

Homework 1

STT 465, Bayesian Statistical Methods

Lowell Monis

September 10, 2025

Question 1

(a) 10 points

Two factories I and II produce phones for brand ABC. Factory I produces 60% of all ABC phones, and factory II produces 40%. 10% of phones produced by factory I are defective, and 20% of those produced by factory II are defective. You purchase a brand ABC phone, and assume this phone is randomly chosen. Suppose the phone is not defective. What is the probability that it came from factory II?

Answer

First, I establish the events for this problem:

D : A phone is defective.

D^c : A phone is not defective.

F_1 : A phone is from factory I.

F_2 : A phone is from factory II.

I now define the probabilities from the problem statement:

$$P(F_1) = 0.60$$

$$P(F_2) = 0.40$$

$$P(D|F_1) = 0.10$$

$$P(D|F_2) = 0.20$$

I need to find the probability that a phone came from factory II, given that it is not defective. Thus, I need to find $P(F_2|D^c)$.

To go ahead with this problem, I am finding the overall probability of a randomly chosen phone being non-defective. I can do this using the Law of Total Probability.

Before that, I need to find conditional probabilities for each factory when the phone is non-defective.

$$P(D^c|F_1) = 1 - P(D|F_1) = 1 - 0.10 = 0.90$$

$$P(D^c|F_2) = 1 - P(D|F_2) = 1 - 0.20 = 0.80$$

The total probability that a randomly chosen phone is not defective can be written as:

$$\begin{aligned} P(D^c) &= P(D^c|F_1) \cdot P(F_1) + P(D^c|F_2) \cdot P(F_2) \\ &= 0.90 \cdot 0.60 + 0.80 \cdot 0.40 \\ &= 0.54 + 0.32 \\ &= 0.86 \end{aligned}$$

Now, I can apply Bayes' theorem to find the probability that the given phone is from factory II if it is not defective:

$$\begin{aligned} P(F_2|D^c) &= \frac{P(D^c|F_2) \cdot P(F_2)}{P(D^c)} \\ &= \frac{0.80 \cdot 0.40}{0.86} \\ &= \frac{0.32}{0.86} \\ &= 0.372093 \approx 0.37 \end{aligned}$$

Answer: There is a 37% chance a randomly selected ABC phone that is not defective was produced in factory II.

(b) 10 points

In answering a question on a multiple-choice test, a student either knows the answer or guesses. Let p be the probability that the student knows the answer and $1 - p$ be the probability that the student guesses. Assume that a student who guesses at the answer will be correct with probability $\frac{1}{m}$, where m is the number of multiple-choice alternatives. What is the probability that a student knew the answer to a question given that he or she answered it correctly?

Answer

First, I establish the events for this problem:

K : The student knows the answer.

G : The student guesses the answer.

C : The student answers the questions correctly.

I now define the probabilities from the problem statement:

$$\begin{aligned} P(K) &= p \\ P(G) &= 1 - p \\ P(C|K) &= 1 \\ P(C|G) &= \frac{1}{m} \end{aligned}$$

$P(C|K) = 1$ because the chances of the student answering the question correctly if they know the answer is 100%.

I need to find the probability that a student knew the answer, given that they answered it correctly. I need to find $P(K|C)$.

To go ahead with this problem, I am finding the overall probability of a student getting the correct answer. I can do this using the Law of Total Probability.

The two ways established to get a correct answer is either by guessing or knowing the answer.

$$\begin{aligned} P(C) &= P(C|K) \cdot P(K) + P(C|G) \cdot P(G) \\ &= 1 \cdot p + \frac{1}{m} \cdot (1 - p) \\ &= p + \frac{1 - p}{m} \\ &= \frac{mp + 1 - p}{m} \end{aligned}$$

Now, I can apply Bayes' theorem to find the probability that the student knew the answer if they got it correct:

$$\begin{aligned} P(K|C) &= \frac{P(C|K) \cdot P(K)}{P(C)} \\ &= \frac{1 \cdot p}{\frac{mp+1-p}{m}} \\ &= \frac{p}{\frac{mp+1-p}{m}} \\ &= \frac{mp}{mp+1-p} \end{aligned}$$

Answer: The probability that a student knew the answer if they got it correct can be expressed as $\frac{mp}{mp+1-p}$, where p is the probability that the student knows the answer, and m is the number of multiple-choice alternatives on the question.

