Problem 1. Short answer: She should switch[1].

There are many different approaches to this problem which give the same answer. However, note that the problem asks "calculate the relevant posterior probabilities", which implies that you should use Bayes' rule.

We can define the following random variables:

- S_i: 1 if she picks door i and 0 o.w.
- O_i: 1 if hosts opens door i and 0 o.w.
- Z_i: 1 if the prize is behind door i and 0 o.w.

Since the order of the door doesn't matter, we can assume that she initially picks door 1 and the host opens door 3 (any orders will work in the same way). We want to compute the posterior $P(Z_1 = 1|S_1 = 1, O_3 = 1)$ and $P(Z_2 = 1|S_1 = 1, O_3 = 1)$. By Bayes' rule:

$$
\begin{aligned}
P(Z_1 = 1|S_1 = 1, O_3 = 1) &= \frac{P(S_1 = 1, O_3 = 1|Z_1 = 1)P(Z_1 = 1)}{\sum_{j\in\{0,1\}} P(S_1 = 1, O_3 = 1|Z_1 = j)P(Z_1 = j)} \\
&= \frac{P(O_3 = 1|S_1 = 1, Z_1 = 1)P(S_1 = 1|Z_1 = 1)P(Z_1 = 1)}{\sum_{j\in\{0,1\}} P(O_3 = 1|S_1 = 1, Z_1 = j)P(S_1 = 1|Z_1 = j)P(Z_1 = j)} \\
&= \frac{\frac{1}{2}\times\frac{1}{3}\times\frac{1}{3}}{\frac{1}{2}\times\frac{1}{3}\times\frac{1}{3} + 1\times\frac{1}{3}\times\frac{1}{3}} = \frac{1}{3}
\end{aligned}
$$

Here we are making use of the fact that S_1 and Z_1 are independent and O_3 is dependent on S_1 and Z_1. (If she picks door 1 and the prize is behind door 2, the host has to pick door 3. Otherwise the host could've picked either door 2 or door 3 with equal probabilities.)

We can compute $P(Z_2 = 1|S_1 = 1, O_3 = 1)$ in the similar way. However, since $P(Z_2 = 1|S_1 = 1, O_3 = 1) = P(Z_1 = 0|S_1 = 1, O_3 = 1)$, we can immediately get that it equals $\frac{2}{3}$, which leads to the conclusion that she should switch.

Problem 2. The likelihood given data $\mathbf{X} = \{X_i\}_{i=1}^N$:

$$
\begin{aligned}
p(\mathbf{X}|\pi) = \prod_{i=1}^N p(X_i|\pi) &= \prod_{i=1}^N \prod_{j=1}^K \pi_j^{\mathbb{1}[X_i=j]} \\
&= \exp\left\{\sum_{i=1}^N \sum_{j=1}^K \mathbb{1}[X_i = j] \ln \pi_j\right\} = \exp\left\{\sum_{j=1}^K m_j \ln \pi_j\right\}
\end{aligned}
$$

where $\mathbb{1}[\cdot]$ is the indicator function and $m_j = \sum_{i=1}^N \mathbb{1}[X_i = j]$ denotes the number of state j that shows up in the data. We can immediately recognize the natural parameters $\ln \pi$ and log-partition function $A(\eta) = 0$. Therefore, the conjugate prior should have the form:

$$
p(\pi|\alpha) \propto \exp\left\{\sum_{j=1}^K (\alpha_j - 1) \ln \pi_j\right\}
$$

which we recognize it as Dirichlet distribution $Dir(\pi; \alpha_1, \cdots, \alpha_K)$[2]. To properly normalize the prior requires a little hard work, since it involves integration over a simplex. You are not required to perform this integral[3]. The complete form of Dirichlet distribution is:

$$
\begin{aligned}
Dir(\pi; \alpha_1, \cdots, \alpha_K) &= \frac{\Gamma(\sum_{j=1}^K \alpha_j)}{\sum_{j=1}^K \Gamma(\alpha_j)} \prod_{j=1}^K \pi_j^{(\alpha_j - 1)} \\
&= \exp\left\{\sum_{j=1}^K (\alpha_j - 1) \ln \pi_j - A(\alpha)\right\}
\end{aligned}
$$

[1] This is the famous Monty Hall problem.

[2] Here we are actually ahead of ourselves and use the natural parameter form of Dirichlet distribution: $\eta_j = \alpha_j - 1$, mainly as a convention.

