Engsci 760 Decision-Making and Modelling Under Uncertainty

Problem Set 1

Oliver Maclaren
oliver.maclaren@auckland.ac.nz

Lecture 1: Basic concepts

Problem 1.

Consider the decision table where the entries represent utilities (good things!)

	State of nature 1	State of nature 2
Decision 1	3	5
Decision 2	1	7

– What does the maximin utility (minimax loss) rule choose?

– Convert this to a regret table. What does the minimax regret rule choose?

Problem 2

Consider the decision table below which is an ordinal transformation (order-preserving transformation) of the previous utility table:

	State of nature 1	State of nature 2
Decision 1	7	11
Decision 2	3	16

– Carry out the same analysis as before (maximin utility, minimax regret) on this new table and the corresponding regret table. Are the decisions the same as before? Which are/are not?

Problem 3

Suppose the new table (relative to Problem 1) was instead

	State of nature 1	State of nature 2
Decision 1	7	11
Decision 2	3	15

– Show that this is given by a *positive linear transformation* (relative-length-preserving transformation) of the original table. Hint: this means the new utilities u' are related to the old utilities by $u' = au + b$ for some $a > 0$ and some b.

– Carry out the same analysis. Are the decisions the same as before? Which are/are not?

Problem 4

Show the following general results underlying the examples above:

– *Any ordinal* (order-preserving) transformation of a utility table doesn't change the resulting maximin utility decision but may change the minimax regret decision.

– *Any positive linear transformation* of a given utility table doesn't change the resulting maximin utility and minimax regret decisions.

Problem 5

Using the handout from Resnik in Lecture 1 as a guide, formulate a *decision tree* for the problem of choosing what to do on a given day described as:

> *Suppose you must first choose between going to the seashore or staying at home. If you go to the seashore, you will wait to determine whether it is raining. If it rains, you will decide whether to fish or stay inside. If you fish and the fishing is good, you will be happy; if the fishing is not good, you will be disappointed. If you stay in, you will feel so-so. On the other hand, if there is no rain, you will sunbathe and be happy. Finally if you stay home, you will feel so-so.*

– Check your answer against Resnik's Figure 1-2.

– Convert this to a table of 'one-time' decisions (strategies) as described by Resnik. Check your answer against his Table 1-6.

Lecture 2: Risk, probability, expected utility

Problem 6

Reconsider the decision table from Problem 1. Assume you now have 'uniform' probabilities over each state of nature and that these are independent of the decision made.

– What does the expected utility decision rule tell you to do?

Problem 7

Do the same expected utility analysis as in Problem 6 but now for the tables shown in Problems 2 and 3 (use the same uniform probabilities as in Problem 6). What patterns do you notice?

Problem 8

What (decision-independent) probabilities would you need to assign to the states of nature for Problem 6 for your decision to change?

Problem 9

Show that for a two-state-of-nature, two-decision problem, with probabilities over states of nature that are independent of decisions, the maximum expected utility decision is the same for a given utility function and any positive linear transformation of it.

Problem 10

Reduce the compound lottery $L(p, L(a), L(q, a, L(r, a, b)))$ with probabilities p, q, r and outcomes a, b down to a simple (non-compound) lottery $L(z, a, b)$ with new probability z. Draw a diagram or series of diagrams to illustrate the reduction process.

Lecture 3: Statistical decision theory (no data and large data limit cases)

Problem 11

Show that under squared loss $l(\theta, X)$ and known distribution over X the best single number summary θ is the population mean of X.

– What is the best summary of the distribution of a function $f(X)$ of X under squared loss $l(\theta, f(X))$. How does this relate to the previous case – do they give 'consistent' answers (hint: consider the cases of linear/nonlinear f)?

– Given a sample of size n what is the best empirical summary of X and $f(X)$ according to the 'empirical risk' approach?

– Roughly, when would you expect empirical risk to give a reasonable answer to the population risk decision problem?

Problem 12

Work through the example problems A, B, C in Lecture 3 (coin flip, no-data quadratic loss, maximum likelihood estimate).

Lecture 4: Statistical decision theory (finite data)

Problem 13

Show that under the squared error loss $l(\delta(X), \theta) = (\delta(X) - \theta)^2$ the risk function $R(\delta, \theta) = \mathbb{E}_{X;\theta}[l(\delta(X), \theta)]$ can be written

$$R(\delta, \theta) = \mathrm{Var}(\delta(X)) + (\mathrm{bias}_\theta(\delta(X)))^2$$

where $\mathrm{bias}_\theta(\delta(X)) = \bar{\delta} - \theta$ is a function of both the expected value of the estimator (here represented with an overbar, i.e. $\bar{\delta} = \mathbb{E}[\delta(X)]$ is the 'population' expectation) and the unknown parameter θ, while $\mathrm{Var}(\delta(X)) = \mathbb{E}[(\delta(X) - \bar{\delta})^2]$ is only a function of the estimator.

