

## 11.2: Mathematical Expectation and General Random Variables

In this unit, we extend the definition and properties of mathematical expectation to the general case. In the process, we note the relationship of mathematical expectation to the Lebesgue integral, which is developed in abstract measure theory. Although we do not develop this theory, which lies beyond the scope of this study, identification of this relationship provides access to a rich and powerful set of properties which have far reaching consequences in both application and theory.

### Extension to the General Case

In the unit on [Distribution Approximations](#), we show that a bounded random variable  $X$  can be represented as the limit of a nondecreasing sequence of simple random variables. Also, a real random variable can be expressed as the difference  $X = X^+ - X^-$  of two nonnegative random variables. The extension of mathematical expectation to the general case is based on these facts and certain basic properties of simple random variables, some of which are established in the unit on expectation for simple random variables. We list these properties and sketch how the extension is accomplished.

#### Definition: almost surely

A condition on a random variable or on a relationship between random variables is said to hold *almost surely*, abbreviated "a.s." iff the condition or relationship holds for all  $\omega$  except possibly a set with probability zero.

#### Basic properties of simple random variables

(E0): If  $X = Y$  a.s. then  $E[X] = E[Y]$ .

(E1):  $E(aI_E) = aP(E)$ .

(E2): *Linearity.*  $X = \sum_{i=1}^n a_i X_i$  implies  $E[X] = \sum_{i=1}^n a_i E[X_i]$

(E3): *Positivity: monotonicity*

a. If  $X \geq 0$  a.s., then  $E[X] \geq 0$ , with equality iff  $X = 0$  a.s. .

b. If  $X \geq Y$  a.s., then  $E[X] \geq E[Y]$ , with equality iff  $X = Y$  a.s. .

(E4): *Fundamental lemma* If  $X \geq 0$  is bounded and  $\{X_n : 1 \leq n\}$  is an a.s. nonnegative, nondecreasing sequence with  $\lim_n X_n(\omega) \geq X(\omega)$  for almost every  $\omega$ , then  $\lim_n E[X_n] \geq E[X]$ .

(E4a): If for all  $n$ ,  $0 \leq X_n \leq X_{n+1}$  a.s. and  $X_n \rightarrow X$  a.s., then  $E[X_n] \rightarrow E[X]$  (i.e., the expectation of the limit is the limit of the expectations).

#### Ideas of the proofs of the fundamental properties

- Modifying the random variable  $X$  on a set of probability zero simply modifies one or more of the  $A_i$  without changing  $P(A_i)$
- Properties (E1) and (E2) are established in the unit on expectation of simple random variables..
- Positivity (E3a) is a simple property of sums of real numbers. Modification of sets of probability zero cannot affect the expectation.
- Monotonicity (E3b) is a consequence of positivity and linearity.

$$X \geq Y \text{ iff } X - Y \geq 0 \text{ a.s. and } E[X] \geq E[Y] \text{ iff } E[X] - E[Y] = E[X - Y] \geq 0$$

- The fundamental lemma (E4) plays an essential role in extending the concept of expectation. It involves elementary, but somewhat sophisticated, use of linearity and monotonicity, limited to nonnegative random variables and positive coefficients. We forgo a proof.
- Monotonicity and the fundamental lemma provide a very simple proof of the monotone convergence theorem, often designated MC. Its role is essential in the extension.

#### Nonnegative random variables

There is a nondecreasing sequence of nonnegative simple random variables converging to  $X$ . Monotonicity implies the integrals of the nondecreasing sequence is a nondecreasing sequence of real numbers, which must have a limit or increase without bound (in which case we say the limit is infinite). We define  $E[X] = \lim E[X_n]$ .

Two questions arise.

Is the limit unique? The approximating sequences for a simple random variable are not unique, although their limit is the same.

Is the definition consistent? If the limit random variable  $X$  is simple, does the new definition coincide with the old?

The fundamental lemma and monotone convergence may be used to show that the answer to both questions is affirmative, so that the definition is reasonable. Also, the six fundamental properties survive the passage to the limit.

As a simple applications of these ideas, consider discrete random variables such as the geometric ( $p$ ) or Poisson ( $\mu$ ), which are integer-valued but unbounded.

#### Example 11.2.1: Unbounded, nonnegative, integer-valued random variables

The random variable  $X$  may be expressed

$$X = \sum_{k=0}^{\infty} k I_{E_k}, \text{ where } E_k = \{X = k\} \text{ with } P(E_k) = p_k$$

Let

$$X_n = \sum_{k=0}^{n-1} k I_{E_k} + n I_{B_n}, \text{ where } B_n = \{X \geq n\}$$

Then each  $X_n$  is a simple random variable with  $X_n \leq X_{n+1}$ . If  $X(\omega) = k$ , then  $X_n(\omega) = k = X(\omega)$  for all  $n \geq k+1$ . Hence,  $X_n(\omega) \rightarrow X(\omega)$  for all  $\omega$ . By monotone convergence,  $E[X_n] \rightarrow E[X]$ . Now

$$E[X_n] = \sum_{k=1}^{n-1} k P(E_k) + n P(B_n)$$

If  $\sum_{k=0}^{\infty} k P(E_k) < \infty$ , then

$$0 \leq n P(B_n) = n \sum_{k=n}^{\infty} P(E_k) \leq \sum_{k=n}^{\infty} k P(E_k) \rightarrow 0 \text{ as } n \rightarrow \infty$$

Hence

$$E[X] = \lim_n E[X_n] = \sum_{k=0}^{\infty} k P(E_k)$$

We may use this result to establish the expectation for the geometric and Poisson distributions.