[3] For those who are interested, here is a nice walk-through: `http://www.youtube.com/watch?v=cv7ESInBhbs`.

where $A(\alpha) = \sum_{j=1}^{K} \ln \Gamma(\alpha_j) - \ln \Gamma(\sum_{j=1}^{K} \alpha_j)$. The posterior follows:

$$p(\pi|\mathbf{X}) \propto \exp\left\{ \sum_{j=1}^{K} (\alpha_j + m_j - 1) \ln \pi_j \right\} = Dir(\pi; \alpha_1 + m_1, \cdots, \alpha_K + m_K)$$

where the natural parameters $\eta_j = \alpha_j + m_j - 1$. Clearly the most obvious characteristic of the parameters of the posterior is that they are simply the parameters of the prior plus the "counts" from data.

Problem 3. (a) The posterior of μ and λ:

$$\begin{aligned}
p(\mu, \lambda|\boldsymbol{x}) &\propto p(\boldsymbol{x}|\mu, \lambda) p(\mu, \lambda) \\
&= \prod_{i=1}^{N} p(x_i|\mu, \lambda) p(\mu|\lambda) p(\lambda) \\
&= \prod_{i=1}^{N} \mathrm{Normal}(x_i; \mu, \lambda^{-1}) \mathrm{Normal}(\mu; 0, a\lambda^{-1}) \mathrm{Gamma}(\lambda; b, c) \\
&\propto \lambda^{\frac{N+1}{2}} \exp\left\{ -\frac{\lambda(N + \frac{1}{a})}{2} \mu^2 + (\lambda \sum_{i=1}^{N} x_i) \cdot \mu - \frac{\lambda \sum_{i=1}^{N} x_i^2}{2} \right\} \lambda^{b-1} \exp\{-c\lambda\}
\end{aligned}$$

The first exponential term can be simplified by "completing the square":

$$\begin{aligned}
& -\frac{\lambda(N + \frac{1}{a})}{2} \mu^2 + (\lambda \sum_{i=1}^{N} x_i) \cdot \mu - \frac{\lambda \sum_{i=1}^{N} x_i^2}{2} \\
= & -\frac{\lambda(N + \frac{1}{a})}{2} \left(\mu - \frac{\sum_{i=1}^{N} x_i}{N + \frac{1}{a}} \right)^2 - \lambda \left(\frac{\sum_{i=1}^{N} x_i^2}{2} - \frac{(\sum_{i=1}^{N} x_i)^2}{2(N + \frac{1}{a})} \right)
\end{aligned}$$

Plug back into the posterior:

$$p(\mu, \lambda|\boldsymbol{x}) \propto \underbrace{\lambda^{\frac{1}{2}} \exp\left\{ -\frac{\lambda(N + \frac{1}{a})}{2} \left(\mu - \frac{\sum_{i=1}^{N} x_i}{N + \frac{1}{a}} \right)^2 \right\}}_{p(\mu|\lambda, \boldsymbol{x}) = \mathrm{Normal}(\mu; \eta, \tau^{-1})} \cdot \underbrace{\lambda^{b + \frac{N}{2} - 1} \exp\left\{ -\left(c + \frac{\sum_{i=1}^{N} x_i^2}{2} - \frac{(\sum_{i=1}^{N} x_i)^2}{2(N + \frac{1}{a})} \right) \lambda \right\}}_{p(\lambda|\boldsymbol{x}) = \mathrm{Gamma}(\lambda; \alpha, \beta)}$$

where

$$\begin{aligned}
\eta &= \frac{\sum_{i=1}^{N} x_i}{N + \frac{1}{a}} \qquad \tau = \lambda(N + \frac{1}{a}) \\
\alpha &= b + \frac{N}{2} \qquad \beta = c + \frac{\sum_{i=1}^{N} x_i^2}{2} - \frac{(\sum_{i=1}^{N} x_i)^2}{2(N + \frac{1}{a})} = c + \frac{1}{2} \left(N \overline{\mathrm{Var}}(x) + \frac{N \bar{x}^2}{aN + 1} \right)
\end{aligned}$$

where $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$ is the empirical mean and $\overline{\mathrm{Var}}(x) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$ is the empirical variance. Therefore, we identify the posterior as the product of $\mathrm{Normal}(\mu; \eta, \tau^{-1})$ and $\mathrm{Gamma}(\lambda; \alpha, \beta)$ where the parameters are given above[4].

[4]Together it is called Normal-Gamma distribution.

