QUEEN MARY, UNIVERSITY OF LONDON

MTH6102: Bayesian Statistical Methods

Solutions of exercise sheet 2

2023-2024

(1) **10 points**. Let X be a discrete random variable with pmf $p(x|\theta)$, $\theta \in \{1, 2, 3\}$. One data point x is taken from $p(x|\theta)$. Find the MLE of θ.

x	$p(x \mid 1)$	$p(x \mid 2)$	$p(x \mid 3)$
0	1/3	1/4	0
1	1/3	1/4	0
2	0	1/4	1/4
3	1/6	1/4	1/2
4	1/6	0	1/4

Solution: For each value of x, the MLE $\hat{\theta}$ is the value of θ that maximises the likelihood $\mathcal{L}(\theta|x) = p(x \mid \theta)$. These values are in the following table.

x	0	1	2	3	4
$\hat{\theta}$	1	1	2 or 3	3	3

Thus, at $x = 2$, $\mathcal{L}(\theta|2) = 0$ when $\theta = 1$ and $\mathcal{L}(\theta|2) = 1/4$ when $\theta = 2$ or $\theta = 3$. So both $\hat{\theta} = 2$ or $\hat{\theta} = 3$ are both maxima.

(2) **20 points.** Let $Y_1, \ldots, Y_n$ be an iid sample from $N(\mu, \sigma^2)$, with both μ and σ^2 unknown.

(a) Find the likelihood and log likelihood functions.

(b) Find the maximum likelihood estimates $\hat{\mu}$ and $\hat{\sigma}$.

Solution: For each sample point, $y = (y_1, \ldots, y_n) \in \mathbb{R}$, the likelihood function of (μ, σ^2), $\mathcal{L}(\mu, \sigma^2 \mid y)$, is the joint density $f(y \mid \mu, \sigma^2)$ of $Y_1, \ldots, Y_n$. By independence,

$$\mathcal{L}(\mu, \sigma^2 \mid y) = f(y \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-(1/2) \sum_{i=1}^{n} (y_i - \mu)^2/\sigma^2\},$$

and the log likelihood is

$$\ell(\mu, \sigma^2 \mid y) = \log \mathcal{L}(\mu, \sigma^2 \mid y) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - (1/2) \sum_{i=1}^{n} (y_i - \mu)^2/\sigma^2.$$

The partial derivatives, with respect to μ and σ^2, are

$$\frac{\partial}{\partial \mu} \ell(\mu, \sigma^2 \mid y) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu),$$

$$\frac{\partial}{\partial \sigma^2} \ell(\mu, \sigma^2 \mid y) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \mu)^2.$$

Setting these partial derivatives equal to 0 and solving for μ and σ^2 yields the solution

$$\hat{\mu} = \bar{y} = n^{-1}\sum_{i=1}^{n} y_i, \quad \hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

Next, we need to verify that this solution is, in fact, a global maximum. First note that if $\mu \neq \bar{y}$, then $\sum_{i=1}^{n}(y_i - \mu)^2 > \sum_{i=1}^{n}(y_i - \bar{y})^2$. Hence, for any value of σ^2,

$$\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\{-(1/2)\sum_{i=1}^{n}(y_i - \bar{y})^2/\sigma^2\} \geq \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\{-(1/2)\sum_{i=1}^{n}(y_i - \mu)^2/\sigma^2\},$$

with equality if and only if $\mu = \bar{y}$. Hence, for any value of σ^2, $\hat{\mu} = \bar{y}$ is indeed a global maximum. Next, having verified that $\bar{y}$ maximises $\mathcal{L}(\mu, \sigma^2 \mid y)$ as a function of μ (for σ^2 fixed), using univariate calculus, it is easy to verify that the function $\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\{-(1/2)\sum_{i=1}^{n}(y_i - \bar{y})^2/\sigma^2\}$, as a function of σ^2, achieves its maximum at $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$. Hence, the estimators $\bar{Y}$ and $n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ are the MLEs of μ and σ^2, respectively.

