Stat 5101 Lecture Slides: Deck 6

Existence of Integrals and Infinite Sums, Countable Additivity and Monotone Convergence, Existence of Moments, Correlation

Charles J. Geyer
School of Statistics
University of Minnesota

Existence of Integrals

Just from the definition of integral as area under the curve, the integral

$$\int_a^b g(x)\,dx$$

always exists when a and b are finite and g is *bounded*, which means there exists a finite c such that

$$|g(x)| \leq c, \qquad a < x < b.$$

In this case

$$\left| \int_a^b g(x)\,dx \right| \leq \int_a^b |g(x)|\,dx \leq c(b-a)$$

Existence of Integrals (cont.)

It is a theorem of advanced calculus (which we will not prove) that every continuous function g having domain $[a, b]$ where a and b are finite is bounded.

So if we know g is continuous on $[a, b]$, then we know

$$\int_a^b g(x)\,dx$$

exists.

It is important that the domain is a *closed* interval. The function $x \mapsto 1/x$ is continuous but unbounded on $(0, 1)$. So continuous on an *open* interval is not good enough.

Existence of Integrals (cont.)

We are worried about non-existence. Clearly

$$\int_a^b g(x)\,dx$$

may fail to exist in one of two cases

(I) either a or b is infinite, or

(II) g is unbounded (meaning not bounded).

Existence of Integrals: Case I

So when does

$$\int_a^\infty g(x)\,dx$$

exist?

In probability theory, we require absolute integrability, so

$$\int_a^\infty |g(x)|\,dx$$

must be finite.

Existence of Integrals: Case I (cont.)

First we do a very important special case. Suppose $a > 0$, then

$$\int_a^\infty x^\alpha \, dx < \infty$$

if and only if $\alpha < -1$.

Existence of Integrals: Case I (cont.)

Case I of Case I, if $\alpha \neq -1$, then

$$\int_a^b x^\alpha \, dx = \left. \frac{x^{\alpha+1}}{\alpha+1} \right|_a^b = \frac{b^{\alpha+1} - a^{\alpha+1}}{\alpha+1}$$

If $\alpha > -1$, then this goes to infinity as $b \to \infty$. If $\alpha < -1$, then this goes to $-a^{\alpha+1}/(\alpha+1)$ as $b \to \infty$.

Case II of Case I, if $\alpha = -1$, then

$$\int_a^b x^\alpha \, dx = \log(x) \Big|_a^b = \log(b) - \log(a)$$

and this goes to infinity as $b \to \infty$.

Existence of Integrals: Comparison Principle

Also obvious from the definition of integral as area under the curve, if

$$|g(x)| \leq |h(x)|, \qquad a < x < b,$$

then

$$\int_a^b |g(x)| \, dx \leq \int_a^b |h(x)| \, dx$$

including when either integral is infinite, that is, when the right-hand side is finite, then so is the left-hand side and when the left-hand side is infinite, then so is the right-hand side.

Existence of Integrals: Case I (cont.)

Suppose $a > 0$, suppose g is continuous on $[a, \infty)$, and suppose

$$\lim_{x \to \infty} \frac{|g(x)|}{x^\alpha} = c$$

exists and is finite. If $\alpha < -1$, then

$$\int_a^\infty |g(x)|\, dx < \infty.$$

Conversely, if $c > 0$ and $\alpha \geq -1$, then

$$\int_a^\infty |g(x)|\, dx = \infty.$$

Existence of Integrals: Case I (cont.)

From the definition of limit, we know there exists a finite r such that

$$\frac{c}{2} \le \frac{|g(x)|}{x^\alpha} \le 1 + c, \qquad x \ge r$$

and we know that

$$\int_a^r |g(x)| \, dx < \infty$$

and

$$\frac{c}{2} \int_r^\infty x^\alpha \, dx \le \int_r^\infty |g(x)| \, dx \le (1 + c) \int_r^\infty x^\alpha \, dx$$

Hence the result about $g(x)$ follows from the result about x^α.