#### Example 11.2.2: $X \sim \text{geometric}(p)$

We have  $p_k = P(X = k) = q^k p$ ,  $0 \leq k$ . By the result of Example 11.2.1.

$$E[X] = \sum_{k=0}^{\infty} k p q^k = p q \sum_{k=1}^{\infty} k q^{k-1} = \frac{p q}{(1-q)^2} = q/p$$

For  $Y - 1 \sim \text{geometric}(p)$ ,  $p_k = p q^{k-1}$  so that  $E[Y] = \frac{1}{q} E[X] = 1/p$

#### Example 11.2.3: $X \sim \text{poisson}(\mu)$

We have  $p_k = e^{-\mu} \frac{\mu^k}{k!}$ . By the result of Example 11.2.1.

$$E[X] = e^{-\mu} \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} = \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu} e^{\mu} = \mu$$

#### The general case

We make use of the fact that  $X = X^+ - X^-$ , where both  $X^+$  and  $X^-$  are nonnegative. Then

$$E[X] = E[X^+] - E[X^-] \text{ provided at least one of } E[X^+], E[X^-] \text{ is finite}$$

**Definition.** If both  $E[X^+]$  and  $E[X^-]$  are finite,  $X$  is said to be *integrable*.

The term integrable comes from the relation of expectation to the abstract Lebesgue integral of measure theory.

Again, the basic properties survive the extension. The property (E0) is subsumed in a more general uniqueness property noted in the list of properties discussed below.

### Theoretical note

The development of expectation sketched above is exactly the development of the Lebesgue integral of the random variable  $X$  as a measurable function on the basic probability space  $(\Omega, \mathcal{F}, P)$ , so that

$$E[X] = \int_{\Omega} X dP \quad (11.2.1)$$

As a consequence, we may utilize the properties of the general Lebesgue integral. In its abstract form, it is not particularly useful for actual calculations. A careful use of the mapping of probability mass to the real line by random variable  $X$  produces a corresponding mapping of the integral on the basic space to an integral on the real line. Although this integral is also a Lebesgue integral it agrees with the ordinary Riemann integral of calculus when the latter exists, so that ordinary integrals may be used to compute expectations.

### Additional properties

The fundamental properties of simple random variables which survive the extension serve as the basis of an extensive and powerful list of properties of expectation of real random variables and real functions of random vectors. Some of the more important of these are listed in the table in [Appendix E](#). We often refer to these properties by the numbers used in that table.

### Some basic forms

The mapping theorems provide a number of basic integral (or summation) forms for computation.

In general, if  $Z = g(X)$  with distribution functions  $F_X$  and  $F_Z$ , we have the expectation as a Stieltjes integral.

$$E[Z] = E[g(X)] = \int g(t)F_X(dt) = \int uF_Z(du)$$

If  $X$  and  $g(X)$  are absolutely continuous, the Stieltjes integrals are replaced by

$$E[Z] = \int g(t)f_X(t) dt = \int uF_Z(du)$$

where limits of integration are determined by  $f_X$  or  $f_Y$ . Justification for use of the density function is provided by the Radon-Nikodym theorem—property (E19).

If  $X$  is simple, in a primitive form (including canonical form), then

$$E[Z] = E[g(X)] = \sum_{j=1}^m g(c_j)P(C_j)$$

If the distribution for  $Z = g(X)$  is determined by a csort operation, then

$$E[Z] = \sum_{k=1}^n v_k P(Z = v_k)$$

The extension to unbounded, nonnegative, integer-valued random variables is shown in Example 11.2.1, above. The finite sums are replaced by infinite series (provided they converge).

For  $Z = g(X, Y)$ ,

$$E[Z] = E[g(X, Y)] = \int \int g(t, u)F_{XY}(dt du) = \int vF_Z(dv)$$

In the absolutely continuous case

$$E[Z] = E[g(X, Y)] = \int \int g(t, u)f_{XY}(t, u)dudt = \int vf_Z(v)dv$$

For joint simple  $X, Y$  (Section on [Expectation for Simple Random Variables](#))

$$E[Z] = E[g(X, Y)] = \sum_{i=1}^n \sum_{j=1}^m g(t_i, u_j)P(X = t_i, Y = u_j)$$

## Mechanical interpretation and approximation procedures

In elementary mechanics, since the total mass is one, the quantity  $E[X] = \int t f_X(t) dt$  is the location of the center of mass. This theoretically rigorous fact may be derived heuristically from an examination of the expectation for a simple approximating random variable. Recall the discussion of the m-procedure for discrete approximation in the unit on [Distribution Approximations](#). The range of  $X$  is divided into equal subintervals. The values of the approximating random variable are at the midpoints of the subintervals. The associated probability is the probability mass in the subinterval, which is approximately  $f_X(t_i)dx$ , where  $dx$  is the length of

the subinterval. This approximation improves with an increasing number of subdivisions, with corresponding decrease in  $\Delta x$ .  $X_s$  is

$$E[X_s] = \sum_i t_i f_X(t_i) \Delta x \approx \int t f_X(t) dt$$

The approximation improves with increasingly fine subdivisions. The center of mass of the approximating distribution approaches the center of mass of the smooth distribution.

