Assignment for **Mathematics for Economists**
Fall 2016

Due date: Mon. Nov. 21.

Readings: CSZ, Ch. 5, Ch. 8.12

We now turn to finishing our coverage of concavity/convexity. There are two parts: Jensen's inequality for concave/convex functions and its implications in expected utility theory and the value of information; constrained optimization. Below you will find a short set of notes on the value of information (in the finite setting where things are easiest).

As to constrained maximization: CSZ §5.8 gives more practice with Lagrangean multipliers; CSZ §5.9 fixes firmly in your mind that the multipliers are the derivatives of the value function. For twice continuously differentiable functions, CSZ §5.10 takes you through a proof of the "alternating signs of principal minors of the Jacobean implies concavity" result. We will cover it in lecture with a quick review of the necessary linear algebra. In the Spring semester, you will be taking Econometrics I, remembering, or learning, or re-learning linear algebra should happen before then.

A Very Short Introduction to the Value of Information

There is a joint distribution pdf, $q(\cdot, \cdot)$, for a random signal, S and a random state X, that is, $q(x, s) = Prob(X = x, S = s)$, everything taking values in finite spaces, $\mathcal{X}$ for X and $\mathcal{S}$ for S. The time line is: $S = s$ is observed for some $s \in \mathcal{S}$, but the value of X is **not** observed, perhaps because it hasn't happened yet; the decision maker picks an action $a \in A$, since it can be, and usually is, a function of s, this is written $a(s)$; the value of $X = x$ is observed; von Neumann-Morgenstern utility of $u(a, x)$ is received. The expected utility maximization problem is to pick a function from observations to actions, $s \mapsto a(s)$, so as to maximize

$$(\ddagger) \quad E\, u(a(S), X) = \sum_{x,s} u(a(s), x) q(x, s). \tag{1}$$

One can solve this problem by formulating a complete contingent plan, $s \mapsto a(s)$, or one can "cross that bridge when one comes to it," that is, wait until $S = s$ has been observed and figure out at that point what $a(s)$ should be.

Let $\pi(s) = \sum_x q(x, s)$, rewrite $E\, u = \sum_{x,s} u(a(s), x)q(x, s)$ as

$$\textstyle\sum_s \pi(s) \sum_x u(a(s), x) \frac{q(x,s)}{\pi(s)}. \tag{2}$$

$\beta(x|s) := \frac{q(x,s)}{\pi(s)}$ is the posterior probability that $X = x$ after $S = s$ has been observed. As is well-known and easy to verify, if $a^*(s)$ solves $\max_a \sum_x u(a, x)\, \beta(x|s)$ at each s with $\pi(s) > 0$, then $a^*(\cdot)$ solves the original maximization problem in ($\ddagger$). In other words, the dynamically consistent optimality of best responding to beliefs formed by Bayesian updating arises from the linearity of expected utility in probabilities.

The pieces of the basic statistical decision model are:

- $\beta_0(x) := \sum_s q(x,s)$, the marginal of q on X, known as the **prior** distribution — as in the probability distribution that the decision maker has prior to observing and incorporating any information contained in the signal;
- $\pi(s) := \sum_x q(x,s)$, the marginal of q on S, the **distribution of the signals**; and
- $\beta(\cdot|s) = \frac{q(\cdot,s)}{\pi(s)}$, the **posterior** distributions, i.e. the conditional distribution of X given that $S = s$.

The martingale property of beliefs is the observation that the prior is the π-weighted convex combination of the posteriors, for all x,

$$E\left(\beta(x|S)\right) = \sum_s \pi(s)\beta(x|s) = \sum_s \pi(s)\frac{q(x,s)}{\pi(s)} = \sum_s q(x,s) = \beta_0(x). \tag{3}$$

To put this more directly, the average of the posteriors is the prior, or, thinking of B as a random vector of beliefs and S as the random signal, this is repeating the law of iterated expectations, $E\left(E\left(B|S\right)\right) = E\,B$.

Beliefs at s, $\beta(\cdot|s)$ belong to $\Delta(\mathcal{X})$ and have distribution π. Re-writing,

> Information is a distribution, $\pi \in \Delta(\Delta(\mathcal{X}))$ having mean equal to the prior.

Blackwell (1950, 1951) showed that all signal structures are equivalent to such distributions.

Now, the value of the information, π, is the increase in expected utility that comes from receiving signals and acting optimally in response. To get at this, for $\beta \in \Delta(\mathcal{X})$, define

$$V(\beta) = \max_{a\in A} \textstyle\sum_x u(a,x)\,\beta(x). \tag{4}$$

Being the maximimum of affine functions, $V(\cdot)$ is convex, this because its epigraph, $\{(\beta,r) : V(\beta) \geq r\}$ is, by the definition, equal to the intersection of a set of closed half-planes.

