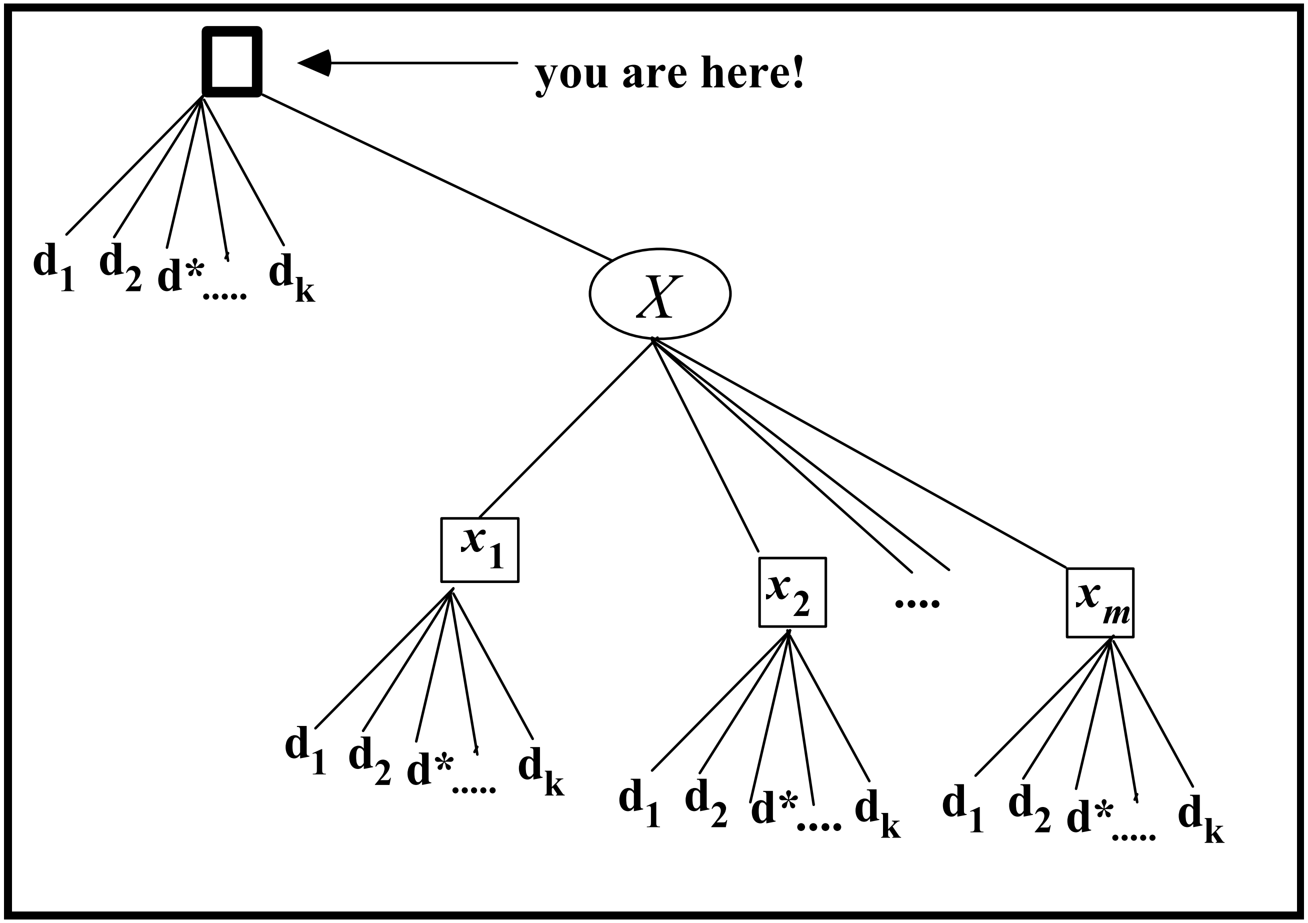
Sequential Decisions

A Basic Theorem of (Bayesian) Expected Utility Theory:

If you can postpone a terminal decision in order to observe, *cost free*, an experiment whose outcome might change your terminal decision, then it is strictly better to postpone the terminal decision in order to acquire the new evidence.

