SR HW 7

Emilio Horner

2022-10-20

7E1.

1. The measure should be continuous 2 It should increase as the events also increases 3. It should be additive

7E2.

```
 p <- c( 0.3 , 0.7 )
-sum( p*log(p) )
```

```
## [1] 0.6108643
```

This question is the same as the example on page 206. The entropy is .61.

7E3.

Suppose a four-sided die is loaded such that, when tossed onto a table, it shows "1" 20%, "2" 25%, "3" 25%, and "4" 30% of the time. What is the entropy of this die?

```
p <- c(0.20, 0.25, 0.25, 0.30)
(X <- -sum(p * log(p)))
```

```
## [1] 1.376227
```

The entropy of the dice is 1.38

7E4.

```
p <- c(1/3,1/3,1/3)
(X <- -sum(p * log(p)))
```

```
## [1] 1.098612
```

The entropy of the dice this time is 1.09.

7M1.

The AIC evaluates how well a model fits the data that the model was generated from. It is calculated from the number of independent variables used in the model and the likelihood estimate of the model. The best fit model is the one that explains the greatest variation using fewest possible independent variables.

The WAIC offers the approximation of the out-of-sample deviance that converges to the cross-validation approximation in a large sample.

WAIC is the more general. To go from WAIC to AIC one needs to assume that the posterior distribution is a multivariate normal distribution, and the priors are flat.

7M2.

Model selection means choosing the model with the lowest criterion value and eliminating the other models. This is not a good practice because one loses the differences in the CV/PSIS/WAIC values. These different values can reveal how confident one is in the model. Also maximizing predictive accuracy is not the same as

causation. Highly confounded models can sometimes make good short term predictions. Instead one should use model comparison which is more general and uses multiple models to understand how variables influence predictions.

7M3.

The information criterion need the same number of observations because the deviance is based on the sum of the observations. More observations will lead to a higher deviance and therefore not be comparable to a model with less observations.

7M4.

As the prior gets more concentrated the number of parameters will decrease. The model essentially becomes less flexible at fitting the sample.

7M5.

Overfitting is when a model learns too much from a sample. The informative prior tells the model to ignore or reduce in value certain unrealistic values.

7M6.

If the priors are overly informative the model would be constrained and would have reduced predictive performance.