Glen Cowan

RHUL Physics

25 August, 2010

Solutions to Statistics Problems for TAE 2010

1(a) We are given the two pdfs

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

with $0 \le x \le 1$, 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

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

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,

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

where $\pi_{\text{s}} = 0.01$ and $\pi_{\text{b}} = 0.99$ are the given prior probabilities. Plugging in the numbers gives

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

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

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

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

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

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(e) To generate values of x following $f(x|b) = 3x^2$ ($0 \leq x \leq 1$) using the acceptance-rejection method, we need to enclose the pdf in a box of height $f_{max} = 3$ and covering the range $0 \leq x \leq 1$, as shown in Fig. 1.
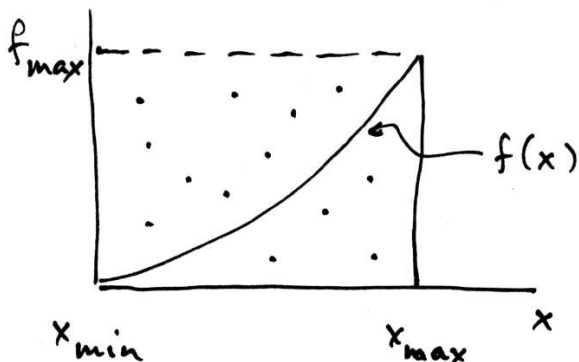

Figure 1: Illustration of the acceptance-rejection method.

To generate x we must (1) generate u uniform between 0 and f_{max}; (2) generate x uniform between $x_{min} = 0$ and $x_{max} = 1$; (3) if $u < f(x)$, accept x, otherwise reject. Points accepted by this procedure will follow the pdf $f(x)$.

To generate x values using the transformation method, we set the cumulative distribution function equal to r where r is uniformly distributed between 0 and 1. The cumulative distribution is

$$F(x|b) = \int_0^x f(x'|b)\, dx' = \int_0^x 3x'^2\, dx' = x^3 \; .$$

Setting this equal to r and solving for x gives the desired transformation

$$x(r) = r^{1/3} \; .$$

By evaluating this function with r values that are uniformly distributed in $[0, 1]$, one obtains x values that follow $f(x|b)$.

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

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