The analysis also provides a value for the new evidence, to answer:

How much are you willing to "pay" for the new information?



An agent faces a current decision:

- **with *k* terminal options $D = \{d_1, ..., d^*, ..., d_k\}$ (d^* is the best of these)**
- **and one sequential option: first conduct experiment *X*, with outcomes $\{x_1, ..., x_m\}$ that are observed, then choose from *D*.**

Terminal decisions (acts) as functions from states to outcomes

The canonical decision matrix: **decisions** × **states**

	s_1	s_2			s_j			s_n
d_1	O_{11}	O_{12}			O_{1j}			O_{1n}
d_2	O_{21}	O_{22}			O_{2j}			O_{2n}
d_m	O_{m1}	O_{m2}			O_{mj}			O_{mn}

$d_i(s_j)$ = outcome $\mathbf{o_{ij}}$.

What are "**outcomes**"?

That depends upon which version of expected utility you consider.

We will allow arbitrary outcomes, providing that they admit a von Neumann-Morgenstern cardinal utility $U(\bullet)$.

A central theme of Subjective Expected Utility [*SEU*] is this:

- axiomatize preference $\leq$ over decisions so that

$$d_1 \leq d_2 \quad \textbf{iff} \quad \Sigma_j\, \mathbf{P}(s_j)\mathbf{U}(o_{1j}) \;\leq\; \Sigma_j\, \mathbf{P}(s_j)\mathbf{U}(o_{2j}),$$

for **one** subjective (personal) probability $\mathbf{P}(\bullet)$ defined over *states*
and **one** cardinal utility $\mathbf{U}(\bullet)$ defined over *outcomes*.

- Then the decision rule is to choose that (an) option that *maximizes SEU*.

Note: In this version of SEU, which is the one that we will use here:

(1) decisions and states are probabilistically independent, $\mathbf{P}(s_j) = \mathbf{P}(s_j \mid d_i)$.

Aside: This is necessary for a fully general *dominance* principle. That is, assume (simple) *Dominance:* $d_1 < d_2$ **if** $\mathbf{U}(o_{1j}) < \mathbf{U}(o_{2j})$ ($j = 1, \ldots, n$).

Note well that if $\mathbf{P}(s_j) \neq \mathbf{P}(s_j \mid d_i)$, then *dominance* may fail.

(**2**) Utility is state-independent, $\mathbf{U_j}(o_{i,j}) = \mathbf{U_h}(o_{g,h})$, if $o_{i,j} = o_{g,h}$.

Here, $\mathbf{U_j}(o_{\bullet j})$ is the conditional utility for outcomes, given state s_j.

(**3**) (Cardinal) Utility is defined up to positive linear transformations, $\mathbf{U}'(\bullet) = a\mathbf{U}(\bullet) + b$ (a > 0) is also the same utility function for purposes of *SEU*.