(b) Take the posterior representation from part (a). To simplify notation we let $\frac{1}{n_0} = N + \frac{1}{a}$. Note that $\tau = \frac{\lambda}{n_0}$. We will show that the predictive distribution[5] is a Student's t.

$$
\begin{aligned}
p(x^*) &= \int_{(0,\infty)} \int_{\mathbb{R}} p(x^*|\mu,\lambda)p(\mu|\lambda,\eta,n_0)p(\lambda|\alpha,\beta)d\mu d\lambda \\
&= \int_{(0,\infty)} p(\lambda|\alpha,\beta) \int_{\mathbb{R}} p(x^*|\mu,\lambda)p(\mu|\lambda,\eta,n_0)d\mu d\lambda \\
&= \int_{(0,\infty)} p(\lambda|\alpha,\beta) \int_{\mathbb{R}} \Big(\frac{\lambda}{2\pi}\Big)^{\frac{1}{2}} \cdot \exp\Big(-\frac{\lambda(\mu-x^*)^2}{2}\Big) \cdot \Big(\frac{\lambda}{2\pi n_0}\Big)^{\frac{1}{2}} \cdot \exp\Big(-\frac{\lambda(\mu-\eta)^2}{2n_0}\Big) d\mu d\lambda \\
&= \int_{(0,\infty)} p(\lambda|\alpha,\beta) \int_{\mathbb{R}} \Big(\frac{\lambda}{2\pi}\Big)^{\frac{1}{2}} \cdot \Big(\frac{\lambda}{2\pi n_0}\Big)^{\frac{1}{2}} \cdot \Big(\frac{\lambda(1+\frac{1}{n_0})}{2\pi}\Big)^{-\frac{1}{2}} \\
&\quad \cdot \underbrace{\Big(\frac{\lambda(1+\frac{1}{n_0})}{2\pi}\Big)^{\frac{1}{2}} \cdot \exp\Big(-\frac{\lambda}{2}(1+\frac{1}{n_0})\Big(\mu - \frac{\lambda x^* + \frac{\lambda\eta}{n_0}}{\lambda(1+\frac{1}{n_0})}\Big)^2}_{\text{Complete a normal density function.}} - \frac{\lambda(x^*)^2}{2} - \frac{\lambda\eta^2}{2n_0} + \frac{(\lambda x^* + \frac{\lambda\eta}{n_0})^2}{2\lambda(1+\frac{1}{n_0})}\Big) d\mu d\lambda \\
&= \int_{(0,\infty)} \frac{\beta^\alpha \lambda^{\alpha-1}\exp(-\beta\lambda)}{\Gamma(\alpha)} \cdot \Big(\frac{\lambda}{2\pi(n_0+1)}\Big)^{\frac{1}{2}} \cdot \exp\Big(-\frac{\lambda(x^*-\eta)^2}{2(n_0+1)}\Big) d\lambda \\
&= \int_{(0,\infty)} \frac{\beta^\alpha \lambda^{\alpha+\frac{1}{2}-1}\exp(-(\beta + \frac{(x^*-\eta)^2}{2(n_0+1)})\lambda)}{\Gamma(\alpha)} \cdot \Big(\frac{1}{2\pi(n_0+1)}\Big)^{\frac{1}{2}} d\lambda \\
&= \frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)} \cdot \Big(\frac{1}{2\pi(n_0+1)}\Big)^{\frac{1}{2}} \cdot \frac{\beta^\alpha}{(\beta + \frac{(x^*-\eta)^2}{2(n_0+1)})^{\alpha+\frac{1}{2}}} \cdot \int_{(0,\infty)} \underbrace{\frac{(\beta + \frac{(x^*-\eta)^2}{2(n_0+1)})^{\alpha+\frac{1}{2}} \cdot \lambda^{\alpha+\frac{1}{2}-1}\exp(-(\beta + \frac{(x^*-\eta)^2}{2(n_0+1)})\lambda)}{\Gamma(\alpha+\frac{1}{2})}}_{\text{Complete a Gamma density function.}} d\lambda \\
&= \frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)} \cdot \Big(\frac{\frac{\alpha}{\beta(n_0+1)}}{\pi \cdot 2\alpha}\Big)^{\frac{1}{2}} \Big(1 + \frac{\frac{\alpha}{\beta(n_0+1)} \cdot (x^*-\eta)^2}{2\alpha}\Big)^{-\alpha-\frac{1}{2}} \\
&= \mathrm{St}(x^*|\eta, 2\alpha, \frac{\alpha}{\beta(n_0+1)}).
\end{aligned}
$$

[5]We simplify the notation to $p(x^*)$.