TOPIC. Expectations, continued. This lecture continues our study of expectations. We first consider some extremal problems whose statement and/or solution involves expectations, then study the notion of a "g-mean", and end with a development of Jensen's inequality.

An extremal characterization of $E(X)$. The following theorem has implications for the game in Example 7.5.

Theorem 1. *Let X be a random variable and let f be the function from $\mathbb{R}$ to $[0,\infty]$ defined by*

$$f(c) = E\big((X-c)^2\big). \tag{1}$$

(a) *If $E(X^2) = \infty$, then $f(c) = \infty$ for all c.* (b) *If $E(X^2) < \infty$, then $f(c) < \infty$ for all c, f is uniquely minimized by*

$$c = \mu := E(X) \tag{2}$$

which exists and is finite, and

$$f(\mu) = E\big((X-\mu)^2\big) = E(X^2) - \mu^2 = \mathrm{Var}(X). \tag{3}$$

Proof • Suppose $f(b) < \infty$ for some $b \in \mathbb{R}$. Since

$$(u+v)^2 \le 2(u^2+v^2)$$

for all real numbers u and v, we have

$$(X-c)^2 = \big((X-b)+(b-c)\big)^2 \le 2\big[(X-b)^2+(b-c)^2\big]$$

By properties $\mathsf{E}_{\le}$ and E_{+} of expectation

$$f(c) = E\big((X-c)^2\big) \le 2\big[E\big((X-b)^2\big) + E\big((b-c)^2\big)\big]$$
$$= 2f(b) + 2(b-c)^2 < \infty$$

for all $c \in \mathbb{R}$, and in particular, $f(0) = E(X^2) < \infty$. This argument also shows that if $f(b) = \infty$ for some b, then $f(c) = \infty$ for all c, and in particular $E(X^2) = \infty$.

• Suppose $E(X^2) < \infty$. Since $|X| \le 1 + X^2$, we have

$$E(|X|) \le 1 + E(X^2) < \infty,$$

i.e., $\mu := E(X)$ exists and is finite. We need to show that $c = \mu$ uniquely minimizes

$$f(c) = E[(X-c)^2] = E[X^2 - 2cX + c^2]$$

Since the three summands X^2, $-2cX$, and c^2 are each integrable, we may continue with

$$f(c) = E(X^2) + E\big(-(2cX)\big) + E(c^2) \qquad \text{(by } \mathsf{E}_{+}\text{)}$$
$$= E(X^2) - 2c\mu + c^2 = [E(X^2) - \mu^2] + (c-\mu)^2.$$

This expression is obviously uniquely minimized by $c = \mu$; moreover the minimum is

$$E(X^2) - \mu^2 = f(\mu) = E\big[(X-\mu)^2\big]. \qquad \blacksquare$$

Example 1. Recall that in the game in Example 7.5, you pay me

$$(F-c)^2 - w$$

where c is your guess, w is my wager, and F is a random number chosen from the F-distribution with 3 and 4 degrees of freedom. To minimize your expected loss $E\big((F-c)^2 - w\big)$, at first sight Theorem 1 seems to suggest that you should guess

$$c = E(F) = 4/(4-2) = 2.$$

However, since $E(F^2) = \infty$ (verify that!), Theorem 1 actually says that your expected loss will be infinite, no matter what you guess, or what I wager. The SLLN guarantees that if we play the game repeatedly using independent draws F_1, F_2, ..., my average fortune

$$\frac{1}{n}\sum\nolimits_{k=1}^{n}\big((F_k-c)^2 - w\big)$$

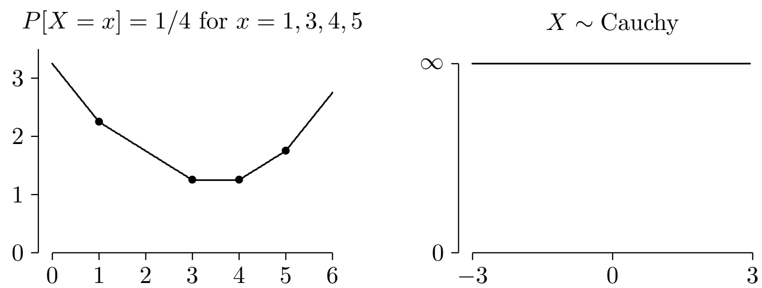
for the first n plays will tend to $E(F-c)^2 - w = \infty$ as $n \to \infty$. I like this game! •