Existence of Integrals: Case I (cont.)

There exists a constant c such that

$$f(x) = \frac{c}{1 + x^2 + 3(x - 1)^4}, \qquad -\infty < x < \infty$$

is a PDF.

Compare

$$\frac{f(x)}{|x|^{-4}} \to \frac{c}{3}$$

as $x \to -\infty$ or $x \to +\infty$. Since $-4 < -1$, it follows that the integral of f is finite.

Existence of Integrals: Case I (cont.)

In the preceding example we used two important principles.

- Constants don't matter.
- In polynomials, only the term of highest degree matters.

Existence of Integrals: Case I (cont.)

In more detail, if c is a constant

$$\int_a^b cg(x)\,dx = c\int_a^b g(x)\,dx$$

and both sides exist or neither does. And

$$\lim_{x\to\infty} \frac{a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k}{x^k} = a_k$$

Existence of Integrals: Case I (cont.)

Returning to our example, suppose X has PDF

$$f(x) = \frac{c}{1 + x^2 + 3(x-1)^4}, \qquad -\infty < x < \infty,$$

then for what positive values of β does $E(|X|^\beta)$ exist?

Compare

$$\frac{|x|^\beta f(x)}{|x|^{\beta-4}} \to \frac{c}{3}$$

as $x \to -\infty$ or $x \to +\infty$. Since $\beta - 4 < -1$, if and only if $\beta < 3$, it follows $E(|X|^\beta)$ exists if and only if $\beta < 3$ (when $\beta > 0$ is assumed).

The Cauchy Distribution

There exists a constant c such that

$$f(x) = \frac{c}{1 + x^2} \qquad -\infty < x < \infty$$

is a PDF.

Compare

$$\frac{f(x)}{|x|^{-2}} \to c$$

as $x \to -\infty$ or $x \to +\infty$. Since $-2 < -1$, it follows that the integral of f is finite.

The Cauchy Distribution (cont.)

In this case we can actually determine the constant.

$$\int_{-t}^{t} \frac{dx}{1 + x^2} = \text{atan}(x)\Big|_{-t}^{t} = \text{atan}(t) - \text{atan}(-t),$$

where atan is the arctangent function, which goes from $-\pi/2$ to $\pi/2$ as its argument goes from $-\infty$ to ∞. Thus

$$\int_{-\infty}^{\infty} \frac{dx}{1 + x^2} = \lim_{t \to \infty} \Big[\text{atan}(t) - \text{atan}(-t)\Big] = \pi$$

and $c = 1/\pi$.

The Cauchy Distribution (cont.)

The distribution with PDF

$$f(x) = \frac{1}{\pi(1 + x^2)}, \qquad -\infty < x < \infty$$

is called the *standard Cauchy distribution*.

The distribution with PDF

$$f_{\mu,\sigma}(x) = \frac{\sigma}{\pi} \cdot \frac{1}{\sigma^2 + (x - \mu)^2}, \qquad -\infty < x < \infty$$

is called the *Cauchy distribution* with location parameter μ and scale parameter σ and is abbreviated Cauchy(μ, σ).

The Cauchy Distribution (cont.)

The Cauchy(μ, σ) distributions are a location-scale family.

The Cauchy(μ, σ) distribution is symmetric about μ, so the parameter μ can be called the *center of symmetry* as well as the *location parameter*.

The Cauchy Distribution (cont.)

If X has the Cauchy(μ, σ) distribution, then for what positive values of β does $E(|X|^\beta)$ exist?

Compare

$$\frac{|x|^\beta f(x)}{|x|^{\beta - 2}} \to \frac{\sigma}{\pi}$$

as $x \to -\infty$ or $x \to +\infty$. Since $\beta - 2 < -1$, if and only if $\beta < 1$, it follows $E(|X|^\beta)$ exists if and only if $\beta < 1$ (when $\beta > 0$ is assumed).

