Principles of Data Science Coursework Report

James Hughes

November 28, 2023

Section A

Part (a)

We begin by showing that both densities s and b are properly normalised in the range $M \in [-\infty, +\infty]$. In the former case, as a first step we use a change of variables $Z = \mu + \sigma M$ such that $\frac{dM}{dZ} = \sigma$, for which the integral limits don't change:

$$\begin{aligned}
\int_{-\infty}^{\infty} s(M;\mu,\sigma) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(M-\mu)^2}{2\sigma^2}\right] dM \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}Z^2) \cdot \sigma dZ \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}Z^2) dZ
\end{aligned}$$

In order to prove that s is properly normalised, we simply need to show that the last expression above evaluates to 1. We do this by computing it's square, which in turn leads to an integral in two dummy variables. Below we then use the transformation to polar coordinates $(X, Y) = \rho(R, \phi) = (R\cos(\phi), R\sin(\phi))$ which has Jacobian matrix

$$\begin{pmatrix} \cos(\phi) & -R\sin(\phi) \\ \sin(\phi) & R\cos(\phi) \end{pmatrix}$$

and hence $|J_\rho(R, \phi)| = R$. Then, denoting the integral in the final expression above by I, we find:

$$
\begin{aligned}
\left[\frac{1}{\sqrt{2\pi}} I\right]^2 &= \frac{1}{2\pi}\left[\int_{-\infty}^{\infty} \exp(-\frac{1}{2}X^2)dX\right]\left[\int_{-\infty}^{\infty} \exp(-\frac{1}{2}Y^2)dY\right] \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\frac{1}{2}(X^2+Y^2))dXdY \\
&= \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty} \exp(-\frac{1}{2}(R^2)) \cdot RdRd\phi \\
&= \left[-\exp(-\frac{1}{2}R^2)\right]_{R=0}^{R=\infty} \\
&= 1.
\end{aligned}
$$

For the background, we note that the density (integrand) b is zero for all $M < 0$, and show

$$
\begin{aligned}
\int_{-\infty}^{\infty} b(M;\lambda) &= \int_{0}^{\infty} \lambda e^{-\lambda M} dM \\
&= \left[-e^{-\lambda M}\right]_{M=0}^{M=\infty} \\
&= 1.
\end{aligned}
$$

Finally this lets us show that the probability density p given is properly normalised over $[-\infty, +\infty]$ because

$$
\begin{aligned}
\int_{-\infty}^{\infty} p(M; f, \lambda, \mu, \sigma)dM &= f\int_{-\infty}^{\infty} s(M;\mu,\sigma)dM + (1-f)\int_{-\infty}^{\infty} b(M;\lambda)dM \\
&= f \cdot 1 + (1-f)\cdot 1 \\
&= 1.
\end{aligned}
$$

Part (b)

In order to ensure that the fraction of signal in the restricted distribution for M remains f, we must normalise the signal and background separately over $[\alpha, \beta]$, we introduce these new restricted distributions as

$$s_r(M;\mu,\sigma) = \frac{s(M;\mu,\sigma)}{\int_\alpha^\beta s(M;\mu,\sigma)dM} \qquad b_r(M;\lambda) = \frac{b(M;\lambda)}{\int_\alpha^\beta b(M;\lambda)dM}$$

We then compute the relevant integrals:

$$\begin{aligned}
\int_\alpha^\beta s(M;\mu,\sigma)dM &= \int_{-\infty}^\beta s(M;\mu,\sigma)dM - \int_{-\infty}^\alpha s(M;\mu,\sigma)dM \\
&= \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\beta-\mu}{\sigma\sqrt{2}}\right)\right] - \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\alpha-\mu}{\sigma\sqrt{2}}\right)\right] \\
&= \frac{1}{2}\operatorname{erf}\left(\frac{\beta-\mu}{\sigma\sqrt{2}}\right) - \frac{1}{2}\operatorname{erf}\left(\frac{\alpha-\mu}{\sigma\sqrt{2}}\right)
\end{aligned}$$

$$\begin{aligned}
\int_\alpha^\beta b(M;\lambda)dM &= \int_{-\infty}^\beta b(M;\lambda)dM + \int_{-\infty}^\alpha b(M;\lambda)dM \\
&= 1 - e^{-\lambda\beta} - (1 - e^{-\lambda\beta}) \\
&= e^{-\lambda\alpha} - e^{-\lambda\beta}
\end{aligned}$$

The properly normalised probability density function for $M \in [\alpha, \beta]$ is then

$$p_r(M;\boldsymbol{\theta}) = \frac{2f}{\operatorname{erf}\left(\frac{\beta-\mu}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{\alpha-\mu}{\sigma\sqrt{2}}\right)} s(M;\mu,\sigma) + \frac{1-f}{e^{-\lambda\alpha} - e^{-\lambda\beta}} b(M;\lambda)$$

where $s(M;\mu,\sigma)$ and $b(M;\lambda)$ are as given on the sheet. Note that the value for $p_r(M)$ is zero when $M \notin [\alpha, \beta]$.

Part (c)

Section B