0

·

5.1 Introduction

- Consider discrete random variable in Chap. 4.
- There also exist random variables whose set of possible values is uncountable.
- Two examples would be the time that a train arrives at a specified stop and the lifetime of a transistor.
- Let X be such a random variable.
- We say that X is a *continuous* random variable if there exist a nonnegative function f, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

  $$P\{X \in B\} = \int_B f(x)dx \qquad (1.1)$$

  The function f is called the *probability density function* of the random variable X (see Fig. 5.1).

- $$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx$$
- Letting $B = [a, b]$,
  $$P\{a \le X \le b\} = \int_a^b f(x)dx \qquad (1.2)$$
- Let $a = b$ in Eq. (1.2), we obtain
  $$P\{X = a\} = \int_a^a f(x)dx = 0$$
- In words, this equation states that the probability that a continuous random variable will assume any fixed value is zero.
- Distribution function:
  $$P\{X < a\} = P\{X \le a\} = F(a) = \int_{-\infty}^{a} f(x)dx$$

Example 5.1a. Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of C?

(b) Find $P\{X > 1\}$.

(a) $C \int_0^2 (4x - 2x^2)dx = 1$

$$C \left[2x^2 - \frac{2x^3}{3} \right] \Bigg|_{x=0}^{x=2} = 1$$

$$C = \frac{3}{8}$$

(b)

$$\begin{aligned} P\{X > 1\} &= \int_1^\infty f(x)dx \\ &= \frac{3}{8} \int_1^2 (4x - 2x^2)dx \\ &= \frac{1}{2} \end{aligned}$$

Example 5.1b. The amount of time, in hours, that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the probability that

(a) a computer will function between 50 and 150 hours before breaking down;

(b) it will function less than 100 hours?

(a) $1 = \int_{-\infty}^{\infty} f(x)dx = \lambda \int_{0}^{\infty} e^{-x/100}dx$

$$1 = -\lambda(100)e^{-x/100}\Big|_0^{\infty} = 100\lambda$$

$$\lambda = \frac{1}{100}$$

$$\begin{aligned} P\{50 < X < 150\} &= \int_{50}^{150} \frac{1}{100} e^{-x/100}dx \\ &= -e^{x/100}\Big|_{50}^{150} \\ &= e^{-1/2} - e^{-3/2} \approx .384 \end{aligned}$$

(b)

$$\begin{aligned} P\{X < 100\} &= \int_{0}^{100} \frac{1}{100} e^{-x/100}dx \\ &= -e^{x/100}\Big|_{0}^{100} \\ &= 1 - e^{-1} \approx .633 \end{aligned}$$

Example 5.1c. The lifetime in hours of a certain kind of radio tube is a random variable having a probability density function given by

$$f(x) = \begin{cases} 0 & x \leq 100 \\ \frac{100}{x^2} & x > 100 \end{cases}$$

What is the probability that exactly 2 of 5 such tubes in a radio set will have to be replaced within the first 150 hours of operation? Assume that the events E_i, $i = 1, 2, 3, 4, 5$, that the ith such tube will have to be replaced within this time, are independent.

- $$\begin{aligned} P(E_i) &= \int_0^{150} f(x)dx \\ &= 100 \int_{100}^{150} x^{-2}dx \\ &= \frac{1}{3} \end{aligned}$$

- From the independence of the events E_i, it follows that the desired probability is

$$\binom{5}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = \frac{80}{243}$$

- The relationship between F and f:

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)dx$$

- Differentiating both sides of the above yields

$$\frac{d}{da}F(a) = f(a)$$

- If ϵ is small, then

$$\begin{aligned} P\{a - \epsilon/2 \le X \le a + \epsilon/2\} &= \int_{a-\epsilon/2}^{a+\epsilon/2} f(x)dx \\ &\approx \epsilon f(a) \end{aligned}$$

- The probability that X will be contained in an interval of length ϵ around the point a is approximately $\epsilon f(a)$.

5.2 Expectation and variance of continuous random variables

- The expected value of a discrete random variable:

$$E[X] = \sum_{x} xP\{X = x\}$$

- If X is a continuous random variable having probability density function $f(x)$, then as

  $f(x)dx \approx P\{x \leq X \leq x+dx\}$ for dx small

- The expected value of X:

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

Example 5.2a. Find $E[X]$ when the density function of X is

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

-

$$\begin{aligned} E[X] &= \int xf(x)dx \\ &= \int_0^1 2x^2 dx \\ &= \frac{2}{3} \end{aligned}$$

Example 5.2b. The density function of X is given by

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $E[e^X]$.

- Let $Y = e^X$.
- For $1 \le x \le e$,

$$\begin{aligned} F_Y(x) &= P\{Y \le x\} \\ &= P\{e^X \le x\} \\ &= P\{X \le \log(x)\} \\ &= \int_0^{\log(x)} f(y)dy \\ &= \log(x) \end{aligned}$$

- By differentiating $F_Y(x)$, the probability density function of Y is given by

$$f_Y(x) = \frac{1}{x} \quad 1 \le x \le e$$

-

$$\begin{aligned} E[e^X] &= E[Y] = \int_{-\infty}^{\infty} x f_Y(x)dx \\ &= \int_1^e dx \\ &= e - 1 \end{aligned}$$

Proposition 5.2.1: If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,

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

An application of Proposition 2.1 to Example 5.2b:

$$E[e^X] = \int_0^1 e^x dx = e - 1$$

Lemma 5.2.1: For a nonnegative random variable Y,

$$E[Y] = \int_0^{\infty} P\{Y > y\}dy$$

Proof:

- Y is a continuous random variable with probability density function f_Y.
- $$\int_0^{\infty} P\{Y > y\}dy = \int_0^{\infty} \int_y^{\infty} f_Y(x)dxdy$$

  where we have used the fact that

$$P\{Y > y\} = \int_y^\infty f_Y(x)dx$$

- Interchanging the order of integration

$$\begin{aligned}\int_0^\infty P\{Y > y\}dy &= \int_0^\infty \left(\int_0^x dy\right) f_Y(x)dx \\ &= \int_0^\infty x f_Y(x)dx \\ &= E[Y]\end{aligned}$$

Proof of Proposition 5.2.1:

- For any function g for which $g(x) \geq 0$, we have from Lemma 5.2.1 that

$$\begin{aligned}E[g(X)] &= \int_0^\infty P\{g(X) > y\}dy \\ &= \int_0^\infty \int_{x:g(x)>y} f(x)dxdy \\ &= \int_{x:g(x)>0} \int_0^{g(x)} dy f(x)dx \\ &= \int_{x:g(x)>0} g(x)f(x)dx\end{aligned}$$

Example 5.2c. A stick of length 1 is split at a point U that is uniformly distributed over (0,1). Determine the expected length of the piece that contains the point p, $0 \leq p \leq 1$.