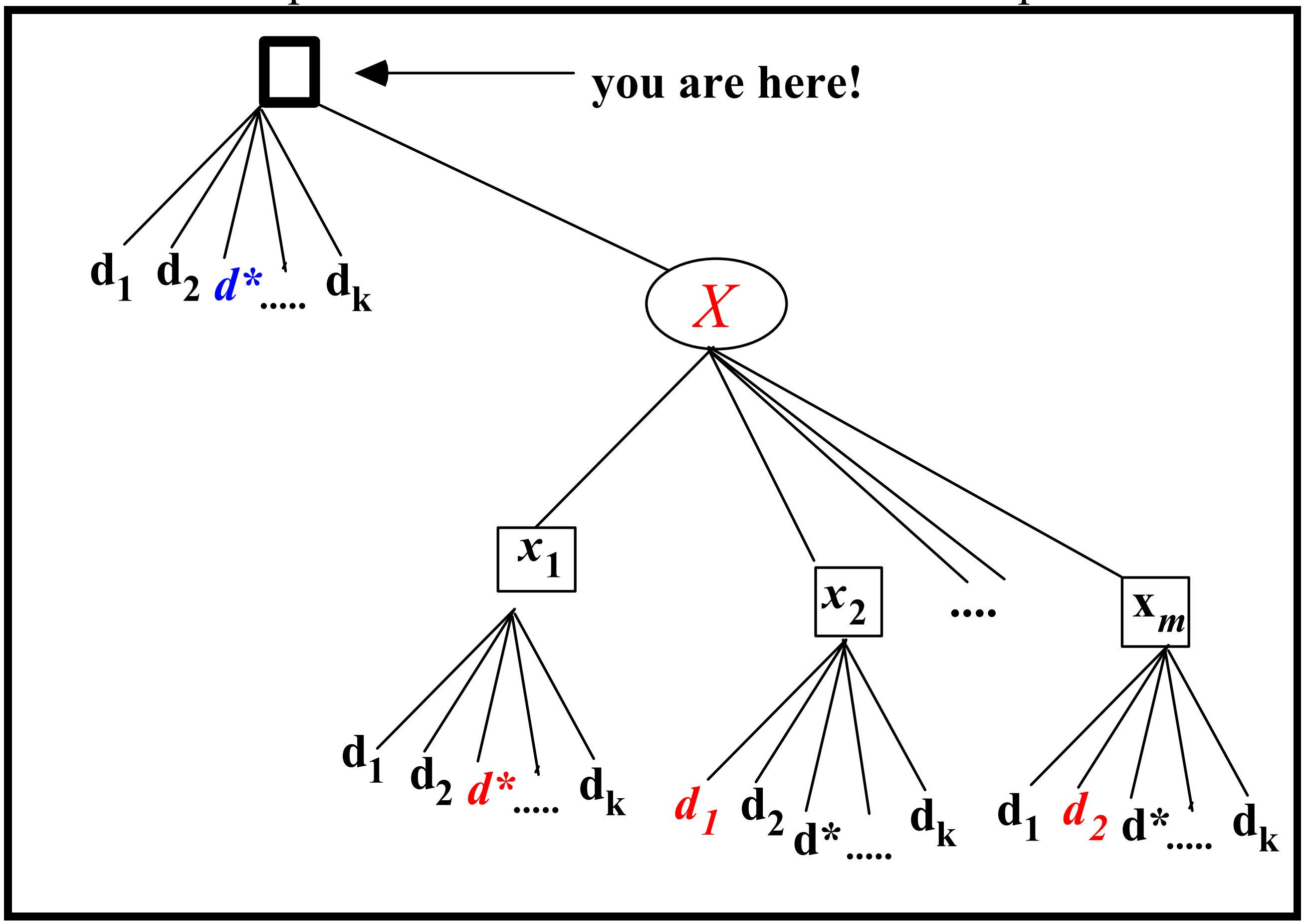
Note: More accurately, under these circumstances with act/state prob. independence, utility is defined up to a similarity transformation: $\mathbf{U_j}'(\bullet) = a\mathbf{U_j}(\bullet) + b_{\mathbf{j}}$.

Defn: The decision problem is said to be in *regret form* when the $b_{\mathbf{j}}$ are chosen so that, for each state s_j, $\max_{\boldsymbol{D}} \mathbf{U}_{\mathbf{j}}'(o_{ij}) = 0$.

Then, all utility is measured as a "loss," with respect to the best that can be obtained in a given state.

Example: *squared error* $(t(X) - \theta)^2$ used as a loss function to assess a point estimate $t(X)$ of a parameter θ is a decision problem in regret form.

Reconsider the value of new, cost-free evidence when decisions conform to *SEU*. Recall, the decision maker faces a choice *now* between *k*-many terminal options ***D*** = {d_1, ..., ***d****, ..., d_k} (***d**** maximizes SEU among these k options) and there is one sequential option: first conduct experiment *X*, with sample space {x_1, ..., x_m}, and then choose from *D*. Options in *red* maximize SEU at the respective choice nodes.



By the law of conditional expectations: $E(Y) = E\,(E\,[Y \mid X]\,)$.

