02477 Bayesian Machine Learning 2025: Assignment 1

This is the first assignment out of three in the Bayesian machine learning course 2025. The assignment is a group work of 3-5 students (make your own groups) and hand in via DTU Learn. The assignment is **mandatory**. The deadline is **2nd of March 23:59**.

Being able to manipulate probability distributions is crucial for probabilistic machine learning. The purpose of this exercise is to become more familiar with probabilistic reasoning and Bayesian computations, e.g.

- basic concepts in Bayesian machine learning (prior, likelihood, posterior)
- the sum rule for marginalization and the product rule constructing joint distributions
- computing the analytical expression for conditional distributions of simple Bayesian models
- manipulating Gaussian distributions

Part 1: The beta-binomial model

Your friend has set up a website for her new business. So far $N = 115$ potential customers has visited her site, but only $y = 4$ customers have completed a purchase. To plan her future investments, she asks you for help to compute the probability that at least one of the next $N^* = 20$ customers will make a purchase. You decide to model the problem using the beta-binomial model with a uniform prior distribution on the probability of making a purchase $\theta \in [0, 1]$:

$$\theta \sim \text{Beta}(a_0, b_0), \tag{1}$$
$$y|\theta \sim \text{Binomial}(N, \theta) \tag{2}$$

where $a_0 = b_0 = 1$.

Task 1.1: Compute the prior mean of θ and provide a 95%-credibility interval for the prior. *Hint: See Section 4.6.6 in Murphy1 for details on posterior/credibility intervals. The book discusses intervals for posterior distribution, but we can also use credibility intervals to summarize prior distributions.*

Solution to task 1.1

For $a_0 = b_0 = 1$, the prior becomes a uniform distribution for $\theta \in [0, 1]$, i.e. $p(\theta) = 1$ for $\forall\theta \in [0, 1]$. Hence, the prior mean is $\mathbb{E}[\theta] = 0.5$ and $[0.025, 0.975]$ is a 95%-credibility interval since

$$p(0.025 < \theta < 0.975) = \int_{0.025}^{0.975} p(\theta) d\theta = \int_{0.025}^{0.975} 1 d\theta = [\theta]_{0.025}^{0.975} = 0.975 - 0.025 = 0.95.$$

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

Solution to task 1.2 Since the prior is a $\text{Beta}(a_0, b_0)$-distribution and the likelihood is a $\text{Binomial}(N, \theta)$-distribution, we know that the posterior is another $\text{Beta}(a, b)$-distribution given by

$$p(\theta|y) = \text{Beta}(\theta|a_0 + y, b_0 + N - y) = \text{Beta}(\theta|1 + 4, 1 + 115 - 4) = \text{Beta}(\theta|5, 112)$$

such that $a = 5$ and $b = 112$. Next, we can compute the posterior mean

$$\mathbb{E}[\theta|y] = \frac{a}{a+b} = \frac{5}{5+112} = \frac{5}{117} \approx 0.043. \tag{3}$$

and the interval can be evaluated using scipy:

$$p(0.014 < \theta < 0.086) \approx 0.95.$$

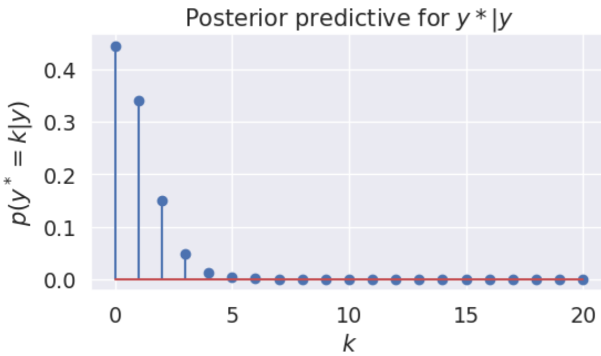

Figure 1: Plot of posterior predictive for $y^*|y$.

```python
from scipy.stats import beta

# hyperparameters, data and posterior parameters
a0, b0 = 1, 1
y, N = 4, 115
a = a0 + y
b = b0 + N - y

# interval
lower, upper = beta.interval(0.95, a=a, b=b)

# print
print(f'Posterior mean: {a/(a+b):4.3f}')
print(f'Posterior 95%-interval: [{lower:4.3f}, {upper:4.3f}]')

```

Listing 1: Code for plotting contours

Let y^* denote the number of purchases during the next $N^* = 20$ visits.

