_1)^T] \\ &= a^T \cdot vI \\ &= va^T \end{aligned} \quad (15)$$

Lets quickly explain some intuition behind step 2 and three in equation (15). We know that from our model that $y = a^T z_2$, so when we take the expected value we get

$$\mathbb{E}(y|z_1) = \mathbb{E}(a^T z_2) = a^T \mathbb{E}(z_2) = a^T z_1 \quad (16)$$

We also take advantage of the fact that $z_2 - z_1$ follow a gaussian noise model with variance vI so we can substitue that in. The uncertainties added in these steps offset each other due to difference in signs. Our final distribution becomes

$$p(y, z_2|z_1) \sim \mathcal{N} \left(\begin{bmatrix} a^T z_1 \\ z_1 \end{bmatrix}, \begin{bmatrix} va^T a + \sigma^2 & va^T \\ va & vI \end{bmatrix} \right) \quad (17)$$

where the covariance term has to be symmetric so we remove the transpose in the left corner.

Task 2.3 Determine the distribution $p(y|z_1)$

We have actually already found the distribution, the mean and variance can be extracted from the distribution we just described, by looking at the mean and variance in the vector and matrix.

$$p(y|z_1) \sim \mathcal{N}(a^T z_1, \sigma^2 + va^T a) \quad (18)$$

Task 2.4 Determine the distribution $p(z_1|y)$

Lets start by writing up bayes theorem for this distribution given as

$$p(z_1|y) = \frac{p(y|z_1)p(z_1)}{p(y)} \quad (19)$$

We again use the equations described in Murphy and we see that for the gaussian linear models posterior distribution described as

$$p(z_1|y) \sim \mathcal{N}(\mu_{z_1|y}, \Sigma_{z_1|y}) \quad (20)$$

We have the following

$$\Sigma_{z|y}^{-1} = \Sigma_z^{-1} + \mathbf{W}^T \Sigma_y^{-1} \mathbf{W} \quad (21)$$

$$\mu_{z|y} = \Sigma_{z|y} \left[\mathbf{W}^T \Sigma_y^{-1} (y - b) + \Sigma_z^{-1} \mu_z \right] \quad (22)$$

We now have to substitute in our known values in these expressions, we start by substituting in our terms for the variance and we end up with

$$\begin{aligned}\Sigma_{z_1|y}^{-1} &= (vI)^{-1} + a^T(va^T a + \sigma^2)^{-1}a, \\ &= \frac{1}{v}I + \frac{aa^T}{va^T a + \sigma^2}.\end{aligned}\quad (23)$$

We flip the exponent so we get

$$\Sigma_{z_1|y} = \left(\frac{1}{v}I + \frac{aa^T}{va^T a + \sigma^2} \right)^{-1}. \quad (24)$$

After having found this term we can now substitute our values into the expression for the mean given in equation (22)

$$\begin{aligned}\mu_{z_1|y} &= \Sigma_{z_1|y} [a^T(va^T a + \sigma^2)^{-1}y + (vI)^{-1} \cdot 0] \\ &= \Sigma_{z_1|y} \left(\frac{a}{va^T a + \sigma^2} y \right)\end{aligned}\quad (25)$$

Leaving us with a final distribution given as

$$p(z_1|y) \sim \mathcal{N} \left(\frac{va}{va^T a + \sigma^2} y, \left(\frac{1}{v}I + \frac{aa^T}{va^T a + \sigma^2} \right)^{-1} \right). \quad (26)$$

Part 3

In part 3 we will be working with the Poisson-Gamma model:

$$\begin{aligned}p(y_i|\lambda) &= \frac{\lambda^{y_i}}{y_i!} \exp(-\lambda), \\ p(\lambda|a_0, b_0) &= \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda^{a_0-1} \exp(-b_0\lambda),\end{aligned}$$

where $y_i \in \mathbb{N}_0$ and $\lambda, a_0, b_0 \in \mathbb{R}^+$.

Task 3.1 Determine the joint distribution $p(\mathbf{y}, \lambda)$.

Collecting a set of N observations of y_i and assuming each of them are independent conditioned on λ , the probability of observing $\mathbf{y} = \{y_i\}_{i=1}^N$ is given by

$$p(\mathbf{y}|\lambda) = \prod_{i=1}^N p(y_i|\lambda)$$

The joint distribution can be written as a conditional distribution

$$p(\mathbf{y}, \lambda) = p(\mathbf{y}|\lambda)p(\lambda)$$

Studying the Poisson-Gamma model we get

$$\begin{aligned}
p(\mathbf{y}, \lambda | a_0, b_0) &= \prod_{i=1}^N \frac{\lambda^{y_i}}{y_i!} \exp(-\lambda) \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda^{a_0-1} \exp(-b_0 \lambda) \\
&= \frac{b_0^{a_0}}{\Gamma(a_0)} \frac{\lambda^{(a_0-1+\sum_{i=1}^N y_i)}}{\prod_{i=1}^N y_i!} \exp(-(N+b_0)\lambda) \\
&\stackrel{(*)}{=} \frac{b_0^{a_0}}{\Gamma(a_0)} \frac{\lambda^{(a_0-1+N\bar{y})}}{\prod_{i=1}^N y_i!} \exp(-(N+b_0)\lambda),
\end{aligned}$$

where we have used the average $\bar{y} = \frac{1}{N} \sum_i^N y_i$ in (*) to simplify the expression.