Question 2

Recall the infection example covered in class: to study the infection rate θ in the city, we randomly sample 20 individuals and record the number of infections y in the sample. Consider the simplified case where the infection rate θ can only take three possible values 0.01, 0.03, 0.05. Namely, the parameter space for θ is $\Theta = \{0.01, 0.03, 0.05\}$. Suppose $y = 0$ and the sampling model is assumed to be the Binomial model.

From the question,

- Sample size: $n = 20$
- Observed $y = 0$
- Parameters $\Theta = 0.01, 0.03, 0.05$
- Binomial pmf (likelihood for observed y):

$$p(y|\theta) = \binom{n}{k} \theta^y (1 - \theta)^{n-y}$$

For $y = 0$ this simplifies to

$$L(\theta) = p(y = 0|\theta) = (1 - 2\theta)^{20-0} = (1 - 2\theta)^{20}$$

Numerical values of estimators (computed exactly from $(1 - \theta)^{20}$):

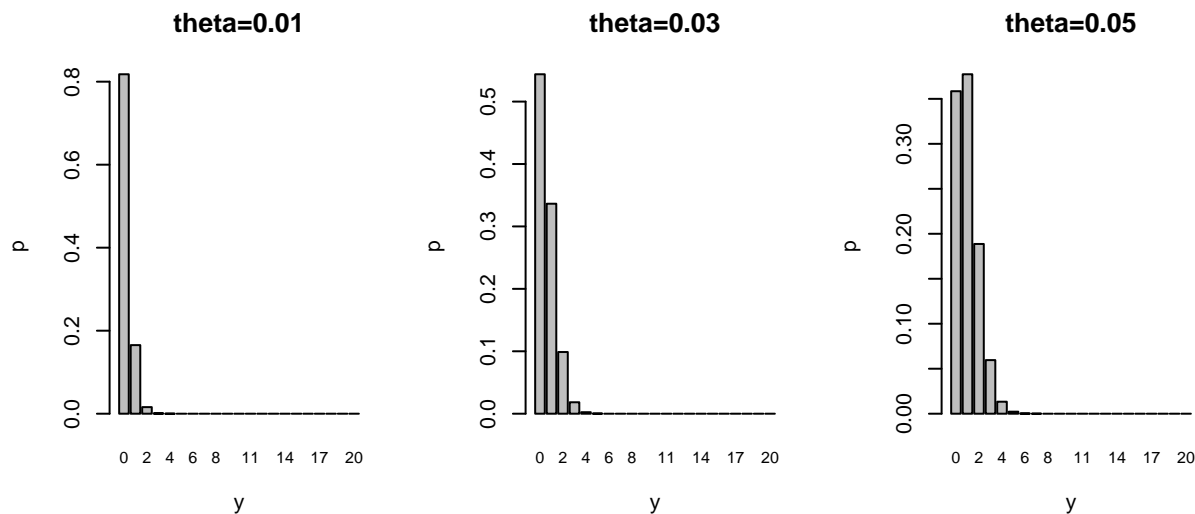
- $L(0.01) = 0.99^{20} \approx 0.8179069$
- $L(0.03) = 0.97^{20} \approx 0.5437943$
- $L(0.05) = 0.95^{20} \approx 0.3584859$

(a) 5 points

Plot probability mass function $p(y|\theta)$ as a function of y for the different values of θ (you should have 3 plots for the three different values of θ).

Answer

```
n<-20
theta<-c(0.01,0.03,0.05)
y<-0:n
pmf<-sapply(theta, function(x) dbinom(y, size=n, prob=x))
colnames(pmf)<-paste0('theta=',theta)
par(mfrow=c(1,3))
for (i in 1:3) {
  barplot(pmf[,i], names.arg=y, main=colnames(pmf)[i],
    xlab=expression(y), ylab=expression(p), cex.names=0.8)
}
```



```
par(mfrow=c(1,1))
```

As θ increases, the pmf shifts right. For the smaller range of θ values, most mass is concentrated at small counts. For $\theta = 0.01$, the mass at $y = 0$ is highest. So, we can make a fair estimation that the MLE will be at 0.01 for a discrete parameter space like this.

