Principles of Data Science Coursework Report

James Hughes

November 28, 2023

1 Section A

1.1 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 $(X, Y) = \rho(R, \theta) = (R\cos(\theta), R\sin(\theta))$ which has Jacobian matrix

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

and hence $|J_\rho(R,\theta)| = R$. Then

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

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

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

1.2 Part (b)

2 Section B