Applications of Integration

6

6.8 Probability

Probability

Calculus plays a role in the analysis of random behavior. Suppose we consider the cholesterol level of a person chosen at random from a certain age group, or the height of an adult female chosen at random, or the lifetime of a randomly chosen battery of a certain type.

Such quantities are called **continuous random variables** because their values actually range over an interval of real numbers, although they might be measured or recorded only to the nearest integer.

Probability

We might want to know the probability that a blood cholesterol level is greater than 250, or the probability that the height of an adult female is between 60 and 70 inches, or the probability that the battery we are buying lasts between 100 and 200 hours.

If X represents the lifetime of that type of battery, we denote this last probability as follows:

$$P(100 \leq X \leq 200)$$

Probability

According to the frequency interpretation of probability, this number is the long-run proportion of all batteries of the specified type whose lifetimes are between 100 and 200 hours.

Since it represents a proportion, the probability naturally falls between 0 and 1. Every continuous random variable *X* has a **probability density function** *f*.

This means that the probability that *X* lies between *a* and *b* is found by integrating *f* from *a* to *b*:

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)\, dx \qquad \boxed{1}$$

Probability

For example, Figure 1 shows the graph of a model for the probability density function *f* for a random variable *X* defined to be the height in inches of an adult female in the United States (according to data from the National Health Survey).

The probability that the height of a woman chosen at random from this population is between 60 and 70 inches is equal to the area under the graph of *f* from 60 to 70.

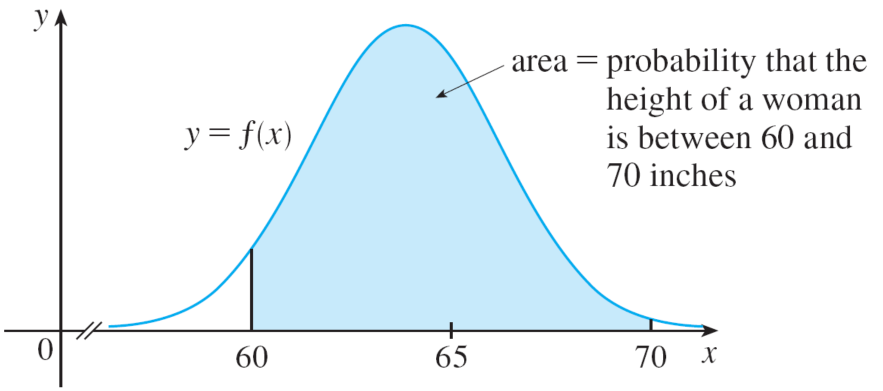

Figure 1

Probability density function for the height of an adult female

Probability

In general, the probability density function *f* of a random variable *X* satisfies the condition $f(x) \geq 0$ for all *x*.

Because probabilities are measured on a scale from 0 to 1, it follows that

2 $$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Example 1

Let $f(x) = 0.006x(10 - x)$ for $0 \leq x \leq 10$ and $f(x) = 0$ for all other values of x.

(a) Verify that f is a probability density function.

(b) Find $P(4 \leq X \leq 8)$.

Solution:

(a) For $0 \leq x \leq 10$ we have $0.006x(10 - x) \geq 0$, so $f(x) \geq 0$ for all x. We also need to check that Equation 2 is satisfied:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{10} 0.006x(10 - x)\,dx$$

$$= 0.006 \int_{0}^{10} (10x - x^2)\,dx$$

Example 1 – *Solution*

cont'd

$$= 0.006\left[5x^2 - \tfrac{1}{3}x^3\right]_0^{10}$$

$$= 0.006\left(500 - \tfrac{1000}{3}\right)$$

$$= 1$$

Therefore *f* is a probability density function.

Example 1 – *Solution*

(b) The probability that *X* lies between 4 and 8 is

$$P(4 \leqslant X \leqslant 8) = \int_4^8 f(x)\,dx$$

$$= 0.006 \int_4^8 (10x - x^2)\,dx$$

$$= 0.006\left[5x^2 - \tfrac{1}{3}x^3\right]_4^8$$

$$= 0.544$$

Average Values

Average Values

Suppose you're waiting for a company to answer your phone call and you wonder how long, on average, you can expect to wait.

