First course in probability and statistics (MS-A0501, MS-A0502, MS-A0503, MS-0504)
Department of Mathematics and Systems Analysis, Aalto University
Jukka Kohonen, Pekka Pere, Natalia Vesselinova.

Exam 21.2.2024 suggested solutions

SOLUTION 1

a) A student passes the exam if 1) the teacher does not ask any question, 2) the teacher asks a single question and the student answers it correctly or 3) the teacher poses two questions and the student answers both of them correctly

$$\begin{aligned} \mathsf{P}(\text{pass}) &= p + p\pi + p\pi^2 \\ &= \frac{1}{3} + \frac{1}{3}\pi + \frac{1}{3}\pi^2 \end{aligned}$$

The three events (0, 1, or 2 questions) are mutually exclusive and collectively exhaustive. The probability of passing the exam is a weighted average of the conditional probability of passing under each of these events, weighted by the probability of each event occurring.

b) The probability to fail – given the adopted policy that all questions must be answered correctly – equals the probability to incorrectly answer a question, $1 - \pi$.

Alternatively, one can arrive at this result by calculating the conditional probability $\mathsf{P}(A \mid B)$, where A denotes the event that the student incorrectly answers the second question and B the event that the teacher asks 2 questions and the student answers correctly the first of them.

$$\mathsf{P}(A \mid B) = \frac{p\pi(1-\pi)}{p\pi} = 1 - \pi.$$

SOLUTION 2

In statistical terms, 20 Bernoulli experiments are carried out. In each experiment a green cube is chosen with probability 0.5. Let Y be the number of green cubes chosen so that $Y \sim \mathsf{Bin}(20, 0.5)$.

a) Six-months-olds pick 17 green cubes with probability **0.0011**:

$$\begin{aligned} \mathsf{P}(Y = 17) &= \binom{20}{17} 0.5^{17} \times 0.5^{(20-17)} \\ &= \frac{20!}{(20-3)! \times 3!} \times 0.5^{17} \times 0.5^{3} \\ &= \frac{20!}{17! \times 3!} \times 0.5^{20} \\ &= \frac{20 \times 19 \times 18}{3!} \times 0.5^{20} \\ &= \frac{20 \times 19 \times 18}{3 \times 2 \times 1} \times 0.5^{20} \\ &= 20 \times 19 \times 3 \times 0.5^{20} \\ &= 1140 \times 0.5^{20} \\ &\approx 0.001087189 \end{aligned}$$

$$\approx \mathbf{0.0011}$$

R command `dbinom(17,20,0.5)` returns the same probability.

b) Six-months-olds pick at least 17 green cubes with probability **0.0013**:

$$\begin{aligned}
\mathsf{P}(Y \geqslant 17) &= \sum_{i=17}^{20} \mathsf{P}(Y = i) \\
&= \sum_{i=17}^{20} \binom{20}{i} 0.5^i \times 0.5^{(20-i)} \\
&= \binom{20}{17} 0.5^{17} \times 0.5^3 + \binom{20}{18} 0.5^{18} \times 0.5^2 \\
&+ \binom{20}{19} 0.5^{19} \times 0.5^1 + \binom{20}{20} 0.5^{20} \times 0.5^0 \\
&\approx 0.001087 + 0.00018 + 0.000019 + 0.0000095 \\
&\approx \mathbf{0.0013}
\end{aligned}$$

The probability can be obtained conveniently by means of the probability of the complementary event

$$\mathsf{P}(Y \geqslant 17) = 1 - \mathsf{P}(Y < 17) = 1 - \mathsf{P}(Y \leqslant 16)$$

with the R command `1-pbinom(16,20,0.5)`. The calculations reveal that it would be very improbable for six-months-olds to choose 17 or more green cubes if they picked a green and red cube with equal probability. This sort of probability calculations and inference underlie the reasoning of the researchers.

