Section 1.9. Some Special Expectations

Note. In this section we use the expectation operator to define the mean, variance, and moment generating function of a random variable.

Definition 1.9.1. Let X be a random variable whose expectation exists. The *mean value* μ of X is defined as $\mu = E[X]$.

Definition 1.9.2. Let X be a random variable with finite mean μ and such that $E[(X-\mu)^2]$ is finite. Then the *variance* of X is $E[(X-\mu)^2]$. The variance is commonly denoted σ^2 on $\mathrm{Var}(X)$. The *standard deviation* of X is $\sigma = \sqrt{\sigma^2} = \sqrt{E[(X-\mu)^2]}$.

Note 1.9.A. We have by the linearity of E (given in Theorem 1.8.2) that

$$\sigma^2 = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + E[\mu^2]$$

$$= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2.$$

Note. The variance "operator" is not linear, but we have the following result related to a special transformation of X.

Theorem 1.9.1. Let X be a random variable with finite mean μ and finite variance σ^2. Then for all constants a and b we have $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

Example. Exercise 1.9.3(a). Let X have distribution $f(x) = 6x(1-x) = 6x - 6x^2$, $0 < x < 1$, zero elsewhere. Compute $P(\mu - 2\sigma < X < \mu + 2\sigma)$.

Definition 1.9.3. Let X be a random variable such that for some $h > 0$, the expectation of e^{tX} exists for $-h < t < h$. The *moment generating function* (or *mgf*) of X is the function $M(t) = E[e^{tX}]$ for $-h < t < h$.

Note. When a moment generating function exists, we must have for $t = 0$ that $M(0) = E[e^{0X}] = E[1] = 1$.

Example. Exercise 1.9.7. Show that the moment generating function of the random variable X having the probability density function $f(x) = 1/3$, $-1 < x < 2$, zero elsewhere, is

$$M(t) = \begin{cases} \frac{e^{2t} - e^{-t}}{3t} & \text{for } t \neq 0 \\ 1 & \text{for } t = 0. \end{cases}$$

Note. The next theorem (the proof of which "is beyond the scope of this text") shows that a distribution that has a moment generating function M is completely determined by M.

Theorem 1.9.2. Let X and Y be random variables with moment generating functions M_X and M_Y, respectively, existing in open intervals about 0. Then $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$ if and only if $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$ for some $h > 0$.

Example 1.9.A. We now illustrate Theorem 1.9.2 with a continuous example. Suppose $M(t) = 1/(1-t)$, $t < 1$. We observe (since the book gives this information...we'll give more information below concerning this) that with $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere, then

$$\int_{-\infty}^{\infty} e^{tx} f(x)\, dx = \int_{0}^{\infty} e^{tx} f(x)\, dx = \int_{0}^{\infty} e^{(t-1)x}\, dx$$

$$= \frac{1}{t-1} e^{(t-1)x}\Big|_0^\infty = \lim_{b\to\infty} \frac{1}{t-1}(e^{(t-1)b} - 1) = \frac{1}{1-t}.$$

Since, by Theorem 1.9.2, f is uniquely determined by M then we must have $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere.

Note 1.9.B. In analysis (see Proposition IV.2.1 of my online notes for Complex Analysis [MATH 5510] on IV.2. Power Series Representation of Analytic Functions) we have that if $\varphi(s,t)$ is continuous on $[a,b] \times [c,d]$ (and real valued for us) and if $g(t) = \int_a^b \varphi(s,t)\, ds$ then g is continuously differentiable on $[c,d]$ and $g'(t) = \displaystyle\int_a^b \frac{\partial \varphi}{\partial t}(s,t)\, ds$ (this is sometimes called Liebniz's Rule). We are interested in differentiating $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$, so the only hypothesis we need is that the probability density function f is continuous. We then have

$$M'(t) = \frac{d}{dt}\left[\int_{-\infty}^{\infty} e^{tx} f(x)\, dx\right] = \int_{-\infty}^{\infty} \frac{\partial}{\partial t}[e^{tx} f(x)]\, dx = \int_{-\infty}^{\infty} x e^{tx} f(x)\, dx;$$

in the discrete case, we have

$$M'(t) = \frac{d}{dt}[M(t)] = \frac{d}{dt}\left[\sum_x e^{tx} p(x)\right] = \sum_x x e^{tx} p(x).$$

With $t = 0$, we have $M'(0) = \mu$ in both the discrete and continuous case. Next,

$$M''(t) = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)\, dx \text{ or } M''(t) = \sum_x x^2 e^{tx} p(x)$$

so that $M''(0) = E[X^2]$ and

$$\mathrm{Var}(X) = \sigma^2 = E[X^2] - \mu^2 = M''(0) - (M'(0))^2.$$

In general, for $m \in \mathbb{N}$ we have

$$M^{(m)}(t) = \int_{-\infty}^{\infty} x^m e^{tx} f(x)\, dx \text{ or } M^{(m)}(t) = \sum_x x^m e^{tx} p(x)$$

so that $M^{(m)}(0) = E[X^m]$.

