Stat 5101 Notes: Expectation

Charles J. Geyer

November 10, 2006

1 Properties of Expectation

Section 4.2 in DeGroot and Schervish is just a bit too short for my taste. This handout expands it a little.

1.1 Fundamental Properties

Property 1 (Absolute Values)**.** *A random variable X has expectation if and only if $|X|$ has expectation.*

This is true just by definition. We define expectations by sums or integrals only if they are *absolutely* summable or integrable (as the case may be).

Property 2 (Additivity)**.** *For any random variables X_1, X_2, ..., X_n*

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) \tag{1.1}$$

assuming the expectations on the right hand side exist.

Property 3 (Homogeneity)**.** *For any random variable X and any constant a*

$$E(aX) = aE(X) \tag{1.2}$$

assuming the expectation on the right hand side exists.

The preceding two properties are well-known properties of summation and integration (sums and constants come out). Hence they apply to expectations defined in terms of sums and integrals.

Property 4 (Constants)**.** *For any constant a*

$$E(a) = a. \tag{1.3}$$

This property is another convention. Taking the constant out of (1.3) gives $aE(1) = a$, so we only need to show $E(1) = 1$, but this holds by definition of p. f. or p. d. f. (they sum or integrate to one, as the case may be).

Property 5 (Monotonicity)**.** *For any random variables X and Y such that $X \leq Y$*

$$E(X) \leq E(Y) \tag{1.4}$$

assuming these expectations exist. For any random variables X and Y such that $|X| \leq |Y|$, if Y has expectation, then X has expectation.

The first part is another well-known property of summation and integration. The second part is true just by definition. The sum or integral defining $E(|X|)$ will converge if the one defining $E(|Y|)$ does.

1.2 Interpretation

Recall that any function of a random variable is a random variable. We haven't put any such "function" random variables in our statements of the fundamental properties, but such variables can be plugged in anywhere we see a random variable giving

$$E\left(\sum_{i=1}^{n} g_i(X)\right) = \sum_{i=1}^{n} E\{g_i(X)\}$$

for any functions g_i (assuming expectations exist) for the additivity property,

$$E\{ag(X)\} = aE\{g(X)\}$$

for any function g (assuming expectations exist) for the homogeneity property,

$$E\{g(X)\} = a$$

for the constant function defined by $g(x) = a$, for all x, for the constants property, and

$$E\{g(X)\} \leq E\{h(X)\}$$

for any functions g and h, such that $g(x) \leq h(x)$, for all x, for the monotonicity property.

1.3 Linear Functions

Corollary 1.1 (Linear Functions)**.** *For any random variable X and any constants a and b*

$$E(aX + b) = aE(X) + b \tag{1.5}$$

assuming the expectation on the right hand side exists.

Proof. This is just additivity (Property 2), homogeneity (Property 3), and constants (Property 4). □

Another way to write (1.5) is to define the linear function $g(x) = ax + b$. Then (1.5) says

$$E\{g(X)\} = g\bigl(E\{X\}\bigr). \tag{1.6}$$

In other words, we can take a *linear function* outside an expectation.

It is a very important "non-property" of expectation that (1.6) is *generally false* for nonlinear functions

$$E\{g(X)\} \neq g\bigl(E\{X\}\bigr).$$

Example 1.1 (Non-Property)**.**
Take

$$g(x) = \frac{1}{x}$$

and

$$f(x) = 2x, \qquad 0 < x < 1.$$

We show that in this case

$$E\left(\frac{1}{X}\right) \neq \frac{1}{E(X)} \tag{1.7}$$

First,

$$\begin{aligned} E(X) &= \int_0^1 x \cdot 2x \cdot dx \\ &= \frac{2x^3}{3}\bigg|_0^1 \\ &= \frac{2}{3} \end{aligned}$$

Second,

$$\begin{aligned} E\left(\frac{1}{X}\right) &= \int_0^1 \frac{1}{x} \cdot 2x \cdot dx \\ &= 2x \Big|_0^1 \\ &= 2 \end{aligned}$$

In fact, we can say more than "generally not equal." We present the following theorem without proof (see and advanced probability book for the proof, if you are interested).

