U.T. Economics Summer 2011 Math Camp

Date: Monday, August 19 through Thursday, August 22

Topics: Random variables, functions of random variables, expectations, Jensen's inequality and expectations of concave functions of random variables
Readings: Casella and Berger, Ch. 1, Ch.2.1-2

Kolomogorov's model of randomness

I. A **probability space** is a triple, $(\Omega, \mathcal{F}, P)$ where
   A. $\Omega \neq \varnothing$
   B. $\mathcal{F} \subset \mathcal{P}(\Omega)$ satisfies
      1. $\varnothing, \Omega \in \mathcal{F}$
      2. $[E \in \mathcal{F}] \Rightarrow [E^c \in \mathcal{F}]$
      3. $[\{E_n : n \in \mathbb{N}\} \subset \mathcal{F}] \Rightarrow [\cup_{n\in\mathbb{N}} E_n \in \mathcal{F} \wedge \cap_{n\in\mathbb{N}} E_n \in \mathcal{F}]$
   C. $P : \mathcal{F} \to [0, 1]$ satisfies
      1. $P(\varnothing) = 0$, $P(\Omega) = 1$, more generally
      2. $[\{E_n : n \in \mathbb{N}\} \subset \mathcal{F}, \wedge(\forall m, n)[E_n \cap E_m = \varnothing] \Rightarrow [P(\cup_{n\in\mathbb{N}} E_n) = \sum_{n\in\mathbb{N}} P(E_n)]$.
      3. A **random variable (resp. vector)** is a mapping $X : \Omega \to \mathbb{R}$ (resp. $\mathbb{R}^\ell$) such that $(\forall x \in \mathbb{R})[X^{-1}((-\infty, r]) \in \mathcal{F}]$ (resp. $(\forall \mathbf{x} \in \mathbb{R}^\ell)[X^{-1}((-\infty, \mathbf{x}]) \in \mathcal{F}]$).

II. Terminology. $(\Omega, \mathcal{F})$ is a **measurable space**, P is a **probability measure**, and random variables are **measurable functions**. The study of such structures (with fewer assumptions on P) is called **measure theory**, and probability theory can sometimes be usefully regarded as a sub-field of measure theory. For many purposes in economics, the measure theoretic aspect of randomness does not matter, and we will spend most of our time working on those parts of economics where this is true.

III. The intended interpretation. Some ω is drawn according to P, which means that "the probability that $\omega \in E$ is equal to $P(E)$", and there is an ongoing dispute about what *that* means; a random phenomenom that we can measure using real numbers or real vectors takes values the $X(\omega)$, which is what we observe.

IV. Informational content of observations. Observing that the number, or vector, $X(\omega)$, is equal to x, or $\mathbf{x}$, leads to the conclusion that the E that happened is a subset of $X^{-1}(x)$, or $X^{-1}(\mathbf{x})$. If $\Omega \subset \mathbb{R}^{\mathbb{N}}$, i.e. the set of all sequences, with typical element, $(\omega_1, \omega_2, \omega_3, \ldots)$, and $X(\omega) = \omega_n$, then $X^{-1}(x)$ is the set of all sequences of the form

$$(\omega_1, \ldots, \omega_{n-1}, x, \omega_{n+1}, \ldots).$$

This means that there is a great deal of uncertainty left after observing $X(\omega)$.

V. Conditional probability. Often, observational information is given using **conditional probabilities** — given that I observed B, the probability of A is $P(A|B)$. When $P(B) > 0$, this is defined as $P(A|B) = \frac{P(A\cap B)}{P(B)}$. If what I observe is

$B = \{\omega : X(\omega) = x\}$, then $P(A|X)$ is a random variable. This is important, conditional probability is itself a random variable.

