6 SUMMARISING CONTINUOUS RANDOM VARIABLES

In this chapter we shall revisit concepts that have been discussed for discrete random variables and see their analogues in the continuous setting. We then introduce generating functions and conclude this chapter with a discussion on bivariate normal random variables.

6.1 EXPECTATION, AND VARIANCE

The notion of expected value carries over from discrete to continuous random variables, but instead of being described in terms of sums, it is defined in terms of integrals.

DEFINITION 6.1.1. *Let X be a continuous random variable with piecewise continuous density $f(x)$. Then the expected value of X is given by*

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx.$$

provided that the integral converges absolutely[1]*. In this case we say that X has "finite expectation". If the integral diverges to $\pm\infty$ we say the random variable has infinite expectation. If the integral diverges, but not to $\pm\infty$ we say the expected value is undefined.*

The next three examples illustrate the three posibilities: the first is an example where expectation exists as a real number; the next is an example of an infinite expected value; and the final example shows that the expected value may not be defined at all.

EXAMPLE 6.1.2. Let $X \sim \text{Uniform}(a, b)$. Then the expected value of X is given by

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_{a}^{b} x \cdot \frac{1}{b-a}\, dx = \frac{1}{2(b-a)}(b^2 - a^2) = \frac{b+a}{2}.$$

This result is intuitive since it says that the average value of a $\text{Uniform}(a, b)$ random variable is the midpoint of its interval. ■

EXAMPLE 6.1.3. Let $0 < \alpha < 1$ and $X \sim \text{Pareto}(\alpha)$ which is defined to have the probability density function

$$f(x) = \begin{cases} \frac{\alpha}{x^{\alpha+1}} & 1 \le x < \infty \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \int_{1}^{\infty} x \cdot \frac{\alpha}{x^{\alpha+1}}\, dx = \alpha \lim_{M\to\infty} \int_{1}^{M} x^{-\alpha} dx = \frac{\alpha}{-\alpha+1}(-1 + \lim_{M\to\infty} M^{-\alpha+1}) = \infty$$

as $0 < \alpha < 1$.

Thus this Pareto random variable has an infinite expected value. ■

EXAMPLE 6.1.4. Let $X \sim \text{Cauchy}(0,1)$. Then the probability density function of X is given by

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2} \text{ for all } x \in \mathbb{R}.$$

Now,

$$E[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\pi(1+x^2)}\, dx$$

Now by Exercise 6.1.10, we know that as $M \to -\infty, N \to \infty$ the $\int_M^N \frac{x}{1+x^2} dx$ does not converge or diverge to $\pm\infty$. So $E[X]$ is not defined for this Cauchy random variable. ■

Expected values of functions of continuous random variables may be computed using their respective probability density function by the following theorem.

THEOREM 6.1.5. *Let X be continuous random variables with probability density function $f_X : \mathbb{R} \to \mathbb{R}$.*

(a) Let $g : \mathbb{R} \to \mathbb{R}$ be piecewise continuous and $Z = g(X)$ Then the expected value of Z given by

$$E[g(X)] \quad = \quad \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

(b) Let Y be a continuous random variable such that (X, Y) have a joint probability density function $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose $h : \mathbb{R}^2 \to \mathbb{R}$ be piecewise continuous. Then,

$$E[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\, dx\, dy.$$

Proof- The proof is beyond the scope of this book. For (a) when g is as in Exercise 5.3.10 then one can provide the proof using only the tools of basic calculus (we will leave this case as an exercise to the reader) ■

We illustrate the use of the above theorem with a couple of examples.

EXAMPLE 6.1.6. A piece of equipment breaks down after a functional lifetime that is a random variable $T \sim Exp(\frac{1}{5})$. An insurance policy purchased on the equipment pays a dollar amount equal to $1000 - 200t$ if the equipment breaks down at a time $0 \le t \le 5$ and pays nothing if the equipment breaks down after time $t = 5$. What is the expected payment of the insurance policy?

For $t \ge 0$ the policy pays $g(t) = \max\{1000 - 200t, 0\}$ so,

$$\begin{aligned} E[g(T)] &= \int_0^{\infty} \frac{1}{5} e^{(1/5)t} \max\{1000 - 200t, 0\}\, dt \\ &= \int_0^{5} \frac{1}{5} e^{(1/5)t}(1000 - 200t)\, dt \\ &= 1000e^{-1} \approx \$367.88 \end{aligned}$$

■

EXAMPLE 6.1.7. Let $X, Y \sim \text{Uniform}(0,1)$. What is the expected value of the larger of the two variables?

We offer two methods of solving this problem. The first is to define $Z = \max\{X, Y\}$ and then determine the density of Z. To do so, we first find its distribution. $F_Z(z) = P(Z \le z)$, but

$\max\{X, Y\}$ is less than or equal to z exactly when both X and Y are less than or equal to z. So for $0 \leq z \leq 1$,

$$\begin{aligned} F_Z(z) &= P((X \leq z) \cap (Y \leq z)) \\ &= P(X \leq z) \cdot P(Y \leq z) \\ &= z^2 \end{aligned}$$

Therefore $f_Z(z) = F_Z'(z) = 2z$ after which the expected value can be obtained through integration

$$E[Z] = \int_0^1 z \cdot 2z \, dz = \frac{2}{3} z^3 \mid_0^1 = \frac{2}{3}.$$

An alternative method is to use Theorem 6.1.5 (b) to calculate the expectation directly without finding a new density. Since X and Y are independent, their joint distribution is the product of their marginal distributions. That is,

$$f(x, y) = f_X(x) f_Y(y) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Therefore,

$$\begin{aligned} E[\max\{X, Y\}] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \max\{x, y\} \cdot f(x, y) \, dx \, dy \\ &= \int_0^1 \int_0^1 \max\{x, y\} \cdot 1 \, dx \, dy \end{aligned}$$

The value of $\max\{x, y\}$ is x if $0 < y \leq x < 1$ and it is y if $0 < x \leq y < 1$. So,

$$\begin{aligned} E[\max\{X, Y\}] &= \int_0^1 \int_0^y y \, dx \, dy + \int_0^1 \int_y^1 x \, dx \, dy \\ &= \int_0^1 xy \mid_{x=0}^{x=y} \, dy + \int_0^1 \frac{1}{2} x^2 \mid_{x=y}^{x=1} \, dy \\ &= \int_0^1 y^2 \, dy + \int_0^1 \frac{1}{2} - \frac{1}{2} y^2 \, dy \\ &= \frac{1}{3} + \frac{1}{3} = \frac{2}{3}. \end{aligned}$$

■

Results from calculus may be used to show that the linearity properties from Theorem 4.1.7 such as apply to continuous random variables as well as to discrete ones. We restate it here for completeness.

THEOREM 6.1.8. *Suppose that X and Y are continuous random variables with piecewise continuous joint density function function $f : \mathbb{R}^2 \to \mathbb{R}$. Assume that both have finite expected value. If a and b are real numbers then*

(a) $E[aX] = aE[X]$;

(b) $E[aX + b] = aE[X] + b$

(c) $E[X + Y] = E[X] + E[Y]$; *and*

(d) $E[aX + bY] = aE[X] + bE[Y]$.

(e) *If* $X \geq 0$ *then* $E[X] \geq 0$.

Proof- See Exercise 6.1.11. ■

We will use these now-familiar properties in the continuous setting. As in the discrete setting we can define the variance and standard deviation of a continuous random variable.

DEFINITION 6.1.9. *Let X be a random variable with probability density function $f : \mathbb{R} \to \mathbb{R}$. Suppose X has finite expectation. Then*

(a) the variance of the random variable is written as $Var[X]$ and is defined as

$$Var[X] = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)dx,$$

(b) the standard deviation of X is written as $SD[X]$ and is defined as

$$SD[X] = \sqrt{Var[X]}$$

Since the above terms are expected values, there is the possibility that they may be infinite because the integral describing the expectation diverges to infinity. As the integrand is strictly positive, it isn't possible for the integral to diverge unless it diverges to infinity.

The properties of variance and standard deviation of continuous random variables match those of their discrete counterparts. A list of these properties follows below.

THEOREM 6.1.10. *Let $a \in \mathbb{R}$ and let X be a continuous random variable with finite variance (and thus, with finite expected value as well). Then,*

(a) $Var[X] = E[X^2] - (E[X])^2$.

(b) $Var[aX] = a^2 \cdot Var[X]$;

(c) $SD[aX] = |a| \cdot SD[X]$;

(d) $Var[X + a] = Var[X]$; *and*

(e) $SD[X + a] = SD[X]$.

If Y is another independent continuous random variable with finite variance (and thus, with finite expected value as well) then

(f) $E[XY] = E[X]E[Y]$;

(g) $Var[X + Y] = Var[X] + Var[Y]$; *and*

(h) $SD[X + Y] = \sqrt{(SD[X])^2 + (SD[Y])^2}$.