It should be clear that a similar argument for  $g(X)$  leads to the integral expression

$$E[g(X)] = \int g(t) f_X(t) dt$$

This argument shows that we should be able to use tappr to set up for approximating the expectation  $E[g(X)]$  as well as for approximating  $P(g(X) \in M)$ , etc. We return to this in Section.

### Mean values for some absolutely continuous distributions

**Uniform** on  $[a, b]$   $f_X(t) = \frac{1}{b-a}$ ,  $a \leq t \leq b$  The center of mass is at  $(a+b)/2$ . To calculate the value formally, we write

$$E[X] = \int t f_X(t) dt = \frac{1}{b-a} \int_a^b t dt = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

**Symmetric triangular** on  $[a, b]$  The graph of the density is an isosceles triangle with base on the interval  $[a, b]$ . By symmetry, the center of mass, hence the expectation, is at the midpoint  $(a+b)/2$ .

**Exponential**( $\lambda$ )  $f_X(t) = \lambda e^{-\lambda t}$ ,  $0 \leq t$  Using a well known definite integral (see [Appendix B](#)), we have

$$E[X] = \int t f_X(t) dt = \int_0^\infty \lambda t e^{-\lambda t} dt = 1/\lambda$$

**Gamma**( $\alpha, \lambda$ )  $f_X(t) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha t^{\alpha-1} e^{-\lambda t}$ ,  $0 \leq t$  Again we use one of the integrals in [Appendix B](#) to obtain

$$E[X] = \int t f_X(t) dt = \frac{1}{\Gamma(\alpha)} \int_0^\infty \lambda^\alpha t^\alpha e^{-\lambda t} dt = \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} = \alpha/\lambda$$

The last equality comes from the fact that  $\Gamma(\alpha+1) = \alpha \Gamma(\alpha)$ .

**Beta**( $r, s$ ).  $f_X(t) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} t^{r-1} (1-t)^{s-1}$ ,  $0 < t < 1$  We use the fact that

$$\int_0^1 u^{r-1} (1-u)^{s-1} du = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}, \quad r > 0, s > 0.$$

$$E[X] = \int t f_X(t) dt = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \int_0^1 t^r (1-t)^{s-1} dt = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \cdot \frac{\Gamma(r+1)\Gamma(s)}{\Gamma(r+s+1)} = \frac{r}{r+s}$$

**Weibull**( $\alpha, \lambda, v$ ).  $F_X(t) = 1 - e^{-\lambda(t-v)^\alpha}$ ,  $\alpha > 0, \lambda > 0, v \geq 0, t \geq v$ . Differentiation shows

$$f_X(t) = \alpha \lambda (t-v)^{\alpha-1} e^{-\lambda(t-v)^\alpha}, \quad t \geq v$$

First, consider  $Y \sim \text{exponential}(\lambda)$ . For this random variable

$$E[Y^r] = \int_0^\infty t^r \lambda e^{-\lambda t} dt = \frac{\Gamma(r+1)}{\lambda^r}$$

If  $Y$  is exponential (1), then techniques for functions of random variables show that  $[\frac{1}{\lambda} Y]^{1/\alpha} + v \sim \text{Weibull}(\alpha, \lambda, v)$ .

Hence,

$$E[X] = \frac{1}{\lambda^{1/\alpha}} E[Y^{1/\alpha}] + v = \frac{1}{\lambda^{1/\alpha}} \Gamma\left(\frac{1}{\alpha} + 1\right) + v$$

**Normal**( $\mu, \sigma^2$ ) The symmetry of the distribution about  $t = \mu$  shows that  $E[X] = \mu$ . This, of course, may be verified by integration. A standard trick simplifies the work.

$$E[X] = \int_{-\infty}^\infty t f_X(t) dt = \int_{-\infty}^\infty (t - \mu) f_X(t) dt + \mu$$

We have used the fact that  $\int_{-\infty}^\infty f_X(t) dt = 1$ . If we make the change of variable  $x = t - \mu$  in the last integral, the integrand becomes an odd function, so that the integral is zero. Thus,  $E[X] = \mu$ .

## Properties and Computation

The properties in the table in [Appendix E](#) constitute a powerful and convenient resource for the use of mathematical expectation. These are properties of the abstract Lebesgue integral, expressed in the notation for mathematical expectation.

$$E[g(X)] = \int g(X) dP \quad (11.2.2)$$

In the development of additional properties, the four basic properties: (E1) Expectation of indicator functions, (E2) Linearity, (E3) Positivity; monotonicity, and (E4a) Monotone convergence play a foundational role. We utilize the properties in the table, as needed, often referring to them by the numbers assigned in the table.

In this section, we include a number of examples which illustrate the use of various properties. Some are theoretical examples, deriving additional properties or displaying the basis and structure of some in the table. Others apply these properties to facilitate computation

### Example 11.2.4: Probability as expectation

Probability may be expressed entirely in terms of expectation.

- By properties (E1) and positivity (E3a),  $P(A) = E[I_A] \geq 0$ .
- As a special cases of (E1), we have  $P(\Omega) = E[I_\Omega] = 1$
- By the countable sums property (E8),

$$A = \bigcup_i A_i \text{ implies } P(A) = E[I_A] = E[\sum_i I_{A_i}] = \sum_i E[I_{A_i}] = \sum_i P(A_i)$$

Thus, the three defining properties for a probability measure are satisfied.

