Choice under Uncertainty and Information

Monday, Jan 5, 2014

1 Expected Utility

In the XVII century, Blaise Pascal and Pierre Fermat thought that risky prospects ought to be assessed based on their expected values. In 1713 Nicolas Bernoulli described the St. Petersburg paradox prompting his cousin, Daniel, to propose expected utility theory as a solution. In 1728, Gabriel Cramer, in a letter to Nicolas Bernoulli, wrote, "the mathematicians estimate money in proportion to its quantity, and men of good sense in proportion to the usage that they may make of it." In 1738, Daniel Bernoulli, proposed that a mathematical function should be used to correct the expected value.

Information, from an economic perspective, has been studied traditionally in the context of choice theory. Given a space of consequences $X = x_1, x_2, .., x_n$ the set of lotteries is represented $\mathcal{P} = \Delta\left(X\right)$. A lottery $p \in \mathcal{P}$ is a vector of probabilities where p_i is the probability of consequence x_i.

Definition 1 *A utility function $U : \mathcal{P} \to R$ has an **expected utility form** if there are numbers $(u_1, u_2, ..., u_n)$ such that for very $p \in \mathcal{P}$*

$$U\left(p\right) = \sum_{i=1}^{n} p_i u_i.$$

Theorem 2 *(VNM 1944) A complete and transitive preference relation on $\mathcal{P}$ satisfies continuity and independence if and only if it admits an expected utility representation $U : \mathcal{P} \to R$.*

2 Comparing Risky Prospects

When can we say that a lottery pays more than another? This question leads to the idea of First Order Stochastic Dominance.

Proposition 3 *The distribution G first order stochastically dominates distribution F if for all x $G(x) \leq F(x)$, or equivalently if for every non decreasing function u, $\int u(x)\,dG(x) \geq \int u(x)\,dF(x)$.*

Proof. Homework. ■

- Two similar but logically distinct ideas: G is more likely to pay more than F, and a decision maker will prefer G to F.

When can we say that a lottery is more risky than another?

- A comparison of variances is insufficient

- Here we develop a measure of risk based for comparing rv's with the same mean

- Without loss of generality consider a family of rv indexed by r on the closed interval $[0, 1]$

- Let $F(x, r)$ be the cdf of x_r. We assume F is twice continuously differentiable

- Intuitively we say that $F(., r_2)$ is more risky than $F(., r_1)$ if the distribution for r_2 is obtained from the distribution for r_1 by displacing weight from the center to the tails, while keeping the mean constant. FIGURE.

Formally $F(., r_2)$ MR $F(., r_1)$ if and only if

- 1. the two distributions have the same mean. This implies

$$\int_0^1 [F(y, r_2) - F(y, r_2)\, dy] = 0$$

  2. $F(., r_2)$ is obtained from $F(., r_1)$ by a succession of displacements of the type illustrated.

3. Clearly the previous two conditions imply that $\int_0^y [F(x, r_2) - F(x, r_1)] \, dx \geq$ $0 \; \forall y \in [0, 1]$.

- On the other hand, Rothschild and Stiglitz (1970) show that (3) and (1) implies (2).

Proof. Homework. ∎

- We thus arrive at the following definition.

Definition 4 *An increase in r is a mean preserving increase in risk iff*

1. $\int_0^1 \frac{\partial}{\partial r} F(x, r) \, dx = 0$

2. $\int_0^y \frac{\partial}{\partial r} F(x, r) \, dx \geq 0 \ \forall y \in [0, 1].$

- This definition induces a partial ordering over distributions with the same mean.

- Rothschild and Stiglitz (1970) show that

$$x_{r_2} \text{ MR } x_{r_1} \Longleftrightarrow Eu(x_{r_2}) \leq Eu(x_{r_1})$$

- for all concave $u(\cdot)$. Moreover,

$$x_{r_2} \text{ MR } x_{r_1} \Longleftrightarrow x_{r_2} = x_{r_1} + \varepsilon$$

- where ε is a noise such that $E[\varepsilon | x_{r_1}] = 0$.