Example 1 (The danger of random drug testing)**.** *At random, you are given a test for any of the many drugs that are illegal in your country. Let B be the event that you have some of the drug in your system, and let A be the event that the test says that you have the drug in your system. A good test has $P(A|B)$, the sensitivity, close to 1, and the false positive rate, $P(A|B^c)$, close to 0. Of interest is $P(B|A)$, the proportion of those accused by the test that actually have the drug in their system.*

If the sensitivity is 0.99 and the false positive rate 0.02, both of which are reasonable figures, then $P(B|A)$ is

$$\frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} = \frac{0.99 \cdot P(B)}{0.99 \cdot P(B) + 0.02 \cdot P(B^c)} = \frac{1}{1 + \frac{0.02}{0.99}\frac{P(B^c)}{P(B)}}.$$

If (say) 99% of the population given the random test do not have any of the drug in their system, $P(B^c) = 0.99$, then $P(B|A) = 1/3$. That is, less than half of those turned up by the test actually have any of the drug in their system. What is happening is that in every (say) 100 people given the test, 99 have no drugs in their system, and 2 of them have false positives, while the 1 with drugs in their system has the true positive. In all, only 1 of the three positives are true positives.

If (say) 99.9% of the population given the random test do not have any of the drug in their system, $P(B^c) = 0.999$, then $P(B|A) \simeq 1/21$. That is, less than 5% of those turned up by the test actually have any of the drug in their system.

If (say) the drug testing is not random, but based on probable cause, then $P(B^c)$ might be 0.2. In this case, $P(B|A) \simeq 0.995$. This is better, but still means there is still a 1 in 200 chance that the test is wrong, and one needs to think about very closely about standards of evidence. [*The preceding is also a cautionary tale about too much preventive medicine, giving tests for rare conditions at random turns up a large number of false positives.*]

VI. Characterizing random variables. For a random variable, X, the **cumulative distribution function (cdf)** is $F_X(x) := P(X^{-1}((-\infty, r]) = P(\{\omega : X(\omega) \in (-\infty, x]\})$, and the **reverse cdf** is $G_X(x) = 1 - F_X(x) = P(X > r)$. Knowing all of the values of $F_X(x)$, $x \in \mathbb{R}$, allows us to calculate the value of $P(X^{-1}(A))$ for all the $A \subset \mathbb{R}$ that we care about.[1] In this sense, the cdf characterizes the random variable X.

VII. Characterizing random vectors. $F_X(\mathbf{x}) := P(\{\omega : X(\omega) \leq \mathbf{x}\})$ has the same properties. However, it tells us a great deal about the joint random behavior of the components of the vector.

Example 2. $F_X(x_1, x_2) = \min\{x_1, x_2\}$ *for* $0 \leq x_1, x_2 \leq 1$ *versus* $F_Y(x_1, x_2) = x_1 \cdot x_2$ *for* $0 \leq x_1, x_2 \leq 1$*. Note that* $F_X(r, 1) = F_X(1, r) = F_Y(r, 1) = F_Y(1, r) = r$ *for*

[1] They are called the "Borel measurable subsets of $\mathbb{R}$," though one can go further to the "analytic subsets of $\mathbb{R}$."

all $0 \le r \le 1$, *so that each component of both vectors has the uniform distribution. However,* X *is uniformly distributed on the diagonal in the square* $[0,1] \times [0,1]$ *while* $P(Y \in A) =$ *area of* A*, i.e. the vector* Y *is uniformly distributed over the square* $[0,1] \times [0,1]$.

VIII. Induced distributions. Every random variable or vector induces a distribution on its range defined by $\mu_X(A) = P(X \in A)$, $A \subset \mathbb{R}$ or $A \subset \mathbb{R}^\ell$.

IX. Expectations. For any (measurable) function $u : \mathbb{R} \to \mathbb{R}$ and any random variable X, we have the new random variable $Y(\omega) := u(X(\omega))$. The **expectation of** Y is $E\,Y = E\,u(X) = \int_\Omega u(X(\omega))\,dP(\omega)$, or, after change-of-variables, $E\,u(X) = \int_\mathbb{R} u(x)\,d\mu_X(x)$. If $u(x) = x$, we write this is $E\,X$. We essentially never have $E\,u(X) = u(E\,X)$ when $u(\cdot)$ is not the identity function.

This is important — we expect $E\,u(X) \neq u(E\,X)$.

X. Expectations of concave functions. The important result here is Jensen's inequality.

Lemma 1. *If* $C \subset \mathbb{R}^\ell$ *is convex,* $P(X \in C) = 1$*, and* $u : C \to \mathbb{R}$ *is concave, then* $u(E\,X) \ge E\,u(X)$.