*Remark.* There are treatments of probability which characterize mathematical expectation with properties (E0) through (E4a), then define  $P(A) = E[I_A]$ . Although such a development is quite feasible, it has not been widely adopted.

### Example 11.2.5: An indicator function pattern

Suppose  $X$  is a real random variable and  $E = X^{-1}(M) = \{\omega : X(\omega) \in M\}$ . Then

$$I_E = I_M(X)$$

To see this, note that  $X(\omega) \in M$  iff  $\omega \in E$ , so that  $I_E(\omega) = 1$  iff  $I_M(X(\omega)) = 1$ .

Similarly, if  $E = X^{-1}(M) \cap Y^{-1}(N)$ , then  $I_E = I_M(X)I_N(Y)$ . We thus have, by (E1).

$$P(X \in M) = E[I_M(X)] \text{ and } P(X \in M, Y \in N) = E[I_M(X)I_N(Y)]$$

### Example 11.2.6: Alternate interpretation of the mean value

$E[(X - c)^2]$  is a minimum iff  $c = E[X]$ , in which case  $E[(X - E[X])^2] = E[X^2] - E^2[X]$

INTERPRETATION. If we approximate the random variable  $X$  by a constant  $c$ , then for any  $\omega$  the error of approximation is  $X(\omega) - c$ . The probability weighted average of the square of the error (often called the *mean squared error*) is  $E[(X - c)^2]$ . This average squared error is smallest iff the approximating constant  $c$  is the mean value.

#### verification

We expand  $(X - c)^2$  and apply linearity to obtain

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

The last expression is a quadratic in  $c$  (since  $E[X^2]$  and  $E[X]$  are constants). The usual calculus treatment shows the expression has a minimum for  $c = E[X]$ . Substitution of this value for  $c$  shows the expression reduces to  $E[X^2] - E^2[X]$ .

A number of inequalities are listed among the properties in the table. The basis for these inequalities is usually some standard analytical inequality on random variables to which the monotonicity property is applied. We illustrate with a derivation of the important Jensen's inequality.

### Example 11.2.7: Jensen's inequality

If  $X$  is a real random variable and  $g$  is a convex function on an interval  $I$  which includes the range of  $X$ , then

#### verification

The function  $g$  is convex on  $I$  iff for each  $t_0 \in [a, b]$  there is a number  $\lambda(t_0)$  such that

$$g(t) \geq g(t_0) + \lambda(t_0)(t - t_0)$$

This means there is a line through  $(t_0, g(t_0))$  such that the graph of  $g$  lies on or above it. If  $a \leq X \leq b$ , then by monotonicity  $E(a) = a \leq E[X] \leq E(b) = b$  (this is the mean value property (E11)). We may choose  $t_0 = E[X] \in I$ . If we designate the constant  $\lambda(E[X])$  by  $c$ , we have

$$g(X) \geq g(E[X]) + c(X - E[X])$$

Recalling that  $E[X]$  is a constant, we take expectation of both sides, using linearity and monotonicity, to get

$$E[g(X)] \geq g(E[X]) + c(E[X] - E[X]) = g(E[X])$$

*Remark.* It is easy to show that the function  $\lambda(\cdot)$  is nondecreasing. This fact is used in establishing Jensen's inequality for conditional expectation.

*The product rule for expectations of independent random variables*

### Example 11.2.8: product rule for simple random variables

Consider an independent pair  $\{X, Y\}$  of simple random variables

$$X = \sum_{i=1}^n t_i I_{A_i} \quad Y = \sum_{j=1}^m u_j I_{B_j} \quad (\text{both in canonical form})$$

We know that each pair  $\{A_i, B_j\}$  is independent, so that  $P(A_i B_j) = P(A_i)P(B_j)$ . Consider the product  $XY$ . According to the pattern described after Example 9 from "Mathematical Expectation: Simple Random Variables."

$$XY = \sum_{i=1}^n t_i I_{A_i} \sum_{j=1}^m u_j I_{B_j} = \sum_{i=1}^n \sum_{j=1}^m t_i u_j I_{A_i B_j}$$

The latter double sum is a primitive form, so that

$$\begin{aligned} E[XY] &= \sum_{i=1}^n \sum_{j=1}^m t_i u_j P(A_i B_j) = \sum_{i=1}^n \sum_{j=1}^m t_i u_j P(A_i)P(B_j) = \left(\sum_{i=1}^n t_i P(A_i)\right) \left(\sum_{j=1}^m u_j P(B_j)\right) \\ &= E[X]E[Y] \end{aligned}$$

Thus the product rule holds for independent simple random variables.

### Example 11.2.9: approximating simple functions for an independent pair

Suppose  $\{X, Y\}$  is an independent pair, with an approximating simple pair  $\{X_s, Y_s\}$ . As functions of  $X$  and  $Y$ , respectively, the pair  $\{X_s, Y_s\}$  is independent. According to Example, above, the product rule  $E[X_s Y_s] = E[X_s]E[Y_s]$  must hold.