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

Solution to task 1.3

The likelihood for y^* is given by

$$p(y^*|N^*, \theta) = \text{Binomial}(y * |N^*, \theta) \tag{4}$$

and the posterior predictive distribution is thus given by

$$p(y^*|y) = \int p(y^*|N^*, \theta)p(\theta|y)d\theta = \int \text{Binomial}(y^*|N^*, \theta)\text{Beta}(\theta|a, b)d\theta \tag{5}$$

From the calculation on slide 13 in Lecture 2, we know that this integral evaluates to

$$p(y^* = k|y) = \int p(y^*|N^*, \theta)p(\theta|y)d\theta = \binom{N^*}{k}\frac{B(a+k, b+N^*-k)}{B(a,b)} = \binom{20}{k}\frac{B(5+k, 132-k)}{B(5,112)}. \tag{6}$$

Figure 1 shows the posterior predictive distribution for $y^* = k|y$ for $k \in \{0, 1, N^*\}$.

```python
from scipy.special import beta as beta_fun
from scipy.special import binom

# helper function for evaluating
def posterior_pred(a, b, Nstar, k):
```

```python
        return binom(Nstar, k) * beta_fun(a+k, b + Nstar - k) / beta_fun(a, b)

    # evaluate probabilities for k in {0, 1, ..., Nstar}
    Nstar = 20
    ks = jnp.arange(0, Nstar+1)
    post_pred_k = lambda k: posterior_pred(a, b, Nstar, k)
    probs = post_pred_k(ks)

    # sanity check: verify that the PMF
    assert jnp.sum(probs) == 1

    # plot
    fig, ax = plt.subplots(1, 1, figsize=(6, 3))
    ax.stem(ks, probs)
    ax.set(xlabel='$k$', ylabel='$p(y^*=k|y)$', title='Posterior predictive for $y*|y$');
```

Listing 2: Code for plotting the posterior predictive in eq. (6)

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

Solution to task 1.4

We have

$$\begin{aligned} p(y^* > 0|y) &= 1 - p(y^* = 0|y) && (7)\\ &= 1 - \binom{N^*}{k}\frac{B(a+k, b+N^*-k)}{B(a,b)} && (8)\\ &= \binom{20}{0}\frac{B(5+0, 132-0)}{B(5,112)} && (9)\\ &\approx 0.55 && (10) \end{aligned}$$

```python
    print(f'Probability af at least one purchase: {1 - post_pred_k(0):3.2f}')
```

Listing 3: Code for computing the probability of at least one purchase

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

Solution to task 1.5

The mean is given by

$$\mathbb{E}[y^*|y] = \sum_{k=0}^{20} kp(y^* = k|y) \approx 0.85$$

The variance is given by

$$\mathbb{V}[y^*|y] = \sum_{k=0}^{20}(k - \mathbb{E}[y^*|y])^2 p(y^* = k|y) \approx 0.95$$

```python
    # compute mean
    ystar_mean = jnp.sum(ks*probs)

    # compute variance
    ystar_var = jnp.sum((ks-ystar_mean)**2 * probs)

    print(f'Mean of post pred.: {ystar_mean:3.2f}')
    print(f'Variance of post pred.: {ystar_var:3.2f}')
```

Listing 4: Code for computing mean and variance of $y^*|y$

Part 2: Linear Gaussian systems

Let $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^2$, and $y \in \mathbb{R}$ be random variables and consider the following linear Gaussian system

$$\boldsymbol{z}_1 \sim \mathcal{N}(\boldsymbol{0}, v\boldsymbol{I}) \tag{11}$$
$$\boldsymbol{z}_2|\boldsymbol{z}_1 \sim \mathcal{N}(\boldsymbol{z}_1, v\boldsymbol{I}) \tag{12}$$
$$y|\boldsymbol{z}_2 \sim \mathcal{N}(\boldsymbol{a}^T\boldsymbol{z}_2, \sigma^2) \tag{13}$$

where $\boldsymbol{a} \in \mathbb{R}^2$ is constant. The joint distribution of $(\boldsymbol{z}_1, \boldsymbol{z}_2, y)$ is given by

$$p(y, \boldsymbol{z}_1, \boldsymbol{z}_2) = p(y|\boldsymbol{z}_2)p(\boldsymbol{z}_2|\boldsymbol{z}_1)p(\boldsymbol{z}_1). \tag{14}$$

To solve this part, you will need the equations for linear Gaussian systems in section 3.3 in Murphy1 as well as the basic rules of probability theory (sum rule, product rule, conditioning).

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

Hints: Compute $p(\boldsymbol{z}_2)$ first. The equations for linear Gaussian systems in Section 3.3 in Murphy1 will be handy.