The proof uses the supporting hyperplane theorem.

XI. Probability density functions. If it exists, the **probability density function (pdf)** is $f_X(x) = F_X'(x)$, i.e. the function with the property that $(\forall x \in \mathbb{R})[F_X(x) = \int_{-\infty}^{r} f(t)\,dt]$. Given a pdf, $P(X \in A) = \int_A f(t)\,dt$, and $E\,u(X) = \int_\mathbb{R} u(t)f(t)\,dt$. The multi-dimensional pdf is defined analogously.

XII. Hazard rates. When pdf's exist, **hazard rates exist**, the **hazard rate at** t is $h_X(t) = \frac{f_X(t)}{1-F_X(t)}$, which can be understood as $\frac{1}{dt}P(t < X < t + dt \big| X > t)$ for small (infinitesimal) $dt > 0$.[2]

XIII. Indicator functions. We have used indicator functions to talk about statements and proofs, they are really useful in probability. The **indicator function of the set** A, given by $1_A(t) = 1$ if $t \in A$ and $1_A(t) = 0$ if $t \notin A$. For example, the density of the uniform distribution on the interval $[0,1]$ has density function $f(t) = 1_{[0,1]}(t)$. For another example, $\mathbb{E}\,1_A(X) = P(X \in A)$.

XIV. The space of random variables when Ω is finite, $\mathcal{F} = 2^\Omega$, and $P(\{\omega\}) > 0$ for each $\omega \in \Omega$. This is $\mathbb{R}^\Omega$ using the notation that Y^X is the set of all functions from X to Y. Taking $2 = \{0,1\}$, we have 2^Ω being the set of indicator functions for the class of all subsets of Ω, which we identify with the class of all subsets of Ω.

---

[2]Here are two hazard rate examples not related to what we are doing here:

- If W is the random waiting time until a randomly chosen unemployed person finds a job, then we expect the hazard rate, $h_W(t)$ to depend positively on the unemployment rate (more jobs are open), to depend negatively on the size of the savings account or other resources (e.g. a working spouse) that the person has (they can be pickier about the job that they accept), to depend positively on the number of dependents they have (more hungry mouths burn to feed).
- If W is the random waiting time until the next accident shows up in a hospital emergency room, then we expect the hazard rate, $h_W(t)$, to be higher on Friday and Saturday nights, to be lower in the hours before dawn, and to peak at rush hours during the weekdays.

A. For $X \in \mathbb{R}^\Omega$, $E\,X := \sum_{\omega\in\Omega} X(\omega)P(\{\omega\})$. Letting $\mu_X(x) = P(X^{-1}(x))$ be the **image distribution of P under the random variable** X, we have the change of variable result, $E\,X = \sum_x x \cdot \mu_X(x)$.
B. For $u : \mathbb{R} \to \mathbb{R}$ and $X \in \mathbb{R}^\Omega$, we can define $Y(\omega) = u(X(\omega))$ so that $E\,Y = \sum_\omega u(Y(\omega))P(\omega)$. We can define the image distribution of μ_X under u or the image distribution of P under Y, and we had better get the same result — $\mu_Y(Y = y) = P(Y^{-1}(y)) = \mu_X(u^{-1}(y))$.
C. Taking $u(r) = |r|$ gives an important class of rv's: $\|X\|_1 := E\,|X|$; $\|X\|_2 := (E\,|X|^2)^{1/2}$; more generally, $\|X\|_p := (E\,|X|^p)^{1/p}$ for $p \in [1, \infty)$; and $\|X\|_\infty = \lim_{p\uparrow\infty} \|X\|_p = \max_\omega |X(\omega)|$.
D. $\mathcal{A} \subset \mathbb{R}^\Omega$ is a **vector algebra** if for all $X, Y \in \mathcal{A}$ and all $\alpha, \beta \in \mathbb{R}$, $\alpha X + \beta Y \in \mathcal{A}$ and $XY \in \mathcal{A}$. Note that $\mathcal{A}$ is a vector subspace of $\mathbb{R}^\Omega$. One of the things that we do with vector subspaces is to project onto them. This is equivalent to solving the problem, for $X \in \mathbb{R}^\Omega$,

$$\min_{Y\in\mathcal{A}} \|X - Y\|_2.$$

The solution to this problem is denoted $E\,(X|\mathcal{A})$, which turns out to be an instance of a crucial object in econometrics.