From this we also see that the posterior must have the same form as the prior, since it is independent of \mathbf{y} , hence they are conjugate wrt. $p(\mathbf{y}|\lambda)$ as expected.

$$p(\lambda|\mathbf{y}) \propto \lambda^{(a_0-1+N\bar{y})} \exp(-(N+b_0)\lambda) \propto \text{Gamma}(a_0 + N\bar{y}, b_0 + N)$$

Task 3.2 Functional form of Gamma distribution.

For numerical stability, the log transform is often used. For a Gamma distribution the log transform is as follows:

$$\begin{aligned}
\log p(\lambda|a, b) &= \log \left(\frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp(-b\lambda) \right) \\
&= \underbrace{a \log b - \log \Gamma(a)}_{\text{constant}} + (a-1) \log \lambda - b\lambda
\end{aligned}$$

Task 3.3 Analytical expression for the posterior distribution.

From Bayes theorem we know that

$$p(\lambda|\mathbf{y}) = \frac{p(\mathbf{y}|\lambda)p(\lambda)}{p(\mathbf{y})},$$

where $p(y)$ is the evidence, acting as a normalisation factor.

$$p(\mathbf{y}) = \int_{\lambda} p(\mathbf{y}|\lambda)p(\lambda) d\lambda$$

We can then evaluate the integral of the joint distribution from task 3.1, since it is just a new gamma function

$$\begin{aligned}
p(y) &= \frac{b_0^{a_0}}{\Gamma(a_0) \prod_{i=1}^N y_i!} \int_{\lambda} \lambda^{(a_0-1+N\bar{y})} \exp(-(N+b_0)\lambda) d\lambda \\
&= \frac{b_0^{a_0}}{\Gamma(a_0) \prod_{i=1}^N y_i!} \frac{\Gamma(a_0 + N\bar{y})}{(b_0 + N)^{a_0 + N\bar{y}}}.
\end{aligned}$$

To ease the notation, we will denote $a_1 = a_0 + N\bar{y}$ and $b_1 = b_0 + N$. Normalising the joint distribution found in task 3.1, we can then derive the posterior distribution

$$\begin{aligned}
p(\lambda|\mathbf{y}) &= \frac{\overbrace{p(\mathbf{y}|\lambda)p(\lambda)}^{p(\mathbf{y},\lambda)}}{p(\mathbf{y})} \\
&= \underbrace{\frac{b_0^{a_0}}{\Gamma(a_0)} \frac{\lambda^{a_1-1}}{\prod_{i=1}^N y_i!} \exp(-b_1\lambda)}_{p(\mathbf{y},\lambda)} \left(\underbrace{\frac{b_0^{a_0}}{\Gamma(a_0) \prod_{i=1}^N y_i!} \frac{\Gamma(a_1)}{(b_1)^{a_1}}}_{p(\mathbf{y})} \right)^{-1} \\
&= \frac{b_1^{a_1}}{\Gamma(a_1)} \lambda^{a_1-1} \exp(-b_1\lambda) \\
&= \text{Gamma}(a_0 + N\bar{y}, b_0 + N)
\end{aligned}$$

We see that this is consistent with the conjugate statement about the posterior in task 3.1.

Task 3.4 Posterior distribution of λ given the data:

$\mathbf{y}_{\text{sample}} = \{y_1 = 7, y_2 = 4, y_3 = 8, y_4 = 11, y_5 = 12\}$, as well as assuming $a_0 = 1$ and $b_0 = 1/10$.

The expected value of the posterior is given by

$$\begin{aligned}
\mathbb{E}[\lambda] &= \int_{\lambda} \lambda \frac{b_1^{a_1}}{\Gamma(a_1)} \lambda^{(a_1-1)} \exp(-b_1\lambda) d\lambda \\
&= \frac{b_1^{a_1}}{\Gamma(a_1)} \int_{\lambda} \lambda^{a_1} \exp(-b_1\lambda) d\lambda \\
&= \frac{b_1^{a_1}}{\Gamma(a_1)} \frac{\Gamma(a_1+1)}{b_1^{(a_1+1)}} \\
&\stackrel{(**)}{=} \frac{a_1}{b_1} \\
&= \frac{a_0 + N\bar{y}}{b_0 + N},
\end{aligned}$$

where in (**) we have used $\Gamma(z+1) = z\Gamma(z)$.