### Example 11.2.10. product rule for an independent pair

For  $X \geq 0$ ,  $Y \geq 0$ , there exist nondecreasing sequences  $\{X_n : 1 \leq n\}$  and  $\{Y_n : 1 \leq n\}$  of simple random variables increasing to  $X$  and  $Y$ , respectively. The sequence  $\{X_n Y_n : 1 \leq n\}$  is also a nondecreasing sequence of simple random variables, increasing to  $XY$ . By the monotone convergence theorem (MC)

$$E[X_n] \nearrow E[X], \quad E[Y_n] \nearrow E[Y], \quad \text{and} \quad E[X_n Y_n] \nearrow E[XY]$$

Since  $E[X_n Y_n] = E[X_n]E[Y_n]$  for each  $n$ , we conclude  $E[XY] = E[X]E[Y]$

In the general case,

$$XY = (X^+ - X^-)(Y^+ - Y^-) = X^+ Y^+ - X^+ Y^- - X^- Y^+ + X^- Y^-$$

Application of the product rule to each nonnegative pair and the use of linearity gives the product rule for the pair  $\{X, Y\}$

*Remark.* It should be apparent that the product rule can be extended to any finite independent class.

### Example 11.2.11: the joint distribution of three random variables

The class  $\{X, Y, Z\}$  is independent, with the marginal distributions shown below. Let

$W = g(X, Y, Z) = 3X^2 + 2XY - 3XYZ$ . Determine  $E[W]$ .

```
X = 0:4;
Y = 1:2:7;
Z = 0:3:12;
PX = 0.1*[1 3 2 3 1];
PY = 0.1*[2 2 3 3];
PZ = 0.1*[2 2 1 3 2];
```

```
icalc3                                % Setup for joint dbn for {X,Y,Z}
Enter row matrix of X-values    X
Enter row matrix of Y-values    Y
Enter row matrix of Z-values    Z
Enter X probabilities    PX
Enter Y probabilities    PY
Enter Z probabilities    PZ
Use array operations on matrices X, Y, Z,
PX, PY, PZ, t, u, v, and P
EX = X*PX'                        % E[X]
EX =      2
EX2 = (X.^2)*PX'                  % E[X^2]
EX2 = 5.4000
EY = Y*PY'                        % E[Y]
EY = 4.4000
EZ = Z*PZ'                        % E[Z]
EZ = 6.3000
G = 3*t.^2 + 2*t.*u - 3*t.*u.*v;  % W = g(X,Y,Z) = 3X^2 + 2XY - 2XYZ
```

### Example 11.2.12. a function with a compound definition: truncated exponential

Suppose  $X \sim \text{exponential}(0, 3)$ . Let

$$Z = \begin{cases} X^2 & \text{for } X \leq 4 \\ 16 & \text{for } X > 4 \end{cases} = I_{[0,4]}(X)X^2 + I_{(4,\infty)}(X)16$$

Determine  $E(Z)$ .

**Analytic Solution**

$$\begin{aligned} E[g(X)] &= \int g(t)f_X(t) dt = \int_0^\infty I_{[0,4]}(t)t^2 0.3e^{-0.3t} dt + 16E[I_{(4,\infty)}(X)] \\ &= \int_0^4 t^2 0.3e^{-0.3t} dt + 16P(X > 4) \approx 7.4972 \text{ (by Maple)} \end{aligned}$$

**APPROXIMATION**

To obtain a simple approximation, we must approximate the exponential by a bounded random variable. Since  $P(X > 50) = e^{-15} \approx 3 \cdot 10^{-7}$  we may safely truncate  $X$  at 50.

```
tappr
Enter matrix [a b] of x-range endpoints [0 50]
Enter number of x approximation points 1000
Enter density as a function of t 0.3*exp(-0.3*t)
Use row matrices X and PX as in the simple case
M = X <= 4
G = M.*X.^2 + 16*(1 - M); % g(X)
EG = G*PX' % E[g(X)]
EG = 7.4972
[Z,PZ] = csort(G,PX); % Distribution for Z = g(X)
EZ = Z*PZ' % E[Z] from distribution
EZ = 7.4972
```

Because of the large number of approximation points, the results agree quite closely with the theoretical value.

### Example 11.2.13. stocking for random demand (see exercise 4 from "Problems on functions of random variables")

The manager of a department store is planning for the holiday season. A certain item costs  $c$  dollars per unit and sells for  $p$  dollars per unit. If the demand exceeds the amount  $m$  ordered, additional units can be special ordered for  $s$  dollars per unit ( $s > c$ ). If demand is less than amount ordered, the remaining stock can be returned (or otherwise disposed of) at  $r$  dollars per unit ( $r < c$ ). Demand  $D$  for the season is assumed to be a random variable with Poisson ( $\mu$ ) distribution. Suppose  $\mu = 50$ ,  $c = 30$ ,  $p = 50$ ,  $s = 40$ ,  $r = 20$ . What about  $m$  should the manager order to maximize the expected profit?

#### PROBLEM FORMULATION

Suppose  $D$  is the demand and  $X$  is the profit. Then

$$\text{For } D \leq m, X = D(p - c) - (m - D)(c - r) = D(p - r) + m(r - c)$$

$$\text{For } D > m, X = m(p - c) + (D - m)(p - s) = D(p - s) + m(s - c)$$

It is convenient to write the expression for  $X$  in terms of  $I_M$ , where  $M = (-\infty, m]$ . Thus

$$\begin{aligned} X &= I_M(D)[D(p - r) + m(r - c)] + [1 - I_M(D)][D(p - s) + m(s - c)] \\ &= D(p - s) + m(s - c) + I_M(D)[D(p - r) + m(r - c) - D(p - s) - m(s - c)] \\ &= D(p - s) + m(s - c) + I_M(D)(s - r)(D - m) \end{aligned}$$

Then  $E[X] = (p - c)E[D] + m(s - c) + (s - r)E[I_M(D)D] - (s - r)mE[I_M(D)]$ .