Returning to "the increase in expected utility that comes from receiving signals and acting optimally in response," if there is no information, the expected utility is $V(\beta_0)$, so the value of π is

$$\textstyle\int_{\Delta(\mathcal{X})} V(\beta)\,d\pi(\beta) - V(\beta_0). \tag{5}$$

An easy observation: if information never leads to a change in the action picked at β_0, then the value of π is 0.

Problems

A. Prove the "easy observation" just given. [If you start working very hard on this problem, then you are doing it the wrong way.]

B. **From Chapter 8.12**: 8.12.3 and 8.12.4.

C. [Infrastructure investment] $X = G$ or $X = B$ corresponds to the future weather pattern, the actions are to Leave the infrastructure alone or to put in New infrastructure, and the signal, s, is the result of investigations and research into the distribution of future values of X. The utilities $u(a,x)$ are given in the following table where c is the cost of the new infrastructure.

	Good	Bad
L	10	6
N	$(10-c)$	$(9-c)$

1. After you see a signal $S = s$, you form your posterior beliefs $\beta(G|s)$ and $\beta(B|s)$. For what values of $\beta(G|s)$ do you optimally choose to leave the infrastructure as it is? Express this inequality as "L if $c > M$," give the value of M in terms of $\beta(G|s)$, and interpret. [You should have something like "leave the old infrastructure as it is if costs are larger than expected gains."]
2. If $c = 0.60$, that is, if the new infrastructure costs 20% of the damages it prevents, give the set of $\beta(G|s)$ for which it optimal to leave the old infrastructure in place. **For the rest of this problem, assume that $c = 0.60$.**
3. Suppose that the original or prior distribution has $\beta(G) = 0.75$ so that, without any extra information, one would put in the New infrastructure. We now introduce some signal structures. Suppose that we can run test/experiments that yield $S = s_G$ or $S = s_B$ with $P(S = s_G|G) = \alpha \geq \frac{1}{2}$ and $P(S = s_B|B) = \gamma \geq \frac{1}{2}$. The joint distribution, $q(\cdot,\cdot)$, is

	Good	Bad
s_G	$\alpha \cdot 0.75$	$(1-\gamma)\cdot 0.25$
s_B	$(1-\alpha)\cdot 0.75$	$\gamma \cdot 0.25$

Give $\beta(\cdot|S = s_G)$ and $\beta(\cdot|S = s_B)$. Verify that the average of the posterior beliefs is the prior, that is, verify that $\sum_s \pi(s)\beta(\cdot|x) = \beta(\cdot)$.
4. Show that if $\alpha = \gamma = \frac{1}{2}$, then the signal structure is worthless.
5. Give the set of $(\alpha, \gamma) \geq (\frac{1}{2}, \frac{1}{2})$ for which the information structure strictly increases the expected utility of the decision maker. [You should find that what matters for increasing utility is having a positive probability of changing the decision.]

D. [The value of repeated independent observations] This problem continues where the previous problem left off. Now suppose that the test/experiment can be run twice and that the results are independent across the trials. Thus, $P(S = (s_G, s_G)|G) = \alpha^2$, $P(S = (s_G, s_B)|G) = P(S = (s_B, s_G)|G) = \alpha(1-\alpha)$, and $P(S = (s_B, s_B)|G) = (1-\alpha)^2$ with the parallel pattern for B.

1. Fill in the probabilities in the following joint distribution $q(\cdot,\cdot)$. Verify that the average of posterior beliefs is the prior belief.

	Good	Bad
(s_G, s_G)		
(s_G, s_B)		
(s_B, s_G)		
(s_B, s_B)		

2. Give $\beta(G|(s_G, s_G))$, $\beta(G|(s_G, s_B))$, $\beta(G|(s_B, s_G))$, and $\beta(G|(s_B, s_B))$.
3. Show that if $\alpha = \beta = \frac{1}{2}$, then the signal structure is worthless.
4. Give the set of $(\alpha, \beta) \geq (\frac{1}{2}, \frac{1}{2})$ for which the information structure strictly increases the expected utility of the decision maker.

5. Explain why the set is larger here than it was in the previous problem.

E. [A more theoretical problem] We say that signal structure A is unambiguously better than signal structure B if every decision maker, no matter what their action set and what their utility functions are, would at least weakly prefer A to B. This problem asks you to examine some parts of this definition.