Proposition 5 *Consider two distributions with the same mean G and F. The distribution G second order stochastically dominates distribution F if for all x, $\int_{-\infty}^{x} G(t)\,dt \leq \int_{-\infty}^{x} F(t)\,dt$ or equivalently if for every concave function u, $\int u\,dG \geq \int u\,dF$.*

Proof. We only prove sufficiency. Integration by parts leads to

$$\int u(x)\,dG - \int u(x)\,dF(x) = \int u'(x)\left[F(x) - G(x)\right]dx$$

Integrating by parts again, and using $\int \left[F(x) - G(x)\right]dx = 0$ (because both distributions have the same mean) yields

$$\int u(x)\,dG(x) - \int u(x)\,dF(x) = \int u''(x)\left[\int_{-\infty}^{x} G(t)\,dt - \int_{-\infty}^{x} F(t)\,dt\right]dx \geq 0$$

where the inequality follows from assumption $u'' \leq 0$ and the hypothesis

$$\left[\int_{-\infty}^{x} G(t) \, dt - \int_{-\infty}^{x} F(t) \, dt \right] \leq 0.$$

This shows that all risk averse decision makers prefer distribution G to F. Homework: prove that if $\int_{-\infty}^{x} G(t) \, dt \leq \int_{-\infty}^{x} F(t) \, dt$ does not hold for all x, then we can find a decision maker (characterized by a concave utility function u) who prefers F to G. ∎

Remark 6 *To demonstrate that the variance is not a good measure of risk we consider an example of two rv's with the same mean x_1 and x_2, where $var(x_1) < var(x_2)$ yet $Eu(x_2) > Eu(x_1)$.*

Example 7 $x_1 = 1$ *(resp 100) with probabilty .8 (resp .2) and $x_2 = 10$ (resp 1090) with probability .1 (resp .99). One can check that $E(x_i) = 20.8$, $var(x_1) = 1568$, and $var(x_2) = 11547$ and $E\log(x_1) = .92 < E\log(x_2) = 2.35$.*

3 Information Structures

3.1 Information Structure without Noise

- Let $(\Omega, \mathcal{O})$ be a measurable space of states of nature.

- The prior information of an agent is represented by a probability measure π on $(\Omega, \mathcal{O})$

- For example $\Omega = \{\omega_1, \omega_2, \omega_3\}$ can be good, average and bad quality

- $\pi(\omega_1)$ is the probability the agent assigns to the product being of good quality

Definition 8 *An information structure* without noise *consists of a space of signals Y and a measurable function φ from the space of states to Y*

INSERT FIGURE

- The function φ defines a partition of Ω, the elements of which are given by

$$O_i = \varphi^{-1}(y_i) \text{ for } y_i \in Y.$$

- A decision maker wishes to maximize his objective function $u(\omega, a)$ with respect to his action $a \in A$ without knowing ω.

- If he is rational, he maximizes expected utility

$$\max_{a \in A} \int_{\Omega} u(a, \omega) \pi(\omega) \, d\pi \tag{1}$$

- Let a^{*0} be the solution to problem (1).

- Let $\mathcal{P}_1 = \{O_1(y), y \in Y_1\}$ be the partition generated by information structure 1.

- EXAMPLE.

- Denote $v(\omega|y)$ the posterior probability distribution. If $\omega \notin O_1(y)$

then $v(\omega|y) = 0$, otherwise

$$v(\omega|y) = \frac{\pi(\omega)}{\int_{O_1(y)} \pi(\tilde{\omega}) \, d\tilde{\omega}}.$$

- Namely, the agent revises his beliefs using Bayes' Theorem.

- For each value of y the agent knows that he will solve the following problem

$$\max_{a \in A} \int_\Omega u(a, \omega) v(\omega|y) \, d\omega$$
$$= \int_\Omega u(a_1^*(y), \omega) v(\omega|y) \, d\omega = V(y).$$

- He can evaluate ex ante the value of having information structure $\mathcal{P}_1$

$$U\left(\mathcal{P}_1\right) = \int_{Y1} V\left(y\right) \pi\left(y\right) dy$$

- where $\pi\left(y\right)$ is the prior probabilty of having the signal y, that is

$$\pi\left(y\right) = \int_{O_1(y)} \pi\left(\omega\right) d\omega.$$

- We say that information structure 1 is better than information structure 2 for the agent if

$$U\left(\mathcal{P}_1, \pi, u\right) > U\left(\mathcal{P}_2, \pi, u\right).$$

- Clearly, this comparision depends on the agent's preferences u and his prior beliefs π.

- But can we compare information structures independently from these characteristics?