Solution to task 2.1

First, we compute the marginal of $\boldsymbol{z}_2$:

$$p(\boldsymbol{z_2}) = \int p(\boldsymbol{z}_2|\boldsymbol{z}_1)p(\boldsymbol{z}_1)d\boldsymbol{z}_1 = \int \mathcal{N}(\boldsymbol{z}_2|\boldsymbol{z}_1, v\boldsymbol{I})\mathcal{N}(\boldsymbol{z}_1|\boldsymbol{0}, v\boldsymbol{I})d\boldsymbol{z}_1 \tag{15}$$

This is linear Gaussian system, and hence, the solution to the integral can be obtained using eq. (3.38) in Murphy1:

$$p(\boldsymbol{z_2}) = \mathcal{N}(\boldsymbol{z}_2|\boldsymbol{I}\boldsymbol{0} + \boldsymbol{0}, v\boldsymbol{I} + v\boldsymbol{I}) = \mathcal{N}(\boldsymbol{z}_2|\boldsymbol{0}, 2v\boldsymbol{I}) \tag{16}$$

We can then compute $p(\boldsymbol{y})$ using the sum rule

$$p(y) = \int p(y|\boldsymbol{a}^T\boldsymbol{z}_2)p(\boldsymbol{z}_2)d\boldsymbol{z}_2 = \int \mathcal{N}(y|\boldsymbol{a}^T\boldsymbol{z}_2, \sigma^2)\mathcal{N}(\boldsymbol{z}_2|\boldsymbol{0}, 2v\boldsymbol{I})d\boldsymbol{z}_2 \tag{17}$$

Using eq. (3.38) once more yields

$$p(y) = \mathcal{N}(y|\boldsymbol{a}^T\boldsymbol{0} + \boldsymbol{0}, \sigma^2 + \boldsymbol{a}^T 2v\boldsymbol{I}\boldsymbol{a}) = \mathcal{N}(y|\boldsymbol{0}, \sigma^2 + 2v\boldsymbol{a}^T\boldsymbol{a}) \tag{18}$$

Task 2.2: Determine the distribution $p(y, \boldsymbol{z}_2|\boldsymbol{z}_1)$.

Solution to task 2.2

Using rules for conditional distributions, we get

$$p(y, \boldsymbol{z}_2|\boldsymbol{z}_1) = \frac{p(y, \boldsymbol{z}_2, \boldsymbol{z}_1)}{p(\boldsymbol{z}_1)} = \frac{p(y|\boldsymbol{z}_2)p(\boldsymbol{z}_2|\boldsymbol{z}_1)p(\boldsymbol{z}_1)}{p(\boldsymbol{z}_1)} = p(y|\boldsymbol{z}_2)p(\boldsymbol{z}_2|\boldsymbol{z}_1) = \mathcal{N}(y|\boldsymbol{a}^T\boldsymbol{z}_2, \sigma^2)\mathcal{N}(\boldsymbol{z}_2|\boldsymbol{z}_1, v\boldsymbol{I}) \tag{19}$$

Task 2.3: Determine the distribution $p(y|\boldsymbol{z}_1)$.

Solution to task 2.3

We have

$$p(y|\boldsymbol{z}_1) = \int p(y, \boldsymbol{z}_2|\boldsymbol{z}_1)\mathrm{d}\boldsymbol{z}_2 = \int \mathcal{N}(y|\boldsymbol{a}^T\boldsymbol{z}_2, \sigma^2)\mathcal{N}(\boldsymbol{z}_2|\boldsymbol{z}_1, v\boldsymbol{I})\mathrm{d}\boldsymbol{z}_2 \tag{20}$$

Again, we recognize this integral as the marginalization in a linear Gaussian system, so we can use eq. (3.38) again to get:

$$p(y|\boldsymbol{z}_1) = \mathcal{N}(y|\boldsymbol{a}^T\boldsymbol{z}_1 + \boldsymbol{0}, \sigma^2 + \boldsymbol{a}^T v\boldsymbol{I}\boldsymbol{a}) = \mathcal{N}(y|\boldsymbol{a}^T\boldsymbol{z}_1, \sigma^2 + v\boldsymbol{a}^T\boldsymbol{a}) \tag{21}$$

Task 2.4: Determine the distribution $p(\boldsymbol{z}_1|y)$.

Hint: Start by using Bayes' rule.