Another extremal theorem. Let X be a random variable with distribution function F and left-continuous representing (or quantile) function R. Consider minimizing

$$f(c) = E(|X - c|)$$

for $c \in \mathbb{R}$. To get some feeling for this, look at these two cases:

Graphs of $f(c)$ versus c



In the left panel, $f(c)$ is finite for all c. Every number c in the range from 3 to 4 is a minimizer; these c's are the medians of X. In the right panel, f is infinite for all c, so every c minimizes f. We are going to show that in general these are the only two possibilities. Recall that m is a median of X if and only if

$$\begin{aligned} &P[X \le m] \ge 1/2 \text{ and } P[X \ge m] \ge 1/2 \\ &\quad\iff R(1/2) \le m \le R(1/2+). \end{aligned} \tag{4}$$

Theorem 2. *Let X be a random variable and set*

$$f(c) = E\big(|X - c|\big) \tag{5}$$

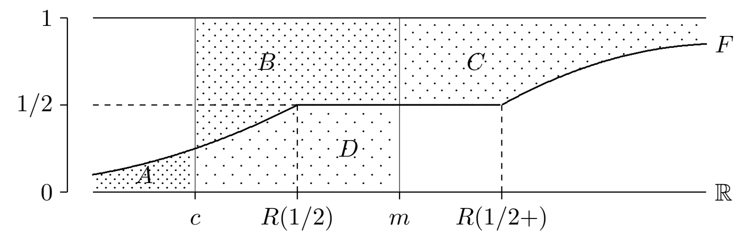
for $-\infty < c < \infty$. (a) *If X is not integrable, then $f(c) = \infty$ for all c.* (b) *If X is integrable, then $f(c) < \infty$ for all c, and c minimizes $f(c)$ if and only if c is a median of X.*

Proof The fact that either $f(c) = \infty$ for all c (and in particular $E(|X|) = \infty$), or $f(c) < \infty$ for all c (and in particular $E(|X|) < \infty$) is easy, and is left to you. For the rest of the argument, suppose X

is integrable. We need to determine the c's that minimize $f(c) := E(|X - c|)$. We can't use the usual calculus technique of solving the equation $f'(c) = 0$ for c because f may not be differentiable. Let m be a median for X. For the time being, suppose $c < m$. Let $U \sim \text{Uniform}(0, 1)$, so $X \sim R(U)$. Then

$$f(c) = E(|R(U) - c|) = \int_0^1 |R(u) - c|\, du = |A| + |B| + |C| \tag{6}$$

where A, B, and C are the regions indicated below:



Similarly,

$$f(m) = E(|R(U) - m|) = |A| + |D| + |C|, \tag{7}$$

with D as indicated above. Since $|A|$ and $|C|$ are finite by (7.13), we may subtract (7) from (6) to get

$$f(c) - f(m) = |B| - |D| \ge 0;$$

the picture shows that equality holds if and only if $R(1/2) \le c$. Similarly, one can show (do it!) that for $c > m$,

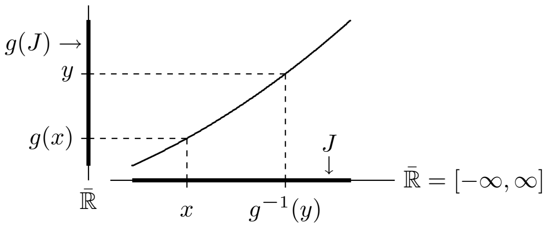
$$f(c) - f(m) \ge 0,$$

with equality if and only if $c \le R(1/2+)$. Consequently c minimizes f if and only

$$R(1/2) \le c \le R(1/2+)$$

i.e., if and only if c is a median of X. ∎

***g*-means.** Let J be a closed subinterval of the extended real-line $[-\infty, \infty]$ and let g be a continuous, strictly monotone mapping from J into $[-\infty, \infty]$. The range $g(J)$ of g is a closed subinterval of $[-\infty, \infty]$. g has an inverse g^{-1} on $g(J)$; g^{-1} is continuous and strictly monotone. This situation is illustrated below:



Now suppose X is a random variable taking values in J. The ***g*-mean of *X*** is defined to be

$$E_g(X) = g^{-1}\big(E\big(g(X)\big)\big); \tag{8}$$

this quantity exists if and only if $g(X)$ has an expectation.