- $L_p(U)$: The length of the substick that contains the point p.
- Fig. 5.2:

$$L_p(U) = \begin{cases} 1 - U & U < p \\ U & U > p \end{cases}$$

-

$$\begin{aligned} E[L_p(U)] &= \int_0^1 L_p(u)du \\ &= \int_0^p (1-u)du + \int_p^1 udu \\ &= \frac{1}{2} - \frac{(1-p)^2}{2} + \frac{1}{2} - \frac{p^2}{2} \\ &= \frac{1}{2} + p(1-p) \end{aligned}$$

- It is interesting to note that the expected length of the substick containing the point p is maximized when p is the midpoint of the original stick.

Example 5.2d. Suppose that if you are s minutes early for an appointment, then you incur the cost cs, and if you are s minutes late, then you incur the cost ks. Suppose that the travel time from where you presently are to the location of your appointment is a continuous random variable having probability density function f. Determine the time at which you should depart if you want to minimize your expected cost.

- X: The travel time.
- If you leave t minutes before appointment, then your cost $C_t(X)$ is given by

$$C_t(X) = \begin{cases} c(t-X) & \text{if } X \le t \\ k(X-t) & \text{if } X \ge t \end{cases}$$

-

$$\begin{aligned} E[C_t(X)] &= \int_0^\infty C_t(x) f(x) dx \\ &= \int_0^t c(t-x) f(x) dx + \int_t^\infty k(x-t) f(x) dx \\ &= ct \int_0^t f(x) dx - c \int_0^t x f(x) dx \\ &\quad +k \int_t^\infty x f(x) dx - kt \int_t^\infty f(x) dx \end{aligned}$$

- The value of t that minimize $E[C_t(X)]$ can now be obtained by calculus.

- $$\begin{aligned}\frac{d}{dt}E[C_t(X)] &= ctf(t) + cF(t) - ctf(t) \\ &\quad -ktf(t) + ktf(t) - k[1 - f(t)] \\ &= (k + c)F(t) - k\end{aligned}$$

- Equating to zero shows that the minimal expected cost is obtained when you leave t^* minutes before your appointment, where t^* satisfies
$$F(t^*) = \frac{k}{k + c}$$

Corollary 5.2.1: If a and b are constants, then
$$E[aX + b] = aE[X] + b$$

- The variance of a continuous random variable is defined exactly as it is for a discrete one.

- If $E[X] = \mu$, then the variance of X:
$$\mathrm{Var}(X) = E[(X - \mu)^2] = E[X^2] - \mu^2$$

Example 5.2e. Find Var(X) for X as given in Example 5.2a.

- $$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x)dx \\ &= \int_0^1 2x^3 dx \\ &= \frac{1}{2} \end{aligned}$$

- Since $E[X] = \frac{2}{3}$, we obtain that
$$\text{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

For constants a and b:
$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

The next few sections are devoted to a study of some of important classes of continuous random variables.

5.3 The uniform random variable

- A random variable is said to be *uniformly* distributed over the interval (0, 1) if its probability density function is given by

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- For any $0 < a < b < 1$,

$$P\{a \leq X \leq b\} = \int_a^b f(x)dx = b - a$$

- The probability that X is in any particular subinterval of (0, 1) equals the length of that subinterval.

- In general, we say that X is a uniform random variable on the interval (α, β) if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

- Distribution function:

$$F(a) = \begin{cases} 0 & a \leq \alpha \\ \frac{a-\alpha}{\beta-\alpha} & \text{if } \alpha < a < \beta \\ 0 & \text{otherwise} \end{cases}$$

Example 5.3a. Let X be uniformly distributed over (α, β). Find (a) $E[X]$ and (b) $\text{Var}(X)$.

- (a)

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{\beta + \alpha}{2} \end{aligned}$$

(b)

– To find $\text{Var}(X)$, we first calculate $E[X^2]$.

$$\begin{aligned} E[X^2] &= \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} x^2 dx \\ &= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} \\ &= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} \end{aligned}$$

– The variance of a random variable that is uniformly distributed over some interval is the square of the length of that interval divided by 12.

$$\begin{aligned}\mathrm{Var}(X) &= \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{(\alpha+\beta)^2}{4} \\ &= \frac{(\beta-\alpha)^2}{12}\end{aligned}$$

Example 5.3b. If X is uniformly distributed over (0, 10), calculate the probability that (a) $X < 3$, (b) $X > 6$, and (c) $3 < X < 8$.

- (a) $P\{X < 3\} = \int_0^3 \frac{1}{10} dx = \frac{3}{10}$

  (b) $P\{X > 6\} = \int_6^{10} \frac{1}{10} dx = \frac{4}{10}$

  (c) $P\{3 < X < 8\} = \int_3^8 \frac{1}{10} dx = \frac{1}{2}$

Example 5.3c. Buses arrive at a specified stop at 15-minute intervals starting at 7 A.M. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that he waits

(a) less than 5 minutes for a bus;

(b) more than 10 minutes for a bus.

- X: The number of minutes past 7 that the passenger arrives at the stop.

(a) Since X is a uniform random variable over the interval $(0, 30)$, it follows that the passenger will have to wait less than 5 minutes if (and only if) he arrives between 7:10 and 7:15 or between 7:25 and 7:30.

- The desired probability is
  $P\{10 < X < 15\} + P\{25 < X < 30\}$
  $= \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}$

(b) He would have to wait more than 10 minutes if he arrives between 7 and 7:05 or between

7:15 and 7:20, and so the probability is

$$P\{0 < X < 5\} + P\{15 < X < 20\} = \frac{1}{3}$$

The next example was first considered by the French mathematician L. F. Bertrand in 1889 and is often referred to as *Bertrand's paradox*. It is is a geometrical probability problem.

Example 5.3d. Consider a random chord of a circle. What is the probability that the length of the chord will be greater than the side of the equilateral triangle inscribed in that circle?