Problems

A. A person accused of a crime by the District Attorney, with the help of the police department, has a probability p, $0 < p < 1$, of being guilty. At the trial,
   i. if the person is guilty, the evidence indicates that they are guilty with probability e_{guilty}, $\frac{1}{2} < e_{guilty} < 1$, and
   ii. if the person is innocent, the evidence indicates that they are guilty with probability $e_{innocent}$, $0 < e_{innocent} < \frac{1}{2}$.
   1. Give and graph the probability, as a function of p, that the person is guilty conditional on the evidence indicating guilt.
   2. Suppose that the jury convicts every time the evidence indicates that the defendant is guilty. If $e_{guilty} = 0.98$ and $e_{innocent} = 0.01$, how reliable must the District Attorney and the police department be in order to have the false conviction rates lower than 0.1% (one in a thousand)?

B. Bill and Harry each tell the truth $\frac{1}{3}$ of the time. Bill says something, Harry hears it and says "Bill just told the truth." What is the probability that Harry actually told the truth? [You need to be explicit about what independence or other assumptions you are making in order to solve this, and learning to be explicit in this way is part of the problem.]

Optimal choice in the presence of randomness

Different actions, $a \in A$, lead to different random variables or vectors, X_a, each of which has an induced distribution μ_a. Each consequence, $X_a(\omega)$, has a utility associated with it, $u(X_a(\omega))$, and actions are ranked by expected utility, $[a \succsim b] \Leftrightarrow$

$[E\, u(X_a) \geq E\, u(X_b)]$. Equivalently, preferences over actions can be regarded as preferences over induced distributions, $[a \succsim b] \Leftrightarrow [\int u(x)\, d\mu_a(x) \geq \int u(x)\, d\mu_b(x)]$.

Note that the preferences over induced distributions are linear in the induced distributions. Let Δ denote the set of induced distributions. Continuous preferences over induced distributions are linear iff they satisfy the **Independence Axiom**: $[\mu \succsim \nu] \Rightarrow (\forall \eta \in \Delta)(\forall \alpha \in (0,1))[(\alpha\mu + (1-\alpha)\eta) \succsim (\alpha\nu + (1-\alpha)\eta)]$.

C. Mary, through hard work and concentration on her education, had managed to become the largest sheep farmer in the county. But all was not well. Every month Mary faced a 50% chance that Peter, on the lam from the next county for illegal confinement of his ex-wife, would steal some of her sheep. Denote the value of her total flock as w, and the value of the potential loss as L, $0 < L < w$, and assume that Mary is a expected utility maximizer with a vNM utility function

$$u(w) = \ln w.$$

1. Assume that Mary can buy insurance against theft for a price of p per dollar of insurance. Find Mary's demand for insurance. At what price will Mary choose to fully insure?
2. Assume that the price of insurance is set by a profit maximizing monopolist who knows Mary's demand function and the true probability distribution of loss. Assume also that the only cost to the monopolist is the insurance payout. Find the profit maximizing linear price (i.e. single price per unit) for insurance. Can the monopolist do better than charging a linear price?

D. Consider a consumer who must allocate her wealth between consumption in two periods, 1 and 2. Assume that the consumer has preferences on consumption streams (c_1, c_2) represented by the utility function

$$U(c_1, c_2) = u(c_1) + u(c_2) \text{ where } u(c_i) = \frac{c_i^{1-a}}{1-a},$$

$0 < a < 1$. Suppose further that she has wealth W at the start of period 1, and receives no other income, so all of her period 2 consumption is supported by saving in period 1, and expects to pay a share t of her savings at the start of period 2 in taxes. Finally, suppose that the tax rate on savings is set by the government at the start of period 2 at the time it is levied, and is uncertain at the time of the saving decision in period 1.

(a) Assume that tax rates are determined according to the density $f(t)$, and carefully write down the consumer's lifetime utility maximization problem.