Let $f(t)$ be the corresponding density function, where t is measured in minutes, and think of a sample of N people who have called this company.

Most likely, none of them had to wait more than an hour, so let's restrict our attention to the interval $0 \leq t \leq 60$.

Average Values

Let's divide that interval into n intervals of length Δt and endpoints 0, t_1, t_2, . . ., t_{60}. (Think of Δt as lasting a minute, or half a minute, or 10 seconds, or even a second.)

The probability that somebody's call gets answered during the time period from t_{i-1} to t_i is the area under the curve $y = f(t)$ from t_{i-1} to t_i, which is approximately equal to $f(\bar{t}_i)\,\Delta t$. (This is the area of the approximating rectangle in Figure 3, where $\bar{t}_i$ is the midpoint of the interval.)

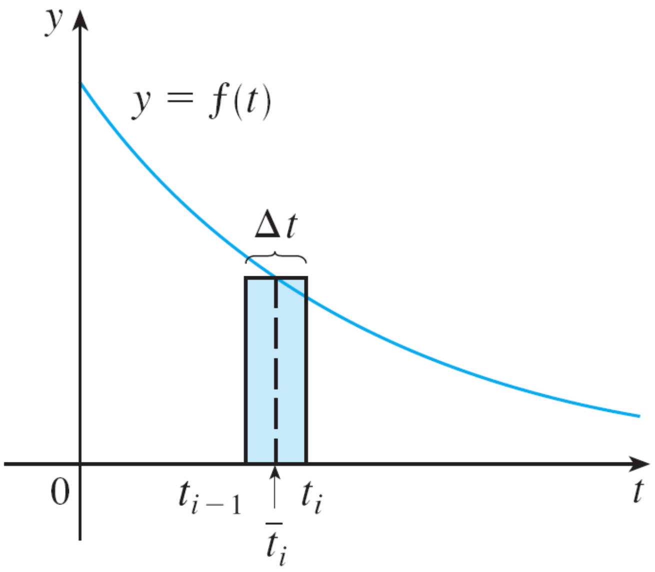

Figure 3

Average Values

Since the long-run proportion of calls that get answered in the time period from t_{i-1} to t_i is $f(\bar{t}_i)\,\Delta t$, we expect that, out of our sample of N callers, the number whose call was answered in that time period is approximately $N\,f(\bar{t}_i)\,\Delta t$ and the time that each waited is about $\bar{t}_i$.

Therefore the total time they waited is the product of these numbers: approximately $\bar{t}_i\,[N\,f(\bar{t}_i)\,\Delta t]$.

Adding over all such intervals, we get the approximate total of everybody's waiting times:

$$\sum_{i=1}^{n} N\bar{t}_i\,f(\bar{t}_i)\,\Delta t$$

Average Values

If we now divide by the number of callers *N*, we get the approximate *average* waiting time:

$$\sum_{i=1}^{n} \bar{t}_i f(\bar{t}_i)\,\Delta t$$

We recognize this as a Riemann sum for the function *t f*(*t*). As the time interval shrinks (that is, $\Delta t \to 0$ and $n \to \infty$), this Riemann sum approaches the integral

$$\int_0^{60} t\,f(t)\,dt$$

This integral is called the *mean waiting time.*

Average Values

In general, the **mean** of any probability density function *f* is defined to be

$$\mu = \int_{-\infty}^{\infty} x\,f(x)\,dx$$

The mean can be interpreted as the long-run average value of the random variable *X*. It can also be interpreted as a measure of centrality of the probability density function.

The expression for the mean resembles an integral we have seen before.

Average Values

If $\mathcal{R}$ is the region that lies under the graph of *f*, we know that the *x*-coordinate of the centroid of $\mathcal{R}$ is

$$\bar{x} = \frac{\int_{-\infty}^{\infty} x\, f(x)\, dx}{\int_{-\infty}^{\infty} f(x)\, dx} = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu$$

because of Equation 2.