(b) 5 points

In classical inference, compute the maximum likelihood estimate for θ .

Answer

Since θ is restricted to parameter space Θ , we pick $\theta \in \Theta$ that maximizes $L(\theta) = (1 - \theta)^{20}$. If θ were allowed to vary continuously $[0, 1]$, the MLE for the Binomial distribution would be $\hat{\theta}_{\text{MLE}} = \frac{y}{n}$.

Because $(1 - \theta)^{20}$ is decreasing in θ on $[0, 1]$, the maximum over the three allowed values is the smallest θ . Thus, $\hat{\theta}_{\text{MLE}} = 0.01$.

(c) 5 points

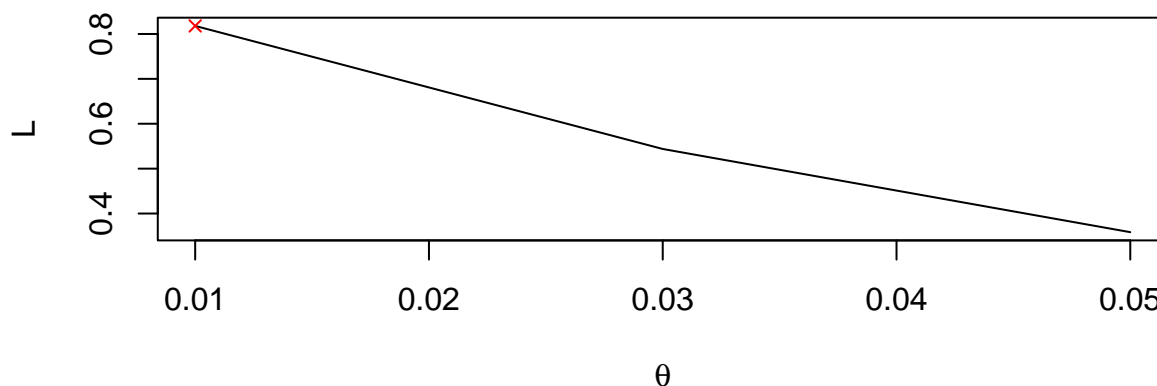
Plot the likelihood function as a function of θ and indicate the maximum likelihood found in part (b) on that plot.

Answer

```
likelihood_fn <- (1 - theta)^20

plot(theta, likelihood_fn, type= 'l', pch = 19, xlab = expression(theta),
      ylab = expression(L), main = "Likelihood Plot")
points(theta[which.max(likelihood_fn)],
       max(likelihood_fn), col = "red", pch = 4, cex = 0.8)
```

Likelihood Plot



The red mark indicates the maximum likelihood found in part (b). The estimator is at 0.01, and the estimate is at $L(0.01) \approx 0.8179069$.

(d) 10 points

In Bayesian inference, let us specify the following prior density $p(\theta)$ for θ :

$$p(0.01) = P(\theta = 0.01) = 0.2, \quad p(0.03) = P(\theta = 0.03) = 0.2, \quad p(0.05) = P(\theta = 0.05) = 0.6.$$

We can use the posterior mode (i.e., the value of θ having the highest posterior probability) to estimate θ . What's this estimate?

Compare the above two estimates (one from classical inference and one from Bayesian inference), and explain the difference.

Answer

Prior:

$$p(0.01) = 0.2, \quad p(0.03) = 0.2, \quad p(0.05) = 0.6$$

Posterior/discrete:

$$p(\theta|y) \propto p(\theta)L(\theta)$$

We now compute the unnormalized posterior:

- $u_1 = 0.2 \times L(0.01) = 0.1635814$
- $u_2 = 0.2 \times L(0.03) = 0.1087589$
- $u_3 = 0.6 \times L(0.05) = 0.2150916$

We compute the sum to find the normalized posterior. The sum is $u = u_1 + u_2 + u_3 = 0.4874318$

Now, we can proceed with normalization:

- $p(\theta = 0.01|y) = \frac{u_1}{u} = 0.3355985$
- $p(\theta = 0.03|y) = \frac{u_2}{u} = 0.2231263$
- $p(\theta = 0.05|y) = \frac{u_3}{u} = 0.4412752$

From the above, the posterior mass is largest at $\theta = 0.05$. Thus, we can conclude that the posterior mode is 0.05.