Summary: If X has the Cauchy(μ, σ) distribution, then $E(X^k)$ exists for *no* positive integer k. The mean does not exist, neither does the variance. Hence μ cannot be the mean, and σ cannot be the standard deviation.

Existence of Integrals: Case I (cont.)

The Maclaurin series

$$e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^k}{k!} + \cdots$$

which we also know as the theorem that a Poisson PMF sums to one, shows that e^x grows faster than any polynomial as $x \to \infty$. Similarly for $e^{\lambda x}$ when $\lambda > 0$.

Hence for any $\beta \in \mathbb{R}$, any $\alpha \in \mathbb{R}$, any $\lambda > 0$, and any $a > 0$

$$\lim_{x \to \infty} \frac{x^\beta e^{-\lambda x}}{x^\alpha} = \lim_{x \to \infty} \frac{x^{\beta - \alpha}}{e^{\lambda x}} = 0$$

and

$$\int_a^\infty x^\beta e^{-\lambda x} \, dx < \infty$$

Existence of Integrals: Case II

Now we turn to case II. The domain of integration is bounded, but the integrand is unbounded.

Again we start with the monomial special case. If $a > 0$, then

$$\int_0^a x^\alpha \, dx < \infty$$

if and only if $\alpha > -1$.

Note that the magic exponent -1 is the same, but the inequality is reversed.

Existence of Integrals: Case II (cont.)

The substitution $x = 1/y$ reduces this to the other case.

$$\int_0^a x^\alpha \, dx = \int_\infty^{1/a} y^{-\alpha} \left(-y^{-2}\right) \, dy = \int_{1/a}^\infty y^{-\alpha-2} \, dy$$

and we already know the latter is finite if and only if $-\alpha - 2 < -1$, which is the same as $-\alpha < 1$ or $\alpha > -1$.

Existence of Integrals: Case II (cont.)

We can move this theorem to any other point. If $a < b$, then

$$\int_a^b (x - a)^\alpha \, dx < \infty$$

if and only if $\alpha > -1$, and

$$\int_a^b (b - x)^\alpha \, dx < \infty$$

if and only if $\alpha > -1$.

Existence of Integrals: Case II (cont.)

And we can analyze other integrals by comparison. Suppose g is continuous on $(a, b]$ and suppose

$$\lim_{x \downarrow a} \frac{|g(x)|}{(x-a)^{\alpha}} = c$$

exists and is finite. If $\alpha > -1$, then

$$\int_a^b |g(x)| \, dx < \infty.$$

Conversely, if $c > 0$ and $\alpha \leq -1$, then

$$\int_a^b |g(x)| \, dx = \infty.$$

Existence of Integrals: Case II (cont.)

The case where g is unbounded at b is an obvious modification. Suppose g is continuous on $[a, b)$ and suppose

$$\lim_{x \uparrow b} \frac{|g(x)|}{(b-x)^{\alpha}} = c$$

exists and is finite. If $\alpha > -1$, then

$$\int_a^b |g(x)|\, dx < \infty.$$

Conversely, if $c > 0$ and $\alpha \leq -1$, then

$$\int_a^b |g(x)|\, dx = \infty.$$

Existence of Integrals: Case I and II Summary

If $g(x)$ is continuous on $[a, \infty)$, then

$$\int_a^\infty |g(x)|\, dx < \infty$$

only if $g(x)$ goes to zero as $x \to \infty$ fast enough, faster than $1/x$.

If $g(x)$ is continuous on (a, b), then

$$\int_a^b |g(x)|\, dx < \infty$$

only if $g(x)$ goes to infinity as $x \to a$ or as $x \to b$ slow enough, slower than $1/(x - a)$ or $1/(b - x)$.

Existence of Integrals: Gamma Distribution

When does there exist a c such that

$$f(x) = cx^{\alpha-1}e^{-\lambda x}, \qquad 0 < x < \infty$$

is a PDF?

We have already been told this is the PDF of the Gam(α, λ) distribution when $\alpha > 0$ and $\lambda > 0$, but we haven't proved it.