Proof- The proof is essentially an imitation of the proofs presented in Theorem 4.1.10, Theorem 4.2.5, Theorem 4.2.4, and Theorem 4.2.6. One needs to use the respective densities, integrals in lieu of sums, and use Theorem 6.1.11 and Theorem 6.1.5 when needed. We will leave this as an exercise to the reader. ■

EXAMPLE 6.1.11. Let $X \sim$ Normal $(0,1)$. In this example we shall show that $E[X] = 0$ and $Var[X] = 1$. Before that we collect some facts about the probability density function of X, given by (5.2.7). Using (5.2.9) with $z = 0$, we can conclude that

$$\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{2} \tag{6.1.1}$$

Observe that there exists $c_1 > 0$ such that

$$\max\{| x |, x^2\} e^{-\frac{x^2}{2}} \leq c_1 e^{-c_1|x|}$$

for all $x \in \mathbb{R}$. Hence

$$\begin{aligned} \int_{-\infty}^\infty | x | \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \leq c_1 \int_{-\infty}^\infty e^{-c_1|x|} < \infty \\ \int_{-\infty}^\infty x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \leq c_1 \int_{-\infty}^\infty e^{-c_1|x|} < \infty \end{aligned} \tag{6.1.2}$$

Using the above we see that

$$E[X] = \int_{-\infty}^\infty x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx < \infty$$

So we can split integral expression in definition of $E[X]$ as

$$E[X] = \int_{-\infty}^0 x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + \int_0^\infty x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Further the change of variable $y = -x$ will imply that

$$\int_{-\infty}^0 x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = -\int_0^\infty y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

So $E[X] = 0$. Again by (6.1.2),

$$Var[X] = \int_{-\infty}^\infty (x - E[X])^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^\infty x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx < \infty$$

To evaluate the integral we make a change of variable to obtain

$$\int_{-\infty}^\infty x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \int_{-\infty}^0 x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + \int_0^\infty x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2 \int_0^\infty x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Then we use integration by parts like Lemma 5.5.4. Set $u(x) = x$ and $v(x) = e^{-\frac{x^2}{2}}$, which imply $u'(x) = 1$ and $v'(x) = -xe^{-\frac{x^2}{2}}$. Therefore for $a > 0$,

$$\begin{aligned} \int_0^a x^2 e^{-\frac{x^2}{2}} dx &= \int_0^a u(x)(-v'(x)) dx = u(x)(-v(x)) \mid_0^a - \int_0^a u'(x)(-v(x)) dx \\ &= a^2 e^{-\frac{a^2}{2}} + \int_0^a e^{-\frac{x^2}{2}} \end{aligned}$$

Using the fact that $\lim_{a\to\infty} a^2 e^{-\frac{a^2}{2}} = 0$ and (6.1.1) we have

$$\begin{aligned} Var[X] &= 2\frac{1}{\sqrt{2\pi}} \int_0^\infty x^2 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{\pi}} \lim_{a\to\infty} \int_0^a x^2 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{\pi}} \lim_{a\to\infty} \left[a^2 e^{-\frac{a^2}{2}} + \int_0^a e^{-\frac{x^2}{2}} dx \right] \\ &= \frac{1}{\sqrt{\pi}} \left[0 + \int_0^\infty e^{-\frac{x^2}{2}} dx \right] = \frac{1}{\sqrt{\pi}} [0 + \sqrt{\pi}] = 1 \end{aligned}$$

Suppose $Y \sim$ Normal (μ, σ^2) then we know by Corollary 5.3.3 that $W = \frac{Y-\mu}{\sigma} \sim$ Normal $(0,1)$. By Example 6.1.11, $E[W] = 0$ and $Var[W] = 1$. Also $Y = \sigma W + \mu$, so by Theorem 6.1.8(b) $E[Y] = \sigma E[W] + \mu = \mu$ and by Theorem 6.1.10 (d) and (b) $Var[Y] = \sigma^2 Var[W] = \sigma^2$. ■

EXAMPLE 6.1.12. Let $X \sim$ Uniform(a, b). To calculate the variance of X first note that Theorem 6.1.5(a) gives

$$E[X^2] \quad = \quad \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx = \int_a^b x^2 \cdot \frac{1}{b-a}\,dx = \frac{1}{3(b-a)}(b^3 - a^3) = \frac{b^2+ab+a^2}{3}.$$

Now, since $E[X] = \frac{b+a}{2}$ (see Example 6.1.2), the variance may be found as

$$Var[X] = E[X^2] - (E[X])^2 = \frac{b^2+ab+a^2}{3} - (\frac{b+a}{2})^2 = \frac{(b-a)^2}{12}.$$

Taking square roots, we obtain $SD[X] = \frac{b-a}{\sqrt{12}}$. So the standard deviation of a continuous, uniform random variable is $\frac{1}{\sqrt{12}}$ times of the length of its interval. ■

The Markov and Chebychev inequalities also apply to continuous random variables. As with discrete variables, these help to estimate the probabilities that a random variable will fall within a certain number of standard deviations from its expected value.

THEOREM 6.1.13. *Let X be a continuous random variable with probability density function f and finite non-zero variance.*

(a) **(Markov's Inequality)** *Suppose X is supported on non-negative values, i.e. $f(x) = 0$ for all $x < 0$. Then for any $c > 0$,*
$$P(X \geq c) \leq \frac{\mu}{c}.$$

(b) **(Chebychev's Inequality)** *For any $k > 0$,*
$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Proof - (a) By definition of μ and assumptions on f, we have

$$\mu = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} xf(x)dx.$$

Using an elementary fact from integrals we know that

$$\int_0^{\infty} xf(x)dx = \int_0^c xf(x)dx + \int_c^{\infty} xf(x)dx$$

We note that the first integral is non-negative so we have

$$\mu \quad \geq \quad \int_c^{\infty} xf(x)dx.$$

As $f(\cdot) \geq 0$, we have $xf(x) \geq cf(x)$ whenever $x > c$. So again using facts about integrals

$$\mu \quad \geq \quad \int_c^{\infty} cf(x)dx = c\int_c^{\infty} f(x)dx = cP(X > c).$$

The last equality follows from definition. Hence we have the result.

(b) The event $(|X-\mu| \geq k\sigma)$ is the same as the event $((X-\mu)^2 \geq k^2\sigma^2)$. The random variable $(X-\mu)^2$ is certainly non-negative, is continuous by Exercise 5.3.9, and its expected value is the variance of X which we have assumed to be finite. Therefore we may apply Markov's inequality to $(X-\mu)^2$ to get

$$P(|X-\mu| \geq k\sigma) \quad = \quad P((X-\mu)^2 \geq k^2\sigma^2) \leq \frac{E[(X-\mu)^2]}{k^2\sigma^2} = \frac{Var[X]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}.$$

Though the theorem is true for all $k > 0$, it doesn't give any useful information unless $k > 1$. ■

EXERCISES

Ex. 6.1.1. Suppose X has probability density function given by

$$f_X(x) = \begin{cases} 1- \mid x \mid & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute the distribution function of X.

(b) Compute $E[X]$ and $Var[X]$.

Ex. 6.1.2. Suppose X has probability density function given by

$$f_X(x) = \begin{cases} \frac{\cos(x)}{2} & -\frac{\pi}{2} \leq x \leq \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute the distribution function of X.

(b) Compute $E[X]$ and $Var[X]$.

Ex. 6.1.3. Find $E[X]$ and $Var[X]$ in the following situations:

(a) $X \sim \text{Normal}(\mu, \sigma^2)$, with $\mu \in \mathbb{R}$ and $\sigma > 0$.

(b) X has probability density function given by

$$f_X(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2-x & 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Ex. 6.1.4. Let $1 < \alpha$ and $X \sim \text{Pareto}(\alpha)$. Calculate $E[X]$ to show that it is finite.

Ex. 6.1.5. Let X be a random variable with density $f(x) = 2x$ for $0 < x < 1$ (and $f(x) = 0$ otherwise).

(a) Calculate $E[X]$. You should get a result larger than $\frac{1}{2}$. Explain why this should be expected even without computations.

(b) Calculate $SD[X]$.

Ex. 6.1.6. Let $X \sim \text{Uniform}(a, b)$ and let $k > 0$. Let μ and σ be the expected value and standard deviation calculated in Example 6.1.12.

(a) Calculate $P(|X-\mu| \leq k\sigma)$. Your final answer should depend on k, but not on the values of a or b.

(b) What is the value of k such that results of more than k standard deviations from expected value are unachievable for X?

Ex. 6.1.7. Let $X \sim \text{Exponential}(\lambda)$.

(a) Prove that $E[X] = \frac{1}{\lambda}$ and $SD[X] = \frac{1}{\lambda}$.

(b) Let μ and σ denote the mean and standard deviation of X respectively. Use your computations from (a) to calculate $P(|X-\mu| \leq k\sigma)$. Your final answer should depend on k, but not on the value of λ.

(c) Is there a value of k such that results of more than k standard deviations from expected value are unachievable for X?

Ex. 6.1.8. Let $X \sim \text{Gamma}(n, \lambda)$ with $n \in \mathbb{N}$ and $\lambda > 0$. Using Example 5.5.3, Exercise 6.1.7(a) and Theorem 6.1.8(c) calculate $E[X]$. Using Theorem 6.1.10 calculate $Var[X]$.

Ex. 6.1.9. Let $X \sim \text{Uniform}(0, 10)$ and let $g(x) = \max\{x, 4\}$. Calculate $E[g(X)]$.

Ex. 6.1.10. Show that as $M \to -\infty, N \to \infty$ $\int_M^N \frac{x}{1+x^2} dx$ does not have a limit.

Ex. 6.1.11. Using the hints provided below prove the respective parts of Theorem 6.1.8.

(a) For $a = 0$ the result is clear. Let $a \neq 0$ and $f_X : \mathbb{R} \to \mathbb{R}$ be the probability density function of X. Use Lemma 5.3.2 to find the probability density function of aX. Compute the expectation of aX to obtain the result. Alternatively use Theorem 6.1.5(a).

(b) Use Theorem 6.1.5(b).

(c) Use the joint density of (X, Y) to write $E[X+Y]$. Then use (5.4.2) an (5.4.3) to prove the result.

(d) Use the same technique as in (b).

(e) If $X \geq 0$ then its marginal density $f_X : \mathbb{R} \to \mathbb{R}$ is positive only when the $x \geq 0$. The result immediately follows from definition of expectation.

Ex. 6.1.12. Prove Theorem 6.1.10.

6.2 COVARIANCE, CORRELATION, CONDITIONAL EXPECTATION AND CONDITIONAL VARIANCE

Covariance of continuous random variables (X, Y) is used to describe how the two random variables relate to each other. The properties proved about covariances for discrete random variables in Section 4.5 apply to continuous random variables as well via essentially the same arguments. We define covariance and state the properties next.