Solution to task 2.4

Bayes' rule yields

$$p(\boldsymbol{z}_1|y) = \frac{p(y|\boldsymbol{z}_1)p(\boldsymbol{z}_1)}{p(y)} \tag{22}$$

This is also a linear Gaussian system, where $p(y|\boldsymbol{z}_1) = \mathcal{N}(y|\boldsymbol{a}^T\boldsymbol{z}_1, \sigma^2 + v\boldsymbol{a}^T\boldsymbol{a})$ takes the role of the likelihood and $p(\boldsymbol{z}_1) = \mathcal{N}(\boldsymbol{z}_1|\boldsymbol{0}, v\boldsymbol{I})$ take the role of the prior. Hence, we can use eq. (3.37) to compute $p(\boldsymbol{z}_1|y)$:

$$p(\boldsymbol{z}_1|y) = \mathcal{N}(\boldsymbol{z}_1|\mu_{z_1|y}, \Sigma_{z_1|y}), \tag{23}$$

where

$$\Sigma_{z_1|y} = \left[\frac{1}{v}\boldsymbol{I} + \boldsymbol{a}(\sigma^2 + v\boldsymbol{a}^T\boldsymbol{a})^{-1}\boldsymbol{a}^T\right]^{-1} \tag{24}$$

$$\mu_{z_1|y} = \Sigma_{z_1|y}\boldsymbol{a}(\sigma^2 + v\boldsymbol{a}^T\boldsymbol{a})^{-1}y \tag{25}$$

Part 3: A conjugate model for count data

In week 1, we studied the Beta-binomial model, which is an example of a so-called **conjugate model**. In this exercise, you will work with another conjugate model, namely the Poisson-Gamma model, where the likelihood takes the form of a Poisson distribution, $y_i|\lambda \sim \text{Poisson}(\lambda)$ with rate $\lambda > 0$, and the prior on λ is a Gamma distribution, i.e. $\lambda \sim \text{Gamma}(a_0, b_0)$, where $a_0 > 0$ and b_0 are hyperparameters known as the *shape* and *rate*, respectively.

The Poisson distribution is discrete distribution, which is often applied to model **count data**, where $y_i \in \{0, 1, 2, \dots\}$ are a non-negative integers. The probability mass function (PMF) for the Poisson distribution is:

$$p(y_i|\lambda) = \frac{\lambda^{y_i}e^{-\lambda}}{y_i!} \tag{26}$$

for $y_i \in \{0, 1, 2, \dots\}$.

The Gamma distribution, e.g. $\lambda \sim \text{Gamma}(a_0, b_0)$, is distribution over the non-negative real line, i.e. $\lambda > 0$ with the following probability density function (PDF):

$$p(\lambda|a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)}\lambda^{a_0-1}e^{-b_0\lambda}, \tag{27}$$

where $\Gamma(x)$ is the gamma function.

Consider now the following model

$$y_i|\lambda \sim \text{Poisson}(\lambda) \tag{28}$$

$$\lambda \sim \text{Gamma}(a_0, b_0), \tag{29}$$

where $\{y_i\}$ are assumed to *conditionally independent* given λ. Suppose we collect N observations such that $\boldsymbol{y} = \{y_i\}_{i=1}^N$

Task 3.1: Determine the joint distribution of $(\boldsymbol{y}, \lambda)$

Solution to task 3.1

Using the product rule and the assumption of conditional independence:

$$p(\boldsymbol{y}, \lambda) = p(\boldsymbol{y}|\lambda)p(\lambda) = p(\lambda)\prod_{i=1}^{N} p(y_i|\lambda) = \text{Gamma}(\lambda|a_0, b_0)\prod_{i=1}^{N}\text{Poisson}(y_i|\lambda). \tag{30}$$

Task 3.2: Show that the functional form of a Gamma distribution is given by $\log p(\lambda|a, b) = (a-1)\log(\lambda) - b\lambda +$ constant.

Solution to task 3.2

$$\begin{aligned}
\log p(\lambda|a, b) &= \log\left[\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}\right] &(31)\\
&= \log\left[\frac{b^a}{\Gamma(a)}\right] + (a-1)\log(\lambda) - b\lambda &(32)\\
&= (a-1)\log(\lambda) - b\lambda + \text{constant} &(33)
\end{aligned}$$

Task 3.3: Derive the analytical expression for the posterior distribution $p(\lambda|\boldsymbol{y})$ and show that it is a Gamma-distribution