If $\lambda > 0$, then we know from applying our theorem about case I that

$$\int_a^\infty x^{\alpha-1}e^{-\lambda x}\,dx < \infty$$

for any real α and any $a > 0$.

Existence of Integrals: Gamma Distribution (cont.)

Still assuming $\lambda > 0$, we need to apply our theorem about case II for the integral

$$\int_0^a x^{\alpha-1} e^{-\lambda x} \, dx.$$

When is that finite?

Since

$$\lim_{x \downarrow 0} \frac{x^{\alpha-1} e^{-\lambda x}}{x^{\alpha-1}} = 1,$$

we have

$$\int_0^a x^{\alpha-1} e^{-\lambda x} \, dx < \infty$$

if and only if $\alpha - 1 > -1$, that is, if and only if $\alpha > 0$.

Existence of Integrals: Gamma Distribution (cont.)

Could we have $\lambda < 0$? No, because then $x^{\alpha-1} e^{-\lambda x} \to \infty$ as $x \to \infty$ and the integral cannot be finite.

Could we have $\lambda = 0$? Then for $a > 0$ we have

$$\int_a^\infty x^{\alpha-1} e^{-\lambda x} \, dx = \int_a^\infty x^{\alpha-1} \, dx$$

finite if and only if $\alpha - 1 < -1$, and we have

$$\int_0^a x^{\alpha-1} e^{-\lambda x} \, dx = \int_0^a x^{\alpha-1} \, dx$$

finite if and only if $\alpha - 1 > -1$. So no α works.

Existence of Integrals: Gamma Distribution (cont.)

Summarizing our analysis

$$\int_0^\infty x^{\alpha-1} e^{-\lambda x} \, dx$$

is finite if and only if $\alpha > 0$ and $\lambda > 0$.

Hence, if X has the Gam(α, λ) distribution, then $E(X^\beta)$ exists if and only if $\alpha + \beta > 0$.

Existence of Sums

We handle infinite sums by comparing them with integrals. Think of an infinite sum as the integral of a step function, and get the following. Suppose

$$\lim_{i \to \infty} \frac{|a_i|}{i^\alpha} = c$$

exists and is finite. If $\alpha < -1$, then

$$\sum_{i=1}^{\infty} |a_i| < \infty.$$

Conversely, if $c > 0$ and $\alpha \geq -1$, then

$$\sum_{i=1}^{\infty} |a_i| = \infty.$$

Existence of Sums (cont.)

For example, there is no constant c such that

$$f(x) = \frac{c}{x}, \qquad x = 1, 2, \ldots$$

is a PMF, but there is a constant c such that

$$f(x) = \frac{c}{x^2}, \qquad x = 1, 2, \ldots$$

is a PMF.

Very similar to "Case I" of existence of integrals.

Deja Vu All Over Again

Now we redo the axioms.

This time, we are very careful about existence of expectations.

Sharpening the Axioms

Our axioms for expectation are

$$E(X + Y) = E(X) + E(Y) \tag{1}$$

$$E(X) \geq 0, \qquad \text{when } X \geq 0 \tag{2}$$

$$E(aX) = aE(X) \tag{3}$$

$$E(1) = 1 \tag{4}$$

Now we add the proviso that in (1), (3), and (4), the expectation on the left-hand side exists if all expectations on the right-hand side exist.

Sharpening the Monotonicity Axiom

If X and Y are nonnegative random variables such that $X \leq Y$, then the expectation of X exists whenever the expectation of Y exists, and

$$E(X) \leq E(Y)$$

We already knew $X \leq Y$ implies $E(X) \leq E(Y)$ *when the expectations exist*, but this tells us something about when they do not, that is, when the right-hand side is finite, then so is the left-hand side and when the left-hand side is infinite, then so is the right-hand side.

Sharpening the Axioms (cont.)

All of these sharpenings of the axioms hold for expectations defined by summation or integration or by a combination of the two.