- The first formulation is as follows:
  - The position of the chord can be determined by its distance from the center of the circle.
  - This distance can vary between 0 and r, the radius of the circle.
  - The length of the chord will be greater than the side of the equilateral triangle

inscribed in the circle if its distance from the center is less than $r/2$.

- Assume that a random chord is one whose distance D from the center is uniformly distributed between 0 and r.
- The probability that it is greater than the side of an inscribed equilateral triangle is

$$P\left\{D < \frac{r}{2}\right\} = \frac{r/2}{r} = \frac{1}{2}$$

- The second formulation of the problem consider an arbitrary chord of the circle; through one end of the chord draw a tangent.
  - The angle θ between the chord and the tangent, which can vary from 0° to 180°, determines the position of the chord (see Fig. 5.4).
  - The length of the chord will be greater than the side of the inscribed equilateral triangle if the angle θ is between 60° and 120°.
  - Assume that a random chord is one whose

angle θ is uniformly distributed between $0°$ and $180°$.

– The desired answer in this formulation is

$$P\{60 < \theta < 120\} = \frac{120 - 60}{180} = \frac{1}{3}$$

5.4 Normal random variables

- X is a normal random variable, or simply that X is normally distributed, with parameters μ and σ^2 if the density of X is given by

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

- The density function is a bell-shaped curve that is symmetric about μ. (see Fig. 5.5).

- The normal distribution was introduced by the French mathematician Abraham de Moivre in 1733 and was used by him to approximate probabilities associated with binomial random variables when the binomial parameter n is large.

- This result was later extended by Laplace and others and now is encompassed in probability theorem known as the central limit theorem.
- The central limit theorem (Chap. 8), one of the two most important results in probability theory, gives a theoretical base to the often noted empirical observation that many random phenomena obey, at least approximately, a normal probability distribution. (The strong law of large number)
- Some examples of this behavior are the height of a man, the velocity in any direction of a molecule in gas, and the error made in measuring a physical quantity.
- To prove that $f(x)$ is indeed a probability density function, we need to show that

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

  – By making the substitution $y = (x -$

$\mu)/\sigma$, we see that

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

– Let $I = \int_{-\infty}^{\infty} e^{-y^2/2}dy$. Then

$$\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} e^{-y^2/2}dy \int_{-\infty}^{\infty} e^{-x^2/2}dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(y^2+x^2)/2}dydx \\
&= \int_{0}^{\infty}\int_{0}^{2\pi} e^{r^2/2} r d\theta dr \\
&= 2\pi \int_{0}^{\infty} re^{-r^2/2}dr \\
&= -2\pi e^{-r^2/2}|_0^{\infty} \\
&= 2\pi
\end{aligned}$$

– $I = \sqrt{2\pi}$.

Example 5.4a. Find (a) $E[X]$ and (b) $\text{Var}(X)$ when X is a random variable with parameters μ and σ^2.

- (a)

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} xe^{-(x-\mu)^2/2\sigma^2}dx$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} [(x-\mu)+\mu] e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu) e^{-(x-\mu)^2/2\sigma^2} dx \\
&\quad + \mu \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y e^{-y^2/2\sigma^2} dy + \mu \int_{-\infty}^{\infty} f(x) dx \\
&= \mu \int_{-\infty}^{\infty} f(x) dx \\
&= \mu
\end{aligned}
$$

(b)

$$
\begin{aligned}
\mathrm{Var}(X) &= E[(X-\mu)^2] \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left[-y e^{-y^2/2} |_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2} dy \right] \\
&= \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\
&= \sigma^2
\end{aligned}
$$

- If X is normally distributed with parameters μ and σ^2, then $Y = \alpha X + \beta$ is normally distributed with parameters $\alpha\mu + \beta$ and $\alpha^2\sigma^2$.
- To show this, suppose $\alpha > 0$. (The verification when $\alpha < 0$ is similar.)
  - F_Y, the cumulative distribution function of the random variable Y, is given by
$$\begin{aligned} F_Y(a) &= P\{\alpha X + \beta \leq a\} \\ &= P\left\{X \leq \frac{a-\beta}{\alpha}\right\} \\ &= F_X\left(\frac{a-\beta}{\alpha}\right) \end{aligned}$$
  - Differentiation yields that the density function of Y is
$$\begin{aligned} f_Y(a) &= \frac{1}{\alpha} f_X\left(\frac{a-\beta}{\alpha}\right) \\ &= \frac{1}{\sqrt{2\pi}\alpha\sigma} \exp\left\{-\left(\frac{a-\beta}{\alpha} - \mu\right)^2 / 2\sigma^2\right\} \\ &= \frac{1}{\sqrt{2\pi}\alpha\sigma} \exp\{-(a - \beta - \alpha\mu)^2 / 2(\alpha\sigma)^2\} \end{aligned}$$

- If X is normally distributed with parameters μ and σ^2, then $Z = (X - \mu)/\sigma$ is normally distributed with parameters 0 and 1.
- Such a random variable Z is said to have the *standard*, or *unit*, normal distribution.
- The cumulative distribution function of a standard normal random variable:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$$

- The value of $\Phi(x)$ for nonnegative x are given in Table 5.1.
- For negative values of x,

$$\Phi(-x) = 1 - \Phi(x) \qquad -\infty < x < \infty$$

- If Z is a standard normal random variable, then

$$P\{Z \leq -x\} = P\{Z > x\} \quad -\infty < x < \infty$$

- Since $Z = (X - \mu)/\sigma$ is a standard normal random variable whenever X is normally distributed with parameters μ and σ^2, it follows that the distribution function of X can be expressed as

$$\begin{aligned} F_X(a) &= P\{X \le a\} \\ &= P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

Example 5.4b. If X is a normal random variable with parameters $\mu = 3$ and $\sigma^2 = 9$, find

(a) $P\{2 < X < 5\}$;

(b) $P\{X > 0\}$;

(c) $P\{|X - 3| > 6\}$.