With Y **the Utility of an option** $U(d)$, **and** X **the outcome of the experiment,**

$$\text{Max}_{d \in D}\; E(U(d)) = E\,(U(d^*))$$

$$= E\,(E\,(U(d^*) \mid X))$$

$$\leq E\,(\text{Max}_{\,d \in D}\; E(U(d) \mid X))$$

$$= U(\text{sequential option}).$$

- **Hence, the academician's** ***first-principle*****:**
  Never decide today what you might postpone until tomorrow
  in order to learn something new.

- $E(U(d^*)) = U$**(sequential option) if and only if the new evidence** Y **never leads you to a different terminal option.**

- U**(sequential option) -** $E\,(U(d^*))$ **is the** ***value of the experiment*****: what you will pay (at most) in order to conduct the experiment prior to making a terminal decision.**

***Example*: Choosing sample size, fixed versus adaptive sampling** **(DeGroot, chpt. 12)**

The statistical problem has a terminal choice between two options, $D = \{ d_1, d_2\}$.

There are two states $S = \{s_1, s_2\}$, with outcomes that form a regret matrix:

$U(d_1(s_1)) = U(d_2(s_2)) = 0$, $U(d_1(s_2)) = U(d_2(s_1)) = -b < 0$.

	s_1	s_2
d_1	0	$-b$
d_2	$-b$	0

Obviously, according to SEU, $d^* = d_i$ if and only if $P(s_i) \geq .5$ (i = 1, 2).

Assume, for simplicity that $P(s_1) = p < .5$, so that $d^* = d_2$ with $E(U(d_2)) = -pb$.

***The sequential option*: There is the possibility of observing a random variable $X = \{1, 2, 3\}$. The statistical model for X is given by:**

$$P(X = 1 \mid s_1) = P(X = 2 \mid s_2) = 1 - \alpha.$$

$$P(X = 1 \mid s_2) = P(X = 2 \mid s_1) = 0.$$

$$P(X = 3 \mid s_1) = P(X = 3 \mid s_2) = \alpha.$$

Thus, $X = 1$ or $X = 2$ identifies the state, which outcome has conditional probability 1-α on a given trial; whereas $X = 3$ is an irrelevant datum, which occurs with (unconditional) probability α.

Assume that X may be observed repeatedly, at a cost of c-units per observation, where repeated observations are conditionally *iid*, given the state s.

- ***First*, we determine what is the optimal fixed sample-size design, $N = n^*$.**
- ***Second*, we show that a sequential (adaptive) design is better than the best fixed sample design, by limiting ourselves to samples no larger than n^*.**
- ***Third*, we solve for the global, optimal sequential design as follows:**
    - **We use Bellman's principle to determine the optimal sequential design bounded by $N \leq k$ trials.**
    - **By letting $k \to \infty$, we solve for the global optimal sequential design in this decision problem.**