Definition. For X a random variable and $m \in \mathbb{N}$, the *mth moment* of the distribution (of X) is $E[X^m]$.

Note. In mechanics, the term "moment" is used when dealing with "twisting forces." See my online notes for Applied Mechanics 1 (Statics) (formerly MATH 2610, but now "Statics" [CEE 2110] in the TTU-ETSU dual degree program in engineering) 4.1. Two-Dimensional Description of the Moment, 4.2. The Moment Vector, 4.3. Moment of a Force About a Line, Chapter 8. Moments of Inertia, my Calculus 2 online notes on 6.6. Moments and Centers of Mass (in two dimensions), and my Calculus 3 online notes on 15.6. Moments and Centers of Mass (in three dimensions). The term "moment" used in the Statics and Calculus notes corresponds to our use of the term when $m = 1$, and the use of the term "moment of inertia" in the Statics and Calculus notes corresponds to our use of the term when $m = 2$. *This* is the reason the term "moment" is as defined previously for us.

Example 1.9.6. Let X be a continuous random variable with probability density function $f(x) = \frac{1}{\pi}\frac{1}{x^2+1}$ for $x \in \mathbb{R}$. This is the Cauchy probability density function introduced in Exercise 1.7.24. Let $t > 0$ be given. If $x > 0$ then by the Mean Value Theorem applied to e^x on the interval $[0, tx]$, there is some $0 < \xi_0 < tx$ such that

$$\frac{e^{tx}-1}{tx-0} = \frac{e^{tx}-1}{tx} = e^{\xi_0} \geq 1.$$

So $e^{tx} \geq 1 + tx \geq tx$. So we have

$$M(t) = \int_{-\infty}^{\infty} e^t \frac{1}{\pi}\frac{1}{x^2+1}\,dx \geq \int_0^{\infty} e^{tx}\frac{1}{\pi}\frac{1}{x^2+1}\,dx$$

$$\geq \int_0^{\infty} \frac{1}{\pi}\frac{tx}{x^2+1}\,dx = \lim_{b\to\infty} \frac{t}{2\pi}\ln(x^2+1)\Big|_0^b = \infty.$$

Since $t > 0$ is arbitrary, the integral does not exist in an open interval of 0. So the moment generating function of the Cauchy distribution does not exist.

Example 1.9.7. Let X have the moment generating function $M(t) = e^{t^2/2}$, $-\infty < t < \infty$. We'll see in Chapter 3 that this is the moment generating function for the standard normal distribution. The Maclaurin series for $e^{t^2/2}$ is

$$M(t) = e^{t^2/2} = \sum_{k=0}^{\infty} \frac{(t/2)^k}{k!} = \sum_{k=0}^{\infty} \frac{t^{2k}}{2^k k!} = \sum_{\ell=0}^{\infty} \frac{M^{(\ell)}(0)}{\ell!} t^\ell.$$

So we see that for ℓ odd, we have $M^{(\ell)}(0) = 0$ and for ℓ even (say $\ell = 2k$) we have $\frac{M^{(2k)}(0)}{(2k)!} = \frac{1}{2^k k!}$ or $M^{(2k)}(0) = \frac{(2k)!}{2^k k!}$. From this we have the moments

$$M^{(2k-1)}(0) = E[X^{2k-1}] = 0 \text{ for } k \in \mathbb{N},$$

$$M^{(2k)}(0) = E[X^{2k}] = \frac{(2k)!}{2^k k!} \text{ for } k \in \mathbb{N}.$$

Note. A function f such that $\int_{-\infty}^{\infty} |f(x)|\,dx < \infty$ is said to be in $L^1(\mathbb{R})$. The Fourier transform of $f \in L^1(\mathbb{R})$ is

$$\mathscr{F}(f) = \hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx} f(x)\,dx.$$

Since we consider f a probability distribution function, then $\int_{-\infty}^{\infty} |f(x)|\,dx = 1$ and so the Fourier transform of all probability density functions is defined. However, we are interested now in the inverse Fourier transform which is defined for all $f \in L^2(\mathbb{R})$ where

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{R} \,\middle|\, \int_{-\infty}^{\infty} |f(x)|^2\,dx < \infty \right\}.$$

For $f \in L^2(\mathbb{R})$, we define the Fourier transform $\hat{f} = \mathscr{F}\{f\}$ using functions in $L^1(\mathbb{R})$ (see Definition 4.11.2 in my online notes for Applied Math 1 [MATH 5610] on 4.11. The Fourier Transform) and we find that $\hat{f} \in L^2(\mathbb{R})$. We then define for $\hat{f} \in L^2(\mathbb{R})$, the inverse Fourier transform $\mathscr{F}^{-1}$ as

$$\mathscr{F}^{-1}\{\hat{f}(k)\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} \hat{f}(k)\,dk.$$

We then have that $\mathscr{F}$ and $\mathscr{F}^{-1}$ are one to one and onto mappings from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$ ($\mathscr{F}$ and $\mathscr{F}^{-1}$ are, in fact, "Hilbert space isomorphisms" of $L^2(\mathbb{R})$ with $L^2(\mathbb{R})$). This is all described in some detail in my online notes for Applied Math 1 on 4.11. The Fourier Transform.