Calling them "axioms" means we assert they also hold for expectation defined any other way. And what would that be?

The answer to that question is really beyond the scope of this course, but we take a brief digression into advanced probability theory to give a hint at the answer.

The Monotone Convergence Axiom

If X_1, X_2, $\ldots$ is an increasing sequence of nonnegative random variables, meaning

$$0 \le X_1(s) \le X_2(s) \le \cdots, \qquad \text{for all } s,$$

then

$$E\left(\lim_{n \to \infty} X_n\right) = \lim_{n \to \infty} E(X_n)$$

so monotone limits can be moved outside expectations.

The Monotone Convergence Axiom (cont.)

The random variable in

$$E\left(\lim_{n \to \infty} X_n\right)$$

is the pointwise limit

$$X(s) = \lim_{n \to \infty} X_n(s)$$

The limit always exists (perhaps $+\infty$) because the limit of a monotone sequence always exists (if $+\infty$ is allowed as a limit).

The Monotone Convergence Axiom (cont.)

In order for this axiom to make sense, we need to define what $E(X)$ means when X is nonnegative and allowed to have the value $+\infty$.

Let

$$A = \{ s \in S : X(s) = \infty \}$$

Then we have two cases. If $\Pr(A) > 0$, then $E(X) = +\infty$. If $\Pr(A) = 0$, then

$$E(X) = E\{X I_{A^c}(X)\}$$

The Monotone Convergence Axiom (cont.)

The monotone convergence axiom is adopted in the vast majority of advanced probability theory, despite its having no motivation other than mathematical convenience.

It has the very great inconvenience that we have to redefine integration in order to make it hold. It does *not* hold for probabilities and expectations defined by the kind of integration taught in first, second, and third year calculus.

Nothing we have said up to here requires the monotone convergence axiom except for the properties of distribution functions and the implications of variance zero discussed again below.

The Monotone Convergence Axiom (cont.)

A convenient shorthand is "uparrow" for monotone convergence. We can shorten the axiom to

$$X_n \uparrow X \quad \text{implies} \quad E(X_n) \uparrow E(X)$$

By subtracting an arbitrary function from both sides of the limit, we see that this holds even if the X_n are not nonnegative, so long as all of the expectations $E(X_n)$ exist.

This also implies

$$X_n \downarrow X \quad \text{implies} \quad E(X_n) \downarrow E(X)$$

with the obvious definition of "downarrow", still assuming all $E(X_n)$ exist.

Continuity of Probability

For events, we write $A_n \uparrow A$ if

$$A_1 \subset A_2 \subset \cdots \quad \text{and} \quad A = \bigcup_{n=1}^{\infty} A_n$$

and $A_n \downarrow A$ if

$$A_1 \supset A_2 \supset \cdots \quad \text{and} \quad A = \bigcap_{n=1}^{\infty} A_n$$

Probability is expectation of indicator functions then implies

$$A_n \uparrow A \quad \text{implies} \quad \Pr(A_n) \uparrow \Pr(A)$$

and

$$A_n \downarrow A \quad \text{implies} \quad \Pr(A_n) \downarrow \Pr(A)$$

This is called *continuity of probability*.

Continuity of Probability and DF

Distribution functions are right continuous

$$F(x) = \Pr(X \leq x) = \lim_{y \downarrow x} F(y)$$

and have left limits

$$F_{-}(x) = \Pr(X < x) = \lim_{y \uparrow x} F(y)$$

and

$$\lim_{y \downarrow -\infty} F(y) = 0$$
$$\lim_{y \uparrow +\infty} F(y) = 1$$

All of these properties follow from continuity of probability and cannot be proved without the monotone convergence axiom.

Countable Additivity

If A_1, A_2, ... are disjoint (mutually exclusive) events, then

$$\Pr\left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \Pr(A_n)$$

This also follows from continuity of probability and cannot be proved without the monotone convergence axiom.