- (a)

$$\begin{aligned} P\{2 < X < 5\} &= P\left\{\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right\} \\ &= P\left\{-\frac{1}{3} < Z < \frac{2}{3}\right\} \end{aligned}$$

$$\begin{aligned}
&= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) \\
&= \Phi\left(\frac{2}{3}\right) - \left[1 - \Phi\left(\frac{1}{3}\right)\right] \approx .3779
\end{aligned}$$

(b)

$$\begin{aligned}
P\{X > 0\} &= P\left\{\frac{X-3}{3} > \frac{0-3}{3}\right\} \\
&= P\{Z > -1\} \\
&= 1 - \Phi(-1) \\
&= \Phi(1) \approx .8413
\end{aligned}$$

(c)

$$\begin{aligned}
P\{|X-3| > 6\} &= P\{X > 9\} + P\{X < -3\} \\
&= P\left\{\frac{X-3}{3} > \frac{9-3}{3}\right\} \\
&\quad + P\left\{\frac{X-3}{3} < \frac{-3-3}{3}\right\} \\
&= P\{Z > 2\} + P\{Z < -2\} \\
&= 1 - \Phi(2) + \Phi(-2) \\
&= 2[1 - \Phi(2)] \approx .0456
\end{aligned}$$

Example 5.4c. An examination is often regarded as being good (in the sense of determining a valid grade spread for those taking it) if the test scores of those taking the examination can be approximated by a normal density function. (In order words, a graph of the frequency of grade scores should have approximately the bell-shaped form of the normal density.) The instructor often uses the test scores to estimate the normal parameters μ and σ^2 and then assigns the letter grade A to those whose test score is greater than $\mu + \sigma$, B to those whose score is between μ and $\mu + \sigma$, C to those whose score is between $\mu - \sigma$ and μ, D to those whose score is between $\mu - 2\sigma$ and $\mu - \sigma$, and F to those getting a score below $\mu - 2\sigma$. (This is sometimes referred to as grading "on the curve.") Since

- $$P\{X > \mu + \sigma\} = P\left\{\frac{X-\mu}{\sigma} > 1\right\}$$

$$= 1 - \Phi(1) \approx .1587$$

- $$\begin{aligned} P\{\mu < X < \mu + \sigma\} &= P\left\{0 < \frac{X - \mu}{\sigma} < 1\right\} \\ &= \Phi(1) - \Phi(0) \approx .3413 \end{aligned}$$

- $$\begin{aligned} P\{\mu - \sigma < X < \mu\} &= P\left\{-1 < \frac{X - \mu}{\sigma} < 0\right\} \\ &= \Phi(0) - \Phi(-1) \approx .3413 \end{aligned}$$

- $$\begin{aligned} P\{\mu - 2\sigma < X < \mu - \sigma\} &= P\left\{-2 < \frac{X - \mu}{\sigma} < -1\right\} \\ &= \Phi(2) - \Phi(1) \approx .1359 \end{aligned}$$

- $$\begin{aligned} P\{X < \mu - 2\sigma\} &= P\left\{\frac{X - \mu}{\sigma} < -2\right\} \\ &= \Phi(-2) \approx .0228 \end{aligned}$$

- Approximately 16 percent of the class will receive an A grade on the examination, 34 percent a B grade, 34 percent a C grade, and 14 percent a D grade; 2 percent will fail.

Example 5.4d. An expert witness in a paternity suit testifies that the length (in days) of pregnancy (that is, the time from impregnation to the delivery of the child) is approximately normally distributed with parameters $\mu = 270$ and $\sigma^2 = 100$. The defendant in the suit is able to prove that he was out of the country during a period that began 290 days before the birth of the child and ended 240 days before the birth. If the defendant was, in fact, the father of the child, what is the probability that the mother could have had the very long or very short pregnancy indicated by the testimony?

- X: The length of the pregnancy.
- Assume that the defendant is the father.
- The probability that the birth could occur within the indicated period is
  $P\{X > 290 \text{ or } X < 240\}$
  $= P\{X > 290\} + P\{X < 240\}$
  $= P\left\{\frac{X-270}{10} > 2\right\} + P\left\{\frac{X-270}{10} < -3\right\}$

$$= 1 - \Phi(2) + 1 - \Phi(3)$$
$$\approx .0241$$

Example 5.4e.

- Suppose that a binary message– either 0 or 1–must be transmitted by write from location A to location B.
- The data sent over the write are subject to a channel noise disturbance, so to reduce the possibility of error, the value 2 is sent over the wire when the message is 1 and the value -2 is sent when the message is 0.
- If x, $x \pm 2$, is the value sent at location A, then R, the value received at location B, is given by $R = x + N$, where N is the channel noise disturbance.
- When the message is received at location B the receiver decodes it according to the following rule:

$$\text{If } R \geq .5, \text{ then 1 is concluded.}$$

If $R < .5$, then 0 is concluded.

- As the channel noise is often normally distributed, we will determine the error probabilities when N is a unit normal random variable.
- There are two types of errors that can occur:
  - One is that the message 1 can be incorrectly concluded to be 0.
  - The other that 0 is concluded to be 1.
- The first type of error will occur if the message is 1 and $2 + N < .5$, whereas the second will occur if the message is 0 and $-2 + N \geq .5$.
- $$\begin{aligned} P\{\text{error}|\text{message is 1}\} &= P\{N < -1.5\} \\ &= 1 - \Phi(1.5) \approx .0668 \end{aligned}$$
- $$\begin{aligned} P\{\text{error}|\text{message is 0}\} &= P\{N \geq 2.5\} \\ &= 1 - \Phi(2.5) \approx .0062 \end{aligned}$$

- The following inequality for $\Phi(x)$ is of theoretical importance:

$$\frac{1}{\sqrt{2\pi}}\left(\frac{1}{x}-\frac{1}{x^3}\right)e^{-x^2/2} < 1-\Phi(x) < \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}$$

for all $x > 0$ (4.4)

- To prove inequality (4.4)
  - Note the obvious inequality

$$(1-3y^{-4})e^{-y^2/2} < e^{-y^2/2} < (1+y^{-2})e^{-y^2/2}$$

$$\int_x^\infty (1-3y^{-4})e^{-y^2/2}dy < \int_x^\infty e^{-y^2/2}dy < \int_x^\infty (1+y^{-2})e^{-y^2/2}dy$$

$$\frac{d}{dy}[(y^{-1}-y^{-3})e^{-y^2/2}] = -(1-3y^{-4})e^{-y^2/2}$$

$$\frac{d}{dy}[y^{-1}e^{-y^2/2}] = -(1+y^{-2})e^{-y^2/2}$$

for $x > 0$,

$$-(y^{-1}-y^{-3})e^{-y^2/2}|_x^\infty < \int_x^\infty e^{-y^2/2}dy < -y^{-1}e^{-y^2/2}|_x^\infty$$

or

$$(x^{-1}-x^{-3})e^{-y^2/2} < \int_x^\infty e^{-y^2/2}dy < x^{-1}e^{-x^2/2}$$

- $1 - \Phi(x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$ for large x.