We can then determine the mean of λ conditioned on the given data. The posterior distribution, given $\mathbf{y}_{\text{sample}}$, is

$$p(\lambda|\mathbf{y}_{\text{sample}}) = \text{Gamma}(a_0 + N\bar{y}, b_0 + N) = \text{Gamma}(43, 5.1),$$

and the mean given the data

$$\mathbb{E}[\lambda] = \frac{a_0 + N\bar{y}}{b_0 + N} = \frac{43}{5.1} = 8.4$$

Task 3.5

```
In [142...]: y = jnp.array([7, 4, 8, 11, 12])
N = len(y)
a0, b0 = 1, 0.1

lambda_ = jnp.linspace(0, 30, 1000)
```

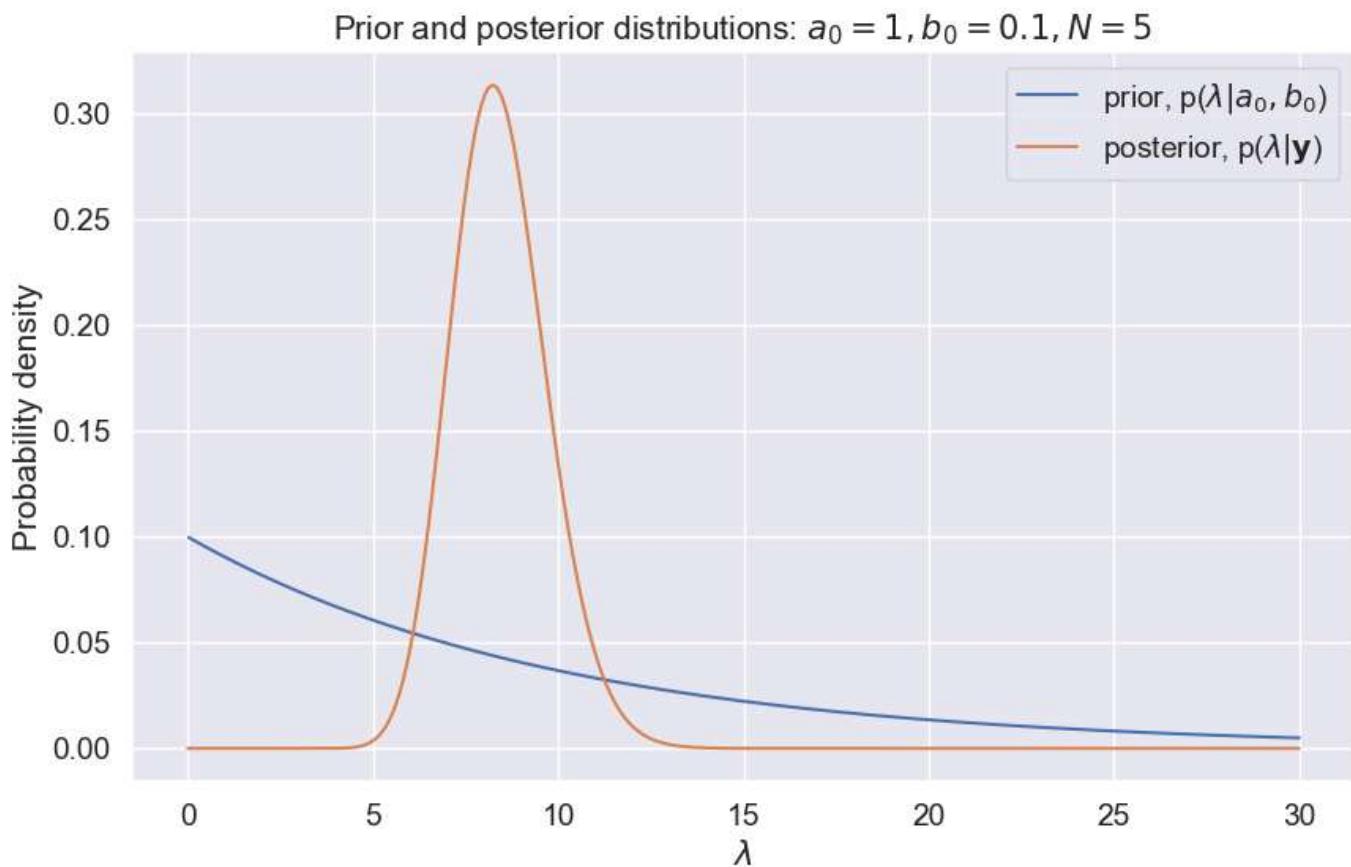
```

def log_Gamma(lamb, y, a, b):

    constant = a * jnp.log(b) - gammaln(a)
    return constant + (a - 1) * jnp.log(lamb) - b * lamb


fig, ax = plt.subplots(1, 1, figsize=(10, 6))
ax.plot(
    lambda_,
    jnp.exp(log_Gamma(lambda_, y, a0, b0)),
    label=r"prior,  $p(\lambda | a_0, b_0)$ ",
)
ax.plot(
    lambda_,
    jnp.exp(log_Gamma(lambda_, y, a0 + y.sum(), b0 + N)),
    label=r"posterior,  $p(\lambda | \mathbf{y})$ ",
)
ax.set_xlabel(r" $\lambda$ ")
ax.set_ylabel("Probability density")
ax.set_title(
    r"Prior and posterior distributions:  $a_0={}$ ,  $b_0={}$ ,  $N={}$ ".format(a0, b0, N)
)
plt.legend()
plt.show()

```



In []: