Statistical Methods for Particle Physics
TAE 2017

Solutions to problems sheet 1

1(a) [4 marks] The exponentially distributed time measurements, $t_1, \ldots, t_n$, and the Gaussian distributed calibration measurement y are all independent, so the likelihood is simply the product of the corresponding pdfs:

$$L(\tau, \lambda) = \prod_{i=1}^{n} \frac{1}{\tau + \lambda} e^{-t_i/(\tau+\lambda)} \, \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\lambda)^2/2\sigma^2} \; .$$

The log-likelihood is therefore

$$\ln L(\tau, \lambda) = -n \ln(\tau + \lambda) - \frac{1}{\tau + \lambda} \sum_{i=1}^{n} t_i - \frac{(y - \lambda)^2}{2\sigma^2} + C \; ,$$

where C represents terms that do not depend on the parameters and therefore can be dropped. Differentiating $\ln L$ with respect to the parameters gives

$$\begin{aligned} \frac{\partial \ln L}{\partial \tau} &= -\frac{n}{\tau + \lambda} + \frac{\sum_{i=1}^{n} t_i}{(\tau + \lambda)^2} \\ \frac{\partial \ln L}{\partial \lambda} &= -\frac{n}{\tau + \lambda} + \frac{\sum_{i=1}^{n} t_i}{(\tau + \lambda)^2} + \frac{y - \lambda}{\sigma^2} \; . \end{aligned}$$

Setting the derivatives to zero and solving for τ and λ gives the ML estimators,

$$\begin{aligned} \hat{\tau} &= \frac{1}{n} \sum_{i=1}^{n} t_i - y \\ \hat{\lambda} &= y \; . \end{aligned}$$

1(b) [4 marks] The variances of $\hat{\lambda}$ and $\hat{\tau}$ and their covariance are

$$\begin{aligned} V[\hat{\lambda}] &= V[y] = \sigma^2 \; , \\ V[\hat{\tau}] &= V\left[\frac{1}{n} \sum_{i=1}^{n} t_i - y\right] = \frac{1}{n^2} \sum_{i=1}^{n} V[t_i] + V[y] = \frac{(\tau + \lambda)^2}{n} + \sigma^2 \\ \mathrm{cov}[\hat{\tau}, \hat{\lambda}] &= \mathrm{cov}\left[\frac{1}{n} \sum_{i=1}^{n} t_i - y, y\right] = -V[y] = -\sigma^2 \; , \end{aligned}$$

For the covariance we used the fact that t_i and y are independent and thus have zero covariance.

1(c) [3 marks] The standard deviations of $\hat{\tau}$ and $\hat{\lambda}$ can be determined from the contour of $\ln L(\tau, \lambda) = \ln L_{\max} - 1/2$, as shown in Fig. 1. The standard can be approximated by the distance from the maximum of $\ln L$ to the tangent line to the contour (in either direction).

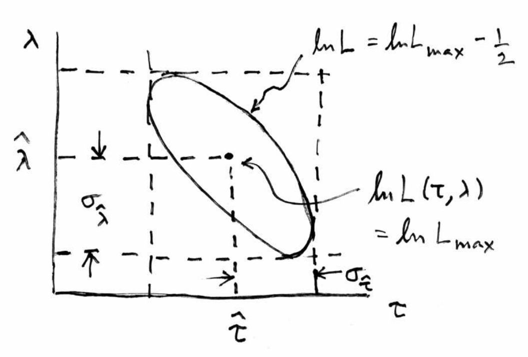

Figure 1: Illustration of the method to find $\sigma_{\hat{\tau}}$ and $\sigma_{\hat{\lambda}}$ from the contour of $\ln L(\tau, \lambda) = \ln L_{\max} - 1/2$ (see text).

If λ were to be known exactly, then the standard deviation of $\hat{\tau}$ would be less. This can be seen from Fig. 1, for example, since the distance one need to move τ away from the maximum of $\ln L$ to get to $\ln L_{\max} - 1/2$ would be less if λ were to be fixed at $\hat{\lambda}$.