(b) Assume that t will take on a value of 1/2 or 0 with equal probability. Find the optimal choice of consumption in period 1. Does this increase or decrease with an increase in the parameter a? Explain.

E. [Higher effort for those who find effort cheaper] A decision maker of type θ, $\theta \in \Theta := [\underline{\theta}, \overline{\theta}] \subset (0,1)$, chooses an effort, $e \in [0, \infty)$. An effort of e costs $c(\theta, e) := (1-\theta)c(e)$ in utility terms where $c(e)$ is an increasing function so that higher types correspond to lower costs and to lower marginal costs. After the effort is exerted, there are two possible outcomes, success, $u(\text{success}) = R$, and failure, with $u(\text{failure}) = r$, $R > r > 0$. We assume that the decision maker is an expected utility maximizer. This means that if success and failure are different amounts of money, attitudes toward risk are already encoded in the numbers R and r.

High efforts do not guarantee success, but they do increase the probability of success. In particular, there is a probability $P(e)$ of success and a probability $1 - P(e)$ of failure. Thus, expected utility, as a function of e, θ, R, and r is

$$U(e, R, r, \theta) = R \cdot P(e) + r \cdot (1 - P(e)) - (1 - \theta)c(e),$$

and we define $V(R, r, \theta) = \max_{e \in [0,\infty)} U(e, R, r, \theta)$.

We assume that:

(a) $P : [0, \infty) \to [0, 1]$ is continuous and non-decreasing,

(b) $c : [0, \infty) \to \mathbb{R}$ is continuous, non-decreasing, $c(0) = 0$, and that for all large enough e and all θ, $(1 - \overline{\theta})c(e) > R - r$. [This has the effect of making the optimal e bounded].

1. Show that the mappings $R \mapsto V(R, r, \theta)$, $r \mapsto V(R, r, \theta)$, and $\theta \mapsto V(R, r, \theta)$ are all at least weakly increasing.
2. Suppose that $R' > R$ and that $e' := e^*(R', r, \theta)$ and $e := e^*(R, r, \theta)$ contain only a single point. Intuitively, a higher reward for success ought to make someone work harder for that success. Show that this is true if $P(\cdot)$ is strictly increasing.
3. Suppose that $\theta' > \theta$ and that $e' := e^*(R, r, \theta')$ and $e := e^*(R, r, \theta)$ contain only a single point. Intuitively, higher types have lower costs and lower marginal costs, so ought to be willing to work harder. Treating θ as the parameter and e as the decision variable, show that this is true.

F. [Neyman-Pearson Lemma] Suppose that a random vector X either has density $f(\mathbf{x}|\theta_0)$ or has density $f(\mathbf{x}|\theta_1)$. We do not observe which value θ takes, we only observe $X = \mathbf{x}$, and after observing it, we must choose between the two values. Let us call the guess that $\theta = \theta_0$ the **null hypothesis**, and $\theta = \theta_1$ the **alternative hypothesis**. A **decision rule** is given by a set of values, $\mathbb{X}_r$, called the "rejection region," and the behavioral rule "reject the null hypothesis if $X \in \mathbb{X}_r$, accept it else."

If we reject the null when $\theta = \theta_0$, we are making what is called a **Type I** error, if we accept the null when $\theta = \theta_1$, we are making what is called a **Type II** error. We let $\alpha(\mathbb{X}_r) = P(X \in \mathbb{X}_r|\theta_0) = \int_{\mathbb{X}_r} f(\mathbf{x}|\theta_0)\, d\mathbf{x}$ denote the probability of Type I error, and we let $\beta(\mathbb{X}_r) = P(X \notin \mathbb{X}_r|\theta_1) = \int_{\mathbb{X}_r^c} f(\mathbf{x}|\theta_1)\, d\mathbf{x}$ denote the probability of a Type II error. Notice that these are both conditional probabilities.