A function g of one real variable is *convex* if it is continuous, twice differentiable except at isolated points, and $g''(x) \geq 0$ wherever the derivative exists.

Examples are $g(x) = x^2$, $g(x) = 1/x$ (the function in our example above), and $g(x) = |x|$.

Theorem 1.2 (Jensen's Inequality)**.** *If g is a convex function, then*

$$E\{g(X)\} \geq g\big(E\{X\}\big) \tag{1.8}$$

provided the expectation on the right hand side exists. The equality is strict unless X is a constant random variable.

Applying Jensen's inequality to $g(x) = 1/x$ gives

$$E\left(\frac{1}{X}\right) > \frac{1}{E(X)},$$

when X is a non-constant, positive-valued random variable, and that certainly agrees with the calculation in Example 1.1.

1.4 Probability is a Special Case of Expectation

Probability is expectation of indicator functions. For any event A

$$\Pr(A) = E(I_A) \tag{1.9}$$

Suppose X is a continuous random variable with p. d. f. f, then the right hand side of (1.9) is

$$\int I_A(x) f(x)\,dx = \int_A f(x)\,dx$$

because $I_A(x)$ is zero when $x \notin A$ (and so such points contribute nothing to the integral) and one otherwise. And the right hand side is what we have been taught is $\Pr(A)$.

When the variable in question is not the "original variable" things become a bit more complicated

$$\Pr(Y \in A) = E\{I_A(Y)\} \tag{1.10}$$

Again suppose X is a continuous random variable with p. d. f. f, and now suppose $Y = g(X)$, then the right hand side of (1.10) is

$$\int I_A[g(x)] f(x)\, dx = \int_{\{\, x : g(x) \in A \,\}} f(x)\, dx$$

Neither side of this equation is nice and clear. Both express the notion that the integration is over the set of points x such that $g(x) \in A$. There is no nice notation for that because it's just an inherently messy concept.

Example 1.2.
Suppose in the notation above

$$f(x) = \tfrac{3}{4}(1 - x^2), \qquad -1 < x < 1$$

and

$$g(x) = (x - \tfrac{1}{2})^2$$

and

$$A = (-\infty, \tfrac{1}{16})$$

So the probability we are trying to calculate is $\Pr(Y < \frac{1}{16})$, which because $Y = g(X)$ is the same as $\Pr\{g(X) < \frac{1}{16}\}$. That event occurs when

$$(x - \tfrac{1}{2})^2 < \tfrac{1}{16}$$

which is the same as

$$|x - \tfrac{1}{2}| < \tfrac{1}{4}$$

which is the same as

$$\tfrac{1}{4} < x < \tfrac{3}{4}$$

Now that we have decoded all the notation, the problem is straightforward

$$\begin{aligned}
\Pr(Y < \tfrac{1}{16}) &= \int_{1/4}^{3/4} f(x)\,dx \\
&= \tfrac{3}{4} \int_{1/4}^{3/4} (1 - x^2)\,dx \\
&= \frac{3}{4} \left[x - \frac{x^3}{3} \right]_{1/4}^{3/4} \\
&= \frac{35}{128}
\end{aligned}$$

2 Existence of Infinite Sums and Integrals

We will do integrals first. The sums we are interested are easily approximated by integrals.

2.1 Boundedness

The first principle of existence of integrals is that bounded functions integrated over bounded intervals always exist. If $-M \le g(x) \le M$ for $a < x < b$, then

$$-M(b-a) \le \int_a^b g(x)\,dx \le M(b-a)$$

by the monotonicity of integration and expectation of constants properties. Therefore problems about existence only arise when we integrate over an unbounded interval or when the integrand is unbounded.