The MLE picked the parameter with the largest likelihood. The Bayesian posterior mode, on the other hand, combines both the likelihood and the prior. Here, the prior assigns 60% of the mass to $\theta = 0.05$. So even though 0.05 has a smaller likelihood than 0.01, the prior weight is large enough to make the posterior highest at $\theta = 0.05$.

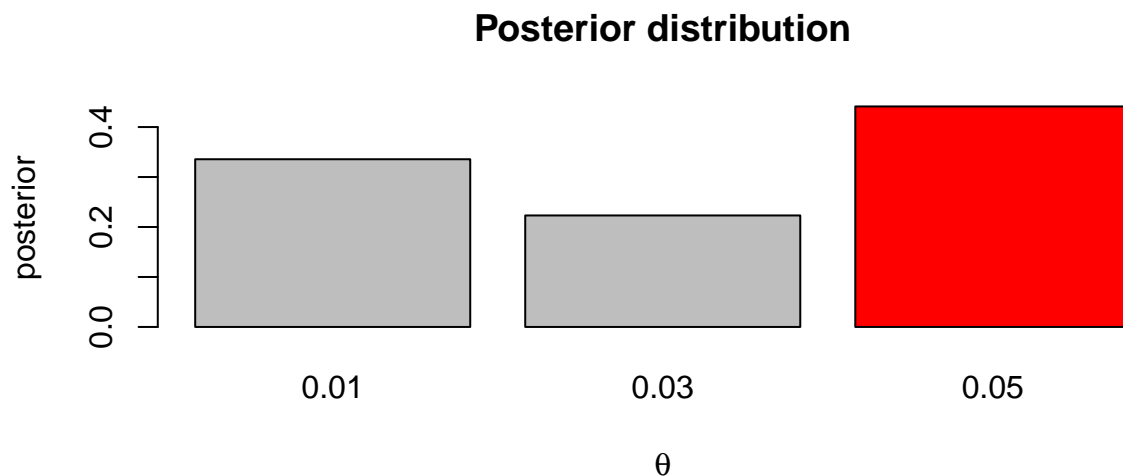
(e) 5 points

Plot the posterior distribution as a function of θ and indicate the posterior mode found in part (d) on that plot.

Answer

```
prior <- c(0.2,0.2,0.6)
unnorm <- prior * likelihood_fn
posterior <- unnorm / sum(unnorm)

barplot(posterior, col = c('grey', 'grey', 'red'), names.arg = theta,
        xlab = expression(theta), ylab = "posterior",
        main = "Posterior distribution")
```



The red bar indicates the posterior mode of 0.05.

(f) 10 points

Now, if we change the prior distribution in (b) to the following:

$$p(0.01) = P(\theta = 0.01) = \frac{1}{3}, p(0.03) = P(\theta = 0.03) = \frac{1}{3}, p(0.05) = P(\theta = 0.05) = \frac{1}{3}.$$

What's the posterior mode now? How does this estimate compare to the maximum likelihood estimate from (a)?

Answer

The posterior is proportional to likelihood times prior. With equal priors as given above, we can reduce the proportion to the following:

$$p(\theta|y) \propto L(\theta)$$

This is true since the prior becomes a part of the constant.

Since this is already normalized, the posterior mode is the θ with the largest likelihood, which according to our calculation using classical inference in (b), is $\theta = 0.01$.

Numerically, this will be:

- $p(\theta = 0.01|y) = \frac{1}{3}L(0.01) = 0.2726356$
- $p(\theta = 0.03|y) = \frac{1}{3}L(0.03) = 0.1812648$
- $p(\theta = 0.05|y) = \frac{1}{3}L(0.05) = 0.1194953$

Therefore, the posterior mode with a uniform prior will be 0.01. This matches the MLE estimated in (a) and confirmed in (b). This is because, intuitively, uniform priors do not give an extra preference, so the preference depends completely on likelihood. This is also probably why this is the same mode as the MLE among the discrete choices.