#### Analytic Solution

For  $D \sim \text{Poisson}(\mu)$ ,  $E[D] = \mu$  and  $E[I_M(D)] = P(D \leq m)$

$$E[I_M(D)D] = e^{-\mu} \sum_{k=1}^m k \frac{\mu^k}{k!} = \mu e^{-\mu} \sum_{k=1}^m \frac{\mu^{k-1}}{(k-1)!} = \mu P(D \leq m - 1)$$

Hence,

$$\begin{aligned} E[X] &= (p - s)E[D] + m(s - c) + (s - r)E[I_M(D)D] - (s - r)mE[I_M(D)] \\ &= (p - s)\mu + m(s - c) + (s - r)\mu P(D \leq m - 1) - (s - r)mP(D \leq m) \end{aligned}$$

Because of the discrete nature of the problem, we cannot solve for the optimum  $m$  by ordinary calculus. We may solve for various  $m$  about  $m = \mu$  and determine the optimum. We do so with the aid of MATLAB and the m-function cpoisson.

```
mu = 50;
c = 30;
p = 50;
```



```
s = 40;
r = 20;
m = 45:55;
EX = (p - s)*mu + m*(s - c) + (s - r)*mu*(1 - cpoisson(mu, m))...
-(s - r)*m.*(1 - cpoisson(mu, m+1));
disp([m;EX]')
45.0000    930.8604
46.0000    935.5231
47.0000    939.1895
48.0000    941.7962
49.0000    943.2988
50.0000    943.6750    % Optimum m = 50
51.0000    942.9247
52.0000    941.0699
53.0000    938.1532
54.0000    934.2347
55.0000    929.3886
```

A direct, solution may be obtained by MATLAB, using finite approximation for the Poisson distribution.

#### APPROXIMATION


```
pctest = cpoisson(mu,100)           %Check for suitable value of n
pctest = 3.2001e-10
n = 100;
t = 0:n;
pD = ipoisson(mu,t);
for i = 1:length(m)                  % Step by step calculation for various m
    M = t > m(i);
    G(i,:) = t*(p - r) - M.*(t - m(i)).*(s - r) - m(i)*(c - r);
end
EG = G*pD';                        % Value agree with theoretical to four decimal
```

An advantage of the second solution, based on simple approximation to  $D$ , is that the distribution of gain for each  $m$  could be studied — e.g., the maximum and minimum gains.

—□

#### Example 11.2.14. a jointly distributed pair

Suppose the pair  $\{X, Y\}$  has joint density  $f_{XY}(t, u) = 3u$  on the triangular region bounded by  $u = 0$ ,  $u = 1 + t$ ,  $u = 1 - t$  (see Figure 11.2.1). Let  $Z = g(X, Y) = X^2 + 2XY$ . Determine  $E[Z]$ .

 Figure 1 is a density drawing, with a horizontal axis labeled as t, and a vertical axis labeled as u. A triangle of width 2, with a base sitting on the horizontal axis, from t=-1 to t=1. The third point of the triangle (the one not on the horizontal axis) is directly above, on the vertical axis. The drawing of the triangle is thus divided in equal halves by the vertical axis. The side of the triangle on the horizontal axis has no direct label. The side of the triangle on the left is labeled u = 1 + t, and the side of the triangle on the right is labeled u = 1 - t. A caption below the triangle reads  $f_{xy}(t, u) = 3u$  on the triangle.

**Figure 11.2.1.** The density for Example 11.2.14.

#### Analytic Solution

$$E[Z] = \int \int (t^2 + 2tu) f_{XY}(t, u) \, du \, dt$$

$$= 3 \int_{-1}^0 \int_0^{1+t} (t^2 u + 2tu^2) \, du \, dt + 3 \int_0^1 \int_0^{1-t} (t^2 u + 2tu^2) \, du \, dt = 1/10$$

## APPROXIMATION

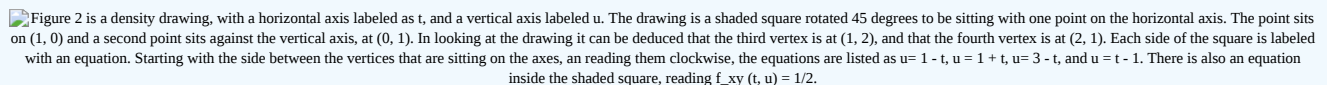
```
tuappr
Enter matrix [a b] of X-range endpoints [-1 1]
Enter matrix [c d] of Y-range endpoints [0 1]
Enter number of X approximation points 400
Enter number of Y approximation points 200
Enter expression for joint density 3*u.*(u<=min(1+t,1-t))
Use array operations on X, Y, PX, PY, t, u, and P
G = t.^2 + 2*t.*u;           % g(X,Y) = X^2 + 2XY
EG = total(G.*P)             % E[g(X,Y)]
EG = 0.1006                  % Theoretical value = 1/10
[Z, PZ] = csort(G,P);        % Distribution for Z
EZ = Z*PZ'                   % E[Z] from distribution
EZ = 0.1006
```