Definition 6.2.1. *Let X and Y be random variables with joint probability density function $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose X and Y have finite expectation. Then the covariance of X and Y is defined as*

$$Cov[X, Y] = E[(X - E[X])(Y - E[Y])] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[X])(y - E[Y]) f(x, y) dx dy, \quad (6.2.1)$$

Since it is defined in terms of an expected value, there is the possibility that the covariance may be infinite or not defined at all. We now state the properties of Covariance.

THEOREM 6.2.2. *Let X, Y be continuous random variables such that they have joint probability density function. Assume that $0 \neq \sigma_x^2 = Var(X) < \infty, 0 \neq \sigma_y^2 = Var(Y) < \infty$. Then*

(a) $Cov[X,Y] = E[XY] - E[X]E[Y]$.

(b) $Cov[X,Y] = Cov[Y,X]$;

(c) $Cov[X,X] = Var[X]$.

(d) $-\sigma_X\sigma_Y \leq Cov[X,Y] \leq \sigma_X\sigma_Y$

(e) *If X and Y are independent then $Cov[X,Y] = 0$.*

Let a, b be real numbers. Suppose Z is another continuous random variable, and $\sigma_z = Var(Z) < \infty$. Further (X,Z), (Y,Z), $(X, aY + bZ)$, and $(aX + bY, Z)$ all have (their respective) joint probability functions. Then

(f) $Cov[X, aY + bZ] = a \cdot Cov[X,Y] + b \cdot Cov[X,Z]$;

(g) $Cov[aX + bY, Z] = a \cdot Cov[X,Z] + b \cdot Cov[Y,Z]$;

Proof- See Exercise 6.2.13 ■

DEFINITION 6.2.3. *Let (X,Y) be continuous random variables both with finite variance and covariance. From Theorem 6.2.2(d) the quantity $\rho[X,Y] = \frac{Cov[X,Y]}{\sigma_X\sigma_Y}$ is in the interval $[-1,1]$. It is known as the"correlation" of X and Y. As discussed earlier, both the numerator and denominator include the units of X and the units of Y. The correlation, therefore, has no units associated with it. It is thus a dimensionless rescaling of the covariance and is frequently used as an absolute measure of trends between the two continuous random variables as well.*

EXAMPLE 6.2.4. Let $X \sim$ Uniform $(0,1)$ and be independent of $Y \sim$ Uniform $(0,1)$. Let $U = \min(X,Y)$ and $V = \max(X,Y)$. We wish to find $\rho[U,V]$. First, $0 < u < 1$

$$P(U \leq u) = 1 - P(U > u) = 1 - P(X > u, Y > u) = 1 - P(X > u)P(Y > u) = 1 - (1-u)^2,$$

as X, Y are independent uniform random variables. Second, for $0 < v < 1$,

$$P(V \leq v) = P(X \leq v, Y \leq v) = P(X \leq v)P(Y \leq v) = v^2,$$

as X, Y are independent uniform random variables. Therefore the distribution function of U and V are given by

$$F_U(u) = \begin{cases} 0 & \text{if } u < 0 \\ 1 - (1-u)^2 & \text{if } 0 < u < 1 \\ 1 & \text{if } u \geq 1. \end{cases} \quad \text{and } F_V(v) = \begin{cases} 0 & \text{if } v < 0 \\ v^2 & \text{if } 0 < v < 1 \\ 1 & \text{if } v \geq 1. \end{cases}$$

As F_U, F_V are piecewise differentiable, the probability density function of U and V are obtained by differentiating F_U and F_V respectively.

$$f_U(u) = \begin{cases} 2(1-u) & \text{if } 0 < u < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and } f_V(v) = \begin{cases} 2v & \text{if } 0 < v < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Thirdly, $0 < u < v < 1$

$$\begin{aligned} P(U \leq u, V \leq v) &= P(V \leq v) - P(U > u, V \leq v) \\ &= v^2 - P(u < X \leq v, u < Y \leq v) \\ &= v^2 - P(u < X \leq v)P(u < Y \leq v) \\ &= v^2 - (v-u)^2, \end{aligned}$$

where we have used the formula for distribution function of V and the fact that X, Y are independent uniform random variables. It is easily seen that $P(U \leq u, V \leq v) = 0$ for all other possibilities of (u, v). As the joint distribution function is piecewise differentiable in each variable, the joint probability density function of U and V, $f : \mathbb{R}^2 \to \mathbb{R}$, exists and is obtained by differentiating it partially in u and v.

$$f(u,v) = \begin{cases} 2 & \text{if } 0 < u < v < 1 \\ 0 & \text{otherwise} \end{cases}$$

Now,

$$\begin{aligned} E[U] &= \int_0^1 u2(1-u)du = u^2 - 2\frac{u^3}{3} \mid_0^1 = \frac{1}{3} \\ E[V] &= \int_0^1 v2vdv = 2\frac{v^3}{3} \mid_0^1 = \frac{2}{3} \\ E[U^2] &= \int_0^1 u^2 2(1-u)du = 2\frac{u^3}{3} - 2\frac{u^4}{4} \mid_0^1 = \frac{1}{6} \\ E[V^2] &= \int_0^1 v^2 2vdv = 2\frac{v^4}{4} \mid_0^1 = \frac{1}{2} \\ E[UV] &= \int_0^1 \left[\int_0^v uv2du\right] dv = \int_0^1 2v \left[\frac{u^2}{2} \mid_0^1\right] dv = \int_0^1 2v\frac{v^2}{2}dv = \frac{v^4}{4} \mid_0^1 = \frac{1}{4} \end{aligned}$$

Therefore

$$\begin{aligned} Var[U] &= E[U^2] - (E[U])^2 = \frac{2}{3} - \frac{1}{9} = \frac{5}{9} \\ Var[V] &= E[V^2] - (E[V])^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18} \\ Cov[U,V] &= E[UV] - E[U]E[V] = \frac{1}{4} - \frac{1}{3}\frac{2}{3} = \frac{5}{36} \\ \rho[U,V] &= \frac{Cov[U,V]}{\sqrt{Var[V]}\sqrt{Var[U]}} = \frac{\frac{5}{36}}{\sqrt{\frac{5}{9}}\sqrt{\frac{1}{18}}} = \frac{1}{2\sqrt{2}} \end{aligned}$$

■

As seen in Theorem 6.2.2 (e), independence of X and Y guarantees that they are uncorrelated (i.e $\rho[X,Y] = 0$). The converse is not true (See Example 4.5.6 for discrete case). It is possible that $Cov[X,Y] = 0$ and yet that X and Y are dependent, as the next example shows.

Example 6.2.5. Let $X \sim$ Uniform $(-1, 1)$. Let $Y = X^2$. Note from Example 6.1.2 and Example 6.1.12 we have $E[X] = 0, E[Y] = E[X^2] = \frac{1}{3}$. Further using the probability density function of X,

$$E[XY] = E[X^3] = \int_{-1}^1 x^3\frac{1}{2} = \frac{x^4}{8} \mid_{-1}^1 = 0.$$

So $\rho[X,Y] = 0$. Clearly X and Y are not independent. We verify this precisely as well. Consider the

$$P(X \le -\frac{1}{4}, Y \le \frac{1}{4}) = P(X \le -\frac{1}{4}, X^2 \le \frac{1}{4}) = P(-\frac{1}{2} \le X \le -\frac{1}{4}) = \frac{1}{8},$$

as $X \sim$ Uniform $(-1,1)$. Whereas,

$$P(X \le -\frac{1}{4})P(Y \le \frac{1}{4}) = P(X \le -\frac{1}{4})P(X^2 \le \frac{1}{4}) = P(X \le -\frac{1}{4})P(-\frac{1}{2} \le X \le \frac{1}{2}) = \frac{3}{8}\frac{1}{2} = \frac{3}{16}.$$

Clearly

$$P(X \le -\frac{1}{4}, Y \le \frac{1}{4}) \neq P(X \le -\frac{1}{4})P(Y \le \frac{1}{4})$$

implying they are not independent. ■

We are now ready to define conditional expectation and variance.

DEFINITION 6.2.6. *Let (X,Y) be continuous random variables with a piecewise continuous joint probability density function f. Let f_X be the marginal density of X. Assume x is a real number for which $f_x(x) \neq 0$. The conditional expectation of Y given $X = x$ is defined by*

$$E[Y \mid X = x] = \int_{-\infty}^{\infty} y f_{Y|X=x}(y)dy = \int_{-\infty}^{\infty} y \frac{f(x,y)}{f_X(x)}dy$$

whenever it exists. The conditional variance of Y given $X = x$ is defined by

$$\begin{aligned} Var[Y|X=x] &= E[(Y - E[Y|X=x])^2|X=x] \\ &= \int_{-\infty}^{\infty} \left(y - \int_{-\infty}^{\infty} y\frac{f(x,y)}{f_X(x)}dy \right)^2 \frac{f(x,y)}{f_X(x)}dy. \end{aligned}$$

The results proved in Theorem 4.4.4, Theorem 4.4.6, Theorem 4.4.8, and Theorem 4.4.9 are all applicable when X and Y are continuous random variables having joint probability density function f. The proofs of these results in the continuous setting follow very similarly (though using facts about integrals from analysis).