- ***The best, fixed sample design*.**

Assume that we have taken $n > 0$ observations: $\tilde{X} = (x_1, \ldots, x_n)$

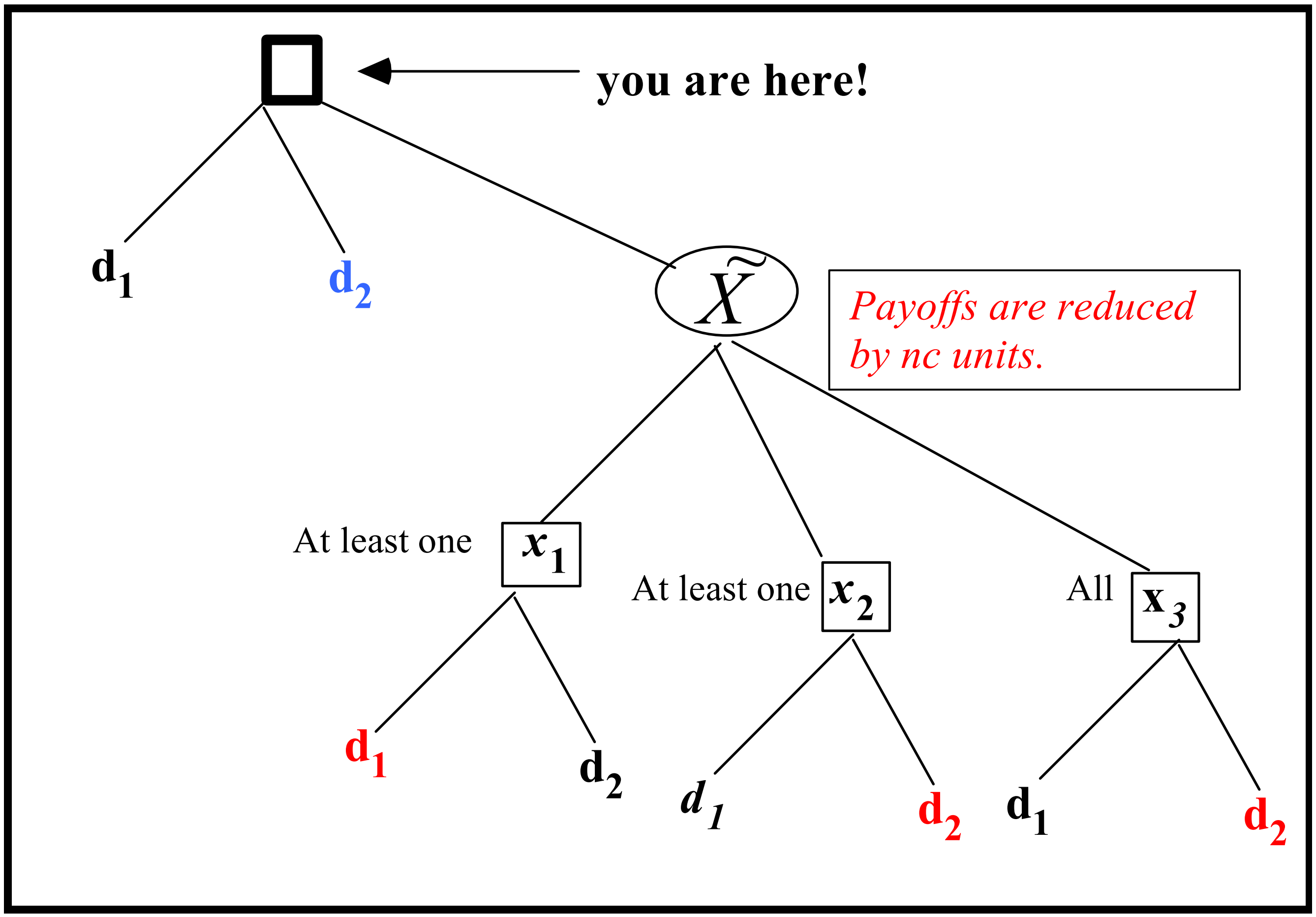
The posterior prob., $P(s_1 \mid \tilde{X}) = 1$ ($P(s_2 \mid \tilde{X}) = 1$ $x_i = 2$) if $x_i = 1$ for some $i = 1, \ldots, n$. Then, the terminal choice is made at no loss, but nc units are paid out for the experimental observation costs.

Otherwise, $P(s_1 \mid \tilde{X}) = P(s_1) = p$, when all the $x_i = 3$ ($i = 1, \ldots, n$), which occurs with probability α^n. Then, the terminal choice is the same as would be made with no observations, d_2, having the same expected loss, $-pb$, but with nc units paid out for the experimental observation costs.

That is, the pre-trial (SEU) value of the sequential option to sample n-times and then make a terminal decision is:

$$E(\text{sample } n \text{ times before deciding}) = -[pb\alpha^n + cn].$$

Assume that c is sufficiently small (relative to $(1-\alpha)$, p and b) to make it worth sampling at least once, i.e. $-pb < -[pb\alpha + c]$, **or** $c < (1-\alpha)pb$



Thus, with the pre-trial value of the sequential option to sample *n*-times and then make a terminal decision:

$$E(\textbf{sample } n \textbf{ times before deciding}) = -[pb\alpha^{n} + cn].$$