Solution to task 3.3

Bayes' rule states that

$$\begin{aligned}
\log p(\lambda|\boldsymbol{y}) &= \log p(\boldsymbol{y}|\lambda) + \log p(\lambda) - \log p(\boldsymbol{y}) &(34)\\
&= \log p(\boldsymbol{y}|\lambda) + \log p(\lambda) + \text{const.} &(35)\\
&= \log\left[\prod_{i=1}^{N}\text{Poisson}(y_i|\lambda)\right] + \log p(\lambda) + \text{const.} &(36)\\
&= \sum_{i=1}^{N}\log\left[\text{Poisson}(y_i|\lambda)\right] + \log p(\lambda) + \text{const.} &(37)\\
&= \sum_{i=1}^{N}\left[y_i\log\lambda - \lambda - \log y_i!\right] + (a_0-1)\log(\lambda) - b_0\lambda + \text{const.} &(38)\\
&= -N\lambda + \log\lambda\sum_{i=1}^{N}y_i + (a_0-1)\log(\lambda) - b_0\lambda + \text{const.} &(39)\\
&= \left(a_0 + \sum_{i=1}^{N}y_i - 1\right)\log\lambda - (b_0+N)\lambda + \text{const.} &(40)
\end{aligned}$$

Comparing the last line with the equation for the functional form of the Gamma distribution and equation coefficient, we can deduce that the posterior $p(\lambda|\boldsymbol{y})$ must be a Gamma distribution with parameters $a = a_0 + \sum_{i=1}^{N} y_i$ and $b = b_0 + N$, i.e.

$$p(\lambda|\boldsymbol{y}) = \text{Gamma}(\lambda|a, b) = \text{Gamma}(\lambda|a_0 + \sum_{i=1}^{N}y_i, b_0 + N). \tag{41}$$

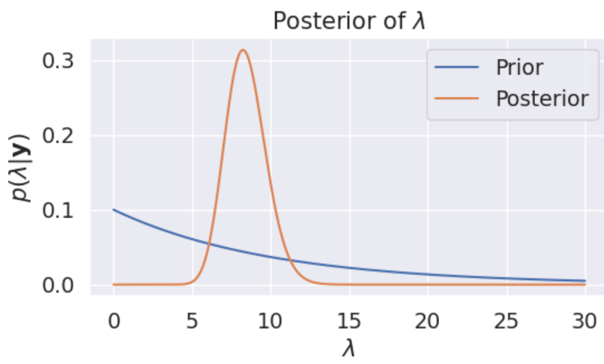

Figure 2: Prior and posterior density for Poisson-Gamma model.

Suppose we observe $y_1 = 7$, $y_2 = 4$, $y_3 = 8$, $y_4 = 11$, and $y_5 = 12$ such that $N = 5$. Assume $a_0 = 1$ and $b_0 = \frac{1}{10}$.

Task 3.4: Determine the posterior distribution for λ given the data above and report the mean.

Hints: If $\lambda \sim Gamma(a, b)$, then $\mathbb{E}[\lambda] = \frac{a}{b}$.

Solution to task 3.4

We have

$$\sum_{i=1}^{5} y_i = 7 + 4 + 8 + 11 + 12 = 42 \tag{42}$$

and thus,

$$p(\lambda|\boldsymbol{y}) = \text{Gamma}(\lambda|a, b) = \text{Gamma}(\lambda|1 + 42, 1/10 + 5) = \text{Gamma}(\lambda|43, 5.1) \tag{43}$$

Task 3.5: Plot $p(\lambda)$ and $p(\lambda|\boldsymbol{y})$ for $\lambda \in [0, 30]$.

Hints: Implement the log density first for numeric stability. The function scipy.stats.gammaln implements the logarithm of the $\Gamma(x)$ function, i.e. $\log \Gamma(x)$.

Solution to task 3.5

```python
from scipy.special import gammaln

# prior and data
y = jnp.array([7, 4, 8., 11., 12.])
N = len(y)
a0, b0 = 1, 0.1

# posterior
a = a0 + jnp.sum(y)
b = b0 + N

# implement density
gamma_logpdf = lambda lam, a, b: a*jnp.log(b) - gammaln(a) + (a-1)*jnp.log(lam) -b*lam
gamma_pdf = lambda lam, a,b: jnp.exp(gamma_logpdf(lam, a, b))
gamma_values = jnp.linspace(0, 30, 1000)

# plot
fig, ax = plt.subplots(1, 1, figsize=(6, 3))
```

```
    ax.plot(gamma_values, gamma_pdf(gamma_values, a0, b0), label='Prior')
    ax.plot(gamma_values, gamma_pdf(gamma_values, a, b), label='Posterior')
    ax.set(xlabel='$\\lambda$', ylabel='$p(\\lambda|\\mathbf{y})$')
    ax.legend()
```

Listing 5: Code for plotting the prior and posterior of Poisson-Gamma model