1. **Conjugate Priors (30)**

(a) Exponential and Gamma

The likelihood is \( P(X \mid \lambda) = \prod_{i=1}^{N} \lambda \exp(-\lambda x_i) \) and the prior is \( p(\lambda \mid \alpha, \beta) = \text{gamma}(\lambda \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta \lambda) \). Let \( X \) denote the observations \( x_1, \ldots, x_N \) and let \( s_N \) denote their sum. Then the posterior is

\[
p(\lambda \mid X) \propto \beta^{\alpha} \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)} \prod_{i=1}^{N} \lambda \exp(-\lambda x_i) = \beta^{\alpha} \frac{\lambda^{\alpha+N-1}}{\Gamma(\alpha+N)} \exp(-\lambda (\beta + s_N)) \propto \text{gamma}(\lambda \mid \alpha + N, \beta + s_N).
\]

Therefore the parameter updates are as follows:

\[
\alpha' \leftarrow \alpha + N, \quad \beta' \leftarrow \beta + s_N
\]

For the prediction distribution we compute the following integral:

\[
p(x_{N+1} \mid x_1, \ldots, x_N) = \int p(x_{N+1} \mid \lambda)p(\lambda \mid x_1, \ldots, x_N)d\lambda
\]

\[
= \int \lambda \exp(-\lambda x_{N+1}) \text{gamma}(\lambda \mid \alpha + N, \beta + s_N)d\lambda
\]

\[
= \frac{(\beta + s_N)^{\alpha+N}}{\Gamma(\alpha+N)} \Gamma(\alpha+N) \int \lambda \text{gamma}(\lambda \mid \alpha + N, \beta + s_N + x_{N+1})d\lambda
\]

\[
= \frac{(\beta + s_N)^{\alpha+N}}{(\beta + s_N + x_{N+1})^{\alpha+N}} \beta + s_N + x_{N+1}
\]

where the penultimate step uses the standard formula \( \alpha/\beta \) for the expected value of a gamma distribution.

(b) Geometric and Beta

The likelihood for a single observation of value \( k \) is \( P(X = k \mid \theta) = (1 - \theta)^{k-1} \theta \) and prior is \( p(\theta \mid a, b) = \text{Beta}(a, b) = a \theta^{a-1} (1 - \theta)^{b-1} \), where \( a \) is the normalization constant. Then the posterior is,

\[
p(\theta \mid X) = a \theta^{a-1} (1 - \theta)^{b-1} (1 - \theta)^{k-1} \theta = a \theta^a (1 - \theta)^{b+k-2} = \text{Beta}(\theta \mid a + 1, b + k - 1)
\]

Therefore the parameter updates are

\[
a' \leftarrow a + 1, \quad b' \leftarrow b + k - 1
\]
For the prediction distribution we compute the following integral:

\[
p(X_2 = \ell \mid X_1 = k) = \int p(X_2 = \ell \mid \theta) p(\theta \mid X_1 = k) d\theta
\]

\[
= \int (1 - \theta)^{\ell - 1} \theta \text{Beta}(\theta \mid a + 1, b + k - 1) d\theta
\]

\[
= \frac{\Gamma(a + b + k)}{\Gamma(a + 1) \Gamma(b + k - 1)} \frac{\Gamma(a + b + k + \ell - 2)}{\Gamma(a + b + k + \ell - 1)} \int \theta \text{Beta}(\theta \mid a + 1, b + k + \ell - 2) d\theta
\]

\[
= \frac{\Gamma(a + b + k)}{\Gamma(b + k - 1)} \frac{\Gamma(a + b + k + \ell - 2)}{\Gamma(a + b + k + \ell - 1)} \cdot (a + 1)
\]

where the penultimate step uses the standard formula \(\alpha/(\alpha + \beta)\) for the expected value of a Beta distribution.

(c) Mixture Prior

The prior is given by the mixture,

\[
P(\theta \mid \gamma_1, \ldots, \gamma_M) = \sum_{m=1}^{M} w_m P(\theta \mid \gamma_m)
\]

Moreover, we are given that \(P(\theta \mid \gamma_m)\) is a conjugate prior for the likelihood \(P(X \mid \theta)\); in other words,

\[
P(\theta \mid X, \gamma_m) = \alpha_m P(X \mid \theta) P(\theta \mid \gamma_m) = P(\theta \mid \gamma'_m).
\]

When we multiply the mixture prior with the likelihood, we get the following posterior:

\[
P(\theta \mid X, \gamma_1, \ldots, \gamma_M) = \alpha P(X \mid \theta) \sum_{m=1}^{M} w_m P(\theta \mid \gamma_m)
\]

\[
= \sum_{m=1}^{M} \alpha w_m P(X \mid \theta) P(\theta \mid \gamma_m)
\]

\[
= \sum_{m=1}^{M} \frac{\alpha w_m}{\alpha_m} P(\theta \mid \gamma'_m)
\]

\[
= \sum_{m=1}^{M} w' P(\theta \mid \gamma'_m).
\]

Therefore we observe that the posterior has the same form as the prior, i.e., a mixture distribution with updated weights and hyperparameters.

2. Bayesian Naive Bayes

(a) Using the results in 1(a), given a single e-mail sample with attribute values \(x_j\) and class \(y\), we perform the following update:

\[
\alpha_{j,y} \leftarrow \alpha + 1
\]

\[
\beta_{j,y} \leftarrow \beta + x_j
\]