Worksheet 1: Practice with Joint Measures

Name:b Aaron Zoll
Due September 8, 2023

You will investigate the rule of succession which states that the probability of success for the $(m+1)$ th trial in a sequence of independent bernoulli random variables $x_1, \ldots, x_{m+1}$, given $k \leq m$ successes in the first m trials is

$$\mathbb{P}\left(x_{m+1}=1 \mid \sum_{j=1}^{m} x_j = k\right) = \frac{k+1}{m+2}.$$

In principle if $x_1, \ldots, x_{m+1}$ are truly independent, you might expect the right hand side to simply be $\mathbb{P}\left(x_{m+1}=1\right)$ which does not depend on k (i.e. the outcomes of $x_1, \ldots, x_m$).

In this worksheet, you will explore which space to make sense of eq. (1). While seemingly elementary, this example provides a clear illustration of Bayesian reasoning in machine learning, where model-defining parameters are defined by a measure instead of as point-values. We shall let $\mathcal{X}=\{0,1\}, x \in \mathcal{X}$ denote an outcome, and let $s_{\mathrm{m}}: \mathcal{X}^{\mathrm{m}} \rightarrow \mathbb{R}$ be defined by $(x_1, \ldots, x_m) \mapsto \sum_j^m x_j$. Implicit in eq. (1) is that $p:=\mathbb{P}(x=1)$ is unknown (i.e. there is some distribution on p) and that given p, then the trials $x_1, \ldots, x_{m+1}$ are independent. To concretize this observation, we let $I=[0,1]$ and suppose that there is joint distribution $\mathbb{P}_{\mathcal{X}^{\mathrm{m}} \times \mathrm{I}}$ according to which $\mathbb{P}_{\mathcal{X}^{\mathrm{m}} \mid \mathrm{I}}$ is independent. In particular,

$$\mathbb{P}_{\mathcal{X}^{\mathrm{m}} \mid \mathrm{I}}\left(\sum_{j=1}^{m} x_j = k \mid i = p\right) = \binom{m}{k} p^k (1-p)^{m-k}.*$$

Because p does not appear explicitly in eq. (1), it must be (being) marginalized away. That means that

$$\mathbb{P}\left(x_{\mathrm{m}+1}=1 \mid s_{\mathrm{m}}\left(x_1, \ldots, x_{\mathrm{m}}\right)=\mathrm{k}\right)=\mathbb{P}_{\mathcal{X} \times \mathrm{I} \mid \mathcal{X}^{\mathrm{m}}}\left(\left\{x_{\mathrm{m}+1}=1\right\} \times \mathrm{I} \mid s_{\mathrm{m}}\left(x_1, \ldots, x_{\mathrm{m}}\right)=\mathrm{k}\right)$$

A reasonable measure to place on the marginal probability $\mathbb{P}_{\mathrm{I}}$ is the uniform one $\mathbb{P}_{\mathrm{I}}((\mathrm{a}, \mathrm{b}))=\mathrm{b}-\mathrm{a}$ for $0 \leq \mathrm{a} \leq \mathrm{b} \leq 1$. With this preamble, you may now proceed.

1. Using the definition of conditional probability, rewrite eq. (2) as a ratio of two joint probabilities. You may use that $\mathcal{X} \times \mathrm{I} \times \mathcal{X}^{\mathrm{m}} \cong \mathcal{X}^{\mathrm{m}+1} \times \mathrm{I}$ to simplify (but be careful to write the correct event in the numerator!). In the denominator, express the marginal probability $\mathbb{P}_{\mathcal{X}}$ in terms of $\mathbb{P}_{\mathcal{X}^{\mathrm{m}} \times \mathrm{I}}$; you will see why in the next problem.

Solution

$$\begin{aligned}
\mathbb{P}_{\mathcal{X}|\mathcal{X}^m}(x_{m+1}=1|s_m(x_1,...,x_m)=k) &= \mathbb{P}_{\mathcal{X}\times I|\mathcal{X}^m}(\{x_{m+1}=1\}\times I|s_m(x_1,...,x_m)=k) \\
&= \frac{\mathbb{P}_{\mathcal{X}\times I\times X^m}(\{x_{m+1}=1\}\times I\times s_m(x_1,...,x_m)=k)}{\mathbb{P}_{\mathcal{X}^m}(s_m=1)} \\
&= \frac{\mathbb{P}_{\mathcal{X}^{m+1}\times I}(\{x_{m+1}=1\}\times\{s_m(x_1,...,x_m)=k\}\times I)}{\mathbb{P}_{\mathcal{X}^m\times I}(\{s_m=1\}\times I)}
\end{aligned}$$

* See Binomial Distribution for more information.

