STAT 530: Decision Theory Starting Points

We start with a *loss function*, $L(\theta, a)$, which reflects the loss (or negative utility) when *action* a is chosen and the *state of nature* is θ. Less grandiosely, we might have that θ is the value of the parameter being estimated and a is the estimate. In such a case, squared-error loss, $L(\theta, a) = (a - \theta)^2$, is one common choice.

Now think about a 'rule' (or estimator!) dictating what action to take based on what data y are observed, i.e., choose $a = \delta(y)$.

The performance of a given $\delta()$ is assessed via its *risk function*:

$$\begin{aligned} R(\theta, \delta) &= E_\theta\{L(\theta, \delta(Y))\} \\ &= \int L(\theta, \delta(y)) p(y|\theta) dy, \end{aligned}$$

reflecting, at each θ, average performance under repeated sampling. The mean-squared-error of an estimator is a prototypical risk function.

If two estimators $\delta_1()$ and $\delta_2()$ have risk functions that do not cross, then the situation is clear - we prefer the one with lower risk function. We might say that δ_1 beats δ_2 if $R(\theta, \delta_1) \leq R(\theta, \delta_2)$ for all θ, with strict inequality for some θ.

This leads to the following terminology:
A rule is *inadmissible* if it is beaten by some other rule.
A rule is *admissible* if it is not beaten by any other rule.

Very often we find ourselves comparing two estimators whose risk functions cross, i.e., neither beats the other. One suggestion is to then compare on the basis of the maximum value of the risk function (so called *minimaxity*). Another suggestion is to compare on the basis of some sort of average (across θ) of the risk function.

The *Bayes risk* of $\delta()$ with respect to the 'prior distribution' $\theta \sim \pi$ is

$$\begin{aligned} r(\pi, \delta) &= E_\pi R(\theta, \delta) \\ &= \int R(\theta, \delta) d\Pi(\theta). \end{aligned}$$

Note that the Bayes risk can be evaluated for any estimation scheme, not just a Bayesian estimation scheme. Note also that the emphasis is now on choosing a 'prior' π reflecting how we wish to weight θ values in determining average performance - not on choosing π to reflect pre-data belief.

Just to confuse things further, when we do perform Bayesian analysis we can talk about yet another form of risk: the *posterior risk* of choosing a particular action. That is

$$\begin{aligned}\rho_\pi(a; y) &= E\{L(\theta, a)|Y = y\} \\ &= \int L(\theta, a)\pi(\theta|y).\end{aligned}$$

The task is then to make connections between how Bayesian and non-Bayesian estimation procedures perform according to these sorts of criteria.