

Solutions to problems on hypothesis tests

1(a) We are given the two pdfs

$$\begin{aligned} f(x|s) &= 3(x-1)^2, \\ f(x|b) &= 3x^2, \end{aligned}$$

with $0 \leq x \leq 1$, and we want to select events of type s by requiring $x < x_{\text{cut}}$, with $x_{\text{cut}} = 0.1$. The probabilities to select events of type s and b are

$$\begin{aligned} P(x < x_{\text{cut}}|s) &= \int_0^{x_{\text{cut}}} f(x|s) dx = (x-1)^3 \Big|_0^{x_{\text{cut}}} = (x_{\text{cut}}-1)^3 + 1 \\ &= (0.1-1)^3 + 1 = 0.271 \\ P(x < x_{\text{cut}}|b) &= \int_0^{x_{\text{cut}}} f(x|b) dx = x^3 \Big|_0^{x_{\text{cut}}} = x_{\text{cut}}^3 \\ &= (0.1)^3 = 0.001 \end{aligned}$$

1(b) The signal purity is the probability for an event to be signal given that it is selected. To find this from the available ingredients we apply Bayes' theorem,

$$P(s|x < x_{\text{cut}}) = \frac{P(x < x_{\text{cut}}|s)\pi_s}{P(x < x_{\text{cut}}|s)\pi_s + P(x < x_{\text{cut}}|b)\pi_b} = \frac{(1 + (x_{\text{cut}} - 1)^3)\pi_s}{(1 + (x_{\text{cut}} - 1)^3)\pi_s + x_{\text{cut}}^3\pi_b},$$

where $\pi_s = 0.01$ and $\pi_b = 0.99$ are the given prior probabilities. Plugging in the numbers gives

$$P(s|x < x_{\text{cut}}) = \frac{0.271 \times 0.01}{0.271 \times 0.01 + 0.001 \times 0.99} = 0.732,$$

1(c) For an event with an observed value of x , the probability that it is background is again given by Bayes' theorem,

$$P(b|x) = \frac{f(x|b)\pi_b}{f(x|b)\pi_b + f(x|s)\pi_s} = \frac{x^2\pi_b}{x^2\pi_b + (x-1)^2\pi_s} = \frac{0.05^2 \times 0.99}{0.05^2 \times 0.99 + 0.95^2 \times 0.01} = 0.215.$$

1(d) The pdf $f(x|b) = 3x^2$ is concentrated towards one, and $f(x|s) = 3(x-1)^2$ towards zero. So if we observe $x = 0.05$, then values of x less than this represent less compatibility with $f(x|b)$. Therefore the p -value of the background hypothesis can be obtained as

$$p = \int_0^x f(x'|b) dx' = \int_0^x 3x'^2 dx' = x^3 = 0.05^3 = 1.25 \times 10^{-4}.$$

This is not the same as the probability for the event to be of type b, but rather the probability, assuming b, to observe x with equal or lesser compatibility with b than what was found with the actual data. Unlike the probability $P(b|x)$ found in (c), the p -value is independent of the prior probability for the event to be of type b.

1(f) We are now given two joint pdfs for x and y ,

$$\begin{aligned} f(x, y|s) &= 6(x-1)^2 y, \\ f(x, y|b) &= 6x^2(1-y), \end{aligned}$$

with $0 \leq x \leq 1$ and $0 \leq y \leq 1$. According to the Neyman-Pearson lemma, the test statistic that gives the highest power for a given significance level test (in this case equivalent to having the highest signal purity for a given efficiency), is given by the likelihood ratio

$$t(x) = \frac{f(x, y|s)}{f(x, y|b)} = \frac{(x-1)^2 y}{x^2(1-y)}.$$