Why are g-means of interest? One answer is that they facilitate making comparisons of the effects of various transformations g_1, g_2, ... on X. The point is that the $E(g_i(X))$'s can be on different scales, whereas the $E_{g_i}(X)$'s are all on the same scale as X.

Another reason g-means are of interest is several common quantities are g-means. From here through (13) below suppose that

$$X \text{ takes values in } J := [0, \infty]; \tag{9}$$

we allow the possibility that $X = \infty$ with positive probability.

• Suppose $g(x) = 1/x$; use the conventions that $1/0 = \infty$ and $1/\infty = 0$. This g is a continuous, strictly decreasing map of J onto itself; moreover $g^{-1} = g$. Hence

$$E_g(X) = \frac{1}{E(1/X)}\,. \tag{10}$$

This is called the ***harmonic mean*** of X; it always exists.

X takes values in $J = [0, \infty]$. $\qquad E_g(X) := g^{-1}\big(E(g(X))\big).$

• Suppose $g(x) = \log(x)$; use the conventions that $\log(0) = -\infty$ and $\log(\infty) = \infty$. This g is a continuous, strictly increasing map of J onto $[-\infty, \infty]$; its inverse is $g^{-1}(y) = e^y$, with the conventions that $e^{-\infty} = 0$ and $e^{\infty} = \infty$. Then

$$E_g(X) = \exp\big(E(\log(X))\big). \tag{11}$$

This is called the ***geometric mean*** of X; it exists if and only if $\log(X)$ has an expectation. For example, suppose $P[X = x_1] = 1/2 = P[X = x_2]$, with $0 \le x_1 < x_2 \le \infty$. If x_1 and x_2 are finite, then $\log(X)$ has an expectation (possibly $-\infty$) and

$$E_g(X) = \exp\big(\log(x_1)/2 + \log(x_2)/2\big) = \sqrt{x_1 x_2}.$$

But if $x_1 = 0$ and $x_2 = \infty$, then $Y = \log(X)$ does not have an expectation since $P[Y = -\infty] = 1/2 = P[Y = \infty]$; in this case the g-mean of X doesn't exist.

• Suppose $g(x) = x$. This is a continuous, strictly increasing map of J onto itself, and $g^{-1} = g$. Here

$$E_g(X) = E(X). \tag{12}$$

This is the ***arithmetic mean*** of X; it always exists.

• Suppose $g(x) = x^2$, with the convention that $\infty^2 = \infty$. This g is a continuous, strictly increasing map of J onto itself; the inverse is $g^{-1}(y) = \sqrt{y}$, with the convention $\sqrt{\infty} = \infty$. Thus

$$E_g(X) = \sqrt{E(X^2)} = \|X\|_2. \tag{13}$$

This is the ***root mean square***, or ***L_2-norm***, of X; it always exists.

Example 2. Consider the power transformations defined on $J = [0, \infty]$ by

$$g_p(x) := \begin{cases} x^p, & \text{if } p \neq 0, \\ \log(x), & \text{if } p = 0, \end{cases} \tag{14}$$

for $-\infty < p < \infty$. We will compute the g_p means of X for a couple of random variables X.

(a) Suppose $X = U$ is standard uniform. Then for nonzero p,

$$E(U^p) = \int_0^1 u^p \, du = \begin{cases} \dfrac{u^{p+1}}{1+p}\Big|_0^1 = \dfrac{1}{1+p}, & \text{if } p > -1, \\ \infty, & \text{if } p \leq -1, \end{cases}$$

so

$$E_{g_p}(U) = \big(E(U^p)\big)^{1/p} = \begin{cases} 1/(1+p)^{1/p}, & \text{if } p > -1, \\ 0, & \text{if } p \leq -1. \end{cases} \tag{15_1}$$

For $p = 0$ we have

$$E\big(\log(U)\big) = \int_0^1 \log(u) \, du = -1 \Longrightarrow E_{g_0}(U) = e^{-1}. \tag{15_2}$$

(b) Suppose $X = F \sim UF(2,2)$; F can be written as the ratio A/B of two independent standard exponential random variables A and B. For $p \neq 0$, we have

$$\begin{aligned} E_{g_p}(F) &= \big(E(A^p)E(1/B^p)\big)^{1/p} \\ &= \begin{cases} 0, & \text{if } p \leq -1, \\ \big(\Gamma(1+p)\Gamma(1-p)\big)^{1/p}, & \text{if } |p| < 1, \\ \infty, & \text{if } p \geq 1, \end{cases} \end{aligned} \tag{16_1}$$