In advanced probability theory, this is taken as an axiom, called the *axiom of countable additivity* and monotone convergence is derived from it (so is called the *monotone convergence theorem*), but that requires a huge amount of work that is far beyond the scope of this course and also goes against our style of emphasizing axioms for expectation and treating probability as a special case of expectation.

Countable Additivity (cont.)

One says conventional advanced probability theory is *countably additive probability theory* to distinguish it from *finitely additive probability theory* which only allows finite additivity

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i),$$

which we derived from $E(X + Y) = E(X) + E(Y)$ by probability being expectation of indicator functions and mathematical induction.

Everything in this course up to now, except the properties of DF just reviewed, and the implications of variance zero holds in finitely additive probability theory.

Almost Surely

A logical expression involving random variables is said to hold *almost surely* if it holds for all outcomes s except in an event A such that $\Pr(A) = 0$.

If X is a nonnegative random variable, then $E(X) = 0$ if and only if $X = 0$ almost surely. Proofs both ways involve monotone convergence.

If $E(X) = 0$, then we can conclude by Markov's inequality that $\Pr(X \geq 1/n) = 0$ for any $n > 0$. The events

$$A_n = \{ s \in S : X(s) \geq 1/n \}$$

increase to

$$A = \{ s \in S : X(s) > 0 \}$$

so continuity of probability implies $\Pr(X > 0) = 0$.

Almost Surely (cont.)

If $X = 0$ almost surely, then the random variables defined by

$$X_n(s) = \begin{cases} X(s), & X(s) \leq n \\ n, & \text{otherwise} \end{cases}$$

have expectation

$$\begin{aligned} E(X_n) &\leq 0 \cdot E\{I_{\{0\}}(X_n)\} + n \cdot E\{I_{(0,n]}(X_n)\} \\ &= n \cdot \Pr(X > 0) \\ &= 0 \end{aligned}$$

and $X_n \uparrow X$, so $E(X) = 0$ by monotone convergence.

Almost Surely (cont.)

A very useful "sanity check" is a special case of this principle.

If X has first and second moments, then $\text{var}(X) = 0$ if and only if X is almost surely constant, in which case the constant is $E(X)$.

This follows from the monotone convergence axiom and cannot be proved without it.

Riemann and Lebesgue Integration

The kind of integration taught in first, second, and third year calculus does not conform to the monotone convergence axiom. In general it is not true that

$$0 \leq f_1(x) \leq f_2(x) \leq \cdots, \qquad \text{for all } x$$

and

$$\lim_{n \to \infty} f_n(x) = f(x)$$

implies

$$\lim_{n \to \infty} \int_S f_n(x)\, dx = \int_S f(x)\, dx \tag{$*$}$$

Although $(*)$ does hold whenever all the integrals are defined, the fact that all of the f_n are functions for which integrals are defined, does not imply that f is such a function.

Riemann and Lebesgue Integration (cont.)

It is a fact, but one that requires a huge amount of work that is far beyond the scope of this course to prove, that one can just take $(*)$ as a *definition* of the integral of f when f is a function whose integral is not defined in first, second, and third year calculus.

To distinguish, one says the definition of integration used in first, second, and third year calculus is *Riemann integration*, and the definition of integration via $(*)$ is *Lebesgue integration*. To further distinguish, fourth, fifth, etc. year calculus are called *real analysis* rather than calculus.

Riemann and Lebesgue Integration (cont.)

Lebesgue integration allows some very weird functions to be integrated.

Let $\{a_1, a_2, \ldots\}$ be an enumeration of the rational numbers in the interval $(0, 1)$, for example

$$\begin{aligned} a_1 &= 1/2 \\ a_2 &= 1/3 \\ a_3 &= 2/3 \\ a_4 &= 1/4 \\ a_5 &= 3/4 \\ &\vdots \end{aligned}$$

Riemann and Lebesgue Integration (cont.)

and let

$$A_n = \{a_1, \ldots, a_n\}$$

then

$$A_n \uparrow A$$

where A is the set of rational numbers between zero and one (exclusive).