(3) **20 points.** In a certain factory, machines D, E and F all produce computer chips of the same type. Of their production, machines D, E and F, respectively produce 2%, 3% and 1% defective chips. Machine D produces 30% of the output of the factory, machine E 25% and machine F the rest.

Suppose one chip is selected at random from the output of the factory and the chip is defective

(a) Use Bayes' theorem to find the probabilities that the chip was manufactured on machines D, E and F.

(b) Identify the data, hypotheses, likelihoods, prior probabilities and posterior probabilities.

(c) Redo the computation of (a) using a Bayesian updating table.

Solution:

(a) Relabel the machines as $1, 2, 3$ and let $\mathcal{H}_1$ be the event that a chip was produced by machine i. Let $\mathcal{D}$ be the event that a particular chip is defective. We have

$$P(\mathcal{H}_1) = 0.3, P(\mathcal{H}_2) = 0.25, P(\mathcal{H}_3) = 0.45.$$

$$P(\mathcal{D} \mid \mathcal{H}_1) = 0.02, P(\mathcal{D} \mid \mathcal{H}_2) = 0.03, P(\mathcal{D} \mid \mathcal{H}_2) = 0.01.$$

Applying the law of total probability, the probability that a random chip is defective is

$$P(\mathcal{D}) = \sum_{i=1}^{3} P(\mathcal{D} \mid \mathcal{H}_i)\, P(\mathcal{H}_i) = 0.006 + 0.0075 + 0.0045 = 0.018.$$

If it is defective, using Bayes' theorem the probability that it was manufactured by machine D (machine 1) is

$$P(\mathcal{H}_1 \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \mathcal{H}_1)\, P(\mathcal{H}_1)}{P(\mathcal{D})} = \frac{0.02 \times 0.3}{0.018} = 0.33.$$

The probability that it was manufactured by machine E (machine 2) is

$$P(\mathcal{H}_2 \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \mathcal{H}_2)\, P(\mathcal{H}_2)}{P(\mathcal{D})} = \frac{0.03 \times 0.25}{0.018} = 0.42.$$

$$P(\mathcal{H}_3 \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \mathcal{H}_3)\, P(\mathcal{H}_3)}{P(\mathcal{D})} = \frac{0.01 \times 0.45}{0.018} = 0.25.$$

(b)
- Hypotheses: We are testing three hypotheses, $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ that a chip was produced by machine 1, 2 and 3, respectively.
- Data: The result of our experiment. In this case, the chip is defective, $\mathcal{D}$=chip is defective.
- Prior probabilities: The prior are the probabilities of the hypotheses before testing the chip. In this, case

$$P(\mathcal{H}_1) = 0.3, P(\mathcal{H}_2) = 0.25, P(\mathcal{H}_3) = 0.45.$$

- Likelihood: The likelihood is the probability that the chip is defective, $\mathcal{D}$ (the data) given that the hypothesis $\mathcal{H}_i$ is true. In this case, there are three likelihoods, one for each hypothesis $\mathcal{H}_i$

$$P(\mathcal{D} \mid \mathcal{H}_1) = 0.02, P(\mathcal{D} \mid \mathcal{H}_2) = 0.03, P(\mathcal{D} \mid \mathcal{H}_2) = 0.01.$$

- Posterior probabilties: The posterior are the probabilities of the hypotheses given the data $\mathcal{D}$ (the chip is defective). In this case

$$P(\mathcal{H}_1 \mid \mathcal{D}), P(\mathcal{H}_2 \mid \mathcal{D}), P(\mathcal{H}_3 \mid \mathcal{D}).$$