In picking $\mathbb{X}_r$, there is a tradeoff between α and β. The Neyman-Pearson Lemma tells us how to optimally make that tradeoff. Let us suppose that we dislike both types of errors, and in particular, that we are trying to devise a test, characterized by its rejection region, $\mathbb{X}_r$, to minimize

$$a \cdot \alpha(\mathbb{X}_r) + b \cdot \beta(\mathbb{X}_r)$$

where $a, b > 0$. The idea is that the ratio of a to b specifies our tradeoff between the two Types of error, the higher is a relative to b, the lower we want α to be relative to β. This problem asks about tests of the form

$$\mathbb{X}_{a,b} = \{\mathbf{x} : af(\mathbf{x}|\theta_0) < bf(\mathbf{x}|\theta_1)\} = \left\{\mathbf{x} : \frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_0)} > \frac{a}{b}\right\}.$$

This decision rule is based on the **likelihood ratio**, and likelihood ratio tests appear regularly in statistics.

1. Show that a test of the form $\mathbb{X}_{a,b}$ solves the minimization problem given above. [Hint: let $\phi(\mathbf{x}) = 1$ if $\mathbf{x} \in \mathbb{X}_r$ and $\phi(x) = 0$ otherwise. Note that $a \cdot \alpha(\mathbb{X}_r) + b \cdot \beta(\mathbb{X}_r) = a \int \phi(\mathbf{x}) f(\mathbf{x}|\theta_0)\, d\mathbf{x} + b \int (1 - \phi(\mathbf{x})) f(\mathbf{x}|\theta_1)\, d\mathbf{x}$, and this is in turn equal to $b + \int \phi(\mathbf{x})[af(\mathbf{x}|\theta_0) - bf(\mathbf{x}|\theta_1)]\, d\mathbf{x}$. The idea is to minimize the last term in this expression by choice of $\phi(\mathbf{x})$. Which $\mathbf{x}$'s should have $\phi(\mathbf{x}) = 1$?]
2. Now suppose that the DM has a prior distribution, $(\mu_0, \mu_1) \gg (0, 0)$ on $\{\theta_0, \theta_1\}$, and two actions available, a_0, a_1. Further suppose that $u(a_0, \theta_0) = r > 0$, $u(a_1, \theta_1) = s > 0$, and $u(a_0, \theta_1) = u(a_1, \theta_0) = 0$. As a function of the value of X, the DM must choose either a_0 or a_1. Find the optimal decision rule.
3. As a function of a and b, find the $\mathbb{X}_{a,b}$ when $X \in \mathbb{R}^n$ is iid Bernoulli(θ), $\theta \in \Theta = \{\theta_0, \theta_1\} \subset (0, 1)$.

G. The basic model of information contains payoff relevant states, $m \in M$, signals, $s \in S$, and actions, $a \in A$. To make our lives simple here, S, M, and A are assumed finite.

The signal-state vector is a random vector with induced distribution μ. The marginal distribution of μ on M is defined by $\mu_M(E) = \mu(E \times S)$ for $E \subset M$ and denoted by $\text{marg}_M(\mu)$ or μ_M. μ_M is the **prior distribution on** M.

The definition for the marginal distribution on S, $\text{marg}_S(\mu) = \mu_S$ is similar.

After $S = s$ is observed, the **posterior distribution on** M is $\text{marg}_M(\cdot|S = s)$. Thus, the posterior distribution, or the conditional distribution, is a random vector taking the value $\text{marg}_M(\cdot|S = s)$ with probability $\mu_S(s)$.

The optimization problem involves $S = s$ being observed, but not m, then $a \in A$ is chosen to maximize $E\,(u(a, m)|S = s)$. Here, expectation is taken with respect to the posterior distribution.

1. Consider the DM facing the problem $\max E\, u(a, Y)$ where $a \in A = \{1, 2, 3\}$, $Y \in M = \{1, 2\}$, $\mu_M(1) = \mu_M(2) = \frac{1}{2}$, and consider three different information structures for this problem:
   - (i) the DM has no information about Y except its distribution, $S = \{1\}$, before picking $a \in A$.
   - (ii) the DM observes S before picking $a \in A$, where the joint distribution of S and Y is given by

$S = 2$	0.15	0.4
$S = 1$	0.35	0.1
	$Y = 1$	$Y = 2$

   - (iii) the DM observes S' before picking $a \in A$, where the joint distribution of S and Y is given by

$S' = 4$	0.05	0.4
$S' = 3$	0.45	0.1
	$Y = 1$	$Y = 2$

The utilities are given by

$a = 3$	0	40
$a = 2$	30	30
$a = 1$	50	0
	$Y = 1$	$Y = 2$

where, e.g. $50 = u(a = 1, Y = 1)$, and so on.