- **then the *optimal fixed sample size design* is, approximately (obtained by treating *n* as a continuous quantity):**

$$n^{*} = \frac{-\log[pb\log(1/\alpha)/c]}{1/\log(1/\alpha)}$$

- **and the *SEU* of the optimal fixed-sample design is approximately**

$$E(\textbf{sample } n^{*} \textbf{ times then decide}) = -(c/\log(1/\alpha))\,[1 + \log[pb\log(1/\alpha)/c]\,]$$

$$> -pb = E(\textbf{decide without experimenting})$$

- **Next, consider the plan for bounded sequential stopping, where we have the option to stop the experiment after each trial, up to *n** many trials.**

At each stage, *n*, prior to the $n^{*\text{th}}$, evidently, it matters for stopping only whether or not we have already observed $X = 1$ or $X = 2$.

- **For if we have then we surely stop: there is no value in future observations.**
- **If we have not, then it pays to take at least one more observation, if we may (if $n < n^*$), since we have assumed that** $c < (1-\alpha)pb$.

If we stop after *n*-trials ($n < n^*$), having seen $X = 1$, or $X = 2$, our loss is solely the cost of the observations taken, *nc*, as the terminal decision incurs no loss.

Then, the expected number of observations *N* from bounded sequential stopping (which follows a *truncated negative binomial* **distn) is:**

$$E(N) = (1-\alpha^{n^*})/(1-\alpha) < n^*.$$

Thus, the *S*ubjective *E*xpected *U*tility of (bounded) sequential stopping is:

$$-[pb\alpha^{n^*} + cE(N)] > -[pb\alpha^{n^*} + cn^*].$$

- **What of the unconstrained sequential stopping problem?**

With the terminal decision problem $D = \{ d_1, d_2\}$, what is the global, optimal experimental design for observing X subject to the constant cost, c-units/trial and the assumption that $c < (1-\alpha)pb$?

Using the analysis of the previous case, we see that if the sequential decision is for bounded, optimal stopping, with $N \leq k$, the optimal stopping rule is to continue sampling until either $X_i \neq 3$, or $N = k$, which happens first. Then, we see that

$E_{N \leq k}(N) = (1-\alpha^k)/(1-\alpha)$ and the SEU of this stopping rule is $-[pb\alpha^k + c(1-\alpha^k)/(1-\alpha)]$,

which is monotone increasing in k.

Thus the global, optimal stopping rule is the unbounded rule: continue with experimentation until $X = 1$ or $= 2$, which happens with probability 1.

$E(N) = 1/(1-\alpha)$ and the SEU of this stopping rule is $-[c/(1-\alpha)]$.

Note: Actual costs here are unbounded!

The previous example illustrates a basic technique for finding a global optimal sequential decision rule:

1) Find the optimal, *bounded* decision rule $\boldsymbol{d}^*_{\boldsymbol{k}}$ when stopping is mandatory at $\boldsymbol{N} = \text{k}$.

In principle, this can be achieved by *backward induction*, by considering what is an optimal terminal choice at each point when $N = \text{k}$, and then using that result to determine whether or not to continue from each point at $N = \text{k-1}$, etc.

2) Determine whether the sequence of optimal, bounded decision rules converge as $\text{k}\to\infty$, to the rule $\boldsymbol{d}^*_\infty$.

3) Verify that $\boldsymbol{d}^*_\infty$ is a global optimum.

Let us illustrate this idea in an elementary setting: the *Monotone* case (Chow *et al*, chpt. 3.5)

- Denote by $\boldsymbol{Y_{d,n}}$ the expected utility of the terminal decision $\boldsymbol{d}$ (inclusive of all costs) at stage $\boldsymbol{n}$ in the sequential problem.
- Denote by $\tilde{\boldsymbol{X}}_{\boldsymbol{n}} = (\boldsymbol{X}_{\mathbf{1}}, \ldots, \boldsymbol{X}_{\mathbf{n}})$, the data available upon proceeding to the n*th* stage.
- Denote by $\boldsymbol{A_n} = \{\tilde{\boldsymbol{x}}_{\boldsymbol{n}} : \mathrm{E}[\boldsymbol{Y_{d,n+1}} | \tilde{\boldsymbol{x}}_{\boldsymbol{n}}] \leq \mathrm{E}[\boldsymbol{Y_{d,n}} | \tilde{\boldsymbol{x}}_{\boldsymbol{n}}]\}$, the set of data points $\tilde{\boldsymbol{X}}_{\boldsymbol{n}}$ where it does *not* pay to continue the sequential decision *one* more trial, from *n* to *n+1* observations, before making a terminal decision.