2. Conveniently, this expression (nearly) collapses the problem to a single computation (of either numerator or denominator). To compute either probability, you will want to condition again, this time as $\mathcal{X}^{\mathrm{m}} \mid \mathrm{I}$, and use the law of total probability. † Express each one:

Solution
Let $A = \{x_{m+1} = 1\}$ and $C = \{s_m = k\}$ Numerator (N):

$$\begin{aligned}
&\mathbb{P}_{\mathcal{X}^{m+1} \times I}(\{x_{m+1} = 1\} \times \{s_m(x_1, ..., x_m) = k\} \times I) \\
&= \int_I \int_{A \times C|I} d\mathbb{P}_{\mathcal{X} \times \mathcal{X}^m|I}(x_1, ..., x_m|i) d\mathbb{P}_I(i) \\
&= \int_I \int_{A|I} \int_{C|I} d\mathbb{P}_{\mathcal{X}|I}(x|i) d\mathbb{P}_{\mathcal{X}^m|I}(x_1, ..., x_m|i) d\mathbb{P}_I(i) \text{ (by independence)} \\
&= \int_I \mathbb{P}_{\mathcal{X}|I}(A|I)\mathbb{P}_{\mathcal{X}^m|I}(C|I) d\mathbb{P}_I(i) \\
&= \int_0^1 \mathbb{P}_{\mathcal{X}|I}(A|i = p)\mathbb{P}_{\mathcal{X}^m|I}(C|i = p) dp \\
&= \int_0^1 p \binom{m}{k} p^k (1 - p)^{m-k} dp \\
&= \int_0^1 \binom{m}{k} p^{k+1} (1 - p)^{m-k} dp
\end{aligned}$$

Similarly, the denominator (D):

$$\begin{aligned}
&\mathbb{P}_{\mathcal{X}^m \times I}(\{s_m(x_1, ..., x_m) = k\} \times I) \\
&= \int_I \int_{C|I} d\mathbb{P}_{\mathcal{X}^m|I}(x_1, ..., x_m|i) d\mathbb{P}_I(i) \\
&= \int_I \mathbb{P}_{\mathcal{X}^m|I}(C|I) d\mathbb{P}_I(i) \\
&= \int_0^1 \mathbb{P}_{\mathcal{X}^m|I}(C|i = p) dp \\
&= \int_0^1 \binom{m}{k} p^k (1 - p)^{m-k} dp
\end{aligned}$$

3. Explicitly compute the inner (conditional) probability (or expectation) for both:
Solution [see above]

4. Show by induction on k that $\int_0^1 p^k(1-p)^{m-k}dp = \frac{k!(m-k)!}{(m+1)!}$.
Solution

$$\begin{aligned}
&\text{For } k=0\\
&= \int_0^1 p^0(1-p)^m dp\\
&= -\frac{1}{m+1}(1-p)^{m+1}\Big|_0^1\\
&= \frac{1}{m+1}\\
&= \frac{0!(m-0)!}{(m+1)!}
\end{aligned}$$

$$\begin{aligned}
&\text{Suppose this holds for k} < \text{m}\\
&\text{For } j=k+1\\
&= \int_0^1 p^j(1-p)^{m-j}dp\\
&= \int_0^1 p^{k+1}(1-p)^{m-k-1}dp\\
&\text{Let } u = p^{k+1},\ dv = (1-p)^{m-k-1}\\
&du = (k+1)p^k,\ v = -\frac{1}{m-k}(1-p)^{m-k}\\
&= uv\Big|_0^1 + \frac{k+1}{m-k}\int_0^1 p^k(1-p)^{m-k}dp\\
&= 0 + \frac{(k+1)k!(m-k)!}{(m-k)(m+1)!}\\
&= \frac{(k+1)!(m-k-1)!}{(m+1)}\\
&= \frac{j!(m-j)!}{(m+1)!}
\end{aligned}$$

5. Show also, using part 4 . and judicious integration by parts that $\int_0^1 p^{k+1}(1-p)^{m-k}dp = \frac{(k+1)!(m-k)!}{(m+2)!}$.
Solution (no by parts needed!)

$$\begin{aligned}
&\int_0^1 p^{k+1}(1-p)^{m-k}dp\\
&= \int_0^1 p^{k+1}(1-p)^{m+1-(k+1)}dp\\
&= \frac{(k+1)!((m+1)-(k+1))!}{((m+1)+1)!}\\
&= \frac{(k+1)!(m-k)!}{(m+2)!}
\end{aligned}$$

6. Use results in parts 3., 4., and 5. to compute the expression (4) and simplifying, conclude (1).
Solution Thus, we get

$$\begin{aligned}
(1) &= \frac{N}{D}\\
&= \frac{\binom{m}{k}\frac{(k+1)!(m-k)!}{(m+2)!}}{\binom{m}{k}\frac{(k)!(m-k)!}{(m+1)!}}\\
&= \frac{(k+1)!(m+1)!}{k!(m+2)!}\\
&= \frac{k+1}{m+2}
\end{aligned}$$