1. For each of the information strutures, show that $\mu_M(E) = \sum_{s \in S} \mu_M(\cdot|S = s)\mu_S(s)$, i.e. the average of the posterior distributions is the prior. [This is a general property, and you should be able to show that it is true for all finite M and S.]
2. Let $(\beta, 1 - \beta) \in \Delta(\{1, 2\})$ be a possible distribution of Y. Give the set of β for which
   a. $\text{argmax}_{a \in A} \int u(a, y)\, d\beta(y) = \{1\}$,
   b. $\text{argmax}_{a \in A} \int u(a, y)\, d\beta(y) = \{2\}$, and
   c. $\text{argmax}_{a \in A} \int u(a, y)\, d\beta(y) = \{3\}$.
3. Graph the function $V : \Delta(\{1, 2\}) \to \mathbb{R}$ given by $V(\beta) = \max_{a \in A} \int u(a, y)\, d\beta(y)$ and show that it is convex.
4. Solve the DM's problem in the cases (i), (ii), and (iii). Show that in each case, the DM's solution $a^*(S)$ or $a^*(S')$, can be had by having the DM calculate their posterior distribution and then maximize according to the rules you just above. Show that this DM prefers (iii) to (ii) to (i).
5. Consider a new DM facing the problem $\max_{b \in B} \int v(b, y)\, d\beta(y)$ where B is an arbitrary set. Define a new function $W : \Delta(\{1, 2\}) \to \mathbb{R}$ by $W(\beta) = \sup_{b \in B} \int v(b, y)\, d\beta(y)$. Show that W is convex.
6. Let η and η' be the distributions of the posteriors you calculated above. To be specific here, η and η' are points in $\Delta(\Delta(\{1, 2\}))$, that is, distributions over distributions. Show that for any convex $f : \Delta(\{1, 2\}) \to \mathbb{R}$, $\int f(\beta)\, d\eta'(\beta) \geq \int f(\beta)\, d\eta(\beta)$.
7. Show that any expected utility maximizing DM with utility depending on Y and their own actions would prefer the information structure S' to the information structure S. [Any DM must have a utility function and a set of options, they have a corresponding $W : \Delta(\{1, 2\}) \to \mathbb{R}$. What property or properties must it have?]

H. Each of the following distributions is given in one way or another. In each case, give the cdf, the reverse cdf, the density, the hazard rate, and the expectation.

1. The uniform distribution on the interval $[a, b]$, $a < b$, which has density
$$f(t) = \tfrac{1}{b-a} 1_{[a,b]}(t).$$
2. The negative exponential distribution with parameter $\lambda > 0$, which has cdf
$$F(t) = 1_{[0,\infty)}(t)[1 - e^{-\lambda t}].$$

3. The Weibull distribution with parameters $\lambda, \gamma > 0$, which is of the form $W = X^\gamma$ where X is a negative exponential with parameter λ. [Hint: for every $t \geq 0$, you know that $P(W \leq t) = P(X^\gamma \leq t) = P(X \leq t^{1/\gamma})$, and you have the cdf for the negative exponential in the previous problem.]
4. The Pareto(α, x) distribution on $[0, \infty)$, $\alpha, x > 0$,

$$F(t) = 1_{[0,\infty)}(t)[1 - \left(\frac{x}{t+x}\right)^\alpha].$$

I. [Discrete time discounting] Suppose that the net benefits for a project are $B_0, B_1, \ldots$ where: net benefits per period are bounded, i.e. there exists a (perhaps quite large) number $\overline{B}$ such that for all t, $|B_t| \leq \overline{B}$; there exists a time period T such that $B_t < 0$ for $t = 1, \ldots, T$ and $B_t > 0$ for $t = T+1, T+2, \ldots$; and for a sufficiently large T', $\sum_{t=0}^{T'} B_t > 0$, i.e. that if we do not discount, then the net benefits outweight the net costs.