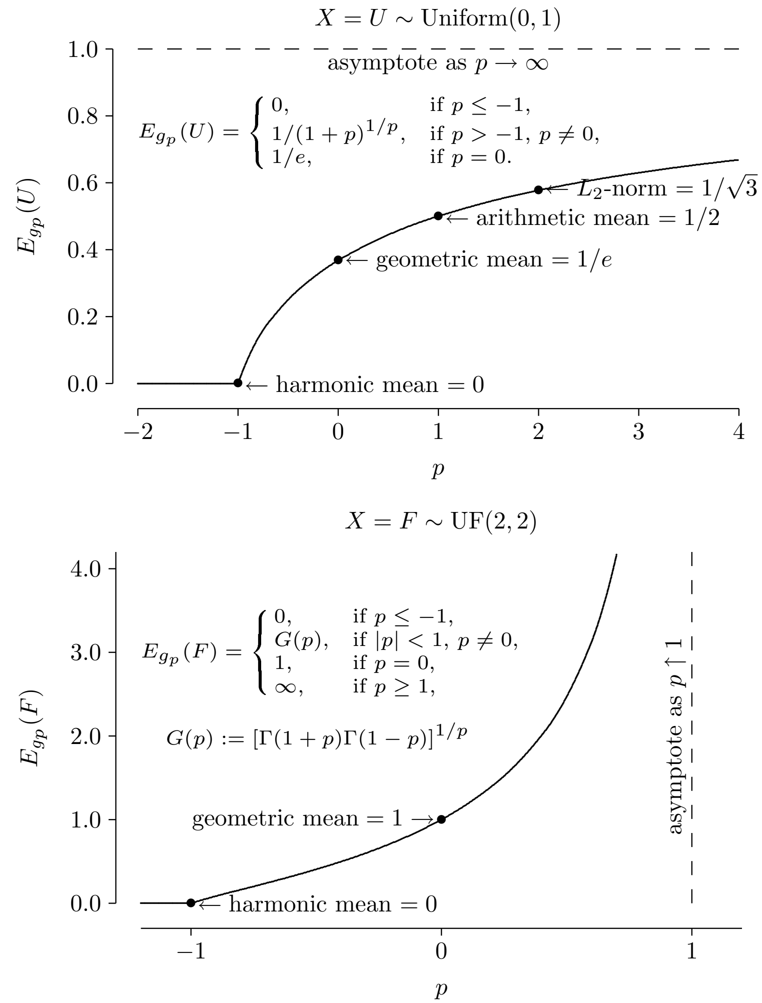
whereas for $p = 0$ we have

$$E_{g_p}(F) = \exp\big(E(\log(A)) - E(\log(B))\big) = e^0 = 1. \tag{16_2}$$

These results are illustrated on the following page. •

Figure 2: g_p-means of X, for the power transformations

$$g_p(x) := \begin{cases} x^p, & \text{if } p \neq 0, \\ \log(x), & \text{if } p = 0. \end{cases}$$

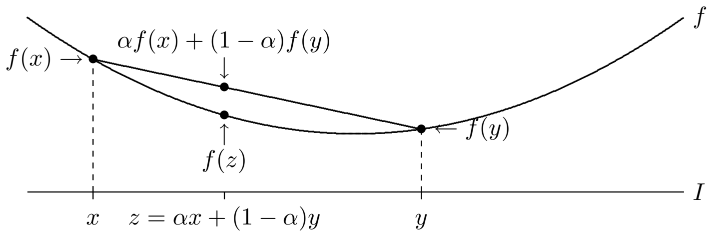


Extrapolating from these two examples, how would you expect the $E_{g_p}(X)$'s to behave for an arbitrary nonnegative random variable X?

Jensen's inequality. Let I be a subinterval of $(-\infty,\infty)$ and let f be a mapping from I to $(-\infty,\infty)$. f is said to be ***convex*** if

$$f\big(\alpha x + (1-\alpha)y\big) \le \alpha f(x) + (1-\alpha)f(y) \quad \text{for all points } x \text{ and } y \text{ in } I \text{ and all } 0 \le \alpha \le 1. \tag{17}$$

(17) says that for x and y in I, the chord from $\big(x, f(x)\big)$ to $\big(y, f(y)\big)$ sits above the graph of f over $[x,y]$, as illustrated below:



There is another interpretation of (17). Let X be a random variable taking the values x and y with probabilities α and $1-\alpha$ respectively. Then the LHS of the inequality in (17) is $f\big(E(X)\big)$, while the RHS is $E\big(f(X)\big)$. Thus (17) says

$$f\big(E(X)\big) \le E\big(f(X)\big) \tag{18}$$

for all random variables taking (at most) two values in I. Jensen's inequality asserts that this relationship holds for every integrable random variable taking values in I:

Theorem 3 (Jensen's inequality). *Let f be a convex function defined on an interval I of $\mathbb{R}$, and let X be an integrable random taking values in I. Then*

(J1) $E(X) \in I$.