5.4.1 The normal approximation to the binomial distribution

- The DEMoivre-Laplace limit theorem states that when n is large, a binomial random variable with parameters n and p will have approximately the same distribution as the normal random variable with the same mean and variance as the binomial.
- This result was proved originally for the special case $p = 1/2$ by DeMoivre in 1733 and was then extended to general p by Laplace in 1812.

The DeMoivre-Laplace limit theorem: If S_n denotes the number of successes that occur when n independent trials, each resulting in a success with probability p, are performed then, for any $a < b$,

$$P\left\{a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right\} \to \Phi(b) - \Phi(a)$$

as $n \to \infty$.

- Poisson approximation and normal approximation.
- The normal approximation will, in general, be quite good for values of n satisfying $np(1-p) \geq 10$.

Example 5.4f. Let X be the number of times that a fair coin, flipped 40 times, lands heads. Find the probability that $X = 20$. Use the normal approximation and then compare it to the exact solution.

- $$\begin{aligned} P\{X=20\} &= P\{19.5 \le X < 20.5\} \\ &= P\left\{\frac{19.5-20}{\sqrt{10}} < \frac{X-20}{\sqrt{10}} < \frac{20.5-20}{\sqrt{10}}\right\} \\ &\approx P\left\{-.16 < \frac{X-20}{\sqrt{10}} < .16\right\} \\ &\approx \Phi(.16) - \Phi(-.16)\} \approx .1272 \end{aligned}$$

- The exact result:

$$P\{X=20\} = \binom{40}{20}\left(\frac{1}{2}\right)^{40} \approx .1254$$

Example 5.4g. The ideal size of a first-year class at a particular college is 150 students. The college, knowing from past experience that on the average only 30 percent of those accepted for admission will actually attend, uses a policy of approving the applications of 450 students. Compute the probability that more than 150 first-year students attend this college.

- X: The number of students that attend.
- X is a binomial(450, .3).

- The normal approximation yields that

$$
\begin{aligned}
P\{X \geq 150.5\} &= P\left\{\frac{X-(450)(.3)}{\sqrt{450(.3)(.7)}} \geq \frac{150.5-(450)(.3)}{\sqrt{450(.3)(.7)}}\right\} \\
&\approx 1-\Phi(1.59) \\
&\approx .0559
\end{aligned}
$$

Example 5.4h. To determine the effectiveness of a certain diet in reducing the amount of cholesterol in the bloodstream, 100 people are put on the diet. After they have been on the diet for a sufficient length of time, their cholesterol count will be taken. The nutritionist running this experiment has decided to endorse the diet if at least 65 percent of the people have a lower cholesterol count after going on the diet. What is the probability that the nutritionist endorses the new diet if, in fact, it has no effect on the cholesterol level?

- X: The number of people whose count is lowered.

- X is a $B(100, 1/2)$.
- The probability that the nutritionist will endorse the diet when it actually has no effect on the cholesterol count:

$$\begin{aligned}\sum_{i=65}^{100} \binom{100}{i} \left(\frac{1}{2}\right)^{100} &= P\{X \geq 64.5\} \\ &= P\left\{\frac{X - (100)(\frac{1}{2})}{\sqrt{100(\frac{1}{2})(\frac{1}{2})}} \geq 2.9\right\} \\ &\approx 1 - \Phi(2.9) \\ &\approx .0019.\end{aligned}$$

Historical notes concerning the normal distribution:

- Abraham De Moire (1733).
- Karl Griedrich Gauss (1777-1855).

5.5 Exponential random variables

- A continuous random variable whose probability density function is given, for some

$\lambda > 0$, by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

is said to be an *exponential* random variable (or, more simply, is said to be exponentially distributed) with parameter λ.

- The cumulative distribution function $F(a)$ of an exponential random variable:

$$\begin{aligned} F(a) &= P\{X \leq a\} \\ &= \int_0^a \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x}\Big|_0^a \\ &= 1 - e^{-\lambda a} \quad a \geq 0 \end{aligned}$$

- Note that $F(\infty) = \int_0^\infty \lambda e^{-\lambda x} dx = 1$.

Example 5.5a. Let X be an exponential random variable with parameter λ. Calculate (a) $E[X]$ and (b) $\mathrm{Var}(X)$.

- (a)

– The density function is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

–

$$\begin{aligned} E[X] &= \int_0^\infty x\lambda e^{-\lambda x} dx \\ &= -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\ &= 0 - \frac{e^{-\lambda x}}{\lambda}\Bigg|_0^\infty \\ &= \frac{1}{\lambda} \end{aligned}$$

- (b)

–

$$\begin{aligned} E[X^2] &= \int_0^\infty x^2\lambda e^{-\lambda x} dx \\ &= -x^2 e^{-\lambda x}|_0^\infty + \int_0^\infty 2xe^{-\lambda x} dx \\ &= 0 + \frac{2}{\lambda}E[X] \\ &= \frac{2}{\lambda^2} \end{aligned}$$

–

$$\mathrm{Var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2$$

$$= \frac{1}{\lambda^2}$$

- The exponential distribution often arises, in practice, as being the distribution of the amount of the time until some specific event occurs.
- The amount of time starting from now until an earthquake occurs, or until a new war breaks out, or until a telephone call you receive turns out to be a wrong number are all random variables that tend in practice to have exponential distributions.

Example 5.5b. Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = \frac{1}{10}$. If someone arrive immediately ahead of you at a public telephone booth, find the probability that you will have to wait

(a) more than 10 minutes;

(b) between 10 and 20 minutes.

- X: The length of the call made by the person in the booth.

(a)

$$\begin{aligned} P\{X > 10\} &= 1 - F(10) \\ &= e^{-1} \approx .368 \end{aligned}$$

(b)

$$\begin{aligned} P\{10 < X < 20\} &= F(20) - F(10) \\ &= e^{-1} - e^{-2} \approx .233 \end{aligned}$$

Memoryless property:

- A nonnegative random variable X is *memoryless* if

$$P\{X > s + t | X > t\} = P\{X > s\} \quad \text{for all } s, t \geq 0 \quad (5.1)$$

- If we think of X as being the lifetime of some instrument, Eq. (5.1) states that the probability that the instrument survives for at least $s + t$ hours, given that it has survived

t hours, is that same as the initial probability that it survives for at least s hours.