THEOREM 6.2.7. *Let (X,Y) be continuous random variables with joint probability density function $f : \mathbb{R} \to \mathbb{R}$. Assume that $h, g : \mathbb{R} \to \mathbb{R}$ be defined as*

$$g(y) = \begin{cases} E[X|Y=y] & \text{if } f_Y(y) > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and } h(y) = \begin{cases} Var[X|Y=y] & \text{if } f_Y(y) > 0 \\ 0 & \text{otherwise} \end{cases}$$

are well-defined piecewise continuous functions. Let $k : \mathbb{R} \to \mathbb{R}$ be a piecewise continous function. Then

$$E[k(X) \mid Y = y] = \int_{-\infty}^{\infty} k(x) f_{X|Y=y}(x)dx, \tag{6.2.2}$$

$$E[g(Y)] = E[X], \tag{6.2.3}$$

and

$$Var[X] = E[h(Y)] + Var[g(Y)]. \tag{6.2.4}$$

Proof- The proof of (6.2.2) is beyond the scope of this book. We shall omit it. To prove (6.2.3) we use the definition of g and Theorem 6.1.8 (a) to write

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f_Y(y)dy = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x f_{X|Y=y}(x)dx\right] f_Y(y)dy$$

Using the definition of conditional density and rearranging the order of integration we obtain that the above is

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x\frac{f(x,y)}{f_Y(y)}dx\right] f_Y(y)dy = \int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f(x,y)dy\right]dx = \int_{-\infty}^{\infty} x f_X(x)dx = E[X].$$

So we are done. To prove (6.2.4), using Exercise 6.2.8

$$h(y) = E[X^2 \mid Y = y] - (E[X \mid Y = y])^2 = E[X^2 \mid Y = y] - (g(y))^2$$

From the above we have,

$$\begin{aligned} E[h(Y)] &= E[X^2] + E[g(Y)^2] \\ Var[g(Y)] &= E[g(Y)^2] - (E[g(Y)])^2 = E[g(Y)^2] - (E[X])^2 \end{aligned}$$

Therefore summing the two equations we have (6.2.4). ■

As before it is common to use $E[X|Y]$ to denote $g(Y)$ after which the result may be expressed as $E[E[X|Y]] = E[X]$. This can be slightly confusing notation, but one must keep in mind that the exterior expected value in the expression $E[E[X|Y]]$ refers to the averge of $E[X|Y]$ viewed as a function of Y.

Similarly one denotes $h(Y)$ by $Var[X|Y]$. Then we can rewrite (6.2.4) as

$$Var[X] = E[Var[X|Y]] + Var[E[X|Y]].$$

EXAMPLE 6.2.8. Let $X \sim$ Uniform $(0,1)$ and be independent of $Y \sim$ Uniform $(0,1)$. Let $U = \min(X,Y)$ and $V = \max(X,Y)$. In Example 6.2.4 we found $\rho[U,V]$. During that computation we showed that the marginal densities of U and V were given by

$$f_U(u) = \begin{cases} 2(1-u) & \text{if } 0 < u < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and } f_V(v) = \begin{cases} 2v & \text{if } 0 < v < 1 \\ 0 & \text{otherwise.} \end{cases}$$

and the joint density of (U,V) was given by

$$f(u,v) = \begin{cases} 2 & \text{if } 0 < u < v < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $0 < u < 1$. The conditional density of $V \mid U = u$, is given by

$$f_{V|U=u}(v) = \frac{f(u,v)}{f_U(u)}, \text{ for } v \in \mathbb{R}.$$

So,

$$f_{V|U=u}(v) = \begin{cases} \frac{1}{1-u} & \text{if } u < v < 1 \\ 0 & \text{otherwise} \end{cases}$$

Therefore $(V \mid U = u) \sim$ Uniform $(u,1)$. So the conditional expectation is given by

$$E[V \mid U = u] = \int_u^1 \frac{v}{1-u}dv = \frac{1-u^2}{2(1-u)} = \frac{1+u}{2}.$$

The conditional variance is given by

$$\begin{aligned} Var[V \mid U=u] &= E[V^2 \mid U=u] - (E[V \mid U=u])^2 \\ &= \int_u^1 \frac{v^2}{1-u} dv - \left(\frac{1+u}{2}\right)^2 \\ &= \frac{1-u^3}{3(1-u)} dv - \frac{(1+u)^2}{4} = \frac{(1-u)^2}{12}. \end{aligned}$$

We could have also concluded these from properties of Uniform distribution computed in Example 6.1.2 and Example 6.1.12. We will use this approach in the next example. ■

EXAMPLE 6.2.9. Let (X,Y) have joint probability density function f given by

$$f(x,y) = \frac{\sqrt{3}}{4\pi} e^{-\frac{1}{2}(x^2-xy+y^2)} \quad -\infty < x, y < \infty.$$

These random variables were considered in Example 5.4.12. We showed there that X is a Normal random variable with mean 0 and variance $\frac{4}{3}$ and Y is also a Normal random variable with mean 0 and variance $\frac{4}{3}$. We observed that they are not independent as well and the conditional distribution of Y given $X=x$ was Normal with mean $\frac{x}{2}$ and variance 1. Either by direct computation or by definition we observe that

$$E[Y \mid X=x] = \frac{x}{2} \quad Var[Y \mid X=x] = 1.$$

We could compute the $Var[Y]$ using (6.2.4), i.e

$$\begin{aligned} Var[Y] &= Var[E[Y \mid X]] + E[Var[Y \mid X=x]] \\ &= Var[\frac{X}{2}] + E[1] \\ &= \frac{1}{4} Var[X] + 1 = \frac{1}{4}\frac{4}{3} + 1 = \frac{4}{3}. \end{aligned}$$

■

EXERCISES

Ex. 6.2.1. Let (X,Y) be uniformly distributed on the triangle $0 < x < y < 1$.

(a) Compute $E[X|Y=\frac{1}{6}]$.

(b) Compute $E[(X-Y)^2]$.

Ex. 6.2.2. X is a random variable with mean 3 and variance 2. Y is a random variable with mean -1 and variance 6. The covariance of X and Y is -2. Let $U = X+Y$ and $V = X-Y$. Find the correlation coefficient of U and V.

Ex. 6.2.3. Suppose X and Y are both uniformly distributed on $[0,1]$. Suppose $Cov[X,Y] = \frac{-1}{24}$. Compute the variance of $X+Y$.

Ex. 6.2.4. A dice game between two people is played by a pair of dice being thrown. One of the dice is green and the other is white. If the green die is larger than the white die, player number one earns a number of points equal to the value on the green die. If the green die is less than or equal to the white die, then player number two earns a number of points equal to the value of the green die. Let X be the random variable representing the number of points earned by player one after one throw. Let Y be the random variable representing the number of points earned by player two after one throw.

(a) Compute the expected value of X and of Y.

(b) Without explicitly computing it, would you expect $Cov[X, Y]$ to be positive or negative? Explain.

(c) Calculate $Cov[X, Y]$ to confirm your intuition.

Ex. 6.2.5. Suppose X has variance σ_X^2, Y has variance σ_Y^2, and the pair (X, Y) has correlation coefficient $\rho[X, Y]$.

(a) In terms of σ_X, σ_Y, and $\rho[X, Y]$, find $Cov[X, Y]$ and $Cov[X + Y, X - Y]$.

(b) What must be true of σ_X^2 and σ_Y^2 if $X + Y$ and $X - Y$ are uncorrelated?

Ex. 6.2.6. Let (X, Y) have the joint probability density function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f_{X,Y}(x, y) = \begin{cases} 3(x + y) & \text{if } x > 0,\ y > 0, \text{ and } x + y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $E[X|Y = \frac{1}{2}]$ and $Var[X|Y = \frac{1}{2}]$

(b) Are X and Y independent ?

Ex. 6.2.7. Suppose Y is uniformly distributed on $(0, 1)$, and suppose for $0 < y < 1$ the conditional density of $X \mid Y = y$ is given by

$$f_{X|Y=y}(x) = \begin{cases} \frac{2x}{y^2} & \text{if } 0 < x < y \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that, as a function of x, $f_{X|Y=y}(x)$ is a density.

(b) Compute the joint p.d.f. of (X, Y) and the marginal density of X.

(c) Compute the expected value and variance of X given that $Y = y$, with $0 < y < 1$.

Ex. 6.2.8. Let (X, Y) have joint probability density function $f : \mathbb{R}^2 \to \mathbb{R}$. Show that $Var[X \mid Y = y] = E[X^2 \mid Y = y] - (E[X \mid Y = y])^2$.

Ex. 6.2.9. For random variables (X, Y) as in Exercise 5.4.1, find

(a) $E[X]$ and $E[Y]$

(b) $Var[X]$ and $Var[Y]$

(c) $Cov[X, Y]$ and $\rho[X, Y]$

Ex. 6.2.10. From Example 5.4.12, consider(X, Y) have joint probability density function f given by

$$f(x, y) = \frac{\sqrt{3}}{4\pi} e^{-\frac{1}{2}(x^2 - xy + y^2)} \quad -\infty < x, y < \infty.$$

Find

(a) $E[X]$ and $E[Y]$

(b) $Var[X]$ and $Var[Y]$

(c) $Cov[X,Y]$ and $\rho[X,Y]$

Ex. 6.2.11. From Example 5.4.13, suppose $T = \{(x,y) \mid 0 < x < y < 4\}$ and let $(X,Y) \sim$ Uniform (T). Find

(a) $E[X]$ and $E[Y]$

(b) $Var[X]$ and $Var[Y]$

(c) $Cov[X,Y]$ and $\rho[X,Y]$

Ex. 6.2.12. From Example 5.4.9, consider the open disk in $\mathbb{R}^2$ given by $C = \{(x,y) : x^2 + y^2 < 25\}$ and $\mid C \mid = 25\pi$ denote its area. Let (X,Y) have a joint density $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = \begin{cases} \frac{1}{|C|} & \text{if } (x,y) \in C \\ 0 & \text{otherwise.} \end{cases}$$

Find

(a) $E[X]$ and $E[Y]$

(b) $Var[X]$ and $Var[Y]$

(c) $Cov[X,Y]$ and $\rho[X,Y]$

Ex. 6.2.13. Using the hints provided below prove the respective parts of Theorem 6.2.2

(a) Use the linearity properties of the expected value from Theorem 6.1.8.