(J2) $E\big(f^-(X)\big) < \infty$ *(in particular, $f(X)$ has an expectation).*

(J3) $f\big(E(X)\big) \le E\big(f(X)\big)$.

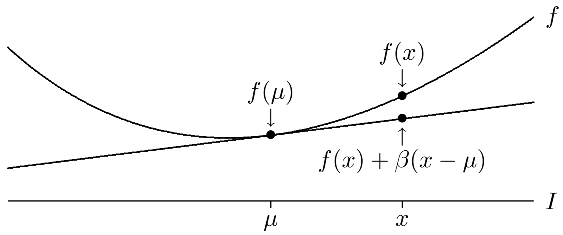
(J1) $E(X) \in I$. (J2) $E\big(f^-(X)\big) < \infty$. (J3) $f\big(E(X)\big) \le E\big(f(X)\big)$.

Proof • (J1) *holds*. There is nothing to prove if both endpoints of I are infinite. Suppose the left endpoint a of I is finite. If $a \in I$, then we have $a \le X$, and so $a \le E(X)$. But if $a \notin I$, then we have $a < X$, and so $a < E(X)$. Similar remarks apply if the right endpoint of I is finite. Consequently $E(X) \in I$ in all cases.

• (J2) *and* (J3) *hold*. Put $\mu = E(X)$. Suppose first that μ lies in the interior of I. According to Exercise 13 there exists a finite number β such that

$$f(x) \ge f(\mu) + \beta(x-\mu) \text{ for all } x \in I, \tag{19}$$

as illustrated below:



Since X takes all its values in I, we have

$$f(X) \ge f(\mu) + \beta(X-\mu) := Y. \tag{20}$$

Taking negative parts in (20) gives

$$f^-(X) \le Y^- \Longrightarrow E\big(f^-(X)\big) \le E(Y^-) < \infty$$

since Y is integrable; thus (J2) holds and $f(X)$ has an expectation. Taking expectations in (20) gives

$$E\big(f(X)\big) \ge E(Y) = f(\mu) + \beta E(X-\mu) = f(\mu),$$

so (J3) holds.

(J1) $E(X) \in I$.   (J2) $E\big(f^-(X)\big) < \infty$.   (J3) $f\big(E(X)\big) \le E\big(f(X)\big)$.
(17): $f\big(\alpha x + (1-\alpha)y\big) \le \alpha f(x) + (1-\alpha)f(y)$.
(19): $f(x) \ge f(\mu) + \beta(x-\mu)$ for all $x \in I$

There is one other possibility, namely, that μ is an endpoint of I. For definiteness, suppose that μ is the left endpoint, say a. Then there might not exist a β such that (19) holds (give an example!). However, in this case we have $X - a \ge 0$ and $E(X-a) = a - a = 0$, so $X = a$ with probability one. Hence $f(X) = f(a)$ with probability one, and (J2) and (J3) hold trivially. ∎

Some addenda: (1) A function $f: I \to \mathbb{R}$ is said to be ***strictly convex*** if strict inequality holds in (17) whenever $x \ne y$ and $0 < \alpha < 1$. If f is strictly convex, one can show that strict inequality holds in (19) for all $x \ne \mu$, and hence that strict inequality holds in (J3) except when X is ***degenerate***, in the sense that there exists a number c (necessarily $= \mu$) such that $X = c$ with probability one.

(2) A function $f: I \to \mathbb{R}$ is said to be ***concave*** if $-f$ is convex, and ***strictly concave*** if $-f$ is strictly convex.

Jensen's inequality has an obvious formulation for concave functions. Simply stated, if f is concave on I and X is an integrable random variable taking values in I, then

$$f\big(E(X)\big) \ge E\big(f(X)\big); \tag{21}$$

moreover, if f is strictly concave, then equality holds in (21) if and only if X is degenerate.