An important way to establish that a function is bounded is the following theorem proved in advanced calculus books

Theorem 2.1. *A real-valued function of one real variable that is continuous on a closed interval* $[a, b]$ *is bounded on* $[a, b]$.

Note that the interval $[a, b]$ must be *closed*, meaning it includes the endpoints (as the square brackets indicate). The function $g(x) = 1/x$ is continuous on the open interval $(0, 1)$ but is not bounded on that interval (it goes to infinity as $x \to 0$).

2.2 The Magic Exponent -1

Theorem 2.2. *Suppose g is a nonnegative function of one real variable that is continuous on $[a, \infty)$ for some real number a. If*

$$\lim_{x \to \infty} \frac{g(x)}{x^\alpha} \tag{2.1}$$

is finite and nonzero, then

$$\int_a^\infty g(x)\,dx \tag{2.2}$$

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

If the limit (2.1) *is zero, then* (2.2) *exists if $\alpha < -1$ (but nothing can be said about nonexistence when $\alpha \geq -1$).*

Proof. The hypothesis about (2.1) and the definition of limit say that for any $\epsilon > 0$ there exists a b and L such that $a < b < \infty$ and $0 \leq L < \infty$ and

$$(L - \epsilon)x^\alpha \leq g(x) \leq (L + \epsilon)x^\alpha, \qquad x > b. \tag{2.3}$$

A continuous function is bounded on bounded intervals (Theorem 2.1), hence

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

exists. When $0 < \epsilon < L$ the sandwiching inequalities (2.3) show that

$$\int_b^\infty g(x)\,dx \tag{2.4a}$$

exists if and only if the integral

$$\int_b^\infty x^\alpha\,dx \tag{2.4b}$$

exists. When $L = 0$ the sandwiching inequalities (2.3) only show that (2.4a) exists if (2.4b) exists (but nothing about "only if").

So when does (2.4b) exist? There are two cases. First case, $\alpha \neq -1$. Then

$$\int_b^\infty x^\alpha\,dx = \lim_{c \to \infty} \int_b^c x^\alpha\,dx = \lim_{c \to \infty} \frac{c^{\alpha+1} - b^{\alpha-1}}{\alpha + 1}$$

and this limit is finite if $\alpha + 1 < 0$ and infinite if $\alpha + 1 > 0$, which agrees with the assertion of the theorem for these cases. Second case, $\alpha = -1$.

$$\int_b^\infty x^{-1}\,dx = \lim_{c \to \infty} \int_b^c x^{-1}\,dx = \lim_{c \to \infty} \log c - \log b$$

and this limit is infinite, which agrees with the assertion of the theorem for this case. And we are done. □

When the limit (2.1) in the hypothesis of the theorem is finite and nonzero, we say $g(x)$ "behaves like" x^α near infinity. In that case, $g(x)$ integrates if and only if $\alpha < -1$. When the limit (2.1) is zero, we say $g(x)$ goes to zero faster than x^α as x goes to infinity. In that case, $g(x)$ integrates if x^α does (but not necessarily vice versa).

What about integration to $-\infty$? The change of variable $y = -x$ makes the theorem apply to that case too. In short, what the theorem says about that case is that if $g(x)$ behaves like $|x|^\alpha$ at minus infinity, then it integrates if and only if $\alpha < -1$. The only difference is the absolute value in $|x|^\alpha$.

A student asked why the theorem only applies to nonnegative functions. That is because our theory is about *absolute integrability.* We say $\int g(x)\,dx$ exists if and only if $\int |g(x)|\,dx$ exists. So all questions about existence are about nonnegative integrands $|g(x)|$. The theorem is just written without explicit absolute value signs.

Example 2.1 (Cauchy Distribution)**.**
Is there a constant c such that the function

$$f(x) = \frac{c}{1+x^2}, \qquad -\infty < x < +\infty \tag{2.5}$$

is a probability density?

