STAT 530: Underpinnings of Loss Function

Feb 10, 2009

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

Recall, interest in **minimizing expected loss**.

min_a
$$\int L(\theta, a)p(\theta|y)d\theta$$

min _{δ} $\int \int L(\theta, \delta(y))p(y|\theta) dy p(\theta) d\theta$
But what is the rationale for having/choosing a loss function in the
first place???

Aside: minimizing expected loss = maximizing expected utility

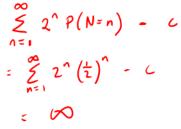
$$l\cdot g = U(\theta, a) = -L(\theta, a)$$

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

St. Petersberg porrdox

Pay \$c. Receive 2^N , where N is the number of independent flips of a fair coin required to obtain one tails.

Expected profit?



Pay any price to play!

What is your preference?

$A = \{$ \$1 million with probability 1

versus

$$B = \begin{cases} $5 \text{ million with probability } 0.10 \\ $1 \text{ million with probability } 0.89 \\ $0 & \text{with probability } 0.01 \end{cases}$$

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

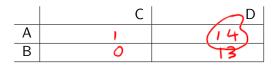
And what about

$$C = \begin{cases} \$1 \text{ million with probability 0.11} \\ \$0 & \text{with probability 0.89} \end{cases}$$

versus

$$D = \begin{cases} $5 \text{ million with probability 0.10} \\ $0 & \text{with probability 0.90} \end{cases}$$

Which cell are you in?



We'll come back to this.

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

Think of each option under consideration as a probability distribution over the set of possible rewards. Write $P_1 \prec P_2$ as preferring $R \sim P_2$ over $R \sim P_1$. Write $P_1 \approx P_2$ as ambivalent between $R \sim P_1$ and $R \sim P_2$ How would the preferences of someone rational behave?

Axiom 1: Rational person can always decide

"don't know" isn't allowed

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

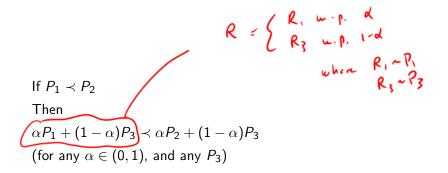
For any P_1 , P_2 , either $P_1 \prec P_2$ or $P_1 \approx P_2$ or $P_2 \prec P_1$

Axiom 2: Rational person is transitive

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

If $P_1 \prec P_2$ and $P_2 \prec P_3$ Then $P_1 \prec P_3$

Axiom 3: Rational person 'handles mixtures' sensibly



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ ○臣 - の々ぐ

Axiom 4: Rational person doesn't believe in infinitely good/bad

If $P_1 \prec P_2 \prec P_3$, then there are numbers α and β , both between 0 and 1, such that $\alpha P_1 + (1 - \alpha)P_3 \prec P_2 \leftarrow P_3$ not in finitely good and $P_2 \prec \beta P_1 + (1 - \beta)P_3 \leftarrow P_1$ not in finitely bad

Provided Axioms 1 through 4 hold, there exists a function u() mapping the set of possible rewards to the real line such that

$$P_1 \prec P_2 \leftrightarrow E_{P_1}\{u(R)\} < E_{P_2}\{u(R)\}$$
(Furthermore $\not D$ unique up to linear transformation)
So a rational person has a utility function, and always prefers the option with higher expected utility!

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

Did anybody say
$$B \prec A$$
 and $C \prec D$?
Say you did. (14)

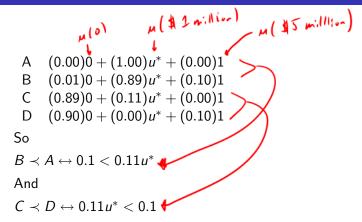
Without loss of generality, scale your utility function so that $\sqrt[4]{0}(\$ \ 0) = 0$ and $\sqrt[4]{1}(\$ \ 5 \ million) = 1$.

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

Let
$$u^* = \cancel{0}(\$ \ 1 \ \text{million})$$
.

What are your expected utilities for A,B,C,D?

Expected Utilities



Holding both preferences simultaneously irrational!

Can regard 'the prior' in a few different ways

- pure subjectivism: represent the investigator's beliefs about parameters prior to seeing the data
- pragmatic subjectivism: make pretty wide, but discount values that all agree are implausible (e.g. Odds-ratio=8)
- objectivism: flat, flat, flat!
- decision-theoretic stance: think of prior as choice about how to weight different parts of the parameter space when evaluating the performance (i.e., Bayes risk) of a procedure.

Also, often useful to think of strength of prior in intuitive terms, e.g., effective sample size

Performance evaluation:

 Can ask about frequentist performance of Bayesian procedure, i.e., what happens if repeatedly simulate Y given fixed θ

- \blacksquare compatibility of θ value and prior plays a role
- Can also aggregate performance across parameter space
 - e.g., Bayes risk
 - e.g., interval coverage w.r.t. joint (θ, Y) sampling

Ability to simulate arbitrarily large sample from $p(\theta|y)$ equates with 'knowing' $p(\theta|y)$.

burn-in, dependence may be nuisances here

Averaging across unobservables - not plugging in estimates.

- think of predicting the next data point
- also think of joint posterior on parameters and unobserved latent variables