SOLUTION 3

a) From the problem statement we have that

$$\mathsf{P}(T|D) = \mathbf{0.95}$$

By the chain rule:

$$\mathsf{P}(T \cap D) = \mathsf{P}(D)\mathsf{P}(T|D) = \frac{1}{100} \cdot 0.95 = \mathbf{0.0095}$$

By the total probability law:

$$\begin{aligned}
\mathsf{P}(T) &= \mathsf{P}(T \cap D) + \mathsf{P}(T \cap D^c) \\
&= \mathsf{P}(D)\mathsf{P}(T|D) + \mathsf{P}(D^c)\mathsf{P}(T|D^c) \\
&= 0.01 \cdot 0.95 + 0.99 \cdot 0.02 \\
&\approx 0.0293
\end{aligned}$$

By the Bayes' rule:

$$\mathsf{P}(D|T) = \frac{\mathsf{P}(T \cap D)}{\mathsf{P}(T)} = \frac{0.0095}{0.0293} \approx \mathbf{0.3242}$$

By complementing the probability to 1:

$$\mathsf{P}(D^c|T) = 1 - \mathsf{P}(D|T) \approx \mathbf{0.6758}$$

By the Bayes' rule:

$$\begin{aligned}\mathsf{P}(D|T^c) &= \frac{\mathsf{P}(D)\mathsf{P}(T^c|D)}{\mathsf{P}(T^c)} \\ &= \frac{\mathsf{P}(D)\mathsf{P}(T^c|D)}{1 - \mathsf{P}(T)} = \\ &= \frac{(1/100) \cdot 0.05}{1 - 0.0293} \\ &\approx \mathbf{0.0005}\end{aligned}$$

(Grade: 1 point question)
b) not graded
c) $\mathsf{P}(D|T)$ (Grade: 0.5 point) d) $0 + + - +$ (Grade: 0.5 point)

SOLUTION 4

a) The test statistic is

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{4.98 - 4}{1.30/\sqrt{289}} \approx \mathbf{12.815}.$$

b) The central limit theorem states that the sample average $\bar{X}$ of independent identically distributed random variables X_i with mean μ and variance σ^2 approximately follows the normal distribution $\mathsf{N}(\mu, \sigma^2/n)$ for large n. It follows that

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{large } n}{\sim} \mathcal{N}(0, 1).$$

According to the question, the assumptions for the central limit theorem apply. The figure 4.98 is a sample average of independent evaluations on the Likert-scale. The standard deviation 1.30 is known. The numerical value of the test statistic, 12.815, can be compared to the standard normal. Its 0.9995-fractile is 3.290527 (`qnorm(0.9995)`). The probability for the test statistic to take a value outside the range $(-3.290527, 3.290527)$ is 0.001 under the null hypothesis. Null hypothesis of mean 4 can be rejected at significance level 0.001, because $12.815 > 3.290527$. Our conclusion is that the people (of the intended target population) like Became Rich people more than "moderately" as described by integer 4.

SOLUTION 5

a) The likelihood function when observing 2 heads and 6 tails is given by

$$\mathcal{L}(\theta) = \theta^2(1 - \theta)^6.$$

The function is continuously differentiable in $[0, 1]$ and zero at the extremes of the interval. We follow the steps we delineated in the lecture namely we take the logarithm of the likelihood

$$l(\theta) = 2\log(\theta) + 6\log(1-\theta).$$

Taking its derivative, we get

$$l'(\theta) = \frac{2}{\theta} - \frac{6}{1-\theta},$$

which is 0 at $\theta = \mathbf{1/4}$. The likelihood is positive here, so the maximum is at $\theta = \mathbf{1/4}$ (not at the endpoints, where it evaluates to 0).

b) The unnormalised posterior is the product of the prior (given in the question) and the likelihood (described in the question and derived in a)):

$$30\theta^2(1-\theta)^2 \cdot \theta^2(1-\theta)^6 = 30 \cdot \theta^4(1-\theta)^8.$$

c) To find the MAP we note that the unnormalised posterior is continuously differentiable in $[0, 1]$, and is zero at the end points. To maximise the function, we can again take the logarithm to ease our job

$$4\log(\theta) + 8\log(1-\theta).$$

Its derivative is

$$\frac{4}{\theta} - \frac{8}{1-\theta},$$

and is zero at $\theta = \mathbf{1/3}$.
Alternatively, we can take the derivative of the unnormalised posterior to get

$$4\,\theta^3(1-\theta)^8 - 8\,\theta^4(1-\theta)^7$$

Setting it to 0, we obtain $1-\theta = 2\theta$, which gives $\theta = \mathbf{1/3}$.
The unnormalized posterior is positive here, so the maximum is at $\theta = \mathbf{1/3}$ (not at the endpoints, where the value is zero).