There are two ways to do this problem. The easy way, using Theorem 2.2 doesn't actually do any integrals or find the value of the constant. It only says that *some* constant works. Obviously (2.5) behaves like $|x|^{-2}$ as x goes to plus or minus infinity because the 1 in the denominator is negligible compared to x^2. Formally

$$\frac{f(x)}{|x|^{-2}} \to c, \qquad \text{as } |x| \to \infty.$$

Applying the theorem and the comment about minus infinity following it, we see that f has finite integral. Hence c can be chosen so that the integral is equal to one. That finishes the easy way.

Now for the hard way. To determine the actual value of the constant, you have to know the indefinite integral

$$\int \frac{dx}{1+x^2} = \tan^{-1} x + \text{constant}$$

and

$$\lim_{x \to \pm\infty} \tan^{-1} x = \pm\frac{\pi}{2}$$

so

$$\int_{-\infty}^{\infty} \frac{dx}{1 + x^2} = \pi.$$

Hence $c = 1/\pi$.

Example 2.2 (Moments of the Cauchy Distribution)**.**
Same density (2.5) as in the preceding example. Now that we know that it is a probability density, we can determine whether some expectations exist. For what positive β do $E(|X|^\beta)$ exist?

First write down the integral that defines this expectation, if it exists,

$$E(|X|^\beta) = \int_{-\infty}^{\infty} \frac{c|x|^\beta}{1 + x^2} \, dx$$

call that integrand something

$$g(x) = \frac{c|x|^\beta}{1 + x^2} \tag{2.6}$$

then we see that g behaves like $|x|^{\beta-2}$ as $|x|$ goes to infinity. Applying the theorem and the comment about minus infinity following it, we see that g has finite integral if and only if $\beta - 2 < -1$, which is when $\beta < +1$.

So that is our answer. $E(|X|^\beta)$ exists when $0 < \beta < 1$ and does not exist when $\beta \geq 1$.

Example 2.3 (A Horrible Distribution)**.**
Is there a constant c such that the function

$$f(x) = \frac{c}{1 + x^6 + |x|^{1/2} + \cos^2(x)}, \qquad -\infty < x < +\infty \tag{2.7}$$

is a probability density? And for what positive real numbers β does $E(|X|^\beta)$ exist?

Now there is only one way to do this problem. The easy way, using Theorem 2.2 doesn't actually do any integrals or find the value of the constant. It only says that *some* constant works. The hard way (trying to actually evaluate some integral) is now impossible. The formula was deliberately chosen to be just too messy to integrate.

Obviously (2.7) behaves like $|x|^{-6}$ as x goes to plus or minus infinity because the 1 in the denominator is negligible compared to x^6, so is $\cos(x)^2$ because $|\cos(x)| \leq 1$, and so is $|x|^{1/2}$ because smaller powers grow slower than larger.

Since $-6 < -1$, the theorem says $f(x)$ does integrate. Hence there exists a c that makes this a probability density even though we have no hope of calculating it, except numerically (for example in Mathematica)

```
In[1]:= h[x_] = 1 / (1 + x^6 + Abs[x]^(1 / 2) + Cos[x]^2)

                        1
Out[1]= -------------------------------
             6                        2
        1 + x  + Sqrt[Abs[x]] + Cos[x]

In[2]:= f[x_] = h[x] / NIntegrate[h[x], {x, -Infinity, Infinity}]

                    0.983593
Out[2]= -------------------------------
             6                        2
        1 + x  + Sqrt[Abs[x]] + Cos[x]

In[3]:= NIntegrate[f[x], {x, -Infinity, Infinity}]

Out[3]= 1.
```

Now obviously $|x|^\beta f(x)$ behaves like $|x|^{\beta-6}$ as x and integrates by the theorem if and only if $\beta - 6 < -1$, which is $\beta < 5$. Again the integrals are impossible to do except numerically (again in Mathematica continuing the calculation started above)

```
In[4]:= NIntegrate[x^2 f[x], {x, -Infinity, Infinity}]

Out[4]= 0.672384
```

So we can say that a random variable with density (2.7) has mean zero (by symmetry, see Section 3 below) and variance $E(X^2) = 0.672384$, and also has third and fourth, but no higher moments. But we can't do anything for this distribution by hand calculation except say which moments exist. But we can easily say which moments exist, which is the point of the example.

