1 Glossary

- Explanatory variable (input variable, predictor, covariate): Often written as x. Generally written as x if the data has been observed, $\tilde{x}$ if it hasn't.
- Outcome variable (output variable, target, response): Often written as y. Generally written as y if the data has been observed, $\tilde{y}$ if it hasn't.
- Data: The combination of input and output variables that have been observed. $D = (x_i, y_i)$.
- Model: Some way (parametric or non) of relating the input and output variables. $y = M(x, \theta)$.

Probabilities

- Conditional probability: The probability of some outcome (e.g. winning the game), given some other outcome (scoring three goals); $P(x|y)$.
- Joint probability: The probability of both events occuring (e.g. winning the game **and** scoring three goals); $P(x, y)$. Note

$$P(x, y) = P(x)P(y|x) = P(y)P(x|y) \tag{1}$$

  from which Bayes' theorem follows.

Basic Bayesian components

- Bayes' rule: The posterior probability of some set of parameters is equal to the prior, π, multiplied by the likelihood, $\mathcal{L}$, scaled by the evidence, $\mathcal{Z}$.

$$P(\theta|D, M) = \frac{\mathcal{L}(\mathcal{D}|\theta, \mathcal{M})\pi(\theta)}{\mathcal{Z}} \tag{2}$$

  Note that this can also be written less specifically as,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

- Likelihood: The liklihood of the data given the model (M), its parameters (θ); $\mathcal{L}(D|M, \theta)$
- Prior probability distribution: The probability of the parameters ($\pi(\theta)$) for the model, before conditioning on the data. This maybe be determined by previous work, or natural limit (some properties, such as variance, must be > 0).
- Posterior probability distribution: The probability of the parameters (θ) for the model, given the data. $p(\theta|D, M)$

- Evidence (model evidence, bayesian evidence, marginal liklihood): The probability of the data, marginalized (integrated) over the domain of the parameter space;

$$P(D|M) \equiv \mathcal{Z} = \int \mathcal{L}(D|\theta, M)\pi(\theta)d\theta \tag{4}$$

  Thinking of this as "model evidence" is nice! This is the total evidence for the model. If we assume that the prior is normalized (integrates to 1), this can be thought of as the average, weighted by the prior.

Advanced Bayesian components

- Prior predictive distribution: The prediction if y for x, given a model and a prior distribution over the parameters for that model. $p(y|x) = \int p(y|x,\theta)p(\theta)d\theta$.
- Posterior predictive distribution: The prediction of $\tilde{y}$ for $\tilde{x}$, given a model conditioned on some data; $p(\tilde{y}|\tilde{x}, D, M)$. In the process of conditioning on the data, we obtain a posterior on the model parameters. Thus, $p(\tilde{y}|\tilde{x}, D, M) = \int p(\tilde{y}|\tilde{x}, \theta, M)p(\theta|D, M)d\theta$, or the probaility of $\tilde{y}$ at a given θ, weighted by the posterior probability of that θ.
- Latent variable
- Bayes Factor: The amount by which the data has changed our opinion about two models. For model comparison, we ask the question,

$$\frac{P(M_1|D)}{P(M_2|D)} \tag{5}$$

  We know from above that

$$\mathcal{Z} \equiv P(D|M) \tag{6}$$

  and so we just need to invert this conditional probability using bayes theorem.

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)} = \frac{\mathcal{Z}P(M)}{P(D)} \tag{7}$$

  and so the ratio of probabilies we initially wanted to know about is (as the $P(D)$ cancels),

$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{\mathcal{Z}_1 P(M_1)}{\mathcal{Z}_2 P(M_2)} \tag{8}$$

  where $P(M)$ is the prior on that model. Ignoring the priors, we get the Bayes Factor, which is just the ratio of evidences

$$K = \frac{\mathcal{Z}_1}{\mathcal{Z}_2} \tag{9}$$

Let's give a full example. Consider subhalo abundance matching, a method of populating dark matter halos with stellar mass. Our target, y, is the stellar mass. We choose an explanatory variable (such as halo mass) that we thing correlates well (covaries) with this - this is x. We have some model M to compute $y|x$. This model parameterizes the SMHM relation and its scatter $y = M(x, \theta)$. To find those other parameters, we find the liklihood of our data given the model and θ. To do this, we also assume a cosmology and that an N-body simulation is a good realization of the universe. The Likelihood compares the summary statistic of the N-body + model and observed data.