Example 3. The function $f: x \rightsquigarrow x^2$ is strictly convex on $I = \mathbb{R}$. For an integrable X, Jensen's inequality implies the familiar inequality

$$\big(E(X)\big)^2 \le E(X^2),$$

with equality iff X is constant with probability one. Replacing X by $|X|$ we get

$$E(|X|) \le \sqrt{E(|X|^2)}\,,$$

i.e., the arithmetic mean of $|X|$ is no greater than its root mean square. •

Exercise 1. This exercise deals with another way to prove the interesting part of Theorem 2. Suppose that X is an integrable random variable and put $f(c) = E(|X-c|)$ for $c \in \mathbb{R}$. Let m be a median of X and suppose that $c < m$. Define functions Δ and ℓ on $\mathbb{R}$ by

$$\Delta(x) = |x-c| - |x-m|, \qquad \ell(x) = \begin{cases} -(m-c), & \text{if } x < m, \\ m - c, & \text{if } x \ge m. \end{cases}$$

Show that $\Delta(x) \ge \ell(x)$ for all x. Use that inequality and properities of expectation to show that $f(c) - f(m) \ge 0$, with equality if and only if c is itself a median of X. ⋄

Exercise 2. Let X be a real random variable and let q be a number lying strictly between 0 and 1. Set $p = 1 - q$. For $-\infty < c < \infty$ put

$$f(c) := E\big(L_c(X)\big)$$

where

$$L_c(x) := q(x-c)^+ + p(x-c)^- = \begin{cases} q|x-c|, & \text{if } x \ge c, \\ p|x-c|, & \text{if } x \le c. \end{cases}$$

(a) Show that if X is not integrable, then $f(c) = \infty$ for all c. (b) Show that if X is integrable, then $f(c)$ is finite for all c, and c minimizes $f(c)$ if and only if c is a q^{th}-quantile for X. (c) Suppose that $X \sim N(0,1)$. Find a simple expression for $\alpha := \inf\{f(c) : c \in \mathbb{R}\} = f(\Phi^{-1}(q))$ in terms of the normal density ϕ. For what q is α the largest? ⋄

Exercise 3. Suppose F has an unnormalized F-distribution with 3 and 5 degrees of freedom. Let g_p be the power transformations defined by (14). Plot the g_p means of F for $-1 \le p \le 1$. Choose an appropriate vertical scale. ◇

Exercise 4. (a) Suppose J is a closed subinterval of $[-\infty, \infty]$ and g is a continuous, strictly monotone function from J to $[-\infty, \infty]$. Let a and b be finite numbers, with $b \neq 0$, and let h be the mapping from J to $[-\infty, \infty]$ defined by $h(x) = a + bg(x)$ for each $x \in J$; note that h is continuous and strictly monotone. Let X be a random variable taking values in J. Show that the h-mean of X exists if and only if the g-mean does, in which case the two are equal: $E_h(X) = E_g(X)$.
(b) Suppose $J = [0, \infty]$ and $h_p(x) = (x^p - 1)/p$ for $p \neq 0$. Find $\lim h_p(x)$ as $p \to 0$. ◇

A real-valued random variable X is said to have ***mode*** x_m if X has a probability mass function, or a density function, say f, and

$$f(x_m) \ge f(x) \tag{22}$$

for all possible values x of X. Note that not all random variables have modes, and that modes may not be unique.

Exercise 5. Find the modes of the following random variables: (i) $X \sim \text{Binomial}(n, p)$; (ii) $X \sim F$ with $m > 2$ and n degrees of freedom. ◇

Exercise 6. Suppose g is a real-valued, continuous, strictly increasing function on an interval $J \subset \mathbb{R}$ and X is a random variable taking values in J. Is there any general relationship between the median of $g(X)$ and g(the median of X)? Ditto, for the mode (assuming X and $g(X)$ have modes)? ◇

Let J be a subinterval of $\mathbb{R}$ and let g be a real-valued, continuous, strictly-increasing function on J. Let X be a random variable taking values in J. The ***g-median of X*** is defined to be

$$\text{Median}_g(X) := g^{-1}\big(\text{Median of } g(X)\big). \tag{23}$$

Similarly, the ***g-mode of X*** is

$$\text{Mode}_g(X) := g^{-1}\big(\text{Mode of } g(X)\big). \tag{24}$$

g-medians always exist; the g-mode of X exists if $g(X)$ has a mode.