So that pretty much takes care of integrals over unbounded intervals. What about the other problem, unbounded integrands? This issue is much less important. If you have lost patience with this subject at this point, don't worry. You have already got the most important issue.

Theorem 2.3. *Suppose g is a nonnegative function of one real variable that is continuous on $[a, c]$ except at one point b. If*

$$\lim_{x \to b} \frac{g(x)}{|x - b|^\alpha} \tag{2.8}$$

is finite and nonzero, then

$$\int_a^c g(x)\,dx \tag{2.9}$$

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

If the limit (2.8) *is zero, then* (2.9) *exists if* $\alpha > -1$ *(but nothing can be said about nonexistence when* $\alpha \le -1$*).*

We won't go through the gory details of the proof. It works just like the proof of Theorem 2.2. We find that if $g(x)$ "behaves like" $|x - b|^\alpha$ near b, then $g(x)$ integrates if and only if $\alpha > -1$. And when $g(x)$ goes to infinity slower than $|x - b|^\alpha$ as x goes to b, then $g(x)$ integrates if x^α does (but not necessarily vice versa).

Example 2.4 (Moments of the Cauchy Distribution Continued)**.**
Now if we are going to consider negative β we also have to worry about the fact that $g(x)$ defined by (2.6) goes to infinity as x goes to zero. We see that

$$\lim_{x \to 0} \frac{g(x)}{|x|^\beta} = 1$$

so Theorem 2.3 says that

$$\int_{-a}^{a} g(x)\,dx$$

exists if and only if $\beta > -1$. We saw in Example 2.2 that integrals over $(-\infty, -a)$ and (a, ∞) exist whenever $\beta < +1$, so they are no problem when β is negative.

Hence, summarizing both Example 2.2 and this example, $E(|X|^\beta)$ exists for the distribution having density (2.5) if and only if $-1 < \beta < 1$.

Note that this does not include any nontrivial integer values (The only integer in this range is zero. We know $|X|^0$ is the constant random variable always equal to 1, because $y^0 = 1$ for all y by convention. $E(1) = 1$ does exist, but is trivial.)

Example 2.5 (Expectation of $1/X$)**.**
Suppose X is a continuous random variable with density f that is continuous at zero with $f(0) > 0$. Then $E(1/X)$ does not exist. Why? The integrand in

$$E\left(\frac{1}{X}\right) = \int_{-\infty}^{\infty} \frac{1}{x} \cdot f(x)\,dx$$

behaves like $|x|^{-1}$ at zero, hence the integral over any integral containing zero does not exist.

This example gives us another view of the "non-property" (1.7). This example shows us that when f is continuous with $f(0) > 0$, that the left hand side of (1.7) does not exist. In contrast, the right hand side does exist whenever $E(X)$ exists and is not equal to zero. The "non-property" holds in a very strong sense when one side of the "not equals" is undefined and the other is defined. Then the two sides are about as unequal as they can be.

2.3 Exponential Tails

One last fact completes our understanding of existence of integrals: e^{-x} goes to zero at infinity faster than any power of x. For positive x

$$x^\beta e^{-x} = \frac{x^\beta}{e^x} = \frac{x^\beta}{1 + x + x^2/2 + \cdots + x^k/k! + \cdots} \le \frac{x^\beta}{x^k/k!}$$

because dropping positive terms from the denominator makes the denominator smaller and the fraction larger.