Define the *Monotone Case* where: $\boldsymbol{A_1} \subset \boldsymbol{A_2} \subset \ldots,$ and $\cup_i \boldsymbol{A_i} = \mathbf{\Omega}$.

Thus, in the monotone case, once we enter the $\boldsymbol{A_i}$-sequence, our expectations never go up from our current expectations.

- **An *intuitive rule* for the monotone case is δ*: Stop collecting data and make a terminal decision the first time you enter the $\boldsymbol{A_i}$-sequence.**

- **An experimentation plan δ is a *stopping rule* if it halts, almost surely.**
- **Denote by $y^- = -\min\{y, 0\}$; and $y^+ = \max\{y, 0\}$.**
- **Say that the *loss is essentially bounded* under stopping rule δ if $E_\delta[Y^-] < \infty$, the *gain is essentially bounded* if $E_\delta[Y^+] < \infty$, and for short say that δ is *essentially bounded in value* if both hold.**

***Theorem*: In the *Monotone Case*, if the intuitive stopping rule δ is essentially bounded, and if its conditional expected utility prior to stopping is also bounded, i.e.,**

$$\text{if } \liminf_n E_\delta[\, Y_{\delta, n+1} \mid \delta(\tilde{x}_n) \text{ is to continue sampling}] < \infty$$

then δ is best among all stopping rules that are essentially bounded.

***Example*: Our sequential decision problem, above, is covered by this result about the Monotone Case.**

Counter-example 1: Double-or-nothing with incentive.

Let $\tilde{X} = (X_1, \ldots, X_n, \ldots)$ be *iid* flips of a *fair* coin, outcomes {-1, 1} for {H, T}:

$$P(X_i = 1) = P(X_i = -1) = .5$$

Upon stopping after the n*th* toss, the reward to the decision maker is

$$Y_n = [2n/(n+1)] \prod_{i=1}^{n} (X_i + 1).$$

In this problem, the decision maker has only to decide when to stop, at which point the reward is Y_n: there are no other terminal decisions to make.

Note that for the fixed sample size rule, halt after *n* flips: $E_{d=n}[Y_n] = 2n/(n+1)$.

However, $E[Y_{d=n+1} | \tilde{x}_n] = [(n+1)^2/n(n+2)]\, y_n \geq y_n$.

Moreover, $E[Y_{d=n+1} | \tilde{x}_n] \leq y_n$ **if and only if** $y_n = 0$,

In which case $E[Y_{d=n+2} | \tilde{x}_{n+1}] \leq y_{n+1} = 0$,

- **Thus, we are in the Monotone Case.**

Alas, the *intuitive rule* **for the monotone case, δ*, here means halting at the first outcome of a "tail" (x_n = -1), with a sure reward $Y_{\delta *}$ = 0, which is the worst possible strategy of all! This is a *proper* stopping rule since a tail occurs, eventually, with probability 1.**

This stopping problem has NO (global) optimal solutions, since the value of the fixed sample size rules have a *l.u.b.* of 2 = $lim_{n\to\infty}$ 2n/(n+1), which cannot be achieved.

When stopping is mandatory at *N* = k, the optimal, *bounded* decision rule,

$$d_k^* = \text{flip k-times,}$$

agrees with the payoff of the truncated version of the intuitive rule:

$$\delta_k^* \text{ flip until a tail, or stop after the } k^{th} \text{ flip.}$$

But here the value of limiting (intuitive) rule, *SEU*(δ*) = 0, is not the limit of the values of the optimal, bounded rules, 2 = *lim* $_{n\to\infty}$ 2n/(n+1).

