Bayes class 8 – Name:

Sean van der Merwe

# Reflection

Let’s start with some basics.

If $X\~N(0,1)$ and $Y\~N\left(0,1\right)$ then what is $P\left(X>Y\right)$? (5 seconds)

If $X\~N(0,1)$ and $Y\~N\left(2,1\right)$ with correlation $-0.5$ then what is $P\left(X>Y\right)$? (2 minutes, no calculator)

If $X\~N(3,1)$ and $Y\~N\left(2,1\right)$ with correlation $-0.5$, and $W\~N(0,3^{2})$ independently, then what is $P\left(X>(Y^{2}+2W)\right)$? (write down the appropriate computer code below).

**Remember**: When we can simplify a problem we must do so, saving time and gaining accuracy. Similarly, when a problem cannot be simplified then we must not waste time trying to do so and be ready to use the tools available. Being able to tell these cases apart quickly saves the most time of all!

# Heteroscedasticity

There are many types of heteroscedasticity. In most undergraduate statistics programmes, students are taught to watch out for some of these, but seldom taught how to work with them. This is a serious shortcoming as understanding uncertainty is at the heart of what makes a good statistician. Furthermore, predictions from an unstable system are bound to lead to trouble.

## Systematic heteroscedasticity

The first case is where we have $Var\left(X\right)=f\left(E\left(X\right)\right)$ or something similar, when the change is variance is directly related to a change in something observed. This can often be addressed with a transformation.

What is usually a good transformation when $Var\left(Y\right)∝E\left(Y\right)$?

What is usually a good transformation when $StdDev\left(Y\right)∝E\left(Y\right)$?

These transformations are special cases of what general transformation?

## Specific heteroscedasticity

The opposite case is when each observation has different variance with no discernible pattern. This case can also present as excess kurtosis or extreme values in one or more directions.

One approach used to address such issues is to adapt the residual distribution, usually replacing a non-robust but powerful distribution (*e.g.* normal) with a more robust alternative (*e.g.* Student-t). The result is that outlying observations become less influential, which is useful when modelling central tendencies.

Another approach is switching from a parametric model to a non-parametric model. This is somewhat similar to the above in that we abandon the usual assumptions (*e.g.* normal) and use a more free specification for the residuals.

What are the downsides to changing the assumed residual distribution?

## Class based heteroscedasticity

This is where groups/classes/treatments/levels of observations have naturally different variation.

The standard approach to addressing this is to model the groups/classes/treatments/levels separately.

What does a separate regression for each group imply about the regression lines?

What is the problem with doing a separate regression for each group?

How do we get one regression line with different variances for each group? Give at least 3 approaches:

Can you see how these three approaches are similar to the questions at the start of the class?