Theorem 2.4. *Suppose g is a nonnegative function of one real variable that is continuous on $[a, \infty)$ for some real number a, and suppose*

$$\lim_{x \to \infty} \frac{g(x)}{x^\alpha e^{-x}} \tag{2.10}$$

is finite for some real number α. Then

$$\int_a^\infty g(x)\, dx \tag{2.11}$$

exists.

Note that now with the e^{-x} factor the value of α in the x^α factor is completely irrelevant.

Example 2.6 (Normal and Other Distributions)**.**
For any $\alpha > 0$ there exists a constant c such that

$$f(x) = c \exp\bigl(-|x|^\alpha\bigr), \qquad -\infty < x < +\infty$$

is a probability density. Moreover, if X is a random variable having density f, then $E(|X|^\beta)$ exists for all $\beta > -1$.

The case $\alpha = 2$ is called the "normal" distribution (Section 5.6 in DeGroot and Schervish). The case $\alpha = 1$ is called the "Laplace" or the "double exponential" distribution (mentioned on pp. 580-581 in DeGroot and Schervish).

2.4 Infinite Sums

Now that we know about integrals, sums are easy. First a sequence s_n cannot be unbounded on a bounded interval. A finite stretch of the sequence $s_i, s_{i+1}, \ldots, s_j$ is automatically bounded (among any finite set of numbers one is largest).

Hence we only have issues analogous to integrals over unbounded intervals $[a, \infty)$. And everything about such infinite sums is exactly analogous to integrals. We see this by just turning the infinite sum into an integral. For any series s_n define a step function

$$g(x) = \sum_{i=1}^{\infty} s_i I_{(i-1,i)}(x).$$

The graph of g is a step function that has height s_i on a step of width one. So the area under g is just the sum of the series

$$\sum_{i=1}^{\infty} s_i = \int_0^{\infty} g(x)\, dx. \tag{2.12}$$

From this analysis we see that the rules for infinite series are just the same as for integrals. If

$$\lim_{n \to \infty} \frac{s_n}{n^\alpha}$$

is finite and nonzero, then the infinite sum (2.12) exists if and only if $\alpha < -1$. And if

$$\lim_{n \to \infty} \frac{s_n}{n^\alpha e^{-n}}$$

is finite, then the infinite sum (2.12) exists regardless of the value of α.

2.5 Existence of Moments

Theorem 2.5. *If $|X - a|^\alpha$ has expectation for any $a \in \mathbb{R}$ and any $\alpha > 0$, then $|X - b|^\beta$ has expectation for and $b \in \mathbb{R}$ and any β such that $0 \le \beta \le \alpha$.*

This theorem is used so much (often without comment) that it is worth describing more wordily.

- We can say "moments of order α exist" without specifying which kind (ordinary, absolute, central, absolute central). Either all moments of order α exist or none exist.

- Existence of moments of order α implies existence of moments of all lower orders (of order β for $0 \le \beta \le \alpha$).

For example, existence of *variance* (a second moment) implies existence of the *mean* (a first moment). We will often say something like "assume X is a random variable with finite variance" leaving it unsaid that the mean of X exists.

Proof. The function $g(x) = |x - a|^\alpha$ is continuous, so we only need to worry about summation or integration, as the case may be, over unbounded intervals. By the trick explained in Section 2.4 above, we can turn the sums into integrals and collapse both cases into one. Now

$$\lim_{|x| \to \infty} \frac{|x - b|^\alpha}{|x - a|^\alpha} = 1$$

regardless of the values of a and b. So by the same argument used in the proof of Theorem 2.2 for any $\epsilon > 0$ there exists a constant c such that

$$(1 - \epsilon)|x - a|^\alpha \le |x - b|^\alpha \le (1 + \epsilon)|x - a|^\alpha, \qquad |x| \ge c$$

from which it is clear (choosing $\epsilon < 1$) that both of $E(|X - a|^\alpha)$ and $E(|X - b|^\alpha)$ exist or neither exists.

