# Worksheet 1: Practice with Joint Measures

Name: 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

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

In principle if $x_1, \ldots, x_{m+1}$ are truly independent, you might expect the right hand side to simply be $\mathbb{P}(x_{m+1} = 1)$ 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 $s_m : \mathcal{X}^m \to \mathbb{R}$ 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 $\mathbb{P}_{\mathcal{X}^m \times I}$ according to which $\mathbb{P}_{\mathcal{X}^m|I}$ is independent. In particular,

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

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

$$\mathbb{P}\left(x_{m+1} = 1 \,| s_m(x_1, \ldots, x_m) = k\right) = \mathbb{P}_{\mathcal{X} \times I|\mathcal{X}^m}\left(\left\{x_{m+1} = 1\right\} \times I \,| s_m(x_1, \ldots, x_m) = k\right) \tag{2}$$

A reasonable measure to place on the marginal probability $\mathbb{P}_I$ is the uniform one $\mathbb{P}_I((a, b)) = b - a$ for $0 \leq a \leq b \leq 1$. 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 $\mathcal{X} \times I \times \mathcal{X}^m \cong \mathcal{X}^{m+1} \times I$ to simplify (but be careful to write the correct event in the numerator!). In the denominator, express the marginal probability $\mathbb{P}_{\mathcal{X}^m}$ in terms of $\mathbb{P}_{\mathcal{X}^m \times I}$; you will see why in the next problem.

$$\tag{3}$$

*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 $\mathcal{X}^m|I$, and use the law of total probability.† Express each one:

$$ \tag{4}$$

3. Explicitly compute the inner (conditional) probability (or expectation) for both:

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

5. Show also, using part 4. and judicious integration by parts that $\displaystyle\int_0^1 p^{k+1}(1-p)^{m-k}dp = \frac{(k+1)!(m-k)!}{(m+2)!}$.

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

†Recalling that probability is an expectation $\mathbb{P}_{\mathcal{X}}(A) = \displaystyle\int_{\mathcal{X}} \mathbb{1}_{x\in A}\, d\mathbb{P}_{\mathcal{X}}(x)$, you may also use the law of total expectation.