Definition 9 *We say that information structure* 1 *is* finer *than information structure* 2 *if the partition generated by structure* 1 *is finer than the one generated by structure* 2, *that is* $\forall O_2 \in \mathcal{P}_2$ *there is* $\left\{ O_i^1 \right\}_{i=1}^{k} : \cup_{i=1}^{k} O_i^1 = O_2.$

Theorem 10 *Information structure* 1 *is finer than information structure* 2 *if and only if for any prior probability distribution* π *and for any utility function* u: $U\left(\mathcal{P}_1, \pi, u\right) \geq U\left(\mathcal{P}_2, \pi, u\right).$

Proof. if (1) is finer than (2), then $\forall y^2$ and O^2, there exists $O_1^1, .., O_k^1 : \cup_{j=1}^{k} O_j^1$. Let $a^{*2}\left(y_2\right)$ be the optimal action given signal y^2. By definition we have

that for all j

$$\max_a \int_{O_j^1} u\left(a, \omega\right) v\left(\omega | y_j^1\right) d\omega \geq \int_{O_j^1} u\left(a^{*2}\left(y^2\right), \omega\right) v\left(\omega | y_j^1\right) d\omega.$$

Let $a^{*1}\left(j\right)$ be the solution to the above problem for $j = 1..., k$. We have

$$\sum_{j=1}^k \pi\left(y_j^1\right) \int_{O_j^1} u\left(a^{*1}\left(j\right), \omega\right) v\left(\omega | y_j^1\right) d\omega$$

$$\geq \sum_{j=1}^k \pi\left(y_j^1\right) \int_{O_j^1} u\left(a^{*2}\left(y_2\right), \omega\right) v\left(\omega | y_j^1\right) d\omega,$$

since this holds for any signal $y^2 \notin Y^2$ the result has been shown. To shown the converse result we must show that for any pair of partitions $(\mathcal{P}_1, \mathcal{P}_2)$ such that neither partition is finer than the other we can find a decision problem in which $\mathcal{P}_1$ is preferred to $\mathcal{P}_2$ and vice versa. FIGURE.
∎

- Prior information can be identified with an uninformative information structure $\mathcal{P}_0$

$$U\left(\mathcal{P}_0, \pi, u\right) = \max_{a \in A} \int_{\Omega} u\left(a, \omega\right) \pi\left(\omega\right) d\pi.$$

- Every information structure is finer than $\mathcal{P}_0$. From Theorem 10 we conclude that an information structure without noise is always valuable to an agent and we can define the value of information structure by

$$V\left(\mathcal{P}_1, \pi, u\right) = U\left(\mathcal{P}_1, \pi, u\right) - U\left(\mathcal{P}_0, \pi, u\right) \geq 0.$$

3.2   Information Structure with Noise

- Consists of a space of signals and a function from Ω to the state of probability measures over Y

- In other words it's given by a conditional probability function $f(y|\omega)$ over Y.

- For example, if y is normally distributed then $f(y|\omega) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\omega)^2}{2\sigma^2}\right)$

- Given $f(y|\omega)$ beliefs are revised according to Bayes, as follows:

$$v(\omega|y) = \frac{f(y|\omega)\,\pi(\omega)}{\int_\Omega f(y|\tilde{\omega})\,\pi(\tilde{\omega})\,d\tilde{\omega}}.$$

- The important point is that the decision maker knows $f(.|.)$, namely he knows the probability the information system will make a mistake.

- Without information the decision maker solves

$$\max_{a \in A} \int_{\Omega} u(a, \omega) \, \pi(\omega) \, d\pi$$

- which yields the optimal decision a^{0*}

- With information structure characterized $f(y|\omega)$, for any value of y he solves

$$\max_{a \in A} \int_{\Omega} u(a, \omega) \, v(\omega|y) \, d\omega$$

- which yields decision $a^*(y)$. By definition of $a^*(y)$ we have

$$\int_{\Omega} u(a^*(y), \omega) \, v(\omega|y) \, d\omega \geq \int_{\Omega} u\left(a^{0*}, \omega\right) v(\omega|y) \, d\omega$$

- therefore

$$\int_Y \int_\Omega u\left(a^*\left(y\right), \omega\right) v\left(\omega|y\right) d\omega \pi\left(y\right) dy$$

$$\geq \int_Y \int_\Omega u\left(a^{0*}, \omega\right) v\left(\omega|y\right) d\omega \pi\left(y\right) dy$$

$$= \int_Y \int_\Omega u\left(a^{0*}, \omega\right) f\left(y|\omega\right) dy \pi\left(\omega\right) d\omega \quad \text{(By Bayes)}$$

$$= \int_\Omega u\left(a^{0*}, \omega\right) \pi\left(\omega\right) d\omega.$$

- Comparing information structures with Noise is more delicate and is the subject of BlacKwell's Theorem.