Each I_{A_n} is Riemann integrable

$$\int_0^1 I_{A_n}(x)\,dx = 0$$

because I_{A_n} is nonzero only on a finite set.

Riemann and Lebesgue Integration (cont.)

Hence by monotone convergence I_A is Lebesgue integrable

$$\int_0^1 I_A(x)\,dx = 0$$

But you can't draw the graph of the function I_A. We have come a long way from the integral is "the area under the curve".

The Monotone Convergence Axiom (cont.)

If you don't remember anything from our introduction of the monotone convergence axiom to here except the continuity properties of DF and the sanity check that $\text{var}(X) = 0$ if and only if X is almost surely constant, that's all right.

The only reason we took this much class time about stuff that is really beyond the scope of this course is that you are liable to stumble over this stuff frequently if you read anything about probability theory except textbooks designed for courses at this level and even they often gratuitously drag in monotone convergence or countable additivity. So you have to know something about them just to avoid mystification. Hopefully, this is enough.

Existence of Moments

For any real number a and any positive integer k, the expectation $E\{(X-a)^k\}$, if it exists, is called the k*-th moment about the point* a.

For any real number a and any positive real number p, the expectation $E\{|X-a|^p\}$, if it exists, is called the p*-th absolute moment about the point* a.

By definition, the k-moment exists if and only if the k-th absolute moment exists.

If p is not an integer, then a^p only makes sense for positive a, and only p-th absolute moments make sense.

Existence of Moments (cont.)

If $0 < q \leq p < \infty$, then

$$\frac{|x - a|^q}{|x - b|^p} \to I_{\{0\}}(p - q), \qquad \text{as } x \to \infty \text{ or } x \to -\infty.$$

Hence there exists an r such that

$$|x - a|^q \leq 2|x - b|^p, \qquad |x| \geq r,$$

from which we conclude: if any p-th absolute moment exists, then all q-th moments exist for $0 < q \leq p$.

Conversely, if any q-th absolute moment fails to exist, then all p-th moments fail to exist for $q \leq p$.

Existence of Moments (cont.)

This means we can say "second moments exist" without specifying which one or bothering to mention that this also implies that first moments also exist and also p-th moments for $0 < p \leq 2$.

Conversely, we can say "second moments do not exist" without specifying which one or bothering to mention that this also implies that third, fourth, fifth, etc. moments do not exist either and also p-th moments for $2 \leq p < \infty$.

Existence of Moments (cont.)

Bounded random variables always have expectation. If

$$|g(x)| \leq c, \qquad \text{for all } x,$$

then

$$E\{|g(X)|\} \leq E(c) = c.$$

Existence of Moments (cont.)

If X and Y both have p-th moments, then so does $X + Y$.

Proof: Define

$$A = \{\, s \in S : |X(s)| \geq |Y(s)| \,\}.$$

Then

$$|X + Y| \leq 2I_A|X| + 2I_{A^c}|Y|$$

hence

$$\begin{aligned} E\{|X + Y|^p\} &\leq 2^p E\{I_A|X|^p\} + 2^p E\{I_{A^c}|Y|^p\} \\ &\leq 2^p E\{|X|^p\} + 2^p E\{|Y|^p\} \end{aligned}$$

By mathematical induction, if X_1, ..., X_n have p-th moments, then so does $X_1 + \cdots + X_n$.

Correlation

If X and Y are non-constant random variables, which implies $\text{sd}(X) > 0$ and $\text{sd}(Y) > 0$, then

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\text{sd}(X)\,\text{sd}(Y)}$$

is called the *correlation* of X and Y or the *correlation coefficient* of X and Y.

If either X or Y is a constant random variable, then $\text{cor}(X, Y)$ is undefined.