### Example 11.2.15. A function with a compound definition

The pair  $\{X, Y\}$  has joint density  $f_{XY}(t, u) = 1/2$  on the square region bounded to  $u = 1 + t, u = 1 - t, u = 3 - t$ , and  $u = t - 1$  (see Figure 11.2.2).

$$W = \begin{cases} X & \text{for } \max\{X, Y\} \leq 1 \\ 2Y & \text{for } \max\{X, Y\} > 1 \end{cases} = I_Q(X, Y)X + I_{Q^c}(X, Y)2Y$$

where  $Q = \{(t, u) : \max\{t, u\} \leq 1\} = \{(t, u) : t \leq 1, u \leq 1\}$ . Determine  $E[W]$ .

Figure 2 is a density drawing, with a horizontal axis labeled as t, and a vertical axis labeled u. The drawing is a shaded square rotated 45 degrees to be sitting with one point on the horizontal axis. The point sits on (1, 0) and a second point sits against the vertical axis, at (0, 1). In looking at the drawing it can be deduced that the third vertex is at (1, 2), and that the fourth vertex is at (2, 1). Each side of the square is labeled with an equation. Starting with the side between the vertices that are sitting on the axes, an reading them clockwise, the equations are listed as u = 1 - t, u = 1 + t, u = 3 - t, and u = t - 1. There is also an equation inside the shaded square, reading  $f_{xy}(t, u) = 1/2$ .

**Figure 11.2.2.** The density for Example 11.2.15

### Analytic Solution

The intersection of the region  $Q$  and the square is the set for which  $0 \leq t \leq 1$  and  $1 - t \leq u \leq 1$ . Reference to the figure shows three regions of integration.

$$E[W] = \frac{1}{2} \int_0^1 \int_{1-t}^1 t \, du \, dt + \frac{1}{2} \int_0^1 \int_1^{1+t} 2u \, du \, dt + \frac{1}{2} \int_1^2 \int_{t-1}^{3-t} 2u \, du \, dt = 11/6 \approx 1.8333$$

## APPROXIMATION

```
tuappr
Enter matrix [a b] of X-range endpoints [0 2]
Enter matrix [c d] of Y-range endpoints [0 2]
Enter number of X approximation points 200
Enter number of Y approximation points 200
Enter expression for joint density ((u<=min(t+1,3-t))& ...
(u>=max(1-t,t-1))/2
Use array operation on X, Y, PX, PY, t, u, and P
M = max(t,u)<=1;
G = t.*M + 2*u.*(1 - M);      % Z = g(X,Y)
EG = total(G.*P)              % E[g(X,Y)]
EG = 1.8340                   % Theoretical 11/6 = 1.8333
[Z,PZ] = csort(G,P);          % Distribution for Z
```

```
EZ = dot(Z,PZ)
EZ = 1.8340
```

```
% E[Z] from distribution
```

### Special forms for expectation

The various special forms related to property (E20a) are often useful. The general result, which we do not need, is usually derived by an argument which employs a general form of what is known as Fubini's theorem. The special form (E20b)

$$E[X] = \int_{-\infty}^{\infty} [u(t) - F_X(t)] dt$$

may be derived from (E20a) by use of integration by parts for Stieltjes integrals. However, we use the relationship between the graph of the distribution function and the graph of the quantile function to show the equivalence of (E20b) and (E20f). The latter property is readily established by elementary arguments.

#### Example 11.2.16. The property (e20f)

If  $Q$  is the quantile function for the distribution  $F_X$ , then

$$E[g(X)] = \int_0^1 g(Q(u)) du$$

#### VERIFICATION

If  $Y = Q(U)$ , where  $U \sim$  uniform on  $(0, 1)$ , then  $Y$  has the same distribution as  $X$ . Hence,

$$E[g(X)] = E[g(Q(U))] = \int g(Q(u)) f_U(u) du = \int_0^1 g(Q(u)) du$$

#### Example 11.2.17. Reliability and expectation

In reliability, if  $X$  is the life duration (time to failure) for a device, the reliability function is the probability at any time  $t$  the device is still operative. Thus

$$R(t) = P(X > t) = 1 - F_X(t)$$

According to property (E20b)

$$E[X] = \int_0^{\infty} R(t) dt$$

#### Example 11.2.18. Use of the quantile function

Suppose  $F_X(t) = t^a$ ,  $a > 0$ ,  $0 \leq t \leq 1$ . Then  $Q(u) = u^{1/a}$ ,  $0 \leq u \leq a$ .

$$E[X] = \int_0^1 u^{1/a} du = \frac{1}{1+1/a} = \frac{a}{a+1}$$

The same result could be obtained by using  $f_X(t) = F'_X(t)$  and evaluating  $\int t f_X(t) dt$ .