(b) Use definition of covariance.

(c) Use the definitions of variance and covariance.

(d) Imitate the proof of Theorem 4.5.7.

(e) Use part (a) of this problem and part (f) of Theorem **??**.

(f) Use the linearity properties of the expected value from Theorem 6.1.8.

(g) Use the linearity properties of the expected value from Theorem 6.1.8.

Ex. 6.2.14. Let X,Y be continuous random variable with piecewise continuous densities $f(x)$ and $g(y)$ and well-defined expected values. Suppose $X \le Y$ then show that $E[X] \le E[Y]$.

Ex. 6.2.15. Let T be the triangle bounded by the lines $y = 0$, $y = 1 - x$, and $y = 1 + x$. Suppose a random vector (X,Y) has a joint p.d.f.

$$f_{(X,Y)}(x,y) = \begin{cases} 3y & \text{if} \\ 0 & \text{otherwise.} \end{cases}$$

Compute $E[Y|X = \frac{1}{2}]$.

Ex. 6.2.16. Let (X,Y) be random variables with joint probability density function $f : \mathbb{R}^2 \to \mathbb{R}$. Assume that both random variables have finite variances and that their covariance is also finite.

(a) Show that $Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$.

(b) Show that when X and Y are positively correlated (i.e. $\rho[X,Y] > 0$) then $Var[X+Y] > Var[X] + Var[Y]$, while when X and Y are negatively correlated (i.e. $\rho[X,Y] < 0$), then $Var[X+Y] < Var[X] + Var[Y]$.

6.3 MOMENT GENERATING FUNCTIONS

We have already seen for the distribution of a discrete random variable or a continuous random variable is determined by its distribution function. In this section we shall discuss the concept of moment generating functions. Under suitable assumptions, these functions will determine the distribution of random variables. They are also serve as tools in computations and come in handy for convergence concepts that we will discuss.

The moment generating function generates or determine the moments which in turn, under suitable hypothesis determine the distribution of the corresponding random variable. We begin with a definition of a moment.

Definition 6.3.1. *Suppose X is a random variable. For a positive integer k, the quantity*

$$m_k = E[X^k]$$

is known as the "k^{th} moment of X". As before the existence of a given moment is determined by whether the above expectation exists or not.

We have previously seen many computations of the first moment $E[X]$ and also seen that the second moment $E[X^2]$ is related to the variance of the random variable. The next theorem states that if a moment exists then it guarantees the existence of all lesser moments.

Theorem 6.3.2. *Let X be a random variable and let k be a positive integer. If $E[X^k] < \infty$ then $E[X^j] < \infty$ for all positive integers $j < k$.*

Proof - Suppose X is a continuous random variable. Suppose $E[X^k]$ exists and is finite, so that $E[|X^k|] < \infty$. Divide $\mathbb{R}$ in two pieces by letting $R_1 = \{x \in T : |x| < 1\}$ and letting $R_2 = \{x \in T : |x| \geq 1\}$. If $j < k$ then $|x|^j \leq |x|^k$ for $x \in R_2$ so,

$$\begin{aligned}
E[|X^j|] &= \int_{\mathbb{R}} |x|^j f_X(x)\,dx = \int_{R_1} |x|^j f_X(x)\,dx + \int_{R_2} |x|^j f_X(x)\,dx \\
&\leq \int_{R_1} 1 \cdot f_X(x)\,dx + \int_{R_2} |x|^k f_X(x)\,dx \\
&\leq \int_{R_1} f_X(x)\,dx + \int_{R_2} |x|^k f_X(x)\,dx \\
&= 1 + E[|X^k|] < \infty
\end{aligned}$$

Therefore $E[X^j]$ exists and is finite. See Exericse 6.3.7 when X is a discrete random variable. ■

When a random variable has finite moments for all positive integers, then these moments provide a great deal of information about the random variable itself. In fact, in some cases, these moments serve to completely describe the distribution of the random variable. One way to simultaneously describe all moments of such a variable in terms of a single expression is through the use of a "moment generating function".

Definition 6.3.3. *Suppose X is a random variable and $D = \{t \in \mathbb{R} : E[e^{tX}]$ exists$\}$. The function $M : D \to \mathbb{R}$ given by*

$$M(t) = E[e^{tX}],$$

is called the moment generating function for X.

The notation $M_X(t)$ will also be used when clarification is needed as to which variable a particular moment generating function belongs. Note that $M(0) = 1$ will always be true, but for other values of t, there is no guarantee that the function is even defined as the expected value might be infinite. However, when $M(t)$ has derivatives defined at zero, these values incorporate information about the moments of X. For a discrete random variable $X : S \to T$ with $T = \{x_i : i \in \mathbb{N}\}$, then for $t \in D$ (as in Definition 6.3.3)

$$M_X(t) = \sum_{i \geq 1} e^{tx_i} P(X = x_i).$$

For a continuous random variable X with probability density function $f_X : \mathbb{R} \to \mathbb{R}$ then for $t \in D$ (as in Definition 6.3.3)

$$M_X(t) = \int_{\mathbb{R}} e^{tx} f_X(x) dx.$$

We compute moment generating function for a Poisson (λ) and a Gamma (n, λ), with $n \in \mathbb{N}, \lambda > 0$.

EXAMPLE 6.3.4. Suppose $X \sim$ Poisson (λ) then for all $t \in \mathbb{R}$,

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} P(X = k) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!} = e^{-\lambda} e^{e^t \lambda} = e^{-\lambda(1+e^t)}.$$

So the moment generating function of X exists for all $t \in \mathbb{R}$. Suppose $Y \sim$ Gamma (n, λ) then $t < \lambda$,

$$M_Y(t) = \int_R e^{ty} \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y} dy = \frac{\lambda^n}{\Gamma(n)} \int_R y^{n-1} e^{-(\lambda - t)y} dy = \frac{\lambda^n}{\Gamma(n)} \frac{\Gamma(n)}{(\lambda - t)^n} = \left(\frac{\lambda}{\lambda - t} \right)^n,$$

where we have used (5.5.3). The moment generating function of Y will not be finite if $t \geq \lambda$. ■

We summarily compile some facts about moment generating functions. The proof of some of the results are beyond the scope of this text.

THEOREM 6.3.5. *Suppose for a random variable X, there exists $\delta > 0$ such that $M_X(t)$ exists $(-\delta, \delta)$.*

(a) The k^{th} moment of X exists and is given by

$$E[X^k] = M_X^{(k)}(0),$$

where $M_X^{(k)}$ denotes the k^{th} derivative of M_X.

(b) For $0 \neq a \in \mathbb{R}$ such that $at, t \in (-\delta, \delta)$ we have

$$M_{aX}(t) = M_X(at).$$

(c) Suppose Y is another independent random variable such that $M_Y(t)$ exists for $t \in (-\delta, \delta)$. Then

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

for $t \in (-\delta, \delta)$.

Proof - (a) A precise proof is beyond the scope of this book. We provide a sketch. Express e^{tX} as a power series in t.

$$e^{tX} = 1 + tX + \frac{t^2X^2}{2} + \cdots + \frac{t^nX^n}{n!} + \ldots$$

The expected value of the left hand side is the moment generating function for X while linearity may be used on the right hand side. So the power series of $M(t)$ is given by

$$M(t) = 1 + t \cdot E[X] + \frac{t^2}{2} \cdot E[X^2] + \cdots + \frac{t^n}{n!} \cdot E[X^n] + \ldots$$

Taking k derivatives of both sides of the equation (which is valid in the interval of convergence) yields

$$M^{(k)}(t) = E[X^k] + t \cdot E[X^{k+1}] + \frac{t^2}{2} \cdot E[X^{k+2}] + \ldots$$

Finally, when evaluating both sides at $t = 0$ all but one term on the right hand side vanishes and the equation becomes simply $M^{(k)}(0) = E[X^k]$.

(b) $M_{aX}(t) = E[e^{(aX)t}] = E[e^{X(at)}] = M_X(at)$.

(c) Using Theorem 4.1.10 or Theorem 6.1.10 (f) we have

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t).$$

■

Theorem 6.3.5 applies equally well for both discrete and continuous variables. A discrete example is presented next.

Example 6.3.6. Let $X \sim$ Geometric(p). We shall find $M_X(t)$ and use this function to calculate the expected value and variance X. For any $t \in \mathbb{R}$,

$$\begin{aligned} M_X(t) &= E[e^{tX}] = \sum_{n=1}^{\infty} e^{tn}P(X=n) = \sum_{n=1}^{\infty}(e^t)^n \cdot p(1-p)^{n-1} = pe^t \cdot \sum_{n=1}^{\infty}(e^t \cdot (1-p))^{n-1} \\ &= \frac{pe^t}{1-e^t(1-p)} \end{aligned}$$

Having completed that computation, the expected value and variance can be computed simply by calculating derivatives.

$$M'_X(t) = \frac{pe^t}{[1-(1-p)e^t]^2}$$

and so $E[X] = M'_X(0) = \frac{p}{p^2} = \frac{1}{p}$. Similarly,

$$M''_X(t) = \frac{pe^t + p(1-p)e^{2t}}{[1-(1-p)e^t]^3}$$

and so $E[X^2] = M''_X(0) = \frac{2p-p^2}{p^3} = \frac{2}{p^2} - \frac{1}{p}$. Therefore, $Var[X] = E[X^2] - (E[X])^2 = \frac{1-p}{p^2}$. Both the expected value and variance are in agreement with the previous computations for the goemetric random variable.