If $0 \le \beta < \alpha$, then

$$\lim_{|x| \to \infty} \frac{|x - b|^\beta}{|x - a|^\alpha} = 0$$

regardless of the values of a and b. So by the same argument used above for any $\epsilon > 0$ there exists a constant c such that

$$-\epsilon|x - a|^\alpha \le |x - b|^\beta \le \epsilon|x - a|^\alpha, \qquad |x| \ge c$$

from which it is clear that $E(|X - b|^\beta)$ exists if $E(|X - a|^\alpha)$ exists. □

3 Symmetry

We say random variables are *equal in distribution*, written

$$X \stackrel{\mathcal{D}}{=} Y.$$

if they both have the same probability distribution (same p. f., same p. d. .f, same d. f., whatever, however you wish to specify the model).

Note that means that all expectations involving these random variables must be the same

$$E\{g(X)\} = E\{g(Y)\}$$

holds for all real-valued functions g such that the expectations exist and either both expectations exist or neither expectation exists. This is because *expectations depend only on distributions not on the variables having those distributions.*

A random variable X is said to be *symmetric about zero* if X and $-X$ are equal in distribution. More generally, a random variable Y is said to be *symmetric about a point* a if $X = Y - a$ is symmetric about zero, that is, if

$$Y - a \stackrel{\mathcal{D}}{=} -(Y - a). \tag{3.1}$$

The point a is called the *center of symmetry* of Y.

Some of the most interesting probability models we will meet later involve symmetric random variables, hence the following theorem is very useful.

Theorem 3.1. *Suppose a random variable Y is symmetric about the point a. If the mean of Y exists, it is equal to a. Every higher odd integer central moment of Y that exists is zero.*

In notation, the two assertions of the theorem are

$$E(Y) = \mu = a$$

and

$$E\{(Y - \mu)^{2k-1}\} = 0, \qquad \text{for any positive integer } k.$$

The theorem doesn't say anything about even integer central moments. Those we have to calculate by doing the sums or integrals, as the case may be, but the theorem does save a lot of needless work calculating sums or integrals for the moments it does say something about.

Proof. The hypothesis of the theorem is that (3.1) holds. Hence

$$E\{(Y - a)^k\} = E\{[-(Y - a)]^k\} = (-1)^k E\{(Y - a)^k\} \tag{3.2}$$

holds for all integers k such that the expectations exist. For even k equation (3.2) tells us nothing since $(-1)^k = 1$ makes both sides trivially the same. But for odd k, since $(-1)^k = -1$, equation (3.2) implies

$$E\{(Y - a)^k\} = -E\{(Y - a)^k\}$$

and the only number that is equal to its negative is zero. Thus we have shown

$$E\{(Y - a)^k\} = 0, \qquad \text{for odd positive integers } k. \tag{3.3}$$

In particular, the $k = 1$ case gives

$$0 = E(Y - a) = E(Y) - a$$

so $\mu = E(Y) = a$. This proves the first assertion of the theorem about $E(Y)$ and plugging $a = \mu$ into (3.3) proves the rest. □

Example 3.1 (Symmetric Binomial Distribution)**.**
The Binomial$(n, 1/2)$ distribution is symmetric about the point $n/2$. This is obvious from the definition. If we swap the meaning of "success" and "failure" the probabilities don't change. This means the Y successes have the same distribution as the $n - Y$ failures, that is,

$$Y \stackrel{\mathcal{D}}{=} n - Y.$$

Subtracting $n/2$ from both sides gives

$$Y - \frac{n}{2} \stackrel{\mathcal{D}}{=} \frac{n}{2} - Y = -\left(Y - \frac{n}{2}\right)$$

and that's symmetry about $n/2$.

Conclusion from the theorem $E(Y) = n/2$ and every odd central moment is zero.

Warning: The binomial distribution is **not** symmetric if the success probability is not 1/2. Swapping "success" and "failure" only leaves the distribution unchanged if the probabilities of success and failure are equal.