So a thin plate in the shape of $\mathcal{R}$ balances at a point on the vertical line $x = \mu$. (See Figure 4.)

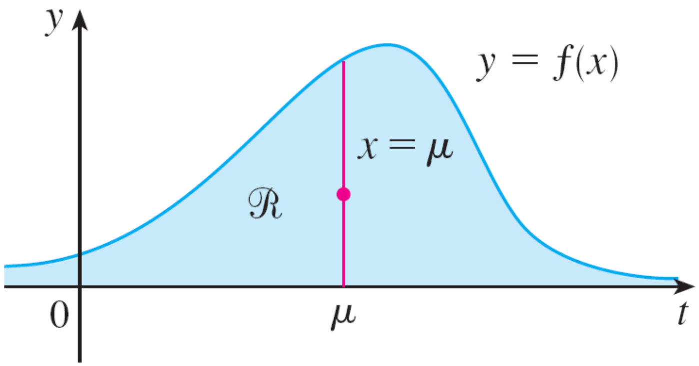

Figure 4

$\mathcal{R}$ balances at a point on the line $x = \mu$

Example 3

Find the mean of the exponential distribution of Example 2:

$$f(t) = \begin{cases} 0 & \text{if } t < 0 \\ ce^{-ct} & \text{if } t \geq 0 \end{cases}$$

Solution:

According to the definition of a mean, we have

$$\mu = \int_{-\infty}^{\infty} t f(t)\, dt$$

$$= \int_{0}^{\infty} tce^{-ct}\, dt$$

Example 3 – *Solution*

cont'd

To evaluate this integral we use integration by parts, with $u = t$ and $dv = ce^{-ct}\,dt$:

$$\int_0^\infty tce^{-ct}\,dt = \lim_{x\to\infty}\int_0^x tce^{-ct}\,dt$$

$$= \lim_{x\to\infty}\left(-te^{-ct}\Big]_0^x + \int_0^x e^{-ct}\,dt\right)$$

$$= \lim_{x\to\infty}\left(-xe^{-cx} + \frac{1}{c} - \frac{e^{-cx}}{c}\right)$$

$$= \frac{1}{c}$$

Example 3 – *Solution*

cont'd

The mean is $\mu = 1/c$, so we can rewrite the probability density function as

$$f(t) = \begin{cases} 0 & \text{if } t < 0 \\ \mu^{-1} e^{-t/\mu} & \text{if } t \geq 0 \end{cases}$$

Average Values

Another measure of centrality of a probability density function is the *median.*

That is a number *m* such that half the callers have a waiting time less than *m* and the other callers have a waiting time longer than *m*.

In general, the **median** of a probability density function is the number *m* such that

$$\int_m^\infty f(x)\,dx = \tfrac{1}{2}$$

This means that half the area under the graph of *f* lies to the right of *m*.

Normal Distributions

Normal Distributions

Many important random phenomena—such as test scores on aptitude tests, heights and weights of individuals from a homogeneous population, annual rainfall in a given location—are modeled by a **normal distribution**.

This means that the probability density function of the random variable *X* is a member of the family of functions

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \qquad \boxed{3}$$

Normal Distributions

You can verify that the mean for this function is μ. The positive constant σ is called the **standard deviation**; it measures how spread out the values of X are.

From the bell-shaped graphs of members of the family in Figure 5, we see that for small values of σ the values of X are clustered about the mean, whereas for larger values of σ the values of X are more spread out.