- Let

$$U\left[Y, f, \pi, u\right] = \sum_Y \pi\left(y\right) \int u\left(a^*\left(y\right), \omega\right) v\left(\omega|y\right) d\omega.$$

Definition 11 *Information structure $\left[Y^1, f^1\right]$ is more valuable than $\left[Y^2, f^2\right]$ iff*

$$U\left[Y^1, f^1, \pi, u\right] \geq U\left[Y^2, f^2, \pi, u\right], \forall u, \pi.$$

Theorem 12 *Blackwell (1951). Information structure $\left[Y^1, f^1\right]$ is more valuable than $\left[Y^2, f^2\right]$ iff $\left[Y^1, f^1\right]$ is* sufficient *for $\left[Y^2, f^2\right]$, namely iff there exists non negative numbers $\beta_{y_k^1, y_{k'}^2}$ such that*

1. $f^2\left(y_{k'}^2|\omega\right) = \sum_{y_k^1 \in Y_1} \beta_{y_k^1, y_{k'}^2} f\left(y_k^1|\omega\right)$ *for all ω and $y_{k'}^2$*

2. $\sum_{y_{k'}^2 \in Y^2} \beta_{y_k^1, y_{k'}^2} = 1$ *for all $y_k^1 \in Y^1$.*

Proof. See Cremer 1982. ∎

This (1) is a generalization of the following idea. Each time y_1 is observed, it is garbled by a stochastic mechanism independent of ω and transformed into a vector of signals in Y^2 via the conditional distribution $p(y_2|y_1)$.

Example 13 *When Y^1, Y^2, Ω have a finite number of elements say 2 elements then*

$$F^1 = \begin{pmatrix} f^1\left(y_1^1|\omega_1\right) & f^1\left(y_1^1|\omega_2\right) \\ f^1\left(y_2^1|\omega_1\right) & f^1\left(y_2^1|\omega_2\right) \end{pmatrix}$$

$$F^2 = \begin{pmatrix} f^2\left(y_1^2|\omega_1\right) & f^2\left(y_1^2|\omega_2\right) \\ f^2\left(y_2^2|\omega_1\right) & f^2\left(y_2^2|\omega_2\right) \end{pmatrix}$$

$$B = \begin{pmatrix} \beta_{y_1^2 y_1^1} & \beta_{y_1^2 y_2^1} \\ \beta_{y_2^2 y_1^1} & \beta_{y_2^2 y_2^1} \end{pmatrix}$$

- B is a Markov probability matrix (if we fix a column and sum across rows we get 1). Condition 1 can be written as

$$F^2 = BF^1.$$

Example 14 *Consider a perfect expert in the example when a good can be either good or bad.*

$$F^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Now consider a Markov Matrix of the form

$$B = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}.$$

Then

$$F^2 = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}$$

So expert 2 is no longer perfect, he errs three out of four times when quality is high and one out of two times when is bad.

3.3 Demski 1973 and the Impossibility of General Accounting Standards

- Demski (1973) asks whether there is an optimal accounting standard defined as a complete ranking of information systems based on (Blackwell) informativeness. As we know from Blackwell (1951), Blackwell's is an incomplete ordering, hence the answer to Demski's question is no.

- For example, we cannot say in general whether a system that provides precise good news is better than one that provides precise bad news: it depends on the decision problem.

- Of course, we know there is a normative criterion to partially ranks information systems in single agent decision settings that holds for all u and π and it's called Blackwell.

3.4   Normally Distributed Signals

- For continuously distributed signals, we say that a signal s is sufficient for signal s' if there is a stochastic transformation $g : S \times S' \to R_+$, where $\int_{Y'} g\left(y', y\right) dy' = 1$ for any $y \in Y$. Assuming that g is integrable

$$f'\left(y'|\omega\right) = \int_Y g\left(y', y\right) f\left(y|\omega\right) dy.$$

For example assume $\omega \sim N\left(\mu, \sigma_\omega^2\right)$ and $y|\omega \sim (\omega, \sigma_\varepsilon^2)$, that is

$$y = \omega + \varepsilon$$

where $\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$ and $\mathrm{cov}(\varepsilon, \omega) = 0$.

- It's clear that y is more informative than y' iff $\tau_\varepsilon \geq \tau_{\varepsilon'}$ because s' can be obtained from s by adding noise.