Counter example 2: For the same fair-coin data, as in the previous example, let

$$Y_n = min[1, \sum_{i=1}^{n} X_i] - (n/n+1).$$

Then $$\mathrm{E}[Y_{d=n+1} | \tilde{x}_n] \leq y_n \quad \text{for all } n = 1, 2, \ldots .$$

Thus, the Monotone Case applies trivially, i.e., δ* = stop after 1 flip.

Then $$SEU(\delta^*) = -1/2 \quad (= .5(-1.5) + .5(0.5)).$$

However, by results familiar from simple random walk,

$$\text{with probability 1, } \sum_{i=1}^{n} X_i = 1, \text{ eventually.}$$

Let *d* be the stopping rule: halt the first time $\sum_{i=1}^{n} X_i = 1$.

$$\text{Thus, } 0 < SEU(d).$$

Here, the Monotone Case does not satisfy the requirements of being essentially bounded for *d*.

Remark: Nonetheless, *d* is globally optimal!

Example: The Sequential Probability Ratio Tests, Wald's *SPRT* (Berger, chpt. 7.5)

Let $\widetilde{X}$ **= (**X_1**, …,** X_n**, …) be *iid* samples from one of two unknown distributions,**

H_0**:** $f = f_0$ **or** H_1**:** $f = f_1$**. The terminal decision is binary: either** d_o ***accept*** H_0

or d_1 ***accept*** H_1**, and the problem is in regret form with losses:**

	H_0	H_1
d_o	0	$-b$
d_1	$-a$	0

The sequential decision problem allows repeated sampling of X, subject to a constant *cost per observation* of, say, 1 unit each.
A sequential decision rule $\delta = (d, s)$**, specifies a stopping size** S**, and a terminal decision** d**, based on the observed data.**

The conditional expected loss for δ $= a\alpha_0 + E_0[S]$**, given** H_0

$= b\alpha_1 + E_1[S]$**, given** H_1

where α_0 **= is the probability of a type 1 error (falsely accepting** H_1**)**

and where α_1 **= is the probability of a type 2 error (falsely accepting** H_0**).**

For a given stopping rule, *s*, it is easy to give the Bayes decision rule

accept H_1 **if and only if** $P(H_0|\tilde{X}_s)a \leq (P(H_1|\tilde{X}_s))b$

and *accept* H_0 **if and only if** $P(H_0|\tilde{X}_s)a > (P(H_1|\tilde{X}_s))b$.

Thus, at any stage in the sequential decision, it pays to take at least one more observation if and only if the expected value of the new data (discounted by a unit's cost for looking) exceeds the expected value of the current, best terminal option. By the techniques sketched here (*backward induction for the truncated problem, plus taking limits*), the global optimal decision has a simple rule:

- **stop if the posterior probability for** H_0 **is sufficiently high:** $P(H_0|\tilde{X}) \geq c_0$
- **stop if the posterior probability for** H_1 **is sufficiently high:** $P(H_0|\tilde{X}) \leq c_1$
- **and continue sampling otherwise, if** $c_1 < P(H_0|\tilde{X}) < c_0$.

Since these are *iid* data, the optimal rule can be easily reformulated in terms of cutoffs for the likelihood ratio $P(\tilde{X}|H_0) / P(\tilde{X}|H_1)$: **Wald's *SPRT***.

A final remark – based on Wald's 1940s analysis. (See, e.g. Berger, chpt 4.8.):

- **A decision rule is admissible if it is not weakly dominated by the partition of the parameter values, i.e. if its risk function is not weakly dominated by another decision rule.**

- **In decision problems when the loss function is (closed and) bounded and the parameter space is finite, the class of Bayes solutions is *complete*: it includes all admissible decision rules. That is, non-Bayes rules are *inadmissible*.**

Aside: For the infinite case, the matter is more complicated and, under some useful conditions a complete class is given by Bayes and limits of Bayes solutions – the latter relating to "improper" priors!

Additional References

Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2^{nd} ed. Springer-Verlag: NY.

Chow, Y., Robbins, H., and Siegmund, D. (1971) *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin: Boston.

DeGroot, M. (1970) *Optimal Statitical Decisions*. McGraw-Hill: New York.