Let $Y \sim$ Normal(μ, σ^2). The density of Y is $f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(y-\mu)^2/2\sigma^2}$. For any $t \in \mathbb{R}$,

$$\begin{aligned} M_Y(t) &= E[e^{tY}] = \int_{-\infty}^{\infty} e^{ty} \cdot \frac{1}{\sigma\sqrt{2\pi}}e^{-(y-\mu)^2/2\sigma^2}\,dy = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}e^{-(y^2-(2\mu y+2\sigma^2 ty)+\mu^2)/2\sigma^2}\,dy \\ &= e^{\mu t+(1/2)\sigma^2t^2}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-(y-(\mu+\sigma^2 t))^2/2\sigma^2}\,dy \\ &= e^{\mu t+(1/2)\sigma^2t^2} \end{aligned} \tag{6.3.1}$$

where the integral in the final step is equal to one since it integrates the density of a Normal$(\mu + \sigma^2 t, \sigma^2)$ random variable. One can easily verify that the $M_Y'(0) = \mu$ and $M_Y''(0) = \mu^2 + \sigma^2$. ■

As with the expected value and variance, moment generating functions behave well when applied to linear combinations of independent variables (courtesy Theorem 6.3.5 (b) and (c)).

EXAMPLE 6.3.7. Supppose we wish to find the moment generating function of $X \sim$ Binomial(n, p). We have seen that such a random variable may arise as the sum of indpendent Bernoulli variables. That is, $X = Y_1 + \cdots + Y_n$ where $Y_j \sim$ Bernoulli(p). But it is routine to compute

$$M_{Y_j}(t) = E[e^{tY_j}] = e^{t\cdot 1}P(Y_j = 1) + e^{t\cdot 0}P(Y_j = 0) = pe^t + (1-p).$$

Therefore by linearity (inductively applying Theorem 6.3.5 (c)),

$$M_X(t) = M_{Y_1+\cdots+Y_n}(t) = M_{Y_1}(t)\cdot\ldots\cdot M_{Y_n}(t) = (pe^t + (1-p))^n.$$

■

Moment generating functions are an extraordinarily useful tool in analyzing the distributions of random variables. Two particularly useful tools involve the uniqueness and limit properties of such generating functions. Unfortunately these theorems require analysis beyond the scope of this text to prove. We will state the uniqueness fact (unproven) below and the limit property in Chapter 8. First we generalize the definition of moment generating functions to pairs of random variables.

DEFINITION 6.3.8. *Suppose X and Y are random variables. Then the function*

$$M(s, t) = E[e^{sX+tY}]$$

is called the (joint) moment generating function for X and Y. The notation $M_{X,Y}(s, t)$ will be used when confusion may arise as to which random variables are being represented.

Moment generating functions completely describe the distributions of random variables. We state the result precisely.

THEOREM 6.3.9. *(M.G.F. Uniqueness Theorem)*

(a) (One variable) Suppose X and Y are random variables and $M_X(t) = M_Y(t)$ in some open interval containing the origin. Then X and Y are equal in distribution.

(b) (Two variable) Suppose (X, W) and (Y, Z) are pairs of random variables and suppose $M_{X,W}(s, t) = M_{Y,Z}(s, t)$ in some rectangle containing the origin. Then (X, W) and (Y, Z) have the same joint distribution.

An immediate application of the theorem is an alternate proof of Corollary 5.3.3 based on moment generating functions.

EXAMPLE 6.3.10. Let $X \sim$ Normal(μ, σ^2) and let $Y = \frac{X-\mu}{\sigma}$. Show that $Y \sim$ Normal$(0, 1)$.

We know X is normal, (6.3.1) shows that the moment generating function of X is $M_X(t) = e^{\mu t + (1/2)\sigma^2 t^2}$, for all $t \in \mathbb{R}$. So consider the moment generating function of Y. For all $t \in \mathbb{R}$

$$\begin{aligned} M_Y(t) &= E[e^{tY}] = E[e^{t(X-\mu)/\sigma}] = E[e^{tX/\sigma}e^{-t\mu/\sigma}] = e^{-t\mu/\sigma}\cdot M_X(\frac{t}{\sigma}) \\ &= e^{-t\mu/\sigma}\cdot e^{\mu(t/\sigma)+(1/2)\sigma^2(t/\sigma)^2} = e^{\frac{t^2}{2}}. \end{aligned}$$

But this expression is the moment generating function of a Normal$(0, 1)$ random variable. So by the uniqueness of moment generating functions, Theorem 6.3.9 (a), the distribution of Y is Normal$(0, 1)$. ■

Just as the joint density of a pair of random variables factors as a product of marginal densities exactly when the variables are independent (Theorem 5.4.7), a similar result holds for moment generating functions.

Theorem 6.3.11. *Suppose (X, Y) are a pair of continuous random variables with moment generating function $M(s, t)$. Then X and Y are indpendent if and only if*

$$M(s,t) = M_X(s) \cdot M_Y(t).$$

Proof - One direction of the proof follows from basic facts about independence. If X and Y are independent, then by Exercise 6.3.4 , we have

$$M(s,t) \quad = \quad E[e^{sX+tY}] = E[e^{sX}e^{tY}] = E[e^{sX}]E[e^{tY}] = M_X(s) \cdot M_Y(t).$$

To prove the opposite direction, we shall use Theorem 6.3.9(b). Let $\hat{X}$ and $\hat{Y}$ be independent, but have the same distributions as X and Y respectively. Since $M_{X,Y}(s,t) = M_X(s)M_Y(t)$ we have the following series of equalities:

$$M_{X,Y}(s,t) = M_X(s)M_Y(t) = M_{\hat{X}}(s)M_{\hat{Y}}(t) = M_{\hat{X},\hat{Y}}(s,t).$$

By Theorem 6.3.9(b), this means that (X, Y) and $(\hat{X}, \hat{Y})$ have the same distribution. This would imply that

$$P(X \in A, Y \in B) = P(\hat{X} \in A, \hat{Y} \in B) = P(\hat{X} \in A)P(\hat{Y} \in B) = P(X \in A)P(Y \in B),$$

for any events A and B. Hence X and Y are independent. ■

Notice that the method employed in Example 6.3.10 did not require considering integrals directly. Since the manipulation of integrals can be complicated (particularly when dealing with multiple integrals), the moment generating function method will often be simpler as the next example illustrates.

Example 6.3.12. Let a, b be two real numbers. Let $X \sim$ Normal(μ_1, σ_1^2) and $Y \sim$ Normal(μ_2, σ_2^2) be independent. Observe that

$$M_{aX+bY}(t) \quad = \quad M_{X,Y}(at, bt)$$

Using Theorem 6.3.11, we have that the above is

$$M_X(at)M_Y(bt) = e^{a\mu_1 t + (1/2)a^2\sigma_1^2 t^2} e^{b\mu_2 t + (1/2)b^2\sigma_2^2 t^2} = e^{(a\mu_1 + b\mu_2)t + (1/2)(a^2\sigma_1^2 + b^2\sigma_2^2)t^2}$$

which is the moment generating function of a Normal random variable with mean $a\mu_1 + b\mu_2$ and variance $a^2\sigma_1^2 + b^2\sigma_2^2$). So $aX + bY \sim$ Normal$(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$. ■

We conclude this section with a result on finite linear combinations of independent normal random variables.

Theorem 6.3.13. *Let $X_1, X_2, \ldots, X_n$ be independent, normally distributed random variables with mean μ_i and variance σ_i^2 respectively for $i = 1, 2, \ldots n$. Let $a_1, a_2, \ldots, a_n$ be real-valued numbers, not all of which are zero. Then then the linear combination $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ is also normally distributed with mean $\sum_{i=1}^n a_i\mu_i$ and variance $\sum_{i=1}^n a_i^2\sigma_i^2$.*

Proof- This follows from the preceeding example by induction and is left as an exercise. ■

EXERCISES

Ex. 6.3.1. Let $X \sim \text{Normal}(0,1)$. Use the moment generating function of X to calcluate $E[X^4]$.

Ex. 6.3.2. Let $Y \sim \text{Exponential}(\lambda)$.

(a) Calculate the moment generating function $M_Y(t)$.

(b) Use (a) to calculate $E[Y^3]$ and $E[Y^4]$, the third and fourth moments of an exponential distriubtion.

Ex. 6.3.3. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables.

(a) Let $Y = X_1 + \cdots + X_n$. Prove that $M_Y(t) = [M_{X_1}(t)]^n$.

(b) Let $Z = (X_1 + \cdots + X_n)/n$. Prove that $M_Z(t) = [M_{X_1}(\frac{t}{n})]^n$.

Ex. 6.3.4. Let X and Y be two independent discrete random variables. Let $h : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. Show that

$$E[h(X)g(Y)] = E[h(X)]E[g(Y)].$$

Show that the above holds if X and Y are independent continous random variables.

Ex. 6.3.5. Suppose X is a discrete random variable and $D = \{t \in \mathbb{R} : E[t^X] \text{ exists}\}$. The function $\psi : D \to \mathbb{R}$ given by

$$\psi(t) = E[t^X],$$

is called the probability generating function for X. Calculate the probability generating function of X when X is

(a) $X \sim \text{Bernoulli}(p)$, with $0 < p < 1$.

(b) $X \sim \text{Binomial}(n,p)$, with $0 < p < 1$, $n \geq 1$.

(c) $X \sim \text{Geometric}(p)$, with $0 < p < 1$.

(d) $X \sim \text{Poisson } (\lambda)$, with $0 < \lambda$.

Ex. 6.3.6. Let $X, Y : S \to T$ be dicrete random variables with the number of elements in T is finite. Prove part (a) of Theorem 6.3.9 in this case.

Ex. 6.3.7. Prove Theorem 6.3.2 when X is a discrete random variable.

6.4 BIVARIATE NORMALS

In Example 6.3.12, we saw that if X and Y are independent, normally distributed random variables, any linear combination $aX + bY$ is also normally distributed. In such a case the joint density of (X,Y) is determined easily (courtesy Theorem 5.4.7). We would like to understand random variables that are not independent but have normally distributed marginals. Motivated by the observations in Example 6.3.12 we provide the following definition.