Let us suppose that the random variable takes on two possible values, x and x', with probabilities $p > 0$ and $p' = (1 - p) > 0$, that there are two possible actions, a and b, that the signals take on two values s and s', that the signal structure is given by

	x	x'
s	αp	$(1-\gamma)p'$
s'	$(1-\alpha)p$	$\gamma p'$

and that the utilities are given by

	x	x'
a	$u(a, x)$	$u(a, x')$
b	$u(b, x)$	$u(b, x')$

We assume that $\alpha > \frac{1}{2}$ and $\gamma > \frac{1}{2}$ and we let $V(\alpha, \gamma)$ denote the maximal expected utility when the signals are distributed as above and the utilities are above.

1. Show that $V(\alpha, \gamma) \geq V(\frac{1}{2}, \frac{1}{2})$, that is, the information structure with $\alpha > \frac{1}{2}$ and $\gamma > \frac{1}{2}$ is unambiguously better than the structure with $\alpha = \gamma = \frac{1}{2}$.
2. Give utilities such that $V(\alpha, \gamma) > V(\frac{1}{2}, \frac{1}{2})$, that is, show that some decision maker strictly prefers the information structure with $\alpha > \frac{1}{2}$ and $\gamma > \frac{1}{2}$ to the structure with $\alpha = \gamma = \frac{1}{2}$.
3. Show that for any utilities, if $\alpha' > \alpha \geq \frac{1}{2}$ and $\gamma' > \gamma \geq \frac{1}{2}$, then $V(\alpha', \gamma') \geq V(\alpha, \gamma)$.
4. Give utilities such that $V(\alpha', \gamma') \geq V(\alpha, \gamma)$ for $\alpha' > \alpha \geq \frac{1}{2}$ and $\gamma' > \gamma \geq \frac{1}{2}$.

F. [Another theoretical problem] A way to make information less valuable is to **scramble** it. Let T be a random signal that has distribution $P(\cdot|S = s)$ and suppose that, conditional on $S = s$, T is independent of X. (Note that this rules out the more valuable case of repeated observations above.) Show that the signal structure T is (weakly) less valuable than the signal structure S. [Do not start this problem until you have found a good definition of conditional independence.]

G. The Joint Stock Act of 1844 (in Britain) allowed for the formation of joint-stock companies that were, for the first time, *not* government granted monopolies. The Limited Liability Act of 1855 meant that one could lose up to and including one's entire investment in a firm, but no more. It is not surprising that this led to more investment. Let X be a random variable taking both negative and positive values and having $E\,X = (1 + r)$ for $r > 0$ being the expected rate of return, and let $Y = \max\{0, X\}$.

1. Show that aY first order stochastically dominates aX for all $a \geq 0$.

2. Show that it is possible to have a random variable X' be riskier than X and have $Y' := \max\{0, X'\}$ first order stochastically dominate $Y = \max\{0, X\}$.
3. Consider the two problems

$\max_{a \geq 0}[u(w_0 - a) + \rho E\, u(w_1 + aX)]$ and $\max_{a \geq 0}[u(w_0 - a) + \rho E\, u(w_1 + aY)]$

where $0 < \rho < 1$. Give general conditions under which the solution to the second problem is larger than the solution to the first problem, and prove that your conditions deliver this conclusion.

H. The decision maker has wealth w_0 at present and next period's wealth is a random variable W_1. Savings can be put into a risk-free asset with rate of return r. Consider the problem

$$\max_{a \geq 0}[u(w_0 - a) + \rho E\, u(W_1 + a(1 + r))]$$

where $0 < \rho < 1$.

1. Give sufficient conditions for a^* to be larger when W_1 becomes riskier and prove that your conditions deliver this conclusion.
2. Now suppose that one can also invest $b \geq 0$ in a risky asset with returns $X \geq 0$ per each unit invested and that $E\,X > (1 + r)$. Consider the problem

$$\max_{a,b \geq 0}[u(w_0 - (a + b)) + \rho E\, u(W_1 + a(1 + r) + bX)]$$

where X is stochastically independent of W_1. Under the conditions you found above, can you tell what happens to a^* and b^* if X becomes riskier? What about if W_1 becomes riskier?

I. **From Chapter 5.8**: 5.8.17, 5.8.18, and 5.8.19.
J. **From Chapter 5.8**: 5.8.20, and 5.8.21.
K. **From Chapter 5.8**: 5.8.22, 5.8.23, and 5.8.24.
L. **From Chapter 5.9**: 5.9.2, 5.9.5, 5.9.6, and 5.9.7.