1. Show that for ρ sufficiently close to but still strictly less than 1, $\sum_{t=0}^{\infty} \rho^t B_t > 0$.

One interpretation of discounting has to do with the ability to invest resources at t and receive a larger set of resources at $t+1$ or later. If this is true, receiving an amount x at t is worth less than receiving the same x at $t+1$ or later. This has to do with $\rho = \frac{1}{1+r}$ where r is the interest rate. Another interpretation of discounting has to do with the uncertainty of future benefits.

Let τ be a random variable taking on the values $0, 1, \ldots$, and having the **geometric distribution**, i.e. $P(\tau = t) = \delta^t(1 - \delta)$.

2. Give the expectation of τ, $E\,\tau$, as a function of δ.
3. Show that τ is "memoryless," i.e. that $P(\tau = t + s | \tau \geq t) = P(\tau = s)$ for $s, t \geq 0$.
4. Give $E \sum_{t=0}^{\tau} B_t$. ($\sum_{t=0}^{\tau} B_t$ is called, for pretty straightforward reasons, a "random sum." Random sums turn out to be very important in many contexts beyond discounting.)

J. [Continuous time discounting] When the continuously compounded interest rate is r, the worth (or value) of a flow of q/year over the next T years is $W(q, r, T) = \int_0^T qe^{-rt}\,dt$. The corresponding total flow is $F(q, T) = \int_0^T q\,dt = qT$. Since $e^{-rt} < 1$ for all $t > 0$, $W(q, r, T) < F(q, T)$. [More generally, if the flow is $q(t)$, the net worth is $\int_0^T q(t)e^{-rt}\,dt$ and the total flow is $\int_0^T q(t)\,dt$.]

1. If $T = 30$ and $W = \$250,000$ and the interest rate is 6%, i.e. $r = 0.06$, what are the corresponding yearly payments, q? What is the difference between the total flow and W? How do your answers change if $T = 15$?
2. Repeat the previous problem with $r = 0.04$, i.e. with a 1/3 reduction in the interest rate.
3. Show that if $W(q^\circ, r, T) = W^\circ$ and T is large, then $q^\circ \simeq rW^\circ$. To put this another way, if T is large, then flow payments corresponding to W° are approximately rW°. [Thus, if $r = 0.05$, then yearly payments corresponding to a value of, say, $W^\circ = \$100,000$, is $q^\circ = \$5,000$.]

K. [Evaluating flows of costs and benefits in continuous time] In a previous homework set analyzing aspects of discounting in the discrete time case, you showed that if B_t is the net benefit at time t, $\sum_{t=0}^{T} B_t > 0$, and for all $t > T$, $B_t \geq 0$, then for ρ close to 1, $\sum_{t=0}^{\infty} \rho^t B_t > 0$. The implication of this is that if we are sufficiently patient, then positive total net benefits implies positive discounted net benefits. In this case, patience corresponds to ρ close to 1. If we think of $\rho = \frac{1}{1+r}$ in the discrete time case, $\rho \simeq 1$ corresponds to $r \simeq 0$. Here I am asking you to show that the same patience argument holds with continuous time discounting, and the result about patience will show up for $r \simeq 0$.

Suppose that $|q(t)| \leq \overline{Q}$ for all $t \geq 0$, that $q(t) < 0$ for $0 \leq t < T'$, that $q(t) > 0$ for all $t > T'$, and that for some (large) T, $\int_0^T q(t)\, dt > 0$.

1. Show that for small enough $r > 0$, $\int_0^T g(t)e^{-rt}\, dt > 0$.
2. Show that for large enough $r > 0$, $\int_0^T g(t)e^{-rt}\, dt < 0$.
3. Suppose that τ is a random variable with the exponential distribution having parameter λ, i.e. with the cdf $F(t) = P(\tau \leq t) = 1 - e^{-\lambda t}$. Show that $E \int_0^{\tau} g(t)\, dt = \int_0^{\infty} g(t)e^{-\lambda t}\, dt$. [Compare this result to the expected value of the random sum $E \sum_{t=0}^{\tau} B_t$ where τ has a geometric distribution.]