DEFINITION 6.4.1. *A pair of random variables (X,Y) is called "bivariate normal" if $aX + bY$ is a normally distributed random variable for all real numbers a and b.*

We need to be somewhat cautious in the above definition. Since the variables are dependent it may turn out that $aX + bY = 0$ or some constant. (E.g: $Y = -X$,or $Y = -X + 2$ with $a = 1, b = 1$). We shall follow the convention that a constant c random variable in such cases is a normal random variable with mean c and variance 0.

If (X, Y) are bivariate normal then as $X = X + 0Y$ and $Y = 0X + Y$ both X and Y individually are normal random variables. The converse if not true (See Exercise 6.4.3). However the joint distribution of bivariate normal random variables are determined by their means, variances and covariances. This fact is proved next.

Theorem 6.4.2. *Suppose (X, Y) and (Z, W) are two bivariate normal random variables. If*

$$\begin{aligned} E[X] = E[Z] = \mu_1, &\qquad E[Y] = E[W] = \mu_2 \\ Var[X] = Var[Z] = \sigma_1^2, &\qquad Var[Y] = Var[W] = \sigma_2^2 \\ &\textit{and} \\ Cov[X, Y] = Cov[Z, W] = \sigma_{12} & \end{aligned} \tag{6.4.1}$$

then (X, Y) and (Z, W) have the same joint distribution.

Proof- As (X, Y) and (Z, W) are bivariate normal random variables, given real numbers s, t $sX + tY$ and $sZ + tW$ are normal random variables. Using (6.4.1) and the properties of mean and covariance (see Theorem 6.2.2) we have

$$\begin{aligned} E[sX + tY] &= sE[X] + tE[Y] = s\mu_1 + t\mu_2, \\ E[sZ + tW] &= sE[Z] + tE[W] = s\mu_1 + t\mu_2, \\ Var[sX + tY] &= s^2 Var[X] + t^2 Var[Y] + 2stCov[X, Y] \\ &= s^2\sigma_1^2 + t^2\sigma_2^2 + 2st\sigma_{12}, \\ &\text{and} \\ Var[sZ + tW] &= s^2 Var[Z] + t^2 Var[W] + 2stCov[Z, W] \\ &= s^2\sigma_1^2 + t^2\sigma_2^2 + 2st\sigma_{12}. \end{aligned}$$

From the above, $sX + tY$ and $sZ + tW$ have the same mean and variance. So they have the same distribution (as normal random variables are determined by their mean and variances). By Theorem 6.3.9 (a) they have the same moment generating function. So, the (joint) moment generating function of (X, Y) at (s,t) is

$$M_{X,Y}(s, t) = E[e^{sX+tY}] = M_{sX+tY}(1) = M_{sZ+tW}(1) = E[e^{sZ+tW}] = M_{Z,W}(s, t)$$

Therefore (Z, W) has the same joint m.g.f. as (X,Y) and Theorem 6.3.9 (b) implies that they have the same joint distribution. ■

Though, in general, two variables which are uncorrelated may not be independent, it is a remarkable fact that the two concepts are equivalent for bivariate normal random variables.

Theorem 6.4.3. *Let (X, Y) be a bivariate normal random variable. Then $Cov[X, Y] = 0$ if and only if X and Y are independent.*

Proof - That independence implies a zero covariance is true for any pair of random variables (use Theorem 6.1.10 (e)), so we need to only consider the reverse implication.

Suppose $Cov[X, Y] = 0$. Let μ_X and σ_X^2 denote the expected value and variance of X and μ_Y and σ_Y^2 the corresponding values for Y. Let s and t be real numbers. Then, by the bivariate

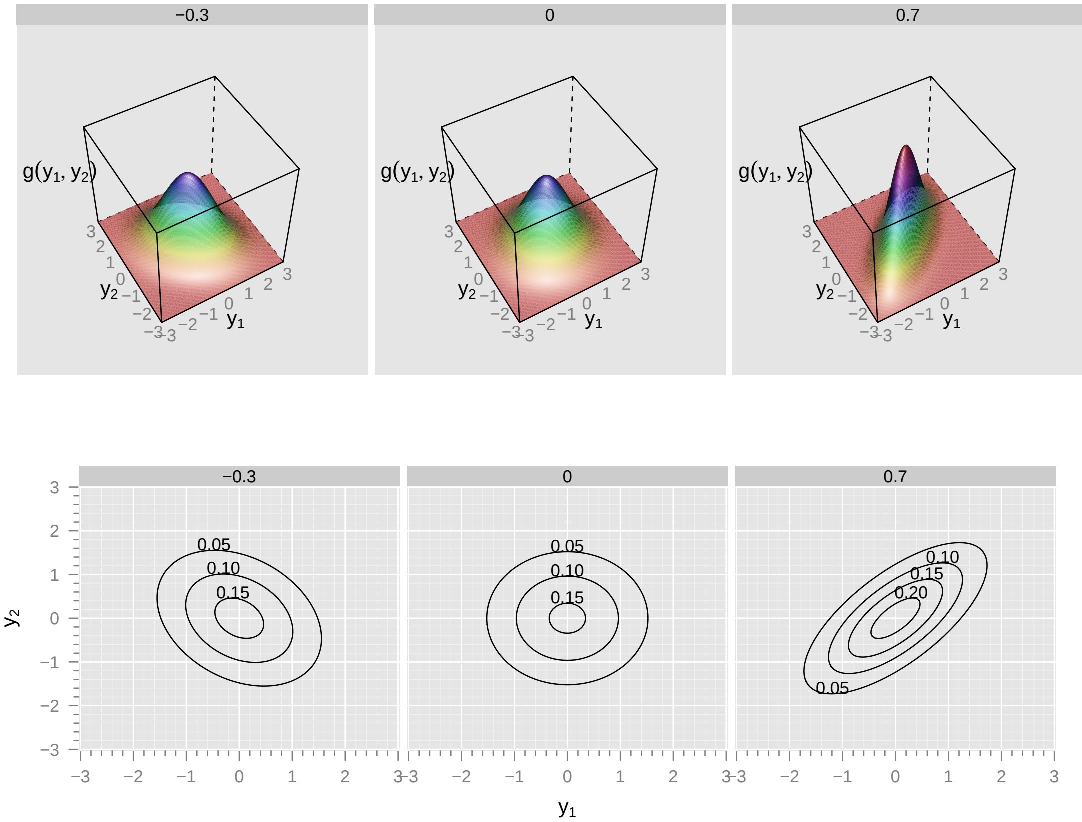

Figure 6.1: The density function of Bivariate Normal distributions. The set of panels on top show a three-dimensional view of the density function for various values of the correlation ρ. The bottom set of panels show contour plots, where each ellipse corresponds to the (y_1, y_2) pairs corresponding to a constant value of $g(y_1, y_2)$.

normality of (X, Y), we know $sX + tY$ is normally distributed. Moreover by properties of expected value and variance we have

$$E[sX + tY] = sE[X] + tE[Y] = s\mu_X + t\mu_Y$$

and

$$Var[sX + tY] = s^2 Var[X] + 2stCov[X, Y] + t^2 Var[Y] = s^2\sigma_X^2 + t^2\sigma_Y^2.$$

That is, $sX + tY \sim \text{Normal}(s\mu_X + t\mu_Y, s^2\sigma_X^2 + t^2\sigma_Y^2)$. So for all $s, t \in \mathbb{R}$

$$\begin{aligned} M_{X,Y}(s,t) &= E[e^{sX+tY}] = M_{sX+tY}(1) = e^{(s\mu_X+t\mu_Y)+(1/2)(s^2\sigma_X^2+t^2\sigma_Y^2)} \\ &= e^{s\mu_x+(1/2)s^2\sigma_X^2} \cdot e^{t\mu_Y+(1/2)t^2\sigma_Y^2} \\ &= M_X(s) \cdot M_Y(t). \end{aligned}$$

Hence by Theorem 6.3.11 X and Y are independent. ■

We conclude this section by finding the joint density of a Bivariate normal random variable. See Figure 6.1 for a graphical display of this density.

THEOREM 6.4.4. *Let* (Y_1, Y_2) *be a bivariate Normal random variable, with* $\mu_1 = E[Y_1], \mu_2 = E[Y_2], 0 \neq \sigma_1^2 = Var[Y_1], 0 \neq \sigma_2^2 = Var[Y_2]$, *and* $\sigma_{12} = Cov[Y_1, Y_2]$. *Assume that the correlation coefficient* $|\rho[Y_1, Y_2]| \neq 1$. *Then the joint probability density function of* (Y_1, Y_2), $g : \mathbb{R}^2 \to [0, \infty)$ *is given by*

$$g(y_1, y_2) = \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right]\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \tag{6.4.2}$$

Proof- Let a, b be two real numbers. We will show that

$$P(Y_1 \le a, Y_2 \le b) = \int_{-\infty}^{a}\int_{-\infty}^{b} g(y_1, y_2)dy_2dy_1. \tag{6.4.3}$$

From the discussion that follows (5.4.1), we can then conclude that the joint density of (Y_1, Y_2) is indeed given by g. To show (6.4.3) we find an alternate description of (Y_1, Y_2) which is the same in distribution. Let Z_1, Z_2 be two independent standard normal random variables. Define

$$\begin{aligned} U &= \sigma_1 Z_1 + \mu_1 \\ V &= \sigma_2(\rho Z_1 + \sqrt{1-\rho^2}Z_2) + \mu_2 \end{aligned} \tag{6.4.4}$$

Let $\alpha, \beta \in \mathbb{R}$. Then

$$\alpha U + \beta V = (\alpha\sigma_1 + \beta\sigma_2\rho)Z_1 + (\beta\sigma_2\sqrt{1-\rho^2})Z_2 + \alpha_1\mu_1 + \beta\mu_2.$$

As Z_1 and Z_2 are independent standard normal random variables by Theorem 6.3.13, $(\alpha\sigma_1 + \beta\sigma_2\rho)Z_1 + (\beta\sigma_2\sqrt{1-\rho^2})Z_2 \sim$ Normal $(0, (\alpha\sigma_1 + \beta\sigma_2\rho)^2 + (\beta\sigma_2\sqrt{1-\rho^2})^2)$. Further using Corollary 5.3.3 (a) we have that $\alpha U + \beta V \sim$ Normal $(\alpha_1\mu_1 + \beta\mu_2, (\alpha\sigma_1 + \beta\sigma_2\rho)^2 + (\beta\sigma_2\sqrt{1-\rho^2})^2)$. As α, β were arbitrary real numbers by Definition 6.4.1, (U, V) is a bivariate normal random variable.