- In other words, if the instrument is alive at age t, the distribution of the remaining amount of time that it survives is the same as the original lifetime distribution (that is, it is as if the instrument does not remember that it has already been in use for a time t).
- The condition (5.1) is equivalent to

$$\frac{P\{X > s + t, X > t\}}{P\{X > t\}} = P\{X > s\}$$

or

$$P\{X > s + t\} = P\{X > s\}P\{X > t\} \tag{5.2}$$

Example 5.5c. Consider a post office that is staffed by two clerks. Suppose that when Mr. Smith enters the system, he discovers that Ms. Jones is being served by one of the clerks and Mr. Brown by the other. Suppose also that Mr. Smith is told that his service will begin as soon as either Jones or Brown leaves. If the amount of time that a clerk spends with a customer is exponentially distributed with parameter λ, what is the probability that, of the three customers, Mr. Smith is the last to leave the post office?

- The answer is obtained by reasoning as follows: Consider the time at which Mr. Smith first finds a free clerk. At this point either Ms. Jones or Mr. Brown would have just left and the other one would still be in service.
- However, by the lack of memory of the exponential, it follows that the additional amount of time that this other person (either Jones or Brown) would still have to spend in the post office is exponentially distributed with parameter λ,
- That is, it is the same as if service for this person were just starting at this point. Hence,

by symmetry, the probability that the remaining person finishes before Smith must equal $\frac{1}{2}$.

Uniqueness of memoryless property:

$$\begin{aligned} \overline{F}(x) &= P\{X > x\} \\ \overline{F}(s+t) &= \overline{F}(s)\overline{F}(t) \\ \overline{F}(x) &= e^{-\lambda x} \end{aligned}$$

Example 5.5d. Suppose that the number of miles that a car run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5000-mile trip, what is the probability that he or she will be able to complete the trip without having to replace the car battery? What can be said when the distribution is not exponential?

- It follows by the memoryless property of the exponential distribution that the remaining lifetime (in thousands of miles) of the battery is exponential with parameter $\lambda = \frac{1}{10}$.

- The desired probability is
$$\begin{aligned} P\{\text{remaining lifetime} > 5\} &= 1 - F(5) = e^{-5\lambda} \\ &= e^{-1/2} \approx .604 \end{aligned}$$
- If the lifetime distribution F is not exponential, then the relevant probability is
$$P\{\text{lifetime} > t{+}5 | \text{lifetime} > t\} = \frac{1 - F(t+5)}{1 - F(t)}$$
where t is the number of miles that the battery had been in use prior to the start of the trip.

Laplace distribution: (Double exponential distribution)

- The distribution of a random variable that is equally likely to be either positive or negative and $|X| \sim \exp(\lambda)$.
- The density function:
$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|} \qquad -\infty < x < \infty$$

- The distribution function:

$$F(x) = \begin{cases} \frac{1}{2}\int_{-\infty}^{x} \lambda e^{\lambda x} dx & x < 0 \\ \frac{1}{2}\int_{-\infty}^{0} \lambda e^{\lambda x} dx + \dfrac{1}{2}\int_{0}^{x} \lambda e^{-\lambda x} dx & x > 0 \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{\lambda x} & x < 0 \\ 1 - \frac{1}{2} e^{-\lambda x} & x > 0 \end{cases}$$

Example 5.5e. Let us reconsider Example 5.4e, which suppose that a binary message is to be transmitted from A to B, with the value 2 being sent when the message is 1 and -2 when it is 0. However, suppose now that rather than being a standard normal random variable, the channel noise N is a Laplacian random variable with parameter $\lambda = 1$. Again suppose that if R is the value received at location B, then the message is decoded as follows:

If $R \geq .5$, then 1 is concluded.

If $R < .5$, then 0 is concluded.

- The noise is Laplace with parameter $\lambda = 1$.

- The 2 types of errors will have probabilities given by

$$\begin{aligned} P\{\text{error} \mid \text{message 1 is sent}\} &= P\{N < -1.5\} \\ &= \frac{1}{2}e^{-1.5} \approx .1116 \\ P\{\text{error} \mid \text{message 0 is sent}\} &= P\{N \geq 2.5\} \\ &= \frac{1}{2}e^{-2.5} \approx .041 \end{aligned}$$

- The error probabilities are higher when the noise is Laplacian with $\lambda = 1$ than when it is a standard normal variable.

5.5.1 Hazard rate functions

- Consider a positive continuous random variable X that we interpret as being the lifetime of some item, having distribution function F and density f.
- The *hazard rate* (sometimes called the *failure rate*) function $\lambda(t)$ of F is defined by

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)} \qquad \overline{F} = 1 - F$$

- $$\begin{aligned} P\{X \in (t, t+dt) | X > t\} &= \frac{P\{X \in (t, t+dt), X > t\}}{P\{X > t\}} \\ &= \frac{P\{X \in (t, t+dt)\}}{P\{X > t\}} \\ &\approx \frac{f(t)}{\overline{F}(t)} dt \end{aligned}$$

- $\lambda(t)$ represents the conditional probability intensity that a t-unit-old item will fail.

- Suppose that the lifetime distribution is exponential. Then, by the memoryless property, it follows that the distribution of remaining life for a t-year-old item is the same as for a new item. Hence $\lambda(t)$ should be constant.

$$\begin{aligned} \lambda(t) &= \frac{f(t)}{\overline{F}(t)} \\ &= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} \\ &= \lambda \end{aligned}$$

The parameter λ is often referred to as the *rate* of the distribution.

- The failure rate function $\lambda(t)$ uniquely determines the distribution F.
- Note that by definition

$$\lambda(t) = \frac{\frac{d}{dt}F(t)}{1 - F(t)}$$

$$\log(1 - F(t)) = -\int_0^t \lambda(t)dt + k$$

or

$$1 - F(t) = e^k \exp\{-\int_0^t \lambda(t)dt\}$$

- Letting $t = 0$ shows that $k = 0$ and thus

$$F(t) = 1 - \exp\{-\int_0^t \lambda(t)dt\}$$

- If $\lambda(t) = a + bt$:
  - $F(t) = 1 - e^{-at-bt^2/2}$
  - $f(t) = (a + bt)e^{-at-bt^2/2}, \quad t \geq 0$
  - The Rayleigh density function if $a = 0$.

