Q1. To find the probability that an employee is a smoker given that he/she uses the health insurance plan, we can use Bayes' theorem. Let's denote:

- AA: Event that an employee is a smoker.
- BB: Event that an employee uses the health insurance plan.

Given:

- $P(B)=0.70P(B)=0.70$ (probability that an employee uses the health insurance plan)
- $P(A|B)=0.40P(A|B)=0.40$ (probability that an employee is a smoker given that he/she uses the health insurance plan)

We can calculate $P(A)P(A)$ (the probability that an employee is a smoker) using the law of total probability and then use Bayes' theorem:
$P(A)=P(A|B)\times P(B)+P(A|\neg B)\times P(\neg B)P(A)=P(A|B)\times P(B)+P(A|\neg B)\times P(\neg B)$
Where:

- $P(\neg B)=1-P(B)P(\neg B)=1-P(B)$ (probability that an employee does not use the health insurance plan)
- $P(A|\neg B)P(A|\neg B)$ is the probability that an employee is a smoker given that he/she does not use the health insurance plan.

Given that this information is not provided, let's assume $P(A|\neg B)=0.20P(A|\neg B)=0.20$ (just for the sake of example). Then:
$P(A)=0.40\times 0.70+0.20\times(1-0.70)=0.28+0.06=0.34P(A)=0.40\times 0.70+0.20\times(1-0.70)=0.28+0.06=0.34$
Now, we can use Bayes' theorem to find $P(A|B)P(A|B)$:
$P(A|B)=P(B|A)\times P(A)P(B)P(A|B)=P(B)P(B|A)\times P(A)$
Substitute the given values:
$P(A|B)=0.40\times 0.340.70=0.1360.70\approx 0.194P(A|B)=0.700.40\times 0.34=0.700.136\approx 0.194$
So, the probability that an employee is a smoker given that he/she uses the health insurance plan is approximately 0.194 or 19.4%.

Q2. The main difference between Bernoulli Naive Bayes and Multinomial Naive Bayes lies in the type of features they are designed to handle:

- **Bernoulli Naive Bayes**: This model is suitable for binary feature vectors, where each feature represents the presence or absence of a particular attribute. It assumes that features are independent binary variables.

- **Multinomial Naive Bayes**: This model is designed for multinomially distributed features, typically used for text classification where features represent word counts or frequencies. It assumes that features are independent and follow a multinomial distribution.
-

Q3. Bernoulli Naive Bayes can handle missing values by treating them as an additional category. When a feature's value is missing for a particular instance, it is considered as a separate category, and the probability of that category is calculated based on the observed frequencies in the training data.

Q4. Yes, Gaussian Naive Bayes can be used for multi-class classification. In Gaussian Naive Bayes, the features are assumed to follow a Gaussian (normal) distribution, and the class conditional probability is modeled using the mean and variance of the features in each class. The model can be extended to multiple classes by computing the class probabilities using Bayes' theorem and selecting the class with the highest probability as the predicted class for a given instance.