Using Theorem 6.1.8 and Theorem 6.1.10 (d) that,

$$\mu_1 = E[U], \mu_2 = E[V], \sigma_1^2, \text{ and } \mathrm{Var}[U].$$

Also in addition, using Exercise 6.2.16 and Theorem 6.2.2 (f), we have

$$\begin{aligned} Var[V] &= \sigma_2^2\rho^2 Var[Z_1] + \sigma_2^2(1-\rho^2)Var[Z_2] + 2(\sigma_2(\rho + \sqrt{1-\rho^2})Cov[Z_1, Z_2] \\ &= \sigma_2^2\rho^2 + \sigma_2^2(1-\rho^2) + 0 = \sigma_2^2 \\ &\quad \text{and} \\ Cov[U, V] &= Cov[\sigma_1 Z_1 + \mu_1, \sigma_2(\rho Z_1 + \sqrt{1-\rho^2}Z_2)] \\ &= \sigma_1\sigma_2\rho Cov[Z_1, Z_1] + \sigma_1\sigma_2\sqrt{1-\rho^2}Cov[Z_1, Z_2] \\ &= \sigma_1\sigma_2\rho + 0 = \sigma_{12}. \end{aligned}$$

As bivariate normal random variables are by their means and covariances (by Theorem 6.4.2), (Y_1, Y_2) and (U, V) have the same joint distribution. By the above, we have

$$P(Y_1 \le a, Y_2 \le b) \quad = \quad P(U \le a, V \le b). \tag{6.4.5}$$

By elementary algebra we can also infer from (6.4.4)

$$Z_1 = \frac{U - \mu_1}{\sigma_1}, \ Z_2 = \frac{V - \mu_2}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho Z_1}{\sqrt{1-\rho^2}}.$$

So

$$\{U \leq a,\ V \leq b\} \quad = \quad \left\{Z_1 \leq \frac{a-\mu_1}{\sigma_1},\ Z_2 \leq \frac{b-\mu_2}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho Z_1}{\sqrt{1-\rho^2}}\right\}$$

So, using this fact in (6.4.5) we get

$$\begin{aligned} P(Y_1 \leq a, Y_2 \leq b) &= P\left(Z_1 \leq \frac{a-\mu_1}{\sigma_1},\ Z_2 \leq \frac{b-\mu_2}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho Z_1}{\sqrt{1-\rho^2}}\right) \\ &= \int_{-\infty}^{\frac{a-\mu_1}{\sigma_1}} \int_{-\infty}^{\frac{b-\mu_2}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho z_1}{\sqrt{1-\rho^2}}} \frac{\exp(-\frac{z_1^2+z_2^2}{2})}{2\pi} dz_2 dz_1 \end{aligned} \tag{6.4.6}$$

First performing a u-substitution in the inner integral for each fixed z_1,

$$z_2 = \frac{y_2-\mu_2}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho z_1}{\sqrt{1-\rho^2}}$$

yields that the inner integral in (6.4.6) for each $z_1 \in \mathbb{R}$

$$\int_{-\infty}^{\frac{b-\mu_2}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho z_1}{\sqrt{1-\rho^2}}} \frac{\exp(-\frac{z_1^2+z_2^2}{2})}{2\pi} dz_2 = \int_{-\infty}^{b} \frac{\exp(-\frac{z_1^2+\left(\frac{y_2-\mu_2}{\sigma_2\sqrt{1-\rho^2}} - \frac{\rho z_1}{\sqrt{1-\rho^2}}\right)^2}{2})}{2\pi\sigma_2\sqrt{1-\rho^2}} dy_2$$

$$= \int_{-\infty}^{b} \frac{\exp(-\frac{1}{2(1-\rho^2)}[(1-\rho^2)z_1^2 + (\frac{y_2-\mu_2}{\sigma_2} - \rho z_1)^2])}{2\pi\sigma_2\sqrt{1-\rho^2}} dy_2$$

$$= \int_{-\infty}^{b} \frac{\exp(-\frac{1}{2(1-\rho^2)}[z_1^2 + (\frac{y_2-\mu_2}{\sigma_2})^2 - 2\rho(\frac{y_2-\mu_2}{\sigma_2})z_1])}{2\pi\sigma_2\sqrt{1-\rho^2}} dy_2.$$

Substituting the above into (6.4.6), we have

$$P(Y_1 \leq a, Y_2 \leq b) \quad = \quad \int_{-\infty}^{\frac{a-\mu_1}{\sigma_1}} \int_{-\infty}^{b} \frac{\exp(-\frac{1}{2(1-\rho^2)}[z_1^2 + (\frac{y_2-\mu_2}{\sigma_2})^2 - 2\rho(\frac{y_2-\mu_2}{\sigma_2})z_1])}{2\pi\sigma_2\sqrt{1-\rho^2}} dy_2.dz_1 \tag{6.4.7}$$

Performing a u-subsitution

$$z_1 = \frac{y_1-\mu_1}{\sigma_1}$$

on the outer integral above we obtain

$$\begin{aligned} &P(Y_1 \leq a, Y_2 \leq b) \\ &= \int_{-\infty}^{a} \int_{-\infty}^{b} \frac{\exp(-\frac{1}{2(1-\rho^2)}[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + (\frac{y_2-\mu_2}{\sigma_2})^2 - 2\rho(\frac{y_2-\mu_2}{\sigma_2})\left(\frac{y_1-\mu_1}{\sigma_1}\right)])}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} dy_2 dy_1 \end{aligned}$$

Thus we have established (6.4.3). ■

EXERCISES

Ex. 6.4.1. Let X_1, X_2 be two independent Normal random variables with mean 0 and variance 1. Show that (X_1, X_2) is a bivariate normal random variable.

Ex. 6.4.2. Let (X_1, X_2) be a bivariate normal random variable. Assume that the correlation coefficient $|\rho[X_1, X_2]| \neq 1$. Show that X_1 and X_2 are Normal random variables by calculating their marginal densities.

Ex. 6.4.3. Let X_1, X_2 be two independent normal random variables with mean 0 and variance 1. Let (Y_1, Y_2) be a bivariate normal random variable with zero means, variances equal to 1 and correlation $\rho = \rho[Y_1, Y_2]$, with $\rho^2 \neq 1$. Let f be the joint probability density function of (X_1, X_2) and g be the joint probability density function of (Y_1, Y_2). For $0 < \alpha < 1$, let (Z_1, Z_2) be a bivariate random variable with joint density given by

$$h(z_1, z_2) = \alpha g(z_1, z_2) + (1 - \alpha) f(z_1, z_2),$$

for any real numbers z_1, z_2.

(a) Write down the exact expressions for f and g.

(b) Verify that h is indeed a probability density function.

(c) Show that Z_1 and Z_2 are Normal random variables by calculating their marginal densities.

(d) Show that (Z_1, Z_2) is not a bivariate normal random variable.

Ex. 6.4.4. Suppose $X_1, X_2, \ldots, X_n$ are independent and normaly distributed. Let $Y = c_1 X_1 + \cdots + c_n X_n$ and let $Z = d_1 X_1 + \cdots + d_n X_n$ be linear combinations of these variables (for real numbers c_j and d_j). Then (Y, Z) is bivariate normal.

Ex. 6.4.5. Prove Theorem 6.3.13. Specifically, suppose for $i = 1, 2, \ldots, n$ that $X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$ with $X_1, X_2, \ldots, X_n$ independent. Let $a_1, a_2, \ldots, a_n$ be real numbers, not all zero, and let $Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$. Prove that Y is normally distributed and find its mean and variance in terms of the a's, μ's, and σ's.

Ex. 6.4.6. Let (X_1, X_2) be a bivariate Normal random variable. Define

$$\Sigma = \begin{bmatrix} Cov[X_1, X_1] & Cov[X_1, X_2] \\ \\ Cov[X_1, X_2] & Cov[X_2, X_2] \end{bmatrix}$$

$$\text{and } \mu_1 = E[X_1], \mu_2 = E[X_2], \mu_{2\times 1} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}.$$

Σ is referred to as the covariance matrix of (X_1, X_2) and μ is the mean matrix of (X_1, X_2).

(a) Compute $\det(\Sigma)$.

(b) Show that the joint density of (X_1, X_2) can be rewritten in matrix notation as

$$g(x_1, x_2) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}\right)$$

(c)

$$A_{2\times 2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \eta_{2\times 1} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$$

such that a_{ij} are real numbers. Suppose we define

$$Y = AX = \begin{bmatrix} a_{11}X_1 + a_{12}X_2 + \eta_1 \\ a_{21}X_1 + a_{22}X_2 + \eta_2 \end{bmatrix}.$$

Then (Y_1, Y_2) is also a bivariate Normal random variable, with covariance matrix $A\Sigma A^T$ and mean matrix $A\mu + \eta$.
Hint: Compute means, variances and covariances of Y_1, Y_2 and use Theorem 6.4.2