Exercise 7. Let J and g be as above. Let a and $b > 0$ be constants put $h(x) = a + bg(x)$ for all $x \in J$. Show that the g-and h-medians of X are the same. Show that the g- and h-modes are the same, provided they exist. ◇

Exercise 8. Let $r > 0$ and let X_r be a Gamma random variable with density $f_r(x) = I_{(0,\infty)}(x)\, x^{r-1} \exp(-x)/\Gamma(r)$.

(a) Find the g_p-means and g_p-modes of X_r for the power transformations g_p in (14).

(b) For $r = 1$, 4, and 16, numerically evaluate and plot the g_p-mean, g_p-mode, and the g_p-median against p in the interval $[-1/6, 1]$, including at least the integral multiples of $1/6$ in the range $[0, 1/2]$. Make a separate plot for each r, but include the mean, median, and mode on the same plot.

(c) For what value of p do you think the distribution of $g_p(X_r)$ is the most nearly symmetric? Why? What bearing does this have on the problem studied in Homework 2?

[Remark: The formula for the g-means when $p = 0$ involves the derivative of the log of the gamma function:

$$\psi(r) = \frac{d}{dr} \log \Gamma(r) = \frac{\Gamma'(r)}{\Gamma(r)} = \int_0^\infty \log(x) f_r(x)\, dx, \tag{25}$$

which is known as the ***digamma*** function. A good reference the properties of ψ, and many other analytic and numerical facts, is the *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun. The name for ψ in MAPLE is `Psi`. SPLUS has a `gamma` function that computes Γ. However, it doesn't have a function to compute ψ; hence the need for the previous reference. $\diamond$

Exercises 9–15 develop some properties of convex functions. In all of them, J is a subinterval of $\mathbb{R}$ ($J = \mathbb{R}$ is allowed) and f is real-valued function on J. Let I be the interior of J.

Exercise 9. Show that if f is convex, then

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(z)-f(x)}{z-x} \le \frac{f(z)-f(y)}{z-y} \tag{26}$$

for all points x, y, and z in J with $x < y < z$. A "proof by picture" is acceptable, provided you draw the right picture and explain how it implies (26). $\diamond$

Exercise 10. Conversely, show that f is convex if

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(z)-f(y)}{z-y} \tag{27}$$

for all points x, y, and z in J with $x < y < z$. $\diamond$

Exercise 11. Show that if f is convex, then

$$(D_+ f)(x) := \lim_{y \downarrow x,\, y > x} \frac{f(y)-f(x)}{y-x} \tag{28}$$

exists and is finite for each point $x \in I$. [$(D_+ f)(x)$ is called the ***right-hand derivative of f at x***.] $\diamond$

Exercise 12. Similarly, show that a convex f has a left-hand derivative $(D_- f)(x)$ at each $x \in I$. Show further that

$$(D_- f)(x) \le (D_+ f)(x) \le (D_- f)(y) \le (D_+ f)(y) \tag{29}$$

for all points x and y in I with $x < y$. Deduce that set of points $x \in I$ at which f is not differentiable is at most countable. $\diamond$

Exercise 13. Show that if f is convex, then for each $x \in I$

$$f(y) \ge f(x) + \big((D_+ f)(x)\big)(y - x) \tag{30}$$

for all $y \in J$. Show further that if f is strictly convex, then strict inequality holds in (30) unless $y = x$. $\diamond$

Exercise 14. Suppose that f is continuous on J and differentiable on I, and f' is nondecreasing on I. Show that f is convex. [Hint: use the mean value theorem to verify (27).] What condition on f' guarantees that f is strictly convex? $\diamond$

Exercise 15. Suppose that f is continuous on J and twice differentiable on I, and $f'' \ge 0$ on I. Show that f is convex. What condition on f'' guarantees that f is strictly convex? $\diamond$

Exercise 16. Show that the function

$$f(x) := \begin{cases} x\log(x), & \text{if } 0 < x, \\ 0, & \text{if } x = 0, \end{cases} \tag{31}$$

is strictly convex on $[0, \infty)$. Show that for any nonnegative numbers $x_1, x_2, \ldots, x_k$ and strictly positive weights $w_1, w_2, \ldots, w_k$ summing to 1, one has

$$f(w_1 x_1 + \cdots + w_k x_k) \le \sum_{j=1}^{k} w_k f(x_k), \tag{32}$$

with equality if and only if the x_i's are all equal. $\diamond$