Correlation (cont.)

$$
\begin{aligned}
0 &\le \operatorname{var}\left(\frac{X}{\operatorname{sd}(X)} \pm \frac{Y}{\operatorname{sd}(Y)}\right) \\
&= \frac{\operatorname{var}(X)}{\operatorname{sd}(X)^2} \pm \frac{2\operatorname{cov}(X,Y)}{\operatorname{sd}(X)\operatorname{sd}(Y)} + \frac{\operatorname{var}(Y)}{\operatorname{sd}(Y)^2} \\
&= 2 \pm 2\operatorname{cor}(X,Y)
\end{aligned}
$$

from which we infer

$$
-1 \le \operatorname{cor}(X,Y) \le 1
$$

which we call the *correlation inequality*.

Correlation Matrix

The matrix with i, j component $\text{cor}(X_i, X_j)$ is called the *correlation matrix* of the random vector $(X_1, \ldots, X_n)$.

Note that the diagonal elements are

$$\frac{\text{cov}(X_i, X_i)}{\text{sd}(X_i)\,\text{sd}(X_i)} = \frac{\text{var}(X_i)}{\text{sd}(X_i)^2} = 1$$

If $\mathbf{M}$ is the variance matrix and $\mathbf{D}$ is a diagonal matrix having the same diagonal elements as $\mathbf{M}$, then the correlation matrix is $\mathbf{D}^{-1/2}\mathbf{M}\mathbf{D}^{-1/2}$, from which we see that a correlation matrix, like a variance matrix, is positive semidefinite.

Correlation Matrix (cont.)

The requirement that a correlation matrix be positive semidefinite is stronger than the correlation inequalities for its components.

In homework problem 4-5 we saw that if $(X_1, \ldots, X_n)$ is exchangeable, then when $i \neq j$

$$\operatorname{cov}(X_i, X_j) \geq -\frac{\operatorname{var}(X_i)}{n-1}$$

and since $\operatorname{sd}(X_i) = \operatorname{sd}(X_j)$ by exchangeability

$$\operatorname{cor}(X_i, X_j) \geq -\frac{1}{n-1}$$

unless $n = 2$ this is stronger than the correlation inequality.

Correlation (cont.)

$$
\begin{aligned}
0 &\le \operatorname{var}\left(Y - \operatorname{cor}(X,Y)\frac{\operatorname{sd}(Y)}{\operatorname{sd}(X)}X\right) \\
&= \operatorname{var}(Y) - 2\operatorname{cor}(X,Y)\frac{\operatorname{sd}(Y)}{\operatorname{sd}(X)}\operatorname{cov}(X,Y) \\
&\qquad + \operatorname{cor}(X,Y)^2\frac{\operatorname{sd}(Y)^2}{\operatorname{sd}(X)^2}\operatorname{var}(X) \\
&= \operatorname{var}(Y) - \operatorname{cor}(X,Y)^2\operatorname{var}(Y)
\end{aligned}
$$

Assuming $\operatorname{var}(Y) > 0$, we have $\operatorname{cor}(X,Y)^2 = 1$ if and only if

$$
Y - \operatorname{cor}(X,Y)\frac{\operatorname{sd}(Y)}{\operatorname{sd}(X)}X
$$

has variance zero hence is an almost surely constant random variable.

Correlation (cont.)

The constant must be

$$E(Y) - \text{cor}(X, Y) \frac{\text{sd}(Y)}{\text{sd}(X)} E(X)$$

Hence we have proved that $\text{cor}(X, Y)^2 = 1$ if and only if

$$Y = E(Y) + \text{cor}(X, Y) \frac{\text{sd}(Y)}{\text{sd}(X)} \bigl[X - E(X) \bigr]$$

almost surely. In short, the correlation of X and Y has the extreme values -1 or $+1$ if and only if Y is a linear function of X (and vice versa).

Correlation (cont.)

In intro statistics we teach

> *Correlation measures linear association. It does not measure nonlinear association.*

We just saw that maximal correlation implies perfect linear association.

Conversely, a long time ago we looked at the example where X is a nonconstant random variable whose distribution is symmetric about zero and $Y = X^2$. Then $\text{cor}(X, Y) = 0$ even though there is perfect, albeit nonlinear, association between the variables.