Example 5.5f. One often hears that the death rate of a person who smokes is, at each age, twice that of a nonsmoker. What does this mean? Does it mean that a nonsmoker has twice the probability of surviving a given number of years as does a smoker of the same age?

- $\lambda_s(t)$ denote the hazard rate of a smoker of age t and $\lambda_n(t)$ that of a nonsmoker of age t, then
$$\lambda_s(t) = 2\lambda_n(t)$$

- The probability that an A-year-old nonsmoker will survive until age B, $A < B$, is
$P\{A\text{-year-old nonsmoker reaches age } B\}$
$$\begin{aligned}
&= P\{\text{nonsmoker's lifetime} > B | \text{ nonsmoker's lifetime} > A\ \} \\
&= \frac{1 - F_{\text{non}}(B)}{1 - F_{\text{non}}(A)} \\
&= \frac{\exp\{-\int_0^B \lambda_n(t)dt\}}{\exp\{-\int_0^A \lambda_n(t)dt\}} \\
&= \exp\{-\int_A^B \lambda_n(t)dt\}
\end{aligned}$$
whereas the corresponding probability for a smoker is, by the same reasoning,

$$
\begin{aligned}
&P\{A\text{-year-old smoker reaches age } B\} \\
&\qquad = \exp\left\{-\textstyle\int_A^B \lambda_s(t)dt\right\} \\
&\qquad = \exp\left\{-2\textstyle\int_A^B \lambda_n(t)dt\right\} \\
&\qquad = \left[\exp\left\{-\textstyle\int_A^B \lambda_n(t)dt\right\}\right]^2
\end{aligned}
$$

- Two people of the same age, one of whom is a smoker and the other a nonsmoker, the probability that the smoker survives to any given age is the square of the corresponding probability for a nonsmoker.

5.6 Other continuous distributions

5.6.1 The Gamma distribution

- A random variable is said to have a gamma distribution with parameters (t, λ), $\lambda > 0$, and $t > 0$ if its density function is given by

$$
f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{t-1}}{\Gamma(t)} & x \geq 0 \\ 0 & x < 0 \end{cases}
$$

where $\Gamma(t)$, called the gamma function, is defined as

$$\Gamma(t) = \int_0^\infty e^{-y} y^{t-1} dy$$

- The integration by parts of $\Gamma(t)$,

$$\begin{aligned}\Gamma(t) &= -e^{-y} y^{t-1}|_0^\infty + \int_0^\infty e^{-y}(t-1)y^{t-2} dy \\ &= (t-1)\int_0^\infty e^{-y} y^{t-2} dy \\ &= (t-1)\Gamma(t-1)\end{aligned}$$

- For integral values of t, say $t = n$,

$$\begin{aligned}\Gamma(n) &= (n-1)\Gamma(n-1) \\ &= (n-1)(n-2)\Gamma(n-2) \\ &= \cdots \\ &= (n-1)(n-2)\cdots 3 \cdot 2\Gamma(1)\end{aligned}$$

- Since $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$, it follows that for integral values of n,

$$\Gamma(n) = (n-1)!$$

- If the events are occurring randomly in time and in accordance with three axioms of Sec. 4.8, then it turns out that the amount of time one has to wait until a total n events has occurred will be a gamma random variable with parameters (n, λ).

- T_n: The time at which the nth event occurs.

$$\begin{aligned} P\{T_n \leq t\} &= P\{N(t) \geq n\} \\ &= \sum_{j=n}^{\infty} P\{N(t) = j\} \\ &= \sum_{j=n}^{\infty} \frac{e^{-\lambda t}(\lambda t)^j}{j!} \end{aligned}$$

- The density function of T_n:

$$\begin{aligned} f(t) &= \sum_{j=n}^{\infty} \frac{e^{-\lambda t} j(\lambda t)^{j-1}\lambda}{j!} - \sum_{j=n}^{\infty} \frac{\lambda e^{-\lambda t} j(\lambda t)^j}{j!} \\ &= \sum_{j=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^j}{j!} \\ &= \frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

This distribution is often referred to in the literature as the n-Erlang distribution.

- If $n = 1$, it is $\exp(\lambda)$.
- If $\lambda = 1/2$ and $t = n/2$, it is χ_n^2.

- The chi-squared distribution often arises in practice as being the distribution of the error involved in attempting to hit a target in n dimensional space when each coordinate error is normally distributed.

Example 5.6a. Let X be a gamma random variable with parameters t and λ. Calculate (a) $E[X]$ and (b) $\mathrm{Var}(X)$.

- (a)

$$\begin{aligned} E[X] &= \frac{1}{\Gamma(t)} \int_0^\infty \lambda x e^{-\lambda x} (\lambda x)^{t-1} dx \\ &= \frac{1}{\lambda\Gamma(t)} \int_0^\infty \lambda e^{-\lambda x} (\lambda x)^{t} dx \\ &= \frac{\Gamma(t+1)}{\lambda\Gamma(t)} \\ &= \frac{t}{\lambda} \end{aligned}$$

(b) $E[X^2] = t(t+1)/\lambda^2$

$$\mathrm{Var}(X) = \frac{t}{\lambda^2}$$

5.6.2 The Weibull distribution

- The Weibull distribution is widely used in engineering practice due to its versatility.
- It was originally proposed for interpretation of fatigue data, but now its use has extended to many other engineering problems.
- It is widely used, in the field of life phenomena, as the distribution of the lifetime of some object.
- The Weibull distribution function:

$$F(x) = \begin{cases} 0 & x \leq v \\ 1 - \exp\{-(\frac{x-v}{\alpha})^{\beta}\} & x > v \end{cases} \tag{6.2}$$

- A random variable whose cumulative distribution function is given by Eq. (6.2) is said to be a Weibull random variable with parameters v, α, and β.

- Differentiation yields that the density is

$$f(x) = \begin{cases} 0 & x \le v \\ \frac{\beta}{\alpha}(\frac{x-v}{\alpha})^{\beta-1} - \exp\{-(\frac{x-v}{\alpha})^{\beta}\} & x > v \end{cases}$$

5.6.3 The Cauchy distribution

- A random variable is said to have a Cauchy distribution with parameter θ, $-\infty < \theta < \infty$, if its density is given by

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2} \qquad -\infty < \theta < \infty$$

Example 5.6b. Suppose that a narrow beam flashlight is spun around its center, which is located a unit distance from the x-axis (see Fig. 5.7). When the flashlight has stopped spinning, consider the point X at which the beam intersects the x-axis. (If the beam is not pointing toward the x-axis, repeat the experiment.)