#### Example 11.2.19. Equivalence of (e20b) and (e20f)

For the special case  $g(X) = X$ . Figure 3(a) shows  $\int_0^1 Q(u) du$  is the difference in the shaded areas

$$\int_0^1 Q(u) du = \text{Area } A - \text{Area } B$$

The corresponding graph of the distribution function  $F$  is shown in Figure 11.2.3(b). Because of the construction, the areas of the regions marked  $A$  and  $B$  are the same in the two figures. As may be seen,

$$\text{Area } A = \int_0^{\infty} [1 - F(t)] dt \quad \text{and} \quad \text{Area } B = \int_{-\infty}^0 F(t) dt$$

Use of the unit step function  $u(t) = 1$  for  $t > 0$  and  $0$  for  $t < 0$  (defined arbitrarily at  $t = 0$ ) enables us to combine the two expressions to get

$$\int_0^1 Q(u) du = \text{Area } A - \text{Area } B = \int_{-\infty}^{\infty} [u(t) - F(t)] dt$$

Figure three contains two graphs. The first graph has a horizontal axis labeled  $t$ , and a vertical axis labeled  $u$ . The large label of the graph reads,  $u = Q(t)$ . A dashed vertical line along  $t = 1$  bounds an increasing curved plot. The curve starts with a vertical asymptote along the vertical axis below the horizontal axis, and as it approaches the horizontal axis, the slope becomes more shallow. The curve's slope shallows until it is midway in horizontal distance between the vertical axis and the dashed vertical line. At this point, the slope begins to increase again, until it reaches a vertical asymptote along the dashed line at  $t = 1$ . The horizontal and vertical axes, along with the curve itself, create a bounded shape. A small right triangle loosely fits this bounded shape, and is labeled as B. The dashed line, horizontal axis, and the segment of the curve above the horizontal axis create a larger bounded shape, and a larger right triangle loosely fits this bounded shape, labeled A. The second graph is roughly similar. The axes are in the same place, but with this figure, a dashed line is now drawn horizontally along  $u = 1$ . A curve of the same shape now begins as a horizontal asymptote along the  $t$ -axis. It increases in slope at an increasing rate for half of the vertical distance and then decreases in slope back to a horizontal asymptote at  $u = 1$ . The same triangles fitting the same bounded regions as in the first figure are used in the second figures, only because of the rotated nature of the new curve, these triangles are rotated in the same fashion.

Figure 11.2.3. Equivalence of properties (E20b) and (E20f).

Property (E20c) is a direct result of linearity and (E20b), with the unit step functions cancelling out.

#### Example 11.2.20. Property (e20d) useful inequalities

Suppose  $X \geq 0$ . Then

$$\sum_{n=0}^{\infty} P(X \geq n+1) \leq E[X] \leq \sum_{n=0}^{\infty} P(X \geq n) \leq N \sum_{k=0}^{\infty} P(X \geq kN) \quad , \text{ for all } N \geq 1$$

#### VERIFICATION

For  $X \geq 0$ , by (E20b)

$$E[X] = \int_0^{\infty} [1 - F(t)] dt = \int_0^{\infty} P(X > t) dt$$

Since  $F$  can have only a countable number of jumps on any interval and  $P(X > t)$  and  $P(X \geq t)$  differ only at jump points, we may assert

$$\int_a^b P(X > t) dt = \int_a^b P(X \geq t) dt$$

For each nonnegative integer  $n$ , let  $E_n = [n, n+1]$ . By the countable additivity of expectation

$$E[X] = \sum_{n=0}^{\infty} E[I_{E_n} X] = \sum_{n=0}^{\infty} \int_{E_n} P(X \geq t) dt$$

Since  $P(X \geq t)$  is decreasing with  $t$  and each  $E_n$  has unit length, we have by the mean value theorem

$$P(X \geq n+1) \leq E[I_{E_n} X] \leq P(X \geq n)$$

The third inequality follows from the fact that

$$\int_{kN}^{(k+1)N} P(X \geq t) dt \leq N \int_{E_{kN}} P(X \geq t) dt \leq NP(X \geq kN)$$

*Remark.* Property (E20d) is used primarily for theoretical purposes. The special case (E20e) is more frequently used.

#### Example 11.2.21. Property (e20e)

If  $X$  is nonnegative, integer valued, then

$$E[X] = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$$

#### VERIFICATION

The result follows as a special case of (E20d). For integer valued random variables,

$$P(X \geq t) = P(X \geq n) \text{ on } E_n \text{ and } P(X \geq t) = P(X > n) = P(X \geq n+1) \text{ on } E_{n+1}$$

An elementary derivation of (E20e) can be constructed as follows.

#### Example 11.2.22. (e20e) for integer-valued random variables

By definition

$$E[X] = \sum_{k=1}^{\infty} kP(X = k) = \lim_n \sum_{k=1}^n kP(X = k)$$

Now for each finite  $n$ ,

$$\sum_{k=1}^n kP(X = k) = \sum_{k=1}^n \sum_{j=1}^k P(X = k) = \sum_{j=1}^n \sum_{k=j}^n P(X = k) = \sum_{j=1}^n P(X \geq j)$$

Taking limits as  $n \rightarrow \infty$  yields the desired result.

**Example 11.2.23. the geometric distribution**

Suppose  $X \sim \text{geometric}(p)$ . Then  $P(X \geq k) = q^k$ . Use of (E20e) gives

$$E[X] = \sum_{k=1}^{\infty} q^k = q \sum_{k=0}^{\infty} q^k = \frac{q}{1-q} = q/p$$

This page titled [11.2: Mathematical Expectation and General Random Variables](#) is shared under a [CC BY 3.0](#) license and was authored, remixed, and/or curated by [Paul Pfeiffer](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.