1(d) [5 marks] The second derivatives of lnL are

$$
\begin{aligned}
\frac{\partial^2 \ln L}{\partial \tau^2} &= \frac{n}{(\tau+\lambda)^2} - \frac{2\sum_{i=1}^n t_i}{(\tau+\lambda)^3} \, , \\
\frac{\partial^2 \ln L}{\partial \lambda^2} &= \frac{n}{(\tau+\lambda)^2} - \frac{2\sum_{i=1}^n t_i}{(\tau+\lambda)^3} - \frac{1}{\sigma^2} \, , \\
\frac{\partial^2 \ln L}{\partial \tau \partial \lambda} &= \frac{n}{(\tau+\lambda)^2} - \frac{2\sum_{i=1}^n t_i}{(\tau+\lambda)^3} \, .
\end{aligned}
$$

Using $E[t_i] = \tau + \lambda$ we find the expectation values of the second derivatives,

$$
\begin{aligned}
E\left[\frac{\partial^2 \ln L}{\partial \tau^2}\right] &= \frac{n}{(\tau+\lambda)^2} - \frac{2n(\tau+\lambda)}{(\tau+\lambda)^3} = -\frac{n}{(\tau+\lambda)^2} \, , \\
E\left[\frac{\partial^2 \ln L}{\partial \lambda^2}\right] &= -\frac{n}{(\tau+\lambda)^2} - \frac{1}{\sigma^2} \, , \\
E\left[\frac{\partial^2 \ln L}{\partial \tau \partial \lambda}\right] &= -\frac{n}{(\tau+\lambda)^2} \, .
\end{aligned}
$$

The inverse covariance matrix of the estimators is given by

$$
V_{ij}^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j}\right]
$$

where here we can take, e.g., $\theta_1 = \tau$ and $\theta_2 = \lambda$. We are given the formula for the inverse of the corresponding 2×2 matrix, and by substituting in the ingredients we find

$$
V = \begin{pmatrix} \frac{(\tau+\lambda)^2}{n} + \sigma^2 & -\sigma^2 \\ -\sigma^2 & \sigma^2 \end{pmatrix}
$$

which are the same as what was found in (c).

2(a) The likelihood function is given by the binomial distribution evaluated with the single observed value n and regarded as a function of the unknown parameter θ:

$$L(\theta) = \frac{N!}{n!(N-n)!} \theta^n (1-\theta)^{N-n} .$$

The log-likelihood function is therefore

$$\ln L(\theta) = n \ln \theta + (N-n) \ln(1-\theta) + C ,$$

where C represents terms not depending on θ. Setting the derivative of $\ln L$ equal to zero,

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \frac{N-n}{1-\theta} = 0 ,$$

we find the ML estimator to be

$$\hat{\theta} = \frac{n}{N} .$$

2(b) We are given the expectation and variance of a binomial distributed variable as $E[n] = N\theta$ and $V[n] = N\theta(1-\theta)$. Using these results we find the expectation value of $\hat{\theta}$ to be

$$E[\hat{\theta}] = E\left[\frac{n}{N}\right] = \frac{E[n]}{N} = \frac{N\theta}{N} = \theta ,$$

and therefore the bias is $b = E[\hat{\theta}] - \theta = 0$. Similarly we find the variance to be

$$V[\hat{\theta}] = V\left[\frac{n}{N}\right] = \frac{1}{N^2} V[n] = \frac{N\theta(1-\theta)}{N^2} = \frac{\theta(1-\theta)}{N} .$$

2(c) Suppose we observe $n = 0$ for $N = 10$ trials. The upper limit on θ at a confidence level of $\mathrm{CL} = 1 - \alpha$ is the value of θ for which there is a probability α to find as few events as we found or fewer, i.e.,

$$\alpha = P(n \le 0; N, \theta) = \frac{N!}{0!(N-0)!} \theta^0 (1-\theta)^{N-0} .$$

Solving for θ gives the 95% CL upper limit

$$\theta_{\mathrm{up}} = 1 - \alpha^{1/N} = 1 - 0.05^{1/10} = 0.26 .$$

3(a) We are given the two pdfs

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

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

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

3(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,

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

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

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

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

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

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

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

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 $P({\rm b}|x)$ found in (c), the p-value is independent of the prior probability for the event to be of type b.

1a) $$\text{p-value} = P(n \geq \overset{n_{obs}=16}{n_{obs}} \mid s = 0,\ b = 3.9)$$

$$= \sum_{n=n_{obs}}^{\infty} \frac{b^n}{n!} e^{-b}$$

$$= 1 - \sum_{n=0}^{n_{obs}-1} \frac{b^n}{n!} e^{-b}$$

$$= F_{\chi^2}(2b;\, 2n_{obs}) \qquad \Big[n_{dof} = 2(\underset{\substack{\uparrow \\ \text{upper lim. of sum}}}{m}+1) = 2n_{obs} \Big]$$

$= 1 -$ `TMath::Prob(7.8, 32)`

$$= 3.58 \times 10^{-6}$$

1b) $$Z = \Phi^{-1}(1-p) = 4.5$$

$=$ `TMath::NormQuantile(1-p)`