- As indicated in Fig. 5.7, the point X is determined by the angle θ between the flashlight and the y-axis, which from the physical situation appears to be uniformly distributed between $-\pi/2$ and $\pi/2$.

- The distribution function of X is thus given by

$$\begin{aligned} F(x) &= P\{X \le x\} \\ &= P\{\tan\theta \le x\} \\ &= P\{\theta \le \tan^{-1} x\} \\ &= \frac{1}{2} + \frac{1}{\pi}\tan^{-1} x \end{aligned}$$

where the last equality follows since θ, being uniform over $(-\pi/2, \pi/2)$, yields that

$$P\{\theta \le a\} = \frac{a - (-\pi/2)}{\pi} = \frac{1}{2} + \frac{a}{\pi} \quad -\frac{\pi}{2} < a < \frac{\pi}{2}$$

- The density function of X is given by

$$f(x) = \frac{d}{dx}F(x) = \frac{1}{\pi(1+x^2)} \quad -\infty < x < \infty$$

- X has the Cauchy distribution.

5.6.4 The Beta distribution

- A random variable is said to have a beta distribution if its density is given by

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$$

- When $a = b$, the beta density is symmetric about $1/2$, giving more and more weight to regions about $1/2$ as the common value a increases.

- When $b > a$, the density is skewed to the left, and it is skewed to the right when $a > b$.

- The relationship between the beta function and the gamma function:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$E[X] = \frac{a}{a+b}$$

$$\mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

5.7 The distribution of a function of a random variable

- Suppose that we know the distribution of X and want to find the distribution of $g(X)$.
- To do so, it is necessary to express the event that $g(X) \leq y$ in terms of X being in some set.

Example 5.7a.

- Let X be uniformly distributed over $(0, 1)$. We obtain the distribution of the random variable Y, defined by $Y = X^n$, as follows: For $0 \leq y \leq 1$,

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{X^n \leq y\} \\ &= P\{X \leq y^{1/n}\} \end{aligned}$$

$$
\begin{aligned}
&= F_X(y^{1/n}) \\
&= y^{1/n}
\end{aligned}
$$

- The density function of Y is given by

$$
f_Y(y) = \begin{cases} \frac{1}{n} y^{1/n-1} & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}
$$

Example 5.7b.

- If X is a continuous random variable with probability density f_X, then the distribution of $Y = X^2$ is obtained as follows: For $y \geq 0$,

$$
\begin{aligned}
F_Y(y) &= P\{Y \leq y\} \\
&= P\{X^2 \leq y\} \\
&= P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y})
\end{aligned}
$$

- Differentiation yields

$$
f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})]
$$

Example 5.7c.

- If X has a probability density f_X, then $Y = |X|$ has a density function that is obtained as follows: For $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{|X| \leq y\} \\ &= P\{-y \leq X \leq y\} \\ &= F_X(y) - F_X(-y) \end{aligned}$$

- On differentiation,

$$f_Y(y) = f_X(y) + f_X(-y) \quad y \geq 0$$

Theorem 7.1: Let X be a continuous random variable having probability density function f_X. Suppose that $g(x)$ is a strictly monotone (increasing or decreasing), differentiable (and thus continuous) function of x. Then the random variable Y defined by $Y = g(X)$ has a probability density function given by

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)]|\frac{d}{dy}g^{-1}(y)| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

where $g^{-1}(y)$ is defined to equal that value of x such that $g(x) = y$.

Proof:

- When $g(x)$ is an increasing function.
- Suppose that $y = g(x)$ for some x. Then, with $Y = g(X)$

$$\begin{aligned} F_Y(y) &= P\{g(X) \leq y\} \\ &= P\{X \leq g^{-1}(y)\} \\ &= F_X(g^{-1}(y)) \end{aligned}$$

- Differentiation gives that

$$f_Y(y) = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y)$$

- When $y \neq g(x)$ for any x, then $F_Y(y)$ is either 0 or 1, and in either case $f_Y(y) = 0$.

Example 5.7d. Let X be a continuous non-negative random variable with density function f, and let $Y = X^n$. Find f_Y, the probability density function of Y.

- If $g(x) = x^n$, then

$$g^{-1}(y) = y^{1/n}$$

and

$$\frac{d}{dy}\{g^{-1}(y)\} = \frac{1}{n} y^{1/n-1}$$

- From Theorem 7.1,

$$f_Y(y) = \frac{1}{n} y^{1/n-1} f(y^{1/n})$$

- If $n = 2$,

$$f_Y(y) = \frac{1}{2\sqrt{y}} f(\sqrt{y})$$

which (since $X \geq 0$) is in agreement with the result of Example 5.7b.

Summary

- A random variable is called *continuous* if there is a nonnegative function f, called the *probability density function* of X, such that for any B
$$P\{X \in B\} = \int_B f(x)dx$$
- If X is continuous, then its distribution function F will be differentiable and
$$\frac{d}{dx}F(x) = f(x)$$
- Expected value of X:
$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$
- $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$
- $\mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$
- Uniform(a, b):
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

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

- $N(\mu, \sigma^2)$:

  $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$

  $\mu = E[X] \qquad \sigma^2 = \mathrm{Var}(X)$

- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.
- If $X \sim N(\mu, \sigma^2)$, then $Z = aX + b \sim N(a\mu + b, a^2\sigma^2)$.
- $\mathrm{Exp}(\lambda)$:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

  $E[X] = \frac{1}{\lambda} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$

- An exponential random variable has the *memoryless* property,

  $$P\{X > s + t \mid X > t\} = P\{X > s\}$$

- Hazard rate:

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \quad t \geq 0$$

- If F is the exponential distribution with parameter λ, then $\lambda(t) = \lambda$.
- Gamma(t, λ):

$$f(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{t-1}}{\Gamma(t)} \quad x \geq 0$$

$E[X] = \frac{t}{\lambda} \qquad \mathrm{Var}(X) = \frac{t}{\lambda^2}$

- Beta(a, b):

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \quad 0 \leq x \leq 1$$

$E[X] = \frac{a}{a+b} \qquad \mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$