Definition. Let f be a probability density function for continuous random variable X and suppose $f \in L^2(\mathbb{R})$. The function

$$\varphi(t) = M(it) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx = \mathscr{F}^{-1}\{f\}.$$

Then φ is the *characteristic function* of the distribution of X.

Note. In the text's Examples 1.9.5 and 1.9.6 it is shown that not every distribution has a moment generating function. However, Hogg, McKean, and Craig claim on page 74 that *every* distribution has a characteristic function. This is a bit suspect since $\mathcal{F}^{-1}$ is only necessarily defined for $f \in L^2(\mathbb{R})$ and all we know is that a probability density function f is in $L^1(\mathbb{R})$. For example,

$$f(x) = \begin{cases} \frac{1}{2}\left(\frac{x^{-1/2}}{1+|\ln x|}\right)^2 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

is in $L^1(\mathbb{R})$ (the integral is 1 so that f is a pdf) but it is not in $L^2(\mathbb{R})$ (see Exercise 7.2.7(b) in Royden and Fitzpatrick's *Real Analysis*, 4th Edition).

Note. Hogg, McKean, and Craig also claim (page 75): "Every distribution has a unique characteristic function; and to each characteristic function there corresponds a unique distribution of probability." Since $\mathscr{F}$ and $\mathscr{F}^{-1}$ are Hilbert space isomorphisms of $L^2(\mathbb{R})$ with itself, then this *is* true for f and $\hat{f}$ in $L^2(\mathbb{R})$... but for f or $\hat{f}$ in $L^1(\mathbb{R})$? In Robert B. Ash's (with contributions from Catherine Doleans-Dade) *Probability and Measure Theory*, 2nd Edition (Academic Press, 2000), the following definition is given.

> **Definition 7.1.1 of Ash.** Let μ be a finite measure on the real Borel sets. The *characteristic function* of μ is the mapping from $\mathbb{R}$ to $\mathbb{C}$ given by $h(u) = \int_{\mathbb{R}} e^{iux}\,d\mu(x)$ where $u \in \mathbb{R}$. Thus h is the Fourier transform of μ. If F is a distribution function corresponding to μ (in which case we write $h(u) = \int_{\mathbb{R}} e^{iux}\,dF(x)$) then h is the *characteristic function of distribution* F, or the *characteristic function of random variable* X if X has distribution function F.

A "finite measure" is a measure for which the universal set (the real numbers, here) has finite measure. So Ash's rigorous presentation of probability also imposes a kind of boundedness. We move on, a bit concerned about the existence of characteristic functions. However, in this class we'll encounter characteristic functions one more time, in Section 5.3. Central Limit Theorem when we give a partial proof of the Central Limit Theorem using moment generating functions (which we have seen do not always exist for a given distribution) and refer to a more advanced course for a proof of the Central Limit Theorem based on characteristic functions (which always exist…we suppose!). The "more advanced course" would be a measure theory based probability class, such as one based on Ash's *Probability and Measure Theory*.

Note. As a final comment, we explain *why* we chose $f(x) = e^{-x}$ in Example 1.9.A when considering the equation

$$\frac{1}{1-t} = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx,\ t < 1.$$

The Laplace transform of f is

$$\mathscr{L}\{f(t)\} = \int_0^{\infty} e^{-st} f(s)\, ds.$$

See my online notes on "A Second Course in Differential Equations" (not an official ETSU class, but possibly corresponding to some of the material covered in Introduction to Applied Math [MATH 4027/5027]) on 6.1. Definition of the Laplace Transform. So to find f in $\int_{-\infty}^{\infty} e^{tx} f(x)\, dx$, where $t < 1$ and $f(x) = 0$ for $x \leq 0$, we need to find f such that $\mathscr{L}\{f(t)\} = \int_0^{\infty} e^{-tx} f(x)\, dx$, $t > -1$. That is (in our example) we need

$$f(t) = \mathscr{L}^{-1}\left\{\int_0^\infty e^{-tx} f(x)\,dx\right\} = \mathscr{L}^{-1}\left\{\frac{1}{1-(-t)}\right\} = \mathscr{L}^{-1}\left\{\frac{1}{1+t}\right\}.$$

From the theory of Laplace transforms, we have $\mathscr{L}^{-1}\left\{\frac{1}{1+t}\right\} = e^{-t}$ (see Example Page 289 Number 12 in my online notes on 6.2. Solutions of Initial Value Problems). This is where the choice of $f(x) = e^{-x}$ comes from in Example 1.9.A.

Revised: 7/2/2021