(c) The Bayesian updating table is

Hypothesis	Prior	Likelihood	Bayes numerator	Posterior
$\mathcal{H}_1$	$P(\mathcal{H}_1) = 0.3$	$P(\mathcal{D} \mid \mathcal{H}_1) = 0.02$	$P(\mathcal{D} \mid \mathcal{H}_1)P(\mathcal{H}_1) = 0.02 \times 0.3 = 0.006$	$P(\mathcal{H}_1 \mid \mathcal{D}) = 0.33$
$\mathcal{H}_2$	$P(\mathcal{H}_2) = 0.25$	$P(\mathcal{D} \mid \mathcal{H}_2) = 0.03$	$P(\mathcal{D} \mid \mathcal{H}_2)P(\mathcal{H}_2) = 0.03 \times 0.25 = 0.0075$	$P(\mathcal{H}_2 \mid \mathcal{D}) = 0.42$
$\mathcal{H}_3$	$P(\mathcal{H}_3) = 0.45$	$P(\mathcal{D} \mid \mathcal{H}_3) = 0.01$	$P(\mathcal{D} \mid \mathcal{H}_3)P(\mathcal{H}_3) = 0.01 \times 0.45 = 0.0045$	$P(\mathcal{H}_3 \mid \mathcal{D}) = 0.25$
Total	1		$P(\mathcal{D}) = 0.018$	1

Law of total probability:

$$P(\text{data}) = P(\mathcal{D}) = \sum_{i=1}^{3} P(\mathcal{D} \mid \mathcal{H}_i)\, P(\mathcal{H}_i) = 0.006 + 0.0075 + 0.0045 = 0.018.$$

Bayes' theorem: posterior= $\frac{\text{prior}\times\text{likelihood}}{\text{total prob. of data}}$

$$P(\mathcal{H}_1 \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \mathcal{H}_1)\, P(\mathcal{H}_1)}{P(\mathcal{D})} = \frac{0.02 \times 0.3}{0.018} = 0.33.$$

$$P(\mathcal{H}_2 \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \mathcal{H}_2)\, P(\mathcal{H}_2)}{P(\mathcal{D})} = \frac{0.03 \times 0.25}{0.018} = 0.42.$$

$$P(\mathcal{H}_3 \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \mathcal{H}_3)\, P(\mathcal{H}_3)}{P(\mathcal{D})} = \frac{0.01 \times 0.45}{0.018} = 0.25.$$

(4) **50 points**. Suppose that you have recently started taking a train to work in a new location. You would like to estimate the probability q that the train arrives no more than 5 minutes late. Based on past experience living in South London, you assign a Beta distribution for q with parameters $\alpha = 5, \beta = 25$ as a prior distribution.

(a) What is the mean of the prior distribution?

Suppose that you observe k late arrivals in n journeys. For this observed data, use digits from your student ID number. Let the last three digits of your ID number be ABC. Then take $n = 10 + AB$ and $k = C$. (E.g. if the ID ends in ...092, then $n = 10 + 09 = 19, k = 2$; if the ID ends in ...374, then $n = 10 + 37 = 47, k = 4$)

(b) What is the maximum likelihood estimate $\hat{q}$ for q?
(c) What is the posterior distribution for q?
(d) What is the mean of the posterior distribution? What is the variance of the posterior distribution?
(e) What would the mean of the posterior distribution be if you had taken as a prior distribution the uniform distribution on $[0, 1]$?

Solution: This question is a direct application of the binomial example with Beta prior distribution from the lectures.

The prior mean for q is

$$\frac{\alpha}{\alpha + \beta} = \frac{5}{30} = 0.167.$$

Combining this prior with a Binomial likelihood, with k late trains observed out of n journeys and $k \sim \text{Binomial}(n, q)$, the result is a $\text{Beta}(k + \alpha, n - k + \beta)$ posterior distribution for q.

The posterior mean for q is

$$E(q \mid k) = \frac{k + \alpha}{n + \alpha + \beta}.$$

The variance of a $\text{Beta}(\alpha, \beta)$ random variable is

$$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Hence the variance of the posterior distribution is

$$Var(q \mid k) = \frac{(k + \alpha)(n - k + \beta)}{(n + \alpha + \beta)^2(n + \alpha + \beta + 1)}.$$

The maximum likelihood estimate is

$$\hat{q} = \frac{k}{n}.$$

If the prior distribution was uniform on $[0, 1]$, this would correspond to a Beta distribution with $\alpha = 1, \beta = 1$, and the posterior distribution would be $\text{Beta}(k + 1, n - k + 1)$. The posterior mean for q would be

$$E(q \mid k) = \frac{k + 1}{n + 2}.$$