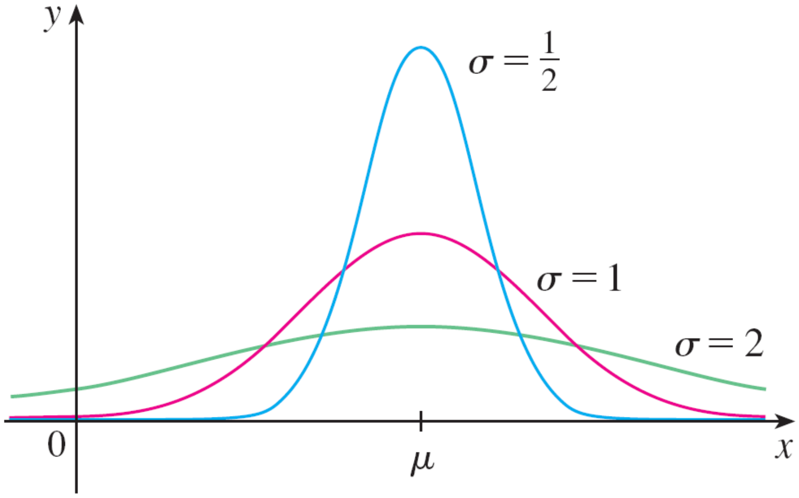

Figure 5

Normal distributions

Normal Distributions

Statisticians have methods for using sets of data to estimate μ and σ .

The factor 1/($\sigma\sqrt{2\pi}$) is needed to make *f* a probability density function. In fact, it can be verified using the methods of multivariable calculus that

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx = 1$$

Example 5

Intelligence Quotient (IQ) scores are distributed normally with mean 100 and standard deviation 15. (Figure 6 shows the corresponding probability density function.)

(a) What percentage of the population has an IQ score between 85 and 115?

(b) What percentage of the population has an IQ above 140?

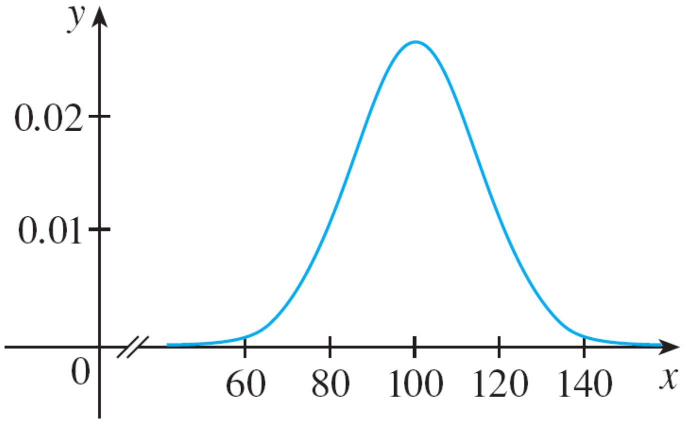

Figure 6

Distribution of IQ scores

Example 5(a) – *Solution*

Since IQ scores are normally distributed, we use the probability density function given by Equation 3 with μ = 100 and σ = 15:

$$P(85 \leqslant X \leqslant 115) = \int_{85}^{115} \frac{1}{15\sqrt{2\pi}} e^{-(x-100)^2/(2\cdot 15^2)}\,dx$$

The function *y* = e^{-x^2} doesn't have an elementary antiderivative, so we can't evaluate the integral exactly.

Example 5(a) – *Solution*

cont'd

But we can use the numerical integration capability of a calculator or computer (or the Midpoint Rule or Simpson's Rule) to estimate the integral.

Doing so, we find that

$$P(85 \leq X \leq 115) \approx 0.68$$

So about 68% of the population has an IQ between 85 and 115, that is, within one standard deviation of the mean.

Example 5(b) – *Solution*

cont'd

The probability that the IQ score of a person chosen at random is more than 140 is

$$P(X > 140) = \int_{140}^{\infty} \frac{1}{15\sqrt{2\pi}} e^{-(x-100)^2/450}\,dx$$

To avoid the improper integral we could approximate it by the integral from 140 to 200. (It's quite safe to say that people with an IQ over 200 are extremely rare.)

Then

$$P(X > 140) \approx \int_{140}^{200} \frac{1}{15\sqrt{2\pi}} e^{-(x-100)^2/450}\,dx$$

$$\approx 0.0038$$

Therefore about 0.4% of the population has an IQ over 140.