Problem 14

Using the previous result, explain the idea of a 'bias-variance trade-off' in general terms. (See next problem for a concrete illustration. See 'miscellaneous' problem further below for more intuition that might help).

Problem 15

Suppose we have $X_1, ..., X_n$ independent and identically distribution (IID) random variables (representing taking n independent samples) from a normal distribution, i.e. $X_i \sim N(\mu, \sigma^2)$ for $i = 1, ..., n$.

Define the *sample* mean as $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (note that here the bar is the sample rather than population mean, in contrast to the problem on deriving the risk function above). Consider two estimators of σ^2, the 'unbiased sample variance' S^2 and the 'maximum likelihood variance estimate' $\hat{\sigma}^2$, respectively:

$$\delta_1(X_1, X_2, ..., X_n) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\delta_2(X_1, X_2, ..., X_n) = \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

It can be shown that $\mathbb{E}[\delta_1] = \sigma^2$ and $\mathrm{Var}(\delta_1) = \frac{2\sigma^4}{n-1}$.

– Find the expected value and variance of the estimator δ_2. Hint – use the following properties of expectations and variances of an arbitrary random variable Y: $\mathbb{E}[aY] = a\mathbb{E}[Y]$ and $\mathrm{Var}(aY) = a^2Y$.

– Hence down the risk functions for each of the estimators, δ_1 and δ_2, in terms of the bias and variance terms. Which of these is unbiased and which biased?

– Which of these estimators is best from the perspective of mean-squared error? How does this relate to the bias-variance trade-off?

– What happens to the bias of the maximum likelihood estimator $\hat{\sigma}^2$ as n (sample size) goes to infinity? What happens to the variance? Hence what happens to the risk?

Problem 16

Suppose we are trying to estimate the mean of a normal distribution 'data model', $N(\mu_d, \sigma_d^2)$, where μ_d is unknown. Suppose as a good Bayesian you make up a convenient prior of the form $N(\mu_p, \sigma_p^2)$ where you (of course) get to choose the prior mean and variance.

Given a single observation of the data, i.e. arising as a realisation x_0 of the random variable $X \sim N(\mu_d, \sigma_d^2)$, it can be shown (see later in course or EngSci 721!) the posterior is a normal distribution with mean $\frac{\sigma_p^2}{\sigma_d^2+\sigma_p^2}x_0 + \frac{\sigma_d^2}{\sigma_d^2+\sigma_p^2}\mu_p$ and variance $\frac{\sigma_d^2\sigma_p^2}{\sigma_d^2+\sigma_p^2}$.

– What is the Bayes estimator under a squared error loss function?

– What is the Bayes estimator under an absolute error loss function, i.e. $l(\delta(X), \theta) = |\delta(X) - \theta|$?

– What is the Bayes estimator as the prior variance goes to infinity? How does this relate to the estimator considered in Lecture 4? What would you guess this 'limiting prior' is (technical note: it is 'improper', i.e. converges to something that no longer integrates to 1, but this is 'allowed' as long as the posterior is proper and/or if the corresponding estimator is well-defined).

Miscellaneous

Problem 17

Given the decomposition of the expected squared loss into variance and bias terms, consider the following cases to better understand the difference between the *best choice of a single number summary of a given (known) distribution* and the *best choice of an estimator of a single (unknown) number*.

Firstly, assume that we are given a known function $\delta(X)$ of X (i.e. you have no choice over this) and a known distribution for X, i.e. $X \sim \mathbb{P}(X)$ for some fixed and known distribution. We want to summarise the distribution of $\delta(X)$ by a single number (this sort of problem is considered in Lecture 3).

– Is this a case of risk or ignorance?

– Show that the expected loss formally 'looks like' the 'transpose' of the risk function for estimators δ above, i.e.

$$R(\theta, \delta) = \mathrm{Var}(\delta(X)) + (\mathrm{bias}_\theta(\delta(X)))^2$$

where here the expectations appearing in the bias and variance terms are with respect to the *known* distribution of X (which has no dependence on θ).

– What are the possible decisions here?

– Show that the best 'decision' is given by $\theta = \bar{\delta} = \mathbb{E}[\delta(X)]$. Hint: consider which part you control and which part is fixed.

Next suppose we are in the setting of statistical decision theory where the goal is to choose an estimator $\delta(X)$ for fixed but unknown θ.

– Explain why we don't necessarily want to choose an estimator such that $\bar{\delta} = \mathbb{E}[\delta(X)] = \theta$, i.e. why we don't necessarily want an unbiased estimator.