- In general, y is more informative than y if and only if ε' has the same distribution as $\varepsilon + \xi$ where ε and ξ are independent.

- So if ε' is normal, then both ε and ξ must be normal for y and y' to be Blackwell comparable (see Lehmann (1988)). This implies than only a normally distributed signal can be Blackwell more informative than other normally distributed signal.

- UNIFORM EXAMPLE.

3.4.1 Sufficient Statistic

The notion of sufficiency should not be confused with the notion of sufficient statistic. Roughly, given a set X of independent identically distributed data conditioned on an unknown parameter θ, a sufficient statistic is a function $T(X)$ whose value contains all the information needed to compute any estimate of the parameter (e.g. a maximum likelihood estimate). Due to the factorization theorem, for a sufficient statistic $T(X)$, the joint distribution can be written as

$$p(X) = h(X)g(\theta, T(X)),$$

From this factorization, it can easily be seen that the maximum likelihood estimate of θ will interact with X only through $T(X)$. Typically, the sufficient statistic is a simple function of the data, e.g. the sum of all the data points. Wikipedia.

4   Integral Precision

Blackwell does not tell us how to recognize empirically the informativeness of an information system. In fact, checking Blackwell conditions is notoriously difficult.

Ganuza and Penalva (2010) provide a measure of informativeness based on the dispersion of conditional expectations called integral precision.

- Let V be a random variable representing the state of nature, i.e., the firm value. Let Y_k be a signal.

- For a given prior $H(v)$ we compare a signal Y_1 with another Y_2 in terms of information content. We say that Y_1 is more precise than

Y_2 if $E[V|Y_1]$ is more disperse than $E[V|Y_2]$. We use the notion of dispersion that the satistics literature refers to as the convex order.

Definition 15 X *is greater than Z in the convex order if for all convex real valued functions ϕ, $E[\phi(X)] \geq E[\phi(Z)]$ provided the expectation exists.*

- If both X and Z have the same finite mean, then $X \geq_{cx} Z$ if and only if X is a mean preserving increase in risk of Z.

Definition 16 Y_1 *is more integral precise than Y_2 if $E[V|Y_1]$ is greater in the convex order than $E[V|Y_2]$.*

- Notice that signals are ordered for a given prior. The prior plays a crucial role in the definition.

- Integral precision is a partial order, and it's consistent with Blackwell in the sense that if two signals are ordered according to Blackwell, they will be equally ordered according to integral precision. There are signals that can be ranked according to integral precision but can't be ranked according to Blackwell.

The notion of integral precision implies that if we think of stock prices as expectations conditional on accounting information, then the informativeness of earnings announcements could be measured by estimating the dispersion of stock prices during earnings announcements. HOMEWORK.

5   Entropy

- Entropy is a measure of the uncertainty behind a random variable. It was developed by Claude Shannon in the 1948.

- Mathematically, it is defined as

$$H\left(X\right) = E[I\left(X\right)] = E\left[-\ln P\left(X\right)\right]$$

- We can think of entropy as the average surprise of the information.

Example 17 *For example, if X is a binomial r.v then*

$$H\left(X\right) = -p_0 \ln p_0 - \left(1 - p_0\right) \ln \left(1 - p_0\right).$$

Entropy is maximized when $p_0 = \frac{1}{2}$.

INSERT PLOT HERE

- Information, according to entropy, must satisfy the following conditions:

  1. $I(p) \geq 0$

  2. $I(1) = 0$

  3. $I(p_1 p_2) = I(p_1) + I(p_2)$ if the two events are independent.

- The information of signal Y can be represented as
$$E_Y [H(X) - H(X|Y)].$$

6 Information and Risk Sharing: The Hirshleiffer 1971 Effect

References

Blackwell, David, "Comparison of experiments," in "Proceedings of the second Berkeley symposium on mathematical statistics and probability," Vol. 1 1951, pp. 93–102.

Demski, Joel S, "The general impossibility of normative accounting standards," *Accounting Review*, 1973, pp. 718–723.

Ganuza, Juan-José and Jose S Penalva, "Signal orderings based on dispersion and the supply of private information in auctions," *Econometrica*, 2010, *78* (3), 1007–1030.

Lehmann, E. L., "Comparing Location Experiments," *The Annals of Statistics*, 1988, *16* (2), pp. 521–533.

Rothschild, Michael and Joseph E Stiglitz, "Increasing risk: I. A definition," *Journal of Economic theory*, 1970, *2* (3), 225–243.